A Software Defined WAN Architecture

Cloud Computing Requires Massive Wide-Area Bandwidth

Google™

- Low latency access from global audience and highest levels of availability
  - Vast majority of data migrating to cloud
  - Data must be replicated at multiple sites
- WAN unit costs decreasing rapidly
  - But not quickly enough to keep up with even faster increase in WAN bandwidth demand

WAN Cost Components

Google

- Hardware
  - Routers
  - Transport gear
  - Fiber
- Overprovisioning
  - Shortest path routing
  - Slow convergence time
  - Maintain SLAs despite failures
  - No traffic differentiation
- Operational expenses/human costs
  - Box-centric versus fabric-centric views

Why Software Defined WAN

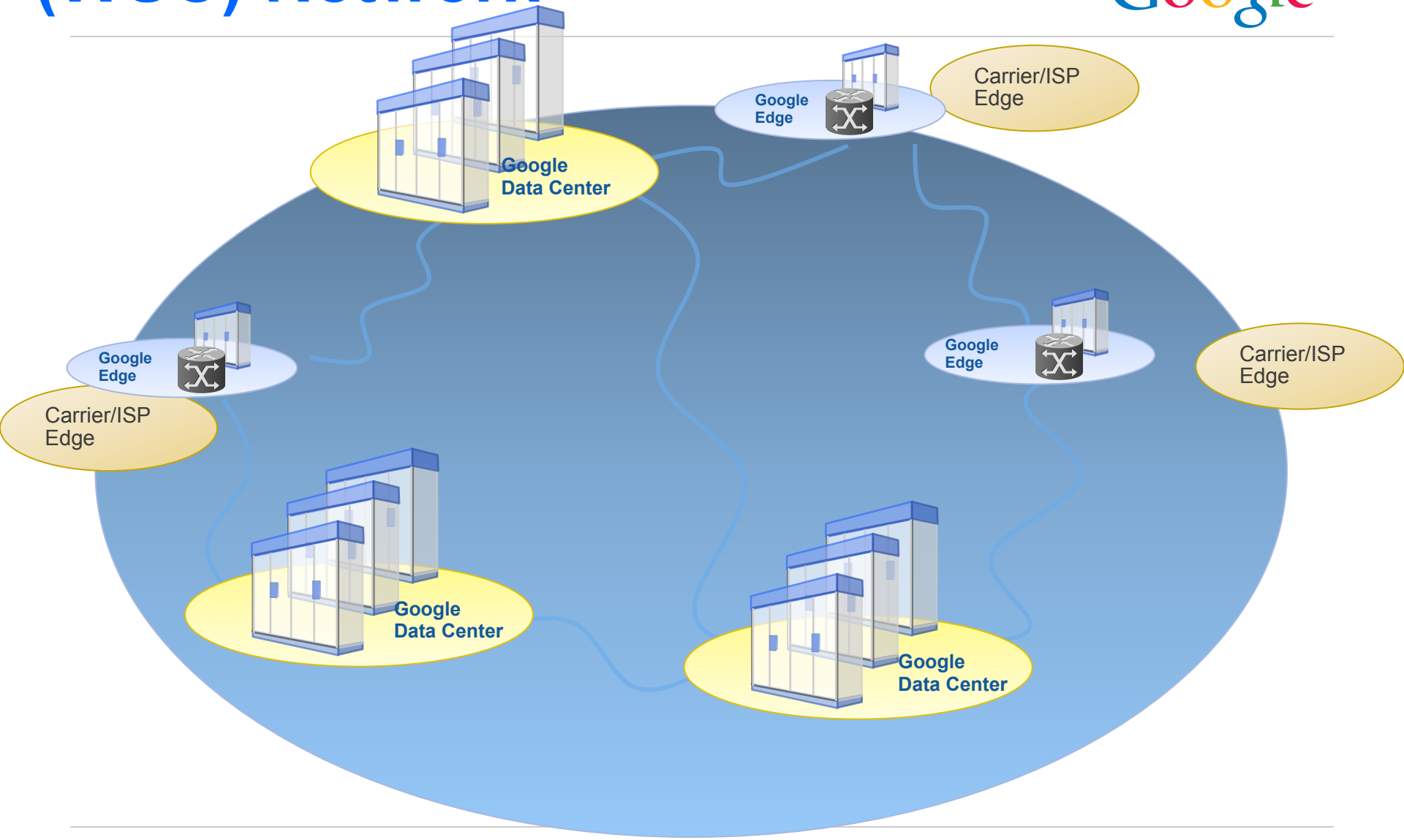
Google™

- Separate hardware from software
  - Choose hardware based on necessary features
  - Choose software based on protocol requirements
- Logically centralized network control
  - More deterministic
  - More efficient
  - More fault tolerant
- Automation: Separate monitoring, management, and operation from individual boxes
- *Flexibility and Innovation*

