How is the Forged Certificates in the Wild: Practice on Large-scale SSL Usage Measurement and Analysis

Mingxin Cui[1,2], Zigang Cao[1,2], and Gang Xiong[1,2]✉

[1] Institute of Information Engineering, Chinese Academy of Sciences
[2] School of Cyber Security, University of Chinese Academy of Sciences
{cuimingxin,caozigang,xionggang}@iie.ac.cn

Abstract. Forged certificate is a prominent issue in the real world deployment of SSL/TLS - the most widely used encryption protocols for Internet security, which is typically used in man-in-the-middle (MITM) attacks, proxies, anonymous or malicious services, personal or temporary services, etc. It wrecks the SSL encryption, leading to privacy leakage and severe security risks. In this paper, we study forged certificates in the wild based on a long term large scale passive measurement. With the combination of certificate transparency (CT) logs and our measurement results, nearly 3 million forged certificates against the Alexa Top 10K sites are identified and studied. Our analysis reveals the causes and preference of forged certificates, as well as several significant differences from the benign ones. Finally, we discover several IP addresses used for MITM attacks by forged certificate tracing and deep behavior analysis. We believe our study can definitely contribute to research on SSL/TLS security as well as real world protocol usage.

Keywords: Forged Certificate · Passive Measurement · SSL MITM.

1 Introduction

Secure Sockets Layer (SSL) and its successor Transport Layer Security (TLS) are security protocols that provide security and data integrity for network communications (we refer to SSL/TLS as SSL for brevity in this paper). An X.509 certificate plays an important role in SSL Public Key Infrastructure, which is the basis of the SSL encryption framework. When establishing SSL connection, the server or/and the client is required to provide a certificate to the peer to prove its identity. Since the widespread use of SSL, issues of certificate come out one after another as well, such as compromised CAs, weak public key algorithm, forged certificates, and so on. In this paper, we focus on forged certificates that mainly used in MITM attacks on HTTPS web services.

When carrying out an SSL man-in-the-middle (MITM) attack, the attacker usually uses a forged certificate to pretend the compromised or malicious server and deceive careless users. And a victim's negligence would then lead to the privacy disclosure and property loss. Since attempts to MITM attacks

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_62

on https-encrypted web sites have never stopped, it's necessary to conduct a comprehensive analysis to study the status quo of forged certificates in the real world.

Many researchers have published their work on the certificate ecosystem, providing different aspects view of X.509 certificates. And studies that try to reveal the negative side of the certificate have never been stopped. However, there're few works focused on forged certificate used by MITM attacks. In this paper, we conduct a comprehensive study of forged certificates in the wild, and the contributions of our work are as follows: First, we implement a 20-month passive measurement to collect the real-world SSL certificates on two research networks to explore the forged certificate issue, which is up to now the largest scale long term study. Second, we analyze the forged certificates against Alexa top 10 thousand web sites by combining both the passive measurement results and the public certificate transparency (CT) logs [14, 18], which is highly representative and comprehensively. Third, we reveal the reasons, preferences of forged certificates, as well as distinct differences between the forged certificates and the benign ones in several attributes, offering valuable insights to researches on SSL/TLS security. Finally, several IP addresses are discovered which are probably used to carry out MITM attacks through a series of forged certificates tracing and traffic behavior analysis.

The remainder of this paper is structured as follows. Section 2 elaborates the related works. Section 3 describes our measurement and dataset used in this paper. We analyze and elaborate how forged certificates performed in the wild in Section 4 and compare to benign ones In Section 5. In Section 6, we try to trace and identify SSL MITM attacks and we conclude our work in Section 7.

2 Related Works

In recent years, many researchers have focused on the measurement of SSL encrypted traffic, and there're two ways to implement this: active measurement and passive measurement. [8] performed 110 scans of the IPv4 address space on port 443 over 14 months to study the HTTPS certificate ecosystem. [7] also implemented scans of the public IPv4 address space to collect data, with the help of ZMap [9]. And the certificate they collected could be found in Censys [1]. What's more, [19] found that a combination of Censys data and CT logs cloud account for more than 99% of their observed certificates. There're many other active scans which provide datasets of certificates, such as Rapid7 SSL [13]. [10] implemented both active and passive measurement to present a comprehensive analysis of X.509 certificates in the wild. The authors conducted HTTPS scans of popular HTTPS servers listed in Alexa Top 1 Million over 1.5 years from 9 locations distributed over the world. They also monitored SSL traffic on a 10 Gbps uplink of a research network.

