Journal of *Cybersecurity and Privacy*

MDPI

Article

An In-Depth Measurement of Security and Privacy Risks in the Free Live Sports Streaming Ecosystem

Nithiya Muruganandham [1], Yogesh Sharma [2] and Sina Keshvadi [1,*]

[1] Department of Engineering, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada
[2] Faculty of Engineering and Applied Science, University of Regina, Regina, SK S4S 0A2, Canada; yogesh.sharma@uregina.ca
* Correspondence: skeshvadi@tru.ca

Abstract

Free live sports streaming (FLS) services attract millions of users who, driven by the excitement of live events, often engage with these high-risk platforms. Although these platforms are widely perceived as risky, the specific threats they pose have lacked large-scale empirical analysis. This paper addresses this gap through a comprehensive study of the FLS ecosystem, conducted during two major international sporting events (UCL playoffs and NHL Stanley Cup Playoffs, 2024–2025 season). We analyze the infrastructure, security threats, and privacy violations that define this space. Analysis of 260 unique domains uncovers systemic security risks, including drive-by downloads delivering persistent malware, and widespread privacy violations, such as invasive device fingerprinting that disregards regulations like the General Data Protection Regulation (GDPR). Furthermore, we map the ecosystem's resilient infrastructure, identifying eight clusters of co-owned domains. These findings imply that effective countermeasures must target the centralized infrastructure and ephemeral nature of the FLS ecosystem beyond traditional blocking.

Keywords: free live streaming; online piracy; malware analysis; web privacy; device fingerprinting; malvertising

1. Introduction

The global shift towards online streaming has fundamentally transformed media consumption, particularly for live sports, creating a market that has long passed traditional broadcast television in both reach and revenue [1]. Concurrent with this growth, a vast and illicit shadow economy of Free Live Sports (FLS) services has emerged and solidified its position within the digital landscape [2]. These FLS platforms form a significant part of a digital piracy ecosystem that attracted over 216.3 billion global visits in 2024 alone [3]. They operate as sprawling directories that connect users to a dynamic and often transient network of unauthorized streams, allowing them to circumvent costly subscriptions and geographic broadcasting restrictions.

While the economic impact of piracy on content creators remains a subject of debate [4], the direct harms to the millions of consumers who engage with these sites, such as pervasive malware and privacy invasions, have received far less large-scale scrutiny. This gap stems partly from the ephemeral nature of FLS infrastructure: sites often exist only during live events and are not linked from main pages, evading detection by automated web crawlers.

Users of these services face a trade-off, exchanging free access for exposure to security risks, privacy violations, and poor performance [2]. Drawing on prior behavioral security work, we hypothesize that users exhibit a form of contextual risk displacement [5],

Check for updates

Academic Editor: Sokratis Katsikas

Received: 30 October 2025
Revised: 23 December 2025
Accepted: 29 December 2025
Published: 1 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

where users, driven by the urgency and excitement of a live game, willingly lower their typical security posture during time-critical sporting events. They may perceive in-browser streaming as a relatively low-risk activity, placing undue trust in the browser's sandbox or their ad-blockers to protect them, an assumption prior work has shown to be a significant factor in the rise of illegal streaming [6]. This behavior is often rationalized by the high price of legitimate content or a sense of social normalization ("everyone does it") [7]. This tendency to prioritize convenience over security [8] creates a uniquely permissive environment, allowing malicious actors to operate with impunity against a user base that is not only vast but also less likely to employ fundamental security precautions like antivirus software [9,10].

While studies have investigated the risks associated with general digital piracy [7,11] and mapped the FLS ecosystem to document the presence of malware [12] and performance issues [2], the unique challenges inherent to the FLS ecosystem remain largely underexplored in a longitudinal study. The real-time nature of live events—especially major ones like the FIFA World Cup, UEFA Champions League playoffs, or NBA/NHL playoffs—creates a distinct operational environment where threat actors can more effectively exploit user urgency compared to platforms hosting static content. This paper addresses this critical gap through a large-scale measurement study of the FLS ecosystem. We collected and analyzed data from 260 unique domains across the 2025 UEFA Champions League playoffs and NHL Stanley Cup Playoffs. The upcoming FIFA World Cup 2026 being hosted across North America, with Canada staging 13 matches, provides us additional motivation to carry out this study, as millions are likely to seek free live streams of these high-profile events. Our main contributions are as follows:

1. Large-Scale Threat Characterization: We perform a broad analysis of security threats across hundreds of FLS sites, revealing a diverse and adaptive landscape of attacks ranging from phishing schemes to sophisticated malware delivery.
2. In-Depth Behavioral Malware Analysis: We document the specific post-infection behavior of malware delivered via FLS platforms, detailing their persistence mechanisms and data exfiltration techniques.
3. A Multi-Vantage Point Privacy Assessment: We conduct a multi-vantage point analysis of the FLS tracking ecosystem from four distinct geographic regions, demonstrating that invasive device fingerprinting is applied uniformly with a disregard for regional privacy laws like the GDPR.
4. Co-ownership and Infrastructure Mapping: We use publisher-specific IDs to uncover eight clusters of co-owned FLS domains, revealing their consolidated and resilient illicit infrastructure.

The remainder of this paper is structured as follows: Section 2 provides background on the FLS ecosystem and reviews related work; Section 3 details our methodology; Section 4 presents our findings on security threats and privacy implications; Section 5 discusses our findings, limitations, and future work; and Section 6 concludes the paper.

2. Background and Related Work

The FLS ecosystem operates in a complex legal gray area, as the act of distributing copyrighted content is illegal, while the legality of simply viewing such streams varies significantly by jurisdiction [13]. It presents unique security and privacy challenges that remain largely underexplored. While significant research has addressed general digital piracy and malvertising, the real-time, event-driven nature of FLS platforms creates a distinct environment where threat actors can effectively exploit user urgency.

