Automated Collection of Redirect Chains from Search Results Pages to Prize Scam Pages

Mizuki Takagi[1*], Konan Nagashima[1], Kazunari Mitani[1], Masatake Shirai[1], Taiichi Saito[1]
[1] Tokyo Denki University, 5, Senju Asahicho, Adachiku, 120-8551, Japan

* Corresponding author: Mizuki Takagi; email: 22kmc16@ms.dendai.ac.jp
Manuscript submitted August 15, 2022; revised September 8, 2022; accepted October 15, 2022.
doi: 10.17706/jsw.18.1.31-43

Abstract: Currently, there is an increasing number of prize scams in which users are tricked into entering their personal information to claim that they have won a smartphone or other device. In addition, the scams are not only led by email or SMS but also by Google searches in some cases. Users who access a linked page (Entrance page) from a search result are redirected to a scam page through several intermediate pages. In SCIS2021 and SCIS2022, we collected and analyzed redirection chain logs to reveal the reality of such a prize scam. First, in SCIS2021, we surveyed to gather records of redirect chains. However, we could conduct only 99 small surveys due to manual collection. Next, in SCIS2022, we obtained 496 redirect chain logs through an automatic display of redirect chain logs but could not implement automatic log determination. Therefore, this paper proposes a tool to automatically collect and determine the records of redirection chains using Power Automate Desktop. Furthermore, we will disclose the scam methods and sound a warning by highlighting such scam campaigns.

Keywords: Power Automate Desktop, Prize scam, Redirect Chain

1. Introduction

Currently, there are increasingly more reports of prize scams that deceptively offer a chance to win a smartphone and then charge for unintended and unrelated services [1-2]. These prize scam pages offer offering "a chance to buy an iPhone at a discount". The Council of Anti-Phishing Japan reports that by entering credit card information users unintentionally sign up for a service that charges a monthly fee. Other organizations such as Trend Micro [3] and PCrisk [4] have reported similar problems. These facts show that prize scams are a worldwide problem.

The Japan Cybercrime Center reported that users are often directed to scam pages when using search engines [5]. JPCERT/CC has published the domain names related to the prize scam on GitHub from June 30, 2021, to the present [6]. We also investigated the prize scams beginning with search engine results and reported our findings in SCIS 2021 and SCIS 2022 [7-8].

In SCIS2021, we presented the overall picture of the prize scam by examining page transitions from links in search results to the scam site.

When a user accesses a seemingly harmless link (a link to an entrance page) in the search results of Google search, a prize scam is triggered. After briefly displaying the entrance page, the browser repeatedly redirects multiple times until it finally reaches the prize scam page. Such page transitions through successive redirects are called a redirect chain.

Fig. 1 shows a redirect chain to a prize scam page starting from a link in the search results, consisting of the following three parts.

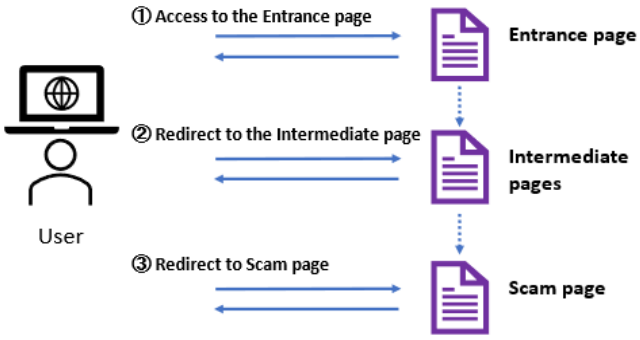

Fig. 1. The flow of winning scams starting from links in search results

Entrance page - is the starting point of the prize scam that appears in search results of Google and other search engines. When users access this page, they are redirected to an intermediate page.

Intermediate pages - are pages in the redirect chain other than the entrance page or the scam page.

Scam page - displays that the user has won a chance to buy an iPhone, and eventually induces the user to enter credit card information.

In this paper, we present the prize scam tricks identified in our previous research, and aim to automatically collect redirect chain logs and conduct a large-scale survey of more than 5,000 cases.

We are also working on reporting to Google and JPCERT/CC the domain names of entrance pages, intermediate pages, and scam pages included in the collected redirect chain logs.

To the best of our knowledge, this is the first comprehensive study of the redirect chains from search results to scam pages. Our main contributions can be summarized as follows

1. We investigated the common structure of sites that provide entrance pages, the use of black hat SEO, and redirection pattern chains, in prize scam that starts with links in search results.

