An Investigation of Cyber Autonomy on Government Websites

Hsu-Chun Hsiao† Tiffany Hyun-Jin Kim‡ Yu-Ming Ku† Chun-Ming Chang†
Hung-Fang Chen† Yu-Jen Chen† Chun-Wen Wang∗ Wei Jeng†

†National Taiwan University ‡HRL Laboratories ∗Cornell University

ABSTRACT

From a national security viewpoint, a higher degree of *cyber autonomy* is crucial to reduce the reliance on external, oftentimes untrustworthy entities, in order to achieve better resilience against adversaries. To probe into the concept of government cyber autonomy, this study examines the *external dependency of public-facing government websites* across the world's major industrialized, Group of Seven (G7) countries. Over a two-year period, we measured HTTPS adoption rates, the autonomy status of CAs, and the autonomy status of CPs on G7 government websites. We find that approximately 85% of web resources loaded by G7 government sites originate from the United States. By reviewing policy documents and surveying technicians who maintain government websites, we identify four significant forces that can influence the degree of a government's autonomy, including government mandates on HTTPS adoption, website development outsourcing, the citizens' fear of large-scale surveillance, and user confusion. Because a government websites are considered critical information infrastructures, we expect this study to raise awareness of their complex dependency, thereby reducing the risk of blindly trusting external entities when using critical government services.

CCS CONCEPTS

• **Security and privacy → Web application security**.

KEYWORDS

Cyber autonomy; government website security; HTTPS adoption

1 INTRODUCTION

A central guideline to cybersecurity is to reduce reliance on *external entities*, especially for those systems under different jurisdictions or governance. From a national security standpoint, it is critical to reduce the reliance on external entities in order to achieve *cyber autonomy*[1]. Recently, cyber autonomy has gained increased attention by worldwide governments due to the revelation of state-sponsored hacktivism [8, 13], including hardware/software shipped with spyware [23, 29], compelled certificates issued for Internet interception [53], and zero-day exploits developed by rival government agencies [26].

To understand how cyber autonomy is practiced by governments worldwide, we investigated a critical facet of a government's cyber autonomy: *external dependency of public-facing government websites*. Since public-facing government websites are the front line for delivering official information and accessing government services, attacks targeting these websites can not only cause widespread panic and social instabilities, e.g., spreading inaccurate information or falsifying alerts [11, 16, 25], but also steal citizens' credentials and sensitive information [17, 19, 22, 24, 30].

In our investigation, we examined the top 100[2] government websites for each of the world's major industrialized countries, known as Group of Seven (G7), in 2017 and 2018. We considered three crucial elements for the cyber autonomy of governments: HTTPS adoption status and dependency on two external entities, namely Certificate Authorities (CAs) and Content Providers (CPs).

HTTPS adoption on government websites can protect sensitive information of their citizens. Unfortunately, our findings reveal the uneven HTTPS adoption rates across G7 countries. In 2017, the US (97%) and the UK (81%) had the highest HTTPS adoption rates, followed by Germany (57%), Canada (53%), France (46%), Japan (40%), and Italy (25%). There has been a steady increase in the adoption rate (13.7% on average) in 2018; a notable increase is by Germany, whose adoption rate has jumped to 90% in 2018. However, among these G7 government websites supporting HTTPS, 91% (in 2017) and 77% (in 2018) are incorrectly configured and thus still vulnerable to attacks. The adoption rates of HSTS, a countermeasure to HTTPS downgrade attacks, remain low (below 50%) in most G7 counties (except the US, reaching 83% in 2018).

CAs are the roots of trust on the Web. With hundreds of eligible CAs to issue certificates, an identified challenge is preventing rogue CAs from spoofing certificates or issuing certificates for phishing domains. For example, CAs can be forced by local governments to issue compelled certificates for HTTPS interception [41]. Among the G7 government websites with valid certificates, approximately 10% are using the Extended Validation (EV) certificates, which are designed to mitigate phishing. France, Japan, and the US operate their own government root CAs (i.e., root CAs run by a government), but their root certificates are untrusted by Mozilla and Apple, causing multiple certificate errors.

CPs are servers that host web resources (e.g., images, JavaScript) for websites. When a website loads resources from untrusted servers, a variety of undesirable consequences can occur, such as including malicious content, executing malicious scripts that secretly steal credentials, mining cryptocurrencies, or sending denial-of-service traffic [44, 47]. Among the 660 websites that we crawled in 2018, 287 accessed resources from oversea countries, and each website made an average of 1.89 requests to load resources hosted on foreign servers. Among 2,160 external resources loaded by the G7 government websites, 912 (42%) were from 126 location-dependent URLs, which were served by different destination servers based on

[1] We use cyber autonomy to refer to "self-controlling and free from external influence" [1] in achieving security protection.

[2] Websites are ranked based on Alexa's statistics. See §2.1.1 for details.

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313645

the source locations. Of the 1,248 location-independent resources, 1,063 (85%) of them originated from the US.

To understand the forces that influence the dependency level of government websites, we also reviewed policy documents and conducted a survey on technicians who maintain government websites. We highlight our findings below:

1. Government mandates help increase autonomy. The HTTPS adoption rates in the countries that have announced mandates over a year are significantly higher than the rest of the G7 countries.

2. Outsourcing undermines cyber autonomy. In our survey, 53 of 63 government websites are created and maintained by third-party contractors. While contractors are legally bound to provide services, they may be unable to respond immediately to attacks.

3. Fear of large-scale government surveillance hinders autonomy. Pinning government websites to government root certificates could eliminate the impact of rogue CAs spoofing government website certificates. However, because some countries have exploited government-issued certificates for Internet surveillance, the trustworthiness of government root CAs has been questioned, and few government root certificates are currently included in all major browsers' trust stores [14].

