Beyond Counting:
New Perspectives on the Active IPv4 Address Space

Philipp Richter
TU Berlin

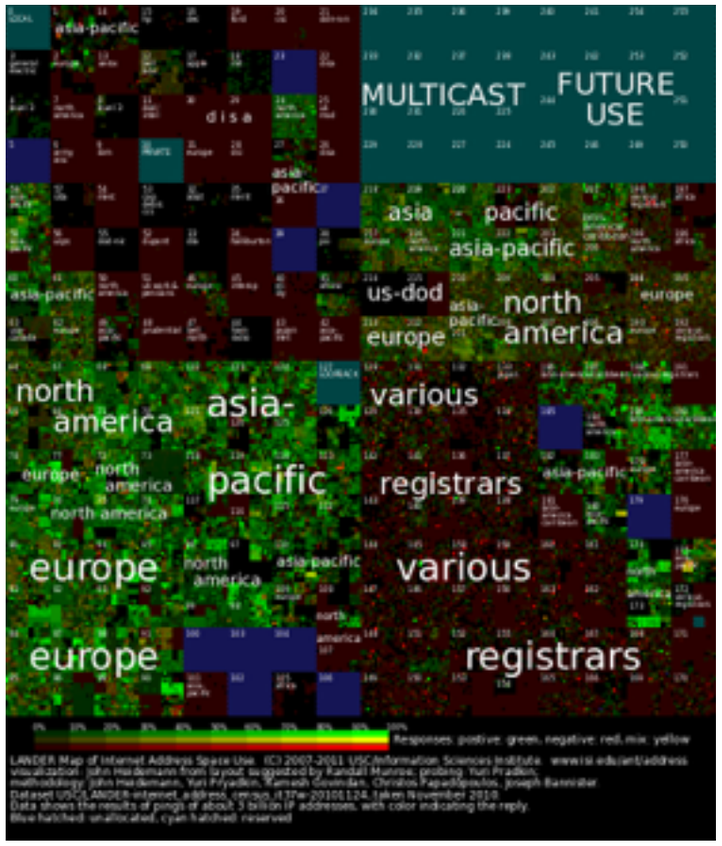
Georgios Smaragdakis
MIT

David Plonka
Akamai

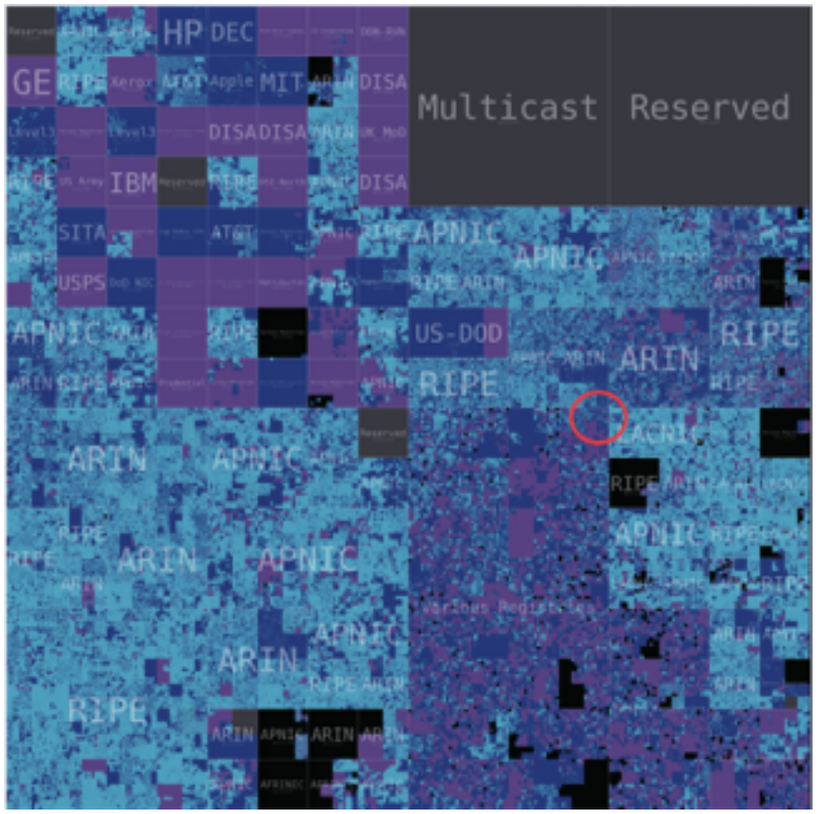
Arthur Berger
Akamai / MIT

ACM IMC 2016. Santa Monica, CA, USA.

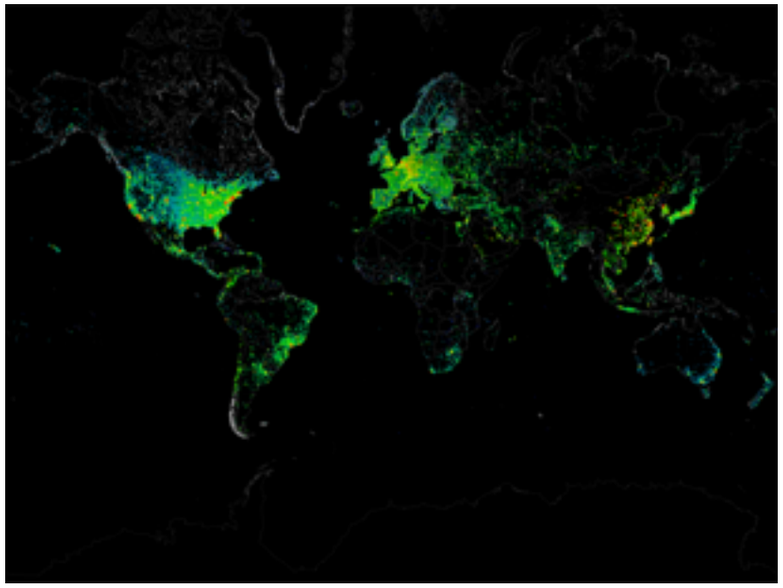
Mapping Address Space Activity



IPv4 Census Visualization 2010
USC, Heidemann et al.



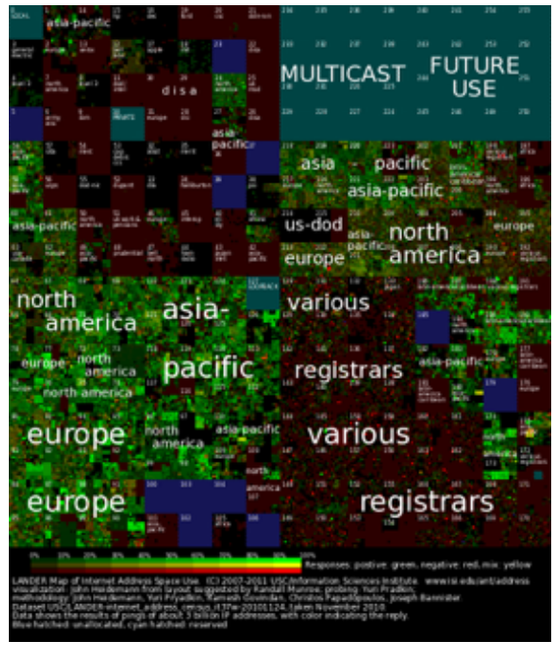
IPv4 Census Visualization 2014
Dainotti et al.



Active IPv4 Addresses 2012
Carna botnet.

Why Mapping Address Activity?

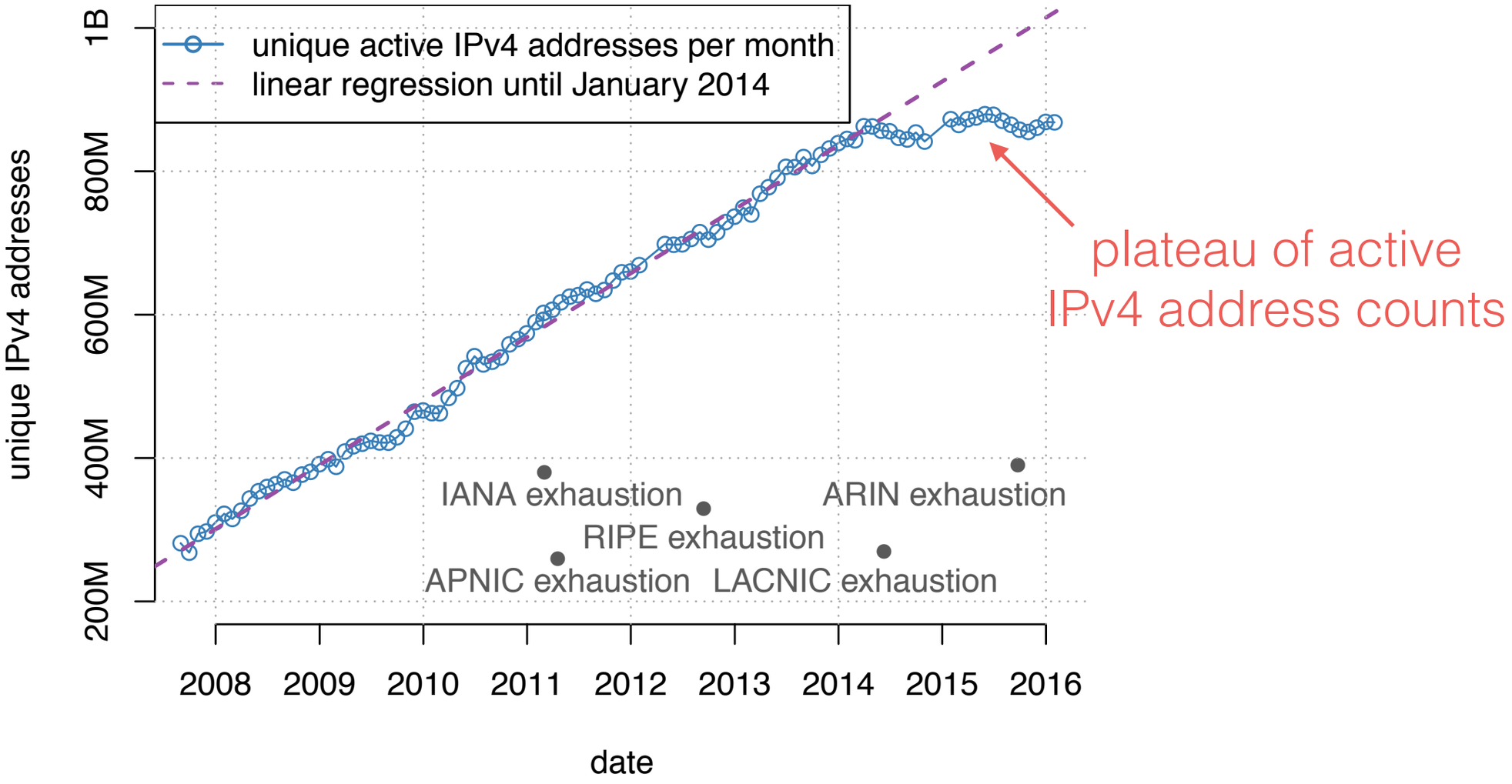
- ➡ Practical applications
  - Security (e.g., host reputation)
  - Operations (e.g., detect network changes)
  - Measurements (e.g., select active targets)



IPv4 Census Visualization
USC, Heidemann et al., 2010.

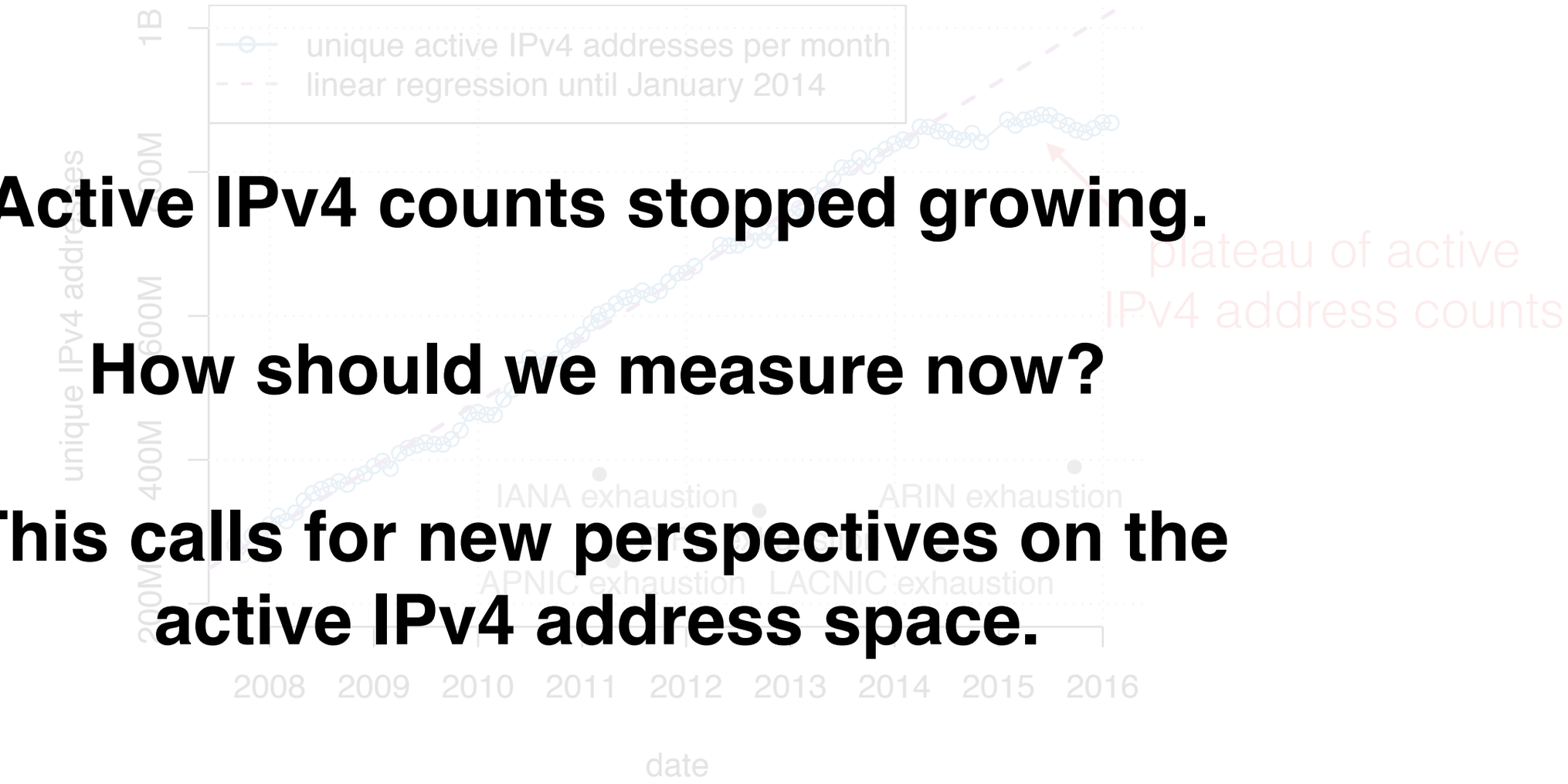
- ➡ Track growth and expansion of the Internet
- ➡ More recently: **IPv4** address exhaustion
  - Measure address space utilization
  - Inform policies for address management

Stagnation of active IPv4 Addresses



Active IPv4 address counts have stagnated since 2014, while IPv6 counts have grown.

Stagnation of IPv4 Address Activity



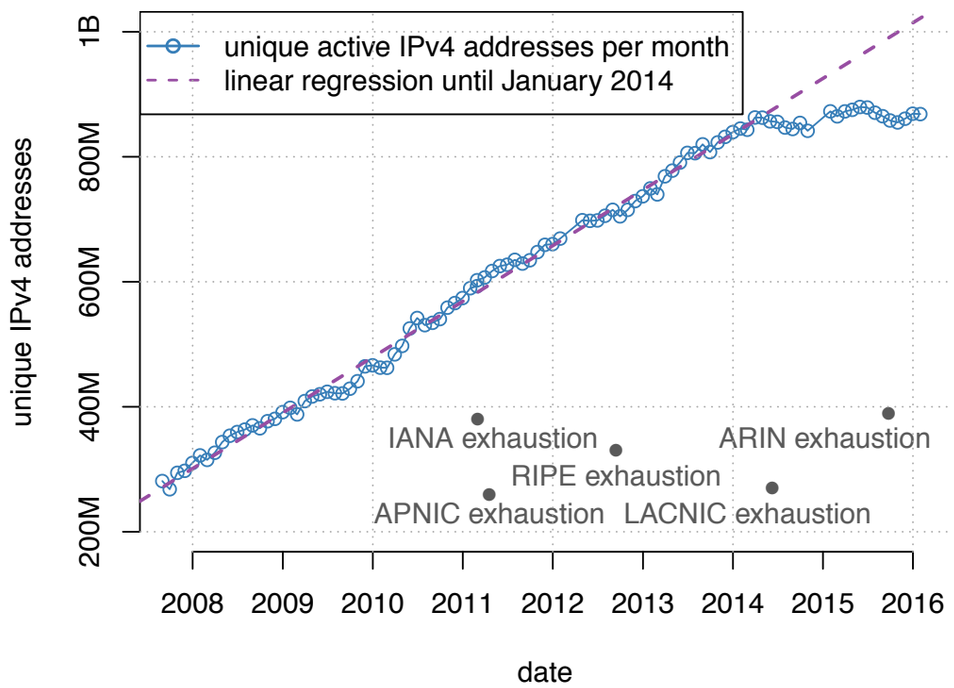
Active IPv4 counts stopped growing.

