DNSFlow- Enabling netflow-like telemetry for DNS

Kyle Creyts, Manish Karir
Merit Network Inc.

Joe Eggleston, Craig Labovitz
DeepField Networks

Intro

- An analogy to (gasp) telecom networks:

- Data plane – scalable telemetry via Netflow
  - Operationally incredibly useful
  - Embedded everywhere
  - Well-developed analysis ecosystem

- Control (signaling) plane – DNS
  - Potential for broad **operational** visibility (beyond dns system itself)
  - Embedded nowhere (some syslog support)
  - Nascent analysis ecosystem

- DNS + Netflow…
  - Lends DNS-based context to analysis of data trafffic

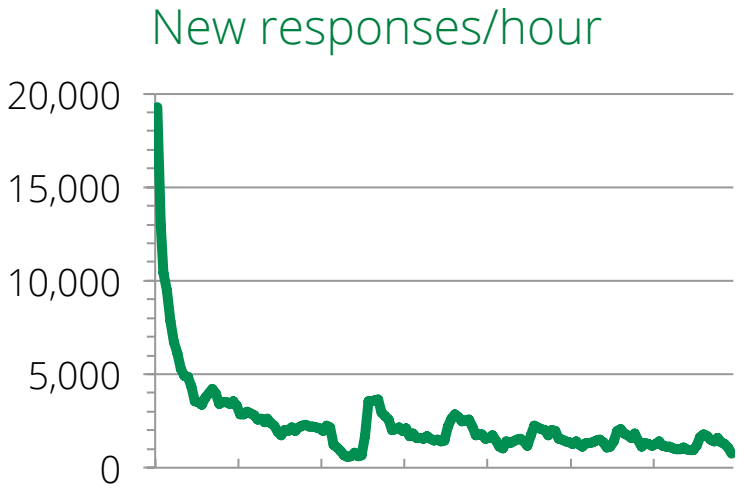
DNSFlow

- DNSFlow – protocol/reference implementation for efficient export of DNS responses
- Primary operational insight from **recursive responses**
  - Also compelling security applications for iterative responses (see ISC Passive DNS)

- Goal – Open access to data
  - Vendor-neutral, open, standards-based, cheap (free), efficient, lightweight, flexible deployment

Challenges

- Scale of data
- But, most queries redundant
  - 1:20 compression ratio possible
  - Even more with smart sampling



