Exploring HTTPS Certificate Ecosystem: Analyzing the Entire IPv4 Address Space

1st Sina Keshvadi
Faculty of Science
Thompson Rivers University
Kamloops, BC, Canada
SKeshvadi@tru.ca

2nd Yogesh Sharma
Faculty of Engineering and Applied Science
University of Regina
Regina, SK, Canada
Yogesh.Sharma@uregina.ca

***Abstract*—HTTPS, a widely adopted protocol for secure communication on the internet, relies on the TLS protocol to ensure encryption and authentication during data transmission. In this study, we conducted a large-scale measurement on the entire IPv4 address space to analyze the TLS certificate ecosystem used in HTTPS. Over eight consecutive days, we found 46.80M hosts with an open 443 port, of which 33.36M (71.2%) successfully completed a TLS handshake, and we collected 27.88M unique SSL/TLS certificates. This paper presents an overview of the certificate status and distribution, including the prevalence of untrusted and expired certificates. We found that TLS 1.2 is still widely used, accounting for 53.80% of all TLS protocol usage, while TLS 1.3 has shown a significant increase in usage, reaching 43.20% of all TLS protocol usage. Our study also investigates the certificate authorities that issued the certificates, revealing a diverse set of organizations, with Let's Encrypt being the most prominent one. We compare our results with a study conducted a decade ago to examine the changes in the TLS certificate ecosystem. The findings propose implications for internet security and highlight the need for improved certificate management and monitoring practices.**

***Index Terms*—Internet Measurement, Internet Security, HTTPS, TLS.**

I. Introduction

The security of online activities has become a crucial concern in today's digital age. To address these concerns, the Hypertext Transfer Protocol Secure (HTTPS) has emerged as a critical security protocol for ensuring secure communication over the internet. HTTPS is built on top of the Transport Layer Security (TLS) protocol, which provides end-to-end encryption for online communication. TLS relies on digital certificates that uses Public Key Infrastructure (PKI) to establish a secure binding between a server's public key and its hostname, thereby providing assurance of the server's identity to the client. The chain of trust in TLS relies on the signing of digital certificates by trusted third-party entities known as Certificate Authorities (CAs) [1].

While the root CA certificate is included in the client's operating system or browser and is signed by a limited set of authorities, intermediate CAs, which can sign trusted certificates for any domain and delegate certificate authority to other entities, are not publicly known. This presents a significant concern for the security of TLS as the chain of trust is only as strong as the weakest intermediate CA. For instance, Turktrust, a Turkish certificate authority, issued intermediate certificates to an unauthorized third party in 2011, allowing the party to issue fraudulent certificates for Google domains [2]. This incident highlighted concerns about the trustworthiness of certain CAs and their ability to secure root certificates. Therefore, to fully understand the security of online communication, it is crucial to study the ecosystem of certificate authorities and the certificates they issue, as well as to identify potential security vulnerabilities and risks in the use of TLS protocols.

In this study, we aimed to gather a comprehensive dataset of all certificates on the internet, which would allow for a better understanding of the ecosystem of certificates and CAs. To achieve this, we used ZMap [3], a high-speed network scanner, to scan the entire IPv4 address space on port 443, which provides HTTPS, for 8 consecutive days from February 01 to February 07, 2023. We identified 46.80 million unique IP addresses where their port 443 was open. We then used ZGrab2 [4], a TLS handshaking tool, to make a connection attempt with these IP addresses and collect information about certificates. Out of the hosts with port 443 open, 33.36 million (71.2%) successfully completed the handshake, and we collected 27.88 million unique certificates during these days.

In our study, we present several key findings regarding the adoption and usage of TLS and certificates on the internet. Our data showed a positive trend towards the adoption of TLS v1.3, with 43.20% of hosts utilizing this protocol. However, we also found that TLS v1.2 remained the most widely used version, accounting for 53.80% of hosts. Surprisingly, we observed that a significant percentage of hosts (3%) still use non-secure versions of TLS, including SSLv3, which has been deemed unsecure for over a decade. Regarding certificate usage, we found that a large majority of certificates were not trusted by browsers, with 85% of the total certificates falling into this category. A particularly interesting observation was that Forinet, a cybersecurity company providing network security appliances, issued 8.66M certificates, none of which were browser-trusted certificates. This trend represents a significant change in the certificate ecosystem.

