A First Look at User-Installed Residential Proxies From a Network Operator's Perspective

Etienne Khan*, Elisa Chiapponi†, Martijn Verkleij*
Anna Sperotto*, Roland van Rijswijk-Deij* and Jeroen van der Ham-de Vos*
*Design and Analysis of Communication Systems (DACS)
University of Twente, Enschede, The Netherlands
Email: {e.khan, a.sperotto, r.m.vanrijswijk, j.vanderham}@utwente.nl, research@mverkleij.nl
†Amadeus IT Group, Biot, France
Email: elisa.chiapponi@amadeus.com

***Abstract*—Residential proxies (RESIP) enable the tunneling of traffic through non-data center Internet connections. Previous research has focused on malicious software on end-user devices that made them part of RESIP networks. This study investigates RESIP networks that users voluntarily join in exchange for monetary rewards, aiming to understand the activities facilitated through these services. We developed a testbed environment to operate and monitor eight different residential proxy applications over 7.5 months, enabling us to collect and analyze 368 GB of proxied network traffic, the majority of which is encrypted.**

In this work, we highlight three distinct case studies that suggest these proxies are used in practices not advertised by the RESIP providers and in one case, shed light on the scale of the proxied campaigns. Firstly, we discuss the use of RESIPs on two dating apps, Tinder and happn, highlighting their likely role in facilitating fraudulent activities. Secondly, an analysis of metadata suggests that RESIPs may play a crucial part in phishing campaigns. Thirdly, a collaboration with a leading technology company in the travel industry allows us to analyze the behavior of web scrapers.

Our results underscore the need for enhanced detection mechanisms to mitigate fraud and protect users.

***Index Terms*—residential proxies (RESIP), web scraping, denial of inventory, phishing, fraud**

I. Introduction

The use of Residential Proxies (RESIP), is widespread through all couleur of Internet users. These proxies tunnel network traffic through residential Internet connections to offer network measurement vantage points from diverse access networks, circumvent bot and unwanted behavior detection, and facilitate significant cybercrime. For example RESIPs have been used to detect end-to-end violations of DNS, HTTP, and HTTPS [1], but also were an integral part of analyzing DNSSEC's public key infrastructure management [2]. Contrary to that type of academic work, web scraping bots employ these proxy networks to extract price information [3], circumvent rate-limited APIs, conduct competitive data mining, and harvest personal and financial data, among other activities [4].

Bot activities from these proxy networks are not limited to passive data harvesting; they also actively interact with their intended targets. For instance, scalper bots purchase high-demand consumer electronics [5], event tickets, limited-edition sneakers, and more, with the goal of reselling these items at higher prices [6]. While this does not necessarily result in an economic shortfall for vendors and producers, it does affect their relationship with legitimate customers who express frustration at being unable to purchase their desired items or having to pay the scalper's premium [7].

Vendors can be attacked by these bots directly through so-called denial of inventory attacks. Bots add items to the check-out process without completing the purchase or committing to the transaction [8]. The intention behind this is to tarnish a business's reputation or influence pricing algorithms [9], [10].

Lastly, cybercriminals take advantage of these network for their activities. For example, the U.S. Department of the Treasury's Office of Foreign Assets Control (OFAC) recently sanctioned a cybercrime network associated with the 911 S5 (residential proxy) botnet [11]. This network enabled cyber-criminals to carry out credit card theft, bomb threats, and child exploitation by supplying proxies that, when traced back, led to the victims computer instead of the criminal's [12].

These examples illustrate the significant security risks posed by RESIP networks, raising concerns about how these residential proxies are sourced and managed. Hence, a core question is: How do RESIP providers procure their residential proxies? Some RESIP providers answer this on their websites by saying that their proxies have been "procured ethically", without providing further details. As this is an unregulated market, RESIP providers are free to fill in their own definition of "ethical procurement". We mostly encounter statements such as that the potential bandwidth sharer needs to be informed beforehand that their bandwidth will be used by third parties and that they will receive a compensation for their participation. And lastly, the bandwidth sharer needs to explicitly agree to be part of the RESIP network. However, this is not necessarily true, as industry reports and academic research have shown. For example, free VPN apps can carry a RESIP proxy component to recruit unknowing users into the RESIP network [13], additionally mobile phone apps [14], [15], hacked routers [16] and IoT devices [17] can serve as payload delivery vehicles.

In this study, we examine residential proxies that are voluntarily user-installed and therefore would be inline with the definition of "ethically procured" which many RESIP providers

978-3-903176-66-9

state. The proxies we examine are all executable binaries (contrary to hidden mobile phone SDKs) and downloadable from the RESIP providers' websites, and they clearly state that the user will share their network with third parties. Users are enticed by the RESIP providers to share their Internet connection with the use colorful language such as "Ready to make sweet money?" or "Youre already paying for an internet connection (or two), so why not turn it into a steady passive income source?". Essentially, all RESIP providers offer economic incentives to join their RESIP network.

To conduct our measurements, we have registered with eight distinct RESIP providers, which explicitly promoted a bandwidth sharing application, and recorded the resulting network traffic by running their proxy applications for several months. We now introduce the term "bandwidth broker (BB)", because in many cases RESIP providers are not transparent about the source of their proxies. Similarly, some of the bandwidth-sharing recruitment sites we identified do not sell RESIPs directly. Instead, they only provide the proxy software and likely sell access to their network to other RESIP providers.

We think that the word broker is a more accurate description of an ecosystem in which a user decides to share their bandwidth for profit, similar to a commodities market.

In this contribution, we examine the traffic flowing through user-installed residential proxy networks. We present three case studies: interactions with dating applications, phishing activities, and sophisticated scraping bots. The final case study is expanded with insights from a targeted technology company. Our findings highlight a growing concern for network operators, who now face malicious activities originating not only from traditional malware but also from user-initiated behaviors facilitated by RESIPs.