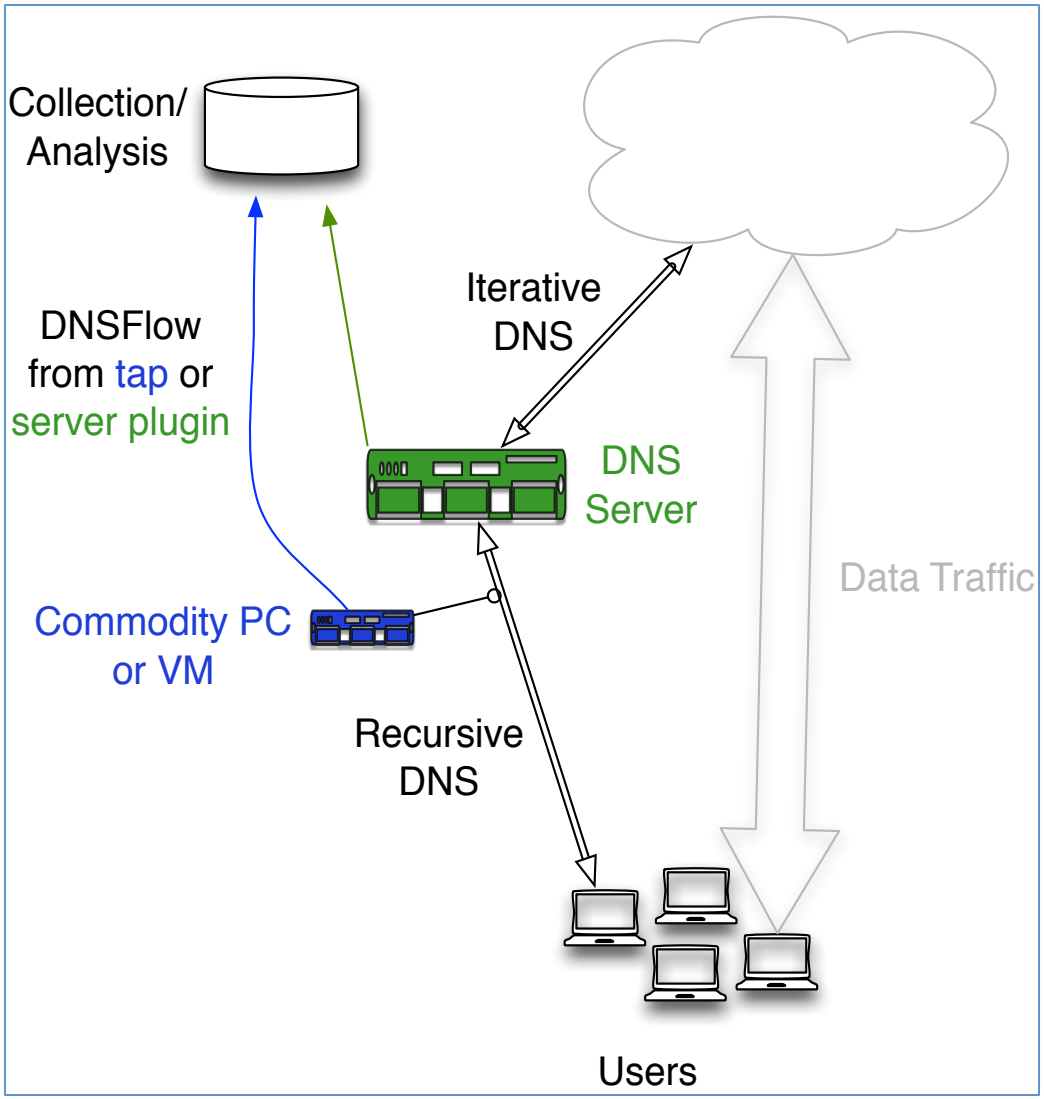
Status

- Open source, version 1
  - Simple (hundreds of lines of ANSI C)
  - Flow like output
  - NMSG (Passive DNS) interoperability layer
- Upcoming version IPFIX
  - Easy integration with existing tools.
- Support for compression, sampling and built-in algorithms identifying redundant queries
- Deployed in several production networks today

- https://github.com/deepfield/dnsflow

Deployment

- Runs on any *nix.
- Anywhere with visibility of recursive responses.
  - Recursive resolver
  - SPAN/mirror port of resolver
- Discussions towards embedding in open source and commercial DNS Servers



