International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I10P108

Volume 72 Issue 10, 43-49, October 2024
© 2024 Seventh Sense Research Group®

Original Article

Vulnerabilities, Privacy Risks, and Ethical Implications of Residential Proxy Services

Sudarshan Kumar Kaushik

Presidium Gurgaon Sector 57. USA.

Corresponding Author : awesomesud347@gmail.com

Received: 23 August 2024 Revised: 27 September 2024 Accepted: 12 October 2024 Published: 23 October 2024

Abstract - *The usage of proxy servers, which act as intermediates between two nodes, has seen a steady increase in use in recent years. These proxy servers effectively create a service called proxies that many proxy providers sell. These proxies come in various forms for various use cases, both ethical and non-ethical. Previous research on proxies focuses' on the detection, blocking, real-world usage, or comparative analysis of different types. This research rather takes a deep dive into a specific type of residential proxy and studies. It offers solutions to both a proxy provider and proxy users regarding the Vulnerabilities, Privacy risks, and Ethical Implications both parties may face when they work together to fulfil an ethical use case. Therefore, this paper studies the existing proxy protocols, encryption protocols, detection methods, selection methods, and ethical standards and creates a system of checks that can be split into two major types: Necessary and Optional. It furthermore finds that a proxy provider and a user must check for the necessary parameters to ensure success rates and ethical usage. Ultimately, this paper aims to bring out the best in residential proxies by empowering both proxy providers and users by providing them valuable data/insight about the current state and deemed industry standard that can be incorporated into their approach to residential proxies.*

Keywords - *IP addresses, Proxies, Proxy providers, Proxy servers, Residential proxies.*

1. Introduction

Proxy services sell access to proxy servers that intermediate traffic between servers and clients on the internet. When clients use these proxy services, their requests for data to web servers are sent to the proxy instead. The proxy then forwards the client's request to the server and the server's response back to the client [1]. Proxies thus help a client to mask their IP address as all communication is done via the proxy server's IP address. Proxies today come in various types- Open, Residential, Datacenter, and Reverse [2]- each aimed at specific clients and use cases. One of the fastest-growing proxy types, and the focus of this paper is residential proxies [3]. These proxies combine the anonymity of a standard proxy with a residential IP address, which makes them appear as unsuspicious internet users and bypass most proxy firewalls.

Hence, residential proxies have proved to be a powerful tool that can be used to achieve both ethical and unethical goals, i.e. Residential proxies power a majority of the traffic for Market Research, Ad Verification, AI dataset scraping, etc [4], while also being used to carry out coordinated DDOS attacks [5]. This dual nature of residential proxies also extends to their sourcing techniques: some proxy services rely on voluntary consent from a residential IP owner to share his/her IP for proxy purposes [6], whereas some proxy services are known to infect and use mobile devices, IoT devices, PUPs as residential proxy hosts. This paper discusses case studies of various residential IPv4-based proxy providers to determine the best types, detection methods [7] [8], selection methods, encryption methods, protocols and ethical implications of residential proxies. It also aims to shine a light on the major research gap addressing the characteristics of high-quality residential proxies/providers to both help select a residential proxy for legitimate use and also help web servers mitigate low quality residential proxies with an illegitimate use case. Throughout the paper, data from/about whois records [9] [10] [11] [12] [13]; geolocation databases [14] [15] [16]; port and service scans; response latency; ICMP response; spam scores and history [17] [18]; success rates; and sourcing methodology will be evaluated for each case/proxy service.

2. Methodology

This study employed a Systematic Literature Review (SLR) as a research approach to collect, synthesize, and appraise the findings of all available evidence on the topic to provide a comprehensive overview of the current state of the knowledge on this topic. This process involved using a systematic approach to identify, evaluate, and synthesize relevant studies to minimize bias and ensure the reliability and validity of the review. An Online search was conducted on Literary Databases to determine the most relevant literature published in English within the last 10 years on this topic. The articles included were on the basis of relevance to the main research focus, time of study, and frequency of practical application of their respective findings/suggestion.