II. Background & Related Work

In this section, we position our work in the context of current RESIP research and provide a short summary of JA4 fingerprinting, which we use in our analysis.

A. Research on RESIPs

In recent years, several studies on RESIP services have been published. In 2019, Mi et al. conducted the first comprehensive study in this area. They purchased access to five RESIP services and sent approximately 62 million labeled requests through the acquired proxies towards destinations under their control. This resulted in a data set containing more than 6.1 million unique residential IPs. In a follow-up step, they probed these IPs for open ports and banners, discovering that more than 200 000 responding hosts were IoT devices. Combined with the fact that 4 out of the 5 measured RESIP services did not specify how they recruit users led them to believe that the devices most likely were compromised [17].

A similar infrastructure has been used to characterize the Chinese RESIP ecosystem and the security risks of the more than 9 million collected IPs [18] as well as validate RESIP detection techniques and reveal insights of the RESIP's inner workings[19], [20]. Moreover, the data set collected in [17] was used to compare RESIP and open proxies [21]. Finally, Khan et al. [22] studied the geo-unblocking capability of commercial VPN services. They exposed how these VPN providers use residential proxies to bypass geoblocking.

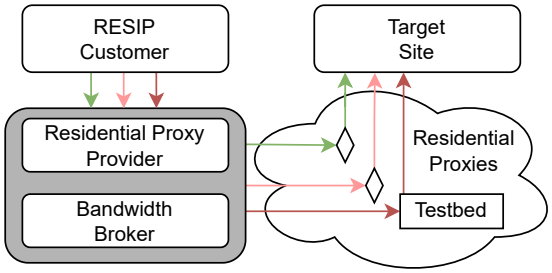

Fig. 1: An overview of the Bandwidth Broker ecosystem and how tunneled traffic flows.

Between 2021 and 2022, studies on mobile devices as RESIP have been published [15], [23]. The focus of these works is to understand how a mobile device becomes part of a RESIP network without the owner's knowledge.

Our work offers a new perspective compared to previous studies. We integrated our testbed into a voluntarily installed RESIP network, where we neither decide on nor control the traffic passing through our machines. By running the executable binaries promoted to regular users, who share their bandwidth in exchange for economic compensation, we recorded every forwarded byte of traffic. This approach provided us with a unique opportunity to understand the types of activities these proxies enable.

Figure 1 presents an overview of the RESIP/BB ecosystem. RESIP customers who wish to leverage residential proxies can either subscribe to the service at a RESIP provider, who does not disclose the source of their proxies, or at a BB, that instead does so. The grey box surrounding these two actors symbolizes their opaque nature as well as their currently unknown relationships with each other. Purchasing proxies allows the RESIP customer to forward traffic to their target site, by sending their data to an ingestion proxy from the RESIP provider or BB. Their traffic may be routed through one or multiple proxies (represented by the diamond shapes in Figure 1) from the RESIP provider's network.

To become part of a RESIP network, we offered the available bandwidth from our testbed to a BB and became one of these proxies. This results in our testbed only ever seeing a portion of the RESIP customer's traffic. This is illustrated in Figure 1 with the differently colored arrows.

Recently, Huang et al. [24] proposed a measurement setup similar to ours to analyze the network traffic produced by RESIP and differentiate it from the traffic created directly by a device based on machine learning. They used PacketStream, IPRoyal, and Honeygain and they deployed nodes in US and China. Differently from them, we propose a European testbed, we analyze a larger number of bandwidth brokers and we are

in the unique position of analyzing some of the data in transit as well as their activity on the target sites (see Section IV-C).

B. JA4 Fingerprints

In our analysis, we use JA4 fingerprints to differentiate between the various TLS clients detected in our network flows [25]. The fingerprint consists of three distinct parts.

The first part contains six human-readable parameters: the detected protocol (TCP/QUIC), TLS version, if a server name indicator was present, the number of transmitted cipher suites as well as extensions and finally the first ALPN value. The second part of the fingerprint is a truncated SHA256 hash of the transmitted ciphers, in a sorted order. The last part of the fingerprint consists of another truncated SHA256 hash. The components for the hash are a sorted list of TLS extension, concatenated with the signature algorithms in order of appearance.

All three parts are then concatenated to form the fingerprint. The fingerprints strive to be unique for each TLS client, but when multiple programs use the same underlying TLS client, i.e. *Google Chrome* and Microsoft's Chromium based browser *Edge* it will result in the same fingerprint. This immediately shows that additional features (such as user-agent) are needed to distinguish between similar software that makes use of the same TLS client engine.

III. Approach

In this section, we describe how we selected the bandwidth brokers on which we perform our measurements, describe our measurement setup and outline our data processing.

A. Bandwidth Broker Selection

To identify BBs for our measurements, we conducted a preliminary study in October 2023. We utilized search engines and used keywords such as "earn money passively" or "bandwidth sharing money". The search results could be categorized into three types: links to BB websites, blog posts discussing various passive income strategies including bandwidth sharing, and social media discussions on platforms like Reddit where users shared their preferred BBs. From these sources, we compiled a list of potential BBs.

A crucial criterion was the availability of the proxy application as a Windows binary, reflecting our goal to mimic an average Internet end-user, most of whom use Windows [26]. Among this category, we chose the BBs that appeared more frequently in our searches. Our final list of BBs (in no particular order) includes: Honeygain,IP Royal Pawns, BrightVPN, earn.fm, PacketStream, Packetshare, Repocket, and Proxyrack. BrightVPN is a product of Bright Data (previously known as Luminati), who also have a dedicated bandwidth sharing app called EarnApp. EarnApps terms of service forbids the usage of their client inside a virtual machine, the method we used for installation, which is why we opted to share our bandwidth through their free VPN application. We did notice that some of these BBs - Honeygain, IP Royal Pawns, Bright Data - appeared more frequently than others, suggesting that they are a bigger actor in this ecosystem.

