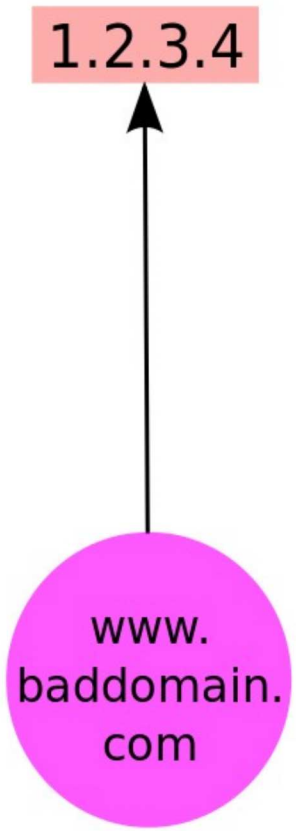
Modeling DNS Agility with DNSMap

… to enable malware detection

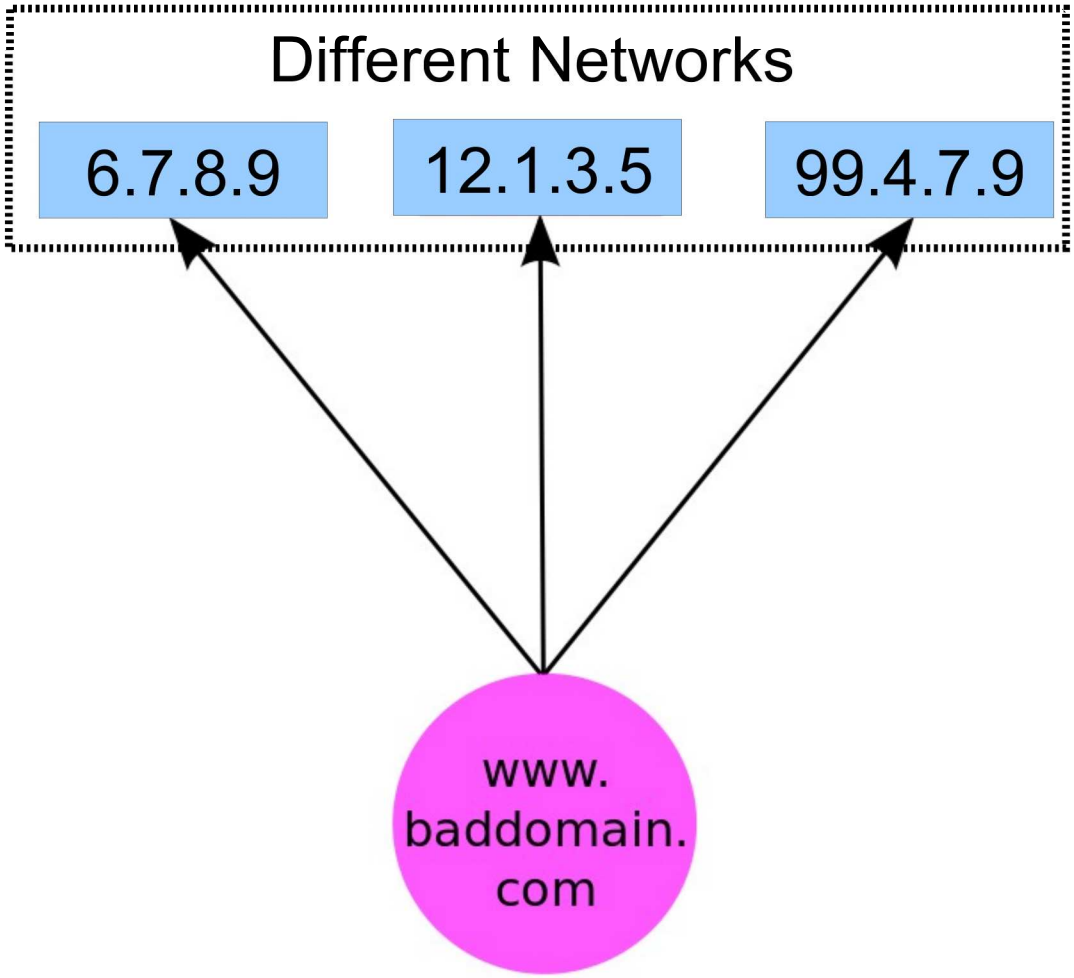
Andreas Berger, Wilfried Gansterer

berger@ftw.at

Malware DNS Activity (1)