Then, the retrieved articles were thoroughly analyzed based on topics of residential proxies: their vulnerabilities, privacy risks, and ethical implications. Furthermore, the discussion centers around the theme of providing/using the right residential proxy. It holds tests to determine the vulnerabilities of residential proxies that may hamper their provisioning or usage. The results are then analyzed based on standard parameters defined by existing IP Databases (IP2location, maxmind) and compiled into a checklist format to enhance clarity on the repercussions of failing each check.

3. Discussion

Residential Proxy servers act as relays of internet connections, wherein they isolate the internet connection to the public web server and proxy server in a way that hides the identity and IP Address of the requesting client. This working of the proxy server makes them a powerful tool that urges us to discuss their types and the vulnerabilities, privacy risks, and ethical implications associated with them.

3.1. Types of Residential Proxies

3.1.1. Rotating

These residential proxy types were the first ones to achieve significantly higher success rates as opposed to non-residential proxies. These proxies borrow the IP Address and bandwidth of a residential internet uplink by asking an internet uplink owner to download a proxy server app [19] [20].

In this way these services control a big pool of IP addresses and sell their business clients bandwidth, which allows the client to access the IP pool to proxy their requests through a large amount of IP addresses. These proxies were named rotating due to their inherent nature of instability in connection, which occurs because they are hosted on the devices of residential internet users (not in data centres) and are not dedicated to one client.

3.1.2. Static

Static residential proxies have recently emerged as a proxy type that takes the success rates of rotating residential proxies and improves on the missing stability and performance. These proxies use similar kinds of IP addresses as rotating res proxies–IP Addresses that are owned by and announced on Residential ISP's ASNs. However, instead of running a proxy server on the end user of ISPs and borrowing their IP address and bandwidth, these proxies are hosted in a datacenter connected to the ISP's ASN Uplink, which guarantees a better performance in latency, uptime and bandwidth

3.2. Vulnerabilities

Residential proxies of both types cater to numerous use cases and different types of websites. Therefore, they need to be checked for possible vulnerabilities, privacy risks and ethical implications that can affect their success rate.

3.2.1. IP Databases- A way to track Proxy Vulnerability

Renowned IP address databases [14] [15] [16] are used by almost every online website as databases to enable filtering features like geo-blocking on their firewalls. These databases have now started accounting for the possibility of an IP address being a proxy. This poses a threat to the effectiveness of residential proxies, as websites widely adopt these IP databases already to enable their firewall features. To avoid being detected and flagged at any IP Database, the proxy's IP address must appear to be of Residential standard to all checks employed by the database. The checks mentioned and discussed have been categorized into two types:

Necessary - These checks are essential from the cybersecurity and network safety standpoint.
Optional - These checks are solely related to the detection of proxies by databases.
Necessary - Port and Service Scans | Special Request Handling
Optional - Whois Records

3.2.2. Port and Service Scans
What are Ports and Port Scanning

The port number is the part of the addressing information used to identify the senders and receivers of messages in computer networking. Different port numbers are used to determine what protocol incoming traffic should be directed to. A port number identifies a specific process to which an Internet or other network message is to be forwarded when it arrives at a server. Ports are identified for each protocol and are considered as a communication endpoint [21]. All services running on a host are bound to a port number, including proxy software. Some common ports for common services are:

7-Echo

20-FTP Data

21-FTP Control

22-SSH

23-Telnet

25-SMTP

37-Time

53-DNS

80-HTTP

110-POP3

115-SFTP

143-IMAP

161-SNMP

443-HTTPS

546-DHCP Client

547-DHCP Server

1080-Socks5 Proxy

8080-Http/s proxy

Port scanning is a methodology wherein the TCP ports of an IPv4 address are scanned using varying techniques with the goal of identifying open ports and services running on the host device(proxy server, in this case) on that port. Port scans usually scan for open, closed and filtered ports. Open ports are generally considered a vulnerability point for a private network, as they reveal information about the services running on a host in a network.

Port and Service Scan Check

In order for residential proxy networks and hosts to avoid detection on the internet, the proxy host must be configured to look the same way as a residential ISP user.

Port Scan: Residential ISP IP vs Residential Proxy Host

It is found that the port scans of residential ISP IPs do not yield any open ports, as the majority of the ISPs use NAT (Network Address Translation) to distribute and share their IPs to multiple users.

