Who are They?
Website Authentication: Certificates and Identity

Milica Stojmenović
Carleton University
Ottawa, Canada
Milica.Stojmenovic@carleton.ca

Robert Biddle
Carleton University
Ottawa, Canada
Robert.Biddle@carleton.ca

ABSTRACT
Users expect to authenticate to websites, but not for websites to authenticate to them. This is the root of phishing and other inducements for users to visit imitation websites, where their credentials are captured for fraudulent use. The limited but potentially useful method for addressing this problem involves website certificates. We have been exploring ways to help users understand these to ensure they interact with only genuine sites. However, recent efforts for "encryption everywhere" are making this more challenging by making certificates without identity easily available and encouraging users to look simply for https. The result is that users trust *all* websites with https, when they shouldn't; websites that do offer identity have no advantage. The laudable promotion of encryption, combined with poor browser support, is making website authentication more difficult.

1. INTRODUCTION
Web users are familiar with authentication. The use of usernames and passwords is almost universal, certainly where websites maintain sensitive user information, and often simply to help websites monitor usage and maintain contact with users. The relationship between websites and users, however, is asymmetric. Users are required to register and authenticate to websites, but websites are not required to authenticate to users. This is the vulnerability that allows phishing and other forms of fraud, where users mistake fraudulent sites for genuine ones, with a range of bad consequences. Of course, websites and users are in a one-to-many relationship, so full mutual authentication is not easily done, especially without significant and burdensome infrastructural change. We suggest that a limited but usefully effective approach is already available: website certificates that carry verified identity information. This approach, however, has in the last year become problematic for a surprising reason: the recent emphasis on "encryption everywhere" is resulting in website identity becoming obscured.

Website certificates connect two human efforts with cryptographic technology, yielding a socio-technical infrastructure. Certificate Authorities (CAs) will actually investigate website operator identity, looking up records and even making phone calls, before issuing certificates that cryptographically encode these details to websites. Browsers can then make this information available to users, who are then able to assess whether the website is genuine. There are, however, several problems. Most starkly, many users are unaware this infrastructure even exists, or have only a weak understanding of how it works. It is therefore unsurprising that users pay little attention to browser indicators [8].

Of course, certificate infrastructure is not a perfect answer to fraudulent websites. There are many CAs and they may face difficulties[1], and websites may be corrupted by attackers, yet retain valid certificates. But certificates have important advantages. Perhaps most importantly, they can identify websites as being operated by genuine organizations in known jurisdictions, and this information can be conveyed by the browser independent of (potentially fraudulent) website content. Of course, the certificate doesn't say whether an organization is trustworthy, but rather that they are who they say that are. The mechanism works instantly, whereas schemes that require fraudulent sites to be detected and reported can take days, while fraudulent sites are typically short-lived[2] and now often use https as well[3]. We feel the certificate infrastructure is worth maintaining and improving. We are especially interested in better notifications and in education, and are studying such improvements [13-14].

To our surprise, however, another effort at improving website security is making the challenge more severe. The laudable emphasis on promoting the use the https protocol is to ensure communication between browsers and websites is encrypted and cannot be read or manipulated in transit. This encryption involves certificates, but in order to make them widely and easily available, and even free of charge, the identity validation is minimal. In particular, only domain validation is done, confirming that the certificate is issued to those who control the website at the indicated domain. Users are encouraged to look for https as an indication of security, but now fraudulent sites also feature https. To support wider encryption, we are in danger of undermining users' ability to determine identity.

The next section discusses certificate types and relevant studies that have suggested alternative designs for certificate interfaces. Then, we discuss frameworks we use to explain our position: mental models, third-party advice, and the Judge-Advisor System.

Who Are You: WAY 2018 Workshop at USENIX Symposium on Usable Privacy and Security (SOUPS) 2018. August 12 – 14, 2018, Baltimore, MD, USA.

