IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 December 2023, accepted 26 January 2024, date of publication 30 January 2024, date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3360108

RESEARCH ARTICLE

Longitudinal Measurement Study of the Domain Names Associated With the Olympic Games

RYO KAWAOKA[1], DAIKI CHIBA[2], (Member, IEEE), TAKUYA WATANABE[3], MITSUAKI AKIYAMA[3], (Member, IEEE), AND TATSUYA MORI[1,4,5], (Member, IEEE)

[1]Computer Science and Communication Engineering Department, Waseda University, Shinjuku-ku, Tokyo 169-8555, Japan
[2]NTT Security (Japan) KK, Chiyoda-ku, Tokyo 101-0021, Japan
[3]NTT, Musashino-shi, Tokyo 180-8585, Japan
[4]NICT Koganei-shi, Tokyo 184-8795, Japan
[5]RIKEN AIP Chuo-ku, Tokyo 103-0027, Japan

Corresponding author: Tatsuya Mori (mori@nsl.cs.waseda.ac.jp)

ABSTRACT In this study, we conducted a comprehensive longitudinal measurement study of domain names associated with major global events. We aimed to understand the registrants' motives, usage, and abuse of these domain names. We specifically focused on the Olympic Games since they attract sustained attention from when the venue is announced to the event's conclusion. Our study focused on the Tokyo, Beijing, and Paris Olympics. Our three-year investigation revealed that the number of Olympic-related domain name (ODN) registrations increased concurrently with the postponement of the 2020 Tokyo Olympics and the diplomatic boycott of the 2022 Beijing Olympics. Furthermore, we discovered a substantial increase in the number of ODNs used for malicious websites just before the games. Many ODNs related to the regional nature of each game were acquired, and several ODNs required close attention from a security perspective.

INDEX TERMS Measurement, domain name, Olympic games.

I. INTRODUCTION

Large-scale events that garner global attention significantly impact society, including users' Internet activities. Conversely, our activities in a society heavily influence such events. The Olympic Games are a quintessential example of a large-scale global event.[1] The Games have various websites dedicated to ticket sales, souvenirs, hotel reservations, broadcast schedules, live video streaming, and preliminary game results. Consequently, the Olympics are more susceptible to cyberattacks [1]. The Tokyo Olympics was postponed from 2020 to 2021 due to COVID-19, resulting in the registration of numerous new domain names, as we will explore later. The Beijing Olympics faced a diplomatic boycott because of human rights concerns, leading to the acquisition of multiple domain names related to the boycott campaign during the event.

We note that the Olympic Games are unique in their global reach and impact, drawing audiences in the billions and commanding a brand value comparable to globally renowned corporations like Apple and Google [2]. As a temporary but intensely focused global event held in different locations every four years, the Olympic Games carry distinctive characteristics that enhance its visibility and reputation. Furthermore, the increased use of social media since the 2004 Summer Olympic Games has allowed global audiences to actively engage with the Games, contributing to its media coverage and enhancing its reach and impact [3].

Building on this background, we conducted an extensive longitudinal measurement study of domain names associated with major global events, focusing, in particular, on the Olympic Games. As quintessential examples, we concentrated on the Tokyo 2020, Beijing 2022, and Paris 2024 Olympic Games. Each Olympic Games event has its official website operated by the Organizing Committee of the host country. In addition to the official site domain

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Agostino Ardagna.

[1]'War' serves as an example of another large-scale event that attracts global attention. Studying domain names associated with wars can be significant due to their worldwide impact and interest. However, this topic falls outside the scope of our current investigation and is marked for future research.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

names, this study examined Olympic-related domain names (ODNs) registered by various stakeholders. These ODNs reflect the intense and temporary engagement associated with each Olympic Games, distinguishing it from permanent entities like well-known brand names. As we elaborate later, an ODN is a domain name that contains the name of the host city, the year of the event, and the term ''olympic'' as substrings. This measurement study aimed to determine the registration patterns of ODNs and their purposes. In addition, we investigated the long-term evolution of ODNs used for malicious purposes, including phishing and malware distribution sites, providing specific examples of such malicious ODNs.

Our analysis of large-scale longitudinal measurements of ODNs is expected to offer valuable insights into the patterns and modus operandi of phishing attacks targeting large-scale, global sporting events like the Olympic Games and prominent global events like the World Expo. This study aims to assist organizations and nations planning to host such large-scale events in the future by promoting security practices to prevent phishing attacks, including managing official domain names and disseminating related information.

Our key findings can be summarized as follows:

- The number of ODN registrations increased concurrently with the postponement of the Tokyo Olympics and the diplomatic boycott campaign of the Beijing Olympics.
- In the period leading up to the games, the number of malicious websites using ODNs increased significantly.
- ODNs related to the regional nature of each game were registered.
- Numerous ODNs required close attention from a security perspective.

The remainder of this paper is organized as follows. Section II describes the measurement methodologies. Section III presents the dataset used in our study. Section IV exhibits the results of our longitudinal analysis of Olympic domain names. Section V presents a discussion of our findings. Section VI provides an overview of related work. Finally, Section VII concludes the paper.

II. MEASUREMENT METHODOLOGIES

A. OVERVIEW

Figure 1 illustrates the workflow of the measurement study. First, ODNs are extracted through keyword matching on more than 260 million domain name data collected from over 1.5K different DNS zones of top-level domains (TLDs), such as country-code TLDs (ccTLDs), generic TLDs (gTLDs), and sponsored TLDs. Next, we applied a DNS A record lookup on the ODNs and accessed the website for each domain name for which we found an IP address associated with that domain name. If the website could be successfully accessed, we collected screenshots, HTML sources, and metadata from the website. Finally, for ODNs whose IP addresses are obtained, we apply VirusTotal, an online scanning service, to detect malicious websites.

In Section V-D, we discuss ethical considerations regarding our measurement study. In the following section, we describe each step in greater detail.

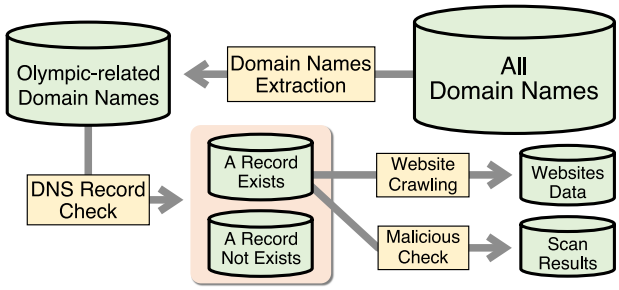


FIGURE 1. Workflow of the measurement study.

B. EXTRACTING ODNS FOR EACH OLYMPIC GAMES

We employed the following two-step strategy to extract the ODNs associated with a particular Olympic Games from numerous domain names. First, the ODNs were extracted from the list of domain names by three methods (Step 1). Next, the extracted ODNs are mapped to each event such that there is no overlap between events, where we refer to the Olympics held in each city as an event (Step 2). These steps are described as follows:

1) STEP 1: ODN EXTRACTION

We used three methods to extract ODNs: keyword matching, typosquatting, and IDN homographs.

a: KEYWORD MATCHING (KEY)

We demonstrate the procedure for extracting ODNs using keyword matching. For a given domain name, the target of keyword matching is the *effective label* (EL) of that domain name, where we define EL as the label next to the public suffix [4]; e.g., in the case of `tokyo2020.org`, the EL is `tokyo2020`.

We detect a domain name as an ODN if its EL satisfies two or more of the following conditions, which we empirically derived through a careful manual inspection[2]:

- EL contains EVENT-NAME.
- EL contains `olympic`.
- EL contains 4-digit numbers from the year of the event to 4 years after the year of the event. (e.g., 2020, 2021, 2022, 2023, and 2024 in the case of the Tokyo Olympics).
- EL contains `ticket` and EVENT-NAME.
- EL contains `ticket` and `olympic`.

The host city name included in the official domain name used for the Olympics is expressed as EVENT-NAME. For example, ''tokyo'' is used for the Tokyo Olympics, ''beijing''

[2]Although we evaluated the impact of keywords variations such as ''olympian'' instead of ''olympic'' and two-digit rather than four-digit years, the results were not significantly affected.

is applied for the Beijing Olympics, and "paris" is utilized for the Paris Olympics. We added "ticket" as a keyword because of the financial incentive for attackers and the high likelihood that they will be used for phishing attacks.

With the above rule, in the case of the Tokyo Olympics, domain names, such as `tokyo-olympic2020[.]example`, are detected as an ODN. Although this rule may result in the false detection of domain names unrelated to the Olympics, a prior manual inspection has revealed that such cases are extremely rare. It should be noted that the four-digit number includes the year of the event up to the year in which the next Olympics have been scheduled to be held.

We note that our keyword matching technique involves concepts related to combosquatting [5], which refers to domain squatting involving combinations of trademarks and other terms. While combosquatting employs dictionary words, our approach can be characterized as a dictionary technique combined with mangling rules for generating keyword variations, focusing on a wider set of keywords associated with major global events, such as city name, year, and terms like "tickets" and "olympics" for the Olympic Games.

b: TYPOSQUATTING (TYPO)

Typosquatting is a technique in which a common website or domain name is intentionally misspelled or incorrectly typed to trick users into visiting a website for fraudulent or malicious activity. Typosquatting is often used to mislead users into accessing fraudulent information or content, or in phishing attacks to steal passwords or personal information. We extract ODNs that target Typosquatting of official Olympic domain names. To expand the range of extraction, we generated a list of Typosquatting candidates based on the official domain name with the year part extended to the year 5 years, which is the same as in keyword matching.

First, based on the official domain names, five base strings were selected for each Olympics; `tokyo2020`, `tokyo2021`, `tokyo2022`, `tokyo2023`, `tokyo2024` for Tokyo Olympics; `beijing2022`, `beijing2023`, `beijing2024`, `beijing2025`, `beijing2026` for Beijing Olympics; `paris2024`, `paris2025`, `paris2026`, `paris2027`, `paris2028` for Paris Olympics.