We also analyzed the distribution of certificates among authorities and found that a small number of organizations controlled a large percentage of all trusted certificates. Additionally, while the emergence of free certificate issuers like Let's Encrypt has led to increased usage of trusted certificates, we still found a significant number of self-issued certificates in our dataset. Furthermore, our analysis of the keys and signatures used to sign leaf certificates revealed that only 36.1% of certificates used the current best practices for signing keys. Finally, we compared our results to a similar study conducted in 2013 [1], highlighting the significant changes and trends that have occurred in the certificate ecosystem over the past decade.

The rest of this paper is organized as follows: Section II provides an overview of background concepts and related works. Section III details the methodology and experimental setup employed to conduct measurements and analyze the data. The findings obtained from our analysis are presented in Section IV. Ethical considerations pertaining to this research are discussed in Section V. Section VI presents the limitations of our study and suggests avenues for future research. Finally, we conclude our study in Section VII.

II. Background and Related Work

A. Background

HTTPS (Hypertext Transfer Protocol Secure) is a protocol designed to provide secure communication over the internet. HTTPS is an extension of the HTTP protocol, which is the foundation of communication on the World Wide Web. The primary difference between HTTP and HTTPS is the use of SSL/TLS (Secure Sockets Layer/Transport Layer Security) protocol for secure communication [5].

TLS [6] is a cryptographic protocol that provides secure communication over the internet by encrypting data in transit. It is the successor to SSL, and TLS v1.2 and v1.3 are currently the most widely used versions of the protocol [7]. The TLS protocol relies on a public key infrastructure (PKI) to establish trust between parties. PKI is a system of digital certificates and certificate authorities that enable secure communication by verifying the identity of communication endpoints.

Certificates are digital documents that contain information about the identity of communication endpoints, such as domain names and public keys. Certificates are issued by certificate authorities (CAs), which are trusted third-party organizations that verify the identity of certificate owners. CAs form a hierarchy, with a small number of root CAs at the top of the hierarchy, issuing certificates to intermediate CAs, which in turn issue certificates to end entities such as websites.

For readers who require a more detailed explanation of the TLS public key infrastructure, we suggest referring to RFC 5280 [8].

B. Related Works

One of the earliest large-scale measurement study of certificates on Internet was the SSL Observatory Project conducted in 2010 by the Electronic Frontier Foundation (EFF) [9]. The project aimed to analyze various aspects of the certificate ecosystem by scanning the entire IPv4 address space over a three-month period. Specifically, the project focused on identifying organizations that controlled a valid signing certificate. To achieve this, they used NMAP to find hosts listening on tcp 443 and collected x.509 certificates used for HTTPS on the internet. The researchers also checked for odd behavior, identified trusted intermediaries such as foreign security agencies and companies, and scrutinized CAs. Although the study was never formally published, the EFF's work provided the first glimpse into the HTTPS certificate ecosystem, which inspired several subsequent studies.

Holz et al. [10] conducted an extensive passive measurement study to analyze the quality of the X.509 Public Key Infrastructure (PKI) used in critical protocols like HTTPS and IMAPS. The authors found that the certification processes of the PKI lacked stringent standards, resulting in numerous certificates that do not meet the requirements of a secure PKI. The study discovered that only 18% of the certificates would be accepted without a warning by a client utilizing the Mozilla Root Store, and the situation was even more dire for self-signed certificates. Additionally, the paper identifies some positive trends, such as the use of secure ciphers with acceptable key lengths and the increasing use of intermediate certificates.

Durumeric et al. [1] investigated into the HTTPS certificate ecosystem by performing 110 Internet-wide scans over a 14-month period. The study investigates the trust relationships among root authorities, intermediate authorities, and the leaf certificates used by web servers, ultimately identifying and classifying more than 1,800 entities that are able to issue certificates vouching for the identity of any website. The authors also uncover practices that may put the security of the ecosystem at risk and identify frequent configuration problems that lead to user-facing errors and potential vulnerabilities.

In a recent study, Holz et al. [7] explored the deployment, uptake, and use of TLS 1.3 in wild. Using a combination of active domain scans, passive monitoring of large networks, and crowd-sourcing efforts on Android devices, the authors track the deployment and adoption of TLS 1.3 from the early design phase to more than a year after standardization. The study shows that, in contrast to TLS 1.2, which took more than five years to adopt due to severe attacks on previous versions, TLS 1.3 is being deployed quickly and without significant security concerns. The study highlights the need for multi-perspective studies on the evolution of the internet and cannot be captured by a single dataset alone.

