The Infrastructure Utilization of Free Contents Websites Reveal their Security Characteristics

Mohamed Alqadhi[1] and David Mohaisen[1]

University of Central Florida

Abstract. Free Content Websites (FCWs) are a significant element of the Web, and realizing their use is essential. This study analyzes FCWs worldwide by studying how they correlate with different network sizes, cloud service providers, and countries, depending on the type of content they offer. Additionally, we compare these findings with those of premium content websites (PCWs). Our analysis concluded that FCWs correlate mainly with networks of medium size, which are associated with a higher concentration of malicious websites. Moreover, we found a strong correlation between PCWs, cloud, and country hosting patterns. At the same time, some correlations were also observed concerning FCWs but with distinct patterns contrasting each other for both types. Our investigation contributes to comprehending the FCW ecosystem through correlation analysis, and the indicative results point toward controlling the potential risks caused by these sites through adequate segregation and filtering due to their concentration.

Keywords: Web security, correlation analysis, free content websites

1 Introduction

The Web has revolutionized the way users spend their time online accessing various types of content, such as books, games, music, movies, and software. For example, many game websites offer users free or paid games. Generally, websites are grouped into two groups. 1. Website content is available for a fee, and these types of websites are known as premium content websites (PCW). 2. Website content for free, where they are known as free content websites (FCWs). Previous studies [9,10] reported that the FCWs tend to be riskier than PCWs in terms of user privacy and security features [5,7,8,9,10,11,19], although a clear understanding of what contributes to this risk is unclear. Given the popularity of these websites and the associated risk [7,8,9], we set out to investigate the network characteristics and the hosting patterns for these websites, including the network size, the cloud service provider (CSP), and the hosting country. We do so to identify the correlation between the security features of those websites and their characteristics in terms of hosting patterns.

Approach. For a complete characterization of the FCW hosting infrastructure and associated patterns, we continue to pursue the following. 1. We identify the size of networks these websites use for hosting in small, medium, and large sizes. 2. We identify and investigate the cloud service providers for these websites and their characteristics. 3. We study the main hosting countries of FCWs to provide a sufficient description of their

hosting characteristics. 4. We perform a correlation analysis to distinguish the security and hosting patterns for FCWs compared to PCWs. 5. We provide a correlation analysis of the hosting patterns of FCW and PCW and their security assessment. In doing this correlation analysis, and in contrast to the PCWs, we hope to shed light on the features that contribute most to explaining such websites' security and privacy risks.

Revealing the correlations between the hosting countries with hosted FCWs and PCWs determines the appropriate action governments should take to improve hosting requirements. Revealing the correlations of malicious websites with the hosting countries will focus the efforts of governments to 1. evaluate their security standards, 2. take a step forward in implementing more stringent security standards to combat malicious websites, 3. and protect the end users by reviewing the privacy and security policies or mutual agreements that any website operating must adhere to in these countries.

Contributions. We used a data set that included 1,562 FCW and PCW obtained from the research work of Alabduljabbar *et al*. [6]. Using Pearson's correlation analysis, we examined the connections between FCW, PCW, network size, hosting CSP, and nations. We analyzed these correlations to find patterns and affinities related to various hosting arrangements. The links between website attributes, network size, hosting providers, and regional distribution are better understood due to this investigation. **(1) Full Comparison.** We provide a comprehensive understanding of the different characteristics of the FCW hosting pattern compared to PCW by studying their correlations with network size, hosting CSPs, and hosting countries. **(2) Systematic Analysis.** We provide a systematic security analysis for FCW and PCW. Analyze the correlations between FCW, PCW, and malicious or benign attributes of content categories. We study the correlations of malicious FCWs and PCWs with hosting infrastructures. **(3) Hosting Correlations** We provide a detailed discussion of different characteristics of the hosting pattern. Derived from the results of a correlation analysis between small, medium, and large sizes of the networks and malicious or benign websites. We provide correlation results of the top hosting CSPs and countries. We discuss whether a strong or weak correlation exists between hosting patterns for specific content categories or security behaviors.

Paper Organization. The rest of the paper reviews the related work 2, followed by research questions, data collection, and the analysis method described in Section 3. The results of the analysis are given in Section 4. The detailed discussion is provided in Section 5. Finally, the concluding remarks and the work summary are in Section6.

2   Related Work

This work provided a detailed correlation analysis for FCWs and PCWs with their different networking hosting patterns and security aspects. Security analysis on FCWs has been established previously by Alabduljabbar *et al*. [6,7,8,9]. The cost of using FCWs has been investigated in [17,18,20,28]. The correlation between FCW security and the use of a specific content management system has been introduced in [10], while other studies performed a correlation analysis on website security, such as [13,15,21,22,27]. Taking into consideration the number of studies and the lack of space, we concentrate solely on a subset of relevant studies to this work and their results.