Result: A WAN that is more efficient, higher performance, more fault tolerant, and cheaper

Google's Software Defined WAN

A Warehouse-Scale-Computer (WSC) Network



Google's WAN

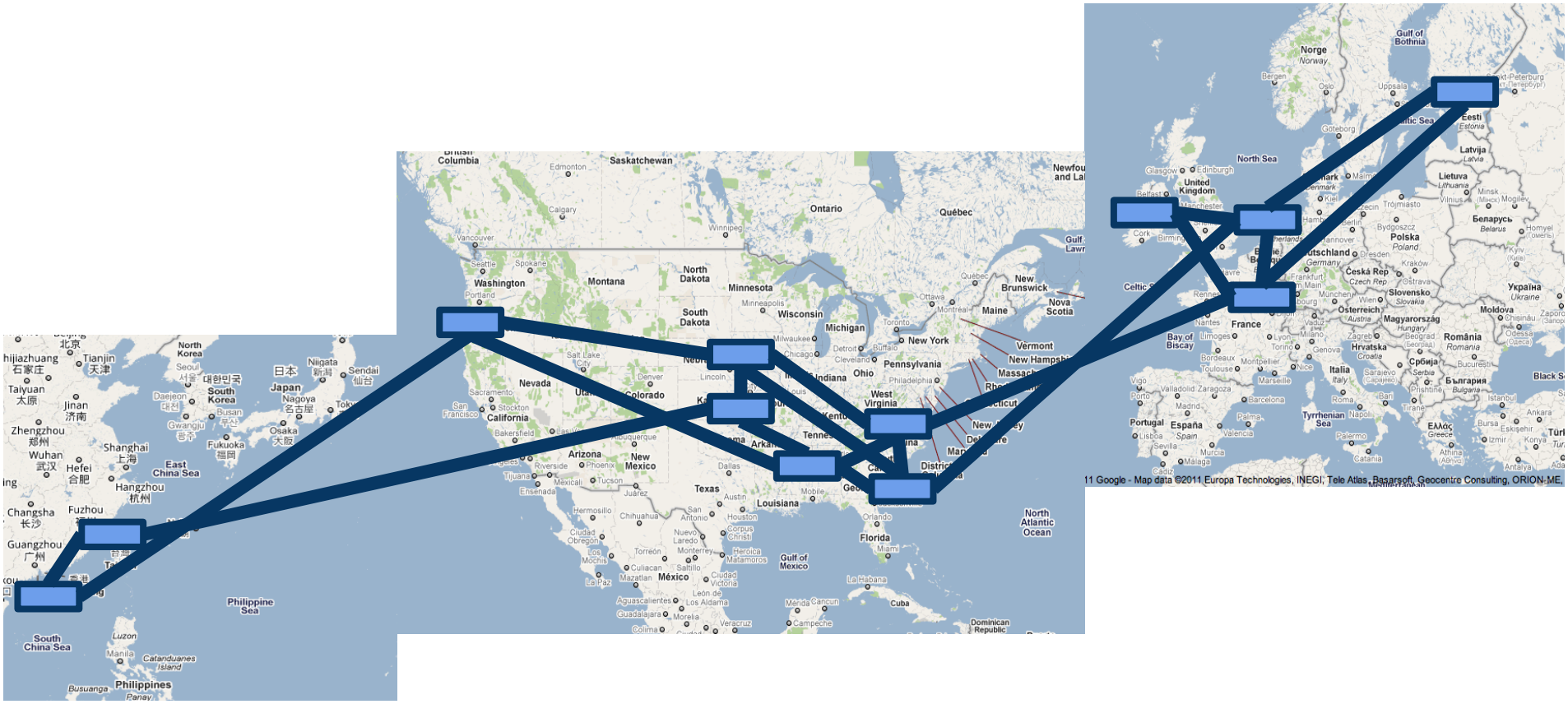
Google™

- Two backbones
  - I-Scale: Internet facing (user traffic)
  - G-Scale: Datacenter traffic (internal)

- Widely varying requirements: loss sensitivity, topology, availability, etc.

