Growth Analysis of a Large ISP

Andrew Ferguson, Jordan Place, and Rodrigo Fonseca

Cogent Communications

1. One of the world's largest IP networks, covering 3 continents
2. Public map (below) provides a static snapshot at the city-level
3. Since Jan. 2012, we made weekly snapshots at the router interface-level

http://www.cogentco.com/en/network/network-map

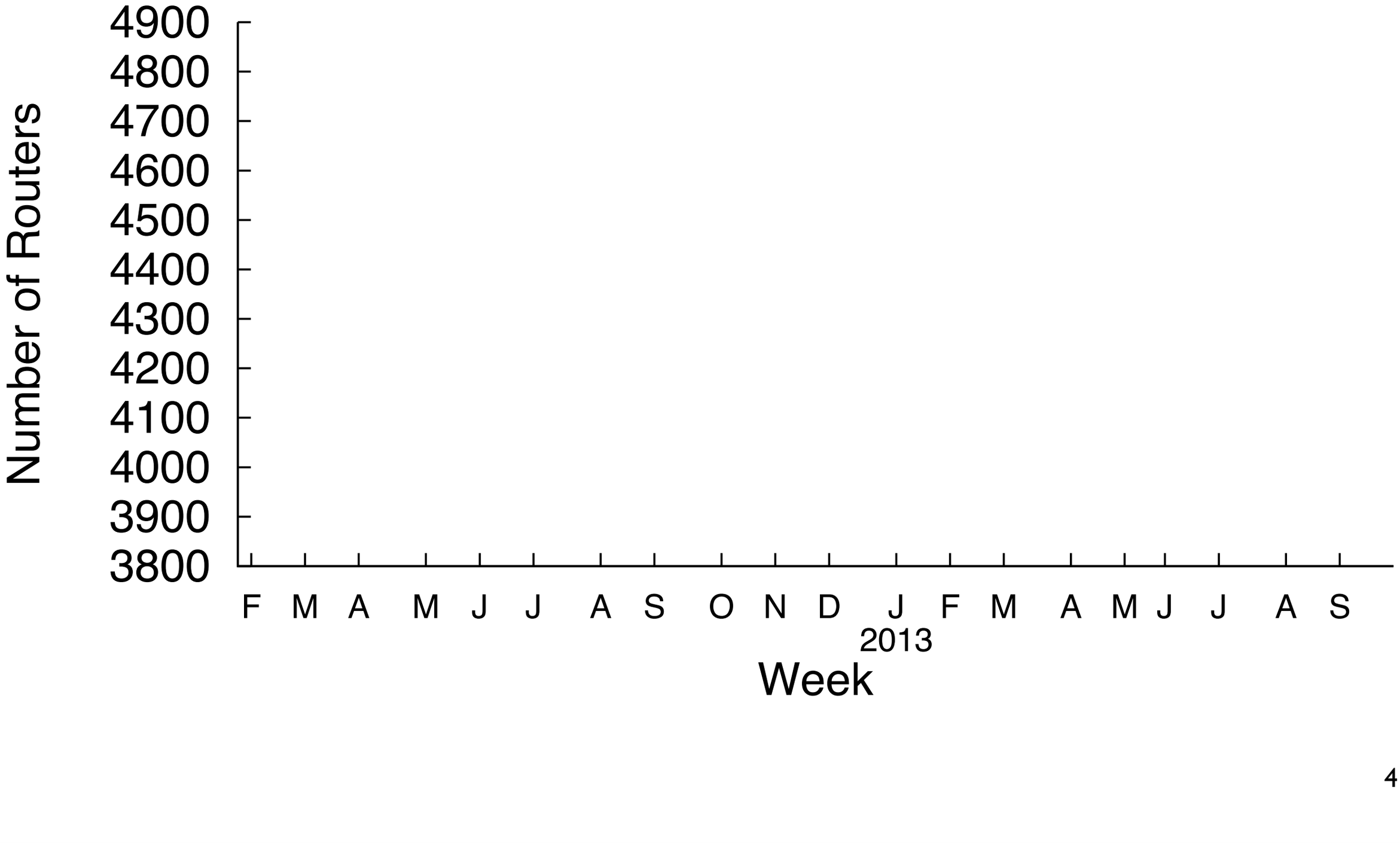
the "Large ISP" is Cogent, we have been measuring it almost every week since January 2012.

these weekly snapshots give us a unique opportunity to examine the dynamics of a large ISP network

Sample Results

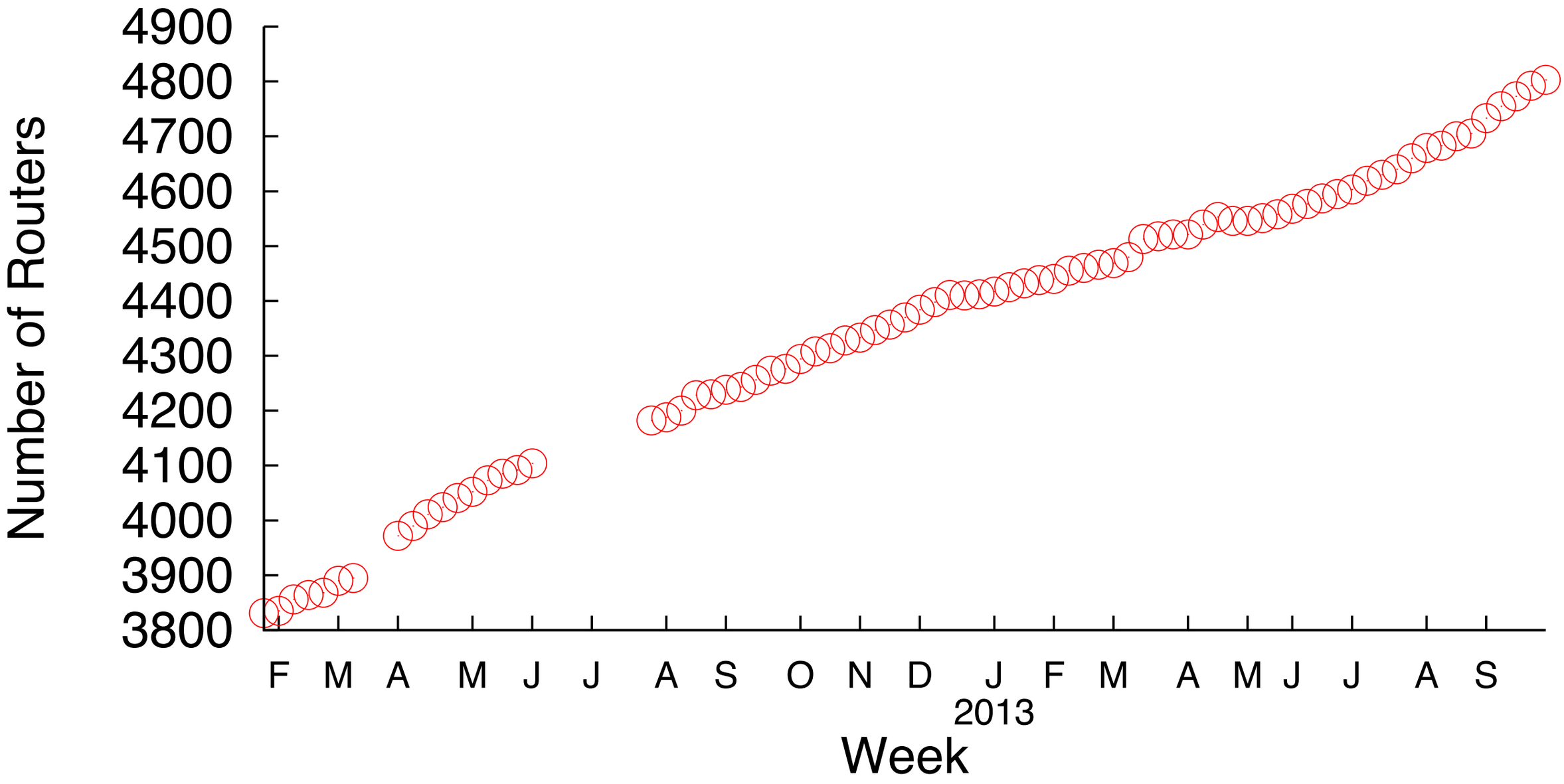
Before I tell you how we collected this data, I want to show you some initial results.

