Understanding the Role of Extended Validation Certificates in Internet Abuse

Brendan Saltaformaggio and Maria Konte
brendan@ece.gatech.edu, mkonte@gatech.edu
June 28, 2019

Abstract

Extended Validation (EV) Certificates play an instrumental role in web security. EV certificates assure the visitors of a website that they are indeed visiting the safe website they intend to, and not an imposter set up by cybercriminals. Previous work [1] has shown that domains who invest in EV certificates are prudent with cyber security practices, and these domains were not found to be associated with phishing sites. Additional previous work [2] on the association between EV certificates and abused domains, motivated us to perform a large-scale in-depth study to investigate and understand any such associations. We cross-correlated abused domains found in our corpus of malware network traffic, blacklists, and underground marketplace communications with domains that have EV certificates. We found that the probability that a domain with an EV certificate is abused or associated with cybercrime is negligible. We found overwhelming evidence that EV certificates are highly indicative of a legitimate domain registered by a legitimate business. This reinforces the notion that browsers should generally err on the side of trusting a website which has invested in an EV certificate, and this trust is the primary benefit that EV certificates provide to their owners. Our future work focuses on designing new security indicators for the browser that better communicates a website’s trustworthiness [3].

1 Introduction

Modern browsers have the capability to inform users whether the website they are visiting is likely safe or potentially malicious. The users rely on browser warnings for an indication of safe browsing. One of the most important browser safety indications that users trust is the existence of SSL certificates for the website they are visiting [4]. As a result, the SSL Certificate providers advertise the effectiveness of SSL certificates, EV certificates in particular, by emphasizing that when users see the green lock icon they feel reassured that they are accessing a safe website.

In this research, we aim to thoroughly study the role of EV certificates in Internet abuse. Previous work [1] shows that domains with EV certificates are safe and they were not found to be associated with phishing sites. Following that study, additional previous work [2] on the association between EV certificates and abused domains found that some connection may exist. This motivated us to perform a large-scale in-depth study to investigate and understand any such associations.
We performed a joint analysis of a unique combination of data sources to investigate if domains with EV certificates are associated with cybercrime. We found that the probability that a domain with an EV certificate is abused or associated with cybercrime is negligible. More specifically out of a sample of approximately 2.6 million domains with EV certificates, we only found 3 domains to be associated with cybercrime, and 379 to be among the domains in our blacklists/malware feeds. Overall, our findings suggest that domains which invest in EV certificates are prudent with cybersecurity practices and highly unlikely to be associated with cybercrime or abuse.

2 Background

Secure Sockets Layer (SSL) and SSL Handshake. SSL encrypts data sent between a client and server or between two servers, preventing cyberactors from sniffing and modifying data. SSL is a standard that creates a secure communication tunnel between a browser and server. SSL uses both public and symmetric key encryption. The public-key encryption is used to agree upon a symmetric key between the two entities. Once agreed, the symmetric key is used to encrypt and decrypt data. The first step for SSL is to perform a SSL handshake connection. The browser will send the server a HELLO message, which will contain a list of browser-supported symmetric-key algorithms. Once received, the server will reciprocate by selecting its preferred symmetric algorithm. The server will also send a SSL certificate. A certificate will contain the server’s ID, public key, and metadata. A certificate authority (CA) provides these certificates which give the browser assurance that they are exchanging information with the correct entity. The browser will validate that the certificate came from an authorized CA. Once validated, the browser will encrypt a symmetric key with the server’s public key and send it to the server. The server will decrypt the agreed-upon symmetric key and will now use this key for encrypting and decrypting data back and forth.

Types of SSL Certificates. The three standard levels of SSL certificates are: (1) Domain Validation (DV): DV certificates provide the lowest assurance for clients. Registration is automated and simply checks that the domain name is registered. Confirmation is done via email response and by setting up a DNS record for the requested website. The processing time is very fast and only takes a few minutes to a few hours. (2) Organization Validation (OV): OV certificates provide higher assurance for clients. Registration is not automated and requires real people to validate the requesting domain. Additional information is needed which includes name, city, state, country, and documentation to verify the registering entity’s identity. The processing time takes longer than DV, and it may range from a few hours to a few days. (3) Extended Validation (EV): EV certificates provide the highest assurance for clients. The CA checks that the requesting business is a legal entity, and the validation requires sufficient disclosure of business information to perform this verification. There is an additional human intervention where the entity is contacted via phone to verify their identity. The processing time may range from a few days to a few weeks [5].