[1] https://scotthelme.co.uk/are-ev-certificates-worth-the-paper-theyre-written-on/

[2] https://www.infosecurity-magazine.com/news/84-of-phishing-sites-last-for-less/

[3] https://info.phishlabs.com/blog/quarter-phishing-attacks-hosted-https-domains

Following this, we present results on the prevalence of each type of certificate of the most popular websites.

2. CERTIFICATES

2.1 Certificate Infrastructure

The Secure Sockets Layer (SSL) and its successor, Transport Layer Security (TLS) use asymmetric cryptography to support both in-transit encryption and some assurance of identity using X.509 certificates. CAs are third-party organizations, who create and issue certificates to be used by websites. The certificates support both encryption and identity at various levels. Identity is supported at different levels according to the processes used by the CA to confirm the identity of the website.

Domain Validated (DV) Certificates merely confirm that the certificate is issued to someone that does indeed control the website, typically determined by a challenge-response process. **Organization Validated (OV)** adds details of the website organization (e.g. company registration), confirmed by CAs. **Extended Validation (EV)** takes this process further and can involve determining geographic location of the organization, confirming contact details, and so on.

Websites store certificates and allow them to be accessed by browsers. Browsers have a critical role because when a user accesses any website with the https protocol, the browser first retrieves the certificate. The browser determines whether the certificate was issued by a recognized CA (itself a cryptographic process). If not, the browser will either disallow access, or require the user to confirm an exception. If the CA is recognized, the browser then uses the certificate to establish encrypted access to the website, and presents identity information to the user.

Fig. 1 shows certificates on a desktop computer (mobile devices are out of scope). Twitter has an EV (top of Fig. 1). As can be seen, the identity (Twitter, Inc.) is presented emphatically to the left of the URL, along with company's national registration jurisdiction (US). Facebook has an OV, which in Google's Chrome browser appears as shown in Fig. 1 in the middle.

No identity information is shown without detail exploration using browser menus. Dior has a DV, which in Google's Chrome browser appears as shown in Fig. 1 at the bottom, and no identity information is available, although this is nowhere made clear. Later versions of Chrome intend on removing the "secure" [4] wording but this still makes DV and OV indistinguishable.

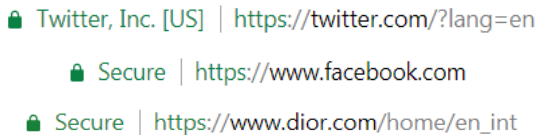

Fig. 1. EV, OV, and DV in Chrome

Our interest is in helping users understand that certificates with identity offer a way to authenticate sites as being what they claim, but this design is challenging in several ways. The EV presentation is a good start, but is not self-evident since users may not understand, and nothing is there to explain: the tool-tip simply says "view site information", and when clicked only claims the site is secure. The OV and DV indicators specifically *say* "secure", and may make users wonder if the EV site is not. And while OV sites do provide identity, it is deeply hidden, and thus OV and DV sites are effectively indistinguishable. But fraudulent website with a free DV certificate would look the same as a valid OV site. For example, secwww.com/facebook (owned by us) shows a valid DV cert. Moreover, free DV certs are being encouraged, even for e-commerce sites. Recently Shopify[5], a service that helps vendors create websites, promotes the fact that it provides DV certificates for all sites.

2.2 Certificate User Studies

An influential study [8] showed that most users do not notice or interact with certificate indicators. Another study [4] investigated how to improve indicators and then implemented the findings in Google's Chrome browser. An important aspect of these studies is that there was no specific focus on *identity*. For example, the second study [4] emphasized having a certificate, thereby offering an encrypted connection, but there was little mention of identity to help avoid fraudulent sites.

There have been studies that did address identity, especially since the introduction of EVs. One study [9] compared the certificate indicator interfaces used by Mozilla Firefox with their own proposed redesign with users. They found that the then-current indicator was too subtle: no one noticed it. Fifteen (out of 28) participants claimed to notice the indicator on the proposed redesign and three included the indicator in their decision making. No participant attempted to interact with the indicator and thus did not see more information. Twenty-two participants preferred the redesign since it was more noticeable and provided information without interaction.

