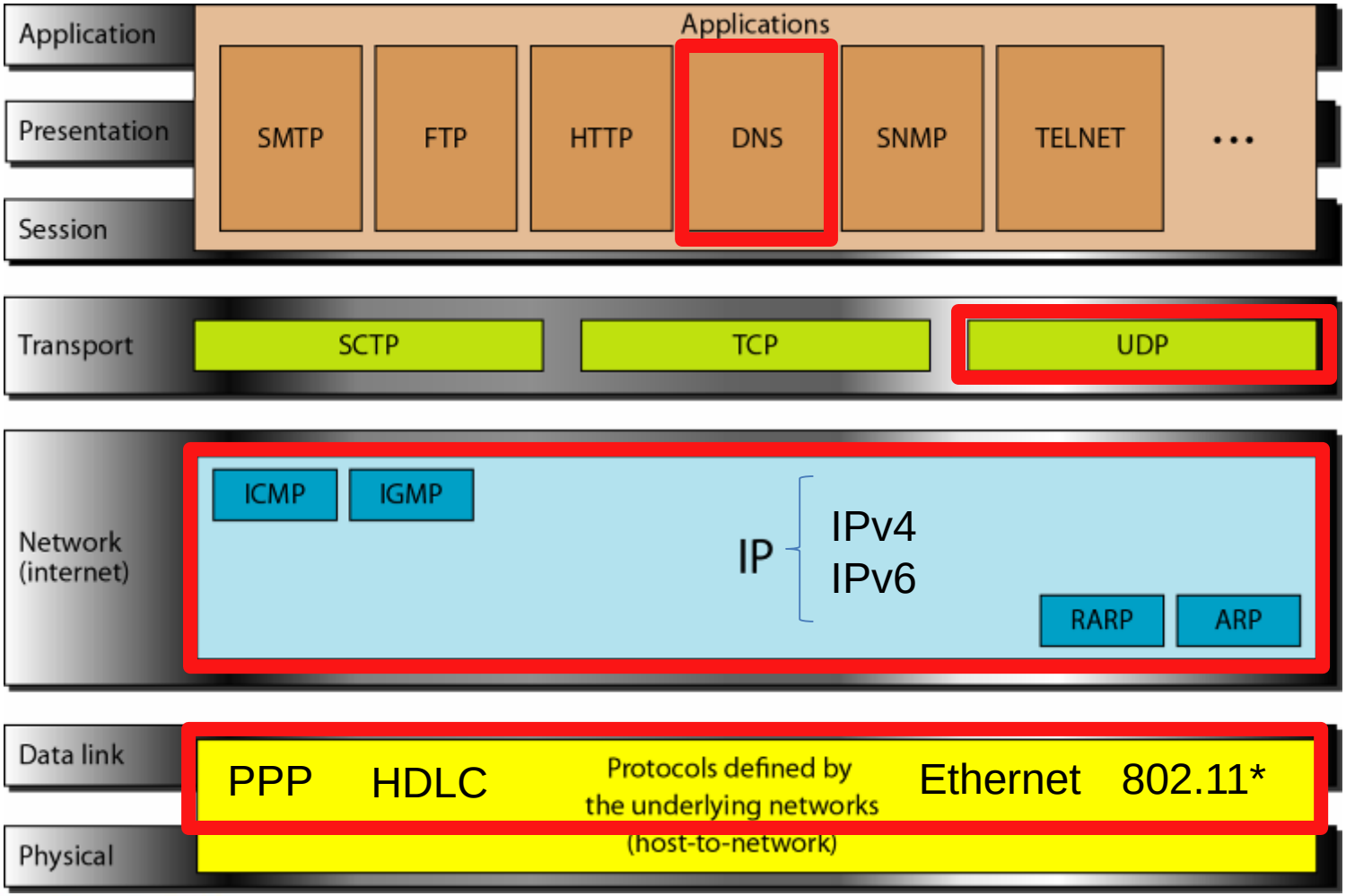
Overview

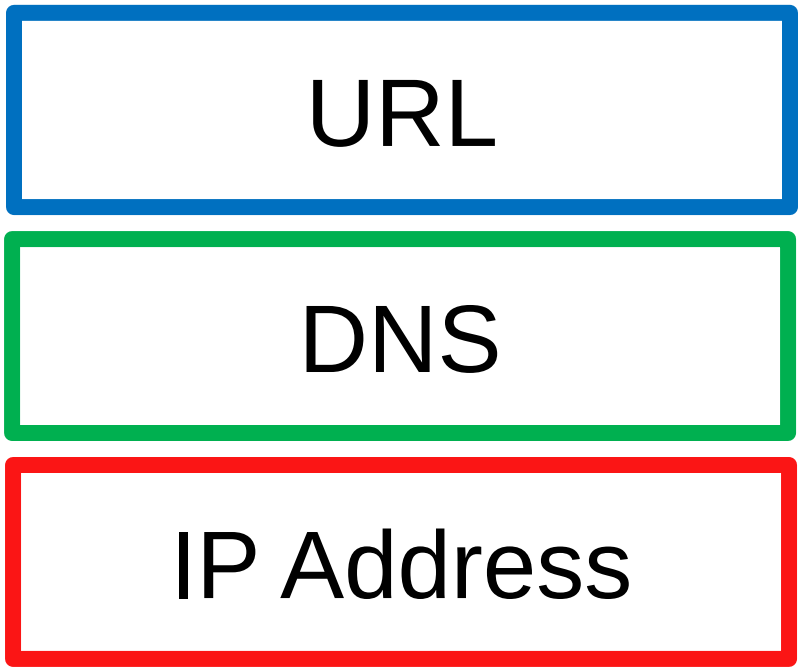
- Link Layer
- Network Layer
  - Addressing
  - Address Resolution (ARP)
  - Fragmentation
  - Intra-AS Routing
    - Distance Vector
    - Link State
  - Multicast Routing
  - IPv6
- Transport Layer
  - UDP
  - DNS

Protocols in the OSI Model



* Figure is courtesy of B. Forouzan

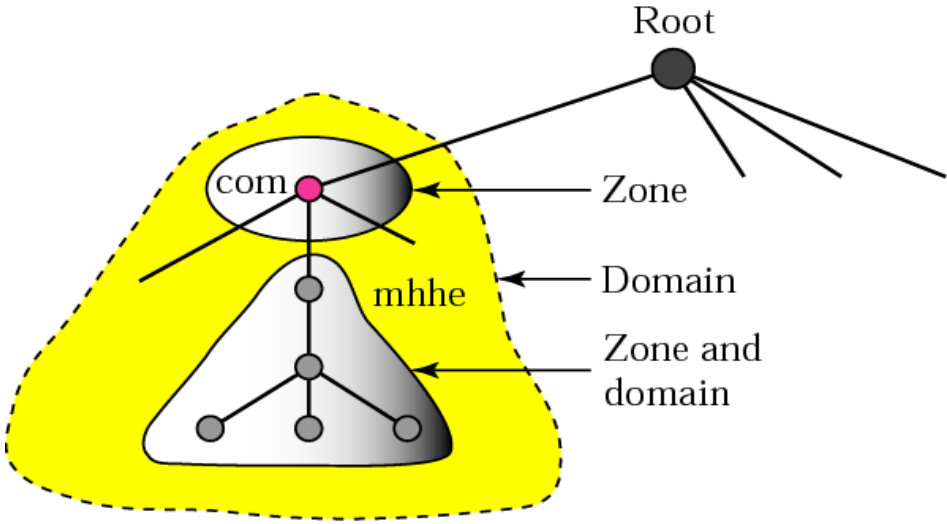
URLs to Names to Addresses

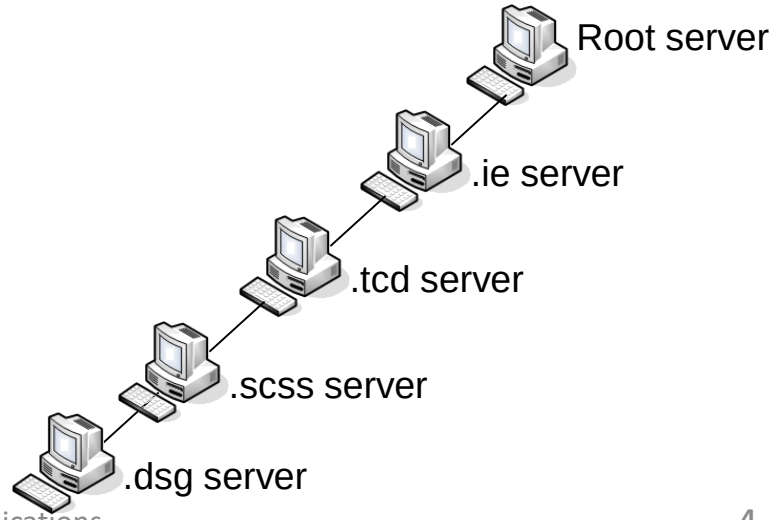


*URL = Uniform Resource Locator

Hierarchy of Name Servers

- Every zone has a DNS server
- DNS server maintain lists of
  - Nodes in the zone
  - References to servers of zones underneath it

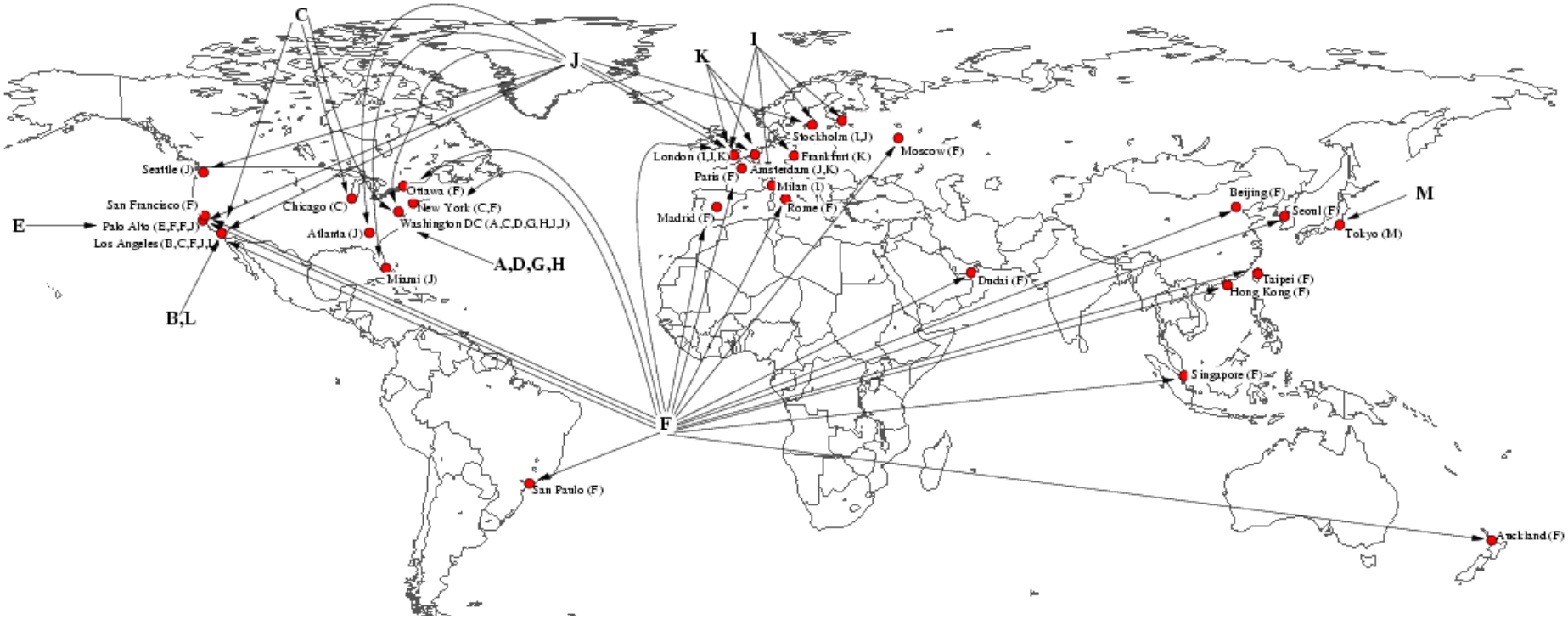




www IN A 134.226.36.14

* Figure is courtesy of B. Forouzan

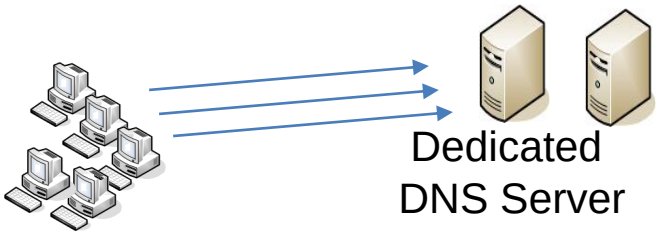
DNS Root Servers - Anycast

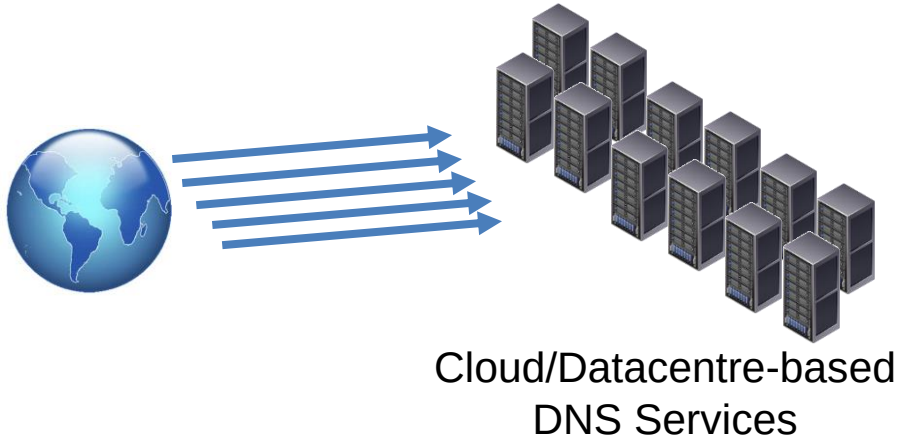


* Figure is courtesy of CAIDA

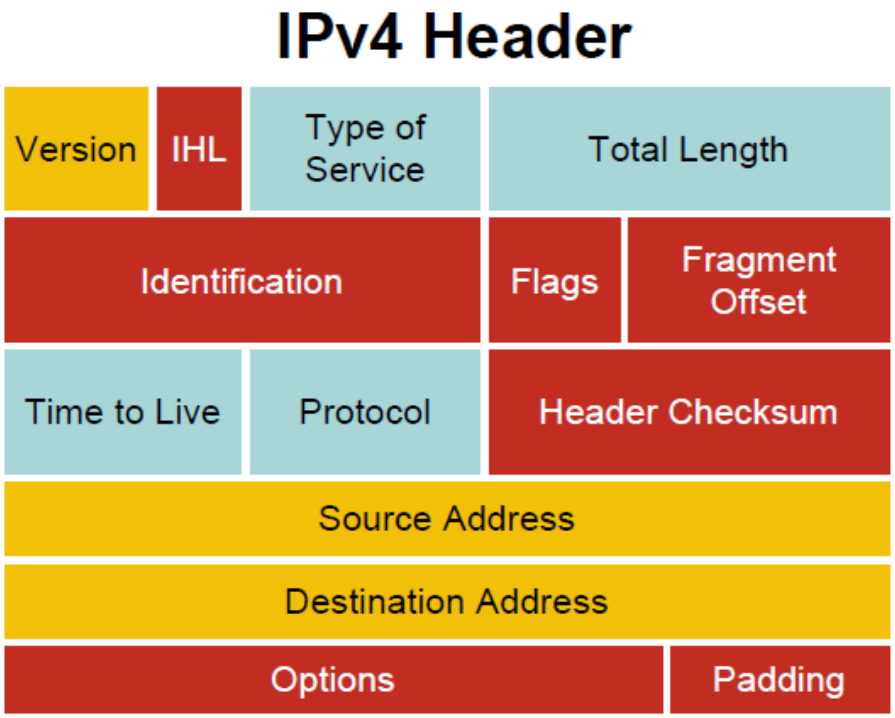
Mirai Bot & The Attack on Dyn

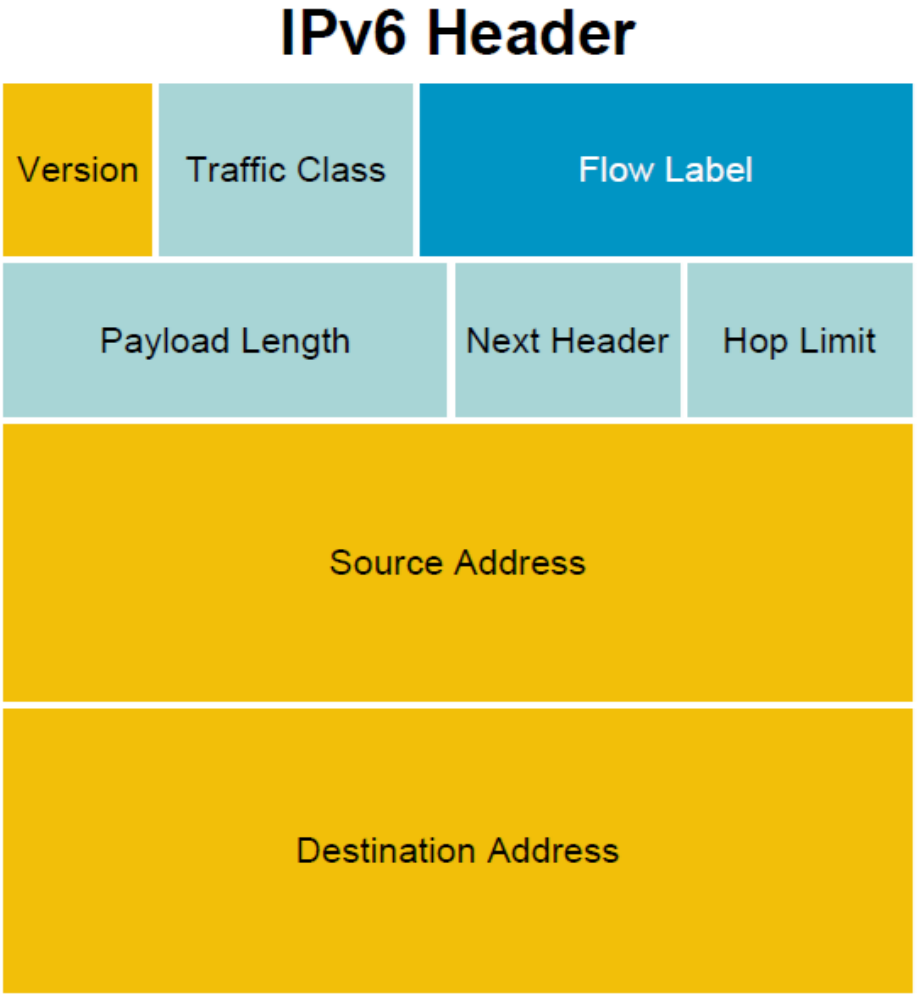
- Good example for scalable DNS services
- … and the vulnerability of these services

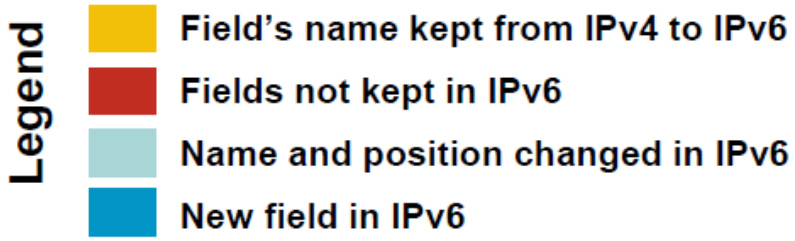




Header Comparison

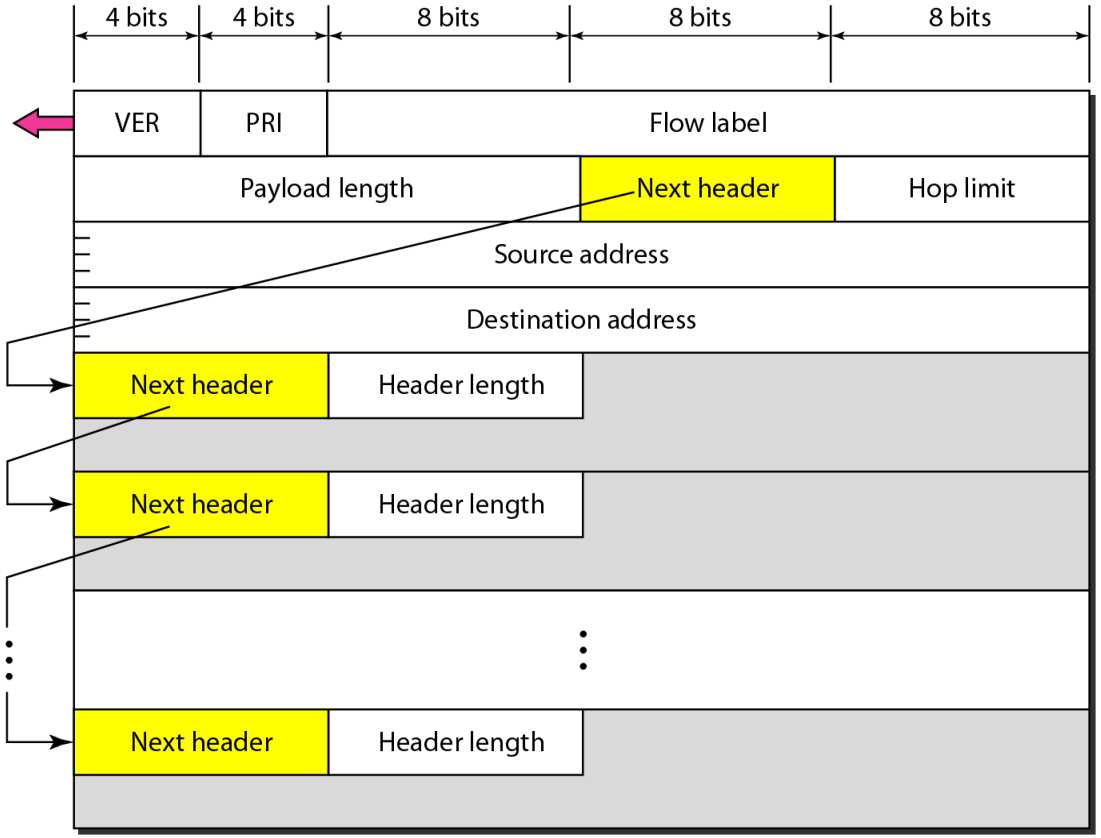






* Figure is courtesy of P. Smith, Cisco

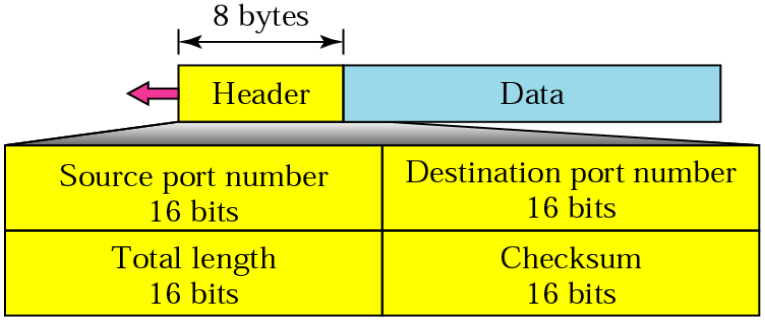
Extension Headers

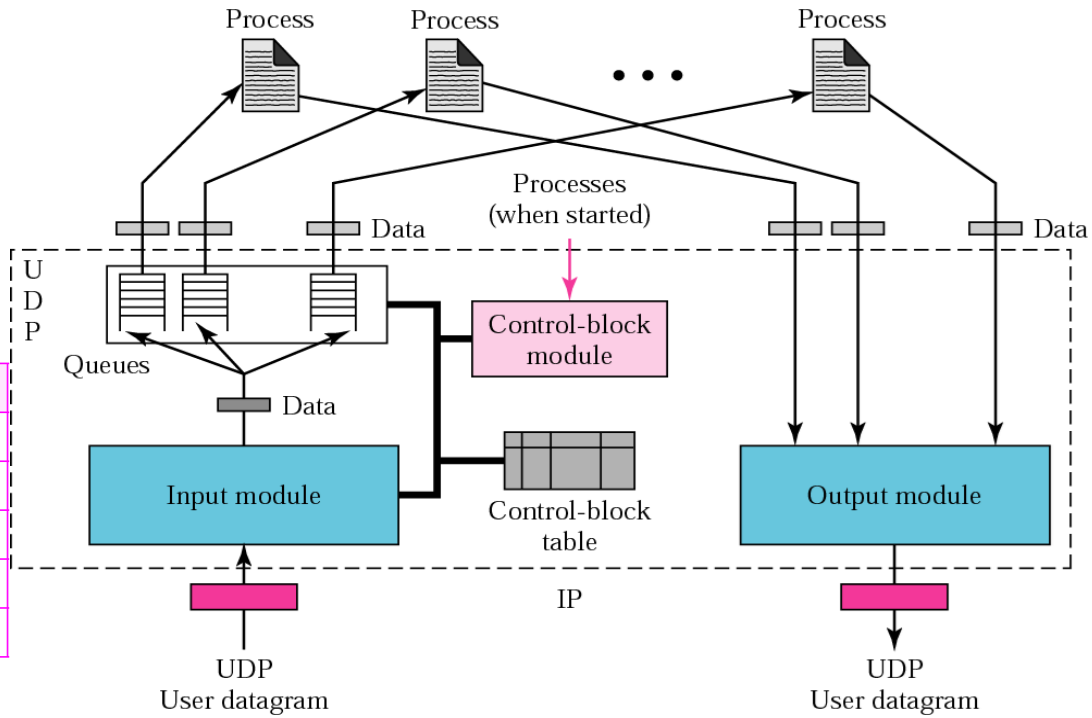


* Figure is courtesy of B. Forouzan

UDP – Portnumber id's Queues

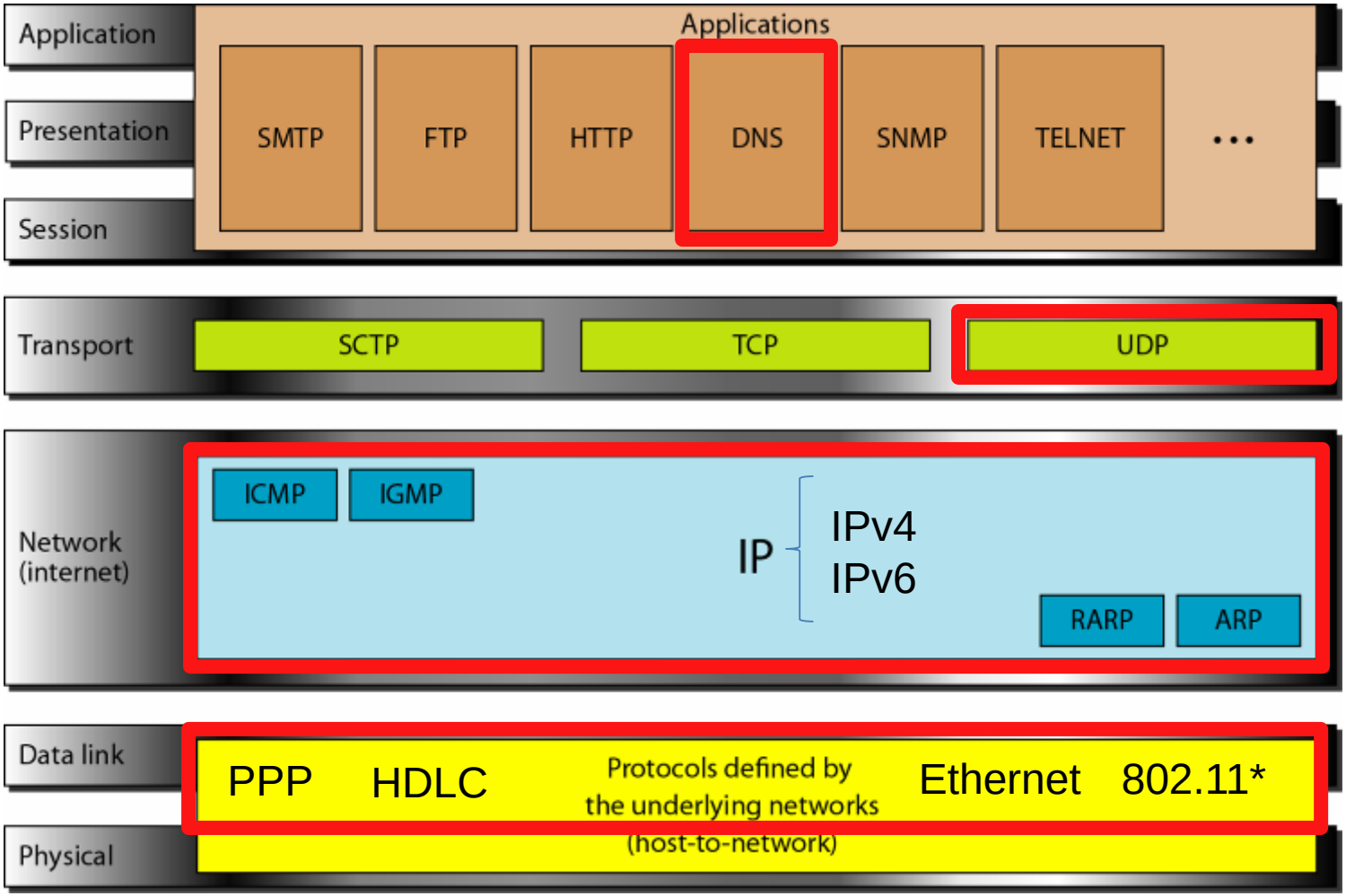
- Connectionless, unreliable protocol
  - No flow and error control
  - Port numbers are used to multiplex data





