INFORMATION SECURITY FOR SOUTH AFRICA

Proceedings of the ISSA 2008
Innovative Minds Conference

7 – 9 July 2008
School of Tourism & Hospitality
University of Johannesburg
Johannesburg
South Africa

Edited by
HS Venter, MM Eloff, JHP Eloff and L Labuschagne

Preface

PROCEEDINGS EDITORS: HS Venter, MM Eloff, JHP Eloff and L Labuschagne
PRODUCTION EDITOR: HS Venter
COVER DESIGNER: HS Venter

Publication Data
ISBN 978-1-86854-693-0
Editors: Hein Venter, Mariki Eloff, Jan Eloff and Les Labuschagne
Title of publication: Proceedings of the ISSA 2008 Innovative Minds Conference
Publisher: ISSA
Place of publication: Pretoria, South Africa
Year of publication: 2008
Edition: First Edition, First Impression

TABLE OF CONTENTS

Introduction

ISSA2008 is the annual conference for the information security community that continues on the successful recipe established in 2000. The upcoming conference is held under the auspices of the University of Johannesburg Business IT Department, the University of Pretoria School of Information Technology and the University of South Africa (Unisa) School of Computing.

The ISSA2008 Conference will run from Monday, 7 to Wednesday, 9 July at the University of Johannesburg's School of Tourism and Hospitality facility, in Auckland Park, Johannesburg, Gauteng, South Africa.

The conference has grown each year in various ways. Not only have delegate and presenter numbers been on the rise, but interest from industry has also grown and been displayed through sponsorship of the conference or aspects thereof. We believe that the quality and relevance of the information presented by industry practitioners and academics has also evolved over the years, as have the opportunities for senior research students to present their research to a critical and representative audience.

Conferences have become a major focus area - and often a money spinner - in many industries, so at any time you will see a number of conferences being advertised in fields such as information security. What sets the ISSA conference apart is that it is not intended to generate a profit for an organisation, and it does not encourage marketing of products and services through presentations. Instead, the proceeds from registration fees are reinvested to ensure that the conference grows each year. In exchange for their investment in the conference, sponsors are afforded an opportunity to present company-specific information that has a bearing on the conference themes, and presentations submitted by potential speakers are sent through a vigorous review process, managed by a team of respected international experts in information security.

We trust that the annual ISSA conference will continue to be recognised as an platform for professionals from industry as well as researchers to share

their knowledge, experience and research results in the field of information security.

To ensure ongoing improvement, we again encourage input from all those interested in the field of Information Security, particularly those who are actively seeking to progress the field, to take part and share their knowledge and experience.

We look forward to seeing old friends and new participants at ISSA2008.

Les Labuschagne

Conference Co-organiser

ISSA 2008 Conference Organisers:
Jan Eloff
Les Labuschagne
Mariki Eloff
Hein Venter
July 2008

Focus

Information security has evolved and in the last few years there has been renewed interest in the subject worldwide. This is evident from the many standards and certifications now available to guide security strategy. This has led to a more clear career path for security professionals.

The convergence of technologies together with advances in wireless communications, has meant new security challenges for the information security fraternity. As hotspots become more available, and more organisations attempt to rid their offices of "spaghetti" so the protection of data in these environments becomes a more important consideration.

It is this fraternity that organisations, governments and communities in general look to for guidance on best practice in this converging world.

Identity theft and phishing are ongoing concerns. What we are now finding is that security mechanisms have become so good and are generally implemented by companies wanting to adhere to good corporate governance, so attackers are now looking to the weak link in the chain, namely the individual user. It is far easier to attack them than attempt to penetrate sophisticated corporate systems. A spate of spyware is also doing the rounds, with waves of viruses still striking periodically. Software suppliers have started stepping up to protect their users and take some responsibility for security in general and not just for their own products.

The conference focuses on all aspects of information security and invites participation across the Information Security spectrum including but not being limited to functional, business, managerial, theoretical and technological issues.

Invited speakers will talk about the international trends in information security products, methodologies and management issues.

In the past ISSA has secured many highly acclaimed international speakers, including:

- **Alice Sturgeon** manages the area that is accountable for identifying and architecting horizontal requirements across the Government of Canada. Her topic made reference to An Identity Management Architecture for the Government of Canada

- **Dr Alf Zugenmaier**, DoCoMo Lab, Germany. His topic was based on Security and Privacy.
- **William List**, WM List and Co., UK. His topic was: Beyond the Seventh Layer live the users
- **Prof. Dennis Longley**, Queensland University of Technology, Australia. His topic was: IS Governance: Will it be effective?
- **Prof. TC Ting**: University of Connecticut, and fellow of the Computing Research Association, United States
- **Prof. Dr. Stephanie Teufel**: Director of the International Institute of Management in Telecommunications (iimt). Fribourg University, Switzerland
- **Rich Schiesser**, Senior Technical Planner at Option One Mortgage, USA
- **Rick Cudworth,** Partner, KPMG LLP, International Service Leader, Security and Business Continuity - Europe, Middle East and Africa

The purpose of the conference is to provide information security practitioners and researchers worldwide with the opportunity to share their knowledge and research results with their peers.

The objectives of the conference are defined as follows:

- Sharing of knowledge, experience and best practice
- Promoting networking and business opportunities
- Encouraging the research and study of information security
- Supporting the development of a professional information security community
- Assisting self development
- Providing a forum for education, knowledge transfer, professional development, and development of new skills
- Promoting best practice in information security and its application in Southern Africa
- Facilitating the meeting of diverse cultures to share and learn from each other in the quest for safer information systems

Conference Committees

General Conference Chairs
Jan Eloff (Department of Computer Science, University of Pretoria)
Les Labuschagne (Department of Business Information Technology, University of Johannesburg)
Mariki Eloff (School of Computing, University of South Africa)

Organising Committee
Adele da Veiga (KPMG, South Africa)
Jan Eloff (Department of Computer Science, University of Pretoria)
Les Labuschagne (Department of Business Information Technology, University of Johannesburg)
Mariki Eloff (School of Computing, University of South Africa)
Hein Venter (Department of Computer Science, University of Pretoria)
Taryn van Olden (Conference Management, Cyan Sky Communication)
Colette d'Hotman de Villiers (Conference Management, ISSA 2008)

Conference Programme Committee
Jan Eloff (Department of Computer Science, University of Pretoria)
Les Labuschagne (Department of Business Information Technology, University of Johannesburg)
Mariki Eloff (School of Computing, University of South Africa)
Hein Venter (Department of Computer Science, University of Pretoria)
Martin Olivier (Department of Computer Science, University of Pretoria)
Barry Irwin (Department of Computer Science, Rhodes University)
Marijke Coetzee (The Academy for Information Technology, University of Johannesburg)
Lynette Drevin (School of Computer, Statistical and Mathematical Sciences, North-West University)

Review Committee

Sampson Asare , Computer Science Department, University of Botswana, Botswana

Hettie Booysen , Comsec, South Africa

Reinhardt Botha , Nelson Mandela Metropolitan University, South Africa

Stelvio Cimato , University of Milano, Italy

Nathan Clarke , University of Plymouth, UK

Marijke Coetzee , University of Johannesburg, South Africa

Adele da Veiga, KPMG, South Africa

Mieso Denko , University of Guelph, Canada

Paul Dowland , University of Plymouth, UK

Lynette Drevin , North West University, South Africa

Jan Eloff, University of Pretoria, South Africa

Mariki Eloff , UNISA, South Africa

Sheikh Muhammad Farhan , University of Engineering and Technology, Taxila, Pakistan

Eduardo Fernandez , Florida Atlantic University, USA

Simone Fischer-Hübner, Karlstad University, Sweden

Dario Forte , University of Milano, Italy

Steven Furnell , University of Plymouth, UK

Anna Granova , University of Pretoria, South Africa

Stefanos Gritzalis , University of the Aegean, Greece

Ajantha Herath , Richard Stockton College of New Jersey Pomona, USA

Suvineetha Herath , Richard Stockton College of New Jersey Pomona, USA

Barry Irwin , Rhodes University, South Africa

Christian Damsgaard Jensen , Technical University of Denmark, Denmark

Karthik Kannan , Purdue University, Indiana, USA

Sokratis Katsikas , University of the Aegean, Greece

Les Labuschagne , University of Johannesburg, South Africa

Dennis Longley , Queensland University of Technology, Australia

Marianne Loock , UNISA, South Africa

Günter Müller, University of Freiburg, Germany

Martin Olivier , University of Pretoria, South Africa

Jacques Ophoff , Nelson Mandela Metropolitan University, South Africa

Rolf Oppliger , eSECURITY Technologies, Switzerland

Mauricio Papa , University of Tulsa, USA

Guenther Pernul , University of Regensburg, Germany

Stephen Perelson , Nelson Mandela Metropolitan University, South Africa, South Africa

Tana Pistorius , UNISA, South Africa

Dalenca Pottas , Nelson Mandela Metropolitan University, South Africa

M Mohsin Rahmatullah, University of Engineering and Technology, Taxila, Pakistan

Pierangela Samarati, University of Milano, Italy

Mikko Siponen , University of Oulu, Finland

Elme Smith , UNISA, South Africa

Graham Teare, KPMG, South Africa

Stephanie Teufel, University of Fribourg, Switzerland

Review Process

The Review Process was undertaken by experienced and well respected Information Security experts. In a blind peer-review process full papers were scrutinised by an international panel of reviewers. The reviewers were asked to provide specific feedback and comments to authors. This feedback was provided to give a perspective on how a paper can be improved for final submission and inclusion in this - the formal conference proceedings.

A 'Call for Papers' was issued towards the end of 2007, inviting anyone interested in making a contribution towards the conference by submitting a short abstract by the end of March 2008. Abstracts were received and subsequently divided into broad topics by the Programme Committee. The abstracts, within a broad field, were forwarded to a review panel in the field to judge on the possible acceptability of the abstract based upon the scope and depth of the subject matter to the conference as a whole. The authors were then requested to submit full papers by the end of April 2008. These draft papers were "anonomised", and then forwarded to two independent reviewers, with the request that the full paper should be reviewed and judged according to a number of criteria. Reviewers were asked to use a 10 point Likert scale to rate the following criteria:

- Originality
- Significance
- Technical Quality
- Relevance

Reviewers were also asked to give an Overall Rating as well as a Confidence in Rating for each the paper. In the next section, reviewers had to qualify their rating by providing a rationale for the Overall Rating given. This was followed by the Reviewer Comments that would assist the authors in improving and correcting their papers. Reviewers were asked to be as comprehensive as possible in this section.

The Programme Committee received the completed review forms from the Reviewers and combined the scores from the reviewers for each paper to determine a whether they would be accepted or not. Only papers with a combined value above a certain threshold were accepted as full papers. In the event where two reviewers differed drastically from one another, the paper was sent to a third reviewer.

The reviewers' comments were forwarded to the author with the request to submit a final revised version of the paper by May 2008. Only those papers which were of an acceptable quality as recommended by both Reviewers are included in the Conference Proceedings as Reviewed Papers.

The review process used is based on what is considered the international de facto standard for blind paper reviews.

Conference Sponsors and Organisers

Sponsors

KPMG

Transformation is central to our values and a fundamental driver of almost all that we do at KPMG. We strongly believe in living our values and our vision for transformation focuses on the full spectrum of empowerment: people (equal opportunities), clients (value creation), and communities (KPMG stakeholders). In addition, we are proud of being South Africans, successfully operating in the global market place. KPMG has transformed itself into a global organisation, a truly multi-disciplinary firm known for its innovative practices. It is at the forefront in delivering industry specific service offerings, ideas, insight and understanding of complex client needs.

Visit the website at: www.kpmg.co.za.

Conference Organisers

University of Johannesburg Department of Business IT

The Department of Business IT has extensive resources to facilitate excellence in terms of the graduate and post-graduate courses on offer. These include the lecturing staff, as well as computer laboratories, libraries and constant exposure to industry through various projects. We offer comprehensive qualifications that include vocational and academic direction of study. The main research focus areas include information security management, IT project management and advanced application paradigms.

Visit the Department's website at: www.uj.ac.za/bit

University of Pretoria Department of Computer Science

The University of Pretoria Computer Science Department offers opportunities for studies in BSc Information Technology and BSc Computer Science on under- and post-graduate levels. These comprehensive, industry-relevant courses cover a number of the facets of information technology.

Visit the website at: www.cs.up.ac.za

University of South Africa School of Computing

The School of Computing at Unisa offers comprehensive tertiary distance education in the computer-related fields. These offerings are available to any student from any part of the world. Students can enrol for a wide range of qualifications from single modules certificates, through to obtaining PhDs in the fields of Computer Science, Information Systems, or Information Technology. The School consists of four departments: Theoretical Computing, Software Development, Software Construction and Systems Organisation. It has access to an excellent range of resources to facilitate both education and research. The specialist areas of research are Information Security, Human-Computer Interaction, Computer Ethics, Theoretical Computing, and Electronic Education Technologies and Strategies.

The School's website can be found at: www.cs.unisa.ac.za

PART 1

REVIEWED RESEARCH PAPERS

PASSWORD MANAGEMENT: EMPIRICAL RESULTS FROM A RSA AND USA STUDY

HA Kruger[*], T Steyn[*], L Drevin[*], BD Medlin[#]

[*]School of Computer, Statistical and Mathematical Sciences
North-West University (Potchefstroom Campus)

[#]Computer Information Systems Department
Appalachian State University

[*]Hennie.Kruger@nwu.ac.za, Tjaart.Steyn@nwu.ac.za,
Lynette.Drevin@nwu.ac.za
Private Bag X6001, Potchefstroom, 2520, South Africa
+27 18 299 2531

[#]Medlinbd@appstate.edu
Boone, NC, USA, 28607
828 262 2411

ABSTRACT

"The state of information security as a whole is a disaster, a train wreck". This view is given by Forte and Power (2007) describing the state of information security towards the end of the first decade of the 21st century. Amongst solutions offered, the view that security programs have to be holistic is proposed indicating that technical controls are of little value without the workforce understanding the risks of their irresponsible behavior. Another solution proposed by them is the role of awareness and education. All levels of users should be targeted letting them understand their role and responsibility in information security. Password related behavior is often highlighted as a key component of information security

awareness. However, studies have shown that password hygiene is generally poor amongst users (Stanton, Stam, Mastrangelo, & Jolton, 2005).

In an effort to identify, categorize and prioritize those factors that may have a significant impact on password behavior, a study was conducted amongst students in South Africa and the United States of America to investigate certain aspects of password management practices. The objective of this paper is to report on the empirical results obtained, using techniques such as cause-and-effect diagrams and Pareto analyses.

KEY WORDS

Password management, ICT Security Awareness, Cause-and-Effect diagrams, Pareto analyses.

PASSWORD MANAGEMENT: EMPIRICAL RESULTS FROM A RSA AND USA STUDY

1 INTRODUCTION

Information security has become an important issue in modern organizations. However, not everybody is convinced that information security and the associated measures implemented are producing the expected results. Forte and Power (2007) states that "the state of information security as a whole is a disaster, a train wreck". They argued that technical controls are of little value without the workforce understanding the risks of their irresponsible behavior and advocate that the role of education and awareness programs should receive more attention.

A project to evaluate security awareness levels of staff was initiated in 2005 and consists firstly of an identification phase where key areas, on which measurements can be taken, were identified. Secondly, knowledge, attitude and behavior of staff may be surveyed to determine their awareness levels pertaining to the identified areas. The assessment of appropriate system generated data also forms part of this phase. Finally, the data may then be used to construct a model that may assist in improving the overall information security culture. These phases are described in detail in Kruger, Drevin and Steyn (2006). During the identification phase, the effectiveness of password management has emerged as an issue that should be evaluated. This is in line with the fact that password related behavior is often highlighted as a key component in security programs. However, studies have shown that password behavior is generally poor amongst users (Stanton, Stam, Mastrangelo & Jolton, 2005). The verification of awareness levels that relate to the effective use of passwords would assist in covering some of the main objectives of any security program e.g. the integrity and confidentiality of data.

This paper reports on the use of cause-and-effect diagrams to identify significant causes of poor password management behavior and to assist in the prioritization of the identified causes, Pareto analyses were used. The

study was conducted amongst students in South Africa and the United States of America and some comparative results and statistics will be presented.

The remainder of the paper is organized as follows. In the next section the methodology used is briefly presented. Section 3 discusses the results obtained with some concluding remarks in the last section.

2 METHODOLOGY

In this study, the effectiveness of password management was described in terms of two categories – *secure* passwords and *confidentiality* of passwords. Both these categories are defined by different criteria e.g. secure passwords may be defined by password length (Pfleeger and Pfleeger, 2007), how regular passwords are changed (Furnell, 2007), etc., while confidentiality may be defined by criteria such as making passwords available to others – by writing it down or telling someone (Pfleeger and Pfleeger, 2007), the use of different passwords for different systems (Furnell, 2007), etc.

To assist in understanding and identifying problems associated with ineffective password management, two cause-and-effect diagrams were constructed for the two categories. A cause-and-effect diagram is a tool that can be used to represent the relationship between some effect that could be measured and the set of possible causes that produce the effect (Berenson and Levine, 1996). The diagrams are constructed by showing the effect or problem on the right hand side of the diagram and the major causes listed on the left hand side. The causes may also be subdivided into a few major categories depending on the problem under investigation. Following a comprehensive process that included literature surveys, brain storming sessions and pilot studies, a list of 23 causes were identified relevant to secure passwords and the confidentiality of passwords. These causes were grouped into main categories with the help of validation techniques such as content validation, reliability tests and construct validation. The final result was a 5-factor instrument (questionnaire) consisting of 23 items derived from the 23 causes for the two categories studied and was defined as follows:

Secure Passwords	**Confidentiality of Passwords**
Attitude/viewpoint – measured by 3 different items	Attitude/viewpoint – measured by 3 different items
Knowledge and Resources – measured by 4 different items	Knowledge and Resources – measured by 3 different items
Expectation and Feedback – measured by 2 different items	Expectation and Feedback – measured by 2 different items
Skills – measured by 1 item	Knowledge related behavior – measured by 1 item
Own perception of behavior – measured by 2 different items	Own perception of behavior – measured by 2 different items

The complete process covering the construction of the cause-and-effect diagrams, the development of the measuring instrument and reliability test results can be found in Kruger, Drevin and Steyn (2008).

3 APPLICATION AND RESULTS

Using the measuring instrument described in section 2, an empirical experiment was conducted at two universities, one in South Africa and the other in the USA, to see how students apply password management principles. A significant user base of students exists at universities and there are a large number of confidential and privacy security issues associated with student users that can directly be linked to passwords and the management of passwords. As with other users, students should be prohibited from accessing systems where test and examination marks can be changed; test and examination papers can be accessed before student assessments take place; or, where fraudulent actions such as altering of financial data can be done. By not keeping a password confidential or making use of passwords that can easily be guessed, considerable financial losses can be incurred by students e.g. when somebody else uses the password to download large files from the Internet. Irregularities during examinations and tests that are done on computers are also likely when students can access other students' work. Apart from the usual dishonest

behavior that should be avoided, it seems to be appropriate to assess the password management knowledge and attitude of young people. They are the business and ICT leaders of the future and should be made aware of the risks and consequences of poor password management.

A simple web application was used to make the questionnaire available to students at the two universities. Although a total of 507 responses were received it was decided to use only those with the field of study in natural sciences and economic and management sciences. In addition, only students in their 3rd year of study or higher were considered. The reason for this selection was to try and ensure that a homogeneous group of students are used to compare the results between the two universities. The final comparison was therefore performed on 193 responses of which 93 were from the South African university and 100 from the university in the USA.

The final results were presented as Pareto charts. A Pareto chart or diagram is a graphical representation in the form of a bar graph that is used to arrange information in such a way that priorities and relative importance of data can be established. It is often used by managers to direct efforts to the biggest improvement opportunity by highlighting the vital few causes in contrast to the trivial many (Pareto diagram, 2007). The charts are constructed by arranging the bars in decreasing order from left to right along the x-axis. Cumulative percentages are then used to assist in analyzing the chart.

Figure 1 contains the Pareto charts for the two universities for the main factors relevant to secure passwords, while figure 2 presents the charts for the confidentiality of passwords. It can be seen from figure 1 that the order of the main factors relevant to secure passwords, is the same for both universities with the factor *Expectation and Feedback* the most significant. This factor was measured by two items – *secure passwords are not compulsory* and *secure passwords are not important*. Looking at figure 2, it is clear that *Expectation and Feedback* – measured by *confidentiality of passwords is not compulsory* and *confidentiality of passwords is not important* is once again the biggest concern when dealing with confidentiality of passwords. Based on this it must be accepted that the current message (feedback) that students, at both universities, receive from management, lecturers, their environment, their peers, etc. is that the use of secure passwords as well as the confidentiality of passwords are not really

important and also not compulsory – it is not really expected from them to use secure passwords or to keep their passwords confidential and compliance of this will not be verified.

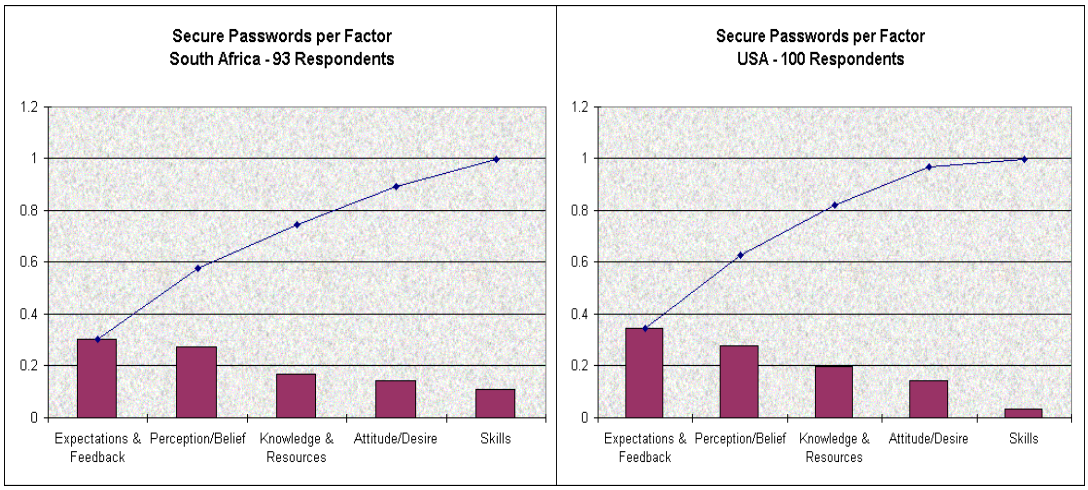


Figure 1 – Pareto charts for secure passwords

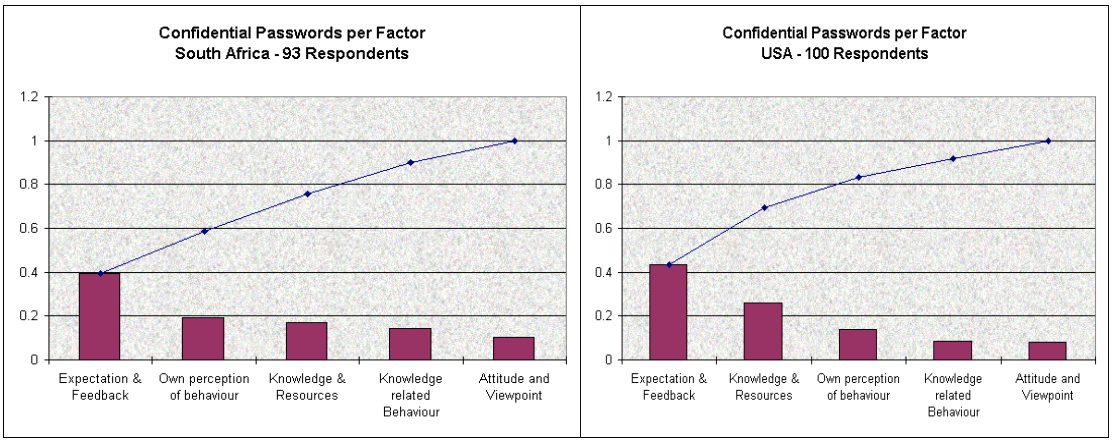


Figure 2 – Pareto charts for confidential passwords

It should also be noted from the Pareto charts that by addressing only the first factor (Expectations and Feedback) about 40% of problems related

to confidentiality of passwords at both universities can be solved. In the case of secure passwords, more than 30% of problems at the two universities can be solved. These facts from the Pareto charts create a perfect opportunity for management to address specific password management issues instead of implementing, for example, a comprehensive and expensive awareness program. Each one of the factors, and the items related to them, was analyzed in a similar way but are not presented here.

Two other interesting observations based on the responses suggest that in some cases students believe that they are complying with good password management principles but their behavior may indicate differently. In the first instance, one of the questions explicitly asked respondents whether they belief that the passwords they are using are secure passwords. Almost half of the respondents (46%) in South Africa stated that they use secure passwords. When checking their passwords and applying two basic rules concerning password length and the use of different character sets, it was found however, that 63% of those who said that they are using secure passwords have passwords with 6, or less, characters or make use of only one character set. At the university in the USA, the figure was much lower at 14% who said that they use secure passwords of which 29% had weak passwords when measured against the same two rules. Another concern when looking at these statistics is that 54% of students at the South African university and an alarming 86% of students at the university in the USA stated (admit) that they do not use secure passwords.

Secondly, another question asked respondents whether they believe that they are keeping their passwords confidential. At the South African university 76% answered that they do keep their password confidential but 15% of the same students also indicated that they would make their passwords available to others when needed. At the university in the USA 77% said that they keep their passwords confidential and only 5% of them would give their passwords to somebody else. These two findings are in line with a similar result described by Albrechtsen (2007). According to Albrechtsen's study users stated that although information security is important, they are not always able to point out practical security actions with which they contribute to information security – basically they are not aware of what they could or should do. This is probably true in this case as well. Students may view passwords as an important issue and they may

believe that they are using secure passwords but they are not always aware of the practical requirements such as password length.

Table 1 lists some other interesting issues when analyzing each factor and is based on the frequency of answers received from students.

Table 1 – Additional findings

	Universities	
Item	**South Africa**	**USA**
I use simple passwords so that it can easily be remembered	55% admit that they use simple passwords	50% admit that they use simple passwords
I know where to get help or information regarding *secure* passwords	37% do not know where to get help	61% do not know where to get help
I know where to get help or information regarding the *confidentiality* of passwords	41% do not know where to get help	61% do not know where to get help
I can define (or explain) the concept "confidentiality of passwords"	15% admit that they cannot explain the concept	33% admit that they cannot explain the concept

In general the overall results revealed the following. The most significant issues, according to the Pareto charts, and to which students should be made aware of include aspects such as:

- Proper use of passwords which include the use of secure passwords and keeping passwords confidential *is compulsory*.
- Passwords are an extremely *important* aspect of ICT security and improper use will degrade the quality of security and increase the probability of a number of security risks.
- The use of simple passwords that can easily be remembered is not acceptable.
- Making passwords available to other people is not allowed.

– Where to get help or information on proper password principles.

Addressing these few simple principles would solve on average more than 60% of the problems related to effective password management. The remaining factors and their associated items can be evaluated in the same manner and simultaneously, or in a follow-up exercise, be addressed. On the positive side of the scale it appears as if students have the necessary skills e.g. they know where and how to physically change passwords; they generally have a positive attitude or viewpoint towards effective password management e.g. they think that it is worthwhile to use secure and confidential passwords and they do not claim that they are too busy to concern themselves with secure and confidential passwords. They also agree in general that passwords should be kept confidential.

4 CONCLUSION

This paper presented a study where cause-and-effect diagrams were used to assist in evaluating password management practices amongst students at two universities – one in South Africa and the other in the USA. Pareto analyses were then used to identify and prioritize significant aspects. Results indicated that students at both universities do not regard the use of secure passwords, or keeping their passwords confidential, as an important aspect; they did not experience it as being compulsory; and, most of them would use simple passwords that can easily be remembered.

The use of cause-and-effect diagrams and the Pareto analyses proved to be extremely helpful in understanding and gaining insight into those factors that have a significant impact on the effectiveness of password management. Results obtained also created an opportunity for directed security awareness programs where efforts can be focused on specific important issues instead of conducting the usual comprehensive programs where aspects that may not be significant are also addressed.

5 ACKNOWLEDGEMENT

This paper is based upon work that is financially supported by the National Research Foundation (NRF). The opinions expressed in this paper are those of the authors.

6 REFERENCES

Albrechtsen, E. 2007. A qualitative study of users' view on information security. *Computers & Security*, 26:276-289.

Berenson, M.L. & Levine, D.M. 1996. *Basic Business Statistics. Concepts and Applications.* Sixth edition. Upper Saddle River, NJ: Prentice Hall.

Forte, D. & Power, R. 2007. The state of information security towards the close of the first decade of the 21st century. *Computer Fraud & Security*, October, 2007.

Furnell, S. 2007. An assessment of website password practices. *Computers & Security*, 26:445-451.

Kruger, H.A., Drevin, L. & Steyn, T. 2006. A framework for evaluating ICT security awareness, *In: Proceedings of the 2006 ISSA Conference, Johannesburg, South Africa,* 5-7 July 2006 (on CD).

Kruger, H.A., Drevin, L. & Steyn, T. 2008. Password management assessment. Technical Report. North-West University, South Africa, FABWI-N-RKW:2008-222**.**

Pareto Diagram. 2007. [Web:] http://mot.vuse.vanderbilt.edu/mt322/Pareto.htm [Date of use: July 2007].

Pfleeger, C.P. & Pfleeger, S.L. 2007. *Security in Computing.* Fourth edition. Prentice Hall.

Stanton, J.M., Stam, K.R., Mastrangelo, P. & Jolton, J. 2005. Analysis of end user security behaviors. *Computers & Security*, 24(2):124-133.

A USER CENTRIC MODEL FOR ONLINE IDENTITY AND ACCESS MANAGEMENT

[1]M. Deas and [2]S. Flowerday

Nelson Mandela Metropolitan University
University of Fort Hare

[1]mbdeas@gmail.com
Tel: +27(0)43 7047071
P.O. Box 15520, Beacon Bay, East London, 5205

[2]sflowerday@ufh.ac.za

ABSTRACT

The problem today is that users are expected to remember multiple user names and passwords for different domains when accessing the Internet. Identity management solutions seek to solve this problem by creating a digital identity that is exchangeable across organisational boundaries. This is done through the setup of collaboration agreements between multiple domains, thus users can easily switch across domains without having to repeatedly sign-on. However, this technology is accompanied by the threat of user identity and personal information being 'stolen'. Criminals make use of fake or 'spoofed' websites as well as social engineering techniques to gain illegal access to a user's information. This problem has been catapulted to the fore by the statement that phishing has increased by 8000% over the period January 2005 to September 2006 (APACS, 2007). Thus, the need for user protection from online threats has drastically increased. This paper examines two processes to protect user login information. Firstly, user's information must be protected at the time of sign-on, and secondly, a

simple method for the identification of the website is required by the user. This paper looks at these processes of identifying and verifying user information followed by how the user can verify the website at sign-on. The roles of identity and access management are defined within the context of single sign-on. Three different models for identity management are analysed, namely the Microsoft .NET Passport, Liberty Alliance Federated Identity for Single Sign-on and the Mozilla TrustBar for website authentication. A new model for the definitive protection of the user in the online environment is proposed based on the evaluation of these three existing models.

KEY WORDS

Identity Management, Authentication Management, Mozilla TrustBar, Liberty Alliance, .NET Passport

A USER CENTRIC MODEL FOR ONLINE IDENTITY AND ACCESS MANAGEMENT

1 INTRODUCTION

The Internet has played a major role in the way people do business and interact socially. Websites are used to sell goods and services online whilst storing sensitive customer information such as credit card details and identity numbers. This information is regularly stored using simplistic user sign-on tools. The use of this technology creates the challenge of how to ensure that the correct authorised user connects to the appropriate online system.

To ensure users are who they claim to be at the time of sign-on, a more advanced authentication tool than that of a single key authentication password, is required. Through the use of dual key authentication over that of single key passwords, higher levels of trust between the user and the website provider are created. A number of users are still naïve as to the potential dangers of the Internet and are unaware that they may be at risk by using websites with simple security measures for client authentication. The threat exists for criminals to make use of fake or 'spoofed' websites and social engineering techniques to gain illegal access to user information and potentially commit identity theft. Although organisations have been set up to standardise the processes of online identity management, none fully protect the user and enforce a dual method of user and website authentication. Because of this the risk still exists that a user's account information can be accessed illegally. It is therefore important that adequate identity management controls are put in place to secure the online user.

The remainder of this paper is organized as follows: Section 2 presents the role of identity management for businesses as a tool to meet legal requirements for client protection and the benefits of identity management to the business. Section 3 investigates the role of access and authentication management, focusing on user issues and trends relating to online systems usage. Section 4 provides an overview of the identity management models implemented by Microsoft Passport .NET, Liberty

Alliance Federated User Identity and the Mozilla TrustBar. Section 5 provides a critical comparison of the models. However, none of the investigated models focus on the issues of the user and the protection of the user within the online environment. Each model focuses exclusively on the sign-on or website identification processes and lacks a wholesome environment within which the user may interact with. Section 6 proposes a model for user centric online protection. This model is based on the use of dual authentication techniques in the form of user authentication by the system, and system authentication by the user.

2 ROLE OF IDENTITY MANAGEMENT

Through the use of identity management, businesses benefit as they draw from best practices and ensure compliance to regulations. Legal requirements for client protection are implemented to provide a code of "best practice" as noted in COBIT and ITIL (Lewis, 2003). The business is ultimately responsible for the use of identity information and is held accountable should that information be used fraudulently. In making use of the identity management life-cycle, the user's account is managed from the time of creation to the time when the user permanently leaves the system. This process includes the removal and addition of system rights (De Leeuw, 2004). Through efficient use of an identity management solution, companies realise the following benefits:

1. Better planning, implementation and management of solutions through a complete user based life-cycle of services.
2. Reduction in costs and complexity, while increasing the rate of return on investment made in identity management.
3. Predictable implementation procedures and efficient business operations, thereby ensuring greater system satisfaction for both users and customers.
4. Manages all four main areas of concern for the business (people, process, practice and platform), when implementing identity management in the organisation (Sun Microsystems, n.d.; Gordon, 2004).

Organisations now view identity management solutions as the answer to a number of security challenges. It is also imperative for organisations to consider how they can take full advantage of the benefits and the value of an identity management system within their business (BMC Software, 2006). Furthermore, the use of effective identity management controls will provide the system user with a secure environment within which they can function. The effectiveness of such a process, however, is only as strong as the level to which access and authentication management controls are applied.

3 ACCESS AND AUTHENTICATION MANAGEMENT

In order to manage a business environment in which multiple users require access to systems over large and distributed networks becomes difficult plus it is essential that the business ensures the users connecting to this environment are whom they claim to be. The Internet has the ability to mask an identity, and this process can be used to perpetrate fraud. Therefore, every action performed online is subject to a degree of risk. This lack of trust has spread into the banking sector. In a recent report by the journal, *Computer Fraud and Security*, it was stated that 52% of respondents were unlikely to sign up to online banking facilities and that 82% of respondents would not respond to any emails from financial firms (Consumers losing trust in online banking: survey, 2007).

In online commerce, customers take on substantial levels of risk when making purchases from an online vendor, because all encounters take place through the vendor website. Customers therefore need to be able to assess the risk involved when purchasing online.

Customers often leave a website when they do not gain a sufficient sense of trust (Chau, Hu, Lee & Au, 2006). Online merchants store large amounts of customer data therefore it is critical for vendors to build strong trusting relationships with their customers. This can be ensured by making use of proper access control procedures to provide minimal risk to the customer. From the perspective of the merchant, there is little concern over the identity of the individual customer, but more concern over their ability to pay for services or goods. If security is breached on the vendor website, it is imperative that accurate logs exist for the auditing of user actions.

The dangers to users in the online environment are summarised as spoofing, phishing and identity theft. By implementing strong controls to ensure that only an authorised user accesses the system, the business risk is

diminished (Rodger, 2004). In order to identify the best methods to protect an online identity from online threats, it is essential to look at international systems for single sign-on (SSO) protection of the user.

4 COMPARISON OF IDENTITY MANAGEMENT SOLUTIONS

The ideal environment for the computer scientist is one in which computer systems know who their users are. The ideology behind this is based on the concept that users should be authenticated as simply as possible. An investigation is performed to determine the best method of implementation, specifically looking at Microsoft's Passport .NET, Liberty Alliance and Mozilla TrustBar.

4.1 Microsoft Passport .NET

The Passport .NET service makes use of SSO Identity Domain. Microsoft is suited to the process of handling an SSO platform as it already provides a large variety of services online for e-mail, online messaging and search facilities. However, issues regarding user privacy and freedom of movement online could be infringed should a single entity take control of all SSO authentications and the information held therein. The process followed for user authentication through the Passport service is shown in Figure 1.

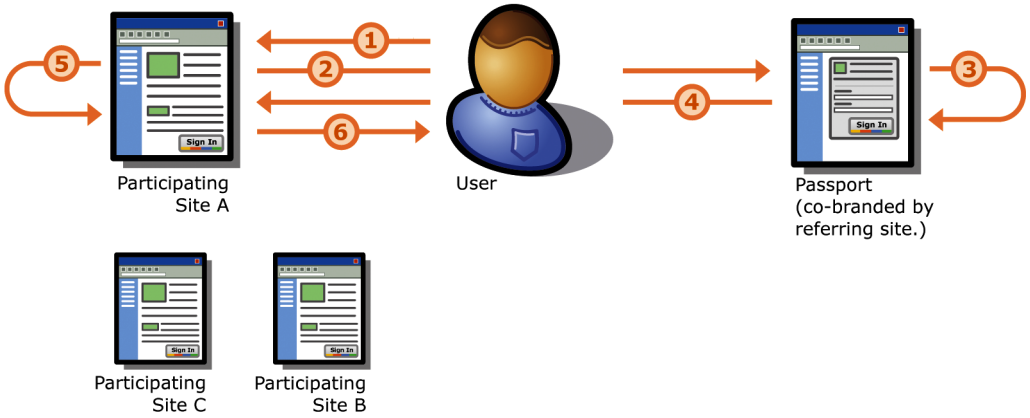


Figure 1 – The Passport Authentication Process (Microsoft, 2004)

1. User browses to participating site or service (Site A in this example) and clicks 'Sign In' button or link.
2. User is redirected to Passport.

3. Passport checks if the user has a 'Ticket Granting Cookie' (TGC) in their browser's cookie file meeting the rules of Site A. If one is detected, they skip to step 4 and do not go through the login process. If the TGC has lapsed based on Site A's time requirements, then the user is redirected to a page asking for their login credentials to be entered correctly in order to proceed.
4. The user is redirected back to Site A with their encrypted authentication ticket and profile information attached.
5. Site A decrypts the authentication ticket and profile information and signs the customer into the website.
6. The user accesses the page, resource or service they requested from Site A.

In concluding Figure 1, no information about a user is shared with Sites B and C unless the user chooses to sign-on to those sites.

A potentially hazardous feature to the user of Passport .NET reported by both Microsoft (2004) and discussed by Kormann and Rubin (2000) is that of the automatic sign-on to Passport. If this option is selected, the username and password of the individual user are stored locally on the individual client's machine. When an automatic sign-on is selected the user will be signed on to the .NET Passport service without intervention. Disconnecting from the Internet or turning the machine off has no effect on the connection of the user to the service. This option exposes a user's account to infiltration potentially exposing sensitive information.

Although a user may use their .NET Passport account at multiple sites, the password is only stored in the .NET Passport database and is only shared with the .NET Passport servers that need to make use of it for authentication. The .NET Passport service contains a feature that, should the user make an error in attempting to sign-on, the system automatically blocks access to the user account for a few minutes. This process stops attempts to gain unlawful access to an account using password cracking software.

Overall, the .NET Passport solution provides a relatively simple solution to the problems experienced by users within SSO. Websites affiliated with the .NET Passport program can opt to have the service manage their user base, shifting the responsibility for this process from

themselves to Microsoft. As previously stated, the main drawback to the .NET Passport solution is the problem of having a single entity responsible for and controlling all identity authentication tasks. This, by itself, increases new risks and issues relating to both privacy and security.

4.2 Liberty Alliance Federated User Identity

The Liberty Alliance is an undertaking by a group of organisations and government agencies to provide a set of open technical specifications for the creation of a federated identity solution. When the Liberty Alliance began their operations, the first phase of development involved the setting up of specifications which enabled simplified SSO for end users. This process became Liberty's Identity Federation Framework (Madsen, 2004).

The Liberty specifications for SSO includes an enabler which provides SSO functionality across different enterprise domains and websites. Pfitzmann (2004) describes the process of Liberty's SSO as follows:

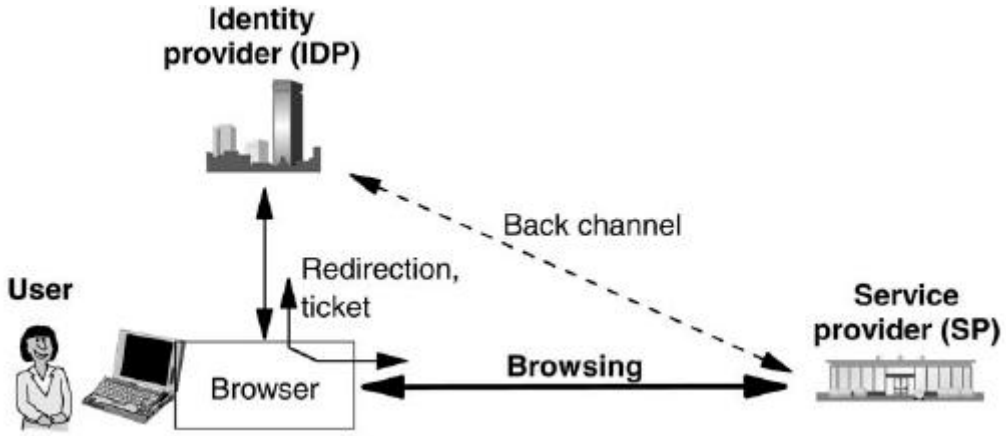


Figure 2 – Browser based SSO (Pfitzmann, 2004)

In Figure 2, a user accesses the service provider whilst browsing online. When submitting a sign-on request, the service provider redirects the browser to the user's identity provider. The user then logs in using a typical username and password. The identity provider redirects the browser back to the service provider with an additional ticket to handle other services, such as data transfer logistics on other channels.

The benefit of using this form of implementation is that the user is not redirected to a separate login page for authentication purposes. Once the user is authenticated by one service within the 'circle of trust', all other websites within that trust domain can verify the user as having been authenticated, eliminating the need for multiple sign-on (Liberty Alliance, 2007).

The Liberty Alliance provides a viable alternative to the solution from the .NET Passport. By having a consortium of companies involved in the setting of standards, a broader level of consensus is achieved and the best solution implemented. Within the realm of SSO the problem however with Liberty Alliance's solution is the lack of ability to provide a scalable solution for a user to connect to a multitude of websites, because each setup of the federated identity solution exists within separate circles of trust. If a user moves between two different trust circles, they will be required to sign-on with different credentials.

4.3 Mozilla TrustBar

A potential solution for the identification of websites is the use of a plug-in toolbar supplied within Mozilla Firefox browsers. By using this plug-in a user can store images mapped to server certificates. Whenever a server certificate is verified, the mapped image is displayed on the toolbar, while the corresponding page is still being loaded (Jøsang & Pope, 2005). Mozilla TrustBar focuses on how to secure the user whilst authenticating websites. Thus, the user is protected from the threats of phishing and website spoofing. The TrustBar attempts to make users more aware of the security behind the web pages they view.

The overall process of the client user authentication on the server attempts to protect against potential eavesdropping and modification by Man in the Middle (MITM) adversaries. Large numbers of financial and other websites make use of Secure Socket Layer (SSL) to authenticate the user. A number of those sites however only make use of SSL protocols once the user has typed in a username and password and then clicked 'submit' (Herzberg, 2005).

This form of implementation has the potential for MITM to redirect the user towards a modified version of the website. Should this occur the user may unknowingly provide login information to a third party. Through the modified page, if the user attempts to login, the user information is sent

back to the MITM. Clearly the traditional approach of signing on does not protect the user from these forms of attack; and the user requires an easier way to verify that he or she is on the intended website.

Herzberg (2005) mentions the ways in which TrustBar provides a solution to the client user problem as follows:

- TrustBar periodically downloads a list of the unprotected websites that are maintained on the Mozilla TrustBar servers. This list stores the unprotected login sites which Mozilla tracks, as well as any alternate login pages for those websites that are protected. This information can be used to redirect the client if an unsafe link is found.
- TrustBar makes allowance for users to assign a logo to websites of their own choosing to visually identify the website. TrustBar tracks changes to websites and displays information in the form of a "Same since" and a date value. After the website changes, a warning is displayed when the page is accessed by the user.

Herzberg and Gbara (2007) provide further uses of the TrustBar for solving the user problem in the following situations:

- In SSL websites, TrustBar shows by default, the name of the organisation that owns the website through the identification of the digital certificate. TrustBar also displays a representation of the logo or the name of the certification authority which issued the certificate.
- TrustBar displays a padlock for all protected websites, and a "No Entry" sign for unprotected websites.

The Mozilla TrustBar's solution to the user's web usage condition is novel. The service provides the client with a free-to-use facility that can make the online user feel more secure. Through providing a visual aid to the user showing the current status of the accessed website, the user's overall experience is improved.

5 COMPARISON OF MODELS

The three reviewed models have different approaches to the handling of identity management. It is not entirely possible to provide a valid

comparison of the .NET Passport system, which was implemented with a singular methodology, to the Liberty Alliance framework. The reason for this is that the latter is not a system, but a set of open technical standards which an organisation can implement. The efficiency of a Liberty Alliance framework implementation is only as strong as the level to which the specifications are applied. Further complicating this analysis is the Mozilla TrustBar. The TrustBar looks at the identity management paradigm from that of the user. TrustBar implements similar steps when performing authentication, but instead of the authentication of the user, the accessed website is authenticated. The consolidated comparisons, where they can be drawn, are shown in Table 1.

Table 1 draws a comparison of the three models into specific sections. The .NET Passport is rooted as a singular entity, which is maintained by Microsoft. All usage of the .NET Passport requires adherence to Microsoft standards by website vendors, stipulated in contracts between Microsoft and these parties. The Liberty Alliance makes use of a set of open specifications that can be implemented in various ways to allow for an SSO environment to be created for users. The SSO facility, however, only applies between websites within the same circle of trust, and should a user move out of the circle, they must resubmit their login credentials. TrustBar takes a different perspective looking at the issue from the user point of view. TrustBar currently works off a single system, which is implemented by Mozilla, handling classification and analysing the security of websites to create a central repository for determining website validity.

The three models analysed provide a significant step towards meeting the overall goal of an integrated system for the protection of the user in the online environment. The following three points summarise and categorise each model:

- Microsoft .NET Passport – provides a solution for broad implementation of identification and verification of users within an SSO environment. It also provides simple integration between vendors, due to a single user identification provider.
-

Table 1 – Comparison of .NET Passport, Liberty Alliance and Mozilla TrustBar

	.NET Passport (Lopez, Oppliger & Pernul, 2004)	**Liberty Alliance** (Olsen & Mahler, 2007)	**Mozilla TrustBar** (Herzberg, 2005)
System	Singular System implemented by Microsoft	Open Specifications. Can be implemented in various ways within multiple different systems	Single System implemented by Mozilla to handle classification of web addresses
SSO	Previously multi-organisation SSO. Since 2003 single-organisation sign-on	Depends on implementation, supports multi-organisation SSO	Single verification of websites accessed by user
Choice of Identity Providers	Microsoft was the only identity provider	Allows for several identity providers so far as they are accepted by the service provider	Mozilla serves as identity provider for authentication of websites
Identifiers	Personal Unique Identifier per user	Unique handle per user per federated pair of website	Unique identifiers per participating website
Responsible Controller	Microsoft and service providers are single data controllers	Controllers or processors? -Service providers within a circle of trust become data controllers "at the time users visit their websites" -However according to the Liberty Alliance, it is possible that some service providers may act as processors	Mozilla and service providers as single data controllers
Contractual Framework	Contract between Microsoft and service provider	Implementation dependant -Contract between every website in a circle of trust -Depending on the type of implementation other models may be possible, such as every participating service provider has a contract with one party which organises and administrates the circle of trust	No contract required, makes use of open source community

- Liberty Alliance – provides a flexible solution for the website vendor through the use of circles of trust. This is limited to providing SSO on a smaller scale because of the limited size of the circle of trust. Each website within the circle of trust provides its own login forms for the user. With a greater level of trust between the websites involved, the ability to audit user movements within the system is increased.
- Mozilla TrustBar – provides a way to identify the website from the user perspective. This is accomplished by the implementation of an easy to use identification and verification process; the user is alerted to potential threats within the websites they are seeking to access.

6 USER CENTRIC ONLINE IDENTITY & AUTHENTICATION MODEL

Although all three models discussed do provide a useful service, none cover all the needs of the user. Although each model focuses on the sign-on or website identification issues, none focus on the issues of the user or protection of the user within the online environment. The issues that need to be addressed for the protection of users in the online environment can be summarised as follows:

- Scam Protection – Users must be aware of potential scams online. Education is the best prevention (Bradley, 2007).
- Spoofing – Users must be aware of fraudulent sites and the risks that can occur should their information be compromised (Herzberg, 2005).
- Multiple Verification – When making use of multiple websites, each with individual login criteria, a facility to improve the user experience through the use of SSO methodology is required. SSO reduces the potential for interception of client login data, and promotes ease of use online (Lopez, Oppliger & Pernul, 2007).

- Credential Security – User credentials must be securely transmitted when authenticating SSO environments.

Each of the models possesses attributes that address some of these user requirements, but they themselves are insufficient.

6.1 Authentication of IT Systems and Users

Authentication procedures in the online world are more complex than their real world counterparts. Through the use of brute force attacks, security controls can be compromised in a short period of time. The use of social engineering techniques can make the process even simpler. When performing the process of converting an offline system to an online version, technical authentication procedures are adapted to the online capabilities frequently without adopting the necessary security measures (FIDIS, 2006). The authentication of the actual website may be adequate, but if users are unable to establish the trustworthiness of the website they are lured to, this authentication is in vain.

If more controls and checks are enforced along with dual authentication by users and systems, a more secure environment for the online user will be ensured. This process can be performed by the authentication of users by IT systems and the authentication of the accessed IT system by the user.

6.1.1 Authentication of Users by IT Systems

From a technical viewpoint, an identity is nothing more than a digital pseudonym representing an individual. Therefore, measures are required to verify that a digital pseudonym belongs to the appropriate authorised person (FIDIS, 2006).

Figure 3 depicts the ways in which an IT system can determine the authenticity of a user. An increase in the number of criteria to be enforced provides for a more comprehensive verification of the user.

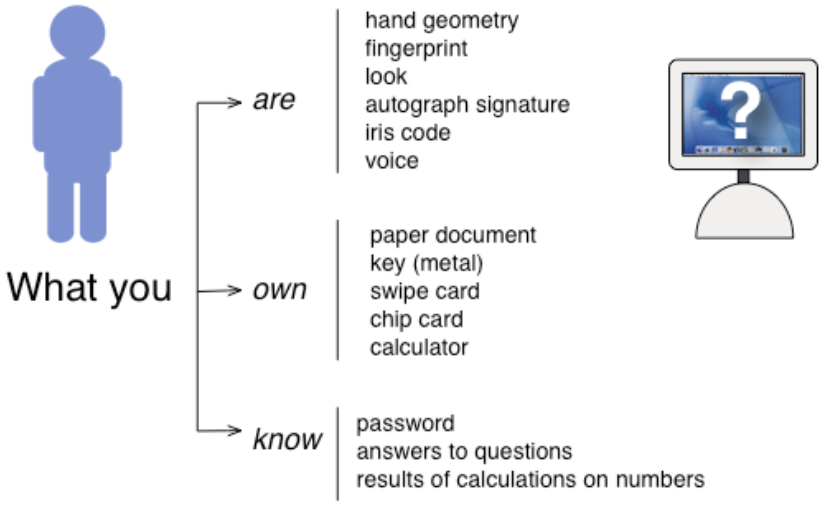


Figure 3 - Authentication by an IT System (FIDIS, 2006)

Based on the above figure, IT systems can recognize a user by their attributes through the use of biometric techniques, what they possess, and what they know. The higher the number of controls implemented using these identification criteria, the higher the level of certainty that the user accessing the system is authorised to do so. Consequently the more criteria used for the authentication process, the higher the levels of trust created between the user and the system.

6.1.2 Authentication of an IT System by a Person

User identity theft is often performed through deceiving the user on a spoofed website. A user enters their login information and attempts to connect, thereby sending their identity data to the perpetrator. To curb this problem, users should authenticate an IT system using the criteria described by the Future of Identity in the Information Society (FIDIS, 2006) which are:

- What the IT system is – By looking at the information contained on the website the user can determine its validity. The immediate method of identifying a website is through the assessment of the website URL. If the URL corresponds to that of the users expected website, they should continue by

determining the validity of the website's digital certificates. In checking the digital certificates the user can determine the validity of the website. This process can be automated through the use of a system such as the Mozilla TrustBar.

- What the IT system knows – Through the registration process the user will set up their initial profile. Some websites may request other personal information relating to the client. The display of this personal information thus verifies the authenticity of a website.

Through the use of both user and system authentication in a dual pronged approach, a user is assured of making use of a valid Internet website.

6.2 A Model for Securing the User's Online Experience

In order to protect the user from threats to their online identity, an approach is required that satisfies both user authentication and website authentication. In Figure 4, a model is proposed which promotes a dual-pronged solution to the protection of user information in the online environment. The model addresses user protection from two angles. The validity of the website is checked and reported to the user. This ensures that the user is attempting to access and authenticate the correct version of the website. Then the process of SSO authentication takes place to allow the user to make use of the benefits of the SSO environment.

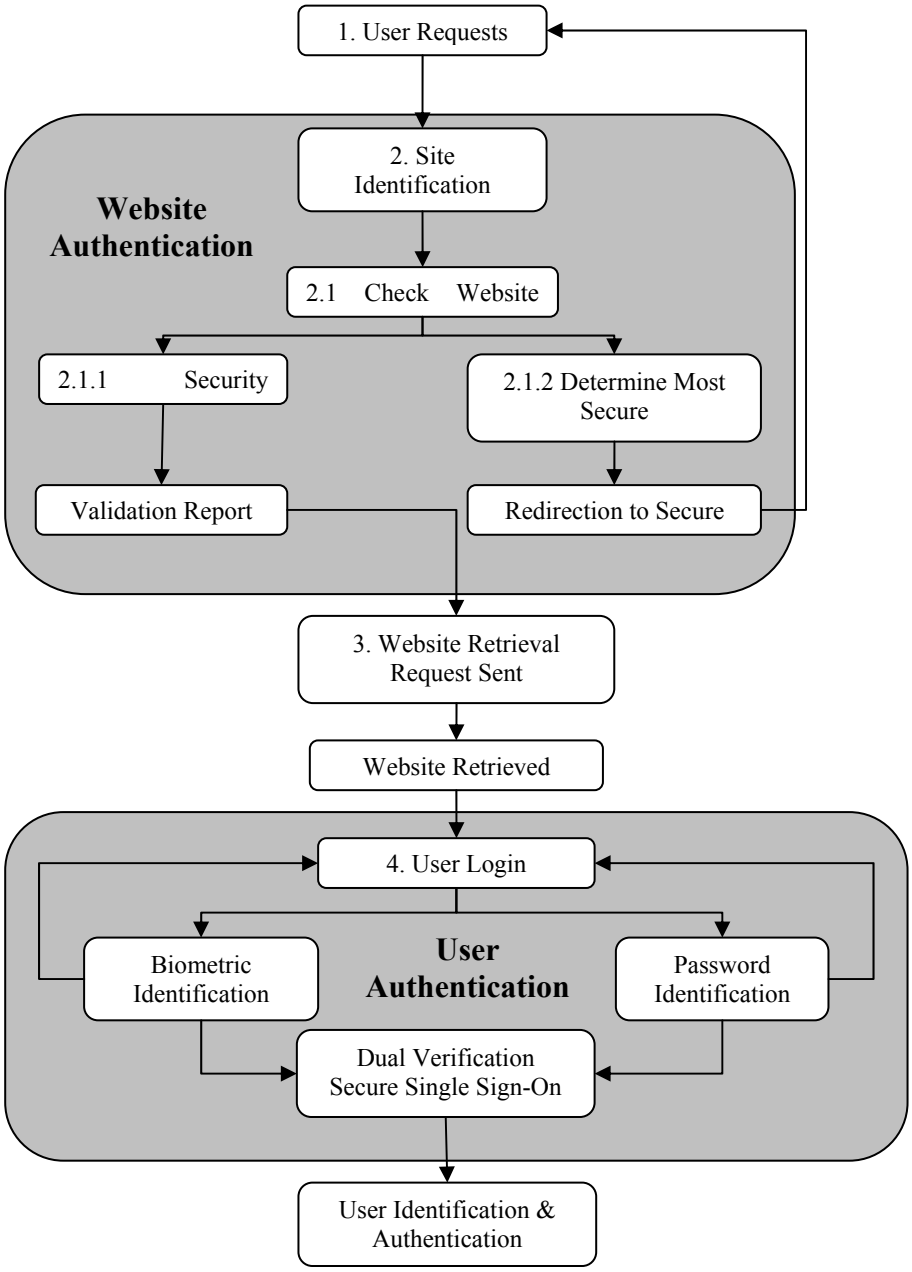


Figure 4 – Model for User Centric Online Protection

The process in the model is expanded as follows:

1. User Requests Website – The user makes use of their browser and enters the URL of the website they intend to visit. By selecting a URL, the process responsible for authenticating the website is initiated.

2. Site Identification Request – When the user performs a request for the website, a request is sent to a repository responsible for the validation of websites. The information in this repository provides the user with information which validates the security of the website.

2.1 Check Website URL – This is accomplished by using the information stored in the repository. These checks are undertaken in order to determine the level of security within the requested website, such as the use of SSL and digital certificates.

2.1.1 Determine Most Secure Site – A number of websites potentially have web pages, which are often more secure but are not set as the default login page. In these cases the repository determines the most secure website. If a more secure inner link for the same domain is found, the repository sends a redirect response to the browser to redirect to the more secure website. Should this occur, then the process from Step 1 reoccurs.

2.1.2 Site Security Check – Based on the URL the repository performs a search determining the validity of the digital certificates, authority of the certificate issuers, and if the site is flagged by the repository as a spoofed website. As a result, a URL validation report is then sent back to the user's browser displaying the results and allowing the user the opportunity to validate the website themselves.

3. Website Retrieval Request Sent – When a user's request for a website is sent, the website is retrieved via HTTP protocols.

4. User Login – Once the page has been loaded with information required for the validation report, the user attempts to login to the SSO website. Incorporated within this process is the use of a biometric input, such as fingerprint identification, along with password identification, which is used to provide a more secure environment. If either of these validation procedures fails, the user is redirected to the initial login page. If the sign-on is successful,

then the user is verified and authenticated within the SSO environment. This process ultimately leads to higher security levels of user identification.

7 CONCLUSION

This paper has discussed the need for a comprehensive model for the protection of users within the online environment. The roles of identity management and the benefits to business were discussed. The role of access and authentication management provided an insight into online user habits with regards to security. The three models .NET Passport, Liberty Alliance Federated Identity and Mozilla TrustBar were examined to determine the processes followed by industry to address identity management. A critical comparison of these models was made which found that none covered all the needs of the user in creating a comprehensive secure environment. A model was then proposed based on the best practices of the industry to promote the use of dual levels of authentication, that is user authentication of the website followed by the website authentication of the user in order to create a secure environment. In using a username and password along with other identification methods, the accuracy of user identification is increased. In addition, the ability of the user to identify the website and verify its authenticity protects the user from the threat of spoofing. This should protect the user from identity theft. An additional benefit of this process is higher levels of trust generated between users and vendors.

8 REFERENCES

APACS (2007) New research reveals that people are still unaware of basic security measures when banking online. Retrieved December 2007 from http://www.apacs.org.uk

BMC Software (2006) Supporting the identity management lifecycle with BMC Identity Management, Technical White Paper. Retrieved July 2007 from http://www.bmc.com

Bradley, T. (2007) Gone Phishing. Retrieved October 2007, from http:/netsecurity.about.com/od/secureyouremail/a/aa061404.htm

Chau, P.Y.K., Hu, P.J., Lee, B.L.P. & Au, A.K.K. (2007) Examining customer's trust in online vendors and their dropout decisions: An empirical study. *Electronic Commerce Research and Applications*, Vol 6(2), pp 171 – 182

Consumers losing trust in online banking: survey (2007) *Computer Fraud & Security*, Vol 2007(2) pp 4

De Leeuw, E. (2004) Risks and threats attached to the application of Biometric technology in National identity management. Retrieved May 2007, from http://secure.gvib.nl/afy_info_ID_1322.htm-ThesisMSIT.zip
FIDIS (2006) D5.2b: ID-related crime: Towards a common ground for interdisciplinary research. Retrieved September 2007, from http://www.fidis.net
Gordon, T. (2004) Quantifiable benefits of implementing identity management systems. Retrieved July 2007, from http://www.isd.salford.ac.uk
Herzberg, A. (2005) Defending users of unprotected login pages with TrustBar 0.4.9.93. Retrieved September 2007, from http://osdir.com/
Herzberg, A. & Gbara, A. (2007) TrustBar: Protecting (even naïve) Web users from spoofing and phishing attacks. Retrieved June 2007, from http://www.cs.biu.ac.il
Jøsang, A. & Pope, S. (2005) User Centric Identity Management, Australian Computer Emergency Response Team Asia Pacific Information Technology Security Conference, Royal Pines Resort – Gold Coast, Australia 22nd-26th May, 2005
Kormann, D.P. & Rubin, A.D. (2000) Risks of the Passport single sign-on protocol. *Computer Networks*, Vol 33(1-6) pp 51-58
Lewis, J. (2003) Enterprise Identity Management: It's About the Business. vol.1, 2 July 2003, Burton Group Directory and Security Strategies Directory and Security Strategies Research Overview. Retrieved November 2007, from www.burtongroup.com
Liberty Alliance (2007) Contractual framework outline for circles of trust. Retrieved July 2007, from http://www.projectliberty.org
Lopez, J., Oppliger, R. & Pernul, G. (2004) Authentication and authorisation infrastructures (AAIs): a comparative survey. *Computers & Security*, Vol 23(7) pp 578 – 590
Madsen, P. (2004) Federated identity and web services. *Information Security Technical Report*, Vol 9(3), pp.56-65
Microsoft (2004) .NET Passport Review Guide. Retrieved August 2007, from http://www.microsoft.com

Olsen, T. & Mahler, T. (2007) Risk, responsibility and compliance in 'Circles of Trust' – Part I *Computer Law & Security Report*, Vol 23(5), pp 342 - 351
Pfitzmann, B. (2004) Privacy in enterprise identity federation – policies for Liberty 2 single sign on. *Information Security Technical Report*, Vol 9(1), pp 45 – 58
Rodger, A. (2004) Access Management the key to compliance. *Card Technology Today*, Vol 16(4), pp 11-12
Sun Microsystems (n.d) Identity management services framework. Retrieved July 2007, from http://www.sun.com/service/identity/

NO AGE DISCRIMINATION FOR BIOMETRICS

[1]M.M. Lessing and L. Weissenberger

[1]Council for Scientific and Industrial Research (CSIR)
[1]marthie.lessing@gmail.com
PO Box 923, Ferndale, 2160, South Africa

ABSTRACT

Biometric advances apply to a range of disciplines to ensure the safety and security of individuals and groups. To stress the value of biometrics, this study focuses on the application of biometric techniques to a vast range of individuals and groups, irrespective of their age. This report covers biometric development within three generations.

For the younger age group, biometrics can play a significant role in ensuring physical safety within the learning and dormitory environment. Additionally, biometrics can assist teachers within this environment to enhance the administration features. This allows more hands-on time for the education of children.

The application of biometrics for adults has made great progression in the last couple of years. The research considers biometric advancements in the areas of travel and immigration, healthcare, law enforcement and banking. For the purpose of this study, adults are considered the individuals and groups in a working environment. Many of these applications are relevant to the younger and more senior generations as well.

Senior citizens can also benefit from biometric applications. In many countries, biometric techniques control the administration of pension funds and general welfare administration.

For each of the biometric applications, this research reviews the application of biometrics, associated advantages and disadvantages, as well as specific implementations. A number of sample applications from all over

the world, illustrates the usability of biometrics for a variety of groups, individuals and disciplines. From this report, it is clear that biometrics is a universal application, used by anyone, anywhere.

KEY WORDS

Biometrics, safety and security, school environment, travel, immigration, healthcare, law enforcement, banking, pension funds.

NO AGE DISCRIMINATION FOR BIOMETRICS

1 BACKGROUND

Biometrics is not a passing fad, and definitely not a new development. The earliest recorded use of biometrics for identification purposes occurred during the 14th century. This is when Chinese merchants stamped children's palm and footprints with ink on paper. The year 1881 was a noteworthy milestone in the advance of biometrics: Alphonse Bertillon developed an anthropometric system that measures and distinguishes between human traits (AB 2006). It is, however, only in the last 120 years that the biometric discipline introduced drastic changes (Arnold 2006).

The application of biometric techniques has slowly infiltrated our daily lives and has established an intimate interdependence between humans and technology. In computer security, biometrics refers specifically to automatic authentication techniques relying on physical measurable features. Biometrics refers to: "… technologies that measure and analyse human body characteristics, such as fingerprints, eye retinas and irises, voice patterns, facial patterns and hand measurements, for authentication purposes" (SearchSecurity.com 2006). The use of biometric applications has become so prevalent that AuthenTec, the world's leading provider of fingerprint sensors and solutions, reported a 67 percent increase of biometric related sales in 2007 (FindBiometrics 2008a).

In No age discrimination for biometrics we attempt to show that biometrics has infiltrated all facets of human life, with specific attention to the security aspects. We discuss biometrics for the young, for adults and for senior citizens and prove that no matter what your age, there is a biometric application that applies to your needs and circumstances.

2 BIOMETRICS FOR THE YOUNG

Children are the world's future, and it is very important to keep them safe. Since majority of children attend schools and many youths colleges and universities, this section addresses the application of biometric techniques in the learning environment.

2.1 Educational institutions

Schools are supposed to be a secure haven where children can achieve their full potential, but are schools as safe as we want to believe? In order to learn and excel, scholars need to feel safe in an environment where they can focus on their studies. Not only the learners' and students' safety is important, but teachers and lecturers also need to know that they are in a protected environment where they can focus on the curriculum.

Existing security solutions are far from perfect. Children have the knack to lose physical keys and magnetic swipe cards, forget passwords or fall victim to the social engineering skills of school bullies to obtain lock numbers. In addition, the current process to gain secure access to school grounds is both time consuming and ineffective. The most probable solution is a reliable access control and identification system, combined with effective entry policies.

2.1.1 Application areas

It is seldom that all the relevant parties immediately accept the introduction of a new technology. Especially in the teaching and learning environment where minors are involved, people become natural sceptics. Yet, with the young generation's outlook on technology and inventions, it is not surprising to find a great variety of biometric applications in school environments.

The best way to introduce biometrics in educational institutions is to prepare an alternative system for individuals uncomfortable with the new system. This will get the approval of the students, their parents and the educators (Goldberg 2003).

The most common application is campus access. Access is limited to preferably one entrance equipped with a biometric scanner to verify the identity of all students, staff and authorised visitors upon entry (Goldberg 2003). The execution of biometric systems in schools should be relatively easy, since children seem to hold no fear of new technology. More than 1,300 UK primary schools' libraries are using fingerprint technology to replace the old-fashioned password systems. More children now borrow books because they want to use the new technology (Out-Law News 2004).

Another application that can speed up scholar attendance registry is placing cameras throughout the school, authenticating students looking into the camera. Roll call information is available immediately, and not only after first break as is presently the norm with class attendance lists (Nixon 2003). Students can also use their biometric features to authenticate themselves on the school network by logging on to a wireless network. The application of biometrics leads to tremendous time saving (Goldberg 2003). For students with medical allergies, biometric devices have proved to be lifesavers. School nurses use fingerprint scanners to identify sick students and their allergies before they administer medication (FindBiometrics 2003).

For schools with a vision to send pupils into the business world equipped with all the necessary skills, biometric school environments are ideal. The school environment introduces the technology to students in an informal environment, and prepares them adequately for biometrics in a more formal corporate world. In addition, the use of biometric technology in the classroom will leave both the scholar and the teacher feeling more secure in the school environment and will enhance the learning experience (BIOMETRICS in Education 2004).

2.1.2 User resistance

The main problem with using biometrics in educational institutions is people's resistance to change (in this instance the school/university staff and parents), their resistance specifically to biometric technologies and the cost implications of such technologies. Though expensive, the arguments in support of biometrics weigh much heavier. Privacy concerns may put a dampener on acceptance rates, but should not prevent biometric technology from reaching its full potential in security enhancements (Goldberg 2003). Educational institutions already hold the responsibility for sensitive data such as identification numbers and special needs information. Biometric systems simply add additional information to these personal data files.

Additionally, there is a worldwide fear by school employees of contracting diseases from using the same scanners as the scholars. Children can pass on childhood diseases that could hold immense danger for adults and their families. The likelihood of open cuts and the transfer of germs is

another concern, but according to Borja (2002) the risk of contamination is minimal to nonexistent.

Advocates in support of biometrics in schools say that the resistance endures because the technology is misunderstood. Finger scanners do not record actual fingerprints and is of no use to law enforcement (Graziano 2003). Once the resistance has subsided, biometrics can enhance the school environment with access control, positive identification and a record of those entering and leaving school buildings. Hopefully the fears of identifying information misuse will soon be forgotten to embrace the benefit of enhanced accountability (Goldberg 2003).

2.1.3 Advantages and disadvantages Application areas

The advantages of a successfully implemented system include saving on administration costs, increased accountability, improved building and data security and improved effectiveness in administrative tasks. With minimal biometric training required, this technology relieves teachers of their administrative tasks, leaving more time spent teaching (BIOMETRICS in Education 2004).

The greatest benefit of biometric authentication is that students can display their fingertips publicly without any threat of compromising their accounts' privacy or the network's security. It is also highly impossible for a student to authenticate using someone else's fingertip, or forgetting their fingertips (Kennard 2002). The most disabling disadvantage of biometric systems in the educational environment, however, remains user resistance by educational administrators.

2.1.4 Biometric implementations in schools worldwide

Biometric implementation on school grounds have been surfacing slowly since late 1997. The following are a few examples of educational institutions that implemented biometric systems successfully:

- At the Kvarnby School, Stockholm it generally takes half a period before teachers could help all children sort out their passwords problems. Some students forget the passwords, while others borrow user names and passwords from other students. Often the default password is never changed and written on the blackboard, completely negating the use of passwords. A fingerprint-based

solution eliminated these identification problems, making the login routines easier and saving valuable classroom time (Security International 2002). At Johnson and Wales University, Denver, students no longer need to worry about lost access cards or residence room keys locked in, nor the fines associated with such occurrences. Students entering the Pulliam residence need only to slide their hand into the biometric hand reader for the door to open. The room doors work on the same system (Johnson & Wales University 2002).

- Since 1993, the scholars at the Penn Cambria schools have registered by means of their fingerprints. Only students who have recently transferred to the district and returning students in the fifth and ninth grades need to reregister due to a growth spurt. At these ages, children's fingers have matured to the point where the scanners can no longer quickly identify them. Of around 3000 students, only about 1% opts out of the scanning system due to religious preferences (Graziano 2003).
- US Biometric Corporation, in conjunction with law enforcement experts, are introducing educational seminars to help tertiary institutions understand biometrics. The idea is to assist campuses to increase security proactively for both students and employees (Business Wire 2008).

2.1.5 Summary

Educational institutions need enhanced security features to safeguard learners/students and staff members. By implementing biometric systems, it is possible to improve the current security systems and to boost supporting actions in and around the study environment. Currently many biometric applications serve the global educational environment, varying from simple identification procedures to intense authentication and verification procedures. The advantages of biometric systems far outweigh the disadvantages thereof, and may even lead to a more technology proficient youth.

3 BIOMETRICS FOR ADULTS

The application of biometrics in adult life is a vast field. This section deals with travel and immigration, healthcare, law enforcement and banking.

3.1 Travel and immigration

In the light of terrorist attacks occurring around the world, governments are looking to intensify security controls, especially in the field of immigration. Biometrics is likely to increase security at immigration control points like border posts and airports.

3.1.1 Application areas

Biometrics can assist the travel and immigration industry in two distinct ways: verifying the identity of visitors and including biometric identification within passports. When validating a visitor's identity, the immigration official takes a digital photo of the person to match against the database. The system performs a matching against wanted or missing persons in an attempt to uncover fraudulent applications (US Department of State 2004:1,12).

Most immigration departments also take fingerprints of the visitor to compare these to fingerprints in the database. If the system does not recognise the person's fingerprint, it sends the prints electronically to an off-site facility for further analysis by an experienced latent fingerprint examiner (Biometric Technology Today 2004:5).

3.1.2 Advantages and disadvantages

Biometric technology strengthened confidence in passports and visas, reduced fraudulent activity and continually assists in fighting terrorism. The inclusion of biometric data within passports will have three security benefits:

- immigration officers can verify whether the passport identifies the bearer sufficiently;
- the movements of travellers can be more easily tracked, enabling officers to identify those who breach the conditions of their visas; and

- passports will be harder to forge (Biometric Technology Today 2004:12).

Proponents of human and civil rights have expressed apprehension that the increased security measures are part of an international attempt to keep track of people's movements. In particular, they have concerns about facial recognition since it discloses a person's ethnic and racial background. They also have reservations about the accuracy and reliability of the technology.

Regarding passports containing biometric data, vendors have had a great deal of difficulty in incorporating the microchip into the pages of the passport. The main concerns are the possibility of illicit people attempting to read the information off the chip from a distance, and the compatibility of readers manufactured by different suppliers (Biometric Technology Today 2005:2).

3.1.3 Biometric implementations in travel and immigration worldwide

The application of biometrics in the travel and immigration industry has vast ranges. The following countries employ biometric applications successfully:

- Russia recently issued the first biometric passports for Russians travelling abroad. This passport includes a special photograph and a microchip for digital finger or retina prints (FindBiometrics 2008d). More than 50 other countries have migrated to the use of biometric passports in the past three years (Wikipedia 2008d).
- Singapore employed the LG IrisAccess technology to definitively authenticate visa holders and allow them to enter the country, introducing timesaving of over 2400 percent (FindBiometrics 2008b).
- In South Africa¸ IDTek was awarded a R250 000 contract by Airports Company SA (ACSA) to install Sagem's fingerprint biometric technology in the restricted personnel areas of OR Tambo International Airport. This application will enforce security and increase physical access control reliability (ITWeb 2006).

3.1.4 Summary

Biometrics is an important step for governments to take to stay ahead of terrorists and other lawbreakers. In this industry, it is still in its infancy with many obstacles to overcome, but the results look promising so far, which bodes well for biometrics within the immigration environment.

3.2 Healthcare Industry

Errors in the medical profession could mean the difference between life and death. In the United States, approximately 115 000 deaths occur each year from misidentifying a patient (Schneider 2005:24). By implementing biometric technology, nurses can reduce this number significantly by checking a patient's fingerprint before performing a surgical procedure.

Another major issue within healthcare is fraud. It is common for multiple individuals to use the same medical aid card to receive health benefits, especially when the cards do not contain a photographic image of the insured person (Messmer 2004:17). With the implementation of biometrics, this kind of fraud would be greatly minimised.

3.2.1 Application areas

abroad often employ iris recognition scanning for system access control, ensuring that only authorised doctors and nurses can view confidential patient records. Additionally, these workstations can be fitted with proximity sensors so that if a user moves away from the terminal, the system will log the person out (Dalton 2004:12). Fingerprints are the most commonly used biometric identifier used within the healthcare industry, since most readers are able to read prints through grime (Schneider 2005:24).

Many hospitals traditionally had multiple systems, each requiring different logon credentials. Medical professionals needed to remember as many as six different passwords. In some cases, systems implemented single sign-on to solve this problem. The drawback of such a system would be that compromise of a single logon password compromises the entire system. Hence, using biometrics for authentication instead of a token or password is becoming very popular in the healthcare environment. Biometric technology is much more secure than tokens or passwords (Mansfield 2003:40).

3.2.2 Advantages and disadvantages

Biometrics within the healthcare industry can improve the quality of healthcare, reduce medical errors and decrease healthcare costs (Schneider 2005:22). It can also assist in providing an audit trail, by identifying which staff member supplied care as well as what type of care was provided to a patient (Beyond doors: Securing records with finger flick 2002). The main advantages of biometrics include the combating of fraud and abuse in health care entitlements programmes, the protection and proper management of confidential medical records, positive identification of patients, and securing medical facilities and equipment (Marohn 2006).

Although fingerprint scanners are the most widely used biometric device used within the healthcare industry, this can be problematic in areas where staff are routinely required to wear gloves (Dalton 2004). These areas employs alternative, often more intrusive biometric technologies.

3.2.3 Biometric Implementations in Healthcare Worldwide

Biometrical implementations in health care have been successful in the following:

- A hospital in New York and an ambulance service in Chicago have both implemented an ultrasound fingerprint scanner used during patient registration. This fingerprint is stored as part of the patient's permanent record to ensure that only that person uses the medical aid card. The system has successfully expanded to control access to cabinets containing narcotics (Messmer 2004).
- A common phenomenon is phantom billing, where health care providers bill for services never rendered. Texas addressed this problem by incorporating a biometric smart card-based program that requires both the medical providers and the recipients to authenticate themselves when checking in for service. This system greatly reduced hospital expenditures and improved program integrity (Marohn 2006).
- In the aftermath of the Florida hurricanes in 2002, the USA initiated the e-Life-Card program. Individuals seeking medical care during the hurricanes experienced significant delays and lack of access to their medical information. The program allows first

responders to access critical information by using patients' fingerprints as authenticator (Marohn 2006).

- Poudre Valley Health System, Colorado, previously used PINs to control access to their newborn nurseries. However, many incidents of unauthorised persons gaining access to this highly restricted area occurred. The facility now uses hand geometry readers instead of PINs (Reynolds 2004:16).
- Australia introduced the MethaDose program, employing iris recognition technology to support the treatment of heroin addicts. The program registers patients to detect duplicate enrolees, and to enable authentication for patients that are unable to claim their identity coherently. Registration includes personal information such as name, biometric data, permitted dosage, last and next scheduled dosage, all stored on a central database. The Netherlands launched a similar program to automate and control distribution of vaccines during epidemics (Marohn 2006).

3.2.4 Summary

It is clear that biometrics can be very beneficial to the medical industry by creating a more convenient working environment for staff, and reducing fraud and medical errors.

3.3 Law Enforcement

The legal implications and red tape of police departments accessing inter-jurisdictional systems creates an ongoing problem in identifying and apprehending criminals. Additionally, perpetrators using multiple identities can fool traditional identification systems. If biometric features are used, law enforcement may have more success in this regard, linking multiple identities to a single person. In many regards, it may be beneficial for law enforcement agencies to share biometric data.

3.3.1 Application areas

The first recorded use of latent fingerprints as a means of identification is in 14th century Persia (Wikipedia 2008a). Since then, police departments relied heavily on latent fingerprints and witness reports to identify people that were present at crime scenes. The Integrated Automatic Fingerprint

Identification Systems (IAFIS) is a database system maintained by die Federal Bureau of Investigation, using a one-to-many matching technology to match fingerprints to individuals. IAFIS contains fingerprint and criminal history information for over 47 million people (Patrick 2007).

Since the 9/11 terrorist attacks, there has been major interest in the use of facial recognition software to identify terrorists and other wanted criminals in public areas such as airports, sports stadiums and correctional facilities. Some patrol cars in the United States were fitted with mobile facial recognition units, giving police officers the ability to verify a person's identity within minutes. This is particularly useful when individuals claim they do not have any form of identification on their person. Law enforcement also uses iris recognition to improve efficiency and safety within correctional facilities (Zalud 2003:30).

3.3.2 Advantages and disadvantages

Biometrics has been invaluable as a unique identifying characteristic in determining when a suspect is using multiple identities or aliases (Biometric Technology Today 2005:12). Facial recognition has increased the speed and efficiency of the booking process at police stations, and aided in distributing images and information to other police departments, correctional facilities and sheriffs' offices (Zalud 2003:31).

A major disadvantage of biometrics is that civil liberties groups believe the use of cameras in public streets to constitute an infringement of privacy. They believe that law enforcement should not violate citizens' rights unless they have legitimate cause to do so (Beckley 2004:16). They also question the reliability, effectiveness and correctness of results (Hudson 2003:1) and that these systems could promote racial profiling. Additionally, the use of cameras in city streets has not been as valuable as anticipated since the technology is most effective when the subject is stationary, at close range and when the light is good (Winton 2004). This limits use of the technology, but future improvements should minimise these restrictions.

3.3.3 Biometric Implementations in Law Enforcement Worldwide

Following are examples of biometric implementations in law enforcement:

- The American serial killer Ted Bundy bit Lisa Levy in her left buttock cheek during one of his attacks, leaving prominent bite

marks. A forensic expert positively matched plaster casts of Bundy's teeth to photographs of Levy's wound, leading in part to his conviction (Wikipedia 2008b). The accuracy of this method is highly criticised, since a study done by the American Board of Forensic Odontology revealed a 63% rate of false identifications (Wikipedia 2008c).

- In Florida, police has deployed mobile facial recognition systems in patrol cars. In six months, police made 37 arrests that would not have been possible previously due to the perpetrator providing false or no identification (Biometric Technology Today 2005:12).
- In the United Kingdom, police convicted Mark Gallagher in 1998 of murdering a 94-year¬old woman. The main incriminating evidence was an ear print found on a window at the murdered woman's home. The judicial system overturned this conviction in 2004 when scientists pronounced the ear print evidence flawed, and DNA evidence incriminated another man for committing the crime (Graham-Rowe 2005).

3.3.4 Summary

Biometrics has assisted police departments and law enforcement agencies to capture criminals that they would not have been able to before implementing the technology. It appears that as biometrics becomes more affordable and flexible, biometrics within law enforcement will play a vital role.

3.4 Banking

Despite years of marketing and hype surrounding biometrics as the answer to all security problems, biometrics is taking off exceptionally slowly in banking environments (Bruno 2001). World wide financial institutions are slowly starting to implement biometrics.

3.4.1 Application areas

The type of biometrics banks should use depends on a variety of factors. Some biometrics, such as retina scanning, is highly accurate. Economically, however, retinal devices are not practical for securing a bank's ATMs, although it may be appropriate to use internally for vault access and

computer networks. A practical mass-market approach to biometrics for banks is devices that rely on existing infrastructure, such as cameras on ATMs (Bruno 2001).

Millions of financial transactions are easily and securely processed using fingerprint technology. A variety of Sagem biometric-based services offer merchants a secure, low-cost payment form that reduces transaction fraud without sacrificing customer convenience (Law Enforcement 2005). Another popular use for banking biometrics is PassVault, made by Diebold. PassVault enables customers to access their safe-deposit box unassisted by bank personnel, by registering their hand or fingerprint scan when applying for a box. Customers enter a PIN and scan their handprint when they want to open the box (Bruce 2001).

An important aspect of biometrics is privacy. To ensure the widespread acceptance and implementation of biometrics, it may be necessary to encrypt the retrieved biometric features. This ensures that someone cannot reconstruct an identifiable fingerprint from an encrypted finger scan stored in the database. Recent incomplete research shows a relationship between personality and the patterns of colours in the iris, igniting a widespread fear that using biometric systems may reveal private information about a person (Patrick 2007).

3.4.2 Advantages and disadvantages

The advantages of biometric systems in the banking environment are numerous, especially for developing countries with newly developing banking networks. These advantages include reducing the surplus of fiduciary money in circulation, and boosts and secures electronic fund transfers and clearing. For many people the most important benefit is ensuring that the right person receives the payment, preventing identity theft and subsequent fraud. Biometric systems also reduce the cost and risk of transporting funds. Developing banking services, and especially encouraging individual savings, can facilitate proper monthly expenditures (Philippe 2004).

Biometrics can be very effective, but is not well suited for users who want to work on multiple machines or in different locations. Many people take work home to do after hours, but without a biometric reader at home, they cannot do their e-commerce transactions (Livewired Communications

2003). Another concern regarding biometrics is their reliability, largely due to media headlines negating the technology in the earliest days of public trials. The most practical disadvantage is the logistics: getting customers to come and register their details (Sturgeon 2005).

3.4.3 Biometric implementations in banks worldwide

Biometric implementation in banks has been successful at the following places:

- Banque Artesia, Amsterdam is using South African company Biometrics.co.za's software to provide banking services to the oil industry in Rotterdam. These high-risk transactions include large sums of money, necessitating an easy to use system that can reliably identify clients before effecting electronic transactions (Burrows 2004).
- The Bank of Tokyo-Mitsubishi, Japan deploys a security system based on vein-pattern recognition at all its branches. The bank's clients use smart Visa credit cards with the customer's vein-pattern information stored on the card's chip to validate their identity when using ATMs (Biometric Technology Today 2004).
- The United States Government Accountability Office reported that the Federal Emergency Management Agency have improperly disbursed more than R7 billion by not validating the identity of aid registrants in the wake of hurricanes Katrina and Rita. One individual received more than R1 million in aid, by registering 13 times using different Social Security numbers (Patrick 2007). ISO published ISO 19092:2008 to increase the security of financial transactions over electronic media. This standard aims to ascertain security requirements for the implementation and management of state-of-the-art biometric identification technology within the financial industry (FindBiometrics 2008c).

3.4.4 Summary

It is crucial that all financial institutions should be as safe as possible. By implementing biometrics worldwide in banks, the country's citizens can rest assured that the economy is safe and stable. A definite boost of confidence in banks is also noticeable. This is due to people who formerly refused to

open bank accounts because there was no foolproof security system in place.

4 BIOMETRICS FOR SENIOR CITIZENS

Senior citizens are generally not up to date with the latest technology trends. However, the use of biometrics in seniors' life can ease many aspects, especially regarding pension allocation.

4.1 Pensioners

According to Joseph Atick, CEO and president of Identix, the public's privacy concerns are a bigger issue than that of cost. He mentioned that senior citizens were more open to biometric technology than the younger generation: they like not having to remember a PIN and want to ensure that their nest eggs are protected (Coogan 2004).

4.1.1 Application areas

Before biometrics made its debut in social welfare, accessing information regarding pension was time consuming and difficult. Senior citizens, who often have difficulty walking, had to go to the appropriate government office and wait in long queues, often in several different offices all over town. Senior citizens can now visit a single biometric kiosk to obtain the relevant information needed to receive pension benefits: databases from the National Institute of Social Security, the National Institute of Employment, the General Treasury of Social Security and the Social Institute for Sea Workers all connect to the system. To use the kiosks, senior citizens have to have a smart card with their name and an ID number, and enrol in the system by scanning either of the index fingers. This ensures that only the enrolled person can receive monetary benefits from the system. In cases of frailty or illness, the system can fingerprint a family member or friend to collect the allowance on behalf of the beneficiary (Gemplus Corporation 2002). The system is also adapted to allow for payment of pensions via ATMs. Using these ATMs, pensioners can access their cash at any time, at the touch of a finger. They now do not need to carry large amounts of cash with them anymore and make them less vulnerable to thievery when travelling home (FindBiometrics 2004).

Before the implementation of the biometric system, it was easy to obtain pension benefits illegally. If someone lost their ID card, an

unauthorised individual could use it to access the cardholder's medical records or pension benefits. The best way to protect against these types of incidents is to combine verification of both the card and the fingerprint (Pronko 1998).

The ideal biometric system to implement for social welfare would involve digitised photographs and hand geometry stored in a central database. A plastic identity card with a magnetic strip will contain the photograph, the client's analogue signature and date of birth, a selection of security features and a thumbprint. Authorised staff members at multiple sites will use data scanned from a person's hand to search the databank for matches and interface with the existing information systems (Davies 1994).

4.1.2 Advantages and disadvantages

The most outstanding benefits of biometric applications regarding social welfare industries are that pensioners receive their benefits in a faster, more convenient and secure way. The synergistic effect of offering welfare and pension payments through biometrics-equipped bank ATM networks offer many benefits. Government can reduce its cost and provide a more efficient and timely service to its constituents; financial institutions can increase the volume of transactions, whilst reducing the unit transaction costs; banks' cumulative revenues can be increased by charging the government agencies for the service. Hopefully, in the long run the public at large can benefit from reduced taxes as a result of a more efficient government (Yanez & Gomez 2004).

By implementing these social assistance cards with smart card technology, the biometric system adds inherent security features. The decent storage capability and the electronically readable format make the smart card the optimal solution to address the social welfare program's major security, financial control and portability concerns (Gemplus Corporation 2002). Disadvantages include that only government agencies may do capturing and storage of fingerprints for it to be considered legal (Yanez & Gomez 2004).

4.1.3 Biometric implementations regarding pension distribution worldwide

- Spain's government incorporated biometric verification units with information kiosks to allow citizens to access personal information, pension and healthcare benefits. The 633 kiosks are located in different government offices in the Andalusia region of Spain, and will eventually be implemented nationwide (Pronko 1998).
- South Africa's government has had difficulty with the payout of pensions to the elderly, especially those in remote areas who often have limited mobility. In 2001, the government started doing pension payments through a mobile van distribution pay points as part of their plan to bring the government services closer to the people. HighTech Laboratories designed the system to use about 500 vans, fitted with ATM-style machines and Identix BioTouch USB fingerprint readers (Identix 2004).
- The Philippine Social Security System launched an identification card system in November 1998 to ensure that members, pensioners and dependants do not enrol using multiple identities (Breedt & Olivier 2004).

4.1.4 Summary

By introducing biometrics in the area of pension retrieving, the lives of the senior citizens become less complicated. The adoption of such implementations has been slow, but it has proved to be successful.

5 CONCLUSION

In No age discrimination for biometrics, we showed that biometrics applies to all facets of human life. The research focused on three distinct age genres, reviewing each area, as well as both successful and unsuccessful implementations. Many of the applications discussed under adult biometrics are applicable to both the younger and older generation, but falls favour to the most prominent category. For example, while children from primary school level upwards use a savings account, and retired people often put their pension in a bank account, usually adults make the most noteworthy financial decisions.

After examining the school environment, the immigration and healthcare industries, law enforcement, the banking environment and pension distribution, it is clear that the advantages far outweigh the disadvantages of using biometrics for security. The common advantages include increased security, reduced fraud, less administration problems created by forgotten passwords, easier employee auditing and logging and significant cost savings (QuestBiometrics 2005). In most cases, the most prominent disadvantage is user resistance, which will significantly lessen as the technology becomes more widely accepted.

6 REFERENCES

AB. 2006. A Short History of Biometrics. Associated Content -The People's Media Company. [Available from: http://www.associatedcontent.com/article/48809/a_short_history_of_biometrics.html (Accessed 7 February 2008)].

Arnold, B. 2006. Caslon Analytics biometrics. [Available from: http://www.caslon.com.au/ biometricsnote1.htm (Accessed 7 February 2008)].

Beckley, A. 2004. The Future of Privacy in Law Enforcement. FBI Law Enforcement Bulletin. [Available from: www.accessmylibrary.com/coms2/summary_0286-31073479_ITM (Accessed 15 August 2005)].

Beyond doors: Securing records with finger flick. 2002. Security, 39(7):57. [Available from http://0-proquest.umi.com.raulib.rau.ac.za:80/pqdweb?did=181577441&sid=7&Fmt=4&clientId= 57200&RQT=309&VName=PQD (Accessed 10 August 2005)].

Biometric Technology Today. 2004. US -Visit awards fingerprint services contract. Biometric Technology Today, 12(10):5.

Biometric Technology Today. 2005. Passport plans in disarray. Biometric Technology Today, 13(5):2.

BIOMETRICS in Education. [Available from: http://webhost.bridgew.edu/jcolby/it525 (Accessed 10 August 2005)].

Borja, R. 2002. Finger-Scanning Technology Monitors School Employees. [Available from: http://www.edweek.org/login.html?source=http%3A%2F%2Fwww.google.co.za%2Fsearch%3Fhl%3D af%26q%3D%2522Finger-Scanning%2BTechnology%2BMonitors%2BSchool%2BEmployees%2522%2BBorja%26meta%3D&destination=http%3A%2F%2Fwww.edweek.org%2Few%2Farticles%2F2002%2F10%2F23%2F08biometric.h22.html&levelId=2100&baddebt=false (Accessed 15 April 2008)].

Breedt, M. & Olivier, M. 2004. Using a central data repository for biometric authentication in passport systems. [Available from: http://icsa.cs.up.ac.za/issa/2004/Proceedings/Full/072.pdf (Accessed 12 February 2008)].

Bruce, L. 2001. Face-scanning, fingerprinting ATMs gain ground. [Available from: http://www.bankrate.com/brm/news/atm/20010302a.asp (Accessed 10 April 2008)].

Bruno, M. 2001. Biometrics Are Too Hot to Handle. [Available from: http://www.encyclopedia.com/ doc/1O8-biometrics.html (Accessed 12 February 2008)].

Business Wire. 2008. US Biometrics Corporation to Conduct The First In Its Series of Biometrics Education Seminars to Universities and Colleges Across The United States. [Available from: http://www.pr-inside.com/us-biometrics-corporation-to-conduct-the-r427644.htm (Accessed 8 February 2008)].

Burrows, T. 2004. Dutch bank using SA biometrics. [Available from: http://www.itweb.co.za/office/ grintek/0308010725.htm (Accessed 10 August 2005)].

Coogan, J. 2004. FACT Act Provision Raises Biometrics' Profile. American Banker. [Available from: http://www.biometricgroup.com/in_the_news/03_17_04.html (Accessed 13 February 2008)].

Dalton, A. 2004. Eye Spy. Hospitals & Health Networks. 78(11):12. [Available from http://0¬proquest.umi.com.raulib.rau.ac.za:80/pqdweb?did=750586381&sid

=6&Fmt=4&clientId=57200&R QT =309&VName=PQD (Accessed 21 August 2005)].

Davies, S. 1994. Touching Big Brother: How biometric technology will fuse flesh and machine. [Available from: http://www.asylumsupport.info/publications/privacy/biometrictechnology.htm (Accessed 13 April 2008)].

FindBiometrics. 2003. BIO-key Scores Straight A's in Education Application. FindBiometrics.com.[Available from: http://www.findbiometrics.com/viewnews.php?id=512 (Accessed 10 March 2008)].

FindBiometrics. 2004. NCR and Bancafe Use Biometric Technology to Reach New Colombian Banking Customers. [Available from: http://www.findbiometrics.com/press-release/1795 (Accessed 12 February 2008)].

FindBiometrics. 2008a. AuthenTec Reports Record Fourth Quarter 2007 Financial Results. [Available from: http://www.findbiometrics.com/press-release/4886 (Accessed 8 February 2008)].

FindBiometrics. 2008b. Case Study: Iris Recognition Enhances Security, Accelerates Traffic, and Reduces Costs at Border Crossing. [Available from: http://www.findbiometrics.com/Pages/ airport_articles/lg-case-study.html (Accessed 8 February 2008)].

FindBiometrics. 2008c. Safer electronic financial transactions with new ISO standard for state-of¬the-art biometric authentication. [Available from: http://www.findbiometrics.com/press-release/4896 (Accessed 8 February 2008)].

FindBiometrics. 2008d. New Russian Biometric Passports With Empty Chips Issued. [Available from: http://www.findbiometrics.com/article/495 (Accessed 8 February 2008)].

Gemplus Corporation. 2002. Aplitec Social Assistance and Pension Card. [Available from: http://www.gemalto.com (Accessed 18 February 2008)].

Goldberg, L. 2003. Creating Safer and More Efficient Schools with Biometric Technologies. [Available from: http://www. thejournal.com/articles/16433 (Accessed 10 March 2008)].

Graham-Rowe, D. 2005. Ear biometrics may beat face recognition. [Available from: http://www.newscientist.com/article.ns?id=dn7672&feedId=online-news_rss20 (Accessed 2 March 2008)].

Graziano, C. 2003. Learning to Live With Biometrics. [Available from: www.wired.com/ news/privacy/0,1848,60342,00.html (Accessed 10 March 2008)].

Hudson, A. 2003. Tampa cops end camera program. The Washington Times. [Available from: http://0-search.epnet.com.raulib.rau.ac.za/login.aspx?direct=true&db=nfh&an=4KB20030821094511 (Accessed 15 August 2005)].

Identix. 2004. South African National Pension Payout Program: Facilitating Entitlement Distribution. [Available from: http://www.ibia.org/membersadmin/casestudy/pdf/9/South%20Africa%20National%20Pension.pdf (Accessed 12 February 2008)].

ITWeb. 2006. Jo'burg utilises biometrics. [Available from: http://www.itweb.co.za/sections/computing/2006/0608301030.asp?S=Biometrics&A=BIO&O=FRGN (Accessed 8 February 2008)].

Johnson & Wales University. 2002. Denver is First University in Colorado to Utilize Biometrics to Gain Access to Dorm Rooms. 2002. [Available from: http://www.jwu.edu/media/pressarc/ 02/co11_21_02.htm (Accessed 10 March 2008)].

Kennard, L. 2002. SCANNING STUDENTS: A Stockholm School Goes Biometric. [Available from: http://support.novell.com/techcenter/articles/nc2002_05b.html (Accessed 11 February 2008)

Law Enforcement. 2005. [Available from: http://www.morpho.com/company/our_markets.html (Accessed 3 December 2007)].

Livewired Communications. 2003. [Available from: http://www.itweb.co.za/office/grintek/ 0308010725.htm (Accessed 10 August 2005)].

Mansfield, S. 2003. Password Proliferation Alleviated. Security, 40(9):39-40. [Available from: www.accessmylibrary.com/coms2/summary_0286-31104683_ITM (Accessed 15 March 2008)].

from: www.networkworld.com/news/2004/121304biometrics.html (Accessed 17 March 2008)].

Nixon, S. 2003. School roll could be replaced with eye scan. [Available from: www.smh.com.au/ articles/2003/03/07/1046826531945.html (Accessed 10 March 2008)].

Out-Law News. 2004. Debunking six myths of biometrics. Out-Law.com. Out-Law News, 09/07/2004. [Available from: http://www.out-law.com/ page-4698 (Accessed 12 February 2008)].

Patrick, A. 2007. Biometrics and Identity Theft. [Available from: http://www.andrewpatrick.ca/ essays/biometrics-and-identity-theft (Accessed 7 February 2008)].

Philippe, H. 2004. SAGEM provides the first biometric system to secure a major banking application. [Available from: www.sagem-ds.com/eng/site.php?spage=03010419 (Accessed 26 March 2008)].

Pronko, N. 1998. Biometrics Protects Government Data. Business Solutions. [Available from: http://www.businesssolutionsmag.com/index.php?option=com_jambozine&layout=article&view=p age&aid=2277&Itemid=5 (Accessed 13 March 2008)].

QuestBiometrics. 2005. Advantages of Biometrics: Why opt for biometric technology? [Available from: http://www.questbiometrics.com/advantages-of-biometrics.html (Accessed 7 February 2008)].

Reynolds, P. 2004. The Keys to Identity. Health Management Technology, 25(12):12-16. [Available from: www.healthmgttech.com/archives/1204/1204the_keys.htm (Accessed 5 February 2008)].

Schneider, J. K. 2005. National health infrastructure prompts need for proper patient identification. Managed Healthcare Executive, 15(8):22-24. Available from: http://managedhealthcareexecutive.

modernmedicine.com/mhe/Hospitals+&+Providers/National-health-infrastructure-prompts-need-for¬pr/ArticleStandard/Article/detail/173306 (Accessed 20 March 2008).

SearchSecurity.com. 2006. Biometrics. [Available from: http://searchsecurity.techtarget.com/ sDefinition/ 0,,sid14_gci211666,00.html (Accessed 7 February 2008)].

Security International. 2002. City of Stockholm School System. Precise Biometrics. [Available from: http://www.security-int.com/categories/fingerprint-identification/city-stockholm-school¬system.asp (Accessed 12 February 2008)].

Sturgeon, W. 2005. 'We want biometrics' say bank customers. [Available from: http://software. silicon.com/security/0,39024655,39130185,00.htm (Accessed 9 April 2008)].

US Department of State. 2004. DoS awards record breaking biometric deal. 2004. Biometric Technology Today, 12(10):1,12.

Wikipedia. 2008a. Fingerprint. [Available from: http://en.wikipedia.org/wiki/Fingerprint# Timeline (Accessed 8 February 2008)].

Wikipedia. 2008b. Ted Bundy. [Available from: http://en.wikipedia.org/wiki/Ted_Bundy (Accessed 8 February 2008)].

Wikipedia. 2008c. Forensic dentistry. [Available from: http://en.wikipedia.org/wiki/Forensic_ dentistry (Accessed 8 February 2008)].

Wikipedia. 2008d. Biometric passport. [Available from: http://en.wikipedia.org/wiki/Biometric_ passport (Accessed 8 February 2008)].

Winton, R. 2004. ID System Gets in Face of Criminals. [Available from: http://pqasb.pqarchiver.com/ latimes/access/770653801.html?dids=770653801:770653801&FMT=ABS&FMTS=ABS:FT&type =current&date=Dec+25%2C+2004&author=Richard+Winton&pub=Los+Angeles+Times&edition= &startpage=B.1&desc=ID+System+Gets+in+Face+of+Criminals%3B+LAP

D+officers+field-test+a+ hand-+held+computer+using+facial+recognition+to+identify+suspects.+Critics+raise+issues+of +privacy+and+reliability (Accessed 12 February 2008).

Yanez, M. & Gomez, A. 2004. ATM & BIOMETRICS: A SOCIO-TECHNICAL BUSINESS MODEL. P7 – 9. University of Miami, School of Business Administration.

Zalud, B. 2003. Facial Makeover. Security, 40:30-32.

THE IP PROTECTION OF ELECTRONIC DATABASES: COPYRIGHT OR COPYWRONG?

Tana Pistorius

University of South Africa

pistot@unisa.ac.za
+27 12 429 8334
Department of Mercantile Law
P.O. Box 392
Pretoria
0003

ABSTRACT

The protection of the intellectual investments embodied in databases is of the utmost importance. Technological innovation has rendered databases vulnerable to unauthorised access, reproduction, adaptation and publication.

The copyright protection of databases is not always adequate to address the protection of non-original databases. Vast collections of data are thus vulnerable to information security threats. The European Union enacted a *sui generis* form of protection for non-original databases. A decade later a review of the first court decisions reveal paltry databases protection. The *sui generis* layer of IP protection in the EU has thus not led to innovation and growth in the European database industry. Courts' restrictions on the protection of "single-source databases" and the interpretation of the substantial investment requirement have contributed to the low level of database right adoption. The action of database owners against deep linking has proved to be much more effective than the database right. South Africa, as developing country, should devise its own strategies to cope with the proliferation of protectionism within the context of the widening digital divide. The database right seems to be "copy wrong" for now.

KEY WORDS

Database right; Copyright protection; originality; sui generis; deep linking

THE IP PROTECTION OF ELECTRONIC DATABASES: COPYRIGHT OR COPYWRONG?

1 INTRODUCTION

Electronic databases are collections of recorded data or information in an electronic or digital form. Databases form the core of information technology. Tremendous resources are often invested to assemble large quantities of information into databases. Still, the resulting products are vulnerable to piracy. Technological innovation has rendered databases vulnerable to unauthorised access, reproduction, adaptation and publication. The possibilities for the creation of recompiled and derived products are beyond the imagination, let alone the knowledge, of the original owner.[1] It has been noted that, from an economic point of view, all electronic databases have two characteristics in common --- "they are costly to produce, but they are easy to reproduce or copy".[2]

2 COPYRIGHT PROTECTION OF DATABASES

Traditional principles of copyright law require a measure of originality in the selection or arrangement of data in a compilation, before it will attract copyright protection. Article 2(5) of the Berne Convention for the Protection of Literary and Artistic Works grants copyright protection to collections of literary or artistic works (such as encyclopaedias and anthologies) which, because of the selection and arrangement of their contents, constitute intellectual creations. This protection arises without prejudice to the copyright in each of the works forming part of such collections. Bare facts cannot be protected by copyright, but compilations of facts are within the

---

[1] Brown, Bryan & Conley 'Database Protection in a Digital World' (1999) 6 *Richmond Journal of Law & Technology* 2.

[2] Nelson 'Recent Development: Seeking Refuge from a Technology Storm: The Current Status of Database Protection Legislation After the Sinking of the Collections of Information Anti-Piracy Act and the Second Circuit Affirmation of *Matthew Bender & Co. v. West Publishing Co* (1999) 6 *Journal of Intellectual Property Law* 453 at 455.

subject matter of copyright protection if these compilations constitute original works of authorship.

Copyright protection has frequently been extended to compilations of non-copyright material because of the labour and skill involved in selecting and arranging the material. For example, protection has been granted to compilations such as a street directory;[3] a list of stock-exchange prices;[4] an alphabetical list of railway stations in a railway guide;[5] a trade catalogue;[6] a racing information service;[7] chronological fixture lists of football clubs;[8] a directory of telefax users;[9] and a catalogue and price list.[10]

Traditional copyright principles require a measure of originality or creativity in the selection or arrangement of data in a compilation, or other indications of creative authorship, for the compilation to attract copyright. The requirement of originality for copyright protection of compilations is interpreted differently in various legal systems. The United Kingdom and Commonwealth courts have favoured the "sweat -of-the-brow" approach to database protection.[11] If an author has expended labour and skill in creating the work, it will enjoy copyright protection, notwithstanding the bland nature of the work.

---

[3] See *Kelly v Morris* (1866) LR 1 Eq 697.
[4] See *Exchange Telegraph Co Ltd v Gregory & Co* [1896] 1 QB 147.
[5] See H Blacklock & Co Ltd v C Arthur Pearson Ltd [1915] 2 Ch 376.
[6] Purefoy Engineering Coy Ld & another v Sykes Boxall & Coy Ld & others (1955) 72 RPC 89 (CA.
[7] See *Portway Press Ld v Hague* [1957] RPC 426.
[8] See *Football League Ltd v Littlewoods Pools* Ltd [1959] Ch 637.
[9] See *Fax Directories (Pty) Ltd v SA Fax Listings CC* 1990 (2) SA 164 (D.
[10] See *Payen Components SA Ltd v Bovic CC & others* 1995 (4) SA 441 (A.
[11] See *Waterlow Publishers Ltd v Rose* The Times 8 Dec 1989; *Waterlow Publishers Ltd v Reed Information Services Ltd* The Times 11 Oct 1990 as quoted by Morton 'Draft EC Directive on the Protection of Electronic Databases: Comfort After *Feist*' (1992) 8 *Computer Law & Practice* 38 at 39 n12; See also Cornish '1996 European Community Directive on Database Protection' (1996-1997) 21 *Columbia-VLA Journal of Law & the Arts* 1 at 2.

Under traditional German copyright principles, most factual databases do not qualify for copyright protection unless their "selection, accumulation and organization" has been the subject of expertise beyond that of the average programmer.[12] In terms of French copyright law, which requires original works to reveal something of the author's own personality, and Dutch Copyright law, most compilations will not enjoy copyright protection[13]

3 THE LEGAL PROTECTION OF DATABASES IN THE EU

The European Union adopted a novel approach in the Council Directive on the Legal Protection of Databases[14] after nearly eight years of deliberation. The Directive provides a two-tier form of protection. It strives to create a harmonised level of copyright protection for "original" databases.[15] A novel "sui generis" right to protect investments in databases was also introduced.[16] Both rights differ in terms of requirements for protection, duration of rights, scope of protection, the exceptions or limitations that apply and the determination of the right holders (both natural and legal).[17]

The Database Directive extends copyright protection to databases that constitute "the author's own intellectual creation" -- databases which evidence some measure of "originality" or "creativity" on the part of the author.[18] Article 5 states that compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such. Article 5 adopts the

---

[12] See *Incassoprogramm* decision of 9 May 1985 of the Federal Supreme Court; See also Pattison [1992] 4 *European Intellectual Property Review* 113 at 113-114)

[13] *Van Dale v Romme* Judgement of 4 January 1991 as quoted by Cornish (1996-1997) 21 *Columbia-VLA Journal of Law & the Arts*; See also Pattison 'The European Commission's Proposal on the Protection of Computer Databases' [1992] 4 *European Intellectual Property Review* 113 at 114 n12-13.

[14] See Council Directive 96/9 of 11 March 1996 On the Legal Protection of Databases 1996 *Official Journal* (L 77) 20 (hereinafter the 'Database Directive'))

[15] See Articles 3-5.

[16] See Articles 7, 10 and 11.

[17] See Articles 6, 8, 9 and 15.

[18] See recital 15 and art 3(1).

approach of the American Supreme Court's decision in *Feist Publications Inc v Rural Telephone Services Co*[19] in which it was held that only the selection or arrangement of a compilation of facts, and not the facts themselves, can be protected under copyright. The Database Directive rejected the traditional approach of the United Kingdom and Ireland and raised the threshold for copyright protection.[20]

The approach chosen in the Directive was to harmonise the threshold of "originality". Those "non-original" databases that did not meet the threshold would be protected by a newly created right. A high standard for originality, akin to that of *droit d'auteur* countries were adopted. This new standard of originality had the effect of protecting fewer databases by copyright (which was now limited to so-called "original" databases).[21] Those databases that fell below the originality bar, but which were created through substantial investment attained a "sui generis" form of protection.

This database right prevents the extraction and reutilisation of the whole or a substantial part of the contents of a non-original database. While "original" databases require an element of "intellectual creation", "non-original" databases are protected as long as there has been "qualitatively or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents" of a database.[22] The "sui generis" right is a Community creation with no precedent in any international convention and no other jurisdiction has adopted the sui generis right. The distinction between "original" and "non-original" databases is alos unique to the European Union.[23]

---

[19] 499 US 340 (1991) at 344—348.
[20] See Brown, Bryan & Conley 1999 *Richmond Journal of Law & Technology* text at n 135.
[21] Commission of the European Communities "First Evaluation Of Directive 96/9/Ec On The Legal Protection Of Databases" *DG Internal Market And Services Working Paper* 12 Dec 2005 available at http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf (accessed 23 April 2008) (hereafter "EU First evaluation") at 3.
[22] *Ibid*.
[23] *Idem* at 4.

In essence, the Directive sought to create a legal framework that would establish the ground rules for the protection of a wide variety of databases in the information age. It did so by giving a high level of copyright protection to certain databases ("original" databases) and a new form of "sui generis" protection to those databases which were not "original" in the sense of the author's own intellectual creation ("non-original" databases).[24] The effect of the Database Directive has recently been evaluated.[25]

All 25 Member States have transposed the Directive into national law.[26] National jurisprudence evidences the adoption of a wide notion of the term "database", embracing listings of telephone subscribers; compilations of case law and legislation; websites containing lists of classified advertisements; catalogues of various information and lists of headings of newspaper articles under its ambit. The European Court of Justice (ECJ) has also embraced a broad interpretation of the definition of "database" in the Directive.[27]

---

[24] *Idem* at 3.

[25] Article 16 of the Database Directive requires the Commission to submit to the European Parliament, the Council and the European Economic and Social Committee a "report on the application of this Directive, in which, *inter alia,* on the basis of specific information supplied by the Member States, it shall examine the application of the sui generis right...this right has led to abuse of a dominant position or other interference with free competition which would justify appropriate measures being taken, including the establishment of non-voluntary licensing arrangements. Where necessary, it shall submit proposals for adjustment of this Directive in line with developments in the area of databases".

[26] EU First Evaluation at 4 notes that Germany, Sweden and the United Kingdom met the deadline of implementation (1 January 1998); Austria and France adopted laws during the course of 1998 whose provisions apply retro-actively from 1 January of the same year. Belgium, Denmark, Finland and Spain implemented in 1998; Italy and the Netherlands in 1999; Greece and Portugal in 2000; Ireland and Luxembourg in 2001. Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia implemented between 1999 and 2003. The EEA countries (Iceland, Lichtenstein and Norway) have also implemented the Directive.

[27] See Case C-444/02 (*Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE - "OPAP"*) n 20, 25.

4 CASE LAW IN THE EU

4.1 Substantial investment

The sui generis provisions of the Database Directive protect the contents of any non-copyrightable database that is the product of substantial investment in obtaining, verifying, or presenting the database's contents.[28] There are no specific standards for determining the substantiality of an investment. The test is quantitative as well as qualitative in nature.[29] The investment may concern the obtaining, verification, or presentation of the content.[30] Not every compilation of information will be considered a "database" for the purpose of the sui generis right. To qualify for protection, a database must be "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means".[31]

The precise meaning of the term “substantial investment” as contained in Article 7 of the Directive has become the focal point of the textual ambiguities of the “sui generis” right. On the one hand, the cost of collecting and maintaining up-to-date information concerning several thousands of real estate properties was held to be a “substantial investment” by a district court of The Hague.[32] On the other hand, the district court of Rotterdam regarded newspaper headlines as a mere “spin-off” of newspaper publishing and the court therefore held that it did not reflect a “substantial investment”.[33] “Spin-off” databases are databases that are by-products of a main or principal activity. Where the database is a single source database it is normally regarded as a spin-off database. In certain Member States, notably the Netherlands, the “spin-off” theory forms a bar against “sui generis” protection for “spin-off” databases.[34]

---

[28] See Recital 39 and art 7(1)).
[29] See Brown, Bryan & Conley op cit text at n150.
[30] See Cornish op cit at 8.
[31] See Brown, Bryan & Conley op cit text at n153.
[32] *NVM v. De Telegraaf*, judgment of 12 September 2000.
[33] *Algemeen Dagblad a.o. v. Eureka*, judgment of 22 August 2000.
[34] See EU First Evaluation at 12.

"Deep-linking" through search engines are another source of divergent case-law.

A deep link is a special for of linking which enables the user to access content on an internal page of a web site, bypassing the home page of the web site.[35] In some cases, the heading, the Internet address (URL) and a brief summary of a press article have been held not to constitute a substantial part of a database[36] and the hyper linking of headings of press articles has been held not to infringe the owner's "sui generis" right.[37] However, in most cases the systematic bypassing of the homepage of the database maker (including banner advertisements) was found to be an infringement of the database maker's "sui generis" right.[38]

A Danish court in *Danish Newspaper Organization v Newsbooster*[39] held so-called "deep linking" is a breach of copyright. The case was brought by the Danish Newspaper Organisation (DNO) against the Newsbooster service, which linked to articles on 28 of the plaintiff's news websites without going through their home pages. The court held that the newspaper articles were copyrightable works. The court held as follows:

"The text collections of headlines and articles, which make up some Internet media, are thus found to constitute databases enjoying copyright protection pursuant to section 71 of the Danish Copyright Act. Under section 71(1) of the Act, the makers of the databases, i.e. the Principals, have the exclusive right protected by the said provision."

On liability for linking, the court held that by means of its search engine, Newsbooster offers its users regular relevant headlines with deep links to articles on Newsbooster's website or in Newsbooster's electronic

---

[35] See Ebersöhn "Hyperlinking and deep-linking" Vol II part 2 *Juta's Business Man's Law* 73-74.

[36] See High Regional Court Cologne, 27 October 2000; District Court Munich, 1 March 2002

[37] See judgment by the German Federal Court of Justice, 18 July 2003 ("*Paper Boy*").

[38] See "*Berlin Online*" – District Court Berlin 8 October 1998; "*Süddeutsche Zeitung*" – Landgericht Köln 2 December 1998.

[39] *Danish Newspaper Publishers' Association v Newshooter.com*; ApS Copenhagen Court, 24 June 2002; Court Journal No F1-8703/2002.

newsletters. These links need to be supplemented and updated on a regular basis and consequently, Newsbooster' s search engine needs to crawl the websites of the Internet media frequently for the purpose of registering headlines and establishing deep links in accordance with the search criteria defined by the users. As a result, Newsbooster repeatedly and systematically reproduces and publishes the Principals' headlines and articles. Newsbooster has a commercial interest in this business and this activity is in conflict with section 71(2) of the Danish Copyright Act.

