Cornell University

A configuration-only approach to shrinking FIBs

Prof Paul Francis (Cornell)

Virtual Aggregation

- An approach to shrinking FIBs (and RIBs)
  - In routers, not in route reflectors
- Works with legacy routers
  - New configuration only
- ISPs can independently and autonomously deploy

Outline

- Project status and immediate goals
- Mechanics
- Evaluation results

Status

- Tested a couple of versions of VA by configuring on Linux and Cisco routers
  - Simple, static, small-scale experiments (~10 routers)
  - Cisco 7301 and Cisco 12000
- Modeled using data from a large ISP
  - (router topology and traffic matrix)

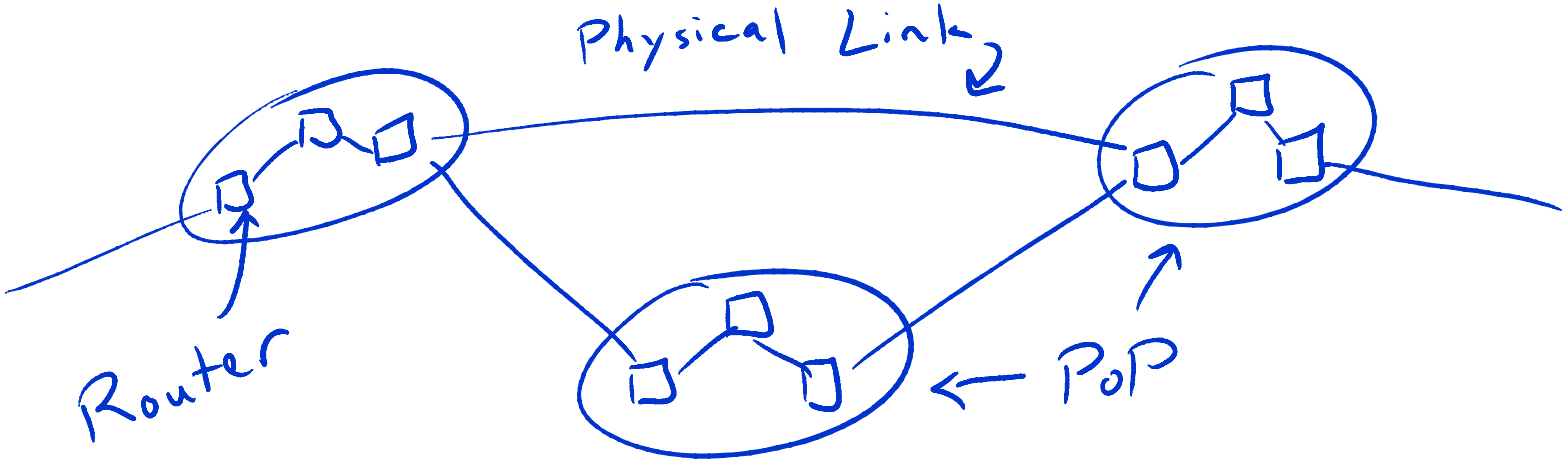
- Have *not* tested on a live network
- Have *not* tested dynamics
- Have *not* tested at large scale

- Cornell owns some IPR....

Goal of this talk

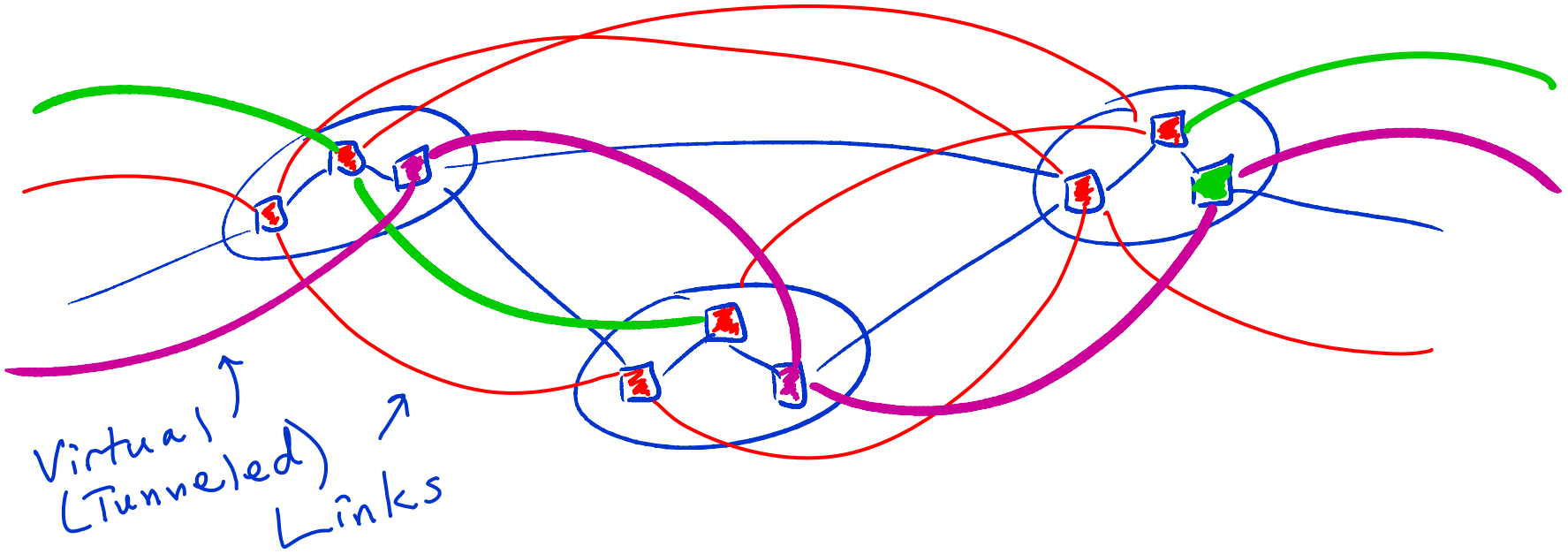
- There are a number of variants of Virtual Aggregation
- I'm looking for a few router management experts to help design the best variant
  - MPLS, route reflectors, Ethernet, filters, aggregation, ...