EV Certificate Authentication Process. The certificate vetting process, according to [6], includes a seven stage authentication process that takes place with the following steps. Given the rigor and information disclosure involved, it may be assumable that cybercriminals would not be willing to go through the vetting process to acquire an EV certificate.
1. **EV Enrollment:** The enrollment procedure verifies that the applying person is indeed an employee of the organization, and he or she is authorized to be proceeding with the certificate purchase.

2. **Organization Authentication:** The second step verifies, via government registration information, that the applying company or organization is a legally registered entity and that it is active in the registered location.

3. **Operational Existence:** This step further verifies that the organization has been in existence for three or more years. If not, then additional documents must be required. This step in particular aims to prevent cybercriminals from quickly setting up shell companies to obtain EV certificates.

4. **Physical Address:** The CA verifies that the organization has a real physical address in the country that it is registered in.

5. **Telephone Verification:** The CA verifies that the organization’s telephone is a working phone number.

6. **Domain Authentication:** The CA verifies that the organization is the rightful owner of the registering domain.

7. **Final Verification Call:** The CA calls the applying organization contact to verify the EV application.

### 3 Methodology and Data Collection

**Data Collection and Pre-Processing.** The collection process is implemented using a modular series of scripts written in Python that are tailored to the unique requirements of each data source we are collecting from. After we collect the data, then we develop our pre-processing scripts to parse the data and represent them in a JSON format.

**Data store.** We import all the data sets into an ElasticSearch instance. For each dataset we build appropriate indices so that we can cross-correlate it with the rest. This framework provides us with the flexibility to easily extend our data sets or import and continuously update our new findings. Each data store has unique requirements, and in the interest of flexibility throughout the course of this study a unique though generally generic API was implemented to handle the insertion of processed data. For the common use case, the data store module parses pre-processed data and writes it to the ElasticSearch using its query language. We use ElasticSearch as a horizontally scalable, and efficiently searchable data archive. Due to its distributed nature, it provides the flexibility necessary to grow easily with the data collected as needed. Since it ingests schema-free JSON documents, it was ideal for data collected from unique sources with equally unique schemas. However, the real advantage is that as we grow a large archive of historical data, it provides a full-text capable search engine powered by Lucene.

**System Interface.** We connect to the system using PKI-enabled SSH connections and implement local-to-remote port forwarding of the web UI and protocols used by the data store. All further interaction with the data is done either through the built-in web UI’s for
the data store or via the terminal. This is likely to change at a later date as the project matures.

**Data Analytics.** For each blacklisted and cyberactor associated domain we identify if it has an SSL certificate, and we pull the certificate if one exists. For the domains that we identify to be malicious and also have an SSL certificate, we collect the organization’s information from PrivCo and Mergent data sources, described further below.

### 3.1 Data Sources

**Underground Marketplaces and Forums:** We have access to a private real-time feed of cyberactor activities in underground marketplaces and forums. This contains posts made by cybercriminals along with information they disseminate themselves: the cyberactor’s name, social media accounts they claim to own, infrastructure data (domains, IPs, prefixes, ASes) they claim to own, and services they advertise (bulletproof hosting, malware kits, DDoS service, fast flux service, etc). We also have access to the historical data of this feed which goes back to 2014. From this feed we extract all of the domains and we add them to our list of malicious domains. We collected a total of 3998 domains associated with cyberactors.

**Global Repository of SSL Certificates:** We have access to an academic source of a global repository of SSL certificates. This feed indexes more than a billion certificates and more than a million certificates are added on a daily basis. For a specific domain we can pull the related SSL certificate information if that exists. At the time of experimentation, there were a total of 2,604,344 domains with EV certificates.

**Blacklists.** We collected domain blacklists for domains observed hosting advertisement and malicious activity from the Squidblacklist and SANS Suspicious Domains blacklists. These lists are updated daily and the main purpose of this feed is to provide data to web filtering platforms. We collected a total of 90,953 domains from the blacklists.

