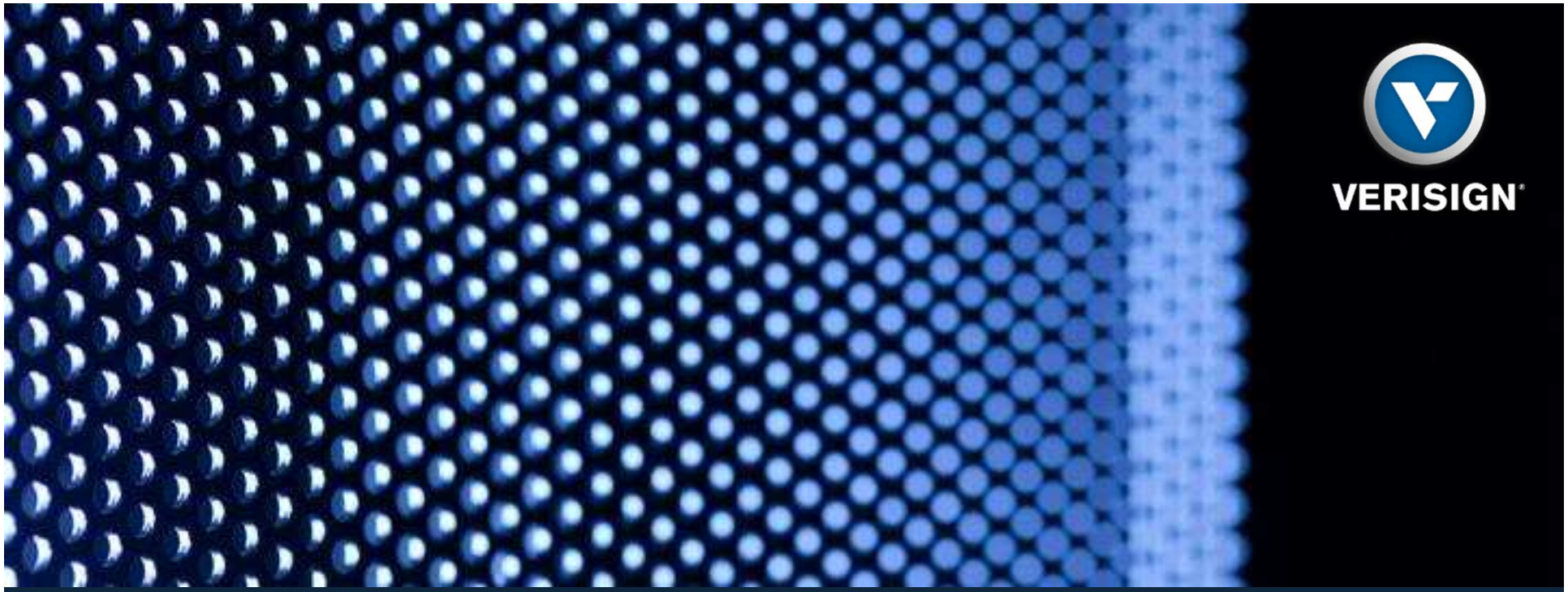


VERISIGN

The Effectiveness of Block Lists to Prevent Collisions

Matthew Thomas

Yannis Labrou

Andrew Simpson

March 2014

About this talk

- Examine the efficacy of block listing based on sampled DNS traffic data in order to prevent potential name collision events.
  - “Day in the life of the Internet” (DITL) Observations
  - Longitudinal study of A+J Root NXDomain Traffic

Data - Collection & Processing

DITL Data

2013 Collisions Project DITL Analysis

- JAS Global Advisors[1] and Demand Media[2] provided an uncomplicated extraction of DITL data for the applied gTLDs by year and by TLD

- Traffic volume and measurements were described in numerous other publications.

- Details: https://www.dns-oarc.net/node/332

[1] Kevin White [2] Roy Hooper

A and J Root NXDomain Data

- NXDomain traffic at Verisign-operated A+J root servers measured from July 16, 2013 until December 31, 2013.

- Contained ~3.6 billion NXD records and ~27.5 million unique second-level domains.*

Data Processing

- Top Level Domain (TLD) Exclusions
  - Limited to applied for gTLDs
  - ".home" and ".corp" removed due to high risk categorization[1]

- Second Level Domain (SLD) Exclusions
  - Chrome 10 character strings[2]
  - Technique based on ICANN published methodologies[3]

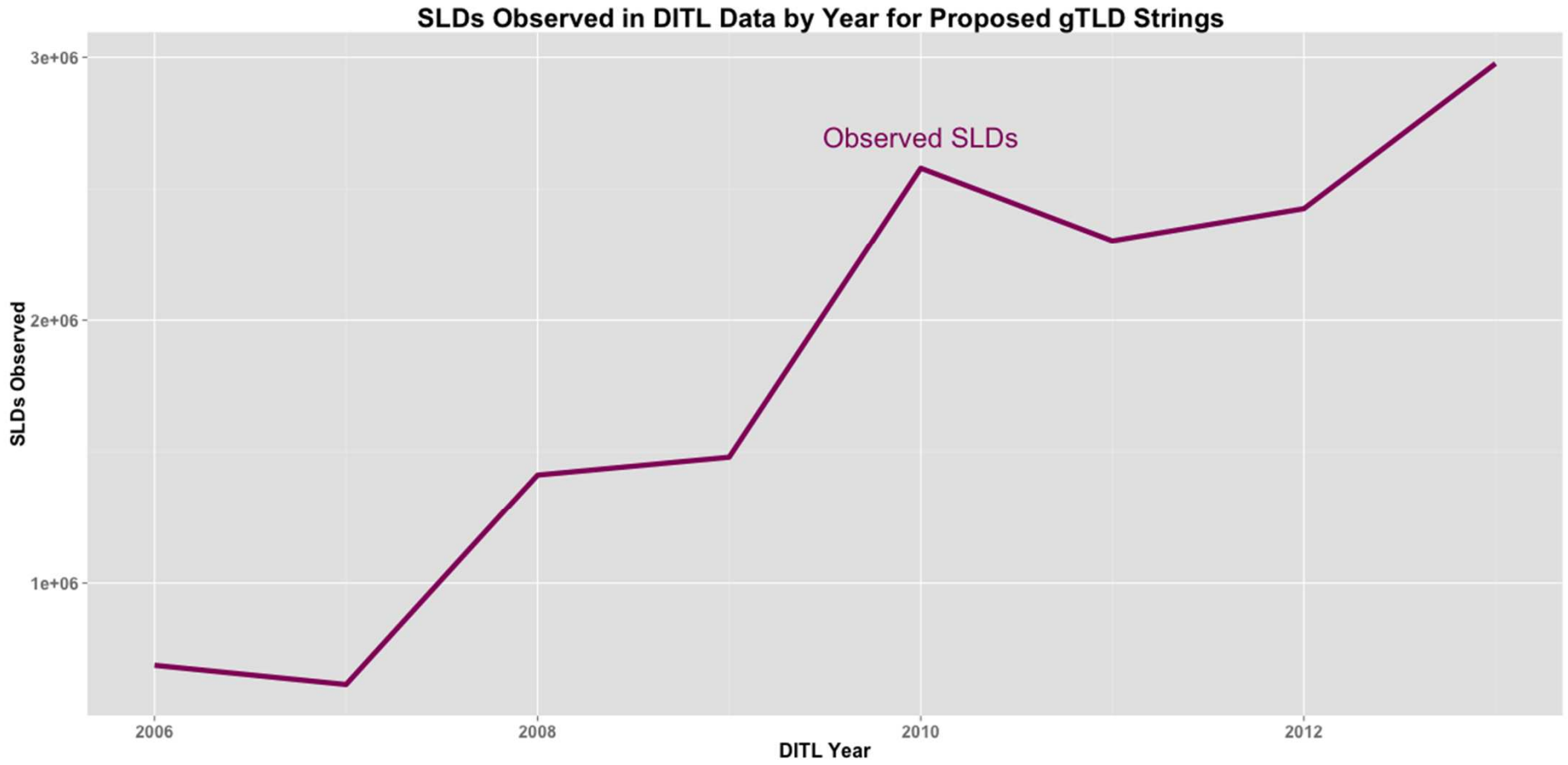
[1] http://www.icann.org/en/news/announcements/announcement-3-05aug13-en.htm

[2] https://isc.sans.edu/diary/Google+Chrome+and+(weird)+DNS+requests/10312

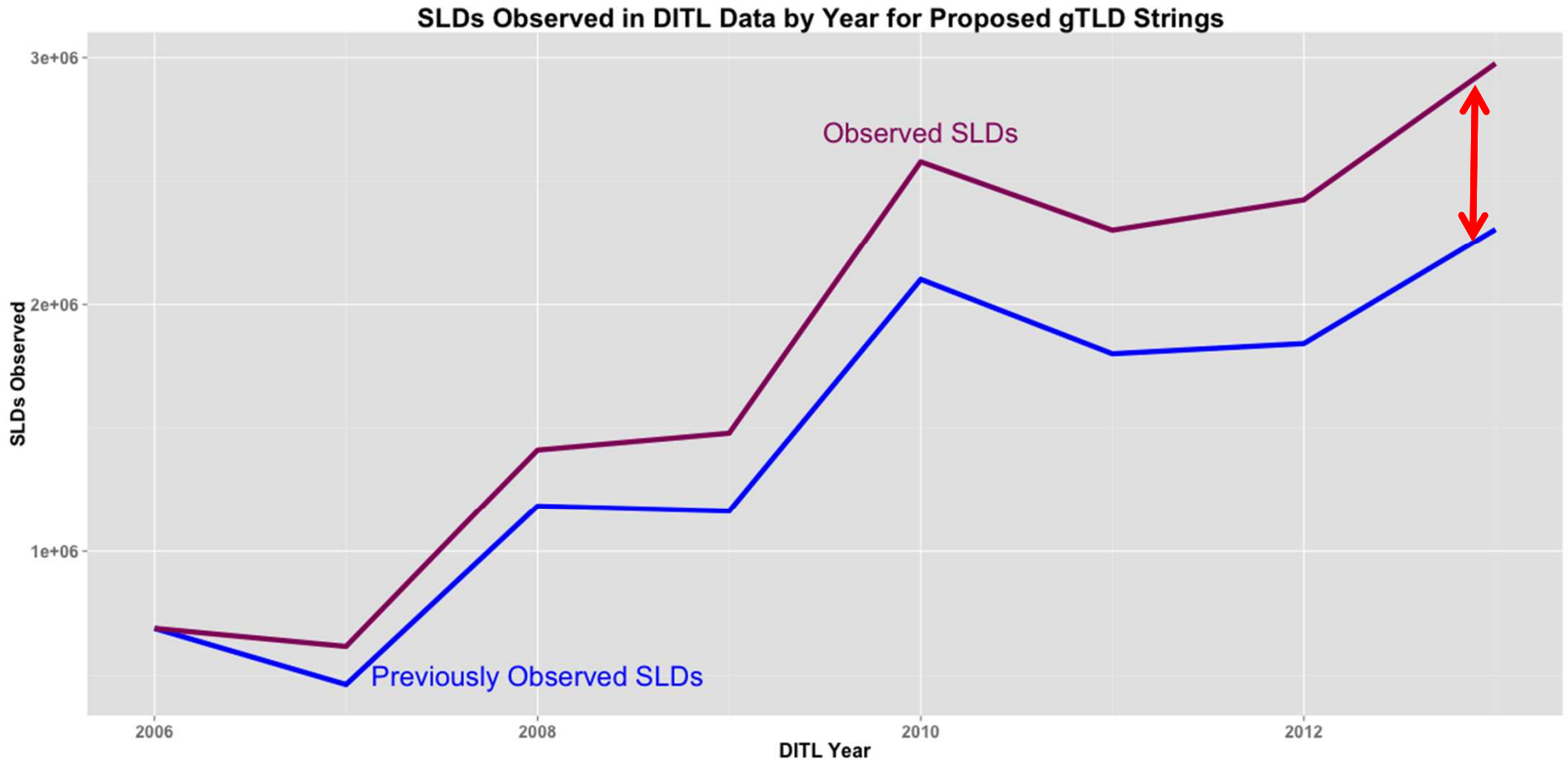
[3] E.g. http://www.icann.org/en/about/agreements/registries/luxury/luxury-apd-report-12nov13-en.htm

DITL Measurements

DITL – Longitudinal SLD Growth



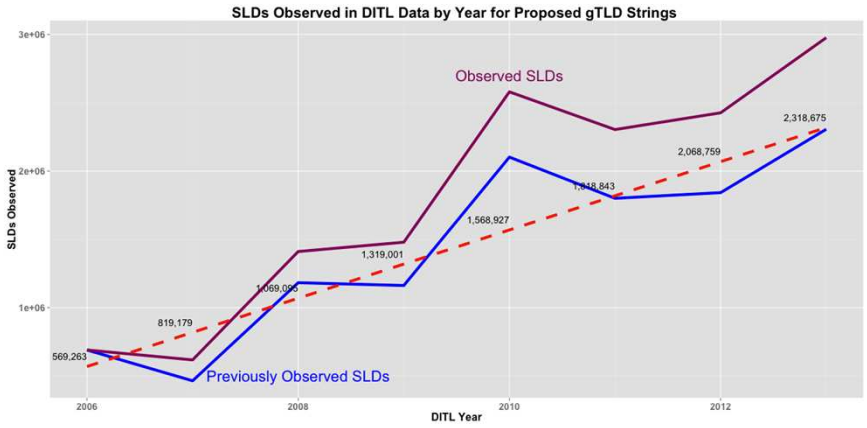
DITL – Longitudinal SLD Growth



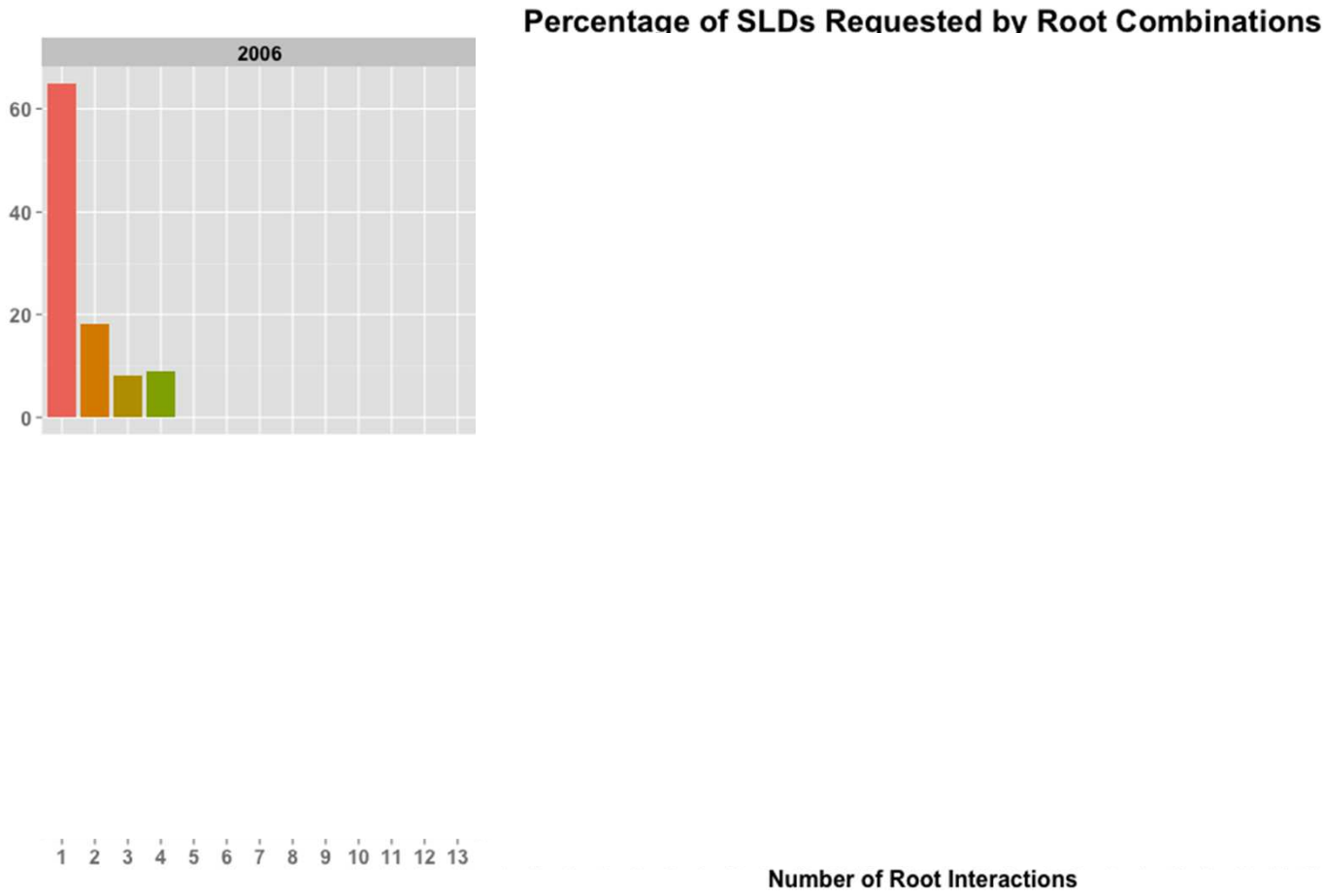
DITL – Longitudinal SLD Growth

- Steady growth rate of new SLDs
- Increasing delta of Observed and Previously Observed
- Early indication of problems using potential block listing due to highly entropic system

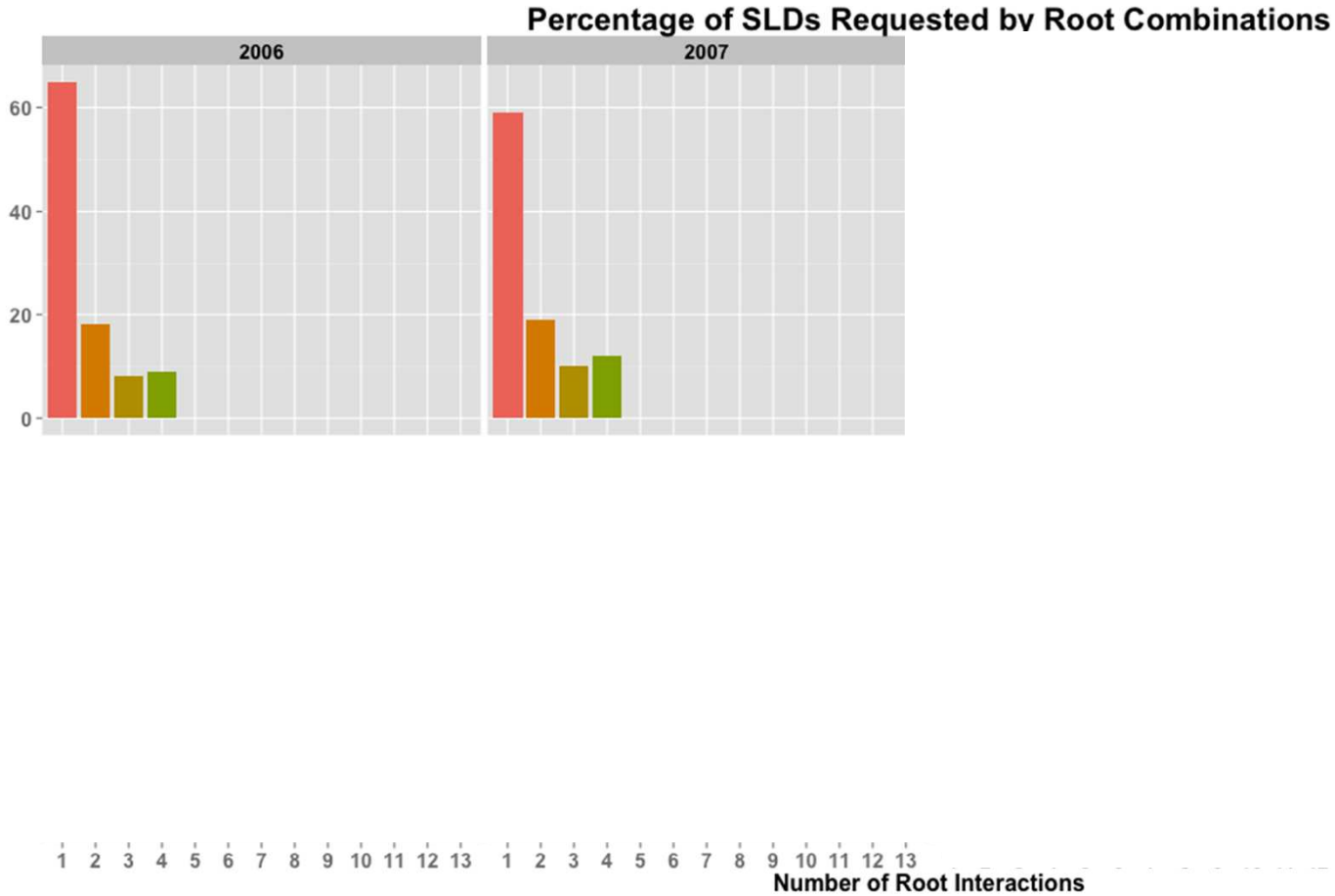
Can we study a subset of roots to measure the growth rate and dynamics of SLDs?



