東京2020オリンピック公式ドメイン名に対する類似ドメイン名の実態調査

河岡 諒[1,a)]　千葉 大紀[2]　渡邉 卓弥[2]　秋山 満昭[2]　鈴木 宏彰[1,†1]　森 達哉[1,3]

概要：2021 年夏に開催された**第 32 回オリンピック競技大会（2020/東京）** の公式サイトに用いられたドメイン名 tokyo2020.org に類似したドメイン名 (以下，類似ドメイン名) を調査した結果を報告する．2020 年 1 月から 2021 年 8 月にかけて長期間収集したドメイン名データを分析した結果，以下の結果を得た．(1) 2021 年 7 月 23 日の開会式の時点で 1,825 件のアクティブな類似ドメイン名が登録されていた．(2) 調査開始の 2020 年 1 月から累計で 4,158 件のユニークな類似ドメイン名が登録された．(3) 1 年の延期決定を受けて，文字列 2021 を含む類似ドメイン名が大量に登録された．(4) 類似ドメイン名を用いた悪性サイトは開会前から常に存在し続けたが，特に開会式が近づいた 7 月に多くの悪性判定サイトが検出された．(5) 悪性判定サイトの多くはライブ配信サイトを偽装したものであった．得られた調査結果をもとに，世界的な巨大イベントにおけるドメイン名の適切な運用・管理方法に関して考察する．

キーワード：ドメイン名, 東京 2020 オリンピック，フィッシング

Analysis of the Domains that Look Like the Domains of Official Tokyo 2020 Olympic Games

Ryo Kawaoka[1,a)]　Daiki Chiba[2]　Takuya Watanabe[2]　Mitsuaki Akiyama[2]　Hiroaki Suzuki[1,†1]
Tatsuya Mori[1,3]

Abstract: This paper reports the results of an analysis of domain names that look similar in appearance to the domain name tokyo2020.org used for the official website of the 32nd Olympic Games (2020/Tokyo) held in the summer of 2021 (hereinafter referred to as "similar domain names"). We analyzed domain name data collected from January 2020 to August 2021, and revealed the following results. (1) At the time of the opening ceremony on July 23, 2021, 1,825 active similar domain names had been registered. (2) A total of 4,158 unique similar domain names have been registered since the measurement started in January 2020. (3) Following the decision to postpone the event for one year, a large number of similar domain names containing the substring 2021 were registered. (4) Malicious websites using the similar domain names have always existed since the beginning of the measurement study. Number of malicious websites increased especially in July when the opening ceremony was approaching. (5) Most of the malicious websites were disguised as live streaming sites. Based on the obtained findings, we discuss the appropriate operation and management of domain names that will be used for large-scale global events held in future.

Keywords: domain name, Tokyo 2020 Olympics, Phishing

1. はじめに

2021 年 7 月 23 日から 8 月 8 日にかけて**第 32 回オリンピック競技大会（2020/東京）** (以降，**東京 2020 オリン**

[1] 早稲田大学 / Waseda University
[2] 日本電信電話株式会社 / NTT
[3] 情報通信研究機構 / NICT
[†1] 現在，NTT コミュニケーションズ株式会社
Presently with NTT Communications Corporation
[a)] k-hsw119@nsl.cs.waseda.ac.jp

ピックと表記[*1]) が開催された．同大会は，COVID-19 の影響により 1 年の延期を経ての開催となった．また，主会場となった国立競技場をはじめ，大多数の会場で無観客となったため，結果として過去に比類のない大規模なオンラインイベントとなった．実際，Akamai Networks 社は，多数のメダル決定がかかった競技最終日に，同社による競技のストリーミング配信トラフィック量が 10 Tbps に達したことを報告している [1]．

オリンピックは世界的に注目が集まる巨大なイベントであるため，チケットやグッズの販売，旅行・宿泊予約，放送・配信，ニュース速報等に関連した，様々なウェブサイトが提供される．その一方で，オリンピックは様々なサイバー攻撃のターゲットとなる [2]．Cloudflare 社は，東京 2020 オリンピックの競技が始まって以降，日本へのサイバー攻撃が前週比で 10 倍以上に増加したことを報告している [3]．

オリンピックに関連したサイバー攻撃の対象は，組織委員会の情報資産，開会式の妨害，大会公式サイトへの攻撃等，様々なチャネルに及ぶ．その中にあって，マスユーザをターゲットとしたサイバー攻撃の典型は，オリンピック公式サイトを騙るフィッシング攻撃である．公式サイトはチケットやグッズの販売の他，スケジュール情報を提供するものであるが，本大会に特異なこととして，チケットの返金が発生したため，オンライントランザクションが発生する機会が増えている．また，無観客開催，および COVID-19 がもたらした「ニューノーマル」により，オンライン配信を視聴する環境が普及し，インターネットのライブ配信を視聴する要求が高まった．したがって，攻撃者がオリンピック公式，およびライブ配信等の関連サイトをターゲットとしたフィッシング攻撃を実行する動機は十分に存在する．