Virtual Aggregation: Basic Idea



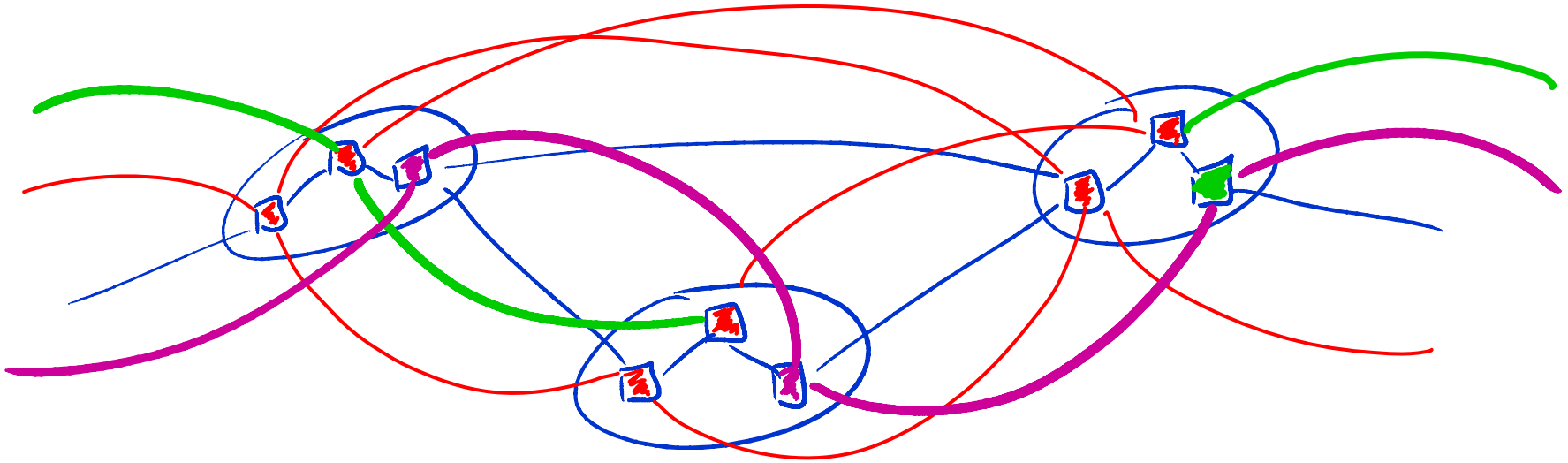
- Goal is to partition the DFZ table among existing FIBs

Virtual Aggregation: Basic Idea



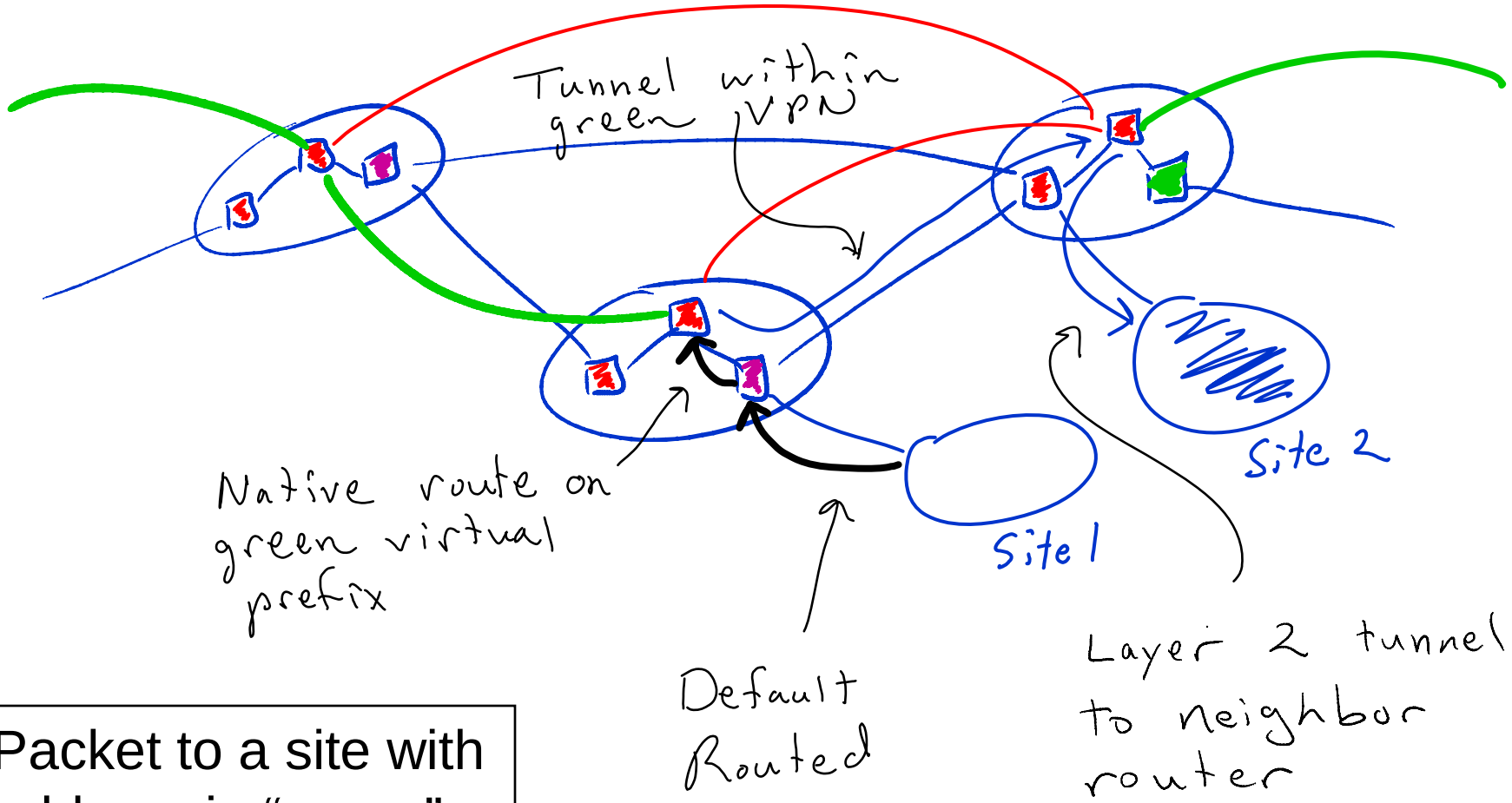
- Divide IP address space into “virtual prefixes” (say /7’s)
- Operate each virtual prefix as a “VPN”
- Assign different routers to different “VPNs”
  - Or even different physical FIBs within a router