Inferred Router Growth

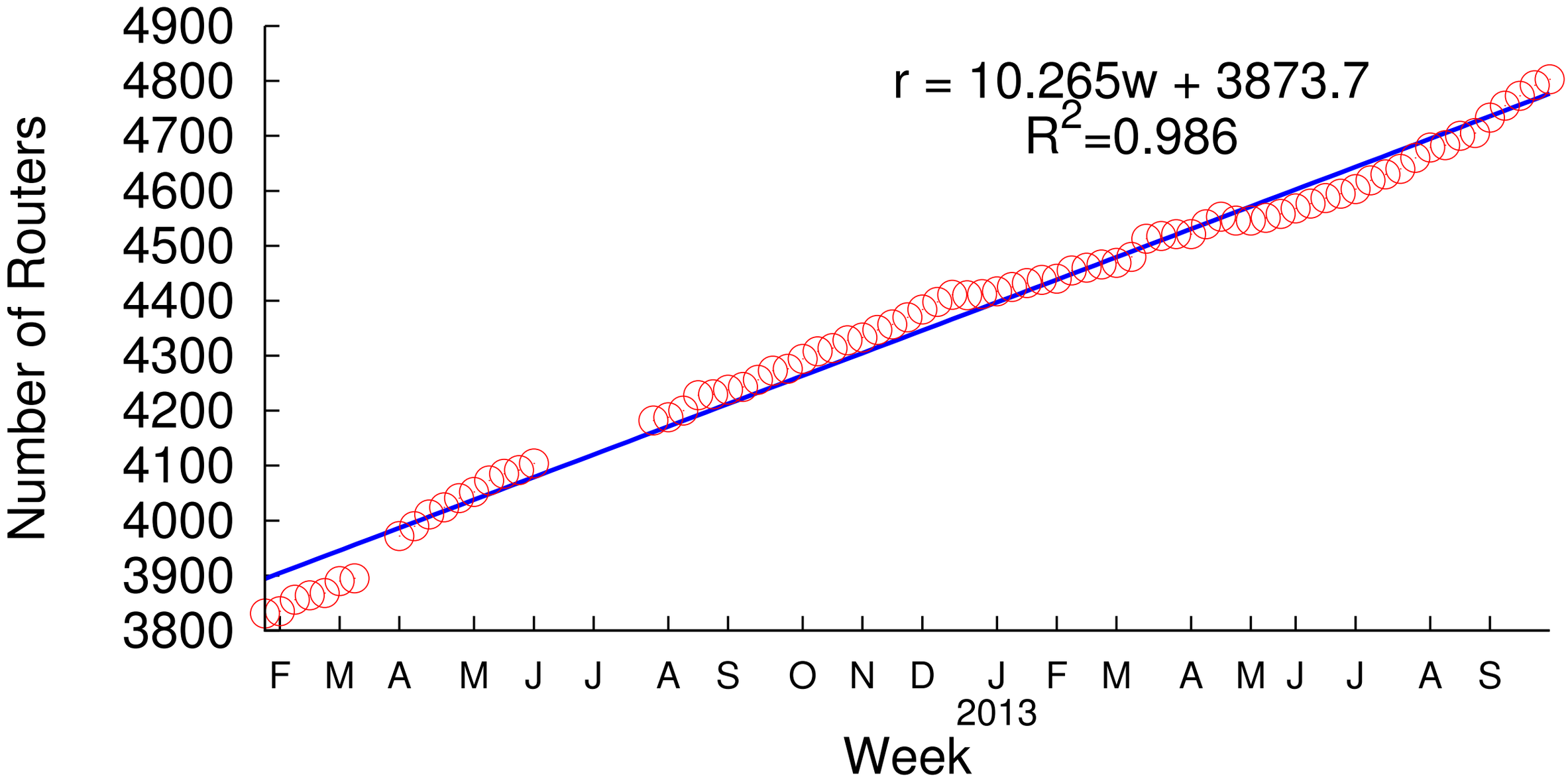


Along the x axis, we have time since January 2012.
Along the y axis, the number of inferred routers.

