The IP Address Structure is Multifractal: An Invariant of Observed Internet Traffic?

Table of Contents

Introduction: Leveraging Prefix Structure for DDoS Detection [4]

- Promises DDoS detection at the prefix level
- “[…] attack and benign sources are not uniformly distributed […], [they] tend to cluster in distinct prefix level structures formally described by multifractal models” (Appendix A)

Introduction: Leveraging Prefix Structure

- Their references regarding:
  - benign traffic: Observed Structure of Addresses in IP Traffic Kohler et al. 2002 and 2006 [1,3]
  - malicious traffic: Toward a Model for Source Addresses of Internet Background Traffic Bradford et al. 2006 [2]

Points of Interest

- Key findings?
- Methods?
- Datasets?

Observed Structure of Addresses in IP Traffic (2002) [1]

- "How does behavior change as aggregation increases?"
  - e.g. from individual TCP flow vs gigabit traffic conglomerate
- Packet distribution among conglomerate's component addresses
- How do those addresses aggregate?

Destination Prefix Aggregation

- 2 packets are in the same aggregate if their prefix matches
- Depending on chosen “resolution” (e.g. /8, /16, /24)

Their Data 1

Trace	Description	Time (hr)	*N*	Packet count	Sampled?
U1	Access link to a large university	~ 4.0	69,196	62,149,043	no
U2	Access link to a large university	~ 1.0	144,244	101,080,727	no
A1	ISP	~ 0.6	82,678	33,960,054	no
A2	ISP	1.0	154,921	29,242,211	no
R1	Link from a regional ISP	1.0	168,318	1,476,378	1 in 256
R2	Link from a regional ISP	2.0	110,783	1,992,318	1 in 256
W1	Access link in front of a large Web server	~ 2.0	124,454	5,000,000	no

Their Data 2

- From 1998 to 2001
- Omnidirectional → only look at destination addresses
- N = number of distinct destination addresses

Method

- Look at R1 distributions of:
  - Packet counts per TCP/UDP flow
  - Destination addresses
  - Destination address aggregates
- Log-log complementary CDF graphs
- Looking at factors affecting aggregate packet counts
- Multifractal modeling
- Testing the model against structural metrics

R1 Log-log complementary CDF of packet counts

- TCP/UDP flow packet counts
- Large aggregates

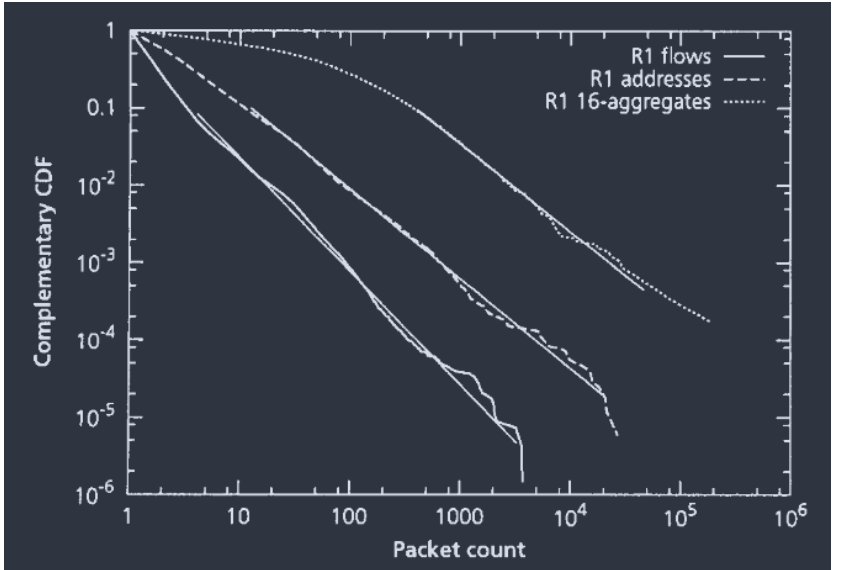


Figure 4—Log-log complementary CDF of packet counts for R1 flows, addresses, and 16-aggregates. All are consistent with power-law distributions. The fit lines have slopes −1.46, −1.16, and −1.13, respectively.