上述した背景を元に，本研究は東京 2020 オリンピック公式サイト `tokyo2020.org` の類似ドメイン名に着目し，特に悪性と分類されるドメイン名の登録状況の長期的な変遷と，会期中の変化，および悪性ドメイン名の具体的な事例を調査する．このような分析は，オリンピックや万博などの世界的に注目度が高い巨大なイベントをターゲットとしたフィッシング攻撃の手口や，パターンを理解するのに有益な知見を与えることが期待される．本研究は，将来にイベントを開催する組織に対し，フィッシング攻撃の被害を防ぐために講じることが望ましいセキュリティプラクティス (取得ドメイン名の管理や，その周知方法なども含む) を提供することを狙いとする．

本研究の主要な貢献，および発見は以下の通りである．

- 1.5 年以上の長期にわたり，世界的な巨大イベント（オリンピック）に関するドメイン名とその悪用を調査・分析した初の研究である
- 類似ドメイン名を用いた悪性サイトはオリンピック開会式の開催日に合わせて急激に増加した
- 類似ドメイン名を用いた悪性サイトの多くはライブ配信サイトであった

[*1] 大会名称は知的財産であり保護対象となるが，論文に記載することに関して，東京オリンピック・パラリンピック競技大会組織委員会ブランド管理部に確認し，問題がないとの回答を得ている．

2. 検出手法とデータの収集

本節は，東京 2020 オリンピック類似ドメイン名の検出方法，類似ドメイン名の収集方法，ならびに類似ドメイン名を用いたウェブサイト情報の収集方法を示す．

2.1 類似ドメイン名の検出方法

東京オリンピック公式サイトの類似ドメイン名を検出する手法として，キーワードマッチングを行う．具体的には，以下に示す条件の内，2 つ以上を満たしたら類似ドメイン名として判定する．

- `tokyo` が含まれる，ただしトップレベルドメイン (TLD) としての `.tokyo` は除く
- `olympic` が含まれる
- `2020`，`2021`，`2022`，`2023`，`2024` が含まれる
- `ticket` が `tokyo` とともに含まれる，ただし TLD としての `.tokyo` は除く
- `ticket` が `olympic` とともに含まれる

上述の条件により，`tokyo-olympic2020[.]net` などのドメイン名が検知される．この条件はオリンピックとはまったく無関係のドメイン名を誤検出する可能性があるが，事前の検査により，そのようなケースはきわめて稀であることが判明している．なお，2020 から 2024 の数字は開催年に相当するが，当初の開催年 2020 に加え，実際の開催年 2021 および，その後のパリ五輪開催予定年の 2024 までの年を含めた．

次に国際化ドメイン名 (IDN) を用いた IDN ホモグラフに対応するために，ドメイン名が ASCII 文字に類似した非 ASCII 文字（ホモグリフ）を含む場合，該当ホモグリフを対応する ASCII 文字に置換する．そのような置換を行うために，先行文献 [4] のホモグリフ文字リスト (simchar.json)，およびと Unicode コンソーシアムが提供する類似した文字の組み合わせリスト [5] を利用する．この処理により，`tōkyō2020[.]com` が `tokyo2020[.]com` に変換され，キーワードマッチングが可能となる．

最後にタイポスクワッティングへの対応を行う．複雑さを避けるため，公式ドメイン名に対するタイポスクワッティングのみを考慮する．具体的には，文字列 `tokyo2020`，`tokyo2021`，`tokyo2022`，`tokyo2023`，`tokyo2024` に対して，以下の操作をすべての組み合わせで繰り返すことにより，タイポスクワッティングリストを生成する．

- 1 文字削除

- 隣り合う文字を置換
- 1 文字選択し，その文字を QWERTY 配列で隣接するキーの文字に置換
- 1 文字選択し，同じ文字を直後に挿入
- 1 文字選択し，その隣接するキーの文字を直後に挿入

上記で生成したタイポスクワッティングリストがドメイン名の部分文字列となっている場合，類似ドメイン名として検出する．例えば，`toyko2020[.]org` などが検知される．

2.2 類似ドメイン名の収集

類似ドメイン名の抽出元のドメイン名のソースとして，大規模な商用データベースである Zonefiles.io [6] を利用した．このデータベースは約 1,500 個の DNS ゾーンファイルを定期的に転送することにより更新されており，約 2.6 億件のドメイン名を含んでいる．このデータベースを毎日ダウンロードし，2.1 節にて説明した検知プログラムを適用することにより，その日に登録されている東京 2020 オリンピック類似ドメイン名を収集した．

2.3 ウェブサイト情報の収集

東京 2020 オリンピック類似ドメイン名の利用実態を調査するために，2.2 節にて収集した類似ドメイン名の DNS レコードの確認と HTTP/HTTPS のアクセスとオンラインスキャンサービスによるスキャンを調査期間中に実施した．