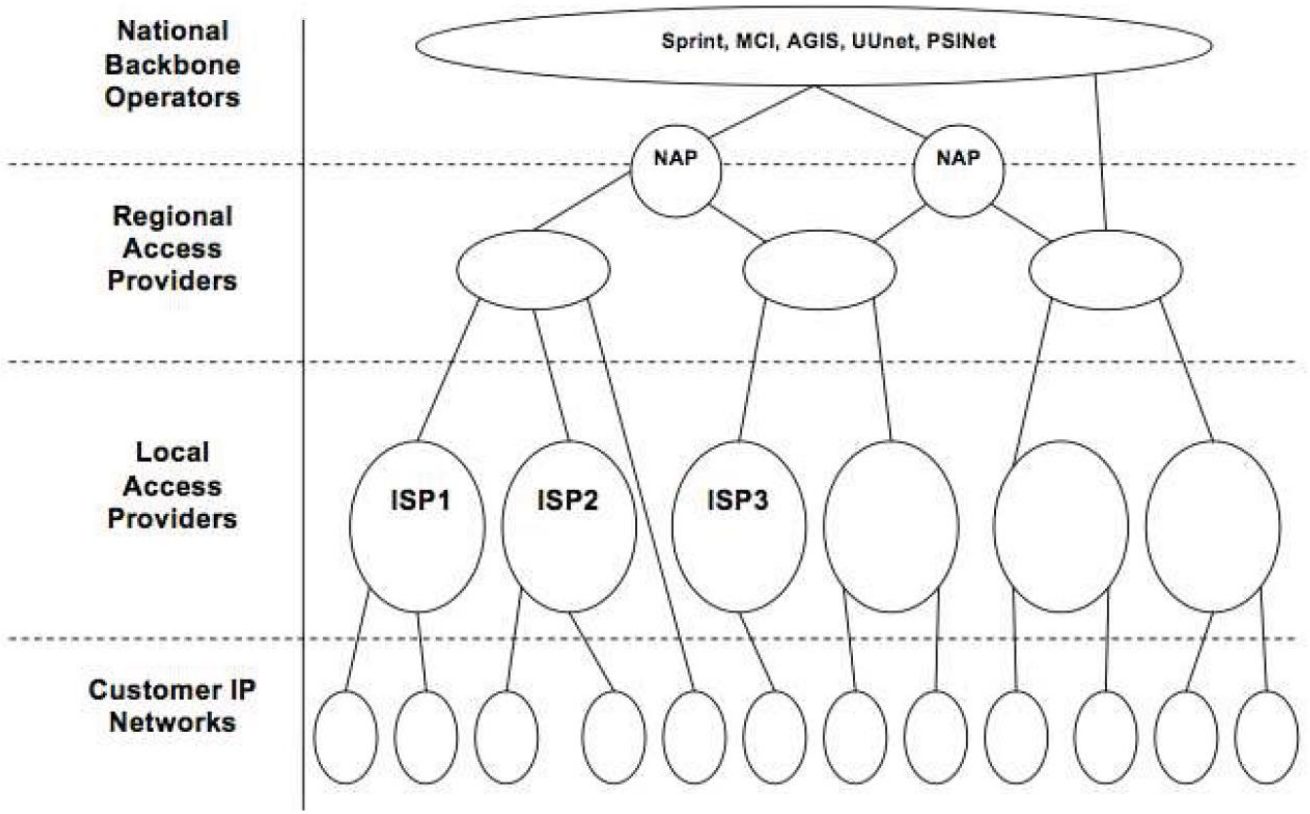
State	*Process ID*	*Port Number*	*Queue Number*
IN-USE	2,345	52,010	34
IN-USE	3,422	52,011	
FREE			
IN-USE	4,652	52,012	38
FREE			

Protocols in the OSI Model



* Figure is courtesy of B. Forouzan

Traditional Logical Internet Topology



* Figure is courtesy of C. Labovitz, et al, SIGCOMM 2010

CS Predictions

- "I think there is a world market for maybe five computers."

  Thomas Watson, President of IBM, 1943

- "There is no reason anyone would want a computer in their home."

  Ken Olsen, Founder of Digital Equipment Corporation, 1977

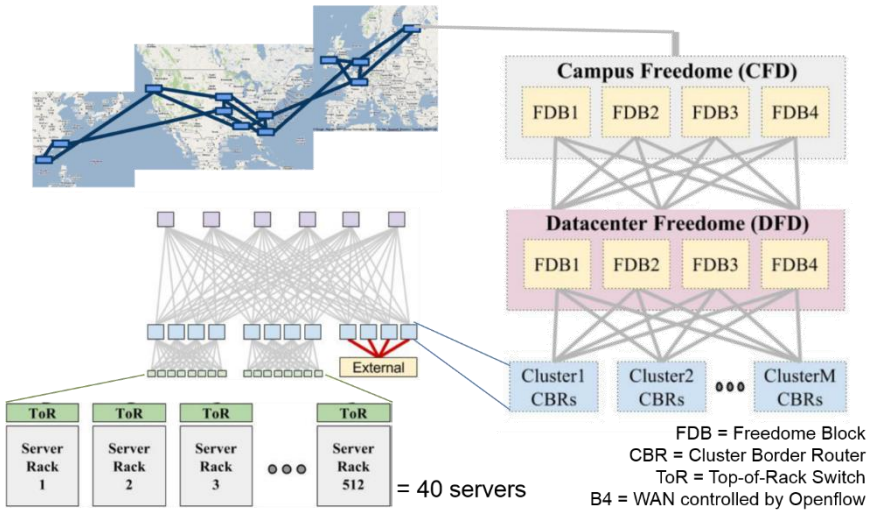
- "I predict the Internet in 1996 will catastrophically collapse."

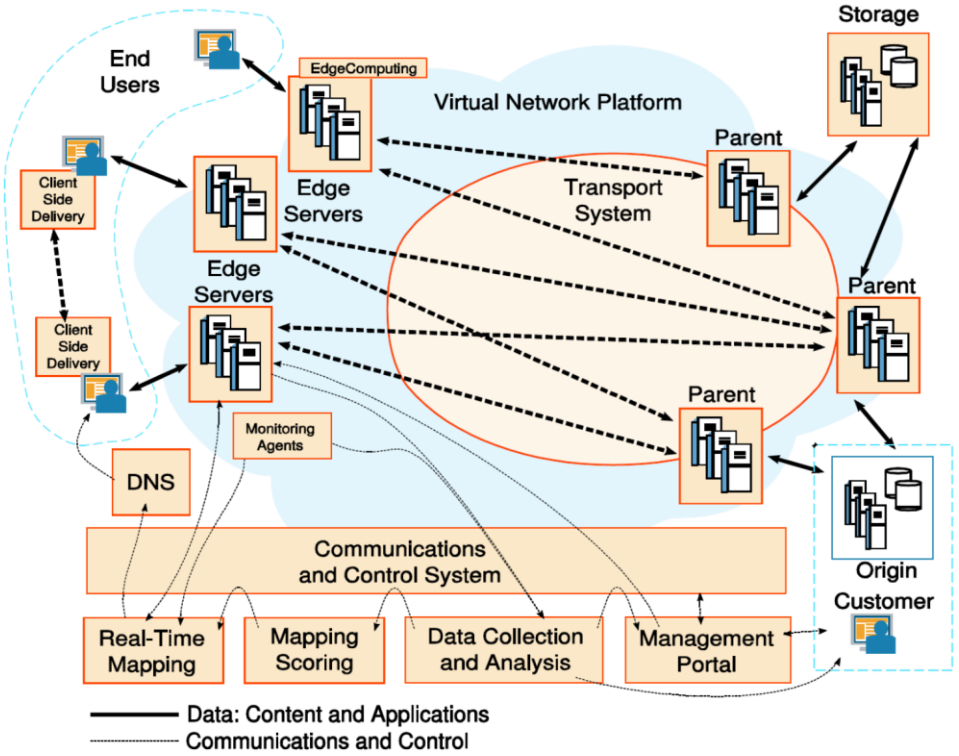
  Robert Metcalfe, 1995

- "IPv6 is dead."

  David Cheriton, 1999

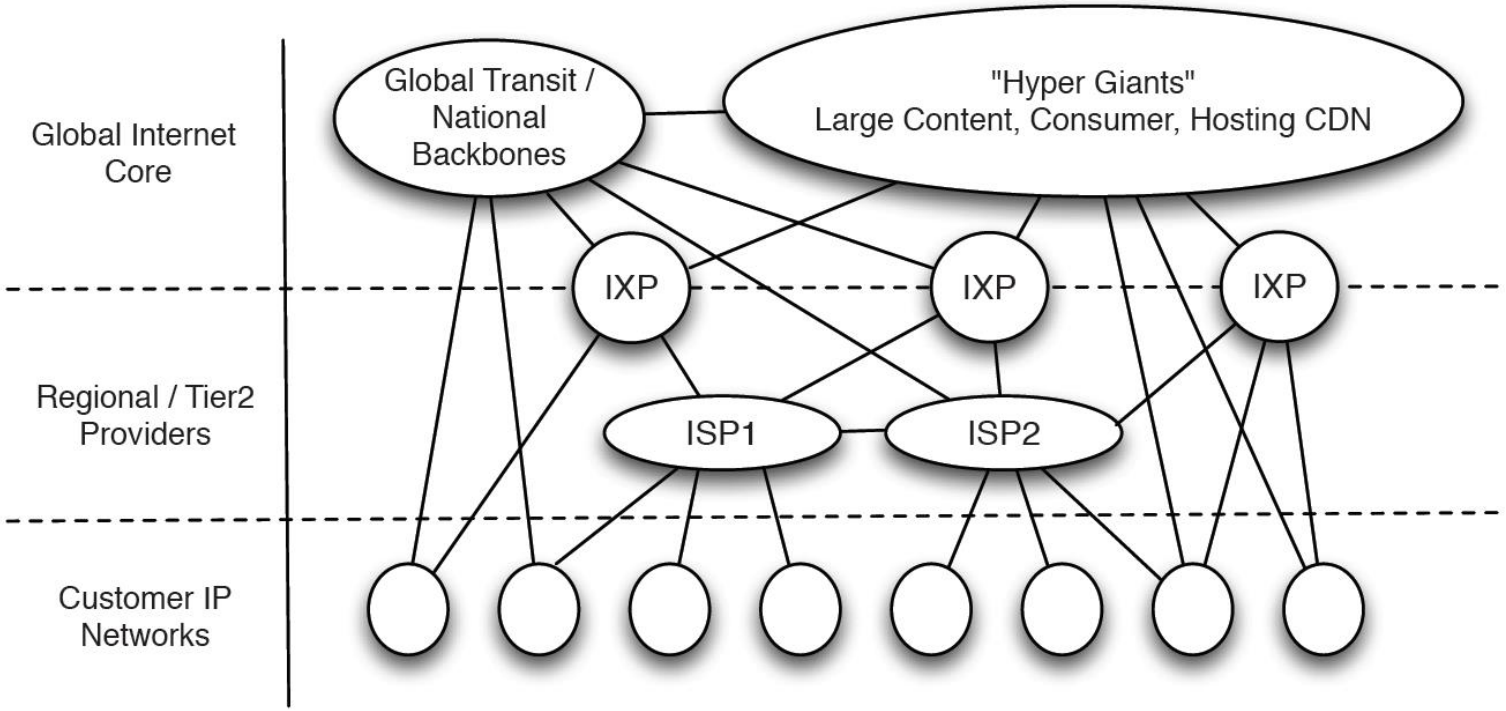
Hyper-Giants





- Google
- Akamai

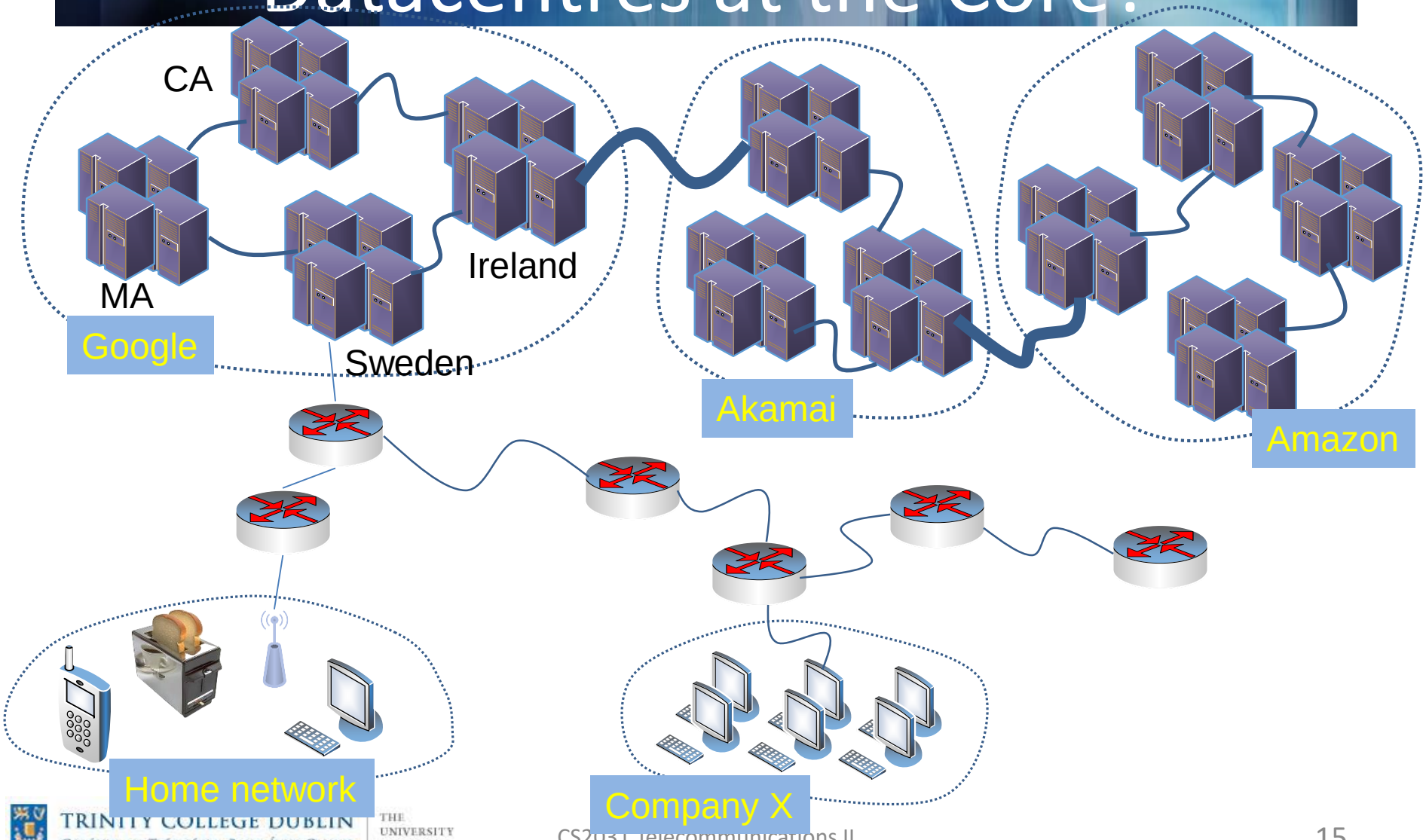
Emerging Logical Internet Topology



- According to a statement by Craig Labovitz in 2014:
  - 30 Entities created 50% of the traffic in the US at peak time

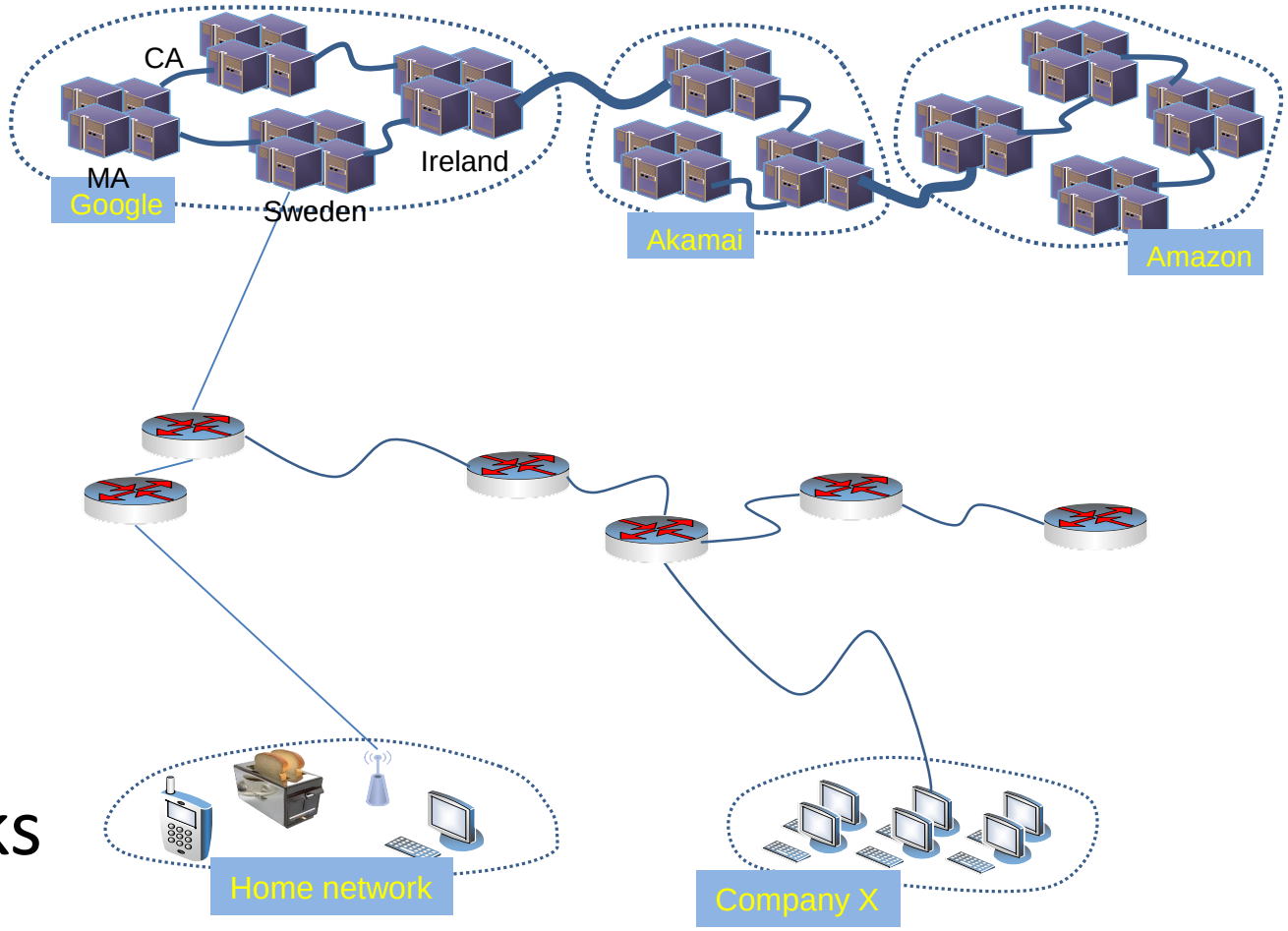
* Figure is courtesy of C. Labovitz, et al, SIGCOMM 2010

Datacentres at the Core?



My view of "future"* networking

- Sets of Datacentres
- Traditional Internet
- Edge networks



*or current?

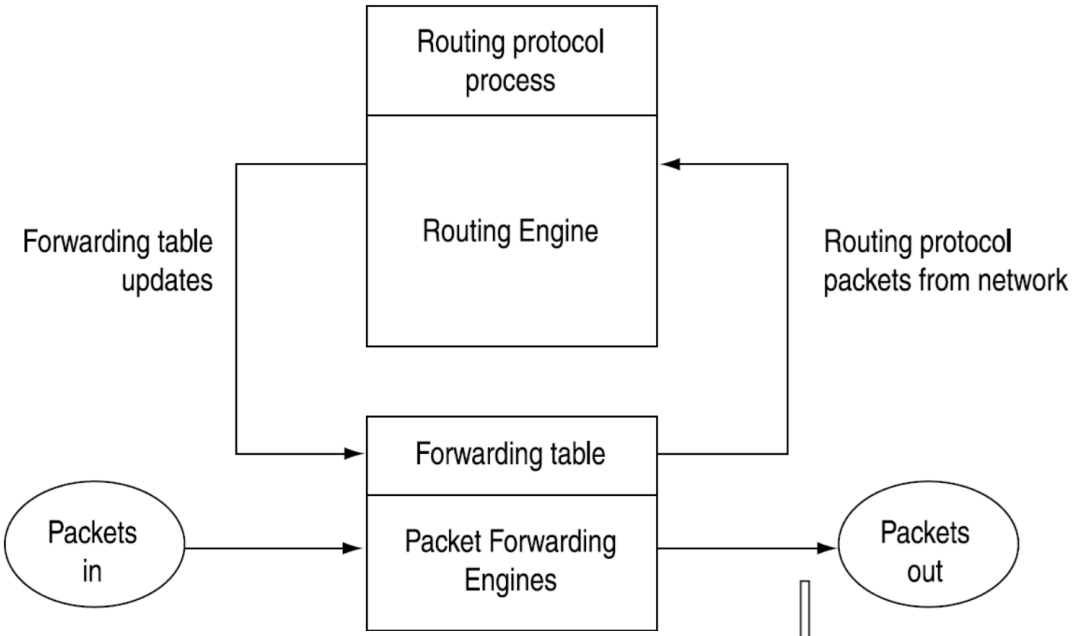
Overview

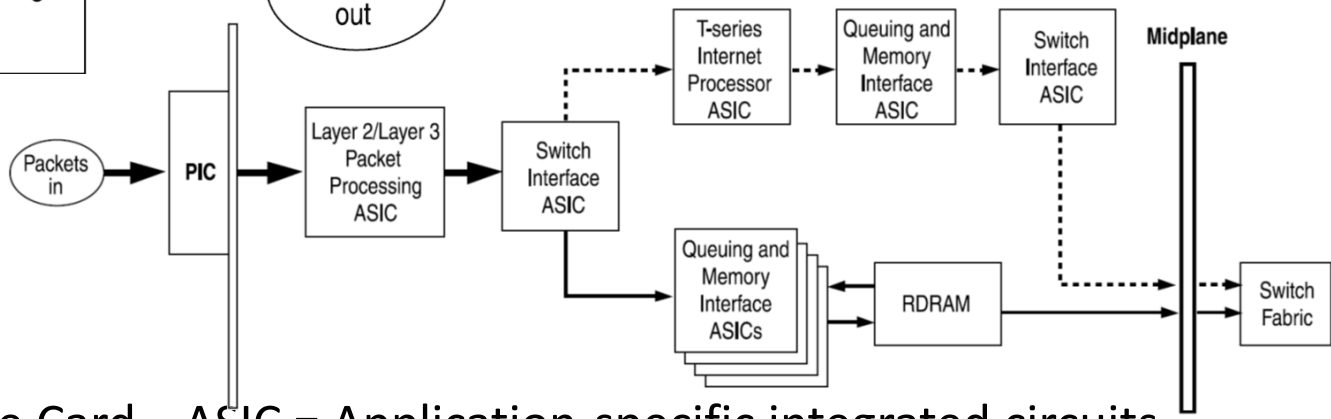
- Link Layer
- Network Layer
  - Addressing
  - Address Resolution (ARP)
  - Fragmentation
  - Intra-AS Routing
    - Distance Vector
    - Link State
  - Multicast Routing
  - IPv6
- Transport Layer
  - UDP
  - DNS
- Software-Defined Networking / Openflow
- CLOS / Fat-tree
- ATM/MPLS

CS2031
Telecommunications II

SDN & Openflow

Switch/Router Internals

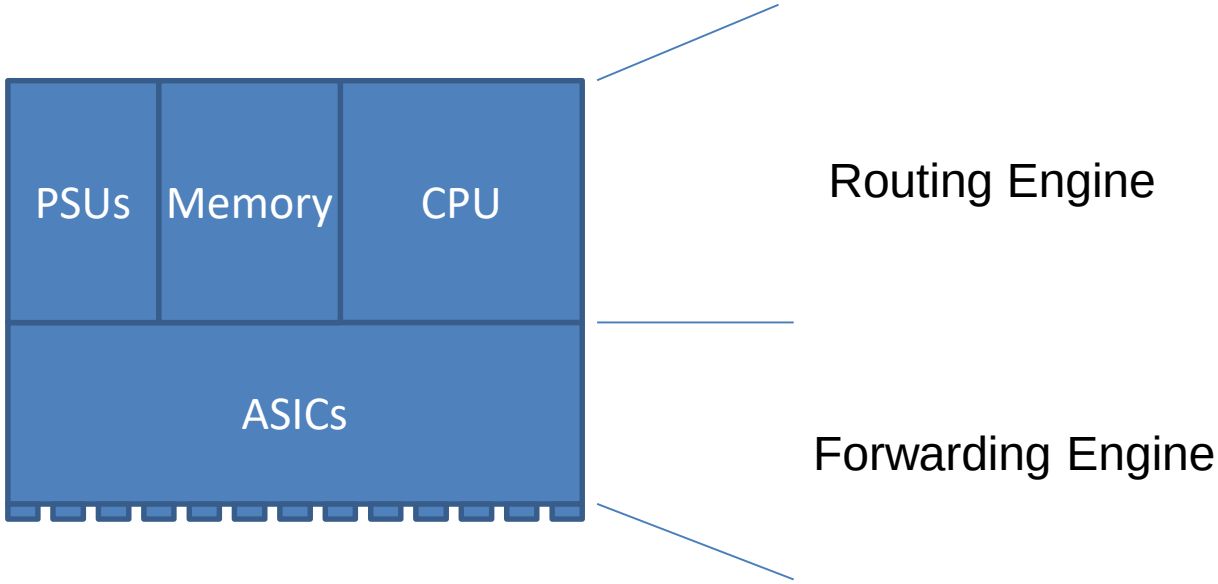




PIC = Physical Interface Card ASIC = Application-specific integrated circuits

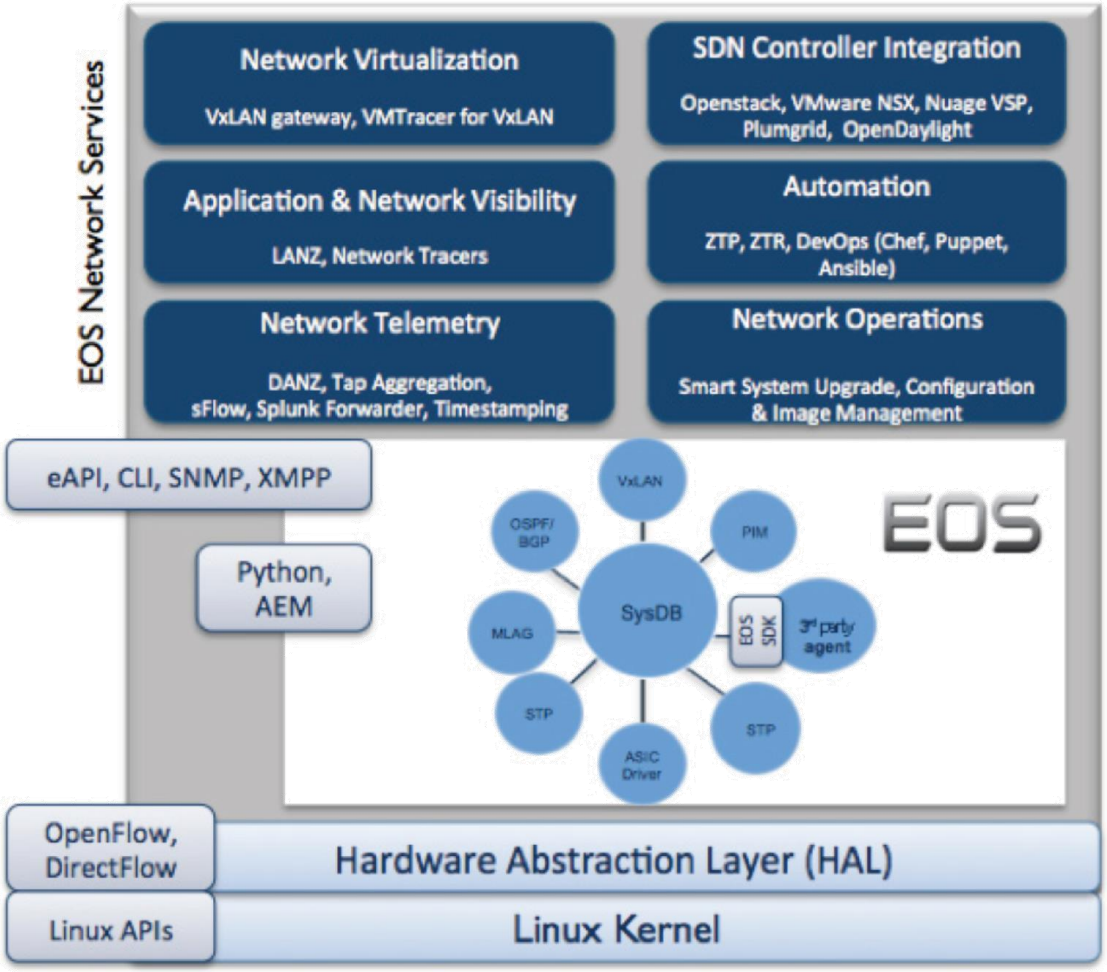
* Figure is courtesy of Juniper Networks, T1600 Core Router Hardware Guide

