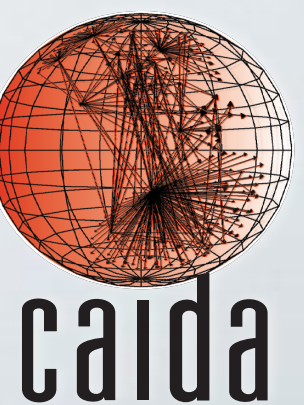


# Machine Learning Internet Router Hostname Patterns

Matthew Luckie - University of Waikato

OARC 35, May 2021



# Network Operators Embed Information in Router Hostnames

|                                                    |    |
|----------------------------------------------------|----|
| 100ge4-1.core3.fmt2.he.net                         | 1a |
| 100ge4-2.core3.fmt2.he.net                         | 1b |
| v1119.core3.fmt2.he.net                            | 1c |
| v1832.core3.fmt2.he.net                            | 1d |
| ge2-9.core1.atl1.he.net                            | 2a |
| ge6-7.core1.atl1.he.net                            | 2b |
| 10ge16-5.core1.ash1.he.net                         | 3a |
| 10ge16-6.core1.ash1.he.net                         | 3b |
| 100ge5-1.core1.ash1.he.net                         | 3c |
| esnet.10gigabitethernet5-15.core1.ash1.he.net      | 4a |
| fastserv.core1.ash1.he.net                         | 5a |
| dren-as668.10gigabitethernet16-4.core1.ash1.he.net | 6a |

It helps themselves  
and third parties  
contextualize Internet  
routing behavior

# Network Operators Embed Information in Router Hostnames

|                                  |            |        |    |
|----------------------------------|------------|--------|----|
| 100ge4-1                         | core3.fmt2 | he.net | 1a |
| 100ge4-2                         | core3.fmt2 | he.net | 1b |
| v1119                            | core3.fmt2 | he.net | 1c |
| v1832                            | core3.fmt2 | he.net | 1d |
| ge2-9                            | core1.atl1 | he.net | 2a |
| ge6-7                            | core1.atl1 | he.net | 2b |
| 10ge16-5                         | core1.ash1 | he.net | 3a |
| 10ge16-6                         | core1.ash1 | he.net | 3b |
| 100ge5-1                         | core1.ash1 | he.net | 3c |
| esnet.10gigabitethernet5-15      | core1.ash1 | he.net | 4a |
| fastserv                         | core1.ash1 | he.net | 5a |
| dren-as668.10gigabitethernet16-4 | core1.ash1 | he.net | 6a |

**Router names:**  
a string common to  
interfaces on the same  
router, and unique  
to each router.

core3.fmt2  
core2.atl1  
core1.ash1



# Network Operators Embed Information in Router Hostnames

|                                                    |    |
|----------------------------------------------------|----|
| 100ge4-1.core3.fmt2.he.net                         | 1a |
| 100ge4-2.core3.fmt2.he.net                         | 1b |
| v1119.core3.fmt2.he.net                            | 1c |
| v1832.core3.fmt2.he.net                            | 1d |
| ge2-9.core1.atl1.he.net                            | 2a |
| ge6-7.core1.atl1.he.net                            | 2b |
| 10ge16-5.core1.ash1.he.net                         | 3a |
| 10ge16-6.core1.ash1.he.net                         | 3b |
| 100ge5-1.core1.ash1.he.net                         | 3c |
| esnet.10gigabitethernet5-15.core1.ash1.he.net      | 4a |
| fastserv.core1.ash1.he.net                         | 5a |
| dren-as668.10gigabitethernet16-4.core1.ash1.he.net | 6a |

