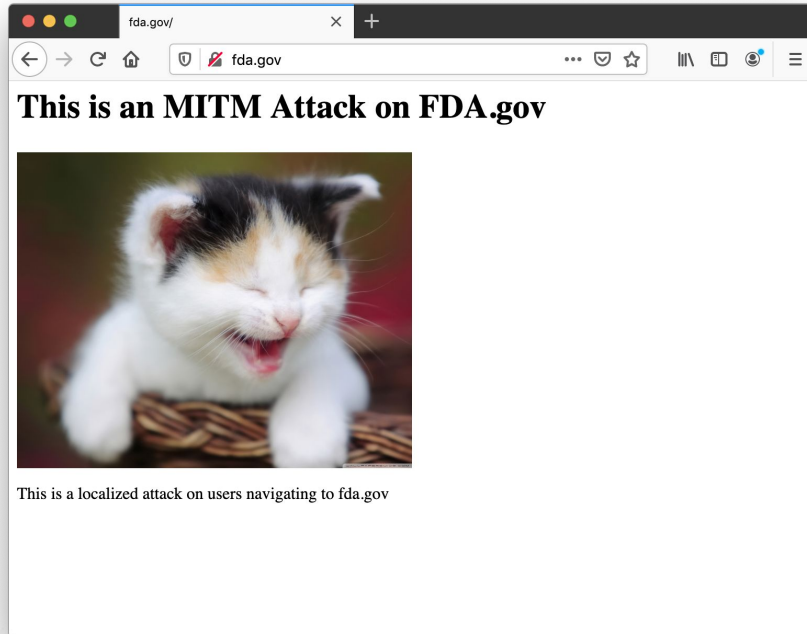# Accept the Risk and Continue: Measuring the Long Tail of Government https Adoption

Sudheesh Singanamalla, Esther Han Beol Jang

Richard Anderson, Tadayoshi Kohno, Kurtis Heimerl

University of Washington

Presented at The Internet Measurement Conference (**IMC'20**)

What is `https`? And Why use it?

This is an MITM Attack on FDA.gov

This is a localized attack on users navigating to fda.gov

Secure version of the `http` protocol

- uses TLS for encryption and authentication
- Default port: 443

Problems with `http`:

- **Lack of privacy/confidentiality**: Users' Internet traffic is visible and can be monitored by an attacker
- **Lack of authentication/identity:** User has no way to validate that the response is actually from the server
- **Lack of integrity:** User cannot validate that the message is not modified.

Motivation: `https` in The Internet Today

Google's HTTPs report[1]

Measures the top 1 million websites on the Alexa top Million list. Published at USENIX Security 2017.

Measuring the Tail

Government websites are critical sites which may not show up in top million datasets. These could include national identity systems, citizen registers, tax, and health information.

1. Felt, Adrienne Porter, et al. "Measuring {HTTPS} Adoption on the Web." 26th USENIX Security Symposium (USENIX Security 17). 2017.

Popular Datasets & New Govt. Dataset

Majestic Million **+** Cisco Million **+** Censys Big Query **+** Alexa Top Million **=** 27,532 unique government websites

# Govt. Websites	Majestic Million	Cisco Million	Tranco Million
Top 1K	56	0	30
Top 10K	508	14	373
Top 100K	2538	433	2351
Top 1M	12445 (**1.24%**)	9296 (**0.93%**)	12293 (**1.23%**)

Chasing the tail…

- Crowdsource unique websites from 23 countries.

27,532 unique government websites **+** amazon mechanical turk **=** **27,794** unique government websites

Chasing the tail...

- Crawl upto 7 levels of Depth.