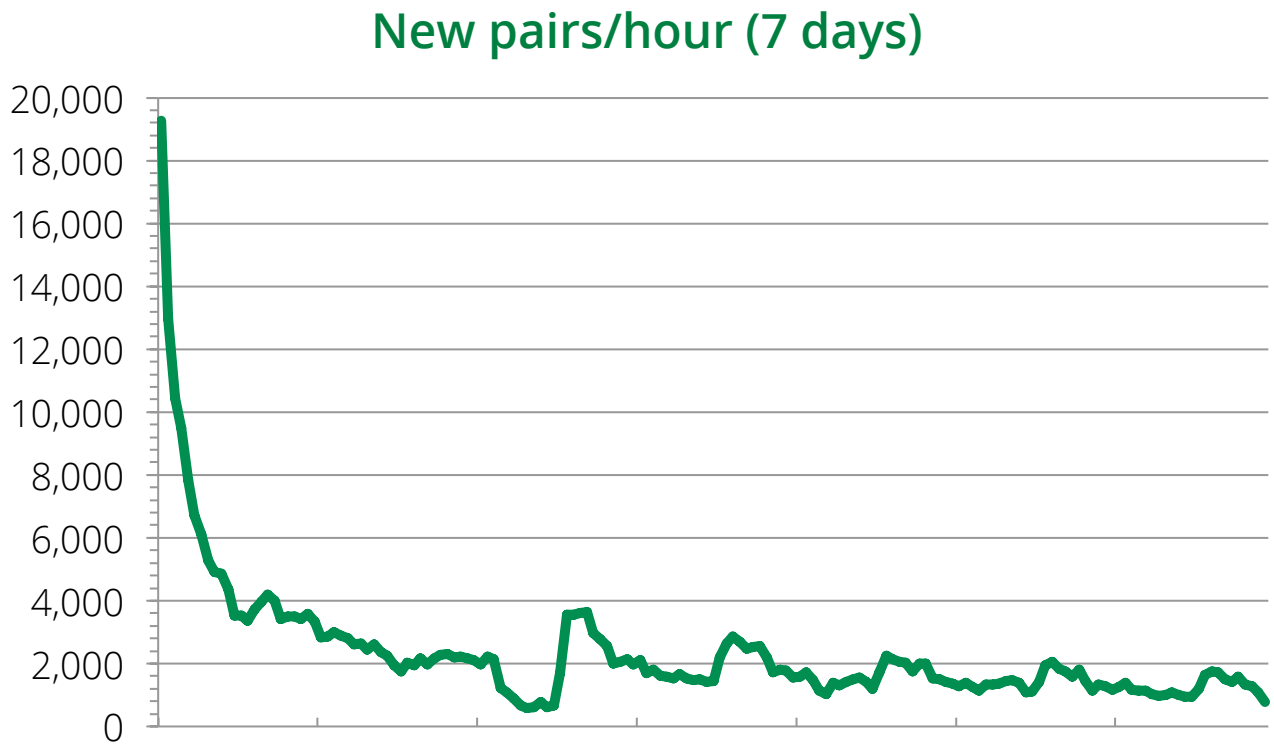
Using DNSFlow Data – Phishing

- Notification of phishing activity
  - used in conjunction with DNS blacklists such as **phishtank**, **malwaredomains.com**, or **SURBL**
  - when a user is successfully phished (opens a link) with a domain in the blacklist, it will incur a DNS lookup
  - that DNS answer will appear in DNSflow
  - can query DNSflow data, see which users (client IPs) were phished, and which domain(s)-IP pair they were victim to

alert_id	set_id	IP f/ response	client IP	timestamp	BL source	name from DNS
7907823	4819173	xxx. 235.133.xxx	10.1.1.2	2012-04-09 00:57:57.379013	malwaredomains	www.sexy-screen-savers.com.
7924562	4829607	xxx. 233.142.xxx	10.1.1.1	2012-04-09 05:45:08.815553	malwaredomains	www.meb.gov.tr.

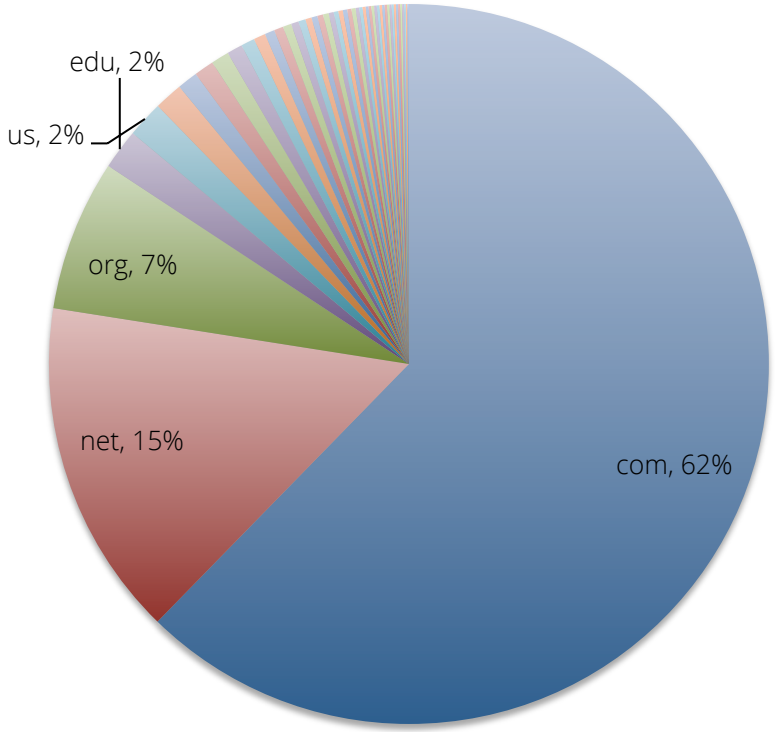
Building DNSMapDB

- Goal: To characterize the DNS-IP mappings as observed at a regional ISP
- Two different types of entries in DNSMapDB – largely-static and largely-dynamic
- After first 7 days we continue to learn new pairs but at a constant rate of roughly 1000pairs/hour