B. Testbed

1) Hardware: The core principle of our testbed is that it should be easily deployable across various infrastructures and networks. Our setup allows us to quickly deploy additional measurement nodes for future measurements as each BB runs on a dedicated Windows VM. The VMs accept incoming connections from our central data collection hub, on fixed intervals. Our central hub then performs several tasks, such as retrieving the captured network traffic, as well as running a general health check on all components. The health check employs multiple strategies to verify whether the components are still operational (e.g. CPU utilization to detect hung processes), checks if the processes are actively running and assess the overall network usage. In case of crash, the system is able to restart and resume operations autonomously. The central hub can also push firewall rules to the measurement VMs, should it be necessary (e.g. in the case of severe abuse such as DDoS). We verified that our VMs can capture sustained network traffic of at most 1 Gbps, should one of the proxies require this amount of bandwidth. We also created an additional Windows VM without any proxy application, which we call the baseline VM. Its purpose is to record the inherent network traffic of a modern Windows installation, which can then be used as a filter to distinguish legitimate Windows traffic on the proxy VMs.

2) Network: In addition to segregating each BB through an individual VM, we allocate up to 1 Gbps of dedicated bandwidth per BB. This allocation is based on claims by some BBs that they support geo-unblocking of multimedia content, as well as findings from [22], which demonstrated that RESIPs are used for high-bandwidth applications like relaying multimedia streams. The hypervisor hosting the VMs connects to the wider Internet via a 10 Gbps interface, providing ample headroom for different BBs to utilize burst bandwidth simultaneously. Furthermore, each BB is allotted a dedicated IPv4 address from a /24 subnet, so that each proxy's interaction online is linked to a single IP address.

Lastly, to relay RESIP traffic, one must be considered residential. While we were not able to deploy our testbed at residential ISPs directly, we were fortunate enough that our /24 IPv4 subnet was considered to be residential according to several different IP metadata providers. In some cases, the online portal or the proxy program confirmed as well that our IP address was seen as residential. Proxyrack considers our device of type "Residential ($0.50)", meaning that every GB of traffic we tunnel would be compensated with 50 US Dollar cents. IP Royal Pawns only tunnels traffic through what they consider residential connections, and we can confirm that we have processed data for them. Honeygain also accepts data center proxies, but they notified us (through the proxy application) that our proxy is eligible for their "Content Delivery" feature. This feature is only accessible to Windows users who are of type residential [27]. Finally, also earn.fm showed in their interface how much traffic we have proxied through a graph with two lines, one line represents the amount

of data center traffic, which in our case constantly rested at zero, and the other represents residential that matched the amount of data we saw being proxied.

The other providers did not specifically state that our IPs were of type residential. However, we recorded outgoing traffic for all of them hence we assume that we fulfilled their requirements.

C. PCAP Processing

We exhaustively collect all network traffic entering and exiting the VMs as PCAPs. We convert the data to a more manageable format on an ad-hoc basis only before analysis. To analyze our data, we create network flows using Tranalyzer, a flow-based traffic analyzer built upon a flexible plugin-based architecture [28]. Tranalyzer's plugins extend the classic 5-tuple packet flow (source IP address, source port, destination IP address, destination port, transport protocol), depending on the loaded plugins. In our study, we make extensive use of the sslDecode plugin to analyze SSL/TLS traffic. Depending on the enabled options, sslDecode can display various properties of TLS connections, such as the list of submitted and ultimately agreed upon TLS cipher(s), the server name indication (SNI), and, in the most recent version also a JA4 fingerprint derived from the SSL/TLS Client Hello and Server Hello records [25]. As we do not perform man-in-the-middle attacks we cannot run any analysis on the payload itself. However, plaintext traffic like HTTP or DNS remains readable, and the payloads are kept for future analysis.

D. Case Study Selection

As mentioned in Section I, BBs often use vague language to entice users into participating in bandwidth sharing, with statements such as: "This bandwidth [...] is used by businesses for various online tasks,"[1] "Your traffic is used by verified and authentic businesses only,"[2] and "We use your bandwidth and IP address to help users and businesses worldwide avoid geo-restrictions, IP bans, and other blocks."[3]

RESIP seller websites offer slightly more transparency, listing use cases like web/price scraping, travel fare aggregation, and circumventing multi-account detection on social media platforms. For example: "Most social media sites have strict limitations on the number of accounts you can create and operate [...] the best way to bypass these limitations is by using a [residential] proxy."[4]

Since most of the traffic we collected is encrypted, we cannot directly verify these claims. Instead, we conducted a best-effort analysis on traffic that appears to fall outside the advertised use cases of the BBs/RESIPs. Traffic to e-commerce websites, as indicated by their FQDN, is likely related to the advertised web/price scraping and was excluded from further analysis.

[1]https://packetstream.io/share-bandwidth/
[2]https://www.packetshare.io/
[3]https://pawns.app/internet-sharing/
[4]https://iproyal.com/other-proxies/facebook-proxy/

TABLE I: Overview of our collected data from each BB including our economic compensation (up until 2024-07-16).

Bandwidth Broker	Start date	Proxied	Flows	Earnings
BrightVPN	2024-03-07	190GB	1.10M	free VPN
earn.fm	2024-04-17	9GB	0.28M	1.69 USD
Honeygain	2024-01-01	48GB	3.90M	20.55 USD
Packetshare	2024-02-27	51GB	2.37M	10.34 USD
PacketStream	2024-04-25	2GB	0.64M	0.21 USD
IP Royal Pawns	2023-11-17	55GB	2.62M	11.48 USD
Proxyrack	2024-01-01	3GB	1.12M	2.07 USD
Repocket	2024-01-01	10GB	1.79M	7.2 USD
Total		368GB	13.82M	53.54 USD

However, we also unexpectedly observed traffic to dating applications. Furthermore, the correlation of certain FQDNs combined with TLS fingerprinting led us to suspect that there is phishing traffic proxied through our nodes. Lastly, our collaboration with a technology company in the travel industry let us look closer into traffic aimed at travel related websites. We not only found bots scraping data, but also sophisticated bots trying to manipulate pricing and availability of items, which we further discuss in Section IV-C.

