

A Configuration-only Approach  
to  
**FIB Reduction**

Paul Francis  
Hitesh Ballani, Tuan Cao  
Cornell

# Virtual Aggregation

- An approach to shrinking FIBs (and RIBs)
  - In interface-card FIB, maybe control-card RIB
- Works with legacy routers
  - New configuration only
- ISPs can independently and autonomously deploy
- IPv4 and IPv6
- FIB-size versus latency/load trade-off

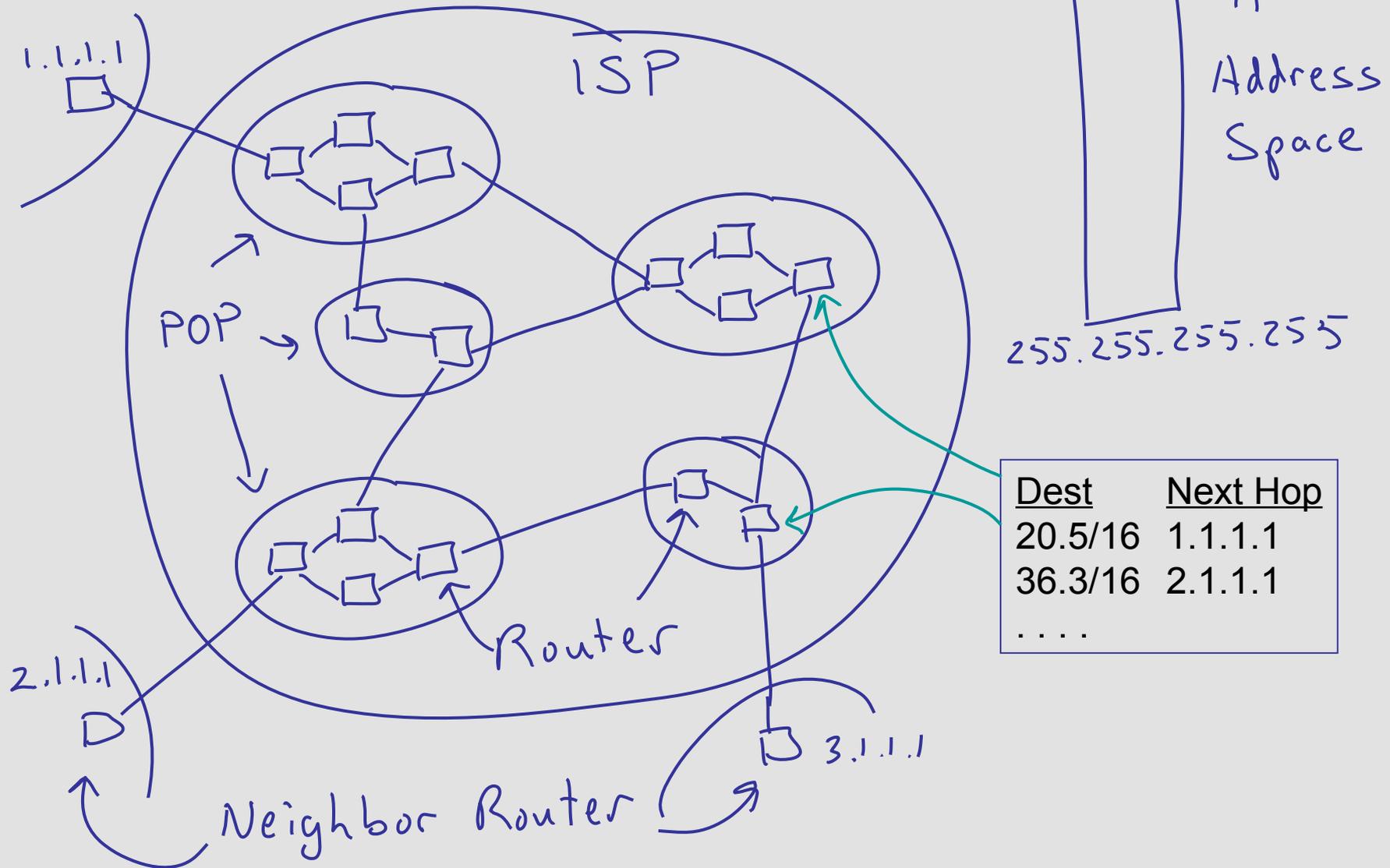
# Outline

- Project status
- Mechanics
- Evaluation results

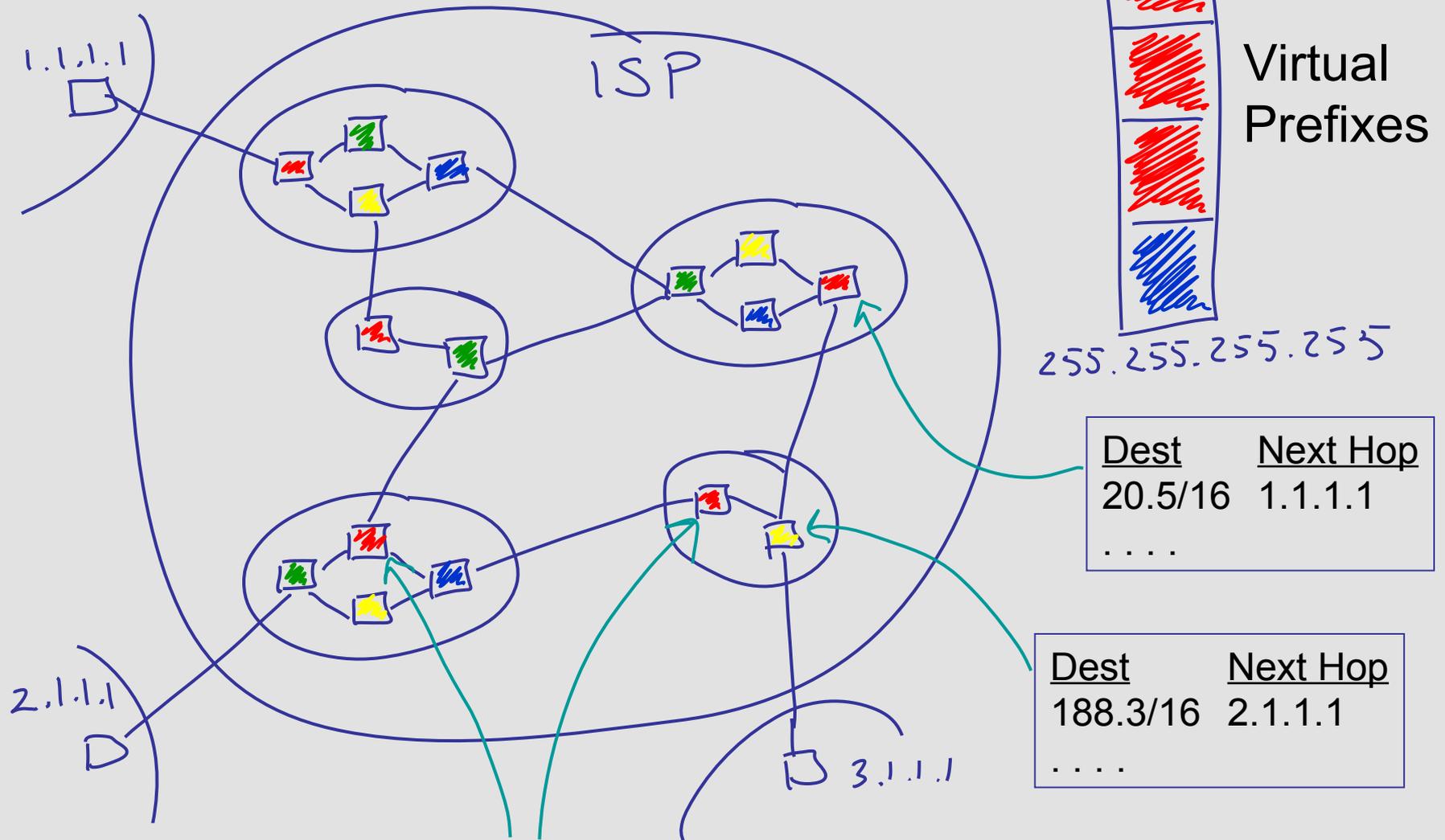
# Status

- Tested a couple of versions of VA by configuring on Linux and Cisco routers
  - Simple experiments (~10 routers)
  - Cisco 7301 and Cisco 12000
- Modeled using data from a large ISP
  - (router topology and traffic matrix)
- Have tested for large routing tables and fail-over
- Have *not* tested on a live network
- Have *not* test IPv6