**PrivCo and Mergent Intellect databases:** We have access to two business databases: (1) Reference USA and (2) Mergent Intellect. Mergent Intellect contains data of more than 70 million US companies and more than 200 million global companies. One of the criteria for obtaining an EV certificate is that the organization must be legally recognized and active through government (U.S. or International) databases. In our analysis, these datafeeds provide us with organization information for domains that have an associated SSL certificate. Our intuition is that by following this information we may identify patterns about how malicious domains register for these higher level certificates.

**Large-Scale Malware Feed:** At Georgia Tech, we have access to a large malware repository, that contains binaries, static analysis results, and network traffic that is generated in the first several minutes of each binary’s execution. From this feed we extract the domain names that appear in the network traffic.
4 Findings

Based on the techniques presented above, we were able to compute a number of metrics to investigate if there is any association between EV certificates and cybercrime.

Firstly, we measure the value that legal entities gain from obtaining an EV certificate. Secondly, we discuss our any findings for EV certificate holding websites hosting malware or being used in cybercrimes. Lastly, for the small fraction of domains that we found to have any connection with cybercrime, we track the specific cyberactors who were actively involved with underground marketplaces and forums or specific cybercrime planning/orchestration activities.

4.1 Correlation Between EV Certificates and Legal Entities

Based on our Internet wide repository of SSL certificates, we observed a total of 2,604,344 domains with EV certificates. We found only a small fraction of these domains to be blacklisted or associated with cybercrime activities, which we will discuss next. Overall, we found overwhelming evidence that EV certificates are highly indicative of a legitimate domain registered by a legitimate business. This reinforces the notion that browsers should more generally assume that a website with an EV certificate is trustable, and this trust is the primary benefit that EV certificates provide to their owners.

4.2 Blacklisted Domains Holding EV Certificates

As a first step, we aimed to measure the number of EV certificates among known malicious domains. From our large pool of EV registered domains, we first turned our attention to measuring those which have been observed serving malware or being used in command and control communications for cyberattacks. We performed a cross-correlation with Georgia Tech’s large-scale malware feed, and we found malware traffic communicating with 271 domains with EV certificates. We have included the list of domains in Appendix A. In addition, we measured the number of EV certificates among domains that were found in the Squid-blacklist and SANS Suspicious Domains blacklists. We found a total of 108 domains on these blacklists which hold EV certificates. We have included that list of domains in Appendix B. As mentioned previously, we were encouraged to see that these findings represent only a small fraction of the total of over 2.6M domains with EV certificates. In Section 4.4, we further provide a statistical analysis of the significance of these domains among the full EV certificate ecosystem.

4.3 Association of EV Certificates With Cybercrime

Next, we focused on understanding if domains with SSL certificates are still found to be abused by cyberactors, without being blacklisted, or if they are associated with cybercrime by any means. We performed cross-correlation of our EV SSL certificate database and any indices extracted from our data set of underground marketplace and forum communication. For the domains that we found to be associated with cybercriminals and holding SSL certificates, we investigated deeper to narrow down the involved cyberactors and the cybercrime activities they are associated with. We only found 3 domains with EV certificates that are
associated with cyberactors who have been actively tracked on underground marketplaces and forums. We found the following domains:

- darktrade.biz (Dark Trade Limited). This domain is associated with Dark Trade Limited. The associated cyberactor handle is: Rajit Bansal. This actor has a long-standing history and strong reputation score at the forum: Hack Forums. The actor has advertised darktrade.biz on cyberforums. This website is speculated to be a get-rich-quick Ponzi scheme for people looking to make money in cryptocurrency. Looking into organization information, the name associated with the company is Daniel Sinclair, which does not match the cyberactor claiming to own this domain. Interestingly, the EV certificate for darktrade.biz is currently expired, and in our future investigation we plan to measure the occurrence of domains presenting expired EV certificates.

