Clean Netherlands (Nederland Schoon)

Tackling internet pollution using science and law enforcement

Maciej Korczyński, Arman Noroozian, Michel van Eeten

Faculty of Technology, Policy and Management

Delft University of Technology

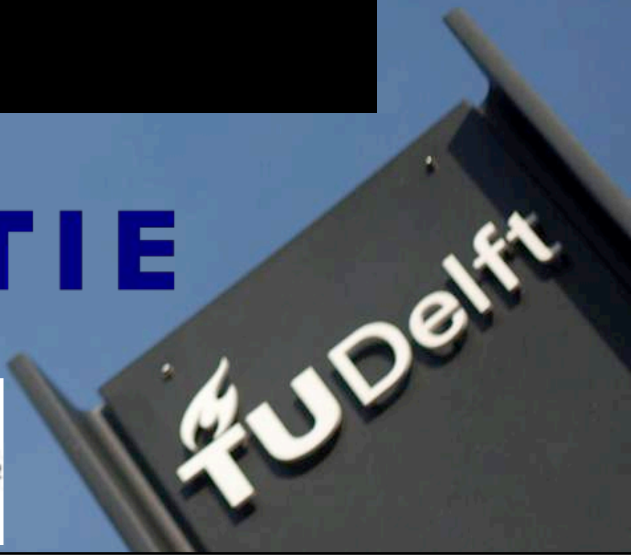
Contact: maciej.korczynski@tudelft.nl

13-14 April 2015

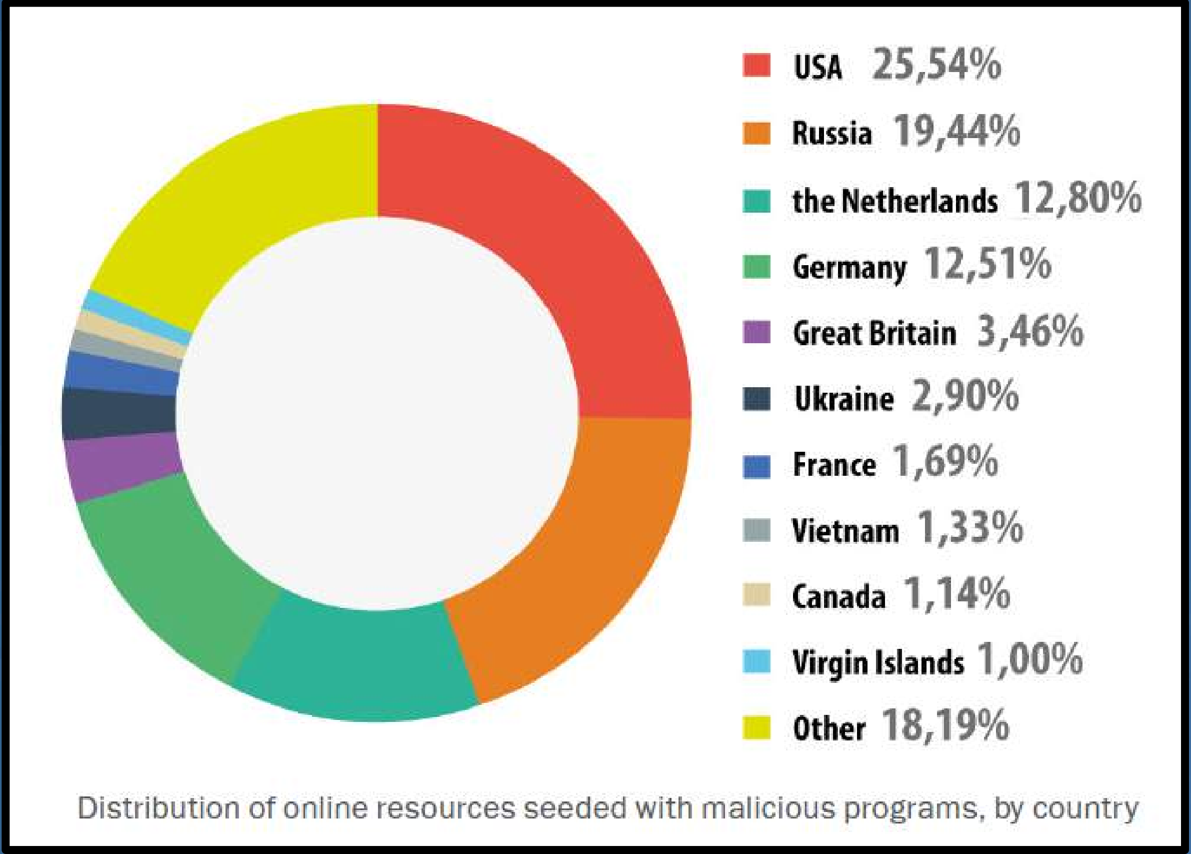
POLITIE

Authority for Consumers & Markets

Openbaar Ministerie



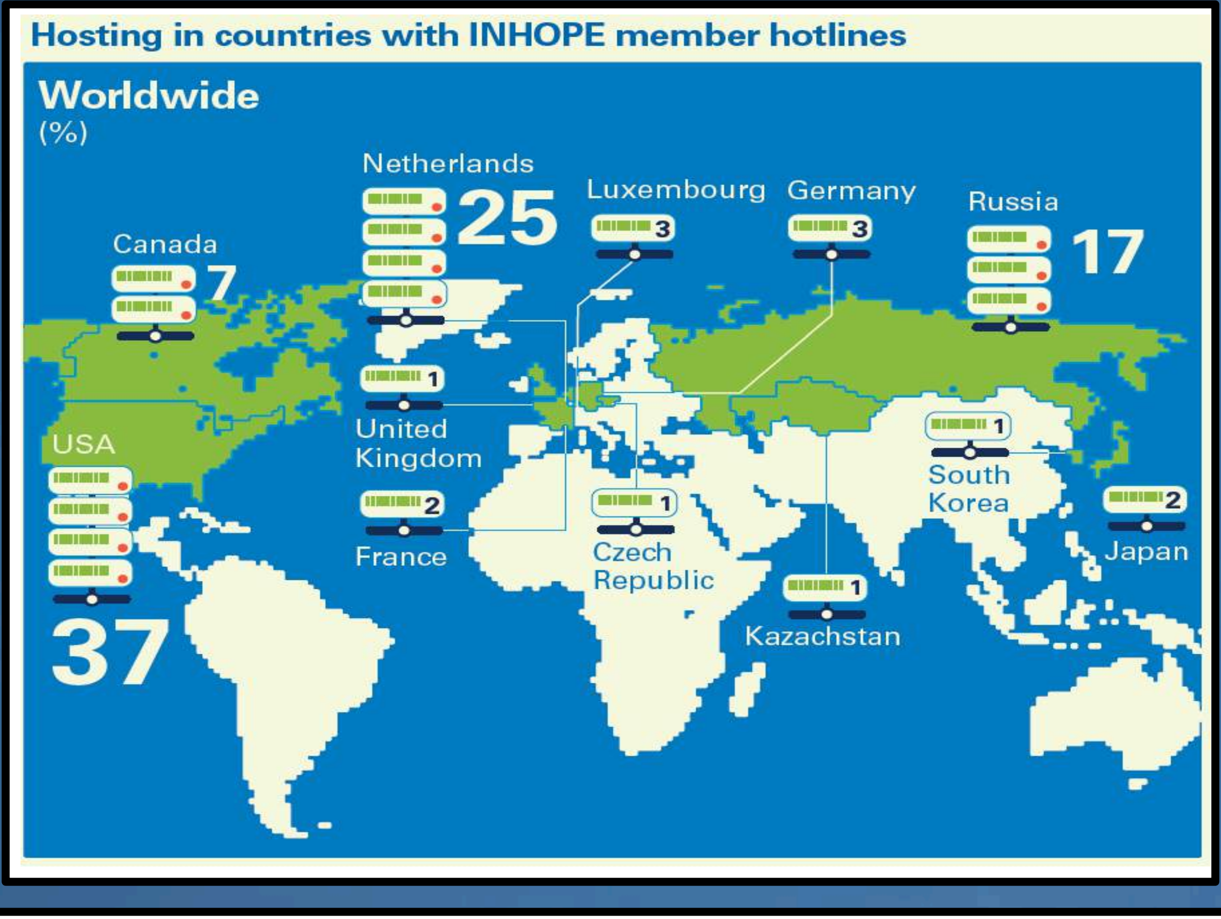
Problem: Malware



Distribution of online resources seeded with malicious programs, by country

Source: http://media.kaspersky.com/pdf/KSB_2013_EN.pdf

Problem: Child pornography



Source: http://inhope.org/Libraries/Infographics/INHOPE-2013-Inforgraphic.sflb.ashx

Coalition

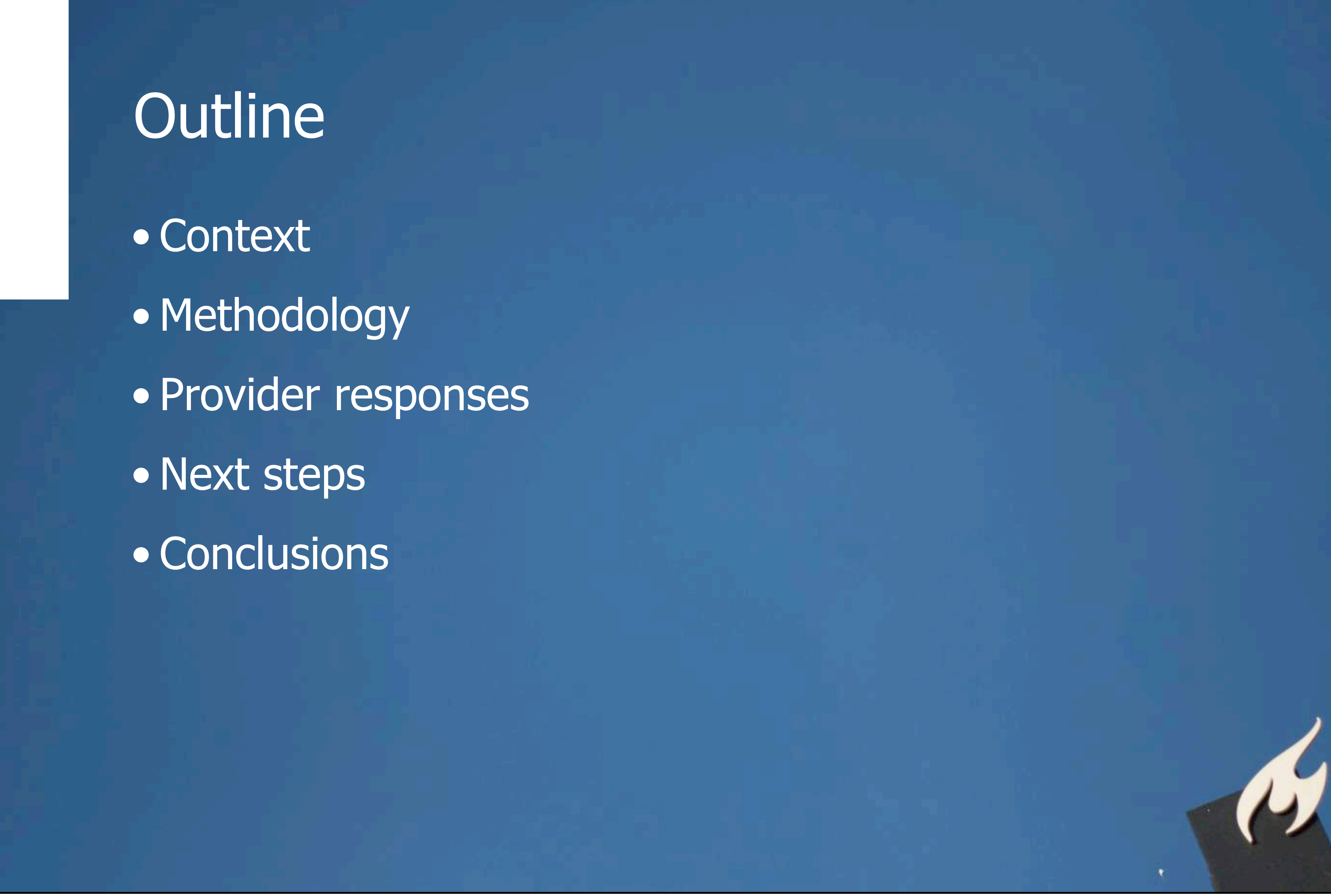
- “Nederland Schoon” is aimed at cybercrime facilitators

- Project goals:
  - build empirically sound ‘pollution map’ on the ASN level
  - research what separates the good from the bad from the mediocre
  - enhance self cleansing ability of NL market by
    - promoting best practices and awareness, and
    - pressuring the rotten apples