Our study presents a unique methodology for scanning the entire IPv4 address space to collect data on certificates using TLS handshakes. This enables us to provide a comprehensive dataset for characterizing the certificate ecosystem, including certificate authorities, TLS version adaptation, and key signature algorithm usage. We also provide empirical evidence of the progress of the HTTPS ecosystem since a previous study conducted in 2013 [1].

III. Methodology

In order to collect a comprehensive dataset of all certificates on the internet, we used ZMap [3], a fast single packet network scanner, to detect Internet-facing systems by scanning the entire IPv4 address space on port 443 for eight consecutive days from February 01 to February 07. We conducted our experiments using two Ubuntu Google Cloud Compute Engine instances in the us-west4-b zone, each equipped with 8 vCPUs. Given that ZMap's performance relies on the computational power and network bandwidth of the system, we conducted an initial scan of 1% of the address space at various scan rates to determine the optimal setting for our measurement environment. Through observation, we determined that a scan rate of 40,000 packets per second yielded the highest hit rate when utilizing our Google Cloud Engine. Although ZMap is capable of scanning the entire IPv4 IP address space in less than an hour, our selected scan rate ensured that each scan was completed within a 4-hour timeframe. After each scan iteration, we compiled a list of IP addresses that responded to SYN scans on port 443.

To perform a TLS handshake with the identified IP addresses and retrieve their respective certificates for in-depth analysis, we employed ZGrab2 [4], a stateful application layer scanner. ZGrab2 facilitated the retrieval of certificate details such as the subject, issuer, key signature algorithm, public key length, and other relevant information, allowing us to gather comprehensive insights into the TLS certificate landscape and assess various aspects of the ecosystem.

In order to mitigate potential concerns regarding the using of ZMap and ZGrab2, which were both developed by the same author and may share vulnerabilities, we devised a scraper script that utilized OpenSSL to scan a subset of 1000 IP addresses identified by ZMap. Subsequently, we compared the outcomes obtained from our script with the results generated by ZGrab2. The two scans yielded highly consistent results, with only minor variations of fewer than 5 certificates. This comparison was significant as it allowed us to verify the accuracy and consistency of our results, and facilitated comparisons with previous studies that had employed OpenSSL as a scanning tool.

We then processed the collected certificates to identify duplicates and certificates that did not comply with the X.509 certificate format. Parsing the results of TLS handshake is challenging due to the diverse information obtained from diverse hosts. We wrote scripts to parse ZGrap output containing handshake transcripts. Finally, we stored our data in Google's BigQuery and used SQL scripts to analyze the data.

To enable the replication of our measurements and the reproducibility of our results, the scripts used in our study, including those for running ZMap, ZGrab2, and analyzing the collected data, have been made available in the project's dedicated GitHub repository [11].

IV. Results

This section presents an analysis of our dataset of internet certificates collected during the measurement process. Subsection IV-A focuses on the hosts discovered through a complete IPv4 address space scan using ZMap, with an emphasis on port 443. Subsection IV-B provides a thorough analysis of the collected certificates by performing TLS handshakes with the responsive IP addresses.

A. Distribution of Active IP Addresses and Open Ports

Using ZMap, we conducted a comprehensive scan of the entire IPv4 address space to identify hosts with an open port 443 for TLS connections. Over the course of eight consecutive days, we collected a vast dataset consisting 46.80M distinct IP addresses with an open 443 port. Among these, 33.36M (71.2%) of those successfully completed a TLS handshake. Our finding suggests an increase in the number of hosts with an open 443 port compared to the previous study conducted a decade ago, which reported 33M hosts with an open 443 port [1]. However, the proportion of hosts that successfully completed a TLS handshake, 71.2%, remained consistent with the previous research, which reported a success rate of 67% for hosts that completed a TCP handshake on port 443. In the following paragraphs, we present several noteworthy observations derived from the pool of IP addresses collected in our study.

1) Analysis of Commonly Used Ports: To better understand the prevalence of open ports, we conducted measurements on other commonly used ports. The main objective of this study was to gain insights into the Internet surface attack landscape (entry points that are visible and accessible to attackers) to provide better protection against potential threats. To minimize scanning traffic, we conducted measurements on only 1% of the random IPv4 address space. Our findings, as shown in Table I, revealed that open port 80 had the highest number of IP addresses with 537.1K, followed by port 443 with 313.7K. Port 7547, which is used by TR-069 protocol for remote management of devices, ranked third, while port 22, used for SSH, and port 5060, used for VoIP/SIP, followed it. Interestingly, we also observed that many IP addresses offer HTTP service on ports other than the default port 80, particularly on ports 8080, 8880, and 8443.