4. User confusion. Partial inclusion of government root CAs in the browsers' trust stores results in certificate warnings, confusing users whether the government websites in question are trustworthy. The lack of distinct government domain indication may result in users falling for attacks that could have been easily identifiable with government-specific domain names.

Contributions. This paper attempts to measure the current practice of building cyber autonomy on government websites. Over a two-year period, we measured HTTPS adoption rates and the autonomy status of CAs and CPs on G7 government websites. We also conducted a survey on technicians who maintain government websites. Our findings indicate the importance of having government mandates to enhance the overall security of government websites, as well as synchronizing autonomy-oriented government policies with corresponding entities (CAs and CPs) such that they can provide consistent, trustworthy services without errors.

2 HTTPS ADOPTION

HTTPS [3] runs HTTP over Transport Layer Security (TLS), which is a fundamental security protocol that enables end-to-end encryption and authentication for HTTP connections. Without HTTPS, traffic can be intercepted by any en-route adversary, as well as off-path adversaries who are capable of hijacking routes (e.g., malicious ISPs). TLS attacks can be categorized into four types based on the following weaknesses [37].

Weaknesses in cryptography: Websites supporting outdated cryptographic algorithms (e.g., RC4 and MD5) or weak key lengths (e.g., RSA_EXPORT and DHE_EXPORT) enables attackers to crack TLS.

Weaknesses in protocol design: attackers can replay, interleave, or manipulate protocol messages to deceive the client or server into an unintended state, such as deliberately downgrading the protocol version or a cipher suite.

Weaknesses in implementation: attackers can exploit vulnerabilities in the TLS software libraries (e.g., OpenSSL).

Oracle attacks: attackers can adaptively interact with a victim running TLS and derive secret information based on responses.

Table 1: The weakness of eleven TLS attacks.

Type	Attack	Weakness
Weakness in cryptography	FREAK Attack [34]	Support RSA_EXPORT
	Logjam Attack [31]	Support DHE_EXPORT
	RC4 Attack [32, 56]	Support RC4 stream cipher
	Sweet32 Attack [35]	Support block ciphers with 64-bit block size and CBC mode
Oracle attacks	BEAST Attack [38]	Support TLS versions earlier than TLS 1.0 and CBC mode of block cipher
	CRIME Attack [39]	Support SSL/TLS compression
	DROWN Attack [33]	Use the same RSA keys in both SSL2.0 and newer versions of SSL/TLS
	POODLE Attack [46]	Support SSL3.0 without using TLS_FALLBACK_SCSV
Weakness in implementation	Heartbleed Attack [6, 40]	Support OpenSSL versions 1.0.1-1.0.1f
Weakness in protocol	ECDH-DH Cross-Protocol Attack [45]	Support ECDH and secp384r1 (ECC parameter) in server side and DH in client side
	Renegotiation Attack [51]	Support insecure renegotiation mechanism

Table 1 summarizes 11 critical TLS attacks examined in our study. Most of the attacks can be prevented by changing TLS configuration or disabling outdated protocol versions. The last two attacks would require updating to the latest protocol version, TLS 1.3.

2.1 Methodology

In this section, we describe our approach of collecting the G7 governments' domain datasets and measuring their security.

2.1.1 URL Collection. Because there is no single source maintaining a list of up-to-date government websites, we combined multiple publicly-available datasets and systematically curated them. For each of the G7 countries, we first crawled the list of domains under the Regional/Continent Name/Country Name/Government category on Alexa [7].

Because the scope and number of websites vary by country,[3] we improved the data quality of the combination of Alexa and public datasets by heuristically restricting domain names to the corresponding country code's top-level domains (cc-TLD) or second-level domain (e.g., gov, go), assuming that the majority of the collected government domains in that country adhere to a naming convention. As we noticed that the Alexa datasets are often incomplete and mixed with non-government sites (such as politicians' websites), we also searched for publicly-available datasets online. Another rationale to do so is to avoid the risk of using single evidence to generate our result [48]. We thus found official lists of government websites for the US, the UK, Canada, and Japan, but not the rest of G7. The list and final dataset we collect are available and can be accessed at the Open Science Framework project page [2].

2.1.2 Website Scanning. For each of the G7's government domains, we formed the request URL by adding four common prefixes: `http://`, `http://www`, `https://`, and `https://www`. Measurements were performed in September 2017 and August 2018. We further examined the resilience to the 11 TLS attacks (Table 1) if the domain supports HTTPS (i.e., at least one of the four requests gets redirected to an HTTPS page). Websites with certificate errors are not classified as HTTPS-supporting.

2.2 Findings

HTTPS adoption rate. Figure 1 shows the HTTPS and HSTS adoption rates of G7 government domains in 2017 and 2018. The colored bars stand for the HTTPS adoption rates, where the gray

[3]The US dataset contains 1,078 federal domains, while the dataset for Japan has only 77 for "the Cabinet, Ministries and Agencies, the Diet, the Supreme Court, and other government agencies". The UK dataset contains 3,631 central and local domains; the Germany data contains 620 federal and 8,468 non-federal domains.

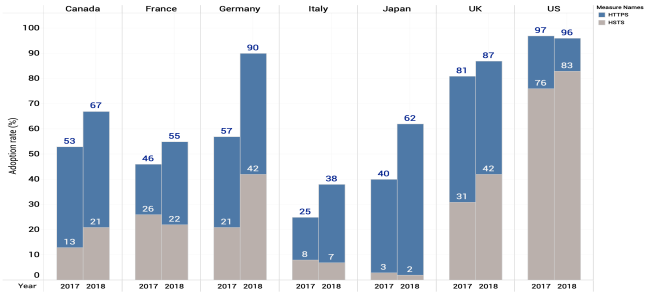

Figure 1: HTTPS and HSTS adoption rates of G7 top 100 government websites in 2017 and 2018

shade within each bar represents its HSTS adoption rates. The countries from the highest to the lowest HTTPS adoption rate in 2017 are as follows: the US (97%), the UK (81%), Germany (57%), Canada (53%), France (46%), Japan (40%) and Italy (25%). Both the US and UK remain equivalently high in 2018, while Germany increases from 53% to 90% within a year. The other countries show an average increase of 14.5% from 2017 to 2018, conforming to the global trend of HTTPS adoption.

