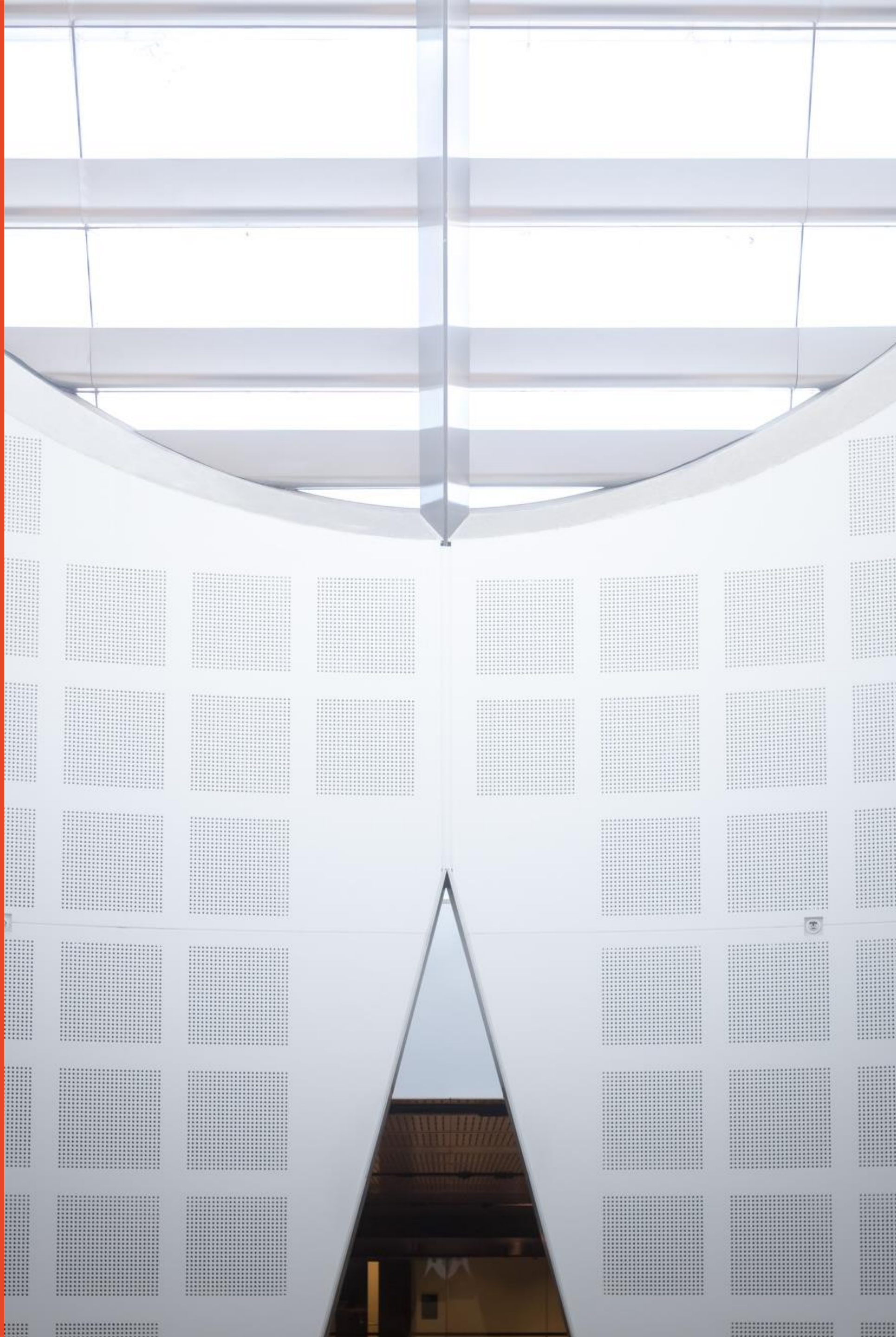
Topic 11: Software Defined WANs

Presented by
Dong YUAN

School of Electrical and Computer Engineering

dong.yuan@sydney.edu.au

THE UNIVERSITY OF SYDNEY



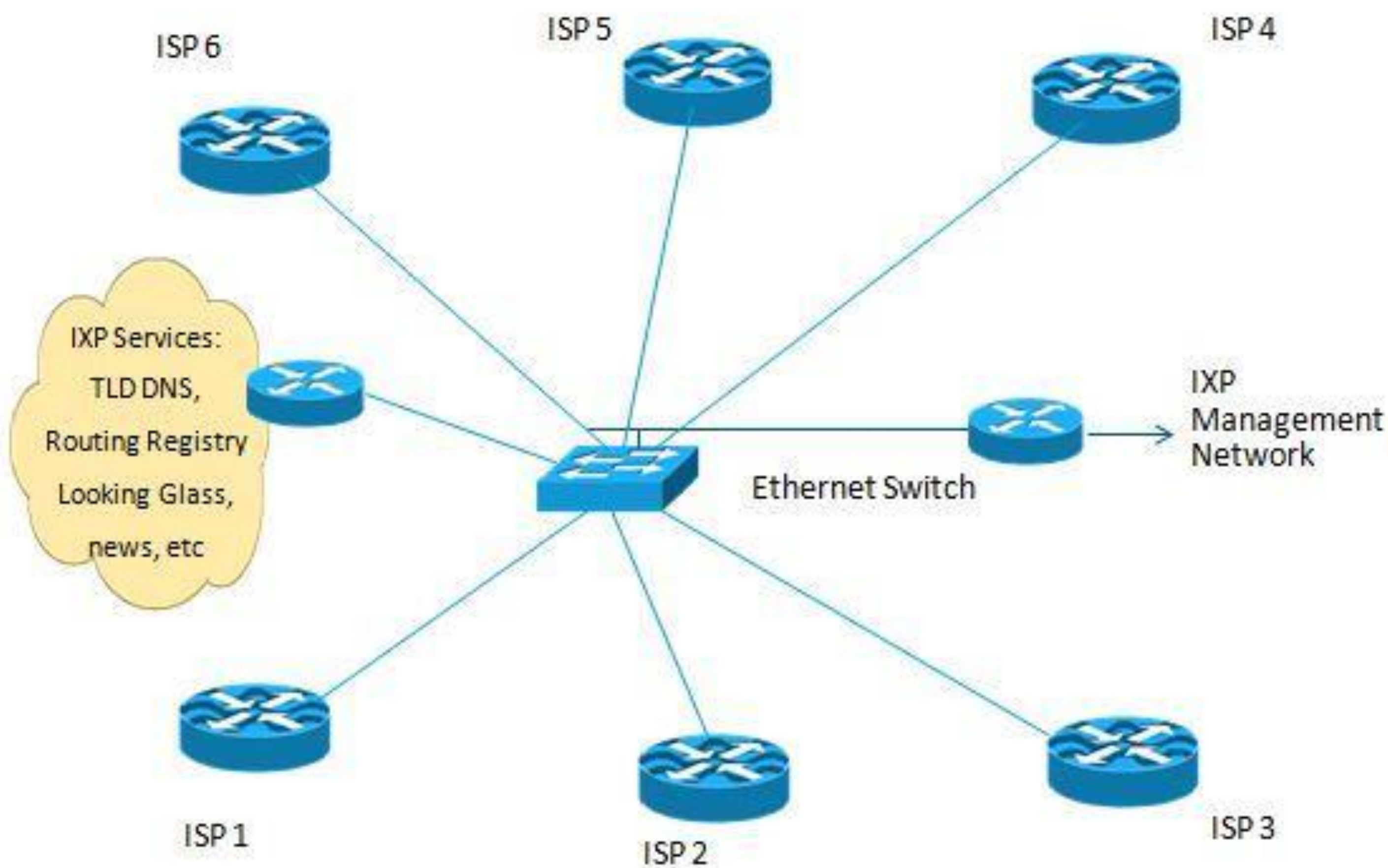
Contents

- Software Defined Exchange (SDX)
  - Gupta, Arpit, et al. "Sdx: A software defined internet exchange." ACM SIGCOMM Computer Communication Review 44.4 (2015): 551-562.

