# 大規模IPv6ネットワークスキャンを見つける

福田 健介

kensuke@nii.ac.jp

国立情報学研究所/総合研究大学院大学

National Institute of Informatics

# Kensuke Fukuda (福田健介)

- Affiliation:
  - National Institute of Informatics (国立情報学研究所)
  - Sokendai (総合研究大学院大学)
- Position: Associate professor
- Research interests: Network measurement, Network security

# Kensuke Fukuda (福田健介)

- Affiliation:
  - National Institute of Informatics (国立情報学研究所)
  - Sokendai (総合研究大学院大学)
- Position: Associate professor
- Research interests: Network measurement, Network security

学生募集してます！

# Detecting IPv6 network scanners

Kensuke Fukuda

National Institute of Informatics / Sokendai

Fukuda, et al. ”Detecting Malicious Activities with DNS Backscatter over Time." In IEEE/ACM Transactions on Networking, vol.25, no.5, pp.3203–3218, 2017.

Fukuda, et al. ”Who Knocks at the IPv6 Doors? Detecting IPv6 Scanning" In ACM Internet Measurement Conference 2018, Boston, MA, 2018. (to appear)

National Institute of Informatics

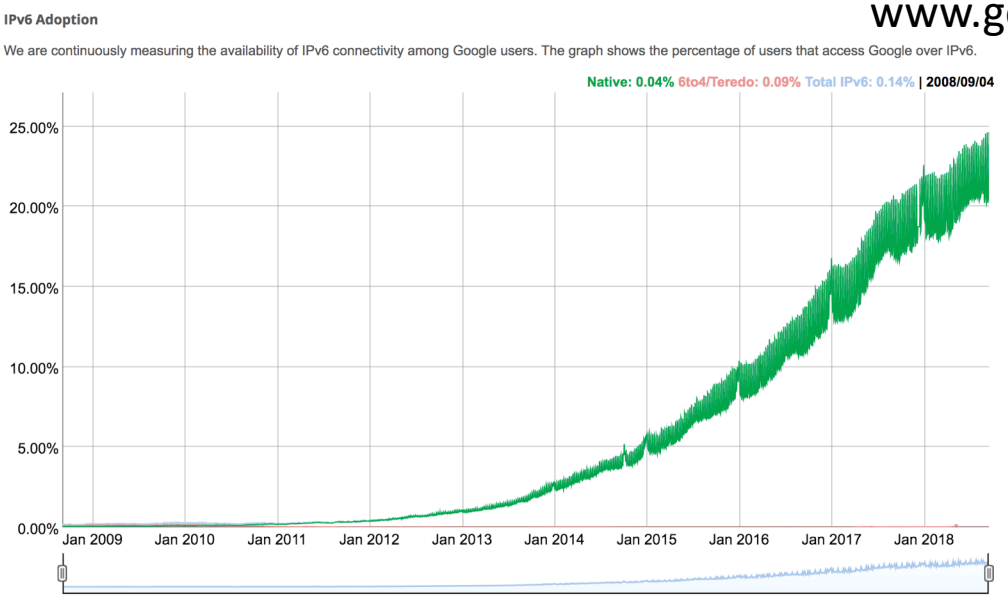
# Today's talk

- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# Deployment of IPv6

- Growth of IPv6 deployment
  - 20% of Google users
  - 25% of ASes announce IPv6 prefix

www.google.com/intl/en/ipv6



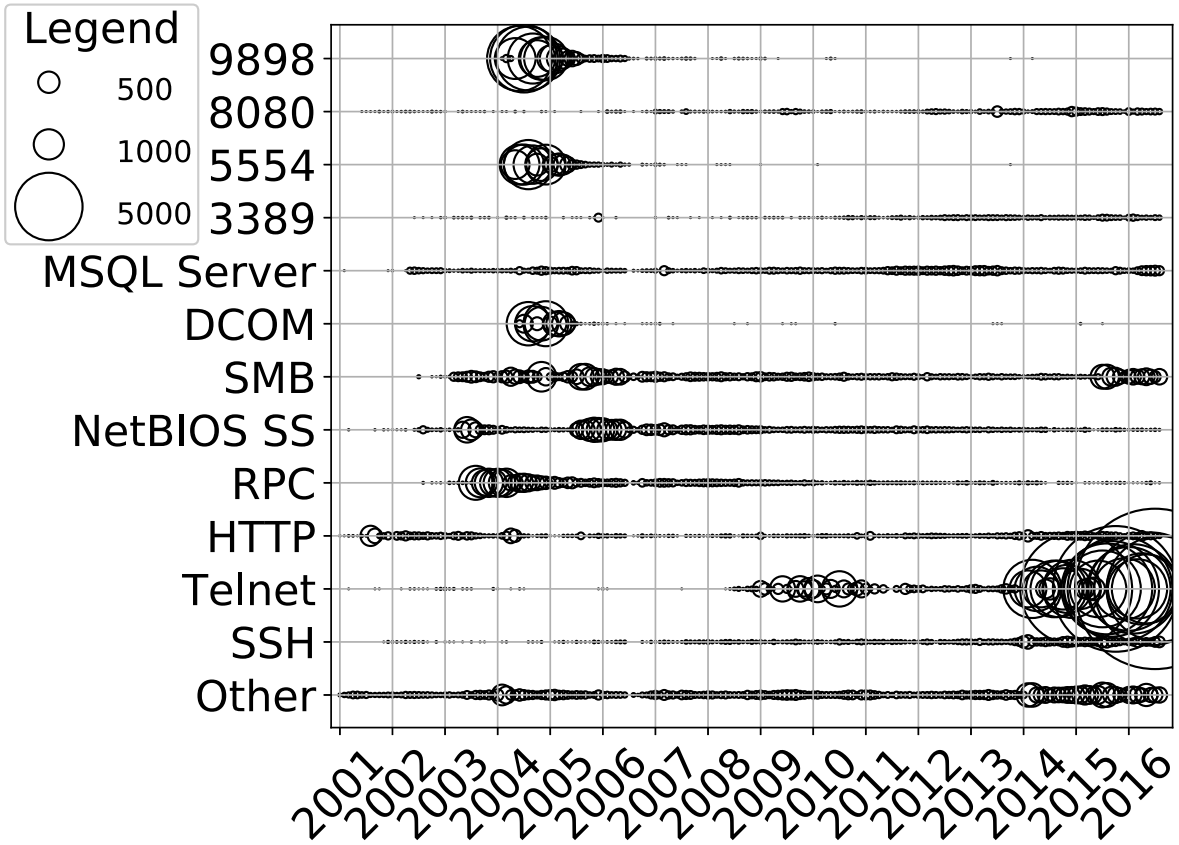

IPv6 security does matter?

# IPv4 scan

- Easy to scan whole IPv4 address space
  - Research purpose
    - Finding vulnerability
    - Detecting outage
  - Other purpose
- De facto scanning tool: Zmap
  - Takes 45 min with a single 10GE port
  - Many security studies used Zmap

Z.Durumeric, et al. “Zmap: Fast Internet-wide Scanning and Its Security Applications.” In USENIX Security Symposium 2013, pp.605-620, Washington D.C, 2013

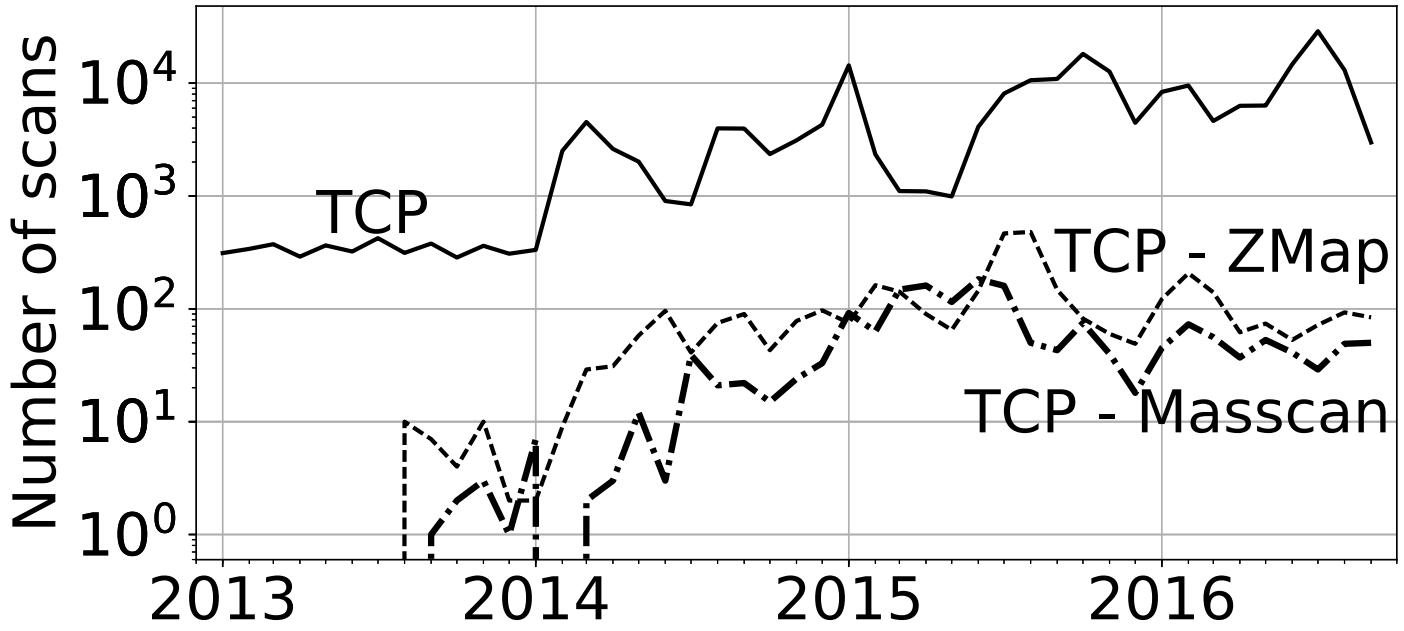
# IPv4 scan in the wild




- Measured at a transit link in WIDE (AS2500)

J.Mazel, et al. “Profiling Internet Scanners: Spatiotemporal Structures and Measurement Ethics.” In TMA 2017, Dublin, Ireland, 2017, IEEE/IFIP

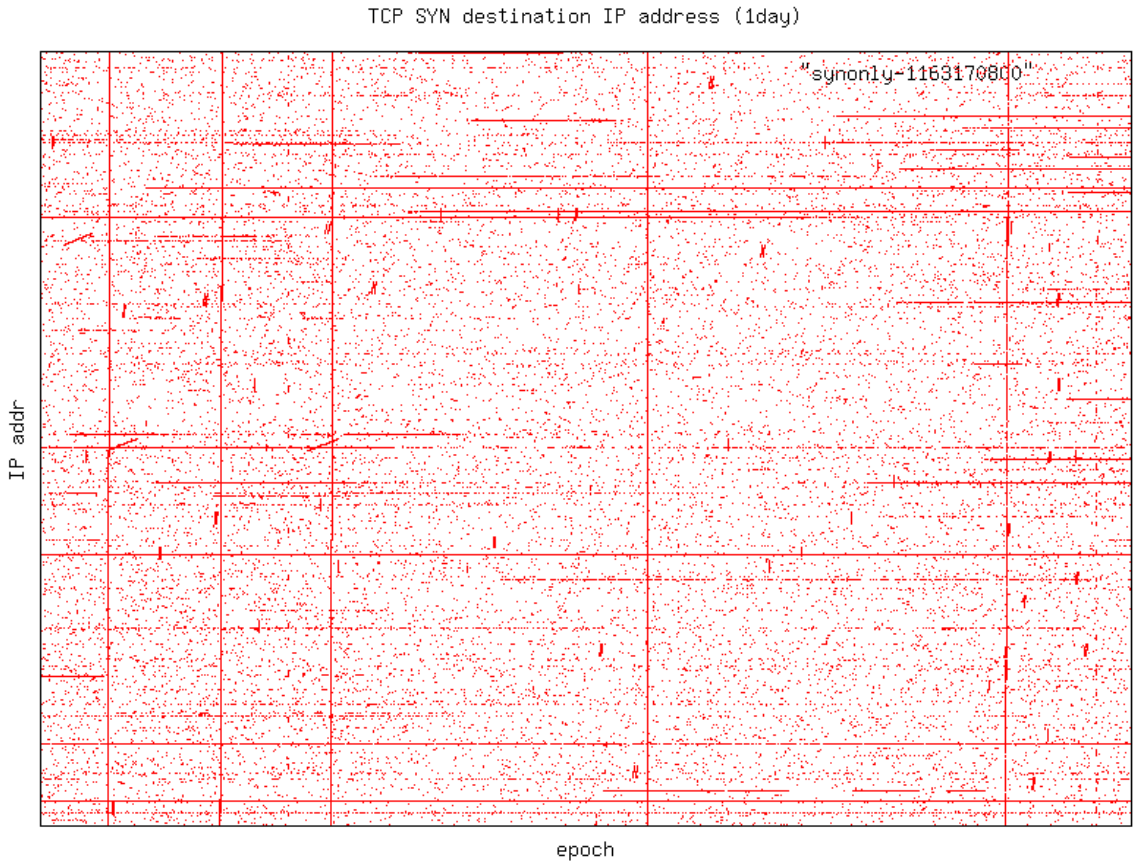
# IPv4 scan in the wild




- > 100 scanning IPs / month with scan tools

J.Mazel, et al. “Profiling Internet Scanners: Spatiotemporal Structures and Measurement Ethics.” In TMA 2017, Dublin, Ireland, 2017, IEEE/IFIP