Table 1: Network sizes and their characteristics. The maximum slash bit is 32 (IPv4). x represents the number of bits and y represents the number of addresses.

Size	Bits in CIDR	# Addresses
Small (SN)	$/24 < x \leq /32$	$2^8 > y \geq 2^0$
Medium (MN)	$/16 < x \leq /24$	$2^{16} > y \geq 2^8$
Large (LN)	$/8 < x \leq /16$	$2^{24} > y \geq 2^{16}$
Very Large (VLN)	$/0 < x \leq /8$	$2^{32} > y \geq 2^{24}$

Security Analysis. Zhao *et al*. [32] investigated the impacts of user-generated content (UGC) and marketer-generated content (MGC) on free content consumption by integrating the literature with research on determinants of physical exercise. Drutsa *et al*. [14] investigated the utility of new data sources to predict video popularity without reliable data from video hosting services. Vasek *et al*. [16] examined the effectiveness of sharing abuse data with web hosting providers to mitigate malicious online activities. Mirheidari *et al*. [23] devised two attacks against web servers exploiting the improper isolation between files on shared web hosting servers. Also, in *et al*. [24] outlined a comprehensive overview of common attacks on shared Web servers.

Correlation Analysis. Several works performed a correlation analysis of the website's security. Visschers *et al*. [13] explores the cost of cybercrime and its relationship to the web security posture. Mezzour *et al*. [22] examines the relationship between social and technological factors and international variations in network-based attacks and hosting. Moreover, Mekovec *et al*. [21] found how user perceptions of security and privacy impact their evaluation of online services using correlation analysis.

Domestic Analysis. Goethem *et al*. [15] presents a large-scale security analysis of 22,851 websites originating in 28 European countries. Furthermore, Raponi and Di Pietro [27] analyzed the password recovery management mechanism of Alexa's top 200 websites, with domains registered in certain European countries. They found that more than 54% of the websites in France, 36% in Italy, 47% in Spain, and 33% in the UK were vulnerable in December 2017.

3 Methodology

3.1 Research Questions

This work aims to derive insightful results of FCW correlations and different hosting patterns compared to PCW. To achieve this goal, we have worked to provide valid answers to the following questions. **RQ1**. What are the main differences between the hosting patterns (networks, hosting CSPs, and countries) of FCWs compared to PCWs? **RQ2**. What type of correlation exists between hosting patterns (networks, hosting CSPs, and countries) and malicious FCWs or PCWs? **RQ3** What are the main correlations of the hosting patterns (networks, hosting CSPs, and countries) of content websites? **RQ4**. What are the implications of FCWs hosting patterns correlation analysis?

3.2 Data Collection Process

We comprehend the research questions by examining multiple datasets. 1. a main dataset of FCWs, PCWs, and their annotations, 2. complementary dataset for augmenting the analysis of the main dataset in terms of security (maliciousness detection), 3. network size classification and, 4. hosting patterns (network, CSP, and country) annotations. In the following, we review these datasets.

Fig. 1: FCWs vs. PCWs.

Free and Premium Websites. We use the dataset of Alabduljabbar *et al*. [7,8,9]. The main criteria for selecting the sample were determined based on popularity, main language, and activities. During the collection time, all selected websites were live. Data were collected using three search engines (Bing, DuckDuckGo, and Google). The classification of content type has been applied manually, whether the website is in FCW, PCW, or (book, game, movie, music, or software) content category.

Malicious Annotation. After collecting the data, the VirusTotal [2] API has been used to determine the security of each website, which is a tool that combines more than 70 scanning engines and is available online. VirusTotal enabled us to detect malicious IPs, domains, or URLs correlated with websites. We used VirusTotal since it is the standard tool used in this domain [30,4,12,25,26,29,31] We broadened the data collected according to the VirusTotal output. Since it gives multiple detection results, we take an entity, website, or IP as malicious if at least one of the returned scan results is at least.

Hosting Patterns Annotation. We analyze the scope of the network infrastructure associated with FCWs using the IP addresses connected to each domain as a feature for analysis. We rely on two major API services–ipdata [1] and IPSHU [3]–to gather pertinent information about the given IP address. The subnet mask is used to determine the size of each website's network. Using the CIDR (Classless Inter-Domain Routing) notation, we classified: small networks (/25 - /32), medium networks (/16 - /24), large networks (/8 -/15), and (anything below /7) very large networks as in Table 1.