2. We proposed a system to automatically collect redirect chains and found 976 entrance pages, 526 intermediate pages, and 970 scam pages.

3. We showed that the prize scam is not a temporary campaign, but a large-scale campaign over a long period of time.

The remainder of this paper is organized as follows: Session 2 present related studies. Session 3 describes the background of this research and the characteristics of each flow of the redirection chain. Session 4 presents the results of previous experiments. Session 5 describes the proposed method in the automatic collection system. Session 6 reports on the results obtained from the experiment. Sessions 6 discusses a discussion of the results acquired and prospects. Finally, Sessions 7 present related studies and briefly summarize the results of this paper.

2. Related Works

Phishing has been studied extensively due to the risk of identity theft. The system proposed in this paper aims to detect prize draw scam pages.

As a study of redirects associated with scam sites, a survey on investigative techniques for redirecting campaigns from fake news sites to scam pages have been reported [9]. Chen et al. proposed methods for collecting redirect paths, cascading nodes, and searching for other domain names hosted with similar IP addresses.

Other studies related to redirection chains in prize draw scams include research studies on malicious redirection and research cloaking [10]. Gianluca et al. investigated how various browsers reach scam pages. They collected redirects to scam pages and analyzed the characteristics of redirects by creating a redirect graph. Furthermore, they propose a detection method for scam pages and implement and evaluate the system as SpiderWeb [11]. Wang et al. investigated how cloaking affects search engines. They also conducted a study on the validity period of cloaked search results. They pointed out that retaining the pages for an extended period is possible. These studies suggest that malicious sites are using redirects and cloaking to avoid detection.

For the prize draw scams investigated in this study, redirects from Google search results were used for inducements, and there is a significant link to SEO poisoning [12]. Leonidas et al. collected data for about four years and reported on changes in SEO poisoning. Host responses to search redirect attacks have also been investigated, showing a link between SEO poisoning and phishing [13]. Liao et al. have researched long-tail SEO. The research shows the effectiveness of long-tail SEO spam and network characteristics of long-tail SEO campaigns.

3. Background

In this session, we present the background information on the redirect chain consisting of entrance pages, intermediate pages, and scam pages.

3.1. Entrance Page

An entrance page is the starting point of a prize scam. When we access a link in search results, the entrance page is displayed for a short time and then we are redirected to an intermediate page. We describe the features of the entrance page.

Entrance Page SEO and Structure - SEO(Search Engine Optimization) is a technique for improving web pages so that they appear at higher rankings in search results. An entrance page has features that abuse this SEO. While Google search results page displays multiple items consisting of URL, page title, and page summary, entrance pages also employ a page structure such that it has multiple page titles and page summaries like Google's search results page. In addition, the site that has an entrance page also hosts many other entrance pages and the pages have links to each other. They are considered to be SEO techniques to retrieve page titles and page summaries from the top of the actual Google search results and to increase the number of links between the entrance pages.

Entrance Page Cloaking - Cloaking is a technique that changes the behavior of a Web page according to the user's environment. We observed the following two cloaking patterns.

1. Normally, no redirection occurs, but it does occur when Referer header in the request is a search engine.
2. Normally, a redirection occurs, but if User-Agent header in the request is Googlebot, it does not occur.

Referer header indicates the URL of the page from which the request came. User-Agent header is a string that identifies the browse [14].

Since site administrators need not search the sites they manage and directly access them, the Referrer is not attached. These cloaking patterns can prevent site administrators and site search engines from finding occurrences of redirections.

3.2. Intermediate Page

Intermediate page is a page in a redirect chain other than the entrance page and the scam page. There may be multiple intermediate pages in a redirect chain, and redirections are performed among them. We analyzed the domain names of the intermediate pages and then identified four redirection patterns on intermediate pages. The behavior of each redirection pattern of redirects is shown below.

In Pattern 1, the user is redirected from the entrance page to the scam page via one intermediate page. Fig. 2 shows an example of Pattern 1.

We here let the domain of the intermediate page be intermediate.biz. When a user accesses the entrance page, the user is redirected to intermediate.biz and finally to the scam page.

In Pattern 2, the user is redirected from the entrance page to the scam page via two intermediate pages. Fig. 2 shows an example of Pattern 2.

We here let the domain names of the two intermediate pages be intermediate1.biz and intermediate2.biz. When a user accesses the entrance page, the user is first redirected to intermediate1.biz and then to intermediate2.biz, and finally to the scam page.

