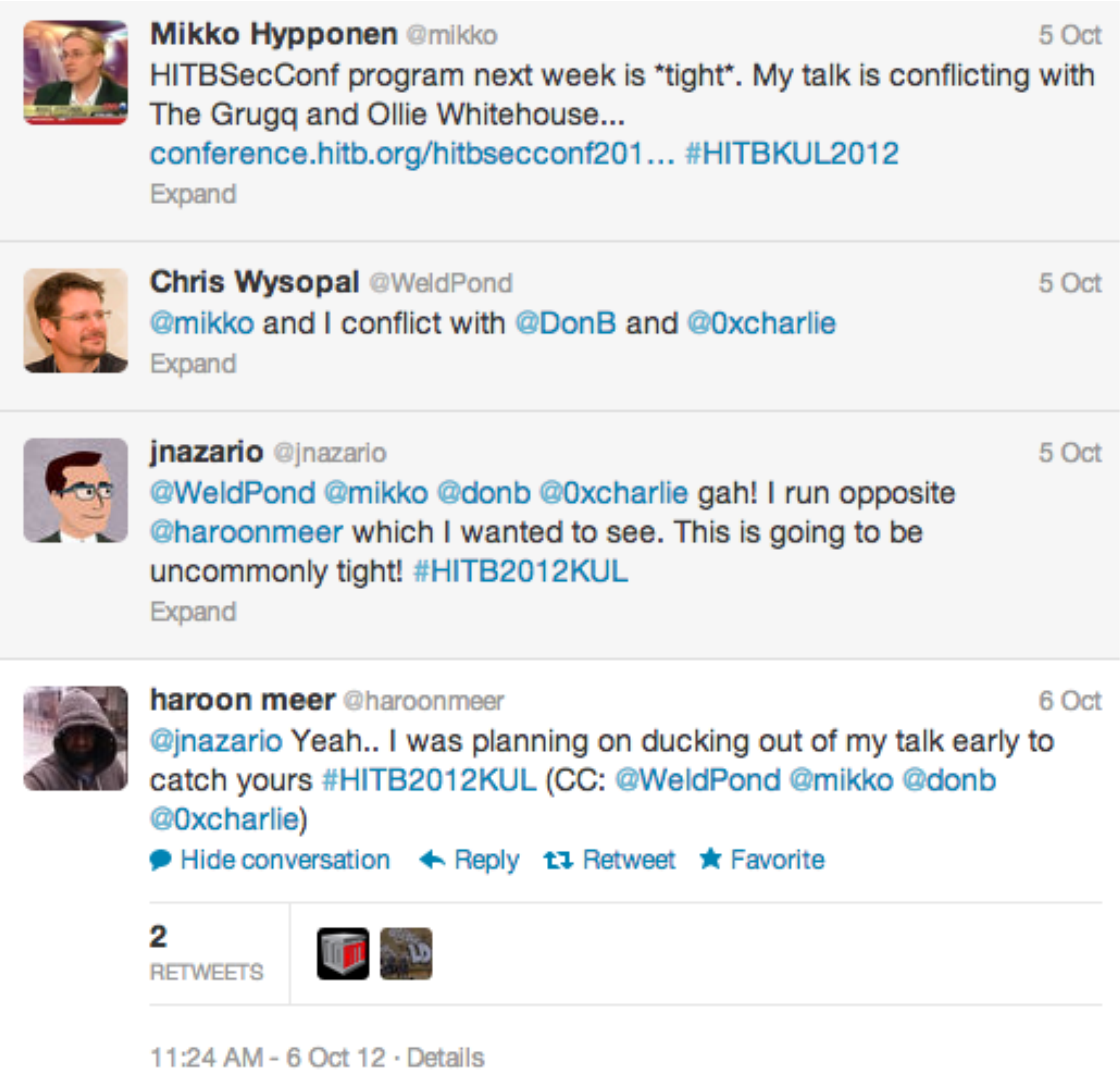
Measuring Botnet Populations

Jose Nazario, Ph.D.
jose@monkey.org
October 2012



Mikko Hypponen @mikko 5 Oct

HITBSecConf program next week is *tight*. My talk is conflicting with The Grugq and Ollie Whitehouse... conference.hitb.org/hitbsecconf201… #HITBKUL2012

Chris Wysopal @WeldPond 5 Oct

@mikko and I conflict with @DonB and @0xcharlie

jnazario @jnazario 5 Oct

@WeldPond @mikko @donb @0xcharlie gah! I run opposite @haroonmeer which I wanted to see. This is going to be uncommonly tight! #HITB2012KUL

haroon meer @haroonmeer 6 Oct

@jnazario Yeah.. I was planning on ducking out of my talk early to catch yours #HITB2012KUL (CC: @WeldPond @mikko @donb @0xcharlie)

2 RETWEETS

11:24 AM - 6 Oct 12 · Details

Overview

- Background
- Implications – Why count?
- Measurement Methodologies
- Limitations and Complications
- Recommendations

Jose Nazario, Ph.D.

- Invincea Labs, 2012-present
  - Prior: Arbor Networks, 2002-2012
- My fourth HITB
  - 2004, 2007, 2010, 2012
- Interests
  - Botnets/malware, large scale trends and data, cyber warfare, etc
- Active with ENISA, FIRST, Honeynet Project

- Ph.D. in Biochemistry

What we Measure

- Trying to measure number of infected devices
  - Affected people, accounts, etc

- What we can measure is number of infected PCs or IPs

- We must estimate to infected population size

Why Count?

