TRUST ON THE WEB

Russell Cloran and Barry Irwin

Rhodes University

{R.Cloran,B.Irwin}@ru.ac.za

ABSTRACT

This paper forms a backdrop for work investigating trust on the semantic web. With the mass of information currently available on the web, and the low barrier to entry for the publication of information on the web, it can be difficult to classify the authority of information found on the web. We use a case study of a suspected phishing scam in South Africa to examine the methods an advanced user may use to verify the authenticity of a web site and the information it published. From this case study, we see that a website which is legitimate may easily appear to be a scam, because of the manner in which information is presented and the failure to use established industry best practices. We discuss a number of ways in which doubt may have been eliminated. We then discuss how a distributed trust system, as favoured by many researchers in trust on the semantic web, may have been implemented in this case to prove the authenticity of the site without the traditional means involving the high cost of a digital certificate from a recognised Certificate Authority.

KEY WORDS

Web credibility, trust

TRUST ON THE WEB

1 INTRODUCTION

In a distributed, open publishing environment such as the web, anybody can publish any information, misinformation, or disinformation. This is being further emphasised with the advent of personal publishing platforms, such as weblogs or blogs, being widely and freely available, allowing users with very minimal technical skill or infrastructure backup to begin publishing content on the Web. When browsing the web it is often left to the reader to decide whether the information being read is correct or not. There are a number of sources of information which can be used to attempt to check the credibility of a website, such as which domain the information is published in, the details of domain registration (obtained by whois), HTTPS certificates and the structure of the World Wide Web (link structure).

This paper uses a case study based around news articles about advanced fee fraudsters, and a website which purported to belong to an organisation which fought the advanced fee fraudsters. This paper aims to describe a number of checks which a user may do to verify the authenticity and credibility of a website, but also shows that these checks are not foolproof.

This paper begins with a brief introductory background, we describe the story surrounding the website which we will investigate and the reasons that it seemed initially suspicious. We then go on to describe a number of ways of verifying the authenticity and credibility of a website. For each check we also describe the scenario which occurred in the case study and suggest some possible ways of solving the problem.

1.1 Background

On 15 July 2004 a South African newspaper, DIE BURGER, published a story titled “Internet-sindikaat kry ook SA kredietinligting” (Internet syndicate obtains SA credit information). An English version of the story appeared on NEWS24.COM under the title of “419 scammers hack eBay” on 26 July 2004. The article claimed that advanced fee fraudsters had hacked into the eBay database, and obtained a list of credit card numbers. The article provided a link to 419legal.org, and a link to a page which allowed the user to check whether their credit card number was on the list of numbers which were illegally obtained. The 419legal.org website claimed to be affiliated to the South African Police Services, and contained SAPS logos in the site banner.

By 28 July the issue had been resolved, and the police had made a statement saying that Rian Visser was a police officer of the SAPS, and that he did run the 419legal.org website, as the website claimed.

1.2 Reason for suspicion

The check on the credit card numbers was performed by allowing the user to enter their full credit card number into a web form. When we initially visited the site, this form submission was across a plain text HTTP connection, but the next day the form was submitted across an HTTPS secured link. This seemed suspicious to us, and to a number of our peers, as it is poor practise to provide credit card details to anybody who is not fully trusted. Investigation of details surrounding the website turned up many loose ends, and not much consistency in the information available. It was suspected that the website may have been the front for a "phishing" operation, a type of attack in which a user is lured into providing details to a third party. Ironically, this was one of the types of operations which the website advertised that it was fighting.

1.3 Special note

We would like to point out that at the conclusion of investigations by both ourselves and the journalists involved, we were happy to find that 419legal.org was a legitimate website run by a police employee, Rian Visser. This paper is not intended to discredit Mr Visser or the organisations in which he is involved: We are satisfied that the website is credible, and indeed run by Mr Visser, himself a credible person involved in the investigation of advance fee fraud.

This paper only uses the 419legal.org site as a case study due to the interest it created amongst the South African Internet community which the authors involve themselves in. Despite a number of loose ends and suspicious information, 419legal.org was shown to be credible, which is part of the interest in the site. In this paper we suggest a number of ways in which the administrators of 419legal.org, or any other website, could improve the credibility of their website through a few simple technical changes.

