An Internet-wide view of large-scale IP Sharing

Vasileios Giotsas, Marwan Fayed, Cefan Rubin
Nick Wood, Antoine Cordelle
Cloudflare

IP-based traffic filtering is a very popular security technique

- **Blocklists** filter requests against known threat actors
- **Rate-limiting** to prevent bots and abuse of services
- **Anomaly detection** based on behavioural fingerprinting of IP activity

IP-based traffic filtering is a very popular security technique

- **Blocklists** filter requests against known threat actors
- **Rate-limiting** to prevent bots and abuse of services
- **Anomaly detection** based on behavioural fingerprinting of IP activity

What about collateral damage?

Large-scale IP sharing (LSS) makes IP-based blocking problematic

- **Carrier-Grade NAT (CGNAT)**:
  - Used by ISPs to manage IP address scarcity
  - IPv4 scarcity leads to higher concentration of users per IP

- **VPN / Proxies:**
  - Chosen by users for privacy, security, or performance

Large-scale IP sharing (LSS) makes IP-based blocking problematic

- **Carrier-Grade NAT (CGNAT)**:
  - Used by ISPs to manage IP address scarcity
  - IPv4 scarcity leads to higher concentration of users per IP

  Users don't use CGNAT by choice, shouldn't be punished for it

- **VPN / Proxies:**
  - Chosen by users for privacy, security, or performance

  Often abused for malicious purposes

"Confirm you are not a Robot" -- CGNAT IP blacklisted? Issues/Problems

(self.tmobileisp)
submitted 1 year ago by LordFlux

For the past couple of weeks, I've been getting a lot of "Confirm you are not a Robot" pages where I have to click a check box or fulfill a Captcha requirement. This is when I do anything -- even a simple Google search.

CGNAT and Google reCAPTCHA haunting me throughout internet

Philips · Dec 13, 2020 · 14 · 11,63

Feb 12, 2021 #3

This is probably due to CGNAT and Captcha not being terribly good at handling IP address sharing on networks.

Google flags Starlink CGNAT IP's

Google services become very hard to use on S
Sometimes pages will die multiple times and c
page load. Where can I tell Google to stop flag
unflagged?

Frequent hCaptchas on certain sites behind Starlink CGNAT

Website, Application, Performance Security

Techjar 1 Feb 2022

Simple as the title. I'm on Starlink, which uses CGNAT, and certain sites with higher security settings are hitting me with hCaptcha pages sometimes multiple times a day. This did not happen with my previous ISP. Presumably this is because of the high frequency and variability of traffic from these IP addresses, and the fact that Starlink is fairly new. Is there anything that be done to reduce the frequency of or eliminate these erroneous captchas?

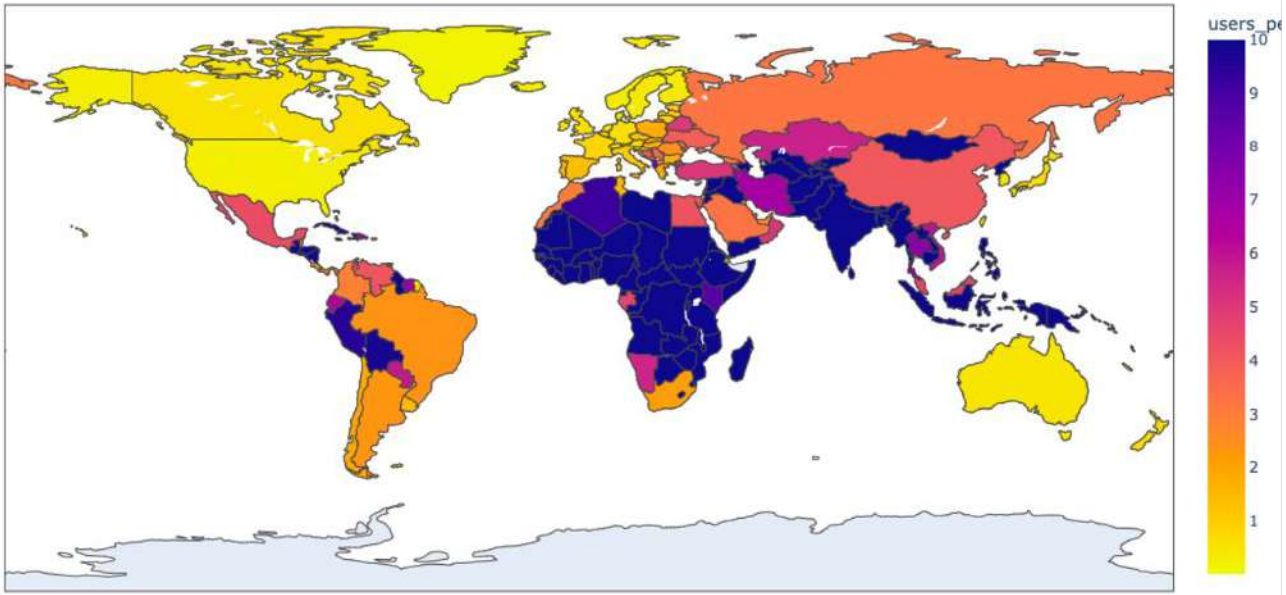
10.2.5 Impact on Anti-Spam Measures

As noted in the technical analysis, there have been reports of anti-spam/anti-abuse measures impacting email clients behind CGN, as a result of mail servers detecting too many sessions from a single IPv4 address.[123]