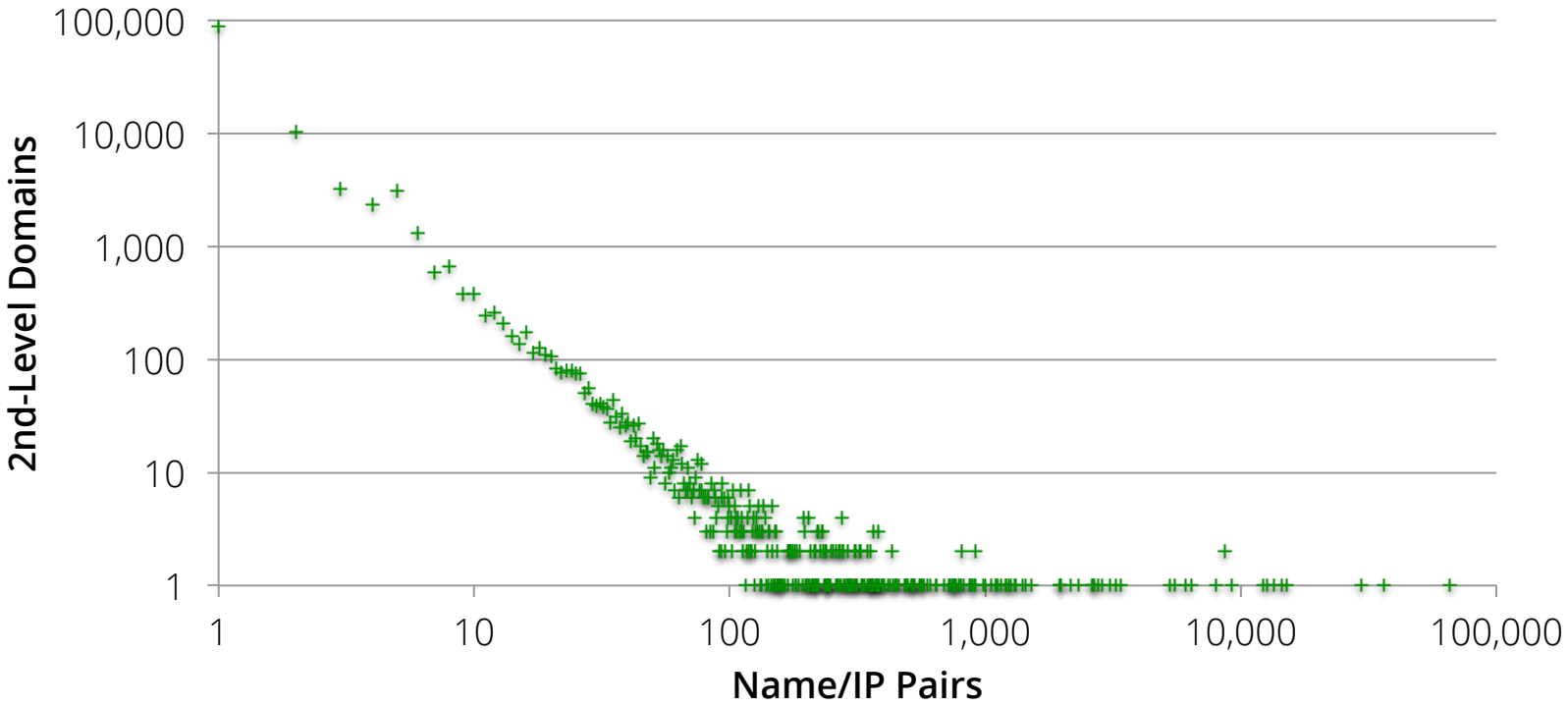
Characterizing DNSMAPdb

- There are two different contributors to DNSMapDB
  - Mostly static infrastructure
  - Largely dynamic mappings which originate from CDNs and massive web hosting – dynamic infrastructure
- Shortest name 2 character and longest 235, but average is 27 characters.
- Total of 308 unique TLDs
- Top 5 TLDs account for almost 90% of the DNS names observed