The typosquatting list is generated by repeating the following operations for all combinations of these base strings:

- Delete one character (e.g. `toky2020[.]org`)
- Replace adjacent characters (e.g. `toyko2020[.]org`)
- Select one character and replace it with the character on the adjacent key in the QWERTY sequence (e.g. `tolyo2020[.]org`)
- Select one character and insert the same character just after it (e.g. `toyyko2020[.]org`)
- Select one character and insert the character of the adjacent key just after it (e.g. `toyklo2020[.]org`)

If a domain name contains the typosquatting list generated above as a substring, it is detected as an ODN.

c: IDN HOMOGRAPH (IDN)

There are IDN homograph attacks, in which a similar-looking domain name is generated using non-ASCII characters, in addition to typosquatting. To support IDN homographs using Internationalized Domain Names (IDNs), if a domain name contains non-ASCII characters (homoglyphs) that are similar to ASCII characters, the homoglyphs are replaced with the corresponding ASCII characters. To perform such substitution, we use the homoglyph character list from the previous study [6] and the list of similar character combinations provided by the Unicode Consortium [7]. With this substitution, for example, `tōkyō2020[.]com` is converted to `tokyo2020[.]com`. Applying the keyword matching described above to the converted domain name enables detecting ODNs using the IDN homograph.

2) STEP 2: EVENT MAPPING

Step 2 aims to map the ODNs extracted in Step 1 for each Olympic event. Intuitively, mapping is straightforward for cases in which the city name is included in the EL, and the mapping is non-trivial for other cases. For example, ODNs extracted when they contain "olympic" and "ticket" as strings, and cases that contain more than one city name, are not self-evident when assigned to Olympic events.

In this study, we developed a rule to map the extracted ODNs to events, as illustrated in Figure 2. First, we check if the ODN contains the EVENT-NAME (tokyo, beijing, or paris). If the ODN contains only a single event name, we mark it as the corresponding event; if the ODN contains multiple event names, we mark it as other. If the ODN does not include the EVENT-NAME, we check if the year is included. If the year is included, we map it according to the year in which each Olympics is held. i.e., if the year is 2020 or 2021, it is mapped to the Tokyo Olympics (taking into account the postponement); if the year is 2022, it is mapped to the Beijing Olympics; and if the year is 2024, it is mapped to the Paris Olympics. If it is none of these, it is marked as other. If the ODN contained neither EVENT-NAME nor year, we checked the ODN's registration date and mapped it to the Tokyo Olympics if it was registered from 2019 through July 2021, to the Beijing Olympics if it was registered from August 2021 through January 2022, and to the Paris Olympics if it was registered after that date. ODNs registered before 2019 were marked as other.

This rule has a few limitations. For example, if the ODN does not contain information on the host city or year of the event, we use WHOIS to extract the registration year of the domain name and apply that information as a clue to mapping the domain name to the event. However, this method does not allow an accurate mapping for cases in which domain names are registered well in advance of the event. We manually

tested the validity of the rules and found no significant problems. This result is consistent with the observation that the number of registrations for ODNs increases as the date of a relevant event approaches.

C. WEBSITE INSPECTION

In the following section, we present methods for investigating websites that use ODNs to clarify their use. In particular, we determined a malicious site for cases in which a working website was found.

First, we checked whether an IP address was assigned to the ODN by conducting DNS lookups. Next, if the ODN has a valid IP address assigned, we send HTTP and HTTPS GET requests to collect response codes and content from the website. Here, the content consists of HTML code, screenshots, and other metadata. When we access a website, there are cases in which redirects occur. In such cases, data were collected from all redirect destinations. There were cases in which connection timeouts or errors related to TLS certificates occurred when accessing the website. In such cases, we recorded the content collection as having failed.

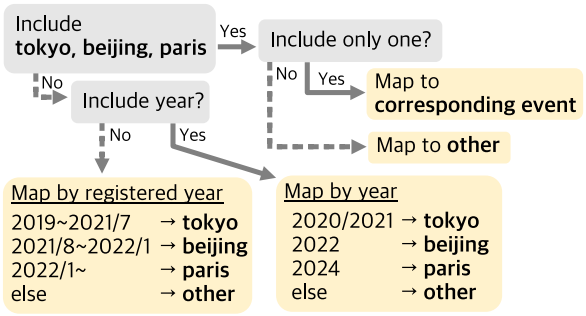


FIGURE 2. Rule for mapping ODNs to events.

We used Google Chrome version 86.0.4240.0. to access the website, with the user agent set to Windows 10 on a desktop (desktop) and iOS 12.2 on an iPhone (mobile). Each website was accessed a total of four times using HTTP and HTTPS and two user agents.

Finally, we applied the online scanning service VirusTotal [8] to the ODNs where IP addresses exist and to all related websites, including those redirected from the landing site. We note that this was the only method used to identify malicious ODNs.

This study is a long-term measurement, and a large number of data is collected over time. We created the dashboard illustrated in Figure 3 to monitor the status of data collection and the current status over time. The dashboard is created with Grafana [9]. This dashboard shows the number of data we have collected at the latest point in time (e.g.. number of active domain names, number of domain names with DNS A records, number of screenshots obtained, number of ODNs, number of ODNs detected by VirusTotal (VT)), as well as a graph showing the number of ODNs over time and the number detected by VT, the top ODN registrar list, and VT detection details. (Rev. 1-(2))

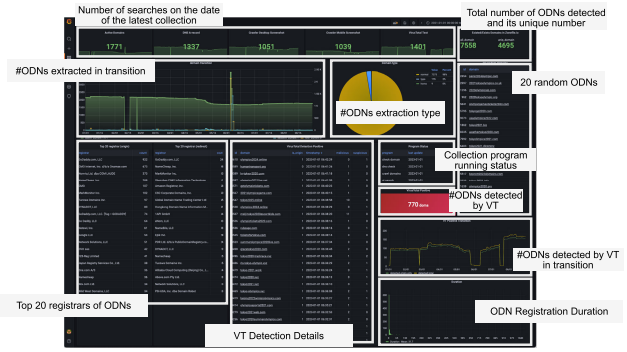


FIGURE 3. Screenshot of collection status monitoring tool built with Grafana.

For reference purposes, the snapshot of the dashboard can be accessed from https://snapshots.raintank.io/dashboard/snapshot/BelRyeYxvJ9vZR2UaJHCewTAfi8FzUkw.

III. DATASET

In this section, we describe the data used for our measurement study, a large database of domain names, and the period and conditions under which we collected the data.

1) DATA SOURCE

This study adopted a commercial domain name database, Zonefiles [10], as a large domain name list. This database contains approximately 260 million domain names collected from 1.5K distinct DNS zones, i.e., classic TLDs, ccTLDs, and various generic/sponsored TLDs. The database is updated daily. One reason for choosing Zonefiles was its extensive coverage of ccTLDs, including.jp,.cn, and.fr, which are pivotal for analyzing the local characteristics of domain registrations pertinent to the host countries of the Olympics. In this study, we fetched the database daily and extracted and analyzed the ODNs using the method described in Section II.

There are two reasons for selecting Zonefiles.io as our primary data source. First, it contains the largest number of domain names (over 239 million domain names as of January 2020) of all the data we could realistically obtain. Second, we considered the local characteristics of the host country and emphasized ccTLD coverage in our measurement of registrations of domain names associated with the three Olympic Games. There was a possibility of using another service, such as OpenINTEL [11], but OpenINTEL only covered 14 ccTLDs and did not include.jp,.cn, and.fr, the host countries of the Olympics as of January 2020. On the other hand, Zonefiles.io, which we used in this study, covered 250 ccTLDs, including all of the above-mentioned.jp,.cn, and.fr. In Zonefiles.io, this high coverage is achieved both by zone transfers and by using its own crawler to scan daily for some gTLDs that do not respond to zone transfers and ccTLDs, which do not allow zone transfers to third parties [12].

2) DATA COLLECTION PERIODS

To study the ODNs of the Tokyo, Beijing, and Paris Olympic Games, our data collection spanned two years, with information collected daily. Given the constraints imposed by COVID-19 and the limitations of the data collection environment, the data collection was divided into four distinct cycles, as summarized in Table 1. It should be noted that there were periods of server or network downtime during each of these cycles. Looking at these periods of inactivity, as illustrated in Figure 4, we did not observe any significant shifts in data trends that would indicate a compromised dataset. In fact, there was a noticeable numerical gap between cycles 2 and 3, with cycle 3 showing a higher number of domains. This increase is attributed to our methodological shift from capturing differential zone file data in Cycle 2 to capturing complete zone file data in Cycle 3. Despite this variance, the overall trend in domain registrations remained remarkably stable, reinforcing our confidence that the conclusions drawn from our data are sound and reliable.

In retrospect, while the continuity of our data collection was interrupted by unexpected server outages and the widespread effects of the COVID-19 pandemic, these did not significantly derail our efforts or the substance of our findings. Such disruptions, while not initially accounted for in our research design, were managed with careful attention to detail, ensuring that the integrity of the study remained intact. The robustness of the trends identified in our analysis is robust. The comprehensive nature of our data set, combined with the methodological thoroughness with which we conducted our research, underscores the trustworthiness of the domain name registration trends at the heart of our research. Following this introspection, we describe the specific data collection methodology used for each cycle.

Cycle 1: On the first day of the survey, January 1, 2020, we extracted all domain names contained in Zonefiles. From the second day onward, we used the differential data acquisition function of Zonefiles to obtain domain names that were added or deleted each day. We can extract the domain names that exist on each day from the data collected using the aforementioned methods. We extracted the ODNs from the domain names that existed each day using the method described in Section II-B. Among the extracted ODNs, those with DNS A records were analyzed daily using VirusTotal (VT).

Cycle 2: The same procedure as that applied during Cycle 1 was used to extract the ODNs that were present each day. VT scans were conducted once every 2 days for ODNs, particularly for those with DNS A records. During Cycle 2, the frequency of VT was reduced because of server resource constraints.