The IPs of FCWs and PCWs are used to determine the hosting CSPs by querying ipdata [1] and IPSHU [3]. They give the CSP name for the hosting site and its longitude and latitude to determine the hosting country for each website. After refining the websites, we found that only 1,509 (96. 6%) websites are online, as appears in Figure 1. Among the findings, 788 FCWs and 721 PCWs were grouped into five categories: books (144 free, 191 premium), games (78 free, 111 premium), movies (310 free, 152 premium), music (80 free, 86 premium), and software (176 free, 181 premium).
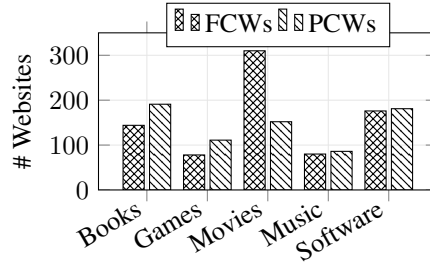
3.3 Correlation Analysis

We aim to determine whether there is a correlation between the distribution of malicious or benign FCWs and PCWs in different network sizes, CSPs, and counties. To quantify the strength of our correlation, we will use the Pearson correlation coefficient, which is calculated using the following formula: $\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$ Here, X represents free, premium, malicious, or benign attributes. On the contrary, Y represents the characteristic

being studied. The numerator of the formula represents the covariance between X and Y, while the denominator represents the product of their standard deviations.

In this study, we used the correlation analysis approach to recognize the patterns and differences between FCW and PCW, across various analysis dimensions. This study uses six main dimensions: Type of website (FCW or PCW), type of content category, maliciousness of websites, network size, CSP, and hosting country. In the following, we define each of those dimensions as appears in the workflow of this analysis in Figure 2.
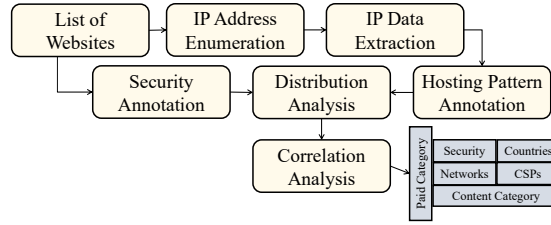

Fig. 2: The workflow of high-level representation of our data extraction.

Content Websites. This feature signifies the type of website, free (FCW) or paid (PCW), that correlates with a specific infrastructure entity (size, CSP, country), security attributes (malicious or benign) and content types. The paid feature of a website is determined by using different search engines, as described in the FCW and PCW data collection and annotation 3.2. We study the correlation of FCWs / PCWs with different infrastructure features, security features, or content categories to determine where to focus the development effort to improve the hosting of more secure FCWs or PCWs.

Content Categories. In this study, the content of the websites is categorized as (books, games, movies, music, and software). We study the correlation between content types and hosting infrastructure patterns. We study their correlation with FCWs / PCWs and their security attributes (malicious or benign). To know the weaknesses of hosting different content categories. Such as the correlation of malicious websites to a specific content type in a specific hosting infrastructure.

Security Attributes. This feature signifies the total association of malicious or benign websites with a specific infrastructure entity (size, CSP, or country). The results of the VirusTotal scan determine the maliciousness of a website as described in the security annotation process in 3.2. We study the correlation of malicious or benign websites with different types of content categories, the type of website (FCW, PCW), network sizes, CSPs, and countries. Inform hosting providers about the risks associated with FCW.

Network Size. The network size dimension represents the number of websites discovered within a specific size of the network (small, medium, large, and very large), as described in Section 3.2. We provide the results of the correlation analysis between FCWs and PCWs, the content categories of websites, and malicious associations with the size of the network. Investigating FCWs' networks is essential to know their weaknesses.

Cloud Service Provider. The CSP indicates the cloud service provider used to host FCW or PCW. We are studying the correlation between the top ten and the other 298 CSPs discovered during this study. We study the correlations of CSPs with the different

content categories and security attributes. Determine the security policies of the CSPs that need to be investigated, altered, or improved.

Country. This feature represents the hosting countries obtained from the hosting patterns annotations 3.2. We found 44 countries with a heavy-tailed distribution. We study the correlation between FCWs/PCWs and their malicious association for different content categories with the hosting countries. To know where we can make any improvements to the security agreements or rules of hosting FCWs.

4 Correlation Results

We investigate the correlations between the analysis dimensions described in Section section 3. Specifically, we examine the correlation between FCW and PCW with other characteristics, including network sizes, CSPs, and countries. We also study the malicious and benign classifications in their hosting infrastructures. Figures 3, 5, and 7, illustrate the correlation between FCW and PCW with various features of the infrastructure that indicate the five categories of content.

