Applications of SDN

B4: THE GOOGLE USE CASE

Applications of SDN

B4: THE GOOGLE USE CASE

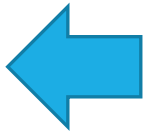
Initial SDN Adopters

Google amazon facebook Microsoft

Cloud providers, to enable efficient upscaling of data centers

Use cases:

- Switching fabrics – SDN within data centers: Handle traffic between racks in DC
  - From black box to white box switches, server interconnect controlled by software
- Virtual networks – SDN to set up, manage, tear down:
  - Easy-to-use VPNs and VLANs VPN L3 / VLAN L2
- Wide area networks – traffic engineering between sites: more efficiently link in WAN
  - Dynamically managing end-to-end paths in links between data centers



Google WAN

Used to inter-DC traffic

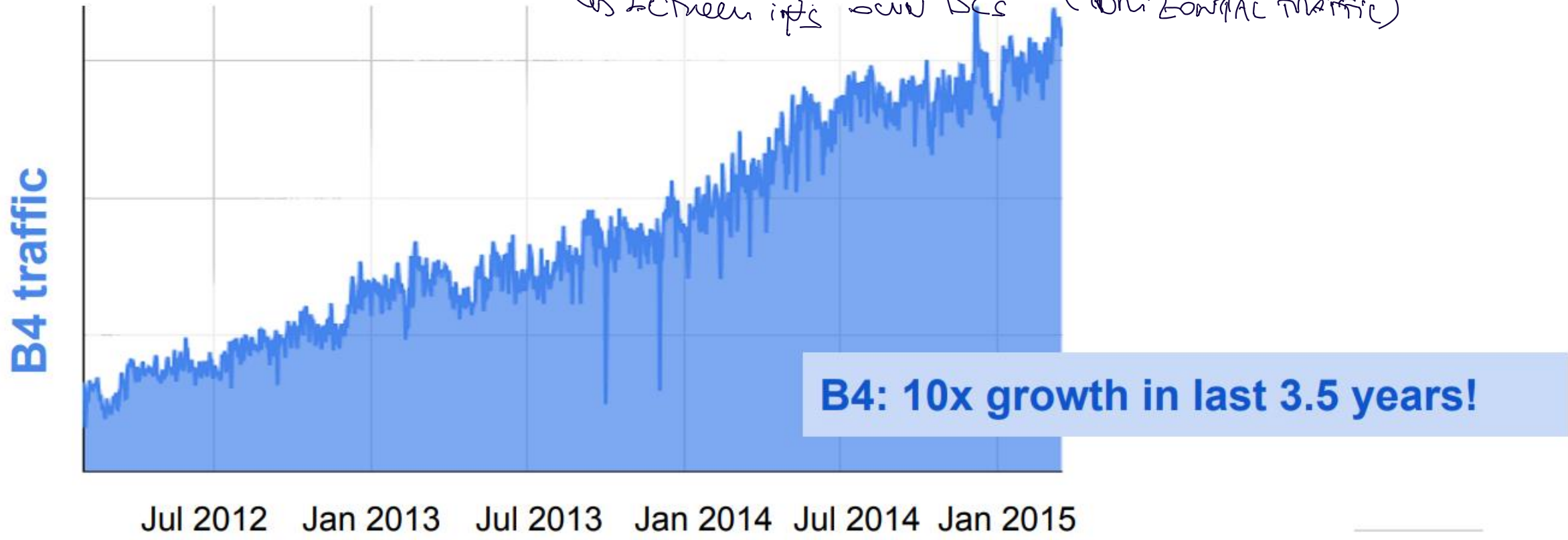
https://www.infinera.com/videos/real-world-transport-sdn-success-cases/tag/software-and-automation/

Two Backbones

Google

Two separate backbones:

- B2: Carries Internet facing traffic → Growing faster than the Internet
- B4: Inter-datacenter traffic → More traffic than B2, growing faster than B2



B4 – Google's software defined WAN

- Google's globally deployed WAN, datacenter backbone
  - private WAN connecting its data centers (NOT: Internet-facing backbone for user traffic)
- Modest number of sites (~ dozen)

Between DCs

Motivation for Backend Backbone

Data centers deployed across the world

- Serve content with geographic locality
- Replicate content for fault tolerance

Need a network to connect these data centers to one another

- Not on the public Internet
- Cost effective network for high volume traffic
- Bursty/bulk traffic (not smooth/diurnal)