In Pattern 3, a user is redirected from an entrance page to a first intermediate page, then to another intermediate page of a subdomain of the first one. After that, redirections are repeated among intermediate pages of sibling domain names of the first one. Then the user is redirected to an intermediate page of a different domain and finally to a scam page. Fig. 2 shows an example of Pattern 3.

We here let the domain of the first intermediate page be intermediate1.biz and the domain of the last intermediate page be intermediate2.biz. When a user accesses the entrance page, the user is first redirected to intermediate1.biz. Next, the user is redirected to 0.intermediate1.biz, a subdomain of intermediate1.biz. Next, the user is redirected to its sibling domain, 1.intermediate1.biz. The redirections among sibling domain names are repeatedly performed. The prefixes of sibling domain names are sequential numbers (0,1,2...) or alphabetic characters (a,b,c...) . Finally, the user is redirected to intermediate2.biz and then to the scam page.

In Pattern 4, a user is redirected from an entrance page to a first intermediate page, then to an intermediate page of a subdomain of the first one. After that, redirections are repeated among intermediate pages of sibling domain names. Then the user is redirected to the other, then to an intermediate page of a subdomain of the other. After that, redirections are repeated among intermediate pages of sibling domain names. Finally, to a scam page. Fig. 2 shows an example of Pattern 4.

We here let the domain names of the three intermediate pages be intermediate1.biz, intermediate2.biz, and intermediate3.biz, respectively. When a user accesses the entrance page, the user is first redirected to intermediate1.biz. Next, the user is redirected to 0.intermediate1.biz, a subdomain of intermediate.biz. Next, the user is redirected to its sibling domain, 1.intermediate1.biz. The redirections among sibling domain names are repeatedly performed. Next, the user is redirected to intermediate2.biz. It also redirects in the same method as described above. As in Pattern 3, contiguous characters are used for the sibling domain names. Finally, the user is redirected to intermediate3.biz and then to the scam page.

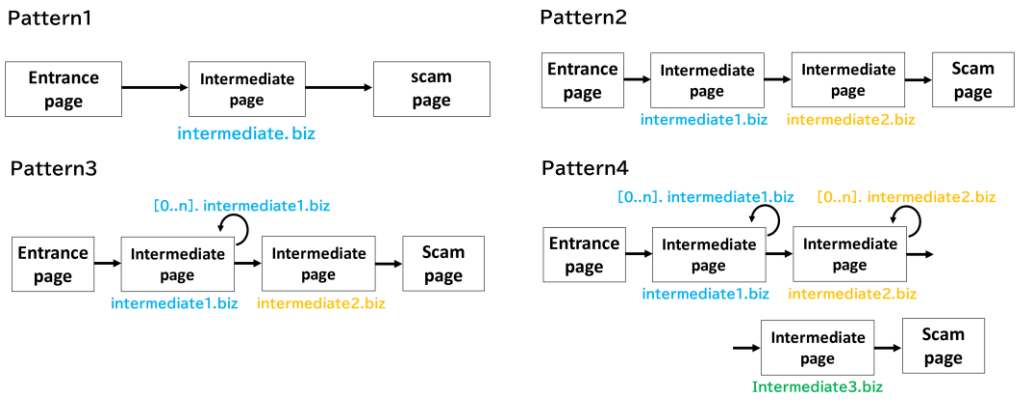

Fig. 2. Example of Each Pattern operation

3.3. Scam Page

The scam page displays that the user has won an iPhone and eventually tries to the user to enter the user to enter credit card information. The scam pages also displayed the common design page. These scam pages work as follows.

- The scam site displays a page says, “you have won an opportunity to get an iPhone at a discount”.
- The scam site displays a page that offers you a discount on iPhones if you participate in a survey or game.
- The scam site displays messages congratulating other users for getting an iPhone.
- A single HTML file performs all of these actions by rewriting the pages by jQuery.

When a user clicks on a button on the scam page to purchase an iPhone at a discounted price, the user is prompted for credit card information after an intermediate page redirects several times.

Google provides Google Safe Browsing as a countermeasure against prize scams [15]. It blocks URLs that are determined to be malicious. Attackers appears to be attempting to circumvent it by changing the domain name of the scam page within a short period of time.

4. Automatic Redirect Chain Collection

To investigate redirect chains in the prize scams, we obtained the browser output HAR files and analyzed them in. The HAR file records the redirections of the browser.