- Prosecution is no goal per se, but not excluded either

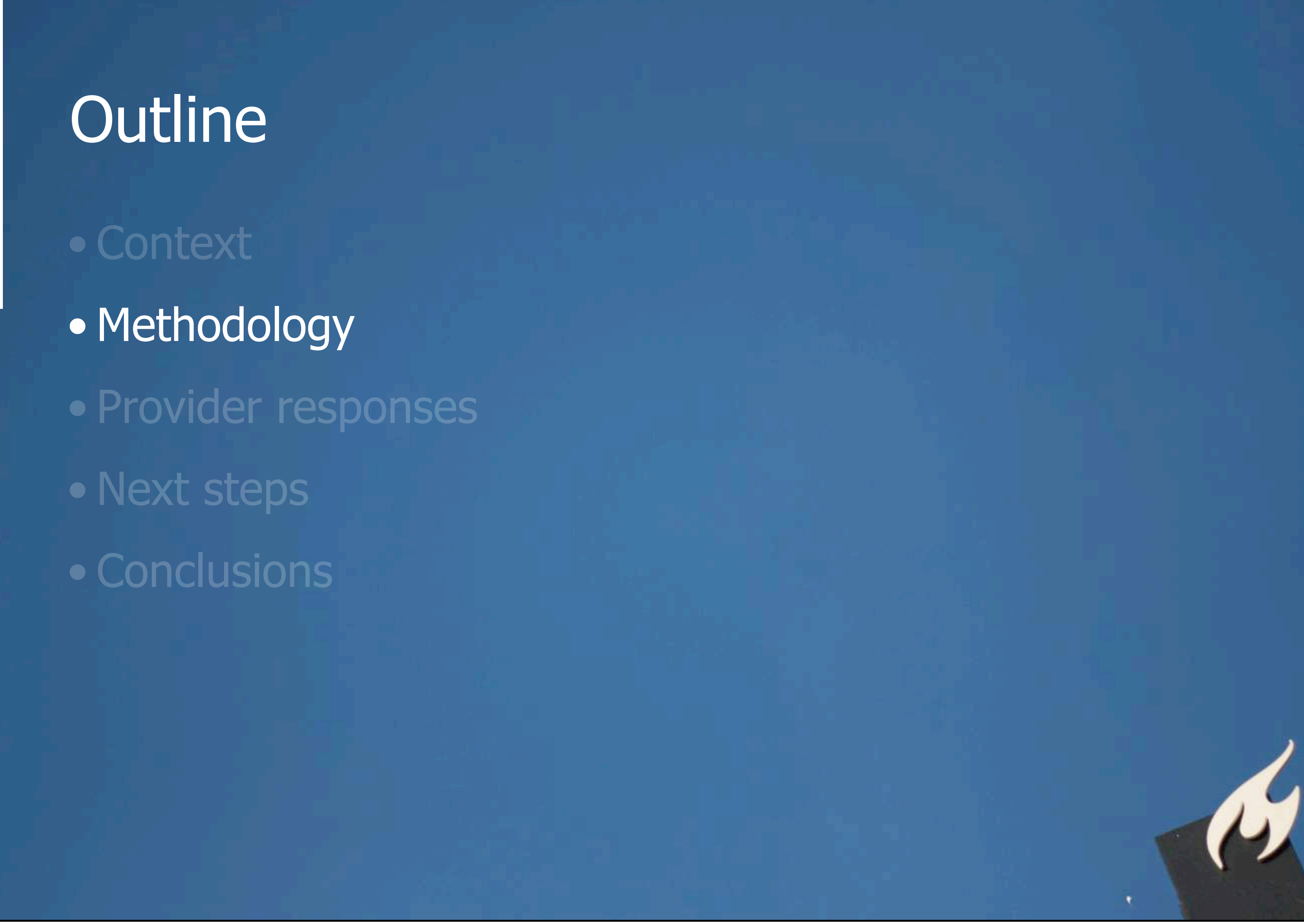
Coalition

- Delft University of Technology
- National Police / High Tech Crime Unit
- ACM (Authority for Consumers and Markets)
- Public Prosecutor



