NTNU

Kunnskap for en bedre verden

Where is the web still insecure? Regional scans for HTTPS certificates

Anushah Hossain, Kristina Nelson, **Tjerand Silde**

Department of Mathematical Sciences,
NTNU Trondheim

NTNU

Disclaimer

This is a **short paper**, based on an **ongoing project**, where we present our **preliminary results**, the **limitations**, and **future work**.

NTNU

Overview

- Introduction
- Background
- Methods
- Results
- Limitations
- Conclusion
- Future work

NTNU

Introduction

- We want to understand web security as it is experienced around the world
- We scan the top 500 most visited sites from nine countries of interest
- We document HTTPS usage, the encryption algorithms, and certificate information, including issuing date and length of validity
- We analyze the trends and security issues, and point to important future work

NTNU

Background I

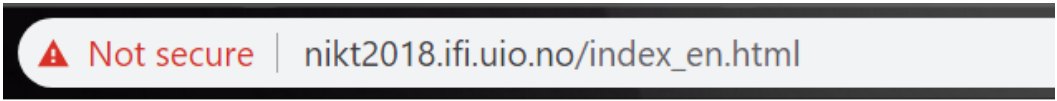
Scott Hanselman

@shanselman

Follow

HTTPS & SSL doesn't mean "trust this." It means "this is private." You may be having a private conversation with Satan.

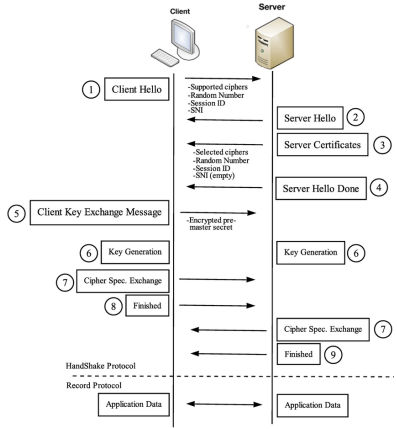
Background II



NIKT 2018 (SVALBARD 18-20.09.18)

DATES **CALLS** **REGISTRATION** **PROGRAM**

Background III



researchgate.net/publication/298065605

NTNU

Background IV

A series of actions taken since 2014 to incentivize HTTPS usage:

- August 2014:
  Google made HTTPS-status a ranking signal for internet searches.
- September 2016:
  Google and Mozilla announced that from January 2017, they will label HTTP pages with password or credit card form fields as “not secure".
- February 2018:
  Google announced that from July 2018, Chrome will mark all HTTP sites as “not secure” (Mozilla Firefox still show an information-button).

Background V

NIST [5] specifies a set of primitives and key sizes considered “secure":

- SHA-256, SHA-384, SHA-512 and SHA-3
- RSA is secure with at least 2048 bit keys
- Elliptic Curves are secure with at least 224 bit keys

We also note that:

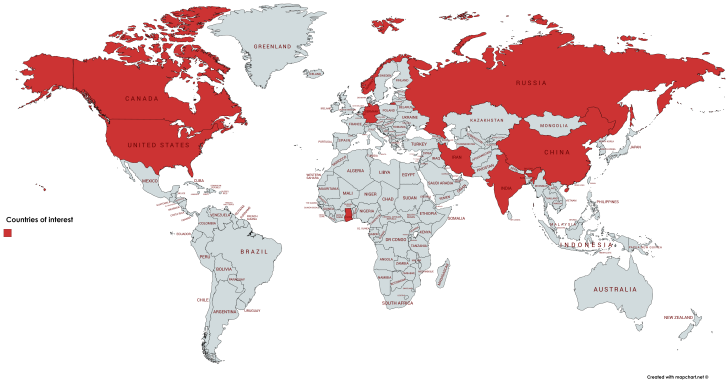
- SHA-1 was broken in 2005 [2] and should not be used
- RSA with 1024-bit keys is considered breakable by an adversary with sufficient computational power [4]