27,794 unique government websites

+

=

843,561 hostnames which filter down to **301,219** unique hostnames and **134,812** unique government websites
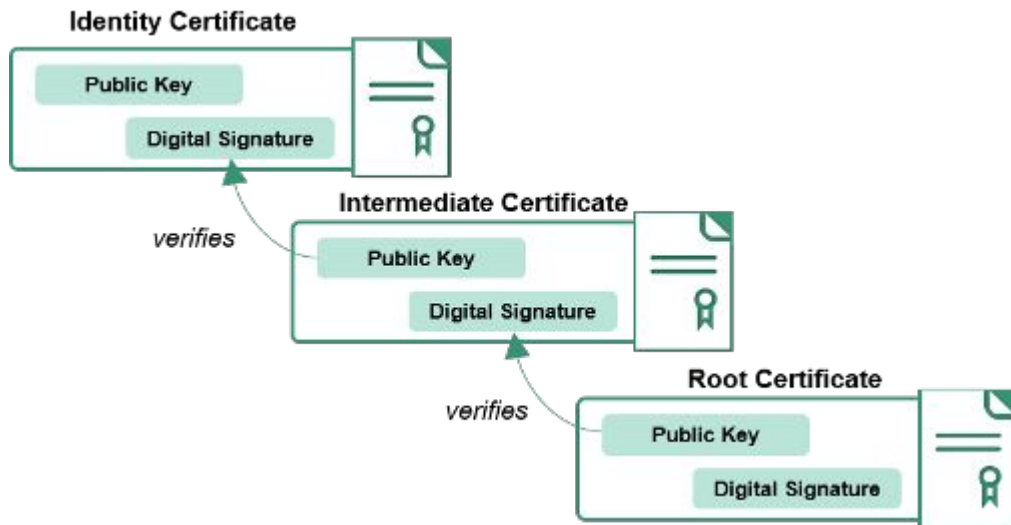
Chasing the tail...

- Explicit whitelist and hand curation from 62 countries.

134,812 unique government websites **+** **=** **135,408** unique government websites

Validating the Certificates

- OpenSSL with the Apple Mac OS trust store imported
- Download the entire certificate chain and validate

Results: At a glance

	Count	%
Total websites considered	**135,408**	**100**
➤ Content served on HTTP only	82,152	60.67
➤ Content served on HTTPS	53,256	39.33
➤ Valid HTTPS Certificates	38,033	71.41
➤ Invalid HTTPS Certificates	15,223	28.58
➤ Hostname Mismatch	5,571	36.59
➤ Unable to get local issuer cert	3,732	24.51
➤ Exceptions	2,619	17.20
➤ Unsupported SSL Protocol	1,929	73.65
➤ Timed out	378	14.43
➤ Connection refused	135	5.15
➤ Connection Reset by peer	141	5.38
➤ Wrong SSL Version Number	11	0.42
➤ TLSv1 Alert Internal Error	9	0.34
➤ SSLv3 Alert Handshake Failure	7	0.26
➤ TLSv1 Alert Internal Proto. V.	8	0.30
➤ Self-signed certificate	2014	13.22
➤ Certificate Expired	838	5.50
➤ Self-signed certificate in chain	347	2.27
➤ Others	102	0.67

Approx.

72%

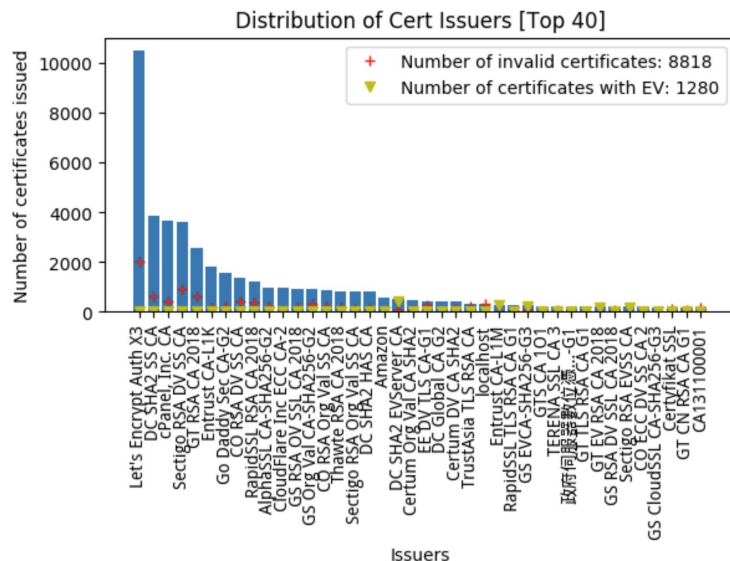Government websites worldwide do *not* have `https`