Inferred Router Growth



Inferred Router Growth

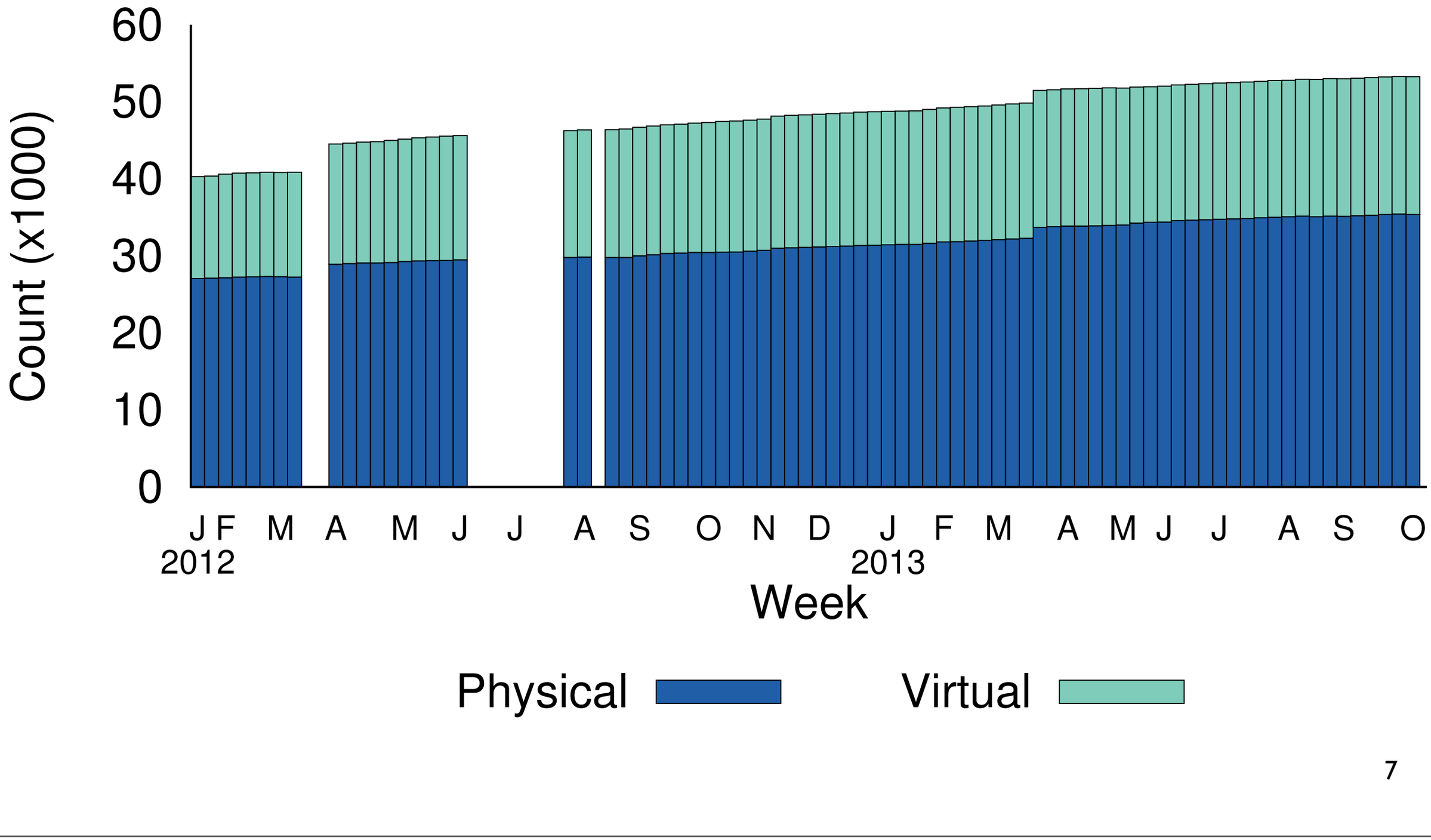


Growth in routers over time -- averaging 11.3 new routers per week. We see some decrease at the end of 2012, which has carried into 2013.

Interestingly, the number of interfaces per router remains roughly constant over this period. We will take a look at the interface growth rate in the next graph.

Before I go on, I want to point out there are two gaps in our graphs, which come from failures in our measurement infrastructure.

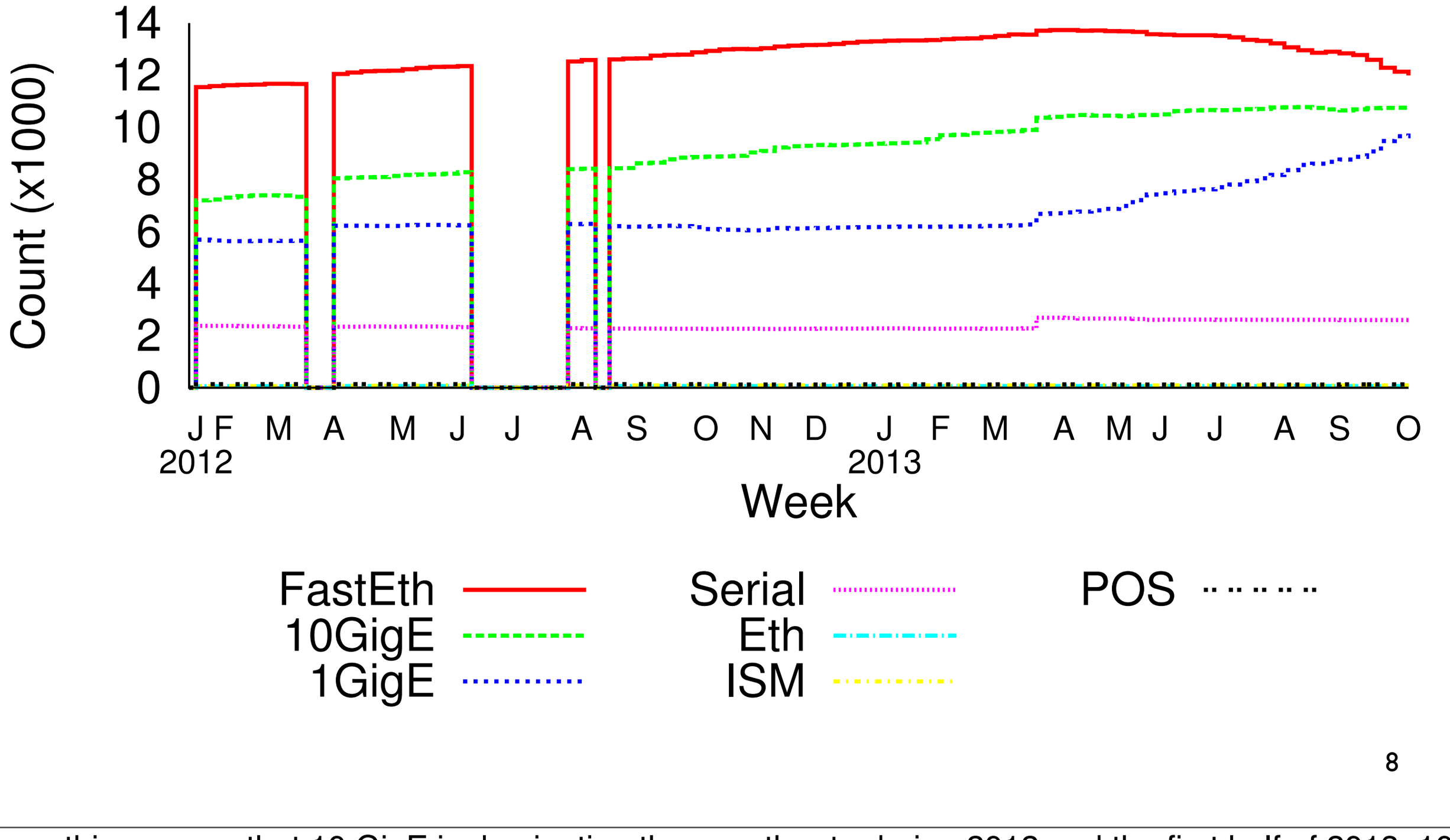
Interface Growth



Virtual interfaces include VLAN, Loopback, Tunnel, and Multilink interface types.
Physical interfaces are 10, 100, 1000, 10k Mbps Ethernet, etc.

