Measuring Adoption of Security Mechanisms in the HTTPS Ecosystem

Authors
Nicholas Hitch, MSc (Royal Holloway, 2022)
Simon Bell, ISG, Royal Holloway

Abstract

Are you responsible for maintaining the security of web applications? Do you want to provide an easy way for security researchers to disclose issues identified on your web applications? Do you have a general interest in HTTPS research? This article focuses on security mechanisms used by websites: Which of the SSL/TLS (encryption protocol that protects data between computers) versions 1.0-1.3 are supported by websites, HTTPS redirection (including HTTPS to HTTP redirection - yes this does happen!) and the use of the security.txt standard (used to inform security researchers how to report security issues they find on websites). The top 1 million websites, as ranked by Tranco (tranco-list.eu), were scanned daily for 16 months in order to capture the extent to which these security mechanisms were in use.[a]

[a]This article is based on an MSc dissertation written as part of the MSc in Information Security at the ISG, Royal Holloway, University of London. The full thesis is published on the ISG's website at `https://www.royalholloway.ac.uk/research-and-teaching/departments-and-schools/information-security/research/explore-our-research/isg-technical-reports/`.

1 Introduction

The world revolves around internet communication, and websites are a big part of that. There have been many security improvements to the technology that powers the HTTPS therefore analysing how these are or are not being utilised is of great interest and value to the information security community.

Security is a battleground with attackers ever changing their techniques and defenders doing their best to thwart the attackers attempts.

HTTPS is a critical security mechanism and was introduced as HTTP lacked the security features needed to protect the data as it travelled across the internet. HTTPS establishes an encrypted communication channel (provided by the SSL/TLS protocols) between the browser and the website such that attackers are unable to see what data is being sent back and forth. Additionally, HTTPS provides a way to detect if changes were made to the data as it travels across the internet.

It has become quite common practice for websites to automatically redirect a user to a HTTPS URL if the user enters or follows a HTTP URL e.g. the URL http://example.com redirecting to https://example.com. This helps to enable the use of the HTTP protocol over a secure channel without the user having to actively seek out the HTTPS URL of a website (which most users are unlikely to do). One of the main reasons is making it as easy as possible for a user to reach the website they are trying to visit.

Almost 2% of the top 1 million websites are still redirecting from HTTPS to HTTP.

The security.txt standard revolves around a text file for the primary purpose of providing security researchers the required information on how to disclose security issues they have identified on websites. A file named "security.txt" is placed in a known location e.g. https://example.com/.well-known/security.txt such that a security researcher can easily find, should the website support the security.txt standard.

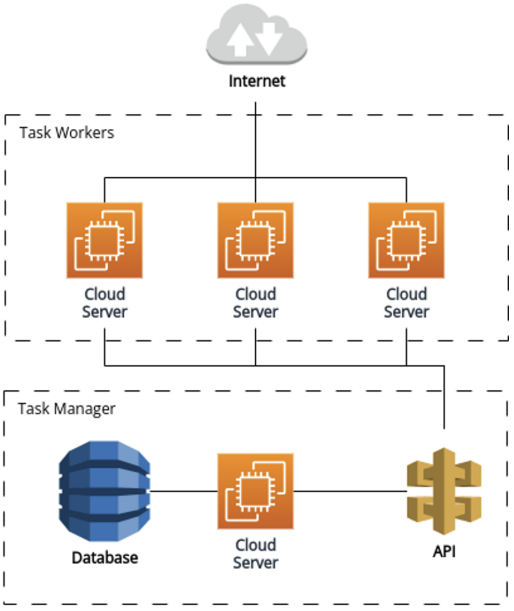

Figure 1: Scanning infrastructure

2 Scanning

The daily 1 million website scanning setup was split into two main sections: the actual scanning of the websites (task workers); and the management of which websites to scan including the storage of the scanning results (task manager).

Each day the task manager (a cloud server e.g. a AWS EC2) downloaded the list of the top 1 million websites from Tranco (tranco-list.eu) and a "scanning task" was created for each website.