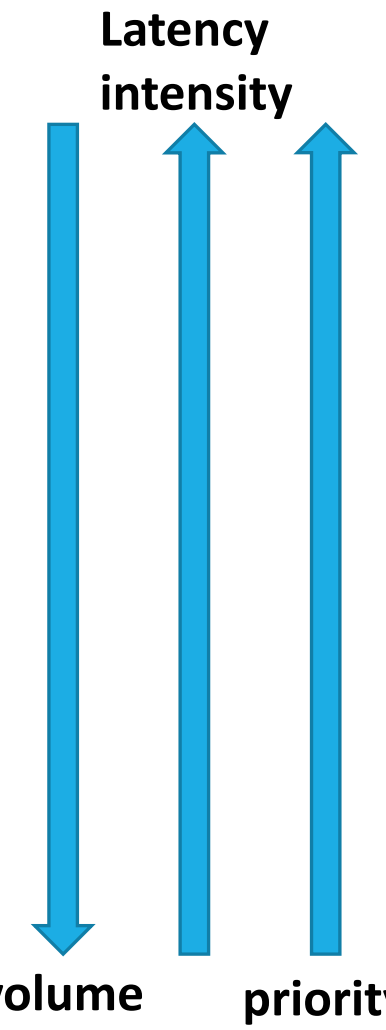
WAN Intensive Apps

YouTube Web Search
Google+ Maps AppEngine
Photos and Hangouts
Android/Chrome Updates

B4 – Google's software defined WAN

Main Traffic flow types

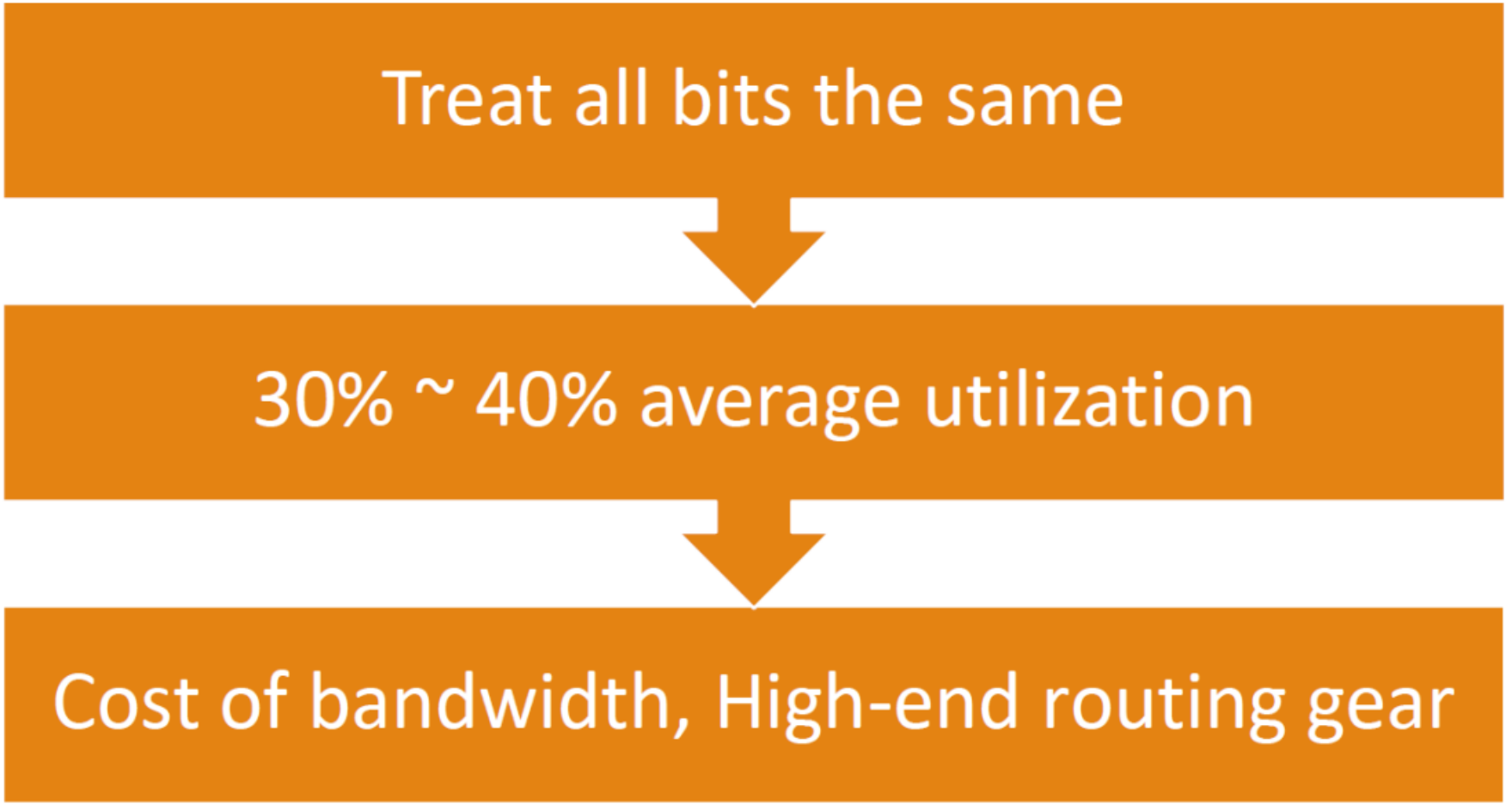
1. **User data copies** to remote data centers for availability/reliability (lowest volume, most latency intensive, highest priority)
2. **Remote storage access** for computation over distributed data sources
3. Large-Scale data push **synchronizing state across multiple data centers** (highest volume, least latency intensive, lowest priority)