Virtual Aggregation: Basic Idea



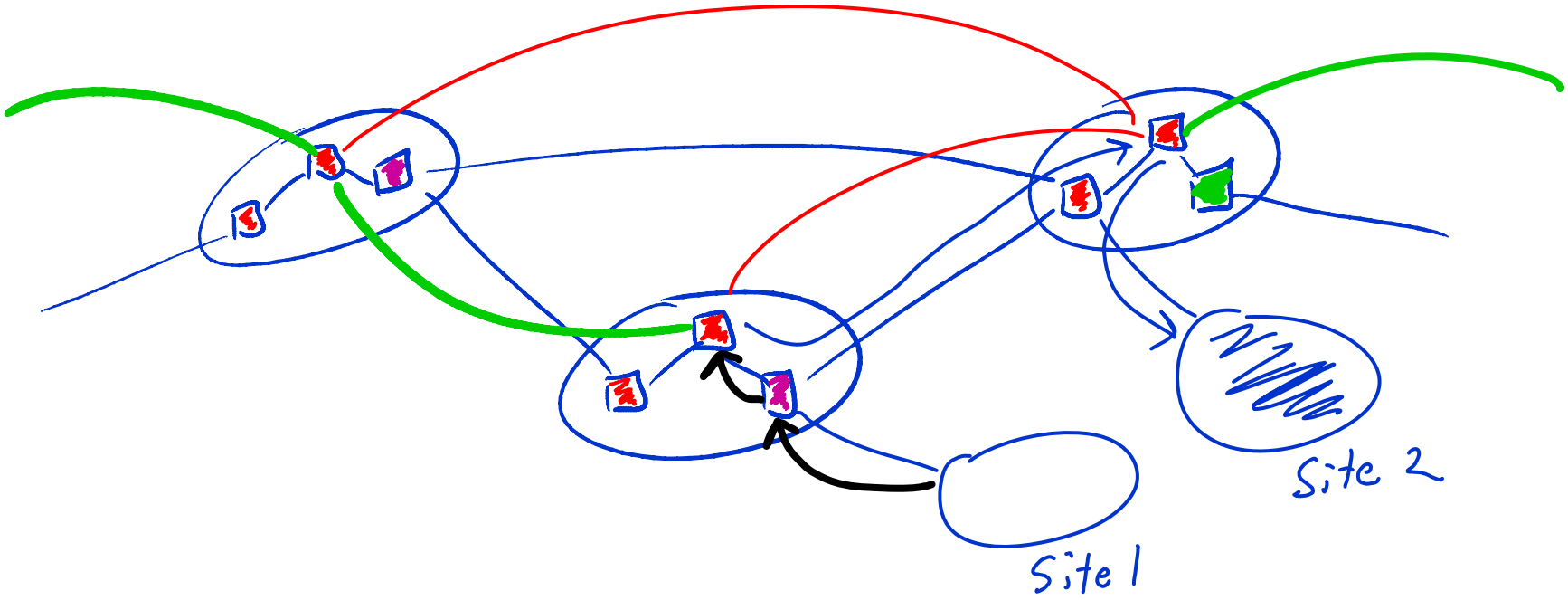
- Each router then knows:
  - Routes to all sub-prefixes within its virtual network
  - Routes to all other virtual networks

Virtual Aggregation: Basic Idea



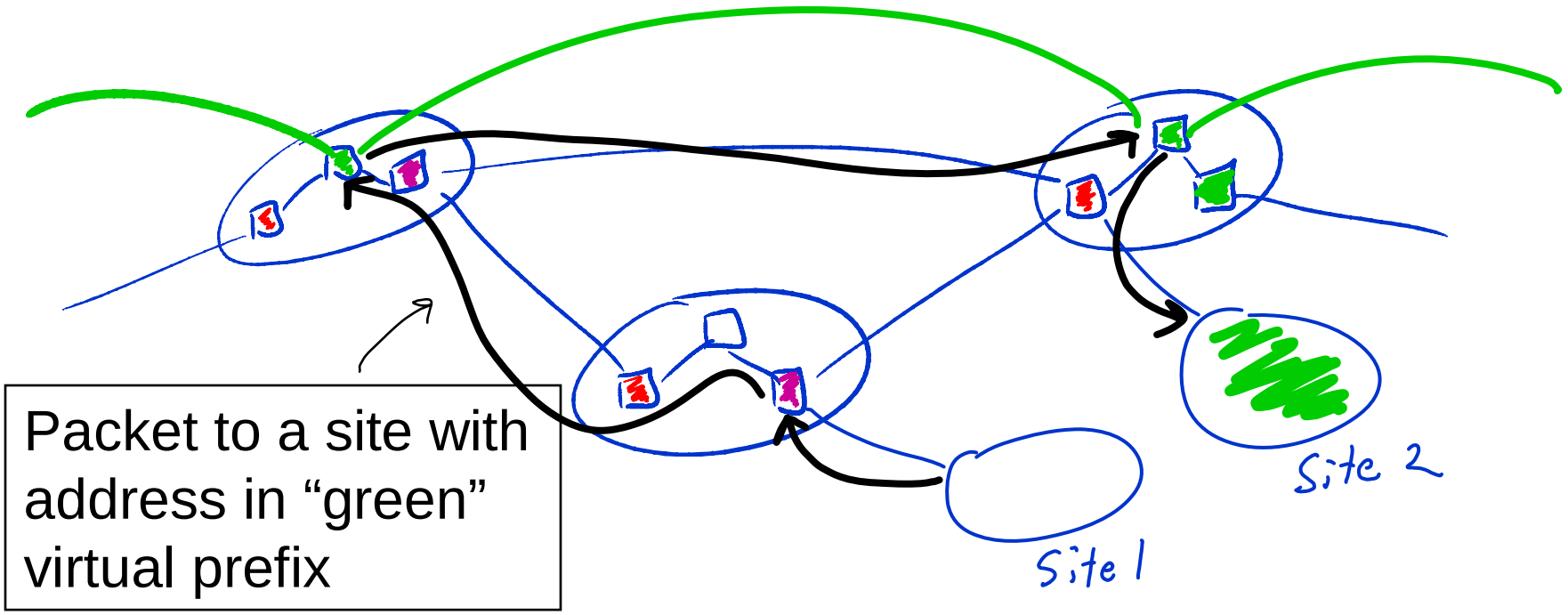
Packet to a site with address in "green" virtual prefix