まず各ドメイン名の DNS の A レコード，NS レコード，SOA レコード，CNAME レコードを確認した．次に，A レコードが存在したものについては IP アドレスが割り当てられているので HTTP/HTTPS のリクエストを送信し，レスポンスやコンテンツの収集を試みた．具体的には，各ホストの 80 番ポートと 443 番ポートに接続可能かを確認し，接続可能であれば HTTP/HTTPS のリクエストを送信し，レスポンスやスクリーンショット，そのほかメタデータの記録を行った．このとき，コネクションタイムアウトや証明書に関するエラーが発生するものも存在したがそれらはコンテンツ収集に失敗したと記録した．ウェブサイトへのアクセスは Google Chrome の version 86.0.4240.0 を用い，ユーザーエージェントを Desktop の Windows 10 に設定したもの (Desktop) と iPhone の iOS 12.2 に設定したもの (Mobile) に設定を行ったもののそれぞれで HTTP/HTTPS のアクセスを行い，一つのドメイン名に対して合計で 4 種類のアクセスを行った．最後に A レコードの存在したドメイン名とウェブサイトへのアクセスにおけるリダイレクト先のドメイン名 (異なるドメイン名にリダイレクトしていた場合のみ) に対してオンラインスキャンサービスの VirusTotal [7] を用いて悪性判定を行い，結果を記録した．

本研究は長期間に渡る調査であり数多くのデータを時系列的に収集するものであった．収集状況や現在の状態を経時的にモニタリングできるように図 1 に示すようなダッシュボードを作成し，適宜状況を確認できるようにした．ダッシュボードの作成には，Grafana [8] を利用した．

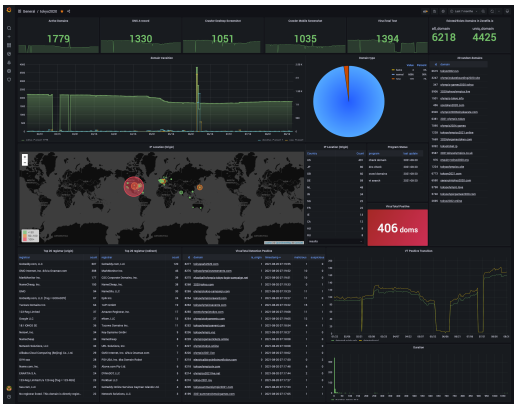

図 1 類似ドメイン名ダッシュボードのスクリーンショット

2.4 調査期間

本研究では 3 つの期間にてデータ収集を行った．データの信頼性や網羅性を向上させるために収集方法やデータソースを順次ブラッシュアップさせており，その方法ごとに 3 つの期間に分かれている．それぞれの期間とその収集条件について説明する．また，マシン関係の事情により収集期間内でもデータ収集が行えていない日付もありるため，収集日数についても記載する．

期間 1: 2020/01/01〜2020/05/15 期間 1 は 2020 年 1 月 1 日から 2020 年 5 月 15 日まで行った調査で，133 日分の収集結果からなる．調査開始初日に Zonefiles.io に含まれる東京 2020 オリンピック類似ドメイン名を抽出し，その日以降は毎日追加されたドメイン名と削除されたドメイン名の差分データを取得し，そのデータから類似ドメイン名を抽出し前日までのデータに追加や削除を行うことで登録数を算出した．また，VirusTotal による悪性スキャンは DNS の A レコードが存在するものに対して毎日行った．

期間 2: 2020/10/01〜2020/12/28 期間 2 は 2020 年 10 月 1 日から 2020 年 12 月 28 日まで行った調査で，89 日分の収集結果からなる．類似ドメイン名の登録数の算出方法は期間 1 と同一であるが，VirusTotal による悪性スキャンは DNS の A レコードが存在するものに対して 2 日に 1 回行った．

期間 3: 2020/12/29〜2021/08/22 期間 3 は 2020 年 12 月 29 日から 2021 年 8 月 22 日まで行った調査で，227 日分の収集結果からなる．この期間では Zonefiles.io に含まれるドメイン名全件のリストを毎日取得し，そのデータから東京 2020 オリンピック類似ドメイン名を抽出することにより登録数を算出した．このデータ利用方法の変更はより網羅性を高めるために行った．また，VirusTotal による悪性スキャンは DNS の A レコードが存在するものに対

して毎日行った．

3. 分析と結果

本節では，収集したデータをもとに登録数や悪性判定数の時系列変化を分析する．また，収集したスクリーンショットからどのようなウェブサイトが存在したのかについてケーススタディとしてライブ配信サイトを紹介する．

3.1 集計方法

ここでは，収集データの集計方法について説明を行う．

まず，収集した類似ドメイン名に対する調査結果を用いて，日毎の登録数・A レコード・ウェブサイトへのアクセス結果の集計を行った．登録数は日毎の類似ドメイン名の数を集計することにより算出し (**登録数**)，A レコードについては DNS の A レコードが存在する類似ドメイン名の数を集計し (**A レコード**)，ウェブサイトへのアクセス結果は Desktop もしくは Mobile のクローラにて HTTP もしくは HTTPS のアクセスにて応答コードとして 200 OK を返したものを集計した (**Website**)．

次に，オンラインスキャンサービスの VirusTotal [7] によるスキャン結果の集計を行った．VirusTotal の URL スキャンでは 1 件の URL のスキャンに対して，約 90 個のオンラインスキャナによる悪性判定結果が得られる．類似ドメイン名に対するスキャン結果を用いて，1 つ以上のスキャナによって悪性判定されたドメイン名の数 ($\mathbf{VT}_{\geq 1}$) と 2 個以上のスキャナによって悪性判定されたドメイン名の数 ($\mathbf{VT}_{\geq 2}$) を集計した．ここで，登録中のドメイン名のうち A レコードのないものは IP アドレスが紐づいておらず利用されていないと判断しスキャンを行わず，2 個以上のスキャナによる結果集計はスキャン結果の確度を高めるために行った．