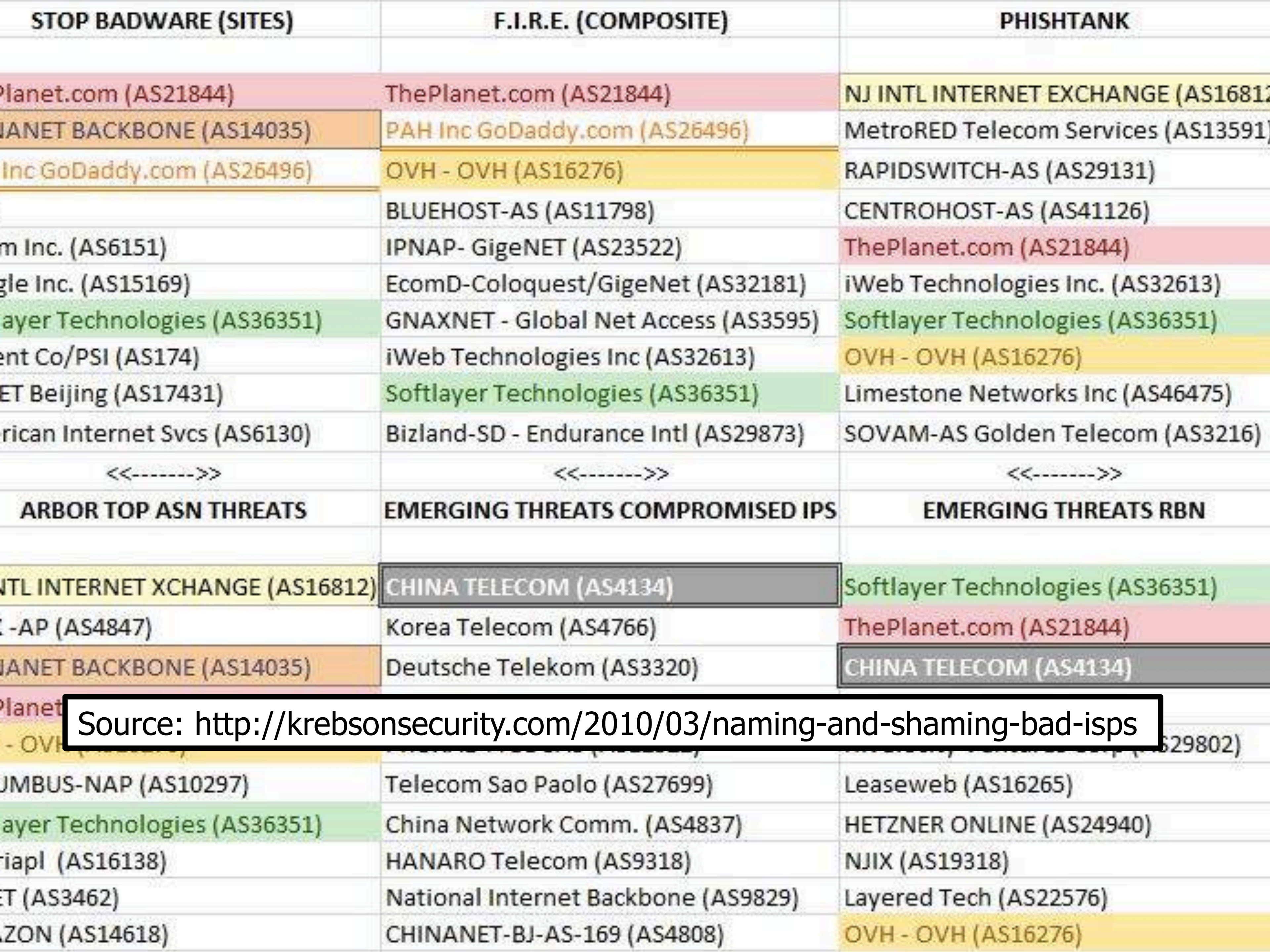
Outline

- Context
- Methodology
- Provider responses
- Next steps
- Conclusions



Outline

- Context
- Methodology
- Provider responses
- Next steps
- Conclusions

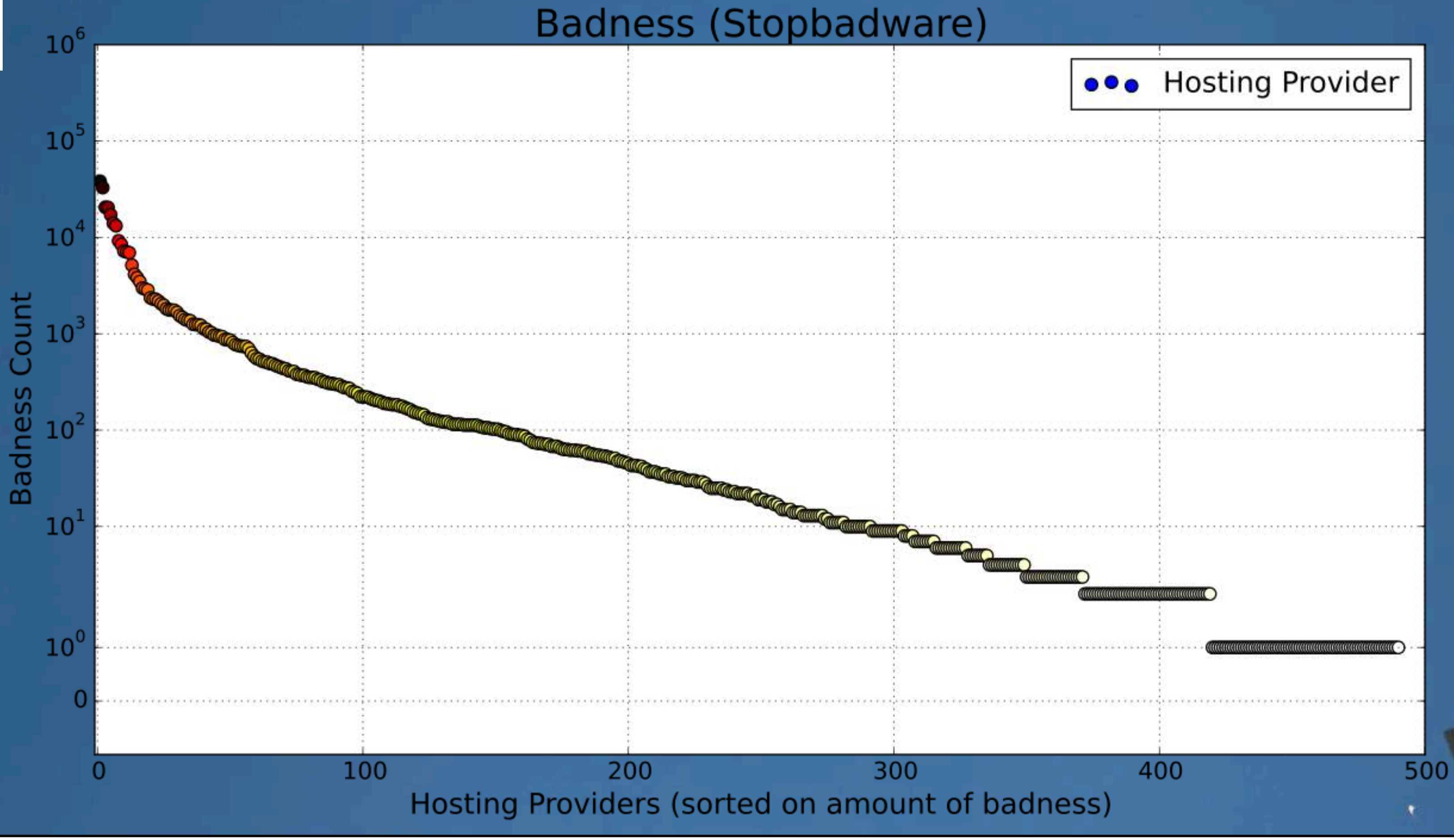


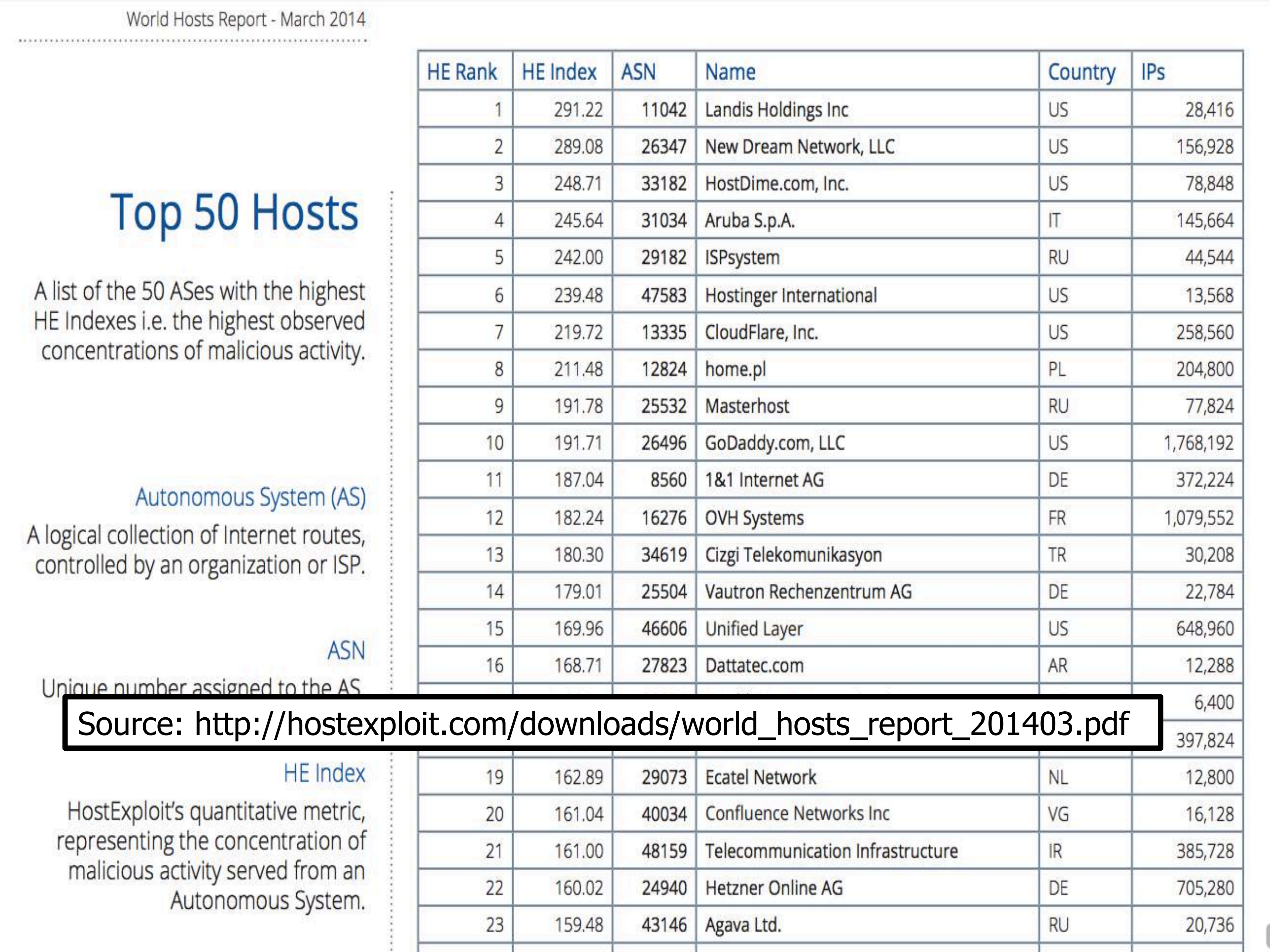
STOP BADWARE (SITES)	F.I.R.E. (COMPOSITE)	PHISHTANK
Planet.com (AS21844)	ThePlanet.com (AS21844)	NJ INTL INTERNET EXCHANGE (AS1681
IANET BACKBONE (AS14035)	PAH Inc GoDaddy.com (AS26496)	MetroRED Telecom Services (AS13591)
Inc GoDaddy.com (AS26496)	OVH - OVH (AS16276)	RAPIDSWITCH-AS (AS29131)
	BLUEHOST-AS (AS11798)	CENTROHOST-AS (AS41126)
m Inc. (AS6151)	IPNAP- GigeNET (AS23522)	ThePlanet.com (AS21844)
gle Inc. (AS15169)	EcomD-Coloquest/GigeNet (AS32181)	iWeb Technologies Inc. (AS32613)
ayer Technologies (AS36351)	GNAXNET - Global Net Access (AS3595)	Softlayer Technologies (AS36351)
ent Co/PSI (AS174)	iWeb Technologies Inc (AS32613)	OVH - OVH (AS16276)