How should we measure now?

This calls for new perspectives on the active IPv4 address space.

Active IPv4 address counts have stagnated since 2014

Questions



- How effectively can we measure IPv4 space activity?
- At what timescales does activity manifest itself?
- What operational practices eventually determine activity?
- Can we extract meaningful address space demographics?

Agenda

- Related Work
- The CDN as an Observatory
- Macroscopic View on Address Activity
- Microscopic View on Address Activity
- Traffic & Devices
- Implications

Related Work Highlights

2008 First active census of the allocated IPv4 space [Heidemann et al.]

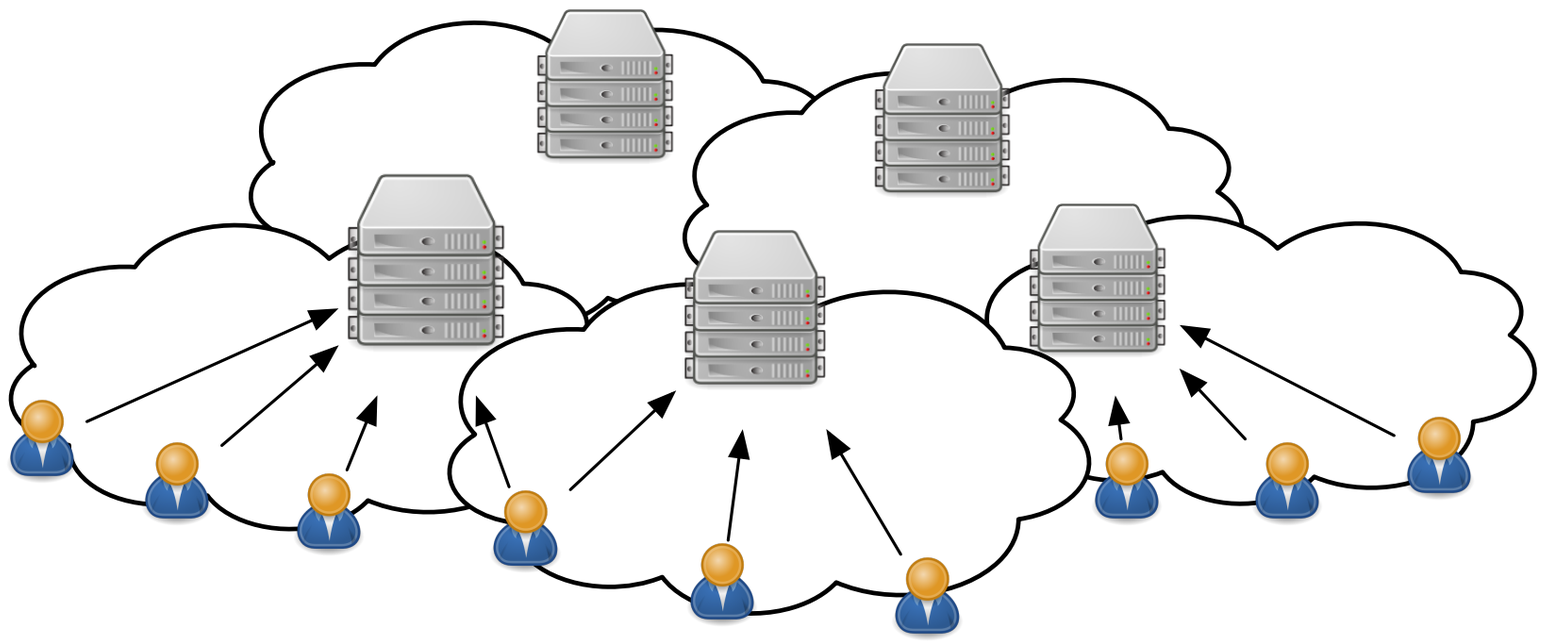
2013 ZMap: Scanning the entire IPv4 space in < 1 hr [Durumeric et al.]

2014 Measure/estimate total number of “used” IPv4 addresses

- ➡ combine active and passive data sources

- ➡ **5.5M active /24 blocks measured, 2013 (7 sources)** [Dainotti et al.]
- ➡ **1.2B active IPv4 addresses estimated, 2014 (9 sources)** [Zander et al.]

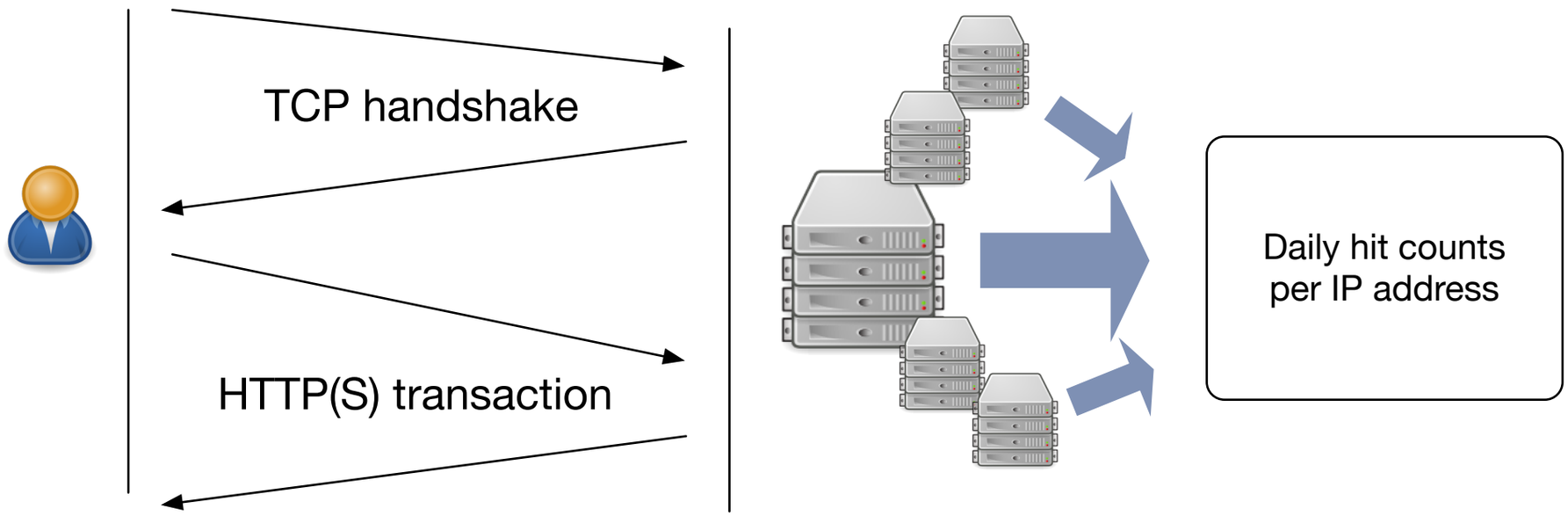
The CDN as Vantage Point



The CDN Vantage Point

- 200,000+ servers in 1500+ ASes in 120+ countries
- Serves Web content, mobile content, software updates, etc.
- **3 trillion requests on a daily basis**

CDN Logs



Our Data: CDN Logs

- HTTP(S) requests per IP address from all 200,000+ servers
- Evidence of Web activity (not affected by spoofing)

2015: A Year of CDN Client Activity

CDN Vantage Point: Advantages

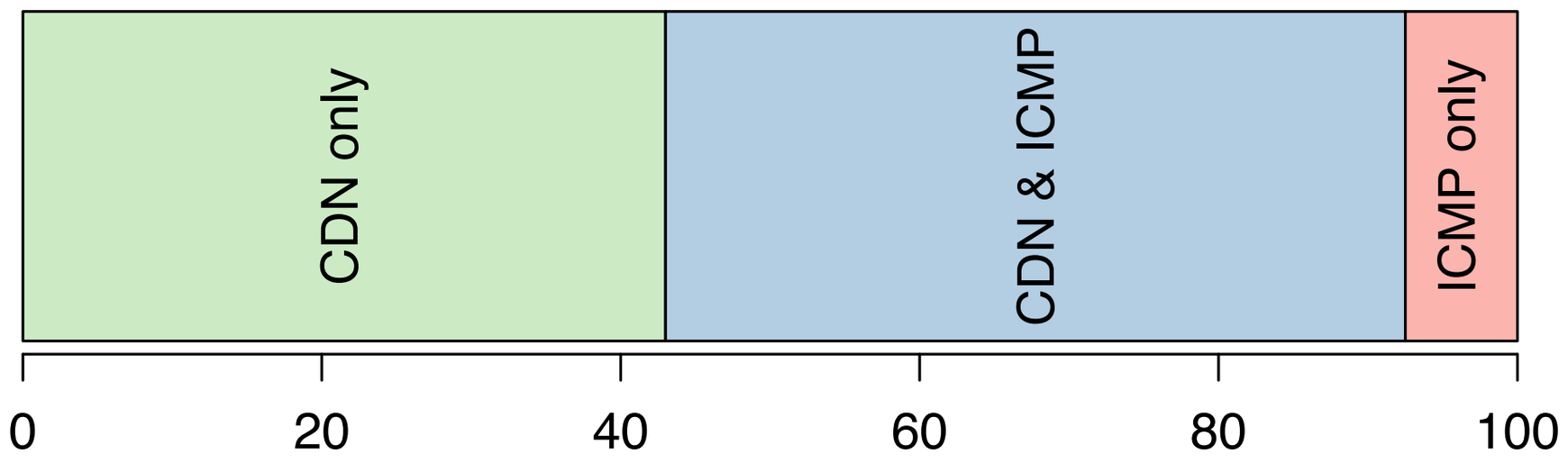
- Year-long observation of Web client address activity
- Granularity: Individual IPv4 address activity on a daily basis

Web Client Visibility (totals for 2015)

- **1.2 billion unique IPv4** addresses (42% of routed)
- **6.5M active /24** address blocks (59% of routed)

Highest number of active IPv4 addresses measured so far.

Visibility: CDN vs. ICMP (8 snapshots, ZMap)



% IPv4 addresses visibly active (N=950, Oct. 2015)

CDN only (~40%)

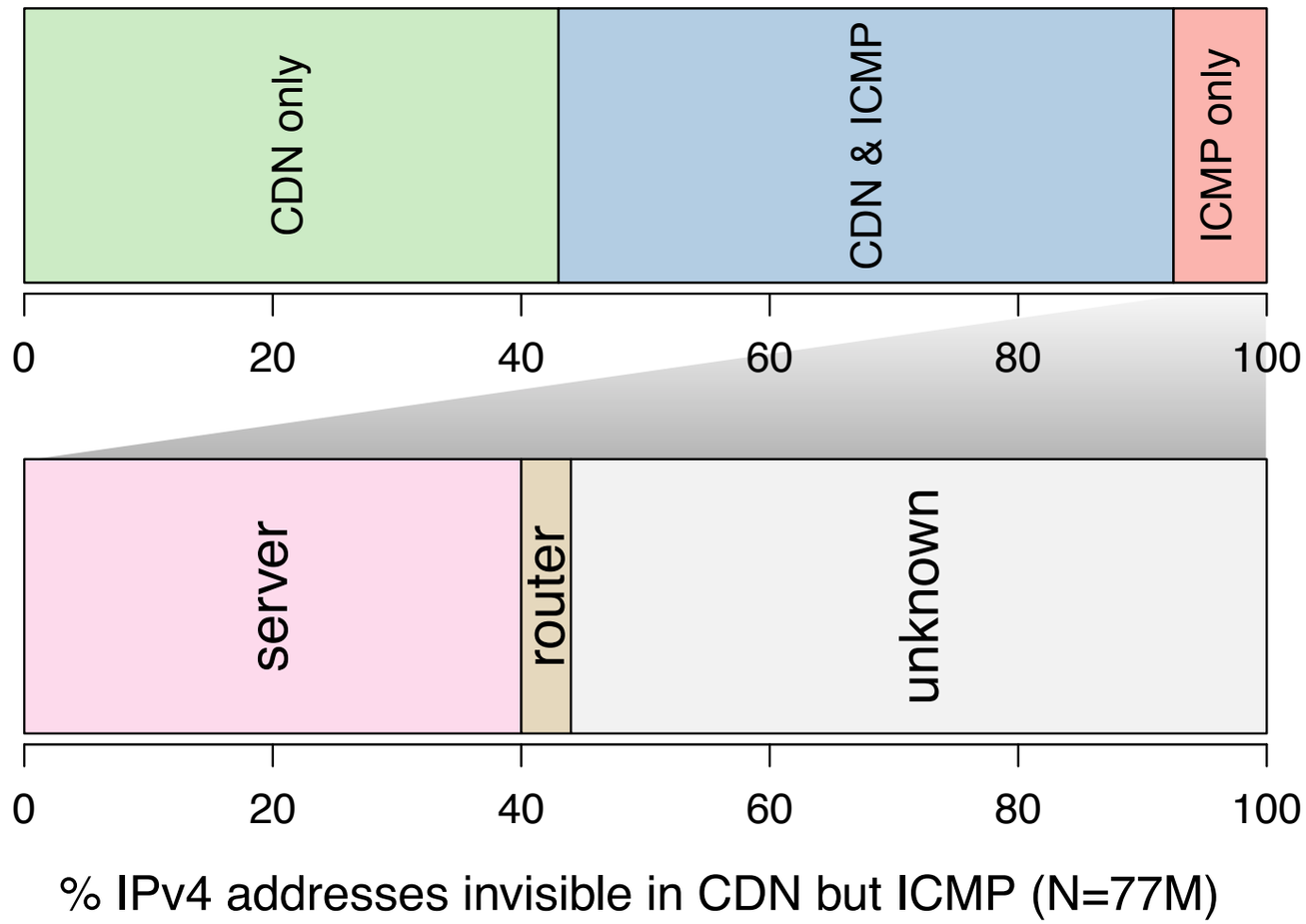
- Only 8 ICMP scans vs. 1 month
  ➡ short-term activity not visible
- Hosts not replying to ICMP
  ➡ firewalls, access control, etc.

ICMP only (~8%)

- Hosts not contacting the CDN
- Infrastructure / servers
- Tarpits

Finding Non-CDN Activity

- Router IP addresses
  - ➡ CAIDA's Ark traceroutes
- Server IP addresses
  - ➡ ZMap Server Scans
- explains 42% of ICMP-only

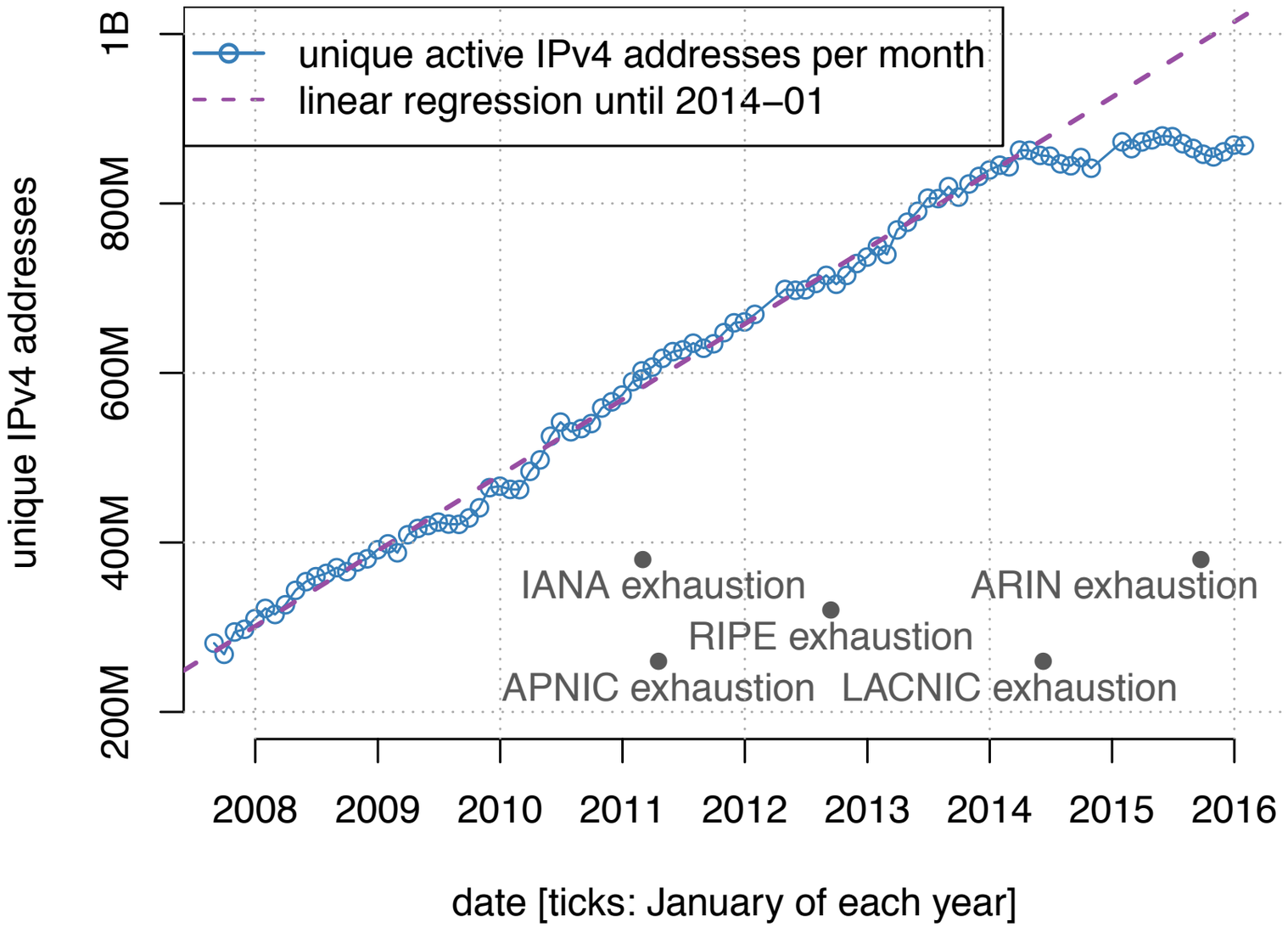


A lot of Non-CDN address activity: "the server side"
Further unseen activity: Future work

