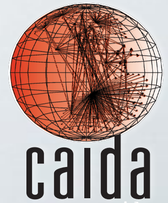
Learning to Extract and Use ASNs in Hostnames

Matthew Luckie - University of Waikato
Alexander Marder - CAIDA / UC San Diego
Marianne Fletcher - University of Waikato
Bradley Huffaker - CAIDA / UC San Diego
k claffy - CAIDA / UC San Diego

IMC 2020



Which AS operates these three routers?

Router #1

173.205.63.202
62.115.147.199
195.22.195.27

Router #2

77.67.94.154
213.248.68.105
63.218.52.253

Router #3

216.221.157.90
64.125.14.5

Which AS operates these three routers?

Router #1
- 173.205.63.202
- 62.115.147.199
- 195.22.195.27

Router #2
- 77.67.94.154
- 213.248.68.105
- 63.218.52.253

Router #3
- 216.221.157.90
- 64.125.14.5

More examples cited in the paper

A growing body of work depends on getting the answer correct

Inferring Persistent Interdomain Congestion

A First Comparative Characterization of Multi-cloud Connectivity in Today's Internet

Latency Imbalance Among Internet Load-Balanced Paths: A Cloud-Centric View

Revealing the Load-balancing Behavior of YouTube Traffic on Interdomain Links

Investigating the Causes of Congestion on the African IXP substrate

Unintended consequences: Effects of submarine cable deployment on Internet routing

Router Ownership Inference Techniques: Limited Validation

Towards an Accurate AS-Level Traceroute Tool SIGCOMM 2003

Scalable and Accurate Identification of AS-Level Forwarding Paths INFOCOM 2004

Approach: Increase congruity between colocated BGP and traceroute paths.

Validation: AT&T router configuration.

Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers PAM 2010

Approach: evaluate heuristics based on degrees.

Validation: Tier-1 and Tier-2 AS, five R&E networks.

Router Ownership Inference Techniques: Limited Validation

bdrmap: Inference of Borders Between IP Networks — IMC 2016

Approach: infer operator AS for first-hop interdomain links.
Validation: Tier-1 network, two access networks, one R&E.

MAP-IT: Multipass Accurate Passive Inferences from Traceroute — IMC 2016

Approach: infer operator AS through graph refinement.
Validation: Two Tier-1 networks (DNS), one R&E.

Pushing the Boundaries with bdrmapIT: Mapping Router Ownership at Internet Scale — IMC 2018

Approach: integrate MAP-IT and bdrmap heuristics.
Validation: Tier-1 network, access network, two R&E.

Which AS operates these three routers?

Router #1: Inferred AS 15133

3257	173.205.63.202
1299	62.115.147.199
6762	195.22.195.27

Router #2: Inferred AS 3491

3257	77.67.94.154
1299	213.248.68.105
3491	63.218.52.253

Router #3: Inferred AS 6461

3257	216.221.157.90
6461	64.125.14.5

origin AS of corresponding prefix

Why is this hard?

It is challenging to infer the operator of an **AS border router** as the router could have IP addresses that belong to their neighbors.

Which AS operates these three routers?

Router #1: Inferred AS 15133

3257	173.205.63.202
1299	62.115.147.199
6762	195.22.195.27

Router #2: Inferred AS 3491

3257	77.67.94.154
1299	213.248.68.105
3491	63.218.52.253

Router #3: Inferred AS 6461

3257	216.221.157.90
6461	64.125.14.5

Overview: bdrmapIT

- **bdrmapIT** uses a complex set of heuristics to infer an AS that operates the router

Which AS operates these three routers?

Overview: bdrmapIT

Router #1: Inferred AS 15133

3257	173.205.63.202
1299	62.115.147.199
6762	195.22.195.27

Router #2: Inferred AS 3491

3257	77.67.94.154
1299	213.248.68.105
3491	63.218.52.253

Router #3: Inferred AS 6461

3257	216.221.157.90
6461	64.125.14.5

- **bdrmapIT** uses a complex set of heuristics to infer an AS that operates the router
  - **Graph refinement**: use AS ownership inferences of subsequent routers to infer owner of a given router

Which AS operates these three routers?

Overview: bdrmapIT

Router #1: Inferred AS 15133

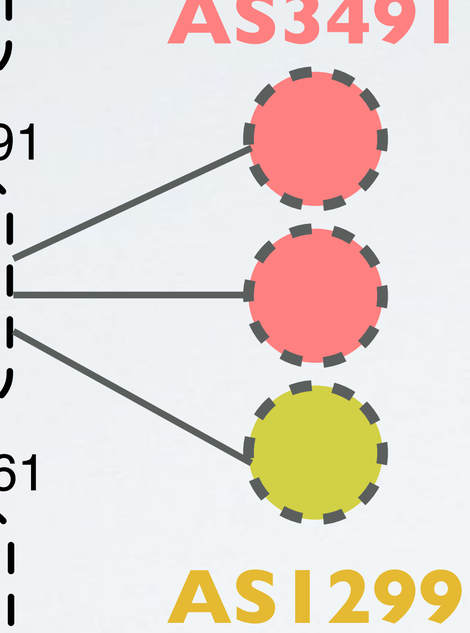
3257	173.205.63.202
1299	62.115.147.199
6762	195.22.195.27

Router #2: Inferred AS 3491

3257	77.67.94.154
1299	213.248.68.105
3491	63.218.52.253

Router #3: Inferred AS 6461

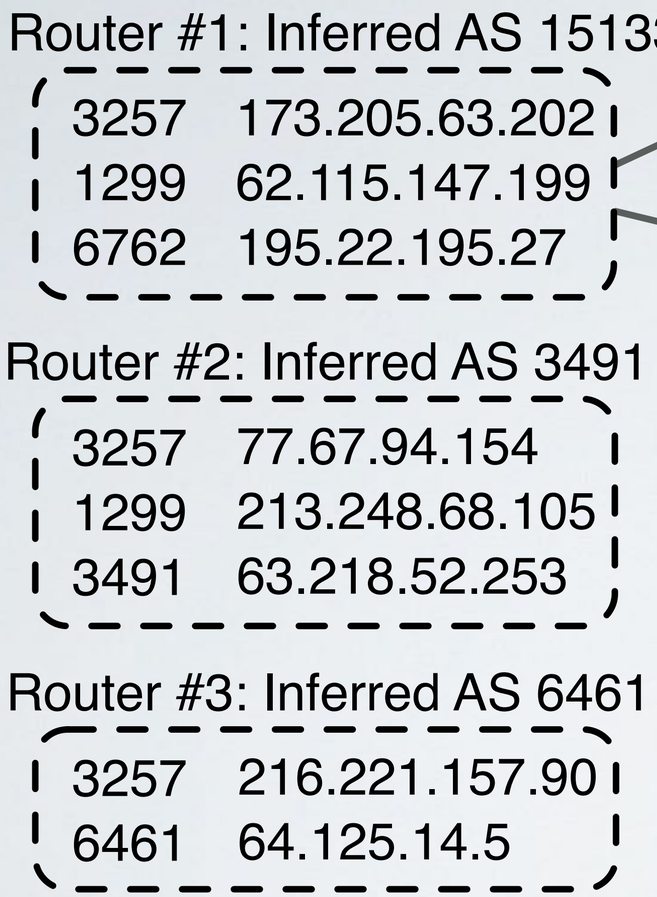
3257	216.221.157.90
6461	64.125.14.5



- **bdrmapIT** uses a complex set of heuristics to infer an AS that operates the router
  - **Graph refinement**: use AS ownership inferences of subsequent routers to infer owner of a given router

Infer **AS3491** operates Router #2 through majority vote

Which AS operates these three routers?



Overview: bdrmapIT

AS15133

AS15133

- **bdrmapIT** uses a complex set of heuristics to infer an AS that operates the router
  - **Graph refinement**: use AS ownership inferences of subsequent routers to infer owner of a given router
  - **Last hop**: use origin ASes of destination IP addresses probed to reason about edge routers with no subsequent routers

Which AS operates these three routers?

Router #1: Inferred AS 15133

3257	173.205.63.202
1299	62.115.147.199
6762	195.22.195.27

Router #2: Inferred AS 3491

3257	77.67.94.154
1299	213.248.68.105
3491	63.218.52.253

Router #3: Inferred AS 6461

3257	216.221.157.90
6461	64.125.14.5

Overview: bdrmapIT

AS15133

- **bdrmapIT** uses a complex set of heuristics to infer an AS that operates the router

AS15133

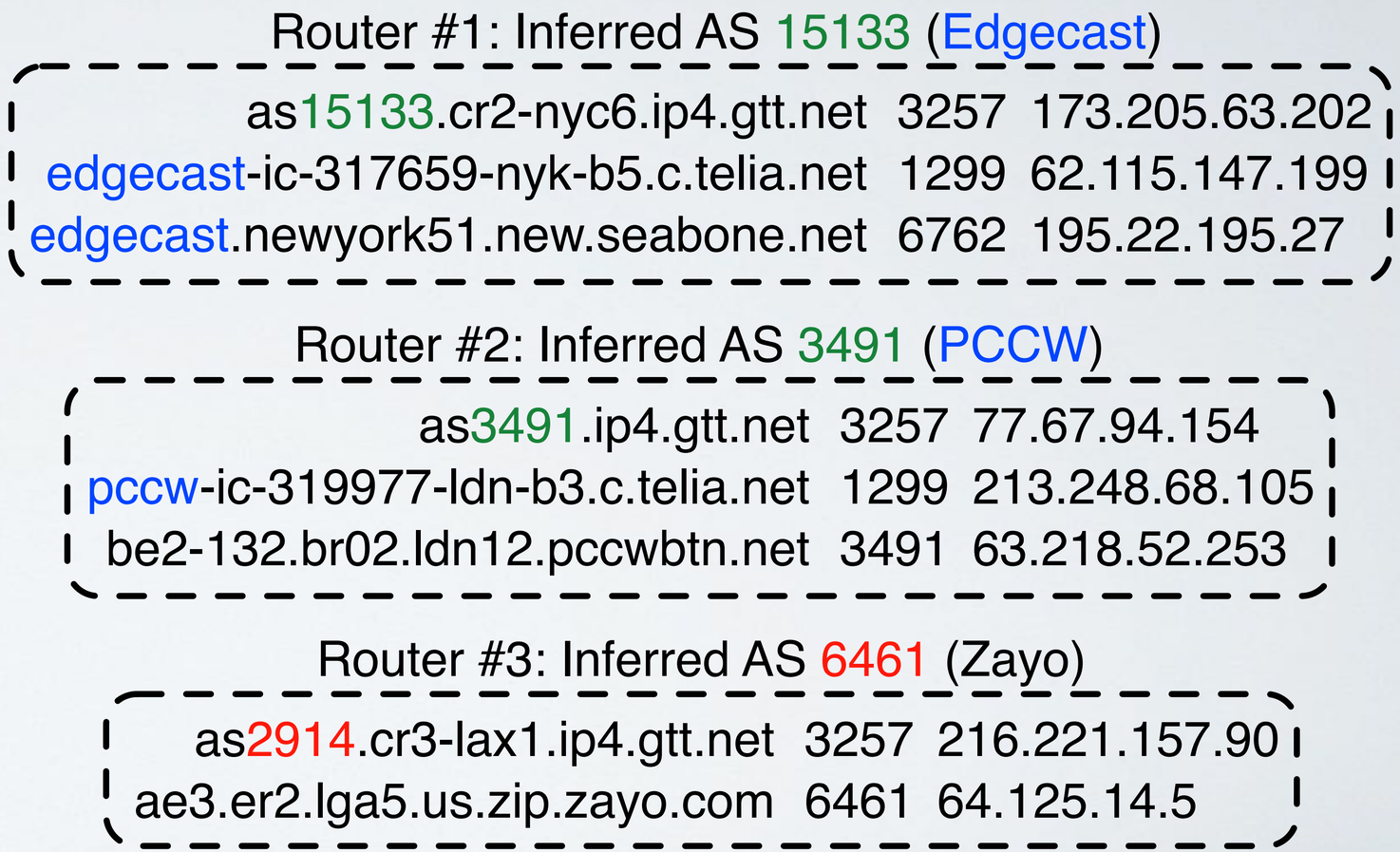
- **Graph refinement**: use AS ownership inferences of subsequent routers to infer owner of a given router