Switches/Routers



*Arista, EOS Whitepaper, 2016

Example Switches/Router OS

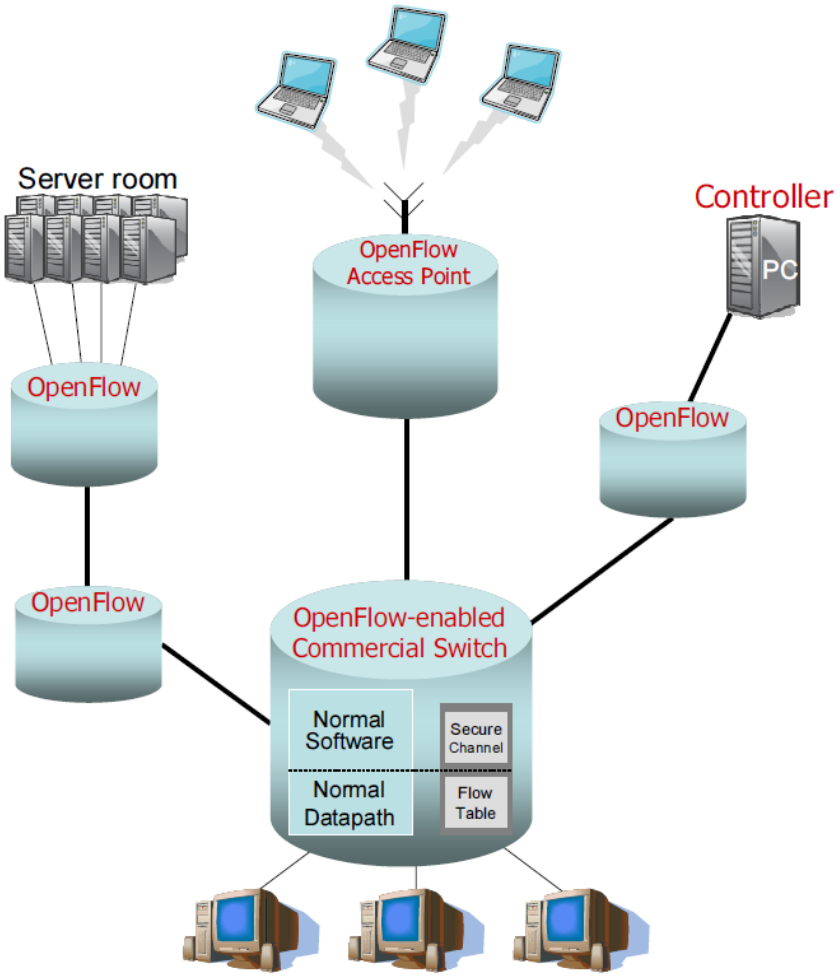


EOS Architecture

*Arista, EOS Whitepaper, 2016

Original Openflow Paper

- Openflow switches
- Controller w/ secure connection
- Configurable flow tables



OpenFlow: Enabling Innovation in Campus Networks

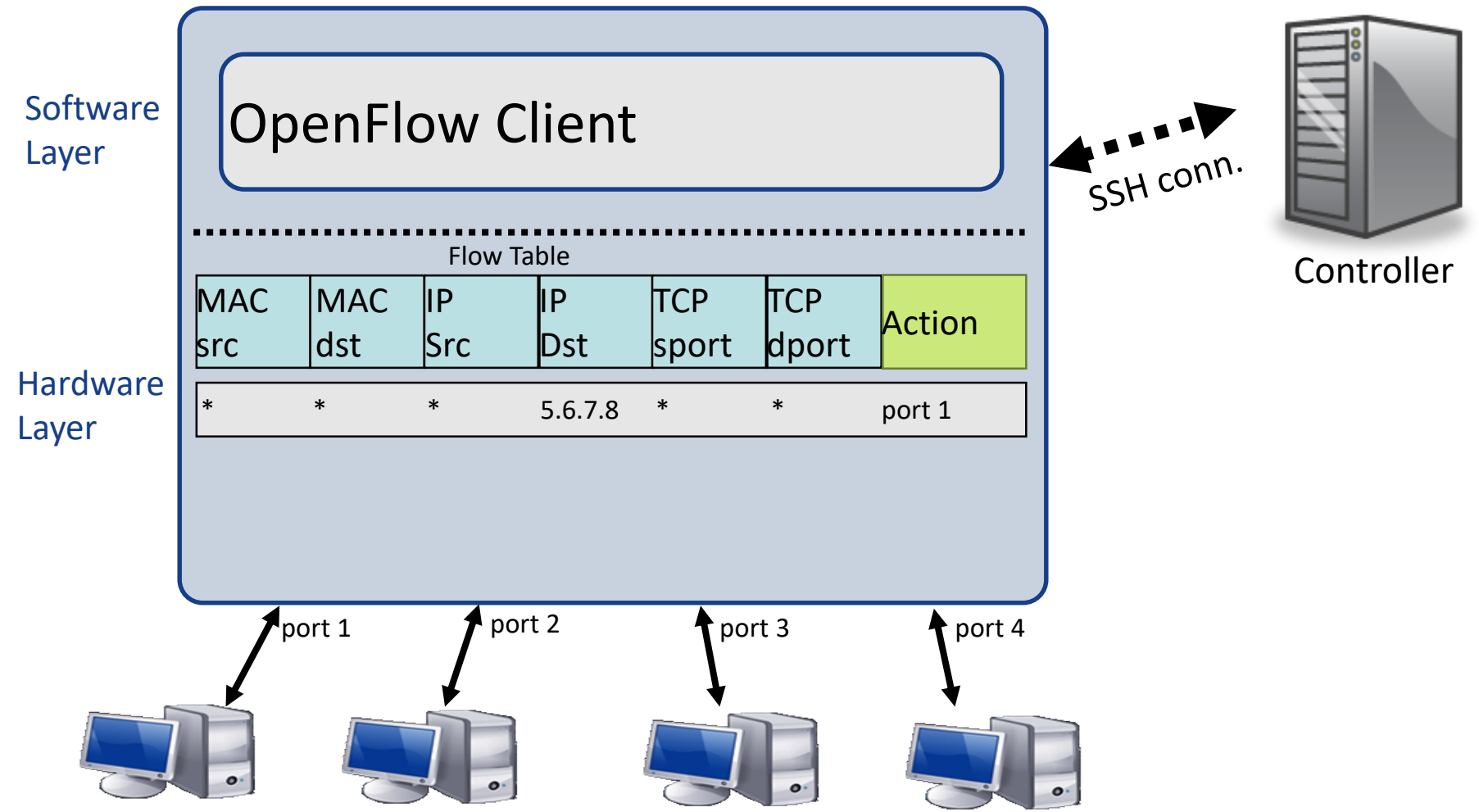
March 14, 2008

Nick McKeown, Stanford University
Tom Anderson, University of Washington
Hari Balakrishnan, MIT
Guru Parulkar, Stanford University
Larry Peterson, Princeton University
Jennifer Rexford, Princeton University
Scott Shenker, University of California, Berkeley
Jonathan Turner, Washington University in St. Louis

Header of 2008 Paper

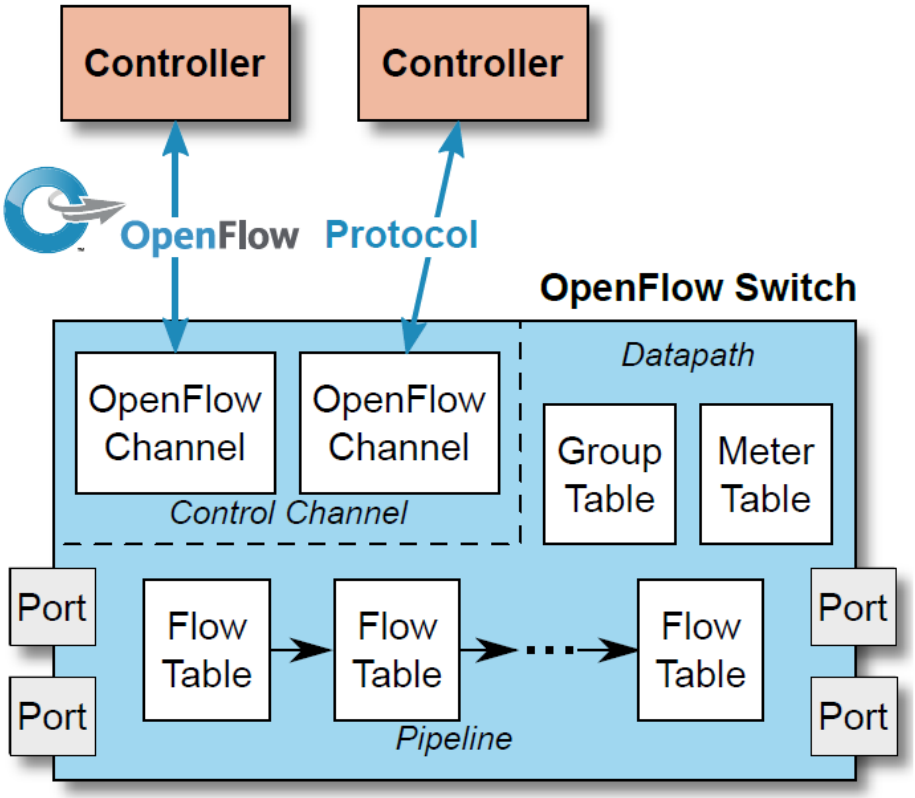
•Nick McKeown et al, SIGCOMM 2008

Openflow Switch



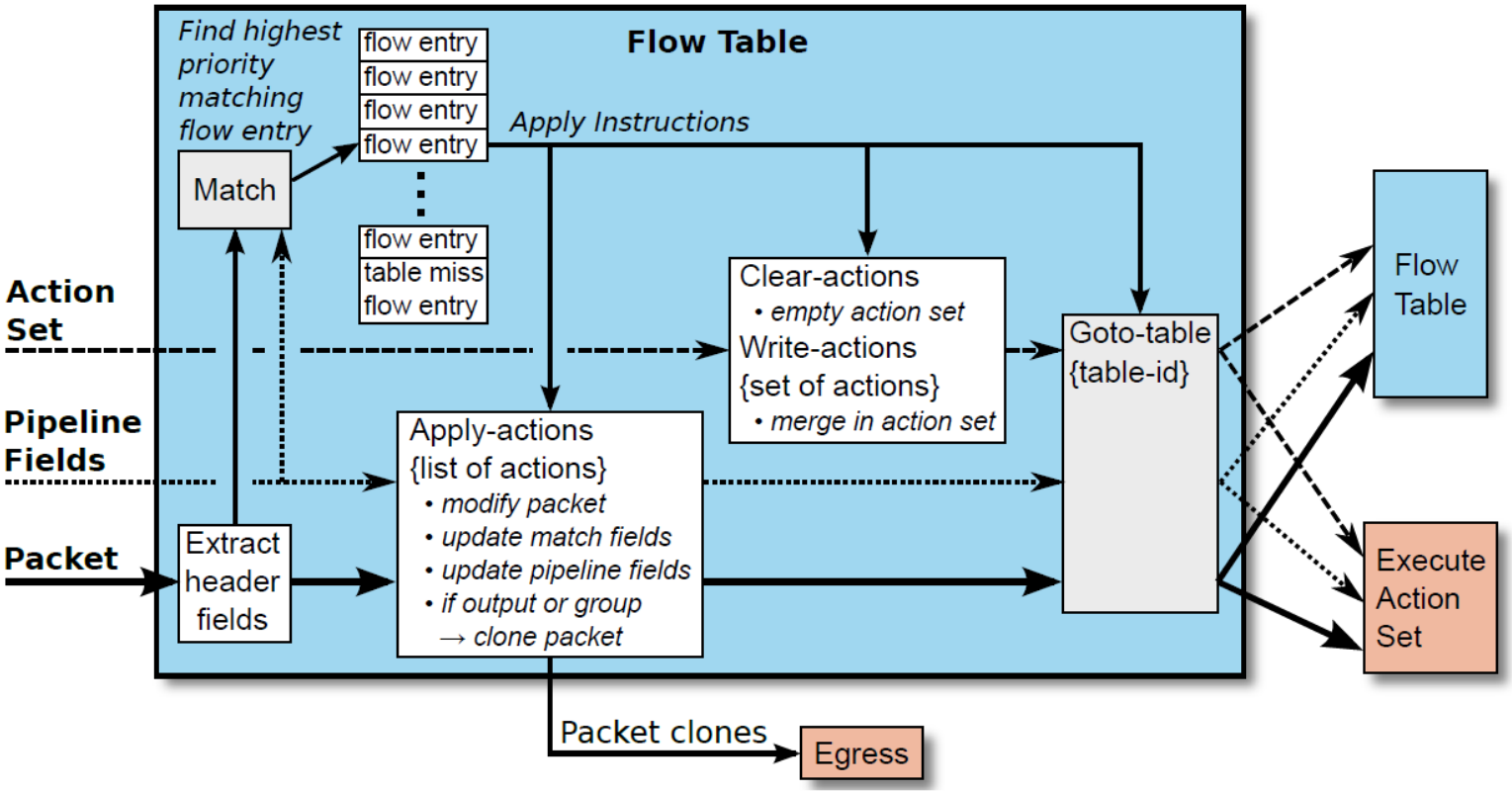
•Yiannis Yiakoumis, et al, HOT-Interconnects 2010

Components of Openflow Switches



OpenFlow Switch Specification,v1.5.1, Open Net. Found.

Flowtable Processing



OpenFlow Switch Specification,v1.5.1, Open Net. Found.

Counters in Flowtables

Counter	Bits	
Per Flow Table		
Reference Count (active entries)	32	*Required*
Packet Lookups	64	*Optional*
Packet Matches	64	*Optional*
Per Flow Entry		
Received Packets	64	*Optional*
Received Bytes	64	*Optional*
Duration (seconds)	32	*Required*
Duration (nanoseconds)	32	*Optional*
Per Port		
Received Packets	64	*Required*
Transmitted Packets	64	*Required*
Received Bytes	64	*Optional*
Transmitted Bytes	64	*Optional*
Receive Drops	64	*Optional*
Transmit Drops	64	*Optional*
Receive Errors	64	*Optional*
Transmit Errors	64	*Optional*
Receive Frame Alignment Errors	64	*Optional*
Receive Overrun Errors	64	*Optional*
Receive CRC Errors	64	*Optional*
Collisions	64	*Optional*
Duration (seconds)	32	*Required*
Duration (nanoseconds)	32	*Optional*
Per Queue		

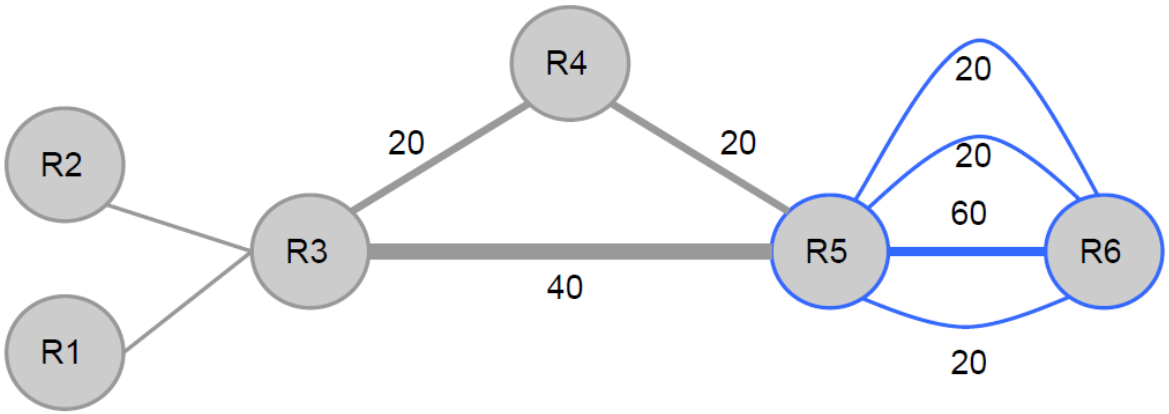
OpenFlow Switch Specification,v1.5.1, Open Net. Found.

Why????

As in: Why Openflow? ☺

Traditional Net. Example

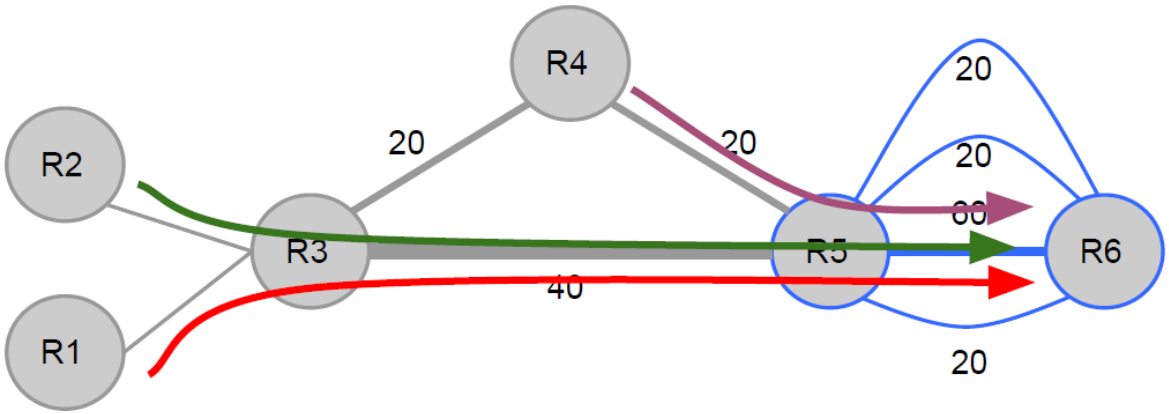
- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20



•Urs Hoelzle, Open Network Summit 2012

Traditional Net. Example

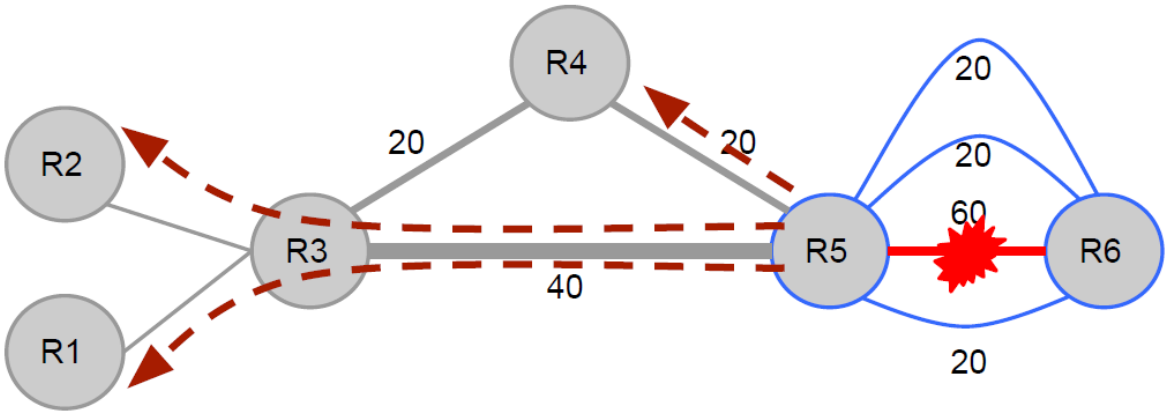
- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20



•Urs Hoelzle, Open Network Summit 2012

Traditional Net. Example

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

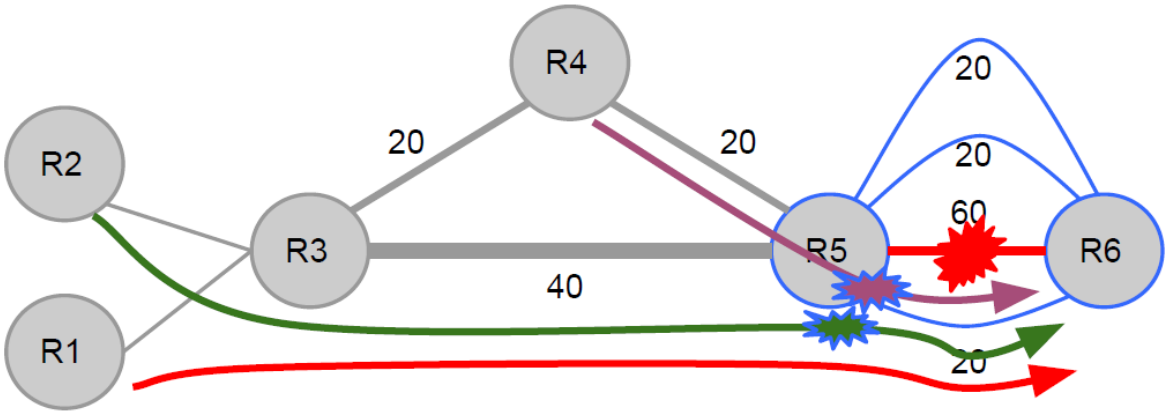


- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path

•Urs Hoelzle, Open Network Summit 2012

Traditional Net. Example

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

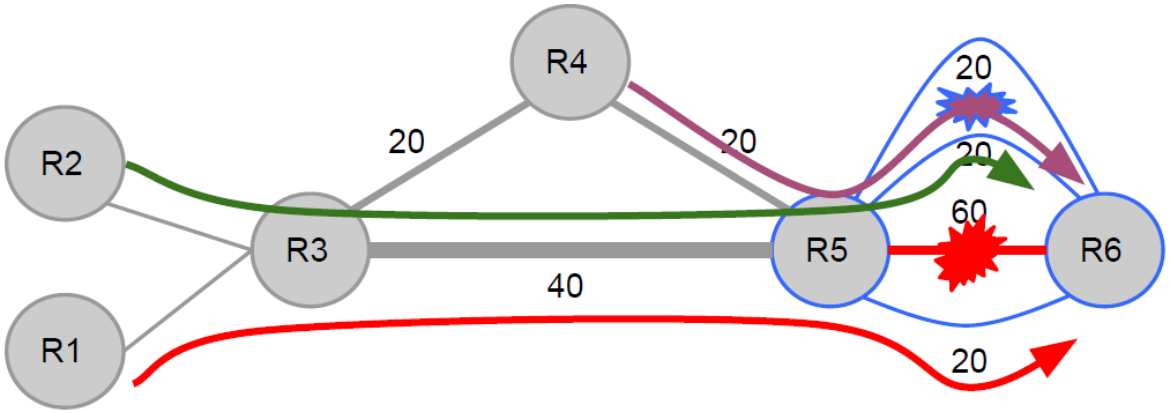


- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1 wins, R2, R4 retry for next best path

•Urs Hoelzle, Open Network Summit 2012

Traditional Net. Example

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

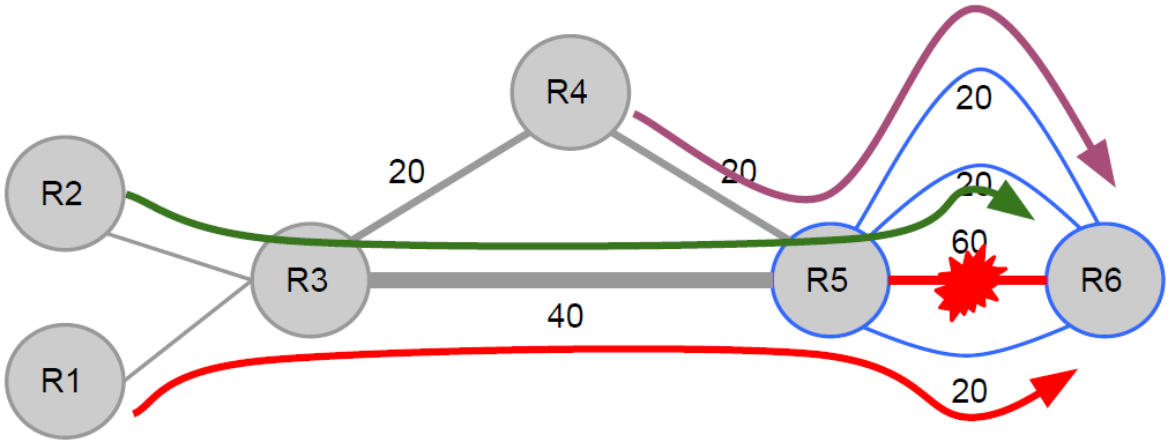


- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1 wins, R2, R4 retry for next best path
  - R2 wins this round, R4 retries again

•Urs Hoelzle, Open Network Summit 2012

Traditional Net. Example

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

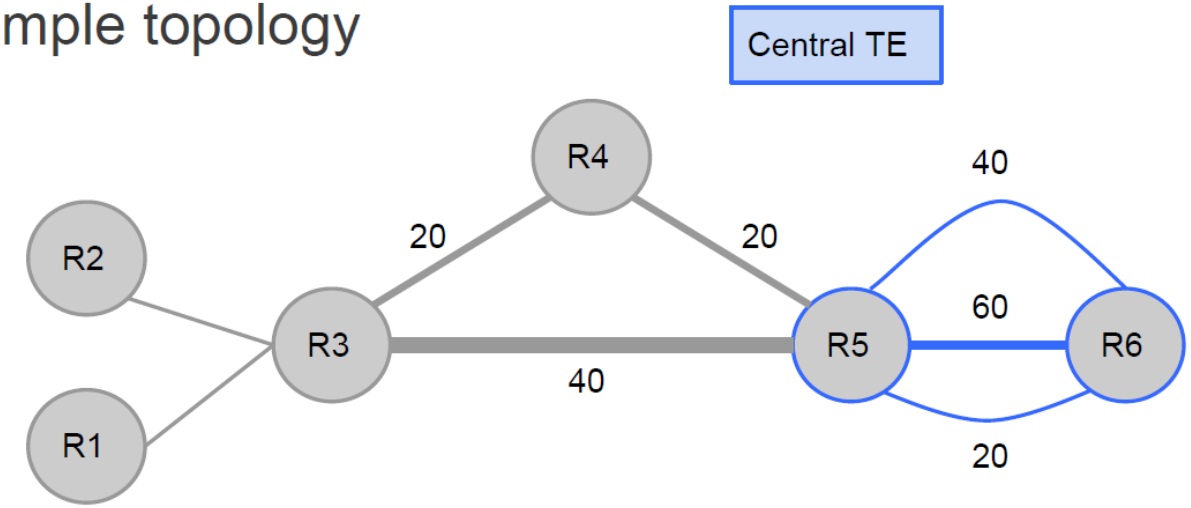


- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1 wins, R2, R4 retry for next best path
  - R2 wins this round, R4 retries again
  - R4 finally gets third best path