However, manual investigation was inefficient, so we developed an automated system collecting redirect chains.

This section describes our automated system, which collects the redirect chain logs such as:

- A link to the entrance page appears in search results.
- Immediately after the entrance page is displayed, the browser is redirected to several intermediate pages and finally to a scam page.

4.1. Development Environment for Automated System

In this section, we describe the development environment for our system. We used Power Automate Desktop (PAD), which is a Robotic Process Automation (RPA) system released by Microsoft Corporation [16]. It is a no-code development environment and provides components for controlling the browser and UI and so on. These components are called action. The user can automate various procedures by combining actions.

4.2. Automatic Collection System in Operation

This section describes the behavior of the automated system.
First, we divide the automation behavior into the following steps and describe the behavior of the automated system at each step.

Collect URLs from Search Result Pages - This step takes a search phrase list as input, and collects the URLs from Google search result pages. The search phrase list is a list constructed from the phrases contained on an entrance page. When a search phrase list is inputted, the automated system behaves as follows.
First, the automated system starts the browser; the browser displays the Google search page. The automated system sends a phrase in the search phrase list to the search engine, displaying the search result page.

To obtain the URL, the automated system executes the script shown in Fig. 3. After, the script shown in Fig. 4 is executed to acquire the page titles.

```
document.querySelectorAll('div.yuRUbf > a');
```

Fig. 3 JavaScript to retrieve a URL

```
document.querySelectorAll('div.yuRUbf > a > h3');
```

Fig. 4 JavaScript to retrieve the page title

The scripts are executed using the action in PDA, "Execute JavaScript function on Web page." The automated system stores, the current search phrase, page title, and URL into a list. After processing the first page, the automated system displays the following search results and repeats collecting URLs and page titles. The browser is closed if the number of URLs collected exceeds 100 or the "Next" button is not displayed. The

automated system repeats the above until the search phrase list is completed. The output result of this step is a list consisting of three items: search phrase, page title, and URL.

Obtain Redirect Chain Logs - In this step, the list obtained in the previous step is used as input. The behavior is as follows. First, the automated system launches a browser and accesses one of the listed URLs by setting the referer header [9] to https://www.google.com. The reason for going through the Google search page is to avoid the cloaking of an entrance page. After accessing the URL, the automated system records the screenshot, the HTML, the results of the dig command, and the URL of the page and waits until the next redirection occurs. When the next redirection occurs, the automated system saves the redirect chain logs and continues to record information on the redirected page. Conversely, if no redirection occurs and a certain amount of time elapses, the browser exits. And the automated system saves nothing. After the browser exits, the automated system obtains and stores the IP address of the final URL obtained by the dig command.

The automated system repeats the above until the end of the URL list. The output result of this step is a log folder with the URL, screenshot, HTML, and IP address of the Web page displayed during the redirection.

Determine if the Scam Page was Reached – The automated system uses the log folder obtained in the previous step and determines whether the page in the log folder is a scam page. It judges that the browser has reached a scam page if the HTML log of the final page contains the following specific words:

- Visitor surveys
- Membership giveaway
- Search contests
- Promotional contests

The PowerShell runs a script that uses Select-String cmdlet to search the HTML log of the final page for the aforementioned specific words. Select-String is a cmdlet that matches to search for text patterns in strings and files [17] and can be used as "Select-String C:¥test.txt -Pattern {specific word}". If the HTML log contains any of the specific words, the script determines the user has reached a scam page. Finally, the script saves only HTML logs that are determined to have reached a scam page.

5. Result

In this section, we present the results of the collection performed by our developed automated system. We collected the redirect chain logs from June 11, 2022, to June 16, 2022.

First, we inputted a search phrase list in Google search and retrieved URLs from Google search result pages. The URL retrieved may display an entrance page. Then we combine a retrieved URL with a search command “site:” and search other URLs in the site that hosts the page of the retrieved URL. If the combined URL is of an entrance page, newly obtained URLs may also be of other entrance pages. Finally, we used the automated system to collect 5,659 URLs.

Next, we used the automated system to collect logs from the list of 5,659 URLs. It accesses the collected URLs and saves redirection chain logs when the redirection occurs. We inputted 5,659 URLs into the automated system and then obtained 2,742 redirect chain logs.

Finally, we used the automated system to determine if the browser reached a scam page for each log in the obtained 2,742 redirect chain logs. The method that determines whether the user has reached a scam page is if the HTML log contains any of the specific words. We input the obtained 2,742 logs into the automated system, and the automated system determined that 970 logs reached scam pages.