# Today: All routers have routes to all destinations

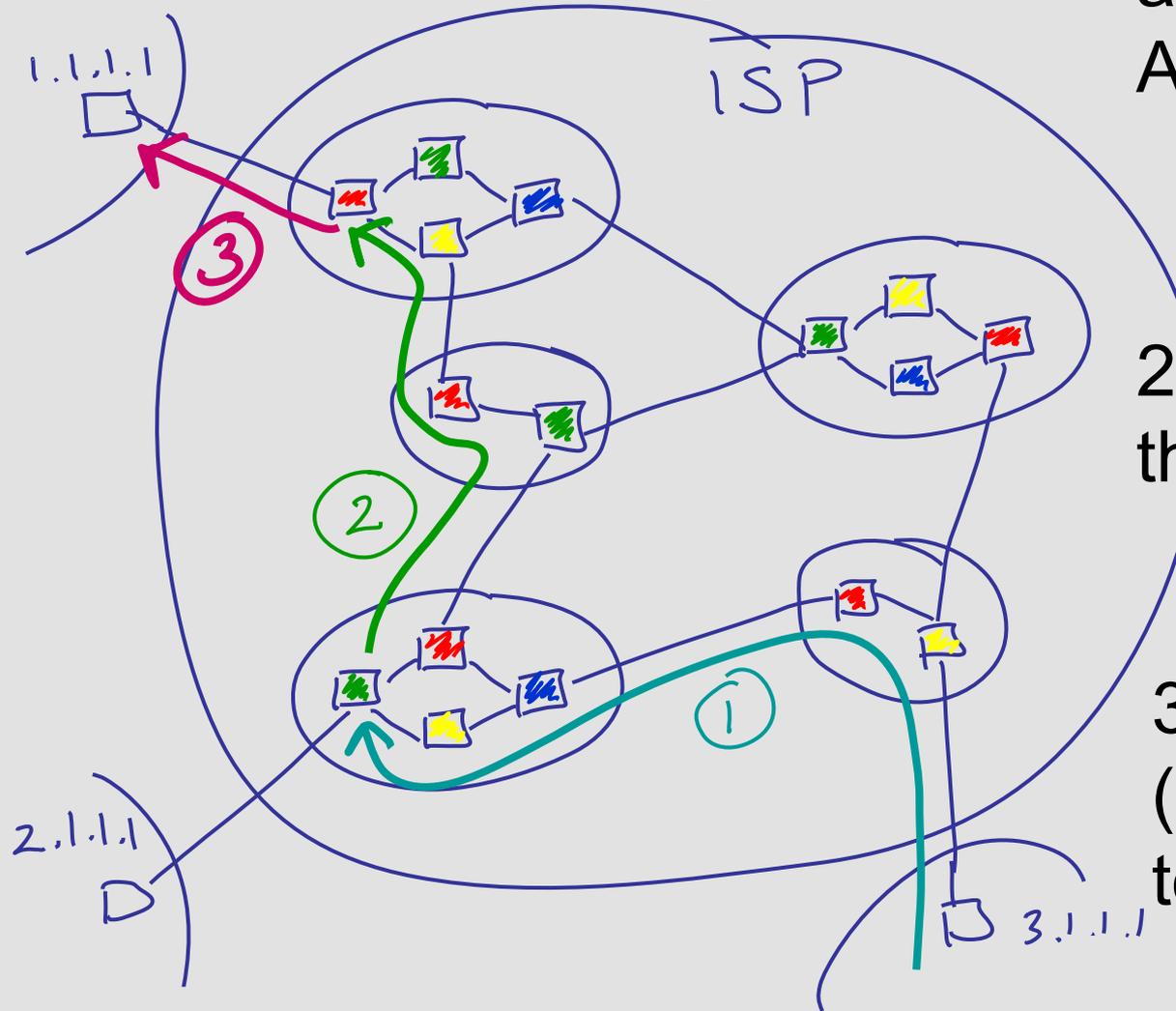


# Virtual Aggregation: Routers have routes to only part of the address space



“Aggregation Point” routers for the red Virtual Prefix

# Paths through the ISP have three components:

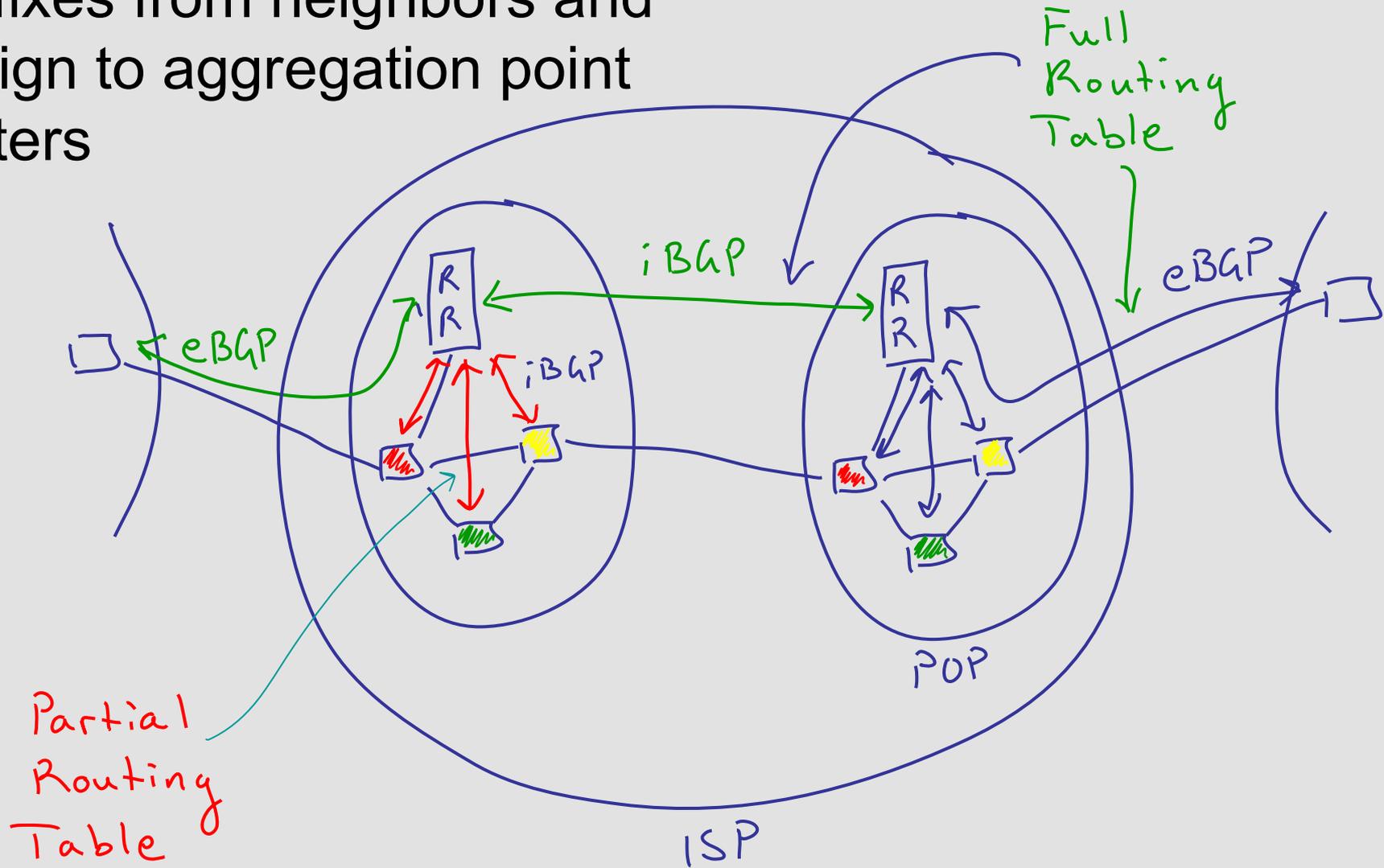


1: Native route to a nearby Aggregation Point

2: MPLS tunnel to the egress router

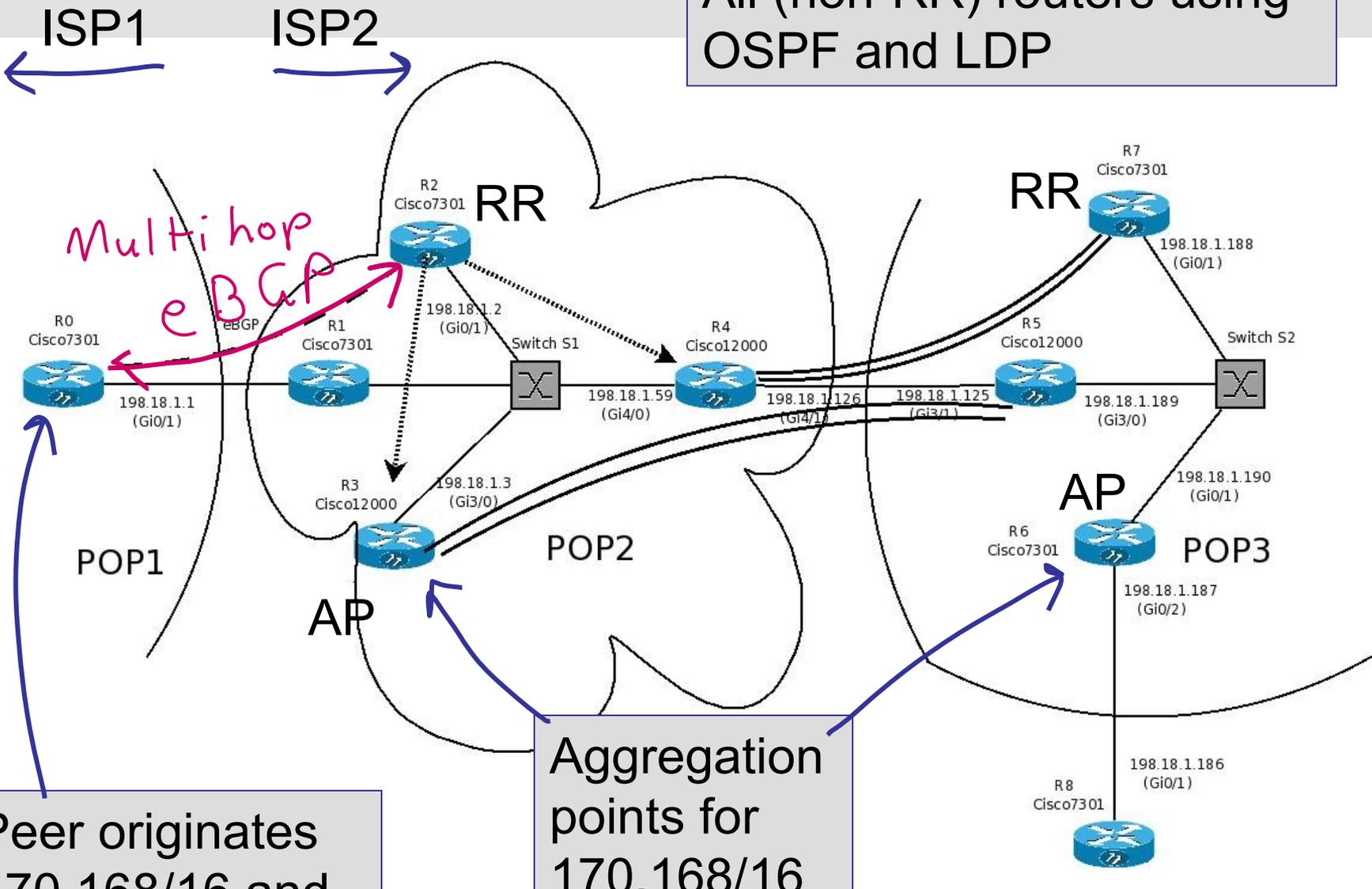
3: Static route (tunnel extension) to neighbor router

Route Reflectors (RR) filter out prefixes from neighbors and assign to aggregation point routers