In the event that an IPv4 address is blocked or blacklisted as a source of spam, the impact on a CGN would be greater, potentially affecting an entire subscriber base. This would increase cost and support load for the ISP, and, as we have seen earlier, damage its IP reputation.

https://oxil.uk/.k-media/b87a9baa644d1c3afd6e3dd458b481f7.pdf

Actions on IP addresses can have disproportionate effects along socio-economic boundaries

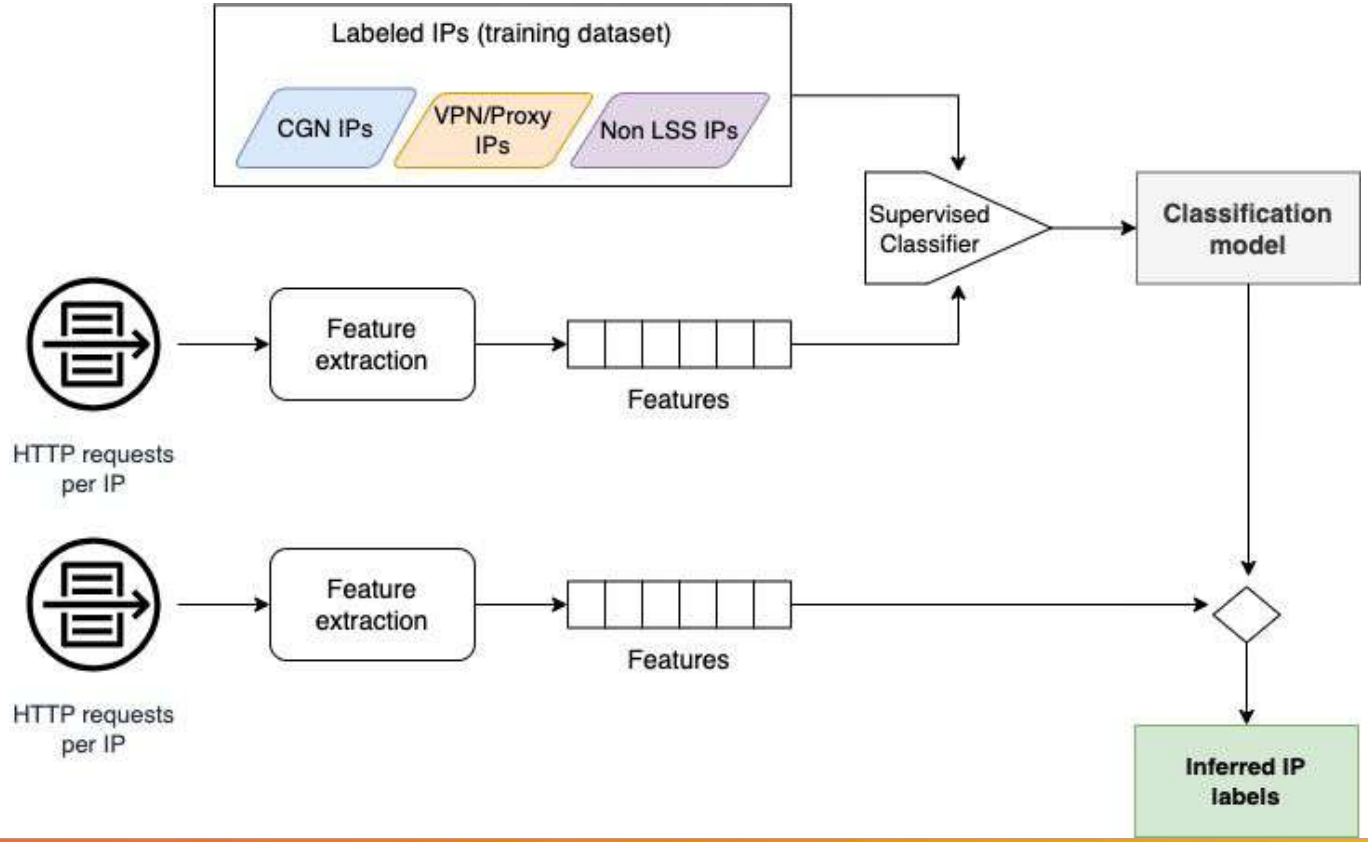


Internet users normalized by the number of IPs registered in the country

Our goal:

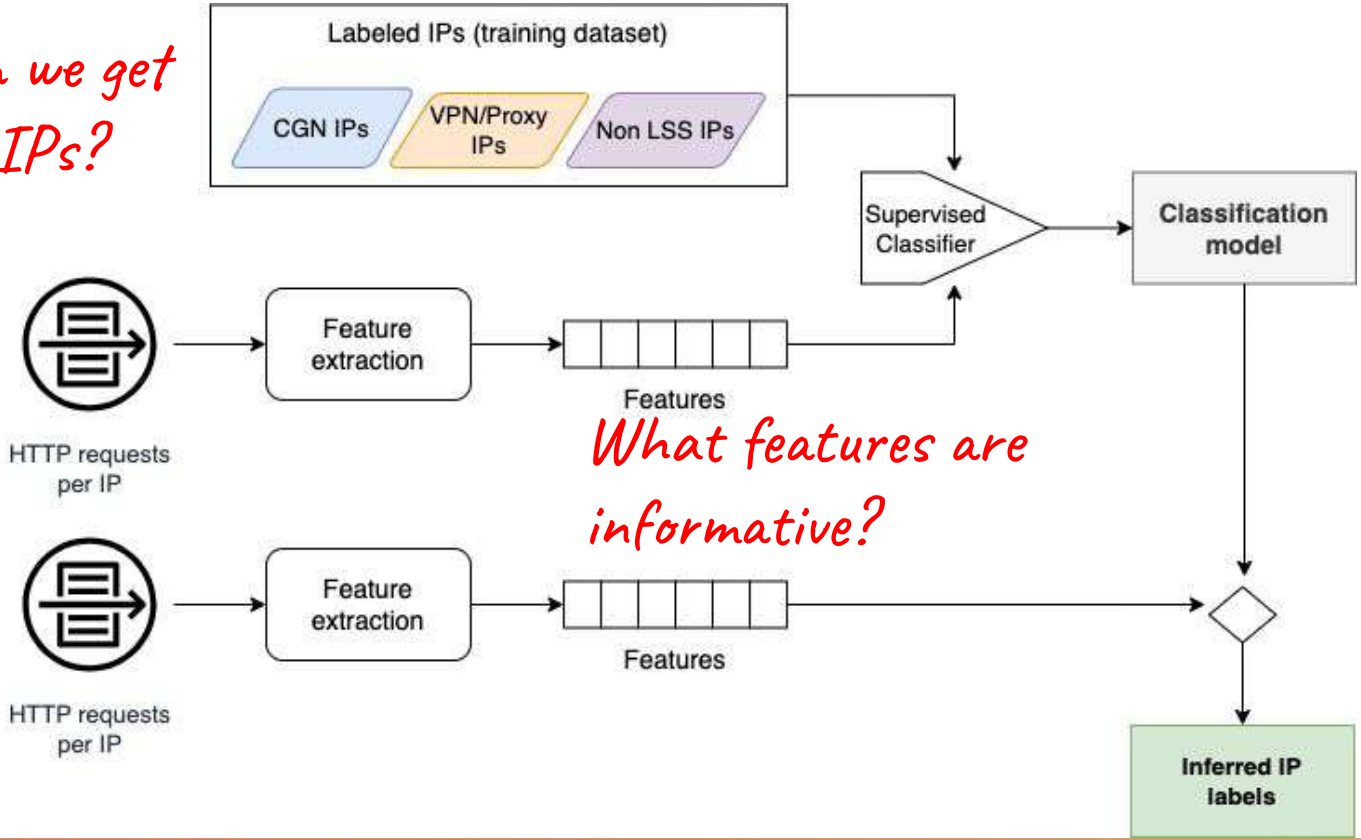
Detect large-scale IP sharing to calibrate traffic filtering and minimize collateral damage