Therefore, we focus on RESIP use cases that clearly breach the terms of service (ToS) of targeted websites, posing risks not only to the platforms but also to users sharing their bandwidth and the network operators managing those networks.

IV. Results

As of 2024-07-16, our measurement setup has collected 368GB of data over a span of 7.5 months. Table I shows the totals for each BB we selected. We used IP Royal Pawns in our preliminary study to test our measurement setup, hence data collection for it started earlier. Some of the later additions are due to those providers not appearing in our previous search (see Section III-A), or not offering a Windows binary at that time. Although we detected occasional traces of other protocols such as HTTP or SMTP, most of the registered traffic is encrypted (around 99%). For this study, we will focus exclusively on TLS-protected traffic and we present three relevant case studies to show potentially malicious activities proxied by RESIPs. The first case study discusses the use of residential proxies on dating applications (Section IV-A). The second case study highlights the use of RESIPs in suspected phishing attempts (Section IV-B). Lastly, we examine the involvement of RESIPs in price scraping campaigns with additional data from an industry partner (Section IV-C).

A. Dating applications

IP Royal's traffic we observed around 20 000 unique FQDNs. Sorting those FQDNs by the number of recorded flows we notice two unexpected sites, "happn" and "Tinder". They are both mobile-based social networks that facilitates romantic and interpersonal connections between users through a location-based matching system.

Recently, dating sites have been at the center of journalistic attention [29], [30] due to an increase in fraud, with some

egregious stories even reaching worldwide audiences through true-crime productions [31]. In fact, the Federal Bureau of Investigation (FBI) [32] and consumer banks issue advisories stating that one should never send money to "matches" on dating sites [33]. The U.S. Department of Homeland Security has stated that [34] professional romance scams often have a transnational component, making the ability to easily create multiple dating platform accounts with residential proxies advantageous to scammers [35].

At their core, dating applications are designed to facilitate meetups between two parties who have "matched" with each other. However, considering the previously mentioned professional scammers and the purpose of RESIPs, which is to spoof a location and present a residential user profile, it is unlikely that a user of the proxy intends to meet the potential "match". We also do not consider RESIPs to be a form of privacy protection for two main reasons: first, seeking privacy on a dating application seems to be an oxymoron; and second, the cost of renting RESIPs is significantly higher than purchasing VPN services for privacy. Crucially, VPN services typically do not allow the creation of new accounts on these platforms, whereas residential proxies do permit this functionality [36].

To get a better picture of the traffic we proxied to Tinder and happn, we plotted all flows on a 24-hour histogram, creating a long-term view of the daily access pattern. Figure 2 shows the result for Tinder (a similar pattern also holds true for happn). Each color represents a weekday, starting with Monday on the bottom in red and ending with Sunday on top in pink. While we do not suggest that one weekday is more active than the other, we want to highlight the time of day during which the requests were made. Most requests for both dating applications occurred between 08:00 and 22:00 UTC, which strongly overlaps with the general day and night rhythm in Western Europe. This suggests interaction with Tinder users who live in a Western European timezone. Since Tinder profiles are available around the clock, this traffic does not appear to be of a general data scraping nature, because we then would expect a more uniformly distributed pattern.

In summary: 1) interactions with matches are very likely not with the goal of meeting *offline*, 2) contrary to (privacy) VPNs, RESIPs allow for the creation of multiple dating platform accounts at a substantially higher cost, and 3) we notice a day and night pattern which leads us to believe that we observe human interaction and not bots. Our observations combined with the law enforcement advisory and news items, let us believe that we are observing one or multiple actors who use RESIPs with malicious intent.

Takeaway: RESIPs possibly support malicious actors in various ways to target users of online dating platforms. RESIPs do that by not only facilitate rapid account creation but also enabling overseas actors to pose as local residents.

B. Phishing

As we mentioned in Section III-C, we also extracted JA4 fingerprints from all the flows we collected. We filtered our data based on the JA4 fingerprint of a popular open-source

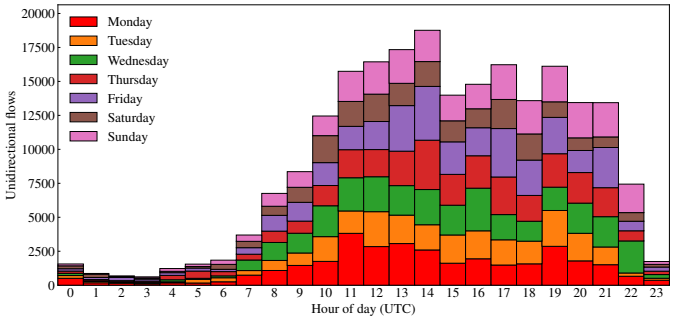

Fig. 2: Tinder access on a 24-hour scale

TABLE II: Number of detected flows with the Evilginx JA4 fingerprint

Bandwidth Broker	Flows	Bandwidth Broker	Flows
PacketStream	286 037	Proxyrack	915
IP Royal Pawns	13 283	Repocket	752
Packetshare	12 050	earn.fm	118
Honeygain	5675	BrightVPN	2

and modular reverse proxy phishing toolkit called Evilginx (JA4: `t13d191000_9dc949149365_e7c285222651`). We verified this fingerprint by running and recording the TLS traffic of a working Evilginx installation. Evilginx operates as shown in Figure 3: ① If a victim clicks on a phishing link (also called a *lure*), ② then Evilginx will open a connection through a residential proxy to the target site. Since Evilginx cannot decrypt TLS, it performs a two-part man-in-the-middle attack by forwarding the retrieved website data from step ② in a new TLS connection to the victim. Any login credentials entered by the victim can be read in plain text by Evilginx, as the initial TLS connection from the victim terminates on the Evilginx server. If the targeted site requires two-factor authentication (2FA), Evilginx forwards the challenge as well, allowing it to extract session cookies, which enable an attacker to log in without credentials and a 2FA challenge (step ③).