HSTS adoption. HTTP Strict Transport Security (HSTS) [4] is a web security policy mechanism that allows a server to declare that it accepts only HTTPS (not HTTP) connections from browsers. HSTS is designed to prevent a man-in-the-middle (MitM) attack, which downgrades HTTPS to plain HTTP, or a *SSL stripping attack*. An HTTP request to an HSTS-enabled site will be automatically upgraded to HTTPS by supported browsers. As shown in Figure 1, the HSTS adoption rates of government websites in Japan (3%) and Italy (8%) are extremely low in 2017, and even decrease by 1% in 2018. France also has a 4% decrease in 2018. While HSTS adoption rates for the US, the UK, Germany and Canada have increased from 2017 to 2018, all except the US remain below 50%.

Resilience to critical TLS attacks. Among these G7 government websites supporting HTTPS, 91% (in 2017) and 77% (in 2018) are incorrectly configured and thus still vulnerable to attacks. Given the 11 TLS attacks (Table 1) and the SSL stripping attack, we quantified a website's resilience level by the number of attacks against which the website can defend. Figure 2 shows the cumulative distribution of government websites resilience levels in each G7 country. More than 80% of US government websites can defend against at least 11 types of attacks. Germany makes significant progress in 2018: more than two-thirds of government websites can defend against more than 10 types, whereas it had 47% in 2017.

Autonomy implications. Increase in HTTPS and HSTS adoption rates on G7 government websites are promising, and as we discuss in §5.1, one catalyst to increase adoption rates is a government mandate. On the other hand, incorrect configurations may end up providing a false sense of security to users while leaving the door open for attackers. Hence, cross-validating implementations and keeping them up to date are crucial for secure cyber autonomy.

3 DEPENDENCY ON CA

Because HTTPS relies on digital certificates for entity authentication, trust in HTTPS is bootstrapped from certificate authorities, which issue certificates that bind a site's public key to the domain name after validating its identity. CAs are structured hierarchically for scalability: *intermediate* CAs can certify other CAs' identities,

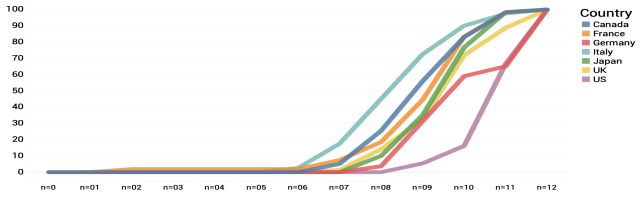

Figure 2: Resilience to 11 TLS attacks and SSL stripping in G7 in 2018

and users can determine which *root* CAs are trusted. Most users trust the root certificates in the default trust store distributed with the underlying operating systems (OS) or browsers.

Unfortunately, with hundreds of CAs eligible to issue certificates, rogue CAs can easily spoof certificates for phishing or HTTPS interception. In addition, issues in certificates and certificate chains undermine the trustworthiness of a website.

3.1 Methodology

Using the G7 governments' top 100 website lists as described in §2.1.1, we analyzed their certificates. For each website, we examined the *country* field of its root certificate to assess its dependency on foreign CAs , whether an Extended Validation (EV) certificate is used, and whether the certificate and its chain are valid.

We consider a certificate to be invalid if any of the following errors is detected: (1) a *name mismatch* error, which occurs when the hostname is inconsistent with the leaf certificate's common name; (2) an *untrusted root CA* error, which occurs when the root certificate's issuer is not in the trust store, potentially resulting in untrustworthy self-signed certificates or certificates signed by untrusted entities; (3) an *invalid time* error, which occurs when the current time is before or after the valid certificate lifetime; or (4) a *certificate issuer error*, which occurs when a leaf certificate's issuer is inconsistent with its parent's subject field.

3.2 Findings

Government root CAs. We analyzed the root certificates included in the trust stores of the major browsers and OS. 168, 359, and 149 trusted root certificates are pre-installed in the latest trust store of Apple macOS 10.13 (High Sierra) [9], Microsoft Windows [20], and Mozilla [21], respectively. Microsoft Windows's trust store contains 53 government root certificates[4] from 26 countries, including France, Japan, and the US. The Mozilla's trust store contains eight government root certificates from Hong Kong, the Netherlands, Spain, Taiwan, and Turkey. Apple macOS's trust store contains only two government root certificates, from Taiwan and Finland. As

[4]We consider that a certificate is issued by a government if it contains "government", "gov", or "federal" in the description.