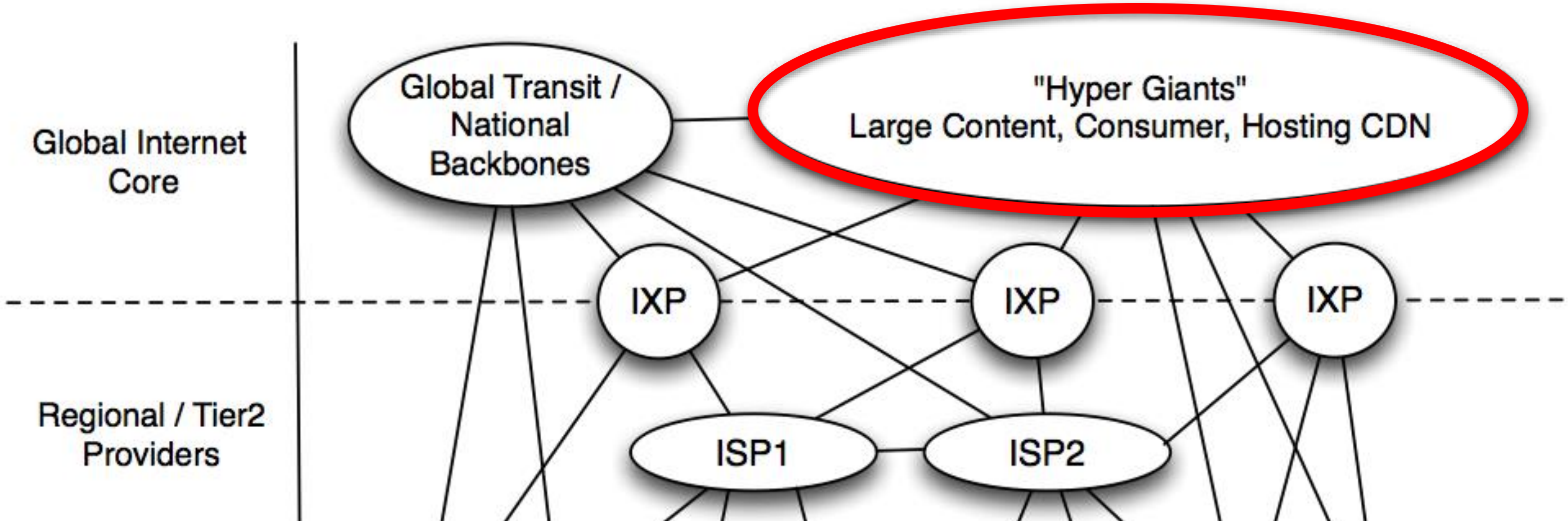
- Software Defined WAN
  - Jain, Sushant, et al. "B4: Experience with a globally-deployed software defined WAN." ACM SIGCOMM Computer Communication Review 43.4 (2013): 3-14.
  - Hong, Mandal, et al. "B4 and After: Managing Hierarchy, Partitioning, and Asymmetry for Availability and Scale in Google's Software-Defined WAN," ACM SIGCOMM 2018, pp 74-87.

Internet exchange

An Internet exchange point (IX or IXP) is a physical infrastructure through which Internet service providers (ISPs) and Content Delivery Networks (CDNs) exchange Internet traffic between their networks (autonomous systems, i.e. AS).



The Interdomain Ecosystem is Evolving ...



Flatter and densely interconnected Internet*

– Routing on Internet relies on the Border Gateway Protocol

*Labovitz et al., *Internet Inter-Domain Traffic*, SIGCOMM 2010

…But BGP is Not

– Routing **only on destination IP prefixes**
  (No customization of routes by application, sender)

– Can only influence **immediate neighbors**
  (No ability to affect path selection remotely)

– **Indirect** control over data-plane forwarding (Indirect mechanisms to influence path selection)

How to overcome BGP's limitations?

SDN for Interdomain Routing

- Forwarding on **multiple header fields** (not just destination IP prefixes)

- Ability to **control entire networks** with a single software program (not just immediate neighbors)

- **Direct control** over data-plane forwarding (not indirect control via control-plane arcana)

How to incrementally deploy SDN for Interdomain Routing?

Deploy SDN at Internet Exchanges

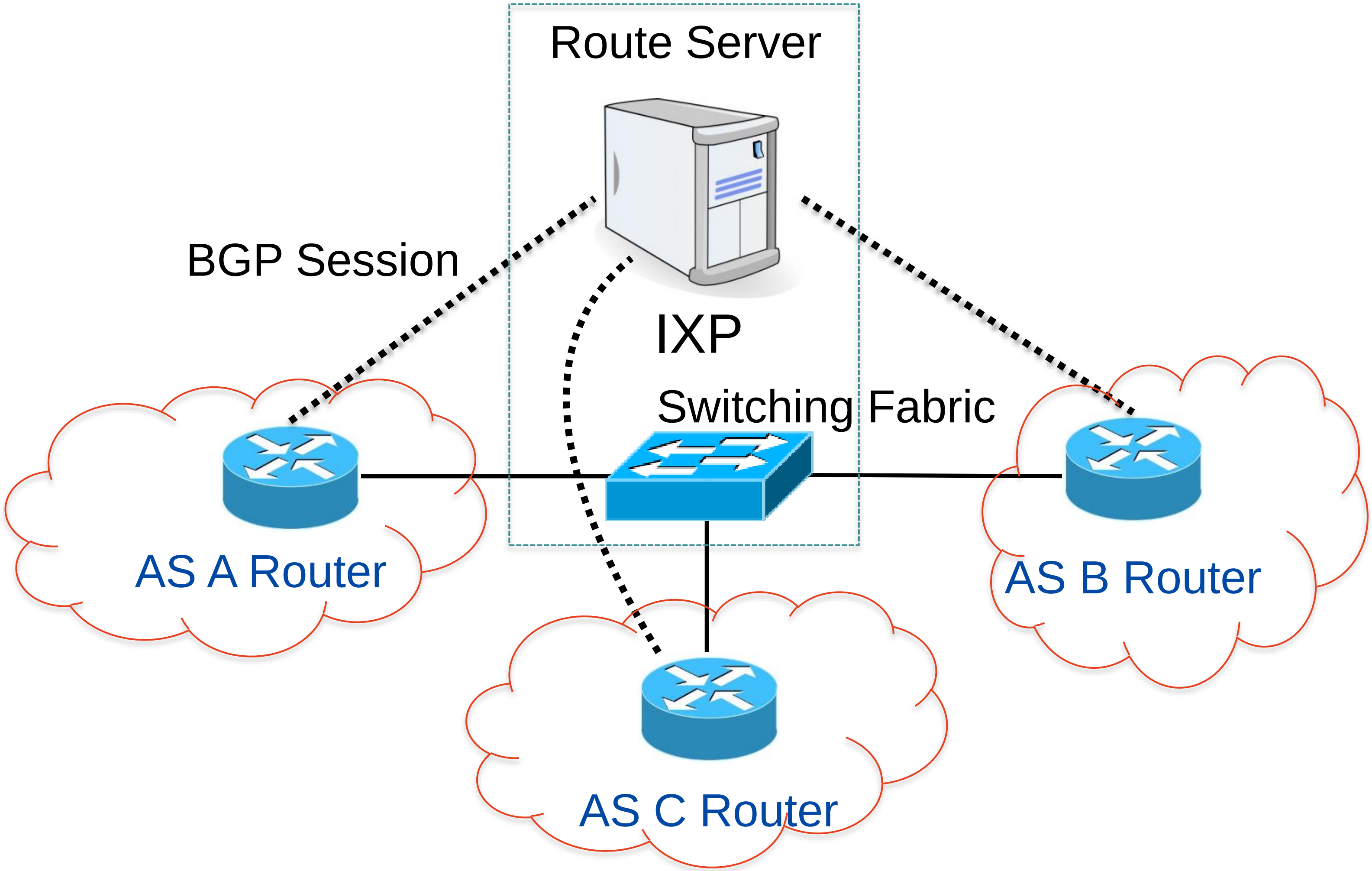
- **Leverage:** SDN deployment even at single IXP can yield benefits for tens to hundreds of ISPs

- **Innovation hotbed:** Incentives to innovate as IXPs on front line of peering disputes

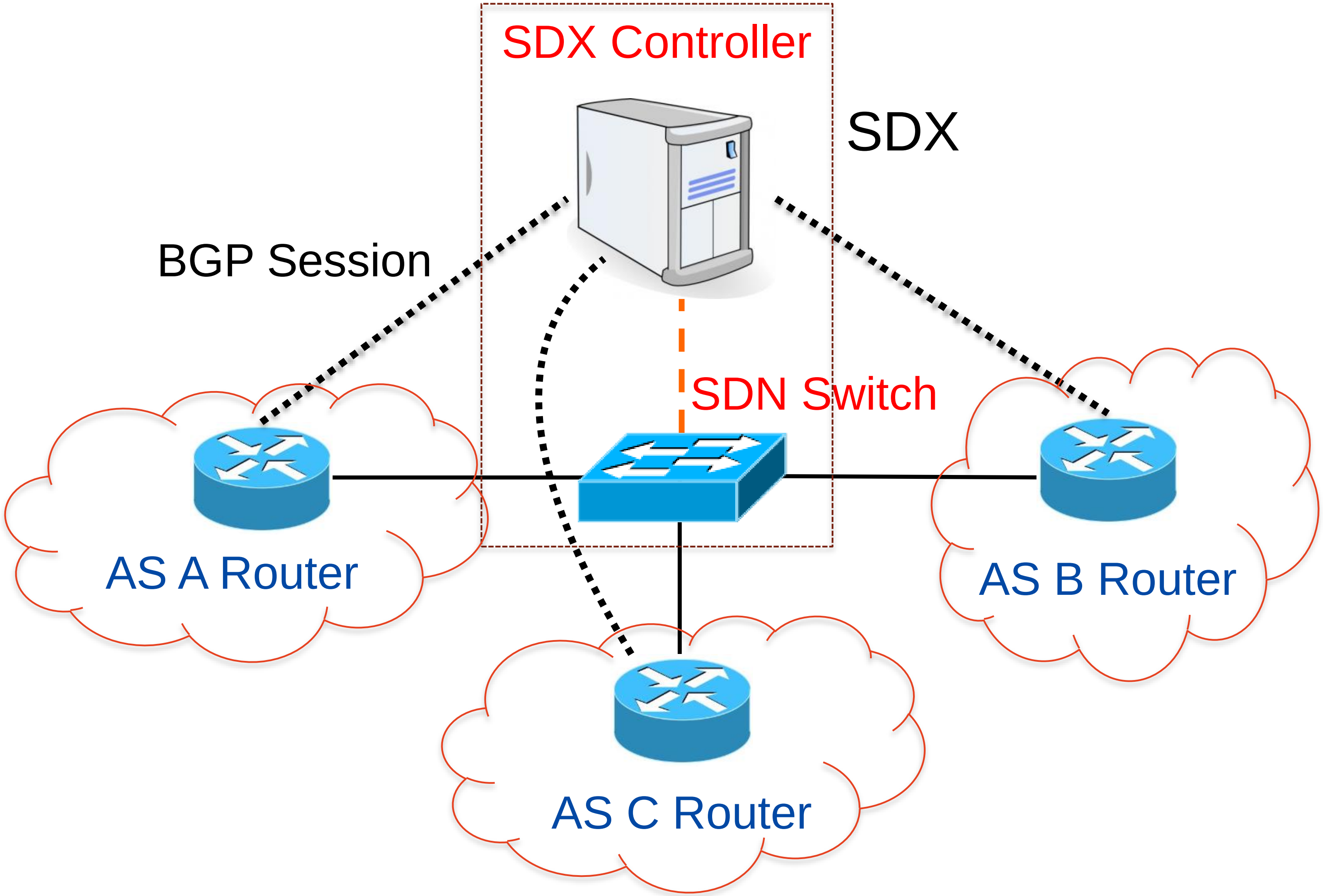
- **Growing in numbers:** ~100 new IXPs established in past three years*