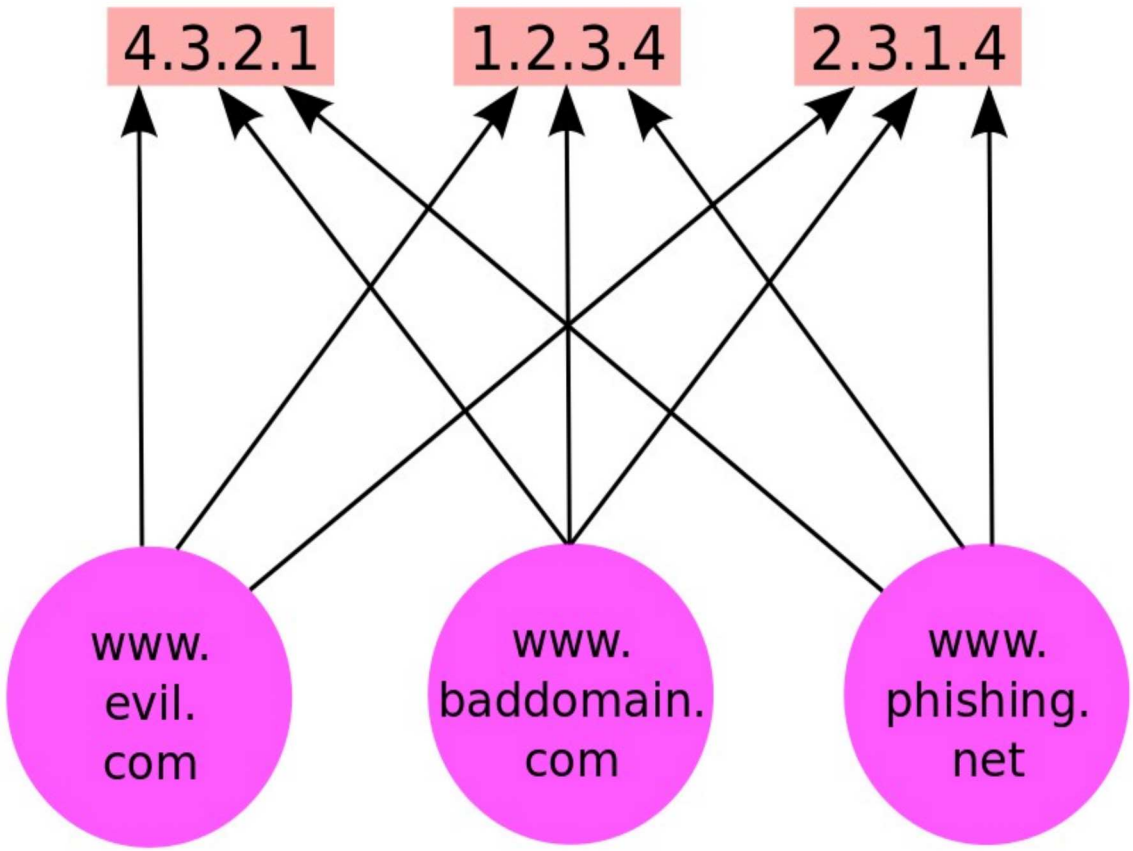


Malware DNS Activity (2)



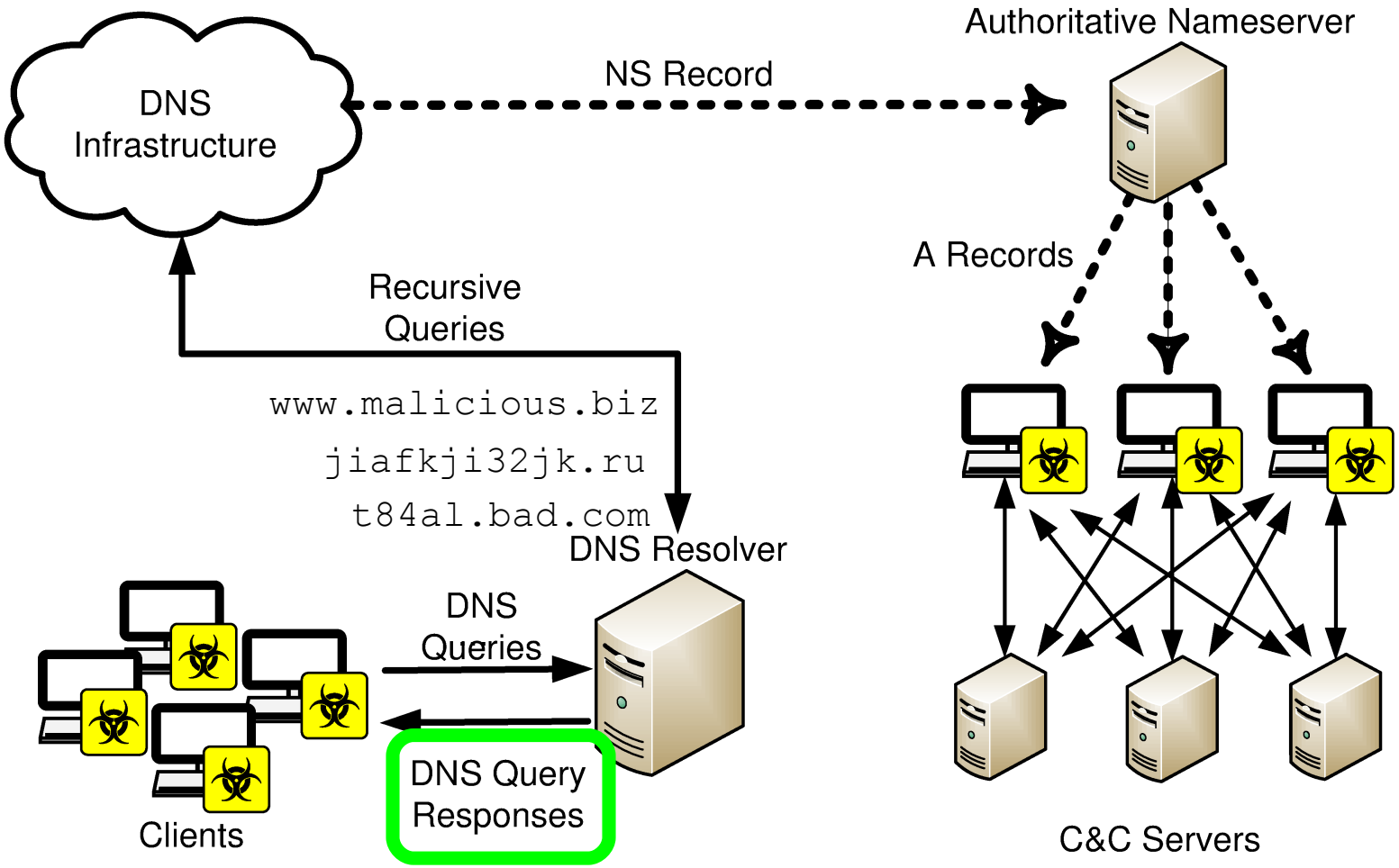
Redundant Hosting & IP Flux

Malware DNS Activity (3)



IP Address Reuse

Malware DNS is often *agile*



Existing Approaches

1. Get DNS query responses
2. Extract features + assess DNS agility
   - #A-records, FQDN entropy, #ASes, TTL, ...
3. Anomaly detector (e.g., machine learning)
4. Detection results: www.malicious.biz is bad!