- Prevalence measurements
  - By geographic region
  - Prioritize efforts
  - Scale of resources needed to gather
- Know when to call it a victory (counts = 0)
- Size of possible impact
  - Financial, attack, etc

Counting Methodologies

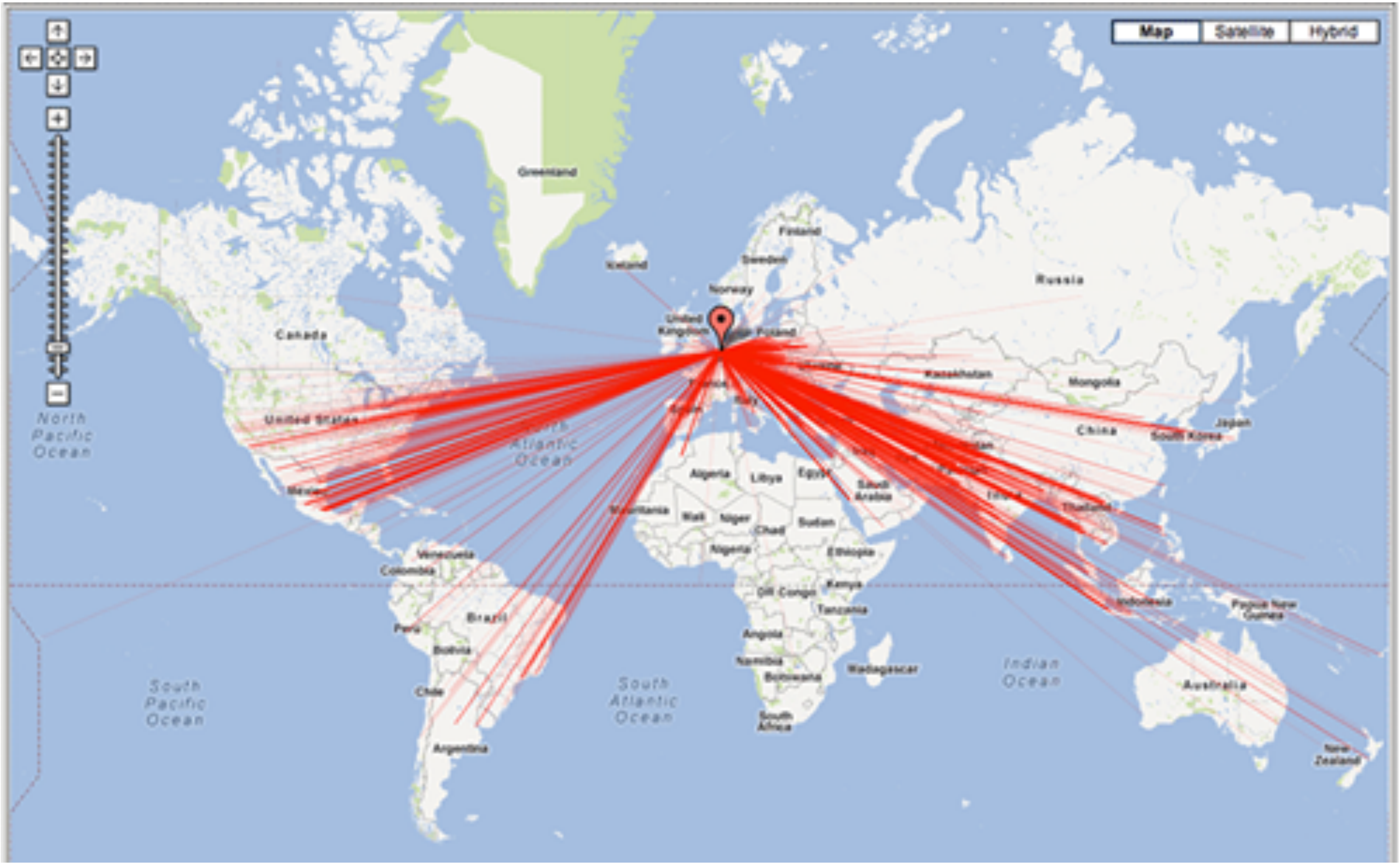
- Sinkholes
- Traffic logs and telltale signs
- Botnet panels
- Darknet monitors
- Direct observation
    - Network
    - Host
    - P2P enumeration

Sinkholes

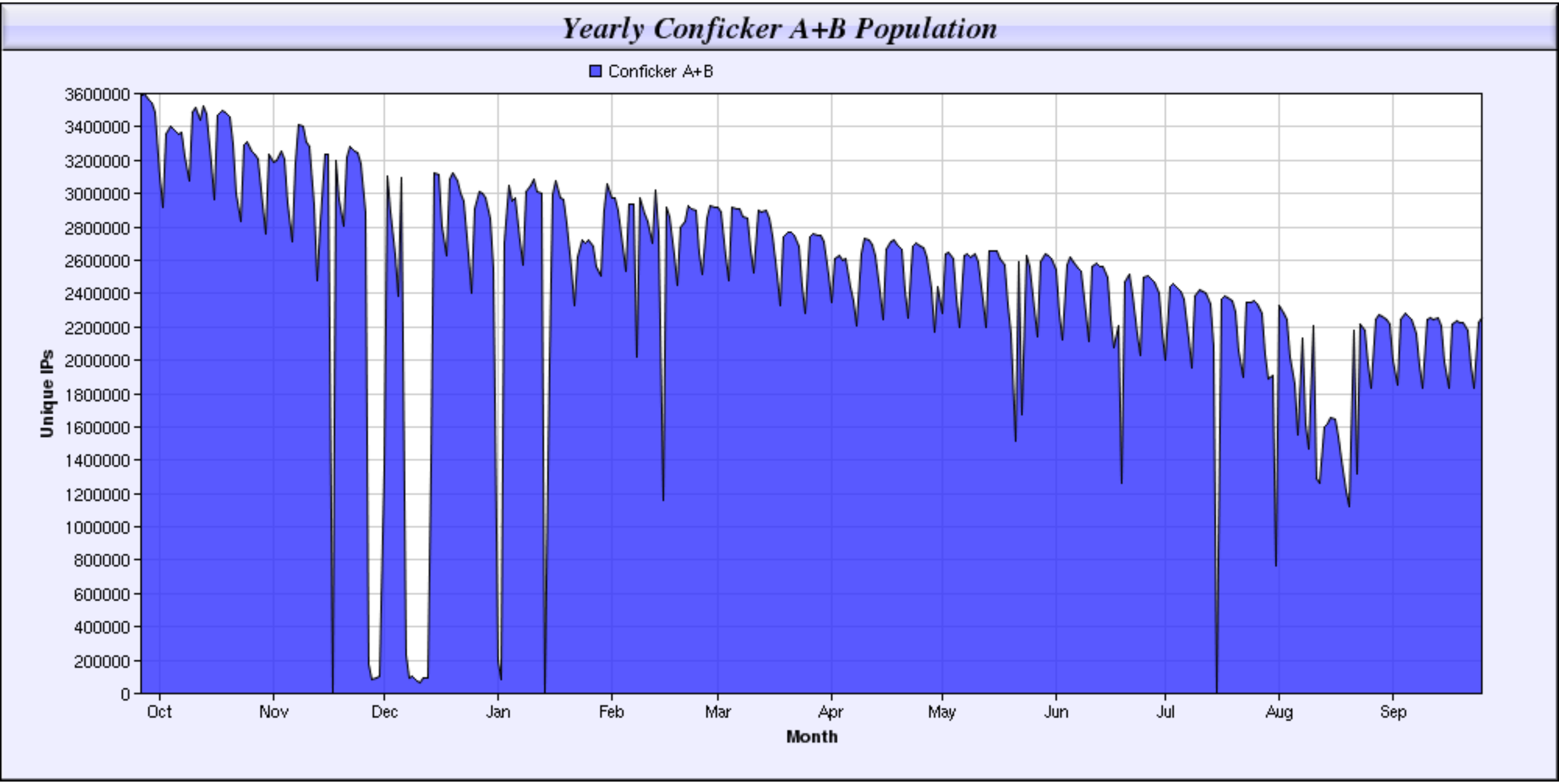
- Redirect botnet command and control (CnC) server to your own host
  - DNS injection
  - P2P injection
  - Route redirection
- Often called "hijacking"
- Count unique IPs per day connecting

