Understanding the Security of Free Content Websites by Analyzing their SSL Certificates: A Comparative Study

Abdulrahman Alabduljabbar
University of Central Florida
Orlando, FL, USA
jabbar@knights.ucf.edu

Runyu Ma
George Mason University
Fairfax, VA, USA
rma5@gmu.edu

Soohyeon Choi
University of Central Florida
Orlando, FL, USA
soohyeon.choi@knights.ucf.edu

Rhongho Jang
Wayne State University
Detroit, MI, USA
r.jang@wayne.edu

Songqing Chen
George Mason University
Fairfax, VA, USA
sqchen@gmu.edu

David Mohaisen
University of Central Florida
Orlando, FL, USA
mohaisen@ucf.edu

ABSTRACT

Online services that provide books, music, movies, etc., for free have existed on the Internet for decades. While there are some common beliefs and warnings that such online services may contain hidden security risks, many ordinary users still visit such websites, making them a convenient vehicle for subsequent exploitation.

In this paper, we investigate and quantify through measurements the potential vulnerability of such free content websites (FCWs). For this purpose, we curated 834 FCWs offering books, games, movies, music, and software. For a comparison purpose, we also sampled a comparable number of premium content websites, where users need to pay for using the service for the same type of content. For our modality of analysis, we use SSL certificates. Namely, we explore SSL certificates' structural and fundamental differences between free and premium content websites.

Through our analysis, we unveil that 36% of the free websites' certificates have major issues, with 17% invalid certificates, 7% expired, and 12% with mismatched domain names. Moreover, although surprisingly, we uncover the usage of ECDSA predominantly among the free websites. Among other observations, we notice that 38% of the FCWs use ECDSA-256, compared to only 20% of their premium counterparts, which provides better security guarantees (and performance) than the common algorithm option and key size (RSA-2048) in premium websites. Our observations raise concerns regarding the safety of using such free services from a transport standpoint and call for in-depth analysis of their risks.

CCS CONCEPTS

• **Information systems → World Wide Web**; **Web mining**; • **Security and privacy → Usability in security and privacy**.

CySSS '22, May 30, 2022, Nagasaki, Japan

ACM ISBN 978-1-4503-9177-1/22/05…$15.00
https://doi.org/10.1145/3494108.3522769

KEYWORDS

Web Security, Free Content Websites, SSL Certificates, Internet Measurements

ACM Reference Format:
Abdulrahman Alabduljabbar, Runyu Ma, Soohyeon Choi, Rhongho Jang, Songqing Chen, and David Mohaisen. 2022. Understanding the Security of Free Content Websites by Analyzing their SSL Certificates: A Comparative Study. In *Proceedings of the 1st Workshop on Cybersecurity and Social Sciences (CySSS '22), May 30, 2022, Nagasaki, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3494108.3522769

1 INTRODUCTION

There has been a recent explosion in popularity and usage of online services and web platforms that deliver content (music, movies, books, etc.). This significant growth in such platform's popularity is in part attributed to the convenience of their use [10, 11, 16–19, 27, 34, 36]. Generally speaking, websites delivering such content are categorized into two groups based on their monetization options: free and premium. The free websites provide free physical or virtual services and are typically run by donations or advertisements [12, 21, 22, 39]. On the other hand, the premium services are either subscription-based or pay-as-you-use. The latter category is, in most cases, strictly mentored to ensure quality, while the free services lack high level of monitoring as they may be user-driven.

The reliance of FCWs on advertisements and user-driven content raises several concerns. For instance, advertisements can be exploited for data and information leakage, in addition to running malicious scripts on the user device [3, 30, 38]. Moreover, the lack of censorship raises security concerns regarding the provided services. For instance, in an attempt to reduce the operational cost of the online service, the service providers may relax their security and privacy requirements or may not use them altogether.

Motivated by these concerns, we explore the fundamental and structural differences between free content and premium websites. In doing this analysis, we use the Secure Sockets Layer (SSL) certificate content. The SSL certificate is a digital authentication method that proves the identity of a website and (eventually) provides an encrypted connection between the client and the server. SSL certificate is a critical element of a website to secure the users' data and protect them against mischievous phishing and skimmers.

To this end, we investigate the validity of the SSL certificate for both free and premium services. In particular, we focus on

understanding the fundamental differences between free and premium content websites in three directions: (i) Errors within the SSL certificate, including unmatched client name, expired certificate, or invalid/vulnerable information and content, (ii) SSL certificate issuer organization analysis, including the most commonly used certificate providers, such as *Cloudflare Inc.*, *Let's Encrypt*, *DigiCert Inc*, and the SSL certification issuer countries distribution analysis (*e.g.,* United States, United Kingdom, and Belgium), and (iii) SSL certificate signature algorithm analysis (*e.g., SHA256 with RSA*, *SHA256 with ECDSA*, and *SHA1 with RSA*).

Understanding the different characteristics of the SSL certificate is crucial for user risk exposure analysis. The most common issues within the SSL certificates are (i) **Untrusted SSL Certificate:** The certificate is not signed by a trusted certificate authority. The website, in this case, publishes a certificate self-signed by the server. (ii) **Domain Name Mismatch:** This happens when the website's URL is different from the domain name in the SSL certificate, which indicates either illegal use of the certificate or inconsistent domain change. (iii) **Mixed content Warning:** This warning is issued when elements among the website content are unsecured, indicating that either such content are malicious or can be exploited. (iv) **SSL Certificate Expired:** Expired SSL certificate may result in out-of-date security practices, causing further exploitation.