Agenda

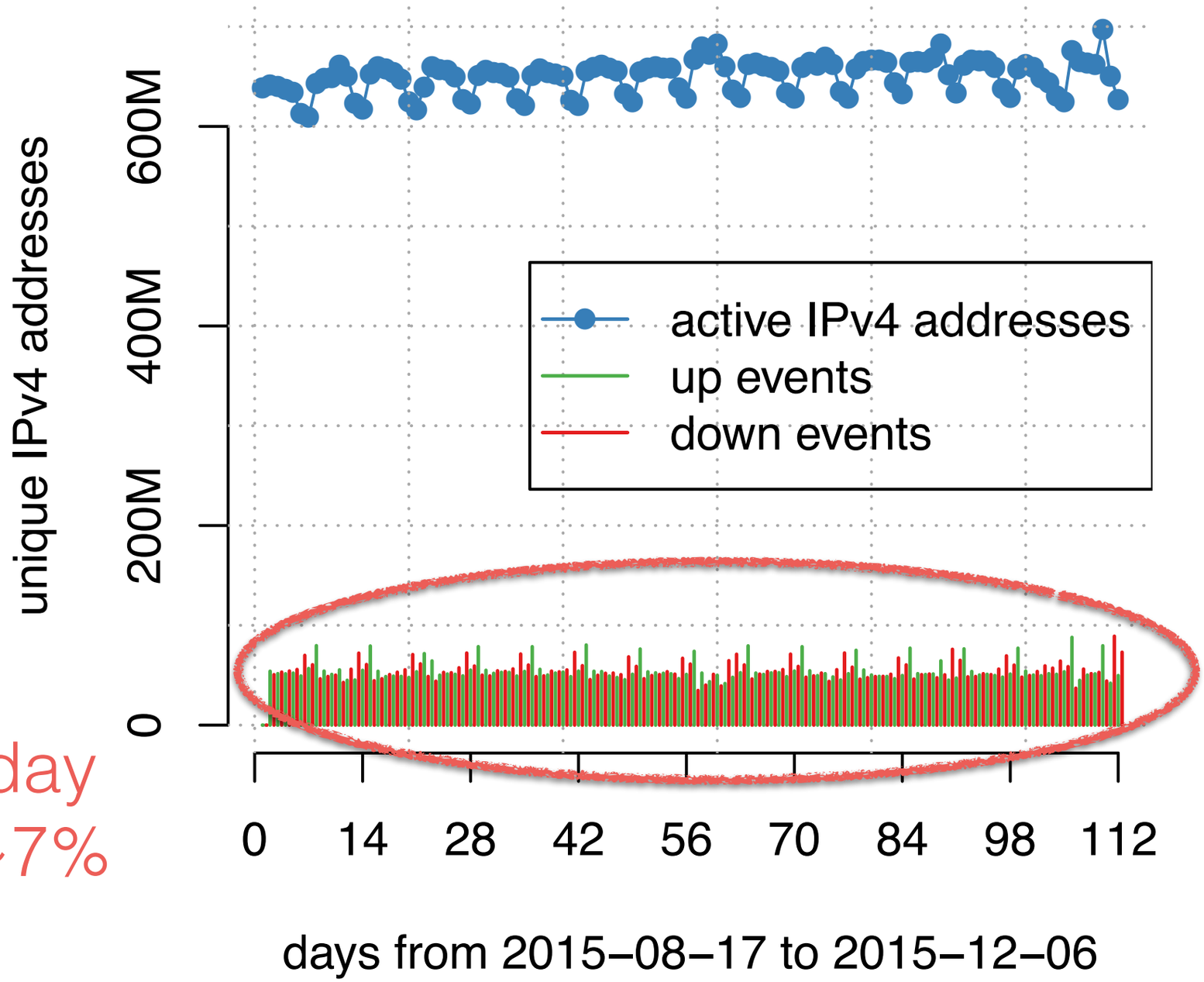
- Related Work
- The CDN as an Observatory
- **Macroscopic View on Address Activity**
- Microscopic View on Address Activity
- Traffic & Devices
- Implications

Peak IPv4 (?)



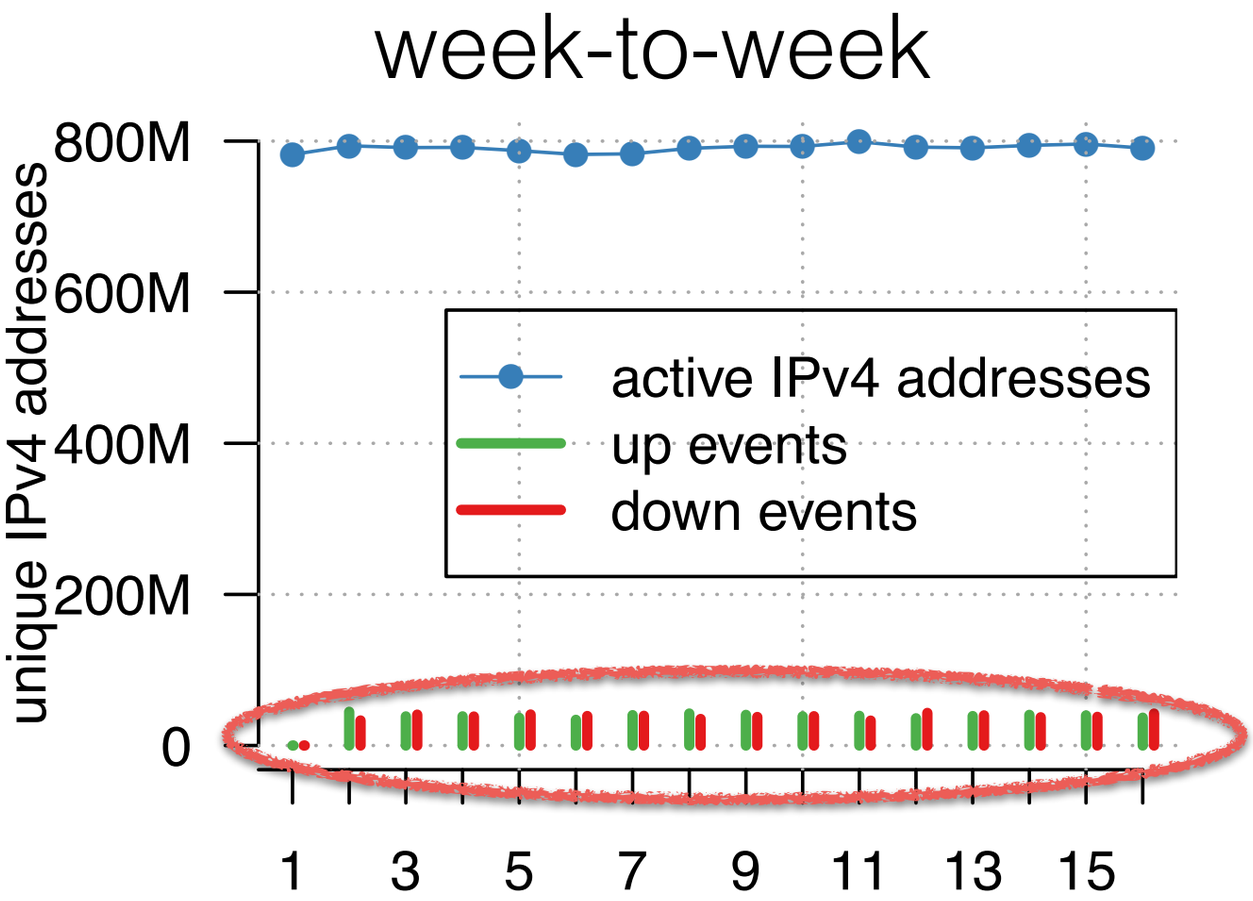
Address counts stagnated.
Do we have a constant set of active IPv4 addresses?