We have a variant that does not require RR's

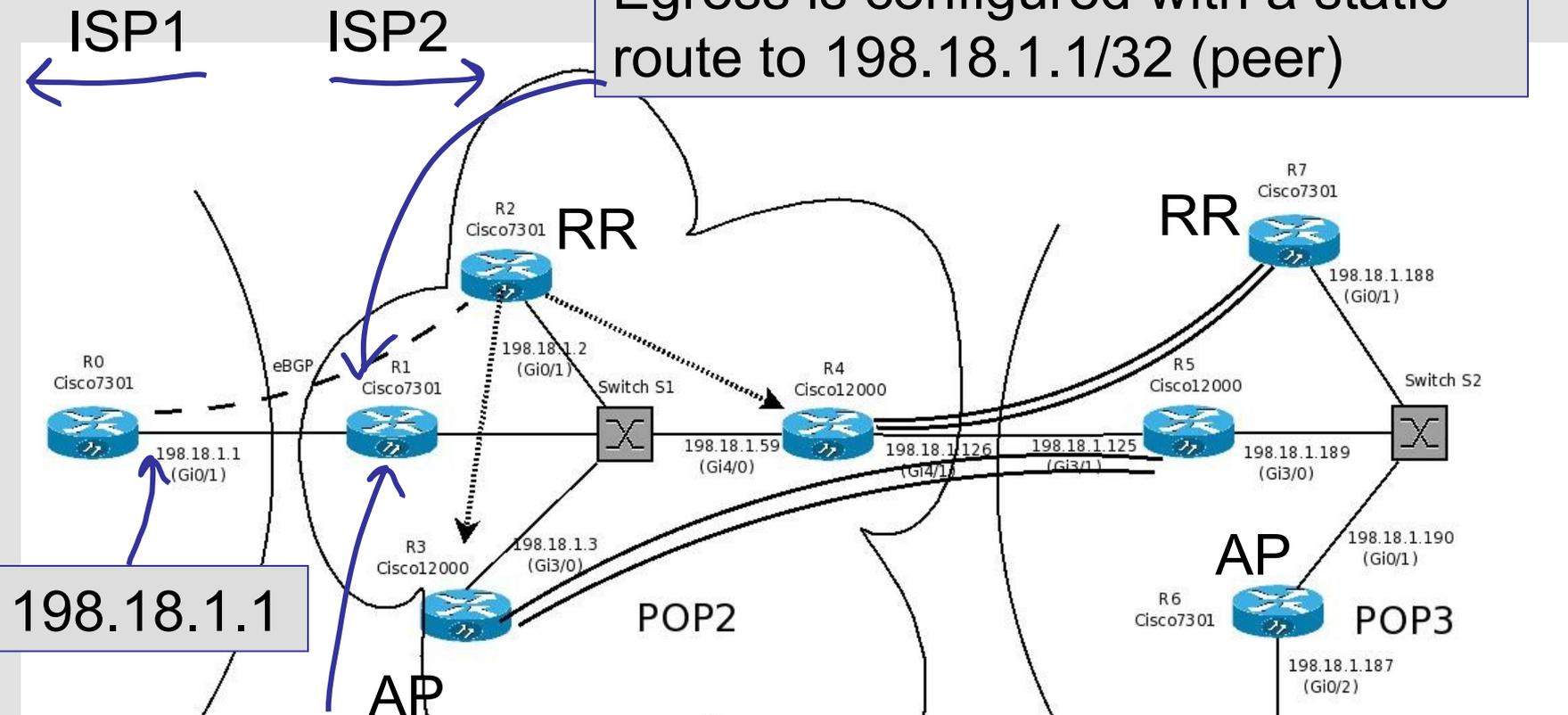
All (non-RR) routers using OSPF and LDP



Peer originates 170.168/16 and 203.250/16

Aggregation points for 170.168/16

Egress is configured with a static route to 198.18.1.1/32 (peer)



198.18.1.1

```
R1#show configuration | include ip route
ip route 198.18.1.200 255.255.255.255
GigabitEthernet0/2 198.18.1.200
```

With OSPF/LDP, all routers get MPLS tunnels to 198.18.1.1/32



ISP1

ISP2

170.168/16 and 203.250/16  
Next-hop = 198.18.1.1 (peer)