The aforementioned reasons motivate for understanding the differences between free and premium services and websites. Toward this goal, our analyses uncover that the two categories are indeed distinguishable, each with shared behavior among its services. Our experimental evaluation shows that 35.85% of the free websites' certificates have significant issues, with up to 17% invalid SSL certificates due to unsecure content and 12% with mismatched domain names. Surprisingly, we uncover the usage of the emerging ECDSA encryption algorithm among the free websites, a faster and more secure option in comparison with the more popular option of RSA-2048 used along with SHA-256 in premium websites.

Contributions and Findings. Starting with a list of 1,562 free and premium services websites obtained from the top results of Google, DuckDuckGo, and Bing search engines, we extract and analyze the SSL certificates toward assessing potential exploitation and risks, across the following verticals.

(1) **SSL Certificate Validity Analysis (§4.1).** We analyze the SSL certificates, extracting existing issues that expose the user to vulnerabilities. We uncover that, on the fundamental level, the free and premium websites are highly distinguishable, with free websites certificates being labeled as 17% invalid, 7% expired, and 12% with mismatched domain names, a ratio that is much higher than its premium counterpart.
(2) **SSL Certificate Issuer Analysis (§4.2).** We analyze the SSL certificate issuing organizations, unveiling that, at the country-level, the SSL certificate issuing organizations are very similar. However, we observed the heavy usage of "Cloudflare" among the free websites (38.22% of the websites), in comparison with only 15.88% among the premium counterparts.
(3) **SSL Certificate Signature Analysis (§4.3).** We study the utilized signature encryption algorithms of the SSL certificates and uncover that on the data encryption level, the premium websites are using RSA with larger public key sizes, in comparison with the emergence of more secure ECDSA algorithm among the free websites (with the used key parameters).

Organization. In section 2, we shed light on the literature of online websites and SSL certificate-level analyses. We discuss our compiled dataset in section 3. Finally, We conclude this work with the main insights and future directions in section 5.

2 RELATED WORK

The related work is divided into two groups: online services and websites analysis, and SSL certificates analysis and measurements. We review the former in §2.1 and the latter in §2.2.

2.1 Online Website Analysis

Online services and web applications are evolving in terms of development and utilization. However, with the evolution of their capabilities, different components in these applications can be compromised, invalidating some security aspects and putting their users at risk, a topic that has been of increasing interest.

For instance, one of the security aspects that is not thoroughly studied in the literature is the validity of the websites' certificate [14]. To address this gap, Chung *et al.* [14] proposed the first in-depth analyses of the invalid certificates in the web public key infrastructure (PKI). The study shows that the vast majority of certificates in the web PKI are invalid. Their study also investigated the source of the invalid certificates, showing that the invalid certificates were mostly generated from end-user devices, with periodic regeneration of new self-signed certificates.

Several studies analyzed the most popular websites' privacy policies [1, 2, 4, 6, 29, 31]. Libert *et al.* [31] evaluated one million popular websites against various privacy-compromising mechanisms, such as data leakage. The privacy can be compromised by running a *JavaScript* code on the users' devices. They show that roughly nine out of ten websites share user data with third-party services without the users' consent, which is a significant concern.

On a similar dataset distribution, Lavrenovs *et al.* [29] (in 2018) conducted a comprehensive assessment of the security for Alexa top-million websites. First, they initiated four types of requests to each website to obtain the HTTP header information. Then, they examined the presence of security-related response header variables, including Strict-Transport-Security, Content-Security-Policy, X-XSS-Protection, X-Frame-Options, Set-Cookie, and X-Content-Type. Their findings demonstrate that 29.1% of HTTPS requests have incorrect Transport Layer Security (TLS) configurations. Moreover, they found that the HTTP Strict Transport Security policy, a critical response that informs browsers a website should only be accessed with HTTPS, is implemented in only 17.5% of the websites. These findings raise worrisome concerns regarding the security policies followed by the Alexa top one million websites.

Exploring environments to evaluate the security flaw in online services and web applications, Alsmadi *et al.* [6] designed a component-based testing mechanism where the goal of their designed mechanism is to test a variety of invalid inputs and investigate the website behavior and security issues raised with such input. Since any invalid input is considered as an attack surface for adversaries to explore, it has to be eliminated, and the right mechanisms have to be implemented to address such an input (*i.e.,*

Table 1: An overview of the collected dataset. The collected URLs are associated with five different categories, and belong to free content and premium websites. Overall, 1,562 websites were analyzed for the purpose of this study.

# URLs	Books	Games	Movies	Music	Software	Overall
Free Content	154	80	331	83	186	834
Premium	195	113	152	86	182	728
Total	349	193	483	169	368	1,562

reject invalid inputs). To do so, the authors proposed several methods for detecting invalid inputs uncovering a large number of SQL injection vulnerabilities.

2.2 SSL Certificate Analysis

SSL certificates are studied in the literature for online risk and vulnerability analysis of websites [5, 8, 9, 13–15, 24–26, 28, 33, 41]. The following is an overview of some recent studies on the topic.

Kim *et al.* [24–26] pioneered the analysis of certificates used for signing malicious codes, highlighting avenues of trust abuse, and quantifying the (in)effectiveness of remedies. Meyer *et al.* [32] analyzed the SSL certificate information and content to differentiate between phishing websites and benign websites. Moreover, Huang *et al.* [23] analyzed forged SSL certificates on the web, showing a small number of the studied SSL connections were tampered with forged SSL certificates, where most of them were related to antivirus software and corporate-scale content filters.

