

# Observations of IPv6 Addresses

David Malone <David.Malone@nuim.ie>

Hamilton Institute, NUI Maynooth, Ireland.

28 April 2008

## IPv6 Chat

- IPv6 talks mention NAT, CIDR and  $2^{128}$  addrs.
- NAT means you get more addresses.
- CIDR means you get more networks.
- 1000 hosts per gram of earth with  $HD = 0.8$ .

## IPv6 More Expressive?

- This means something for IPv6 addresses.
- No NAT: can see end host address.
- Standard(-ish) boundaries at /64 (and /48).
- Many addresses facilitates logical assignment.

## Some examples

- 2001:0770:0010:0300:0000:0000:86e2:510b
- 2001:770:10:300:0:0:86e2:510b
- 2001:770:10:300::86e2:510b
- 2001:770:10:300::134.226.81.11
- fe80::21e:52ff:fec8:84b2

## Plan

- Automate assignment of attributes.
- Collect sets of IPv6 addresses.
- See what patterns of usage look like.
- Datasets: mirror server, .ie ccTLD server, traceroute6.

## Previous IPv6 Work

- CAIDA: topology measurements.
- Huston/Döring/Massar: BGP routing studies.
- Savola/Kei/Yamazaki: 6to4 traffic.
- WIDE: Traffic data collection.
- Cho/Luckie/Huffaker: IPv4/IPv6 relative performance.
- Bieringer: `ipv6calc`.

## Observing Network Part

- Like IPv4: registry based.
- Global addresses to RIRs.
- Also: 6bone, 6to4, teredo.
- Special addresses: loopback, unspec.
- Special blocks: mapped, ULA, link-local, site-local, multicast.

## Observing Host Part

- Autoconf: look for ff:fe and set bit.
- ISATAP: Look for 0[02]00 and 5efe.
- v4based: last 32 bits look like v4 address.
- low: only low byte set.
- wordy: feed:deb:dead:c0de
- privacy: few words and large numbers.



# Words

00ad 00ba 00be 00d0 00da 00ed 0ace 0ada 0add 0ade 0b00 0b0a 0b0b 0baa 0bad 0bea 0bed  
0bee 0c00 0c0b 0c0d 0cab 0d0b 0d0c 0d0d 0d0e 0dab 0dad 0deb 0dee 0ebb 0f00 0f0b 0f0d  
0f0e 0fad 0fae 0fed 0fee abba b00b b0b0 b0de baba babe bade baff bead beef c0c0 c0ca c0d0  
c0da c0de c0ed c0ff cafe cede d00b d0d0 d0de dada dead deaf deed f00d f0ad face fade faff  
feed 1337 0000 1111 2222 3333 4444 5555 6666 7777 8888 9999 aaaa bbbb cccc dddd eeee  
ffff 00ff abab