In the following sections, we will describe the results in the following order: entrance page, intermediate page, and scam page.

5.1. Results of This Study

We investigated the domain names of entrance pages, intermediate pages, and scam pages in 2,742 logs obtained by the automated system. We also examined the ownership of IP addresses of the domain names using DB-IP for our survey.

Entrance Page Results - We found that the obtained 2,742 logs contain 976 entrance page URLs. We also dropped duplicate domain names from the 976 URLs, resulting in 303 domain names. Furthermore, we

excluded duplicates from the 303 IP addresses associated with the domain names, leaving us with only 33 IP addresses.

We investigated the owners of these 33 IP addresses and their details. The results are shown in Table 1.

Intermediate Page Results - We found that the obtained 2,742 logs contain 526 intermediate page URLs. We also dropped duplicate domain names from these 526 URLs, resulting in five domain names. Furthermore, we excluded duplicates from the IP addresses associated with the 21 domain names, resulting in only 19 IP addresses.

We investigated the owners of these 19 IP addresses and details. The results are shown in Table 2.

Scam Page Results - We found that the obtained 2,742 logs contain 970 scam page URLs. We also dropped duplicate domain names from these 970 URLs, resulting in 604 domain names. Furthermore, we excluded duplicates from the IP addresses associated with the 604 domain names, resulting in only 33 IP addresses.

We investigated the owners of these 33 IP addresses and details. The results are shown in Table 3.

As a result, 29 of the 33 IP addresses were found to be in the range of 5.189.217.0/24 and three in the range of 103.253.43.0/24, and one in 141.95.108.187. 5.189.217.0/24 was an IP address owned by a hosting company named Fast Content Delivery LTD. 103.253.43.0/24 was owned by a hosting company named Tele Asia Limited. 141.95.108.187 was an IP address owned by a hosting company named OVH GmbH.

Table 1. IP Address Owners and Details Entrance Page

IP address owner	Number of IP addresses owned
Cloudflare, Inc.	129
Scalaxy B.V.	9
Relevate	5
Relcom Bn	2
Chernyshov Aleksandr Aleksandrovich	2
LLC Vpsville	2
Dataline Ltd	1
DigitalOcean, LLC	1
Fast Content Delivery LTD EU block	1
Ihor Hosting LLC	1
Jsc Iot	1
Melbikomas UAB	1
Trader soft LLC	1
vds	1

Table 2. IP address owners and details intermediate pages

IP address owner	Number of IP addresses owned
Cloudflare, Inc.	9
Fast Content Delivery LTD EU block	7
Hetzner Online GmbH	3

Table 3. IP address owners and details scam pages

IP address owner	Number of IP addresses owned
Fast Content Delivery LTD EU block	29
Tele Asia Limited	3
OVH GmbH	1

6. Discussion

In this section, we discuss the state of the prize scam campaigns we have investigated, an evaluation of the automated system we have developed, countermeasure, and future efforts to address issues.

First, we discuss the status of prize scam campaigns by comparing our previous results with this study's results (See Appendix for our previous results). We see in our current results that the number of domain names and IP addresses excluding duplicates is small compared to the number of URLs in the log. Likewise, our previous results similarly excluding duplicates also has shown small compared to the number of URLs in the log. We saw similar facts in the previous results. From these observations, we suspect that the prize scam campaign is being conducted by a small number of people. In our previous results, we observed redirection patterns of multiple intermediate pages. However, we observed a drastic decrease in the number of intermediate pages observed in this study compared to our previous results. The number of pages in this study is 526 intermediate pages, compared to 4,854 intermediate pages in the results of SCIS2022 in Appendix. In addition, we observed several redirection patterns in previous studies. But most of the redirection patterns we observed in this study consist of one or two intermediate pages which are immediately redirected to the scam page. We observed that the page appearance of intermediate pages and scam pages remained unchanged. From these observations and the above suspicion that a small number of people manage a prize scam campaign, we suspect that the attacker changed patterns to reduce operating costs. Furthermore, the most important point to focus on in this study is the IP addresses of scam pages that we collect in our previous studies and this study. As a result, both are IP addresses owned by a hosting company called Fast Content Delivery LTD EU block. Therefore, we consider that the current prize scam campaign is still ongoing from 2020.