As illustrated in Figure 1, the FLS ecosystem is fundamentally structured around five interconnected elements: *Users*, who seek free access; *Content Providers*, who source and redistribute streams; *Content Hosting Platforms*, which serve as repositories for these illicit streams; *Aggregators*, which act as central hubs compiling links; and *Ad Servers*, which form the financial backbone of this ecosystem [12,14]. The entire operation is predominantly financed by a vast and lucrative advertising network. Studies estimate that ad-supported piracy generates over $2.5 billion in annual revenue [3,4], a figure consistent with earlier reports showing top piracy sites earned hundreds of millions from advertising alone [8]. This robust economic model has drawn scrutiny from policy advocates, who call for greater accountability from the ad-tech companies and major brands whose advertising budgets fuel this illicit market [15].

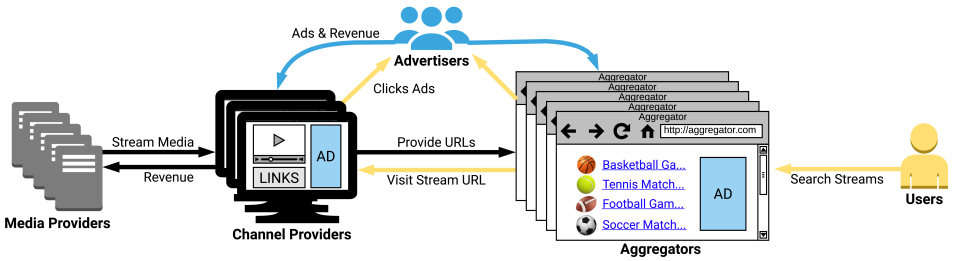


Figure 1. The ecosystem of illegal live sports streaming services. Adapted from Hsiao et al. [12].

A primary threat within the FLS ecosystem is "malvertising", where legitimate ad networks are co-opted to deliver malware and scams [13,16]. This phenomenon is facilitated by systemic vulnerabilities in the programmatic advertising infrastructure, particularly within Real-Time Bidding (RTB) and header bidding protocols. In the RTB ecosystem, where advertisers bid on impressions in milliseconds, the sheer complexity of redirections among diverse stakeholders allows attackers to inject malicious content into legitimate networks undetected [17]. Siby et al. [17] demonstrated that the intricate graph of trackers makes it computationally difficult to trace these malicious flows. Furthermore, the industry shift towards *header bidding*—where publishers query multiple exchanges simultaneously—has expanded the pool of third-party actors executing code on user devices, significantly increasing the attack surface for such injections [18]. Research on digital piracy platforms confirms this danger, finding that a majority of ad interactions redirect to malicious websites [12]. Our work builds directly on these findings by performing sandboxed dynamic analysis of the final malware payloads.

Beyond direct security threats, FLS platforms engage in aggressive user tracking. While foundational work focused on third-party cookies [6], recent studies have examined the proliferation of deceptive consent interfaces. For instance, Bouhoula et al. [19] revealed that over 65% of legitimate sites still collect data after users reject cookies. Similarly, Sanchez-Rola et al. [20] mapped evasive fingerprinting techniques on phishing pages. We adapt these insights to the FLS ecosystem, revealing that FLS operators bypass these nuances entirely by failing to implement any consent mechanisms whatsoever.

While digital piracy is a well-studied field, our work distinguishes itself from recent literature in three key areas. First, regarding infrastructure, whereas Sheaib et al. [7] mapped technical video origins in movie streaming, our analysis of Google Publisher IDs uncovers financial consolidation, revealing how operators use sport-specific brand diversification to build resilience. Second, concerning threat delivery, we move beyond the server-side evasion techniques analyzed by Lee et al. [9] to map client-side vectors, documenting how professional, high-traffic aggregators co-opt legitimate ad networks to deliver specialized malware, a contrast to the transient, zero-cost phishing sites analyzed by Saha et al. [21]. Finally, we complement prior Quality of Service (QoS) measurements by Keshvadi et al. [2]

by shifting focus to the deep-seated security and privacy threats that coexist with these streaming services.

3. Methodology

3.1. FLS Site Identification and Data Collection

Identifying a comprehensive set of FLS domains is challenging due to their transient nature. To address this, we developed a structured, three-phase collection pipeline (Figure 2) that implements the logic detailed in Algorithm 1.

Algorithm 1 Automated FLS Link Discovery and Filtering (Implementation: `scripts/1_collect_links.py` and `scraper.py` [22])

Require: Event Configurations $\mathcal{C}_{events}$, Blocklist $\mathcal{B}$
Ensure: Verified FLS Domains $\mathcal{D}_{final}$
$\mathcal{D}_{db} \leftarrow \emptyset$ {Persistent Database}
{**Phase 1: Discovery**}
$\mathcal{S}_{agg} \leftarrow$ LoadSeedsFromSocialAndSearch()
{**Phase 2: Automated Crawling & Filtering**}
while Event is Live **do**
  {Continuous monitoring (15-min intervals)}
  for aggregator $A \in \mathcal{S}_{agg}$ **do**
    LaunchHeadlessBrowser(A)
    $DOM \leftarrow$ RenderPage(A)
    $Links \leftarrow$ ExtractHREF(DOM)
    $A_{dom} \leftarrow$ GetRegisteredDomain(A)
    for $url \in Links$ **do**
      $u_{dom} \leftarrow$ GetRegisteredDomain(url)
      {Filter internal links and known ad networks}
      if IsValid(url) **AND** $u_{dom} \neq A_{dom}$ **AND** $u_{dom} \notin \mathcal{B}$ **then**
        if $url \notin \mathcal{D}_{db}$ **then**
          $\mathcal{D}_{db} \leftarrow \mathcal{D}_{db} \cup \{url\}$
        end if
      end if
    end for
  end for
  Sleep(15 min)
end while
{**Phase 3: Verification**}
$\mathcal{D}_{final} \leftarrow \emptyset$
for $d \in \mathcal{D}_{db}$ **do**
  if ManualVerifyStream(d) **then**
    $\mathcal{D}_{final} \leftarrow \mathcal{D}_{final} \cup \{d\}$
  end if
end for
return $\mathcal{D}_{final}$