Factors affecting aggregate packet counts

- Address packet counts
- Address structure
- Correlation between these two

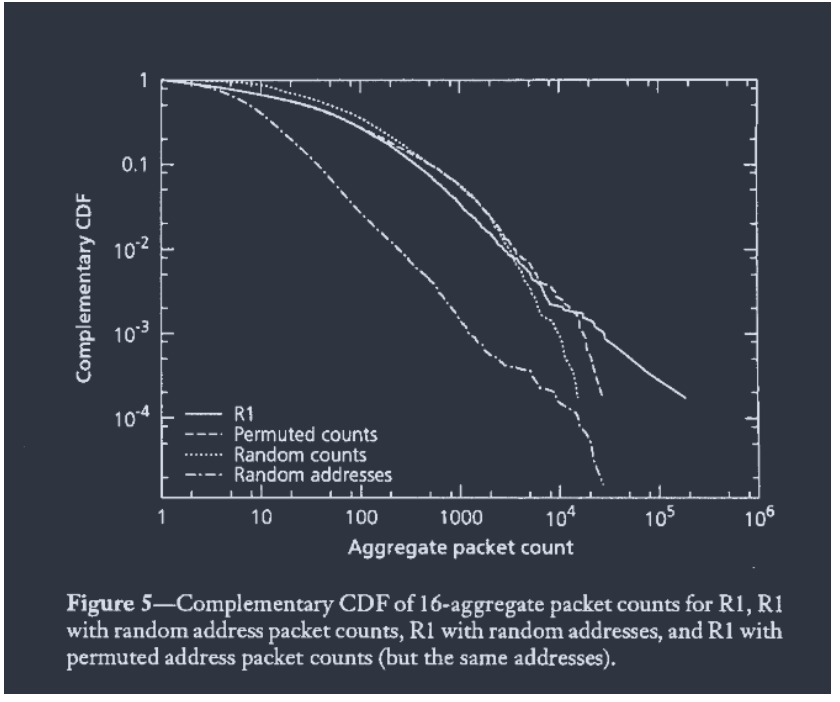


Figure 5—Complementary CDF of 16-aggregate packet counts for R1, R1 with random address packet counts, R1 with random addresses, and R1 with permuted address packet counts (but the same addresses).

Multifractal Model 1

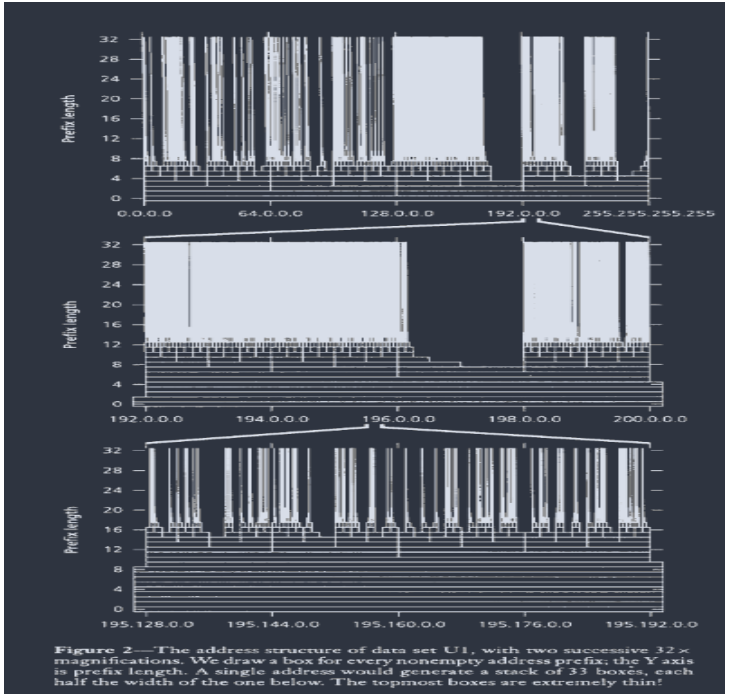


Figure 2—The address structure of data set U1, with two successive 32× magnifications. We draw a box for every nonempty address prefix; the Y axis is prefix length. A single address would generate a stack of 33 boxes, each half the width of the one below. The topmost boxes are extremely thin!