More than

60%

Serve content only using `http`

More than
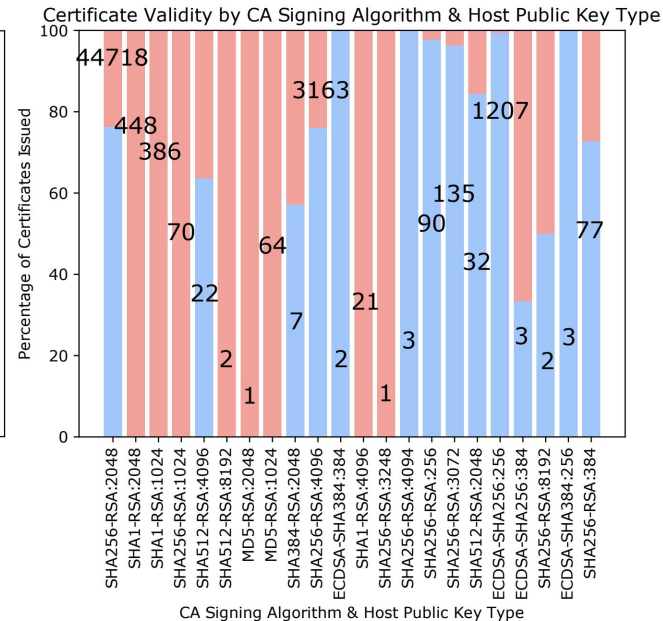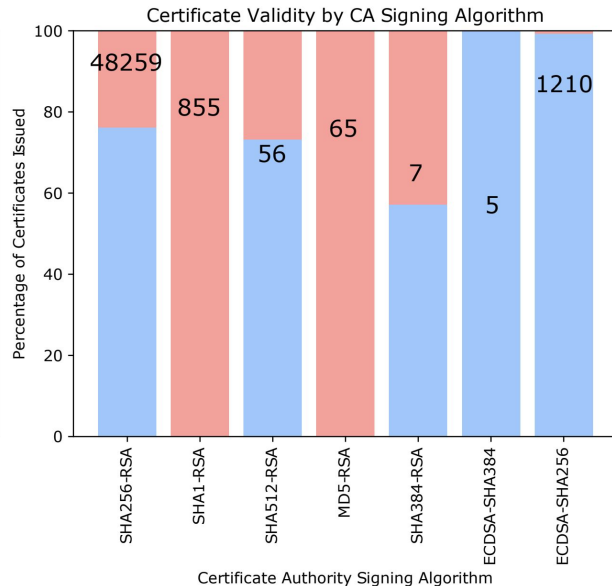
11%
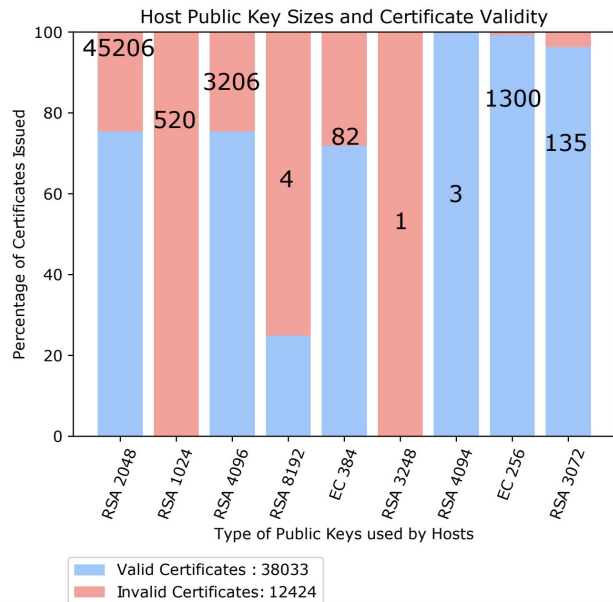
Websites result In an **invalid** https connection

9

Validity by Certificate Authorities

Distribution of Cert Issuers [Top 40]

- Free CAs like Let's Encrypt are the leading certificate providers
  - 80% validity
  - 20% invalidity
    - Hostname mismatch
    - Expiry
    - Self signed certs.