# IPv4 scan in the wild




- Source: /18 IPv4 darknet

- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# Difficulty in IPv6 scan

- Huge address space
  - IPv4: $4.3 \times 10^9$ -> IPv6: $3.4 \times 10^{38}$
- SLAAC: Stateless Address Auto Configuration
  - EUI-64 based (RFC4291)
  - Semantically Opaque Interface Identifiers (RFC7217)

Random probing is not efficient!

Question: How to generate target IPv6 addresses?

# Making target hitlists for scan

- Passive data collection
  - Traffic data
- Active data collection
  - Alexa top 1M, rDNS (IPv4 -> A -> AAAA -> IPv6), traceroute, Zone files,
- Target generation
  - rDNS scan
  - Generating plausible addresses

# Passive and Active measurement

| Characteristic | Active sources | Passive sources | Traceroutes | CAIDA [5] |
|---|---|---|---|---|
| File size | 75MB | 5.4GB | 2.4MB | 40MB |
| Unique input lines | 2.7M | 149M | 1.3M | 618k |
| Unique targets | 2,699,573 | 148,631,234 | 109,554 | 102,580 |
| Unique ASes | 5,750 | 8,219 | 4,170 | 5,488 |
| Unique announced prefixes | 8,602 | 17,554 | 5,367 | 9,269 |
| AS coverage | 56.46% | **80.71%** | 41.00% | **53.90%** |
| ASes unique to source | 128 | 1,276 | 14 | 147 |
| Normalized ASes | 1,918.33 | 3,684.67 | 1,158.83 | 1,873.17 |
| Prefix coverage | 33.37% | **68.09%** | 20.76% | **36.00%** |
| Prefixes unique to source | 346 | 5,798 | 53 | 514 |
| Normalized prefixes | 3,199.25 | 10,302.58 | 1,569.92 | 3,681.25 |
| ICMPv6 response rate | 75.5% | 13.3% | n/a | 42.0% |
| Combined unique IPs | | **149,619,624** | | |
| Combined AS coverage | | **8,531 (83.77%)** | | |
| Combined prefix coverage | | **18,502 (71.77%)** | | |

O. Gasser, et al. “Scanning the IPv6 Internet: Towards a Comprehensive Hitlist.” In TMA 2016., Louvain La Neuve, Belgium, 2016. IFIP

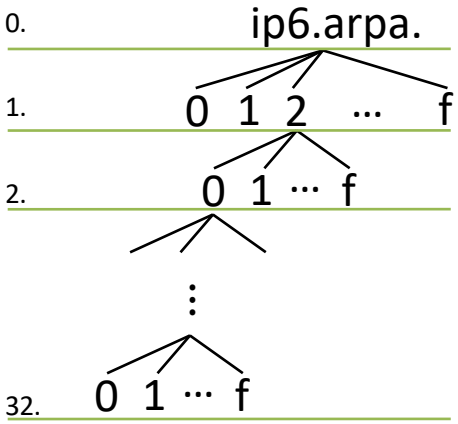
# Passive and Active measurement

| Characteristic | Active sources | Passive sources | Traceroutes | CAIDA [5] |
|---|---|---|---|---|
| File size | 75MB | 5.4GB | 2.4MB | 40MB |
| Unique input lines | 2.7M | 149M | 1.3M | 618k |
| Unique targets | 2,699,573 | 148,631,234 | 109,554 | 102,580 |
| Unique ASes | 5,750 | 8,219 | 4,170 | 5,488 |
| Unique announced prefixes | 8,602 | 17,554 | 5,367 | 9,269 |
| AS coverage | 56.46% | **80.71%** | 41.00% | **53.90%** |
| ASes unique to source | 128 | 1,276 | 14 | 147 |
| Normalized ASes | 1,918.33 | 3,684.67 | 1,158.83 | 1,873.17 |
| Prefix coverage | 33.37% | **68.09%** | 20.76% | **36.00%** |
| Prefixes unique to source | 346 | 5,798 | 53 | 514 |
| Normalized prefixes | 3,199.25 | 10,302.58 | 1,569.92 | 3,681.25 |
| ICMPv6 response rate | 75.5% | 13.3% | n/a | 42.0% |
| Combined unique IPs | | **149,619,624** | | |
| Combined AS coverage | | **8,531 (83.77%)** | | |
| Combined prefix coverage | | **18,502 (71.77%)** | | |

O. Gasser, et al. “Scanning the IPv6 Internet: Towards a Comprehensive Hitlist.” In TMA 2016., Louvain La Neuve, Belgium, 2016. IFIP

# Reverse DNS (rDNS) scan

- Crawling PTR registered name




1.0.0.2.ip6.arpa. -> NOERROR
0.1.0.0.2.ip6.arpa. -> NOERROR
0.0.1.0.0.2.ip6.arpa. -> NOERROR
⋮
1.1.0.0.2.ip6.arpa. -> NXDOMAIN -> (ignore subtree)
2.1.0.0.2.ip6.arpa. -> NOERROR
0.2.1.0.0.2.ip6.arpa. -> NOERROR
⋮

- Found 5.8M new addresses (from 73K /32 seeds)

T. Fiebig, et al. “Something from nothing (There): Collecting global IPv6 datasets from DNS,” In PAM’17, pp.30-43, Sydney, Australia, 2017.

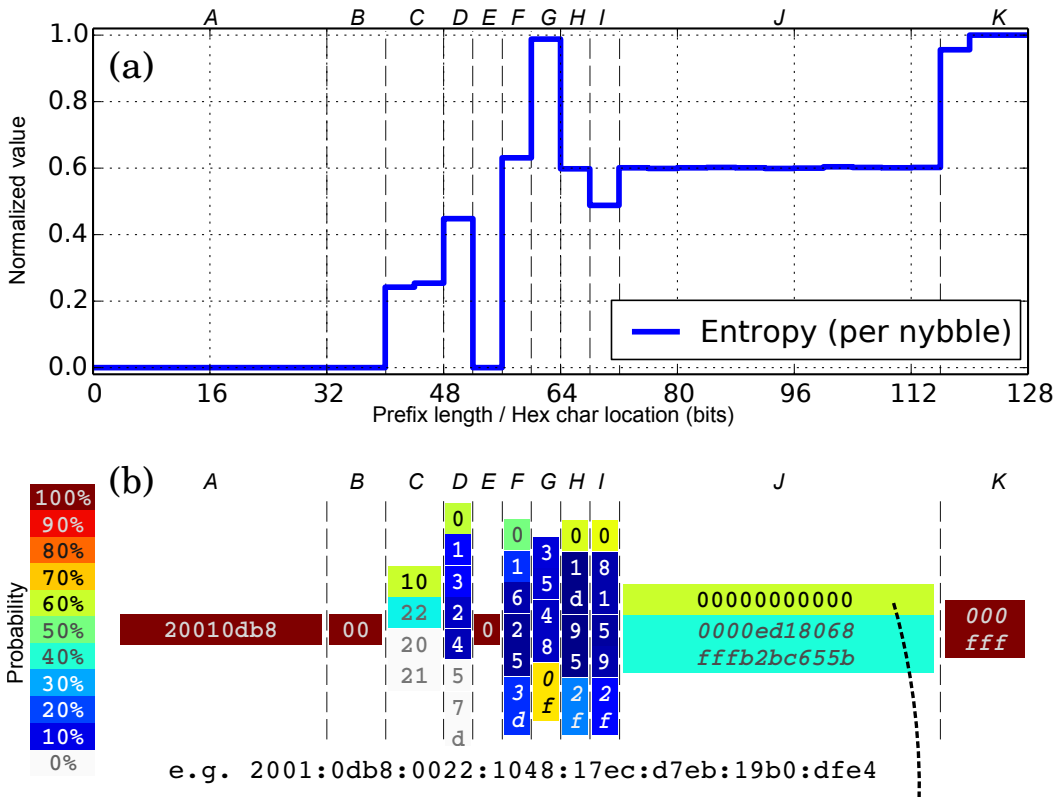
# Note on rDNS scan

- How often PTRs are registered in v6?
- Checking route advertised 175K /32 prefixes (2018.09)
  - No error: 7K
  - Serv fail: 33K
  - NX domain: 135K
- Sparsely registered!
  - Usage is limited?
  - Registration is limited?

新津, et al. “大規模IPv6アドレス収集手法の検討.” 信学技法 2018.09

# Generating plausible addresses

- Bayesian inference with nibble-based entropy




P. Foremski, et al. "Entropy/ip: Uncovering structure in ipv6 addresses." In IMC'16, pp.167-181, Santa Monica, USA, 2016.

# Generating plausible addresses

- Clustering dense part of nibbles

```
2 : : 1 : 1 0 0
2 : : 1 : 1 0 1
2 : : 1 : 2 0 0
2 : : 1 : 2 0 5
2 : : 3 : 1 0 a
2 : : 3 : 1 0 c
2 : : 3 : 2 0 f
```

2::?:?0?

A.Murdock, et al. ”Target Generation for Internet-wide IPv6 Scanning." In IMC’17, pp.167–181, London, UK, 2017.

