Poster: On the (In)Security of Government Web and Mail Infrastructure

Evan Lam*, Richard Anderson*, Kurtis Heimerl*, Yurie Ito†, Jonathan Joseph de Koning†,
Adam M. Lange†, Jarrod O'Malley†, Adam Shostack*†, Arastoo Taslim†, Sudheesh Singanamalla*†
* University of Washington, † CyberGreen Institute

***Abstract*—Government web infrastructure is a critical part of today's Internet and the functioning of society. Citizens' interactions with digital government infrastructure needs to be secure since they might contain important and sensitive information. These interactions can be through various web applications providing digital public services, or through communication mechanisms such as email. Government websites and mail servers typically form the long tail of today's Internet and do not appear on large top million Internet datasets making them very understudied. DNS infrastructure forms the center piece for citizens to interact with government services allowing resolution of IP addresses, and enabling email communication and sender policy enforcement between mail service providers. In this poster, due to their inter-dependent nature, we present a comprehensive security evaluation of government web infrastructure covering both web and mail services in addition to understanding the security of the DNS services they rely on. We open source our implementation of the security scanner to the community, invite collaborators to engage with the data periodically scanned, and release the largest public dataset of government hostnames.**

I. Introduction

The usage of Transport Layer Security (TLS) enables secure Internet interactions. Due to its application-neutral nature, it has been adopted to not just secure communications to websites but also to secure communications between mail servers on the Internet through the SMTP STARTTLS extension. Authoritative DNS name servers responsible for the government domains provide the necessary information to client requests. Plain text DNS responses sent from these authoritative nameservers are vulnerable to tampering by network adversaries. However, such attacks on the integrity and authenticity of the responses can be addressed through the usage of DNS security extensions (DNSSEC). In this work, we curate multiple large datasets, filter for government hostnames and release the largest known dataset focused on governments. We build high performance network scanners to measure the adoption and usage of DNSSEC by various domains, in addition to understanding their usage of TLS on web and mail servers and draw insights about their similarity.

II. Dataset & Scanner Tools

We begin our efforts by creating a large dataset of government hostnames. We first synthesize and merge 17 publicly available top million datasets and databases published in previous research efforts [2], [1], and perform a one time extension of the dataset by identifying Subject Alternative Names (SANs) from their TLS certificates. To ensure we only include government hostnames, we crosscheck the eTLD of each hostname against a list of pre-generated and well known government eTLDs. This list was compiled by combining country domains (such as .ly, .us, .au) with common government extensions (such as .gov, .gob, and .go). Since a few governments such as the Canadian provincial governments do not follow this exact format, we manually add them along with special federal (*.fed*), and military (*.mil*) ccTLDs – resulting in the final dataset containing 401216 hostnames.

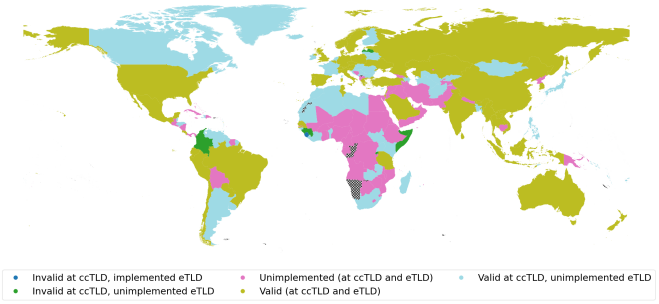


Fig. 1. The categorization of DNSSEC implementation at ccTLD (e.g. au. and .br) and eTLD (e.g. gov.au. and gov.br) levels per country. For countries with multiple government eTLDs *(eg. .gov.us and .gov)* most popular is chosen.

In collaboration with a non-profit partner CyberGreen, we built a custom high performance Internet security scanner that identifies the usage and validity of DNSSEC for a given hostname, performs TLS scans on websites and mail servers returned from the authoritative DNS nameservers. These scans are performed periodically from AWS cloud instances in Singapore with a reserved static IP address and associated reverse DNS PTR records. Due to the sensitive and probing nature of these scans, we duly notify AWS trust and safety teams about the intent to run the scans, provide appropriate opt-out and informational pages about the scanner. As a part of this poster, we open source our scanner implementation written in Golang, the extended dataset, and invite future collaborations from Internet measurement and security communities to leverage the results for longitudinal studies.

