Using digital certificates to identify web site owners and protect against phishing

The solution to phishing already exists, so why is it never used?

Edwin Aldridge
August 2005

Abstract

Phishing exploits the ordinary Internet user's inability to be sure that a web site which they have been induced to visit is actually operated by the company or organization whose name appears on screen.

Yet the technology and infrastructure to verify precisely this not only exists but is deployed and used daily by every popular browser. Digital certificates can provide browsers with a reliable source of information about site owners which users can access via the padlock icon, although the information is not as useful as it might be and the padlock display is unintelligible.

Current anti-phishing solutions make use of established pattern recognition and blacklisting approaches but are not expected to be fully effective on their own and are widely seen as being components of an integrated solution.

Digital certificates, on the other hand, are designed to provide an effective whitelist of assured identities. There is also a clear synergy between the roles of certification authorities and traditional trusted third parties such as banks, auditors and online business information providers which means that existing certification services can readily be leveraged to provide the consumer with evidence of legitimacy and good standing in addition to mere identity.

This paper explores the reasons why this approach has never been seriously explored and proposes that a user friendly version of the padlock, providing trustworthy and helpful information about web site owners in plain language, would be a practical tool with which consumers could protect their own interests.

Empowering the consumer in this way would not eliminate the dangers of phishing but it would certainly change the nature of the risks and responsibilities.

Many organizations and agencies would have an interest in promoting such a solution.

Background

Phishing is the practice of luring the unwary to fake web sites for the purpose of either stealing their money or stealing their personal information - so that their money can be stolen later. This is made possible by the ease and speed with which false names and false branding can be put up, moved around and taken down in a virtual world and because of the difficulty of policing the dark corners of a space as vast as the Internet.

But practically all browsers implement *digital certificate* technology, which was specifically designed to identify the parties to a network connection and which is deployed ubiquitously, in conjunction with a network security protocol called SSL[i], to implement 'secure' web pages.

When browsing a secure web page the digital certificate contents can usually be viewed by clicking on a tiny padlock icon located at one corner of the browser window. But this feature is little understood and mostly overlooked.[ii] Those users who are aware of the padlock icon usually just check whether it is open or locked, the implication being that if the padlock is locked the site is safe. [iii]

However the browser software cannot tell whether the site owner details in the digital certificate match the company name and branding displayed on screen. Only the user can tell this by opening the padlock and checking the certificate. But no-one ever does.

Taking the end user's point of view it is obvious why - the display is hopelessly geeky (see Figure 1 and Figure 2) and difficult even for the knowledgeable to interpret.

Nonetheless, even today it is often possible to confirm that a site genuinely belongs to a given company (in this example a software vendor called Altova) and that the certification authority which issued the certificate (Verisign) has checked this in some way.