- **Last hop**: use origin ASes of destination IP addresses probed to reason about edge routers with no subsequent routers

Infer **AS15133** operates Router #1 as it is in the path to AS15133

Our approach: use information in hostnames

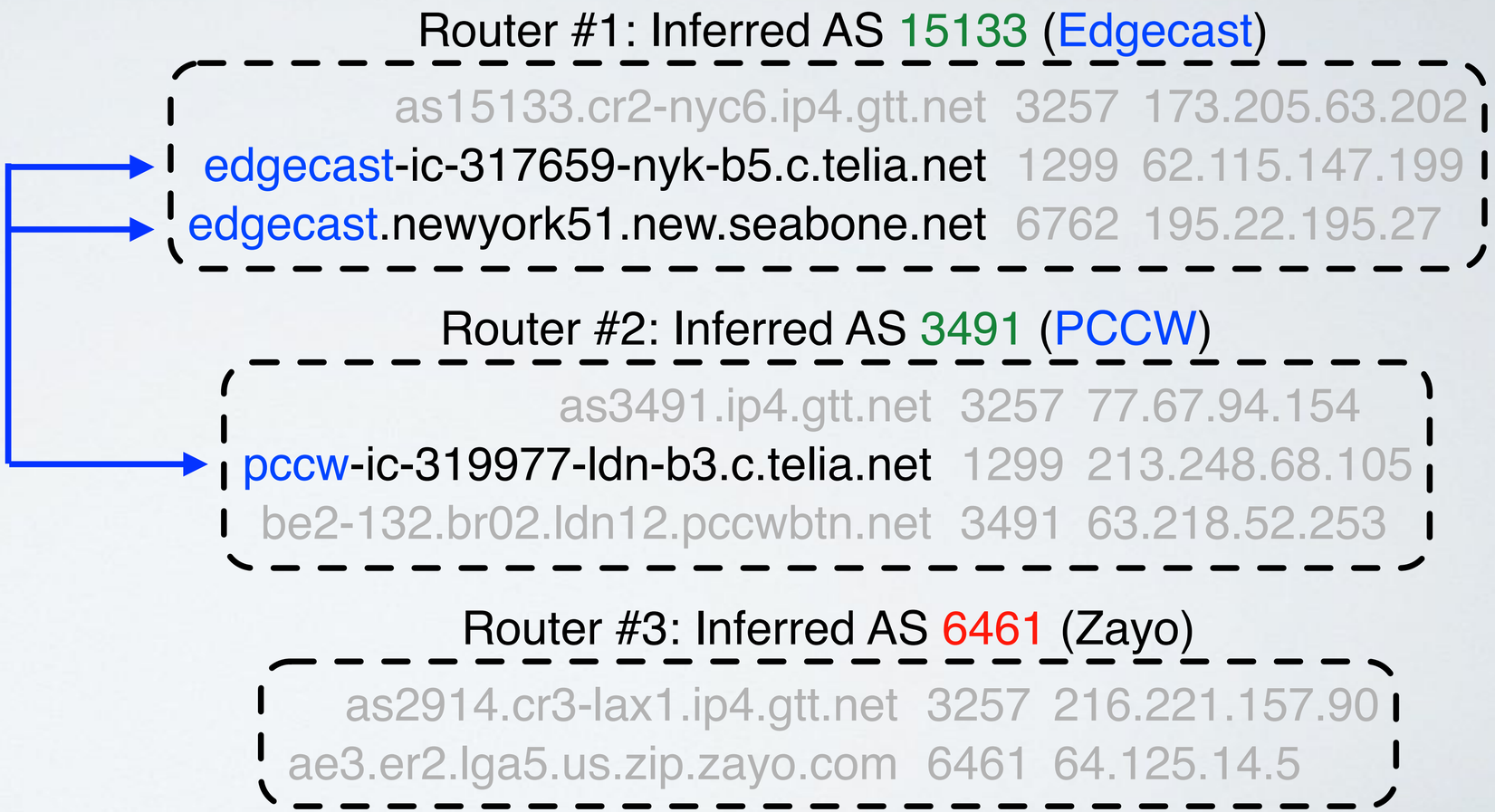
Some operators embed information in hostnames because it helps them debug their networks



Our approach: use information in hostnames

Some operators embed the **name** of the neighbor network using the address.

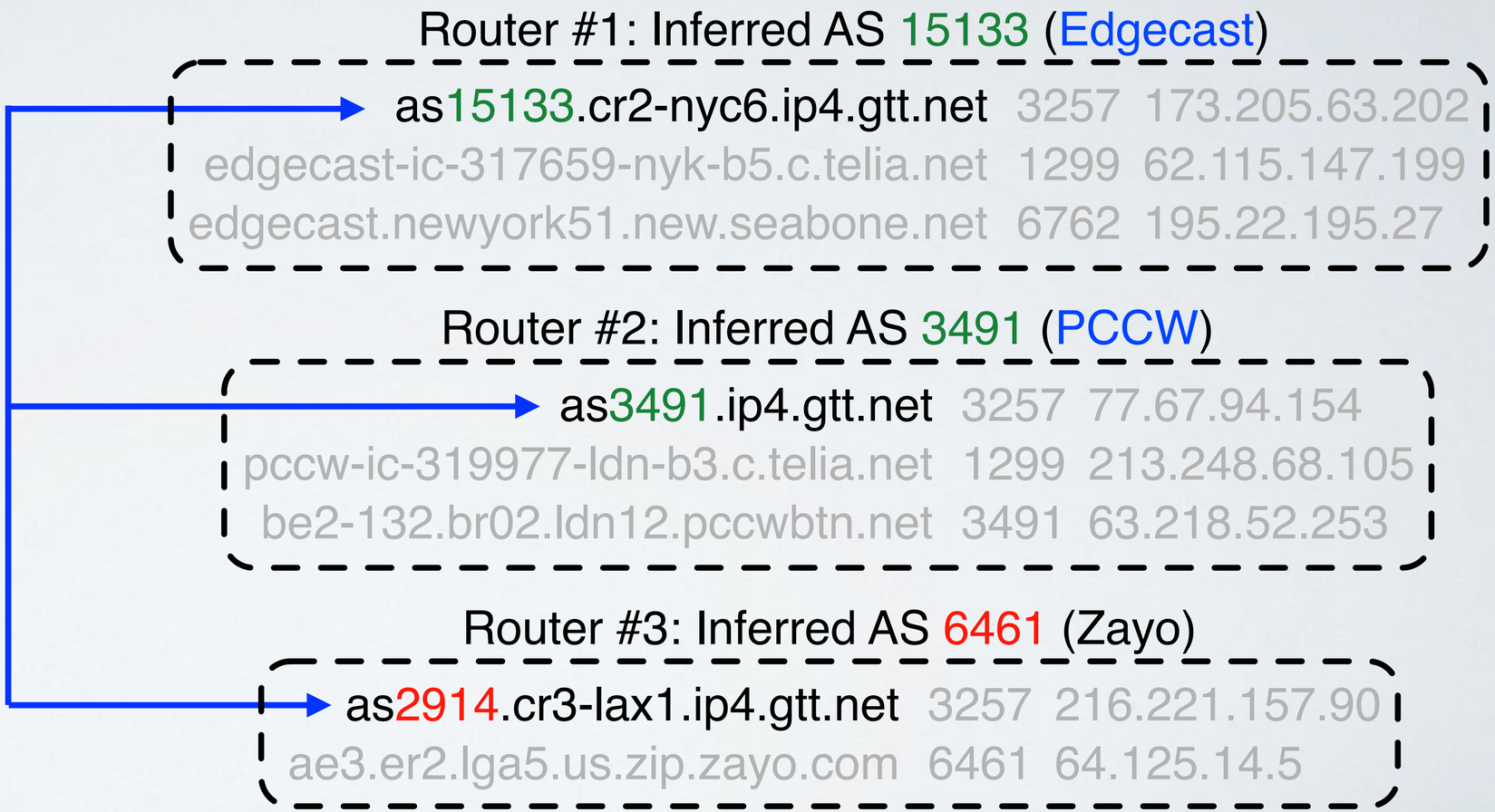
e.g. telia.net, seabone.net



Our approach: use information in hostnames

Some operators embed the **ASN** of the neighbor network using the address.

e.g. gtt.net



Our approach: use information in hostnames

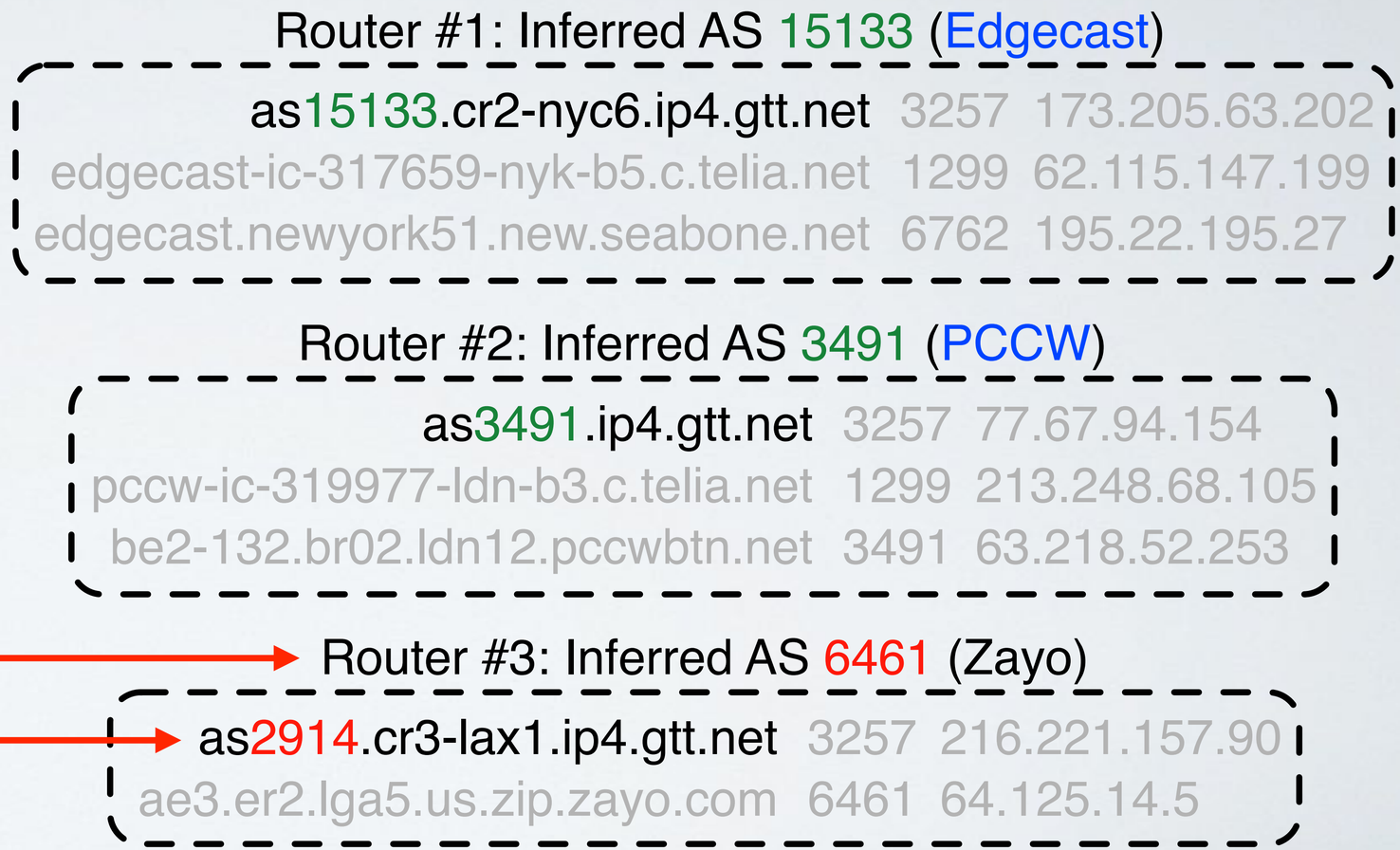
How DNS Misnaming Distorts Internet Topology Mapping

USENIX ATC 2006

Issue:

Is the inferred AS incorrect?