4.1 General Correlation Results

FCWs strongly correlate with malicious websites, and most FCWs reside in medium and small networks. PCWs are more prevalent in large networks. We also found a strong relationship between the top hosting CSPs and malicious FCWs. However, the relationship between the top countries and PCWs is more diverse. Interestingly, countries correlated with hosting FCWs are found to be more related to hosting malicious websites. The following are the most noticeable insights from the correlation analysis.

Malicious or Benign. We notice that FCWs are mostly correlated with the malicious attribute. As appears in Figure 3, a negative correlation coefficient varies between 0.13 and 0.46. The strongest correlation is found on software websites. The weakest correlation is found on movie websites. Unlike the malicious website, we found a strong correlation between PCWs and benign attributes. The highest positive correlation coefficient is 0.46 on software websites, and the lowest was discovered on movie websites.

Network Correlation. Per Figures 3 and 4, we observe a strong correlation between FCW and medium networks, especially in games, movies, and music websites. There is a weak correlation between small networks and FCWs. Compared to the large network that strongly correlates with PCWs. Noticeably, there are no correlations with very large networks. Moreover, benign attributes are correlated with large and small networks. The malicious attribute has a strong correlation with medium networks.

CSPs Correlations. The top ten CSPs strongly correlate with the FCWs. Although some CSPs show a strong correlation with premium websites, most CSPs correlate strongly with benign attributes. Only two of the top ten CSPs have a strong correlation with malicious websites. Some of the content categories in the top ten CSPs indicate a weak correlation with malicious websites, as we can see in Figures 5, and 6.

Countries Correlations. The top hosting countries strongly correlate with benign websites. Similar to the other countries. In the opposite direction, some content categories indicate a strong correlation to malicious websites in several countries. Such as the

games websites in the United States and Belgium as in Figure 8. However, the FCWs strongly correlate with the United States, Germany, Australia, France, and other countries. Especially for the categories of books, movies, and software as in Figure 7.

Furthermore, since we provided a summary of the most important findings of the results from the correlation analysis, the following will be a detailed analysis of the network size, CSPs, and countries' correlation to FCWs and PCWs.

Corr	(+)	(-)	SN	MN	LN	VLN
Books	0.02	−0.02	−0.08	−0.07	0.11	0.05
Games	0.32	−0.32	0.02	−0.2	0.21	0
Movies	0.13	−0.13	−0.01	−0.22	0.23	0
Music	0.24	−0.24	−0.01	−0.18	0.18	0
Software	0.46	−0.46	−0.13	−0.08	0.16	0
Overall	0.2	−0.2	−0.05	−0.14	0.17	0.03

Fig. 3: General networks. Red and blue indicate a vital contribution FCWs and PCWs respectively.

Corr	SN	MN	LN	VLN
Books	−0.04	0.1	−0.09	−0.04
Games	−0.12	0.22	−0.19	0
Movies	−0.03	0.08	−0.08	0
Music	−0.07	0.17	−0.16	0
Software	−0.06	0.21	−0.2	0
Overall	−0.05	0.14	−0.13	−0.02

Fig. 4: The correlation of malicious vs. benign networks. Red for malicious and green for benign.

4.2 Networks Correlations

General Networks. Figure 3 shows the relationship between FCWs and PCWs in different network sizes, indicating their correlation with various content categories. The results indicate a strong relationship between the book FCWs, predominantly hosted in small and medium networks. In contrast, large networks strongly correlate with PCWs. Unlike the book category, the games, movies and music categories show a weak correlation with small networks, which varies between FCWs and PCWs.

Figure 3 illustrates the associations between FCWs and PCWs with different networks. The correlation highlights the connections between FCWs and PCWs in different categories that use different network sizes and their malicious or benign classification. Reflecting the network correlation observed earlier in Figure 3, malicious attributes emerge as a significant factor. However, the distinction here is the pronounced association between malicious attributes and FCWs, suggesting that medium and small networks are more strongly linked to malicious factors than other websites.

Malicious Networks. The heat map shown in Figure 4 illustrates the associations between malicious websites and various characteristics of the size of the network. This correlation highlights the connections between malicious and benign websites in different categories and network sizes. For example, red indicates a high concentration of malicious websites and green represents predominantly benign content. On examination, we observe that most malicious websites are found in medium-sized networks, while small and large networks mainly consist of benign websites. Furthermore, there is a notable correlation between the malicious attribute and the category of games in medium networks, with similar patterns observed on music and software websites. On the contrary, the "Other" categories exhibit a weaker likelihood of maliciousness.