Table 2: Certificate analysis and Domestic CA of top 100 government sites in G7, 2018

Country	w/ HTTPS protocol	Valid Cert	Domestic CA	EV Cert
Canada	70	67	0 (0%)	7 (10.4%)
France	65	55	21 (38.1%)	6 (10.9%)
Germany	92	90	44 (48.8%)	6 (6.6%)
Italy	44	38	9 (23.7%)	4 (10.5%)
Japan	65	62	27 (43.5%)	7 (11.3%)
UK	91	87	9 (10.3%)	10 (11.5%)
US	96	96	89 (92.7%)	10 (10.4%)
Total/Average	523/74.7	495/70.7	200/28.6	51/7.3

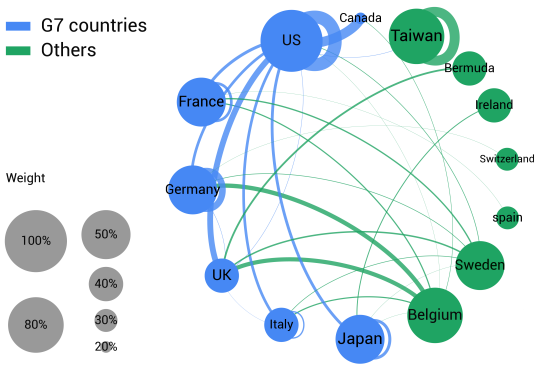

Figure 3: Dependency on root CAs by countries.

some government root certificates are not included in all trust stores, using these certificates will trigger warnings on some browsers. In the 2017 dataset, 11 Japanese government websites relied on ApplicationCA2 Root, a Japanese government root CA that is untrusted by Mozilla and Apple macOS.

Certificate type. Table 2 summarizes the number of websites with valid certificates, domestic root CAs, and EV certificates in 2018. The percentages in the last two columns are computed over websites with valid certificates. Among G7 government websites with valid certificates, about 10% are using Extended Validation (EV) certificates, which are designed to mitigate phishing. Interestingly, despite its great share of the CA market, Let's Encrypt (LE)[5] certificates are rarely used by government websites. As of November 2018, LE issues more than 60% of certificates on the Internet [10]. In our 2018 dataset, there are 8 from Germany and 2 from Italy, and each of the remaining G7 countries has only one government website using LE certificates.

Root CA dependency. Governments can reduce the root of trust and increase legal accountability by restricting CAs to those registered in the same country (e.g., via DNS-based Authentication of Named Entities). Hence, we examined the reliance of the government websites on domestic CAs. A root certificate's country is determined by the country field in its certificate information.

Figure 3 uses a circular layout for visualizing the relationship between the G7 countries' government websites and their root CAs' countries of origin as of August 2018. For instance, the clockwise curve from Country A to Country B represents higher dependency of A's government websites on B's root CAs. A larger circle area of a country indicates that more websites rely on the CAs in the country. Among all G7 government sites we analyzed, the most popular root CAs and their registered countries are GlobalSign Root CA in Belgium (accounting for 16% of 496 G7 sites with valid certificates), AddTrust External CA Root in Sweden (6%), DigiCert Global Root CA in the US (34%), and Entrust Root Certification Authority in the US (11%). Among G7, all but the US have more than 50% of government websites using root CAs registered in foreign countries in 2018. Ninety out of 96 US government sites adopt US-based root CAs, including one using Let's Encrypt and three using Entrust.net CA, the third most popular CA in G7.

The validity of certificates and certificate chains. Table 3 summarizes common misconfigurations. In 2017, 45% of France's Top 100 government sites suffered from name mismatch errors. Several certificates used `www.snakeoil.dom` or `ssl3.ovh.net` as their common name, which are used as examples in online tutorials. Japan has the highest percentage of untrusted root certificates: 11 (out of 12) use the same root certificate, ApplicationCA2 Root, a Japanese government root CA not trusted by all major browsers.

Autonomy implications. While governments are moving towards CA autonomy by restricting the usage of external CAs and promoting government root CAs, we observed inconsistencies across the trust stores of the browsers and OS. Although different browser and OS vendors have different policies, such inconsistencies imply some concerns about including the government root CAs to the trust store. Hence, further analysis must be conducted to resolve the conflicts, which may require cooperation among the browser and OS vendors. As we discuss in §5.3, skepticism about government surveillance activities on its citizens can be addressed by restricting the government root CAs' usage to government websites only, or by delegating the management of government root CAs to non-profit organizations within the same jurisdiction.

4 DEPENDENCY ON CP

Content Providers (CP) provide web resources (e.g., images and JavaScript) that can be included on websites. However, when a site loads resources from untrusted servers, it may include malicious content or even execute malicious scripts [44, 47].

4.1 Methodology

We loaded each website from the top 100 government website dataset for each country using Google Chromium Version 68.0.3440.106 for 20 seconds, and extracted the request URLs logged by the browser's network panel. As some content may be served by nearby servers, we loaded these websites from servers located in Asia and North America in October 2018 to capture such location-dependent behaviors. We fetched the IP from the remote address field in each request header, and used pygeoip GeoIP API along with the Maxmind geolocation database[6] to identify its country of residence.

Since we are interested in external CPs that may provide untrusted resources, we removed URLs with the same domain name (and thus are considered local), and focused on URLs that retrieve the most common external resources such as image, JavaScript, and JSON files [43] from foreign servers, as these resources can be used to exploit the website [44, 47]. Our `CP_ASIA` and `CP_NA` datasets contain 2,160 and 2,185 URLs for crawling from Asia and North America, respectively. Among these total 4,345 URLs, 1,960 (45.2%), 2,151 (49.5%), and 231 (5.3%) are for JavaScript, image, and JSON files, respectively.

[5] Let's Encrypt is a non-profit certificate authority that issues free certificates.

[6] The Maxmind database has 96-98% accuracy in the country level [49] and has been used to validate the IP geolocation information released by Regional Internet Registries(RIR) [57].

Table 3: G7 certificate scanning result for the indicated issues. A lower percentage is better in terms of security.