- Very common

Khelios Sinkhole from Kaspersky



1 Year of Conficker Sinkhole Data

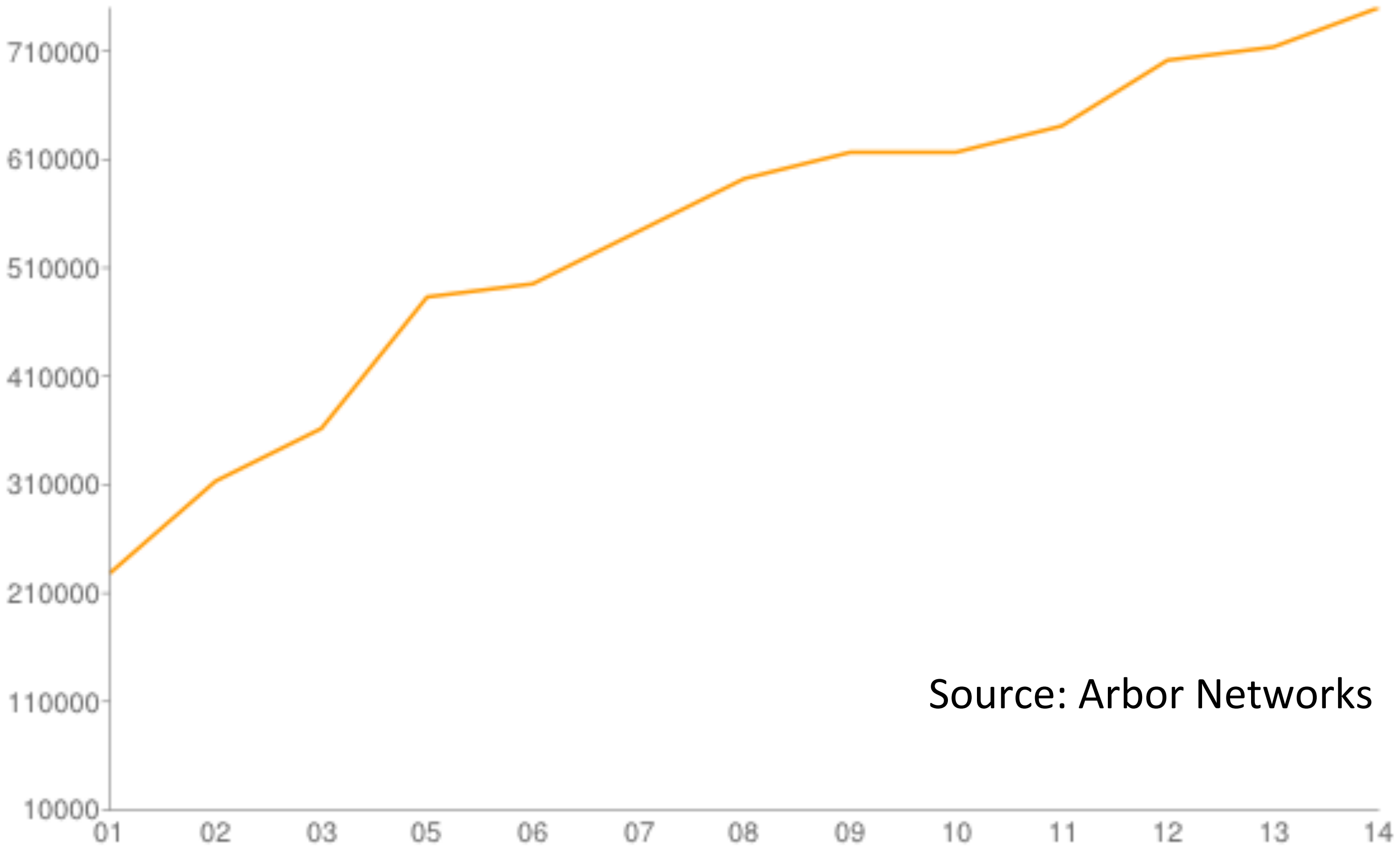


Traffic Logs

- Assume some feature to count on
  - Unique identifier per client IP
  - Hostname, MAC address
- Infection count (e.g. q=N in Conficker)