TABLE I
Top-10 Open Ports by Number of IP Addresses

Port Number	Number of IP Addresses
80	537.1K
443	313.7K
7547	254.9K
22	217.6K
5060	97.8K
500	91.0K
21	83.1K
123	74.2K
25	67.4K
8443	61.0K

2) TLS Version: The secure and reliable transmission of information over the Internet depends heavily on the maintenance of an up-to-date and secure TLS protocol. Table II presents a summary of the usage of TLS protocols observed in our dataset. The results indicate that TLS 1.2 is still widely used, accounting for 53.80% of all TLS protocol usage. In contrast, TLS 1.3 has shown a significant increase in usage, reaching 43.20% of all TLS protocol usage. This figure is much higher than the 23.6% reported in a passive measurement study conducted in April 2018 [12]. However, the usage of TLS 1.0 and TLS 1.1 is relatively low, with only 2.72% of all TLS protocol usage. These results highlight the importance of upgrading to the latest and most secure protocol, TLS 1.3, as usage of older protocols such as TLS 1.0 and TLS 1.1 poses a security risk due to known vulnerabilities.

TABLE II
TLS PROTOCOL USAGE

Protocol	Count	Percentage
TLS 1.2	25,217,498	53.80%
TLS 1.3	20,234,285	43.20%
TLS 1.0	1,269,285	2.21%
TLS 1.1	338,343	0.71%
SSLv3	36,689	0.08%

3) Geographical Distribution of IP Addresses with Open 443 Port: In order to understand the geographical distribution of IP addresses with open 443 port, we obtained data on the number of IP addresses with open 443 port by country. This information was particularly relevant as we aimed to investigate the country of certificate authorities in the subsequent sections. Table III displays countries with the highest number of responding IP addresses. The United States had the highest number of responding IP addresses with 49.4 million, followed by Germany with 16.28 million, China with 13.55 million, and UK with 13.48 million. These findings suggest that the distribution of responding IP addresses is not evenly spread across the globe, with certain countries having a higher number of responding IP addresses than others. It's important to note that the numbers presented here represent the number of IP addresses that responded to the scan, not the total number of IP addresses in each country. There could be many factors such as differences in internet infrastructure, security measures, or firewall configurations that affect out results.

TABLE III
NUMBER OF IPS WITH OPEN PORT 443 BY COUNTRY (TOP 10)

Country	IPv4 Addresses	Open Port 443
United States	49.41M	16.36M
Germany	16.28M	3.56M
China	13.55M	2.31M
United Kingdom	13.48M	1.70M
South Korea	12.73M	810.33K
Italy	9.45M	1.30M
Canada	9.23M	898.98K
Japan	7.17M	1.95M
France	7.72M	1.45M
Japan	7.17M	1.25M

B. Certificates

Each day we received an average of 8.9M certificates and we collected 27.88M unique certificates in overall during a period of 8 days between Feb 01, 2023 to Feb 07, 2023. In the subsequent paragraphs, we present our analysis of the certificates gathered during our measurement.

1) Certificates Status and Distribution: Table IV presents an overview of the status and distribution of certificates collected in our analysis. Surprisingly, we observed that the majority of certificates were untrusted, accounting for 85% of the total certificates, which is a concerning observation. We investigated untrusted certificate authorities, which will be discussed in the next subsection. Furthermore, while 66% of certificates were unexpired, there were still 9.35 million expired certificates in circulation. In addition, we observed a significant number of self-signed certificates (3.62M) and a relatively small number of revoked (24.14K) and precertificates (59.55K). Precertificates are intermediate certificates that are signed by a CA and can be used to generate end-entity certificates, allowing for faster certificate issuance and deployment.

It is worth noting that some of the expired certificates may have been intentionally left in place to support legacy systems or devices that cannot be easily updated. Additionally, some network devices, such as routers or firewalls, may use self-signed certificates or certificates signed by a private CA for internal communication, which can also contribute to the high number of untrusted and expired certificates. However, the presence of such a large number of untrusted and expired certificates still poses a significant security risk, as it may allow attackers to perform man-in-the-middle (MITM) attacks, intercept sensitive data, and compromise the security of Internet traffic.