3.1.1. Phase 1: Discovery

We began by identifying a seed set of aggregator websites, which serve as directories to FLS domains. We focused on two major sporting events: the UEFA Champions League (UCL) playoffs and the NHL Stanley Cup Playoffs (2024–2025 season). We compiled an initial list of seed aggregators by triangulating across three primary sources. First, we monitored event-specific subreddits (e.g., `r/soccerstreams`, `r/puckstreams`) and related Discord servers where users actively share links in real-time. Second, adopting the approach from Saha et al. [21], we queried the X (Twitter) API for hashtags such as `#soccerstreams` and `#NHLstreams` during live games. Finally, to capture publicly indexed sites, we executed 20 distinct Google search queries (e.g., "watch soccer free online") before and during the

first matches. This discovery phase yielded 42 unique aggregator domains for the UCL event and 35 for the NHL event.

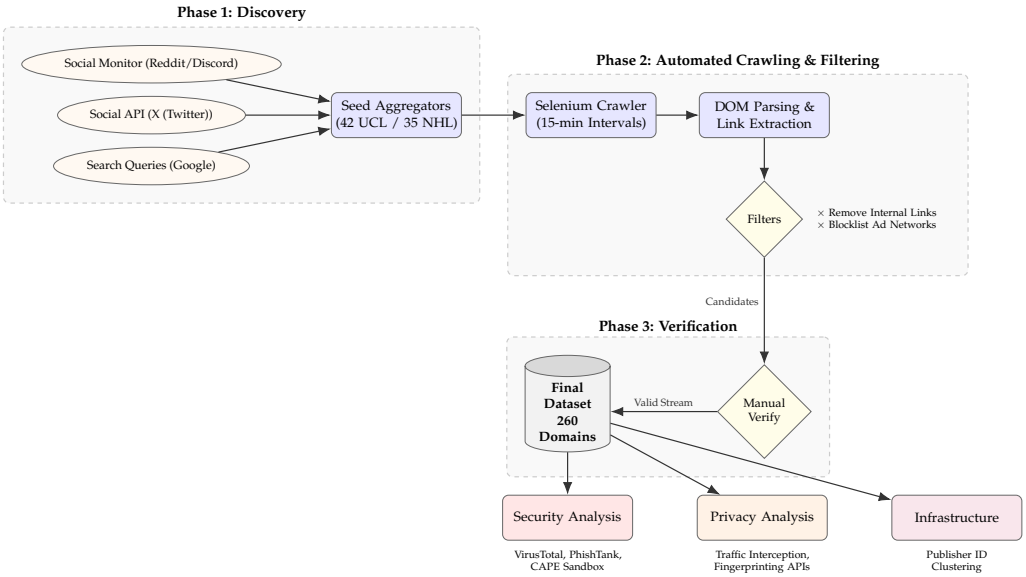


Figure 2. End-to-end data collection and analysis pipeline. The process comprises three phases: (1) Discovery of seed aggregators via social monitoring and search queries; (2) Automated crawling and filtering of streaming links using Selenium; (3) Manual verification yielding final dataset, used by the security, privacy, and infrastructure analysis modules.

3.1.2. Phase 2: Automated Crawling & Filtering

Since FLS links are often ephemeral, we developed an automated crawling engine (Algorithm 1) to continuously monitor the seed aggregators. The script ran in 15-min intervals throughout all knockout-stage games, using Selenium to control a headless Chrome browser. This allowed us to parse the fully rendered Document Object Model (DOM) and extract dynamically loaded links [12]. To isolate external video sources, we applied a two-step filtering process. First, we removed internal navigation links by excluding URLs where the registered domain matched the aggregator. Second, we filtered out URLs belonging to a manually curated blocklist of over 50 known advertising networks (e.g., `doubleclick.net`) and social media platforms.

3.1.3. Phase 3: Verification

The automated phase produced a set of candidate domains which required final verification. All remaining candidate URLs were manually visited by a researcher to verify that they were active FLS sites providing a video stream. To ensure the reliability of this classification process, a second author independently verified a random sample of 10% of the candidate domains. This verification step yielded a 100% agreement rate. This process resulted in a final verified dataset of 260 unique FLS domains (190 UCL, 70 NHL).

Our data collection continued until we reached a point of saturation, where new iterations of our crawling pipeline yielded diminishing returns in identifying unique domains. For example, during the final rounds of the UCL playoffs, almost all links extracted from aggregators pointed to domains we had already cataloged. While this dataset of 260 domains is not an exhaustive census of the entire web (excluding private or invite-only IPTV networks), we argue that it accurately represents the effective public-facing FLS ecosystem accessible to the average user.

3.2. Measurement Environment and Ethics

All data collection and analysis scripts were written in Python v3.12 and run on a machine with Ubuntu 22.04. To analyze web traffic, we used a rooted Google Pixel 5 smartphone, which allowed us to install a system-level certificate for Charles Proxy (v4.6.4) [23] for HTTPS decryption. Network traffic was captured using Wireshark, and redirection URLs for phishing analysis were captured with Burp Suite. Automated website interactions and DOM analysis were performed using Selenium with the Chrome WebDriver. To address potential geo-specific behaviors, we conducted our analysis from four distinct vantage points (VPs): our university lab in Canada, and AWS EC2 instances in US East (N. Virginia), Frankfurt (Germany), and Singapore. These locations were chosen to represent diverse regulatory environments, most notably the GDPR in Europe [24].

Our study involved analyzing publicly accessible web content and did not involve interaction with human subjects. The "victims" in our threat analysis were our own controlled devices and accounts. As such, this study was exempt from Institutional Review Board (IRB) review. We adhered to the ethical standards and guidelines established in similar measurement studies [2,7,12,14]. Specifically, we did not exploit vulnerabilities in the FLS sites themselves, nor did we interact with other users or distribute any collected malware.