- Can help give some better numbers

Conficker Counts



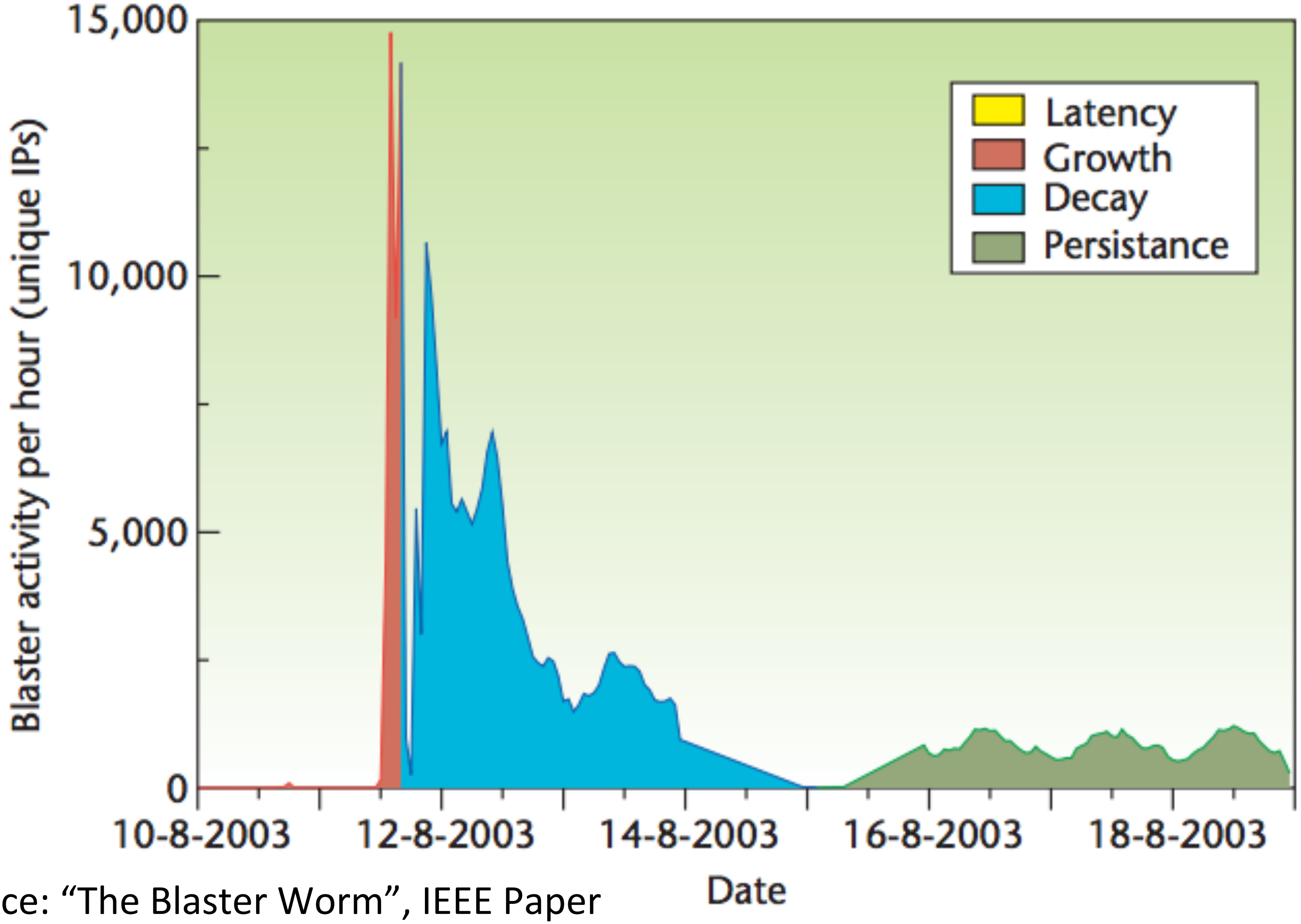
Source: Arbor Networks

- Used “q” value per client IP
  - “q” was used to report victim counts
- Summed values per day

Darknet Monitoring

- Monitor large, unused IPv4 address space blocks
  - Contiguous or disparate
- Fingerprint bot specific signs
  - TCP/IP service
  - Exploit attempts

Blaster Worm Example (2003)



