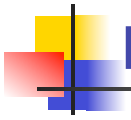
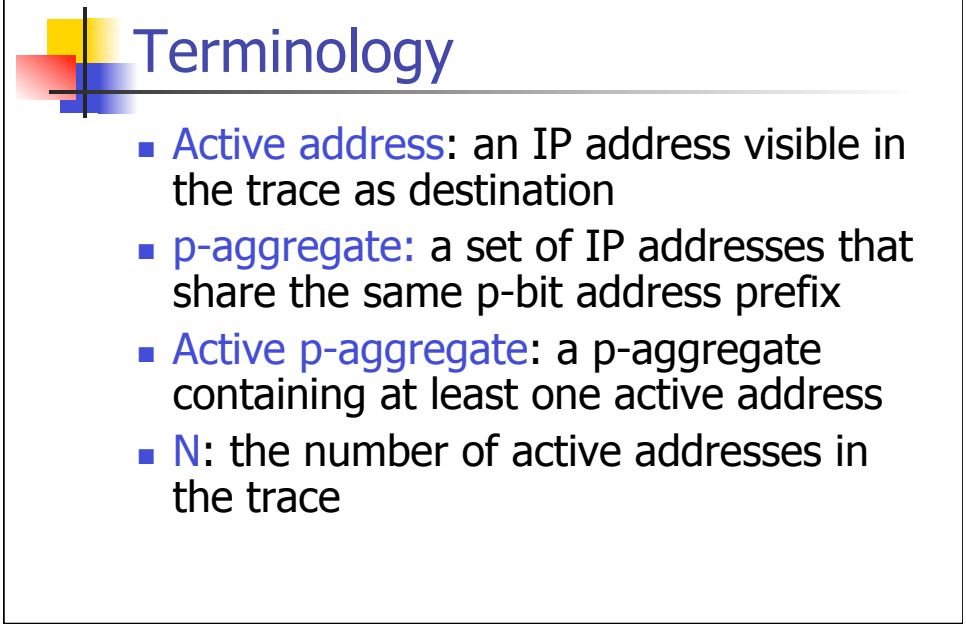
Observed Structure of Addresses in IP Traffic

CSCI 634, Fall 2010



More specific

- Structural characteristics of destination IP addresses seen on Internet links
  - Traces are omni-directional
  - Set of source addresses is roughly equal to set of destination addresses
- Destination address prefix based aggregation


Terminology
Active address: an IP address visible in the trace as destination
p-aggregate: a set of IP addresses that share the same p-bit address prefix
Active p-aggregate: a p-aggregate containing at least one active address
N: the number of active addresses in the trace

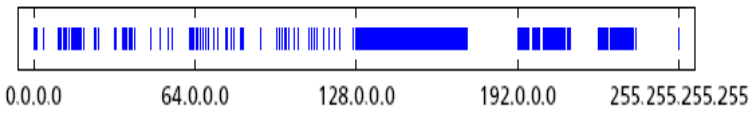
Problem statement

- How can we model the set of destination IP addresses visible on the access links?
- In particular, how can we model the addresses aggregate (address structure)?

Address structure

- The arrangement of active addresses in the address space

Example from a 4-hour trace at a university access link:



Trace collection

Name	Description	ΔT	# pkts	N	
U1	large university access link	~ 4 h	62M	69,196	
U2	large university access link	~ 1 h	101M	144,244	
A1	ISP	~ 0.6 h	34M	82,678	
A2	ISP	1 h	29M	154,921	
R1	link from regional ISP	1 h	1.5M	168,318	§
R2	link from regional ISP	2 h	1M	110,783	§
W1	large Web site access link	~ 2 h	5M	124,454	

- Collected between 1998 and 2001

  Most anonymized while preserving prefix and class relationships

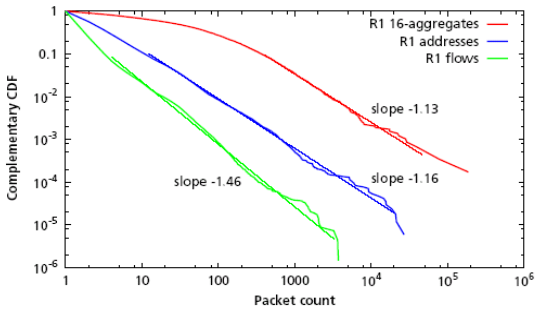
  § means sampled (1 in 256)

Importance of address structure

- A stable short-term fingerprint of a site
  - Different sites have different address structure
- Address structure is the most important factor affecting aggregate packet count distribution

Packet count distribution

- No. of packets per flow; (heavy-tail)
- per destination address; (heavy-tail)
- per destination address aggregate (MORE heavy-tail)