The RESIP from step ② serves two purposes. Firstly, it circumvents proxy/VPN detection at the target site, because often logging in from known proxy or data center IPs will trigger security mechanisms like a denied login and will very likely alert the user by email of a suspicious login attempt. Secondly, it masks the attacker's identity, because multiple other proxies or VPNs may be put between the Evilginx server and the RESIP. Before we continue we want to bring the reader's attention back to step ②: It is Evilginx that opens a connection to the target site, therefore the TLS client's fingerprint we record on our proxy belongs to the Evilginx software. The author of Evilginx recently mentioned that during internal testing he found that his phishing toolkit was automatically blocked on some sites and only after changing the TLS client parameters (and thus also the resulting JA4 fingerprint) would it successfully work again. Consequently, the upcoming *Evilginx Pro* version will feature a customizable JA4 fingerprint [37].

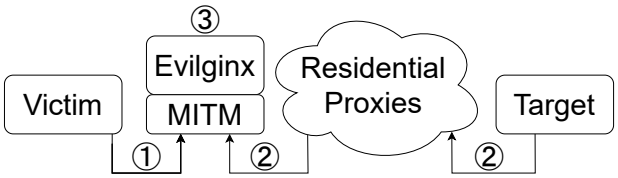

Fig. 3: An overview of Evilginx' reverse-proxy phishing flow.

In Table II we present the number of flows with the Evilginx JA4 fingerprint per BB. PacketStream shows a significantly higher number of flows with its fingerprint compared to the other providers and these flows account for roughly 44% of PacketStream's total flows (as described in Table I). The encryption of TLS data limits the amount of supporting evidence we can extract from the flow itself when it comes to user behavior at a target website, which hampers our analysis capabilities. However, a field that is included in the TLS handshake is the server name indicator (SNI), which contains a FQDN. Looking at PacketStream's flows, we can see that 99.9% of those flows contain the FQDN: `client.packetstream.io`. We believe this is an endpoint where the PacketStream proxy binary fetches configuration data or sends liveness checks of the proxy system to their proxy management back-end. Analysis of PacketStream's proxy binary and Evilginx' source code show that both are developed in the Go programming language which explains the JA4 fingerprint overlap. We consequently believe that FQDNs are valuable supporting evidence to characterize flows, and as a result we focused our analysis on flows from the other Bandwidth Brokers.

Evilginx uses files called *phishlets*, which contain phishing-target specific configuration, including the FQDN on which the authentication for the website is hosted. We non-exhaustively searched for FQDNs from publicly available *phishlets*[5] in our data. Packetshare flows that match the Evilginx JA4 signature and phishing FQDNs show connections to `account.booking.com`, `outlook.office365.com`, `discord.com`, `paypal.com`, and more. While some of these targets have some similarity with possible scraping targets, these FQDNs redirect to a login page or otherwise static landing page. Online banking portals, cryptocurrency exchanges, and entertainment sites were among the results too.

We strongly believe that malicious use of residential proxies is evident from the detected flows and the diversity of targeted domains. These findings highlight the role user-installed RESIPs play in enabling phishing campaigns. Furthermore, it highlights previously undisclosed risks for bandwidth sharers and the operators whose network they use. We want to underscore the necessity for improved detection mechanisms and vigilant monitoring by network operators.

Takeaway: User-installed RESIPs are very likely being leveraged to conduct sophisticated phishing attacks, posing a significant threat to online security.

[5]https://github.com/An0nUD4Y/Evilginx2-Phishlets

C. Sophisticated Scraping Bots

Our final analysis focuses on traffic directed toward travel industry-related websites. We reached out to two key players in the travel industry; one international and one regional. We collaborated with the international player, one of the worlds leading technology companies in the travel sector (hereafter referred to as TC), which put us in the unique position to analyze proxied connections towards targeted sites. We presented our findings to the regional player, and they were able to confirm our observations and conclusions.

The travel industry is a prime target for bots, with numerous sophisticated campaigns reportedly originating from residential IPs [38]. Analysis of traffic data at TC revealed that these sophisticated campaigns make substantial use of RESIPs for their activities.

TC protects travel websites that provide pricing information and booking services. A typical user flow involves accessing a landing page to adjust parameters like date and number of travelers, which affect booking prices. Subsequently, a new page shows the results and matching prices. The TC uses a third-party anti-bot product to extract parameters from requests and block those matching bot signatures. Simple rules filter naive and moderately complex bots, but sophisticated campaigns require manual intervention. TC analysts analyze and correlate parameters with domain-specific information to produce new bot signatures daily.

To look for our proxies' involvement, we filtered all requests received at TC by our proxies' IP addresses. Matches were found for our BrightVPN, Honeygain, and Packetshare proxies. They accessed five protected domains and performed web scraping for price extraction. The following subsections describe the different types of activity found.

1) Direct Targeting: In this subsection, we analyze three campaigns employing a technique termed "direct targeting". Bots using direct targeting bypass the non-malicious user flow by constructing specific queries with the required parameters to access pricing pages directly. Each of our proxies initiated one or two direct requests to travel sites. However, the corresponding bot signatures sent a significantly larger number of requests, as detailed in Table III. The table indicates the number of requests seen per IP from the perspective of TC showing the vast distribution of the IPs among ASes and countries.

These requests were blocked by TC due to matching manually created bot signatures. After the time window of observation, we did not encounter any traffic matching these bot signatures. However, in two out of three cases, we observed other bot signatures triggering on the same domains with similar traffic shape. This leads us to believe that the scrapers changed their approach in order to prevent detection while keeping the campaign active.