Or is the hostname stale?



Our approach: use information in hostnames

^as(\d+)\.+\.gtt\.net$

- **use heuristic inferences to label a training set**
- **automatically learn regexes extracting ASNs**
- **automatically vet extracted ASNs different to inferred ASNs**
- **use extracted ASNs to improve heuristic inferences**

Router #1: Inferred AS 15133 (Edgecast)

as15133.cr2-nyc6.ip4.gtt.net 3257 173.205.63.202
edgecast-ic-317659-nyk-b5.c.telia.net 1299 62.115.147.199
edgecast.newyork51.new.seabone.net 6762 195.22.195.27

Router #2: Inferred AS 3491 (PCCW)

as3491.ip4.gtt.net 3257 77.67.94.154
pccw-ic-319977-ldn-b3.c.telia.net 1299 213.248.68.105
be2-132.br02.ldn12.pccwbtn.net 3491 63.218.52.253

Router #3: Inferred AS 6461 (Zayo)

as2914.cr3-lax1.ip4.gtt.net 3257 216.221.157.90
ae3.er2.lga5.us.zip.zayo.com 6461 64.125.14.5

Related Work: undns

SIGCOMM 2003

^be-\d+\.(cor|bdr)\d+\.([a-z]{3})\d+\.[a-z]{2,3}\.vocus\.net\.au$

be-102.cor01.per02.wa.vocus.net.au
be-103.cor01.per02.wa.vocus.net.au

be-102.cor02.mel07.vic.vocus.net.au
be-151.cor02.mel07.vic.vocus.net.au

be-100.bdr01.syd03.nsw.vocus.net.au
be-101.bdr01.syd03.nsw.vocus.net.au

be-100.bdr02.syd03.nsw.vocus.net.au
be-101.bdr02.syd03.nsw.vocus.net.au

as38883.cust.bdr02.syd03.nsw.vocus.net.au
te-0-0-0-bdr1-syd-eqx.firenet.net.au

```
type=1 {
  cor "backbone"
  bdr "gateway"
}
loc=2 {
  per "Perth, Australia"
  mel "Melbourne, Australia"
  syd "Sydney, Australia"
}
```

Hand-crafted regexes built by manually interpreting hostnames.

Hand-crafted rules to interpret extracted output.

Related Work: DRoP

CCR 2014

([a-z]{3})[^a-z]+[a-z]+[0-9]*\.vocus\.net\.au$

be-102.cor01.per02.wa.vocus.net.au
be-103.cor01.per02.wa.vocus.net.au

be-102.cor02.mel07.vic.vocus.net.au
be-151.cor02.mel07.vic.vocus.net.au

be-100.bdr01.syd03.nsw.vocus.net.au
be-101.bdr01.syd03.nsw.vocus.net.au

be-100.bdr02.syd03.nsw.vocus.net.au
be-101.bdr02.syd03.nsw.vocus.net.au

as38883.cust.bdr02.syd03.nsw.vocus.net.au
te-0-0-0-bdr1-syd-eqx.firenet.net.au

- DRoP automatically infers regexes that extract apparent **location identifiers** from hostnames for routers
- Trained using measurements of RTT and topological distance between known landmarks and routers

Related Work: Hoiho

^be-\d+\.([a-z]+\d+\.[a-z]+\d+\.[a-z]+)\.vocus\.net\.au$

be-102.cor01.per02.wa.vocus.net.au
be-103.cor01.per02.wa.vocus.net.au

be-102.cor02.mel07.vic.vocus.net.au
be-151.cor02.mel07.vic.vocus.net.au

be-100.bdr01.syd03.nsw.vocus.net.au
be-101.bdr01.syd03.nsw.vocus.net.au

be-100.bdr02.syd03.nsw.vocus.net.au
be-101.bdr02.syd03.nsw.vocus.net.au

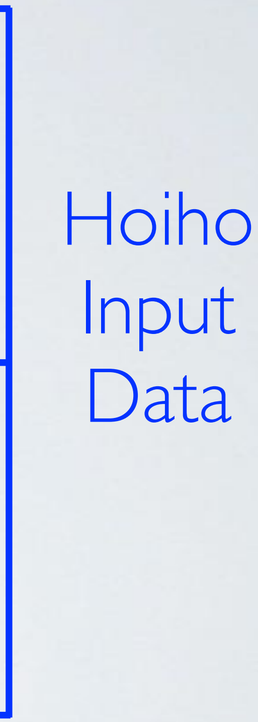
as38883.cust.bdr02.syd03.nsw.vocus.net.au
te-0-0-0-bdr1-syd-eqx.firenet.net.au

- Hoiho automatically infers regexes that extract apparent **router names** from hostnames for routers
- Trained using router alias inferences from MIDAR and Mercator

CAIDA Internet Topology Data Kit (ITDK)

Heavily curated router-level topology dataset published roughly twice a year

- IPv4 Routers, with aliases inferred by MIDAR and Mercator
- Links between routers
- Router geolocation
- Router ownership
  - RouterToAsAssignment (July 2010 - February 2017)
  - bdrmapIT (August 2017 - January 2020)
- DNS hostnames
- 17 ITDK datasets between July 2010 to January 2020



Approach by example

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

Goal: learn regex to extract ASNs from these hostnames

Green labels: training and embedded ASN congruent **(TP)**

Red labels: training and embedded ASN incongruent **(FP)**

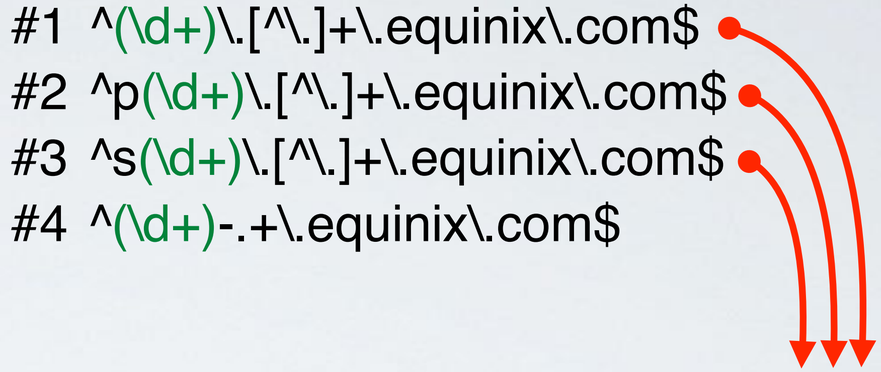
Approach by example: build base regexes

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

		True Positives
#1	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
#2	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
#3	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
#4	^(\d+)-.+\.equinix\.com$	h, i, j, k