Cost is a major issue (high growth rate of WAN traffic)

- WAN links are typically utilized 30-40% 2-3 x bandwidth overprovisioning

Traditional WAN routing



B4 requirements

1. **Elastic bandwidth demand**
   - Applications can tolerate periodic failures or temporary bandwidth reductions
2. **Moderate number of sites**
   - Few dozen of sites, no need for large routing tables
3. **End application control**
   - Fine-grained application control (no need for link overprovisioning)
4. **Cost sensitivity**
   - Private intercontinental links cost, so it needs to be used at its full capacity (100%)

Main design decisions

1) **separate HW from SW** : simple/low cost HW and put intelligence in SW
   - Customize routing and monitoring protocols to requirements
   - Rapid custom protocol development
   - Leverage powerful compute in Google servers

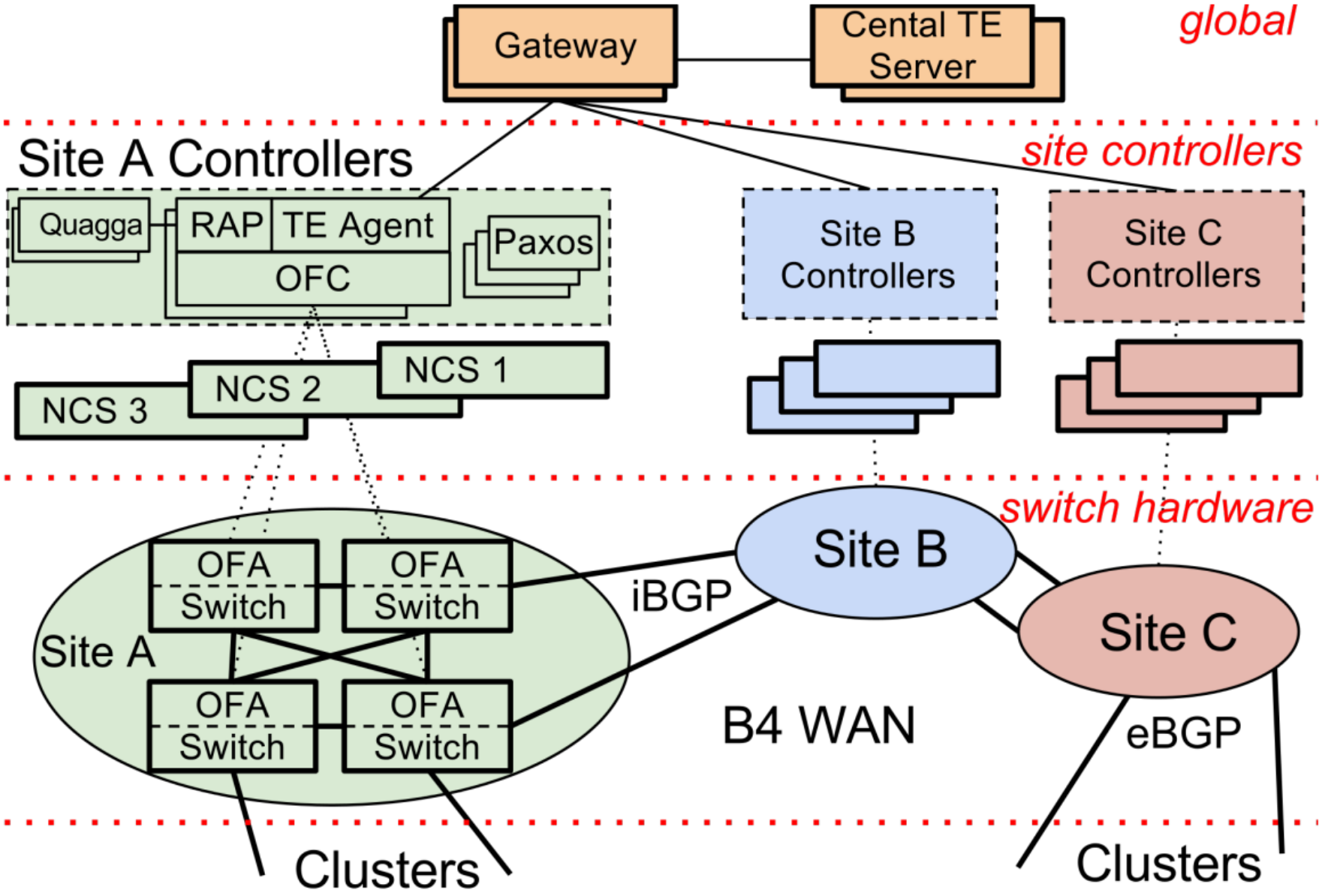
2) **low cost switching hardware**
   - Edge application control limits need for large buffers
   - limited number of sites -> no need of large forwarding tables