Figure 1 Example of valid certificate details as displayed by the padlock

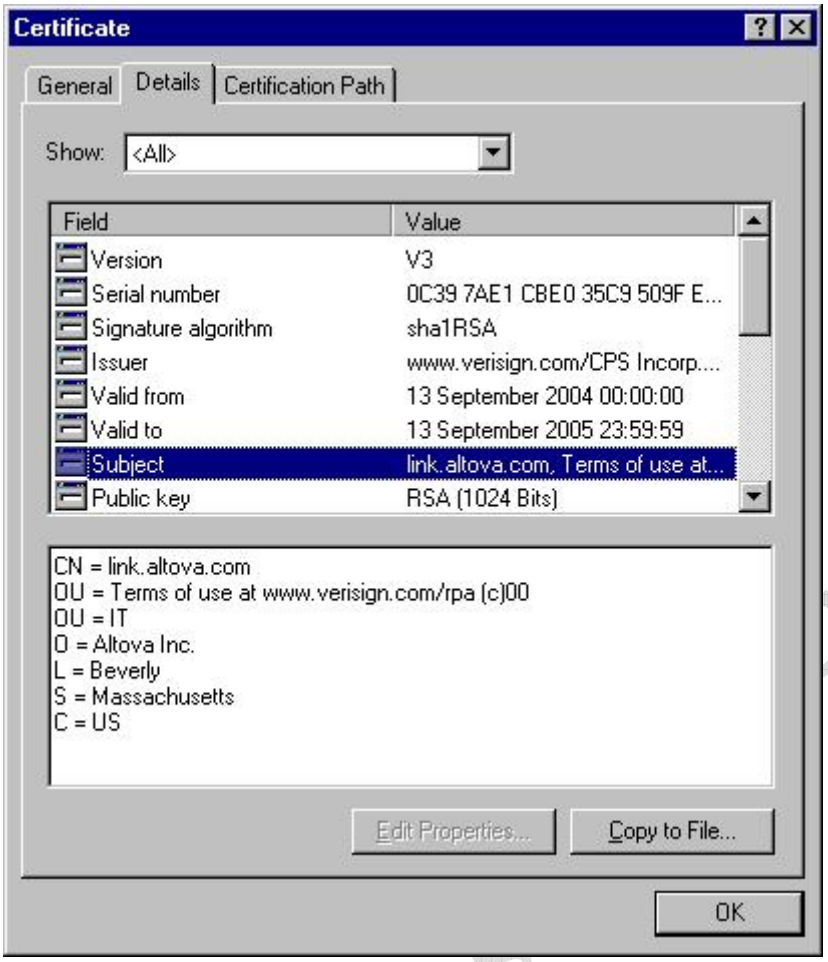


Figure 2 Further details from a digital certificate as displayed by the Padlock

The question arises as to why this functionality, which directly addresses the defining problem of phishing, has been overlooked.

Automate or bust

Wherever technology permits, an automated solution has to be preferable to one which depends on the user, as this proposal would. But it is generally accepted that there is no single technical solution to the problem of phishing and for that reason alone serious consideration should be given to the approach of helping the user to help themselves. Nonetheless, for many, involving the end user may still seem to be a feeble resort.

It is therefore important to remember whose interest this is in. Banks may bear most of the financial burden but it is always in the consumer's interest to protect their own money. Obtaining restitution for fraudulent losses is far from guaranteed and can cause considerable inconvenience to say the very least. [iv]

By the same token there is no reason to suppose that members of the public cannot make informed decisions if they are provided with the right information in a sensible format and there have been many examples of successful advertising campaigns making people aware of good safety and security practices. [v]

There is no reason why safety on the Internet cannot be promoted in the same way, especially when regulators, financial institutions and security vendors all have an interest.

The main tests which can be automated by the browser have long since been implemented, but these amount to little more than some technical checks on certificate validity plus a swift comparison of the DNS name conventionally recorded in the certificate against the URL of the web page. [vi] The rest of the certificate information is unused and many types of certificate no longer even identify the organization to which they have been issued [vii].

Unfortunately, the many public conflicts over ownership and legitimacy of DNS names attest to their unreliability as do the equally well publicized security problems with the DNS system itself. These will not be rehearsed here but suffice to say that control of a DNS name is no guarantee of rightful ownership.

In passing, it is also worth noting that a legitimate DNS name need not adequately indicate the name of the owning organization. The name www.nwolb.com [viii] belongs to NatWest PLC, a UK retail bank, but who would know this coming to the site for the first time? The site bears the bank's branding, it correctly shows customers' account details, it allows money transfers, but there is no easy way of knowing that this is not a long established man-in-the-middle attack.

The practical effect of this is that, without checking the certificate ourselves, we can be reasonably sure that a web site belongs to the same organization which controls the associated DNS name but we have no assurance that they have any right to that name. [ix] There is nothing whatsoever to tell us if the organization is legitimate.

Preferred approaches

The industry's main response to phishing has been directed towards the proven, or at least established, methods of blacklisting and filtering. Matching against lists of known, suspicious 'signatures' is the core technology of anti-virus software, intrusion detection systems, web site filtering and email compliance checking products. The same technology is now being recruited to provide anti-spam and anti-phishing solutions.

Great energy is being put into establishing global systems for scanning email to detect and prevent phishing attacks, but the central weakness with this is that it pits the intelligence which programmers are able to encode into pattern recognition software against the human intelligence of email writers and recipients. [x] This is the Red Queen's Race from Lewis Carroll's *Through the Looking-Glass* in which you have to run as fast as you can to stay still.

Another approach, mainly to the problem of unsolicited business mail, has been to press for statutory remedies. Whilst this is an entirely understandable course of action, it will surely only have limited success while the Internet is global and legislation is enacted locally.

