

UNCOVERING BLACK-HAT SEO BASED FAKE E-COMMERCE SCAM GROUPS FROM THEIR REDIRECTORS AND WEBSITES

A PREPRINT

 **Makoto Shimamura**
Cyber Security Institute
Trend Micro Incorporated, Japan

Shingo Matsugaya
Cyber Security Institute
Trend Micro Incorporated, Japan
Japan Cybercrime Control center, Japan

Keisuke Sakai
Cyber Security Control Task Force
Kanagawa Prefectural Police, Japan

Kosuke Takeshige
Community Safety Department
Chiba Prefectural Police, Japan

 **Masaki Hashimoto**
Faculty of Engineering and Design
Kagawa University, Japan

June 12, 2025

ABSTRACT

While law enforcements agencies and cybercrime researchers are working hard, fake E-commerce scam is still a big threat to Internet users. One of the major techniques to victimize users is luring them by black-hat search-engine-optimization (SEO); making search engines display their lure pages as if these were placed on compromised websites and then redirecting visitors to malicious sites. In this study, we focus on the threat actors conduct fake E-commerce scam with this strategy. Our previous study looked at the connection between some malware families used for black-hat SEO to enlighten threat actors and their infrastructures, however it shows only a limited part of the whole picture because we could not find all SEO malware samples from limited sources. In this paper, we aim to identify and analyze threat actor groups using a large dataset of fake E-commerce sites collected by Japan Cybercrime Control Center, which we believe is of higher quality. It includes 692,865 fake EC sites gathered from redirectors over two and a half years, from May 20, 2022 to Dec. 31, 2024. We analyzed the links between these sites using Maltego, a well-known link analysis tool, and tailored programs. We also conducted time series analysis to track group changes in the groups. According to the analysis, we estimate that 17 relatively large groups were active during the dataset period and some of them were active throughout the period.

Keywords Fake E-commerce sites · Website defacement · Black-hat SEO · Maltego · Link analysis · Japanese keyword hack

1 Introduction

While law enforcements agencies and cybercrime researchers are working hard, fake E-commerce scam is still a big threat to Internet users. The number of fake E-commerce sites (hereafter referred to as “fake EC sites”) that aim to defraud people out of their money or steal their personal information has been increasing, resulting in significant financial damage to society. Many reports indicate that financially-motivated threat actors actively continue scams with fake EC sites [1, 2, 3, 4]. Additionally, in Japan, the number of reported fake EC sites is on the rise. According to a report from Japan Cybercrime Control Center (JC3) [3], 47,278 fake EC sites were reported to JC3 in 2023, while 28,818 sites in 2022. Therefore, we need to develop countermeasure to protect Internet users from fake EC sites.

SEO poisoning, or black-hat SEO, using malware installed into compromised websites is one of the major techniques for luring users to fake EC sites, where “SEO” stands for “search engine optimization”. This technique has been observed by multiple security vendors [5, 6, 7, 8]. These malware are installed into compromised websites to intercept web server requests and return arbitrary contents. By doing so, threat actors can send crafted sitemap to search engines,

leading them to index generated lure pages as if these were part of the compromised websites. Consequently, search engine users are directed to visit these sites. The malware then intercepts the request handler and redirect user's browser to fake EC sites. In this paper, we refer to these compromised websites as "redirectors", since they redirect users from search engine results to fake EC sites, and the malware conducts such redirection as "SEO malware". Particularly, the technique which uses Japanese keyword and redirects to Japanese fake EC site is known as "Japanese Keyword Hack" [7, 8].

We previously analyzed the relationship between some malware families used for this strategy [9] using information extracted from destined websites. But the approach lacked completeness because the collection of malware samples was done manually and therefore not all malware families related to black-hat SEO were analyzed. Thus there is an important research question still in a fog; *how many threat actor groups conduct fake EC scam in Japan?* To uncover this, we analyze fake EC sites collected in JC3 who monitors cybercrime for a long time. They collected 692,865 fake EC sites trawled from many redirectors over two years, and the collected fake EC sites are filtered by internal confirmation process in JC3, including manual analysis. Thus we expect we can identify more groups than the previous approach. We use data collected in JC3 from May 20, 2022 to Dec. 31, 2024 as a dataset in this study.

To identify groups, we analyzed links between malware family names, fake EC domains, and Matomo [10] servers as key entities in our previous study in [9]. But to do the same with the JC3 dataset, we cannot identify SEO malware installed in the redirectors. Thus we analyze links only information extracted from fake EC sites in this study. By analyzing links between entities, we can identify clusters of fake EC sites, whose number approximates the number of threat actor groups. Actually, redirectors can be included for analysis in theory. But we found many redirectors returned only one to three of fake domains during preliminary observation and thus there were difficulty to form groups with redirectors.

Next, based on the groups identified by the link analysis phase, we also conduct a time series analysis, which is described in the later section of this paper. The time series analysis will show trends in group activities and may reveal hidden relationship between groups; for example, if a group become inactive and another group become active instead, we can suspect the latter could be a successor of the former. This hypothesis could be confirmed by characteristics of the two groups. Moreover, if a group operates actively in long time, the group is worth to focus because we can expect the actors behind it will not be inactive soon.