Corr	CF	AZ	LW	TR	GO	ST	LS	AK	FS	MS	Or
Books	−0.07	0.19	−0.16	−0.16	0.14	−0.14	−0.06	0.1	0.11	0.07	−0.04
Games	−0.2	0.26	−0.15	−0.09	0.12	0	−0.15	0.21	0.14	0.06	−0.07
Movies	−0.17	0.39	−0.2	−0.18	0.09	−0.16	−0.15	0.18	0.15	0.08	0.06
Music	−0.17	0.39	−0.05	−0.11	−0.01	−0.2	−0.08	0.11	0.06	0.04	−0.1
Software	−0.27	0.13	−0.18	−0.09	0.07	0	−0.17	0.13	0.07	0.16	0.2
Overall	−0.15	0.26	−0.18	−0.16	0.08	−0.15	−0.15	0.15	0.11	0.09	0.05

Fig. 5: Most used CSP's analysis. The color indication is similar to 3. The top hosting CSPs are, "Cloudflare"(CF), "Amazon"(AZ), "Liquid Web"(LW), "Trellian"(TR), "Google"(GO), "Sp-Team"(ST), "LeaseWeb"(LS), "Akamai"(AK), "Fastly"(FS), "Microsoft"(MS), Other CSPs(Or).

Corr	CF	AZ	LW	TR	GO	ST	LS	AK	FS	MS	Or
Books	0.58	−0.2	−0.02	−0.09	−0.06	−0.08	−0.03	−0.01	−0.08	0.12	−0.28
Games	0.7	−0.22	0.14	0.08	−0.13	0	0.06	−0.22	−0.15	−0.07	−0.44
Movies	0.11	−0.15	0.05	0.02	0.04	0.08	0.02	−0.05	−0.06	−0.02	−0.05
Music	0.52	−0.17	0.12	−0.07	−0.08	−0.05	−0.05	−0.07	−0.12	−0.08	−0.16
Software	0.58	−0.16	0.15	−0.01	−0.09	0	−0.04	−0.11	−0.1	−0.13	−0.38
Overall	0.48	−0.18	0.06	−0.03	−0.05	−0.01	−0.01	−0.08	−0.09	−0.04	−0.25

Fig. 6: Malicious vs. benign hosting CSPs. The colors are similar to 4.

4.3 Cloud Service Providers

General Correlations. Figure 5 illustrates the associations between FCWs and PCWs and the CSPs most commonly used in the top hosting countries. The correlation shows the connections between FCWs and PCWs over the top-used CSPs. Furthermore, we found that most PCWs are associated with CSPs that report the lowest malicious activity. We notice that FCWs are used primarily with the most malicious websites that host CSPs, as appears in Figure 6. In contrast, PCWs are used primarily with the least malicious websites hosting CSPs. Most PCWs, whose categories are games, movies, and music, are hosted by "Amazon", while "Cloudflare" hosts most FCWs of books and software FCWs. Finally, the general categories are highly distributed.

Malicious Correlations. Figure 6 shows the relationship between malicious websites and the top hosting CSPs. The correlation indicates the relation between malicious and benign websites on the most used CSPs. We notice that the highest concentration of malicious websites strongly correlates with the CSPs "Cloudflare" and "Liquid Web", while benign websites are primarily associated with the other CSPs. Interestingly, book websites exhibit a strong malicious relationship with "Microsoft" CSP, which is known to have one of the lowest reported percentages of malicious activity. The Movies also display multiple malicious correlations with the top six CSPs compared to the others.

4.4 Countries Correlation

General Correlations. Figure 7 shows the relationship between FCW and PCW in the top 10 hosting countries. The correlation indicates the relationship between FCW and PCW in the top hosting countries. For example, we noticed a strong relationship between PCWs in the movie category and the United States and fewer correlations with

Corr	US	BE	NL	DE	AU	FR	CN	GB	CA	IE	Or
Books	0.03	−0.17	0.04	−0.17	−0.13	−0.06	0.13	0.11	0.11	0.08	−0.03
Games	0.1	−0.25	0.11	0	−0.02	0.03	0.03	0.09	0.09	0.09	−0.2
Movies	0.29	−0.16	−0.07	−0.22	−0.15	−0.07	0.15	0.09	0	0.16	−0.08
Music	0.14	−0.1	−0.04	−0.25	−0.05	0.04	0.11	0.13	−0.08	0.15	−0.07
Software	0.19	−0.29	−0.08	0.04	−0.05	0.01	0.07	−0.05	0.04	0.09	0.03
Overall	0.17	−0.19	−0.03	−0.16	−0.13	−0.02	0.11	0.07	0.05	0.12	−0.04

Fig. 7: Top hosting countries (Alpha-2). The colors are similar to 3.