Virtual Aggregation: Basic Idea



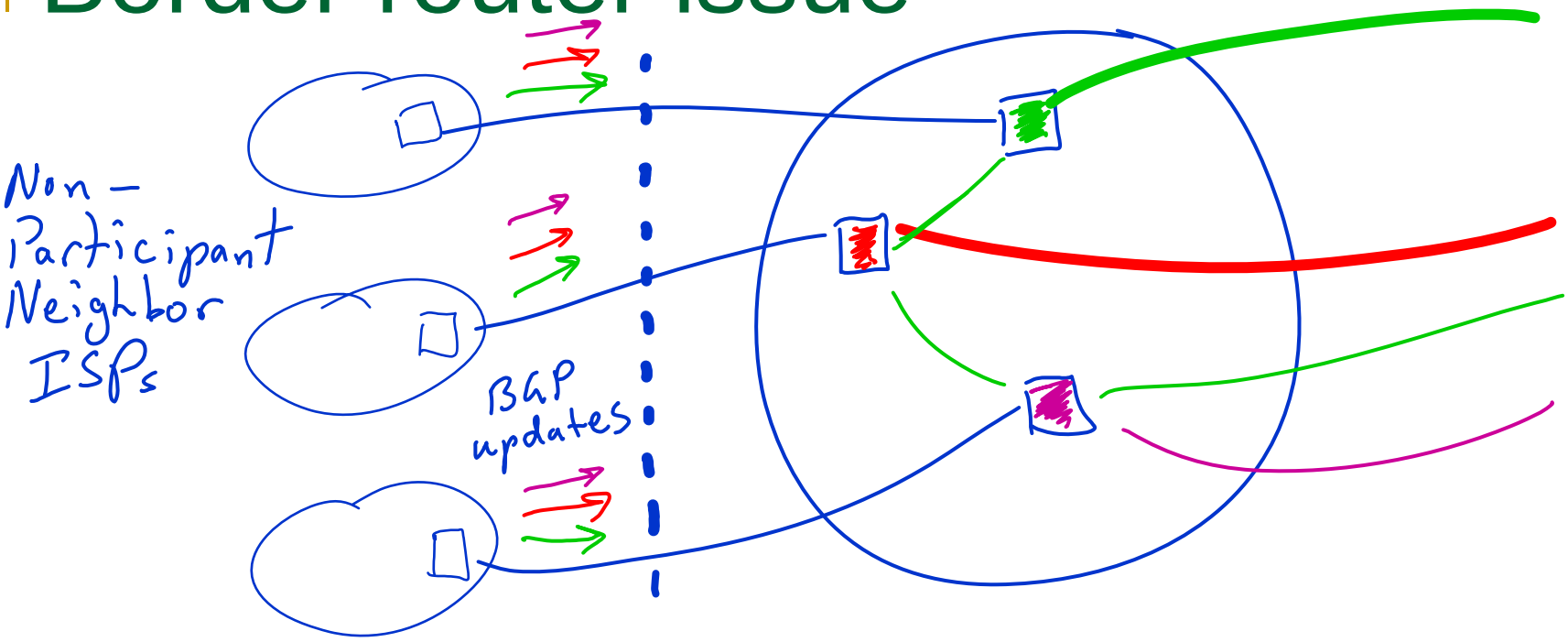
- Path length can increase
  - Not so bad if each virtual prefix has a member router in each POP
- Load can increase

Virtual Aggregation: Basic Idea



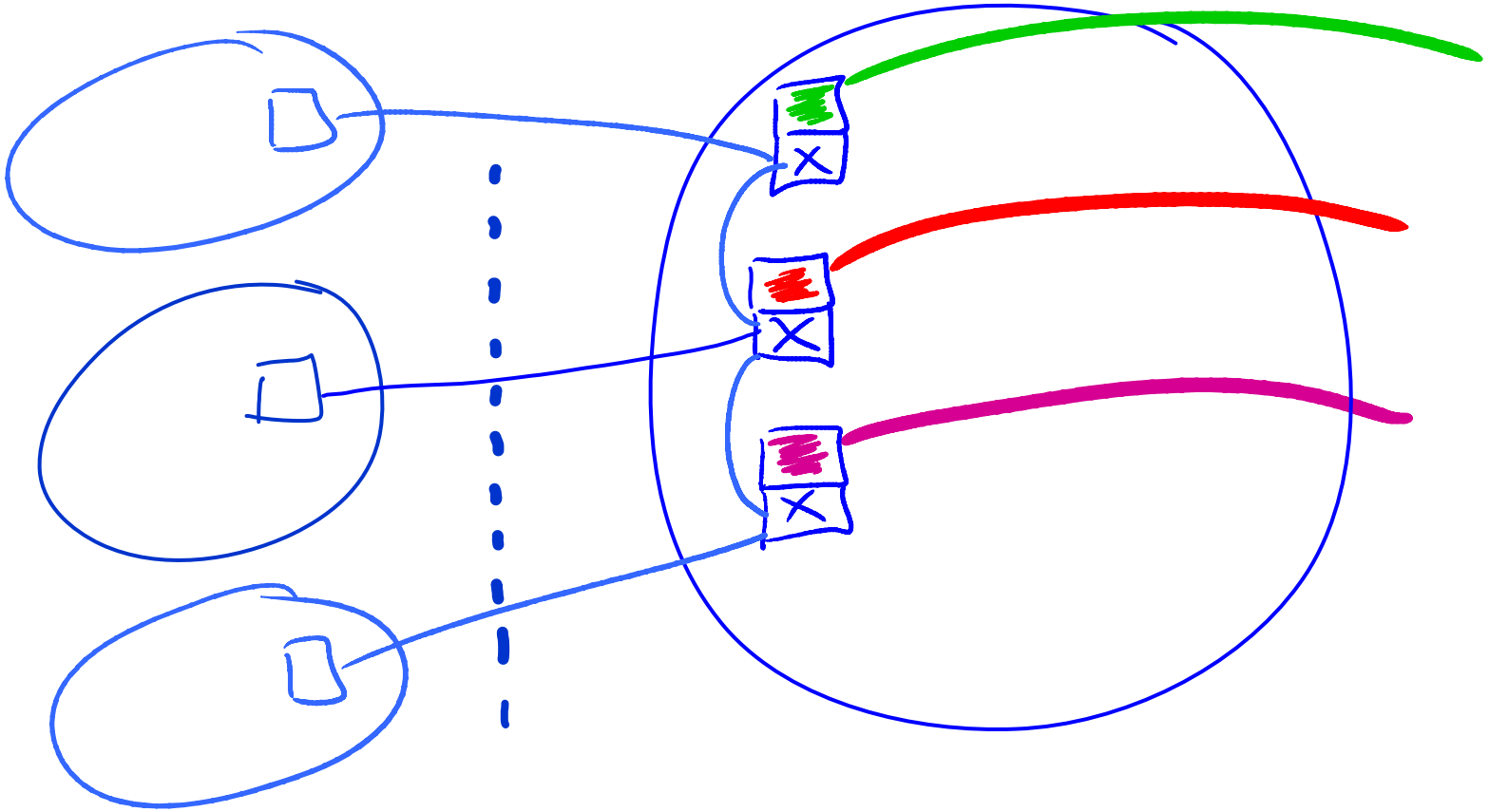
- Path length increase
- Can be significant if a POP does not have a member router for a given virtual prefix