There're also many works focused on the negative side of the certificate. [16] identified web-fraud using attributes extracted from certificates. [5] implemented an empirical study of certificates for depository institutions and showed the bad

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_62

condition of bank websites in disposing SSL. [6] proposed a machine-learning approach to detect phishing websites utilizing features from their certificates. [2] studied invalid certificates in the wild, and revealed that most of the invalid certificates could be attributed to the reissues of a few types of end-user devices. Comparison between valid and invalid certificates was conducted as well. Several studies [4, 11] detected SSL MITM attacks using different methods. However, they didn't focus on forged certificates used by MITM attackers. [12] implemented a method to detect the occurrence of SSL MITM attacks on Facebook, and their results indicated that 0.2% of the SSL connection they analyzed were tampered with forged SSL certificates. Their work only concentrated on Facebook, could not provide an overall view of forged certificates.

3 Measurement And Datasets

In this section, we describe our passive measurement and the datasets. The methodology of identifying forged certificates is elaborated as well.

3.1 Passive Measurement

In order to study the forged certificates, we implemented a passive measurement on two large research networks from November 2015 to June 2017. These networks could provide 100 Gbps bandwidth. Our program collected certificates and SSL sessions statistical information after an anonymous processing. Useful data would be added into the corresponding datasets, namely *DsCrt* and *DsCnn* respectively.

Table 1. Overview of Certificate Dataset

CERT TYPES		#(CERT)	#(ISSUERS)	#(SUBJECTS)
forged certs	selfsigned	107,306	922	922
	un-selfsigned	2,759,980	215,236	3,988
	totally	2,867,286	216,154	4,165
benign certs		1,910,385	180	11,707
totally		4,777,671	216,243	12,012

3.2 Datasets

Based on the measurement, we made up two datasets to store the certificates information and SSL sessions statistical information separately.

DsCrt contained all certificates we collected during the 20-month long measurement. Excluding tiny errors due to the high-speed network environment, we totally obtained 188,064,507 unique certificates, including 3,359,040 CAs (both root and intermediate) and 184,705,467 leaves. We extracted and identified these leaf certificates using the methodology mentioned in Section 2.3 and harvested 2,867,286 forged ones in 4,777,671 certificates that claimed to belong to Alexa Top 10k domains. After the identification, all of the gathered certificates were parsed into json format completely, referring to *Censys* data format [7]. The parsed attributes of a certificate include but are not limited to *sha1 value*, *signature algorithm*, *public key information*, *issuer*, *subject*, *validation period*, *extensions*, and so on.

DsCnn recorded the statistical information of SSL sessions detected during our measurement. For each SSL session, we stored server IP, server port, and some basic statistics such as bytes, packets, and packet interval, and of course the corresponding certificate SHA1 string. Server IP and port might help to trace the suspicious MITM attacks, and the basic statistics could be used to train machine-learning models in the future work.

3.3 Identifying Methodology

When identifying a leaf certificate was benign or not, we utilized the CAs in the Chrome root store (as of July 1, 2017). Since the measurement lasted such a long time, we ignored validation errors only due to expiration time. Thus, we recognized a leaf certificate was benign if the root CA of the corresponding certificate chain was credible. Otherwise it's not. For the latter one, we then checked if it was self-signed, and then labeled the certificate using "*is_benign*" and "*is_selfsigned*" attributes.

Since many web service providers use self-signed certificates due to the balance of cost and safety, and the compromise or abuse of root and intermediate CAs, it's really hard to identify whether a certificate was forged or not, especially for a self-signed one or a website in obscurity. Hence we chose the domains listed in Alexa Top 10k as target, and studied forged certificates of these well-known web services (if provided SSL encryption) picking the public CT logs [18] as a benchmark. For a certain certificate, if it wasn't included in any public CT logs, we regarded it as a forged one. Based on this constraint, we extracted 4,777,671 certificates which claimed to belong to Alexa Top 10k domains, and verified them with the help of CT logs included in Chrome. Finally we harvested 2,867,286 forged certificates of 4,165 different web services or domains after the verification.

3.4 Ethical Considerations

Considering the privacy and ethical issues in the passive measurement, we implement an anonymous process while dealing with the data. The client IP of each connection has been anonymized before our collection in the measurement system. Thus, we do not know the real client IP address of each SSL session. We focus on certificates and corresponding servers, but not the user privacy.