2 VERIFYING THE AUTHENTICITY OF A WEBSITE

2.1 Web search

2.1.1 Names involved

A simple first step in finding out about a website may be to find the names of the people involved on the web site which is being checked, and then searching for these names on the web. Google is currently one of the most popular web searching tools in the world, if not the most popular. It is well recognised as being an impartial search engine which turns up accurate results. We can thus use Google to search for names, given that we know the names of the people who created or administer the website being checked.

This method is similar to the method described by Reagle[1], who suggests that "cryptographic signatures themselves might not be necessary to make a reasonable trust evaluation about a statement that has had time to grow into the tangled root structure of the Web." Whilst Reagle is talking about Semantic Web technologies, his suggestion may be taken further to simple facts which may be written about on the web, such as names.

At first glance this method seems to be a good first step, but without suitable cryptography is is conceivable that an attacker may set up a website which takes advantage of an already well established name on the web. In this point, we diverge slightly from Reagle's hypothesis about Semantic Web technologies - the statement which we can verify through preponderance could be taken to be "a person with this name is involved in the subject with that this website deals with", and in most cases this statement will be implicit. Reagle's suggestion may not be entirely appropriate here, because there should not be a large amount of information on the web saying that the person who created, or was mentioned in, other websites of the same topic is the same person involved in this website, the only link would be the weak link of the same name.

Case study We questioned the claims of the owners of the website on its own forums[1], as well as on a page on one of the author's blogs. We received the response that we should simply "Google for [the name involved]" and we would easily see that this person was a credible and well known fighter against advanced fee fraud.

Searching for "Rian Visser" revealed a number of web pages by and about Rian Visser. It appeared that there were two people who featured highly: A Dutch author of childrens' books, and a South African who was involved in fraud (especially advanced fee fraud) investigation. This second profile fitted the person who claimed to be the author of 419legal.org perfectly.

Whilst a network of web pages and sites such as those found on Google would've been difficult to fake, we found it plausible that a fraudster could take advantage of the wealth of information already on the Web, and create a website which claimed to be owned by this person.

Solution A solution to this problem is to take advantage of some sort of cryptography. The author of the web page need not have been certified by people who could have reached through a trust network (such as a PGP web of trust), or even by anyone else at all, but the fact that they would've been able to sign the new document with the same key as the large body of information which we found by and about them on the web would've convinced us that an attack such as that described above was not being carried out. Other mechanisms, such as Thawte's free Web of Trust product[2] could have provided an even better assurance of the identity of the website administrators.

2.1.2 Web linking

The design of the Web is such that a link may be created from any page to any other page, but this link is a one way link only. This is part of the power of the Web; anybody can link to anybody else, but nobody can force anybody to link back to them. Some sort of semantic value is therefore encoded in a link, and this can be thought of as a vote for the page which the link is directed at. This hidden semantic value was recognised and harnessed by the creators of Google, now a popular and prominent Web search engine. There is value for a user in being able to find other pages which link to a page. This gives some idea of related pages which the author of the page in question may not have linked to, and an idea of the popularity or perhaps even credibility of the page.

Google provides a number of advanced queries, one of which allows a user to search for web pages which link to a particular page, and another which allows a search within a certain sub-domain.

[1] http://419legal.org/index.php (Unfortunately we do not think that this URL will be long lasting, it has already changed once since we have known about it)

[2] http://www.thawte.com/wot/

Case study Searching for "link:419legal.org" on Google at the time of the investigation offered a large number of results for websites about advance fee fraud which linked to 419legal, however a search for "site:gov.za 419legal" provided only one result, which was not in the SAPS domain. It was expected that a website which was affiliated to the SAPS would have had a link from the SAPS website, however no such link was found.

The large number of links from other websites dealing with advance fee fraud provided evidence which strengthened the belief that 419legal.org was a credible website.

Solution If 419legal.org was linked to from a web page which should have been undeniably trusted, such as the SAPS web page on advance fee fraud, doubt as to the authenticity of 419legal.org would have been greatly diminished.

2.2 Domain registration

2.2.1 DNS

The Domain Name System is a hierarchical system which allows administrators of a domain to delegate responsibility for a sub-domain to another authority. An interesting implication of this is that there is a path of delegation which could be used to create a trusted path, similar to chained certificates in a PKI. Whilst is is not necessarily true that the path can be trusted for commercial domains such as .COM or .CO.ZA, where very little fact checking occurs there is a stronger trust implication with government domains such as .GOV.ZA. This means that it may be reasonable for a user to assume that a website within a government domain is accurate, because the administrators of that website are trusted by the administrators of the domain.