Overview of inference methodology



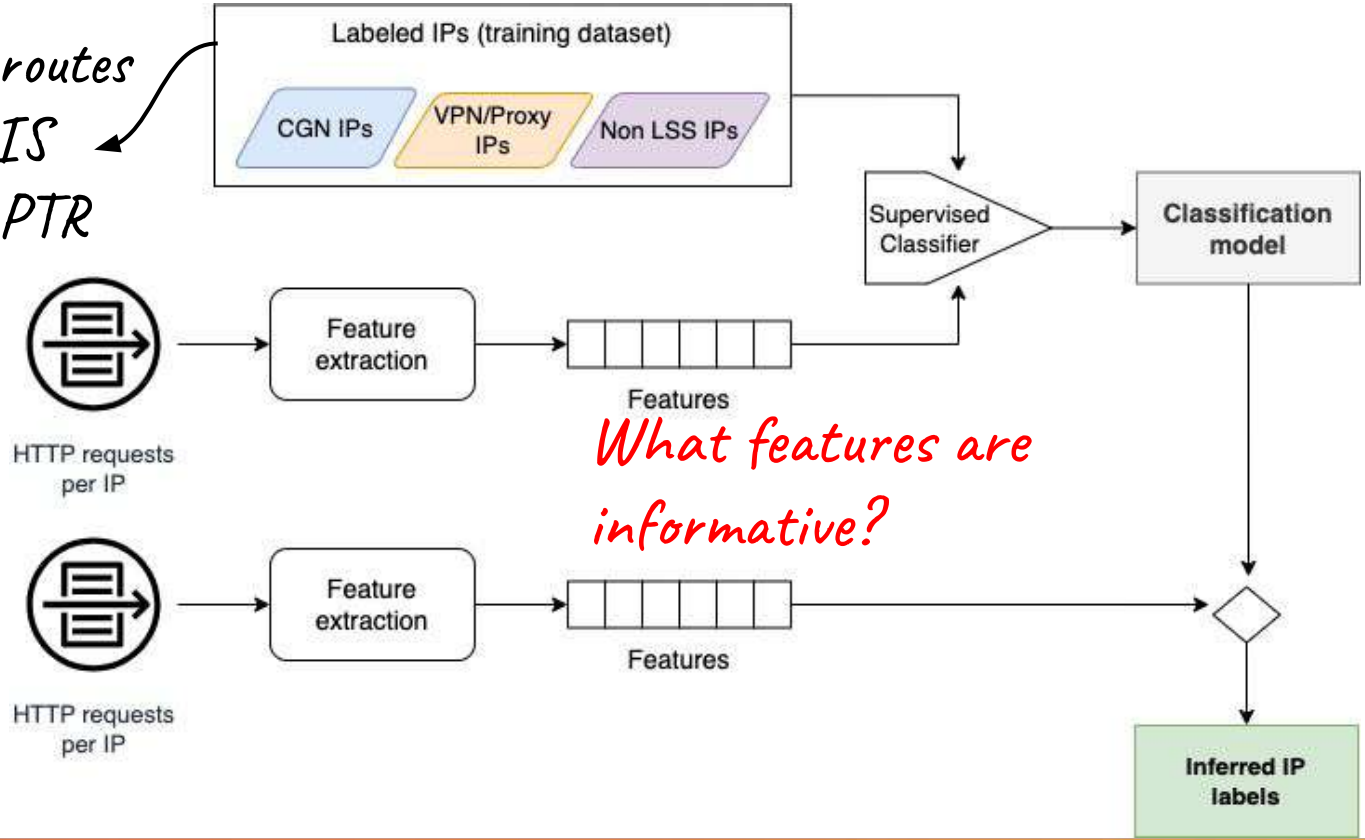
Overview of inference methodology

How can we get labelled IPs?



Overview of inference methodology

1. Traceroutes
2. WHOIS
3. DNS PTR



Constructing a training dataset of labeled IPs

Distributed traceroutes from RIPE Atlas

```
traceroute to 1.1.1.1 (1.1.1.1), 64 hops max, 40 byte packets
 1  my.meraki.net (192.168.128.1)  4.690 ms  1.626 ms  1.673 ms
 2  mvx-177-92-65-193.mundivox.com (177.92.65.193)  2.250 ms  2.132 ms
 3.293 ms
 3  10.11.106.254 (10.11.106.254)  4.101 ms  4.558 ms  4.256 ms
 4  100.64.12.94 (100.64.12.94)  4.901 ms  3.768 ms  6.577 ms
 5  100.64.12.178 (100.64.12.178)  5.615 ms  4.198 ms  5.552 ms
 6  100.67.36.233 (100.67.36.233)  3.963 ms  5.325 ms  4.371 ms
 7  64.191.232.248 (64.191.232.248)  17.432 ms  20.750 ms  14.755 ms
 8  172.68.16.89 (172.68.16.89)  6.857 ms
    172.68.16.99 (172.68.16.99)  6.376 ms
    172.68.16.107 (172.68.16.107)  5.176 ms
 9  one.one.one.one (1.1.1.1)  3.270 ms  3.551 ms  3.650 ms
```

Constructing a training dataset of labeled IPs

WHOIS data

```
inetnum:        154.72.13.0 - 154.72.13.255
netname:        ORG-USS1-AFRINIC
descr:          THIS REOURCE IS USED TO CGNAT OUR MOBILE
 SUBSCRIBERS TO GO TO INTERNET
country:        ST
admin-c:        JT26-AFRINIC
tech-c:         JT26-AFRINIC
status:         ASSIGNED PA
mnt-by:         UNITEL-STP-MNT
source:         AFRINIC # Filtered
```