Table 2. Top 20 Issuers of Forged Certificates

No.	ISSUER CN	#(CERT)	RATIO	No.	ISSUER CN	#(CERT)	RATIO
1	([0-9a-z]{16})	998959	34.84%	11	UBT (EU) Ltd.	15715	0.55%
2	mitmproxy	993585	34.65%	12	SSL-SG1-HK1	15350	0.54%
3	FortiGate CA	114276	3.99%	13	Lightspeed Rocket	14827	0.52%
4	(selfsigned)	107306	3.74%	14	thawte 2	14512	0.51%
5	Cisco Umbrella Secondary SubCA *-SG	52196	1.82%	15	10.1.100.51	13917	0.49%
6	Phumiiawe	34742	1.21%	16	Pifbunbaw	12630	0.44%
7	www.netspark.com	28128	0.98%	17	192.168.1.1	11597	0.40%
8	samsungsemi-prx.com	20174	0.70%	18	Essentra	11512	0.40%
9	DO_NOT_TRUST _FiddlerRoot	19921	0.69%	19	Bureau Veritas	10776	0.38%
10	(null)	18712	0.65%	20	michael.aranetworks .com	10497	0.37%

4 Forged Certificates Status in Quo

Based on the measurement and identifying methodology, we obtained 2,867,286 forged certificates of 4,165 different web services/domains. These forged certificates contained 107,306 self-signed ones of 922 different subjects. Others belonged to 215,236 different issuers of 3,988 unique subjects. Details could be seen in Table 1.

In this section, we studied the forged certificates comprehensively, including issuers, subjects, public key, validity period, lifetime, and so on. We also compared these features of forged certificates to the benign ones', tried to reveal the significant difference between them.

4.1 Issuers of Forged Certificates

We firstly analyzed the issuers of these forged certificates and tried to find out (or determine) the main causes. Table 2 lists the top 20 issuers CommonName (CN) [3] of forged certificates. No.1 issuer CN indicated 189,912 issuers whose CNs satisfied the regular expression of [0-9a-z]{16}. No.4 issuer CN indicated all self-signed forged certificates. No.10 indicated the corresponding certificates didn't have the attribute of issuer CN.

These issuers could be divided into several classes. Some issuers are related to security products or anti-virus software, such as *FortiGate CA* (firewall), *Cisco Umbrella Secondary SubCA *-SG* (secure internet gateway), and *www.netspark.com* (content filter). Meanwhile, some others might be used for malicious services like MITM attacks, such as "[0-9a-z]{16}". According to the attributes of issuers, the certificates they've issued, and the corresponding servers, we roughly classified these issuers into 4 classes: *SecureService*, *Research*, *Proxy*, and *Suspicious*, which were presented in Table 3. Issuers classified as

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_62

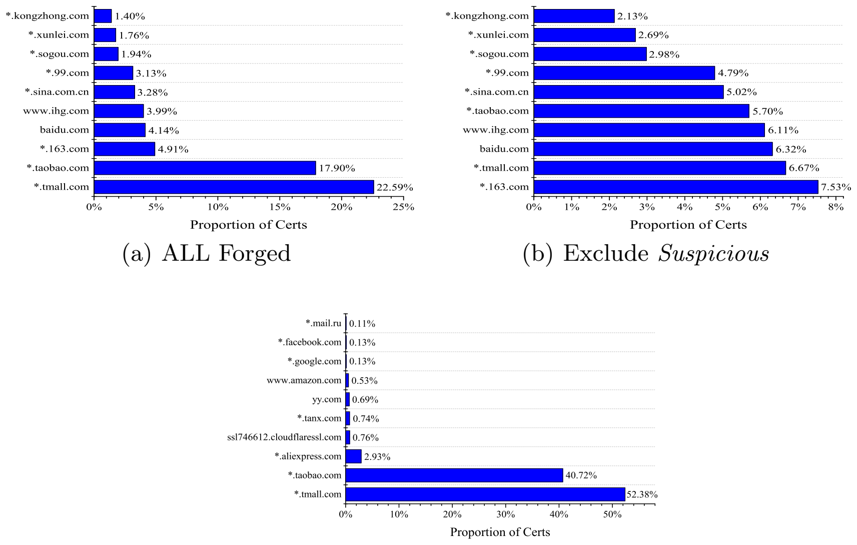
Table 3. Rough Classification of Issuers

	SecureService	Research	Proxy	Suspicious
examples	FortiGate CA	mitmproxy	PERSONAL Proxy CA	([0-9a-z]{16})
	Cisco Umbrella Secondary SubCA *-SG	DO_NOT_TRUST_FiddlerRoot	EBS_FG_CA_SSLProxy	
	Lightspeed Rocket	colegiomirabal.edu	ProxySchool12	thawte 2