•Urs Hoelzle, Open Network Summit 2012

Topology with Central TE

- Simple topology



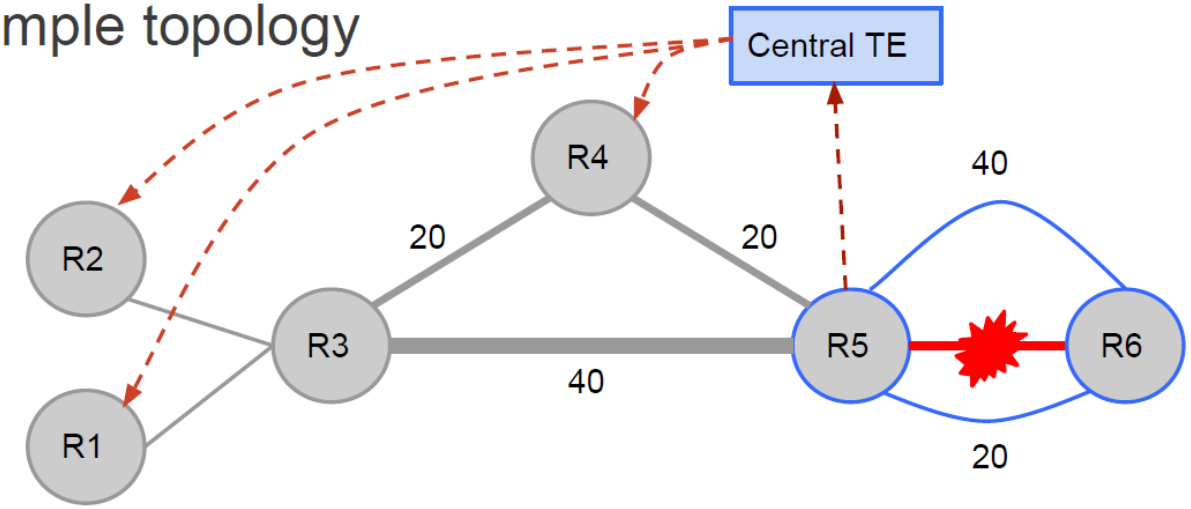
- Flows:
  - R1->R6: 20; R2->R6: 20; R4->R6: 20

TE = Traffic Engineering

•Urs Hoelzle, Open Network Summit 2012

Topology with Central TE

- Simple topology

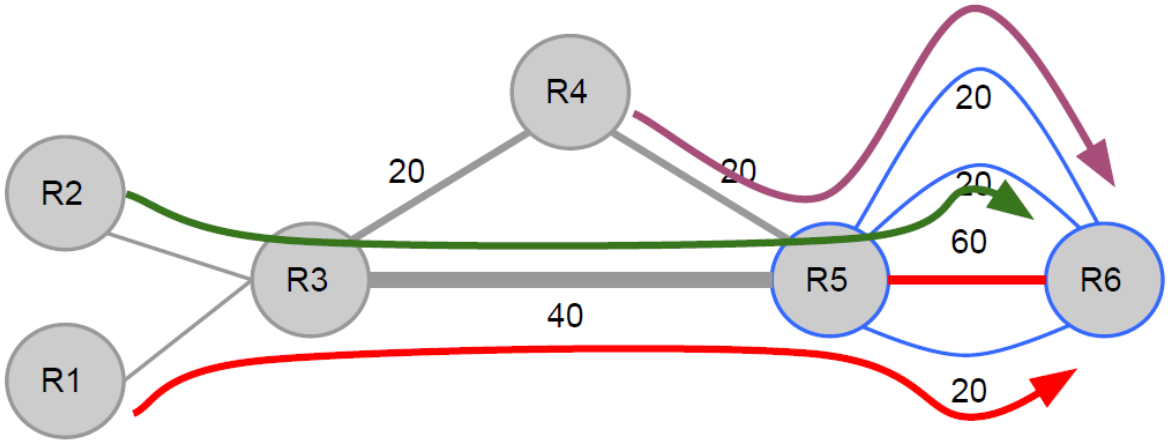


- Flows:
  - R1->R6: 20; R2->R6: 20; R4->R6: 20
- R5-R6 fails
  - R5 informs TE, which programs routers in one shot

•Urs Hoelzle, Open Network Summit 2012

Topology with Central TE

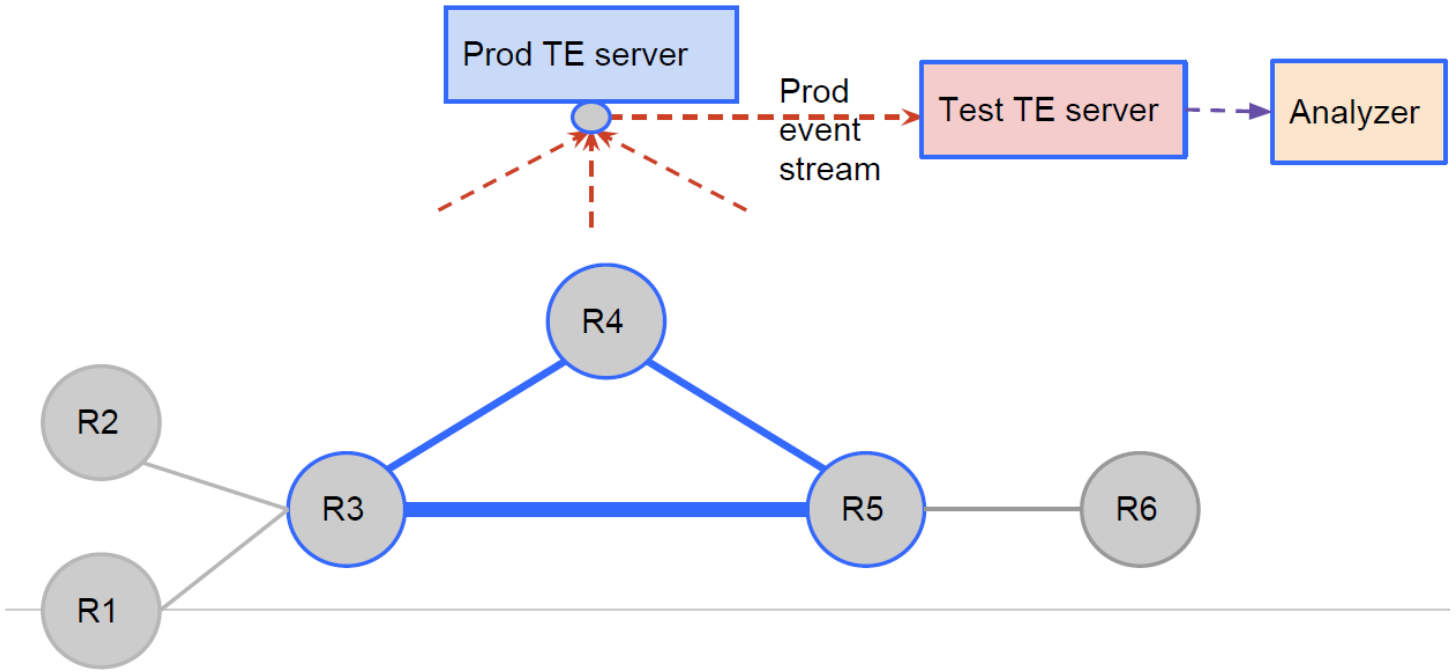
- Simple topology



- 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

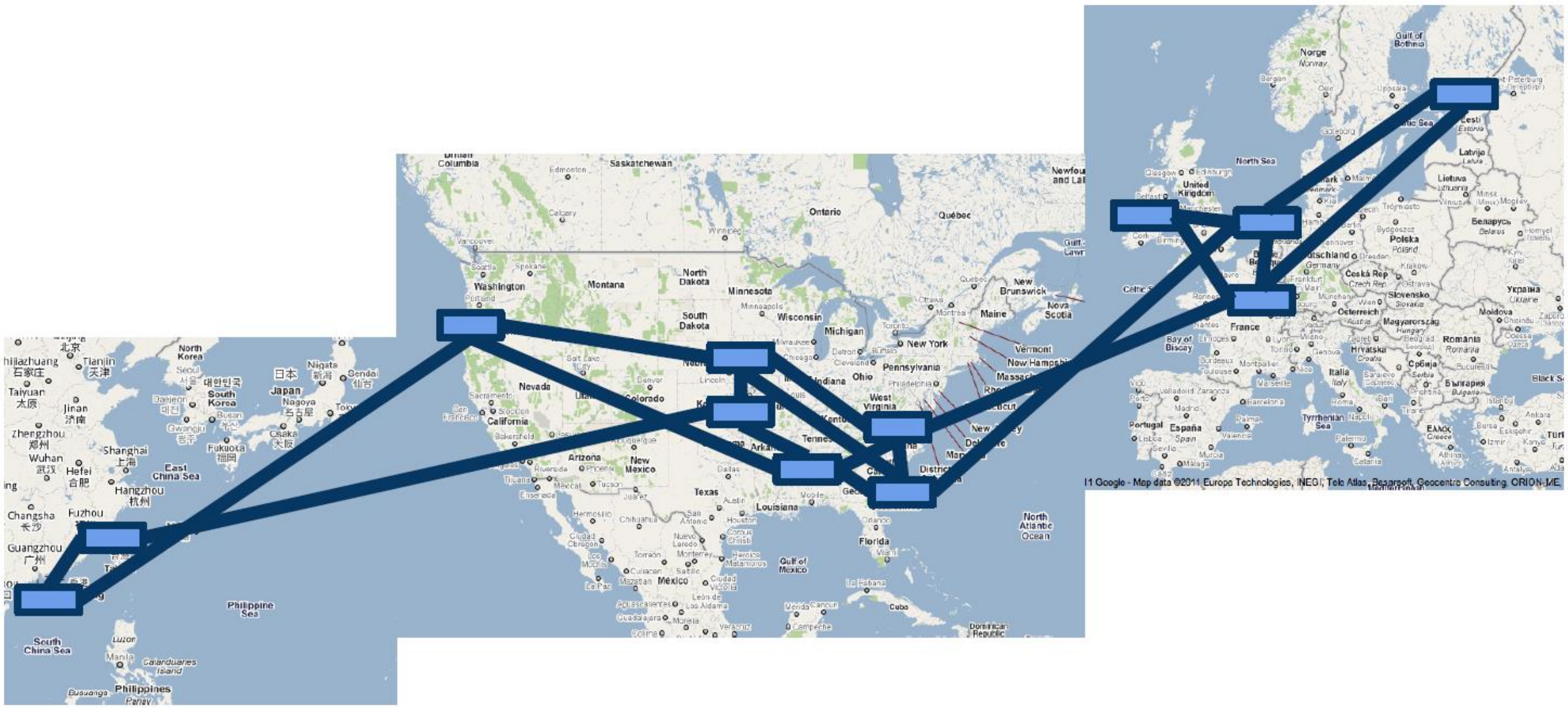
•Urs Hoelzle, Open Network Summit 2012

Traffic Engineering with Analyzer



•Urs Hoelzle, Open Network Summit 2012

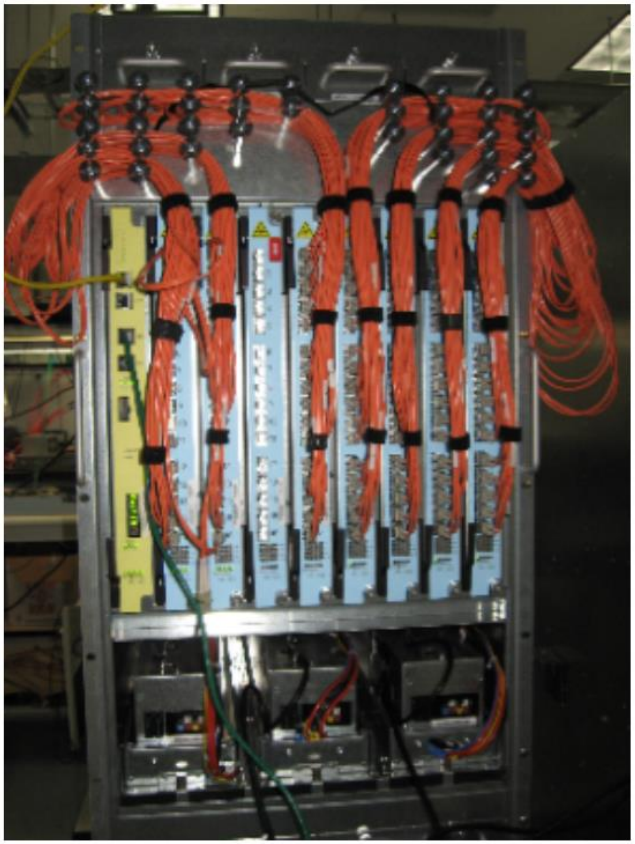
Google's WAN



•Urs Hoelzle, Open Network Summit 2012

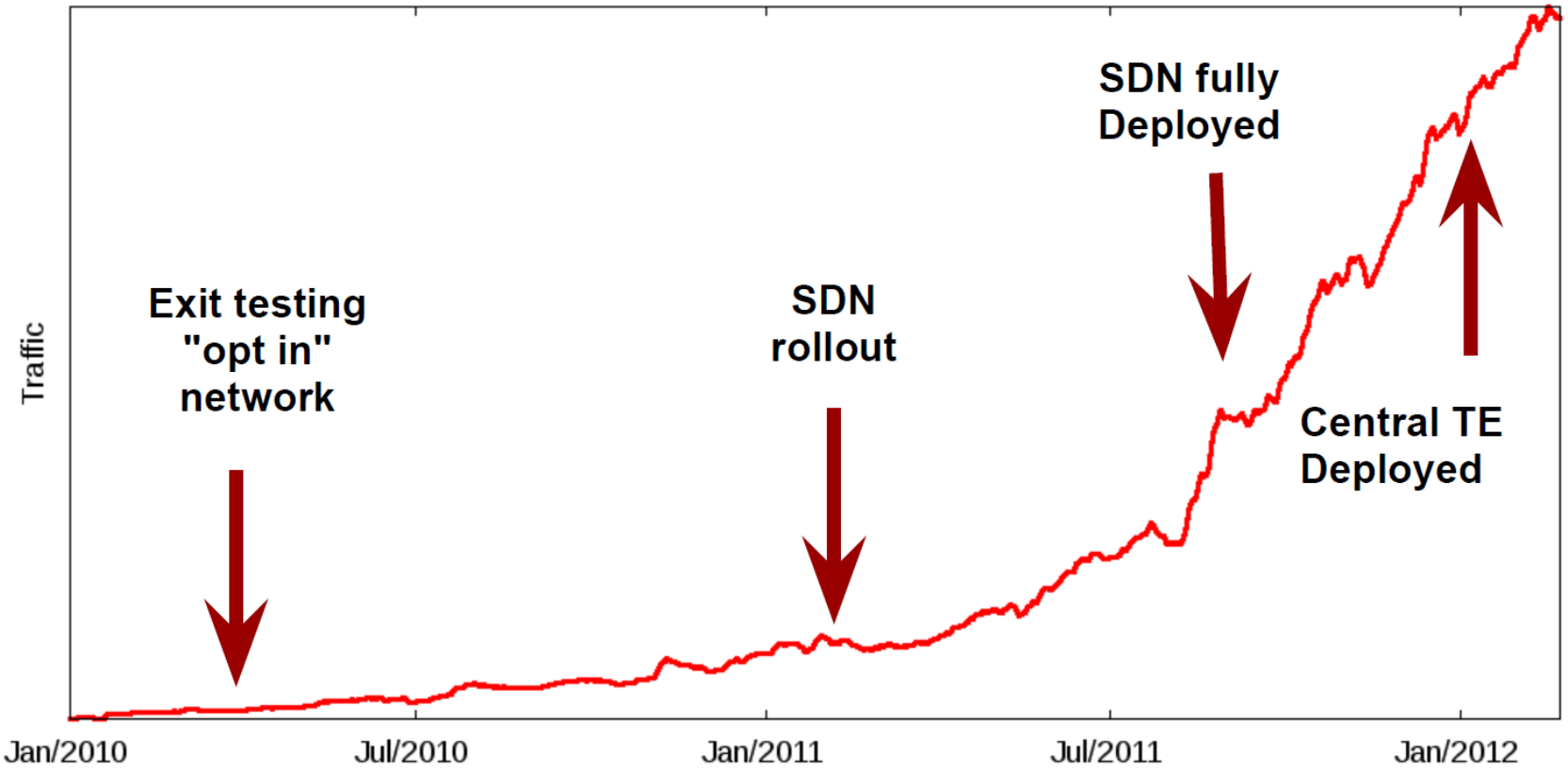
Google's Hardware

- Built from merchant silicon
  - 100s of ports of nonblocking 10GE
- OpenFlow support
- Open source routing stacks for BGP, ISIS
- Does not have all features
  - No support for AppleTalk...
- Multiple chassis per site
  - Fault tolerance
  - Scale to multiple Tbps



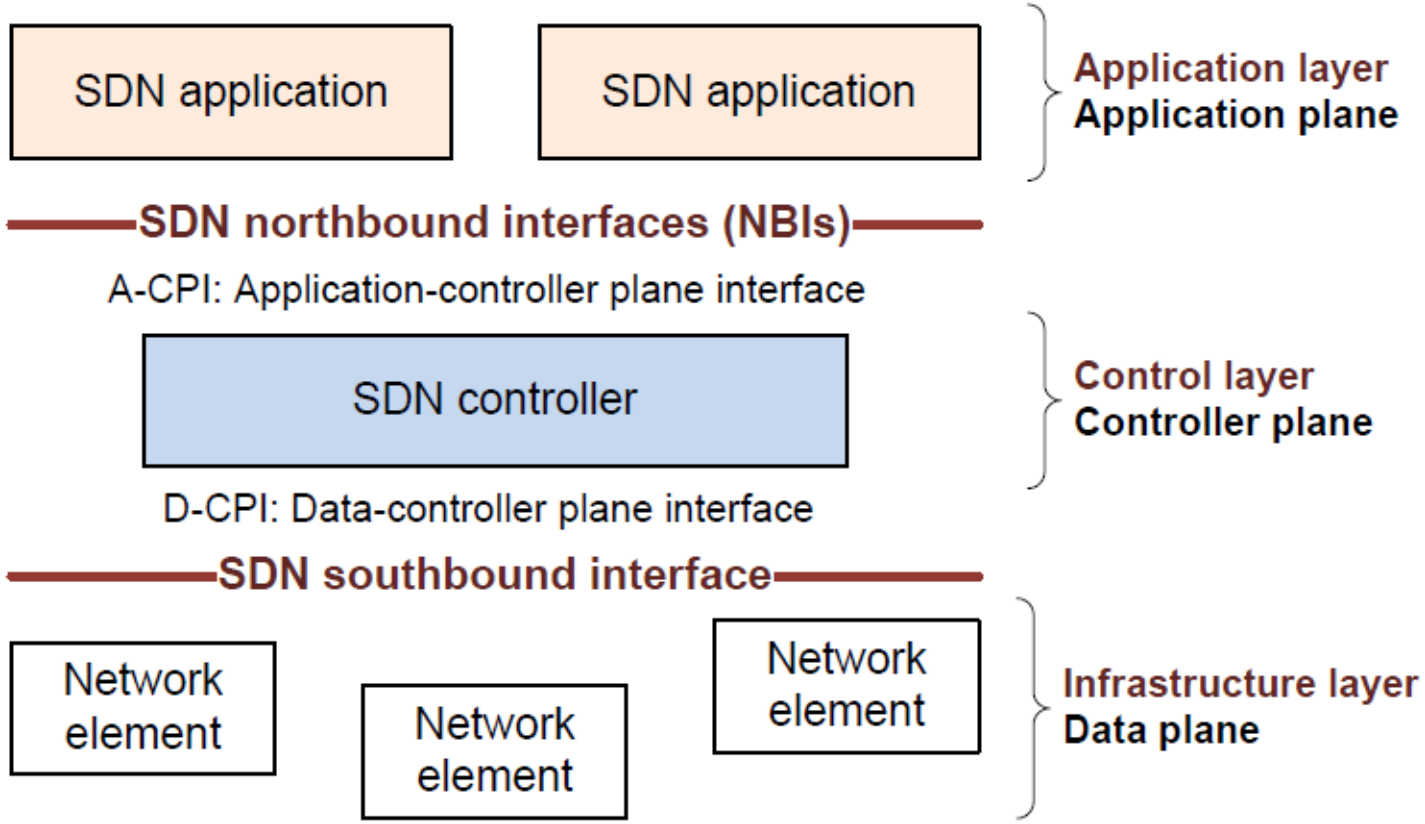
•Urs Hoelzle, Open Network Summit 2012

Timeline of Deployment



•Urs Hoelzle, Open Network Summit 2012

SDN Architecture



SDN Overview

Specifies behavior

Control Program

Abstract Network Model

Compiles to topology

Network Virtualization

Global Network View

Transmits to switches

Network OS

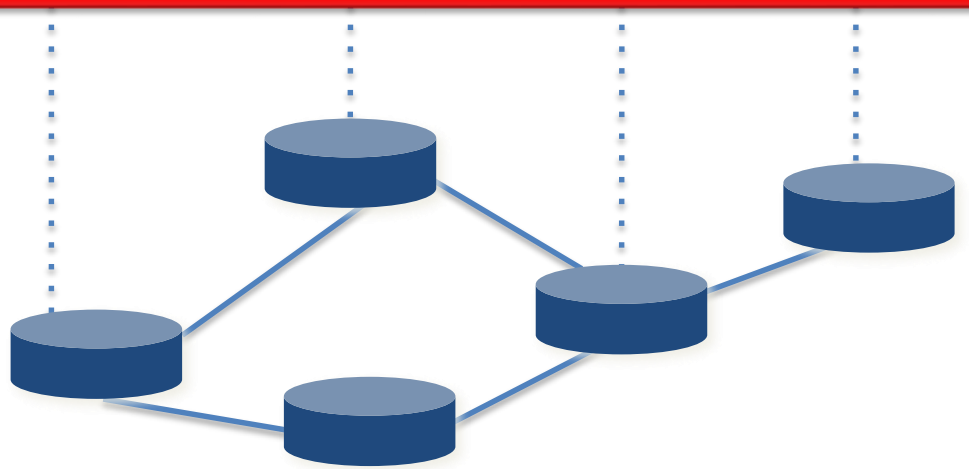
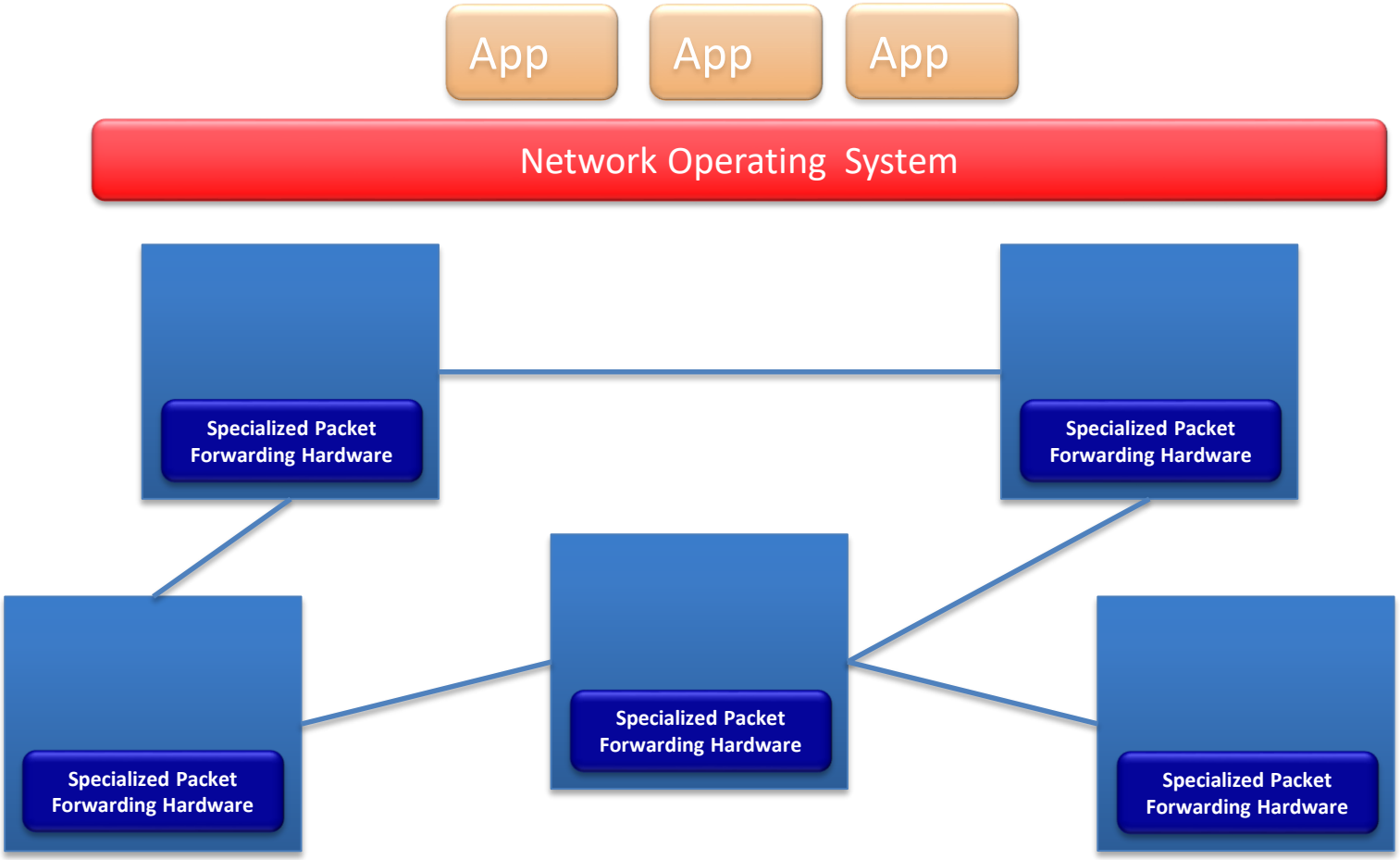
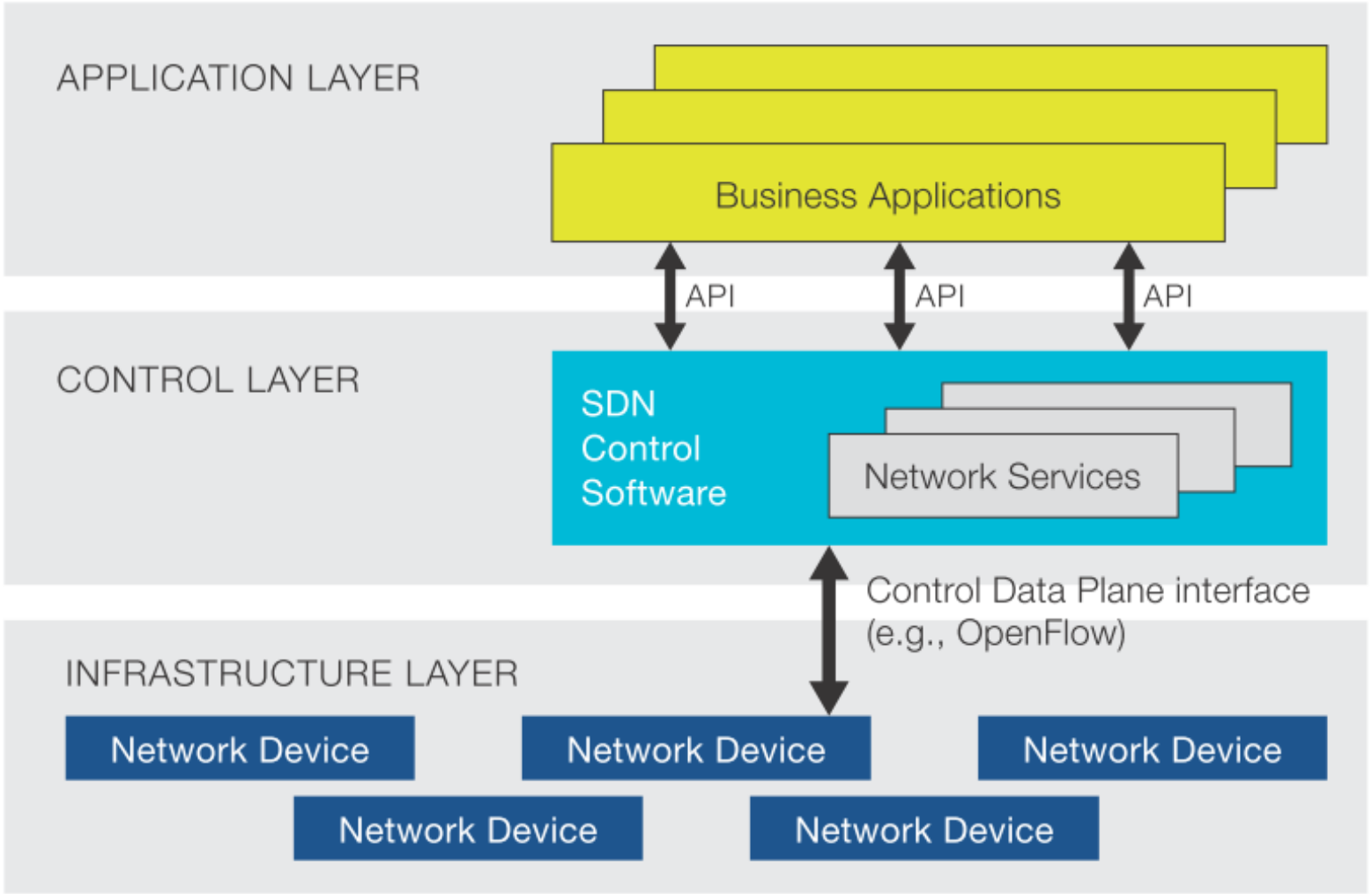