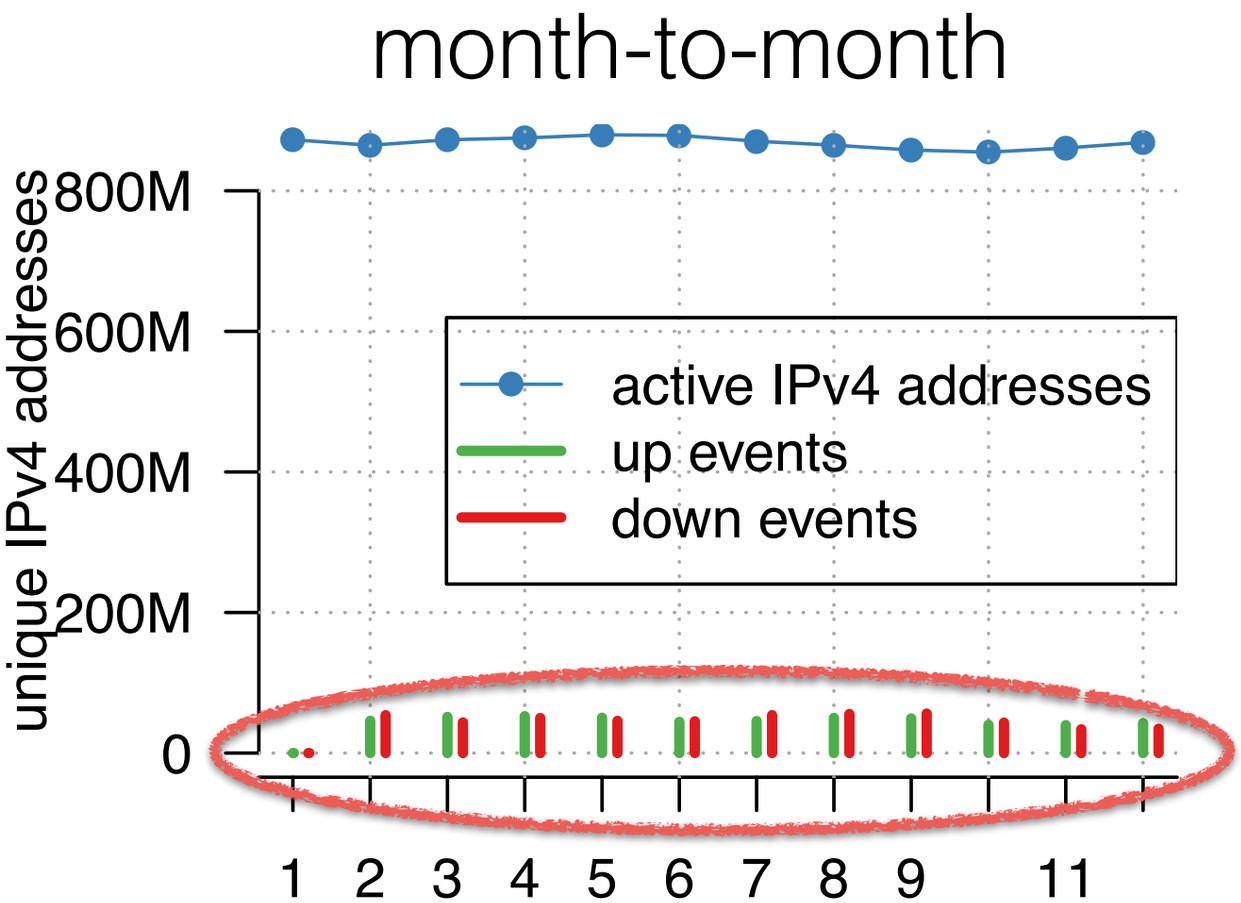
Daily IPv4 Activity and Churn



day-to-day churn: ~7%

Weekly, Monthly Activity and Churn

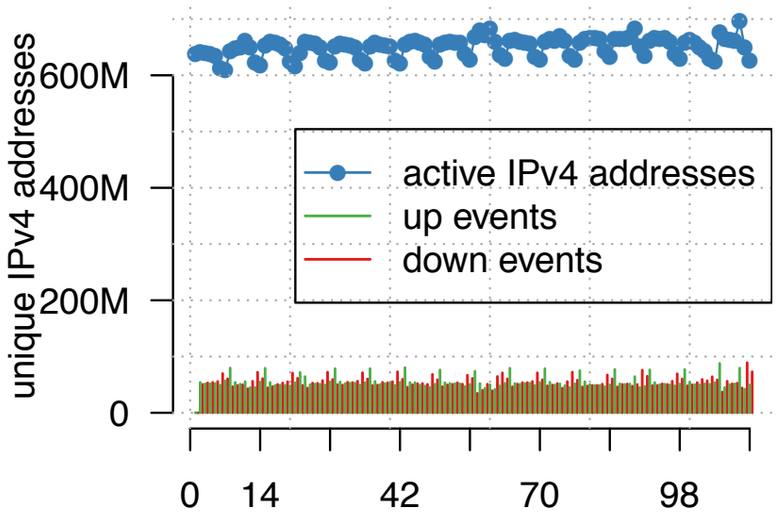


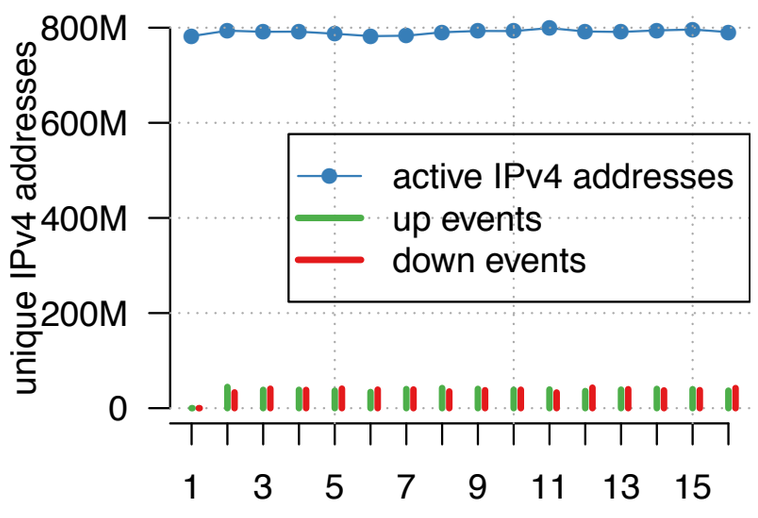


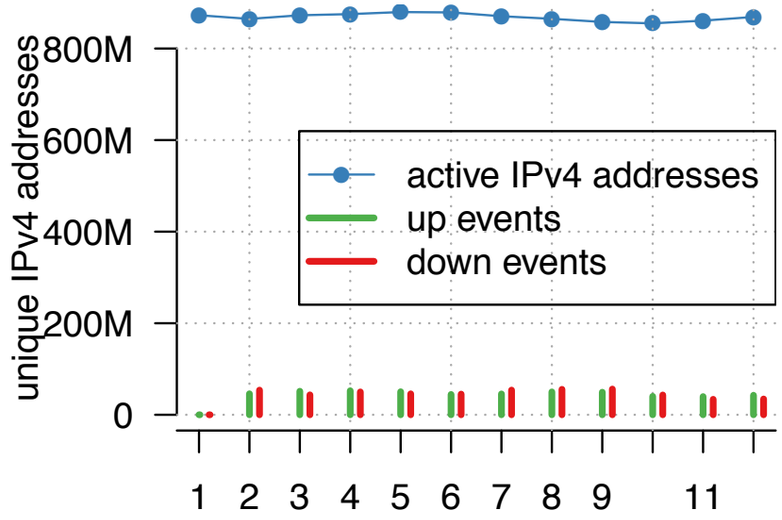
week-to-week
churn: ~5%

month-to-month
churn: ~5%

Churn on All Timescales



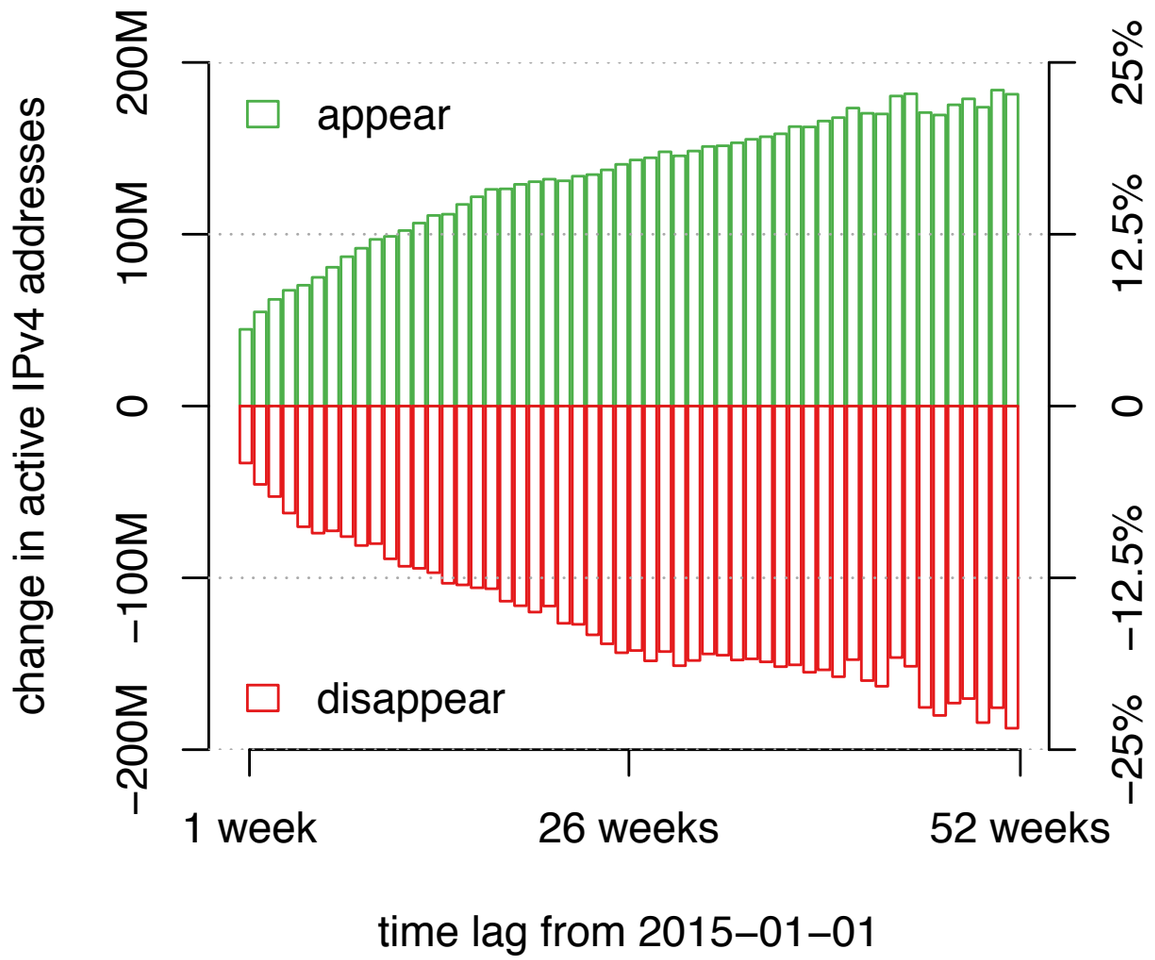




Active IPv4 address population is in constant churn day-to-day, week-to-week, month-to-month

Long-term Effect of Address Churn

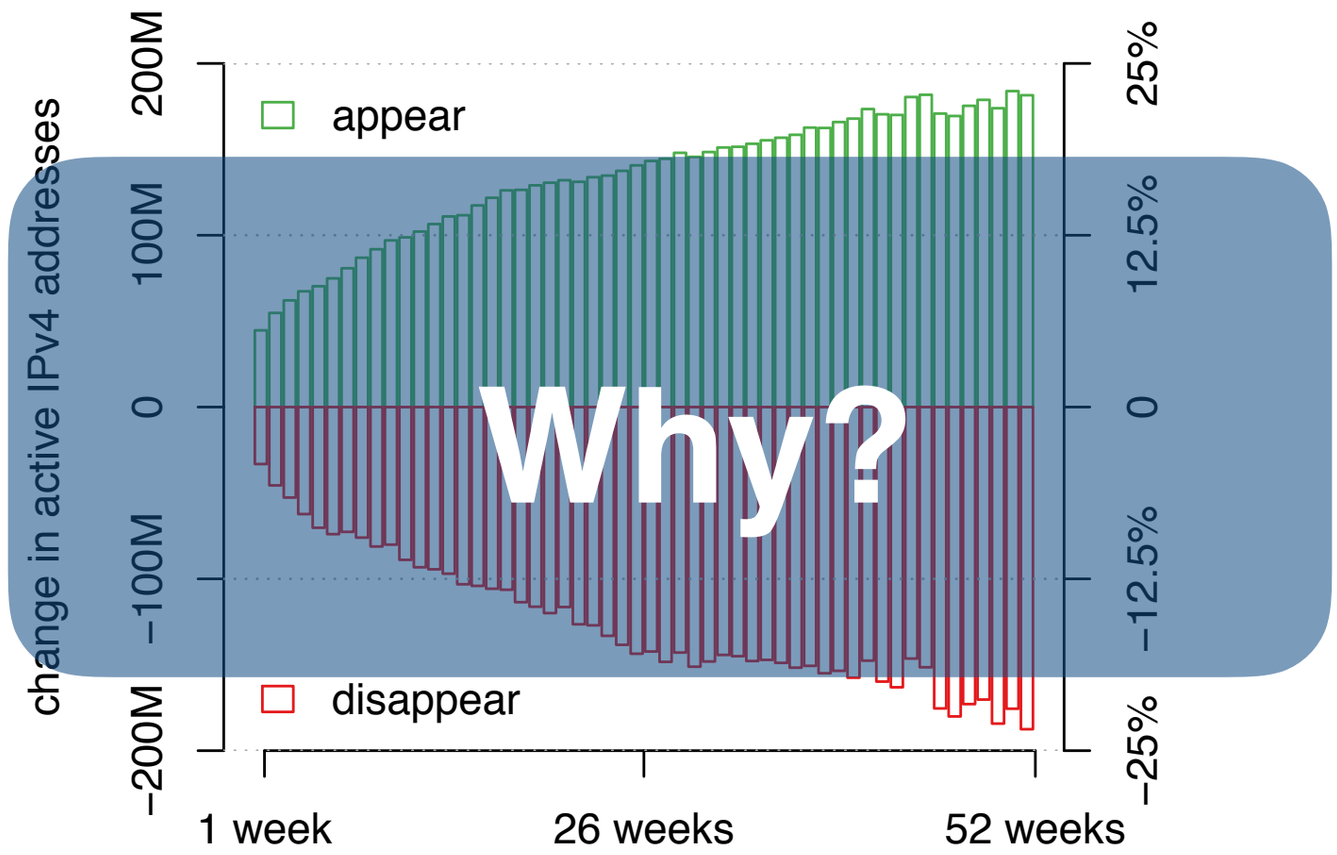
we compare weekly activity against the **first** week of 2015



Over the course of one year, 25% of the active IP address pool changed.

Long-term Effect of Address Churn

we compare weekly activity against the **first** week of 2015

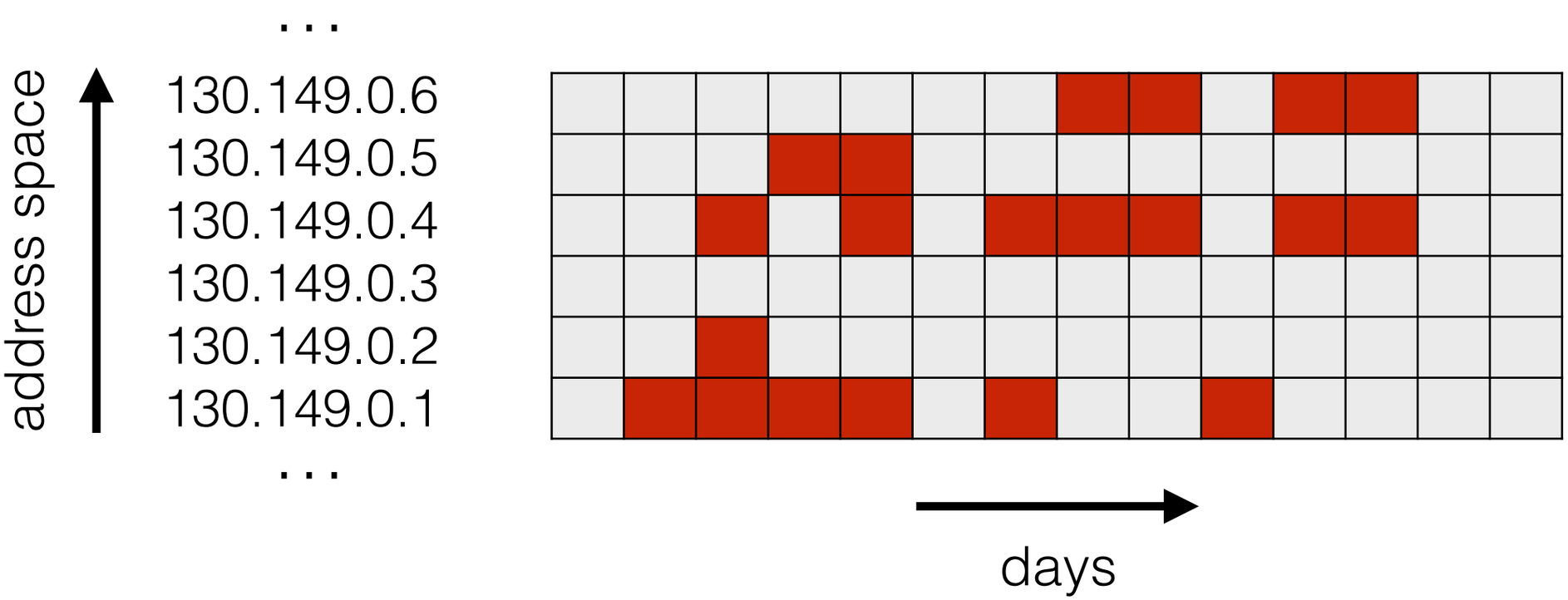


Over the course of one year, 25% of the active IP address pool changed.

Agenda

- Related Work
- The CDN as an Observatory
- Macroscopic View on Address Activity
- **Microscopic View on Address Activity**
- Traffic & Devices
- Implications

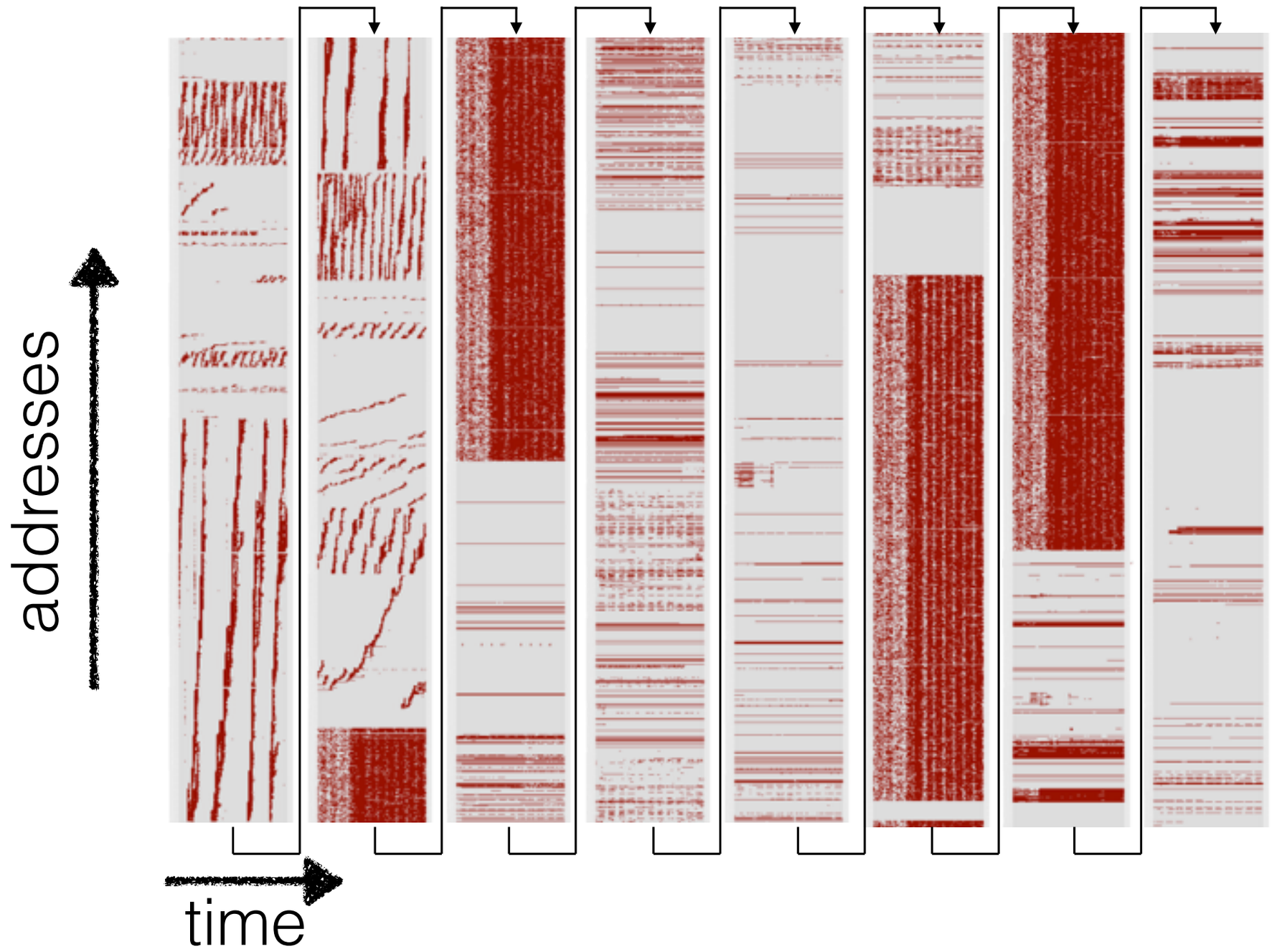
Address Activity Matrix



for each day on which an IP address
was active (requested content), we draw a red dot.

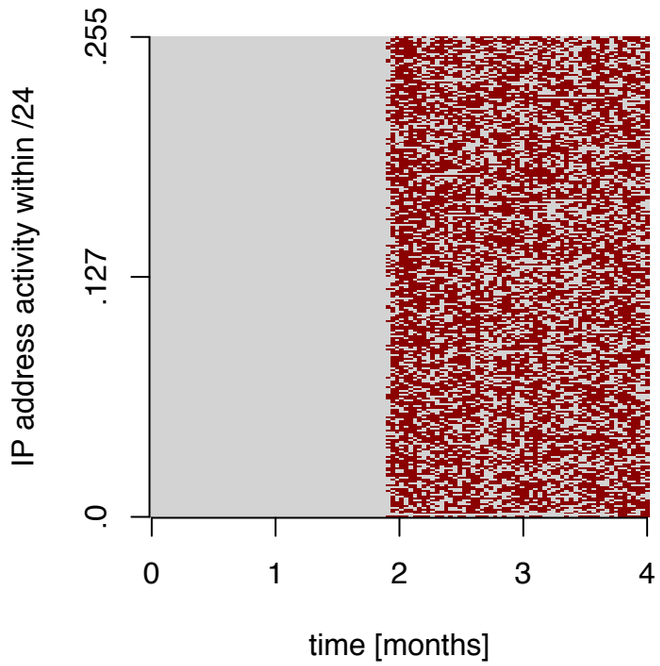
Address Activity Matrix at Scale (“Bacon Strips”)