SecureService were those deployed in security products for security protection or content audit. A *Research* issuer mainly claimed to belong to a research institute or a university, or indicated to a famous tool used to analyze HTTPS traffic. Actually, these tools might be used for malicious services, but we roughly classified the corresponding issuers to *Research* considering the mainly usage and the diversity of certificates faked by them. We simply considered an issuer claimed to be a proxy (mainly contained the word "proxy" in the Common Name) as *Proxy*. This category had the lowest priority due to its simplest classification basis. *Suspicious* issuers referred to those we suspected faking certificates for MITM attacks. The reason we named it *Suspicious* but not *Malicious* was that we could not directly prove that the corresponding services were malicious, and it's really hard to confirm. We determined an issuer to be a suspicious one based on three aspects: 1) the certificates faked by this issuer limited to several species, mainly for financial types; 2) the account of corresponding SSL sessions was much less than the average; 3) the account of corresponding servers for an issuer (or a series of similar issuers) was much less than the average. Examples of these classes could be seen in Table 3 as well.

4.2 Preference of Forged Certificates

In addition from the issuer's perspective, we also analyzed the forged certificate in the view of the subject, trying to figure out the preference of forged certificates. We analyzed subjects in four classes mentioned above, and found the preference of *Suspicious* issuers performed differently from the others. As showed in Fig.1 (c), more than 96% forged certificates issued by *Suspicious* issuers are related to three shopping websites which belonged to one e-commerce company Alibaba. **.tmall.com* and **.taobao.com* are mainly for domestic users, and **.aliexpress.com* mainly provides service for global consumers. Without considering *Suspicious* issuers, the top 10 forged certificates each represents a well-known web service in different fields, and they perform relatively average in the count, as showed in Fig.1 (b). The reason for this result is obvious: criminals are more interested in the wallets of victims. And the other classes of issuers do not have a clear tendency when forging certificates. Their preferences only related to the popularity of each web service.



(a) ALL Forged

(b) Exclude *Suspicious*

(c) *Suspicious*

Fig. 1. Preference of Forged Certs. Abscissa axis lists the Top 10 subjects of forged certificates in different classes, vertical axis indicates the percentage of all forged certificates owned by each subject.

5 Comparison with Benign Certificates

We studied several attributes of forged certificates, including security attributes (such as certificate version, signature algorithm, and public key information), validity period, and lifetime. Compared to benign certificates, the forged ones performed extremely different in many aspects.

5.1 Security Attributes

We selected version, signature algorithm, public key algorithm and public key length to characterize the security of a certificate. Fig.2 shows the comparison of these attributes between forged and benign certificates.

Version: There're three versions of the X.509 certificate: *v1*, *v2*, and *v3*. *v1* certificates were deprecated long time ago, considering the security. And *v2* certificates were even not widely deployed on the Internet. *v3* certificate is currently the most widely used in the wild, and our measurement result confirmed this. Forged and benign certificates performed similarly in the statistical characteristics of the *version* attribute, both of which had more than 99% of *v3* certificates. The tiny difference is that compared to benign ones, more forged certificates were *v1* (nearly 1%). What's more, we find 102 benign certificates and 18 forged