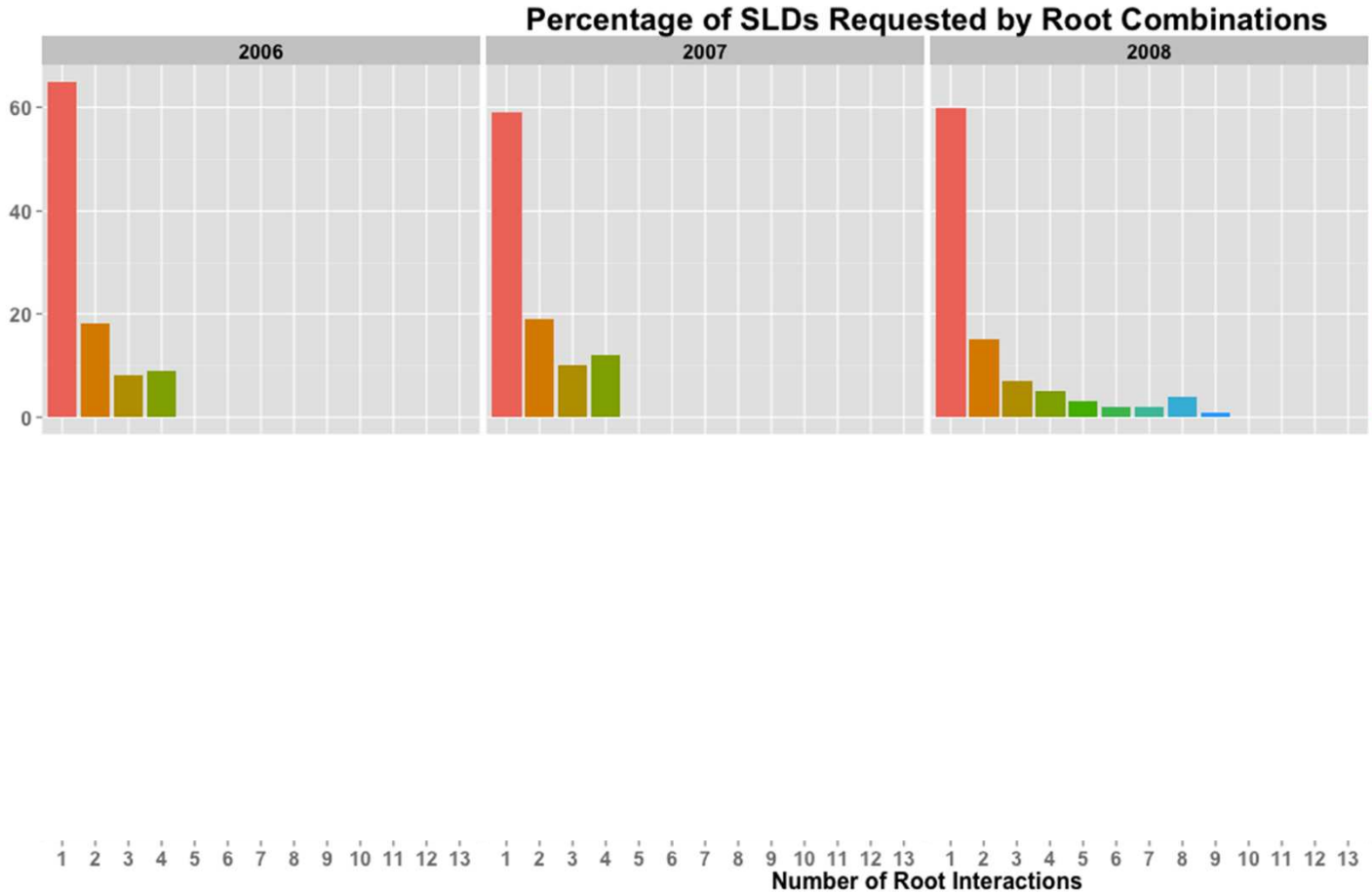
DITL – SLD Root Affinity



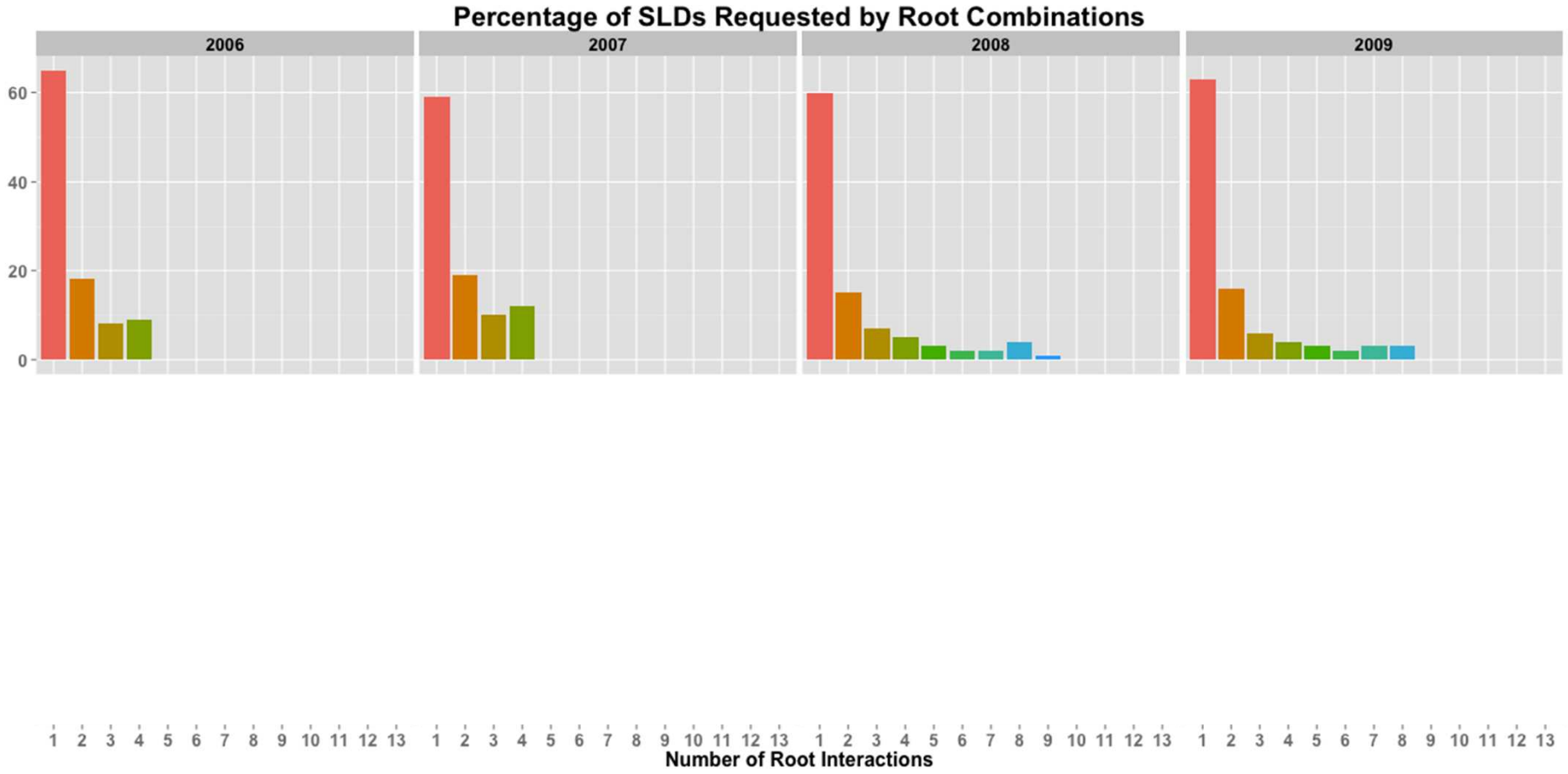
DITL – SLD Root Affinity



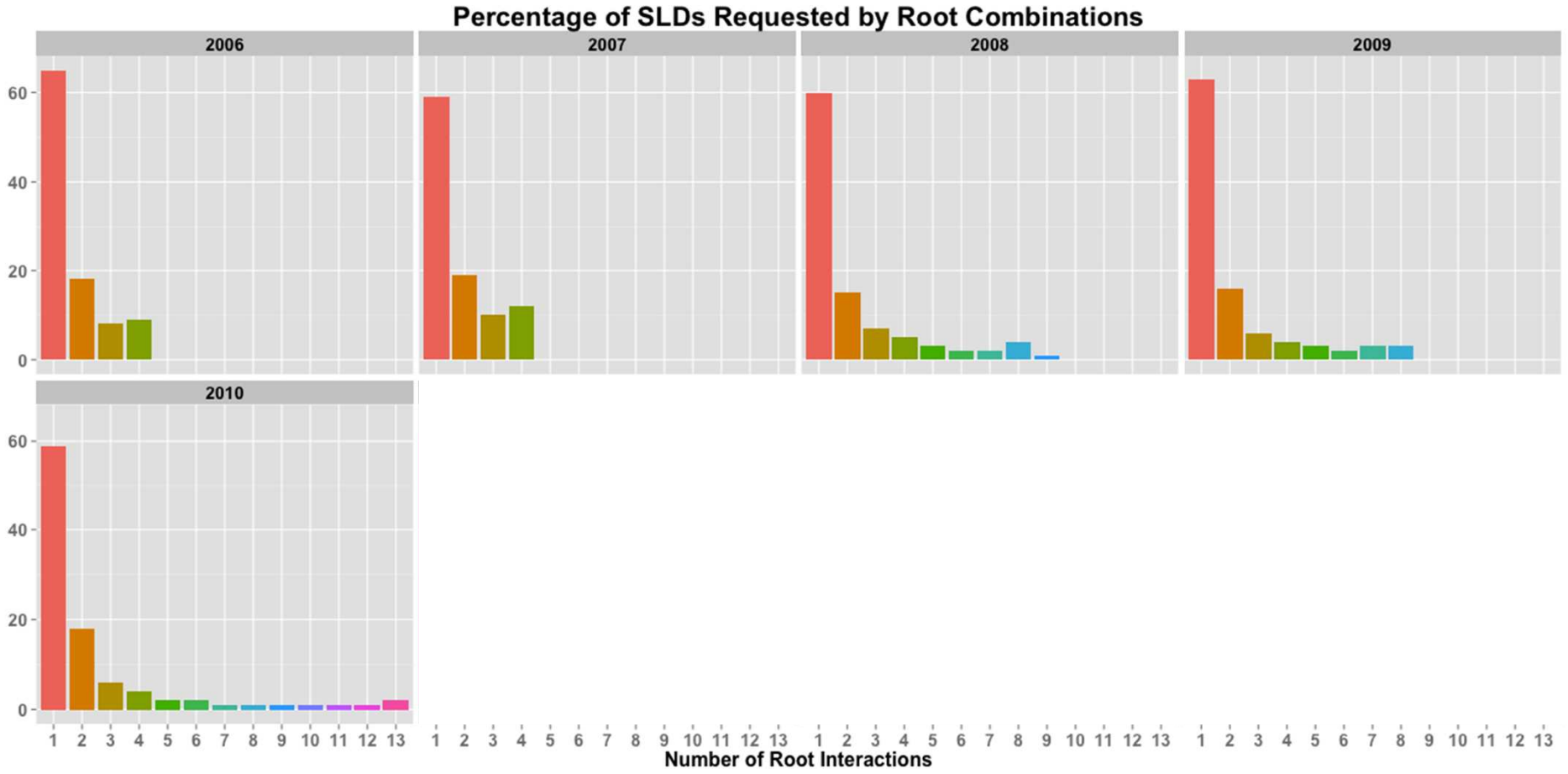
DITL – SLD Root Affinity



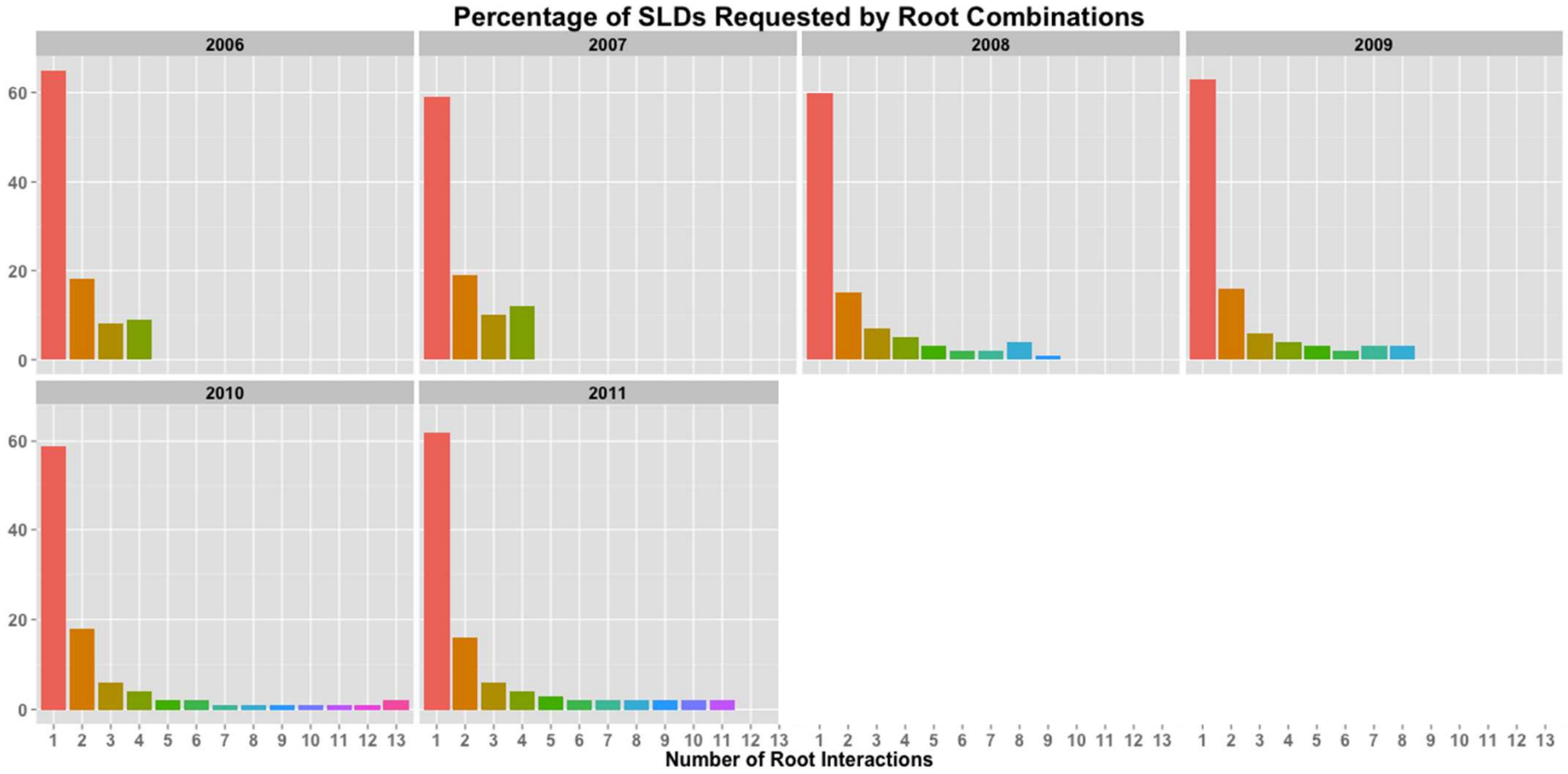
DITL – SLD Root Affinity



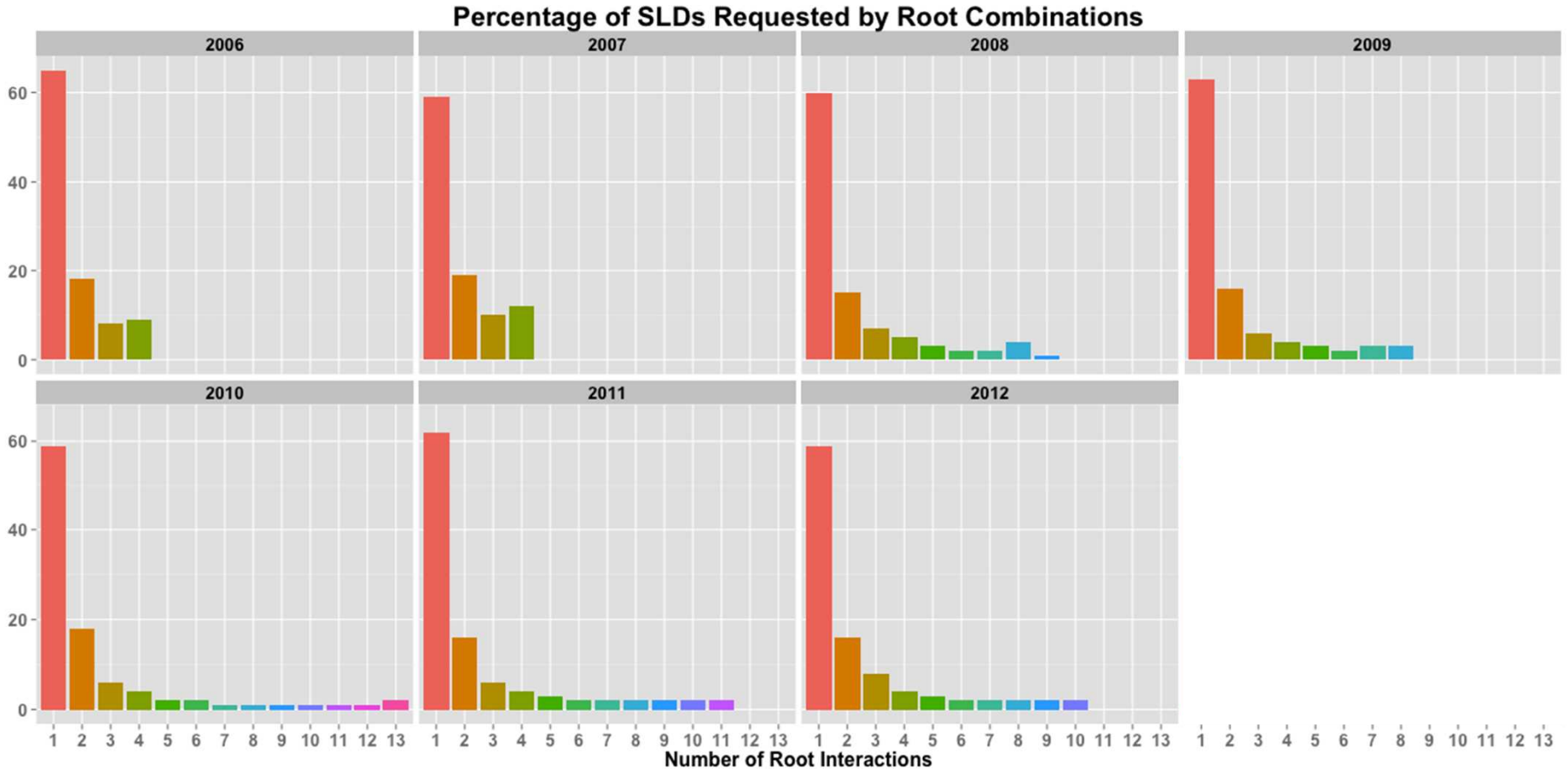
DITL – SLD Root Affinity



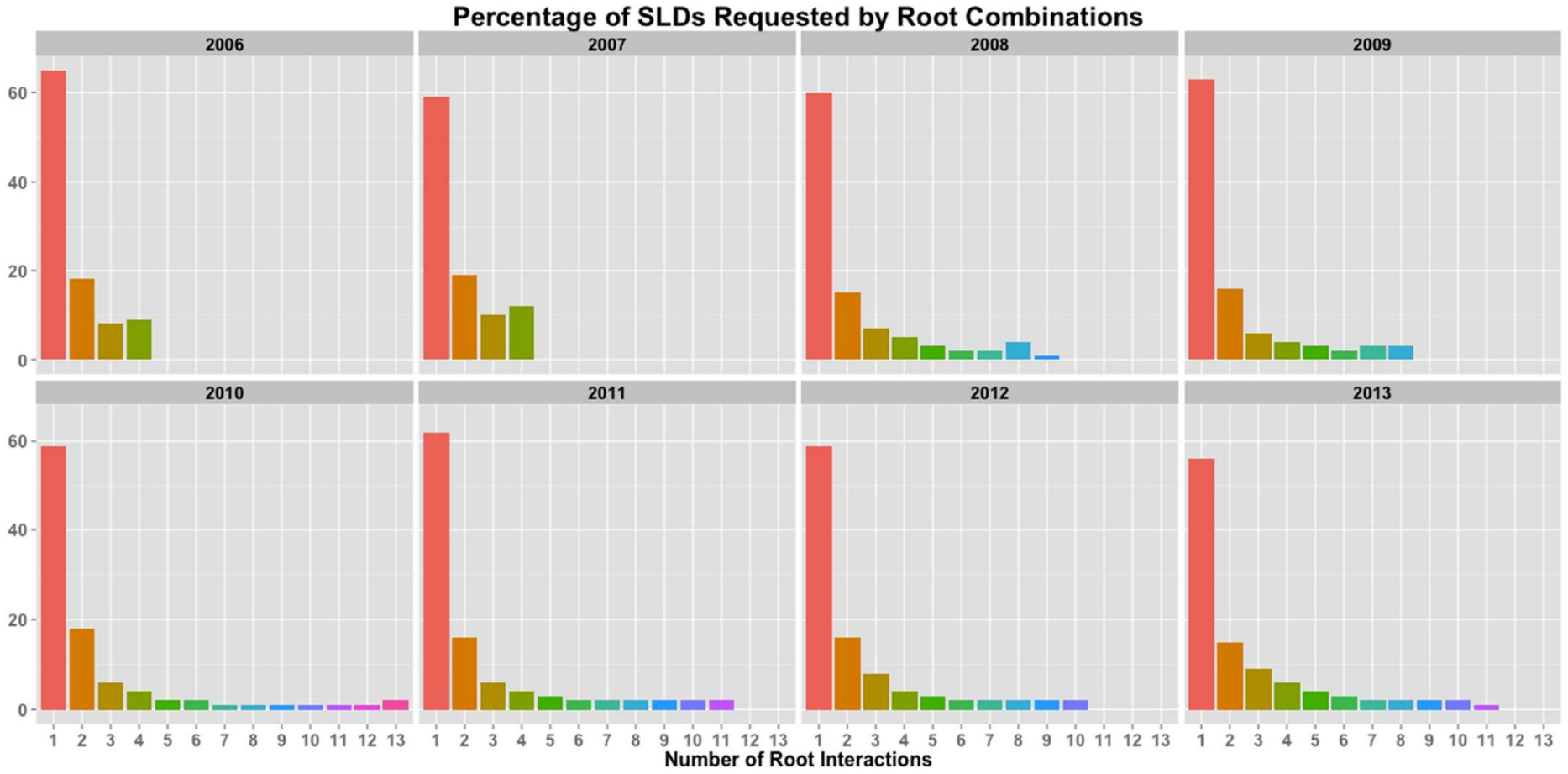
DITL – SLD Root Affinity



DITL – SLD Root Affinity



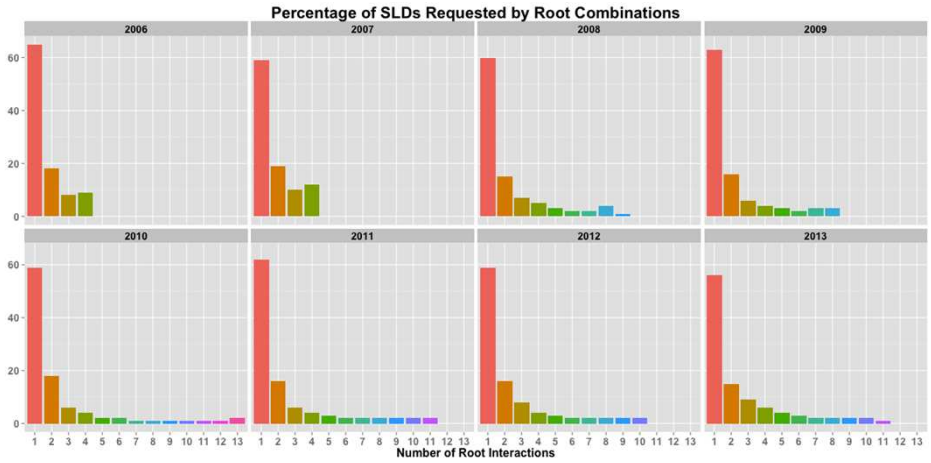
DITL – SLD Root Affinity



DITL – SLD Root Affinity

- Observational sampling at a specific subset of roots would be biased and of limited value for block listing purposes.