For BrightVPN, we observed one matching request (2024/05/16, 15:08 UTC). On that day, the same bot signature blocked 9099 requests from 9021 IPs, between 15:00 and

TABLE III: Number of IPs and corresponding number of direct requests, distribution among ASes and countries for the bot signature matching the IPs of our proxies.

BrightVPN IP		Honeygain IP		Packetshare IP	
IPs	**Requests**	**IPs**	**Requests**	**IPs**	**Requests**
8949	1	13116	1	2876	1
70	2	123	2	74	2
2	fewer than 7	1	5	5	3
ASes	**Countries**	**ASes**	**Countries**	**ASes**	**Countries**
1736	153	1567	134	890	122

16:00 UTC. Over the next 36 hours, TC blocked traffic targeting the domain with bot signatures different from the initial one but with a similar timing pattern. This suggests they belonged to the same campaign. In total, they observed 7 different bot signatures, each appearing approximately 46 minutes after the previous one was blocked by analysts.

For Honeygain, we found two requests sent at the same time (2024/06/05, 04:44 UTC), showing possible retransmission. On that day, the bot signature matched 13 367 requests. Unlike BrightVPN, this campaign lasted for 15 hours without any subsequent bot signatures showing the same traffic pattern. However, during the campaign, TC observed frequent parameter rotations (on average every 23 minutes) suggesting the actor tried to evade detection before ceasing attempts.

For Packetshare, we registered a single request (2024/04/04, 12:41 UTC). On that day, the rule blocked 3039 requests from 2955 IPs in a 3-hour window. Over the next 5 days, analysts discovered and blocked 13 bot signatures with a similar traffic pattern. On average, the bot changed its signature approximately every 6.5 hours after being blocked.

This data highlights the widespread impact of scraping campaigns employing residential IPs. It reveals that the actors behind these campaigns target specific pages directly, continuously rotating residential IP addresses in an attempt to evade detection. Furthermore, it indicates that they actively monitor results and promptly alter bot signatures upon detection to maintain operational effectiveness.

2) RTT Detection and Signature Evasion: The third party anti-bot product in use at TC implements a recently published technique for RESIP detection based on round trip time (RTT) [19]. The use of this technique shows additional scraping campaigns that were not blocked by any other bot signature. Our matching proxies were from BrightVPN and Honeygain.

For BrightVPN, our proxy tunneled a total of 78 requests in 13 minutes directly to the pricing page (2024/05/21, 04:30-04:43 UTC). TC observed an already ongoing campaign targeting the same domain, but assumes that these two events are not linked. The ongoing campaign started one day earlier and spiked in traffic only on full hours of the day, furthermore no more than 3 requests per IP were observed. Our identified requests however, arrived outside the previously stated peak times and considering the number of requests, it seems unlikely that the two events were linked. More likely, this was an isolated phenomenon; perhaps an individual user testing before launching a stronger campaign, or someone experimenting with a residential IP service for the first time.

For Honeygain, we observed two instances where the bot followed the complete user flow instead of directly targeting the pricing page. The first one occurred on 2024/05/01, 18:50-19:00 UTC, during which 32 flows were sent within 10 minutes. The second instance took place on 2024/05/29, 22:00-22:10 UTC, with 57 flows sent within 10 minutes. In both cases, all requests were flagged by RTT detection.

These two occurrences displayed similar timing patterns, but the number of requests varied. In total, the RTT detection flagged requests from 2024/04/29 to 2024/06/06 (39 days), as originating from RESIPs for the domain under study. However, not all flagged requests followed the 10-minute interval pattern, and the majority of IPs made only few requests. This leads to believe that multiple scraping campaigns were running in parallel, with our proxies participating in only one of them.

If we filter only for IPs that appeared in 10-minute intervals with multiple flows, we observe that the specific campaign was sporadically active over the 39 days. In addition to the previously mentioned occurrences, we identified 21 other instances, bringing the total to 23. These occurrences never happened on the same day and involved 17 different IPs. Four of these IPs appeared twice. Each IP sent between 31 and 65 requests within the 10-minute intervals.

In both the BrightVPN and Honeygain cases, we observed variations in parameters typically used for bot signatures, such as TLS settings. This supports the idea that manual attempts were made to find parameters that could avoid detection, suggesting the actor was testing different bot versions.

Takeaway: RESIPs are a preferred vehicle for sophisticated bot campaigns, where actors draw on a vast pool of fresh IPs that aid in rotating bot signatures to evade detection.

V. Limitations

As noted in Section II-A, our perspective on ongoing campaigns is inherently limited, as it is based on a single node from a single ISP per RESIP provider. As such, we can confirm the specific behavior of the traffic we observed exiting our node but cannot generalize our findings to a percentage of the overall traffic egressing the BB. Moreover, there could be campaigns not captured by our testbed but still proxied by these providers. As a matter of fact, only due to our collaboration with the TC (see Section IV-C), we were able to assess the following: While none of the IPs of our testbed participated in a denial of inventory attack, the RESIP RTT detection method identified some residential IPs attempting this attack against the TC's customers. Analyzing these IPs with the IP reputation service Spur [39], we found that the attacking IPs overlap with seven of the BBs we analyzed (all except earn.fm, which is not identified by Spur yet). To achieve a more comprehensive understanding, it would be beneficial to have measurement proxies in various access networks.

As discussed in Section III-B2, BBs that indicate on their online portals whether our supplied bandwidth is classified as *data center* or *residential* have consistently categorized

our bandwidth as *residential*. However, we do not know if BBs further differentiate the proxied traffic based on specific characteristics of the residential environment, such as the ISP. For example, the US PacketStream node of Huang et al. [24] generated 2.7 TB of data in 6 months. Even though we measured a significantly larger number of BBs, our cumulative relayed bandwidth stays far below that amount.

Although this limitation restricts our view of the global landscape, the use cases presented in this work provide a solid foundation for understanding the different types of traffic proxied by these providers.