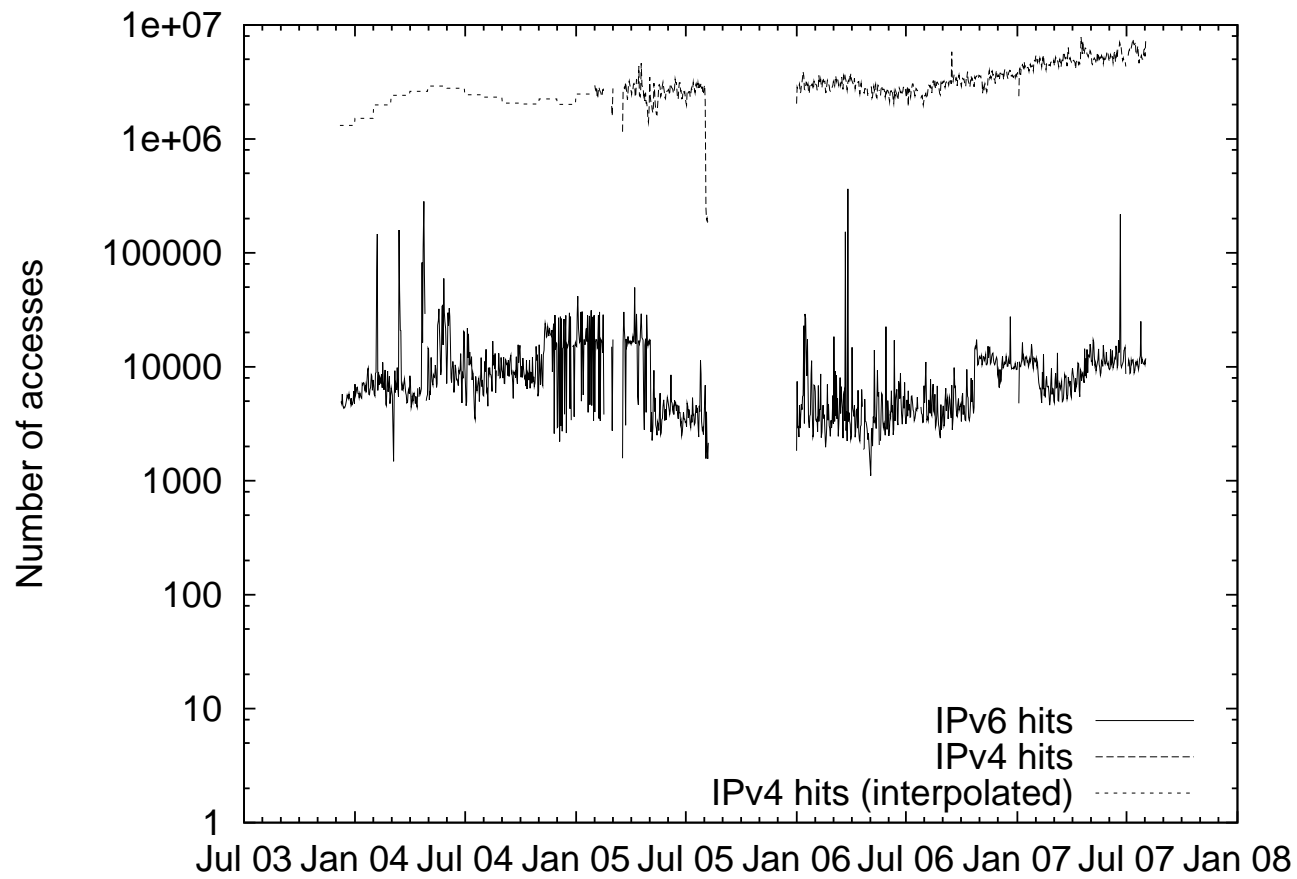
# Privacy

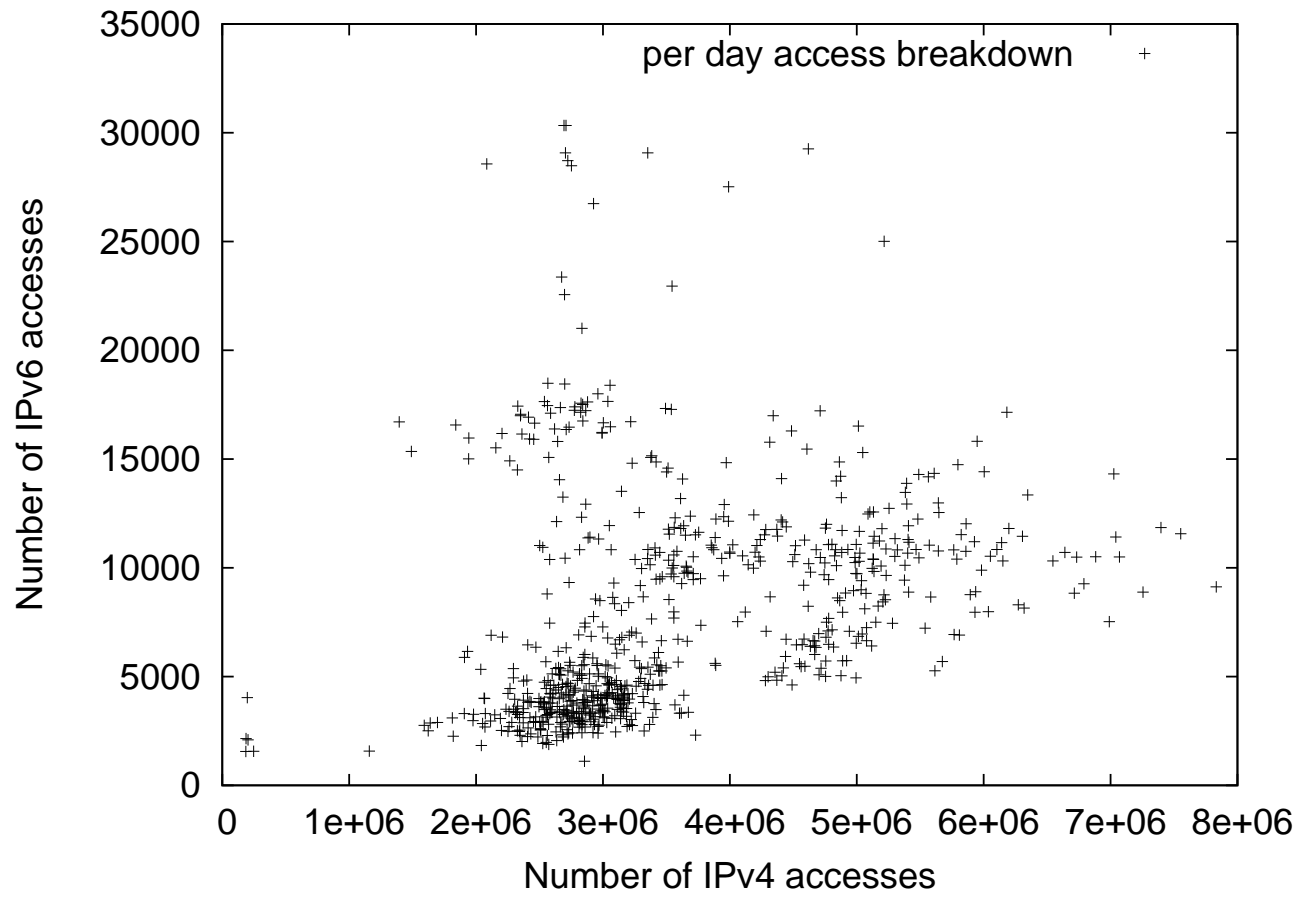
- have the 6th bit clear,
- have between 27 and 35 set bits,
- first half has between 9 and 21 set bits,
- second half has between 10 and 22 set bits,
- must not have two or more ‘words’.

$$\frac{1}{2^{63}} \sum_{\substack{9 \leq i \leq 21, 10 \leq j \leq 22 \\ 27 \leq i+j \leq 35}} \binom{31}{i} \binom{32}{j} \approx 0.7335. \quad (1)$$

