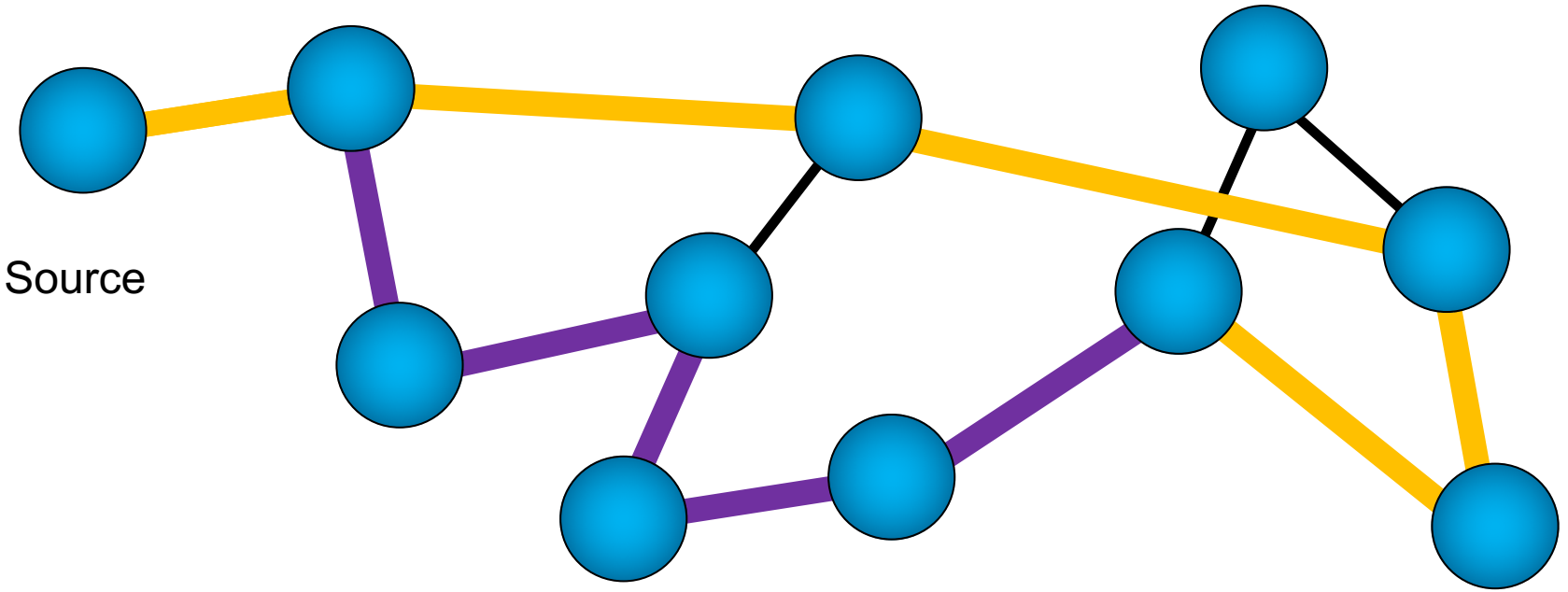
Tunnels, NFV, and Middleboxes

Craig A. Shue, Ph.D.
Associate Professor

Tunneling

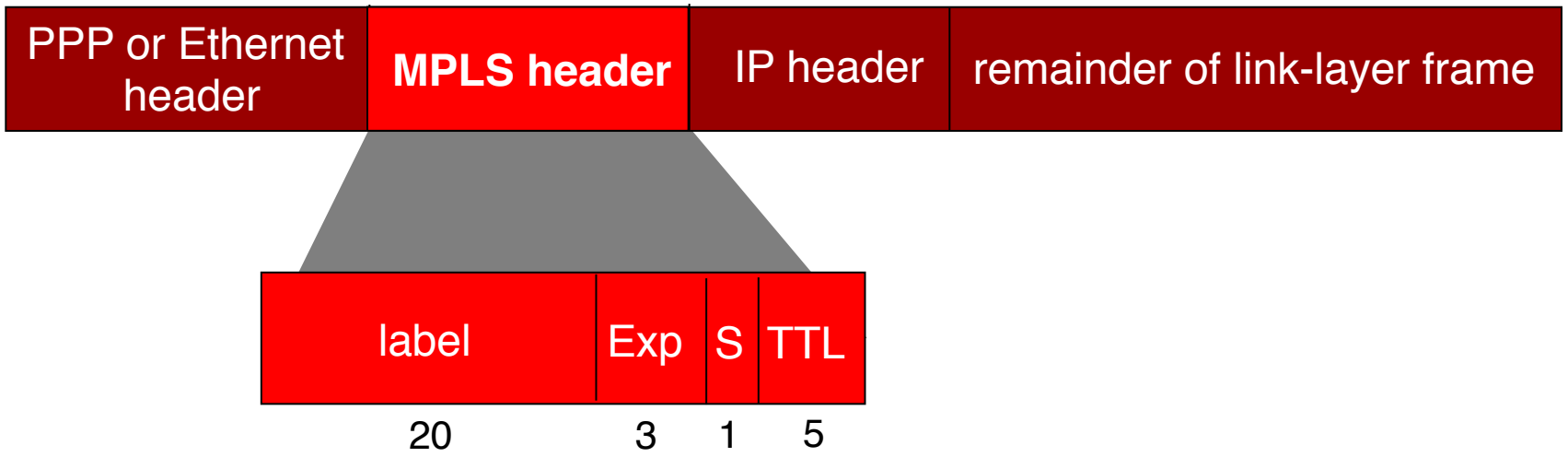
- The notion of routing traffic between two points independently of the packet's source or destination