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_62

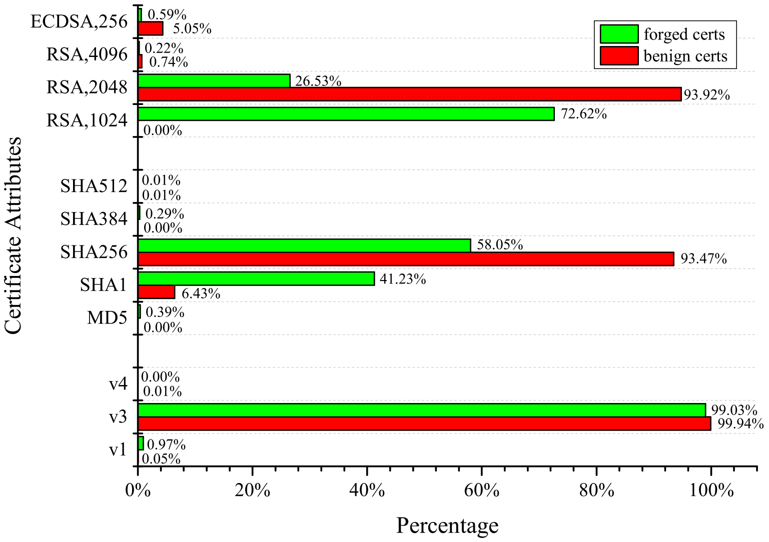
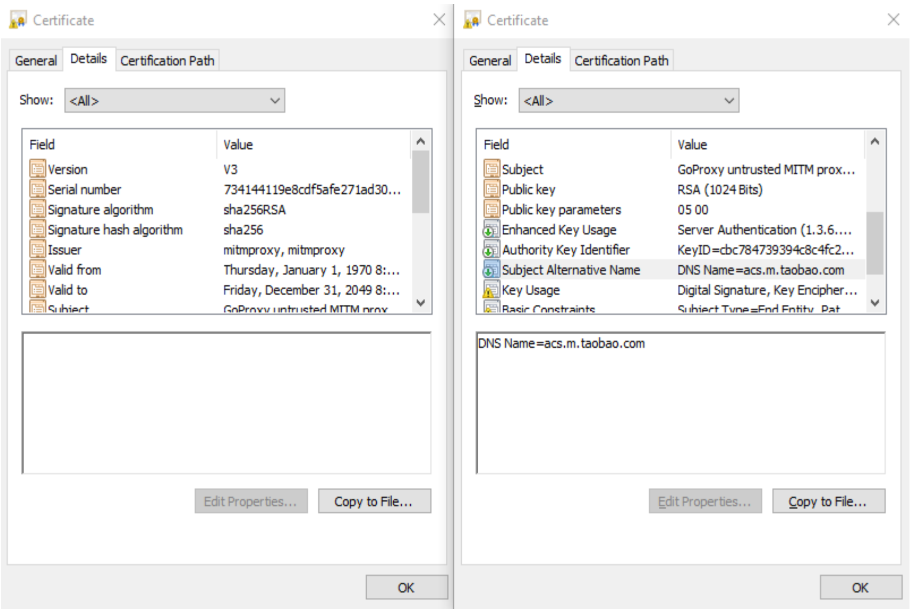


Fig. 2. Comparison of Security Attributes Between Forged and Benign Certs

ones using *v4*. Since the subscript of *version* started from 0, which meant that *"version=0"* indicated the *v1* certificates, we speculated that most of the *v4* certificates might be ascribed to misoperation.

Signature Algorithm: It's well known that MD5 and SHA1 algorithms were both cracked many years ago. Especially the SHA1 algorithm, which was once widely used, is still widely used in the wild, due to our measurement result. As shown in Fig.2, more than 90% benign certificates used SHA256, while only 6.43% used SHA1. Well for the forged certificates, only 58.05% used SHA256, much less than the benign ones. And surprisingly to us, 41.23% of the forged certificates was still using SHA1 algorithm. It was obvious that forged certificates had a much lower security in the attribute of the signature algorithm. We then analyzed forged certificates which used SHA1 algorithm, and found out more than 80% SHA1 certificates could be attributed to *mitmproxy*. *mitmproxy* [17] is a free and open source HTTPS proxy. It's widely used for HTTPS proxies, security experiments, and of course MITM attacks. We found nearly 950,000 forged certificates signed by *mitmproxy* with the signature algorithm of SHA1 and 1024-bit RSA public key, while others used the signature algorithm of SHA256. Fig.3 is an example of certificate signed by *mitmproxy*. We found that though most of this kind of certificates used SHA1 certificates, they have changed to use SHA256 recently. We speculate that this may be attributed to version updates or source code rewrites.



Field	Value
Version	V3
Serial number	734144119e8cdf5afe271ad30...
Signature algorithm	sha256RSA
Signature hash algorithm	sha256
Issuer	mitmproxy, mitmproxy
Valid from	Thursday, January 1, 1970 8:...
Valid to	Friday, December 31, 2049 8:...
Subject	GoProxy untrusted MITM prox

Field	Value
Subject	GoProxy untrusted MITM prox...
Public key	RSA (1024 Bits)
Public key parameters	05 00
Enhanced Key Usage	Server Authentication (1.3.6....
Authority Key Identifier	KeyID=cbc784739394c8c4fc2...
Subject Alternative Name	DNS Name=acs.m.taobao.com
Key Usage	Digital Signature, Key Encipher...
Basic Constraints	Subject Type=End Entity, Pat