3) **centralize traffic engineering (TE)**
   - Allow more optimal and faster TE than distributed control routing
   - Share bandwidth among competing applications
   - TE server is an application on top of an SDN ~~distributed~~ control layer

4) **increase link utilization**
   - Efficient use of expensive long haul transport
   - Largest bandwidth consumers adapt to available bandwidth

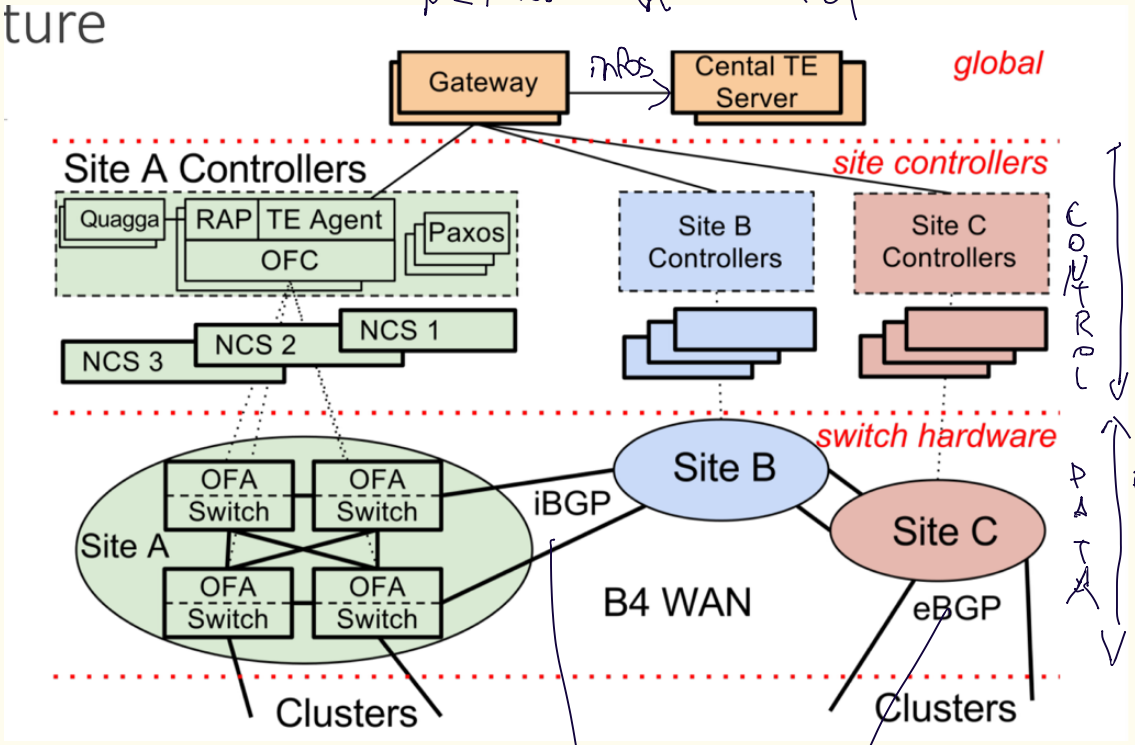
Traffic Engineering: techniques to optimize and control traffic flows in a telecommunication network in order to ensure a maximum throughput and a sufficient QoS level for your flows

B4 architecture overview



ture

network-app. control



- SITE A, B, C = Datacenters
- Set of OpenFlow switche in each site
- OFA: openflow agent that communicates with the controller
- each SITE has a SITE CONTROLLER

need ≠ site controller:
multiples SDN controller (replication, performance, ...) reliability...