Traffic Engineering with MPLS In the Internet

MultiProtocol Label Switching (MPLS)

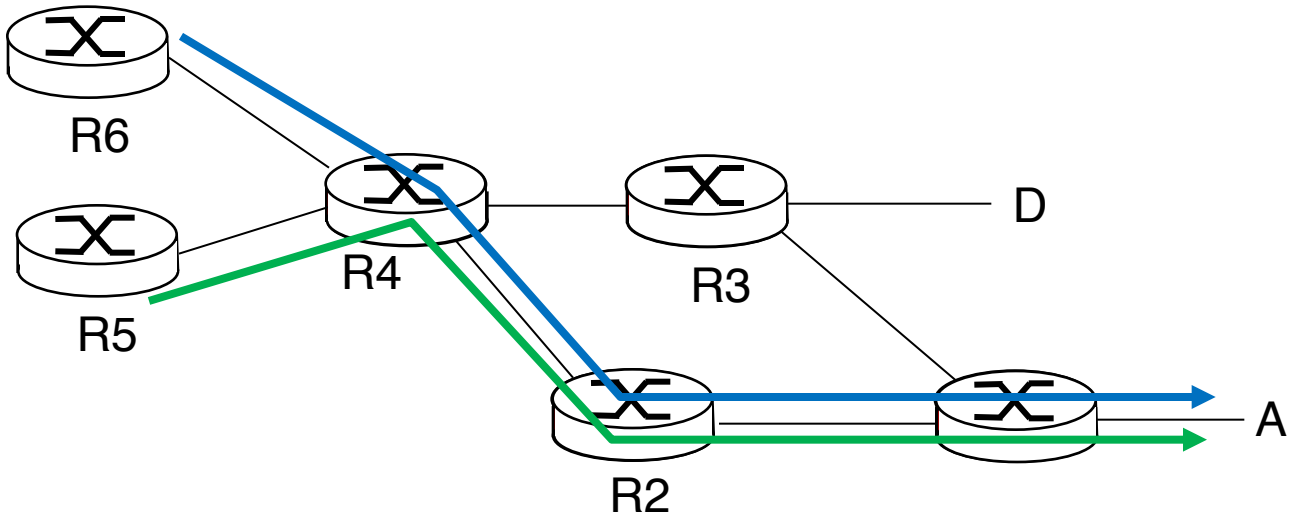
- Initial goal: high-speed IP forwarding using fixed length label (instead of IP address)
  - Fast lookup using fixed length identifier (rather than shortest prefix matching)
  - Borrowing ideas from Virtual Circuit (VC) approach
  - IP datagram still keeps IP address



MPLS Capable Routers

- Called label-switched router
- Forward packets to outgoing interface based only on label value (*don't inspect IP address*)
  - MPLS forwarding table distinct from IP forwarding tables
- *Flexibility:* MPLS forwarding decisions can *differ* from those of IP
  - use destination *and* source addresses to route flows to same destination differently (traffic engineering)
  - re-route flows quickly if link fails: pre-computed backup paths (useful for VoIP)