Constructing a training dataset of labeled IPs

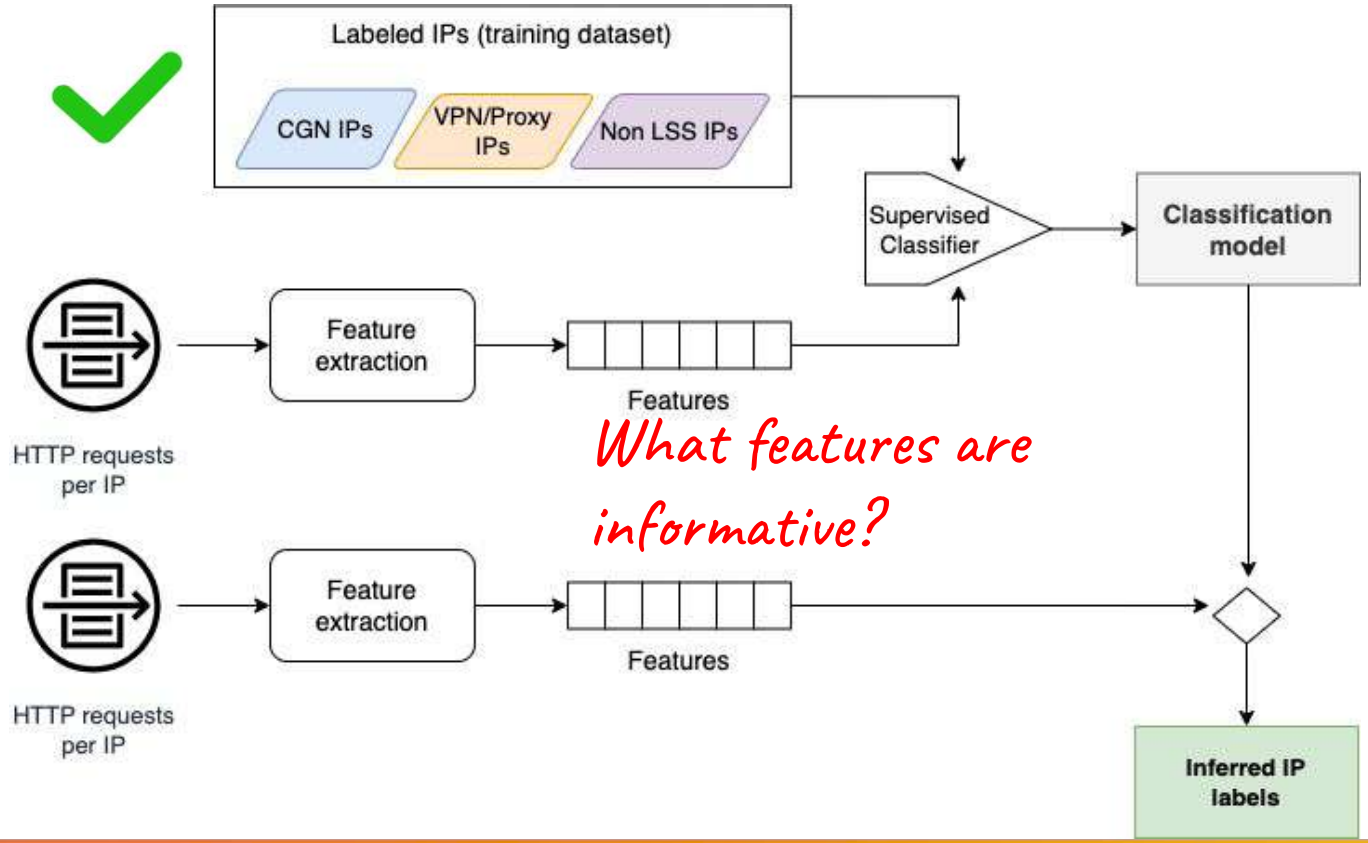
DNS PTR records

IP	PTR
23.134.17.0	23-134-17-0.cgnat-ipv4.missouricom.com
23.134.17.1	23-134-17-1.cgnat-ipv4.missouricom.com
23.134.17.2	23-134-17-2.cgnat-ipv4.missouricom.com
23.134.17.3	23-134-17-3.cgnat-ipv4.missouricom.com
23.134.17.4	23-134-17-4.cgnat-ipv4.missouricom.com
23.134.17.5	23-134-17-5.cgnat-ipv4.missouricom.com
23.134.17.6	23-134-17-6.cgnat-ipv4.missouricom.com

Composition of our training dataset per label

Reference Dataset	Addresses Found	ASes
CGNAT IPs	215,770	1,496
VPNs & Proxies	179,448	306
Non LSS IPs	878,560	2,602

Overview of inference methodology

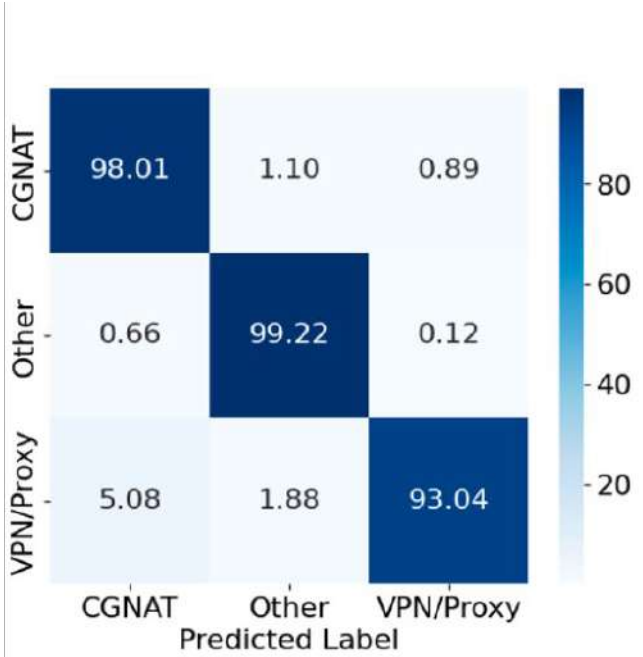


Key discriminating features per IP and per /24 prefix

- Diversity of User-Agents
- Diversity of TLS signatures
- Source port distribution
- TCP RTT variability
- Diversity of destination hosts
- TLS/TCP RTT difference
- Number of requests