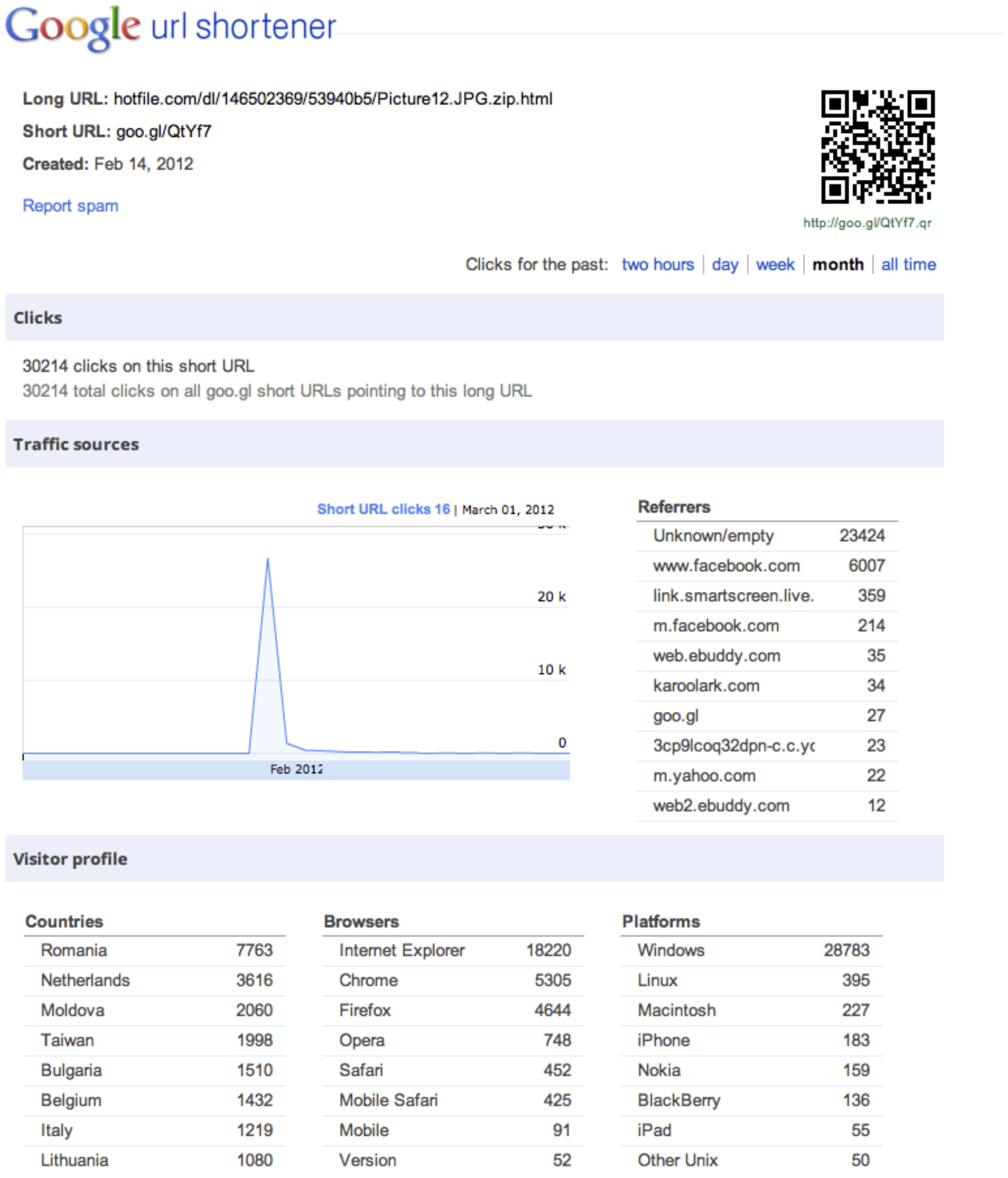
Source: “The Blaster Worm”, IEEE Paper

URL Shorteners

- Use case: malicious links spammed using a link shortener

- Services used to map long URLs to shorter one
  - Great for space-limited uses
  - Great for obfuscating malware/intent
- Several provide statistics we can openly view

- Limitations
  - Some click out of research but not to get infected
  - Unknown infection/block rates



Google url shortener

Long URL: hotfile.com/dl/146502369/53940b5/Picture12.JPG.zip.html

Short URL: goo.gl/QtYf7

Created: Feb 14, 2012

Report spam

http://goo.gl/QtYf7.qr

Clicks for the past: two hours | day | week | **month** | all time

Clicks

30214 clicks on this short URL

30214 total clicks on all goo.gl short URLs pointing to this long URL

Traffic sources

Short URL clicks 16 | March 01, 2012

Referrers	
Unknown/empty	23424
www.facebook.com	6007
link.smartscreen.live.	359
m.facebook.com	214
web.ebuddy.com	35
karoolark.com	34
goo.gl	27
3cp9lcoq32dpn-c.c.yc	23
m.yahoo.com	22
web2.ebuddy.com	12

Visitor profile

Countries	
Romania	7763
Netherlands	3616
Moldova	2060
Taiwan	1998
Bulgaria	1510
Belgium	1432
Italy	1219
Lithuania	1080

Browsers	
Internet Explorer	18220
Chrome	5305
Firefox	4644
Opera	748
Safari	452
Mobile Safari	425
Mobile	91
Version	52

Platforms	
Windows	28783
Linux	395
Macintosh	227
iPhone	183
Nokia	159
BlackBerry	136
iPad	55
Other Unix	50

- Example from a drive by download using goo.gl link

- Shows countries, referrers, platforms, etc

Direct: Network Flows

- Count traffic to designated CnCs
  - Upstream
  - Aggregate of multiple views

- Pretty rare, people just take down CnC instead

Direct: Host Views

- Microsoft has the best option here
  - Count reports from Windows Defender/etc uses
  - Distribute tool globally, get unique identifier for host