# Summary: Hitlist generation

| | **Data source** | **#Addresses** |
|---|---|---|
| Gasser et al. [7] | traffic, traceroute, DNS AAAA/PTR | 150M |
| Defeche et al. [8] | BitTorrent peers | 1.5M |
| | **Method** | **#Addresses (seeds)** |
| Fiebig et al. [9] | rDNS scan | 5.8M (73K) |
| Foremski et al. [10] | Entropy | 770K (10K) |
| Murdock et al. [11] | Clustering | 55M (3.0M) |

新津, et al. “大規模IPv6アドレス収集手法の検討.” 信学技法 2018.09

- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# IPv6 scan response

| Characteristic | Active sources | Passive sources | Traceroutes | CAIDA [5] |
|---|---|---|---|---|
| File size | 75MB | 5.4GB | 2.4MB | 40MB |
| Unique input lines | 2.7M | 149M | 1.3M | 618k |
| Unique targets | 2,699,573 | 148,631,234 | 109,554 | 102,580 |
| Unique ASes | 5,750 | 8,219 | 4,170 | 5,488 |
| Unique announced prefixes | 8,602 | 17,554 | 5,367 | 9,269 |
| AS coverage | 56.46% | **80.71%** | 41.00% | **53.90%** |
| ASes unique to source | 128 | 1,276 | 14 | 147 |
| Normalized ASes | 1,918.33 | 3,684.67 | 1,158.83 | 1,873.17 |
| Prefix coverage | 33.37% | **68.09%** | 20.76% | **36.00%** |
| Prefixes unique to source | 346 | 5,798 | 53 | 514 |
| Normalized prefixes | 3,199.25 | 10,302.58 | 1,569.92 | 3,681.25 |
| ICMPv6 response rate | 75.5% | 13.3% | n/a | 42.0% |
| Combined unique IPs | | **149,619,624** | | |
| Combined AS coverage | | **8,531 (83.77%)** | | |
| Combined prefix coverage | | **18,502 (71.77%)** | | |

O. Gasser, et al. “Scanning the IPv6 Internet: Towards a Comprehensive Hitlist.” In TMA 2016., Louvain La Neuve, Belgium, 2016. IFIP

# IPv6 scan response

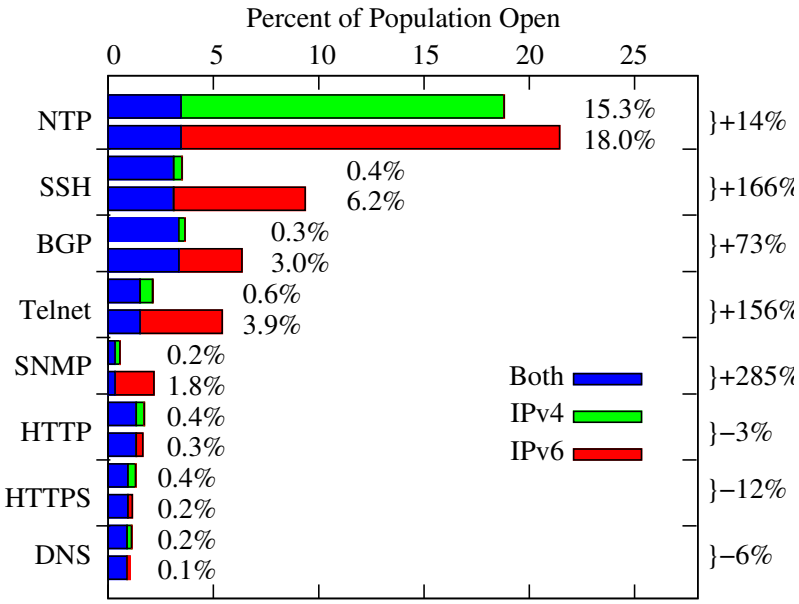
| Characteristic | Active sources | Passive sources | Traceroutes | CAIDA [5] |
|---|---|---|---|---|
| | 75MB | 5.4GB | 2.4 | |
| | 2.7M | 149M | 1. | |
| | 2,699,573 | 148,631,234 | 109, | |
| | 5,750 | 8,219 | 4, | |
| Unique anno prefixes | 8,602 | 17,554 | 5,367 | ,269 |
| AS coverage | 56.46% | **80.71%** | 41 % | **53.90%** |
| ASes unique to source | 128 | 1,276 | 14 | 147 |
| Normalized ASes | 1,918.33 | 3,684.67 | 1,158.83 | 1,873.17 |
| Prefix coverage | 33.37% | **68.09%** | 20.76% | **36.00%** |
| Prefixes unique to source | 346 | 5,798 | 53 | 514 |
| Normalized prefixes | 3,199.25 | 10,302.58 | 1,569.92 | 3,681.25 |
| ICMPv6 response rate | 75.5% | 13.3% | n/a | 42.0% |
| Combined unique IPs | | **149,619,624** | | |
| Combined AS coverage | | **8,531 (83.77%)** | | |
| Combined prefix coverage | | **18,502 (71.77%)** | | |

Coverage is so-so, but more responses

Coverage is good, but a few responses

O. Gasser, et al. “Scanning the IPv6 Internet: Towards a Comprehensive Hitlist.” In TMA 2016., Louvain La Neuve, Belgium, 2016. IFIP

# IPv4/v6 scan response




- Ping to 25K Dual stack routers
- More open ports in IPv6

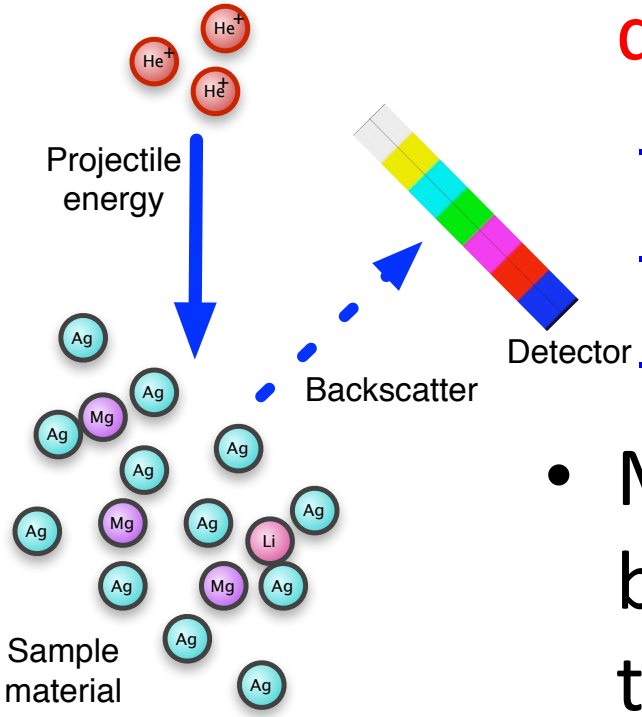
J.Czyz et al. “Don’t Forget to Lock the Back Door! Characterization of IPv6 Network Security Policy.” In NDSS 2016., San Diego, 2016. Internet Society

- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# Detecting Network-wide Scans

- Passive backbone/IXP traffic collection
- Darknet (aka network telescope)
  - Routed but no legitimate hosts
- Distributed firewall logs (e.g., SANS)
- DNS backscatter

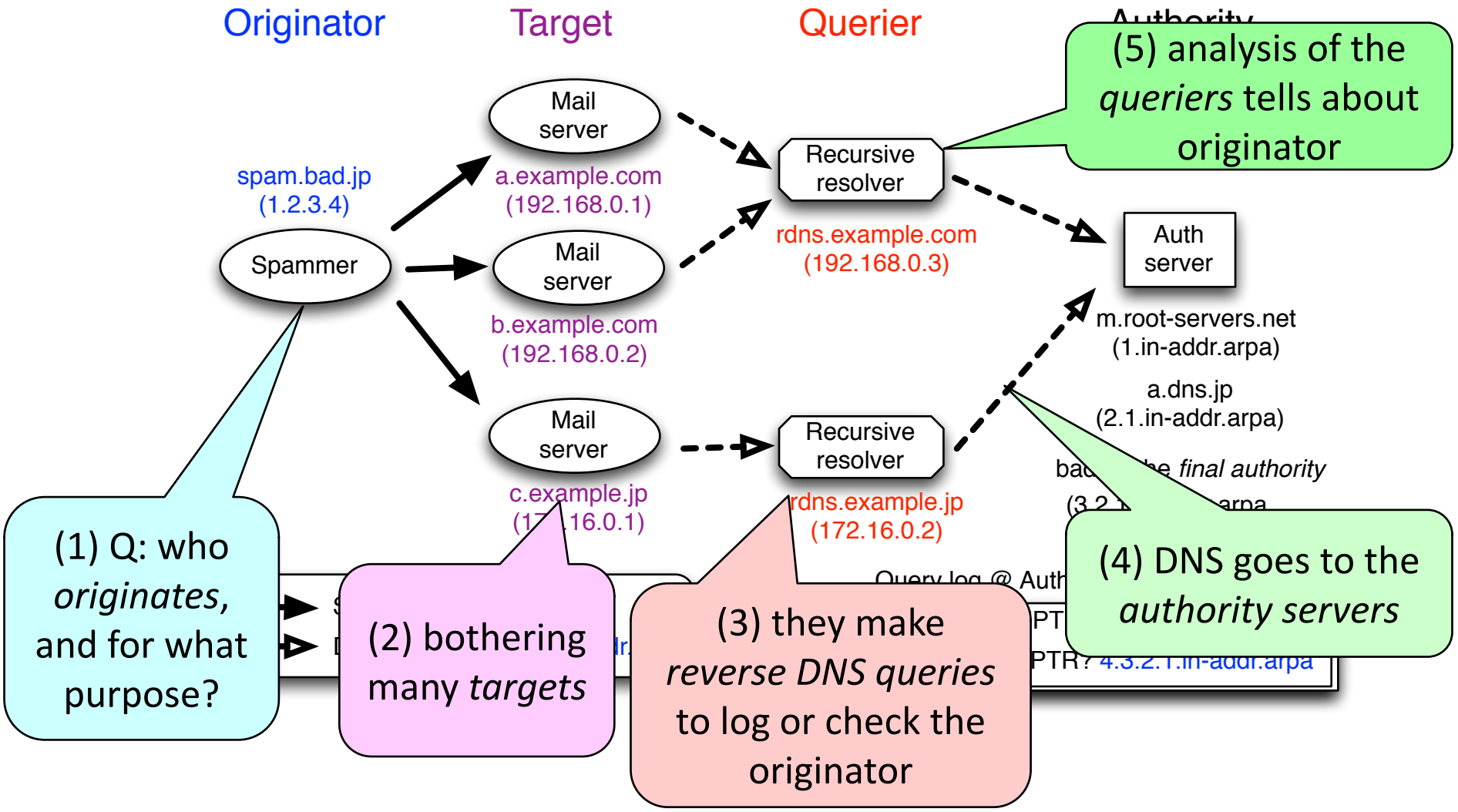
# Key idea of DNS backscatter




- Large event triggers reverse DNS queries near target automatically
  - SMTP server: hostname of spammer
  - Firewall: hostname of scanner
  - Web server: hostname of web crawler
- Many reverse DNS queries (DNS backscatter) at auth server are hint to identify events

Fukuda, et al. ”Detecting Malicious Activities with DNS Backscatter over Time." In IEEE/ACM Transactions on Networking, vol.25, no.5, pp.3203–3218, 2017.

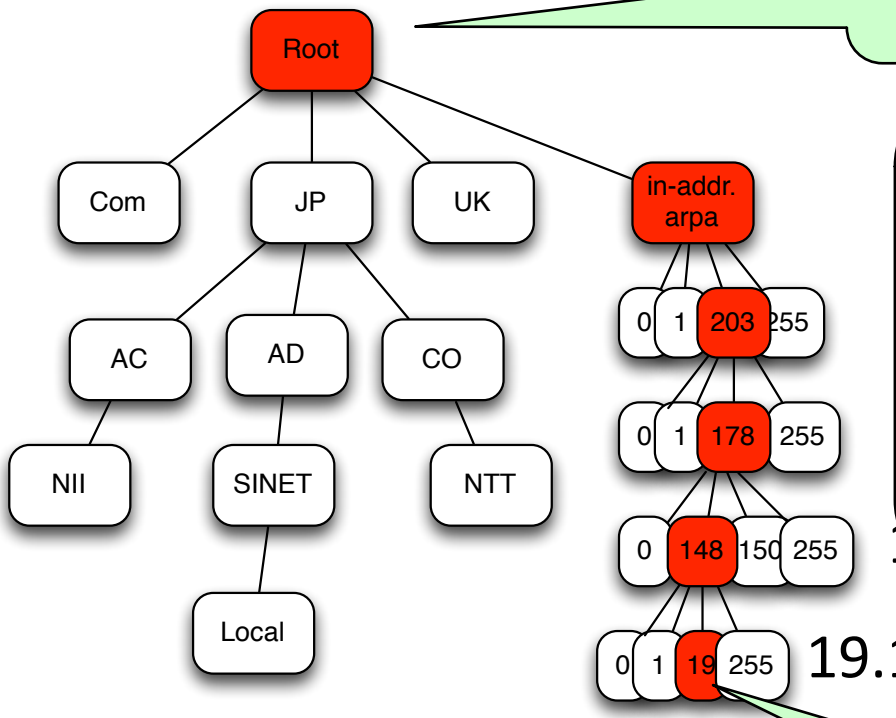
# Detecting Events through DNS backscatter