The court ruled that Newsbooster is prohibited from offering a search service with deep links from the websites newsbooster.dk and newsbooster.com directly to the plaintiffs' news articles; reproducing and publishing headlines from the Internet versions of newspaper articles; distributing electronic newsletters with deep links directly to the newspaper articles; and reproducing and distributing headlines from the newspapers.[40] A similar ruling was made in *Copiepresse v Google Inc*.[41]

Four cases concerning single-source databases of sports information in the areas of football and horseracing have been referred to the ECJ. The cases were referred from national courts in Greece, Finland, Sweden and the United Kingdom. The ECJ gave its judgments in these cases on 9 November 2004.[42] With respect to the extensive lists of runners and riders drawn up by the *British Horseracing Board* (the "*BHB*") in its function as the governing body for the British horseracing industry, the ECJ simply stated that:

---

[40] The quotations from the court's ruling were obtained from "Translation of pages 29 - 42 of the ruling made by the Bailiff's Court on 5 July 2002 at http://www.newsbooster.com/?pg' judge&lan' eng) accessed on 22 July 2007.

[41] Court of First Instance, Brussels, 5 September 2006. A copy of the decision is available at http://www.chillingeffects.org/international/notice.cgi?action=image_7796 (as at 9 Oct 2006); *Contra Algemeen Dagblad BV et al v Eureka Internetdiensten* 2000 District Court of Rottendam) discussed by Ebersöhn vol 11 Part II *Juta's Business Law* at 76.

[42] Cases C-46/02 (*Fixtures Marketing Ltd v Oy Veikkaus Ab*); C-203/02 (*The British Horseracing Board Ltd and Others v William Hill Organisation Ltd*); C-338/02 (*Fixtures Marketing Limited v. AB Svenska Spel*) and C-444/02 (*Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairouAE - "OPAP"*) available at www.curia.eu.int (accessed 23 April 2008).

"The resources used to draw up a list of horses in a race and to carry out checks in that connection do not constitute investment in the obtaining and verification of the contents of the database in which that list appears"

4.2 Obtaining or creating data for database

The ECJ thus distinguishes between the resources used in the "creation" of materials that make up the contents of a database and the *obtaining* of such data in order to assemble the contents of a database. Only the latter activity is protected under the "sui generis" right. This leaves little protection for bodies like the *BHB,* which "create" the data that makes up the contents of their database. Arguably, other industries like the publishers of directories, listings or maps, remain protected as long as they do not "create" their own data but *obtain* these data from others. The ECJ distinction between "creation" and *obtaining* of data means that sports bodies such as the *BHB* cannot claim that they *obtained* the data within the meaning of the Directive. Therefore, such bodies cannot license their own data to third parties.[43]

While going against the Commission's original intention of protecting "non-original" databases in a wide sense, the judgements have the merit of pointing to the serious difficulties raised by attempting to harmonise national laws by recourse to untested and ambiguous legal concepts ("qualitatively or quantitatively substantial investments in either the obtaining, verification or presentation of contents").

The ECJ's judgment would probably apply to the databases created by broadcasting organisations for the purposes of scheduling programmes: they would not be able to assert a "sui generis" right in the contents of such databases. In addition, the European Court ruled that on-line betting activities on football matches and horse races carried out by betting companies such as *Svenska Spel* or *William Hill* was not infringing in nature. The Court noted that such use did not affect the whole or a substantial part of the contents of the plaintiffs' databases, and they

[43] Müller & Munz "Recent Case Law from Germany Concerning the Database Right" Vol 12 (No 2) 2007 *Communications Law* at 70-71.

therefore did not prejudice the substantial investment of the latter in the creation of their databases.

In *British Horse Racing Board v. William Hill*[44] the British Court dismissed the *BHB*'s arguments aimed at showing that its database was protectable by the "sui generis" right under Article 7(1) of the Directive. The court held that the scope of the "sui generis" protection does not include the "creation" of the underlying data.[45] A soccer fixture list would usually not be protected under the "sui generis" right.

5 THE POSITION OBTAINING IN SOUTH AFRICA

In terms of section 1(1) of the Copyright Act[46] the definition of a literary work includes tables and compilations, including tables and compilations of data stored or embodied in a computer or a medium used in conjunction with a computer. This clearly includes electronic databases. The South African legislature has thus opted for the protection of electronic databases as a form of compilation, which is a species of literary work.[47]

Dean[48] submits that under South African law an electronic database, like any other work, should be "original". No higher standard or level of creativity is required. As noted above, in US law, a minimal degree of creativity or so-called "creative spark" is required to satisfy the originality requirement. In South Africa, on the other hand, creativity is not required to make a work original – the so-called "sweat of the brow" is sufficient. The requirement of originality is satisfied solely by the fact that the contents of a

---

[44] Case No: A3/2001/0632 *The British Horseracing Board Limited; The Jockey Club; Weatherbys Group Limited and William Hill Organization Limited*.

[45] For example, the national football bodies establish the annual "football calendar" by pairing the teams, setting up home and away matches. It comprises the basic activity of organising soccer tournaments, involves the "creation" of data. The collection and verification of the data in order to set up the fixture list is only a by-product of this basic activity, but the by-product requires relatively little investment.

[46] Act 98 of 1978 as amended by section 50(*e*) of the Intellectual Property Laws Amendment Act 38 of 1997.

[47] See Dean O. H. (2003). *Handbook of South African Copyright Law* (Revision service 11) Johannesburg, Juta & Co Ltd at 1-8 to 1-8A, 1-14.

[48] Dean Handbook at 1-8A.

particular compilation must have been independently collected through the author's own skills or labour, and not copied from another.[49] In *Haupt t/a Softcopy v Brewers Marketing Intelligence (Pty)* Lt[50]d (supra) Streicher JA confirmed that, as our present Copyright Act originated from UK law, creativity is not a requirement for copyright in SA law. The court then confirmed the test for originality in SA copyright law to be as follows:

"Save where specifically provided otherwise, a work is considered to be original if it has not been copied from an existing source and if its production required a substantial (or not trivial) degree of skill, judgment or labour."[51]

The "sweat of the brow" doctrine is still firmly entrenched in South African copyright law.

Electronic databases were protected by copyright prior to the 1997 Amendment Act, as the material embodiment requirement could be met by digital embodiment. The South African database owner is in an advantageous position: the originality requirement is set so low that both original and non-original databases qualify for protection. The Database Directive has not been an outstanding success and the repeal of the sui generis database right has even been proposed.[52] International instruments are not likely to follow.

Stone and Kernick[53] note that a comprehensive database, which contains the entire universe of relevant data, may be commercially useful, but is not copyrightable, as "selection" requires the exercise of creative judgment in culling facts, and not using the relevant universe. In essence, this amounts to the protection of means to access information. Policy considerations underlying the regulation of access to information and access to knowledge should be heeded. It can never been seriously proposed that

---

[49] See *Waylite Diary CC v First National Bank Ltd*.

[50] (2006) 4 SA 458 (SCA).

[51] *Supra* at 473A-B

[52] See conclusion to the EU First Evaluation.

[53] Stone & Kernick 'Protecting Databases: Copyright? We don't Need No Stinkin' Copyright' (1999) 16 *The Computer Lawyer* 17.

information itself should be protected (except by the law regarding trade secrets).

Information technology has become an indispensable development tool, and a crucial means of information and knowledge exchange.[54] Electronic databases are the tools that provide information about information; they are regarded as the new building blocks of knowledge.[55] Their importance cannot be too heavily underscored as they form the core of information technology and all information systems.[56] The copyright protection of such comprehensive databases remains problematic.

This may be especially problematic for digital databases such as those accessed through the Internet, since their very appeal is their all-inclusiveness.[57] Copyright law has emerged as one of the most forceful means of regulating the flow of ideas and knowledge-based products.[58] "Sui generis" protection comes close to protecting data as property. There is a long-standing principle that copyright should not be extended to cover basic information or "raw" data. However, as evidenced by the ECJ's differentiation between the "creation" of data and its *obtaining* demonstrate, the "sui generis" right comes precariously close to protecting basic information.[59]

---

[54] Sun "Copyright law under siege: An inquiry into the legitimacy of copyright protection in the context of the global divide" 2005 (36) *International Review of Industrial Property and Copyright Law* 192.

[55] Pistorius "Copyright in the Information Age: The Catch-22 of Digital Technology" 2006 (2) *Critical Arts* 47 at 54.

[56] Bastian 'Protection of 'Noncreative' Databases: Harmonization of United States, foreign and international law' (1999) 22 *Boston College Environmental Affairs Law Review* 425 at 426; Lavenue 'Database rights and technical data rights: the expansion of intellectual property for the protection of databases' 38 (1997) *Santa Clara L Rev* 1.

[57] See Brown, Bryan & Conley *Richmond Journal of Law & Technology*, text at note 93.

[58] See Sun 2005 op cit IIC at 211.

[59] See Fieldhouse & Bolton "Copyright? Wrong! – Copyright protection of computer programs as literary works" 2003 *Copyright World* 22 at 25.

South Africa, as developing country, should devise its own strategies to cope with the proliferation of protectionism within the context of the widening digital divide.[60]

6 EVALUATION OF THE VALUE OF THE DATABASE RIGHT

Introduced to stimulate the production of databases in Europe, the "sui generis" protection has had no proven impact on the production of databases.[61] Nevertheless, as the figures discussed below demonstrate, there has been a considerable growth in database production in the US, whereas, in the EU, the introduction of "sui generis" protection appears to have had the opposite effect. With respect to "non-original" databases, the assumption that more and more layers of IP protection means more innovation and growth appears not to hold up.[62]

7 CONCLUSION

It has been noted that there is a risk that national courts applying the European Court's case law will conclude that relatively little of the investment in establishing a database appears to have been in collecting and verifying the information displayed on a website containing data on e.g. real estate or job advertisements.[63] On the other hand, the ECJ's narrow

---

[60] Pistorius "Developing Countries and Copyright in the Information Age – The Functional Equivalent Implementation of the WCT″ 1-27 (2) *Potchefstroom Electronic L. J.* (2006) *at:* http://www.puk.ac.za/opencms/export/PUK/html/fakulteite/regte/per/issues/2006_2__Pistorius_art.pdf (accessed Feb. 28 2007).

[61] According to the Gale Directory of Databases, the number of EU-based database "entries" was 3095 in 2004 as compared to 3092 in 1998 when the first Member States had implemented the "sui generis" protection into national laws (EC First consultation at 20). It is noteworthy that the number of database "entries" dropped just as most of the EU-15 had implemented the Directive into national laws in 2001. In 2001, there were 4085 EU-based "entries" while in 2004 there were only 3095 (EU First consultation at 20). The "sui generis" right has helped Europe to catch up with the US in terms of investment but, at the same time, that the "sui generis" right did not help to significantly improve the global competitiveness of the European database sector. The data taken from the *GDD* reveal that the economic gap with the US has not been reduced (EC First consultation at 23).

[62] EU First Evaluation at 24.

[63] EU First Evaluation at 20.

interpretation of the "sui generis" protection for "non-original" databases where the data were "created" by the same entity as the entity that establishes the database would put to rest any fear of abuse of a dominant position that this entity would have on data and information it "created" itself (so-called "single-source" databases).[64]

The interpretation of the ECJ may also allay the fear of those who believed that the Directive would lock up information otherwise publicly available, at least with respect to those databases which contain data "created" by the database maker himself.[65] It is noteworthy that the ECJ and some national judges appear to fear that the balance between users and rightholders is inappropriate. Indeed, the interpretation adopted by the European Court may have been influenced by the concern that the "sui generis" right might otherwise significantly restrict access to information. Thus, for instance, the ECJ has ruled that the mere act of consultation of a database is not covered by the database maker's exclusive rights.[66]

As Brown, Bryan and Conley[67] so eloquently put it:
"Sweat equity is all that is left".

8 REFERENCES

Bastian 'Protection of 'Noncreative' Databases: Harmonization of United States, foreign and international law' (1999) 22 *Boston College Environmental Affairs Law Review* 425

---

[64] *Ibid*; See also Müller & Munz op cit 2007 *Communications Law* 70-71.

[65] EU First Consultation at 22.

[66] "However, it must be stressed that the protection of the *sui generis* right concerns only acts of extraction and re-utilisation as defined in Article 7(2) of the directive. That protection does not, on the other hand, cover consultation of a database. Of course, the maker of a database can reserve exclusive access to his database to himself or reserve access to specific people. However, if he himself makes the contents of his database or a part of it accessible to the public, his *sui generis* right does not allow him to prevent third parties from consulting that base", case C-203/02, n. 54, 55.

[67] 1999 *Richmond Journal of Law & Technology*, text at note 205.

Brown, Bryan & Conley 'Database Protection in a Digital World' (1999) 6 *Richmond Journal of Law & Technology* 2

Commission of the European Communities "First Evaluation Of Directive 96/9/Ec On The Legal Protection Of Databases" *DG Internal Market And Services Working Paper* 12 Dec 2005 available at http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf

Cornish '1996 European Community Directive on Database Protection' (1996-1997) 21 *Columbia-VLA Journal of Law & the Arts* 1

Council Directive 96/9 of 11 March 1996 On the Legal Protection of Databases 1996 *Official Journal* (L 77) 20 (hereinafter the 'Database Directive'))

Dean O. H. (2003). *Handbook of South African Copyright Law* (Revision service 11) Johannesburg, Juta & Co Ltd

Ebersöhn "Hyperlinking and deep-linking" Vol II part 2 *Juta's Business Man's Law* 73

Fieldhouse & Bolton "Copyright? Wrong! – Copyright protection of computer programs as literary works" 2003 *Copyright World* 22

Lavenue 'Database rights and technical data rights: the expansion of intellectual property for the protection of databases' 38 (1997) *Santa Clara L Rev* 1

Morton 'Draft EC Directive on the Protection of Electronic Databases: Comfort After *Feist*' (1992) 8 *Computer Law & Practice* 38

Müller & Munz "Recent Case Law from Germany Concerning the Database Right" Vol 12 (No 2) 2007 *Communications Law* at 70

Nelson 'Recent Development: Seeking Refuge from a Technology Storm: The Current Status of Database Protection Legislation After the Sinking of the Collections of Information Anti-Piracy Act and the Second Circuit Affirmation of *Matthew Bender & Co. v. West Publishing Co* (1999) 6 *Journal of Intellectual Property Law* 453

Pattison 'The European Commission's Proposal on the Protection of Computer Databases' [1992] 4 *European Intellectual Property Review* 113.

Pistorius "Developing Countries and Copyright in the Information Age – The Functional Equivalent Implementation of the WCT″ 1-27 (2) *Potchefstroom Electronic L. J.* (2006) *at:*

http://www.puk.ac.za/opencms/export/PUK/html/fakulteite/regte/per/issues/2006_2__Pistorius_art.pdf
Pistorius "Copyright in the Information Age: The Catch-22 of Digital Technology" 2006 (2) *Critical Arts* 47
Stone & Kernick 'Protecting Databases: Copyright? We don't Need No Stinkin' Copyright' (1999) 16 *The Computer Lawyer* 17
Sun "Copyright law under siege: An inquiry into the legitimacy of copyright protection in the context of the global divide" 2005 (36) *International Review of Industrial Property and Copyright Law* 192
Court cases:
"*Berlin Online*" – District Court Berlin 8 October 1998
Algemeen Dagblad a.o. v. Eureka, judgment of 22 August 2000
Algemeen Dagblad BV et al v Eureka Internetdiensten 2000 District Court of Rottendam)
C-203/02 (*The British Horseracing Board Ltd and Others v William Hill Organisation Ltd*)
C-338/02 (*Fixtures Marketing Limited v. AB Svenska Spel*)
Case C-444/02 (*Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE -"OPAP"*)
Case No: A3/2001/0632 *The British Horseracing Board Limited; The Jockey Club; Weatherbys Group Limited and William Hill Organization Limited*
Cases C-46/02 (*Fixtures Marketing Ltd v Oy Veikkaus Ab*)
Court of First Instance, Brussels, 5 September 2006. A copy of the decision is available at http://www.chillingeffects.org/international/notice.cgi?action=image_7796 (as at 9 Oct 2006)
Dale v Romme Judgement of 4 January 1991
Danish Newspaper Publishers' Association v Newshooter.com; ApS Copenhagen Court, 24 June 2002; Court Journal No F1-8703/2002.
District Court Munich, 1 March 2002
Exchange Telegraph Co Ltd v Gregory & Co [1896] 1 QB 147.
Fax Directories (Pty) Ltd v SA Fax Listings CC 1990 (2) SA 164 (D.
Football League Ltd v Littlewoods Pools Ltd [1959] Ch 637.
H Blacklock & Co Ltd v C Arthur Pearson Ltd [1915] 2 Ch 376.
High Regional Court Cologne, 27 October 2000
Incassoprogramm decision of 9 May 1985 of the Federal Supreme Court

Judgment by the German Federal Court of Justice, 18 July 2003 ("*Paper Boy*").
Kelly v Morris (1866) LR 1 Eq 697.
NVM v. De Telegraaf, judgment of 12 September 2000.
Payen Components SA Ltd v Bovic CC & others 1995 (4) SA 441 (A.
Portway Press Ld v Hague [1957] RPC 426.
Purefoy Engineering Coy Ld & another v Sykes Boxall & Coy Ld & others (1955) 72 RPC 89 (CA.
"*Süddeutsche Zeitung*" – Landgericht Köln 2 December 1998.
Waterlow Publishers Ltd v Reed Information Services Ltd The Times 11 Oct 1990
Waterlow Publishers Ltd v Rose The Times 8 Dec 1989;

THE PRINCIPLE OF SECURITY SAFEGUARDS: ACCIDENTAL ACTIVITIES

Rasika Dayarathna

Department of Computer and Systems Sciences (DSV), Stockholm University/Royal Institute of Technology
Forum 100; 164 40 Kista, Stockholm- Sweden.
si-ika@dsv.su.se

ABSTRACT

The principle of information security safeguards is a key information principle contained in every privacy legislation measure, framework, and guideline. This principle requires data controllers to use an adequate level of safeguards before processing personal information. However, privacy literature neither explains what this adequate level is nor how to achieve it. Hence, a knowledge gap has been created between privacy advocates and data controllers. This paper takes a step to bridge the aforementioned knowledge gap by presenting an analysis of how data protection and privacy commissioners have evaluated the level of adequacy of security protection given to personal information in selected privacy invasive cases. This study addresses security measures used to protect personal information against accidental incidents. This analysis also lays a foundation for building a set of guidelines for data controllers on designing, implementing, and operating both technological and organizational measures used to protect personal information.

KEY WORDS

Information privacy, information security, accidental disclosure, accidental loss, personal information.

THE PRINCIPLE OF SECURITY SAFEGUARDS:

ACCIDENTAL ACTIVITIES

1 INTRODUCTION

Privacy principles are the basic building blocks of privacy standards which include privacy directives, legislation measures, guidelines, frameworks and industry best practices. One of the key information privacy principles is information security safeguards. According to the Organisation for Economic Cooperation and Development (OECD), the principle of security safeguards states that "personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data"(OECD, 1980). The EU Directive 95/46/EC mentions this principle in Articles 17 and 25. Article 17 prohibits the processing of personal information without providing an adequate level of protection for personal information and Article 25 prohibits the transferring of personal information to a third country that does not have an adequate level of protection for personal information. Every year a significant number of complaints pertaining to the violation of this principle are received by privacy and data protection commissioners. For example, during the period of 2004-05, the Hong Kong privacy commissioner received 131 complaints; this amounts to 14% of the total cases (HG-Annual Report, 2006). This paper presents the measures suggested by privacy and data protection commissioners for the protection of personal information against accidental incidents.

What does this principle state? Is it synonymous with information security? According to the explanatory notes of the OECD's privacy guidelines (OECD, 1980), privacy and security are two different things. Information security focuses on providing confidentiality, availability, and integrity to informational assets of organizations. In contrast, the principle of security safeguards in information privacy focuses on achieving a "reasonable" or "adequate" level of protection, not "perfect" or "maximum" protection for personal information of natural persons. Natural persons include customers, employees, employers, and other stakeholders.

Personal information often falls into organizational information assets. However, personal information belongs to outsiders who have given their personal information to organizations for specific purposes and time period. Therefore, this author argues that extra care must be taken in respect to personal information. In order to provide better protection, there should be appropriate security standards. According to Iachello (2003), existing multi-national and domestic security standards and best practices have not sufficiently covered information privacy aspects.

Another conflicting aspect is that information security heavily focuses on protecting informational assets from external parties. However, reported cases have shown that a large number of information privacy threats are posted by insiders including organizations themselves (Muelle & Rannenberg, 1999). As a result of focusing heavily on outsider attacks, the current evaluation schemes do not provide adequate attention for multilateral security. In certain cases, it can be seen that information security and privacy lead to conflicting situations. For example, some information security requirements such as keeping backups in many locations or monitoring employees' activities conflict with information privacy requirements. In addition to that, information privacy legislation measures give certain inalienable rights to data subjects such as accessing personal information and making corrections (Opinion 1/98, 1998). Unless data controllers take appropriate measures, exercising these rights threatens information assets. Another conflicting issue is that information security entails and eagerness to acquire more personal information, but legal privacy legislation measures prohibit excessive use of personal information. For instance, the use of fingerprints found on a student's canteen was considered to be privacy invasive by the Swedish data protection commissioner.

1.1 Advantage

There is a dilemma that states technologists can not precisely understand what legal advocates and legislators say (Dempsey & Rubinstein, 2006). One aim of this study is to present legal privacy requirements imposed in the principle of security safeguards in an understandable manner to technologists. It is also expected that this will assist technologists to precisely understand their legal privacy obligations in designing and operating information systems. The main aim of this paper is to understand

the notions of ‘adequate’ or ‘reasonableness’ mentioned in privacy standards. This understanding is necessary for choosing appropriate organizational and technological measures for protecting personal information.

1.2 Methodology

Without defining what ‘adequate’ or ‘reasonable’ means in privacy legislation measures, competent bodies are given a mandate to decide whether a measure is adequate/reasonable or not. This paper followed the Common Law tradition, which analyzes and interprets previously given decisions and judgments by legal authorities in judging a present case. It is expected that analyzing and interpreting verdicts given by data protection and privacy commissioners sheds lights on understanding what an adequate or reasonable level is. There are cases that fall into one or more information privacy principles. The criterion used to identify whether a case relates to this principle is an allegation that an event occurred due to the lack of appropriate organizational and technological measures.

1.3 Materials and Methods

This study covers some national data protection legislation measures, regional directives, privacy guidelines, and frameworks introduced by leading privacy organizations and verdicts given by selected data protection and privacy commissioners. The studied directives, frameworks, and guidelines are Article 17 of EU Data Protection Directive 95/46/EC, Principle 7 of the Asia-Pacific Economic Cooperation (APEC) Privacy Framework (APEC, 2005), Section 16 of the APEC Privacy Charter (Greenleaf & Waters, 2003), Article 7 of the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CECPI/APPD) ¬Convention No 1981, the AICPA/CICA Privacy Framework (AICPA/CICA, 2004) introduced by the American Institute of Certified Public Accountants (AICPA), the Canadian Institute of Chartered Accountants (CICA), and OCED Privacy Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (OECD, 1980). The EU Directive and CECPI/APPD cover the whole Europe and APEC Privacy Framework Charter cover some Asia-Pacific countries while the OECD covers a large number of industrialized countries. The

AICPA/CICA Privacy Framework is the base for the 'WebTrust' web privacy seal, which is one of the leading online privacy seals.

First, the principle of information security was analyzed and functional requirements imposed by the principle were identified. In this examination, six privacy directives, legislation measures, and frameworks were used. Second, verdicts given by data protection and privacy commissioners were scrutinized to identify underlying privacy threats. Based on the identified threat, a verdict was placed under one of the identified functional requirements derived in stage 1. Then, the recommended organizational or technological measures were identified and presented accordingly. There are certain cases where the commissioners have not suggested protection measures. In those cases, appropriate measures are suggested to give a rough picture of possible solutions. However, these measures are not comprehensive. In addition, some expected privacy threats are given for cases where there is no involvement of ICT but may be interpreted similarly.

1.4 Analysis

Seven legal privacy threat categories were identified in analyzing the principles: accidental loss and disclosure, unauthorized access, use, destruction, alteration and disclosure. The criterion applied in differentiating accidental and unauthorized activities is discussed in Section 1.5.

Table 1 shows the high-level requirements imposed in the principle. On the horizontal axis, high-level requirements are given under three categories: accidental, unauthorized, and others. The first category, accidental, covers all kinds of accidental privacy breaches which include access, use, destruction, loss, alterations, and disclosures. The second category covers unauthorized access, use, destruction, alterations, and disclosure. The last category covers any other kind of misuse. For example, the EU Directive 95/46/EC states that measures should be taken as a protection from all other unlawful forms of processing. The vertical axis provides the identification of the studied privacy literature (data protection directives, frameworks, charter, and guidelines). When a high-level requirement is explicitly mentioned in a given piece of literature, the corresponding box is marked with 'Y'; otherwise, it is left blank. Cases

where high-level requirements are given in a similar term are presented with superscripts and discussed in the legend.

Table 1: High-level requirements imposed in the principle of security of safeguards according to the studied international privacy literature.

	Accidental						Unauthorized					Other
	Access	Use	Destructio	Loss	Alteration	Disclosure	Access	Use	Destructio	Alteration	Disclosure	
CECPI/APPD			Y	Y			Y	Y	Y	Y	Y	
OECD				Y			Y	Y	Y	Y[3]	Y	
EU DPA			Y	Y	Y		Y		Y[4]	Y	Y	Y[5]
APEC Charter	Y	Y		Y	Y[1]	Y	Y	Y		Y[3]		Y
APEC Privacy				Y			Y	Y	Y	Y[3]	Y	Y
AICPA				Y			Y	Y[2]	Y	Y	Y	

Superscripts stand for: Y1 accidental modification, Y2 misuse, Y3 unauthorized modification, Y4 unlawful destruction and Y5 all other unlawful forms of processing. APEC Charter and APEC Privacy stand for the APEC Privacy Charter and the APEC Privacy Framework respectively.

All studied literature emphasizes the accidental loss component of the principle. Only the European literature mentioned accidental destruction. The EU Directive mentions "alteration" without specifying whether it refers to accidental or unauthorized alteration. The predecessor of the APEC Privacy Framework, APEC Privacy Charter, specifically mentions accidental access, use, and disclosure. Taking the above points into account, the accidental threat category has been divided into accidental loss and destruction and accidental disclosure.

1.4.1 Factors affecting the adequate level of protection

Legal privacy literature presents factors that affect adequate levels of protection. Table 2 presents the factors given in EU Directive 95/46/EC, the Canadian Personal Information Protection and Electronic Documents Act, the old Swedish Data Protection Ordinance, and the PISA (Privacy Incorporated Software Agents) project documentation. The old Swedish Data Protection Ordinance and the PISA documentation have defined privacy risk classification schemes based on these factors.

Table 2: Factors affecting the level of reasonableness as defined in the principle of security safeguards

	EU Directive	Canadian Law	PISA	Swedish Ordinance
Nature of PI	Yes	Yes	Yes	Yes
Cost factor	Yes			
State of the art	Yes			
Processing risk	Yes		Yes	
Amount of PI		Yes	Yes	
Distribution		Yes		
Format		Yes		
Storage method		Yes		

PI stands for personal information. According to Article 8 of the Directive 95/46/EC, sensitive data are racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, health, and sex life. The Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) states the sensitivity of personal data depends on the context. However, it gives medical records and income records as examples for sensitive personal data. According to privacy legislation measures, sensitive personal information should be given additional protection.

Data protection legislation measures insist on taking organizational and technological measures to protect personal information, but they do not mention what those measures are. However, Section 4.7.3 of the PIPEDA sheds lights on those measures. Instead of giving precise definitions, the act

gives some examples. Examples given for physical access are locked filing cabinets and restricted access to offices; organizational measures are security clearances and limiting access on a "need-to-know" basis; technological measure are the use of passwords and encryption. Based on the above, these measures can be explained. Organizational measures cover all administrative measures such as drafting policies, recruiting people, allocating resources, and providing training with special focus on data protection. Physical measures to cover access to office premise and locations where information system and personal information reside. Technological measures include all measures used to protect personal information and system based on data and software.

1.4.2 Accidental and Unauthorized Activities

Some privacy literature discusses accidental disclosure under the heading of unauthorized disclosure. However, there is a marginal gap between accidental and unauthorized activities. The criteria applied in this paper for distinguishing accidental activities from unauthorized activities are the intention and motive of parties involved. The violation of explicit instructions to follow certain procedures or not to perform certain activities falls into unauthorized activities. In addition to that, performing with the knowledge of the negative consequences that could result from an activity and deliberately neglecting to implement appropriate preventive measures come under the unauthorized category. Accidental activities are mainly due to human errors. Privacy invasive activities carried out with innocent mind and without knowing negative consequences fall into the accidental category. This distinction is important for designing and developing protection measures.

2 ACCIDENTAL ACTIVITIES

Accidental activities take two forms: accidental loss and accidental disclosure of personal information. Section 2.1 and Section 2.2 are dedicated to the accidental loss and the accidental disclosure of personal information respectively.

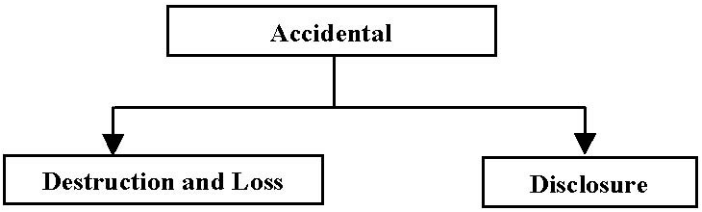


Figure 1: Accidental activities

2.1 Accidental Destruction and Loss

This sub-section discusses the accidental destruction of physical media and the accidental loss of physical media. In the former case, the physical device is with custody but data cannot be recovered. In the second case, the physical media is not in the custody.

It is the duty of data controllers to take appropriate measures to protect physical devices from accidental destruction. This is very important in information privacy since data controllers are required to maintain up-to-date personal information. This position is also stressed in Article 6 of the EU Directive 95/46/EC. The physical storage media range is from papers to cutting edge storage devices. The Australian privacy commissioner stated that the ACT Department of Corrective Services had not taken precautionary measures against deterioration of thermal papers (Privacy Commissioner Tenth Annual Report- Australia, 1998).

There were no cases reported in the studied literature on accidental destruction of technological storage devices. However, all information security standards state the importance of protecting physical storage media from accidents. Common measures are taking backups and storing them in protected places.

According to the OECD privacy guidelines, another cause for losing personal information is the loss of physical devices that contain personal information. The loss of physical devices takes two forms. One is revealing personal information contained in the media and the other one is the mere loss of physical media. Information privacy professionals are keen on the loss of physical media that may reveal personal information. For example, when it is very clear that the motive of a theft is the monitory value of the equipment and not the value of personal information contained therein, there is no information privacy threat. The Australian federal privacy

commissioner took a soft approach (Privacy Commissioner Ninth Annual Report-Australia, 1997) in a case where there was no evidence of accessing sensitive personal information contained in stolen hard drives. On the other hand, if there was a possibility of leaking personal information, the commissioners would have taken it seriously. A Hong Kong bank collected applications for credit cards together with copies of national identity cards on a public holiday. The officer responsible for handing over the collected documents to the bank accidentally left the collected forms and copies of identity cards behind on the bus while taking them home. The applicants complained to the Hong Kong privacy commissioner (HG-ar0304-7, 2004). The commissioner insisted that the bank take proper security safeguards in handling personal information. One of the proposed measures is handing over collected applications to the nearest, safest place.

Media often reports stolen laptops that contain personal information. In a Canadian case, the commissioner suggested some precautionary measures that include implementing proper access control mechanisms and encrypting data (PIPEDA 289,2005). Some other organizational measures are to limit taking laptops containing personal information out of office premises, requiring a prior approval before taking the laptops out of the office, verifying the appropriateness of measures taken before granting approvals, and preventing employees from leaving laptops unattended, specially in vehicles where they can be seen.

2.2 Accidental Disclosure

It can be seen in the following section that causes for accidental disclosure are a lack of knowledge and awareness, human errors, carelessness, and negligence. The Canadian privacy commissioner has stated that it is a duty of data controllers to taka appropriate measures to prevent unauthorized disclosure resulting from employees' mistakes. In taking preventive measures, the controllers have to take into account the sensitivity of personal information and the possibility of disclosure (PIPEDA 180, 2003). In addition to the above-mentioned points, all unexpected situations that lead to the disclosure of personal information are discussed in this sub section.

In many cases, accidental disclosures take place when there is a transmission of personal information. Several cases have been reported on revealing personal information in the conventional postal mail system. This

is due to sending sensitive information in unsealed envelopes (PIPEDA-154, 2001), sending mail to wrong recipients (PIPEDA-28, 2002), printing sensitive information on envelopes, and placing sensitive information in a visible manner through envelope windows (Settled case 9, 2003). The Lithuanian data protection commissioner has insisted on sending public utility service bills in sealed envelopes (WP-29, 2006). To guarantee all mail is properly sealed, the Canadian privacy commissioner suggested checking seals on outgoing message at an outside facility (PIPEDA 197, 2003). Sending an email message is risky since an ordinary email message goes in a clear text format. The Dutch data protection commissioner has advised Dutch libraries to send encrypted email messages to their library members because this communication carries personal information, particularly preferences on library books (WP-29, 2006).

Sending messages to unintended recipient is a serious issue. This is common in the case of facsimile communication. This is largely due to many people sharing a fax machine and not having a cover to conceal the content. In reported case in Hong Kong, a fax copy containing sensitive personal information was sent to a wrong fax number. The sender's sensitive personal information was leaked since the message was collected by an unintended recipient (HG-ar0102-5, 2001). This case highlights the importance of dialling the correct number of the receiving fax machine. A Canadian employee alleged that his employer had intercepted and read a fax receipt. After the inquiry, the privacy commissioner appreciated the guidelines given for fax users on the company's internal web site. It advised fax users to make sure not to leave the document in the sending fax machine and not to send fax messages to unattended fax machines (PIPEDA 251, 2003).

Today, email is the most common means of communication. It is empowered with a number of new features that are not available through conventional communication means. Some features are mail forwarding, replying, and forwarding to multiple recipients. A company sent an email to 618 recipients about a photography contest. Since all mail recipients' addresses were placed in the "to" field, everyone got to know other members in the programme. The company was instructed to create a group email address for all recipients and send mails to that group email address instead of putting all email address in the "to" field (PIPEDA 277,2003). In this case, only the group email address appears. Other possible

vulnerabilities are forwarding to the wrong email addresses, forwarding without deleting sensitive personal information, and placing sensitive personal information in the "subject" field. The latter is especially concerning because the content of the subject field never gets encrypted even in encrypted email messages. Care must be taken in sending electronic documents since it is possible to include personal information in a hidden manner.

There is a possibility of sending an email message to a wrong recipient. It was reported that an email address was assigned to two persons at different times. In this case the parties, who knew the previous email holder, were not informed about the subsequent change. Without knowing the change of holders, an email message containing personal information was sent. This message went to the second owner who was not the intended receiver. Consequently, the sender's personal information was revealed (Computable, 2007). The Italian data protection authority has insisted that police authorities use digital identities of recipients (WP-29, 2006). In addition to that, the double verification system suggested by the Canadian privacy commissioner for the postal mail system (PIPEDA 28, 2002) provides protection from sending email messages to unintended recipients.

Revealing previous users' information is another threat to information privacy. This could happen due to the improper design of data collecting and recording procedures and technological vulnerabilities. One means of collecting users/visitors information is asking them to fill out a row in a registry. In such a data collecting system, there is a possibility of revealing previous users' information. This issue is highlighted in a Canadian case (PIPEDA 304, 2005). In this case, visitors were asked to write their names at the entrance to a movie theatre. The Canadian privacy commissioner ruled out this procedure since it led visitors to notice the previous users' information and asked the movie theatre to give each visitor a form to write down particulars. It seems this problem was solved in the electronic data collecting system. However, there are some reported cases where this problem occurred in a different manner due to the lack of awareness, poor designing of information systems, and negligence. For example, it can be seen that many users leave their computers, web browsers, and sensitive accounts such as email accounts, bank accounts open without properly logging off. Some users are not aware of threats and others simply ignore this for convenience. Leaving without proper logging off is a problem in

publicly accessible computers, particularly machines in cyber cafes. Possible means of overcoming this problem is proper awareness campaigns and trainings on possible threats and protection measures. Proper design of technological solutions could solve these kinds of vulnerabilities to a great extent.

Even though there are decisions that state it is not necessary to send registered mail (PIPEDA 43, 2002), the Australian privacy commissioner insisted on getting signatures on a delivery receipt (J v Superannuation Provider, 2006). Therefore, it can be expected that getting an email delivery report would give a kind of guarantee that the message was been delivered to the intended recipient. However, there is a possibility of sending an acknowledgement by a third party.

Another identified information privacy threat is using collected personal information for training, educational, and promotional campaigns. Unless the collected personal information is carefully scrutinized and de-identified, it poses threats to data subjects. A company developed a case study based on the facts collected from a couple for marketing purposes. In the process of building the marking plan, some identifiable information was not taken off. Subsequently, the couple realized their personal information was contained in the marketing plan. They then complained to the privacy commissioner. In the inquiry, the company admitted its mistakes and promised to take precautionary measures. One of those measures is not to build further marketing plans based on particulars of customers (NZPrivCmr2-26280, 2002). A possible threat in the digital world is using databases containing personal information for educational, training, and testing purposes.

Another important area is providing access to data repositories that contain personal information to IT service agencies. It is the responsibility of the principle, not the agency, to protect personal information (PrivCmrNZ6-2663, 1994). Outsourcing business processes that contain personal information is a serious threat. It is suggested to have strict contractual terms with data processors. According to Article 16 and 25 of EU Directive 95/46/EC, sending personal information without having a proper contractual term with data processors is prohibited.

Some reported cases highlight the limitation of technological systems. An erroneous match took place since a mother and her son have the same initials in addition to surname and address. This erroneous matching

disclosed the son's personal information. Thereafter, the company decided to put the son's name to a manual monitor list where an employee checked the entries manually each month (PIPEDA 150, 2003).

Taking proper administrative measures is essential for protecting personal information from accidental disclosures. Some recommended measures are having a close place to deal with customers (PIPEDA 237,2003), having working desks with raised barriers at chest level to prevents seeing personal information and instructing employees not to speak loudly (PIPEDA 245, 2003). It is also essential to follow proper security procedures when discarding personal information. In one case, personal information of a worker was revealed because the organization had failed to follow proper procedures in discarding binders. The investigation showed the limitations of security procedures (PIPEDA 228, 2003).

Not only organizations but also data subjects have a duty to help keep data subjects' personal information secure. Some times, data subjects put their personal information at risk by failing to follow given security procedures. In a Canadian case, a worker accompanied her co-worker to a clinic's reception area and later alleged that the receptionist disclosed her personal information to the accompanied co-worker. The commissioner turned down the complaint since the clinic had followed proper security procedures (PIPEDA 237, 2003). Another case was turned down since a user had chosen her mother's maiden name for her password despite the instruction given to her not to choose an easy-to-guess password (PIPEDA 315, 2003).

3 DISCUSSION

This study presented an analysis of decisions given on the principle of security safeguards, particularly accidental incidents by data protection and privacy commissioners. It covered the Ninth Annual Report of the Article 29 Working Party on Data Protection and decisions published online by the Canadian, Hong Kong, New Zealand, and Australian privacy commissioners. This study shed light on how to understand the legal privacy obligations of data controllers in case of accidental incidents. Furthermore, it presented some implementation details of technological measures and appropriate organizational measures for managing technological measures.

Unlike existing privacy frameworks, best practices, and guidelines, the recommendations presented in this paper are meant to have legally binding

effects. This is because these guidelines were derived from verdicts given by data protection and privacy commissioners along with other legally competent tribunals. Therefore, it can reasonably be said these guidelines are mandatory legal privacy requirements. However it should not be expected that these guidelines have the same level of legally binding effects in all jurisdictions as there is no universal harmonized data protection regime. Because all of the studied verdicts were geared toward protecting personal information, it can reasonably be claimed that adhering to measures presented in this paper will contribute to the enhancement of information privacy.

4 REFERENCES

1 [AICPA/CICA] Assurance Services Executive Committee of the AICPA and the Assurance Services Development Board of the CICA. (2004). AICPA/CICA Privacy Framework. New York: Author.

2 APEC. (2005). APEC Privacy Framework -APEC Secretariat-. Retrieved November 30, 2007, from www.apec.org.

3 Computable. (2007). Online identiteit gestolen- 19th of October- Computable. Retrieved November 30, 2007, from www.computable.nl/nieuws.

4 Dempsey, J. X., & Rubinstein, I. (2006). Guest Editors' Introduction: Lawyers and Technologists-Joined at the Hip? IEEE Security and Privacy, 4(3), 15-19.

5 GH-2004005. (2004). Senders of information through fax: must ensure no unauthorized or accidental access to the information by unrelated parties. Retrieved November 30, 2007, from http://www.pcpd.org.hk/english/casenotes/

6 Greenleaf, G., & Waters, N. (2003). The Asia-Pacific Privacy Charter Retrieved November 30, 2007, from http://www.worldlii.org/int/other/PrivLRes/2003/1.html#Heading1

7 HG-Annual Report. (2006). Personal DataAnnual Report 2004-05 Retrieved November 30, 2007, from http://www.pcpd.org.hk/english/publications/annualreport2005_4.html

8 HG-ar0102-5. (2001). Wrongful transmission of subscribers' personal data by fax. Retrieved November 30, 2007, from www1.pco.org.hk/textonly/english/casenotes/case_complaint2.php?id=156

9 HG-ar0304-7. (2004). Retrieved November 30, 2007, from

www1.pco.org.hk/textonly/english/casenotes/case_complaint2.php?id=202
10 Iachello, G. (2003). Protecting Personal Data: Can IT Security Management Standards Help? In Proceedings of the 19th Annual Computer Security Applications Conference (Vol. 8, pp. 266 - 275). Washington, DC: IEEE Computer Society.
11 J v Superannuation Provider. (2006). Improper disclosure of personal information and failure to take reasonable steps to protect, and correct personal information. Retrieved November 17, 2007, from www.privacy.gov.au/act/casenotes/2005.html
12 Muelle, G., & Rannenberg, K. (1999). IT Security and Multilateral Security In K. Rannenberg,
A. Pfitzmann & G. Müller(Eds.) Multilateral Security in Communications. Boston, MA: Addison¬Wesley-Longman.
13 NZPrivCmr2-26280. (2002). Clients object to their financial report being used for marketing purposes by life assurance company. Retrieved November 30, 2007, from
http://www.privacy.org.nz/library/
14 OECD. (1980). OECD Guidlines on the protection of Privacy and Transborder Flows of Personal Data. from http://www.oecd.org/document /20/0,2340,en_2649_201185_15589524_1_1_1_1,00.html
15 Opinion 1/98. (1998). Platform for Privacy Preferences (P3P) and the Open Profiling Standard (OPS). Retrieved September 04, 2007, from
http://ec.europa.eu/justice_home/fsj/privacy/workinggroup/wpdocs/2007_en .htm
16 PIPEDA-154 (2001). Couple dismayed at receiving unsealed envelope from bank. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/cf-dc_030415_1_e.asp
17 PIPEDA 43 (2002). Credit card fraud victim questions bank's use of first-class mail as privacy safeguard. Retrieved November 30, 2007, from http://www.privcom.gc.ca/
18 PIPEDA 150 (2003). Credit agency accused of improper disclosure of personal information. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/ cf-dc_030411_2_e.asp
19 PIPEDA 180 (2003). Bank uses tape-recording of customer's call for unidentified training purpose; connects another customer to the recording. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/ cf-dc_030710_02_e.asp

20 PIPEDA 197 (2003). Individual alleged bank sent personal information in unsealed envelopes. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/ cf-dc_030801_02_e.asp
21 PIPEDA 227 (2003). Mass mailout results in disclosure of contest entrants e-mail addresses. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/ cf-dc_040902_02_e.asp
22 PIPEDA 228 (2003). A transportation company disclosed an employee's personal information without consent. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/ cf¬dc_031104_03_e.asp
23 PIPEDA 237 (2003). Individual accuses employer of disclosing personal information to co¬workers. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/cf¬dc_031120_02_e.asp
24 PIPEDA 245 (2003). Bank alleged to have unnecessarily collected and improperly disclosed personal information. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/ cf-dc_031203_e.asp
25 PIPEDA 251 (2003). A question of responsibility. Retrieved November 30,2007,from http://www.privcom.gc.ca/cf-dc/2003/cf-dc_031212_04_e.asp
26 PIPEDA 254 (2003). Daughter racks up long-distance charges; mom blames phone company. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2003/cf-dc_031223_e.asp
26 PIPEDA 289 (2005). Stolen laptop engages bank's responsibility. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2005/289_050203_e.asp
27 PIPEDA 304 (2005). Movie theatre chain strengthens personal information handling practices. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2005/304_20050607_e.asp
28 PIPEDA 315 (2005). Web-centred company's safeguards and handling of access request and privacy complaint questioned. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf¬dc/2005/index2-5_e.asp
29 PIPEDA_28 (2002). Bank sends customers' pay stubs to wrong party. Retrieved November 30, 2007, from http://www.privcom.gc.ca/cf-dc/2002/cf-dc_020104_e.asp
30 Privacy Commissioner Ninth Annual Report- Australia. (1997). Retrieved November 30, 2007, from http://www.privacy.gov.au/publications/97annrep.pdf

31 Privacy Commissioner Tenth Annual Report- Australia. (1998). Retrieved November 30, 2007, from http://www.privacy.gov.au/publications/98annrep.pdf
32 PrivCmrNZ6-2663. (1994). Woman complains process server revealed debt details at old address Retrieved November 30, 2007, from http://www.privacy.org.nz/library/
33 Settled case 9. (2003). Windows reveal too much information. Retrieved November 30, 2007, from http://www.privcom.gc.ca/ser/2004/s_040706_e.asp
34 Waters, N., & Greenleaf, G. (2001). Privacy Law and Policy Reporter. Retrieved November 30, 2007, from http://www.austlii.org/au/journals/PLPR/2004/36.html
35 WP-29. (2006). Ninth Annual Report of the Article 29 Working Party on Data Protection. Retrieved November 4, 2007, from http://ec.europa.eu/justice_home/fsj/privacy/workinggroup/wpdocs/2007_en.htm

COMPUTER MONITORING IN THE 21ST CENTURY WORKPLACE

Prathiba Mahanamahewa

Course Director: Faculty of Law, University of Colombo

mahanamahewa@yahoo.com

94 1 2716210

Munidasa Kumaratunga Mawatha
Colombo 03
Sri Lanka

ABSTRACT

This paper attempts to lay the foundation for future research into an area that has been called the "hottest workplace privacy topic of the next decade." The existing empirical studies and the literature reviewed of this area suggest that the latest intrusive monitoring technologies which have been introduced to the current workplace has undoubtedly created an unwanted and unexpected imbalance and developed a wide gap in the 21st century employer/employee relationship. The paper argues for the introduction of Privacy enhancing technologies empowered with legal instruments in protection of workplace privacy. In addition, the paper is of the view that employees' awareness and training on workplace privacy policy developments are decisive factors to achieve this objective and this in turn creates trust and confidence and beneficial to both employees and employers in the current workplace. The paper proposes a contractarian framework to protect employers' interests and employees' on-line rights.

This paper suggests that employees' views and opinions are more important in computer monitoring to develop a privacy policy in the workplace. To attain these objectives an empirical survey was conducted in

five government sector organizations in Sri Lanka to gather factual information and to examine attitudes, beliefs and opinions on computer monitoring. The results of the study could be used as guide for policy-makers and for legislatures involved in drafting privacy legislation, and associated policies relevant to the workplace.

KEY WORDS

Electronic Privacy; Information Privacy; Data Protection; E-Policies; Workplace

COMPUTER MONITORING IN THE 21ST CENTURY WORKPLACE

1 INTRODUCTION

It is no secret that governments worldwide are going "online" (i.e., accessing the Internet and establishing Web sites) at a very rapid rate. The United States leads all countries with most of their agencies online. Canada and Australia online agencies follow the United States. In Sri Lanka, the government embarked on an ambitious program that established e-mail and Internet in all government sector organisations under the guise of World Bank in 2003. Under this project public sector employees are equipped with a computer and wide access to e-mail and Internet in the workplace. This transformation of workplace to on-line environment has raised numerous privacy related questions for both employers and the employees. In particular, the issue of e-mail and Internet usage and employee monitoring in the workplace is a significant matter. It was demonstrated here that many countries around the world are competing rapidly to maintain a policy to govern this issue in respect of employee privacy in the workplace. Finding the balance point for many public managers means developing, implementing, and enforcing an acceptable use policy for the e-mail and Internet. But what are the key components of such a policy? How and why do the components vary from organisation to organisation? This paper analyses the issue of electronic surveillance in the workplace and its impact on employee privacy rights. This paper will commence with a discussion of E-government and its various stations and in particular its implementation in Sri Lanka. It will then consider the information privacy as a human right in international and regional instruments. This will be followed by an evaluation of the regulatory framework for protection of privacy in United Kingdom and Sri Lanka. Finally it highlights the importance of an e-policy in an organisation to balance the competing interests of employer and employee.

2 E-GOVERNMENT

E-Government initiatives can be seen to operate at various levels (O'Flaherty, 2000). The first level comprises simple government to citizen communication through which government information such as reports, policy documents, legislation and case law is made available direct to the public through electronic means. In the second stage, citizen to government communication becomes possible allowing citizens to make electronic submissions concerning government proposals for example or to provide government agencies with new information about themselves, such as change of address, by electronic means. Third-level services facilitate more complex interactive transactions. These often involve legally binding procedures and/or online payments. Examples of such transactions include voter and motor vehicle registration or the submission of formal objections to applications for building permits. Fourth level services focus on the delivery of access to a wide range of government services across a whole government administration through a single contact point. At the fifth stage, yet to be fully realized in practice, government applications become intertwined with commercial applications and users are facilitated in building their own interfaces designed around their personal interactions with both government services and commercial entities.

3 INTERNET USAGE AND ELECTRONIC PRIVACY

There are four primary categories of Internet usage: sending and receiving electronic mail (Known as e-mail), accessing and posting documents on the World Wide Web, sending and retrieving computer files (known as file transfer protocol or FTP),and joining electronic discussion groups (such as news groups, listservs, and Internet relay chat groups). E-mail is the most widely used Internet service, although many users are active in all categories. In general the workplace presents a unique arena for privacy analysis. Two competing interests exist in the employment context: the employer's right to conduct business in a self-determined manner is matched by the employee's privacy interests or the right to be let alone.

For managers, monitoring is necessary. It is argued that workplace e-mail and Internet monitoring are the most effective means to ensure a safe and secure working environment and to protect employees. In addition, some contend that monitoring may boost efficiency, productivity and

customer service and allows more accuracy to evaluate performance (DeTienne, 1993; Sipior and Ward, 1995; Orthmann, 1998; Sipior et al., 1998). The impact of monitoring of these workplace relationships is the focus of this thesis. If used reasonably it may enhance efficiency without "trenching on" employees rights.

However, critics of monitoring point to research evidencing a link between monitoring and psychological and physical health problems, increased boredom, high tension, extreme anxiety, depression, anger, serve fatigue and musculoskeletal problems (Amick and Smith 1992; Kidwell and Bennett, 1994; Chalykoff and Kochan, 1989; OTA, 1987; Working Women Education Fund 9to5, 1990; Stanton, 2000). More seriously, critics point to violations of their fundamental right to privacy (Stone et al., 1983; Bylinsky, 1991; Culnan, 1993; Smith, 1993; Vest et al., 1995; Alge, 2001). Unless an acceptable remedy is soon found, workplace productivity may rapidly deteriorate and employee morale may disintegrate.

4 INFORMATION PRIVACY AS A HUMAN RIGHT

4.1 International Instruments

The most significant international human rights instrument is that of the Universal Declaration of Human Rights of 1948 (UDHR). Its provisions which deal expressly with privacy are set out in article 12, which states:

> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

In almost identical terms, article 17 of the International Covenant on Civil and Political Rights, 1966 (ICCPR) provides that:

> No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home correspondence, nor to unlawful attacks upon his honour and reputation.

Everyone has the right to the protection of the law against such interference or attacks.

4.2 Regional Instruments

Whereas the above provisions are framed essentially in terms of a prohibition on "interference with privacy", the equivalent provisions of

article 8 of the European Convention on Human Rights, 1950 (ECHR) are phrased in terms of a right to "respect for private life":

> Everyone has the right to respect for his private and family life, his home and correspondence.
>
> There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

The Charter of Fundamental Rights of the European, 2000 Union reaffirms the recognition of fundamental rights in the context of EU. Article 7 of the Charter states that

> Everyone has the right to respect for his or her private and family life, home and communication.

Article 8 of the Charter states that

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis for the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

The intention of the drafters of this article was to follow the traditional wording of article 8 (ECHR) while at the same time adapting the former to modern developments and technological change. This was done by replacing the term 'correspondence' (article 8, ECHR) with 'communications'. Article 7 guarantees protection against the intervention or interference of public authorities in the private sphere.

Article 8 of the Charter recognises Data Protection as an innovative fundamental right. The Draft Treaty establishing a Constitution for Europe ('European Constitution'), as proposed by the European Convention on the

Future of Europe reproduces article 7 of the Charter of Fundamental Rights under article 7 and article 8. Article 50 of the Draft Treaty is intended to establish a single legal basis for the protection of personal data, both for the protection of data which is processed by the European institutions. The protection of privacy may take on new meaning as a consequence of the Charter of Human Rights and the adoption of privacy provisions in a future European Constitution. Other than article 11 of the Inter-American Convention on Human Rights, the major regional human rights catalogue omits express protection for privacy or private life.

These international and regional instruments recognise privacy as a fundamental human and civil right. If article 12 of the Universal Declaration of Human Rights is taken in conjunction with article 8 of the Convention for the protection of Human Rights and Fundamental Freedoms, as well as with the concepts outlined by international organisations and individual countries, a fairly clear and broad definition of privacy can be identified, setting a standard of privacy that clearly protects the individual. That which is private should be respected, only to be breached in the case of very clearly set criteria, a notion reinforced with the European Convention of human rights. It is against these fundamental codes and declarations of human rights that this consideration of e-Work and monitoring is set. All actors considered are clearly covered by the definition and should, therefore, be respectful of and compliance with the protection provided by them. Although electronic surveillance is yet to be considered under the ICCPR, it has been taken up under the equivalent privacy right (article 8) contained in the ECHR, as well as in the draft European Constitution.

4.3 The Council of Europe and the Data Protection Directives

The Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data 1981 seeks to protect individual rights and freedoms. The convention is particularly relevant because it provides for a right to privacy. In 1995, the European Union enacted the Data Protection Directive (95/46/EC) in order to harmonize member states' laws in providing consistent levels. The Directive provides for a basic or fundamental level of privacy that not only reinforces current data protection law, but also establishes a range of new rights. The twin objectives of the Directive expressed in Article 1 were: to protect the rights of individuals with respect to the processing of their personal data; and to

facilitate the free movement of personal data between member states. The first objective received much attention and it was the second that hold out the prospect of major economic benefit. The EU Directive (1995) is motivated by economic considerations, particularly the need to harmonise data privacy laws in the Union. However, the Directive also stresses the importance of fundamental human rights. The economic impact of the EU Directives has been far greater than any other instrument given its legal effect within the EU and its approach towards third countries. One of the fundamental economic objectives of the Directive was to enhance the free flow of data within the EU by removing barriers caused by internal borders.

In 1997, the European Union supplemented the 1995 directive by introducing the Telecommunications Privacy Directive (97/66/EC). On June 25, 2002 the European Union Council adopted the new privacy and Electronic Communication Directives (2002/58/EC). In the context of the spread of ICT at work and its associated risks, new concerns are arising in respect of the relationships between employers and employees to address these special issues related to workplace privacy. The European Commission is due to enact a Directive on workplace data protection in 2004 or 2005. The next section analyses information privacy defined in agreements of International organisation.

5 INTERNATIONAL ORGANISATIONS

5.1 Organisation for Economic Cooperation and Development (OECD)

In the late 1980s, the OECD issued a set of guidelines concerning the privacy of personal records. Although broad, the OECD guidelines set up important standards for future governmental privacy rules. Unsurprisingly, the organization had economic considerations in mind when it issued its guidelines. These guidelines underpin most current international agreements, national laws, and self-regulatory policies, and are voluntary and address the collection limitation principle, purpose specification principle, use limitation principle, openness principle and accountability principle. Clarke (1988) argues that the expression of privacy guidelines was shown to have been motivated by the protection of business activities, rather than of peoples privacy.

The primary reference for privacy protection is located in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data in 1980 and Implementing the OECD 'Privacy Guidelines' in the Electronic Environment: Focus on the Internet', Committee for Information, Computer and Communications Policy. This instrument is a set of guidelines. It is not a convention; and it merely "recommends" that member countries consider the principles into their domestic legislation. Greenleaf (1996) contends that existing legislations having incorporated privacy guidelines do not provide sufficient protection against new monitoring technologies coupled with highly bureaucratic administrative practices.

5.2 International Labour Organisation (ILO)

The ILO has no convention to protect privacy but has adopted a Code of Practice on Protection of Workers Personal Data 1997, which covers general principles about protection of personal data and specific provisions regarding the collection, security, storage, use and communication of such data. Unlike other ILO instruments, the code is not legally binding like other international treaties. It provides employers and workers with the basis for rules to be designed by them. The code was intended to provide guidance in the development of legislation, regulations, collective agreements, work rules, policies and practical measures in the workplace.

According to an ILO survey (ILO, 1993), workers in industrialized countries are gradually losing privacy in the workplace as technological advances allow employers to monitor nearly every facet of time on the job as a remedial measure and to protect employee privacy. The ILO introduced, therefore, guidelines on employee monitoring at the workplace. To further protect workplace privacy the ILO introduced a code of practice called the Protection of Worker's Personal Data (1997). Its purpose is to provide guidance on the protection of workers personal data and is not as a binding force. The code does not replace national laws, regulations or other accepted standards. It can be used in the developments of legislation, regulations, collective agreements, work rules, policies and practical measures at enterprise level. According to the ILO Code, secret monitoring is permitted only if there is suspicion on reasonable grounds of criminal activity or other serious wrongdoing. The ILO recognises that workers rights to privacy should be treated as a fundamental human rights issue, but the new technology can pose dangers to privacy, even as it is improving all

of our lives. The ILO calls the problem the "chemistry of intrusion", a combination of threats to informational privacy, increasing encroachments on physical privacy and increased physical surveillance.

There is a certain consistency among these principal instruments. Each seeks to establish consistent rules to protect the recognized right to privacy in order to pre-empt incompatible national rules that would damage the economic benefits of free flow of information.

It is now a quarter of a century since key data privacy instruments were adopted by the OECD (1980) and Council of Europe (1981). These were followed by the United Nations Guidelines (1990) and EU Directives (1995). Most of these instruments have had reviews of one or another sort. Nonetheless, there are people who wonder whether the various national laws, and these instruments, really achieve their objectives of protecting privacy and whether achieving the supposed benefits is worth the cost. The time to define the exact nature and extent of privacy protections is long overdue. Unless privacy is asserted as a human right, fundamental protections for individuals and institutions will decline leading to a breakdown of social and economic processes.

The EU Directives requires that "Each member State shall provide that one or more public authorities are responsible for monitoring the application within its territory of provisions adopted by the Member states pursuant to the Directive" (EU 1995, Article 28.1). The OECD Guidelines fail to require the creation of privacy protection agency. However, the public expectation is that a specialist body will exist to supervise government agencies and corporations, and the OECD Guidelines fail to fulfil that expectation. The OECD Guidelines fail to specify the measures needed to ensure that privacy protection regime is achieved. The OECD Guidelines appear to be silent on this matter. They clearly need to be enhanced to require the privacy protection agency to make the maximum information available to the public, and to establish working relationships with privacy advocates and representatives of the public. Therefore, need to define a new the exact nature and extant of privacy protections.

6 LEGISLATION TO PROTECT PRIVACY IN UNITED KINGDOM

The UK does not have a written constitution. Nor, until recently, could it be said to recognize a generic concept of "constitutional rights". The Human

Rights Act 1998 (UK), which came into force on 2 October 2000, recognizes a right to privacy. Article 8 has broad application, and provides a concrete right for individual at work. In the absence of any widespread recognition of a common law tort of invasion of privacy, several British legislatures have attempted to create a statutory protection of privacy. Workplace privacy is regulated by two sets of legislations. Namely, the Regulation of Investigatory Powers Act 2000 (RIPA) and the Data Protection Act 1998 (DPA). The main purpose of the Regulation of Investigatory Powers Act (RIPA) 2000 is to ensure that the relevant investigatory powers are used in accordance with human rights. The data Protection Act 1998, designed to implement EC Directive 95/46 on personal data, came into effect in March 2000, and establishes a separate statutory regime which governs the "processing" of "personal data". Therefore, even if employers obtain personal data as a result of an interception authorised under the Regulation of Investigatory Powers Act (or outside the scope of that Act), they must also ensure that they comply with requirements of the data Protection Act.

7 IS THERE ANY LEGAL PROTECTION FOR DATA PRIVACY IN SRI LANKA?

The concept of privacy is not clearly defined unlike the European Union, where most people seem to know what to expect, which makes the work of the judicial bodies easier as issues of interpretation are quickly settled. Conventionally, general privacy concerns have been addressed through the law of torts (breach of confidentiality, trespass and nuisance) and criminal law. Like many other Commonwealth countries the common law of Sri Lanka is based on principles of English law. In addition, some of the principles of law of contract are governed by Roman-Dutch law like in South Africa. The lack of a framework on data protection prevents e-business from the European Union and affects Sri Lanka's economy. Therefore, privacy norms and procedures are expected to arrive from the United Kingdom.

Therefore, legislative measures or other measures, such as the adoption of "Codes of Practice", embodying privacy principles would ensure workplace privacy protection on employees' personal information. This would mean that Sri Lanka is in a similar position with the West. The

question remains whether these arrangements can meet the extra demands brought about by electronic communications.

Information about an individual's tastes and leisure activity has economic value, and the exchange of such information helps to grease the economy. Sri Lanka has never banned the sale of such data, despite the potential impact on privacy. There are, however, many different levels of legal protection for privacy when websites and e-commerce firms, without consent, use private information for commercial purposes. No comprehensive protection exists. In many countries there is a general law that governs the collection, use and dissemination of personal information by the public and private sectors. An oversight body then ensures compliance. This is the preferred model for most countries adopting data protection laws and was adopted by the EU to ensure compliance with its data protection regime.

8 E-POLICIES

Organisations without such policies run the risk of being sued for actions of an employee. Policies "create clear standards to prevent employment disputes and ensure consistent supervisory administration of employment relations". The specific policy selected depends on the culture of the workplace, but most policies have common elements.

The common policy components are:

- Cautioning employee about the risks of using e-mail and the Internet.
- Informing employees:
  - that e-mail is irretrievable
  - that Internet activities can be traced by third parties
  - of downloading procedures and the risk of viruses
  - of all prohibitions of inappropriate and illegal uses
  - that the employer can be held liable for activities of the employee, and
  - that their electronic actions can be so identified with the employer.
- Include information designed to curtail employee conduct for which the company may be liable, namely: defamation, harassment, and discrimination; copyright and patent infringement.
- Establish limits on what may be downloaded from the Internet or exchanged via e-mail.

- Remind users that text, graphics, and software that appear to be freely available on the Internet are often subject to intellectual property laws that limit copying, distribution, and use.
- Confidential Information
- Use technological means to prevent trade secret and confidential files from being transmitted.
- Mandate the use of encryption software or ban the transmission of sensitive information altogether.
- Create an approval and clearing policy for information to be published on the web
- Establish monitoring procedures and inform employees about the details of such monitoring
- If applicable, clarify that incidental personal use is a privilege, which can be revoked for abuse or excessive use.

9 CONCLUSION

It is well established that neither constitution nor statutory law addresses the new privacy issues associated with technology and the old common law does not clearly cover the area of privacy in question in Sri Lanka. Therefore a gap exists between the time when a new communication technology is created and the time when a statute is designed by state legislature to cover the new technology. This paper contends that a modern computerised workplace reduces the arbitrary powers enjoyed by the employers and reduces their ability to act against the employee unilaterally and effectively. Hence, we can design a incentive-compatible, benefit maximising contact between managers and employees based on the following principles: employee participation in defining privacy policies; full disclosure of all implementation schemes pursuant to these policies; and employer monitoring to ensure compliance with such policies. Finally, this has been endorsed by the Article 29 – European Union Data protection Working Party (Article – 29 EU Data Protection Working Party, 2002, p.24) in their statement specifically: “A blanket ban on personal use of the Internet by employees may be considered to be impractical and slightly unrealistic as it fails to reflect the degree to which the Internet can assist employees in their daily life”.

10 REFERENCES

O'Flaherty (2000) 'Privacy Impact Assessments: an essential tool for data protection' A presentation to a plenary session on New Technologies, Security and Freedom' at the 22nd Annual Meeting of Privacy and Data Protection Officials held in Venice, September 27-30 2000.

DeTienne, (1993) "Big brother or friendly coach? Computer monitoring in the 21st century" The Futurist, Vol. 27, p. 33.

Sipior and Ward, (1995) "The ethical and legal quandary of email privacy" Communications of the Association for Computing Machinery, Vol. 38, p. 8.

Orthmann, (1998) "Workplace computer monitoring" Employment Testing - Law and Policy Reporter, Vol. 12, p. 182.

Amick and Smith (1992) "Stress, computer-based work monitoring and measurement systems" Applied Ergonomics, Vol. 23, p. 6.

Kidwell and Bennett, (1994) "Employee Reactions to Electronic Control Systems" Group and Organization Management, Vol. 19, p. 203.

Chalykoff and Kochan, (1989) "Computer-aided monitoring: Its influence on employee job satisfaction and turnover" Personnel Psychology, Vol. 42, p. 807.

OTA, (1987) Office of Technology Assessment. (1987) The electronic supervisor: New technology, new tensions (U.S Government Printing Office: Washington DC).

Working Women Education Fund 9to5, (1990) "Stories of mistrust and manipulation: The electronic monitoring of the American workforce" (Author: Cleveland, OH).

Stanton, (2000) "Reaction to Employee Performance Monitoring: Framework, Review, and Research Directions" Human Performance, Vol. 3, p. 85.

Stone et al., (1983) "A Field Experiment Comparing Information-Privacy Values, Beliefs, and Attitudes Across Several Types of Organizations" Journal of Applied Psychology, Vol. 68, p. 459.

Bylinsky, (1991) "How companies spy on employees" Fortune, Vol. 124, p. 131.

Culnan, (1993) "How did they get my name? An exploratory investigation of consumer attitudes toward secondary information use" MIS Quarterly, Vol. 17, p. 341.

Smith, (1993) "Privacy policies and practices: Inside the organizational maze" Communications of the ACM, Vol. 36, p. 105.

Vest et al., (1995) "Factors influencing managerial disclosure of AIDS health information to co-workers" Journal of Applied Social Psychology, Vol. 25, p. 1043.

Alge, (2001) "Effects of Computer Surveillance on Perceptions of Privacy and Procedural Justice" Journal of Applied Psychology, Vol. 86, p. 797.

A COLLABORATIVE DISTRIBUTED VIRTUAL PLATFORM FOR FORENSIC ANALYSIS OF MALICIOUS CODE

Leonard Shand[1] and Theodore Tryfonas[2]

[1]Independent e-commerce and e-forensics specialist
Monmouthshire
United Kingdom
leonard@lenshand.net

[2]Lecturer, Faculty of Engineering
University of Bristol
United Kingdom

theodore.tryfonas@gmail.com

ABSTRACT

Malicious software is prevalent in many forms with the potential for many types of malware to be downloaded while browsing the Internet using an unprotected system. The potential impact can be irreparable harm to a computer file system or even place a person in a situation where they could be charged for a criminal act, if the perpetrator assumes control of their system. Understanding contemporary forms of malware is crucial in order to prepare better defences against it as well as investigate related incidents and claims. Therefore forensic analysis of specific malware, requires specialised tools and techniques and is of significant importance for information security professionals.

In an effort to facilitate the process of forensic analysis of malicious and hostile code we intend to develop a system whereby specific malware can be identified, classified and the malware and detailed forensic analysis stored in a searchable database. The research results would assist computer forensics expert witnesses and infosecurity specialists, to determine the potential role, and impact on a case of certain malware types found to be present on a computer under examination.

To this end, we first research on different types of malware and obtain a selection of malware samples as a specimen to investigate. We create an environment containing suitable investigative tools with which to analyse malware and devise a virtual testing utility platform (containing networking settings, software etc.) to conduct examinations. Experts can use the virtual infrastructure provided to analyse malware and then log their analysis results, notes and experiences in a bespoke on-line collaborative web accessed database. In there experts can log their findings and further produce analytical aids including the behavioural profile of the malware inspected, and potentially be others analysing the same types of malicious code.

KEY WORDS

Malware analysis, computer forensic examination, virtual platform, knowledge management system

A COLLABORATIVE DISTRIBUTED VIRTUAL PLATFORM FOR FORENSIC ANALYSIS OF MALICIOUS CODE

1 INTRODUCTION

Malware has traditionally been associated with viruses and Trojans and is a software program designed to disrupt computer systems. Of late, however, malware has become more insidious and stealth-like and the means of detection have become more and more difficult. Programmers of malware are becoming increasingly adept at modifying malware and the proliferation of source code on the Internet has enabled them to inspect the operation of the malware in much greater detail [1].

When a forensic investigator investigates a computer system that has been used in a suspected criminal activity, he needs to determine, amongst other criteria, whether such activity was the result of the perpetrator's actions or whether the computer system and/or any related software could have been instrumental in causing the offence.

The purpose of the research presented in this paper is to assist the forensic investigator to be able to see at a glance whether there are indications/evidence that malware could be the cause of the suspected crime or whether the perpetrator is responsible. The overall project aims at exploring the following:

- The current status regarding malware and previous research into its analysis.
- The information required for determining whether malware could cause a criminal act to be perpetrated.
- Safe methods and environments for malware analysis.
- Investigating the initial requirements of a prototype tool for collaborative malware analysis and develop a web-based database application for it.

Our system is collaborative in the following sense: we assume the involvement of three distinct roles in malware analysis. That would be the anti-virus researcher/analyst, a person of highly technical orientation and skills, at least in the areas of operating systems and networking. Also, a computer forensics examiner, a person primarily competent in the use of forensic analysis applications such as EnCase and AccessData, as well as aware of current and relevant computer incident related legislation. Finaly, a systems administrator who is responsible for maintaining the web application. Facilitating the colaboration between the three roles and allowing for seamless exchange of information between them, and particularly towards the forensic examiner, is deemed essential. Of course the three roles mentioned here need not necessarily be distinctive individuals, but a malware analyst can also be a system administrator or forensic examiner and vice versa.

The paper then is structured as follows. Section two reviews the types of malware that a forensic examiner may come across during a computer examination as well as related concepts of malware impacts and protection. The third part describes the creation of a safe environment for the collection and analysis of malware. Specialist software was identified which could be used to analyse the malware in the safe environment and choices as to the most appropriate set of tools will be chosen in this report that will best suit the testing and analysis of the malware. Also in part three we discuss the design and implementation of a web-based application, the implementation of the safe environment and the analysis of malware. Part four examines the potential of this system in use and how it could assist a group of forensic investigators in collaborating and sharing knowledge on malware incidents. Finally we conclude the paper identifying areas for further development.

2 MALICIOUS CODE AND PROTECTION

2.1 Types, Impact and Controls

Under UK law, having malware in your possession is not necessarily an offence, but the dissemination of that material in any form is (the UK Computer Misuse Act, section 3). Using the malware in a concerted effort to do damage is a crime and as such is punishable under one or more of the pieces of legislation mentioned above. Apparently, knowing that your computer could possibly be used, or has been used in a malware threat could make you liable. It appears that by not properly protecting your computer

with all manner of anti-malware software, you could possibly be prosecuted criminally for recklessness and civilly for negligence. While this has never been tested in a court of law, the possibility is there.

The computer forensic investigator needs to determine what, if any, information on the system can be used as evidence for or against a user under investigation. The search for and inspection of malware is fraught with difficulty. There are three main categories of malware that any computer user could encounter in their life. The severity of malware is regarded as one of benign to destructive or dangerous, depending on the role of the malware. Malware can infect a system or the files on that system.

Viruses are the most common form of malware. A virus is a program which infects a computer system by installing themselves on it and then replicating. Most known viruses are caught by up to date antivirus software and are not as much of a threat as they used to be [2]. There are three distinct types of virus:

- File infectors: Two types of file infectors have been identified. The first is a virus that infects and attaches itself to files and then executes each time the file is run or opened. The second type does not change the file in any way, but alters the route in which a file is opened. Performance is variable as some viruses will actually impact on the performance of the entire system quite considerably.
- Boot sector: Boot sector viruses infect the boot sector of discs. They then replicate when booted from that disc. This type of virus may have no noticeable impact on the performance of the system.
- Macro: At present this is the most common form of virus [3]. Macro viruses use the program's own macro programming language (Visual Basic for Applications: VBA) to allow execution. The infection takes place by writing itself into the normal.dot file. Any time any Microsoft Office application is run, the virus is spread by infecting the document that is being created or read. Due to other applications' ability to open most formats of Office files, the virus can be spread to other computer platforms too. This is the case with Macintosh, DEC, Linux and Windows.

Worms are the oldest form of malware, even though they did not start out that way. The first implementation of a worm was by John F Shock and

Jon A Hupp, researchers at Xerox PARC (Palo Alto Research Center) in 1978. Its purpose was to search out other computer hosts, find idle processors on the network and assign them tasks, sharing the processing load, and so improving the 'CPU cycle use efficiency' across an entire network and then copy itself and self destruct after a programmed interval [4].

Worms travel through networks and the Internet by various means. They do not attach themselves to other programs. Worms almost always cause harm, even if it is only consuming bandwidth. While many worms are designed to spread, many have payloads. A payload is code designed to do more than just spread. Payloads can have the following actions: delete files on a host system, encrypt files in a cryptoviral extortion attack, send documents via e-mail, and install a backdoor in the infected computer to allow the creation of a "zombie" machine under control of the worm author.

When these machines are networked they are often referred to as botnets and are very commonly used by spam senders for sending junk email or to cloak their website's address. Backdoors can be exploited by other malware, including worms. While many worms are malevolent, there are a few that are used for good intent. One such example is those used to update or patch systems or applications such as Windows [5]. The biggest issue with worms is that they do their task without the explicit consent of the user and could generate considerable network traffic.

Trojan horses derive their name from Homer's reference to the wooden horse of the Greeks in Troy. Its purpose is to infect a system under the disguise of a useful or required application or file. Most Trojans have a payload and are perpetrated from elsewhere to gain access to a system in order to gain full control of it as well as giving itself access to files and data on it. Trojans usually consist of two parts: a client and a server.

When the server is installed, it allows the remote client software to send commands to the server. This notifies the remote attacker, who can then upload and download files, can delete and create files and folders and can control most of the machine. Most Trojans will notify the remote attacker that the server is running. This action is mostly done via IRC (Internet Relay Chat) [6]. The Trojan infection process is explained in Figure 1.

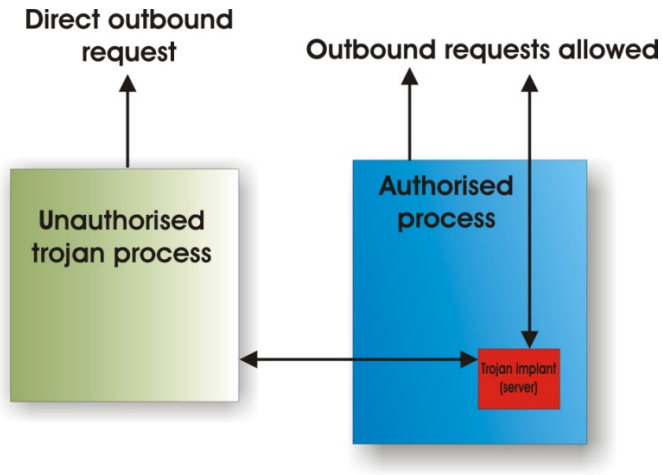


Figure 1 Trojan infection

Malware affects the system in many ways. Most notably are changes in system behaviour which include:

- attempt to connect to websites
- open file shares
- send email
- open other communication channels with remote systems
- launching new services and/or opening listening ports on a system that wait for remote commands
- modifying start-up settings to ensure that it will always run each time the system reboots
- modifying registry setting in Windows

In many cases the user is not even aware that malware exists on the system. It has been noted that even with antivirus software, a firewall and anti-spyware installed, malware can exist on the system without detection [7].

While a lot of users are computer literate enough to know that prevention is better than cure, many users still do not adequately protect their systems. A survey conducted by Schwartz Communications, Inc.

indicates that although many users have antivirus software installed, this software is not updated to an acceptable standard. And while it would be safe to assume that antivirus software will protect a user from most known viruses, for Trojans and spyware, this alone is not sufficient. An adequate firewall and dedicated anti-spyware software is also required. Of course, while this will provide protection for the user, it is not the panacea that most users will perceive as.

According to CERT [8] the following steps should be taken before connecting a new computer to any network:

- Connect the new computer behind a network (hardware-based) firewall or firewall router
- Turn on the software firewall included with the computer, if available
- Disable nonessential services, such as file and print sharing
- Download and install software patches as needed

While the above holds true, the user should attempt to download and install patches or service packs for the relevant operating system, before connecting it to the network by making use of an existing networked computer system. In this way, the user is assured that the system is up to date and that only software patches then need to be updated.

Anti-virus packages make use of various methods to detect malware. According to Aycock [9] there are three main tasks an anti-virus package should perform:

- Detection
- Identification
- Disinfection

Detection of malware is usually by its signature. This signature may be able to be in the form of a combination of bits or it could be a complete cryptographic payload. It is important that the anti-virus package correctly identify malware and to produce as few false positives as possible. Once the malware has been identified, the user should be presented with an option of whether the malware needs inoculation, quarantine or deletion.

2.2 The Need for Collaborative Forensic Analysis of Malware

In the case of Regina v Caffrey, Aaron Caffrey was acquitted only after a lengthy and costly court case. The Register [10] reported: “A forensic

examination of Caffrey's PC found attack tools but no trace of Trojan infection".

From the discussion on malware and of related court cases, it becomes apparent that, a tool that could assist forensic examiners in the task of identifying, analysing and reporting on malware found on a suspect machine, is deemed to be extremely useful. The application proposed here is a collaborative system in the form of a web-based database which will allow the forensic examiner to look up several aspects of malware, including:

- Name
- Aliases
- File names associated with the malware
- Exploit/means of attack
- Action of the malware
- Which parts of specific legislation could be used for an arrest
- Any registry keys potentially affected

With this information in hand, the examiner could be assisted in determining if the user knowingly had malware on the system at the inferred time or whether the malware came to be on the machine by other devious means.

3 SYSTEM DESIGN SPECIFICATIONS AND IMPLEMENTATION DETAILS

3.1 Virtual Analysis Environment

As malware can be detrimental in its operation, it is important that a safe method be devised to capture, analyse and disassemble malware. In this chapter we will consider the design specifications for our proposed tool based on the nature of malware and the requirements of collaboration, as discussed earlier.

While it would be easy to set up a computer system without any connectivity to the Internet or an internal network, many malware types attempt to dial out, scavenge the network or use subversive methods of informing a server or remote attacker. It is therefore necessary to implement a system that would have a form of network connectivity. Virtual environments have become quite widespread in the past decade. A virtual environment is one that runs on a computer system and allows the user to

create virtual machines within this environment. The virtual machine software allows a single computer to allow many more operating systems to be run in the same environment.

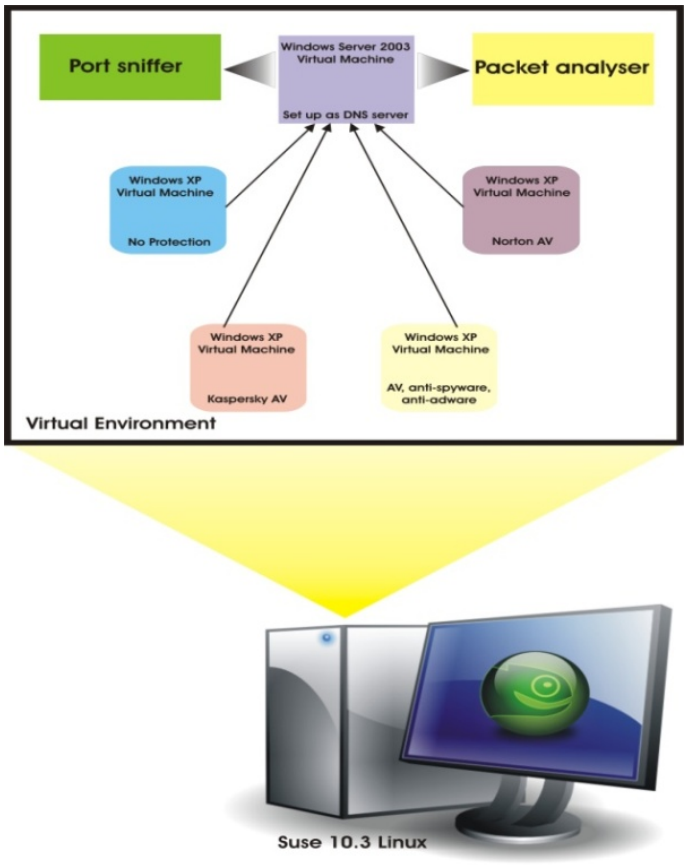


Figure 2 An envisaged virtual system

Using virtualisation, a number of technologies could be implemented within a single machine environment as detailed in Figure 2. By making use of a virtual machine system, licensing issues with proprietary software needs to be taken into account. While the software is running within a virtual environment the software must be appropriately licensed as should the operating system of the base machine. Each installation essentially needs its own license. In the case of Figure 2, The Microsoft Windows 2003 server

would require a license and its minimum 5 user connection license and the four Windows XP workstations would each require their own license.

Various software packages are required to analyse the malware. The software can be categorised as follows:

- System software – the base operating system
- Test bed software – Vmware
- Test bed system software – Windows Server 2003 and Windows XP Professional SP2
- Network analysis tools
- File analysis and disassembly tools
- Registry tools

There is a plethora of analysis tools that a forensic analyst could use for testing for the presence of malware, examining and disassembling malware and analysing malware behaviour. As this project is concerned with determining how the malware could have infested itself on a machine and what actions the malware will perform, only specific tools will be required.

Network analysis tools have been on the market as long as we have had networks. Most network administrators use tools every day to analyse their company's network. These tools would include, but are not limited to:

- Nessus – a vulnerability scanner (www.nessus.org)
- nMap – a utility for network exploration (http://insecure.org/nmap/)
- PRTG traffic grapher – monitors bandwidth usage (http://www.paessler.com/prtg)
- Ethereal – Now Wireshark - a protocol analyser (http://www.ethereal.com/)

As all of these programs are either free or can be downloaded and fully function as a trial, most of them will be used in the analysis to determine if the malware is making use of the network in any way.

File analysis tools are used to determine the code of the file. It is necessary to determine if the source code could be viewed in any form to

investigate how the malware operates. For this purpose OllyDBG can be used (available from http://www.ollydbg.de/ as shareware). Registration is free and once the registration form has been emailed, the program may be used freely. The IDA Pro disassembler and debugger has a graphical based interface and focuses on fundamental analysis of files (http://www.datarescue.com/idabase/idadown.htm).

Also, the Windows registry is a complex directory which stores information regarding the Windows operating system and installed applications. Many malwares update the registry in some fashion and a means of determining what has been changed and by which application is key. While regedit.exe could be used for searching the registry and manually changing keys, it is very difficult to determine changes. For this a snapshot of the registry is needed. A program that claims to be able to take snapshots of the registry is Registry Workshop from www.torchsoft.com.

All the above tools can be used to some extent for each piece of malware under analysis. As an innocent person could be wrongfully convicted by false information provided by the web-based application, it is necessary to determine exactly what the malware is capable of and how it operates. Therefore all means necessary to determine the modus operandi of the malware and possibly its origins are vital.

3.2 Content Management System

The overall system enables *malware analysts* to record their findings and enter those into a database. *Computer examiners* can then search the database for relevant information in a quick and efficient manner. The application allows an analyst to quickly search for an item and then attempt to identify it on the suspect system (within the safe environment), as one of its goals is to record malware, its associated behaviours and actions.

Reviewing the results of analysis of malware is crucial to the investigator in a criminal case. Not only does the information have to be current, it needs to be accurate and must also allow any one with some computer knowledge to be able to find the information – e.g. an investigative officer not necessarily expert in computing. The application allows the investigator to use various methods by which to search for specific items such as:

- Type of malware
- Name of malware
- Possible crime committed
- Registry key
- Infection method

Aside from the functional and non-functional requirements, it is critical to develop an application that is appealing and easy to use by the intended users. In order to achieve this, the system stores and displays all malware data in an appealing and usable manner for the user. This is achieved by attempting to make the system as easy to learn and use as possible, but also by providing the most possible detail in a single screen.

In the application the following tasks are provided for: record malware and its attributes, record malware behaviour and infection methods, record Acts of Parliament and statutes, record possible crimes, amend all the above, view malware and its attributes, view malware behaviour and infection methods, view Acts of parliament and statutes, view possible crimes and query data held by the system. Creating new records and amending existing records is available only to authorised skilled analysis personnel, while viewing and searching of the records is provided to authorised law enforcement agencies and personnel.

Finally, MySQL is used for the database implementation. The ER diagram of Figure 3 represents diagrammatically the initial structure of the database. Although there are two types of users, namely analysts who would record data into the database and users who would only perform searches, a decision was made to remove unnecessary redundancy from the database design. Therefore a single Users table has been created which has a 'level' for the user. This will allow only persons with the requisite authority to access the data input section of the web application. Figure 3 reflects the simplest structure of the database.

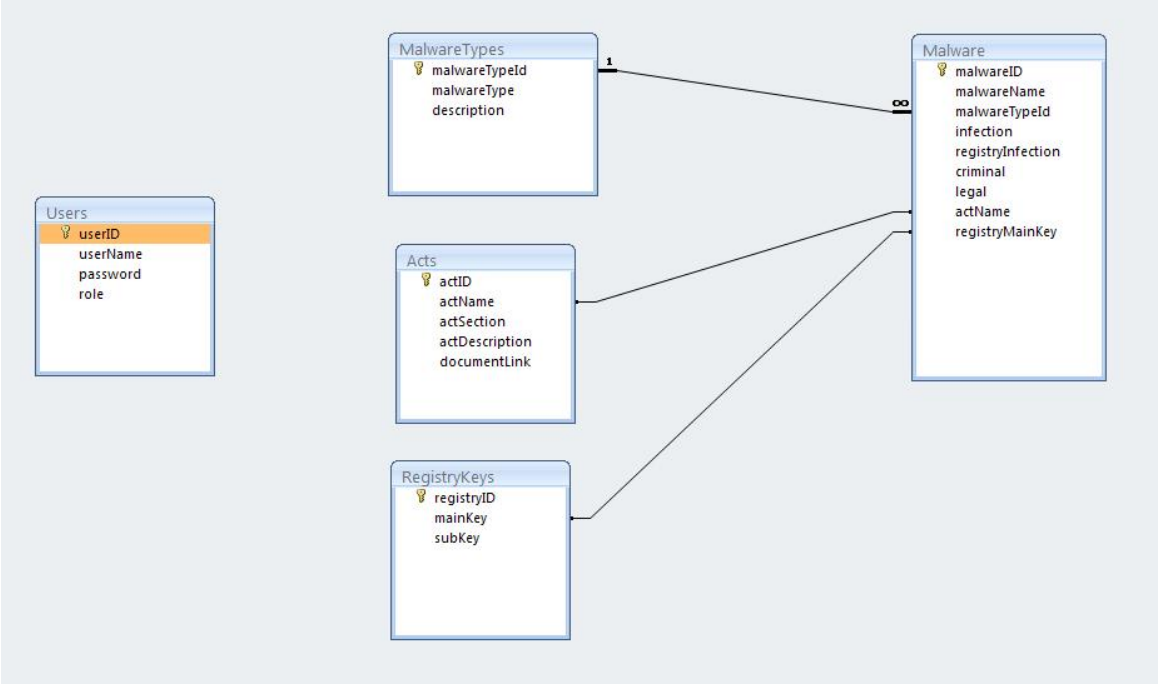


Figure 3 ER diagram for the web application database

4 USING THE SYSTEM PROTOTYPE

Our Malware Analysis Tool (MAT) is a prototype web application built upon a platform that matches the specifications as discussed in the previous part, which allows investigators to search for entries on items of malware (as for example it can be seen in figures 4 and 5) and use this information to assist them in determining whether there is malware present on the suspect machine.

The two key components of this environment are the content management system, enabling the sharing of knowledge, and the secure testing environment. The latter has been tested with various types of malware including Trojans and rootkits and appears to be solid with no leakage out of the environment. In this environment, the malware analyst is at the minute required to analyse the malicious code and subsequently manually input the data into the MAT analysis form which is then posted to the MAT database. Our ultimate goal is to implement a collection of scripts

and macros which automatically record the results as the analysis is taking place in the safe environment. Once the analysis is complete, the form can then be automatically uploaded to the MAT database.

Figure 4 Specific Malware listing

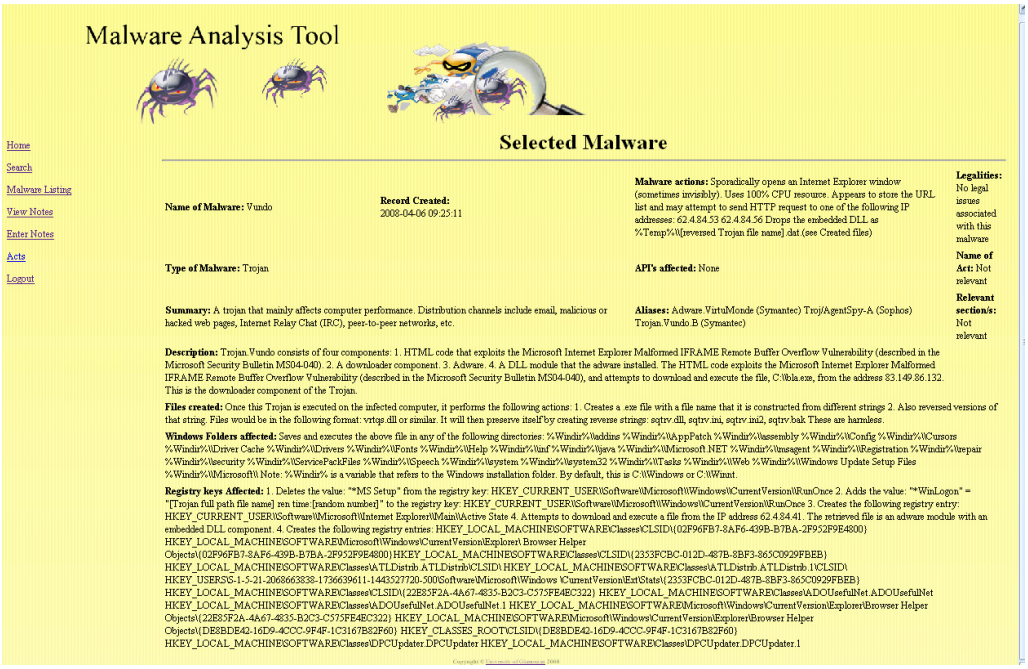


Figure 5 Listing of case under investigation

5 CONCLUSIONS AND FURTHER WORK

Not only is forensic examination of malware a problem for computer users, but also for law enforcement and forensic practitioners [11, 12]. As such, it shows that no sooner has one form of malware been discovered and analysed, then another shows up with distinctly differing properties. What was developed initially as a tool to assist researchers has been used by others for more sinister means. Malware is insidious at its best and performs its task without the users' consent, or claimed to be so [13, 14].

Finding, tracking, capturing and preventing malware infestations on a computer system are akin to war being waged. Therefore, when a crime is committed, it is up to the forensic investigator to search for the quickest means to determine who the guilty party is. Forensics analysts do not have the manpower to spend countless hours analysing and disassembling malware to come to a decision. A platform that would enable collaboration between malware analysts and forensic examiners could assist the latter in determining how the crime took place and whether it would be possible to convict a suspect. The objective would be to reduce the time spent searching for malware and create an easy reference for the investigator. To this end, the Malware Analysis Tool makes accessible all the information required by the investigator in a searchable format. Not only does it perform well as a reference tool for all types of malware, it also speeds up searches for registry keys, malware, affected APIs and a host of other Window's related areas.

As mentioned previously a future improvement to the manual process of analysis of malware and simultaneously inputting the data into the MAT would be to have a program which automatically records the results as the analysis is taking place, extracting the required data as the process continues. Once the analysis is complete, the form could be automatically uploaded to the MAT database. Ideally the format of the data form would also include a XML version, as this information would be accessible in more ways than just a web page.

6 REFERENCES

[1] Griffin, Brad (6 November 2006) An Introduction to Viruses and Malicious Code [online] available from <http://www.securityfocus.com/infocus/1188> [5 October 2007]

[2] Brain, Marshall (n. d.) How computer viruses work [online] available from <http://computer.howstuffworks.com/virus.htm/printable> [17 November 2007]

[3] Griffin, Brad (6 November 2000) An Introduction to Viruses and Malicious Code [online] available from http://www.securityfocus.com/infocus/1188, http://www.securityfocus.com/infocus/1189, http://www.securityfocus.com/infocus/1190 [5 October 2007]

[4] PARC (n. d.) Innovation Milestones [online] available from <http://www.parc.com/about/history/default.html> [18 November 2007]

[5] AntiVirusWorld.com (n. d.) Computer Worm [online] available from <http://www.antivirusworld.com/articles/computer-worm.php> [17 November 2007]

[6] GFi (n. d.) How do Trojans Work? [online] available from <http://kbase.gfi.com/showarticle.asp?id=KBID001671> [17 November 2007]

[7] PrevX (n. d.) MSSPA.EXE [online] available from <http://www.prevx.com/filenames/2190786876915996974-X1/MSSPA.EXE.html> [17 November 2007]

[8] CERT/CC (15 December 2003) Before You Connect a New Computer to the Internet [online] available from <http://www.cert.org/tech_tips/before_you_plug_in.html> [17 November 2007]

[9] Aycock, John (2006) Computer Viruses and Malware. New York: Springer

[10] Leyden, John (17 October 2003) Caffrey acquittal a setback for cybercrime prosecutions [online] available from <http://www.theregister.co.uk/2003/10/17/caffrey_acquittal_a_setback/> [17 November 2007]

[11] Bradbury D (2006), 'The metamorphosis of malware writers', Computers & Security, Volume 25, Number 2, 89-90.

[12] Forte D (2005), 'Spyware: more than a costly annoyance', Network Security, Volume 2005, Issue 12, 8-10.

[13] George E (2003), Case Note 'UK Computer Misuse Act- the Trojan virus defence Regina v Aaron Caffrey, Southwark Crown Court, 17 October 2003', Digital Investigation, Volume 1, Number 2, 89.

[14] Brenner S and Carrier B with Henninger J (2005), 'The Trojan Horse Defense In Cybercrime Cases', Santa Clara Computer and High Technology Law Journal, Vol 21, Issue 1.

THE USE OF FILE TIMESTAMPS IN DIGITAL FORENSICS

Renico Koen[1], Martin S. Olivier[2]

[1]ICSA
University of Pretoria
South-Africa
[2]ICSA
University of Pretoria
South-Africa

[1]renico.koen@gmail.com, [2]martin@mo.co.za

ABSTRACT

Digital evidence is not well perceived by the human senses. Crucial pieces of digital evidence may simply be missed by investigators as the forensic significance of seemingly unimportant pieces of collected data may not be fully understood. This paper will discuss how abstract pieces of information may be extracted from seemingly insignificant evidence sources such a file timestamps by making use of correlating evidence sources. The use of file timestamps as a substitute for missing or corrupt log files as well as the information deficiency problem surrounding the use of timestamps will be discussed in detail. A prototype was developed to help investigators to determine the course of event as they occurred according to file timestamps. The prototype results that were obtained as well as prototype flaws will also be addressed.

KEY WORDS

Digital Forensics, Event Reconstruction, Reco Platform, Timestamps.

THE USE OF FILE TIMESTAMPS IN DIGITAL FORENSICS

1 INTRODUCTION

Digital evidence is not well perceived by the human senses [10]. Crucial pieces of digital evidence may simply be missed due to the fact that examiners do not fully comprehend how seemingly useless pieces of data can be converted to evidence of high value. This situation may be very problematic for digital investigators as it may help to create an incomplete picture of digital crimes under inspection [2]. It is therefore extremely important to examine all evidence, no matter how insignificant it may seem.

If an investigation team can understand an intruder's modus operandi, it may be possible to determine various attributes describing the intruder, such as skill level, knowledge and location [3]. Security mechanisms such as log files will usually be used to determine the actions of the intruder. Unfortunately it is possible that active security systems on the compromised system may be configured incorrectly or disabled completely [9]. In such circumstances investigators will have to turn to alternative sources of digital evidence.

File timestamps may serve as a worthy alternative, as timestamp information may be viewed as a simplistic log of events as they occurred. Although file timestamp information may be considered one-dimensional in a sense that it only records the time of the very last action that was performed on a file, it may still be a valuable source of evidence when very few alternatives remain. Unfortunately the processing of file timestamp information may be complicated by the sheer volume of available timestamps that should be processed.

The overabundance of digital evidence that need to be processed in small amounts of time could be described as an audit reduction problem [4]. The audit reduction problem describes the situation in which the presence of too much information obscures the focus point of investigations. Audit reduction would therefore be prevalent in digital evidence analysis due to the masses of files that needs to be inspected, spurred on by massive storage capacities of modern storage devices.

File timestamps analysis is an excellent example of the audit reduction problem: modern hard drives storage capacity may be anywhere in between

the gigabyte to terabyte ranges; a very large number of files may be found on these devices — each file having different timestamp information associated with it. Although most of the file timestamps would be irrelevant to a case, a few may still be the key to its successful resolution. If these timestamps are simply overlooked, an incorrect conclusion could potentially be reached which may have dire consequences in store for the accused as well as the investigation team.

This paper will discuss the use of timestamps as a supplement or alternative to log files when log files are not available. The information deficiency problem, which describes the situation in which not enough information is available to allow investigators to get a clear picture of forensic significant events, will be discussed. This is done to inform the user of possible problems that may be experienced with alternative evidence sources. The concept of synergy applied to digital data is proposed as a solution to the information deficiency problem. The principle should allow investigators to use various insignificant evidence sources to generate abstract forms of information that are considered to be of forensic value. The paper is structured as follows. Section 2 will discuss the importance of file timestamps. Section 3 will focus on file timestamps related to incident phases. Section 4 will introduce the information deficiency problem and section 5 will discuss a possible solution to the problem. Section 6 will discuss the development of a prototype, section 7 will discuss the results obtained and section 8 will discuss the prototype flaws. Finally, section 9 will describe future work and section 10 will discuss the conclusion.

2 FILE TIMESTAMPS AS A SOURCE OF EVIDENCE

Attackers may try to delete or alter log files in an attempt to cover their tracks; fortunately pieces of information may still remain due to a lack of skills or access rights [9]. As an example, consider the use of well-known UNIX commands such as cat and grep. The attacker may use these two commands to remove identifying information from a system log file. A clever attacker may even change the log file's modification date after the alteration as not to arouse any suspicion from the system administrator. With the system log files compromised, investigators will have to find an alternative source of evidence as compromised evidence sources may not be credible in a court of law.

Fortunately there exists a less obvious source of digital evidence — file

timestamps. Consider the example mentioned previously: the attacker used a combination of well-known tools such as the cat and grep commands to remove identifying information from the system log file. Very few attackers would actually reset the file access timestamps that were created when the shell command was executed. Even if they did manage to modify the file access times, they would have used a tool to do so. This means that although the commands used by the attacker do not have valid timestamps associated with it, a valid timestamp would be left somewhere on the system by the attacker, unless the command was executed from a read-only medium.

From the discussion it should be obvious that only extremely skilled attackers would be able to access a system without leaving a single trace; less skilled attackers are bound to leave small pieces of evidence behind that may ultimately be used to identify the responsible parties.

Popular file systems such as FAT, NTFS and EXT store file timestamps to keep record of:

- The file creation time
- Last time the file was accessed
- The last time the file was modified

These timestamps are updated by the underlying operating system when appropriate, but skilfully written applications also have the ability to manipulate timestamps as they require. Applications have different approaches concerning the management of timestamps. As an example, consider two well-known UNIX applications, namely cp and tar. When a file is copied using the cp command, the resulting creation and modification timestamps of the destination file would indicate the time that the cp command was executed. This is not the case with the tar command. When a compressed archive is created, the relevant files, along with their timestamps, are stored in a compressed archive. It should therefore be noted that some applications will posses timestamp modification capabilities which may have a negative effect on the timestamp analysis process. This topic will be discussed further in section 8.

3 TIMESTAMPS AND INCIDENT PHASES

Three digital evidence stages have been identified by Koen and Olivier [6] which classify evidence according to its temporal relationship with a digital incident. The identified stages are as follows:

- Pre-incident
- Incident
- Post-incident

The pre-incident stage focuses primarily on forensic readiness. Forensic readiness describes the extent to which a system is able to supply forensically-sound information to aid the digital investigation process [7]. Special software and hardware can be installed to monitor user actions and minimize the likelihood that the users of these systems can participate in mischievous activities without being noticed through policy management and the enforcement of restrictions. Suspicious activities may be captured and logged as required. The incident stage is concerned with the capture of digital evidence while a crime is being committed. The incident stage is primarily responsible for the capture and archiving of events as they occur in real time. The last stage is the post-incident stage in which the entire suspect and/or victim system's state is captured and analyzed after the digital crime has been committed. The phase is characterized by the mass-archiving of the states of the systems involved in the digital crime in an attempt to determine how the systems were used and by whom.

The information supplied by timestamps is very limited in a sense that a timestamp only records the last time a specific activity took place. To simplify this discussion, it will be assumed that a file will only have a single timestamp associated with it. Although this is not the case in reality, the principle will stay the same for timestamp-based information.

The most accurate timestamp from an evidence timeline classification point-of-view would be the timestamp recorded in the pre-incident stage as a timestamp with a time earlier than the incident means that the file in question was used before the incident occurred. This means the file may have executed an action on files involved with the incident, but it could only have done so up until the point that it was last loaded in memory. Timestamps captured

in the incident stage indicate that the files in question were used during the incident stage, but could also have been used during the pre-incident stage. The situation gets worse in the post-incident stage: files with timestamp in this stage may have had actions performed on them during any one of the phases. An information deficiency problem therefore exists with regards to timestamps and the incident stage and especially the post-incident stage.

For analysis purposes it will have to be assumed that evidence had actions performed on it in every stage prior to its current incident stage. A solution to the information deficiency problem may be to introduce additional evidence sources in an attempt to build a timelines that indicate upper and lower bound incident stages in which actions were performed on the object in question.

4 APPLICATIONS AND FILE TIMESTAMP RELATIONSHIPS

In order for a timestamp to change an action is needed. The action will have to be triggered by an application or device driver resident in memory at the time of change. For this discussion it is assumed that three types of timestamps exist, namely the creation, modification and access timestamp and that the operating system alone can modify file timestamp values. The value of the timestamp is not important in this example as its meaning is largely dependent on the application that triggered the event. What should be considered important is the fact that an executable code that triggered the event to be executed should have been active in physical memory prior to triggering the event. This means that the file in question should have been loaded into memory, thus modifying its file access timestamp. An executable that accesses or modifies a file should therefore have an file access timestamp which is smaller than the file in question's timestamp (create, read or modify depending on the action performed). The following macro can be defined to determine if an application's create, access or modify time has been edited:

$$touched(f) = ceil\ (\ create(f),\ access(f),\ modify(f)\ )$$

Using the defined macro, the following condition should therefore hold:

$$access(executable) <= touched(file)$$

Unfortunately due to the information deficiency problem identified previously, a piece of executable code may be loaded again in the future which means that the stated condition will not hold anymore as the access time of the executable code changed. The following situation may therefore exist:

access (executable) <= touched(file) or access (executable) >= touched(file)

This basically means that it would not be possible to pinpoint the application responsible for the modification of a file as not enough information exists. If the timestamp found on an executable piece of code shows that the executable was last accessed before a file timestamp was last modified it does not necessarily rule out the executable as the accessory or modifier of the file in question as some sections of code may stay resident in memory for a period of time before it actually accessed the file. It can therefore be concluded that application/file timestamp relationships is of very little forensic significance on its own; some additional form of information is needed to help to rule out executables that could not have modified the file in question. The executable access timestamp cannot be used to help rule out the application associated with it as the application may have been resident in memory for some time before it triggered the modification of a file's timestamps. If it were possible to prove that the application in question was removed from memory some time after its file access timestamp indicated, it may be enough to rule out the application as the trigger source.

As an example, consider the diagram illustrating the executable access timestamps in the different incident stages (see figure 1). Various evidence artefacts have been organized according to file creation timestamp dates. As discussed previously, application/file timestamp relationships are not of forensic significance on their own; executable 1, 2 and 3 could therefore individually have created files A, B, C, D and E. It is therefore not possible to rule out any executables from the equation.

Imagine the intruder managed to reboot the system in question during the incident phase. Knowledge of this event may help to place an upper-bound on the last possible time that executable 1 could have had an effect on the file timestamps of the listed artefacts. File access information informs us that executable 1 was last executed during the pre-incident phase; a system log file (collaborating evidence) shows us that the system went offline during the incident phase. Executable 1 was not loaded again after the system went back online after the reboot. It can therefore be concluded that executable

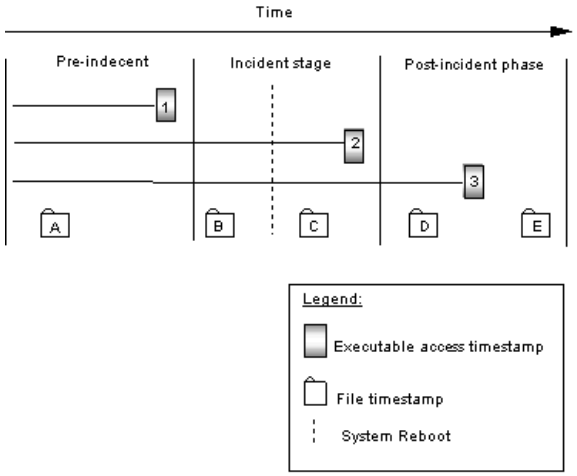


Figure 1: Files organized according to timestamp information.

1 did not have an effect on the timestamps of the listed artefacts after the reboot. With enough collaborative evidence at hand it may be possible to narrow the list of possible executables down substantially which may have been responsible for triggering an event that modified timestamp information. This example relied on the knowledge that a system rebooted. Normally such information will be gathered from a system log file, but in the absence of credible log files, investigators may once again need to turn to file access timestamps as an indicator of system events. When a system boots, various executables are loaded as services. These executables are usually only loaded once and stay loaded until a system halts or reboots. By looking at the access timestamps of these services it may actually be possible to determine when the system booted. This method will be discussed further in section 6.

5 SOLUTION TO THE INFORMATION DEFICIENCY PROBLEM

Synergy describes the situation in which the whole is greater than the sum of the parts [11]. Although the discussed events may be seen as insignificant on their own, their importance may increase when their collective importance is realized, therefore when a state of synergy is achieved.

Consider the example in figure 1 again: each of the events that caused changes in timestamp information associated with the files is of very little forensic value when considered on their own. Even the timestamp that indicated that the system in question performed a boot operation would seem relatively useless on its own as it does not convey any useful information other than a system boot took place. The real value in the timestamp information lies in the fact it represents events that took place. On a higher level these events may be related with one another to create an abstract view of the events as they occurred.

The example in the previous section illustrates that it may be possible to extract useful information from seemingly useless data when viewed on its own. A file's access timestamp may have very little importance on its own; its importance is directly related to the importance of the event that it represents. A principle based on synergy that focuses on the creation of abstract evidence information from insignificant pieces of data may therefore be formulated as follows:

Event data is generated when a significant digital event occurs. Although the generated event data is of little value when viewed independently, collectively event data can produce information that can help investigators to deduce relationships between events to produce abstract views of the evidence at hand.

Investigators usually have lots of complex questions to answer in a short period of time [3]; the possibility therefore exists that evidence may be overlooked as investigators focus their attention to evidence that seems more important in an attempt to save valuable time. Identifying the relationships that may exist between seemingly unimportant pieces of digital evidence may be an extremely tedious task to perform. As Adelstein [1] points out, it is not feasible for investigators to manually analyze storage devices with storage capacities in access of gigabytes as there is just too much data to process. Without some form of automated processing the benefit obtained as a result of time invested by investigators would be minimal due to the sheer volumes of data that needs to be processed.

6 PROTOTYPING

A prototype has been created based on application/timestamp relationships discussed previously in an attempt to illustrate the defined principle in action. The prototype was developed to extract information from Linux-based EXT2/3 file system storing ordinary files, applications and system logs. The prototype was built under the assumption that the file timestamps have not been tampered with. It has also been assumed that the executable access time indicated the last time the application was loaded by the operating system. File creation timestamps were ignored as it is assumed that file access and modification times will always be larger than a file's creation time.

Casey [2] proposed a certainty scale that may be used to determine the level of trust that can be placed in the information deduced by the investigators by examining the forensic evidence at hand. Evidence that appears highly questionable will have a low certainty level associated with it while evidence that can be correlated with other captured evidence sources will receive a higher certainty rating. Casey's certainty scale can be used in addition to the defined principle to increase the level of trust experienced with extracted information; evidence which can be correlated with other sources of information may experience a higher degree of certainty.

Relating Casey's work and the defined principle to timestamp information it can be assumed that timestamp information that is correlated with timestamp information from the same disk image will have a lesser degree of certainty than timestamp information that may be related to some other form of evidence, such as system logs. The prototype was built with the purpose of identifying the last possible time that an application could have been loaded in memory, known as the last possible execution time. This was done in an attempt to determine which files could have been modified by the application in question. The last possible execution time is determined in one of two ways: by correlating an application's access timestamp with system log entries or by correlating an application's access timestamp with the access timestamps of system applications and/or files that are accessed on system boot or shutdown events.

The first method would obviously be the better choice for the correlation of evidence as it contains a rich source of system-related history information. To determine the last possible time an application could have been in memory is simple: use the application in question's access timestamp and search for the earliest system halt or reboot event that occurred after the

access time in the log file. The time specified in the log for the halt or reboot event would therefore serve as the last possible execution time as the executable was never accessed again after that specific point in time. The second method may serve as an alternative to log files in situations when it has become evident that the system log files have been tampered with or in environments where no log files exist. When an operating system boots or halts, it will load various system applications and access stored settings, changing their accessed timestamps. The timestamps viewed on their own are insignificant, but when used to determine when a system was turned on or off, it may be of great value to forensic investigators. As an example, consider the sequence of events that occurs when a standard Linux system boots. The first process created by the kernel executes the /sbin/init application. When the /sbin/init application starts, it reads the /etc/inittab file for further instructions. By simply checking the accessed timestamps of either one of the two files it would be possible to determine the last time that a system booted. It can be argued that the information is also obtainable from alternative sources (such as the /proc/uptime file), but in situations where the alternative is damaged or simply does not exist, timestamps will have to suffice. Calculating the last possible execution time for the second technique is similar to the method used to determine the last possible execution time for the first method: determine an application's accessed timestamp information and determine the last time a system booted or halted by looking at the applications and files associated with the system boot or halt operations.

The prototype reads disk images to produce XML files containing timestamp information. These XML files are then converted to scatter charts to improve the way timestamp information is perceived by the human senses. The prototype depends on two freely available libraries, namely the Reco Platform [5] and JFreeChart [8]. The design is illustrated in figure 2.

The Reco Platform supplies low-level EXT2/3 support to the system while the JFreechart library supplies the graphing functionality required by the application. The prototype source code has been released under the GNU GPL license and is available on Sourceforge [5]. The next section will discuss the results that were obtained using the developed prototype in more detail.

7 RESULTS

The prototype was tested using Linux (Fedora Core 4). A disk image was made and last possible execution times were computed for each application

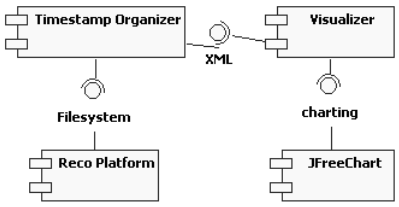


Figure 2: The prototype design.

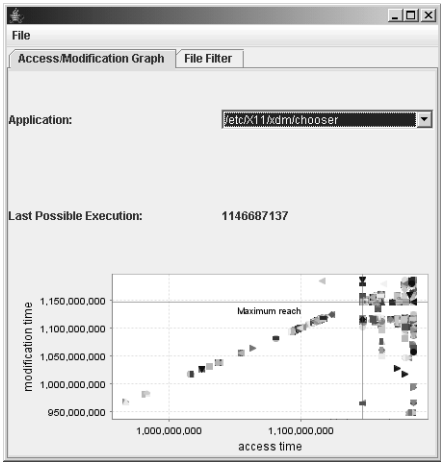


Figure 3: A screenshot of the prototype.

using both methods described previously to produce separate XML files. A scatter chart was constructed using each detected file's modification and access times as coordinate values. A selected application's last possible execution time was plotted as horizontal and vertical lines to indicate the reach (in terms of what the application could have modified) of the application in question. Figure 3 illustrates the produced scatter chart as well as the horizontal and vertical lines indicating the maximum reach of the application in question.

The user is allowed to select an application of interest in a dropdown control populated with a list of applications. The application's last possible execution time is computed and plotted on the scatter chart upon selection.The last possible execution time, access time and modification times are repre-

sented by an integer value; the integer value is a timestamp that describes the amount of seconds that have elapsed since January 1,1970 (which means that the values could easily be manipulated using a function such as ctime) when the event in question occurred or should occur. In the example (figure 3) the last possible execution time for the application */etc/X11/xdm/chooser* was calculated to be 1146687137 seconds since 1 January 1970. Translated to human-understandable terms, the last possible execution time for the application in question is Tuesday, May 2, 2006 at 23:58:57. The application cannot be responsible for any file access or modification operations performed after the last possible execution time, represented by the horizontal and vertical lines on the graph. Any files outside of the horizontal and vertical lines will therefore have been accessed or modified by other applications.

By simply looking at the generated chart it is possible to visually detect which files could have been modified by the application in question. Due to the sheer magnitude of the amount of files that are stored on a disk drive, a file filter functionality has been added to the prototype to search for files with timestamps conforming to specific criteria. Determining the names of the files that could have been modified by the application in question was as simple as submitting a filter query that contained the last possible execution time of the application in question.

A comparison between the two techniques used to determine an application's last possible execution time yielded the results that were expected: since system log files contain detailed history information, more accurate last possible execution times could be calculated leading to more accurate results. File access timestamps contain only the last time the file was accessed and can therefore be compared to a log file containing entries which date back to the last time a system in question was booted. This implies that the method could work with the same efficiency as the first method in a scenario where a system rarely goes offline. However, this method would be very inaccurate for systems that goes offline frequently.

8 CRITICISMS

As discussed in section 2, some applications have the ability to modify timestamps. The work in this paper assumed that the timestamps are modified by the operating system only and did not take into account that applications may manipulate the proposed analysis method by changing file timestamps to render the method invalid. In reality, interpreted meaning of a timestamp

is therefore largely dependent on the way in which the application responsible for the creation or modification of a file manages timestamp information.

It has also been assumed that applications will be stored on a writable medium; an application's timestamp information will therefore be updated each time the application is loaded into memory. This may not necessarily be the case as it is possible in UNIX environment to mount file systems in read-only mode. This means that an application's file access time will not change rendering the method described in this paper useless.

