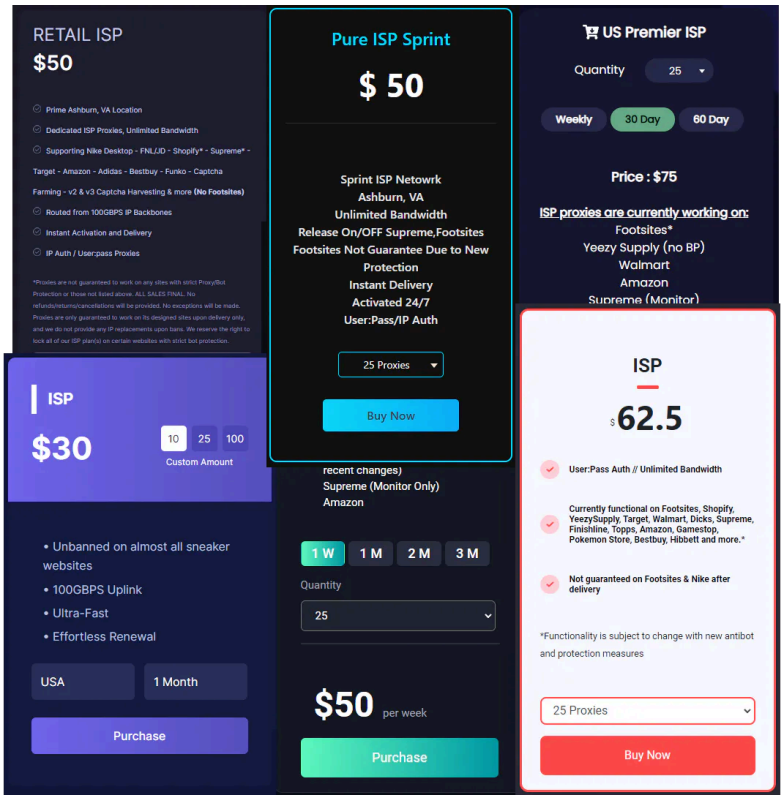
Detecting Residential Proxy Networks

11/23/2021

In recent years we have observed the rise of so called 'residential' proxy networks. These are proxy networks custom tailored to appear as a residential connection, while being hosted inside of a data center on a server.



Such a service generally charges around **$2.00 - $3.00** for access to a single residential proxy IP address for a term of one month, and there are at least dozens of services offering such proxies.

Why?

The motiviation behind such services is primarily to support the 'scalping' market. Scalping is the act of purchasing many limited stock, in demand products, most commonly items such as shoes, GPU's, gaming consoles and anything else that might

have a low supply with high demand, in order to resell it for a profit. Indeed we can see these services openly advertising their proxies are capable of succesfully circumventing anti-scalper checks on the most popular websites.

These proxy networks directly hurt your average, every day consumer, while bringing in a sizeable income to the operators, who resell the same products the consumer originally wanted to buy at an increased price.

Detection

Websites targeted by scalping have already taken comprehensive measures to stop such activity. The most common and effective check to filter out malicious IP's is by inspecting the source network they originate from on the internet, referred to as ASN (autonomous system number). While many hosting services exist that would offer a great network to host proxies, they are easily identifiable, and are not effective at accessing websites scalpers are interested in.

Scalpers are one step ahead. They find hosting providers willing to announce their proxy IP's directly on common residential networks to appear legitimate, or even contract with residential networks directly and colocate server infrastructure in a data center. The most common networks used for this purpose are:

- Sprint AS1239
- Lumen / Level3 AS3356
- AT&T AS7018
- Windstream Communications AS7029
- Comcast AS7922
- Cox Communications AS22773

As you can imagine, differentiating legitimate and proxy traffic becomes a lot more cumbersome when they are both coming from the same residential provider. Generally at this point, the main form of detection employed to combat these proxies is analyzing historical patterns. A proxy network will have repeated history of accessing the website and initiating a high volume of purchase activity that is outside the realms of a typical user.