- E.g., Antonakakis et al.: NOTOS; Bilge et al.: EXPOSURE; Perdisci et al.: FluxBuster

Agile == Malicious?

- No! → CDNs, cloud services, data centers, ...

More complications

- Wildcards: <random-prefix>.example.com
- Limited visibility: mysite.blogspot.com @ 5 IPs
- Constant changes: new services, others disappearing

- But: benign services show **some** kind of stability [TMA'12]
  - Some DNS mappings are highly stable
  - Some use (fixed) **range** of IPs + stable FQDN **pattern** (*.example.com)
  - What counts is relation between FQDN and IP
  - Are benign mappings **more** stable than malicious ones?

Objectives

- Understand which services map where
  - What are **typical** FQDNs for a **specific** IP address?
  - When this is known, we find truly agile mappings (rather) easily
- Realtime detection of significant changes
  - What's significant:
    - New FQDN-IP mapping
    - difference(FQDN, previously seen FQDNs at this IP) > threshold

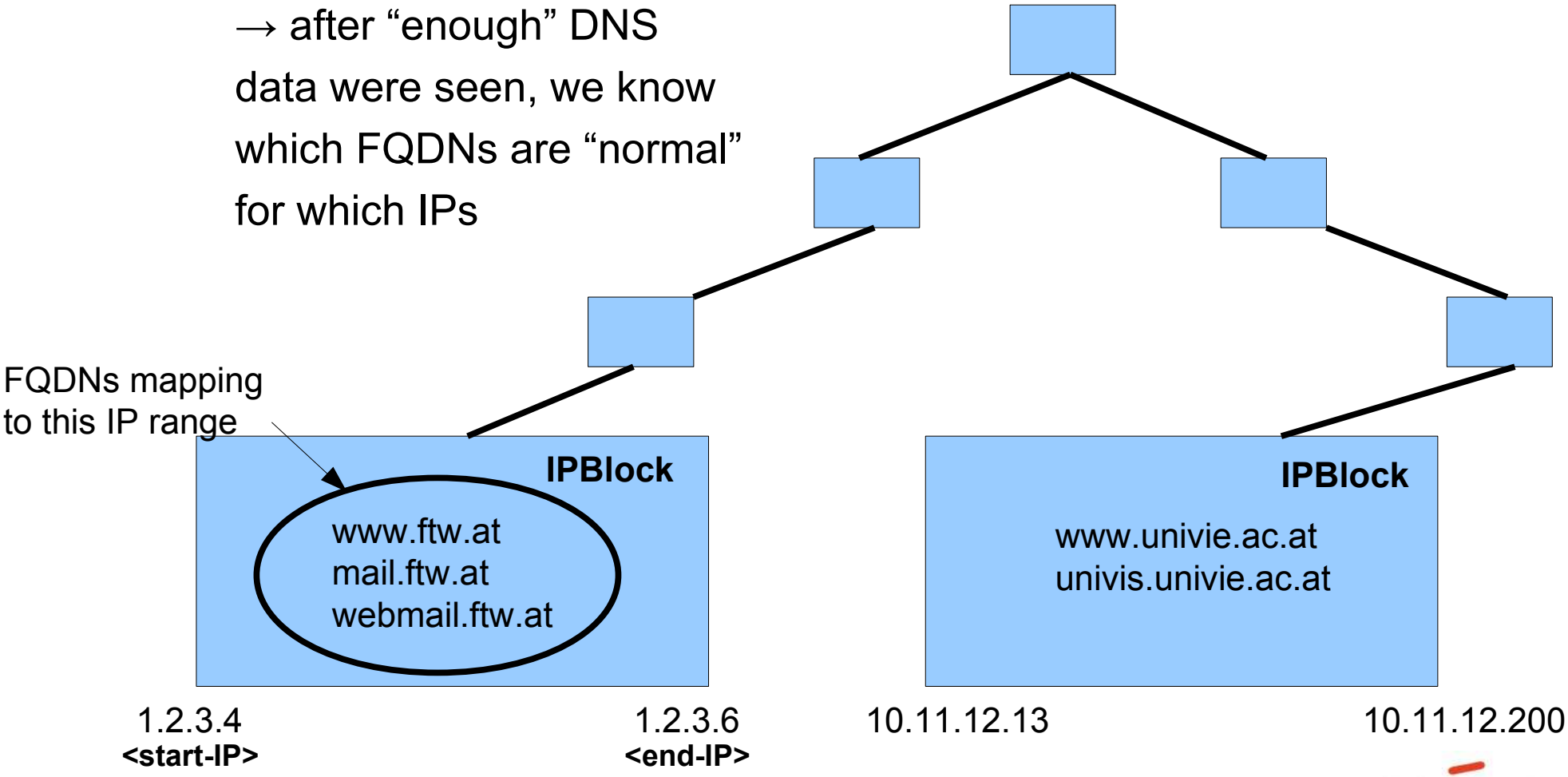
Domain Divergence [0,1]

	d_1	d_2	d_3	d_4
d_1=www.ftw.at	0.0	0.38	0.72	0.93
d_2=mail.ftw.at	0.38	0.0	0.89	0.84
d_3=www.facebook.com	0.72	0.89	0.0	0.50
d_4=01371354742.67.channel.facebook.com	0.93	0.84	0.50	0.0

DNSMap: Extracting DNS Mappings

- Idea: build a FQDN-IP map

  → after “enough” DNS data were seen, we know which FQDNs are “normal” for which IPs