3.3. Security Threat Analysis

We submitted the full URL of every collected FLS domain to the VirusTotal API [25]. A URL was classified as hosting malicious JavaScript only if more than 5 security vendors flagged it as malicious. This conservative threshold ($N > 5$) was chosen based on best practices established in prior measurements [14] to filter out false positives often generated by heuristic engines on 'grayware' sites. To detect phishing, all redirection URLs captured by our crawler were logged and cross-referenced with the PhishTank blacklists [26]. To assess the persistence and detection lag of these threats, we performed a final longitudinal query of all unique URLs against the PhishTank database in early July 2025 (approximately one month after the conclusion of the NHL and UCL finals).

To investigate high-severity payload behaviors, we captured all executable files (e.g., .exe, .dmg, .apk) triggered by drive-by downloads. We executed each binary in a high-interaction CAPE Sandbox [27] environment running a Windows 10 guest OS. We generated detailed execution reports documenting file system modifications, registry changes, and network callbacks (including DNS requests and contacted C2 server IPs).

3.4. Privacy Violation Analysis

To identify requests belonging to advertising or tracking services, we matched every third-party domain against a combined blocklist derived from EasyList and Disconnect [28,29]. To differentiate between tracking methods, we performed static analysis on the DOM source code captured by our crawler. We utilized BeautifulSoup to extract content from all `<script>` tags and applied keyword matching to detect the presence of specific API calls associated with browser fingerprinting. We detected *Canvas Fingerprinting* by identifying scripts that invoke `canvas.toDataURL` or `getImageData`. *WebRTC Fingerprinting* was identified by the presence of keywords accessing the `localDescription` of an `RTCPeerConnection` object. We also monitored for other high-entropy keywords often used in fingerprinting scripts, including `getClientRects` and `audioContext`. This static analysis approach identifies the presence of fingerprinting capabilities within the site's code, providing a lower-bound estimate of privacy risks. Finally, to uncover hidden infrastructure relationships, we extracted Google Analytics and AdSense publisher IDs from all HTTP responses and page sources [7]. Sites sharing the same ID were grouped into clusters, indicating common ownership.

4. Analysis and Findings

The FLS ecosystem represents a significant portion of internet traffic. During UCL 2025, we observed that over 17.5% of the aggregators had more than 10 million visits between April and June 2025, with approximately 60% averaging 3.83 million site visits. This massive user base is exposed to a diverse range of security and privacy threats. Table 1 provides a summary of these threats, detailing their prevalence and observed vectors. We elaborate on these findings in the following subsections.

Table 1. Prevalence of security and privacy threats detected across the ecosystem. Prevalence percentages refer to the proportion of the 260 FLS domains containing the specific threat code. Confidence Intervals (C.I.) are calculated at the 95% confidence level.

Threat Category	Prevalence	95% C.I.	Observed Behavior/Vector
Security Threats			
Malicious JavaScript	31.5%	±5.6%	Obfuscated scripts injecting overlay ads and redirecting to external malware sources.
Phishing Redirects	*51 Chains*	–	Redirection chains leading to confirmed phishing pages (via PhishTank).
Ad-Blocker Evasion	12.0%	±4.0%	Scripts containing logic to detect ad-blockers and block video playback.
Drive-by Downloads	*Observed*	–	Automatic download of payloads (e.g., `HD-StreamPlayer.exe`) triggered by ad interactions.
Privacy Violations (Static Detection)			
Canvas Fingerprinting	12.0%	±4.0%	Scripts invoking high-entropy extraction APIs (`canvas.toDataURL` or `getImageData`).
Font Fingerprinting	8.0%	±3.3%	Scripts containing keywords for font enumeration (e.g., `measureText`).
WebRTC Leakage	5.0%	±2.7%	Presence of `RTCPeerConnection` logic capable of leaking local IP addresses.

4.1. Security Analysis

Our analysis reveals a diverse and adaptive threat landscape within the FLS ecosystem, ranging from direct malware delivery to sophisticated social engineering. None of the threats detailed in this section were observed on the legitimate broadcasters (TSN and Sportsnet).

4.1.1. Drive-By Downloads and Malware Behavior

The most severe threat observed was the delivery of malware via drive-by downloads. Over the course of the data collection period, we successfully captured and analyzed 5 unique malicious payloads, ranging from infostealers to obfuscated droppers. We identified two primary delivery vectors: user-initiated downloads triggered by deceptive UI elements like "play" buttons, and covert downloads that occurred automatically without any user interaction. In a notable instance of the latter, we observed a multi-stage attack triggered by a malicious script embedded in a third-party advertising iframe. Upon loading, the script executed a `window.location.replace()` command pointing to the payload URL, effectively bypassing user interaction prompts. This forced the browser to drop two payload files (e.g., `1420e1ce...tmp`) directly into the system's `C:/Users/administrator/AppData/Local/Temp` directory. Detailed analysis of the captured payload (MD5: `0a3970cad0e2e7aad82103d56cae7592`) revealed the use of critical Windows API calls, specifically `LdrLoadDll` and `NtCreateFile`. The malware utilized these calls to perform DLL injection, loading malicious modules into `advapi32.dll`—a technique frequently associated with persistent adware and infostealers [30]. The payloads from these downloads had varied objectives, from adware and scareware to the infostealer detailed in the following case study.

Case Study: HD-StreamPlayer.exe

To illustrate the concrete harm from these downloads, we analyze HD-StreamPlayer.exe, a 1.2 MB executable recovered during the NHL Finals. Upon execution, the program created a mutex object to prevent multiple instances from running, copying itself to the user's AppData directory (%AppData%\Roaming\RenderSvc\stream.exe), and modifying the Windows Registry (`HKCU\...\Run\RenderSvc`). Subsequently, the malware contacted a server at 185.17.1.88 and transmitted a detailed profile of the victim's machine, including their public IP address, Windows OS version, and system user names. Table 2 maps the observed adversary behaviors, ranging from the initial drive-by compromise to data exfiltration, against the MITRE ATT&CK framework [31].