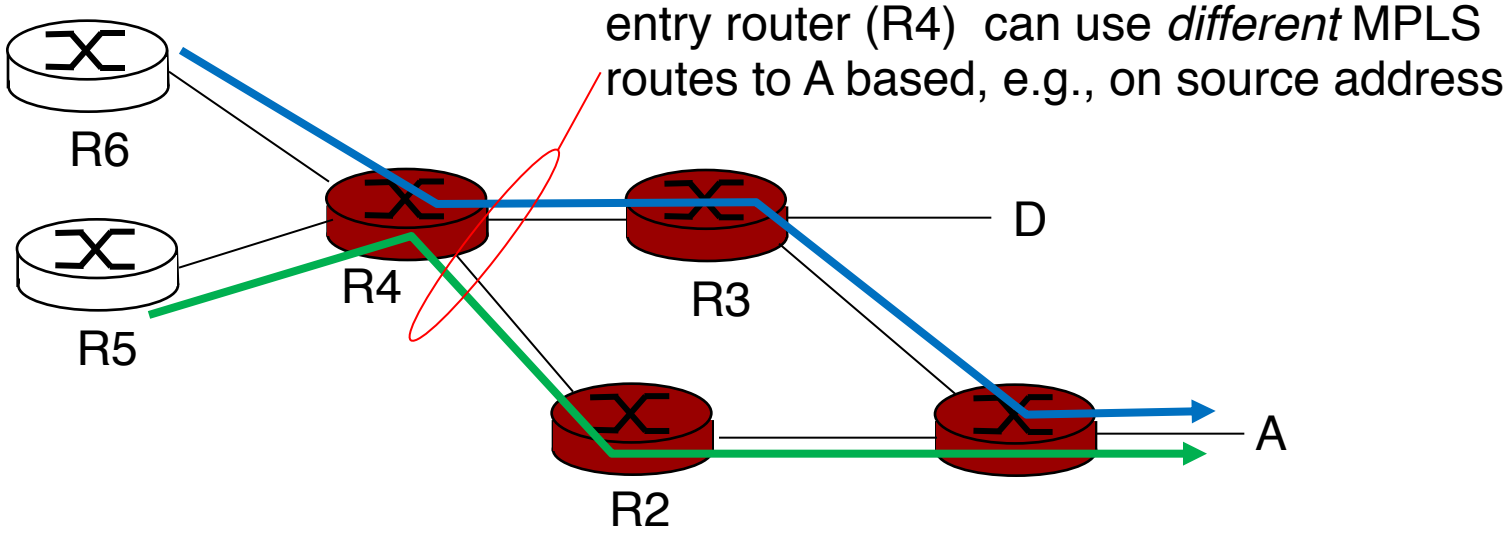
MPLS versus IP paths



- *IP routing: path to destination determined by destination address alone*

IP router

MPLS versus IP paths



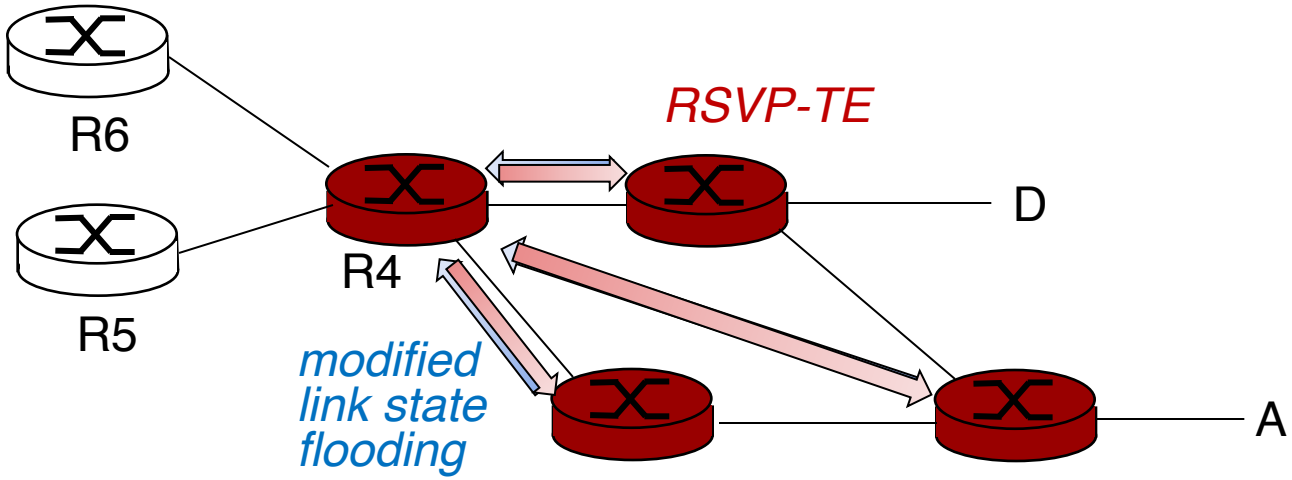
- *IP routing: path to destination determined by destination address alone*
- *MPLS routing:* path to destination can be based on source *and* dest. address
  - *fast reroute:* precompute backup routes in case of link failure