20k adjacent IP addresses (in active /24s), University Network



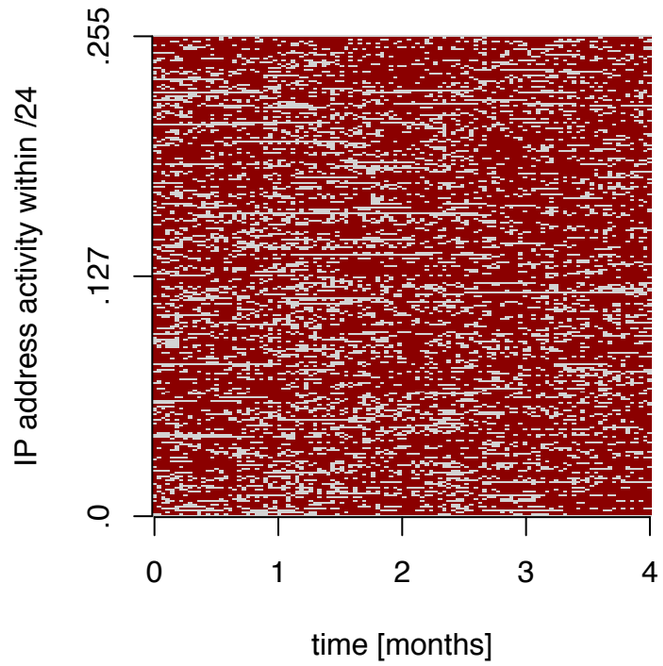
Address Activity and Reasons for Churn

Non-regular Activity, Operational Changes



blocks go into/out of use
activity changes substantially

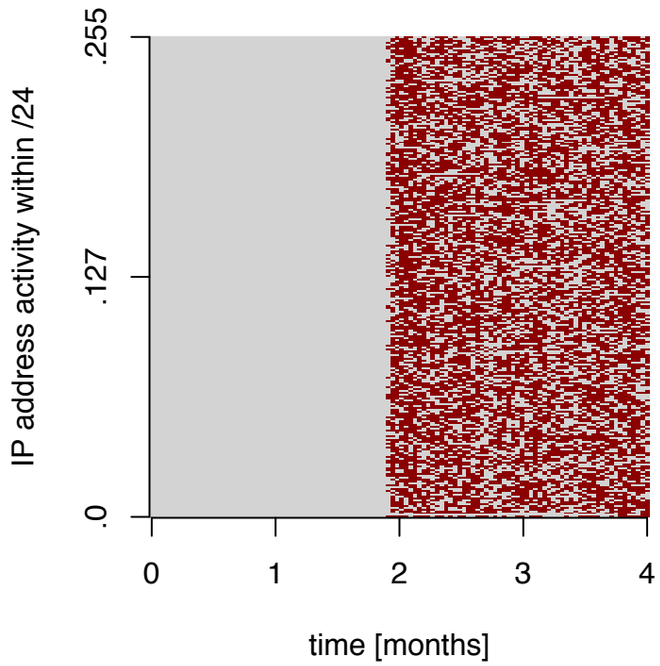
“in situ” - Regular Activity



address assignment practice
and
user behavior

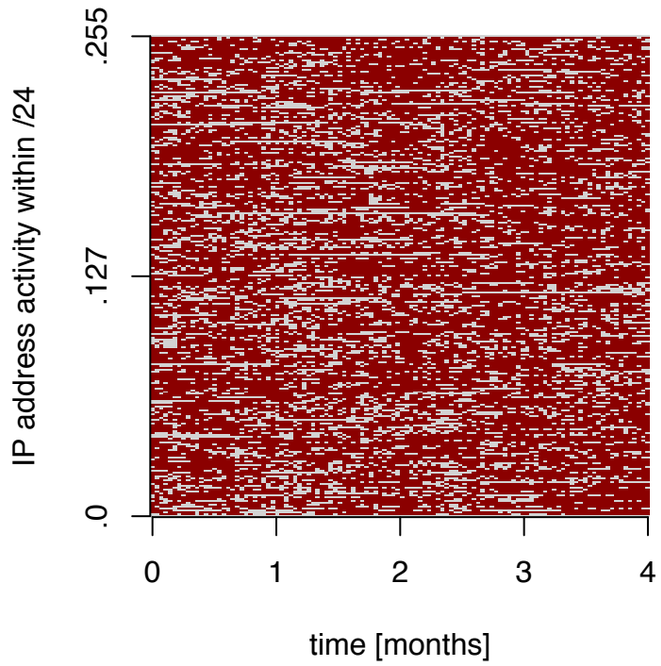
Address Activity and Reasons for Churn

Non-regular Activity, Operational Changes



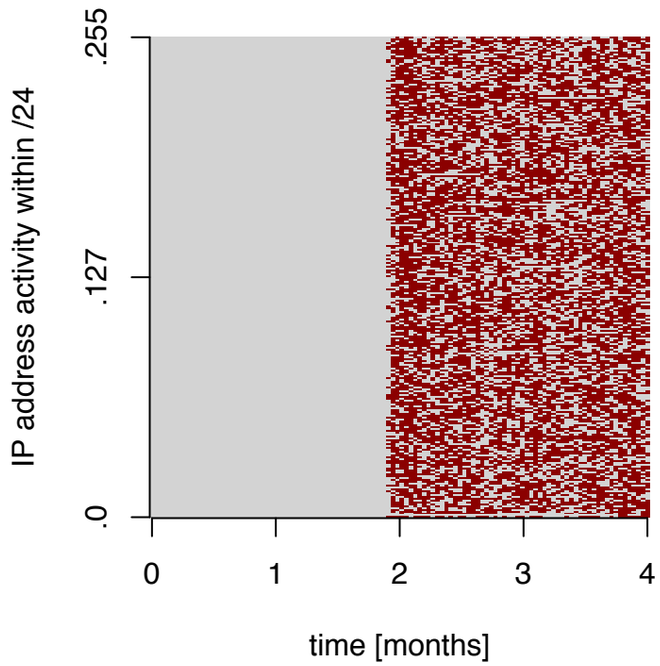
blocks go into/out of use
activity changes substantially

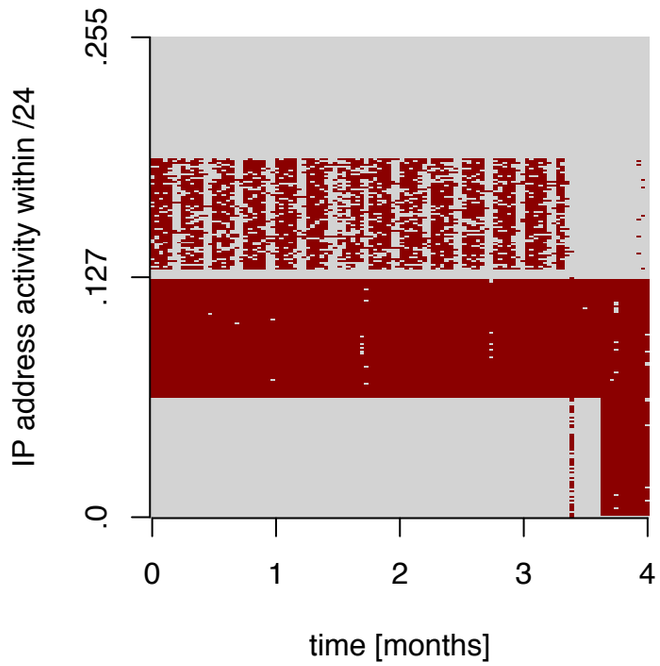
“in situ” - Regular Activity

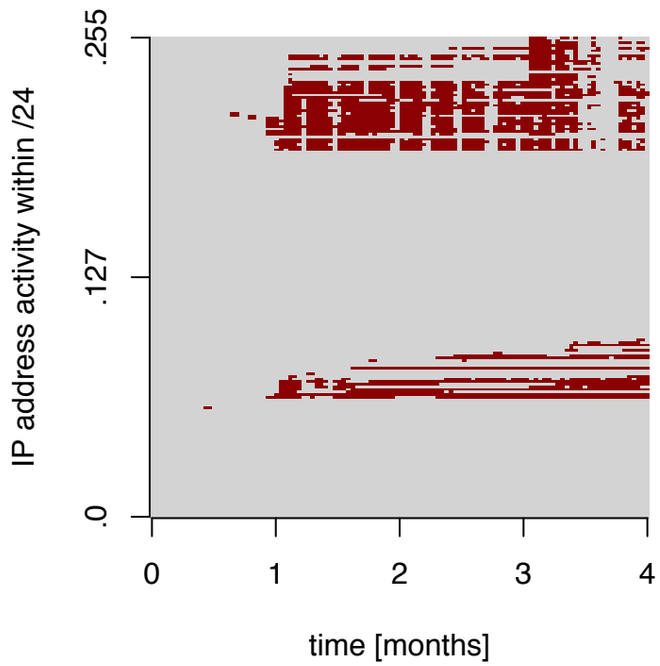


address assignment practice
and
user behavior

Non-regular Activity







Major changes in activity

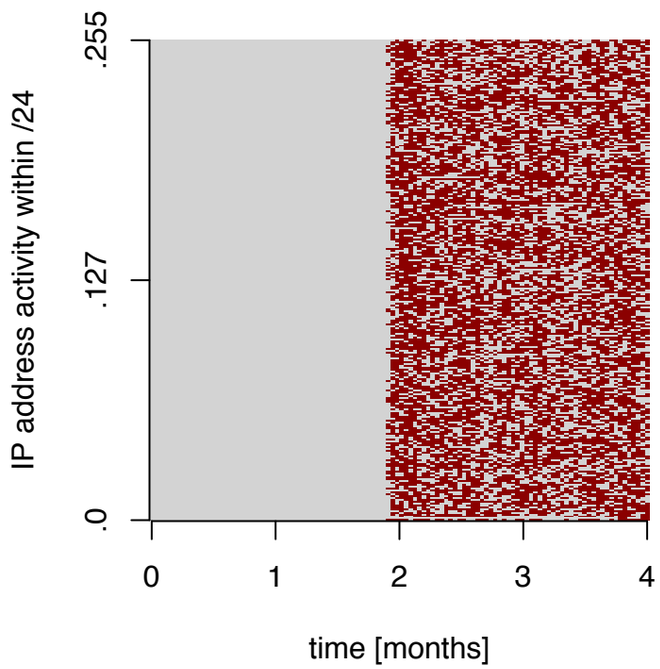
➡ All ranges (intra-/24 up to entire /16 address blocks)

➡ Hardly visible in the global routing table

Up to 10% of active /24 blocks show a major change over the course of 4 months

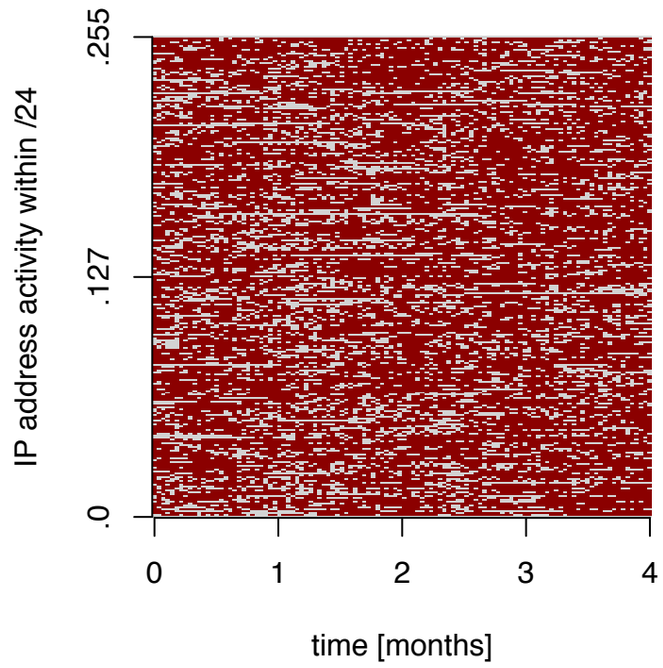
Address Activity and Reasons for Churn

Non-regular Activity: Operational Change



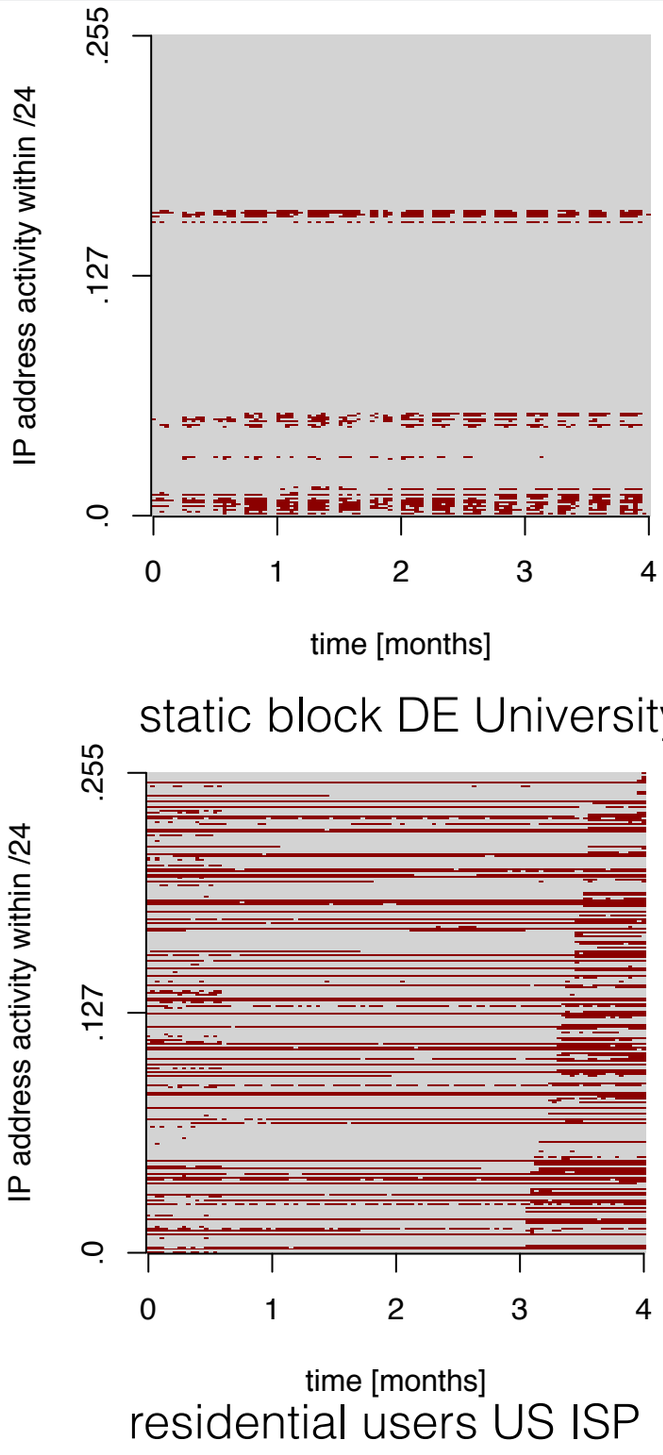
blocks go into/out of use
activity changes substantially