Border router issue



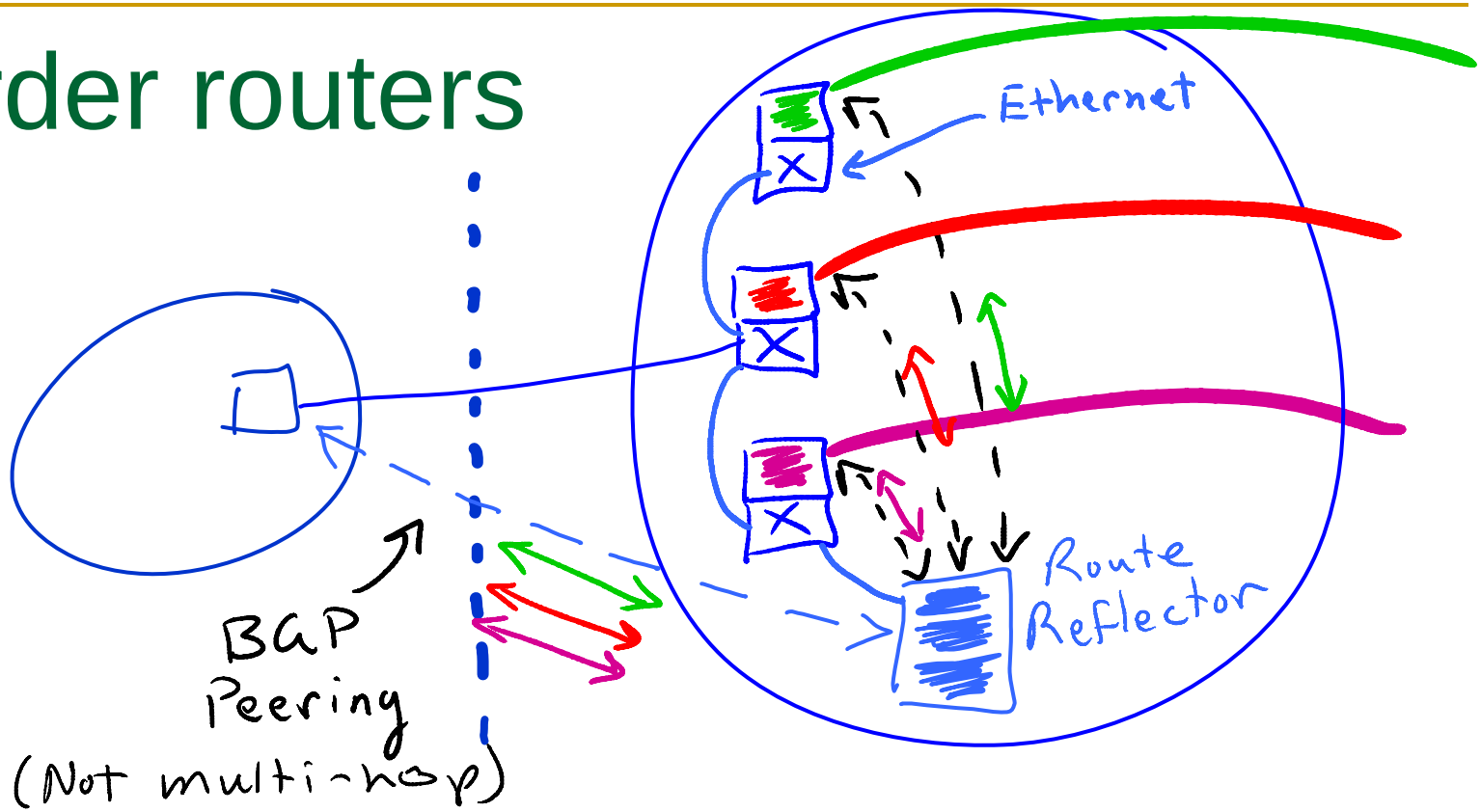
- Problem is that border routers need full routing tables to peer with non-participant neighbor ISPs