III. Results

Of the 401216 hostnames in our final dataset, we successfully resolve the `A` (IPv4) records for 304774 (75.96%) hostnames in our DNS scans, and are able to resolve 308241 hostnames for either IPv4 and IPv6 records, and obtain 59533 unique mail exchange MX records before attempting to make TCP and TLS handshakes. 251819 (75.91%) of these hostnames are served through HTTPS enabled webpages using port 443. Our results indicate that approximately 25% of the government websites across the world still do not use TLS.

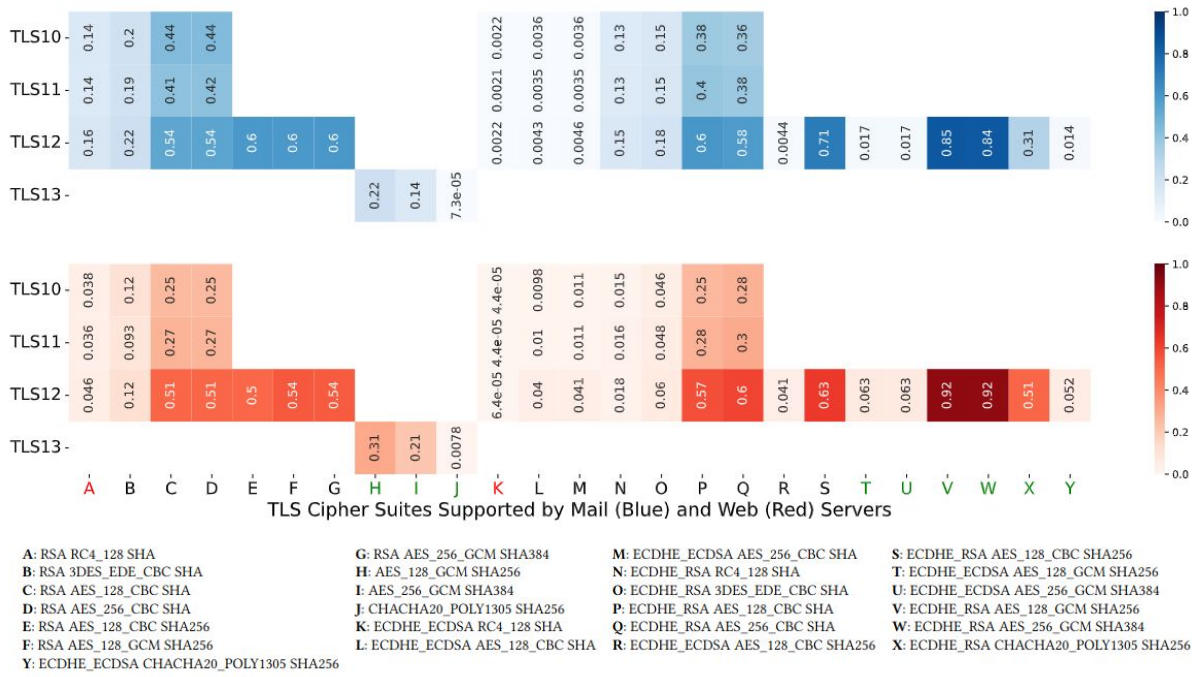


Fig. 2. Heatmaps indicating the percentage of government hostnames supporting various TLS cipher suites grouped by TLS protocol version. The blue heat map on top indicates the protocol usage for 16830 mail servers, and red heat maps indicate the protocol and cipher suite support for 251819 web servers. The cipher suites with red labels (A, K) are classified as insecure, black labels are weak, and the green labels as the secure cipher suites. The key is provided below and includes the key exchange, encryption, authentication, and MAC algorithms used in the cipher suites.