NCS: network control server, host controller

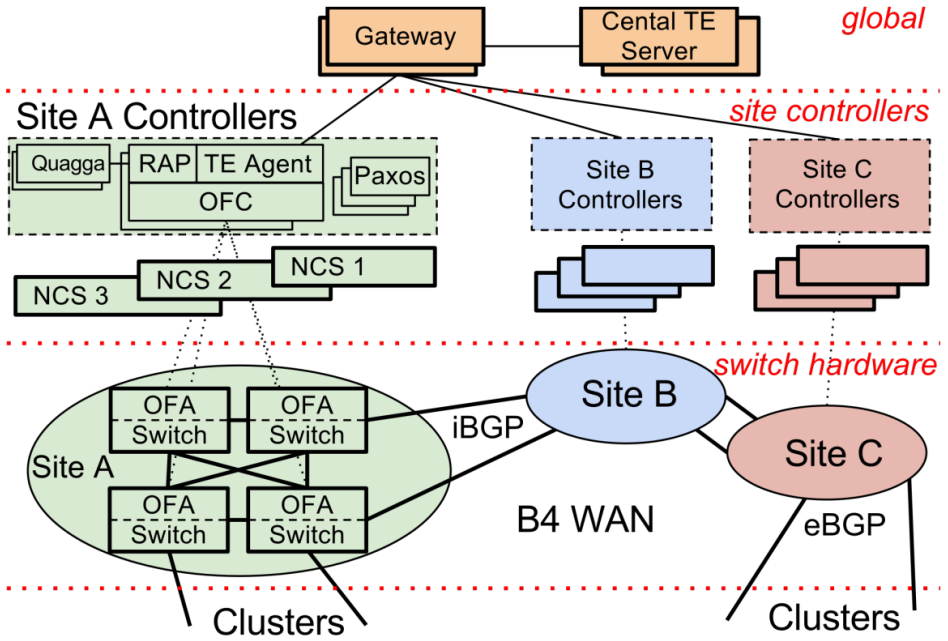
internal / external BGP session

=> DIFFERENT FLOW: routes provided by controller & by BGP

(high prior)

level (priority)

B4 Architecture overview



Switch hardware

- Forwards traffic
- No complex control software
- Google custom designed with commodity silicon

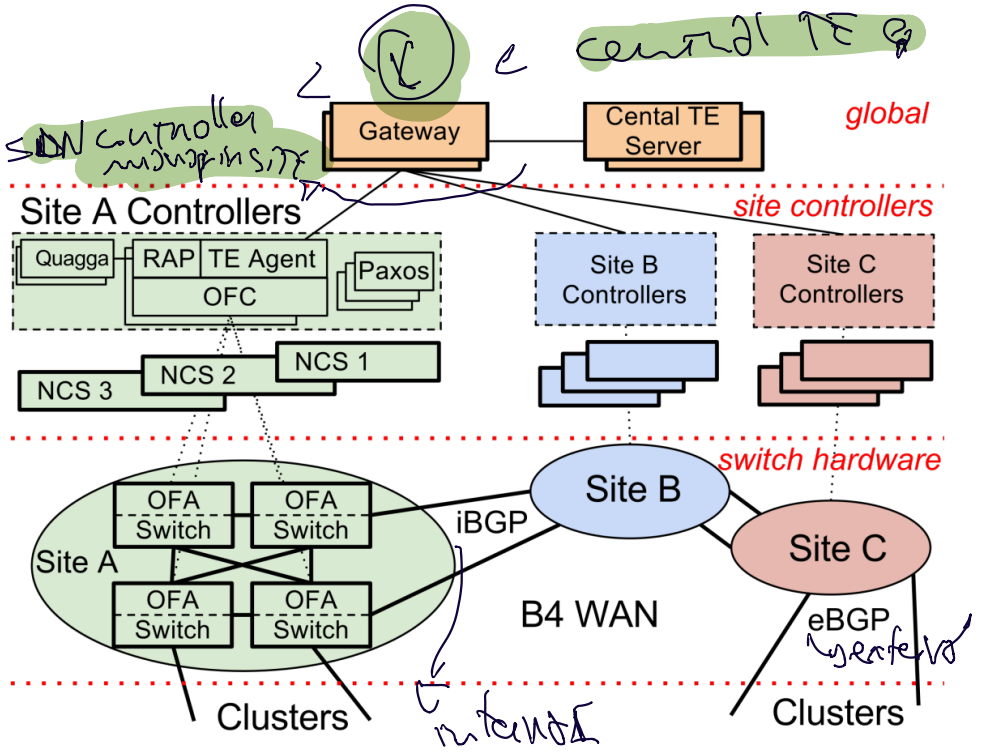
Control layer

- Network Control Servers (NCS) hosting both OpenFlow controllers (OFC) and Network Control Applications (NCAs).
- OFCs maintain network state based on NCA directives and switch events
- OFCs instruct switches to set forwarding table entries based on this changing network state