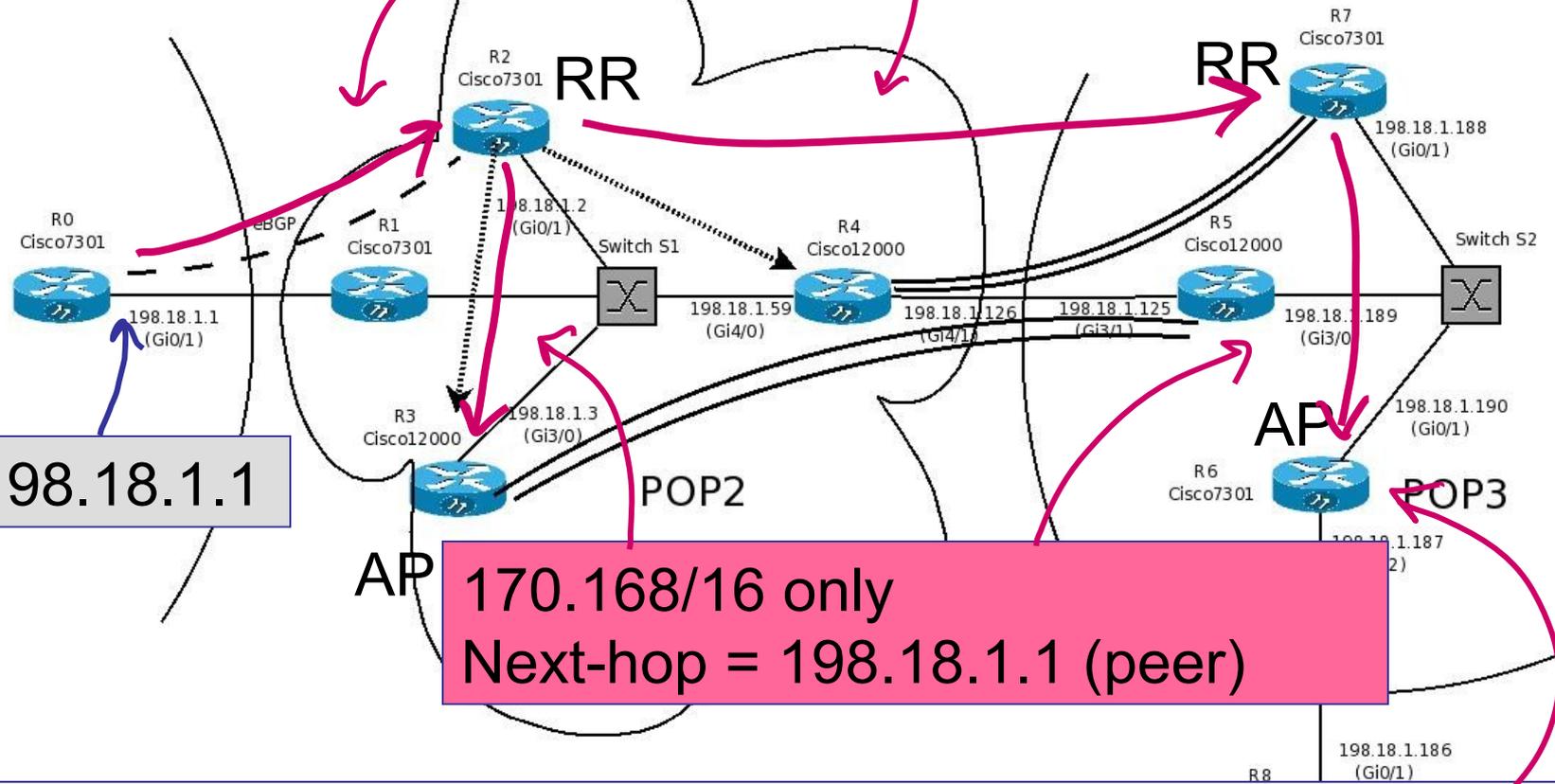
198.18.1.1

170.168/16 only  
Next-hop = 198.18.1.1 (peer)

```

R6#show ip bgp
BGP table version is 6, local router ID is 6.6.6.6
  Network          Next Hop           Metric LocPrf Weight Path
 *>i170.168.0.0    198.18.1.200       0      100      0 1 i