A. *DNSSEC Adoption and Usage Needs Improvement*

We compare the adoption of DNSSEC at the ccTLD level (eg. *.uk*), with that of the respective country government eg. *.gov.uk* which we term as the *eTLD* for this section and present a map as shown in Figure 1 to understand the adoption of DNSSEC at each country and their respective governments' hostnames. We observe 57 countries (light green) have valid DNSSEC enrollments, 68 (light blue) have DNSSEC enabled at the TLD but not at the government eTLD, 68 others (magenta) with the majority in Africa and the Arabian region have not enabled DNSSEC at either level, and 5 countries (dark green) have invalid DNSSEC entries at the ccTLD. Sierra Leone is the only country which has invalid DNSSEC ccTLD entries but attempts to use DNSSEC for their eTLDs. Of the 304774 hostnames, 269171 (67.09%) hostnames do not have DNSSEC enabled. Of the remaining 35603 hostnames that have DNSSEC enabled, we identify that 23676 hostnames are invalid, reducing the number of hostnames with valid DNSSEC records to 11927, a mere 2.97% of the dataset and much below the current DNSSEC adoption rate of $\leq$7% for `.com`.

B. *TLS Protocol Usage Differs Between Mail and Web Servers*

Of the HTTPS supporting hostnames, 32.67% of hostnames support TLS 1.0, 35.08% support TLS 1.1, 99.62% support TLS 1.2, and 59.76% support TLS 1.3. 500 hostnames solely support TLS 1.0, and 23 support only TLS 1.1 – both of which have been deprecated. TLS 1.2 is supported by 89.09% of the mail hosts, followed by 39.91% and 37.3% of the mail hosts supporting TLS 1.0 and TLS 1.1 respectively. Only 25.8% of the mail servers support TLS 1.3. We observe that despite being a smaller set of hosts compared to the web, 924 mail hosts only support TLS 1.0 accounting for 5.49% of the mail hosts indicating a stark contrast with web server infrastructure where only 0.19% of the hosts support TLS 1.0. A larger proportion of mail servers configured by governments tend to use insecure and deprecated TLS protocols compared to their web counterparts and are also shown in Figure 2. A higher percentage of mail hosts compared to web servers, use weak TLS cipher suites based on RSA and cipher block chaining (C, D) possibly because they are supported by TLS 1.0, TLS 1.1, and TLS 1.2. The continuing support for TLS 1.0 and TLS 1.1 puts these web servers at risk of various downgrade attacks such as BEAST, CRIME, and SLOTH.

IV. Conclusion

When understanding the security of web infrastructure, we posit that it is valuable to consider the entire ecosystem – especially for the long tail of the Internet. In this poster, we present a preliminary attempt to reason about and understand security by considering website, mail, and DNS infrastructure simultaneously. Our measurements reveal insights about the global extent of insecurity and inconsistency in the use of secure protocols dependent on the application. We contribute an open-source scanner and the largest curated dataset focused on government hostnames.

References

[1] N. Samarasinghe, A. Adhikari, M. Mannan, and A. Youssef, "Et tu, brute? privacy analysis of government websites and mobile apps," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 564–575.

[2] S. Singanamalla, E. H. B. Jang, R. Anderson, T. Kohno, and K. Heimerl, "Accept the risk and continue: Measuring the long tail of government https adoption," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 577–597.

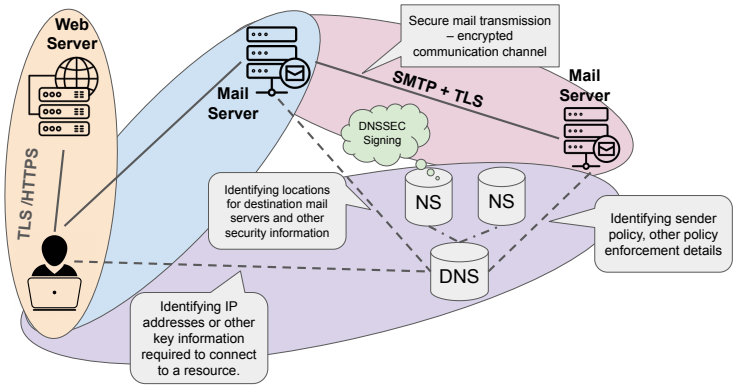
On the (In)Security of Government Web and Mail Infrastructure