# Different Authorities See Different Amount of Backscatter

Ex: 203.178.18.19 ->

More events and more attenuation



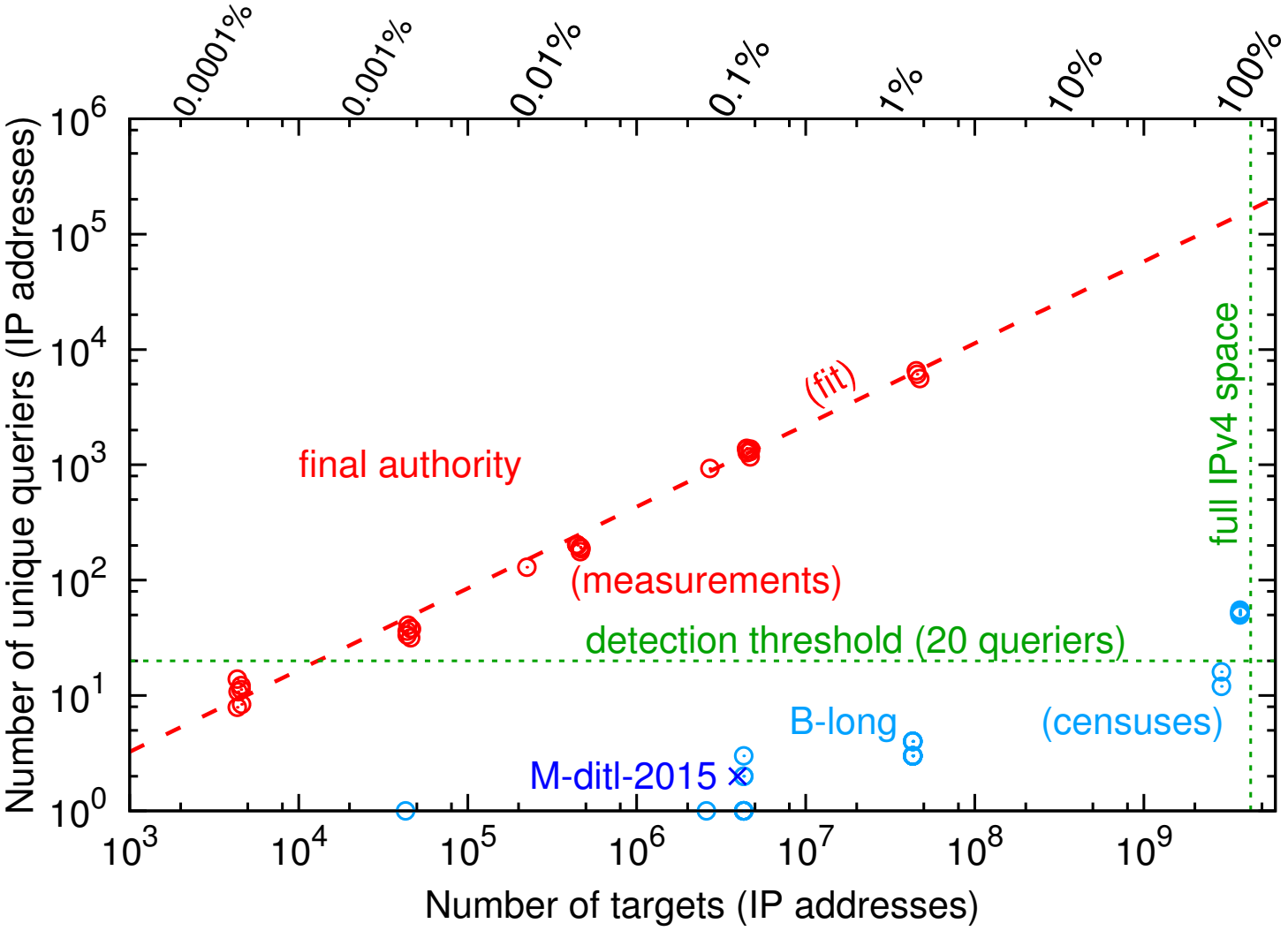

Zone -> Authority

Tradeoff between visibility of events & attenuation of signal

19.148.178.203.in-addr.arpa -> Final Auth

Less events and less attenuation

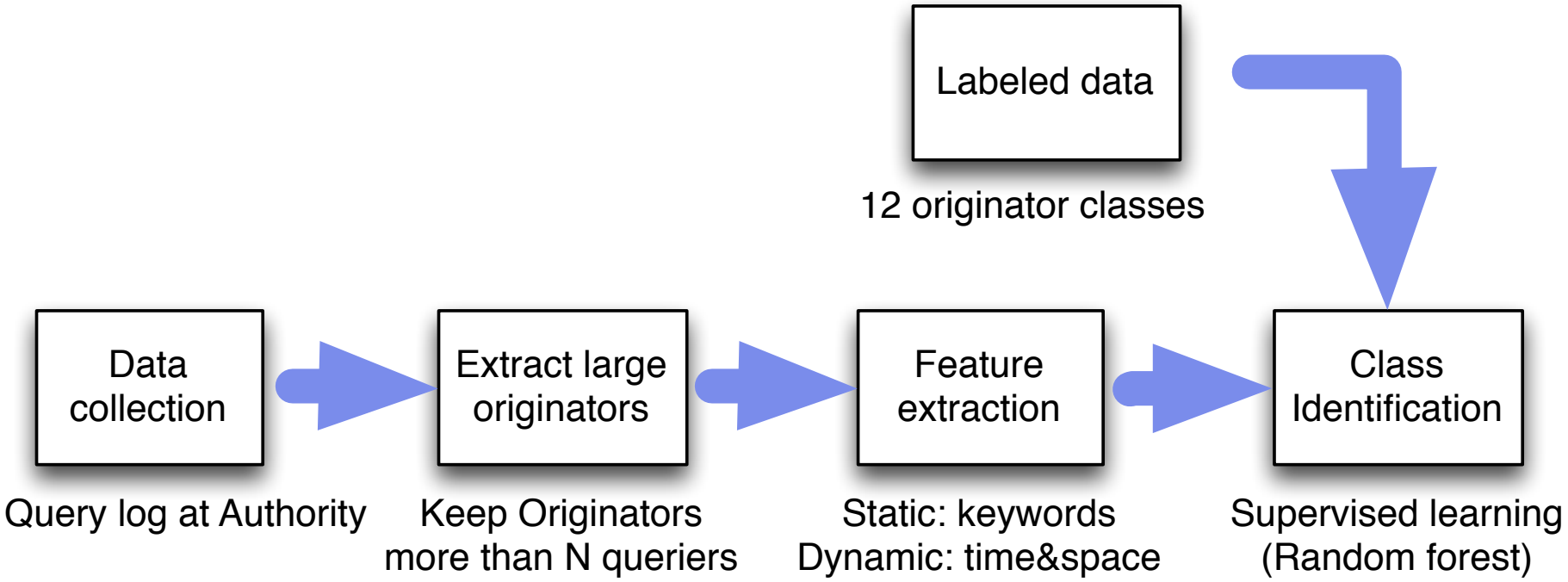
# IPv4 backscatter sensitivity



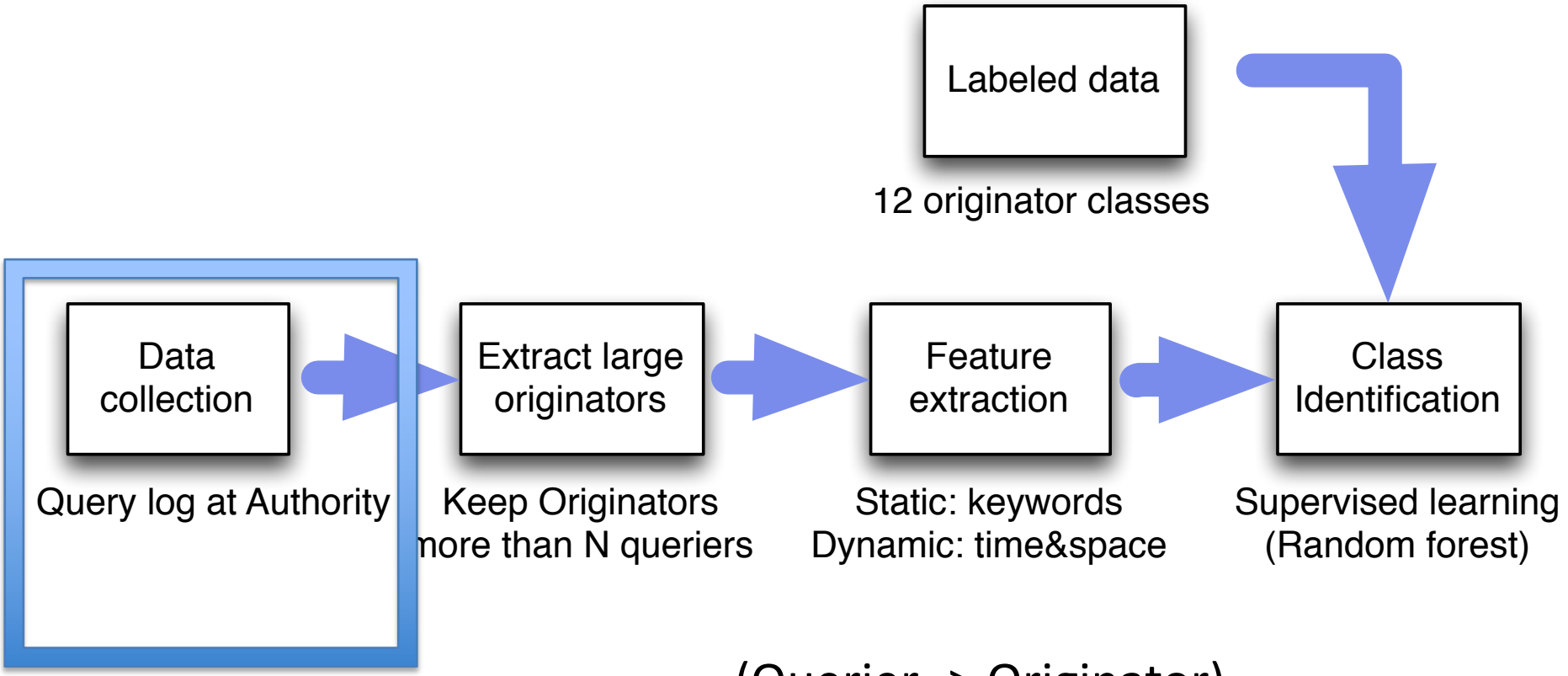
# Advantages

- Deployable
  - Centralized monitoring at DNS authority
- Privacy friendly
  - Information is on queriers NOT originators
  - Reverse queries are generated automatically
  - Focus on large events (ignore small users)
- Robust against malicious originators
- Can infer different class of originator (e.g., scanner) with Machine Learning

# Identification process



# Identification process




Query logs: from any server

2 Root DNS

JP-DNS

Local authority

(Querier -> Originator)