IPBlock Example

- From 74.125.100.**16** to 74.125.100.**39**
- Domain clusters for this IPBlock:

v1.lscache3.c.youtube.com: [v11.lscache3.c.youtube.com,v15.lscache3.c.youtube.com, v6.lscache3.c.youtube.com,v21.lscache3.c.youtube.com, ...],

safebrowsing-cache.google.com: [safebrowsing-cache.google.com],

v1.lscache3.googlevideo.com: [v1.lscache3.googlevideo.com]

Cluster labels

Clustered FQDNs

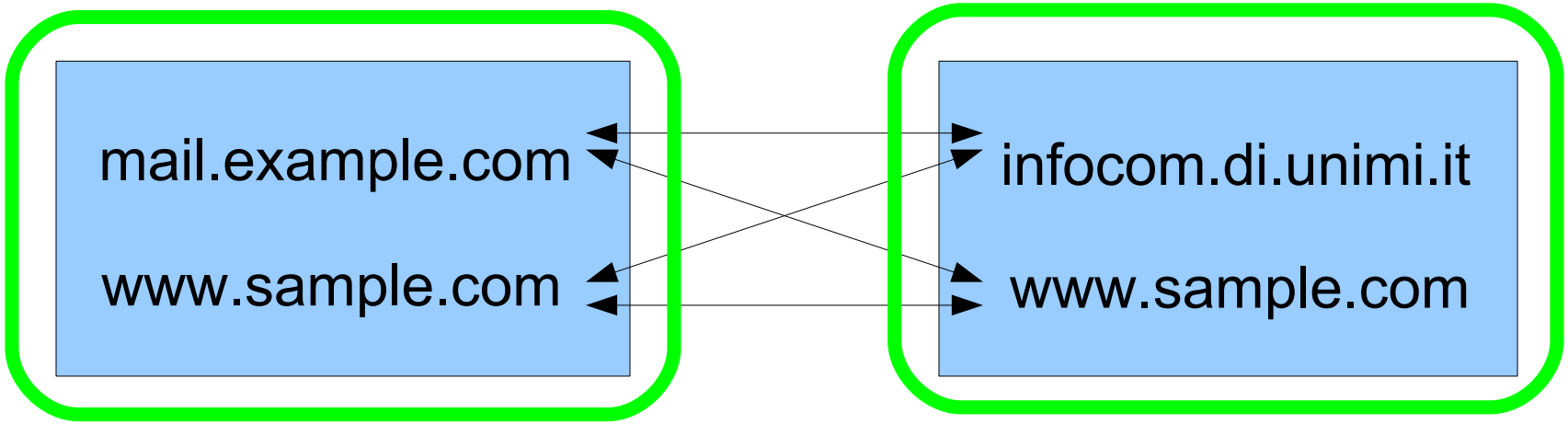
≤ threshold

Evaluation

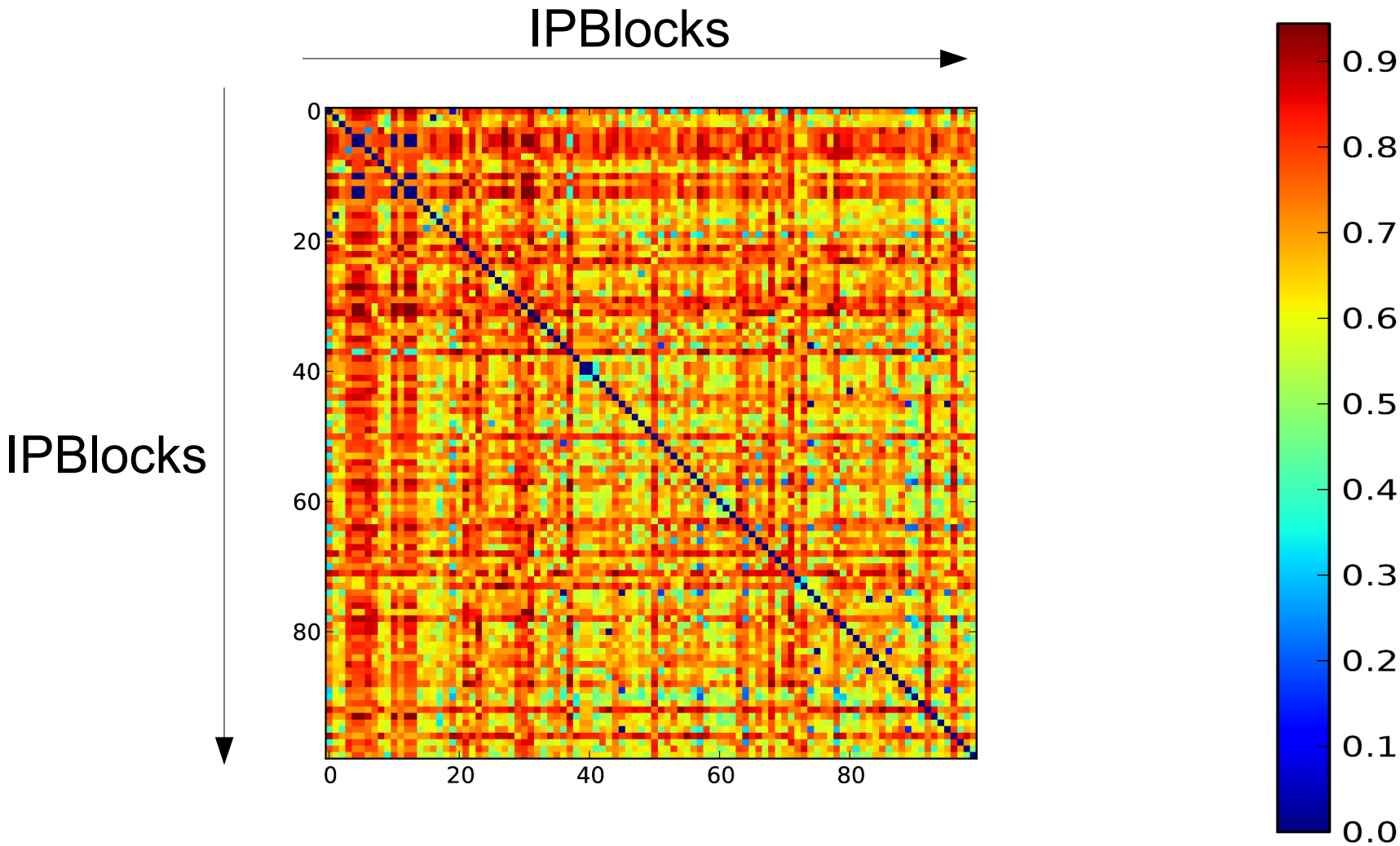
- Some details first
  - IPBlocks are merged and split as we go on
  - Remove FQDNs not seen mapping to an IPBlock within **two days**
- Data set
  - 2 weeks of DNS traffic from access provider
  - ~100k customers, ~1.1 Billion query responses
  - Processing takes around 8 hours
- → Evaluate DNSMap representation after these 2 weeks
  - 475k IPBlocks, 2.3M FQDNs