Another concern is that an application may have accessed or modified a suspicious file prior to its last possible time of execution; if the suspicious file was accessed or modified again some time later in the future (presumably after the application in question's last possible time of execution), the timestamp may be labelled as being out of reach of the application in question. Technically this is true as the file was last modified by another application, but this situation may not always be desirable. A way to overcome this problem is to divide application timestamps into the various incident stages discussed in section 3. Only applications with access timestamps falling in the incident and post-incident phase will have to be considered for inspection as it can be assumed that applications with last possible execution times falling in the pre-incident stage were not involved with the incident in question.

9 FUTURE WORK

A complex application would typically touch various files while it is executing. A typical scenario would be where the application in question first accesses its configuration files and then data files. By describing an application's actions formally, it may be possible to create a profile that accurately describes an application's file access characteristics.

Another topic that requires attention is the inspection of the file access of an operating system's boot process. When an operating system performs the boot process, various files will be accessed. Different operating systems would access different files which creates the possibility that the file access operations performed by an operating system could potentially be used as a fingerprint to help operating system identification in circumstances in which conventional methods are not deemed appropriate. The described process could potentially be improved by adding the concept of a termination signature. The termination signature describes the characteristics of an application when it terminates, in other words what actions it takes just before

it terminates. If such a signature can be incorporated into the concepts described in this paper, more accurate results may be obtained.

10 CONCLUSION

This paper discussed how timestamps could be used to rule out files that could not have been modified by distinct applications based on an application's calculated last possible execution time. A principle was introduced based on the concept of synergy claiming that insignificant pieces of event datum may collectively be of significant forensic importance. A prototype was constructed based on this principle, using timestamps as a source of insignificant evidence. The prototype calculated various applications' last possible execution times and visually depicted the information in a manner that can easilly be understood by the observer. The prototype helped to visualize abstract digital data which are not well-perceived by the human senses to help investigators to easily understand the produced data as well as its importance. Unfortunately the method used by the prototype is not absolute in a sense that it cannot successfully be applied to all environments under all conditions. It has become evident that a great need exists for ways in which digital evidence can be visualized. More research will have to be conducted to find ways to visualize digital information to allow investigators to easily understand digital evidence at hand.

References

[1] Adelstein, F. Live forensics: diagnosing your system without killing it first. *Commun. ACM 49*, 2 (2006), 63–66.

[2] Casey, E. Uncertainty, and loss in digital evidence. *International Journal of Digital Evidence 1*, 2 (2002).

[3] Casey, E. Investigating sophisticated security breaches. *Commun. ACM 49*, 2 (2006), 48–55.

[4] Corey, V., Peterman, C., Shearin, S., Greenberg, M. S., and Bokkelen, J. V. Network forensics analysis. *IEEE Internet Computing 6*, 6 (2002), 60–66.

[5] Koen, R. Reco platform homepage. Online: http://sourceforge.net/projects/reco, June 2007.

[6] Koen, R., and Olivier, M. An open-source forensics platform. In *SAICSIT '07. Proceedings of the Annual SAICSIT conference* (2007).

[7] Mohay, G. Technical challenges and directions for digital forensics. In *SADFE '05: Proceedings of the First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05) on Systematic Approaches to Digital Forensic Engineering* (Washington, DC, USA, 2005), IEEE Computer Society, p. 155.

[8] ORL. JFreechart. Online: http://www.jfree.org/jfreechart, Online: July 2007.

[9] Stallard, T., and Levitt, K. Automated analysis for digital forensic science: Semantic integrity checking. In *ACSAC '03: Proceedings of the 19th Annual Computer Security Applications Conference* (Washington, DC, USA, 2003), IEEE Computer Society, p. 160.

[10] Wang, S.-J. Measures of retaining digital evidence to prosecute computer-based cyber-crimes. *Comput. Stand. Interfaces 29*, 2 (2007), 216–223.

[11] Wikipedia. Synergy. Online: http://en.wikipedia.org/w/index.php?title=Synergy, July 2007.

UML MODELLING OF DIGITAL FORENSIC PROCESS MODELS (DFPMs)

Michael Köhn[1], J.H.P. Eloff[2], MS Olivier[3]

[1,2,3]**Information and Computer Security Architectures (ICSA) Research Group**
Department of Computer Science
University of Pretoria
South Africa

[1]**mkohn@cs.up.ac.za,** [2]**eloff@cs.up.ac.za,** [3]**molivier@cs.up.ac.za**

ABSTRACT

A number of forensic processes have been used successfully in the field of Digital Forensics. The aim of this paper is to model some of these processes by using the Unified Modeling Language (UML) - specifically the behavioural Use Cases and Activity diagrams. This modelling gives a clear indication of the limitations of these processes. A UML-based comparison is made of two prominent DFPMs that are currently available in the literature. This is followed by a newly proposed DFPM as developed by the authors.

KEY WORDS

Digital Forensics, Digital Forensic Process Model, Process Modelling, Unified Modelling Language, UML

UML MODELLING OF DIGITAL FORENSIC PROCESS MODELS (DFPMs)

1 INTRODUCTION

The authors of this paper argue that a Digital Forensic Process Models (DFPM) in particular and the field of digital forensic investigations in general can benefit from the introduction of a formal modelling approach. In this paper we propose that UML [1] would be a suitable paradigm for modelling forensic processes. Most of the modelling representations for forensic investigations found in the current literature are made in a rather informal and intuitive way [**?**, 2]. Thus it is argued that because of the value of a forensic investigation and the formal field of forensic investigation can benefit from introducing a formal modelling approach. Some of these formal modelling approaches include Z-specification, relational algebra and UML modelling. UML modelling is the vehicle chosen for this paper because it provides a structured and behavioural approach that is needed for a forensic investigation. UML is an accepted formal specification for the modelling of processes. This paper will focus on modelling two existing DFPMs, that of Kruse [3] and that of the United States Department of Justice (USDOJ) [4]. The UML that will be used will be limited to Use Case and Activity Diagrams.

Digital forensics has experienced a number of rapid advances to date. This can be seen in the tools that have been developed for forensic investigations such as Encase [1] and Forensic Tool Kit (FTK) [2]. These tools try to encompass the whole digital forensic process into one tool. Encase, which has done this with great success has been accepted in the United States and other countries as a reliable forensic investigation tool [5]. A number of the tools that do not form part of the greater investigation are nevertheless of some use and do assist. Knoppix [3] is one such tool that offers limited forensic capability. In the event of encountering a computer that is turned off, it could aid the investigator in possibly finding material without tampering with the integrity of the data. From this it is clear that a digital forensic investigation is made up of multiple facets, which include technology, procedure and legal components. Thus it seems that there is a need for an integrated DFPM.

[1] Encase online: http://www.guidancesoftware.com/
[2] Access Data online: http://www.accessdata.com/
[3] Knoppix online: http://www.knoppix.org/

A number of DFPMs that have been developed since 2000 aim to assist the investigator in reaching a conclusion upon completion of the investigation. DFPMs used in investigations with success include — but are not limited to — those proposed by Kruse [3], the United States Department of Justice (USDOJ) [4], Casey [6], Reith [7] and Ciardhuin [2].

According to the Oxford online dictionary, the term forensic is defined as "relating to or denoting the application of scientific methods to the investigation of crime" and "of or relating to courts of law"[4]. From this definition it is clear that the ultimate goal of a digital forensic investigation is to present some form of evidence in a court of law using the correct legal procedures with scientific backing.

Closer examination of DFPMs reveals no apparent problem, but a number of questions do arise. Who are the actors that will interact with the system or defined process? Are the role players clearly defined? Do some of these models have short comings? Is it possible to combine some features of existing DFPMs in order to construct an ideal DFPM? To answer these questions, a formal way of comparison is needed to explore some of these problems.

The remainder of the current paper is structured as follows. Section 2 presents some background to the paper and refers to related work performed with regard to forensic processes. In section 3 the Kruse and USDOJ DFPM is modelled in UML using Activity and Use Case Diagrams. Some comments are also made on these two DFPMs. Section 4 contains the result of a brief comparison between the Kruse and USDOJ DFPMs. Section 5 introduces a new integrated model called InteDFPM, which combines the Kruse and USDOJ DFPMs. The paper is concluded in Section 6.

2 BACKGROUND AND RELATED WORK

Digital forensics has been accepted as the process of "analytical and investigative techniques used for the preservation, identification, extraction, documentation, analysis and interpretation of computer media (digital data) which is stored or encoded for evidentiary and or/or root cause analysis" [8]. Most of the proposed DFPMs use some elements of the above definition as point of departure for the development of such a process, such as [3, 4, 6, 9, 7, 2]. These DFPMs are listed in Figure 1. The names of the DFPMs are given in the left margin, while the processes included in each of these models are

[4]The Oxford Dictionary: http://www.askoxford.com

listed along the top.

	Acquire	Authenticate	Analyze	Collection	Examination	Reporting	Recognition	Identification	Individualisation	Reconstruction	Preservation	Classification	Presentation	Decision	Preparation	Approach Strategy	Returning Evidence	Awareness	Authorization	Planning	Notification	Transpoprtation	Storage	Hypothosis	Proof/defence	Dissemination
Kruse	*	*	*																							
USDOJ			*	*	*	*																				
Casey							*			*	*	*														
DFRWS			*	*	*			*			*		*	*												
Reith			*	*	*			*			*		*		*	*	*									
Ciardhuain				*	*								*					*	*	*	*	*	*	*	*	*

Figure 1: Current DFPMs

The investigation phase of the process constitutes the main focus of most DFPMs. In [4, 9, 7] examination, analysis and collection are included, as this is where most of the activities taking place as part of the investigation are conducted. This focus on investigation is dangerous for a number of reasons. Forensics generally should have a goal of presenting evidence in some form and providing some factual basis to substantiate the investigation's finding.

In the analysis of some of the DFPMs as seen in Figure 1 one can clearly see the additions that have been made over time. These DFPMs have become increasingly complex. The terminology used in the models is a factor that contributes to creating this unnecessary complexity. Many terms are quite similar to those used in other DFPMs to describe a similar concept. For example, 'Acquire' used in the Kruse DFPM and 'Collect' used in the USDOJ DFPM would probably amount to the same process — the activities may overlap in many respects.

On examining Figure 1, the reader may agree that there is indeed a need to refine these DFPMs in order to create an integrated model that encapsulates components derived from the given/selected few DFPMs.

3 UML MODELLING

For the purposes of this paper we will be modelling the Kruse and USDOJ DFPMs. The two different types of behavioural UML models that are used

will be the Activity and Use Case Diagrams. Only a high-level system depiction will be presented in all diagrams.

3.1 Kruse

The Kruse model of computer forensics consists of three main processes or phases. The first is acquire the evidence while ensuring that the integrity of the data is maintained. Secondly, authenticate the acquired data, while checking the integrity of the extracted data against the original data. Authentication in digital forensics is usually done by comparing data of the original MD5 hash with the copied MD5 hash [10]. Thirdly, analyse the data without tainting the integrity of the data. This process involves the most intense part of the investigation into the Kruse model.

It is also worth mentioning that the Kruse DFPM is designed specifically for computer-related crimes [3].

3.1.1 UML Activity Diagram

The Kruse DFPM Activity diagram is represented in Figure 2.

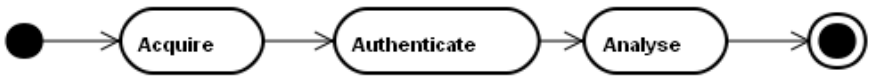


Figure 2: Kruse Activity diagram

The three processes follow one after the other: Acquire, Authenticate and then Analyse. These processes commence with a starting state and end with a finishing state.

3.1.2 UML Use Case Diagram

The Kruse DFPM Use Case is represented in Figure 3. This figure also depicts the different role players.

The three main role players that interact with the system are the Investigator, the Prosecution and the Defense. The Investigator can be specialised to a First Responder, which can be any one of the following: Emergency Response Team or System Administrator. The Prosecution and the Defence will be role players in a criminal matter only. The system consists of three

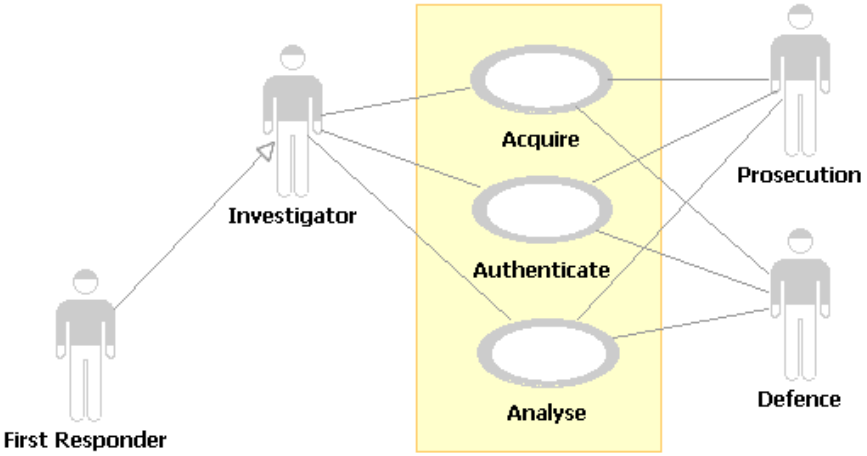


Figure 3: Kruse Use Case Diagram

Use Cases: Acquire, Authenticate and Analyse. The system boundary is depicted by the large rectangle containing the three use cases.

3.1.3 Comments on Kruse DFPM

It should be noted that this is truly an oversimplification of the Kruse DFPM. Each of the use cases in Figure 3 and the activity diagram in Figure 2 will be expanded to include subprocesses.

The activity diagram is clear and it is obvious to see that an investigation starts, runs its course and stops. The main concern is that no real evidence document or report is generated during the investigation. The Kruse DFPM however states in its specifications that documentation and chain-of-custody reports should be maintained during each of the processes.

The use case clearly indicates that the investigator will interact with each one of the processes. Kruse states that in many instances the investigator will not be the same person. The 'Acquire' activity is always encountered by the First Responder and the other two use cases can be performed in a laboratory environment. The court is mentioned throughout the specification, but there is no clear interaction with the system.

3.2 United States Department of Justice (USDOJ)

The USDOJ [4] model accounts for four phases namely collection, examination, analysis and report. The collection phase involves searching for the evidence, recognising that the evidence would be applicable to the specific case, collecting the evidence, while documenting every step taken in the process. The main aim of the second phase, examination, is to reveal any hidden or obscure data. The origin of the original data and its significance are important in providing a visual output that will be used in the analysis process. The third phase involves analysis and the visual product of the examination process is the input to this analysis. Here a case will be built and evidence will be constructed to prove the particular crime. Baryamureeba [**?**] states that the analysis phase will also determine the probative value, which would actually be the function of the courts. The outcome of this phase would be to produce evidence that would serve to prove the elements of a specific crime. Every step is also documented throughout. The final phase in the USDOJ model is the report phase. During this phase a complete report will be compiled to document the process followed from the beginning of the investigation. The product will be the final evidence report presented in court. Contained in this document is the chain-of-custody report, complete investigation documentation and presentable evidence.

One of the design principles in the USDOJ DFPM is to abstract the process from any specific technology [4].

3.2.1 UML Activity Diagram

The Activity Diagram of the USDOJ DFPM is given in Figure 4.

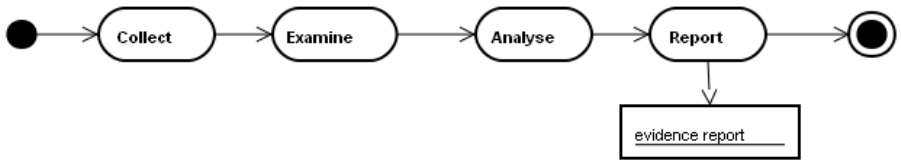


Figure 4: USDOJ Activity diagram

The process commences with a starting state. The data is collected from the digital device, after which it is examined and then analysed. During the

report phase, an evidence report is created as an object output. After the completion of the evidence report, the process stops.

3.2.2 UML Use Case Diagram

The Use Case diagram of the USDOJ DFPM can be seen in Figure 5.

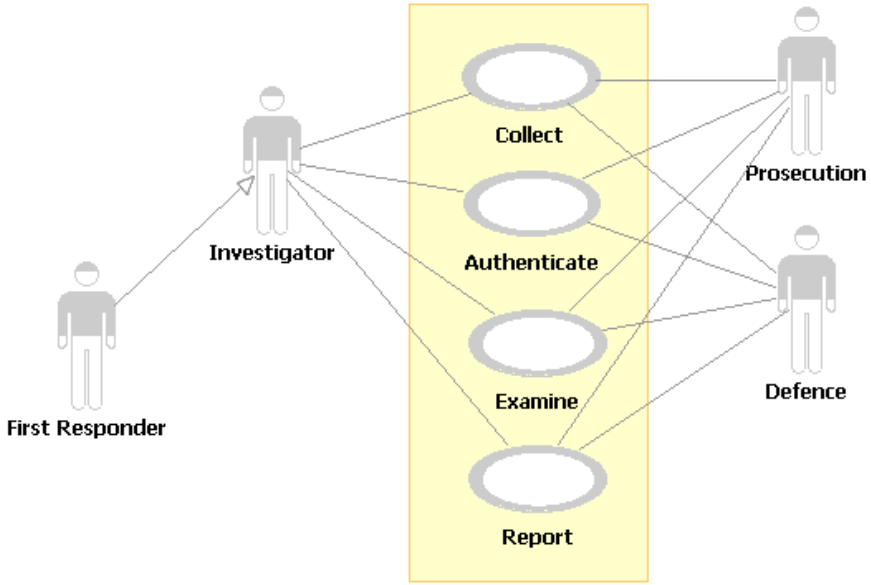


Figure 5: USDOJ Use Case Diagram

In Figure 5 there are three actors: the Investigator, the Prosecution and the Defence. The First Responder is a specialisation of Investigator. An Investigator can be any one of the following: police officer, manager or a forensic investigator. The DFPM is specifically set up for First Responders. There are four use cases in the system, namely, Collection, Examination, Analysis and Reporting.

3.2.3 Comments on the USDOJ DFPM

In the USDOJ Activity diagram, the processes are executed one after the other. There is one apparent difference, which involves the fact that during the Reporting process an evidence report is generated as an output. This

will ultimately be used in a matter before the court. The evidence report will contain all the evidence collected during the investigation, including the chain-of-custody document and presentable evidence. It should be noted that the current paper will not consider what a court considers to be presentable evidence.

The Use Case diagram in Figure 5 does not show the court as a role player that interacts with the system. In the USDOJ specification the court is often mentioned, but no emphasis is placed on the fact that the court ultimately will evaluate the presented evidence report in its finding. There is also no clear indication as to how and when the court must evaluate the document. Nevertheless, an important contribution by the USDOJ DFPM is the fact that an evidence report document is in fact produced.

4 COMPARISON BETWEEN THE TWO DFPMs

Similarities between the Kruse and USDOJ DFPMs are apparent: Firstly, although the models use different terminology ('Acquire' and 'Collect') to describe the first phase, the processes are actually the same. For our purposes we will refer to both as 'Collect' in the remainder of the paper. Secondly, both models have an 'Analysis' phase, resulting in an Analyse process.

There are however also a number of significant differences that cannot be ignored. These include the fact that Kruse's DFPM explicitly validates the integrity of the data in an authentication process, while the USDOJ DFPM includes an examination process. The latter might not always be needed, as data is often hidden and obscured from an investigator. This process will also compromise the integrity of the data. Finally, the DFPM of the USDOJ includes the compilation of a report process, while the Kruse DFPM does not.

5 InteDFPM: INTEGRATED DFPM

The Kruse and USDOJ DFPMs have been modelled using UML Activity and Use Case diagrams. In this section we propose to integrate and expand the two DFPMs into a combined DFPM containing the best elements of both DFPMs. This combined model is called the InteDFPM.

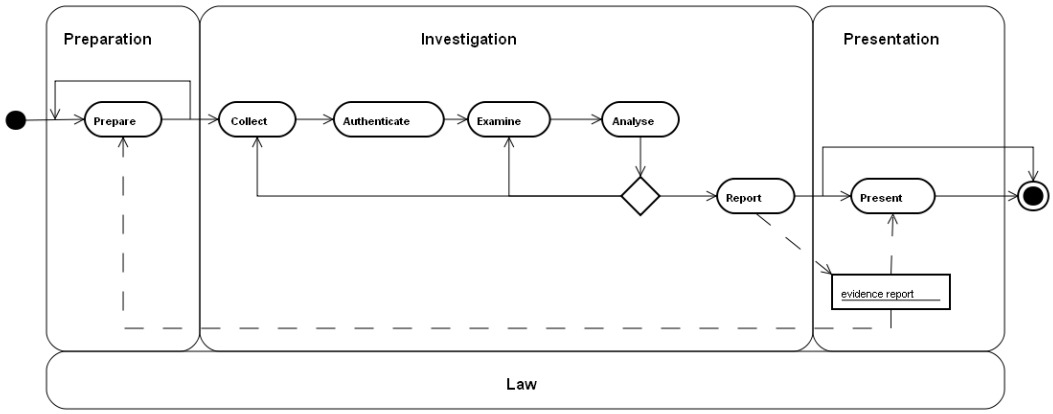


Figure 6: InteDFPM Activity Diagram

5.1 UML Activity Diagram

Figure 6 shows the InteDFPM superimposed on a framework proposed by Köhn et al [11]. This framework has three phases: Preparation, Investigation and Presentation. Note that the sub-processes are not included. The law is the foundation for this framework as illustrated by the row along the bottom of Figure 6. The implication is really to ensure that everything is based on sound legal principles so as to withstand legal scruiteny in court.

Two processes have been added to the Activity diagram to integrate with the Köhn framework. These are 'Prepare' in the Preparation phase and 'Present' in the Presentation phase.

The whole process is triggered by a criminal action (not indicated in Figure 6), which constitutes the starting point. Prepare is the first step and will not be elaborated on here. The rest of the processes follow logically — from Prepare to Collect, Authenticate, Examinate and then Analyse. Authentication, is included between the Collection and Examination steps to ensure the data integrity of the data before the Examination is started. Examination can modify the contents of the data such as in the case of hidden files, compressed files and other forms of data obfuscation. The data has to be authenticated before any of this happens. If this is not done, there might be a dispute in court concerning the validity of the material.

A decision point follows the Analysis process. The primary investigator will consider whether to examine more data or to collect more data from the

original source. Once this decision point has reached depletion an evidence report is compiled as part of the Report process. This process will include the compilation of presentable evidence, chain-of-custody reports and complete documentation compiled during the investigation phase. The evidence document is the output of the Investigation phase.

Eventually the evidence report will be an input to the Presentation process. This is where the court will also have the opportunity to evaluate the evidence. It should be noted that the present process can be excluded in the event of not finding sufficient evidence or other relevant factors.

The court finding will be an input to further investigations. This will help investigators to prepare for unforeseen factors that were previously unknown.

5.2 UML Use Case Diagram

Figure 7 illustrated the Integrated Use Case Diagram for the combined Kruse and USDOJ DFPMs.

Figure 7 corresponds to a large extent with the separate Kruse and USDOJ Use Case diagrams. Collect, Authente, Examine, Analyse and Report are the required use cases.

The system will interact with the following role players: the Investigator can be specialised to be either a First Responder or Other. A specialised Investigator can be any type of Investigator specified by a number of DFPMs. The Investigator will interact with almost all the use cases. It must be noted that it is not always the same person investigating the data. Thus the Investigator does not remain the same person throughout the course of the investigation.

The Prosecution and Defense will be interested in the steps taken in each of the use cases. The Court will examine the evidence report generated by the Report use case. The Court's interaction will change when there is a dispute about the steps taken during investigation. In such a case the Court will evaluate all the use cases. Ultimately, the Court will be interested only in the findings presented in the evidence report, and it will reach a finding based on the presented evidence. The Court will also determin the admissibility and weight of each of the pieces of evidence included in the evidence report.

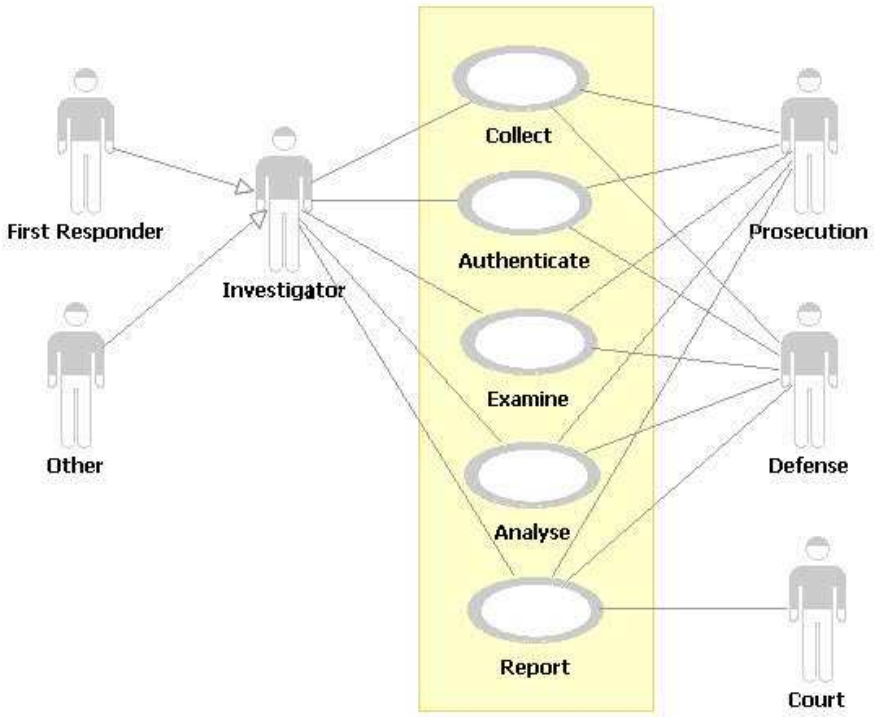


Figure 7: InteDFPM Use Case diagram

6 CONCLUSION

The aim of this paper was to model two DFPMs from the current literature. Activity and Use Case diagrams from the behavioural UML specification were used for this purpose. An Integrated DFPM (InteDFPM) was proposed by combining the Kruse and USDOJ DFPMs, after which the InteDFPM was superimposed on a framework proposed by Köhn [11]. The InteDFPM Use Case Diagram was also presented.

By modelling the DFPMs using UML, it becomes clear that there are a number of shortcomings in the design of the DFPMs. Who are the role players that interact with the system? Neither the Kruse DFPM nor the USDOJ DFPM makes any definitive statement on who the role players should be, except that there must be an Investigator. Furthermore, both DFPMs use different terminology. These problems have been addressed in the paper.

A very important action that is missing both in the above architecture and in the DFPM is the criminal act itself. Future work should explore the possibility of including the criminal act and subsequently including it into the InteDFPM. Other DFPMs should also be investigated for possible incorporation into the InteDFPM.

References

[1] G. Booch, J. Rumbaugh, and I. Jacobson, *The Unified Modeling Language User Guide*. Addison Wesley, 1999.

[2] S. O. Ciardhuain, "An extended model of cybercrime investigations," *International Journal of Digital Evidence*, vol. 3, 2004.

[3] W. Kruse and J. Heiser, *Computer Forensics: Incident Responce Essentials*. Addison Wesley, 2002.

[4] Technical Working Group for Electronic Crime Scene Investigation, *Electronic Crime Scene Investigation: A Guide for First Responders*, United States Department of Justice, 2001.

[5] T. Wilsdon and J. Slay, "Towards a validation framework for forensic computing tools," in *First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05)*, 2005, pp. 48–55.

[6] E. Casey, *Digital Evidence and Computer Crime*. Elsevier Acedemic Press, 2004.

[7] M. Reith, C. Carr, and G. Grunsch, "An examination of digital forensic models," *International Journal of Digital Evidence*, vol. 1, 2002.

[8] S. van Solms, C. Louwrens, C. Reekie, and T. Grobler, "A control framework for digital forensics," in *IFIP 11.9*, 2006.

[9] Digital Forensics Research Workshop, *A Road Map for Digital Forensics Research*, 2001. [Online]. Available: http://www.dfrws.org/2001/dfrws-rm-final.pdf

[10] A. Culley, "Computer forensics: past, present and future," *Digital Forensics*, vol. 8, pp. 32–36, 2003.

[11] M. D. Köhn, J. H. P. Eloff, and M. S. Olivier, "Framework for a digital forensic investigation," in *Information Security South Africa (ISSA)*, H. S. Venter, Ed., 2005.

SPAM OVER INTERNET TELEPHONY AND HOW TO DEAL WITH IT

Dr. Andreas U. Schmidt[1], Nicolai Kuntze[1], Rachid El Khayari[2]

[1]Fraunhofer-Insitute for Secure Information Technology SIT
Rheinstrasse 75, Germany
[2]Technical University Darmstadt
Germany

[1]{andreas.schmidt|nicolai.kuntze}@sit.fraunhofer.de,
[2]rachid.el.khayari@googlemail.com

ABSTRACT

In our modern society telephony has developed to an omnipresent service. People are available at anytime and anywhere. Furthermore the Internet has emerged to an important communication medium.
These facts and the raising availability of broadband internet access has led to the fusion of these two services. Voice over IP or short VoIP is the keyword, that describes this combination.
The advantages of VoIP in comparison to classic telephony are location independence, simplification of transport networks, ability to establish multimedia communications and the low costs.
Nevertheless one can easily see, that combining two technologies, always brings up new challenges and problems that have to be solved. It is undeniable that one of the most annoying facet of the Internet nowadays is email spam. According to different sources email spam is considered to be 80 to 90 percent of the email traffic produced.
Security experts suspect that this will spread out on VoIP too. The threat of so called voice spam or Spam over Internet Telephony (SPIT) is even more fatal than the threat that arose with email spam, for the annoyance and disturbance factor is much higher. As instance an email that hits the inbox at 4 p.m. is useless but will not disturb the user much. In contrast a ringing phone at 4 p.m. will lead to a much higher disturbance.
From the providers point of view both email spam and voice spam produce unwanted traffic and loss of trust of customers into the service.

In order to mitigate this threat different approaches from different parties have been developed. This paper focuses on state of the art anti voice spam solutions, analyses them and reveals their weak points. In the end a SPIT producing benchmark tool will be introduced, that attacks the presented anti voice spam solutions. With this tool it is possible for an administrator of a VoIP network to test how vulnerable his system is.

KEY WORDS

SPAM, Internet Telephony, VoIP, SPIT, attack scenarios

SPAM OVER INTERNET TELEPHONY AND HOW TO DEAL WITH IT

1 INTRODUCTION

In the following sections we will discuss the problematic of SPAM over Internet Telephony. The first section will deal with a general SPIT explanation and classification, followed by a scientific SPIT threat analysis.
In the second section the state of the art in SPIT prevention mechanisms will be presented and their weaknesses analysed. In the last section we will take a look at our SPIT benchmark tool (SXSM - SIP XML Scenario Maker) and how this tool can exploit the weaknesses of anti SPIT mechanisms.

2 SPIT VERSUS SPAM

The focus of this paper is set on the topic of so called SPAM over Internet Telephony (SPIT). The first aspect to mention is, that although SPIT contains the phrase "SPAM" and has some parallels with email spam, it also has major differences. The similarity is that in both cases senders (or callers) use the Internet to target recipients (or callees) or a group of users, in order to place bulk unsolicited calls [3]. The main difference is that an email arrives at the server before it is accessed by the user. This means that structure and content of an email can be analysed at the server before it arrives at the recipient and so SPAM can be detected before it disturbs the recipient. As in VoIP scenarios delays of call establishment are not wished, session establishment messages are forwarded immediately to the recipients. Besides this fact the content of a VoIP call is exchanged not until the session is already established. In other words if the phone rings it is too late for SPIT prevention and the phone rings immediately after session initiation, while an email can be delayed and even, if it is not delayed, the recipient can decide if he wants to read the email immediately or not.
In addition to these aspects another main difference between spam and SPIT is the fact, that the single email itself contains information, that can be used for spam detection. The header fields contain information about sender, subject and content of the message. A single SPIT call in contradiction is technically indistinguishable from a call in general. A SPIT call is initiated and answered with the same set of SIP messages as any other call.

3 INTUITIVE SPIT DEFINITION

SPIT is described very similar in different publications and the descriptions can be summarised as "unwanted" , "bulk" or "unsolicited" calls. In [2] e.g. SPIT is defined as "unsolicited advertising calls", which is already a special form of SPIT. In [3] SPIT is defined as "transmission of bulk unsolicited messages and calls" which is a more general definition than the first one, as it doesn't characterise the content and includes also messages. Note that with this definition it is not clear, if the term "messages" is used in order to generalise the type of messages that are sent (e.g. "SIP INVITE" or "SIP OPTIONS" messages) or, if it is used in order to include SPAM that is sent over Instant Messages (SPIM = SPAM over Instant Messages). The most precise definition is found in [1] where "Call SPAM" (as the authors call it) is defined as "a bulk unsolicited set of session initiation attempts (e.g., INVITE requests), attempting to establish a voice, video, instant messaging, or other type of communications session". The authors of [1] go even one step further and classify that "if the user should answer, the spammer proceeds to relay their message over the real-time media." and state that this "is the classic telemarketer spam, applied to SIP". We can easily see, that the presented definitions so far are very similar, but differ in their deepness.

4 SPIT ANALYSIS

The problem with the definitions above, is that they are either to specific or to general. In order to find a more precise definition, we have to analyse how SPIT is put into execution and what the goal of the initiator of SPIT is.

In practice the initiator of SPIT has the goal to establish a communication session with as much victims as possible in order to transfer a message to any available endpoint. The attacker can fulfil this via three steps. First the systematic gathering of the contact addresses of victims. Second is the establishment of communication sessions with these victims and the third step is the sending of the message.

In the following we will not only discuss, why the process of information gathering is part of the SPIT process, but we will also see that it is the basis of any SPIT attack. In order to contact a victim, the attacker must know the SIP URI of the victim. We can differ permanent SIP URIs (e.g. sip:someone@example.com) and temporary SIP URIs (e.g. sip:someone@192.0.2.5).

4.1 Information gathering

At first we will take a look at information gathering of permanent SIP URIs. If an attacker wants to reach as many victims as possible he must catalogue valid assigned SIP URIs. The premisses for the **Scan attack** are the possession of at least one valid account and knowledge about the scheme of SIP URIs of the targeted platform (e.g. provider).
Let us assume the attacker has a valid SIP account at SIP provider "example.com" and he wants to scan the provider's network, in order to achieve a list of valid permanent SIP URIs. Let us also assume that the provider "example.com" distributes SIP URIs that correspond to the following scheme: The user name of the SIP URI is a phone number that begins with the digits "555" followed by 4 more random digits. All phone numbers from "5550000" to "5559999" are valid user names of this provider. As the attacker has now knowledge about all valid user names, he must find out which of them are already assigned to customers and which of them are not. The attacker can now step through the whole list of valid SIP URIs and send adequate SIP messages to each URI and receive information about the status of the tested URI. The simplest way is sending an INVITE message to each SIP URI and analyse the answer of the SIP Proxy. If the SIP URI is not assigned, the SIP Proxy may answer with a "404 Not found" response, if the SIP URI is assigned but the user is not registered at the moment, the SIP Proxy may answer with a "480 Temporarily unavailable" response and if the SIP URI is assigned and the user is registered, the call will be established and answered with a "200 OK" response. When the attacker has stepped through the whole list and marked all possible SIP URIs, he has a list of assigned SIP URIs, that can be used for future attacks. Note that it is not necessary that the scan attack must be fulfilled with an INVITE message, we just discussed this way as the simplest way, because it already leads to the desired session establishment. The attacker could also use an OPTIONS request or a REGISTER request and analyse the reaction of the Proxy. Mainly the implementation of the targeted Proxy decides on which message will grant the desired information. Some Proxies e.g. respond to all OPTIONS requests with a "200 OK" message, even in case of an invalid or unassigned SIP URI.
Now we will take a look at gathering of temporary SIP URIs. Temporary SIP URIs consist of the user name part and the host part. The user name part is usually a string or a phone number and the host part is the IP, where the endpoint can be reached directly. If an attacker has already generated a list

of valid assigned SIP URIs, he now additionally needs the corresponding IP addresses of the SIP URIs. In some Proxy implementations the temporary SIP URIs are published in the "Contact" header of the response message to a request. In this case the desired information is achieved in the same way as the permanent SIP URIs. If the proxy does not provide the IP address in the SIP responses, the attacker must use a more complex method to achieve the desired information. Let us assume this time that "example.com" is an Internet Service and VoIP provider. The provider assigns IP addresses of the range 192.0.2.5-192.0.2.155 to his customers and SIP URIs with the same scheme as described above (555XXXX). Let us assume The customers have hardphones (e.g. analogue telephone attached to VoIP ready router or Analog telephony adapter). With this knowledge the attacker can step through the list of IP addresses and try sending an adequate SIP request (INVITE,OPTIONS) directly to the endpoint (e.g. to UDP Port 5060) and analyse the responses in the same way as described above. The attacker can populate a list of temporary SIP URIs. Note that the temporary SIP URIs are only valid for a short time period (max. 24 hours), as customers are usually forced to disconnect their internet connection after a certain period. Although this procedure is harder to fulfil than the first one, it has the major advantage, that the attacker doesn't need valid accounts as premiss. Because the Proxy is not involved and SIP messages are sent directly to the victim, the attacker can use any SIP identity he wishes as source address. The client can not verify the identity, as nearly all existing implementations of clients accept SIP messages from any source.
Now that we have seen how lists of permanent or temporary SIP URIs can be achieved, we will discuss the usage of them.

4.2 SPIT session establishment

When the attacker has collected a large number of contact addresses, he can begin session establishment to the victims. Which list he must use (temporary or permanent URIs) depends on the communication infrastructure he wants to use. We can distinguish two possible ways of session establishment: The attacker can establish a session with sending an INVITE message via Proxy, which we can call **SPIT via Proxy** or he can establish a session with sending an INVITE message directly to the endpoint without involving the Proxy, which we can call Direct IP Spitting. For SPIT via proxy the attacker only needs a list of permanent SIP URIs and for Direct IP Spitting he needs

the list of temporary SIP URIs. Again for SPIT via Proxy the attacker needs at least one valid user account and for Direct IP Spitting he doesn't need a valid account at all.

4.3 SPIT media sending

The last step of the SPIT process is the media sending after the session has been established. Which type of media is sent, depends on the scenario in which the SPIT attack takes place. The best scenario classification can be found in [2] and defines three types of SPIT scenarios:

- Call Centres: In Call centres a computer establishes a call to an entry of the catalogue and then dispatches the call to a call centre agent who will then talk to the callee.
- Calling Bots: A calling bot steps through the list of gathered information, establishes a session and then sends a prerecorded message.
- Ring tone SPIT: Some VoIP telephones come pre-configured in a way that they accept a special SIP header information called "Alert-info" which may contain an URL pointing to a prerecorded audio file somewhere on the Internet. Obviously, this can be used to play advertising messages before the call has even been accepted by the user just as the phone is ringing. An adaption of this method could be a SPIT attack where the attacker just wants to let the victims phone ring, in order to disturb the victim. In this special case no media is sent at all and the session is terminated as soon as the phone rings (e.g. when a "180 Ringing" is received). Obviously this is the most annoying facet of SPIT.

4.4 SPIT summary

As we can see now the SPIT process is very complex and has different aspects which have to be considered in order to develop countermeasures. The general definitions that we discussed in the first section are insufficient as a basis of discussion and do not cover all facets of the problem. In general we can say, that Spitting describes the systematic scanning of a VoIP network with the target of gathering information about available user accounts and the systematic session establishment attempts to as many users as possible in order to transfer any kind of message.

5 SPIT COUNTERMEASURES AND THEIR WEAKNESSES

In the following sections we will discuss state of the art SPIT prevention mechanisms in order to point out their advantages and disadvantages. The countermeasures are ordered by type and not by publication. As a matter of fact most publications define a set of countermeasures as a solution to mitigate SPIT. Nevertheless we will discuss every method on it's own and not the orchestration of different mechanisms. Note that only those techniques are listed, that have crystallised in research.

5.1 Device Fingerprinting

The technique of active and passive device fingerprinting is presented in [4] and is based on the following assumption:
Having knowledge about the type of User Agent that initiates a call, helps finding out whether a session initiation attempt can be classified as SPIT or not. The assumption is based on the analogy to e.g. HTTP based worms. As described in [4] these types of worms have different sets of HTTP headers and different response behaviour, when compared to typical Web browsers. So if we can compare the header layout and order or the response behaviour of a SIP User Agent with a typical User Agent, we can determine if the initiated session establishment is an attack or a normal call.
The authors describe two types of techniques that can be used for that purpose "Passive and Active Device Fingerprinting".

5.1.1 Passive Fingerprinting

The e.g. INVITE message of a session initiation is compared with the INVITE message of a set of "standard" SIP clients. If the order or appearance of the header fields does not match any of the standard clients, the call is classified as SPIT. The fingerprint in this case is the appearance and the order of the SIP header fields. The authors of [4] present a list of collected fingerprints of standard hard and soft phones.

5.1.2 Active Fingerprinting

User Agents are probed with special SIP messages and the responses are analysed and compared with the response behaviour of standard clients. The fingerprint in this case is the returned response code and the value of certain

header fields. If the fingerprint doesn't match any of the standard clients, the call is classified as SPIT. The authors recommend the sending of specially crafted standard compliant and non compliant OPTIONS requests, in order to analyse the response behaviour of a client.

5.1.3 Weakness of Device Fingerprinting

The weakness of passive fingerprinting is described by the authors of [4] themselves. As passive fingerprinting only analyses the order and existence of the header fields of an INVITE message, an attacker simply needs to order the header fields in the same way as one standard client. In that case the passive fingerprinting mechanism can't detect the attack.
We can state nearly the same for active fingerprinting, as an attacker only needs to behave like one standard client when receiving unexpected or non standard compliant SIP messages. It is very simple for an attacker to develop an attacking SIP client that behaves exactly like a standard client, as he can use the same SIP Stack or imitate the behaviour of SIP Stack of a standard client. We can call this attack **Device Spoofing** and any attacker, who is able to spoof a device can nott be identified.
As Device Fingerprinting is discussed as a server side anti SPIT mechanism, it is useless against Direct IP Spitting as the clients don't have any chance to verify the fingerprint of the attacking client.
In the end we will take a look on practical issues of Device Fingerprinting. When we take a look at today's VoIP universe, we will find out that there exist a vast variety of hard- and softphones. Each of this phones has it's own SIP Stack and even within a product family header layouts and behaviour differ even between two versions of the same device. The result is, that an administrator who uses Device Fingerprinting in order to protect his system, must always keep the list of fingerprints up to date. Comparing the INVITE message of a caller with an old or incomplete fingerprint list, can lead to blocking the call although the call is not a SPIT call. Let us e.g. assume that a caller uses a standard client and that the manufacturer sends out a firmware upgrade, that makes major changes to the SIP Stack. Any calls of this user are blocked or marked as SPIT, until the administrator of the VoIP network updates the fingerprint list and this procedure will repeat any time a new firmware version is rolled out or new clients are released.
Taking it even one step further, we can see, that as more and more clients and versions are released, the fingerprint list will become wider and wider

and in the end nearly any combination of e.g. header fields will be present in the list. The main problem of device fingerprinting is that it is derived from a HTTP security technique. In that scenario only few clients (web browsers) from few developers exist, in contradiction to the VoIP world.

5.2 White Lists, Black Lists, Grey Lists

The White List technique is presented e.g. in [2] [1] and works as follows: Each user has a list of users that he accepts calls from and any caller who is not present in the list will be blocked. In addition the private White Lists can be distributed to other users. If e.g. a caller is not present in the White List of the callee, White Lists of other trusted users can be consulted and their trusted users (up to a certain level), however this technique needs additional mechanisms. Black Lists are the contradiction of White Lists and contain only identities, that are already known as spammers. Any call from a caller whose identity is present in the callee's Black List is blocked. Even Black Lists can be implemented as distributed Black Lists, where a callee can consult the Black Lists of other users. Grey listing works as follows: On initial request of an unknown user (not in White List) the call is rejected and the identity is put on the Grey List. As stated in [2] in case the caller tries calling back within a short time period, the call will be accepted. An adaption of this technique is described in [1] as Consent Based Communication. In case of Consent Based Communication the call of an unknown caller is initially blocked and put on the Grey List. The callee can consult the Grey List and decide, if he will accept future calls from this identity or block it permanently.

5.2.1 Weaknesses of White Lists, Black Lists, Grey Lists

Black Lists can not really be viewed as a SPIT countermeasure, because additional methods are needed to classify a caller a Spitter. A Black List on server side would require e.g. statistical methods for classifying a caller as Spitter. In case of a client side Black List, the user must mark a caller as a Spitter, e.g. after receiving an initial SPIT call from this caller. Both server side and client side Black List are very useless against Direct IP Spitting for different reasons. Server sided Black Lists are bypassed by Direct IP Spitting, because the SIP messages are sent directly to the client. Client sided Black Lists are circumvented by Direct IP Spitting, because the caller can take on any identity in order to place calls. So if one identity is blocked he can simply switch the Identity. We can call this attack **SIP Identity Spoofing** and

any attacker who can spoof SIP identities, can easily bypass Black Lists. White Lists are at first sight harder to circumvent than Black Lists, because the attacker has no knowledge about the entries of the White List of the victim. So even if he wants to spoof an identity, the attacker doesn't know which identity he must take on, in order to place a successful call. In case of Direct IP Spitting the attacker could simply try out all existing accounts with a brute force attack until he finds out which identities are not blocked. A less exhausting procedure can be performed in case of distributed or imported white lists [2]. In that scenario the attacker needs one valid account. After adding the victim to the attacker's white list, he can now select that he wants to import the white list of the victim. So he can get access to all entries of the victim's white list and can spoof these identities e.g. in a Direct IP Spitting attack. The Grey List mechanism can be bypassed the same way as White List mechanisms, as it just represents a mechanism that allows first time contact. All in all we can say, that any attacker who is able to perform SIP Identity Spoofing, can bypass Black Lists, White Lists and Grey Lists. In the end we will take again a look at the practical side of the presented mechanisms. The concepts of Black, White and Grey Listing are derived from the Instant Messaging world, where it is a matter of course, that users first ask for permission, before they are added to another user's buddy list and only buddies can communicate with each other. When a user receives a communication request, he receives the profile of the other user containing e.g. nick name, email address, full name or even profile photo. On basis of this information, the user can decide and is able to decide, if he wants to accept messages in future from that party or not. Taken to the VoIP scenario this mechanism seems very impractical as the introduction problem has to be solved. Let us assume e.g. an employee of a bank wants to call one of his customers. In case of white listing the call can not be successfully routed to its target, as customers usually don't have the phone numbers of employees of their home bank listed in the White List. The decision basis for accepting or rejecting a call is simply the phone number that is sent by the caller. If the call is rejected at first (Grey listing) the callee must decide if he wants to accept future calls and he must base this decision on the phone number. We can easily see that this fact is very impractical.

5.3 Reputation Systems

Reputation based mechanisms are described in [5] or in [1] and can be summarised as follows: After receiving a call, the callee can set a reputation value for the caller, that marks this caller as Spitter or not. This reputation value must be assigned to the identity of the caller and can be used for future session establishment requests. This technique can be used e.g. as attachment to Grey listing [1] in order to provide a better decision basis. The authors of [5] explain that the user feedback can be used additionally for calls that were not detected by other SPIT preventing components. The way the reputation value is generated can differ. The SPIT value can be e.g. an additional SIP header, or included in a special error response code or distributed via SIP event notification mechanism. Reputation systems can be either based on negative or positive reputation values. This means that in first case only Spitters are marked with negative values or in second case "normal" callers are marked with positive values.
An adaption of this method can be found in [6] where user feedback is combined with statistical values in order to calculate a reputation value. The reputation value is e.g. composed of a value representing the number of times an identity occurs in other users' Black Lists, call density, call length or similar statistic values. The assumption behind this approach is that the calculated value will differ much between "normal" users and spitters.

5.3.1 Weakness of Reputation Systems

Reputation systems that are based on negative reputation can be bypassed in same way as Black Lists [1]. A user with a negative reputation can be viewed as globally blacklisted as his calls are blocked e.g. for any user (this depends on the policy that is used). Nevertheless an attacker that is black listed simply needs to gain access to a new "clean" account. In case of a SPIT value as SIP header, the SPIT value can be spoofed by the attacker (e.g. with Direct IP Spitting) and we can call this attack **SIP Header Spoofing**. The attacker can simply set or change values of header fields, when he uses Direct IP Spitting.
In addition an attacker can create several accounts with the aim of pushing the SPIT value of one account up or down (depending on implementation). This attack can be called **Reputation Pushing or Pulling**.
Again we will also take a closer look of practical issues of the anti SPIT mechanism. At first we must admit, that Reputation systems are more auxiliary

features than SPIT blocking mechanisms. The reason for this argumentation is, that the user must classify a call as SPIT via a button or by entering a value. This value is used for future decisions on that SIP identity. So initially SPIT is not prevented by this technique. Then the SPIT value of an identity has to be shown to callees, so that they can decide about accepting or rejecting the call. Let us assume a Spitter has achieved a SPIT value or SPIT probability of e.g. thirty percent and then calls a victim. What should happen now? When the call is forwarded to the user and the value is e.g. shown in the display of the callee's phone, he can decide to accept or reject the call on a better decision basis. The problem is that anyhow his phone rings and that is what should be prevented. He could have just picked up the call and listened the first 5 seconds to know that it is SPIT. So the SPIT value didn't just add one percent of benefit. On top of this fact attackers could misuse the scoring system and create enough accounts in order to threaten "normal" users with collectively giving them negative reputation [1]

5.4 Turing tests, Computational Puzzles

Turing test are tests where the caller is given a challenge, that a human can solve easily and that is hard to solve for a machine. Therefore Turing tests or CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) are tests, that countermeasure Calling Bot attacks in VoIP scenarios. Turing tests in VoIP scenario work as follows: On initial call establishment attempt, the caller is transferred to an interactive System where he is challenged with a task e.g. dialing 5 digits that he is hearing (so called Audio CAPTCHA). While the numbers are read out background music or any other kind of noise is played, so that speech recognition systems can't be used to solve the task. A human caller in contradiction will solve the task without difficulties and only if the task is solved, the call will be forwarded to its destination. Turing tests can be used in combination with white lists, solving the introduction problem as described in [8].
Computational Puzzles seem at first sight very similar to the Turing tests concept. As described in [7] a SIP Proxy or User Agent Server can request from a User Agent Client (caller) to compute the solution to a puzzle. The goal of this method is to raise CPU costs of a call and so reduce the number of undesirable messages that can be sent. Turing test in contradiction have the goal to block non-human callers, as described above. According to [7] the

puzzle, that has to be solved, could be finding a pre-image that will SHA1 hash to the correct image. This means that the UAC will be challenged with a SHA1 hash of a value and the UAC must find out (by computing it) which value has been hashed.

5.4.1 Weakness of Turing tests and Computational Puzzles

Turing tests seem at first sight very effective for SPIT prevention in combination with white lists, but nonetheless have weak points. The first approach of bypassing Audio CAPTCHA is relaying the CAPTCHA to human solvers. An attacker could pay cheap workers, who are only hired to solve Audio CAPTCHA. In countries with cheap labour this would raise the costs per call only marginally [1]. In order to reduce the costs, an attacker could even e.g. set up an adult hotline and could dispatch Audio CAPTCHA to the customers of this service. This technique is known from visual CAPTCHA where the images from CAPTCHA protected sites are copied and relayed to a high traffic site owned by the attacker. All in all we can state, that an attacker who can detect CAPTCHA and relay it to human solvers is able to bypass Turing tests and we can call this attack **CAPTCHA Relay Attack**.
Computational Puzzles can not be viewed as SPIT prevention mechanisms, as attackers usually possess high computational power. So Circumventing a system protected by Computational Puzzles, doesn't even demand a special attack. The attacker just needs sufficient CPU power.
In the end again we will take a look at some practical issues of the described techniques. As far as Turing tests are concerned, we can see, that this method is very intrusive. User Interaction is forced every time a caller is not present in the White List of a callee.
The difficulty with Computational Puzzles is, that different VoIP endpoints have different abilities in computational power. So if the task is to hard to solve (consumes too much CPU power), session establishment will be delayed very much for e.g. a low-end cell phone, while attackers with high CPU power PCs won't be concerned much. With this fact Computational Puzzles are very ineffective and contra productive, as they only bother "normal" users.

5.5 Payments at risk

Payments at risk mechanisms can be used in order to demand payment from an unknown caller. In [1] this technique is described as follows: If user A wants to call user B, he must first send a small amount of money to user B.

When User B accepts the call and confirms that the call is not a SPIT call, the amount will be charged back to user A. With this technique it is possible to raise costs for SPIT callers while keeping "normal" calls cheap. In [1] it is described as an auxiliary technique that solves the introduction problem of White lists, tis means, that payment is only required for callers who are not on the White list of callee. In general the payment could be demanded for every call, but this would make the telephony service more expensive.
An adaption of this method is described in [9], here the Payment technique is used in combination with a SPIT prediction value that is computed at server side. If the SPIT likelihood is high the call is rejected, if the SPIT likelihood is small the call is forwarded to the callee and if the SPIT likelihood value is in between payment is demanded automatically. Only if the payment is fulfilled the call will be forwarded to its target. The difference between the two approaches is, that in the first case the payed amount is only charged back for non SPIT calls and in the second case, callers who reject payment are treated as Spitters.

5.5.1 Weakness of Payment at risk

In which way Payment at risk can be bypassed depends mainly on the way it is implemented. As described demanding payment for each call won't be very realistic, because this would require a high administrative overhead and more costs for service providers. Let us assume Payment at risk combined with White listing as in the first example, so that payment is only required for callers that are not present in the callee's White List. In this case a caller could simply spoof identity as described in the section about White List.
In the second scenario, where Payment at risk is combined with a Reputation system, the attacker just needs to achieve an adequate reputation value, as described in the corresponding section.
Let us even assume, that Payment at Risk is used for every call. Even In that case an attacker could circumvent it, by impersonating as another user, so that he can establish calls and shift the costs on to "normal" customers. In which way this kind of **SIP Identity Hijacking** attack is fulfilled is an other question and out of scope for now.
Besides the technical aspects, practical issues of Payment at Risk are numerous. At first the relative high costs, that are required for micropayment will must be viewed, the inequities in the value of currency between sender and recipient [1] and the additional interactions that a user must take (e.g.

confirming a call from an unknown party as non SPIT).

5.6 Intrusion Detection Mechanisms, Honey phones

Intrusion Detection Systems are (generally described) systems, that can be used for detection of any kind of abnormal behaviour within a e.g. network and so reveal attacks. An implementation of this technique is presented in [10] based on the Bayes inference approach combined with network monitoring of VoIP specific traffic. The Intrusion Detection System is designed as a defence mechanism against different VoIP specific attacks including scan attacks and SPIT attacks. For every attack a conditional probability table (CPT) is defined for variables such as request intensity, error response intensity, parsing error intensity, number of different destinations, max number of dialogues in waiting state, number of opened RTP ports, request distribution and response distribution. Let us look at e.g. the CPT for the number of different destinations variable: For a SPIT attack the likelihood of having more than 7 different destinations is set to 1 and the likelihood of having up to 7 different destinations is set to 0. The concept behind this technique is, that the different attacks affect these variables in different ways, e.g. a SPIT attack usually has a higher probability of a higher number of destinations than normal traffic. So a belief of a network trace can be calculated with the aid of likelihood vectors that were defined in the CPT. In the end the trace can be categorised as an attack or normal trace (refer to [10] for detailed description).
Honey phones can be used as part of an Intrusion Detection Systems as described in [2] [11] and can be viewed as VoIP specific Honeypots. A Honeypot represents a part of a network that is not accessible by "normal" users and therefore any access to the honeypot can be viewed as an attack. VoIP specific honeypots can be used in order to detect Scan attacks or SPIT attacks. As described in [11] the Honeypot is implemented as a complete parallel VoIP infrastructure, that is logically and physically separated from the normal network and so simulates a whole VoIP network. Let us assume a Scan attack as described earlier. When the attacker sends e.g. OPTIONS or INVITE requests to valid assigned permanent URIs they are forwarded through the normal SIP network (Proxy, UAC), but when the attacker tries to send an OPTIONS request to an unassigned or invalid SIP URI the request will be forwarded to the Honeypot, where the requests can be monitored and treated adequately. The authors of [11] propose call analysis in order to determine

attack characteristics, interaction with the originator in order to determine the source of the attack and blocking of the calls, as adequate treatment. The monitoring system of this approach works as follows: A day is divided into sections of specified time (e.g. one hour). For each section a predefined metric is calculated (e.g. number of calls, number of different recipients, average duration of a call) matching predefined events (e.g.call). In the learning phase (e.g. a month), daily statistics are built to extract a long term account profile (e.g. daily average of the number of calls for each section). In the detecting stage (e.g. a day), a short term profile is compared to the long term one by using an appropriate distance function (e.g. Euclidean distance, quadratic distance, Mahalanobis distance). A recent profile which is quite different from the long term one indicates possible misuse. Another method is to study non stationary features of an account, for example the distribution of calls over all callees or the shape of the callees' list size over all dialed calls. By comparing changes of a distribution over the time by using of an appropriate distance function (e.g. Hellinger distance), sudden bursts may be detected and treated as abnormalities [11].

5.6.1 Weakness of Intrusion Detection Mechanisms, Honey phones

Intrusion Detection Systems base on the assumption, that the characteristics of attacks differ much from characteristics of normal calls. At first sight this assumption seems logic, as e.g. within a SPIT attack, the attacker calls hundreds or thousands of victims within an hour, while a normal user wouldn't even send out one percent of this amount of calls. Nevertheless the attacker has two possibilities in order to bypass detection by an Intrusion Detection System. The first is to align his behaviour with the behaviour of normal users, e.g. adjust the call rate to 5 calls per hour. Obviously this technique is hard to fulfil, because this would make an attack very inefficient as it would consume too much time, but on the other hand the goal of a spitter is not to reach as much users as possible within the shortest time period. Reaching e.g. thousand users with a call rate of 5 calls per hour would take approximately 8 days. We can call this technique **Call Rate Adaption**, this means that an attacker is able to adjust his call rate (e.g. number of calls per time slot, number of simultaneous calls). As the call rate is not the only variable that is used in order to detect abnormal behaviour an attacker can use a second technique in order to not be detected by Intrusion Detection Systems. The attacker can use different accounts for his attacks, so

that statistic values are spread over several accounts. Let us assume that an attacker has one hundred valid user accounts. With this amount of accounts he can partition the targeted user accounts into one hundred groups and use only one account per group. The users from group one are only called with account one and so on. It is harder for a monitoring system to detect attacks that are originated from different sources, as there must be a technique to correlate partial attacks to one complete attack. This technique can be called **Account Switching**, as the attacker switches the used account while he is performing an attack.
Honeypots are very effective against scan attacks as anyone who tries to reach invalid or unassigned identities, will be trapped and so Honeypots are very effective against SPIT. When the Spitter can't scan the network for assigned and unassigned numbers, he is forced to view all numbers as assigned. When he views all numbers as assigned, he will sooner or later step into the trap, because he will establish calls to endpoints, that are part of the Honeypot. Nevertheless attackers can trick the Honeypot mechanism with **SIP Identity Hijacking**. When an attacker impersonates the accounts of normal users and then performs SPIT attacks with this normal accounts, he will access end points in the Honeypot system with normal accounts. So the assumption that accesses to the Honeypot are only established by attackers is lapsed.
In the end we will take again a look at the practical issues of the presented solutions. The practical problem with intrusion detection systems in general is, that they base on statistical assumptions that are not verified. The questions that has to be solved is: Where is the borderline between normal usage and abnormal usage? The publishers state that statistical values are assumed or derived from attack characteristics, but in order to reduce the rate of false negative and false positive classifications, the knowledge basis must be precise. So we can say that what we lack, is knowledge of SPIT characteristics as we nowadays can't really distinguish SPIT from normal traffic unless the SPIT attacks are excessive. Honeypots have the disadvantage, that they only detect access to invalid or unassigned accounts, this means that an attacker who only accesses valid accounts won't be handled by a honeypot.

5.7 Summary

We can finally say, that we have seen SPIT countermeasures with different weak points. All of the presented ideas have technical and practical weak

points, that can be exploited by attackers in order to circumvent these technique. An attacker who is able to perform Device Spoofing, SIP Identity Spoofing, SIP Header Spoofing, Reputation Pushing or Pulling, CAPTCHA Relay Attack, SIP Identity Hijacking, Call Rate Adaption, and Account Switching has a good repertoire, that enables him to bypass any of the presented techniques or combinations of them. An attacker now needs a tool that aggregates the presented attacks.

6 SIP XML SCENARIO MAKER

In the following sections we will introduce our SPIT producing benchmark tool, that implements the presented attacking techniques.

6.1 Technical Basis

SXSM is based on SIPp developed by HP [12]. SIPp is an Open Source test tool and traffic generator for SIP. SXSM expands SIPp with the ability to quickly create custom SIP scenarios via a graphical user interface (GUI), execute created scenarios and evaluate the result of the execution. The functionality is fulfilled by two different editing modes and one execution mode.

6.2 Message Editor

The message editor delivers the basis functionality, the ability to create custom SIP Messages. SIP messages are the smallest elements of SIP scenarios, as SIP scenarios are sequences of SIP messages.
Within the message editor, the user can configure the layout of each and every SIP message, that can be used in the scenario editor (explained later). Additionally the messages can be grouped into sets, so that the user can create e.g. two differently composed INVITE messages and put one into set A and one into set B. Later the user can distinguish the two INVITE messages, because they are in different sets. In the message editing mode the user can even compose or modify existing SIPp control messages (e.g. pause commands) that can be used in order to control the behaviour during execution. SXSM comes preconfigured with a set of standard messages that can be used as orientation in order to compose own messages.

6.3 Scenario Editor

The scenario editor is the core element of SXSM. In this mode the user can create SIP scenarios, based on the bricks created in the message editor. Even the scenarios can be grouped in different sets. The user must chose a name for a scenario and can then select from the list of messages, the messages he wants to add to the scenario and in which order they should appear. Afterwards the user can edit the scenario, that is presented as an XML file, in detail. Let us assume the user wants to create a scenario where an INVITE message is sent, then a "100 Trying" is received, then a "180 Ringing" is received then a "200 OK" is received. In this case the user simply selects these messages and adds them to the scenario, saves the scenario file and work is done.

6.4 Shoot Mode

The shoot mode represents the execution mode. In this mode the user can put previously created SIP scenarios into a sequence, execute them one after the other and evaluate the results presented.
The user selects scenarios from the scenario list and adds them to the shoot list, then he configures the call rate (calls per time period), then he enters information about the target (remote IP, remote port) and about himself (local IP, local port). After this information has been provided, he must provide information about the SIP identities, that should be used for the execution. The user can choose, if he wants to use fixed values for both source and target SIP URI or inject values from an external CSV file. In first case he must provide a user name for the targeted SIP URI and this user name will be used for each and every call, that is executed through the shoot list. If the user wants to inject values from an external CSV, he must specify the location of the file. After all parameters have been set, the user can execute the shoot list. SXSM then feeds SIPp with the input data and waits until the scenarios are executed. When the execution is fulfilled, SXSM evaluates the exit codes, that were generated by SIPp. The following exit codes are considered:

- 0: All calls were successful
- 1: At least one call failed
- 97: exit on internal command. Calls may have been processed. Also exit on global timeout

- 99: Normal exit without calls processed
- -1: Fatal error

Based on the exit code a success rate is calculated and displayed. If e.g. all scenarios of the shoot list completed with exit code "0" then the success rate is "100". Additionally log files will be presented for each scenario, that can be used for debugging purpose in case of failed scenarios.

6.5 Using SXSM as attack tool

As SXSM is implemented within a very broad and modular context, it can be used for all SIP testing purposes and in special as a SPIT producing attack tool. In the following we will discuss how the different attacks, that were presented in the previous sections, can be put into practice.

6.5.1 Device Spoofing

The Device Spoofing attack is an attack, that has two facets. As we discussed earlier device fingerprints can be derived from the layout of the SIP messages or from the behaviour. The layout of SIP messages can be manipulated within the message editor of SXSM. If a user wants to imitate the message layout (presence and order of SIP headers) of a device, he simply needs to create a new set of SIP messages, name the set after the device and create the desired SIP messages, according to the wished layout.
The behaviour of the client can be manipulated within the scenario editor. The ability to create scenarios that contain branch points eases this process. A scenario can then contain a section for every message that can be received. The user must only put tests into the scenario with the scheme "if message x is received jump to section y".

6.5.2 SIP Identity Spoofing

Identity Spoofing in its simple form is provided by inserting the wished SIP URI in the "From" and "Contact" header of the SIP messages. If the user wants to inject the SIP URI from an external CSV file he must specify this in the scenario file. The user simply needs to put the expression "[fieldn]" where n represents a number (the column of the CSV file) at the position where the user name is usually placed in the "From" or "Contact" header according to the following scheme:

From: [field0] <sip:[field1]@[local_ip]:[local_port]>;
Here the first column of the "caller.csv" file contains the name of the identity and the second column contains the user name part of the SIP URI.

6.5.3 SIP Header Spoofing

SIP Header Spoofing can be fulfilled with two different approaches. The first one is to create custom SIP messages with the message editor and set headers and header values as wished. The second is using standard SIP messages and change values of headers with the detail view of the scenario editor. The first one should be preferred, if the user wants to create a lot of scenarios with the same header value and the second variant should be used, if the user wants to tweak values only once in a while.

6.5.4 Call Rate Adaption

The Call Rate Adaption attack can be fulfilled within the shoot mode. The user just needs to add a scenario to the shoot list and adjust the values for call rate. He can put e.g. "Scenario X" into the shoot list and set the call rate to 10 times per second and stop as soon as 100 calls have been finished. With this method it is possible to control in detail how many times in what time period a scenario is executed. As one and the same scenario can be present several times in the shoot list, the user can define the behaviour very precisely. So he can e.g. determine that "Scenario X" should be executed 100 times with a call rate of 10 calls per second and then 20 times with a call rate of 2 calls per second. Note that the phrase "call" means one pass of scenario from beginning to end and does not mean that a call is actually placed. A scenario could e.g. consist of sending an OPTIONS request and receiving the answer.

6.5.5 Account Switching

The Account Switching attack is a special form of SIP Identity Spoofing and can be fulfilled by providing an external CSV file with appropriate data. Let us assume the user wants to place 100 calls to hundred different targets with ten different SIP identities. The CSV file for the callee should contain hundred rows and each row should contain the user name of the targeted URI. The CSV file for the caller should contain 10 rows and each row should contain one of the ten user names, that should be used as source. Including

the SIP identities for the caller can be fulfilled with the mechanism described in the section about SIP Identity Spoofing. Including the SIP identities of the target can be configured in the shoot mode by selecting the external CSV file as target.

6.5.6 Reputation Pushing or Pulling

Reputation Pushing or Pulling is very dependent of the implementation, but can be fulfilled in a generic way. The user simply needs to create two scenarios. One, that sends out a call and one, that receives a call. The receiving scenario should include e.g. a positive reputation value into the BYE message. Note, that this is the point where it is implementation specific, as it depends on the implementation of the Reputation System where the reputation value must be put. Then the user must launch two instances of SXSM and shoot out the calling scenario with one instance and the receiving scenario with the other instance. Combining this technique with Account Switching for the call receiving side can lead to the desired effect of Reputation Pushing or Pulling.

6.5.7 SIP Identity Hijacking

SIP Identity Hijacking is again very implementation dependant. but we can take a look at a simple attack derived from [13]. The registration hijacking attack is presented as follows:

1. Disable the legitimate user's registration. This can be done by:
    - performing a DoS attack against the user's device
    - deregistering the user (another attack which is not covered here)
    - Generating a registration race-condition in which the attacker sends repeatedly REGISTER requests in a shorter timeframe (such as every 15 seconds) in order to override the legitimate user's registration request.
2. Send a REGISTER request with the attacker's IP address instead of the legitimate user's

With SXSM this process could be put into practice, by creating scenarios for each of the presented steps and execute them one after the other, but as the Session Hijacking attack has a variety of aspects, that have to be considered we will not discuss this matter in detail for now.

6.5.8 CAPTCHA Relay Attack

The CAPTCHA Relay Attack can be fulfilled with the Third Party Call Control (3PCC) mechanism. With this mechanism it is possible for SIPp (and therefore for SXSM) to create a communication session with several remote endpoints and so relay e.g. calls. The procedure is fulfilled as follows. The Attacker calls the victim, who sends the Audio CAPTCHA. In response the attacker calls human solver and "REFER"s the victim to him. As the result the victim accepts and the human solver solves CAPTCHAs.
As this attacks is as good as the used CAPTCHA detecting algorithm, the technique must be adapted to future implementations of both CAPTCHA generators and detecters.

7 CONCLUSION

This paper described the main aspects, that should be considered by developers of anti SPIT solutions. We saw, that a precise problem definition is mandatory for SPIT research, as we can only cover all aspects, if we clearly know which problems have to be faced. Then we saw, that state of the art anti SPIT mechanism still embody both technical and practical weak points, that can be viewed as vulnerabilities. In the last part of this paper we got an impression of how this vulnerabilities can be abused by attackers, with a simple attack tool, whose power lays in full control over the SIP protocol. The next step of research should consider the presented attacks and develop mechanism that blank out these vulnerabilities.

References

[1] J. Rosenberg, C. Jennings, RFC 5039 - *The Session Initiation Protocol (SIP) and Spam*, IETF, 2008.

[2] M. Hansen, M. Hansen, J. Möller, T. Rohwer, C. Tolkmit and H. Waack, *Developing a Legally Compliant Reachability Management System as a Countermeasure against SPIT*, 2007.

[3] S. Dritsas, J. Mallios, M. Theoharidou, G.F. Marias and D. Gritzalis, *Threat Analysis of the Session Initiation Protocol Regarding Spam*, IEEE, 2007.

[4] H. Yany, K. Sripanidkulchaiz, H. Zhangy, Z. Shaez and D. Saha, *Incorporating Active Fingerprinting into SPIT Prevention Systems*, 2007.

[5] M. Stiemerling S. Niccolini, S. Tartarelli, *Requirements and methods for SPIT identification using feedbacks in SIP. Internet-draft*, 2008.

[6] F. Wang, Y. Mo, B. Huang, *P2P-AVS: P2P Based Cooperative VoIP Spam Filtering*, 2007.

[7] C. Jennings, *Computational Puzzles for SPAM Reduction in SIP. Internet-draft*, 2008.

[8] H. Tschofenig, E. Leppanen, S. Niccolini, M. Arumaithurai, *Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) based Robot Challenges for SIP. Internet-draft*, 2008.

[9] S. Liske, K. Rebensburg, B. Schnor, *SPIT-Erkennung, -Bekanntgabe und -Abwehr in SIP-Netzwerken*, 2007.

[10] M. Nassar, R. State, O. Festor, *Intrusion detection mechanisms for VoIP applications*, 2007.

[11] M. Nassar, S. Niccolini, R. State, T. Ewald, *Holistic VoIP Intrusion Detection and Prevention System*, 2008.

[12] SIPp at Sourceforge, http://sipp.sourceforge.net/index.html.

[13] Two attacks against VoIP, http://www.securityfocus.com/infocus/1862.

SPAM CONSTRUCTION TRENDS

Barry Irwin[1], Blake Friedman[2]

Rhodes University
Department of Computer Science
South Africa

[1]b.irwin [at] ru.ac.za, [2]blakef [at] rucus.ru.ac.za

ABSTRACT

This paper replicates and extends *Observed Trends in Spam Construction Techniques: A Case Study of Spam Evolution.* A corpus of 169,274 spam email was collected over a period of five years. Each spam email was tested for construction techniques using SpamAssassin's spamicity tests. The results of these tests were collected in a database. Formal definitions of *Pu and Webb's* co-existence, extinction and complex trends were developed and applied to the results within the database. A comparison of the *Spam Evolution Study* and this paper's results took place to determine the relevance of the trends. A geolocation analysis was conducted on the corpus, as an extension, to determine the major geographic sources of the corpus.

KEY WORDS

Spam, Geolocation

SPAM CONSTRUCTION TRENDS

1 INTRODUCTION

Unsolicited commercial email, more commonly known as spam, has placed an increasing burden on global human, computational and bandwidth resources. There is little argument over the proliferation of spam, which has seen significant increases in the quantity and frequency of its distribution into users' inboxes [2, 4]. Current estimates of the scale of the spam problem have identify that up to 80% [10] of all attempts to send email are spam related. The advent of filters which adapt to statistically identifiable components of spam has been met with spammers using increasingly complex construction techniques [1]. Spam has been shown to have a detrimental effect on the end user's perception of the integrity of email and their overall Internet experience [2]. Due to the quantity of spam and the effect this is having, there is a need to improve upon existing anti-spam techniques.

Significant research has been conducted into methods of spam detection, however little attention has been given to the analysis of spam construction trends, particularly the continuity of these techniques [5]. An understanding of whether spam emails' structures significantly vary is a critical factor in dealing with spam. Changes in the structure of spam emails, over a period, can be used to ratify specific anti-spam efforts' effect. This paper extends the framework developed by Pu and Webb [11], hereon referred to as the *Spam Evolution Study*, to further the analysis of spam construction trends.

A large corpus of spam emails was collected and processed through SpamAssassin [9] using a distributed processing architecture. SpamAssassin identifies the components which make up each spam email using a number of rule based *spamicity* tests. A complete history of the relative frequency of each component over the period of a corpus of emails is developed. Each component is then classed using Pu and Webb's original trends: co-existence, extinction and complex. Formalised descriptions of these trends are developed. The results of our trend analysis is then compared to Pu and Webb's results. Further extensions are made by associating each spam email with its geopolitical origin, based on the IP addresses of the sending *mail transfer agent* (MTA).

2 THE CORPUS

Two significant corpora were collected, combined and analysed. The first corpus consisted of a personal MTA's spam collection of 101,170 cataloged spam emails. This corpus was collected between July 2003 and July 2007, using a combination of hand sorting, Bayesian filters, *DNS blacklists* (DNSBL) and SMTP protocol conformity tests to updated the corpus. The second corpus consisted of 68,104 spam emails, collected from January 2006 until August 2007. This corpus represents a user base of approximately 3,000 schools users. This corpus is particularly of interest, as it contains spam which has evaded a far-side MTA performing DNSBL and SMTP protocol conformity tests. A large portion of this corpus consisted of spam containing MIME-encoded viruses, amounting to 2.4Gb of decompressed data.

Emails which originated from local hosts, as well as erroneous files, were removed from the combined corpora of 201,288 emails. The final size of the combined corpora is 169,274 spam emails. As with the *Spam Evolution Study* fluctuations in the quantity of spam are normalised. The normalisation is performed by dividing the spamicity count by the total number of messages per month, determining the relative state of the various spamicity tests each month.

3 DISTRIBUTED PROCESSING ARCHITECTURE

SpamAssassin 3.2.3 [9] formed the most computationally intensive portion of the study. SpamAssassin is the open-source project used in the *Spam Evolution Study*, and was the basis for characterising the various components of a spam email. SpamAssassin uses a number of methods to evaluate the likelihood of an email being spam, these include header and text analysis, DNSBL, statistical and collaborative filtering. These methods are collectively referred to as *spamicity* tests. Initial testing indicated that SpamAssassin was prohibitively computationally intensive when applied to the complete corpus. A distributed processing architecture was developed to decrease the analysis period. The architecture distributes the corpus amongst a number of processing nodes, each running a SpamAssassin instance. Spam emails are then processed in parallel on each node. The results are then submitted by each node to a database for analysis. The details of the implementation and performance of the architecture are further described in [3].

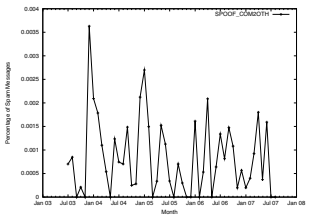


Figure 1: Complex

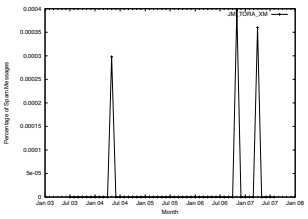


Figure 2: Extinction

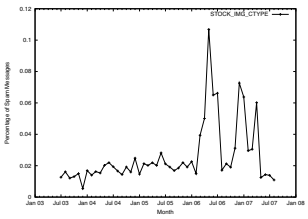


Figure 3: Co-existence

4 SPAM MODEL

Formal definitions of the *Spam Evolution Study's* trends are developed in this section. An iterative process, built on a testing framework, was used to develop the algorithms and extract the trend groups from the corpus. The framework generated a graph of each spamicity test. Each graph depicted the spamicity test's frequency, as a percentage of the total number of email for each month, over the duration of the study. Examples of these graphs can be seen in figures 1, 2 and 3. Further details of this framework are found in [3]. Each graph was categorised based on the trend algorithms, and a comparison to the data would follow to determine the accuracy of the trend.

4.1 Environment Model

To allow for more formal definitions of these algorithms, further definitions of the environment are required. The *months* during which the testing took place occurred between the start month 1 until the final month M, and are defined as $months := \{m \in \mathbb{N} | 1 \leq m \leq M\}$. The total period, P_{tot}, describes the entire testing period, which is defined as $P_{tot} := \bigcup_{i=1}^{M} \#P_i$. The sub-period, which is to say the days within a month, is defined as $P_t := \{n \in \mathbb{N} | n_t, \ldots, n_{t+1}\}$where $t \in months$. During this period, we measure each spamicity test out of a possible set of spamicity tests. We shall refer to a particular spamicity test s where $s \in spamicity$ and $spamicity$ is defined as the set of all spamicity test, $spamicity := \{BAD\_CREDIT, HELO\_OEM, \ldots\}$. Emails are, for the purposes of this analysis, *only seen as subsets of spamicity tests.* A particular email is referred to within the period of the testing, denoted by e_t, where $t \in P_{tot}$, such that $e_t \subset spamicity$. It is also useful to view a particular spamicity test's frequency on a particular month as a percentage of the total number of emails during this month. The values represented in figures 1, 2 and 3 use the frequency function $f(s,t)$,which is defined as $f(s,t) = \frac{\sum_{i \in P_t} \#(e_i \cap \{s\})}{\#P_t}$.

4.2 Complex Trend

The complex trend "combine different trends or contain high variability" [11]. The complex trend's algorithm would have to identify fluctuations between monthly results and mixed candidate spamicity tests.

The complex trend is predominantly identified by fluctuations between each months proportional appearance. The difference between the percentage of email which contains test s in month n and $n+1$ would have to be measured for the duration of the testing. The cumulative value is represented by the function $c(s)$ defined as:

$$c(s) = \sum_{n}^{M-1} |f(s,n) - f(s,n+1)| \tag{1}$$

The set of all complex spamicity tests C for a given spamicity s is then defined as:

$$C(s) = \{s \in \mathit{spamicity} | c(s) \geq \mathit{min\,bound}\} \cup \{\ \mathit{spamicity} \backslash E \backslash X\}$$

Where E is the set of co-existent spamicity tests and X the set of extinct spamicity test, the definitions of which will follow. The value of $\mathit{min\,bound} = 8.4$, which was determined from the ordered-by-magnitude results of $c(s)$ for all elements of *spamicity*. Values above the *min bound* were found to clearly indicate a significantly increased quantity of fluctuation. This process is elaborated in [3].

4.3 Co-Existence Trend

The second trend, "co-existence, [was] indicated by a sustained population of a strain of spam, particularly through the end of the study period" [11]. The "co-existence group consists of curves that remain flat" [11], indicating that there must be little fluctuation in the month-to-month values. The co-existence trend algorithm was required to identify a consistently sustained population, and react to variations from the sustained population, particularly towards the end of the study period.

In considering co-existence, it was found that grouping certain ranges and assigning a collective value was reasonable. Spamicity tests which were found in $(0\% \ldots 80\%]$ of the emails in a given month were considered viable co-existent candidates. A particular spamicity test's appearance in 80% and above emails for a month was considered a fluctuation, and carried a lesser

weighting. Spamicity tests which were not found in a month were negatively weighted, particularly if this occurred in the final month of testing. A failure to appear in the final month resulted in the exclusion of a spamicity test from the co-existent group. The grouping is represented in the bucket function $b(s,t)$ with s being a spamicity test, where $s \in spamicity$, and t is a month in the testing period, where $t \in P_{tot}$. The bucket function is defined as:

$$b(s,t) = \begin{cases} 1 & \text{if } f(s,t) > 0.8, \\ 10 & \text{if } 0.1 < f(s,t) \leq 0.8, \\ 5 & \text{if } 0 < f(s,t) \leq 0.1, \\ -10 & \text{if } f(s,t) = 0, \\ -1000 & \text{if } f(s,t) = 0 \text{ and } t = M \end{cases}$$

The bucket function is then applied to the entire range of the corpus, and each months value is adjusted to give greater weighting to the latter range of the corpus. The co-existence function $e(s)$ for a particular spamicity test is defined as:

$$e(s) = \sum_{n}^{M} \frac{b(s,n)}{(M-n+1)^2} \tag{2}$$

The set of all co-existent spamicity tests, E, is defined as:

$$E \quad = \quad \{s \in spamicity | e(s) > accept\,bound, c(s) \leq min\,bound\}$$

This set excludes all spamicity tests which display a high degree of fluctuation, and are considered complex. The *accept bound* responds to the bucket function, where $accept\,bound = 0$.

4.4 Extinction Trend

The final trend is "extinction, indicated by the population of a strain of spam declining to zero or near zero during the study period" [11]. Extinction presented significant problems in attempts to define a reasonable algorithm, and because of this it is based off the two existing algorithms. The definition requires that extinct spamicity tests identify a consistently sustained population and have no monthly population or decline to a near-zero population.

	Main Corpus		Spam Evolution Study		Relative Difference
Trend	#	%	#	%	%
Co-existent	197	31	64	13	18
Extinction	316	51	236	48	3
Complex	111	18	195	39	21

Table 1: Comparison of the distribution of the spamicity tests amongst the trends.

As has already been shown in section 4.2, a value greater that *min bound* for the $c(s)$ function indicated a high degree of fluctuation in the monthly spamicity test results. Values less than or equal to *accept bound* for the $e(s)$ function indicate a spamicity test which has significantly declined for periods, or is consistently absent. The set of all extinct spamicity test is defined as:

$$X = \{s \in spamicity | e(s) \leq accept\,bound \text{ and } c(s) \leq min\,bound\}$$

5 SPAM EVOLUTION ANALYSIS

A comparison between the *Spam Evolution Study*'s distribution and the distribution of the corpus is shown in table 1. The corpus has approximately 82% of the tests falling under the co-existent and extinct trends. The *Spam Evolution Study* has approximately 61% of the spamicity tests falling under similar trends. The two corpora do not reflect a similar distribution of the spamicity tests outside of the complex trend. The differences between the two corpora's co-existent and extinct trends shows that over a longer period extinction is more prominent than co-existence.

The maximum range for each spamicity test and the average range indicates a correlation between the extinction and complex trends of both corpora. The majority of these two trends are found in the $[0.0 \ldots 0.1)$ range. This is to say that the majority of these tests, which identify the corpus' emails as spam, are dispersed over less than 10% of the corpus emails on average or at a maximum each month. The corpus' co-existence trends, in particular, show a significantly higher proportion located in this low range. This is not in keeping with the *Spam Evolution Study's* co-existence trend, which is dispersed amongst the higher ranges of both the maximum and average spamicity test results.

Assuming that SpamAssassin is able to consistently identify the components of a spam email using its spamicity tests, the locality of the majority of

spamicity tests in range could be caused by two conditions. Firstly the types of spam captured are from a large number of spammers, or secondly spammers employ a diverse number of techniques, or both. In either instance the average and maximum distribution suggests a large and varied number of spamicity tests per an email in the corpus. This is reciprocated by further analysis which shows that an average of 8.96 spamicity tests are found for every email in the corpus.

One hypothesis for the dominance of the extinction spamicity tests is a natural extension of the evolutionary metaphor used by *Pu and Webb*. All spamicity tests inevitably tend towards extinction, while some may co-exist for longer periods: their existence relies on their evolving beyond the means of their respective spamicity tests. This evolution implies that the older spamicity test must adjust to these variations, resulting in their older form's extinction. We see a reflection of this behavior in the difference between spamicity test from one version of SpamAssassin and another. The above findings indicate that the trends specified in the *Spam Evolution Study* are relevant to the corpus. There are issues which mitigate these findings in a direct comparison to the *Spam Evolution Study*, which will be discussed. It does, however, hold that the process used by the *Spam Evolution Study* still has relevance in analysing the corpus.

6 DIFFERENCES TO THE SPAM EVOLUTION STUDY

The structure of the corpus was significantly varied from the *Spam Evolution Study's* corpus in two respects: quantity and period. The corpus has an average of 3,385 spam email for each month, while the *Spam Evolution Study* has 38,889 spam emails for each month. 634 Spamicity tests were applied to the corpus, while 495 spamicity tests were applied to the *Spam Evolution Study'*s corpus. If we assume the average of 8.96 spamicity tests per an email applies to both corpora, this would result in the *Spam Evolution Study* being significantly more viable and representative of spam in the wild.

The limited number of sources which make up the corpus, could have unfairly weighted certain tests, favoring specific trends. The *Spam Evolution Study's* use of the SpamArchive project allowed for a significantly more diverse series of sources. The diversity of sources increases the probability of *Pu and Webb's* results reflecting the state of spam in the wild.

The version of SpamAssassin utilised, further reduces the comparative value of this study. The *Spam Evolution Study* does not specify the exact spamicity tests it utilised, however a brief comparison between the spamicity tests of

SpamAssassin 3.1.x and 3.2.x shows significant differences. SpamAssassin 3.1.x contains 795 test and 3.2.x contains 746 tests. Only 383 of the original tests are found in the newer version, which was utilised in this paper.

The specific algorithms utilised by *Pu and Webb* to differentiate between the spamicity trends were not published. Accordingly this paper developed its own algorithms; this is the most significant variation from the *Spam Evolution Study*.

The comparison of this study and the *Spam Evolution Study* is severely limited by the above variations. More specifically the structure of the corpus, the version of SpamAssassin and the trend algorithms introduce a number of limitations to any direct comparison of this paper's results.

7 GEOLOCATION

Geographic location, or geolocation, is the mapping of an IP addresses to a series of geographic co-ordinates. The mapping of an IP address to a country was considered an adequate degree of granularity.

The IP addresses stored in a spam email must be considered unreliable. An RFC 2822 [8] email header should contain a number of *received* fields, in which the IP address of a connecting MTA is stored. Spammers abuse the standard, and often include a number of forged received fields to exploit anti-spam filters. For this reason only the IP addresses associated with connections to reliable MTAs can be trusted. A reliable MTA is defined as the border MTA, which updates an email's header with the first verifiable received field. The border MTA for both corpora was easily determined, although the structure of the anti-spam solutions was such that the connecting IP addresses of non-routable, local and far-side MTAs had to be removed. For example the far-side MTA, which is located in the United States was *one* of the border MTAs for the schools corpus. The border MTA was followed by a number of internal MTAs which append additional received fields. These fields had to be removed from consideration, with only the IP address recorded by border MTA being used for geolocation.

Once an authentic IP address had been obtained, its geolocation had to be determined. The open-source HostIP [6] database was used as a reference, and a customised local implementation was configured to map IP addresses directly to countries. The appropriately selected IP address is then mapped to a country. The email's geographic location is then updated in the database for representation and analysis.

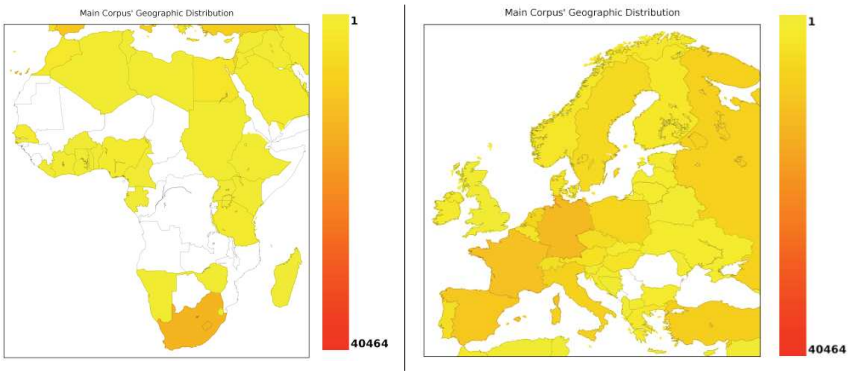


Figure 4: The distribution of the corpus over the African, Europe and the world.

Geolocation data was represented using map projections. A Miller cylindrical map projection was used to graphically display quantitative data. An example of this is the distribution of the corpus' sources of spam from Africa, seen in figure 4.

7.1 Results

The locality of the corpus, with the visualisation of the quantity of spam detected in specific regions, is an excellent tool for the analysis of a spam corpus. The geographic origin of an email was a factor which the original *Spam Evolution Study* was unable to explore due to corpus' structure. This paper uses African and European projections as examples. The top five contributive countries, accounting for the majority of spam in the corpus, are listed in table 2, and should be compared to the projection in figure 4.

Country	# Spam Emails	% of Corpus	Cumulative %
United States	40464	23.904%	23.904%
Taiwan	22359	13.209%	37.113%
United Kindom	17066	10.082%	47.195%
Korea, Republic of	15557	9.190%	56.385%
China	12429	7.343%	63.728%

Table 2: The top five spamming countries in the corpus.

A projection of Africa is shown in figure 4. It is clear that both South African and Egypt are the primary sources of spam in the continent. Most surprising is the lack of spam from central and western Africa, which is the largest

continuously populated region in the corpus to be spam-free. Continental Europe is widely dispersed, and was a significant contributor to the corpus. With the exception of Montenegro, Serbia and Romania every country in Europe contributed.

8 CONCLUSIONS

This paper replicated the *Spam Evolution Study*, and presented map projections of the collected spam corpus. A corpus of 169,274 spam emails was collected. The corpus was analysed using SpamAssassin and a distributed processing network. The results were further evaluated by dividing each tests into the three trends of co-existence, extinction and complex. These trends were formalised as an extension to the original study. The trends were found to be applicable to the corpus, however a number of variations from the *Spam Evolution Study* reduced the comparative value of these findings. Geographic projections were created, using data collected from the corpus and findings detailed.

9 FUTURE WORK

The corpus is limited to emails which have been distributed to South African MTAs. This underutilised the distributed architecture which was specifically designed to handle a significantly larger corpus. One of the early limitations of this study is the relatively small scale of the corpus when compared to other studies [7, 11, 12]. Future research into the effects of geolocation on the evolution of spam construction would be benefited by applying this study on a substantially larger and wider ranging corpus. A closer analysis of particular provinces and states within countries could be performed.

The linking of the developed state of a country to the quantity of spam it produces would be a particularly challenging and interesting extension. An extension of this study could be conducted on further research into selecting the grouping of the various geographic locations of identified spam. One interesting possibility would be the use of spam construction techniques to probabilistically determine the identity and locations of botnets.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support of our colleagues in the Department of Computer Science. We would also like to acknowledge the support of Telkom SA, Business Connexion, Comverse

SA, Stortech, Tellabs, Amatole, Mars Technologies, openVOICE and THRIP through the Telkom Centre of Excellence in the Department of Computer Science at Rhodes University.

REFERENCES

[1] James Carpinter and Ray Hunt. Tightening the net: A review of current and next generation spam filtering tools. *Computers & Security*, 25(8):566–578, November 2006.

[2] Deborah Fallows. Spam: How it is hurting email and degrading life on the internet. Technical report, PEW Internet & American Life Project, October 2003.

[3] Blake Friedman. A Formalised Replication and Extension of Observed Trends in Spam Construction Techniques: A Case Study of Spam Evolution. Honours Dissertation, Rhodes University, November 2007.

[4] Tom Gillis. Internet security trends for 2007: A report on spam, viruses and spyware. Technical report, IronPort, 2007.

[5] Joshua Goodman, Gordon V. Cormack, and David Heckerman. Spam and the ongoing batle for the inbox. *Communications of the ACM*, 50(2):25–33, February 2007.

[6] Simon Gornal. Hostip.info. http://www.hostip.info/, 2007.

[7] Geoff Hulten, Anthony Penta, Gopalakrishnan Seshadrinathan, and Manav Mishra. Trends in spam products and methods. In *CEAS 2004 - First Conference on Email and Anti-Spam*, 2004.

[8] J. Klensin. Simple Mail Transfer Protocol (SMTP). RFC 2821, April 2001.

[9] Justin Mason. SpamAssassin. http://spamassassin.apache.org/, 2007.

[10] Messaging Anti-Abuse Working Group. Email metrics program: The network operators' perspective. 3rd and 4th Quarters 4, MAAWG, March 2006.

[11] C. Pu and S. Webb. Observed trends in spam construction techniques: A case study of spam evolution. In *CEAS 2006 - Third Conference on Email and Anti-Spam*, July 2006.

[12] B. Taylor. Sender reputation in a large webmail service. In *CEAS 2006 - Third Conference on Email and Anti-Spam*, July 2006.

APPLICATION OF MESSAGE DIGESTS FOR THE VERIFICATION OF LOGICAL FORENSIC DATA

Pontjho M. Mokhonoana[1], Martin S. Olivier[2]

[1]Information and Computer Security Architectures Research Group
Department of Computer Science
South Africa

[1]pontjho@tuks.co.za, [2]martin@mo.co.za

ABSTRACT

A message digest is a fixed length output produced by applying a cryptographic algorithm on input binary data of arbitrary length. If the input data changes even by one bit, the generated message digest will be completely different from the original. This is used in digital investigations to verify that stored digital evidence has not been tampered with.
This technique has been applied successfully on physical disk images because there is only one continuous stream of data. However, this is not applicable to logical disk images where there is no obvious or standard method of concatenating the data to produce an output message digest. This paper describes the difficulties that complicate the computation of a message digest for logical data. In addition, a candidate process for calculating a verification value for computer forensic evidence for logical data, regardless of its underlying representation is given. This method is presented in the context of cellphone forensics.

KEY WORDS

Computer Forensics, Verification, Cryptographic Hash, Message Digest

APPLICATION OF MESSAGE DIGESTS FOR THE VERIFICATION OF LOGICAL FORENSIC DATA

1 INTRODUCTION

Current best practices for dealing with digital evidence advocate that when evidence is acquired, the first to the last bit of the data on the device be copied to create a physical disk image [Jansen and Ayers, 2006]. This has the advantage of allowing the recovery of deleted or partially overwritten data. The physical image also facilitates the simplicity of the process of the computation of the associated message digest of that image since the one way hash is based on a single stream of binary data.

It is however not always possible to obtain a physical image of a device's data; this is usually the case when extracting data from Small Scale Digital Devices [Harrill and Mislan, 2007] such as cellular phones or a Subscriber Identity Module (SIM) card. In other cases, it is preferable to obtain a logical image when only the logical data is required. Creating a physical image makes very little sense since that would add unnecessary additional processing and storage overheads.

When looking at logical data, the picture gets a little more complicated since a simple reordering of the logical data items will affect the produced hash. This complexity of verifying logical data makes it difficult to verify data retrieved from mobile devices using most of the acquisition methods [Mokhonoana and Olivier, 2007], since they produce a logical image.

For example, a phone book entry will consist of the name of the person, their phone number and usually other contact details such as the email or secondary number. There is no defined order in which such items should be read meaning that if a hash value were to be computed on that data, the output produced by different tools on the same content may be different depending on the order in which the data was read.

To give an idea of how the logical image from a smartphone would look like, figure 1 is given below. Each of the contact entries could in turn have the name, surname, office number, mobile number, email as well as other attributes.

One way to get around this is to compute the hash for each data item in the logical image. Such a solution has a number of drawbacks:

- It would complicate the process of verifying the authenticity of the log-

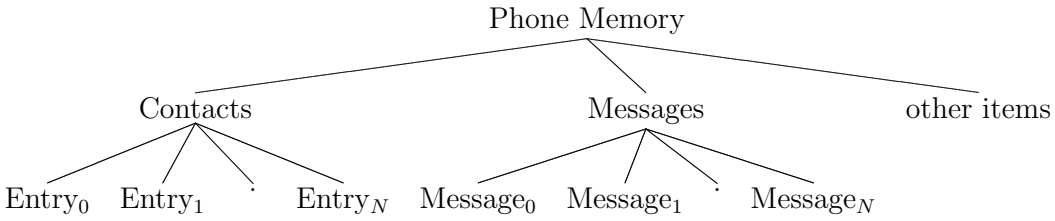


Figure 1: Sample Data From a Mobile Phone

ical image since it would increase the algorithmic complexity of comparing the hashes.

- If there is a large number of data items in the logical image, the hashes could take up a large amount of disk space.
- It does not take into account the attributes of the logical data items. If it does, then there is the question of the ordering and representation of the attributes.

To solve the above mentioned problems, the Sorted Vector Hashing (SVH) is porposed in this paper. The goal of the algorithm is to enable the computation of a hash value based on the logical content of a file rather than just its binary data. This will facilitate the comparison of the content of different data items such as an email, even though they are stored in different mail formats.

The aim of this paper is to present a method for producing a message digest for a logical image file that also takes into account the attributes of the data items and the potentially different representations of the attributes. The method achieves the goal by allowing the message digest computation to be performed on items at any granular level and putting it together with a simple algorithm which will be discussed in more detail later in the paper.

2 RELATED WORK

Most of the computer forensics tools employ the use of hashes to verify their images. However for logical data, different tools use different schemes to do the verification. *EnCase* [Guidance Software, 2008a] uses a proprietary L01 format for the storage of logical data. For that reason, it is difficult to determine how its verification works.

The *Forensic Toolkit* [Guidance Software, 2008b], which uses FTK Imager to create images employs the use of AD1 image containers to store logical data. The format does not have built-in mechanisms for the verification. Instead, another file (called a hash list) with the hashes of all the individual files is created. The hashes are only based on the filestream content and not their associated attributes.

TULP 2G [van den Bos and van der Knijff, 2005] is an open source tool for acquiring and decoding data from electronic devices. It uses an XML format for the storage of case related data. A sample of the data is given in figure 2. To calculate the hash of the case items or the entire case, the formulas in figure 3 are used.

In these formulas, the hash of an item is computed by concatenating its metadata with the string content. The hash is then performed on that resultant output. A similar process is used to compute the hash of the investigation and that of the case.

```
<Case Name="SampleCase" Creator="…" DateCreated="…" DateModified="…" MD5="…" SHA1="…">
    <Notes>…</Notes>
    <Item Name="SampleItemA" DateCreated="…" DataType="…" StorageType="…" ItemType="…" MD5="…" SHA1="…">
        …
    </Item>
    <Investigation Name="SampleInvestigation1" Creator="SampleAuthor" DateCreated="…" MD5="…" SHA1="…">
        <Notes>…</Notes>
        <Item Name="SampleItem1" DateCreated="…" DataType="…" StorageType="…" ItemType="…" MD5="…" SHA1="…">
            …
        </Item>
        <Item Name="SampleItem2" DateCreated="…" DataType="…" StorageType="…" ItemType="…" MD5="…" SHA1="…">
            …
        </Item>
        …
    </Investigation>
    <Investigation Name="SampleInvestigation2" Creator="…" DateCreated="…" MD5="…" SHA1="…">
        …
    </Investigation>
    <Item Name="SampleItem3" DateCreated="…" DataType="…" StorageType="…" ItemType="…" MD5="…" SHA1="…">
        …
    </Item>
    …
</Case>
```

Figure 2: Tulp 2G Data Format

The problem with the L01 format is the lack of documentation around how it works. That makes it difficult for other tools to implement and verify the results produced by a tool. FTK's hash list is quite simple to verify in that it is a plaintext file which could be read and verified by a human. The first problem is that it only considers the file data. If the file's metadata, such as creation or modification date, is altered, that change will not be detected. The second problem is that because the hash of every single file has to be

$$Hash(Item) = Hash(Name + DateCreated + DataType + StorageType + ItemType + StringContent)$$

$$Hash(Investigation) = Hash\left(Name + Creator + DateCreated + Notes + \sum_{i=1}^{\#Items} Hash(Item[i])\right)$$

$$Hash(Case) = Hash(Name + Creator + DateCreated + DateModified + Notes + \sum_{i=1}^{\#ItemChildenOfCase} Hash(Item[i]) + \sum_{j=1}^{\#Investigations} Hash(Investigation[j]))$$

Figure 3: Calculation of Hash

specified, the hash list gets large very quickly, making it difficult for human verification. The Tulp2G XML format is an improvement on the above in that it is open and takes all attributes into account when calculating the hash. However, it is designed for a specific application and is not flexible enough to handle other applications. If other attributes were introduced, it would not be able to handle them without changing the algorithm used to calculate the hash. In addition, it does not address the ordering problem discussed above.

3 SORTED VECTOR HASHING

A logical image will consist of the following: Item, Attributes and Children. The item is a logical entry in the image. For example on a logical filesystem image, the item could be a file. The attribute will be the filename, date or owner. The children could be other files (if this file is a folder) or data streams in the file. Most file systems allow a single child in the latter vase (but note that NTFS can have multiple streams).

The way that a hash is calculated depends on what you are trying to verify. For example when trying to calculate the hash of a folder, the order of the children is not important. Actually, from a logical perspective, there is no universally defined way of sorting the contents of a folder. That is, the following should hold for a folder f, with files $f1$, $f2$ and $f3$: $h(f) = h(f1, f2, f3) = h(f2, f1, f3) = ... = h(f3, f2, f1)$. In this text, this is referred to as commutativity. However when calculating the hash of the attributes, the order is important. For example a file will have a creation, modification and access date. Now even though there also is no obvious order in which the hash must be applied to the attributes, it must be applied

uniformly to yield the same output. That is: if $a = \{a1, a2, a3\}$ is the set of attributes then $h(a) = h(a1, a2, a3)$, but

$$h(a1, a2, a3) \neq h(a2, a1, a3) \ldots \neq h(a3, a2, a1)$$

The process for performing the hash must be able to cater for the two scenarios above, one where the order is not important and the other where it is. To achieve this, the Sorted Vector Hashing (SVH) algorithm was conceived and its aim is to generate a single hash from a number of logical items. The algorithm assumes that the sort order is not important for children but is for attributes.

To compute the hash on the data represented in the tree on figure 1, one would start at the terminal nodes in the tree and work one's way up to the root. Each node in the tree has zero or more attributes and children. Stated formally:

$$I = (\langle a_1, a_2, \ldots, a_m \rangle \langle c_1, c_2, \ldots, c_n \rangle) \tag{1}$$

I represents an arbitrary node in the tree. a_i and c_i represent an arbitrary attribute and child of I respectively. Note that the inner text content of a node is treated as an attribute.

The ordering of the attributes must be preserved. For that reason computing the hash(*H(x)*) of the attributes is straightforward since it only involves concatenating the hashes of the respective attributes and then calculating the hash over the resultant output. This is referred to as Unsorted Vector(UV). Therefore the hash of the attributes is:

$$H(I_a) = UV(I_a) = H(\biguplus_{i=1}^{m} H(a_i)) \tag{2}$$

where $\biguplus$ is used to denote concatenation.

The ordering of the children however is not important and as a result the oder in which the happen to be must not affect the output hash. To achieve this, the child hashes are computed and then the hashes are sorted. It is only after that that the resultant output is concatenated and hashed. To state this formally, assume $H(c_i) = y_i$. Then, construct a sorted vector, $\langle z_1, \ldots, z_n \rangle$, such that

$$\forall y_i \exists z_j \text{ where } z_j = y_i \tag{3}$$

and

$$(\forall z_i, i < n) z_i \leq z_{i+1} \tag{4}$$

$$\text{Then } H(I_c) = SV(I_c) = H(\biguplus_{j=1}^{n} z_i) \tag{5}$$

To get the hash of I, the hash of its attributes is concatenated with that of the children and then rehashed. That is:

$$H(I) = UV(H(I_a) \biguplus H(I_c)) \tag{6}$$

Note that equation 6 is applied recursively to the nodes in the tree starting from the root. To compute the hash of the child, the equation is applied where the child would then be I.

4 EVALUATION AND FUTURE WORK

To evaluate the algorithm, it was tested against a number of sample XML documents where a node represents an item. The XML attributes represent the item's attributes and the child elements the item's children. If the algorithm works, it must fulfil the following requirements:

1. The names of the tags should not affect the output
2. The ordering of the attribute tags should affect the output
3. The ordering of the child elements should not affect the output
4. The values of the child or attribute nodes should affect the output

A number of these sample XML documents were used to test if the algorithm behaved as it should. These are given in figures 4, 5, 6 and 7. The hash values for the documents are summarised in the tables 1 and 2. In the following paragraphs, the results from applying the algorithm are discussed. It is also described how the requirements stated above are satisfied by the algorithm.

The algorithm described in the previous section does not take into account the names when computing the output. Therefore the first requirement is obviously satisfied.

```
<?xml version="1.0"?>
<image date="16/03/2008" examiner="Jack Sparrow">
    <files>
        <file path="C:/multimedia/file1.jpg" create_date="7 Jan 2006" modified_date="8 Jan 2007">
            abc...xyz
        </file>
        <file path="C:/multimedia/file2.jpg" create_date="10 Jan 2008" modified_date="11 Jan 2008">
            zyx...cba
        </file>
    </files>
    <phonebook>
        <contact name="John" lastname="Doe" email="" fax="011 111 1123" />
        <contact name="Mary" lastname="Doe" email="mary@doe.com" fax="011 112 1123" />
    </phonebook>
</image>
```

Figure 4: Sample 1

```
<?xml version="1.0"?>
<image date="16/03/2008" examiner="Jack Sparrow">
    <files>
        <file filepath="C:/multimedia/file2.jpg" create_date="10 Jan 2008" modified_date="11 Jan 2008">
            zyx...cba
        </file>
        <file filepath="C:/multimedia/file1.jpg" create_date="7 Jan 2006" modified_date="8 Jan 2007">
            abc...xyz
        </file>
    </files>
    <phonebook>
        <entry firstname="Mary" lastname="Doe" email="mary@doe.com" fax="011 112 1123" />
        <entry firstname="John" lastname="Doe" email="" fax="011 111 1123" />
    </phonebook>
</image>
```

Figure 5: Sample 2

In sample 4, the order of the attributes is changed. Since in this paper the names of the tags are not considered, the ordering of the attributes is important in order to preserve the semantic meaning of the attributes. Therefore, sample 4 would be considered different from 1. Since the hash is different, it satisfies the second requirement.

Sample 1 and sample 2 are supposed to represent the same logical data. The only differences are the ordering of the children of the contact and file node as well as the names of some of the tags. The reader's attention is drawn to the fact that their hashes are exactly the same. This is consistent with the third requirement that the ordering of the child elements should not affect the output.

In sample 3 the value of the contact node is changed. The name, lastname and email attribute values are changed to *"Jack", "Johnson"* and *""* respectively. For that reason, the hash produced is different, thus satisfying requirement 4.

```
<?xml version="1.0"?>
<image date="16/03/2008" examiner="Jack Sparrow">
    <files>
        <file path="C:/multimedia/file1.jpg" create_date="7 Jan 2006" modified_date="8 Jan 2007">
            abc...xyz
        </file>
        <file path="C:/multimedia/file2.jpg" create_date="10 Jan 2008" modified_date="11 Jan 2008">
            zyx...cba
        </file>
    </files>
    <phonebook>
        <contact name="Jack" lastname="Johnson" email="" fax="011 111 1123" />
        <contact name="Mary" lastname="Doe" email="mary@doe.com" fax="011 112 1123" />
    </phonebook>
</image>
```

Figure 6: Sample 3

```
<?xml version="1.0"?>
<image date="16/03/2008" examiner="Jack Sparrow">
    <files>
        <file path="C:/multimedia/file1.jpg" create_date="7 Jan 2006" modified_date="8 Jan 2007">
            abc...xyz
        </file>
        <file path="C:/multimedia/file2.jpg" create_date="10 Jan 2008" modified_date="11 Jan 2008">
            zyx...cba
        </file>
    </files>
    <phonebook>
        <contact name="Jack" email="" fax="011 111 1123" lastname="Johnson" />
        <contact name="Mary" email="mary@doe.com" fax="011 112 1123" lastname="Doe" />
    </phonebook>
</image>
```

Figure 7: Sample 4

From the analysis of the hashes, one can conclude that the algorithm works as expected — two logical items must produce the same hash if and only if their content is the same.

5 CONCLUSION

In this paper, the challenges of calculating a hash for logical items are discussed. The current tools which handle logical data are analysed and their limitations discussed. The biggest contribution of this paper is the introduction of a method for verifying logical data that overcomes some of the limitations of the other tools. The method was tested on a small dataset and the results confirm that it works.

The work presented is an important aspect of the research into mobile forensics since it enables reliable verification and comparison of logical data from potentially different sources. It is hoped that this will contribute to a

Sample	*MD*5
1	85 19 86 F2 24 B2 70 EA F4 3B F8 93 F7 E6 79 71
2	85 19 86 F2 24 B2 70 EA F4 3B F8 93 F7 E6 79 71
3	78 77 A6 E5 FF C6 33 B4 A8 F7 82 AC 73 D9 9F EE
4	96 E1 6F FD 78 05 E8 5B F8 5A AD 05 90 63 81 BB

Table 1: MD5 Hash Summary

Sample	*SHA*1
1	17 04 13 47 FA 0E 1A A4 C0 CC A4 B6 0A 50 90 F9 4F 35 1C D3
2	17 04 13 47 FA 0E 1A A4 C0 CC A4 B6 0A 50 90 F9 4F 35 1C D3
3	7B 5F B0 8F 83 EC 07 61 68 6D DC 6E B9 BD 09 8A 05 11 E9 A9
4	0A 89 2D 95 0B D7 D5 51 5C 8F 2F B5 5B F9 89 50 06 24 7B FE

Table 2: SHA1 Hash Summary

standardisation of verification methods in computer forensics.

References

[Guidance Software, 2008a] Guidance Software (2008a). Encase. www.guidancesoftware.com.

[Guidance Software, 2008b] Guidance Software (2008b). Forensic toolkit. www.accessdata.com.

[Harrill and Mislan, 2007] Harrill, D. C. and Mislan, R. P. (2007). A small scale digital device forensics ontology. *Small Scale Digital Forensics Journal*, 1(1):1–7.

[Jansen and Ayers, 2006] Jansen, W. and Ayers, R. (2006). Guidelines on cell phone forensics. Technical report, National Institution of Standards and Technology.

[Mokhonoana and Olivier, 2007] Mokhonoana, P. and Olivier, M. S. (2007). Acquisition of a symbian smart phones content with an on-phone forensic tool. In *Proceedings of the Southern African Telecommunication Networks and Applications Conference 2007 (SATNAC 2007)*, Mauritius.

[van den Bos and van der Knijff, 2005] van den Bos, J. and van der Knijff, R. (2005). TULP2G – an open source forensic software framework for acquiring and decoding data stored in electronic devices. *International Journal of Digital Evidence*, 4(2).

A PROOF-OF-CONCEPT IMPLEMENTATION OF EAP-TLS WITH TPM SUPPORT

Carolin Latze and Ulrich Ultes-Nitsche

University of Fribourg

{carolin.latze | uun}@unifr.ch
Boulevard de Pérolles 90
1700 Fribourg
Switzerland

ABSTRACT

Many people who have tried to configure their IEEE 802.11 enabled mobile phones to connect to a public wireless hotspot know one of the major differences between IEEE 802.11 networks and 2G: the missing standardized login process. While the 2G standard covers all aspects of the communication process, first IEEE 802.11 standards only targeted the data transmission. Due to this lack of standards for authentication, the login process and the missing secure subscriber identification, a number of different, mostly incompatible, login procedures have been established that are all far away from being as usable, comfortable and secure as 2G methods. This is why the authors of this paper propose to use EAP-TLS, which is a well established, secure and scalable authentication protocol, in combination with identities provided by a Trusted Platform Module (TPM) in order to archieve a high comfort for the user

This paper describes the concept, presents a Linux based implementation, and evaluates the approach in a testbed.

KEY WORDS

Security, Authentication Protocols, TPM, EAP-TLS

A PROOF-OF-CONCEPT IMPLEMENTATION OF EAP-TLS WITH TPM SUPPORT

1 INTRODUCTION

From the user's point of view, GSM networks are one of the simplest and most comfortable networks that exist. There is no need to configure anything, just buy a phone and Subscriber Identity Module (SIM) and use it. That is completely different for 802.11 networks as they do not provide a standardized authentication protocol. When GSM was released, it included an identity management and based on those identities one authentication protocol used in every GSM-enabled device. By contrast, at the time 802.11 was released, the standard concentrated on data transmission and not on identity management and authentication. All the authentication protocols that exist today came later and - as shown in [1] – are not comparable to GSM authentication. With the emergence of Trusted Platform Modules (TPMs), for the first time there exists a kind of integrated hardware token and identity provider in the world of computer networks comparable to the SIM card in GSM networks. In 2007, the authors of this paper proposed to use these TPMs with EAP-TLS to build a secure scalable and user-friendly authentication protocol for 802.11 networks, which is as comfortable as the GSM authentication protocol [1]. This paper presents a concrete realization of the proposed protocol, providing a proof-of-concept prototype of the protocol showing how easy it is to modify existing EAP-TLS implementations to use this new approach. Furthermore, the work shows, that this protocol may be used on every kind of 802.11 enabled device.

This paper is structured as follows: an introduction into EAP-TLS with TPM support is given, followed by a detailed description of the proof-of-concept prototype. The paper concludes with improvements for the presented prototype and conclusions.

2 EAP-TLS WITH TPM SUPPORT

In 2007, the authors of this work proposed to extend EAP-TLS with TPM support to implement a user-friendly, secure and scalable authentication protocol for 802.11 networks [1].

EAP-TLS refers to the integration of the Transport Layer Security (TLS) protocol within the Extensible Authentication Protocol (EAP). It is one of the most secure authentication protocols for 802.11 networks when using the *mutual* authentication mode. But using mutual authentication in EAP-TLS means that the client needs its own certificate, which is too complicated for naïve users. In 2002, the Trusted Computing Group (TCG) released the main specification of the new Trusted Platform Module (TPM) [2], which states that the TPM will come with a predefined certificate infrastructure explained in the next section. This TPM certification process relies on the fact, that TPMs may be uniquely identified. As such a certificate belongs to the TPM and not to the real person, it is possible to automate the certificate retrieval, which makes it much easier to use for naïve users. Because of this possibility to facilitate the certificate retrieval without compromising its security, the authors of this paper propose to use the TPM certificates in conjunction with EAP-TLS to implement a highly secure, scalable and user-friendly authentication protocol.

3 PROOF-OF-CONCEPT IMPLEMENTATION

The following sections show a proof-of-concept implementation of the authentication protocol presented above. It starts with an overview over the architecture and goes on with a detailed description of the components needed in this architecture.

3.1 Architecture Overview

In general, the architecture consists of three components: client, authentication server and Privacy CA as shown in *Figure 1*. The client is the one, which wants to be authenticated, the authentication server authenticates the client and the Privacy CA issues the certificates and provides a verification service. Such a Privacy CA will most likely be deployed by every

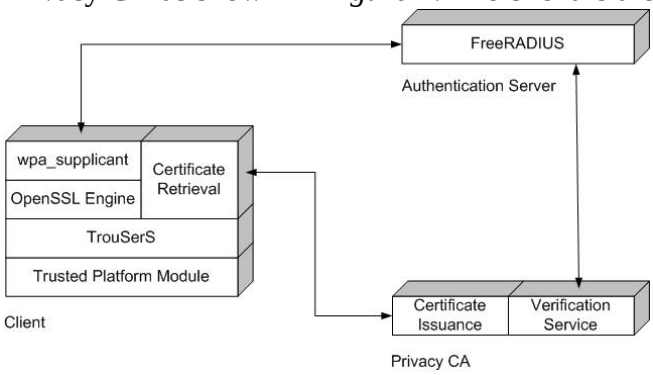


Figure 1 Architecture Overview

operator the enables EAP-TLS authentication using TPMs at its public wireless hotspots.