Next, we discuss countermeasures against prize scams. First, we thought filtering these IP addresses would be the most straightforward countermeasure because the IP addresses of the aforementioned scam pages were from the same hosting company. However, filtering using IP addresses has the possibility of over-filtering such that a legitimate web site is also erroneously blocked when the domain name of the site is assigned to the same IP address. Other methods, such as domain name filtering, can also be easily circumvented. Therefore, we consider that effective action is to detect and stop redirects to scam pages. The method is to use the contents of the page. In Section II described, we described the structure of the redirect chain such that intermediate pages and scam pages have a common page design essentially, even for different domain names. Therefore, we consider the possibility that these pages load common external resources, and that monitoring the loading of such external resources allows us to detect and stop redirects to scam pages.

Finally, we discuss the automated redirect chain automated system we developed. First, we only collected 99 cases because we collected them manually in our previously published SCIS2021. On the other hand, we collected 3,310 cases in this study. Therefore, we consider this automated system to be a useful system that collects more efficiently than manual collection. However, we still have to manually analyze the collected IP addresses and domain names in this study as in previous studies, which takes longer than the 496 logs collected in SCIS2022.

In addition, two problems arose when we developed our automated system. The first is reCAPTCHA when searching. reCAPTCHA is a service from Google that helps protect websites from spam and abuse. The main purpose of a reCAPTCHA is to block spambots. The problem was that when we collected URLs from search result pages, a reCAPTCHA occurred, which had not occurred in previous studies. The reason for the reCAPTCHA was that the automated system was determined to be a bot because the search speed of the automated system was too fast. We attempted to avoid this problem by having the automated system perform a delay process during the search. We have succeeded in avoiding this problem by setting the delay time at search to about 3 minutes. The second is to speed up the redirection of intermediate pages. The problem was that the redirection was faster than the automated system collected. Therefore, some redirect chain logs could not be obtained properly. This is another problem that we did not observe in our previous study. Therefore, we obtained the redirect chain logs pages that redirect fast, by setting the execution delay to 20 milliseconds. Based on the above, our future tasks are to automate the analysis of the collected data and to stabilize the automated system.

7. Conclusion

In this paper, we observed the redirect chain in the prize scam, in which a browser briefly displays the entrance, redirects multiple times displaying the intermediate pages and finally reaches the prize scam page. We observed several redirection patterns in previous studies. But most of the redirection patterns we observed in this study were one or two intermediate pages immediately redirected to the scam page. However, we observed that intermediate pages and scam pages design remained unchanged.

Next, we developed an automated system that automatically collects redirection chain logs performed by prize scams. The automated system automatically collected the redirect chain logs from the entrance page, intermediate pages, and scam pages in the prize scam. We also analyzed the redirect chain logs from our collection. Specifically, we obtained 2,472 redirect chain logs from the 5,659 URLs collected. The logs obtained consisted of 976 entrance pages, 526 intermediate pages, and 970 scam pages. Comparing our previous results with this study's results, both many more pages are related to the prize scam, but the number of domain names and IP addresses excluding duplicates is small compared to the number of URLs in the log. In addition, we observed a decrease in the number of intermediate pages observed in this study compared to our previous results, and the same hosting company had owned the IP addresses of the scam pages since 2020.

Finally, as a future effort, we will consider making fixed-point observations of the prize scams. This way, we will be able to identify and disclose the latest reality of prize scams and alert and prevent them. We will consider reporting malicious domain names in the collected redirect chains to JPCERT/CC, hosting service and CDN providers, and Google.

Appendix

In this section, we review our findings in SCIS2021 and SCIS2022; in SCIS2021, we collected the redirect chain logs from December 30, 2020, to January 5, 2021, and in SCIS2022, from December 30, 2020, to January 3, 2022.

SCIS2021 Results of Study

We investigated the domain names of the entrance page, intermediate page, and scam page in 99 logs of HAR file obtained manually. HAR file is an abbreviation of "HTTP Archive File". It is a file containing a log of trace information from within a browser session. It records web requests made by the browser to the website page, including request and response headers, the body, and the time it takes to load the assets. We also investigated the ownership of the IP addresses by using DB-IP.

Entrance Page Results – We found that the obtained 99 logs of HAR File contained 99 URLs for the entrance page. When duplicate domain names were omitted from 99 URLs, 99 domain names were included. Furthermore, duplicated IP addresses were excluded from the 99 domain names, and 62 IP addresses were retained.

We investigated the owners of these 62 IP addresses and their details. The results are shown in TABLE 4 below.