Cycle 3: During Cycle 3, all domain names provided by Zonefiles were collected daily instead of obtaining the differences, and ODNs were extracted using the same procedure as that used during Cycles 1 and 2. This change was made owing to the possibility that data omissions may occur when differences are acquired. VT scans were conducted daily for ODNs, particularly for those with DNS A records. Since this period includes the Olympic period, we classified the website screenshots obtained during this period as a detailed analysis.

Cycle 4: Finally, we observed a trend in ODNs for an additional period after Tokyo and Beijing ended, and the Olympics attention subsided. The data collection methodology for this period was the same as in Cycle 3, but no analysis was conducted on website screenshots.

TABLE 1. Data collection cycles.

Cycle	Event	Period (YYYY/MM/DD)	# days
Cycle 1	Tokyo	2020/01/01–2020/05/15	133
Cycle 2	Tokyo	2020/10/01–2020/12/28	89
Cycle 3	Tokyo	2020/12/29–2022/04/30	451
Cycle 3	Beijing	2021/10/12–2022/04/30	180
Cycle 3	Paris	2021/10/12–2022/04/30	179
Cycle 4	Tokyo	2022/05/01–2022/09/30	144
Cycle 4	Beijing	2022/05/01–2022/09/30	132
Cycle 4	Paris	2022/05/01–2022/09/30	131

IV. MEASUREMENT STUDY

A. BASIC STATISTICS OF ODNS

Table 2 presents the number of unique ODNs extracted for each event. #ODNs (IP) and #ODNs (Web) represent the number of unique ODNs for which an A record existed and the number of unique ODNs that responded to an HTTP/HTTPS request, respectively. #ODNs (VT) represents the number of unique ODNs detected by two or more engines in VirusTotal.[3] ODNs that reappeared after disappearing during the measurement period were counted as a single ODN.

TABLE 2. Statistics of the collected ODNs.

Event	#ODNs	#ODNs (IP)	#ODNs (Web)	#ODNs (VT)
Tokyo	2,377	1,990	1,972	207
Beijing	583	549	537	31
Paris	1,181	778	776	36

The 1-year postponement is one of the reasons for the particularly large number of ODNs for the Tokyo Olympics. As demonstrated later, there are indications that domain names such as `tokyo2021` were registered in large numbers for speculative purposes when the Tokyo Olympics were postponed.

ODNs listed in the top three row categories in Table 5 are considered to have been acquired for speculative purposes. Here, speculative registration refers to acquiring domain names expected to appreciate in value for later resale or

[3]The VT detection results may contain false positives. We restricted our results to those detected by two or more engines to eliminate this effect.

monetization. While such domains may eventually be used for malicious websites, speculative registration alone does not necessarily imply malicious intent. The ODNs in the other categories are considered to have been acquired to operate functional websites.

It can also be seen that the number of ODNs for the Summer Olympics is greater than the number of ODNs for the Winter Olympics. For example, there are more ODNs for the Paris Olympics than for the Beijing Olympics, despite the Paris Olympics having yet to be held. These observations reflect that the Summer Olympics are generally larger than the Winter Olympics. In the case of the Tokyo/Beijing Olympics, the numbers of participating countries, athletes, and events were 205/91, 11,000/2,897, and 339/109, respectively [13], [14].

For the ODNs, we aggregated their public suffixes as effective TLDs (eTLDs). Publicsuffix list [4] was used to extract eTLDs. The results are listed in Table 3 (a) for Tokyo, Table 3 (b) for Beijing, and Table 3 (c) for Paris. Based on the information presented in the tables, two conclusions can be drawn. First, legacy TLDs such as com/org/net/info are the most frequently used TLDs for all Olympic Games. Second, in the Tokyo and Paris Olympics, TLDs related to the host city are used (e.g. tokyo, jp, paris, fr).

The number of ODNs per extraction type is shown in Table 4. The results show that the number of ODNs using typosquatting is larger than the number of ODNs using IDN homographs for all Olympics. Conversely, compared to the number of ODNs extracted by keyword matching, the number of ODNs extracted by typosquatting and IDN homographs is very small: approximately 3.6% ((80+2)/2295) for the Tokyo Olympics, 3.9% ((21+1)/561) for the Beijing Olympics, and 8.7% ((84+11)/1086) for the Paris Olympics.

B. LONGITUDINAL ANALYSIS

For each event, the time evolution of the number of active ODNs is shown in Figures 4, 5, and 6. The figures also present the time evolution of the number of ODNs detected as malicious/suspicious websites. Note that some of the spikes in the figures are measurement outliers owing to database synchronization glitches. The number of registered ODNs shows a unique pattern for the Tokyo and Beijing Olympics, as described below. For the Paris Olympics, which will take place in 1.5 years, the number of registered ODNs has not changed significantly and is on a gradual upward trend.

For the Tokyo Olympics (Figure 4), the number of ODN registrations increased drastically during Cycle 1. A closer inspection revealed that many ODNs containing words related to the 2021 hosting of the games, such as postpose, next, 2021, etc., were registered. In other words, this surge is due to the large number of domain name registrations that refer to the 2021 Olympics because the Olympics were postponed by one year due to COVID-19. Thus, many ODNs were registered for speculative purposes in April 2020, when the postponement was officially announced by the Tokyo Olympics Organizing Committee.

Since then, the number of ODNs has been on a gradual downward trend as previously registered domain names have expired. However, it is interesting to note that, as of April 2022, there are still more than 1,000 ODNs for the Tokyo Olympics. Regarding the Beijing Olympics (Figure 5), the number of registered ODNs increased sharply in November 2021. This rise was because of the diplomatic boycott of the Beijing Olympics owing to human rights issues, and a large number of ODNs associated with such a political movement were registered. The above finding can also be seen in the analysis of the character strings comprising the ODNs, which will be discussed later (Section IV-D).

For the Tokyo and Beijing Olympics, which have already been held, the number of ODNs detected by VirusTotal increased as the games approached. During the games, ODNs with a high confidence level of maliciousness appeared, particularly those with a VT detection count of 4 or higher. Interestingly, when the number of domains detected by VT in the ODNs for the Beijing Olympics increased, a similar trend was observed for the Paris Olympics.

Finally, the measurement results in Cycle 4 show that ODNs have decreased after the end of the Olympic games (Tokyo and Beijing) as expected. This observation suggests that domain names acquired for speculative purposes are being abandoned after the event. Conversely, the number of domain name registrations is increasing for the upcoming Paris Olympics (Figure 6). We can expect that the ODNs for the Paris Olympics will exhibit a similar time variation pattern as the ODNs for the Tokyo and Beijing Olympic Games.

C. BREAKDOWN OF THE ODN WEBSITES

To understand the usage of ODNs, we visually categorized the ODN websites by using their screenshots. We crawl ODN websites daily using the method presented in Section II-C. Next, we extract screenshots of websites that responded with an HTTP status code of 200 OK. For efficiency, we computed the SHA-3 hash value of the screenshots and aggregated images with the same hash value. Finally, we manually classified the screenshot images using a custom GUI classifier we developed with pyQt5 [15]. We developed a custom GUI to assist in the manual classification of screenshots. We exhibit the GUI in Figure 7. In this GUI, the screenshot is displayed in the center of the screen and can be zoomed in and out. We further reduced the time required for classification by allowing it to be operated using only the keyboard.

We classified the screenshots collected in Cycle 3 to 7 categories according to the aforementioned procedure. The meaning of each category is as follows:

- **Screenshot not Obtained**: No screenshots obtained with web crawling.
- **No Valid Content**: Screenshots of error messages, domain name purchase pages, etc., where it is assumed that no user owns the domain name.

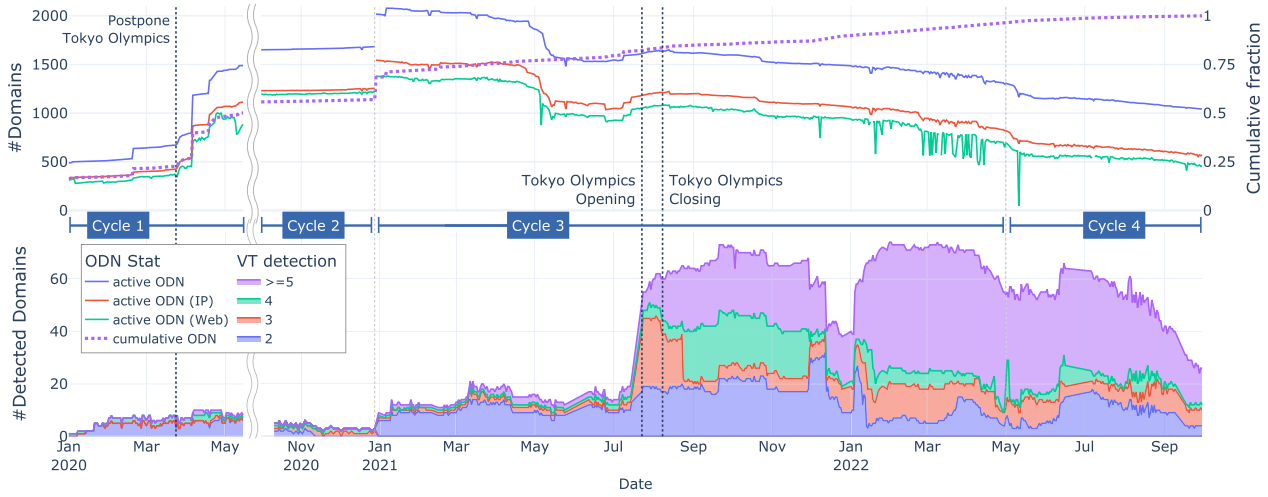


FIGURE 4. Time evolution of registered ODNs (top) and those detected by VT (bottom) for Tokyo Olympics.