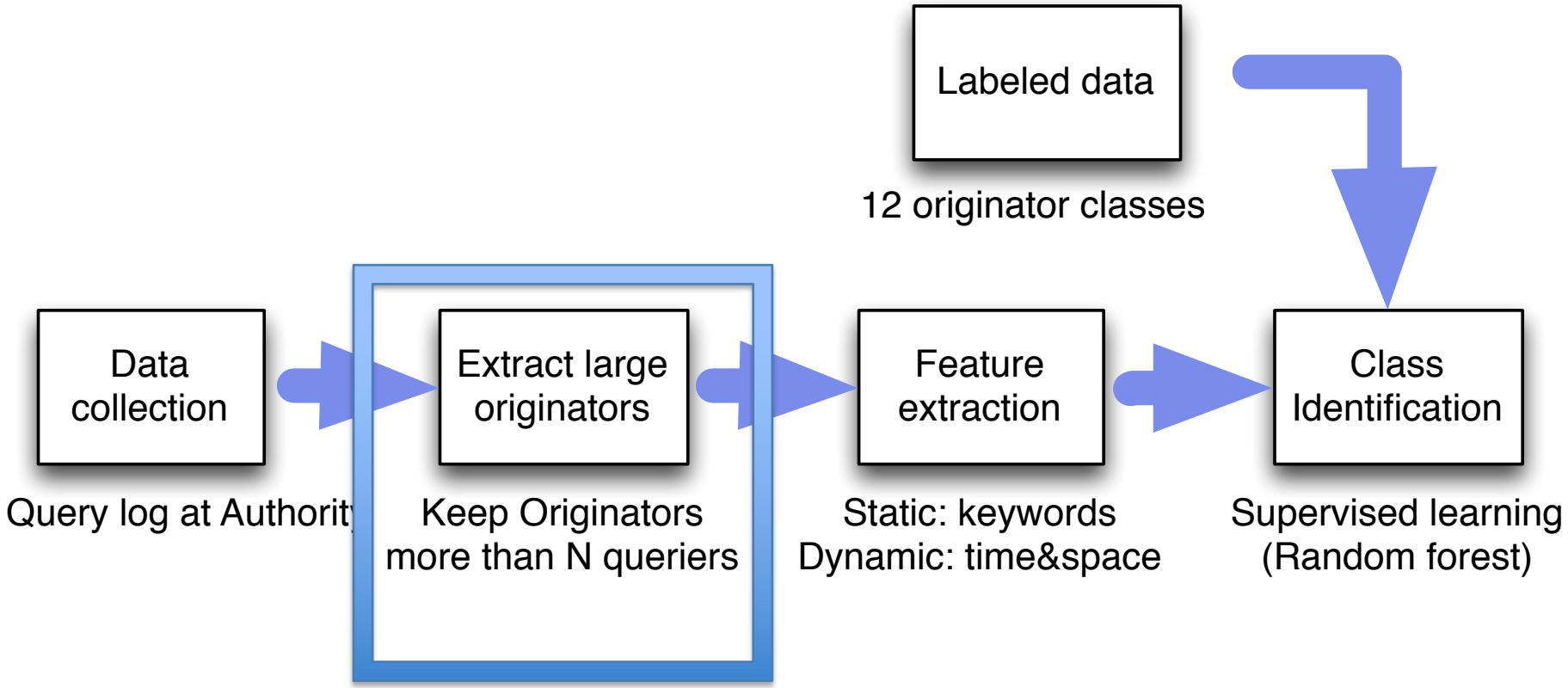
mail1.example.com -> 1.2.3.4

fw0.foo.jp -> 5.6.7.8

spam.good.jp -> 1.2.3.4

ns0.example.jp -> 7.8.9.10

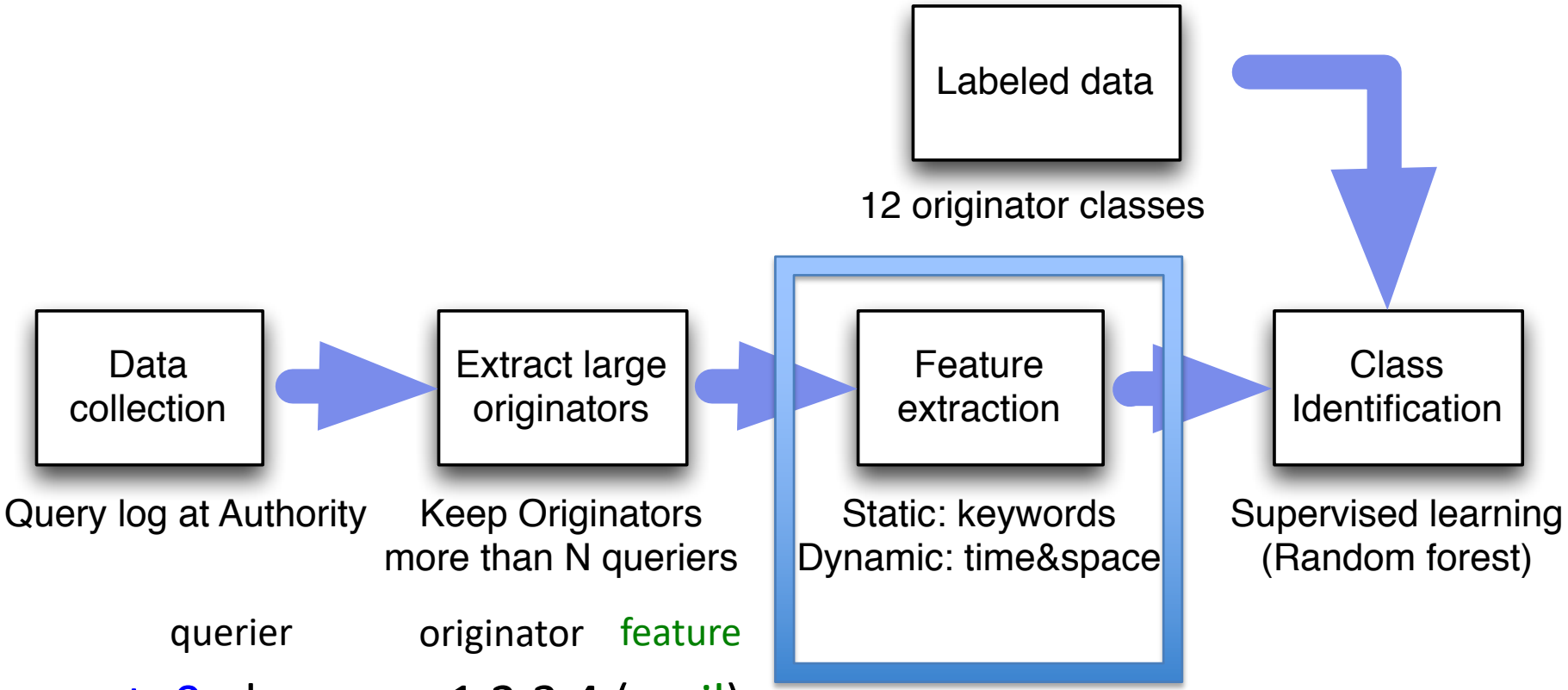
# Identification process




Query logs: Large originators
(N > 20)

(Originator -> #Uniq Queriers)
1.2.3.4 -> 50 (Keep)
5.6.7.8 -> 10 (Drop)
7.8.9.10 -> 5 (Drop)

# Identification process




querier originator feature

smtp0.ok.com -> 1.2.3.4 (mail)

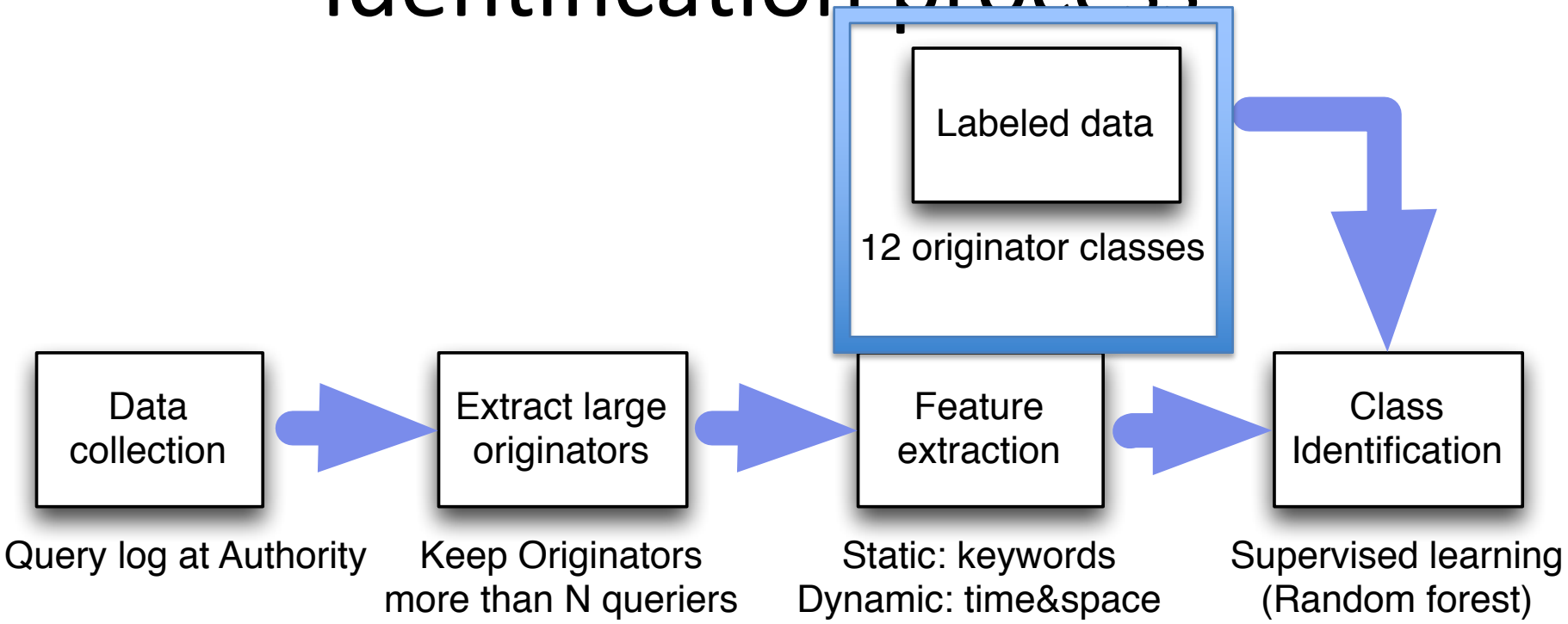
mail1.example.com -> 1.2.3.4 (mail)

spam.foo.jp -> 1.2.3.4 (spam)

ns0.bar.jp -> 1.2.3.4 (ns)

Feature vector: 1.2.3.4: <mail 50%, ns 25%, spam 25%, #AS, qps...>

# Identification process

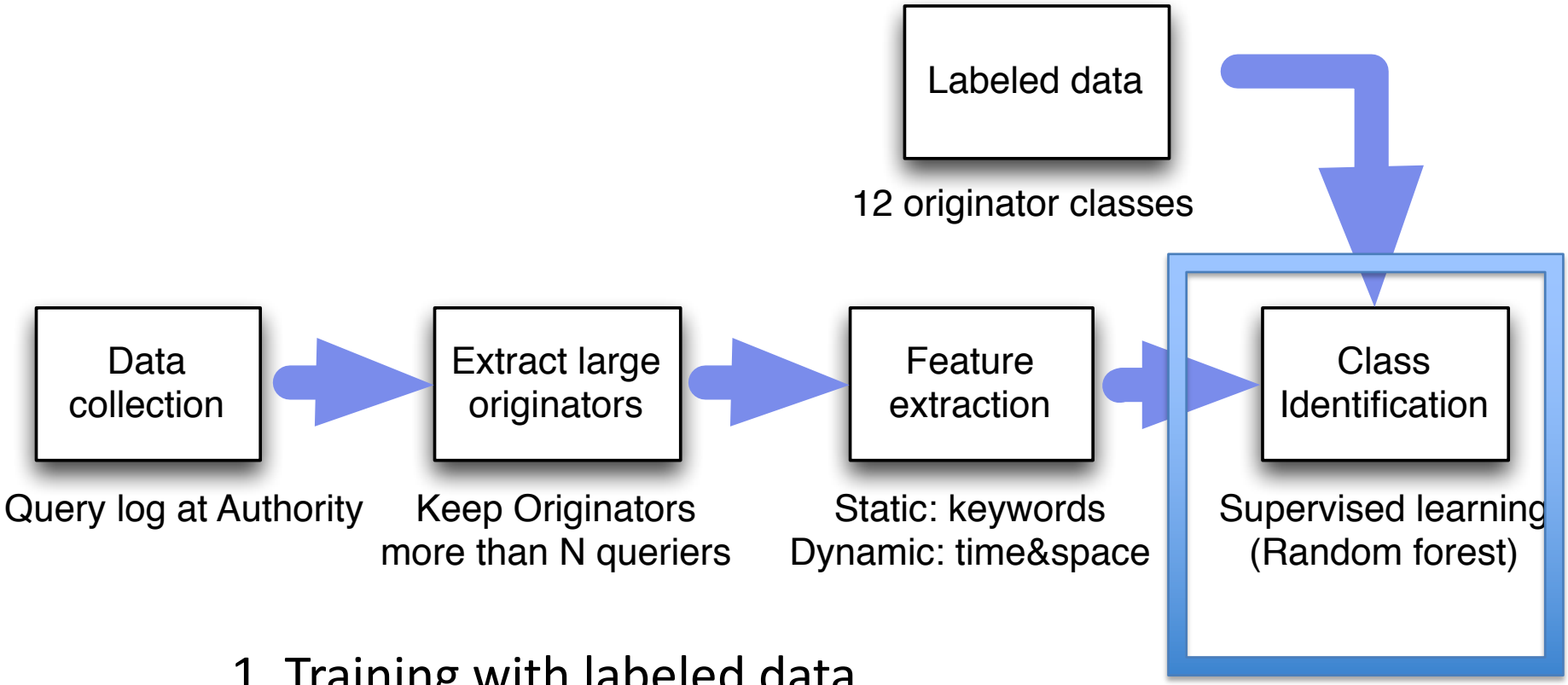



Ground truth from dozen of public sources (darknet, DNLBL…)