Intermediate Page Results – We found that the obtained 99 logs of HAR file contained 400 intermediate page URLs. 68 domain names were included when duplicate domain names were omitted from these 400 URLs. Furthermore, 36 IP addresses were retained when duplicated IP addresses were omitted from the 68 domain names.

We investigated the owners of these 36 IP addresses and their details. The results are shown in TABLE 5 below.

Scam Page Results – We found that the obtained 99 logs of HAR file contained 99 scam page URLs. When duplicate domain names were omitted from these 99 URLs, 16 domain names were included. Furthermore, excluding duplicated IP addresses from the 16 domain names, 13 IP addresses were retained.

We investigated the owners of these 13 IP addresses and their details. We found that all 13 IP addresses were within the range of 5.189.217.0/24. 5.189.217.0/24 is an IP address owned by a hosting company called Fast Content Delivery LTD.

TABLE 4. IP address owners and details of entrance pages in SCIS2021

IP address owner	Number of IP addresses owned
Cloudflare, Inc.	29
Contabo GmbH	15
EuroByte LLC	7
Ruweb	4
Ovh Sas	2
Webdia Sa	2
Zomro B.V.	2
Microsoft Corporation	1

TABLE 5. IP address owners and details of intermediate pages in SCIS2021

IP address owner	Number of IP addresses owned
DigitalOcean, LLC	10
Clundfare, Inc.	7
Fast Content Delivery LTD	5
Linode, LLC	4
Ovh Sas	3
Hetzner Online GmbH	2
LeaseWeb Netherlands B.V.	1
Relink LTD	1
Ihor Hosting LLC	1
Complat-telecom It-Itd	1
ripe	1

SCIS2022 Results of Study

We investigated the domain names of the entrance page, intermediate page, and scam page in 496 logs obtained using automated collection tool. The automatic collection tool is a prototype system of the system completed in this study. We also investigated the ownership of the IP addresses by using DB-IP.

Entrance Page Results – The 496 obtained logs contained 496 URLs for the entrance page. When duplicate domain names were omitted from 496 URLs, 216 domain names were included. Furthermore, duplicated IP addresses were excluded from the 216 domain names, and 168 IP addresses were retained.

We investigated the owners of these 168 IP addresses and their details. The results are shown in TABLE 6 below.

Intermediate Page Results – The obtained 496 logs contained 4854 intermediate page URLs. 129 domain names were included when duplicate domain names were omitted from these 4854 URLs. Furthermore, 23 IP addresses were retained when duplicated IP addresses were omitted from the 129 domain names.

We investigated the owners of these 23 IP addresses and their details. The results are shown in TABLE 7 below.

Scam Page Results - The 496 logs contained 496 scam page URLs. When duplicate domain names were omitted from these 496 URLs, 232 domain names were included. Furthermore, excluding duplicated IP addresses from the 232 domain names, 19 IP addresses were retained.

We have examined the ownership and details of these 19 IP addresses. We found that 18 of the 19 IP addresses are in the 5.189.217.0/24 range, and one is in the 31.44.185.251 range. 5.189.217.0/24 is an IP address owned by a hosting company named Fast Content Delivery LTD. 31.44.185.251 is owned by a hosting company called WebLine LTD.

TABLE 6. IP address owners and details of entrance pages in SCIS2022

IP address owner	Number of IP addresses owned
Contabo GmbH	134
Cloudflare, Inc.	29
Amazon Technologies Inc	1
Ovh Sas	1
EuroByte LLC	1
Abdilaziz Uulu Zhusup	1
Mnet	1

TABLE 7. IP address owners and details of intermediate pages in SCIS2022

IP address owner	Number of IP addresses owned
DigitalOcean, LLC	10
Fast Content Delivery LTD	7
Clundflare, Inc.	3
The Constant Company, LLC	1
DataWeb Global Group B.V.	1
Flyservers S.A.	1

Conflict of Interest

This research paper presentation and registration were funded by Tokyo Denki University. All authors declare that they have no conflicts of interest.

Author Contributions

Takagi, Nagashima and Mitani conducted the research; Takagi, Nagashima and Mitani analyzed the data; Shirai and Saito supervised the whole research; Takagi and Nagashima wrote the paper; all authors had

approved the final version.