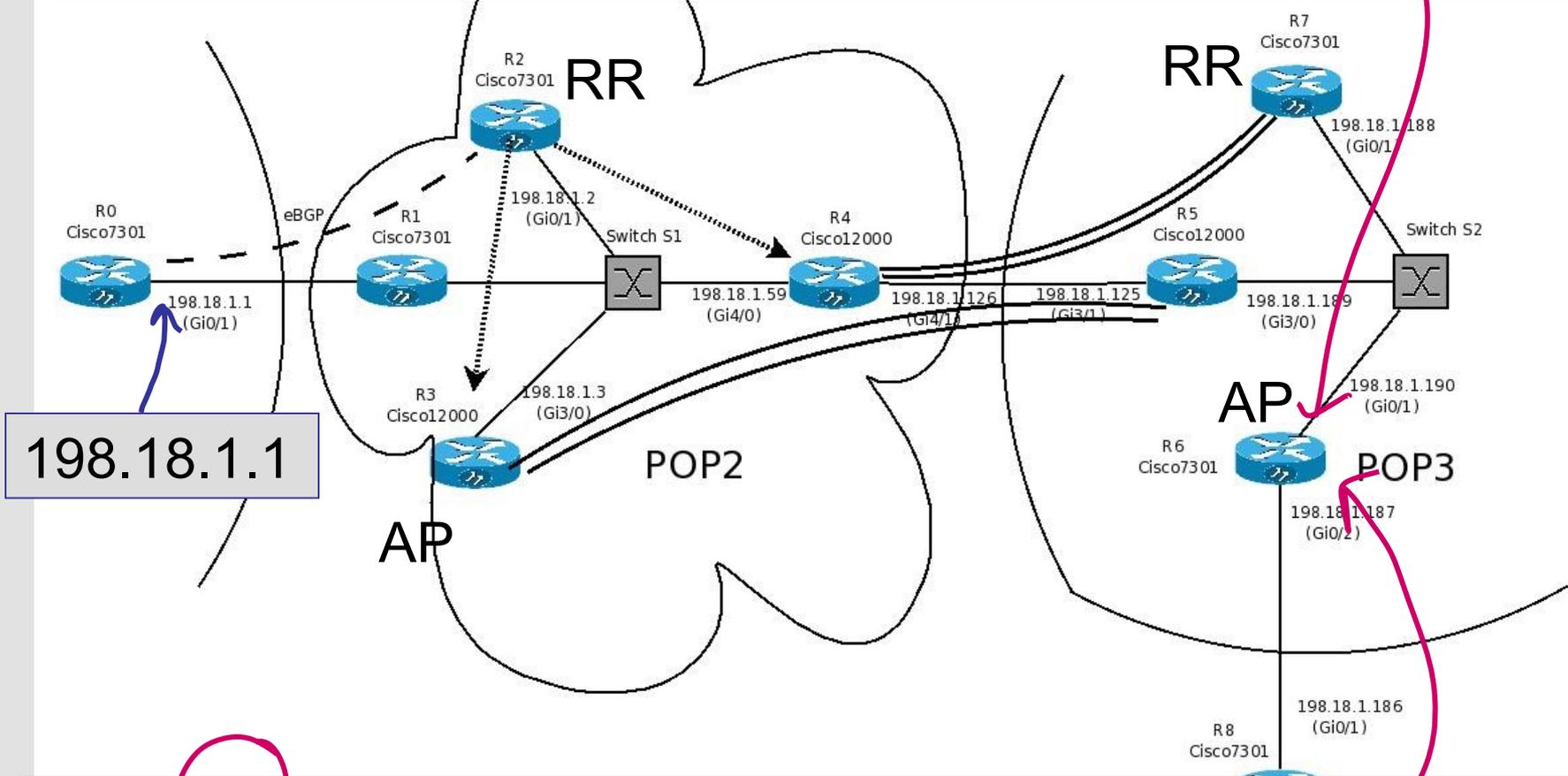
```



```
R6#show ip bgp
```

```
BGP table version is 6, local router ID is 6.6.6.6
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*>i170.168.0.0	198.18.1.200	0	100	0	1 i

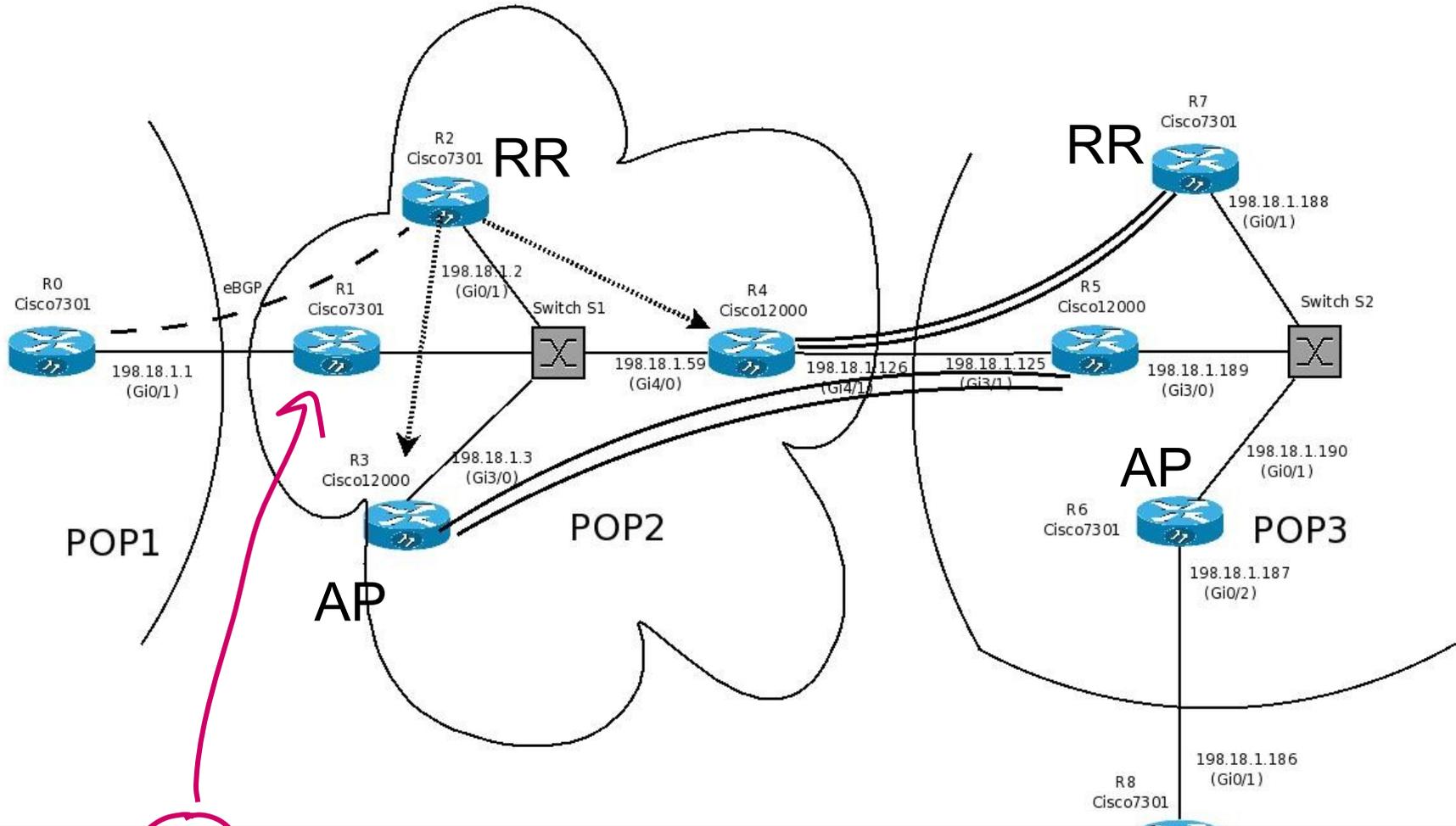