VI. Ethical Considerations

The work described in this paper has presented us with several different ethical dilemmas that affect many of the different stakeholders involved. In this section, we aim to provide an overview of the ethical challenges that we identified and explain how we handled them to minimize impact on the relevant stakeholders.

A. Ecosystem

The ecosystem of RESIP providers is very diverse, as are the usages of these services. As described in Section I these residential proxies are used in academic and marketing research. However, these services are also used for geo-unblocking or otherwise hiding the origin of the original user. Some of the RESIP providers do customer screening in the form of interviews, and/or requiring a detailed usage plan for their service. BBs on the other hand, do not require this.

The BBs claim that their RESIP are "procured ethically" by requiring an active installation by the user, in return for compensation. However, most ISPs state in their terms of service that it is not allowed to resell the internet connection service. Users installing the proxy software are probably unaware they are breaching their ToS, nor are users warned about this during the registration process.

Many service providers, such as streaming platforms, restrict access to residential networks to prevent VPNs from accessing content across regions due to licensing. RESIP providers offer a way to bypass these restrictions.

B. Protecting Stakeholders

We created accounts on the BB platforms as regular users, but did not inform the BB that we were doing this for measurement, as we did not want to influence the results. Nearly all traffic that we encountered used TLS, meaning that it was only possible to see the destination host, but not the traffic itself, see also Section III-C.

With our testbed we are providing capacity to BBs, but are doing so to be able to measure and analyse this traffic so that others can learn from this. We worked with ISPs and impacted service providers to gain a better understanding of the traffic flows going through, and to be able to characterise possible malicious traffic, documenting our efforts, so that others can learn from our analysis.

It was impossible for us to request informed consent from the users sending traffic through our proxies as there was no way for us to contact them directly. Additionally, for all traffic we observe, the source is of the BB, not the user, so we are unable to trace back the original user. To mitigate any impact on these individuals, we only examined the hosts they were connecting to, and the size of the traffic, and only produce aggregated results here. We secured our testbed so that only the researchers had access to the data, and further restricted access to just our university's network.

An approval record from our institution's ethics board is available under registration number ECIS-230340.

VII. Conclusion

In this study, we analyzed the traffic of voluntarily installed residential proxies (RESIPs) and their usage patterns. We highlighted three different use cases based on the data we collected: Their use on online dating apps, phishing campaigns as well as advanced web scraping. Our findings indicate that RESIPs can be used to create multiple accounts and potentially engage in fraudulent activities on the dating platforms. Furthermore, they allow criminals to execute advanced phishing campaigns, as even two-factor authentication can be circumvented. Lastly, RESIPs enable malicious actors to perform advanced scraping attacks, by easily rotating among many different residential IP addresses. This work significantly advances our understanding of how RESIPs facilitate such activities and the associated risks.

By shedding light on the misuse of RESIPs, we contribute to better security practices and inform both platform providers and network operators about the correlated potential threats. These findings aid in developing better detection and prevention strategies against RESIP-based fraud, underscoring the need for improved security measures and continued research.

References

[1] T. Chung, D. Choffnes, and A. Mislove, "Tunneling for Transparency: A Large-Scale Analysis of End-to-End Violations in the Internet," in *Proceedings of the 2016 Internet Measurement Conference*. ACM, pp. 199–213.

[2] T. Chung, R. van Rijswijk-Deij, B. Chandrasekaran, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson, "A Longitudinal, End-to-End View of the DNSSEC Ecosystem." USENIX Association, pp. 1307–1322.

[3] E. Chiapponi, M. Dacier, O. Thonnard, M. Fangar, M. Mattsson, and V. Rigal, "An industrial perspective on web scraping characteristics and open issues," in *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S)*. IEEE, pp. 5–8.

[4] Imperva. Bad Bot Report 2021: The Pandemic of the Internet — Imperva. Blog. [Online]. Available: https://www.imperva.com/blog/bad-bot-report-2021-the-pandemic-of-the-internet/

[5] M. Kan. Inside the GPU Shortage: Why You Still Can't Buy a Graphics Card. PCMAG. [Online]. Available: https://www.pcmag.com/news/inside-the-gpu-shortage-why-you-still-cant-buy-a-graphics-card

[6] HUMAN Security, "2022 Automated Fraud Benchmark Report." [Online]. Available: https://www.humansecurity.com/hubfs/HUMAN_Report_2022-Automated-Fraud-Benchmark-Report.pdf

[7] S. E. Needleman, "Desperate Parents Turn to Shopping Bots to Hunt for Hottest Christmas Gifts." [Online]. Available: https://www.wsj.com/articles/desperate-parents-turn-to-shopping-bots-to-hunt-for-hottest-christmas-gifts-11637417633

[8] OWASP. OAT-021 Denial of Inventory — OWASP Foundation. [Online]. Available: https://owasp.org/www-project-automated-threats-to-web-applications/assets/oats/EN/OAT-021_Denial_of_Inventory.html

[9] HUMAN Security. What are Denial of Inventory and Scalping Attacks? — Detection & Prevention. HUMAN. [Online]. Available: https://www.humansecurity.com/learn/topics/what-are-denial-of-inventory-and-scalping-attacks

[10] V. Shetty. Tis the Season for Denial of Inventory Attacks. Arkose Labs. [Online]. Available: https://www.arkoselabs.com/blog/season-denial-of-inventory/

[11] U.S. Department of the Treasury. Treasury Sanctions a Cybercrime Network Associated with the 911 S5 Botnet. U.S. Department of the Treasury. [Online]. Available: https://home.treasury.gov/news/press-releases/jy2375

[12] Federal Bureau of Investigation (FBI). Guidance on the 911 S5 Residential Proxy Service. Guidance on the 911 S5 Residential Proxy Service. [Online]. Available: https://www.ic3.gov/Media/Y2024/PSA240529

[13] D. Goodin. US sanctions operators of free VPN that routed crime traffic through user PCs. Ars Technica. [Online]. Available: https://arstechnica.com/security/2024/05/us-sanctions-operators-of-free-vpn-that-routed-crime-traffic-through-user-pcs/

[14] HUMAN Security. The Impact of Residential Proxy Networks: PROXYLIB. HUMAN. [Online]. Available: https://www.humansecurity.com/learn/blog/the-impact-of-residential-proxy-networks-proxylib

[15] X. Mi, S. Tang, Z. Li, X. Liao, F. Qian, and X. Wang, "Your Phone is My Proxy: Detecting and Understanding Mobile Proxy Networks," in *Proceedings 2021 Network and Distributed System Security Symposium*. Internet Society.