Let’s dive into the physical interfaces

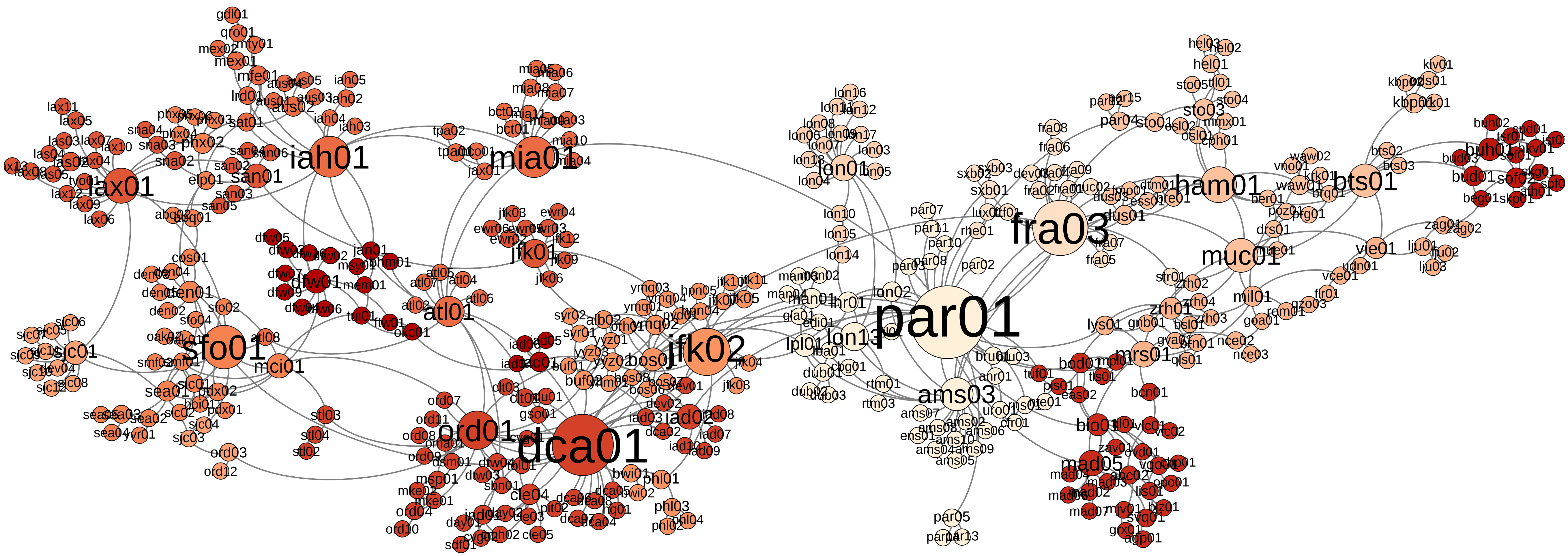
Physical Interface Breakdown



From this, we see that 10 GigE is dominating the growth rate during 2012 and the first half of 2013. 100 MbE is still the dominant product. This corresponds with information in the financial reports Cogent files with the government, but not with the detail we show above.

An incredible trend we've discovered since the camera-ready deadline is the replacement of 100 Mbps infrastructure with 1 Gbps infrastructure.

Visualization of Inferred Paths



1. Infer connection between two routers sharing appropriate /30 subnets
2. Nodes are sized according to the number of paths passing through them
3. Layout above is force-directed (no geographical information used)

This is a static snapshot of the network for the week of April 7, 2013. We constructed graphs like these for each week in our dataset, and explore the graph evolution in our paper.

How did we do this?

So, how did we build this time-series data set?

Cogent's DNS Records

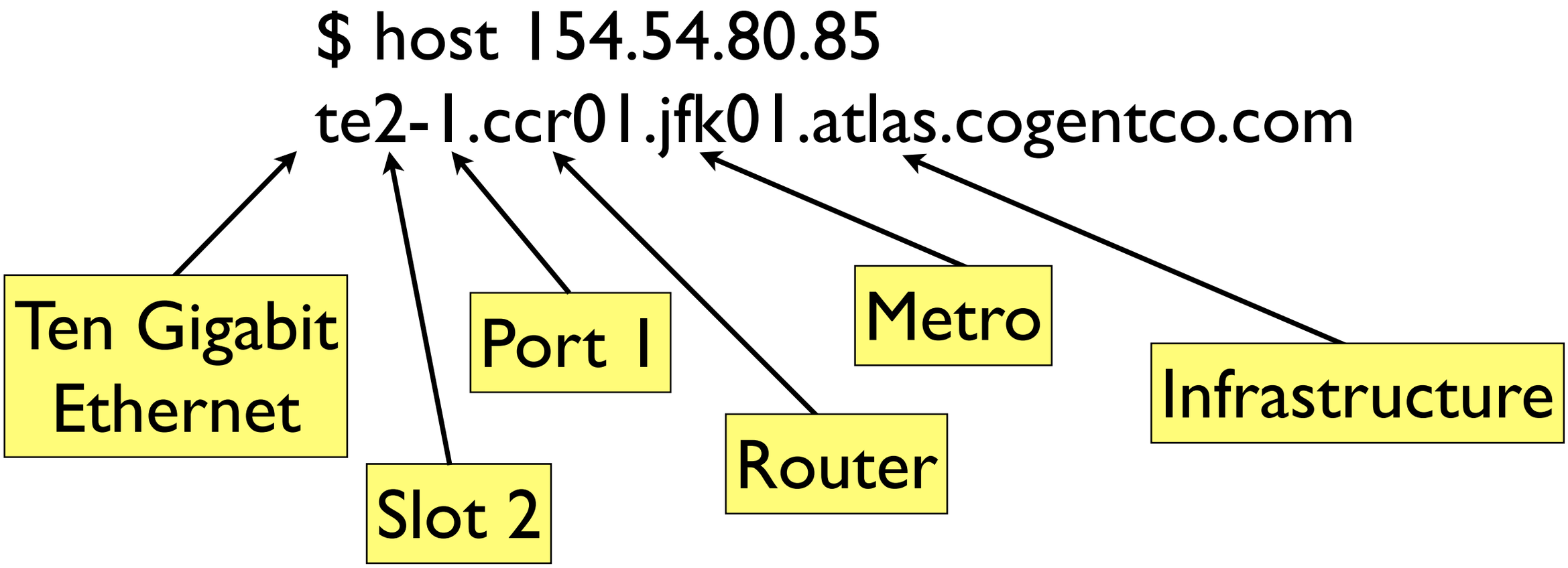
```
$ host 154.54.80.85
te2-1.ccr01.jfk01.atlas.cogentco.com
```


Like many ISPs, Cogent includes information about their network in their DNS records, which is particularly useful for traceroute and other debugging tasks.

This reverse DNS entry is rich with information if we make some educated guesses, based on Cisco naming conventions.

Cogent only provides reverse DNS records (not forward DNS), but with a lucky guess (or some brute force) we can determine if there is an interface next to it.