Another study [1] compared the usability of a new interface for EV to the existing one, by examining which features of the interface users could understand: whether the different levels of authentication were clear and if the users could distinguish between website identity and encryption. They found that the existing interface used technical terms, unfamiliar to the typical user. The results also showed that participants could draw the correct conclusions using the alternative design. They correctly determined the ownership of the website and the privacy of transmitted data. Participants were also more certain about their decisions using the alternative design.

3. THEORETICAL OVERVIEW

3.1 Mental Models

Mental Models are a combination of our perceptions and ideas [10]. They help us make sense of our surroundings. We suggest a desirable mental model for certificates in Fig. 2 [14]. In this model, users interact with the Internet through the browser. They load a website that appears to belong to the intended organization

[4] https://blog.chromium.org/2018/05/evolving-chromes-security-indicators.html

[5] https://www.shopify.ca/blog/

– how can they be sure? Certificates can confirm these details, shown by browsers and issued by trusted CAs.

Mental models are developed over time, as our experiences diversify and we learn more about our environments. Third party experiences are also helpful when learning what we should and should not do, and can help us make better decisions online.

3.2 Trusted Third-Party Advice

Most online consumers actively seek out and accept third-party advice [11]. When purchasing online, 97% of users rely on feedback before purchasing from an unfamiliar seller [7]. A review system is also in place in hotel booking [15] and movie box office [3] systems. Third-party reviewers are regarded as the most credible, objective, and influential since they are seen as unbiased [6]. This is not always the case, as corporations do seed messages out through influencers (ex. paid reviews). A third-party guarantee is relied on in the Judge-Advisor System (JAS), a well-studied and accepted model for decision-making, discussed next.

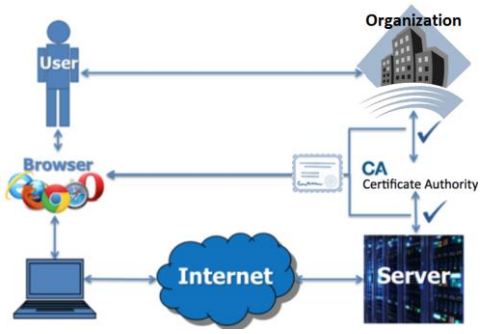

Fig. 2. Key players and process of website identity determination.

3.3 Judge-Advisor System (JAS)

According to the JAS, people seek advice from other, independent, people that have more experience with the topic [12]. There are two parties involved: people looking for advice and those experienced enough to give it. The Advisor's role is to give suggestions to the judge, who then makes the decision. Regardless of the source and credibility of the Judge, helpful advice is often taken when users are deciding on downloading potentially unsafe software [5]. If we apply JAS to certificates, then the person seeking advice before a decision is a user and the advisor is a CA.

4. CERTIFICATE LANDSCAPE

To investigate the certificate landscape, we obtained lists of the most popular websites, and then looked at the certificates they use. For lists of popular websites, we used the Alexa service API in autumn 2017. The popularity of websites is typically influenced by the geographic location of the user, so we retrieved a list restricted by country, beginning with our country, Canada, and retrieved the top 1,000 used domains. We then created scripts using OpenSSL to access the certificates used by each domain. To determine whether a site was OV, we looked for information identifying the owner; to determine which were EV, we looked at the policy information and checked for an Object Identifier (OID) indicating to a browser whether they used extended validation. We also looked for website redirection, where the domain requested by the user is replaced by the site itself, and where that happened we used the replacement to access the certificate.

Within the top 1000 sites used in Canada, we found that there were 106 EV certs, 488 OV certs, 301 DV certs, and 105 sites with no cert. Roughly speaking: *10% EV, 50% OV, 30% DV, and 10% without*. From these numbers alone, we can draw some conclusions. For example, approximately 60% have certificates that provide identity information, which might seem encouraging.

One potential caveat is that only 297 use the entered domain directly: the others all redirect. Most redirect to subdomains (especially www), but 35 redirect to other domains.