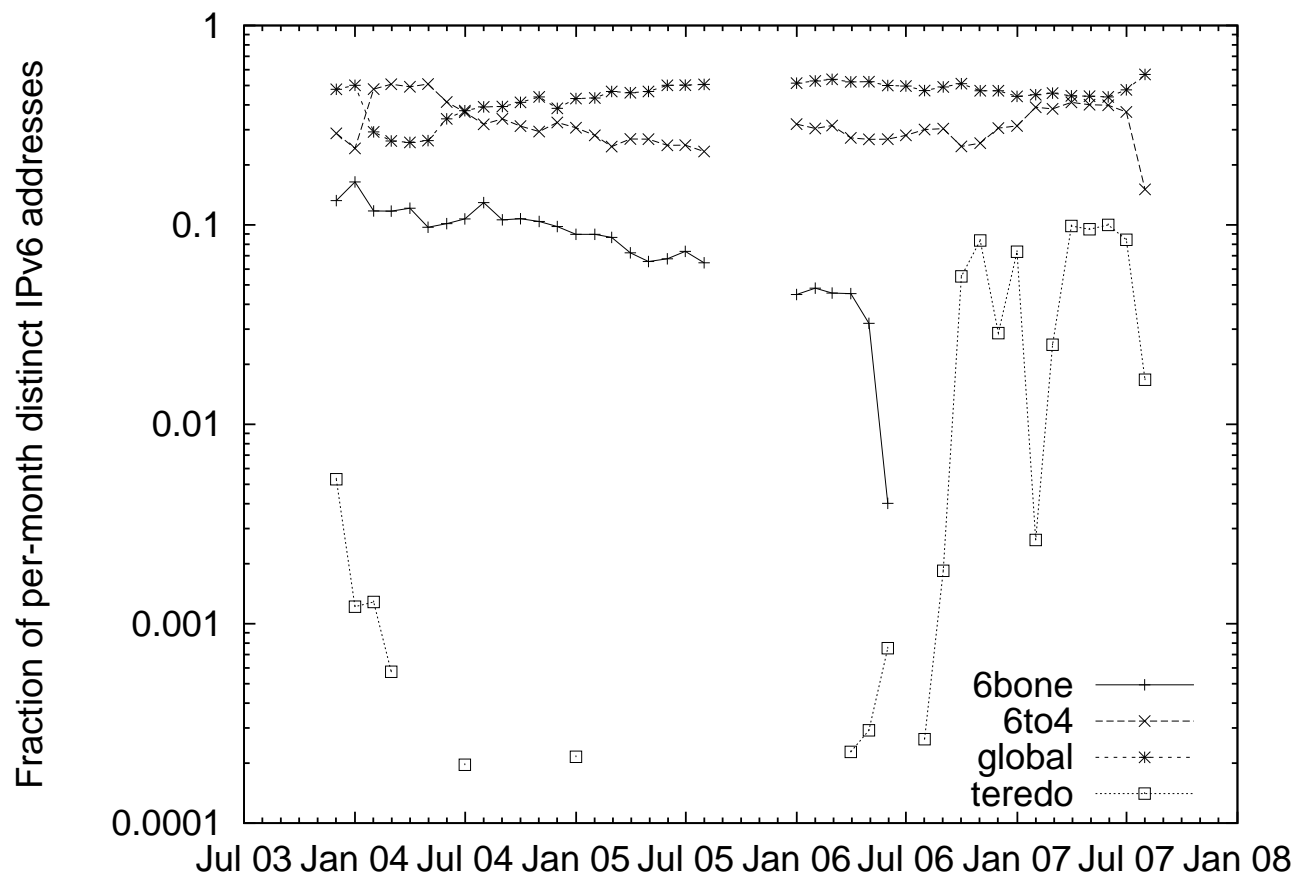
## Dataset: ftp.heanet.ie

- Busy mirror server (sourceforge, Linux distros, putty,)
- Data from Dec 2003 to Aug 2007: over 1300 days.
- Mostly full Combined log file format.
- Some gaps — how to normalise?