ET Beijing (AS17431)	Softlayer Technologies (AS36351)	Limestone Networks Inc (AS46475)
rican Internet Svcs (AS6130)	Bizland-SD - Endurance Intl (AS29873)	SOVAM-AS Golden Telecom (AS3216)
<<------->>	<<------->>	<<------->>
ARBOR TOP ASN THREATS	EMERGING THREATS COMPROMISED IPS	EMERGING THREATS RBN
NTL INTERNET XCHANGE (AS16812)	CHINA TELECOM (AS4134)	Softlayer Technologies (AS36351)
-AP (AS4847)	Korea Telecom (AS4766)	ThePlanet.com (AS21844)
IANET BACKBONE (AS14035)	Deutsche Telekom (AS3320)	CHINA TELECOM (AS4134)
Planet		
- OV	[illegible]	S29802)
UMBUS-NAP (AS10297)	Telecom Sao Paolo (AS27699)	Leaseweb (AS16265)
ayer Technologies (AS36351)	China Network Comm. (AS4837)	HETZNER ONLINE (AS24940)
riapl (AS16138)	HANARO Telecom (AS9318)	NJIX (AS19318)
ET (AS3462)	National Internet Backbone (AS9829)	Layered Tech (AS22576)
ZON (AS14618)	CHINANET-BJ-AS-169 (AS4808)	OVH - OVH (AS16276)