Border routers



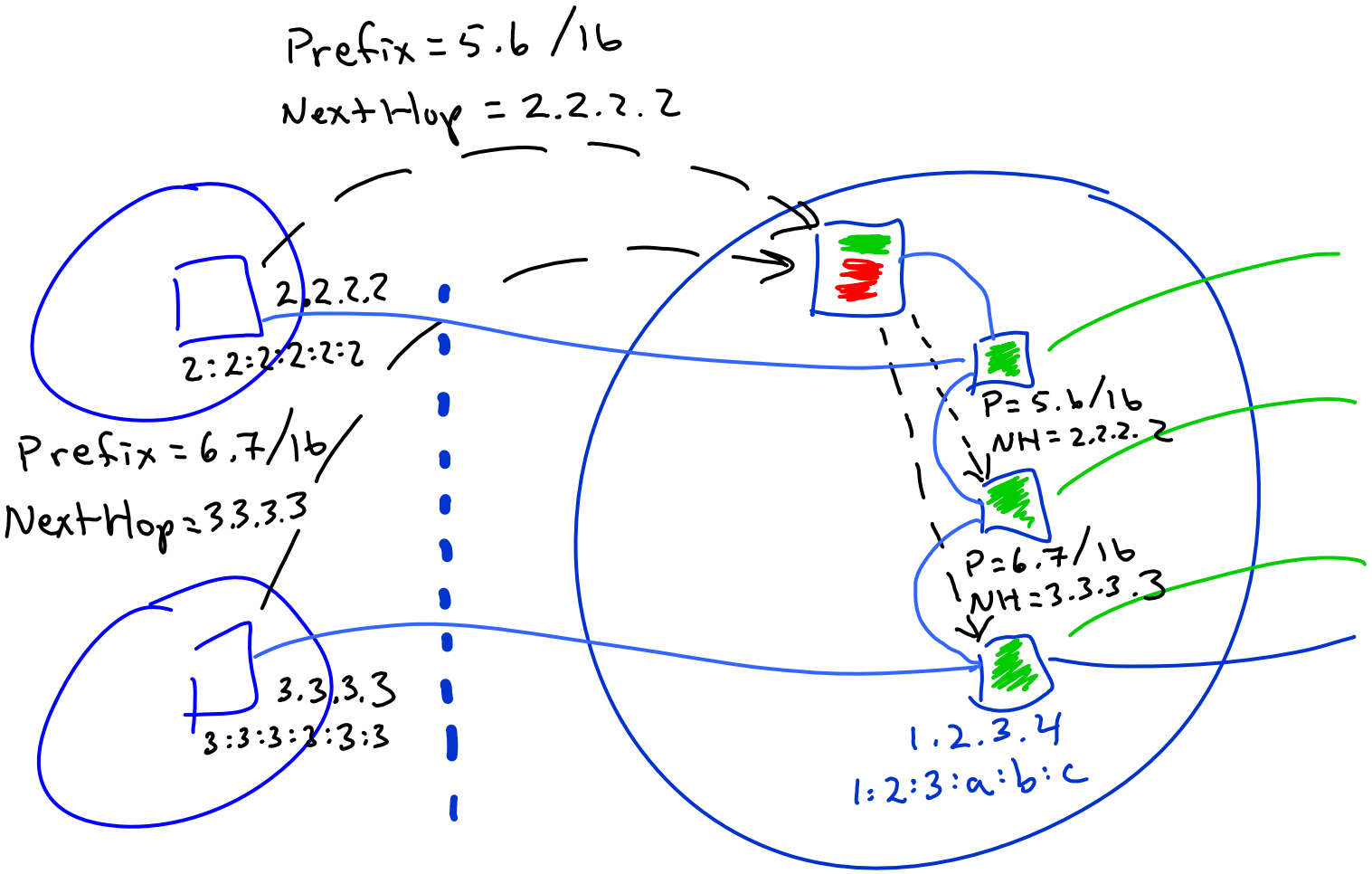
- We exploit the fact that routers can also operate as Layer-2 switches