- U1 address structure
- /0 → /5 → /10
- Self-similar

Multifractal Model 2

- Check the multifractal spectrum of a data set
- Develop using Cantor dusts, extended to Cantor measures
- Matches their real data “fairly” well at all scaling exponents

Multifractal Model for A2

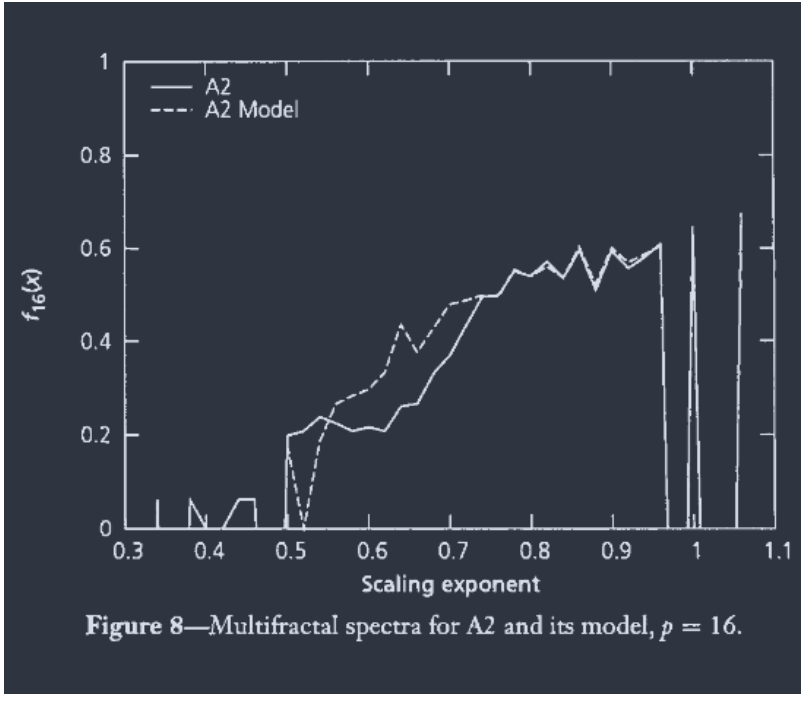


Figure 8—Multifractal spectra for A2 and its model, $p = 16$.

Potential Cause for this Multifractal Structure

- Multiplicative process/cascade
- IP address Allocation
  - (ICANN → RIRs → …)
  - Left-to-right allocation

Structural Metrics of Aggregates

- Active aggregate counts and Aggregation ratio
- Discriminating prefixes
- Aggregate population distributions
  - Limitation of their model: fails to match clustering for A2 and U1
  - sparse clusters at mid-range prefixes (/18 to /25) + dense host concentration at long prefixes (/27+)

Aggregation Ratio as a Fingerprint

- Characterization of traffic visible at a location
- Stable curve (40 min+)
- Segment traces with the same N:
  - U2, A1 and A2 (10 sections each)
  - Half the addresses change from section to section
  - Aggregation ratio curve remains stable for

Key findings

- Address structure is the strongest factor affecting the packet count distribution for medium-sized aggregates (/16)
- Address structures are well modeled by multifractals
- Address structure differs between sites and is stable over short time scales

Toward a Model for Source Addresses of Internet Background Radiation (2006) [2]

- Initial investigation of the network locations of hosts that generate malicious background radiation
- “Where does Internet background radiation come from?”
- Characterize background radiation **source addresses** across the IPv4 address space for /8, /16 and /24 aggregates

Their Data

- 10.12. till 16.12. 2004
- Dshield: aggregated IDS, Firewall logs
- Active iSink telescopes:
  - Campus: 16.000 addresses
  - Provider: 16 million addresses

Name	#Scans	#Unique Source Addresses
Dshield	296,004,577	7,694,291
Campus	36,774,175	448,894
Provider	212,481,885	2,355,150

Table 1. Details of the data sets used in this study which were collected from Dec 10-16 2004.