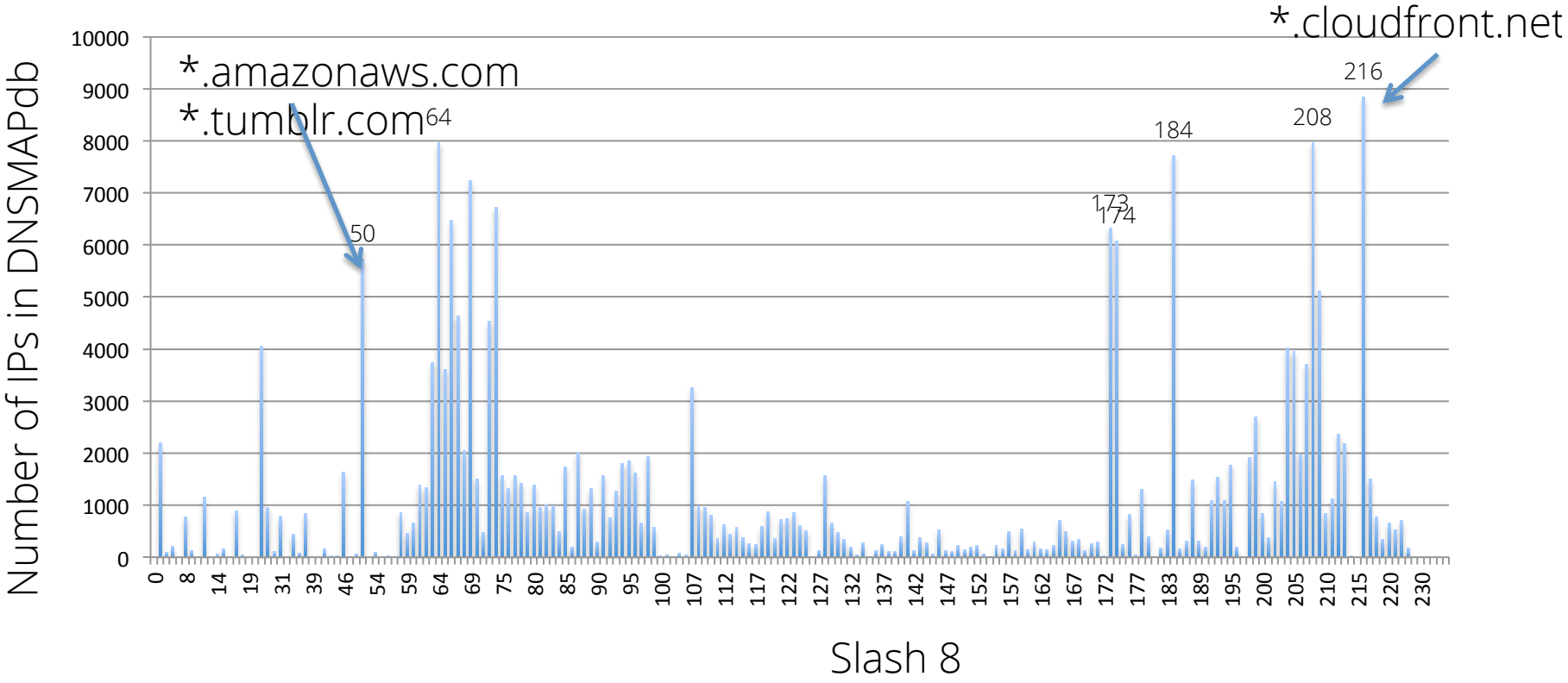
Second level domain distribution

- How big is a given second level domain – most have very few name-IP mappings < 60 second level domains have more 10 or more mappings
- Looking at the other end of the spectrum very few have large number of entries but how big do these get? – Atleast 40 have more than 1000 entries each - top 8 have 10K or more each