## Geolocation:

a string identifying  
approximately the  
location of the router

fmt - Fremont, CA

atl - Atlanta, GA

ash - Ashburn, VA

# Network Operators Embed Information in Router Hostnames

|                                                    |    |
|----------------------------------------------------|----|
| 100ge4-1.core3.fmt2.he.net                         | 1a |
| 100ge4-2.core3.fmt2.he.net                         | 1b |
| v1119.core3.fmt2.he.net                            | 1c |
| v1832.core3.fmt2.he.net                            | 1d |
| ge2-9.core1.atl1.he.net                            | 2a |
| ge6-7.core1.atl1.he.net                            | 2b |
| 10ge16-5.core1.ash1.he.net                         | 3a |
| 10ge16-6.core1.ash1.he.net                         | 3b |
| 100ge5-1.core1.ash1.he.net                         | 3c |
| esnet.10gigabitethernet5-15.core1.ash1.he.net      | 4a |
| fastserv.core1.ash1.he.net                         | 5a |
| dren-as668.10gigabitethernet16-4.core1.ash1.he.net | 6a |

## Speed:

a string identifying  
the serialization rate  
of the interface

100ge – 100Gbps Eth

10ge – 10Gbps Eth

ge – 1Gbps Eth

10gigabitethernet



# Network Operators Embed Information in Router Hostnames

|                                                    |    |
|----------------------------------------------------|----|
| 100ge4-1.core3.fmt2.he.net                         | 1a |
| 100ge4-2.core3.fmt2.he.net                         | 1b |
| v1119.core3.fmt2.he.net                            | 1c |
| v1832.core3.fmt2.he.net                            | 1d |
| ge2-9.core1.atl1.he.net                            | 2a |
| ge6-7.core1.atl1.he.net                            | 2b |
| 10ge16-5.core1.ash1.he.net                         | 3a |
| 10ge16-6.core1.ash1.he.net                         | 3b |
| 100ge5-1.core1.ash1.he.net                         | 3c |
| esnet 10gigabitethernet5-15.core1.ash1.he.net      | 4a |
| fastserv core1.ash1.he.net                         | 5a |
| dren-as668.10gigabitethernet16-4.core1.ash1.he.net | 6a |

**Network names:**  
a string identifying the  
name of the network  
that operates the router.

ESnet - AS293  
FastServ - AS29889  
DREN - AS668

# Network Operators Embed Information in Router Hostnames

|                                                    |    |
|----------------------------------------------------|----|
| 100ge4-1.core3.fmt2.he.net                         | 1a |
| 100ge4-2.core3.fmt2.he.net                         | 1b |
| v1119.core3.fmt2.he.net                            | 1c |
| v1832.core3.fmt2.he.net                            | 1d |
| ge2-9.core1.atl1.he.net                            | 2a |
| ge6-7.core1.atl1.he.net                            | 2b |
| 10ge16-5.core1.ash1.he.net                         | 3a |
| 10ge16-6.core1.ash1.he.net                         | 3b |
| 100ge5-1.core1.ash1.he.net                         | 3c |
| esnet.10gigabitethernet5-15.core1.ash1.he.net      | 4a |
| fastserv.core1.ash1.he.net                         | 5a |
| dren-as668.10gigabitethernet16-4.core1.ash1.he.net | 6a |

## ASNs:

A number identifying the ASN of the network that operates the router.

DREN – AS668



# Hoiho

*(Holistic Orthography of Internet Hostname Observations)*

|                                  |            |        |    |
|----------------------------------|------------|--------|----|
| 100ge4-1                         | core3.fmt2 | he.net | 1a |
| 100ge4-2                         | core3.fmt2 | he.net | 1b |
| v1119                            | core3.fmt2 | he.net | 1c |
| v1832                            | core3.fmt2 | he.net | 1d |
| ge2-9                            | core1.atl1 | he.net | 2a |
| ge6-7                            | core1.atl1 | he.net | 2b |
| 10ge16-5                         | core1.ash1 | he.net | 3a |
| 10ge16-6                         | core1.ash1 | he.net | 3b |
| 100ge5-1                         | core1.ash1 | he.net | 3c |
| esnet.10gigabitethernet5-15      | core1.ash1 | he.net | 4a |
| fastserv                         | core1.ash1 | he.net | 5a |
| dren-as668.10gigabitethernet16-4 | core1.ash1 | he.net | 6a |