The task workers (cloud servers e.g. a AWS EC2) would continually poll the task manager web application (API) for new scanning tasks. The results from a scanning task were sent to the task manager web application (API) and stored in a database. This is shown in Figure 1.

Once the day's scanning tasks were completed, all the results were downloaded via the task manager web application (API) and stored on a file storage server. All the task results were then deleted from the task manager database ready for the next day's scanning.

After 16 months of scanning once per day, the results on the file storage server were processed and analysed. A select few security mechanisms were analysed. A selection of those will now be presented in this article. Please see the full report for all the security mechanisms that were analysed.

3 HTTP(S) Redirection

HTTP(S) redirection is where a website redirects from one URL to another. In this research it was looking to see if a website redirected to a HTTPS URL if given a HTTP URL (e.g. http://www.example.com redirecting to https://www.exmaple.com) and the reverse; if a website redirected to a HTTP URL if given a HTTPS URL (e.g. https://www.example.com redirecting to http://www.exmaple.com).

If a HTTP (not HTTPS) response was received or no response was received, additional requests were made in an attempt to get a response. The details of this can be found in the full report.

The research by Buchanan et al [1] found that in August 2015 only 6.7% of sites when accessed over HTTP were redirected to HTTPS and this increased to 24.78% in May 2017. This research finds that

in November 2020 the number of sites redirecting from HTTP to HTTPS further increased to 56% and continued this trend to reach 64.5% in January 2022 as shown in table 1 and figure 2.

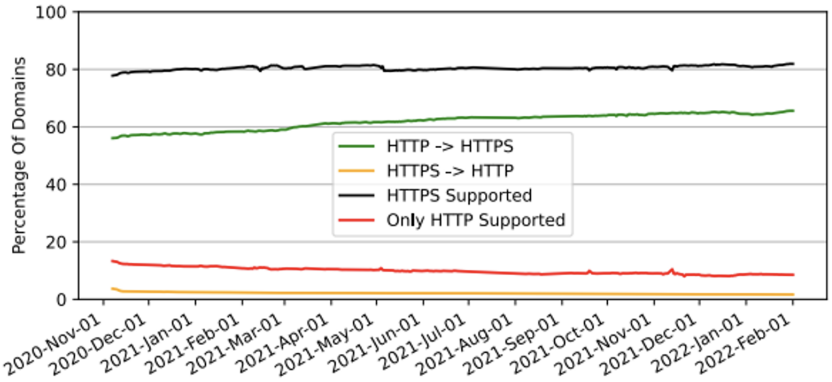

Figure 2: HTTP(S) Redirections for Top 1 Million websites

Table 1: HTTP(S) Redirection

Redirection	**Aug 2015** Buchanan et al [1]	**May 2017** Buchanan et al [1]	**Nov 2020**	**Jan 2022**
HTTP > HTTPS	6.7%	24.78%	56%	64.5%
HTTPS > HTTP	NA	NA	3.5%	1.7%

The rate of the upward trend of HTTP to HTTPS redirection is slowing. However, it is hoped that more sites continue to implement redirection such that future research does not show a plateau before nearly all sites in the top 1 Million redirect to HTTPS.

To some surprise there are a number of sites that redirect from HTTPS to HTTP, 3.5% in November 2020 and down to 1.7% in January 2022. It is assumed that this is done on purpose, and that as the trend is on the decline the reasons for implementing HTTPS to HTTP redirection are being overcome or are not relevant anymore.

There are a number of sites that only support HTTP and do not support HTTPS, around 16% in November 2020 and down to around 9% in January 2022 again trending in the right direction.

4 SSL/TLS Version Usage

Negotiated TLS Versions: When a website was being scanned, a request was made to each website allowing any of the TLS versions from 1.0 to 1.3 to be used and figure 3 shows which versions the scanner and website agreed upon using (which should be the highest version that the website supports). None of the old legacy SSL versions were supported by the client used to make the request to the website (i.e. SSLv3, SSLv2 . . . .).