Finally, the most effective solution at the moment is probably the extension of brand protection services, searching the Internet for spoof websites which are then shut down by the corporate lawyers. But this again is a never ending race.

Historical use of certificates

The main interest in digital certificate technology so far has been for authenticating people rather than servers. Many organizations have implemented Public Key Infrastructures (PKI) [xi] for managing digital certificates but these have been almost exclusively for the purpose of authenticating people, usually customers or members of staff.

For most Internet based businesses a server certificate is simply something which has to be purchased in order to run a 'secure' web site – a budget line item rather than a potential solution to anything.

So called *mutual authentication*, where the user authenticates the server as well as the server authenticating the user, is more talked about than acted on, largely because on internal networks, where most PKIs have been implemented, it has simply not been seen as a requirement.

Identity and Trust

For organizations to know their customers and staff is an entirely different matter than for the public to know an organization. Customers have accounts and employees are on the payroll. The organization already knows them and knows how far to trust them.

But the converse problem is not so easy. The practicalities of commercial and civil trust are understood by one set of people and the technicalities of digital identity, as mediated by digital certificates, are understood by another - in an entirely different industry.

In the real world of commerce there are many organizations and structures which enable trust and help people to understand and manage the risk of doing business. Banks provide *letters of credit*, practical guarantees to importers and exporters that the buyer will pay when the manufacturer has shipped. Banks

also give credit references, business information services give corporate histories and profiles, reputable law firms act as go betweens and accountants do independent audits. Trust in retailers is enabled largely by consumer protection regulation and enforced by trading standards organizations and the courts.

But the world of logical trust between networked computer systems relies on digital certificates which use cryptographic methods based on mathematical theory. And if this is not enough to deter most people, the mathematical concepts are far beyond simple high school mathematics. [xii]

It is not necessary to be a mathematician to understand how digital certificates work and there are many good explanations available [xiii]. The details are not relevant here, but it is worth saying that the underlying principles are often explained by reference to a world where a number of characters, conventionally called Alice and Bob [xiv], wish to exchange secret messages whilst being sure that those messages come from who they purport to come from, rather than from, say, Mallory (a malicious third party).

In this rather simple world, intended to clarify technical mechanisms rather than model business requirements, there is only one Alice and only one Bob. Bob knows Alice and Alice knows Bob. But in the real world we need to ask who exactly *is* Alice, where does Bob live and can I trust him with my money? [xv] Clearly, these questions cannot be addressed by cryptography.[xvi]

Linking the digital and real worlds

The link between digital identities and the real people and organizations is made during the certificate issuing process. It is at this point that real identity is established and digital identifiers (keys) exchanged.

The procedures used are defined in a document known as the Certification Practice Statement (CPS)[xvii] which is usually referenced in the digital certificate. A typical CPS purports to inform the 'relying party' (you, the user) exactly how the certificate authority authenticated the applicant when the request was received. But this could scarcely be less accessible. The URL might be recorded in the digital certificate, but thanks to the padlock's obscurity this is far from obvious. The documents themselves[xviii] tend to be lengthy and legalistic but light on practical detail - often promising suitable identity checks but leaving it to the reader to guess exactly what this amounts to. [xix]

As far as the author is aware, no major certificate authority ascertains the standing or legitimacy of the organizations to which it issues certificates.

Rehabilitating the technology

The result of all this is a technology and infrastructure which has only a very small intersection with the day to day realities of Internet commerce. Everything needed is there, but it belongs to the nerds.

The first step towards rehabilitating the technology is to understand the practical problem which the public faces when using a web site for the first time.

There are two possibilities - you may recognize the name of the organization or you may not. If you do, you will have some background knowledge which will enable you to decide whether to entrust it with your card details and your money. All you need from the technology is an assurance that you are truly dealing with the organization which appears on screen.

If you do not already know the organization then you have to obtain background information somehow. Some certificates help to a limited extent by giving the organizational name and country of origin. The country is particularly important because doing business across legal jurisdictions affects the practical ability to obtain restitution – one of the key ingredients of trust. Also different countries have different levels of regulation and *passing off* [xx] is easier in some jurisdictions than others. [xxi]

But to find out about the standing or legitimacy of an organization still requires some kind of research, whether formal or informal. This may be by asking other people, searching the web, or using traditional trusted third parties such as a business information providers, banks, etc.