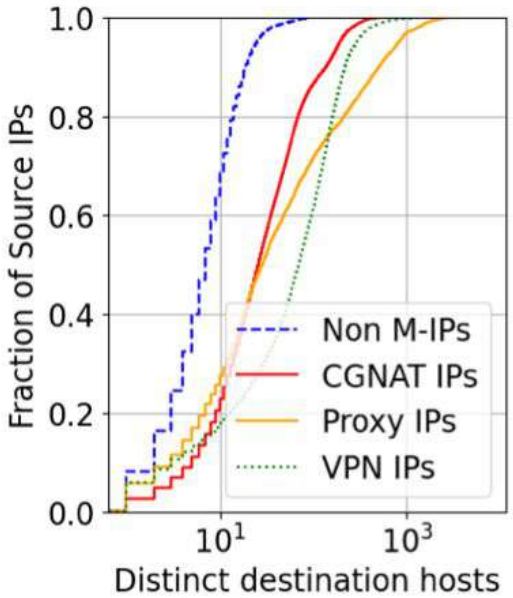
Multi-Class Classification Model

- XGBoost classifier with 97% F1-score
- 98% accuracy on test set
- 10-fold cross-validation with 0.994 AUC
- Independent validation with 96% accuracy on SOCKS proxy dataset and dataset from mobile broadband provider

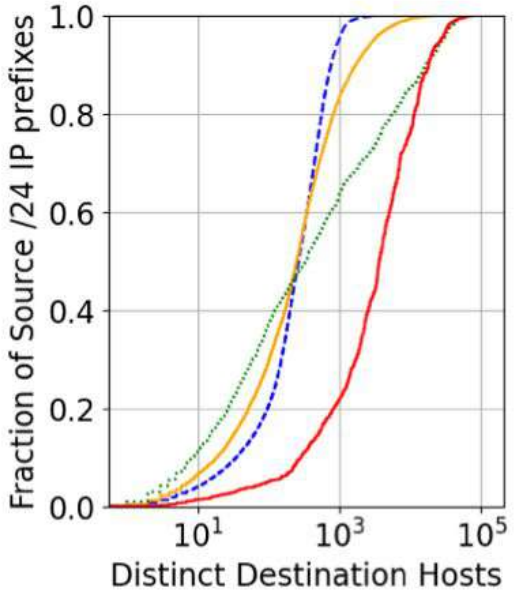


Confusion matrix of IP type inferences.

Comparing the HTTP features per IP and per /24 prefix helps to distinguish VPN/Proxies from CGNs



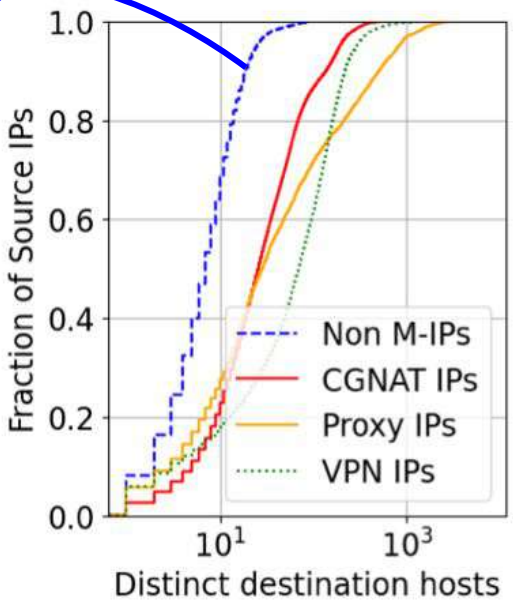
per IP address



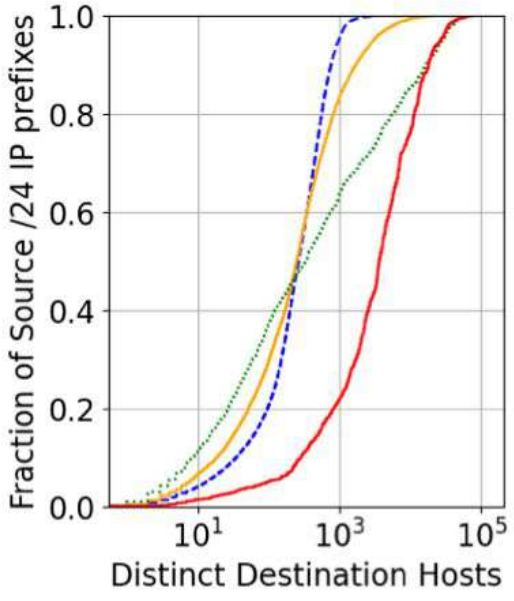
per /24 prefix

Comparing the HTTP features per IP and per /24 prefix helps to distinguish VPN/Proxies from CGNs