The contributions of this research are threefold:

1. We estimated number of threat actor groups conducting fake EC scam with black-hat SEO in Japan, based on a large dataset with in-depth analysis.
2. Based on the groups identified, we conducted time series analysis to enlighten groups active in the latest situation.
3. We also show the effectiveness of our analysis with a case study; multiple groups use new strategy of refund scam on the top of fake EC scam.

The rest of this paper are constructed as follows. We describe the background and related works in Section 2 to clarify our research area. Section 3 explains the dataset used in this study. Section 4 presents the result of our link analysis to identify groups. Section 5 offers the result of time series analysis based on groups identified in the previous analysis. In Section 6, we give a case study of our analysis from warnings in Japanese Consumer Affairs Agency. We then discuss observations, limitations, and future research directions in Section 7. Finally we summarize this paper in Section 8.

2 Related works

Kodera, *et al.* [11] collected and analyzed fake EC sites from search engines using titles common in them and domain names of reported URLs to open blocklists such as URLHaus [12] and OpenPhish [13]. Their results show 99.8% of Japanese fake EC sites, that are actually referred as "redirectors" in this paper, redirect users to fake EC sites only when accessing from search engine result.

Yang, *et al.* [14] proposes a technique to detect black-hat SEO for Chinese illegal websites. The target SEO technique of the research is based on a slight defacement in contents of well-known websites, and thus it is different technique from the one focused in this paper.

Zhang, *et al.* [15] proposes a novel technique to detect compromised websites used for black-hat SEO. This could help our study to collect fake EC sites because it can detect a large number of compromised sites in a week, but we need more consideration as black-hat SEO is a common technique. i.e.) we need to develop an effective filter to detect compromised sites for SEO malware families.

Table 1: Used dataset from JC3, from May 20, 2022 to Dec. 31, 2024

| Type | # of entities |
|---------------------------------------|---------------|
| Fake EC domain | 105,286 |
| Email address | 13,456 |
| Matomo Server | 36 |
| 51.la ID | 4,958 |
| Total # of entities for Maltego graph | 123,736 |

To the best of our knowledge, none of these related studies tried to identify threat actors behind. In this study we try to identify threat actor groups using dataset of redirectors and destined sites. Our goal is to create building blocks to attribute fake EC scam targeting Japan to threat actors. Our previous study [9] tried to analyze relationship between fake EC sites and SEO malware families, using Matomo servers and email addresses. The study successfully identified fake EC site groups related to SEO malware families, but as we mentioned in the Introduction, there are lack of completeness because we could not collect “all” SEO malware families used in the wild. Instead, we use a large dataset from JC3 in this study to identify threat actors behind Japanese fake EC scams.

3 Dataset

JC3 dataset includes 692,865 fake EC sites mounted on 105,286 domains and 13,456 email addresses, 36 Matomo Server URLs, and 4,958 IDs of 51.la related to them which is summarized in Table 1. The fake EC sites are collected from redirectors by pretending accesses from search engine crawlers and victim users. From the fake EC sites, they extracted email addresses, Matomo Servers and 51.la IDs by regular expression match and a decode program if an email address is encoded by CloudFlare’s Web Application Firewall [16].

4 Link Analysis

4.1 Preliminary group identification

To analyze relationship between redirectors and fake EC sites, we follow our previous approach [9], i.e., we use Maltego [17], a well-known link analysis tool, to visualize relationships as a graph. We define three links to create a graph as followings, depicted in Figure 1;

- **Link #1) Domain → Mail address** A fake EC site may have multiple mail addresses linked to threat actors. If the same mail address is used in multiple sites, we regard the sites as linked.
- **Link #2) Domain → Matomo server** A fake EC site may have zero or one Matomo server to send information of visitors. If the same Matomo server is used in multiple sites, we regard these sites as related.
- **Link #3) Domain → 51.la ID** 51.la is a well-known access analyzer service often used in fake EC sites. A fake EC site may have zero or one 51.la ID to send information of visitors. If the same ID is used in multiple sites, we regard these sites as related.

Figure 2 is a graph based on the entire dataset, created by Maltego. It shows the entire relationship between fake EC sites, its E-mail address, Matomo servers and redirectors in the dataset. The graph is shown in “Circular Layout” to consider link density. As you can see, there are too many entities and links. Maltego was too heavy to visualize over 100,000 nodes. Thus we created a program to detect groups from the graph file.

The Python code conducts preliminary identification of possible groups from the graph. The code is shown as Listing 1. With the code, we detected 1,118 groups from the dataset. We then filter out groups that have less than 200 domains or less than 2,000 sites considered as small groups and less priority for analysis. For the rest of this paper focus on the left 8 groups listed in Table 2. Note that the group IDs were assigned by the order of domain numbers. G7, G8, and G10 are dropped from focus as they have less than 2,000. There are cases that multiple subgroups are considered as one group even very weakly connected with just one link due to the algorithm. We will separate them into subgroups by removing such weak links in the next subsection. Note that it is possible that an actor operates a massive number of sites in small numbers of domains and we mistakenly drop such “large” groups, but we surely confirmed dropped 1,110 groups are not the case.