IP-only router

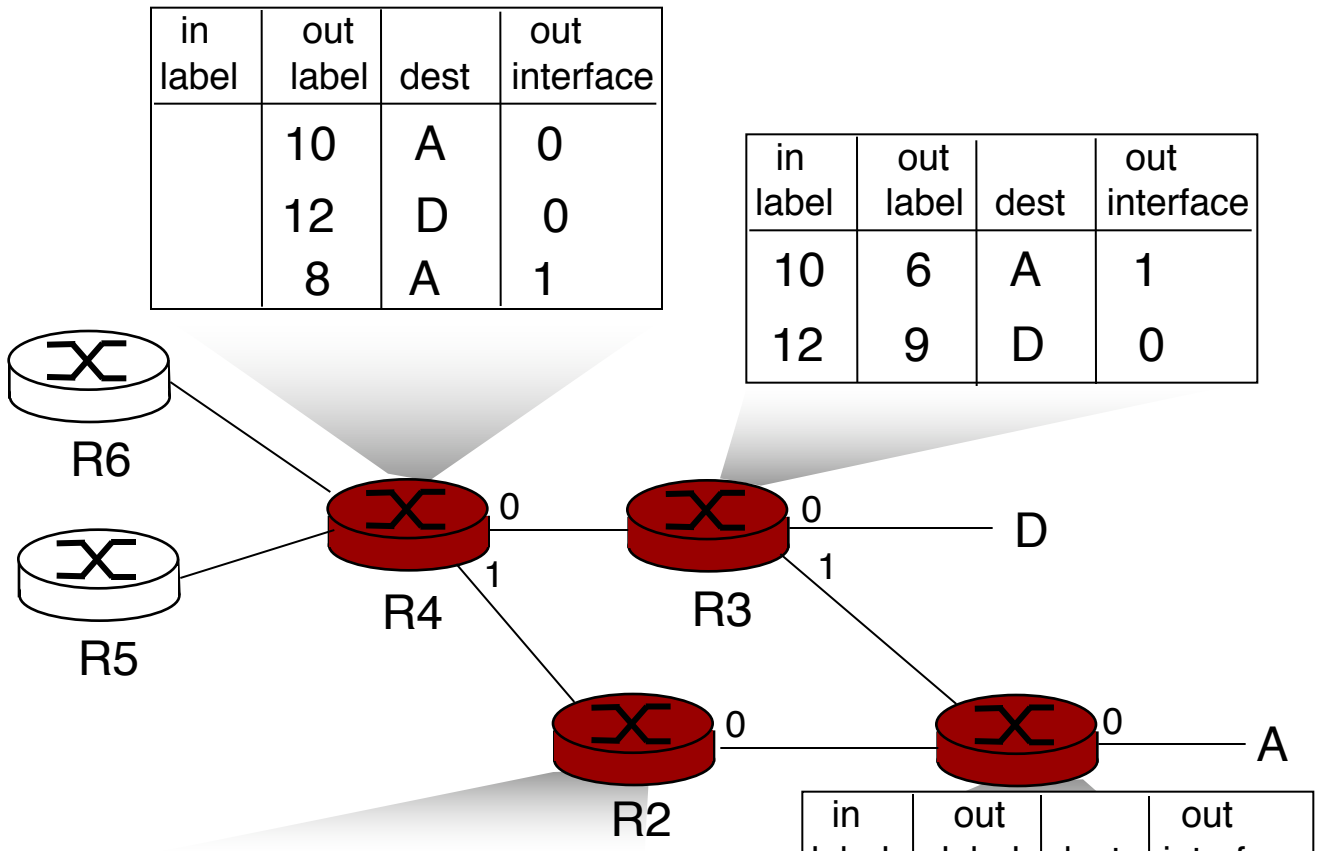
MPLS and IP router

MPLS signaling

- Modify OSPF, IS-IS link-state flooding protocols to carry info used by MPLS routing,
  - Link bandwidth, amount of "reserved" link bandwidth
- *Entry MPLS router uses RSVP-TE signaling protocol to set up MPLS forwarding at downstream routers*



MPLS forwarding tables



Why bother?

- Avoid congested nodes
- Route to underutilized paths

Constraint-Based Routing

- Add additional constraints to routing decision
  - Not just shortest path
- Set the capacity of the link and reserve slices
- Allows calculation of traffic and reservations before traffic actually flows

MPLS Design Issues

- Geographical scope
- Participating routers
- Hierarchy
- Bandwidth requirements of LSPs
- Path attribute
- Priority
- Number of LSPs between pairs
- Affinity of LSPs and links (bias to regions)
- Adaptability and resilience of the LSPs

MPLS Considerations

- Guess and Check: WANDL Simulator
  - Explore changes before production impact
- Periodically updating LSP bandwidth
  - Cron jobs that update small groups at a time
- Offline systemic constraints analysis
  - Globally optimal rather than local with partial information

Offline LSP Calculation

- Sort LSPs from highest priority to lowest
- Iteratively:
  - calculate LSP by pruning unusable links
    - Insufficient bandwidth, excess delay, wrong color
  - Compute optimal path for LSP
  - Deduct resources for consumed elements
- For backups, compute resources using remainders, except exclude all primary elements
- Optimality is NP-complete (bin packing reduces to it)

Quality of Service

- Differentiated Service Fields populated at network edge
- In routing, preferentially weight higher priority
- Premium traffic preferred over best effort

- Remember, which queue a packet is pulled from immediately impacts bandwidth, not just latency

Network Function Virtualization

- Take a physical device's actions and convert it to a VM performing the same tasks
  - Physical intrusion detection system becomes VM inspecting traffic
  - Physical firewall becomes VM with iptables
  - …
- These devices, VM or physical, are in the middle of the connection
  - Hence "middleboxes"