[16] S. Fadilpai. Thousands of Asus routers taken over by malware to form new proxy service. TechRadar. [Online]. Available: https://www.techradar.com/pro/security/thousands-of-asus-routers-taken-over-by-malware-to-form-new-proxy-service

[17] X. Mi, X. Feng, X. Liao, B. Liu, X. Wang, F. Qian, Z. Li, S. Alrwais, L. Sun, and Y. Liu, "Resident Evil: Understanding Residential IP Proxy as a Dark Service," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, pp. 1185–1201.

[18] M. Yang, Y. Yu, X. Mi, S. Tang, S. Guo, Y. Li, X. Zheng, and H. Duan, "An Extensive Study of Residential Proxies in China," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. ACM, pp. 3049–3062.

[19] E. Chiapponi, M. Dacier, O. Thonnard, M. Fangar, and V. Rigal, "BAD-PASS: Bots Taking ADvantage of Proxy as a Service," in *Information Security Practice and Experience*, C. Su, D. Gritzalis, and V. Piuri, Eds. Springer International Publishing, vol. 13620, pp. 327–344.

[20] E. Chiapponi, M. Dacier, and O. Thonnard, "Inside Residential IP Proxies: Lessons Learned from Large Measurement Campaigns," in *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, pp. 501–512.

[21] J. Choi, M. Abuhamad, A. Abusnaina, A. Anwar, S. Alshamrani, J. Park, D. Nyang, and D. Mohaisen, "Understanding the Proxy Ecosystem: A Comparative Analysis of Residential and Open Proxies on the Internet," vol. 8, pp. 111 368–111 380.

[22] E. Khan, A. Sperotto, J. Van Der Ham, and R. Van Rijswijk-Deij, "Stranger VPNs: Investigating the Geo-Unblocking Capabilities of Commercial VPN Providers," in *Passive and Active Measurement*, A. Brunstrom, M. Flores, and M. Fiore, Eds. Springer Nature Switzerland, vol. 13882, pp. 46–68.

[23] A. Tosun, M. De Donno, N. Dragoni, and X. Fafoutis, "RESIP Host Detection: Identification of Malicious Residential IP Proxy Flows," in *2021 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, pp. 1–6.

[24] R. Huang, D. Zhao, X. Mi, and X. Wang. Shining Light into the Tunnel: Understanding and Classifying Network Traffic of Residential Proxies. [Online]. Available: http://arxiv.org/abs/2404.10610

[25] J. Althouse, "FoxIO-LLC/ja4," FoxIO. [Online]. Available: https://github.com/FoxIO-LLC/ja4

[26] StatCounter. Desktop operating system market share 2013-2024. Statista. [Online]. Available: https://www.statista.com/statistics/218089/global-market-share-of-windows-7/

[27] Honeygain. Honeygain Explains: Content Delivery. Honeygain. [Online]. Available: https://www.honeygain.com/blog/honeygain-explains-content-delivery/

[28] S. Burschka and B. Dupasquier, "Tranalyzer: Versatile high performance network traffic analyser," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, pp. 1–8.

[29] M. de Koning, "gaan we niet doen, tiger, zegt ze nog. en dan laat ze zich toch door haar tinder-date overhalen." [Online]. Available: https://www.nrc.nl/nieuws/2024/02/02/gaan-we-niet-doen-tiger-zegt-ze-nog-en-dan-laat-ze-zich-toch-door-haar-tinder-date-overhalen-a4187930

[30] M. Biino. How a romance scammer defrauded 3 Tinder dates out of over $100,000. Business Insider. [Online]. Available: https://www.businessinsider.com/romance-scammer-peter-gray-tinder-fraud-2024-5

[31] J. Dilillo. Who Is The Tinder Swindler? Netflix Tudum. [Online]. Available: https://www.netflix.com/tudum/articles/who-is-tinder-swindler-real-shimon-hayut

[32] Federal Bureau of Investigation (FBI). Romance Scams. Federal Bureau of Investigation. [Online]. Available: https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-scams-and-crimes/romance-scams

[33] HSBC UK. How To Avoid Romance Scams — Banking Scams - HSBC UK. [Online]. Available: https://www.hsbc.co.uk/help/security-centre/how-to-avoid-romance-scams/

[34] U.S. Department of Homeland Security. Looking for love online? Protect Yourself Against Romance Scams — Homeland Security. [Online]. Available: https://www.dhs.gov/hsi/insider/romance-scams-protect-yourself

[35] "How to Register and Login to Multiple Tinder Accounts Without Being Detected? @Vmlogin." [Online]. Available: https://www.youtube.com/watch?v=nyXUMTW0_ec

[36] Smartproxy. Buy the Best Tinder Proxy For Your Tinder Bots. Buying the Best Tinder Proxy For Your Tinder Bot. [Online]. Available: https://smartproxy.com/blog/tinder-proxy

[37] "Kuba Gretzky: Keynote: A Smooth Sea Never Made a Skilled Phisherman." [Online]. Available: https://www.youtube.com/watch?v=Nh99d3YnpI4

[38] Imperva. 2023 Imperva Bad Bot Report. Resource Library. [Online]. Available: https://www.imperva.com/resources/resource-library/reports/2023-imperva-bad-bot-report/

[39] Spur Intelligence Corporation. Beat fraud, boost revenue - Spur. [Online]. Available: https://spur.us/