Border routers

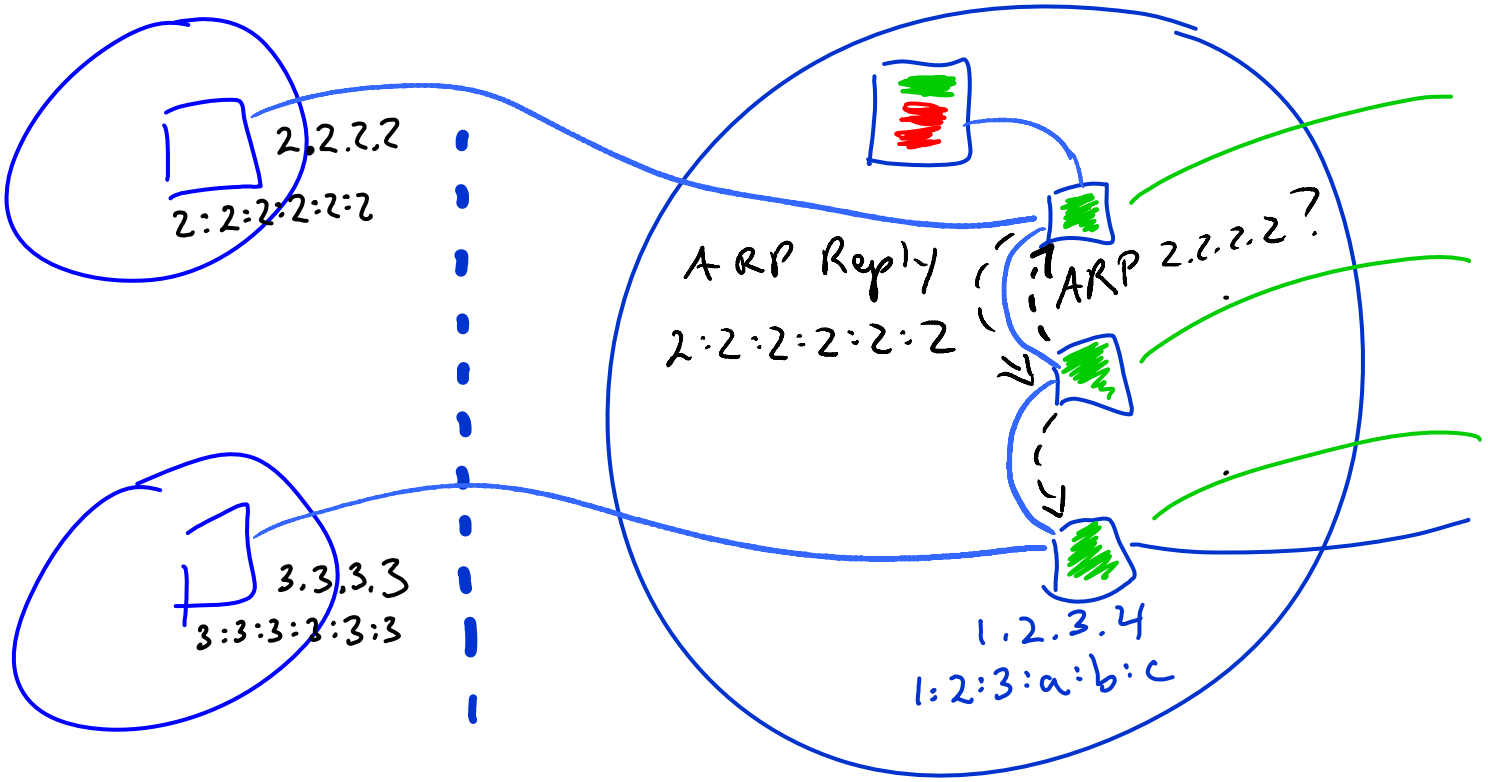


- Peer using a Route Reflector (RR), which aggregates on behalf of routers
- RR using BGP Next-Hop attribute to refer peer to the appropriate router
- Layer 2 is used to tunnel outgoing packets to neighbor router

Border routers



Border routers



Increase in path length and router load

- Increase in router load has two causes:
  - Increase in path length (router hops) means more traffic per router
  - For legacy routers, tunneling is a more expensive operation

- Results shown here for configuration with IP-in-IP tunnels at each PoP
  - In practice, will probably use MPLS from aggregating router to the egress
  - This will improve load numbers significantly

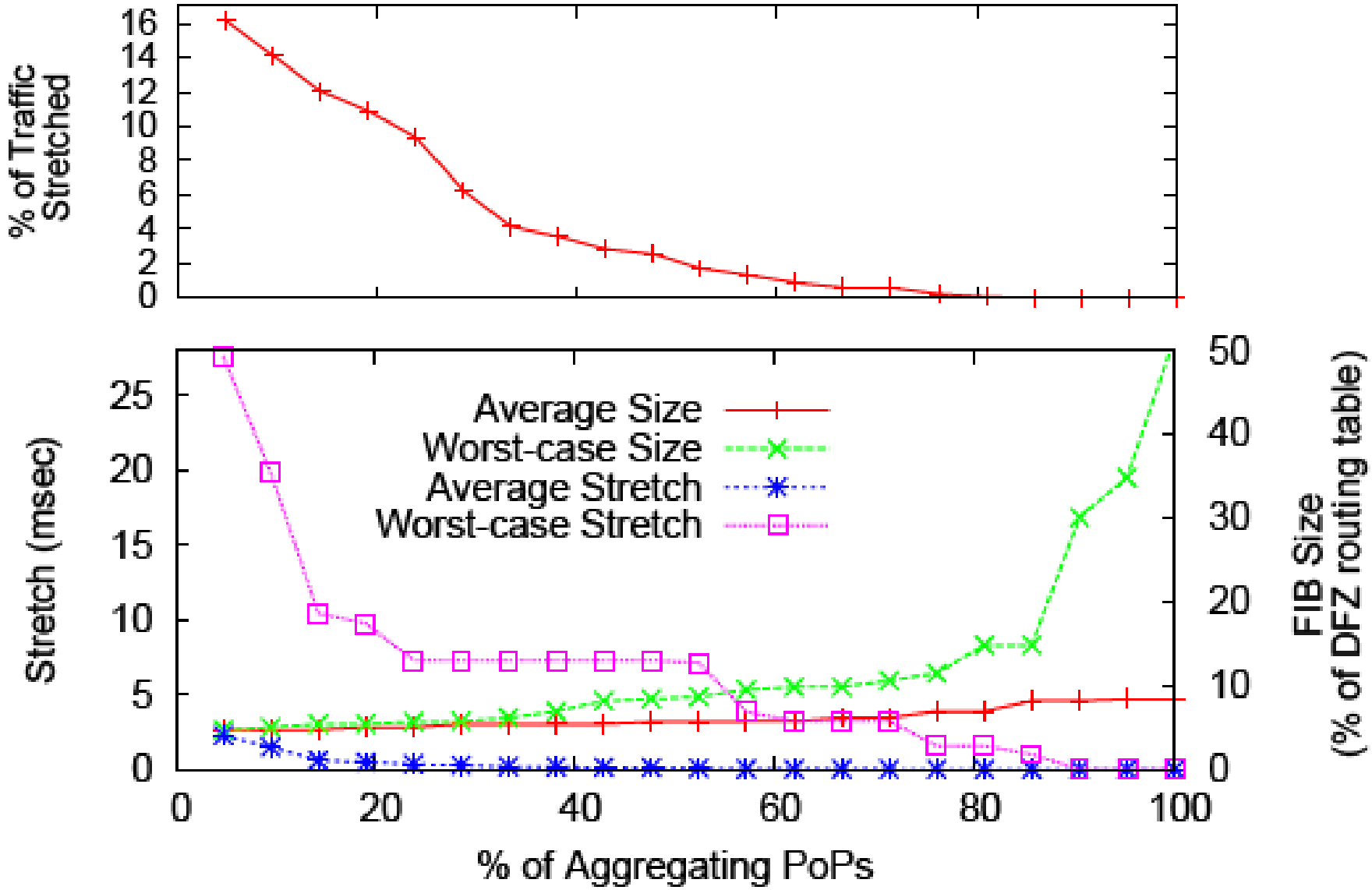
Path length / Router load solution

- Basic idea is to exploit the fact that traffic distribution follows a power law
  - 90% of traffic goes to 10% of destination prefixes
- Route packets for popular prefixes natively

- Monitor traffic matrix to find popular prefixes
- Periodically (weekly?) update aggregation filters to let popular prefixes slip through