Making Middleboxes Someone Else's Problem: Network Processing as a Cloud Service

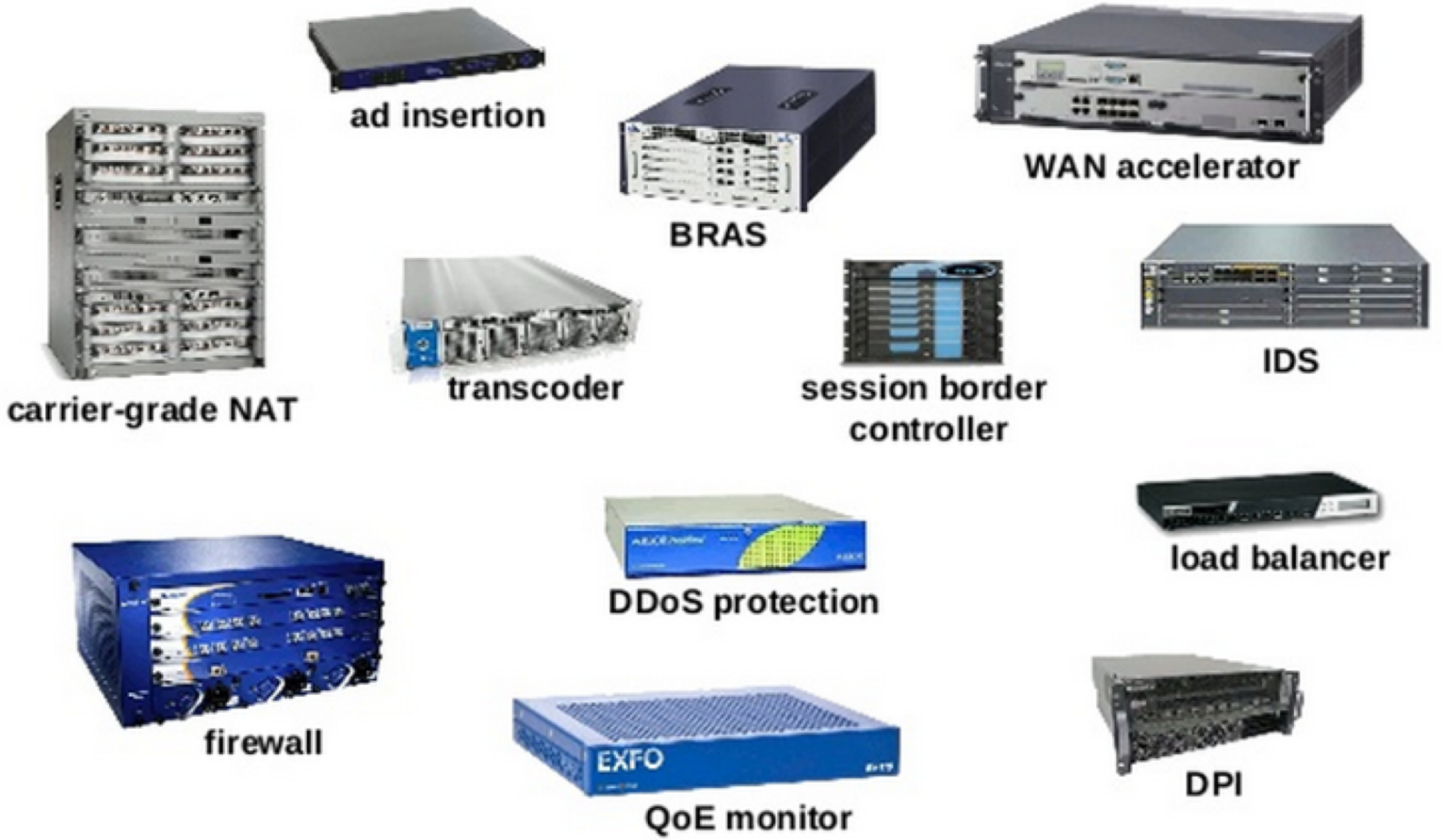
Justine Sherry, Shaddi Hasan, Colin Scott, Arvind Krishnamurthy, Sylvia Ratnasamy,
Vyas Sekar

ACM SIGCOMM 2012

Slide credit: Curtis Taylor

Background

- What is a middlebox?



https://cdn2.hubspot.net/hub/214114/hubfs/blog-images/middleboxes.png?t=1448342962696&width=750&height=444

Problem and Motivation

- Need middlebox infrastructure, but:
  - Expensive
  - Complex to manage
  - New failure modes
- Cloud computing could “outsource”
  - Decreasing cost
  - Reducing management complexity
  - Better fault-tolerance
  - Scalability

Main Contributions

1. Survey of 57 enterprise to understand
   a. the nature of real-world middlebox deployments (*e.g.*, types and numbers of middleboxes)
   b. “pain points” for network administrators
   c. failure modes
2. `APLOMB` architecture – the Appliance for Outsourcing Middleboxes

Survey Results Summary (1)

- Dataset
  - 19 small (fewer than 1k hosts) networks
  - 18 medium (1k-10k hosts) networks
  - 11 large (10k-100k hosts) networks
  - 7 very large (more than 100k hosts) networks
- Nature of MBes
  - Number of MBes is on par with the number of routers
  - Average very large: 2,850 L3 routers and 1,946 MBes
  - Average small: 7.3 L3 routers and 10.2 MBes

Survey Results Summary (2)

- Pain points
  - Management – heterogeneous MBes requires broad expertise (i.e., more personnel)
    - Even small networks (10s MBes) required a management team of 6-25 personnel
  - Upgrading and Vendor Interaction
    - Upgrade median – every 4 years
    - Dealt with average of 4.9 vendors per upgrade

Survey Results Summary (3)

- Failure modes

	Misconfig.	Overload	Physical/Electric
Firewalls	67.3%	16.3%	16.3%
Proxies	63.2%	15.7%	21.1%
IDS	54.5%	11.4%	34%