- Widely varying traffic characteristics: smooth/diurnal vs. bursty/bulk

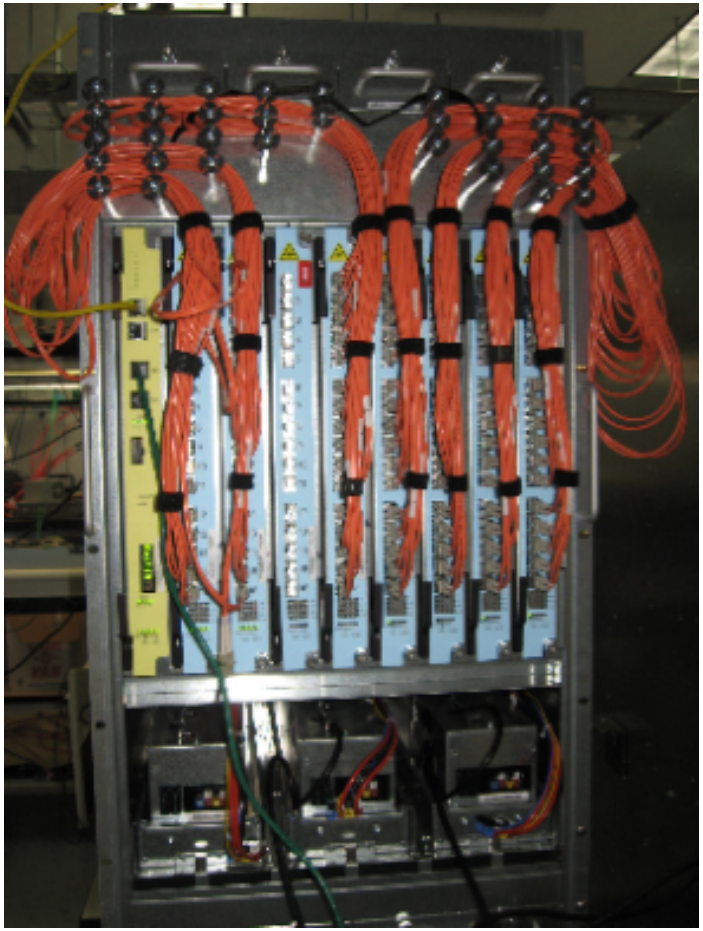
Google's Software Defined WAN



G-Scale Network Hardware

Google

- 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



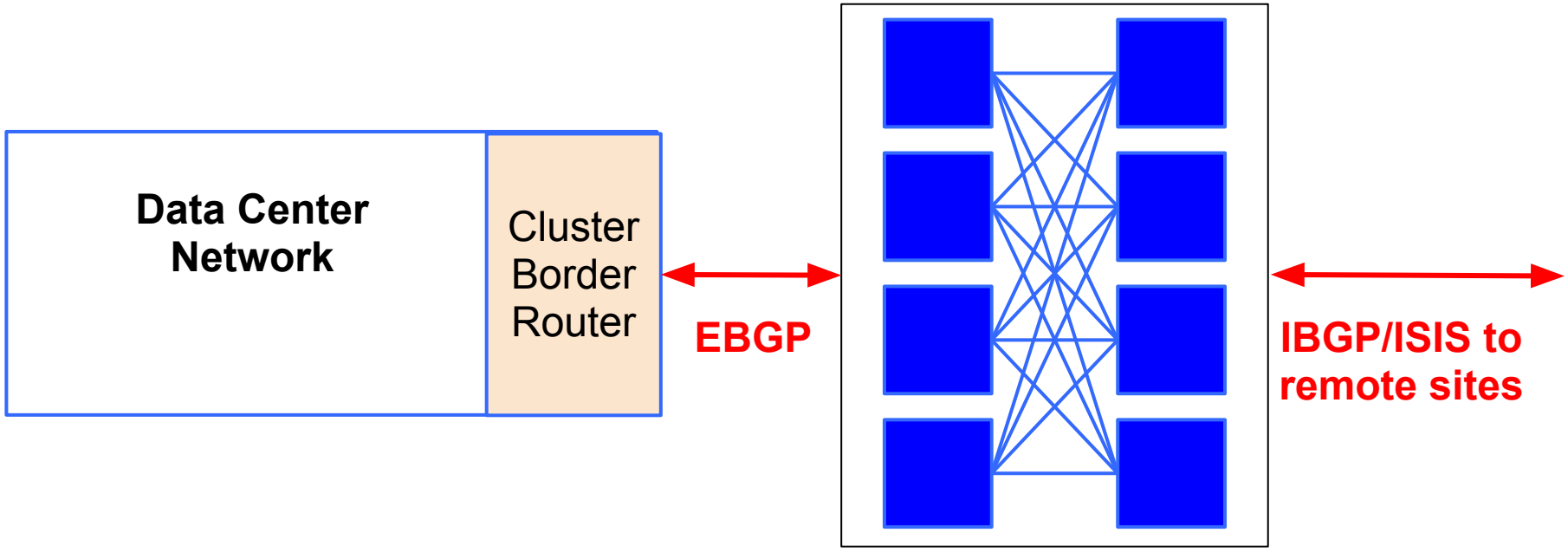
G-Scale WAN Deployment

Google™

- Multiple switch chassis in each domain
  - Custom hardware running Linux
- Quagga BGP stack, ISIS/IBGP for internal connectivity