Performance measurements

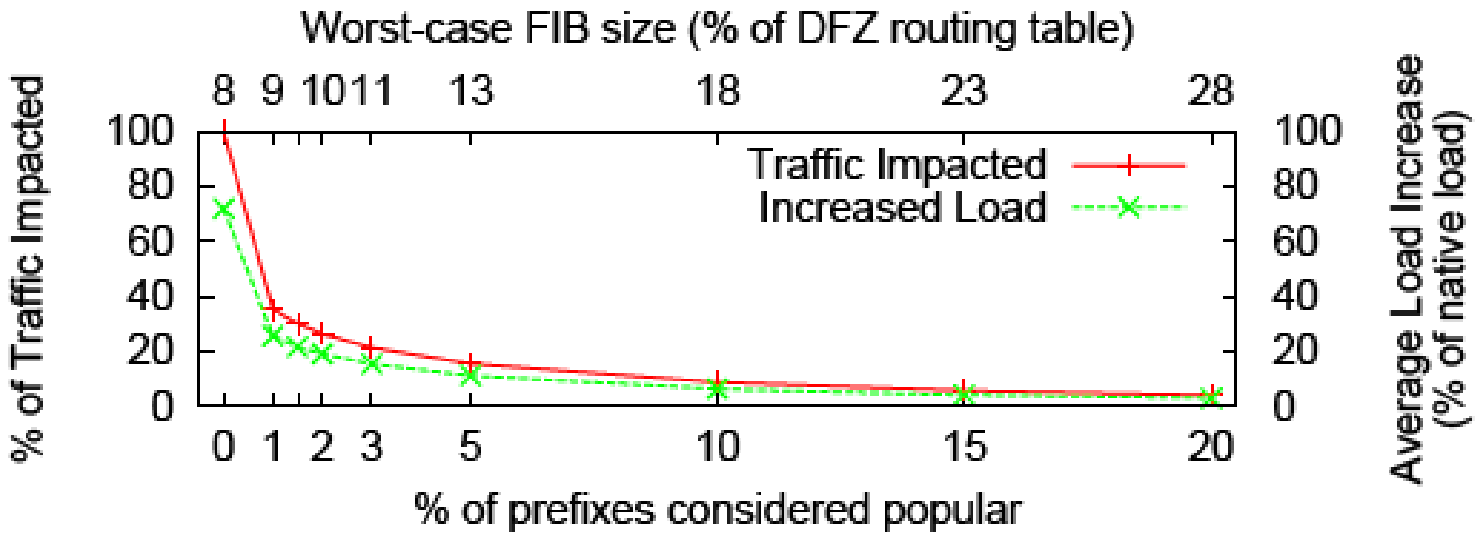
- Use data from a large tier-1 ISP ("BIG-ISP")
  - Router- and Pop-level topology, traffic matrix
- Define PoP as:
  - "Aggregating PoP": Has an aggregating router for each virtual prefix (two, actually)
  - "Non-aggregating PoP": No aggregating routers, only carry routes to virtual prefixes
- Control and measure:
  - % of natively routed prefixes (highest volume)
  - Stretch (in absolute terms, ms)
  - Increase in load
  - % of PoPs that are aggregating
  - FIB size

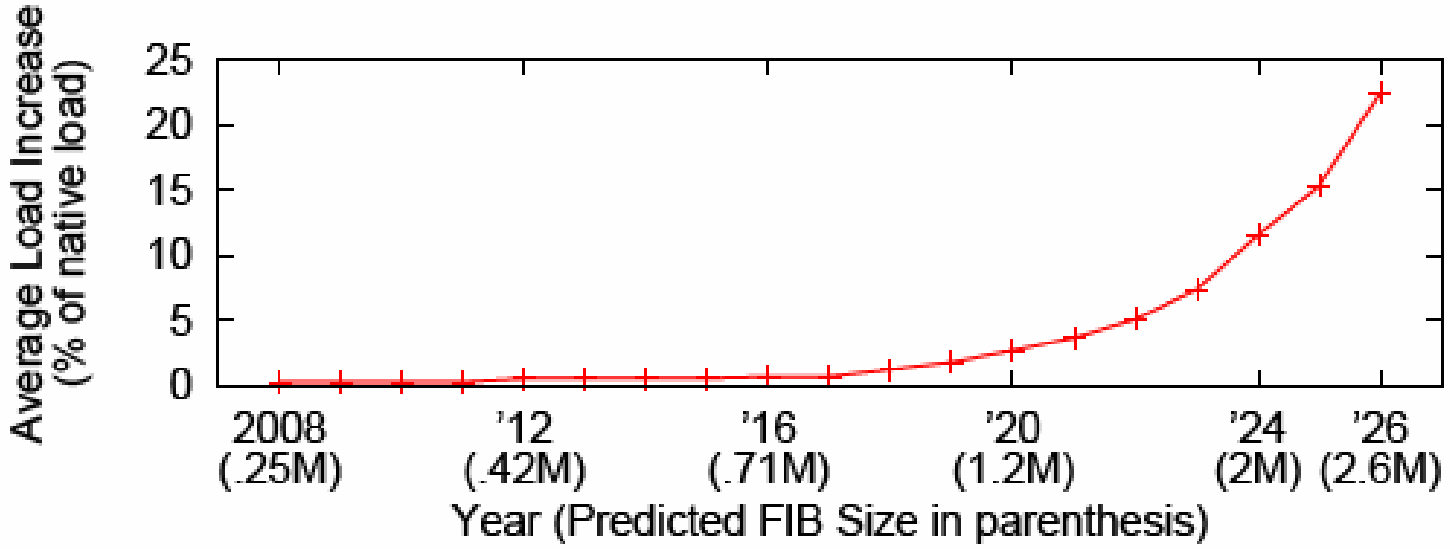


1.5% of prefixes are natively routed

One FIB per router

Load





(Note that traffic increases normally at roughly 50% to 100% per year)

Conclusions and future work

- Appears very promising
  - Big reduction in table size, buys years of continued growth
- Looking for participation
- Need to experiment with different, bigger, and more dynamic configurations
- Need to build a “planning tool”:
  - Input = traffic engineering data
  - Output = specific configurations and performance estimates