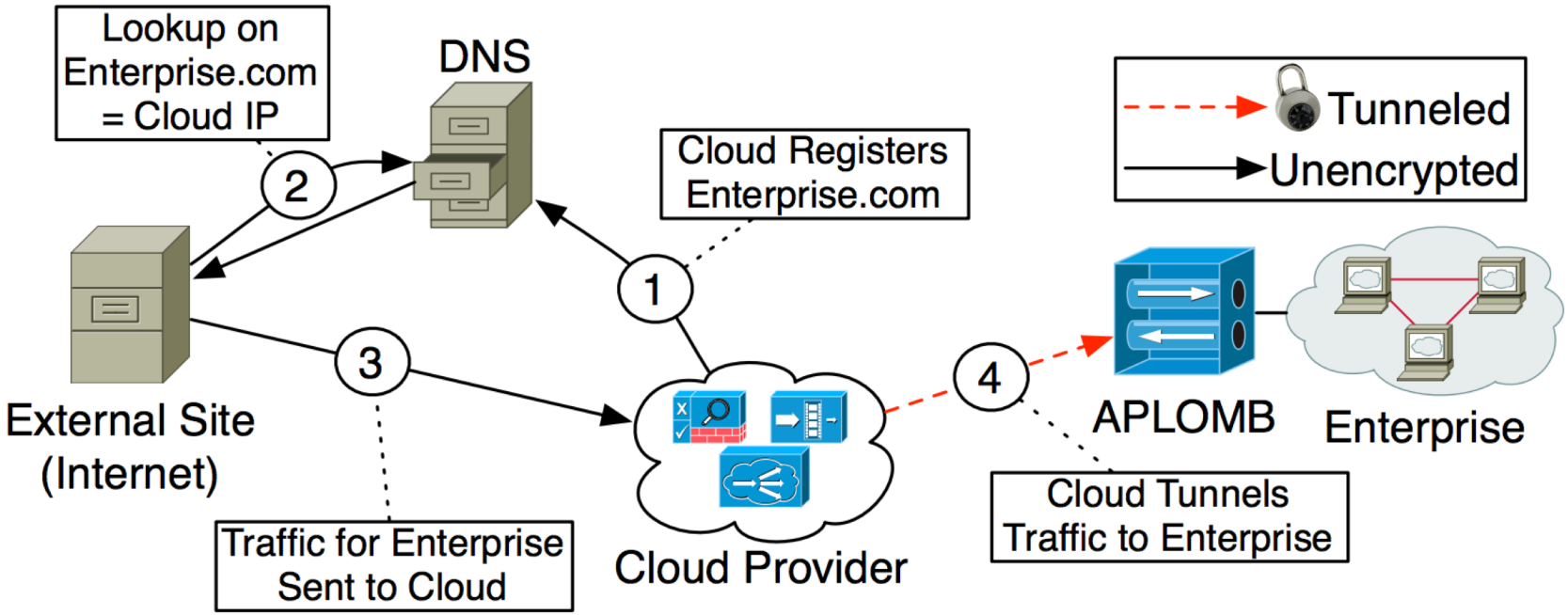
Table 1: Fraction of network administrators who estimated misconfiguration, overload, or physical/electrical failure as the most common cause of middlebox failure.

APLOMB (1)

- An architecture that enables enterprise networks to put MBes in the cloud
- Three design challenges APLOMB meets:
  1. Functional equivalence
  2. Low complexity at the enterprise
  3. Low performance overhead*

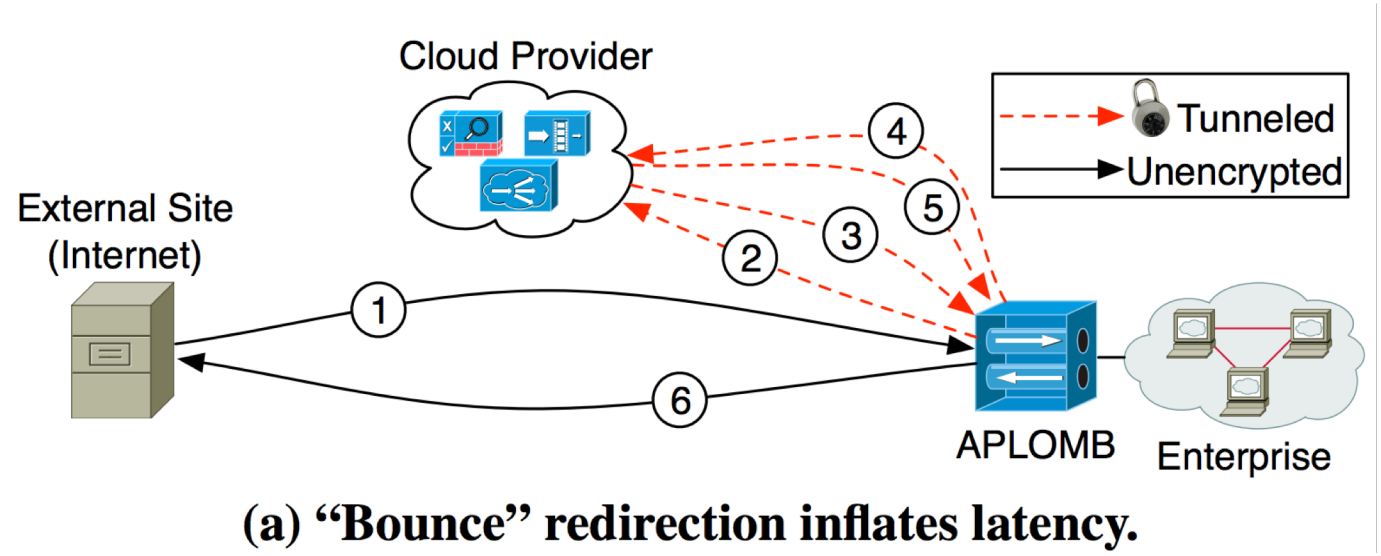
APLOMB (2)