Table 1. Top 20 websites with EV.

Rank	Website	Organization
9	twitter.com	Twitter
20	td.com	The Toronto-Dominion Bank
28	t.co	Twitter
41	apple.com	Apple Inc.
44	bmo.com	Bank of Montreal
46	wordpress.com	WordPress
49	bestbuy.ca	Best Buy Co
52	github.com	GitHub
58	tripadvisor.ca	TripAdvisor LLC
61	buzzfeed.com	BuzzFeed
64	dropbox.com	Dropbox
74	airbnb.ca	Airbnb
76	rumble.com	Rumble Inc.
77	chaturbate.com	Chaturbate LLC
79	steamcommunity.com	Valve Corp.
85	theglobeandmail.com	The Globe and Mail Inc.
142	battle.net	Blizzard Entertainment
148	steampowered.com	Valve Corp.
151	gouv.qc.ca	Gouvernement du Quebec
158	shopify.com	Shopify Inc.

This can be reasonable, even for non-subdomains, where national domains redirect to more general ones or vice-verse (e.g. google.com redirects to google.ca) or older domains redirect to newer ones (e.g. blogspot.com redirects to blogger.com). Without identity information, and with weak understanding of subdomains, users have little support to determine whether redirection is legitimate.

With only EV sites supplying users with identity information in Google Chrome, it is interesting to consider which sites have EV certs. Looking at the 106 EV sites from our sample, most represent well-known organizations. The top 20 are shown in the Table 1. We can see a mix of international Internet services (e.g. Twitter, Wordpress, Github) and Canadian businesses (e.g. BestBuy, Globe and Mail), and banks (e.g. Toronto-Dominion Bank, Bank of Montreal). It is a diverse set, and both government sites and "adult" sites are included. Perhaps more interesting are the sites that are at the top of the popularity list, yet do not have EV certs. We show the top 20 in the table below. It is an impressive list, including Google, YouTube, Facebook, Reddit, Wikipedia, and Amazon. All these sites, and many others down

the list, have only OV certs, and therefore do have identification, but none easily available to the user in Google Chrome.

All the discussion above relates to the most popular websites in Canada. We also looked at other countries. We used a list of countries with the largest number of Internet users, beginning with China, India, and so on. We then applied the same procedure, first using Alexa to find the most popular 1000 sites in those countries, and finally using a script with OpenSSL to retrieve and categorize their certificates. The results are shown in Fig. 3, shown at the bottom. As can be seen, the results do vary, for example with the United Kingdom, the United States, and Germany having the highest proportion of EV certs, while China and Iran having the lowest. But the general pattern is surprisingly consistent. Taking the mean proportions across all 21 countries, we have 7% EV, 32% OV, 42% DV, and 19% with no certificate.

Table 2. Top 20 websites and their certificates.

Rank	Certificate Type	Organization
1	google.ca, OV	Google Inc
2	youtube.com, OV	Google Inc
3	google.com, OV	Google Inc
4	facebook.com, OV	Google Inc
5	reddit.com, OV	Reddit Inc.
6	wikipedia.org, OV	Wikimedia Foundation
7	amazon.ca, OV	Amazon
8	live.com, OV	Microsoft Corporation
9	twitter.com, EV	Twitter
10	yahoo.com, OV	Yahoo! Inc.
11	netflix.com, OV	Netflix
12	kijiji.ca, OV	eBay
13	instagram.com, OV	Facebook
14	imgur.com, OV	Imgur
15	diply.com, OV	GoViral Inc.
16	amazon.com, OV	Amazon.com
17	linkedin.com, OV	LinkedIn Corporation
18	twitch.tv, OV	Twitch Interactive
19	pornhub.com, OV	MG Freesites Ltd.
20	td.com, EV	The Toronto-Dominion Bank

While only small proportion of websites have EV certificates, many other websites do have OV certificates that carry identity. The problem is that through browser design and the newly widespread use of free DV certificates, the advantage – to websites and to users – is lost.