```
R6#show mpls forwarding-table
```

Local tag	Outgoing tag or VC	Prefix or Tunnel Id	Bytes switched	Outgoing interface	Next Hop
26	26	198.18.1.200/32	0	Gi0/1	198.18.1.189

ISP1

ISP2



```
R1#show mpls forwarding-table
```

Local tag	Outgoing tag or VC	Prefix or Tunnel Id	Bytes switched	tag	Outgoing interface	Next Hop
27	Untagged	198.18.1.200/32	31518		Gi0/2	198.18.1.200

# Operation without RR's

- Cisco has a config trick whereby entries in the RIB are not installed in the FIB
  - Set admin-distance to 255
- Using this trick, each router selectively filters what goes into the FIB
- Tested at scale with failover

# Additional Load and Latency

- Paths can be longer with Virtual Aggregation
- More or less, depending on location of Aggregation Point

# Minimizing Overhead

Traffic volume follows a power-law distribution

95% of traffic goes to 5% of prefixes

This has held up for years

Install “Popular Prefixes” in routers

On a per-POP or per-router basis

Different POPs have different popular prefixes

Popular prefixes are stable over weeks

# Performance Study

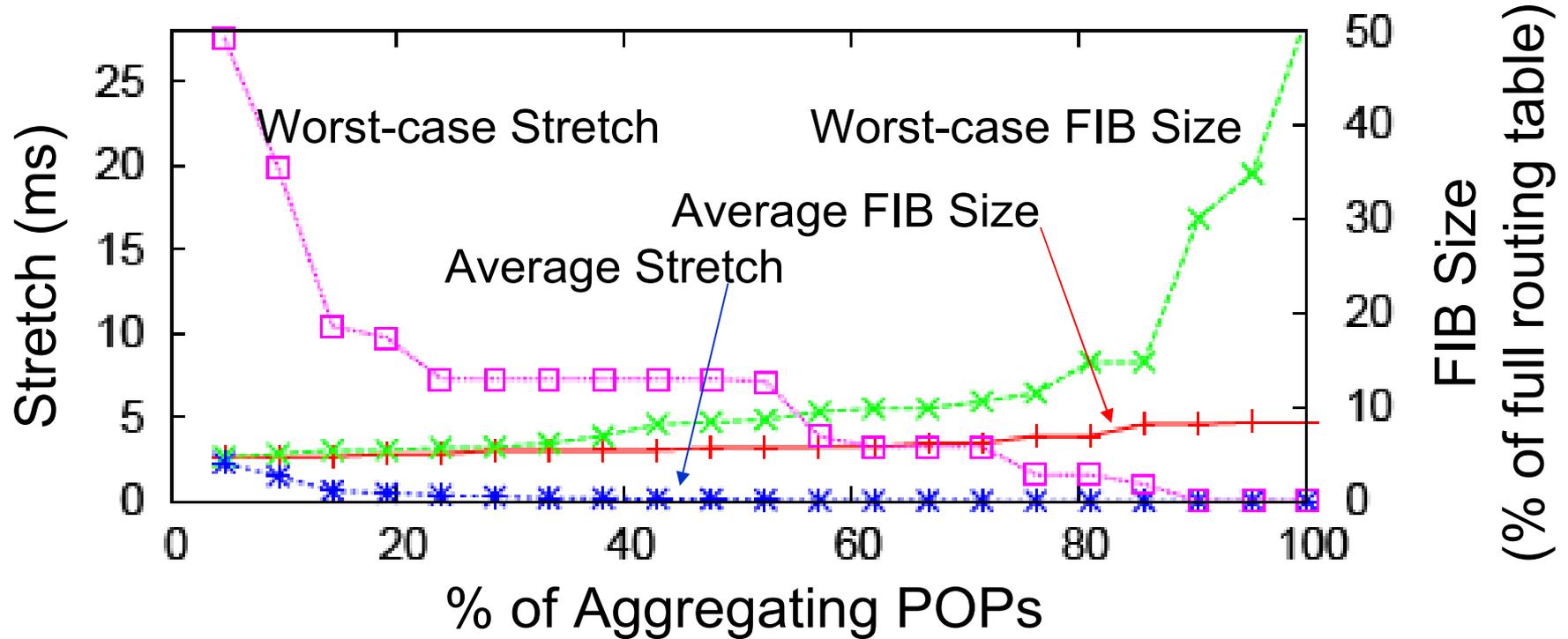
Data from a large tier-1 ISP

Topology and traffic matrix

Vary number of Aggregation Points (AP)  
and number of popular prefixes

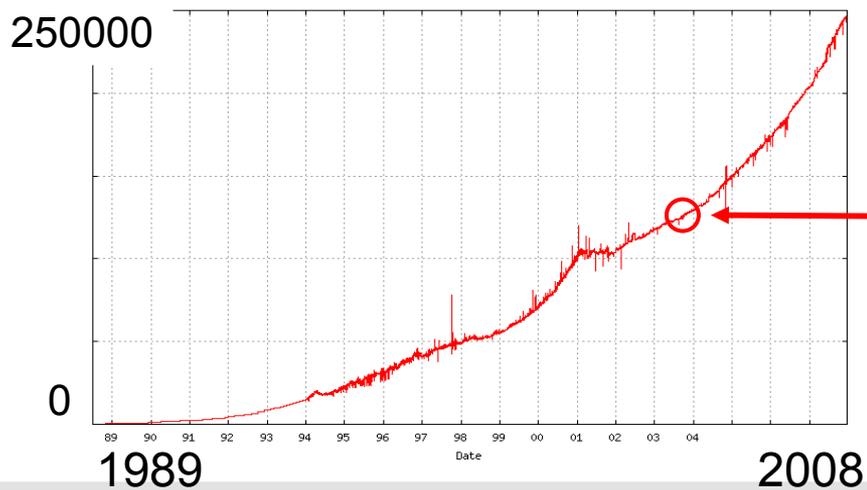
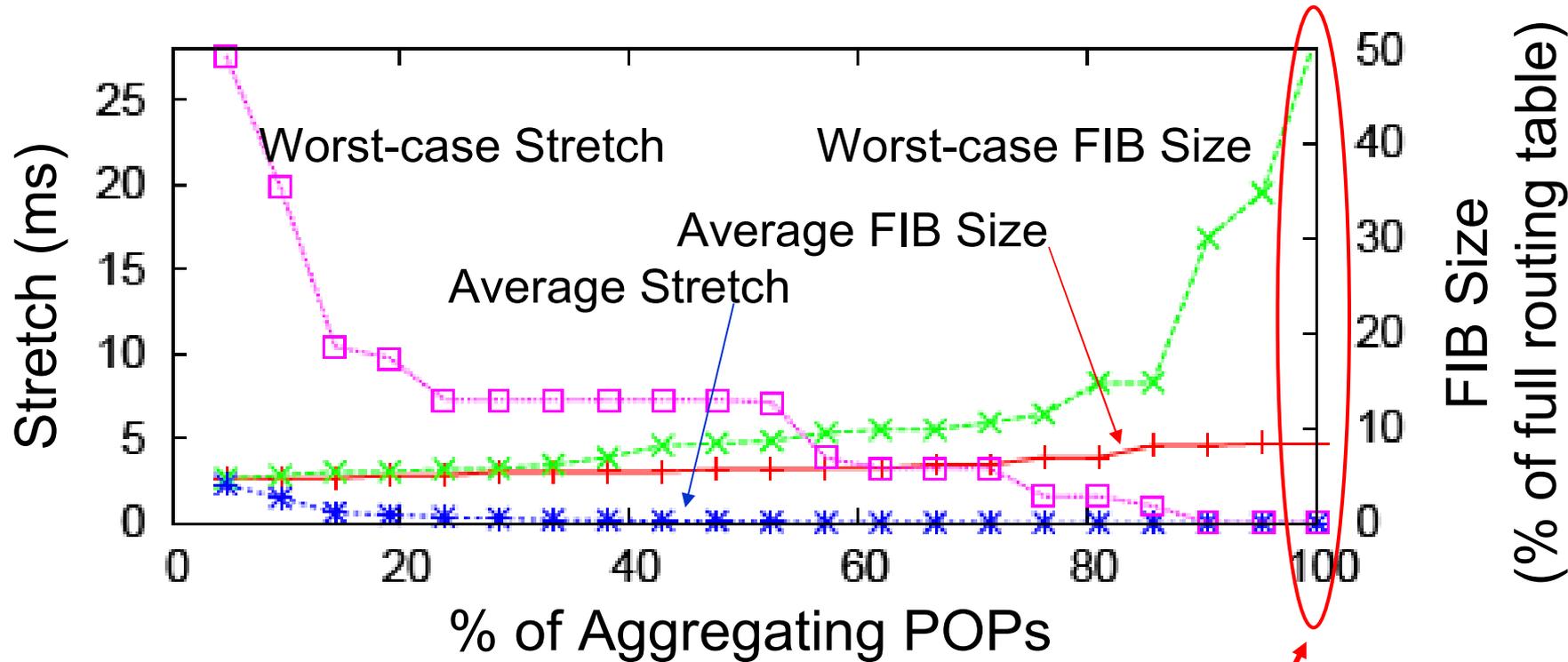
Naive AP deployment: A POP has either  
(redundant) AP's for all virtual prefixes, or  
no virtual prefixes

Naive popular prefixes deployment: same  