- Considered 3 design possibilities. DNS-based redirection was the optimal choice

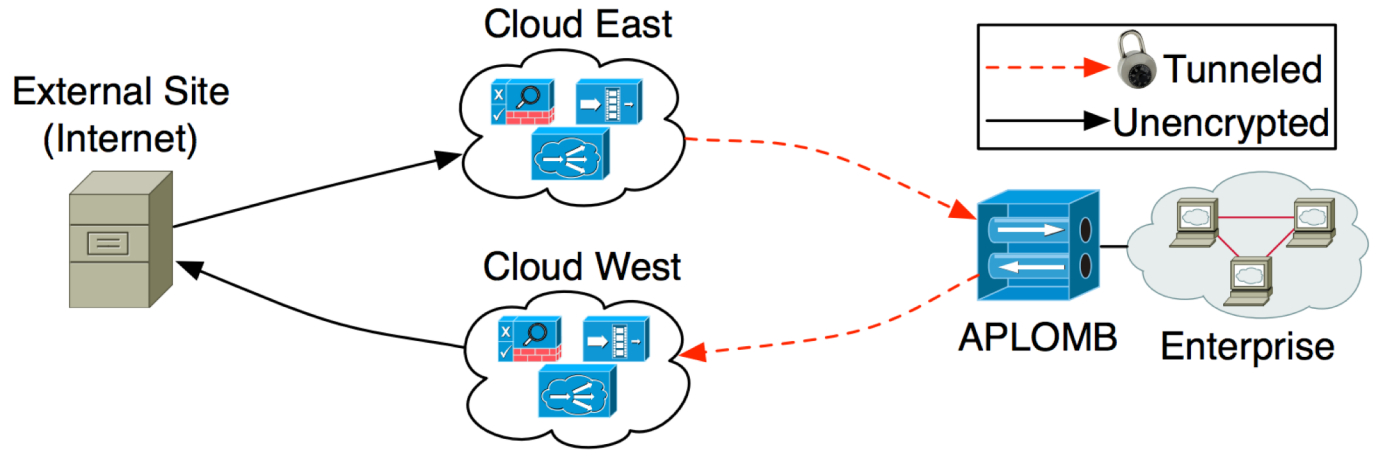


(c) DNS-based redirection minimizes latency and allows providers to control PoP selection for each request.

Other Design Considerations



(a) "Bounce" redirection inflates latency.



(b) Direct IP redirection in multi-PoP deployments cannot ensure that bidirectional traffic traverses the same PoP.

APLOMB Performance

- Deployed middleboxes unknown:

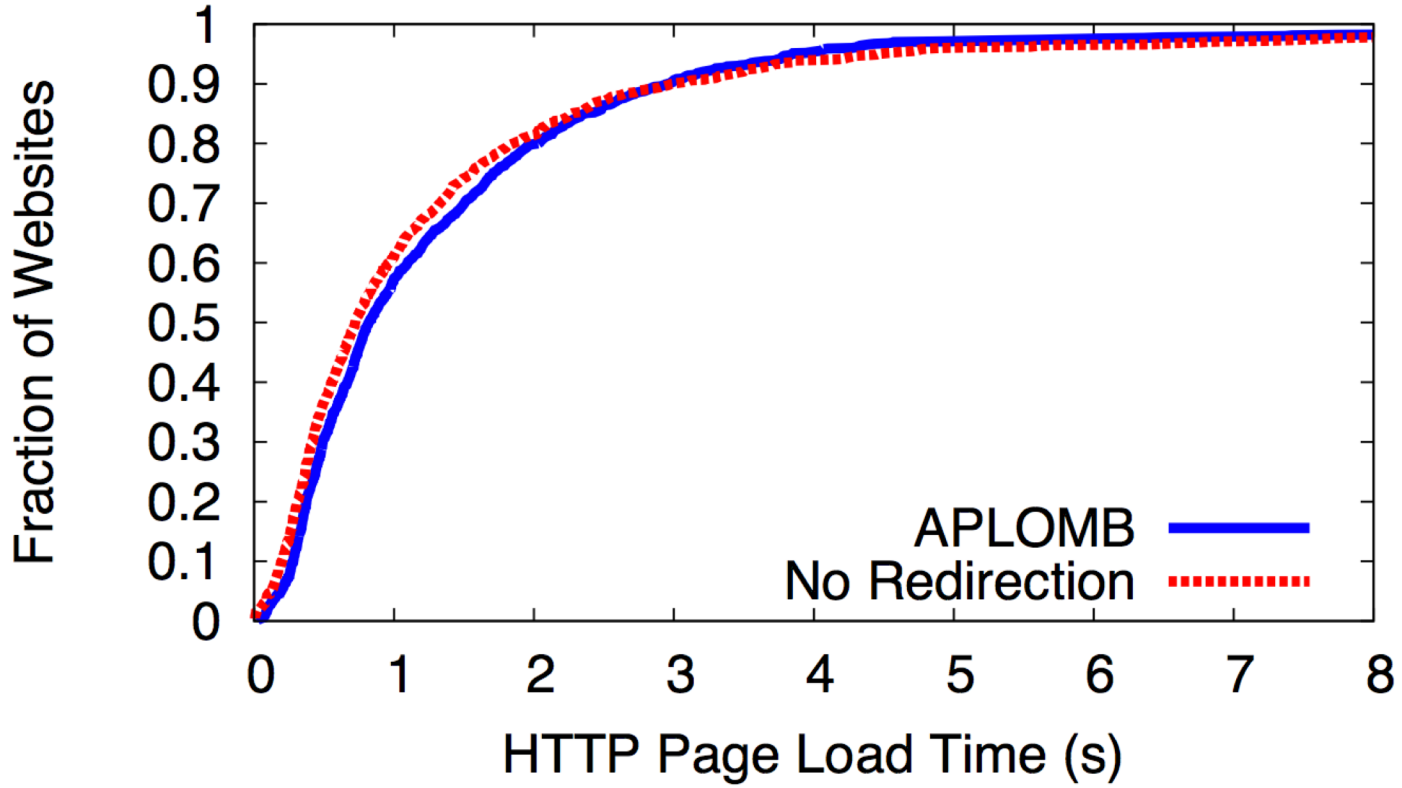


Figure 12: CDF of HTTP Page Load times for Alexa top 1,000 sites with and without APLOMB.

APLOMB Applicability (1)

- Based on survey numbers
- APLOMB ~50% of MBes can be outsourced
- APLOMB+ ~90%
  - Some MBes must necessarily reside at the source. APLOMB+ implements protocol-agnostic compression
  - Remaining 10% are internal firewalls

APLOMB Applicability (2)

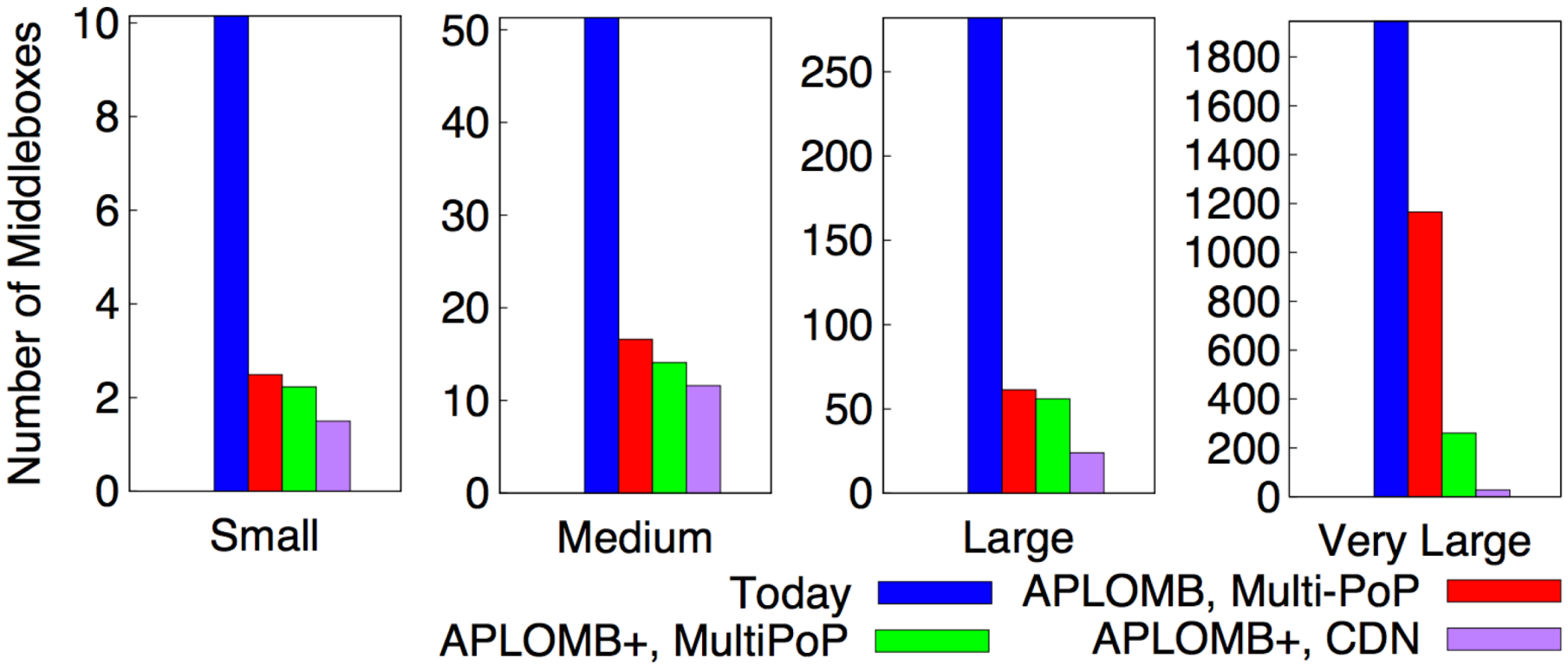


Figure 9: Average number of middleboxes remaining in enterprise under different outsourcing options.

Critique

- Ultimately, DNS-based redirection only works when the enterprise controls a set of domain names for incoming requests
- Requires complex, physical gateway for deployment
- Question remains: Can this work for smaller networks?