3.2 開会式当日と閉会式当日の結果と累積数

集計結果について，区切りとなる開会式当日と閉会式当日についてのデータ，また，2021 年 8 月 22 日時点での累積数について表 1 に示す．ここで，累積に関してはドメイン名に対してユニークを取り，**A レコード**，**Website**，$\mathbf{VT}_{\geq 1}$，$\mathbf{VT}_{\geq 2}$ が 1 度でも条件を満たしたものをカウントした．さらに，2021 年 8 月 22 日時点までに観測された類似ドメイン名に対して，その検知タイプごとの件数を表 2 に示す．

表 1 から東京 2020 オリンピック終了時点で 1,854 件の類似ドメイン名が登録されており，累積では 4,000 件を超える類似ドメイン名が観測されている．また，表 2 より，その多くは通常のキーワードマッチングで検出されたドメイン名であることも確認できる．一方，数としては少ないがより高度な類似ドメイン名として，タイポスクワッティングおよび IDN ホモグラフを利用したケースも観測された．

表 1 開会式当日と閉会式当日の結果と累積数

項目	開会式当日 [a]	閉会式当日 [b]	累積 [c]
登録数	1,825	1,854	4,158
A レコード	1,371	1,405	2,506
Website	1,218	1,253	2,458
$\mathrm{VT}_{\geq 1}$	148	162	435
$\mathrm{VT}_{\geq 2}$	48	57	94

[a] 2021/07/23　[b] 2021/08/08

[c] 2021/08/22 までのユニークな累積数

表 2 類似ドメイン名の検出結果

検出タイプ	件数
キーワードマッチング	4,060
タイポスクワッティング	96
IDN ホモグラフ	2
合計	4,158

3.3 類似ドメイン名登録数の遷移

以下では，1 年 8 ヶ月に渡る観測期間全体での長期的な傾向，ならびに開催期間における傾向を分析する．

長期的な傾向　調査期間全体 (2020/01/01〜2021/08/22) における類似ドメイン名の登録数の時系列を図 2 に示す．本調査を開始した 2020 年 1 月はオリンピック開催予定日の約半年前にあたるが，この時点では約 900 件の類似ドメイン名が登録されていた．なお，本調査に先駆けて著者らが 2019 年 5 月に実施した予備調査 [9] では，956 件の類似ドメイン名を観測しており，当初の大会開催予定の 1 年前から，半年前にかけては大きな変化はなかったことがわかる．図より，登録されたドメイン約 2/3 に対して IP アドレスが割り当てられ，ウェブサイトとしてアクセス可能な状態にあること，そしてその比率は全体を通じて大きな変化はなかったことがわかる．

調査開始の 2020 年 1 月の時点では，COVID-19 の影響は限定的であったため，オリンピックは計画通り開催される予定であった．その後，COVID-19 の世界的なパンデミックが進み，2020 年 3 月 24 日に東京オリンピック・パラリンピックの延長方針が合意された．同時期に大量の類似ドメイン名が登録されており，特に 4 月 6 日には 467 件の登録があった．この内 321 件は 2021 という文字列を含んでいたため，1 年後の延期を受けた投機目的のドメイン名登録であった可能性が高い．その後はデータが取得できていない期間が半年ほど続いたが，類似ドメイン名の登録数は 2,000 件弱で推移しており，大きな変化はなかったと考えられる．

前述したように，2021 年 1 月以降の**期間 3** では，より網羅的なデータ収集方法に変更したため，登録数にギャップが生じているが，登録数のトレンドとしては大きな変更はなく，期限を迎えたドメイン名が徐々に登録削除となり，