*https://prefix.pch.net/applications/ixpdir/summary/growth/

Background: Conventional IXPs



SDX = SDN + IXP


SDX Controller
SDX
BGP Session
SDN Switch
AS A Router
AS B Router
AS C Router

SDX Opens Up New Possibilities

- More flexible **business relationships**

  Make peering decisions based on time of day, volume of traffic & nature of application

- More direct & flexible **traffic control**

  Define fine-grained traffic engineering policies

- Better **security**
  - Prefer "more secure" routes
  - Automatically blackhole attack traffic

SDX Enables Innovations at IXPs

- Application-specific peering
  - Video traffic via Comcast, non-video via AT&T

- Inbound traffic engineering
  - Divide traffic by sender or application

- Dropping of attack traffic
  - Blocking unwanted traffic in middle of Internet

Building SDX is Challenging

- Programming **abstractions**

  How networks define SDX policies and how are they combined together?

- **Interoperation** with BGP

  How to provide flexibility without breaking global routing?

- **Scalability**

  How to handle policies for hundreds of peers, half million prefixes and matches on multiple header fields?

Scalability Challenges

- **Reducing Data-Plane State:** Support for all forwarding rules in (limited) switch memory

- **Reducing Control-Plane Computation:** Faster policy compilation

SDX Platform

- Running code with full BGP-integration

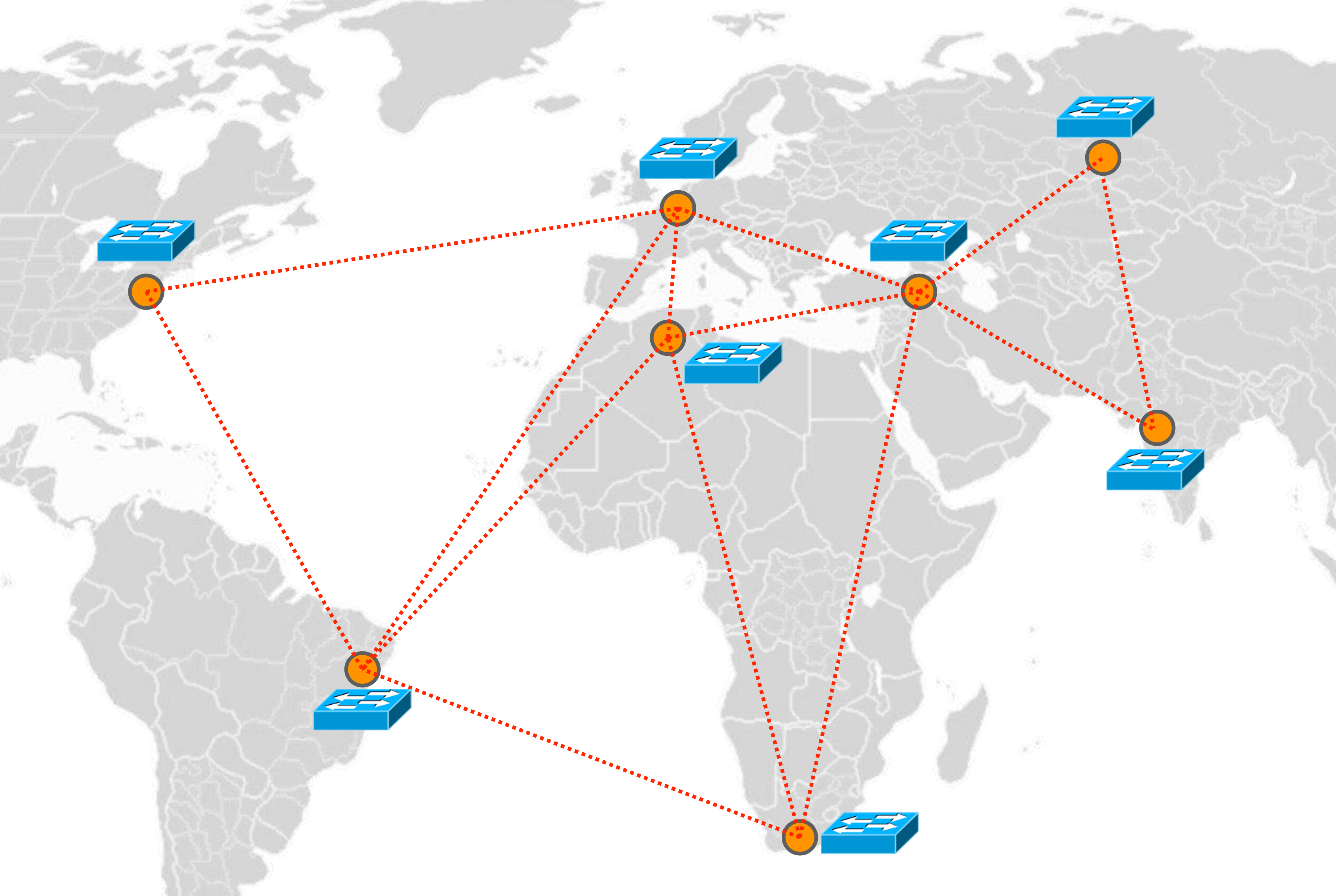
- SDX Testbeds:
  - Uses Transit Portal
  - Emulates edge routers (Mininet)

Github repo: https://github.com/sdn-ixp/sdx/

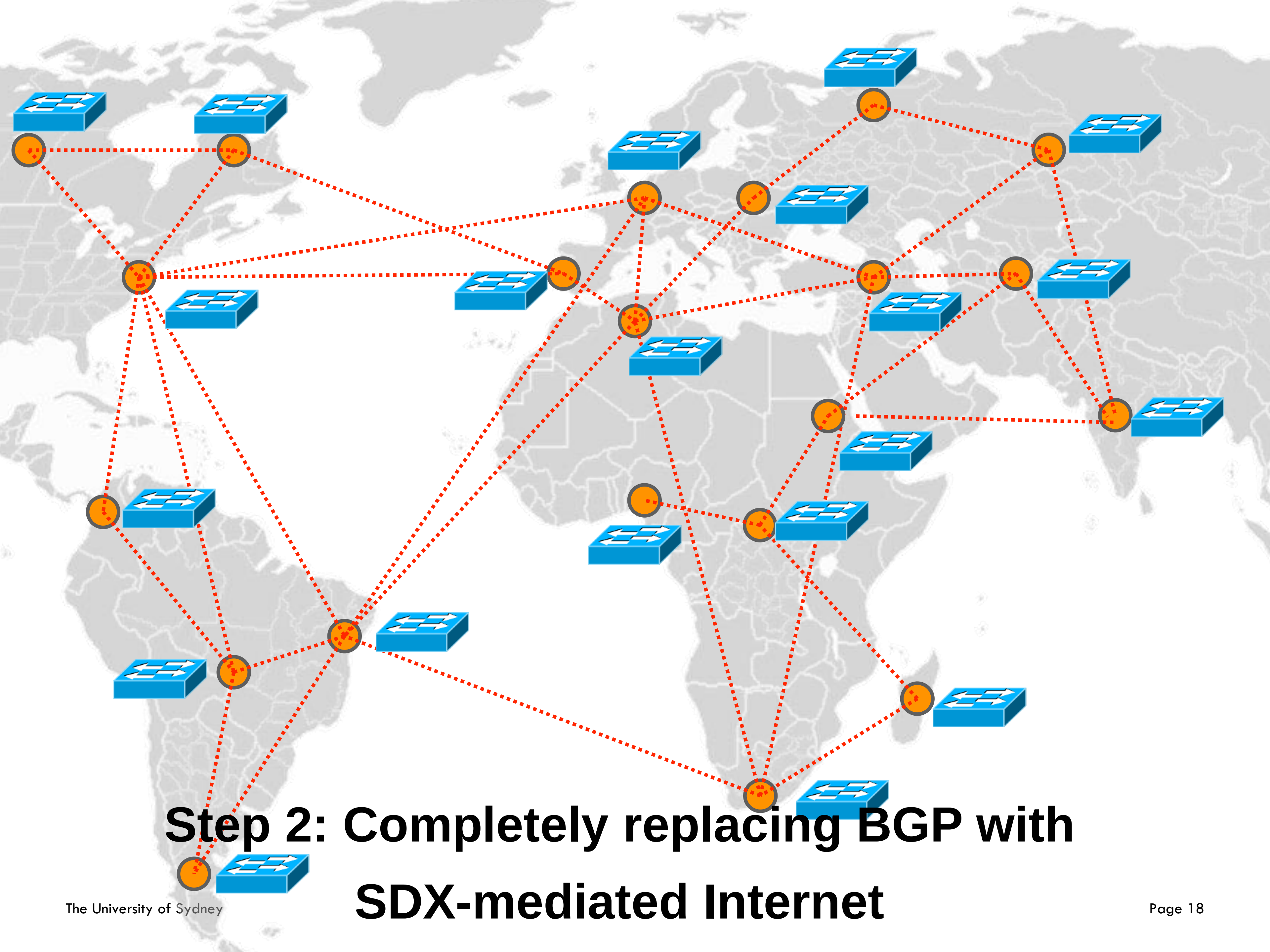
Next Steps

Building SDX-mediated Internet

SDX currently considers a single deployment



Step 1: Interconnecting SDX platforms



Step 2: Completely replacing BGP with SDX-mediated Internet

Challenges?