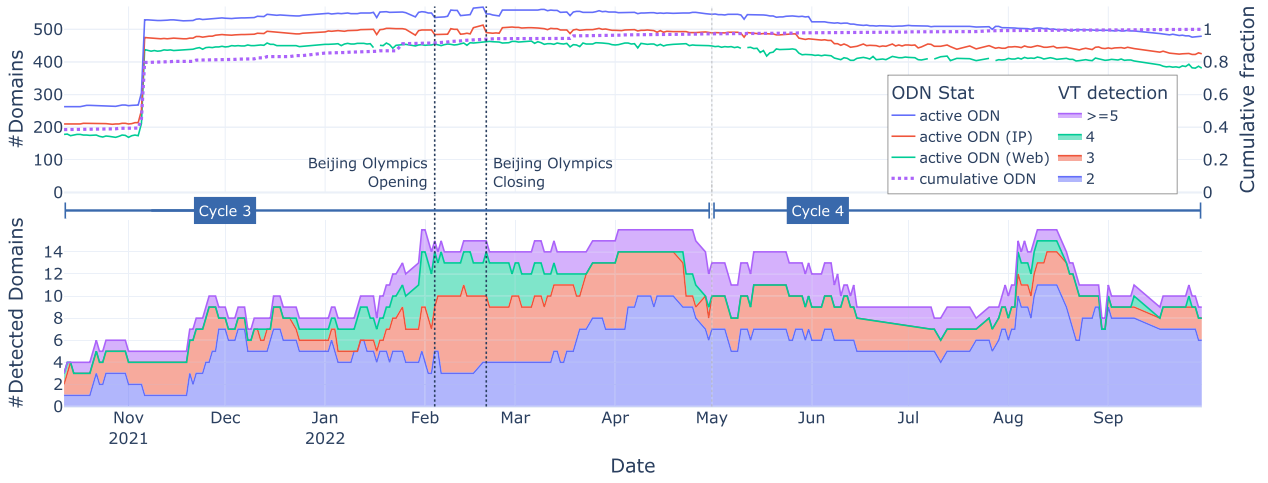


FIGURE 5. Time evolution of registered ODNs (top) and those detected by VT (bottom) for Beijing Olympics.

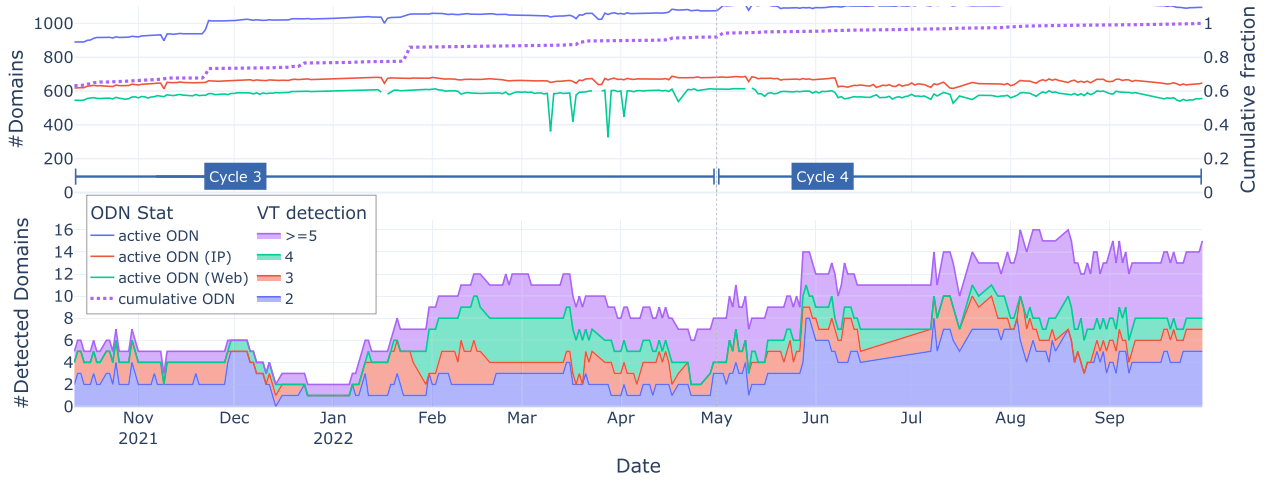


FIGURE 6. Time evolution of registered ODNs (top) and those detected by VT (bottom) for Paris Olympics.

TABLE 3. Top 10 eTLDs.

(a) Tokyo.

Rank	eTLD	count
1	com	1,055
2	org	256
3	net	122
4	info	87
5	tokyo	86
6	xyz	37
7	jp	36
8	site	34
9	co.uk	31
10	live	44

(b) Beijing.

Rank	eTLD	count
1	com	209
2	org	38
3	net	29
4	info	24
5	live	21
6	online	20
7	xyz	20
8	shop	14
9	vip	13
10	club	13

(c) Paris.

Rank	eTLD	count
1	com	501
2	org	125
3	net	85
4	info	44
5	co.uk	26
6	paris	24
7	fr	23
8	biz	22
9	tickets	16
10	store	16

TABLE 4. The number of ODNs per extraction type.

Extraction Type	Tokyo	Beijing	Paris
Keyword (Key)	2,295	561	1,086
Typosquatting (Typo)	80	21	84
IDN homograph (IDN)	2	1	11
Sum	2,377	583	1,181

- **Trivial Websites (index of/advertisement)**: Screenshots showing an index of advertisements, etc., which infer that the user who owns the domain name exists but the website is under construction or the domain name is not in operation.
- **Not Related to Olympics**: Screenshots confirm that the website is operational and that the website delivers content not related to the Olympics
- **Related to Other Olympics**: Screenshots confirm that the website is operational, and the website delivers content related to the Olympic Games to which the ODN is not mapped
- **Related to Olympics**: Screenshots confirm that the website is operational, and the website delivers content related to the Olympic Games to which the ODN is mapped
- **Very Similar to Official Olympics websites**: Screenshots confirm that the website is operational and the website is very similar to the official Olympic website.

The classification results are listed in Table 5. Note that if the website category of the ODN changed during the measurement period, a more relevant category (i.e., lower category in Table 5) was assigned. For example, if an ODN was initially domain parking (Trivial websites) and subsequently the content for the Olympics was posted (Related to Olympics), the use of that ODN was determined to be "Related to Olympics."

ODNs listed in the top three row categories in Table 5 are considered to have been acquired for speculative purposes. Here, speculative registration refers to acquiring domain names expected to appreciate in value for later resale or monetization. While such domains may eventually be used for malicious websites, speculative registration alone does not necessarily imply malicious intent. The ODNs in the other categories are considered to have been acquired to operate functional websites.

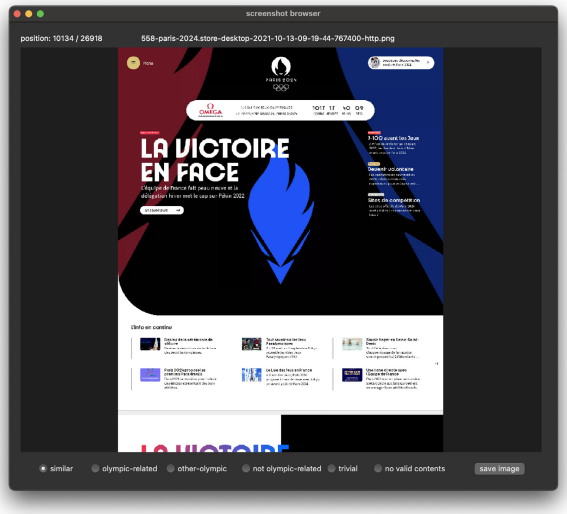


FIGURE 7. Custom GUI for Screenshot Classification.

Although most ODNs are acquired for speculative purposes, can be observed that other ODNs are used for websites related to the Olympics and those that distribute content similar to official Olympic websites. We found a malware distribution website whose appearance is similar to the official website of the Beijing Olympics. Other similar websites existed that redirected to the official site of the Olympics and the official site of the International Olympic Committee (IOC). Although these similar websites appear harmless, we caution against their potential for later exploitation.

We conducted a detailed content analysis of the ODN websites classified as "related to Olympics." The results are listed in Table 6. Across the three Olympics, many websites provided summarized information about the Olympic competitions, such as competition schedules, results, and participating teams (Olympic info.). In addition, each Olympics has specific usages, e.g., live streaming sites for the Tokyo

TABLE 5. Breakdown of the ODN websites.

	Tokyo				Beijing				Paris			
Category	Key	Typo	IDN	Sum	Key	Typo	IDN	Sum	Key	Typo	IDN	Sum
Screenshot not Obtained	281	2	1	284	316	4	0	320	105	10	0	115
No Valid Content	999	24	1	1024	94	3	0	97	384	21	4	409
Trivial Websites (index of/advertisement)	159	6	0	165	18	4	0	22	38	10	0	48
Not Related to Olympics	185	18	0	203	13	4	0	17	55	28	0	83
Related to Other Olympics	8	0	0	8	3	0	0	3	0	0	0	0
Related to Olympics*	209	2	0	211	44	0	1	45	34	1	0	35
Very Similar to Official Olympics Websites	5	0	0	5	9	0	0	9	10	1	0	11
Sum	1,846	52	2	1,900	497	15	1	513	626	71	4	701

*A website that describes the Olympic but does not look like an official Olympic website.

TABLE 6. Categorization results of Olympic-related websites.

(a) Tokyo.

Category	Key	Typo	IDN
Olympic Info.	116	2	0
Live Streaming	38	0	0
Travel	21	0	0
Ticket	11	0	0
Game	8	0	0
Accommodations	5	0	0
Others	10	0	0
Sum	209	2	0

(b) Beijing.

Category	Key	Typo	IDN
Olympic Info.	18	0	1
Boycott	11	0	0
Travvel	3	0	0
Sale Goods	3	0	0
Accommodations	2	0	0
Live Streaming	2	0	0
NFT	2	0	0
Game	2	0	0
Ticket	1	0	0
Sum	44	0	1

(c) Paris.

Category	Key	Typo	IDN
Olympic Info.	19	1	0
Travel	11	0	0
PCR	1	0	0
Accommodations	1	0	0
Game	1	0	0
Live Streaming	1	0	0
Sum	34	1	0

Olympics (38/211) and boycott-related websites for the Beijing Olympics (11/45).