Note: The CAs issuing certs differ by country.

Certificate Validity & Common Errors

Certificate Reuse

- **Incorrect use** of wildcard certificates
  - `*.portal.gov.bd` applied on all `*.gov.bd`
- Use of **web server default** certificates
  - "`localhost`"
  - "`example.com`"
    - Used across **58 hostnames** across **24 countries.**
  - Probably from a popular question-answer website
    - Allows the ability to **intercept, decrypt and modify** `https` traffic.
    - **Indistinguishable** if users add certificate to **allowed browser exceptions**

In Depth Case Studies: USA and ROK

1. Both countries have **similar HDI scores** and **Internet adoption rates** but have a differing `https` adoption
   - USA : 81.12%
   - ROK : 37.95%
2. **Technical sophistication** of both countries biases them towards higher `https` adoption numbers compared to the rest of the world.
3. ROK recently moved out of its own NPKI infrastructure to use global standards, and USA mandates government websites to have `https`. [Congress S.2749 116-192]
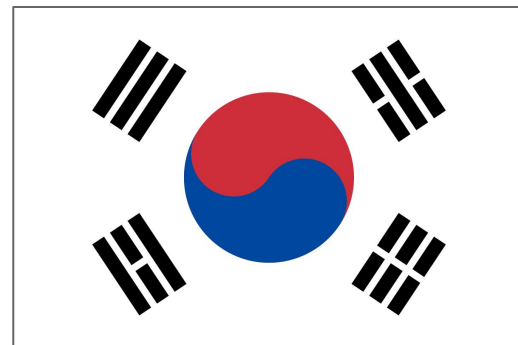
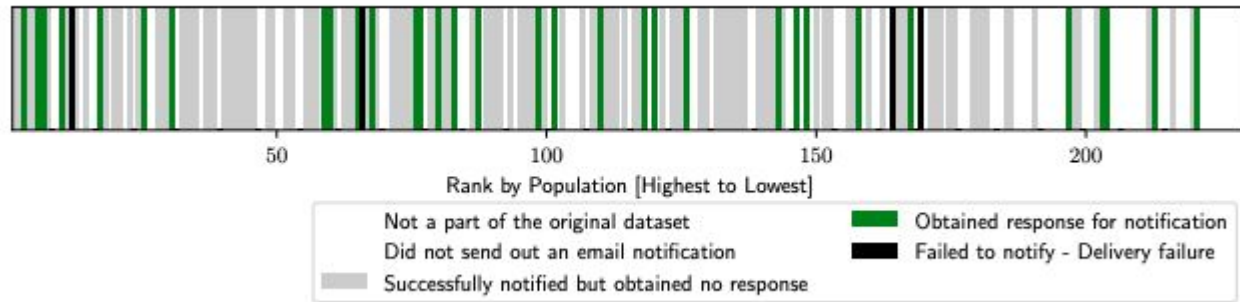**Takeaway**: `https` adoption in government websites is below expectations worldwide.