**Goal: automatically  
learn regular  
expressions that  
extract features from  
router hostnames.**

$^{\wedge}[^{\wedge}\backslash.]^{+}\backslash.(\text{core}\backslash\text{d}^{+}\backslash.[\text{a-z}]\backslash\text{d}^{+})\backslash.\text{he}\backslash.\text{net}\$$



# Key challenge #1: naming heterogeneity

```
┌───────────────────────────────────────────┐
│      as15133.cr2-nyc6.ip4.gtt.net      │
│ edgecast-ic-317659-nyk-b5.c.telia.net │
│ edgecast.newyork51.new.seabone.net    │
└───────────────────────────────────────────┘
┌───────────────────────────────────────────┐
│      as3491.ip4.gtt.net                │
│ pccw-ic-319977-ldn-b3.c.telia.net      │
│ be2-132.br02.ldn12.pccwbtn.net         │
└───────────────────────────────────────────┘
┌───────────────────────────────────────────┐
│      as2914.cr3-lax1.ip4.gtt.net      │
│ ae3.er2.lga5.us.zip.zayo.com          │
└───────────────────────────────────────────┘
```

**Operators decide their  
own conventions  
independently**

e.g. some networks embed  
ASNs, some embed AS names

# Key challenge #1: naming heterogeneity

```
as15133.cr2-nyc6.ip4.gtt.net  
edgecast-ic-317659-nyk-b5.c.telia.net  
edgecast.newyork51.new.seabone.net  
as3491.ip4.gtt.net  
pccw-ic-319977-ldn-b3.c.telia.net  
be2-132.br02.ldn12.pccwbtn.net  
as2914.cr3-lax1.ip4.gtt.net  
ae3.er2.lga5.us.zip.zayo.com
```

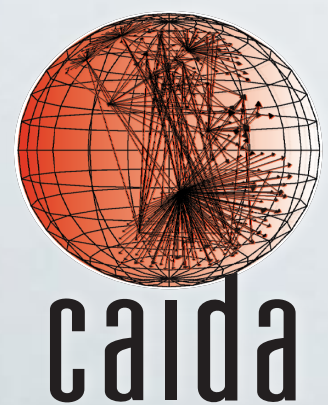
**Operators decide their  
own conventions  
independently**

Even geolocation codes: e.g.  
nyc, nyk, new all for New York.

# Key challenge #2: almost no ground truth to train a model

- **CAIDA Internet Topology Data Kit (ITDK)**
- **Heavily curated router-level topology dataset published roughly twice a year**
- Routers, with aliases inferred using heuristics
- Links between routers inferred using traceroute
- Router ownership inferred using heuristics
- DNS hostnames (PTR records)
- 19 ITDK datasets between July 2010 to April 2021

Hoiho  
Input  
Data





# Hoiho approach by example: extracting ASNs

| training<br>ASN | hostname<br>(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)        |

# Hoiho approach by example: extracting ASNs

| training<br>ASN | hostname<br>(PTR record)        |     |                                     | True Positives |
|-----------------|---------------------------------|-----|-------------------------------------|----------------|
| 109             | 109.sgw.equinix.com             | (a) | #1 ^(\d+)\.[^\.]+\..equinix\.com\$  | a, b, c        |
| 714             | 714.os.equinix.com              | (b) | #2 ^p(\d+)\.[^\.]+\..equinix\.com\$ | d, f           |
| 714             | 714.me1.equinix.com             | (c) | #3 ^s(\d+)\.[^\.]+\..equinix\.com\$ | e, g           |
| 714             | p714.sgw.equinix.com            | (d) | #4 ^(\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) |                                     |                |