Corr	US	BE	NL	DE	AU	FR	CN	GB	CA	IE	Or
Books	0.11	0.12	−0.09	−0.13	−0.09	0.06	−0.01	0.02	0.02	−0.05	−0.08
Games	0.31	0.27	−0.17	−0.15	0.01	−0.04	−0.1	−0.16	−0.09	−0.09	−0.21
Movies	−0.06	0.05	0.01	0.02	0.01	0.04	−0.03	−0.04	−0.04	−0.06	0.08
Music	0.11	0.2	−0.09	0.03	−0.08	0.02	0.05	−0.08	−0.1	−0.1	−0.11
Software	−0.04	0.29	−0.08	−0.12	−0.03	0.05	−0.06	0.05	−0.11	−0.08	−0.1
Overall	0.05	0.2	−0.07	−0.07	−0.04	0.04	−0.03	−0.01	−0.05	−0.07	−0.06

Fig. 8: Malicious vs. benign hosting countries (Alpha-2). Colors are similar to 4.

other categories. Moreover, we observed that most of the top hosting countries exhibit strong relationships with FCWs, especially those reported to be highly malicious. For example, countries such as China, the UK, Canada, and Ireland show weak relationships with PCWs, but surprisingly, it is more vital than their relationship with the FCWs. Furthermore, we noticed a high concentration of FCWs in the game and software categories in Belgium, which are reported to be the most malicious websites. Simultaneously, a strong association can be observed between FCWs in movies and music categories and Germany, which is reported to have a low level of malicious activity.

Malicious Correlations. Figure 8 shows the correlation between malicious and benign websites with the top ten hosting countries. The correlation highlights the relationship between malicious and benign websites in the top hosting countries. The heat map reveals a strong relationship between the United States and Belgium that hosts malicious websites, particularly on books, games, and software websites. Most other countries have a strong connection to benign websites.

5 Results Discussion

In this section, we will discuss the key insights of the correlation analysis. Highlighting the answers to the research questions. We will list the challenges we encountered during the study. Finally, we will shed light on the limitations and recommendations.

5.1 Results Takeaway

To sum up the key results of the correlation analysis, we will highlight the insights that provide detailed answers to the research questions as follows.

Free or Premium. The correlation analysis results answer **RQ1**.

1. We notice the difference in the hosting patterns of FCWs and PCWs in their common network size, the top CSPs, and most of the hosting countries. 2. FCWs have

a weak correlation to small networks, whereas PCWs have no correlation to small networks. 3. FCWs have a strong correlation with "Cloudflare", "Liquid Web", "Trilian", "SP-Team", and "LeaseWeb" CSPs. Although PCWs seem to have a strong correlation with the other top ten hosting CSPs. 4. Some of the top hosting countries show a strong correlation with FCWs. On the contrary, the other top hosting countries have a strong correlation with PCWs. 5. The results of FCWs depict certain hosting patterns that are uniquely different from PCWs. Indicating the differences in their security behavior.

Malicious or Benign. The results also provide answers to **RQ2** where we find the network hosting patterns for malicious websites. 1. Malicious websites show a strong association with FCW, while benign websites strongly correlate with PCW. 2. In general, malicious websites have a strong correlation with the medium size of the networks. The benign websites are strongly correlated with large networks and weakly with small networks. 3. The top ten CSPs and the other hosting CSPs show a significant correlation to benign websites. Some of the top ten CSPs have a strong correlation with hosting malicious websites. 4. Hosting countries seem to have a significant correlation with benign websites. In the opposite direction, some of the content categories exhibit a strong correlation with malicious attributes. Especially in the top two hosting countries.

Hosting Patterns. The results of studying the different categories of website content in the different hosting patterns give significant answers to **RQ3**. 1. Different content categories show a different level of correlation to malicious and benign attributes. Games and software websites exhibit a strong correlation with malicious FCWs. 2. Small networks have a weak correlation with all FCW content categories. Medium networks have a strong correlation with FCW games, movies, and music websites. Large networks show a lower correlation with such categories in PCWs. 3. We notice the differences in the correlation with the top hosting CSPs. We found a strong correlation between all content categories and FCWs in the top hosting CSPs. Other top CSPs strongly correlate with PCWs. Such as "Amazon", "Akamai", "Fastly", and "Microsoft". 4. There is a weak correlation between all categories of PCWs and some of the top hosting countries. In contrast, we found a strong correlation of FCWs and categories hosted in "Belgium".

Results Implications. The implications of the previous findings provide answers to **RQ4**. 1. Isolation of FCWs that use small networks may be considered an applicable solution to mitigate FCW and PCW risks. 2. Addressing malicious environments within CSPs is one of the most effective solutions to reduce risk exposure. 3. Taking legal action to force such CSPs to improve their security could be a viable solution to secure the network. 4. FCWs and PCWs are concentrated in medium networks, the same as malicious content websites. This implies the need for a better solution than isolating these networks. 5. Games and software content are the most correlated with malicious websites. This implies the serious need to develop security scanning tools specialized in detecting malicious code that may be injected into software or game FCWs.