Cogent's DNS Records



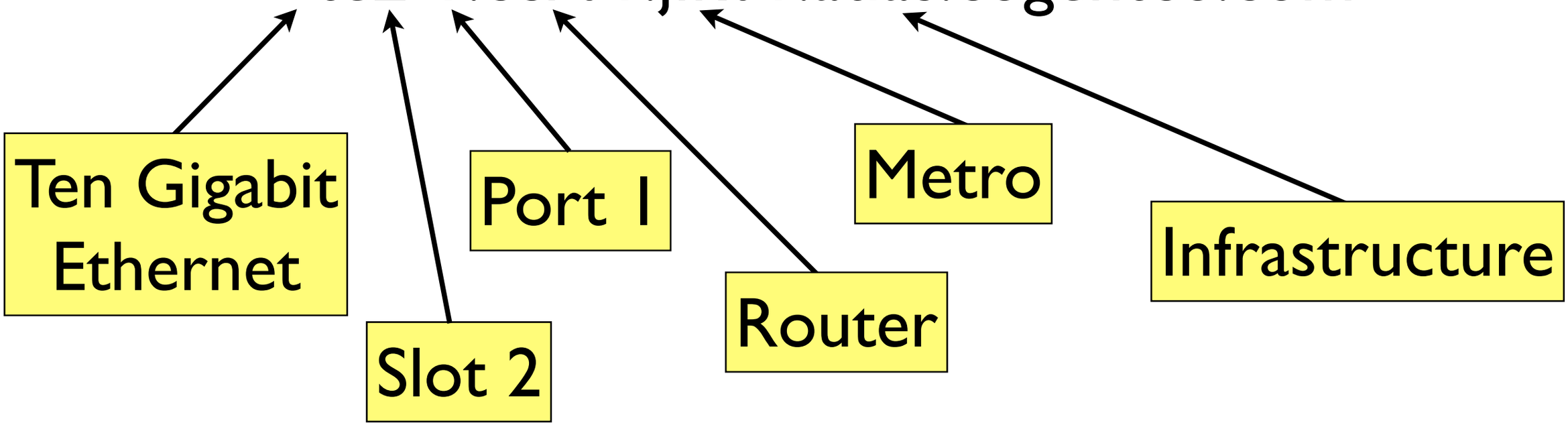
Like many ISPs, Cogent includes information about their network in their DNS records, which is particularly useful for traceroute and other debugging tasks.

This reverse DNS entry is rich with information if we make some educated guesses, based on Cisco naming conventions.

Cogent only provides reverse DNS records (not forward DNS), but with a lucky guess (or some brute force) we can determine if there is an interface next to it.

Cogent's DNS Records

$ host 154.54.80.85
te2-1.ccr01.jfk01.atlas.cogentco.com



$ host 154.54.25.17
te2-2.ccr01.jfk01.atlas.cogentco.com

Like many ISPs, Cogent includes information about their network in their DNS records, which is particularly useful for traceroute and other debugging tasks.

This reverse DNS entry is rich with information if we make some educated guesses, based on Cisco naming conventions.

Cogent only provides reverse DNS records (not forward DNS), but with a lucky guess (or some brute force) we can determine if there is an interface next to it.

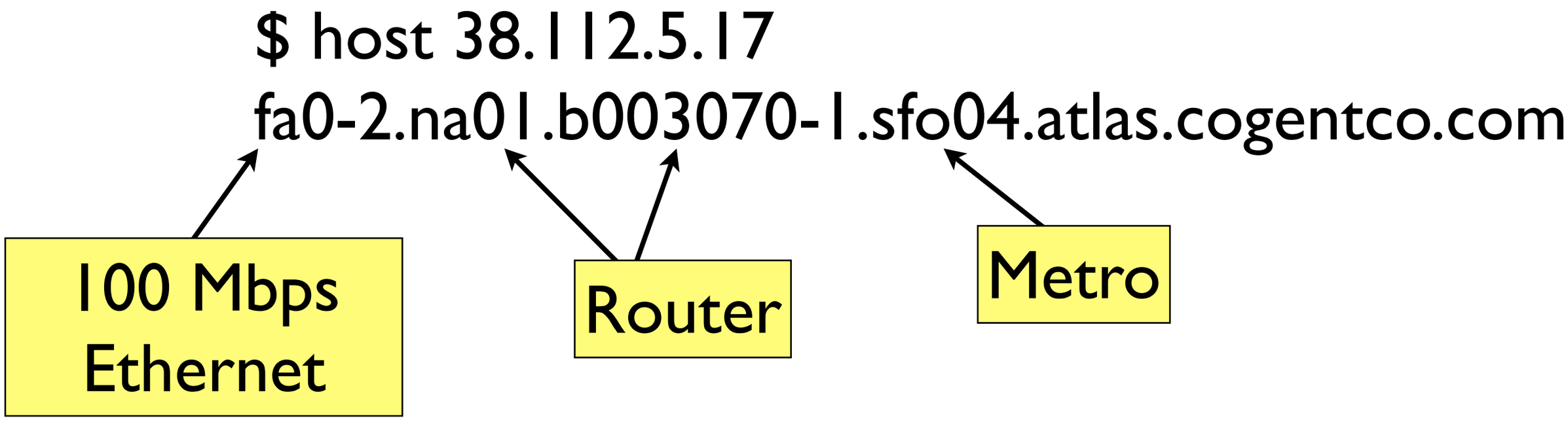
Cogent's DNS Records (2)

```
$ host 38.112.5.17
fa0-2.na01.b003070-1.sfo04.atlas.cogentco.com
```


Let's take a look at another example.

This time, let's check the other useable address in the same /30 subnet.

Cogent's DNS Records (2)

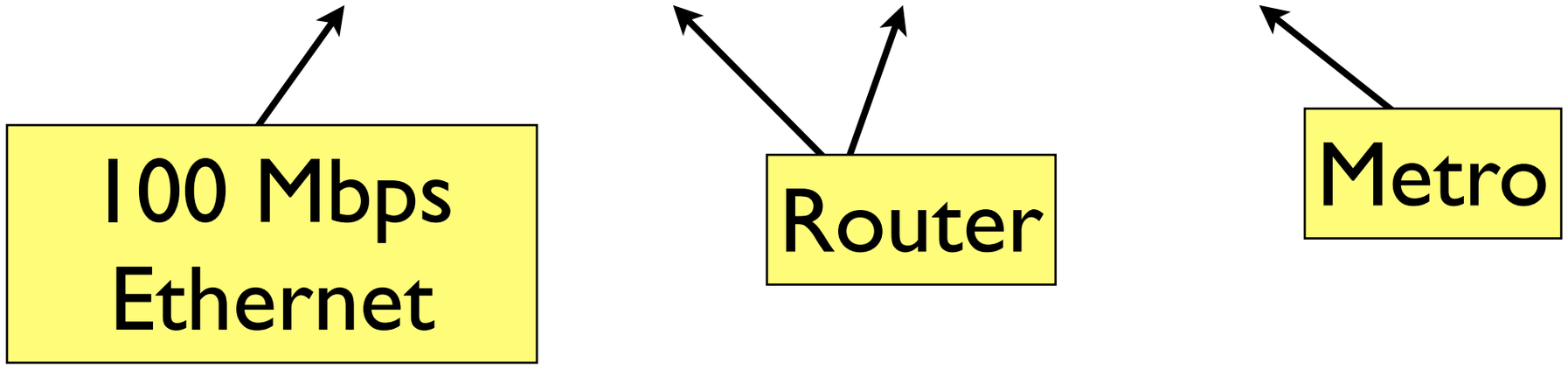


Let's take a look at another example.

This time, let's check the other useable address in the same /30 subnet.

Cogent's DNS Records (2)

$ host 38.112.5.17
fa0-2.na01.b003070-1.sfo04.atlas.cogentco.com



$ host 38.112.5.18
Tetratech.demarc.cogentco.com

Let's take a look at another example.