Country	Domains w/ cert.		Name mismatch		Issuer error		Untrusted root		Invalid time	
	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018
Canada	53	70	2%	4%	0%	0%	4%	0%	0%	0%
France	46	65	45%	15%	9%	0%	13%	6%	6%	3%
Germany	57	92	10%	2%	0%	0%	0%	0%	0%	0%
Italy	25	44	25%	13%	0%	0%	15%	0%	12%	4%
Japan	40	65	0%	4%	0%	0%	28%	28%	1%	1%
UK	81	91	2%	3%	0%	0%	1%	0%	0%	0%
US	97	96	0%	0%	0%	0%	0%	0%	0%	0%

Table 4: # websites loading resources from oversea servers.

country	ca	fr	de	it	jp	UK	US	total/Average
CP_ASIA	48	51	26	46	30	55	31	287/41
CP_NA	52	49	28	55	29	61	12	287/40.9

We say that a request URL is *location-aware* if it is routed to different destination servers when it is loaded from different source locations. A location-aware URL could have a negative impact on cyber autonomy because the user cannot be assured that the resource is always loaded from domestic servers. We identified 126 location-aware URLs that appear in both the CP_ASIA and CP_NA datasets, and in total 912 (of out 2,160; 42%) external resources were loaded from these location-aware URLs. Excluding the location-aware ones, the number of external JavaScript, image, and JSON files are 1,292 (51%), 1182 (46%), and 74 (3%), respectively.

4.2 Findings

Websites loading resources from foreign servers. Among the 660 websites that were successfully crawled, each website made an average of 1.89 requests to load resources hosted on foreign servers, with a median of 1. Table 4 summarizes the percentage of websites, among the top 100 government websites in each G7 country, that load resources from foreign servers. All G7 countries have a non-negligible degree of dependency, from 8% (US) to 23% (Japan, Germany) and more than 40% (the rest of G7), when users access government websites from Asia. The percentage reduces drastically for US government websites when accessed from North America. However, most G7 countries tend to have a high degree of dependency regardless of access location.

We found several government websites hosted in foreign countries. For example, the National Film Board of Canada (www.nfb.ca) is hosted in the US and loads resources from US-based servers.

CP dependency by country. The three most common destinations of external resources are the US (1,173), the Philippines (269), and Taiwan (225) in the CP_ASIA dataset, with a total of 2,160 external resources, and the US (1,915), Ireland (120), and the Netherlands (72) in the CP_NA dataset. These results align with common beliefs that the US hosts a large portion of web resources and that Content Distribution Network (CDN) servers are prevalent. Figure 4 shows the CP dependency by country, utilizing the similar circular layout as Figure 3. The orange lines represent the location-aware resources, whose IP addresses were different in the two datasets collected from different locations. Of the 1,248 location-independent resources, 1,063 (85%) of them originate from the United States.

Autonomy implications. In general, we observed greater reliance on US content providers for all G7 governments. Regarding location-aware URLs, a majority of the government websites may be loading content from foreign CPs, possibly to tolerate the latency. However, proving better usability with short delays have negative effects towards having cyber autonomy. Further investigation on the trade-offs between latency and autonomy may be needed to analyze the benefits of risking security for enhanced time delays.

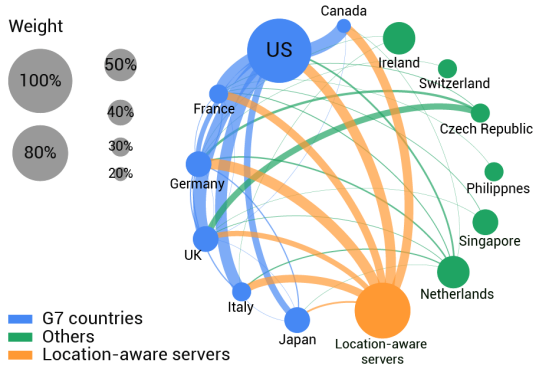

Figure 4: Dependency on external resources by countries.

5 DISCUSSION

5.1 Government Policies

Government policies play an important role in cyber autonomy. We found that at least four of the G7 countries have mandates that promote HTTPS adoption on government websites, and some have policies limiting the use of weak cryptographic algorithms and parameters for HTTPS and certificates. However, we did not find policies that restrict access to external resources. Considering the issuance or enforcement date of their security policies, we found that countries with established HTTPS adoption policies indeed exhibit a significantly higher level of security than others in terms of HTTPS deployment. The adoption rates of the US and UK were the highest in 2017, and Germany has improved in 2018. We briefly describe the findings on G7's HTTPS policies below.

The US. On June 8, 2015, the White House Office of Management and Budget memorandum M-15-13 was released to "Require Secure Connections across Federal Websites and Web Services" [5]. The memorandum provides a timeframe for compliance, requesting that all federal websites and services upgrade to HTTPS-only (with HSTS) by the end of 2016. Note that this memorandum applies to federal sites only, not to the state or local sites. On October 16, 2017, the Department of Homeland Security issued Binding Operational Directive 18-01, requiring federal agencies to remove support for outdated cryptographic algorithms and protocols. There is no restriction on acceptable CAs for federal domains.

The UK. The Digital Service Standard and Service Manual in the UK requires all Government Digital Service (GDS) websites to be accessible only through HTTPS starting October 1, 2016 and provides guidelines to configure HTTPS and HSTS [18].

Germany. In January 2018, the Federal Office for Information Security (BSI) issued BSI-TR-02102-2 (TLS), a technical guideline providing recommendations for TLS usage, and stating that the "Federal Administration is obliged to encrypt all data traffic over unprotected networks using SSL/TLS" [27].

Japan. The Japanese government announced a standard security guideline [15] in July 2018, which explicitly states that servers should install legitimate certificates for supporting TLS protocol.

EU. Enforced on May 25 2018, the General Data Protection Regulation (GDPR) is a new regulation that aims to enhance personal privacy on the Internet. Adopting HTTPS is considered a baseline security measure for protecting the PII during transmission.