# Hoiho approach by example: extracting ASNs

| training<br>ASN | hostname<br>(PTR record)            |                                                                        | True Positives      |
|-----------------|-------------------------------------|------------------------------------------------------------------------|---------------------|
| 109             | 109.sgwg.equinix.com (a)            | #1 $^(\d+)\backslash.[^\.]+\backslash.equinix\backslash.com\$$         | a, b, c             |
| 714             | 714.os.equinix.com (b)              | #2 $^p(\d+)\backslash.[^\.]+\backslash.equinix\backslash.com\$$        | d, f                |
| 714             | 714.me1.equinix.com (c)             | #3 $^s(\d+)\backslash.[^\.]+\backslash.equinix\backslash.com\$$        | e, g                |
| 714             | p714.sgwg.equinix.com (d)           | #4 $^(\d+)\-.\+ \backslash.equinix\backslash.com\$$                    | h, i, j, k          |
| 714             | s714.sgwg.equinix.com (e)           |                                                                        |                     |
| 24115           | p24115.mel.equinix.com (f)          | #5 $^(?:pls)?(\d+)\backslash.[^\.]+\backslash.equinix\backslash.com\$$ | a, b, c, 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)        |                                                                        |                     |



# Hoiho approach by example: extracting ASNs

| training<br>ASN | hostname<br>(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

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

a, b, c, d, e, f, g

```
#6 ^(?:pls)?(\d+)\.[a-z\d]+\..equinix\.com$
```

a, b, c, d, e, f, g

# Hoiho approach by example: extracting ASNs

| training<br>ASN | hostname<br>(PTR record)            |                                                                                                                                                                    | True Positives      |
|-----------------|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 109             | 109.sgw.equinix.com (a)             | #1 $^(\backslash d+)\backslash.[^\backslash.]+\backslash.equinix\backslash.com\$$                                                                                  | a, b, c             |
| 714             | 714.os.equinix.com (b)              | #2 $^p(\backslash d+)\backslash.[^\backslash.]+\backslash.equinix\backslash.com\$$                                                                                 | d, f                |
| 714             | 714.me1.equinix.com (c)             | #3 $^s(\backslash d+)\backslash.[^\backslash.]+\backslash.equinix\backslash.com\$$                                                                                 | e, g                |
| 714             | p714.sgw.equinix.com (d)            | #4 $^(\backslash d+)-\backslash.+ \backslash.equinix\backslash.com\$$                                                                                              | h, i, j, k          |
| 714             | s714.sgw.equinix.com (e)            |                                                                                                                                                                    |                     |
| 24115           | p24115.mel.equinix.com (f)          | #5 $^(?:pls)?(\backslash d+)\backslash.[^\backslash.]+\backslash.equinix\backslash.com\$$                                                                          | a, b, c, 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)       | #6 $^(?:pls)?(\backslash d+)\backslash.[a-z\backslash d]+\backslash.equinix\backslash.com\$$                                                                       | a, b, c, d, e, f, g |
| 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)            | #7 $^(?:pls)?(\backslash d+)\backslash.[a-z\backslash d]+\backslash.equinix\backslash.com\$$<br>$^(\backslash d+)-\backslash.+ \backslash.equinix\backslash.com\$$ | a, b, c, d, e, f, g |
| 8075            | 8074.hkg.equinix.com (o)            |                                                                                                                                                                    | h, i, j, k          |
| 55923           | 45437-sy1-ix.equinix.com (p)        |                                                                                                                                                                    |                     |



# Summary

- We have publicly released the source code implementation
  - <https://www.caida.org/tools/measurement/scamper/>
- We have publicly released inferred regexes, as well as webpages demonstrating how each regex applied to the training data
  - <https://www.caida.org/publications/papers/2019/hoiho/>
  - <https://www.caida.org/publications/papers/2020/hoiho/>
- We welcome feedback on applying our techniques in other domains