NTNU

Methods I

- We selected nine countries that range in geography, income level, and political regime.
- We scraped the top five hundred most visited sites for each country from the Alexa top sites service.
- We collected the site listing data from Alexa on March 26.
- We extracted certificate information on April 14.
- We used the OpenSSL python library [3] for extraction
- We recorded information about the certificate issue and expiration dates, signing algorithm, encryption algorithm and key sizes
- We also recorded HTTP Strict Transport Security (HSTS) usage

Methods II



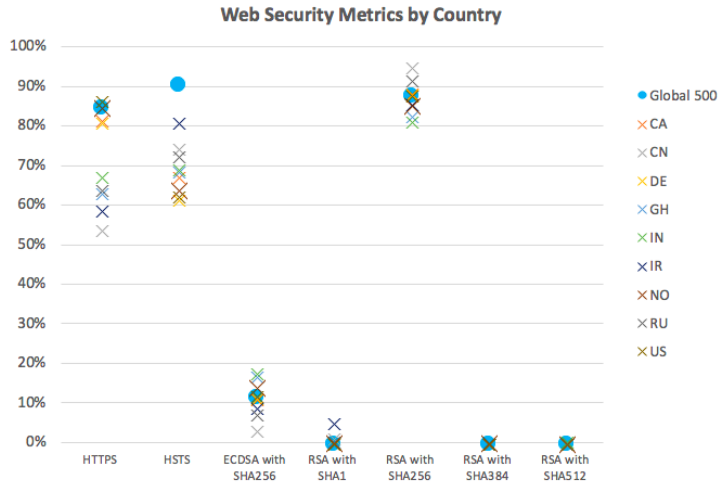
Canada, China, Germany, Ghana, India, Iran, Norway, Russia and USA

Results I

		Global 500	Canada	China	Germany	Ghana	India	Iran	Norway	Russia	USA
HTTPS		85%	82%	54%	81%	63%	67%	59%	85%	64%	87%
HSTS (% of HTTPS sites)		91%	67%	74%	62%	69%	69%	81%	64%	73%	63%
Signing Algorithm	ecdsa-with-SHA256	12%	11%	3%	12%	17%	18%	9%	14%	7%	12%
	sha1WithRSAEncryption	0%	0%	1%	0%	0%	0%	5%	0%	1%	0%
	sha256WithRSAEncryption	88%	89%	95%	88%	83%	81%	86%	85%	92%	88%
Encryption Algorithm, Key Size	EC256	17%	14%	10%	15%	20%	20%	11%	17%	10%	15%
	RSA2048	79%	84%	87%	78%	76%	77%	83%	75%	82%	83%
	RSA4096	3%	2%	2%	6%	4%	2%	4%	7%	7%	2%
Average Certificate Length (months)		19	20	23	20	18	18	36	21	20	19
Total Site Count		500	500	500	500	500	500	500	500	500	500

We include a Global 500 column of the top sites overall, as ranked by Alexa top sites, as a point of comparison for the country results.

Results II


Web Security Metrics by Country
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
HTTPS
HSTS
ECDSA with SHA256
RSA with SHA1
RSA with SHA256
RSA with SHA384
RSA with SHA512
Global 500
CA
CN
DE
GH
IN
IR
NO
RU
US