Case study In the case of 419legal, it was claimed that the website was run by a detective in the South African Police Service. It may have been reasonable to assume that if this website was a legitimate part of the SAPS they would have either obtained space on the SAPS website, or a sub-domain under the .GOV.ZA domain or one of its sub-domains. Even a domain registered within the .ZA top level domain may have seemed more legitimate than the .ORG sub-domain which was registered. Because the domain was registered within the .ORG top level domain it was impossible to verify the identity of the registrant through DNS or whois, which we discuss in Section 2.2.2.

Solution Creating your website in the correct domain can mitigate any confusion surrounding your website. If a user-friendly domain name is required, then techniques such as HTTP redirects may be employed to allow users to use the name which is easy to remember, but actually host the site on the domain, which will become the canonical name for the site.

If this is not desirable, or not possible, even a link from a page on a web site in a valid domain might be sufficient. This is discussed above in Section 2.1.2.

2.2.2 whois

The whois protocol was originally designed to run off one central server. As the size of the Internet became such that this was not possible, other protocols were developed, such as rwhois. The nature of the service has remained the same, it is a means of querying a server for information about a particular object, be it a person or an Internet hostname.

whois servers, in general, use a non-standard format to record information (there is no widely published standard format); the information is intended to be human readable, and not necessarily machine processable. For example Uniform, the authority for the .co.za domain, accept an email form as a part of the registration, and simply use the relevant parts of this form as the information provided by the whois server. As mentioned in Section 2.2.1, the DNS can be thought of as a path of trust, and since domain authorities often run whois servers for the domains which they control, whois provides the means to check information of domain registrants.

Domain authorities for large domains, such as .org or .co.za, often do not check information of domain registrants. Registrars, which provide information to the domain authority in some domains, may also not be trusted to check information thoroughly. The utility of the whois servers provided by domain authorities is thus limited because the trust chain discussed in Section 2.2.1 does not always properly hold up.

Case study The 419legal.org site was registered in the .org domain, a domain which uses registrars to do registrations. The information in the registration is not included in full, but in summary, the registrant was “D. Squire”, the organisation which the domain was registered for was “E-Payments” and the physical address was recorded as an address in Hillary, Durban. The email address provided was in the e-payments.co.za domain, and the telephone number provided was a number which appeared to be a telephone number for Kloof, an area of Durban which is not near Hillary.

This information did not match any of the other information which we had, and visiting the website “http://www.e-payments.co.za/” did not provide any more useful information. We believed at the time that “D. Squire” had registered the domain on behalf of Rian Visser, or 419legal.org, an assumption which we still believe to be true. It would have been possible, however, for “D. Squire” to act as an agent for 419legal.org, and register the domain in their name, or in the personal name of one of the people highly involved in setting it up. The fact that “D. Squire” had registered the domain in his name, in fact, provided a possibility for contacting “D. Squire” by telephone to verify the information about 419legal.org. This path, however, is still open to attack.

Solution Information published by whois servers should not be trusted as authoritative or correct. In verifying the credibility of a website, checking whois information could confirm information already held, but it should not be regarded as holding much weight.

Stricter checking by domain authorities or registrars could be implemented to improve this situation, but this may prove prohibitively expensive.

Registration in the name of the person holding the domain would go one step towards improving the apparent transparency of the information provided, but may not add any real value to the credibility of the website. As discussed above, registration in an agent's name may in fact improve the situation by allowing checking through a different avenue.

2.3 Professionalism

Professionalism is a something which is hard to pinpoint. It is an overall impression created by the website in question; the manner in which things are done. A study on web credibility found that a large percentage of users look at the overall design of a website as an indication of its credibility[2]. Website design is not the only point of professionalism, others include consistency over pages on the website and quality of information provided. The Stanford Guidelines for Web Credibility suggests 10 ways to improve the credibility of a website[3], many of which discuss the professionalism of the site.

Case study Although this is subjective point, the 419legal.org did not leave an impression of professionalism in its design. Images were poorly scaled, colours were poorly selected and the layout was not optimal. The information provided on the site was also in English, presumably to improve international communication, but a lot of it was clearly written by a person who did not speak English as their first language.