What is Software-Defined WAN (Virtualisation approach)

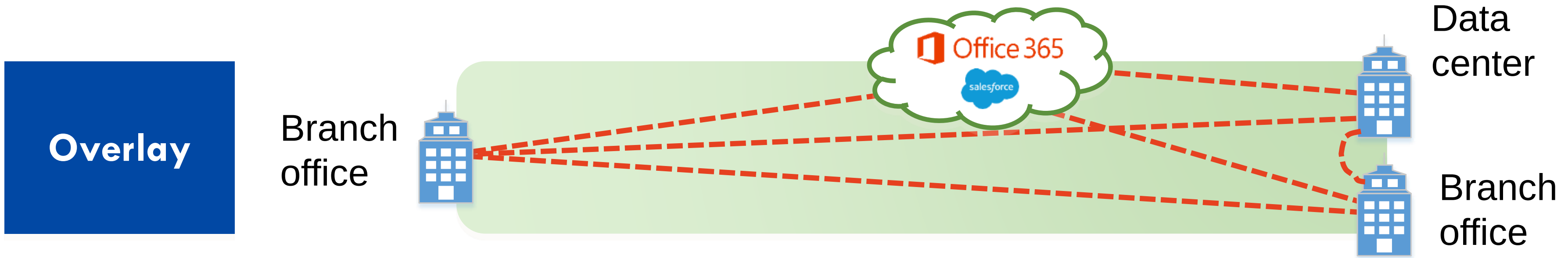
- A method for enabling flexibility, cost reduction, and redundancy by virtualizing WAN connections
- Flexibility:
  - Virtualized WAN connections enable scalable and dynamic routing, which optimizes performance.
- Cost Reduction:
  - Optimization enables businesses to leverage inexpensive broadband connections.
- Redundancy:
  - When multiple WAN connections are available customers have built-in redundancy for mission critical cloud-based applications.

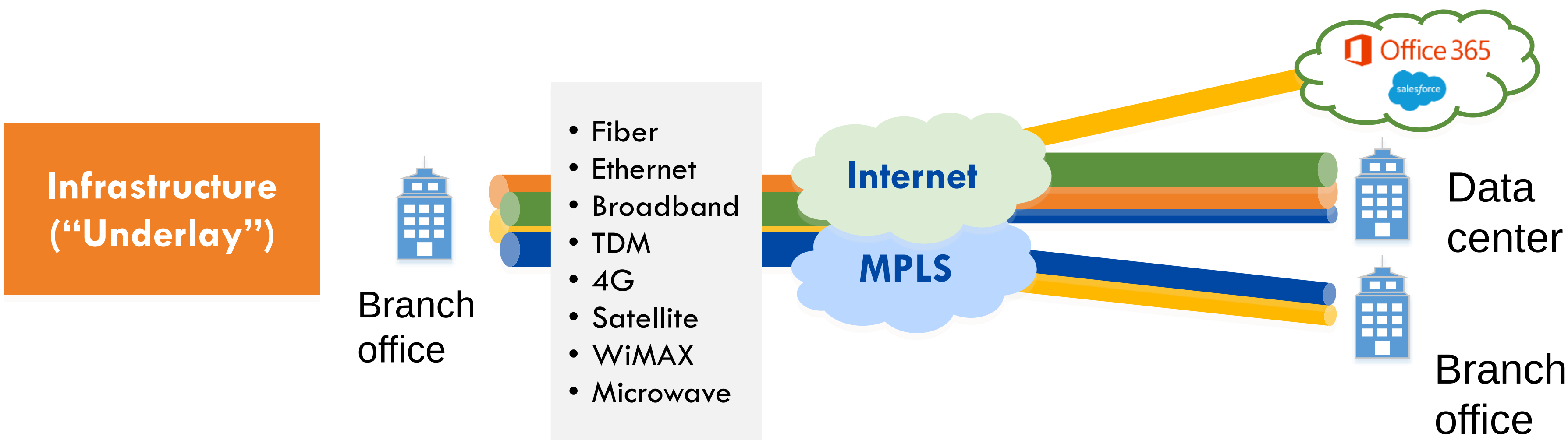
SDN –> SD-WAN

	What does the acronym stand for?	*What does it mean?*
SDN	Software-Defined Networking	Separating the **control and data planes** to create centrally-controlled, programmable networks
NFV / VNF	Network Function Virtualization / Virtual Network Function	A Virtual Network Function is a **virtualized task** formerly performed on proprietary, dedicated hardware. NFV is the practice of utilizing VNFs
SD-WAN	Software-Defined Wide Area Network	Loosely applying SDN concepts to the WAN to create a **centrally-controlled overlay network** that intelligently uses a variety of infrastructure options