But Wait ... What about Non-Gov Websites?
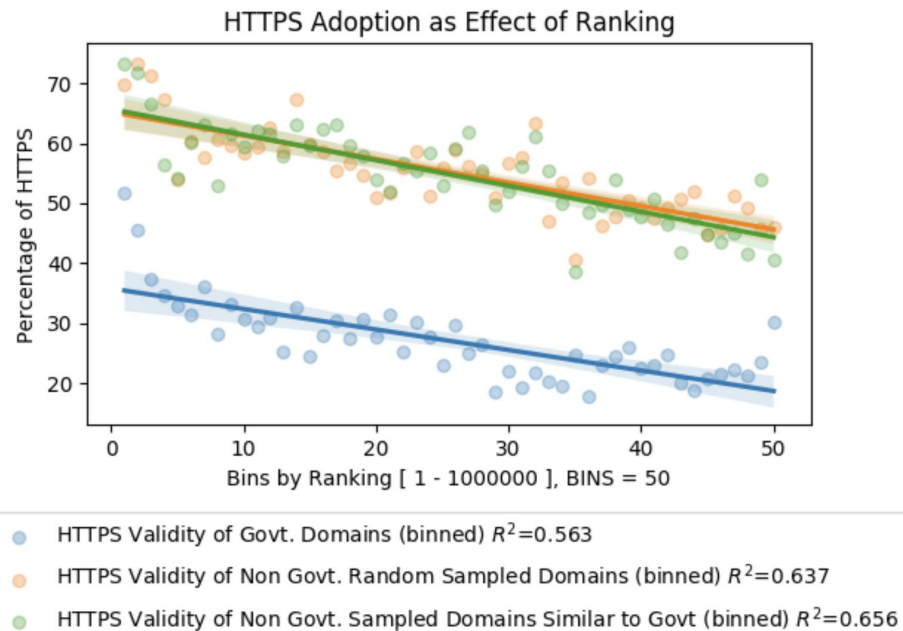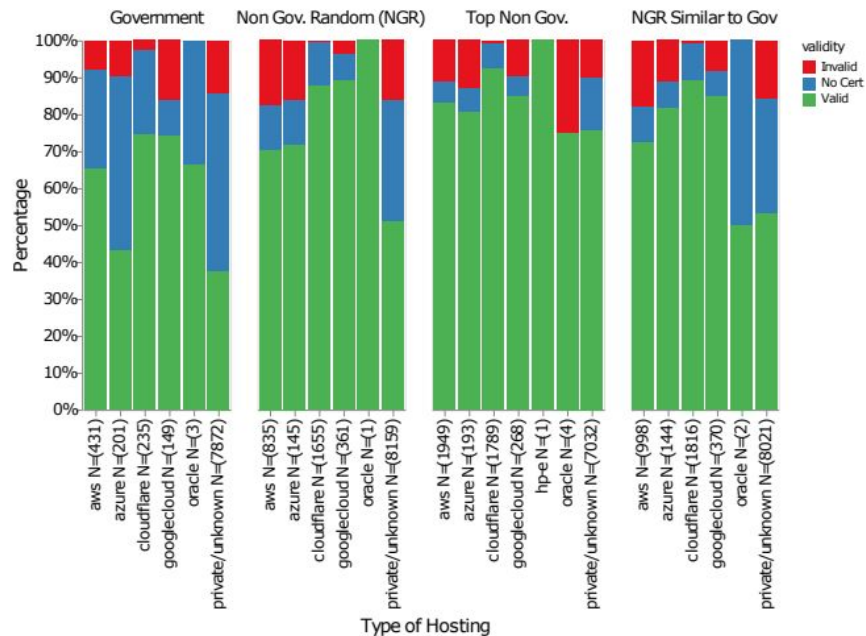
Takeaway: Higher public cloud services usage and higher `https` adoption and validity in Non-Gov Websites.

Responsible Disclosures and Notifications

- **Controlled issuance** of Government domains make it easier to reach the country government registrars
- **Higher response rate** (~22%) compared to direct notification studies in the past (~5.8%)
- 39 countries who **proactively engaged**.

Rank by Population [Highest to Lowest]

Not a part of the original dataset
Did not send out an email notification
Successfully notified but obtained no response
Obtained response for notification
Failed to notify - Delivery failure

Impact of Notifications

- Scanned the reported websites **2 months** later
  - Silently updated with no response
  - Unavailable websites back online
  - `http-only` traffic upgraded to `https:`
    - \> 10% improvement in 62 countries
    - \> 40% improvement in 7 countries.

We weakly attribute this to the disclosure and notifications.

Future Work

Making .gov More Secure by Default

When the public sees information on a .gov website, they need to trust that it is official and accurate. This trust is warranted, because registration of a .gov domain is limited to bona fide US-based government organizations. Governments should be easy to identify on the internet and users should be secure on .gov websites.

.gov

1. S.2749 - DOTGOV Online Trust in Government Act of 2019
2. Encourage the usage of DNSSEC signed CAA records and HSTS Preloading
3. Encourage domain registrars to implement safeguards from domain names which could impersonate government domains.
4. Improve `https` adoption.