Among trusted certificates, domain validated (DV) certificates constituted the vast majority (77%), followed by organization validated (OV) certificates at 16%. Notably, only 2.7K (0.01%) certificates were of the highest validation level, extended validation (EV) certificates. Our analysis also revealed the presence of several certificates that were not labeled as any of the common types: DV, OV, or EV certificates. One possible explanation for this is that these certificates may be part of specialized certificate types for specific purposes or industries, such as code signing or email signing certificates, which may not fall neatly into the DV, OV, or EV categories.

TABLE IV
CERTIFICATES OVERVIEW

Certificate Type	Count
Total	27.88M
Trusted	4.18M
Untrusted	23.70M
Unexpired	18.53M
Expired	9.35M
Self-Signed	3.62M
Revoked	24.14K
Precertificates	59.55K
Domain Validated (DV)	12.84M
Organization Validated (OV)	658.44K
Extended Validation (EV)	2.7K

2) Certificate Authorities: Table V presents a breakdown of the organizations and their signed leaf certificates, along with the percentage of trusted certificates. From 27.88M unique certificates that was collected from hosts that completed a TLS handshake, a total of 4.42M (15.8%) were browser trusted certificates. Surprisingly, This percentage is lower than the 48% reported in a previous study [1], which is surprising given the introduction and growth of "Let's Encrypt", a free certificate authority that is well-sponsored by big tech companies. One would expect this percentage to have increased over the past decade. The table shows that Let's Encrypt accounting for for 37.18% of the signed leaf certificates, of which 57.01% are trusted.

Fortinet follows closely with 31.08% of the signed leaf certificates, none of which are trusted. Fortinet is a cybersecurity company that provides various network security appliances, including firewalls and VPN gateways. It issues its own certificates for various purposes, such as for SSL VPN connections, secure communication between Fortinet devices, and ensuring that the certificates are trusted by its own devices and are configured to work properly with the specific features and functions of its products. This practice can, however, raise vulnerabilities, since untrusted certificates can create security risks. Notably, after conducting our measurement and while writing our observations for publication, FortiGuard Lab issued an alert regarding an improper certificate validation vulnerability that could "allow a remote and unauthenticated attacker to perform a Man-in-the-Middle attack" [13].

Other certificate authorities have a smaller percentage of signed leaf certificates, with varying levels of trust. Notably, there has been a significant drop in the proportion of certificates issued by commercial CAs such as GoDaddy, Symantec, and DigiCert, since the previous study conducted by Durumeric et al. (2013) [1]. This highlights the growing impact of Let's Encrypt on the ecosystem of certificates and certificate authorities.

TABLE V
ORGANIZATION AND SIGNED LEAF CERTIFICATES

Organization	Signed Leaf Certs	Trusted Certs
Let's Encrypt	10.36M (37.18%)	2.52M (57.01%)
Fortinet	8.66M (31.08%)	0 (0%)
self-sigend	3.62M (12.98%)	0 (0%)
Google Trust Services LLC	1.26M (4.52%)	335.84K (7.59%)
cPanel, Inc.	509.30K (1.83%)	341.78K (7.73%)
Cloudflare, Inc.	479.08K (1.72%)	476.33K (10.77%)
WeLinkGame	259.25K (0.93%)	0 (0%)
Amazon	185.90K (0.67%)	177.81K (4.02%)
ZeroSSL	184.82K (0.66%)	95.95K (2.17%)
DigiCert Inc	181.31K (0.52%)	171.62 (3.88%)
Acme Co	163.32K (0.59%)	0 (0%)
GoDaddy.com, Inc.	42.87K (0.15%)	41.60K (0.94%)
Others	1.97M (7.08%)	260.75K (5.89%)
Total	**27.88M (100%)**	**4.42M (100%)**

3) Self-Signed Certs: As shown in Table V, a significant number of certificates (3.62 million or 12.98%) were found to be self-signed, meaning they were signed by the domain owner rather than a trusted Certificate Authority (CA). The use of self-signed certificates is generally discouraged in favor of obtaining certificates from trusted CAs, as it can pose security risks and create confusion for users. Despite this, some organizations may choose to use self-signed certificates for various reasons, such as for internal testing or development purposes, or if they are unable or unwilling to obtain certificates from a trusted CA. We observed that the Chinese online gaming company WeLinkGame and "Acme Co" (a fictional company commonly used in examples and illustrations) were the most common issuers of self-signed certificates, accounting for 7.16% and 4.51% of all self-signed certificates, respectively.