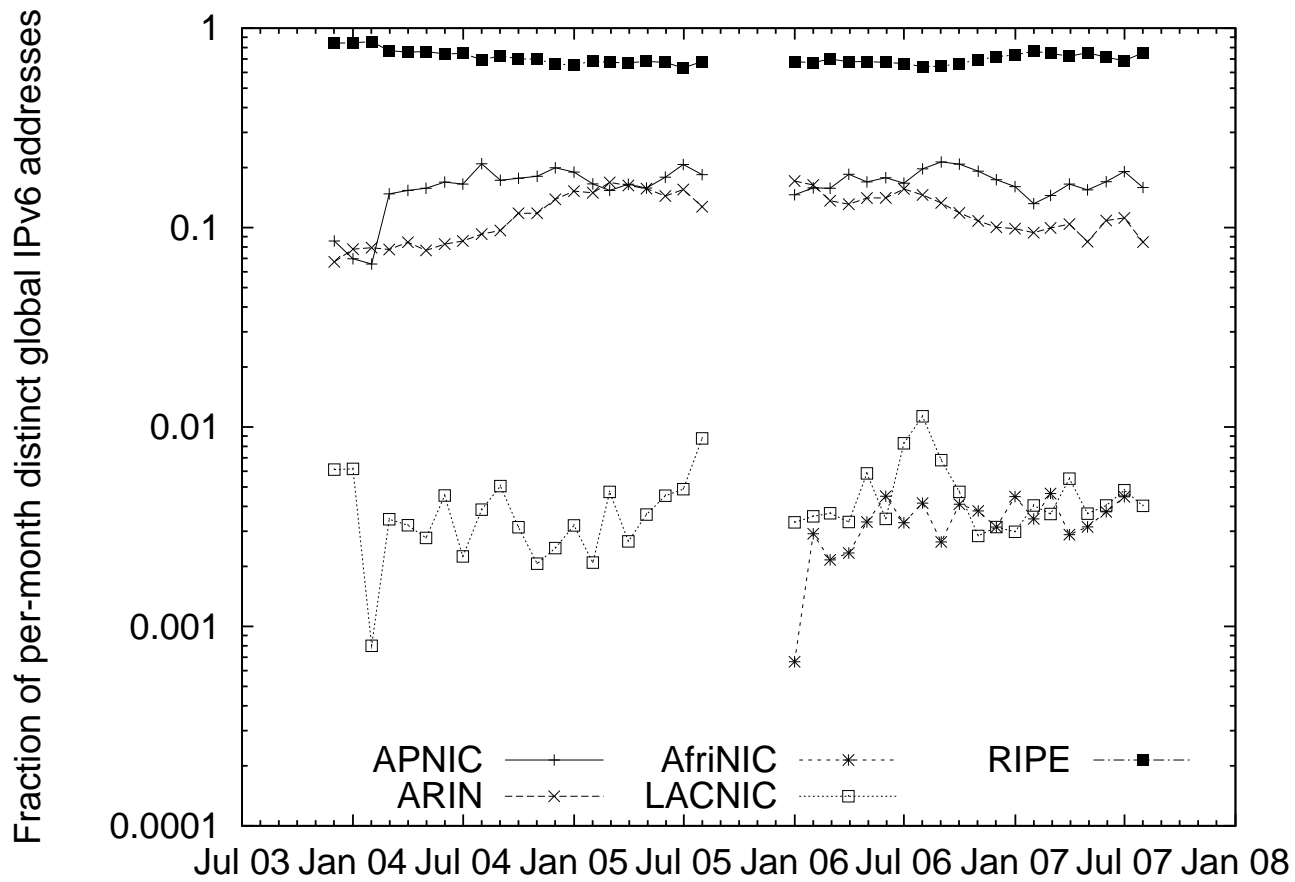




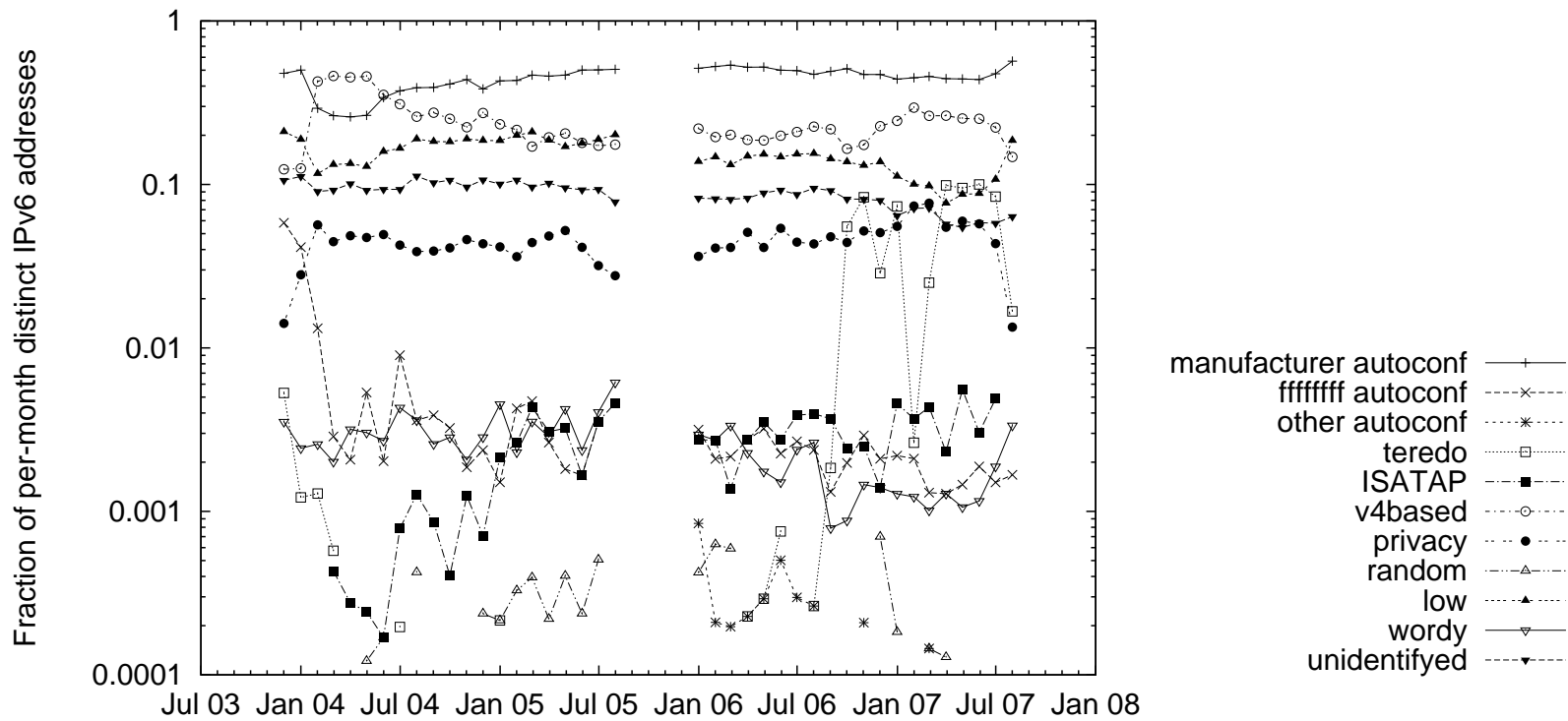
# Results: Prefix



# Results: Global by RIR



# Results: Host ID





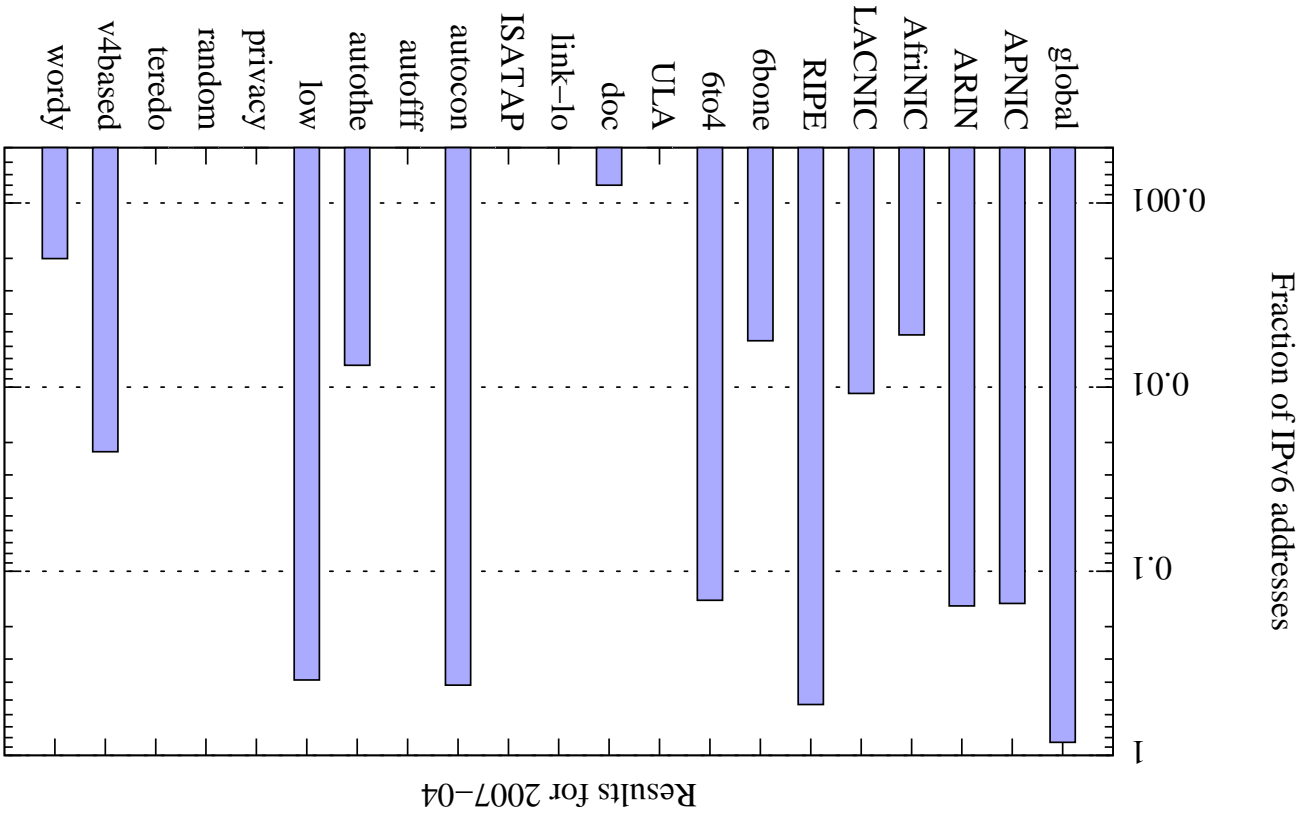