- onionbit.com (nCrypt & Privacy). This domain is associated with nCrypt & Privacy Services S.L.. The associated cyberactor is: Loren Minel Andronie (p0s3id0n). Business records for onionbit.com list the founder as Loren Minel Andronie, which matches the cyberactor observed on underground forums. The registrant of the WHOIS record for onionbit.com is also the same cyberactor. This domain provides encrypted web mail services. The domain itself seems to be a legit website. This can be a case when cyberactors have a verifiable organization/business and collaborate with others who perform malicious activity behind it. Adding further suspicion to this case: the EV certificate presented by onionbit.com is actually for ico1.xcrypt.club. The vetting process to obtain an EV certificate should have caught this discrepancy, so we expect the certificate switch occurred after it was obtained legitimately. This is something we marked for future investigation, i.e., if the original EV vetting checks hold over time for each domain that our data set finds presenting an EV certificate.

- bmocareers.com (Bank of Montreal). The domain, bmocareers.com, is associated with the Bank of Montreal. The associated cyber-actor is: Lihwak. This actor has been known to be involved in banking fraud. On cybercrime forums, Lihwak has made posts and comments directed towards requests for credit reports and other personally identifiable information of US nationals. For this particular domain, we do not suspect that Bank of Montreal is a malicious domain. However, the reports imply that accounts associated with Bank of Montreal have been compromised. This example shows that even though Bank of Montreal uses an EV certificate, cyberactors still target Bank of Montreal and their domain may be targeted by cyberattacks in the future.

4.4 Statistical Significance of our Findings

In this section, we evaluate the statistical significance of suspicious domains among the full EV certificate ecosystem, by calculating the p-value using the z-test statistic.

Our null hypothesis is \( H_0(\text{NULL}) : p_0 \geq 0.00013 \). This represents the probability that an EV SSL certificate is associated with bad domains mentioned in underground forums or found among blacklists/malware traffic. Therefore, our alternative hypothesis is \( H_a(\text{Alternative}) : p_0 < 0.00013 \).

We assume \( \alpha = 0.01 \). We also know that our sample size is \( n = 2,604,344 \). Therefore, the proportion estimate is \( \hat{p} = \frac{3+108+271}{2,604,344} \). We calculate \( Z = \frac{p_0 - \hat{p}}{\sqrt{\frac{p_0(1-p_0)}{n}}} = -2.3497 \).
Then, we calculate the p-value as $P(Z < -2.3497) = 0.0094$ by looking up the Z table. We observe that the p-value (0.0094) is less than alpha. Thus, we reject the null hypothesis in favor of the alternative hypothesis. So, we conclude that $p_0$ is less than 0.00013 with significance level $\alpha = 0.01$.

So, the probability that an EV SSL certificate is associated with bad domains is less than 0.00013 or less than 0.013 %. Which means that EV SSL certificates are highly unlikely to be linked to domains that are associated with underground forums and marketplaces or malware/cybercrime activities.

### 4.5 Abused Domains Versus Intentionally Harboring Cybercrime

From our analysis, we were able to find two types of cybercrime associated domains with EV certificates: a) Domains that are already blacklisted and likely run by suspicious cyberactors. For example, these domains maybe registered and set up by suspicious cyberactors and even associated with businesses that are run by the cyberactors. b) Domains that are linked with legitimate businesses but they are heavily abused by cybercriminals, and therefore they show up as such in underground forum discussions.

Moving forward, and by leveraging our initial observations, we plan to build models that can learn the difference between the two categories based on the usage statistics of the domain and therefore reliably differentiate between abused domains and those intentionally harboring cybercrime. For example, domains that are run by cybercriminals often leverage some of the existing DNS hosting infrastructure they have set up for other suspicious domains they also run. In other words, they tend to not set everything up from scratch, but favor reusing their existing infrastructure (in intelligent ways that intend to frustrate detection). These behaviors can provide indicators of domains which are intentionally harboring cybercrime.

### 5 Future Work

Our future work focuses on designing new security indicators for the browser that better communicate a website’s trustworthiness [3]. Given that our data overwhelmingly shows that domains which invest in EV certificates tend to be trustworthy, we hope to better communicate the value of a domain’s EV certificate with website visitors. Our main goal is to design security indicators with two main properties: a) They are understood by non-experts by clearly communicating to the users whether a website can be trusted, and b) They can draw the users’ attention so that domains can make the most of the EV certificate’s value. Security indicators are commonly seen in major browser displays as locks, shields, or other symbols. Researchers have shown that, unfortunately, browser users do not always understand or notice them, and our future work with focus on these issues.