- High root affinity may prove useful to study a SLD's longitudinal patterns by sampling from a specific root.

Do specific roots exhibit higher levels of affinity that may influence root sampling?



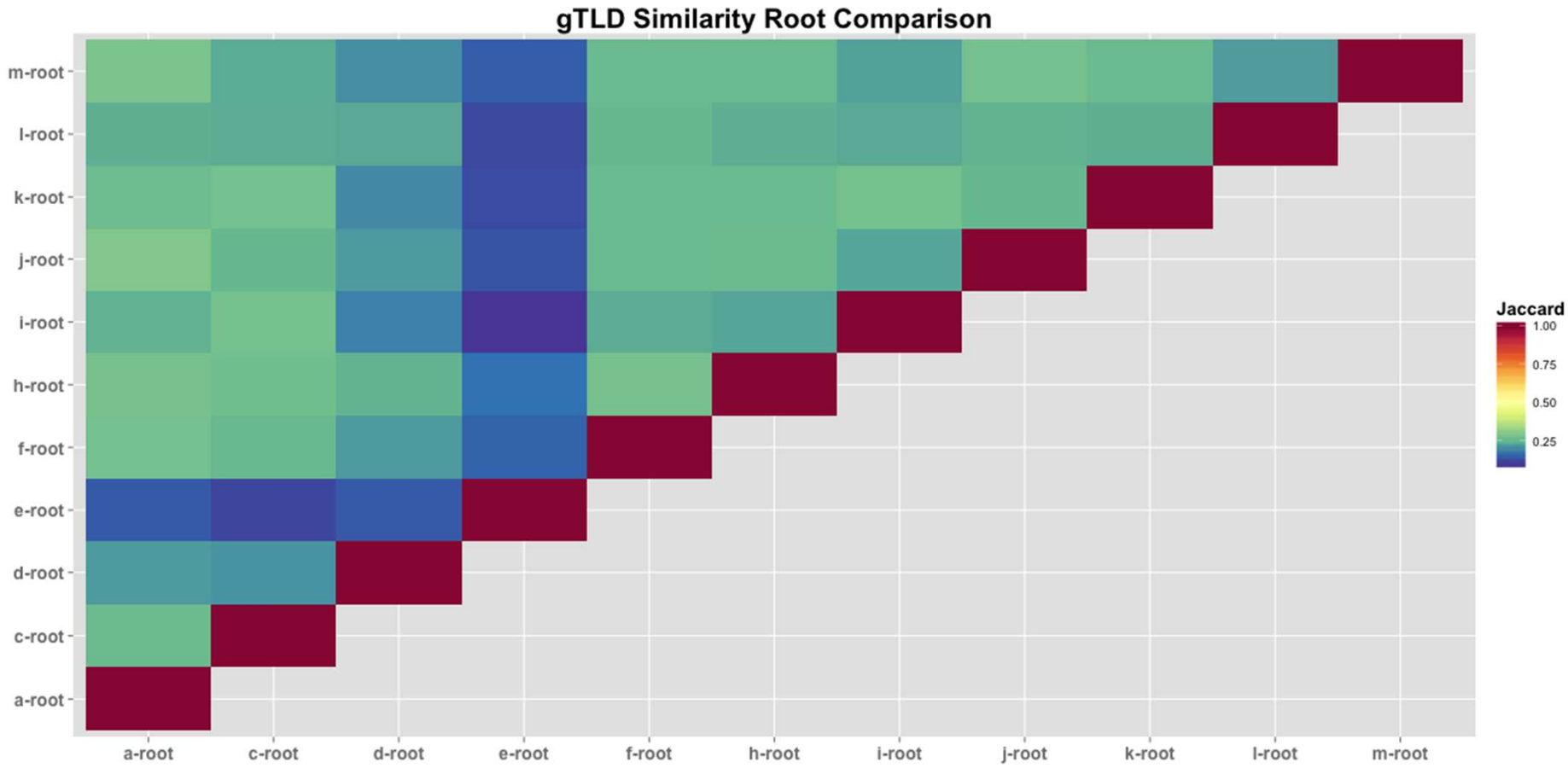
DITL – Intra-Root Affinity

- Similarity function is a real-valued function that quantifies the similarity between two entities.
- Jaccard Index is a statistic for comparing the similarity and diversity of sample sets.

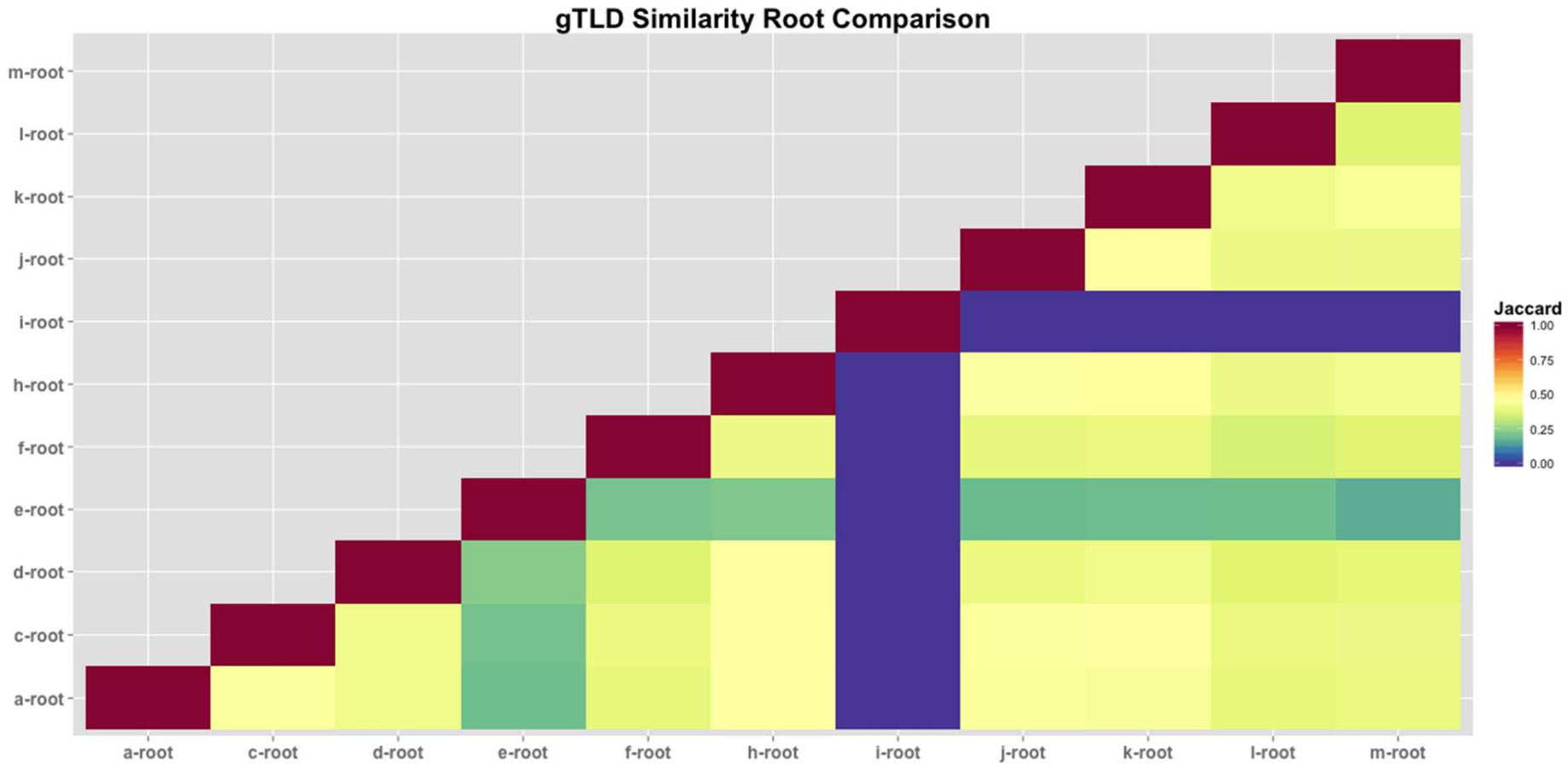
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \qquad 0 \leq J(A,B) \leq 1.$$

- Similarity matrix is a matrix of scores that represent the similarity between a number of data points.