*Figure courtesy of Scott Shenker



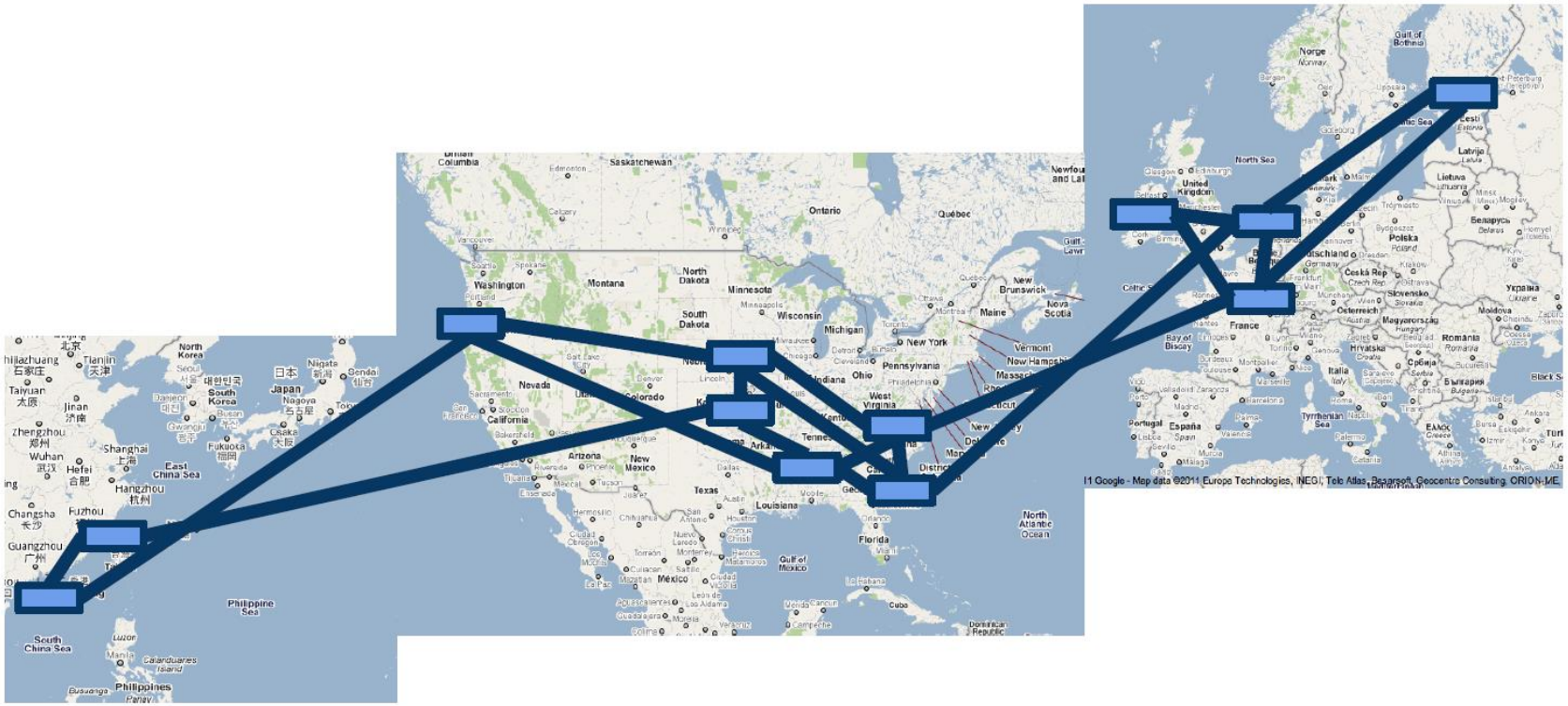
•Yiannis Yiakoumis, et al, HOT-Interconnects 2010

Openflow SDN Architecture



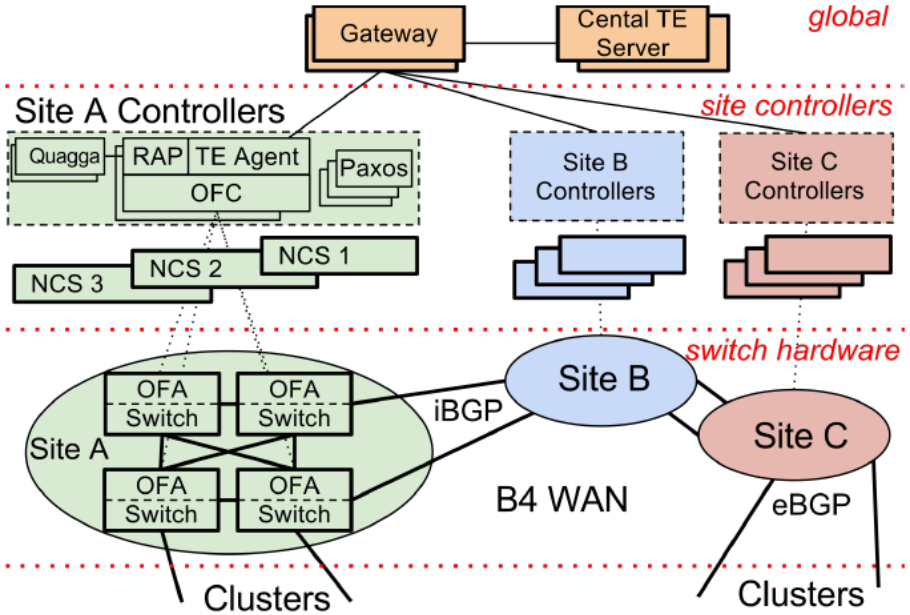
OpenFlow Whitepaper, Open Net. Found.

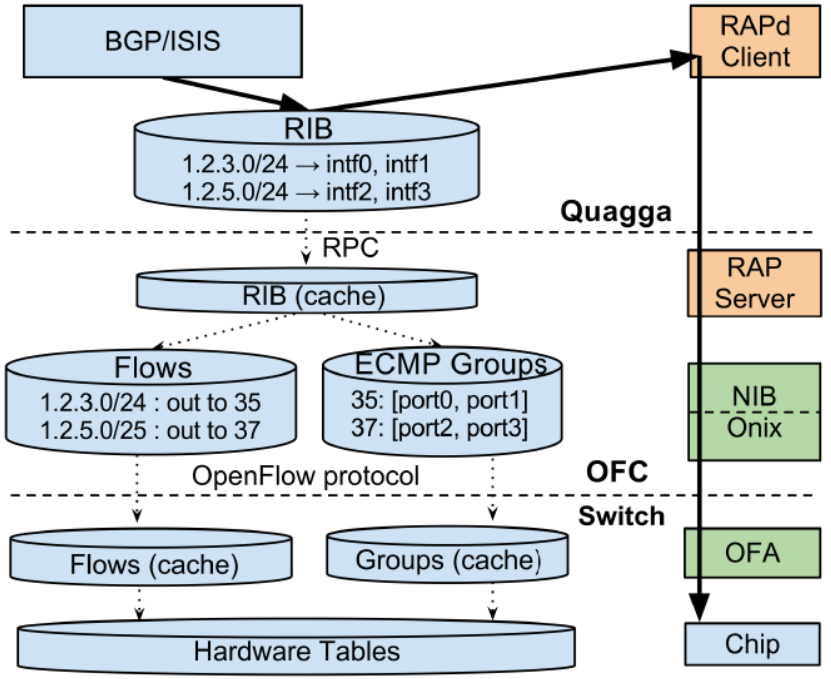
Google's WAN



•Urs Hoelzle, Open Network Summit 2012

Openflow in B4



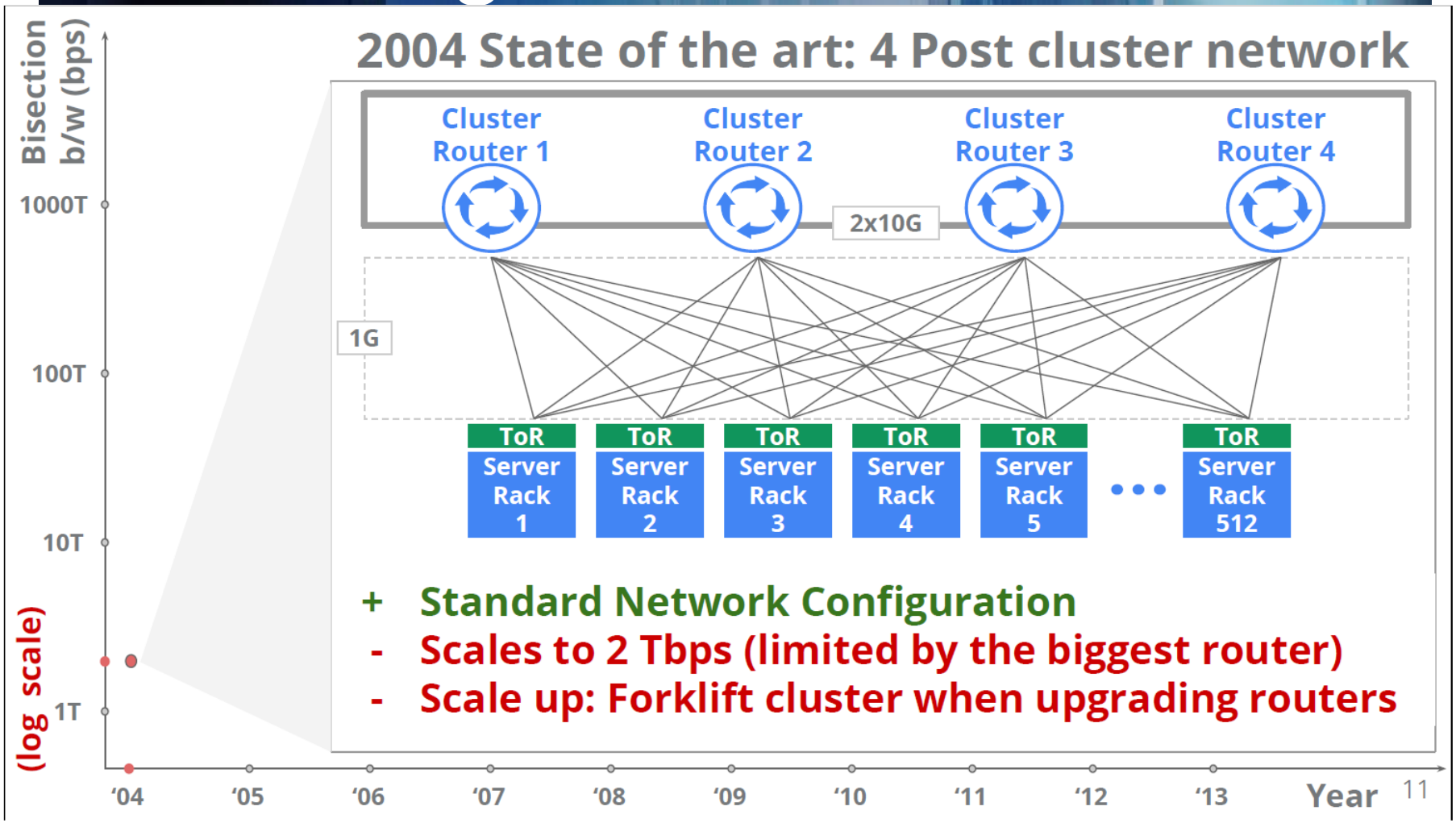


B4: Experience with a Globally Deployed Software Defined WAN, Sushant Jain, et al, 2013

CS2031
Telecommunications II

Clos / Google's Infrastructure

Google's Infrastructure



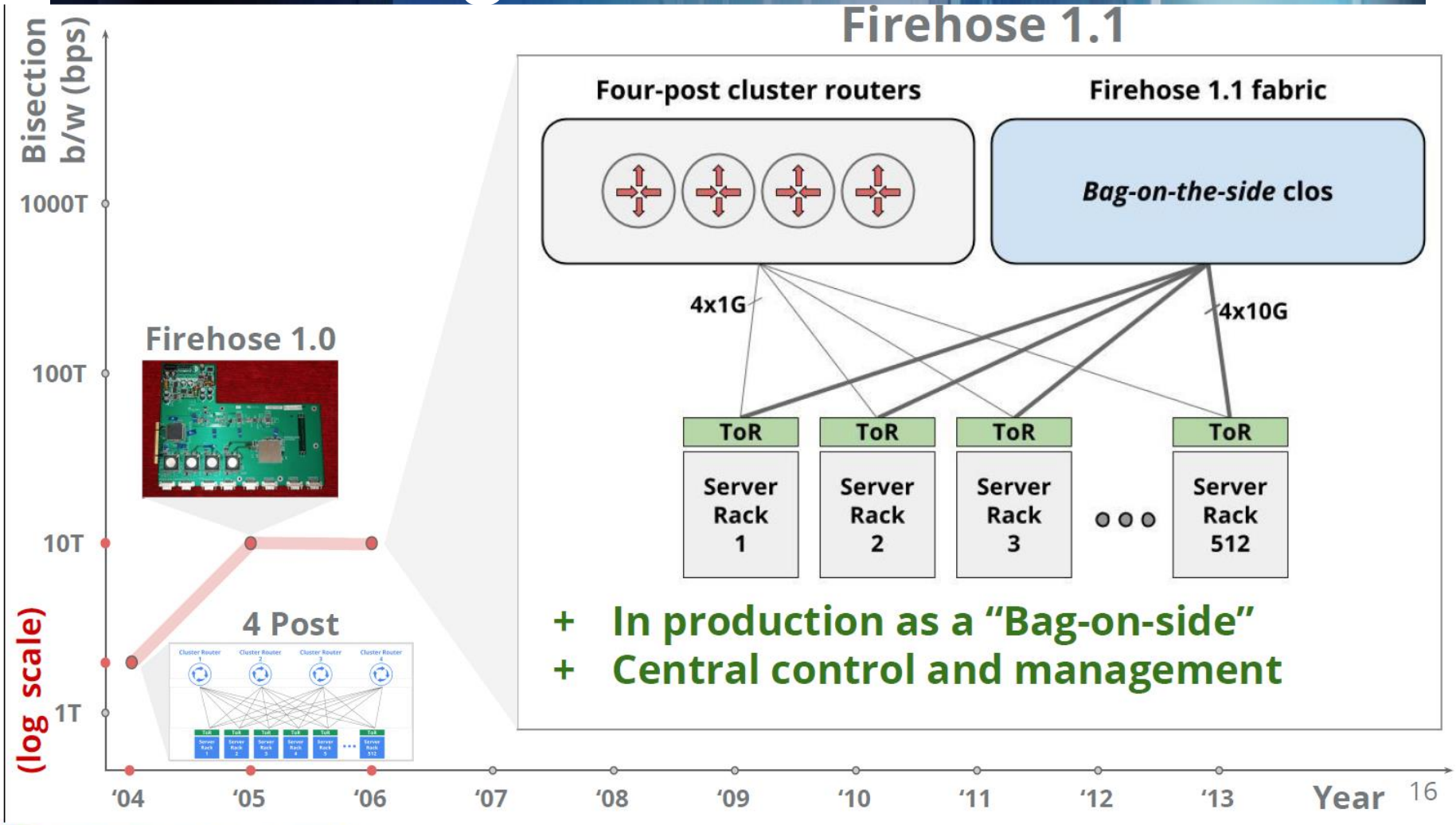
11

Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

The Obligatory Datacentre Pic

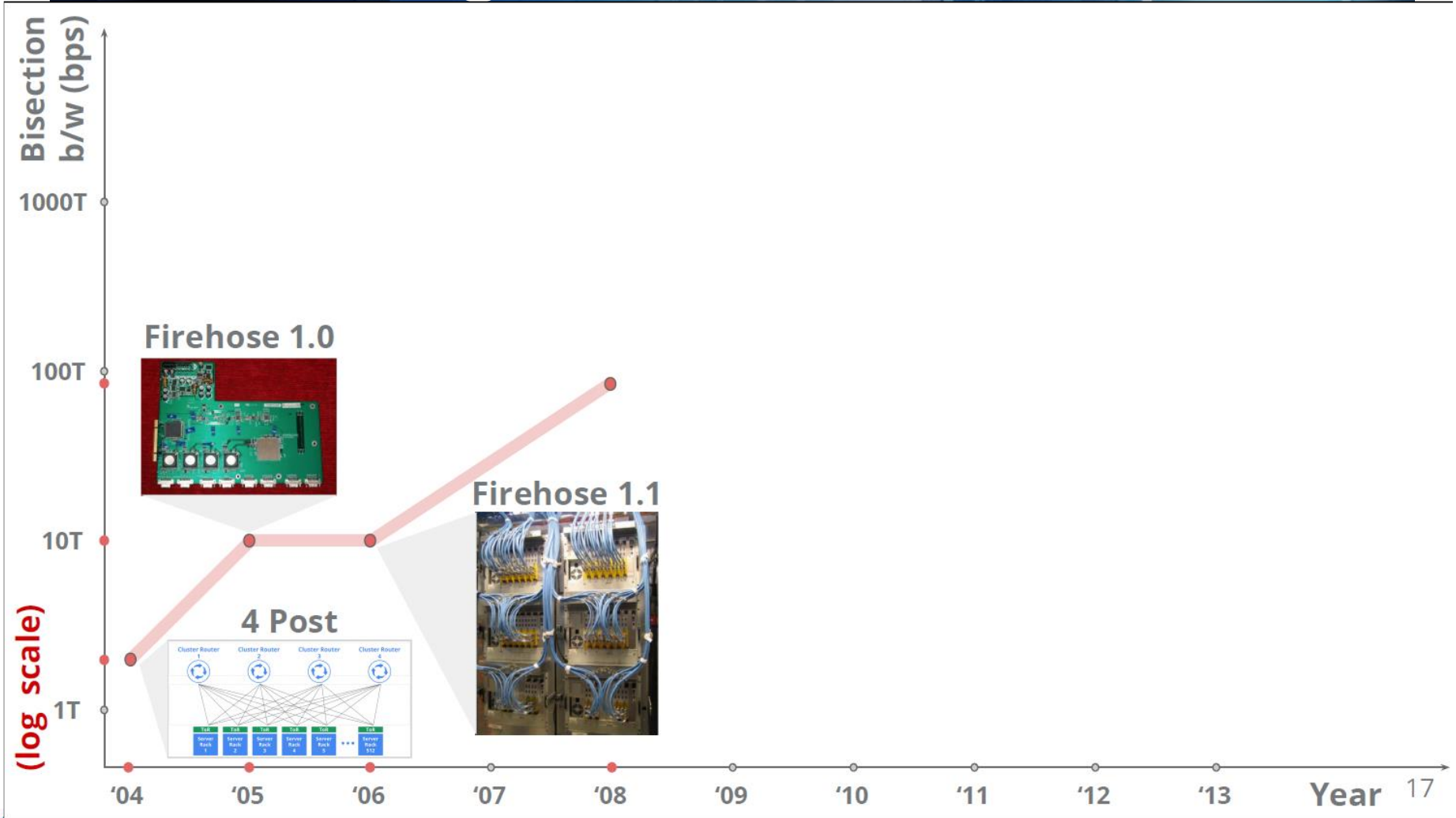
http://swanand18.blogspot.ie/2013/10/inside-facebooks-data-center-near.html

Google's Infrastructure



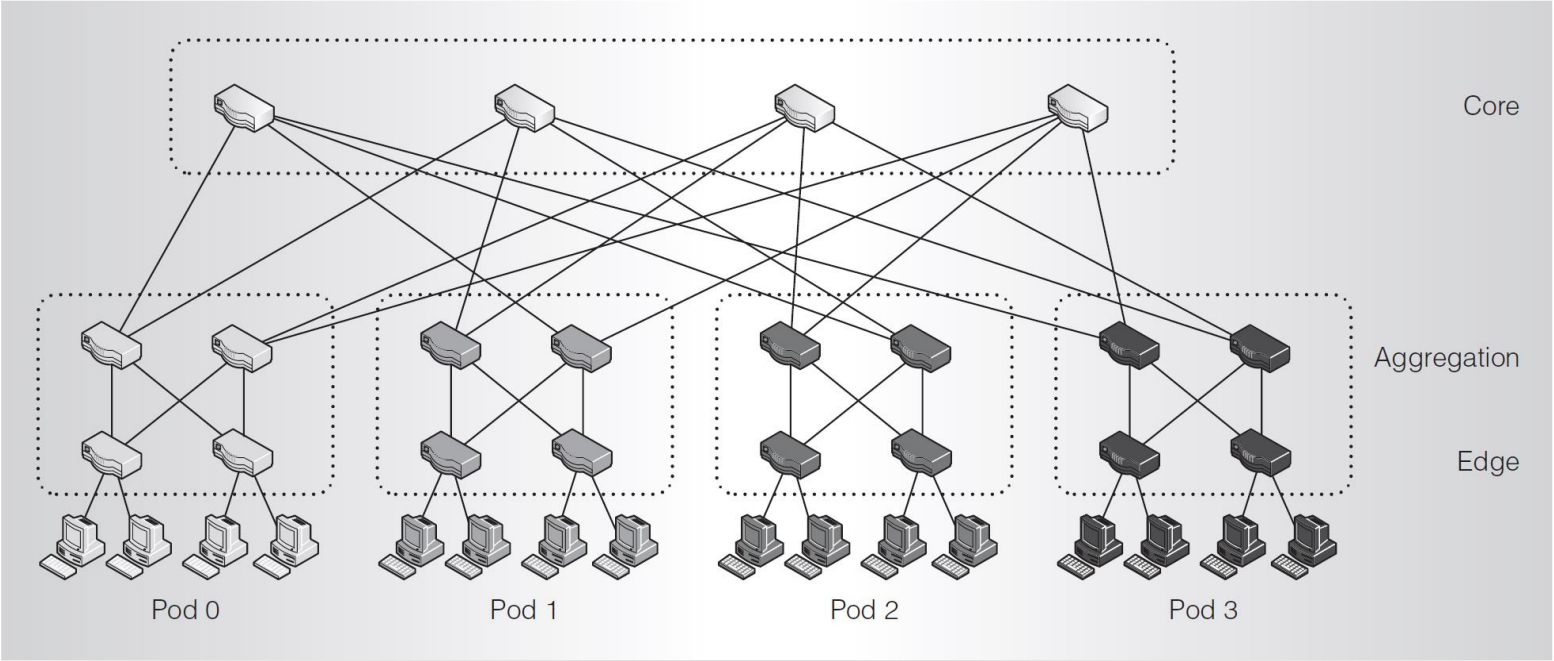
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

Google's Infrastructure



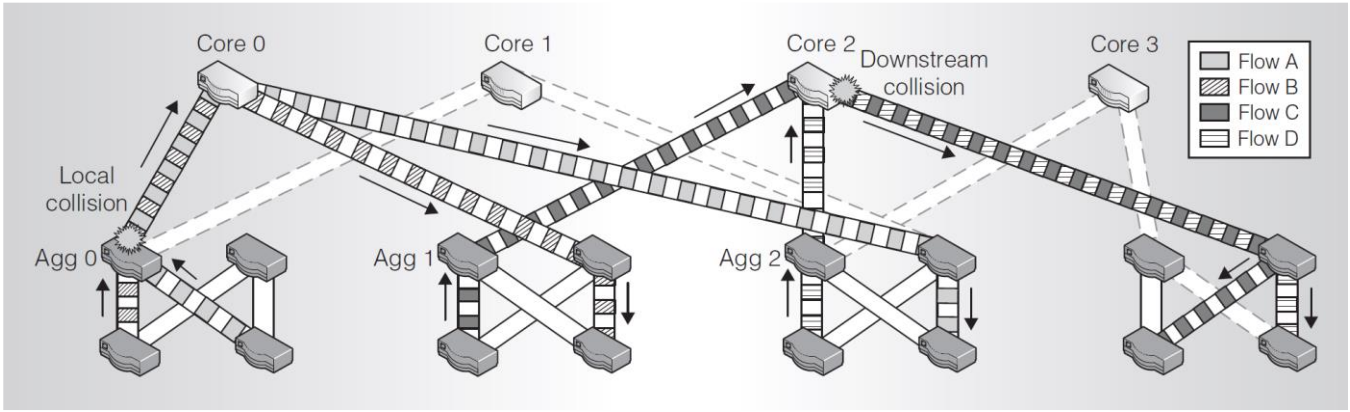
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

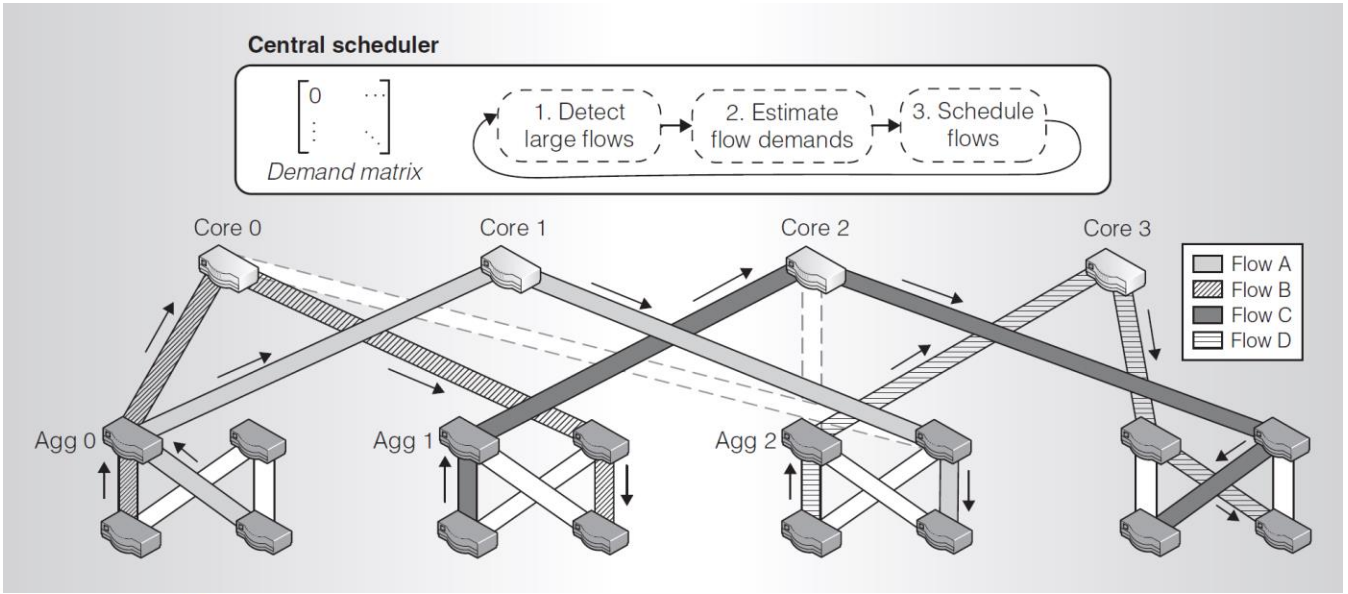
Fat-tree Structure



Scale-Out Networking in the Data Center, Amin Vahdat, et al , 2010

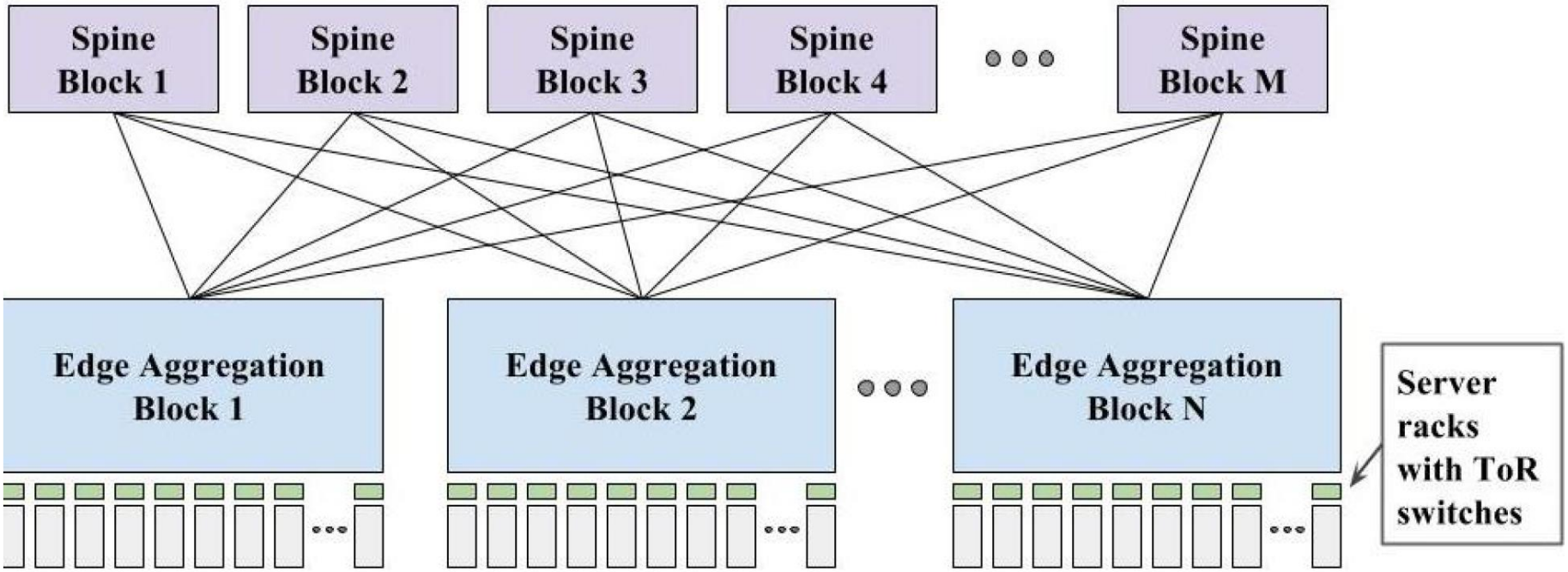
Equal-cost multipath (ECMP) forwarding collisions