Approach by example: merge regexes

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

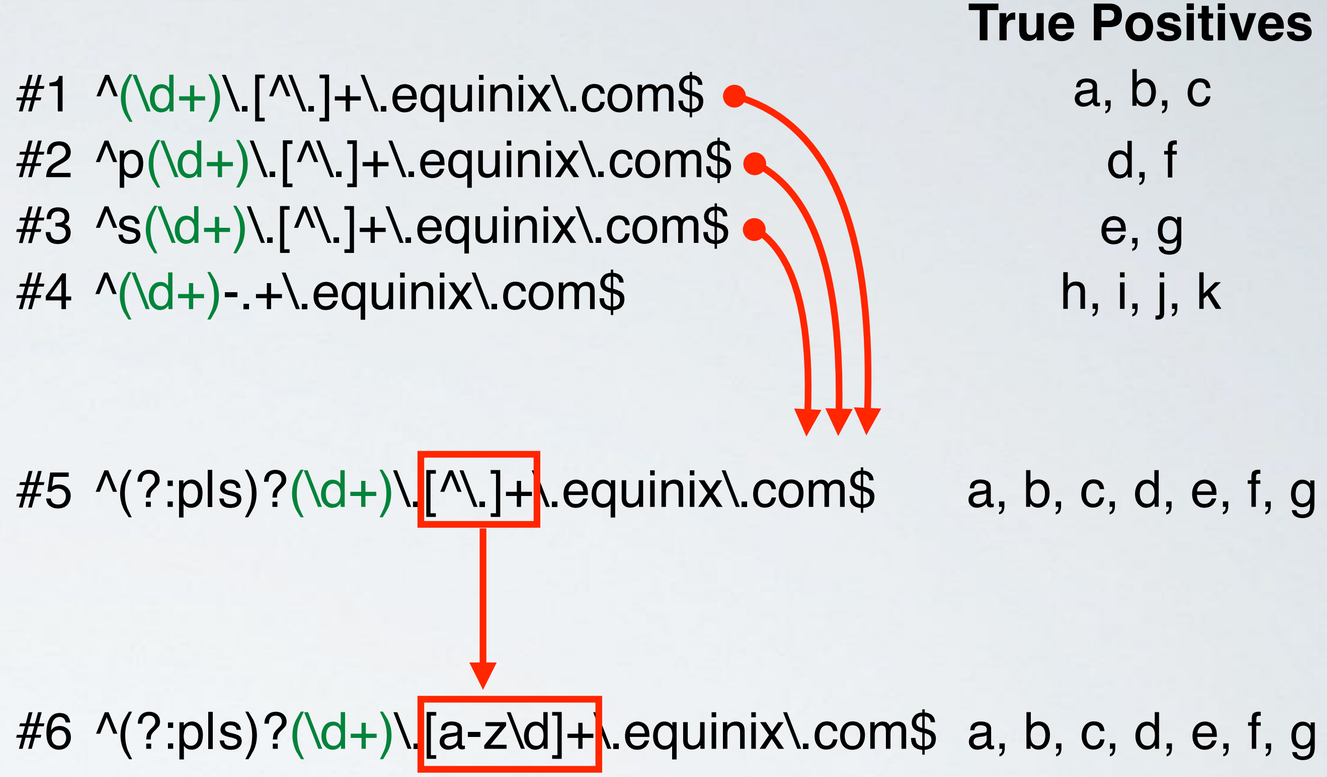


#5 ^(?:pls)?(\d+)\.[^\.]+\.equinix\.com$

True Positives
a, b, c
d, f
e, g
h, i, j, k
a, b, c, d, e, f, g

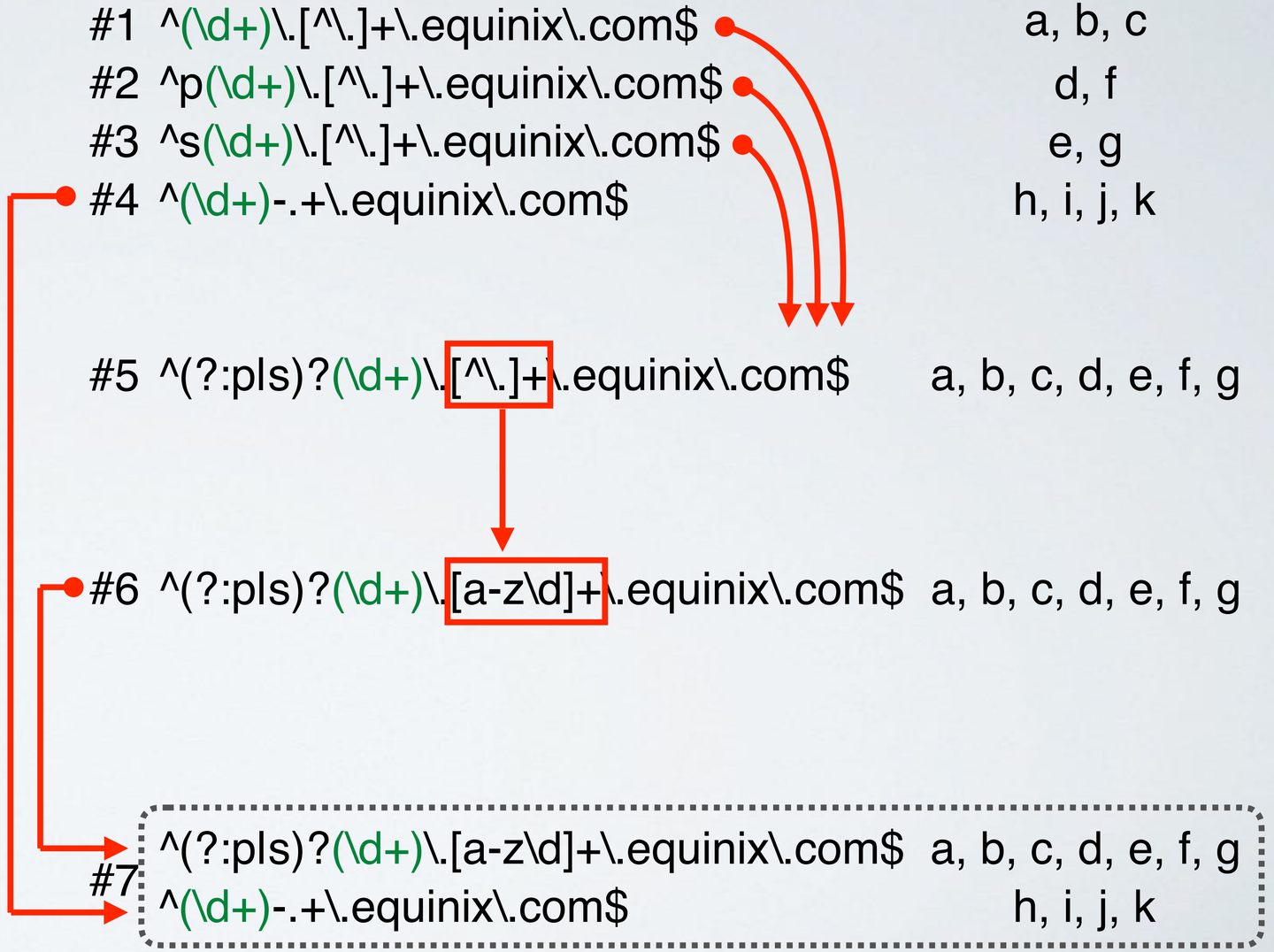
Approach by example: embed character classes

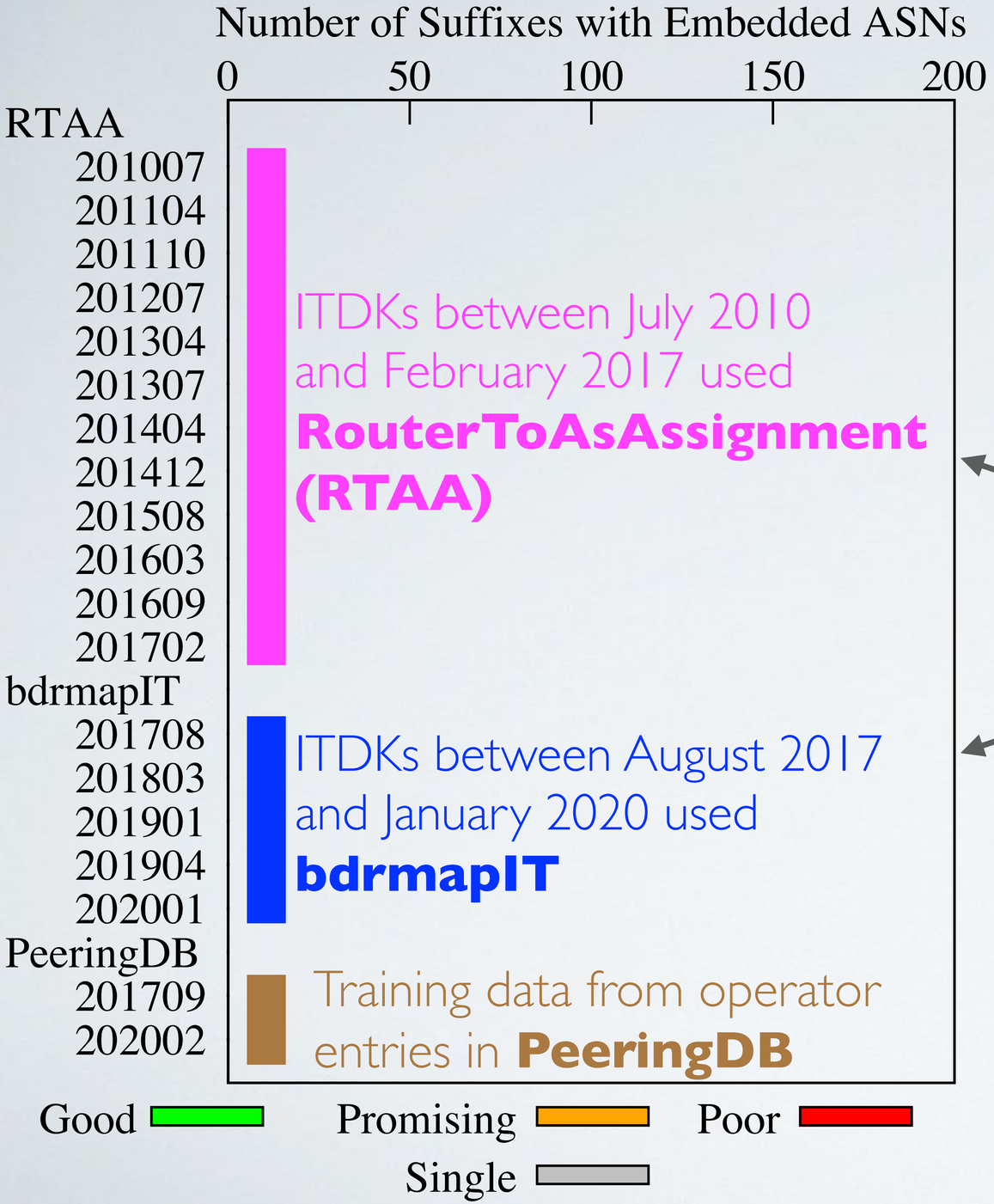
training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)



Approach by example: build sets

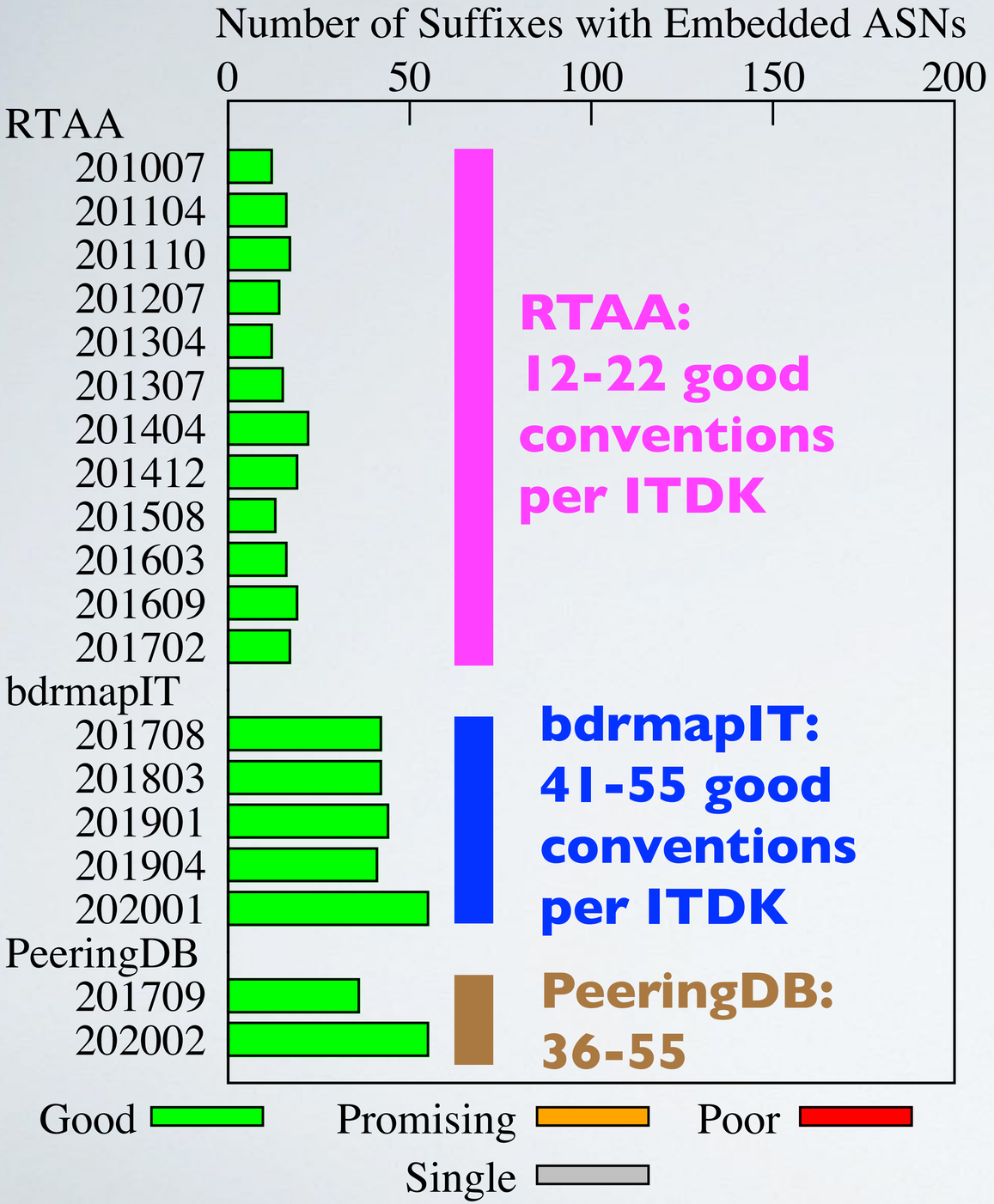
training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)