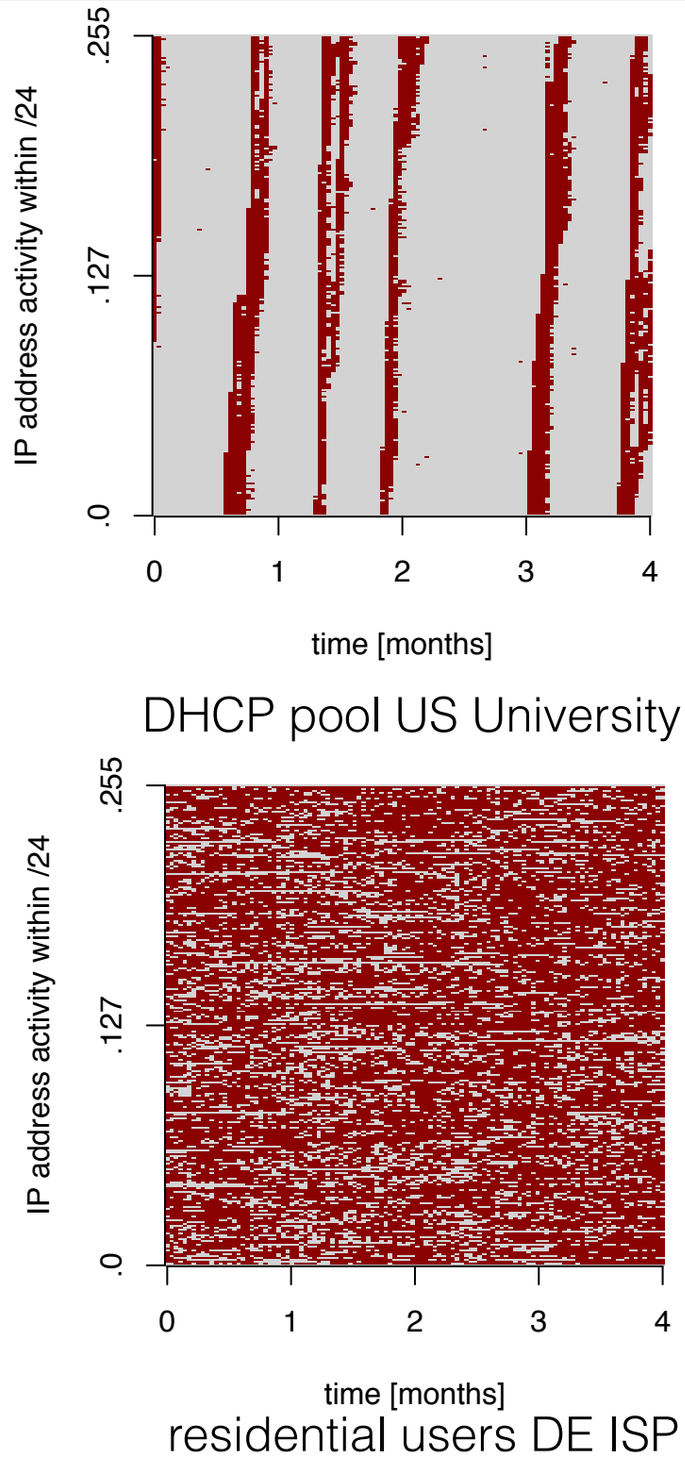
“in situ” - Regular Activity



address assignment practice
and
user behavior

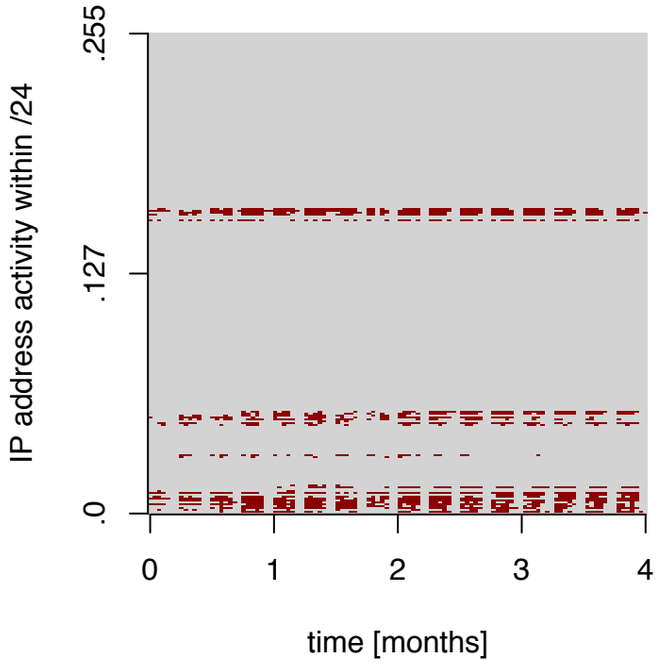
Patterns: Regular Address Activity



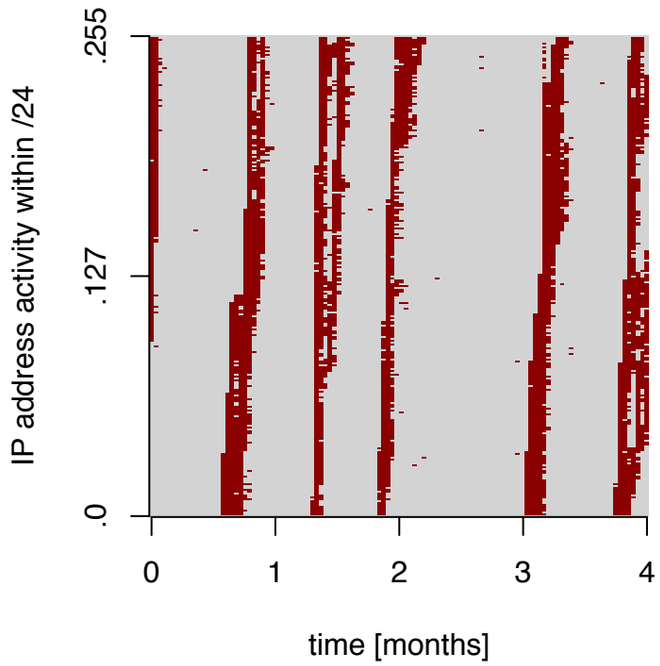


Metric 1: Filling Degree per /24

Number of active IP addresses per /24 [1…256]



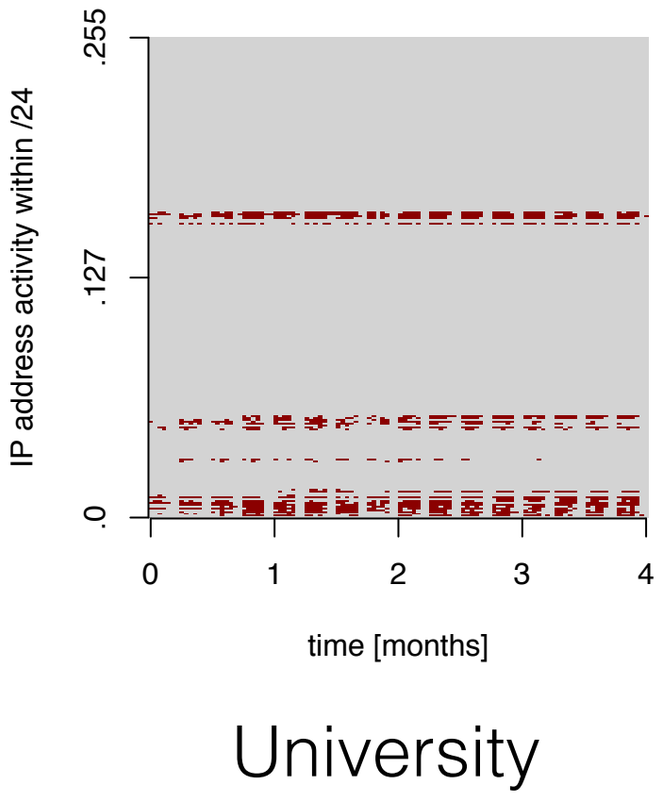
low
(degree = 29)



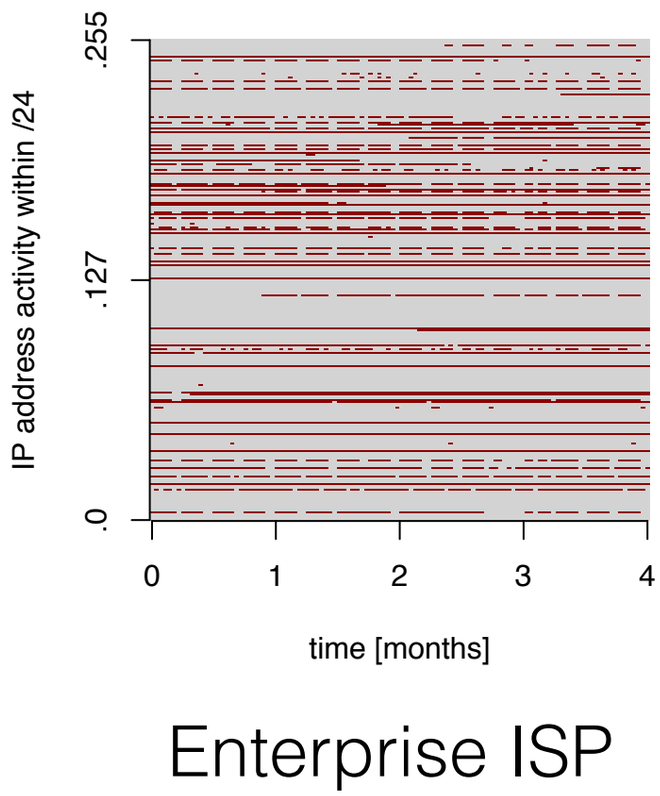
high
(degree = 254)

Typically high (max) for dynamically assigned blocks
Typically lower for statically assigned blocks

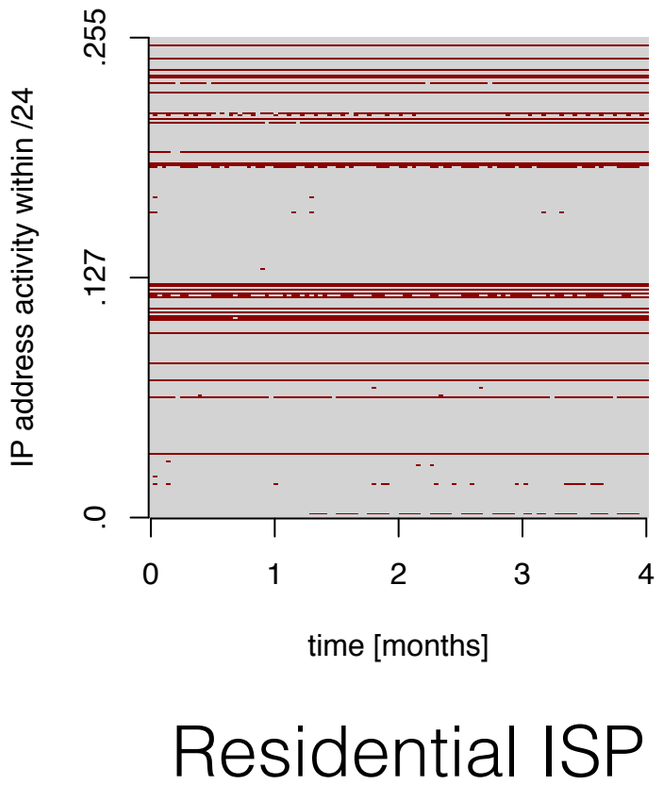
Patterns: Static Address Blocks



University



Enterprise ISP

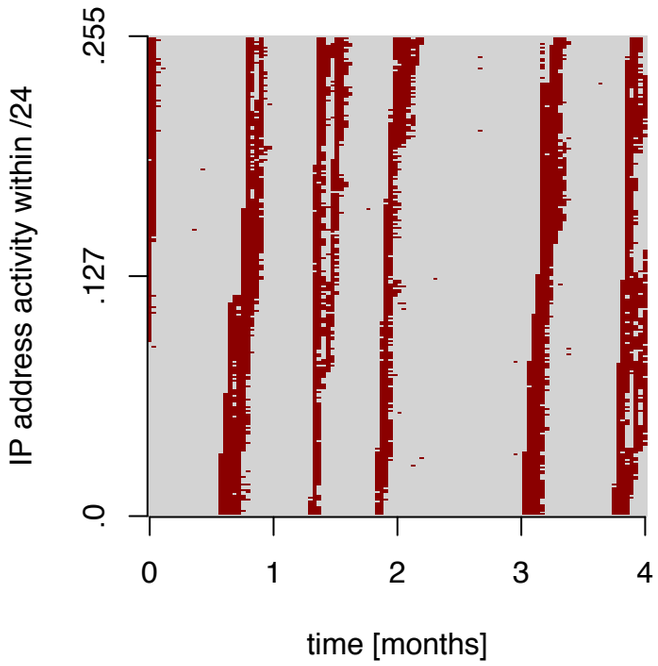


Residential ISP

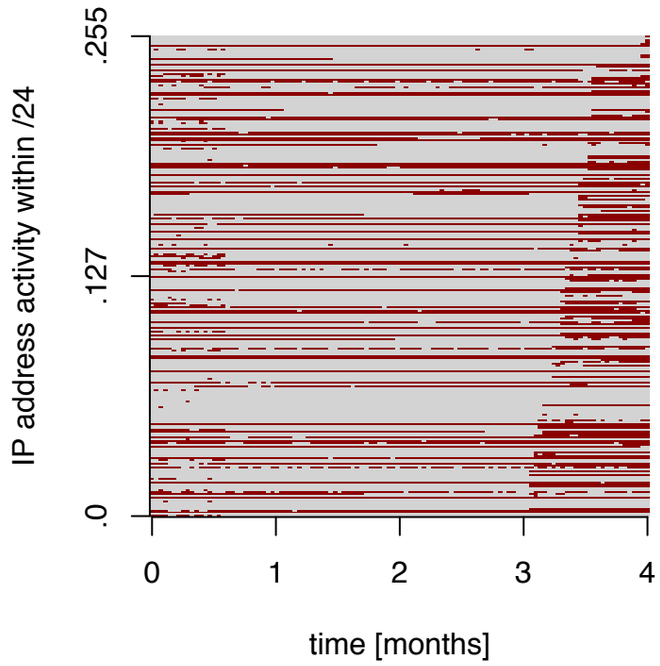
activity depends on manual address assignment

most static address blocks show “activity gaps”

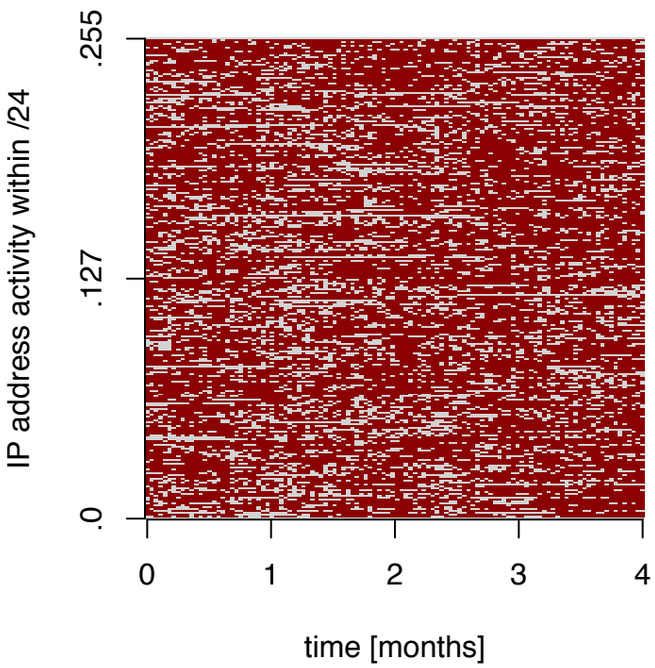
Patterns: Dynamic Address Blocks



DHCP pool US University

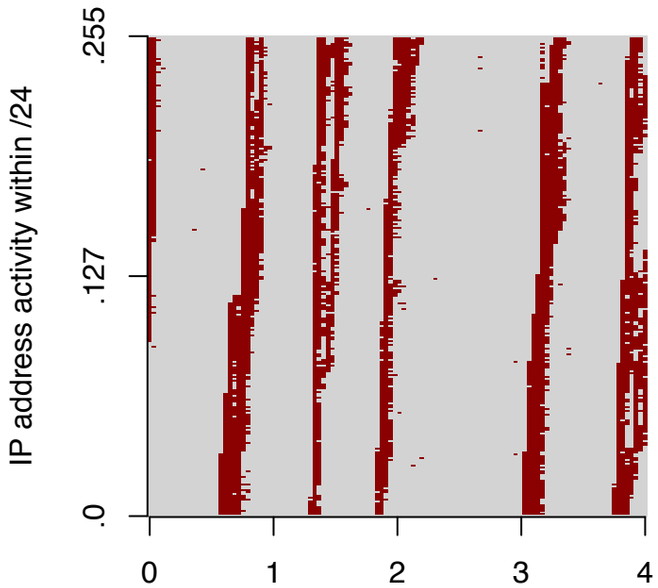


residential users US ISP

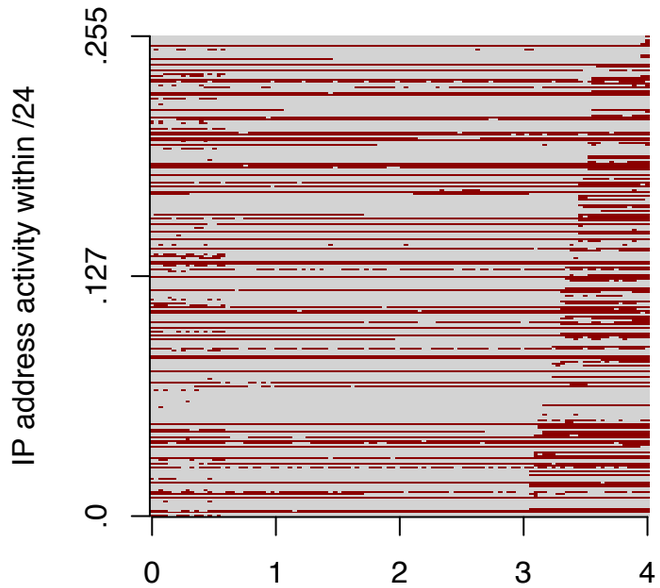


residential users DE ISP

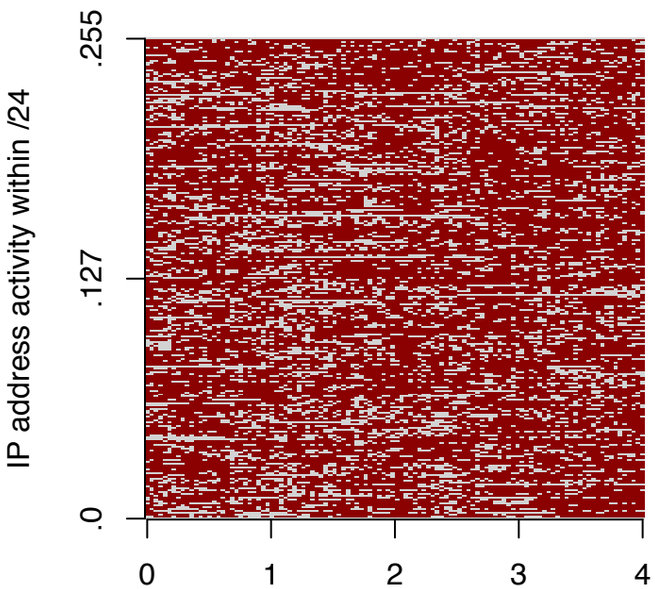
Metric 2: Spatio-temporal Utilization



DHCP pool US University



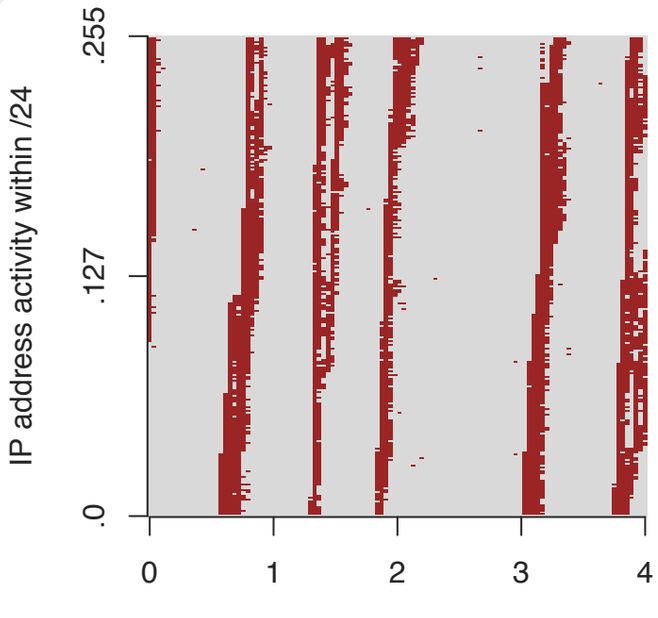
residential users US ISP



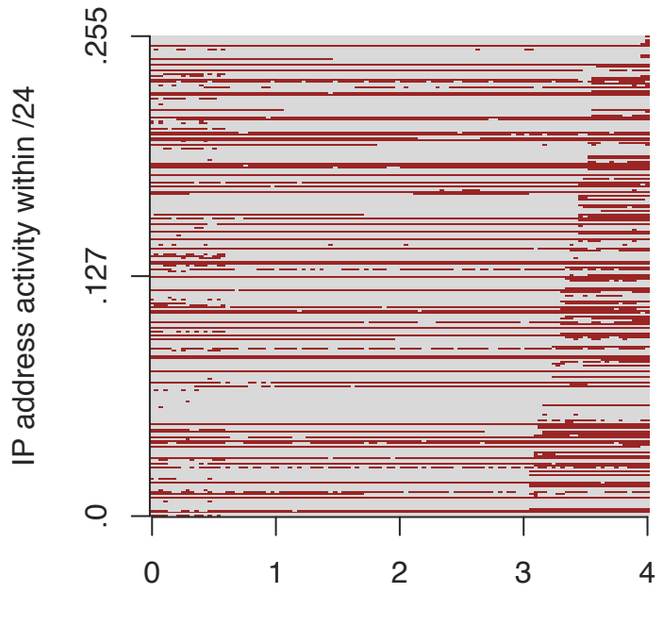
residential users DE ISP

$$\frac{\textbf{\textit{sum(<active IP, day>)}}}{\textbf{\textit{sum(all possible <active IP, day>)}}} = \frac{\textbf{\textit{red}}}{\textbf{\textit{red + grey}}}$$