4) Key Distribution for Trusted Signing Certificates: At the time of writing, the recommended security protocol for key algorithms is ECDSA, followed by RSA keys of at least 2048 bits as a secondary option [14]. Table VI presents the distribution of certificate key types used by various issuers. It is evident from the table that RSA keys are currently the most widely used, accounting for over 59% of all certificates issued. The majority of RSA certificates have a key length of 2048 bits, while only a small proportion of certificates have key lengths of either 1024 or 3072 bits. Additionally, 36.1% of certificates have a 256-bit ECDSA key. ECDSA is known for its efficiency and security and is widely used in digital signatures and key exchange, including SSL/TLS certificates, SSH keys, and blockchain technologies. The two main issuers of ECDSA certificates are Fortinet and Let's Encrypt, with Fortinet accounting for the majority of ECDSA certificates. Additionally, it is worth noting that a significant proportion of non-standard key algorithms such as DSA, EdDSA, or different key sizes for RSA and ECDSA keys are used by self-signed certificates.

A decade ago, a scan conducted by Durumeric et al. [1] revealed that 98.7% of certificates used the compromised SHA-1 algorithm. However, since the time of that study, the SHA-1 algorithm has been compromised. Our current study did not encounter any certificates signed using SHA-1. Additionally, the prior study found no leaf certificates signed by ECDSA keys, with only 0.3% of chains containing an ECDSA key. In contrast, our study discovered a substantial portion of certificates (36.1%) signed using ECDSA (256-bit) keys, suggesting an increasing adoption of ECDSA in the past decade. Further, the earlier study identified a small percentage of certificates ($< 1\%$) signed using insecure MD5 and MD2 algorithms, which we did not observe in our study. These findings underscore the improvements in the certificate ecosystem over the past decade, with the adoption of more secure key algorithms and the retirement of compromised ones.

V. ETHICAL CONSIDERATION

Our study follows the principles of informed consent [15] and ethical best practices [16], [17]. No individuals were involved in our measurement. We utilized ZMap [3], a tool designed to avoid scanning sensitive networks, and we took into account other ethical considerations during the scan. To

TABLE VI
DISTRIBUTION OF CERTIFICATE KEY TYPES

Key Type	Certs	Main Issuers
ECDSA (256-bit)	10.05M (36.1%)	Fortinet (82.75%) Let's Encrypt (9.01%)
RSA (2048-bit)	14.63M (52.5%)	Let's Encrypt (50.08%) Google Trust Services (8.53%)
RSA (4096-bit)	1.69M (6.1%)	Let's Encrypt (89.34%) ZeroSSL (3.8%)
RSA (1024-bit)	746.73K (2.7%)	Archer C5 (19.12%) EX221-G2u (8.93%)
ECDSA (384-bit)	468.04K (1.7%)	Let's Encrypt (39.17%) Exim Developers (5.79%)
RSA (3072-bit)	231.22K (0.8%)	Let's Encrypt (79.20%) Exim Developers (9.56%)
Other	441.53K (1.58%)	self-signed (78.97%) 2Wire (0.11%)

ensure the privacy of individuals, we refrained from collecting any personal or sensitive information, including client IP addresses and traffic payloads. Furthermore, we implemented measures to minimize the potential risk of causing harm to online servers during our active scans.

VI. LIMITATIONS AND FUTURE DIRECTIONS

Our study has certain limitations that should be acknowledged and addressed in future research endeavors. Firstly, we were unable to cover TLS certificate usage in the IPv6 address space due to the extensive size of the IPv6 address pool. Future research should consider the collection of IPv6 addresses through passive measurement techniques over an extended period of time. Second, our analysis was confined to the use of certificates on port 443, and it is possible that there are other servers behind each IP address that use certificates on different ports. Finally, our analysis was limited to the use of certificates and did not consider other aspects of TLS security, such as the use of strong cipher suites, certificate pinning, or other security features that can help mitigate attacks against the TLS protocol.

In terms of future work, one area of interest is to determine the reasons behind untrusted and expired certificates. Furthermore, more studies are required to examine the potential of automated techniques, such as machine learning, to improve the detection of malicious certificates and reduce the likelihood of attacks on the TLS protocol.