Method

- Examine the distribution of the number of unique source hosts per source network (mostly /24 granularity)
- Spatially adaptive density estimation technique
- Multifractal modeling

Density Plots

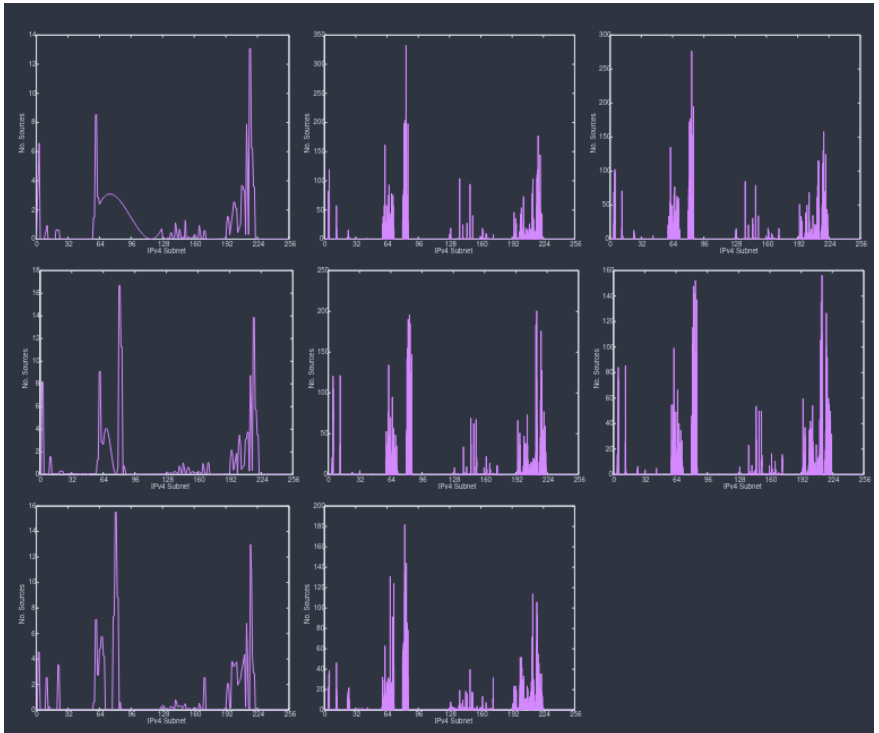


Fig. 1. Density plots for Campus (top), Provider (middle) and DSHIELD (bottom row) at /8 (left column), /24 (middle column) and /24-all (spoofed + non-spoofed) (right column). We do not have ability to obtain non-spoofed TCP DSHIELD data. Left and middle columns for Campus and Provider sites only include verifiably non-spoofed TCP source addresses. X-axis: Subnet, Y-axis: No. observed source addresses

- Similar for all 3 data sources at all aggregation levels
- /24 focus of subsequent analysis

Temporal Characteristics

- Density plots per data set broken down into daily aggregates
- Source densities consistent on a daily basis

Fig. 2. Density plots for the Campus(top), Provider(middle) and DSHIELD(bottom) data set in /24 aggregates for three days (left-right: Dec 11, 13, 15 2004) of the measurement period. X-axis: Subnet, Y-axis: No. observed source addresses

Density Broken Down by Destination Port

- Source address densities change across ports
- Sparsity and burstiness are maintained across data sets

Multifractal Modeling 1

- Use the Method Kohler et al. proposed to calculate the multifractal spectrum for the campus data set
- At medium scales it shows properties of multifractal scaling
- Similar results for other data sets

Multifractal Modeling 2

- Propose an algorithm for generating random IP distributions with multifractal characteristics
- Random cascade model
- Re-allocates mass probablistically based on a parametric density function

Cause of the Multifractal Structure?

- Refer to Kohler et al. (address allocation)
- Occurrence of Infected hosts dependent on the network “type”:
  - Well managed networks: likely to be zero
  - ISPs with home users: random
  - Poorly managed networks: complete

Key findings

- Source addresses of Internet background radiation are sparse, bursty and show consistent structure across different network monitors over time
- They form tight clusters
- They can be modeled using multifractals
- New method to model traffic

Implications

- Assumption: most background radiation is generated by systems that have been compromised
  - Implies Machines:
    - are currently vulnerable
    - Probable targets
    - Most likely part of a botnet

Observed Structure of Addresses in IP Traffic (2006) [3]

- Slight update to their 2002 paper:
  - Mention [2] multifractal modeling method
  - Compare new 2005 traces from with 2001 traces

- →structural characteristics “not far from those observed at that laboratory in 2001”
-