5. DISCUSSION

In the previous sections, we have shown how the certificate infrastructure *should* be able to help users detect imposter sites and shown that many websites already have the certificates necessary. The next step relates to the browser interface, and how well it supports users in this process. For example, one issue is that users need to distinguish advice from browsers (and so the CA) from anything shown on a webpage (which can be manipulated by fraudulent websites). But there are several other issues.

As we showed in Section 2, the interface of Google Chrome makes it clear when a site has an EV cert, and shows the organization name and jurisdiction clearly. Other major browsers do similarly, but we focus on Chrome as it is now by far the most dominant browser for individual use[6].

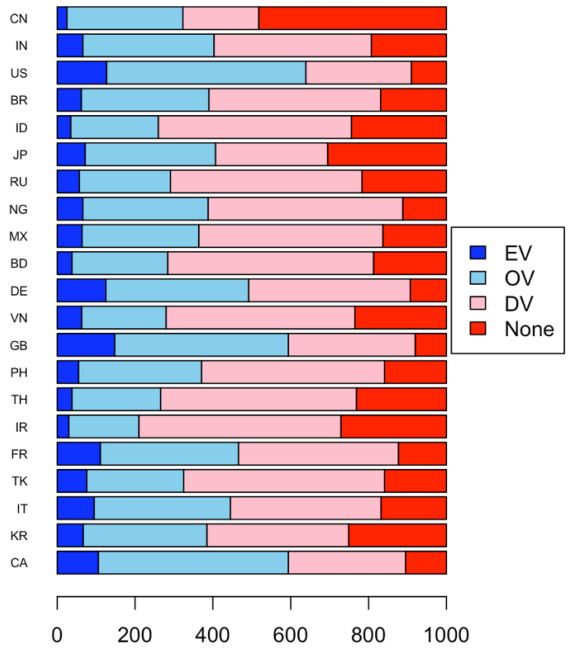

Fig. 3. Certificate types for top 1000 used websites in the 21 countries with most Internet users.

As shown in Section 4, however, EVs are a small proportion, whereas OVs are much more common, and also carry identity information. The typical interaction design to make this information accessible is for the user to click an indicator, such as the lock symbol, whereupon the identity information can (eventually) be seen. For some browsers this is not easy for users, especially because technical details of no concern to most users make the identity hard to notice. Google Chrome required several steps for the identity to be detectable by clicking the indication, then the certificate, then the details, and then noticing that an Organization is shown. For a DV, the user has to notice the *absence* of an Organization. For most of the 2017 the situation was worse still, and certificate information was not available by clicking on the indicator all: users had to navigate other menus. In 2018 this was corrected to the procedure described above, but it is still insufficient. Mozilla Firefox, the 2nd most popular browser, is in some ways worse: for OVs, clicking on the indicator and "more information" yields "This website does not supply ownership information", when in fact is does – as yet a further click would reveal, if one knew what to look for. There is also clearly a need for better literacy among web users, and there are many stakeholders that should be concerned. In particular, we notice that CAs focus their attention almost exclusively on websites, and

[6] http://gs.statcounter.com/browser-market-share

ignore users: this means that users do not understand the value of the CA's work, or even recognize their names.

This is the status quo: the most dominant browser only makes identity easily accessible for sites with EV certs – only about 10% of all sites. All sites with OV certs are shown in a manner identical to those with DV certs, which carry no information about the website's organization. Another recent development makes the situation more problematic. In an effort to increase browser-website encrypted communication, DV certs are now free to obtain, through services such as *Let's Encrypt* (letsencrypt.org). In itself this is laudable, but a consequence is that imposter sites can now obtain for free certificates that in Google Chrome are effectively indistinguishable from legitimate sites.