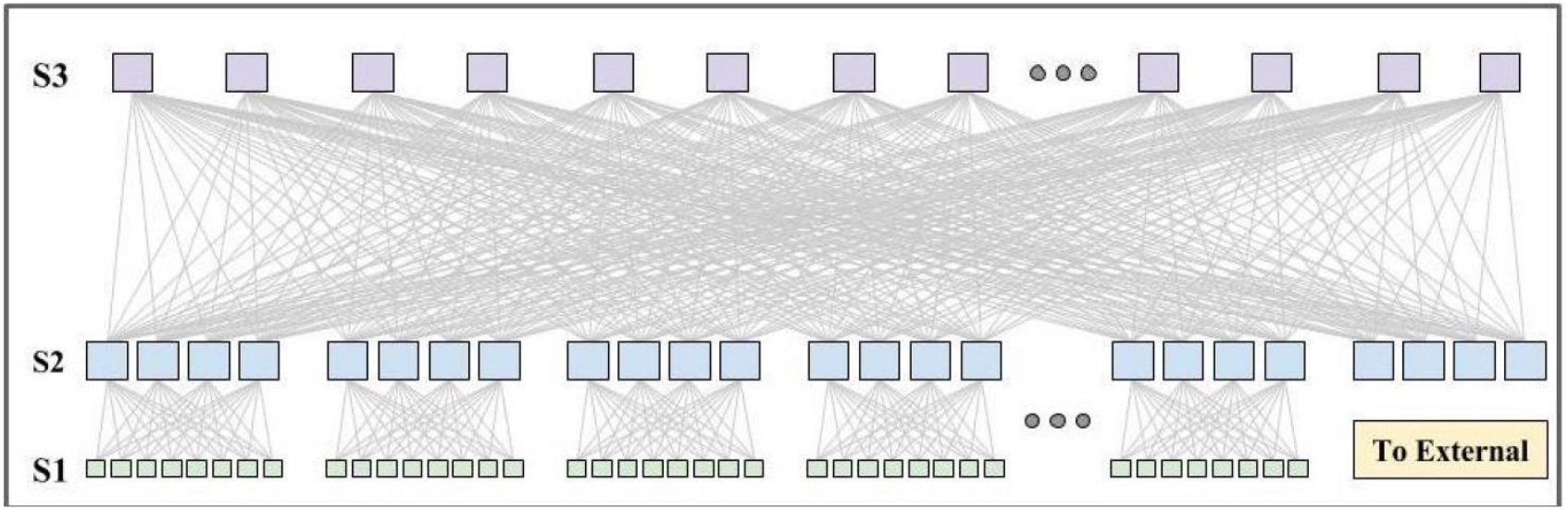
Scale-Out Networking in the Data Center, Amin Vahdat, et al , 2010

CLOS Architectures



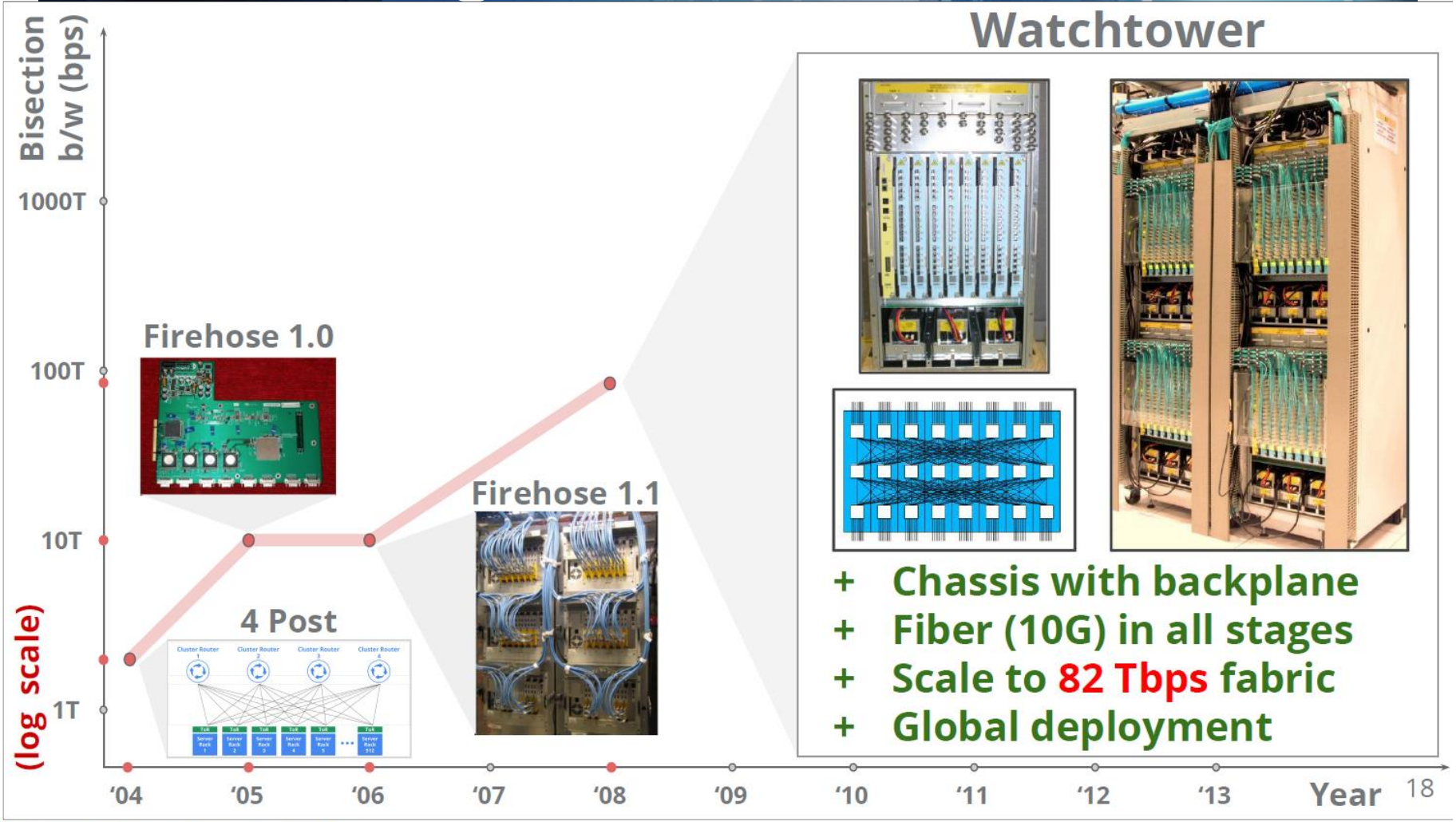
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

CLOS Architectures



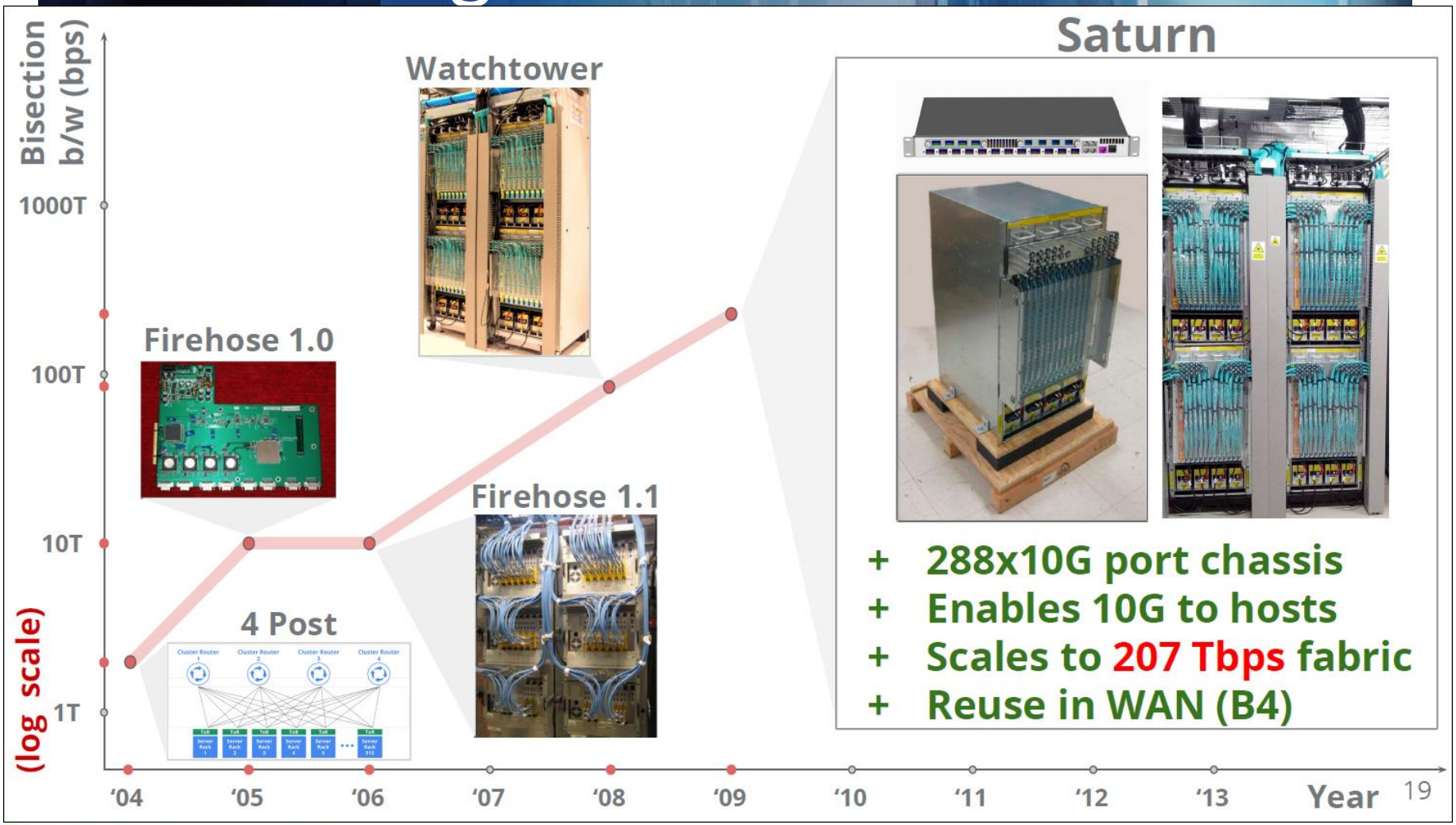
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

Google's Infrastructure



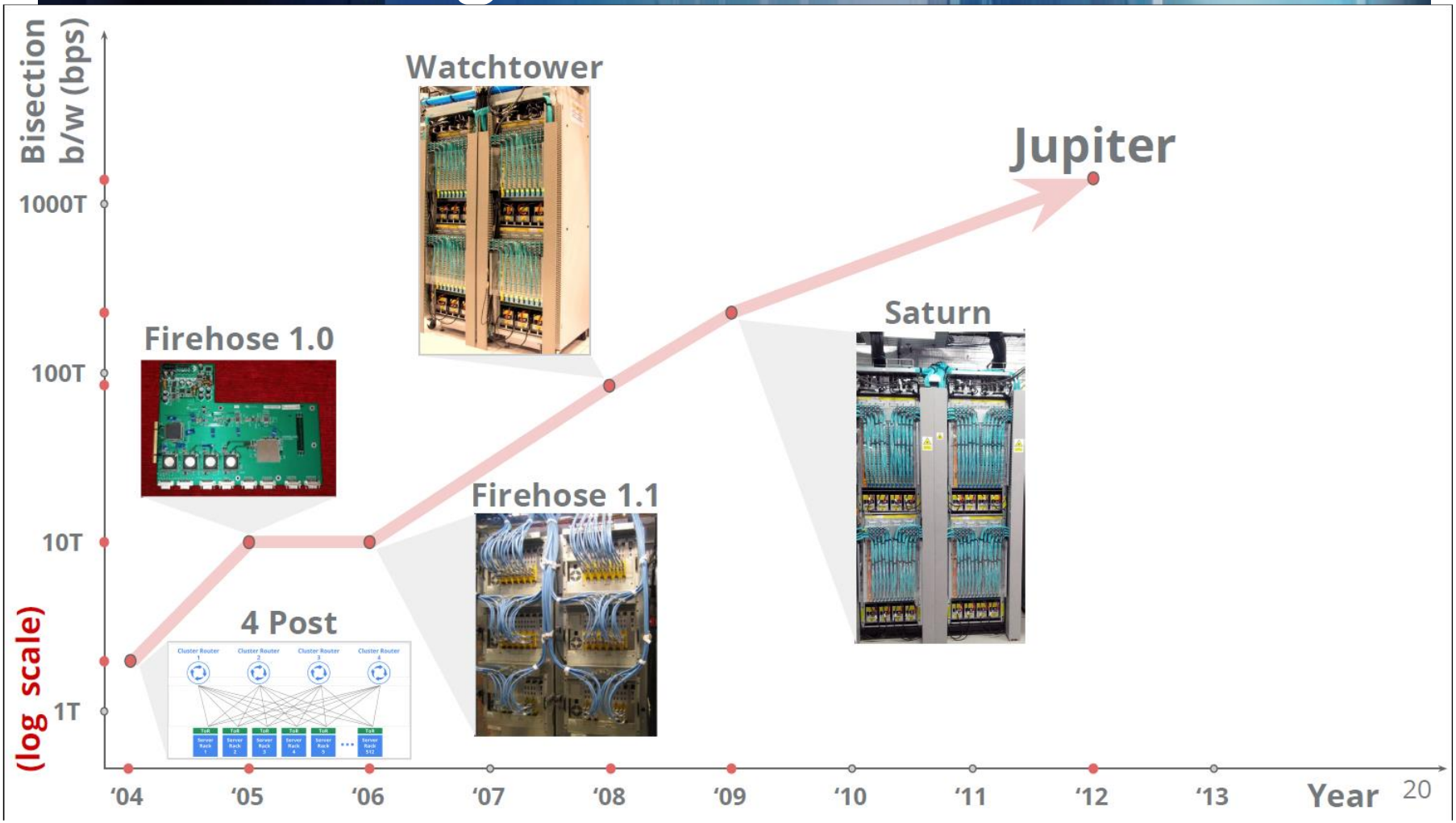
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

Google's Infrastructure



Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

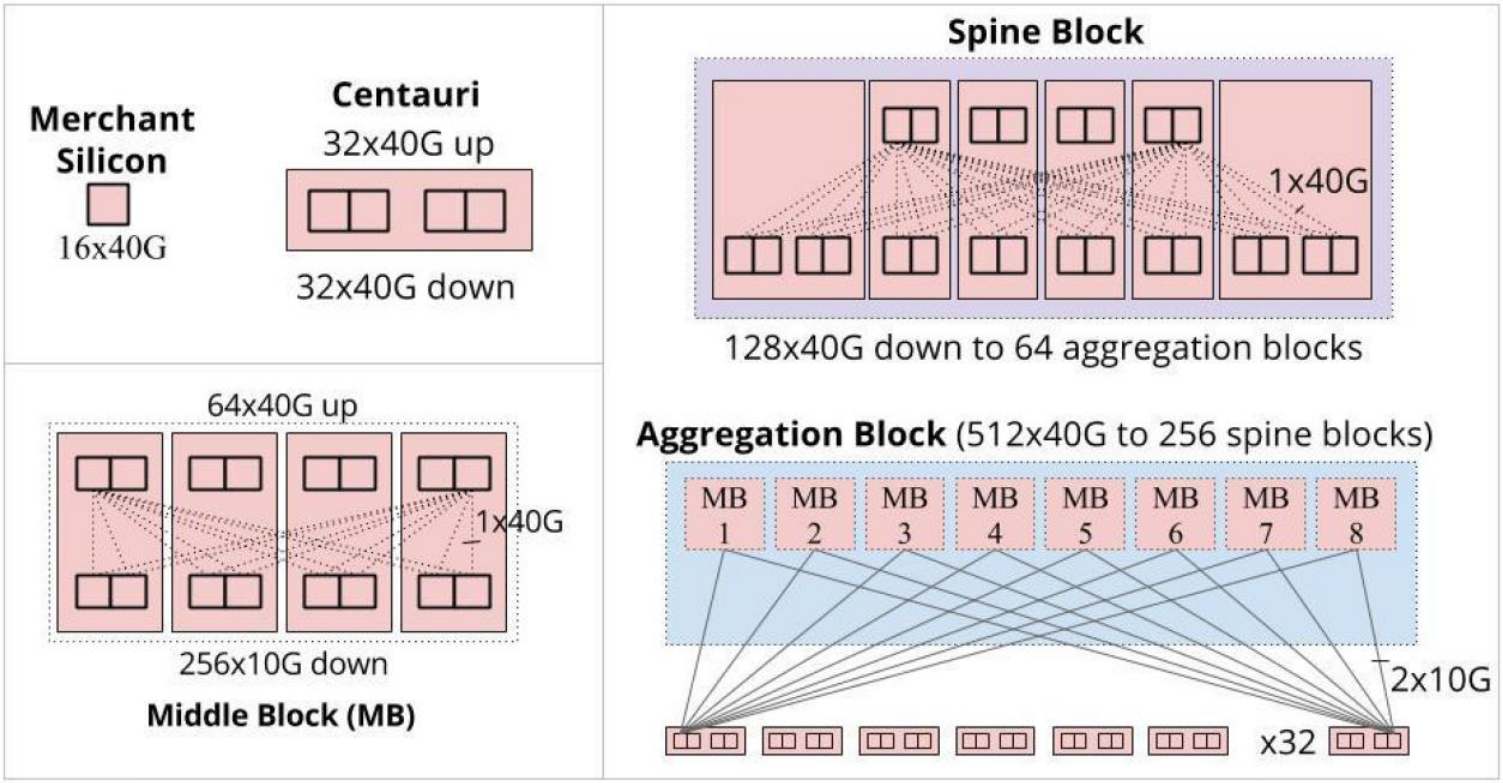
Google's Infrastructure



Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

Google's Infrastructure

Jupiter topology

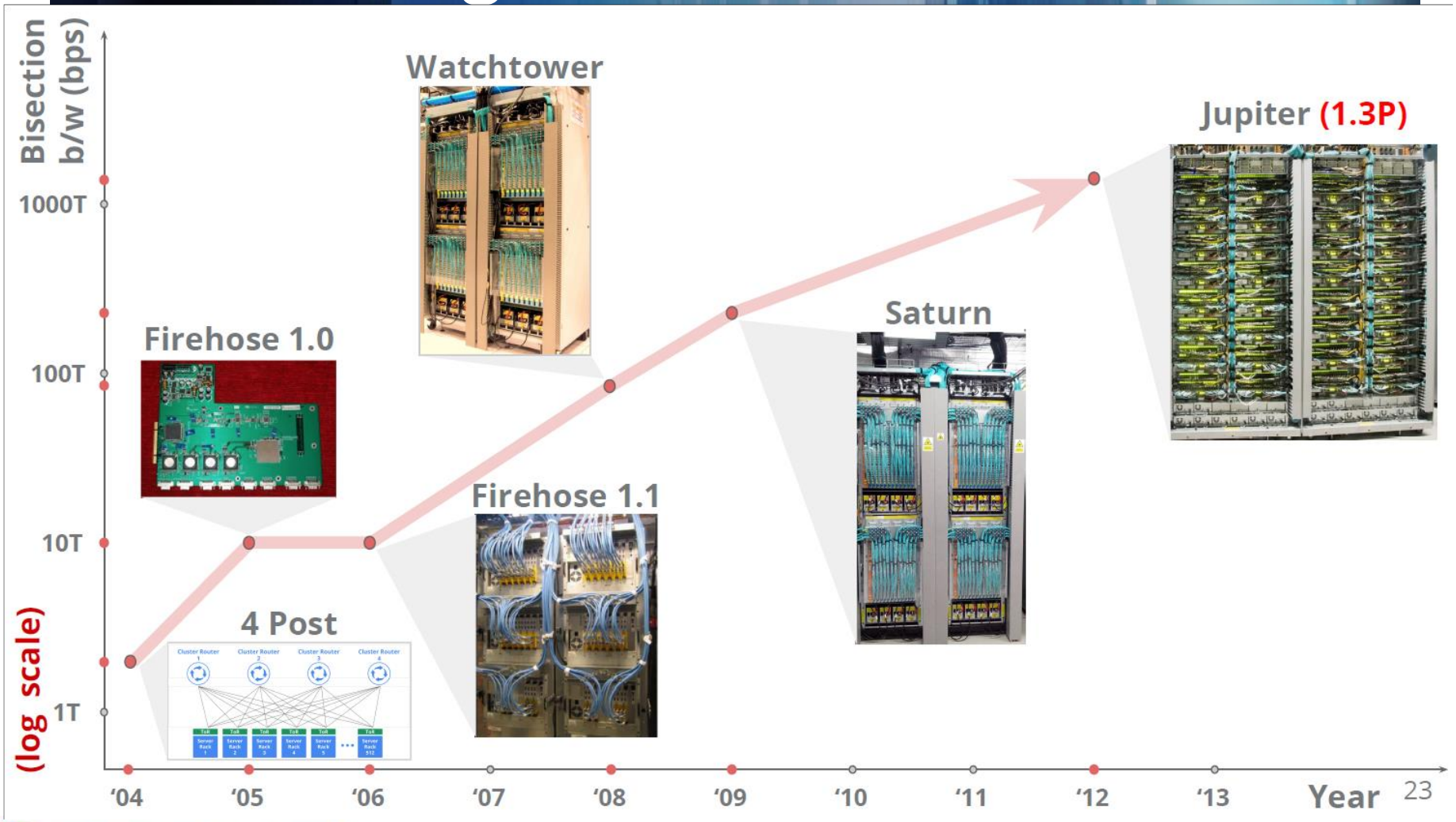


+ Scales out building wide 1.3 Pbps

Google

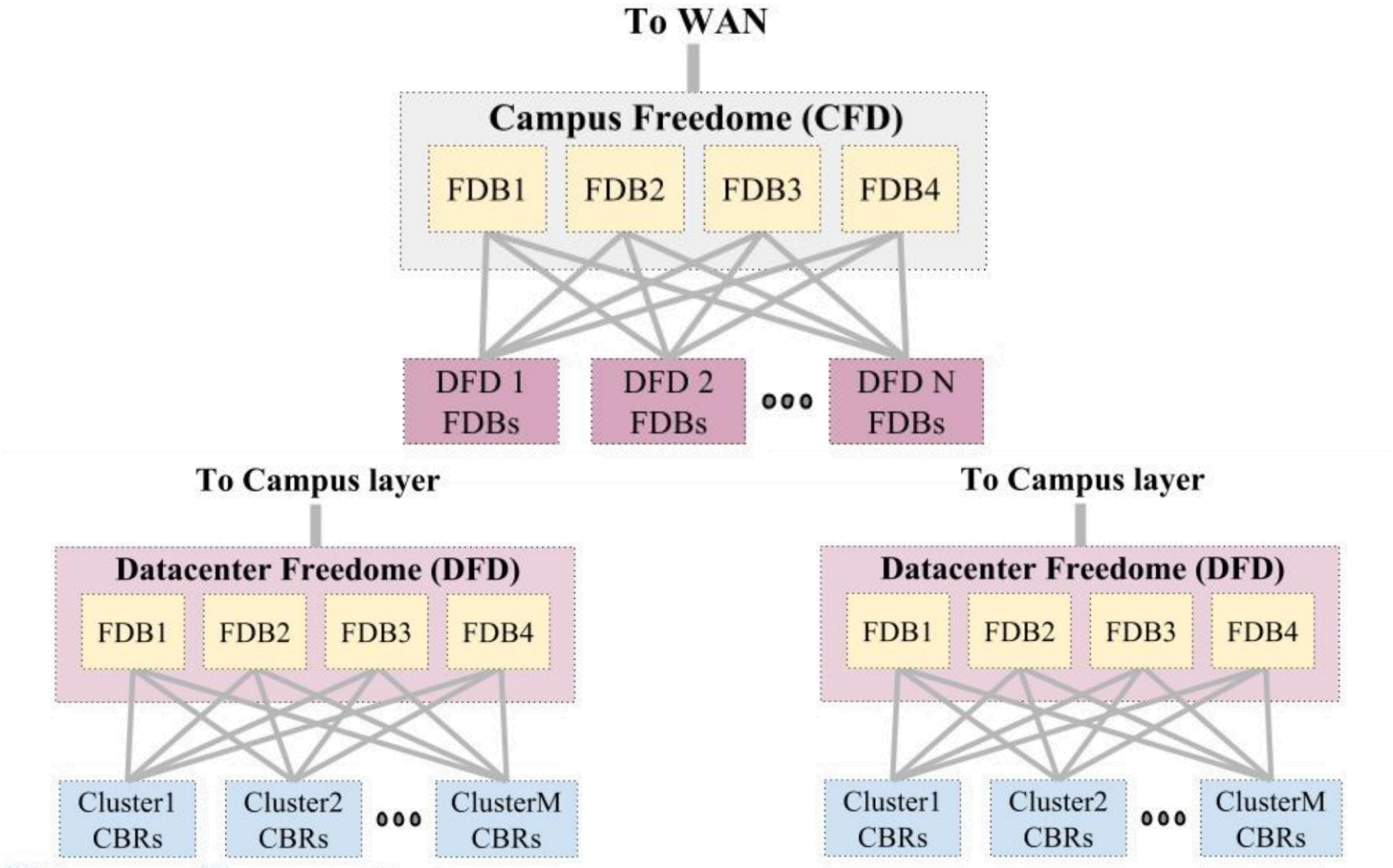
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

Google's Infrastructure



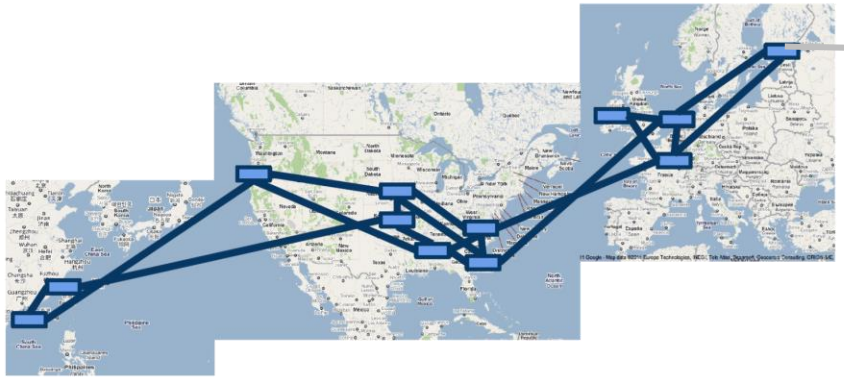
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