Factors affecting aggregate packet counts

- Address packet counts
  - No. of packet per destination address
- Address structure
  - No. of active addresses per aggregate
- Correlation between these two

Semi-experiments

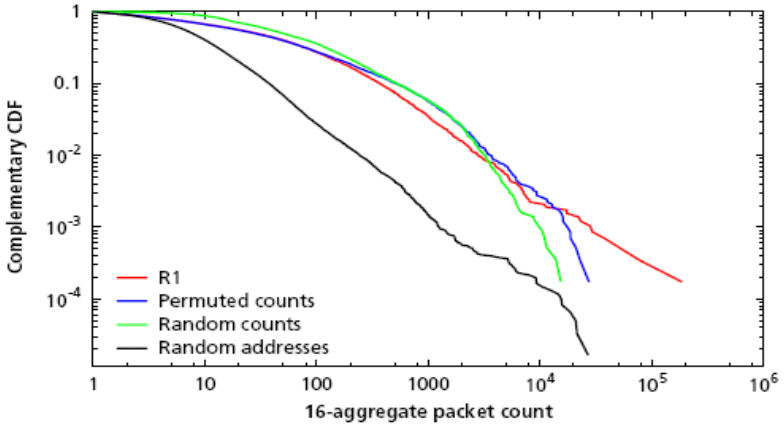
- Manipulate the data, destroying one factor at a time; see which factors impact aggregate packet counts
- "Random counts": destroy per-address packet counts

  Replace the (heavy-tailed) per-address packet count distribution with a uniform distribution over [0, 17.54]
- "Random addresses": destroy address structure

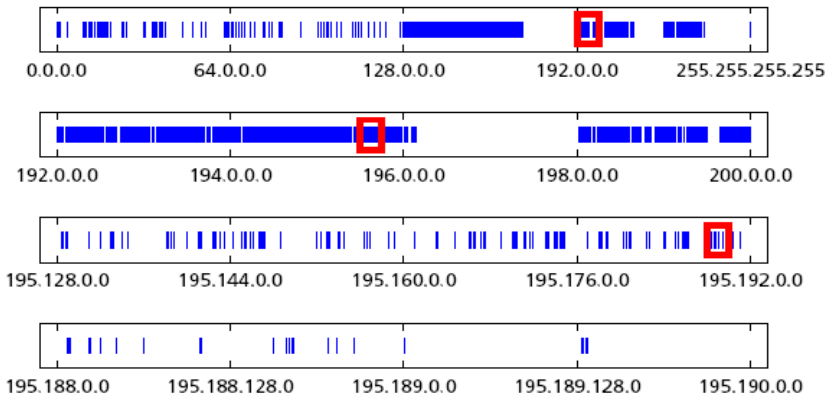
  Replace address structure with a uniform random distribution over the entire IP address space
- "Permuted counts": destroy correlation

  Permute per-address packet counts among the active addresses

Address structure matters most



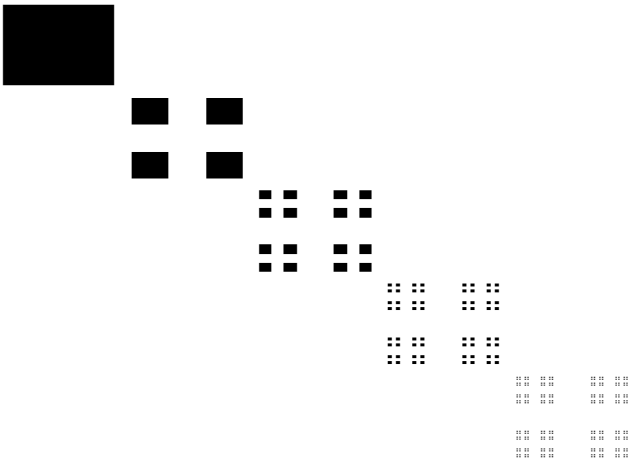
Tour of U1's address structure



Address structure looks self-similar

- Interesting structure all the way down
  - Visually self-similar characteristics
- Validate the intuition: multi-fractal model for address structure
  - An address structure viewed as a subset of the unit interval [0,1)
  - Cantor dust with two parameters
    - Dimension: active p-aggregates
    - Mass: active addresses within each prefix aggregate

Canonical Cantor Dust



Dimension for address space

- Let n_p equal the number of active /*p*s in a trace

  $n_{32} = N$

  $n_p \leq n_{p+1} \leq 2n_p$

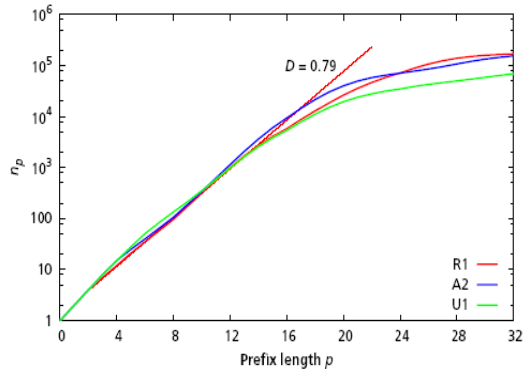
  each /*p* contains and is covered by 2 disjoint /$(p+1)$s