Table 2. Classification of observed FLS threats mapped to the MITRE ATT&CK Framework. Behaviors of captured binaries and malicious scripts were analyzed against standard adversary tactics.

Tactic	Technique ID	Technique Name	Observed Evidence in FLS Ecosystem
Initial Access	T1189	Drive-by Compromise	Malicious JavaScript in ad iframes triggering `window.location.replace` to force downloads without user interaction.
	T1566.002	Phishing: Spearphishing Link	Deceptive "Play" buttons and overlay ads tricking users into manually downloading malware.
Execution	T1204.002	User Execution: Malicious File	Users manually launching `HD-StreamPlayer.exe` believing it is a required codec update.
	T1059.007	Command and Scripting Interpreter	Obfuscated JavaScript (e.g., `tag.min.js`) executing in the browser to inject ads/redirects.
Persistence	T1547.001	Registry Run Keys	Malware adds entries to `HKCU\...\Run` to ensure execution on startup.
Privilege Escalation	T1055.001	Process Injection: DLL Injection	Dropped `.tmp` payloads utilizing `LdrLoadDll` to inject malicious code into `advapi32.dll`.
Defense Evasion	T1027	Obfuscated Files or Information	Heavily obfuscated JavaScript and packed binaries used to evade signature-based AV detection.
	T1562.001	Impair Defenses: Disable Tools	Scripts detecting ad-blockers and blocking content playback until protection is disabled.
Discovery	T1082	System Information Discovery	Malware profiling the victim's OS version, Hostname, and User Name immediately upon execution.
	T1016	System Network Config Discovery	Scripts (WebRTC) and binaries enumerating local IP addresses and Network Context.
Exfiltration	T1041	Exfiltration Over C2 Channel	Continuous transmission of fingerprinted data to tracking APIs via HTTP GET parameters.

Other Malware Payloads

Beyond data exfiltration, other manually executed files revealed different monetization schemes. Some instances involved adware that forcibly redirected the user's browser to gambling and adult websites. In other cases, we found scareware that displayed deceptive warnings of a virus infection to pressure users into installing fake McAfee antivirus products.

4.1.2. Malicious and Evasive JavaScript

Malicious JavaScript was the most prevalent threat, detected on 32.6% of UCL sites and 28.6% of NHL sites. A Chi-square test of independence performed on these proportions yielded no statistically significant difference ($\chi^2(1, N = 260) = 0.38, p > 0.5$). This statistical uniformity indicates that threat actors deploy consistent attack infrastructure regardless of the specific sporting event or target demographic. These scripts often employed obfuscation to hide their intent. Their primary goals were to generate fraudulent ad impressions, redirect users, and track interactions with pop-ups using cookies (document.cookie). A manual examination of `tag.min.js`, a script found on several sites hosted at `alwingulla.com`, revealed functions specifically designed to dynamically insert ads into the DOM and monitor user interactions. This script also initiated network requests to `my.rtmark.net`, a domain associated with the distribution of unwanted browser extensions and other malicious programs. In addition, we observed evasive techniques like homepage cloaking, where a benign homepage masks the illicit streams served on deep, unlinked paths to evade casual inspection.

4.1.3. Malvertising, Phishing, and Scams

The FLS ecosystem's primary revenue model is aggressive advertising, which exposes users to a chaotic and high-risk environment. Our manual inspection revealed that virtually any interaction with a page—clicking the video player, adjusting volume, or even clicking empty space—triggered a barrage of pop-up ads. These redirects led to a wide range of high-risk destinations, including gambling sites, adult content, and cryptocurrency scams.

While the FLS domains themselves did not host phishing content, this threat was delivered via malicious redirects. To investigate this, we analyzed all redirect URLs captured by Burp Suite during our manual interactions. We regularly checked the captured URLs against the PhishTank database [26] throughout our data collection period. This longitudinal checking revealed a significant detection lag. While 37 URLs were initially flagged as phishing, this number grew to 51 by the end of our study, which indicates that many phishing sites were live and victimizing users before they were added to popular blacklists.

Finally, we observed several instances of fake antivirus scams designed to deceive users into installing malware. One notable example involved the popular aggregator `vipleague.im`, which redirected users to a page impersonating a McAfee online scanner. This fraudulent page displayed a fabricated scan report with fictitious virus warnings to alarm the user. Curiously, while similar scams often lead to a fake checkout page, this particular instance ultimately redirected to the legitimate McAfee checkout page, perhaps in an attempt to gain affiliate revenue.

4.2. Privacy Analysis

This section analyzes the privacy-violating behaviors of FLS platforms, comparing them to legitimate services where relevant.

4.2.1. Unauthorized Data Collection and Tracking

In contrast to prior work that found regional differences in tracking [7], our multi-vantage point analysis indicates a uniform deployment of tracking infrastructure. We compared the rendered DOM and third-party script references captured from our remote vantage points (USA, Germany, Singapore) against our primary dataset. We found that FLS operators served identical HTML structures and third-party tracking tags to all locations (100% code similarity for the analyzed components). This suggests that these platforms do not implement geo-fencing for compliance.