Towards accurate malicious SSL certificate detection, Ghafir *et al.* [20] studied the command and control communication channels of malicious SSL certificate services and their generated patterns and traces. Their experimental evaluation highlighted the successful detection of malicious SSL certificates using blacklisting information, the associated IP addresses, and practices. More recently, Wang *et al.* [40] statically and dynamically analyzed the SSL certificates to extract potential exploitation and vulnerabilities within Android applications. Their analysis showed that 11.07% of the studied applications are prone to man-in-the-middle and phishing.

In this work, we explore the SSL certificate-based fundamental structural differences between free content and premium websites, by understanding their patterns for modeling and characterization.

3 DATASET OVERVIEW

We compiled a list of 1,562 free content (834) and premium (728) websites for our analyses. When selecting the websites, we considered the following factors: (i) selecting the most popular websites, *e.g.,* websites that appear in the top results by Google, DuckDuckGo, and Bing search engines, and (ii) maintaining a balanced dataset. In addition, we verified and labeled each website manually. The compiled websites are then categorized into five groups based on the provided content: books, games, movies, music, or software. Table 1 shows the distribution of the dataset.

Our dataset is then augmented with various SSL certificate attributes, including SSL certificate validation, issuer organization, issuing country, and signature algorithm. Such information was retrieved using APIVoid [7], a framework that provides cyber threat detection and analysis, and OpenSSL [35], a command-line tool to retrieve SSL certificates chains from target websites and parse them to a readable format. We analyze whether the websites' associated SSL certificates are expired, invalid, or unmatched (with respect to the domain name they are used for). In particular, we focus on information that reflects users' exposure to risk, including:

(1) Unmatched hostname in the certificate.
(2) Expired certificate.
(3) Invalid/fabricated certificate.
(4) The certificate validity.

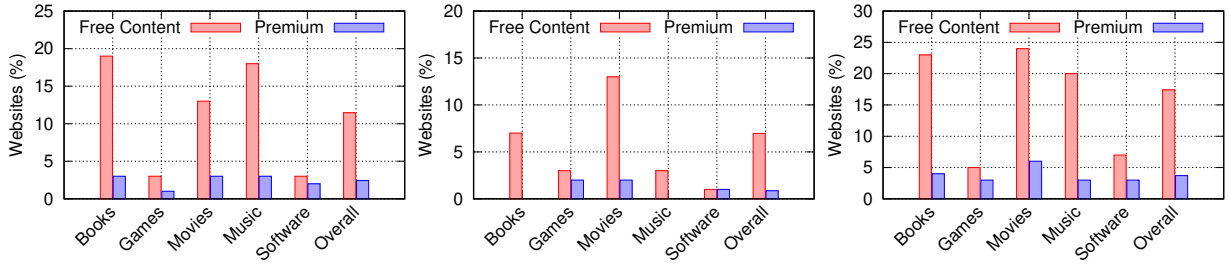
4 SSL CERTIFICATE ANALYSIS OF FCWS

In this section, we analyze the fundamental characteristics of FCWs' and premium websites' SSL certificates, including their validity analysis, issuer organization and country distribution, and signature algorithm and public key size (security parameters).

4.1 SSL Certificate Validity Analysis

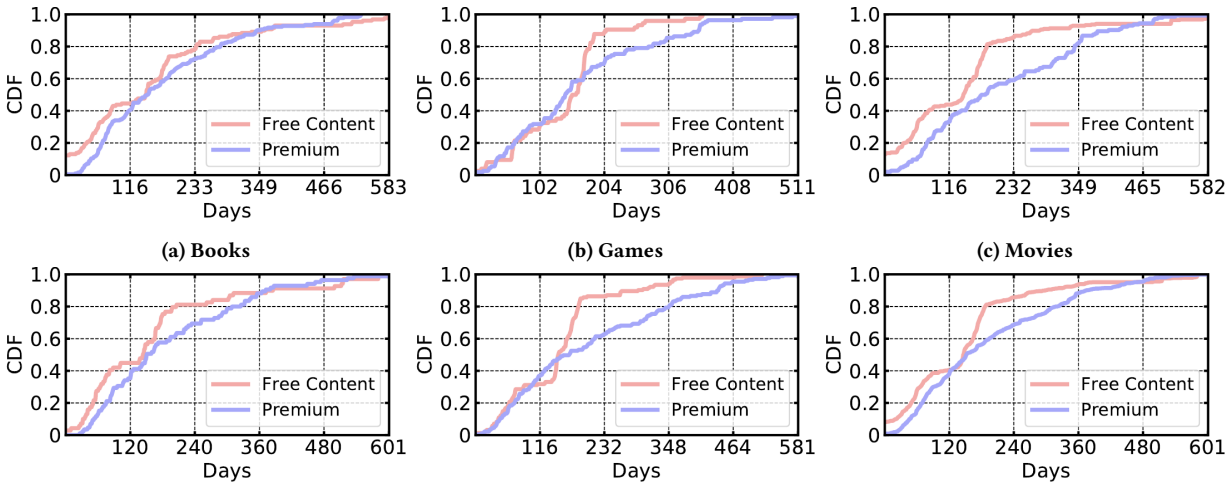
Among the compiled free websites, we notice a significant portion (35.85%) of them have issues with their certificates (*i.e.,* 11.47% unmatched name, 6.97% expired, and 17.42% invalid certificate), compared to only 6.99% of the premium websites' SSL certificates. To better understand the implications of such issues, we divided our analysis into four directions as follows.