Source: coevolve.com

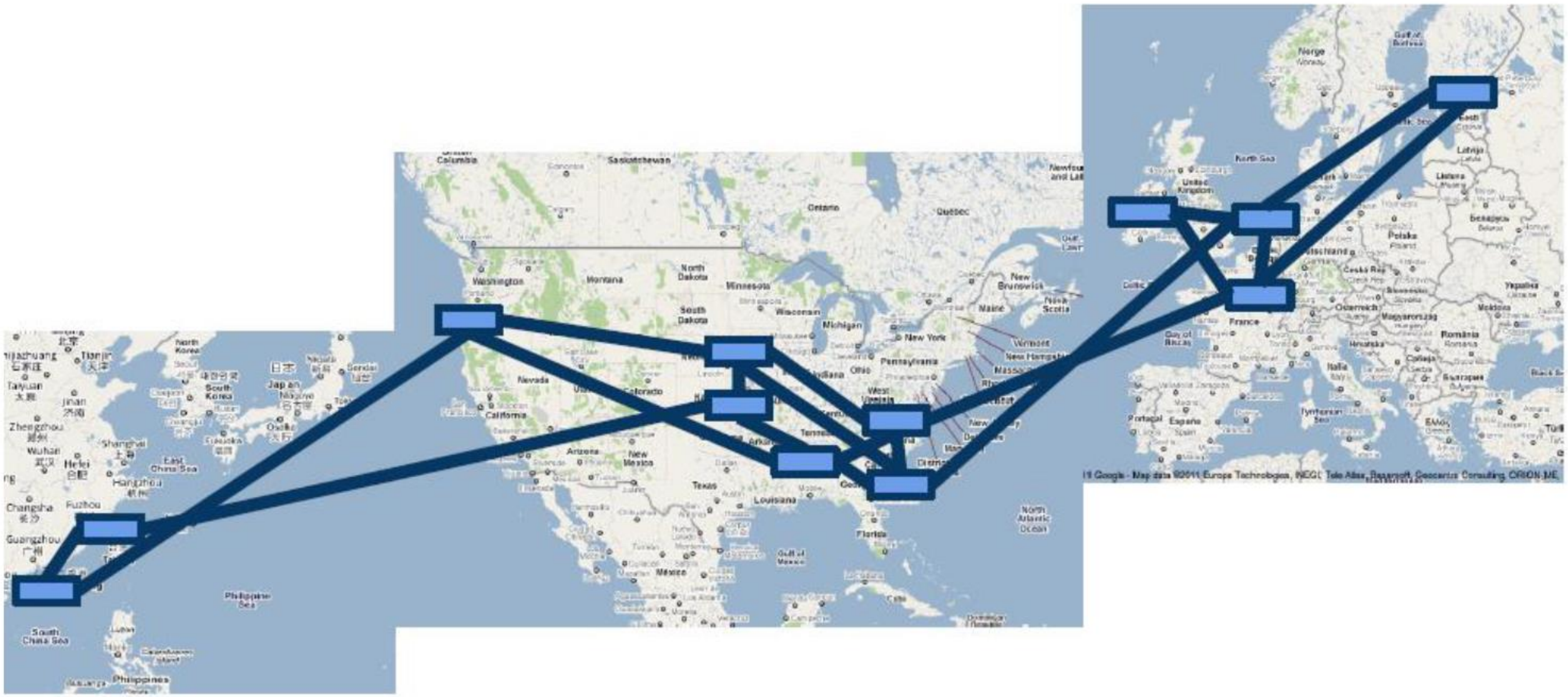
SD-WAN





Source: coevolve.com

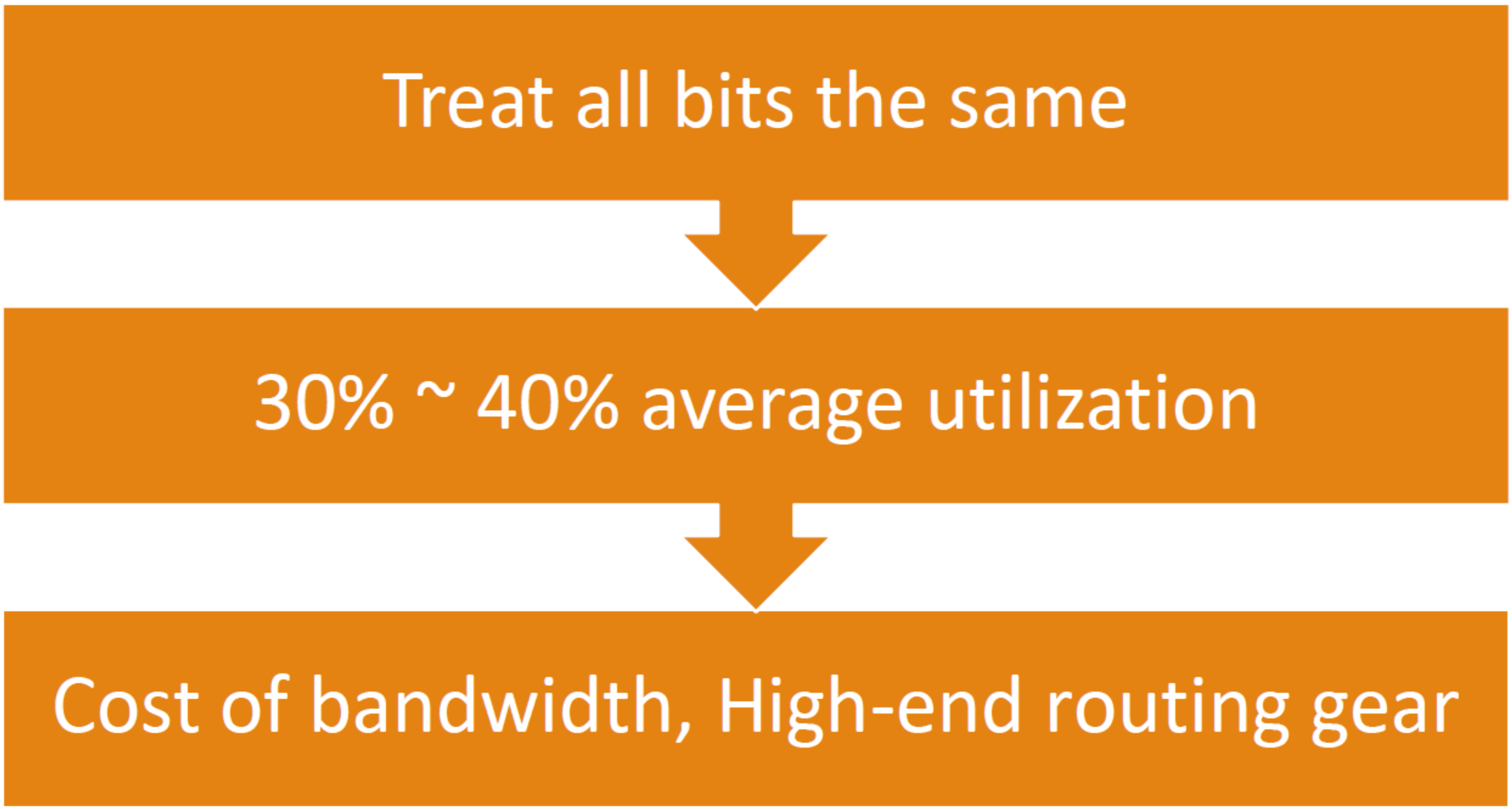
B4: Google's Software Defined WAN



B4: Google's Software Defined WAN

- Google's private WAN connecting its data centers
  - Elastic bandwidth demands
    - Can tolerate periodic failures with temporary BW reduction
  - Small number of sites
    - Allows special optimization
  - Complete control of end application
    - Application priorities and control bursts
  - Cost Sensitivity
    - Unsustainable cost projection with traditional approach (2-3x cost of a fully utilized WAN).

Traditional WAN routing (packet-based)



Traffic priority

- User data copies to remote data centers for availability and durability (lowest volume, most latency intensive, highest priority)
- Remote storage access for computation over distributed data sources
- Large-Scale data push synchronizing state across multiple data centers (highest volume, least latency intensive, lowest priority)

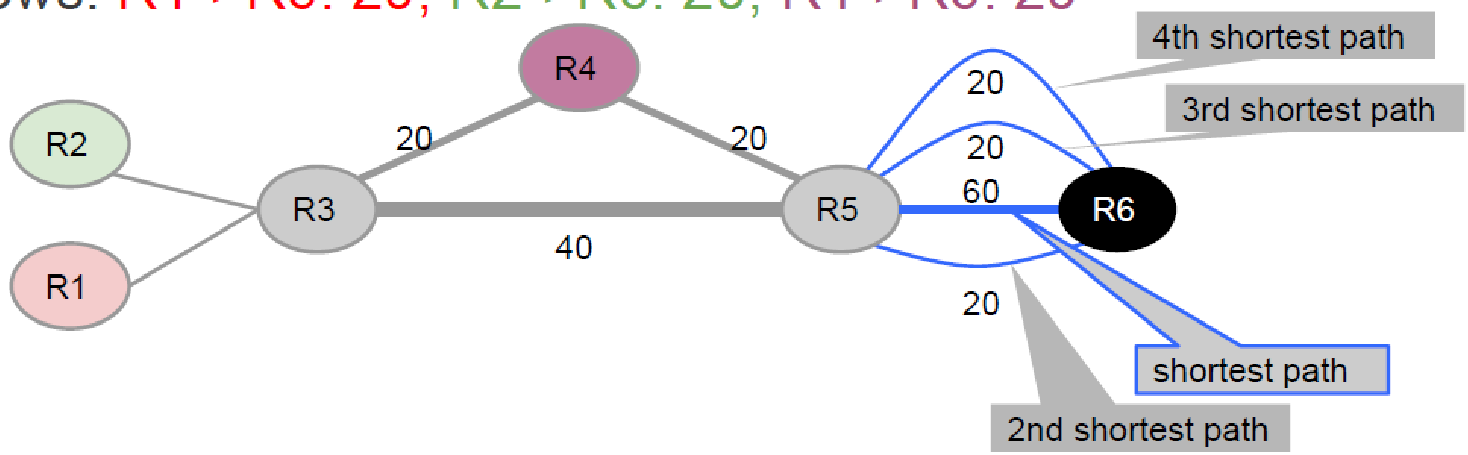
- Centralized traffic Engineering (TE)
  - Near 100% utilization
  - Fast, global convergence for failures.

B4 design decisions

- B4 routers built from merchant switch silicon
  - Low router cost → scale network capacity
- Drive links to 100% utilization
  - High average bandwidth over predictability: largest bandwidth consumers can adapt to bandwidth availability
- Centralized traffic engineering
  - Application classification and priority
  - Faster, deterministic global convergence for failures
- Separate hardware from software
  - Customized routing
  - Easier to protect against common case software failures

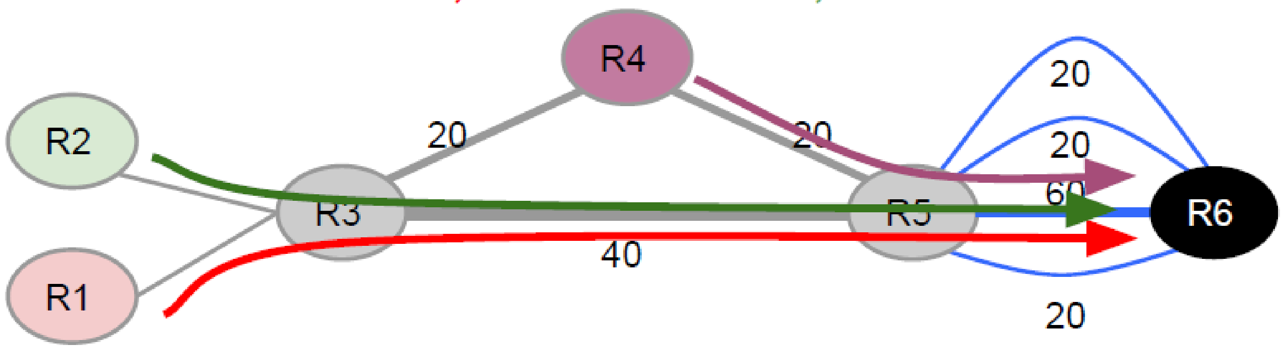
Centerlized TE: convergence after failure

- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20



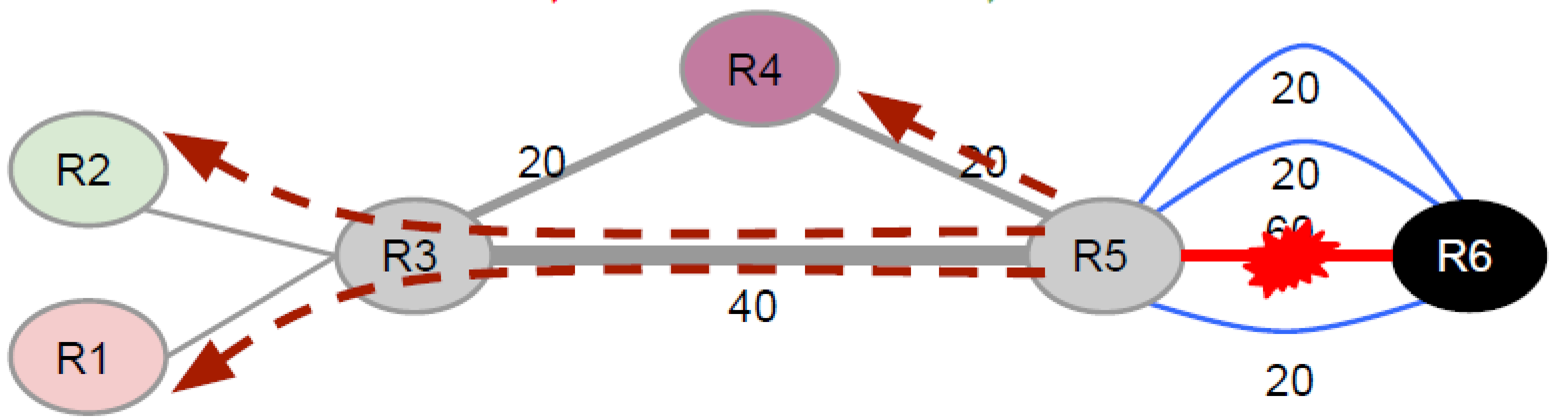
Centerlized TE: convergence after failure

- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20



Centerlized TE: convergence after failure

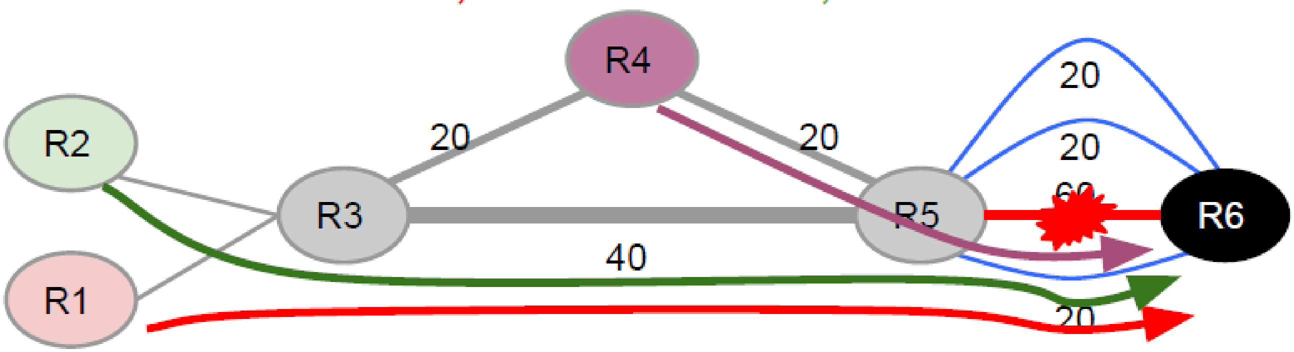
- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20



- R5-R6 link fails
  - R1, R2, R4 *autonomously* find next best path

Centerlized TE: convergence after failure

- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20

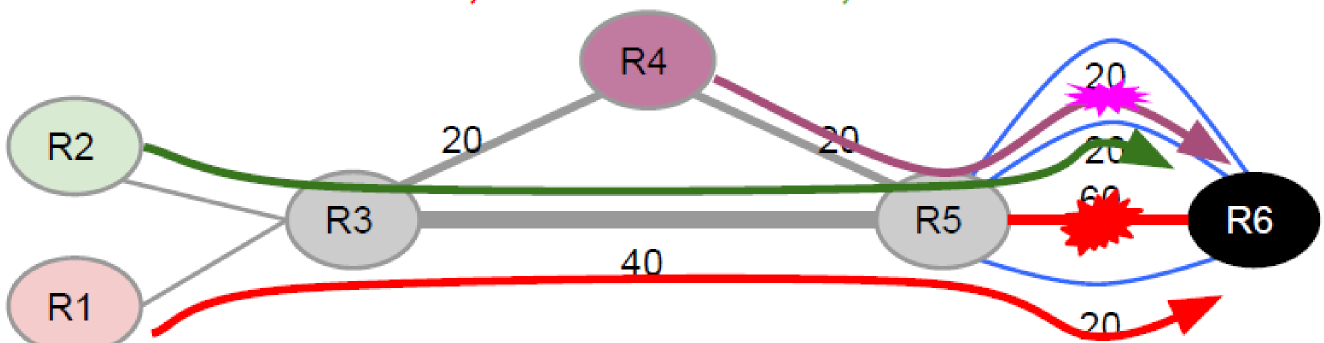


No Traffic Engineering

- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1, R2, R4 push **20** altogether

Centerlized TE: convergence after failure

- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20

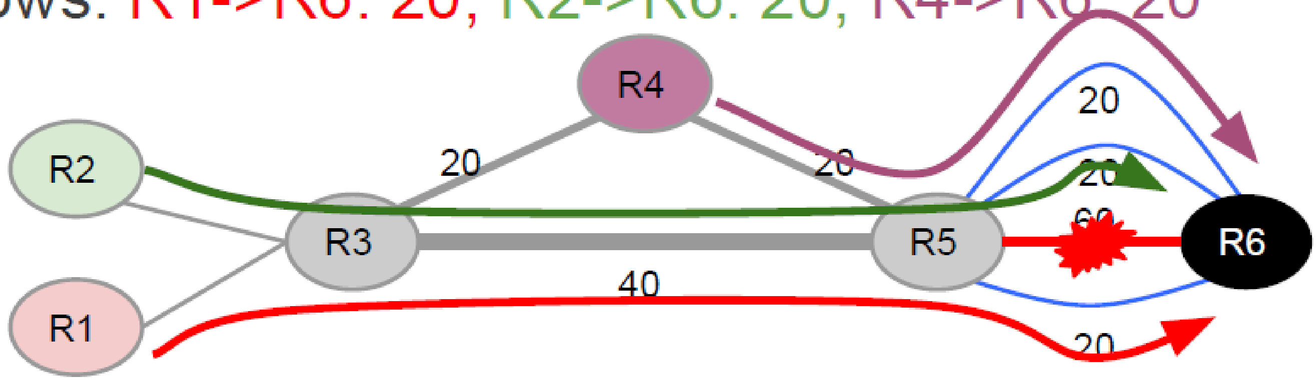


Distributed Traffic Engineering Protocols

- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1 wins, R2, R4 retry for next best path
  - R2 wins this round, R4 retries again

Centerlized TE: convergence after failure

- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20

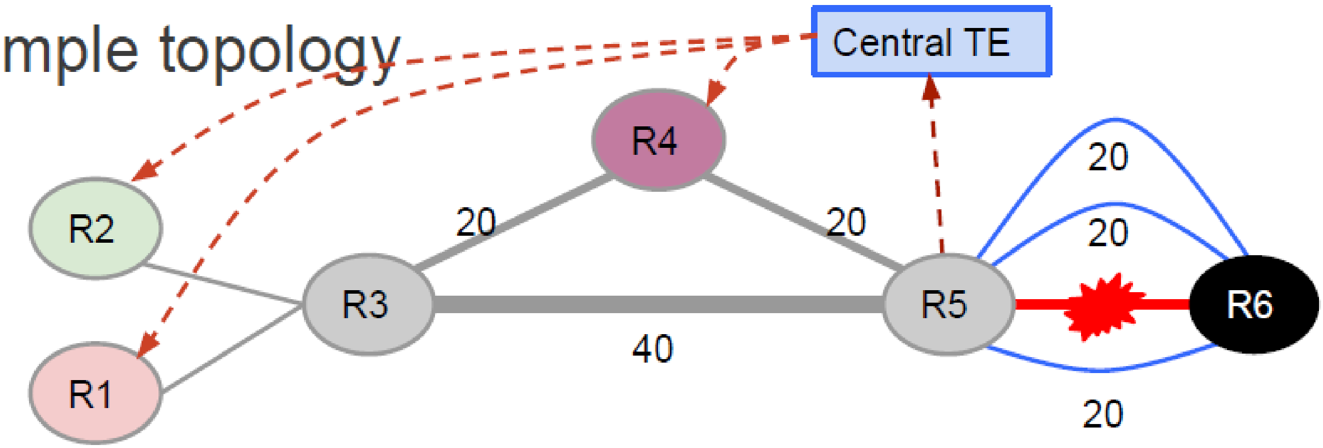


Distributed Traffic Engineering Protocols

- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1 wins, R2, R4 retry for next best path
  - R2 wins this round, R4 retries again
  - R4 finally gets third best path!