B4 Architecture overview

Global layer

- Logically centralized applications: SDN Gateway and a central TE server)
- SDN Gateway abstracts details of OpenFlow and switch hardware from the central TE server.
- Each B4 site consists of multiple switches with potentially hundreds of individual ports



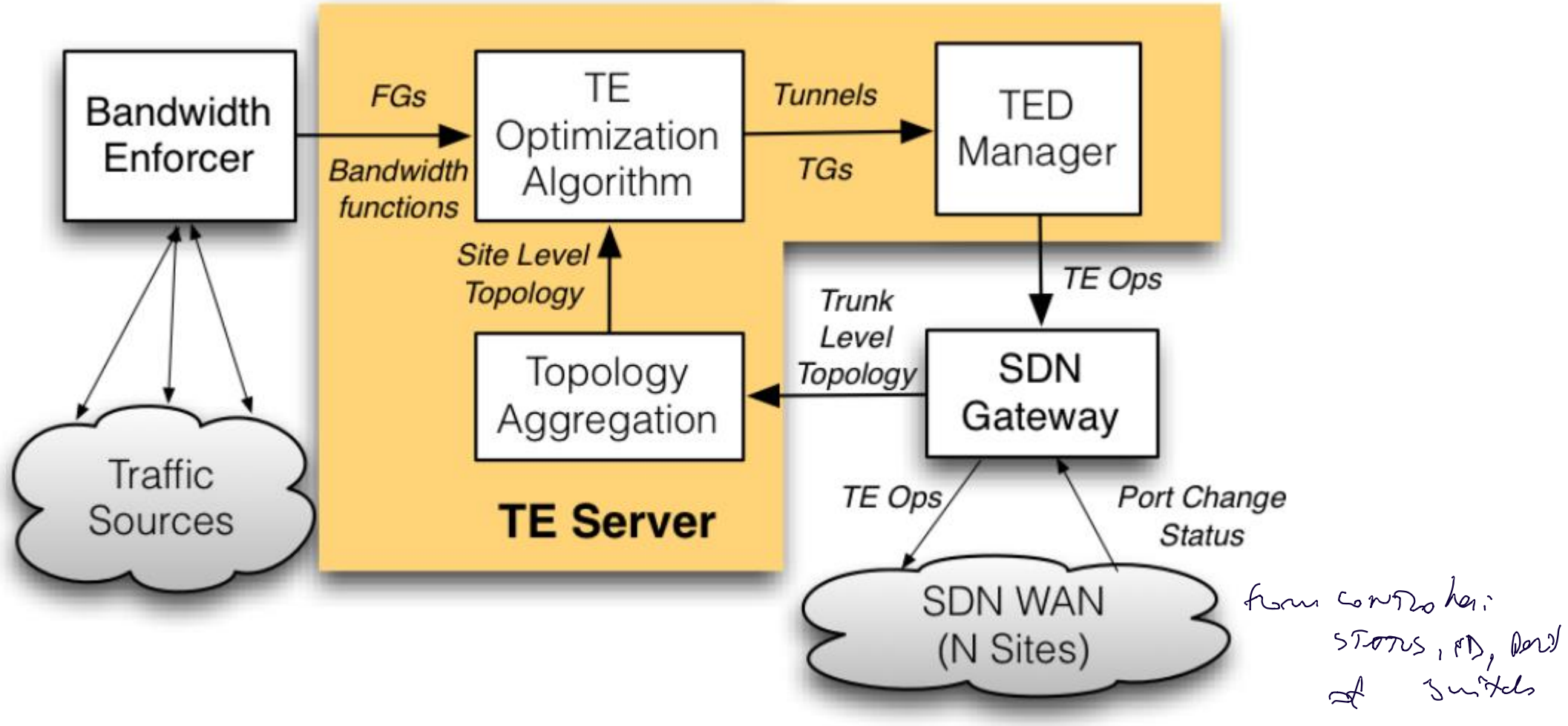
Traditional WAN integrated with SDN

Each cluster contains a set of BGP routers that peer with B4 switches at each WAN sites

Why BGP?