Evaluation

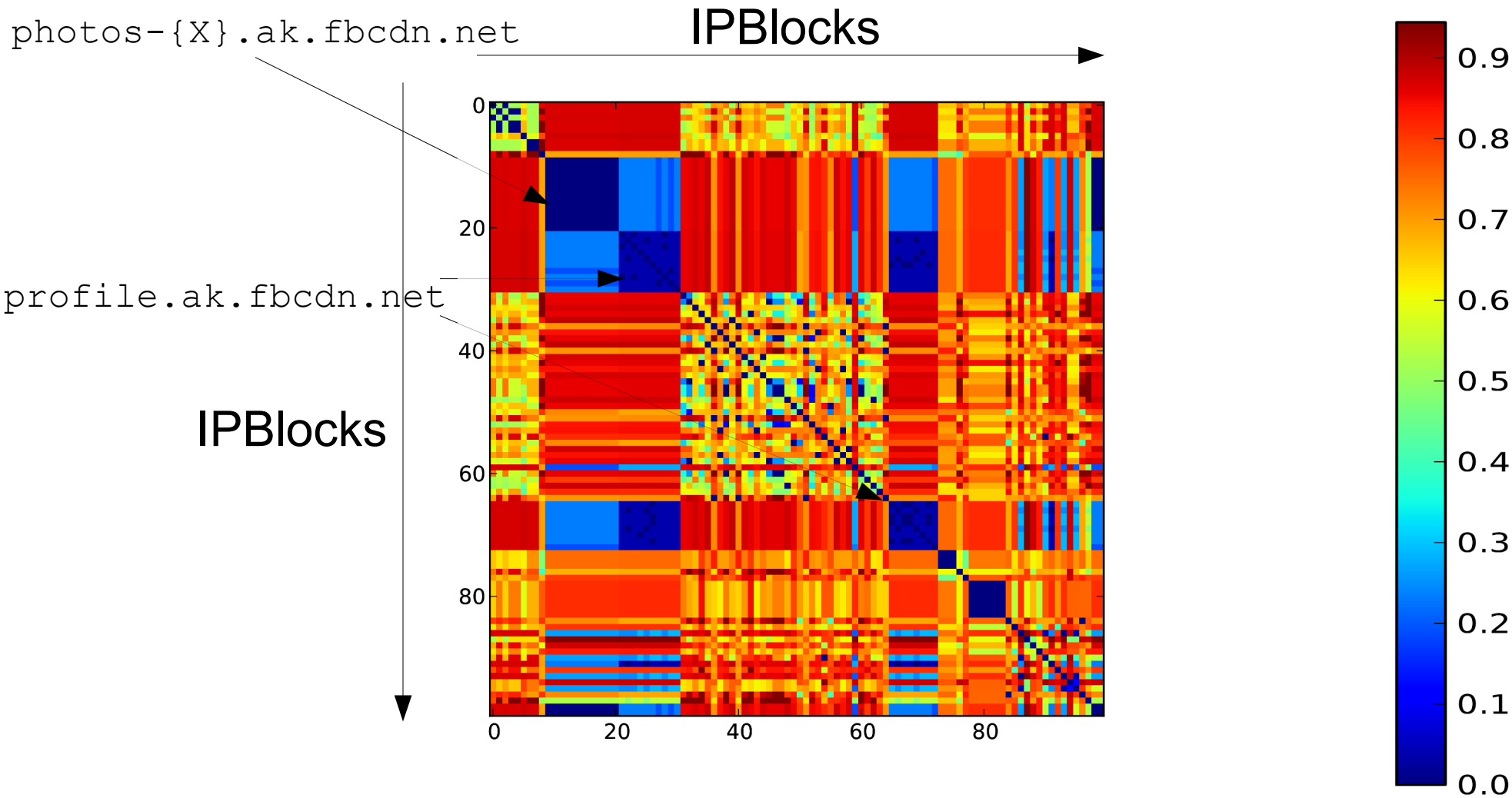
- Goal: understand how similar IPBlocks are to each other
  - Are there sub-ranges for particular services?
  - Or are large networks a big mess where everything maps everywhere?
- IPBlock mutual similarity = $\frac{\mu + \omega}{2}$