Table 3 describes Matomo server domains used in fake EC sites in the groups. As readers can see, sites in group G3 uses a Matomo server in la51[.]xyz, which is the same Matomo domain used in G2. Thus we believe these two groups

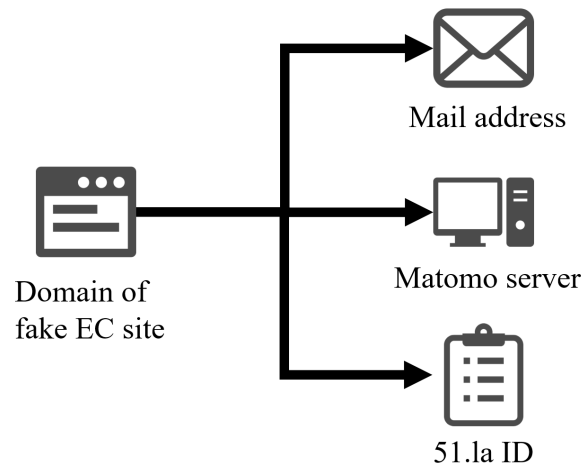


Figure 1: Entity connections to create a Maltego graph for link analysis

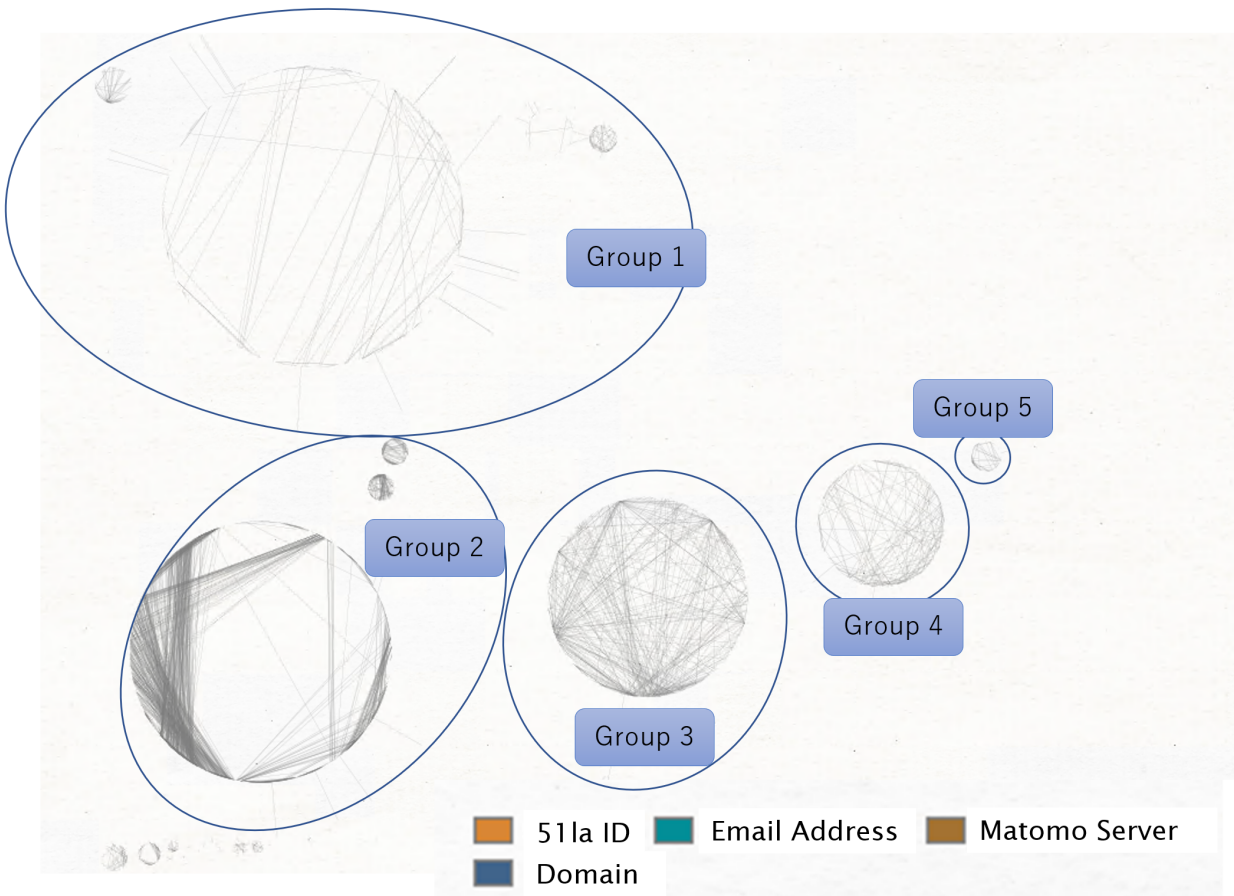


Figure 2: A maltego graph shows relationship of fake EC sites

Listing 1: Preliminary grouping algorithm in Python

```

1 # The entire graph is expressed as "links": a dictionary of links
2 # (sourceNode -> [targetNode1, targetNode2, ...]), and
3 # reverse links (targetNode -> [sourceNode1, sourceNode2, ...])
4 # for all links in the dataset
5
6 def detect_one_group(key, links):
7     group = set()
8     group.add(key)
9     prev_g_len = 0
10    # collect all links related to nodes in group
11    while prev_g_len < len(group):
12        prev_g_len = len(group)
13        newg = group.copy()
14        for src, dsts in links.items():
15            if src in group:
16                newg |= dsts
17        group = newg
18    return group
19
20 detected_groups = []
21 while True:
22     try:
23         key = list(links.keys())[0] # a node to begin with
24     except: # If unable to select node, finish group detection
25         break
26     detected_group = detect_one_group(key, links)
27     for k in detected_group:
28         del(links[k]) # delete already used links
29     detected_groups.append(detected_group)
30
31 print(f'Detected_{len(detected_groups)}_groups')
```