- operator familiarity with the protocol
- used before introducing SDN
- enabling a gradual rollout

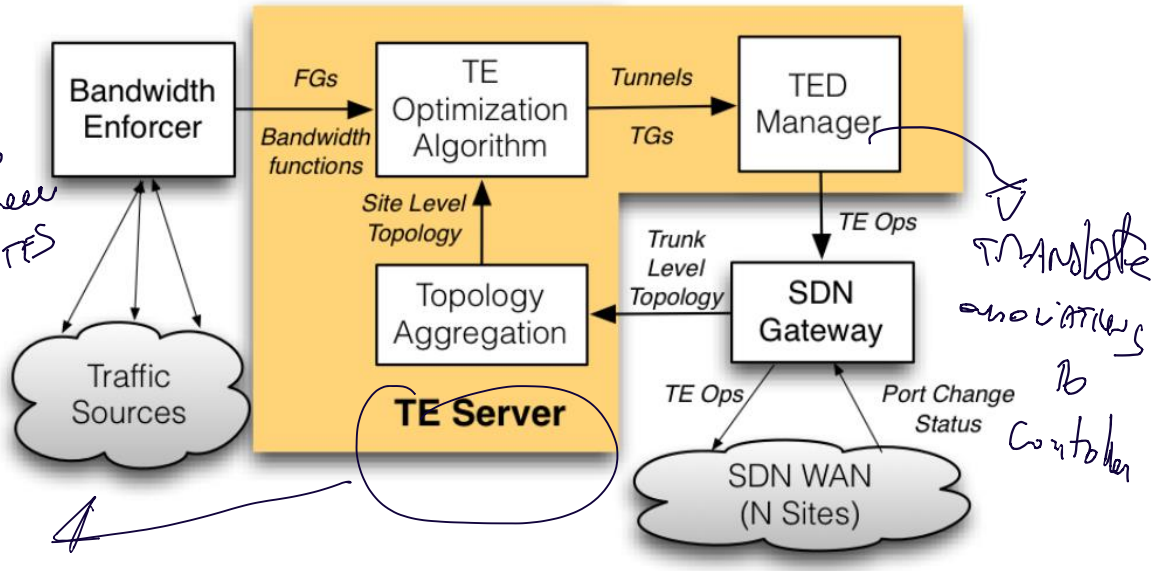
B4 – Traffic Engineering



B4 – Traffic Engineering

TE Server

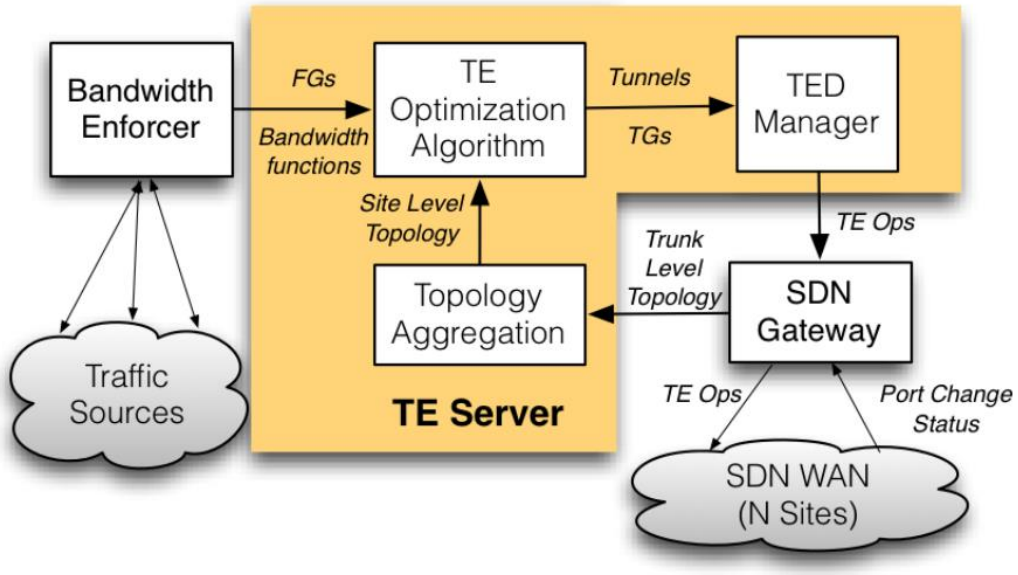
- aggregates trunks to compute site-site edges.
- This abstraction reduces the size of the topology graph in input to the TE Optimization Algorithm
- Applications are aggregated as a flow group (FG) defined as {source site, dest site, QoS} tuple
- A Tunnel (T) represents a site-level path in the network, e.g., a sequence of sites
- Tunnel Group (TG) maps FGs to a set of tunnels



B4 – Traffic Engineering

Bandwidth Enforcer

- Specified the bandwidth allocation to a specific application (Bandwidth functions)
- These functions are derived from administrator-specified static weights