The client has to request a new identity before connecting to the EAP-TLS secured network. Afterwards, he may start an EAP-TLS authentication with the authentication server. During this EAP-TLS authentication process, the server has to verify the client's certificate. As TPM certificates are slightly different to X.509 certificates, there has to be a verification service, which does the job for the EAP-TLS authenticator.

3.1.1 The TPM

The Trusted Platform Module (TPM) as specified by the Trusted Computing Group (TCG) [3] is a module, which amongst others, provides cryptographic functions and secure storage of keys and signatures. A very important feature of the TPM is that this module can be uniquely identified. It is equipped with a so called Endorsement Key Pair, which is unique. The private part of this key pair is never released from the TPM. The possibility to manage several identities (many different X.509 identity certificates) by the TPM makes the module very useful for different applications. The user might for instance use one identity for her e-banking account and the other one for an authentication as proposed in this paper. Using different identities for different purposes makes the user untraceable. Such a TPM identity must be signed by a certification authority (CA), which means that the CA is the only authority except the TPM's owner that is able to map the identities to a genuine TPM. To certify such a TPM identity, the CA has to check several certificates provided by the TPM manufacturer. That is why a special CA, called Privacy CA is needed to issue TPM identity certificates.

3.1.2 An Open Source TCG Software Stack: TrouSerS

The Trusted Platform Module (TPM) as described in [2] has to provide several cryptographic methods and protected storage. But it also has to be cheap to build to make it a ubiquitary device. That is why the TCG decided to distinguish between methods, that have to run in a protected environment and those that may run in a software-only environment called TCG Software Stack (TSS).

There are already some implementations of the TCG Software Stack, for instance the closed source NTRU TSS [4], an open source Java TSS

implementation called jTSS [5] and an open source IBM implementation called TrouSerS [6]. Given by the applications that had to be modified for the prototype, there was the constraint to use C as programming language and since TrouSerS is the only open source C implementation of the TSS, the authors decided to use it in their prototype implementation.

TrouSerS comes with a persistent storage file to store certain uncritical information on the hard disc in order to save memory on the TPM. This file contains for instance information about whether a key needs authentication or not (but it does not contain the authentication secret!) and may contain the public portion of keys, whose private key resides inside the TPM. To access those keys in the persistent storage, so called `UUID`s are used. There are certain predefined `UUID`s, for instance `{0, 0, 0, 0, 0, {0, 0, 0, 0, 0, 1}}` for the Storage Root Key (SRK), which is the parent for all keys stored in a given TPM.

3.1.3 The Privacy CA and TPM Certificates

As stated above, the TPM may manage different identities. This section describes the identity retrieval as specified in [2].

After having created a new RSA key that will serve as identity key, the TCG Software Stack (TSS) has to collect all information needed to request a new identity using `Tspi_TPM_CollateIdentityRequest`. This information includes: The endorsement credential, which identifies the TPM uniquely, the conformance credential, which states that the TPM is genuine, the platform credential, which states that the platform is genuine and the public portion of the newly generated identity key.

At the time of writing this paper, there exist only two implementations of such a Privacy CA: One in Java, implemented by the University of Graz [7] and freely available to install anywhere and another one available online [8]. The Java implementation uses the XML Key Management Protocol XKMS [9] for the communication between client and server, whereas the online PCA uses a REST-style AP [10]. The authors decided to use the latter one as this online PCA maps directly into client methods provided by TrouSerS.

After having verified all those certificates and public key information sent by the client, the PCA has to sign the identity certificate and send it back to the client. The client is now able to use this certificate. But identity

keys are special purpose keys, which cannot be directly used for TLS authentication. They are meant to be used to certify new keys, which may then be used for different purposes. Although the identity certificate itself is a valid X.509 certificate, those new certificates are not valid X.509 certificates anymore. The problem lies in the `basicContraint` extension of the identity certificate. According to the TCG, the extension has to be set to `CA:false` [2]. This means, that there is no possiblity to create a valid X.509 certificate beneath the identity certificate. The reason for this constraint lies in the structure of X.509 certificates:

```
Certificate ::= SEQUENCE {
        tbsCertificate       TBSCertificate,
        signatureAlgorithm   AlgorithmIdentifier,
        signatureValue       BIT STRING
}
TBSCertificate ::= SEQUENCE {
        version             EXPLICIT Version DEFAULT v1,
        serialNumber        CertificateSerialNumber,
        signature           AlgorithmIdentifier,
        issuer              Name,
        validity            Validity
        subject             SubjectPublicKeyInfo,
        issuerUniqueID      IMPLICIT UniqueIdentifier OPTIONAL,
        subjectUniqueID     IMPLICIT UniqueIdentifier OPTIONAL,
        extensions          EXPLICIT Extensions OPTIONAL
}
```

After having specified all the `TBSCertificate` values, the structure gets hashed and signed with the CA's key. The signature itself will be filled in to the `Certificate→signatureValue` field. The certified TPM keys are a bit different. The `TCPA_CERTIFY_INFO` structure, returned by the method used to certify keys (`Tspi_Key_CertifyKey`) looks like this:

```
typedef struct tdTPM_CERTIFY_INFO {
        TPM_STRUCT_VER              version;
        TPM_KEY_USAGE               keyUsage;
        TPM_KEY_FLAGS               keyFlags;
        TPM_AUTH_DATA_USAGE         authDataUsage;
        TPM_KEY_PARMS               algorithmParms;
        TPM_DIGEST                  pubkeyDigest;
        TPM_NONCE                   data;
        BOOL                        parentPCRStatus;
```

```
        UINT32                      PCRInfoSize;
        [size_is(pcrInfoSize)] BYTE* PCRInfo;
} TPM_CERTIFY_INFO;
```

This structure will also be hashed and signed by the identity key, but obviously, the values to be signed are completely different from those in X.509 certificates. However in order to use these so called certified keys in a TLS authentication, there has to be a possibility to transmit their public portion to the TLS server. As the TLS standard [11] requires X.509 certificates, the authors decided to implement the proof-of-concept prototype using slightly modified X.509 certificates in order to be able to integrate an unmodified TLS implementation. Therefore new X.509 extensions have been defined, which are simple aliases of the `nsComment` extension to transmit the `TCPA_CERTIFY_INFO` structure within an X.509 container. Those certificates cannot be verified by standard TLS implementations like OpenSSL [12]. In order to overcome this problem, the authors propose to use a verification service provided for instance by the PCA as explained below.

3.1.4 The Concept of an OpenSSL Engine

From OpenSSL version 0.9.6 on came a new concept called OpenSSL Engine [13]. Engine objects represent acceleration hardware or hardware tokens like smart cards to be used with OpenSSL. In 2007, IBM also provided an OpenSSL engine for TPMs [14]. This proof-of-concept prototype makes use of this OpenSSL TPM engine in order to integrate the TPM into the wpa_supplicant.

3.1.5 An Open Source EAP Peer – wpa_supplicant

There are several EAP peer implementations for every operating system like the open source XSupplicant [15] and wpa_supplicant [16] for Linux/Unix and Windows XP. Furthermore, there are also closed source application like the Cisco Secure Services Client [17] or the integrated Microsoft Windows client [18]. The authors decided to use wpa_supplicant, since this is the standard client in many Linux installations.

wpa_supplicant supports a wide range of actual wireless and wired authentication methods like WEP, WPA, WPA2 and many different EAP methods. For the cryptographic methods, it uses OpenSSL by default. It also

comes with engine support at least for smart cards. In order to enable also tpm engines, a new config option has been defined:

```
tpm_engine_path=/usr/local/lib/openssl/engines/libtpm.so
```

Furthermore, there has to be some tpm engine specific initialization code:

```
static int
tls_engine_load_dynamic_tpm(const char *tpm_so_path){
        char *engine_id = "tpm";
        const char *pre_cmd[] = {
                "SO_PATH", NULL /* tpm_so_path */,
                "ID", NULL /* engine_id */, "LIST_ADD", "1",
                "LOAD", NULL, NULL, NULL
        };
        if (!tpm_so_path) return 0;
        pre_cmd[1] = tpm_so_path;
        pre_cmd[3] = engine_id;
       wpa_printf(MSG_DEBUG, "ENGINE: Loading TPM Engine from
       %s",tpm_so_path);
return
tls_engine_load_dynamic_generic(pre_cmd, NULL, engine_id);
}
```

Basically that is all to integrate the OpenSSL TPM engine into wpa_supplicant. In order to use the TPM engine and certificates stored in the TPM, the client software has to specify the `tpm_engine_path`, `engine_id="tpm"`, the key's `UUID` as `key_id` and the TPM's owner password as `pin` in the wpa_supplicant's configuration file.

3.1.6 An Open Source EAP Authentication Server – FreeRADIUS

Similar to the different EAP supplicant implementations, there are also several implementations of authentication servers. There are comercial products from Cisco [19] or Microsoft [20], but in the open source community, the most widely deployed EAP authentication server is FreeRADIUS [21]. This is a modular authentication server, implemented in the C programming language. Authentication methods are implemented as modules, which makes it easily extensible.

As written above, the first prototype uses sligthly modified X.509 certificates, which allows using the EAP-TLS module with only minor changes. The only thing that needs to be changed is the certificate

verification as the TPM certificates are no valid X.509 certificates anymore. In order to be able to verify those certificates, the Privacy CA has to provide a verification service. The FreeRADIUS server tries to verify the client certificate as always using OpenSSL. In case OpenSSL cannot verify the certificate, the server has to open the certificate's extension to determine whether it is a TPM certificate or not. In case it came from a TPM, FreeRADIUS sends the TPM relevant X.509 extensions to the verification service, which replies with SUCCESS or FAILURE. Based on this reply, the FreeRADIUS server decides to authenticate the client or not. In order to avoid attacks on this verification process, the FreeRADIUS server and the verification service have to communicate over SSL using mutual authentication.

3.1.7 The Verification Service

In a valid setup, the client knows its certificate chain from root to its own certificate. In the proposed setup, the client's chain looks like this: `Privacy CA's root certificate → client's identity certificate → client certificate used for authentication`. As written above, the identity certificate comes with the `CA:false` constraint, which means, that this chain is not a valid X.509 chain. The valid part ends with the identity certificate. In a valid setup, the client sends its whole chain to the server, which is then able to verify the client's certificate easily, but in the prototype, the client will only send the last certificate and not the whole chain. But the Privacy CA (PCA) knows the whole chain! That means the establishement of a verification service at PCA solves the problem. In order for the verification service to be able to map the first part of the chain (`Privacy CA's root certificate → client's identity certificate`) to the client'certificate, it has to know the serial number of the approproate identity certificate. This number will be sent in the client's extensions. Using this number and the special TPM extensions, the PCA is able to verify the client's certificate.

4 IMPROVEMENTS

The usage of invalid X.509 certificates is probably not the best choice. Therefore, the next prototype will work with a new kind of certificates, designed for the special needs of the TPM:

```
Certificate ::= SEQUENCE {
```

```
        parentSerialNumber    CertificateSerialNumber,
        pubKey                OCTET STRING,
        tpmCertificate        TPMCertificate,
        signatureValue        BIT STRING
}
TPMCertificate ::= SEQUENCE {
        versionMajor                    OCTET,
        versionMinor                    OCTET,
        versionRevMajor                 OCTET,
        versionRevMinor                 OCTET,
        keyUsage                        OCTET STRING,
        keyFlags                        OCTET STRING,
        authDataUsage                   OCTET,
        algorithmID                     OCTET STRING,
        encScheme                       OCTET STRING,
        sigScheme                       OCTET STRING,
        parmSize                        INTEGER,
        parms                 [0]       OCTET STRING OPTIONAL,
                --If not present,parmSize MUST be 0--
        pubkeyDigest                    OCTET STRING,
        nonce                           OCTET STRING,
        parentPCRStatus                 BOOLEAN,
        PCRInfoSize           [1]       INTEGER OPTIONAL,
                --If not present,parentPCRStatus MUST be FALSE—
        PCRInfo               [2]       OCTET STRING OPTIONAL,
                --If not present,parentPCRStatus MUST be FALSE—
}
```

Such a new certificate represents the `TPM_CERTIFY_INFO` structure perfectly. As this new certificate knows the serial number of its X.509 parent certificate, sending valid chains becomes possible again, even if those chains are no X.509 chains anymore! The new chain looks like this: `X.509 PCA Root Certificate → X.509 Identity Certificate → TPM Certificate` and will be stored in a file called `<name>.tpm`.

However, the Transport Layer Security protocol (TLS) requires X.509 certificates to work correctly [11], which means that a new protocol must be defined when using special TPM certificates. Therefore, the next version of the authentication protocol will be adapted and then called EAP-TPM.

5 EVALUATION AND CONCLUSION

The implementation has shown that existing EAP-TLS implementations may be adopted very easily in order to provide TPM support. On the client

side, the supplicant has to support TPM access to hold the private key in a secure environment. That is very similar to a smart card based setup, which is already supported by many supplicants. Things are bit more complicated on the server side. Due to the fact, that the certificates used in this implementation are invalid X.509 certificates, the server needs a new verification method. In order to avoid that the client has to send its identity certificate explicitly to the server, a verification service has been implemented, which is located at the Privacy CA. The Privacy CA already knows the identity certificates of its clients. The reason, why the authors decided not to send the identity certificate is the following: Using invalid X.509 certificates for the authentication is not desired for a productive version of the protocol. Those certificates are only used in this first prototype, which should show a proof-of-concept in a short time. Inserting a new message into the SSL handshake to transmit the identity certificate explicitly does not make sense since it had to be reverted for later versions. That is why it has been decided to use an external verification service instead. Furthermore, a new certificate type has been presented that is able to handle TPM certified keys and will be used in the next version of this new authentication protocol called EAP-TPM.

This prototype runs on a standard Linux system. The only applications that need to be modified are the supplicant and the RADIUS server, which means, that this prototype may run on every TPM equipped Linux based system, no matter whether it is a fully deployed computer or an embedded device.

Having such an authentication scheme will help to make 802.11 enabled mobile phones as popular as GSM phones. Furthermore, as this protocol runs also on normal computers, it will encourage more users to use public hotspots since it has never been so comfortable before.

6 REFERENCES

[1]**Latze, Carolin, Ultes-Nitsche, Ulrich und Baumgartner, Florian.** Strong Mutual Authentication in a User-Friendly Way in EAP-TLS. *Proceedings of the 15th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2007).* Split - Dubrovnik, Croatia : s.n., 2007.

[2]**Trusted Computing Platform Alliance (TCPA).** *Main Specification Version 1.1b.* 2002.

[3] The Trusted Computing Group. [Online] https://www.trustedcomputinggroup.org/home.
[4]NTRU - Products - Trusted Computing. [Online] http://www.ntru.com/products/tcg_ss.htm.
[5]**TU Graz.** Trusted Computing for the Java (tm) Platform. [Online] http://trustedjava.sourceforge.net.
[6] TrouSerS - The Open Source TCG Software Stack. [Online] http://trousers.sf.net.
[7]**TU Graz.** OpenTC PKI. [Online] http://opentc.iaik.tugraz.at.
[8]**Finney, Hal.** Privacy CA. [Online] http://www.privacyca.com.
[9]**W3C.** The XML Key Management Protocol. [Online] http://www.w2c.org/TR/xkms2.
[10]**Fielding, Roy Thomas.** *Architectural Styles and Design of Network-based Software Architectures.* s.l. : University of California, Irvine, 2000.
[11]**Dierks, T und Allen, C.** *The TLS Protocol - Version 1.0.* 1999. RFC 2246.
[12] OpenSSL. [Online] http://www.openssl.org.
[13]**Messier, Matt, Viega, John und Chandra, Pravir.** *Network Security with OpenSSL.* 2002.
[14] OpenSSL TPM Engine. [Online] 2007. http://sourceforge.net/project/showfiles.php?group_id=126012.
[15]IEEE 802.1X Open Source Implementation. [Online] http://open1x.sourceforge.net.
[16]Linux WPA/WPA2/IEEE 802.1X Supplicant. [Online] http://hostap.epitest.fi/wpa_supplicant/.
[17]Cisco Secure Services Client. [Online] http://www.cisco.com/en/US/prod/collateral/wireless/ps6442/ps7034/product_data_sheet0900aecd805081a7.html.
[18]Windows XP Wireless Auto Configuration. [Online] http://technet.microsoft.com/en-us/library/bb878124.aspx.
[19]Cisco Secure Access Control Server for Windows. [Online] http://www.cisco.com/en/US/products/sw/secursw/ps2086/index.html.
[20]Microsoft Internet Authentication Service. [Online] http://technet.microsoft.com/en-us/network/bb643123.aspx.
[21]The FreeRADIUS Project. [Online] http://www.freeradius.org.

COLLECTIVE IMPROVISATION:

COMPLEMENTING INFORMATION SECURITY

FRAMEWORKS WITH SELF-POLICING

[1]Kennedy Njenga

[2]Irwin Brown

[1]Department of Business Information Technology, University of Johannesburg; Tel 011 559 1253 Email: knjenga@uj.ac.za

[2] Department of Information Systems, University of Cape Town; Tel 021 650 4260 Email: Irwin.Brown@uct.ac.za

ABSTRACT

The approach to information security governance has predominantly followed a functionalist paradigm with emphasis placed on formalized rule structures and policy frameworks. The alternative socio-organisational (reflexive) approach has in the recent past grown in prominence due to the emergent socio-organizational aspect of technologies and processes. This paper challenges the epistemology of the functionalist approaches which assumes predictability. Information security practitioners realize that much of their activities are adapted to fit emergent changes. The aim of this paper is to explore an antidote to functionalist structured approaches by conceptualizing collective improvisation and *self-policing*. A case study approach that incorporates grounded theory techniques is employed for this purpose. Tentative findings reveal that collective improvisation is most pronounced in activities related to operational activities in governance. The implications of these and other findings are also discussed.

KEY WORDS

Information Security Governance, Collective Improvisation, Self-Policing

COLLECTIVE IMPROVISATION: COMPLEMENTING INFORMATION SECURITY FRAMEWORKS WITH SELF-POLICING

1 INTRODUCTION

When the United States congress enacted the Sarbanes-Oxley Act of 2002 ("SOX") to protect investors and combat corporate crime, what followed was an active role by corporate directors and by extension, security practitioners who became mandated to improve corporate governance and information security governance. Von Solms & Von Solms (2004) have called for broader responsibilities by management regarding information security.

According to Von Solms (2006), corporate governance consists of structured frameworks for internal controls and policies that are directed and managed by organizations. Information security governance is seen as a subset of organizations' overall corporate governance program. Structured frameworks in information security governance include CoBIT, King, COSO, and ISO 17799 (explained further in the subsequent sections). The design of many of these frameworks can be explained by understanding the functionalist paradigm and approach which is evidenced by numerous publications that offer normative guidelines for implementing and managing secure information systems (Baskerville 1988; Straub & Welke 1998).

In recent times, Hu *et al.* (2007) has argued for a more coherent socio-organizational framework that explains deviation from a 'functionalist only' approach. They propose a holistic framework that takes into account practitioners' unique reflexive behaviour. Reflexivity refers to the reconfiguration of normative orientations that guide actors and organisations (Beck 1997). Ogus (2000) talks of reflexivity in terms of reforming the conventional structures of 'command and control' governance. This paper introduces an insightful alternative by proposing a multi-faceted approach that includes reflexivity and collective improvisation into the domain of

information security governance. Improvisation, derived from the Latin word '*improviso*' is defined as 'situated performance where thinking and action occur simultaneously and on the spur-of-the-moment' (Ciborra 1999). According to Ciborra (1999), collective improvisation refers to the combined improvisational effort of several individuals or organizations. The motivation for this research is of interest since the current thinking regarding information security governance is not well-known. Ciborra *et al.* (2000) has documented improvisation in organisations and explains it as a simultaneously structured and unpredictable, often emergent and opaque phenomenon. The nature of this paper extends an analytical understanding of collective improvisation in information security activities and proposes the following research question that contextualises the issue;

How is collective improvisation manifested in information security governance activities?

In addition, the paper aims at exploring how collective improvisation influences practitioner's actions towards understanding information security governance issues. The paper makes a theoretical contribution by arguing that practitioners' engagement with policy is essentially driven by novelty and reflexivity and expressed as *self-policing*. *Self-policing* is a concept that often leads to less enforcement activity and deterrence (Innes 1999).

The paper is structured into six main sections. This first section has introduced and set the context for research. In the second section the functionalist approach is introduced. The third section presents a multi faceted improvisational approach. The fourth section describes the research methodology. In this section, the use of grounded theory techniques is explained and justified. The fifth section presents and discusses research findings. In the final sixth section the paper is concluded by deriving implications for IS practitioners and researchers.

2 PREDICTIVE KNOWLEDGE: THE FUNCTIONALIST PARADIGM

Information security researchers have recognized the significance of well planned sound information security policies that focus on clear methodologies and programmes (Von Solms & Von Solms 2005; Schultz 2005). In their studies in Information Security, Dhillon & Backhouse (2001) have noted the dominance of the functionalist approach that emphasizes

formalized rule structures in designing and managing security. It is the notion of predictive knowledge that has influenced the functionalist approach to formulating policies for monitoring and control (Wheeler and Venter 2006). Predictive knowledge hence reinforces the functionalist paradigm when viewing designers and practitioners as solely technical experts (Wheeler and Venter 2006).

Predictive knowledge assumes the intent by users to follow order, maintain status quo and reinforce rational choice (Wheeler and Venter 2006). The functionalist structured approaches to information security have generated interest among information security researchers (e.g. Straub and Welke 1998; Siponen 2000; Von Solms and Von Solms 2005; Vorster and Labuschagne 2006) and is characterised by the use of many policies, frameworks and standards meant to foster order and control. **Figure 1-1** points to one of the many structured functionalist approaches to information security governance noted by the researcher.

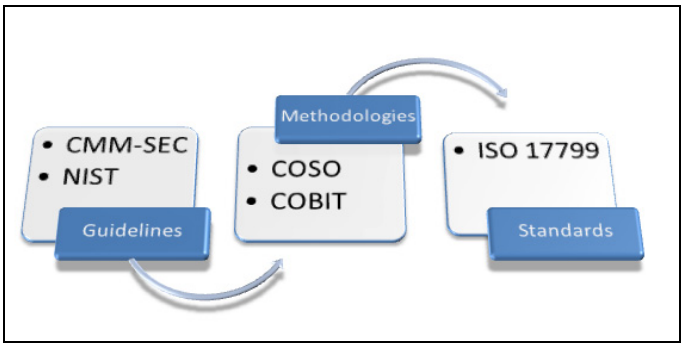


Figure 1-1 Structured Functionalist Approach to Information Security Governance

2.1 Guidelines

Information security governance is endowed with rich functionalist guidelines which provide direction for the activities or process to achieving set goals. The Capability Maturity Model for Security (CMM-SEC) is an example of a guideline that defines the process an enterprise must go through to move from limited security capabilities to increasingly optimizing protection postures (Burton Group 2005). The National Institute of Standards and Technologies (NIST) has issued guidelines under the

banner of the *Risk Management Guide for Information Technology Systems* in its Special Publication 800-30 (NIST SP 800-30). The ISO/IEC Guide 73 released jointly by the International Organization of Standards (ISO) and the International Electro-technical Commission (IEC) provides specific guidance on terms and definitions of concepts related to risk management.

2.2 Methodologies and Frameworks

There are a variety of functionalist methodologies in use to identify, measure, control and monitor information security risks. These stem either from government regulations e.g. Sarbanes-Oxley Act (SOX) or industry recommendations such as CoBIT™, COSO, (Committee of Sponsoring Organisations of the Treadway, Commission, (*Internal Control—Integrated Framework*, 1992). There is also Turnbull in the UK, CoCo in Canada, KING II, in South Africa and the IT Infrastructure Library (ITIL) for IT service management.

2.3 Standards

Standards are also functionalist in nature and are continually developed for the purpose of serving as measures for organizations to achieve desirable ends. They fall short of the main purpose of guidelines and frameworks since these do not show how to achieve stated ends. The latter assist organizations by showing how these stated ends may be achieved. An example of a prominent standard in use in South African is the ISO/IEC 17799 standard adopted from the British Standard BS 17799.

It is only by a closer examination of the information security risk management process and specifically the policy adoption process does one realize that information security activities are guided by an approach that is multi-faceted and not restricted to only blindly following frameworks, guidelines or standards in a purely functionalist manner.

3 REFLEXIVITY AND IMPROVISATION

Even as early as the 1970's scholars and researchers of systems thinking (Cleveland 1973) generated ideas of the need for holistic systems approaches to management. Cleveland (1973) argued on the need to match what became known as unsystematic reality with 'constructive ambiguity'. This argument opposed the functionalist premise while proposing that the management systems then, were too exact, too clear and therefore too rigid.

In the present times, the same attributes are still common in many organizations and have been instrumental in shaping and monitoring policy. The main problem with this thinking then and now is that in an effort to build efficient systems, scholars have been tempted to analyse and view everything systematically, while avoiding the soft socio-cognitive aspects of purposes and meaning. The gap in approach has been filled presently by studies relating to reflexivity and improvisation (Ciborra 1999; Ogus 2000). Much has been written concerning improvisation, strategy formulation and implementation (Perry 1991). These studies acknowledge actions that provide for reflexivity, in the sense that activities could be done in more that one way and each way finely fitting the situation (Scribner 1984). *Self Policing* is seen as the expression of reflexivity and increases efficiency in governance in two ways; one, remediation is achieved early; two, there is reduction in enforcement effort (Innes 1999). In an effort to understand this soft discourse, the next section presents the methodology that was used.

4 RESEARCH METHODOLOGY

A single case research strategy was employed, which was exploratory, interpretive and contextual. It sought to generate new insights into the phenomenon of collective improvisation in information security. As a pointer, this researcher drew a level of comfort from the interpretive paradigm. The researcher was able to identify, examine and evaluate the phenomenon of collective improvisation through the subjects' eyes and from the subjects' perspective (Hu *et. al*. 2007; Strauss & Corbin 1998). The interpretive paradigm permitted the researcher to provide useful insights that integrated the technical and the sociological human aspects of information security.

4.1 Grounded Theory Techniques

The researcher used grounded theory techniques to inductive derive a framework that emphasizes the fit between data and 'reality'. Grounded theory techniques, (Glaser & Strauss 1967; Glaser 1978; Strauss 1987; Strauss & Corbin 1990) formed a basis for content analysis of raw data and proved an attractive means for inductive reasoning. It should be noted that grounded theory has been used successfully in both organizational and information systems research (Orlikowski 1993; Sarker *et. al.* 2001; Trauth & Jessup 2000; Urquhart 1997).

4.2 Data Collection

Gathering primary data on information security proved to be challenging. What was experienced confirmed the findings of Kotulic & Clark (2004) namely that organisations are reluctant to share information about security policies with individuals from outside the company. The primary data was gathered and consisted of a series of 11 in-depth interviews with senior practitioners. The single organization was a large multi-national corporation.

5 RESULTS AND INTERPRETATION

The researcher used ISO 17799 domains to establish **Units of Analysis** or activities common in information security governance that employed a high degree of collective cognitive abilities. The researcher then interviewed practitioners engaged in these activities. The recorded interviews were transcribed and arranged into themes related to each of these units for analysis. Codes were derived from the transcripts that would help establish the level of conceptual density of instances of reflexivity and collective improvisation in these units. High level concepts were derived from these codes. What followed was the deriving of still even higher level categories from the concepts related to collective improvisation in each of the units. **Table 1** shows a mapping of the units of analysis to the ISO 17799 structured domains.

Table 1. Mapping ISO 17799 Domains to Research Units of Analysis

CORE InfoSecurity Management Activities ISO 17799 Sections			Re-search ed	Unit of Analysis
	Section	Type of Activity (Domain)		
	1	Introduction Reference text n/a	n/a	n/a
IDENTIFY	2	Introduction Reference text n/a	n/a	n/a
	3	Security policy	Yes	3 Information Security Policy
	4	Security organisation	*No	
	5	Information Asset Classification and Control	Yes	1 Assets control
ANALYSE	6	Personnel Security	*No	
	7	Physical and Environmental Security		2 Information Architecture Security
	8	Communications and Operations Management	*No	
RESPOND	9	Access Control	*No	
	10	System Development and Maintenance	Yes	4 Event Monitoring
	11	Business Continuity and Management	Yes	6 Business Continuity
	12	Compliance with legal requirements	Yes	5 Governance and Regulatory Compliance

* No – *activities that were deemed to lack any depth in cognitive- reflexivity were not researched on*

The understanding and integration of concepts and categories was done iteratively (Glaser & Strauss 1967). **Table 2** below shows the process

undertaken by the researcher to analyse data using grounded theory techniques.

Table 2. Research Process

RESEARCH PROCESS			
Process 1	**Analyse data relating to the** first unit of analysis to conceptualise improvisation	Use open coding	Develop concepts, and categories relating to improvisation in information security activities
Process 2	Theoretical sampling	Literal and theoretical replication across cases (go to process 3 until theoretical saturation)	Confirms, extends, and sharpens theoretical framework by analysing the rest of the units of analysis
Process 3	**Analyse data relating to** the subsequent other units of analysis to conceptualize improvisation	Use open coding	Develop concepts, and categories relating to improvisation in information security activities
Process 4	Explore relationships between concepts and	Use axial coding	Develop connections

RESEARCH PROCESS			
	Categories from all units of Analysis	Use selective coding	between a category and its sub-categories Integrate categories to build theoretical framework
Process 5	Reaching closure	Theoretical saturation when possible	Ends process when marginal improvement becomes small

5.1 Interpretation

Over 200 codes were generated and 19 independent, contextual concepts relating to collective improvisation were identified. Each unit of analysis was analysed independently. From analysing the concepts, it was discovered that collective improvisation was more conceptually dense, i.e. occurred in many instances on the operational based domains listed in ISO 17799. A summary of results that analysed the level of conceptual density is shown in **Table 3** below.

Table 3. Mapping of Concepts with Collective Improvisation

Units of Analysis Activities related to;	**Sub categories** **(Collective Improvisation)**	**Core Categories**	*Conceptual Density of Concepts*	*Concepts*
1 **Assets control**	☑	**Strategic**	1	***Being practical***
		Tactical		
		Operational		
2 Information Architecture Security	☑☑☑	**Strategic**	3	Exceptionality, Being inventive, Rational adoptive
		Tactical		
		Operational		
3 Information Security Policy	☑	**Strategic**	1	Being quick-witted
	☑	**Tactical**	1	Lateral thinking
		Operational		
4 Event Monitoring		**Strategic**		
	☑	**Tactical**	1	Being ingenious
	☑	**Operational**	1	Being capable
5 Governance and Regulatory Compliance	☑☑	**Strategic**	2	Getting by, Being practical
		Tactical		
		Operational		
6 Business Continuity	☑☑☑☑	**Strategic**	4	Being quick-witted, Being innovative, Lateral thinking, Being ingenious
		Tactical		

Units of Analysis Activities related to;	**Sub categories** **(Collective Improvisation)**	**Core Categories**	*Conceptual Density of Concepts*	*Concepts*
	☑☑☑☑☑	**Operational**	5	Being resourceful, Managing, Being inspired, Quick reaction, Exceptionality
	19		*19*	

5.2 Axial Codes: Establishing the Principle of Self Policing by Substituting Frameworks

Tentative findings reveal that collective improvisation is most pronounced in activities related to operational activities (specifically business continuity) in governance. Collective improvisation was particularly expressive in *self-policing* where practitioners were at operational level collectively vigilant in extending *self-policing* procedures to deter, quickly investigate and contain threats to information. Using Axial Coding the researcher was able to draw relationships between various core categories. This enabled the researcher to come up with the diagram below. **Figure 1-2** shows the relationships between the units of analysis and the core categories (Strategic, Tactical and Operational).

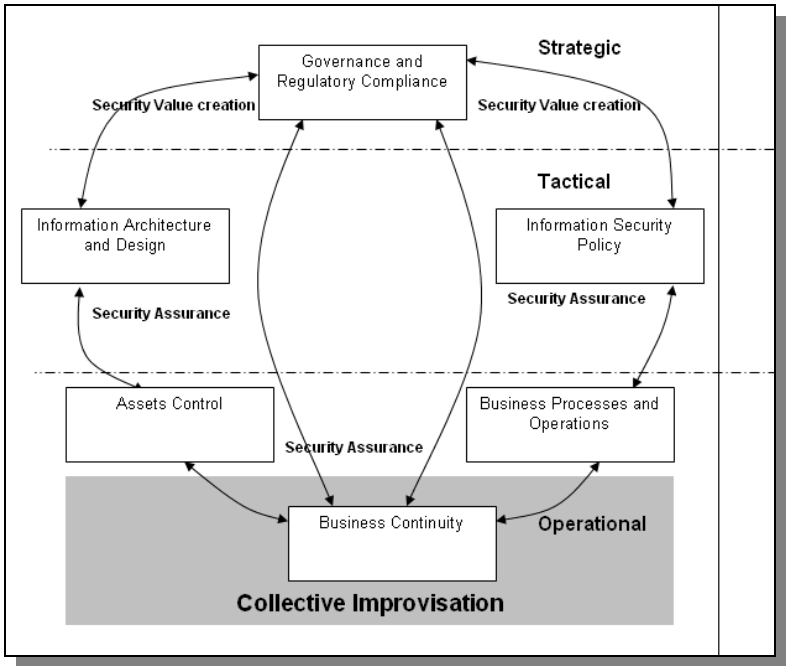


Figure 1-2 Relationship between Categories: Conceptual Density of Collective Improvisation

Figure 1-2 illustrates the conceptual density of collective improvisation as exemplified in areas where practitioners collectively engaged innovatively. Deeper insights reveal that the internalized knowledge of these practitioners was expressive in these innovative engagements temporarily substituting frameworks with their own *self-policing* initiatives. *Self-policing* and reflexivity made it necessary for the practitioners to be adaptive to contingencies.

6 CONCLUSION

To conclude, the paper has generated new insights and suggested a holistic understanding of a wide spectrum of socio-cognitive issues related to information security governance. An important part of this discourse was introducing the idea of reflexivity and collective improvisation and the role it played in information security governance. Such a role was manifested as *self-policing.*

Through a case study research, the paper has proposed a framework for understanding collective improvisation in information security governance. By understanding the proposed framework, practitioners will be able to appreciate a multi-faceted approach to Information Security governance. As laid down by the paper, the framework proposed should accommodate reflexive ways of dealing with 'intractable problems'; away from narrow structured based approaches. Reflexivity should be accommodated in the planning, management and monitoring of information security within an organisation.

7 REFERENCES

Baskerville, R., (1988) "Designing Information Systems Security" John Wiley & Sons, New York, NY.

Beck, U. (1997), The reinvention of Politics: Rethinking Modernity in the Global Social Order, Polity Press, Cambridge.

Burton Group. (2005) Security and Risk Management Strategies, "A Systematic, Comprehensive Approach to Information Security". Version 1.0 http://www.burtongroup.com/Content/doc.aspx?cid=644

Ciborra, C. (1999) A theory of information systems based on improvisation, in Rethinking Management Information Systems (Eds: W. Currie & R. Galliers), Oxford University Press, Oxford.

Ciborra C.; Braa K.; Cordella A.; Dahlbom b.; Hanseth O.; Hepso V.; Ljungberg J.; Monterio E.; and Simon K. A. (2000) 'From Control to Drift'., Oxford: Oxford University Press.

Cleveland H., (1973) "Systems, Purposes and the Watergate" Operations Research, Vol. 21: 5 pp 1019-1023

Dhillon, G. and Backhouse, J. (2001) "Current Directions in IS Security Research: Toward Socio-organizational Perspectives," Information Systems Journal, Vol. 1:1 pp. 11-12.

Glaser, B., G. and Strauss A (1967) "The Discovery of Grounded Theory: Strategies for Qualitative Research", Aldine Publishing Co, Chicago IL.

Glaser, B., G. (1978) "Theoretical Sensitivity: Advances in the Methodology of Grounded Theory", Sociology Press, CA.

Hu, Q., Hart, P., and Donna Cooke, D., (2007) "The role of external and internal influences on information systems security – a neo-institutional perspective", Journal of Strategic Information Systems Vol. 16 pp. 153–172.

Innes, R., (1999) "Self-Policing and Optimal Law Enforcement When Violator Remediation is Valuable" Journal of Political Economy Vol 107:6 pp. 1305-1325

Kotulic, A.G. and Clark, J.G. (2004) "Why there aren't more information security research studies", Information & Management Vol. 41:5 pp. 597-607.

National Institute of Standards and Technology (NIST): US Department of Commerce "Risk Management Guide for Information Technology Systems" Special Publication 800 -30

http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf

Ogus, A. (2000), "Self-regulation", in B. Bouckaert et G. De Geest (eds.), Encyclopedia of Law and Economics, Volume V: The Economics of Crime and Litigation, Edward Elgar, Cheltenham, pp. 587-602.

Orlikowski, WJ (1993) "CASE Tools as Organizational Change: Investigating Incremental and Radical Changes in Systems Development" MIS Quarterly Vol. 17:3 pp. 309-340

Perry L.T., (1991) "Strategic Improvising: How to formulate and Implement Competitive Strategies in Concert" Organisational Dynamics (19:4) pp. 51-64

Sarker, S., Lau, F. and Sahay, S. (2001), "Using an adapted grounded theory approach for inductive theory building about virtual team development," The DATA BASE for Advances in Information Systems Vol 32:1 pp. 38-56.

Schultz E. (2005) "Security dilemmas with Microsoft's Internet Explorer". Computers and Security, Vol 24:3 pp. 175-176

Scribner, S. (1984) Studying working intelligence. In B. Rogoff & J. Lave (Eds.), Everyday Cognition: Its Development in Social Context pp. 9-40. Cambridge: Harvard University Press.

Siponen, M., T. (2000) "A Conceptual foundation for organisational Information security awareness"; Information Management and Computer Security Journal Vol 8:1 pp 31-41.

Straub, D.W. and Welke, R.J., (1998) 'Coping with Systems Risk: Security Planning Models for Management Decision Making': MIS Quarterly, Vol. 22:4 pp. 441-464.

Strauss, A., (1987) "Qualitative Analysis for Social Scientist" Cambridge University Press, Cambridge UK.

Strauss, A. and Corbin, J. (1998) "Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory" Sage Publications, Thousand Oaks, CA.
Strauss, A. and Corbin, J. (1990), "Basics of Qualitative Research: Grounded Theory Procedures and Techniques" Sage, Thousand Oaks, CA
Trauth, E.M. and Jessup, L.M. (2000) "Understanding computer-mediated discussions: positivist and interpretive analyses of group support system use," MIS Quarterly Vol. 24:1 pp. 43-79.
Urquhart, C. (1997), "Exploring analyst-client communication: using grounded theory techniques to investigate interaction in informal requirements gathering", in Lee, A.S., DeGross, J.I. and Liebenau, J. (Eds), Information Systems and Qualitative Research, Chapman & Hall, London pp. 149-81.
Von Solms, B., Von Solms, R., (2004) "Ten deadly sins of security management" Computers & Security Vol. 23 pp. 371–376.
Von Solms B and Von Solms R (2005) 'From Information security to…business security'? Computer and Security Vol 24:4 pp 271-273.
Von Solms, B. (2006). "What every Vice-Chancellor and Council Members should know about the use of ICT" Proceedings of the Conference on Information Technology in Tertiary Education, Pretoria, South Africa.
Vorster A., and Labuschagne L. (2006). "A new comparison framework for information security risk analysis methodologies", South African Computer Journal, Vol 37 pp. 98 – 106.
Wheeler M., and Venter H. (2006). " Change Management: A case study at the University of Pretoria", Proceedings of the Conference on Information Technology in Tertiary Education (CITTE) Pretoria, South Africa.

AN INTRODUCTION TO STANDARDS RELATED TO INFORMATION SECURITY

Johann Amsenga

Eclipse RDC, a division of Armscor Business (Pty) Ltd
amsenga@acm.org
+27 12 671 6915
P.O. Box 7036, Pretoria, Republic of South Africa, 0001

ABSTRACT
"The good thing about standards is that there are so many of them." This humorous comment is often made when some well meaning team member wants to solve a problem by referring to a standard. This may be true, but what is also true is that information systems are becoming more "complex" (in the vaguest sense of the word), that systems and the information processed are more distributed and that the requirement for access is more demanding. Also, the requirement for access is not limited to, or from, a specific or single site, organisation or even country. This has a huge effect on interfacing requirements, information storage and presentation formats and, of course, security.

Adopting internationally recognised standards is a definite route to solve a lot of these problems. Standards are a mechanism for different stakeholders to refer to a common, trusted reference. Standards provide a common technological language, thus enabling a system stakeholder to provide definitions for terms used in a project, and to qualify vague expressions such as "complex".

The South African Bureau of Standards (SABS) is the recognised national institution for the promotion and maintenance of standards in South Africa. The SABS prepare and publish South African National Standards (identified by the letters SANS) reflecting national consensus on a wide range of subjects. A business unit of the SABS, Standards South Africa (StanSA), administers more than 450 technical committees and subcommittees to produce standards. The SABS is a member body of the

International Organisation for Standardisation (ISO) and participates actively in a number of their committees.

This tutorial provides a short introduction to International and South African National Standards related to Information Security. Some of the existing standards are highlighted and the development process is introduced. The tutorial focuses on ISO/IEC International Standards and the national adoption or development by StanSA.

KEY WORDS
Information Security, Security Requirements, System Life Cycle, International Standards, National Standards, SABS, StanSA, ISO/IEC, Development of Standards

AN INTRODUCTION TO STANDARDS RELATED TO INFORMATION SECURITY

1 INTRODUCTION

Information systems are becoming more complex – systems and the information processed are more distributed and the requirement for access is more demanding. Also, the requirement for access is not limited to, or from, a specific or single site, organisation or even country. This has a huge effect on interfacing requirements, information storage and presentation formats and, of course, security.

Adopting internationally and nationally recognised standards is a definite route to solve a lot of these problems. Standards are a mechanism for different stakeholders to refer to a common, trusted reference, and provide a common technological language.

The trusted reference and technological language provided by International Standards are especially important in the information security environment where many organisations view information security as new technology or an uncharted domain. These organisations often have to rely on so called security experts or, even worse, self proclaimed gurus. Security related standards can help these organisations to see through the "buzz" words and to better understand the role and place of security and the related technologies.

This paper provides an introduction to International and South African National Standards related to Information Security. Some of the existing standards are highlighted and the development process is introduced. The paper focuses on ISO/IEC International Standards and the national adoption or development by StanSA.

2 STANDARDS – PURPOSE AND ADVANTAGES

The South African Bureau of Standards (SABS) describes a standard as follows:

A Standard is a published document which lists specifications and procedures established to ensure that a material, product, method or service is fit for its purpose and perform in the manner it was intended for.

Standards define quality and establish safety criteria. Conformance to standards ensures quality and consistency.

The World Trade Organisation (WTO) defines a standard in its Agreement on Technical Barriers to Trade (TBT) as:

A document approved by a recognised body, that provides, for common and repeated use, rules, guidelines or characteristics for products or related processes and production methods, with which compliance is not mandatory. It may also include or deal exclusively with terminology, symbols, packaging, marking or labelling requirements as they apply to a product, process or production method.

Life is too short to reinvent the wheel. In the ever progressing world of information technology, it is good to know that a lot of the work has already been done. However, it is difficult to known which of the vast amount of resources can be trusted. International security related standards are developed by experts in the field, checked by the national standards bodies of many countries, and are internationally accepted and proven. National standards, on the other hand, also take the relevant country's environment into account. International standards adopted as a national standard, provides the best of both worlds.

3 INFORMATION SYSTEMS AND INFORMATION SECURITY

It is important to remember that security must be a stated requirement for any information system. Security must have the same importance as other requirements, such as functionality and usability. Also, security is a quality attribute. Security cannot be viewed in isolation. It is as much part of a system as any other component, and influences the concept, development, production, utilisation, support and retirement of a system just like any other requirement and limitation placed on the system. It is imperative that security be taken into account in all processes and stages of a system's life cycle, and that security must be managed.

From these observations, it is clear that information security cannot be addressed by applying information security standards alone. The information security standards must be used together with information system, management related and quality standards.

It is for this reason that this paper introduces not only information security standards, but also system life cycle related standards, quality

management and quality evaluation standards, and information security management standards.

4 INFORMATION SECURITY STANDARDS

Probably the most well known information security standard was the ISO/IEC 17799, which was adapted from the British standard BS 7799. This standard now forms part of the ISO/IEC 27000 family of standards that addresses Information Security Management Systems (ISMS), and has been renumbered to ISO/IEC 27002.

ISO, together with IEC, published a whole portfolio of standards related to generic methods, techniques and guidelines for information, IT and communication security. This includes the areas of security management, conformance assessments and security evaluation criteria. Work continues in the maintenance of these standards, as well as the development of new standards.

4.1 Information Security Management Systems

Information security is a fundamental component of governance and social responsibilities of organisations. Organisations are expected, and sometimes legally obliged, to implement and manage information security. Information security management systems are addressed by the ISO/IEC 27000 family of standards.

Examples of standards published or being developed in this category are:

- ISO/IEC 27000: Information technology – Security techniques – Information security management systems – Overview and vocabulary
- ISO/IEC 27001: Information technology – Security techniques – Information security management systems – Requirements
- ISO/IEC 27002: Information technology – Security techniques – Code of practice for information security management
- ISO/IEC 27003: Information technology – Security techniques – Information security management system implementation guidance
- ISO/IEC 27004: Information technology – Security techniques – Information security management measurements
- ISO/IEC 27005: Information technology – Security techniques – Information security risk management

- ISO/IEC 27006: Information technology – Security techniques – Requirements for bodies providing audit and certification of information security management systems

4.2 Cryptography and Security Mechanisms

A number of standards are produced by ISO/IEC that covers cryptographic and non-cryptographic techniques and mechanisms for use in security services. The techniques and mechanisms include:

- Confidentiality
- Entity authentication
- Non-repudiation
- Hash functions
- Digital signatures
- Key management

Examples of standards published or being developed in this category are:

- ISO/IEC 9797-1: Information technology – Security techniques – Message Authentication Codes (MACs) – Part 1: Mechanisms using a block cipher
- ISO/IEC 9798-1: Information technology – Security techniques – Entity authentication – Part 1: General
- ISO/IEC 9979: Information technology – Security techniques – Procedures for the registration of cryptographic algorithms
- ISO/IEC 10118-1: Information technology – Security techniques – Hash-functions – Part 1: General
- ISO/IEC 11770-1: Information technology – Security techniques – Key management – Part 1: Framework
- ISO/IEC 15846-1: Information technology – Security techniques – Cryptographic techniques based on elliptic curves – Part 1: General
- ISO/IEC 18033-3: Information technology – Security techniques – Encryption algorithms – Part 3: Block ciphers

4.3 Security Evaluation Criteria

The standards for IT Security evaluation and certification of IT systems, components, and products are also very important. These standards include consideration of computer networks, distributed systems, associated application services, etc. Distinction is made on three aspects:

- Evaluation criteria
- Methodology for the application of the criteria
- Administrative procedures for evaluation, certification an accreditation schemes

Examples of standards published or being developed in this category are:

- ISO/IEC 15408-1: Information technology – Security techniques – Evaluation criteria for IT security – Part 1: Introduction and general model
- ISO/IEC 15408-2: Information technology – Security techniques – Evaluation criteria for IT security – Part 2: Security functional requirements
- ISO/IEC 15408-3: Information technology – Security techniques – Evaluation criteria for IT security – Part 3: Security assurance requirements
- ISO/IEC 15443-1: Information technology – Security techniques – A framework for IT Security assurance – Part 1: Overview and framework
- ISO/IEC 15443-2: Information technology – Security techniques – A framework for IT Security assurance – Part 2: Assurance Methods
- ISO/IEC 15443-3: Information technology – Security techniques – A framework for IT Security assurance – Part 2: Analysis of Assurance Methods
- ISO/IEC 18045: Information technology – Security techniques – A framework for IT Security assurance – Methodology for IT Security Evaluation

4.4 Security Controls and Services

With the growing requirements for standards and guidelines addressing services and applications supporting the implementation of ISO/IEC 27001 control objectives and controls, a number of standards are being produced by ISO/IEC within the context of an overall internal control structure.

Examples of standards published or being developed in this category are:

- ISO/IEC 18043: Information technology – Security techniques – Selection, deployment and operations of intrusion detection systems
- ISO/IEC 18044: Information technology – Security techniques – Information security incident management

- ISO/IEC 24762: Information technology – Security techniques – Guidelines for information and communications technology disaster recovery services
- ISO/IEC 27033-1: Information technology – Security techniques – Network Security – Part 1: Guidelines for network security
- ISO/IEC 27034-1: Information technology – Security techniques – Guidelines for Application Security – Part 1: Overview and Concepts

4.5 Identity Management and Privacy Technologies

ISO/IEC develops and maintains standards and guidelines addressing security aspects of identity management, biometrics and the protection of personal data.

Examples of standards published or being developed in this category are:

- ISO/IEC 24760: Information technology – Security techniques – A framework for identity management
- ISO/IEC 29100: Information technology – Security techniques – A privacy framework
- ISO/IEC 29101: Information technology – Security techniques – A privacy reference architecture
- ISO/IEC 29115: Information technology – Security techniques – Entity authentication assurance

5 INFORMATION SYSTEM ENGINEERING STANDARDS

The introduction to ISO/IEC 15288 (Systems and Software Engineering – System life cycle processes) states the following.

The complexity of man-made systems has increased to an unprecedented level. This has led to new opportunities, but also to increased challenges for the organisations that create and utilise systems. These challenges exist throughout the life cycle of a system and at all levels of architectural detail. They arise from several sources:

- There are inherent differences among the hardware, software and human elements from which systems are constructed.
- Almost every present-day system contains, and/or is modelled and supported by computer-based technology.
- There is a lack of harmonization and integration of the involved disciplines, including science, engineering, management and finance.

There is therefore a need for a common framework to improve communication and cooperation among the parties that create, utilise and manage modern systems in order that they can work in an integrated, coherent fashion.

The standards produced by ISO/IEC in the domain of systems and software engineering, such as ISO/IEC 15288, concern those systems that are man-made and may be configured with one or more of the following: hardware, software, data, humans, processes (e.g. processes for providing service to users), procedures (e.g. operator instructions), facilities, materials and naturally occurring entities.

Standards such as ISO/IEC 15288 provide a common process framework covering the life cycle of man-made systems. This life cycle spans the conception of ideas through to the retirement of a system. It provides the processes for acquiring and supplying systems. In addition, this framework provides for the assessment and improvement of the life cycle processes.

ISO/IEC 15288 also provides processes that support the definition, control and improvement of the life cycle processes used within an organisation or a project. Organisations and projects can use these life cycle processes when acquiring and supplying systems.

All these aspects are applicable to information security. Information security always forms part of a bigger system, and should thus be part of the consideration of the whole system, throughout the full life cycle, starting with the requirements imposed by security during the requirements analysis stages. Implementing information security results in the creation of a system in its own right, subject to all processes and standards of handling man-made systems. Information systems and information security cannot and should not be viewed as two mutually exclusive subjects.

As with information security, ISO, together with IEC, publishes a whole portfolio of standards related to processes, supporting tools and supporting technologies for the engineering of software products and systems. Work continues in the maintenance of these standards, as well as the development of new standards.

Since the scope of this work is so vast, examples of only a few of the areas addressed are given here.

5.1 Software Product Measurement and Evaluation

Standards and technical reports for software products evaluation and metrics for software products and processes. The software product quality requirements and evaluation (SQuaRE) series of standards are being developed in this area. Examples are:

- ISO/IEC 25000: Software Engineering -Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE
- ISO/IEC 25001: Software engineering -Software product Quality Requirements and Evaluation (SQuaRE) - Planning and management
- ISO/IEC 25020: Software engineering -Software product Quality Requirements and Evaluation (SQuaRE) - Measurement reference model and guide

5.2 Life Cycle Management

Standards and technical reports on Life Cycle Management.

- ISO/IEC 12207: Systems and Software Engineering - Software Life Cycle Processes
- ISO/IEC 15026: Systems and Software Engineering - Systems and Software Assurance
- ISO/IEC 15288: Systems and software engineering - System life cycle processes
- ISO/IEC 15939: Systems and software engineering - Measurement process
- ISO/IEC 16085: Systems and software engineering -Life cycle processes -Risk management

6 OTHER STANDARDS

Due to the practical limitations of this paper, other applicable standardisation efforts of ISO and IEC are only listed.

6.1 Financial Services

Standards in the field of banking, securities and other financial services.

6.2 Quality Management and Quality Assurance

Standards for quality management, including generic quality management systems and supporting technologies. The well-known ISO 9000 series falls within this scope.

6.3 Telecommunications and Information Exchange Between Systems

Standards for telecommunications dealing with the exchange of information between open systems including system functions, procedures, parameters and equipment, as well as the conditions for their use.

6.4 Cards and Personal Identification

Standards in the area of identification and related documents, cards, and devices associated with their use in interindustry applications and international interchange.

6.5 Automatic Identification and Data Capture Techniques

Standards for data formats, data syntax, data structures, data encoding, and technologies for the process of automatic identification and data capture and of associated devices utilised in inter¬industry applications and international business interchanges.

6.6 Biometrics

Standards for generic biometric technologies pertaining to human beings to support interoperability and data interchange among applications and systems.

7 DEVELOPMENT OF INTERNATIONAL STANDARDS

A number of organisations develop international standards – ITU, IEEE, IEC, to name but a few. Increasingly, the SANS are being harmonised with international standards in order to facilitate trade.

The International Organisation for Standardisation (ISO) produces voluntary consensus standards through its decentralised global system of standardisation. This paper focuses on ISO because of the active involvement of the SABS in ISO and ISO/IEC committees.

ISO states the follow regarding the development of standards.

ISO is a network of the national standards institutes of 157 countries, on the basis of one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system. ISO is a non-governmental organisation, its members are not, as is the case in the United Nations system, delegations of national governments. Nevertheless, ISO occupies a special position between the public and private sectors. This is because, on the one hand, many of its member institutes are part of the governmental structure of their countries, or are mandated by their

government. On the other hand, other members are in the private sector. Therefore, ISO is able to act as a bridging

organisation in which a consensus can be reached on solutions that meet both the requirements of business and the broader needs of society, such as the needs of stakeholder groups like consumers and users.

Although ISO standards are voluntary, the fact that they are developed in response to market demand, and are based on consensus among the interested parties, ensures widespread applicability of the standards. Consensus, like technology, evolves and ISO takes account both of evolving technology and of evolving interests by requiring a review of its standards at least every five years to decide whether they should be maintained, updated or withdrawn. In this way, ISO standards retain their position as the state of the art, as agreed by an international cross-section of experts in the field.

Technical committees are established to serve specific industries or generic subjects, in order to develop International Standards or other ISO publications appropriate to the needs of that sector. Subcommittees are established and dissolved by the parent technical committee, subject to ratification by the technical management board. A subcommittee is set up to focus on specific parts of the overall standards requirement. Technical committees or subcommittees may establish working groups for specific tasks. All national bodies have the right to participate in the work of technical committees and subcommittees, either as a P-member, an O-member or a L-member.

P-members participate actively in the work, and have an obligation to vote on all questions formally submitted for voting within the technical committee or subcommittee, on enquiry drafts and final draft International Standards, and to participate in meetings. O-members follow the work as observers, and may receive committee documents and have the right to submit comments and to attend meetings. L-members (liaison members) has no power of vote, but has some options to attend meetings and receive documents.

A few of the relevant technical committees are listed below.

- ISO/IEC JTC 1 (Information Technology)
- ISO TC 68 (Financial services)
- ISO TC 176 (Quality management and quality assurance)

JTC 1 is a Joint Technical Committee of ISO and the International Electrotechnical Committee (IEC), and is responsible to develop standards and technical reports for information technology. A few of the relevant subcommittees of JTC 1 are listed below.

- ISO/IEC JTC 1 SC 6 (Telecommunications and Information Exchange Between Systems)
- ISO/IEC JTC 1 SC 7 (Systems and Software Engineering)
- ISO/IEC JTC 1 SC 17 (Cards and Personal Identification)
- ISO/IEC JTC 1 SC 25 (Interconnection of Information Technology Equipment)
- ISO/IEC JTC 1 SC 27 (Security Techniques)
- ISO/IEC JTC 1 SC 31 (Automatic identification and data capture techniques)
- ISO/IEC JTC 1 SC 37 (Biometrics)

7.1 Information Security Standards

SC 27 of JTC 1 is responsible for standardisation of generic methods and techniques for IT Security. This includes:

- identification of generic requirements (including requirements methodology) for IT system security services,
- development of security techniques and mechanisms (including registration procedures and relationships of security components),
- development of security guidelines (e.g. interpretative documents, risk analysis), and
- development of management support documentation and standards (e.g. terminology and security evaluation criteria).

SC 27 consists of the following working groups:

- WG 1 (Information security management systems)
- WG 2 (Cryptography and security mechanisms)
- WG 3 (Security evaluation criteria)
- WG 4 (Security controls and services)
- WG 5 (Identity management and privacy technologies)

7.2 Systems and Software Engineering Standards

SC 7 of JTC 1 is responsible for standardisation of processes, methods and supporting technologies for the engineering and management of software and systems throughout their life cycles.

SC 7 consists of the following working groups:

- WG 1a (ICT Governance)
- WG 2 (Systems and Software Documentation)
- WG 4 (Tools and Environment)
- WG 6 (Software Product Measurement and Evaluation)
- WG 7 (Life Cycle Management)
- WG 10 (Process Assessment)
- WG 19 (Open distributed processing and Modelling Languages)
- WG 20 (Software Engineering Body of Knowledge)
- WG 21 (Asset Management)
- WG 22 (Vocabulary)
- WG 23 (System Quality Management)
- WG 24 (SLC Profiles and Guidelines for VSE)
- WG 25 (IT Service Management)
- WG 26 (Software testing)
- WG 42 (Architecture)

8 DEVELOPMENT OF SOUTH AFRICAN STANDARDS

The South African Bureau of Standards (SABS) is the recognised national institution for the promotion and maintenance of standards in South Africa. The SABS prepare and publish South African National Standards (identified by the letters SANS).

As with ISO, standards are developed by committees. Committees can be technical committees (TCs), subcommittees (SCs) of technical committees, or working groups (WGs). A business unit of the SABS, Standards South Africa (StanSA), administers more than 450 technical committees and subcommittees to produce standards.

Technical committees are constituted to be representative of valid national interests in the standardisation of products or processes. Membership is preferably on the basis of organisation, association or forum representation as opposed to an individual basis. An organisation can join a technical committee or subcommittee as a P-member or an O-member.

A P-member participates actively in the work, and has an obligation to respond to documents circulated for comment, voting or both, and to participate in and vote at meetings. An O-member follows the work as an

observer. Such a member will receive committee documents and may submit comments and participate in meetings, but may not vote.

The development of South African standards is funded by the state. StanSA acts as a facilitator in the development and maintenance of South African standards, and also as the publisher of standards.

Two options are available when considering a new national standard:

- Adopt without change an international or regional standard. This has the advantage that the resulting adopted standard is produced cheaply, quickly and easily. It might not, however, represent fully the needs and requirements of the South African market.
- Develop a South African standard containing at least some different requirements. This has the advantage that a more focused standard can be achieved that addresses local needs well. The development of such a standard is costly and time consuming, and the result of the process may well be a re-invention of the wheel.

Nearly all the information security related SANS standards are adopted from ISO/IEC International Standards or ISO/IEC Technical Reports.

The underlying principles of the preparation of national standards and other normative documents published by StanSA are described in SANS 1-1:2003 Standards for Standards, Part 1: The development of national standards and other normative documents.

A few of the relevant technical committees are listed below. The stated responsibilities have been taken from the scope of each committee as provided on its website.

- StanSA TC 71 (Information Technology) is responsible for standardisation and dissemination of information in the field of information technology and electronic data interchange.
- StanSA TC 74 (Communication Technology) is responsible for standardisation in the field of communication technology of consumer and professional electronics.
- StanSA TC 168 (Banking Sector), which is responsible for standardisation in the field of the banking, securities and other financial services.
- StanSA TC 176 (Quality Assurance and Quality Management Matters) is responsible for standardisation in the field of quality assurance and

quality management including generic quality management systems (QMS) and supporting technologies.

- StanSA TC 178 (Risk Management) is responsible for standardisation in the field of organisation wide risk management in accordance with good corporate governance and other risk management best practices.
- • StanSA TC 179 (Security Management) is responsible for standardisation of systematic approaches to security management in various fields.
- StanSA TC 5120.14 (Security) is responsible for standardisation in the field of security in terms of entrance control and the storage of valuables and the minimisation risk.

It may seem strange that the committees for Security Management and Security are also listed. However, it must be remembered that physical security also plays a role in information security.

8.1 Information Security Standards

SC 71F (Information Security) is a subcommittee of TC 71. This committee is responsible for Standardisation in the field of information security, including guides and codes of practice intended to assist organisations to develop security standards and effective security management practices, as well as specifications to support certification of companies as a means to promote confidence by other organisations and consumers.

The activities of StanSA SC 71F and its working groups correspond to those of ISO/IEC JTC 1 SC 27.

8.2 ICT Systems and Software Engineering Standards

SC 71C (ICT Systems and Software Engineering) is a subcommittee of TC 71, and is responsible for Standardisation in the field of systems and software engineering, excluding hardware. This may be expanded as standardisation of processes, supporting tools and supporting technologies for the engineering of software products and systems, and the development of a unified set of systems and software engineering standards widely accepted by the intended class of users.

The activities of StanSA SC 71C and its working groups correspond to those of ISO/IEC JTC 1 SC 7, although because of resource limitations, not all SC 7 working groups are always addressed.

9 HOW TO BECOME INVOLVED

StanSA participate in technical committees and subcommittees of ISO and the International Electrotechnical Commission (IEC). The views of local stakeholders are gathered through local technical committees and subcommittees, and conveyed to the appropriate international committees of ISO and IEC. This is an extremely important function, as it ensures that, wherever possible, local considerations are incorporated into international standards during their formation.

Because StanSA is committed, wherever possible, to adopting international standards for local use, it is vital that international standards accommodate the needs of local stakeholders. Also, through this participation, South Africa can influence the contents of international standards.

It is important that South African organisations participate in standards committee work, to ensure that their views are known.

Organisations wishing to be part of this exciting opportunity, should contact the SABS, or the author.

10 CONCLUSION

This paper only touched the tip of the iceberg where international security standards are concerned. For more information on ISO and SANS standards, a visit to the respective web sites are recommended. Also, ISO is only one of a number of international organisations developing standards or recommendations related to information security, for example the IEEE and ITU-T.

11 REFERENCES

SABS, Web page of the South African Bureau of Standards, http://www.sabs.co.za, Visited June

2007 ISO, Web page of the International Organisation for Standardisation, http://www.iso.org, Visited June 2007

ISO, My ISO Job - Guidance for delegates and experts, 2005 ISO, Joining In – Participating in International Standardization, 2007 Standards South Africa, SANS 1-1, Standards for standards - Part 1: The Development of National

Standards and other Normative documents, Edition 1, 2003 ISO, ISO/IEC Directives Part 1, Procedures for the Technical Work, Edition 5, 2004 ISO/IEC JTC1, ISO/IEC JTC 1/SC27 N5757, Directives, Edition 5, Version 3.0, 2007

EMERGING FRAMEWORK FOR THE EVALUATION OF OPEN SOURCE SECURITY TOOLS

E. Biermann & JC Mentz

Tshwane University of Technology

biermanne@tut.ac.za
012 382 4743
F'SATIE, Private Bag X680, Pretoria, 0001,
mentzjc@tut.ac.za
012 382 4312
Department Enterprise Applications, Private Bag X680, Pretoria , 0001

ABSTRACT

The drive from the South African Government towards the adoption of open source software across all platforms, incurred a number of research and development questions. The open source domain provides especially SMME's with options to implement high quality software that are financially viable. Although software costs is a major factor within providing proper working environments, specific security issues pertaining to open source needs to be addressed. With the opening of networks as well as the availability of information, companies need not only implement security policies, but also constantly upgrade implementations. The study of open source security issues as well as the actual evaluation of tools therefore becomes essential.

The purpose of this paper is to study the security issues within the open source environment and looking specifically at the use of security software originating from the open source domain. We provide details and results of surveys conducted around the adoption of security tools within South

African companies. The study leads to us proposing a emerging framework for the evaluation of open source security tools.

KEY WORDS

Open source software, security, framework, evaluation, tools.

EMERGING FRAMEWORK FOR THE EVALUATION OF OPEN SOURCE SECURITY TOOLS

1 INTRODUCTION

Interest in open source software (OSS) has grown significantly within South Africa (SA) during the last couple of years. This intensification is partly due to the drive from the SA government towards the adoption of OSS within both the Government and the private sector (FOSS, 2006).

Hoepman & Jacobs (2007) defines OSS as "*software for which the corresponding source is available for inspection, use, modification and redistribution by the user"*. Dimaio (2007) states that the change to OSS is beneficial in terms of cost implications; fast implementation time; tailored applications as well as providing a shortcut to technological independence. As with any new paradigm, disadvantages or challenges are also a reality. Challenges or problems that are present within the open source domain mainly evolve around security issues (Mookhey, 2004).

Industry and academia are divided into two main outlooks or groups when it comes to the security of open source software. On the one side are those stating that the openness of the code automatically decreases the security of the application or tool. This group states that the openness of the code leads to vulnerabilities being easier recognised and misused by attackers (Williams & Danahy, 2006; Hoepman & Jacobs, 2007). Attackers are also provided with a complete view of the product, including its vulnerabilities (Ford, 2007). Research conducted by Fortify (Chess, Lee & West, 2007) indicates that a poorly designed software built process may allow for an attacker to insert malicious code within the developed product. Any developer may contribute to OSS projects and no skills selection are required which may lead to un-secure code (Lawton, 2002). As no standardized quality control seems to be present within the development of OSS, this may result in the code not developed with security issues in mind

(Hoepman & Jacobs, 2007), or malicious code can be inserted within the developed product (Chess, Lee & West, 2007).

The second group beliefs that the publishing of the source code adds to providing more secured programs or applications. Arguments published include: the availability of source code means that there is complete disclosure on how a specific software or feature or section is implemented (Ford, 2007). Hoepman & Jacobs (2007) state the free distribution of source code allows for the independent evaluation of that specific software by external parties. Williams & Danahy (2006) point out that the first step in assuring whether applications are secure is to study the source code. This leads to the identification of security vulnerabilities, design flaws as well as policy violations. Security flaws are rectified faster as the open source development domain see the fixing of bugs as a major interest rather than developing new features or a new version (Lawton, 2002). The likelihood of patching bugs within the software increases within the OSS domain thus making it easier to repair holes (Hoepman & Jacobs, 2007). In the case that a vulnerability becomes known within the proprietary domain, the client is dependent on the specific vendor to develop and publish suitable patches or solutions. Within the open domain, this is however not the case (Ford, 2007). Finally Hoepman & Jacobs (2007) states that the distribution of the source code forces programmers to produce quality code, especially since it will be evaluated by a world-wide audience.

Arguments from both these camps hold value, and our focus is not in proving either. We rather set our focus on the utilisation and effectiveness of security tools developed within the open source domain. Evaluation results of open source security tools in terms of set standards and procedures seems lacking from the open source environment. Security experts within companies that need to implement security solutions have thus no means of determining which security tools are currently utilised effectively by companies. Also lacking is specific technical test results for open source security tools.

In this paper we set to determine the use of open source security tools within SA (specifically Gauteng) companies. This as well as an intensive study into open source security tools lead us to proposing an evaluation framework for open source security tools. This research is guided by the following question: in the quest for providing secure systems and networks,

what OSS tools are available and how can they be evaluated for the quality of protection they provide?

The paper is organised as follows: Section 2 describe current evaluation methods utilised to evaluate the usefulness of open source security tools; as well as a description of our evaluation process. Section 3 provides a categorization of security tools according to the results from the industry surveys. Section 4 details the evaluation criteria and framework while Section 5 portrays results from our partially implemented framework. The paper concludes within Section 6.

2 SECURITY TOOLS EVALUATION

The evaluation of security tools in both the open source and proprietary domains are done in a number of different ways. For example McGann & Sicker (2005) mentions that such tools need to be evaluated in terms of robustness, ease of use, documentation, usefulness and actual functionality. Actual functionality refers to whether claims made by the developer/s are valid. Wilander & Kamkar (2003) focuses on a specific category of tools and evaluating the category by simulating a range of possible attacks.

In the evaluation of security tools it is vital to determine the specific security category for which the tool is developed. In specifying this, the classified category can then be described in terms of minimum security features which the security tool need to satisfy. The evaluation of tools within a specific category can then be achieved by determining which security feature/s it adheres to. The institute for security and open methodologies (ISECOM) has developed an Open-Source Security Testing Methodology Manual (OSSTMM2 – see http://www.isecom.org/osstmm/) that describes a methodology for conducting security testing for organizations. The dimensions of the OSSTMM security testing process include visibility, access, trust, authentication, non-repudiation, confidentiality, privacy, authorisation, integrity, safety and alarm. In addition to this base list the software should also be tested in terms of its quality for example number of internal errors (Li *et al.*, 2006).

In order to work towards determining whether open source software with all its various security issues can be utilised effectively within the security tools environment, we have to follow a detailed process (see Figure 1).

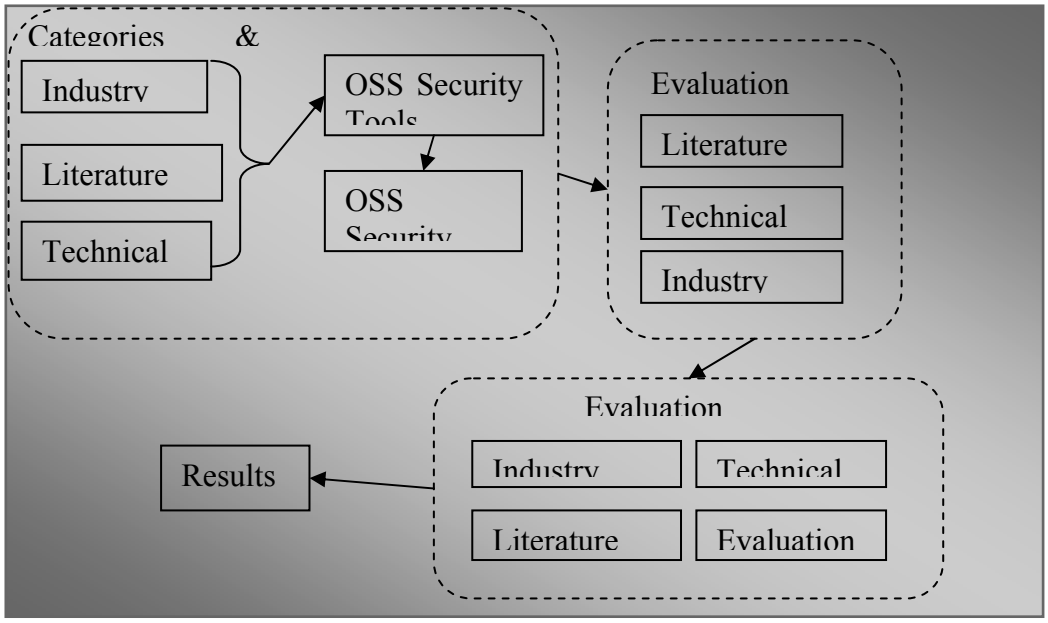


Figure 1: Evaluation Process

The process consists firstly of defining the different categories in which open source software tools can be classified. This is achieved by studying the outputs from industry surveys as well as a detailed literature and technical study. After establishing the categories we set to list the available tools within the different categories. In working towards setting an evaluation framework the aspects necessary to evaluate the different tools need to be defined. These aspects are used to firstly set the evaluation framework and finally evaluate the tools.

3 CATEGORIES & LIST OF TOOLS

Two preliminary surveys were conducted within the Gauteng region (SouthAfrica) during 2006 & 2007 targeting a total of 208 companies. The focus of these surveys was to determine amongst others the adoption of OSS security tools. A total of 192 companies responded and 47% of the 192 respondents indicate that they make use of open source security tools. 44% of these utilize OSS firewalls, 26% utilize OSS network monitoring tools and only 19% utilize OSS anti-malicious tools. A staggering 78% of the respondents within the survey regarded the security of Linux as inherently

adequate. Although this large percentage deems Linux as sufficient, 95% stated that they experience security threats on a daily basis.

Results from the industry surveys as well as a detailed literature and technical study were used to define the following categories of open source security tools within the network domain:

3.1 Category A: Scanning & Monitoring Tools

The aim of a scanning tool is to search and detect systems that have not been configured with security in mind or that have not implemented security patches for specific software as vulnerabilities are exposed (Mookhey, 2004). Monitoring tools refers to software that are utilized to continuously monitor networks in order to detect vulnerabilities in real-time. Tools within this category are further refined to include *port scanners, network vulnerability scanners, web vulnerability scanners, password vulnerability testers* and *network monitoring.*

Port scanners remotely scan a target host and determine open ports. These types of applications are normally lightweight and are a valuable tool in determining unattended server applications (Poole, 2003). Network vulnerability scanners are used to determine incorrect network configurations (Mookhey, 2004), while web vulnerability scanners automatically scan and evaluate web servers and web applications for possible vulnerabilities. Password crackers or password vulnerability testers are programs focused around guessing passwords and comparing them to an illegally obtained password file off-line (Poole, 2003). Network monitoring tools are software tools that are used to scan or track inside the boundaries of a network (Tomsho, Tittel & Johnson, 2003).

3.2 Category B: Analysis Tools

A network analyzer is a tool that enables a person to listen to network traffic that originated from or is destined to the local network. These types of programs are able to intercept and decode all traffic routed along a network as well as display the actual content. The content displayed is for example the IP addresses (source and destination); protocol type and the contents of each of the seven protocol layers (Poole, 2003). The main type of analysis tools is packet sniffers, which collects copies of network packets and analyzes them to provide information that can be used to diagnose and resolve networking issues (Whitman & Mattord, 2004).

3.3 Category C: Intrusion Detection & Prevention Tools

An intrusion detection system operates on the notion of a burglar alarm, activated upon detecting changes or violations within the network configuration. Different types of intrusion detection systems exist, namely network based that are used to protect network information assets as well as host-based to protect server or host information assets. Intrusion detection systems operate on either a signature-based or anomaly-based detection methods. Signature based systems establishes signatures of different attacks and threats; all network traffic is then compared against these signatures for possible attacks. Anomaly-based systems collect data from normal traffic and establish a base-line against which network traffic are compared (Whitman & Mattord, 2004).

3.4 Category D: Firewalls

A firewall is a device (hardware or software) that prevents specific type of information from moving between the un-trusted network (outside the organization) and the trusted network (inside the organization).The advancements in firewall technology has led to defining three different type of firewalls, referred to as generations (Whitman & Mattord, 2004). They are first generation (packet filtering firewalls); second generation (application level firewalls) and third generation (stateful inspection firewalls).

3.5 Category E: Anti-Malicious Tools

Malicious software increased tremendously with the introduction and opening of networks. Security specialists have to guard constantly against malicious code such as viruses, worms and Trojans. Anti-malicious software is the most utilized OSS security software and various tools exist.

3.6 Category F: Cryptography Tools

Cryptography tools refer to specific encryption and decryption tools used to protect data. Tools that can provide cryptographic functions range from example protecting specific communication sessions; encryption of files; encryption of hard disks as well as wireless sessions. It is therefore not feasible within this study to provide a list of specific tools per category of protection due to the vast array of cryptographic possibilities. We rather focused on providing three tools that are used mostly in providing general cryptographic functions.

The surveys as well as the literature and technical study led to an extensive list of OSS tools currently available and used. Due to size restrictions within this paper it is not possible to provide the list (the complete list is available from the authors).

4 EVALUATION FRAMEWORK

This section describes a two level framework to guide the selection of a short list of security tools for further research and evaluation. The first level of the framework deals with aspects related to the accessibility of tools and the second, more detailed level addresses the effectiveness of the tools. Each level consists of a set of criteria that address its intent. The framework is structured in this way to reflect a requirement of widespread use. Before any tool is eligible to be tested for its efficiency at protecting a system it would be necessary to be widely used first. The profile of the average user of open source software for the purpose of this research is therefore not limited to people with technically advanced computer skills. As such the general user is regarded as able to find, download, install and configure on the level of capability similar to general computer literacy.

4.1 Level 1

To fulfil the requirement that a security tool must be widely used, a number of aspects are identified. These include availability, version, platform, interface, download size, available documentation and support.

Availability: this aspect indicates the ease of acquiring the software. This does not only relate to well known web sites but also to the same piece of software being available on multiple download locations. High availability shows that the software is easy to get hold of and by implication an indicator of a large user base.

Version: software change over time and as the developer participation of a particular tools might be large the versions of the program can become confusing. This is seen especially in descriptions such as pre-release and release versions. In addition to this multiple versions with minor variations may exist. The importance of this indicator is that the user will be able to identify which versions are available and to make a choice between a stable or a developing version.

Platform: many different operating systems are available to computer users. A user interested in a tool must be able to identify the platform for

which it is designed. The wide spread use of a tool is also influenced by the same tool available for different platforms.

Interface: with the introduction of Windows 95 the computer user changed to a GUI user. Although a graphical user interface is the standard for program usage there is still the capability to use it from a command line which is predominately text based.

Download size: the size of the program will affect the choice made by the user. A very large file will take longer to download and cost more. At the same time a program that needs to be distributed by any other means than download (for example mailing a digital media such as a compact disk), might make the user think twice before choosing to use it.

Available documentation: this aspect is critical if the user is unfamiliar with the installation and configuration of the software. The more complete and easily available the documentation the wider it will be used.

Support: in conjunction with the documentation, support for a particular product is useful in the event that problems are experienced or additional information are required.

The application of the level 1 criteria resulted in a shortlist of tools that represent those most used for security purposes. The tools on this list has not been evaluated for the quality of security protection that it provides and for that purpose a second level of evaluation are required.

4.2 Level 2

The tool as an application forms an integral part of security and the methodology is instructive towards the compilation of a set of aspects for level 2 testing. The ability of the specific security tool to assist in the creation of a secure network forms the basis of level 2 testing. In specific the following aspects are included:

Functionality: this refers to the actual functionality of the tool in relation to claims made by the developer/s. The documentation from level 1 is analysed and it is determined whether the tool actually include the stated functionalities.

Protection: the protection ability is evaluated according to the category in which the tool is classified. Each category is defined by a set of minimum protection abilities and the security tool is tested according to these defined abilities.

Interoperability: security tools are generally created to only provide protection according to a set category. A requirement for such tools is its

ability to operate successfully with security tools from the same as well as other categories.

Usability: the ease of use as well as the usefulness of the tool. The level of difficulty to install, configure and maintain the tool is evaluated. Also included is to determine the existing need for such a specific tool.

Simulation: a simulated test bed or test environment is required in which current threats and attacks can be simulated. The security tool is then evaluated within this simulated environment for real-time attacks and vulnerabilities.

5 RESULTS

The results from the questionnaires showed a wide variety of tools in use. This is consistent with the milieu of the open source domain but makes the analysis and subsequent answer of the research question challenging. The evaluation of the extensive list of tools according to the first level of the framework was completed utilising a system of weights. For example a weight of 1 was assigned for each platform on which the tool can be implemented and 1 weight was assigned if telephonic support was available. This approach was followed to ensure that the tool that is most widely used in terms of level 1 criteria would make it to the short list of security tools that will be tested in the second level of the framework. Evaluating the extensive list of OSS security tools against the first level of our emerging framework led to the results displayed in Table 1.

Table 1: First level Evaluation Results

Category	Security Tool
A: Port Scanners	*Nmap -Network Mapper (http://insecure.org/nmap/):* Rapidly scan small and large networks. A number of tools also make use of the functionality provided by Nmap, for example XNmap and Nessus
	Angry IP Scanner (*http://www.angryziber.com/*: A very fast scanner that scans IP addresses in any range as well as their ports.
	Unicornscan (http://www.unicornscan.org): This tool provides an interface for introducing small stimuli and measuring the response from TCP/IP enabled devices or networks.
A: Network Vulnerability Scanners	*X-Scan (*http://www.xfocus.org*):* A general network vulnerabilities scanner that can be utilized to scan for network vulnerabilities by using a multi-threading method.
	SARA (http://www.www-arc.com/sara/): The Security Auditor's Research Assistant (SARA) is A third generation Unix-based security analysis tool. This scanner is derived from the famous SATAN (Security Administrators Tool for Analyzing Networks) and features extensive usability and auditing capabilities.
A: Web Vulnerability	*Nikto (http://www.cirt.net):* Performs comprehensive tests against web servers for multiple items, including over 3300 potentially dangerous files/CGIs.

Scanners	*Webscarab*(*http://www.owasp.org/index.php/OWASP_WebScarab_Project*): This tool analyses applications that are communicating via the HTTP and HTTPS protocols.
	Wikto (*http://www.sensepost.com/research/wikto/*): Built for the .NET 2 framework and contains a built-in web spider for directory discovery purposes.
A: Password Vulnerability Testers	*John the Ripper* (*http://www.openwall.com/john/*): A fast password cracker that is available for different UNIX distributions, Windows, DOS, BeOS and OpenVMS.
	Cain & Abel(http://www.oxid.it): A password recovery tool for MS Windows. It allows easy recovery of various types of passwords by sniffing the network.
	Ophcrack (*http://sourceforge.net/projects/ophcrack/*): A Windows password cracker based on a time-memory trade-off using rainbow tables.
A: Network Monitoring Tools	*Nagios* (*http://www.nagios.org*): A powerful network monitoring tool that are used for detecting specific network problems. Included in distributions such as Debian, Fedora and Suse.
	EtherApe(*http://sourceforge.net/project/showfiles.php?group_id=2712*): A graphical network monitoring tool, featuring Ethernet, IP, TCP, FDDI and Token Ring modes.
	WireShark(*http://www.wireshark.org*):Network protocol analyzer for UNIX, OS X and Windows. It allows for the examination of data from a live network or from a captured file on disk. WireShark was known as Ethereal up to 2006.
B: Packet Sniffers	*Snort* (*http://www.snort.org*): Performs real-time traffic analysis and packet logging on IP networks.
	Kismet (*http://www.kismetwireless.net/*): Wireless network detector and sniffer for 802.11 layer 2 wireless networks.
	TCPDump (*http://tcpdump.org*): An IP sniffer that requires few system resources. It is a specialized sniffer used for detecting network problems.
C: Intrusion Detection & Prevention	*Snare* -System Intrusion Analysis and Reporting Environment (*http://www.intersectalliance.com*): This is a series of log collection agents that facilitate centralized analysis of audit log data.
	OSSEC (*http://www.ossec.net*): This tool performs functions such as log analysis; integrity checking; time-based alerting as well as active responses.
	Tripwire (*http://sourceforge.net/projects/tripwire/*): A tool used by security administrators to determine the integrity of files as well as possible modifications or tampering of specific files is the file integrity scanner (Poole, 2003).
D: Firewalls	*Smoothwall Express* (*www.smoothwall.org*): Smoothwall includes traffic shaping, VPN capability as well as proxy and DHCP server capabilities.
	IPCop (*www.IPCop.org*): IPCop includes a whole range of services including traffic shaping on outgoing connections and a built-in DHCP and proxy server.
	Netdefender (*http://www.codeplex.com/netdefender/*): A specialized open source firewall. It operates by blocking all communication on specified ports after the rules have been setup.
E: Anti-Malicious Tools	Clam AV (*www.clamav.org*): : This application is specifically designed for scanning e-mail gateways for malicious code. Virus scans take place from the command line in a terminal window.
	ClamWin (*www.clamwin.com*): Windows version of the ClamAV engine. It seperates the processes of scanning for viruses on harddisk and scanning for viruses in program memory.
	Winpooch (*http://sourceforge.net/projects/winpooch/*): Windows watchdog that detects and monitors changes in the system and effectively blocks spyware.
F: Cryptography Tools	*GnuPG*(*http://www.gnupg.org*): *Gnu Privacy Guard* features a complete implementation of the OpenPGP standard and allows for the encryption and the digital signing of data and communication.

	OpenSSL (http://www.openssl.org): This tool implements the Secure Sockets Layer (SSL v2/v3) and Transport Layer Security (TLS v1) protocols.
	TrueCrypt (http://www.truecrypt.org): Disk encryption software that creates a virtual encrypted disk within a file and mounts it as a real disk. It is able to encrypt entire hard disk partitions or storage devices such as USB flash drives.

It has to be noted that an alarming number of open source tools become proprietary after a few versions of being publicly available (see for example Nessus at www.**nessus**.org). This means that once a company is utilising a specialised tool in providing a secure working environment, the possibility that newer versions will become available at an additional price is increasing.

Results displayed only reflects implementation of Level 1 of our proposed framework. Level 2 aspects is not finalised and the implementation is set to be completed in July 2008.

6 CONCLUSION

This paper addresses the research question by clearly identifying that a number of open source security tools are being used to protect systems. The application of the 1st level of the testing framework resulted in a shortlist of tools that can be identified as the most widely used security tools. This result is significant in that it shows the security awareness of open source developers and it suggests that open source operating systems might not be inherently as secure as is claimed. In addition the results indicate user security awareness.

Further research is in progress as to the evaluation of the quality of the protection that the shortlist of security tools provides to software systems. To that effect this paper highlighted the possible aspects for a second level of security tool testing. Finally the completed security evaluation framework will play a significant role in the protection of systems using open source tools by promoting a high quality of development as well as the development of a standard for the evaluation of such tools.

7 ACKNOWLEDGEMENTS

This research was fully funded by Eskom, Research & Innovation Department. They also fund the research and implementation of Level 2 of the framework.

8 REFERENCES

CHESS, B., LEE, FD. & WEST, J. 2007. Attacking the Build through Cross-Build Injection. [Online] Fortify Software. Available at: http://www.fortifysoftware.com

DIMAIO, A. 2007. Open Source Software in Government: How much open and how much source? Gartner Symposium ItxPO 2007.

FORD, R. 2007. Open vs. Closed: Which is more secure? *ACM Queue Magazine February*.

FOSS. 2006. Policy of Free and Open Source Software use for South African Government. [Online]. Available from: http://www.doc.gov.za/index.php?option=com_docman&task=doc_view&gid=49

HOEPMAN, J. & JACOBS, B. 2007. Increased Security Through Open Source. *Communications of the ACM*, January, Vol. 50, No.1.

LAWTON, G. 2002. Open Source Security: Opportunity or Oxymoron? *IEEE Xplore*, Volume 35, Issue 3, March 2002, Pg 18-21.

LI, Z., TAN, L., WANG, X., LU, S., ZHOU, Y., and ZHAI, C. 2006. Have things changed now?: an empirical study of bug characteristics in modern open source software. In *Proceedings of the 1st Workshop on Architectural and System Support For Improving Software Dependability*.

McGANN, S. & SICKER, D. 2005. An analysis of security threats and tools in SIP-based VoIP systems. *Presented at the 2nd Annual Workshop VoIP Security.,* Washington, DC, June.

MOOKHEY, K.K. 2004. Open Source Tools for Security and Control Assessment. *Information Systems Control Journal* , Volume 1.

POOLE, O. 2003. *Network Security: A practical guide.* Butterworth-Heinemann, Oxford. ISBN 0-7506-50338.

TOMSHO, G., TITTEL, E. & JOHNSON, D. 2003. *Guide to Networking Essentials*. Third Edition. Thomson Course Technology. ISBN 0-619-13087-3.

VIEGA, J. 2004. Open Source Security: Still a Myth. O'Reilly [Online]. Available from: http://www.oreilly.com

WHITMAN, M.E. & MATTORD, H.J. 2004. *Management of Information Security.* Thomson Course Technology. ISBN 0-619-21515-1.

WILANDER, J. & KAMKAR, M. 2003. A comparison of publicly available tools for dynamic buffer overflow prevention. In *Proceedings of the 10th Annual Network and Distributed Systems Security Symposium.*

WILLIAMS, J. & DANAHY, J. 2006. "Opening the black box" A Source Code Security Analysis Case Study. Aspect Security Inc. & Ounce Labs Inc.

SOCIAL ASPECTS OF INFORMATION SECURITY

Evangelos D. Frangopoulos[1 *], Mariki M. Eloff [1 **], Lucas M. Venter[1 **]

[1] School of Computing, University of South Africa (UNISA).
[*] 215, Alexandras Avenue, Athens, GR 11523, Greece.
Tel./fax: +30 210 6428-483. eMail: vfrangopoulos@hol.gr
[**] TvW 8 Theo van Wijk Building, UNISA, Pretoria, South Africa.
Tel.: +27 12 429-6368. eMail: {ventelm,eloffmm}@unisa.ac.za

ABSTRACT
Social Engineering (SE) threats have constituted a reality for Information Technology (IT) systems for many years. Yet, even the latest editions of the generally accepted Information Security (IS) standards and best practices directives do not effectively address the Social Engineering aspect of IS defences.

SE attacks target the human element of IS by exploiting human relations to the maximum possible extent. The social relations between interacting individuals who are involved in an Information Security Management System (ISMS) structure, combined with the frequently unpredictable fashion that humans act and react to stimuli, provide opportunities that Social Engineers may and do exploit. In the ongoing effort against Social Engineering attacks, if the social elements of IS are ignored, fallacious working assumptions may be made. These inadvertently result in the creation of insufficient controls against identified SE threats. Hence, simply put, Information Security scientists can no longer afford to ignore the nature of the social structures that govern all aspects of human relations, and in particular those that lie within the context of an ISMS.

This paper attempts to strengthen the pursued research on SE threat identification and control, by applying sociological principles to IT and ISMSs, thus bringing into the light their nature as social structures. This constitutes part of a larger effort by the authors to systematically identify and subsequently cater for SE threats to IS, in the context of which the social foundations of IS are examined.

KEY WORDS
Information Security, social aspects, social engineering, ISMS, objective reality, subjective reality, Actor-Network Theory, black box

SOCIAL ASPECTS OF INFORMATION SECURITY

1 INTRODUCTION

Social Engineers are frequently successful in exploiting the social relations between the individuals who operate within the context of an Information Security Management System (ISMS) structure, aided by the sometimes unpredictable fashion that humans act and react to stimuli. Although great effort has been invested in forming Information Security (IS) standards and procedures, these, so far, prove inadequately equipped to ensure Information Security against Social Engineering (SE) attacks. It is stipulated that the design flaws do not result from the standards' structures being technically incomplete. Despite being complete from a technical viewpoint, Information Security standards do not encompass provisions for the idiosyncratic nature of the human element, especially within a social context. By providing some insight on the social mechanisms at work in the development and function of an ISMS, certain design flaws of the related standards and procedures may be brought to light and steps be taken towards rectifying them.

The average person's notion of *Information* Security stems from the general idea of Security. Security in general, on the other hand, has been traditionally related to the police, law enforcement, the military etc. In many modern languages, even the word for "security" is used to signify the police force in general or one of their main branches dealing with public safety. Furthermore, whenever and wherever it was needed, security has always been applied in a stern, bureaucratic way, actually taking advantage of bureaucracy and the hierarchical structures associated with it. By using such hierarchical structures, the application of security is achieved through regulation and control (Foucault, 1989, p.65). This mentality is accurately expressed in the age-old saying: "*To trust is good but to control is better*". The idea of security has been applied to material and immaterial issues alike since the birth of the first human societies. Be it the protection of gathered sustenance supplies and, later, capital (material) or the protection of information and even life itself (immaterial), security against the ever-present foe has been one of our most basic needs. As the bureaucratic

application of security has constituted standard practice for a long time, long before the arrival of the computer, it was the obvious step forward to achieve the security of (non-computerised) information in the same way. Furthermore, with the evolution of computer systems as information-handling devices, the existing principle was simply extended to include Information Technology (IT) Security by adding more appropriate controls.

It can thus be safely deduced that any modern ISMS implementation still relies on bureaucracy for its fundamental functions. It could even be argued that a bureaucratic structure through which regulation and control are applied, is a necessary pre-requisite for an ISMS to exist, on the assumption that the imposed technical and physical controls can mitigate all identified risks. However, it must be stressed that the current bureaucratic system was conceived, defined and described by Max Weber in the late 19th and early 20th centuries and still functions along the prescribed way (Bottomore, 1990, p. 203). This, in principle, should constitute an indisputable oxymoron as the futility of attempting to secure Information in the 21st century by using 19th century models and tools is obvious. Consequently, the controls existing within this context may prove inadequate in today's terms.

In the following sections of this paper an attempt is made to first establish the ISMS as a social construct and then analyse it by applying traditional sociological principles to it. This is followed by the application of Actor-Network Theory (ANT) principles to the ISMS, in an effort to better identify those social aspects of IS that may help significantly the ongoing effort against SE attacks. In particular, section 2 discusses the current ISMS practices from a modernist viewpoint. In Section 3 the ISMS as a social construct is investigated. Sections 4 and 5 discuss the Objective and Subjective realities of the ISMS. Sections 6,7 and 8 approach ISMSs from an Actor-Network Theory viewpoint. Section 9 examines Powerplay within the ISMS and, finally, the concluding remarks are given in section 10.

2 CURRENT PRACTICE - THE MODERNIST APPROACH TO THE ISMS

Information Systems are designed and built in a purely deterministic fashion. They are created to bring order to organisations by forcing human actions to take place within the strict context and limits of ordered workflow

implementations. Such strict implementations ensure that human actions are disciplined and unambiguous and that the results of those actions are predictable, clear-cut and exact and, if necessary, securely leading to further pre-defined actions.

In transcribing the processes of the analogue world into workflows for computer-based Information Systems, all uncertainty must be eradicated. The tools of the trade for such an accomplishment are business process analysis, flowcharts and, of course, Boolean logic. In this way, all processes and user actions are transcribed into algorithmic sequences of exact questions strictly requiring unequivocal "yes/no" replies.

All of the above ideally lead to the design and implementation of an Information System which has all ambiguity removed from it and is no more and no less than a finite-state system. All state transitions must be fully reproducible and all user actions must be clear and exact. Such an implementation would thus lead to business practices that are also clear, exact and deprived of all ambiguity. (The feasibility of such a system is unimportant for the present discussion).

As the Information Security Management System must form an integral part of the Information System, the above notions are extended to cover Information Security Management as well. The ISMS is thus covered by the same providence and governed by the same principles described above.

Stemming from the concept of Reason as this was set forth during Enlightenment (Mendelssohn et al., 1989, p.28), rational knowledge is assumed to possess an objective existence which is independent of the observer's posture. This forms the basis of Modernism (Deligiorgi, 1996, p. 18) which builds intellectual structures on rational knowledge and through these promotes innovation and progress. In the context of Modernism, the complexity of intellectual structures is anything but limited as even large-scale processes can be described through modernistic methods and principles.

Indubitably, Modernism has actually been the motive power behind the industrial revolution that resulted in modern technology. Information Technology is clearly modernistic as its very nature requires the observer to be detached from the system being observed. In their inspired paper, Low et al. (1996) argue that software engineering is at present solely viewed from a modernistic perspective. This principle can easily be expanded to

encompass the whole of the Information Technology construct. IT Systems are thus confronted as objective entities that are exact, discreet, identifiable, predictable and independent from the observer.

This leads to Information Systems being viewed as machines that function in a precise, repeatable and predictable way.

Gareth Morgan, in his book "Images of Organization" (1996), discusses a number of ways to view and understand organisations which he calls "Metaphors". The first of these metaphors calls for the organisation to be viewed as a machine with interchangeable components, which is firmly set on a goal. According to this metaphor, human and technological components form a stable machine that operates in a repetitive, predictable and secure way. This is achieved by having rational actors make rational decisions with predictable, reproducible and unambiguous effects in a purely modernistic fashion.

For such a system to function, everything must fall in its place in a larger, well-described framework. Such a framework can only be created by the existence of processes that are governed by standardisation, control and regulation. The interlocking components of the machine are thus combined together according to a complex blueprint and their roles in the machine are fully prescribed.

Hence, all systemic issues are addressed in a default manner within procedures that result from the application of current analysis and design techniques to any IT-related project. Tools and techniques used in system analysis, such as top-down or bottom-up design methods, data-flow diagram methodologies etc (Schach, 2005; Whitten & Bentley, 2007) fully comply with the modernist approach. It must also be noted that all of the above are governed by strict standards leading to normalisation and making control, regulation and evaluation possible.

Furthermore, as businesses and organisations do not just rely on their IT department for number-crunching but are instead built around a skeleton and nervous system formed by that department, it is not unusual for global change and business process re-organisation to initiate within the IT department. The reason for such a decision is that IT is the one centre of operations that is de facto regulated and aligned to processes governed by standards, thus forming a solid and flexible platform to build upon. Information Systems thus tend to dictate the way that an organisation evolves and govern its responses to the ever-changing business demands.

To drive the above points home, one only has to consider the various issues that lead to successful Information Security management by today's standards:

Use of rules and regulations aiming to provide a secure environment.
Commitment of everyone involved to a set of prescribed guidelines or policy. This in effect constitutes behaviour control.
Use of technical measures for controlling the application of (a) and the upholding of (b) above.
Use of non-technical measures to complement (c) above.
De facto existence of a technocratic elite of Information Security professionals that oversees the application of (a), (b), (c) and (d) above.

On closer inspection, the above list reveals three important issues:

First, the above points are by definition dealt with in ISO/IEC standards 17799:2005 (ISO/IEC, 2005a) -corrected and renumbered in July 2007 as 27002:2005 (ISO/IEC, 2005f)- and 27001:2005 (ISO/IEC, 2005b). This proves the modernist character of these standards which may well be inadequate for today's challenges.
Second, the above five points and perhaps more significantly point (e) show that an ISMS is indeed a social construct that has to be examined in detail.
Third, as a whole, points (a) to (e) above form the modernist blueprint for an organisation viewed as a well-oiled machine according to Morgan's (1996) metaphor of "*organisation as machine*" discussed earlier. Furthermore, these points conform to bureaucratic definitions as presented by Max Weber a century ago. Max Weber is assumed to have written "*Wirtschaft und Gesellschaft*" (Economy and Society) between 1910 and 1914. This work was first published around 1922, after the author's death in 1920 (Oakes, 1998) and has watermarked all organisational efforts ever since. Using the translation -obtained from L. Ridener's (1999) website- for "*Wirtschaft und Gesellschaft*" (part III, chap. 6, pp. 650-78), the first of the characteristics of bureaucracy is described as:

I. There is the principle of fixed and official jurisdictional areas, which are generally ordered by rules, that is, by laws or administrative regulations.
1. The regular activities required for the purposes of the bureaucratically governed structure are distributed in a fixed way as official duties.
2. The authority to give the commands required for the discharge of these duties is distributed in a stable way and is strictly delimited by rules

concerning the coercive means, physical, sacerdotal, or otherwise, which may be placed at the disposal of officials.
3. Methodical provision is made for the regular and continuous fulfilment of these duties and for the execution of the corresponding rights; only persons who have the generally regulated qualifications to serve are employed.

As ISMSs currently adopt the above principles, their nature becomes fundamentally bureaucratic, thus causing a deficiency in the level of democratic processes within the organisation structure that are deemed necessary by prevailing trends in management. Bureaucracy pre-supposes strict hierarchical structures of a vertical nature while, today, the push is towards flat, horizontal organisational structures, the governing principles of which were described by Ostroff and Smith (1992).

According to Dhillon and Backhouse (2000), the fast progress of the electronic age and the evolution of IT have caused the emergence of new organisational structures. Consequently, the traditional hierarchical organisations are being transformed into loosely coupled networks that are characterised by co-operation on a horizontal level rather than hierarchical control in a vertical direction. As a result, direct interpersonal and inter-organisational communication, connectivity and the sharing of information have seriously augmented in volume compared to the time when the traditional organisational models based on hierarchy were solidly and exclusively in place.

Hence, the inadequacies of the current bureaucratically-built ISMS are bound to create opportunities for social engineers to thrive in. The assumption that all members of an organisation will play their ISMS-prescribed roles flawlessly during an attack, because of bureaucratic pressure, is wildly optimistic at best. Furthermore, bureaucracy may even hinder essential practices such as reporting of security-related incidents. This will come as a direct result of the inconvenience caused to the person reporting the incident by necessary paperwork etc.

3 THE ISMS AS A SOCIAL CONSTRUCT

Bruno Latour, in his two books, "Science in Action" (1987) and "Laboratory Life" (1986), among other things discusses how Science and Technology affect social constructs and how they are in turn affected by them. This strengthens the idea that all systems that are based on science and/or technology constitute social constructs and should be treated as such. An

ISMS, comprising both human as well as technological components, is thus indeed socially constructed.

In their book "The Social Construction of Reality", which was first published in 1966, Berger and Luckmann (1991) provided one of the definitive works on Social Constructionism. The functionalist interpretations presented by Berger and Luckmann can be readily applied to the ISMS structure in an effort to analyse and understand the social construction of such systems, as has been attempted by Albrechtsen (2004).

Although it may sound oversimplified, for the purposes of this analysis it suffices to concentrate on the discussion of Berger and Luckmann on the dual nature of societal objective and subjective reality. The notion of **Objective reality** concerns the production and maintenance of a shared sense of reality. This reality is ultimately constructed through the processes of externalisation, habitualisation, institutionalisation and legitimation. On the other hand, **Subjective reality** according to Berger and Luckmann (1991, p.167) differs from objective reality in the sense that it refers to the reality "*as apprehended in the individual consciousness rather than on reality as institutionally defined*". In other words, subjective reality is the sense of the socially created objective reality that each individual human being acquires as its own (internalises). This acquisition takes place mainly through the process of secondary socialisation.

Through the application of Burger and Luckmann's principles to ISMS structures, some of the system's inherent shortcomings can be identified and perhaps catered for. In this sense it was decided to follow the same structure as the one adopted in Berger and Luckmann's (1991) book, in order to properly present the application of their principles to ISMSs.

Thus, the social construct of the ISMS as an objective reality and then as a subjective one, according to Burger and Luckmann's work, will be discussed in the next two sections.

4 THE OBJECTIVE REALITY OF THE ISMS: EXTERNALISATION, HABITUALISATION, INSTITUTIONALISATION AND LEGITIMATION

The first step in the social construction of Information Security objective reality is that of externalisation. **Externalisation**, is defined in (Berger & Luckmann, 1991, p.70): "*Human being is impossible in a closed sphere of quiescent interiority. Human being must ongoingly externalize itself in*

activity". Externalisation as such, is an anthropological necessity originating from human biological pre-disposition. Human beings must continually externalise themselves through activity. Furthermore, (Berger & Luckmann, 1991, p.122): "*As man externalizes himself, he constructs the world into which he externalizes himself. In the process of externalization, he projects his own meanings into reality.*" The inherent instability of the human organism makes it imperative that humans produce for themselves a consistent and stable environment for conduct and social order in general. It is exactly such a need that is covered by the creation of an ISMS. Externalisation with respect to ISMSs has taken place through the evolution of the notion of security and the measures that are taken for ensuring it in general. As the threats particular to Information Systems were identified, it became obvious that if left uncontrolled, these threats would result in Information System chaos and disarray. As a result, action against the threats was taken by appropriate controls being applied etc. Hence, a computer user who decides to turn off and secure a PC when unattended, to set up password protection of files and systems or to make backup copies of a day's work is actually externalising.

According to Berger and Luckmann (1991, p.70), **Habitualisation** denotes the principle that "any action that is repeated frequently becomes cast into a pattern, which can then be reproduced with an economy of effort and which, ipso facto, is apprehended by its performer as that pattern". Human actions have an innate tendency to habitualise. Hence, all the actions that are taking place as a result of Externalisation with respect to ISMSs, eventually fall into a pattern that helps the individual go automatically through the motions necessary to apply essential controls. Thus, the simple examples of actions described above, after a certain point in time, are carried out as a matter of course. The user who free-mindedly decided to go through these motions, having established that these are good and effective things to do against data loss or compromise, incorporates them into a daily routine. This way, the necessity of such actions does not have to be re-examined every time they are carried out.

Habitualisation is the first and necessary step towards **Institutionalisation**. As can be found in Berger and Luckmann's work (1991, p.72), Institutionalisation "*occurs whenever there is a reciprocal typification of habitualised action*". They further go on to state that "*any such typification is an institution*" and that "*the institution posits that*

actions of type X will be performed by actors of type X". Finally they claim that "*institutions further imply historicity and control*.". Habitualised actions regarding social relationships form the basis for the creation of institutions that in turn enforce action. The interesting turn takes place as the established institution is "*objectified*" by bequeathing it to the subsequent generation that did not invent it initially. For the new generation, this socially created institution appears as a fully objective reality and, as such, is taken for granted. This is why Institutions always have a history, of which they are the products. "*It is impossible to understand an institution adequately without an understanding of the historical process in which it was produced*" (Berger & Luckmann, 1991, p.72). Institutions thus, by definition, control human conduct by setting up predefined patterns thereof. Shifting back to the ISMS paradigm, Institutionalisation takes place when the actions of individual user(s) like the ones described above, give rise to and become parts of an Information Security Policy.

Legitimation is defined (Berger & Luckmann, 1991, p.110) as "a 'second-order' objectivation of meaning. Legitimation produces new meanings that serve to integrate the meanings already attached to disparate institutional processes". The purpose of legitimation is to explain and validate the existing institutions. This is an important process if the presence of institutions is to be seen by individuals as subjectively plausible. If this is achieved, then the institutions themselves become acceptable. Legitimation is viewed as a 'second-order' objectivation in juxtaposition to the 'first-order' objectivation. 'First order' objectivation denotes the process by which principal meanings are attached to the institutional directives themselves. Legitimation is thus a 'second order objectivation' process in the sense that through it, the institutional directives are explained and justified via the application of cognitive and normative elements. This means that through legitimation actors are told not only how things should be done but also why it should be so and what things are in the first place. In this sense, legitimation provides a balanced combination of knowledge and values. Legitimation in ISMS comes in the form of Information Security standards and guidelines. IS standards such as the prevailing ISO/IEC 27002 (ISO/IEC, 2005f), 27001 (ISO/IEC, 2005b), 13335 (ISO/IEC, 1997; 1998; 2000; 2001; 2004), 15408 (ISO/IEC, 2005c; 2005d; 2005e) and the like, by means of their existence, legitimise the institutional directives of IS. It must be highlighted though, that IS standards effectively incorporate a high level

of formalism in IS management, at the same time bringing forth its bureaucratic nature that is largely based on control and regulation.

In order to better demonstrate how the creation of the social construct of the ISMS as an objective reality is effected, one may consider that aspect of an ISMS that deals with the protection of data against loss or corruption: the creation of backup copies of data.

Making copies of important documents must be as old as writing itself. It is at least as old as the ancient Egyptian civilisation. The fact that surviving hieroglyphics have been identified as copies of important legal manuscripts (University College London, 2003), shows that by making copies for safekeeping, the Ancient Egyptians externalised themselves by taking positive action against whatever they perceived as a threat that might result in the loss of important information. Quite interestingly, the control they created, i.e., making copies, has been very effective, as we are still able to obtain the data the ancient scribes tried to preserve thousands of years earlier. This externalisation, has changed in form over the millennia: During the middle ages it was monks who preserved whatever information they saw fit to preserve, by making elaborate, hand-written copies of it, and, later, typography made the production of copies even easier. However, in essence, the action of making copies of important information has always been the result of the same principal externalisation. It is exactly this externalisation that gave rise to the action taken by today's PC users, that of making backup copies of their computer data. The only difference from ancient times is that these data backups are now stored in electronic form.

In the given context, Habitualisation is portrayed by the fact that the need for data backup is never challenged. Data backup is nowadays considered necessary by any type of user and in any type of data processing system. Computer users routinely create backup copies of their data, and even those who don't, know that they should. Furthermore, computer systems can be programmed to automatically perform these routines with minimal intervention by the user. Again, not only these automated procedures are never challenged, but even more so, it is inconceivable not to incorporate such procedures in a system.

Subsequently, such actions and procedures are Institutionalised by being incorporated in an Information Security Policy. No Information Security Policy is complete without a section on data backup procedures. By being incorporated in an Information Security Policy, the data backup

procedures -not just the principle of data backup- become part of the subsequent user generation's objective reality that is taken for granted and as such remains unchallenged.

Finally, by explaining and validating the institutionalised procedures in an IS standard or set of recommended practices, Legitimation occurs and the social construct of the creation of backup copies is complete.

By expanding the above example to cover all aspects involved in an ISMS, the socially constructed objective reality of the ISMS is effected.

5 THE SUBJECTIVE REALITY OF THE ISMS

As it has already been discussed, Subjective reality is that "version" of objective reality that is internalised by individuals through secondary socialisation. Berger and Luckmann (1991, p.150) define socialisation in general as "*the comprehensive and consistent induction of an individual into the objective world of a society or a sector of it*". Primary socialisation takes place during childhood. It is the process through which people first become members of society. Secondary socialisation is "*any subsequent process that inducts an already socialised individual into new sectors of the objective world of his society*". This is effectively the process of internalising institutional directives. Within this process, an individual acquires behaviours and knowledge that are specific to the role the individual is called to assume within the society. A typical example of secondary socialisation is the educational process.

To shift the notion of the subjective reality into the context of ISMSs, it must be first considered that the socially constructed objective reality of an ISMS has evolved from existing objective realities in the pre-computer era and the relevant security efforts. As such, it relies heavily on a bureaucratic infrastructure and in turn offers a number of Information Security solutions. The ISMS objective reality is internalised as a subjective reality by all those who actually follow the offered Information Security solutions. "Those who follow the offered solutions" can be identified as three major groups in any type of organisation: a) the Information Security professionals who are responsible for carrying out the ISMS development, design, evaluation, maintenance and operation, b) the Management and c) the end-users.

The three identified groups have differences in interests, perspectives, goals and agendas. It is these differences that warrant the division into

groups. The segregation of the three groups is more important than it may be assessed at first, as it severely affects the secondary socialisation process and the way subjective ISMS reality is internalised by each group. As Berger and Luckmann put it (1991, p.158): "*Secondary socialisation requires the acquisition of role-specific vocabularies, which means, for one thing, the internalisation of semantic fields structuring routine interpretations and conduct within an institutional area*". Hence, different roles result in (or require) different role-specific vocabularies and may lead into a lack of common ground that the three groups can share. This, in turn, inhibits communication and co-operation between the groups. Berger and Luckmann (1991, p.158) give a good (and frequently adopted) example to clarify the point: "*a differentiation may arise between foot soldiers and cavalry*". In that example, the cavalry have their own language and employ their own methods for achieving their goal that the foot soldiers do not comprehend, as they don't need to. However, the foot soldiers have every confidence in the cavalry's actions that always get them out of a dire position.

In the case of the three groups involved in an ISMS structure (IS professionals, Management and End-users) the case is quite similar. Bearing in mind that in most cases the group of IS professionals is a subgroup of the organisation's IT professionals or a group that has evolved from IT, the Management rarely fully understands what the IS professionals do and how they do it. Nevertheless, management trusts the IS professionals with the 'crown jewels' of the organisation. Furthermore they assume that the IS professionals will keep the end-users in check with respect to Information Security. Again, management has a rather vague notion on how this is accomplished, generally assuming that technological measures applied by the IS professionals will do their work for them. Thus, it is not unusual for the IS professionals to be under-powered to carry out their work.

The disparity between the subjective reality internalised by the two groups, creates a serious gap of understanding between them with respect to IS. On the other hand, the end-users group view the Management group with respect to IS as being very remote and detached from practical issues, feeling that it is they, the end-users, that are overburdened by security measures and who are also frowned upon when something goes wrong. The end-users also view the IS professionals with scepticism, more-or-less as a 'necessary evil'. Although the end users do place their confidence in the IS

professionals' abilities to help avoid disaster or rectify situations that have gone astray, they also view them as 'techno-mages' performing black art and not doing any 'real' work within the organisation, as the product of their work is neither always tangible nor consistent in volume. Sometimes, the IS professionals are compared to a cruise-ship's doctor who is not busy unless a crisis situation brews. The doctor is certainly not needed every hour of every day on the ship but when the need arises, it is absolutely essential that he is present. Again, mentality gaps with respect to IS are created between End-users and IS professionals as well as End-users and Management. Lastly, in the case of the IS professionals' group, the situation is also quite complicated. Sometimes there is a tendency to deal with Management on a competitive basis, always struggling for more of the power that is in principle denied to them. If that is not the case, there is always the case of differing mentalities as management officials view the world under a different light compared to computer engineers and scientists who usually fill the ranks of IS professionals.

To further aggravate things, when IS professionals have to deal with the inability and, worse, with the reluctance of members of the other groups to internalise the ISMS objective reality in the same sense as they do, the IS professionals may develop a tendency to dispraise the other groups as conglomerations of technologically ignorant people. The gap in the internalisation of the ISMS reality is thus enlarged and the common effort towards the mitigation of IS threats becomes even more difficult to achieve. (It is interesting at this point to note that what is described by Leiwo and Heikkuri (1998) as an ethical divide between hackers and IS personnel is really also a result of the differences in the two groups' subjective realities).

All in all, the above analysis provides the theoretical justification of what is being described as "*lack of IS culture*" in organisations. What is lacking though, is not IS culture per se but the common internalisation of the objective reality regarding IS. The push towards "*holistic*" security is based on the creation of such a common ground that is necessary to advance understanding and co-operation between the organisation's groups towards attaining the required level of IS. By attempting to establish an IS culture, what is in effect being done is moving towards bringing together the naturally diverging agendas towards IS of the different groups. This, though, can not be attained by simply bringing each of the groups to the same level of expertise that each of the other groups has attained in their respective

fields. That would be a futile exercise, as experience is not easily or efficiently transferable.

As we currently stand though, differences between the groups within an organisation remain very severe and the main problem lies with the fact that each group can not identify with the methods and tactics imposed by the other group(s) with respect to IS. As IS professionals are responsible for IS within the organisation, they are the ones who set the pace by defining the essential directives and practices. The other groups although in theory are bound to follow the IS directives (top-level management commitment to the security policy is essential as is strict control of end-user compliance), in practice they usually fail to do so.

It is exactly this difficulty in common acceptance and internalisation of the security effort by all members of an organisation that creates innumerable security holes and provides social engineers with the opportunity for successful attacks.

6 ACTOR-NETWORK THEORY AND THE ISMS

In his "Science on Action", Bruno Latour (1987) brings forth the "Actor-Network Theory" (ANT) and in "Reassembling the Social", Latour (2005) redefines the notion of "the Social" and provides a fresh view of ANT as the "*sociology of associations*". ANT, considered as a subset of Social Constructionism, originated in the field of science studies. It is described as a 'material-semiotic' method used to map relations that occur simultaneously between people and/or objects (hence its 'material' nature) and between immaterial concepts (thus 'semiotic'). As a result, any system in the context of which the interactions between people, their ideas and their technological tools involve simultaneous material and semiotic relations, forms a single "*network*" for the purposes of ANT. The banking system is traditionally used as an obvious example to demonstrate a typical ANT network. Even everyday activities like driving to work every morning can be examined under the light of ANT. The network in that case comprises people, their behaviour on the road, their cars, the road network, the traffic regulations, the Highway Code and the interactions between all of those components.

In the Information Technology sector in general and in ISMSs in particular, interactive relationships exist between the management, IS professionals, end-users, technological solutions, equipment, security policy, bureaucracy, administrative practices and the experiences,

behaviours and ambitions of all individuals involved. Therefore, the ISMS makes a prime subject for study from the ANT viewpoint. Tatnall & Gilding (1999) and Albrechtsen (2004) present strong cases for examination through ANT of Information Systems Research and Information Security Management respectively. Their arguments certainly hold true for the particular case of ISMSs under examination in the context of this work.

Latour's view of the world as a network of "*actants*" (human and non-human actors) connected by complex links and relations, makes ANT useful in examining the reasons behind the success or failure of systems, technologies, scientific theories and social endeavours, as the direct result of changes in their network integrity. ANT does not give answers to the question of why a network is formed in a particular fashion. It is rather a tool for examining how actor-networks get formed and subsequently either hold their form and integrity or fall apart. In ANT, one of the central issues is the study of the forces that hold the network together.

In the interest of clarity, a few points must be clarified before attempting to apply ANT to ISMSs regarding "*actors*" and the notions of "*black boxes*", "*inscription*" and "*translation*".