(1) **SSL Mismatched Domain.** Mismatched domain indicates either (i) impersonation of another website, or (ii) inconsistent website migration and domain transfer, and both indicate a lack of rigorous security practices by the service providers. Figure 1a shows the vast discrepancy between free and premium websites in the context of SSL mismatching. For instance, 19% of the "Books" related free FCWs have unmatched domain names, in comparison with only 3% of the premium counterpart. Along with the five categories, 12% of the FCWs SSLs have domain mismatch, compared to only 2.3% of the premium websites.
(2) **SSL Validity Days.** We analyze the SSL certificate validity (*i.e.,* number of days until the SSL certificate becomes invalid if not renewed), in Figure 2. Notice that the validity days for premium websites are noticeably higher than their free websites, particularly for "Movies" and "Software" websites.
(3) **SSL Expired.** SSL expiration may not directly affect users' security or usage experience, but is an indication that the used data encryption may be out-of-date, increasing the risk of future exploitation. In line with our previous observation, we notice that the expired SSL certificates (*i.e.,* $\leq$ 0 validity days) for FCWs are significantly more than for the premium websites—*i.e.,* 7% of the FCWs' SSL certificates are expired, in comparison with only 1% of the premium websites—as shown in Figure 1b. This may be attributed to the fact that FCWs operators are not renewing the SSL certificate for being unwilling to increase their operational costs. Nonetheless, this practice leads to potential risks regarding user information and data privacy.
(4) **SSL Invalid.** Unlike the previous issues, invalid SSL indicates the usage of vulnerable and insecure elements within the website content. This, in turn, can highly affect the users' data and client safety with potential data and resources exploitation. Toward understanding the risks associated with FCWs, Figure 1c shows the percentage of websites associated with invalid SSL



(a) SSL Mismatched Domain (b) SSL Expired (c) SSL Invalid

Figure 1: The SSL certificate analysis results. We observe that almost 36% of the FCWs have problematic SSL certificates (unmatched, expired, or invalid) compared to 7% in premium websites.



(a) Books (b) Games (c) Movies

(d) Music (e) Software (f) Overall

Figure 2: The CDF of SSL certificate validity days. The premium websites SSL certificates are valid over extended period of time, unlike the FCWs counterparts, where multiple instances are expired.

certificates. In particular, 17% of the FCWs' certificates are invalid, in comparison with only 4% of the premium websites. This gap is even higher for the "Movies" category, with 24.5% and 6% invalid free and premium websites SSL certificates, respectively.

Key Takeaway: On the SSL certificate fundamental level, the free and premium websites are highly distinguishable, where the free websites' certificates are 17% invalid, 7% expired, and 12% with mismatched domain names.

4.2 SSL Certificate Issuer Analysis

A certificate authority (CA) is an organization that issues digital certificates by signing with their private key. To further understand the characteristic differences between free and premium websites, we analyze the hosting platforms and their country-level distribution.

(1) **SSL Certificate Issuer Organization.** Table 2 and Table 3 show the distribution of the free and premium websites certificates' issuing organizations. We found the FCWs heavily use "Cloudflare" for their SSL certificates (38.22% of websites), in comparison with only 15.88% of the premium websites. Moreover, while "DigiCert" is not commonly used for free websites, with only 5.19% of the websites' SSL certificates associated with the organization, it is commonly used among the premium websites, with 22.89% of their certificates issued by the organization.

(2) **SSL Certificate Issuing Country.** Next, we explored the country-level distribution of the SSL certificate issuing organizations, shown in Table 4. Notice that, for both categories, the United States dominates the distribution, with 86.88% and 83.31% of the free and premium websites SSL certificate issuing organizations, respectively. Overall, 94.56% of the issuing organizations are located within the United States and the United Kingdom.

Key Takeaway: On country-level, the SSL certificate issuing organizations are very similar. However, we uncover the heavy usage of "Cloudflare" among the free websites (38.22% of the websites), in comparison with only 15.88% among the premium counterparts.

4.3 SSL Certificate Signature Analysis

In a website's SSL certificate, the key is split into two pieces. One piece is used to encrypt a message and the other is used to decrypt it. These keys allow exchanging information over unsecured channels. Alternatively, the decryption (private) key is used for signing messages and the encryption (public) key is used for signature verification by the recipient. The strength of the encryption

Table 2: A comparison between free content and premium websites (%) in terms of SSL certificate issuer organizations.

Free Content Websites						
Issuer Organization	Books	Games	Movies	Music	Software	Overall
Cloudflare_ Inc.	24.80	59.72	35.29	27.94	48.39	38.22
Let's Encrypt	35.20	27.78	32.94	39.71	30.97	33.04
Sectigo Limited	15.20	4.17	11.76	4.41	5.81	9.48
DigiCert_ Inc.	6.40	2.78	6.27	10.29	1.29	5.19
cPanel_ Inc.	4.00	1.39	3.53	0.00	3.23	2.96
Cisco	0.80	1.39	5.49	0.00	0.00	2.37
GoDaddy.com_ Inc.	4.80	1.39	0.78	4.41	1.94	2.22
Others	8.80	1.39	3.92	13.24	8.39	6.52
Premium Websites						
Issuer Organization	Books	Games	Movies	Music	Software	Overall
DigiCert_ Inc.	22.16	25.23	24.11	20.24	22.47	22.89
Cloudflare_ Inc.	18.92	23.42	7.09	14.29	15.73	15.88
Let's Encrypt	19.46	12.61	13.48	15.48	14.04	15.31
Amazon	14.59	7.21	13.48	13.10	8.99	11.59
GoDaddy.com_ Inc.	7.57	3.60	8.51	10.71	12.92	8.87
Sectigo Limited	7.57	10.81	5.67	3.57	12.36	8.44
GlobalSign nv-sa	5.95	8.11	10.64	11.90	3.37	7.30
Others	3.78	9.01	17.02	10.71	10.11	9.73
All Websites						
Issuer Organization	Books	Games	Movies	Music	Software	Overall
Cloudflare_ Inc.	21.29	37.70	25.25	20.39	30.93	26.86
Let's Encrypt	25.81	18.58	26.01	26.32	21.92	24.02
DigiCert_ Inc.	15.81	16.39	12.63	15.79	12.61	14.19
Sectigo Limited	10.65	8.20	9.60	3.95	9.31	8.95
Amazon	9.35	4.37	5.81	9.21	5.41	6.70
GoDaddy.com_ Inc.	6.45	2.73	3.54	7.89	7.81	5.60
GlobalSign nv-sa	3.87	4.92	4.04	8.55	2.40	4.22
Others	6.77	7.10	13.13	7.89	9.61	9.46

Table 3: The difference between free content and premium websites (%) in terms of SSL certificate issuer organizations.