Table 2: Detected groups (G7, G8, and G10 are omitted due to small # of sites)

| Group ID | # of domains | # of sites | # of email addresses | # of Matomo servers | # of 51.la IDs | Remarks |
|----------|--------------|------------|----------------------|---------------------|----------------|---|
| G1 | 38,698 | 159,727 | 3,777 | 6 | 3,462 | Has subgroups |
| G2 | 37,665 | 335,787 | 6,517 | 21 | 24 | Has subgroups |
| G3 | 4,897 | 38,888 | 410 | 1 | 0 | Has subgroups, Possible relationship with G2 |
| G4 | 1,587 | 7,145 | 83 | 1 | 0 | |
| G5 | 1,361 | 5,514 | 385 | 3 | 46 | |
| G6 | 1,343 | 15,433 | 381 | 0 | 105 | |
| G9 | 352 | 13,080 | 37 | 0 | 39 | |
| G11 | 260 | 2,569 | 8 | 1 | 0 | |
| Subtotal | 86,163 | 581,430 | 11,598 | 33 | 3,676 | |
| Coverage | 81.84% | 83.92% | 86.19% | 91.67% | 74.14% | |

are highly related or the same actor group, but we could not confirm relationships from domains, email addresses, and 51.la IDs in them.

4.2 Detailed identification of subgroups

As we stated in the previous subsection, there are cases that multiple groups are considered as one group even very weakly connected with just one link due to the algorithm. To split a group into subgroups, we apply the algorithm described in Listing 1 to all cases where the removal of any single entity except for Matomo server from the group

Table 3: Detected groups and Matomo servers

| Group ID | Matomo servers (domain only) |
|----------|---|
| G1 | soupn[.]xyz, vxsem[.]xyz, uwmoon[.]xyz, heww[.]xyz |
| G2 | omtag[.]top, ockercsgre[.]top, utermcux[.]top, la51[.]xyz, dvdmoney[.]top, |
| G3 | axya[.]xyz, phoenixforce[.]xyz, vhuhuzce[.]xyz, alljecknet[.]com, gyfast[.]top, gens2[.]top |
| G4 | la51[.]xyz |
| G5 | onlinea[.]online |
| G9 | piwikcontrol[.]info, piwikfile[.]info, matomotogo[.]site |
| | https[.]or[.]ke, oknice03[.]top |

Table 4: Number of sites in detected subgroups after first-stage filtering (more than 200 domains)

| Subgroup ID | # of sites | Subgroup ID | # of sites | Subgroup ID | # of sites |
|-------------|------------|------------------------------|------------|-------------|------------|
| G2-1 | 206,886 | G5 | 5,514 | G1-13 | 373 |
| G1-1 | 94,627 | G11 | 2,569 | G1-17 | 363 |
| G2-2 | 61,310 | G1-6 | 2,065 | G1-16 | 356 |
| G2-3 | 42,961 | — Threshold: (2,000 sites) — | | G1-15 | 348 |
| G2-4 | 20,382 | G1-14 | 1,901 | G1-11 | 340 |
| G1-3 | 17,075 | G1-18 | 1,330 | G1-20 | 311 |
| G3-1 | 16,507 | G1-10 | 917 | G1-23 | 286 |
| G6 | 15,433 | G1-5 | 879 | G1-22 | 285 |
| G9 | 13,080 | G1-7 | 871 | G1-25 | 243 |
| G1-4 | 11,248 | G1-19 | 669 | G1-24 | 228 |
| G1-2 | 10,115 | G1-12 | 596 | G1-26 | 211 |
| G3-2 | 9,939 | G1-21 | 590 | G1-28 | 210 |
| G3-4 | 6,297 | G1-8 | 464 | G1-27 | 201 |
| G3-3 | 6,116 | G1-9 | 448 | | |

graph and extract subgroups that have over 200 domains. Note that we except for Matomo servers here because we believe it is less likely that multiple actors share a Matomo server; if multiple groups share Matomo servers, it is not natural there are many Matomo servers working. We describe the result in Table 4. In the table, group G1 is separated into 28 subgroups named as “G1-N” where N is the rank of domain numbers within subgroups of G1. Similarly, group G2 and G3 are also separated into subgroups. We applied a two-stage filtering process: first, we extracted subgroups with over 200 domains (shown in Table 4), then further filtered to focus on subgroups with over 2,000 sites for detailed analysis (shown in Table 5). Note that there are cases removing a entity results to three or more groups, and some of the groups are smaller than the threshold, 200 domains. We drop such small subgroups if it is the case, rather than merging them with other subgroups. Thus the total size of subgroups may not equal to the size of main group. For example, sum of the numbers of sites for G3-* is 38,859 whereas the number of sites related to G3 is 38,888.