Claims So Far

- Address structures in general Internet traces is multifractal
- They differ between sites and are stable over time windows
- Sources of Internet Background radiation cluster in specific networks
- They also have a multifractal structure

The Multifractal IP Address Structure: Physical Explanation and Implications (2025) [5]

- Extend the prior work of Kohler et al.
- Root cause analysis of the multifractal structure
- Adapting existing methods for IP address analyzation

Their Data 1 Traces

Tol/Tol Instances	Date	Duration	Type	IP Count	Packet Count
Tol: Backbone-link (uni-directional)					
CAIDA-dir-A	2019-01-17 5:00AM	900s	pcap	2,381,306	566M
CAIDA-dir-B	2019-01-17 5:00AM	900s	pcap	3,863,987	1.23B
Tol: Backbone-link (bi-directional)					
MAWI-20150202	2015-02-02 2:00PM	900s	pcap	5,571,625	99M
MAWI-20150710	2015-07-10 2:00PM	900s	pcap	4,415,599	191M
MAWI-20151002	2015-10-08 2:00PM	900s	pcap	4,818,370	135M
MAWI-20180316	2018-03-16 2:00PM	900s	pcap	4,567,614	69M
MAWI-20180807	2018-08-07 2:00PM	900s	pcap	4,635,311	73M
MAWI-20181107	2018-11-07 2:00PM	900s	pcap	4,677,191	80M
MAWI-20190901	2019-09-01 2:00PM	900s	pcap	4,689,835	87M
MAWI-20210110	2021-01-10 2:00PM	900s	pcap	265,794	52M
MAWI-20210614	2021-06-14 2:00PM	900s	pcap	180,532	74M
MAWI-20211212	2021-12-12 2:00PM	900s	pcap	149,572	55M
Tol: Access-link (UCSB) (incoming)					
UCSB-20220428	2022-04-28 12:15	900s	pcap	194,498	1.02B
UCSB-20220921	2022-09-21 19:25	900s	pcap	112,164	1.05B
UCSB-20221205	2022-12-05 20:15	900s	pcap	186,575	1.1B
Tol: Access-link (UO) (incoming)					
UO-20181106	2018-11-16 00:00	1 day	netflow	1,805,085	9B
UO-20190829	2019-08-29 00:00	1 day	netflow	1,497,311	12B
UO-20200213	2020-02-13 00:00	1 day	netflow	786,607	36B
UO-20211106	2021-11-06 00:00	1 day	netflow	1,288,775	26B
UO-20220517	2022-05-17 00:00	1 day	netflow	668,817	23B

Table 5. Our collection of recent traffic traces.

- more recent internet traces for their investigation:
- 900 second pcaps
  - CAIDA-dir-A/B
  - MAWI 2015/2018/2019/2021
  - UCSB (santa barbara campus network)
- 1 day netflow