"*Actors*" are, first of all, assumed to lie within the network of relations. Second, all actors are assumed to be shaped through their relations with one another. Third, it is assumed that there is no difference in the abilities of actors, irrespective of their form, nature or function. Fourth, as soon as an actor engages with an actor-network it too becomes part of that network and is actively introduced in the network's web of links and relations.

"*Black boxes*" are used by Latour (1987) to describe an entity (material or immaterial, human or non-human etc) that has been thoroughly dealt with, examined and transcribed into a particular known function where the output is a direct and predictable result of its input. If x and y denote input and output respectively, a black box can be seen as the function $y = f(x)$. These black boxes can represent various constructs such as a) the actions of users in an Information System, b) a known and generally accepted theory or practice, c) applied technologies etc. Hence, actors in an ANT network can be considered as black boxes and whole networks can also be black-boxed and viewed as entities with specific input/output transfer functions. When "opening up" such a black-boxed network, it can be viewed as a collection of other, smaller black boxes interconnected to

and interacting with one another. This notion helps both in employing a divide-and-conquer approach to dealing with ANT networks, as well as explaining the tendency of taking things "for granted".

"*Inscription*", according to Hanseth and Monteiro (1998, ch.6), "*refers to the way technical artefacts embody patterns of use".* In the same work, they also quote Akrich (1992, p.205) who makes the following statement regarding inscription: "*Technical objects thus simultaneously embody and measure a set of relations between heterogeneous elements*". Hence, Inscription is the process through which a 'pattern of use' or 'action' is coded or embedded in an artefact. However, this does not necessarily signify a strictly deterministic process. Artefacts can either be seen as "*determining their use*" or, on the contrary, be "*flexibly interpreted and appropriated*" (Hanseth & Monteiro, 1998, ch.6). Thus, inscription can be seen as the process through which, the designer's expectations including the desired form of future 'patterns of use' or 'actions' are involved in the development and use of the technology that is expected to enforce them. At the same time though, a feedback path exists as this technology definitively contributes in shaping the designer's expectations.

Insofar "*Translation*" is concerned, Latour (1987) postulates that in the context of ANT, stability and social order are dynamically and continually negotiated as a social process of aligning interests. This is achieved through "*translation*". According to Law (1992, p.366) translation "*generates ordering effects such as devices, agents, institutions, or organisations*". In simpler terms, according to Singleton and Michael (1993), translation is "*the means by which one entity gives a role to others*". Furthermore, in the context of Information Systems, "*In ANT terms, design is translation*" according to Hanseth and Monteiro (1998, ch.6), who go on to explain that interests of all actors involved in the network are translated into specific "needs" according to typical ideal models. Furthermore, the specific needs are translated into more general and unified needs that, through further translation, result into one, all-encompassing solution/system. When the solution/system enters production mode, it becomes adopted by the involved individuals by translating the solution/system into the context of their specific roles.

Translation is of paramount importance to the well being of ANT networks, as through the process of translation, the integrity of the network is maintained. This is achieved by the perpetual occurrence of translations

along links, in order to maintain the network's functionality and thus ensure its success. As translations along the links pre-suppose communication among actors, the overall process of translation and communication leads to power relations among human and non-human actors. ANT is thus perfectly equipped to deal with power relations in ISMSs, something that can not be efficiently done using the frameworks discussed so far. This ISMS 'Powerplay' will be later discussed in detail.

7 BLACK BOXES IN THE ISMS

ISMSs are full of black boxes. This is primarily done in an attempt to break large and complex problems into smaller, more manageable morsels. Through the process of dealing separately with every individual vulnerability, devising an appropriate control for it and including this as a solution in the ISMS, the vulnerability and its control are effectively black-boxed. This black box is then assumed to have a known transfer function and as such it interacts in a predictable fashion with other entities in the ISMS, becoming effectively an actor of the ISMS network. Hence, in the context of an ISMS, technology constitutes a black-boxed actor in its own right.

From the ANT viewpoint, the users involved in the ISMS are also considered as black boxes. The conformance of their actions to the enforced directives is supposed to be unquestionable and their actions rational, governed by the ISMS rules and human logic. Thus, with an assumed stable transfer function, the black-boxing of human actors is complete. In the extended sense, groups of users with common characteristics and/or roles can also become larger black boxes that are more than the sum of their constituent individual user black boxes. The reason for this is that the black box for the group does not merely contain the user black boxes but, instead, also contains their relations and translations between them. From an ANT perspective, the user group is a stand-alone network which can nevertheless be itself black-boxed for the purposes of the larger ISMS network.

Expressing almost everything in terms of black boxes facilitates the breakdown of problems and the synthesis of a solution such as the one provided by an ISMS. The down side of this process is that simplifying assumptions must occasionally be made in order to "*close the lid*" on black boxes. In the ISMS context the most dangerous such assumption is that the humans can be viewed as rational actors -the equivalent of black boxes with

known transfer functions. The fallacy in this assumption comes in total support of an earlier statement presented in this work in the discussion of the modernist view of ISMSs according to which "The assumption that all members of an organisation will play their ISMS-prescribed roles flawlessly during an attack is wildly optimistic at best".

The problem lies in the fact that according to ANT, if the operation (or transfer function) of a black box is proven to be inaccurate, the lid of the black box must be "re-opened" and the black box definition be revisited. Consequently, the links or relations of that black box actor with other nodes as well as the relevant translations running along those links must also be re-examined and amended. To aggravate things, the larger black box that contains the amended entities (smaller black boxes and the relations between them) must also have its lid opened and its operation re-evaluated.

This approach provides a more systematic view of the shortcomings of the modernist view of a mechanistically designed ISMS where all constituent parts are supposed to execute their function flawlessly in a fully predictable manner. It goes to prove that a wrong design assumption at the basic level of user behaviour may lead to the collapse of the whole system. The ISMS may fail to protect the Information if a single user in a critical position falls prey to the attacking Social Engineer.

The only way to avoid such design flaws as much as possible, is to constantly keep re-evaluating the validity of the user black boxes and be ready to re-define the black boxes to any extent required, in order to cater for their shortcomings. The current tendency is to bundle all users under the lowest level of generic incompetence with respect to Information Security and, based on that assumption, attempt to "idiot-proof" systemic functions and operation. This simplistic approach is definitely ignoring the following facts: a) that users are neither simple-minded nor ignorant by default, b) that users may indeed yield under the pressure of a Social Engineering attack but they can also be the only effective means of defence against such attacks and c) that the level of resistance of users against Social Engineering attacks can be raised through training and the promotion of a security-aware culture. By looking at user behaviour in detail, new black box definitions for users will arise, with more appropriate controls for user-related vulnerabilities.

One issue that ANT is particularly capable of analysing is the relation between technical and non-technical actors. In this sense, ANT can provide

a really good insight of how technical measures can be used to control non-technical vulnerabilities. In other words, how technical measures can be employed to steer the users' behaviour in such a way that it becomes resistant to Social Engineering threats. Extensions of this notion can have many repercussions, one of which is that political decisions can be inscribed in any solution/system in the form of a technical measure able to actively affect the organisation's culture-building effort and direct the human element towards a particular goal.

Black boxes can also help in providing an insight on the (previously discussed) issue that was raised by Berger and Luckmann on the differentiation of role-specific vocabularies between groups (Berger & Luckmann, 1991, p.158) and the resulting lack of common ground, communication and co-operation between the groups. Individual group members actually view other groups as black boxes and do not attempt to "open the lid" on them.

In similar fashion, technological issues and solutions remain in tightly closed black boxes for the majority of users who simply assume that these black boxes magically "do their job". This may lead to overconfidence on the part of users. Hence, the users become complacent, lowering their level of alertness as well as their defences. This is not unlike what can be observed when a user installs an antivirus solution on a PC and automatically assumes that the PC is fully protected against all Internet threats. What most users do not realise is that this sense of protection may become a false one if, for example, the scope of the solution is not understood, if regular virus list updates are not carried out or if the users themselves take such actions that compromise the integrity and effectiveness of the solution.

Through the above discussion it is made clear that Actor-Network Theory, through the use of 'black boxes' a) comes in direct support of the corollaries of Social Constructionism regarding ISMSs, b) goes further into providing better understanding of the issues involved and c) may even lead the way into devising appropriate solutions.

8 INSCRIPTION AND TRANSLATION IN THE ISMS

The notions of Inscription and Translation certainly help in the formal analysis of phenomena present in ISMSs. It was stated earlier in this text that "Inscription is the process through which a 'pattern of use' or 'action' is

coded or embedded in an artefact ". (An example of this statement can be obtained by considering how traffic rules are embedded in the traffic lights' patterns at a crossroad). In the case of the ISMS, the 'artefacts' of the previous statement are the technical and non-technical measures that are applied in an effort to reduce vulnerabilities. These artefacts ensure, among other things, that the human element of the ISMS behaves in a particular and predictable manner. In the context of the ISMS, a technical measure would be the use of passwords for logging-on to systems. A non-technical measure on the other hand would be the requirement for a user to not disclose and adequately protect his/her password and, on a different note, the administrative directives that govern reporting of possible social engineering attacks.

According to the already stated definition of translation by Singleton and Michael (1993), as "*the means by which one entity gives a role to others*", the above technical and non-technical measures seriously affect the behaviour of other actors (human users in this case) in the ANT-defined ISMS network.

For example, users are not accepted into a system if they do not use a password that uniquely identifies them and sets their rights properly on the system. Thus, the password infrastructure technical artefact defines the behaviour of the user to the extent that a password *must* be used. Having said that, the fact that a password infrastructure does exist as a technical measure, does not mean that users will not write down their passwords in obvious places or that they will not voluntarily share them and thus, in effect, compromise the system. If this technical measure is supported by the non-technical administrative measure of establishing serious penalties for such negligent behaviour, the overall result will indeed be better password protection.

On the other hand, assuming that a system-wide, password-strength checking algorithm is not in place, only a non-technical measure / artefact / directive may enforce the use of strong passwords. Such a non-technical measure also defines the behaviour of users, but to a different extent than a technical measure does. Directives of this type should be followed but, as practice shows, are not *necessarily* followed by all users.

The same holds true as far as SE attack reporting is concerned. There is no way that a user can be *forced* to take such reporting action. It is rather an issue of having convinced the users beforehand as to the importance of

reports been filed in the case that a SE attack is suspected. Ultimately, unless this type of behaviour becomes the users' "second nature" in their everyday dealings, SE attacks will remain unnoticed. The responsibility for such a goal remains with the management that must promote the appropriate security culture and thus effectively establish yet another, very important, non-technical measure.

As standard procedure, when a new or amended security policy is effected, all office workers sign statements that they have been duly notified of this and thus the security policy is considered to be active. As organisations are feeling the pressure to adopt IT methods in order to become more efficient or more competitive, the integration of IT into the business process is not always a carefully planned one, especially with respect to security. Even if this is not so and the new security policy is indeed a carefully produced one, the hysteresis involved in the office workers' understanding and internalisation of the new situation, usually lies at the basis of the inefficiency or even of the de facto demise of any security policy. Office workers may well be acquainted with the security requirements governing physical access or those requirements relevant to protecting a filing cabinet. They usually, though, understand very little regarding the security of an IT system and consider this to solely be of interest to, as well as the responsibility of, the IT department. Having being notified of and having signed documents pertaining to the new security policy, does not actually make the average worker more security-aware neither does it help in altering the office workers' day-to-day activities towards achieving a higher level of IT security. Combining this with the fact that the average office worker is the first weak link that the Social Engineer will attempt to exploit on the way to the primary target, clearly demonstrates the gravity of the situation. Hence, once again, the need for the promotion of a security culture that appropriately caters for the IT-based organisational reality is brought forward as an indispensable non-technical measure.

Strong incentives and counterincentives can support non-technical measures, as can additional technical measures. An example of such a technical measure would indeed the upgrade of a system to include a password-strength-checking mechanism that rejects weak passwords.

Thus, technical and non-technical measures can come in efficient reciprocal support, effectively dissolving the idea that IS is either a purely technical or purely administrative issue.

Furthermore, an ISMS that is realised under the assumption that users are rational actors, is probably doomed by design. The reason for such a failure is that the assumption of a fully rational and predictable behaviour by the human users involved, leads to the adoption of a minimal set of inscriptions. This would in turn produce inadequate or incomplete translations. Thus the deciding question in this case would be what the full set of inscriptions and translations for a given ISMS is.

Unfortunately, there is no deterministic way of identifying every potentially vulnerable aspect of an organisation and incorporating it in the design of an appropriate ISMS, especially when Social Engineering is factored in. On a more optimistic view though, more SE vulnerabilities can be identified if the diverging subjective realities of the users are acknowledged and examined.

From that point onwards, the greater the number of SE vulnerabilities that are catered for in the context of an ISMS, the harder it will be for the next Social Engineer to mount a successful attack, especially when the Plan-Do-Check-Act (PDCA) cyclic process for the ISMS' continual improvement is adopted.

The diagram of Figure 1 should help in visualising the effect that a correctly implemented PDCA cycle may have on the divergence of the users' subjective realities.

As it can hopefully be seen, the PDCA cycle causes the users to espouse more of the actual policy directives as their own subjective reality (hence the double-shaded area increases) and thus the opportunity for a Social Engineer to act, diminishes.

9 POWERPLAY WITHIN THE ISMS

Having dealt so far with the shortcomings of the modernist approach to Information Security and having identified the inherent difficulties stemming from the differences of individual groups within an organisation, it would be naïve to ignore the repercussions that the balance of power in the context of an ISMS has on its own functionality and effectiveness, as well as on the organisation in general.

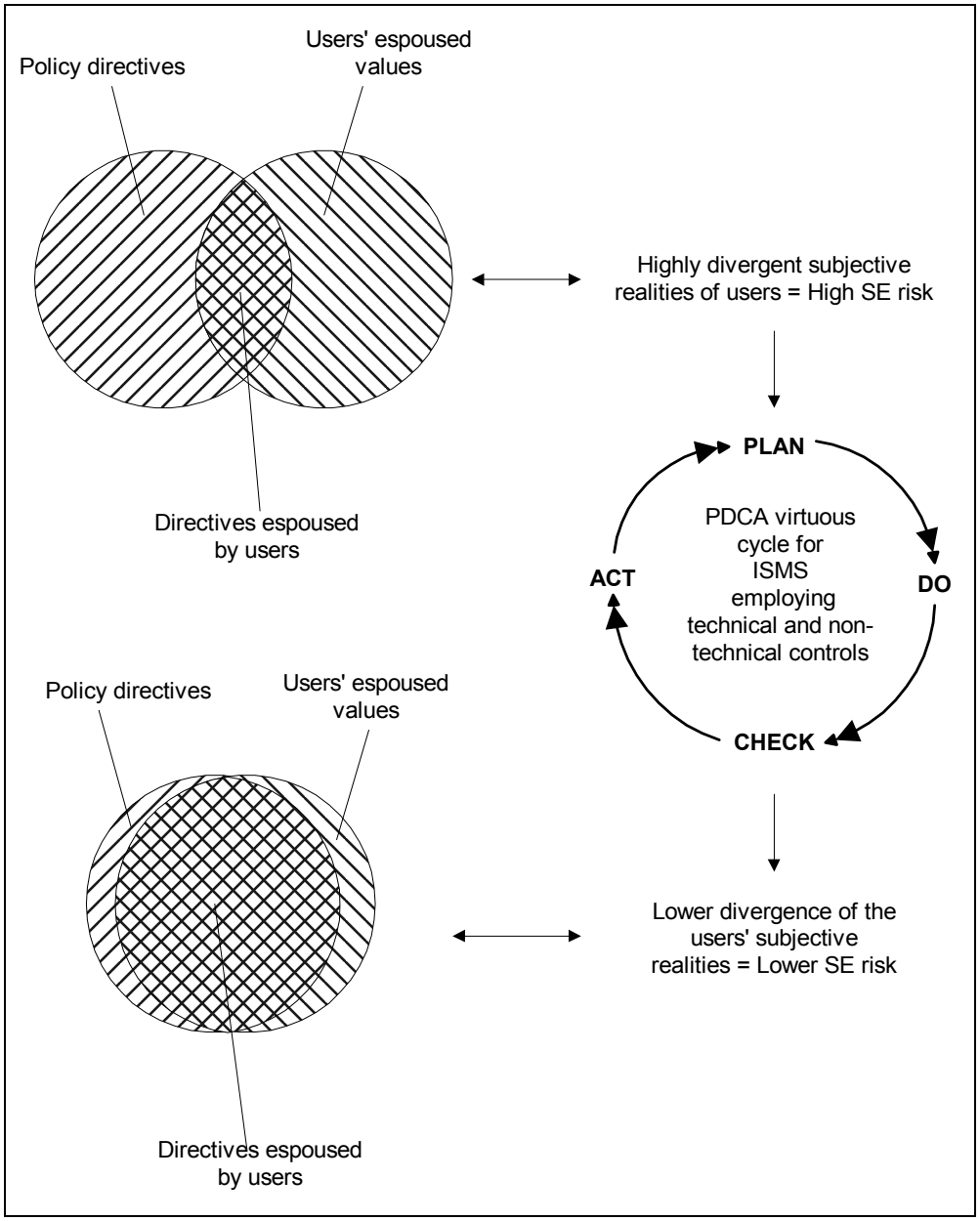


Figure 1: Effect of PDCA cycle on users' diverging subjective realities

"*Power*" is generally accepted to be the ability of an individual or a group of people to realize their own will in communal action, even against the resistance of others (Giddens, 2001, p.420). In viewing the ISMS as a social construct, it has to be taken for granted that the individual groups involved in its operation will ultimately fight for power. The Marxist view is that the struggle for power always has economic motives and in particular the possession of goods and opportunities for income. Also according to Marx, a grouping of people constitutes a "*class*" and class action ensues, when a class becomes conscious of its interests, in the context of its relation, as a class, to other classes (Giddens, 2001, p. 669). Weberian theory gives a more refined view of power and classes that aptly conforms to any bureaucratic system, including ISMSs: According to Weber, the Marxist view of a single source for power is dogmatic. Instead of having motives of a strictly economic nature, Weber argues, that individuals seek power for its own sake due to its intrinsic values and the social honour it carries (Bottomore, 1990, p.238). This notion is then taken one step further and Weber sets the foundation for the "*politics of power*" (Doujon, 1990, p.13).

Regarding classes, Weber introduces an additional structural category, that of the "*status group*". Marxist classes are defined with respect to their place in the market or in the process of production. Furthermore, classes may or may not exist as communal groupings. In contrast to those, Weberian status groups are, in principle, communities formed and held together by commonly accepted values, shared beliefs, similar lifestyles and, most importantly, by the social status, esteem and prestige conferred upon them by others (Giddens, 2001, p.285). Thus, "*social distances*" are established between status groups. Furthermore, according to Weber, status groups are independent of class divisions. Status may vary independently of class.

When a status group gradually develops the idea that the magnitude of the social distance between it and the next superordinate group is too great and that it should be diminished or even nullified, conflict takes place. This conflict ultimately upsets the existing stratification until a generally acceptable equilibrium point defining subordination and superordination is reached. When such a point is reached, conflict subsides and tranquility returns, with members of groups accepting their position and assuming their place in the hierarchy. When the situation is such that warrants the ascension of a group to a higher status stratum, conflict eventually begins

again and the cyclic procedure re-iterates itself. During the time of tranquility (which is the usual case), subordination tends to become more prominent. Under those circumstances, the members of the subordinate group tend to acknowledge the authority that the members of the superordinate group exercise over them. Furthermore, the members of the subordinate group usually become fearful of displeasing those that are higher in hierarchy than themselves. It is this fear of displeasing one's superiors that is frequently exploited by Social Engineers during their attacks.

What can be seen clearly at this point is the obvious need for an equilibrium point to be reached in the social distances between the groups. This equilibrium point should neither be unstable, thus leading to perpetual conflict between groups, nor predispose members of one subordinate group to carry out orders supposedly coming from their superordinates, in an automatic and mindless fashion. Social Engineers are very apt in using authority, fear and intimidation to their advantage and would thrive in either of the two situations.

In the particular case of the ISMS, the stratification phenomenon and the separation of the individuals involved into various users' groups, is justified not only by the divergence of the groups' interests, but also by the distinction in the life-styles, views of the world and postures of their constituents. As IS professionals seek the status and authority to carry out their mission, management group members fear that this may constitute a flanking attack against their own hard-earned status. The highly technical nature of the means employed by the IS professionals in the line of their work, is seldom fully understood by management. This makes members of the management group feel insecure and even aggravates the chance for conflicts between the groups.

Additionally, the group of IS personnel, frequently, does not occupy a clearly defined position in the organisation's hierarchy. In effect, this creates a two-fold status problem for the IS experts group. The first facet of the problem is that high-ranking officials may disregard the security-related control attempted by the IS personnel. This disregard can be passive, in the sense that high-ranking officials may simply ignore the efforts of IS personnel to control them, or active, through intimidation and commination of the IS personnel. Secondly, as long as the higher status of the management group in the hierarchy is undisputed, members of the

management group may use the vagueness of the IS group's status to their advantage by discreetly fuelling the status struggle of the lower-ranking groups in the organisation, as part of a typical divide-and-conquer strategy that results in the strengthening of their own status. As a result, the members of the IS group are viewed by members of the other groups as 'floating' within the organisational structure, not having any particular role or real control over the other groups' members' actions. This fuels inter-group competition, and in effect further undermines the IS group's role while crippling the IS effort. A Social Engineer will definitely make the most of such a situation, either by using the weaker spots in the crippled security system or by actively (and carefully) assuming the role of a high-ranking official in order to achieve the SE objective through intimidation or by otherwise using the status of the assumed role.

The above analysis follows the modernist view of power and although useful in analysing the social structure of an ISMS, it would be unacceptable to ignore the post-modernist view of power that can also apply to ISMSs. The best known such view of power is presented by Foucault, a self-pronounced champion of post-modernism, throughout his works (1988, p.39; 1989, p.65; 2005). Foucault views power as one of many societal controls aiming at a variety of targets from production for financial gain to disciplinary systems to normalisation procedures, all the while being dispensed through historical institutions and exalted by definitions of normal vs. abnormal. Translating this into the reality of the ISMS, power can be seen as originating from the set of technical and non-technical controls that effectively influence the behaviour and actions of the human actors. In effect, power in the ISMS is stemming from the conglomeration of tools, instruments, techniques and procedures that are defined in it.

The fact that ISMS implementations are currently highly technological in nature, has the effect that power is *de facto* passed to the IS professionals who have the responsibility of specifying, designing and implementing the ISMS as well as maintaining its operation. In ANT terms, the IS professionals are responsible for the inscription and translation of the bulk of the effort towards IS. It is interesting to note that apart from the technical controls which are obviously within the scope of the IS professionals' work, non-technical controls have both technological and administrative inscription components which also require the extensive involvement of IS professionals. The controlling artefacts of an ISMS are the fruits of the IS

professionals' efforts and mentality. These artefacts thus function as conduits for the power of the IS professionals which permeates all aspects of the organization, not just the ones related to the ISMS at hand.

Using the barrier of technology, the group of IS professionals can effectively create an impenetrable perimeter, that neither end-users nor management can break through. This may lead to inadequate ISMS inscription and translation as groups other than that of the IS professionals are isolated from the ISMS design process. For efficient and generally acceptable ISMSs to exist, they should not be designed by IS professionals alone but with the active participation of all groups within the organisation. Every ISMS inadequacy is bound to be exploited by the Social Engineer under the proper circumstances. Hopefully, if all groups participate in the creation of the ISMS, it will be easier for members of groups other than the IS professionals to espouse the directives of the ISMS (or in ANT terms "*internalise*" those directives), and make the ISMS function more efficiently. The possible disadvantage to this is that there may exist a higher level of conflict between the groups during the design phase of the ISMS. Care should be taken for such a situation not to become explosive and either hinder the creation of the ISMS or produce an ISMS with severe design flaws.

Either the absence of an ISMS altogether, or the existence of a flawed one, will give ample opportunity for the Social Engineer to act.

10 CONCLUDING REMARKS

By attempting to create a security policy that governs any kind of hierarchical structure, complex interactions come into existence. The social construct underlying the hierarchical structure affects, or even defines, the design, functionality and efficiency of the security policy. On the other hand, the security policy itself affects and transforms the dynamic relationships within the social construct. When this mechanism is set in motion and until an equilibrium point is eventually reached, a period of tumult may be incited. Inconspicuous vulnerabilities that are due to purely sociotechnical reasons arise during such periods, leading to a significant drop in the efficiency of the security policy. Consequently, a Social Engineer may find ample opportunity to mount successful attacks. Furthermore, there is always a possibility that some of the vulnerabilities of the described type are not identified and may thus remain unmitigated for a

long period of time after the initial establishment of the security policy. Thus, emphasis must be placed in the effort to identify these 'socially-induced' vulnerabilities and establish controls for them, if SE attacks are to be repelled.

The study presented in this paper actively supports the research towards combating Social Engineering threats, by providing an insight into the socially-defined opposing forces and interactions within an ISMS that Social Engineers attempt to exploit.

11 REFERENCES

AKRICH, M. 1992. The De-Scription of Technical Objects. Bijker, W. and Law, J.(Eds.). Second printing,1997. *Shaping technology/Building society studies in sociotechnical change.* pp. 205-224. Cambridge, MA: MIT Press.

ALBRECHTSEN, E. 2004. *Information managed securely? An approach to the social construction of information security management* [online]. Term paper, Norwegian University of Science and Technology. Available from URL: http://www.iot.ntnu.no/users/albrecht/rapporter/OTE_paper_Eirik_Albrechtsen.pdf. [Last access on May 4, 2005]

BERGER, P. L. and LUCKMANN, T. 1991. The social construction of reality. A treatise in the sociology of knowledge. London: Penguin Books.

BOTTOMORE, T. B. 1990. Κοινωνιολογία - κεντρικά προβλήματα και βασική βιβλιογραφία. [Greek] [Sociology - A Guide to Problems and Literature]. Translated from English by D. G. Tsaoussis. Athens: Gutenberg

DELIGIORGI, A. 1996. *Ο Μοντερνισμός στη Σύγχρονη Φιλοσοφία : Η αναζήτηση της χαμένης ενότητας.* [Greek] [Modernism in Contemporary Philosophy : The search for the lost unity]. Athens: Αλεξάνδρεια [Alexandria].

DHILLON, G. and BACKHOUSE, J. 2000. Information System Security Management in the New Millenium. In: *Communications of the ACM.* **43**(7) 125-128.

DOUJON, J.-P. 1990. *Histoire des faits économiques et sociaux.* [French] [History of economic and social events]. Grenoble: Presses Universitaires de Grenoble.

FOUCAULT, M. 1988. *Τι είναι Διαφωτισμός;* [Greek] [What is

Enlightenment?]. Translated from French by Stefanos Rozanis. Athens: Εκδόσεις 'Ερασμος [Erasmos Publications]
FOUCAULT, M. 1989. *Επιτήρηση και Τιμωρία: Η γέννηση της φυλακής* [Greek] [Discipline and Punishment: The birth of prison]. Translated from French by Kate Chatzidimou and Ioulietta Ralli. Athens: Κέδρος - Ράππα [Kedros - Rappa]
FOUCAULT, M. 2005. *Εξουσία, Γνώση και Ηθική* [Greek] [Power, Knowledge and Morality]. Translated from French by Zissis Sarikas. Athens: Ύψιλον [Ypsilon]
GIDDENS, A. 2001. *Sociology* (4th edition). Oxford: Blackwell Publishing Ltd.
HANSETH, O. and MONTEIRO, E. 1998. *Understanding Information Infrastructure*. (e-Book) [online]. Available from URL: http://heim.ifi.uio.no/~oleha/Publications/bok.html [Last access on Aug 28, 2006].
ISO/IEC. 1997. International Standard ISO/IEC TR 13335-2:1997. Information technology - Guidelines for the management of IT security - Part 2: Managing and planning IT security. Geneva: ISO Copyright Office.
ISO/IEC. 1998. International Standard ISO/IEC TR 13335-3:1998. Information technology -- Guidelines for the management of IT Security -- Part 3: Techniques for the management of IT Security. Geneva: ISO Copyright Office.
ISO/IEC. 2000. International Standard ISO/IEC TR 13335-4:2000. Information technology -- Guidelines for the management of IT Security -- Part 4: Selection of safeguards. Geneva: ISO Copyright Office.
ISO/IEC. 2001. International Standard ISO/IEC TR 13335-5:2001. Information technology -- Guidelines for the management of IT Security -- Part 5: Management guidance on network security. Geneva: ISO Copyright Office.
ISO/IEC. 2004. International Standard ISO/IEC 13335-1:2004. Information technology -- Security techniques -- Management of information and communications technology security -- Part 1: Concepts and models for information and communications technology security management. Geneva: ISO Copyright Office.
ISO/IEC. 2005a. International Standard ISO/IEC 17799:2005. Information technology -- Security techniques -- Code of practice for information security management. Geneva: ISO Copyright Office.

ISO/IEC. 2005b. International Standard ISO/IEC 27001:2005. Information Technology - Security techniques - Information security management systems- Requirements. Geneva: ISO Copyright Office.
ISO/IEC. 2005c. International Standard ISO/IEC 15408-1:2005. Information technology -- Security techniques -- Evaluation criteria for IT security -- Part 1: Introduction and general model. Geneva: ISO Copyright Office.
ISO/IEC. 2005d. International Standard ISO/IEC 15408-2:2005. Information technology -- Security techniques -- Evaluation criteria for IT security -- Part 2: Security functional requirements. Geneva: ISO Copyright Office.
ISO/IEC. 2005e. International Standard ISO/IEC 15408-3:2005. Information technology -- Security techniques -- Evaluation criteria for IT security -- Part 3: Security assurance requirements. Geneva: ISO Copyright Office.
ISO/IEC. 2005f. International Standard ISO/IEC 27002:2005. Information technology -- Security techniques -- Code of practice for information security management. Geneva: ISO Copyright Office.
LATOUR, B. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.
LATOUR, B. 1987. Science in action: How to Follow Scientists and Engineers Through Society. Cambridge, MA: Harvard University Press.
LATOUR, B. 2005. Reassembling the Social. An introduction to Actor-Network-Theory. Oxford: Oxford University Press.
LAW, J. 1992. Notes on the theory of the actor-network: ordering, strategy, and heterogeneity. *Systems Practice*. **5**(4) 379-393.
LEIWO, J. and HEIKKURI, S. 1998. An Analysis of Ethics as Foundation of Information Security in Distributed Systems. In: *Thirty-First Annual Hawaii International Conference on System Sciences*-Volume 6. IEEE.Also: [online] available from URL: http://computer.org/publications/dlib/ [Last access on June 27, 2005].
LOW, J. et al. 1996. Read this and change the way you feel about software engineering. *Information and Software Technology* 38,77-87.
MENDELSSOHN, M. et al. 1989. *Τι είναι Διαφωτισμός;* [Greek] [What is Enlightenment?]. Translated from German by N. M. Skouteropoulos. Athens: Εκδόσεις Κριτική [Kritiki Publications].
MORGAN, G. 1996. *Images of Organization* (2nd ed.). Thousand Oaks,

CA: Sage Publications, Inc.
OAKES, G. 1998. On the Unity of Max Weber's Methodology. In: *International Journal of Politics, Culture, and Society*. **12**(2) 293-306
OSTROFF, F. and SMITH, D. 1992. The Horizontal Organization. *McKinsey Quarterly*. **1** 148-168.
RIDDENER, L. R. 1999. *Dead Sociologists Society - Max Weber - Bureaucracy*. [online]. Available from URL: http://www2.pfeiffer.edu/~lridener/DSS/Weber/BUREAU.HTML [Last access on July 5, 2006].
SCHACH, S R. 2005. Object-oriented and Classical Software Engineering. 6th ED., McGraw-Hill
SINGLETON, V. and MICHAEL, M. 1993. Actor-Networks and Ambivalence: General Practitioners in the UK Cervical Screening Programme. *Social Studies of Science*. **23** 227-264.
TATNALL, A. and GILDING, A. 1999. Actor-Network Theory and Information Systems Research. In: *Proceedings of the 10th Australasian Conference on Information Systems*. p.955-966
UNIVERSITY COLLEGE LONDON, 2003. *Digital Egypt for Universities - A learning and teaching resource for higher education - Law in ancient Egypt* [online]. Available from URL: http://www.digitalegypt.ucl.ac.uk/administration/law.html [Last access on Apr. 20, 2008]
WEBER M. 1978. *Economy and Society. Edited by Guenther Roth and Claus Wittich*. [Wirtschaft und Gesellschaft]. Berkeley: University of California Press.
WHITTEN, J. L. & BENTLEY, L. D. 2007. *Systems Analysis & Design for the Global Enterprise*. Seventh Edition. McGraw-Hill.

CONSULTED BIBLIOGRAPHY

FOUCAULT , M. 1985. La vie: l'expérience et la science [French] [Life: experience and science] In: *Dits et Ecrits,* t.IV, p.763-776
HACKING, I. 1999. *The social construction of what?* Cambridge, MA: Harvard University Press.

THE IMPACT OF INFORMATION SECURITY AWARENESS TRAINING ON INFORMATION SECURITY BEHAVIOUR: THE CASE FOR FURTHER RESEARCH

AT Stephanou[1], R Dagada[2]

[1, 2] University of the Witwatersrand
tony.stephanou@gmail.com
rabelani.dagada@wits.ac.za

ABSTRACT

Information Security awareness initiatives are seen as critical to any information security programme. But, how do we determine the effectiveness of these awareness initiatives? We could get our employees to write a test afterwards to determine how well they understand the policies, but this does not show how it affects the employee's on the job behaviour. Does awareness training have a direct influence on the security behaviour of individuals, and what is the direct benefit of awareness training? This paper represents a study in progress that aims to answer the question: to what extent does information security awareness training influence information security behaviour?

Research carried out on information security has traditionally been slanted towards technical aspects of security, typically rooted in computer science and mathematics. Security was traditionally seen as a service to be provided and not something that was influenced by users. However, it was soon recognised that focusing on technical issues alone is inadequate. Technologies meant to provide security ultimately depend on the effective implementation and operation of these technologies by people. Thus awareness of policies is needed by all individuals in an organisation to ensure that policies are well understood and not misinterpreted. Some

researchers have maintained that educating users is futile mainly because it is believed that it is difficult to teach users complex security issues and secondly, because security is seen as secondary by the user they will not pay enough attention to it. This paper reflects research in progress and discusses some of the problems with existing information security awareness research and proposes a model to be tested for examining the impact of information security awareness training on information security behaviour.

KEYWORDS

Information security, behavioral information security, awareness initiatives, on the job behaviours, policies, and further research.

THE IMPACT OF INFORMATION SECURITY AWARENESS TRAINING ON INFORMATION SECURITY BEHAVIOUR: THE CASE FOR FURTHER RESEARCH

1 INTRODUCTION AND BACKGOUND

Information Technology systems are dependant on people. Schneier (2003:10) maintains that information security is more about behaviour than anything else, i.e. getting people to behave in a certain way. It is people's intentional and unintentional actions that cause adverse consequences that security wants to prevent. Despite the hype from vendors about the need for security products many critical security activities have not and cannot be automated. Technologies meant to provide security ultimately depend on the effective implementation and operation of these technologies by people. This means that organisations are dependant on people to achieve a secure environment. Since humans are seen as the "weakest link" in the information security chain (Schneier, 2000; Stanton et. al. 2003:1; Katsikas, 2000:130; van Niekerk & von Solms, 2004:2; von Solms, 2000:618), there is a clear requirement to ensure users are trained correctly in terms of information security policies. The goal is to ensure that users use the necessary policies and to ensure that they are not misused or misinterpreted, thereby ensuring the effectiveness of policies (Siponen, 2000:31). Security awareness efforts are seen as the "first line of defence" (OECD, 2002:10). On the other hand, Van Niekerk & von Solms (2004), argue that awareness initiatives while necessary are not sufficient to obtain the desired results, while other authors simply consider educating users futile (Ranum, 2005; Evers, 2007; Nielsen, 2004).

Well established security management standards such as the SABS ISO/IEC 17799 and the OECD guidelines for information systems security also promote the importance of making people aware of security issues. The 2007 Computer Security Institute (CSI) Survey reported a substantial

increase in the importance of security awareness perceived by those surveyed. In the 2006 CSI survey, on average, respondents felt that their organisations were under investing in awareness at that time (Computer Security Institute, 2006). These results imply that organisations do realise the importance of security awareness efforts. Thus the need for information security is well established, but there is inadequate research on the behavioural aspects of awareness initiatives (Schultz, 2004:1; Siponen, 2001:24; Srikwan & Jakobsson, 2007:2; Van Niekerk & von Solms, 2004).

Despite the understanding that awareness is important, it is not beyond doubt whether a clear message is being communicated to users in the first place (Gaunt, 2000:152-153). This is especially true for dynamic, complex threats such as phishing attacks. Srikwan & Jakobsson (2007), for example, doubt whether a clear message is being communicated to users with respect to identity theft, specifically on what to do and why it must be done – even though a vast amount of guidance on this subject is being directed at users. South African banking clients for example are frequently warned about the threat of phishing scams (via email, SMS and so on). Are these interventions having an effect? Perhaps, there may be too much information for lay people to digest and security practitioners may be unwittingly shooting themselves in the foot.

With all this emphasis on awareness, the question one has to ask is: to what end? In other words, does making users more aware lead to more secure behaviour and therefore contribute to a more "secure" organisation or, are awareness campaigns doomed to fail?

The purpose of this paper is two-fold. Firstly, it will be demonstrated that there is a shortage of in-depth information security awareness research and that behavioural concepts are not properly taken into account for security awareness programmes. Next, this paper represents research in progress aimed at explaining and answering some of the questions raised above. A theoretical model is put forward proposing how a particular security awareness approach affects behaviour. This will help scholars and practitioners understand why an awareness initiative is expected to have certain results on security behaviour. The theory proposed will be then be tested empirically using a pretest-posttest experimental design. The authors believe that the contribution of this research is significant in the following ways: The research is a case study that will use system generated data to measure actual user behaviour before and after the security awareness

training intervention in order to determine effectiveness of the training. Therefore the perceptions of users about their own behaviour will not be relied upon. Existing research has used interviews, surveys and "participatory observation" to make conclusions about end-user behaviours in this regard. The research will measure a subset of behaviours required by a typical Acceptable Usage Policy, whereas much of the existing and recent research with respect to awareness training effectiveness has focused on phishing related threats. The research in progress intends to not only demonstrate the impact of security awareness training on user behaviour but to also contribute towards a set of instruments that could be used in future research for behavioural measurement. Finally, the study underway, uses as a foundation, the user behaviour taxonomy developed by Stanton et. al. (2005) in an effort to begin to consolidate the security awareness research landscape and move towards a common understanding and language of what "security behaviour" means.

Various branches relating to information security awareness research currently exist. The landscape of information security awareness research can be categorised as follows:

Figure 1 demonstrates one way of making sense of the available information security research. Most of the research work can be placed into one of these categories. This paper will not discuss research focused on computer abuse or the insider threat.

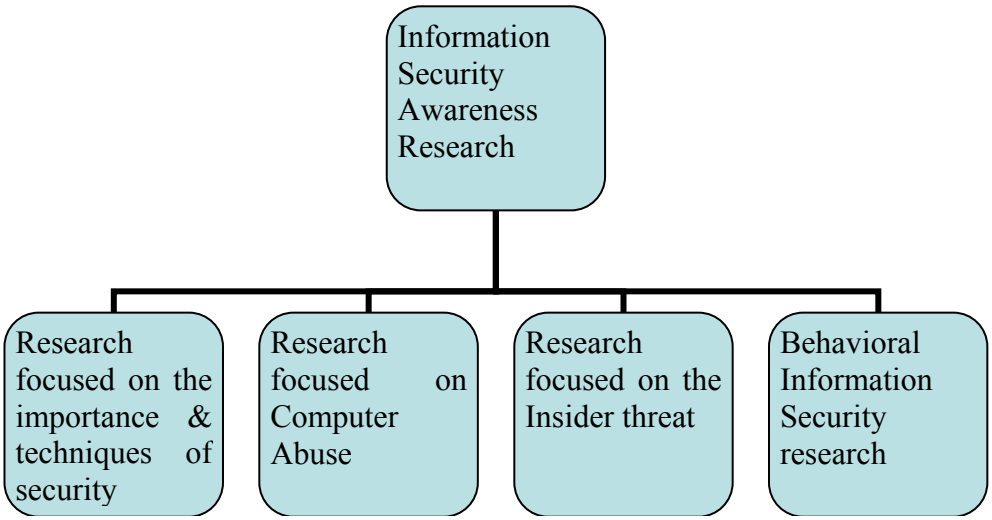


Figure 1: Information Security Awareness Landscape

2 RESEARCH FOCUSED ON THE IMPORTANCE & TECHNIQUES OF SECURITY AWARENESS

Most of the research concentrates on the importance of awareness initiatives (Nosworthy, 2000; Furnell et. al., 2000; von Solms, 2000; von Solms, 2001; Siponen, 2001; Janczewkski & Xinli, 2002) and awareness techniques (Furnell et. al. 1997; Gaunt, 1998; Gaunt, 2000; van Niekerk & von Solms, 2004; Trompeter & Eloff, 2001; Katsikas, 2000; Johnson, Eloff & Labuschange, 2003; Thompson & von Solms, 1998). Some of this research, is not necessarily based on a theoretical model, but instead simply provides guidance on what methods to use. Sommers & Robinson (2004:379) show how an awareness video and a quiz can be used to train students at a university. However, the researchers admitted that they had no way of measuring the effectiveness of this intervention. A video was simply shown and respondents were required to take a quiz afterwards. McCoy and Fowler (2004:349) also deployed a security awareness campaign at a University campus. They too however, did not use any metrics and found this to be a difficult task to carry out – thus implying the importance for this piece of research. Other researchers have also demonstrated approaches for information security awareness programmes such as Perry (1985:94-95), Spurling (1995:20) and Parker (1998:466).

So even though methods may be used to make users aware, recipients of the message may not apply what they know whether they understand the message or not. Some of the reasons for this are because security technologies are difficult to use and consequently not used very well. For example, Furnell, 2005:274 demonstrated the difficulty that users have in finding, understanding and using security features in Microsoft Word. In another case, Whalen & Inkpen (2005:137) measured eyeball tracking of users when using web browsers and concluded that although some security information is viewed (indicating that users were "security aware"), users do not interact with it in order to fully understand its implications. The study also found that users tend to stop looking for security information once they have logged into a site (Whalen & Inkpen 2005:143).

Srikwan & Jakobsson (2007), argue that educational efforts generally expect too much from the audience while others – in an effort to make the message more palatable – simplify the message to such an extent that the meaning is diluted. Without an adequate understanding of security requirements and their support, security processes are bound to be

ineffective (Van Niekerk & von Solms, 2004). For example, a well-crafted incident management process is useless if an employee is not aware of firstly what a security incident looks like and then how to respond to the incident when one is recognised. Ultimately, security education in this context becomes inadequate. Thus security awareness practitioners need to ensure that there is a connection made between what a user knows and what the appropriate behaviour expected from them is. In order for security to be enhanced they need to be told not only what to do but why they should do it.

The problem may be more complex than originally anticipated by security practitioners. Perhaps the solution is not only to deploy awareness campaigns and educate users, but more related to the notion of the ability of users to understand risk and make trade-offs (Schneier, 2003:17) and naturally wanting to be helpful (Mitnick & Simon, 2003). Most of the time people are told what to do without explaining why they need to do this. This is linked to people's understanding of threats. If they are able to understand the underlying threat then they will be able to look for patterns and consequently mitigate any threat posed (Srikwan & Jakobsson 2007).

Security education may inadvertently also have the opposite effect intended and enhance the level of risk that users expose themselves to. For example, if users are instructed to explicitly not share their credit card details to anyone requesting it via email and the attack is changed so that this information is requested telephonically then users could be at risk for simply following what they were told to do. In essence the message needs to be simple enough to capture the problem without losing the complexity of the threat. This is particularly true for education about phishing attacks (Srikwan & Jakobsson 2007).

Despite these challenges, Kumaraguru et. al (2007) showed that security awareness material – when used - can be effective. They found that online material that informs users about the threats of phishing was highly effective – resulting in users getting better at identifying phishing sites. They also call for looking at more effective techniques to deliver the awareness message, getting users to actually read and absorb the material and, ensuring more work is done on the quality of awareness materials presented.

Jagatic et. al. (2007:96) also used contextual training. They demonstrated that a large amount of information (accessible via social networking sites on the Internet) was easily obtainable and could effectively

be used for phishing attacks. The researchers also wanted to measure how social context information could influence the success of phishing attacks. The difference with this research is that they tricked their users by spoofing emails that looked like it came from friends in their social network. The number of students that fell prey to the (harmless) phishing attack was 72% (out of 487 targeted students) – this was much higher than anticipated (Jagatic et. al., 2007:97).

To summarise, previous research on information security awareness has been skewed towards awareness techniques, computer abuse and insider threats. Although recent research has started examining the effectiveness of security awareness the focus has been on phishing threats, which has shown the effectiveness of class-room based training, phishing tests, email based training and web-based awareness material. Measuring the effectiveness of overall security awareness and examining behavioural aspects have been largely neglected. In addition, very few theoretical models have been presented and used to explain and test security behaviours.

3 BEHAVIORAL INFORMATION SECURITY

The importance of getting people to act correctly has always been implied by previous research work. However, a few years ago there has been more explicit focus on behavioural aspects of security. Behavioral information security is a branch of information security research which examines what motivates security related behaviours of computer users. Recent work in behavioral information security has shown: how employee job attitude relates to information security behaviours (Stanton et. al., 2003); what categories of information security behaviours exist (Stanton, et. al., 2005); what influences information security behaviours (Leach, 2003) and, how attitudes and intentions are significant factors in explaining why some employees do not comply with information security policies (Pahnila et. al., 2007).

The study underway described in this paper adopts the model proposed by Stanton et. al (2005) in order to make conclusions about whether awareness training has an effect on specific behaviour categories. This model states that all security behaviour can be plotted on a behavioural continuum. On one level behaviour is categorised based on a user's intentions: from malicious to neutral to benevolent intentions. On another level behaviour can be categorized based on the level of expertise held by

the user ranging from novice to expert and something in between the two. This produces a two-factor taxonomy of user security behaviours yielding six broad behaviour categories as shown in figure 2 below.

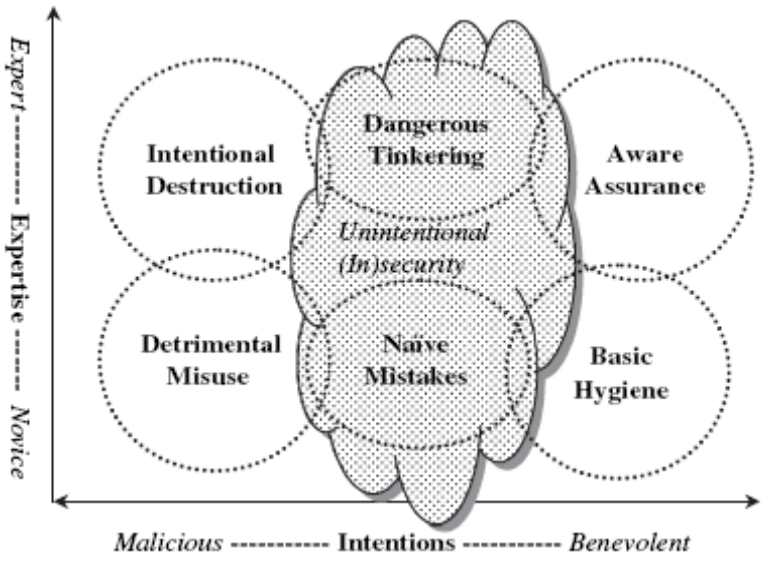


Figure 2. Two-factor taxonomy of end-user security behaviours (Stanton et. al 2005).

Using the model above, information security behaviours can be mapped against two-dimensions, i.e. the level of expertise the end-user possesses and the behavioural intent held by the end-user. The outcome is six different behavioural categories, which the researchers show, most security behaviours will be able to fit into (Stanton et. al, 2005:131). Putting this in context, the goal of security awareness initiatives is to move the intentions of employees towards the right-hand side of the chart. Thus Stanton et. al. (2005:132) provides a practical framework for categorising information security behaviours. This model now lays a foundation for the measurement of security behaviours. An illustrative example of the above taxonomy is shown in the table below:

Table 1.Examples of behaviours that that require low levels of expertise.

Behaviour	Intent	Expertise
Employee sends pornographic material to colleagues.	Malicious	Low
Employee shared password with his wife.	Neutral	Low
Employee chooses a strong password.	Benevolent	Low

Stanton et. al. using simple correlation, showed that good password practices (such as changing passwords frequently and choosing strong passwords) was associated with training and awareness, employees' knowledge of being monitored and organisational benefits, perceived by employees (2005:124,131). A positive correlation does not mean that training and awareness caused these good password practices though. These password practices are known as naïve end-user security behaviours. These behaviours are characterised by individuals with a low level of expertise and with neutral intentions (neither malicious nor benevolent). Interestingly this same piece of research work did not find any correlation with another type of naïve security behaviour – that of sharing one's password. They concluded that there is no evidence that password sharing behaviour is associated with training, awareness, organisational rewards and knowledge of being monitored (Stanton et. al, 2005).

Additional research is needed in this area and is called for explicitly by Stanton et al. (2005). Secondly, different techniques will be used in the current study which may yield different results as those obtained from Stanton et. al. (2005). Vroom & von Solms (2004:191-192,194, 197) have also recognised the importance of human behaviour in the security chain but from an auditing perspective. The argument put forward is that although auditors express an opinion on an organisation's financial and IT arrangements, employee behaviour – which is a key aspect of information security - is not measured. They claim further that the reason that end-user behaviour is often neglected is because it is so difficult to measure and will inevitably be flawed. Auditing end-user behaviour is compared to carrying out employee performance appraisals and the resultant flaws associated with such activity namely: reliability and validity factors. They believe there are too many factors that may interfere with "auditing" the employee

accurately. Thus an alternative approach for auditing behaviour is put forward by them. They proposed that a better approach is to attempt to change organisational culture one level at a time and thereby influence end-user behaviour.

The implications of the Vroom & von Solms' work on this study are significant. Showing that behaviours can be measured in this context, adds a new dimension to the notion put forward by Vroom & von Solms. In addition, the techniques used and the lessons learned will form the groundwork for further research work to take place. Gaunt (2000:151,157), believes that information security awareness initiatives, while important, do not guarantee that staff will comply to appropriate security behaviours. Referring to the health care community he argues that a security culture needs to be entrenched for security to be effective. This requires amongst other things, strong commitment from senior management, clear lines of accountability and responsibility.

According to Gaunt's studies (2000:152-153), a number of obstacles need to be overcome to ensure security measures are effective and a culture of security is instilled. These include: Getting users to change their behaviour to a more secure form may be difficult especially if they have been used to using computer systems in an insecure way. Enforcing stronger security measures may in reality cause more reluctance by employees to change their behaviour. In addition to this, employees may view security measures as impractical and a hindrance to their work. Being unaware exactly what is required of them may also cause employees to become reluctant to embrace security.

Inconsistent application of policies among or within organisation's may lead to frustration by employees and thus undermine the effectiveness of the policies.

Gaunt research, while providing insight into obstacles, also indicates the complexity of the problem and its behavioural aspects. Pahnila et. al. (2007) demonstrate the complexity of security behaviour by arguing that compliance to policy is in fact made up of the intentions and attitudes of employees (which themselves are determined by various factors). They therefore recommend that promoting positive social pressure on employees with respect to compliance to security policies (for example, by all levels of management and peers within organisations) promotes actual security compliance. This should be done by explicitly stating what is required and,

by showing what needs to be done. This is inline with research carried out by Leach (2003). One of the factors that influence user security behaviour is what they are told. In most organisations this takes the form of security policies and security awareness initiatives (Leach, 2003:686). Another influencing factor in this regard is what employees see around them. Employees are strongly influenced by their peers and the messages that are released by the organisation whether internally or externally. If they see inconsistencies and contradictions between the message and the actual behaviour of the organisation, this will ultimately influence their behaviour (Leach, 2003:687).

4 THE NEED FOR FURTHER RESEARCH

According to Dhillon (1999), increasing awareness of security issues is the most cost-effective control that an organization can implement. Research that contributes to the effectiveness of awareness will ultimately benefit organisations as a whole as it will allow them to focus on techniques that improve their employees' intentions and ultimately encourage end-user security behaviours towards a more benevolent state. The research by Stanton et. al. (2005:132) implies that further research is needed in this respect as existing research does not address this appropriately. Diverse methods for measuring these different behaviours are also called for. This is needed since some behaviour may be easier to measure than others. Instruments that measure the behaviour of a database administrator (high technical expertise) that possesses malicious intent may be much more difficult than measuring behaviours that are more naïve in nature such as abuse of Internet access for example.

Kruger & Keaney (2005) developed a prototype for measuring the effectiveness of a security awareness program that was delivered in a global organisation. The model developed was based on three dimensions that could be measured i.e. what a person knows (knowledge), how they feel about a topic (attitude) and, what they do (a person's intention to act in a certain manner). These dimensions were measured to determine the effectiveness of their awareness programme. Information was gathered using questionnaires (including assessing behaviour) although they suggested using system data at a later stage. Thus actual behaviours of the employees were not measured to determine whether a difference was made. Kruger et. al. (2006) also recommends that system data be gathered to

supplement employee surveyed data and propose a basic list of source data from systems that could be used and for what purpose.

In some cases, researchers have however measured end-user behaviour directly, but this has been mainly geared towards how they respond to Internet-based threats, for example, the work carried out by Kumaraguru et. al. (2007), Jagatic et. al. (2007) and Whalen & Inkpen (2005). However the instruments used in these studies to measure certain behaviours may not be appropriate and practical for organisations to implement, such as those used by Whalen & Inkpen (2005). Learning science principles should be used (such as providing immediate feedback when incorrect behaviour is observed) and emphasis should be placed on the quality of awareness material as well as unique ways to deliver the message to end-users (Kumaraguru et.al., 2007). This is important since a lot is expected from users during awareness initiatives i.e. their time and attention, as well as expecting them to absorb the message. Srikwan & Jakobsson (2007), call for educational efforts to demonstrate and place emphasis on the link between behaviour and the outcome of that behaviour as they contend that mechanisms that support such a link "appears to offer significant benefits". Users must understand not only what they must do but why (Srikwan & Jakobsson, 2007:5).

The subject-expectancy effect, where a research subject expects a certain result, and therefore unconsciously affects the outcome of the results, are experienced by many surveys, such as the CSI survey mentioned above. Another example is the PayPal survey (PayPal, 2007) which provides a very good online questionnaire for users to test their understanding of phishing threats and how they work. Once again this type of survey however, does not measure actual behaviour.

Puhakainen (2006:69,139), points out that the only empirical evidence that does exist (with respect to information security awareness research) shows the practical effectiveness of deterrence. Further empirical evidence showing the effectiveness of security awareness training or awareness campaigns is not available, even though the effectiveness of training and campaign activities has been shown in other fields (for example, in cases where AIDS training has been a successful intervention).

Furthermore, scholars have pointed out that only a few existing studies are theoretically grounded (Puhakainen, 2006:149; Pahnila et. al. (2007)) and more work is needed in this regard. Security awareness research in this

context can be categorised as follows: conceptual models providing practical guidance for security awareness, theoretical models without empirical support and, theoretical models with empirical support (Pahnila et. al (2007)).

In an attempt to address the shortcomings and limitations of existing research, Puhakainen (2006) therefore developed three design theories to explain and improve IS Security behaviour. One of the design theories for IS Security awareness training was tested in two organisations. The research showed that the developed theory was relevant for developing practical security awareness training programmes. The researchers relied on the feedback from users, their colleagues and what they observed to determine the effectiveness of the security awareness training programme. This programme was shown to: achieve positive results, change user attitudes and, make users more conscious about their behaviour. The author calls for more practical studies in this regard (Puhakainen, 2006:106, 114, 139).

5 THEORETICAL UNDERPINNING

As mentioned above, a theory of security awareness is needed for researchers and practitioners to understand the expected outcomes of a particular awareness initiative and why this occurs. The model to explain security awareness training is based on work carried out by Nonaka & Takeuchi (1995). They argue that there are two types of knowledge and both are needed to help explain organisational learning, i.e. tacit knowledge and explicit knowledge. They propose that an organisation learns by oscillating between the two types of knowledge (Nonaka & Takeuchi, 1995:61). Tacit knowledge is not tangible and is subjective since it is that which is possessed by employees of the organisation. This includes individual beliefs, experiences and understandings of the organisation and what the organisation requires from them. Explicit knowledge on the other hand is codified, formal and easily expressed. Examples of this are organisational policies and pamphlets. Nonaka & Takeuchi (1995:70, 71) argue that the learning path in an organisation follows four cyclical stages:

- Employees share tacit knowledge;
- Tacit knowledge is made explicit by formalising it (e.g. policies);
- Formalised knowledge is disseminated (e.g. awareness activities) and,
- Employees “learn by doing” and thus explicit knowledge is made tacit by employees internalising it.

The cycle then starts from stage 1 again and follows an infinite loop. The study in progress described in this paper proposes a theoretical model to help explain how awareness training influences behaviour. The study in progress aims to show that in order to ensure appropriate security behaviour, employees need explicit knowledge of security policies and tacit knowledge on how to enact the appropriate security behaviour.

Figure 3 below puts the model in context and shows the actual mechanisms that will be tested. Firstly, users will undergo security awareness training (1). This will be in the form of security awareness material that will be exposed to users showing correct and incorrect behaviours. Thus the security message will be made explicit and disseminated to users (2). As argued above, explicit knowledge also needs to be made tacit by users internalising it. So, after the awareness material is presented, users will be required to write a short test that will measure to what extent the message has been internalised (3). Thereafter, the actual behaviour of respondents are measured to test whether their actual behaviour has changed due to awareness training (4) and, whether internalized knowledge (comprehension) is needed for appropriate behaviour (5).

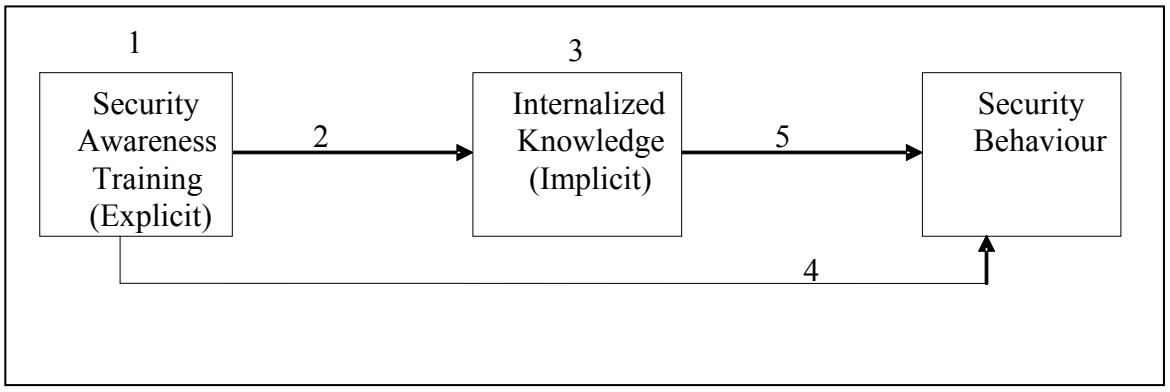


Figure 3. Theoretical model explaining how security awareness training affects behaviour.

6 RESEARCH AGENDA AND IMPLICATIONS

The previous section presents a theoretical model explaining how the authors expect security awareness training to affect behaviour. This section will put forward a research agenda for scholars and practitioners to explore further.

Security awareness training should influence all employees within an organisation to ensure the appropriate behaviour is enacted by all and thereby achieve compliance to information security policies. To confirm this, the following questions should be further explored: In terms of explicit knowledge, what type of security awareness training is more likely to influence behaviour i.e. how important is the quality of the awareness material and the mechanism of delivery? How could practitioners more easily deliver the awareness message to ensure greater participation from end-users? Standardised, cost-effective and automated mechanisms for gathering system generated data (especially for behaviours requiring high levels of expertise) and the feasibility of such mechanisms require additional investigation. In terms of implicit knowledge, further standardised mechanisms should be explored to determine how best to measure implicit knowledge taking into account the role of learning science principles. What are the most effective learning principles and under what conditions are they effective? Status of employees within the organisation and the role that plays in awareness training is important to determine in future research. Once users fully comprehend policies, are the same types of interventions necessary to sustain the required behaviours? This is important as it will likely determine how often awareness interventions are required. Longitudinal studies in this regard would be necessary. An understanding of the influence of factors such as user attitude, perceptions and corporate politics on internalisation of the security awareness message and subsequent behaviour is also needed. Finally, further research is needed on a taxonomy of security behaviours, building on the work of Stanton et. al. (2005).

The implications for practitioners are potentially significant. In order for organisations to implement affective Information Security an understanding from all employees within an organisation is needed. In addition, compliance to these policies is necessary and in some cases needs to be demonstrated by the Information Security function or Risk Management function within an organisation to justify their activities. The outcome of the current study will potentially provide pragmatic guidance for practitioners

when designing and implementing their information security awareness programmes.

7 CONCLUSION

There is a shortage of research on behavioural information security and theoretical models explaining how awareness training affects behaviour. The study in progress builds on existing behavioural information security research and puts forward a theoretical model, based on an organisational learning model. This theoretical model explains how organisational learning takes place, showing that both explicit knowledge and implicit knowledge is needed. The research underway will test the proposed model using system-generated data as indicators of behaviour in a pretest-posttest experimental design. Only a subset of behaviours (based on a typical Acceptable Usage Policy) that require low technical expertise on the part of the end-user will be tested. The objective of this research is to determine the effectiveness of information security awareness training on subsequent behaviour by users in the study. Such a model could help scholars and practitioners understand why an awareness initiative is expected to have certain results on security behaviour and consequently, provide practitioners with practical guidance for their information security programmes.

8 REFERENCES

Computer Security Institute (CSI). 2006. Virus Attacks Named Leading Culprit of Financial Loss by U.S. Companies in 2006 CSI/FBI Computer Crime and Security Survey [online]. [Accessed 9th August 2006]. Available from World Wide Web: <http://www.gocsi.com/press/20060712.jhtml>

Computer Security Institute (CSI). 2007. The 12th Annual Computer Crime and Security Survey [online]. [Accessed: 2007]. Available from World Wide Web: http://www.gocsi.com/forms/csi_survey.jhtml

Dhillon, G. 1999. Managing and controlling computer misuse. Information Management & Computer Security, Vol. 7, Issue 4, pp/ 171-175.

Evers, J. 2006. Security Expert: User education is pointless [online]. [Accessed 2007]. Available from World Wide Web: <http://www.news.com/Security-expert-User-education-is-pointless/2100-7350_3-6125213.html?tag=item>

Furnell, S., Sanders, PW., Warren, MJ. Addressing IS security training and awareness within the European healthcare community. In Proceedings of Medical Informatics Europe '97. 1997.

Furnell, S. M., Gennatou, M., Dowland, P.S. Promoting security awareness and training within small organizations. Proceedings of the First Australian Information Security Management Workshop, Geelong, Australia, 2000.

Furnell, SM., Why users cannot use security. Computers & Security 24,

4, 2005, 274-279.

Gaunt, N., Installing an appropriate IS security policy in hospitals. International Journal of Medical Informatics, 1998, 131-134.

Gaunt N. 2000. Practical approaches to creating a security culture. International Journal of Medical Informatics, 60(2), pp 151-157.

Jagatic, TN., Johnson, M., Jakobsson, M., Menczer, F. Social Phishing.

Communications of the ACM. Vol. 50, Issue 10, 2007, pp. 96 – 100.

Janczewski L, Xinli Shi F. 2002. Development of Information Security Baselines for Healthcare Information Systems in New Zealand. Computers & Security, Vol. 21, No. 2, pp 172 – 192.

Johnston, J., Eloff, JHP., Labuschagne, L. Security and human computer interfaces. Computers & Security, Volume 22, Issue 8, December 2003, Pages 675-684.

Katsikas SK. 2000. Health care management and information systems security: awareness, training or education? International Journal of Medical Informatics, Vol. 60, pp 129-135.

Kruger HA, Drevin L, Steyn T. A framework for evaluating ICT security awareness. Proceedings of the ISSA 2006 from Insight to Foresight Conference 5 July – 7 July 2006, Balalaika Hotel, Sandton, South Africa.

Kruger HA, Kearney WD. Measuring information security awareness: A West Africa Gold Mining environment case study. Peer-reviewed Proceedings of the ISSA 2005 New Knowledge Today Conference 29 June – 1 July 2005, Balalaika Hotel, Sandton, South Africa.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L., Hong, J., Nunge, E., Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System. Conference on Human Factors in Computing Systems archive. Proceedings of the SIGCHI conference on Human factors in computing systems. San Jose, California, USA, 2007, pp. 905 – 914.

Leach J. 2003. Improving user security behaviour. Computers & Security, Vol. 22, No. 8, pp 685-692.

McCoy, C., Fowler, RT. You are the key to security: establishing a successful security awareness program. In the proceedings of the SIGUCCS'04, Baltimore, Maryland, October 10-13, 2004, pp. 346-349.

Mitnick, KD., Simon WL., 2002. The Art of Deception: Controlling the Human Element of Security. John Wiley & Sons.

Nielsen, J. 2004. User education is not the answer to security problems [online]. Accessed [Accessed: 2007]. Available from World Wide Web: <http://www.useit.com/alertbox/20041025.html>

Nonaka, I, Takeuchi H. (1995), The Knowledge Creating Company, New York: Oxford University Press.

Nosworthy JD. 2000. Implementing Information Security In The 21st Century – Do You Have the Balancing Factors? Computers & Security, Vol. 19, pp 337 – 347.

OECD, *OECD Guidelines for the Security of Information Systems and Networks: Towards a Culture of Security* [online]. 2002. [Accessed: 2006]. Available from World Wide Web: <http://www.oecd.org/dataoecd/16/22/15582260.pdf>

Parker, DB. 1998. Fighting Computer Crime: A new Framework for Protecting Information. USA: John Wiley & Sons.

Pahnila, S., Siponen, M., Mahmood, A. Employees' Behavior towards IS Security Policy Compliance. Proceedings of the 40th Hawaii International Conference on System Sciences, 2007.

PayPal. 2007 [online]. Can you spot phishing [Accessed 2008]. Available from World Wide Web: <https://www.paypal.com/fightphishing >

Perry, WE. 1985. Management Strategies for Computer Security. USA:

Butterworth Publishers.

Puhakainen, P. 2006. *A Design theory for Information Security Awareness.* Ph.D. thesis, University of Oulu.

Ranum, M. 2005. *The six dumbest ideas in computer security* [online]. [Accessed: 2007]. Available from World Wide Web: <http://www.ranum.com/security/computer_security/editorials/dumb/>

Schneier B. (2003), Beyond Fear, Copernicus Books, New York

Schneier B. (2000), Secrets & Lies, Wiley Computer Publishing, New York

Schultz E. Security training and awareness—fitting a square peg in a round hole. Computers & Security, Vol. 23, Issue 1, 2001, pp 1 – 2.

Siponen MT. A conceptual foundation for organizational information security awareness. Information Management & Computer Security, Vol. 8, Issue 1, 2000, pp. 31-41.

Siponen, MT. 2001. Five dimensions of Information Security Awareness. Computers and Society, Vol. 32, Issue 2, 2001, pp 24-29.

Sommers, K. Robinson, B., Security awareness training for students at Virginia Commonwealth University. In the proceedings of the SIGUCCS'04, Baltimore, Maryland, October 10-13, 2004, pp. 379-380.

Spurling, P., "Promoting security awareness and commitment", Information Management & Computer Security, 3, 2, 1995, 20-26.

Srikwan, S., Jakobsson, M. 2007. Using cartoons to teach Internet Security [online]. [Accessed 2007]. Available from World Wide Web: <http://www.informatics.indiana.edu/markus/documents/security-education.pdf>

Stanton, J. M., Stam, K. R., Guzman, I., and Caldera, C. 2003. Examining the linkage between organizational commitment and information security. Proceedings of the IEEE Systems, Man, and Cybernetics Conference. Washington, DC.

Stanton, J.M., Stam, K.R., Mastrangelo,. Jolton, J. Analysis of end user security behaviours. Computers & Security, Vol. 24, 2005, pp 124-133.

Thomson ME, von Solms R. Information Security Awareness: Educating your users effectively. Information Management & Computer Security, Vol. 6, Issue 4, 1998, 167-173.

Trompeter, CM., Eloff, JHP. A framework for the implementation of socio-ethical controls in information security. Computers & Security, Vol. 20, 2001, pp 384-391.

Van Niekerk J, von Solms R. Organisational learning models for information security. Peer-reviewed Proceedings of the ISSA 2004 enabling tomorrow conference 30 June – 2 July 2004, Gallagher Estate, Midrand.

Von Solms B. Information Security – The Third Wave. Computers & Security, Vol. 19, 2000, pp 615 – 620.

Von Solms B. Information Security - A multidimensional Discipline. Computers & Security, Vol. 20, 2001, pp 504 – 508.

Vroom C, Von Solms R. 2004. Towards information security behavioural compliance. Computers & Security, 23, pp 191 – 198.

Whalen T., Inkpen, KM., Gathering evidence: use of visual security cues in web browsers. ACM International Conference Proceeding Series, Vol. 112. In Proceedings of the 2005 Conference on Graphics interface, Victoria, British Columbia, May 09 - 11, 2005, pp. 137–144.

LESSONS LEARNT IN THE PROCESS OF COMPUTERIZATION, AUTOMATION AND MANAGEMENT OF ICT SECURITY IN THE DEVELOPING WORLD: A CASE STUDY OF THE UNIVERSITY OF DAR ES SALAAM, TANZANIA

Geoffrey Karokola[1] and Louise Yngström[2]

Department of Computer and System Sciences Stockholm University/Royal Institute of Technology
Forum 100, SE-164 40 Kista, Sweden Tel: +46 (0)8 16 1697, Fax: +46 (0)8 703 90 25 E-mails: {karokola1, louise2}@dsv.su.se

ABSTRACT
This paper intends to discuss and sift out current and important challenges in Information and Communication Technology (ICT) security for developing countries in the Sub-Saharan Africa where Tanzania will be taken as a case study. As a background we analyze lessons learnt in the processes of computerization, automation and the management of ICT security at the University of Dar es Salaam (UDSM) since it is one of the first higher learning institutions in Tanzania. The backbone of UDSM currently connects more than three thousand workstations and twenty five heavy duty servers that are centrally managed and which support different institutional core services.

In the evolution process of computerization and automation of Information and Communication Technology (ICT) at the UDSM that started way back in the early 1990's ICT security was of no priority. While in the western world computerization and automation processes have gradually been incorporating security into ICT infrastructures, developing countries have not experienced a similar evolution – neither in technical nor in practical circumstances. In practice, developing countries need to conform to international developments within ICT security at the same time as they are trying to conform to their own environments and also learn about

the totally new situation created. Simultaneously there are also local and specific restrictions – well known by the developing countries -but usually not experienced by the developed world.

KEY WORDS
Challenges, Lessons learnt, ICT Security, Information Security, Automation, Computerisation, Managing, Developing World

LESSONS LEARNT IN THE PROCESS OF COMPUTERIZATION, AUTOMATION AND MANAGEMENT OF ICT SECURITY IN THE DEVELOPING WORLD: A CASE STUDY OF THE UNIVERSITY OF DAR ES SALAAM, TANZANIA

1 INTRODUCTION

Information and Communication Technology (ICT) is considered to be a major driving force of globalised and knowledge based society in a modern world. As technology remains dynamic, protection of information asserts has became very challenging. A number of attacking techniques exists including denial of service attacks, cross site scripting, content spoofing, phishing, man-in¬the-middle, and brute-force. Therefore, proper protection of information assets in ICT infrastructures is needed. ICT security is considered to be part of that, where confidentiality, integrity and availability to information assets are the three pillars of major concern.

In developed countries the evolution process of computerisation and automation of ICT infrastructures gradually integrates ICT security. However, staring from early 1990's most of developing countries, particularly sub-Saharan Africa has experienced hasty un-secure evolution processes in computerisation and automation of ICT infrastructures [1]. There are critical factors in the developing world that negatively influence the process including lack of awareness and security culture, lack of knowledgeable and experienced human resources in managing ICT facilities, and un-secure integration of ICT security. While in the western world computerization and automation processes have gradually been incorporating security into ICT infrastructures, developing world has neither technical nor practical experience in similar evolution. As a result many organisation and institutions in developing world experienced losses of potential synergies [2, 3].

In this study, University of Dar es Salaam (UDSM) being one of the key player in ICT security in Tanzania and one of the first and leading higher learning institutions is taken as a case study. Furthermore, we

analyze and sift out challenges and lessons learnt in the processes of computerization, automation and managing ICT security at UDSM.

The paper is organised as follows: the background to the studied environment and an ICT security overview in Tanzania is given in chapter one; chapter two presents methodology; chapter three presents ICT security implementation status, chapter four presents challenges and counter¬measures; chapter five presents discussion and lessons learnt. Lastly conclusion and recommendations are given in chapter six.

1.1 Background to the Studied Environment

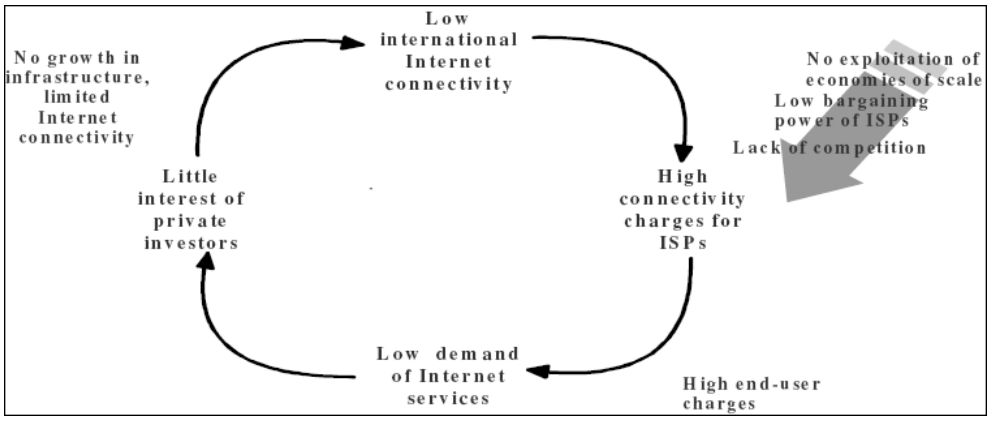
The University of Dar es Salaam (UDSM) is one of the first and leading public higher learning institutions in Tanzania. UDSM was firstly established in 1961 as an affiliated college of the University of London. In 1963 UDSM become the constituent college of University of East Africa and in1970 an independent university [1, 7, 12]. The primary objectives of UDSM were: to transfer knowledge from one generation to another; to establish a place in Tanzania where frontiers of knowledge would be advanced through research; and to be a place where professional training of human resource would be conducted [1, 7, 9]. The university started with only one faculty, the faculty of Law. Gradually it expanded to include more than five campuses; six faculties, four centres, and four institutes – at the main campus alone; and two constituent colleges [1, 7, 12]. In terms of student's enrolment, in 2006/2007 the university has a total of 18,342 students with females constituting 36.1% of all undergraduates [5].

The computerization, automation and integration of ICT security process at the university started in the early 1990's. Apart from many challenges posed as a result of this process over 16 years (1990's – 2007), great achievements were realised -as it is presented and discussed in the paper. The university recognition in offering high quality education and related ICT-based services is recognised amongst the African countries; in the African Universities ranking of 2005 conducted by Webometrics Ranking of World Universities: Cybermetrics Lab, National Research Council, Spain, UDSM appeared to be in the thirteenth position using ICT for teaching and learning [14]. All ICT-based services at UDSM are centrally coordinated and managed by the University Computing Centre (UCC) [1, 7, 9].

1.2 ICT Overview in Tanzania

Confidentiality, Integrity and service Availability are the three pillars for ICT security and special attention needs to be paid to bandwidth being one of the major contributing factors. In most African countries the internet connectivity to the outside world is satellite dependent. As a result of bandwidth charges in Africa being very high, automatically the integration of ICT security in universities and quality service delivery is highly affected [15].

According to the International Telecommunication Union (ITU) report, they argue that higher bandwidth prices in African countries are notably influenced by certain bottlenecks elements. Lack of infrastructures, unfavourable regulatory environment, and uncompetitive market structure were mentioned as influencing elements [4]. The report also gave the economics and consequences of high bandwidth prices to end-users, which are summarised and presented in the figure below.



Source: ITU report on improving IP connectivity in the least developing countries [4]

From these facts, it is obviously that with the existing wider economical gap between developing and developed countries, these higher bandwidth charges are unaffordable, and widens the digital divide. Cheaper and affordable bandwidth is necessary for African HEI's to excel.

The government through the Ministry of Communication and Transport (MoCT) and Tanzania Communications Regulatory Authority (TCRA) formally Tanzania Communication Commission (TCC) is also

advocating availability of cheaper and affordable bandwidth where as regulations permit international VSAT data service providers [2, 6]. To-date Tanzania has a total of fifty-three registered ISP's companies: eight companies with Network Facilities Licenses; eight with Network Service Licenses; and thirty-seven with Application Service Licenses [6]. These ISP's are offering internet services via VSATs', Wireless, leased lines, and dialup.

With regards to computers, the importation of ICT facilities, including computers started in early 1970's. However, there were a number of problems associated with operations, maintenance and management of ICT facilities. As a result the government experiences heavy financial losses. To stop the losses in 1974 the government decided to ban the importation of computers and its related equipments [2]. In early 1980's the ban were lifted. Most of computers and ICT related facilities were then imported by the government, private companies and few individuals [3]. In order to promote the growth, use and affordability of ICT facilities in the country, in early 2000 the government decided to wave taxes to the importation of computers. Since then the dependences on ICT to operate core services in all sectors increased hasty. Similarly status of ICT security awareness among people and its integration to support core business is fairly high.

2 METHODOLOGY

The study is based on the literature review, research work and findings from the four PhD graduates on ICT security paradigm in the studied environment [3, 8, 10, 11]. The study also grips authors working experience of nearly ten years as a forefront in the implementation and management of ICT security in the area.

3 ICT SECURITY IMPLEMENTATION STATUS

This chapter presents UDSM ICT-based services growth trends over the span of nearly sixteen years (1990's -2007) and how ICT security was integrated in the evolution process.

3.1 ICT Network Infrastructure and its Facilities

Having a secure, reliable, and well managed ICT network infrastructure in any organisation is a necessity for high quality service delivery. Security mechanisms including intrusion detection systems (IDS), firewalls, routers,

and virtual LAN (VLAN) are used to secure network infrastructures from attackers who exploit network vulnerabilities.

UDSM being the place where professional training of human resource is conducted -was not left out in ICT development arena. During 1990's – 2007 university progressively implemented the state of the art ICT network infrastructure that consists of gigabit speed optical fibre backbone, wireless links, and structured LAN's at UDSM main campus, constituent's colleges and at its institutes (UCLAS, MUCHS, DUCE and IJMC)[1]. To enhance distance learning -videoconferencing facilities were also installed [1, 7, 9, 12, 17]. Some of the areas covered in the process were:

- Optic fibre backbone: all buildings, including student's halls of residence and Public access Rooms (PAR)[2] at UDSM main campus were linked to the university optical fibre backbone. Also UDSM backbone was extended to UCLAS located more than 2km from the main campus. Optical fibre backbone(s) were also installed at UCLAS, MUCHS, and DUCE.
- Point to point wireless link networks: Other institutes and hall of residents (MUCHS, DUCE, IJMC and Mabibo hostel) located far from the UDSM main campus were connected to UDSM backbone via wireless microwave links with 11-23Mbps capacity.
- Wireless Access Points (Wi-Fi): Both outdoors and indoors WPA's were installed at the UDSM main campus and DUCE for creating flexibility and mobility to staff and students.

[1] UCLAS - University College of Lands and Architectural Studies; MUCHS - Muhimbili University College of Health Sciences; DUCE - Dar es Salaam University College of Education; IJMC – Institute of Journalism and Mass Communication.

[2] Rooms located to the student's halls of residence providing ICT related services to students; services include internet access and printing.

- Videoconferencing: Four UDSM main campus lecture theatres were installed with video conferencing facilities, and two sets of mobile facilities are available for use. However, in the process ICT security implementation and integration to automated services were observed to be of ad-hoc character. To-date UDSM backbone network infrastructure is believed to be one of the best heterogeneous network in higher learning institutions in eastern Africa.

3.2 Computers and its Facilities

Existing state of the art network infrastructure will be curtailed without well secured end-user computers and servers' machines -that allows academia and other university staff to access and utilize fully available ICT-based services and resources.

Taking advantage of the government decision on computer importation ban lifting and tax waving [2], UDSM has installed more than 3,000 computers at her main campus. Computers of different brands include Dell, Mac, Sun, Compaq, HP, and Siemens [1, 7, 12]. The table below delineates the trend of ICT equipments growth and its distribution within UDSM main campus.

Table 1: Trend of ICT facilities growth (Computers). Source: UDSM ICT policy, Master plan & UDSM website [1, 7, 12]

Year	Average Number of Computers	Location and Usage Description
1990	17	Mostly located at administration building, deans offices, and head of departments. Very few were located in faculties computer labs, and main library
1995	200	Nearly all administration building offices, dean's offices, head of departments, and academia offices. Faculties labs, few in departmental labs, and main Library
2002	2400	Nearly all administration building offices, dean's offices, head of departments and sections, and academia offices. Faculties' labs, departmental labs, and main Library. PAR's at student's hall of residence: hall 1, 5, 7 and Mabibo hostel each with at least twenty computers and one printer.
2007	3,200	More computers deployment were in offices and departmental new Labs *(AVU-LC, computer science department etc)*

Apart from the progress made, protection of these ICT facilities was affected by a number of issues including lack of enforcing security measures, maintenance culture, and ICT facilities failures. As a result

confidentiality, integrity and availability of sensitive information assets stored in these computers/ servers was jeopardised.

3.3 Bandwidth and Utilisation Status

UDSM was the first HEI in Tanzania to have dial-up connection (from London) that was purely used for sending and receiving emails once a day. The service turned out to be not only a burden to the university as a result of higher telephone connection charges but also the limited number of ICT services offered to the community [7]. As an alternative to that, UDSM gradually managed to upgrade her bandwidth from different providers as delineated in Tables 2 below.

However, running cost remains a challenge, for instance UDSM used to pay monthly subscription fee amounting to 9,000 US$ for 1/2Mbps from TTCL. Currently UDSM is paying around 11,000 US$ (subsidised rate) for 1.5/7.5Mbps from AVU [1]. Following that -to maximize bandwidth utilization, UDSM employed a lop-sided link bandwidth strategy "thin up/fat down" as less bandwidth is required to send data to the Internet and more to receive large data [15]. The table below summarises bandwidth upgrading trend at UDSM.

Table 2: Trend of bandwidth growth over the period starting from early 1990's to date. Source: UDSM ICT Policy Master plan, UCCICT and PHEA [1, 7, 9, 15]

Year	Bandwidth (Mbps)	Total Bandwidth (Mbps)	Connection Type	ISP	Usage Description
1990/93	< 0.024	< 0.024	Dial-up	Heath net	Only for sending and receiving emails
1993/97	0.256/0.512	0.768	Leased Line	TTCL	Internet /Email, Research, Library and few networked computers,
1998/2000	0.512/1.024	1.536	Leased Line	TTCL	Internet /Email, Research, Library and networked computers,
2001/06	1/2	3	Leased Line	TTCL	Internet /Email, Research, Library, online services, and networked computers
2006 –	1.5/7.5	9	VSAT	AVU	Internet /Email, Research, Library, online services , and networked computers

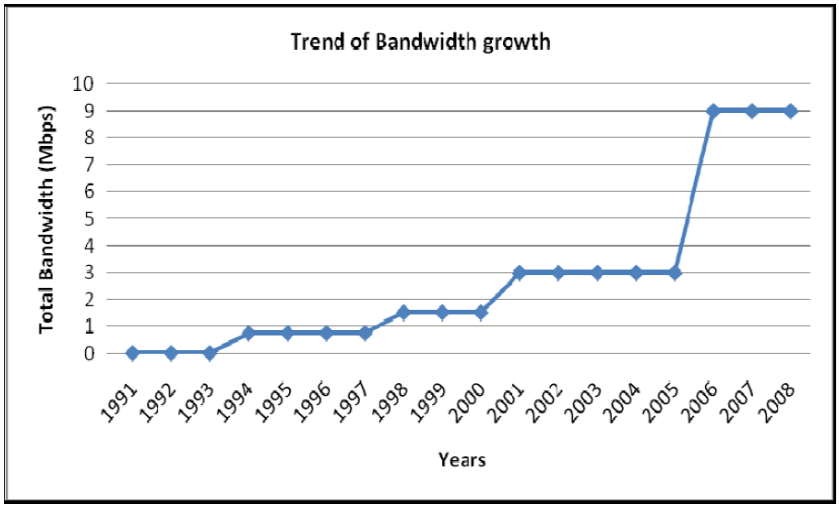


Figure 1: Bandwidth growth starting from 1990's to date

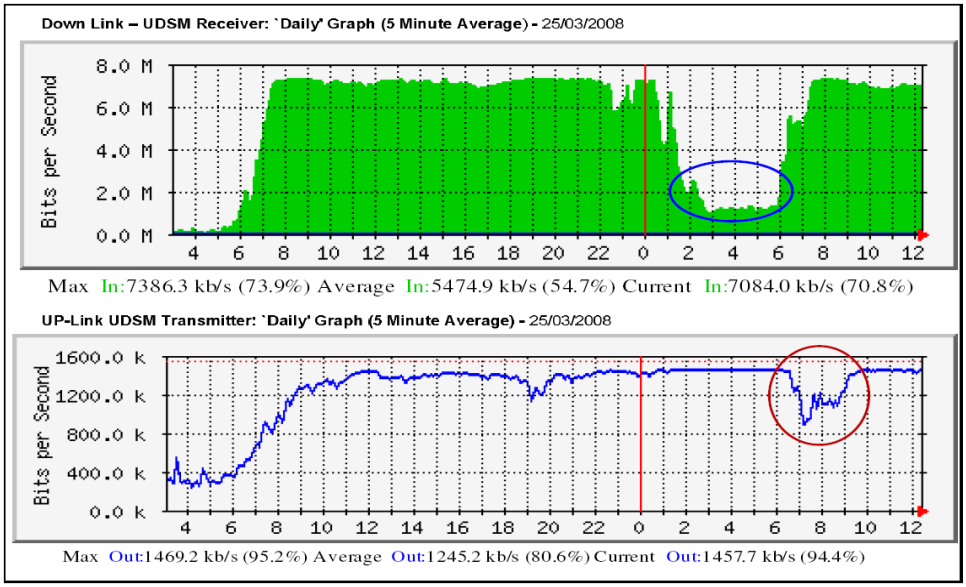


Figure 2: MRTG graphs showing a daily UDSM bandwidth utilisation

Figure 1 shows the bandwidth growth trend while figure 2 shows MRTG daily bandwidth monitoring utilisation graph with down-link of 7.5Mbps and up-link of 1.5Mbps which is over 73% and 95% respectively saturated[3]. UDSM being connected to the outside world via internet was a start for facing new avenues including enhancing teaching, learning, research functions, and library services. However, this process also opens doors to security threats and attacks. A number of measures are taken to minimize these new risks.

3.4 Software's -Operating Systems, Application and Anti-virus

Operating systems (OS) and application programs are the most critical programs for ICT facilities. Malicious code, corruption/destruction of data, and unauthorized change of access rights and privileges are few examples of attacks. To enhance security to these critical assets -awareness, knowledge and technical expertise, and ICT security policies – defining (what, why, how, do's, and don'ts) need to be in place. Patching of operating systems, application programs and use of antivirus solutions are part of security measures that protect ICT assets. As mentioned earlier, the growth in numbers of ICT facilities meant increased software requirements and security demands. Also, more proprietary software's platforms were needed, which demanded UDSM to pay more license fees to vendors while at the same time more bandwidth was required to facilitate downloading and

[3] The circled area in both graphs shows that usage decreases only during the night (from 1 to around 6 hrs). The interpretation tells that most of students do use their laptops to access internet via wireless access points installed across campus while others (including staff family member) who lives in campus do go to Public Access Rooms (PAR) located within the student's hall of resident. Usually at UDSM main campus most of student's departmental computer labs are closed at 20.00hrs with the exception of main library which is closed at 22.00 hrs.

updating of software patches to both servers supporting core services and to end-user computers.

To cut down running cost on software's license fees, UDSM decided to opt for open -access platforms software's – including Linux as OS and Star-office as application programs. This decision was also included in the UDSM ICT policy and master plan [1, 7]. To-date, a fairly larger number of end-user computers and servers are running on open -access platforms [1].

3.5 Information Systems, Online and Library Services

Proper protection procedures against threats that may cause risks to information assets need to be in-place. At UDSM, computerisation and automation process of ICT-based core services and integration of ICT security was not an easy undertaking. As discussed earlier – lack of awareness, knowledge, technical experience and expertise on ICT security were among the major constraints to the process. Thus to ensure that information systems are securely implemented -UDSM had to train her IT staff and was sometimes forced to hire experienced experts / personnel's from abroad [1, 7, 9, 16].

To-date UDSM network infrastructures including existing information systems that are accessible online have fairly high security level. Attackers usually use techniques including cross-site scripting, cookies poisoning & hijacking, content spoofing, denial of service attacks, phishing, man-in-the-middle, or brute-force to exploit vulnerabilities and cause risk/ damage to information assets. Some of the security techniques implemented at UDSM to protect information systems include encryptions, authentication, transport layer security protocols, public key infrastructures – PKI, and virtual private networks. Also automatic backup and disaster recovery solutions were implemented to ensure data availability.

Some of the existing information systems that supports teaching, learning research functions, library and administration services are:

- Teaching/learning: Online systems like Blackboard etc.
- Online Laboratories: iLab – based on real time
- Online Library Services: Online Public Access Catalogue (OPAC), Library Information System (LIBIS), UDSM virtual library, Database for African Theses and Dissertation (DATAD), and Internet search engines and information gateways

- Admission and Examinations: UDSM has developed in-house Academic Registry Information System (ARIS) based on open – access
- Administration: Human Resource Information System (HURIS), and Financial Information System (FIS)
- E-mails and Intranet Services: A number of email services are available.

Apart from success made, still UDSM is facing a number of security challenges that needs special attention.

4 CHALLENGES AND COUNTERMEASURES

The evaluation of computerisation, automation and management of ICT security at UDSM was affected by a number of challenges. These challenges affected much of the fusion and integration process of ICT security with ICT network infrastructure. We categorise the current critical challenges in the following manner: Awareness and capacity building, social-culture, economical, regulatory, technological, power supply, bandwidth and countermeasures.

4.1 Awareness and Capacity Building

Under this category the following were identified as the most critical issues: lack of ICT security awareness and culture among end-users; lack of user knowledge to proper use of ICT facilities including computers – that leads to violation of security procedures; lack of knowledge, practical and technical experiences to IT staff on the higher level implementation and management of both ICT network infrastructures and ICT security.

Others were: lack of technical expertise in defining security requirements and specifications for software's and hardware's before procuring and/or development; inadequate ability to quickly adjust and respond to rapid ICT technological changes – mostly of technological changes occurs in software's development and change of versions; and lack of expertise in maintaining different ICT components/ equipments – notably these requires specialised skills.

4.2 Social-culture

Social-culture behaviour was also seen to be among the challenges in the studied environment. These include: Presence of vandalism on ICT infrastructures and facilities that cause(s) heavy financial loss and service interruption – theft of ICT facilities including network components; and

lack of maintenance culture of ICT facilities among end-users, management and/or decision makers.

4.3 Economical

In this class the following challenges exist: Presence of high experienced IT staff turnover – as a result of high market demand; limited funding that would support proper implementation and management of ICT securities from the institutions and/or government; presence of higher software maintenance and license fees for use of proprietary software's[4]; and presence of higher bandwidth charges[5].

4.4 Regulatory

Regulatory issues were also of concern. Lack of properly defined ICT security policies, procedures and guidelines; and presence of limited support and commitment from the government and/or regulatory bodies on ICT related issues; are some of the challenges that do exist.

4.5 Power Supply

Stable and reliable power supply is a prerequisite needed to smoothly run and operate ICT facilities/ equipments, else survivability of ICT equipments and service availability are jeopardised. At UDSM lack of reliable power supply from the national grid that may cause facilities and systems failure and service interruption, and presence of limited number of un-interrupted power supplies (UPS) to support computers and its facilities are existing challenges.

---

[4] For instance - UDSM pays every year an annual subscription fees for HURIS, FIS and LIBIS amounting to 7,260.00 US$, 26,303.00 US$ and 5,193.51 Euro respectively [7]

[5] "... in Europe and North America a bandwidth that cost 100 US$ a month would cost African universities more that 10,000 US$ a moth..." [15]

4.6 Technology and Bandwidth

ICT is dynamic, this creating a lot of challenges to the ICT security paradigm. These challenges include: presence of malicious codes and alike, including attackers / hackers – exploits network vulnerabilities and cause heavy damages/losses to valuable information assets; limited bandwidth ¬that affects core services availability; and Systems complexity – as most of the systems became more complex, they become more demanding and require high skills to use. Others are presence of fake ICT related equipments in the market and/or vendors – leading to systems failures, service interruption and financial losses; and rapid ICT technological changes.

4.7 Countermeasures to Critical Challenges

To address cited critical challenges, UDSM implemented various countermeasures to mitigate risks and associated damages. These countermeasures are presented and discussed in this section.

4.7.1 Awareness and Capacity Building

As presented, awareness and capacity building is the cross cutting factor that influenced ICT security at UDSM. The programmes for awareness creation are in place. For instance during the start of new academic year all new admitted students are oriented with “don’ts” and “do’s” on use of ICT facilities at UDSM including security issues. However the challenge remains as the students have different backgrounds on ICT knowledge. Also seminars and short-courses related to ICT awareness and use at UDSM are conducted on regular basis to staff at all levels [1, 9].

To build capacity for IT staff, UDSM trained a number of IT staff within and outside the country including professional training[6] and at academic level[7]. Best-practice-study-tours and in¬house were also introduced [1, 9]. For the past five years, ICT security courses have been integrated to some diploma, degree, postgraduates and masters programmes. Also computer literacy courses are mandatory to every degree program at UDSM. Some of the programmes at UDSM that have included ICT security in their curricular continuum are listed below [5, 9, 13]:

- Faculty of Informatics and Virtual Education (FIVE): Certificate and Diploma in Computer science(1/2 yrs); BSc in and BSc with Computer Science (3 yrs); BSc in Electronics and Communication (3 yrs); MSc in Computer Science (2 years); MSc. in Electronic Science and Communication (2 yrs); MSc in Health Informatics (2 yrs); and PhD (3/4 yrs)
- Faculty of Science: Postgraduate Diploma in Scientific Computing (1 yr);
- Faculty of Electrical and Computer Systems Engineering (ECSE): BSc. in Computer Eng. and IT (4 yrs); BSc. in Telecommunications Engineering (4 yrs); Postgraduate Diploma in Electronic and IT (1 yr); MSc in Electronic Engineering and IT (2 yrs); and PhD (4 yrs)
- Faculty of Commerce and Management: Postgraduate Diploma in ICT Policy and Regulation; and Masters in ICT Policy and Regulation (2 yrs)
- University Computing Centre (UCC): Certificate and Diploma in Computing and IT (1/2 yrs). Professional courses: CISCO

[6] For instance -in July 2004, four IT staffs were trained in India at CCNP level; December 2005, three IT staffs were trained in MySQL database administration in Singapore.
[7] For instance -in 2000 seven IT staffs were trained to licentiate and four to PhD level in Sweden.

Internetworking (CCNA, CCNP); Microsoft (MCSA, MCSE etc); IT Essentials; Programming; Oracle (OCDBA, OCP etc) and professional certification.

In addition, the completions of four PhD's (2000 – 2007) in the area of ICT security, their findings and developed model/frameworks have contributed to improvements of ICT security status, not only at UDSM but also in the country [3, 8, 10, 11].

As of today, UDSM has good experience, technical capacity, and expertise in implementing and managing ICT security in network infrastructures. These achievements are due to rising awareness among UDSM community members, recruiting some graduates from the above listed programmes, involvement of the mentioned PhD's graduates and large support from the UDSM management and community.

4.7.2 Social-culture issues

UDSM introduced policies on physical security for protecting her ICT network infrastructures and its facilities. Checkpoints were introduced at the main gates and main buildings entrances where people declare their ICT-related belongings. The numbers of incidents have reasonably gone down. In addition, culture on adhering to maintenance schedules of ICT facilities/ equipments among decision makers is fairly high.

4.7.3 Economical issues

IT staff turnover and retention, UDSM has relatively increased IT staff salaries, as a result staff turnover ratio has significantly gone down to 4.7%. However, the challenge remains as the market salaries are still at high - which could impact staff retention strategies.

To reduce huge amount paid to vendors as maintenance and licenses fees for use of proprietary software's; UDSM has set a policy to migrate from proprietary to open-access software's. To date most of end-user computers (particularly in computer labs) are running on Linux and star-office programs. At the moment (2008) 98% of all servers at UDSM are running on open-access platforms. UDSM is now developing her own information system using open-access platforms; a good example is the developed academic registration information systems (ARIS).

Internet bandwidth charge still remains high despite efforts made by various existing initiatives. To-date UDSM is paying more than 11,000.00 US$ at subsidised rate a month for a total bandwidth of 9Mbps.

4.7.4 Policies and Regulatory

The first UDSM ICT policy and ICT master plan was developed in 1995. Likewise national ICT policy was developed in 2003. The existence of these documents has contributed to the improved of ICT security at UDSM. However, as ICT is very dynamic, challenges remains on proper translation of policy documents in to actions. Also to avoid ad-hoc, timely updating of policy documents to match with the current changes still needs great attention.

4.7.5 Power Supply issues

To mitigate the risks to stable and reliable power supply UDSM installed single and centralised un¬interrupted power supplies (UPS) in offices, computer labs and server rooms. Also automatic standby power generators were installed to strategic areas like theatre rooms, main library, administration and some faculty/departmental buildings. However, the challenge remains on the sustainability of maintaining and running these generators as they require periodical maintenance ¬fuel and spare parts – that requires funding.

4.7.6 Bandwidth and Technological issues

A rapid technological change has forced ICT-based service users always to be at alarming state. UDSM faced a lot of critical technological challenges as presented in section 4.6. However to mitigate the risk UDSM implemented a number of strictly measures to secure her network infrastructure and critical assets against threats that may exploit vulnerabilities and cause risk to information assets. Implementation of tools (hardware's and software's) and configuration techniques including intrusion detection systems (IDS), firewalls, routers, intelligence switches and virtual LAN (VLAN) as part of measures to secure the network infrastructure is done.

Also to enhance confidentiality, integrity, authenticity, authentication, accountability, and non-repudiation to web-based information systems security techniques including -encryption, authentication, TLS, public key infrastructures – PKI were also implemented. Furthermore, automatic and manual data back-up mechanisms, including disaster recovery solutions are

in place. Monitoring and management of network infrastructure is now automated. Tools like "What's-Up-Gold" are in use. The use of such tools has significantly simplified management of network infrastructures, where from one central location IT staffs are able to monitor the entire network depending on the configuration.

For control of viruses and malicious codes, UDSM has implemented a university wide Antivirus solution "escan corporate solution" which is centrally accessed. In addition, all up-to-date patches are also centrally kept and accessed -this technique facilitates easy access and timely availability to networked end-users computers, hence international bandwidth serving.

Addressing the bandwidth problem, despite of efforts made by UDSM to manage the little bandwidth it has, still challenge remains as ICT-based core services are affected. Some of counter¬measures that are in place include:

- Use of bandwidth manager: Internet bandwidth is now allocated per network segments (sub¬networks), this technique also facilitates retention of any un-usual generated traffic not to affect the rest of the network
- Traffic divergence: All UDSM local traffic generated from accessing of local emails and websites are routed through TIX[8]
- Patches and Updates: patches and updates files for operating systems, application programs, and ant-viruses are kept and accessed locally at the central servers. Authorised users are allowed to update their computer patches from the central servers (locally).

[8] Tanzania Internet Exchange (TIX) is a national internet exchange centre that keeps local traffic (local emails and websites) local. This leave international bandwidth to be used for other services.

- Blocking of international online web-based emails: yahoo, hotmail and alike are blocked during working hours where bandwidth is mostly needed for supporting core services. After working hours these web-based emails are allowed when much of the bandwidth is un-used.

Despite of all efforts made to manage sufficiently the little bandwidth UDSM has, still more bandwidth[9] is needed (approximately six times of the current bandwidth) to support more than 3,000 networked computers together with automated ICT-based core services.

5 DISCUSSION AND LESSONS LEARNT

The overall goal for implementing, fusion and managing ICT security into network infrastructures is to secure critical information assets. However, from the analyses we have seen that there are a number of current critical challenges that affect proper and secure implementation, fusion and management of ICT security paradigm. Furthermore, we have seen how UDSM critically addresses these challenges, though fairly few still remain. Based on the analyses and discussions on current challenges at UDSM – we sift out the following as lessons learnt:

- Higher bandwidth charges in Africa, undermines automation process of ICT-based core services and quality service delivery. Due to the limited bandwidth most HEI has, proper measures and policies on its utilisation are required. Bandwidth management techniques such as these implemented at UDSM (discussed in section 4.7.6) may be applied.
- Automation of network infrastructure management at UDSM enabled IT staff to easily manage and monitor the entire network from a single

[9] "...An institution with an average of 3,000 networked computers with automated ICT-based core services requires at least 66Mbps..." [15].

location. As discussed in section 4.7.6 – this has led to time and cost serving, and improved efficiency in managing ICT Security.

- From the discussion presented in section 4.7.1, we have seen that in order to create awareness and building internal capacity, special initiatives are needed. UDSM integrated computers and ICT security related courses/programmes at different levels within university academic curricular. These efforts facilitated also the generation of more IT security specialists.
- The study revealed that defining ICT security requirement specifications for hardware's and/or software's products remains a challenge. Proper attention should be given on developing measures to build internal capacity in the area.
- The discussion (section 3.4 and 4.7.6) shows that so far UDSM has successful implemented security mechanisms. The challenge remains as technology keeps changing – new threats and risk are always at alarming states.
- Dependence on proprietary software undermines development of HEI as much financial resources are required for payment of maintenance and annual subscription fees[10]. Therefore, for the survival of HEI, the migration strategies to open-access software are necessary.
- From the discussion (section 3.5) we have seen that UDSM managed to develop in-house ARIS system using open-access software. Thus HEI(s) could build their internal capacity to develop software in-house.
- Fusion of both ICT security policies to networks, email acceptable use etc, are necessary for enhancing information security at HEI. Top

[10] UDSM is paying every year an annual subscription fees for HURIS, FIS and LIBIS amounting to 7,260.00 US$, 26,303.00 US$ and 5,193.51 Euro respectively.

management support is highly needed for successful implementation and enforcement of these policies.

- ICT-based services should be available when needed. From the discussion in section 4.5, we have seen that UDSM managed to install automated standby power generators in strategic areas and most of end-users computers are connected on UPS(s) -single and centralised ones. Thus HEI(s) in Africa should invest in emergency power supplies.
- Selling of fake ICT equipments (especially in Africa) is emerging which leads to systems failure. The study revealed that -the problem is very challenging and needs to be handled collectively.
- IT staff turnover ratio at UDSM has significantly gone down to 4.7% by July 2004. Good salary and a conducive working environment are contributing factors to the success. HEI(s) in Africa should develop IT staff retention strategies.

In order to be able to visualise the relationship between specific elements in our analysis and discussion-it would be fruitful to identify in particular what actions the university itself can be and is in control of and what actions and items the university cannot control by itself but only influence within a longer time frame. Such an analysis is provided.

The proposed framework is based on internal and external elements. The internal elements are these activities that UDSM has control of itself; while the external ones are these that UDSM has no control of (influencing factors from the environment). The input elements are processed and the output is called ICT security management. Since there is no single solution for ICT security management, hence the input -output process is repetitive.

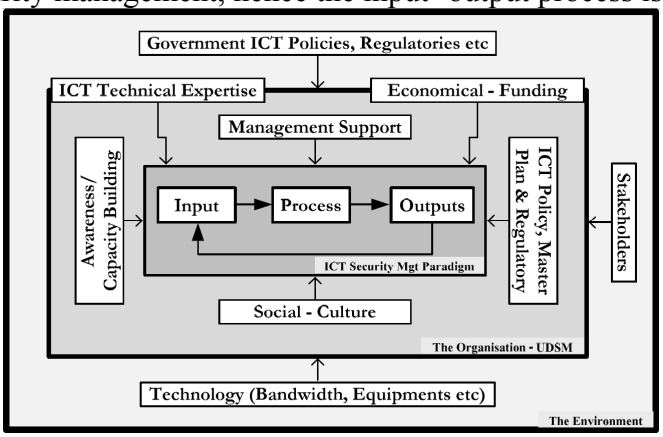


Figure 3: Proposed ICT Security Management Framework based on UDSM experience

6 CONCLUSION AND RECOMMENDATIONS

As information and communication technology is considered to be a major driving force of globalised and knowledge based society in the modern world – proper fusion and integration of ICT security to network infrastructures should be given higher attention. In the paper, the process of computerisation, automation and management of ICT security from early 1990 were presented. In the analysis -challenges were categorised into: Awareness and capacity building, Social-culture, Economical, Regulatory and Policies, Power supplies, and Technology and Bandwidth. Counter¬measures were discussed at length and lessons learnt were presented. In addition, the framework based on UDSM experience in ICT security management was developed and presented.

Generally we have seen that for better ICT service delivery -ICT security is highly needed for protection of critical information assets. Therefore, it is recommended that special attention should be given to the addressed issues that are still affecting ICT security paradigm. Furthermore, we believe that the presented UDSM experience in the area could also be adopted by other universities in the developing world.

7 REFERENCES

[1] UDSM-ICTP, University of Dar es Salaam, ICT Policy, May 2006

[2] TZ-ICT, Tanzania National ICT Policy, March 2003. Available at http://www.tanzania.go.tz/
[Accessed on 25th March, 2008]

[3] Bakari, Jabiri, "A Holistic Approach for Managing ICT Security in Non-Commercial Organisations: Case Study in a Developing Country", PhD Thesis, Department of Computer and Systems Sciences, University of Stockholm and the Royal Institute of Technology, Stockholm, 2007. ISBN: 91-7155-383-8

[4] ITU, "Improving IP connectivity in the least developing countries" (2002)

[5] UDSM-DUS, Directorate of undergraduate studies -UDSM "Undergraduate Programmes and Admission procedures" (2007)

[6] TCRA, Tanzania Communications Regulatory Authority website: http://www.tcra.go.tz
[Accessed on 25th March, 2008]

[7] UDSM-ICTM, University of Dar es Salaam, ICT Master Plan (2008 – 2012), September 2007

[8] Tarimo, Charles, "ICT Security Checklist for Developing Countries: A Social-Technical Approach", PhD Thesis, Department of Computer and Systems Sciences, University of Stockholm and the Royal Institute of Technology, Stockholm, 2006. ISBN: 91-7155-340-1

[9] UCC-ICT, University of Dar es Salaam-Computing Centre website: http://www.ucc.co.tz, [Accessed on 27th March, 2008]

[10] Casmir, Respickius, "A Dynamic and Adaptive Information Security Awareness (DAISA) Approach", PhD Thesis, Department of Computer and Systems Sciences, University of Stockholm and the Royal Institute of Technology, Stockholm, 2005. ISBN: 91-7155-154-9

[11] Chaula, Job, "A Social-Technical Analysis of Information Systems Security Assurance", PhD Thesis, Department of Computer and Systems Sciences, University of Stockholm and the Royal Institute of Technology, Stockholm, 2006. ISBN: 91-7155-339-8

[12] UDSM, University of Dar es Salaam website: http://www.udsm.ac.tz, [Accessed on 25th March, 2008]

[13] UDSM-DPS, Directorate of Postgraduate studies -UDSM "Postgraduate Programmes and Admission procedures" (2008/09)

[14] Webometrics Ranking, http://www.studysa.co.za/contentpage.aspx?pageid=4150 [Accessed on 25th March, 2008]

[15] PHEA, http://www.foundation-partnership.org/pubs/pdf/more_bandwidth.pdf

[16] UDSM-Lib, University of Dar es salaam, Library: http://library.udsm.ac.tz

[17] Luhanga, M., and Mashalla, J., "Reforms and innovations in higher education: A reflection on the initiatives and lessons at the University of Dar es Salaam in Tanzania". A paper prepared for a Nuffic Conference on Higher Education, (2005)

IMMUNE SYSTEM BASED INTRUSION DETECTION SYSTEM

Christoph Ehret, Ulrich Ultes-Nitsche

University of Fribourg

Department of Computer Science, University of Fribourg,

Boulevard de Pérolles 90,

CH-1700 Fribourg, Switzerland

{christoph.ehret,uun}@unifr.ch

ABSTRACT

The threats and intrusions in IT systems can basically be compared to human diseases with the difference that the human body has an effective way to deal with them, what still need to be designed for IT systems. The human immune system (HIS) can detect and defend against yet unseen intruders, is distributed, adaptive and multilayered to name only a few of its features. Our immune system incorporates a powerful and diverse set of characteristics which are very interesting to use in the design of Intrusion Detection Systems (IDS). The authors propose therefore a hybrid intrusion detection system which combines host based and network based components but giving the focus to the host based intrusion detection as it is similar to the HIS. The proposed intrusion detection system will use the concepts of the artificial immune systems (AIS) which is a promising biologically inspired computing model based on the HIS. This paper presents an intrusion detection system based on the model of the human immune system and which will use the artificial immune systems paradigm. Furthermore the paper will also introduce some yet unused AIS concepts that can be applied to improve the effectiveness of IDS.

KEY WORDS

Intrusion detection systems, immune system, artificial immune system.

IMMUNE SYSTEM BASED INTRUSION DETECTION SYSTEM

1 INTRODUCTION

Intrusion detection systems (IDS) are nowadays very important for every IT company which is concerned with security and sensitive systems. Even if a lot of research was already done on this topic, the perfect IDS has still not been found and it stays a hot and challenging area in computer security research. Recently a new approach started to make its way to intrusion detection, namely the immune system. It has a lot of interesting features we would like to find in IDS. A new artificial intelligence paradigm was created from the immune system, namely the artificial immune system; this paradigm is rather new compared to neural networks or fuzzy logic, but it is very promising for different areas in computer science. If we abstractly compare the way an intrusion detection system and the human immune system work, we can actually find quite some similarities. Within this context it is normal to use as much similarities as possible to improve IDS and to see how we can implement the different features; this is where the artificial immune systems paradigm will help.

In this paper, we describe work in progress on artificial-immune-system inspired IDS. Its main purpose is to motivate the new paradigm and highlight the benefit one expect from that paradigm. The paper is structured as follows. First, we present the common design of the intrusion detection systems. Next, we give a brief overview of the immune system followed by a brief introduction to artificial immune systems. Then we discuss similarities between IDS and the immune system and their impact on advanced IDS. Finally, we conclude with presenting future work.

2 INTRUSION DETECTION SYSTEMS

An intrusion detection system can be compared with a house burglar alarm: if somebody tries to enter illegally in the house, one of the sensors will detect it what will trigger the alarm bell and alert the house owner and the police. Similarly, if somebody tries to compromise the confidentiality, the integrity or the availability of a computer system or network, or tries to

break the security protections, an intrusion detection system will alert the system owner and the security team [1].

Intrusion detection is the process of monitoring and analysing events of a computer system or network and tries to find intrusions. Events like trying to break into a system from the Internet using software exploits or trying to gain higher privileges on a system are representative events that will be recognized as an intrusion. Highly sensitive systems that have to be protected against 0-days attacks or critical systems with high availability needs, which cannot be patched very often, are typical systems that need an IDS. It is important to understand that the goal of an IDS is not to prevent an attack, but to detect it as quickly as possible and alert the right people who can then take the appropriate measures if a system was compromised; automatic measures can sometimes also be used by the IDS.

2.1 Placement

The placement or audit source location is one of the IDS taxonomies [2] the authors will focus on. There are two different strategies where to place intrusion detection systems: on a host or on a network node. Both placement strategies have their advantages and disadvantages.

A host-based IDS (HIDS) is often an application installed on the host for monitoring purposes, like Snort [3], Samhain [4] or Prelude [5]. It analyses events from running applications, the operating system, network packets or logs and if an intrusion is detected, an alarm event is sent to a central monitoring instance.

A network-based IDS (NIDS), often a commercial product installed on some special hardware, is positioned on a network node. It captures and analyses network packets that go through the node it monitors. One single NIDS or sensor, intelligently placed, can monitor several hosts independently of their operating system [7]. The captured network packets are analysed locally and if an attack is detected, an alarm event is sent to a central monitoring instance.

Table 1 lists in parallel the advantages and disadvantages of both host based and network based IDS, regarding several typical features.

Table 1. Advantages and disadvantages of HIDS and NIDS

Features	**HIDS**	**NIDS**
Management	Harder to manage due to the heterogeneity of the environment and its high number in large networks with many hosts	Simple to manage due to its homogeneity and a few NIDS are sufficient to monitor a large network with many hosts
Analyse encrypted network traffic	YES	NO
IDS evasion techniques	Harder to perform than on NIDS[6]	Evasion techniques like fragmentation will easily work with NIDS when they have no possibility to reconstruct locally the fragmented network packets
Knows if an attack was successful or not on a host	YES	NO
Protection against targeted attacks	Can be disabled during the attack of a host or by specific denial-of-service attacks	Easier than HIDS to protect against targeted attacks and can run in stealth mode
Detects large network attacks	NO	YES
Uses computing resources of the monitored host	YES	NO

It is not easy to decide between HIDS and NIDS which one is better or suites best our needs, but the trend is to integrate both or to design hybrid IDS that have both components [5][8]. Table 1 will help us to understand the proposed IDS design presented later.

2.2 Detection mechanisms

The detection mechanisms or algorithms represent another IDS taxonomy the authors will focus on. There are two different detection mechanisms IDS can use to find intrusions or attack attempts: the misuse detection and the anomaly detection.

The misuse detection approach, the most used in commercial products, monitors and analyses system events looking for a known event or sequence of events that represents an attack; this event or sequence of events is stored in the form of a signature. One disadvantage of the misuse detection is that if the signatures database is not up to date or if a new attack is used for which there exists no signature yet, the IDS will find nothing suspicious. On the other hand theses signatures permit to define an attack precisely and to give it a name, what really helps system administrators without great security background to understand what happened and if needed inform the security team. Another problem that can exist is if the signatures are too specifically bound to a given attack, the IDS will not be able to detect variants of the attack. It is nevertheless due to this specificity that the false positive or false alarms rate is very low.

The anomaly detection approach detects unusual behaviours, i.e. anomalies, like a great CPU consumption that lasts longer than usual, a high network traffic from the secretary's computer at 4am or the number of files accessed by a user in a given period of time. In order to detect anomalies, the detection system needs to create a normal behaviour profile and train the system on it. The anomaly detection can then use statistical measures or rules and compare the results with the profile; if there are differences, an anomaly was detected. This detection approach is rarely used in commercial products but is of great interest in the research area of intrusion detection. One advantage of this detection mechanism is that it is possible to detect yet unknown attacks and generate immediately a signature from it for the misuse detection. The false positive and false negative rate of the anomaly detection is unfortunately much higher than with the misuse detection.

3 IMMUNE SYSTEM

3.1 Overview

The human immune system (HIS) is quite complex and elaborate. The defence of the HIS is organised in different layers, mainly the exterior defences, which are biochemical and physical barriers like for example skin or bronchi, the physiological barrier, where pH and temperature provide inappropriate living conditions for pathogens, the innate system and finally the adaptive system. Every layer has different defence mechanisms and stops different types of pathogens. The innate and adaptive systems are again divided into several different cells, as we can see it on Figure 1.