Finally, we identified and focused 17 groups shown in Table 5. In the following, we will analyze further for the groups.

5 Time Series Analysis

We conducted time series analysis with the identified groups to track changes. We listed up entities in groups and labeled them. Then we counted how many entities were observed by month. We split the dataset by months. Because due to the collection period, data of the first month (10 days) and data of the last month (3 days) are not comparable with data of other periods. Thus we remove the periods from time series analysis.

The timecharts of numbers of domains in groups are depicted in Figure 3, Figure 4, Figure 5, and Figure 6 respectively. In Figure 3, G1-1 is active throughout the period. On the other hand, G1-2, G1-3, G1-6 shrink their activity in May, 2023. G1-3 appeared in Jun. 2023 to replace them.

Figure 4 indicates subgroups of G2 work in relatively long term. On the other hand Figure 5 illustrate subgroups of G3 operate in a short term, i.e.) less than 4 months.

In Figure 6, we can find G5 and G6 already stopped their activities. Notice that group G2-3, G9 and G11 become active in 2024 to the end date of the dataset. We think they are relatively new groups and worth to track.

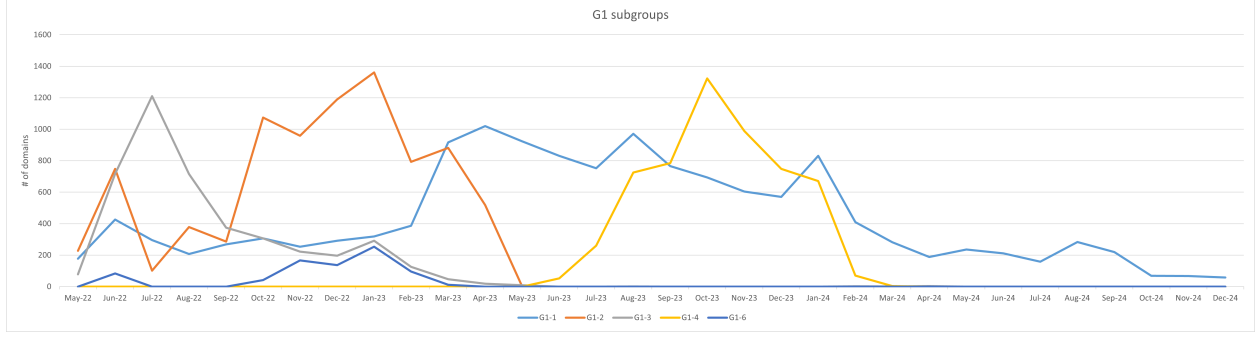


Figure 3: Time-chart of fake EC domains in subgroups of G1

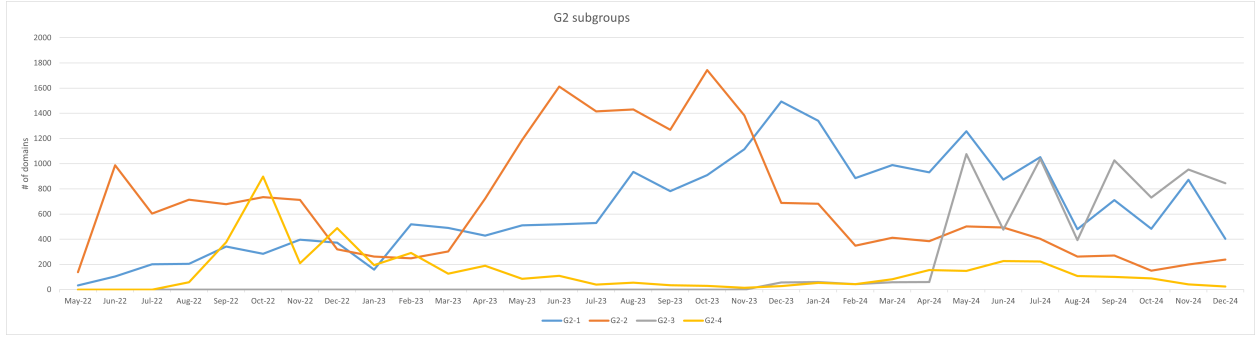


Figure 4: Time-chart of fake EC domains in subgroups of G2.

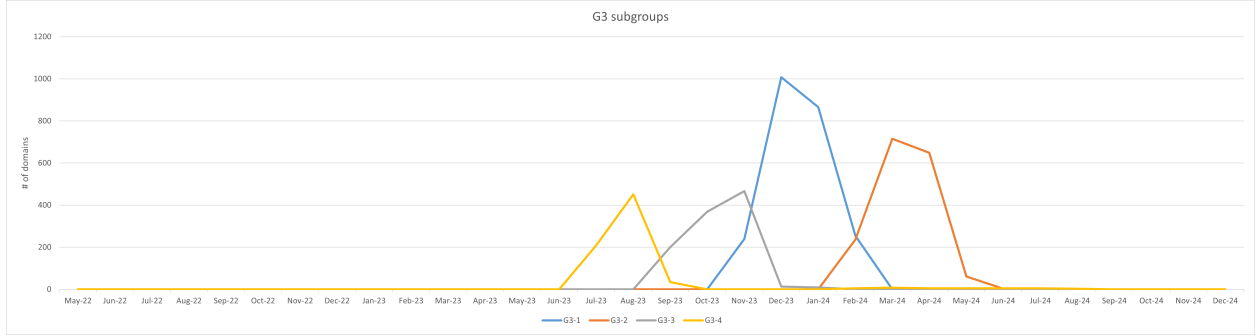


Figure 5: Time-chart of fake EC domains in subgroups of G3.