The solution

This paper proposes that, used effectively, digital certificates could constitute a practical whitelist [xxii] of organizations which the user can consult when accessing a web site.

The padlock interface would have to be redesigned to display whatever useful information a certificate contains in sensible and intelligible form. It would also need to make it clear to the user how rigorously the web site owner has been authenticated, if at all (implying an independent analysis of the CPS).

This work could be done by browser vendors enhancing their existing padlocks or by independent software companies developing a new browser add-in. Consumer anti-virus vendors would be well placed to provide such a tool as an addition to their desktop security product suites. The tool would have to be marketed, but many different agencies would have an interest in promoting its use, not just the vendor.

Supplementary information could be provided in collaboration with the certification authorities to make up for the paucity of information currently held in digital certificates. This should start with references from the public databases used to verify the original certificate request and could be extended to include references to other reputable information sources such as stock exchange listing, professional associations, Hoovers, etc.

The typical data flows are shown in blue in Figure 3. The web site owner applies for a certificate and provides the information requested by the certification authority (CA). The CA then checks this information against appropriate public databases and, if successful, issues a certificate to the web site owner. This is sent to the browser by the web server at the start of each SSL connection.

The proposed additional information flows are shown in red. The padlock tool displays the certificate contents to the user along with any available supplementary information obtained online from the tool vendor. The user may then use those references directly to determine the legitimacy and standing of the web site owner if required.