	Free Content		Premium		
Issuer Organization	#	%	#	%	Diff (%)
Cloudflare_ Inc.	258	38.22	111	15.88	+22.34
Let's Encrypt	223	33.04	107	15.31	+17.73
Sectigo Limited	64	9.48	59	8.44	+01.04
DigiCert_ Inc.	35	5.19	160	22.89	-17.70
cPanel_ Inc.	20	2.96	7	1.00	+01.96
Cisco	16	2.37	0	0.00	+02.37
GoDaddy.com_ Inc.	15	2.22	62	8.87	-06.65
Amazon	11	1.63	81	11.59	-09.96
GlobalSign nv-sa	7	1.04	51	7.30	-06.26
Google Trust Services LLC	1	0.15	8	1.14	-01.00
Entrust_ Inc.	0	0.00	12	1.72	-01.72
Others	25	3.70	41	5.87	-02.16

(alternatively, signature) is determined by two factors: (i) the used algorithm and (ii) the used key size.

Signature Algorithms. Table 5 and Table 6 show the different signature algorithms used by free and premium websites to sign data (*i.e.,* payload). We observe that while 60.74% of the FCWs use SHA256 with RSA signature mechanism (hash-then-sign), the majority (83.26%) of the premium websites use this algorithms combination (~23% difference). This is mainly attributed to it being the traditional go-to algorithm solution adopted by service providers. On the other hand, 38.37% of the FCWs rely on the newer and faster ECDSA (Elliptic Curve Digital Signature Algorithm) algorithm, which uses shorter keys for the same security level as in RSA with larger keys. In comparison, ECDSA is used by only 16.60% of

Table 4: A comparison between free content and premium websites(%) in terms of SSL certification issuer countries.

Free Content Websites						
Issuer Country	Books	Games	Movies	Music	Software	Overall
United States (US)	82.17	100.00	86.30	86.30	91.28	86.88
United Kingdom (UK)	14.73	0.00	13.70	6.85	4.70	10.08
Belgium (BE)	0.78	0.00	0.00	4.11	1.34	1.12
Austria (AT)	0.00	0.00	0.00	0.00	1.34	0.64
Self-Sign	0.78	0.00	0.00	1.37	1.34	0.64
Australia (AU)	0.78	0.00	0.00	0.00	0.00	0.32
Netherlands (NL)	0.78	0.00	0.00	0.00	0.00	0.16
China (CN)	0.00	0.00	0.00	1.37	0.00	0.16
Premium Websites						
Issuer Country	Books	Games	Movies	Music	Software	Overall
United States (US)	84.78	80.81	81.56	80.00	82.12	83.31
United Kingdom (UK)	7.61	10.10	7.80	3.53	13.41	8.83
Belgium (BE)	5.98	8.08	10.64	11.76	3.35	6.26
France (FR)	0.54	0.00	0.00	3.53	0.00	0.64
China (CN)	1.09	0.00	0.00	0.00	0.56	0.48
Japan (JP)	0.00	0.00	0.00	1.18	0.00	0.16
Italy (IT)	0.00	0.00	0.00	0.00	0.56	0.16
Austria (AT)	0.00	1.01	0.00	0.00	0.00	0.16
All Websites						
Issuer Country	Books	Games	Movies	Music	Software	Overall
United States (US)	83.71	86.99	83.18	82.91	86.28	85.10
United Kingdom (UK)	10.54	6.85	9.81	5.06	9.45	9.46
Belgium (BE)	3.83	5.48	7.01	8.23	2.44	3.69
France (FR)	0.32	0.00	0.00	1.90	0.00	0.32
China (CN)	0.64	0.00	0.00	0.63	0.30	0.32
Japan (JP)	0.00	0.00	0.00	0.63	0.00	0.08
Italy (IT)	0.00	0.00	0.00	0.00	0.30	0.08
Austria (AT)	0.00	0.68	0.00	0.00	0.61	0.40
Australia (AU)	0.32	0.00	0.00	0.00	0.00	0.16
Netherlands (NL)	0.32	0.00	0.00	0.00	0.00	0.08
Self-Sign	0.32	0.00	0.00	0.63	0.61	0.32

Table 5: A comparison between free content and premium websites (%) in terms of SSL certificate signature algorithms.

Free Content Websites						
Signature Algorithm	Books	Games	Movies	Music	Software	Overall
SHA256 with RSA	73.60	38.89	64.71	69.12	50.32	60.74
SHA256 with ECDSA	24.80	59.72	35.29	27.94	48.39	38.22
SHA1 with RSA	0.80	1.39	0.00	1.47	0.00	0.44
SHA384 with RSA	0.80	0.00	0.00	1.47	0.65	0.44
SHA384 with ECDSA	0.00	0.00	0.00	0.00	0.65	0.15
Premium Websites						
Signature Algorithm	Books	Games	Movies	Music	Software	Overall
SHA256 with RSA	80.54	75.68	92.20	83.33	83.71	83.26
SHA256 with ECDSA	19.46	23.42	7.80	16.67	16.29	16.60
SHA384 with RSA	0.00	0.90	0.00	0.00	0.00	0.14
SHA1 with RSA	0.00	0.00	0.00	0.00	0.00	0.00
SHA384 with ECDSA	0.00	0.00	0.00	0.00	0.00	0.00

the premium websites. We note that, while ECDSA is a newer and more efficient algorithm adopted by the newer free websites, recent studies suggest that it is more vulnerable to attacks [37] than the traditional RSA algorithm with post-quantum adversary.