Easy to distinguish Non-Shared IPs from the rest



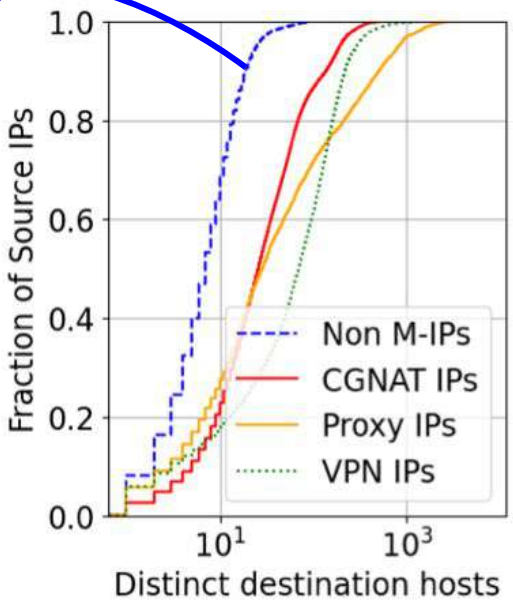
per IP address



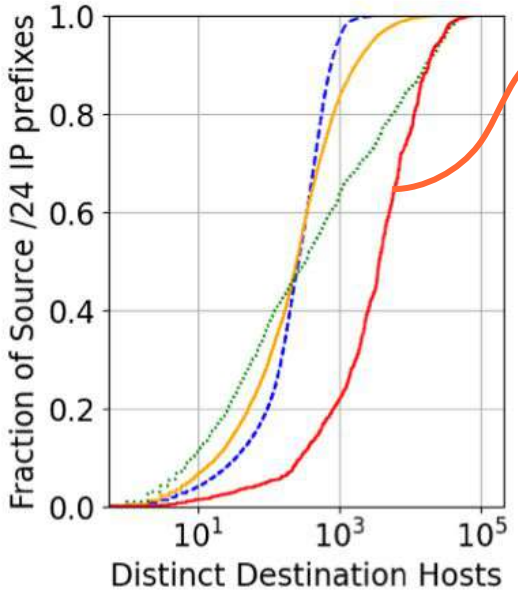
per /24 prefix

Comparing the HTTP features per IP and per /24 prefix helps to distinguish VPN/Proxies from CGNs

Easy to distinguis Non-Shared IPs from the rest



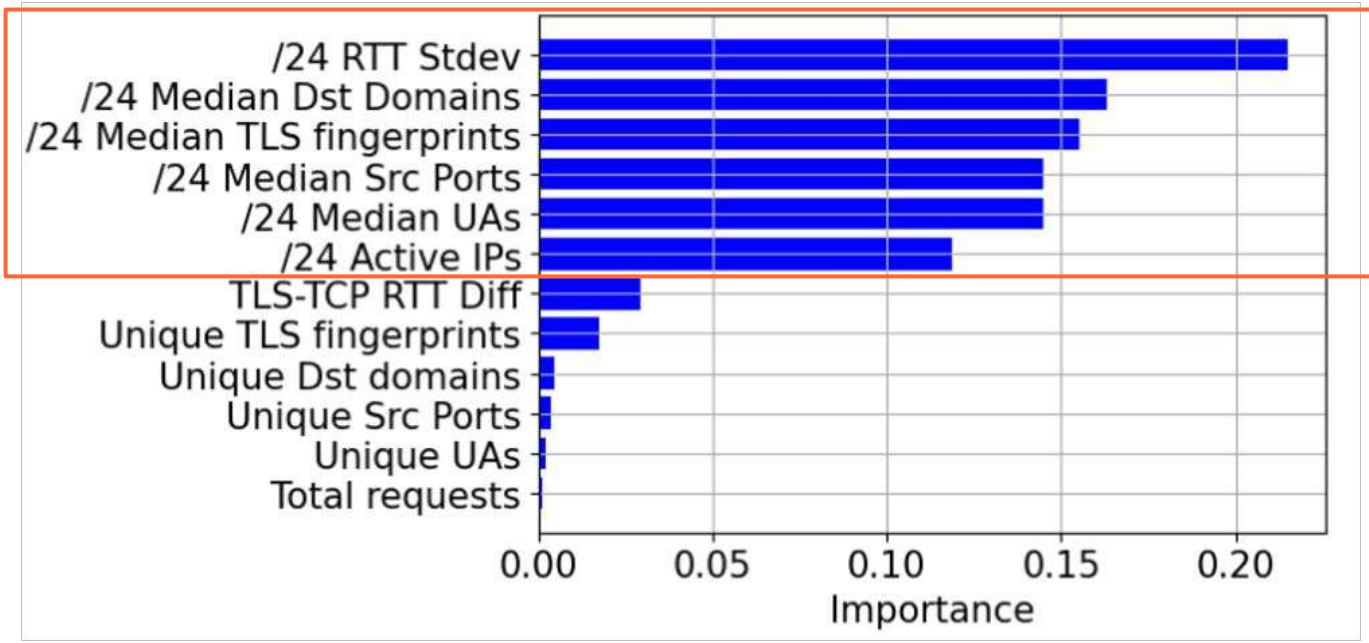
per IP address



per /24 prefix

Easy to distinguis CGNAT IPs from the rest

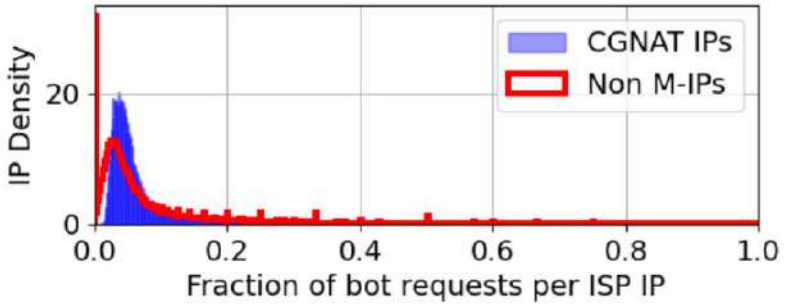
/24 Prefix Features Dominate Classification