The secure page for checking credit card details was hosted on a different site, as discussed in Section 2.4. This is also a matter of professionalism. Although a certificate just for 419legal.org may have been prohibitively expensive, an operation which wished to give a professional image would have found a way to cover the costs.

Since the incident the web page has changed drastically, and URLs for pages have also changed. This is a point of professionalism, discussed by the creator of the web, Tim Berners-Lee in a short article called "Cool URIs Don't Change"[4].

Solution Although costly, investing in a well designed site can improve its credibility. This also applies to copy which will appear on the site. By improving design, transparency, and quality of information, 419legal.org would have shown that they were serious about their website, and thus improved the credibility of it and of themselves. Making use of an HTTPS certificate, as discussed below, would also improve the professionalism of the site.

2.4 HTTPS and PKI

HTTP over TLS[5], often referred to as "HTTPS" for the URI protocol identifier it uses, provides a means for transporting HTTP[6] traffic over TLS[7]. HTTP over TLS requires that a client match a certificate to the host name being queried, unless there is a good reason that that can not or should not be done, for example, the server is on a dynamic IP address and it does not have DNS which can be frequently updated. Certificate Authorities issue certificates for a domain, and check the details of the person or entity requesting the certificate. This means that there is now a verified link between the domain name and the person or company's information provided by the certificate authority, a trusted third party. HTTP over TLS also provides end to end encryption, which means that information can not be read by a third party while it is in transit.

Case study 419legal.org provided a facility to check whether a credit card number existed in the database of stolen credit card numbers that they held. The web page providing this facility[3] was hosted on WWW2.SECURESITESERVER.CO.UK, and the certificate was issued to a company called DONHOST LTD. There was no way to check this further relationship on the web.

Solution TLS certificates should be issued to the organisation which is going to be the end user. This allows a user to check that the website does belong to the organisation which claims to own the website. Using HTTP over TLS as it was used on 419legal.org added confusion, however at least end to end encryption is provided.

3 CONCLUSION

The case study shows how a web page which is poorly set up can do damage to its credibility despite the fact that it is valid. We discussed the use of search engines, the structure of the domain name system and domain registration information as ways of checking the validity of a website. It was shown that HTTP over TLS can be poorly used, thus diminishing its utility in verifying the authenticity of a website.

3.1 A web of trust

There are only a few ways in which the authenticity of a website can currently be established, outside of the use of potentially costly PKI technologies such as TLS. Many of these means require good technical knowledge to use, and are thus inaccessible to most users. It is desirable to be able to verify the authenticity of a website, so that users of the Web can know whether they can trust the information that they are reading, and we demonstrate in this paper that the web is still far from that.

Acknowledgments

This research was supported by DAAD, the NRF and was conducted at the Center of Excellence in Distributed Multimedia at Rhodes University, Grahamstown.

References

[1] J. M. Reagle, “Finding Bacon’s Key: Does Google Show How the Semantic Web Could Replace Public Key Infrastructure?,” 2002. http://www.w3.org/2002/03/key-free-trust.html.

[2] B. Fogg, C. Soohoo, D. Danielson, L. Marable, J. Stanford, and E. R. Tauber, “How do people evaluate a web site’s credibility?,” Oct. 2002. http://www.consumerwebwatch.org/dynamic/web-credibility-report-evaluate.cfm.

[3] B. J. Fogg, “Stanford guidelines for web credibility,” May 2002. http://www.webcredibility.org/guidelines/.

[3] https://www2.securesiteserver.co.uk/cbranch/card1.asp

[4] T. Berners-Lee, “Cool URIs don’t change,” 1998. http://www.w3.org/Provider/Style/URI.

[5] E. Rescorla, “Http over tls,” Tech. Rep. Internet RFC 2818, IETF, May 2000. http://www.ietf.org/rfc/rfc2818.txt.

[6] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Peach, and T. Berners-Lee, “Hypertext transfer protocol – HTTP/1.1,” Tech. Rep. Internet RFC 2616, IETF, June 1999. http://www.ietf.org/rfc/rfc2616.txt.

[7] T. Dierks and C. Allen, “The TLS protocol,” Tech. Rep. Internet RFC 2246, IETF, Jan. 1999. http://www.ietf.org/rfc/rfc2246.txt.