### 6 Conclusion

In this research, we conducted an initial investigation into understanding the role of SSL certificates in Internet abuse. We collected a unique combination of data sets that included: communication on underground marketplaces and forums, blacklisted domains, domains found in malware traffic, SSL certificates, and organization information. We constructed a scalable data store which enabled cross-data-set analysis to investigate if cybercriminals are
using/abusing EV certificates, if a legitimate legal entity benefits by investing in this indicator of trust for their website, and finally to understand the likelihood that an EV certificate registered business is committing cybercrimes. Our findings show that, the vast majority of EV certificates are indicative of legitimate domains. We investigated a total of approximately 2.6 million domains with EV certificates. Out of this sample, we only found a negligible number of domains to be associated with Internet abuse. We conclude that EV certificates are highly indicative of legitimate domains registered by legitimate business. Therefore, users benefit by noticing and using the browser security indicators as a guide to trust domains with EV SSL certificates.

References


A Domains With EV Certificate Found In Malware Traffic

aliakmon.cperi.certh.gr
updates1.safer-networking.org
y-center.ru
familysealrings.com
jbwere.com.au
hewitt.ca
sezz.be
wsrl.nl
showcast.com.au
brycen.co.jp
www.traction-software.co.uk
golfsmith.com
updates4.safer-networking.org
idcband.co.uk
mail.irf.se
hoyes.com
sauder.com
www.prolinkedcare.com
corryfcu.org
fnfg.com
xpressconnect.blackpool.ac.uk
pix.bit.ly
web04.magicjack.com
www.krollontrack.com
4everyware.nl
villanisalumi.it
parex.lv
www.sunnysoft.cz
www.dnsdun.com
mail.uclan.ac.uk
bailliegifford.com
gnosticteachings.org
images.smartname.com
connect.easymarkit.com
oldmutual.com
wpi-wireless-setup.wpi.edu
hotmail.co.jp
tfs.ca
visuals.co.uk
www.responsivedata.com
mma.edu
victoryrecords.com
acerbis.it
www.orka-iceberg.com
mpif.org
birch.com
www.jixunjsq.com
southedge.net
www.vix.at
img.tenpay.com
posten.se
laholm.se
wareconsult.com
alumni.jwu.edu
mail.ruc.dk
mail5.vn.fi
block.io
i-nexus.com
mail.jwu.edu
psbank.com.ph
www.cpfl.com.br
starcostumes.com
secure.tibia.com
www.regalassets.com
mx2.handelsbanken.se
ybbsmtp.mail.yahoo.co.jp
www.saredrogarias.com.br
math.northwestern.edu
unitedsavingscu.org
cat.eduroam.org
adaptoutdoors.com
waterchina.com
www.ufsexplorer.com
orbisinc.com
smtp1.horizon-bcbsnj.com
login.launchpad.net
isom.org
ftb.com
bnpparibasfortis.com
loverslane.com
cloudpath.miracosta.edu
npfe.ru
bluewin.com
securemessage.aessuccess.org
www.alteraeon.com
obvion.nl
www.go.vn
extended-validation-ssl.websecurity.symantec.com
www.naja7host.com
mail.brandeis.edu
azimut.it
jzk.pl
kecgate05.infosys.com
my.smart.com.ph
mail.emmanuel.edu
parliament.fi
smtp.brandeis.edu
mhbbottrop.de
actide.com
web03.magicjack.com
bendigobank.com.au
tilgroup.com
decon.unipd.it
www.elpro.si
www.bridge-of-love.com
www.acrosoftware.com
hmailgw1.hersheys.com
test123.com
firsthorizon.com
secure.accasoftware.com
care.citrixonline.com
ghgroup.com
bankofmissouri.com
macquarie.com.au
pfister.ch
crystalink.co.uk
bellalei.com
gecapital.com
bitext.com
helpdesk.grantmsp.com
csbp.com.au
lcdls.symantec.com
www.jdyou.com
remotely.com.au
enklare.se
continfo.com
laposte.com
trimbos.nl
badgermeter.com
dbb.su.se
walletmix.com
correctproducts.com
q8.dk
esmoke.net
corphthis.com
nerdlink.support
secure.esupport.com
spirit.nl
sensical.net
certh.gr
constangy.com
amcor.com
toppromotions.com
onebeacon.com
11.osdimg.com
mfn.unipmn.it
remote.celeratec.com
barnsley.ac.uk
th-witt.com
codesector.com
dmell-seg-03.o2.com
dg.com
www.propersoft.net
ubercpm.com
moneynetint.com
sec.smtp.chebucto.ns.ca
app.tanwan.com
ccaq.com
store.wotrus.com
cashrun.com
tools.cadren.com
mmile.com
www.myropcb.com
sanpaloimi.com
store.toonboom.com
hol.co.uk
allianz.gr
www.marymaxim.com
esppos.com
mail.utu.fi
support.reliablecomputersinc.com
www.getcashhelp.com
deltalloyd.nl
creditplus.de
my.ispsystem.com
hexui.com
www.amyuni.com
sparda-west.de
hbfuels.com
www.fnw.us
mail1.qmul.ac.uk
my.tesco.com
code.poptm.com
money.v2profit.com
vp.pl
bookcloseouts.com
www.slimwareutilities.com
www.a-q-f.com
vivastay.com
gate-pri.osfi-bsif.gc.ca
host.do
sw.ca
warcococonstruction.com
onsite-ocsp.verisign.com
mail.ucbscz.edu.bo
arla.se
quick.ru
sydneyaquarium.com.au
privacyharbor.com
ionion.ath.hcmr.gr
www.michaelsutter.com
www.loveandseek.com
reciva.se
www.revouninstallerpro.com
marketing.citrixonline.com
static.yunaq.com
troy.k12.mo.us
kleemann.gr
mail.dmea.com
www.pchealthboost.com
mazda.de
www.jzk.pl
interflora.no
www.venturesonsite.com
forge.gridforum.org
rvcschools.org
www.lostpassword.com
coop.nl
kcda.org
alteraeon.com
dmmlw-seg-01.o2.com
magiclife.com
asbbank.co.nz
hdvest.com
B Domains With EV Certificates Found In Blacklists