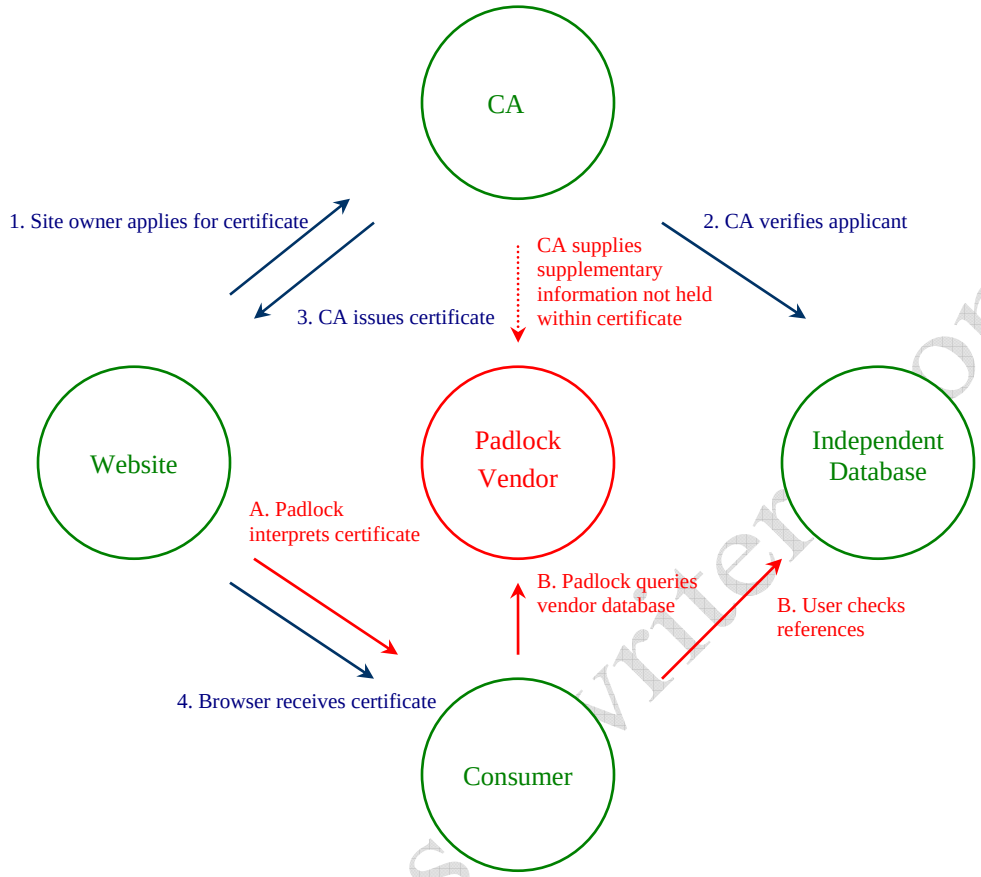


Figure 3 Information flow

Conclusion

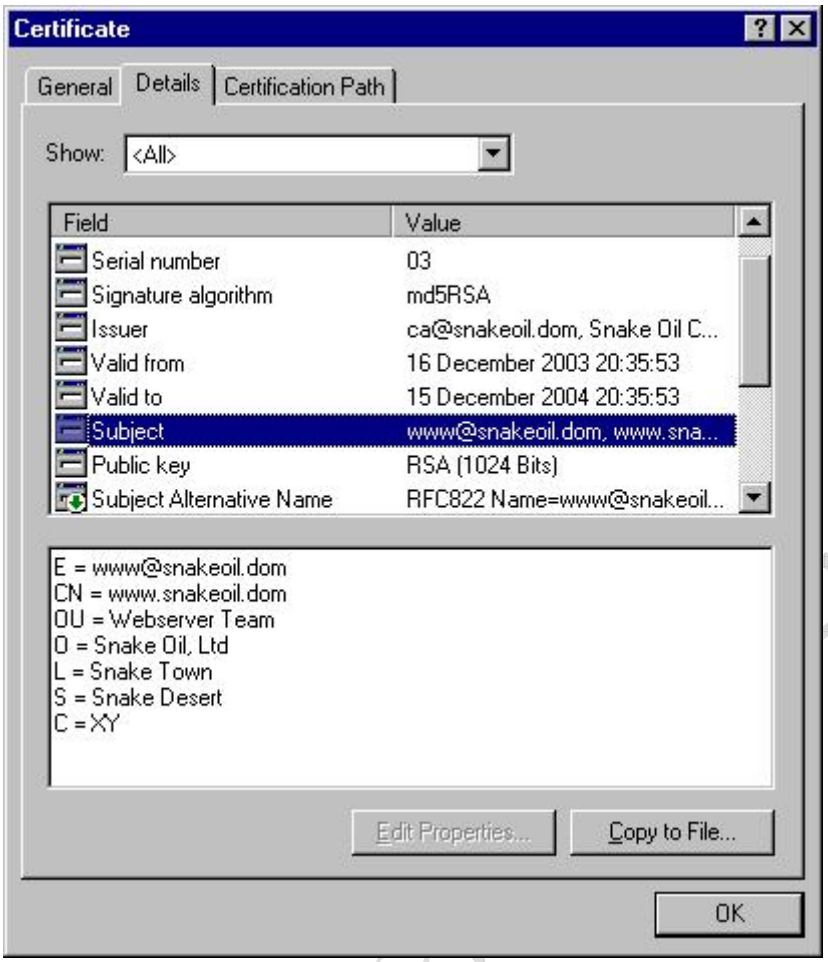
Making digital certificates more accessible to consumers could bring about a number of beneficial changes. At the very least, exposing the uneven quality of digital certificates to public scrutiny is likely to force improvement in certification practice and certificate information content.

At best, this would convert the existing stock of server certificates into an effective white-list of organizations whose trustworthiness has been assured, to a greater or lesser extent, by an alliance between digital certification authorities and traditional trust brokers. It is difficult to guess exactly how trust and commerce on the Internet would be affected but the synergy between old and new is clear.

---

[i] For convenience SSL (Secure Sockets Layer) is taken to include the newer TLS (Transport Layer Security) protocol.

[ii] To appreciate the obscurity of digital certificates and how unlikely it is that a customer would ever check the padlock, see the certificates details below which came from a web site (http://www.ellafinance.com/) which, on the face of it, appears to be an otherwise quite professionally produced advance fee fraud.



[iii] In one of the episodes in the Hitch Hiker's Guide to the Galaxy series by Douglas Adams the two heroes, Arthur Dent and Ford Prefect, find themselves trapped in the cockpit of a space craft which they have stolen. Also in the cockpit is the spacecraft's owner, an alien which has transformed itself into a carbon copy of the Ravenous Bugblatter Beast of Traal. Arthur, who is a native of Earth and inexperienced in such matters asks "Is it safe?" to which Ford Prefect replies "Yes. *It's* perfectly safe".

