



---

# How To Source Proxies For Data Scraping

---

**Neil Emeigh**

**Extract Summit 2022**

**[www.rayobyte.com](http://www.rayobyte.com)**



## Who this is for:

1. Scraping companies
2. Proxy companies
3. Techies that like true crime stories

# What you will learn

---



How we built a proxy company



How to avoid bans (from a proxy perspective)



Legal issues in today's proxy industry



Actionable takeaways



Competition: a rising  
tide lifts all boats

---

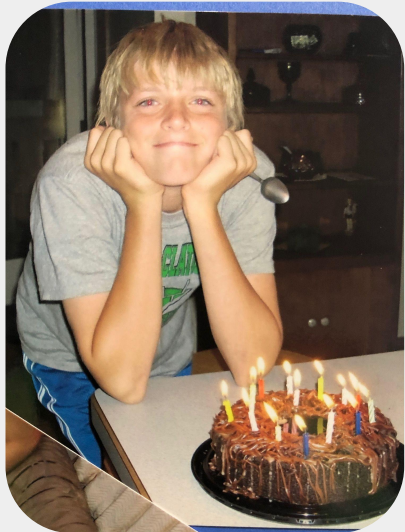


# Building a Proxy Company



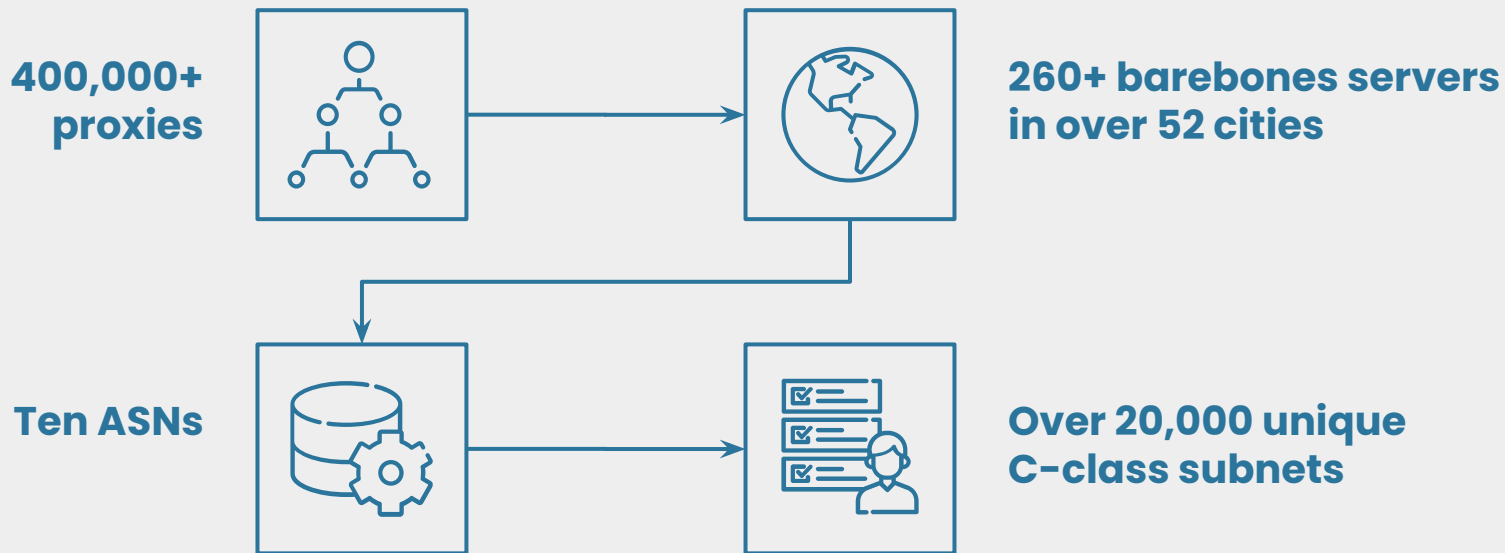
# My Story

---



# Our infrastructure: by the numbers

---



# Buying IPs

Vs

# Leasing IPs

+ We can select the best, most authoritative IPs	+ Cheaper
+ We can put IPs on a clean ASN	+ Easier to manage
+ One-time payment	+ Setup is quicker
- Harder to set up and manage	- Less control
- More expensive	- Prices Fluctuate
- Harder to liquidate	



# Diversity is key

Network Resources	A.B.C	A.B.C
EIL-79 (NET-104-244-100-0-1)	104.244.100.0	0 - 104.244.103.255
EIL-79 (NET-130-254-96-0-1)	130.254.96.0	0 - 130.254.127.255
EIL-79 (NET-130-44-200-0-1)	130.44.200.0	0 - 130.44.203.255
NET-138-229-96-0-20 (NET-138-229-96-0-1)	138.229.96.0	0 - 138.229.111.255
NET-139-180-0-0-21 (NET-139-180-0-0-1)	139.180.0.0	0 - 139.180.7.255
NET-139-180-224-0-21 (NET-139-180-224-0-1)	139.180.224.0	0 - 139.180.231.255
NET-139-60-101-0-24 (NET-139-60-101-0-1)	139.60.101.0	0 - 139.60.101.255
EIL-79 (NET-141-164-80-0-1)	141.164.80.0	0 - 141.164.95.255