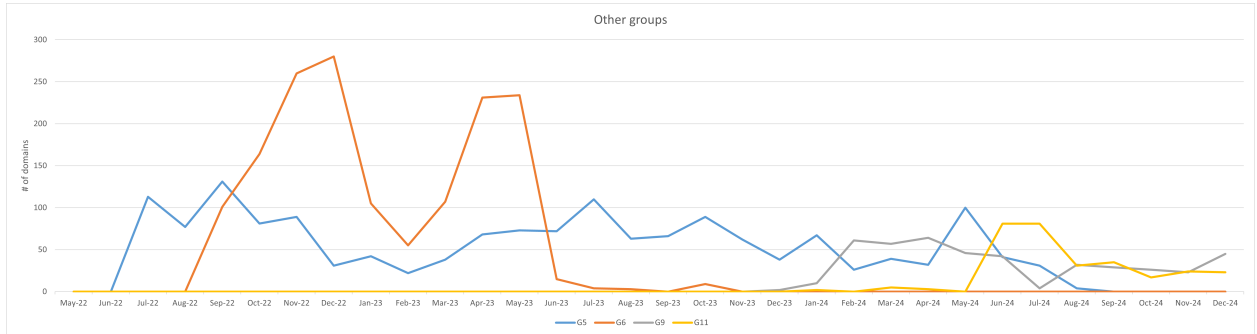


Figure 6: Time-chart of fake EC domains in other groups.

Table 5: Detected subgroups after cutoff

| Group ID | # of fake EC domains | # of email addresses | # of Matomo servers | # of 51.la IDs | Matomo servers (domain only) |
|----------|----------------------|----------------------|---------------------|----------------|---|
| G1-1 | 11,822 | 2,345 | 0 | 2,291 | |
| G1-2 | 8,262 | 336 | 0 | 0 | |
| G1-3 | 3,854 | 411 | 3 | 0 | soupn[.]xyz |
| G1-4 | 2,115 | 43 | 3 | 0 | heww[.]xyz, uwmoon[.]xyz, vxsem[.]xyz |
| G1-6 | 696 | 52 | 0 | 0 | |
| G2-1 | 16,314 | 4,403 | 5 | 23 | la51[.]xyz, gyfast[.]top, omtage[.]top |
| G2-2 | 12,498 | 819 | 5 | 0 | axya[.]xyz, vhuhuzce[.]xyz |
| G2-3 | 5,372 | 469 | 0 | 0 | |
| G2-4 | 2,986 | 823 | 5 | 1 | gens2[.]com, utermcux[.]top, alljecknet[.]com, ockercsgre[.]top |
| G3-1 | 1,828 | 114 | 1 | 0 | la51[.]xyz |
| G3-2 | 1,523 | 109 | 0 | 0 | |
| G3-3 | 853 | 99 | 0 | 0 | |
| G3-4 | 678 | 86 | 0 | 0 | |
| G5 | 1,361 | 385 | 3 | 46 | piwikcontrol[.]info, piwikfile[.]info, matomotogo[.]site |
| G6 | 1,343 | 381 | 0 | 105 | |
| G9 | 352 | 37 | 3 | 39 | https[.]or[.]ke, oknice03[.]top |
| G11 | 260 | 8 | 0 | 0 | |

6 Case Study: Fake EC sites and Refund Scam

In February 28th, 2025, Japanese Consumer Affairs Agency (CAA) released a public warning for “refund scam” in malicious EC sites [18]. The scam actor offers a victim for refund to her order with saying the ordered product is out of stock and not able to sell. Then the actor sends message to the victim with a code for a payment application for refunding procedure, but in fact the code is to send money to the actor. The actor then deceive her to push some buttons to send money with the code. The authority reportedly says this scam strategy incurs financial damage of over 680 million yen in about 5,500 cases and publicized 4 sites [19].

We analyzed whether the sites listed in the warning are fake EC sites, and we can attribute to groups if so. The named sites are *rdpgk[.]minimumrisk[.]shop*, *oggi[.]ayzgyonsale[.]shop*, *madrk[.]cnhmxbest[.]shop* and *qbague[.]voidnetwork[.]shop*. In the following discussion, we refer them as site ID 1 to 4, respectively. Note that we neutralize all their URLs to prevent readers from unintended access in this paper.

Our analysis reveals that the domains are fake EC sites because they are in the dataset. The detailed analysis result is summarized in Table 6. We got exact matches of sites ID 2 and 4 in the dataset, whereas we got no match of sites ID 1 and 3. But we easily identified ID 1 as G1-1 because many sites in *minimumrisk[.]shop* were linked to it. We also conclude the site ID 3 belong to the group G2-1 without burden because sites that share the same domain were in it.