popular prefixes in all routers

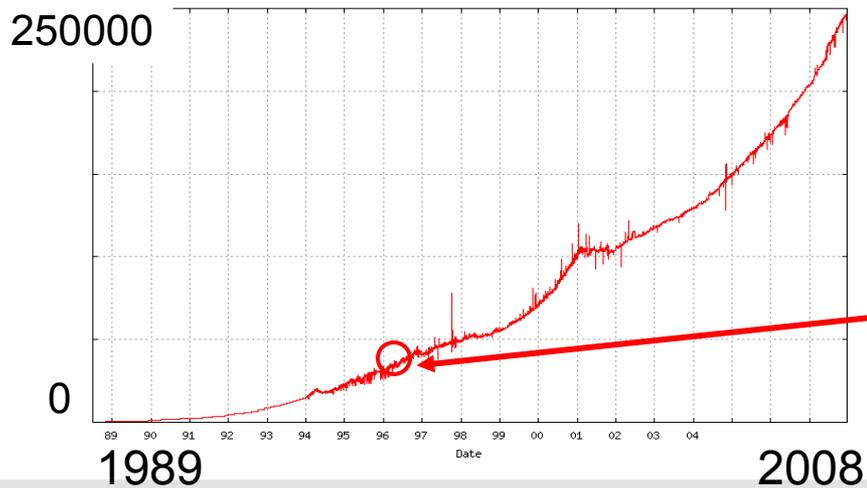
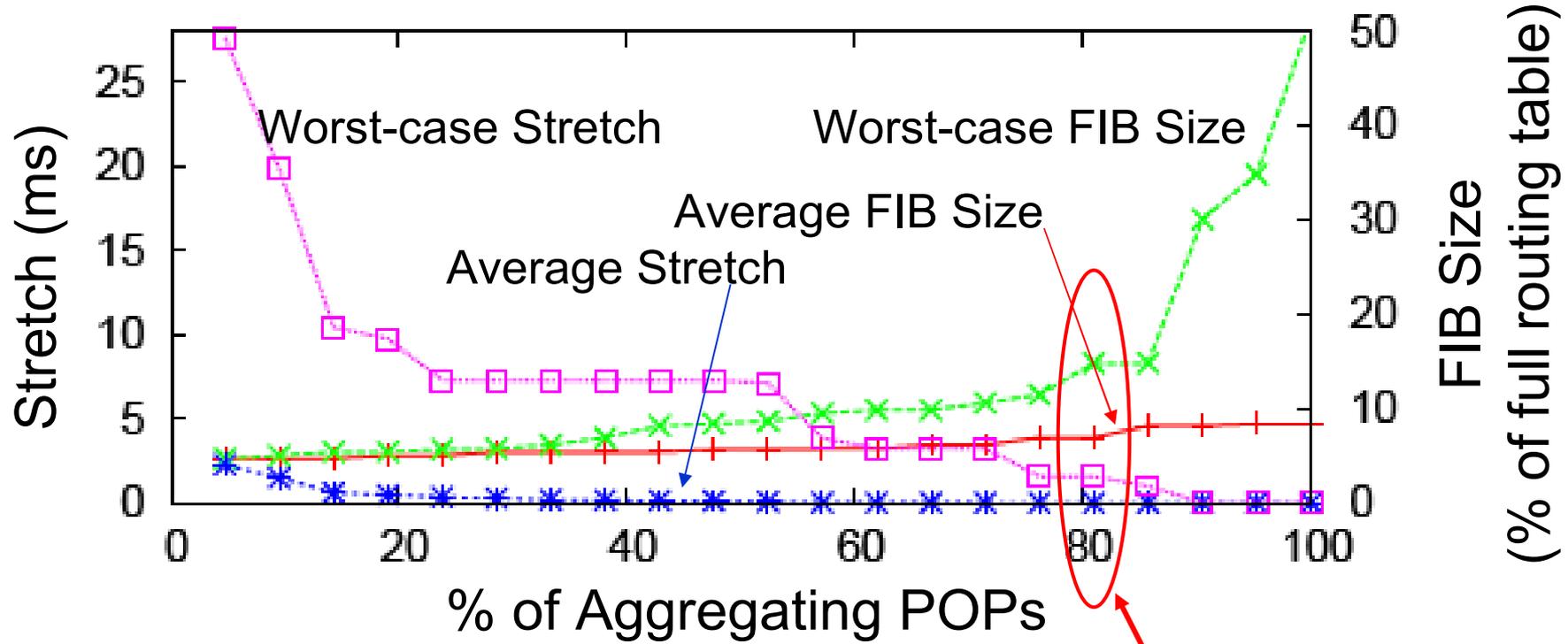


Install 1.5% of popular prefixes in all routers

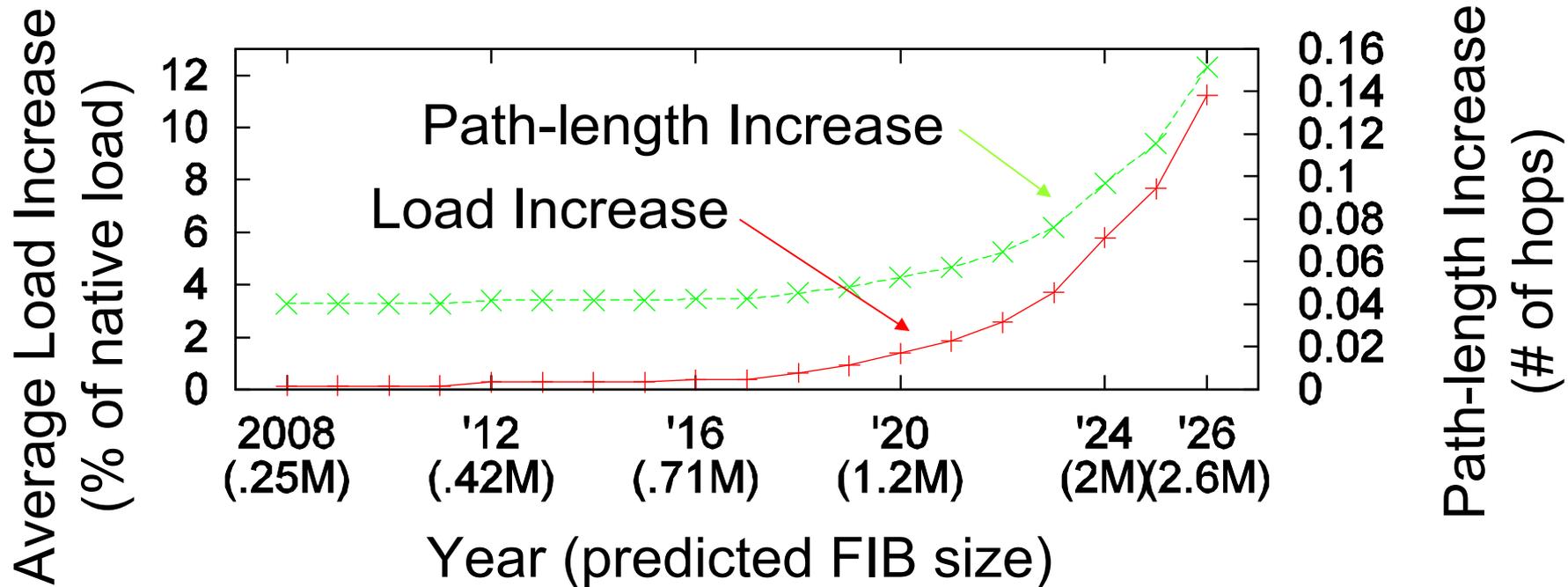
Stretch versus FIB size



Cuts FIB in half (2004 level), virtually no stretch

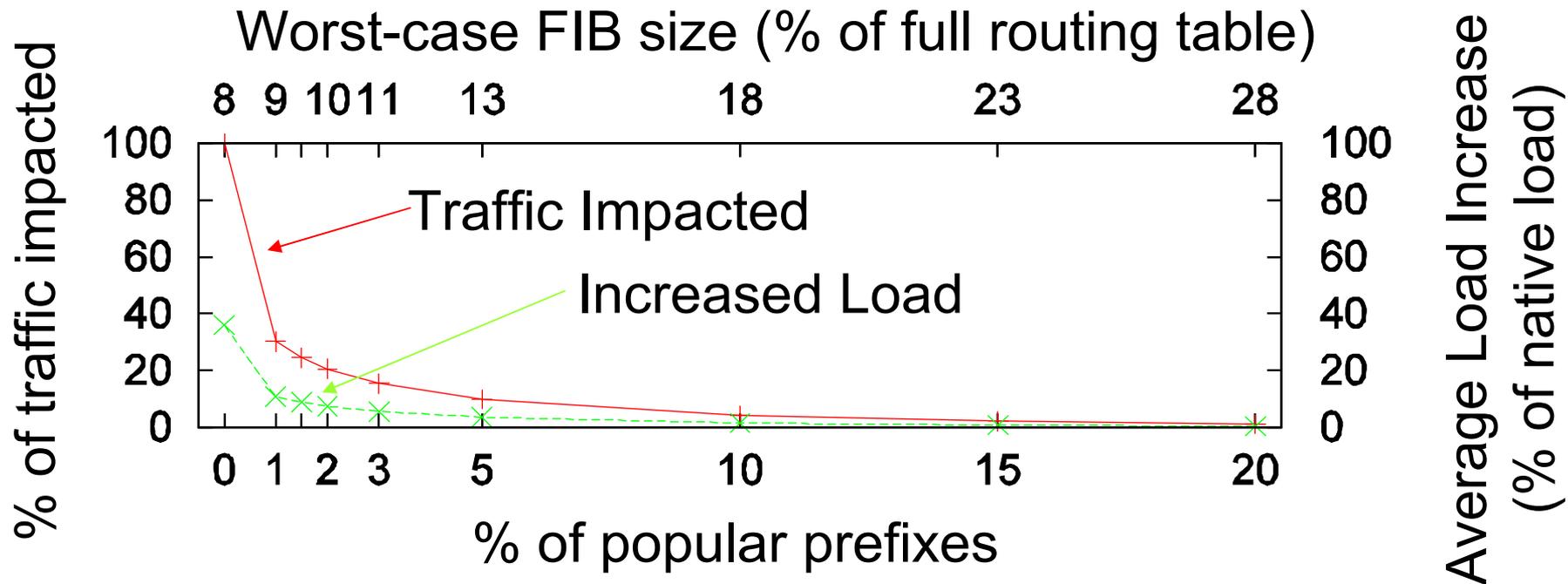


Cuts FIB five times (1996 level), worst case 2ms stretch



Assume 240K FIB entries (current routers)

Load and path-length over time



Roughly 50 % aggregating POPs

Load versus FIB size

# Next Steps

Build a “planning tool”

Determine best configuration

*Deploy on an ISP*

Work out cooperative ISP model

Eliminate need for full RIB anywhere

<http://www.cs.cornell.edu/people/francis/va-wp.pdf>