DITL – Intra-Root Affinity :: SLDs



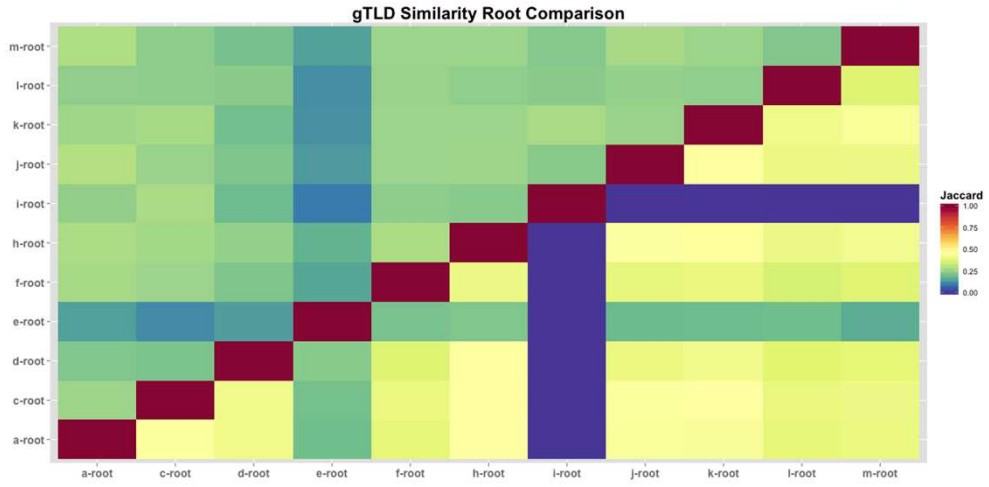
DITL – Intra-Root Affinity :: /24 Networks



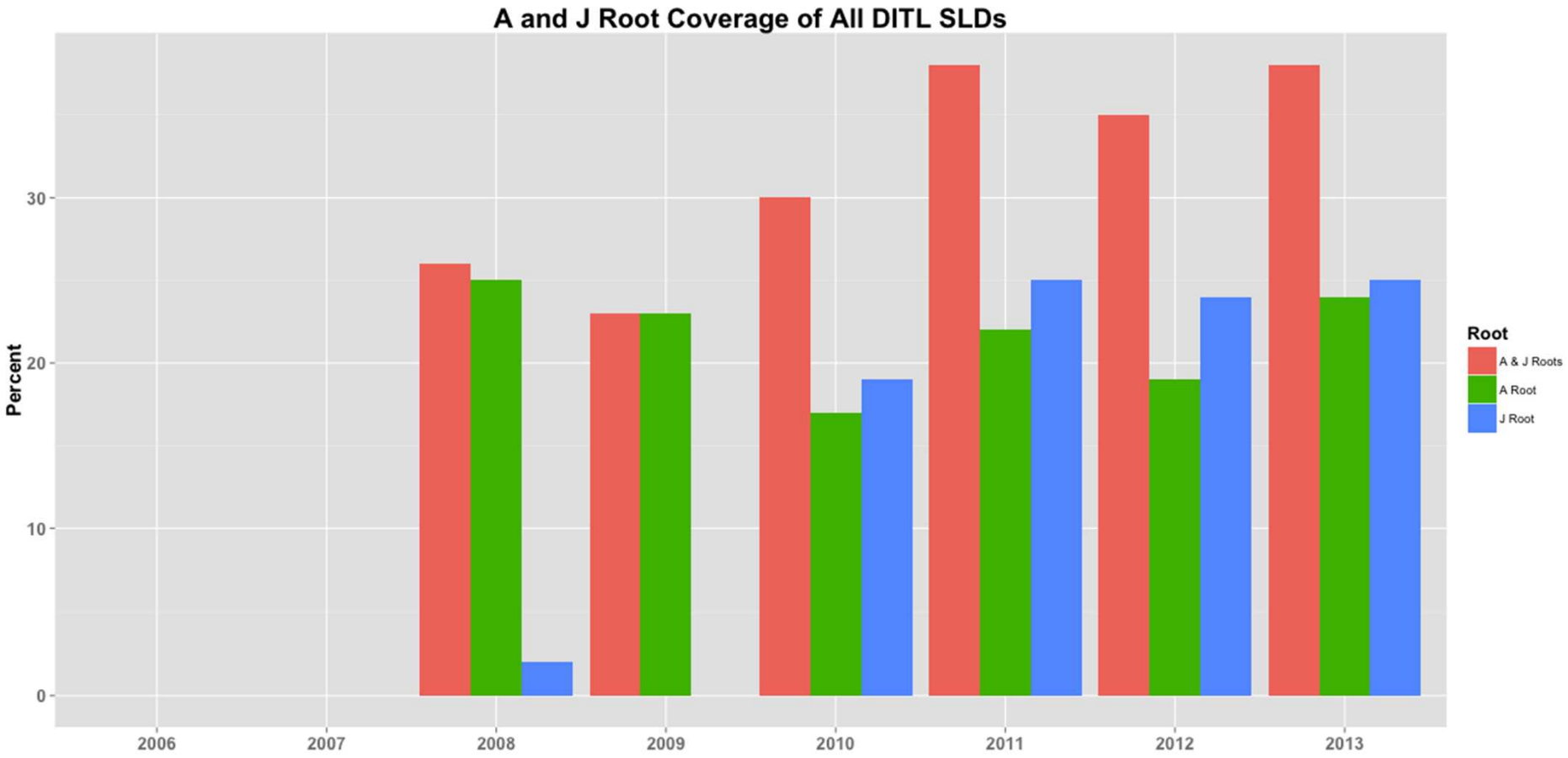
DITL – Intra-Root Affinity

- No inter-root affinity for either specific SLDs or recursive name server traffic.

How representative are the A+J roots of the root NXD traffic overall?



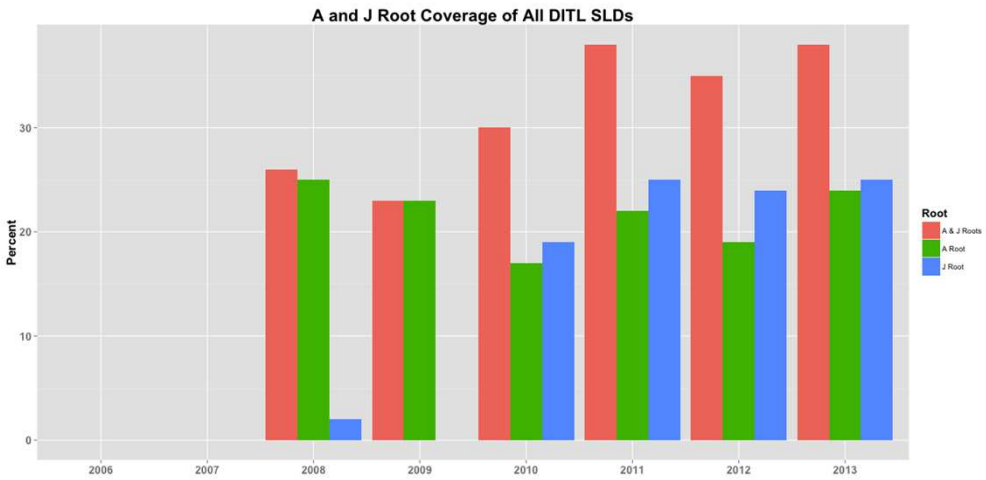
Longitudinal Inspection Using A+J Roots



Longitudinal Inspection Using A+J Roots

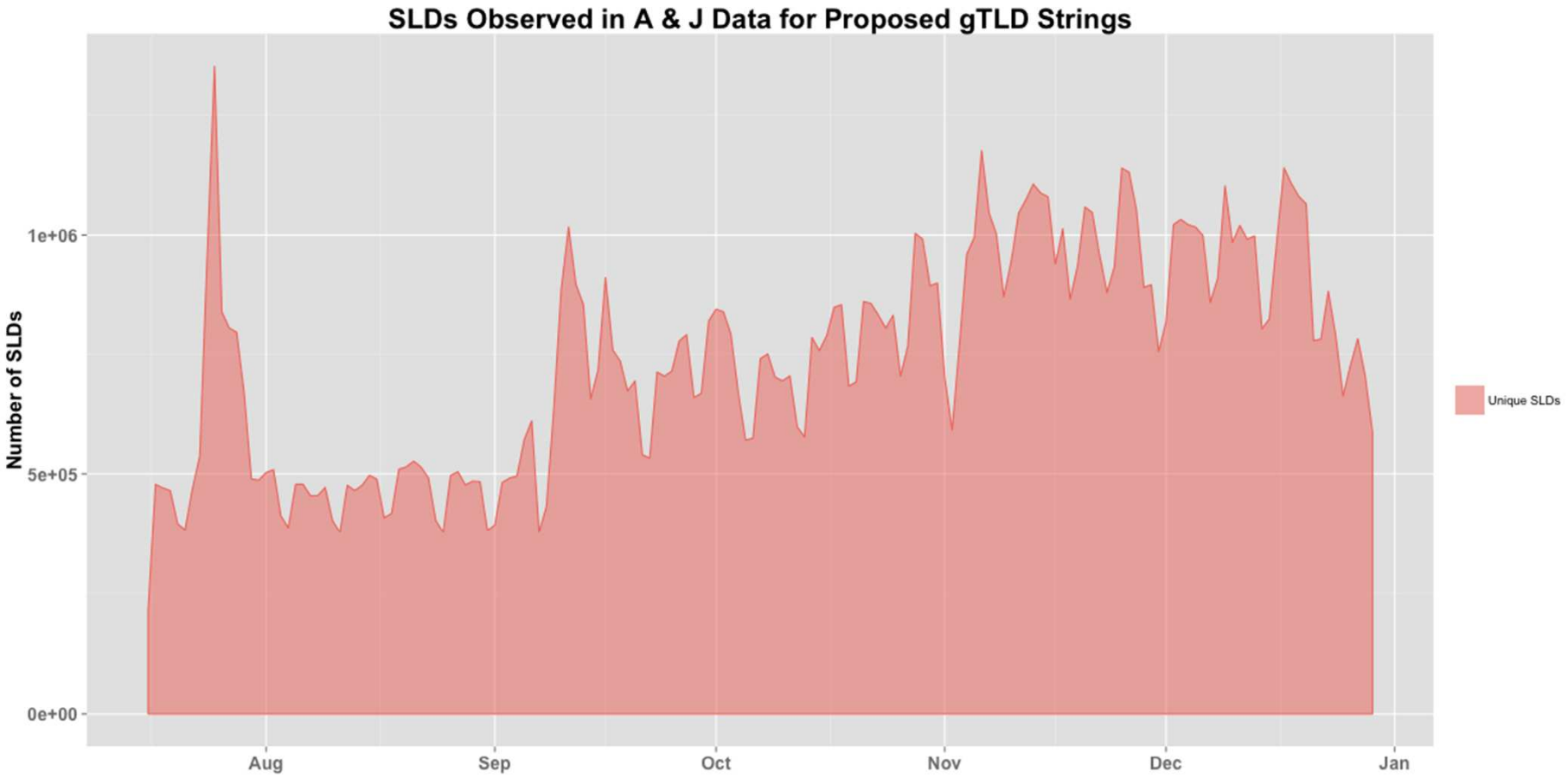
- On an annual basis, A+J combined observe just under 40% of all the SLDs observed across all roots
- Individually A and J each observe ~23% of all SLDs
- Corroborates intra-root affinity measures

What is the SLD growth rate of Observed and Previously Observed SLDs over a longitudinal period?

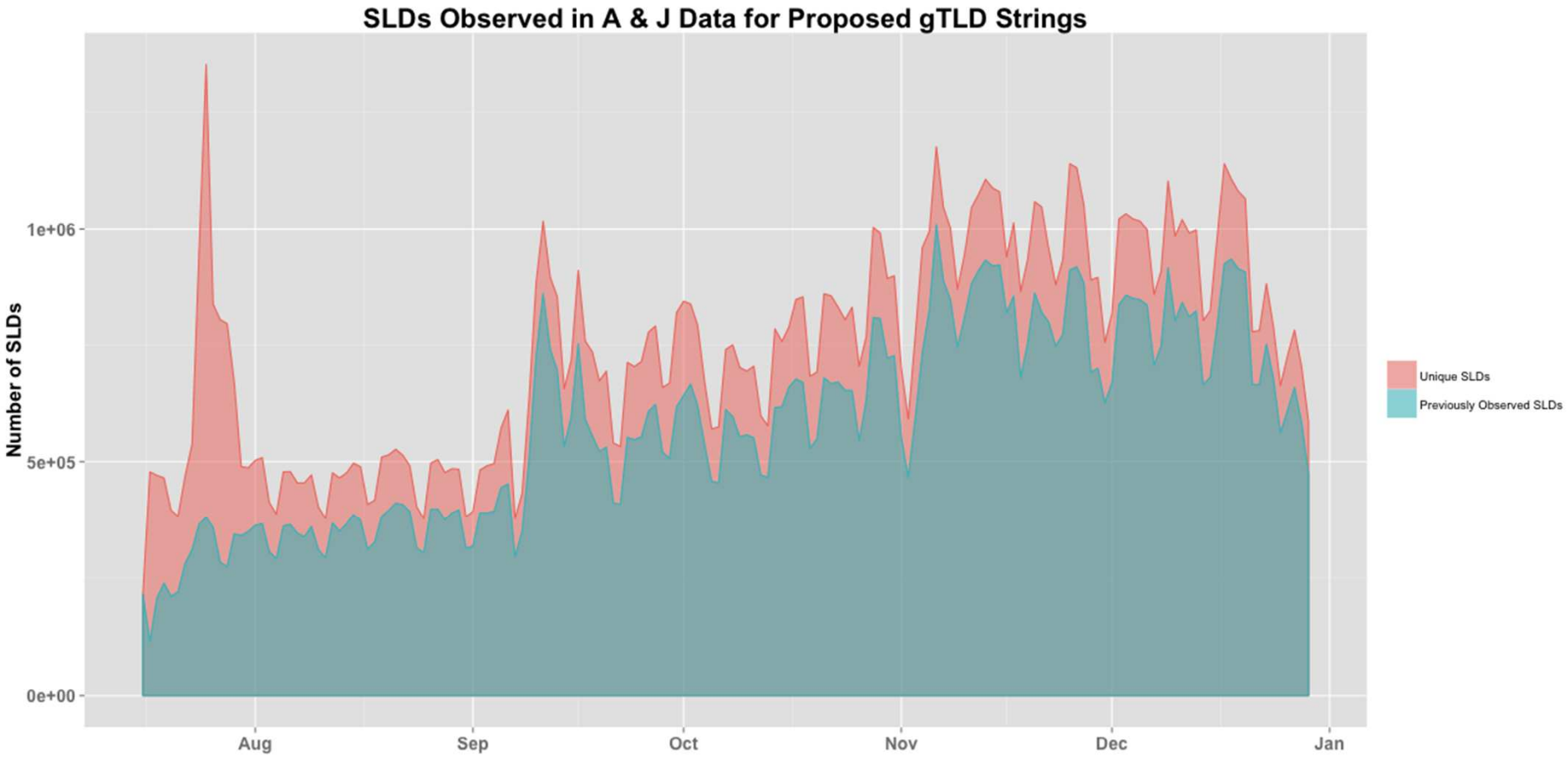


A+J Root Measurements

Longitudinal Inspection Using A+J Roots



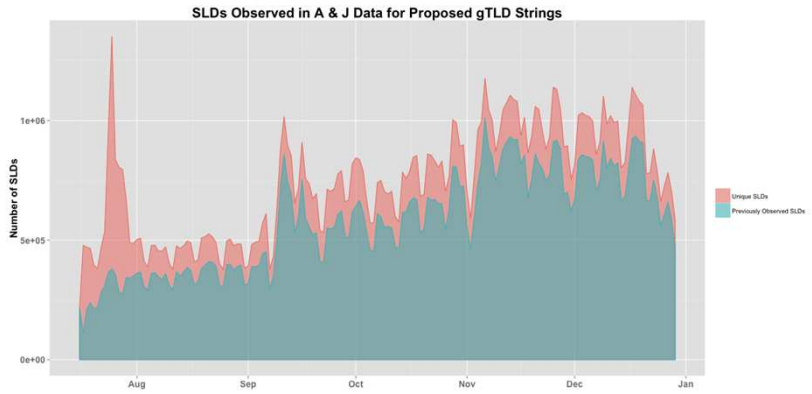
Longitudinal Inspection Using A+J Roots



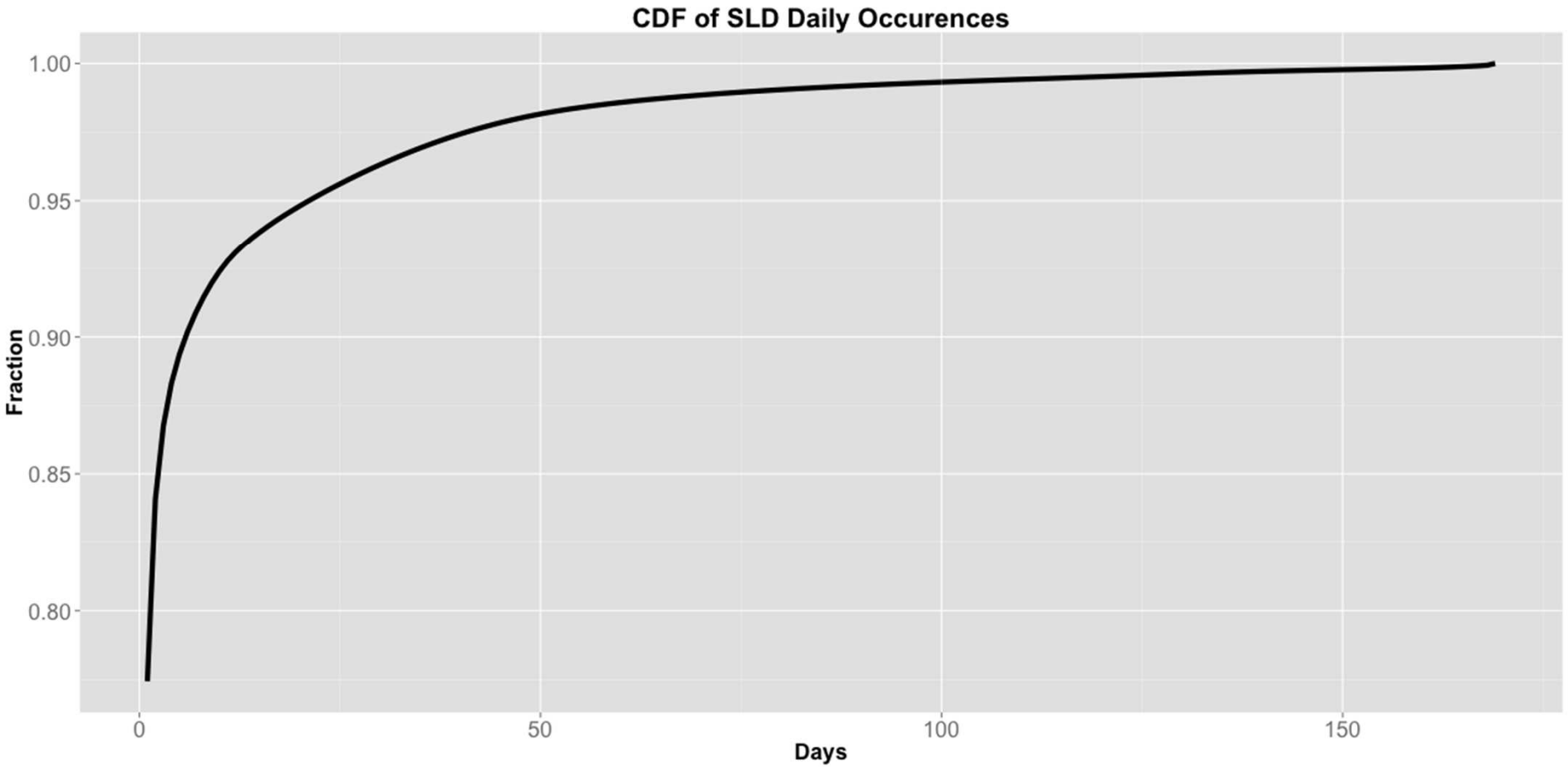
Longitudinal Inspection Using A+J Roots

- Average percentage of new SLDs on a given day is 22.5%
- Same trend seen in year over year DITL measurements.
- Highly entropic SLD universe: any small collection window will only account for a small percentage of SLDs over the subsequent period of time.
- Pattern is so consistent that any collection period will always have a large number of never seen before SLDs.

How frequently do SLDs occur?