Evan Lam[W], Richard Anderson[W], Kurtis Heimerl[W], Yurie Ito[*], Jonathan Joseph de Koning[*], Adam M. Lange[*], Jarrod O'Malley[*], Adam Shostack[W*], Arastoo Taslim[*], Sudheesh Singanamalla[W*]

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

CyberGreen

Motivation



1. Prior research efforts focus on a subset of the ecosystem (colored areas). It is important to consider the entire ecosystem while studying infrastructure security.
2. Very few studies focus on the long tail of the Internet. Government web infrastructure is critical but not popular and hence does not appear on top Internet datasets and is hence very understudied.

Dataset Synthesis and Expansion

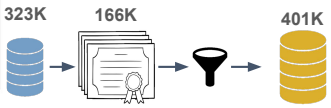
Due to the nature of government website naming schemes, identifying government hostnames is nontrivial. Thus we employ a stringent filtering technique. (`.gov.ccTLD`, `.gouv.`, `.guv.`, `.gub.`, `.go.`, `.gob.` etc..,)

Using TLS SAN fields we expand the hostname dataset by **~160k**

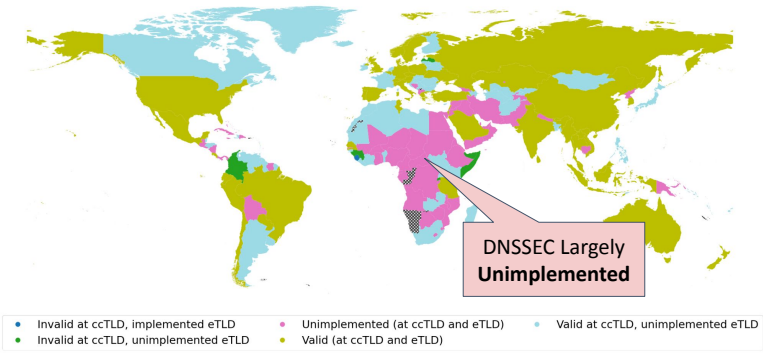
Dataset	#Hostnames	#Domains	% of Dataset
Et Tu Brute	149687	47533	100%
Gov-HTTPS	135408	25742	100%
Domcop 10M	121098	46803	1.21%
Tranco 1M	29334	29334	2.93%
CrUX	29071	3070	2.91%
Majestic 1M	11600	11538	1.16%
Umbrella 1M	11108	3251	1.11%
CISA	9444	9444	100%
GSA	9185	7747	100%
Alexa 1M	9133	9041	0.91%
Cloudflare Radar	7578	7541	0.75%
Gov Cookies	5812	1476	100%
Gov UK Domains	3853	3853	100%
Dot Mil	2776	100	100%
Builtwith 1M	2088	2074	0.20%
Army Mil	301	3	100%
Know Nepal	66	47	100%
Combined Dataset	323164	113762	
SAN Expansion	166304	113762	
Final Dataset	**401216**	115679	

Table 1: Public datasets used to create the final dataset

Open Source Dataset and Scanner

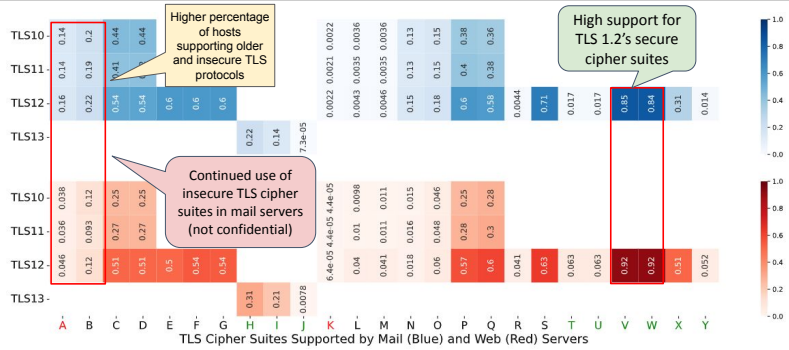


Status of DNSSEC Globally



World map of DNSSEC implementation at eTLD (e.g. `.gov.au`) and ccTLD (eg.`.au`). The most popular eTLD + ccTLD combinations for each corresponding country.