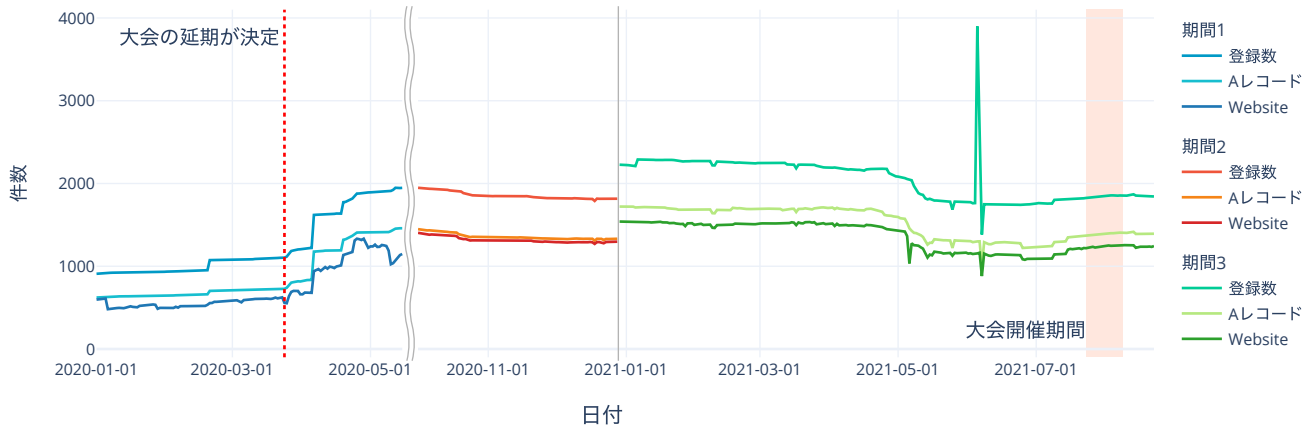

図 2　アクティブな類似ドメイン名登録数の変遷．2021 年 1 月以降の**期間 3** では，データ集計方法を変更したため，ギャップが生じていることに注意

緩やかな減少傾向が続いた．とくに 2021 年 5 月に登録数が顕著に減少しているのは，2020 年 4 月に大量登録されたドメイン名の多くが期限までに登録更新をしなかったため，自動更新猶予期間 (Auto Renew Grace Period) を経て登録削除に至ったと考えられる．実際，2021/5/7 から 2021/5/17 にデータベースから削除された 294 件のうち，81%が 2020 年の 3 月末から 4 月末にかけて登録されたものであった．なお，2021 年 6 月 5 日に大きなスパイクが観測されたが，このスパイクはデータソースの不備に起因するものであると考えられる．同日のデータソースには平常時の約 1.7 倍のドメイン名が含まれていたが，そのうちの大多数は翌日以降削除されたことが確認された．

上述したように類似ドメイン名の登録数はしばらく減少傾向が続いたが，大会開催 3 週間前の 7 月頃から増加傾向に転じた．とくに 7 月 9 日に顕著な増加傾向が続き，以降で見るように大会期間中も増加した．

大会開催期間の傾向　オリンピックの開催期間は 2021 年 7 月 23 日 (開会式当日) から 2021 年 8 月 8 日 (閉会式当日) までの 17 日間であった．この期間における 2 日おきの集計データを表 3 に示す．開会式の日と閉会式の日で比較すると，類似ドメイン名の登録数，および有効なウェブサイトの数がいずれも 30 件程度増加した．この期間に新規登録されたドメイン名の内，9 件は悪性判定されたが，その内の 3 件は五輪サッカーのカードゲームを装ったフィッシングサイト[*2]，2 件はパーキングドメイン，その他はアクセス字に空白やデフォルトのページであった．

3.4　悪性判定サイトの遷移

類似ドメイン名の内，悪性判定されたドメイン名の数を集計した結果を図 3 に示す．一般に $VT_{\geq 1}$ は 1 件のスキャナに検知された結果に大きく影響を受けるため，該当するスキャナの検知ルールの変更により，大きく変化する．一方，$VT_{\geq 2}$ は検知するスキャナの数が増えた分，比較的安定した統計を示すことが知られている．実際，グラフからもそのような傾向が見てとれる．以降は，$VT_{\geq 2}$ を用いて，悪性判定サイトを分析する．

表 3　登録数と VirusTotal 検知数の遷移: 大会中の結果

日付	登録数	A レコード	Website	$VT_{\geq 2}$
2021/07/23	1,825	1,371	1,218	48
2021/07/25	1,835	1,379	1,227	53
2021/07/27	1,843	1,388	1,226	53
2021/07/29	1,851	1,393	1,239	54
2021/07/31	1,852	1,395	1,241	56
2021/08/02	1,853	1,398	1,251	57
2021/08/04	1,856	1,399	1,246	59
2021/08/06	1,853	1,404	1,249	59
2021/08/08	1,854	1,405	1,253	57

悪性サイトは観測期間を通じて増減を繰り返しているが，全般的な傾向に関しては緩やかな増加傾向にあった．そして，大会開催が近づいた 7 月以降に急激な増加があることが見て取れる．収集したスクリーンショットの分析により，多くの悪性判定サイトは，オリンピックの動画配信やライブ配信サイト，スポーツに関する情報を提供するサイトと見てとれた．実際，84 件の悪性判定サイトの内，21 件はライブ配信関連であった．動画やライブ配信に関連した悪性判定サイトが多い事実は，東京 2020 オリンピックにおいてほぼ全ての競技が無観客開催であり，テレビやインターネットを用いて観戦するケースが多い状況を反映した結果であると考えられる．検出した悪性判定サイトの具体的な事例は次節に示す．