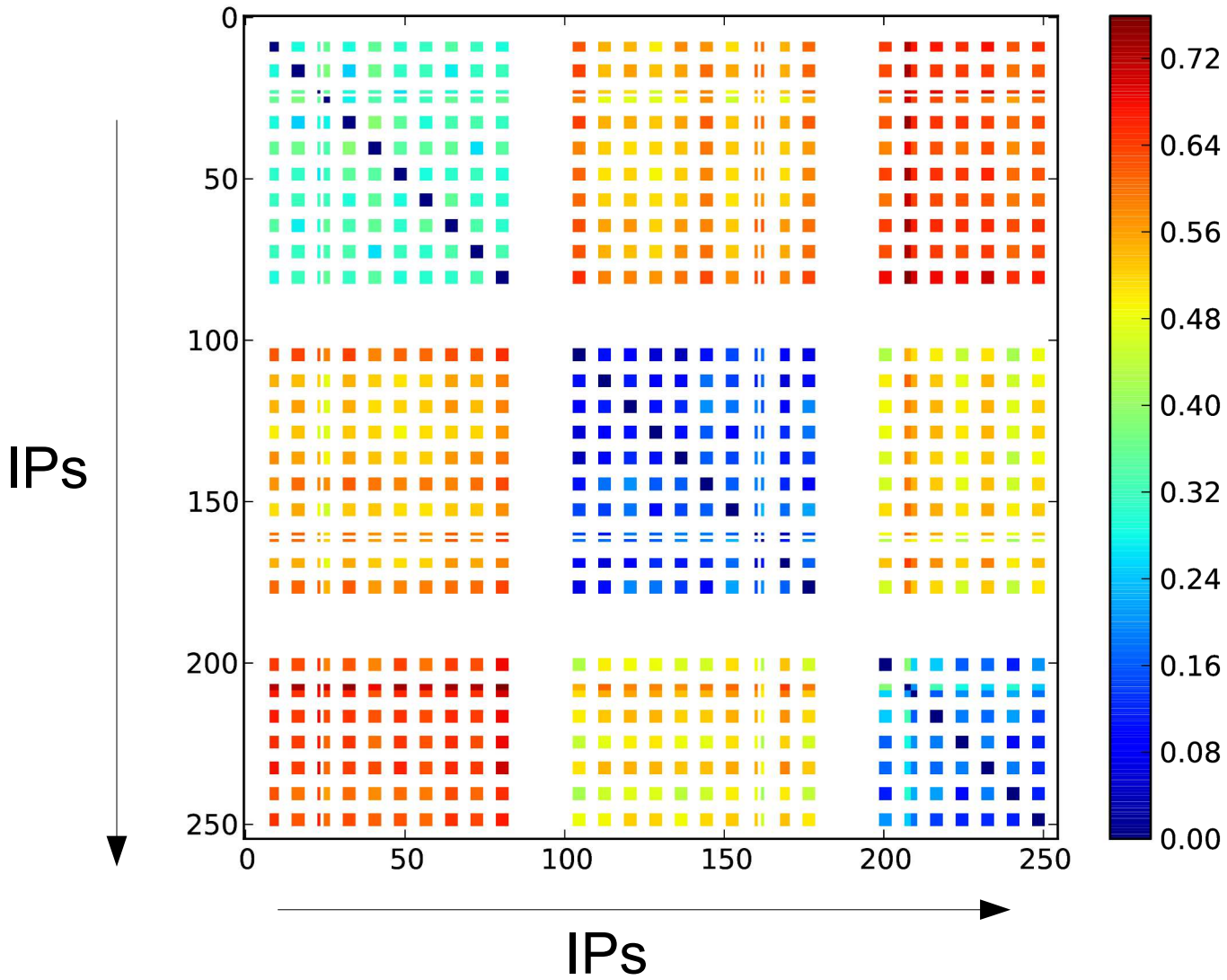
Amazon (101 IPs)



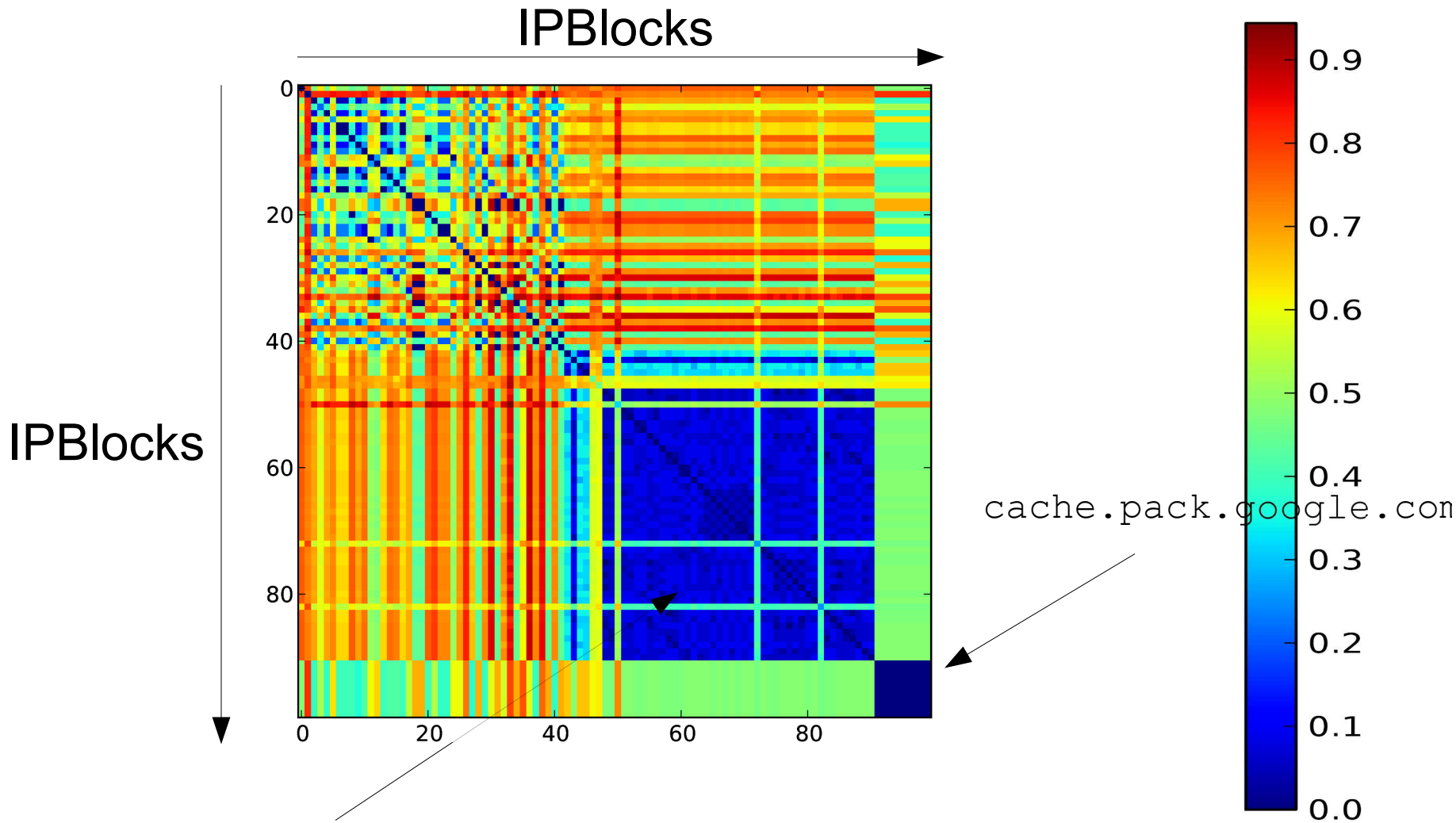
Akamai (164 IPs)