Comparison of TLS in Web + Mail



A: RSA RC4_128 SHA
B: RSA 3DES_EDE_CBC SHA
C: RSA AES_128_CBC SHA
D: RSA AES_256_CBC SHA
E: RSA AES_128_CBC SHA256
F: RSA AES_128_GCM SHA256
Y: ECDHE_ECDSA CHACHA20_POLY1305 SHA256
G: RSA AES_256_GCM SHA384
H: AES_128_GCM SHA256
I: AES_256_GCM SHA384
J: CHACHA20_POLY1305 SHA256
K: ECDHE_ECDSA RC4_128 SHA
L: ECDHE_ECDSA AES_128_CBC SHA
M: ECDHE_ECDSA AES_256_CBC SHA
N: ECDHE_RSA RC4_128 SHA
O: ECDHE_RSA 3DES_EDE_CBC SHA
P: ECDHE_RSA AES_128_CBC SHA
Q: ECDHE_RSA AES_256_CBC SHA
R: ECDHE_ECDSA AES_128_CBC SHA256
S: ECDHE_RSA AES_128_CBC SHA256
T: ECDHE_ECDSA AES_128_GCM SHA256
U: ECDHE_ECDSA AES_256_GCM SHA384
V: ECDHE_RSA AES_128_GCM SHA256
W: ECDHE_RSA AES_256_GCM SHA384
X: ECDHE_RSA CHACHA20_POLY1305 SHA256

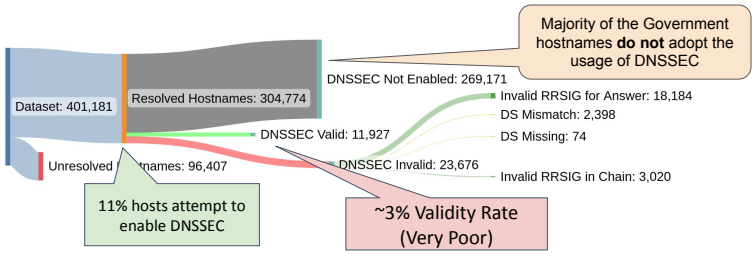
Heatmaps indicating the percentage of government hostnames supporting various TLS cipher suites grouped by TLS protocol version. The blue heat map on top indicates the protocol usage for 16830 mail servers, and red heat maps indicate the protocol and cipher suite support for 251819 web servers. The cipher suites with **red** labels (A, K) are classified as insecure, **black** labels are weak, and the **green** labels as the secure cipher suites..

Reason for Invalidity	Web TLS (N=40590, 16.11%)	Mail TLS (N=4956, 29.44%)
Hostname Mismatch	19085 (47.01%)	3449 (69.59%)
Self Signed Certificates	10473 (25.80%)	225 (4.53%)
Expired Certificates	8176 (20.14%)	667 (13.45%)
Missing SAN Names	2790 (6.87%)	544 (10.97%)

A serious disparity in TLS certificate validity between Web and Mail

Results

Poor DNSSEC Adoption and Use in Govt. Hosts Globally



Conclusion

1. We posit that it is valuable to consider the entire ecosystem -- especially for the long tail of the Internet.
2. Our measurements reveal insights about the global extent of insecurity and inconsistency in use of secure protocols dependent on the application.
   a. The adoption of DNSSEC in government domains/hostnames is low.
   b. Government managed mail and web servers are configured differently with different TLS protocols and cipher suites supported.
   c. More government mail servers serve invalid or self-signed certificates compared to web content servers.
3. We contribute a high performance open source scanner and the largest curated dataset focused on government hostnames.