One popular site, `buffstreams.is`, employed a sophisticated data exfiltration technique designed to evade analysis. Instead of firing multiple distinct tracking pixels, the site utilized a *request tunneling* approach. It constructed a single, massive HTTP request (observed payload size: 16,029 bytes) to a data collection endpoint. Decodification of this payload revealed that it encapsulated 21 distinct sub-requests (comprising 18 GET and 3 POST operations) targeting a syndicate of advertising and analytics services. By bundling these requests into a single serialized payload, the operator effectively masks the destination of the individual trackers from simple DNS-based blockers. The structure of this tunneled request, reconstructed from our captured traffic logs, is schematically represented in Figure 3. This aggregation technique allows FLS operators to tunnel analytics data regarding ad impressions and user behavior through a single endpoint, effectively bypassing ad-blockers that filter based on specific third-party domains.

```
// 1. The Outer Shell (Massive Tunneled Request)
POST /jserrors/1/?data=%7B%22metrics%22%3A%5B... HTTP/1.1
Host: bam.nr-data.net
Referer: https://buffstreams.is/live-stream/

// 2. The Decoded Payload (Representative Schema)
{
  "context": {
    "client": {
      "userAgent": "Mozilla/5.0 ... Android 7.0; Nexus 6...",
      "timeZone": "America/Vancouver",
      "connectionType": "CONN_CELLULAR_4G",
      "gl": "CA"
    }
  },
  "events": [
    {
      // Event 1: Amazon Ad System Bid
      "method": "GET",
      "host": "aax.amazon-adsystem.com",
      "pathname": "/e/dtb/bid",
      "status": 200
    },
    {
      // Event 2: Criteo User Sync
      "method": "POST",
      "host": "bidder.criteo.com",
      "pathname": "/cdb",
      "metrics": {"txSize": 601, "duration": 518}
    },
    {
      // Event 3: AppNexus (Xandr) Pixel
      "method": "GET",
      "host": "secure.adnxs.com",
      "pathname": "/jpt",
      "status": 200
    },
    {
      // Event 4: Verizon AdTech
      "method": "GET",
      "host": "adserver-us.adtech.advertising.com",
      "pathname": "/pubapi/3.0/...",
      "status": 200
    }
  ]
}
```

Figure 3. Schematic of the observed request tunneling technique. A single request encapsulates bidding and tracking calls to major ad-tech providers (Amazon, Criteo, AppNexus), hiding them from user-side blocking tools.

4.2.2. Cookie Proliferation and Lack of Consent

FLS sites make extensive use of tracking cookies without user consent. Our analysis of the 260 FLS domains identified over 1500 unique third-party trackers, with some individual sites setting as many as 70 trackers. Concerningly, not a single FLS site in our dataset provided a clear cookie consent banner or a readily accessible privacy policy. The legitimate broadcasters, TSN and Sportsnet, which, despite also lacking initial consent banners, set

a comparable number of cookies (68 and 62, respectively) that were not associated with malicious ad networks.

4.2.3. Centralized Infrastructure and Brand Diversification

By extracting Google Publisher IDs, we identified 8 distinct clusters of co-owned domains. While prior work on movie streaming sites identified clustering primarily as a mechanism for redundancy (mirroring identical content) [7], our analysis of the FLS ecosystem reveals a distinct brand diversification strategy. For instance, we uncovered a single cluster of 12 domains (sharing one unique AdSense ID) that all targeted North American sports fans, yet each utilized distinct branding and URLs. This indicates that operators manage consolidated portfolios of domains to target specific user communities and, crucially, to ensure service continuity as if one domain is targeted by a takedown notice during a game, the underlying infrastructure remains intact, and users can be shifted to co-owned backup domains.

4.2.4. Web Security Posture and Safeguard Effectiveness

In contrast to earlier studies [2], we found the vast majority of FLS sites (98%) now use HTTPS. This is likely a response to modern browsers actively flagging HTTP sites as "Not Secure". However, manual testing using Mozilla Observatory revealed that 92% of FLS sites failed to implement a Content Security Policy (CSP), and 85% lacked HTTP Strict Transport Security (HSTS), with most receiving a failing "F" score.

Regarding user-side safeguards, ad blockers like uBlock Origin proved effective at preventing most ads, but 12% of sites, often the most popular, actively detected and blocked access if an ad blocker was enabled. Our manual analysis of these blocking events indicates that the sites rely on *DOM-based Content Occlusion*. Unlike simple warning banners, these sites employ scripts that detect the suppression of ad elements and respond by dynamically injecting a full-screen modal overlay (typically with a high `z-index`). This overlay physically obstructs the video player and intercepts mouse events, effectively rendering the site unusable until the ad-blocking software is disabled. To contextualize the risks documented in the FLS ecosystem, we performed a structured comparison against two legitimate Canadian broadcasters, TSN and Sportsnet. We selected these providers as they hold the exclusive broadcasting rights for the UCL and NHL in our primary vantage point (Canada), effectively representing the entire population of legitimate alternatives available to local users. Table 3 summarizes the key differences in security posture and tracking behaviors.

While one legitimate broadcaster (TSN) also received poor scores (graded "F" during the study period), the other (Sportsnet) received a "B". TSN's low grade resulted from a combination of missing `HSTS`, missing `X-Frame-Options`, and a permissive `Content Security Policy` (CSP) that explicitly allows `'unsafe-inline'` scripts. However, while TSN's policy is permissive, 92% of FLS domains failed to implement any CSP. This creates an unrestricted browser environment on FLS platforms that allows arbitrary third-party scripts to execute without validation. Quantitatively, legitimate broadcasters employ a comparable volume of third-party tracking cookies to the most aggressive FLS sites. However, the nature of this tracking differs fundamentally: legitimate sites rely on standard analytics (e.g., Adobe Analytics) rather than the malware-distributing networks (e.g., `my.rtmark.net`) found on FLS platforms. While our sample size of legitimate broadcasters ($N = 2$) precludes significance testing for binary threat vectors, the difference in *tracking volume* allows for statistical comparison against the FLS distribution. We calculated the z-score of the legitimate broadcasters' cookie volume ($M = 65$) against the FLS dataset baseline ($\mu \approx 6.2$, $\sigma \approx 12.4$). The legitimate sites returned a z-score of 4.7, placing them

in the 99th percentile of tracking intensity. Legitimate broadcasters are distinct outliers in terms of surveillance volume. They engage in significantly higher quantities of user tracking than the average piracy site, albeit using regulated commercial networks rather than malicious vectors.

Table 3. Security and Privacy Comparison: FLS Ecosystem vs. Legitimate Broadcasters.