<Port Scan>

```
Note: The host seems down. If it is really up but blocking our ping probes, try -Pn

Nmap done: 1 IP address (0 hosts up) scanned in 3.29 seconds.
```

<IP2location Database Info>

Usage Type: ISP/MOB

Anonymous Proxy: No

Proxy Usage type: ISP/MOB

Proxy Type: NA

Proxy ASN: NA

Threat: NA

Last Seen: NA

Whereas a detected residential Proxy returns open ports such as ports 25, 8080, 1080, and 443, which are default service ports

<Port Scan>

```
nmap scan report for <IP>
Host is up (0.28s latency).
PORT STATE SERVICE VERSION
18080/tcp open squid-proxy
| fingerprint-strings:
|  FourOhFourRequest:
|    HTTP/1.0 400 Bad Request
|    Content-Type: text/plain; charset=utf-8
|    X-Content-Type-Options: nosniff
|    Date: Fri, 30 Aug 2024 12:01:24 GMT
|    Content-Length: 13
|    Request
|    GenericLines, Help, Kerberos, LDAPSearchReq, LPDString, RTSPRequest, SSLSessionReq, TLSSessionReq, TerminalServerCookie:
|    HTTP/1.1 400 Bad Request
|    Content-Type: text/plain; charset=utf-8
|    Connection: close
|    Request
|  GetRequest, HTTPOptions:
|    HTTP/1.0 400 Bad Request
|    Content-Type: text/plain; charset=utf-8
|    X-Content-Type-Options: nosniff
|    Date: Fri, 30 Aug 2024 12:00:53 GMT
|    Content-Length: 13
```

<Ip2location

Database information>

Usage Type: ISP/MOB

Anonymous Proxy: Yes

Proxy Usage type: ISP/MOB

Proxy Type: RES

Proxy ASN: Yes

Threat: NA

Last Seen: Yes

An undetected Residential Proxy, does not return any open port with the exception of 1 port between the 50000-65000 range.

```
<Port Scan>

Nmap scan report for IP

Host is up (0.30s latency).

PORT      STATE    SERVICE VERSION

56600/tcp filtered unknown

Service detection was performed. Please report any incorrect results at https://nmap.org/submit/ .

Nmap done: 1 IP address (1 host up) scanned in 0.99 seconds.
```

<Ip2location Database Info>

Usage Type: ISP/MOB

Anonymous Proxy: No

Proxy Usage type: ISP/MOB

Proxy Type: NA

Proxy ASN: NA

Threat: NA

Last Seen: NA

Hence, for the purpose of residential proxies, it is plausible to close all commonly used service ports, including the one mentioned above, and use a usual port number above the 50000 mark. This methodology works because IP Databases are unable to scan all ports of an IP address due to being limited by both Computing power and IP resources.

3.3. Special Requests Handling

While proxying traffic, a proxy server might come across special packets or requests that may not be handled properly when proxied or are requests that are not necessary to answer. Thus, proxy servers need to be configured to handle these requests to remain undetected and ensure the flow of traffic.

3.3.1. ICMP Requests

The Internet Control Message Protocol is used to diagnose networks and report errors. The ICMP protocol requires a host to be online as it uses the host's responses to function. Common ICMP functions include ping, traceroute, echo, etc. The ICMP Protocol thus can be used to determine types of public networks, including the likes of proxy networks. To ensure the undetectability of a residential proxy, we need to mimic the ICMP policies of residential ISP networks. Residential ISP Networks use Network Address Translation (NAT) to distribute IP Addresses to their end users or hosts. This means ICMP requests to any Residential ISP IP are dropped, as there is no destination host. To mimic the same behavior, a proxy server or its firewall must be configured to drop all ICMP packets.

3.3.2. Unsolicited Requests

In an attempt to recognize a Proxy IP address, IP Databases may also engineer complex requests and contact the proxy server. This might lead to the proxy server replying to the request, hinting at its presence in the IP Database. One such request may be a connection initiation request with the wrong credentials to the proxy server. In the event of this request a proxy server might reply with an authentication error, as is done by many popular proxy software, squid, dante-proxy, etc. To evade detection, the proxy server should be configured to use authentication and respond only to authenticated requests.