5.2 Outsourcing

To understand the factors that influence the dependency level of government websites, we conducted two online surveys aimed at

web technicians who maintain government websites in Taiwan[7] in September 2017 and September 2018. To recruit more survey participants, we contacted 161 government agencies (not limited to the top 100). Among 73 websites with valid responses, only three (4%) are self-developed by the institutions themselves; the rest have been outsourced to third-party companies. Outsourcing the website to third-party contractors is reasonable for small government agencies, which often lack sufficient resources and budgets to hire dedicated staff specialized in network security and web development. Unfortunately, outsourcing development to third-party contractors may introduce an additional barrier in communicating required security features, thereby further complicating the security configuration and maintenance of government websites. When the majority of the websites are developed by third-party contractors, technical specification plays an important role in defining the requirements in website development. However, 8 of the 70 websites did not put HTTPS in their specification.

Our survey results also indicate that government website administrators consciously consider whether a relevant mandate exists when prioritizing their tasks. Six out of 23 participants (26%) reported that the main reason for not fully upgrading to HTTPS is beause there is "no mandate from the supervising agencies", possibly reducing the priority of HTTPS adoption given other tasks. This echoes our observation that government policies can expedite the deployment of security measures.

5.3 Fear of large-scale Surveillance

Although pinning government websites to government root CAs could eliminate reliance on hundreds of CAs in the browser's default trust store, governments root CAs may be abused for large-scale surveillance [12, 14, 28]. Due to this privacy concern, several government root certificates are excluded from trust stores. For example, the government root certificates of France, Japan, and the US are included in Microsoft Windows' trust store, but not in Apple macOS or Mozilla. However, such partial inclusion may confuse users who see inconsistent certificate warnings. Another complication is that certificate pinning mechanisms are not widely supported on the web. HTTP Public Key Pinning (HPKP) is rarely used by websites and has been deprecated by the latest version of Chrome. DNS-based Authentication of Named Entities (DANE) has not been fully supported by major browsers either.

5.4 Usability Issues

For cyber autonomy, it is crucial that the citizens can distinguish government websites from potentially spoofed websites. Some countries use government-specific (top level) domains, such as .gov for the US and UK, and .go.jp for Japan; hence, users can easily recognize if the website in question belongs to the government of the representing country. Unfortunately, without such a domain requirement, users may fall for attacks that might have been easily identifiable with government-specific domain policies.

Another usability issue arises when governments support their own root CAs for cyber autonomy, but the CAs are not recognized as trustworthy by the browsers or OS. As reported in §3.2, Japan operates its government root CAs, GPKI ApplicationCA2 Root and Japanese Government ApplicationCA, but they are only included in Microsoft Windows' trust store, not Apple macOS or Mozilla. Consequently, users experience frequent certificate errors when they use Safari or Mozilla to access a Japanese government website whose certificate is endorsed by either of these two CAs. Such frequent errors may result in a user (1) ignoring certificate warnings [54], or (2) losing trust in the government websites.

[7]The HTTPS adoption rate among the top 100 gov websites in Taiwan is 87% in 2018, and among the 68 websites with valid certificates, 63 (92.6%) are using the domestic CAs, and 60 (88%) using the government root CAs. As for foreign resources, there were only one URL to Germany and four to US loaded by the top 100 government websites.

6 RELATED WORK

To the best of our knowledge, this is the first study examining the external dependency of public-facing government websites; prior work focuses on one aspect of dependency of popular websites.

A large body of research investigates the current practices of HTTPS. These studies usually scan the IPv4 address space or popular domains, or analyze collected datasets to uncover statistical facts and correlations [36, 50]. Felt et al. [50] observed an increase of HTTPS support from 2016 to 2017. In addition to HTTPS adoption rates, we also investigated websites vulnerable to known TLS attacks and discuss the effectiveness of government policies on prompting HTTPS adoption. To identify HTTPS deployment challenges encountered by security experts, Krombholz et al. [42] performed a lab experiment with 28 technically-competent students and interviewed seven security auditors. Our survey targeted government website operators instead of developers, and the results suggest infrastructural and administrative challenges that government websites may encounter when using HTTPS.

Fadai et al. [41] analyzed the trust stores of multiple browsers and operating systems and identified CAs owned by companies and governmental institutions that might not be trustworthy. Vallina-Rodriguez et al. [55] compared the trust stores on Android devices with different OS versions and manufacturers. We focused on analyzing government root certificates and the certificates used by government websites.

Simeonovski et al. [52] proposed to model web dependencies using a property graph, and applied the model to analyze Alexa's Top 100k domains. Kumar et al. [43] also investigated the dependencies on external services of Alexa's Top Million domains. They found that the median of external resources loaded by websites is 23, which is much higher than for government websites. Such differences suggest the importance of performing in-depth studies based on website types, because each type of website may have unique challenges and incentives for security deployment.

7 CONCLUSIONS AND FUTURE WORK

This paper attempts to measure the current practice of building cyber autonomy on government websites. Our findings indicate the importance of having government mandates, as well as synchronizing autonomy-oriented government policies with corresponding entities. An interesting future direction is to investigate dependency on software (e.g., web frameworks and libraries), hardware (e.g., servers and routers), and network (e.g., routing paths), as well as to assess a government's cyber autonomy from other complementary perspectives. We hope that our findings serve as a stepping stone towards building cyber autonomy.

ACKNOWLEDGMENT

This work was financially supported by Ministry of Science and Technology in Taiwan, under Grant MOST108-2636-E-002-005, MOST108-2636-H-002-002, and MOST107-3017-F-002-004, as well as the Grant (#107L900204) by the Ministry of Education (MOE) in Taiwan.

REFERENCES

[1] [n. d.]. autonomy. In *The Oxford Dictionary*. https://en.oxforddictionaries.com/definition/autonomy

[2] [n. d.]. Data collected in this study. https://osf.io/r3qke/?view_only=36da78809786467394 89469e2f4593e8.