- Pro: Most direct measurement
- Con: Not accessible to very many people

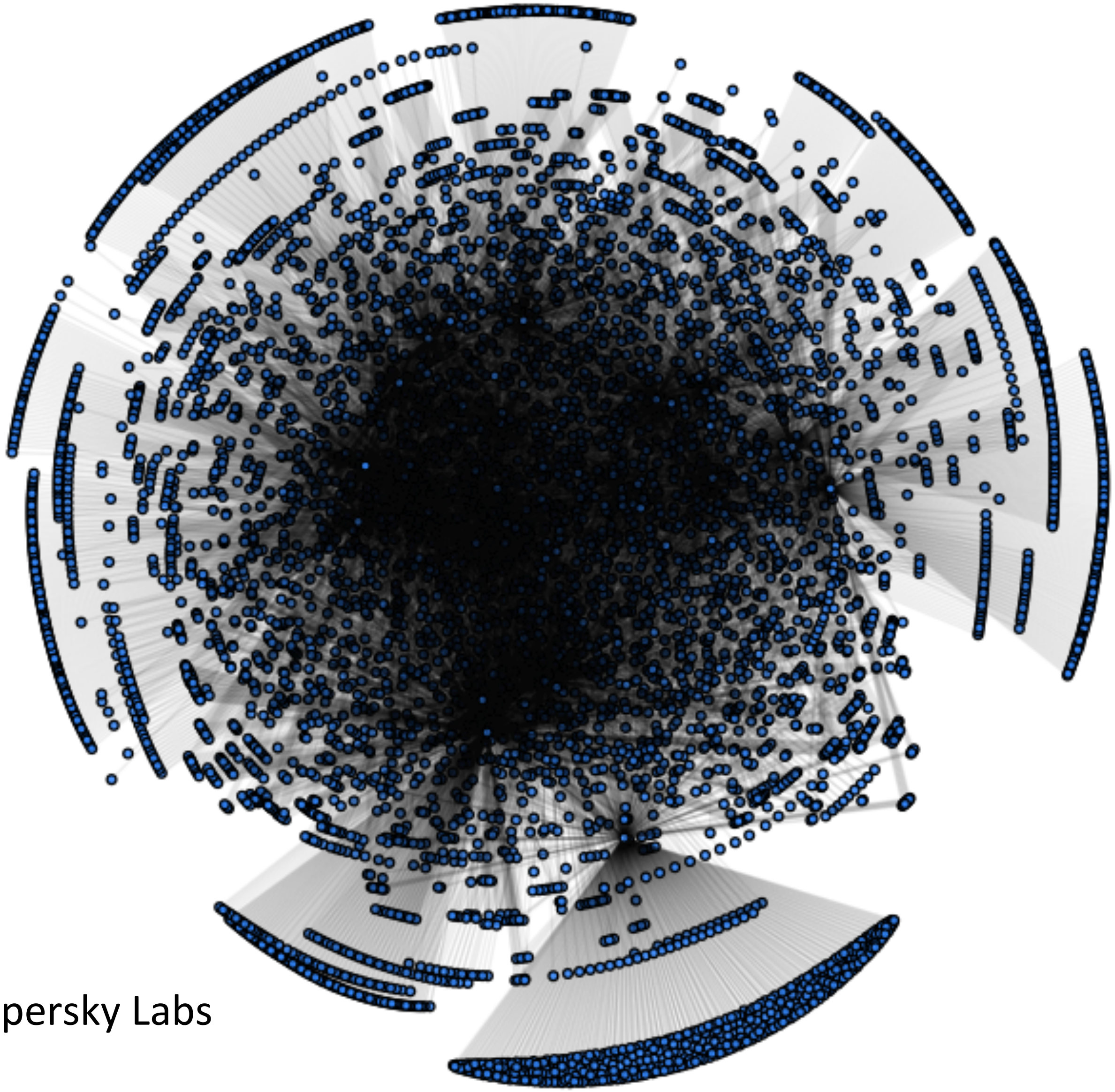
Direct: P2P Enumeration

- “Crawl” the P2P network (for P2P bots)
  - Record list of IPs seen over time
  - Receive updated peer lists