DNS Name=acs.m.taobao.com

Fig. 3. An Example of *mitmproxy* Signed Certificate

Public Key Information: The situation of public key information is similar to the signature algorithm to some extent, as shown in Fig.2. Though 1024-bit RSA key was not secure any more several years ago, 72.62% of forged certificates still use it to encrypt their SSL connection. Similarly, 45% of these insecure forged certificates can be attributed to *mitmproxy*, due to the reason mentioned above. And 48% of them were issued by *([0-9a-z]{16})*. What's more, we found that all forged certificates issued by *([0-9a-z]{16})* had 1024-bit RSA public keys and were signed by SHA1 signature algorithm. For benign certificates, 2048-bit RSA public key accounts for the mainstream with a percentage of 93.92%. And different from the status of signature algorithm, 5.05% proportion of *ECDSA,256* in benign certificates indicates the trend of more secure public key algorithm. Actually, many famous companies, such as Google, Facebook, and Alibaba, have changed their public key algorithm to the more secure Elliptic curve cryptography (ECC) algorithm.

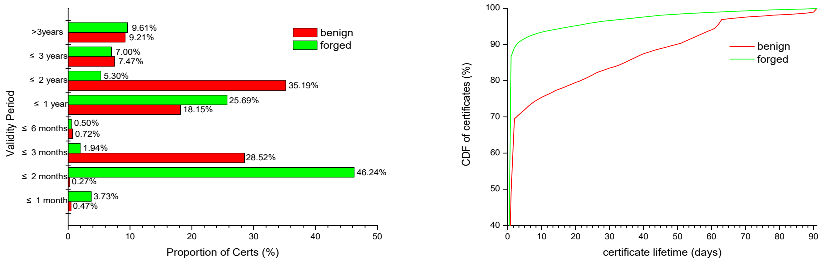
According to the results above, we can conclude that the security attributes of forged certificates performed much worse than the benign ones. Since forged certificates rarely considering security issues, the conclusion is in line with our expectation. What's more, we found that the attributes of forged certificates

DOI: 10.1007/978-3-319-93713-7_62

are closely related to the top issuers listed in Table 2. And this conclusion also applied to the attribute of the certificate validity period.

5.2 Certificate Validity Period:

We calculated the validity period (the time between the certificates' *Not Before* and *Not After* attributes) of each certificate, and found a significant difference between forged certificates and the benign ones (shown in Fig.4(a)). For benign certificates, most of their validity period were located in three intervals: 2˜3months, 6months˜1year, and 1˜2years. While for forged ones, most of their validity period were located in two intervals: 1˜2months and 6months˜1year. In detail, 46.07% forged certificates were valid for 32 days, and 11.86% were valid for 1 year. We then studied the three significant differences shown in Fig.4 (a), revealed that different validity periods of certificates might be caused by different issuers. We found that almost all forged certificates (more than 99.99%) which owned a 32-day validity period were issued by *mitmproxy*. What's more, 99.99% of benign certificates with the validity period of 84day were issued by *Google Internet Authority G2*, and more than 99.96% with the validity period of 90-day were issued by *Let's Encrypt Authority X3*. Certificates validated for 2 years (730 days) cloud be attributed to issuers belonged to *DigiCert* (46.44%), *Symantec* (12.98%), *VeriSign* (7.27%), and so on.



(a) Validity Period of Certs

(b) Lifetime of Certs

Fig. 4. Validity Period and Lifetime of Certs

5.3 Certificate Lifetime:

In order to compare the lifetime (the days between the first time and the last time a certificate exposed in our sight) of forged and benign certificates, we selected a three-month period, from April 1 to June 30, to implement continuous observation. During the 91-day observation, we obtained that the average lifetime of forged certificates was 3.59 days, while for authorized ones it was 12.02 days.

DOI: 10.1007/978-3-319-93713-7_62

The result demonstrated that forged certificates had a much shorter lifetime than the benign ones, as shown in Fig.4 (b). We speculated that the MITM attackers need to update their forged certificates frequently to evade the detection of security products like IDS and firewall, or to replace the blacklisted ones. Fig.4 (b) also shows that more than 85% forged certificates only appeared once in our 91-day observation. This could be attributed to security tests or researches which only performed once.

5.4 Conclusion:

According to the comparisons mentioned above, we cloud conclude that most forged certificates didn't care about security as the widely use of unsafe signature algorithm and public key algorithm. However, what they most concerned was evading the detection of anti-virus software. Hence the lifetime of forged certificates was much shorter than benign ones, as attackers required to update certificates frequently.

6 Tracking MITM Attacks

While the use of forged certificates can be diverse, MITM attacks directly threatened users' privacy and security. Thus tracking MITM attacks is necessary, and many researchers have focused on this issue. In this paper, we discovered a MIT-M attack and performed a tracking with the help of forged certificate attributes and SSL session statistics.