Key Size. The other factor in enhancing the strength of the encryption and signature is the key size. A larger key size exponentially increases the time needed to crack and decrypt the encrypted information (and conversely for the signature forgery). Figure 3 shows the commonly used key sizes among the websites. We note that the "Firefox" Internet browser no longer supports a key size of less than 2048. We observe that the shorter keys (256, for example) are associated with the ECDSA algorithm choice. However, we note

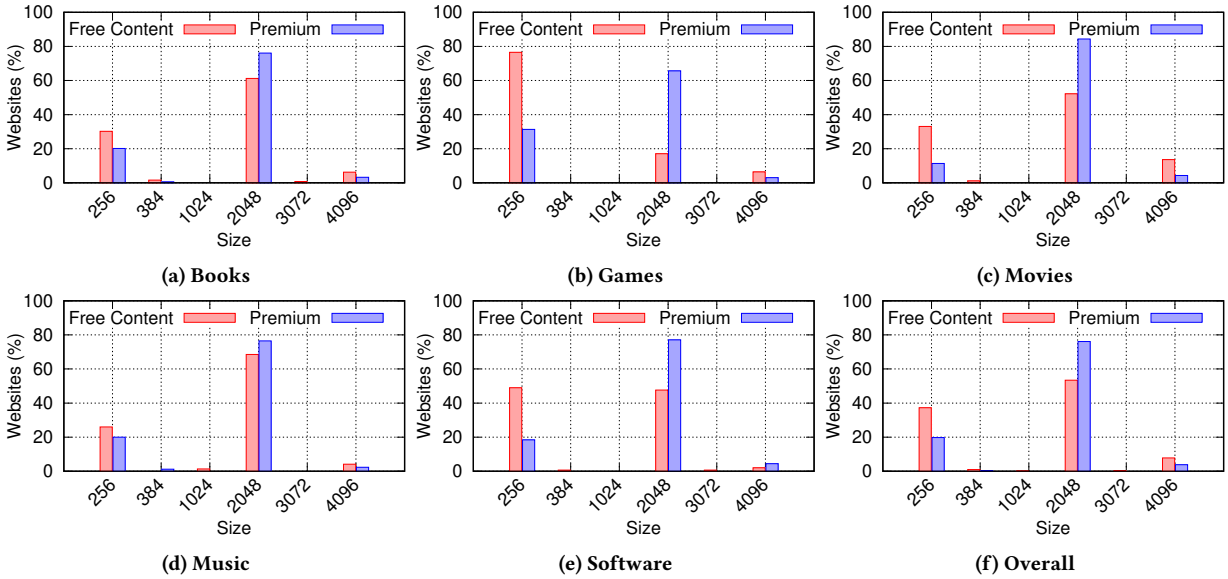


Figure 3: The key size analysis results. We observed that while majority of websites uses the key size of 2048, the portion of FCWs using key size of 256 is significantly higher than premium websites, particularly in "Games" and "Software" categories.

Table 6: The difference between free and premium content websites (%) in terms of SSL certificate signature algorithms.

	Free Content		Premium		
Signature Algorithm	#	%	#	%	Diff (%)
SHA256 with RSA	410	60.74	582	83.26	-22.52
SHA256 with ECDSA	258	38.22	116	16.60	+21.63
SHA1 with RSA	3	0.44	0	0.00	+00.44
SHA384 with RSA	3	0.44	1	0.14	+00.30
SHA384 with ECDSA	1	0.15	0	0.00	+00.15

that ECDSA-256 has the same security level as RSA-3072. Among the free websites, 38% are using this key size and algorithm choice, making them more secure than the majority of websites utilizing RSA key of 2048 bits (in comparison to only 19% of premium websites). Across all categories, the premium websites usage of keys size 2048 is significantly higher, showing that they might not be as secure—although theoretically, since such an insecurity is only possible with a post-quantum adversary.

A plausible explanation of the choice of algorithm and key size is that FCWs are emerging, often with a short life time, making them apt to the utilization of new algorithms, in contrast to well-established premium services deployed for many years where key-rollover and algorithm update are costly.

Key Takeaway: On the data signature algorithms, more premium websites are still using RSA while more FCWs have adopted the more recent ECDSA algorithm, possibly due to the more recent emergence of FCWs, making them easier to adopt new technologies.

5 CONCLUSION & FUTURE DIRECTIONS

The Internet is the dominant channel for marketing, promotion, and communication, particularly via providing online physical and digital services. Recent years have witnessed the rise of websites that provide content for free, dubbed free content websites. In this work, we explored the unique SSL certificate characteristics of premium and FCWs to understand their commonalities and differences. Understanding the potential risks associated with invalid SSL certificates, including, but not limited to, untrusted SSL certificates, expired certificates, mixed content warnings, or invalid/vulnerable information and content. Through our analysis, we highlight that 35.85% of the free websites' certificates have significant issues, with 17% invalid, 7% expired, and 12% with mismatched domain names. Moreover, we uncover the usage of the emerging ECDSA encryption algorithm among the free websites with a key size that would seem to provide better (classical) security and performance than that of the algorithm (and associated key size) used in premium websites; with 38% of FCWs using the ECDSA key size of 256, in comparison with only 20% of their premium counterparts.