- Requires that you know the protocol
- Easily thwarted with strong crypto

- Storm worm, Miner botnet, etc

Miner P2P Botnet



Source: Kaspersky Labs

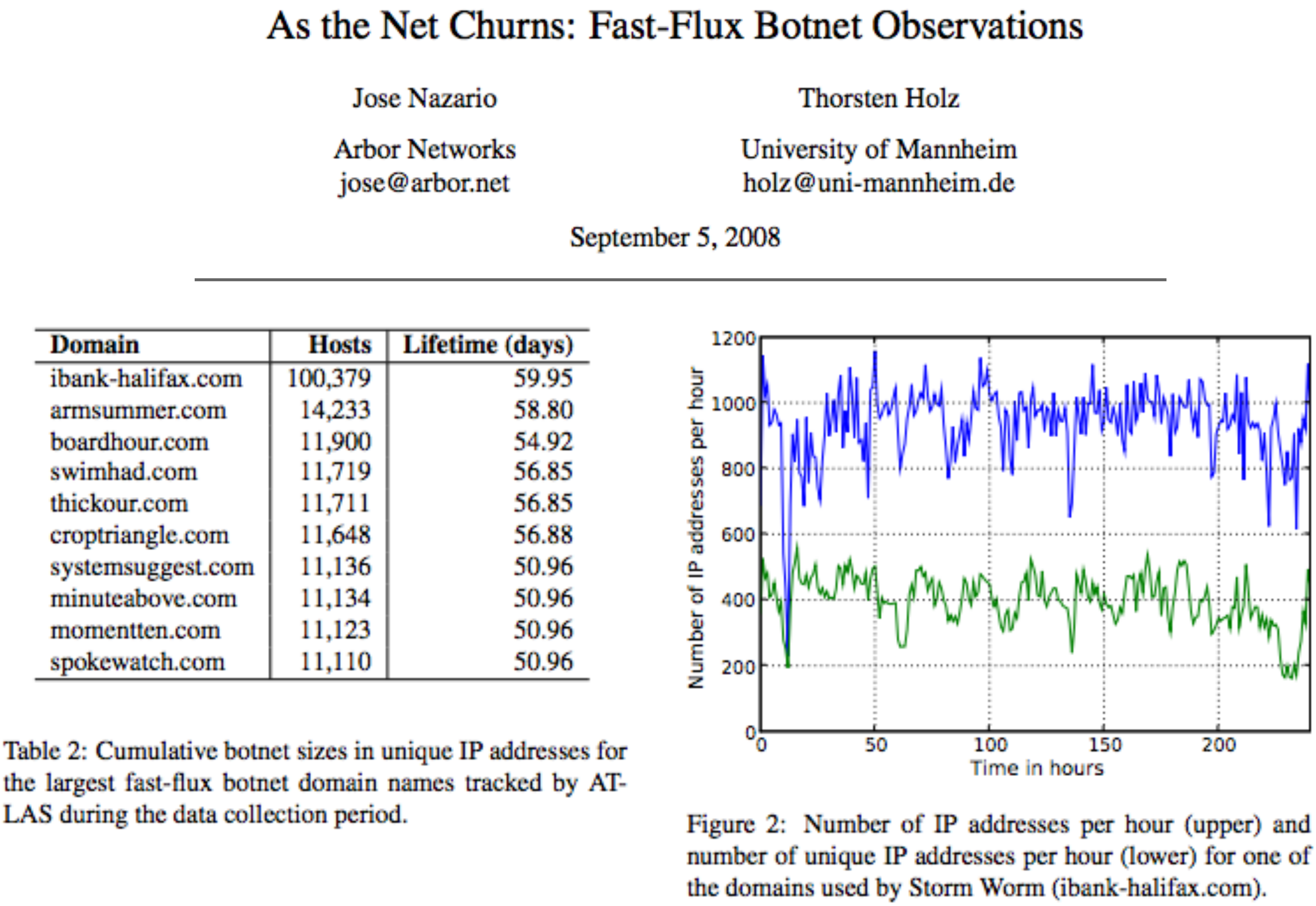
Limitations

- Network visibility
- Redirection by ISP
- DNS blacklists
- Offline hosts
- Inaccurate reporting by the bot

Complications

- DHCP
  - Overcounting: 1 IP does not equal 1 host
  - We estimated 10% volatile DHCP (<24h lifetime)
- NAT
  - 10-100:1 ratio seen in the wild
  - Blaster example (2003)
    - Arbor estimate (IEEE paper): 800,000 hosts
    - Microsoft measurements: 8 million hosts
- Opt-out
  - Many people actively disable updates or reporting
    - Privacy concerns, piracy, etc

2008 Fast Flux Study: 1% visibility via DNS



As the Net Churns: Fast-Flux Botnet Observations

Jose Nazario

Arbor Networks
jose@arbor.net

Thorsten Holz

University of Mannheim
holz@uni-mannheim.de

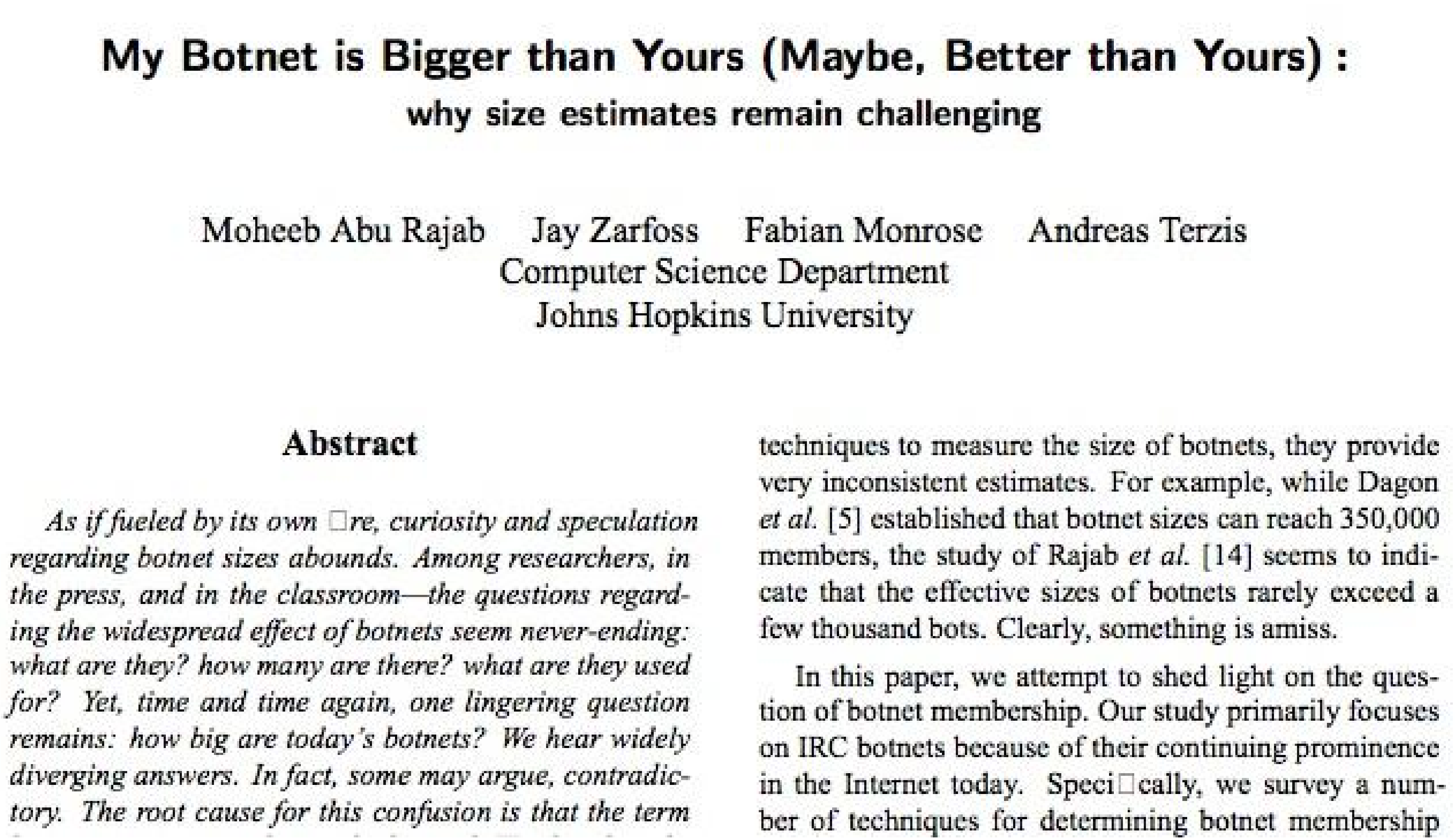
September 5, 2008

Domain	Hosts	Lifetime (days)
ibank-halifax.com	100,379	59.95
armsummer.com	14,233	58.80
boardhour.com	11,900	54.92
swimhad.com	11,719	56.85
thickour.com	11,711	56.85
croptriangle.com	11,648	56.88
systemsuggest.com	11,136	50.96
minuteabove.com	11,134	50.96
momentten.com	11,123	50.96
spokewatch.com	11,110	50.96

Table 2: Cumulative botnet sizes in unique IP addresses for the largest fast-flux botnet domain names tracked by ATLAS during the data collection period.