(Class: ad-tracker, cdn, cloud, crawler, dns, mail, scan, spam…)

| Originator | Feature vector | class |
|---|---|---|
| 11.12.13.14 | <mail 45%, ns 20%, spam 5%,…> | mail |
| 21.22.23.24 | <mail 60%, ns 15%, spam 5%,…> | mail |
| 31.32.33.34 | <mail 45%, ns 15%, spam 15%,…> | spam |

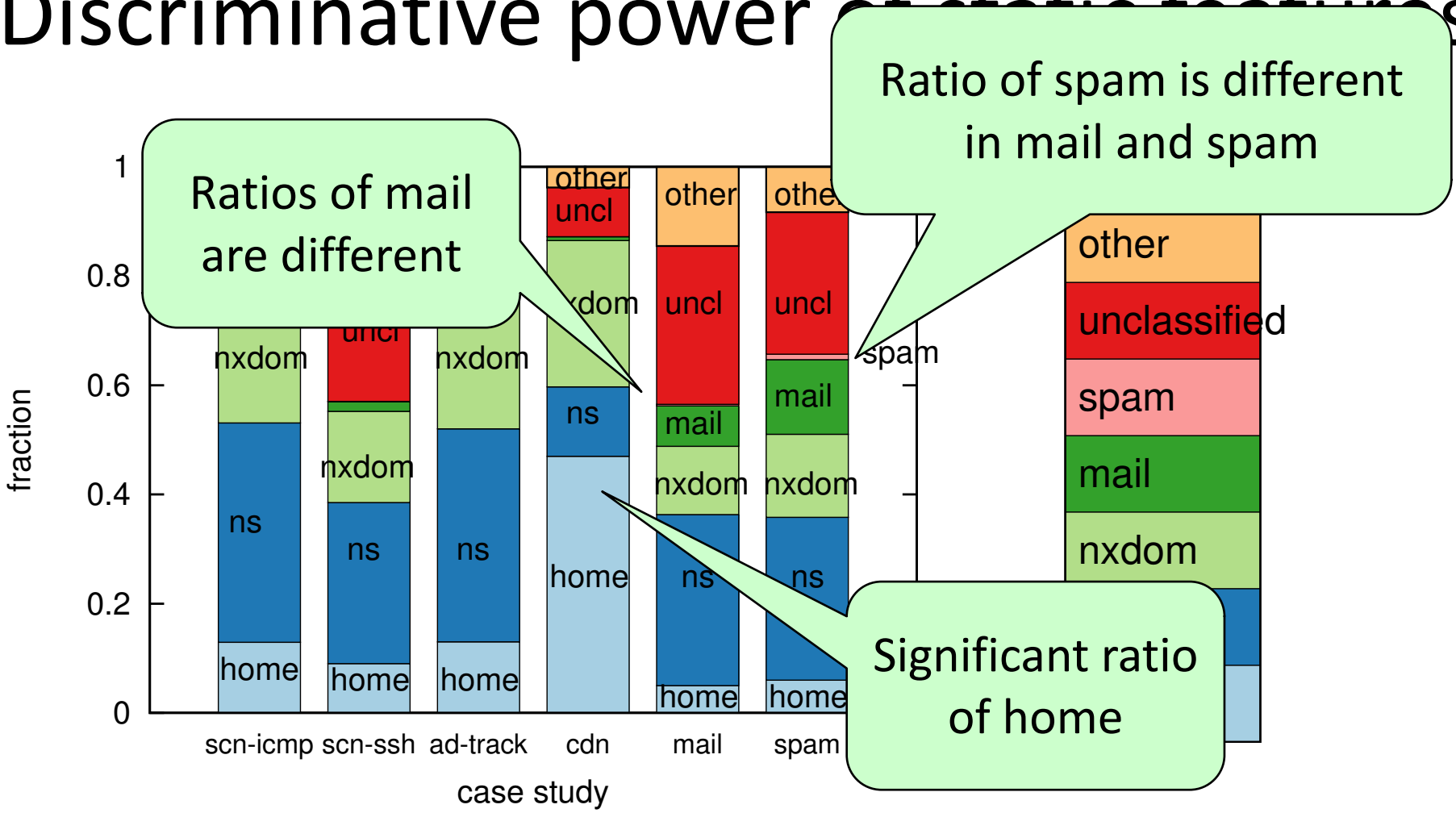
# Identification process




1. Training with labeled data
2. Classify test data with classification matrix

Originator 1.2.3.4 <mail 50%, ns 25%, spam 25%, #AS, qps…>
-> spam

# Discriminative power of static features




Different mixes of features allow distinguishing different classes of events

# Picking the best ML algorithm

| dataset | algorithm | accuracy | precision | recall | F1-score |
|---|---|---|---|---|---|
| | CART | 0.66 | 0.63 | 0.60 | 0.61 |
| JP | **RF** | **0.78** | **0.82** | **0.76** | **0.79** |
| ditl | SVM | 0.73 | 0.74 | | |
| B | CART | 0.48 | 0.48 | 0.45 | |
| post- | **RF** | **0.62** | **0.66** | **0.60** | |
| ditl | SVM | 0.38 | 0.50 | 0.32 | |
| | CART | 0.53 | 0.52 | 0.49 | 0.51 |
| M | **RF** | **0.68** | **0.74** | **0.63** | **0.68** |
| ditl | SVM | 0.60 | 0.6 | | |
| | CART | 0.61 | 0.6 | | |
| M | **RF** | **0.79** | **0.8** | | |
| sampled | SVM | 0.72 | 0.7 | | |

RandomForest is best

Hope to improve with better training data

- Cross validation with 3 ML algorithms
- Num classes: 12, labeled data:200-800
- Precision: 70-80% (imbalanced dataset problem)

- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# How to adapt to IPv6

- Number of queriers is much smaller in v6
  - ML doesn't work well
  - More aggregation (1day -> 1week)
- We directly infer the type of originator
  - Originator's Keyword and AS
    - smtp.foo.bar -> mail
  - Originator-Querier relation
    - All belongs to the same AS -> not network-wide events
  - Matching with Blacklists
    - Spam, scan, etc

Fukuda, et al. "Who Knocks at the IPv6 Doors? Detecting IPv6 Scanning" In IMC'18, Boston, MA, 2018. (to appear)

# Classification

- Major services: Google, MS, FB, Yahoo (by ASN)
- CDN: Akamai,,, (by ASN)
- DNS: Zone files, keyword
- NTP: NTP pool, keyword
- Mail/Web
- Iface: traceroute, keyword
- Tunnels: 6to4, Teledo
- Spam: Blacklist
- Scan: Blacklist, backbone data

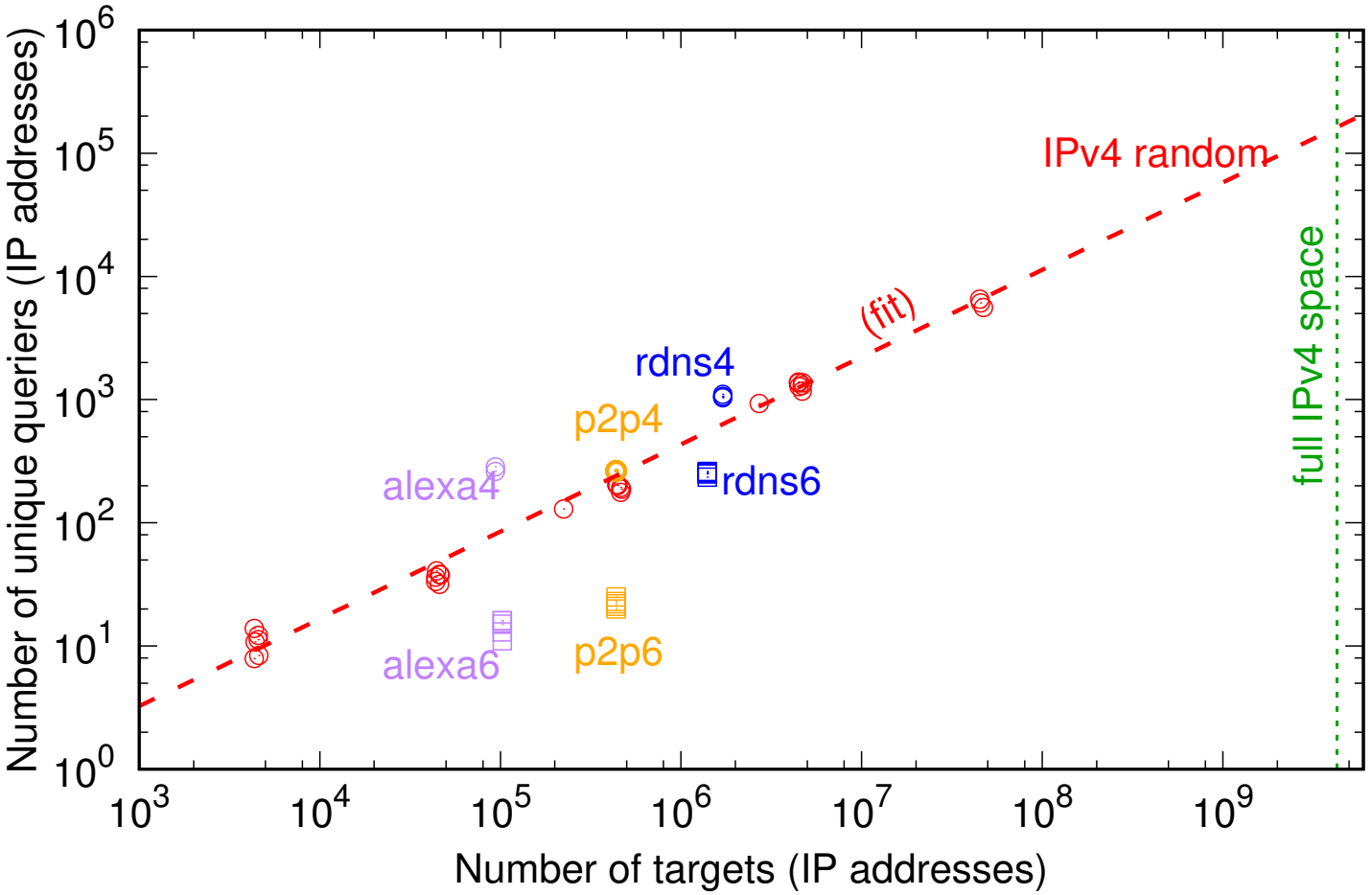
- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# Analyzing IPv6 backscatter sensitivity