decisionlogic.com
biologic.biz
drummersdream.com.au
alt-energy.biz
christopherco.com
archku.ac.bd
bitages.com
bitwave.biz
brandywinematerials.com
exofinancialgroup.com
jovkar.com
auzonet.net
flyfishusa.com
drap-house.fr
blazingboost.com
amateurgolftour.net
dickensonworld.com
extraessay.com
biomeq.com.vn
infomonsta.net
beautycommunity.co.th
ingersollrandmexico.com
endowise.com
confydo.com
amg.biz
ckmack.com
ingetrol.cl
bravocapital.biz
adcube.com.tr
barclaysclub.com
aptian.net
enjoy-your.life
balharbourshops.com
drivehq.com
bioinfomedical.com
carfax.com
ectotrust.com
goodly.pro
carecompare.com
antojese.com
ascentive.com
blindtrack.co.uk
easymining.biz
fbm.com.tr
electronicscity.com
flex.ru
arian.fund
ftec.ai
ibx.key.com
alilaguna.it
download.getjar.com
cryptogolden.com
broker-insight.com
applemountain.net
aplusglass-parebrise-anet.fr
cryptobillion.com
bithonest.biz
banner.casino.williamhill.es
bitcoinclub88.com
insights.abnamro.nl
cloudme.com
click4support.net
britbit.biz
cartalibra.it
ekokond.ru
bitvillage.biz
hadeplatform.com
copetran.com.co
johnsmustang.com
juntadebeneficencia.org.ec
deverellsmith.com
click2sell.eu
bgloanandjewelry.com
ebayshares.com
kykeon-eleusis.com
cheapwritingservice.com
ambis.biz
araiautohelmets.com
ipstresser.com
challengestrata.com.au
essaydoc.com
bizzillion.com
bitxxa.com
cryptoclone.com
dled.ru
houzz.es
alphainfosystem.com
athensheartcenter.com
ddfutures.com
bitways.biz
beds2buy.co.uk
contact-media.co.uk