From Table 4, ODNs of Typo and IDN are fewer than those of Keywords. Furthermore, checking the specific websites for ODNs of typo and IDN, Table 5 shows that there are only a few malicious ODNs that truly aim at the Olympics, and there are websites with a different purpose in the domain names that are actually in operation (Not Related to olympics). These are, for example, websites related to government administration and elections for the Tokyo Olympics, websites related to international conferences for the Beijing Olympics, and websites related to fashion, bicycle racing, and car racing for the Paris Olympics.

D. WORD ANALYSIS OF ODNS

A frequency analysis of keywords and an analysis for accompanying words that are not keywords were conducted for the ODNs extracted on a keyword matching to clarify the character strings that compose the ODNs.

1) KEYWORD FREQUENCY ANALYSIS

The ODNs contain at least two of the four keywords for the ODNs extracted by keyword matching. Therefore, to clarify which keywords are used simultaneously, a keyword co-occurrence analysis was performed and visualized using Upset [16]. In the Upset figure, the leftward-facing bar chart shows the number of each keyword, while the upward-facing bar chart shows the number of keywords circled together just below it. Figures 8, 9, and 10 show the results for Tokyo, Beijing, and Paris, respectively. By using the Upset figure, we can easily understand, for example, that for "Tokyo," the number of ODNs in which "olympic" is used is 1,002, and the number of ODNs in which "Olympic," "tokyo," and "20XX" are used together is 249.

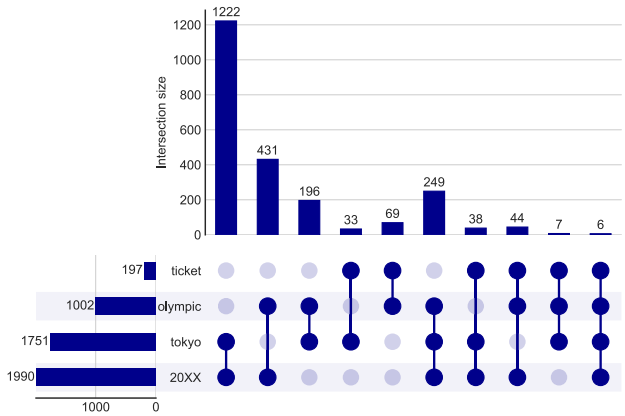


FIGURE 8. Keyword Frequency - Tokyo.

From Figures 8, 9, and 10, there are three findings. First, it can be visually understood that the combination of "EVENT-NAME and 20xx" is overwhelmingly used in the Tokyo and Paris Olympics, while this is not in the Beijing Olympics. Second, only a few ODNs use all four keywords (ticket, olympic, EVENT-NAME, 20xx) in the case of the Tokyo and Paris Olympics, but not in the Beijing Olympics.

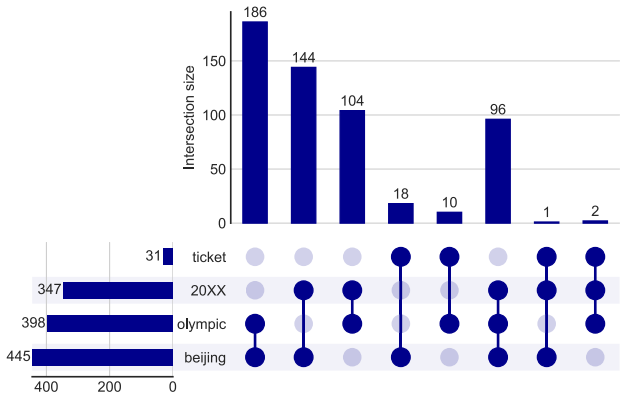


FIGURE 9. Keyword Frequency - Beijing.

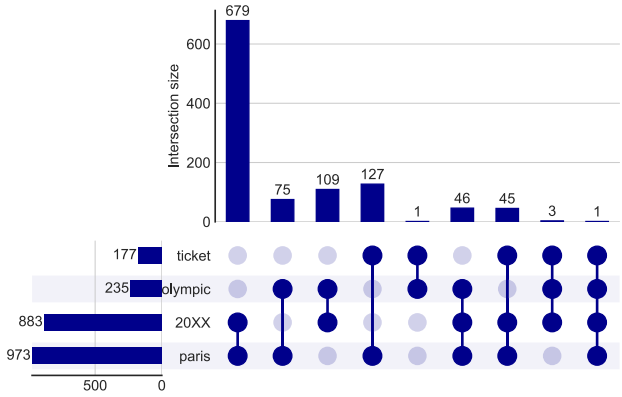


FIGURE 10. Keyword Frequency - Paris.

Third, the combination of ''ticket, olympic, and EVENT-NAME'' is used only in the Tokyo Olympics.

2) COMPOSING WORDS OF ODNS

We analyze the words that make up the ODN. For ELs in the ODN, we extracted words by splitting the string into words using the keywords used for detection or hyphens as separators. From the extracted words, stopwords such as prefixes and prepositions were removed using the list provided by nltk [17], a natural language processing tool. The English stopword list was applied to the Tokyo and Beijing Olympics, and the English and French stopword lists were applied to the Paris Olympics.

The results are listed in Table 7. As a result, we found that frequently used words are the name of the host country (''Japan,'' ''China,'' ''France''), the season in which the event is held (''summer,'' ''winter''), the name of accommodation facilities (''hotel,'' ''billets''), and the specific words that express the Olympic games used in the host country; i.e., analysis of composing words revealed the emergence of keywords based on host city language, such as 'gorin'' for Tokyo and ''jo'' for the Paris Olympics. These observations imply that by carefully building the list of keywords that reflect the cultural/geographical properties of events, we can expect to extract event-related domain names effectively.

E. MALICIOUS ODNS

We performed a manual analysis of malicious ODNs detected by VirusTotal using the screenshots of corresponding websites. We performed our analysis using the ODNs collected in Cycle 3.

First, we aggregated the detection categories for ODNs detected by two or more engines in Virustotal. The results are listed in Table 8. We have adopted the most frequently suggested detection category from the scan results for each day. If the detection categories were split in equal numbers, we marked them as Undetectable. The results illustrate that, across the three Olympics, the most common detection categories are malicious and phishing.

Next, we classified the screenshots collected by web crawling to understand the ODNs detected in VT from a web perspective.

The results of the classification are shown in Table 9. Note that this classification targets those detected by VirusTotal shown in Table 8; since there are no detections on IDNs, they are not included in the analysis and not listed in Table 9. The meaning of each category in Table 9 is as follows:

- **No Screenshot**: No screenshots obtained with web crawling.
- **HTTP Error**: A screenshot of HTTP error messages such as 404, 503, etc.
- **Temporary Page**: A screenshot showing an index of or under construct, etc., which infers that the web page is still under development.
- **Adv. (parking, expire)**: A screenshot showing an advertisement, which is displayed by the domain registrar for reasons such as parking or expiration.
- **Adv. (Other)**: A screenshot showing an advertisement that is not by the domain registrar.
- **Activity (Olympic)**: A screenshot showing content related to Olympic-related activities (e.g., flag relays, Olympic postponement information, Olympic-related symposiums, etc.).
- **Activity (non-Olympic)**: A screenshot showing content about non-Olympic-related activities (e.g., labor strikes, theme parks, Tokyo human rights posters, etc.).
- **Ticket**: A screenshot showing content about Olympic ticket sales.
- **Crypto**: A screenshot showing content about cryptocurrency.
- **Gamble**: A screenshot showing content about gambling.
- **Game**: A screenshot showing content about the game (e.g., sports games, card games, etc.).
- **Live**: A screenshot showing content about live streaming of the Olympic games.
- **Blog**: A screenshot showing a personal blog.
- **Malicious**: A screenshot showing malicious content (e.g., computer virus warnings, anti-virus download pages, etc.).

TABLE 7. Composing words of ODNs.

	Tokyo		Beijing		Paris	
Rank	word	count	word	count	word	count
1	games	109	winter	151	club	64
2	summer	58	boycott	112	jo	54
3	japan	55	games	90	games	51
4	shop	21	wintergames	8	leclub	32
5	ne	20	2008	8	boutique	20
6	jo	19	move	5	experience	20
7	live	16	special	5	shop	19
8	readysteady	15	china	5	france	15
9	plus1	15	art	4	jeux	13
10	updates	12	sgames	4	htraffic	13
11	torchrelay	12	game	4	hospitalite	12
12	mascot	12	kexing	3	hospitality	12
13	lympics	11	bj	3	ing	12
14	game	10	accommodation	2	billetterie	12
15	fieldcast	10	cincinnati	2	billet	12
16	mascots	10	meta	2	billets	12
17	steady	10	news	2	jeuxolympiques	10
18	ready	10	susagames	2	summer	10
19	jp	10	gear	2	marathon	8
20	form	9	stop	2	tous	8
21	info	9	sgame	2	disneyland	8
22	london	8	mds	2	boston	7
23	miraitowa	8	s2008	2	sgames	7
24	slive	8	gamer	2	filming	7
25	escortkids	8	museum	2	marathonpourtous	7
26	booking	8	reservations	2	business	6
27	someity	8	know	1	masters	6
28	crowd	8	live	1	rolex	6
29	recruiting	8	s2028	1	equestrian	6
30	monitor	8	20	1	expo	6

- **Travel/Accommodation**: A screenshot showing content about travel or accommodation information.
- **Boycott**: A screenshot showing content related to the Olympic boycott.
- **Mascot**: A screenshot showing content related to Olympic mascot sales (e.g., sales of Bing Dwen Dwen).
- **Loan**: A screenshot showing the loan application form (e.g., car loan application form).
- **Clothing Sales**: A screenshot showing content about clothing sales (e.g., sales of T-shirts.).

We found several malicious ODN websites with the appearance of a live-streaming website or a boycott campaign website. As mentioned in Section IV-C, many ODNs were registered for live streaming the Tokyo Olympics due to the fully online nature of the event, and for the Beijing Olympics, we observed many ODN registrations associated with the boycott movement. We find that 60.5% (23/38) of ODNs associated with live streaming and 45.5% (5/11) of the ODNs associated with the boycott movement were detected by VirusTotal. These results suggest that attackers are likely to register ODNs associated with events that have a significant social impact and leverage them to attract victims to their malicious sites. We have not yet identified such an event for the coming Paris Olympics, but a similar event may be observed in the future. The social events leading up to the Paris Olympics and the resulting increase in domain name registrations will need to be closely monitored.