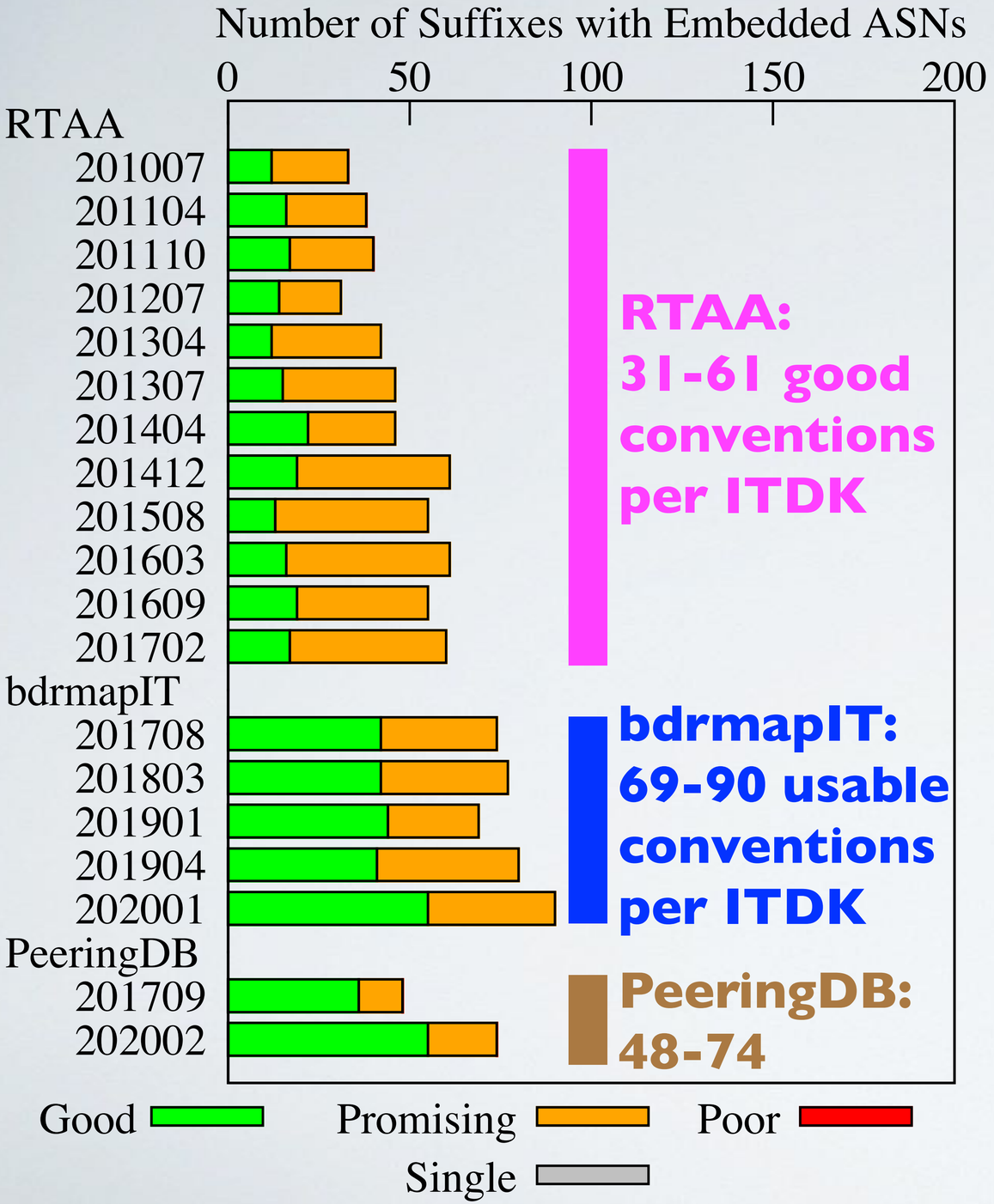
Heuristic-based Router Ownership Inference Methods

Heuristic Method Progress



- **Good conventions:** PPV > 80%,
  >= 3 uniq ASNs congruent w/ training data

3x more conventions using bdrmapIT than RTAA

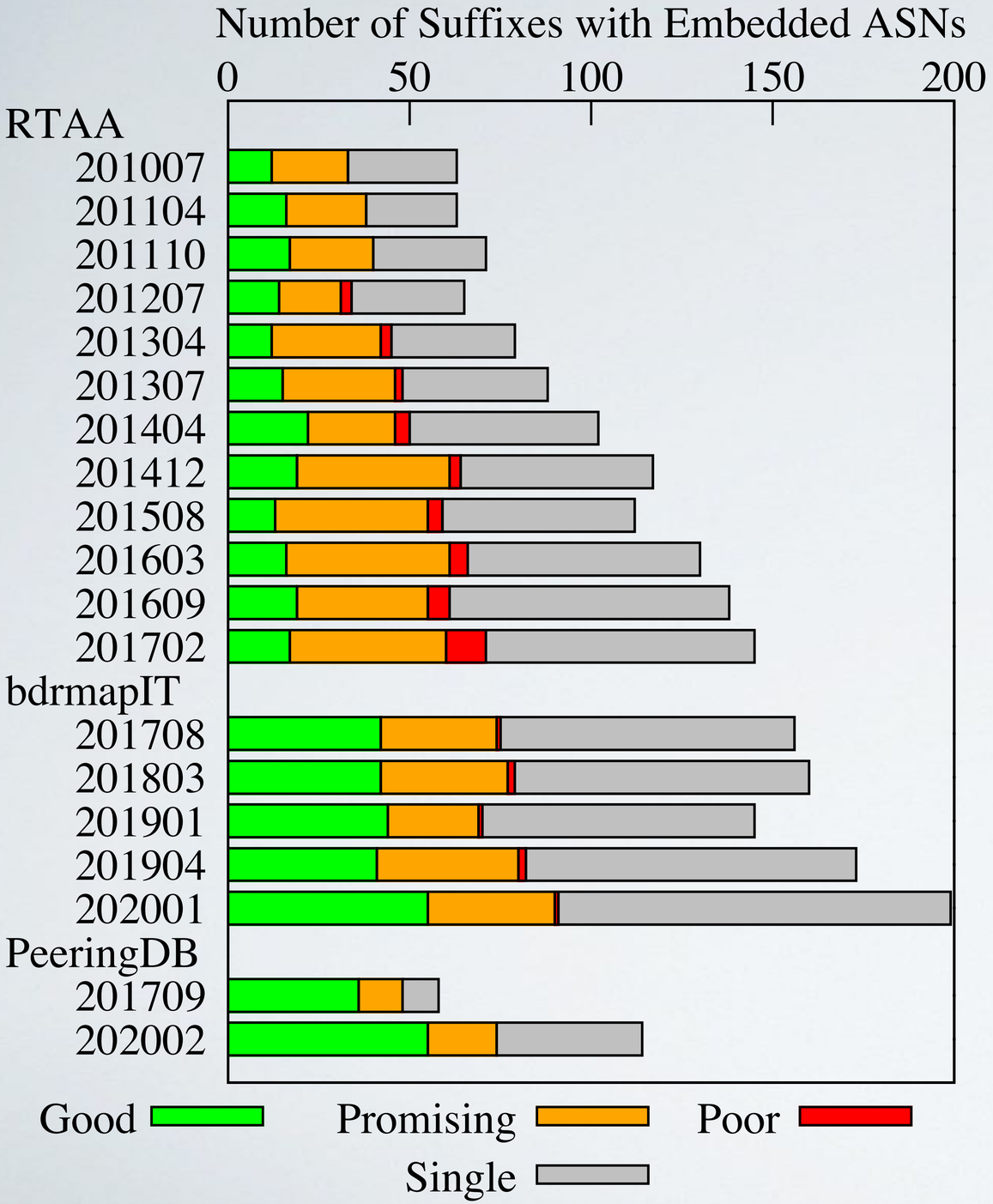


Heuristic Method Progress

- **Promising conventions:** PPV > 50%, >=2 uniq ASNs congruent w/ training data
- **Good** and **Promising** are ***Usable***

We inferred "usable" conventions for 206 suffixes in total.

The number of suffixes we detected embedding an ASN increased over time.



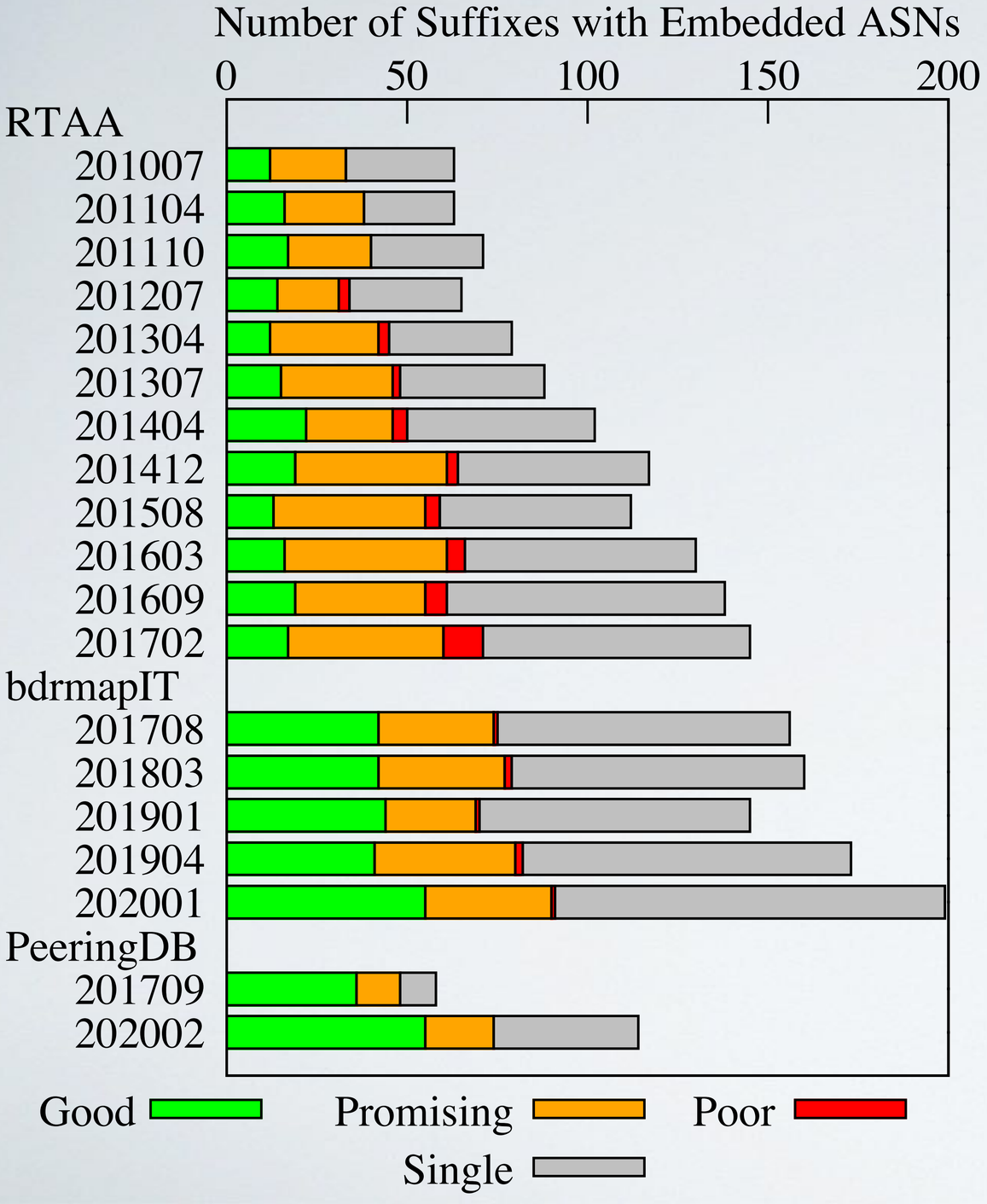
Naming Convention Challenge

- **Single conventions:** organization only embeds their own ASN, even for addresses assigned to a neighbor router

training ASN	hostname (PTR record)	
15576	ge0-2.01.p.ost.ch.as15576.nts.ch	(a)
15576	lo1000.01.lns.czh.ch.as15576.nts.ch	(b)
15576	te0-0-24.01.p.bre.ch.as15576.nts.ch	(c)
44879	01.r.cba.ch.bl.cust.as15576.nts.ch	(d)
51768	02.r.czh.ch.sda.cust.as15576.nts.ch	(e)
206616	01.r.cbs.ch.wwc.cust.as15576.nts.ch	(f)