References

[1] BBB Warning: You got lucky and won a raffle from Amazon!? Watch out; it's another scam. [Online]. Available: https://www.bbb.org/article/news-releases/23914-bbb-warning-you-won-raffle-from-amazon-watch-out-its-another-scam

[2] Consultations regarding registration and billing on pay sites disguised as iPhone winners (in Japanese). [Online]. Available: https://www.ccj.kokusen.go.jp/jri_sysi?page=iPhone

[3] "You've WON an iPhone 13 Pro!" — 6 Hot Raffle Text Scams (BestBuy, Walmart, T-Mobile, Costco, AT&T, and Amazon Survey). [Online]. Available: https://news.trendmicro.com/2021/09/24/youve-won-an-iphone-13-pro-6-hot-raffle-text-scams-best buy-walmart-t-mobile-costco-att-and-amazon-survey/

[4] What is "Amazon loyalty program" scam?. [Online]. Available: https://www.pcrisk.com/removal-guides/18879-amazon-loyalty-program-pop-up-scam

[5] iPhone Prize Scam (Video Commentary) (in Japanese). [Online]. Available: https://www.jc3.or.jp/threats/examples/article-198.html

[6] JPCERTCC/Lucky-Visitor-Scam-IoC. [Online]. Available: https://github.com/JPCERTCC/Lucky-Visitor-Scam-IoC

[7] Masatake, S., Yoshihiro, O., Kazunari, M., Haruki Kobayashi, W, T., & Taiichi, S., A survey on redirect chains from google search to prize draw scam, (in Japanese) *Proceedings of 2021 Symposium on Cryptography and Information Security.*

[8] Masatake, S., Kazunari, M., & Taiichi, S., Automated log collection of redirect chains from Google search to prize draw scam. (in Japanese) 2022 Symposium on Cryptography and Information Security.

[9] Chen, Z., & Freire, J., Discovering and Measuring Malicious URL Redirection Campaigns from Fake News Domain names. *Proceedings of 2021 IEEE Security and Privacy Workshops* (SPW), 2021, pp. 1-6.

[10] Gianluca, S., Christopher, Kruegel., & Giovanni, V. 2013. Shady paths: Leveraging surfing crowds to detect malicious web pages. Association for Computing Machinery, New York, NY, USA, 133–144.

[11] David, Y. W., Stefan, Sa., & Geoffrey, M. V., 2011. Cloak and dagger: dynamics of web search cloaking. Association for Computing Machinery, New York, NY, USA, 477–490.

[12] Nektarios, L., Tyler, M., & Nicolas, C., 2014. A nearly four-year longitudinal study of search-engine poisoning. Association for Computing Machinery, New York, NY, USA, 930–941.

[13] Liao, X. J., Chang, L., Damon, M., Elaine, S., Shuang, H., & Raheem, Beyah., 2016. Characterizing long-tail seo spam on cloud web hosting services. *Proceedings of International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva*, CHE, 321–332.

[14] Mozilla MDN Web Docs, "Referer." [Online]. Available: https://developer.mozilla.org/en/docs/Web/HTTP/Headers/Referer

[15] Google Safe Browsing. [online]. Available: https://safebrowsing.google.com/

[16] Robotic Process Automation (RPA) | Microsoft Power Automate. [Online]. Available: https://powerautomate.microsoft.com/en-us/robotic-process-automation/

[17] Select-String - PowerShell - Microsoft Docs. [Online]. Available: https://docs.microsoft.com/en-us/powershell/module/microsoft.powershell.utility/select-string?view=powershell-7.2

Mizuki Takagi is a Bachelor of Engineering degree from Japan's School of Information and Communication Engineering. He is currently enrolled in the Department of Information and Communication Engineering at the same graduate school.

Konan Nagashima graduated from the Department of Information and Communication Engineering at Tokyo Denki University in Tokyo, and received a bachelor of engineering degree. He is currently a graduate student in the Department of Information and Communication Engineering at the university.

Kazunari Mitani graduated from Tokyo Denki University with a bachelor's degree in Information and Communication Engineering and entered graduate school to work as a system engineer while aiming to become a cyber security engineer.

Masatake Shirai is a Bachelor of Engineering degree from Japan's School of Information and Communication Engineering. He is currently enrolled in the Department of Information and Communication Engineering at the same graduate school.

Taiich Saito has a bachelor degree and a master degree of science from Waseda University, Tokyo, Japan, and a doctor degree of engineering from Chuo University, Tokyo, Japan. Since 2008, he has been a professor at Tokyo Denki University. His research work focuses on cybersecurity and software vulnerability.