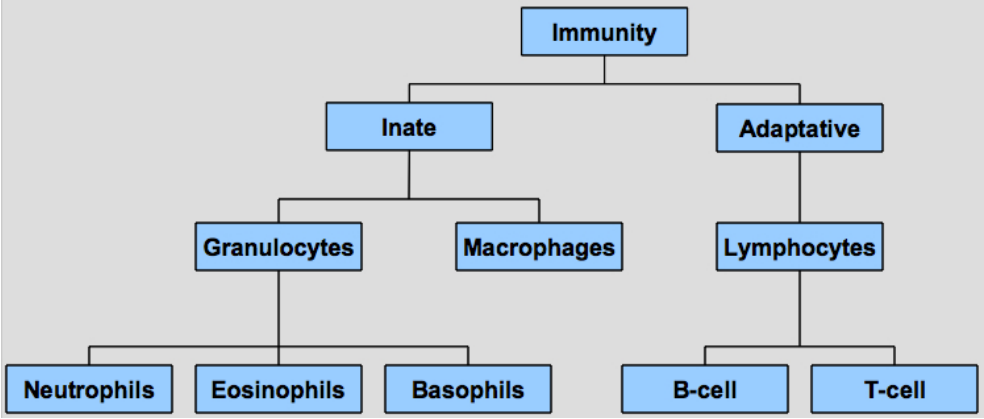


Figure 1. Major immune cells and their classification

Every leukocyte has very specific functions, like for example the Neutrophil[1] which migrates to sites of inflammation or infection and ingests micro organisms or particles, destroys them and dies, or the Eosinophil[2] which is responsible to combat parasites and is the main effector in allergic responses and in asthma. The B- and T-cells are the actors of the adaptive system; they are responsible to detect yet unknown pathogens, produce the

[1] The Neutrophils constitute the majority of blood leukocytes and are part of the phagocyte cells

[2] The Eosinophil constitutes 1-5% of blood leukocytes

specific antibodies and destroy them. Every B- and T-cells have different detectors, called epitopes, which interact with different kind of pathogens.

In order to improve the diversification, new B- and T-cells die and are created with randomly generated receptors every day, what modifies continuously the set of possible detected pathogens. There is a great interaction between all the different cells of the HIS; some immune cells secrete special substances that will attract some other type of immune cells, or some are responsible to produce an inflammation what will allow more immune cells to reach this particular region.

For more information on the different leukocytes and their role within the HIS consult [9].

3.2 Artificial Immune Systems

We can find quite different definitions of an artificial immune system (AIS) in the literature; one possible definition could be "*Artificial immune systems (AIS) are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving*" [10]. The artificial immune system paradigm is rather recent comparing to other artificial intelligence paradigms like Neural Networks, Fuzzy Logic or the genetic algorithms. AIS began in 1986 with Farmer, Packard and Perelson's paper on immune networks [11], but there was only in the mid-90's that it kept the attention of scientists.

What do we need if we want to implement an AIS framework? If we abstract the immune system in a simplistic way we have a population of different types of immune cells and interactions between them through receptors. For our AIS we therefore need to have a population, defined as a set, a way to describe each element of the set, its length, and a way to measure an interaction. To describe the population we will use the concept of *shape space* (S); it is used in immunology to quantitatively describe the interactions between immune cells and antigens. An element of S is described by a set of N_p parameters (length, width, charge, ...). To cover the whole shape-space, we actually need to generate $N = k^L$ different elements, where k is the size of the alphabet, L the length of one element of the set, and N is called the *potential repertoire*. As we have seen, one antibody can detect pathogens with similar structure, i.e. it is not bound to only one specific pathogen (imagine the number of antibodies we would need if each

could detect only one given pathogen). For that we will introduce the notion of *coverage*; $C = \sum_{i=0}^{L-\varepsilon} \frac{L!}{i!*(L-i)!}$ gives us the number of antigens covered by one antibody, where L is the string length of the antibody and ε the *cross-reactivity threshold.* The cross-reactivity threshold characterizes the fact that each antibody interacts with all antigens whose complement lies within a small surrounding region. The minimum elements necessary to cover the shape-space S is therefore given by $N_m = ceil\left(\frac{N}{C}\right)$, where N is the potential repertoire and C the coverage. The interaction, i.e. the affinity between an antibody and an antigen, both of length L, is evaluated with a distance measure between their attribute strings S^L x $S^L \rightarrow \mathbb{R}^+$. To measure the distance, the Euclidean, Manhattan or Hamming distance functions are often used. Finally, the training phase is often done like in the immune system using the negative selection [12] improved sometimes with some genetic algorithms. In the immune system, T-cells are trained in the thymus and selected or matured using the negative selection process depending if they reacted or not to *self* cells; if T-cells recognized the own cells (self-cells) as intruders they will not be selected and will not survive the training phase.

The application domain of AIS is becoming quite large. It is used for example in computer security, data analysis, search and optimization methods, agent-based systems, or autonomous navigation and control systems.

4 IMMUNE SYSTEM ANALOGY TO IDS

The human immune system has abstractly quite some similarities with intrusion detection systems, what the authors think make it naturally a good candidate as model for IDS design. The innate system of the human immune system can be compared with the misuse detection of the IDS; both uses pattern recognition based respectively on memory cells or signatures database to detect intrusions. The adaptive system can be compared with the anomaly detection where both can detect yet unseen attacks and where their sensors have to go through a training phase. Following the immune system model, the authors propose an IDS that uses both misuse and anomaly detection, quite the contrary of traditional IDS design that uses either misuse or anomaly detection. The misuse detection part will contain only the signatures for the running services and the anomaly detection sensors will

be able to generate automatically new signatures of detected and yet unknown attacks.

Each immune system protects a particular body and is also located in that same body. If we compare this to IDS placement strategy we clearly have a host-based IDS. Therefore the authors propose a HIDS with the possibility to send newly generated signatures to other hosts on a same LAN. Thanks to this feature, we include two important characteristics of the immune system that are distributivity and diversity. Moreover this permits us to abstract a LAN as a body and each host of this LAN becomes an immune cell.

One of the seven IDS requirements reported in Kim [13] is *efficiency*. An IDS has of course to be simple and not use too many resources on the monitored system; to this statement we would append "when nothing anomalous happens on this system". What happens to a human being when he has a cold with fever and a nasty headache? He stays in bed and tries to recover as quick and good as possible; he perhaps boils some water for his tea or eats a little bit but that is all he will do until he has recover strength. The authors propose to build an IDS that follows this principle: when something anomalous happens on a system it will slow down its normal functioning and give more resources to the IDS in order to find the problem, possibly fix it and avoid on the same way that the hypothetic attack can spread too quickly. This will also help the response team to take appropriate measures.

5 CONSEQUENCES FOR ADVANCED IDS

Lundin and Jonsson identified nine research issues in the intrusion detection area [14]: foundations, data collection, detection methods, reporting and response, IDS environment and architecture, IDS security, testing and evaluation, operational aspects and social aspects. The authors of the paper focus their research using the AIS paradigm on the issues *detection methods* and *IDS security*. The detection method issue is simple to implement with AIS using the negative selection we have seen previously or using ideas from the danger model, another model in immunology we have not described here and that is out of scope of this paper. This other model is quite promising and has yet not been used often in AIS [15]. As we have seen in the previous section, we will introduce a new intrusion detection

approach using both misuse and anomaly detection with automatic signature generation. This approach was partially implemented in ADENOIDS [16], but it was limited to generate signatures for buffer overflows and was done at a high level of abstraction. Furthermore using both detection approaches together with the co-stimulation mechansim of the immune system will help to reduce the false positives. We will have to go much deeper at a level quite similar to the way the immune system works and auto-generation of signature files for SNORT is foreseen.

The *IDS security* issue can be implemented in artificial immune systems using multiple independent sensors in different places of the system like for example with agents or the co-stimulation mechanisms we have seen in section 3. With this issue we have the multilayered or defence in depth feature of the immune system, the diversity of different kinds of detectors and we can minimize or avoid single points of failure.

6 CONCLUSION

The immune system is complex but very powerful; it can detect a lot of different types of pathogens, even unknown one, and thanks to a strong interaction between all the different actors of the immune system the pathogens can be destroyed. As the immune system has some very interesting features, a new artificial intelligence paradigm called the artificial immune system was created from it. Computer security, especially the antivirus and IDS fields, is of course an interesting candidate to apply AIS. The immune system itself is actually a very interesting approach to intrusion detection.

We discussed in this paper an immune-system-inspired approach to intrusion detection. The similarities between the tasks of the human immune system and intrusion detection systems suggest that IDS can be improved by converting concepts from the biological to the digital world. Clearly, we must abstract from the concrete biological principals to benefit from them in intrusion detection. It was the purpose of this paper to discuss these necessary abstractions. Interaction between misuse and anomaly detection, distributivity, avoiding single points of failure, and locality, possibly affecting only single processes, are what we have extracted as main features of immune-system-inspired IDS.

Our current research in direction focuses on identifying good anomaly detection methods for IDS. This includes particularly reducing the number of false positives in the potentially applicable methods, as they are usually the limiting factor in misuse detection, and not the false negatives that are much easier to control.

7 REFERENCES

[1] R. Bace and P. Mell. “Intrusion Detection Systems”, NIST Special Publication 800-31. 2001.

[2] H. Debar, M. Dacier and A. Wespi. “Towards a taxonomy of intrusion-detection systems”. *Computer Networks*, April 1999.

[3] Snort
URL: http://www.snort.org

[4] Samhain
URL: http://www.la-samhna.de/samhain

[5] Prelude
URL: http://prelude-ids.org

[6] T. H. Ptacek and T. N. Newsham. “Insertion, Evasion, and Denial of Service: Eluding Network Intrusion Detection”, Report from Secure Networks Inc., 1998.

[7] J. McHugh. “Intrusion and intrusion detection”, *International Journal of Information Security*, vol. 1, pp. 14-35, 2001.

[8] S. Northcutt and J. Novak. “Network Intrusion Detection”, Sams, Third Edition, 2002.

[9] I. Roitt, J. Brostoff and D. Male. “Immunology”, Mosby, Sixth Edition, 2001.

[10] L. N. de Castro and J. Timmis. “Artificial Immune Systems: A New Computational Intelligence Approach”, Springer, 2002.

[11] J.D. Farmer, N. Packard and A. Perelson. "The immune system, adaptation and machine learning", in *Physica D*, vol. 2, pp. 187-204, 1986.

[12] S. Forrest, A. Perelson, L. Allen and R. Cherukuri. "Self-Nonself Discrimination in a Computer", in *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, 1994.

[13] J.W. Kim. "Integrating Artificial Immune Algorithms for Intrusion Detection", PhD thesis, University College London, 2002.

[14] E. Lundin and E. Jonsson. "Survey of Intrusion Detection Systems", Technical Report 02-04, Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden, 2002.

[15] U. Aickelin, P. Bentley, S. Cayzer, J. Kim and J. McLeod. "Danger Theory: The Link between AIS and IDS?", in *Proceedings of the Second International Conference, ICARIS 2003* (LNCS 2787), pp. 147-155, Springer, 2003.

[16] F. S. de Paula, L. N. de Castro and P. L. de Geus. "An Intrusion Detection System Using Ideas from the Immune System", in *Evolutionary Computation*, vol. 1, pp. 1059-1066, 2004.

THE INFORMATION SECURITY OF A BLUETOOTH-ENABLED HANDHELD DEVICE

Frankie Tvrz[1] and Marijke Coetzee[2]

[1]Department of Business Information Technology
[2]Academy for Information Technology
University of Johannesburg

[1]frankie.tvrz@sita.co.za
[2]marijkec@uj.ac.za

ABSTRACT

Bluetooth connectivity allows workers to access information anywhere, including both personal and corporate information. Software and applications have been specifically developed for handheld devices such as PDAs, giving users a high level of usability and functionality. The goal of this paper is to present an information security evaluation of a Bluetooth enabled handheld device, such as a PDA. The use of Bluetooth wireless technology and functionality provides added benefits, but also brings new information security threats to organisation's information assets. The research attempts to understand the implications of using a Bluetooth enabled handheld device in both public and private environments. Five high-level layers are defined for this discussion. Security risks are evaluated based on current research into vulnerabilities, attacks and tools that exist to compromise a Bluetooth enabled handheld device. Possible recommendations to mitigate identified security risks are also suggested.

KEY WORDS:
Bluetooth, information security, layered approach, vulnerabilities, attacks

THE INFORMATION SECURITY OF A BLUETOOTH-ENABLED HANDHELD DEVICE

1 INTRODUCTION

Wireless technologies are deployed in both public and private networks, and may even be preferred over traditional wired networks [STAN02]. Bluetooth [GEHR04] gives mobile workers the ability to create ad-hoc connections with mobile devices, corporate networks, and Internet hotspots.

This offers mobility and convenience of use for Bluetooth enabled handheld devices such as PDAs (Personal Digital Assistant) and smart phones [GALL04]. Market research has indicated that Bluetooth-enabled devices will experience a 60% compound annual growth rate between 2003 and 2008. Bluetooth usage continues to increase especially in Bluetooth enabled handheld devices [BRIT07]. It was predicted that Bluetooth enabled devices would increase from 316 million units in 2005 to 866 million in 2009 [SDAA07]. The proliferation of Bluetooth enabled PDAs, smart phones and laptops bring Bluetooth past the enterprise door into the corporate network environment, usually without the knowledge of the corporation [HICK06].

Information security risks are introduced as people are utilising an easy to use technology such as Bluetooth, not really designed for information security, on handheld devices that are becoming more sophisticated in informal and corporate environments.

The aim of this paper is to provide an information security evaluation of a Bluetooth enabled handheld device, such as a PDA. Section 2 gives a basic background to the environment in which Bluetooth is used. Section 3 describes a layered approach to evaluating the Bluetooth handheld device. Section 4 describes the risks of using a Bluetooth enabled handheld device, and section 5 evaluates how information security services are implemented. Section 6 concludes the paper.

2 BACKGROUND

An IT consultant John uses his PDA for both corporate and private use, as shown in figure 1. He stores confidential client information on his PDA such as technical documents, minutes of meetings, calendar items and e-mails. His Bluetooth enable handheld device is a PDA that is a highly functional pocket sized computing device consisting of a small liquid crystal display, an operating system, a processor and memory. The PDAs utilises the Windows Mobile [PRIC03] operating system. John may unknowingly bring malicious software from the wireless public environment into the wired corporate network environment. To be able to comprehensively discuss the information security risks presented by the Bluetooth enabled handheld device of John, a layered approach is defined next.

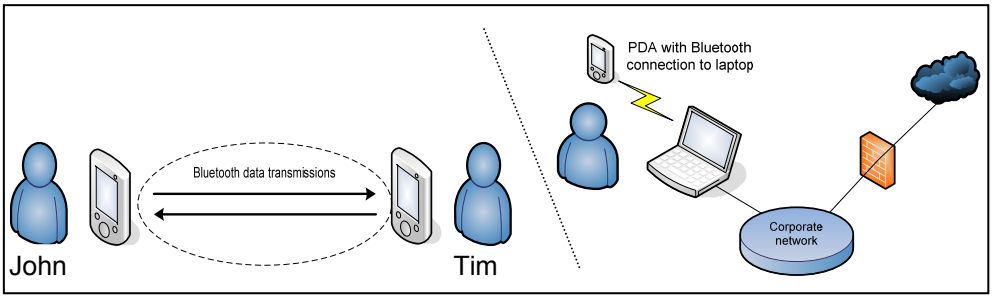


Figure 1 – Bluetooth use in public and corporate networks

3 THE BLUETOOTH HANDHELD DEVICE

Layering is a method of combining different information security components to provide layers of protection. This assists in creating a defensive barrier. Layered information security improves the information security mechanism and increases the difficulty of compromising the handheld device. In contrast, a vulnerability or poor information security configuration in a layer could allow an attacker with a possible unauthorised access point.

In order to organise and structure the discussion on the information security of the Bluetooth enabled handheld device of John, the following layers are defined, as shown in figure 2:

- *Physical (Bluetooth)* – the implementation of Bluetooth in hardware;

- *Bluetooth(software)* – software implemented in, and defined over the physical device to allow wireless interconnectivity between devices;
- *Operating System* - the main control program of the handheld device that enables hardware and loaded applications, including Bluetooth configurations and services, to function correctly;
- *Applications* – programs dependent on the operating system such as e-mail, word processing, and calendar items;
- *User* – considered the administrator of the handheld device as it is under his/her full control. He/she configures features on the handheld device such as its information security, operating system and applications.

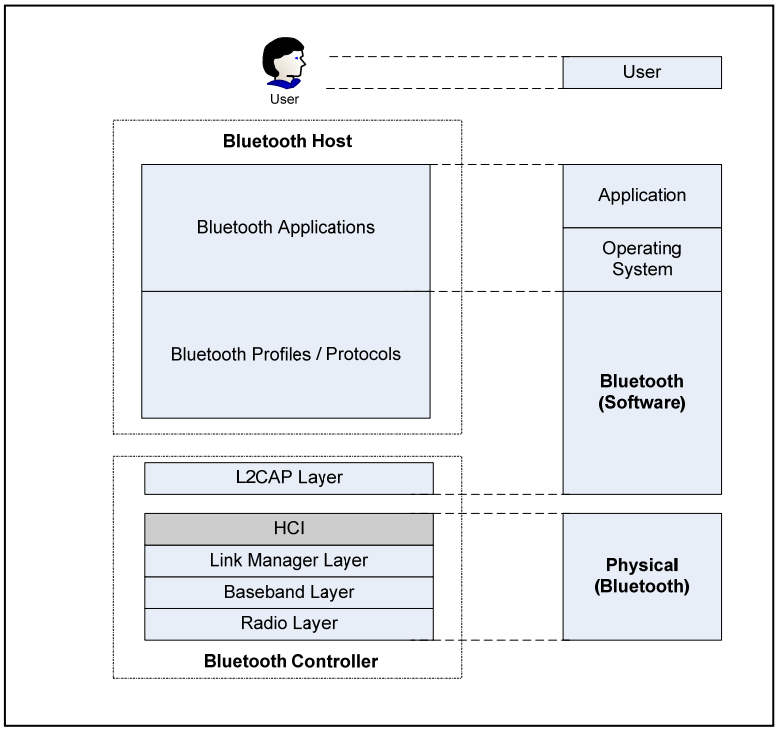


Figure 2 – Layers of the Bluetooth enabled handheld device

It is also important to further define the six layers of the Bluetooth architecture, shown in figure 2 [MCDE05] [ANAN01] [INSI06]:

- *Radio Layer* – the Bluetooth transceiver that defines the transmission and receiving of packets over the physical channel;

- *Baseband Layer* – enables devices search for and connect to other devices by enabling the physical radio link between the devices compromising the piconet;
- *Link Manager Layer* – manages the properties of the air-interface link between devices such as bandwidth allocation for data, bandwidth reservation for audio traffic, authentication by means of challenge response, trust relationships between devices, encryption of data and control of power usage;
- *Host Controller Interface (HCI)* – provides a standard interface for upper level applications to access the lower;
- *Logical Link Control and Adaptation Protocol (L2CAP)* – allows multiple protocols and application to share the air-interface;
- *Profiles* – profiles describe how the technology is used in different scenarios;

Possible information security risks, presented by each of theses layers are discussed next.

4 INFORMATION SECURITY RISK OF THE BLUETOOTH ENABLED HANDHELD DEVICE

An information security risk is the likelihood that an accidental or intentional threat will compromise vulnerabilities within the Bluetooth enabled handheld device. [SANS07], and bring new information security threats to organisation's information assets. Many vulnerabilities and attacks exist on a Bluetooth enabled handheld device that can be used to compromise its information security. In order to gain perspective on the risks that a Bluetooth enabled handheld device presents, Figure 3 gives high-level view of theses aspects. This may assist users to understand and be aware of the inherent risks in using a handheld device within a Bluetooth communication environment.

- *The physical layer* is inherently vulnerable to physical based attacks and manipulation of the Bluetooth enabled handheld device. The frequency hopping technique employed by *Bluetooth hardware* to prevent eavesdropping can be circumvented. Interference from other applications and implementation of the Bluetooth stack on different operating systems pose risks. At the *Baseband* layer inquiry scans can be used to discover

Bluetooth devices which could allow anonymous information gathering to be performed providing the Bluetooth device address, manufacturer and Link Manager protocol version information. Devices can be set to being discoverable or non-discoverable which can have an influence on Bluetooth based attacks. At the *Link Manager layer,* weaknesses exist within Bluetooth authentication, as input parameters are the Bluetooth device address and the user PIN, sent in clear text. This makes Bluetooth authentication vulnerable to eavesdropping and brute force attacks. Encryption provided by Bluetooth only encrypts the packet payload but the packet header and access code are not encrypted, allowing eavesdropping to be performed.

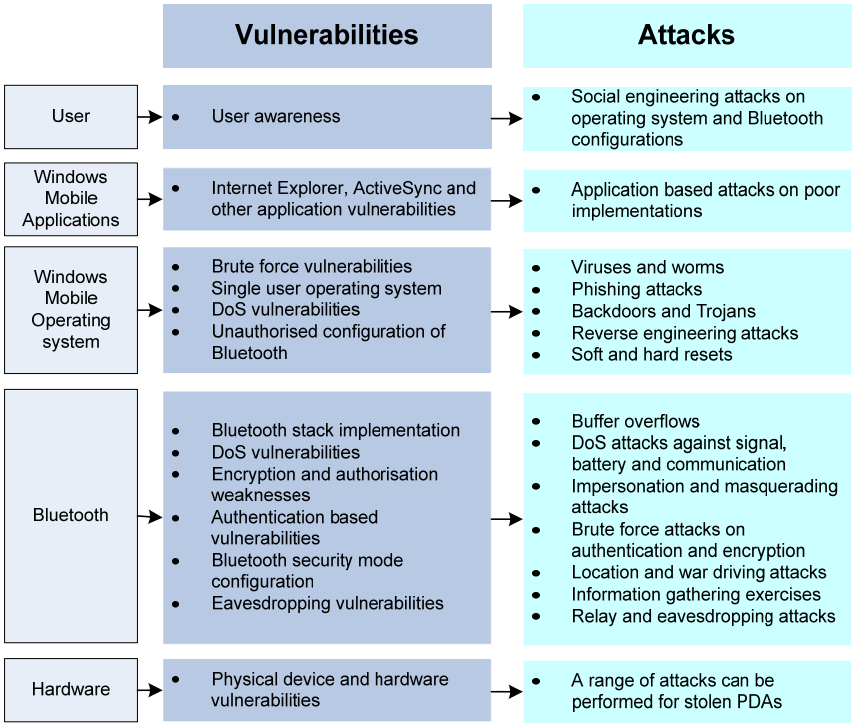


Figure 3: Overview of attacks and vulnerabilities of the Bluetooth enabled handheld device

- *The Bluetooth layer:* The *L2CAP* layer is responsible for maintaining connections between communicating devices and is vulnerable to denial

of service (DoS) attacks. Multiple vulnerabilities exist which allowing a number of attacks, with the complexity of the attack ranging from difficult to effortless execution. Attacks ranging from simple Denial of Service (DoS) attacks against the Bluetooth signal to detailed brute force attacks against the authentication mechanisms can be successfully performed due to inherent vulnerabilities within Bluetooth and its implementation. *Bluetooth profiles* are dependent on the manufacturers of the Bluetooth enabled handheld device and the information security of the protocols used outside of the Bluetooth Specification for the secure operation of the Bluetooth profile.

- *The operating system layer* can be targeted as another method of gaining access to or through the Bluetooth configuration. Vulnerabilities exist because of the single user operating system allowing brute force attacks. The increased threat of malicious software such as viruses and worms can be exploited by attackers. User identification and authentication is not performed, although a power on password is required when using the PDA. Bluetooth configurations are stored in the registry of the operating system and are dependent on operating system controls to protect the Bluetooth information security settings. If the operating system were to be compromised then Bluetooth configurations could be changed through the Windows Mobile operating system.
- *The application layer* has implementation vulnerabilities due to poor application programming, allowing application based attacks which can be used to gain unauthorised remote access to the operating system, and then to Bluetooth.
- *The user layer* is dependent on the user's actions, which would affect the information security of all the layers of the Bluetooth enabled handheld device. The user could be unaware of information security practices to ensure a secure operating environment for the Bluetooth enabled handheld device.

Next, Bluetooth information security services are evaluated that can be used to mitigate theses risks.

5 INFORMATION SECURITY SERVICES OF A BLUETOOTH ENABLED HANDHELD DEVICE

Information security services are the measures that are employed to prevent the unauthorised use, misuse, modification or denial of the use of assets [BRAU00]. Bluetooth information security is designed so that the end user is able to configure and manage the information security options for communication [GEHR04]. Table 1 shows how each layer implements identification, authentication, authorisation and confidentiality. It also indicates how each layer supports the next and whether information security mechanisms provided by one layer possibly supports the controls in the next layer of the Bluetooth enabled handheld device.

Each of the information security services is now evaluated. The integrated implementation of the four information security services is discussed, as it applies across all the layers of the Bluetooth enabled handheld device, and main concerns are highlighted.

Identification *is the act of identifying an entity such as a user, application or device to the Bluetooth enabled handheld device and its applications, so that it can recognise the entity and distinguish it from others.* The main concern is that when a handheld device is accessed via Bluetooth, the user is not identified, but rather the device making the connection. The Bluetooth device address is not used by the operating system or by any application. At the user layer, a concern is that the owner of the device may not be aware that the other user is not identified. Identification is thus not implemented in an integrated manner across the layers of the handheld device and does not comply with the definition given above. Only device based identification is performed. It would be very important to ensure that a user is fully aware of this, and must understand how Bluetooth identification takes place when another entity makes a Bluetooth connection to access services that are provided.

Authentication *verifies the identity of the user, process or Bluetooth enabled handheld device, as a prerequisite to allowing access to resources offered on the device.* The physical, Bluetooth, operating system and application layer of the Bluetooth enabled handheld device are responsible for their own authentication mechanisms. Bluetooth authentication is inherently weak and can be compromised. Furthermore, the operating system does not use the authentication performed by Bluetooth, neither do

Table 1: Information security services

Information Security Service	Physical	Bluetooth	Operating System	Applications	User
Identification	MAC address Bluetooth device address (BD_ADDR)	Bluetooth device address (BD_ADDR)	None, unless third party software is used or integration to the corporate network	Username unless third party software is used, or integration to corporate network	Awareness that device, not user is identified
Authentication	None	Windows Mobile information security policies Initialisation key and Link key based on random number, pin and Bluetooth device address	Windows Mobile information security policies Power-on password	Password integration to server based applications Application execution authentication	Awareness of power-on password or strength of pin number
Authorisation	Windows Mobile information security policies Remote and local storage card wipe	Windows Mobile information security policies Bluetooth security manager consisting of trust relationships, device and service database	Windows Mobile information security policies Application permissions Certificates	Windows Mobile information security policies Application permissions Certificates	Awareness of Windows Mobile information security policies and Bluetooth security mode configuration
Confidentiality	Windows Mobile information security policies Encryption for storage cards	Windows Mobile information security policies Bluetooth encryption algorithms Encryption key	Windows Mobile information security policies Stream based encryption Block cipher encryption One-way hashing algorithm Digital signatures	Windows Mobile information security policies Application based encryption SSL encryption	Awareness of Windows Mobile information security policies, Bluetooth security mode configuration and use of encryption programs

the applications. At the user layer, the user might not be aware that he is not authenticated at the Bluetooth or operating system layer. Bluetooth authenticates the handheld device and the operating system authenticates via a power-on password. He also needs to understand that authentication is based on the link key and not the pin entered by the user when performing Bluetooth authentication. Only device based authentication is performed for the Bluetooth and operating system layers which does not fully comply with the definition given above. The user needs to understand how Bluetooth authentication is performed when gaining access to Bluetooth services, especially since the information security risk is increased when Bluetooth authentication is performed in public environments.

Authorisation *is the process of determining whether the user or Bluetooth enabled handheld device can be granted access to services offered by the host device.* Authorisation mechanisms provided by information security policies can be used on the physical, Bluetooth, operating system and application layers. However, these authorisation mechanisms are not integrated across the layers of the Bluetooth enabled handheld device. The main concern is that detailed authorisation controls are not provided for services offered by Bluetooth, access is based on the first two layers of the Bluetooth enabled handheld device. If devices have paired, they may trust each other, and may be granted access to any service exposed by Bluetooth. Authorisation is not implemented in an integrated manner across the layers of the Bluetooth enabled handheld device, but only partially complies with the definition above. It is important for the user to be aware that Bluetooth device based authorisation is used and when other entities make a Bluetooth connection to access services provided by the Bluetooth enabled handheld device.

Confidentiality *is the property that guarantees that Bluetooth communicated information or information stored on the handheld device is not made available to unauthorised individuals, entities or processes.* Confidentiality can be enforced across all the layers through the use of information security policies, if handheld devices are managed centrally by the organisation. From when a Bluetooth connection is made, the information security policy could enforce that all services offered through Bluetooth use encryption. However the confidentiality mechanisms provided by each layer for the Bluetooth enabled handheld device are independently enforced. The main concern is that inherent information

security risks exist within the Bluetooth specification which are present when using Bluetooth on the handheld device. Confidentiality is not adequately enforced by the Bluetooth specification allowing a number of attacks to be performed. The operating system does not use the encryption provided by Bluetooth when making a Bluetooth connection, neither does the application. At the user layer, he may not be aware that each layer of the Bluetooth enabled handheld device uses their own confidentiality mechanisms which each require to be configured to ensure that data transmitted and stored is encrypted. Confidentiality is not implemented in an integrated manner across the layers of the Bluetooth enabled handheld device and partially complies with the definition given above. It is important that the user understands how Bluetooth encryption is performed, when Bluetooth devices make connections to services offered.

From this evaluation, it is clear that Bluetooth negatively affects the information security of a handheld device such as PDA, as it provides wireless connectivity that is not integrated into the different layers of the Bluetooth enabled handheld device. Fundamental weaknesses exist within the identification and authentication information security weaknesses, impacting on the authorisation information security mechanism. Inherent weaknesses exist within the Bluetooth specification and implementation of Bluetooth on the handheld device, this has led to vulnerabilities and attacks that can be performed successfully to compromise the information security services of the Bluetooth enabled handheld device.

6 CONCLUSION

It is clear that the Bluetooth enabled handheld device presents a new risk within the information security realm. New threats have been created when introducing the handheld device to Bluetooth connectivity within the public and private environments.

Information security risks have been identified on all layers forming part of the Bluetooth enabled handheld device, which have led to the compromise of a number of information security services. The identified vulnerabilities and attacks could be used by to bypass information security mechanisms of all the layers of a Bluetooth enabled handheld device such as a PDA.

Information security services are not adequately addressed, and cannot sufficiently protect assets stored on the device. A compromised device may also be used to gain entry to a corporate network and the information that it stores. The information security risk of using a Bluetooth enabled handheld device should thus be clearly understood by the organisation before introducing it into the corporate network. The user also needs to understand the actions that can be performed to ensure that information security risks are mitigated against, to operate the handheld device in a secure Bluetooth environment.

7 REFERENCES

ANAN01 ANAND N. 2001. An Overview of Bluetooth Information security. SANS Institute 2003.

BRAU00 BAUKNECHT K. 2000. 5 Information Security Services ISO 7498/2. Website. http://www.ifi.unizh.ch/ikm/Vorlesungen/sec/02.pdf. 25 February 2004

BRIT07 Study Says Bluetooth Entering the Mainstream, http://www.brighthand.com/default.asp?newsID=10643, accessed 25 March 2008

GALL04 GALLEGOS F, Auditing Wireless Telecommunications: An Issue of Standards, Information Systems Control Journal, Volume 3 of 2004

GEHR04 GEHRMANN C, PERSSON J, SMEETS B. 2004. Bluetooth Information security, Norwood: Artech House computing library

HICK06 Mobile Computing News, http://searchmobilecomputing.techtarget.com/news/article/0,289142,sid40_gci1179892,00.html , accessed 25 March 2008

INSI06 INSIGHT CONSULTING. 2006. How can Bluetooth services and devices be effectively secured? Computer Fraud & Information security Journal. January 2006

MCDE05 McDERMOTT-WELLS P. 2005. What is Bluetooth? , Potentials, IEEE, Volume 23, Issue 5, of December.

PRIC03 PRICE R. 2003. PDA as a Threat Vector. SANS Institute 2003

SANS07 Glossary of Terms Used in Information security and Intrusion Detection. http://www.sans.org/resources/glossary.php?portal=c43474178943e08ef4a460dfb96fb20f#i . SANS institute. Accessed 25 July 2007

SDAA07 Mobile Phone Market Pushes Growth of Bluetooth Chip Market, http://searchmobilecomputing.techtarget.com/news/article/0,289142,sid40_gci1179892,00.html, accessed 25 March 2008

STAN02 STANLEY R.A. 2002. Wireless LAN Risks and Vulnerabilities. Information Systems Audit and Control Foundation (ISACA).

A NOVEL SECURITY METRICS TAXONOMY FOR R&D ORGANISATIONS

Reijo Savola

VTT Technical Research Centre of Finland

P.O. Box 1100, FI-90650 Oulu, Finland
+358 40 569 6380
Reijo.Savola@vtt.fi

ABSTRACT

In order to obtain evidence of the security and privacy issues of products, services or an organization, systematic approaches to measuring security are needed. In this study we survey the emerging security metrics approaches from the academic, governmental and industrial perspectives. We aim to bridge the gaps between business management, information security management and ICT product security practices. If appropriate *security metrics* can be to offer a quantitative and objective basis for security assurance, it would be easier to make business and engineering decisions concerning information security. We believe that being able to express a high-level taxonomy of security metrics will help the actual process of developing feasible composite metrics even for complex situations. A well-defined taxonomy can be used to enhance the composition of feasible security metrics all the way from business management to the lowest level of technical detail. Information security management, business management and, on the other hand, software security and network security engineering have been handled as separate areas. Common metrics approaches can be used to bridge the gaps in between.

KEY WORDS

Information security metrics, security assurance, information assurance, taxonomy

A NOVEL SECURITY METRICS TAXONOMY FOR R&D ORGANISATIONS

1 INTRODUCTION

The field of defining security metrics systematically is young and the current practice of information security is still a highly diverse field, and holistic and widely accepted approaches are still missing. In order to make advances in the field of measuring, assessing or assuring security, the current state of the art should be investigated and structured in a clear way.

The main contribution of this study is an initial proposal for a security metrics taxonomy for the ICT product Research and Development (R&D), supported with a literature survey of the current state of the art in industry strength and academic approaches to measuring security. Section 2 discusses the characteristics of security metrics and Section 3 proposes a taxonomy for security metrics, and finally, Section 4 gives conclusions.

2 CHARACTERISTICS OF SECURITY METRICS

It is helpful to notice the difference between metrics and measurements. Measurements provide single-point-in-time views of specific, discrete factors, while metrics are derived by comparing two or more measurements taken over time with a predetermined baseline [25]. Furthermore, according to Alger [1], measurements are generated by counting, whereas metrics are generated from analysis. According to Jelen [25], a good metric is Specific, Measurable, Attainable, Repeatable and Time-dependent ("SMART"). Payne [31] remarks that truly useful security metrics indicate the degree to which security goals, such as data confidentiality, are being met. Security metrics are used for decision support and very often these decisions are actually risk management decisions – aiming at mitigating, cancelling or neglecting security risks. Consequently, many metrics that might be useful for different purposes will be associated with risk analysis in a direct or indirect way. Security metrics and measurements can be used for decision support, especially in *assessment* and *prediction*. When using metrics for prediction, mathematical models and algorithms are applied to the collection

of measured data (e.g. regression analysis) to predict the security behaviour of an organization, a process or a product in the future. It is important to clearly know the entity that is the target of measurement because otherwise the actual metrics might not be meaningful. FIPS Publication 199 [11] presents a mechanism for investigating confidentiality, integrity and availability separately, emphasizing the *potential impact* assessment. In general, the security measurements can be based on the above-mentioned widely known objectives, augmented with some objectives such as non-repudiation, depending on the needs of situation.

Security and trust metrics can be obtained at different levels within an organization or a technical system. Detailed metrics can be aggregated and rolled up to progressively higher levels. As Yee [46] states, a multi-faceted or multi-dimensional security measure is needed. Security metrics properties can be quantitative or qualitative, objective or subjective, static or dynamic, absolute or relative, or direct or indirect. According to ISO 9126 standard [18], a direct measure is a measure of an attribute that does not depend upon a measure of any other attribute. On the other hand, an indirect measure is derived from measures of one or more other attributes.

2.1 On the Feasibility of Measuring Security

The feasibility of measuring security and developing security metrics to present actual security phenomena has been criticized in many contributions. In designing a security metric, one has to be conscious of the fact that the metric simplifies a complex socio-technical situation down to numbers or partial orders. McHugh [28] and McCallam [27] are skeptical of the side effects of such simplification and the lack of scientific proof. Bellovin [5] remarks that defining metrics is hard, if not infeasible, because an attacker's effort is often linear, even in cases where exponential security work is needed. Another source of challenges is that luck plays a major role [9] especially in the weakest links of information security solutions. Security metrics are difficult because the discipline of measuring security itself is still in the early stages of development. As yet, there is no common vocabulary and few documented best practices to follow. Those pursuing the development of a security metrics program should think of themselves as pioneers and be prepared to adjust strategies as experience dictates [31].

2.2 Related Work: Earlier Security Metrics Taxonomies

The WISSSR workshop [13] did not propose any specific security metric taxonomy. Instead, the workshop was intuitively organized into three tracks: technical, operational and organizational. Technical metrics are "used to describe, and hence compare, technical objects, e.g., algorithms, specifications, architectures and alternative designs, products, and as-implemented systems". Operational metrics are "used to describe, and hence manage the risks to, operational environments." Organizational metrics are "used to describe, and to track the effectiveness of, organizational programs and processes." In general, there would seem to be an intuitive understanding among the workshop participants that these three tracks would provide a useful basis for a taxonomy of security metrics [36].

Vaughn et al. [44] propose a taxonomy for information assurance metrics consisting of two distinct categories: (i) organizational security metrics and (ii) metrics for Technical Target of Assessment (TTOA). As Seddigh *et al*. [35] conclude, this taxonomy is a valuable contribution, but further work is required to make it applicable to an IT organization.

The U.S. National Institute of Information Standards and Technology (NIST) presents its security metrics taxonomy in NIST Special Publication 800-26 [38] and 800-55 [39]. The taxonomy is comprehensive, presenting three categories (management, technical, and operational) and 17 sub-categories. This taxonomy has been written from the point of view of an organization, and technical metrics category assesses the level of technical security controls in the organization rather than the technical security level of specific products, as does TTOA in Vaughn *et al*.'s taxonomy.

Seddigh *et al*. introduce an information assurance metrics taxonomy for IT Network assessment in [36]. Their taxonomy has three categories – security, Quality of Service (QoS) and availability – based on their novel definition of information assurance. Under each of these three they consider technical, organizational and operational metrics.

The Institute for Information Infrastructure Protection (I3P) [14] is also carrying out work on creating a taxonomy for security metrics from the process control systems perspective. Stoddard *et al.*, in [37], propose an initial security metrics taxonomy for process control systems based on the WISSSR workshop taxonomy and ISO/IEC 17799 [23] and ANSI/ISA-TR99.00.01-2004 [2] standards.

3 PROPOSED SECURITY METRICS TAXONOMY

The most direct factor contributing to the quality of the taxonomy is the quality of the *corpus* or source material [45]. As a source material, we present a survey of security metrics in this study.

3.1 Business Level Security Metrics

The highest category (root node) of our taxonomy is the security metrics for business management (Fig. 1). Business goals steer the security and trust management work and, accordingly, security and trust metrics should be defined in such a way that they are *aligned to the business goals* of a company or a collaborating value net of businesses. One way of establishing an overall metrics process is to begin with the business goals and demonstrate the alignment of lower level security management objectives within that context. Note that in any organisation, e.g. a government organization, "business goals" can be replaced by major goals that are specific to that organization (e.g. defined by legislation).

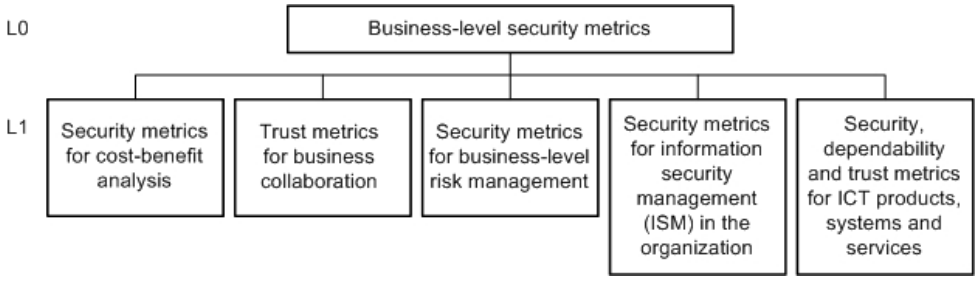


Figure 1. Business-level security metrics

Security ROI (Return On Investment) is quickly gaining popularity because it is a simple metric. Security ROI is defined by Blakley [7] as the amount of this annual benefit over its cost. Trust management is a relatively new research field that aims at understanding, modelling and controlling trust phenomena. Trust evaluation functions, such as Toivonen *et al.*'s work [40] have been defined to set a basis for trust quantification. Basili's [4] Goal/Question/Metric (GQM) approach can be used for establishing a metrics process (or program), beginning with the business goals. Note that regardless of the methodology used, developing business-relevant metrics needs commitment from the business management.

3.2 Metrics for Information Security Management in Organisation

Fig. 2 shows the taxonomy of the security metrics for information security management in the organization. In principal, we here follow the taxonomy definitions of [38] and [36].

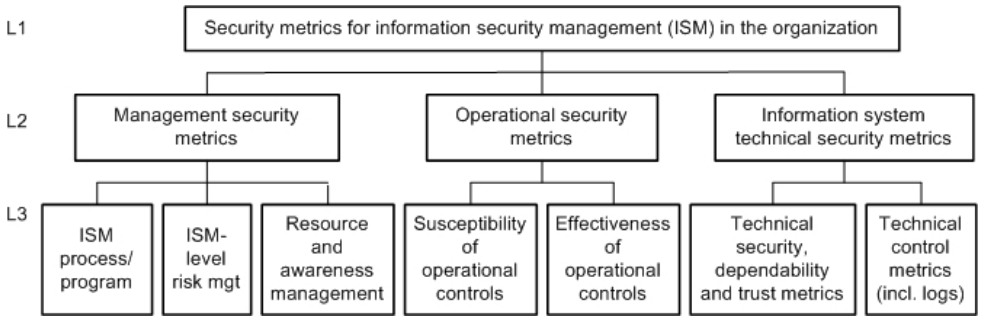


Figure 2. Metrics for ISM

These security metrics support evaluation of the security controls, plans and policies, as well as certification and accreditation activities. Human resource assessment is typically concentrated on training and security awareness polls, and evaluation of the human resource assignments [33]. Operational metrics address the susceptibility and effectiveness of operational security practices (or controls) [36]. They typically concentrate on incident response, the archiving process and the maintenance process of SW, HW and networking equipment. Furthermore, security documentation, data integrity and contingency planning are evaluated [36].

Technical SDT (Security, Dependability and Trust) metrics can be a subset or an instance of the SDT metrics for product life cycle management. NIST SP 800-26 [38] gives guidelines on security self-assessment of information technology systems based on the U.S. Federal IT Security Assessment Framework. NIST SP 800-53A [32] represents assessment methods and procedures for a minimum level due diligence for organizations assessing the security controls in their information systems. NIST SP 800-55 [39] provides guidance on how an organization, by using metrics, identifies the adequacy of in-place security controls, policies, and procedures. An example of an implementation metric is *percentage of NIST SP 800-53A control families for which policies exist*. Effectiveness and efficiency metrics are used to monitor the results of security control

implementation for a single control or across multiple controls. For example, *percentage of security incidents caused by improperly configured access controls* relies on information from or about several controls. NIST SP 80-100 [8], the information security guide for managers, contains a section on security measurements. The Federal Information Processing Standards (FIPS) Publication 199 [11] establishes security categories for both information and information systems. According to [11], the potential impact can be classified as low, moderate or high. According to Lennon of NIST [26], "the universe of possible metrics, based on existing policies and procedures, will be quite large. Metrics must be prioritized."

The Information Security Forum (ISF) [15] is a member-driven non-profit forum that has established the "Standard of Good Practice" (SOGP) [16] and the accompanying "Information Security Status Survey". The survey measures compliance with SOGP and ISO/IEC 17799. ISF offers a benchmark comparison to the members on the total or by business sector. ISF has also developed a simpler metric called "Security Health Check".

3.3 Security Metrics for ICT Products, Systems and Services

Probably the most challenging category of our taxonomy is the security, trust and dependability metrics for products, systems and services, see Fig. 3. For the basic concepts and taxonomy of dependable and secure computing, see the study by Avižienis *et al.* [3].

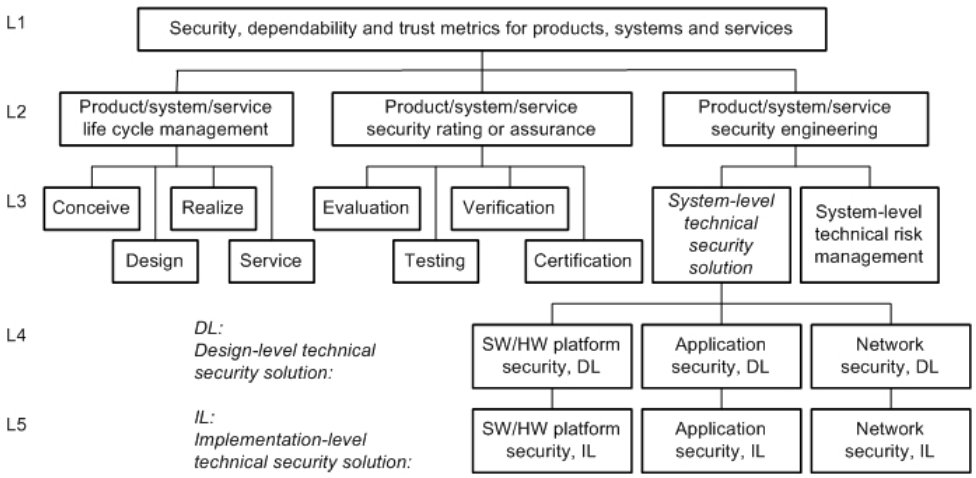


Figure 3. SDT metrics for products, systems and services

During the Conceive phase, the security requirements define the basis for measuring security later by comparing the requirements and actual design or system [34]. The Design phase incorporates activities such as architectural and lower-level design, testing, analysis and validation. As an example of product life-cycle security metrics, Systems Security Engineering Capability Maturity Model (SSE-CMM) ISO/IEC standard 21827 [24] contains security metrics for maturity assessment of the security level of security engineering processes and results of them. The resulting standards are the basis of evaluations by neutral third parties besides manufacturers and procurers. The most widely known of such efforts is the Common Criteria (CC) ISO/IEC 15408 international standard [22]. The CC standard is based on a combination of several other standards for information security, including TCSEC (Trusted Computer System Evaluation Criteria) [41], ITSEC (Information Technology Security Evaluation Criteria) [17], CTCPEC (Canadian Trusted Computer Product Evaluation Criteria) [10] and FC (Federal Criteria for Information Technology Security) [43]. Interpretations of the TCSEC have been published to apply them to other contexts such as the TNI (Trusted Network Interpretation of the TCSEC) [42]. The ISO/IEC technical report ISO/IEC 9126-1 [18] defines a quality model for software, and reports ISO/IEC 9126-2 [19], ISO/IEC 9126-3 [20] and ISO/IEC 9126-4 [21] provide a suggested set of software quality metrics for external, internal and "quality in use" metrics respectively. The particular benefit of this series of reports lies in the overall quality of products – not especially in security.

By "technical security solution" we mean the actual constructs of the system. SDT metrics for system-level technical security solution can be detailed into respective design-level metrics emphasizing either (i) SW/HW platform security, (ii) application security or (iii) network security. Design-level security engineering metrics can be detailed into appropriate implementation-level metrics, mainly representing *vulnerability metrics*. According to CVSS (Common Vulnerability Scoring System), a vulnerability is defined as a bug, flaw, behaviour, output, outcome or event within an application, system, device, or service that could lead to an implicit or explicit failure of confidentiality, integrity or availability [35]. The Forum of Incident Response and Security Teams (FIRST) acts as the custodian of CVSS [12]. NIST's Software Assurance Metrics and Tool Evaluation (SAMATE) project [6] seeks to help answer various questions

on software assurance, tools and metrics. The metrics work being carried out in SAMATE is concentrating on metrics and measures for the software itself and SSA (Software Security Assurance) tools. OWASP (Open Web Application Security Project) [30] is an active discussion and development forum on security metrics. MITRE provides standardized languages as a means for accurately communicating the information and encouraging the sharing of the information with users by developing repositories [29].

4 CONCLUSIONS

"Measuring security" – obtaining enough evidence to be able to make informed decisions on information security issues – is one of the major challenges in information security. Security metrics is an emerging research area rapidly gaining momentum. Unless we are able to measure security phenomena on an adequate level, there will be no advancing leaps in the actual information security field. In this study, we have proposed a high-level taxonomy for security metrics, especially intended for the metrics development for industrial companies producing ICT products. The results of this study can be utilized in the future efforts to form a unified hierarchical security metrics system for ICT industry.

5 ACKNOWLEDGEMENTS

This research has been supported by the European Commission under the 7th Framework Programme through the GEMOM (Genetic Message Oriented Secure Middleware) STREP project, Grant Agreement No. 215327.

6 REFERENCES

1. Alger, J. I.: On Assurance, Measures, and Metrics: Definitions and Approaches. Proc. of Workshop on Information Security System Scoring and Ranking (WISSSR), Williamsburg, Virginia, May, 2001 (2002)
2. ANSI/ISA-TR99.00.01-2004: Security Technologies for Manufacturing and Control Systems Standards. ANSI, Washington, D.C. (2004)
3. Avižienis, A., Laprie, J.-C., Randell, B., and Landwehr, C.: Basic Concepts and Taxonomy of Dependable and Secure Computing. IEEE Trans. on Dependable and Secure Computing. Vol. 1, No. 1. (2004)
4. Basili, V. R. and Weiss, D. M.: A Methodology for Collecting Valid Software Engineering Data. IEEE Transactions on Software Engineering SE-10(6): 728-738, Nov. (1984)

5. Bellovin, S. M.: On the Brittleness of Software and the Infeasibility of Security Metrics. IEEE Security & Privacy, Jul/Aug, p. 96 (2006)
6. Black, P. E.: SAMATE's Contribution to Information Assurance. IAnewsletter, Vol. 9, No. 2 (2006)
7. Blakley, B.: An Imprecise but Necessary Calculation. Secure Business Quarterly: Special Iss. on Return on Security Investment, 1(2), Q4, (2001)
8. Bowen, P., Hash, J., Wilson, M.: Information Security Handbook: A Guide for Managers. National Institute of Standards and Technology Special Publication 800-100 (2006)
9. Burris, P., King, C.: A Few Good Security Metrics. METAGroup (2000)
10. Canadian System Security Centre: The Canadian Trusted Computer Product Evaluation Criteria, Version 3.0e, January 1993, 233 p. (1993)
11. FIPS Publication 199: Standards for Security Categorization of Federal Information and Information Systems. Federal Information Processing Standards Publication (2004)
12. FIRST: Forum of Incident Response and Security Teams. http://www.first.org/cvss/
13. Henning, R. et al.: Proceedings of Workshop on Information Security System, Scoring and Ranking – Information System Security Attribute Quantification or Ordering (Commonly but improperly known as "Security Metrics"), MITRE, Williamsburg, Virginia, May, 2001 (2002)
14. I3P: Institute for Information Infrastructure Protection. www.thei3p.org
15. Information Security Forum (ISF): www.securityforum.org
16. Information Security Forum (ISF): The Standard of Good Practice (SOGP). http://www.isfsecuritystandard.com/index_ns.htm (2005)
17. Information Technology Security Evaluation Criteria (ITSEC) Version 1.2, Commission for the European Communities (1991)
18. ISO/IEC 9126-1:2001: Software Engineering – Product Quality – Part 1: Quality Model. International Organization of Standardization (2001)
19. ISO/IEC 9126-2:2003: Software Engineering – Product Quality – Part 2: External Metrics. International Organization of Standardization (2003)
20. ISO/IEC 9126-3:2003: Software Engineering – Product Quality – Part 3: Internal Metrics. International Organization of Standardization (2003)
21. ISO/IEC 9126-3:2004: Software Engineering – Product Quality – Part 4: Quality-in-Use Metrics. International Organization of Standardization (2004)

22.ISO/IEC 15408-1:2005: Common Criteria for Information Technology Security Evaluation – Part 1: Introduction and General Model. International Organization of Standardization (2005)
23.ISO/IEC 17799:2005: Information Technology – Security Techniques – Code of Practice for Information Security Management. International Organization of Standardization (2005)
24.ISO/IEC 21827:2003: Information Technology – Systems Security Engineering – Capability Maturity Model (SSE-CMM). International Organization of Standardization (2003)
25.Jelen, G.: SSE-CMM Security Metrics. NIST and CSSPAB Workshop, Washington, D.C., June (2000)
26.Lennon, E. B. (Ed.): IT Security Metrics. ITL Bulletin, August 2003. National Institute of Standards and Technology (2003)
27.McCallam, D.: The Case Against Numerical Measures of Information Assurance. Proc. of Workshop on Information Security System Scoring and Ranking (WISSSR), ACSA and MITRE, Williamsburg, Virginia, May, 2001 (2002)
28.McHugh, J.: Quantitative Measures of Assurance: Prophecy, Process or Pipedream? Proc. of Workshop on Information Security System Scoring and Ranking (WISSSR), ACSA and MITRE, Williamsburg, Virginia, May, 2001 (2002)
29.MITRE: Making Security Measurable. http://makingsecuritymeasurable.mitre.org/
30.OWASP: Open Web Application Security Project. http://www.owasp.org/
31.Payne, S. C.: A Guide to Security Metrics. SANS Institute Information Security Reading Room, June (2006)
32.Ross, R., Johnson, A., Katzke, S., Toth, P., Rogers, G.: Guide for Assessing the Security Controls in Federal Information Systems. NIST Publication 800-53A (2006)
33.Sademies, A.: Process Approach to Information Security Metrics in Finnish Industry and State Institutions. VTT Publications 544. 89 p. + app. 2 p. (2004)
34.Savola, R. and Röning, J.: Towards Security Evaluation based on Evidence Information Collection and Impact Analysis. Suppl. Proc. of the 2006 Int. Conference on Dependable Systems and Networks (DSN),

Workshop on Empirical Evaluation of Dependability and Security (WEEDS), June 25-28, 2006, Philadelphia, PA, pp. 113-118.
35. Schiffman, M.: A Complete Guide to the Common Vulnerability Scoring System (CVSS). White paper.
36. Seddigh, N., Pieda, P., Matrawy, A., Nandy, B., Lambadaris, I., Hatfield, A.: Current Trends and Advances in Information Assurance Metrics. Proc. of the 2nd Annual Conference on Privacy, Security and Trust (PST 2004), Fredericton, NB, Oct. (2004)
37. Stoddard, M. et al.: Process Control System Security Metrics – State of Practice. I3P Institute for Information Infrastructure Protection Research Report No. 1, Aug. (2005)
38. Swanson, M.: Security Self-Assessment Guide for Information Technology Systems. NIST Special Publication 800-26, Nov. (2001)
39. Swanson, M., Bartol, N., Sabato, J., Hash, J., Graffo, L.: Security Metrics Guide for Information Technology Systems. National Institute of Standards and Technology Special Publication 800-55, Jul. (2003)
40. Toivonen, S., Lenzini, G., Uusitalo, I.: Context-Aware Trust Evaluation Functions for Dynamic Reconfigurable Systems. Models of Trust for the Web (MTW 06) Workshop, May 22-26, 2006, Edinburgh, Scotland.
41. United States Department of Defense: Trusted Computer System Evaluation Criteria (TCSEC) "Orange Book", DoD Standard, DoD 5200.28-std (1985)
42. United States National Computer Security Center: Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria – Version 1; NCSC-TG-005 (1987)
43. United States National Institute for Standards and Technology and National Security Agency, Federal Criteria for Information Technology Security – Draft Version 1.0, 2 volumes (1993)
44. Vaughn, R., Henning, R. and Siraj, A.: Information Assurance Measures and Metrics: State of Practice and Proposed Taxonomy. Proc. of 36th Hawaii Int. Conf. on System Sciences HICSS 03. (2003)
45. Vogel, C.: Cognitive Engineering. Masson, Paris, France (1988)
46. Yee, B. S.: Security Metrology and the Monty Hall Problem. Proc. of Workshop on Information Security System Scoring and Ranking (WISSSR), ACSA and MITRE, Williamsburg, Virginia, May, 2001 (2002)

BLOOM'S TAXONOMY FOR INFORMATION SECURITY EDUCATION

Johan van Niekerk[1], Rossouw von Solms[2]

[1]Nelson Mandela Metropolitan University
South Africa
[2]Nelson Mandela Metropolitan University
South Africa

[1]johan.vanniekerk@nmmu.ac.za, [2]rossouw.vonsolms@nmmu.ac.za

ABSTRACT

The importance of educating organizational end users about their roles and responsibilities towards information security is widely acknowledged. However, many current user education programs have been created by security professionals who do not necessarily have an educational background. The nature of such programs is thus not always properly understood. This lack of understanding could result in the ineffectiveness of security guidelines or programs in practice. This paper attempts to provide additional understanding of these programs through an examination of the revised version of Bloom's taxonomy. The paper show how this taxonomy could be applied to information security education.

KEY WORDS

Information Security, Information Security Education, Awareness, Bloom's Taxonomy

BLOOM'S TAXONOMY FOR INFORMATION SECURITY EDUCATION

1 INTRODUCTION

In recent years information technology has become such an intrinsic part of modern business that some authors no longer see the use of information technology as a strategic benefit. Instead, it can be argued that information technology is a basic commodity, similar to electricity, and that the lack of this commodity makes it **impossible** to conduct business (Carr, 2003). It is therefor vital for organizations to ensure that they have continuous access to this valuable commodity. The process of ensuring this continuous access is known as information security.

Humans, at various levels in the organization, play a vital role in the processes that secures organizational information resources. Many of the problems experienced in information security can be directly contributed to the humans involved in the process. Employees, either intentionally or through negligence, often due to a lack of knowledge, can be seen as the greatest threat to information security (Mitnick & Simon, 2002, p. 3). It is thus imperative for organizations that are serious about the protection of its information resources to be serious about the education of its employees. The aim of corporate information security education should be to ensure that each and every employee in the organization knows his/her responsibility towards information security.

This need to educate organizational users about their roles and responsibilities towards information security is in fact a well established idea. Most major information security standards address this need in some form. For example, the ISO/IEC standard 13335-1 states that organizations cannot protect the integrity, confidentiality, and availability of information in today's highly networked systems environment without ensuring that each person involved shares the security vision of the organization, understands his/her roles and responsibilities, and is adequately trained to perform them (ISO/IEC TR 13335-1, 2004, p. 14). In order to assist in ensuring information security, individual users thus needs **knowledge** regarding their specific role in the security process. This knowledge can be provided via education, training and awareness campaigns.

Most current information security educational programs are constructed

by information security specialists who do not necessarily have a strong educational background. Puhakainen (2006, pp. 33-56) reviews 59 current approaches to security awareness, most of which are not based on pedagogical theories. Puhakainen (2006, p. 56) also argues that there is a need for theory-based security approaches. These approaches should also be practically effective. The nature of security educational or awareness issues are often not understood, which could lead to programs and guidelines that are ineffective in practice (Siponen, 2000). A formally trained educationalist might, for example, raise the question whether or not **knowledge** is in fact enough. In Bloom's taxonomy, which is a well know and widely accepted pedagogical taxonomy, knowledge only comprises the very first, and lowest, level of education (Sousa, 2006, pp. 248-255). One could argue that this level of comprehension is in fact not adequate for most humans who play a role in the information security process. Similarly, the traditional approach of classifying the requisite information security educational needs as a continuum consisting of either awareness, training or education, might also be too simplistic.

This paper will attempt to provide a more pedagogically sound interpretation of the educational needs of humans involved in information security processes, based on their respective roles and responsibilities towards security, through the incorporation of Bloom's revised taxonomy (Anderson et al., 2001) as a pedagogical framework.

2 RESEARCH PARADIGM AND RATIONALE

The work in this paper is based on qualitative, or phenomenological-, research methods, as described in Creswell (1998). This paper should thus be seen as "an inquiry process of understanding based on distinct methodological traditions of inquiry that explore a social or human problem" (Creswell, 1998, p. 15). The research presented here does not attempt to define *new* knowledge, but rather to provide a more formalized understanding of information security *awareness*, *training* and *education*. As far as could be determined, the application of Bloom's Taxonomy, both the original and the revised versions, specifically to *information security* education has never been published before. It is the authors' belief that the use of this taxonomy could improve the understanding of the pedagogical issues that **should** be considered in any educational program, amongst information security specialists.

Since education, as a field of study, is normally seen as a ""human science" it was deemed fitting to also "borrow" the research paradigm used in this paper from the humanities. Most current work dealing with information security education see this education as a continuum consisting of three main levels, namely; awareness, training and education (Schlienger & Teufel, 2003),(Van Niekerk & Von Solms, 2004),(NIST 800-16, 1998, pp. 15-17). This continuum is used by many information security specialists when constructing information security educational campaigns. These specialists may not necessarily be educationalists. In order to ensure a rigorous research approach, this paper will thus revisit even concepts with a seemingly obvious meaning. The description and discussion of these concepts is deemed necessary because there might exist differences between the ontologies commonly adhered to by information security specialists and researchers from the educational sciences. The primary purpose of this paper is to encourage information security specialists to "borrow" from the humanities when engaged in activities that deals with humans. It can be argued that for most security education programs more knowledge of the underlying theoretical background can help both practitioners and scholars to understand why a particular information security awareness approach is expected to have the desired impact on users security behavior (Puhakainen, 2006, p. 139). It is believed that adherence to sound pedagogical principles when constructing information security educational campaigns, could improve the efficiency of such campaigns.

3 AWARENESS, TRAINING AND EDUCATION

As mentioned earlier, most current work dealing with information security education see this education as a learning continuum that "starts with awareness, builds to training, and evolves into education" (NIST 800-50, 2003, p. 7). NIST 800-16 (1998, pp. 15-17) provides more detail on the various levels of this continuum and describes these levels as follow:

- Awareness: The main purpose of awareness campaigns is to make employees "*aware*" of information security. In other words, these campaigns focus attention on security. This is normally done using techniques that can reach broad audiences. Awareness campaigns are generally aimed at **all** employees in the organization and aims to equip employees with enough knowledge to enable them to recognize poten-

tial security threats. Awareness is not training.

- Training: Training is more formal than awareness and have the goal of building employee knowledge and skills to facilitate the *secure* performance of the employee's normal tasks. Training strives to produce security skills and competencies that are relevant to the specific employee and needed in the performance of the employee's duties. "The most significant difference between training and awareness is that training seeks to teach skills that allow a person to perform a specific function, while awareness seeks to focus an individuals attention on an issue or set of issues."

- Education: "The Education level integrates all of the security skills and competencies of the various functional specialties into a common body of knowledge, adds a multi-disciplinary study of concepts, issues, and principles (technological and social), and strives to produce IT security specialists and professionals capable of vision and pro-active response."

In the current information society, educational or awareness issues affect almost all organizations. Despite this fact the nature of these programs are still not well understood and this often leads to ineffective security guidelines or programs (Siponen, 2000). Many organizations have some form of *awareness* program but often do not augment these with supporting training and/or education programs. The terms *awareness* and *education* are also often used interchangeably. It is not uncommon to hear security specialists talk about "awareness campaigns", when the campaigns actually focus on the training or education levels of the continuum. The purpose of these campaigns is often listed as instilling security **knowledge**, or fostering a **culture** of information security amongst organizational end-users (Van Niekerk & Von Solms, 2006). As mentioned earlier, the term knowledge only describe the lowest level of Bloom's taxonomy of the cognitive domain. From an educational viewpoint one could thus argue that the terminology used lacks rigor. This lack of rigor could contribute to the fact that the nature of awareness and educational issues is often misunderstood. One model that could possibly provide such rigor is Bloom's taxonomy.

4 BLOOM'S TAXONOMY OF THE COGNITIVE DOMAIN

Bloom's taxonomy is possibly one of the best known and most widely used models of human cognitive processes. Bloom's model was originally developed in the 1950's and remained in use more or less unchanged until fairly recently (Sousa, 2006, p. 249). A revised version of the taxonomy was published in Anderson et al. (2001). This revised taxonomy has become accepted as more appropriate in terms of current educational thinking (Sousa, 2006, pp. 249-260). Both versions of Bloom's taxonomy consist of six levels which increases in complexity as the learner moves up through these levels. Figure 1 shows both versions of this taxonomy.

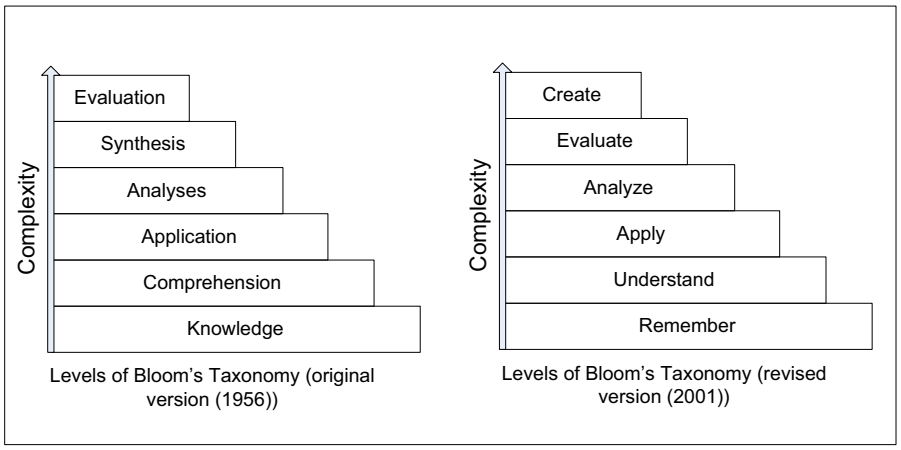


Figure 1: Blooms Taxonomy, Original and Revised (Adapted from Sousa (2006) pp. 249-250)

There are two main differences between the original and the revised versions of the taxonomy. Firstly, the revised version uses descriptive verbs for each level that more accurately describes the intended meaning of each level. Secondly, the revised version has swapped the last two levels of the original version around. This was done because recent studies have suggested that generating, planning, and producing an original "product" demands more complex thinking than making judgements based on accepted criteria (Sousa, 2006, p. 250). The hierarchy of complexity in the revised taxonomy is also less rigid than in the original in that it recognizes that an individual may move among the levels during extended cognitive processes. This pa-

per will focus on the revised version of the taxonomy. Wherever this paper mentions Bloom's taxonomy, it should be assumed that the revised version is intended, unless otherwise stated. The following is a brief explanation of each of the six levels of this revised taxonomy (Sousa, 2006, pp. 250-252):

- Remember: Remember refers to the rote recall and recognition of previously learned facts. This level represents the lowest level of learning in the cognitive domain because there is no presumption that the learner understands what is being recalled.

- Understand: This level describes the ability to "make sense" of the material. In this case the learning goes beyond rote recall. If a learner understands material it becomes available to that learner for future use in problem solving and decision making.

- Apply: The third level builds on the second one by adding the ability to use learned materials in *new* situations with a minimum of direction. This includes the application of rules, concepts, methods and theories to solve problems within the given domain. This level combines the activation of procedural memory and convergent thinking to correctly select and apply knowledge to a completely new task. Practice is essential in order to achieve this level of learning.

- Analyze: This is the ability to break up complex concepts into simpler component parts in order to better understand its structure. Analysis skills includes the ability to recognize underlying parts of a complex system and examining the relationships between these parts and the whole. This stage is considered more complex than the third because the learner has to be aware of the thought process in use and must understand both the content and the structure of material.

- Evaluate: Evaluation deals with the ability to judge the value of something based on specified criteria and standards. These criteria and/or standards might be determined by the learner or might be given to the learner. This is a high level of cognition because it requires elements from several other levels to be used in conjunction with conscious judgement based on definite criteria. To attain this level a learner needs to consolidate their thinking and should also be more receptive to alternative points of view.

- Create: This is the highest level in the taxonomy and refers to the ability to put various parts together in order to formulate an idea or plan that is new to the learner. This level stresses creativity and the ability to form *new* patterns or structures by using divergent thinking processes.

Educational taxonomies, such as Bloom's taxonomy, are useful tools in developing learning objectives and assessing learner attainment (Fuller et al., 2007). All well known educational taxonomies are generic. These taxonomies rely on the assumption that the hierarchy of learning outcomes apply to all disciplines (Fuller et al., 2007). Bloom's taxonomy would thus apply equally to a more traditional "subject", such as zoology, as to organizational information security education.

5 BLOOM'S TAXONOMY FOR INFORMATION SECURITY EDUCATION

Learning taxonomies assist the educationalist to describe and categorize the stages in cognitive, affective and other dimensions, in which an individual operates as part of the learning process. In simpler terms one could say that learning taxonomies help us to "understand about understanding" (Fuller et al., 2007). It is this level of meta-cognition that is often missing in information security education. According to Siponen (2000) awareness and educational campaigns can be broadly described by two categories, namely framework and content. The framework category contains issues that can be approached in a structural and quantitative manner. These issues constitute the more explicit knowledge. The second category, however, includes more tacit knowledge of an interdisciplinary nature. Shortcomings in this second area usually invalidate awareness frameworks (Siponen, 2000). How to really motivate users to adhere to security guidelines, for example, is an issue that would form part of this content category.

It has been shown that even in cases where users have "knowledge" of a specific security policy, they might still willfully ignore this policy because they do not understand *why* this policy is needed (Schlienger & Teufel, 2003). Answering the question "why" not only increase insight but also increases motivation (Siponen, 2000). Simply informing employees that "this is our policy", or "you just have to do it", which is often the traditional approach, is not likely to increase motivation or attitudes (Siponen, 2000). Learning is a

willful, active, conscious, and constructive activity guided by intentions and reflections (Garde et al., 2007). According to most constructivist learning theories, learning should be learner-centered (Garde et al., 2007). In an organizational information security educational campaign, the learners **must** include each and every employee. It is also important to realize that the campaign has to be **successful for each and every learner** (Van Niekerk & Von Solms, 2004).

In order to ensure successful learning amongst all employees, it is extremely important to fully understand the educational needs of individual employees. According to Roper, Grau, and Fischer (2005, pp. 27-36) managers often attempt to address the security education needs of employees without adequately studying and understanding the underlying factors that contribute to those needs. It has been argued before that educational material should ideally be tailored to the learning needs and learning styles of individual learners (Van Niekerk & Von Solms, 2004)(NIST 800-16, 1998, p. 19). One could also argue that awareness campaigns that have not been tailored to the **specific** needs of an individual, or the needs of a **specific target audience**, will be ineffective. It is in the understanding of these needs, that a learning taxonomy can play an important enabling role.

Information security specialists should use a taxonomy, like Bloom's taxonomy, before compiling the content category of the educational campaign. The use of such a taxonomy could help to understand the learning needs of the target audience better. It could also reduce the tendency to focus only on the framework category of these campaigns. For example, simply teaching an individual what a password is, would lie on the *remember*, and possibly *understand* level(s) of Bloom's taxonomy. However, the necessary information to understand *why* their own passwords is also important and should also be properly constructed and guarded might lie as high as the *evaluate* level of the taxonomy. An information security specialist might think that teaching the users what a password is, is enough, but research have shown that understanding *why* is essential to obtaining buy-in from employees. It is this level of understanding that acts as a motivating factor and thus enables behaviour change (Siponen, 2000)(Schlienger & Teufel, 2003)(Van Niekerk & Von Solms, 2004)(Roper et al., 2005, pp. 78-79).

The use of an educational taxonomy in the construction of information security educational programs requires that both the content and the assessment criteria for this program is evaluated against the taxonomy in order to ensure that learning takes place at the correct level of the cognitive do-

Level	Terms	Sample Activities
Create	imagine	Pretend you are an information security officer for a large firm. Write a report about a recent security incident.
	compose	Rewrite a given incident report as a news story.
	design	Write a new policy item to prevent users from putting sensitive information on mobile devices.
	infer	Formulate a theory to explain why employees still write down their passwords.
Evaluate	appraise	Which of the following policy items would be more appropriate. Why?
	assess	Is it fair for a company to insist that employees never use their work email for personal matters? Why or Why not?
	judge	Which of the security standards you have studied is more appropriate for use in the South African context? Defend your answer.
	critique	Critique these two security products and explain why you would recommend one over the other to a customer.
Analyze	analyze	Which of the following security incidents are more likely?
	contrast	Compare and contrast the security needs of banking institutions to those of manufacturing concerns.
	distinguish	Sort these security controls according to the high level policies that they address.
	deduce	Which of these procedures could derive from the given policy.
Apply	practice	Use these mnemonic techniques to create and recall a secure password.
	calculate	Calculate how secure the following password is.
	apply	Think of three things that could go wrong should your password be compromised.
	execute	Use the given tool to encrypt the following message.
Understand	summarize	Summarize the given security policy in your own words
	discuss	Why should non alpha-numeric characters be used in a password?
	explain	Explain how symmetric encryption works.
	outline	Outline your own responsibilities with regards to the security of customer account information.
Remember	define	What is the definition of a security incident?
	label	Label each of the threats in the given picture.
	recall	What is social engineering?
	recognize	Which of the pictures shows someone "shoulder surfing"?

Table 1: Bloom's Taxonomy for Information Security adapted from Anderson et al., 2001

main. The reference point for any educational program should be a set of clearly articulated "performance objectives" that have been developed based on an assessment of the target audience's needs and requirements (Roper et al., 2005, p. 96). Correct usage of an educational taxonomy not only helps to articulate such performance objectives but, more importantly, helps the educator to correctly gauge the needs and requirements of the audience. An example of how Bloom's revised taxonomy could be used in an information security context is supplied in Table 1. This example is not intended to be a definitive work, but rather to serve as an example or starting point for information security practitioners who want to use Bloom's taxonomy when constructing awareness and educational campaigns. It should however be clear that this taxonomy could easily be used to categorize most, if not all, information security educational needs effectively. Once categorized according to a taxonomy like Bloom's taxonomy, it should also be easier to find related information regarding pedagogical methods suitable to assist learners in attaining the desired level of cognitive understanding.

6 CONCLUSION

This paper suggested that information security educational programs would be more effective if they adhered to pedagogical principles. It was specifically suggested that the common categorization of security educational needs into the broad categories of awareness, training, and education, is not ideal. Instead an educational taxonomy, like Bloom's taxonomy should be used to accurately define the security education needs of organizational users. Through the use of such a taxonomy certain common weaknesses in current security awareness and educational programs might be addressed.

An example of how Bloom's taxonomy might be applied to information security concepts was provided. The primary weakness of this paper is the lack of empirical evidence to support the suggested use of Bloom's taxonomy. Future research in this regard should therefor focus on addressing this weakness. It has been argued before that security practitioners who engage in research or activities that relate to the human sciences should not re-invent the wheel, but should rather "borrow" from the humanities when appropriate. This paper is one such an attempt, to "borrow" from the humanities.

References

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., et al. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives, complete edition* (L. Anderson & D. Krathwohl, Eds.). Longman.

Carr, N. G. (2003). IT Doesn't Matter. *Harvard Business Review*, 41–49.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions. thousand oaks, ca: Sage, 1998.* Thousand Oaks, CA: Sage.

Fuller, U., Johnson, C. G., Ahoniemi, T., Cukierman, D., Hernán-Losada, I., Jackova, J., et al. (2007). Developing a computer science-specific learning taxonomy. *SIGCSE Bull.*, *39*(4), 152–170.

Garde, S., Heid, J., Haag, M., Bauch, M., Weires, T., & Leven, F. J. (2007). Can design principles of traditional learning theories be fullfilled by computer-based training systems in medicine: The example of campus. *International Journal of Medical Informatics*, *76*, 124–129.

International Standards Organization. (2004). *ISO/IEC TR 13335-1:2004 Guidelines to the Management of Information Technology Security (GMITS). Part1: Concepts and models for IT security. ISO/IEC, JTC 1, SC27, WG 1.*

Mitnick, K., & Simon, W. (2002). *The art of deception: Controlling the human element of security.* Wiley Publishing.

National Institute of Standards and Technology. (1998). *NIST 800-16: Information Technology Security Training Requirements: A Role- and Performance-Based Model. NIST Special Publication 800-16, National Institute of Standards and Technology.*

National Institute of Standards and Technology. (2003). *NIST 800-50: Building an Information Technology Security Awareness and Training Program. NIST Special Publication 800-50, National Institute of Standards and Technology.*

Puhakainen, P. (2006). *A design theory for information security awareness.* Unpublished doctoral dissertation, Acta Universitatis Ouluensis A 463, The University of Oulu.

Roper, C., Grau, J., & Fischer, L. (2005). *Security Education, Awareness and Training: From Theory to Practice.* Elsevier Butterworth Heinemann.

Schlienger, T., & Teufel, S. (2003). Information security culture - from

analysis to change. *Information Security South Africa (ISSA), Johannesburg, South Africa*.

Siponen, M. (2000). A conceptual foundation for organizational information security awareness. *Information Management & Computer Security*, *8*(1), 31-41.

Sousa, D. A. (2006). *How the brain learns* (3rd ed.). Corwin Press.

Van Niekerk, J., & Von Solms, R. (2004). Corporate information security education: Is outcomes based education the solution? *10th IFIP WG11.1 Annual Working Conference on Information Security Management, World Computer Congress (WCC), Toulouse, France*.

Van Niekerk, J., & Von Solms, R. (2006). Understanding information security culture: A conceptual framework. *Information Security South Africa (ISSA), Johannesburg, South Africa*.

7 ACKNOWLEDGEMENTS

The financial assistance of National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the National Research Foundation.

TOWARDS A FRAMEWORK FOR A NETWORK WARFARE CAPABILITY

N Veerasamy and JPH Eloff

Council for Scientific and Industrial Research

University of Pretoria

nveerasamy@csir.co.za

PO Box 395, Pretoria, 0002

ABSTRACT

Information warfare has surfaced as an emerging concept that affects not only military institutions but ordinary organisations as well. Information warfare in itself consists of various components ranging from its electronic and psychological aspects to its network enabled capabilities and functionality (network warfare). Various computer and information security practices form part of network warfare techniques. Whilst various information and security practices are well-known and applied by many, there is a need for a more structured approach to understanding the various techniques required for a network warfare capability.

A conceptual framework describing the most important network warfare techniques and considerations is proposed. This paper addresses the requirements for a network warfare capability and will look at the high-level approach, constraints, focus areas, levels, techniques and objectives. The framework therefore intends to present a more conceptual and structural examination of network warfare requirements and techniques. It should therefore provide a good baseline when establishing the capability or determining the practical consequences in any sector.

KEY WORDS

Information warfare, network warfare, framework, capability

TOWARDS A FRAMEWORK FOR A NETWORK WARFARE CAPABILITY

1 INTRODUCTION

As the world has moved into the Information Age, there is an increased need for the protection of the precious commodity information. The new emphasis is on Information warfare which is a modern type of conflict in which groups try to secure their own resources and thus prevent adversaries from denying and exploiting their information which would otherwise minimise capabilities. Information warfare refers to actions taken by the opposition to abuse information processing functionality to their benefit. Information warfare at its simplest level is the use of computers to attack an adversary's information infrastructure while protecting one's own information infrastructure [1].

Due to the increased use of computers and the connectivity afforded by networks, information can easily be stored and transported. The need also rises to properly protect these resources. Many users make use of global network connections to communicate and exchange information. However, there also exists the underworld community of hackers and abusers who seek to damage, destroy and deny access to information.

Networks have now become the battleground for various forms of attacks as vicious users attempt to deny and exploit networked resources. Network warfare is thus a form of information warfare in which the connectivity afforded by networks is utilised to carry out exploits on information. "The term netwar refers to an emerging mode of conflict (and crime) at societal levels, short of traditional military warfare, in which the protagonists use network forms of organisation and related doctrines, strategies and technologies attuned to the information age."

Network warfare is thus not only a military strategy but also application to the ordinary user domain as personal attacks and corporation exploitations are commonplace. Hacking attempts and the release of malware affect the average consumer and now ordinary users have become targets for netwar personal and social modes of attack. Protective software,

like anti-virus and firewalls all try to protect the corruption of information across networks. Thus, it can be seen that netwar is relevant to all users of computers and the Internet due to the large number of exploits and defense mechanisms in place.

Network warfare can be seen to encompass various computer and information security principles and techniques. Most computer and information security plans focus on the various security technologies that should be implemented. Whilst these techniques would fit into a network warfare capability plan, other aspects covering the objectives and other strategic factors of netwar have not been fully explored. Further investigation into the requirements and objectives of network warfare is necessary to understand the form of conflict that is being played out across the global community.

A key objective of a netwar capability will be protection and preservation of integrity of information. Previous research into a scheme of transferring data has been carried out to demonstrate the objective of protecting data and thus creating a stealthy means of transportation. The proof-of concept is explored in [2] and [3]. However, it has been identified that a framework of the high-level area of network warfare would be useful in identifying further requirements, objectives and influential considerations. Such a framework has been proposed in this paper.

This paper addresses the requirements for a network warfare capability and will look at the high-level approach, functional activities, constraints, capability requirements and objectives. The framework therefore intends to present a more conceptual and structural examination of network warfare requirements and techniques. It should therefore provide a good baseline when establishing the capability and determining the practical consequences in any sector.

The remainder of this paper is structured as follows; the background section provides an introduction topic of network warfare as a component of information warfare. Section 3 describes the need for understanding network warfare. Thereafter, the framework is introduced in section 4. The framework is further explored in Section5, before the conclusion is given in Section 6.

2 BACKGROUND

This section contains a brief introduction to network warfare as a facet of information warfare. We merely wish to provide the context of the concepts of network warfare and thus elaborate on the initial purpose of this paper. More detailed overviews can be found in other literature [[4],[5] and[6]].

Information warfare consist of the activities carried out in various domains (social, personal political and the military) that seeks to destroy, damage or deny information resources as well as the various defensive measures that are employed to prevent such attacks. Simply put, information warfare implies a range of measures or "actions intended to protect, exploit, corrupt, deny, or destroy information or information resources in order to achieve a significant advantage, objective, or victory over an adversary" (Alger, 1996, p. 12).[7].

The exploitation of information can have various consequences ranging from the psychological ramifications to the impact on control and management processes and the economic effects. The findings of the United States Air Force Armstrong Laboratory show that information warfare has the following unfolding types: command and control warfare, intelligence warfare, electronic warfare, psychological warfare, hacker warfare, economic information warfare and cyberwarfare [6]. Dai further argues that information warfare is composed of six 'forms': operational security, military deception, physcological war, electronic war (EW), computer network war and physical destruction [[8] in [9]]. Thus it can be seen that computers, networks, hacking and cyberspace have an active role to play in information warfare. Network warfare can be seen to encompass the various previously categorised computer and security related warfare concepts into a single branch that deals with information security at a computer and network level. Furthermore network warfare is not simply about the technological solutions that can be used to wage or defend against attacks. Network warfare also entails the high-level approach, strategy and plans to best protect, recover or attack if necessary. The term netwar connotes that the information revolution is as much about organisational design as about technological prowess and that this revolution favours whoever masters the network form [10].

Network warfare relates to the various types of network crime that takes place on computers and through networks and cyberspace to the many defense mechanisms that are deployed to prevent attacks and thus protect information. Arquilla and Rondfeldt explain that "many writers enamored of the flashy, high-tech aspects of information revolution have often depicted netwar as a term for computerised aggression waged via stand-off attacks in cyberspace- that is, as a trendy synomyn of infowar, information operations, "strategic information warfare" Internet war, "hackitivism, 'cyberterrorism, cybotage,etc' [4]". In many cases ordinary users are totally unaware that they are being hacked through cyberspace. The motive of perpetrators is unknown to the beguiling user: fun, profit or challenge are possible answers. A worrying aspect identified by Annual Review of Institute for Information Studies is that "hacking for fun" is being supplanted by hacking for profit as freelancers, businesses, governments and intelligence agencies turn to computer networks to facilitate both legitimate and criminal activities [11]. Therefore, in order to create an awareness of offensive and defensive approaches to network warfare, an understanding of high-level tasks and objectives is necessary. Users, companies and institutions need to be alerted of the new face of warfare that is not only being played out between military forces but also affects their personal and corporate activities.

Network warfare can be seen from different approaches which are often difficult to distinguish between. This blurring of offense and defense reflects a broader feature of netwar: It tends to defy and cut across standard spatial boundaries, jurisdictions, and distinctions between state and society, public and private, war and crime, civilian and military, police and military, and legal and illegal [10]. Offensive and defensive methods, legal and civilian boundaries, geography and physical limitations are all issues that are to be considered against the context of netwar. Network warfare is a multi-faceted issue that is facing the global community due to the increased ease and convenience of use of various computing and networking technologies.

The rest of the paper will examine the exact considerations that affect building a network warfare capability. The focus will turn to elaborating on all the key issues that could impact on a network warfare capability. However, first a brief motivation to understanding network warfare will be provided.

3 MOTIVATION

Users of computers and the Internet may be unaware of the various exploits that are taking place across cyberspace. Awareness needs to be created on the ease and type of netwar techniques in the ordinary user domain. Examples of attacking network warfare techniques range from web site defacement to malware that is unleashed on the Internet. Defensive network warfare techniques include the use of scanners and intrusion detection software to detect unwarranted actions on networks.

However, the use of various computer and security techniques does not fully describe the high-level objectives and considerations required to establish a netwar capability. "Although information warfare would be waged largely, but not entirely through the communication nets of a society or its military, it is fundamentally not about satellites, wires and computers. It is about influencing human beings and the decisions they make [12]. This highlights the need to realise that network warfare is not simply an issue of which technologies to deploy but has deeply embedded in its roots the requirement to formulate a strategy to influence the thought processes and thus the control and organisational structure to ensure that suitable management is applied and maintained. The network warfare capability would be quite limited if the focus was merely placed on the technologies. A greater understanding of the topic from a strategic point of view is required. This will aim to ensure that all influential factors have been considered.

Network warfare will become increasingly important due to the growing dependency on computers and networks. Certain security precautions and measures need to be instilled to try and prevent severe damage (financial and reputation for example).To provide a more thorough understanding of network warfare, a framework, considering key issues, has bee proposed. It is hoped that the framework will offer a more structured and formal overview of the topic so as to highlight the impacting factors and needs.

4 FRAMEWORK

In this Section, the framework is introduced, which allows for exploration of the topic thereof. Each component of the framework will be further discussed.

The framework is given in Figure 1. It mainly consists of three sections, the planning considerations, the techniques and the objectives which will each be discussed in Sections 5, 6 and 7 consecutively.

We consider four planning considerations: constraints/implications, target/focus, levels, and approach. Each planning consideration in turn consists of applicable sub-items. The planning considerations provide the context for which the network warfare techniques and objectives are trying to achieve. Various computer and network security techniques and principles are applicable to network warfare. However, a more formal and structured overview is shown to generate further discussion in the field.

The network warfare techniques are essential functionality that form part of a network warfare capability. The framework considers two locations of network warfare activities: own (interior measures) and foreign (outside zone) systems. The techniques described are not all-inclusive but offer examples of the type of activities that will be performed at different sites. The techniques provide an overview of the classes of activities that contribute to a network warfare capability.

Two categories of network warfare objectives are proposed: attacking (offensive) and protective (defensive). This serves to distinguish and classify the motives of the actors in network warfare events. An attacking objective implies an intention of damage, destruction or failure. A protective approach aims to prevent, defend and recover from harmful action.

The contribution of the framework lies in the taxonomy of the techniques as well as the overview of the additional considerations. A discussion elaborating on the various components of the framework therefore follows.

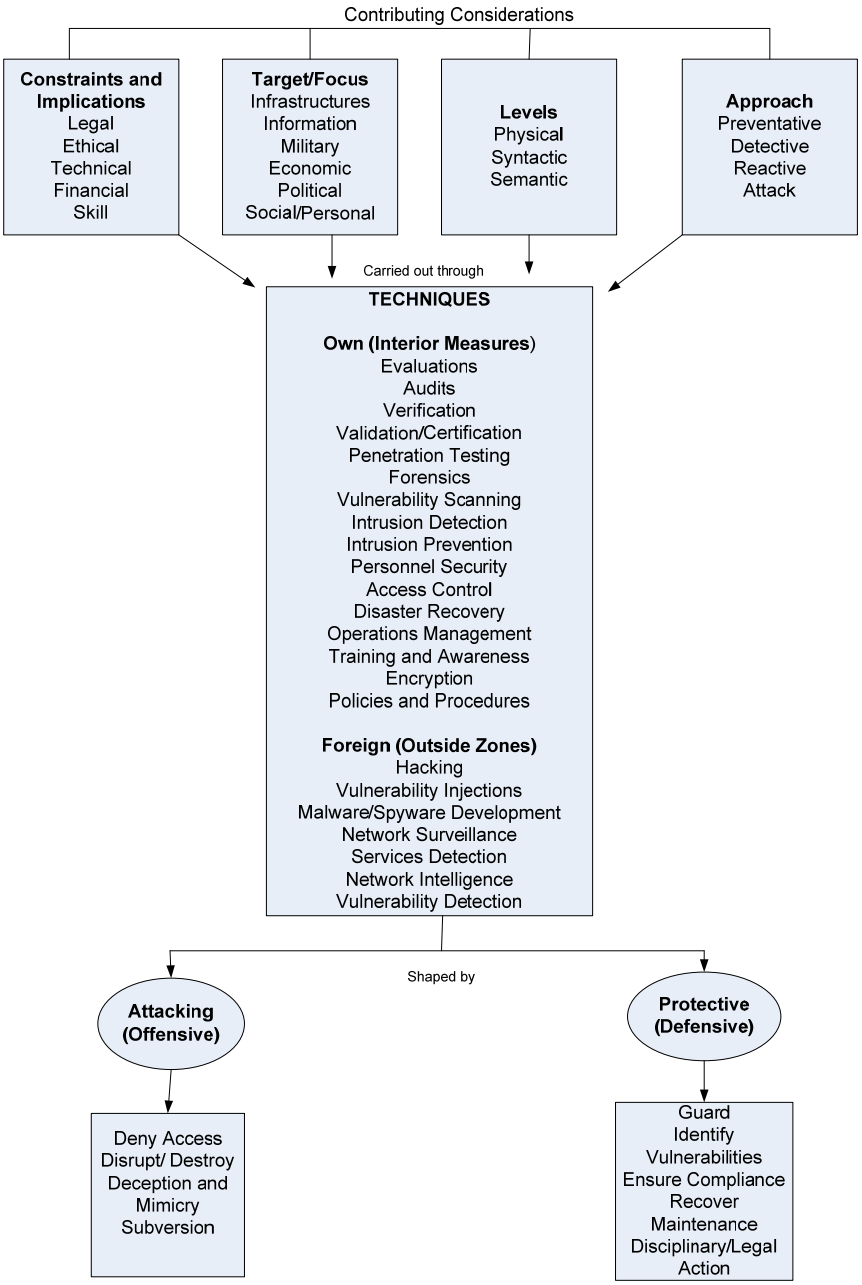


Figure 1. Conceptual Framework for Network Warfare

5 CONTRIBUTING CONDITIONS

This Section addresses the groups of contributing conditions that affect network warfare. Later sections will delve into detailed aspects of network warfare. These conditions explain the context of network warfare. Findings are drawn from an overview of literature as well as practical experience that helped identify the different contextual paradigms of network warfare. These conditions present the different approaches to the topic of network warfare as well as influential considerations in the conceptual framework.

5.1 Constraints and Implications

Several factors can constrain network warfare and have associated consequences. These include the legal issues, ethical dilemmas, technical solutions, financial impact and skill/manpower investment. Logical constraints/implications have been grouped together in the discussion that follows.

5.1.1 Legal Ethical Issues

As network warfare becomes more malicious, legal and ethical boundaries can be crossed. Alger proposes network warfare can cause potentially serious social problems and create novel challenges for the criminal justice system [7]. Criminal activities should not be condoned but the underlying causes of crime also needs to be understood. Ethics and morals play a significant role in determining the personality traits of an individual. Users will need to balance ethical dilemmas before engaging in offensive network warfare. Computers and network are powerful tools and great harm can be caused with them. Motive, attitude, values, upbringing, experience and culture can all impact the approach an individual can have when using computer resources. Socially a culture of responsibility and accountability needs to be instilled to ensure that users take precaution when using potentially harmful tools. A moral sense needs to be developed to ensure that users do not engage in malicious activity. On the other hand in the military context, officers will need to be trained to engage in offensive techniques. If one considers the mindset of a terrorist, a soldier, a network administrator or an activist, very different judgements and behaviour will be found. Network warfare is thus a double-edged sword and depending on the context, different actions (sometimes harmful with different legal and ethical implications) will be deemed necessary.

Computers and networks afford global connectivity with the blurring of physical and judicial boundaries. Detection of offences and the enforcing of laws thus becomes complex with the application of different rules and the difficulty of demonstrating proof aggravating the situation. Network warfare actions can thus cross the legal boundaries, as users participate in various malicious or surveillance tasks. In the military domain, however, different laws may be applicable. The Geneva Convention, and its interpretations allow for different treatment for soldiers, who are protected under its terms, and spies, for whom it offers no such protection [13]. In this way, different legal requirements can exist in the different societal and military domains.

5.1.2 Financial Impact

The release of a single virus can have a serious economic impact. A survey by TrustSecure/ICSA Labs in 2003 determined that the remediation cost of the MS Blaster worm was approximately $475 000 per company [14]. The results of the surveyed respondents in the CSI/FBI investigation showed a total loss of over $52 million (in 2006) for the many exploits ranging from viruses, theft and misuse [15]. One need only look at these incidents and figures to realize that the financial implications of security exploits have a considerable impact on the industry. Disruption to services and loss of availability leads to denial of accessibility to business operations or services which in turn has the impact of loss of productivity and thus business and revenue. Restoration; reparation and protection activities require additional resources (tools, hardware, software, etc) and staff, which too have financial costs involved. The financial consequences of security exploits is one of the most significant issues facing businesses when drawing up their IT budgets and business plans as the monetary amount to be spent, has to be determined, as well as factoring in potential losses. The investment in personnel, tools and skill development plays a significant role in any budget.

5.1.3 Technical Solution and Skill/Manpower Investment

Closely related to the financial costs of security exploits is the investment in additional staff and equipment. Defence and protections systems are implemented by network staff. The rising number of security exploits means that additional controls and personnel have been dedicated to preventing, detecting or repairing attacks. A large number of technologies like firewalls, software (anti-virus, anti-spy ware, and intrusion detection), encryption, and

biometrics are being utilized. Security evaluation tasks like auditing, penetration testing and monitoring require the use of both tools and human interaction to make informed decisions and actions. Employees will also need to be trained in the use and deployment of the various technologies. As the number of exploits rise sharply so does the requirements for improved security. This will entail the increased investment in the appropriate tools, equipment and employees.

5.2 Target/Focus

Network warfare has been shown to be applicable in various domains and is not just limited to military-based conflict. Information warfare may very soon become relatively commonplace: military, corporate/economic, community/social, and personal [16]. Different focus areas of network warfare are thus evident. Molander et. al also talk of different strategic targets which include information and the different information infrastructures (military, physical, economic, political and social). Social wars are being waged as users maliciously try steal identities, carry out fraud or create disinformation on the World Wide Web. Economic attacks seek to blockade economic information flow and thus impact markets [6]. Politicians seek to maintain friendly relations with allies and protect their reputations. Network warfare thus stretches its reach to various targets that have a global impact due to the organisational, private and peace-keeping efforts.

5.3 Levels

According to Libicki, from an operational point, systems can be attacked at a physical, syntactic and semantic level [5]. This in turn implies the existence of protective mechanisms at these levels. At a physical level, network warfare refers to the destruction of equipment and resources so as to substantially ruin/damage information in a tangible format or to prevent a future reproduction of its contents. Syntactic relates to the conformity to a systematic and orderly arrangement [13]. This implies the disruption to the organisational structures, for example causing a denial of service or interruption in data flow. Semantic affects the meaning of what computers receive from elsewhere [5]. This is linked to the receipt of correct data and the various means in which data can be poisoned and thereafter the continued spread as other devices are infected. Syntactic weapons, like

viruses, may be used to corrupt networked systems by destroying or degrading code or data; semantic weapons are used to affect and exploit the trust users have in the information system and the network, as well as to affect their interpretation of the information it contains [11]. The levels represent an aspect of viewing the type of impact network warfare techniques can have.

5.4 Approach

A functional paradigm of defensive information warfare is best described by the following actions: protect, detect and react [17]. Network warfare can thus defensively be approached from a preventative, detective or reactive point of view, as well as an attacking mode when looking at the opposite perspective. An attacking approach will seek to wreak damage, disruption or interruption to the system. Protective mechanisms endeavour to prevent/detect misdemeanours and also formulate a means to stop or recover from attacks. The factors relating to the approach taken to network warfare represent the high-level classification of the objectives. Specific techniques are needed to achieve each approach and individual objectives. This will be discussed next.

5.5 Techniques

The previous section described various conditions that can contribute to network warfare tactics and strategies. By keeping these considerations in mind, an understanding into the application areas of network warfare can be gained.

This section elaborates on specific techniques that can be used to carry out network warfare. Techniques represent the various ways and procedures that are followed in order to accomplish a complex task [18]. Network warfare is thus the complex task that can be carried out through various methods depending on the various contributing conditions and objectives. Although this section is by no means a complete listing of all possible network warfare techniques, it does cover a significant aspect of network warfare practices. A distinction has been drawn between techniques on own systems and those executed on foreign systems. The division is due to the differing strategic goals, information gathering purposes, legal implications and high-level objectives. A closer examination of each technique indicates the underlying goal which differs for interior and

exterior requirements. In each case, a different objective is trying to be achieved. The types of objectives will be further explored in next section.

5.5.1 Own (Interior Measures)

Companies, individuals and institutions often implement preventative, detective and reactive measures on their own system to protect, alert and recover from attacks. The International Information Systems Security Certification Consortium (ISC²) is a corporation that has developed a security certification program for information systems security practitioners worldwide. According to the ISC² a number of The Certified Information Systems Security Professional (CISSP) certification as endorsed by ISC², consists of domains that make up a Common Body of Knowledge (CBK). The domains that make up the CBK cover security topics like: Management, Cryptography, Operations, Disaster Recovery, Law and Physical Security [19]. The breakdown of security into the various domains is indicative that security has a very wide range of considerations Evaluations, audits and verifications seek to ensure that specific standards/measures are being adhered to in an effort offer proactive security and thus compliance and certification. Vulnerability scanning and intrusion detection activities aim to find vulnerabilities before/whilst they are being exploited to prevent further damage. Penetration testing is authorised attempts to determine whether the security controls in a system can be bypassed or if exploitable avenues are present. Forensics represents a branch of computer security searching for evidence of wrongful actions which can be utilised to hold users accountable for their behaviour. Disaster recovery planning ensures that crucial data is backed up and that a proper command structure is followed to get critical systems operational again after a crisis. Access control (biometrics, password policies, logon, auditing, physical security) aims to ensure that only authorised users are allowed entry into the systems and networks. Encryption obscures the contents of the data to protect its confidentiality. Policies, procedures, operations management and training seek to guide users to best practices relating to computer, information and network security which in turn will instil awareness on the topic.

The discussed techniques demonstrate the various means in which interior security measures can be implemented. Based on different objectives various protective, detective and reactive techniques will be used

to establish the capability. Measures taken on foreign systems will be discussed next.

5.5.2 Foreign (Outside Zones)

Hacking attempts are often targeted at outside systems. Motives often stem from profit and fun to political and military intentions. Another harmful netwar technique is vulnerability injections, for example exploiting a database query language vulnerability to insert incorrect data. Further malicious examples of targeting foreign systems include the development of malware and spyware. Security bulletins and web sites are filled with notifications of exploit and patch releases. More passive techniques used on outside systems to gain network intelligence include network surveillance (studying the behaviour of the enemy), services detection (to identify possible critical targets) and vulnerability detection (discover exploitable avenues).

Various techniques have been shown to form part of a network warfare capability. The execution of each task thereof depends on the underlying objective. To provide insight into the intentions of the various techniques, a high-level explanation of network warfare objectives follow.

5.6 Objectives

The previous sections addressed factors that can influence network warfare, as well as various techniques that can be employed to carry out network warfare. This section focuses on the issues of identifying the purpose and reasoning behind network warfare.

The objectives of network warfare have been divided into two categories: attacking (offensive) and protective (defensive). This shows two different mindsets: malicious versus maintaining security. As with any form of warfare, forces may have to attack to create advantage, as well as defend to prevent damage. Bhalla talks of two aspects of information warfare: defensive and offensive [20]. In a similar way network warfare has offensive components and a defensive strategy. The specific objective under each categorisation will be discussed next.

5.6.1 Attacking(Offensive)

“The objectives of information warfare can be masking or unmasking of facts, exploitation, deception (such as disinformation), disruption or denial

of service, and destruction of information [18]". This shows that the main aims of an offensive network warfare strategy would be deny access to a service, damage/destroy information, deception/mimicry, and subversion (insertion of malicious data). Denial-of-service attacks try to interrupt the use of specific systems. Breaking into machines (physically and electronically) to delete/alter data are forms of information damage and destruction. Unauthorised modification of data affects the accuracy of its contents. Various malicious modes of subversion have been unleashed in cyberspace (worms, viruses, Trojans, malware, spyware). The attacking objectives described are indicative of the offensive portion of computer and network security and thus shows how these malicious intentions are a core aspect of network warfare as a whole.

5.6.2 Protective (Defensive)

From a protective point of view, network warfare attempts will aim to secure the system from attacks. It is shown that defensive objectives include: guarding, vulnerability identification, recovery, maintenance and disciplinary/legal action. Guarding the system will seek to offer protection and thus prevent (and detect) attacks. In a similar manner, vulnerability identification aims to identify possible ways of exploitation. Maintenance consists of ensuring that the specific technologies are performing their defensive roles as well instilling good practices in users so that they remain aware of the risks of poor security. Disciplinary/legal action ensures that users are held accountable for their actions. Network warfare protective mechanisms/techniques aim to ensure that the system is secure and try to guard against malicious activity.

6 CONCLUSION

This paper addresses network warfare as an influential consideration facing global users of computers, networks, the Internet and cyberspace in general. Network warfare forms a critical branch of Information Warfare. The focus of Network Warfare lies heavily in the in the computer and network means through which information can be attacked and the various ways of protecting such resources. Various computer and network security issues form part of network warfare. However, other considerations too were shown to impact the area of network warfare. A more structured means of elucidating the field of network warfare was therefore required. Through an

analysis of the topic, it was revealed that network warfare can be executed through various techniques with different objectives, approaches, constraints and target and levels.

This paper took a high-level look into network warfare and a proposed a framework. The framework aims to present a more conceptual and structural examination of network warfare requirements and techniques. It should therefore provide a good baseline when establishing the capability or determining the practical consequences in any sector.

The framework proposes contributing considerations, techniques and objective. Four groups of contributing considerations are addressed: constraints/implications, target/focus, level and approach. Network warfare is often only linked to its military context. It has been shown that network warfare is applicable to many other domains, including social, political and economic. An investigation of computer and network technologies revealed a number of enabling techniques for network warfare. As network warfare involves a form of conflict, with any battle there exists an offensive component and a defensive aspect. Different offensive and defensive techniques were thus identified and discussed.

Further research into understanding other areas that impact network warfare can be incorporated into the design of the framework. The framework, by itself is a good starting point for placing the concept of network warfare into context. It is hoped that further analysis, can provide the ability to extend the framework.

7 REFERENCES

[1] *Information Warfare, Are you at risk?*, A J Elbirt, IEEE Technology and Society Magazine, 2003/2004.

[2] *Stealthy Network Transfer of Data*, N Veerasamy & CJ Cheyne, Proceedings of World Academy of Science, Engineering and Technology, Vol 25, November 2007.

[3] *Stealthy Network Transfer of Data*, N Veerasamy & CJ Cheyne, International Journal of Computer Science and Engineering, Vol 1 no 3, Summer 2007.

[4] *Networks and Netwars*, J Arquilla & D Ronfeldt, Rand, 2001.

[5] *What is Information Warfare?*, M Libicki, Strategic Forum Number 28, May 1995

[6] *Situation Awareness, Information Dominance and Information Warfare,* MR Endsley & WM Jones, Tech-Report 97-01 , February 1997.

[7] *Information warfare: Its Application in Military and Civilian Contexts*, B Cronin & H Crawford, School of Library and Information Science, Indiana University, Indiana USA.

[8] *On Integrating Network Warfare and Electronic Warfare*, Q Dai, Zholgguo Junshi Kexue (China Military Science), February 2002, 112-117 as translated by the foreign Broadcast Information Service (FBIS) Web site.

[9] *Chinese and American Network Warfare*, T L Thomas, JHQ, issue 38.

[10] *The Advent of Netwar,* J Arquilla & D Rondfeldt, RAND, 1996.

[11] *The Promise of Global Networks*, Nortel Networks and Aspen Institute, Institute for Information Studies, 1999.

[12] *Information Warfare*, G J stein, Airpower Journal, No 1, pp 30-39, Spring 1995. [27]

[13] *Syntactic and Technique,* The Free Online Dictionary, Available online from http://www.thefreedictionary.com, Accessed 27 February 2009.

[14] *Cert Statistics,* Carnegie Mellon University, Available online http://www.cert.org/stats/cert_stats.html, Accessed 1 June 2007.

[15] *CSI/FBI Computer Crime and Security Survey,* Computer Security Institute (CSI), Available online from http://i.cmpnet.com/gocsi/db_area/pdfs/fbi/FBI2006.pdf, Accessed 1 June 2007.

[16] *Conflict and the Computer: Information Warfare and Related Ethical Issues,* S Nitzberg, in Proceedings of National Information Systems Security Conference, National Institute of Standards and Technology,1998, Available online from csrc.nist.gov/nissc/1998/proceedings/paperD7.pdf

[17] *Defensive Information Warfare,* B Panda & J Giordano, Communications of the ACM, vol. 42 no. 7, pp 31-32, July 1999.

[18] *Revolution in Information Affair*, MW Wik, Available online from: http://www.kkrva.se/Links/Infokrig/Wik1.html, Accessed 13 February 2008.

[19] *CISSP All-in-One Certification Guide* ,Harris, S., Mcgraw-Hill/Osborne, 2002.

[20] *Is the Mouse Click Mighty Enough to Bring Society to its Knees,* N Bhalla, Computers & Security, vol. 22, issue 4, pp 322-336, May 2003.

PART 2

RESEARCH IN PROGRESS PAPERS

BCP/DRP CASE STUDY: ADAPTING MAJOR INCIDENT HANDLING RESPONSE FRAMEWORKS TO A CORPORATE ENVIRONMENT

Pieter Blaauw

Pick 'n Pay Information Systems
pblaauw@pnp.co.za

ABSTRACT

Business continuity is defined as "The degree to which an organization may achieve uninterrupted stability of systems and operational procedures during and after a disruptive event". Business continuity and disaster recovery plans and the very need for them became to a frightening reality after September 11 2001. Never was the emphasis on disaster recovery and business continuity placed so heavily on business following a world event, and it has shaped how modern business approach these two interwoven aspects of their business.

Fortunately not every business suffers a major disaster in its lifetime. On the contrary, many businesses today, large or small, face what can better be described as challenges, but these can have far reaching effects on the bottom line. The need for incident handling frameworks in business has thus come to light, and this paper will look at how one corporate in South Africa has adapted a major incident handling framework to their unique environment, with the hope that it can shed some light for other corporations to adapt to disruptive events in their business. The paper will cover severity levels, minor incidents, major incidents and the processes followed in the incident handling framework.

KEYWORDS

Major Incident Handling, Minor Incidents, Major Incidents

BCP/DRP CASE STUDY: ADAPTING MAJOR INCIDENT HANDLING RESPONSE FRAMEWORKS TO A CORPORATE ENVIRONMENT

1 INCIDENT MANAGEMENT

An incident is any event that disrupts a companies normal operations, no matter how small. Incidents happen on a daily basis across companies but the majority of these incidents have a very low impact. Nevertheless all incidents need to be managed properly to ensure that the effect on the company is minimized. Incident handling is not simply a technical function either. Policies and procedures need to be in place prior to the incident(s) taking place.

1.1 Incident classification

There are several ways that companies classify incidents and their severity levels. Some companies use the red, orange and green systems, others use a numerical system. Within Pick 'n Pay Information Systems, a numerical system is used, where the incident is classified depending on its severity and impact on the company, with a severity level 1 being only a notification of a possible problem, and a severity level 4 being a major system or service outage.

This numerical system does serve as a guide only and each incident is dealt with based on it's impact on the organization and not the underlying Service Level Agreement or business process. This ensures that incidents are escalated quickly and gives the organization time to react to the incident should the severity of it require such action.

	Routine Incident
Severity 1	Routine Daily Issue Zero or Minor Business Impact Localized scope, closed within a few hours
	Significant Incident
Severity 2	Service degradation or Outage. Impact to several users. Escalation of issue longer than a business day.
	Serious incident
Severity 3	Serious service degradation or outage affecting a entire line of business
	Severe Incident
Severity 4	Severe outage affecting one or multiple areas of business Threat to actual loss of reputation, trading loss, and or settlement impact. Requires senior management control

1.2 Minor Incidents

Minor incidents occur on a regular basis. On their own they cause minimal disruption to the company. However, there needs to be an appropriate response to them since they can have a far reaching impact if not checked.

1.3 Major Incidents

As incidents increase in severity, and as such their impact on the company, so the response to them should similarly increase. Routine and Minor incidents usually have a fairly standard response, according to the policies and procedures of the company, and more severe incidents are handled by enhancing the standard business arrangements.

The most severe incidents (Severity 4 in our table) i.e. those affecting the widest area in business, requires a more structured incident handling process more complicated than the normal business processes, referred to as a Major Incident Handling Framework.

2 MAJOR INCIDENT HANDLING

2.1 Process

Major Incident Handling frameworks are management structures applied to the response plans and arrangements for the most severe incidents in a business. The precise structure and response procedures varies from business to business, and also from business unit to business unit. However, in any other company, like Pick 'n Pay, the objective of these procedures remains the same, to ensure there is an effective and appropriate response to all major incidents that can affect the business.

Business units in any company, like Pick 'n Pay, have different priorities and business drivers. It stands to reason that because of this, each unit's MIH process will vary according to their circumstances and priorities. There are some shared features and roles in every MIH, no matter where it is applied and what the company may be.

Incident Owner: The individual or business unit responsible for the resolution of the incident in the business. If the resolution lies outside the company, in this case, Pick 'n Pay, it lies with the management responsible for the SLA with the outside solution provider.

Communications: In any event or incident, communication is crucial in the handling of the event. In a major incident handing framework it provides for the accurate and timely communication of information to all the affected parties inside and outside of the business. Communication provides those with a role in the incident handling procedures with all the information they require and relieves those who are dealing with the incident with having to provide the rest of the organization with information they require.

Technical Recovery: Where the cause of the incident is within the organization, it's important that it gets resolved within the organization. Technical recovery relates to the underlying technical issues and persons involved in recovering from the incident.

Business Recovery: A major incident will disrupt the organization's business. The extent of the interruption may be sufficient to warrant the implement and certain plans and arrangements to manage the disruption in order to ensure that critical business processes are maintained as far as possible during the disruption.

Due to the differences in organizations the way that these procedures happen varies from company to company. The following activities needs to be covered in each process though.

Assessment: Once an incident has occurred it needs to be assessed in terms of its impact on the company and its operations and business processes. A severity level will then be assigned to the incident and the severity level then determines the response to the incident. An important aspect to keep in consideration is that a incident's severity level can change during its life.

Warm-up: If an incident has been classed as a Severity 3 or 4, and it has been determined that the full MIH process needs to be invoked, a warm-up period is required. This is to assemble the required team and assign the correct roles to each member of that team.

Incident Management Cycle: Once a MIH process has been established, it will consist of a set of procedures that will repeat itself through a cycle. How often and when the cycle repeats itself depends on the incident and the organization.

Assessment: Asses the situation and device a plan

Activities: Implement the plan

Review: Look at the action to determine if the plan was able to resolve the situation

Communication: Update those involved in the previously defined steps / plan

Cool Down: Once the incident has been resolved it is necessary to stand down the MIH process. This means communicating with all the involved parties that the incident has been resolved.

Post Incident Review: On severity 4 incidents its recommended that a Post Incident Review take place to establish what the cause was and how corrective actions can be put in place to prevent a repeat of similar incidents.

3 WATCH STATES

Watch States are used in MIH to heighten awareness in three situations

There is a reason to believe that an incident may occur

There is a reason to believe that an incident which has occurred may escalate further

An incident that has occurred has not been fully closed

The purpose of a Watch State is to monitor the situation closely and facilitate the response and escalation to the next level. When a watch state is declared, consideration needs to be given to:
Who should be advised of the situation
Who should be advised that they may required to respond to the situation
Preparations that need to be carried out to deal with a possible escalation

Once an incident has been declared and the use of the Watch List implemented it needs to be reviewed on a ongoing basis. The nature of the incident will determine how often. As an example, in Pick 'n Pay, a Point of Sale incident will be reviewed far more frequently than a e-mail incident, due to the very nature of the business and the affect it has on the business.

4 INCIDENT ESCALATION

As seen from the previous sections, of the paper, an incident can be classified at any one of the four levels of severity. During the MIH process the incident can then be escalated to another level as deemed necessary. While in theory the incident owner should be responsible for escalating (and de-escalating) the incident between levels, in Pick 'n Pay a single problem manager is responsible for this task, and consults with various people before taking the necessary action.

By using a Major Incident Handling framework, making these decisions with the aid of the BCM or BCC and other stakeholders, and defining actions that need to be taken becomes that much easier and simpler.

5 POST INCIDENT REVIEW

Incidents represent a disruption to any company's normal operations. It is therefore in the company’s best interest to minimize the number of incidents and their severity. It stands to reason that after every severity 3 or 4 a Post Incident Review needs to take place to identify the cause of the incident and if possible, take action to reduce the chances and risk of the incident repeating itself.

The extent of the review will be determined by the extent of the severity and the affect it had on the business. For small severity 1 and 2 incident just recording the time and incident may be enough. Should there be a small number of the same incidents it can identify a trend which could then allow for corrective action to be taken.

For more severe incidents (especially severity 4), for example a Point of Sale failure, a more in-depth review needs to take place, to identify the underlying cause. If possible this cause should then be addressed to prevent a possible repeat of the incident.

6 REFERENCES

EPT Consulting
(http://www.etpconsulting.co.uk/Business%20Continuity/business¬continuity-glossary.htm)

Rittinghouse & Ransome, Business Continuity and Disaster Recovery for Infosec Managers, 2005

Hunton, Bryant & Bagranoff, Core Concepts of Information Technology Auditing, 2004

USING OBJECT-ORIENTED CONCEPTS TO DEVELOP A CONCEPTUAL MODEL FOR THE MANAGEMENT OF INFORMATION PRIVACY RISK IN LARGE ORGANISATIONS

Kamil Reddy and H.S. Venter

Information and Computer Security Architecture Research Group
University of Pretoria

{kreddy, hventer}@cs.up.ac.za

ABSTRACT

In this paper we present a conceptual model for the management of information privacy risk in large organisations. The model is based on the similarities between the concepts of departments in large organisations and the object-oriented computer programming paradigm. It is a high-level model that takes a holistic view of information privacy risk management, and, as such, identifies risk in both manual and automated processes during the acquisition, processing, storage and dissemination of information. While conceptual in nature, the model is well suited to practical implementation due to the structure it derives from the object-oriented paradigm. The practical application of the model is demonstrated by way of an example scenario.