This time, let's check the other useable address in the same /30 subnet.

Cogent's DNS Records (2)

$ host 38.112.5.17
fa0-2.na01.b003070-1.sfo04.atlas.cogentco.com

100 Mbps Ethernet

Router

Metro

Pair in /30 Subnet

$ host 38.112.5.18
Tetratech.demarc.cogentco.com

California engineering firm

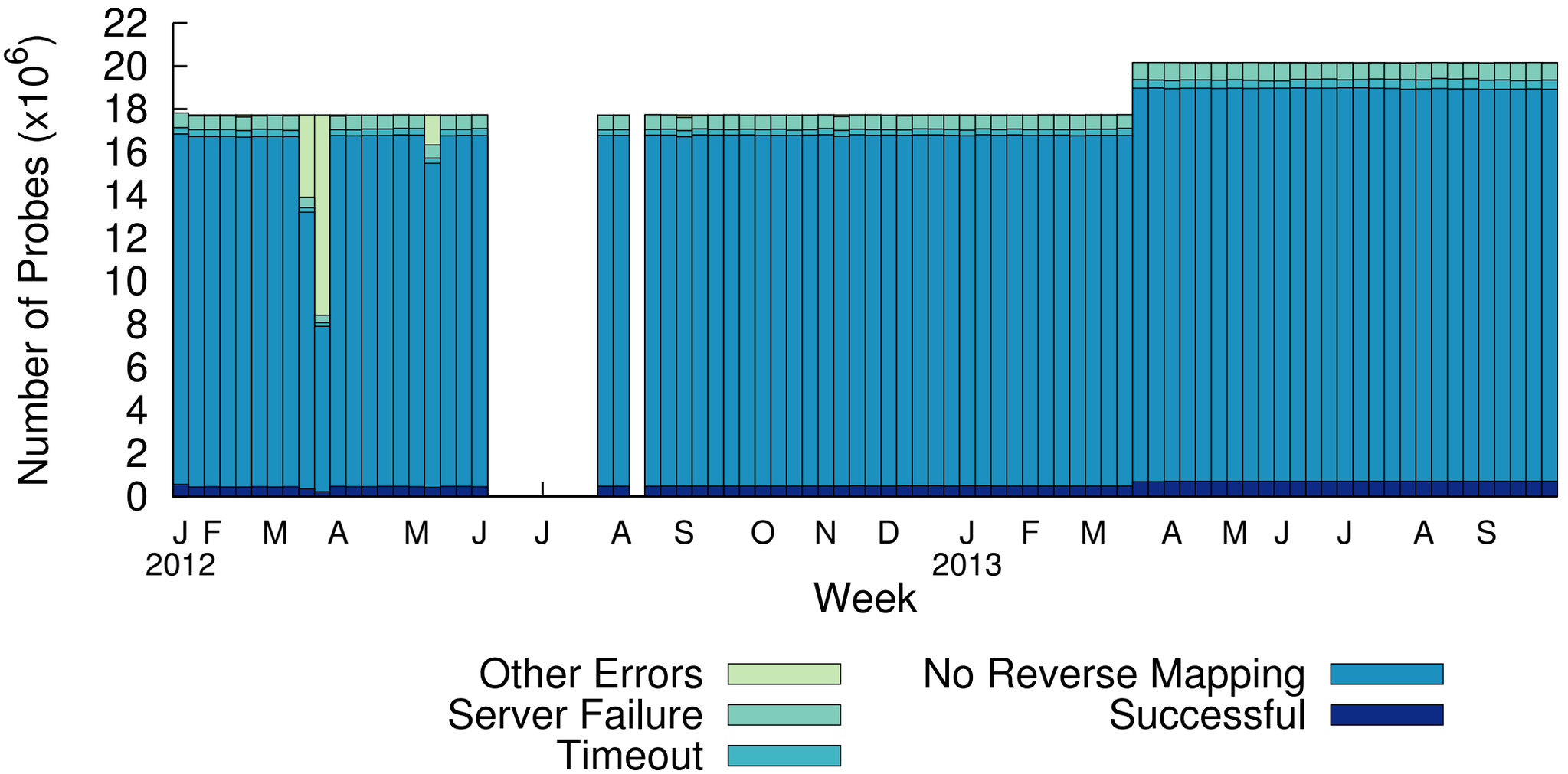
Related business entities

Let's take a look at another example.

This time, let's check the other useable address in the same /30 subnet.

Weekly Surveys

1. Perform 20+ million reverse DNS queries weekly for Cogent-owned IPs
2. Issued from ~100 PlanetLab locations across the globe
3. Also run iffinder on the previous week's discovered interfaces (~55k)



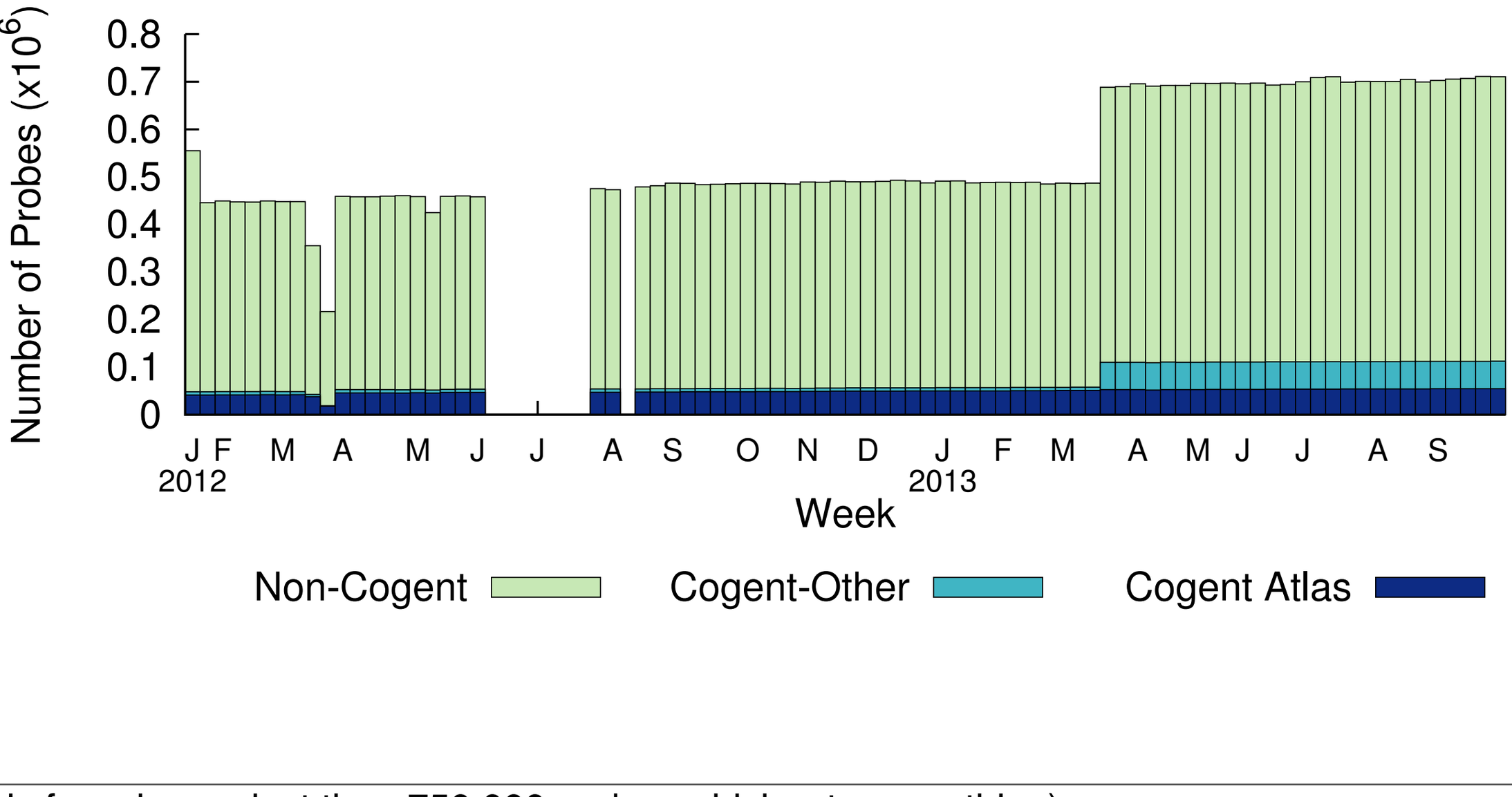
(Explain that we found IP addresses by looking in ARIN & RIPE databases)