Akamai IPs (one /24)

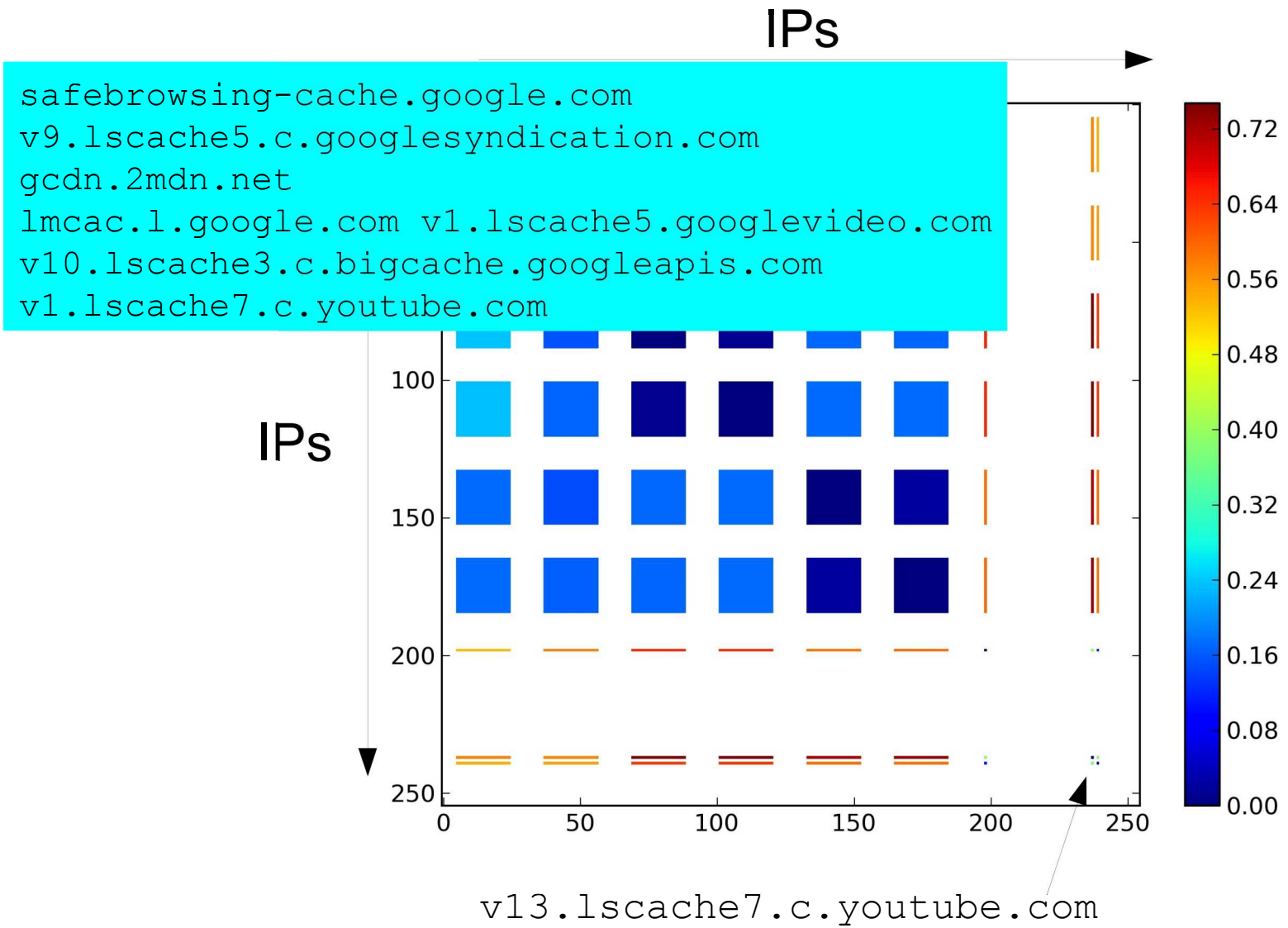


Google (271 IPs)