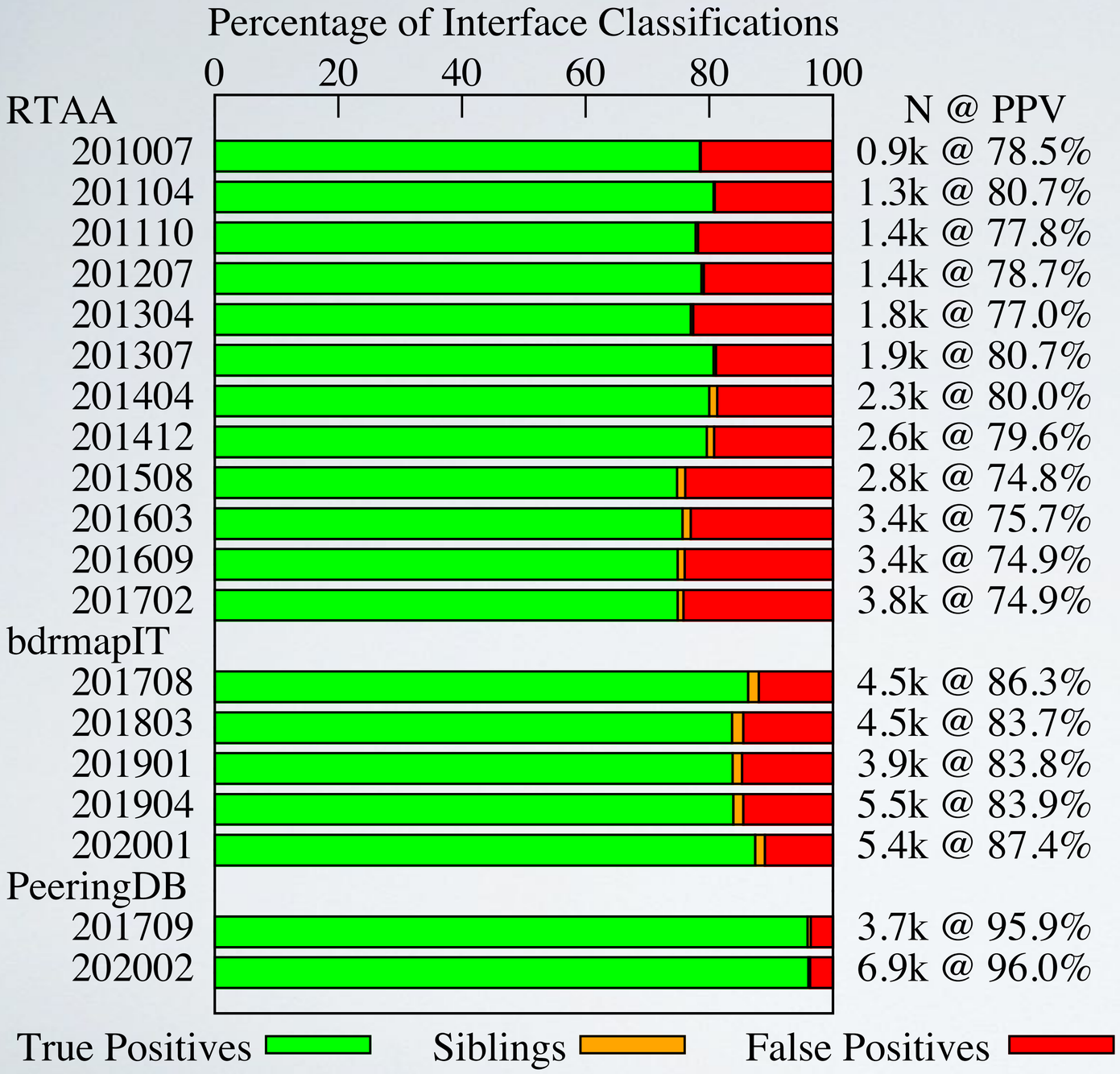
as(\d+)\.nts\.ch$

Validation



We validated 4 conventions with 4 network operators, who confirmed we captured their convention.

Hostname Congruity

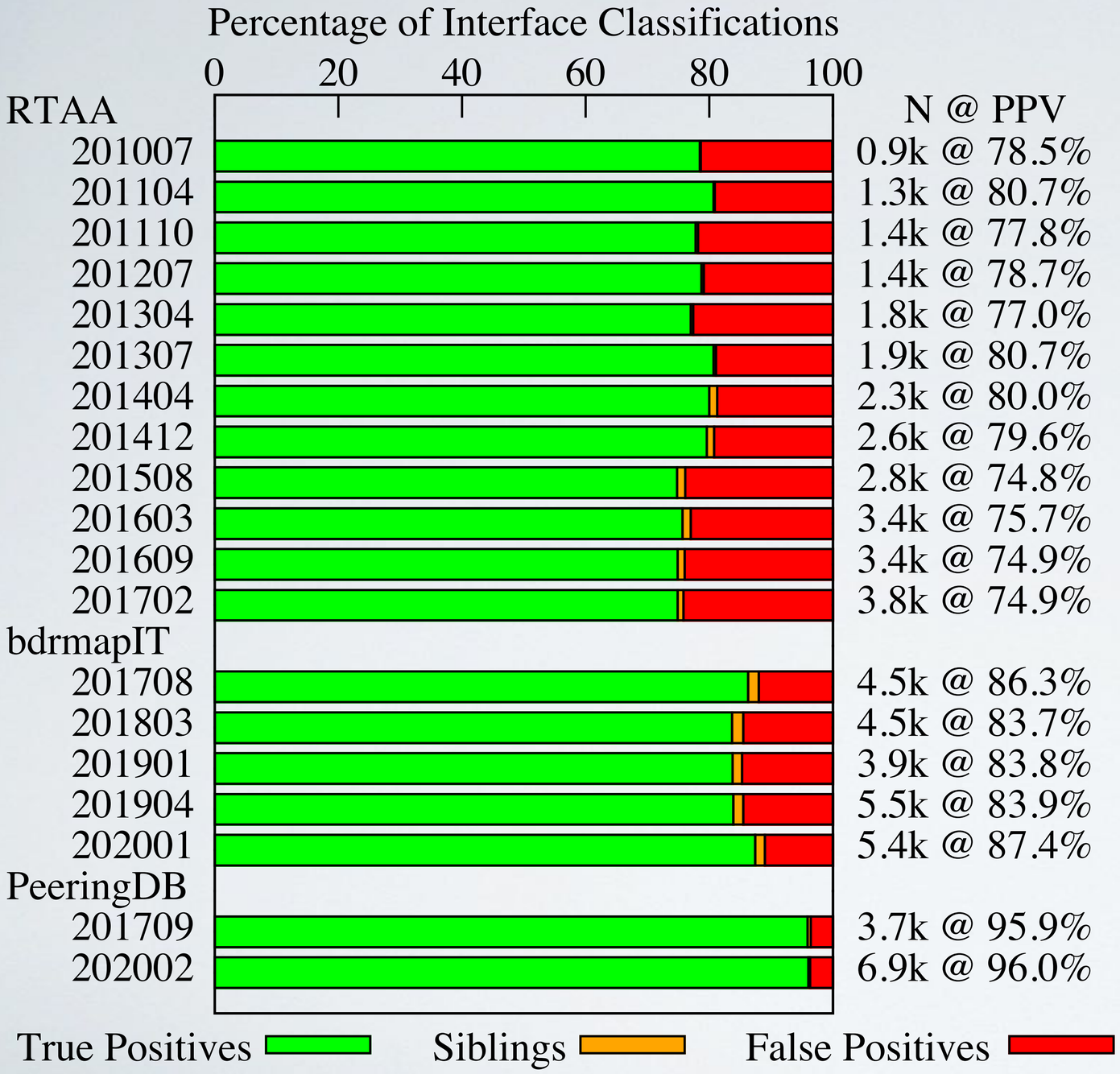


- **RouterToAsAssignment**
  PPV 74.8% — 80.7%

- **bdrmapIT**
  PPV 83.7% — 87.4%

- **PeeringDB: PPV 96.0%**

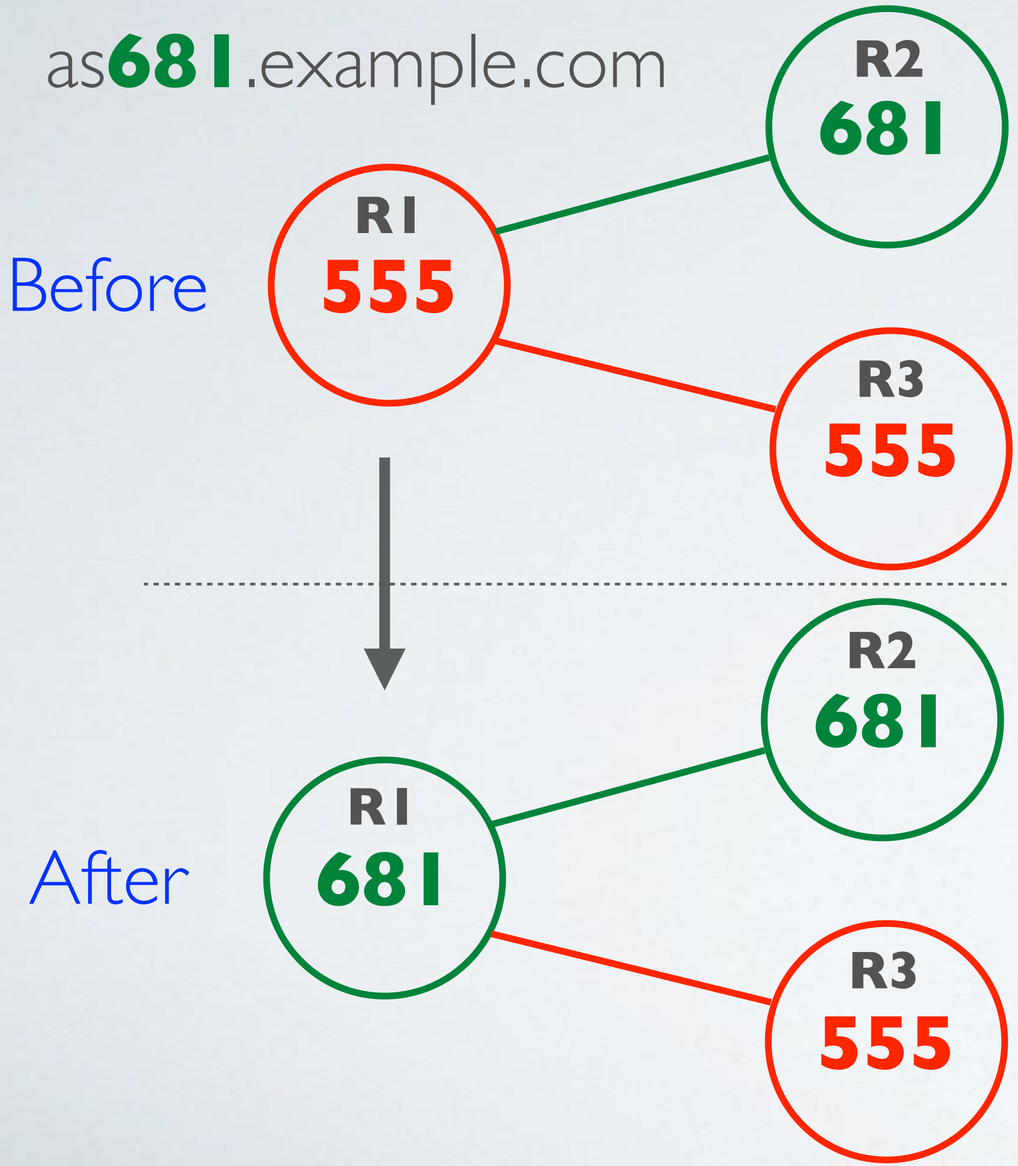
Hostname Congruity



- **RouterToAsAssignment**
  PPV 74.8% — 80.7%
- **bdrmapIT**
  PPV 83.7% — 87.4%
- **PeeringDB: PPV 96.0%**

We hypothesized that while some incongruent hostnames were stale, more were correct, and heuristic methods inferred the wrong ASN