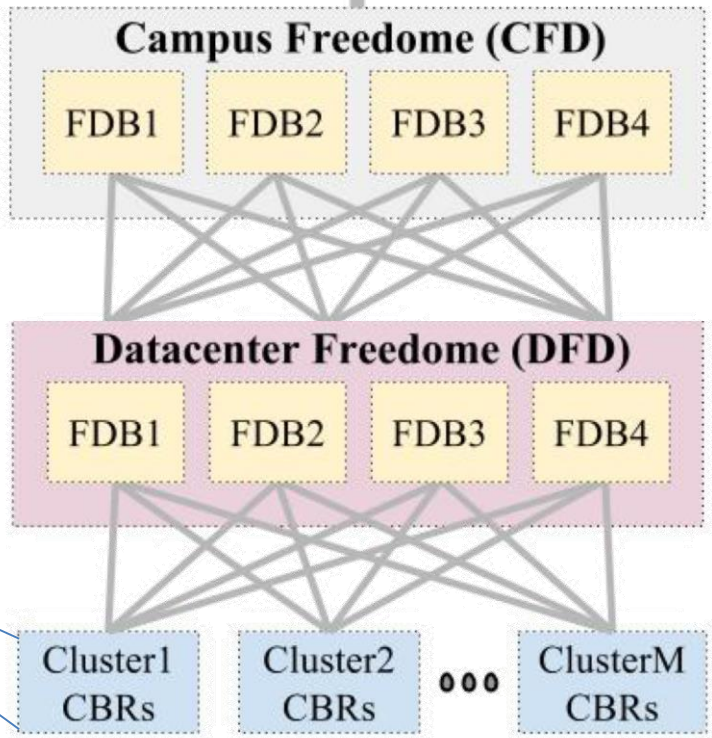
Google's Infrastructure

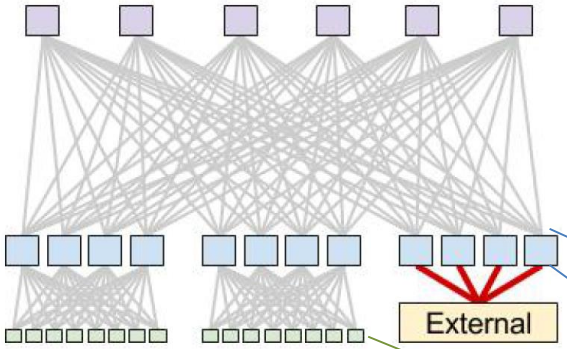


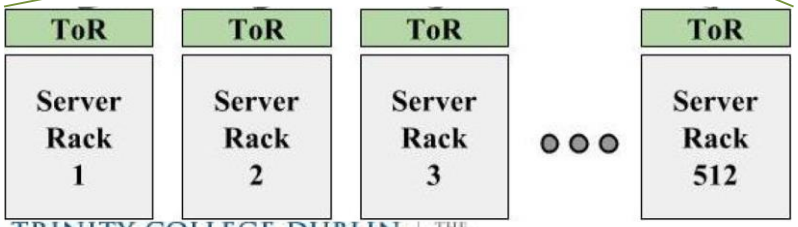
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network, Arjun Singh, et al, 2015

Google's B4 to Jupiter







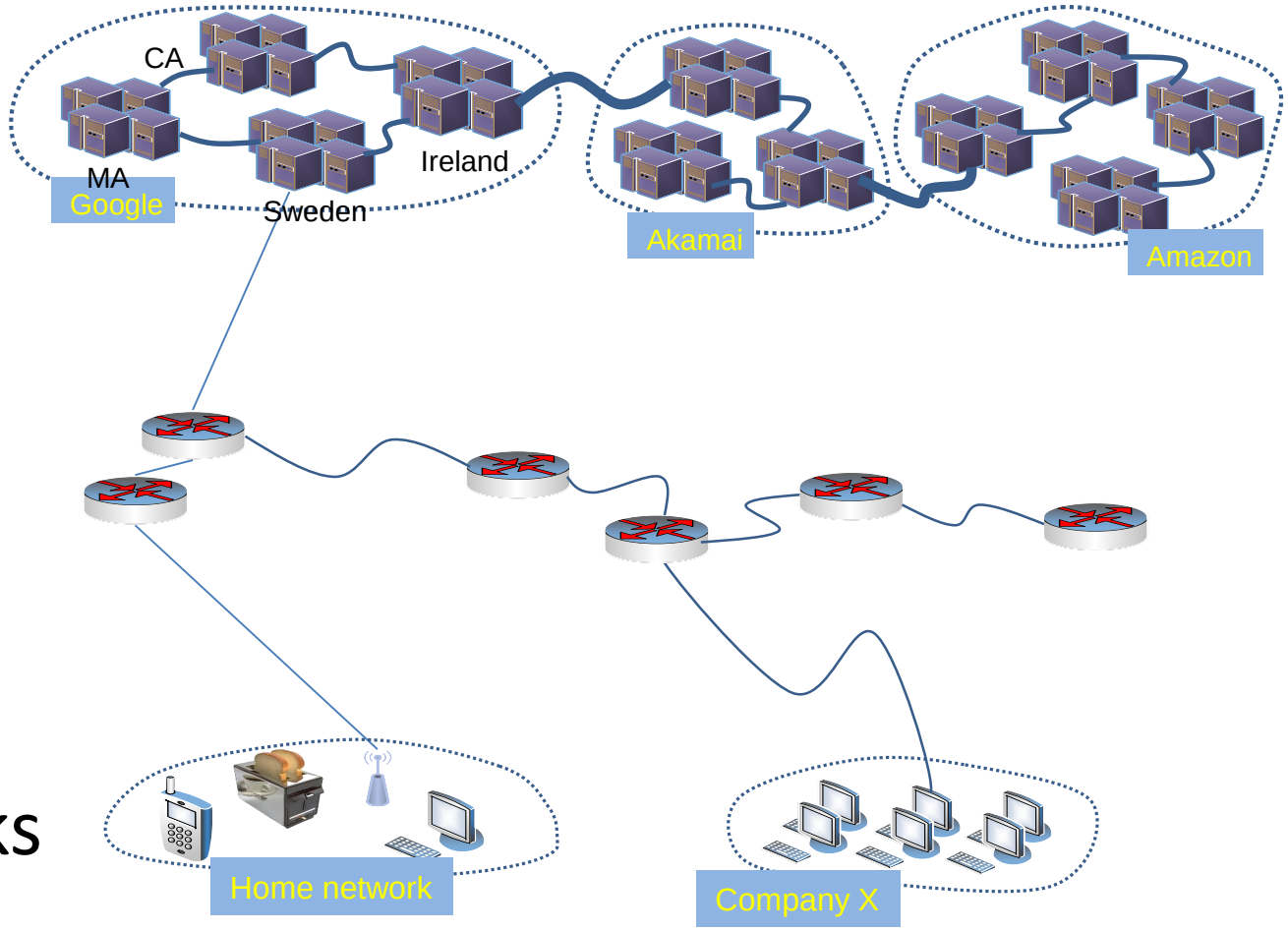


= 40 servers

FDB = Freedome Block
CBR = Cluster Border Router
ToR = Top-of-Rack Switch
B4 = WAN controlled by Openflow

My view of "future"* networking

- Sets of Datacentres
- Traditional Internet
- Edge networks



*or current?

9.12. & 16.12.

No Lectures & Tutorials

Akamai & IPv6

- Akamai will offer IPv6 services to its entire customer base in April - a long-awaited move that will be a major boon to the adoption rate of the next-generation internet protocol.

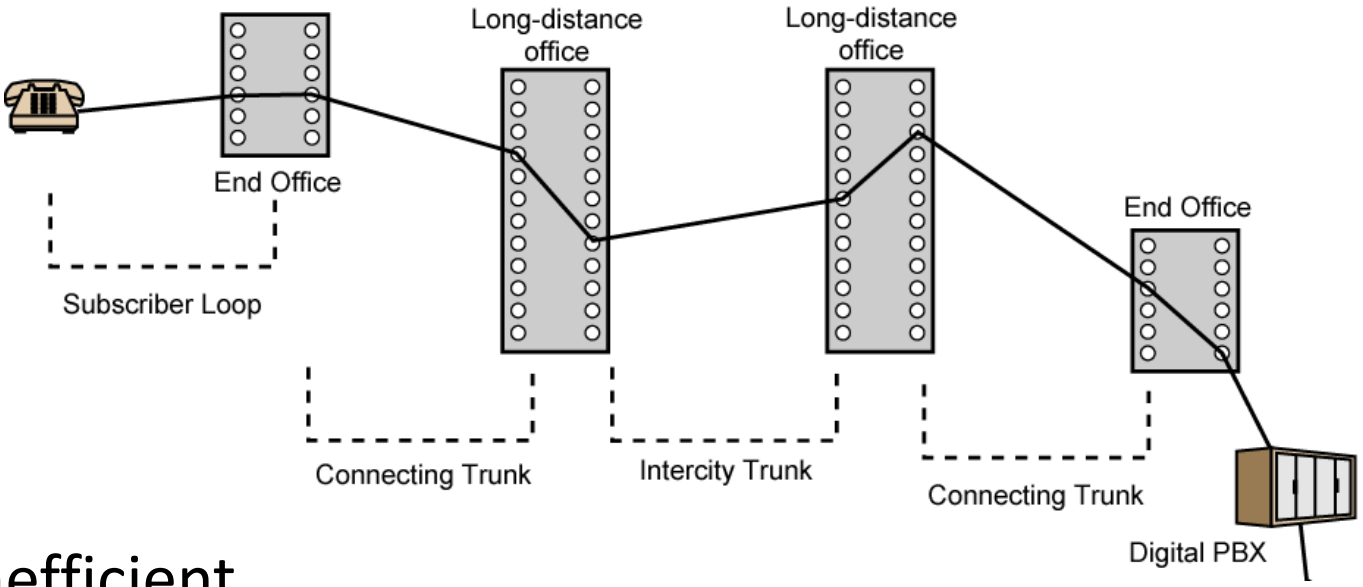
  Carrying between 20% and 30% of the internet's web traffic on any given day, Akamai is the world's largest content delivery network (CDN). Akamai's engineering team has been working for two years to upgrade its 95,000 servers in 71 countries connected by 1,900 networks to support IPv6.

- Mar 27, 2012
  http://www.techworld.com/news/apps/ipv6-akamai-launch-new-internet-protocol-service-in-april-3347188/

CS2031
Telecommunications II

Circuit Switching & Packet Switching

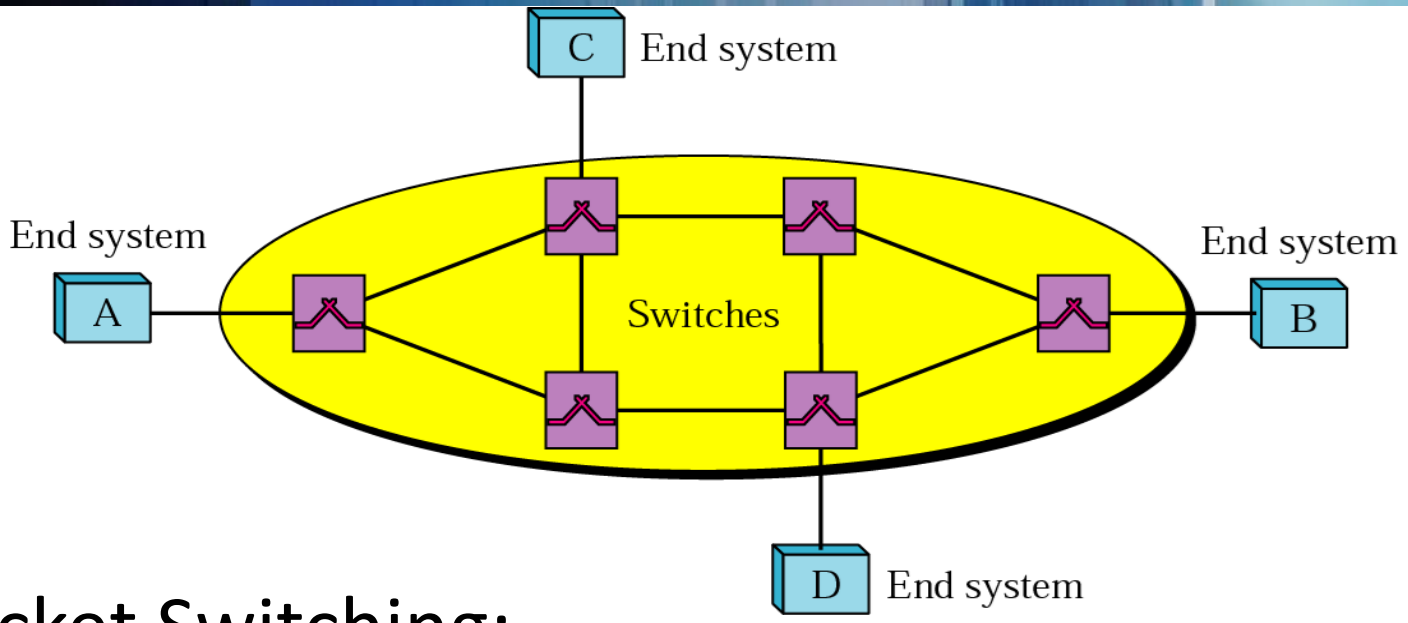
Public Circuit Switched Network



- Inefficient
  - Channel capacity dedicated for duration of connection
  - If no data, capacity wasted
- Set up of connection takes time
- Once connected, transfer is transparent

* Figure is courtesy of A. Tanenbaum

Switched Networks

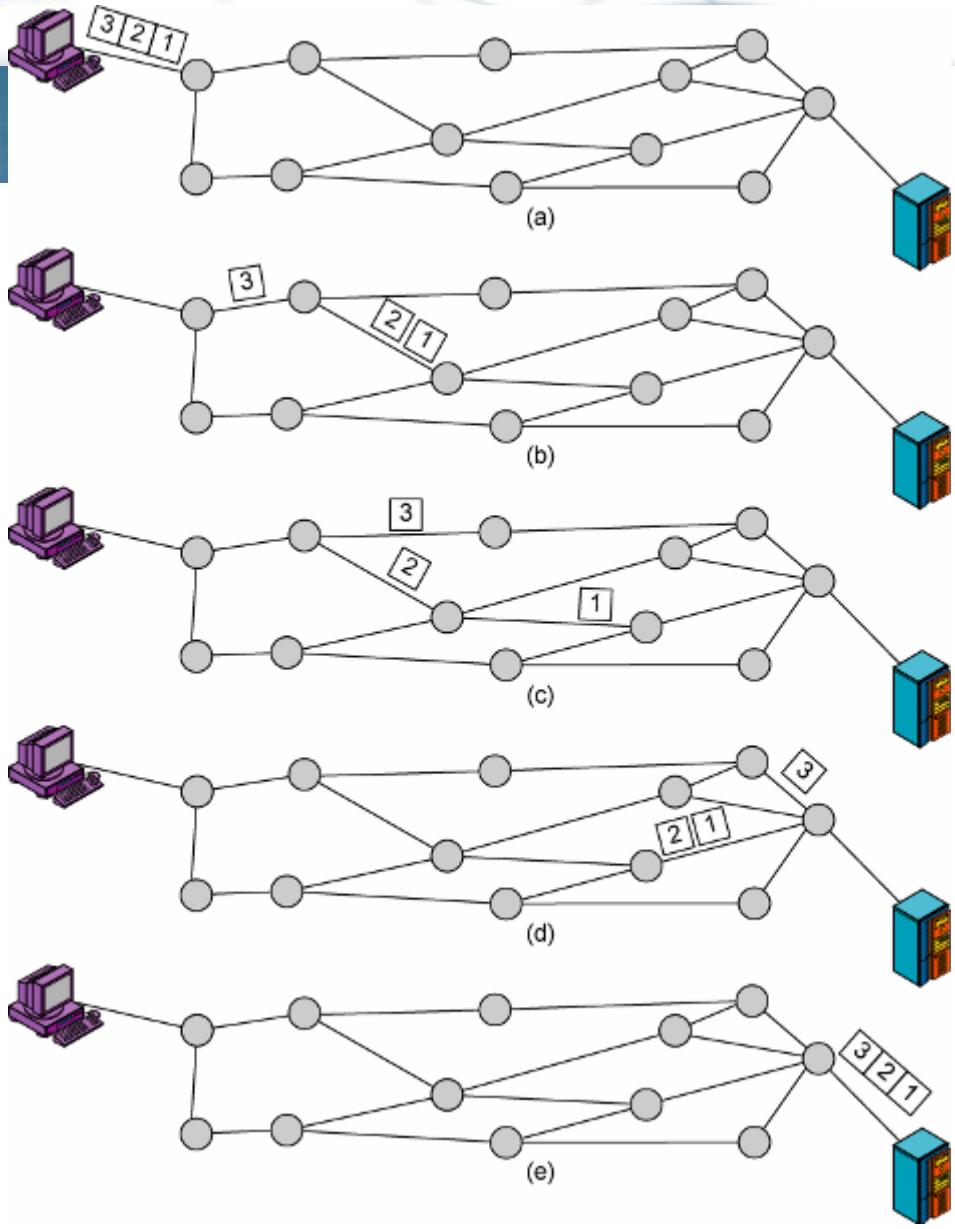


- Packet Switching:
  - Switching decisions are made on individual packets
- Virtual Circuit Switching:
  - A circuit is setup explicitly for individual connections

* Figure is courtesy of B. Forouzan

Packet switching

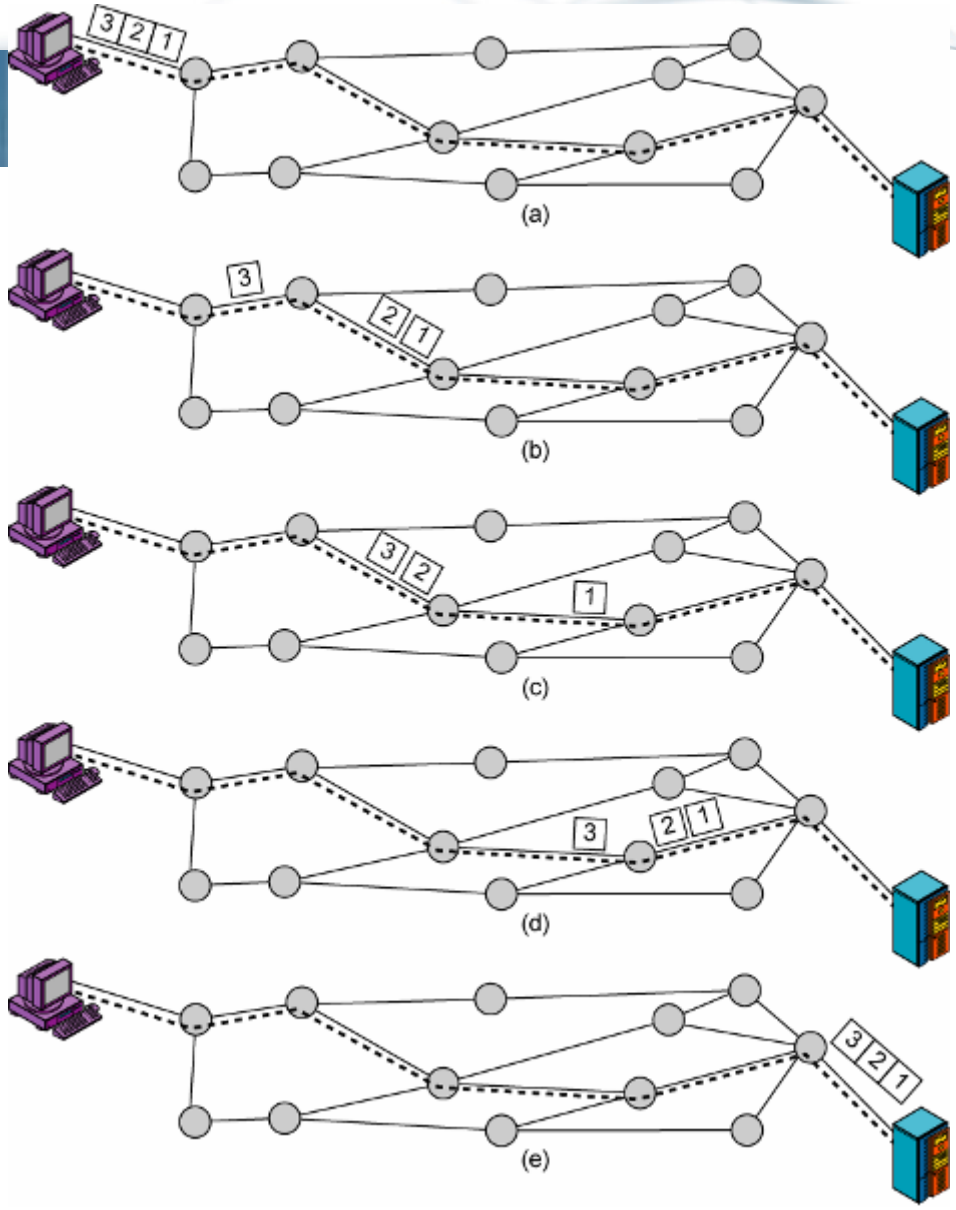
- Frames can be transferred over different paths in the network
- Reliability is generally delegated to higher layers
- Order is not necessarily maintained



* Figure is courtesy of B. Forouzan

Virtual Circuits

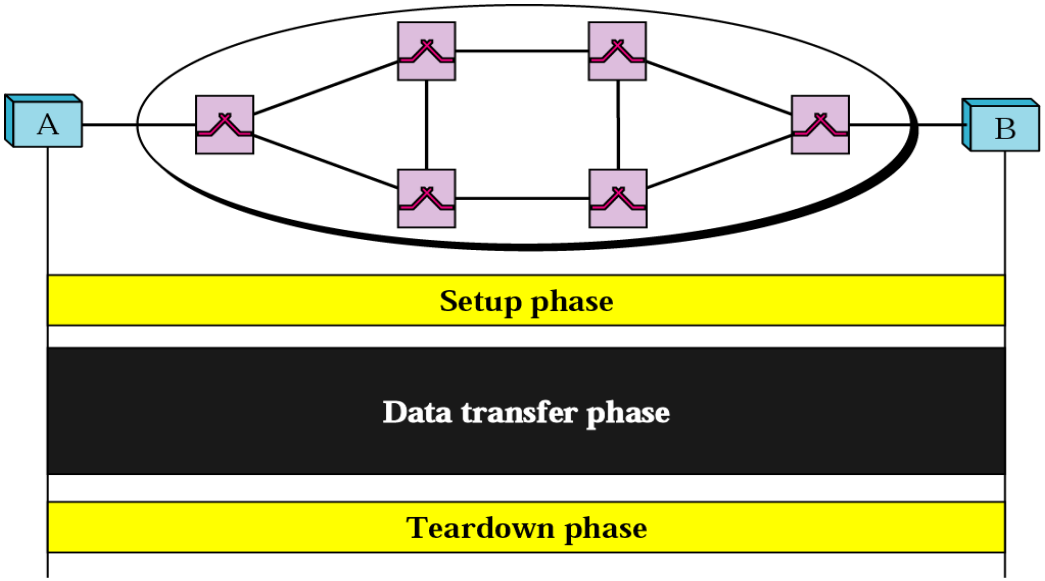
- Connection-oriented communication
- Connection is established before communication
- The network maintains order



* Figure is courtesy of B. Forouzan

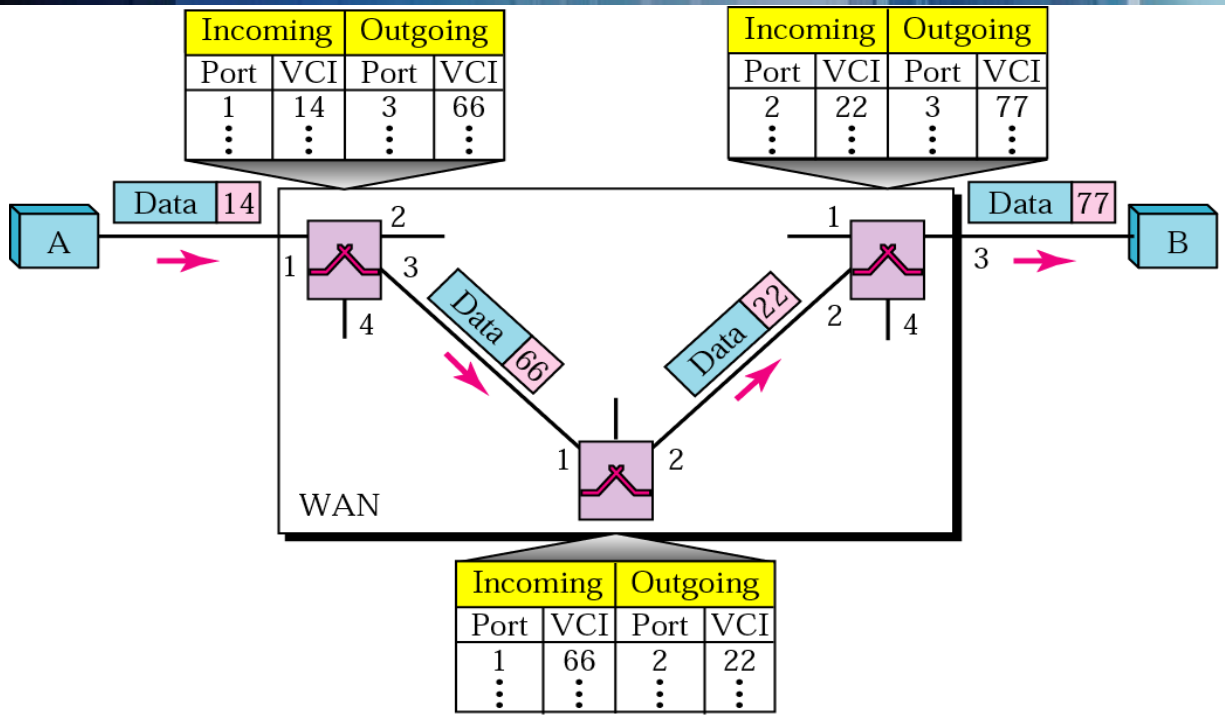
Phases for Virtual Circuit Communication

- Three phases
  - Connection setup
  - Data Transfer
  - Connection termination



* Figure is courtesy of B. Forouzan

Virtual Circuit Switching

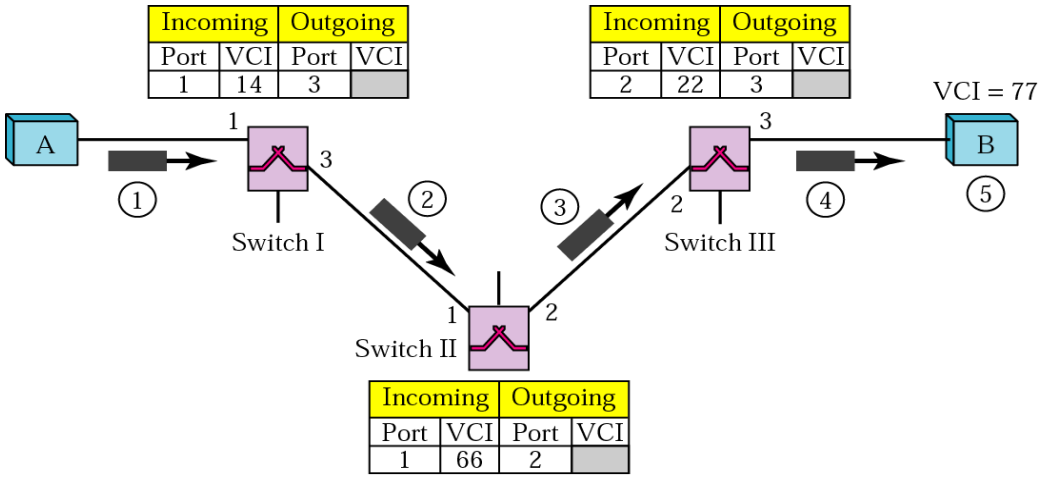


- Every switch maintains a table
  - For duration of communication one entry for incoming and outgoing line
  - Incoming and outgoing line are identified by port number and virtual circuit identifier

* Figure is courtesy of B. Forouzan

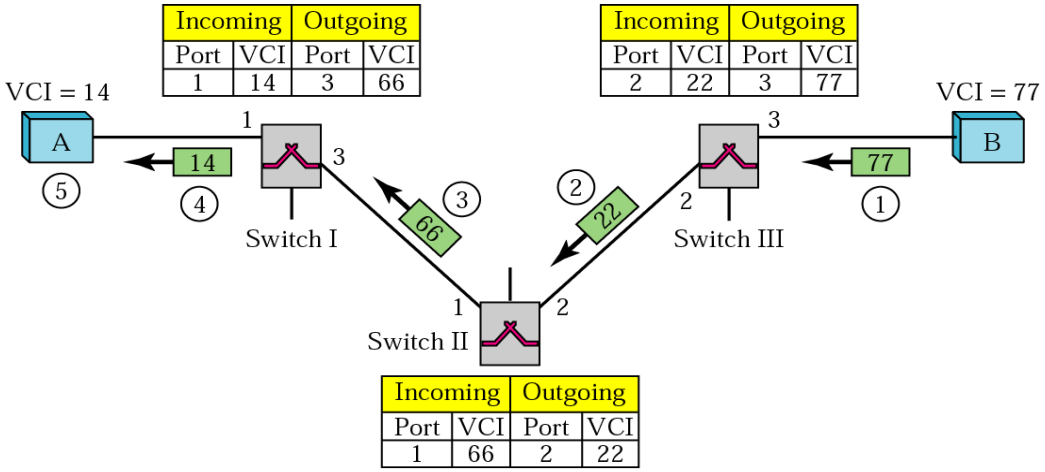
Setup Phase

- Setup request



Incoming		Outgoing	
Port	VCI	Port	VCI
1	66	2	

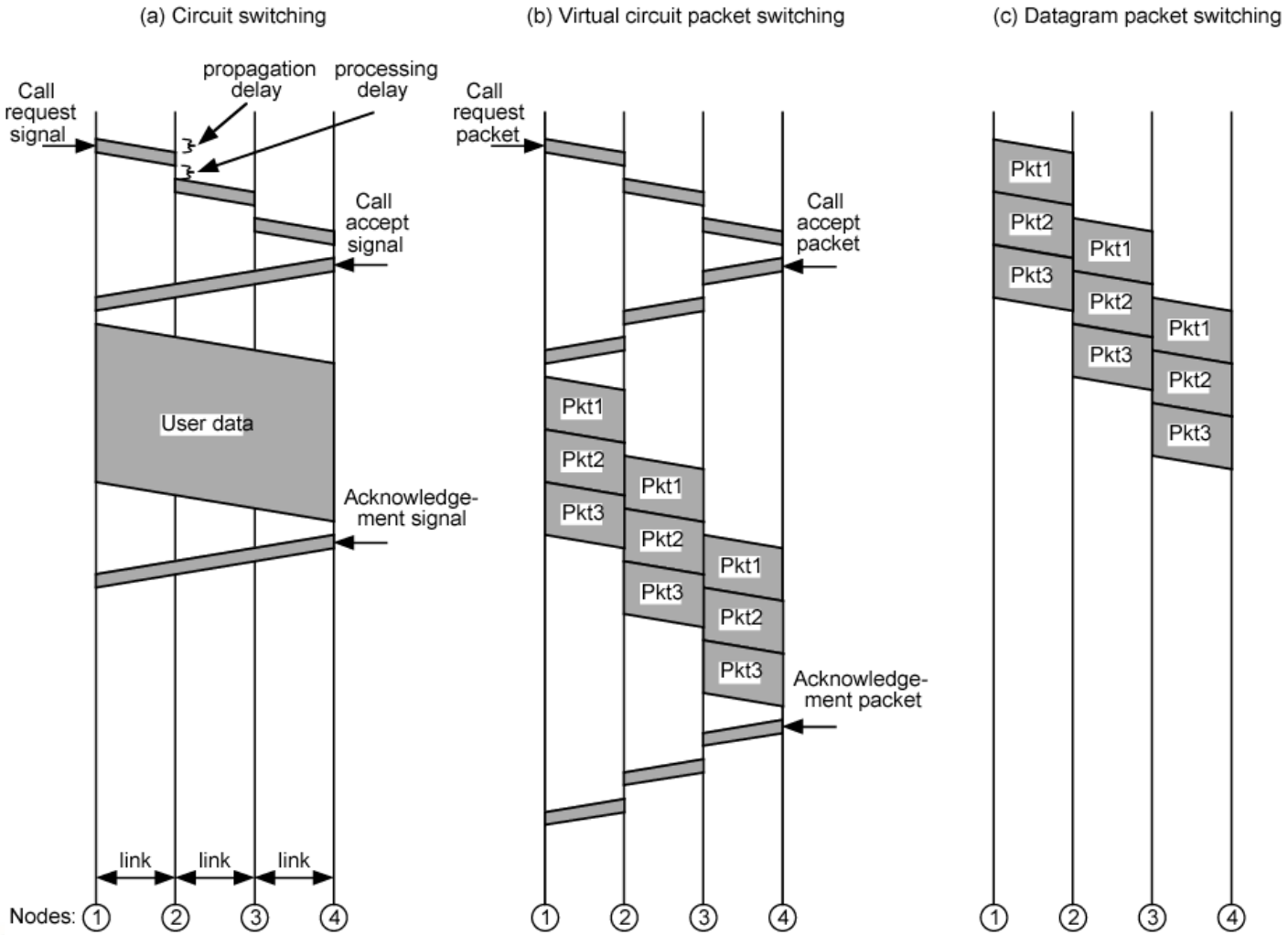
- Setup acknowl.