Table 4. Suspicious Servers

SERVER IP	#(SESSIONS)	#(PORTS)	#(CERTS)	LIFETIME
195.154.161.209	869949 (54.46%)	500 (2876-3375)	49743 (65.88%)	04/01/2017-06/30/2017
62.210.69.21	680721 (42.61%)	500 (3336-3835)	46844 (62.04%)	04/01/2017-06/30/2017
195.154.161.44	38208 (2.39%)	500 (2601-3100)	13129 (17.39%)	04/01/2017-06/30/2017
195.154.161.172	8366 (0.52%)	500 (4606-5105)	4059 (5.38%)	04/08/2017-05/29/2017

When performing a deep analysis, forged certificates with similar issuers that satisfied the regular expression of *[0-9a-z]*{*16*} caused our attention. According to our analysis, more than 96% certificates faked by these issuers belonged to three famous e-commerce websites in China, 52.38% for **.tmall.com*, 40.72% for **.taobao.com*, and 2.93% for **.aliexpress.com*. And all three websites belonged to the same company, Alibaba, the most famous e-commerce company in China. The first two websites mainly served domestic users, while the last one provided global online shopping services. And that's why the forged certificates of this website were far less than the former ones. According to the reasons above, we speculated that these issuers were related to the MITM attacks targeting the

DOI: 10.1007/978-3-319-93713-7_62

Table 5. Information of Suspicious Servers

SERVER IP	LOCATION		ISP	ORGANIZATION	DOMAIN
195.154.161.209	France	48.8582, 2.3387	ONLINE S.A.S.	Iliad-Entreprises	poneytelecom.eu
62.210.69.21	France	48.8582, 2.3387	Free SAS	ONLINE SAS	poneytelecom.eu
195.154.161.44	France	48.8582, 2.3387	ONLINE S.A.S.	Iliad-Entreprises	poneytelecom.eu
195.154.161.172	France	48.8582, 2.3387	ONLINE S.A.S.	Iliad-Entreprises	poneytelecom.eu

online shopping service of Alibaba. Thus we analyzed the statistical SSL session information of corresponding certificates to verify our suspicion.

We selected 3 months connection data from 04/01/2017 to 06/30/2017, and extracted server information of the forged certificates which met the above conditions. Finally, we obtained 230 unique servers from 1,597,532 SSL sessions. Surprisingly, most sessions (99.98%) and certificates (99.77%) belonged to four server IPs, as showed in Table 4. We then looked up the information of these IPs with the help of MAXMIND [15], finding that they might locate in a same position and belonged to a same organization, as showed in Table 5. Thus, we suspected these IP addresses should be attributed to MITM attacks for the reasons below:

1. These IPs covered almost all forged certificates and SSL sessions.
2. Forged certificates used by these IPs are mainly related to 3 domains, which provided online shopping services. Obviously, MITM attackers cared more about the victim's wallet.
3. Each IP was served on 500 different and contiguous ports, and the number of sessions for each (*IP,port,cert*) triple per day is much less than proxy-used forged certificates.
4. These IPs belong to a same organization and locate in the same country.
5. When searching *poneytelecom.eu* on Google, many results indicated that this domain/organization was related to web fraud behaviors.

7 Conclusion

In this paper, we implemented a 20-month passive measurement to study the status quo of forged certificates in the wild. Based on CT logs provided by Chrome, we identified forged certificates of the web services listed in Alexa Top 10k, and finally gathered 2,867,286 forged ones. We analyzed the forged certificates in the view of both issuer and subject. Based on the analysis of issuers, we revealed the causes of forged certificates by roughly classifying them into four categories. *SecureService* certificates were mainly deployed in security products for security protection or content audit. *Research* indicated issuers that claimed to belong to a research institute or a university, or famous tools used to analyze HTTPS traffic. *Proxy* might be used for proxy servers, and *Suspicious* issuers referred to those we suspected faking certificates for MITM attacks. The study of the subject showed us the preference of forged certificates. While *Suspicious* mainly focused on financial related services, others preferences only related to the popularity of each web service. When comparing forged certificates to the benign ones, we found significant differences in security attributes, validity period, and lifetime. Benign certificates used more secure signature and public key algorithm to ensure the security of SSL encrypted connection, while forged ones performed quite terrible. Forged certificates also had a much shorter validity period and lifetime. What's more, we found the validity period of a certificate, no matter forged or benign, was related to its issuer. At last, we traced a series of *Suspicious* forged certificates and harvested four malicious server IP addresses through traffic behavior analysis.