[iv] Claiming that someone else has used your credit card is not without danger. Banks often have to deal with false claims and there have been a few cases where prosecutions for attempted fraud have been brought against perfectly innocent and respectable customers. See Ross Anderson's account of the successful 1994 prosecution bought by the Halifax Building Society against one John Munden, a police constable who had complained that he had not made any of six ATM transactions which appeared on his bank statement. Munden served nearly four years in prison before his eventual acquittal.

[v] Everyone in the UK over a certain age will remember "Clunk click every trip!" the slogan used to remind motorists to fasten their seatbelts.

[vi] WholeSecurity market a browser plug-in called Web Called-ID which performs a real-time analysis of pages displayed in the browser to detect the exploits typically used by phishing sites. But even with the name Caller-ID the emphasis seems to be more on automated testing for suspicious behaviour (more pattern matching and blacklisting) than on positive confirmation of trustworthiness. So the product itself does not change the thrust of this argument, although given sufficient uptake it should certainly raise the bar for phishers.

[vii] See also What's the point of security if you do not know who its for by Dr Colin Walter, Head of Cryptography, Comodo Inc.

[viii] Presumably www.nwolb.com stands for NatWest Online Banking

[ix] There are other relevant aspects of security beyond the scope of this paper which materially affect the decision to enter information at the keyboard. The most significant of these are the questions of who really controls the user's PC and who really controls the web server. The former is set to become a serious problem, if it is not already. The latter is actually addressed by some site certification services whose information could usefully be incorporated into the padlock enhancements proposed here.

[x] Being somewhat skeptical, the author believes that anti spam filters will finally succeed the day a regular expression passes the Turing Test.

[xi] Digital certificates are signed by a certificate issuing authority. The authority's own certificate may be signed by a high level authority and so on, up the chain to a so called 'root' certification authority which signes its own certificates. This structure and the technology which support it is collectively known as a PKI.

[xii] The cryptography which underlies digital certificates relies on certain types of problem which mathematicians rather charmingly refer to as *hard*.

[xiii] See the RSA Crypto FAQ for a detailed explanation of cryptographic concepts and their application.

[xiv] Alice, Bob and other characters are introduced at http://en.wikipedia.org/wiki/Alice_and_Bob

[xv] More information on the private lives of Alice and Bob can be found here

[xvi] To add to this confusion, product marketing plays on the metaphysical connotations of the problem. In the world of 'security solutions' it seems that we are not dealing with mere names and credentials. We are dealing with *identity* itself. It is not just that someone could appropriate our bank account and steal our money. We apparently risk the fate of the Schwarzenegger character Adam Gibson in the movie 6th Day, who returns home to find that a clone has taken his place and is living in his house with his wife and children, while he hides in the garden wondering who he truly is.

[xvii] There is another formal document called the Certificate Policy which describes how the certificate is intended to be used and how the key should be protected. Like the CPS this may be referred to in the digital certificate.

[xviii] See the Baltimore CPS for a typical example of a certification practice statement.

[xix] A well publicized incident occurred in 2001 when Verisign issued two Microsoft certificates to an impostor. It seems that the certificate request was taken to be legitimate on the basis that valid corporate credit card details were provided in payment. These turned out to have been stolen. See "Microsoft Vexed by Falsified Certs" at http://www.theregister.co.uk/2001/03/23/microsoft_vexed_by_falsified_certs/

[xx] *Passing off* is the term used where one business pretends to be another for the purpose of stealing custom.

[xxi] It would be a bad mistake to suppose, as some certification authorities appear to imply, that being officially registered somehow makes a company legitimate. Companies are cheap and simple to set up in most legal jurisdictions and bogus companies are often created as part of a fraudulent enterprise. By contrast, listing on stock exchanges can usually be taken to imply good standing.

[xxii] Discussion of digital certificates can get bogged down on the subject of *certificate revocation*. If for some reason a valid certificate can no longer be used, say because the key information has been stolen, it is recorded in a *certificate revocation list*. In theory this blacklist should be consulted every time a certificate is verified but although the technology exists the implementation is patchy. With server certificates this is rarely a problem precisely because almost no-one relies on them to authenticate a server. Raising the visibility of server certificates would certainly change this and force a practical solution to the problem.