3.3.3. Localhost Requests

Localhost is a hostname corresponding to the local host's loopback address. Any traffic sent to the localhost returns to the host without passing through any network. Some software often uses localhost requests to verify/determine the host machine. This poses a threat to hosts using proxies as, if a localhost request is proxied through a proxy server, the proxy server's local host might respond to it before proxying the response back to the host machine. This would mean a program running on a host machine will be communicating with the proxy server via the localhost destination. It is essential that the proxy server ignores all localhost requests and that the user host does not forward localhost requests to the proxy server.

3.4. Whois Records

The Whois database is hosted by the five Regional Internet Registries (RIRs) [9] [10] [11] [12] [13] and contains the ownership records of non-legacy IP [22] addresses. The database is widely used for identification, contact, and verification of IP ownership/lease. A few databases [14] [15] [16] use whois records along with the aforementioned methodologies to detect residential proxies. Specifically, databases look out for the Autonomous System Number (ASN) the IP Address is hosted on, and the current ownership/lease records.

Take a look at the who is records for a non-proxy residential IP Address:

```
OrgID:       ATTMO-3

OrgName:       AT&T Mobility LLC

AS:          AS20057

CanAllocate:
Street:       7277 164th Ave NE
```

Street: ATTN: IP Management

City: Redmond

State/Prov: WA

Country: US

PostalCode: 98052

Comment: For policy abuse issues, contact: abuse@att.net

Comment: Send all subpoena, Internet, Court order related matters to:

Comment: ATT National Compliance Center

Comment: 11760 US Hwy 1, Suite 600

Comment: North Palm Beach, FL 33408

Comment: Phone Number: 1-800-635-6840

Comment: Fax Number: 1-888-938-4715

Comment: Email: compcent@att.com

RegDate: 2008-10-10

Updated: 2021-06-26

OrgAdminHandle: MWE117-ARIN

OrgTechHandle: IPADM2-ARIN

OrgAbuseHandle: ATTAB1-ARIN

Source: ARIN

It is observed that the IP is hosted on a residential ASN and is registered under a residential ISP organization, as seen in block:

OrgID: ATTMO-3

OrgName: AT&T Mobility LLC

Hence, On the GEOIP Database IP2LOCATION under the field *Usage Type* the same IP address is classified as

(MOB)Mobile ISP

To ensure the same classification, any Residential Proxy IP Address must be hosted on a Residential ISP ASN and owned by a Residential ISP.

For private or non-residential IP spaces to classify as ISPs on the GEOIP database, the addresses must utilize the reassignment record provided by whois and reassign the addresses to the respective ISP whose ASN they are hosted on.

4. Testing and Result

Testing for the above Necessary and Optional Checks on a variety of IP addresses yields the following results:

Table 1. Testing and Results

Port and Service	Special Requests	Who is Records	Detected as Proxy	Detected non-Resi	Detected as Spam
			✓	✓	✓
✓			✓	✓	
	✓		✓	✓	✓
		✓	✓		✓
✓	✓			✓	
	✓	✓	✓		✓
✓		✓	✓		✓
✓	✓	✓			

The results thus obtained are consistent with the measures taken to prevent detection and being classified as spam. Furthermore, handling port and service scans and special requests remains crucial to the security of a proxy network.

5. Privacy Risks

5.1. Understanding Risks- If any

Residential proxy services are used by various businesses for the collection of web data. Their commendable success rates and speed attract usage from individuals as well. However, proxy services have evolved as a means of collecting public web data and have rarely been used or tested for use by an individual or any entity dealing with private information.

When it comes to privacy, Residential proxies are comparable to an upcoming service-Residential VPN. On drawing a comparison between the encryption, protocols, and use cases:

5.1.1. Residential Proxies

- Use lightweight unencrypted protocols like https and socks5 to ensure minimal latency
- Are mostly used by/targeted towards business clients which fetch public web data to gather some intelligence.

5.1.2. Residential VPNs

- Use tunnelling protocols that encrypt traffic, which allows them to bypass almost all ISP and geo filters for a trade-off in bandwidth and latency.
- Are mostly used by/targeted towards individual clients who intend to use the service to go about their daily web business.

It is quite clear that residential proxies can pose a significant privacy risk to individual users who intend to handle sensitive personal information, as residential proxies have been developed to be used by businesses to fetch already public web data.