Automatically vetting ASNs in hostnames



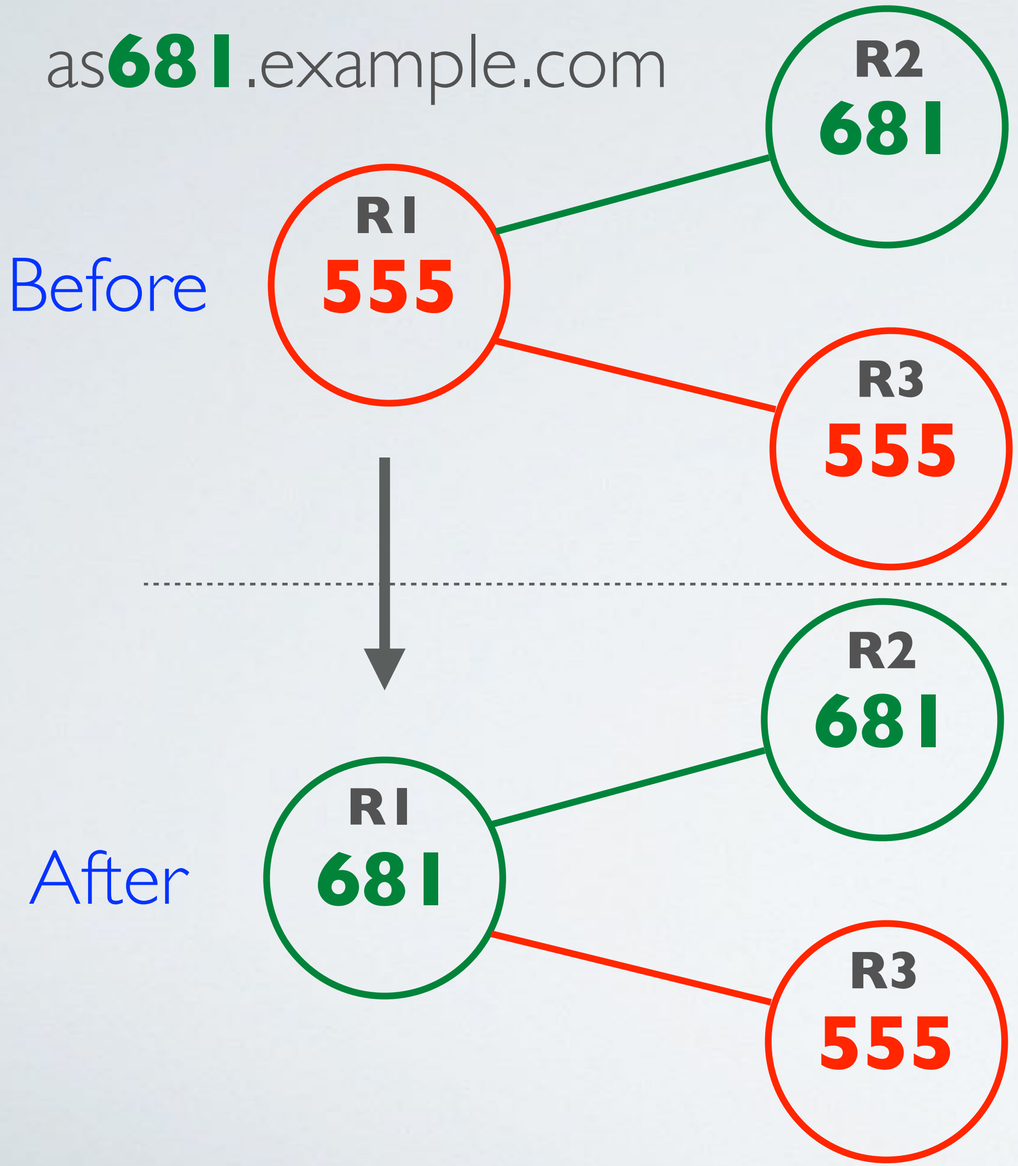
- Accept hint if ASN in hostname is ASN of

  (a) a subsequent router

  (b) originating a destination prefix

  (c) a provider of (a) or (b)

Automatically vetting ASNs in hostnames



- Accept hint if ASN in hostname is ASN of

  (a) a subsequent router

  (b) originating a destination prefix

  (c) a provider of (a) or (b)

- This heuristic increased congruity between inferences and hostnames in Jan 2020 ITDK from 87.4% to 97.1%

Validation: distinguished stale from correct hostnames for 92.5% of hostnames

Evidence-Based Router Ownership Inference

Suffixes:

	PeeringDB	bdrmapIT	bdrmapIT
76telecom.com.br		202001	201904
absolutok.com		202001	201904
accessdigital.com.au			201904
acesso10.net.br			201904
actix.net.au			
akl-ix.nz	202002	202001	201904

Evaluation against training data:

```
                10835
69.174.23.114 + as10835-gw.cr0-phx1.ip4.gtt.net

                10835
69.174.23.110 + as10835.cr3-sea2.ip4.gtt.net
144.224.113.150   sl-gigli-1022680-0.sprintlink.net
 209.193.81.117

                1215
69.22.153.234 + as1215.xe-7-0-6.ar2.sjc1.us.as4436.gtt.net
 148.87.37.227    sjc-ext-border-rtr-2-v100.oracle.com
```

- We publicly release the source code implementation
  - https://www.caida.org/tools/measurement/scamper/
- We publicly release inferred regexes, as well as webpages demonstrating how each regex applied to the training data
  - https://www.caida.org/publications/papers/2020/hoiho/
- We publicly release our extensions to bdrmapIT to vet ASN annotations and integrate them into bdrmapIT inferences
  - https://github.com/alexmarder/bdrmapit

Acknowledgments

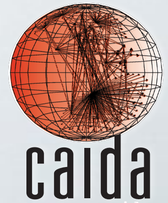
- We thank Young Hyun and Ken Keys for assistance with the ITDK, and the anonymous reviewers for their helpful comments.

- This work was supported by NSF OAC-1724853, and by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD via contract number 70RSAT18CB0000013, but this paper represents only the position of the authors.

Learning to Extract and Use ASNs in Hostnames

Matthew Luckie - University of Waikato
Alexander Marder - CAIDA / UC San Diego
Marianne Fletcher - University of Waikato
Bradley Huffaker - CAIDA / UC San Diego
k claffy - CAIDA / UC San Diego

IMC 2020



BACKUP SLIDES

Towards Evidence-Based Router Ownership Inference

- Router Ownership Inference is critical to a growing body of work
- Heuristic methods often involve judgement calls by their designers
  - Validation is difficult: prior work
    - up to 7 networks
    - mostly Tier-1 and R&E networks.
  - Difficult to share validation data with other researchers
- This work unlocks a new source of validation data
  - For Jan 2020 ITDK, 90 networks, of different types and locations

High-level Approach

- Infer if an operator embeds information **identifying the ASN that operates the router** in PTR hostname records for router interfaces
- **Input**:
  - Mozilla public suffix list to identify where domains can be registered (.net, .org, .nz, .co.nz, .geek.nz)
  - Hostnames for router interfaces observed by traceroute (PTR records)
  - Router alias inferences from MIDAR, mercator
  - Router ownership inferences from bdrmapIT, RouterToAsAssignment
- **Output**: regular expressions that extract router operator ASN

Approach by example

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

Goal: learn regex to extract ASNs from these hostnames

Green labels: training and embedded ASN congruent **(TP)**

Red labels: training and embedded ASN incongruent **(FP)**

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)		
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$
714	714.os.equinix.com	(b)	^(\d+)\..+\.equinix\.com$
714	714.me1.equinix.com	(c)	
714	p714.sgw.equinix.com	(d)	
714	s714.sgw.equinix.com	(e)	
24115	p24115.mel.equinix.com	(f)	
24115	s24115.tyo.equinix.com	(g)	
22282	22822-2.tyo.equinix.com	(h)	
24482	24482-fr5-ix.equinix.com	(i)	
54827	54827-dc5-ix2.equinix.com	(j)	
55247	55247-ch3-ix.equinix.com	(k)	
2906	netflix.zh2.corp.eu.equinix.com	(l)	
19324	ipv4.dosarrest.eqix.equinix.com	(m)	
8075	8069.tyo.equinix.com	(n)	
8075	8074.hkg.equinix.com	(o)	
55923	45437-sy1-ix.equinix.com	(p)	

Build base regexes using components that exclude specified punctuation. Include literals surrounding ASN.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)		
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$
714	714.os.equinix.com	(b)	^(\d+)\..+\.equinix\.com$
714	714.me1.equinix.com	(c)	
714	p714.sgw.equinix.com	(d)	^p(\d+)\.[^\.]+\.equinix\.com$
714	s714.sgw.equinix.com	(e)	^p(\d+)\..+\.equinix\.com$
24115	p24115.mel.equinix.com	(f)	
24115	s24115.tyo.equinix.com	(g)	
22282	22822-2.tyo.equinix.com	(h)	
24482	24482-fr5-ix.equinix.com	(i)	
54827	54827-dc5-ix2.equinix.com	(j)	
55247	55247-ch3-ix.equinix.com	(k)	
2906	netflix.zh2.corp.eu.equinix.com	(l)	
19324	ipv4.dosarrest.eqix.equinix.com	(m)	
8075	8069.tyo.equinix.com	(n)	
8075	8074.hkg.equinix.com	(o)	
55923	45437-sy1-ix.equinix.com	(p)	

Build base regexes using components that exclude specified punctuation. Include literals surrounding ASN.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)		
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$
714	714.os.equinix.com	(b)	^(\d+)\..+\.equinix\.com$
714	714.me1.equinix.com	(c)	^p(\d+)\.[^\.]+\.equinix\.com$
714	p714.sgw.equinix.com	(d)	^p(\d+)\..+\.equinix\.com$
714	s714.sgw.equinix.com	(e)	^s(\d+)\.[^\.]+\.equinix\.com$
24115	p24115.mel.equinix.com	(f)	^s(\d+)\..+\.equinix\.com$
24115	s24115.tyo.equinix.com	(g)	
22282	22822-2.tyo.equinix.com	(h)	
24482	24482-fr5-ix.equinix.com	(i)	
54827	54827-dc5-ix2.equinix.com	(j)	
55247	55247-ch3-ix.equinix.com	(k)	
2906	netflix.zh2.corp.eu.equinix.com	(l)	
19324	ipv4.dosarrest.eqix.equinix.com	(m)	
8075	8069.tyo.equinix.com	(n)	
8075	8074.hkg.equinix.com	(o)	
55923	45437-sy1-ix.equinix.com	(p)	

Build base regexes using components that exclude specified punctuation. Include literals surrounding ASN.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)		
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$
714	714.os.equinix.com	(b)	^(\d+)\..+\.equinix\.com$
714	714.me1.equinix.com	(c)	^p(\d+)\.[^\.]+\.equinix\.com$
714	p714.sgw.equinix.com	(d)	^p(\d+)\..+\.equinix\.com$
714	s714.sgw.equinix.com	(e)	^s(\d+)\.[^\.]+\.equinix\.com$
24115	p24115.mel.equinix.com	(f)	^s(\d+)\..+\.equinix\.com$
24115	s24115.tyo.equinix.com	(g)	
22282	22822-2.tyo.equinix.com	(h)	^(\d+)-.+\.equinix\.com$
24482	24482-fr5-ix.equinix.com	(i)	^(\d+)-[^-]+\.equinix\.com$
54827	54827-dc5-ix2.equinix.com	(j)	^(\d+)-[^\.]+\.equinix\.com$
55247	55247-ch3-ix.equinix.com	(k)	
2906	netflix.zh2.corp.eu.equinix.com	(l)	
19324	ipv4.dosarrest.eqix.equinix.com	(m)	
8075	8069.tyo.equinix.com	(n)	
8075	8074.hkg.equinix.com	(o)	
55923	45437-sy1-ix.equinix.com	(p)	