Source: http://krebsonsecurity.com/2010/03/naming-and-shaming-bad-isps

Mapping abuse

- Just count it





World Hosts Report - March 2014

Top 50 Hosts

A list of the 50 ASes with the highest HE Indexes i.e. the highest observed concentrations of malicious activity.

Autonomous System (AS)

A logical collection of Internet routes, controlled by an organization or ISP.

ASN

Unique number assigned to the AS

HE Index

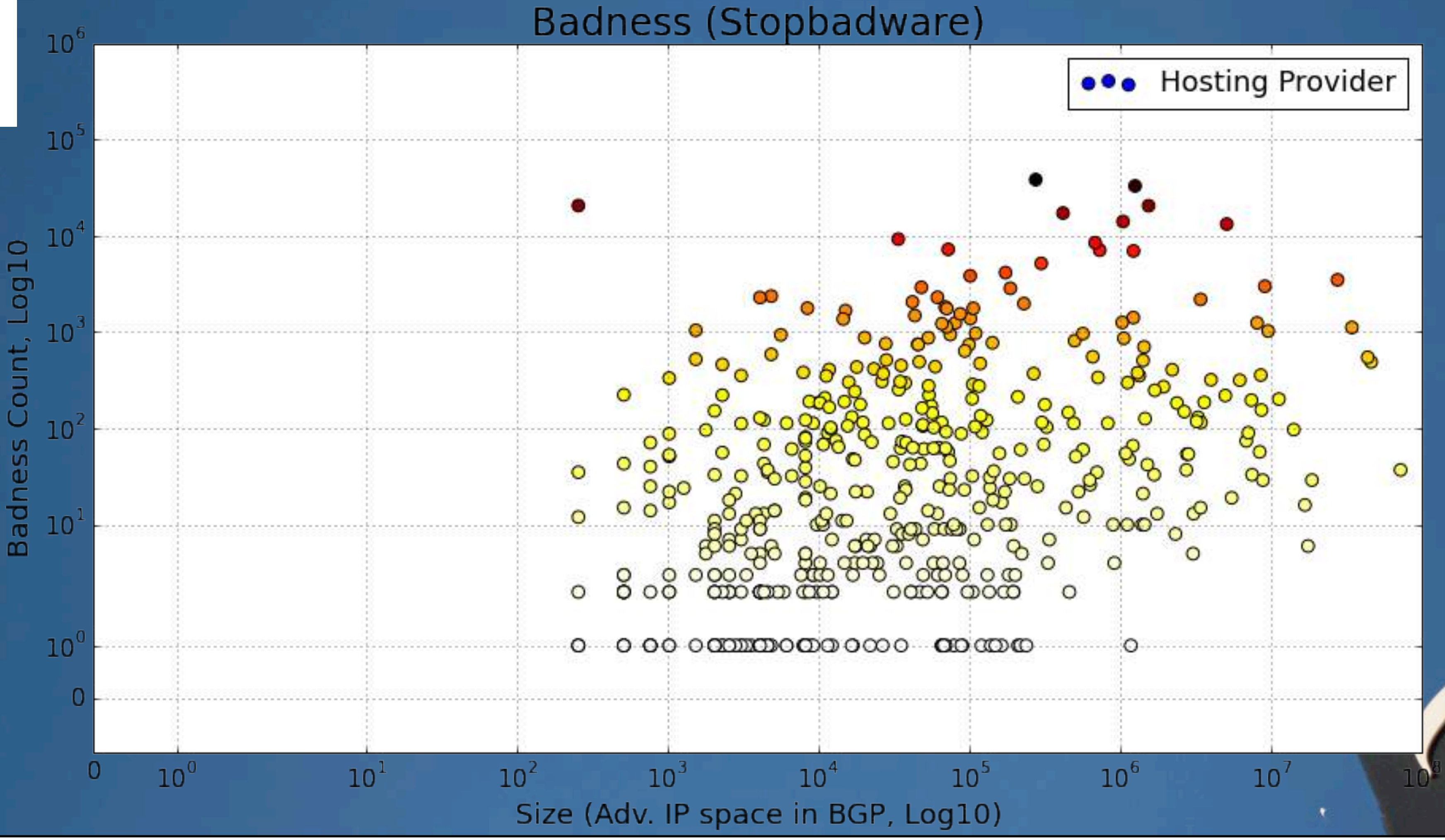
HostExploit's quantitative metric, representing the concentration of malicious activity served from an Autonomous System.

HE Rank	HE Index	ASN	Name	Country	IPs
1	291.22	11042	Landis Holdings Inc	US	28,416
2	289.08	26347	New Dream Network, LLC	US	156,928
3	248.71	33182	HostDime.com, Inc.	US	78,848
4	245.64	31034	Aruba S.p.A.	IT	145,664
5	242.00	29182	ISPsystem	RU	44,544
6	239.48	47583	Hostinger International	US	13,568
7	219.72	13335	CloudFlare, Inc.	US	258,560
8	211.48	12824	home.pl	PL	204,800
9	191.78	25532	Masterhost	RU	77,824
10	191.71	26496	GoDaddy.com, LLC	US	1,768,192
11	187.04	8560	1&1 Internet AG	DE	372,224
12	182.24	16276	OVH Systems	FR	1,079,552
13	180.30	34619	Cizgi Telekomunikasyon	TR	30,208
14	179.01	25504	Vautron Rechenzentrum AG	DE	22,784
15	169.96	46606	Unified Layer	US	648,960
16	168.71	27823	Dattatec.com	AR	12,288
					6,400
					397,824
19	162.89	29073	Ecatel Network	NL	12,800
20	161.04	40034	Confluence Networks Inc	VG	16,128
21	161.00	48159	Telecommunication Infrastructure	IR	385,728
22	160.02	24940	Hetzner Online AG	DE	705,280
23	159.48	43146	Agava Ltd.	RU	20,736