6. CONCLUSION

Fraudulent websites that mimic familiar legitimate sites are a severe threat and can be used to capture credentials and manipulate communications. Many ways for users to detect such sites are only heuristic, and do not work for high-quality attacks. We suggest that the existing website certificate infrastructure is the right approach to authenticate identity. It is a socio-technical system, linking the efforts of Certificate Authorities to investigate real-world identity in jurisdictions, and the cryptographic technology that makes identities easy to check and difficult to forge. Moreover, the basic model involved has been established for some time as JAS: when people need to make judgements on important topics, they seek expert impartial advisors.

Website certificate infrastructure is not new and is not perfect. However, we feel it is under-appreciated, and can be improved. In this paper, we have outlined the status quo. An important strength is that many websites already have certificates that confirm identity. An important weakness that needs attention is that users may not know the infrastructure exists or how to use it. However, before such an effort is worthwhile, we also need to consider the supporting elements that need to be in place. In particular, both CAs and browsers have roles to play. For CAs, we suggest they need to educate the public, not just website owners, about their role. For browsers, they need their design to make website identity easily available, understandable, and distinguishable.

7. REFERENCES

[1] Biddle, R. Sobey, J., Whalen, T., Oorschot P. V., & Patrick, A. (2009). Browser Interfaces and Extended Validation SSL Certificates: An Empirical Study. In ACM workshop on Cloud Computing Security.

[2] Dhamija R., Tygar, J. D. & Hearst. M. (2006). Why phishing works. In Proc. of the *SIGCHI Conference on Human Factors in Computing Systems (CHI '06*), R. Grinter, T.s Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson (Eds.). ACM, New York, NY, USA, 581-590.

[3] Duan, W., Gu, B., & Whinston, A. B. (2008). The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry. *Journal of Retailing, 84*, 233–242.

[4] Felt, A. Porter, Reeder, R.W., Ainslie, A., Harris, H., Walker, M., Thompson, C., Acer, M. E., Morant, E., & Consolvo, S. (2016). Rethinking Connection Security Indicators. In *SOUPS*, pp. 1-14.

[5] Freitas, B., Matrawy, A., & Biddle. R., (2016). Online Neighborhood Watch: The Impact of Social Network Advice on Software Security Decisions. *Canadian Journal of Electrical and Computer Engineering*, 39(4), 322-332.

[6] Kamins, M. A., Folkes, V.S., Perner, L., & Kamins. M. A. (1997). Consumer responses to rumors: Good news, bad news. *Journal of Consumer Psychology*, 6(2), 165-187.

[7] Pavlou, P.A. & Dimoka. A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4), 392-414.

[8] Schechter, S. E., Dhamija, R., Ozment, A., & Fischer, I. (2007). The Emperor's New Security Indicators. In Proceedings of the IEEE Symposium on Security and Privacy (SP '07). IEEE Computer Society, Washington, DC, USA, 51-65.

[9] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In Proceedings of the 3rd symposium on Usable privacy and security (*SOUPS '07*). ACM, New York, NY, USA, 88-99.

[10] Sinreich, D., Gopher, D., Ben-Barak, S., Marmor, Y., & Lahat, R. (2005). Mental models as a practical tool in the engineer's toolbox. *International Journal of Production Research*, 43(14), 2977–2996.

[11] Smith, D., Menon, S., & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interacting Marketing*, 19(3), 15-37.

[12] Sniezek J. A. & Buckley, T. (1995). Cueing and cognitive conflict in judge–advisor decision making. *Organizational Behavior and Human Decision Processes*, 62, 159–174.

[13] Stojmenović, M. & Biddle, R. (2018). Hide-and-Seek with Website Identity Information. In proc. Privacy, Security, & Trust, Belfast, UK, in press.

[14] Stojmenović, M., Oyelowo, T., Tkaczyk, A., & Biddle, R. (2018). Building Website Certificate Mental Models. Persuasive Technology, April '18, Waterloo, Canada. Lecture Notes in Computer Science, 10809. Springer, 242-254.

[15] Ye, Q., Law, R., & Gu, B. (2009). The Impact of Online User Reviews on Hotel Room Sales, *International Journal of Hospitality Management*, 28(1), 180–182.