POSTER: A Comprehensive Study of Forged Certificates in the Wild

Mingxin Cui
Institute of Information Engineering, CAS
School of Cyber Security, University of Chinese Academy of Sciences
cuimingxin@iie.ac.cn

Zigang Cao
Institute of Information Engineering, CAS
School of Cyber Security, University of Chinese Academy of Sciences
caozigang@iie.ac.cn

Gang Xiong*
Institute of Information Engineering, CAS
School of Cyber Security, University of Chinese Academy of Sciences
xionggang@iie.ac.cn

Junzheng Shi
Institute of Information Engineering, CAS
School of Cyber Security, University of Chinese Academy of Sciences
shijunzheng@iie.ac.cn

ABSTRACT

With the widespread use of SSL, many issues have been exposed as well. Forged certificates used for MITM attacks or proxies can make SSL encryption useless easily, leading to privacy disclosure and property loss of careless victims. In this paper, we implement a large scale of passive measurement of SSL/TLS and analyze the forged certificates in the wild comprehensively. We measured SSL/TLS connection for 16 months on two large research networks, which provided a total of 100 Gbps bandwidth. We gathered nearly 135 million leaf certificates and studied the forged ones. Our findings reveal main reasons of signing forged certificates, and show the preference of them. Finally, we find out several suspicious servers that might be used for MITM.

CCS CONCEPTS

• **Networks** → *Network measurement*; • **Security and privacy** → *Network security*;

KEYWORDS

Forged Certificate, SSL MITM, Passive Measurement

1 INTRODUCTION

SSL/TLS (we refer to SSL/TLS as SSL for brevity in this paper) is the most widely used encryption protocol to ensure the security of network communication. X.509 certificate plays an important role in SSL PKI (Public Key Infrastructure), which is the basis of SSL encryption framework. With the widespread use of SSL, many issues have been exposed as well. In this paper, we focus on the status quo of forged certificates.

A certificate is used to identify the peer server or client in SSL handshake period. An SSL MITM attack usually uses a forged certificate to deceive careless users, leading to the privacy disclosure and property loss of the victims. The number of forged certificates is increasing as the widespread use of HTTPS, so it is necessary to conduct a comprehensive study of the status quo of forged certificates.

Though many researchers have published their works on the certificate ecosystem [2] [3] [4] and MITM of SSL [1] [5], there are few papers focusing on the forged certificates in the wild. Huang et al [6] implemented a method to detect the occurrence of SSL MITM attack on Facebook and analyzed forged SSL certificates of Facebook in 2014, but they only studied the forged certificates of Facebook.

In this paper, we conduct a comprehensive study of forged certificates in the wild. We implemented a 16-month passive measurement from November 2015 to February 2017 to collect the real-world SSL certificates on two large research networks. These two networks can totally provide over 100Gbps bandwidth, covering more than 17 million IPv4 addresses and nearly 30 million actual users. During the measurement, we collected nearly 135 million leaf certificates totally, including more than 64 million forged ones. After an in-depth analysis and traceability, we find that though many forged certificates can be attributed to antivirus software, security gateways, content filters, and proxies, which has been mentioned in [6]. Besides, we find a considerable number of forged certificates are issued by a serial of similar self-signed CAs. The analysis result shows that certificates related to finance prefer to be forged.

Our contributions can be summarized as follows. Firstly, we conducted a large scale of passive measurement to draw an overall scene of forged certificates in the wild. Secondly, we analyze the forged certificates from three aspects: CA, certificate, and server. Our analysis reveals the main causes and the preference of forged certificates. Thirdly, we find out several suspected SSL MITM attacks and trace out the suspicious servers.

[*] Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CCS '17, October 30-November 3, 2017, Dallas, TX, USA
© 2017 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-4946-8/17/10.
https://doi.org/10.1145/3133956.3138851

Table 1: Certificates Dataset Size

Certificate Types	#(Unique Certificates)	Percentage
CAs	879,707	0.65%
Leaf Certs(Not Forged)	70,459,293	51.86%
Leaf Certs(Forged)	64,528,037	47.49%
Totally	135,867,037	100%

Table 2: Top 20 Issuers of Forged Certificates

Cert Issuer CN	
Cisco Umbrella Secondary SubCA *	FortiGate CA
Zscaler Intermediate Root CA (*.net)	Websense
Sophos Web Appliance	mitmproxy
OpenDNS Intermediate nrt-SG	Web Gateway
Avella School District Proxy CA	Essentra
Sophos SSL CA_C01001BKK84Y2F1	*.securly.com
Bureau Veritas	Self-signed
Lightspeed Rocket	Egedian
DO_NOT_TRUST_FiddlerRoot	Gadang Proxy CA
McAfee Web Gateway	SSL-SG1-HK1

The remainder of this paper is structured as follows. Section 2 elaborates our measurement framework and the details of collected data. We show the real-world forged certificate situation and try to trace and identify SSL MITM attacks in Section 3. Section 4 concludes this paper and list the future work we intend to do.