Source: http://hostexploit.com/downloads/world_hosts_report_201403.pdf

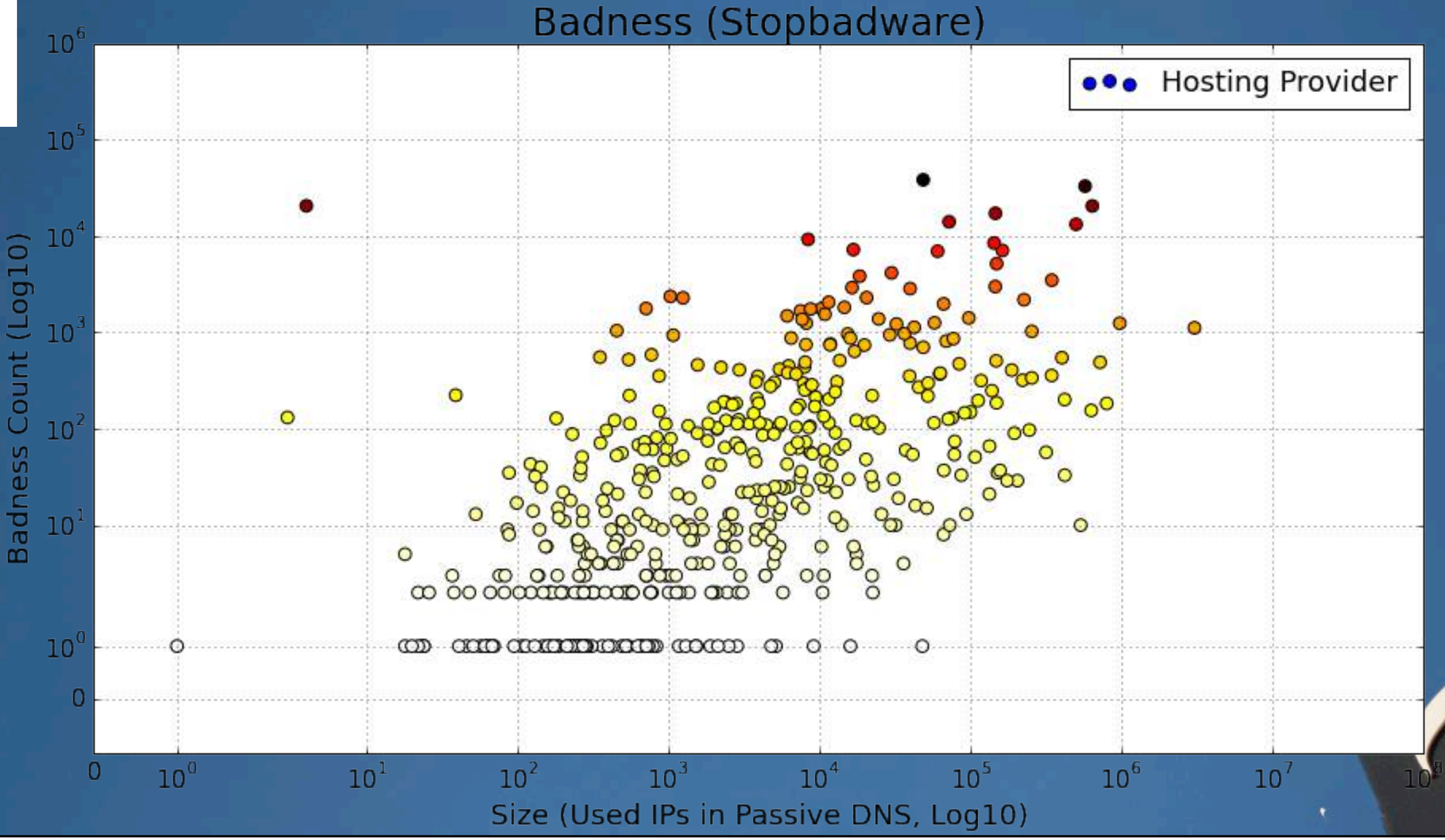
Size matters

- Abuse mapped against # advertised IP space



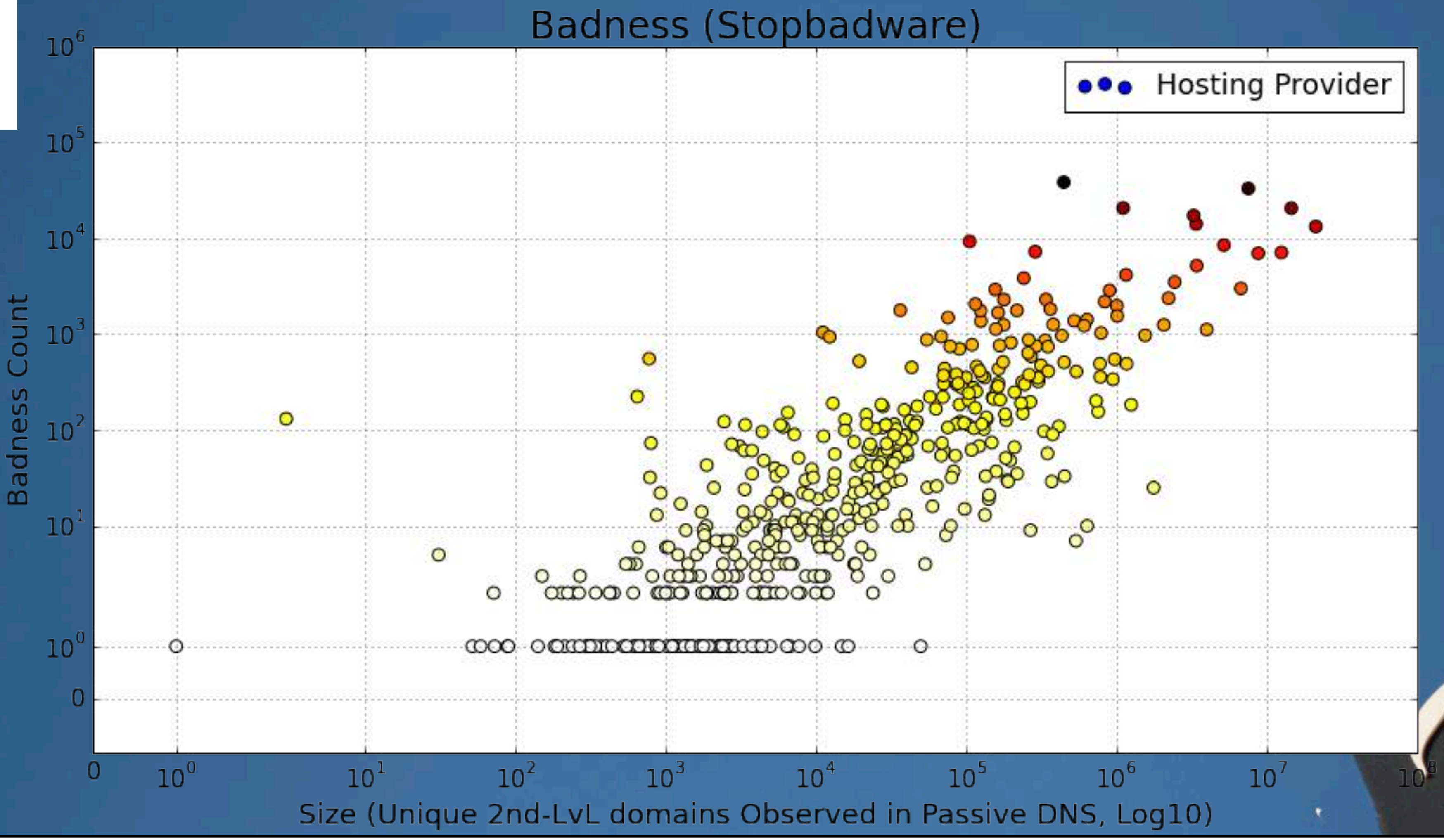
Size matters

- Abuse mapped against # observed IP space (in pDNS)