Centerlized TE: convergence after failure

- Simple topology

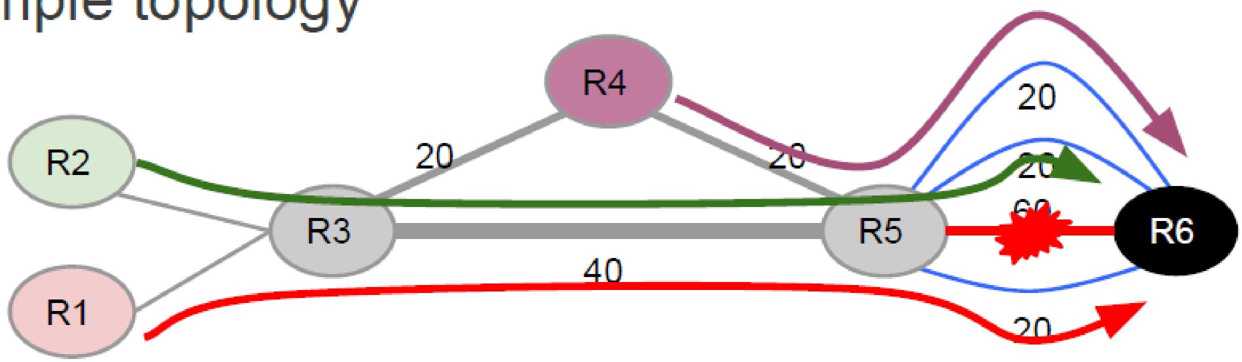


Centralized Traffic Engineering Protocols

- Flows:
  - R1->R6: 20; R2->R6: 20; R4->R6: 20
- R5-R6 fails
  - R5 informs TE, which programs routers in one shot

Centerlized TE: convergence after failure

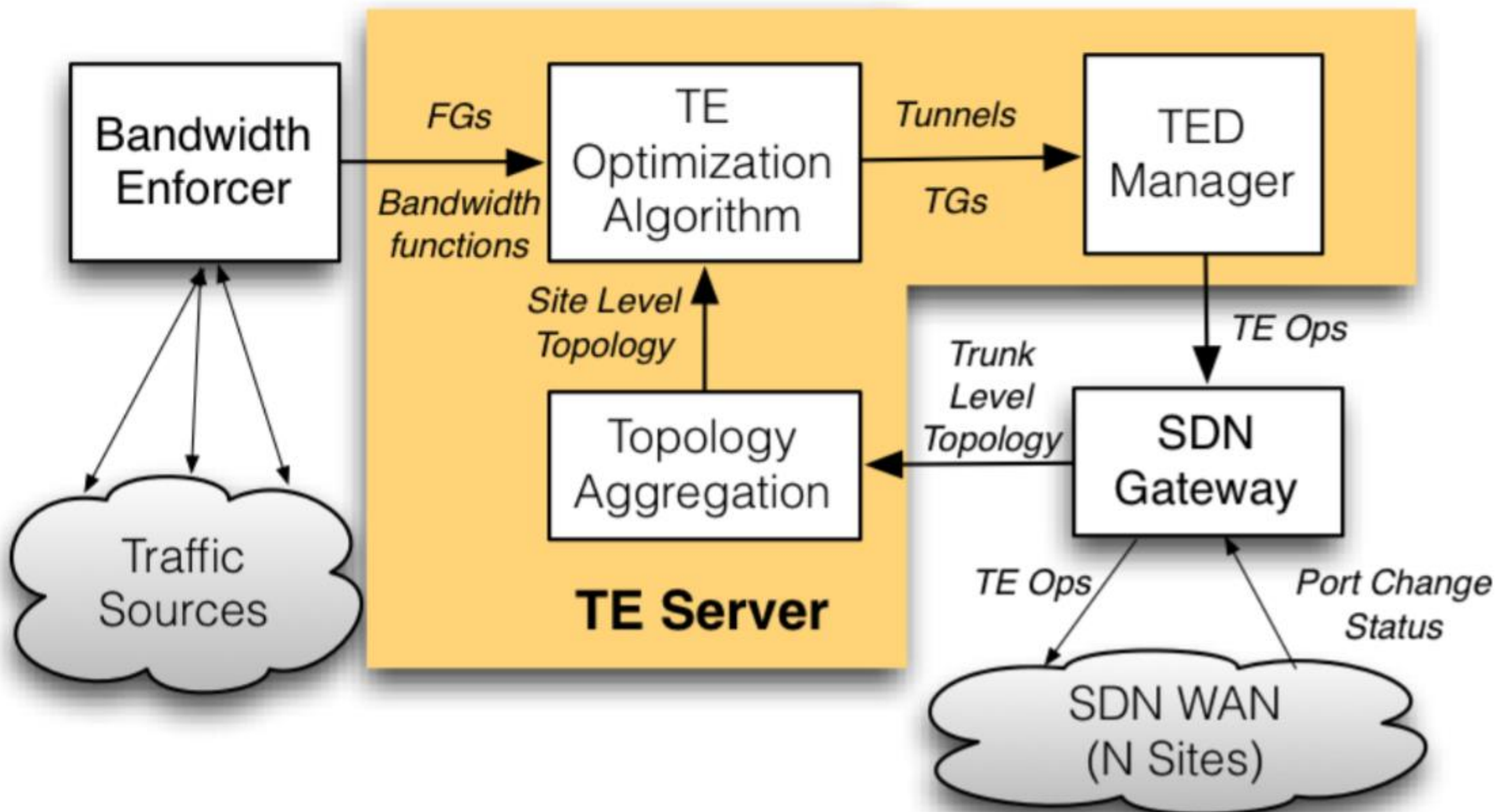
- Simple topology



Centralized Traffic Engineering Protocols

- Flows:
  - R1->R6: 20; R2->R6: 20; R4->R6: 20
- R5-R6 link fails
  - R5 informs TE, which programs routers in one shot
  - Leads to faster realization of target optimum

Traffic Engineering Overview



Advantage of Centralized TE

- Better network utilization with global pictures
- Converges faster to target optimum on failure
- Allows more control and specifying intend
  - Deterministic behavior simplifies planning .vs. over provisioning for worst cast variability
- Can mirror production event streams for testing
- Controller uses modern server hardware – better performance (50x!)

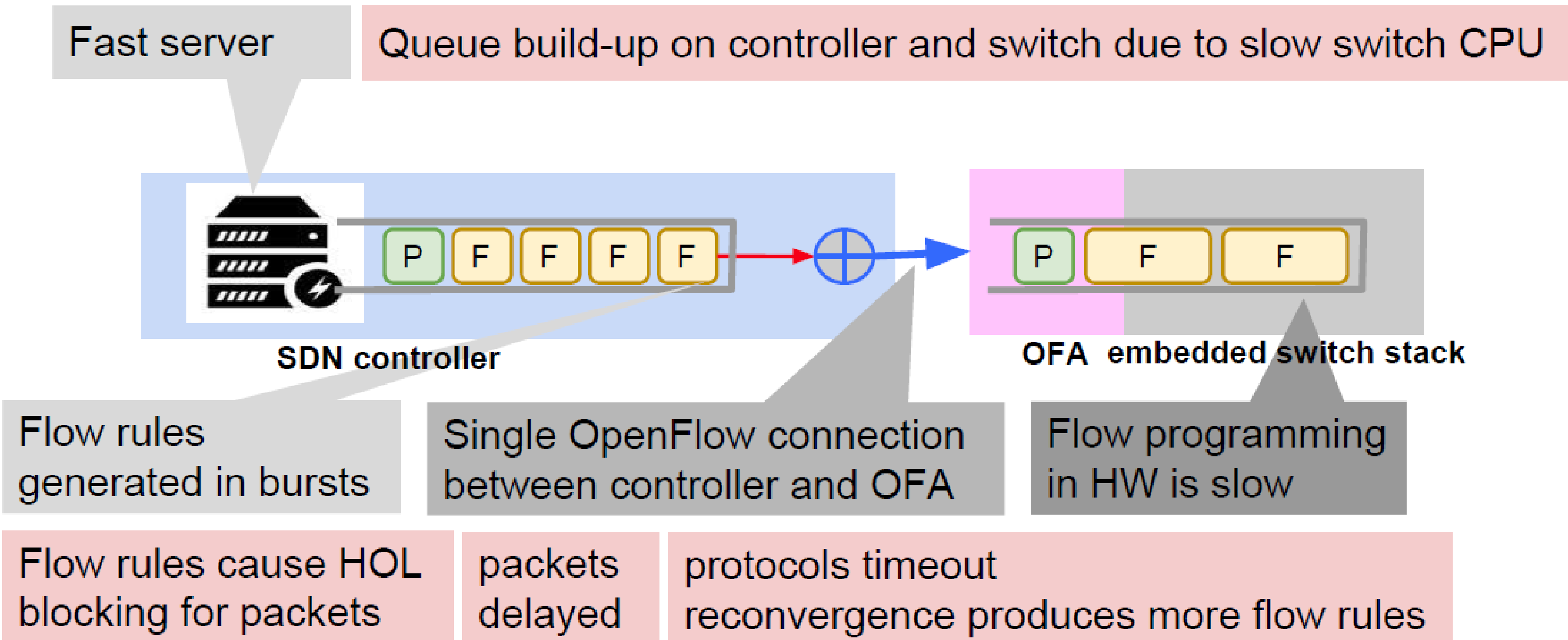
TE optimization algorithm

- Target: Achieve max-min fairness
  - Approximate solution with bandwidth function (See the paper for details)
    - **Tunnel Selection** selects the tunnels to be considered for each flow group
    - **Tunnel Group Generation** allocates bandwidth to FGs using bandwidth functions to prioritize at bottleneck links
    - **Tunnel Group Quantization** changes split ratios in each FG to match the granularity supported by switch hardware tables.