The downside to this approach is any given proxy network is practically guaranteed months of unimpeded access while a fingerprint is built. In this timeframe it will purchase thousands of products and the damage will already be done, by the time it is detected and banned from future access the proxy operators will have already cycled a new block of fresh IP space, and the cycle repeats indefinitely. If websites had a method of detecting these proxies before they ever made a request, such proxy networks would become obsolete overnight, toppling a multi-million dollar industry while simultaneously giving legitimate users access to the in demand products they want.

You can run, but you can't hide 🕵️

Proxy operators ran to residential networks after being blocked from hosting networks and had great success in doing so. However in the process they opened themselves up to a novel fingerprinting technique that is impossible to mask thanks to the way the internet operates via the border gateway protocol (BGP).

Remember how earlier I mentioned any given proxy subnet becomes detectable after a few months due to repeated use? This forces proxy operators to be constantly churning IP space, in fact seeing a proxy IP remain active for a period of over 6 months is extremely unlikely. This weakness can be exploited for detection. By analyzing historical routing history through a service such as RIPEstat it is possible to detect freshly announced IP space and block it's access.

I looked at all current IPv4 prefixes announced by Sprint and it didn't take long to find a suspicious IP block. For this example I chose 23.247.244.0/22 because it was labeled "IPXO LLC", more on that later.

Gotcha!

This IP space was announced on the Sprint network on October 28th, less than a month from the time of this publication. A network such as Sprint is very unlikely to announce an IP block as small as a /22 (1,024 IPs) and assign it to end customers. Huge tier 1 networks like Sprint have millions of free IPv4 address space aquired in the early days of the internet, they have no need to aquire IP space that doesn't directly belong to them, or re allocate smaller sized blocks.

However so far this is just speculation. While exceedingly unlikely Sprint decided to lease a /22 subnet and announce it in the past month for any type of legitimate usage on the Sprint network (cell towers, home connections, etc), it isn't impossible. Luckily, there is a definitive way to verify that this network is being used for proxies.

Nmap to the rescue

If this subnet is in fact running a proxy network, they would need to open a port on any given IP address to accept incoming proxy connections. So let's port scan a random IP in this subnet and see if any suspicious ports are open...

Interesting, we can observe a randomly selected IP in this subnet has no ports open except 63981. It is common for proxy IP's to not have any other ports open such as SSH or HTTP, since they only exist to serve a single purpose, proxying traffic. So, let's give it a try, let's send an HTTP request to this IP on port 63981 and see what we get back.

Say Cheese! 📷

We have confirmed this subnet is being used to operate a proxy network.

We can observe the returned headers to see the operator is running squid proxy software. Squid is an easy to setup forward proxy software that is most commonly associated with these networks.

Squid configuration often calls for the operator to specify a server hostname that is returned in the headers of a request, in this case the operator made it easy for us and labeled his server as **ISP_PROXIES**. ISP Proxies are a common industry term that refer to a proxy network operating under residential ISP networks.

Other fingerprints 🐾

While short lived, freshly announced subnets are the most prominent indicator of a proxy network, it is not a bulletproof method. As you can imagine it is entirely possible to mistake a legitimate network as malicious, and fingerprinting the bad guys doesn't help much if the good guys are mixed in!

We can supplement subnet history with some other factors to make it nearly impossible for a legitimate residential network to slip through. The second most prominent indicator of proxy networks is the source of the subnet. We can safely assume massive residential networks have plenty of IP space and won't be needing to lease subnets from any 3rd parties. Proxy operators need to source their IP space from somewhere, and residential networks aren't keen to give out large swaths of short lived address allocations without proper justification (which proxy operators don't have). This narrows down their choices drastically, and a huge amount of proxy networks source IP space from the same few IP brokers:

- Heficed
- IPXO
- LogicWeb
- CloudInnovation

By running a WHOIS check against a suspicious IP and checking for common IP broker values, we can determine beyond a reasonable doubt if there is a proxy on the other end of that connection.