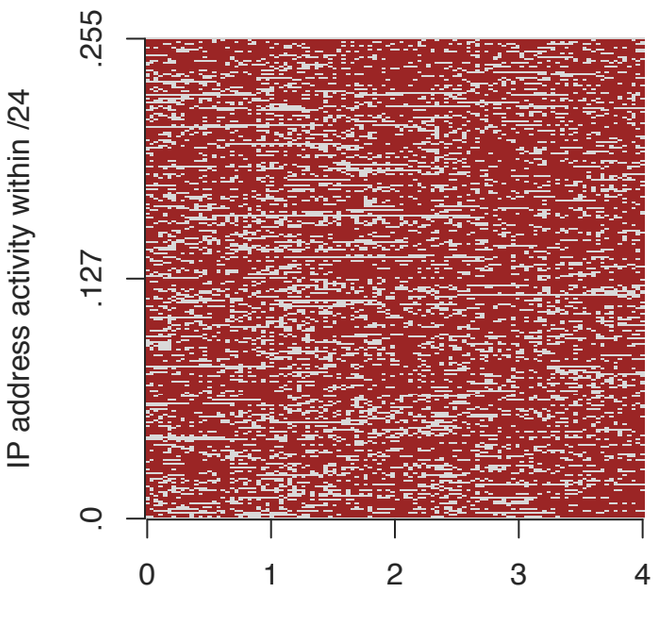
Patterns: Dynamic Address Blocks



spatiotemporal = 18%



spatiotemporal = 26%

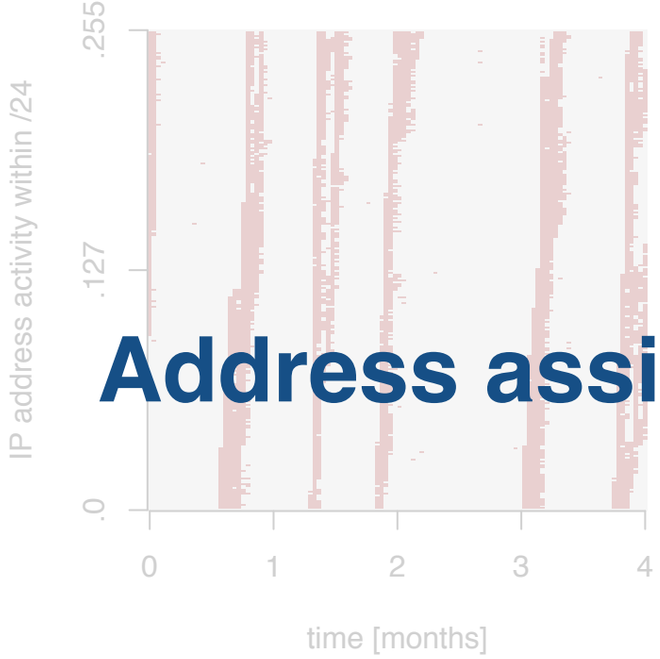


spatiotemporal = 75%

some ~30% of dynamic address blocks (reverse DNS) show low spatiotemporal utilization, overprovisioning

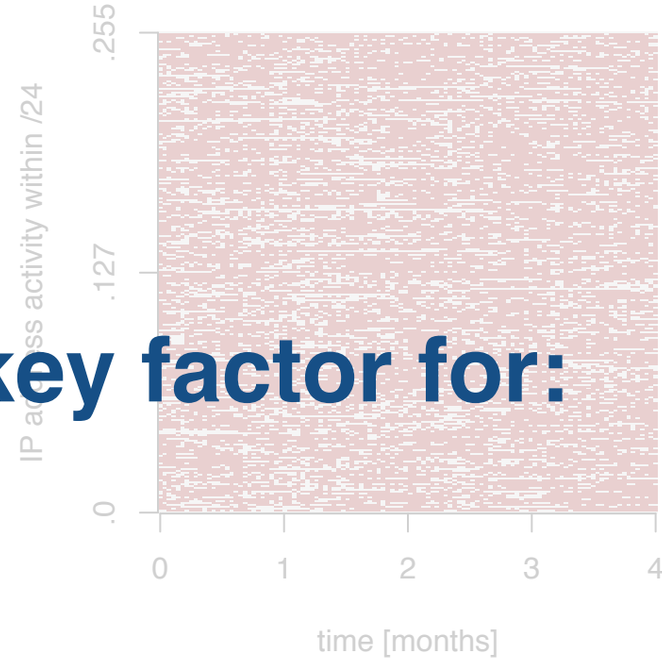
activity and utilization depends on pool size and lease time

Utilization: Dynamic Address Blocks



University
(round robin)

Residential ISP
(long lease time)



Residential ISP
(24hr lease time)

Address assignment practice is a key factor for:

Address activity
Address churn
Address space utilization

utilization depends on pool size configuration

some ~30% of dynamic address blocks (reverse DNS) show low spatiotemporal utilization, overprovisioning

Agenda

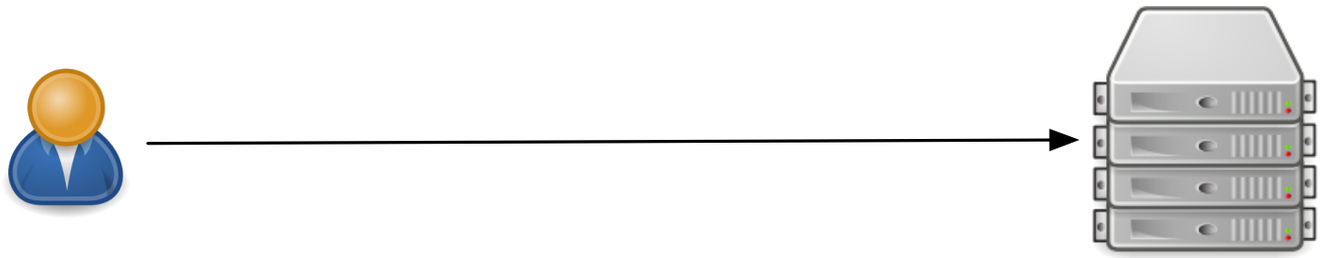
- Related Work
- The CDN as an Observatory
- Macroscopic View on Address Activity
- Microscopic View on Address Activity
- **Traffic & Devices**
- Implications

Traffic & Devices

Estimating Devices Per Address Block

- We sampled HTTP User-Agents of Web requests
- Provide a relative host count metric per address block

```
Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/41.0.2228.0 Safari/537.36
```

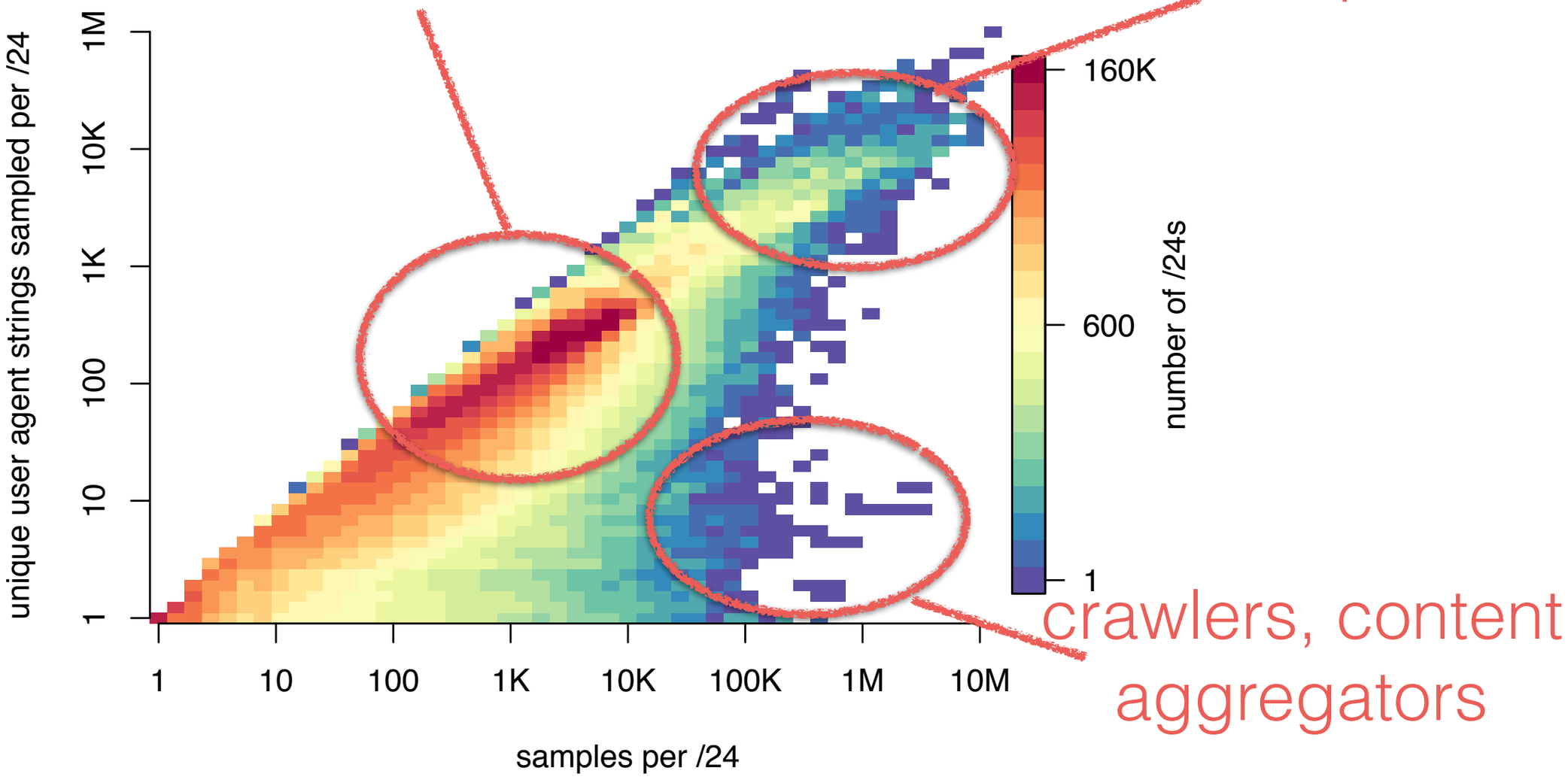


- Rationale: High number of unique User-Agent Strings per IP address block suggests more users/devices