NTNU

Results III

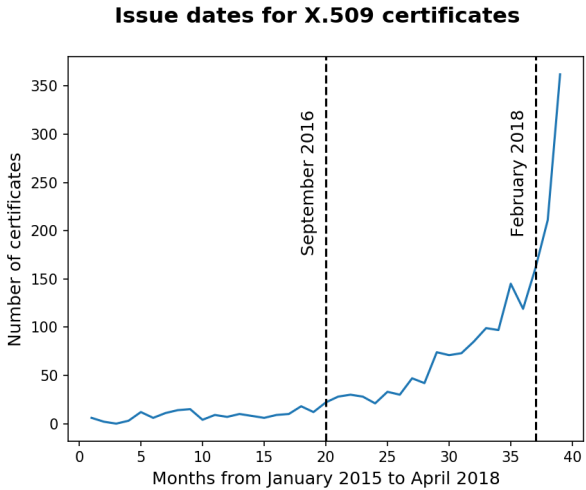
- The United States (87%), Norway (85%), Canada (82%) and Germany (81%) have the highest percentages of top sites using HTTPS
- China has the lowest fraction with only 54%, followed by Iran with 59%
- In Iran only 59% of top sites use HTTPS, but as much as 81% use HSTS

NTNU

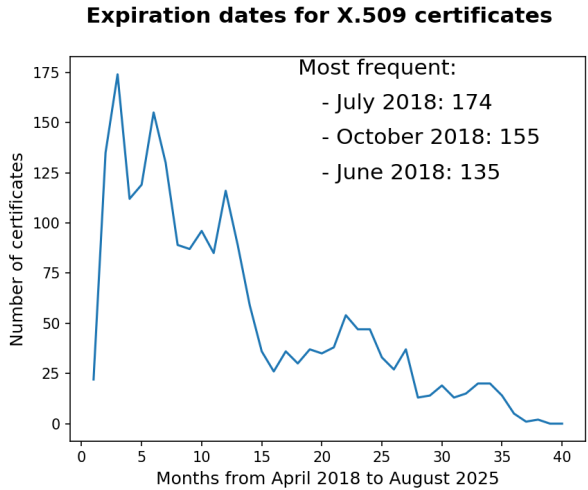
Results IV

- RSA with SHA256 is the most common signing algorithm used, followed by ECDSA with SHA256.
- Twenty-six unique websites still use SHA1 for signings and nine websites still use RSA with key size 1024 bits.
- Iran has the highest percentage of visited sites using the SHA1 hash function, possibly reflecting insecure local content.
- Several certificates issued in China and Iran were valid for 30 to 100 years, but these was marked as insecure in major browsers.

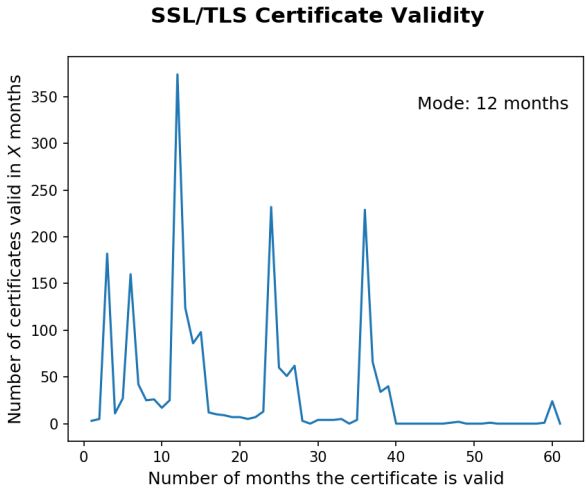
Results V


Issue dates for X.509 certificates
Number of certificates
Months from January 2015 to April 2018
September 2016
February 2018
0
50
100
150
200
250
300
350
0
5
10
15
20
25
30
35
40

Results VI


Expiration dates for X.509 certificates
Most frequent:
- July 2018: 174
- October 2018: 155
- June 2018: 135
Number of certificates
Months from April 2018 to August 2025
0
25
50
75
100
125
150
175
0
5
10
15
20
25
30
35
40

Results VII


SSL/TLS Certificate Validity
Mode: 12 months
Number of certificates valid in X months
Number of months the certificate is valid
0
50
100
150
200
250
300
350
0
10
20
30
40
50
60

NTNU

Limitations I

According to Scheitle et al. [1] it is unclear how representative Alexa listings are of the entire web, as they are based on data collected from opt-in browser extensions.