2 MEASUREMENT AND DATASET

2.1 Passive Measurement Framework

In order to study the malware and Trojan viruses spreading in the SSL encrypted networks, we implemented passive measurement on two large research networks, namely *China Education and Research Network* and *China Science and Technology Network* from May 2016 to May 2017. These networks could provide 61,440 Mbps and 47,104 Mbps bandwidth respectively, covering more than 17 million IPv4 addresses and nearly 30 million users. Our program collected the certificates and connection information (mainly the servers and connection counts) after an anonymous processing. Useful data would be added into the corresponding dataset. The framework of our passive measurement is shown in Figure 1.

2.2 Dataset

During the measurement, we collected nearly 136 million unique certificates, including nearly 135 million end-entity ones and 879,707 CAs. We then gathered more than 64 million forged certificates by validating certificate chains, as shown in Table 1. We parsed the forged certificates and extracted basic information to compose our dataset. And for each SSL connection, we stored the server ip, server port, and the basic statistics as well.

Table 3: Top 10 Forged Certificates

Cert Subject CN	#(Forged Certs)
*.tmall.com	131,091
*.taobao.com	119,496
*.aliexpress.com	13,211
img.alicdn.com	10,290
yy.com	6,961
www.amazon.com	6,137
api.paypal.com	5,440
*.tanx.com	4,803
mobile.paypal.com	4,745
ru.aliexpress.com	4,684

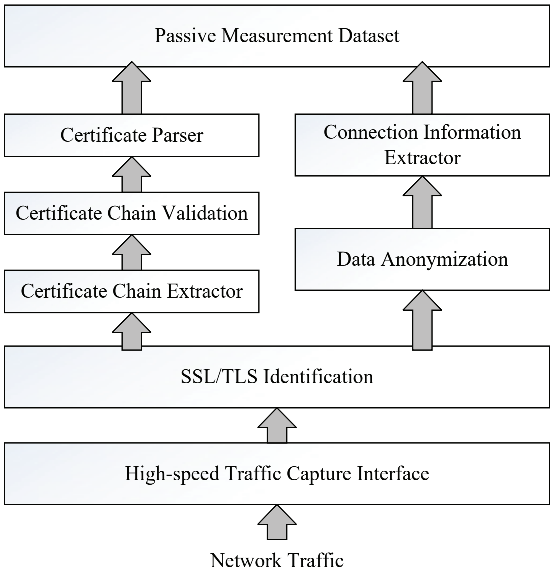
Passive Measurement Dataset
Certificate Parser
Connection Information Extractor
Certificate Chain Validation
Data Anonymization
Certificate Chain Extractor
SSL/TLS Identification
High-speed Traffic Capture Interface
Network Traffic

Figure 1: Passive Measurement Framework

2.3 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 encrypted before collected by the measurement system. Thus, we do not know, and actually do not care about, the real client ip address of each SSL connection. We focus on the certificates and corresponding servers, not the user privacy.

3 FORGED CERTIFICATE IN THE WILD

3.1 Issuers of Forged Certificates

We analyzed the forged certificates and tried to find out the main causes. Table 2 lists the top 20 issuers that issued most forged certificates. These issuers could be divided into several classes. Some issuers are related to security products or antivirus softwares, such as *Cisco Umbrella Secondary*

Table 4: Suspicious Malicious MITM Servers

Suspicious Server	#port	#(port,cert)	#connection	Location	Organization	Domain
62.210.69.21	500(2876-3375)	9,067	61,932	France	ONLINE SAS	poneytelecom.eu
62.210.169.111	25 (4772-4796)	268	93,931	France	ONLINE SAS	poneytelecom.eu
195.154.161.44	441(2604-3100)	1,521	2,142	France	liad-Entreprises	poneytelecom.eu
195.154.161.209	500(3336-3835)	9,118	66,125	France	liad-Entreprises	poneytelecom.eu

SubCA *, and *Zscaler Intermediate Root CA (*.net)*, and *Sophos Web Appliance*. Some issuers are used for content filters, such as **.securly.com* and *Egedian*. Some issuers refer to SSL proxies. And there is no doubt that lots of forged certificates are issued for MITM attacks.