V. DISCUSSION

A. IMPLICATIONS FOR THE EVENT ORGANIZERS

In the following, we explore the concept of ideal domain names for the official Olympic Games games websites based on the measurement findings we revealed. We then extend our findings to the case of generic large-scale global events.

1) IDEAL DOMAIN NAMES FOR OFFICIAL OLYMPIC WEBSITES

This study has discovered phishing sites that utilized ODNs that closely resembled the official domain names of the respective Olympic Games. A unique challenge in protecting against the abuse of ODNs is that they are registered by various stakeholders, making it challenging to implement a brand protection strategy. Consequently, it is challenging for ordinary users to verify the authenticity of domain names.

To effectively prevent damage caused by malicious sites that abuse ODNs similar to official sites, one measure would

TABLE 8. Categorization results for malicious types of websites detected by VirusTotal.

(a) Tokyo.

VT Category	Key	Typo	IDN
Malicious	74	2	0
Phishing	56	0	0
Malware	26	2	0
Suspicious	7	0	0
Spam	5	0	0
Undetectable	35	0	0
Sum	203	4	0

(b) Beijing.

VT Category	Key	Typo	IDN
Phishing	12	3	0
Malicious	9	0	0
Malware	4	0	0
Spam	1	0	0
Suspicious	1	0	0
Undetectable	1	0	0
Sum	28	3	0

(c) Paris.

VT Category	Key	Typo	IDN
Malicious	13	1	0
Phishing	11	1	0
Suspicious	2	0	0
Malware	2	1	0
Spam	1	0	0
Undetectable	4	0	0
Sum	33	3	0

TABLE 9. Categorization results of websites detected by VirusTotal.

(a) Tokyo.

Websites Category	Key	Typo
No Screenshot	4	0
HTTP Error	14	0
Temporary Page	27	0
Adv. (parking, expire)	66	4
Adv. (Other)	30	0
Live	23	0
Game	7	0
Activity (Olympic)	6	0
Blog	6	0
Malicious	5	0
Travel/Accommodation	4	0
Login Page	3	0
Activity (non-Olympic)	2	0
Gamble	2	0
Ticket	1	0
Crypto	1	0
Clothing Sales	1	0
Loan	1	0
Sum	203	4

(b) Beijing.

Websites Category	Key	Typo
No Screenshot	2	0
HTTP Error	2	0
Temporary Page	1	1
Adv. (parking, expire)	9	0
Adv. (other)	2	0
Boycott	5	0
Gamble	0	2
Mascot	2	0
Ticket	1	0
Crypto	1	0
Activity (non-Olympic)	1	0
Game	1	0
Live	1	0
Sum	28	3

(c) Paris.

Websites Category	Key	Typo
No Screenshot	1	0
HTTP Error	0	1
Temporary Page	4	0
Adv. (parking, expire)	15	0
Adv. (other)	4	1
Blog	4	0
Activity (non-Olympic)	2	0
Malicious	2	0
Activity (Olympic)	1	1
Sum	33	3

be to limit the number of official domain names used by the Organizing Committee for the Olympics. The Tokyo Olympics provides a good example of why this is important, as confusion arose when there were two official domain names, `tokyo2020.jp` and `tokyo2020.org`. Having too many official domain names can create confusion and make it difficult for users to distinguish between official and non-official sites. To avoid this, it is desirable to establish a consistent operation for domain names. This would enable users to easily recognize the official domain names and avoid confusion caused by similar-sounding domain names.

One possible solution to establish long-term consistency and make it easy for users to identify official sites while maintaining a degree of flexibility for each host city's domain name operations would be to create a generic top-level domain (gTLD) specifically for the Olympics (e.g., `.olympic`) and impose restrictions on domain name registration. This would limit registration to the Organizing Committee and licensed organizations only.

2) EXTENSIONS TO GENERIC MAJOR GLOBAL EVENTS

This research primarily focused on ODNs; however, our approach can be extended to other large-scale events by modifying the domain name extraction and event mapping techniques we developed in this study. For instance, we can extract relevant domain names for events like the FIFA World Cup, which garners a similar level of attention as the Olympics, by adjusting the keyword used for domain name extraction from "olympic" to "fifa" or "worldcup" and modifying the year and location parameters accordingly.

For example, the domain name `qatar2022.qa` was used for the 2022 FIFA World Cup in Qatar, and we can utilize the techniques outlined in this research to investigate and develop countermeasures against domain name registrations targeting this tournament. After mapping the extracted domain names to their respective events, we can conduct a similar investigation procedure, procedure as described in Section III.

We believe that adopting a similar approach by utilizing `.fifa` as the official domain name for the FIFA World Cup

is promising, akin to the proposal for a generic top-level domain for the Olympics we have discussed earlier. This could ensure consistency in the domain names used for official FIFA World Cup websites across different host countries and strengthen the FIFA brand protection efforts. However, the potential challenges and limitations, such as cost and the need for widespread adoption, must be accounted fortaken into consideration. Further research and discussion among relevant stakeholders, including FIFA, Organizing Committees, and ICANN, would be necessary to evaluate the feasibility and potential impact of this proposal.

B. COUNTERMEASURES AGAINST MALICIOUS DOMAIN NAMES FOR LARGE-SCALE EVENTS

According to the results of this study, domain name registrations and malicious domain names related to an event tend to increase just before the event is held. Therefore, countermeasures against malicious domain names should be enhanced as the event approaches. At the same time, some domain names are registered and used long before the event, even if it is still a considerable amount of time away; thus, away, and thus it is desirable to initiate countermeasures as early as possible after the decision to hold the event is made.

As we have revealed, incidents related to events of significant societal impact, such as the postponement of the Tokyo Olympics or the boycott movement for the Beijing Olympics, can lead to the registration of domain names related to the incident. Therefore, it is important to pay close attention to such incidents related to events and to conduct quick investigations and implement countermeasures accordingly.

Finally, in this study, we discovered a group of ODNs used for malicious activities. By using the MITRE ATT&CK framework [18] to analyze phishing attacks involving these malicious ODNs, we gain valuable insights into the methods and tactics used by cybercriminals. This knowledge helps us develop stronger defense strategies. Examples of the framework's usage can be found in the Appendix. We note that this framework is not limited to ODNs and can be readily applied to general global large-scale events. Additionally, by compiling observed malicious domain names used for the events, we can develop a blocklist to protect against the phishing attacks using such domain names, while also serving as Indicators of Compromise (IoC) for monitoring traffic from infected systems.

C. LIMITATIONS

For this study, we utilized Zonefiles as our data source for domain names. This database was constructed by collecting domain names from various DNS zones. However, one disadvantage of Zonefiles is that it does not provide information on subdomain names. To collect subdomain names, Certificate Transparency (CT) log data, which records issued TLS certificates [19], [20], [21], could be utilized. However, analysis of ODNs using CT logs is left as a topic for future research.

When collecting website content, we accessed only the root directory of each website. Therefore, if there was no index file in the root directory, some website content may have been missed. We expect that, in combination with a web search engine, such a limitation can be addressed [22].

D. ETHICS

We ensured that no personally identifiable information was used in this study, and we solely collected and analyzed publicly available information, such as DNS records and web content associated with Olympic-related domain names.

To minimize the load on the investigated websites, web crawling was designed to limit the number of requests for web content collection. Specifically, our crawling applied at most four patterns of access per day to the top page of each website, with at least a 2-hour interval, using two protocol types (HTTP or HTTPS) and two device types (desktop or mobile). Web crawling was performed for three events, and we did not deduplicate the investigated domain names. Thus, the websites of the domain names found for different events were accessed 12 times per day. We note that, even under this worst-case scenario, the load on each website should be negligible.

VI. RELATED WORK

In this section, we categorize the related work of our study. Table 10 compares the type, which indicates whether each study or solution is an academic research or a commercial solution, the primary methodology, and the focus of the study.

First, we discuss prior academic research from the perspective of conducting measurements on domain name registration. Domain names related to high-profile events often receive active registrations for speculative purposes such as resale or serving ad content. Coull et al. [23] characterized speculative domain name registrations and verified their availability by applying rules to identify current events from popular Google search queries. However, their analysis was limited to short-term events and a limited data source. Conversely, our study extracted domain names from a large database over a period of 1.5 years, providing a more comprehensive understanding of the registration of domain names related to major global events. Crises facing humanity, such as COVID-19 [24], [25] and Ebola [38], offer significant opportunities for domain name monetization. For instance, Pletinckx et al. [24] utilized the DomainTools Threat List to evaluate domains related to COVID-19 and found that these domains were registered to redirect commercial services or present affiliate links to Shopify or Amazon. Similarly, Kawaoka et al. [25] analyzed around 1.65 million domain names to identify trends in the registration of domain names associated with COVID-19 and their purposes. Our study, in contrast to previous research, demonstrated an increase in speculative domain name registrations as major global events approached despite

TABLE 10. Comparison with other prior studies and commercial solutions.

Study/Solution	Type	Methodology	Focus
Our Study	Academic	Measurement	Domain name registration associated with major global events over time
Coull et al. [23]	Academic	Measurement	Domain name registration associated with short-term hot topics/events
Pletinckx et al. [24]	Academic	Measurement	Domain name registration associated with COVID-19
Kawaoka et al. [25]	Academic	Measurement	Domain name registration associated with COVID-19
Hao et al. [26]	Academic	Measurement	Malicious domain name registration with DNS lookup patterns
Korczynski et al. [27]	Academic	Measurement	Malicious domain name registration per TLDs
Tian et al. [27]	Academic	Measurement	Squatting domain registration targeting specific brands
Feng et al. [28]	Academic	Detection	Phishing website
Tang et al. [29]	Academic	Detection	Phishing website
Abdelnabi et al. [30]	Academic	Detection	Phishing website
Lin et al. [31]	Academic	Detection	Phishing website
Liu et al. [32]	Academic	Detection	Phishing website
VirusTotal [8]	Commercial	Online Scan	Malicious website
urlscan [33]	Commercial	Online Scan	Malicious website
PhishTank [34]	Commercial	Threat Feed	Phishing website
OpenPhish [35]	Commercial	Threat Feed	Phishing website
Google Safe Browsing [36]	Commercial	Browser Blocklist	Malware and phishing website
Microsoft Defender SmartScreen [37]	Commercial	Browser Blocklist	Malware and phishing website

utilizing a similar approach of keyword matching. Therefore, our findings represent a departure from past studies and will offer unique and valuable insights for countries and organizations preparing to host large-scale events in the future.