Size matters

- Abuse mapped against # 2nd level domains



Towards badness metrics

1. Count badness per AS across different data sources
2. Normalize for the size of the AS (in 3 ways)
3. Rank ASes on amount of badness
4. Aggregate rankings (Borda count)
5. Identify ASes with consistently high concentrations of badness

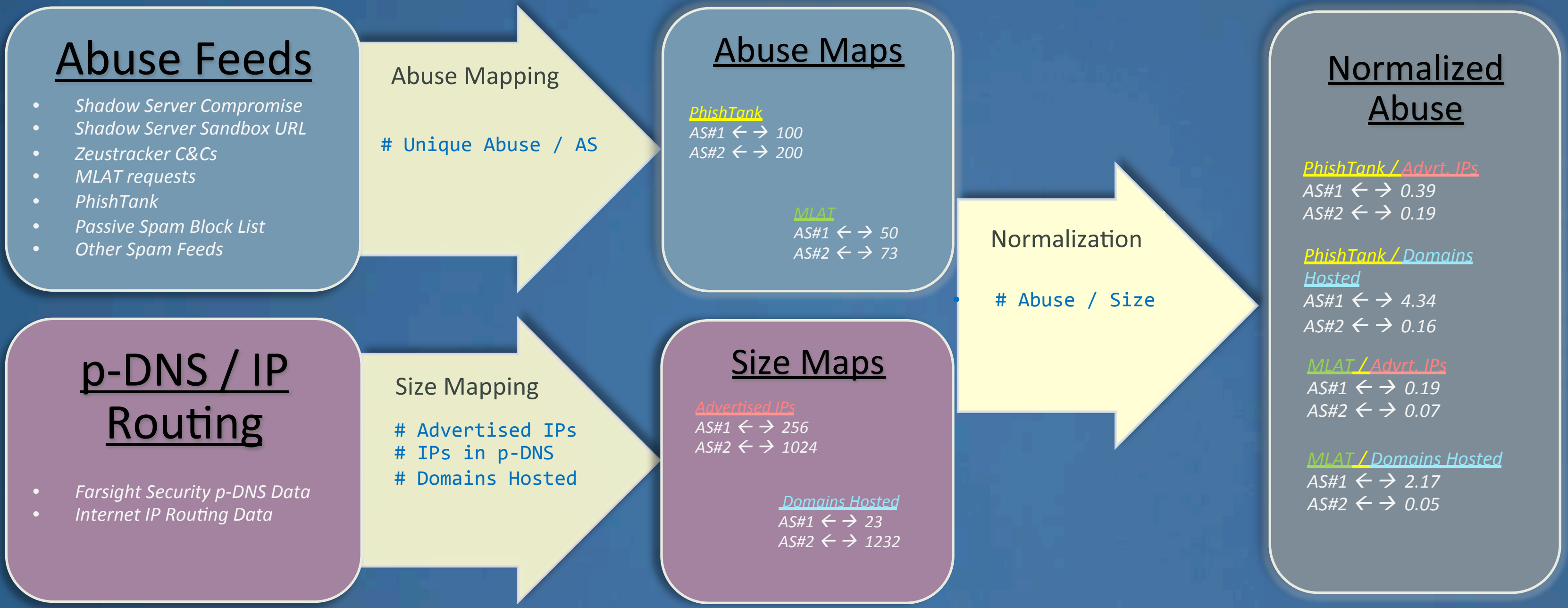
Data sources

- Abuse
  - StopBadware
  - Shadowserver Compromised Website
  - Shadowserver Sandbox URL
  - Zeustracker C&Cs (Abuse.ch)
  - Mutual Legal Assistance Treaty (MLAT) requests
  - Dutch Child Pornography Hotline
  - PhishTank
  - Anti-Phishing Working Group
  - Passive Spam Block List (PSBL)
  - Private Spam trap

- IP Routing Data
  - Python pyasn library

- Passive DNS (pDNS)
  - Farsight Security
  - 750 million unique 2nd Level Domains
  - 93 million unique IPv4 Addresses

Methodology



Abuse Feeds

- *Shadow Server Compromise*
- *Shadow Server Sandbox URL*
- *Zeustracker C&Cs*
- *MLAT requests*
- *PhishTank*
- *Passive Spam Block List*
- *Other Spam Feeds*

p-DNS / IP Routing

- *Farsight Security p-DNS Data*
- *Internet IP Routing Data*

Abuse Mapping

```
# Unique Abuse / AS
```

Size Mapping

```
# Advertised IPs
# IPs in p-DNS
# Domains Hosted
```

Abuse Maps

PhishTank
AS#1 ← → 100
AS#2 ← → 200

MLAT
AS#1 ← → 50
AS#2 ← → 73

Size Maps

Advertised IPs
AS#1 ← → 256
AS#2 ← → 1024

Domains Hosted
AS#1 ← → 23
AS#2 ← → 1232

Normalization

```
# Abuse / Size
```

Normalized Abuse

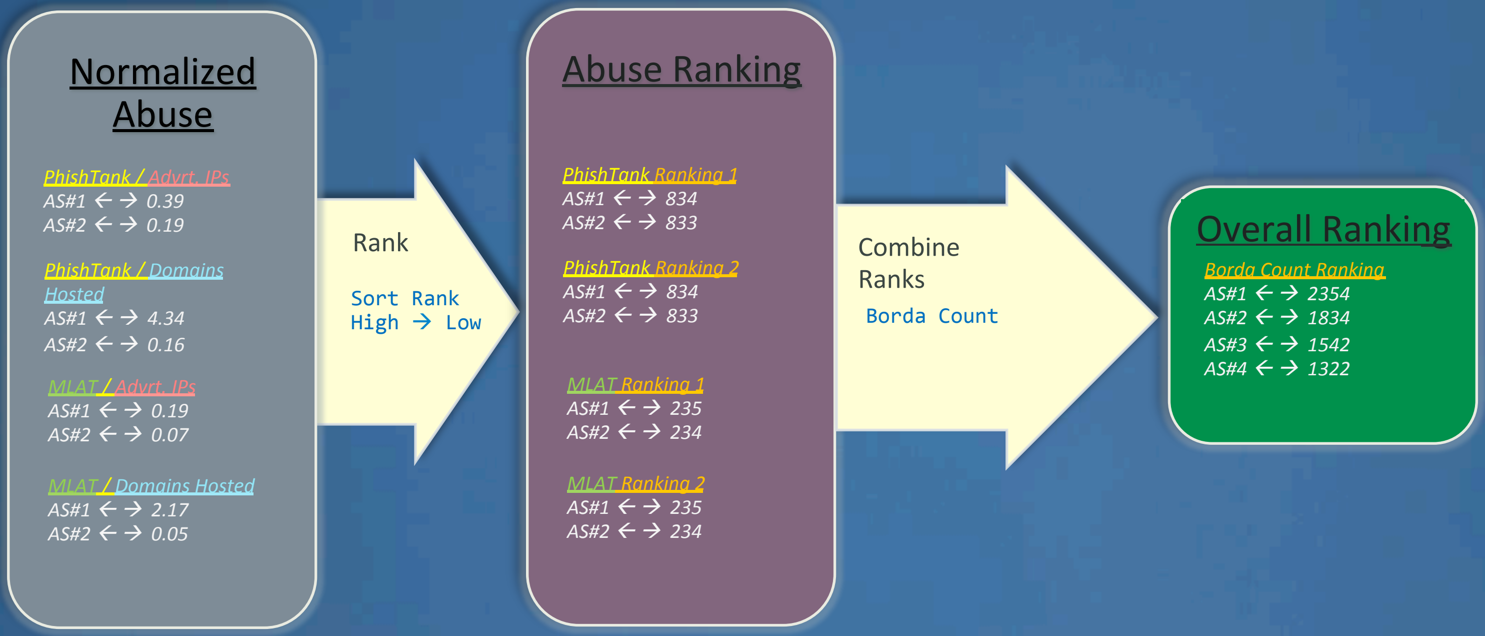
PhishTank / Advrt. IPs
AS#1 ← → 0.39
AS#2 ← → 0.19