Other errors include ID mismatches, replies from unexpected sources, corrupt responses, etc.

(Lastly,

Weekly Surveys

1. Perform 20+ million reverse DNS queries weekly for Cogent-owned IPs
2. Issued from ~100 PlanetLab locations across the globe
3. Also run iffinder on the previous week's discovered interfaces (~55k)



(This is focusing on just the ~750,000 probes which return anything)

(Also, be sure to explain the big jump in March when we added the additional 3 million IP address)

1. Can we believe this data?
2. How high is the coverage?

Validation

There are two key parts we want to validate: how believable is this data? and how much of Cogent's network did we cover?

The short answers are: "Yes" and "Quite high" for details, see our paper

Validation Approaches

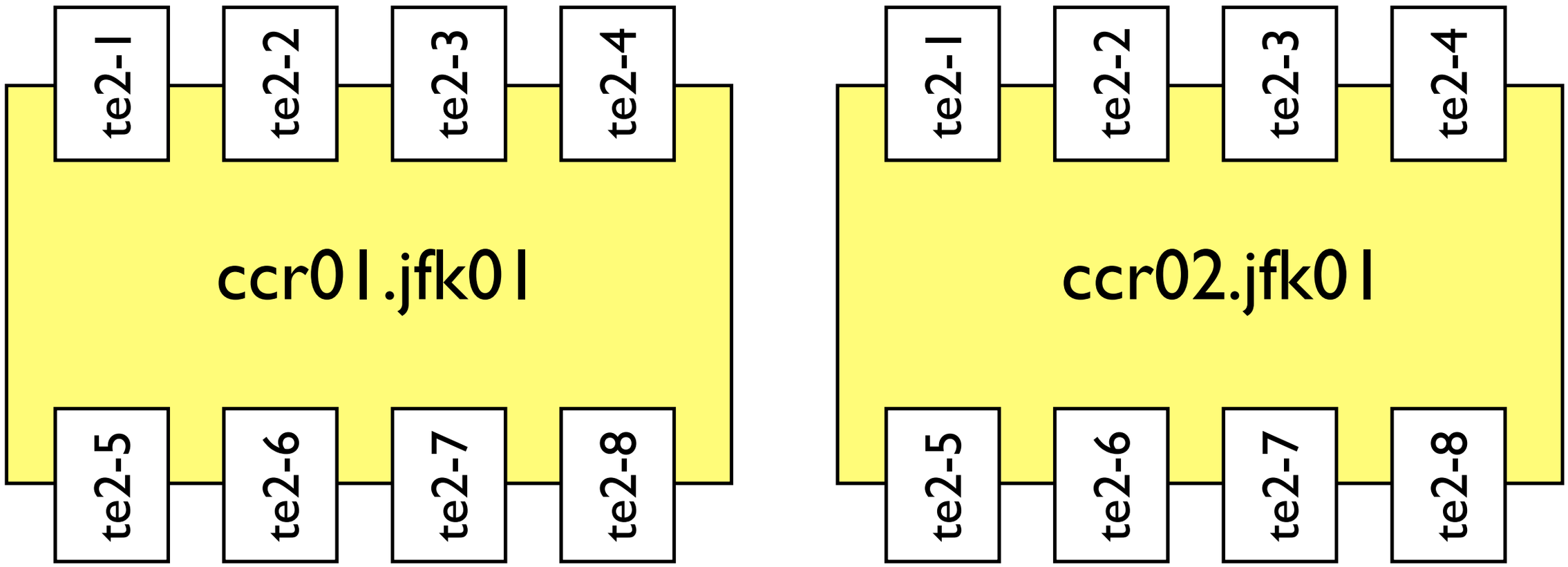
1. Compare with iffinder
2. Check Cogent's public information
3. Use complete set of IPv4 DNS records

Comparison with iffinder

iffinder — a well-known solution to the “alias resolution” problem with a low rate of false positives