This paper contributes by addressing the absence in the literature of freely available models for the holistic management information privacy risk in large organisations.

KEY WORDS

Information Privacy, Information Privacy Risk, Information Privacy Management

USING OBJECT-ORIENTED CONCEPTS TO DEVELOP A CONCEPTUAL MODEL FOR THE MANAGEMENT OF INFORMATION PRIVACY RISK IN LARGE ORGANISATIONS

1 INTRODUCTION

Many organisations acquire, store, process or disseminate information related to individuals. These organisations are often bound by law [1, 2] to protect the interest individuals have in accessing, controlling, or significantly influencing, the veracity and use of their information. This interest is termed *information privacy* [3]. Where information privacy is not adequately protected by an organisation, affected individuals may seek legal recourse against the organisation. This may result in the organisation suffering financial loss and damage to their reputation. *Information privacy risk* (IPR) is the collective term for risks that lead to such breaches of information privacy.

Large organisations generally require a more coordinated and formal approach to their operations than smaller ones [4, 5]. The effective management of IPR in large organisations is therefore particularly important. In this paper, we present a high-level conceptual model that can be used to assist in the management of IPR. As it is a high-level model, it is designed for use by those charged with the overall management of privacy protection within an organisation or department. The model is based on the similarities between departments in large organisations and the object-oriented computer programming paradigm. It is holistic because it addresses IPR in both manual and automated processes during the acquisition, processing, storage and dissemination of information. In order to address the various types of information privacy breaches, the model makes use of the Organisation for Economic Cooperation and Development's Guidelines on the Protection of Privacy and Transborder

Flows of Personal Data (OECD guidelines) [6]. The OECD guidelines set out principles for the ethical handling of private information. The principles contained in the OECD guidelines form the basis of information privacy law in most countries [7, 8].

The rest of this paper is structured as follows. Section 2 describes the related work in the literature. Section 3 provides background on the object-oriented programming paradigm and the OECD guidelines. It also defines what we term the object analogy. The model is described in Section 4. A hypothetical scenario that uses the model is provided in Section 5. Section 6 consists of a discussion of the model. The paper is then concluded in Section 7.

2 RELATED WORK

In this section we discuss the related work found in our review of the literature.

Our search of the literature revealed only a single example of a privacy management model. This model, called *Privacy by 3PT*® [9], is the proprietary work of a company called the Corporate Privacy Group and is hence not freely available for use or public scrutiny. As such, it was not possible to analyse the model in detail. From the company's own literature on the model [9], it focuses on people, policies, procedures and technologies as distinct areas of concern. We do not devote further attention to the details of this model due to the lack of publicly available information. From the information that is available, our model differs in its areas of focus. Our model is also less prescriptive with regard to implementation steps and methods.

Karjoth and Schunter [10] developed a privacy policy model for the specification and enforcement of organisation-wide privacy policies. Their work was extended to form IBM's Enterprise Privacy Authorization Language (EPAL) [11]. EPAL is a formal language and is concerned with the enforcement of policies within information technology (IT) systems [11]. Our work differs from EPAL because our work is applicable at a higher level, and is not concerned only with IT systems.

® 'Privacy by 3PT' is a registered trademark of the Corporate Privacy Group

Casassa Mont [12, 13] also addresses the management of private information in organisations through the use of privacy obligations. *Privacy obligations* are policies that specify the duties and expectations under which organisations must manage private information [12]. Although privacy obligations are considered in EPAL, he develops them in greater detail [13]. Casassa Mont's work is also at the system level and therefore different to our approach.

Biskup and Brüggemann [14, 15] developed DORIS (*Datenschutz-orientiertes Informationssystem*). DORIS is a prototype implementation of a system based on *The* P*ersonal Model of Data*, a model also developed by Biskup and Brüggemann [14, 15]. In The Personal Model of Data the world consists only of entities called 'persons*'*. '*Persons*' represent individuals in the real world. DORIS uses objects to represent 'persons'. The objects consist of attributes and methods. Attributes correspond to an individual's knowledge of themselves in the real world. Methods correspond to actions taken by the individual in the real world. Biskup and Brüggemann also develop a data model, data manipulation language and rights-based privacy policy that are used in the DORIS system. We do not elaborate on these due to space restrictions.

Our work is similar to that of Biskup and Brüggemann in that we also make use of objects. It differs, however, because we make use of similarities between the concepts of organisational departments and objects, while Biskup and Brüggemann uses objects to model individuals. In our work we also go into greater detail regarding the object metaphor. Unlike our work, which is a high-level model, theirs is restricted to enforcing privacy within a single system. Another significant difference is that Biskup and Brüggemann take a view of privacy that is limited to ensuring appropriate access to private information. Their work does not consider the other aspects of information privacy as espoused in the OECD Guidelines.

3 BACKGROUND

In this section we present the background necessary to understand our model and the rationale behind it. We discuss the object oriented programming paradigm, organisational departments, the object metaphor and the OECD guidelines.

3.1 The Object-Oriented Programming Paradigm

We divide our discussion of object-oriented programming (OOP) paradigm into a discussion of the concepts behind the paradigm and brief example of how it is used.

3.1.1 Defining OOP Concepts

There is no single set of concepts that is universally accepted as making up the OOP paradigm [16]. It is, however, most commonly characterised as consisting of three concepts: objects, classes and inheritance [17]. An *object* can be defined as "an individual, identifiable item, either real or abstract, which contains data about itself and descriptions of its manipulations of the data" [16]. The data contained in an object are called *attributes*, while the manipulation of the data is achieved through *methods*. An object's *set* methods are used to input data or to change existing data in the object. An object's *get* methods, on the other hand, may used by other objects to retrieve data from the object.

The concept of *encapsulation* ensures that access to an object's attributes and methods from 'outside' the object is strictly limited according to the definition of each attribute and method. We mention encapsulation in addition to the three concepts listed above because it is also often associated with the OOP paradigm [16] and it is relevant to our model.

We use the definitions in Armstrong [16] to define a *class* as an abstraction of an object that defines the common structure and behaviour shared by a set of objects. An object which belongs to a class is thus a 'concrete' instance of the class. The verb *instantiate* is used to denote the creation of an object from a class definition. *Constructors* are special methods used to instantiate objects. Attribute values may be set at the time of instantiation using a constructor. The accessibility of an object's attributes or methods, required for encapsulation, is specified in an object's class definition.

3.2 The Object Analogy

In a large organisation private information is typically used by one or more departments, for example, the finance and marketing departments. In each department the information may be stored as well as manipulated. By manipulated we mean received, processed, disseminated or any combination

thereof. Departments generally also use a fixed number of known methods for storing and manipulating information. This is analogous to an object in the sense that information in an object can be stored using attributes, and manipulated using methods. The direct analogy between objects and departments is termed the *object analogy*. The object analogy is illustrated in Table 1.

Table 1 – The Object Analogy

Department	**maps to**	**Object**
Type of department	→	Class
Department	→	Object
Information	→	Attributes
Receipt of information	→	Get and set methods, constructors
Processing	→	Methods that change attribute values
Storage	→	Variables for storing attributes
Dissemination	→	Get and set methods, constructors
Use of appropriate information handling methods only (Controls)	→	Encapsulation

As discussed earlier, the information contained in an object, as well as the methods used to manipulate the information, are strictly defined in the object's class definition. In addition to this, encapsulation ensures that only the appropriate, predefined, methods are used to manipulate the information. Viewing departments as objects therefore requires that: 1) all information in a department must be defined and, 2) all methods used for manipulating the information in the department must be defined. Since our model is only concerned with information privacy, this requirement applies only to private information and the methods used to manipulate private information.

The definition of all private information and related methods is the first step in the protection of information privacy. This is because it is impossible to protect information if one does not know it exists, or, if one does not know where or how it is stored and used. Once all private information and related methods in a department are defined, controls may be used to protect the information. Eloff and von Solms [18] define *controls* as measured steps taken to achieve a specific objective. In our case, the objective is to limit breaches of information privacy. Their definition, however, is not detailed enough for our purposes. Hence, we adapt the definition in the COBIT framework [19] to define *information privacy controls* (IPCs) as the policies, procedures and practices designed to provide reasonable assurance that information privacy breaches will be prevented, or detected and corrected. IPCs correspond to encapsulation in the OOP paradigm.

3.3 The OECD Guidelines

The OECD guidelines contain a set of principles referred to as the Fair Information Principles (FIPs). The FIPs provide guidance on the ethical handling of private information. The FIPs were first published in 1973 in a report by the United States Department of Health, Education, and Welfare [20]. Only four principles were listed, but these have since been developed to eight in the OECD guidelines. Globally, information privacy law is based on the FIPs [7, 8]. Due to space restrictions, we do not elaborate on the principles. For the same reason we do not list them here as they are listed in our model.

4 CONCEPTUAL MODEL

In this section we present our conceptual model. We describe each of the four elements in the model. These are: attributes, methods, controls and relationships. We then describe the interrelation between the elements in Figure 1 and show the full specification for the model in Figure 2.

4.1 Attributes

Our model is based on the object analogy. As such, we view a department as an object. The private information used by a department is represented by attributes. Attributes are classified as *electronic* if they are stored in electronic form or manual if they are stored manually (e.g. on paper). In

addition, attributes may be classified as *abstract* if they refer to entities which information pertains to. For example, a paper curriculum vitae (CV) would be defined as manual attribute. The job applicant to whom the CV belonged would be defined as an abstract attribute. This is because the job applicant himself is not stored by the department.

4.2 Methods

The different tasks related to private information in a department are represented by methods. In Solove's Taxonomy of Privacy [21], tasks belong to one of the following three categories: 1) information collection, 2) information processing and 3, information dissemination. We use this notion to classify all methods in the model as *input*, *processing*, or *output* methods. Input methods indicate how information enters a department, while output methods represent how information is passed from a department to outside entities. The term 'outside entities' may refer to another department in the same organisation, or it may refer to another organisation or individual. Processing methods represent the different ways in which the information can be used by a department. Since methods relate to information, each method is related to an attribute in the model. All methods are classified as either manual or electronic.

4.3 Controls

To ensure that there is a reasonable chance that methods do not result in an information privacy breach, we introduce the *controls* element to the model. IPCs are specified here for each method. Each of the IPCs protects one or more of the FIPs in the OECD guidelines. Input, processing and output methods each have a subset of the FIPs associated with them. For example, input methods must have controls for the following FIPs: Openness, Collection Limitation, Purpose Specification, Data Quality, Security and Accountability. By ensuring that controls enforce the FIPs, the model protects organisations against IPR.

4.4 Relationships

The final element of the model is *relationships*. The relationships element describes the flow of private information between the department and outside entities. The information flow consists of the name of the outside entity and a description of the information being transferred. Information is

represented by attributes, therefore attributes are included in the description of a relationship. However, sometimes only specific pieces of information in an attribute may be transferred to or from an outside entity. For example, take the case of an organisation that outsources its customer service function to an outside entity. It may only send the entity a list containing customer names and telephone numbers rather than the complete customer record for each customer. In the model we call these specific pieces of information *attribute primitives*. In the previous example, a name and telephone number are considered attribute primitives. Attribute primitives are specified in the model using capital letters and quotes to differentiate them from ordinary attributes. Thus, the attribute primitives for a name would be specified as "NAME".

The interrelation between the different elements is shown in Figure 1. Each element is represented by a block in the diagram. The block for the controls element interfaces with IPR. It is placed at the interface to signify that its purpose is to protect the attributes and methods from IPR. The attributes block is contained within the methods block to indicate that attributes should only be accessible via methods. Although there is only a single relationship element in the model, there are two blocks for relationships in Figure 1. This is to make provision for input and output relationships between multiple outside entities. Remember that outside entities may be other departments within the same organisation, or they may be other organisations or individuals.

The full specification of the model is shown in detail in Figure 2. The figure is based on a Unified Modelling Language (UML) class diagram. It differs because the 'operations' section of a UML class diagram is called 'methods' in our model. It also has additional sections for controls and relationships. To use the model in a department, one would define specifications for the attributes, methods, controls and relationship in the department. The specifications must follow the format dictated in Figure 2. A strict format is used to allow for easy parsing for implementation on a computer system. In the model specification in Figure 2, the '|' symbol is used to denote the Boolean OR function and the '+' symbol is used to denote the Boolean AND function.

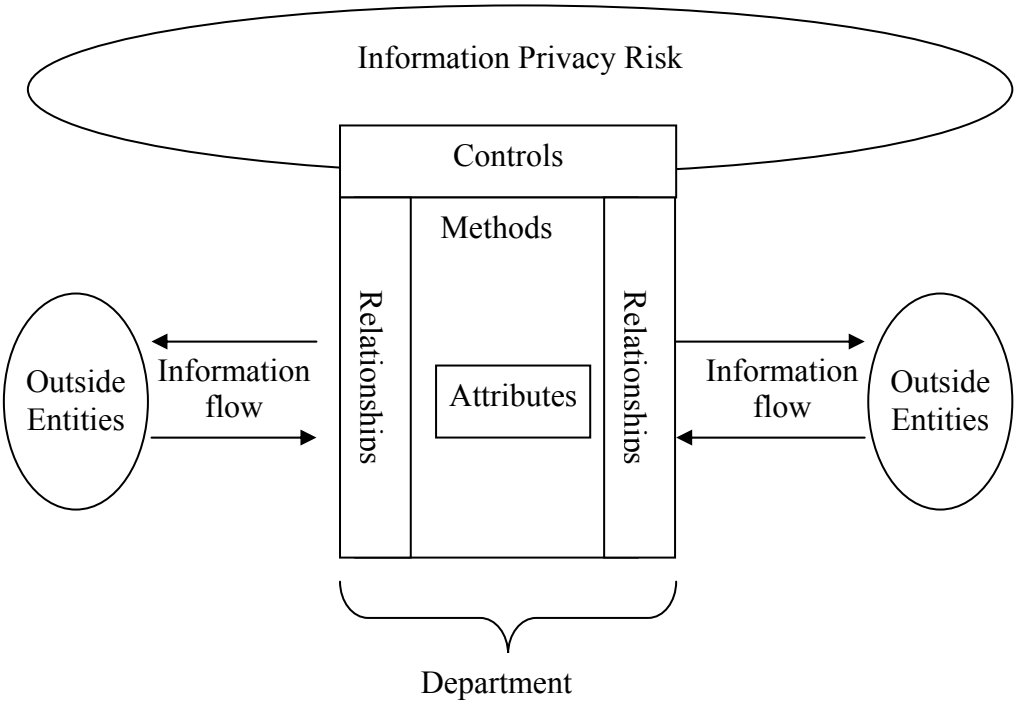


Figure 1 – Interrelation between the Elements in the Model

Figure 2 – Specification of the Conceptual Model for a specific department

Attributes
All private information held by the department is listed here in the form <attribute name_[E|M|A] : *description*> where:
E denotes an electronic copy of the information
M denotes a manual/hardcopy copy of the information
A denotes an abstract entity type to which information pertains, e.g. employee
Methods
All methods for inputting, processing and outputting private information are listed here in the form < [I|O|P]_method name_attribute name_[E|M] : *description* > where:
I denotes an input method or means used to receive private information
O denotes an output method or means used to disseminate private information
P denotes a processing method or means used to process or store private information
E denotes an electronic method
M denotes a manual method
Relationships
All relationships with outside parties are listed here in the form < [I|O]_entity name_[F|T]_{comma delimited list of attributes and/or attribute primitives} : *description* > or starting with [I+O], where :
I denotes an input relationship where private information is received from the named entity
O denotes an output relationship where private information is disseminated to the entity
F denotes another department within the organisation
T denotes a third party (another organisation or individual)
Use of the '+'operator indicates that an input and output relationship exists with the entity.
Controls
All IPCs used for each method above are listed here as < [O|CL|PS|UL|DQ|IP|S|A]_control name - method name : *description*> or starting with [O+CL+PS+UL+DQ+IP+SA], where the letters O,CL,PS,UL,DQ,IP,S,A correspond to the FIP the control is addressing.
Use of the AND operator '+' indicates that more than one principle is being addressed by the control. The FIPs are: O – Openness, CL – Collection Limitation, PS – Purpose Specification, UL – Use Limitation, DQ – Data Quality, IP – Individual Participation, S – Security, A – Accountability
Input methods must have controls that ensure the following: CL, PS, DQ, {S, A,O}*
Output methods must have controls that ensure the following: UL, DQ, IP, S, A, {O}*
Processing methods must have controls that ensure the following: UL, DQ, IP, S, A, {O}*

* FIPs in curly brackets are optional since it may not always be practical to consider them for each control

5 EXAMPLE SCENARIO

In this section we provide a practical scenario which makes use of our conceptual model. We assume the existence of a privacy management system based on our model. We make this assumption because our model is conceptual – in order to be implemented practically, a means is required to record the specifications of attributes, methods, controls and relationships.

In this scenario we consider a job applicant, Bob. Bob wishes to work at company X. In order to work at company X Bob must undergo a psychometric test[1]. The test is performed by a psychologist who requires Bob's permission to give the results to company X for the sole purpose of his job application. Bob grants his permission by signing a permission form, which is faxed by company X's human resources (HR) department. He then undergoes the test, which is performed at the psychologist's rooms. The results are emailed to the company X's HR department. The results reveal that Bob has a personality type that is easily stressed. Since Bob's job does not involve a high degree of stress he is given the job.

As an employee of company X Bob is eligible for medical aid or health insurance. It is the policy of company X, and part of Bob's employment contract, that company X pay a fixed amount towards his health insurance. It is also part of Bob's employment contract that he must insure his health through company H. This is due to the fact that company X has negotiated preferential rates with company H. Bob duly applies for health insurance from company H. In evaluating Bob's application, company H requests the results of Bob's psychometric test, since they know it is company X's policy to have psychometric tests performed on job applicants. The request is made directly to the HR department without Bob's knowledge. From the results of the psychometric test company H discovers Bob has a personality type that is easily stressed. Accordingly, they increase the premiums he must pay for his health insurance. They do this because they argue that a person who is stressed easily is more susceptible to stress-related illnesses. Bob sees that his insurance premiums are more than the standard rate and enquires about the reason. Company H

[1] This is a test based on psychometric theory. Such tests are usually designed to determine personality characteristics, aptitude, intelligence, and other psychological traits.

informs him of the reason. Bob then asks company H how they acquired the information about him. Company H notifies him and he sues his employer for breaching his privacy. Specifically, for using his information for a purpose he had not agreed to.

We now show how the model could have been used by company X to avoid such a situation. Due to space restrictions we do not define all attributes, methods, controls and relationships. We only include those necessary to protect psychometric test results and those relevant to providing a better understanding of the model. All definitions are from the point of view of company X's HR department.

We start by defining attributes for the psychometric test results (note that numbering attribute definitions is not required by the model but we do this for referencing purposes):

(1) Applicant PsychTestPermForm_M : *Manual document containing applicant's permission to use psychometric test results for health insurance*

(2) Applicant PsychTest_E : *E-mail copy of a job applicant's psychometric test results*

(3) Applicant PsychTest_M : *Manual document containing applicant's psychometric test results*

We define two attributes (2 and 3) for the test results since the test results are sometimes printed and stored in a manual file. In (2) we see the definition for the email received from the psychologist and in (3) we see the definition for the manual printout. Note the '_E' and '_M', as well as the descriptions, are used to differentiate the two. In (1) we also define the form that an applicant must sign to grant company X permission to give the results of the test to company H.

We now specify the methods related to these attributes:

(4) I_Receive Applicant PsychTest_E_E : *Receive applicant psychometric test results by e-mail*

(5) P_Print and Store Applicant PsychTest_E_M : *Print and store applicant psychometric test results in manual file*

(6) O_Send PsychTest_E : *E-mail applicant permission form to company H*

In (4) we define an input method for the receipt of the test result email defined in (2). The '_E' again specifies that this method is electronic. In (4) the processing method for printing out the email from (3) is defined. This results in the creation of the manual test results defined in (5). The fact that the email is stored in a manual file is noted by the '_M' in (4). The output method for e-mailing the permission form is also defined in (6).

We now define a control related to these methods:

(8) UL_Signed Applicant PsychTestPermForm_M - O_Send PsychTest_E: *Have applicant sign permission form for the purpose of giving test results to company H.*

The single control in (8) protects the Use Limitation principle in the OECD guidelines. This can be seen by the 'UL_' at the beginning of the definition. The Use Limitation principle states that personal information should not be made available for uses other than those specified at the time of collection. In our scenario we recall that Bob agreed to provide company X with the results of his psychometric test for the sole purpose of his job application. The control defined in (8) states that a signed permission form is required before Bob's test results may be sent via email as defined in (6). This control is sufficient to limit the risk of company X giving company H Bob's test results without his consent. If Bob does not want company H to have the results of his test, he need not sign the permission form. Thus, the likelihood of a situation such as the one in our scenario is significantly diminished.

Additional controls may also be defined to further protect Bob's psychometric test results. For example:

(9) S_Lock Applicant Files Cabinet_M - P_Print and Store Applicant PsychTest_E_M : *Lock manual files used to store applicant test results in a filing cabinet*

This control protects the Security principle in the OECD guidelines. The Security principle states that personal information should be protected by reasonable safeguards to against its loss, unauthorised access, destruction, use, modification or disclosure [6].

The relationship between company X's HR department and company H, which is the subject of our scenario can, be defined by the following:

(10) O_Company H_T_{Applicant PsychTest_E} : *E-mailing of applicant psychometric test results to company H*

The 'O_' at the beginning of the definition indicates that it is an output relationship. In other words, information is being disseminated from the HR department. 'Company H' is the name of the entity the relationship is with. The 'T_' specifies that the relationship is with a third party, that is, with another organisation or individual. 'Applicant PsychTest_E' is the name of the attribute being disseminated in the output relationship – see (2) for the definition of this attribute. The HR department's name is not included in this definition. This is because all the definitions in our scenario up to this point are from the HR department. All relationships are therefore defined with respect to the HR department. A corresponding relationship definition from the appropriate department in company H will look like this:

(11) I_Company X HR Dept_T_{Applicant PsychTest_E}

The only difference in this case is that (11) is an input relationship since the test results are received from company X's HR department.

6 DISCUSSION

In this section we undertake a general discussion of the conceptual model. We provide further rationale for our choice of the elements that make up the model, namely attributes, methods, controls and relationships. Furthermore, we discuss some potential uses for the model.

In order to use the model, an organisation must define the attributes, methods, relationships and controls as required by the model. Ideally, this must be done for each department that deals with private information. As mentioned earlier, the definition of attributes and methods is the first step to protecting information privacy. This is because organisations cannot protect information if they do not know it exists, or, if they do not know where or how it is stored and used. The model is holistic in that attributes may be either electronic or manual. This is important since private information exists in both forms in large organisations.

Defining attributes and methods only maps out what needs to be protected to reduce IPR. It does not specify the means by which protection will be achieved. The purpose of the controls element of the model is to specify such means. It does this by ensuring that organisations have

controls in place to protect each of the FIPs. Since the model is a high-level, conceptual model, it does not dictate what these controls should be. The choice of control is left up to the organisation (e.g. role-based access control or policy based controls such as Karjoth and Schunter [10] may be used as technical controls). Knowledge of the FIPs is thus required in order to implement the model. We do not believe this is a problem for two reasons. Firstly, the FIPs are freely available [6]. Secondly, we believe the FIPs are sufficiently straightforward to understand, especially given the expertise available in large organisations. While the model does not dictate the use of specific controls, it does show which FIPs controls should protect for input, processing and output methods. Organisations are thus able to determine if controls are missing for the various aspects of information privacy defined in the FIPs. This is important because where FIPs are not protected, this results in increased IPR.

Information flow refers to the movement of attributes from one entity to another. It is an important aspect of information privacy. Inappropriate information flows can result in breaches of information privacy [22]. The control of information flows is therefore important in reducing IPR. In our model information can flow in and out of a department only via input and output methods. As discussed, controls protect the privacy of information 'flowing' through these methods. The relationships element of our model explicitly defines the relationships between a department and outside entities. This is done because the 'level of granularity' of our model is the department. That is, our model describes only a single department at a time. The relationships element thus provides a mechanism to link multiple departments by the flow of attributes between them. In other words, it allows an organisation to map inter-departmental flows of private information. It also allows organisations to map the information flows between themselves and other organisations and individuals.

It is important to note that the effectiveness of the model is dependant on accurate and complete information regarding attributes, methods, controls, and relationships. Regular updating of the model is therefore necessary because it is possible that methods, attributes, controls and relationships will change over time. The likelihood of such changes should determine the frequency with which the model is updated.

It is also important to note that the model is a high-level, conceptual management model. Its purpose is to provide guidance about the management of IPR and not to dictate specific controls or methods. Due to the structure of the model, it is easy to record the specifications for attributes, methods, controls and relationships electronically. This may be done using object or relational databases. Once recorded, application systems can be designed to interface with the databases for the purpose of managing and maintaining the model, for example, to add new controls or methods. An application system may also enforce the model's rules. An example of this would include warning a user that controls enforcing certain FIPs are missing with respect to a given method. Information flows can easily be determined with an application system by interrogating the relationships element of each department in the database. From this it will be possible to construct visual maps of information flows. An application system based on the model would, in essence, be a privacy management application. As such, it would primarily be of use to those responsible for the management of IPR. In large organisations, this may be a chief privacy officer, chief information officer or the internal audit head. Due to the rigorous specification of methods and controls, the model may also be used as the basis for privacy audit systems.

A standardisation of the model may allow for a uniform way of representing the private information in an organisation, the processing and protection thereof, as well the information flow both within and between organisations. A standard means to represent private information and its processing, protection and flow is useful in the privacy audit domain. This is because a privacy audit of an organisation will require knowledge of the private information in an organisation as well the controls employed to protect information privacy. In certain countries larger scale audits, or investigations, are carried out by privacy commissions [2]. In these countries privacy commissions are usually statutory bodies charged with protecting information privacy. If a standard means exists to represent private information and its processing, protection and flow, this will make investigations by commissions easier. The reason for this is that commissions can use applications to automatically interrogate information made available from the privacy management systems of large organisations.

7 CONCLUSION

In this paper we have addressed the need for large organisations to manage information privacy appropriately. We have done so by presenting a high-level, conceptual model for the management of IPR. The model is based on the similarity between departments in organisations and objects in object-oriented programming languages. We have provided a detailed specification of the model and discussed the various elements it is comprised of. In order to demonstrate its practical significance, we have also presented a scenario in which the model is used.

Future work on the model may include adding a 'personnel' element to the model. This element will explicitly define the roles and responsibilities of individuals within a department with regard to preventing IPR. Research will be required to determine how this element will link with the original elements in the model and what, if any, modifications to the original elements are necessary.

Further research is also needed in the practical implementation of the model. Such research may be achieved through a case study in which the model is applied to departments in a real organisation. To fully understand the potential benefits and pitfalls of a practical implementation in a real organisation, it will be necessary to develop and deploy a prototype privacy management application system based on the model.

8 REFERENCES

1. A. Daniel Oliver-Lalana, “Consent as a Threat. A Critical Approach to Privacy Negotiation in e-Commerce Practices”, In: *TrustBus 2004*, *LNCS,* Vol. 3184, S. Katsikas., J. Lopez, G. Pernel (eds.), pp. 110-119, Springer, Heidelberg, 2004

2. *Discussion Paper 109, Project 124, Privacy and Data Protection*, South African Law Reform Commission, Pretoria, 2005

3. R. Clarke, “Introduction to Dataveillance and Information Privacy, and Definitions of Terms”, Department of Computer Science, Australian National University, 2006. Available: http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html

4. A. Ghobadian and D. Gallear, "TQM and organization size", *International Journal of Operations & Production Management*, Vol. 17, No. 2, pp.121-63, 1997

5. S.E. Chang and C.B. Ho, "Organizational factors to the effectiveness of implementing information security management", *Industrial Management & Data Systems*, Vol. 106, No. 3, pp. 345-361, 2006

6. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, Organization for Economic Cooperation and Development, Paris, France, 1980. Available at: http://www.oecd.org/document/57/0,3343,en_2649_201185_1815186_1_1_1_1,00.html

7. C.S. Powers, P. Ashley, M. Schunter, "Privacy Promises, Access Control, Privacy Management", In: *Proceedings of the 3rd International Symposium on Electronic Commerce*, North Carolina, USA, 2002.

8. R. Gellman, "Does Privacy Law Work?", In: *Technology and Privacy: The New Landscape*, P.B. Agre and M Rotenberg (eds), pp. 194, The MIT Press, 1998

9. R. Purcell, "Privacy by 3PT®: A Management Model", *Corporate Privacy Group*, Nordland, Washington, USA. Available: http://www.corporateprivacygroup.com/CPG_3PTMGMTMODEL.pdf

10. G. Karjoth and M. Schunter, "A Privacy Policy Model for Enterprises", In: *Proceedings of the 15th IEEE workshop on Computer Security Foundations*, Nova Scotia, Canada, pp. 271, 2002

11. P. Ashley, S. Hada, G. Karjoth, C. Powers, M. Schunter, "Enterprise Privacy Authorization Language (EPAL 1.2)", *International Business Machines Corporation*, 2003. Available: http://www.zurich.ibm.com/security/enterprise-privacy/epal/Specification/

12. M. Casassa Mont, "Dealing with Privacy Obligations: Important Aspects and Technical Approaches", In: *TrustBus 2004*, *LNCS,* Vol. 3184, S. Katsikas., J. Lopez, G. Pernel (eds.), pp. 120-131, Springer, Heidelberg, 2004

13. M. Casassa Mont, "Towards Scalable Management of Privacy Obligations in Enterprises", In: *TrustBus 2006*, *LNCS,* Vol. 4083, S. Fischer-Hübner et al. (eds.), pp. 1-10, Springer, Heidelberg, 2006

14. J. Biskup and H.H. Brüggemann, "The Personal Model of Data: Towards a Privacy-Oriented Information System", *Computers & Security*, Vol. 7, No. 6, pp. 575-597, 1988

15. J. Biskup and H.H. Brüggemann, "The Personal Model of Data Towards a Privacy-Oriented Information System", In: *Proceedings of the Fifth International Conference on Data Engineering*, California, USA, 1989

16. D.J. Armstrong, "The Quarks of Object-Oriented Development", *Communications of the ACM*, Vol. 49, No. 2, 2006

17. L.F. Capretz, "A Brief History of the Object-Oriented Approach", *ACM SIGSOFT Software Engineering Notes*, Vol. 28, No. 2, 2003

18. M.M. Eloff and S.H. von Solms, "Information Security Management: A Hierarchical Framework for Various Approaches", *Computers & Security*, Vol. 19, No. 3, pp. 243-256, 2000

19. *COBIT 4.0*, IT Governance Institute, pp. 14, Rolling Meadows, Illinois, 2005. Available: http://www.itgi.org/template_ITGI.cfm?template=/ContentManagement/ContentDisplay.cfm&ContentID=27263

20. *Privacy Online: Fair Information Practices in the Electronic Marketplace – A Report to Congress*, Federal Trade Commission, pp. 3-4, Washington D.C, USA, 2000. Available at http://www.ftc.gov/reports/privacy2000/ privacy2000.pdf

21. D.J. Solove, "A Taxonomy of Privacy", *University of Pennsylvania Law Review*, Vol. 154, No. 3, 2006

22. L.C.J. Dreyer and M.S. Olivier, "An information flow model for privacy (InfoPriv)", In: *Database Security XII: Status and Prospects*, S. Jajodia (ed), pp. 77–90, Kluwer Academic Publishers, 1999

BPMN AS A BASE FOR CALCULATING THE TARGET VALUE OF EMPLOYEES' SECURITY LEVEL

Jan Schlüter[1], Stephanie Teufel[2]

[1,2]**iimt, University of Fribourg, Switzerland**

[1]**jan.schlueter@unifr.ch,** [2]**stephanie.teufel@unifr.ch**

ABSTRACT

In this paper we aim to create a new model which uses the Business Process Modelling Notation (BPMN) as the base for calculating the target value of the employees' security level. It has to be assumed that all processes, at least those which interact with the information technology directly, are written down in BPMN or a fully-convertible notation respectively describing language. It is important that bigger companies and public authorities do fulfil this requirement.
The problem is that it is hard to get an overview about the information system access and privileges of each employee in bigger companies. The approach currently used in "secaliser", our initial project, expects that the information technology affinity is based on each employee's job position. In many real life situations this condition cannot be fulfilled. Due to the fact that we try to optimise the trainings in an economic way, there is only a small range between the necessary and the optimised security level. This is the reason why it is so important to enhance the exactness of our calculation. The goal of the described model is to make conclusions concerning the specific should-be security level of each employee, based on comprehensible data, which is extracted from the company processes.

KEY WORDS

BPMN, security level analysis, employee, process analysis

BPMN AS A BASE FOR CALCULATING THE TARGET VALUE OF EMPLOYEES' SECURITY LEVEL

1 INTRODUCTION

Nowadays, business becomes more and more complex. The same applies to the structure of the information systems, because they try to copy a model of the complex business. Of course, many of these complexity problems do not affect the information security directly, though there are some security relevant factors which increase disproportionately high with growing business complexity. [8]

In the past, there were many technical ways on how to protect information from being accessible, to be changed, or to be taken away, but those systems do not help us to go the first step: finding out who needs which permissions and how to structure the way the access permissions are allocated. Regarding a suggestive restructuring of business processes in order not to give away the same permissions to a larger than necessary group of employees, is an absolutely important step that is completely untended. [2]

It must not be disregarded that there are different kinds of information to manage. Some information is stored in file systems and the access to the different shares is limited to authorised users. Due to the fact that those mechanisms are implemented in most common operating systems, they are well-known – as well by end users as system administrators – and caused by the simple permission structure these kinds of information can be easily managed, and the access permissions can be clearly arranged. Much more complex in managing the access are other systems which have, in most cases, not such a clear structured base, this also afflicts plain inheritances like normal folder structures do. Although relational storage models can also be managed in a descriptive way according to tables – even if the access permissions can be allocated here much more sophisticatedly – role based access models are not as widespread as in file systems. However, there are other platforms which cannot be managed as well in a descriptive way as relational and folder based structures, for example, hardly adaptable and closed third party software. Beside the fact that the company cannot guarantee the permission system inside the software, it could be hard to find out the kind of information each employee has contact with. Due to the fact that big-

ger software systems often directly use the permission system of a relational database management system, this problem affects smaller software products much more than the bigger ones. Therefore many niche software products are affected, but these software programs may contain the most critical data for the business. [2]

2 PROCEEDING

Due to the fact that not all information is security sensitive on the same level, it becomes important to group the different kinds of information with same security levels and rate those groups. These information groups are the initial point for our considerations: In an internal feasibility study, we checked out different ways on how to get information of the employee's specific security affinity, and the security level we should use for the optimisation. An approach which uses the Business Process Modelling Notation (BPMN) to check all data the employee has contact with, is able to change, or delete, is the most promising one. In addition to the comparatively simple as-is state analysis, this approach gives the possibility to monitor changes in the different processes and to react on those changes contemporarily. Due to the power of BPMN, especially the artefact-components, there are several ways to bind processes, persons, and data-permissions with each other, without breaching the current working draft from 3rd May 2004. [5, 3]

2.1 Grouping information

To get an overview of the permissions an information (or data) directly or indirectly implicates, it is important to collect all available information. The groups which should be created, as described before, have to be divided into two different groups, namely: *Security Sensitive* and *Data Equivalency/Implication*.

Security Sensitive

It makes sense to create groups and directly link them to a security sensitive level. This level is a simple number and represents the security importance of the grouped information. Furthermore, due to gaps between the security sensitive levels of the different groups, it may be necessary to devide critical and non-critical information. Anyhow, it will be hard to scale the groups

in a fair way, because the different information might be very hard to compare with each other, but this problem has to be solved by the company themselves.

Data Equivalence/Implication

That there are different information in the same group does not declare the information as equivalent, because completely different information which may belong together or not, can be on the same security level. For example, *street* and *zip code* as part of an address data set, or two non-correlating information like a *social security number* and an *image*. The reason why we do not only use equivalence groups and link them – of course not unique – to the security sensitivity level is self-explanatory when trying to build up the first business process model: Indeed, equivalent information or information which implicate other information will normally be in the same group, but also, in this opposite reflection, the statement is not universally valid. In our own tests, we exposed that it might be useful to adapt the view in some cases and to divide own personal information from the personal information of a third party.

To mark information as privately accessible is particularly suitable for every kind of identification of the person which is not done automatically, for example, at the cashier's desk where the banker does not ask you for any information of your banking account because they personally knows the (manually) identified person. To describe these social problems with business process models only would be very weak in practice.

As not to breach the BPMN standard, we decided to do the distinction between the two kinds of personal information very simply and just appended a wildcard to the information name which is explicit, not being accessed by any other than the belonging person. This simple approach does only influence the naming convention and not the standard itself.

2.2 Grouping employees

In order to make it possible to point out the employees who do have more permissions than other employees in the same position, it is useful to also group those persons who can be compared with each other. A comparison may be useful with either employees in the same job position or with those who are located at the same place. The employee groups will only be used

to display the result better and to highlight some irregularities. Neither the calculation nor the results are affected by these groups.

3 CREATE A BUSINESS PROCESS MODEL

As written in section 1, the business process model should be created using the BPMN standard. When analysing the different business process model standards, it came out that the functional range of BPMN is much wider than others like the event driven process chains (EPC). Due to the fact that the BPMN standard is administered by the Object Management Group since 2005, and due to their experiences with the Unified Modelling Language (UML), it can be expected that the BPMN functional range for information systems can be increased in the near future. [7, 6, 5]

Normally, bigger companies and public authorities already have at least some of their business processes written down. Of course it will not make sense to create hundreds of processes again that are already written down, but due to the complexity of the BPMN standard it is assumed that it is possible to import other business process formats into BPMN with no or a very small information loss. Common used process description standards like EPC as part of the ARIS Framework do have such a little modelling complexity that the transformation can be done without any information loss (all elements of EPCs are also part of the BPMN standard). Therefore it is only required to add additional information to the already existing processes. [10, 4] Of course it is not realistic for every branch that all processes are already written down, but especially for those processes which concern security sensitive information the assumption is reasonable.

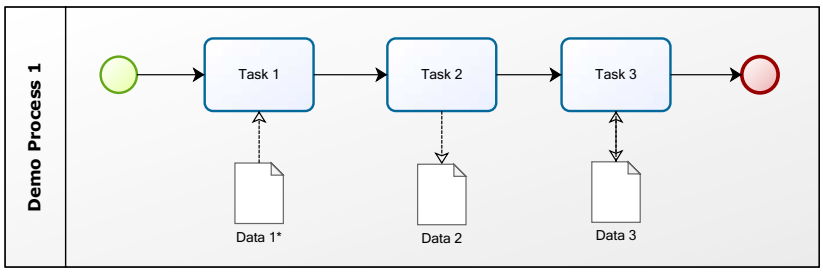


Figure 1: Demo Process 1

Figure 1 shows a demo process created in BPMN containing three *Tasks* and three *Data Objects* in one *Pool*. Except for the wildcard appended to *Task 1*, this is a sample business process which can be found in companies (for example in most applications that manage data permissions like CMS systems), although many companies are using simple EPCs until today[1]. Analysing this process could proceed straight forward, because there are no conditions or anything else to combine with. Even the three Data Objects are not connected through any other *Task* or *Message Flow*. As described in 2.1, the wildcard is normally used in *User Tasks* only, which are manually performed. Inexactnesses like this, which do not need to be caused by the company itself, but maybe by the use of weak tools or export mechanisms will make it difficult to get an overview about the business processes. Therefore it is absolutely necessary that after importing into or creating business processes, the complete process has to be rechecked for a consistent and correct use of BPMN elements and our adapted naming convention. The following description gives a short summary about what happens in *Demo Process 1* for those who are not familiar with BPMN:

- The *Process Sequence Flow* starts, initiates three *Tasks* and ends.
- *Task 1* seems to be a *User Task*, because only the owner of the data (remember the wildcard) is able to read (incoming arrow) the *Data Object*.
- *Task 2* is a *Task* which directly writes (outgoing arrow) into the *Data Object*.
- *Task 3* is a *Task* which reads and writes (incoming and outgoing arrow) the *Data Object*. Typical use of this read and write actions are normally conditional updates or the use in *Sub Processes*[2].

The way how the processes should be created is free to the company, but of course it is useful to use one of the well-established standards for all processes. Due to the fact that the process in detail is much more interesting for the analysis (very general business process parts almost contain no data information), we recommend to use a bottom-up approach. This course of action will ensure that not all data sets are defined from the beginning

[1] against definition, our demo process does neither start nor finish with an event
[2] a non-atomic process

on, but that most of the defined assignments are already meaningful while the business process is not complete. This could lead to a first rough restructuring of the business models while still modelling business processes. [7]

3.1 A sample process for analysing

The problem in Figure 1 was the presumable non exactness of the used BPMN items.

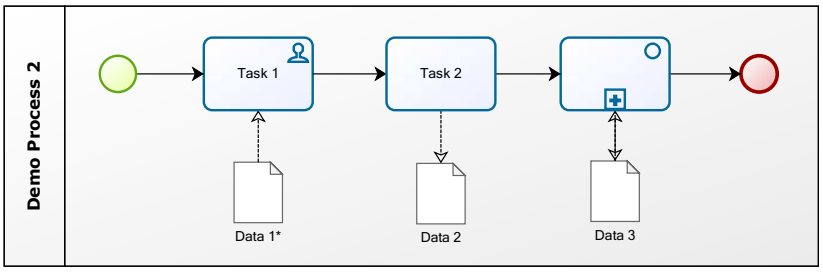


Figure 2: Demo Process 2

The *Demo Process 2* in Figure 2 corrects this problems and now represents a good base for further adaptions. Neither the *Sequence Flow* nor the *Tasks* respective *Sub Processes* have changed their intent until now, or will change it in the following steps.

In the following step we group our *Data Objects*. Figure 3 shows the result of two groups being created:

- *Data 1* and *Data 3* are implicating each other. This means that the *Data Objects* themselves are not equivalent, but that a change of *Data 1* induces a change of *Data 2* and conversely.
- *Data 1* and *Data 2* are assigned to *security level 2*, which means that the *Data Objects* contain security sensitive information.

Note: Even if *Data 1* and *3* are implicating each other this does not mean that those *Data Objects* have to be security sensitive on the same level in general. *Implicating Data Objects* are related to each other and this relation does not need to be connected directly across one object. Also

objects (in context of business model representation), can relate to each other and their attributes may be transitively connected to each other, which by the way may cause the relation between the objects. Comparing this general model with relational database management systems, in combination with data consistency systems as used in enterprise programming, for example Hibernate and Java, will help us to understand the coherences. Due to complexity reasons, our example does not contain any *Message Flows* which would necessarily be building a real business process model, but is uncared for in our analysis. [1]

Considering Figure 3, it will become clear that it will be impossible to either visualise all equivalent and implicating *Data Objects* as one visual group, or to group the different security relevant *Data Objects* to only one unique visual group with the same security level. With only three Data Objects and two Groups, the Diagram has to tend towards to the bottom in order to visualize that *Data 2* is not part of the *Implicating Data Group*. Due to this, there are three different ways:

1. Keep it like it is and only very simple group configurations can be set. This approach would not need any further adaptions, but will lead to a point where you cannot model your business processes in an accurate way, and moreover, the business developing process would take some more time, because doing the layout of the *Data Objects* will become hard.

2. Keep the editor like it is today and just name belonging groups identically. So it could happen that there are multiple groups with the same security level containing different *Data Objects*. This approach would be easy to implement, but cause a loss of the general diagram overview at more complex diagrams.

3. Completely rebuild the editor and integrate different views. In the default mode, every *Data Object* is a member of the different groups, which are listed below each object, and a second view mechanism waits for one special group being selected and arranges all items in a way that they can be displayed as a group. This approach is much more complex and causes to a loose of controlling the business process layout. Also this implementation is much more difficult than the other two options.

Due to the fact that we try to model the business processes of bigger companies, the first option does not fit to our mission. The second, quite

simple to adapt approach should be implemented anyway, because this is the simplest suitable solution and the only one which gives the process creator the possibility to layout everything manually[3]. The third option is the most beneficial one, but doing some tests in automatically arranging and doing the layout of the different groups caused some problems. A problem we cannot currently solve suitably is to arrange groups across different *Pools* and *Lanes*, while we have no problem with *Sub Processes*, which can be, in contrast to the *Pools* and *Lanes*, easily expanded.

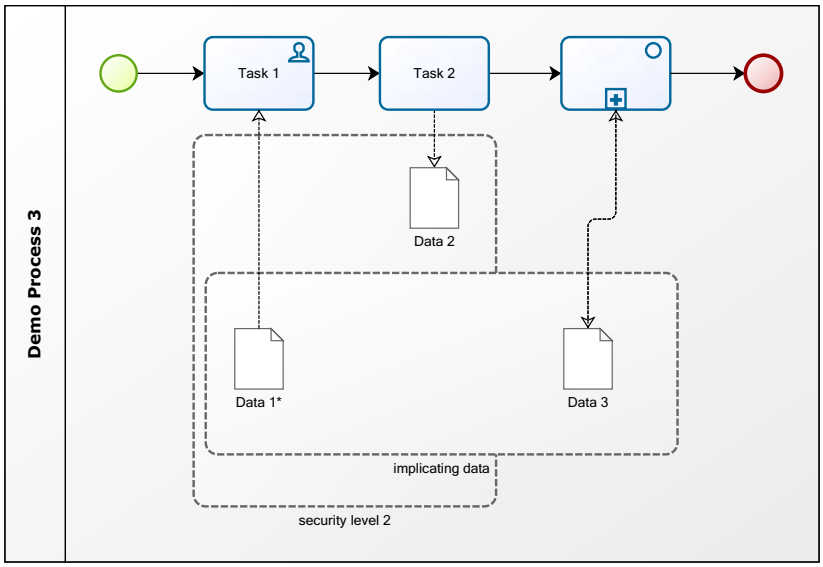


Figure 3: Demo Process 3

Applying a view only for grouping *security level 2* is shown in Figure 4. Of course the example process is very simple and even without focussing on the one group, the process was clearly arranged before, but this would change in bigger processes and it has to be noted that it is not possible to visualise all groups at the same time.

[3]it would almost be impossible to layout every view of option 3 manually, because you have to update everything after just a minor change

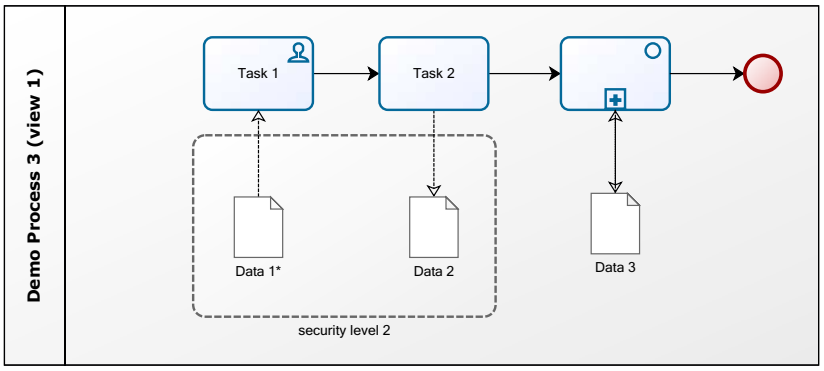


Figure 4: Demo Process 3 (view 1)

3.2 Assigning tasks

BPMN offers the possibility to assign tasks with either *Entities* or to *Roles*, where *Roles* can be either employees, job positions, or other identifying attributes to represent one person or a group of persons. Unfortunately these assignments are hidden attributes of the BPMN standard and will not be printed out, but it is arguable whether additional information in the diagram would make the diagram really more informative or just confusing. However, every *Task* can be linked to a *Role* and due to this we get a direct link from the employee to the necessary user permissions.

Like managing file permissions, it is recommended not to use a certain employee to assign user permissions, but to use user groups[4]. Comparing the employees who are members in the different groups is from the technical point of view very simple, but allows the management to find out which employee has more or less user permissions in comparison to other employees in the same job position.

4 SECURITY LEVEL

The outcome of the described approach will give us information about each employee's security level. This security level is distinct for every employee, and is not mandatorily based on job positions or any other common infor-

[4]these groups have nothing to do with those in 3.1

mation which is currently be used in companies. Especially when planning trainings for employees, it is very important that the distribution of the available resources makes sense. Dedicating too few resources for certain employees could result in security leaks, too many resources could waste resources which will in common result in too few resources for the more important employees. To find the right balance is very hard. [9]

4.1 Procedure

Before starting, we remember our two groups from 2.1 namely *Security Sensitivity* and *Data Equivalence/Implication*. Because it does not make any sense to factor multiple equivalent or implicating security sensitive data objects twice, we will have to find the highest rated security sensitive information which is accessible by each employee per *Data Equivalent/Implicating* group.

Table 1 will give an example were the *Level* column describes a *security sensitivity* group and *Equivalence* a *data equivalence/implicating* group.

Level 1	Level 2	Level 3	Equivalence 1	Equivalence 2
a	b	c	a	d
d	e	f	c	b
g	h	i	e	
j	k	l		

Table 1: Group Table

First of all we need to mention that every task (a to l) is accessible by the sample employee. Second, we see that we can delete every task in our table which has a more security sensitive task in the same equivalence group. Having a look at the highest security level of each equivalency group in Table 1 will show that the tasks a, c and d are redundant because each of those have a more security sensitive task in their equivalent group. The result of this simplification is shown in Table 2.

As shown in Table 2 there are nine security sensitive tasks left. Those tasks can be accumulated which result in a target value of security level of 19.

Level 1	Level 2	Level 3
~~a~~	b	~~c~~
~~d~~	e	f
g	h	i
j	k	l

Table 2: Security Table

5 PROBLEMS

It will be hard for the company to cover the complete business with well defined business processes. The initial effort is very high and increases with the complexity of the business, or in other words: with the number of items in the business process.

Furthermore, even this approach will lose clarity when there are too many groups, or the employees are too far-scattered into these groups so that there are no patterns to identify manually. However, the final step, namely the analysis and evaluation of the results, has to be done manually.

In addition to the described features, it would be nice to help analyse the results and to recommend ways on how to change the business processes to increase security, but this feature can only be developed after having some companies which have produced sample process data which can be analysed manually. These general ideas of improvement have to be transferred into automatic algorithms.

6 CONCLUSION

Business Process Models as being used in modern companies can be used for much more than only to display and analyse workflows or for the accreditation of the business model.

First of all, by means of the company's business processes, it can be detected automatically which user permissions an employee needs to have to do this work correctly. This approach works across the boundaries of isolated information systems and gives a detailed overview over all accessible information inside the company. Worthy of mention is that the model uses the real company view and not a model being distorted by information permission systems. However, transferring the user permission information automatically into the individual permission systems should be possible with most

information systems. The other systems could be set up manually whenever the access to the corresponding information changes. Of course this approach does not help directly to improve the permission systems of the information systems, but it shows up which different roles should be available with which specific user permissions. Furthermore, the problems of the companies current permission systems will come, out and it is possible to point out where the detailed problems in the allocation of user permissions occur and which software has to be adapted to make it possible to protect information that is not necessary to access for certain groups, but security sensitive.

In a second step, we can use this detailed information to get information about the target value of employees' security level. As described in [9], it is absolutely necessary to have information about each employee's security affinity in order to plan company trainings correctly and to think about restructuring either certain parts of the business processes or of the company's organisation structure.

In bigger companies it is not that there is no information which concerns the security level of an employee, but that this information is included in business processes which are at present insufficiently used.

References

[1] Bauer, C., and King, G. *Java Persistence with Hibernate*, revised ed. Manning Publications, 2006. ISBN 978-1932394887.

[2] Benantar, M. *Access Control Systems: Security, Identity Management and Trust Models*, 1st ed. Springer, 2005. ISBN 978-0387004457.

[3] Jeston, J., and Nelis, J. *Business Process Management, Second Edition: Practical Guidelines to Successful Implementations*, 2nd ed. Butterworth-Heinemann, 2008. ISBN 978-0750686563.

[4] Madison, D. *Process Mapping, Process Improvement and Process Management*, 1st ed. Paton Press, 2005. ISBN 978-1932828047.

[5] Object Management Group. `http://www.omg.org/spec/BPMN/1.1`. Website, 2008. last visited: 25.4.2008.

[6] Object Management Group. `http://www.omg.org/spec/UML/2.1.2`. Website, 2008. last visited: 25.4.2008.

[7] Scheer, A.-W., Kruppke, H., Jost, W., and Kindermann, H., Eds. *Agilität durch ARIS Geschäftsprozessmanagement: Jahrbuch Business Process Excellence 2006/2007*, 1st ed. Springer, 2006. ISBN 978-3540333586.

[8] Schlüter, J., Novy, B., Teufel, S., and Marx-Gomez, J. Automatisierte erstellung von wissensbilanzen. In *Multikonferenz Wirtschaftsinformatik 2008* (2008), M. Bichler, T. Hess, H. Krcmar, U. Lechner, F. Matthes, A. Picot, B. Speitkamp, and P. Wolf, Eds., GITO-Verlag. ISBN 978-3-940019-34-9.

[9] Schlüter, J., and Teufel, S. secalyser - a system to plan training for employees. In *Proceedings of the International Symposium on Human Aspects of Information Security & Assurance (HAISA 2008)* (2008), S. Furnell and N. Clarke, Eds., vol. 2. to be published.

[10] Weske, M. *Business Process Management: Concepts, Languages, Architectures*, 1st ed. Springer, 2007. ISBN 978-3540735212.

TOWARDS A CONTEXT-AWARE ACCESS CONTROL FRAMEWORK IN WEB SERVICE TRANSACTIONS

Carina K Wangwe, Mariki M Eloff, Lucas M Venter

University of South Africa

carina.wangwe@gmail.com, +255 754 600512, Box 60049 Dar es Salaam
eloffmm@unisa.ac.za, +27 12 4296330, Box 392 UNISA 0003 SA
ventelm@unisa.ac.za, +27 12 4296330, Box 392 UNISA 0003 SA

ABSTRACT

Interoperability across heterogeneous domains has become a reality through technologies such as Service Oriented Architectures and Web Services. These technologies have been put to use in e-Government and e-Business, enabling services to transact without human intervention. Such transactions, however, raise security concerns, as a human response to an authorization or access request can take into consideration semantics and the context in which the request is being made, while a machine to machine decision to grant access would rely on how well the XML based security policies have captured all semantic and contextual considerations.

This paper proposes a context-aware access control framework in a web services environment. The framework is based on the Organization for Advancement of Structured Information Standards (OASIS) for web services security and access control and extends these to include semantic interpretation of security attributes. Furthermore, the framework addresses contextual information that would affect an access control decision, in a web service transaction, such as legal or regulatory requirements.

KEY WORDS

Access Control

Proceedings of ISSA 2008

A CONTEXT – AWARE FRAMEWORK FOR ACCESS CONTROL ASSERTIONS IN WEB SERVICE TRANSACTIONS

1 INTRODUCTION

With any collaboration, it is crucial to have unambiguous communications between the collaborators, to ensure that no information is either wrongly withheld or provided based on an ambiguous request.

For Web Service transactions, one way to achieve such communication is the use of a semantic framework to provide a basis for interpretation of access control requests depending on the context of the transaction within a given domain. Furthermore, where laws and regulations exist that govern the transaction, these have to be taken into consideration when applying the access control or authorisation policy. The framework would thus include an access control mechanism, semantic interpretation of access requests, a context service and a repository of relevant laws and regulations.

The Organisation of Advancement of Structured Information Standards (OASIS) has adopted standards such as the Extensible Access Control Markup Language (Oasis 2005a) and the Security Assertion Markup Language (Oasis 2005b) to address access control across heterogeneous domains. The Extensible Access Control Markup Language (XACML) is a policy language which uses XML statements to present access control policies while the Security Assertion Markup Language (SAML) is an XML-based security specification schema for exchanging authentication and authorization information. XACML and SAML both have extensibility mechanisms which allow them to be used for different implementation. Use of these standards alone does not however ensure the

correct access control decisions in interacting web services. There is a need to ensure that those XML tags passed to request access are correctly interpreted in the context of the transaction.

The use of ontologies in web services has been promoted by the World Wide Web Consortium (W3C) which has recommended the Web Ontology Language (OWL) as a general ontology for the semantic web (W3C, 2004). OWL is based on the Resource Description Framework (RDF) schema which was an earlier specification from W3C. The ontology serves the purpose of clearly defining terms that are used in a transaction, and enables a semantic evaluation of terms to determine similar meaning. Specific ontologies based on OWL or RDF have been proposed by Ceravolo (2003), Domingue et.al. (2004), and Dritsas et.al. (2005) for the e-Government domain.

For a specific ontology to be used, the context of the transaction must be taken into consideration. Context defines the conditions that must or must not hold in order for an authorisation policy to apply (McDaniel, 2003). Contextual information may include the location of the requester and the provider of the service or the time when the transaction is taking place. For transactions that are taking place in an E-Government or E-Business environment, the legal context may also be necessary. All contextual information needs to be captured and combined so as to act as input into the access control decision.

This paper presents a framework that comprises of a context service, ontological mapping mechanism and a legal repository which together with extended markup languages, support correct access control decisions in interacting web services. The remainder of the paper is structured as follows: Section 2 describes existing access control models for web services. Section 3 proposes a context –aware framework while section 4 looks at related work in this area and we conclude and look at further work in Section 6.

2 ACCESS CONTROL IN WEB SERVICE TRANSACTIONS

A major requirement of an access control model for web services is the handling of the dynamic nature of the transactions. Web services interact across disparate computing platforms, in different geographical locations and with different regulatory compliance requirements. In subsequent sub

sections, we describe some access control models that have been proposed or implemented for web services.

2.1 Role Based Access Control (RBAC)

RBAC uses roles as a basis for access control decisions and was designed specifically with enterprise organisation structure in mind. RBAC allows the specification of security roles that map naturally to an organisation's authorisation structures. However RBAC does not entirely suit web service transactions and its weakness in open environments were identified by De Capitani di Vimercati and Samarati (2005). Several studies have subsequently been done to extend the RBAC model in order to address some of the weaknesses (Demchenko et.al, 2007).

2.2 Attribute Based Access Control ABAC

In recent years, there has been a shift to looking at attributes as a basis for access control in a web services environment. (Coetzee and Eloff, 2007; Damaini et. al, 2005; Shen and Hong, 2006; Yuan and Tong, 2005). Attributes describe the characteristics of the requester, and may be a combination of identity and role. Attributes may be subject attributes, resource attributes or environment attributes. The ABAC model comprises of an Attribute Authority, Policy Enforcement Point, Policy Decision Point and Policy Authority.

It has been recognized that there is still a need for the usage of semantics and or ontologies to ensure correct access control decisions with the ABAC model, and some research to that end has been done. (Preibe et.al; 2006; Warner et.al, 2007).

2.3 Context Aware Access Control

Both RBAC and ABAC paradigms do provide ways to include contextual information (Bacon et.al, 2002; Huselboch et.al., 2005; Strembeck and Neumann, 2004). However other access control models that focus primarily on context have been proposed. These include:

2.3.1 Governance Based Access Control

The idea as presented by the Centre for Governance Institute (2005) is that transactions in which information is shared must be governed by the relevant legislation to which the organizations sharing the information are accountable. Thus any request for information is checked against the exiting laws or regulations before it is granted.

2.3.2 Session Based Access Control (SBAC)

In session based access control, the context of a transaction is limited to a session. Access to resources is based on the attributes of the subjects and the properties of the objects but the rights that can be applied at a given time are limited based on the context defined by the access session (Fernandez and Pernul, 2006)

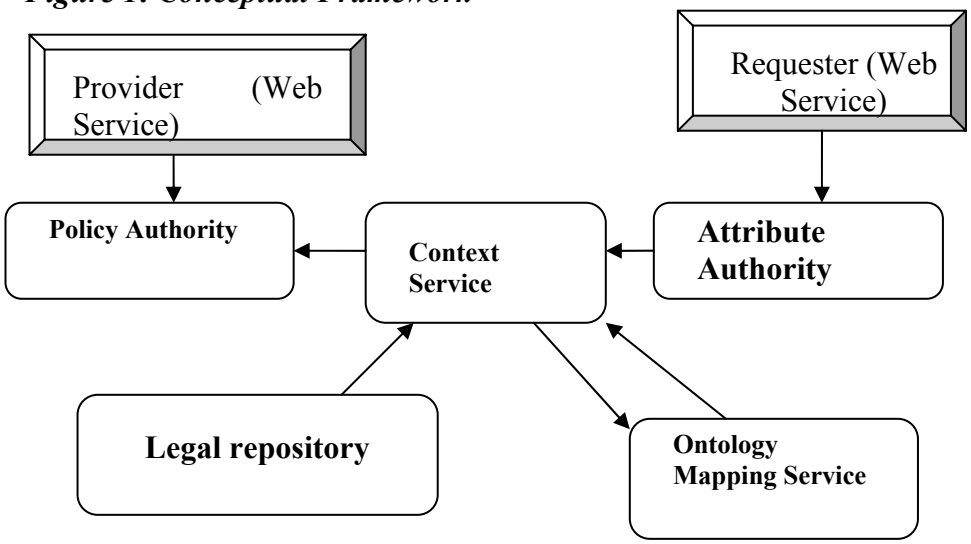
2.3.3. Location-Based Access Control (LBAC)

LBAC takes requester's physical location into account when determining their access privileges. The physical location may be combined with other attributes related to identity or role of the requester. Ardagna et.al (2006) propose combining location with user credentials to support access control decisions.

3 PROPOSED FRAMEWORK

In order to achieve correct access control decisions in the context of a web service transaction, we propose a framework based on the ABAC model. The proposed framework is illustrated in Figure 1 below:

Figure 1: Conceptual Framework



Each of the components of the framework works as follows:

i) Policy Authority

The policy authority contains the Policy Decision Point (PDP) and Policy enforcement points that evaluate the requester's attributes against the providers XACML policy. In order to evaluate the compliance with legal requirements XACML is extended to include a function that accepts environment attributes and compares against relevant laws and regulations within the legal repository. This operation will be stated as a XACML obligation in the Provider's policy. If there is no legal requirement for a particular transaction, then the request is granted provided the other requirements of the policy are met.

ii) Attribute Authority

The attribute authority issues SAML assertions to the requester. The attribute assertions correspond to the subject, resource and environmental attributes of the requester. If there is a legal requirement on the requester's

side that has to be complied with, this requirement is passed in a SAML condition statement.

iii) Ontological mapping service

The ontological mapping services checks the semantics of the requester's attributes match with those in the provider's policy. A mechanism to conduct such a mapping has been described by Patil et.al (2007). If unknown vocabularies are used, ontology mediators may be used (Kolter, et.al, 2007).

iv) Legal repository

The legal repository contains laws and regulations that apply to different transactions. The legal repository contains the conditions in which a transaction is considered legal or illegal. The legal repository is a database which with several indexes to allow multiple matching by the Context Service.

v) Context Service

The context service is a key element of the framework and is adapted from Lei et.al. (2002). The role of the context service is to combine the results from the ontological mapping mechanism and the legal repository into an environmental attribute that is then passed to the attribute authority for authorisation and access control decisions to be made. To illustrate how the framework could be applied, consider the following illustrative example:

A request for information is made in a criminal investigation where a national of Country A is suspected of committing a crime in Country B; and the suspected criminal is now in resident in Country C. In order for the service in Country C to decide whether to authorise access to the information the following requirements must be met:

- The penalty for the crime in Country C must be evaluated against the penalty for the crime in country A. If conviction may result in a death penalty, then Country C must refuse to provide information.
- The crime committed in Country B must be interpreted in the context of the laws of country C.

- Laws of country A must be examined to see if they have any relevance in the crime and or penalty for the crime

Thus for this example the service provider would need access to a legal repository of the countries' laws and also to the ontological mapping mechanism to make semantic comparisons as to whether or not all necessary conditions to grant the requested information hold.

4. RELATED WORK

There are various studies that have been done in relation to context – aware and or semantic – aware authorisation and access control. The studies that are pointed out below are those that address context in access control decisions with some reference to semantics.

Demchenko et al. (2007) use XACML to handle policy and base on RBAC with a Domain Resource Management model. The study argues that domain based access control provides several benefits including dynamic context management. However interpretation of attributes is not addressed by the study. Toninelli et.al (2006) also draw inspiration from the RBAC model and associate the context in which a subject transacts directly with the role that the subject plays in that transaction.

Hu and Weaver (2006) look at the healthcare domain and provide a formal definition of context and context constraints. The definition of context is restricted to time, location, user type, object type and object ID. Context is built into the policy language and WS policy is used for the implementation.

Kolter et al. (2007) describe a semantic aware security architecture which includes an ontological mapping mechanism. The architecture is based on the ABAC model, but does not specifically address how contextual attributes would be handled.

Our work, as presented in Section 3 above, takes into consideration both semantics and contextual information with emphasis on legal requirements.

5. CONCLUSION AND FURTHER WORK

We have presented a framework that comprises of a context service, ontological mapping mechanism and a legal repository which together with extended markup languages support corrects access control decisions in interacting web services. The inclusion of a legal repository make the framework especially useful for e-Government or e-Business transactions that take place across two or more legal domains where different regulations may apply to the transaction. Thus combine with the ontologically mapping mechanism that address semantic interpretation of attributes, the framework lays a basis for correct access control decisions based on the context of the transaction.

Future work shall include formalising a model based on the proposed framework and evaluating the framework in against requirements for access control architectures (Keromytis and Smith, 2007) when the framework is implemented in a practical setting.

6. REFERENCES

Ardagna, C.A., Cremonini, M. & Damiani, E. (2006). *Supporting Location – Based Conditions in Access Control Policies.* Proceedings of ASIACCS'06 held in Taipei. ACM.

Bacon, J., Moody, K. & Yao, W. (2002). A Model of OASIS Role-Based Access Control and Its Support for Active Security. *ACM Transactions on Information Security and Systems Security*, 5(4): 492:540.

Centre for Governance Institute (CGI). (2005). Governance Based Access Control (GBAC): Enabling improved information sharing that meets compliance requirements. Available from http://www.cgi.com/cgi/pdf/cgi_whpr_63_gbac_e.pdf. (Accessed 1 April 2008).

Ceravolo, P. (2003). *Managing identities via interactions between ontologies.* Proceedings of the OTM Workshop held in Catania.

Coetzee, M & Eloff, JHP. (2007). A Trust and Context Aware Access Control Model for Web Service Conversations. *Lecture Notes in Computer Science*, 4657:115:124

Damiani, E., de Capitani di Vimercati, S., & Samarati, P. (2005). *New Paradigms for Access Control in Open Environments*, Proceedings of the fifth IEEE International Symposium on Signal Processing and Information Technology. IEEE.

De Capitani di Vimercati, S. & Samarati, P. (2005). New Directions in Access Control. In Cyberspace *Security and Defense: Research Issues*. Edited by Kowalik, J & A Sachenko, A. Kluwer Academic Publisher.

Demchenko, Y., Gommans, L., & de Laat, C. (2007). Role Based Access Control Model for Distributed Multidomain Applications. In *New Approaches for Security, Privacy and Trust in Complex Environments*. Edited by Venter, H. Eloff, M., Labuschagne, I.., Eloff, J., & von Solms, R. IFIP International Federation for Information Processing.

Domingue, J., Gutierrez, L., Cabral, L., Rowlatt, M., Davies, R., & Galizia, S. (2004,). WP9: Case Study eGovernment D9.3 e-Government Ontology. Available from http://www.dip.deri.org/documents/D9-3-improved-eGovernment.pdf (Accessed 14th March 2008)

Dritsas, S., Gymnopoulos L., Karyda M., Balopoulos, T., Kokolakis, S., Lambriniudakis C., & Gritzalis S. (2005). *Employing Ontologies for the Development of Security Critical Applications: The secure e-poll paradigm*. Proceedings of the International Conference on eBusiness, eCommerce and EGovernment held at Turku. IFIP.

Fernandez & Pernul, (2006) *Patterns for Session Based Access Control.* Proceedings of Pattern Languages of Programming Conference held at Portland.

Keromytis, A.D & Smith J.M. (2007). Requirements for Scalable Access Control and Security Management Architectures. *ACM transactions on Internet Technology (* 7) 2.

Hu, J. & Weaver A.C. (2006) Dynamic , Context – Aware Access Control for Distributed HealthCare Applications . Available at http://www. cs.virginia.edu/papers/p1-hu-dynamic.pdf

Huselbosch, R.J., Salden, A.H., Bargh, M.S., Ebben, P.W.G, & Reitsma, J. (2005) *Context Sensitive Access Control*. Proceedings of SACMAT'05 held in Stockholm. ACM.

Lei, H., Sow, D.M. Davis, J.H., Banavar, G. & Ebling, M.R. (2002). The Design and Applications of a Context Service. *ACM SIGMOBILE Mobile Computing and Communications Review* (6) 4: 45:55.

McDaniel, P. (2003). *On Context in Authorization Policy.* Proceedings of SACMAT 2003 held at Como, Italy. ACM.

OASIS, (2005 a), XACML v2.0 Documentation. Available at http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml (Accessed 15th March, 2008)

OASIS, (2005 b) SAML v2.0 Documentation. Available at http://docs.oasis-open.org/security/saml/v2.0/ (Accessed 15th March 2008).

Patil, V., Mei, A. & and Mancini, L. (2007). *Addressing Interoperability issues in access control models.* Proceedings of ASIACCS'07 held at Singapore. ACM.

Priebe, T., Dobmeier, W. & Kamprath, N (2006), *Supporting Attribute-based Access Control with Ontologies.* Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06) held at Vienna. IEEE Computer Society.

Shen, H & Hong, F (2006). *An Attribute – Based Access Control Model for Web Services*. Proceedings of the Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06) held at Taipei.

Strembeck, M. & Neumann, G, (2004). An Integrated Approach to Engineer and Enforce Context Constraints in RBAC Environments. *ACM Transactions on Information and System Security*, (7) 3: 392:427.

Toninelli, A., Montanari, R., Kagal, L., & Lassila, O. (2006). A Semantic Context – Aware Framework for Secure Collaborations in Pervasive Computing Environments. *Lecture Notes in Computer Science*, (4273): 473-486

Warner, J., Atluri, V., Mukkamala, R., &Vaidya, J. (2007). *Using semantics for automatic enforcement of access control policies among dynamic coalitions*, In Proceedings of SACMAT 2007.

W3C (2004). OWL Web Ontology Language Overview. Available from http://www.w3.org/TR/owl-features (Accessed 2nd May 2008)

Yuan, E. & Tong, J. (2005). *Attribute Based Access Control (ABAC) for Web Services.* In Proceedings of the IEEE International Conference on Web Services (ICWS'05) held at Orlando. IEEE Computer Society.

CONSIDERING CONTRACTS FOR GOVERNANCE IN SERVICE-ORIENTED ARCHITECTURES

Jacqui Chetty[1] and Marijke Coetzee[2]

[1]Department of Business Information Technology
[2]The Academy for Information Technology
University of Johannesburg

[1]jacquic@uj.ac.za
(011) 559 1177
[2]marijkec@uj.ac.za
(011) 559 2907

ABSTRACT
Service Oriented Architecture (SOA) is a design paradigm that enables applications to be built from business processes. Services, service-orientation and related technology give organisations the ability to gain a competitive edge. This however, does not come without cost, due to the fact that organisations develop services quickly, very often without much thought to their management and maintenance. SOA governance is considered a subset of IT governance, to control the design and execution of services. Governance is a multifaceted concept and is addressed at strategic, operational and technical levels. The focus of vendor driven approaches to SOA governance currently is at the technical level, mainly to control the life-cycle of services and their associated policies.

To gain an insight into this level of SOA governance, this research investigates vendor approaches. In order to identify deficiencies in current approaches, the SOA reference model from OASIS is also used to identify

components that ideally need to be governed. From this comparison, additional aspects are identified that need to be addressed by SOA governance. It becomes clear that the governance of service execution is critical in ensuring effective service-oriented architectures. This is particularly prevalent in aspects like security, where actions cannot be ambiguous as they are likely to affect the service execution outcome. This research proceeds to identify service contracts and the enforcement thereof as a means to comprehensively govern service interaction. The paper finally proposes a high-level contract management framework.

KEY WORDS

Service Oriented Architecture, governance, SOA reference model, service contract

CONSIDERING CONTRACTS FOR GOVERNANCE IN SERVICE-ORIENTED ARCHITECTURES

1 INTRODUCTION

Service Oriented Architecture (SOA) (Brown, et al. 2006) is a paradigm for organising and utilising distributed capabilities that may be under the control of different ownership domains, and implemented using a variety of technology stacks. SOA is a holistic approach to designing systems in a distributed environment, where integration is mandatory. Organisations gain a competitive edge by exposing their capabilities or business functions as services, to be re-used for different applications and purposes. Services are well-defined, self-contained, and do not depend on the context or state of other services. Even though SOA is protocol independent, web services technology (Graham, et al. 2002) is becoming the most common implementation of SOA.

SOAs develop in an evolutionary manner, and are different for each organisation and even for each department in an organisation. While most organisations commence their SOA drive with a pilot project, they quickly begin initiatives that span multiple departments or business organisations. Left ungoverned, an SOA could allow anyone to deploy a new service, or invoke and orchestrate any other combination of services. SOA governance is consequently introduced to manage and control the increasing number of services, in order to ensure reuse and consistency, and avoid duplication of work.

Typically, SOAs are governed via the integration of a variety of vendor-oriented governance solutions. These solutions can introduce problems like the violation of principles of SOA. SOA governance should ideally ensure that services are controlled, that they behave in the way they should, and are not misused. In order to assess whether a given service is behaving as it should when it is invoked, service contracts can be used to establish the reference

points for monitoring and reporting by the SOA execution environment. By tracking the actual performance of a service and comparing it to the requirements specified in the service contract, non-compliant services can be identified and timely remedial action can be taken.

This paper considers service contracts as an important element of SOA governance. Section 2 provides a background on SOA governance. Section 3 discusses SOA governance technologies to identify deficiencies in current approaches. To identify additional aspects to be governed by SOA governance, section 4 evaluates the SOA reference model from OASIS. This evaluation identifies that for instance, security policies may be applied ambiguously when services interact. To address this concern, section 5 defines the concept of the service contract. Next, a high-level governance framework is introduced that places a central focus on the use of service contracts for governance. Finally, the paper is concluded.

2 SOA GOVERNANCE

Governance is a set of processes, policies, behaviours, laws and institutions (Von Solms & Von Solms, 2006). These entities influence the approach of developing organisation strategies and objectives into a framework that consists of directives, policies, standards and procedures; implementing this framework operationally; and incorporating metrics that measure the level of compliance regarding the framework. New rules introduced by governance frameworks are forcing companies to rethink how they govern their IT processes. For instance, governance may require of publicly registered companies to show the effectiveness of their internal control structures and reporting procedures. This means organisations need to both control and validate human-to-machine, as well as machine-to-machine service-based interactions.

IT governance is considered a subset of governance (Carter, 2007). It is a framework that consists of processes, organisational structures and leadership, to ensure that the IT systems of an organisation align itself with the strategies and objectives of the organisation (Von Solms & Von Solms, 2006). In its turn, SOA governance can be seen as an extension of IT governance (Carter, 2007).

The term is commonly used to refer to the technology associated with SOA infrastructure such as web service management and security tools, and different Universal Description, Discovery and Integration (UDDI) (Clement, et al. 2004) implementations. The aim of SOA governance is to *control*, *enforce* and *monitor* services throughout their life-cycle (Marks & Bell, 2006), thus ensuring that the services can be reused in an accountable manner across domains of control. SOA governance includes all aspects of IT governance, and also special relationships, policies, and processes to address unique SOA aspects and artefacts. SOA governance is addressed by strategic and operational considerations, and technical mechanisms (Erl, 2008). As governance is a multifaceted concept that is addressed from various angles, the focus of this research is placed at the technical level, to provide metrics and measurements to assist governance decision-making.

The focus of this level of governance is to ensure that services are controlled, behave in the way they should, and are not misused. If a service is designed for a specific purpose and set of consumers, audit logs can for instance prove that the service behaved correctly when the service was invoked. Services should also be available, perform as intended, and be secure. Services that do not comply with these requirements are not governed, and they will inevitably be misused, become unreliable and insecure.

A first aspect of SOA governance to be investigated is the artifacts that are to be governed. In order to address this, the focus is now placed on vendor-oriented SOA governance. The next section provides an overview regarding approaches that are followed, and the artefacts governed by AmberPoint (Amberpoint, 2008), IBM (IBM, 2008) and HP (Systinet, 2008).

3 VENDOR-ORIENTED SOA GOVERNANCE

The vendors approach SOA governance by explicitly addressing visibility, control and trust (Amberpoint, 2008; IBM, 2008; Systinet, 2008). Table 1 consists of a column for each of the three vendors that provide a basic view of their approach to SOA governance. From their perspective, visibility ensures that services are monitored across the SOA lifecycle and that business flows are tracked to assess the business impact. Control is seen as ensuring that systems

deliver a level of quality of service (QoS) that is expected from them within rules and regulations. If consumers are assured of the quality, predictability and transparency of terms and conditions of services, they can trust such services. A main focus of SOA governance is thus to guarantee trustworthy services that can be reused with a high level of assurance. Services need to be managed at design-, run-, change-, and life-time cycle (Marks & Bell, 2006). In order to achieve this, governance vendors employ mechanisms such as registries, repositories, policies, and lifecycle management. Each of theses mechanism is now briefly discussed to highlight the role that they play in governance.

Table 1: Vendors' approach to SOA governance

	AmberPoint	**IBM**	**HP Systinet**
APPROACH	Visibility, Control	Visibility, Control	Visibility, Control , Trust
REGISTRIES	Registry/repository integrated for interoperability	Registry/repository integrated for interoperability	Registry/repository integrated for interoperability
REPOSITORIES	Used during development		
POLICIES			
High-level Man.	Yes	Yes	Yes
3-tiers	Yes	Yes	Yes
Description	WSDL, WS-Policy	WSDL, WS-Policy	WSDL, WS-Policy
Basic interaction	SOAP, WS-Addressing, WS-Notification	SOAP, WS-Addressing, WS-Notification	SOAP, WS-Addressing
Security	WS-Policy, WS-Security, No mention	WS-Policy, WS-Security, WS-Secure Conversion	WS-Policy, WS-Security, WS-Secure Conversion
Reliability	WS-Reliability	WS-ReliableMessaging	WS-ReliableMessaging
Trust	WS-Trust, WS-Federation	WS-Trust, WS-Federation	WS-Trust, WS-Federation
SLA's	WSLA (web service level agreement)	WSLA (web service level agreement)	
Composition	WS-BPEL	WS-BPEL	WS-BPEL
LIFECYCLE MANAGEMENT	Development → Staging → Production	Plan → Define → Enable → Measure	

Registries: The registry is the first and foremost enabling technology for SOA governance. It is a dynamic record of the SOA environment that is used to control and monitor services. A registry holds metadata about services such as its history, who is allowed to make changes to it, who has access to it and how it can be used. A registry that has become an industry standard is Universal Description, Discovery and Integration (UDDI) (Alencar, et al. 2003).

Repositories: Repositories govern the life cycle of services to ensure that a service's records are kept at all stages of its lifecycle. The repository keeps a record of all source code that the organisation develops and provides an audit trail of any previous versions, therefore controlling and monitoring services. The repository can be a separate element or form part of a registry.

Policies: Policies govern the behavior of a service by supplying the rules and constraints that a service needs for successful interaction (Erl, 2006). These characteristics include behavior, preferences, technical limitations and quality of service. Rules and constraints are machine-to-machine specifications that are expressed programmatically as assertions and grouped into various combinations (Erl, 2008). Table 1 illustrates that policies are dealt with at different organisational tiers such as management, architectural and technical. Policy management systems ensure that policies comply with organisational standards, are visible and that they are associated with services. Vendors support the definition of a variety of policies to address aspects such as security, trust, reliability, service-level agreements and composition, as shown in table 1.

Lifecycle management: SOA Lifecycle Management assists with governance by monitoring and controlling policies and processes across the complete SOA lifecycle (Marks & Bell, 2006). It ensures that any changes to a service is monitored and controlled to ensure that the quality of the service remains consistent. Without it, policies may be violated, which may result in noncompliant inefficient services.

By using the abovementioned mechanisms to implement SOA Governance, developers thus have visibility to available services. With a registry and/or a repository in place, they are able to get detailed information

about services by means of their metadata attributes that detail all aspects of the service. In addition, by managing all services aspects in a central location, lifecycle management, change management and impact analysis are facilitated.

An analysis of the table and vendors' approaches identifies that:

- Vendors generally approach SOA governance from their own perspective. Some follow a registry-based approach to control services and policies, and others govern the execution of service interaction. Organisations attempting to address governance comprehensively thus need to integrate various tools to be able to do so.
- Policies are the key element to vendor approaches. Various different types of policies are defined in machine-readable syntax, and are automatically associated with services to control service interaction.
- Vendors support interoperability by adhering to WS specifications.
- Organisations implementing service-oriented systems can be locked into the approach of a specific vendor. This may be to the detriment of the implementation of service-oriented principles such as loose coupling and composability.

In order to further identify SOA aspects and artefacts that need to be governed, the following section investigates the SOA Reference Model. This may identify additional elements that need to be employed to strengthen governance.

4 SOA REFERENCE MODEL

The SOA Reference Model (RM), based on the OASIS SOA RM v1.0 (Brown, et al. 2006) is an abstract framework that focuses on describing services, and the significant relationships and key concepts between them. Key concepts related to the SOA Reference Model namely, visibility, interaction and real world effect are described next.

Visibility: Visibility is when a service consumer has a description of the service and the necessary rules that apply to the service, available to them.

Interaction: Interaction is characterized by actions that occur from passing information between services in the form of messages, or by altering the state of a shared resource. The structure and semantics of exchanged messages is described by an Information Model. A Behaviour Model gives an understanding of service actions, responses, and temporal dependencies between actions on the service. The essence of interaction is grounded in a particular execution context.

Execution context is the agreed upon elements and conditions under which interaction can take place (Brown, et al. 2006) within a specific instantiation of a service (Estes, et al. 2006). Different instances of the same service thus have different execution contexts. The execution context may also evolve during a service interaction. The outcome of execution context is either a change of state or the exchange of information. This is referred to as the real world effect.

Real World Effect: The real world effect is a change of state that has occurred by services participating in the exchange of messages.

To gain an understanding of these concepts, consider the following example: There exists a ServiceA, whose service description is made available to others. Its associated policy, Policy1 is also available to service consumers. *Visibility* is thus an aspect that is addressed by current SOA governance technology through for instance, registries.

Furthermore, Policy1 contains 2 rules. Rule 1 states that if the service consumer is internal to the organisation, a username/password parameter is sufficient, but no QoS guarantees are applicable; alternatively, rule 2 states that if the service consumer is external to the organisation, a certificate must be presented, credit card details will be encrypted and QoS guarantees are applicable. Policy1 governs the interactions of ServiceA with its consumers, but may also be applicable to many other services. It is now possible that a service consumer, external to the organisation, supplies a username/password, and is granted access to ServiceA unintentionally. This happens because the service consumer has not agreed to a service contract, and is choosing to follow rule 1. Consequently, in this *interaction*, ServiceA may be improperly used and

successive service interactions containing credit card details may be unencrypted. This highlights the fact that a policy may be applied improperly, as the consumer has not agreed to use rule 2, and the execution context of ServiceA may differ from one instantiation to the next. It may also differ for different types of service consumers. If this interaction is not actively monitored, the fact that policies are not properly applied may go unnoticed, and the interaction is not adequately controlled. Current SOA governance technology does not sufficiently address this problem.

Finally, the *change of state* occurs for example if ServiceA is accessed to reserve a seat on a flight. This results in ServiceA reserving a seat and receiving money, and the service consumer receiving a reserved seat in exchange for money. To ensure that proper governance of service interaction takes place, the change in state also needs to be monitored. If governance of the visibility, interaction, and change in state is not performed, the result is that a service can be misused; timely remedial action does not occur; or an invalid change of state has occurred. This highlights the following:

- Policies are not agreed to by service consumers, can be applied ambiguously by enforcement points, leading to an improper change in state.
- Different instances of the same service have different execution contexts.
- The execution context of a service interaction needs to be actively monitored.

Using policies for governance cannot prevent this situation from occurring. The next section introduces the service contract, to identify the role that it may play to strengthen SOA Governance.

5 SERVICE CONTRACTS

Service contracts form the foundation for communication between services and therefore represent the most fundamental architectural element of an SOA (Erl, 2008). It supports the relationship between a service and its consumer, and can assist to establish an agreement, and maintain trust between parties. It is not required for the agreement to be entered into legally or to be explicitly

negotiated (OASIS SOA Reference Model Technical Committee, 2006; Jencmen & Yehudai, 2006).

Service contracts are typically unique to a specific service/consumer relationship. It contains formal policies, as well as agreements that are unique to the parties. Furthermore, only semantic information that the organisation wants to make public forms part of the contract (Erl, 2008). A service contract is said to be in place when a valid interaction has taken place (OASIS SOA Reference Model Technical Committee, 2006). Because consumers may vary, there may be multiple service contracts for a single service. SOA governance consequently becomes a process that produces services with a service contract that can be trusted.

Different combinations of policies, applicable to a service, are attached to its service contract. A policy combination that suits given parties is chosen, and the said parties are in agreement regarding the chosen policy combination. Next, a definition of both a policy and service contract is given to distinguish between these concepts.

Policy: A policy is the rules and constraints that govern different aspects of service interaction such as security or reliability. It can be applied to any number of contracts. Examples of policy statements include:

- All interactions with services must be secured with SSL.
- All users should be authenticated with encrypted passwords.
- The service should be available 95% of time.

Service contract: A service contract provides a precise and unambiguous agreement as to how a service and its consumer will interact. It provides a formal definition of the functional and non-functional aspects of the service. The functional aspects include the service endpoint, service operations, input and output messages supported by each operation, and the data representation model of each message's content. The non-functional aspects include the rules and constraints that govern the interaction of service operations. It can also include higher business-level characteristics that are not fundamental to the

service interaction, such as legal requirements. A service contract thus consists of various types of policy statements that can be considered as the clauses of the service contract.

Current standards and technology is widely available to support basic forms of service contracts. For web services, a service contract is collectively viewed as the technical service description defined by WSDL (Christensen, et al. 2001), XSD schemas (Davidson, et al. 1999) and a set of policy documents. Specifications that are used to define policy documents include WS-Policy (Bajaj, et al. 2006), which is used as a container for specifying a range of policy considerations. Specifications such as WS-Security (Hallam-Baker, et al. 2006), Web Services Business Process Execution Language (WS-BPEL) (Alves, et al. 2007), Web Service Level Agreements (WSLA) (Dan, et al. 2003) and Web Service Offerings Language (WSOL) are used to specify a variety of non-functional requirements. Future developments such as the Ontology Web Language for Services (OWL-S) (Burstein, et al. 2004) aim to provide a better language for defining service contracts.

The following section describes a framework for SOA governance that centrally positions service contracts in its approach.

6 GOVERNANCE-BY-CONTRACT FRAMEWORK

As stated, the aim of SOA governance is to *control*, *enforce* and *monitor* services throughout their life-cycle, ensuring that they can be reused in an accountable manner and across domains of control. To address this, the framework for governance-by-contract consists of two phases. The first phase addresses *control* by service contract design, and the second, *enforce* and *monitor* by the enablement of governance-by-contract. The framework does not aim to replace current SOA governance technology, but rather aims to define an approach to using such technology. The main focus of the framework is to address the role that service contracts can play to strengthen SOA governance. The first phase to be addressed is service contract design.

6.1 Service contract design

For governance, *control* means to ensure that adequate measures are in place to provide assurance that objectives will be achieved and undesirable events will be prevented or detected and corrected (IT Governance Institute, 2007). For SOA governance this means creating, implementing and managing policies and service contracts to provide rules and constraints for a service and its consumers to follow. Also, because the service contract is shared amongst service consumers, its design is particularly important. Service consumers agreeing to the service contract become dependent on its definition. Therefore, service contracts need to be carefully designed, maintained and versioned after their initial release.

Service contracts, designed with a view on service governance should be created as follows:

- Standardise the vocabulary that will be used to describe policies and service contracts.
- Design the functional interface of the service.
- Design the non-functional requirements of the service such as security, reliability, or service-level agreements. This process should formally consider governance frameworks that the organisation complies with such as Cobit (IT Governance Institute, 2007).
- Identify each possible execution context required for a service interaction. Service consumer or group of consumers may require different levels of, for instance, service-level agreements or security.
- Identify policies required by each execution context of a service.
- Associate policies to the service contract for a specific execution context.

Because a service contract is specific to the interaction between a consumer and the service, it can establish reference points for monitoring and tracking whether or not service consumers are abiding by the requirements specified in the service contract. Therefore, designing a service contract with governance in mind will strengthen the governance process.

Furthermore, the framework is based on the notion of varying levels of service contracts. For example, the service contract of a service that provides weather reports for portal applications do not need a high level of governance, but that same service may need strict governance if it is being used by a military system. The weather service used by portal applications interacts with a basic service contract that consists of functional specifications. Consequently, a low level of SOA governance needs to be implemented. For instance, the visibility of the service can be ensured through a registry. On the other hand, the weather service used by a military system needs to be protected by associating information security policies to its service contract. In this case, governance of the service execution is required to ensure that rules and constraints attached to the service contract are properly applied. As more non-functional aspects are added to a service contract, the required degree of governance thus increases.

Policies are applied to each service contract according to the non-functional requirements of the service. Previous research identified that service contracts can be structured according to such aspects (Jencmen, 2006; Cubera, 2007). The framework now proposes that service contracts are structured according to three high-level categories, namely:

- *Basic:* A basic contract addresses the functional aspects of a service such as how to locate the service and what the service is about. Such a contract is used when a service has minimum requirements with respect to governance, as it has little impact on the performance of the organisation.
- *QoS:* A QoS service contract addresses non-functional aspects such as security, reliable delivery, and performance. There are a variety of QoS aspects that can be included to increase the quality of the service. Services with such requirements have a significant impact on the performance of the organisation and need to be measured to ensure that they meet their requirements. These types of service contracts differ for service consumers and may be negotiable.
- *Behavioural:* To consider the dependencies between the functions provided by the service, the behavioural service contract defines the expected behaviour of

a service participating in a conversation with others. The conversation can be an orchestration or a composition of services. This contract includes requirements to ensure that a service will behave appropriately in a sequential context. As conversations take place across different domains, governance of these aspects is vital to maintain trust between a service and its consumer.

Establishing levels of service contracts to assist with governance is a challenge that will require significant attention in the future. Service contracts cannot be developed in isolation, but their development must be guided by current governance frameworks. The next paragraph addresses the second phase of governance-by-contract.

6.2 Enablement of governance-by-contract

The quality of service execution can be seen as a reflection on the level of SOA governance. If the health of a service degrades during service execution, the consumer is directly affected. To ensure the health of a service, service contract clauses are *enforced* and *monitored* by applicable enforcement and governance points.

Enforce means to compel components to abide by the rules and constraints (Hawkins, 1995). For SOA governance this means implementing mechanisms to coerce a service and its consumers to abide by the rules and constraints of service contracts. This means to implement the logic for the various governance aspects such as enforcement of encryption requirements, exceptions, events, or counters, as defined by the service contract.

Monitor means to ensure that the right things are done and that these are in line with policies (IT Governance Institute, 2007). For SOA governance this means confirming whether or not service contracts are being properly applied. Monitoring is performed by an external point to monitor the operational state of the service. This is done by observing the change in the state of real world values. A monitor has a predefined set of rules, defined according to the service contract, which would observe when values cross certain thresholds and the QoS of the service deteriorates. The monitor would then raise an alert and

provides feedback to the organisation so as to assist with governance. The monitor is passive and would not actively manipulate the service.

Although there are many current SOA governance technologies to govern services and their execution, there are no standards or methodologies to capture governance requirements from which to build a formal service governance model. The governance-by-contract approach is a first step to ensure that the service contract is not to be circumvented (Erl, 2008) by discouraging the improper application of policy rules, and the misuse of a service or an invalid change in state.

7 CONCLUSION

SOA governance is a very important and current topic that is being addressed by the IT community. It resides at the intersection between a new technology, namely SOA, and IT governance. To ensure the success of SOA, firm and consistent governance is needed.

Current approaches to SOA governance may lead to policies being applied ambiguously when service interaction occurs. To address this problem, service contracts are created for specific consumers or groups of consumers and these are agreed upon. The proposed governance-by-contract framework identifies how service contracts are designed with governance in mind, and includes mechanisms to control, enforce and monitor services. The governance-by-contract approach addresses aspects such as security requirements, service-level agreements based on QoS and key process indicators, and performance management, as set out in the service contract. The framework does not aim to replace current SOA governance technology, but rather seeks to use this technology to approach governance comprehensively.

This paper has introduced the concept of governance-by-contract. There is still much work to be done regarding the design of service contracts and their enablement. Future research aims to investigate, for example, information security governance frameworks in order to define a formal approach to defining the information security policies of a service contract and its enablement.

8 REFERENCES

Alencar, P., Cowan, D. & Kalali, B., (2003), A Service-Oriented Monitoring Registry. *ACM Digital Library*, 107-121, Oct., 2003.

Alves, A., Arkin, A., Askary, S., Barreto, C., Bloch, B., Curbera, F., Ford, M., Goland, Y., Guízar, A., Kartha, N., Liu, C.K., Khalaf, R., König, D., Marin, M., Mehta, V., Thatte, S., van der Rijn, D., Yendluri, P. & Yiu, A. (Editors). (2007). Web Services Business Process Execution Language Version 2.0. Available from: http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html. (Accessed 10 March 2008).

Amberpoint. (2008), Available from: http://www.amberpoint.com. (Accessed 15 October 2007).

Bajaj, S., Box, D., Chappell, D., Curbera, F., Daniels, G., Hallam-Baker, P., Hondo, M., Kaler, C., Langworthy, D., Nadalin, A., Nagaratnam, N., Prafullchandra, H., von Riegen, C., Roth, D., Schlimmer (Editor), J., Sharp, C., Shewchuk, J., Vedamuthu, A., Yalçinalp, U. & Orchard, D. (2006). Web Services Policy 1.2 - Framework (WS-Policy). Available from: http://www.w3.org/Submission/WS-Policy.

Brown, P. F., Hamilton, B. A. Laskey, K., MacKenzie, C. M., McCabe, F. & Metz, R. (Editors). (2006). OASIS: Reference Model For Service-Oriented Architecture 1.0. Available from: http://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf. (Accessed 20 September 2007).

Burstein, M., Hobbs, J., Lassila, O., Martin, D., (editor), McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, P., Sirin, E., Srinivasan, N. & Sycara, K. (2004). OWL-S: Semantic Markup for Web Services. Available from: http://www.w3.org/Submission/OWL-S. (Accessed 26 March 2008).

Carter, S. (2007). *The new language of business: SOA and Web 2.0.* IBM Press.

Christensen, E., Curbera, F., Meredith, G. & Weerawarana, S. (2001). Web Services Description Language (WSDL) Version 1.1. Available from: http://www.w3.org/TR/wsdl. (Accessed 14 November 2007).

Clement, L., Hately, A., Rogers, T. & von Riegen, C. (Editors). (2004). OASIS: UDDI Version 3.0.2. Available from: http://uddi.org/pubs/uddi_v3.htm (Accessed 17 April 2008).

Curbera, F, (2007), Component Contracts in Service-Oriented Architectures, *Computer*, vol. 40, no. 11, pp. 74-80, Nov., 2007.

Dan, A., Franck, R., Keller, A., King, R.P. & Ludwig, H. (Editors). (2003). Web Service Level Agreement (WSLA) Language Specification. Available from: http://www.research.ibm.com/wsla/WSLASpecV1-20030128.pdf. (Accessed 20 February 2008).

Davidson, A., Fuchs, M., Hedin, M., Jain, M., Koistinen, J., Lloyd, C., Maloney, M. & Schwarzhof, K. (1999). Schema for Object-Oriented XML 2.0. Available from: http://www.w3.org/TR/NOTE-SOX/. (Accessed 4 October 2007).

Erl, T. (2006), *Service Oriented Architecture: Concepts, Technology, and Design.* New York: Prentice Hall

Erl, T. (2008), SOA Principles of Service Design: Indiana:Prentice Hall

Estes, K., Kesavarapu, K., Shipe, B. & Strom M. (2006). *Service-Oriented Architecture*. Unpublished manuscript.

Graham, S., Gottschalk, K., Kreger, H. & Snell, J. (2002), Introduction to Web services architecture, IBM Systems Journal, Volume 41, Number 2

Hallam-Baker, P., Kaler, C., Monzillo, R. & Nadalin, A. (Editors). (2006). Web Services Security: SOAP Message Security 1.1. Available from: http://www.oasis-open.org/committees/download.php/16790/wss-v1.1-spec-os-SOAPMessageSecurity.pdf. (Accessed 10 January 2008).

Hawkins, J.M., (compiled by), (1995). *The South African Oxford School Dictionary*. 5th edition. Cape Town: Oxford University Press.

IBM. (2008). Available from: http://www-306.ibm.com/software/solutions/soa. (Acessed 28 November 2007).

IT Governance Institute. (2007). *Cobit 4.1*. Illoinois: IT Governance Institute.

Jencmen, A. & Yehudai, A. (2006), Fortified Web Services Contracts for Trusted Components, *Computer,* pp. 919-926, Sep., 2006

Marks, E.A. & Bell,, M. (2006), *Service-oriented Architecture A Planning and Implementation Guide for Business and Technology.* New Jersey:Wiley

OASIS SOA Reference Model Technical Committee (2006). Policies and Contracts. Available from: http://wiki.oasis-open.org/soa-rm/TheArchitecture/PoliciesAndContracts (Accessed 25 January 2008).

Systinet. (2008). Available from: (https://h10078.www1.hp.com/cda/hpms/display/main/hpms_content.jsp?zn=bto&cp=1-11-130-27%5E1461_4000_100__). (Accessed 20 October 2007).

Von Solms, S.H. & Von Solms, (2006). *Information Security Governance*. Unpublished draft.

A CANONICAL IMPLEMENTATION OF THE ADVANCED ENCRYPTION STANDARD ON THE GRAPHICS PROCESSING UNIT

Nick Pilkington[1], Barry Irwin[2]

Rhodes University
Department of Computer Science
South Africa

[1]n.pilkington@ru.ac.za, [2]b.irwin@ru.ac.za

ABSTRACT

This paper will present an implementation of the Advanced Encryption Standard (AES) on the graphics processing unit (GPU). It investigates the ease of implementation from first principles and the difficulties encountered. It also presents a performance analysis to evaluate if the GPU is a viable option for a cryptographics platform. The AES implementation is found to yield orders of maginitude increased performance when compared to CPU based implementations. Although the implementation introduces complications, these are quickly becoming mitigated by the growing accessibility provided by general programming on graphics processing units (GPGPU) frameworks like NVIDIA's Compute Uniform Device Architechture (CUDA) and AMD/ATI's Close to Metal (CTM).

KEY WORDS

Cryptography, AES, GPU, GPGPU, Offload, Rijndael, OpenGL, CG

A CANONICAL IMPLEMENTATION OF THE ADVANCED ENCRYPTION STANDARD ON THE GRAPHICS PROCESSING UNIT

1 INTRODUCTION

General programming on graphics processing units (GPGPU) refers to non-graphics related programming operations being performed on the graphics processing unit (GPU) rather that on the CPU. This programming paradign opens up many possibilities for increased performance by utilising the specialised processing nature of the GPU. There are currently two frameworks available to program GPUs, namely Compute Uniform Device Architechture (CUDA) [3] from NVIDIA and Close to Metal (CTM) [2] from AMD/ATI. These frameworks currently only support their native GPU architechture and as a result are not portable across all hardware. GPGPU can, however, be acheived from first principles in a general way that allows the code to be executed on a wide range of different hardware configurations. This method will be explained in section **??**. This approach requires that manufacturer specfic caveats and optimizations cannot be taken advantage of, since a canonical implementation is being presented general applicability is more important that specifically optimised performance. There are a number of popular cryptographic algorithms in use in computing today including AES [6], Triple DES [9], and Blowfish [12]. Rijndael (AES) was selected for sample implemetation as it is the FIPS accepted Advanced Encyption Standard [6]. This paper seek to investigate, in detail, the implementation of AES on a GPU, more specifically it will be concerned purely with the encryption process as the decryption process is similar. It also presents a performance analysis of the implementation in comparison to CPU based implementations and discussion to substantiate the results. Finally sample applications and proposed future derivative works are suggested.

2 AES ENCRYPTION

Advanced Encryption Standard (AES) is a symmetric key cryptographic algorithm also known as Rijndael designed by Vincent Rijmen and Joan Daemen in 1998, it was subsequently adopted as the Advanced Encryption Standard in 2002. AES is a block cipher which means that is encrypts data in

Table 1: Key-Block-Round Combinations

Type	Key Length	Block Size	Number of Rounds
AES-128	4	4	10
AES-192	6	4	12
AES-256	8	4	14

Algorithm 1 AES Encryption Pseudocode

```
KeyExpansion
Initial Round
AddRoundKey

for N = 1 to Rounds-1
  SubBytes
  ShiftRows
  MixColumns
  AddRoundKey

SubBytes
ShiftRows
AddRoundKey
```

finite blocks as opposed to operating on a stream like Trivium [7]. The block of data to be encrypted is termed the state. In AES the state is a 4x4 matrix of bytes (figure 1). The state paired with an encryption key of a certain length form the inputs for the AES algorithm. AES is comprised of four different stages, which together represent a single round. Each stage performs some operations on the current state. The number of rounds varies with different implementations of AES (table 1). This paper implements AES-128. Algorithm 1 gives the pseudocode for the AES Encryption process and a depiction of an encryption stage is shown in figure 2. It should be noted that the final round of the encryption process varies from the rest as the mix columns stage is ommitted.

Figure 1: AES Encryption State

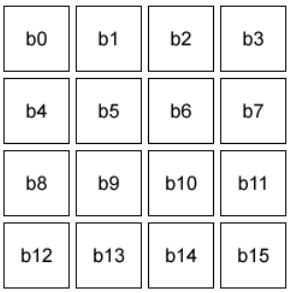
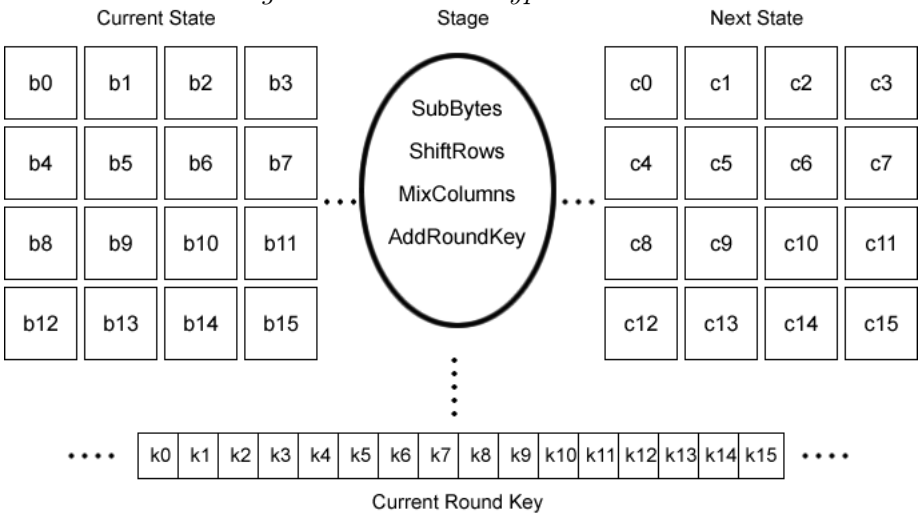


Figure 2: AES Encryption Round



3 GENERAL PROGRAMMING ON GRAPHICS PROCESSING UNITS

Until the third generation of GPUs was released in 2001, GPUs were not programmable and were merely configurable to a limited degree. The advent of this generation exposed areas of the graphics pipeline to programmers allowing them execute custom code on the GPU. This is achieved through pixel shaders which execute once on each rendered pixel in the viewport. Colour, vertex and normal data does not need to be interpreted geometrically but are in fact just arrays of numbers. Rendering an NxN quad onto the screen call the execution of any mapped pixel shaders on each of the N^2 elements of the quad and their output value overwrites the value currently at that position in the quad. Once a problem has been formulated in terms of shaders and rendering it can be mapped and solved on the GPU. A thorough treatise of the basics of GPGPU and how it can be achieved from first principles is given in [10].

4 APPROACH

At the time of writing there are three different shader languages available for programmable shaders: HLSL [4], GLSlang [11] and Cg [1]. Cg and the OpenGL API were used for this implementation. The basis for the AES encryption algorithm is rooted deeply in algebra and the technical specifics of the algorithm [6] have been omitted from this paper. This section will present a high level view of each of the four stages of the encryption process and how each was modelled on the GPU. Each stage of the encryption process was implemented in a separate shader. The four shaders were each executed in order ten times on the initial state to encrypt the data.

4.1 Encryption Stages

Key Expansion

The first stage in the AES encryption process is to expand the key to ten times its original size such that there is a key for each round of the algorithm [6]. This is a pre-process to the encryption process and as such the expanded key was precomputed.

Substitute Bytes

The substitute bytes step of the algorithm replaces each byte in the current state with a corresponding byte using an 8-bit Sbox [6]. The Sbox represents a non-linear transformation. This transformation can be represented in matrix form as:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

The actual transformations to generate the Sbox do not need to be computed explicitly. As the values are constant for a given initial b vector. There are 256 different different b vectors and as a result 256 corresponding Sbox transformed values. The operation can be viewed as a table look up. Thus the resulting look up values for all 2^8 initial b vectors can be computed and stored in a 16x16 texture. The GPU indexes into this texture using the current byte in the state and received the Sbox transformed value, which is then written into its place.

Shift Rows

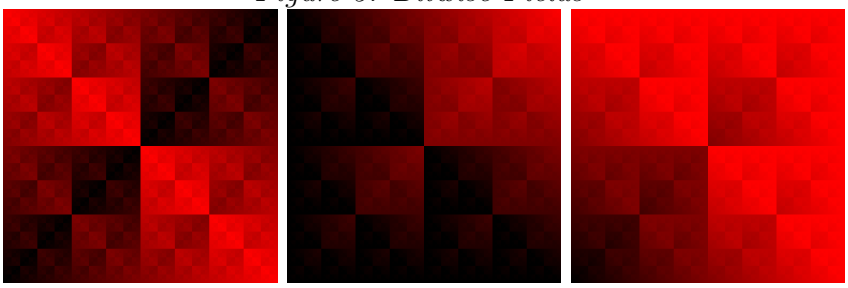
The shift rows operation cycles the bytes in each row cyclically left. The first row is not shifted, the second row is shifted one position, the third row two positions and finally the forth row three positions [6]. The shift operation is performed on the GPU by offsetting the current fragment shaders texture coordinates based on its row and performing a single texture look up on its own texture.

Mix Columns

The mix columns stage operates on each of the four columns of the state. Each column of the state is representative of a four-term polynomial over the Galois field GF(2^8) [6], this polynomial is multiplied modulo $x^4 + 1$ with the fixed polynomial $a(x)$, given by:

$$a(x) = \{03\}x^3 + \{01\}x^2 + \{01\}x^1 + \{02\}$$

Figure 3: Bitwise Fields



(a) XOR. (b) AND. (c) OR.

This can be written as a logical bitwise matrix multiplication in the form $s(x)' = a(x) \oplus s(x)$. Shader languages like Cg do not have support for logical operations [8]. Although reservation has been made for the corresponding symbols &, | and ^ [8], they had not been implemented at time of writing. This makes a seemingly trivial task like a logical XOR impossible to perform without some other mechanism in place. In order to provide this functionality, a look up table of values was precomputed and stored in a texture. A 256x256 texture was used and its red, green and blue colour channels corresponded to the XOR, OR and AND operations respectively. These are all binary operations and the x and y indices of the texture correspond to the operands and the values stored in each channel to the resulting binary operations value. When a bitwise operation needed to be performed the two operands were scaled to the texture coordinate range of $[0.0 \ldots 1.0]$ and a dependent texture look up was performed on the texture. The resulting colour channel could then be read to give the XOR, OR or AND of the operands respectively. Figure 3 depicts the red, green and blue channels respectively. The limitation of this implementation is the range of values of the operands. A single 256x256 texture was used and since AES operates within this range of values, these constraints were not problematic.

Add Round Key

The add round key stage XORs the current key with the state. With bitwise operations the XOR operation can be implemented as the XOR between the current byte of the state and the corresponding byte of the key.

4.2 Algorithm Execution

With each of the operations of AES implemented, the whole encryption process can be achieve by encoding the initial state and expanded round key into textures. A 4x4 pixel quad was then rendered to the screen with the initial sub bytes fragment shader bound. This produced the output for the first stage of the AES encryption. The contents of the frame buffer were then copied back into the texture using a render to texture feedback mechanism after which the shift rows fragment shader was loaded and another 4x4 pixel quad rendered. This process was replicated for the mix columns and add round key stages to yield one iteration of the AES encryption. Since ten iterations were required, the whole process is preformed 10 times giving the encrypted state. Care was taken to treat the final iteration correctly, since the add round key operation does not take place here [6].

4.3 State Tiling

GPU shader operations take place in parallel [10]. Since the only data dependence in AES encryption is that the stages of the encryption take place in order there is no reason to limit processing to a single state per rendering if more than one can be represented. If a single 4x4 texture were used to represent the current state it would utilise less than 0.0016% of a 1024x1024 view space. For this reason $65,536$ state were tiled across the view port to enable complete utilisation of the view space as in figure 4.

5 TESTING CONFIGURATION

The GPU implementation was benchmarked for speed and accuracy. Its speed was measured by how much data it could encrypt per second. This amount was measured as the average amount of data encrypted per second over a 60 second run. All runs were executed on the machine specification detailed in table 2. The encryptions were performed on deterministically random data. The results of each stage of the AES encryption process were cross validated against OpenSSL's AES implementation [5] to ensure correctness.

Figure 4: State Tiling in the View port

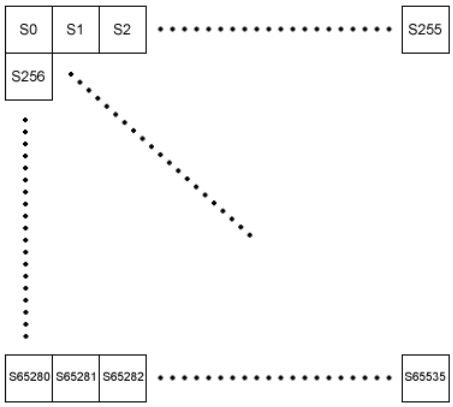


Table 2: Test Platform Configuration

Category	Details
Processor	Intel Core 2 Duo (1.86Ghz)
Memory	2048MB DDR2 (400Mhz)
Graphics	NVIDIA GeForce 7900 GT (256MB)
Mainboard	Intel Corporation Q965
Hard drive	80GB SATA
Operating System	Windows XP Service Pack 2

Table 3: AES Encryption Rate

Type	Encryptions per Second (16-byte state)
CPU	7254.25
GPU	25449.65

Table 4: Maximum Encryption Rate

Type	Encryption Rate (Mb/s)
CPU	2.32
GPU	12.00

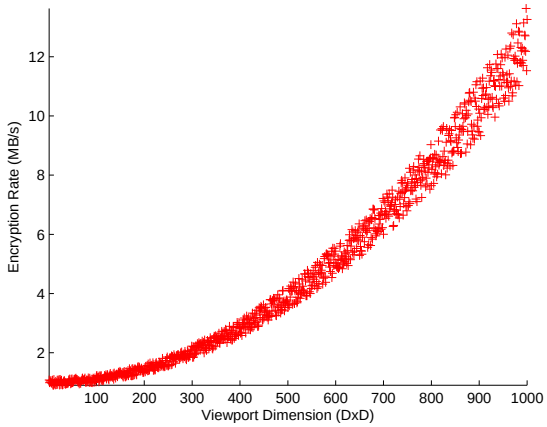
6 RESULTS

Tables 3 and 4 show the average number of complete AES encryptions performed on both the CPU and GPU and the average encryption rate.

7 PERFORMANCE ANALYSIS

Considering the results of both the CPU and GPU implementations in table 4 the GPU outperforms the CPU by 5.17 times. It is important to gain a deeper understanding of what causes this vast performance increase. In general GPUs are slower than CPUs on the clock speed basis the performance gain is not due to this. Figure 5 shows the results of the GPU implementation of AES encryption with increasing numbers of states tiled into the the view port. It may seem retrogressive to look at performance results with smaller tiling, but it is instructive in understanding how the results in table 3 are achieved. It can be seen from figure 5 that data volume is not bottle-necking the encryption process, as when more data is tiled into the view port the encryption rate increases. GPUs perform well on large streams on uniform data and this statement is mirrored by the graph. Using a view port of 1024x1024 and tiling the states, as detailed in subsection 4.3, allowed all of them to be encrypted in parallel. This allowed for far more data to be encrypted per rendering. A CPU cannot do this, thus gains nothing from being passed more concurrent data as it all needs to be processed sequentially anyway. Figure 5 implies that more blocks will yield even higher encryption rates. There is a limit to the size of the renderable surface while maintaining a 1:1 aspect ratio. This problem can be circumvented in a number of ways

Figure 5: AES Encryption Rate



but these methods are not general and as a result is one of the advantages of GPGPU frameworks like CUDA and CTM.

8 CONCLUSION

The results in section 6 show that high performance encryption is possible on the GPU. The unoptimized implementation used exhibited large performance increases over the CPU implementation. The results show that this performance increase is due to the parallel processing nature of the GPU and its ability to operate on more than one data item concurrently. By tiling more that one state into the view port the GPU is able to take advantage of the per-stage parallelism of AES and yield large performance gains. As mentioned on the outset the implementation was restricted to a canonical method such that it could illustrate a proof of concept that is invariable across different hardware configurations. In recent months a large emphasis has been placed on the computing power of GPUs and as a result general computing framework have been released from both NVIDIA and ATI. These allow for more fine grained thread control of the execution of the code which is beyond the OpenGL implementation presented here. The advantage of the implementation presented here is that is that it achieves the same ends as an implementation on CUDA or CTM would but without the abstraction layer that masks the finer implementation details. The developments in CUDA and CTM have largely eclipsed this kind of GPGPU development,

however it still remains important as a foundation of understanding. This implementation paves the way for implementing further mainstream cryptographic algorithms like 3DES and Blowfish on the GPU and making similar performance analyses.

References

[1] The cg language. Tech. rep., NVIDIA Corporation (Available Online: http://developer.nvidia.com/).

[2] Close to the metal project, amd/ati. Available Online: http://sourceforge.net/projects/amdctm/.

[3] Cuda, nvidia corporation. Available Online: http://www.nvidia.com/.

[4] High level shading language, microsoft corporation. Available Online: http://msdn.microsoft.com/.

[5] The openssl project. Available Online: http://www.openssl.org/.

[6] *Federal Information Processing Standards Publication 197, ADVANCED ENCRYPTION STANDARD (AES).* 2001.

[7] Christophe De Cannière, B. P. Trivium specifications. Available Online: http://www.ecrypt.eu.org/stream/ciphers/trivium/trivium.pdf.

[8] Fernando, R., and Kilgard, M. *The Cg Tutorial*. Addison-Wesley Professional, 2003.

[9] Kammer, R. G. *Federal Information Processing Standards Publication, DATA ENCRYPTION STANDARD (DES).* U.S. Department of Commerce/National Institute of Standards and Technology, 1999.

[10] Pilkington, N. An investigation into general processing on graphics processing units [unpublished]. Department of Computer Science, Rhodes University, South Africa.

[11] Rost, R. The opengl shading language. Available Online: http://www.opengl.org/.

[12] SCHNEIDER, B. The blowfish encryption algorithm. http://www.schneier.com/blowfish.html.

INVESTIGATING THE FACTORS IMPACTING THE ADOPTION OF BIOMETRIC TECHNOLOGY BY SOUTH AFRICAN BANKS

Antonio Pooe and [1]Les Labuschagne

University of Johannesburg

[1]LesL@uj.ac.za

ABSTRACT

This paper investigates the cause for the slow adoption of biometric authentication in the South African (SA) banking sector and constitutes exploratory research.