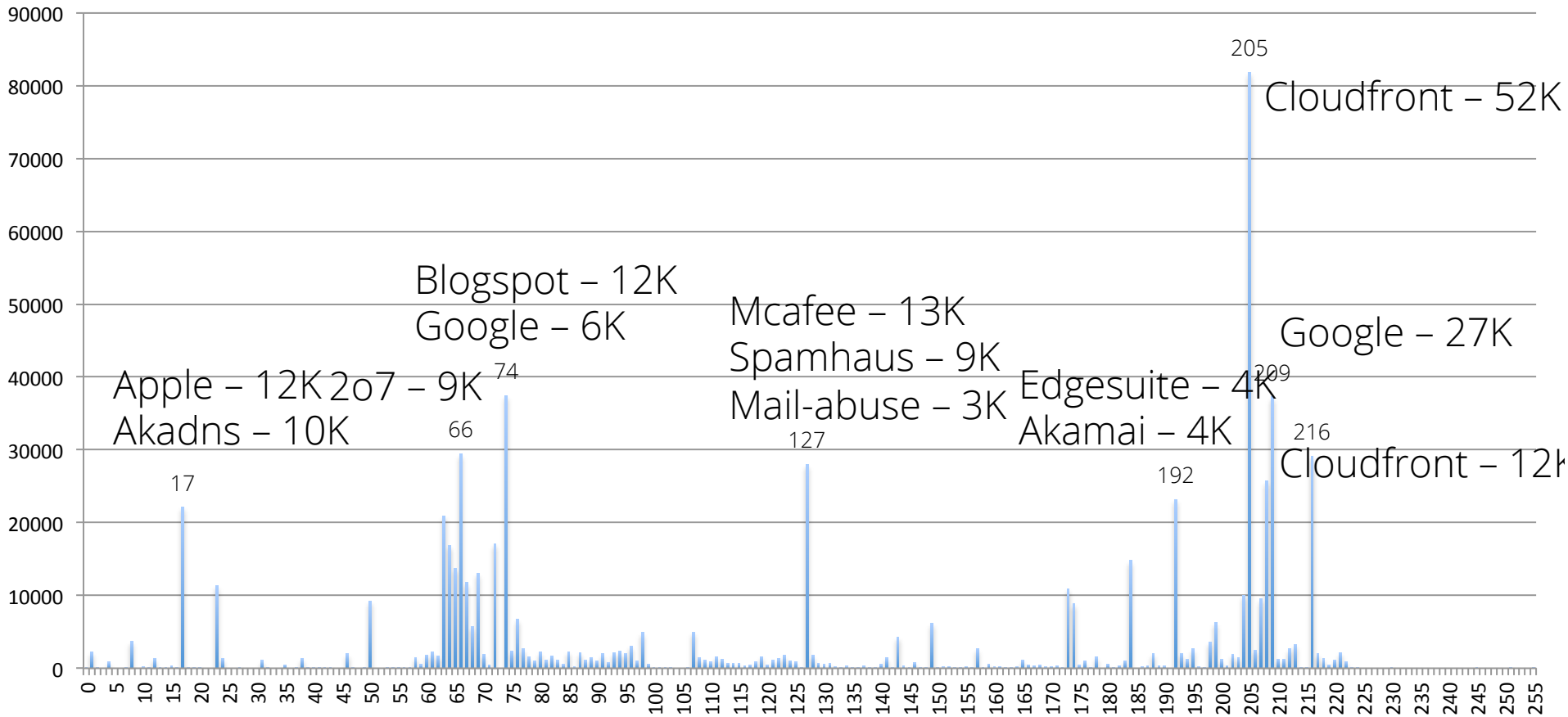
IP Address distribution by /8

- Distribution of 200K unique IPs in DNSMapDB by slash8



DNSMapDB Entries by /8

- Primary contributors to DNSMapDB entries



Thank You

Kyle Creyts <kcreyts@merit.edu>
Joe Eggleston <joe@deepfield.net>