## Following Hosts

- 38495 different autoconf host IDs.
- 3304 in more than one subnet.
- Tend to be 6to4 prefixes.
- Only 148 moved more times than seen.
- Three moved regularly: 46652, 26107 and 2598.
- Looks like multihoming.

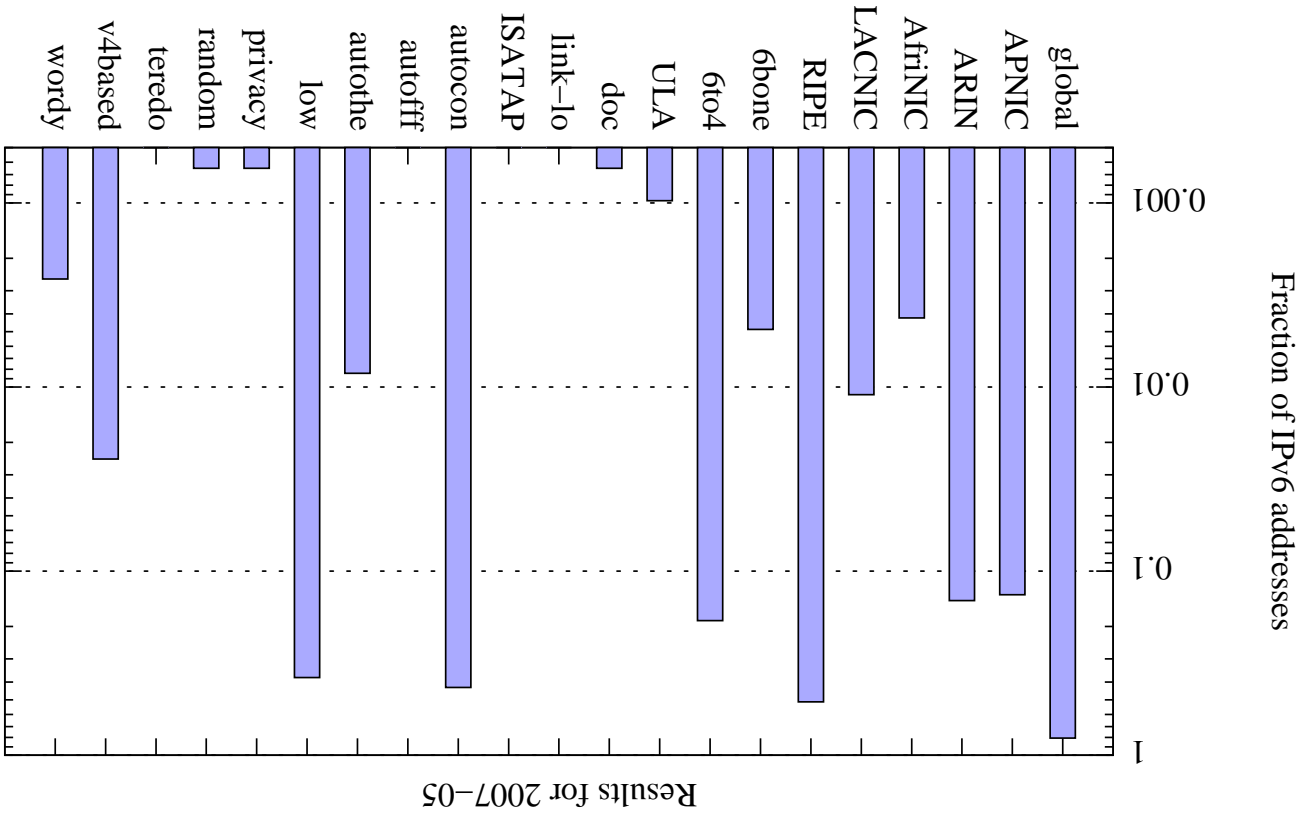
## IEDR Data

- Two months of data from ccTLD.
- Server both IPv4 and IPv6.
- Only IPv6 analysed.