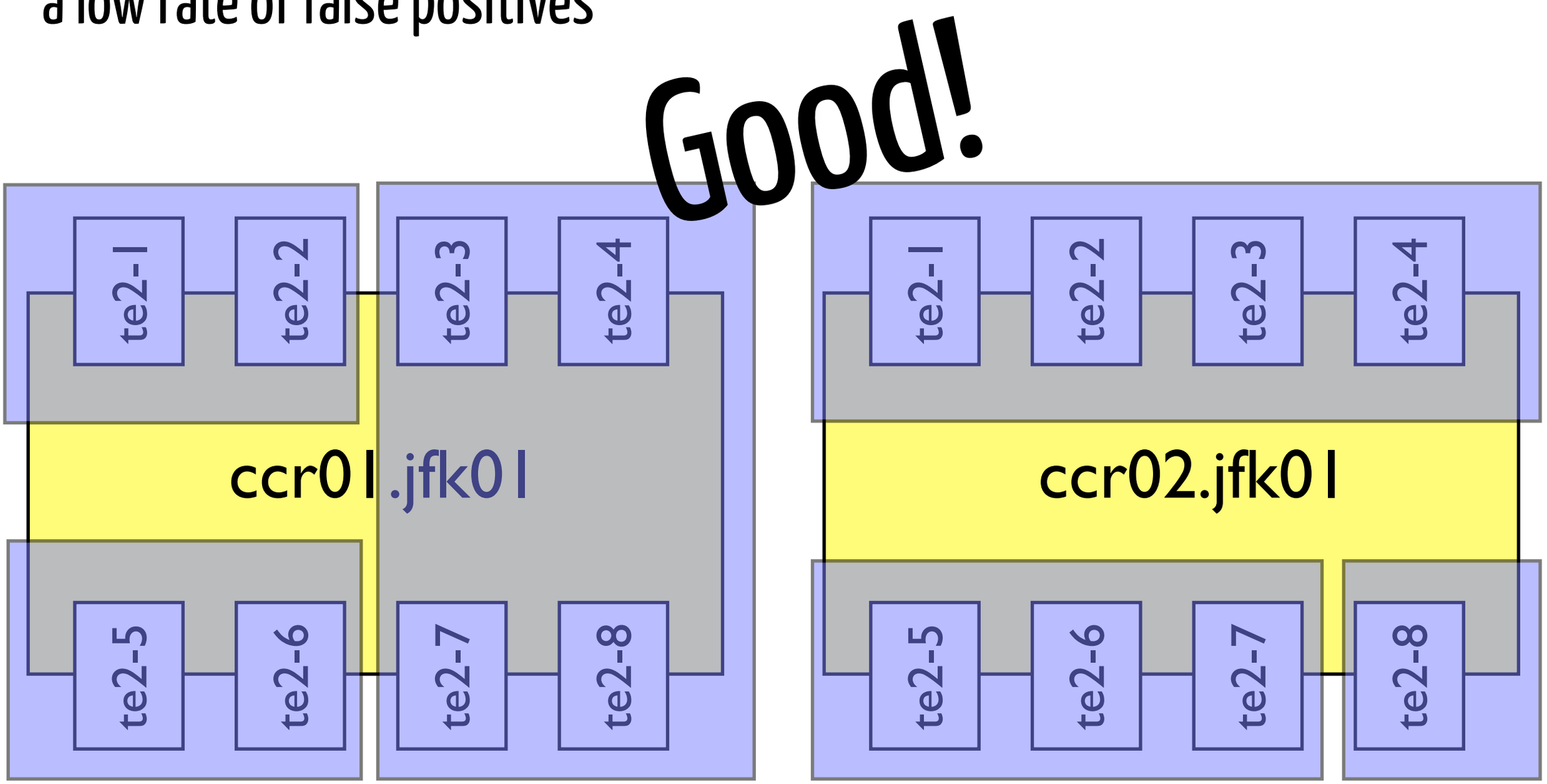
Comparison with iffinder

iffinder — a well-known solution to the “alias resolution” problem with a low rate of false positives



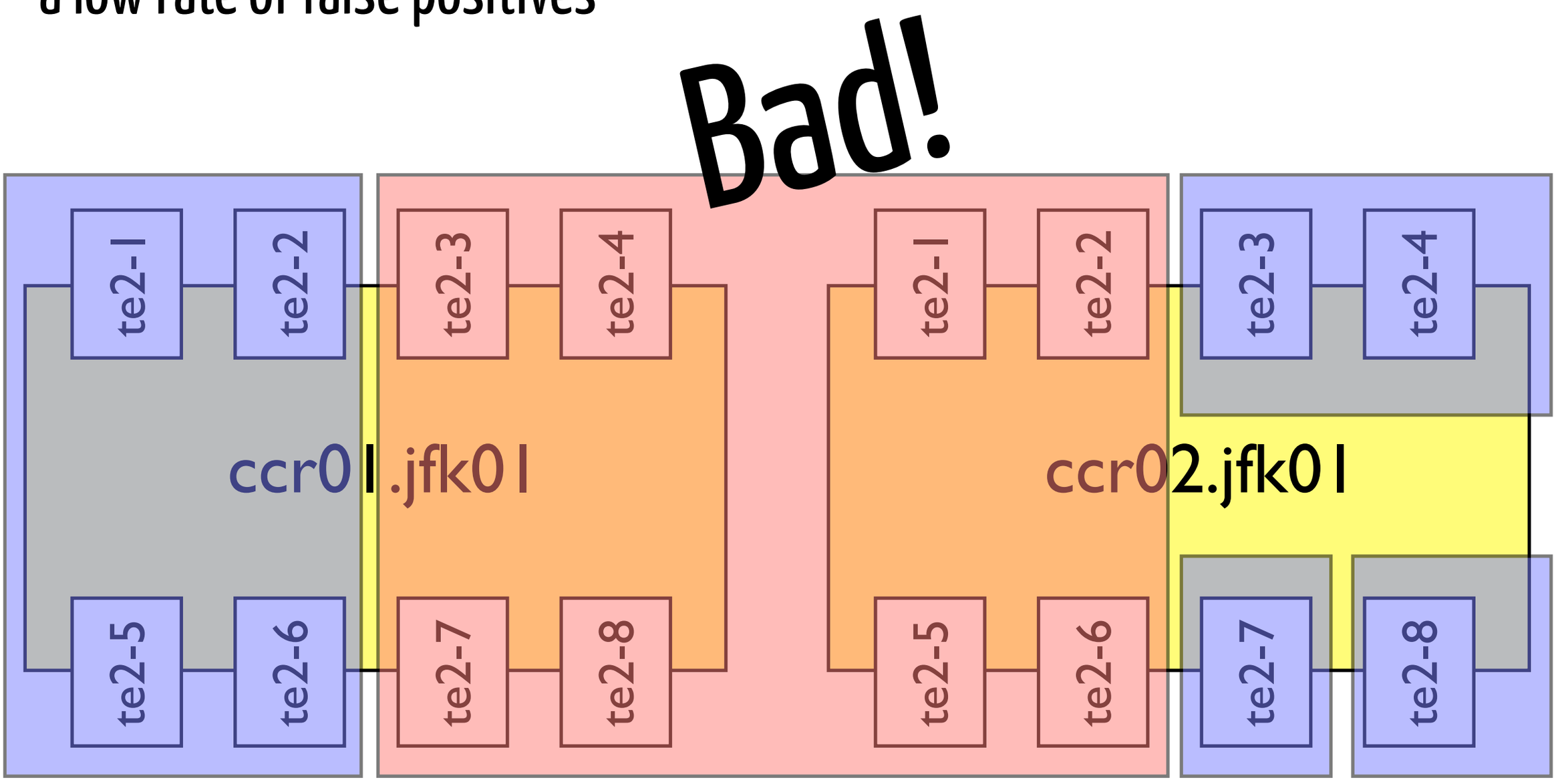
Comparison with iffinder

iffinder — a well-known solution to the “alias resolution” problem with a low rate of false positives



Comparison with iffinder

iffinder — a well-known solution to the "alias resolution" problem with a low rate of false positives



Each week, less than 1% of the inferred routers had a bad comparison with iffinder, such as the one shown above. And furthermore, 95.8% of those discrepancies were resolved by the next week, suggesting that it was just a transient discrepancy.

systems.cs.brown.edu/cogent

Andrew Ferguson
adf@cs.brown.edu

(sell the data more!!)

So far, we have just scratched the surface of looking at Cogent's network. Our complete data set is available online at this address.

I'm happy to take your questions at this time...

Co-authors

- Jordan Place
- Rodrigo Fonseca

systems.cs.brown.edu/cogent

Andrew Ferguson
adf@cs.brown.edu

… or you can direct them to my collaborators as well.

Thank you very much!