Incoming		Outgoing	
Port	VCI	Port	VCI
1	66	2	22

* Figure is courtesy of B. Forouzan

Event Timing

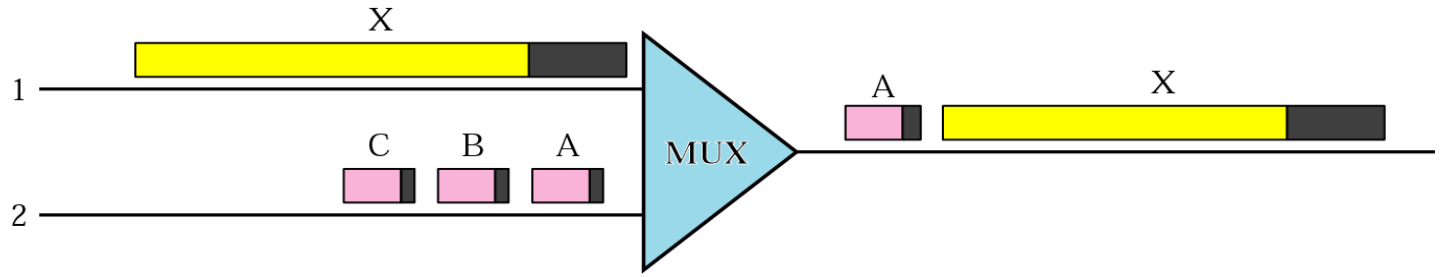


* Figure is courtesy of W. Stallings

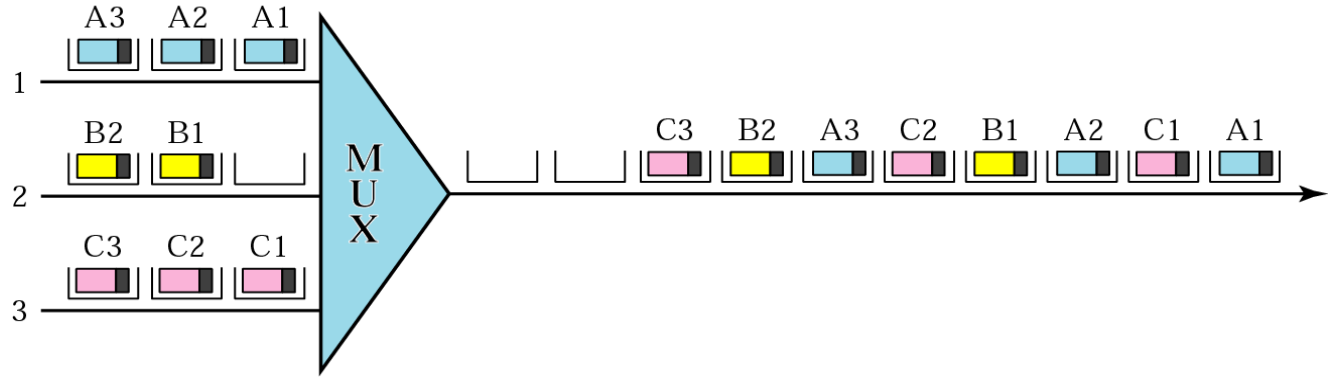
Asynchronous Transfer Mode (ATM)

- Example of virtual circuit switching
  - Cell-Switching
- Similarities between ATM and packet switching
  - Transfer of data in discrete chunks
  - Multiple logical connections over single physical interface
- In ATM flow on each logical connection is in fixed sized packets called cells
- Minimal error and flow control
  - Reduced overhead

Motivation for ATM



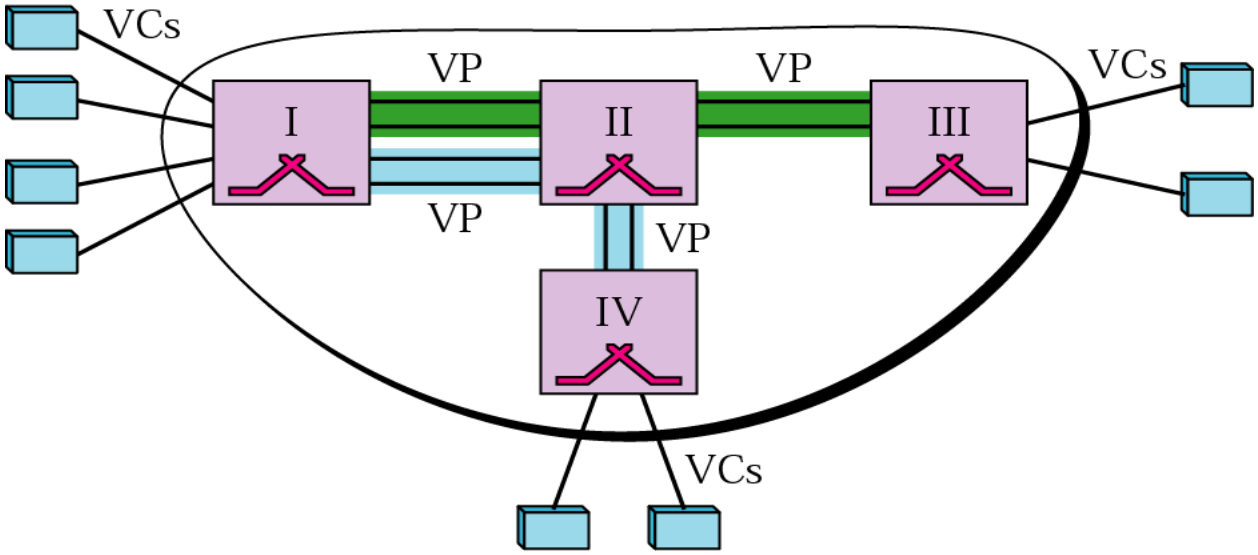
- Frames at a switch may be handled in any order and occupy switch for underspecified time



- Small, fixed-size frames allow simple, fast switches

* Figure is courtesy of B. Forouzan

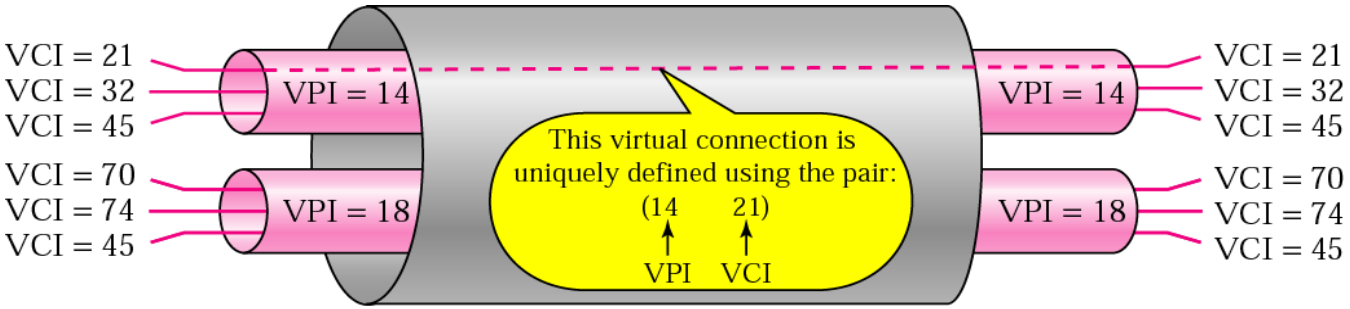
Virtual Circuits / Virtual Paths



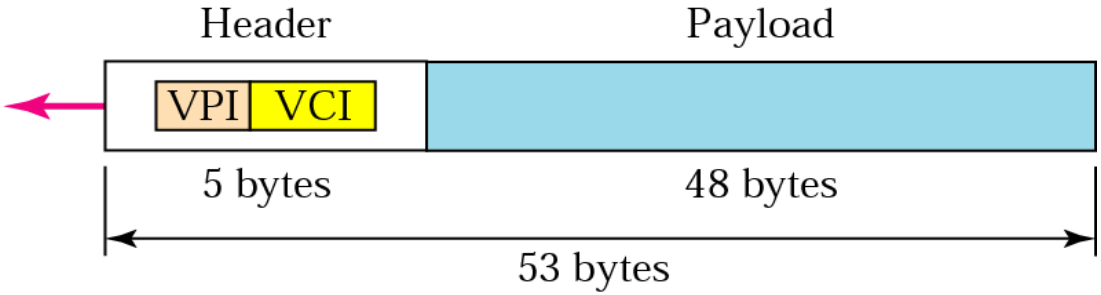
- Virtual circuits are collected into virtual paths

* Figure is courtesy of B. Forouzan

ATM Packet



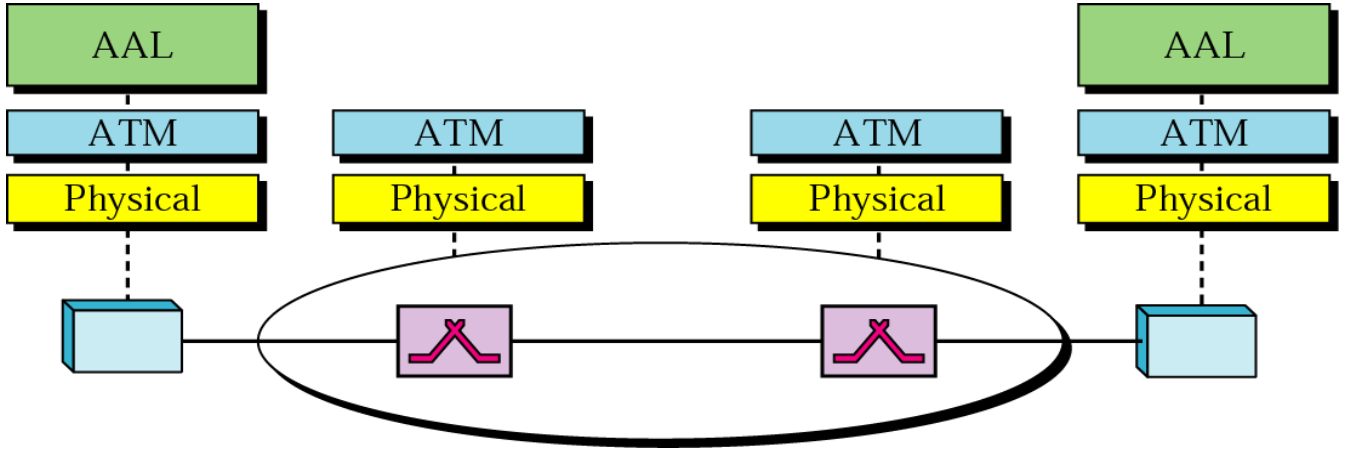
- Connection is specified by combination of Virtual Path ID and Virtual Circuit ID



- Every frame is exactly 53 bytes

* Figure is courtesy of B. Forouzan

Application Adaptation Layer (AAL)



- ATM defined a number of AALs for various purposes (each has its own header format):
  - AAL1: Constant bit rate e.g. multimedia
  - AAL2: Variable-data-rate
  - AAL3/4: Connection-oriented data services
    - Sequencing and Error Control
  - AAL5: Simple and efficient adaptation layer (SEAL)

* Figure is courtesy of B. Forouzan

ATM – It Didn't Happen

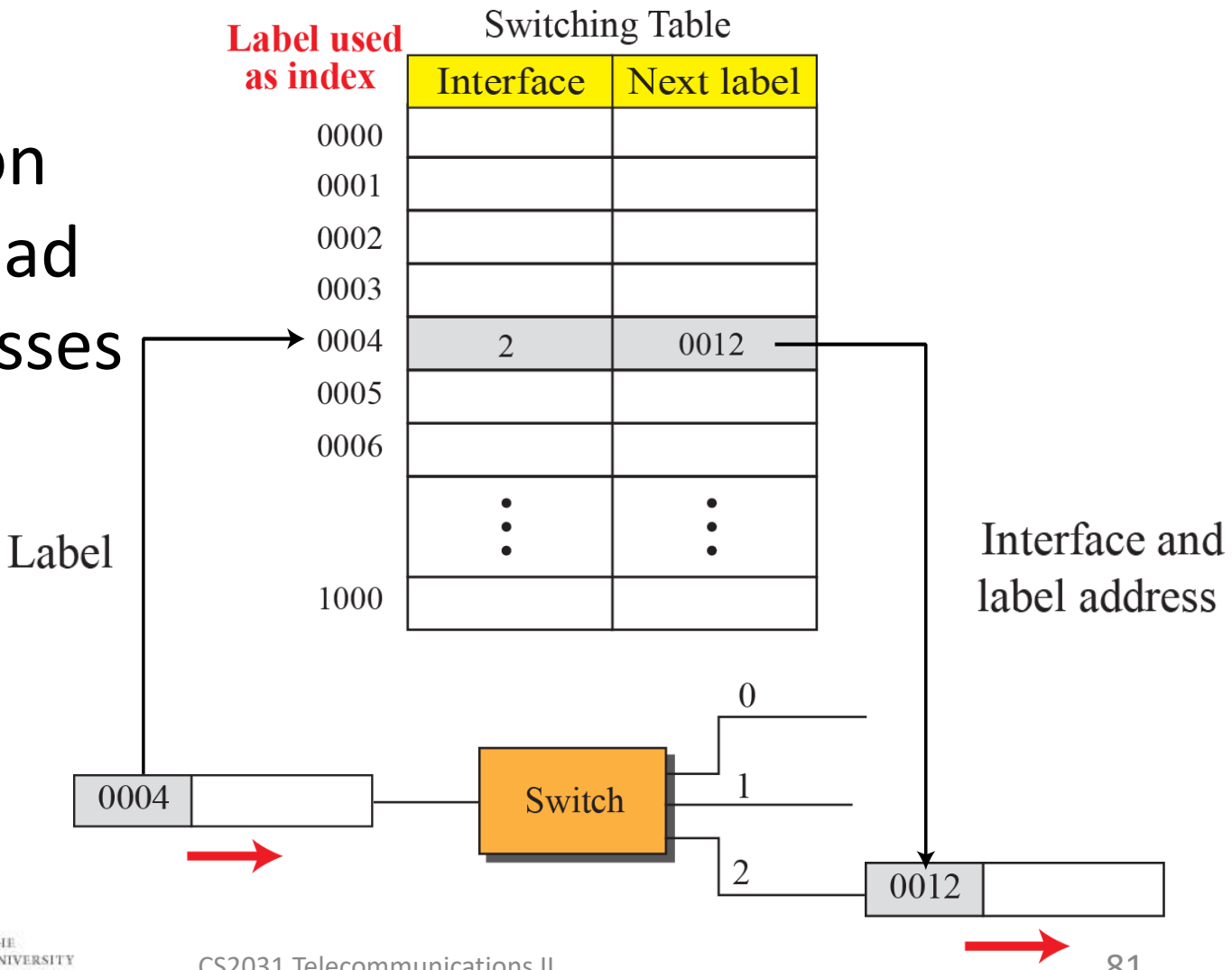
- From Tanenbaum:

  "ATM was going to solve all the world's networking and telecommunications problems by merging voice, data, cable television, telex, telegraph, carrier pigeons, …"

- It didn't happen:
  - Bad Timing
  - Technology
  - Implementation
  - Politics

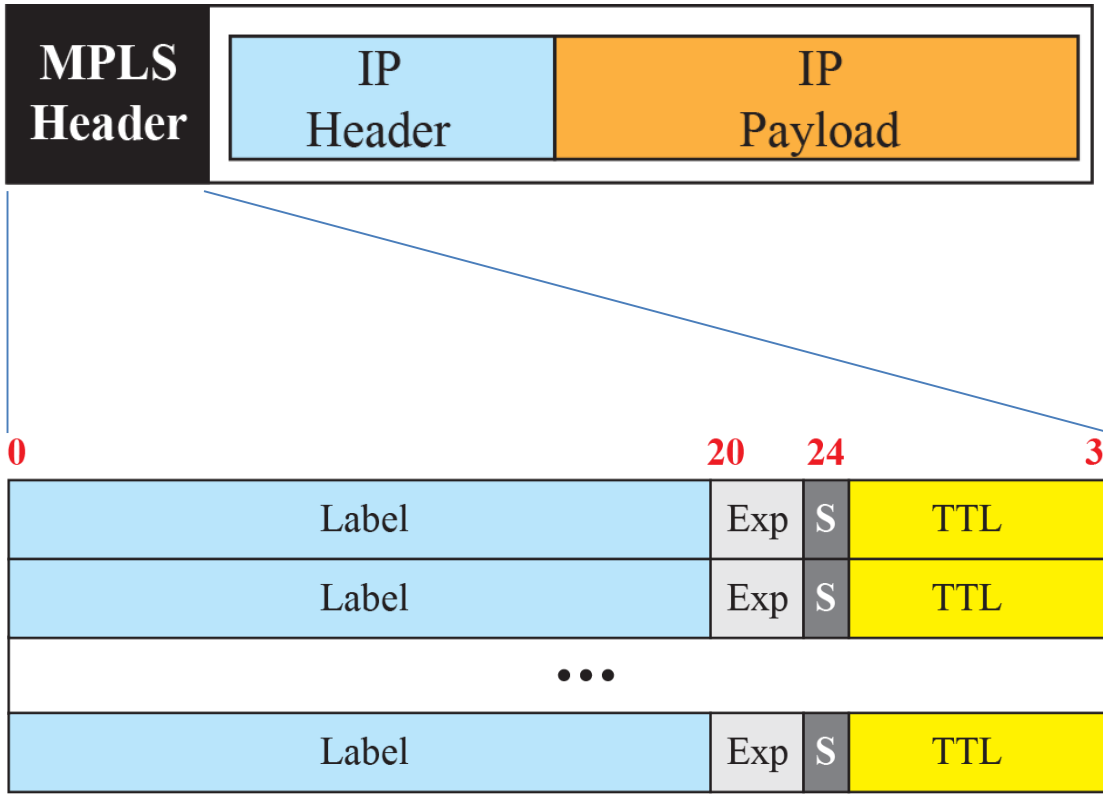
Multiprotocol Label Switching (MPLS)

- Enables switching on labels instead of IP addresses

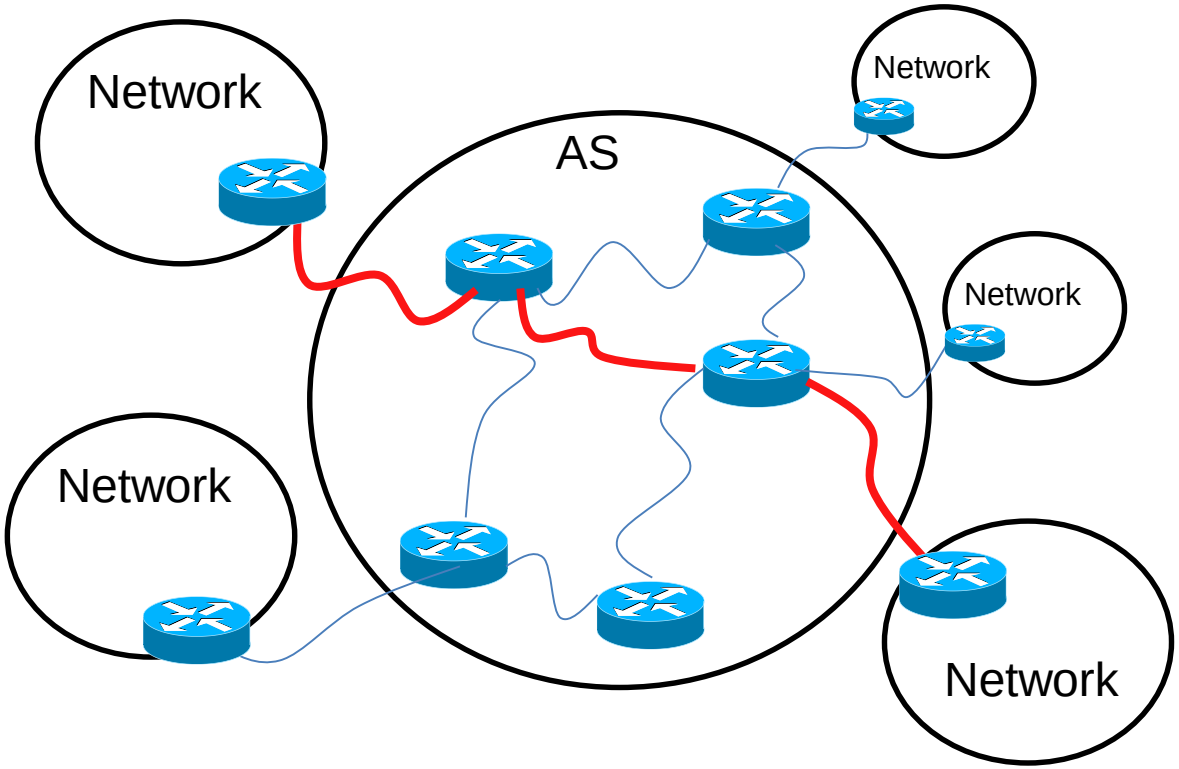


MPLS Header

- MPLS header as stack of labels

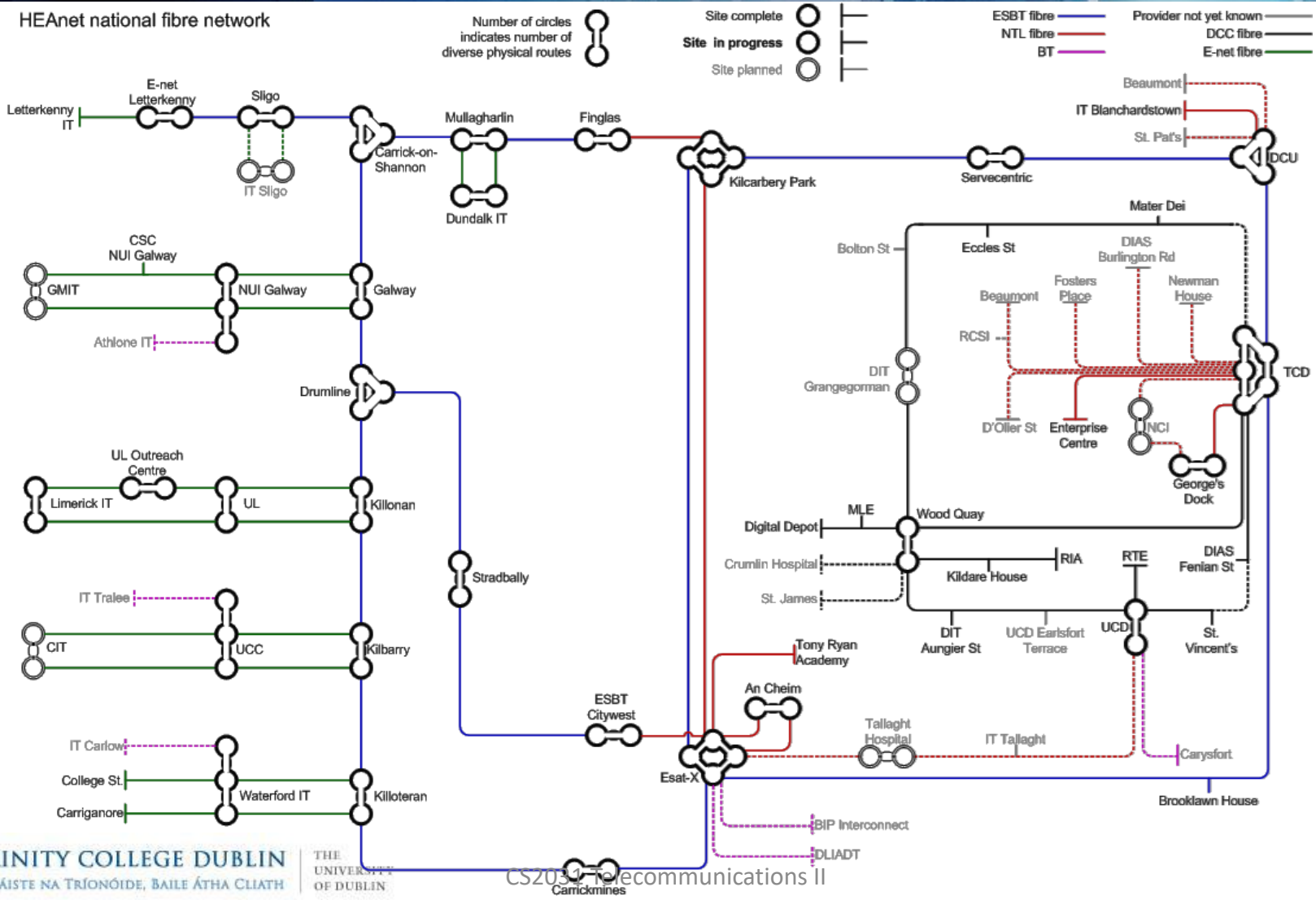


MPLS Use Case



- Creating a virtual network

HEAnet Fibre Network



Summary:Virtual Circuit Switching – ATM

- Virtual Circuit Switching
  - Preplanned route established before any frames sent
  - Call request and call accept frames establish connection (handshake)
  - Each frame contains a virtual circuit identifier instead of destination address
  - No routing decisions required for each frame
  - Clear request to drop circuit
  - Not a dedicated path
- Asynchronous Transfer Mode (ATM)
  - Example for virtual circuit switching
  - Cells consist of 5-byte header and 48-byte payload
  - Circuits identified by virtual circuit ID and virtual path ID
  - Application adaptation layer (AAL) for specific application areas