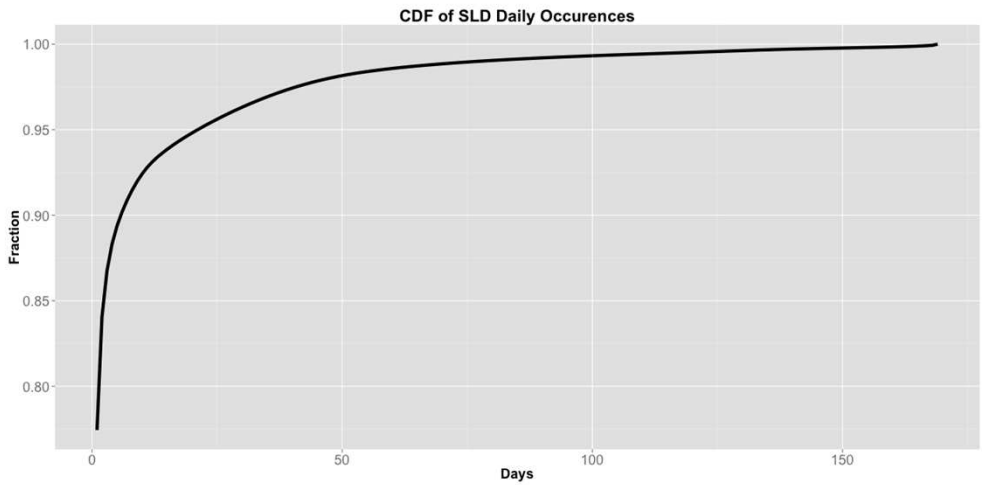
A+J SLD Daily Occurrence Frequencies



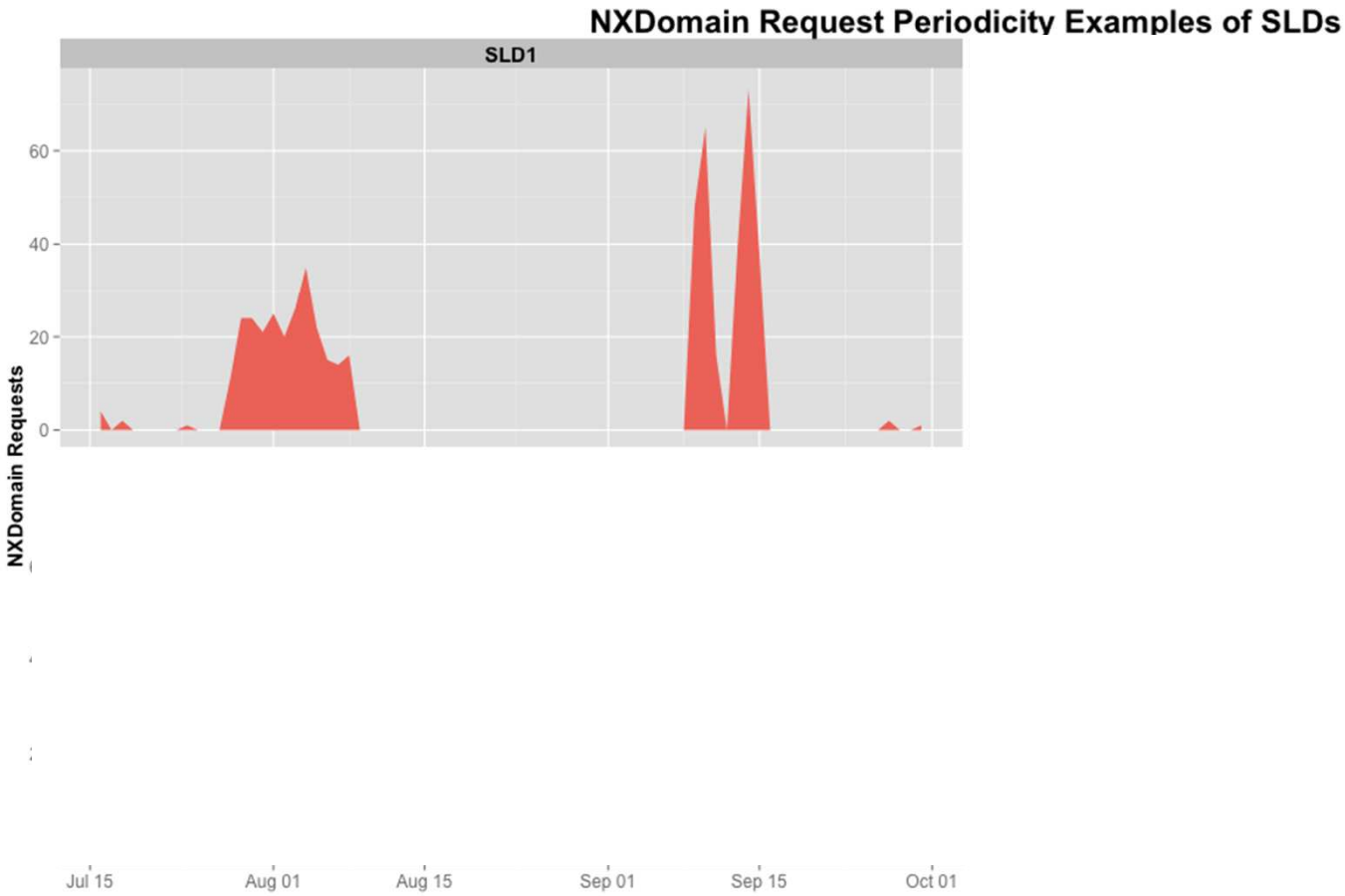
A+J SLD Daily Occurrence Frequencies

- Nearly 80% of the observed SLDs appear on only one day
- Only 5% of SLD's (~ 1.375 million) appeared on more than 20 days

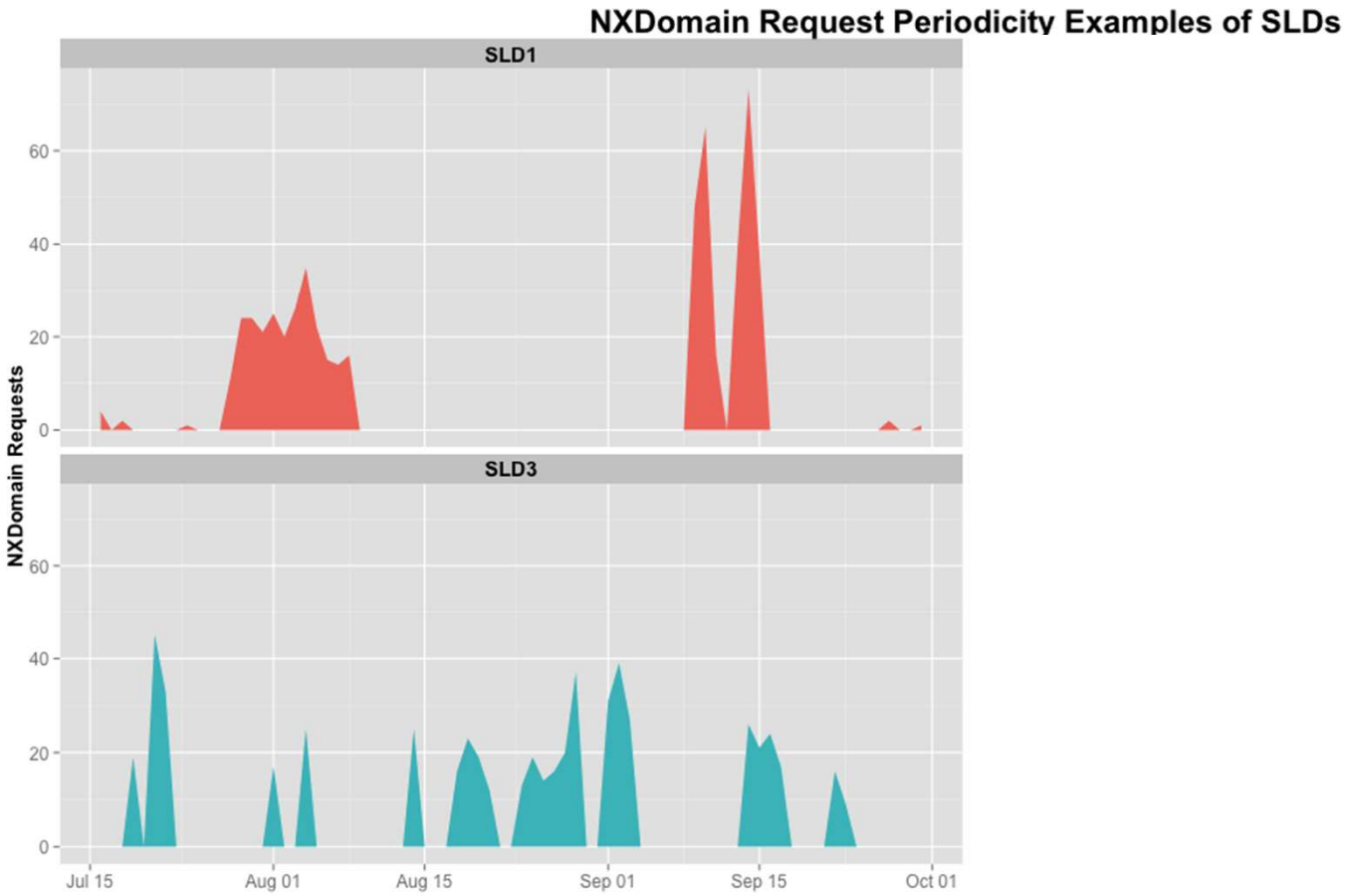
What temporal patterns do non-singleton SLDs exhibit?



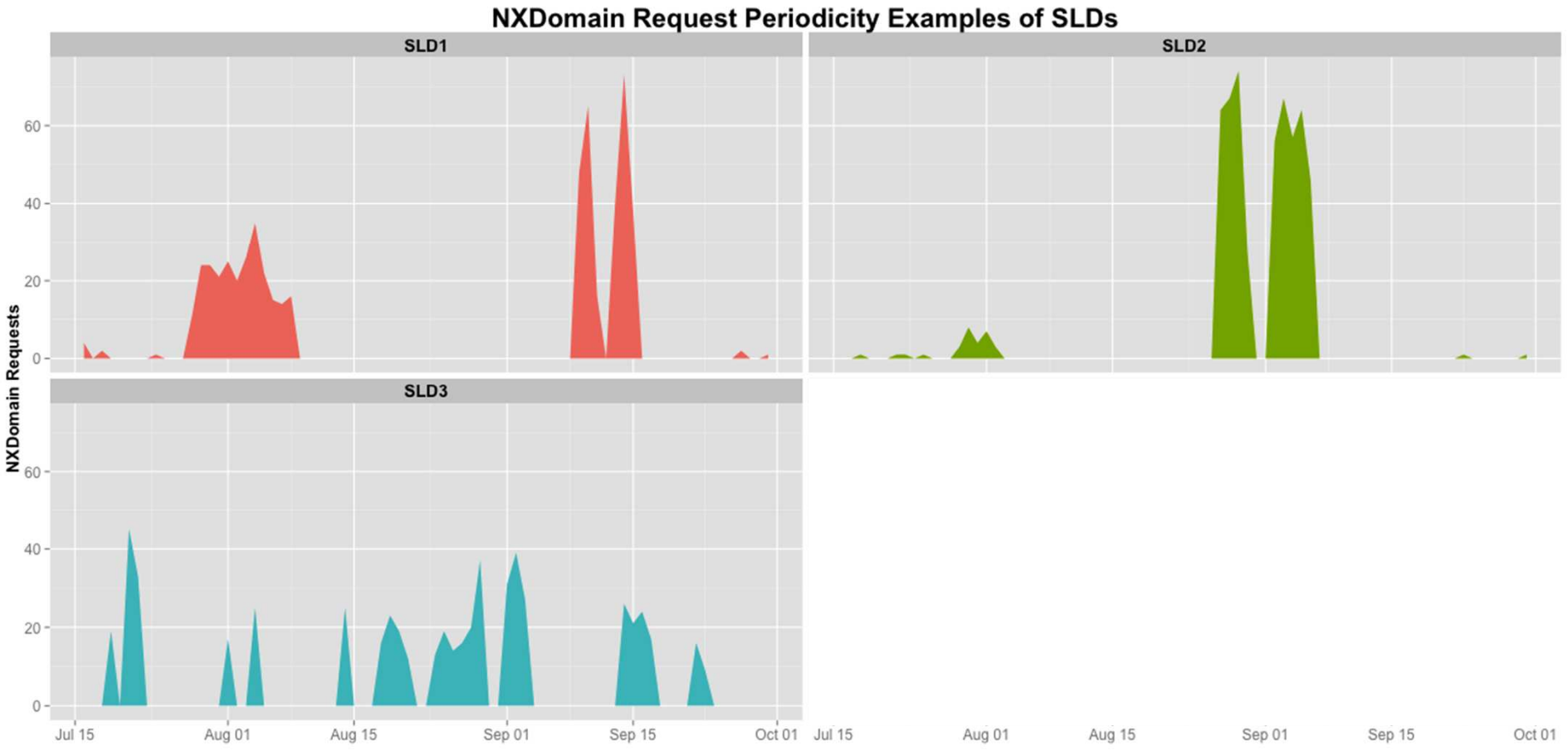
A+J SLD Periodicity



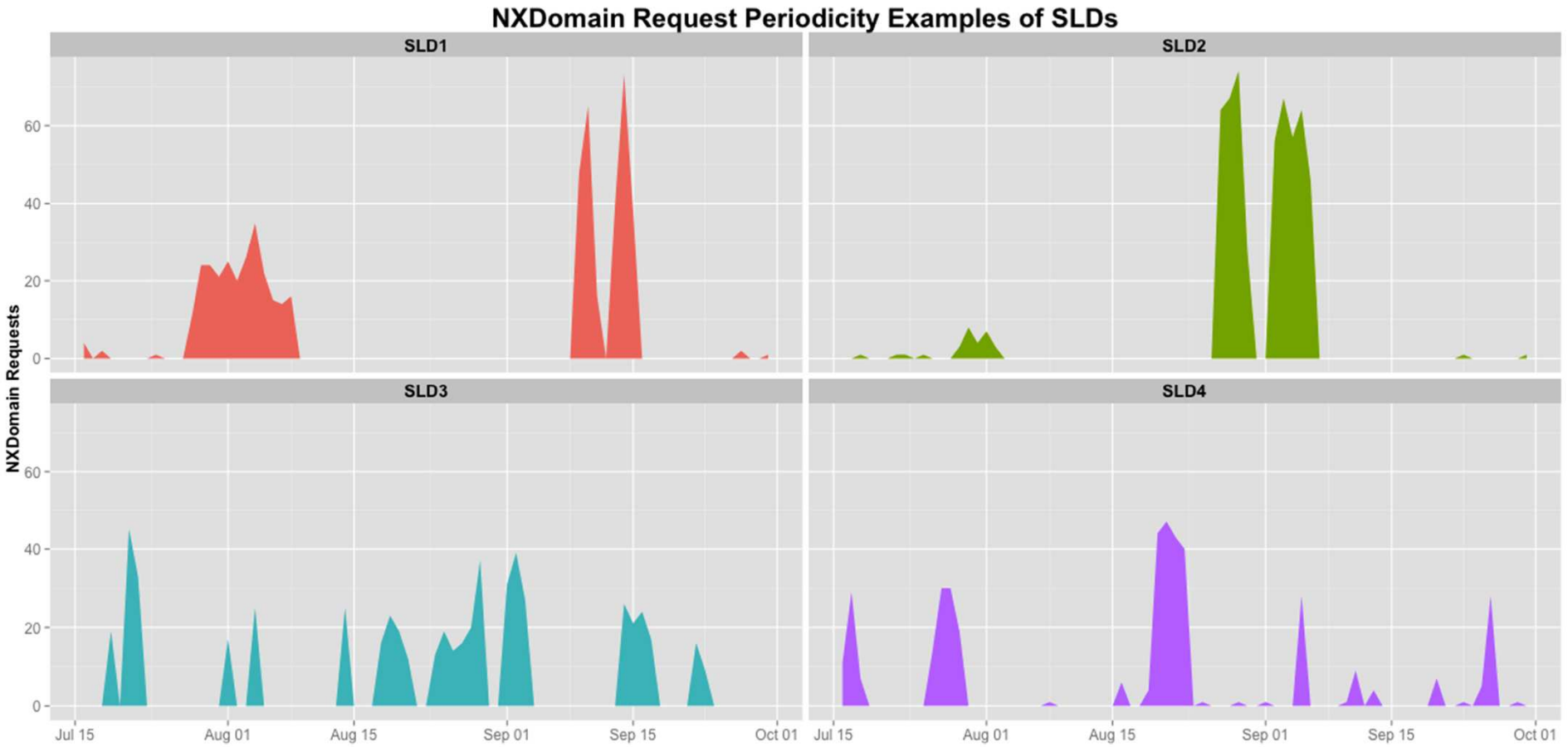
A+J SLD Periodicity



A+J SLD Periodicity



A+J SLD Periodicity



A+J SLD Periodicity

- Given a sequence of NXD requests for a given SLD:

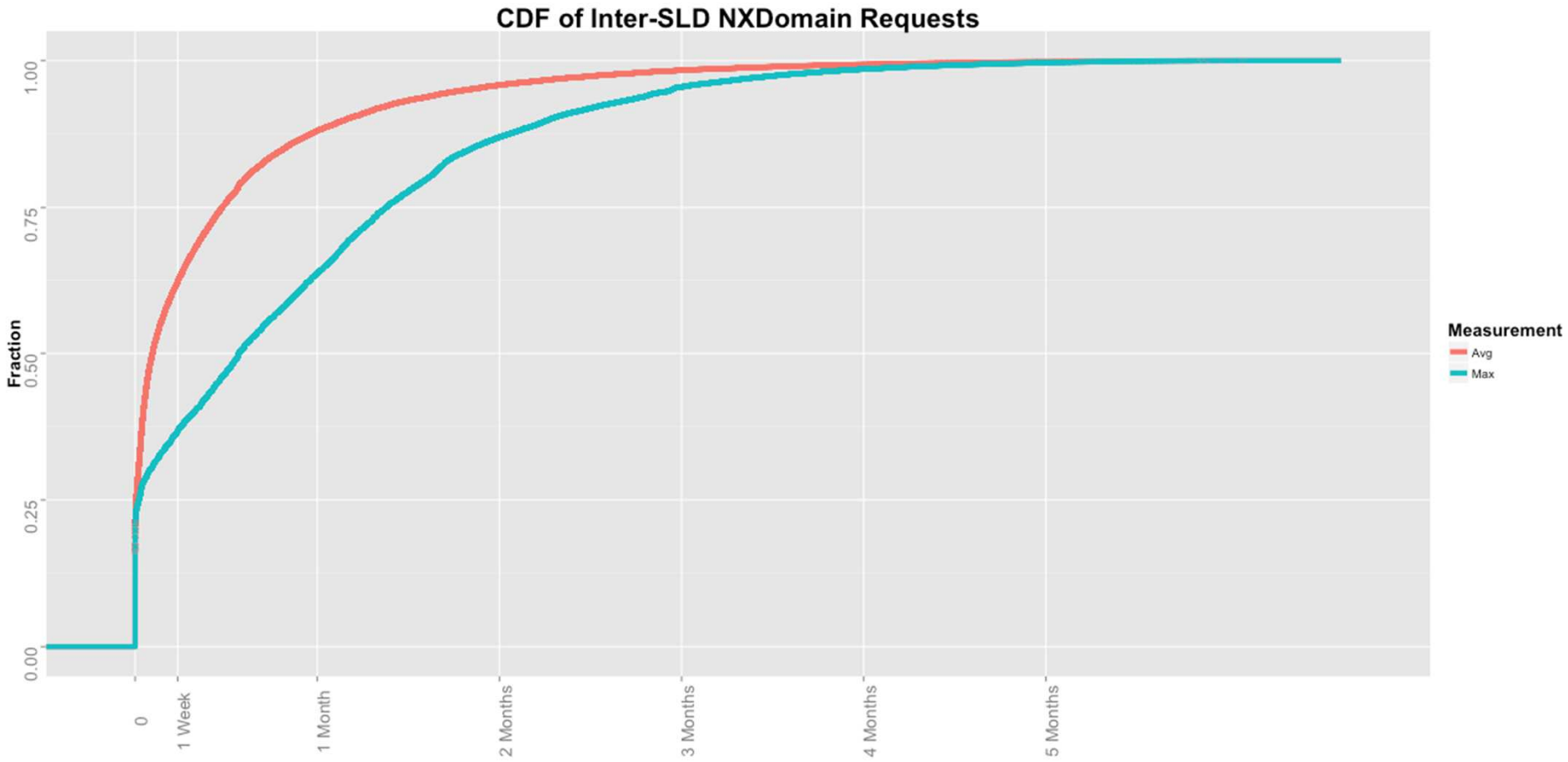
$$\Delta_{ki} = \tau_i(\varepsilon_k) - \tau_{i-1}(\varepsilon_k)$$

$$\mu_k = \frac{\sum_{i=1}^{n} \Delta_{ki}}{n}$$

ε_k : measured domain
τ_i : time of measured request
τ_{i-1} : time of last measured request

- Alternatively, we may look for the maximum value in the distribution to better size our collection window.