Toward a safe and secure web environment, we highlight that FCWs would highly benefit from consistent monitoring and management, particularly with the increasing trend of invalid SSL certificates among them (although not the only risk). Our observations raise concerns regarding the safety of using such free services, especially when such usage could put users at risk, and call for in-depth analysis of their potential risks, ramifications, and remedies.

Future Directions. To understand the risks of FCWs, we will examine various features of FCWs. We will further investigate the URLs and website content toward their malicious activities.

Acknowledgement. This research was supported by Global Research Laboratory (GRL) Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2016K1A1A2912757). A. Alabduljabbar is also supported in part by the Saudi Arabian Cultural Mission (SACM). R. Ma and S. Chen are supported in part by a Commonwealth Cyber Initiative grant and NSF grant CNS-2007153. The authors would like to thank the anonymous reviewers for their feedback.

REFERENCES

[1] Abdulrahman Alabduljabbar, Ahmed Abusnaina, Ülkü Meteriz-Yildiran, and David Mohaisen. 2021. Automated Privacy Policy Annotation with Information Highlighting Made Practical Using Deep Representations. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS'21)*. 2378–2380.

[2] Abdulrahman Alabduljabbar, Ahmed Abusnaina, Ülkü Meteriz-Yildiran, and David Mohaisen. 2021. TLDR: Deep Learning-Based Automated Privacy Policy Annotation with Key Policy Highlights. In *ACM Workshop on Privacy in the Electronic Society (WPES'21)*. ACM, 103–118.

[3] Abdulrahman Alabduljabbar, Runyu Ma, Sultan Alshamrani, Rhongho Jang, Songqing Chen, and David Mohaisen. 2022. Poster: Measuring and Assessing the Risks of Free Content Websites. In *Network and Distributed System Security Symposium, (NDSS'22), San Diego, California, April 24-28, 2022*. The Internet Society.

[4] Abdulrahman Alabduljabbar and David Mohaisen. 2022. Measuring the Privacy Dimension of Free Content Websites through Automated Privacy Policy Analysis and Annotation. In *Companion Proceedings of the Web Conference 2022*.

[5] Omar Alrawi and Aziz Mohaisen. 2016. Chains of Distrust: Towards Understanding Certificates Used for Signing Malicious Applications. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11-15, 2016, Companion Volume*. ACM, 451–456.

[6] Izzat Alsmadi and Fahad Mira. 2018. Website security analysis: variation of detection methods and decisions. In *2018 21st Saudi Computer Society National Computer Conference (NCC)*. 1–5.

[7] APIVoid. 2022. A framework provides JSON APIs useful for cyber threat analysis, threat detection and prevention. https://app.apivoid.com/

[8] Adam Bates, Joe Pletcher, Tyler Nichols, Braden Hollembaek, Dave Tian, Kevin R. B. Butler, and Abdulrahman Alkhelaifi. 2014. Securing SSL Certificate Verification through Dynamic Linking. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*. ACM, 394–405.

[9] Jake Berkowsky and Thaier Hayajneh. 2017. Security issues with certificate authorities. In *8th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2017, New York City, NY, USA, October 19-21, 2017*. IEEE, 449–455.

[10] Anam Bhatti, Hamza Akram, Hafiz Muhammad Basit, Ahmed Usman Khan, Syeda Mahwish Raza, and Muhammad Bilal Naqvi. 2020. E-commerce trends during COVID-19 Pandemic. *International Journal of Future Generation Communication and Networking* 13, 2 (2020), 1449–1452.

[11] Timm Böttger, Ghida Ibrahim, and Ben Vallis. 2020. How the Internet reacted to Covid-19: A perspective from Facebook's Edge Network. In *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*. ACM, 34–41.

[12] Miguel Carvajal, José A García-Avilés, and José L González. 2012. Crowdfunding and non-profit media: The emergence of new models for public interest journalism. *Journalism practice* 6, 5-6 (2012), 638–647.

[13] Yuting Chen and Zhendong Su. 2015. Guided differential testing of certificate validation in SSL/TLS implementations. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2015, Bergamo, Italy, August 30 - September 4, 2015*. ACM, 793–804.

[14] Taejoong Chung, Yabing Liu, David R. Choffnes, Dave Levin, Bruce MacDowell Maggs, Alan Mislove, and Christo Wilson. 2016. Measuring and Applying Invalid SSL Certificates: The Silent Majority. In *Proceedings of the 2016 ACM on Internet Measurement Conference, IMC 2016, Santa Monica, CA, USA, November 14-16, 2016*. ACM, 527–541.

[15] Jeremy Clark and Paul C. van Oorschot. 2013. SoK: SSL and HTTPS: Revisiting Past Challenges and Evaluating Certificate Trust Model Enhancements. In *2013 IEEE Symposium on Security and Privacy, SP 2013, Berkeley, CA, USA, May 19-22, 2013*. IEEE Computer Society, 511–525.

[16] Rahul De', Neena Pandey, and Abhipsa Pal. 2020. Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *Int. J. Inf. Manag.* 55 (2020), 102171.

[17] Muhammed Elhadi, Ahmed Alsoufi, Ahmed Msherghi, Entisar Alshareea, Aimen Ashini, Taha Nagib, Nada Abuzid, Sanabel Abodabos, Hind Alrifai, Eman Gresea, et al. 2021. Psychological health, sleep quality, behavior, and internet use among people during the COVID-19 pandemic: a cross-sectional study. *Frontiers in psychiatry* 12 (2021).