5.2 Limitations and Recommendations

Limitations. Initially, our main data set consisted of 1,562 FCW and PCW. However, after the network annotation process, we found that only 1,509 websites were operating, suggesting a decrease over time. Thus, longitudinal analysis is required to gain insight

into changes in an operation performed on these websites. The top hosting CSPs discovered during this study are widely spread. Where some of these CSPs have different companies, we combined all of the companies of the same entity into one CSP. For example, "Amazon" CSPs provide their services regionally, such as Amazon Data Services Canada and Amazon Data Services France. Consequently, all these CSPs were aggregated into one entity "Amazon". For further analysis of their service distribution, it is imperative to conduct further investigation to ensure their security.

Recommendations. Based on the findings, our recommendations to system administrators are to apply stronger security protocols. To protect their networks from malicious activities. In particular, organizations must prioritize segmenting medium-sized networks as they are often malicious. Moreover, analyzing the CSPs used by FCWs and PCWs can aid in determining which CSPs have a higher number of malicious websites than good ones. This indicates where legal action need to be considered if necessary. General observation suggests that improvements should be made by developing these aspects, reducing malicious websites, and strengthening overall network security.

6 Conclusion

The correlations between FCWs, PCWs and their hosting habits in network size, CSP, and hosting countries have been revealed by this research. Our investigation has shown a significant association between FCWs and medium networks, suggesting that these networks tend to host malicious websites. Additionally, we have identified some CSPs that may need to increase their security requirements, as they are significantly correlated with hosting more malicious content categories. Furthermore, our research shows a notable association between the nations where FCWs are hosted, pointing to the need for more stringent laws and other countermeasures to address dangerous websites.

References

1. —: Reliable IP ddress Data (2022), last access December 14, 2022
2. —: Analyze suspicious files and URLs to detect types of malware automatically (2022), last access December 14, 2022
3. —: IP Address Lookup Tools (2023), last access January 19, 2023
4. Adeniran, A., Mohaisen, D.: Measuring cryptocurrency mining in public cloud services: A security perspective. In: CSoNet. pp. 128–140. Springer-Verlag, Berlin, Heidelberg (2023)
5. Akhawe, D., Barth, A., Lam, P.E., Mitchell, J.C., Song, D.: Towards a Formal Foundation of Web Security. In: Proceedings of the 23rd IEEE Computer Security Foundations Symposium, CSF. pp. 290–304 (2010)
6. Alabduljabbar, A., Abusnaina, A., Meteriz-Yıldıran, Ü., Mohaisen, D.: TLDR: Deep Learning-Based Automated Privacy Policy Annotation with Key Policy Highlights. In: ACM WPES. pp. 103–118 (2021)
7. Alabduljabbar, A., Ma, R., Alshamrani, S., Jang, R., Chen, S., Mohaisen, D.: Poster: Measuring and assessing the risks of free content websites. In: NDSS (2022)
8. Alabduljabbar, A., Ma, R., Choi, S., Jang, R., Chen, S., Mohaisen, D.: Understanding the Security of Free Content Websites by Analyzing their SSL Certificates: A Comparative Study. In: CySSS@AsiaCCS. pp. 19–25 (2022)