Various definitions of biometrics are analysed to determine the common elements. Based on these elements, a new definition is proposed.

This study is limited to the use of biometric technology within the financial services sector. Within the said sector, specific focus is placed on the four leading SA banks. A survey was conducted and forty usable responses were received. The initial results of the survey are analysed and interpreted in this article. The survey also provides insight into current and future biometric technologies used for authentication purposes within the financial services sector.

The value of this article is that is provides insight into the current state of biometric technology in SA.

KEY WORDS

Biometrics, authentication, financial sector, banks, empirical research, questionnaire.

INVESTIGATING THE FACTORS IMPACTING THE ADOPTION OF BIOMETRIC TECHNOLOGY BY SOUTH AFRICAN BANKS

1 INTRODUCTION

The complexity surrounding the challenges in information security continues to grow. These challenges are brought about by ever-increasing incidents of unlawful activity on the internet and viruses that are now propagating at unprecedented speeds. In addition to this, criminal exploits such as "Nigerian scams" also known as "419 scams" and other forms of email fraud and intolerable spam irk computer users around the world (Skalak et al., 2007).

There has also been a significant tightening of legislation around privacy and confidentiality of personally identifiable financial, health or other sensitive information. These governance and legislative requirements bring about a different way of thinking when using and deploying technology in general, especially for security experts whose responsibility it is to put systems in place that meet these legal requirements (Whitman, 2006).

Identification and authentication have typically been achieved by individuals displaying a document such as a licence or a passport, that is, something they have in their possession. In some cases, a user has also been required to produce a password or a personal identification number (PIN), that is, something they know. In digital environments, it is more common to use something you have, a username together with something you know, a password (Nanavati et al., 2002; Layton, 2007).

The authentication challenges that arose from using only something you know and have were brought about by attempts to remember the latter. These attempts ranged from the password or passkey being written down, shared with colleagues, being attacked by an intruder through guess work or social engineering, being attacked using brute force and many other ways. As the sophistication and number of attacks increases, more secure and

accurate measures or authentication are required leading to the investigation of something the user 'is': biometrics (Krutz et al., 2003).

Financial institutions are particularly vulnerable when it comes to authentication as several cases of unauthorised account access have been reported in the media (Da Silva, 2007). Several international banks have already adopted biometrics as an authentication mechanism (Krawczyk et al., 2005) yet SA banks seem to lag behind this trend. The goal of this paper is to report on the findings of the empirical research that was conducted to establish the reasons for this slow adoption.

This paper represents exploratory research. Devlin (2006) suggests that this approach has the goal of formulating problems more precisely, obtaining insight and forming a hypothesis. This type of research is usually small-scale and undertaken to define the exact nature of the problem with a view to gain better understanding of the environment within which the problem exists.

The research problem is, therefore, the slow adoption of biometric technology by SA banks.

The objectives of this paper are to:

I. establish if bank employees have ever been exposed to biometric technology. This can be exposure from within the workplace or external to their organisations. This information is used to determine how the lack of exposure to biometrics technology affects their opinions on whether this technology can work for banking applications or not.

II. capture the perceptions and opinions of the respondents with regards to the future use of biometric authentication in their organization. These views provide insight to the level of awareness and buy-in on biometric authentication and identify the problem areas affecting adoption.

III. measure the participating banks' interest in biometric technology. The survey ascertains if the organization has or is investigating biometric authentication. This information is helpful in determining if the participating banks are planning to deploy biometric authentication and obtaining information relating to areas where this technology is most likely to be deployed.

The survey is limited to the use of biometric technology within the financial services sector in SA. Within the said sector, participation was limited only to the four leading banks, namely Standard Bank of South Africa Limited, ABSA Bank, First National Bank and NedBank (STD Bank, (2008); ABSA, (2008); Nedbank, (2008); FNB, (2008)).

The majority of the questions adopted a bipolar scaling method which uses a five point Likert scale (Dawes 2008, pg 61-77). The questionnaire consists of the following five main sections:

1. Background
2. General Knowledge of Biometrics
3. Organisational Research
4. Current Usage
5. Perceptions

Following is a short explanation of the purpose of each section.

1.1 Background

The purpose of this section is to capture the background of the respondent, including limited biographical data. This yielded valuable information relating to ethnic or gender preferences.

1.2 General Knowledge of Biometrics

This section establishes if the respondent is aware of or has used biometrics before. The aim is to observe from the data gathered if knowledge and previous exposure changes the perceived usability and value of biometric technology.

1.3 Organisational Research

The aim of this section is to measure the participating banks' interest in biometric technology. The study ascertains if the organization has or is currently investigating biometric authentication as a viable alternative to current information security mechanisms.

1.4 Current Usage

This section establishes if the organization is currently using biometric authentication as opposed to just investigating it in the previous section.

This information is helpful in determining if the participating banks are planning on deploying biometric authentication and obtaining information relating to areas where this technology is most likely going to be deployed.

1.5 Perceptions

The aim of this section is to capture the perceptions and opinions of the respondents with regards to the future use of biometric authentication in their organisation. These views provide insight into the level of awareness and buy-in on biometric authentication.

The next section analyses various definitions for biometrics to determine the main components.

2 BIOMETRIC AUTHENTICATION

At the core of security services are identification and authentication, authorization, confidentiality, integrity and non-repudiation (Reid 2004:9). All these services are interrelated and interdependent. The focus of this paper is on the identification and authentication service.

As Reid (2004:5) defines it, biometrics is a physical or psychological trait that can be measured, recorded, and quantified. In so doing, the trait can be used to obtain a biometric enrolment thus determining, with a degree of certainty, that someone is the same person in future biometric authentications based on their previous enrolment authentications.

Another view on the definition of biometrics is that of Azari (2003:112-113) wherein he states that a biometric is some measurement of the biological characteristics of an (human) individual. Under this definition, there are many forms of biometric data for which capture and verification is possible via some device. Fingerprints, voice recognition, and retinal face or hand scanning are all feasible with current technology. However, the nature of biometric data is such that there are significant risks associated with its capture and use in a secure environment.

Nanavati et al. (2002:9) offers a more simplified definition wherein he states that biometrics is the automated use of physiological or behavioural characteristics that determine or verify identity.

Several aspects of the three definitions as presented above require elaboration. It is interesting to note that in all the noted definitions, there are

a few common and uncommon terms or views of what makes up a definition of biometrics.

I. Biological

It is apparent that biometrics has something to do with biological or, in other words, physical and/or psychological/physiological traits, and is the starting point for the definition of biometrics. This trait is one that fits back into the three pillars of authentication (Reid, 2004:9). This trait is something the user is, and can be used on its own or along with something the user knows or has.

II. Measurable

The two definitions by Reid (2004) and Azari (2003) as presented above speak of the biological trait being measurable. This suggests that there must be some level of uniqueness in the biometric trait for it to be measurable (uniqueness thereof), and be used for authentication. This measurement is then used to compare the user's presented biometric to the stored or trusted biometric trait.

III. Recording or Enrolment

This term is unique to the definition by Reid (2004). It suggests that there is a point where the biometric trait is recorded for future use. The use of this recorded biometric trait is for comparisons between this known biometric trait and an unknown biometric trait that will need to be authenticated. During the enrolment phase, the individual's biological trait is converted into a digital string called a template. The engine that performs the conversion is then referred to as a biometric algorithm. This enrolment process is the key to the performance and accuracy of the biometric application ("biometric system").

IV. Automation

Unique to the definition by Nanavati et al. (2002) is reference to the notion of automation. This refers to the comparisons of the stored template and the live or presented template, that take place for the purposes of authentication. This suggests that if this comparison process is manual, then it does not qualify as a biometric process.

V. Determination or verification

Nanavati et al. (2002) further speaks of a process of determination or verification. These terms are unique to this definition. Determining versus verifying identity represents a fundamental distinction in biometric usage. Determining is also referred to as identification, and is a process whereby a one-to-one (1:1) matching or comparison takes place during authentication. On the other hand, verification is a process whereby one live template is matched against a database of many stored templates, represented as (1:N).

From the analysis of the definitions as discussed above, a new definition for the purposes of this article is proposed. Biometrics is the automated use of physiological or behavioural trait/s that can be measured and recorded, to determine or verify an individual's identity. Physiological traits include fingerprints, palm veins, eye retina, eye iris, hand measurements and facial patterns. Behavioural traits include the way the individual walks or gait, typing patterns, signature and the way a person speaks.

Because a person cannot leave their eye or hand on a computer monitor as they would a written down username and/or password; or forge their Deoxyribonucleic acid (DNA) as they would an Identity Document, biometric technology is therefore said to offer better security in applications across the board (Real Time North America n.d.).

With the great number of biometric solutions available in the market, the challenge arises in selecting the correct technology to address a specific need.

Care should be taken when selecting the specific biometric solution or combination thereof, to address a specific need in order to archive maximum security benefits. Failure to do so may result in catastrophic failures, huge financial losses and may even give birth to a national security nuisance (Garfinkel, 2005).

A requirement specific to the banking sector is that authentication of clients is allowed from within the same bank and from other banks' clients accessing shared banking resources.

The next section explains the research design that was followed to conduct empirical research.

3 RESEARCH DESIGN

A questionnaire was administered online to gather data from various respondents representing the different banks. The main reason for adopting this method over traditional methods of self-administration was to speed up the distribution of the questionnaire and collection of data.

As explained by Greenfield (2002:178-179) and Devlin (2006:131-135), internet-based surveys can be conducted in two ways:

I. By using an email to distribute and collect questionnaires. The format for such a method could be in one or more of the following:
   - Plain text questions inserted as part of the email;
   - The actual email message formatted in HTML;
   - A formatted questionnaire send as an email attachment; and
   - An interactive questionnaire from an executable file that can be sent as an attachment to the email.

II. By using web pages. This method entails the administration of the questionnaire through internet web pages. There are many applications available that facilitate the design and administration of online questionnaires.

For purposes of this study, a combination of the two discussed methods was used. The URL of the hosting website was sent to the respondents via email with a brief explanation of the purpose of the survey. In this way, participants can be informed that the questionnaire is available online. Using a web site can then simplify and automate the collection of data and monitoring of the progress of the respondents in completing the survey. This is important given the project time constraints and the need to speedily reach groups of respondents in different locations (Williman 2005:289).

Weekly follow-up emails were then sent to the respondents to encourage them to complete the survey before the deadline. At the conclusion, respondents received emails thanking them and informing them that the survey period had expired.

The following paragraph reports on the initial findings of the survey.

4 SURVEY FINDINGS

A total of two hundred and twenty invitations were sent out to individuals within the four banks and forty useable responses were received. This is an eighteen percent response rate and deemed sufficient for the purposes of investigative research (Educational Benchmarking Knowledge Base, 2005). Following are some of the findings:

4.1 Background

The purpose of this section was to capture the background of the respondent, including limited biographical data. Analysis of the data shows that 70% of the participants are males, suggesting that this might be a male dominated industry. Furthermore, that 65% of the respondents are above the age of 30 while the rest are between the ages of 21 and 30.

The distribution in age differences will allow for the capturing of views from different generations. Further analysis of these results will yield useful information on whether or not the age or the respondent affects opinions on the use of new technologies such as biometrics.

On ethnicity, figure 1 shows an evenly spread distribution representative of the South African ethnic population.

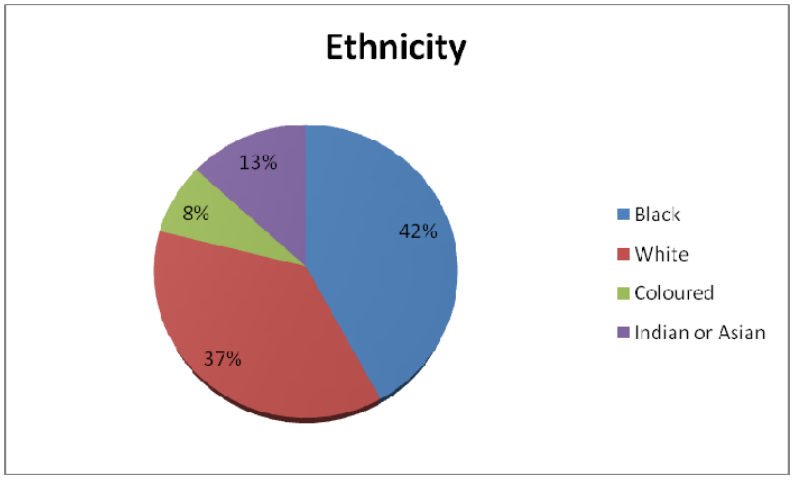


Figure 1. Ethnicity of respondents

Further analysis of this data will show if ethnicity has any impact on the acceptance of biometric technology.

4.2 General Knowledge of Biometrics

This section was aimed at establishing the awareness and experience of respondents with biometrics.

Findings show that 22.5% of respondents had never used biometric technology before. Across the different types of biometric technologies available, the top three with which respondents are familiar are fingerprint (26%), Voice/Speech (13%) and Signature (13%).

Half of the respondents indicated that they seldom use biometric technology. This data was further analysed to establish if it has any bearing on the respondent's confidence in the technology. Further findings show that despite this high percentage of respondents who seldom use biometric technology, 49% of the respondents agree that biometric technology can increase security in the Information Technology sector, while 41% strongly agree with this. Together this represents 89% of the respondents as shown in figure 2.

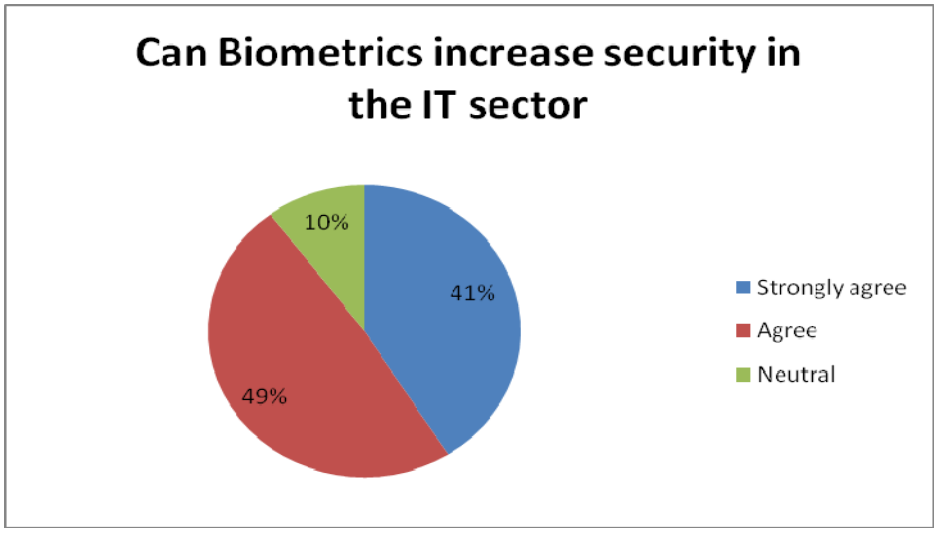


Figure 2. Relationship between biometrics and increased security

Further analysis of this data shows that a lack of exposure to biometric technology does not negatively affect personal views on its usefulness. This technology remains favoured as a possible solution to current authentication challenges.

4.3 Organisational Research

The aim of this section was to measure the participating banks' interest in biometric technology.

Results show that the investigation of biometric technology has exponentially grown in the last 12 months, when compared to the previous five years (figure 3). When comparing data relating to investigations conducted from the last 12 months to that of 24 months ago, analysis shows a growth of 34.2% in investigations into biometric technology.

Findings further show that the three most favoured technologies are still Fingerprint, Voice/Speech and Signature. This relates back to section 4.2 that shows that these are the same technologies that respondents have been exposed to before. The growth of interest in Palm scanning also increased steadily in the last five years.

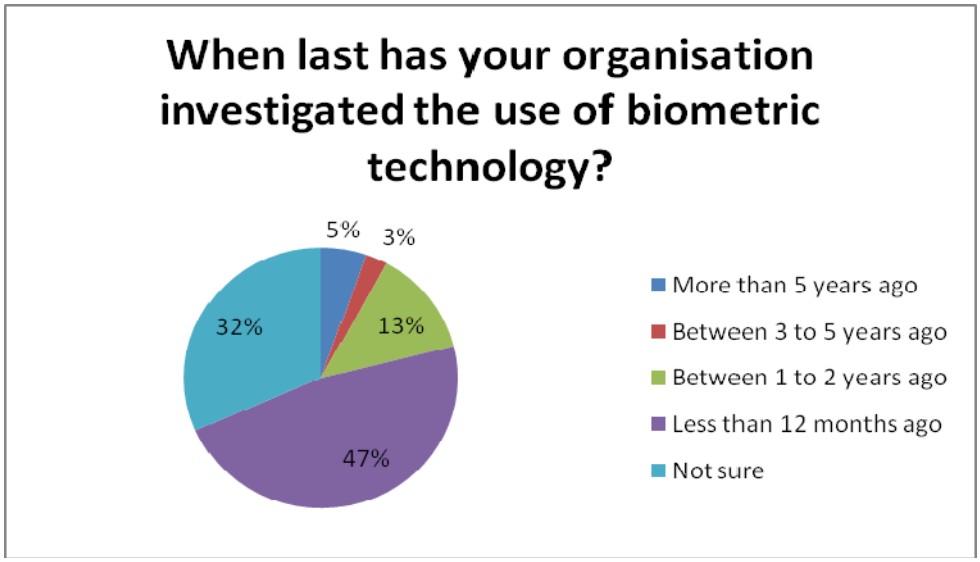


Figure 3. Investigation of biometric technology

Furthermore, 17% of respondents indicated that their organisations were not investigating biometrics. This data needs to be analysed further to establish the possible reasons for this.

4.4 Current Usage

This section was aimed at establishing if the organization is currently using biometric authentication.

Findings show that biometric technology is considered a solution for authentication by the majority of respondents. The areas where this technology is likely to be used include internet banking, telephone banking, branch network and community banking.

Favoured technologies for the future are still Fingerprint, Voice/Speech, Signature and Palm scanning as shown in figure 4. This relates back to sections 4.2 and 4.3.

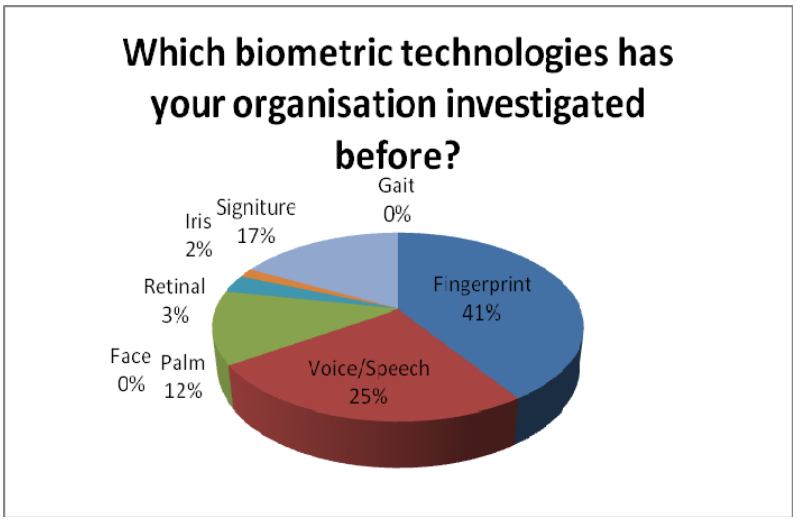


Figure 4. Investigation of biometric technologies

Further analysis is required to establish if these same technologies will be favoured for future use.

4.5 Perceptions

The aim of this section was to capture the perceptions and opinions of the respondents with regards to the future use of biometric authentication in their organization.

Findings show that biometrics is considered for use across different banking channels (figure 5).

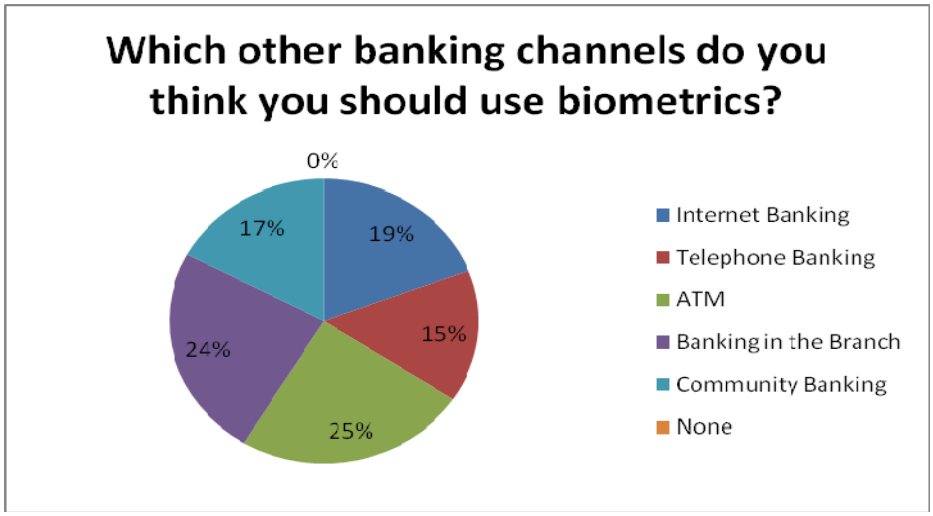


Figure 5. Banking channels that could benefit from biometrics

In the Internet Banking channel, Fingerprint, Signature and Voice/Speech scanning are seen as alternatives despite the technical barriers that could exist to deploy the suggested biometric technology.

For Telephone banking, Voice/Speech is seen as the biometric alternative, while for the ATM and Branch Networks, Fingerprint, Face, Retinal and Palm scanning are close favourites.

The Community banking channel shows great potential for the use of biometrics, with Fingerprint, Signature, Face, Retinal and Palm being suggested alternatives.

When it came to the issue of what other factors could be impacting the adoption of biometric technology by local banks, standards; bank legacy

systems; bank culture; and human cultural habits were seen as possible negative adoption factors. This is graphically illustrated in figure 6.

On the other hand, legislation and the maturity of biometric technology were not seen as negative factors to the adoption of biometric technology in the banking sector.

5 CONCLUSION

The aim of this empirical research was to capture the facts, opinions and perceptions of the respondents on the use of biometric technology and the factors influencing its adoption in order to formulate the problem more precisely, obtain insight and formulate a hypothesis. The initial results have confirmed the original problem statement and have provided current insight into the industry. The hypothesis that follows from this is, therefore, that the slow adoption is caused by a combination of several factors rather than the technology itself.

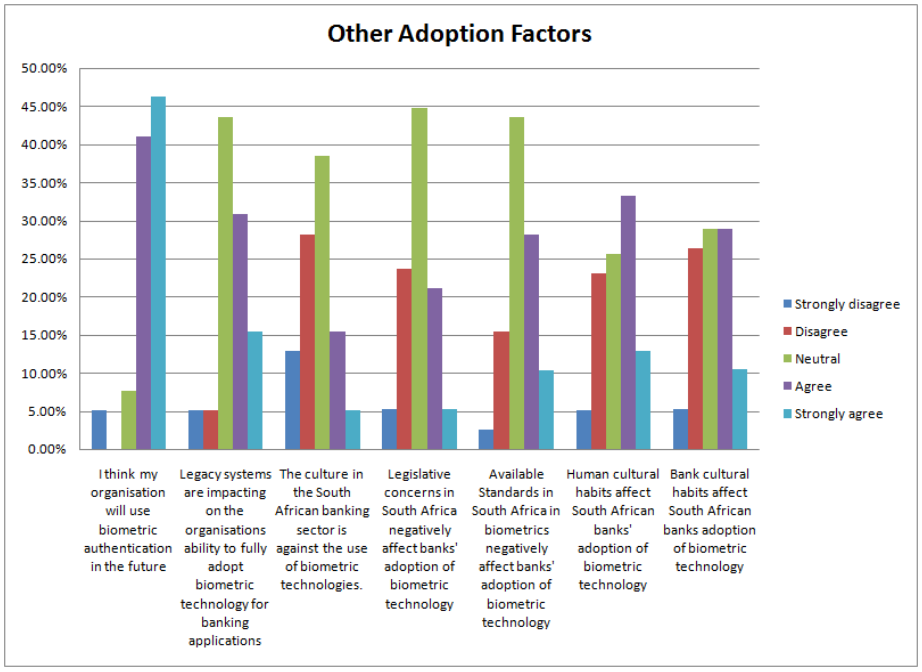


Figure 6. Factors impacting the adoption by SA banks

By means of the first section of the questionnaire, data gathered and analysed show that several bank employees have been exposed to biometric technology before. Though some appear to have never used this technology before, this does not affect their opinion on biometrics as a possible alternative to current security challenges in the banking sector.

It was also established that there is a definite interest in the use of biometric technology across different banking channels. Findings showed that the participating local banks have and are investigating biometric authentication and that these investigations were not limited to any particular biometric trait.

Perceptions and opinions of the respondents with regard to the future use of biometric authentication in their organizations were also successfully captured. These views provided an idea of the level of awareness and buy-in on biometric authentication and where the problem areas affecting adoption exist.

Future research includes further analysis of the data to determine various correlations. This will provide further insight into the problem of slow adoption.

6 REFERENCES

ABSA. 2008. *Absa in context* [Online]. South Africa: ABSA. Available from http://www.absa.co.za/absacoza/content.jsp?VGN_C_ID=d9615591b1a4ff00VgnVCM100000ce17040aRCRD&VGN_CI_ID=5b32515f3a2f1010VgnVCM100000ce17040aRCRD. [Accessed 27 February 2008].

Azari, R 2003. *Current security management and ethical issues of information technology.* IRM Press, London.

Devlin, A 2006. *Research Methods: Planning, Conducting nad Presentation Research*. Thompson Wasworth, USA.

Dawes, John 2008. *Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales*, International Journal of Market Research, 50 (1), 61-77.

Da Silva, I. 2007. *Internet banking fraud on the increase* [online]. South Africa: Marketing Community of South Africa. Available from:

http://www.bizcommunity.com/Article/196/16/19132.html. [Accessed 24 April 2008].

Devlin, A 2006. *Research Methods: Planning, Conducting and Presentation Research*. Thompson Wasworth, USA.

Educational Benchmarking Knowledge Base. 2005. *Determining an Acceptable Survey Response Rate* [Online]. Educational Benchmarking Knowledge Base. Available from: http://kb.webebi.com/article.aspx?id=10007&cNode=5K3B4O. [Accessed 24 March 2008].

FNB. 2008. *Company profile* [Online]. South Africa: FNB. Available from http://www.firstrand.co.za/default.asp?action=3. [Accessed 27 February 2008].

Garfinkel, S. 2005. *Authentication Battle* [Online]. Framingham: CSO Security and Risk. Available from http://www.csoonline.com/read/090105/authentication.html. [Accessed 19 December 2006].

Greenfield, T 2002. Research Methods for Past graduates, Arnold, London.

Krawczyk, S and Michaud, C. 2005. *Biometrics in the banking industry* [Online]. Michigan: Michigan State University. Available from: http://www.cse.msu.edu/~cse891/Sect601/CaseStudy/BiometricsBankingIndustry.pdf. [Accessed 24 April 2008].

Krutz, Ronald L and Vines, R 2003. *The CISSP Prep Guide*. Wiley, Indianapolis.

Layton, T P. 2007. Information Security: Design, Implementation, Measurement, and Compliance. Auerbach publications, Florida.

Nanavati, S, Thieme, M, Nanavati, R 2002. *Biometrics: Identity verification in a networked world.* John Wiley & Sons Inc, Canada.

Nedbank. 2008. *Nedbank group profile* [Online]. South Africa: Nedbank. Available from http://www.nedbankgroup.co.za/financials/nedbank_ar06/o/group_glance.asp. [Accessed 27 February 2008].

Standard Bank. 2008. *About us*[Online]. South Africa: Standard Bank. Available from http://www.standardbank.co.za/site/investor/aboutus_about.html. [Accessed 25 February 2008].

Skalak, S and Nestler, C. 2007. *Global economic crime survey 2007* [Online]. Germany: PriceWaterhouseCoopers. Available from: http://www.pwc.sport.hu/extweb/insights.nsf/docid/625C5CD467FC47768525736E0054D07A. [Accessed 25 April 2008].

Whitman, E, Mattord, H, 2006. *Reading and Cases in the management of information security*. Thompson Course Technology, Canada.

Williman, N 2005. *Your Research Project*. Sage Publications, London.

AN EVALUATION OF SCAN-DETECTION ALGORITHMS IN NETWORK INTRUSION DETECTION SYSTEMS

Richard J Barnett[1], Barry Irwin[2]

Rhodes University
Department of Computer Science
South Africa

[1]barnettrj@acm.org, [2]b.irwin@ru.ac.za

ABSTRACT

Network Intrusion Detection Systems are becoming more prevalent as devices to protect a network. However, the methods they use for some forms of detection are flawed. This paper builds upon existing research by van Riel and Irwin which illustrated these flaws in Snort and Bro's scan-detection engines. Indeed, it has been ascertained that a number of different scanning techniques are not identified by either Snort or Bro.

This paper highlights current research into the improvement of these scan-detection algorithms and presents insight into how this research is being conducted at Rhodes University. This research will improve on the scan-detection engines in Snort and Bro, permitting them to be used in a production environment without fear of succumbing to the false negative problem which currently exists.

KEY WORDS

Network Security, Intrusion Detection, Port scanning, Snort, Bro

Proceedings of ISSA 2008

AN EVALUATION OF SCAN-DETECTION ALGORITHMS IN NETWORK INTRUSION DETECTION SYSTEMS

1 INTRODUCTION

This paper describes current research being performed by the Security and Networks Research Group, in the Department of Computer Science at Rhodes University. It expands on research already performed in the Department in previous years, and in particular picks up on problems highlighted by van Riel and Irwin in [5].

Network Intrusion Detection Systems (NIDS) are more and more frequently becoming valuable aids to network administrators in the constant battle against attacks on current network centric computing. Indeed, whilst firewalls are now standard in the design of networks, network administrators need to know if and when an attack breaches that first line of defence. NIDS alert the administrator to any abnormal happenings inside a network. Most current NIDS rely on signature based detection on traffic flows through the network. The Open Source NIDS Snort [2] and Bro [1] use this method. [7]

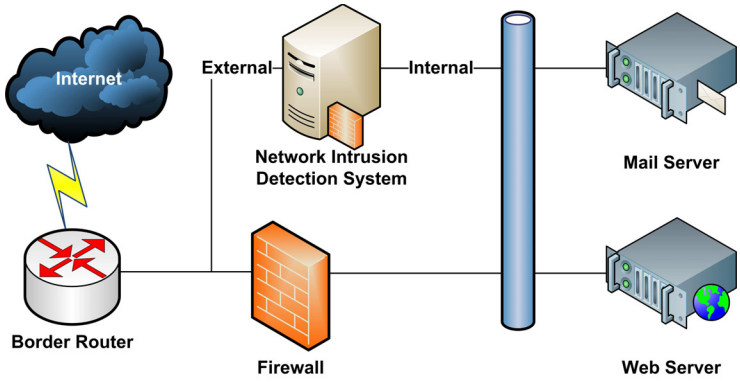
Figure 1 shows a typical placement of a NIDS inside a network. A NIDS would usually sit on the inside of the firewall and would sit on a span port of the local switch. This would give it the ability to view all traffic destined for all hosts on the network. In this case, it is illustrated by a web and mail server. The NIDS would be configured to apply rules targeting web and mail traffic.

Current NIDS can suffer from two major problems. False positives (which occur when a NIDS alerts on traffic which is benign) and false negatives (when a NIDS does not alert but an intrusion has occurred) are significant problems which can render a NIDS useless, either by wasting administrators time or by lulling them into a false sense of security.

Port scanning is a frequently used tool for identifying specific vulnerabilities in networked hosts, and is usually a precursor to further intrusion attempts. It is reported that the current (very large) volume of network attacks and specifically scanning activity may be "the tip of a very large iceberg" [14]. Despite the fact that most current NIDS make use of signature detection, both Snort and Bro have the additional capability to perform scan-detection.

van Riel and Irwin [5] (in the Department of Computer Science) have identified a number of flaws with the algorithms in both Snort and Bro. This was

Figure 1: A NIDS Inside a Network

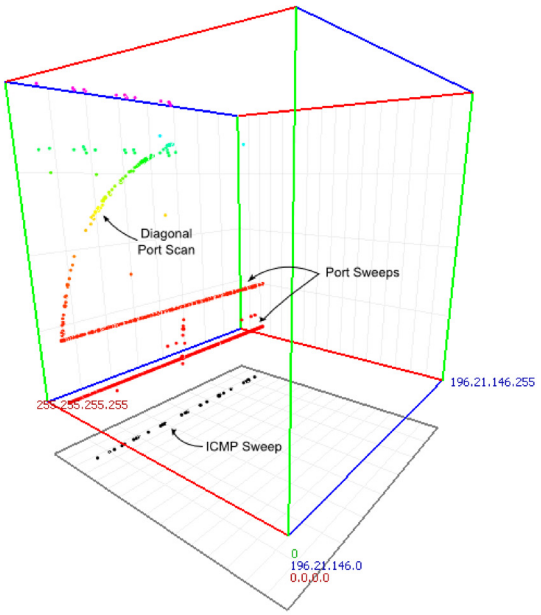


done by developing [13] and using a graphical tool developed for the analysis of network traffic, specifically traffic from network telescope captures. The tool, *InetVis*, presents traffic in a 3D space with a a time delay animation. These flaws present a variety of challenges and question the usefulness of the Snort and Bro algorithms. Figure 2 shows three different possible scans identified using *InetVis*. A horizontal black line with is an incomplete ICMP sweep, two red lines with represent port sweeps, and a curved multicoloured line which represents a diagonal scan. This permits scans to be rapidly identified in the telescope capture. These scans can then be isolated from the capture file with *InetVis* and then be replayed to Snort and Bro. This permitted van Riel and Irwin to identify the flaws in Snort and Bro with some efficiency.

New research is underway to investigate the possibility of extending the Snort and Bro algorithms to work in an efficient and effective manner. This research aims to increase the usefulness in deploying a NIDS and monitoring scans.

The remainder of this paper is structured as follows: Section 2 provides a fairly comprehensive look at some related work in the field of network intrusion detection and scan-detection. Section 3 will discuss how the authors plan on investigating improvements to the algorithms in Snort and Bro. Section 4 will investigate the outcomes that this research hopes to produce, and finally Section 5 will present the conclusions that can be drawn from the research performed so far.

Figure 2: Scanning Techniques Identified in InetVis



2 RELATED WORK

There is a wide variety of literature available in network intrusion detection and port scanning. This section looks at a selection of existing research which may be beneficial for the authors attempts at improving scan-detection in existing NIDS. This section is broken down into work related to port scanning (which will be presented in Section 2.1), work relating to statistical analysis of packets (in Section 2.2) and finally work related to anomaly detection in Section 2.3.

2.1 Port Scanning and Scan Detection

Port scanning can be identified by classifying both connection attempts and the hosts which make such attempts. Allman *et al.* [3] propose a method of classifying connections into three categories. "Good" connections, which are those which are successful. "Bad" connections, those that do not lead to established connections, and "unknown" connections, which are those which cannot be identified for any reason. Hosts can be classified into either "good" hosts, or "bad" hosts. Allman *et al.* observe that there could be numerous methods of performing such categorisation, but that a simple and effective

method is to classify hosts which make a majority of "good" connections as "good" and hosts which make a majority of "bad" connections as being "bad". From this, it is possible to classify scanning activity as traffic which is "bad" and originates from a "bad" host.

Through the use of *InetVis*, van Riel and Irwin [5] have identified a number of different scan types which include diagonal scanning, step scanning and what they refer to as a "creepy crawly" scan. Some of the scans types that have been identified are slow and long running, and exhibit timing and destination address and port selection intended to avoid observation.

2.2 Statistical Analysis in NIDS

A number of authors have investigated the possibility of introducing statistical analysis into NIDS to improve their reliability, whilst decreasing their overhead. Crotti *et al.* [4] propose the use of statistical fingerprinting to quickly identify the contents of a given packet. Their research suggests the use of *Probability Density Functions* to perform the fingerprinting. They do, however, note that one of the primary problems with such an approach is that the contents of packets are required to train the system and that most publicly available packet traces remove all useful application level information for security reasons.

A similar method is proposed by Karamcheti *et al.* [6], who propose the use of *inverse distributions* to classify packets. This method relies on separating each packet into a number of sub-strings and performs comparisons on these sub-strings against known traffic samples. The relationship between the two can then be fitted to the *inverse distribution* to determine the nature of the packet. As with Crotti *et al.*, this solution suffers from the need to have full packet traces to perform seeding.

However, this use of full statistical fingerprinting can, itself, be a limiting factor for NIDS. Ramaswamy *et al.* [11] propose an alternative method. Their method uses *approximate* fingerprinting which makes use of a sliding window across the packet contents to fingerprint the contents. This method may produce false positives, but will never produce false negatives. Therefore, matching packets can then be analysed in more detail making use of more traditional methods.

These methods all require the use of normalised traffic, Rubin *et al.* [12] propose a very sophisticated algorithmic approach which they call *protomatching* to make a single pass over unnormalised network traffic. Because of this, this method can perform well in high throughput network environments.

2.3 Anomaly Detection

Despite the usefulness of each of the approaches discussed above, they all focus quite extensively on signature based NIDS, and look at the contents of a given packet in relation to a rule. Kompella *et al.* [8] and Krügal *et al.* [9] both discuss alternative methods for performing anomaly detection.

Firstly, Krügal *et al.* [9] discuss a method of performing service specific anomaly detection making use of a two part system including a packet processing unit and a statistical processing unit. The packet processing unit performs stream reassembly and other normalisation tasks, and could easily be a NIDS such as Snort or Bro. The statistical engine performs the anomaly detection and can be a separate application, or could be integrated into a NIDS as a plug-in. The statistical analysis proposed by Krügal *et al.* includes analysis on the type of request, the length of the request and the payload distribution of the packet.

The alternative proposed by Kompella *et al.* [8] suggest the use of an "intelligent" data structure which is called a *Partial Completion Filter (PCF).* This is suggested as a possible method to perform scan-detection (amongst other uses) and is designed to be scalable as it uses a largely fixed amount of memory, as it stores only a count of seen packets which match against a hash function. By having several hash functions and counters and a trigger condition on each counter, it is possible to build a sophisticated *PCF* which scales and targets a number of possible scan areas.

There are two traditional approaches to performing scan-detection [8]. The first being that a number of events during a given interval are counted, the second that the number of failed connections in an interval are counted. This corresponds well with work done by Levchenko *et al.* [10] who make a similar assertion. Further to that, however, Levchenko *et al.* discuss the assumption that port scanning requires per-flow state to be stored, which does (as has been seen) not scale effectively. Their paper proves mathematically that ingress detection of port scanning cannot be performed without maintaining state. They do, however, prove that egress detection of scanning activity does not rely on the presence of state tables.

In this context, Levchenko *et al.* define ingress detection as that which looks at incoming connections, and egress detection as that which identifies the TCP RST packets which are transmitted when a connection to a closed port is attempted. (Or the ICMP "Port Unreachable" packet which is sent for UDP connection attempts.) This method suffers from the possibility that host-layer firewalls may prevent such packets from being transmitted.

3 RESEARCH APPROACHES

Initially, the authors intend on verifying the effectiveness of Snort and Bro's scan-detection algorithms in alerting on a variety of scanning techniques. Using the graphical tool *InetVis* [13], the current flaws in the Snort and Bro algorithms will be verified. The results of this process will be used in conjunction with detailed packet analysis and statistical processing, to develop techniques for detecting such scans.

The most common scans, port scans and sweeps are well defined and in the simple case are easy to identify. This is also the case with the related techniques ICMP scanning and other host reachability scanning. Pseudo-random phenomena, which could be attributed to backscatter and the results of network configuration errors, can also be the result of significantly more complex scanning techniques which are more difficult to identify.

A number of a constraints involved in the processing of network traffic are anticipated. These constraints involve the physical system resources required in looking for components of scans, specifically the memory required to detect long, slow running scans and the processing power required to scan large volumes of traffic. Pseudo-random scans are particularly difficult to isolate with limited resources, as a lot of memory is required within a non-trivial temporal reference frame. Experiments on the performance of each of the developed methods when provided with large volumes of data - as would be the case on a high throughput link - will be investigated. This experimentation will permit the authors to determine which scan-detection techniques will scale and which will not.

This will permit algorithms to be developed which solve the false negative problem and which perform in a scalable manner. These methods are expected to involve statistical analysis, multidimensional matrix operations and projections. The integration of these methods and techniques into Snort and Bro is the ultimate outcome of this research as discussed in the next Section.

4 RESEARCH OUTCOMES

This research plans to take validated results from prior research at Rhodes University to identify problems with current scan-detection algorithms in use in common NIDS. Having identified and validated the problems in existing algorithms, the authors aim to adapt and improve these algorithms.

These algorithms will be used to develop new plug-ins for both Snort and

Bro. Whilst the authors intend on producing plug-ins for just two NIDS, the result could be easily modified to produce plug-ins for any system.

As the architectures of Snort and Bro are quite different, the process which will be taken to build plug-ins for them will also be somewhat different. The Snort plug-in is best suited to being developed as a C++ shared object which will act as a Snort Preprocessor. This can be integrated into the Snort pipeline in a suitable place to allow for maximum performance [7]. Bro has a much more flexible customisation system. The use of the Bro language will be considered and used if possible. If a more complex solution is required a Bro *analyser* is an alternative which would be developed in C++ [1].

In addition to the Snort and Bro plug-ins, the authors intend on producing a hardware accelerated version of the plug-ins which can be deployed as an alternative in environments where high volume networking prevents the CPU from efficiently scanning all traffic.

5 CONCLUSIONS

This paper has illustrated current research into the improvement of scan-detection at Rhodes University. The research is still in its infancy, but is anticipated to improve the state of scan-detection engines in NIDS when complete. The authors intend on developing new techniques for scan-detection based on the existing Snort and Bro plug-ins, and a number of other methods. A variety of methods are being investigated and are likely to include a significant statistical analysis component. Existing research suggests this is a positive approach.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support of their colleagues in the Department of Computer Science. We would also like to acknowledge the support of Telkom SA, Business Connexion, Comverse SA, Stortech, Tellabs, Amatole, Mars Technologies, openVOICE and THRIP through the Telkom Centre of Excellence in the Department of Computer Science at Rhodes University. Finally, the authors would also like to acknowledge the support of the National Research Foundation.

REFERENCES

[1] Bro intrusion detection system - bro overview. Online: `http://www.bro-ids.org/`, Accessed: 28/01/2008.

[2] Snort - the de facto standard for intrusion detection/prevention. Online: `http://www.snort.org/`, Accessed: 28/01/2008.

[3] Allman, M., Paxson, V., and Terrell, J. A brief history of scanning. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2007), ACM, pp. 77–82.

[4] Crotti, M., Dusi, M., Gringoli, F., and Salgarelli, L. Traffic classification through simple statistical fingerprinting. *SIGCOMM Comput. Commun. Rev. 37*, 1 (2007), 5–16.

[5] Irwin, B., and van Riel, J.-P. Inetvis: a graphical aid for the detection and visualisation of network scans. In *Conference on Vizualisation Security (VizSec2007)* (2007).

[6] Karamcheti, V., Geiger, D., Kedem, Z., and Muthukrishnan, S. Detecting malicious network traffic using inverse distributions of packet contents. In *MineNet '05: Proceedings of the 2005 ACM SIGCOMM workshop on Mining network data* (New York, NY, USA, 2005), ACM, pp. 165–170.

[7] Kohlenberg, T., Alder, R., Dr. Everett F.Carter, J., Foster, J. C., Jonkman, M., Marty, R., and Poor, M. *Snort Intrusion Detection and Prevention Toolkit*. Syngress Publishing Inc., 2007.

[8] Kompella, R. R., Singh, S., and Varghese, G. On scalable attack detection in the network. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2004), ACM, pp. 187–200.

[9] Krügel, C., Toth, T., and Kirda, E. Service specific anomaly detection for network intrusion detection. In *SAC '02: Proceedings of the 2002 ACM symposium on Applied computing* (New York, NY, USA, 2002), ACM, pp. 201–208.

[10] Levchenko, K., Paturi, R., and Varghese, G. On the difficulty of scalably detecting network attacks. In *CCS '04: Proceedings of the 11th ACM conference on Computer and communications security* (New York, NY, USA, 2004), ACM, pp. 12–20.

[11] Ramaswamy, R., Kencl, L., and Iannaccone, G. Approximate fingerprinting to accelerate pattern matching. In *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2006), ACM, pp. 301–306.

[12] Rubin, S., Jha, S., and Miller, B. P. Protomatching network traffic for high throughputnetwork intrusion detection. In *CCS '06: Proceedings of the 13th ACM conference on Computer and communications security* (New York, NY, USA, 2006), ACM, pp. 47–58.

[13] van Riel, J.-P., and Irwin, B. Inetvis, a visual tool for network telescope traffic analysis. In *Afrigaph '06: Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa* (New York, NY, USA, 2006), ACM, pp. 85–89.

[14] Yegneswaran, V., Barford, P., and Ullrich, J. Internet intrusions: global characteristics and prevalence. In *SIGMETRICS '03: Proceedings of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems* (New York, NY, USA, 2003), ACM, pp. 138–147.

ENABLING USER PARTICIPATION IN WEB-BASED INFORMATION SECURITY EDUCATION.

Ryan Goss[1], Johan van Niekerk[2]

[1]**Nelson Mandela Metropolitan University**
South Africa
[2]**Nelson Mandela Metropolitan University**
South Africa

[1]**ryan@goss.co.za,** [2]**johan.vanniekerk@nmmu.ac.za**

ABSTRACT

The greatest threat to Information Security are the employees within an organization. Many security controls rely on the user in order to be effective. It is thus vital to educate users about their role(s) in security. Many companies cannot afford, in terms of time or finances, to replace employees during training periods. The Web has long since been identified as a viable alternative to traditional training. To a certain extent, using the Web as a training platform depends on user buy-in. Web 2.0, which involves users and is largely user driven, is one way in which such buy-in could be obtained. This paper will discuss both the data acquisition and storage of Information Security principles to a centralized knowledge store, from which Web based Security Education technologies can draw inference. These web based security education applications should involve the user, thereby securing their buy-in and adding to the overall effectiveness of the training program. The use of Resource Definition Framework (RDF), SPARQL Protocol And RDF Query Language (SPARQL) and the Semantic Web will be discussed as possible solutions to the storage and transport of represented knowledge between multiple systems.

KEY WORDS

Information Security, Information Security Education, Web 2.0, Semantic Web

ENABLING USER PARTICIPATION IN WEB-BASED INFORMATION SECURITY EDUCATION.

1 INTRODUCTION

Information Systems have become a crucial tool to the success of an organization and thus need to be protected. At the same time these systems should provide adequate access to the information for the users within the organization. The protection of information resources is also known as information security, and is often described as the CIA, or Confidentiality, Integrity and Availability triangle. These three objectives fall in line with the fundamental goals of Information Communication Technology (ICT) security (Kruger, Drevin, & Steyn, 2006). Federal agencies or even Information Technology (IT) administrators cannot protect the integrity, confidentiality or availability of information in today's highly networked systems without first ensuring that each and every user of the system is aware and acting on their responsibilities within the information system (NIST 800-16, 1998). Mitnick and Simon (2002) argues that the greatest threat to Information Security are the users within an organization. In fact, this is so true that the Computer Security act of 1987 (Public law [P.L.] 100-235) required that each user within a federal agency be subjected to periodic training in both security awareness and computer best practises (NIST 800-16, 1998). The document further stipulates that these requirements include all users, from upper management to standard employees and even anyone involved within the operation of a federal computer system within the agency.

Potentially the most difficult part of security education process is ensuring user buy-in. The presentation aspect of an educational system should therefore be on-going, creative, motivational, eye-catching and intuitive, with the objective of focusing the learners attention so that the learning will be incorporated into conscious decision making (NIST 800-16, 1998). Another potential problem is that in the modern, competitive world, organizations cannot afford to take their users out of the operations of the organization for sustained periods of time without the company suffering. For sometime now, Hypermedia (Web-based) education systems have stepped up to the

plate as a solution to this problem. These systems have become known as E-Learning systems. Adaptive e-learning branched off from static, e-learning systems and provided an alternative to the "one-size fits all" e-learning scenario by allowing the system to evolve its interface and the content displayed by learning from interrogating the user and building a user model. Web-base Training (WBT), the delivery of instruction or learning content over the Internet and/or an organization's intranet is fast becoming popular amongst organizations (Lee, Chamers, & Ely, 2005). Lee et al. (2005) argues that motivation is a key element in the user acceptance of a training platform. Without user acceptance, the educational platform and the content it attempts to deliver will become nothing more than another neglected, eye-catching, yet useless lesson (Lee et al., 2005).

This paper argues that the use of recent developments in technology, such as Web 2.0 and the semantic web, make it possible to keep the user involved and motivated throughout a training program. This increased attention could enable users to better learn both simple and complex information security principles over mediums such as the World Wide Web or the organizational intranet. This paper also discusses methodologies which can be implemented to share stored knowledge, about the user profile and the information security principles being taught, amongst such systems.

2 RESEARCH PARADIGM AND RATIONALE

The purpose of this phenomenological study is to highlight the importance of the role that the user plays in information security within an organization, and to present methods the organization could employ to strengthen this "human factor". The paper is presented using argumentation theory as discussed in Van Eemeren (2001). This theory is concerned with the arts and sciences of civil debate, communication and persuasion. The paper does not necessarily cover new concepts, but rather serves to highlight various pre-existing technologies and how they are employed together to work towards the development of a successful, user-driven, Information Security Education platform.

As far as could be determined by the author, the use of Web 2.0 based

technologies coupled with the successful data sharing process for inter-educational knowledge base access (Web 3.0) for information security education systems has yet to be published. It is the author's belief that the sharing of such information by utilizing a de-centralized, generic knowledge base hosting data pertaining to information security principles and the user's profile, from which various educational tools draw inference would be a large asset in the struggle towards educating users. Creating interconnected, multi-platform compatible knowledge base access mediums will greatly aid in the strengthening of the "human factor" within information security.

The aim of this paper is therefore to show that Web 2.0, its knowledge representation storage mechanisms and transfer of this knowledge base between multiple systems is firstly possible and secondly will aid in the strengthening of the human factor within an information security environment.

3 WEB-BASED EDUCATION SYSTEMS

Hypermedia or Web-based educational systems, as mentioned previously, is by no means a technology in its infancy. In fact, ever since the establishment of the World Wide Web, scientists and scholars have been using the medium to promote information in static form to users of various web sites. Hypermedia offers a multimedia information environment, supports non-linear access to information, and provide a means of interaction with the user, all at the same time integrating the various information formats into a common display (Liaw, 2001). The rapid growth and development of the World Wide Web has been the main driving factor in the rapid migration of educational systems to hypermedia based applications (Liaw, 2001). Liaw (2001) continues to state that some of the potential benefits of hypermedia based applications would include: allowing the learner to structure their learning approach, the ability to pursue cross-references and to "remember" various aspects of the learning session.

Based on human cognition, computer assisted learning environments such as Hypermedia, are based on constructivist learning theory. Variations of this theory include social constructivism, which focuses more on the social context of learning as well as "cognitive constructivism" which which states that learners construct their own knowledge of the world through assimilation and accommodation (Liaw, 2001). Constructivism learning theory's educational

ideology is based on the learner constructing their own knowledge. This knowledge may be constructed through discovery, exploration and investigation (Cook, 2006). The teacher within a constructivist learning environment should structure the learning process so that they become a "co-constructor" of the knowledge being constructed by the learner, thus forming a partnership between both the student and the teacher (Cook, 2006).

In order for web-based systems to accommodate such learning processes, they are required to adapt to the learner, their specific needs for constructing knowledge, as well as the method of presentation of such knowledge for the learner to review. Adaptive e-learning has been around for some time now and addresses this very need. Adaptive e-learning systems can be broken into two main parts: adaptive content generation and adaptive interface design or presentation. Adaptive content generation is concerned with what content to show the learner; the learner should not necessarily be shown content that they are already familiar with. Adaptive presentation involves the user interface adapting to the preferences of the particular user, so as to avoid the heterogeneous "one size fits all" approach to education. The National Institute of Standards and Technology (NIST) IT Security Training requirements document requires that security awareness and training presentations should be designed with recognition that users practice *acclimation* or a tendency to tune-out if the stimulus or "attention-getter" is used repeatedly. Presentations should therefore be ongoing, creative and motivational with focus on the user to consciously start incorporating new knowledge into their existing behavioural pattern by way of assimilation (NIST 800-16, 1998). Adaptive hypermedia systems are perfectly aligned to allow for this constant changing presentation to occur and to assist the educational system in firstly providing the correct knowledge whilst at the same time keeping the user's attention and adapting to their individual learning style.

It is essential that adaptive e-learning systems collect and model user information so as to allow for the system to adapt to the user's characteristics and preferences (Froschl, 2005). An adaptive e-learning system should also have a strong knowledge base, from which the system draws inference. The user model is compared against this knowledge base or "domain model" and it is from this comparison that similarities are drawn and progress of the

learner is quantified. In order to build an Information Security educational adaptive e-learning suite, an extensive and accurate knowledge base is required containing various principles from within the subject domain.

New movements such as Web 2.0 have recently come to light in the struggle to keep the user involved in the training program, thereby ensuring their buy-in and allowing them to effectively participate in the information security training exercise. This participation includes both the the learning from existing information, as well as contribution of their own ontologies pertaining to particular information security principles.

4 WEB 2.0 BASED SYSTEMS

Web 2.0 is a term coined in the first O'Reilly Media Web 2.0 Conference in 2004. It is loosely defined as a business revolution within the computer industry caused by the movement to the Internet as a platform and designed to harness collective intelligence (Needleman, 2007). Web 2.0 is not a technology, but rather a way of thinking whereby users generate content which is published, used and managed through network applications in service-orientated architecture (Judicibus, 2008). Web 2.0 enabled websites also boast a host of advantages over standard "Web 1.0" websites. These include:

- *The user as a contributor*: The user is encouraged to participate in book reviews, commenting on articles, uploading multimedia such as photographs etc. Acting on what was previously discussed, this aids in the necessity to involve the user, thereby ensuring their attention whilst using the system.
- *Trust and collaboration*: Services such as wikipedia which are based on the concept that any user can add an entry and any other user can edit it (Needleman, 2007).
- *Multi-platform applications, above the level of any single device*: The World wide web provided a platform for content delivery over multiple devices. Web 2.0 takes this one step further with mobile devices

contacting remote servers, using the PC as a docking station and local cache during the transaction (Needleman, 2007).

- *Cost Reductions*: Not only are Web 2.0 applications relatively inexpensive to deploy, but in most cases Web 2.0 extensions can be added to non-Web 2.0 products to further reduce costs. For example, wikis could be deployed for users to build up knowledge bases and documentation with relatively little investment from the organization (Zambonini, 2006).

A web based system which actively involves the user as both a contributor and a casual browser could solve many of the obstacles faced by existing educational systems. Information Security Education does not always hold the interest of the users who are to take the courses and therefore whatever can be done to aid in the stimulation of the user and therefore the learning experience would be a huge asset to the training program. Many web-based information security education or awareness systems exist, however these systems operate within the confines of closed environments. One of the major downfalls of Web 2.0 technologies is in their inability to store information in a computer-readable format and therefore data-acquisition and sharing amongst various Web 2.0 websites is hindered. Whilst the majority of Web 2.0 applications typically provide some form of proprietary Application Program Interfaces (API) access to their underlying knowledge store, in order for a remote application to access this knowledge, the accessing application should have extensive parsing ability for the remote API set, with programs often traversing large eXtensible Markup Language (XML) trees to recover the required data (Heath & Motta, 2007). A storage and transport medium needs to be identified which will solve the problem of data storage, facilitating the interoperability of many of these potential Web 2.0 learning environments, thereby allowing the user access to a wealth of information and training material, all from a single website. One such technology, proposed by the World Wide Web Consortium (W3C) is already gaining wide acceptance - namely the Semantic Web.

5 SEMANTIC WEB AS A KNOWLEDGE TRANSPORT AND STORAGE SYSTEM

Breners-Lee (1998) described the Web as being an information space, whos goal is to be useful not only for human to human communication, but also

that machines would be able to participate and help. Breners-Lee (1998) further discusses that one of the major obstacles has been that information on the web has in the past been designed for human consumption and even if the data was represented in a technically sound manner, the structure of such representation would not be evident to a robot browsing the web. One of the core goals of the semantic web is to bring progressively more meaning to the information published on the web (Java et al., 2007). The semantic web encapsulates information with a collection of metadata which describes this information. Using standardized query languages such as RDF and Web Ontology Language (OWL) allows machines and human readers alike access to the information. The machine readers have access to the underlying metadata, whilst for the human readers, this information is masked so as to hide the underlying architecture and merely provide the information requested. Machines being exposed to the metadata will benefit from the deep semantic annotations in their application-orientated task processing (Java et al., 2007).

The semantic web therefore provides a near perfect platform for the development of shareable knowledge models on particular problem domains for the construction of knowledge base systems on an open environment such as the Internet (Chan, 2007). Such knowledge base systems enable semantic web agents to draw inference in common formats, thereby allowing for the ease of distribution and querying of remote knowledge stores without having to locally store the data. The various engines require a common protocol for data acquisition and transfer. Some examples of such protocols are RDF and SPARQL. The exact operation of these protocols is beyond the scope of this paper.

In order for the semantic web to facilitate the process of knowledge storage and retrieval for Information Security Educational applications, a suitable front end environment needs to be created in order for the user to be able to contribute to the knowledge store. One of the greatest downfalls of the Semantic Web is the lack of intuitive interface design for creating, modifying and querying data within the grid.

This system should be able to translate the information from the user to a semantic web based format (such as RDF), thereby enabling remote user applications to share in the accumulated ontology. One such front end has already been discussed : Web 2.0. The problem therefore becomes whether Web 2.0 and Semantic Web technologies can co-exist in order to promote user involvement and support within an Information Security Education System, thereby attending to both the presentation and data retrieval aspects of a successful adaptive e-learning system.

6 WEB 2.0 AND THE SEMANTIC WEB

Web 2.0 has aided in the contribution of an unprecedented volume of knowledge to the World Wide Web, through simple yet engaging interfaces, allowing the user to contribute to a vast number of subject domains (Heath & Motta, 2007). Heath and Motta (2007) continues to describe these heterogeneous knowledge stores as using techniques that do not facilitate the scaling beyond a handful of sources. The semantic web on the other hand provides the key to large-scale data integration, yet lacks the interactive user interfaces necessary to allow for contributions by non-specialists (Heath & Motta, 2007). The perfect hypermedia educational system should therefore provide an interface using Web 2.0 technologies, yet store the knowledge acquired in RDF data sets, ready to be shared via an underlying semantic web. This provision for contribution of knowledge by users would aid in the development of Information Security Education systems whereby experts contribute knowledge which would span world wide for various other educational systems to access. The following two sub-issues deserves special attention:

1. *Contributor Credibility*
   All users contributing to this wealth of knowledge should be rated to ensure the validity of such data. As the proposed educational system will be based on the semantic web, an RDF or Friend Of A Friend (FOAF) object would be built for each user and other users could rate this user, increasing their credibility or score on the system. A user with a high score could be said to be credible, conversely one with a low score would be deemed an amateur whos contributions should be questioned or confirmed by a higher ranking contributor. Harnessing the power of the semantic web, a particular user may already maintain

an existing RDF describing themselves on a remote site also supporting semantic web standard query languages. In this case, a user may link their profile to the remote FOAF object Unique Resource Identifier (URI), thereby allowing the Information Security Education system access to additional information about the user, such as qualifications, employment details, location or whatever the user has decided to publish in their RDF object. This remote access ability allows the base system to capture only minimal information about the user onto its local data store, encouraging the user to rather link to an alternate URI for the enhancement of their profile.

2. *Tagging, not classifying*
Heath and Motta (2007) describes a method of tagging Web 2.0 data, now encapsulated in RDF format, instead of requiring the user to link the new knowledge under a particular heading or category. This ensures ease of contribution by the user, since the information supplied no longer needs to be fixed within the confines of a particular category. Knowledge which may not easily be classified is now easily tagged and stored in the underlying RDF database (Heath & Motta, 2007). Data about tags associated with particular Information Security principles would be described using the Tag Ontology and published on the website in Hyper Text Markup Language (HTML) (for human readers) and via the website's SPARQL endpoint, enabling machines access to the knowledge store. Each Information Security Principle is able to be tagged numerous times, thereby allowing web searches more accuracy whilst querying the knowledge store. Having tags also allows for related principles to be displayed to the user whilst they browse the site or provides a semantic pathway to machines traversing the data.

   As the website learning environment would be Web 2.0 powered, each Information Security Principle would have a section where users could post their views on the principle, argue for or against its validity and generate a discussion around the subject area. These discussions would too be published in RDF format to the underlying semantic web.

   As each user on the website has a FOAF object, stored on the central server, any Information Security application drawing inference from this knowledge store is able to keep track of the progress of the user, using user modeling techniques. These user models can then be con-

verted to RDF and linked against the FOAF object for a particular user. The ability to track the progress of the education of the user ensures feedback to organizations pushing for user training in the field of Information Security, thereby strengthening the human factor. Organizations could flag certain tags within the system and ensure that their users complete all training related to these tags. The system will keep track and quantify the results of the training and give detailed reports back to the organization as to the understanding of the employee.

Future information security education systems, which incorporates concepts from both Web 2.0 and the Semantic Web should thus, for best results, exhibit the following characteristics:

- An intuitive, user-involved and morphing interface
- A common knowledge storage format
- A common interface for querying stored knowledge
- Knowledge and data contributions by users of the site
- A simplistic classification of submitted user information

From the above it should be clear that the Web 2.0 philosophy is well suited to use for interface and interaction design in information security educational systems. Similarly, the Semantic Web would be an appropriate methodology for common data (knowledge) storage and querying, using the Tag Ontology for classification of the data. The way forward is therefore quite clear - toward a third generation, Web 3.0, educational system where various *Mashups* are able to interconnect and share ontologies and knowledge stores, enabling better access to the knowledge and a more customized system for all users. Web 3.0 based educational systems should focus on the backend transports and storage layers, rather than primarily the front end as has been the case on the Web as of late. Nova Spivak of Radar Networks describes Web 3.0 as the next big step in Internet development which is still in its infancy and should be mainstream as of 2010. Spivak is ambitious in his discussion as to what follows the Web 3.0 era - Web 4.0. Spivak finishes by stating that in the world of Web 4.0, users will benefit from distributed searches, intelligent personal agents, semantic databases etc truely working towards 'The WebOS'. Future research regarding Information Security

Education using the latest Web technologies could include an investigation into the movement toward distributed searches, a discussion of the technical storage mechanisms for converting Web 2.0 input into RDF format for use within the semantic web and the underlying technical implementation for communication on the network.

7 CONCLUSION

This paper introduced the idea of implementing a hypermedia Information Security Education platform, powered by Web 2.0 and Semantic web technologies to get the best of user interaction and knowledge base sharing across multiple systems - giving rise to a Web 3.0 training platform.

It was argued that by using Web 2.0, non-specialists could generate semantic annotations suitable for use within a semantic web. The use of Web 2.0 ensures user buy-in to the training experience, thereby keeping their attention and educating them in the various Information Security Principles.

It was further argued that these principles could be captured by any user, however the credibility of such input would be based on a scoring facility, indicated by the credibility and acceptability of each contributor. The comments facility was also discussed to initiate inter-user communication and arguing, ensuring a better understanding of each principle, rather than a blind acceptance of it. This aids in the embedding of such knowledge in the day to day actions of the user, greatly adding to the effectiveness of security awareness campaigns and overall organizational security practises.

Whilst this paper does not, at a technical level, provide a solution on how to use Web 2.0 to enable user-participation in information education, it does show that such an approach is definitely possible. Future efforts in line with this research will be aimed at delivering the more technical hands-on parts of this solution. It should also be clear that the idea of ensuring user buy-in into security education programmes by implementing a Web 2.0 interface still needs to be tested empirically.

References

Breners-Lee, T. (1998). Semantic web roadmap. *[WWW document]. URL http://www.w3.org/DesignIssues/Semantic.html, Sited 2 June 2008.*.

Chan, C. (2007). Development of an ontology for an industrial domain. *Intl Journal of Cognitive Informatics and Natural Intelligence*, *1*(3).

Cook, P. (2006). The project approach: An appreciation for the constructivist theory. *Published by the Forum on Public Policy*.

Froschl, C. (2005). *User modeling and user profiling in adaptive e-learning systems*. Unpublished master's thesis, Graz University, Austria.

Heath, T., & Motta, E. (2007). Ease of interaction plus ease of integration: Combining web 2.0 and the semantic web in a reviewing site. *Web Semantics: Science, Services and Agents on the World Wide Web*.

Java, A., Nirneburg, S., McShane, M., Finin, T., English, J., & Joshi, A. (2007). Using a natural language understanding system to generate semantic web content. *Intl Journal on Semantic Web and Information Systems*, *3*(4).

Judicibus, D. de. (2008). World 2.0. *[WWW document]. URL http://lindipendente.splinder.com/post/15354690/World+2.0., Sited 23 April 2008.*.

Kruger, H., Drevin, L., & Steyn, T. (2006). A framework for evaluating ict security. *Information Security South Africa Conference*.

Lee, D., Chamers, T., & Ely, T. (2005). Web-based training in corporations: design issues. *Intl Journal of Instructional Media*, *32*(1).

Liaw, S. (2001). Designing the hypermedia-based learning environment. *Intl Journal of Instructional Media*, *28*(1).

Mitnick, K., & Simon, W. (2002). *The art of deception: Controlling the human element of security*. Wiley Publishing.

Needleman, M. (2007). Web 2.0 and lib 2.0 - what is it? (if its anything at all). *Serials Review*.

NIST 800-16: Information Technology Security Training Requirements: A Role- and Performance-Based Model. NIST Special Publication 800-16, National Institute of Standards and Technology. (1998).

Van Eemeren, F. (2001). Crucial concepts in argumentation theory. *Amsterdam University Press*.

Zambonini, D. (2006). Why you should let web 2.0 into your hearts. .

8 ACKNOWLEDGEMENTS

The financial assistance of National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the National Research Foundation.

VISUAL CORRELATION IN THE CONTEXT OF POST-MORTEM ANALYSIS

Michael Hayoz and Ulrich Ultes-Nitsche

Research group on telecommunications, networks & security
Department of Informatics, University of Fribourg, Switzerland

{michael.hayoz | uun}@unifr.ch, Bd de Pérolles 90, CH-1700 Fribourg

ABSTRACT

One of the biggest challenges in the field of digital forensics lies in the ability to bring all potential evidence into a chronological correlation, and to draw appropriate conclusions to allow plausible and reproducible chains of activity. The ever-growing size of storage devices results in a considerable amount of information to process. Analysing forensic data, as a general rule, is a time-critical process. The output of current forensic tools mostly has the form of exhaustive lists and tables and can be difficult to manage and interpret. These constraints leave forensic specialists with a need for improvement in the way they handle huge amounts of suspect data. The present paper introduces an attempt to optimize the post-mortem analysis by means of visualization. The approach uses output data of current forensic tools and allows investigators to visually build correlations, with the aim of getting hints as to where it would make sense to start looking for evidence. Formerly unrelated primitive objects are visually classified and aggregated to more complex objects through attributes. As a result, disk-images can be searched for occurrences of patterns similar or close to the ones specified. Search results can be of different type; possible examples are graphs of statistical distributions or even self-organizing maps.

KEY WORDS

digital forensics, post-mortem analysis, visual correlation

VISUAL CORRELATION IN THE CONTEXT OF POST-MORTEM ANALYSIS

1 INTRODUCTION

The field of digital forensics is considered as a branch of common forensic science, and has increasingly detached itself from the broader area of computer security to become a self-contained forensic discipline over the past ten years. Digital forensics are defined in several ways. (Computer Legal Experts, 2007) states it as being *"*[...] *the application of computer investigation and analysis techniques in the interests of determining potential legal evidence"*.

The process of forensic analysis aims at answering questions about former system states and events by reproducing chains of digital activity. These chains of activity are the result of bringing potential legal evidence into a chronological progression and represent one of the most difficult tasks for an expert. It is fairly easy to collect information from a system; the complexity lies in the ability to correlate bits and pieces into a reproducible sequence of past events. Several tools and toolkits have been developed to assist forensic experts and security specialists in their daily work, and have proven their reliability during the process of forensic analysis. There are both open and commercial products available, *The Sleuth Kit* (Carrier, The Sleuth Kit) and *EnCase* (EnCase Forensic, 2008) being two prominent and widely used examples. The authors emphasize that the approach at hand will focus on open source tools only, for the time being.

1.1 Motivation

A forensic investigation is a time critical process. In most cases, external circumstances determine the time available to experts to find supportive evidence. Efficiency and an intuitive handling of large sets of data are of prime importance to the process of forensic analysis. Most security incidents implicate more than just one storage medium. Discussions with Swiss security and forensic experts (Bundeskriminalpolizei, 2008) have shown that whenever a set of storage media has to be examined, it is done

sequentially. Either, because there is not enough equipment at hand or simply, because qualified human resources are low.

The data that eventually represent evidence, are but a small fraction of the data stored on a disk. Furthermore, the capacities of storage media keep increasing, which makes it even more difficult for specialists to know where it makes sense to start looking for evidence among the data to examine. Many of the forensic tools currently in use generate their results as lists or tables, whose length depends on the number of matchings found after applying one or several filters these tools provide. This leads to the fact, that the process of forensic analysis becomes more and more complex in terms of getting a quick overview of the data, and the efficiency of the way that data is being processed.

Studies (Miller, 1956) show that the ability of the human brain to understand complex structures and the relations they induce can be significantly increased through visual stimuli. Hence the approach presented in this paper builds upon the assumption that there is a need for a simplified, assistive means, which allows for a coherent view on the structures and relations of data of different type through the use of abstract visualization. The focus of this work is set to the so-called *post-mortem* analysis, which will be discussed in more detail in Section 3.

The next section will give a brief overview of the process of forensic analysis and describe the different phases it consists of. Section 3 will outline the post-mortem analysis to introduce the context of the presented approach. Section 4 gives an insight into the basic features of common open source forensic tools. The fifth section will present the approach the authors suggest, and the last section will conclude with a brief summary and outline future work.

2 THE PROCESS OF FORENSIC ANALYSIS

The process of forensic analysis defines a sequence of actions to be taken in the event of IT security incidents, e.g. where one or several computers have either been used as a target, or as a means to commit a crime. As a general rule, the authors divide this process into the following 4 phases:

1. Coverage of the crime scene – covering a crime scene goes beyond the seizure of suspect hardware. The surroundings have to be given just as much

attention. For detailed information on crime scene investigation, refer to (Fisher, 2000).

2. Data acquisition – if the suspect system is still running, all volatile data (this concerns all data held in RAM at runtime and temporary files on the hard drive) is to be recovered, if possible without changing the system's actual state. For further details on *live acquisition*, refer to (Carrier, 2005). The second step during data acquisition is called *forensic duplication* and consists of creating exact copies (images) of all hard drives and related media like USB sticks, CD-ROMs etc. to a clean hard drive. This process can be performed locally or over a secured network channel. In depth information on disk imaging can be found in (Carrier, 2005).

3. Post-Mortem analysis – all acquired data images are examined and searched for supportive evidence within a secure environment. This analysis is always performed on copies, never on the original data. Post-mortem analysis will be discussed in more detail in the next section.

4. Consolidation of the investigation's results – all potential evidence is put in chronological correlation, which allows for investigators to draw appropriate conclusions, in order to rebuild plausible and reproducible chains of activity for further use before court.

3 POST-MORTEM ANALYSIS

A post-mortem or *dead analysis* is performed on copies of duplicates gathered during data acquisition. Forensic experts can work without the pressure of a live system, since there is always a backup of the original image available, if necessary. Data acquisition, as a general rule, is done at the disk-level. Loss of possible evidence has to be avoided; this is why disk images should not be created at the volume, file or application levels. For example, if the data would be acquired at the file level, non-allocated space would not be copied and hence make a recovery of deleted files impossible. Potential evidence is lost at every level of abstraction; therefore data, as a rule of thumb, should be acquired at the disk level in order to save every byte that may contain evidence. However, there are situations, in which an investigator might decide to duplicate data at a higher level. It is up to the

expert in charge to decide, where evidence is most likely to be found. This decision mostly depends on the expected type of attack, and the experience of the specialists assigned to the case.

A post-mortem analysis examines all the data gathered from a suspect system for potential leads and evidence. This analysis is done on all possible levels, spanning from the application level down to the disk level, if applicable. Points of interest are unallocated space on hard drives (including slack space), MAC-times (last modification, access, change), swap space, hidden files, deleted files, the structure and content of unknown binaries, log files and operating system related information (kernel version, loaded modules, registry information on Windows etc.), to name a few. It is highly recommended to start with recovering deleted information when conducting a post-mortem analysis (Jones, Bejtlich, & Rose, 2005). Most perpetrators make sure to delete all information relevant to an investigation, before leaving a system. Furthermore, experience shows that a set of suspect data can be reviewed more efficiently, if it is previously reduced to what is relevant to the process of finding evidence. There is much more to say about post-mortem analysis, but doing so would be out of the scope of this paper. Suffice it to say, that all the steps of such an analysis can be performed with the aid of current forensic tools. Detailed information on how to conduct a post-mortem analysis can be found in (Farmer & Venema, 2005).

As aforementioned, the phase of post-mortem analysis sets the foundation for the approach suggested in this paper. The next section will briefly discuss the information one can extract with most of the current open source forensic tools.

4 OPEN SOURCE FORENSIC TOOLS

Ever since digital forensics became important to criminal investigation, people have been working on tools to assist specialists and simplify the task of finding relevant information on corrupted systems. Most of these tools significantly improved the process of forensic analysis, mostly by providing scripts to automate or partly automate the acquisition, recovery and analysis of suspect data. Authorities work with both commercial and open source toolkits. Discussions and experience reports show, that one of the most recurrent drawbacks of current forensic tools is the graphical user interface (GUI) or lack thereof. If available, these GUIs are often complex in their

usability and make it difficult to get a fast overview of relevant information. However, efforts have been made to address the issue: Brian Carrier's *Autopsy* (Carrier, Autopsy Forensic Browser, 2008) tool is a notable example to account for these efforts.

Most toolkits are available as a live CD and assemble a collection of useful tools for live and dead data acquisition, as well as tools for forensic analysis for both Unix-based (The UNIX system, 2008) and Windows operating systems. Many suppliers rely on Linux distributions with good hardware detection capabilities, such as debian (debian, 2008) or KNOPPIX (KNOPPIX, 2008), which builds upon debian. These distributions are very convenient in that they allow for an immediate forensic analysis environment to be set up. The live CD can be mounted on a still running system, commonly referred to as a *smoking gun*, and an incident response can be performed out of the box. Statically pre-compiled binaries are used in order to avoid the execution of any system binaries, which might have been tampered with, root kits being a current example. An investigator can mount a system's partitions in read-only mode; execution of system binaries is prevented as well. Most live CDs provide a host of utilities to extract valuable runtime information, such as RAM content, process information, network information (open sockets etc.) and other temporary data, which would be lost after a system shutdown.

Tools for data acquisition are indispensable for any forensic toolkit. As a general rule, a "good" toolkit will allow experts to duplicate both dynamic and static data to any clean hard drive or over an encrypted network channel. As for the analysis of acquired disk images, the possibilities are far-reaching. Data recovery on different disk levels, timeline analysis (through data timestamps), analysis of unknown binaries and meta-data analysis, are but a few of the possibilities offered to specialists. EnCase has become the state of the art solution for digital forensics among all available commercial products on the market. Its feature set is impressive indeed, but many of the available products developed in the open source community can hold their ground and offer a huge potential for both research and development alike.

The next section discusses a new approach to assist forensic specialists in their work during a post-mortem analysis.

5 VISUAL CORRELATION

This section introduces an approach to optimize the post-mortem analysis by means of visualization. Based on the facts stated throughout this paper, the authors suggest to make use of the advantages brought into play by visual and interactive assistance to simplify the process of rebuilding past chains of activity. These chains eventually result from correlating initially unrelated data and the appropriate interpretation of an investigator, which relies on past experience, to obtain legal evidence. The present research is in its early stages; a prototype implementation is not yet available but will be established in the near future.

The authors propose a GUI, which uses any set of results from current forensic tools as input. This makes sense because all suspect data has already been reduced to a subset of relevant data at this point. As a first step, the input data needs to be pre-processed in order to be graphically displayed. This is done through a logical interface, which recognizes the type of information contained in the input set and abstracts it to classes of graphical entities. For example, if the input data contains information on i-nodes, access times, log files and system processes, the pre-processing will find out about four different types of information. From a graphical point of view, the interface will display these four entities, each of which stands as a representative for its respective type of information (e.g. i-NODE, MAC, LOG and PROC). This abstraction is needed, because displaying every single entity of the input set would result in an unreadable and hence unusable mix-up.

Each decision one makes depends on former action. The same holds for any event on a computer system. Getting back to the suggested approach, the process of correlating entities can be performed the very same way. Before getting into any further detail, another assumption needs to be made. Every data structure is described through a set of properties. These can be meta-data, file names, file extensions, file content, network class, process information, to name just a few. So each of the representative classes can be assigned a set of possible attributes to describe them. An investigator can now use this information to visually build correlations between different classes. Back in the GUI, a selection of attributes can be made for each of the initial graphical entities. The next step consists of aggregating several representatives to form a possible correlation. Consider the following

example to clarify this process. The GUI initially displays 3 representative classes, according to the content found in the input set. These entities are IP, @ and LOG. The investigator would like to find out, if a specific e-mail address can be put in relation with one or several IP addresses. So he/she first selects the e-mail address from the @ class' *sender / recipient* attribute and proceeds the same way to select a range of IP addresses suggested by the IP class. Both graphical entities are then visually correlated, e.g. graphically connected to each other, to form what the authors call a *visual pattern* or *interrogator*. This process of visual correlation is illustrated in Figure 1.

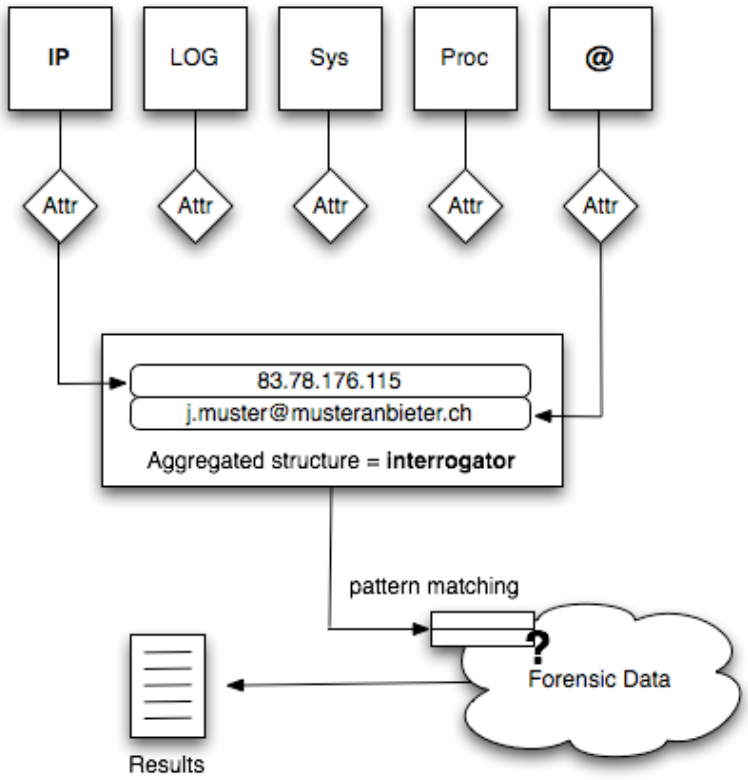


Fig. 1 The process of visual correlation

On a logical level, an interrogator is nothing else than a pattern, which is eventually applied to the input data set. The forensic data is being searched for matching or similar patterns. In the above example, the result might contain one or several entries from a log file, which states that J. Muster has sent or received e-mail on the machine with IP 83.78.176.115. This new information might give the investigator a new lead and will determine the next steps, which might either consist of building a whole new pattern or reusing the previous one to correlate even further, e.g. to refine or restructure the actual interrogator using additional attributes and classes. Another focus will be set on reusability. Most attacks follow specific patterns and stand out through unusual activity. Interrogators can be saved for reuse, or even previously specified and applied to a certain category of forensic data with similar structure.

In a sense, this concept allows to ask questions in a top-down manner, and to reformulate these questions, should the answer be unsatisfying for the examiner. The authors believe that this approach will increase the efficiency with which investigators correlate suspect data and interpret it to retrace past events and system states. The concept presented above provides a simplified view on structures and relations of suspect data, and removes a part of the complexity when it comes to evaluating long lists of results.

6 CONCLUSION

The authors have introduced a new approach to the process of post-mortem analysis, more precisely, to the correlation of initially unrelated data. They propose a visual concept to assist investigators in the process of reproducing chains of previous system activity. The complexity and effort, needed to process large sets of data, is reduced through abstraction. Content of input data sets is classified and displayed by means of graphical class representatives. Visual correlation allows forensic experts to easily specify search patterns, which can be applied to forensic data.

The suggested GUI is comparable to the evidence finding process in common forensic science. Potential evidence of different type (fingerprints, pictures, textile fragments, DNA etc.) is collected and correlated, to verify if there is any relation between them that might be used to incriminate a suspect.

6.1 Future Work

As aforementioned, this work is current research in its early stages. A first prototype has to be developed to deliver a proof of concept. This will allow verifying the use and the applicability of the authors' assumptions. The logical interface, which pre-processes and classifies input data sets as graphical representatives, is about to be specified. Each of the possible representatives with their respective set of attributes will be specified with XML Schema (Vlist, 2002). The prototype will be implemented with the Java programming language (Flanagan, 2005) to allow for maximum portability on different platforms.

Different possibilities to report results are also being considered. First of all, it has not yet been decided how results are being processed and displayed to the user. It might be of interest to offer the ability to choose from different representations. One might include a plug-in mechanism to provide a flexible means to add new reporting types at a later point. Possible types could be graphs and diagrams to visualize the number and relations of particular matchings. Statistical distributions and even self-organizing maps (Kohonen, 2007) might be taken into account. The proposed approach leaves room for discussion, but first meetings with researchers and professionals alike have shown that there clearly is a need for visual concepts in the field of digital forensics. The authors currently seek feedback from law enforcement agencies to test the applicability of their approach to real digital forensic investigations.

7 REFERENCES

(2008). From Bundeskriminalpolizei: http://www.fedpol.admin.ch/fedpol/de/home/fedpol/organisation/bundeskriminalpolizei.html

Carrier, B. (2008). *Autopsy Forensic Browser*. From Autopsy Forensic Browser: http://www.sleuthkit.org/autopsy/index.php

Carrier, B. (2005). *File System Forensic Analysis.* Amsterdam: Addison-Wesley Longman.

Carrier, B. (n.d.). *The Sleuth Kit*. Retrieved 2007 from The Sleuth Kit: http://www.sleuthkit.org/sleuthkit/

Computer Legal Experts. (2007). From Computer Legal Experts: http://www.computerlegalexperts.com/

debian. (2008). From debian: http://www.debian.org/

EnCase Forensic. (2008). From EnCase Forensic: http://www.guidancesoftware.com/products/ef_index.asp

Farmer, D., & Venema, W. (2005). *Forensic Discovery.* Amsterdam: Addison-Wesley Longman.

Fisher, B. A. (2000). *Techniques of Crime Scene Investigation.* CRC Press.

Flanagan, D. (2005). *Java in a Nutshell - A Desktop Quick Reference.* O'Reilly Media.

Geschonneck, A. (2006). *Computer-Forensik.* Dpunkt Verlag.

Jones, K. J., Bejtlich, R., & Rose, C. W. (2005). *Real Digital Forensics.* Amsterdam: Addison-Wesley Longman.

Miller, G. A., (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97. Retrieved 2007 from Psychology Department of University of Toronto.

KNOPPIX. (2008). From KNOPPIX: http://www.knoppix.org/

Kohonen, T. (2007). *Self-Organizing Maps.* Berlin: Springer.

The UNIX system. (2008). From The UNIX system: http://www.unix.org/unix03.html

Vlist, E. v. (2002). *XML Schema.* O'Reilly Media.

LOCATION AND MAPPING OF 2.4 GHZ RF TRANSMITTERS

Daniel Wells[1], Ingrid Siebörger[2] and Barry Irwin[3]

Rhodes University
Department of Computer Science
South Africa

[1]g03w0418@campus.ru.ac.za, [2]i.sieborger@ru.ac.za,
[3]b.irwin@ru.ac.za

ABSTRACT

This paper describes the use of a MetaGeek WiSpy dongle in conjunction with custom developed client-server software for the accurate identification of Wireless nodes within an organisation. The MetaGeek WiSpy dongle together with the custom developed software allow for the determination of the positions of Wi-Fi transceivers to within a few meters, which can be helpful in reducing the area for physical searches in the event of rogue units. This paper describes the tool and methodology for a site survey as a component that can be used in organisations wishing to audit their environments for wireless networks. The tool produced from this project, the WiSpy Signal Source Mapping Tool, is a three part application based on a client-server architecture. One part interfaces with a low cost 2.4 GHz spectrum analyser, another stores the data collected from all the spectrum analysers and the last part interprets the data to provide a graphical overview of the Wi-Fi network being analysed. The location of the spectrum analysers are entered as GPS points, and the tool can interface with a GPS device to automatically update its geographical location. The graphical representation of the 2.4 GHz spectrum populated with Wi-Fi devices (Wi-Fi network) provided a fairly accurate method in locating and tracking 2.4 GHz devices. Accuracy of the WiSpy Signal Source Mapping Tool is hindered by obstructions or interferences within the area or non line of sight.

KEY WORDS

Rogue Access Points, Spectrum Analysis, Trilateration, Wi-Fi

LOCATION AND MAPPING OF 2.4 GHZ RF TRANSMITTERS

1 INTRODUCTION

Wireless networking has brought computer networks into a new, exciting and hostile environment. Factors that need to be considered and understood during implementation of Wi-Fi networks include interference sources and security protocols. Setting up a Wireless Local Area Network (WLAN) is relatively simple, allowing users to achieve mobility, but in some cases, the default security configuration on the devices leads to inferior security measures being implemented. Security leads to a higher implementation complexity and so can sometimes be avoided by the average user.

IEEE 802.11b/g/n Wi-Fi specifications use the 2.4 GHz frequency band. As these technologies become increasingly popular for the home and business, the 2.4 GHz spectrum is becoming cluttered, therefore a need for optimal use of the medium is required. Wi-Fi throughput can be increased by selecting the least utilised Wi-Fi channel, minimising interferences and removing rogue access points (APs). By combining the frequency VS signal amplitude data from three (or more) 2.4 GHz spectrum analysers it is possible to locate 2.4 GHz interference sources and transmitting Wi-Fi devices. The data from the spectrum analysers is combined to produce a graphical display of a Wi-Fi network and devices are located using the method of trilateration [15]. The WiSpy Signal Source Mapping (SSM) Tool was developed to meet this goal.

The graphical display enables users of the tool to discover the approximate locations of 2.4 GHz transmitters and interferences sources. The tool allows users to gain optimal use of the frequency by minimising interference and improves security by providing a close approximation of the physical location of (rogue) Wi-Fi devices. Such a tool can potentially prove invaluable for the auditing and planning of wireless networks within an organisation.

This paper presents the MetaGeek WiSpy spectrum analyser together with the client-server application that was developed. The paper is divided into two logical parts beginning with sections 2 and 3 which discuss related work and introduce the WiSpy SSM Tool. The second part, sections 4 and 5, describe testing and results and discuss relevant conclusions.

2 RELATED WORK

The IEEE 802.11 (Wi-Fi) family of technologies have been adopted on a global scale, and installed in equipment ranging from desktops and laptops to mobile phones, security cameras and home entertainment systems [13]. Security in wireless networking has had to overcome hurdles. Default settings on hardware are most frequently set to least secure operating modes, which not only aids the end-user in setting up their network but also the attacker who wants to take control of it. The responsibility placed on the user ranges from specifying types of security protocols used and specifying passwords for Access Points (APs) and clients to managing a Public Key Infrastructure. A more secure network is more complicated to configure, leading to strong Wi-Fi security solutions being out of reach by the typical end-users. However in any network a trade-off exists between security and usability, but in wireless networking where there is a significant lack of any physical barriers to access, a strong security implementation in crucial [2].

The original and still widely used Wi-Fi security protocol WEP, requires clients and APs to share a single secret key which is used to encrypt all datalink layer communication [1]. The goals of WEP are to provide confidentiality, integrity and access control (C.I.A) and at the time of its release it provided these, but after much scrutiny by cryptologists and attackers the protocol was discovered to have many flaws. Wi-Fi Protected Access (WPA) was introduced to address all the known vulnerabilities of WEP and does so with a minimised impact on network performance [16]. WPA2 is the latest version of WPA with even more security features than WPA.

Apart from which security protocol is utilised, many other factors can influence network performance. Wi-Fi (specifically IEEE 802.11b/g/n) propagates over a cluttered frequency of 2.4 GHz. Typically interference can be separated into two broad categories; traffic from adjacent Wi-Fi networks and that arising from any other transmitters operating in the same frequency band [11]. Adjacent Wi-Fi networks are of the most concern to those living or working in densely populated areas, or multi-tenant office buildings. Some typical devices which cause interference are a range of cordless phones, any Bluetooth device, cordless headsets, wireless bridges, cordless video-game controllers and microwave ovens. A microwave oven can create interference from up to 50 feet (15 meters) away and incur relatively high packet retransmission [6]. Obstructions between antennas also leads to reduced throughput

because the radio link depends on the energy diffracted around the object rather than direct radiation [4]. Wireless Denial-of-Service (WDOS) attacks exists where custom designed transmitters output onto a particular frequency and transmit either Gaussian white noise or a high amplitude signal to effectively prevent any wireless transmissions occurring in a given radius. WDOS devices are illegal, yet plans and kits can easily be viewed or purchased on the Internet [5]. A simpler form of DOS floods the WLAN with associate messages, which prevents any host from sending data or connecting to the AP[14].

Specific concepts and terminology are important in helping understand how one is able to pinpoint the location of 2.4 GHz signal sources. Signal strength in a Wi-Fi network is measured using dBm (decibel milliwatts), which is measured on a logarithmic scale [3]. Wi-Fi devices will be marked with a receive sensitivity and a transmitter power output in this scale. This measurement is particularly useful when working out the distance a signal has traveled, if known at what strength the signal was transmitted. Another important concept is the method of trilateration, similar to triangulation in that it uses the location of known points to discover the position of another point in space [10]. Trilateration uses known distances, not angles, from three points to an unknown point to discover the exact location of the unknown point. Trilateration can be imagined as circles originating from each known point where the radius of the circle is the distance to the unknown point. Where the circles intersect provides the location of the unknown point [10]. Three known points provide the ability to use the method of trilateration, by using more than three allows the accuracy of the method to increase.

A tool to speed up the process of analysing interference and evaluating frequency usage is a spectrum analyser. Although most spectrum analysers on the market are incredibly expensive and bulky, this project utilised a low-cost device with the form factor of a typical USB flash drive. The MetaGeek WiSpy 2.4 GHz Spectrum Analyser takes measurements of signal strength (amplitude in dBm) across radio frequency (2400 - 2483 MHz), and costs $199 USD each [9]. The WiSpy device has a receive sensitivity of -90 dBm, can make approximately five sweeps (obtain frequency VS amplitude data) per second and operates as a low-speed USB Human Interaction Device (HID) [7]. Due to the nature of HID devices, multiple operating systems can use the device with standard drivers. This is the device on which this project was rooted, although with minor modifications any spectrum analyser operating

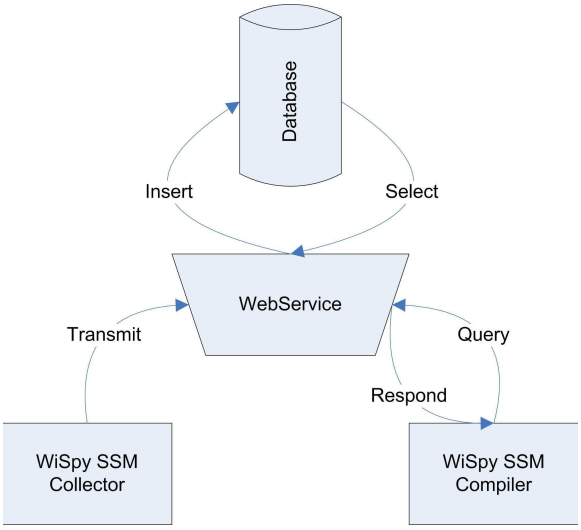


Figure 1: Design of WiSpy SSM Tool

in the 2.4 - 2.5 GHz range should work.

Using the WiSpy together with the custom client-server software tool and the method of trilateration, WiFi transmitters (including rogues) can be tracked and found. The following section describes the tool and its features.

3 WISPY SSM TOOL

The system created was named the WiSpy SSM Tool, SSM for Signal Source Mapping. The system was developed in two parts (applications) with a webservice to connect them and hold the main data store, Figure 1 provides an overview of the system. The first part of the solution is the collecting client; it interfaces with the spectrum analyser and transmits data to the webservice. The collector part of the solution is similar to the software which is packaged with the spectrum analyser, MetaGeek Chanalyzer [8], although the packaged application has no real-time method for extraction of signal data.

No limit exists on how many collecting clients can be present, as more collecting clients will achieve a higher accuracy when discovering the location of 2.4 GHz devices. The webservice receives the data from the collecting client

and stores it in a local lightweight database. The second part, the compiling client, sends queries to the webservice for data which responds if it has data to match the specific query. The compiling client compiles and sorts the data chronologically to graphically display the surrounding 2.4 GHz signals. Each individual part is discussed in further detail in the subsequent sections. Greater detail can be seen in [15].

3.1 WISPY SSM COLLECTOR

This application, in essence, interfaces to the WiSpy spectrum analyser, displaying a line graph of the current signal amplitude VS frequency graph and transmits this data to the webservice to be stored. Ideally this data should be transferred to the webservice over a wired network, a Wi-Fi transmission of data would affect the spectrum analyser signal data collection. In addition to the signal data, the related time, location and node information is also transmitted to the webservice. The data is collected in real time and not modified in any way and temporarily stored in batches to be sent to the webservice. The location is handled as GPS coordinates and the application provides additional functionality to interface with a GPS device to automatically update this field. By combining automatic GPS location updates with the application, roaming collecting nodes are possible. Also, if no Internet or network connectivity is present, data can be directly serialized to a file to be transmitted at a later time. All data is stored and transmitted as XML. This application is not resource intensive and can therefore run minimalistically and unobtrusively at any machine, at any point on the network.

3.2 ASP.NET WEBSERVICE

The webservice provides the interface to a database from which the two applications send and request signal data. The webservice receives requests and responds to them and is stateless. SQLite [12] was the database chosen as it is a light weight solution, perfectly suited for a service where minimal amounts of space are available; it has a small code footprint and provides the necessary data types and operations for this project. Data types of type TEXT and REAL where used, and tables and data are manipulated using standard SQL statements. The database is stored in a single disk file, it has a simple and easy to use API, is self contained and the source code is dedicated to the public domain.

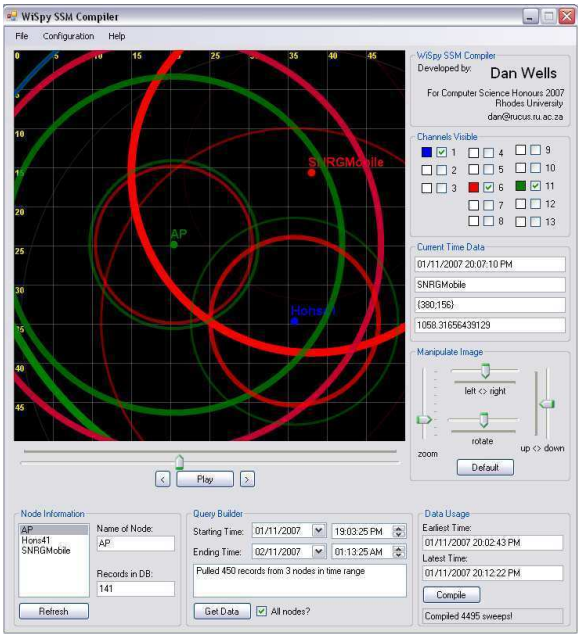


Figure 2: Screenshot of WiSpy SSM Compiler in use

3.3 WISPY SSM COMPILER

Once the signal data has been collected by numerous WiSpy SSM Collectors and stored in the database via the webservice, it needs to be processed and meaningfully displayed in order to discover the location of 2.4 GHz devices. The WiSpy SSM Compiler interfaces with the webservice to provide a list of all the nodes present in the database, and the user has the option of selecting all the nodes or a subset of the nodes to query for data. The user selects a time range from which they would like to view data, and the query is sent to the webservice. Once data is returned it is sorted by time and ready to be viewed by either replaying it in real time or quickly skipping through it using the slider. The display can be rotated and scaled to the users preferences to aid in locating devices. A screenshot of the compiler can be seen in Figure 2.

The data is displayed graphically on a scale grid, the scale can be modified to the users preference by setting latitude, longitude and the width of the display. The signal data is drawn to screen using circles for each Wi-Fi

channel (1-13) that originates from the node location. The user has the option of selecting which channels they would like to view, perhaps only showing the most popular channels (1, 6 and 11) or a specific channel. The larger the circle the further the signal is transmitted from its source to the collecting node, and the smaller the circle the closer the transmitted signal is to the collecting node.

$$\frac{P_{rx}}{P_{tx}} = \frac{G_{tx} \times G_{rx} \times c^2}{(4 \times \Pi \times d \times f)^2} \tag{1}$$

$$d = \frac{\sqrt{\frac{G_{tx} \times G_{rx} \times c^2}{\frac{P_{rx}}{P_{tx}}}}}{4 \times \Pi \times f} \tag{2}$$

The equation used to calculate the distance is shown in equation (1). The symbols used in the signal equation are as follows: P_{rx} is the received power (in watts). P_{tx} is the transmitted power (in watts). G_{tx} is the gain of the transmitting antenna. G_{rx} is the gain of the receiving antenna. c is the speed of light (3×10^8). pi (Π) is approximated to 3.14159. d is the distance between the receiving and transmitting antennas. f is the frequency (in Hz). As d is the variable we will be attemping to discover, equation (2) is simplified for d.

The equation used to calculate the distance is for the ideal line-of-sight scenario, which almost never holds in a real-life environment. In reality, the antenna gains will be hard to quantify (for different APs) and multipath propagation of the signal and obstructions will have unpredictable effects [4]. Any other 2.4 GHz signal sources in the area will also have unpredictable effects, for example, a transmitting Bluetooth device in the area could skew the results showing the AP to be slightly off course to where it really is located.

Once the data has been drawn to the screen it needs to be analysed and understood. With multiple collecting nodes present and displaying their signal data, simultaneous and synchronised, 2.4 GHz signal sources can be visualised and located. Firstly, the user needs to choose which Wi-Fi channel(s) they wish to view, with all channels selected the view can be cluttered. The channels to view can be decided by quickly running through all the data and

seeing which channels are mostly used, and then by deselecting the undesired channels. The user can then begin to locate Wi-Fi devices, by using the method of trilateration, as discussed in section 2.

In the next section results from numerous test cases are analysed and evaluated. In addition to results, typical output from both the WiSpy SSM Collector and WiSpy SSM Compiler are shown and discussed.

4 TESTING AND RESULTS

This section evaluates the toolset developed in order to determine its effectiveness. Results of both component applications (the Collector and Compiler) are discussed.

The experiments were conducted by utilising multiple APs from different vendors, and were configured in such a way that the APs were transmitting the majority of the time. The test setup had an AP connected directly to a personal computer (PC) with an additional PC four meters away, the second PC was installed with a Wi-Fi card and a network was created with the two PCs. Tests were conducted by uploading files from the PC at the AP to the second PC with the Wi-Fi card. The environment was evaluated beforehand to remove as many as possible interference sources which could skew the results. There was line of sight between all Wi-Fi devices and the collecting nodes. All results discussed here were from collecting nodes at fixed locations, although an evaluation with GPS dynamic location updates was also successfully conducted.

4.1 WISPY SSM COLLECTOR RESULTS

Initially the WiSpy SSM Collector was tested to ascertain whether the data passed onto the webservice was accurate and meaningful. Three test cases are discussed, each with a constant file download taking place at a set distance of five meters but on different Wi-Fi channels. These parameters were set to test whether similar signal strength was received from different frequencies but over the same distance. The figures (Figures 3-5) discussed are output from the collector application. These have been cropped from the actual application display for the sake of clarity.

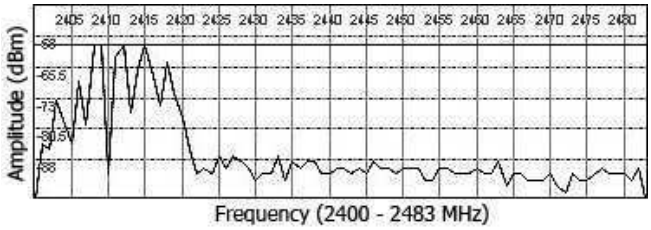


Figure 3: WiSpy SSM Collector - Channel 1 Download

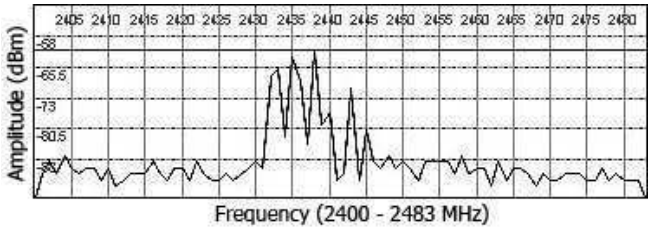


Figure 4: WiSpy SSM Collector - Channel 6 Download

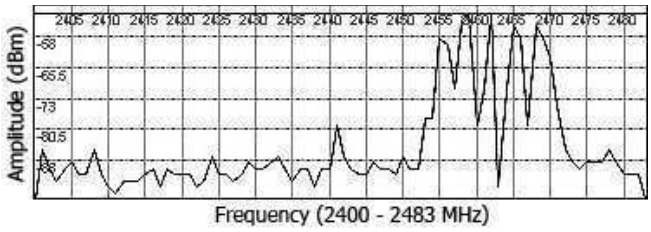


Figure 5: WiSpy SSM Collector - Channel 11 Download

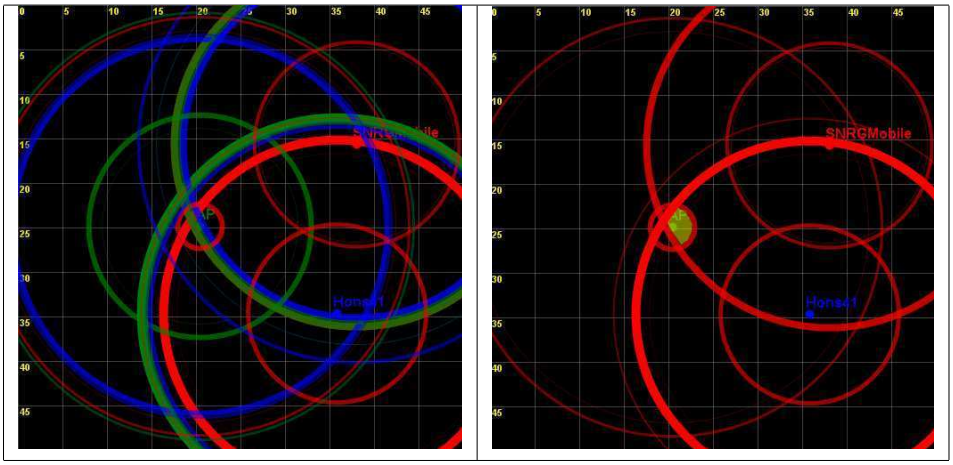
The frequency (in MHz) runs along the x-axis and received power is shown along the y-axis (in dBm). Figure 3 shows high activity centered around 2412 MHz, which demonstrates a Wi-Fi channel 1 download, which was the test case. Each Wi-Fi channel is 22 MHz wide and this is captured correctly. Figure 4 shows a Wi-Fi channel 6 download and Figure 5 shows a Wi-Fi channel 11 download. A simple test using a laptop and the collector application was conducted by initially standing near the transmitting AP and then moving further away from it. As expected, the signal strength reduced as the distance between the AP and the spectrum analyser increased – the signal would have to travel further and would therefore incur free space loss. Using equation (2) we confirmed that for a particular signal strength received the distance at which the signal was transmitted can be calculated.

4.2 WISPY SSM COMPILER RESULTS

Once the data from the WiSpy SSM Collector was confirmed to be accurate, evaluation of the WiSpy SSM Compiler was initiated. In these test cases, intermittent and irregular small file transfers were chosen over large file downloads as we wanted to mimic real world Wi-Fi usage in an office or production environment. The scale in all the following results is in meters. In figures 6-8, the brightest and thickest circles show the last signal data to be displayed. Where the most current circles intersect, an area is highlighted in yellow to suggest a device is in that approximate location.

4.3 WISPY SSM COMPILER RESULT SET 1

In Figure 6, Result 1A displays a typical WiSpy SSM Compiler output which is showing the most commonly used Wi-Fi channels; 1, 6 and 11. The display is cluttered with overlapping colours and circles. By quickly running through the data and analysing it, the user can decide which channel(s) they wish to view more closely. Figure 6 Result 1B displays the same point of time as Result 1A, but only Wi-Fi channel 6 is shown. If we consider the area of intersection, this result is very accurate, as the AP was two meters away from the WiSpy SSM Collector at the 'AP' node.



Result 1A - Cluttered	Result 1B - Only Channel 6

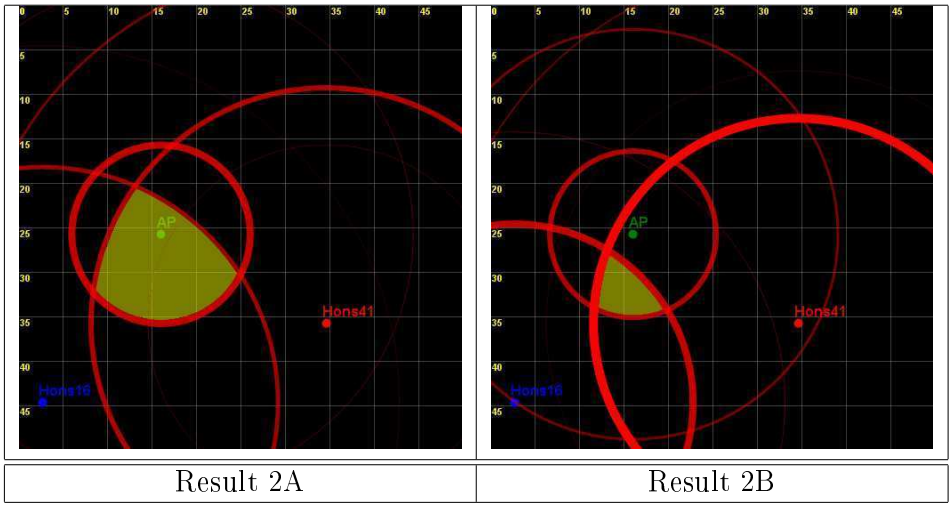
Figure 6: WiSpy SSM Compiler - Result Set 1

Looking closely at Figure 6 Result 1B we see smaller red circles originating from the 'SNRGMobile' and 'Hons41' nodes, suggesting the signal is originating closer to them that where the AP is located. As both these circles are of a similar brush width and brightness, they were collected around the same time, it is possible that interference could have occurred within this area to skew the result.

4.4 WISPY SSM COMPILER RESULT SET 2

Figure 7 shows a different physical layout of WiSpy SSM Collectors. This result set is also based on a Wi-Fi channel 6 network. The area of intersection of Result 2A (highlighted in yellow) is larger than the previous test case (Result set 1) but shows a fairly accurate display of where the AP may be. Result 2A provides an area where the AP is actually located and a person physically walking around the area could potentially see the AP. Figure 7 Result 2B was run under the same conditions as Result 2A, except that it is displaying a different point in time. Although Result 2B shows a smaller intersection area than Result 2A, the AP is not located within this area. It is possible that a potential interference source, not present during the experimental time of Result 2A, but later present during the experimental time of Result 2B could account for the later more inaccurate results. Again,

a person walking around this area could potentially see the AP. For the duration of this test, similar results to the above were obtained.



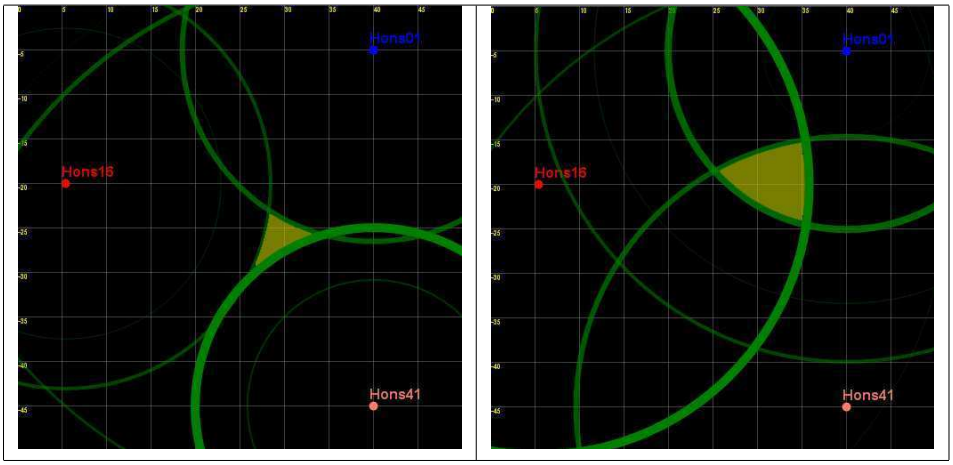
Result 2A	Result 2B

Figure 7: WiSpy SSM Compiler - Result Set 2

4.5 WISPY SSM COMPILER RESULT SET 3

Two results were obtained under a new physical layout as seen in Figure 8. Wi-Fi channel 11 was used in this result set and a WiSpy SSM Collector was not placed near the AP for these results. Instead the three collecting nodes where situated around the AP and all at approximately equal distances from it. In Figure 8 Result 3A, the highlighted area in yellow displays the area where the AP is most likely situated. Result 3A and Result 3B provide very similar areas of intersection and for the duration of this experiment the majority of the results suggested this highlighted area to be the location of the AP. The suggested area by the WiSpy SSM Compiler was a fairly accurate representation of where the AP was in fact located.

In the three result sets (Figures 6-8) we demonstrate the area highlighted in yellow. Using this information and an accurate knowledge of the sampling points or Collecting nodes (such knowledge can be obtained by GPS or building plans) this zone of interest can be determined, and allow for a closer physical inspection of the area.



Result 3A	Result 3B

Figure 8: WiSpy SSM Compiler - Result Set 3

5 CONCLUSIONS

With the rise of mobile users utilising the 802.11b/g/n Wi-Fi technologies, performance needs to be maintained and security needs to be infallible as much as possible. Network administrators need to keep a look out for rogue APs that are not authorised to be on the secure network. An administrator utilising a low-cost 2.4 GHz spectrum analyser can detect interference sources and choose the least cluttered channel for their Wi-Fi network as well as locate potential rogue APs within their networks. The WiSpy SSM Collector application was developed to reach these goals and was evaluated to be successful. By combining multiple WiSpy SSM Collectors (a minimum of three) around the Wi-Fi network, the administrator can fairly accurately locate where particular Wi-Fi devices are physically situated using the WiSpy SSM Compiler. By using more than three WiSpy SSM Collectors, the accuracy of the tool will increase.

The WiSpy SSM Tool can be used in many different settings. The tool allows hunting of rogue Wi-Fi APs and other 2.4 GHz RF sources such as Bluetooth devices or even mundane sources of interference such as microwaves. From a planning perspective, this tool provides the network administrator with a Wi-Fi site that can be used to assist in planning the Wi-Fi network prior to installation, or expansion..

Future work for this project include developing the application in Open Source Software to be ported onto the Linux and FreeBSD operating systems. Templates for types of interferences could be implemented into the WiSpy SSM Collector to automate detection of specific interference sources such as Bluetooth devices, microwaves, cordless phones and adjacent Wi-Fi networks. The WiSpy SSM Compiler could be further developed to display the full spectrum of signal data from each node on demand (similar to the line graph produced in the Collector). This additional functionality would provide the administrator with all the information they need at a central point. The WiSpy SSM Compiler could also integrate an option for under laying an image of the area under investigation, for example an image with the layout of an office, or perhaps a town map, even potentially be extended to produce 'kml' outputs for integration with the popular Go ogle Earth application, for mapping on a much wider scale.

ACKNOWLEDGMENT

The authors acknowledge the financial and technical support for this project from Telkom SA, Amatole Telecommunications, Business Connexion, Comverse, Mars Technologies, OpenVoice, Stortech, Tellabs and THRIP through the Telkom Centre of Excellence in the Department of Computer Science at Rhodes University.

References

[1] Arbaugh, W. A., Shankar, N. and Wan, J. Your 802.11 wireless network has no clothes. *Department of Computer Science, University of Maryland* (2001).

[2] Balfanz, D., Durfee, G., Grinter, R. E., Smetters, D. K. and Stewart, P. Network-in-a-box: How to Set Up a Secure Wireless Network in Under a Minute. *Palo Alto Research Center* (2004).

[3] Bardwell, J. *I'm Going To Let My Chauffeur Answer That: Math and Physics for the 802.11 Wireless LAN Engineer*. 2003.

[4] Button, D. Tech articles: Effect of obstructions on RF signal propagation. Online: http://www.emswireless.com/english/Tech_Articles/tech_art03.asp, Accessed: 19/03/2007, 1999.

[5] Farpoint Group. Evaluating interference in wireless LANs: Recommended practice. *Fairpoint Group Technical Note* (2006).

[6] Geier, J. Performing radio frequency site surveys to effectively support VoWLAN solutions. *Helium Networks* (2006).

[7] MetaGeek. Wi-Spy Hardware Interface Specification. Online: http://www.metageek.net/products-wi-spy-24x/development-specifications, Accessed: 05/06/2007, 2006.

[8] MetaGeek. WiSpy V1 Spectrum Analyser. Online; http://www.metageek.net/products/wi-spy, Accessed: 04/03/2007, 2006.

[9] MetaGeek. MetaGeek Store. Online: https://www.metageekstore.com/, Accessed: 04/03/2007, 2007.

[10] Murphy, W. S. and Hereman, W. Determination of a position in three dimensions using trilateration and approximate distances. *Colorado School of Mines* (1999).

[11] Rose, C., Ulukus, S. and Yates, R. Wireless systems and interference avoidance. *WINLAB, Department of Electrical and Computer Engineering, Rutgers University* (2000).

[12] SQLite. SQLite Home Page. Online: http://www.sqlite.org/, Accessed 01/09/2007, 2007.

[13] Tropos Networks. 802.11 Technologies: Past, Present and Future. Online: http://www.tropos.com/pdf/technology_briefs/tropos_techbrief_wi-fi_technologies.pdf, Accessed 22/10/2007, 2007.

[14] van Rensburg, J. J., Irwin, B. Wireless Security Tools. Proceedings of the ISSA 2006 from Insight to Foresight Conference, 2006.

[15] Wells, D. IEEE 802.11 Signal Source Mapping using Low Cost Spectrum Analysers. Department of Computer Science, Rhodes University, 2007.

[16] Wi-Fi Alliance. Wi-Fi Protected Access: Strong, standards-based, interoperable security for today's Wi-Fi networks. Online: http://www.54g.org/pdf/Whitepaper_Wi-Fi_Security4-29-03.pdf, Accessed: 04/04/2007, 2003.