[18] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, Oliver Hohlfeld, and Georgios Smaragdakis. 2020. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*. ACM, 1–18.

[19] Bhavna Galhotra and Ayushi Dewan. 2020. Impact of COVID-19 on digital platforms and change in E-commerce shopping trends. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 861–866.

[20] Ibrahim Ghafir, Vaclav Prenosil, Mohammad Hammoudeh, Liangxiu Han, and Umar Raza. 2017. Malicious SSL Certificate Detection: A Step Towards Advanced Persistent Threat Defence. In *Proceedings of the International Conference on Future Networks and Distributed Systems, ICFNDS 2017, Cambridge, United Kingdom, July 19-20, 2017*. ACM, 27.

[21] Kathryn Greenhill and Constance Wiebrands. 2012. No library required: the free and easy backwaters of online content sharing. *VALA 2012: eM-powering eFutures* (2012).

[22] Frank Hecker. 1999. Setting Up Shop: The Business of Open-Source Software. *IEEE Softw.* 16, 1 (1999), 45–51.

[23] Lin-Shung Huang, Alex Rice, Erling Ellingsen, and Collin Jackson. 2014. Analyzing Forged SSL Certificates in the Wild. In *2014 IEEE Symposium on Security and Privacy, SP 2014, Berkeley, CA, USA, May 18-21, 2014*. IEEE Computer Society, 83–97.

[24] Doowon Kim, Haehyun Cho, Yonghwi Kwon, Adam Doupé, Sooel Son, Gail-Joon Ahn, and Tudor Dumitras. 2021. Security Analysis on Practices of Certificate Authorities in the HTTPS Phishing Ecosystem. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*. 407–420.

[25] Doowon Kim, Bum Jun Kwon, and Tudor Dumitraş. 2017. Certified malware: Measuring breaches of trust in the windows code-signing pki. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 1435–1448.

[26] Doowon Kim, Bum Jun Kwon, Kristián Kozák, Christopher Gates, and Tudor Dumitraş. 2018. The Broken Shield: Measuring Revocation Effectiveness in the Windows code-signing PKI. In *27th USENIX Security Symposium (USENIX Security 18)*. 851–868.

[27] Julia Koch, Britta Frommeyer, and Gerhard Schewe. 2020. Online shopping motives during the COVID-19 pandemic—lessons from the crisis. *Sustainability* 12, 24 (2020), 10247.

[28] Bumjun Kwon, Sanghyun Hong, Yuseok Jeon, and Doowon Kim. 2021. Certified Malware in South Korea: A Localized Study of Breaches of Trust in Code-Signing PKI Ecosystem. In *International Conference on Information and Communications Security*. Springer, 59–77.

[29] Arturs Lavrenovs and F. Jesus Rubio Melon. 2018. HTTP security headers analysis of top one million websites. In *10th International Conference on Cyber Conflict, CyCon 2018, Tallinn, Estonia, May 29 - June 1, 2018*. IEEE, 345–370.

[30] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 674–686.

[31] Timothy Libert. 2015. Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites. *CoRR* abs/1511.00619 (2015).

[32] Ulrike Meyer and Vincent Drury. 2019. Certified Phishing: Taking a Look at Public Key Certificates of Phishing Websites. In *Fifteenth Symposium on Usable Privacy and Security, SOUPS 2019, Santa Clara, CA, USA, August 11-13, 2019*. USENIX Association.

[33] Mishari Al Mishari, Emiliano De Cristofaro, Karim M. El Defrawy, and Gene Tsudik. 2012. Harvesting SSL Certificate Data to Identify Web-Fraud. *Int. J. Netw. Secur.* 14, 6 (2012), 324–338.

[34] Galit Nimrod. 2020. Changes in internet use when coping with stress: older adults during the COVID-19 pandemic. *The American journal of geriatric psychiatry* 28, 10 (2020), 1020–1024.

[35] OpenSSL. 2022. A robust, commercial-grade, full-featured Open Source Toolkit for the Secure Sockets Layer (SSL) protocol. https://www.openssl.org/

[36] Manav Raj, Arun Sundararajan, and Calum You. 2020. COVID-19 and digital resilience: Evidence from Uber Eats. *arXiv preprint arXiv:2006.07204* (2020).

[37] Martin Roetteler, Michael Naehrig, Krysta M Svore, and Kristin Lauter. 2017. Quantum resource estimates for computing elliptic curve discrete logarithms. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 241–270.

[38] Muhammad Saad, Aminollah Khormali, and Aziz Mohaisen. 2019. Dine and Dash: Static, Dynamic, and Economic Analysis of In-Browser Cryptojacking. In *2019 APWG Symposium on Electronic Crime Research, eCrime 2019, Pittsburgh, PA, USA, November 13-15, 2019*. IEEE, 1–12.

[39] Ronald Snijder. 2010. The profits of free books: an experiment to measure the impact of open access publishing. *Learn. Publ.* 23, 4 (2010), 293–301.

[40] Yingjie Wang, Guangquan Xu, Xing Liu, Weixuan Mao, Chengxiang Si, Witold Pedrycz, and Wei Wang. 2020. Identifying vulnerabilities of SSL/TLS certificate verification in Android apps with static and dynamic analysis. *J. Syst. Softw.* 167 (2020), 110609.

[41] Liang Zhang, David R. Choffnes, Dave Levin, Tudor Dumitras, Alan Mislove, Aaron Schulman, and Christo Wilson. 2014. Analysis of SSL certificate reissues and revocations in the wake of heartbleed. In *Proceedings of the 2014 Internet Measurement Conference, IMC 2014, Vancouver, BC, Canada, November 5-7, 2014*. ACM, 489–502.