9. Alabduljabbar, A., Mohaisen, D.: Measuring the Privacy Dimension of Free Content Websites through Automated Privacy Policy Analysis and Annotation. In: Companion of The Web Conference, WWW. pp. 860–867 (2022)
10. Alaqdhi, M., Alabduljabbar, A., Thomas, K., Salem, S., Nyang, D., Mohaisen, D.: Do Content Management Systems Impact the Security of Free Content Websites? A Correlation Analysis. In: CSoNet (2022)
11. Alrawi, O., Mohaisen, A.: Chains of Distrust: Towards Understanding Certificates Used for Signing Malicious Applications. In: Proceedings of the 25th International Conference on World Wide Web,(WWW). pp. 451–456 (2016)
12. Baek, S., Jung, Y., Mohaisen, A., Lee, S., Nyang, D.: Ssd-insider: Internal defense of solid-state drive against ransomware with perfect data recovery. In: 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS). pp. 875–884. IEEE (2018)
13. Chen, P., Visschers, J., Verstraete, C., Paoli, L., Huygens, C., Desmet, L., Joosen, W.: The relationship between the cost of cybercrime and web security posture: a case study on Belgian companies. In: ACM ECSA. pp. 115–120 (2017)
14. Drutsa, A., Gusev, G., Serdyukov, P.: Prediction of Video Popularity in the Absence of Reliable Data from Video Hosting Services: Utility of Traces Left by Users on the Web. CoRR **abs/1611.09083** (2016)
15. van Goethem, T., Chen, P., Nikiforakis, N., Desmet, L., Joosen, W.: Large-Scale Security Analysis of the Web: Challenges and Findings. In: Springer TRUST. pp. 110–126 (2014)
16. Marie Vasek and Matthew Weeden and Tyler Moore: Measuring the impact of sharing abuse data with web hosting providers. In: Proceedings of the Workshop on Information Sharing and Collaborative Security, ACM. pp. 71–80 (2016)
17. Hu, T., Tripathi, A.K.: Is There a Free Lunch? Examining the Value of Free Content on Equity Review Platforms. In: 16th Workshop on e-Business Digital Transformation: Challenges and Opportunities. pp. 79–86 (2017)
18. Hu, T., Tripathi, A.K., Berkman, H.: The Value of Free Content on Social Media: Evidence from Equity Research Platforms. In: 18th Workshop on e-Business Digital Transformation: Challenges and Opportunities. vol. 403, pp. 123–130 (2019)
19. Kosba, A.E., Mohaisen, A., West, A.G., Tonn, T., Kim, H.K.: ADAM: Automated Detection and Attribution of Malicious Webpages. In: Proceedings of the 15th International Workshop on Information Security Applications, WISA. pp. 3–16 (2014)
20. Lee, D., Nam, K., Han, I., Cho, K.: From free to fee: Monetizing digital content through expected utility-based recommender systems. Inf. Manag. **59**(6), 103681 (2022)
21. Mekovec, R., Hutinski, Z.: The role of perceived privacy and perceived security in online market. In: IEEE MIPRO. pp. 1549–1554 (2012)
22. Mezzour, G., Carley, K.M., Carley, L.R.: Global Variation in Attack Encounters and Hosting. In: ACM Proceedings of the Hot Topics in Science of Security. pp. 62–73 (2017)
23. Mirheidari, S.A., Arshad, S., Khoshkdahan, S., Jalili, R.: Two novel server-side attacks against log file in Shared Web Hosting servers. In: Proceedings of The 7th International Conference for Internet Technology and Secured Transactions, ICITST IEEE. pp. 318–323 (2012)
24. Mirheidari, S.A., Arshad, S., Khoshkdahan, S., Jalili, R.: A Comprehensive Approach to Abusing Locality in Shared Web Hosting Servers. CoRR **abs/1811.00922** (2018)
25. Mohaisen, A.: Towards automatic and lightweight detection and classification of malicious web contents. In: Third IEEE Workshop on Hot Topics in Web Systems and Technologies, HotWeb. pp. 67–72. IEEE Computer Society (2015). https://doi.org/10.1109/HOTWEB.2015.20, https://doi.org/10.1109/HotWeb.2015.20
26. Mohaisen, A., Alrawi, O., Mohaisen, M.: AMAL: high-fidelity, behavior-based automated malware analysis and classification. Comput. Secur. **52**, 251–266 (2015)

27. Raponi, S., Pietro, R.D.: A Longitudinal Study on Web-Sites Password Management (in)Security: Evidence and Remedies. IEEE Access pp. 52075–52090 (2020)
28. Roy, S.S., Karanjit, U., Nilizadeh, S.: A Large-Scale Analysis of Phishing Websites Hosted on Free Web Hosting Domains. CoRR **abs/2212.02563** (2022)
29. Saad, M., Khormali, A., Mohaisen, A.: Dine and dash: Static, dynamic, and economic analysis of in-browser cryptojacking. In: APWG eCrime. pp. 1–12 (2019)
30. Thomas, M., Mohaisen, A.: Kindred domains: detecting and clustering botnet domains using DNS traffic. In: 23rd International World Wide Web Conference, WWW. pp. 707–712. ACM (2014). https://doi.org/10.1145/2567948.2579359, https://doi.org/10.1145/2567948.2579359
31. Wang, A., Mohaisen, A., Chang, W., Chen, S.: Delving into internet ddos attacks by botnets: Characterization and analysis. In: 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2015, Rio de Janeiro, Brazil, June 22-25, 2015. pp. 379–390. IEEE Computer Society (2015). https://doi.org/10.1109/DSN.2015.47, https://doi.org/10.1109/DSN.2015.47
32. Zhao, K., Zhang, P., Lee, H.: Understanding the impacts of user- and marketer-generated content on free digital content consumption. Decis. Support Syst. **154**(154), 113684 (2022)