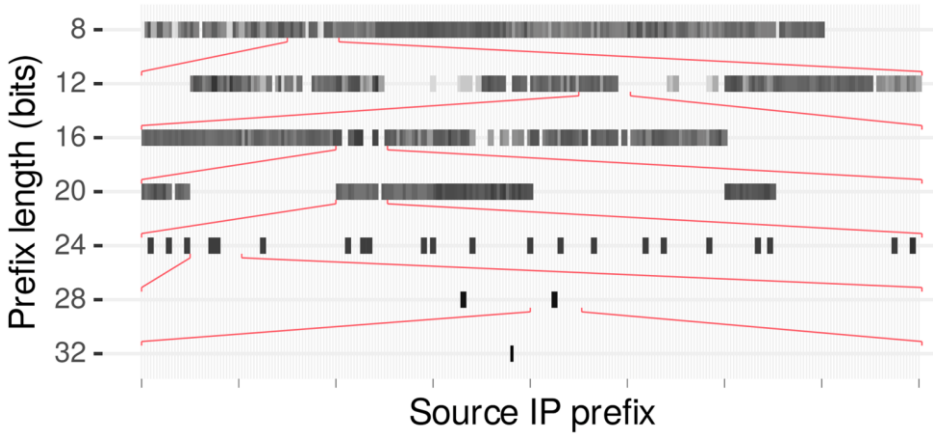
Their Data 2

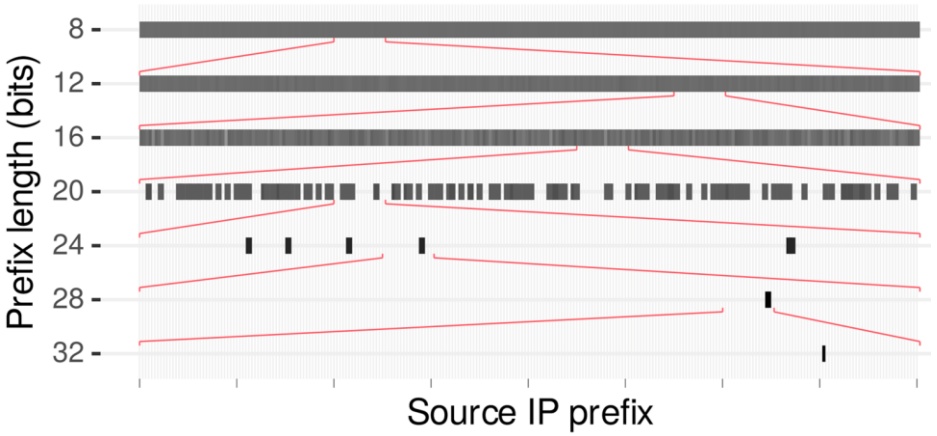
- IANA and bulk WHOIS records from each of the RIRs
  - publicly-available address allocation datasets

Method

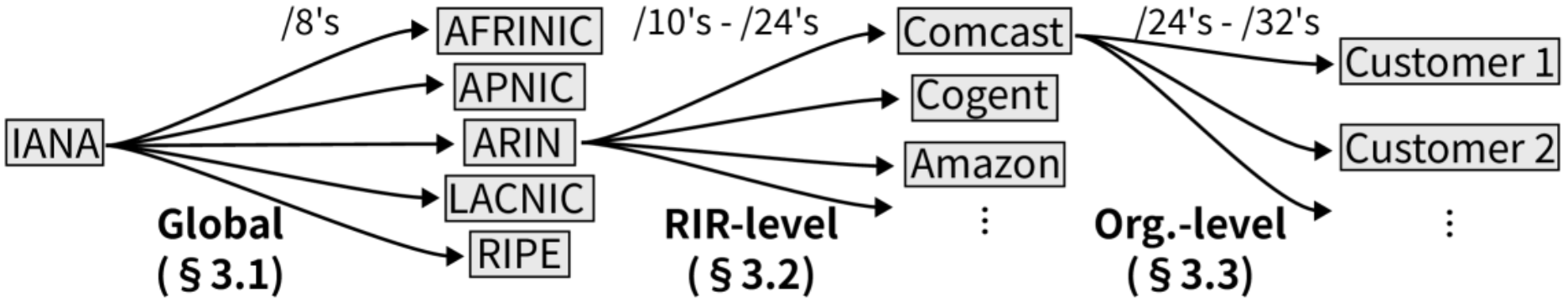
- Compare 500k CAIDA vs 500k synthetic uniformly distributed IPs
- Investigating the IP allocation process:
  - Extracted bulk WHOIS network records from 4 RIRs (~8.7M IPv4, ~1.3M IPv6)
  - Filtered using IANA allocation ranges to remove placeholders and unmanaged blocks
  - Converted non-standard integer ranges into largest valid CIDR prefixes (applied to ~0.7% of records)

Compare 500k CAIDA vs 500k synthetic uniformly distributed IPs





Investigating the IP Allocation Process 1



Investigating the IP Allocation Process 2

- there are not only 3 levels (IANA, RIR, ORG)
- each level has additional iterations
- Yields a cascade with ~8 iterations on average

Key Findings

- IP address Allocation process is a conservative cascade and responsible for the multifractal structure
- Developed a toolbox for analyzing finite and discrete sets of IP addresses

Discussion

Sources

- [1] 2002 Kohler et al. Observed Structure of Addresses in IP Traffic: 10.1145/637201.637242
- [2] 2006 Barford et al. Toward a Model for Source Addresses of Internet Background Radiation: https://pages.cs.wisc.edu/~pb/sources_final.pdf
- [3] 2006 Kohler et al. Observed Structure … : 10.1109/TNET.2006.886288
- [4] 2024 Misa et al. Leveraging Prefix Structure to