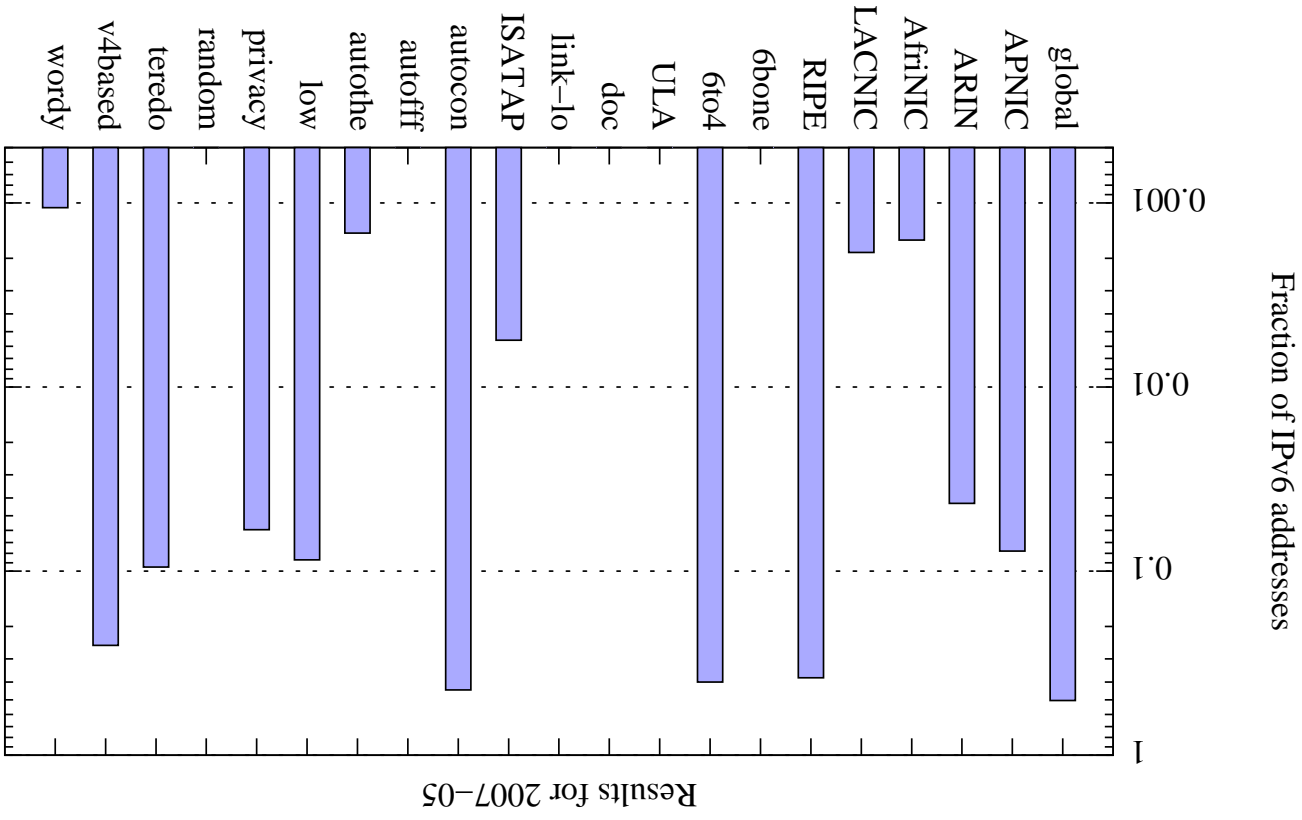
# Results: IE DR 2007-4



# Results: IE DR 2007-4



# Results: HEAnet 2007-5



## Conclusions

- Working techniques address analysis.
- We can get more from IPv6 host ID.
- We see differences across groups.
- We see trends across time.
- We see consistence accords measurement points.

## Future Work

- Subnet allocation within /48.
- Using co-located hosts to improve results.
- Autoconf tracking and vendor analysis.
- Anonymisation question?
- Useful for service adaptation?
- Explicit validation (privacy overestimate).