3.5　悪性サイトの事例: ライブ配信サイト

東京 2020 オリンピック類似ドメイン名に特有な悪性サイトとして数が多かった，偽ライブ配信サイトの例を示

[*2] https://twitter.com/KesaGataMe0/status/1421174931576872960 で情報が提供されている．

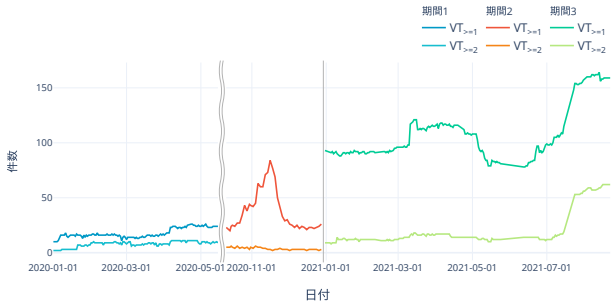

図 3 VirusTotal による検出結果

す．図 4，図 5，図 6 に観測した偽ライブ配信サイトのスクリーンショットを示す．

図 4 はアカウント登録機能を持つ偽ライブ配信サイトで 2021 年 6 月 27 日にデータソース内に登場した．スクリーンショットを時系列で見ると，7 月 24 日まではドメインパーキングのページが表示されており，25 日に現在のウェブサイトの構築途中のページが表示され，翌日から図 4 に示すウェブサイトが表示されていた．VirusTotal による検知数を見ると，7 月 22 日までは検知数は 0 であったが，23 日に初めて検知数が 3 となり，29 日以降さらに検知数が増加し，8 月 4 日以降は 10 以上のスキャナによって検知されている．

図 5 も偽ライブ配信サイトであり，2021 年 7 月 13 にデータソース内に登場した．時間軸でデータを見ると，7 月 24 日に DNS の A レコードが登録され，同日に現在のウェブサイトの構築途中のページが表示された．翌日以降は図 5 のウェブサイトが表示されるようになった．このウェブサイトは，大会開催期間中の 7 月 26 日以降，Desktop と Mobile で表示内容を分けていたことである．具体的には Desktop では矢印の画像のみが表示されるページを提供していたが，Mobile では図 5 に示すコンテンツが表示された．VirusTotal による検知数を見ると，7 月 24 日から 28 日までは検知数は 2 であったが，7 月 30 にかけて 4 に増加し，7 月 31 日以降は 10 以上のスキャナにより検知されている．

図 6 は VPN 経由で視聴する偽ライブ配信サイトであり，2021 年 6 月 28 日にデータソース内に登場した．登場したその日から図 6 に示すコンテンツを表示している．VirusTotal による検知数は 7 月 7 日初めて 1 つのスキャナに検知され，7 月 15 日にかけて 10 に増加し，その日以降は 10 以上のスキャナにより検知されている．本ウェブサイトは，ライブ配信を視聴するために VPN 接続を促す特徴があり，VPN サービスのアフィリエイトプログラムを通じて偽ライブ配信のマネタイズを実現している．

以上で示したように，悪性サイトの登場から活動開始に至る状況，およびオリンピックの開催期間に合わせてライブ配信サイトを装った悪性判定サイトが準備された様子を

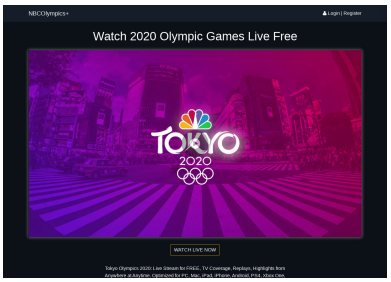

図 4 tokyo---olympics[.]org のスクリーンショット

図 5 olympics----tokyo[.]com のスクリーンショット

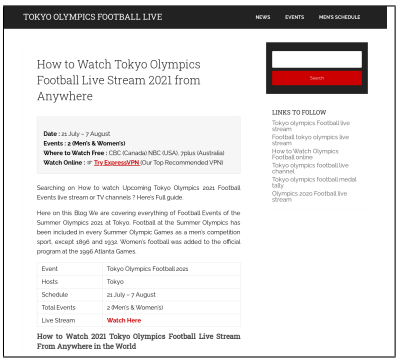

図 6 tokyoolympicsfootballlive[.]com のスクリーンショット

捉えることができた．

4. 議論

4.1 公式ドメイン名の運用・管理に関する考察

本研究の調査により，類似ドメイン名を用いた悪性サイトは大会開催前から存在し，開催が近づくにつれて増加することがわかった．類似ドメイン名はその定義から公式サイトのドメイン名と類似しているため，一般のユーザがドメイン名から真贋性を判定することは困難である．類似ドメイン名を用いた悪性サイトによる被害を防止する有効な対策の 1 つは，組織委員会が利用するドメイン名を限定した上で，公式ドメイン名の周知徹底図ることである．一方，公式ドメイン名が増えると混乱を招くため，情報提供，チケット予約管理，グッズ販売等，用途別のドメイン名はそのサブドメインとして運用することが望ましい．このような運用により，正規のドメイン名とそうではないドメイン名の識別が容易となる．2024 年に開催予定のパリ 2024 オリンピックの公式サイトはドメイン名 paris2024.org を採用しているが，エントリサイトは www.paris2024.org，プレスは press.paris2024.org，物

販は boutique.paris2024.org などのような使い分けがなされている．

東京オリンピック組織委員会は，当初公式サイトは tokyo2020.org であると周知したが，tokyo2020.jp との使い分けが不透明であったり，2021 年 4 月 29 日に国際オリンピック委員会 (IOC) が管理する olympics.com に転送されるような運用に変更されたため [10]，一貫性がある運用とは言い難い状況にあった．パリオリンピックに関しても，2021 年 8 月現在において boutique.paris2024.org は shop.olympics.com に転送される状況にある．今後はこのような公式ドメイン名の運用を一貫性がある形で統一し，周知していくことが肝要である．また，現時点で olympic.com など，IOC の公式ドメイン名に対する類似ドメイン名が多数存在することにも注意が必要である．開催地ごとの運用に合わせた一定の自由度を保ちつつ，長期的な一貫性を確立し，かつ公式サイトとしての識別容易性を実現するためには，オリンピック専用の gTLD を作り，ドメイン名登録に制限をつける（組織委員会および許可を得た組織に限定する）ことも 1 つの解決策であろう．