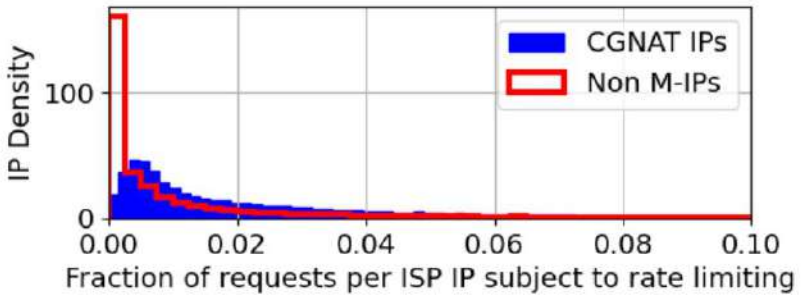
Operational Implications

- CGNAT IPs generate proportionally **16X more requests** than non-shared-IPs

- CGNAT IPs are **3X more likely to be rate-limited** despite similar bot scores



(a) Bot scores are comparable at the median; non-CGNs have a longer tail.

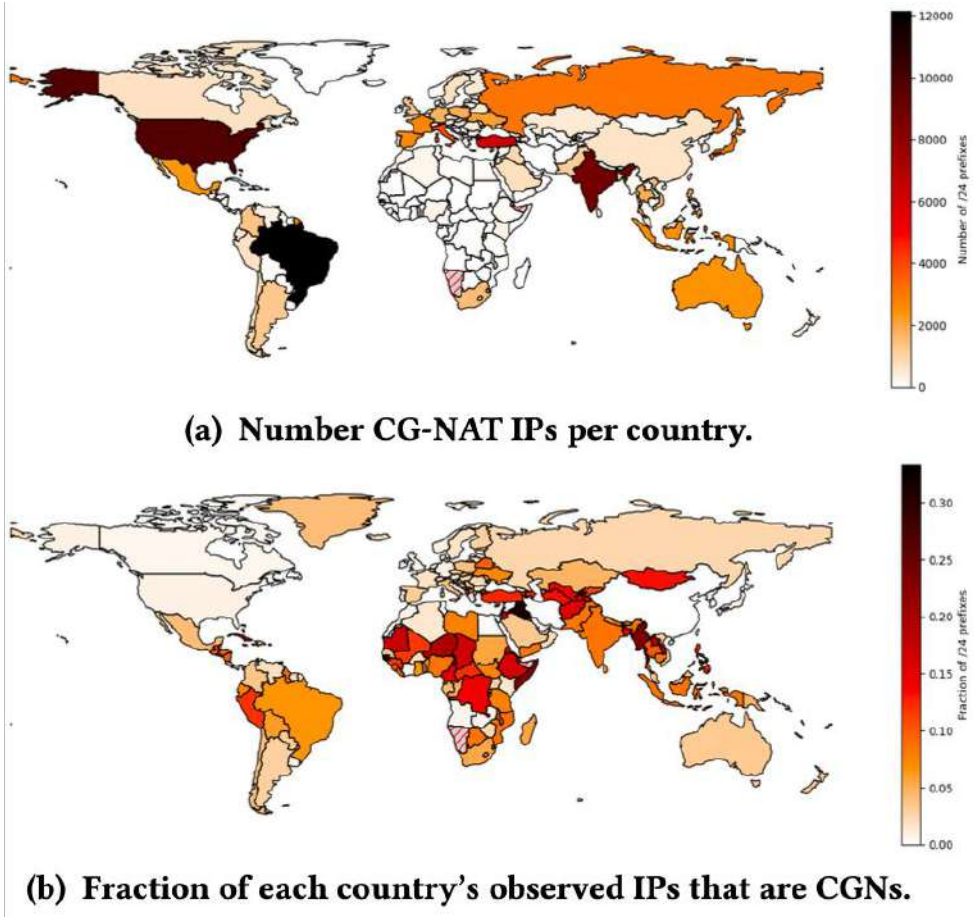


(b) Fractions of requests per CGN IPs are three times more likely to be rate-limited by customers of the CDN.

Global Distribution of CGNATs

- Highest raw numbers:
  Brazil, India, US

- Highest proportion of country's IPs:
  Africa, Central and South-East Asia



Conclusions

- Proper detection and classification of multi-user IPs can prevent collateral damage from IP filtering
- Important to detect CGNATs for equitable security measures
- Diversity, not just volume, is key for identification

vasilis@cloudflare.com


@GVasilis

giotsas.com