Build base regexes using components that exclude specified punctuation. Include literals surrounding ASN.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)		
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$
714	714.os.equinix.com	(b)	^(\d+)\..+\.equinix\.com$
714	714.me1.equinix.com	(c)	^p(\d+)\.[^\.]+\.equinix\.com$
714	p714.sgw.equinix.com	(d)	^p(\d+)\..+\.equinix\.com$
714	s714.sgw.equinix.com	(e)	^s(\d+)\.[^\.]+\.equinix\.com$
24115	p24115.mel.equinix.com	(f)	^s(\d+)\..+\.equinix\.com$
24115	s24115.tyo.equinix.com	(g)	^(\d+)-.+\.equinix\.com$
22282	22822-2.tyo.equinix.com	(h)	^(\d+)-[^-]+\.equinix\.com$
24482	24482-fr5-ix.equinix.com	(i)	^(\d+)-[^\.]+\.equinix\.com$
54827	54827-dc5-ix2.equinix.com	(j)	
55247	55247-ch3-ix.equinix.com	(k)	^(\d+)-[^-]+-[^\.]+\.equinix\.com$
2906	netflix.zh2.corp.eu.equinix.com	(l)	^(\d+)-[^-]+-[^-]+\.equinix\.com$
19324	ipv4.dosarrest.eqix.equinix.com	(m)	
8075	8069.tyo.equinix.com	(n)	
8075	8074.hkg.equinix.com	(o)	
55923	45437-sy1-ix.equinix.com	(p)	

Build base regexes using components that exclude specified punctuation. Include literals surrounding ASN.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

```
^(\d+)\.[^\.]+\.equinix\.com$
^(\d+)\..+\.equinix\.com$
^p(\d+)\.[^\.]+\.equinix\.com$
^p(\d+)\..+\.equinix\.com$
^s(\d+)\.[^\.]+\.equinix\.com$
^s(\d+)\..+\.equinix\.com$
^(\d+)-.+\.equinix\.com$
^(\d+)-[^-]+\.equinix\.com$
^(\d+)-[^\.]+\.equinix\.com$
^(\d+)-[^-]+-[^\.]+\.equinix\.com$
^(\d+)-[^-]+-[^-]+\.equinix\.com$
```

Evaluate regexes according to congruity with training data.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)			True Positives
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
714	714.os.equinix.com	(b)	^(\d+)\..+\.equinix\.com$	a, b, c
714	714.me1.equinix.com	(c)	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
714	p714.sgw.equinix.com	(d)	^p(\d+)\..+\.equinix\.com$	d, f
714	s714.sgw.equinix.com	(e)	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
24115	p24115.mel.equinix.com	(f)	^s(\d+)\..+\.equinix\.com$	e, g
24115	s24115.tyo.equinix.com	(g)	^(\d+)-.+\.equinix\.com$	h, i, j, k
22282	22822-2.tyo.equinix.com	(h)	^(\d+)-[^-]+\.equinix\.com$	h
24482	24482-fr5-ix.equinix.com	(i)	^(\d+)-[^\.]+\.equinix\.com$	i, j, k
54827	54827-dc5-ix2.equinix.com	(j)	^(\d+)-[^-]+-[^\.]+\.equinix\.com$	h, i, j, k
55247	55247-ch3-ix.equinix.com	(k)	^(\d+)-[^-]+-[^-]+\.equinix\.com$	h, i, j, k
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

Evaluate regexes according to congruity with training data.

Stage 1: Generate Base Regexes

training ASN	hostname (PTR record)			True Positives
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
714	714.os.equinix.com	(b)	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
714	714.me1.equinix.com	(c)	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
714	p714.sgw.equinix.com	(d)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	s714.sgw.equinix.com	(e)		
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

Stage 2: Merge Regexes

training ASN	hostname (PTR record)		regex	True Positives
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
714	714.os.equinix.com	(b)	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
714	714.me1.equinix.com	(c)	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
714	p714.sgw.equinix.com	(d)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	s714.sgw.equinix.com	(e)		
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

Merge regexes that differ by a single simple string.

Stage 2: Merge Regexes

training ASN	hostname (PTR record)			True Positives
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
714	714.os.equinix.com	(b)	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
714	714.me1.equinix.com	(c)	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
714	p714.sgw.equinix.com	(d)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	s714.sgw.equinix.com	(e)		
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)	^(?:pls)?(\d+)\.[^\.]+\.equinix\.com$	a, b, c, d, e, f, g
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

Merge regexes that differ by a single simple string.

Stage 2: Merge Regexes

training ASN	hostname (PTR record)			True Positives
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
714	714.os.equinix.com	(b)	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
714	714.me1.equinix.com	(c)	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
714	p714.sgw.equinix.com	(d)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	s714.sgw.equinix.com	(e)	^(?:pls)?(\d+)\.[^\.]+\.equinix\.com$	a, b, c, d, e, f, g
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)	^(?:pls)(\d+)\.[^\.]+\.equinix\.com$	d, e, f, g
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

Merge regexes that differ by a single simple string.

Stage 2: Merge Regexes

training ASN	hostname (PTR record)			True Positives
109	109.sgw.equinix.com	(a)	^(\d+)\.[^\.]+\.equinix\.com$	a, b, c
714	714.os.equinix.com	(b)	^p(\d+)\.[^\.]+\.equinix\.com$	d, f
714	714.me1.equinix.com	(c)	^s(\d+)\.[^\.]+\.equinix\.com$	e, g
714	p714.sgw.equinix.com	(d)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	s714.sgw.equinix.com	(e)	^(?:pls)?(\d+)\.[^\.]+\.equinix\.com$	a, b, c, d, e, f, g
24115	p24115.mel.equinix.com	(f)	^(?:pls)(\d+)\.[^\.]+\.equinix\.com$	d, e, f, g
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

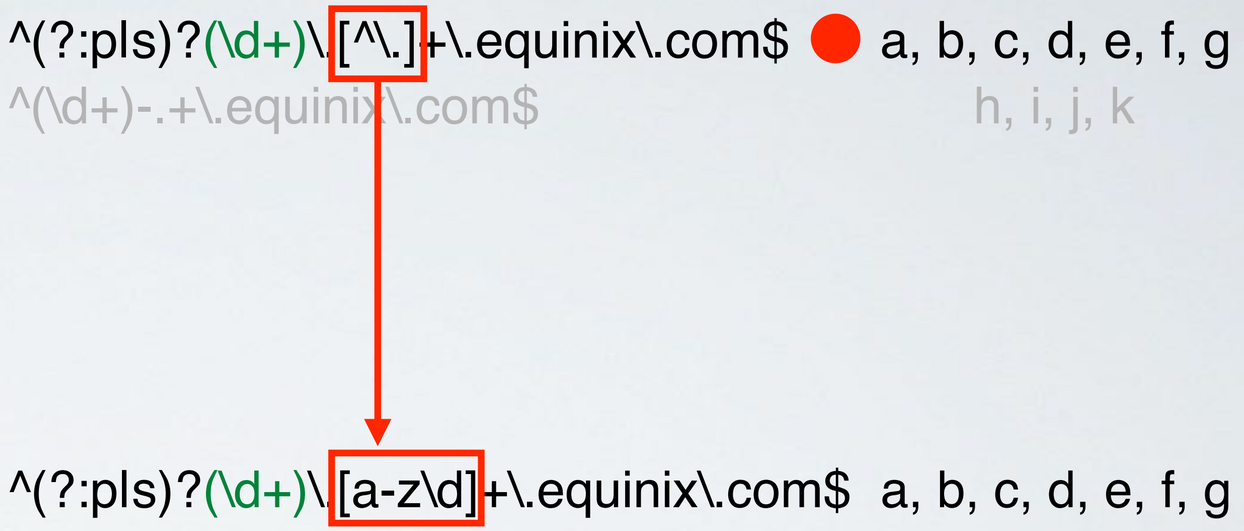
Stage 2: Merge Regexes

training ASN	hostname (PTR record)			
109	109.sgw.equinix.com	(a)	^(?:pls)?(\d+)\.[^\.]+\.equinix\.com$	a, b, c, d, e, f, g
714	714.os.equinix.com	(b)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	714.me1.equinix.com	(c)		
714	p714.sgw.equinix.com	(d)		
714	s714.sgw.equinix.com	(e)		
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

True Positives

Stage 3: Embed Character Classes

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)



Embed character class components to replace less specific components.

Stage 3: Embed Character Classes

training ASN	hostname (PTR record)			
				True Positives
109	109.sgw.equinix.com	(a)	^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
714	714.os.equinix.com	(b)	^(?:pls)?(\d+)\.[^\.]+\.equinix\.com$	a, b, c, d, e, f, g
714	714.me1.equinix.com	(c)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	p714.sgw.equinix.com	(d)		
714	s714.sgw.equinix.com	(e)		
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

Stage 3: Embed Character Classes

training ASN	hostname (PTR record)			
109	109.sgw.equinix.com	(a)	^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
714	714.os.equinix.com	(b)	^(\d+)-.+\.equinix\.com$	h, i, j, k
714	714.me1.equinix.com	(c)		
714	p714.sgw.equinix.com	(d)		
714	s714.sgw.equinix.com	(e)		
24115	p24115.mel.equinix.com	(f)		
24115	s24115.tyo.equinix.com	(g)		
22282	22822-2.tyo.equinix.com	(h)		
24482	24482-fr5-ix.equinix.com	(i)		
54827	54827-dc5-ix2.equinix.com	(j)		
55247	55247-ch3-ix.equinix.com	(k)		
2906	netflix.zh2.corp.eu.equinix.com	(l)		
19324	ipv4.dosarrest.eqix.equinix.com	(m)		
8075	8069.tyo.equinix.com	(n)		
8075	8074.hkg.equinix.com	(o)		
55923	45437-sy1-ix.equinix.com	(p)		

True Positives

Stage 4: Build Regex Sets

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

	True Positives
^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
^(\d+)-.+\.equinix\.com$	h, i, j, k

^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
^(\d+)-.+\.equinix\.com$	h, i, j, k

Build naming conventions made up of multiple regexes to capture diversity within a single suffix.

Stage 4: Build Regex Sets

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

	True Positives
^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
^(\d+)-.+\.equinix\.com$	h, i, j, k
^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
^(\d+)-.+\.equinix\.com$	h, i, j, k

Select Best Convention

training ASN	hostname (PTR record)	
109	109.sgw.equinix.com	(a)
714	714.os.equinix.com	(b)
714	714.me1.equinix.com	(c)
714	p714.sgw.equinix.com	(d)
714	s714.sgw.equinix.com	(e)
24115	p24115.mel.equinix.com	(f)
24115	s24115.tyo.equinix.com	(g)
22282	22822-2.tyo.equinix.com	(h)
24482	24482-fr5-ix.equinix.com	(i)
54827	54827-dc5-ix2.equinix.com	(j)
55247	55247-ch3-ix.equinix.com	(k)
2906	netflix.zh2.corp.eu.equinix.com	(l)
19324	ipv4.dosarrest.eqix.equinix.com	(m)
8075	8069.tyo.equinix.com	(n)
8075	8074.hkg.equinix.com	(o)
55923	45437-sy1-ix.equinix.com	(p)

	True Positives
^(?:pls)?(\d+)\.[a-z\d]+\.equinix\.com$	a, b, c, d, e, f, g
^(\d+)-.+\.equinix\.com$	h, i, j, k

7 unique ASNs

11 TPs

3 FPs

78.6% PPV

Taxonomy of Naming Conventions

		Usable	Single
Simple	^as(\d+)\.example\.com$	17.7%	4.6%
Start	^as(\d+)\.[a-z]+\.example\.com$	50.8%	23.1%
End	^[a-z\d]+\.as(\d+)\.example\.com$	10.8%	43.1%
Bare	(\d+)\.[a-z]+\d+\.example\.com$	5.4%	7.7%
Complex		15.4%	21.5%

Operators that labelled the neighbor ASN usually put the ASN at the start of the hostname.

Key Results - vetting incongruent hostnames

	Correct ASN		Incorrect ASN	
	Used (TP)	Not Used (FP)	Used (FP)	Not Used (TN)
Transit Provider	6	0	0	0
European ISP	4	0	0	0
Large ISP	4	0	2	36
Regional US ISP	6	0	0	1
Asia-Pacific IXP	3	1	0	1
PeeringDB (23)	322	26	6	49
Total:	345	27	8	87
	372		95	