オリンピックの起源は約 2800 年前の古代に遡ることができ，今後も人類の歴史が続く限り継続すると考えられる世界規模のイベントである．IOC が保有するドメイン名に関する類似ドメイン名，各開催地に固有なドメイン名等あわせて長期的に監視を続けていくこと，運用・管理のベストプラクティスを共有していくことが必要であろう．

4.2 制約

以下では，本研究の制約事項や調査範囲の限界を議論する．

類似ドメイン名の検出方法　本研究では 2.1 節に示したキーワードマッチングにより，東京オリンピック類似ドメイン名を抽出した．この条件では，見逃されるドメイン名が存在するケースがある，例えば文字列 ticket のみを含むドメイン名で東京 2020 オリンピックのチケットに関する内容のコンテンツを配信するドメイン名などは含まれない．また，国内では民放オリンピック公式動画サイト gorin.jp が広く認知されているが，このような日本語に特有なオリンピック関連ドメイン名に対する類似ドメイン名は今回の検出対象外である．英語以外の単語を条件に取り込むことは，今後の課題である．類似ドメイン名を抽出する条件やアルゴリズムは，さらに向上することができる余地があり，今後の課題としたい．

コンテンツ収集方法　ドメイン名から URL を作成し，HTTP/HTTPS によるウェブサイトへのアクセスとそのコンテンツ収集を行う段階でも制約が存在する．ウェブサイトへのアクセスはそのルートディレクトリにのみ接続を試みコンテンツの収集を行ったため，ルートディレクトリにインデックスファイルがない場合はウェブサイトのコンテンツを見逃している可能性がある．これは URL のパスを探索することにより低減できる可能性はあるが，それを行うことは一般的には困難である．

データソース　この研究ではドメイン名のデータソースとして Zonefiles.io の提供する大規模なドメイン名のデータベースを利用した．このデータベースは DNS のゾーンファイルを転送することにより作られているが，そのためサブドメインに関する情報が含まれていない．これらの問題は別のデータソースを活用することで解決できる可能性がある．例えば，サブドメインに関する情報は CT-Log などの TLS 証明書のデータを利用することで取得することが可能な場合もある．

また，ドメイン名が登録されているかは Zonefiles.io のデータベースに存在するかで行ったが，厳密にはゾーンファイルに登録状況が反映された上で Zonefiles.io に転送されるまでの時間がかかるため，数日の誤差が生じる可能性がある．ドメイン名の登録日に関しては WHOIS の登録日を確認することで正確な日付を知ることができるが，本調査では登録状況を日々モニタリングすることが主眼であったためデータベースに存在するかで判断した．

4.3 倫理的配慮

我々の研究では，公開されている DNS レコードやドメイン名に対応するウェブコンテンツの分析を行い，個人を特定できるような情報は利用しなかった．また，ウェブクローリングの過程ではコンテンツ収集に必要な最低限のアクセスに留めた．具体的には HTTP および HTTPS の正当なリクエストを Google Chrome の Desktop 版と Mobile 版それぞれで送信することで 1 回のクローリングにおける 1 サイトあたりのアクセス数を最大 4 回に限定し，1 日 1 回のみのクローリングにすることによりウェブサイトに悪影響を与えないよう配慮した．

5. 関連研究

5.1 イベント駆動型のドメイン名登録

現在進行中のイベントに関連したドメイン名を早期に取得する戦略は，ドメイン名ビジネスにおいてよく知られた手法である．実際にこのような技術の特許がドメインレジストラによって出願されている [11]．発生したイベントに関連するドメイン名登録はドメイン名ビジネスでは広く行われている手法であるが，このトピックに関連するセキュリティの観点での研究はほとんど行われていない．Coull ら [12] は，発生したイベントを Google 検索の人気なクエリから見つけるルールを導き出し，投機的なドメイン名登録の特徴付けを行い，さらにドメイン名取得の実現可能性を検証している．彼らの用いた手法は特定の短期間のイベントを想定したものであるが，我々の研究は大規模なドメイン名のデータベースからドメイン名を抽出し，約 1.5 年

間に渡り調査を行った点でこの先行研究とは異なっている．

スポーツや国際的なイベントに関するセキュリティの研究として，Nakano ら [13] は，スポーツ大会や新型 iPhone の発売に代表されるような現実世界のイベントと同期してソーシャルプラットフォーム上で拡散される攻撃を精度よく検知する手法を提案している．

世界的影響がきわめて高い事象として，COVID-19 に注目したドメイン名の研究事例として文献 [14] が挙げられる．同研究では，約 165 万件の COVID-19 に関連するドメイン名の分析を通して COVID-19 関連ドメイン名の登録実態や利用目的を調査している．文献 [14] と本論文は，特定の文字列を含むドメイン名の調査という観点で共通点が多いが，オリンピックには明確なイベントの開催日がある点，およびイベントの性格がまったく異なる点に大きな差異があり，結果として観測された結果や得られた知見もまったく異なるものである．