Limitations II

It is hard to determine usage of HSTS because of the variety of ways sites are able to deliver HSTS. Chrome, Firefox and other major browser vendors have begun shipping a hard-coded “preload” list of HSTS websites.

NTNU

Limitations III

We manually verified a subset of websites with valid certificates and websites with errors, and found both false positive and false negative results compared with the Python script.

NTNU

Limitations IV

If majority of HTTPS sites visited belonging to global companies and global sites tend to use HSTS, this would reflect in a higher relative HSTS rate.

NTNU

Conclusion I

Our results show significant regional variation and suggest that users from China, Ghana, Iran, and Russia are relatively more susceptible to eavesdropping or corrupted data when sending information over the internet.

NTNU

Conclusion II

These initial results suggest that web security is improving, but the benefits are not yet evenly distributed globally.

NTNU

Conclusion III

Knowledge of where the web is insecure, as experienced by a country's users, can help policy makers and other stakeholders place targeted pressure on the sites in question to implement HTTPS and HSTS, or recommend stronger encryption algorithms.

NTNU

Conclusion IV

The majority of action incentivizing web security has come from private sector actors, as we've seen in the success of browser policies and cost-decreasing initiatives such as Let's Encrypt

Conclusion V

We expect the percentage of websites using HTTPS to increase significantly in the coming months, given past responsiveness to browser policy

NTNU

Future work I

Rewrite from Python to Go to

- collect more information
- improve stability
- improve speed

Future work II

More detailed handshake

- public key algorithm* and key-size*
- symmetric key algorithm, mode and key-size
- integrity algorithms and hash-functions*

Future work III

Scan for

- Certificate Transparency
- Certificate Revocation

NTNU

Future work IV

Extend the project to

- cover an extended list of countries of interest
- connect websites to country of origin
- understand the relationship between site popularity and security
- compare with other top website rankings as
  - Similarweb
  - Quantcast
  - Majestic Million

Future work V

Long term scanning, to

- better understand the trends over time
- being able to detect change in security
- get more stable / accurate results

NTNU

Future work VI

Extend project to also include information about

- security headers (Referrer-Policy, X-XSS-Protection, etc.)
- updated connections using TLS 1.3
- Session Resumption and 0-RTT

NTNU

Future work VII

Compare with other works, as for example

- Google Transparency Report on `transparencyreport.google.com`
- Troy Hunt and Scott Helme on `whynohttps.com`
- Internet-Wide Scan Data Repository on `scan.io`
- Bank Grade Security on `bankgradesecurity.com`
- HTTPS-Norge by NRK Beta on `nrkbeta.no/https-norge`

Code, Documentation and Raw Data

security-scan

Tool for scanning websites and check their security.

Log

- 17.09.18: Improved documentation of code and created new issues for further improvment
- 17.09.18: Uploaded raw data from September 2018 scan
- 17.09.18: Repaired wrong tag of certificate dates in August 2018 scan
- 08.08.18: Uploaded raw data from August 2018 scan
- 08.08.18: Uploaded short paper, data from previous scans and relevant resources
- 08.08.18: Uploaded all code to scan and obtain raw data

Installation

- Download and install `Go` from golang.org
- Fork and download this repository
- Use command line to go to your local version of the repository
- Type `make run` to run new scan

Check out `https://github.com/tjesi/security-scan` for details

NTNU

References

- Scheitle et al. *Structure and Stability of Internet Top Lists*. arxiv.org/pdf/1802.02651.pdf. 2018.
- CWI and Google. *Shattered*. shattered.io. 2017.
- The pyOpenSSL developers. *pyOpenSSL documentation*. pyopenssl.org/en/stable. 2018.
- Matthew Green. *How does the NSA break SSL?*. blog.cryptographyengineering.com/2013/12/03/how-does-nsa-break-ssl/. 2013.
- NIST. *Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths*. dx.doi.org/10.6028/NIST.SP.800-131Ar1. 2015.