As shown in figure 3, November 2020 appears to show that it was just before this time that TLS 1.3 started to become the dominant TLS version negotiated, replacing TLS 1.2 as the dominant version. This assumption is enhanced by the research of Holz et al [2] which showed that in November 2019 from the analysis of passively collected TLS connection details, around 79% of connections were using TLS 1.2 with a downward trend and around 19% of connections using TLS 1.3 with an upward trend.

Supported TLS Versions: When a website was being scanned, a TLS request to each website was made for each of the following versions: 1.0, 1.1, 1.2 and 1.3. This was done to determine which TLS versions are still being supported.

TLS 1.2 is supported by over 99% of websites. The reason for this is likely due to TLS 1.2 being the

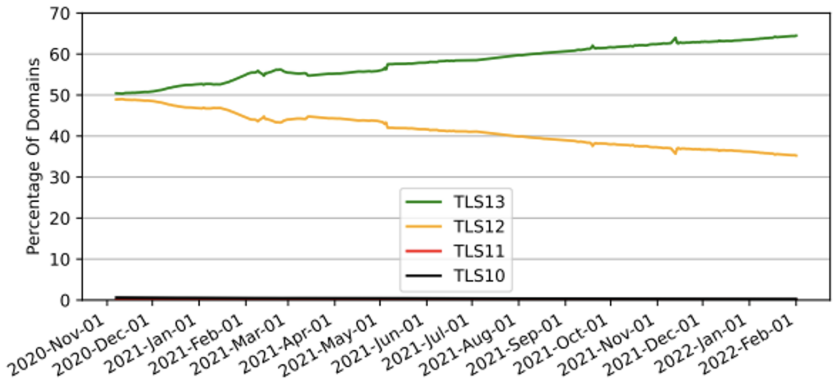

Figure 3: TLS version negotiated for Top 1 Million websites that support HTTPS

Table 2: HTTP(S) Redirection

TLS version	**Jul 2013** Kontzias et al [2]	**Aug 2014** Kontzias et al [2]	**Nov 2020** Holz et al [3]	**Jan 2022**
TLS 1.0	∼73%	∼48%	<1%	<1%
TLS 1.1	∼17%	<1%	<1%	<1%
TLS 1.2	∼1%	∼48%	∼79%	∼35.2%
TLS 1.3	NA	NA	∼19%	∼64.4%

most recent TLS versions supported by some legacy devices and is likely to remain like this for many years to come. As one might expect, TLS 1.3 is on an upward trend going from 50% in November 2020 and reaching 64% in January 2022.

TLS 1.0/1.1 are still widely supported and look to follow the same downward trend with 44% and 48% in January 2022 respectively. As they follow the same trend it is likely due to when software libraries and or applications are upgraded, rather than manual configuration changes to remove support for TLS 1.0 and TLS 1.1.

For the 0.2% of websites that negotiate with TLS 1.0 or TLS 1.1 they did not support any version higher than what they negotiated with. TLS 1.0 and TLS 1.1 should not be used to adhere to best security practices due to several factors including: unsafe encryption standards, vulnerable protocols and other vulnerabilities.

5 Security.txt

The “security.txt” mechanism is to aid in informing security researchers how to disclose security issues found on systems such as a domain hosting a HTTP(S) sever (website).

When a website was being scanned, a request was made for the security.txt file (e.g. https://example.com/.well-known/security.txt).

Figure 5 is split into the rank groupings of the websites scanned, as determined by Tranco (tranco-list.eu). The most highly ranked have the most adoption with security.txt usage dropping as the ranks increase. The trend is moving in the right direction with the highest value reaching over 18% usage by the top 100 ranked websites.

A possible reason for this could be that the highest ranked websites attract engineers who follow the latest developments in the HTTPS ecosystem as well as these organisations having the resources available to be able to implement them.