5.2 悪意のあるドメイン名とウェブサイト

悪意のあるドメイン名の登録と初期の活動を観測する方法については，多くの研究が行われてきた [15], [16], [17]．Hao ら [15] は，DNS のインフラや DNS 名前解決のパターンが正規のドメイン名と悪意のあるドメイン名では大きく異なることを明らかにした．Korczynski ら [16] は，11 個の脅威情報フィードから得られたドメイン名に対して，対応する WHOIS 情報・ウェブコンテンツ・DNS レコードを収集し分析を行うことで，新しい gTLD でのスパムドメインが増加していることを示した．我々の調査はこれらの先行研究における手法を参考にしながら行った．これらの研究は，特定の所有者の存在するブランド名をドメイン名に含む偽ドメイン名をキーワードマッチングにより抽出・分析している．我々の研究は，単純なキーワードのマッチングに加え，IDN ホモグラフ，およびタイポスクワッティングに拡張したドメイン名の抽出を行っている点に技術的な差異がある．

6. まとめ

東京 2020 オリンピック公式サイト `tokyo2020.org` の類似ドメイン名の登録状況を 2020 年 1 月から 2021 年 8 月にかけて調査した．本大会は COVID-19 の影響を受けて 1 年延期されたため，変則的な状況にあったが，調査により大会開催に先駆けて多数の類似ドメイン名が登録されること，延期決定のタイミングで投機目的と考えられるドメイン名登録が多数発生したこと，類似ドメイン名を用いた悪性判定サイトは開催日が近づくタイミングで急増することが明らかになった．また，オンライン需要の高まりを受けて，悪性判定サイトの中にはライブ配信サイトが散見されたことも特筆すべき調査結果であった．類似ドメイン名は公式サイトに使われるドメイン名との見分けがつきにくいため，一般のユーザがその真贋性を見分けることは困難である．本調査で得られた知見・考察が，オリンピックをはじめ，将来開催される世界的な大規模イベントにおけるドメイン名の適切な運用・管理に資することを期待する．

謝辞 東京オリンピックの類似ドメイン名の調査を開始するきっかけを作っていただいた NHK の斉藤直哉氏，および本論文の執筆に着手するきっかけを作っていただいた共同通信社の角亮太氏に感謝します．本調査の初期検討時に貴重な助言を頂いた日本レジストリサービスの米谷嘉朗氏に感謝します．

参考文献

[1] Akamai: Tokyo Streaming Traffic Runs Rings Around Rio, `https://blogs.akamai.com/2021/08/tokyo-streaming-traffic-runs-rings-around-rio.html`.

[2] Redscan: Dangerous games: the cyber security threats to the Olympics, `https://www.redscan.com/news/cyber-security-threats-tokyo-olympics-2020/`.

[3] ITmedia: 日本へのサイバー攻撃、五輪開始後は約 10 倍に　組織委「対策を徹底する」, `https://www.itmedia.co.jp/news/articles/2107/29/news132.html`.

[4] Suzuki, H. et al.: ShamFinder: An Automated Framework for Detecting IDN Homographs, *Proc. ACM IMC*, p. 449–462 (2019).

[5] Suignard, M.: UNICODE SECURITY MECHANISMS, (online), available from ⟨`http://unicode.org/reports/tr39/`⟩ (2019).

[6] Zonefiles.io: Lists of all domains updated daily, `https://zonefiles.io/` (2021).

[7] VirusTotal: Inspects items with over 70 antivirus scanners, `https://www.virustotal.com/`.

[8] Grafana: Dashboard anything. Observe everything, `https://grafana.com/`.

[9] 早稲田大学森達哉研究室：東京五輪公式サイトに類似したドメイン名の調査分析，`https://nsl.cs.waseda.ac.jp/tokyo2020/`.

[10] 東京 2020 組織委員会：東京 2020 公式ウェブサイト URL 変更のお知らせ，`https://olympics.com/tokyo-2020/ja/news/website-url-change`.

[11] Lee, Y.: Generating domain names relevant to current events, US Patent 20100146119A1 (2008).

[12] Coull, S. E. et al.: Understanding Domain Registration Abuses, *Proc. IFIP SEC*, pp. 68–79 (2010).

[13] Nakano, H. et al.: Detecting Event-synced Navigation Attacks across User-generated Content Platforms, *Proc. IEEE COMPSAC* (2021).

[14] Kawaoka, R. et al.: A First Look at COVID-19 Domain Names: Origin and Implications, *Passive and Active Measurement*, Springer International Publishing, pp. 39–53 (2021).

[15] Hao, S. et al.: Monitoring the Initial DNS Behavior of Malicious Domains, *Proc. ACM IMC*, p. 269–278 (2011).

[16] Korczynski, M. et al.: Cybercrime After the Sunrise: A Statistical Analysis of DNS Abuse in New gTLDs, *Proc. ACM ASIACCS*, p. 609–623 (2018).

[17] Tian, K. et al.: Needle in a Haystack: Tracking Down Elite Phishing Domains in the Wild, *Proc. ACM IMC*, p. 429–442 (2018).