Metric	FLS Ecosystem (n = 260)	TSN (Official)	Sportsnet (Official)
Mozilla Observatory Score	F (Most Sites)	F	B
HTTPS Implementation	98%	Yes	Yes
Content Security Policy	8% (Rare)	Missing	Present
Cookie Volume (Max)	≈70	68	62
Malicious Trackers	Detected	None	None
Malware Vectors	Present	Absent	Absent
Ad-Block Detection	12% (Active Block)	None	None

5. Discussion

A central theme emerging from our research is the contradiction between the objective dangers of the FLS ecosystem and the subjective sense of security among its users. On one hand, our technical analysis paints a picture of a uniformly hostile environment. The security risks are not incidental but systemic, ranging from covert technical attacks, such as drive-by downloads delivering infostealer malware and obfuscated JavaScript (present on nearly a third of analyzed domains), to overt social engineering schemes like phishing redirects and fake antivirus scams. Simultaneously, the ecosystem operates with a complete disregard for user privacy and regional laws like the GDPR, demonstrated by uniformly aggressive device fingerprinting and the continuous exfiltration of sensitive user data across all geographic vantage points.

Limitations and Future Work

Our study has several limitations that open avenues for future research. First, our reliance on manual discovery via social media (Reddit, Discord, X) introduces a selection bias towards popular, publicly accessible aggregators. While this methodology was chosen to accurately model the discovery process of a typical user, it inherently excludes niche, invite-only, or private streaming communities (e.g., private IPTV networks) that may operate with different infrastructure or risk profiles. Therefore, our findings should be interpreted as characterizing the public-facing segment of the FLS ecosystem rather than its entirety. Second, our data collection was focused on two specific, albeit major, sporting events (soccer and hockey). Third, the dynamic malware analysis was constrained by the number of samples recovered from the wild. The delivery of executable payloads (drive-by downloads) proved to be a rare event compared to other threats. Consequently, our deep-dive analysis focuses on the 5 unique payloads captured during the observation window. While these samples provide detailed insights into specific attack chains, they represent a specific subset of the broader threat landscape.

Future work could expand this to other popular sports (e.g., basketball, cricket, tennis) to identify sport-specific threats or infrastructure. This study focuses exclusively on web-based FLS, a parallel ecosystem of dedicated streaming hardware and applications (e.g., Kodi) presents a different set of security and privacy risks worth exploring in future work. The malware analysis, while in-depth, was based on a limited number of captured samples. A longitudinal study tracking the evolution of malware payloads over time would be highly valuable. Finally, since the entire FLS economy is predicated on advertising, a deep investigation into the ad-tech supply chain, mapping the specific advertisers, networks,

and financial flows, is a critical next step to understanding and disrupting the monetization strategies that fuel these illicit services.

6. Conclusions

This paper presented a large-scale measurement study of the public-facing Free Live Sports Streaming ecosystem. Our examination of 260 domains reveals a mature and resilient illicit economy that poses significant risks to users. Conducted during two major international sporting events, this study demonstrates a diverse and adaptive threat landscape, with systemic risks ranging from drive-by downloads delivering infostealer malware to widespread phishing and social engineering scams. Our multi-vantage point analysis revealed that FLS operators uniformly deploy aggressive device fingerprinting and data exfiltration techniques, ignoring regional regulations like the GDPR. By analyzing publisher-specific identifiers, we uncovered evidence of a consolidated infrastructure, with clusters of co-owned domains that enhance the ecosystem's resilience against takedown efforts.

Current mitigation strategies, which primarily rely on domain blocking, have proven ineffective due to the ecosystem's resilience. Based on our data, we propose three targeted interventions:

- Our identification of co-owned clusters (e.g., 12 domains sharing a single AdSense ID) implies that enforcement should shift from blocking ephemeral domains to sanctioning the underlying financial identifiers. Blacklisting specific Publisher IDs across ad-tech exchanges would demonetize entire portfolios of sites simultaneously.
- The prevalence of malvertising and drive-by downloads indicates a failure in the ad supply chain. Policy interventions must enforce stricter "Know Your Customer" (KYC) protocols for ad exchanges, holding them accountable for vetting the creative content delivered through their networks.
- Since users might exhibit contextual risk displacement during live events, static blocklists (like PhishTank) are often too slow (as evidenced by the detection lag we measured). Browser vendors should implement heuristic warnings specifically for live streaming signatures, such as the simultaneous execution of high-entropy canvas fingerprinting and aggressive overlay ads, to warn users in real-time.
- To combat the drive-by downloads we observed, browser vendors should implement heuristics to detect and block invisible overlays (e.g., transparent `<div>` elements intercepting clicks) commonly used to hijack user interactions on video players.

Ultimately, this study highlights a need for increased awareness among users and more robust technical and policy interventions to disrupt the financial incentives that fuel this illicit economy.