[3] Accessed: 2017-10-30. HTTP Over TLS. https://tools.ietf.org/html/rfc2818.

[4] Accessed: 2017-10-30. HTTP Strict Transport Security (HSTS). https://tools.ietf.org/html/rfc6797.

[5] Accessed: 2017-10-30. The HTTPS-Only Standard.

[6] Accessed: 2017-10-30. The Heartbleed Bug. http://heartbleed.com/.

[7] Accessed: 2018-11-05. Alexa. https://www.alexa.com.

[8] Accessed: 2018-11-05. An update on state-sponsored activity. https://www.blog.google/technology/safety-security/update-state-sponsored-activity/.

[9] Accessed: 2018-11-05. ApplemacOS 10.13 (High Sierra). https://support.apple.com/en-us/HT208127.

[10] Accessed: 2018-11-05. As of November 2018, it issues more than 60% ofcertificates on the Internet. https://censys.io/certificates/report?q=tags%3Atrusted&field=parsed.issuer.organization.raw&max_buckets=10.

[11] Accessed: 2018-11-05. County Election Websites Can Be Easily Spoofed to Spread Misinformation. https://www.darkreading.com/vulnerabilities-threats/county-election-websites-can-be-easily-spoofed-to-spread-misinformation/d/d-id/1333132.

[12] Accessed: 2018-11-05. French gov used fake Google certificate to read its workers' traffic. https://www.theregister.co.uk/2013/12/10/french_gov_dodgy_ssl_cert_reprimand/.

[13] Accessed: 2018-11-05. Germany, seeking independence from U.S., pushes cyber security research. https://www.reuters.com/article/us-germany-cyber/germany-seeking-independence-from-u-s-pushes-cyber-security-research-idUSKCN1LE1EX.

[14] Accessed: 2018-11-05. Government Certification Authorities. https://wiki.mozilla.org/CA:GovernmentCAs#Concerns_about_Government_CAs.

[15] Accessed: 2018-11-05. Guidelines for formulating measures standards by government agencies (Heisei 30th edition) . https://www.nisc.go.jp/active/general/pdf/guide30.pdf.

[16] Accessed: 2018-11-05. Hawaii missile alert: False alarm sparks panic in US state. https://www.bbc.com/news/world-us-canada-42677604.

[17] Accessed: 2018-11-05. How to spot a fake government website. https://www.gov.sg/news/content/how-to-spot-a-fake-government-website.

[18] Accessed: 2018-11-05. Managing service domains. https://www.gov.uk/service-manual/technology/managing-domain-names.

[19] Accessed: 2018-11-05. Microsoft says it has found a Russian operation targeting U.S. political institutions. https://www.washingtonpost.com/business/economy/microsoft-says-it-has-found-a-russian-operation-targeting-us-political-institutions/2018/08/20/52273e14-a4d2-11e8-97ce-cc9042272f07_story.html?noredirect=on&utm_term=.db741731087b.

[20] Accessed: 2018-11-05. Microsoft Windows (as of January 30, 2018). https://gallery.technet.microsoft.com/Trusted-Root-Certificate-70150b50.

[21] Accessed: 2018-11-05. Mozilla (as of October 19, 2017). https://wiki.mozilla.org/CA/Included_Certificates.

[22] Accessed: 2018-11-05. myGov scam tricking victims into handing over bank details through cloned website. https://www.abc.net.au/news/2018-07-05/mygov-scam-clones-government-website-medicare-phishing-email/9942908.

[23] Accessed: 2018-11-05. NSA reportedly intercepting laptops purchased online to install spy malware. https://www.theverge.com/2013/12/29/5253226/nsa-cia-fbi-laptop-usb-plant-spy.

[24] Accessed: 2018-11-05. People in France warned over scam versions of government websites. https://www.thelocal.fr/20180917/people-in-france-warned-over-scam-versions-of-government-websites.

[25] Accessed: 2018-11-05. Syrian hackers claim AP hack that tipped stock market by $136 billion. Is it terrorism? https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/?noredirect=on&utm_term=.266ad973fec7.

[26] Accessed: 2018-11-05. THE LEAKED NSA SPY TOOL THAT HACKED THE WORLD. https://www.wired.com/story/eternalblue-leaked-nsa-spy-tool-hacked-world/.

[27] Accessed: 2018-11-05. The State of IT Security in Germany 2017 . https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Securitysituation/IT-Security-Situation-in-Germany-2017.pdf.

[28] Accessed: 2018-11-05. Turkish government agency spoofed Google certificate "accidentally". https://arstechnica.com/information-technology/2013/01/turkish-government-agency-spoofed-google-certificate-accidentally/.

[29] Accessed: 2018-11-05. U.S. ban on sales to China's ZTE opens fresh front as tensions escalate. https://www.reuters.com/article/us-china-zte/u-s-ban-on-sales-to-chinas-zte-opens-fresh-front-as-tensions-escalate-idUSKBN1HN1P1.

[30] Accessed: 2018-11-05. Washington State system hacked, data of thousands at risk. https://www.reuters.com/article/us-usa-hack-washingtonstate/washington-state-system-hacked-data-of-thousands-at-risk-idUSBRE9480YY20130509.

[31] David Adrian, Karthikeyan Bhargavan, Zakir Durumeric, Pierrick Gaudry, Matthew Green, J Alex Halderman, Nadia Heninger, Drew Springall, Emmanuel Thomé, Luke Valenta, et al. 2015. Imperfect forward secrecy: How Diffie-Hellman fails in practice. In *Proceedings of ACM SIGSAC Conference on Computer and Communications Security*.