VII. CONCLUSION

In this paper, we presented a comprehensive analysis of certificate ecosystem on the IPv4 address space, focusing on their usage, distribution, and security implications. First, we observed that a majority of TLS certificates on the Internet were untrusted, accounting for 85% of the total certificates. This is a concerning observation, as untrusted certificates pose a significant security risk, allowing attackers to perform man-in-the-middle attacks and compromise the security of Internet traffic. Notably, during the writing of this paper, Fortinet issued an alert regarding vulnerabilities in their certificate systems, underscoring the importance of the study's findings. Second, we found that the distribution of certificate types was heavily skewed towards domain validated (DV) certificates, which indicates that there is a lack of adoption of EV certificates, which may have implications for the overall security of the web. Third, we observed a significant number of expired certificates in circulation, accounting for 33.5% of the collected certificates. While some of these certificates may have been intentionally left in place to support legacy systems or devices, their presence still poses a security risk, as they can be used by attackers to perform attacks. Finally, our analysis revealed several free certificate authorities like Let's Encrypt and ZeroSSL are issuing significant proportion of certificates, which has had great impact on usage of certificate issued by commercial CAs.

REFERENCES

[1] Z. Durumeric, J. Kasten, M. Bailey, and J. A. Halderman, "Analysis of the https certificate ecosystem," in *Proceedings of the 2013 conference on Internet measurement conference*, 2013, pp. 291–304.

[2] M. Coates, "Revoking trust in two turktrust certificates," 2013.

[3] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications." in *USENIX Security Symposium*, vol. 8, 2013, pp. 47–53.

[4] "ZGrab 2.0," https://github.com/zmap/zgrab2, accessed: [May 7, 2023].

[5] S. Satija and R. Chatterjee, "Blindtls: Circumventing tls-based https censorship," in *Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet*, 2021, pp. 43–49.

[6] E. Rescorla, "Rfc 8446: The transport layer security (tls) protocol version 1.3," 2018.

[7] R. Holz, J. Hiller, J. Amann, A. Razaghpanah, T. Jost, N. Vallina-Rodriguez, and O. Hohlfeld, "Tracking the deployment of tls 1.3 on the web: A story of experimentation and centralization," *ACM SIGCOMM Computer Communication Review*, vol. 50, no. 3, pp. 3–15, 2020.

[8] S. Boeyen, S. Santesson, T. Polk, R. Housley, S. Farrell, and D. Cooper, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile," RFC 5280, May 2008. [Online]. Available: https://www.rfc-editor.org/info/rfc5280

[9] P. Eckersley and J. Burns, "An observatory for the ssliverse," 2010.

[10] R. Holz, L. Braun, N. Kammenhuber, and G. Carle, "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*, 2011, pp. 427–444.

[11] Sina Keshvadi, "Github repository: Https certificate ecosystem," https://github.com/Keshvadi/HTTPS-Certificate-Ecosystem-IPv4-Analysis.git, 2023.

[12] P. Kotzias, A. Razaghpanah, J. Amann, K. G. Paterson, N. Vallina-Rodriguez, and J. Caballero, "Coming of age: A longitudinal study of tls deployment," in *Proceedings of the Internet Measurement Conference 2018*, 2018, pp. 415–428.

[13] "Fortianalyzer & fortimanager - lack of client-side certificate validation when establishing secure connections with fortiguard to download outbreakalert," https://www.fortiguard.com/psirt/FG-IR-22-502, accessed May 10, 2023.

[14] F. Boudot, P. Gaudry, A. Guillevic, N. Heninger, E. Thomé, and P. Zimmermann, "The state of the art in integer factoring and breaking public-key cryptography," *IEEE Security & Privacy*, vol. 20, no. 2, pp. 80–86, 2022.

[15] E. Kenneally and D. Dittrich, "The menlo report: Ethical principles guiding information and communication technology research," *Available at SSRN 2445102*, 2012.

[16] S. Keshvadi, M. Karamollahi, and C. Williamson, "Traffic characterization of instant messaging apps: A campus-level view," in *2020 IEEE 45th Conference on Local Computer Networks (LCN)*. IEEE, 2020, pp. 225–232.

[17] S. Keshvadi and C. Williamson, "An empirical measurement study of free live streaming services," in *Passive and Active Measurement: 22nd International Conference, PAM 2021, Virtual Event, March 29–April 1, 2021, Proceedings 22*. Springer, 2021, pp. 111–127.