Second, we outline prior academic research conducting measurements of malicious domain name registration. Hao et al. [26] revealed that the DNS infrastructure and name resolution patterns for malicious domain names differ significantly from those for legitimate domain names. Korczynski et al. [27] analyzed domain names from 11 threat information feeds using WHOIS information, web content, and DNS records and showed an increase in spam domains in newer gTLDs. Tian et al. [39] scanned domains using an approach detecting five types of squatting and identified 657 thousand domains that potentially impersonate 702 popular brands. They further classified web pages using visual analysis and OCR and found 1,175 phishing pages. In contrast to these studies, which focus on the analysis of domain names that imitate specific brands, we analyzed the Olympic Games, which are significant events, to reveal the evolution of domain registrations over time.

Third, numerous academic studies have attempted to detect phishing websites. The following are some recent academic contributions. Feng et al. proposed Web2Vec [28], a phishing website detection method that treats the URL, HTML page content, and DOM structure as strings and utilizes a representation learning technique to learn the web page representation. Tang et al. introduced a deep learning framework incorporating browser plug-ins for detecting phishing websites [29]. Abdelnabi et al. developed VisualPhishNet [30], a framework based on a triplet convolutional neural network for identifying novel-looking phishing websites. Lin et al. proposed Phishpedia [31], a deep learning-based system accurately recognizing and detecting legitimate website logos appearing on phishing websites. Liu et al. presented a deep vision-based approach for detecting phishing websites by inferring the intent of a website's appearance and the placement of UI elements on its screenshots [32]. While these studies focus on detecting brand-specific phishing websites or phishing websites in general, our research is distinct due to its emphasis on long-term, first-time measurements of domain names and their characteristics in relation to a major global event, the Olympic Games.

Fourth, we identify some typical examples of commercial solutions that implement countermeasures against malicious websites, including phishing. VirusTotal [8], which we utilized in Sec. IV, is a tool facilitating online scans of websites or domain names using multiple commercial security products. Urlscan [33] is a specialized website analysis service that scrutinizes website behavior in greater detail. PhishTank [34] and OpenPhish [35] are threat feeds concentrating on phishing websites, providing real-time information on phishing sites detected. Google Safe Browsing [36] and Microsoft Defender SmartScreen [37] are blocklist-based solutions pre-installed in Google Chrome and Microsoft Edge browsers, blocking known malware distribution sites and phishing sites. Our study supplements these investigations by not directly offering them as a commercial solution. Rather, the findings on domain names and websites resulting from our analysis of major global events can be utilized to determine domain names warranting more frequent monitoring in these solutions and to generate rules for detection and blocking.

In addition to these related studies, the following represents related research in our methodology: In the context of our extensive measurements of large TLD datasets, the need for high-performance, scalable infrastructures has become increasingly apparent. Van Rijswijk-Deij et al. [40] explored the challenges of scaling active DNS measurements to cover

expansive TLDs such as.com, a domain with over 123 million names. This work is particularly relevant, as it highlights the importance of balancing extensive data collection with the need to avoid overburdening the global DNS infrastructure. Furthermore, in the context of analyzing domains equivalent to well-known brand names, Pochat et al. [41] examined the reliability of website popularity rankings. They found that even well-established rankings are prone to significant variation and manipulation, which could skew the results of security and privacy research. They introduced Tranco, a more robust, manipulation-resistant ranking list tailored for research, which could improve the validity of studies related to the popularity of branded domains.

VII. CONCLUSION

We conducted a longitudinal measurement study of domain names associated with three Olympic Games hosted in Tokyo (2020/2021), Beijing (2022), and Paris (2024). Our study revealed that the number of domain name registrations increased around the time of the postponement of the Tokyo Olympics and the diplomatic boycott of the Beijing Olympics. We also observed a significant increase in the number of domain names used for malicious websites shortly before the Games. Furthermore, we found that many domain names related to the regional nature of each game were registered and required close attention from a security standpoint. Based on our findings, we propose the creation of an Olympic-specific gTLD (e.g., `.olympic`) to be used exclusively by official Olympic websites and trusted companies and organizations for the Olympic Games held in each respective country. We also note that organizations planning future events should pay attention to regional keywords and domain registration patterns specific to the locality when making security preparations.

APPENDIX
ANALYZING PHISHING ATTACKS USING ODNS WITH THE MITER ATT&CK FRAMEWORK

This appendix presents an example of analyzing the phishing attack utilizing ODNs and the alignment of this analysis with the MITRE ATT&CK framework. This framework allows us to thoroughly understand the Tactics, Techniques, and Procedures (TTPs) employed in a phishing attack scenario that leverages ODNs to direct users towards malicious websites.

A. TACTICS

Tactics signify the adversary's primary objectives or goals. Under the "Resource Development" tactic [42], particularly "Acquire Infrastructure" (T1583) [43], the adversary targets high-profile events like the Olympics. They anticipate that the broad appeal and global attention of these events will attract a larger user base for potential phishing attacks. This tactic capitalizes on the trust and recognition associated with the Olympics, thereby increasing the chances of user interaction and victimization to the phishing scheme.

B. TECHNIQUES

Techniques are the specific methods the adversary uses to achieve their tactical goals. In this scenario, one possible technique could be "Spearphishing via Service" (T1566.03) [44], where the attacker sends phishing emails, seemingly originating from a reliable source, to deceive users into visiting a malicious website.

C. PROCEDURES

Procedures encompass the detailed steps an adversary takes using a technique to meet their tactical objective. In this scenario, the procedure might involve:

1) "Domain Registration" (T1583.001) [45]: The adversary registers a domain name seemingly related to the Olympics.
2) "Compromise Infrastructure" (T1583.002) [46]: The adversary crafts an email seemingly originating from a legitimate Olympic-related entity.
3) "Drive-by Compromise" (T1189) [47]: The adversary embeds a link in the email leading users to an attacker-controlled website.
4) "Content Spoofing" (CAPEC-148) [48]: The adversary designs the phishing site to mimic an authentic Olympic-related website, tricking users into providing sensitive information.

Aligning our analysis with the MITRE ATT&CK framework allows us to provide a more structured and comprehensive understanding of phishing attacks leveraging ODNs.

REFERENCES

[1] Redscan. (2021). *Dangerous Games: The Cyber Security Threats To the Olympics*. [Online]. Available: https://www.redscan.com/news/cyber-security-threats-tokyo-olympics-2020/
[2] G. Morley. (2012). *Is the Olympics Worth More Than Google*. [Online]. Available: https://edition.cnn.com/2012/07/25/sport/olympics-london-2012-google-apple/index.html
[3] A. N. Geurin and M. L. Naraine, "20 years of Olympic media research: Trends and future directions," *Frontiers Sports Act. Living*, vol. 2, pp. 1–12, Sep. 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fspor.2020.572495
[4] *Public Suffix List*. Accessed: Feb. 3, 2024. [Online]. Available: https://publicsuffix.org/
[5] P. Kintis, N. Miramirkhani, C. Lever, Y. Chen, R. Romero-Gómez, N. Pitropakis, N. Nikiforakis, and M. Antonakakis, "Hiding in plain sight: A longitudinal study of combosquatting abuse," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2017, pp. 569–586.
[6] H. Suzuki, D. Chiba, Y. Yoneya, T. Mori, and S. Goto, "ShamFinder: An automated framework for detecting IDN homographs," in *Proc. Internet Meas. Conf.*, Amsterdam, The Netherlands, Oct. 2019, pp. 449–462, doi: 10.1145/3355369.3355587.
[7] M. Suignard. *Unicode Security Mechanisms*. Accessed: Feb. 3, 2024. [Online]. Available: http://unicode.org/reports/tr39/
[8] *VirusTotal*. Accessed: Feb. 3, 2024. [Online]. Available: https://www.virustotal.com/
[9] *Grafana*. Accessed: Feb. 3, 2024. [Online]. Available: https://grafana.com/
[10] *Zonefiles*. Accessed: Feb. 3, 2024. [Online]. Available: https://zonefiles.io/
[11] OpenINTEL. (2021). *Active DNS Measurement Project*. [Online]. Available: https://openintel.nl/coverage/

[12] Zonefiles. (2021). *Bringing Domain Data To Professionals. ZoneFiles in Numbers.* [Online]. Available: https://zonefiles.io/in-numbers/

[13] International Olympic Committee. (2021). *Tokyo 2020 Facts and figures*. [Online]. Available: https://olympics.com/ioc/tokyo-2020-facts-and-figures

[14] International Olympic Committee. (2022). *Beijing 2022 Facts and Figures*. [Online]. Available: https://olympics.com/ioc/beijing-2022-facts-and-figures

[15] Riverbank Computing Limited. *Python Bindings for the Qt Cross Platform Application Toolkit*. Accessed: Feb. 3, 2024. [Online]. Available: https://www.riverbankcomputing.com/software/pyqt/

[16] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, "UpSet: Visualization of intersecting sets," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1983–1992, Dec. 2014, doi: 10.1109/TVCG.2014.2346248.

[17] S. Bird, E. Loper, and E. Klein, *Natural Language Processing With Python*. Sebastopol, CA, USA: O'Reilly Media, 2009.