Unique User-Agents per Address Block

majority of blocks
residential users

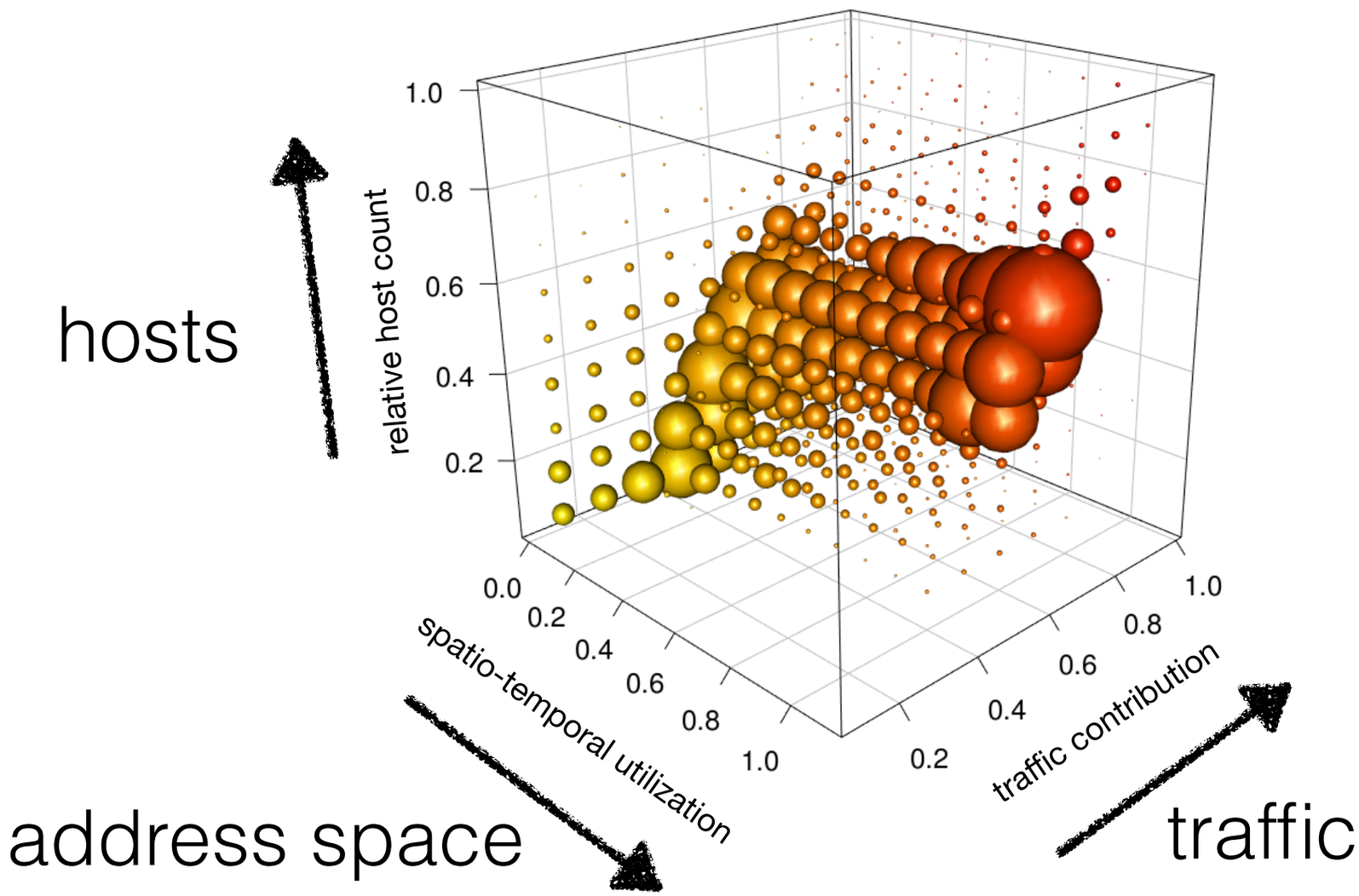
gateways
CGNAT, mobile, proxies



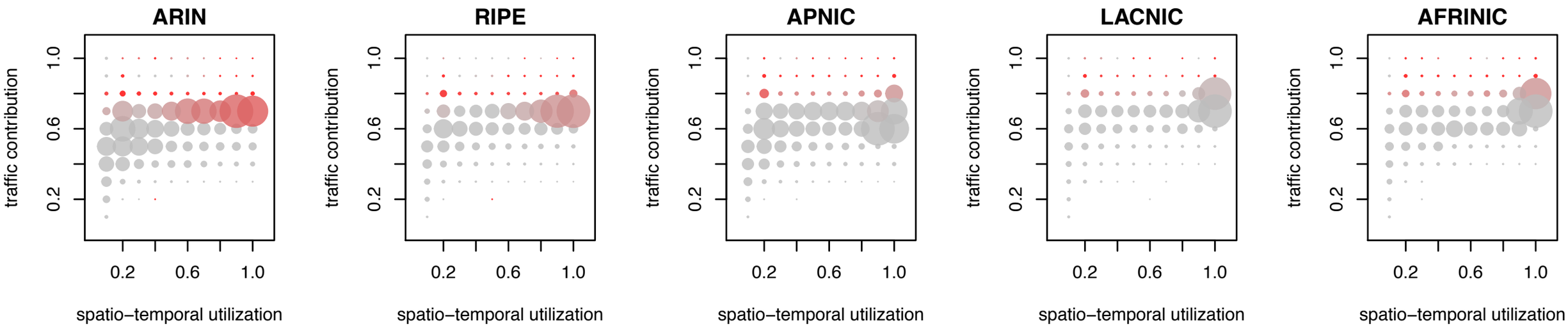
crawlers, content
aggregators

Increasing concentration of traffic on heavy-hitter IPs

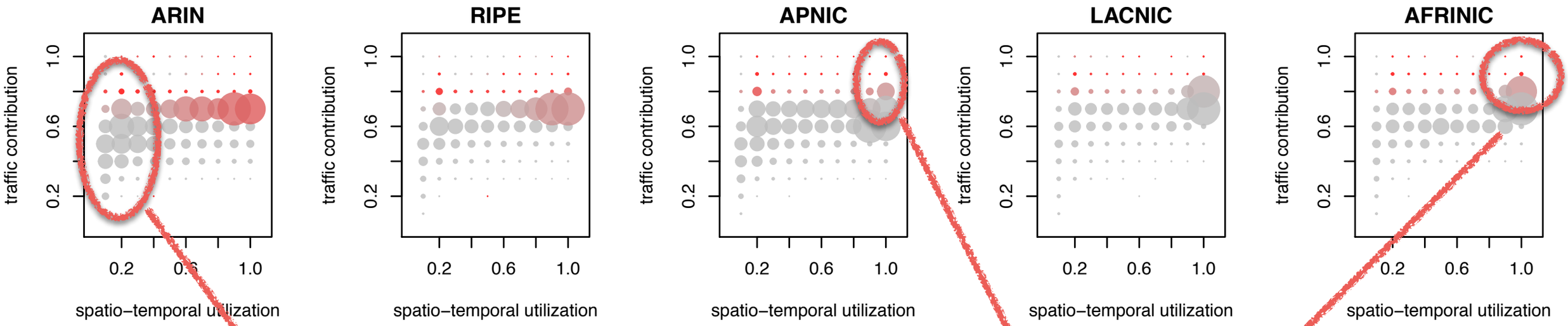
IPv4 Demographics



IPv4 Demographics



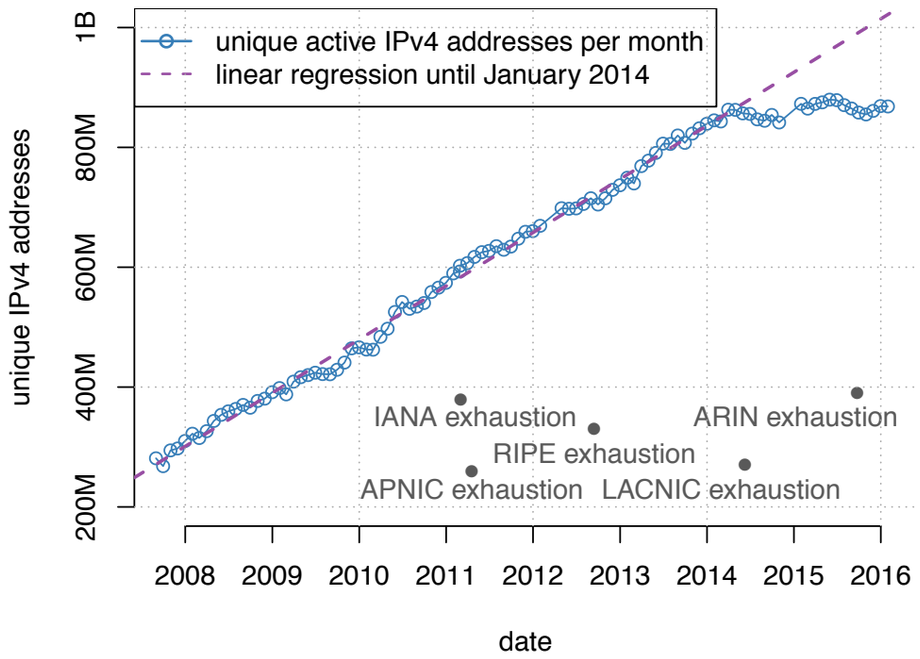
IPv4 Demographics

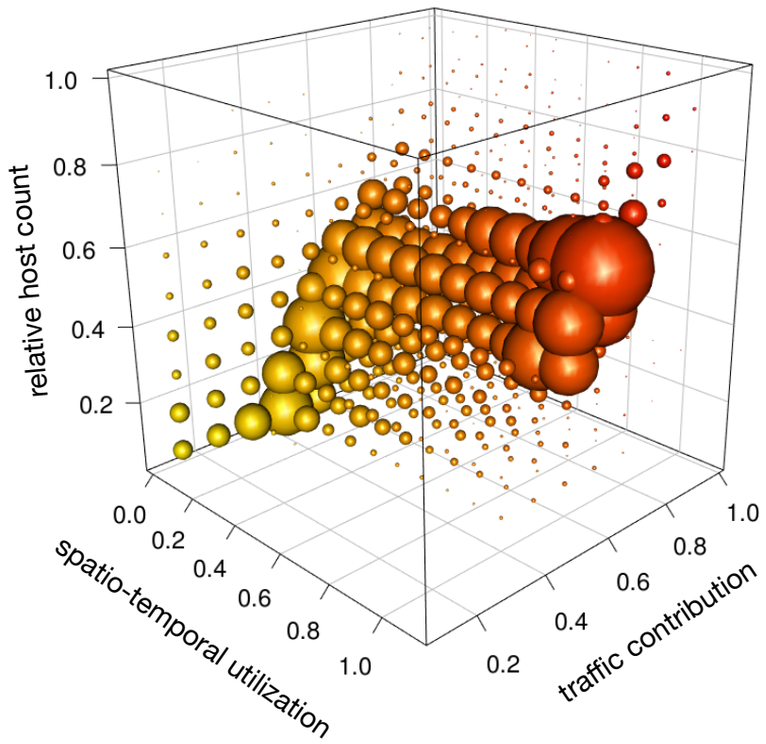


low-utilization blocks
in the north American region
(LEGACY address blocks)

pronounced gateway deployment
in the Asian and African region

Increasingly Complex Situation of IPv4 Space Usage



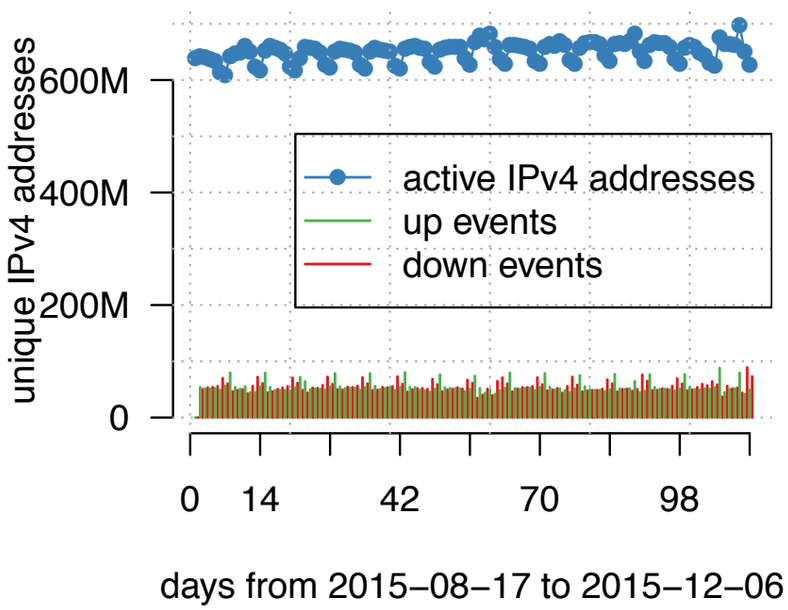


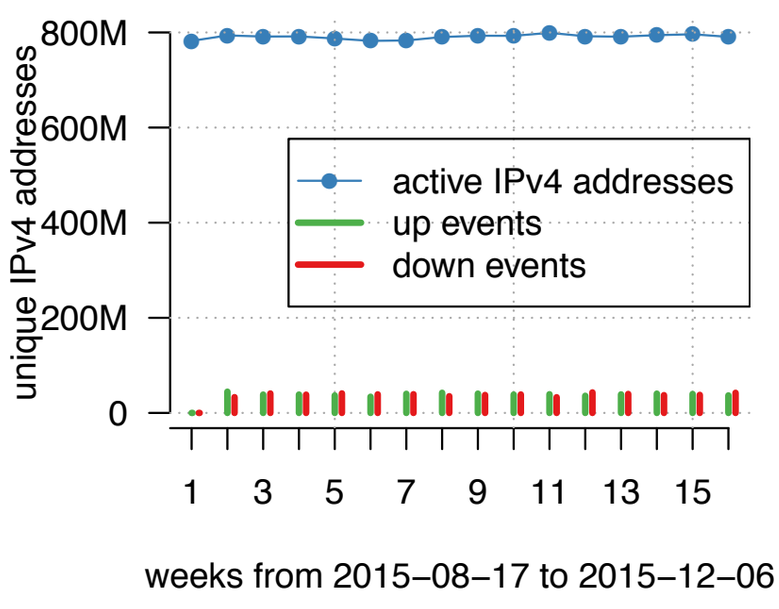
- **Active IPv4 address counts stagnated since 2014**

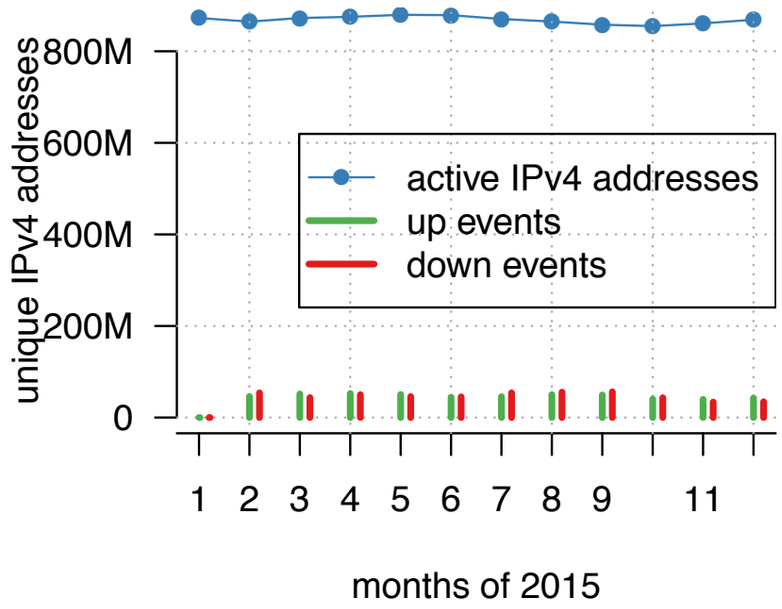
➡ **New perspectives on the active IPv4 space**

- Spatiotemporal utilization
- Traffic
- Relative host counts

Active IPv4 Address Population Highly Dynamic





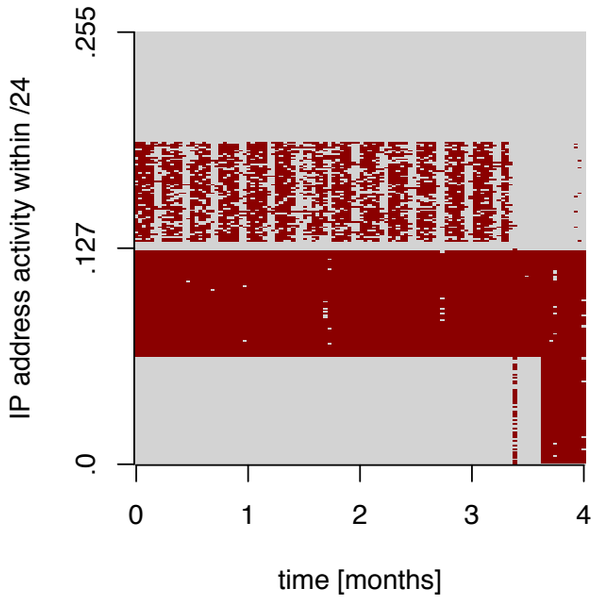


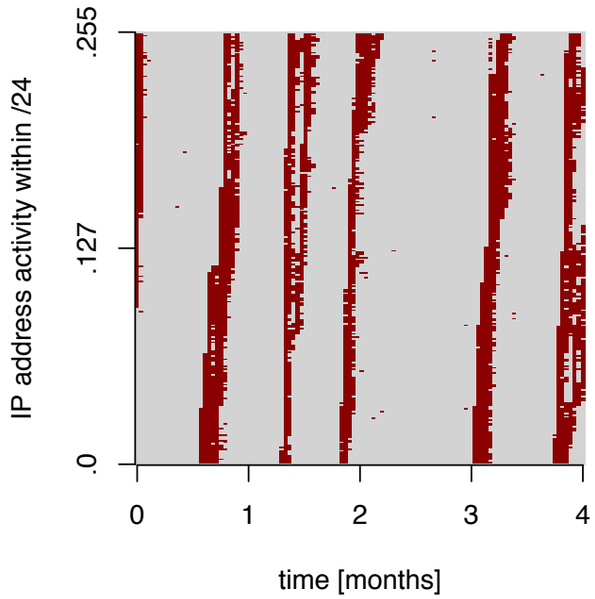
- **Constant churn on all timescales. Up to 25% change within a year**

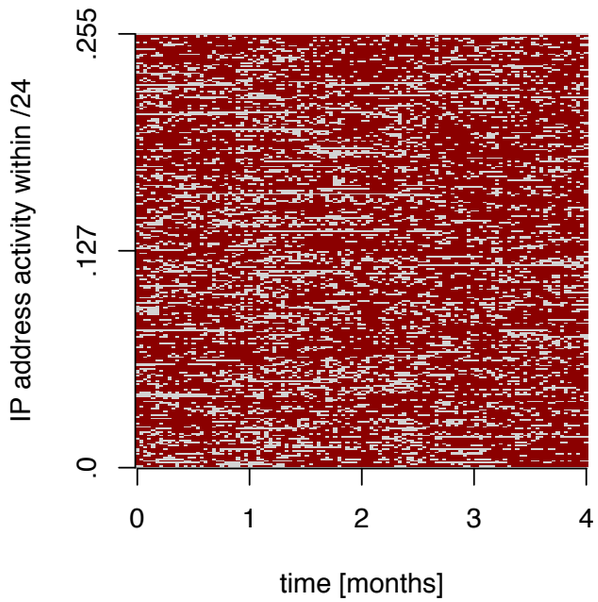
  ➡ **Internet Measurements need to be qualified by time period**

  ➡ **Challenge for operational systems (e.g., host reputation)**

Addressing Explains (some) Dynamics



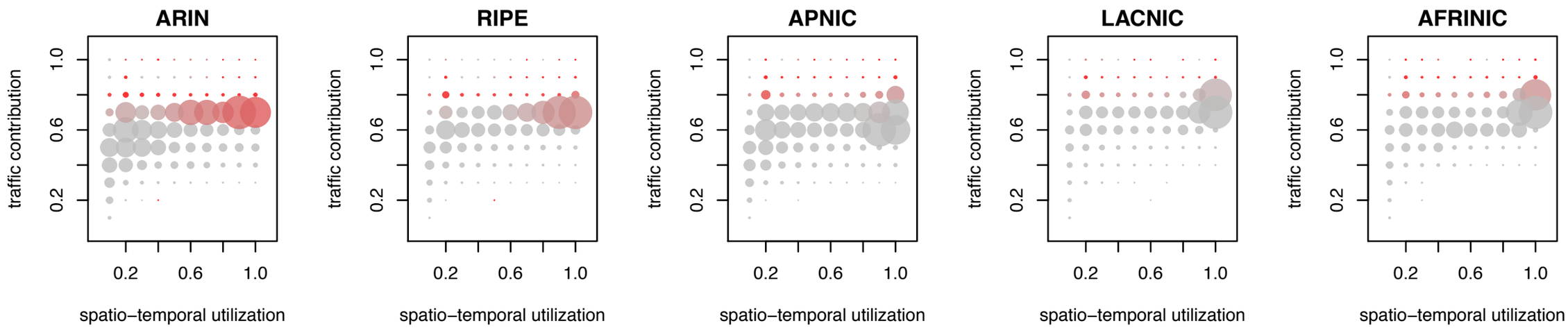




- **Effect of addressing mechanisms and network restructurings**

➡ **Inform measurements and operational systems**

IPv4 Space Utilization seen from a CDN



➡ **Lage portions with little utilization, efficiency gains possible**

➡ **Concentration of Web traffic on fewer heavy hitter IP addresses**

➡ **Address [re-] assignment practices, Internet Governance**

Thank you!