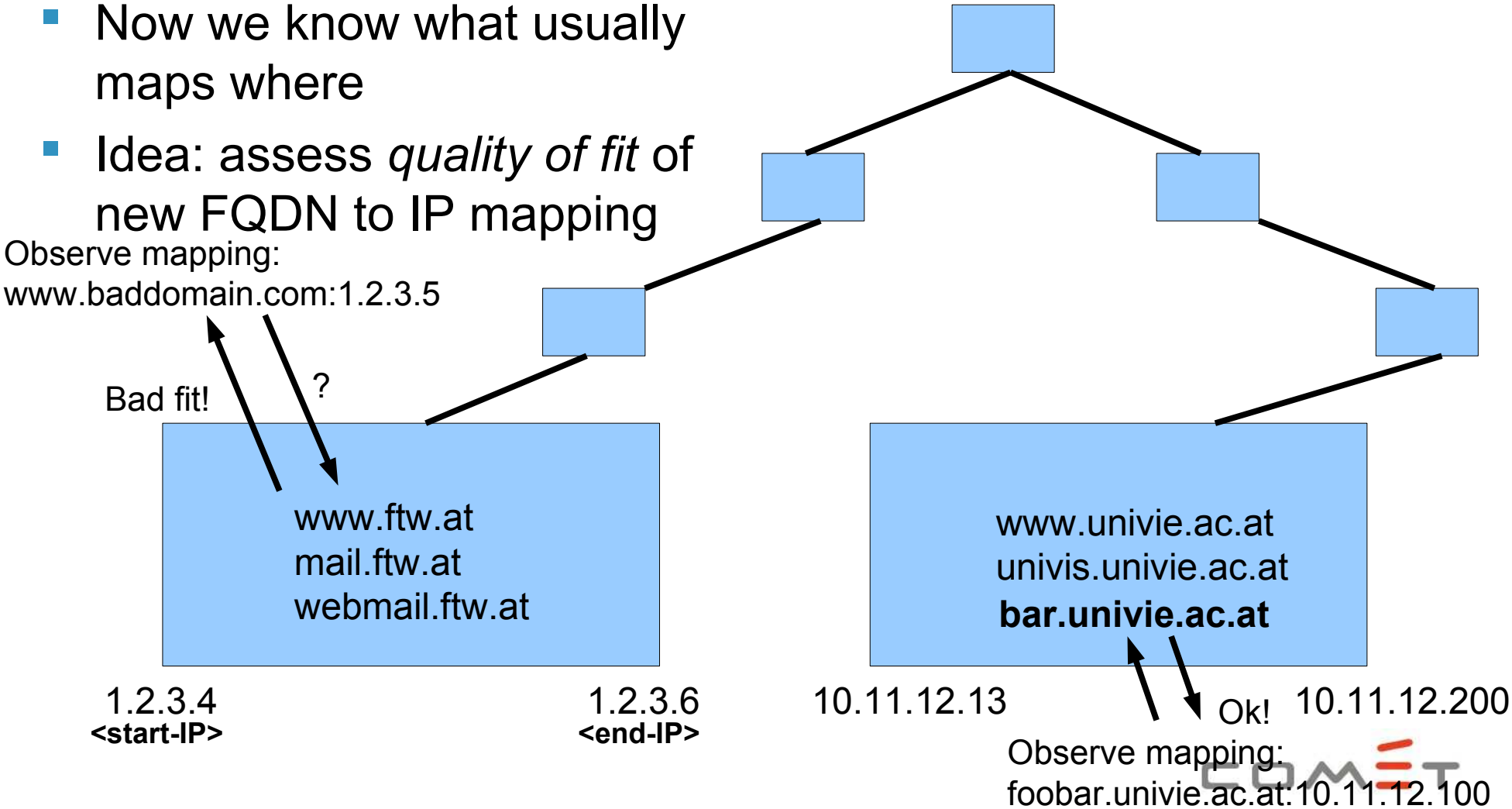
v{N}.lscache{N}.c.youtube.com

Google (one /24)



Detecting significant changes

- Now we know what usually maps where
- Idea: assess *quality of fit* of new FQDN to IP mapping



Example: (many!) change events

Timestamp | domainName | IP | divergence to closest cluster

- 1321884742 www.getontracks.org 72.2.114.102 0.663557628248
- 1321884742 www.cvo6.de 217.119.54.152 0.489795918367
- 1321884742 dcubed.ca 216.119.134.2 0.729411764706
- 1321884742 actioncomplete.com 69.89.12.255 -1
- 1321884742 www.magaze.it 62.149.142.46 0.452898550725
- 1321884742 www.extgtd.de 87.237.122.170 0.548433048433
- 1321884742 igtd.pl 212.85.96.95 0.661016949153
- 1321884742 www.gtd-php.com 69.163.249.44 -1
- 1321884742 www.simplegtd.com 184.173.73.183 0.632277834525
- 1321884742 kinkless.com 64.13.192.116 0.692348849652
- 1321884742 www.davidberman.com 64.26.179.14 -1
- 1321884742 www.bat-groupware.at 81.223.238.66 0.483630952381
- 1321884742 www.rory.de 82.165.97.3 0.675562969141
- 1321884743 www.bulawayo24.com 46.32.227.120
- 1321884743 www.african-bulletin.com 206.188.192.232 0.626077286628
- 1321884743 www.slatch.de 85.13.134.12 0.54347826087
- 1321884743 www.thinkingrock.com.au 113.20.8.41 0.576311988445
- 1321884743 www.nms20.at 85.13.131.16 0.441919191919
- 1321884743 peters-dahlien.at 88.198.125.2 0.553140096618
- 1321884743 www.feuerschutztuer.net 91.186.20.31 0.624017957351

A glimpse at malware detection

- Further analyze last 24h of events
  1. Remove all FQDNs which mapped to only one AS
  2. Find IPs with most alerts
  3. Manually check top-10 (*projecthoneypot.org*)

Type	μ	# FQDNs	Type	μ	# FQDNs
Honeypot?	1.0	61	Malicious	1.0	10
Honeypot?	1.0	61	Legitimate	**373.1**	8
Malicious	1.08	12	Legitimate	**207.4**	5
Malicious	1.09	11	Malicious	1.1	10
Malicious	1.11	10	Malicious	1.0	10

Thanks!
Questions?