3.2 Preference of Forged Certificates

We analyze the forged certificates in two aspects: subject and issuer. From the subject point of view, we reveal the preference of forged certificates. Table 3 shows the top 10 forged certificates. Most of these certificates are related to finance. The reason is obvious: MITM attackers are more interested in the wallets of the victims. This law also applies to the whole dataset. Taking the usage of SSL proxies into account, it's not surprising that many unknown certificates have been forged.

We also find a considerable number of forged certificates are issued by a serial of similar self-signed CAs. The issuer CN of these forged certificates follows the regular expression ”[0-9a-z]{16}”, such as *000b3ae6c82f4368* and *ffec5faf668ae0d6*. More than 400,000 forged certificates are issued by this kind of CAs. And they account for more than 75% of forged certificates in Table 3, involving all items.

3.3 Suspicious MITM Servers

Forged certificates with similar issuer CNs mentioned above caused our attention. We suspected that these forged certificates were issued by the same series of MITM attacks. Thus we analyzed the connection information of the corresponding certificates to verify our suspicion. We randomly selected 3 days connection data on 05/31/2016, 07/31/2016, and 02/28/2017, and extracted the corresponding server information. We obtained totally 796 servers, 2,458 server (*ip,port*) tuples and 21,361 (*ip,port,certificate*) triples from 229,817 connections. We suspected four ip addresses shown in Table 4 should be attributed to MITM attacks for the reasons below:

(1) These ips belong to a same Organization and locate in the same country.
(2) These ips covered more than 93% triples and more than 97% connections.
(3) The corresponding port number of each ip is incremented continuously, and the number of connections for each (*ip,port,certificate*) triple is less than 10.
(4) The forged certificates used by these servers are mainly related to less than 15 domains, most of which provided financial services.

We would focus on these servers and try to verify our suspicion.

4 CONCLUSION AND FUTURE WORK

In this paper, we implemented a large scale of passive measurement to study the status quo of forged certificates. We collected more than 64 million forged certificates and conducted an in-depth study. Our analysis revealed the causes and preference of forged certificates. We also detected several suspected SSL MITM attacks due to the analysis and traced out the suspicious servers.

Future Work. Up to now our focus was on the analysis of forged certificates based on the passive measurement. A passive HTTPS scan of Alexa Top 1 million domains would be implemented next. Combined with the results of passive and active measurement, we could get a more accurate and real scene of forged certificates in the wild. What's more, we would train a machine learning model to detect SSL MITM attacks using the dataset we've collected.

ACKNOWLEDGMENTS

This work is supported by The National Key Research and Development Program of China (No. 2016YFB0801200) and The National Natural Science Foundation of China (No. 61602472, No. U1636217).

REFERENCES

[1] Italo Dacosta, Mustaque Ahamad, and Patrick Traynor. 2012. Trust no one else: Detecting MITM attacks against SSL/TLS without third-parties. In *European Symposium on Research in Computer Security*. Springer, 199–216.

[2] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J Alex Halderman. 2015. A search engine backed by Internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 542–553.

[3] Zakir Durumeric, James Kasten, Michael Bailey, and J Alex Halderman. 2013. Analysis of the HTTPS certificate ecosystem. In *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 291–304.

[4] Ralph Holz, Lothar Braun, Nils Kammenhuber, and Georg Carle. 2011. The SSL landscape: a thorough analysis of the x. 509 PKI using active and passive measurements. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 427–444.

[5] Ralph Holz, Thomas Riedmaier, Nils Kammenhuber, and Georg Carle. 2012. X. 509 forensics: Detecting and localising the SSL/TLS men-in-the-middle. *Computer security–esorics 2012* (2012), 217–234.

[6] Lin Shung Huang, Alex Rice, Erling Ellingsen, and Collin Jackson. 2014. Analyzing forged SSL certificates in the wild. In *Security and privacy (sp), 2014 ieee symposium on*. IEEE, 83–97.