- Custom IPv6 network scanner
  - Multiple proto/service (ICMP, TCP22, TCP80,UDP53, UDP123)
  - Uniq source IP for each target
- Local authoritative server
  - TTL = 1s
- Three hitlists

| Label | # addrs | Description |
|---|---|---|
| Alexa | 10k | Alexa 1M; servers |
| rDNS | 1.4M | Reverse DNS |
| P2P | 40k | P2P Bittorrent; clients |

# V4/V6 backscatter sensitivity




- x10 smaller DNS backscatter

# Application response

| **type** | **icmp6** (ping) | | **tcp22** (ssh) | | **tcp80** (web) | | **udp53** (DNS) | | **udp123** (NTP) | |
|---|---|---|---|---|---|---|---|---|---|---|
| queries | | | | … | 1476509 | 100% | … | | | |
| expected reply | 928953 | 62.9% | 410421 | 27.8% | 661182 | 44.8% | 69965 | 4.7% | 140893 | 9.5% |
| other reply | 145264 | 9.8% | 205446 | 13.9% | 201627 | 13.7% | 672171 | 45.5% | 371044 | 25.1% |
| no reply | 402292 | 27.2% | 860642 | 58.3% | 613700 | 41.5% | 734373 | 49.4% | 964572 | 65.3% |
| expected v4 reply | - | 57.8% | - | 30.0% | - | 35.4% | - | 6.3% | - | 5.9% |

- Source: rDNS
- ICMP6 > tcp80 > tcp22 > udp123 > udp53
- No significant difference between v4 and v6

# Backscatter and response

| | **icmp6** (ping) | | **tcp22** (ssh) | | **tcp80** (web) | | **udp53** (DNS) | | **udp123** (NTP) | |
|---|---|---|---|---|---|---|---|---|---|---|
| v6 backscatter | 1809 | (0.12%) | 774 | (0.05%) | 1020 | (0.07%) | 653 | (0.04%) | 746 | (0.05%) |
| w/expected reply | 1371 | 75.8% (0.09%) | 365 | 47.2% (0.03%) | 597 | 58.5% (0.04%) | 137 | 21.0% (0.01%) | 134 | 18.0% (0.01%) |
| w/other reply | 44 | 2.4% (0.002%) | 94 | 12.1% (0.006%) | 87 | 8.5% (0.006%) | 265 | 40.6% (0.02%) | 183 | 24.5% (0.01%) |
| w/no reply | 394 | 21.8% (0.03%) | 315 | 40.7% (0.02%) | 336 | 32.9% (0.02%) | 251 | 38.4% (0.02%) | 429 | 57.5% (0.03%) |
| v4 backscatter | 4478 | (0.30%) | 2731 | (0.18%) | 3094 | (0.21%) | 3961 | (0.27%) | 4045 | (0.27%) |

- Backscatter: how often firewall logs?
- V4 backscatter is x3-5 larger than v6
- Backscatter with expected reply: ok but logged

- Introduction
- Network scans: state of the art
  - Hitlist generation
  - IPv4/IPv6 sensitivity
- Finding IPv6 scanners with DNS backscatter
  - DNS backscatter
  - How to adapt to IPv6
  - Measurement results
    - Sensitivity
    - Detecting IPv6 scanners

# Datasets

- DNS backscatter:
  - B-root DNS server
  - Full capture (31M uniq querier-originator pairs)
- Backbone: MAWI traffic repository
  - Transit link in AS2500 (WIDE)
  - Tcpdump in 15 min each day
- Darknet: SINET darknet
  - v6 /37 advertised from AS2907 (SINET)

# Classification results (B-root)

| Category | Count (mean/week) | % total |
|---|---|---|
| **Services:** | | |
| **Content Provider** | **4722** | **70.24** |
| Facebook | 3653 | 54.34 |
| Google | 727 | 10.82 |
| Microsoft | 329 | 4.89 |
| Yahoo | 13 | 0.19 |
| **CDN** | **286** | **4.25** |
| **Well-known service** | **815** | **12.12** |
| DNS | 337 | 5.01 |
| NTP | 414 | 6.16 |
| mail (SMTP) | 42 | 0.62 |
| web (HTTP) | 22 | 0.33 |
| **Minor service** | **268** | **3.99** |
| other services | 83 | 1.23 |
| qhost | 185 | 2.75 |
| **Routers:** | | |
| **Router** | **288** | **4.28** |
| iface | 256 | 3.81 |
| near-iface | 32 | 0.48 |
| **Tunnel** | **216** | **3.21** |
| Teredo/6to4 | 207 | 3.08 |
| tor | 9 | 0.12 |
| **Potential Abuse:** | | |
| **Abuse** | **128** | **1.90** |
| spam | 17 | 0.25 |
| scan | 16 | 0.24 |
| unknown (potential abuse) | 95 | 1.41 |
| **Total** | **6723** | **100.00** |

# Classification results (B-root)

| Category | Count (mean/week) | % total |
|---|---|---|
| **Services:** | | |
| **Content Provider** | **4722** | **70.24** |
| Facebook | 3653 | 54.34 |
| Google | 727 | 10.82 |
| Microsoft | 329 | 4.89 |
| Yahoo | 13 | 0.19 |
| **CDN** | **286** | **4.25** |
| **Well-known service** | **815** | **12.12** |
| DNS | 337 | 5.01 |
| NTP | 414 | 6.16 |
| mail (SMTP) | 42 | 0.62 |
| web (HTTP) | 22 | 0.33 |
| **Minor service** | **268** | **3.99** |
| other services | 83 | 1.23 |
| qhost | 185 | 2.75 |
| **Routers:** | | |
| **Router** | **288** | **4.28** |
| iface | 256 | 3.81 |
| near-iface | 32 | 0.48 |
| **Tunnel** | **216** | **3.21** |
| Teredo/6to4 | 207 | 3.08 |
| tor | 9 | 0.12 |
| **Potential Abuse:** | | |
| **Abuse** | **128** | **1.90** |
| spam | 17 | 0.25 |
| scan | 16 | 0.24 |
| unknown (potential abuse) | 95 | 1.41 |
| **Total** | **6723** | **100.00** |

DNS backscatter mainly detect content provider and CDN (benign)

Traceroute triggers many backscatters

Clear abuse and potential ones

# Scanners confirmed at MAWI

| | IP | MAWI | | | Backscatter | Dark | ASN | info |
|---|---|---|---|---|---|---|---|---|
| | | #days | port | scan type | #weeks | #weeks | | |
| (a) | 2001:48e0:205:2::/64 | 6 | TCP80 | Gen | 1 (5) | 1 | 40498 | New Mexico Lambda Rail |
| (b) | 2a02:418:6a04:178::/64 | 2 | ICMP | rand IID | 2 (4) | 0 | 29691 | Nine, CH |
| (c) | 2a02:c207:3001:8709::/64 | 2 | TCP80 | rand IID | 2 (2) | 0 | 51167 | Contabo, DE |
| (d) | 2a03:f80:40:46::/64 | 2 | ICMP | rDNS | 2 (3) | 0 | 5541 | ADNET-Telecom, RO |
| (e) | 2405:4800:103:2::/64 | 2 | ICMP | rDNS | 0 (4) | 0 | 18403 | FPT-AS-AP, VN |
| (f) | 2a03:4000:6:e12f::/64 | 1 | ICMP | rDNS | 0 (0) | 0 | 197540 | NETCUP-GmbH, DE |
| (g) | 2800:a4:c1f:6f01::/64 | 1 | ICMP | rDNS | 0 (0) | 0 | 6057 | ANTEL, UY |

- 4 scanners are detected both MAWI and backscatter
- 3 small scanners are missed in backscatter
- Darknet only finds 1 scanner

# Scanners confirmed at MAWI

Research scanner @ Berkeley

| | IP | MAWI | | | Backscatter | Dark | ASN | info |
|---|---|---|---|---|---|---|---|---|
| | | #days | port | scan type | #weeks | #weeks | | |
| (a) | 2001:48e0:205:2::/64 | 6 | TCP80 | Gen | 1 (5) | 1 | 40498 | New Mexico Lambda Rail |
| (b) | 2a02:418:6a04:178::/64 | 2 | ICMP | rand IID | 2 (4) | 0 | 29691 | Nine, CH |
| (c) | 2a02:c207:3001:8709::/64 | 2 | TCP80 | rand IID | 2 (2) | 0 | 51167 | Contabo, DE |
| (d) | 2a03:f80:40:46::/64 | 2 | ICMP | rDNS | 2 (3) | 0 | 5541 | ADNET-Telecom, RO |
| (e) | 2405:4800:103:2::/64 | 2 | ICMP | rDNS | 0 (4) | 0 | 18403 | FPT-AS-AP, VN |
| (f) | 2a03:4000:6:e12f::/64 | 1 | ICMP | rDNS | 0 (0) | 0 | 197540 | NETCUP-GmbH, DE |
| (g) | 2800:a4:c1f:6f01::/64 | 1 | ICMP | rDNS | 0 (0) | 0 | 6057 | ANTEL, UY |

- 4 scanners are detected both MAWI and backscatter
- 3 small scanners are missed in backscatter
- Darknet only finds 1 scanner

# Scanners confirmed at MAWI

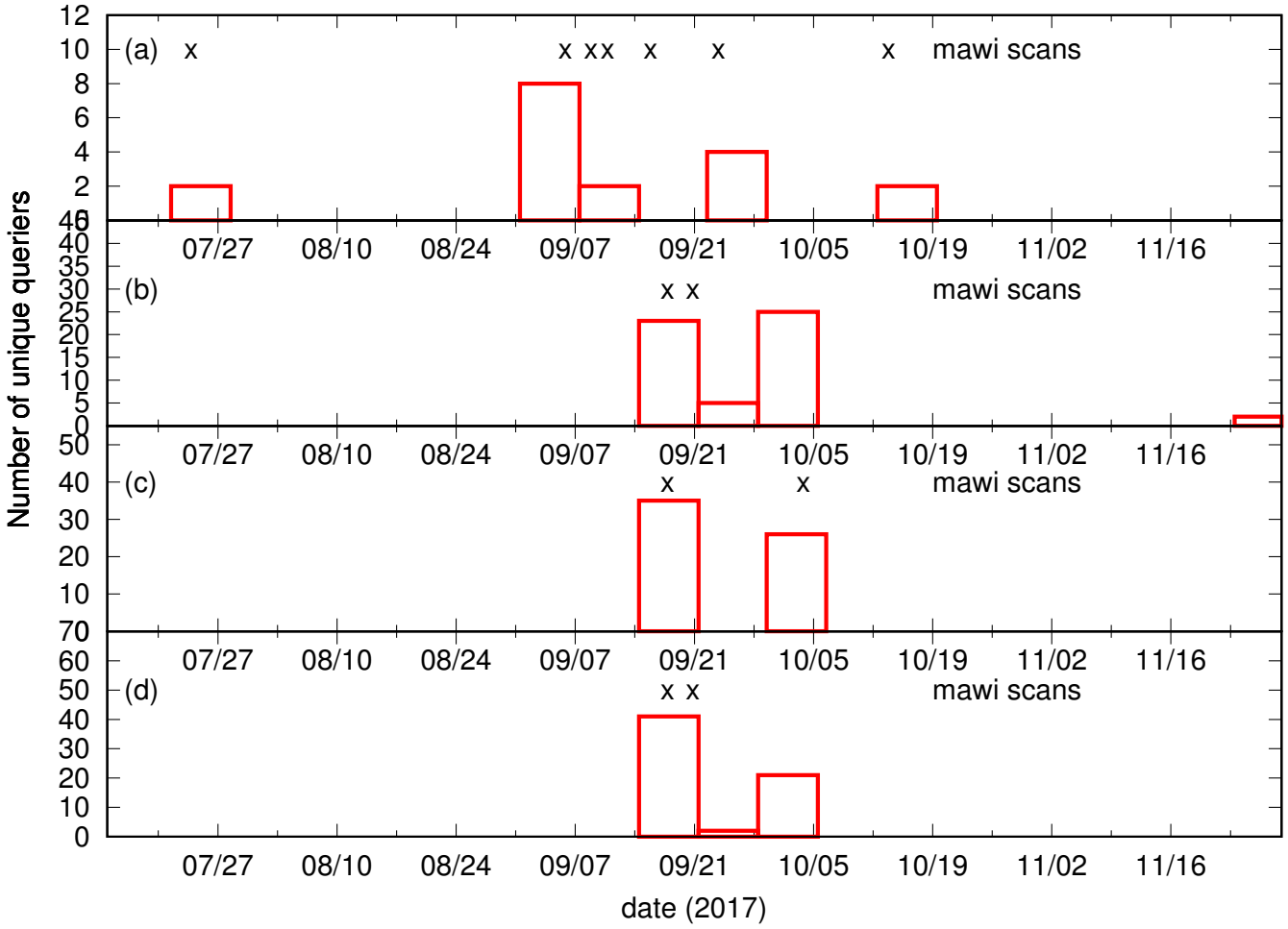
| | IP | MAWI | | | Backscatter | Dark | ASN | info |
|---|---|---|---|---|---|---|---|---|
| | | #days | port | scan type | #weeks | #weeks | | |
| (a) | 2001:48e0:205:2::/64 | 6 | TCP80 | Gen | 1 (5) | 1 | 40498 | New Mexico Lambda Rail |
| (b) | 2a02:418:6a04:178::/64 | 2 | ICMP | rand IID | 2 (4) | 0 | 29691 | Nine, CH |
| (c) | 2a02:c207:3001:8709::/64 | 2 | TCP80 | rand IID | 2 (2) | 0 | 51167 | Contabo, DE |
| (d) | 2a03:f80:40:46::/64 | 2 | ICMP | rDNS | 2 (3) | 0 | 5541 | ADNET-Telecom, RO |
| (e) | 2405:4800:103:2::/64 | 2 | ICMP | rDNS | 0 (4) | 0 | 18403 | FPT-AS-AP, VN |
| (f) | 2a03:4000:6:e12f::/64 | 1 | ICMP | rDNS | 0 (0) | 0 | 197540 | NETCUP-GmbH, DE |
| (g) | 2800:a4:c1f:6f01::/64 | 1 | ICMP | rDNS | 0 (0) | 0 | 6057 | ANTEL, UY |