[18] MITRE Corporation. *ATT&CK: A Knowledge Base for Adversarial Tactics, Techniques, and Procedures*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/

[19] Q. Scheitle, O. Gasser, T. Nolte, J. Amann, L. Brent, G. Carle, R. Holz, T. C. Schmidt, and M. Wählisch, "The rise of certificate transparency and its implications on the Internet ecosystem," in *Proc. Internet Meas. Conf.*, Oct. 2018, pp. 343–349. [Online]. Available: https://dl.acm.org/citation.cfm?id=3278562

[20] Y. Sakurai, T. Watanabe, T. Okuda, M. Akiyama, and T. Mori, "Discovering HTTPSified phishing websites using the TLS certificates footprints," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Genoa, Italy, Sep. 2020, pp. 522–531, doi: 10.1109/EuroSPW51379.2020.00077.

[21] Y. Sakurai, T. Watanabe, T. Okuda, M. Akiyama, and T. Mori, "Identifying the phishing websites using the patterns of TLS certificates," *J. Cyber Secur. Mobility*, vol. 10, no. 2, pp. 451–486, Apr. 2021, doi: 10.13052/jcsm2245-1439.1026.

[22] W. Aqeel, B. Chandrasekaran, A. Feldmann, and B. M. Maggs, "On landing and internal web pages: The strange case of jekyll and hyde in web performance measurement," in *Proc. ACM Internet Meas. Conf.*, Oct. 2020, pp. 680–695, doi: 10.1145/3419394.3423626.

[23] S. E. Coull, A. M. White, T.-F. Yen, F. Monrose, and M. K. Reiter, "Understanding domain registration abuses," *Comput. Secur.*, vol. 31, no. 7, pp. 806–815, Oct. 2012, doi: 10.1016/j.cose.2012.05.005.

[24] S. Pletinckx, G. H. Jansen, A. Brussen, and R. van Wegberg, "Cash for the register? Capturing rationales of early COVID-19 domain registrations at Internet-scale," in *Proc. 12th Int. Conf. Inf. Commun. Syst. (ICICS)*, May 2021, pp. 41–48.

[25] R. Kawaoka, D. Chiba, T. Watanabe, M. Akiyama, and T. Mori, "A first look at COVID-19 domain names: Origin and implications," in *Proc. Int. Conf. Passive Active Netw. Meas.*, in Lecture Notes in Computer Science, vol. 12671, O. Hohlfeld, A. Lutu, and D. Levin, Eds. Cham, Switzerland: Springer, 2021, pp. 39–53 doi: 10.1007/978-3-030-72582-2_3.

[26] S. Hao, N. Feamster, and R. Pandrangi, "Monitoring the initial DNS behavior of malicious domains," in *Proc. ACM SIGCOMM Conf. Internet Meas. Conf.*, P. Thiran and W. Willinger, Eds., Nov. 2011, pp. 269–278 doi: 10.1145/2068816.2068842.

[27] M. Korczynski, M. Wullink, S. Tajalizadehkhoob, G. C. M. Moura, A. Noroozian, D. Bagley, and C. Hesselman, "Cybercrime after the sunrise: A statistical analysis of DNS abuse in new gTLDs," in *Proc. Asia Conf. Comput. Commun. Secur.*, Incheon, Republic Korea, J. Kim, G. Ahn, S. Kim, Y. Kim, J. López, and T. Kim, Eds., May 2018, pp. 609–623, doi: 10.1145/3196494.3196548.

[28] J. Feng, L. Zou, O. Ye, and J. Han, "Web2Vec: Phishing webpage detection method based on multidimensional features driven by deep learning," *IEEE Access*, vol. 8, pp. 221214–221224, 2020, doi: 10.1109/ACCESS.2020.3043188.

[29] L. Tang and Q. H. Mahmoud, "A deep learning-based framework for phishing website detection," *IEEE Access*, vol. 10, pp. 1509–1521, 2022, doi: 10.1109/ACCESS.2021.3137636.

[30] S. Abdelnabi, K. Krombholz, and M. Fritz, "VisualPhishNet: Zero-day phishing website detection by visual similarity," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, J. Ligatti, X. Ou, J. Katz, and G. Vigna, Eds., Oct. 2020, pp. 1681–1698 doi: 10.1145/3372297.3417233.

[31] Y. Lin, R. Liu, D. M. Divakaran, J. Y. Ng, Q. Z. Chan, Y. Lu, Y. Si, F. Zhang, and J. S. Dong, "Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages," in *Proc. 30th USENIX Secur. Symp. (USENIX Secur.)*, M. Bailey and R. Greenstadt, Eds. Berkeley, CA, USA: USENIX Association, 2021, pp. 3793–3810. [Online]. Available: https://www.usenix.org/conference/usenixsecurity21/presentation/lin

[32] R. Liu, Y. Lin, X. Yang, S. H. Ng, D. M. Divakaran, and J. S. Dong, "Inferring phishing intention via webpage appearance and dynamics: A deep vision based approach," in *Proc. 31st USENIX Secur. Symp. (USENIX Security)*, Boston, MA, USA, K. R. B. Butler and K. Thomas, Eds., 2022, pp. 1633–1650. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/liu-ruofan

[33] *Urlscan.io*. Accessed: Feb. 3, 2024. [Online]. Available: https://urlscan.io/

[34] *PhishTank*. Accessed: Feb. 3, 2024. [Online]. Available: https://phishtank.org/

[35] *OpenPhish*. Accessed: Feb. 3, 2024. [Online]. Available: https://openphish.com/

[36] *Google Safe Browsing*. Accessed: Feb. 3, 2024. [Online]. Available: https://safebrowsing.google.com/

[37] *Microsoft Defender SmartScreen*. Accessed: Feb. 3, 2024. [Online]. Available: https://learn.microsoft.com/en-us/windows/security/threat-protection/microsoft-defender-smartscreen/microsoft-defender-smartscreen-overview

[38] Chris Welch. (2014). *This Man Just Sold Ebola.Com for $200000*. [Online]. Available: https://www.theverge.com/2014/10/24/7058543/this-man-just-sold-ebola-for-200000

[39] K. Tian, S. T. K. Jan, H. Hu, D. Yao, and G. Wang, "Needle in a haystack: Tracking down elite phishing domains in the wild," in *Proc. Internet Meas. Conf.*, Boston, MA, USA, Oct. 2018, pp. 429–442. [Online]. Available: https://dl.acm.org/citation.cfm?id=3278569

[40] R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras, "A high-performance, scalable infrastructure for large-scale active DNS measurements," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 6, pp. 1877–1888, Jun. 2016.

[41] V. Le Pochat, T. van Goethem, S. Tajalizadehkhoob, M. Korczynski, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2019, pp. 1–15.

[42] *Resource Development*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/tactics/TA0042/

[43] *Acquire Infrastructure (T1583)*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/techniques/T1583/

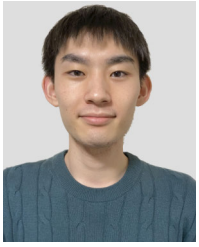
[44] *Spearphishing via Service (T1566.03)*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/techniques/T1566/003/

[45] *Domain Registration (T1583.001)*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/techniques/T1583/001/

[46] *Compromise Infrastructure (T1583.002)*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/techniques/T1583/002/

[47] *Drive-by Compromise (T1189)*. Accessed: Feb. 3, 2024. [Online]. Available: https://attack.mitre.org/techniques/T1189/

[48] *Content Spoofing (CAPEC-148)*. Accessed: Feb. 3, 2024. [Online]. Available: https://capec.mitre.org/data/definitions/148.html



RYO KAWAOKA received the B.E. and M.E. degrees in computer science and engineering from Waseda University, in 2021 and 2023, respectively. He has been engaged in the measurement study of DNS.

DAIKI CHIBA (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in computer science from Waseda University, in 2011, 2013, and 2017, respectively. Since 2013, he has been with Nippon Telegraph and Telephone Corporation (NTT), where he has been engaged in research on cyber security through data analysis. He is currently the Senior Manager of NTT Security (Japan) KK, Tokyo, Japan. He is a member of IEICE.

TAKUYA WATANABE received the B.E. and M.E. degrees in computer science and engineering and the Ph.D. degree in engineering from Waseda University, in 2014, 2016, and 2020, respectively. He was with Nippon Telegraph and Telephone Corporation (NTT), in 2016, he was engaged in research of system security and privacy from an attacker's perspective, especially web and mobile. He is currently with the Cyber Security Project of the NTT Social Informatics Laboratories.

MITSUAKI AKIYAMA (Member, IEEE) received the M.E. and Ph.D. degrees in engineering from the Nara Institute of Science and Technology, in 2007 and 2013, respectively. He was with Nippon Telegraph and Telephone Corporation (NTT), in 2007, he was engaged in research and development on cybersecurity. He is currently a Senior Distinguished Researcher with the NTT Social Informatics Laboratories. His research interests include cybersecurity measurement, offensive security, and usable security and privacy. He is a Senior Member of IPSJ and a member of SIGCHI and IEICE. He received the Cybersecurity Encouragement Award from the Minister for Internal Affairs and Communications, in 2020, and the IPSJ/IEEE Computer Society Young Computer Researcher Award, in 2022.

TATSUYA MORI (Member, IEEE) received the B.E. and M.E. degrees in applied physics and the Ph.D. degree in information science from Waseda University, Tokyo, Japan, in 1997, 1999, and 2005, respectively. He is currently a Professor with Waseda University. He joined the NTT Laboratory, in 1999. Since then, he has been engaged in the research of measurement and analysis of networks and cyber security. From March 2007 to March 2008, he was a Visiting Researcher with the University of Wisconsin-Madison. Since May 2018, he has been a Guest Researcher with the Cybersecurity Research Institute, National Institute of Information and Communications Technology (NICT). Since April 2019, he has been a Visiting Researcher with the Center for Advanced Intelligence Project, RIKEN (RIKEN AIP). He received the Telecom System Technology Award from TAF, in 2010, and the Best Paper Awards from IEICE Transactions, IEEE/ACM COMSNETS, ATIS, NDSS, and EuroUSEC, in 2009, 2010, 2017, 2020, and 2021, respectively. He is a member of ACM, IEICE, and IPSJ.