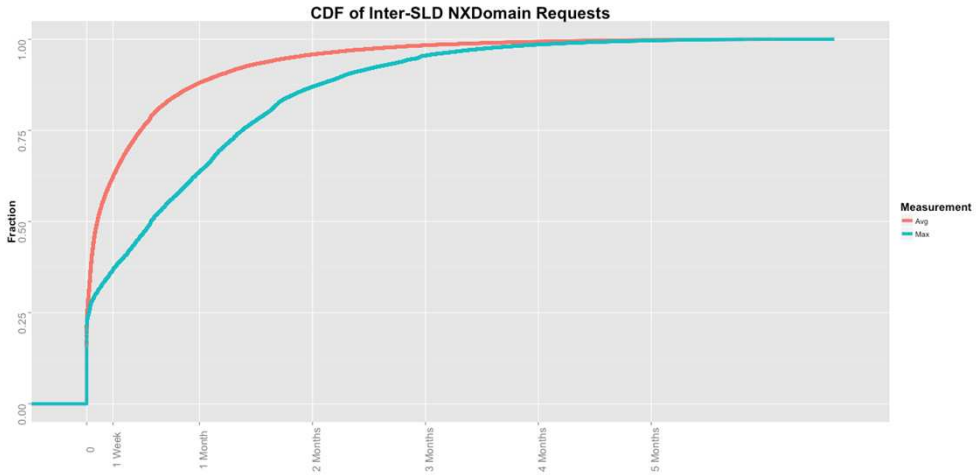
A+J SLD Periodicity



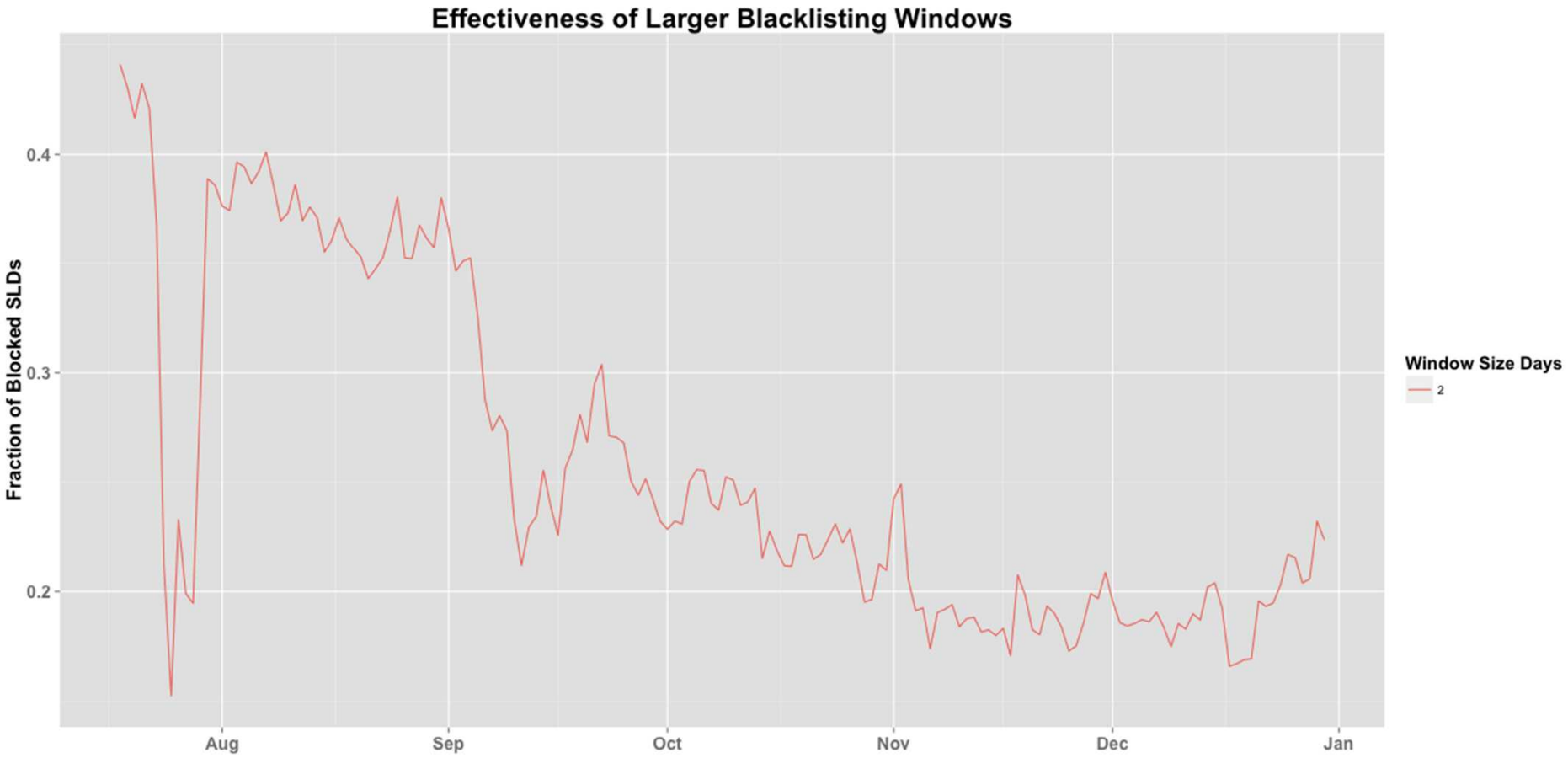
A+J SLD Periodicity

- Many SLDs exhibit some form of “burstiness”.
- 37% of domains exhibit average inter-query period of 1 week or longer.

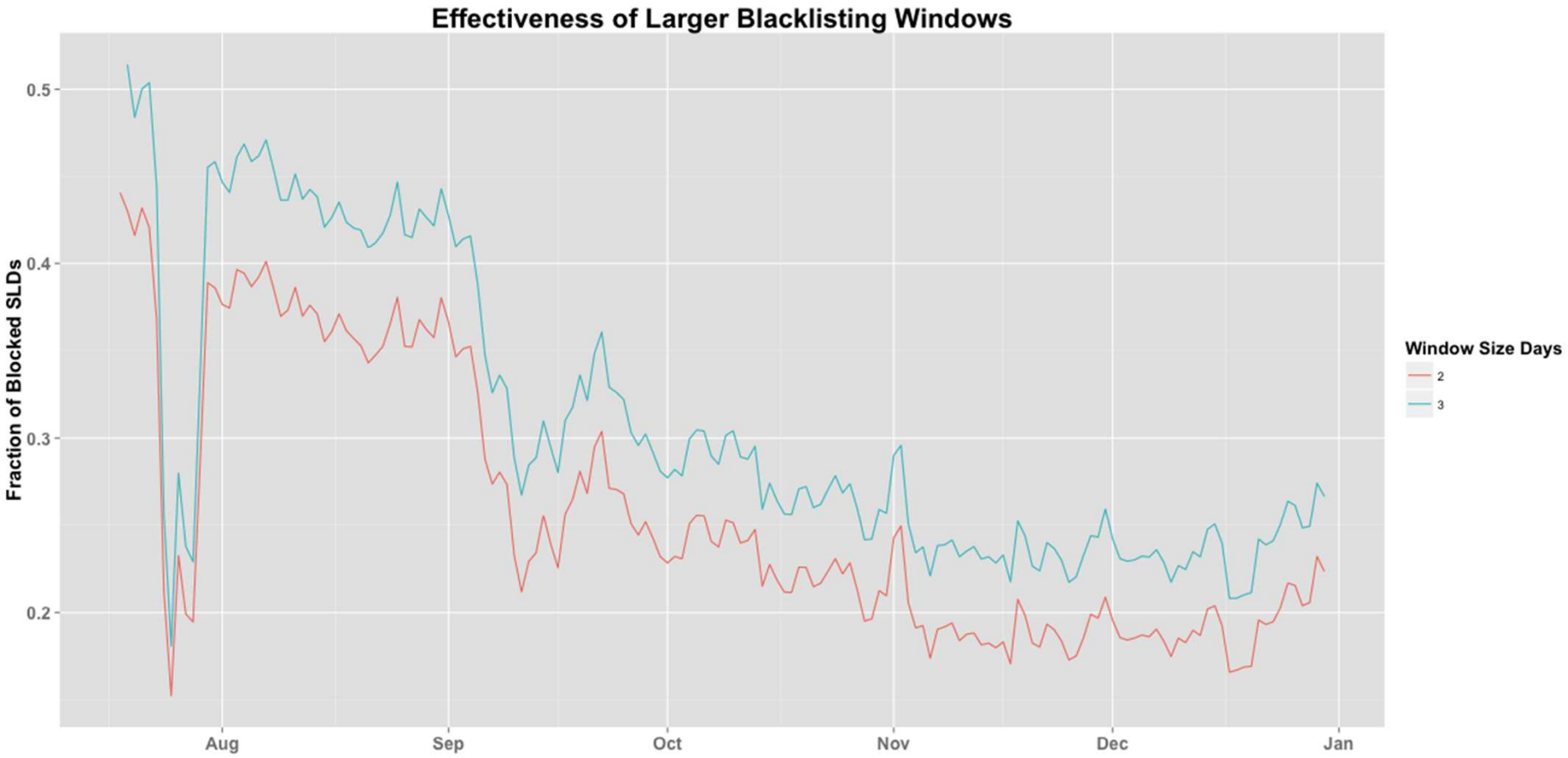
Do larger collection windows increase the efficacy of block listing?



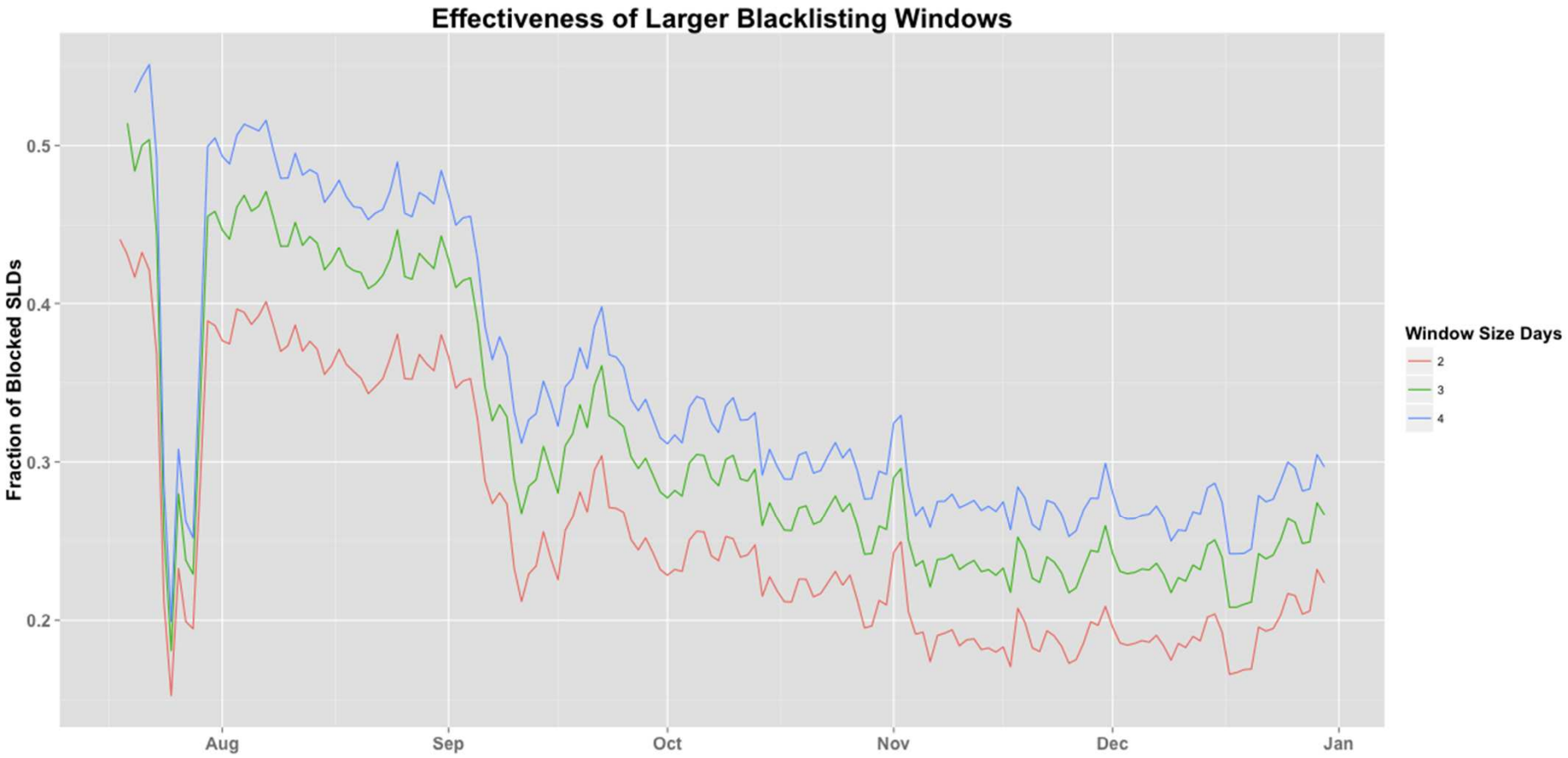
Effectiveness of Larger Block Listing Windows



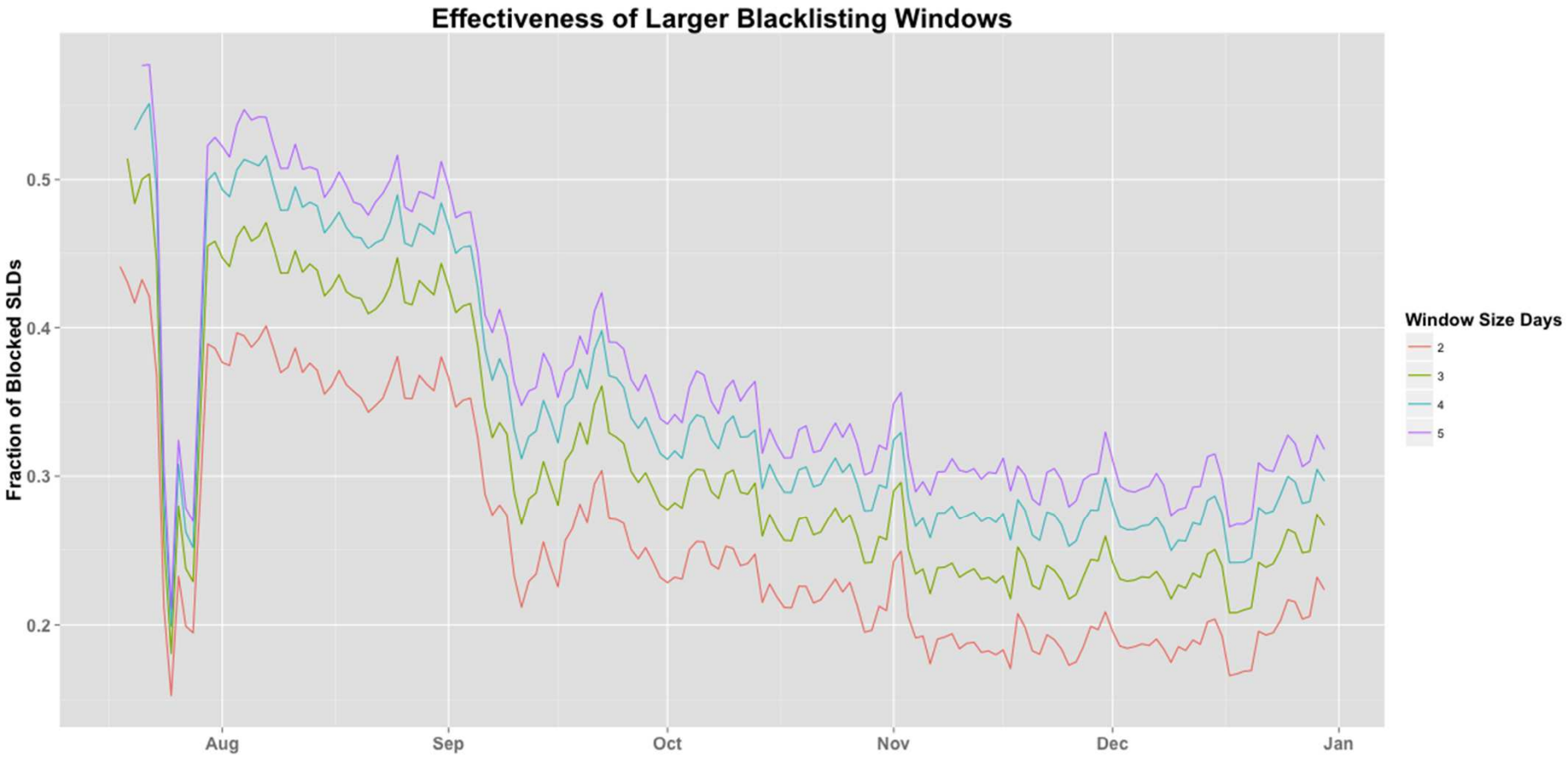
Effectiveness of Larger Block Listing Windows