- Then $D = \lim_{p \to \infty} \frac{\log n_p}{p \log 2}$

  But $p \leq 32$ here, and expect sampling effects for high p

  Examine medium p to see if the limit exists

Fitness with prefix aggregates



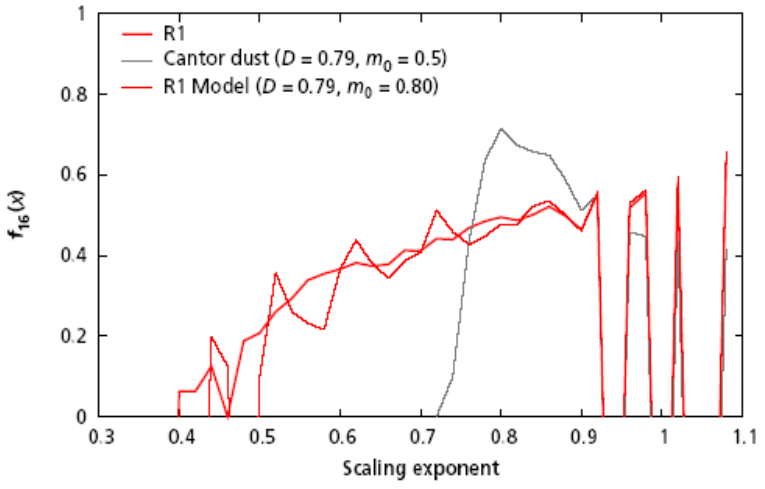
Why multi-fractal

- Mono-fractal only captures global scaling behavior of aggregate counts
- Address structure has different local scaling behavior (active addresses)
- Besides (capacity) dimension, introduce another parameter: mass

Multi-fractal Model

- Start with a cantor dust with dimension D
  - Repeatedly remove middle subinterval with proportion $h = 1 - 2^{1-1/D}$
- Unequally distribute a unit of mass between subintervals
  - Unequal distribution of mass leads to different local scaling behaviors

The model fits well



Causes

- Cascade effects:
  - A recursive subdivision plus a rule for distributing mass
- Procedure of address allocation
  - ICANN allocates short prefixes to providers
  - Providers allocates less shorter prefixes to customers
  - Share the same rule: left-to-right allocation

Is the model useful?

- Certainly the model doesn't *look* like real data:

How do we know whether we've captured relevant properties?

- Develop application metrics for address structures

  Contrast metrics among traces

  Compare with model

How densely addresses are packed?

- Metrics: active p-aggregate counts for prefix p
  - n_p again equals the number of active /ps in a trace
  - n_p measures how densely addresses are packed

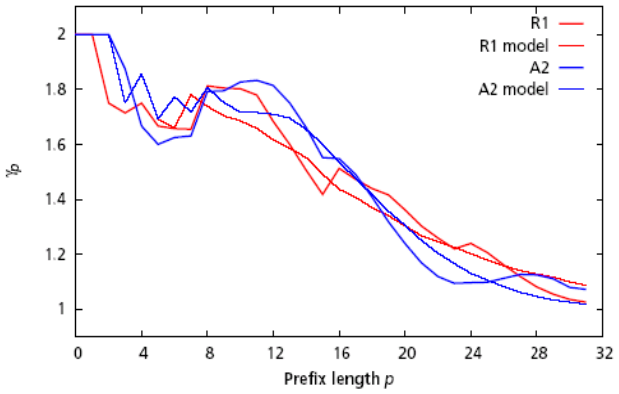
    If $N = 2^{16}$ and $n_{16} = 1$, addresses are closely packed

    If $N = 2^{16}$ and $n_{16} = 2^{16}$, addresses are well spread out

    Useful for algorithms keeping track of aggregates—shows how many aggregates there tend to be
  - $\gamma_p = n_{p+1}/n_p$ more convenient for graphs