We can also factor in the size of the subnet the IP currently belongs to. Proxy operators will inherently lease smaller sized subnets, while legitimate residential connections will commonly originate in massive parent subnets, ranging from /15 all the way to /8.

We can safely assume no proxy network has the budget or capability to lease a subnet of greater size than /16 (/15 is 131,072 IP's). For this reason we can outright disregard any connections coming from subnets of this length or larger, while further inspecting smaller sized subnets.

The last potential fingerprint is going back to the basics, checking ASN's. While proxy operators have moved off commonly associated hosting ASN's to avoid detection, the IP's they lease have not. Much of the leased IP address space on the internet today has some history of being announced on a hosting network at one point or another. Residential networks do not.

We can employ the same historical BGP lookup we used earlier to check if the IP has ANY history of being used on a common hosting network. We can safely assume any legitimate home connections have no such history and thus block any networks that have any history at all of being used on a hosting ASN.

Does it work?

I have developed a Proxy Detector Tool that utilizes all the techniques discussed above to identify if an IP belongs to a proxy or not. Let's test it out with the known proxy we discovered earlier:

Success! We have succesfully fingerprinted a proxy connection, but what about legitimate home connections, will my personal home IP address register as a proxy?

Correct again! In fact, I tested over 100 known home IP's, spanning many countries around the world, belonging to different autonomous networks, and residing in subnets of all variety of sizes (even /24), not a single false positive was registered.

What about IPv6?

IPv6 is commonly restricted on a large portion of websites proxy access is required to, simply because it is harder to fingerprint for the website operators. However IPv6 doesn't have any inherent differences in fingerprinting than IPv4, we just have to recalculate subnet lengths for more thorough checks.

How big is the problem?

Massive.

I set out to analyze the entire Sprint network to see just how infested it has become. The results are staggering.

You can download the raw data here

- At the time of my test, Sprint was announcing **1816** seperate subnets, for a total of **19,197,696** unique IPv4 addresses.
- Of these, **1275** subnets had an active proxy network, totalling **1,291,008** unique IPv4 addresses.

- **6.72%** of all IP's announced on the Sprint network are being used to proxy traffic.
- **70%** of all subnets announced on the Sprint network are being used to proxy traffic!
- If we calculate how much revenue is being generated by these proxies at the low end of $2.00 / IP, we can estimate a monthly revenue of **$2,582,016** going to proxy networks on the Sprint network alone!

These results were aquired with small scale, non intrusive scanning techniques, only testing IPv4, and only testing a single residential network. From this data we can infer the true scale of these networks is massive, spanning millions of unique IP's and thousands of subnets.

I was able to grab metadata to further analyze the distribution of these networks.

Server label distribution

Proxy software often calls for the operator to configure a hostname / label to be associated with the server that is sent in headers, this chart demonstrates how many

times each label was returned from a given subnet and can correlate a single operator between multiple subnets.

Proxy Server Distribution

Based on the type of response a server generates, we can infer the software being used to deploy the proxy network. An overwhelming amount of these proxy networks use squid. Squid is free, open source, and takes little technical skill to setup. What's interesting about it's massive adoption is the drawbacks it imposes, squid is not intended to be used as a forward proxy for a high volume of IP's. Updating configurations would be cumbersome and not easily scalable, this can demonstrate a lack of technical capability on behalf of the network operators who choose to not deploy a custom solution.

Subnet Sizes

As we can expect, the vast majority of these networks operate under very small sized subnets. What is interesting are the 2 /17 subnets, and even a single /16. This indicates a single entity is responsible for a massive amount of IP's.

Conclusion

Residential networks have become infested with proxy networks which are directly driving the scalping industry. If e-commerce websites implement the solutions outlined above, they can make a substantial impact in the battle against bots, scalpers, and automated checkout software as a whole.

Now that a light has been shined on this dark corner of the internet, it will be interesting to see what action might be taken by the networks harboring these proxies, what measures might be implemented by websites to combat them, and where the proxy operators might flock to next in this ever changing game of cat and mouse.

Written by Rasbora

rasbora.dev
rasbora@rasbora.dev
11/23/2021