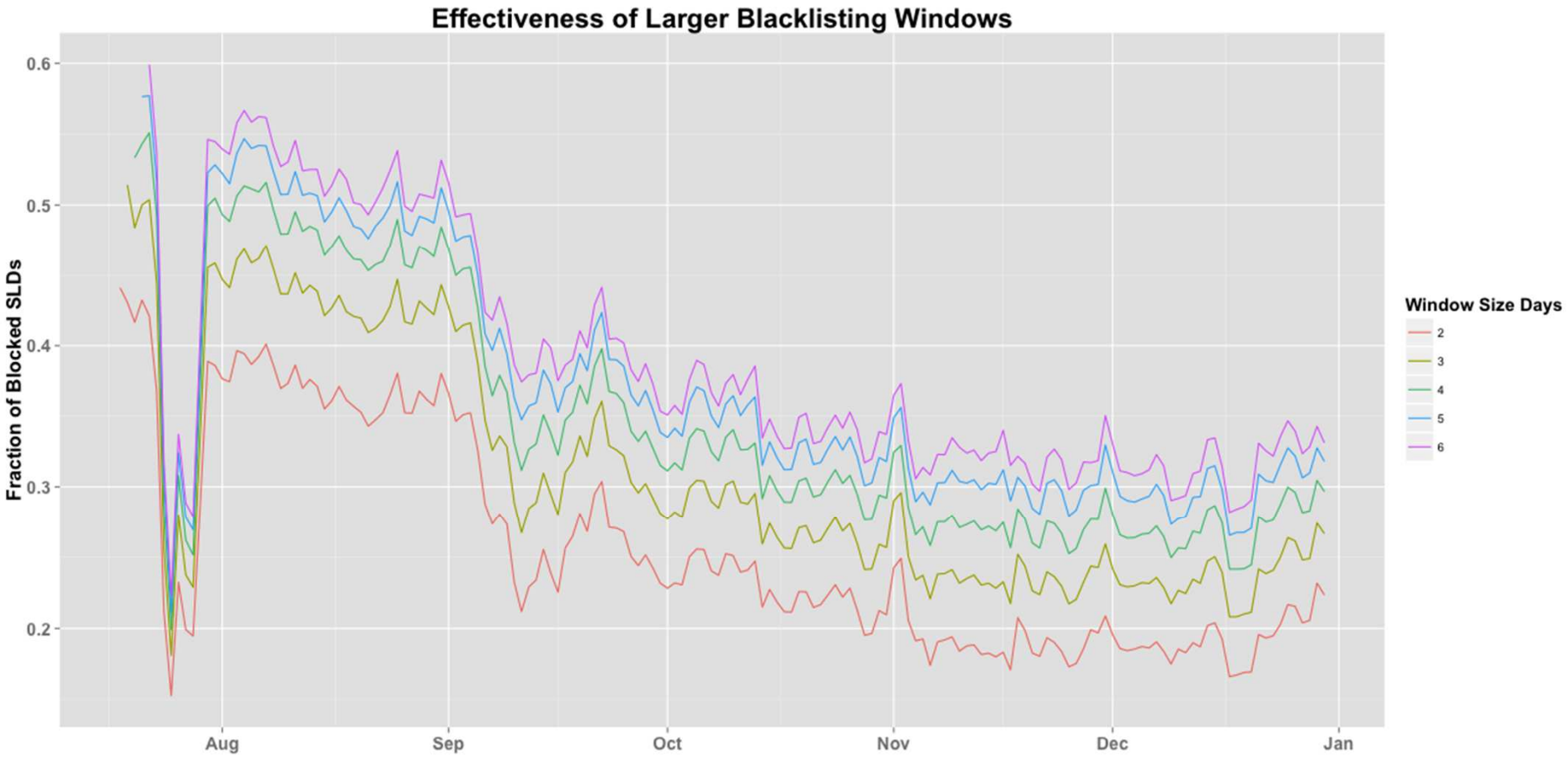
Effectiveness of Larger Block Listing Windows



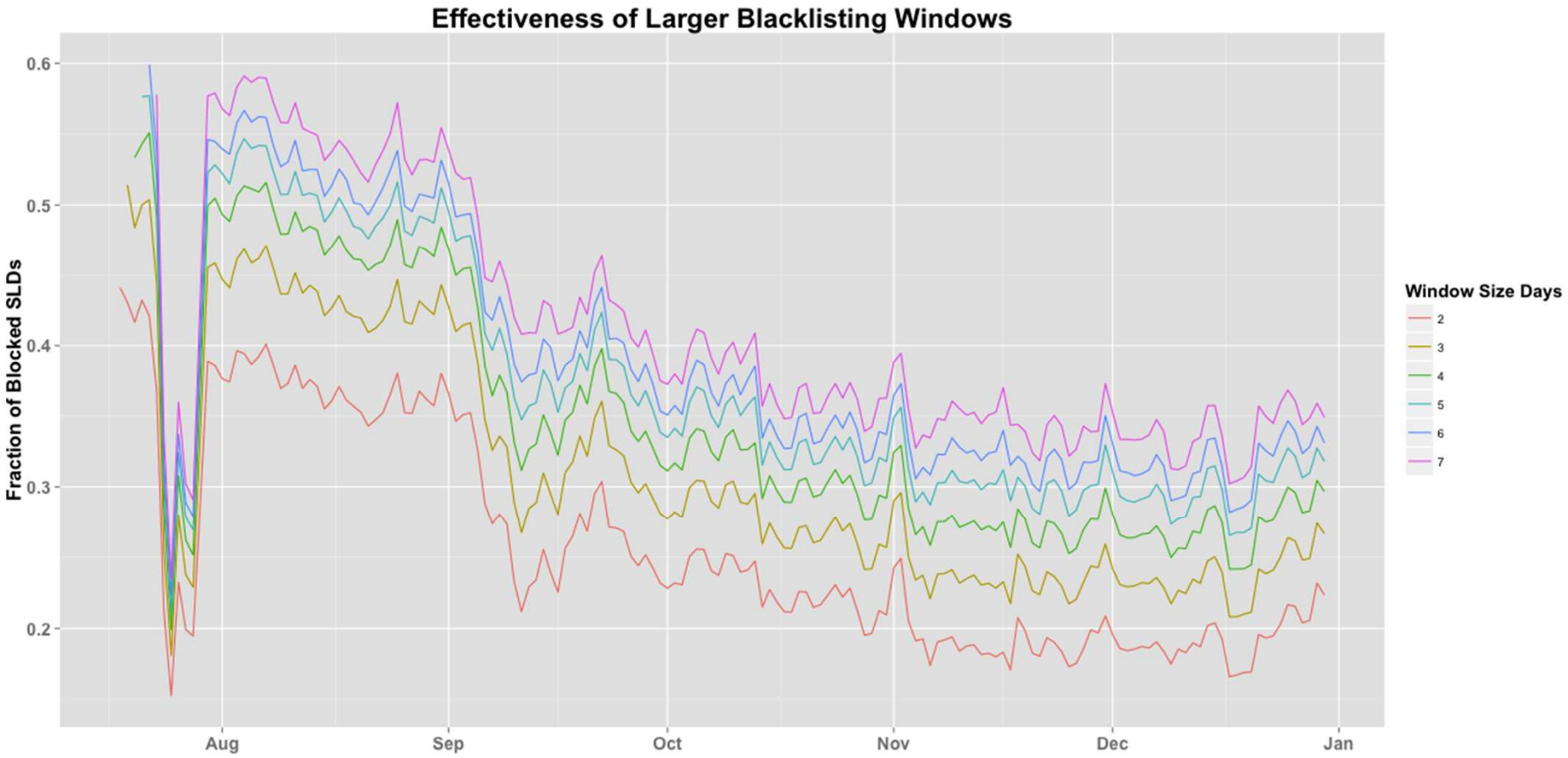
Effectiveness of Larger Block Listing Windows



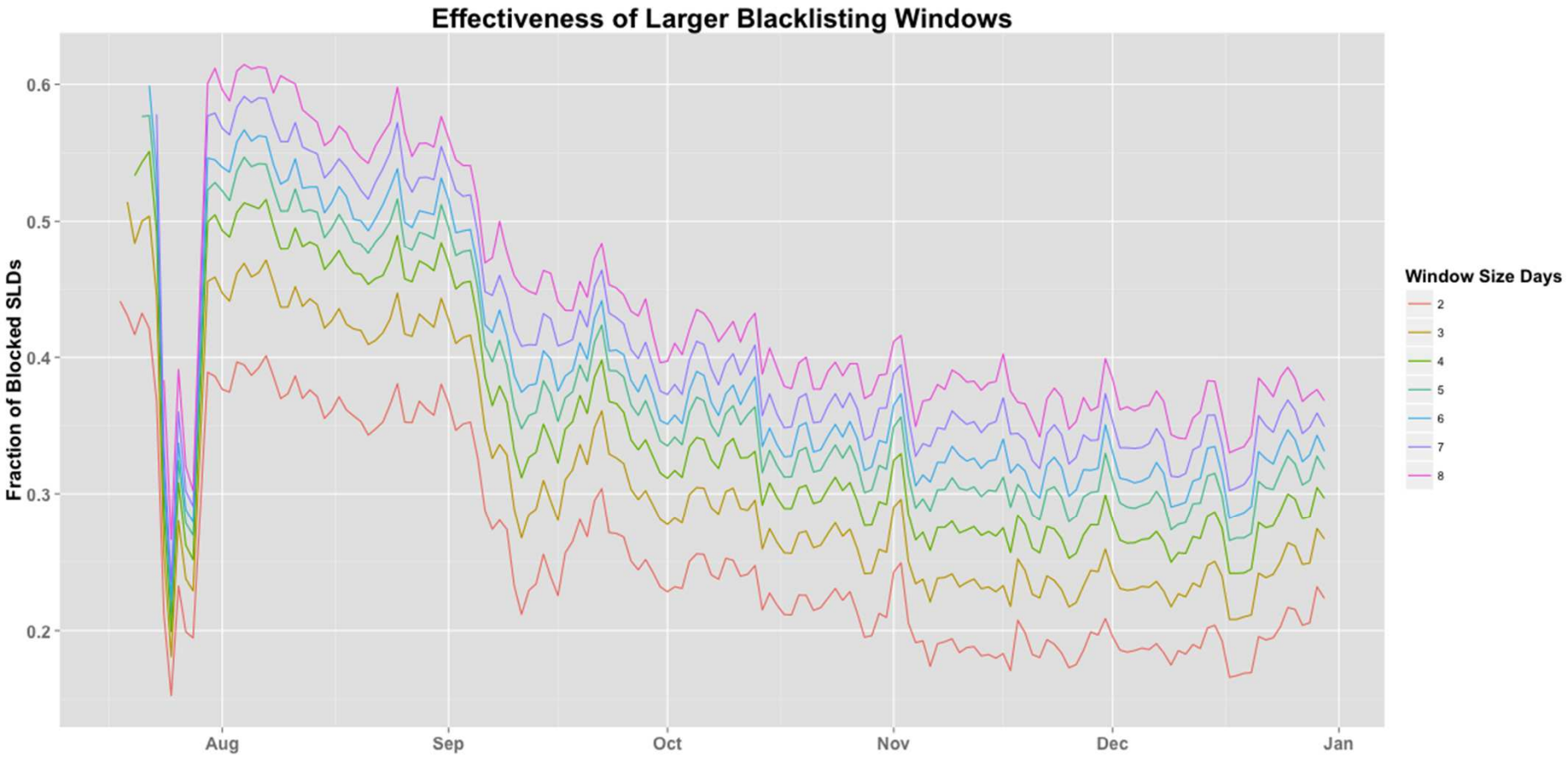
Effectiveness of Larger Block Listing Windows



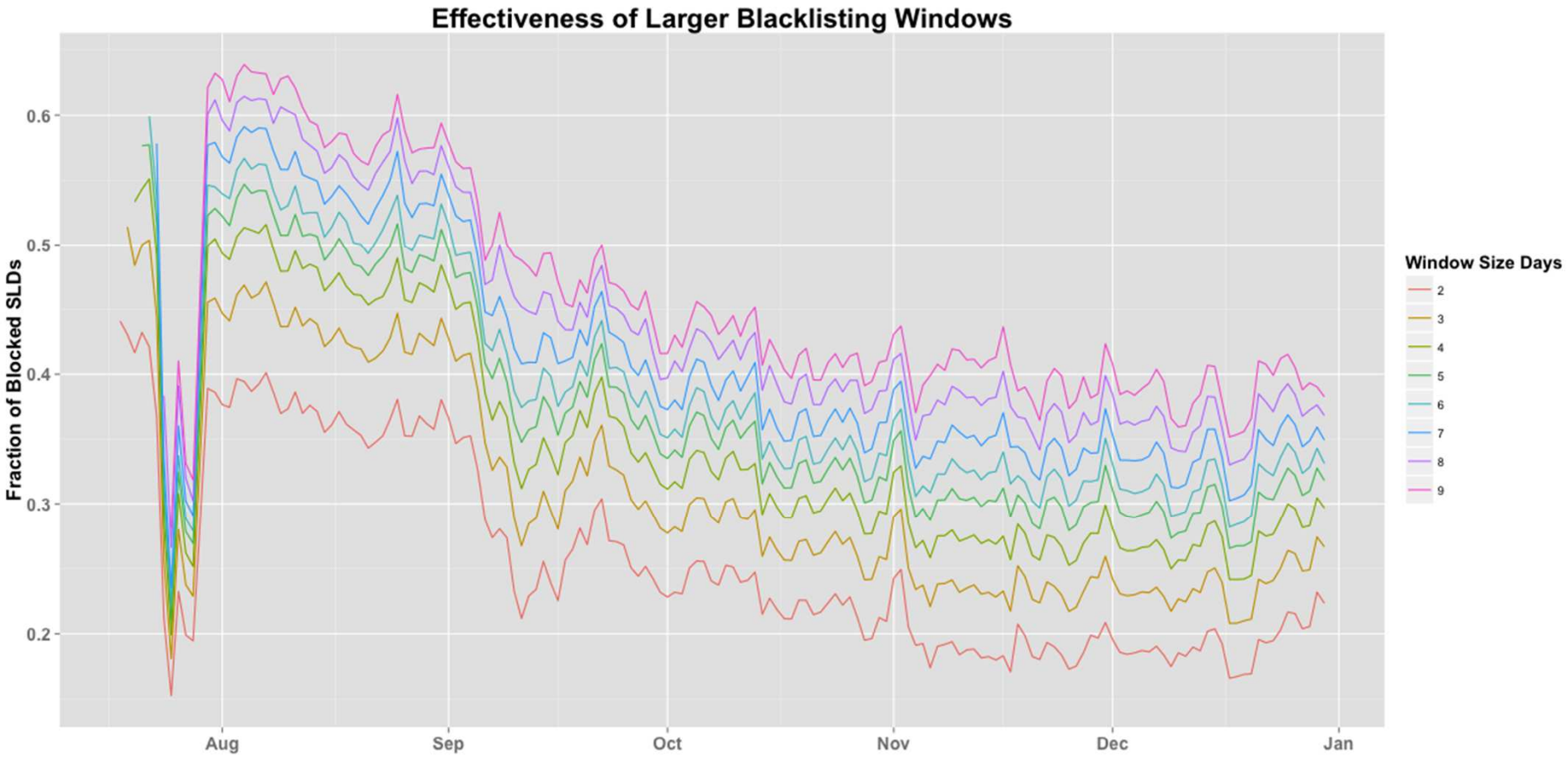
Effectiveness of Larger Block Listing Windows



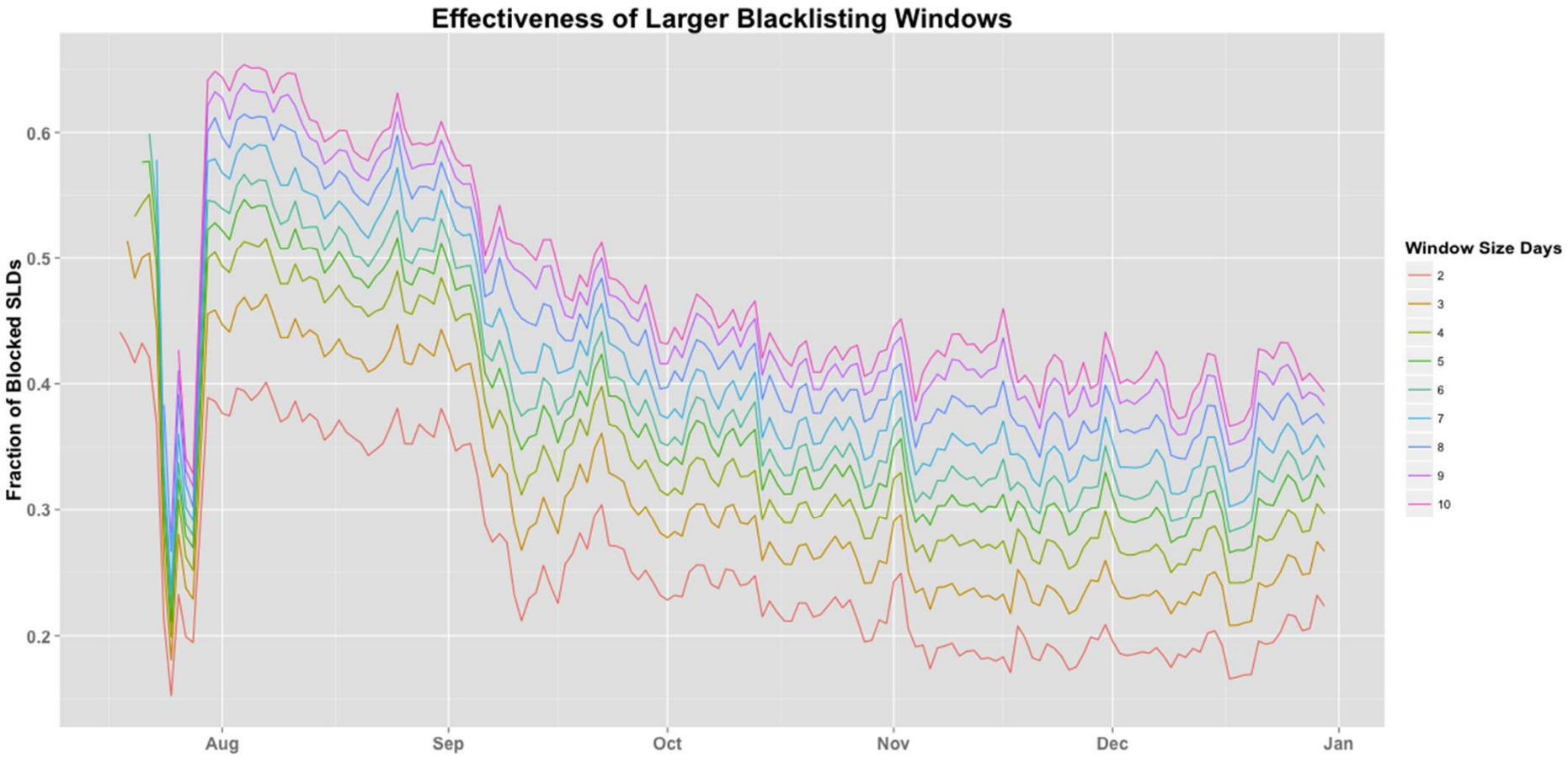
Effectiveness of Larger Block Listing Windows