5.2. Proxies + Encryption - Technically Possible?

Yes. There are numerous implementations of encryption standards within proxies, yet residential proxy services have adopted none. Shadowsocks is one such implementation. Shadowsocks is a proxy server software that uses the socks5 protocol and allows the user to implement AEAD encryption. This software has been continually tested for success rates and has been widely used in China since 2012 to circumvent internet censorship.

6. Ethical Implications

Residential proxies have emerged as one of the most powerful web data tools which are continually being used to gather large amounts of web intelligence for businesses. However, residential proxies have often been unethically used, as seen in 911 S5 Botnet [23], which featured 19 million compromised IP addresses turned into proxies used to carry out a variety of cyber scams and attacks. Furthermore, the unethical use of a residential proxy often traces back to its unethical sourcing. Therefore it is important that providers and consumers alike maintain ethical practices when it comes to sourcing, distributing and usage of residential proxies.

6.1. Ethical Sourcing

As per the current industry standard, ethical sourcing of a proxy refers to when the owner of a residential IP address consents for their IP Address's use as a residential proxy via a proxy software provided by a proxy service. This practice is usually implemented by an EULA (End User Licence Agreement) between the proxy service and the IP owner. [24] Proxy services have also tried mass implementations of such proxy software by the distribution of Proxy Software Development Kits (SDK) to existing apps with an unmonetized user base.[25] [26] These SDKs generally keep in regard the ethical sourcing policies. Unethically sourced proxies, however, have taken the SDK approach already, by embedding proxy software into Potentially Unwanted Programs (PuPs) while ignoring user consent whatsoever. These proxies are often leased and used to run invasive bots, which are referred to as botnets. Recently, governments around the world have put an emphasis on taking down botnets, as not only do they affect their target, but they also put millions of unsuspicious users at risk.[27] [28]

6.2. Distribution and Usage

The complex sourcing structure of ethical residential proxies inevitably calls for an elaborate distribution process that ensures the proxies are used for an ethical use case. If not, the unethical use of a consented proxy may lead to stern legal action against the owner. This makes it a responsibility for both a proxy service and their clients to vet each other and follow ethical practices. Hence, proxy services should adopt a distribution structure, which includes vetting potential clients and their use case's authenticity [29]; tailor-making proxies as per the user's needs can be achieved by blocking ports like SSH and SMTP when not required and by rate limiting requests to preserve IP scores. Furthermore, the users of proxy services shall ensure they use scientific scraping practices [30] to preserve the integrity of the proxy networks. Such practices are not limited to but include proxy rotators, robot file configurators, and fingerprinting.

7. Conclusion

This paper offers a comprehensive literature analysis of the vulnerabilities, privacy risks, and ethical implications associated with residential proxies, which is a powerful, fast growing tool with multiple use cases. The paper tests residential proxies from existing residential proxy providers for vulnerability to suggest necessary/mandatory and optional checks to both providers and users alike when dealing with residential proxies. It employs a standard literature review approach to privacy risks and ethical implications and finds the industry standard for practices like the distribution and usage of residential proxies. Moreover, compared to existing studies, which mainly discuss the different types and classifications of proxies, this review and analysis aim to provide supporting evidence for such classifications made by popular IP databases and have been found to hold regardless of proxy region.

Funding Statement

The author funded the research and the publication of the article.

References

[1] Zahra Nezhadian, "*Predicting Content Manipulations by Open Web Proxies*," Master Thesis, University of Saskatchewan, pp. 1-48, 2022. [Google Scholar] [Publisher Link]

[2] Jinchun Choi et al., "Understanding the Proxy Ecosystem: A Comparative Analysis of Residential and Open Proxies on the Internet," *IEEE Access*, vol. 8, pp. 111368-111380, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Nihal Abdurahiman, "Towards Residential Proxies Detection: An Experimental Analysis in the Android Environment," Thesis, Hamad Bin Khalifa University, pp. 1-6, 2021. [Google Scholar] [Publisher Link]

[4] Residential Proxies, Bright Data. [Online]. Available: https://brightdata.com/blog/guest-post/what-is-a-residential-proxy