[32] Nadhem J AlFardan, Daniel J Bernstein, Kenneth G Paterson, Bertram Poettering, and Jacob CN Schuldt. 2013. On the Security of RC4 in TLS.. In *Proceedings of USENIX Security Symposium*.

[33] Nimrod Aviram, Sebastian Schinzel, Juraj Somorovsky, Nadia Heninger, Maik Dankel, Jens Steube, Luke Valenta, David Adrian, J Alex Halderman, Viktor Dukhovni, et al. 2016. DROWN: Breaking TLS Using SSLv2. In *Proceedings of USENIX Security Symposium*.

[34] Benjamin Beurdouche, Karthikeyan Bhargavan, Antoine Delignat-Lavaud, Cédric Fournet, Markulf Kohlweiss, Alfredo Pironti, Pierre-Yves Strub, and Jean Karim Zinzindohoue. 2015. A messy state of the union: Taming the composite state machines of TLS. In *Proceedings of IEEE Symposium on Security and Privacy (SP)*.

[35] Karthikeyan Bhargavan and Gaëtan Leurent. 2016. On the practical (in-) security of 64-bit block ciphers: Collision attacks on HTTP over TLS and OpenVPN. In *Proceedings of ACM SIGSAC Conference on Computer and Communications Security*.

[36] Li Chang, Hsu-Chun Hsiao, Wei Jeng, Tiffany Hyun-Jin Kim, and Wei-Hsi Lin. 2017. Security Implications of Redirection Trail in Popular Websites Worldwide. In *Proceedings of the 26th International Conference on World Wide Web (WWW)*.

[37] Jeremy Clark and Paul C van Oorschot. 2013. SoK: SSL and HTTPS: Revisiting past challenges and evaluating certificate trust model enhancements. In *Proceedings of IEEE Symposium on Security and Privacy (SP)*.

[38] Thai Duong and Juliano Rizzo. 2011. Here come the ⊕ ninjas. *Unpublished manuscript* (2011).

[39] Thai Duong and Juliano Rizzo. 2012. The CRIME attack. In *Presentation at ekoparty Security Conference*.

[40] Zakir Durumeric, James Kasten, David Adrian, J Alex Halderman, Michael Bailey, Frank Li, Nicolas Weaver, Johanna Amann, Jethro Beekman, Mathias Payer, et al. 2014. The matter of heartbleed. In *Proceedings of ACM Internet Measurement Conference (IMC)*.

[41] Tariq Fadai, Sebastian Schrittwieser, Peter Kieseberg, and Martin Mulazzani. 2015. Trust me, I'm a Root CA! Analyzing SSL Root CAs in Modern Browsers and Operating Systems. In *International Conference on Availability, Reliability and Security (ARES)*.

[42] Katharina Krombholz, Wilfried Mayer, Martin Schmiedecker, Edgar Weippl, Wilfried Mayer, Martin Schmiedecker, and Edgar Weippl. 2017. "I Have No Idea What I'm Doing" - On the Usability of Deploying HTTPS. In *Proceedings of USENIX Security Symposium*.

[43] Deepak Kumar, Zane Ma, Ariana Mirian, Joshua Mason, J Alex Halderman, and Michael Bailey. 2017. Security Challenges in an Increasingly Tangled Web. In *In Proceedings of the World Wide Web Conference (WWW)*.

[44] Bill Marczak, Nicholas Weaver, Jakub Dalek, Roya Ensafi, David Fifield, Sarah McKune, Arn Rey, John Scott-Railton, Ron Deibert, and Vern Paxson. 2015. An analysis of china's "great cannon". *FOCI. USENIX* (2015), 37.

[45] Nikos Mavrogiannopoulos, Frederik Vercauteren, Vesselin Velichkov, and Bart Preneel. 2012. A cross-protocol attack on the TLS protocol. In *Proceedings of ACM SIGSAC conference on Computer and communications security*.

[46] Bodo Möller, Thai Duong, and Krzysztof Kotowicz. 2014. *This POODLE Bites: Exploiting The SSL 3.0 Fallback*. Technical Report. Google.

[47] Nick Nikiforakis, Luca Invernizzi, Alexandros Kapravelos, Steven Van Acker, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2012. You are what you include: large-scale evaluation of remote javascript inclusions. In *Proceedings of the 2012 ACM conference on Computer and communications security*.

[48] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhooby, Maciej Korczynskiz, and Wouter Joosen. 2019. TRANCO: A Research-Oriented Top Sites Ranking Hardened Against Manipulation.

[49] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP Geolocation Databases: Unreliable?. In *Proceedings of ACM SIGCOMM*.

[50] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. 2017. Measuring HTTPS Adoption on the Web. In *Proceedings of USENIX Security Symposium*.

[51] Marsh Ray and Steve Dispensa. 2009. Renegotiating TLS. IETF-76.

[52] Milivoj Simeonovski, Giancarlo Pellegrino, Christian Rossow, and Michael Backes. 2017. Who controls the internet?: Analyzing global threats using property graph traversals. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee.

[53] Christopher Soghoian and Sid Stamm. 2011. Certified lies: Detecting and defeating government interception attacks against SSL (short paper). In *International Conference on Financial Cryptography and Data Security*.

[54] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. 2009. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *Proceedings of the 18th USENIX Security Symposium*.

[55] Narseo Vallina-Rodriguez, Johanna Amann, Christian Kreibich, Nicholas Weaver, and Vern Paxson. 2014. A Tangled Mass: The Android Root Certificate Stores. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*.

[56] Mathy Vanhoef and Frank Piessens. 2015. All Your Biases Belong to Us: Breaking RC4 in WPA-TKIP and TLS.. In *Proceedings of USENIX Security Symposium*.

[57] Sebastian Zander. 2012. *How Accurate is IP Geolocation Based on IP Allocation Data?* Technical Report. Swinburne University of Technology.