Author Contributions: Conceptualization, N.M., S.K. and Y.S.; methodology, N.M. and S.K.; software, Y.S.; validation, N.M.; formal analysis, Y.S.; investigation, N.M.; resources, N.M. and S.K.; data curation, N.M.; writing—original draft preparation, N.M.; writing—review and editing, S.K. and Y.S.; visualization, N.M.; supervision, S.K.; project administration, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) [grant number RGPIN-2023-05343].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The code and datasets supporting this study are available at https://github.com/Keshvadi/fls-paper (accessed on 17 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lee, C.C.; Nagpal, P.; Ruane, S.G.; Lim, H.S. Factors affecting online streaming subscriptions. *Commun. IIMA* **2018**, *16*, 2. [CrossRef]
2. Keshvadi, S.; Williamson, C. An empirical measurement study of free live streaming services. In Proceedings of the International Conference on Passive and Active Network Measurement, Virtual, 29 March–1 April 2021; Springer: Cham, Switzerland, 2021; pp. 111–127.
3. MUSO. *2024 Piracy Trends and Insights*; Technical report; MUSO: Mumbai, India, 2025.
4. Blackburn, D.; Eisenach, J.A.; Harrison, D. *Impacts of Digital Video Piracy on the US Economy*; Nera Economic Consulting, Global Innovation Policy Center: Washington, DC, USA, 2019.
5. Armenia, S.; Ferreira Franco, E.; Nonino, F.; Spagnoli, E.; Medaglia, C.M. Towards the definition of a dynamic and systemic assessment for cybersecurity risks. *Syst. Res. Behav. Sci.* **2019**, *36*, 404–423. [CrossRef]
6. Acar, G.; Eubank, C.; Englehardt, S.; Juarez, M.; Narayanan, A.; Diaz, C. The web never forgets: Persistent tracking mechanisms in the wild. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014; pp. 674–689.
7. Sheaib, H.; Feldmann, A.; Dao, H. Unmasking the Shadows: A Cross-Country Study of Online Tracking in Illegal Movie Streaming Services. In Proceedings of the Proceedings on Privacy Enhancing Technologies Symposium, Washington, DC, USA, 14–19 July 2025; pp. 125–139.
8. Belanger, F.; Hiller, J.S.; Smith, W.J. Trustworthiness in electronic commerce: The role of privacy, security, and site attributes. *J. Strateg. Inf. Syst.* **2002**, *11*, 245–270. [CrossRef]
9. Lee, W.; Hur, J.; Kim, D. Beneath the phishing scripts: A script-level analysis of phishing kits and their impact on real-world phishing websites. In Proceedings of the 19th ACM Asia Conference on Computer and Communications Security, Singapore, 1–5 July 2024; pp. 856–872.
10. Telang, R. *Does Online Piracy Make Computers Insecure? Evidence from Panel Data*; SSRN: Rochester, NY, USA, 2018.
11. Ibosiola, D.; Steer, B.; Garcia-Recuero, A.; Stringhini, G.; Uhlig, S.; Tyson, G. Movie pirates of the caribbean: Exploring illegal streaming cyberlockers. In Proceedings of the International AAAI Conference on Web and Social Media, Palo Alto, CA, USA, 25–28 June 2018; Volume 12.
12. Hsiao, L.; Ayers, H. The price of free illegal live streaming services. *arXiv* **2019**, arXiv:1901.00579. [CrossRef]
13. Digital Citizens Alliance; White Bullet. *Breaking (B)ads: How Advertiser-Supported Piracy Helps Fuel A Booming Multi-Billion Dollar Illegal Market*; Technical report; Digital Citizens Alliance: Washington, DC, USA, 2021.
14. Rafique, M.Z.; Van Goethem, T.; Joosen, W.; Huygens, C.; Nikiforakis, N. It's Free for a Reason: Exploring the Ecosystem of Free Live Streaming Services. In Proceedings of the NDSS, San Diego, CA, USA, 21–24 February 2016.
15. Cooper, S. Online Ads Supporting Copyright Piracy Need to Be Stopped. Free State Foundation: United States of America, 2021. Available online: https://coilink.org/20.500.12592/j4m8xf (accessed on 17 October 2025).
16. Iqbal, U.; Shafiq, Z.; Qian, Z. The ad wars: Retrospective measurement and analysis of anti-adblock filter lists. In Proceedings of the 2017 Internet Measurement Conference, London, UK, 1–3 November 2017; pp. 171–183.
17. Siby, S.; Iqbal, U.; Englehardt, S.; Shafiq, Z.; Troncoso, C. {WebGraph}: Capturing advertising and tracking information flows for robust blocking. In Proceedings of the 31st USENIX Security Symposium (USENIX Security 22), Boston, MA, USA, 10–12 August 2022; pp. 2875–2892.
18. Veale, M.; Borgesius, F.Z. Adtech and real-time bidding under European data protection law. *Ger. Law J.* **2022**, *23*, 226–256. [CrossRef]
19. Bouhoula, A.; Kubicek, K.; Zac, A.; Cotrini, C.; Basin, D. Automated {Large-Scale} analysis of cookie notice compliance. In Proceedings of the 33rd USENIX Security Symposium (USENIX Security 24), Philadelphia, PA, USA, 14–16 August 2024; pp. 1723–1739.
20. Sanchez-Rola, I.; Bilge, L.; Balzarotti, D.; Buescher, A.; Efstathopoulos, P. Rods with laser beams: Understanding browser fingerprinting on phishing pages. In Proceedings of the 32nd USENIX Security Symposium (USENIX Security 23), Anaheim, CA, USA, 9–11 August 2023; pp. 4157–4173.
21. Saha Roy, S.; Karanjit, U.; Nilizadeh, S. Phishing in the free waters: A study of phishing attacks created using free website building services. In Proceedings of the 2023 ACM on Internet Measurement Conference, Montreal, QC, Canada, 24–26 October 2023; pp. 268–281.
22. Muruganandham, N.; Keshvadi, S. Code and Data for: An In-Depth Measurement of Security and Privacy Risks in the Free Live Streaming Ecosystem. 2025. Available online: https://github.com/Keshvadi/fls-paper (accessed on 17 October 2025).
23. XK72 Ltd. Charles Web Debugging Proxy. Available online: https://www.charlesproxy.com (accessed on 9 July 2025).
24. Regulation, P. General data protection regulation. *Intouch* **2018**, *25*, 1–5.
25. Google. VirusTotal. Available online: https://www.virustotal.com (accessed on 9 July 2025).
26. Cisco. PhishTank: Join the Fight Against Phishing. Available online: https://phishtank.org (accessed on 9 July 2025).

27. The CAPE Sandbox Developers. CAPE Sandbox. Available online: https://capesandbox.com (accessed on 9 July 2025).
28. Eyeo GmbH. EasyList and EasyPrivacy Filter Lists. Available online: https://easylist.to (accessed on 9 July 2025).
29. Disconnect, Inc. Disconnect Malvertising and Tracking Protection Lists. Available online: https://disconnect.me (accessed on 9 July 2025).
30. mrd0x. MalAPI.io: Windows API Mapping for Malware Research, 2024. Available online: https://malapi.io (accessed on 9 July 2025).
31. Pennington, A.; Applebaum, A.; Nickels, K.; Schulz, T.; Strom, B.; Wunder, J. *Getting Started with ATT&CK*; MITRE Corporation: McLean, VA, USA, 2019. Available online: https://attack.mitre.org/resources (accessed on 17 October 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.