# How To Avoid Bans



**Control**

**VS**

**Diversity**

# Example: using 256 IP addresses to scrape Google SERP



# Example: using 256 IP addresses to scrape Google SERP



Google will  
ban your  
subnet  
after 200  
requests  
per hour!

## The Control Option: Buy One Class C Subnet

11.111.222.0

11.111.222.1

11.111.222.2

11.111.222.3

11.111.222.4

11.111.222.5

11.111.222.6

11.111.222.7

11.111.222.8

...

11.111.222.255

**Result: Banned by Google  
after 200 requests in an  
hour!**



# The Diversity Option: Buy 1 IP From 256 Subnets

12.111.201.1

11.101.222.1

21.11.212.1

11.100.2.1

22.121.201.1

11.211.222.1

12.102.111.1

11.1.211.1

21.22.110.1

...

22.121.211.1

**Theoretical result:  
Banned by Google after  
51,200\* requests in an  
hour!**

\*256 times 200

# The Diversity Option: Buy 1 IP From 256 Subnets



# The “Noisy Neighbor” Problem

12.111.201.1      **12.111.201.2**

11.101.222.1

21.11.212.1

11.100.2.1

22.121.201.1

11.211.222.1

12.102.111.1

11.1.211.1

21.22.110.1

...

22.121.211.1

**Result: Banned by Google  
before you even get to  
make a request!**

# ASNs – Autonomous System Numbers

Active Aut-Num

Add New

Search...



AS64267

MAINT-AS64267

Edit Delete

AS54252

MAINT-AS64267

Edit Delete

AS397630

MAINT-AS64267

Edit Delete

Related Aut-Num

Home

AS397630

MAINT-AS64267

AS54252

MAINT-AS64267

AS64267

MAINT-AS64267

# ASNs – Autonomous System Numbers

Active Aut-Num		<a href="#">Add New</a>	
		<input type="text" value="Search..."/>	
AS64267		MAINT-AS64267	<a href="#">Edit</a> <a href="#">Delete</a>
AS54252		MAINT-AS64267	<a href="#">Edit</a> <a href="#">Delete</a>
AS397630		MAINT-AS64267	<a href="#">Edit</a> <a href="#">Delete</a>

Related Aut-Num		<a href="#">Home</a>	
AS397630		MAINT-AS64267	
AS54252		MAINT-AS64267	
AS64267		MAINT-AS64267	

Diversity factors: Quantity + Quality

# ASN Quality

---



AFRINIC Ban Speed:  
Already Gone!



Average data center ban  
speed: 200 requests per  
hour



Verizon/Comcast Ban  
Speed: 2000 requests per  
hour





# Legal Adventures in the Proxy Industry



# The industry at large

---

# AFRINIC – Corruption and Chaos





# AFRINIC – Corruption and Chaos

2018: Sexual harassment and intimidation  
scandal



# AFRINIC – Corruption and Chaos

2018: Sexual harassment and intimidation scandal

2019: the “Great African IP Heist” – 4,100,000 addresses stolen



# AFRINIC – Corruption and Chaos

2018: Sexual harassment and intimidation scandal

2019: the “Great African IP Heist” – 4,100,000 addresses stolen

2021: AFRINIC CEO allegedly involved in heist





# AFRINIC – Corruption and Chaos

2018: Sexual harassment and intimidation scandal

2019: the “Great African IP Heist” – 4,100,000 addresses stolen

2021: AFRINIC CEO allegedly involved in heist

2022: Mauritania prepares to arrest AFRINIC members



# My FBI Story



# Subpoenas Are Part of the Job



A perspective view of a server room aisle. The aisle is flanked by rows of server racks on both sides. The racks are filled with various electronic components, and some have glowing purple lights. The floor is a light-colored tile with dark lines. The ceiling has some equipment and cables. The overall lighting is a mix of blue and purple, creating a futuristic or high-tech atmosphere. A white horizontal line is drawn across the middle of the image, and a white vertical line is on the left side, intersecting the horizontal line.

**Take This Home With  
You**

# Keep these in mind when sourcing proxies:

---

Subnet diversity is key

Be aware of “noisy  
neighbors”



# Any questions?

**Neil Emeigh**

**[neil@rayobyte.com](mailto:neil@rayobyte.com)**

**Extract Summit 2022**

**[www.rayobyte.com](http://www.rayobyte.com)**