B4 SDN architecture

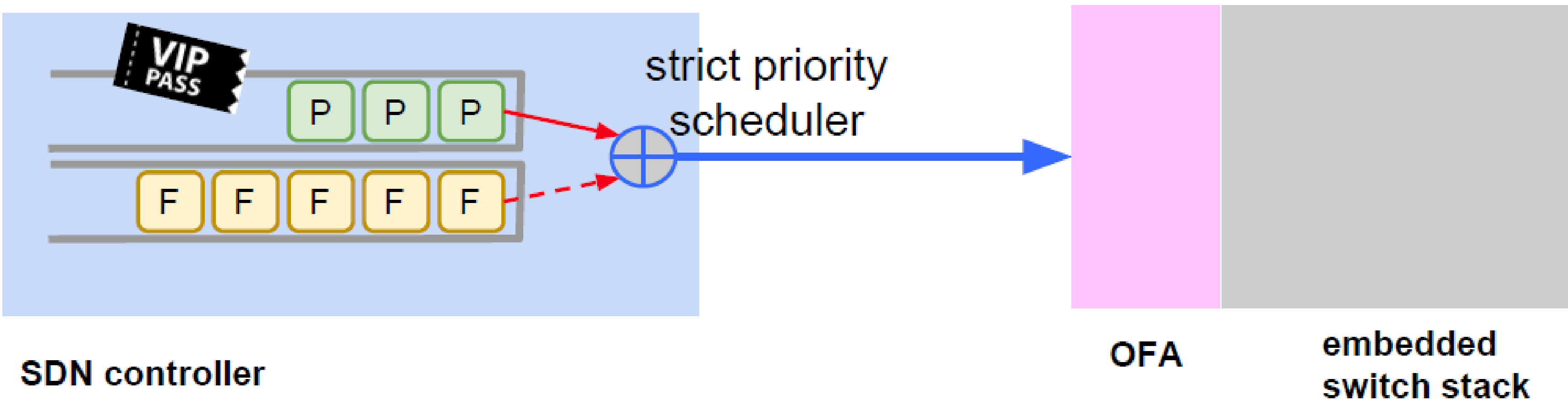
- Switch hardware (Google custom designed with commodity silicon)
  - Forwards traffic
  - No complex control software
- OpenFlow controllers (ONIX based)
  - Maintain network state based on network control application directive and switch events
  - Instruct switches to set forwarding entries
- Central application
  - Central control of the entire network

Message backlogged and delayed



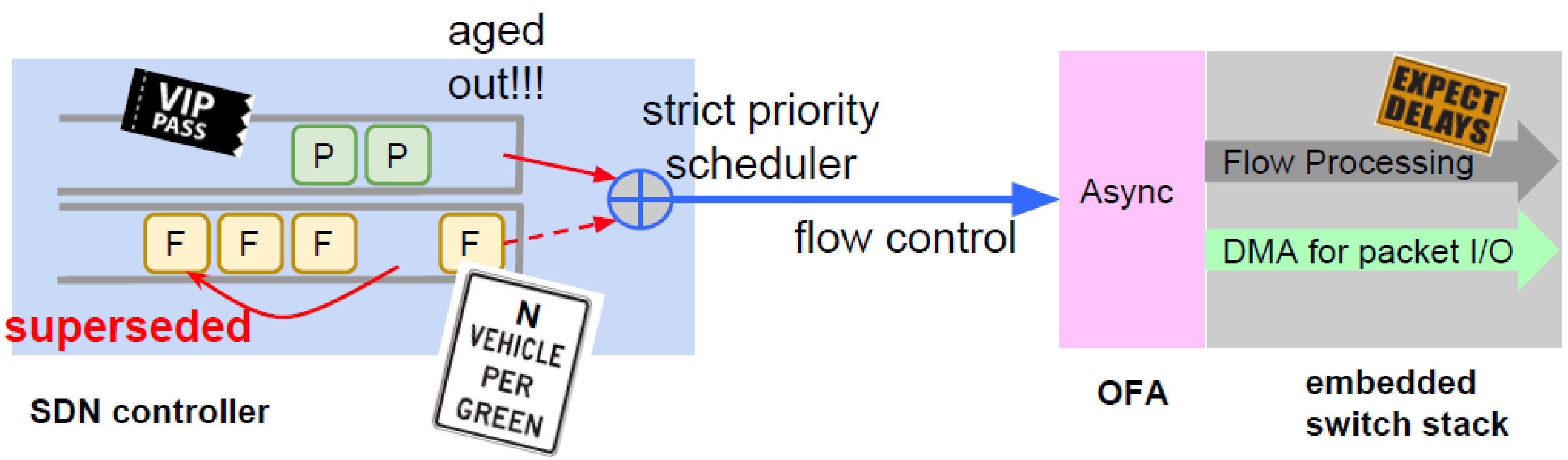
Vicious cycle of protocol instability

Lesson: Mitigation with flow control



- Separate queue for packet IO and flow request
- Strict priority for packet IO over flow programming

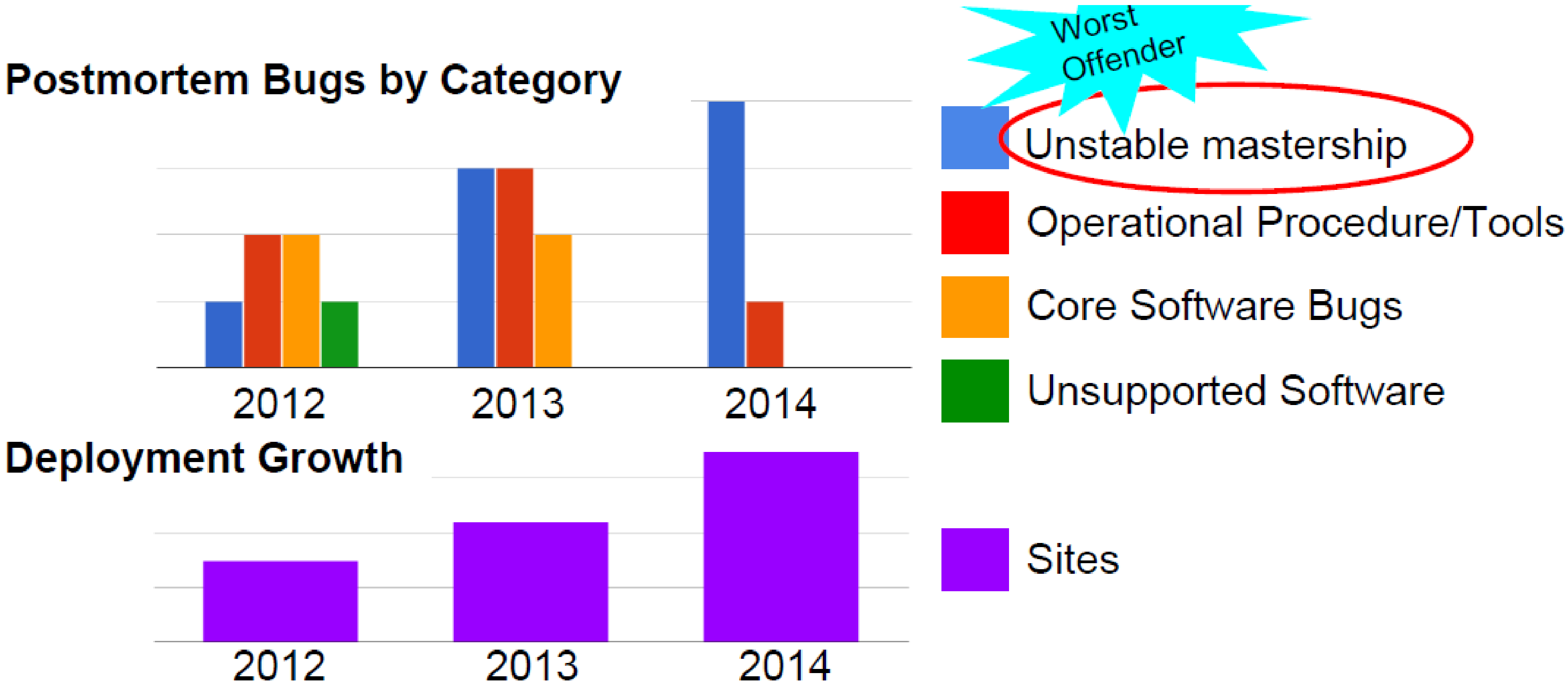
Lesson: Mitigation with flow control



- Separate queue for packet IO and flow request
- Strict priority for packet IO over flow programming
- Limit queue depth in OFA: token based flow control
- Systematics queue drop discipline

- Asynchronous OFA
- Packet IO out of flow processing pipeline

Issues with B4 from availability: Outages



Issues with B4 on scaling

- Flat Topology scales poorly
  - As B4 grows: more sites deployed
  - As compute per site grows: more capacity required per site

- Solving capacity asymmetry problem in hierarchical topology is key to achieve high availability at scale

- Larger switches or more switches
  - Larger switches: loss of large capacity on switch failure
  - More switches: more nodes and links to manage
    - TE will hit scaling issues, converge too slowly.

Challenges?

- High Availability Requirements
- Example applications with different service level objectives (SLO)
  - Search ads, DNS, WWW
  - Proto service backend, Email
  - Ads database replication
  - Search index copies, logs
  - Bulk transfer

Challenges?

Key pain points in the SD-WAN industry

LACK OF:

- A common definition of a SD-WAN framework & services
- Single service orchestration of multiple SD-WAN vendor implementations
- A certification framework for baseline SD-WAN compliance

Industry associations.

e.g., a global industry association of network, cloud, and technology providers.

Conclusions

- SDN is driving significant changes in how networks are **architected, managed and updated**
- "True" SDN deployments have mostly been in carrier or very large enterprise environments where there is a return on the engineering investment
- SD-WAN has seen a **rapid increase in enterprise adoption** but is only loosely related to the textbook definition of SDN; it is delivering a different set of benefits in many cases
- Virtualization in the network is **long overdue** – it is one of the last components of the IT stack to see this change
- Network-based services will become **increasingly sophisticated** as enterprises adopt more SD-WAN at the edge, combined with other NFVs for common functions

Other Applications

There are many!

- Cloud management
- Home network
- Content Delivery Networks
  - Video content distribution

- Mobile and Wireless SDN
  - Hot research direction

Thank you!

References:

https://www.scs.gatech.edu/news/195201/free-online-sdn-course

https://www.sdxcentral.com/sdn/?c_action=num_ball

https://www.opennetworking.org/

THE UNIVERSITY OF SYDNEY