We also show the column to show whether the domains are related to SEO malware in our previous study [9] for readers’ information. The sites ID 2 and 3 are related to malware B, and ID 4 is related to malware E. The site ID 1 is not connected to any known SEO malware family but the result infers it could be related to a variant of malware E because sites ID 1 and 4 are linked to the same group G1-1.

As a result of the analysis above, group G1-1 and G2-1 are related to refund scam cases at least. Surprisingly we find they are the two largest groups. In this case study, the result infers the strategy of refund scam with payment application is possibly shared within multiple fake EC scam groups. As we show in this case study, analysis allows us to attribute a scam strategy with fake EC to specific groups once new strategy comes up.

7 Discussion

7.1 Necessity for grouping fake EC sites

The grouping of fake EC sites contributes more detailed analysis, prevention and protection of the threat. For example, grouping enables us to identify unique characteristics linked to a group. Actually, in Section 4, we identified some

Table 6: Analysis result of URLs related to refund scam publicized from a Japanese CAA warning

| Site ID | Site URL | Group identified | Match in dataset | Related SEO malware in our previous study [9] | Related data first seen |
|---------|-----------------------------|------------------|------------------|---|-------------------------|
| 1 | rdpgk[.]minimumrisk[.]shop | G1-1 | Domain | N/A | Sep. 30th, 2024 |
| 2 | oggi[.]ayzgyonsale[.]shop | G2-1 | Site | Malware B | Oct. 31st, 2024 |
| 3 | madrk[.]cnhmxbest[.]shop | G2-1 | Domain | Malware B | Sep. 6th, 2024 |
| 4 | qbague[.]voidnetwork[.]shop | G1-1 | Site | Malware E | Oct. 2nd, 2024 |

small number of groups use Matomo for analyze visitors, whereas some others use 51.la. More detailed analysis in each group could shed more lights to fake EC actors, but we leave it as our future work. We believe our study in this paper contributes to create a building block to further analysis.

7.2 Remarks in small groups

We dropped groups that have less than 200 domains or 2,000 sites because of limited efforts. But there are some remarkable findings about the dropped groups during study.

- **Preference of access analyzer** In Table 2, we find dropped groups use only 3 Matomo servers, but 1,282 51.la IDs which is about 26%. It indicates smaller groups like to use 51.la rather than Matomo. From this result, we get to think small groups may have common tool and manuals that use 51.la.

7.3 Possible countermeasures

As we pointed out in the above sections, some large group uses Matomo. So the use of suspicious Matomo servers listed in Table 3 can be an indicator of fake EC sites. We think replacing Matomo server incurs some burdens on threat actors because relatively small numbers of Matomo server is used while they operate very large numbers of fake EC sites.

On the other hand, unfortunately, there are some difficulties to use 51.la IDs as indicators of fake EC sites. As we show in Table 5, group G1-1 uses 2,291 51.la IDs as of this analysis, and still the number is increasing. So listing all 51.la IDs related to fake EC sites are more difficult than enumerating used Matomo servers. We doubt some actors may use a dashboard app which can integrate multiple 51.la IDs into one view, as 51.la offers API [20] to export data in realtime.

7.4 Ethical consideration

From the viewpoint of research ethics, we thoroughly checked this paper prior to publication. This study implicitly collects PII (personally identifiable information) including business representatives' names, mail addresses, addresses of fake EC sites, and also information related to PII such as domain names. Most of these information are absolutely fake, or do not allow us to identify threat actors behind them. However, these information could possibly be linked to actual legitimate personnel if a threat actor copied them from other legitimate sites to impersonate, or steal identities from innocent persons. Thus we just show processed data sufficient for discussion in this paper except for URLs in the case study; we did not include any PII and raw data. We also masked URLs, domain names and product names of fake EC sites to avoid conflicts because they might include real trademarks or copyrights. Note that we believe mentioning original site URLs in Section 6 is no problem because they are confirmed as malicious sites and publicized by the Japanese authority.

On the other hand, we mention some product names like "Matomo" and "51.la" in this paper, and the Matomo server domains used by threat actors. We think these are important and necessary to include so that readers can reproduce and verify our analysis. The inclusion of these product names in this paper is not likely to damage their reputation.

The redirectors in the dataset are rather victims of website defacement than threat actor related servers. Thus we did not identify any redirectors in this paper. JC3 shares the list of redirectors to law enforcement agencies to reach to administrators of the sites and suggest them to fix, but it is still on the way because the number is huge and often administrators are unreachable, and new redirectors are frequently discovered in their monitoring process. Finally, we must state that JC3 conducts data collection from redirectors in a non-disruptive manner; the request rate is enough low (a request in 5 seconds) not to incur denial-of-services. We believe the access rate is lower than legitimate search engine crawlers and generally acceptable.

8 Summary

In this paper, we analyzed relationship between fake EC sites collected in JC3 to identify threat actor groups. Based on the fake EC site dataset collected from May 20, 2022 to Dec. 31, 2024, we identified 1,118 possible groups and then filtered to 17 groups based on their size for more detailed analysis. Then we conducted time series analysis for the 17 groups to analyze how long they are active. The result of time series analysis shows that some groups were active throughout the period of the dataset and some other groups were vanished in a short term. Also we give a case study using shared URLs from Japanese Consumer Affairs Agency and shows new refund scam strategy are used in multiple fake EC groups.