Figure 2: Number of IP addresses per hour (upper) and number of unique IP addresses per hour (lower) for one of the domains used by Storm Worm (ibank-halifax.com).

Paper from HotBots 2007



My Botnet is Bigger than Yours (Maybe, Better than Yours) : why size estimates remain challenging

Moheeb Abu Rajab Jay Zarfoss Fabian Monrose Andreas Terzis
Computer Science Department
Johns Hopkins University

Abstract

As if fueled by its own fire, curiosity and speculation regarding botnet sizes abounds. Among researchers, in the press, and in the classroom—the questions regarding the widespread effect of botnets seem never-ending: what are they? how many are there? what are they used for? Yet, time and time again, one lingering question remains: how big are today's botnets? We hear widely diverging answers. In fact, some may argue, contradictory. The root cause for this confusion is that the term

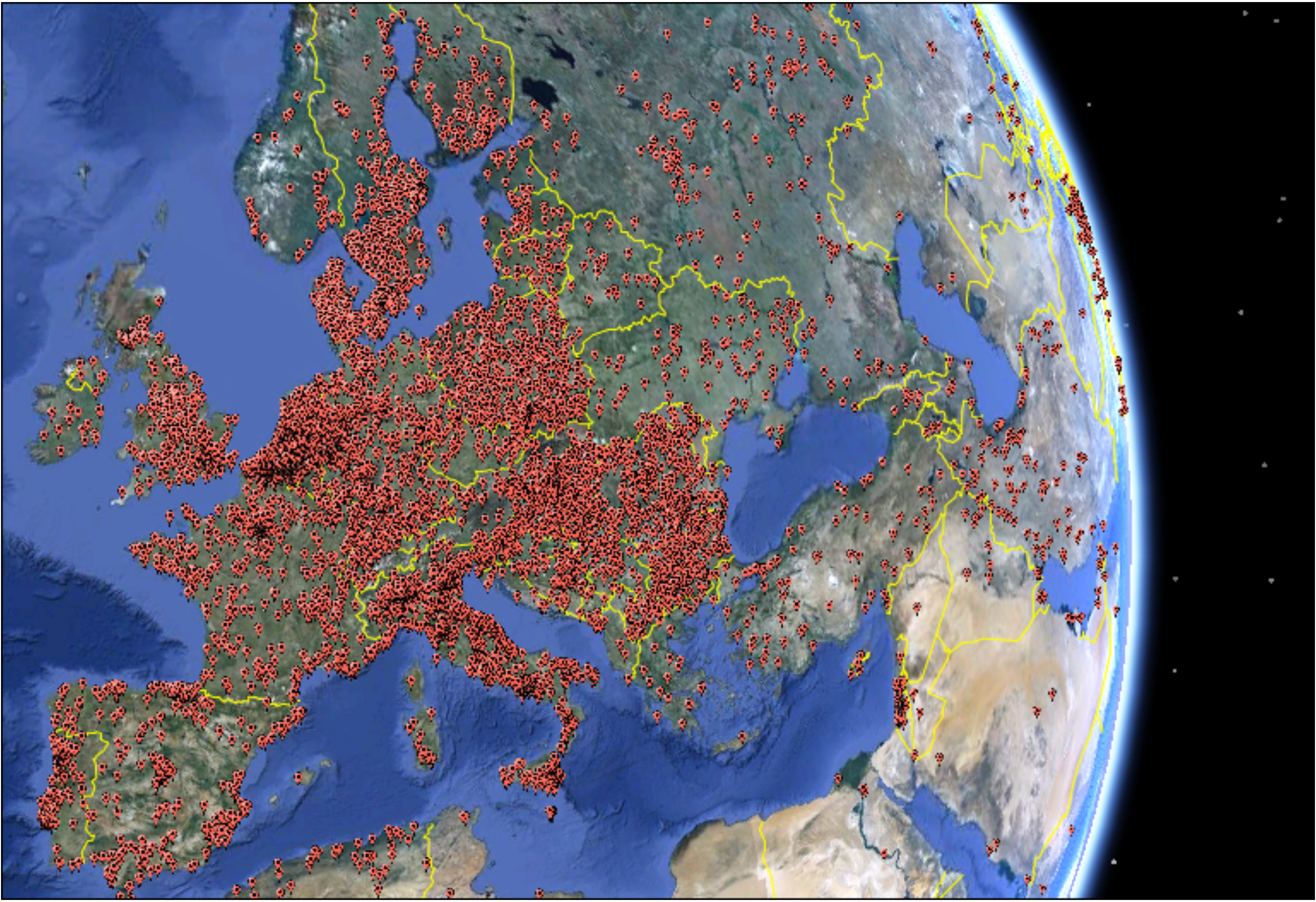
techniques to measure the size of botnets, they provide very inconsistent estimates. For example, while Dagon *et al.* [5] established that botnet sizes can reach 350,000 members, the study of Rajab *et al.* [14] seems to indicate that the effective sizes of botnets rarely exceed a few thousand bots. Clearly, something is amiss.

In this paper, we attempt to shed light on the question of botnet membership. Our study primarily focuses on IRC botnets because of their continuing prominence in the Internet today. Specifically, we survey a number of techniques for determining botnet membership

Other Uses of Botnet Infection Data

- Notifications
  - Very big in the operational security community
    - DCWG, CWG, FBWG, etc
    - Cleanup, etc
  - Global efforts
    - US IBG, AU iCode, NL, DE, JP, etc

- Visualizations - pretty art
  - Great for demos, education
  - Also see http://www.vizsec.org/#program



Source: http://www.f-secure.com/weblog/archives/00002430.html

Thank You

jose@monkey.org
@jnazario