Acknowledgments. This work is supported by The National Key Research and Development Program of China (No.2016QY05X1000 and No.2016YFB0801200) and The National Natural Science Foundation of China (No.61602472). Research is also supported by the CAS/SAFEA International Partnership Program for Creative Research Teams and IIE, CAS international cooperation project.

References

1. Censys: Censys. https://censys.io/, accessed July 21, 2017
2. Chung, T., Liu, Y., Choffnes, D.R., Levin, D., Maggs, B.M., Mislove, A., Wilson, C.: Measuring and applying invalid SSL certificates: The silent majority. In: Proceedings of the 2016 ACM on Internet Measurement Conference, IMC 2016, Santa Monica, CA, USA, November 14-16, 2016. pp. 527–541 (2016)
3. Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., Polk, W.T.: Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile. RFC **5280**, 1–151 (2008)
4. Dacosta, I., Ahamad, M., Traynor, P.: Trust no one else: Detecting MITM attacks against SSL/TLS without third-parties. In: Computer Security - ESORICS 2012 - 17th European Symposium on Research in Computer Security, Pisa, Italy, September 10-12, 2012. Proceedings. pp. 199–216 (2012)
5. Dong, Z., Kane, K., Camp, L.: Phishing in smooth waters: The state of banking certificates in the us (2014)
6. Dong, Z., Kapadia, A., Blythe, J., Camp, L.J.: Beyond the lock icon: real-time detection of phishing websites using public key certificates. In: 2015 APWG Symposium on Electronic Crime Research, eCrime 2015, Barcelona, Spain, May 26-29, 2015. pp. 1–12 (2015)
7. Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., Halderman, J.A.: A search engine backed by internet-wide scanning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12-6, 2015. pp. 542–553 (2015)
8. Durumeric, Z., Kasten, J., Bailey, M., Halderman, J.A.: Analysis of the HTTPS certificate ecosystem. In: Proceedings of the 2013 Internet Measurement Conference, IMC 2013, Barcelona, Spain, October 23-25, 2013. pp. 291–304 (2013)
9. Durumeric, Z., Wustrow, E., Halderman, J.A.: Zmap: Fast internet-wide scanning and its security applications. In: Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013. pp. 605–620 (2013)

10. Holz, R., Braun, L., Kammenhuber, N., Carle, G.: The SSL landscape: a thorough analysis of the x.509 PKI using active and passive measurements. In: Proceedings of the 11th ACM SIGCOMM Internet Measurement Conference, IMC '11, Berlin, Germany, November 2-, 2011. pp. 427–444 (2011)
11. Holz, R., Riedmaier, T., Kammenhuber, N., Carle, G.: X.509 forensics: Detecting and localising the SSL/TLS men-in-the-middle. In: Computer Security - ESORICS 2012 - 17th European Symposium on Research in Computer Security, Pisa, Italy, September 10-12, 2012. Proceedings. pp. 217–234 (2012)
12. Huang, L., Rice, A., Ellingsen, E., Jackson, C.: Analyzing forged SSL certificates in the wild. In: 2014 IEEE Symposium on Security and Privacy, SP 2014, Berkeley, CA, USA, May 18-21, 2014. pp. 83–97 (2014)
13. Labs, R.: Rapid7 labs - ssl certificates. https://opendata.rapid7.com/sonar.ssl/, accessed July 21, 2017
14. Laurie, B.: Certificate transparency. Commun. ACM **57**(10), 40–46 (2014)
15. MAXMIND: Geoip2 precision service demo. https://www.maxmind.com/en/geoip2-precision-demo/, accessed September 17, 2017
16. Mishari, M.A., Cristofaro, E.D., Defrawy, K.M.E., Tsudik, G.: Harvesting SSL certificate data to identify web-fraud. I. J. Network Security **14**(6), 324–338 (2012)
17. mitmproxy: mitmproxy. https://mitmproxy.org/, accessed October 10, 2017
18. Transparency, C.: Certificate transparency - known logs. https://www.certificate-transparency.org/known-logs, accessed July 23, 2017
19. VanderSloot, B., Amann, J., Bernhard, M., Durumeric, Z., Bailey, M., Halderman, J.A.: Towards a complete view of the certificate ecosystem. In: Proceedings of the 2016 ACM on Internet Measurement Conference, IMC 2016, Santa Monica, CA, USA, November 14-16, 2016. pp. 543–549 (2016)