[5] DDoS Attacks Against Hungarian Media Traced to Proxy Infrastructure "White Proxies", Qurium, The Media Foundation, 2023. [Online]. Available: https://www.qurium.org/weaponizing-proxy-and-vpn-providers/ddos-attacks-traced-to-proxy-infrastructure-white-proxies/

[6] Xianghang Mi et al., "Your Phone is My Proxy: Detecting and Understanding Mobile Proxy Networks," *Proceedings of ISOC Network and Distributed System Security Symposium*, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sergey Frolov, Jack Wampler, and Eric Wustrow, "Detecting Probe-Resistant Proxies," *Network and Distributed Systems Security (NDSS) Symposium 2020*, San Diego, CA, USA, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] J.M. Hoogstraaten, "*Evaluating Server-Side Internet Proxy Detection Methods*," Master Thesis, pp. 1-80, 2018. [Google Scholar]

[9] Whois Search, APNIC. [Online]. Available: https://www.apnic.net/about-apnic/whois_search/

[10] Using Whois, American Registry for Internet Numbers. [Online]. Available: https://www.arin.net/resources/registry/whois/

[11] Ripe Network Coordination Center, RIPEstat. [Online]. Available: https://stat.ripe.net/widget/whois

[12] Whois, Lacnic. [Online]. Available: https://www.lacnic.net/1040/2/lacnic/whois

[13] AFRINIC The Internet Numbers Registry for Africa. [Online]. Available: https://afrinic.net/

[14] IP Geolocation, IP2Location. [Online]. Available: https://www.ip2location.com/

[15] Create Smarter, Safer Digital Experiences with Accurate Data, MaxMind. [Online]. Available: https://www.maxmind.com/en/home

[16] IP Geolocation API and Database, DB-IP. [Online]. Available: https://db-ip.com/

[17] IP Reputation Check, IP Address Reputation Lookup & API, IPQS. [Online]. Available: https://www.ipqualityscore.com/ip-reputation-check

[18] IP Address Reputation, Spamhaus Project. [Online]. Available: https://www.spamhaus.org/ip-reputation/

[19] What is Honeygain, Honeygain. [Online]. Available: https://www.honeygain.com/what-is-honeygain/

[20] Earn Passive Income while your Devices Rest, EarnApp. [Online]. Available: https://earnapp.com/bandwidth

[21] What is Ports in Networking?, GeeksforGeeks, 2023. [Online]. Available: https://www.geeksforgeeks.org/what-is-ports-in-networking/

[22] Edvinas Račkauskas, IP Legacy Space Explained, IPXO, 2021. [Online]. Available: https://www.ipxo.com/blog/ip-legacy-space/

[23] 911 S5 Botnet Dismantled and Its Administrator Arrested in Coordinated International Operation, Office of Public Affairs, 2024. [Online]. Available: https://www.justice.gov/opa/pr/911-s5-botnet-dismantled-and-its-administrator-arrested-coordinated-international-operation

[24] End User License Agreement, EarnApp, 2024. [Online]. Available: https://earnapp.com/user-agreement

[25] Extra $200 CPM from One Click!, Bright SDK. [Online]. Available: https://bright-sdk.com/

[26] Earn an Extra $500 CPM from Your App!, Honeygain. [Online]. Available: https://www.honeygain.com/sdk/

[27] Joseph Demarest, Taking Down Botnets, Federal Bureau of Investigation, 2014. [Online]. Available: https://www.fbi.gov/news/testimony/taking-down-botnets

[28] Largest Ever Operation Against Botnets Hits Dropper Malware Ecosystem, Europol. [Online]. Available: https://www.europol.europa.eu/media-press/newsroom/news/largest-ever-operation-against-botnets-hits-dropper-malware-ecosystem

[29] Nadav Roiter, How Bright Data's KYC-First Approach has Helped Pioneer one of the Safest, Legally Compliant, and Ethical Data Collection Networks, Bright Data. [Online]. Available: https://brightdata.com/blog/why-brightdata/bright-datas-kyc-helped-pioneer-data-collection-networks

[30] Gulbahar Karatas, 7 Web Scraping Best Practices You Must Be Aware of in '23, AIMultiple Research, 2024. [Online]. Available: https://research.aimultiple.com/web-scraping-best-practices/