    $N = \prod_{1 \le p < 32} \gamma_p$

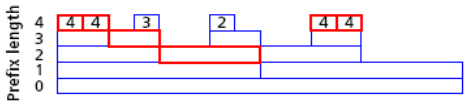
Aggregation ratio γ_p



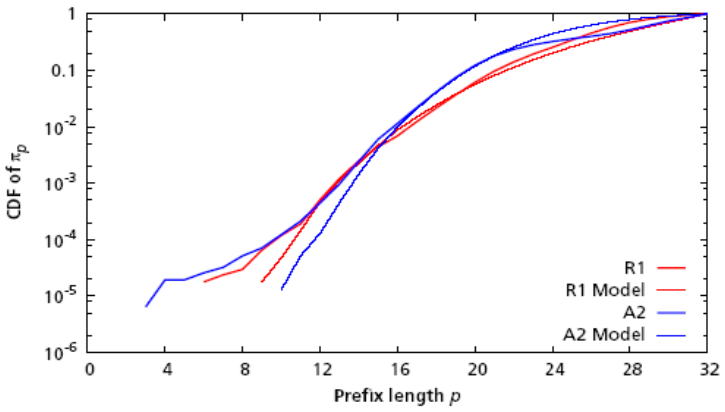
Characterize address separation

- Metrics: discriminating prefixes of an active address a
  - The prefix length of the largest aggregate whose only active address is a
  - πp : number of addresses that have discriminating prefix p

Example with 4-bit addresses:



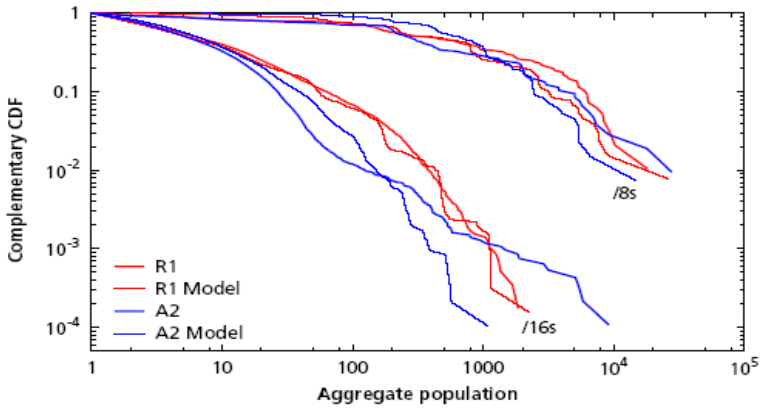
CDF of address discriminating Prefix counts



Aggregate population distribution

- Population of an aggregate is the number of active addresses contained in it
- Aggregates exhibit a wide range of population
- Aggregate population distributions are the most effect test to differentiate address structures

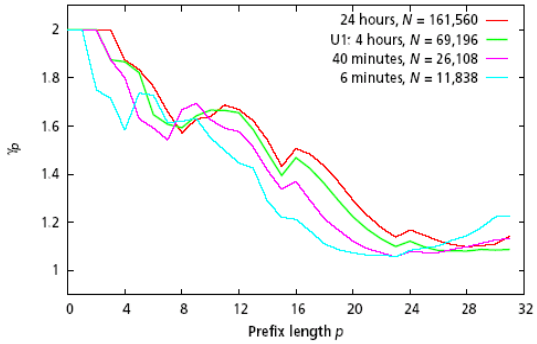
CDF of aggregate population distribution



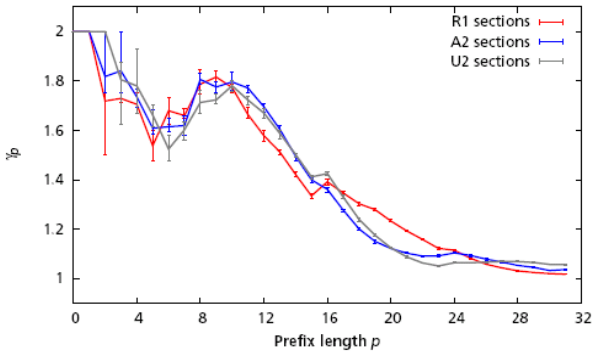
Properties of γp

- Sampling effect
  - How does the shape of the γp curve depend on N?
- Short-term stability
  - Is γp stable over short time period?

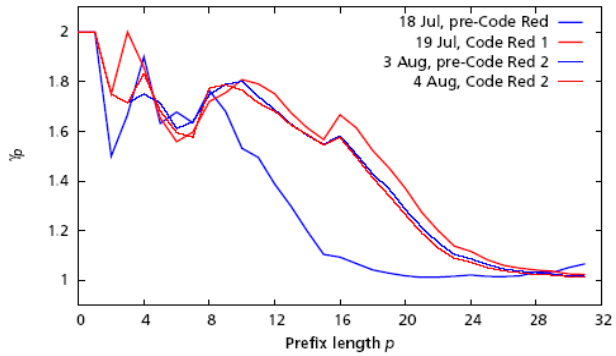
Similar curves



Relatively stable



Worm changes the shape



Worm changes aggregate packet count