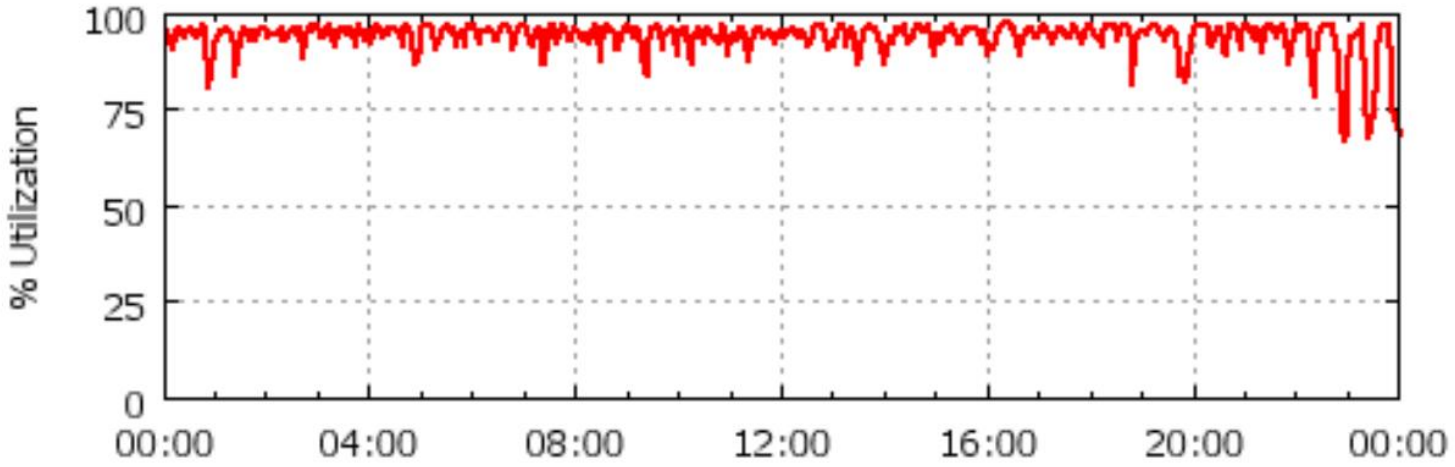
TE Server outputs the Tunnel Groups and, by reference, Tunnels and Flow Groups to the SDN Gateway.

The Gateway forwards these Tunnels and Flow Groups to OFCs that in turn install them in switches using OpenFlow

At the end…

OpenFlow has helped Google **improving B4 backbone[a]**

- Utilization near 100% for elastic loads
- Mirror production events for testing on virtualized switches
- Reduced complexity
- Reduced costs



Traffic on all links between two B4 sites

Follow-up Improvements[b]

- hierarchical topology
- decentralized TE algorithms

References

a. Sushant Jain et al. “B4: experience with a globally-deployed software defined wan”. In Proc. of the ACM SIGCOMM 2013 ACM, New York, NY, USA, 3-14

b. Hong CYY et al.,. B4 and after: managing hierarchy, partitioning, and asymmetry for availability and scale in google's software-defined WAN. In Proc. Proc. of the ACM SIGCOMM 2018 Aug 7 (pp. 74-87).

Additional use cases

Transport network and entreprise network services

- https://www.infinera.com/control-automation/
- https://www.telstra.com.au/business-enterprise/products/networks/sdn/telstra-programmable-network
- https://www.blueplanet.com/resources/analysys-mason-telefonica-germany-is-partnering-with-blue-planet-to-execute-its-ifusion-transport-sdn-strategy.html
- https://www.timenterprise.it/connectivity/rete-fissa/tim-sdwan

References

a. Sushant Jain et al. “B4: experience with a globally-deployed software defined wan”. In Proc. of the ACM SIGCOMM 2013 ACM, New York, NY, USA, 3-14

b. Hong CYY et al.,. B4 and after: managing hierarchy, partitioning, and asymmetry for availability and scale in google's software-defined WAN. In Proc. Proc. of the ACM SIGCOMM 2018 Aug 7 (pp. 74-87).

c. Lessons Learned from B4, Google’s SDN WAN, Subhasree Mandal, July 9, 2015

Switches with OpenFlow support - examples

Cisco Catalyst 9300 series

https://www.cisco.com/c/en/us/support/docs/switches/catalyst-9300-series-switches/217210-understand-openflow-on-catalyst-9000-ser.html

Noviflow

https://noviflow.com/noviswitch/

Juniper

https://www.juniper.net/documentation/us/en/software/junos/sdn-openflow/topics/concept/junos-sdn-openflow-supported-platforms.html