- rDNS hitlists: 4 scanners
- Rand IID hitlists: 2 scanners (e.g., 2001:db8:1::10)
- Generative hitlists: 1 scanner (6Gen)

# (Why darknet misses?)

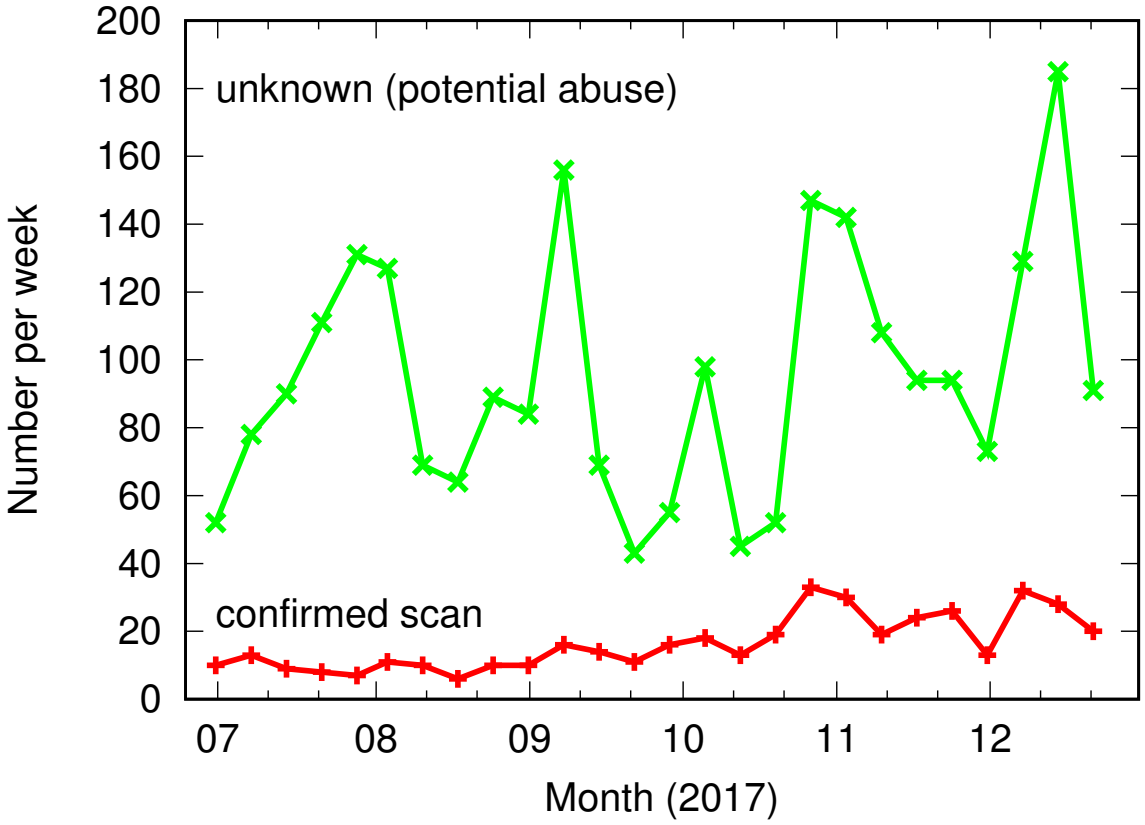
- Current scanners rely on hitlists
  - Never scan IP addresses not in hitlists
- Require to register to hitlists
  - Add v4 PTR, A, and AAAA records!
  - Add v6 PTR records!

# MAWI scan and backscatter



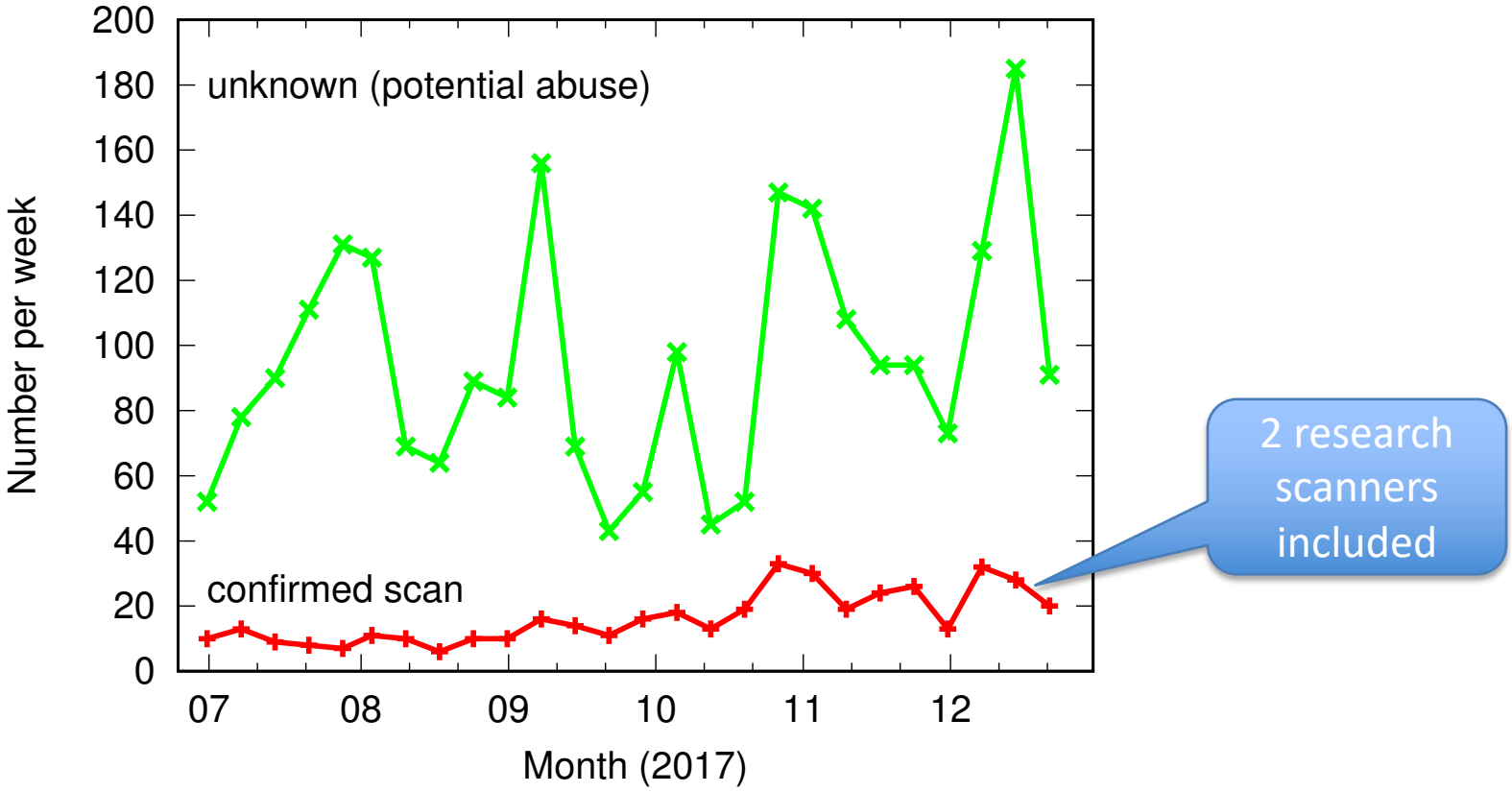

- Strong temporal correlation!

# Abuse over time




- The number of scans increases over time

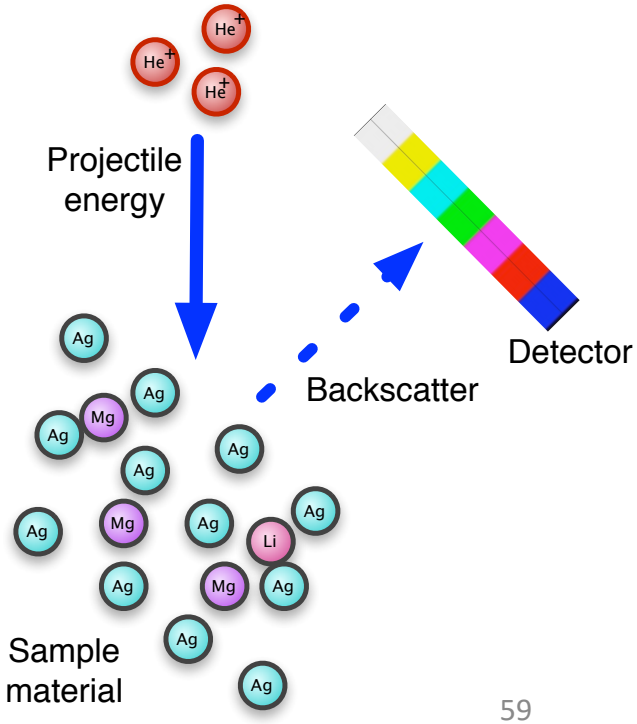
# Abuse over time




- The number of scans increases over time

# Conclusion

- IPv6 scanning: state of the art
  - Hitlist generation
  - Comparison with IPv4 scans
- DNS backscatter - a new data source for Internet-wide events
  - Adapt to IPv6
  - Works well
- IPv6 scanners increase over time



# Thanks to


- JSPS Kakenhi Grant (15KK0019, 18H03237)
- Collaborators:
  - John Heidemann (USC/ISI)
  - Yudai Aratsu (U.Tokyo)
- Others:
  - Johan Mazel (ANSSI, FR)
  - Romain Fontugne (IIJ)
  - Kenjiro Cho (IIJ)
  - Yoshinobu Matsuzaki (IIJ)