We hope the groups identified based on this analysis can help law enforcement to track their activities.

References

- [1] Bayse, “Customers of hundreds of major retail brands targeted by years-long attack,” 2024, [Accessed Jun., 2024]. [Online]. Available: https://www.bayse.io/blog/major_brands_consumer_fraud
- [2] M. Marx *et al.*, “BogusBazaar: A criminal network of webshop fraudsters,” 2024, [Accessed May., 2024]. [Online]. Available: <https://www.srlabs.de/blog-post/bogusbazaar>
- [3] Japan Cybercrime Control Center, “Statistics of malicious shopping sites in 2023,” Jun. 2024, (in Japanese), [Accessed Jun., 2024]. [Online]. Available: <https://www.jc3.or.jp/threats/topics/article-555.html>
- [4] E. T. R. Team, “Inside intelligence center: Financially motivated chinese threat actor silkspecter targeting black friday shoppers,” Nov. 2024, [Accessed Feb., 2025]. [Online]. Available: <https://blog.eclecticicq.com/inside-intelligence-center-financially-motivated-chinese-threat-actor-silkspecter-targeting-black-friday-shoppers>
- [5] LAC Co., Ltd., “Lac security insight vol.2 2022 autumn,” 2022, (In Japanese), [Accessed Apr., 2024]. [Online]. Available: https://www.lac.co.jp/lacwatch/pdf/20221214_lsi_vol2.pdf
- [6] Trend Micro Research, “Grouping actors that operate fake E-commerce sites targeting Japanese,” Oct. 2023, (In Japanese), [Accessed Jan., 2024]. [Online]. Available: https://www.trendmicro.com/ja_jp/research/22/j/seo-poisoning.html
- [7] Google, “Fix the japanese keyword hack,” 2015, [Accessed Apr., 2024]. [Online]. Available: <https://web.dev/articles/fix-the-japanese-keyword-hack>
- [8] A. Martori, “How to find & fix the japanese keyword hack,” 2020, [Accessed Apr., 2024]. [Online]. Available: <https://blog.sucuri.net/2020/04/japanese-keyword-hack.html>
- [9] M. Shimamura, S. Matsugaya, K. Takeshige, K. Sakai, and M. Hashimoto, “An analysis of the relationship between black-hat seo malware families leveraging information from redirected fake e-commerce scam sites,” in *2024 7th IEEE Conference on Dependable and Secure Computing*, Nov. 2024.
- [10] “Matomo - The Google Analytics alternative that protects your data — matomo.org,” [Accessed Jan., 2024]. [Online]. Available: <https://matomo.org/>
- [11] H. Kodera, T. Koide, D. Chiba, K. Aoki, and M. Akiyama, “Understanding attacks with fake shopping websites,” *IPSJ Journal*, vol. 62, no. 9, pp. 1523–1535, sep 2021, (In Japanese).
- [12] abuse.ch, “URLhaus - Malware URL exchange,” [Accessed Jul., 2024]. [Online]. Available: <https://urlhaus.abuse.ch/>
- [13] OpenPhish, “OpenPhish - Phishing Intelligence,” [Accessed Jul., 2024]. [Online]. Available: <https://openphish.com/>
- [14] R. Yang, X. Wang, C. Chi, D. Wang, J. He, S. Pang, and W. C. Lau, “Scalable detection of promotional website defacements in black hat SEO campaigns,” in *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Aug. 2021, pp. 3703–3720. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity21/presentation/yang-ronghai>
- [15] J. Zhang, C. Yang, Z. Xu, and G. Gu, “PoisonAmplifier: A guided approach of discovering compromised websites through reversing search poisoning attacks,” in *Proceedings of the 15th International Symposium on Research in Attacks, Intrusions and Defenses (RAID’12)*, September 2012.
- [16] “Email Address Obfuscation - Cloudflare Web Application Firewall (WAF) docs — developers.cloudflare.com,” [Accessed Jan., 2024]. [Online]. Available: <https://developers.cloudflare.com/waf/tools/scrape-shield/email-address-obfuscation/>

-
- [17] “Homepage — maltego.com,” [Accessed Jan., 2024]. [Online]. Available: <https://www.maltego.com/>
- [18] G. o. J. Consumer Affairs Agency, “Warning about businesses that use code payment services such as “XX Pay” to request money transfers instead of refunds, disguised as refund procedures on online shopping sites,” February 2025, [Accessed Mar., 2025] (In Japanese). [Online]. Available: <https://www.caa.go.jp/notice/entry/041215/>
- [19] Kyodo News, “Consumer Affairs Agency releases name of site disguised as refund, 600 million yen in lost funds,” March 2025, [Accessed Mar., 2025] (In Japanese). [Online]. Available: <https://nordot.app/1268157726111236224?c=302675738515047521>
- [20] G. o. J. Consumer Affairs Agency, “51LA,” [Accessed Mar., 2025] (In Chinese). [Online]. Available: <https://web.51.la/doc/>