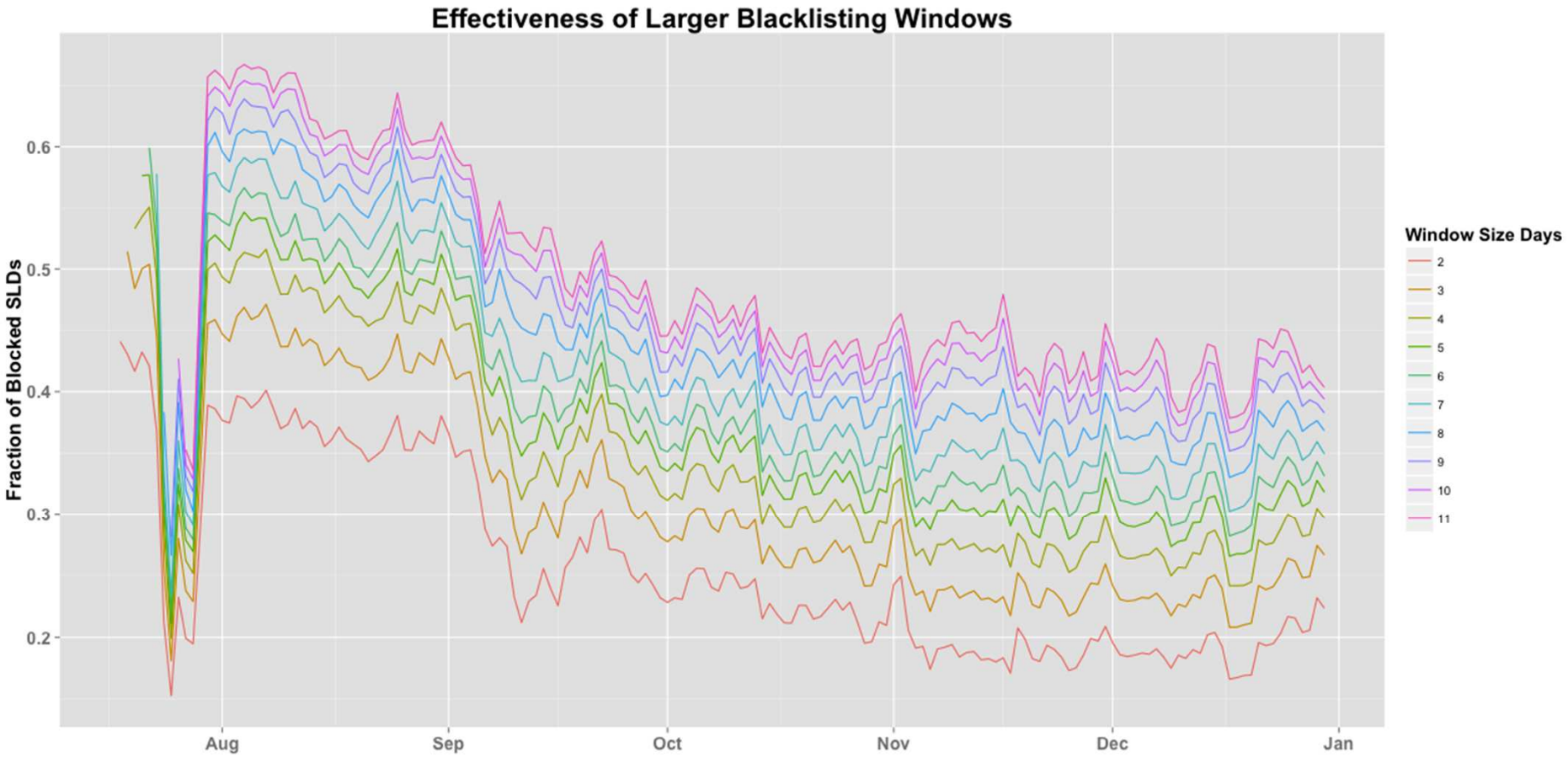
Effectiveness of Larger Block Listing Windows



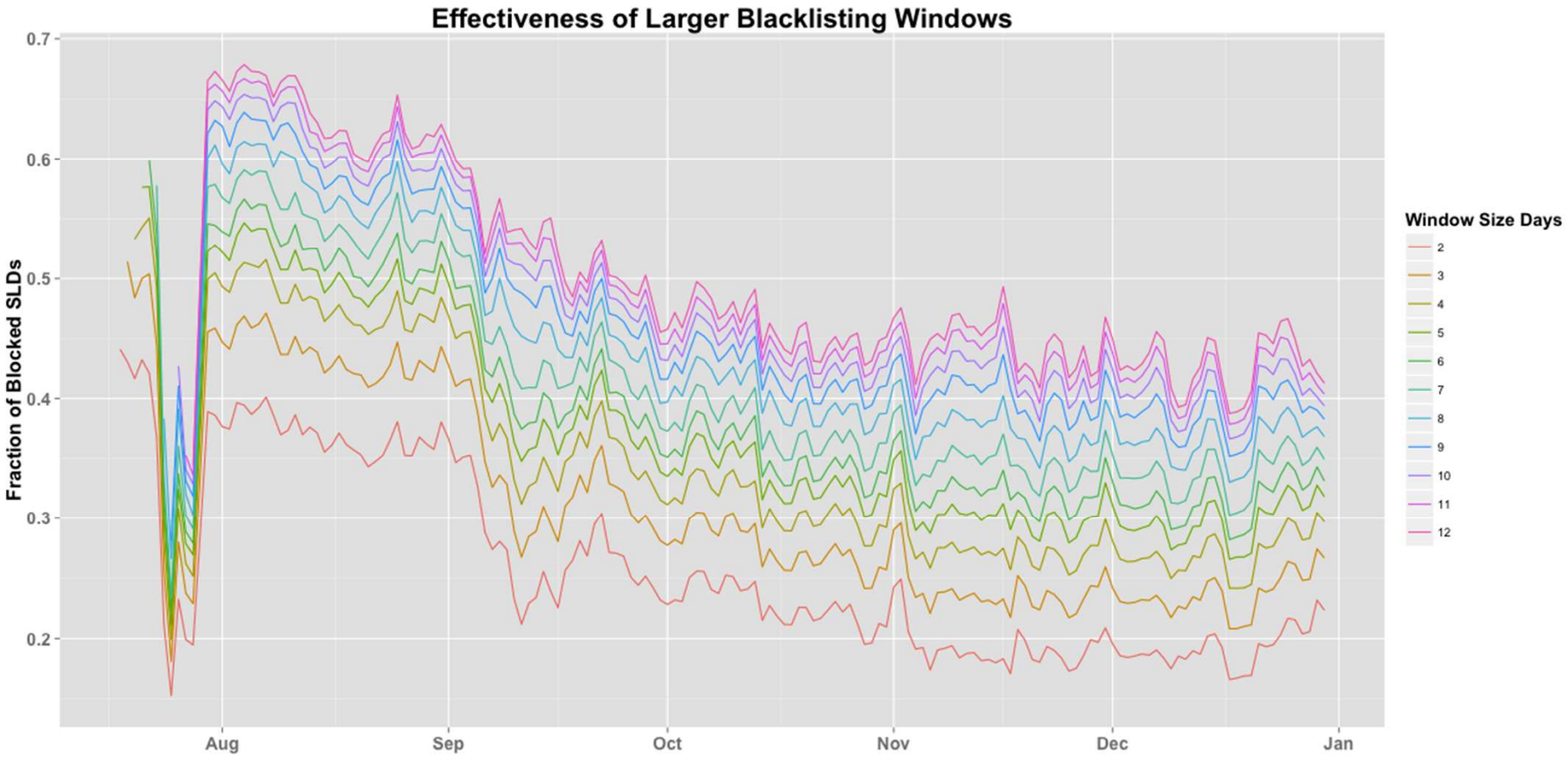
Effectiveness of Larger Block Listing Windows



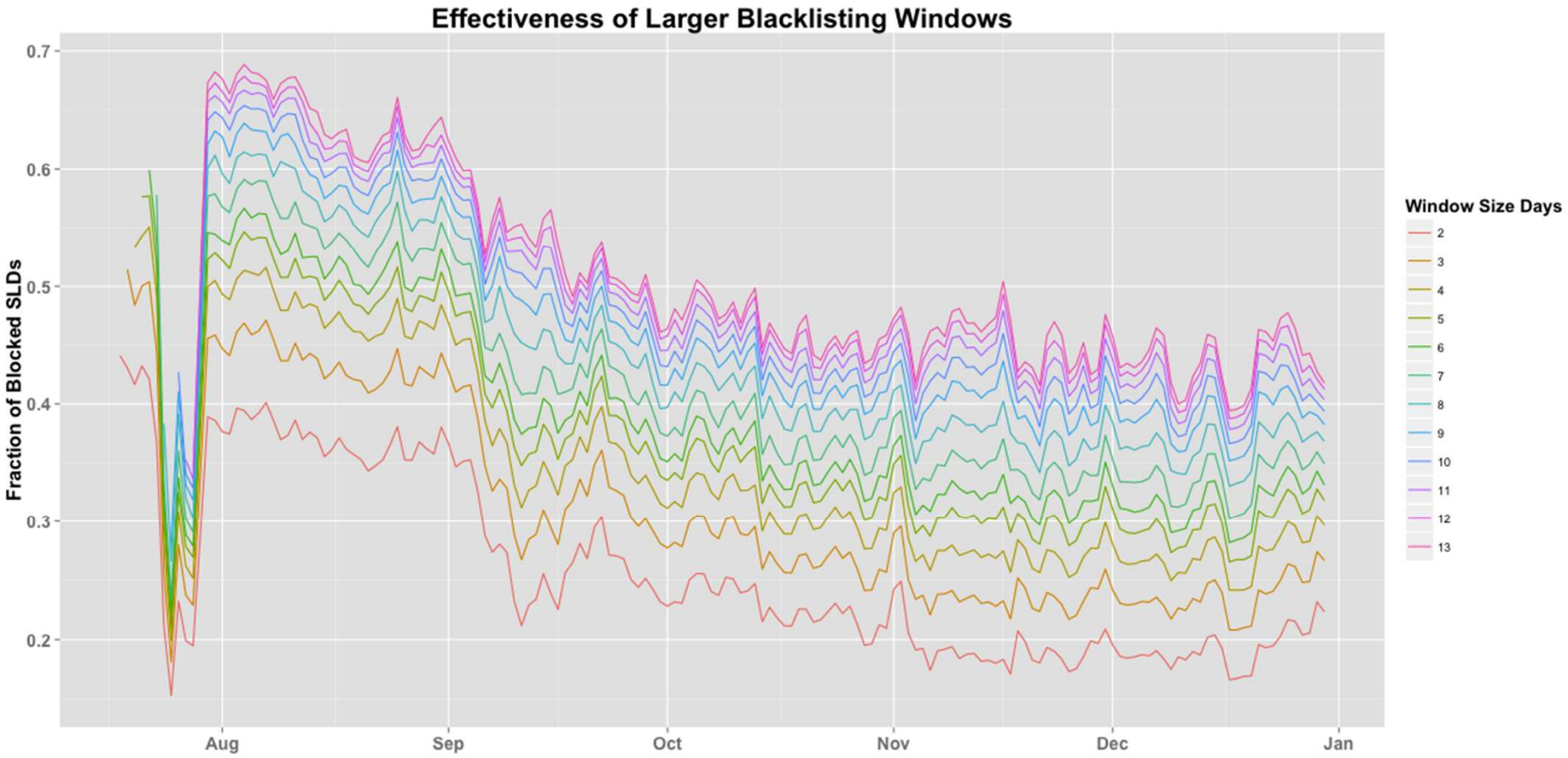
Effectiveness of Larger Block Listing Windows



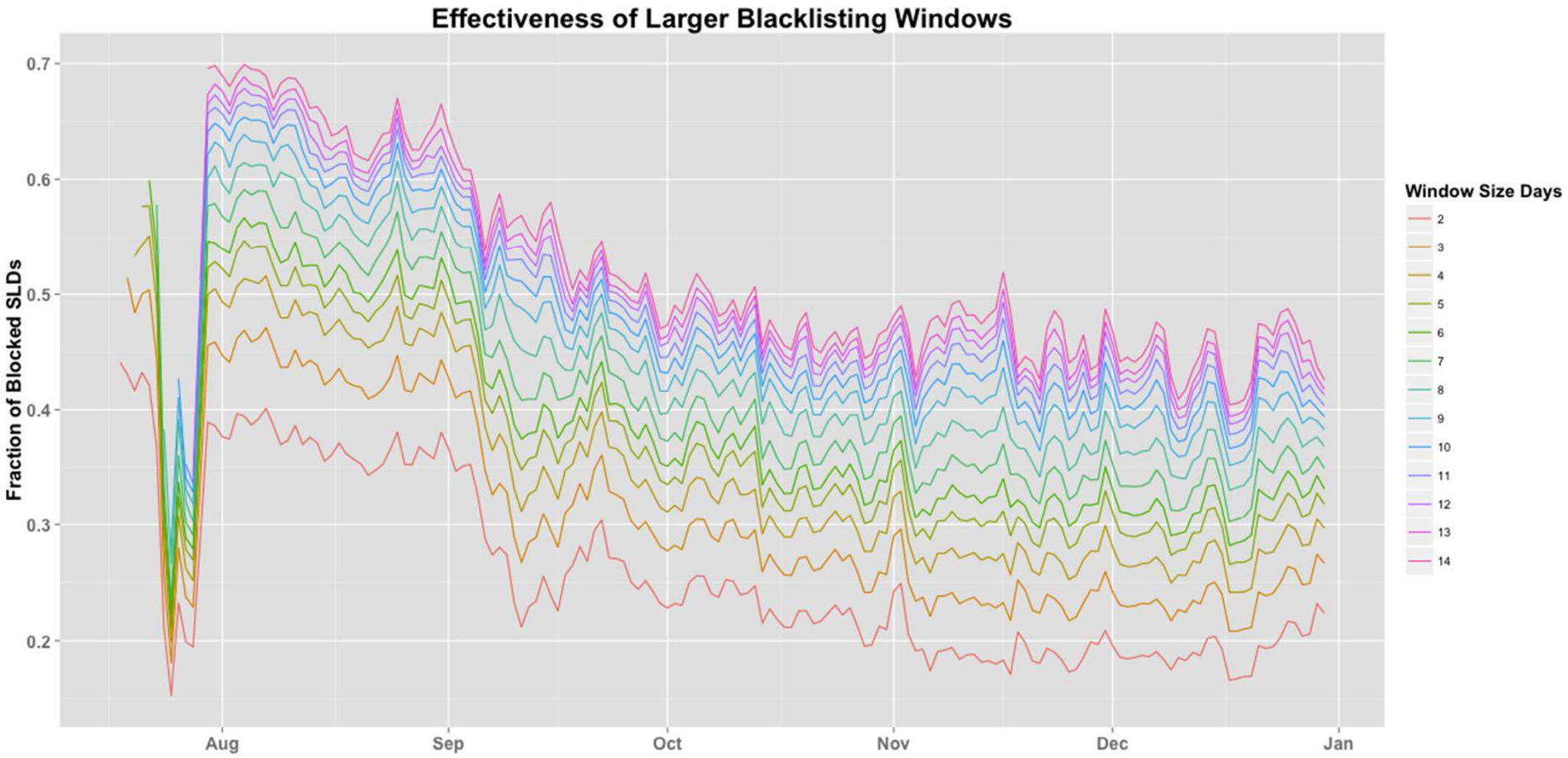
Effectiveness of Larger Block Listing Windows



Effectiveness of Larger Block Listing Windows

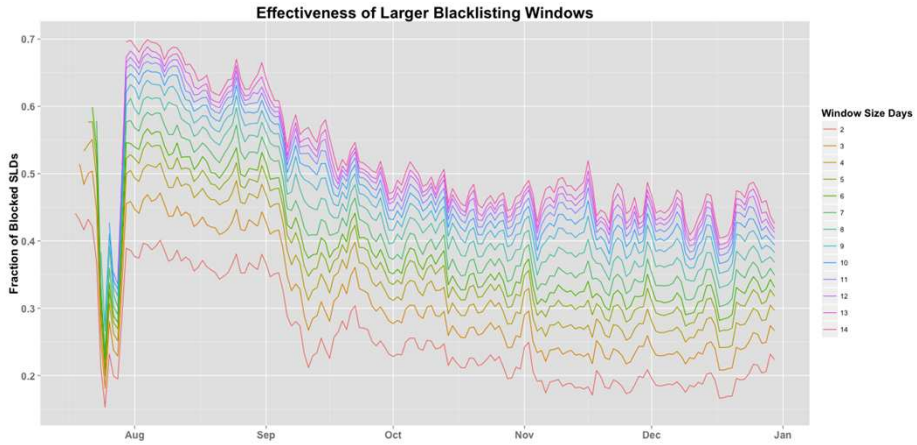


Effectiveness of Larger Block Listing Windows



Effectiveness of Larger Block Listing Windows

- With larger window sizes, the percentage of blocked SLDs increases but the effect of that increase asymptotically approaches an upper bound.
- For an given window size, the ratio diminishes with time – not unexpected due to highly entropic SLD universe.



Concluding Remarks

- Block Listing SLDs to prevent name collisions based on sampled DNS data appears to be an ineffective approach.
  - Highly dynamic and evolving SLD universe.
  - Strong SLD-root affinity will require all root data sets.
  - Temporal patterns exhibited by SLDs require longer observational windows, yet provide diminishing returns as time continues.
  - Alternative methodologies should be explored in conjunction or in place of DNS sampled data block listing.

powered by

VERISIGN™