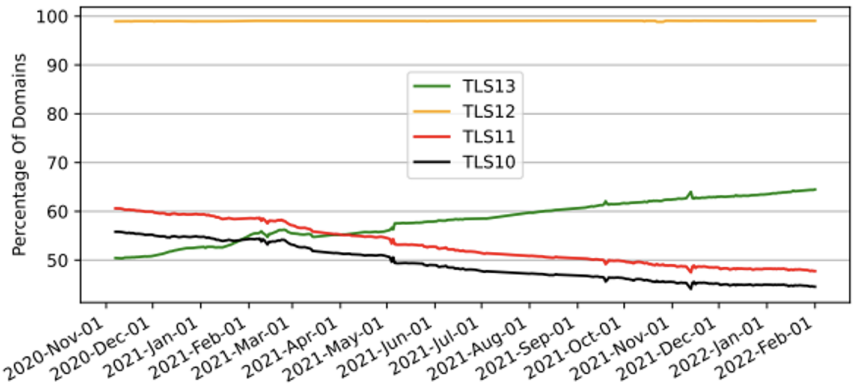

Figure 4: TLS Versions Supported

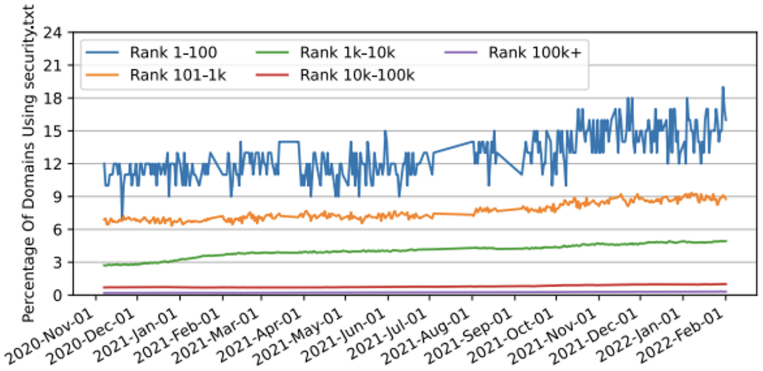

Figure 5: security.txt (only /.well-known path) by rank

6 Conclusion

Throughout our study the number of websites supporting HTTPS remained at a near constant level of around 80%, indicating that a plateau of websites supporting HTTPS has been reached.

With TLS 1.3 use climbing from 50% in November 2020 to 64% in January 2022 (as shown in figure 3), of websites scanned that support HTTPS, there is strong support for the latest TLS versions and this trend looks as though it will continue to rise providing more security for users browsing the internet.

The security.txt mechanism adoption is still in its infancy with adoption at 0.5% of websites scanned as of January 2022. The website's security team contact details are the primary piece of information that security researchers are looking for in order to report a security vulnerability.

The goal of this mechanism is admirable and its relevance looks quite promising as the majority of security.txt files found have a contact field defined. There are a number of areas of future work identified in the research which can be found at the end of the full report.

References:

[1] W. J. Buchanan, S. Helme, and A. Woodward, "Analysis of the adoption of security headers in HTTP," IET Inf. Secur., vol. 12, no. 2, pp. 118–126, Mar. 2018.

[2] 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 Proceed- ings of the Internet Measurement Conference 2018, ser. IMC '18. New York, NY, USA: Association for Computing Machinery, pp. 415–428, Oct. 2018.

[3] 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 experimen- tation and centralization,” Comput. Commun. Rev., vol. 50, no. 3, pp. 3–15, Jul. 2020

Biographies

Nicholas Hitch is a site reliability engineer and an information security enthusiast with a focused interest in the HTTPS ecosystem. Nicholas has recently completed an MSc in Information Security at Royal Holloway University of London.

Dr Simon Bell is a module lead, tutor, and project supervisor on the Information Security Group’s distance learning Information Security MSc programme. His research interests include data-driven approaches to measure cyber attacks on the internet, phishing detection, blocklist effectiveness, honeypots, and malware analysis.

Series editor: Dr Maryam Mehrnezhad, ISG, Royal Holloway