PhishTank / Domains Hosted
AS#1 ← → 4.34
AS#2 ← → 0.16

MLAT / Advrt. IPs
AS#1 ← → 0.19
AS#2 ← → 0.07

MLAT / Domains Hosted
AS#1 ← → 2.17
AS#2 ← → 0.05

Methodology (Continued)



Normalized Abuse

PhishTank / Advrt. IPs
AS#1 ← → 0.39
AS#2 ← → 0.19

PhishTank / Domains Hosted
AS#1 ← → 4.34
AS#2 ← → 0.16

MLAT / Advrt. IPs
AS#1 ← → 0.19
AS#2 ← → 0.07

MLAT / Domains Hosted
AS#1 ← → 2.17
AS#2 ← → 0.05

Rank

Sort Rank
High → Low

Abuse Ranking

PhishTank Ranking 1
AS#1 ← → 834
AS#2 ← → 833

PhishTank Ranking 2
AS#1 ← → 834
AS#2 ← → 833

MLAT Ranking 1
AS#1 ← → 235
AS#2 ← → 234

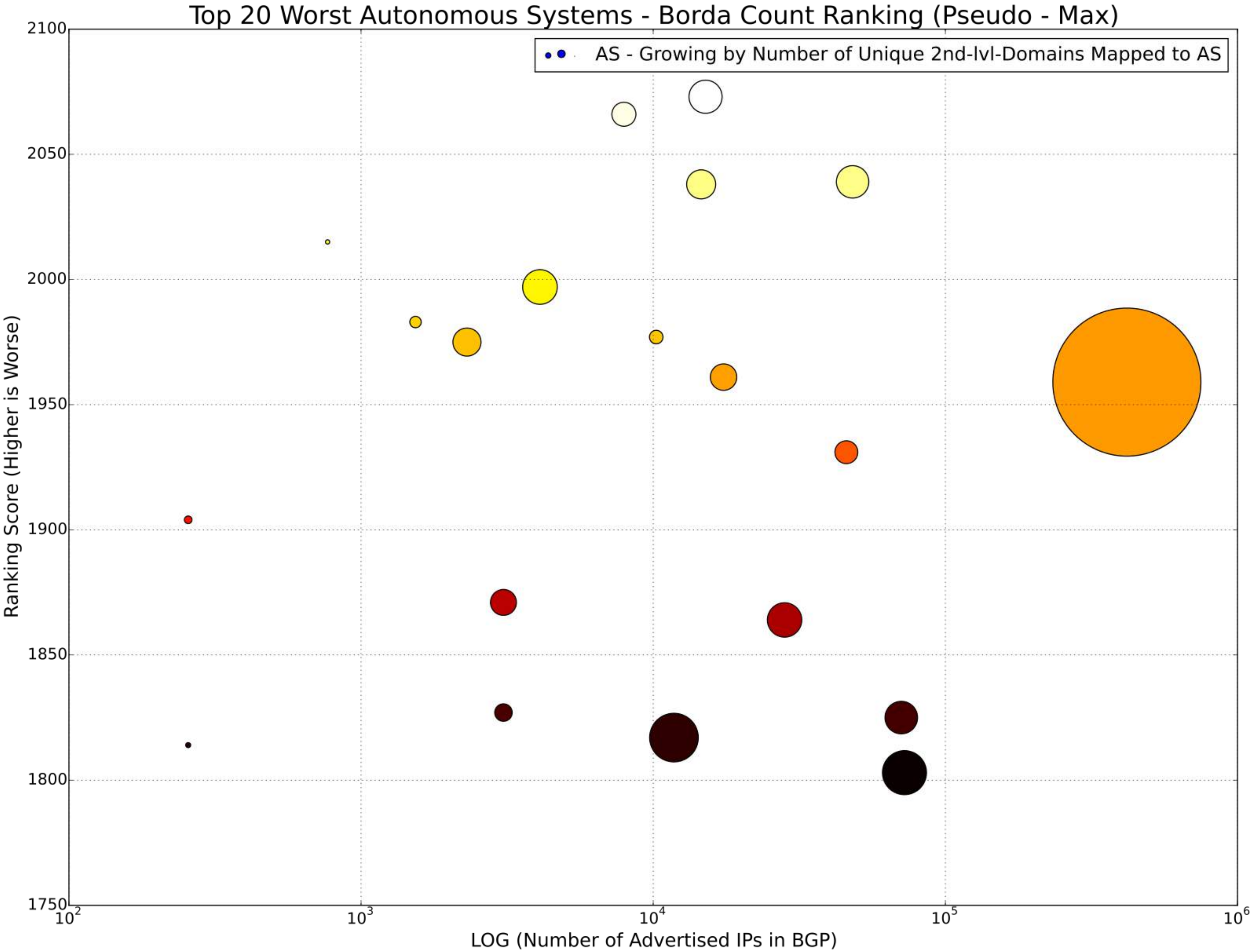
MLAT Ranking 2
AS#1 ← → 235
AS#2 ← → 234

Combine Ranks

Borda Count

Overall Ranking

Borda Count Ranking
AS#1 ← → 2354
AS#2 ← → 1834
AS#3 ← → 1542
AS#4 ← → 1322

Top 20 Worst Autonomous Systems - Borda Count Ranking (Pseudo - Max)
AS - Growing by Number of Unique 2nd-lvl-Domains Mapped to AS
Ranking Score (Higher is Worse)
2100
2050
2000
1950
1900
1850
1800
1750
LOG (Number of Advertised IPs in BGP)
10^2
10^3
10^4
10^5
10^6

Methodology: what is next?

- Measuring uptimes: how quickly does the hosting provider act?
- Get more comprehensive coverage of abuse data
- Separating negligent from criminal
- Developing an approach for identifying criminal hosting, in collaboration with police (“bullet proof hosting providers”)

Outline

- Context
- Methodology
- Provider responses
- Next steps
- Conclusions

Web hosting provider responses

- AS level measurement is adequate
  - indicates feeling of ‘ownership’ of the problems
- Type of service

Outline

- Context
- Methodology
- Provider responses
- Next steps
- Conclusions

Next steps

- Talks with hosting providers with high concentrations of badness
- Infer determining factors (if any)
- Continue measurements
- If necessary; interventions

Outline

- Context
- Methodology
- Provider responses
- Next steps
- Conclusions

Conclusions

- Project aims to
  - measure ‘pollution’
  - get more parties closer towards that mean
  - direct focused pressure to outlying polluters
  - no intention of naming and shaming
- Limitations
  - project does *not* measure intent
  - that is measured by proxying ‘response’ → follow up work
  - some data sources are best effort, e.g. GeoIP

Questions?

Contact information:

Maciej Korczyński

Faculty of Technology, Policy and Management
Delft University of Technology

maciej.korczynski@tudelft.nl
http://mkorczynski.com

Acknowledgements

The research leading to these results was funded by NWO and SIDN

NWO
Netherlands Organisation for Scientific Research

SIDN labs
Internet Research & Innovation

We would like to thank Paul Vixie and Eric Ziegast from Farsight Security for the great support!

FARSIGHT SECURITY