Mixed SDN Deployment

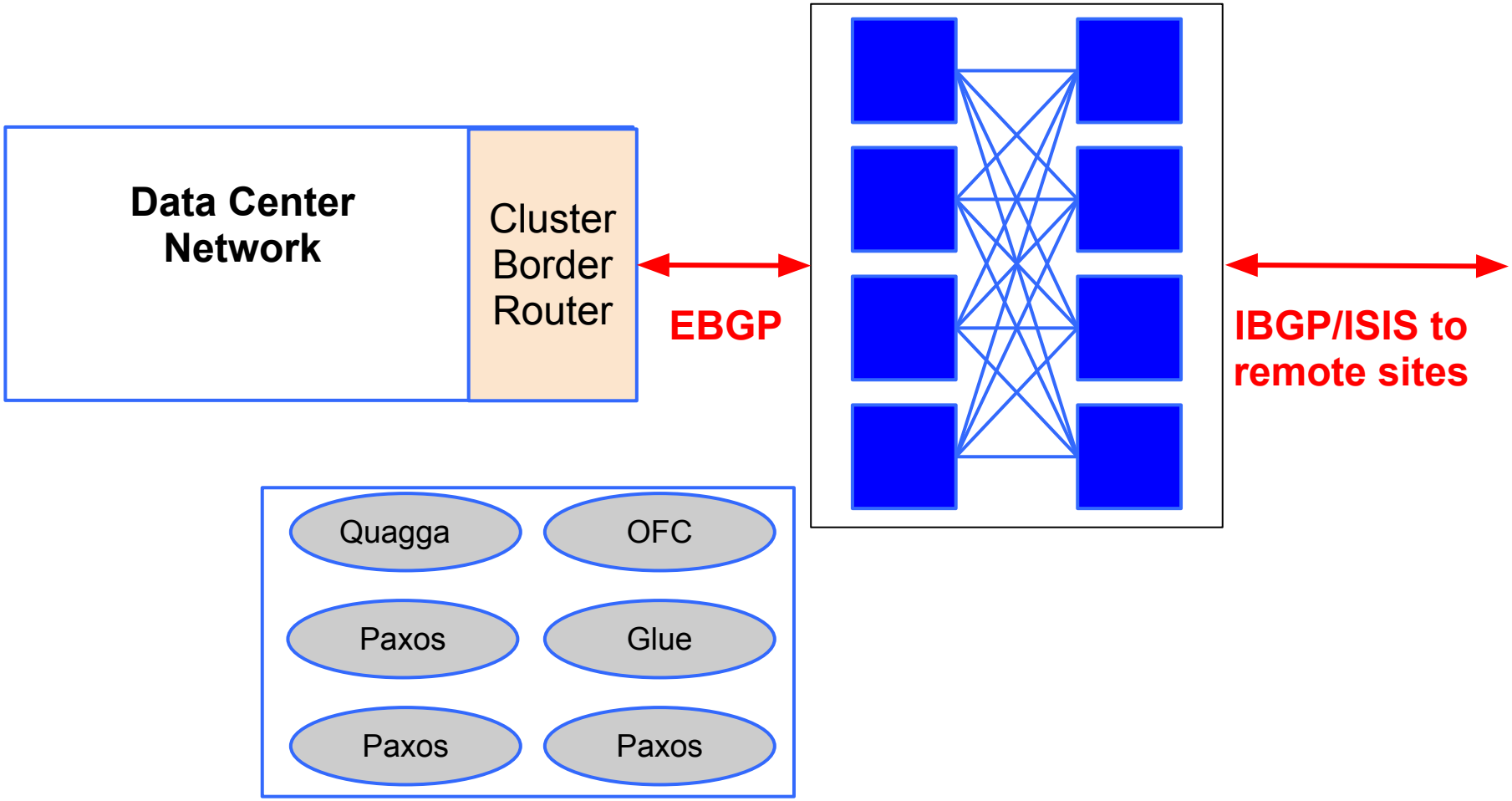
Google™



(not representative of actual topology)

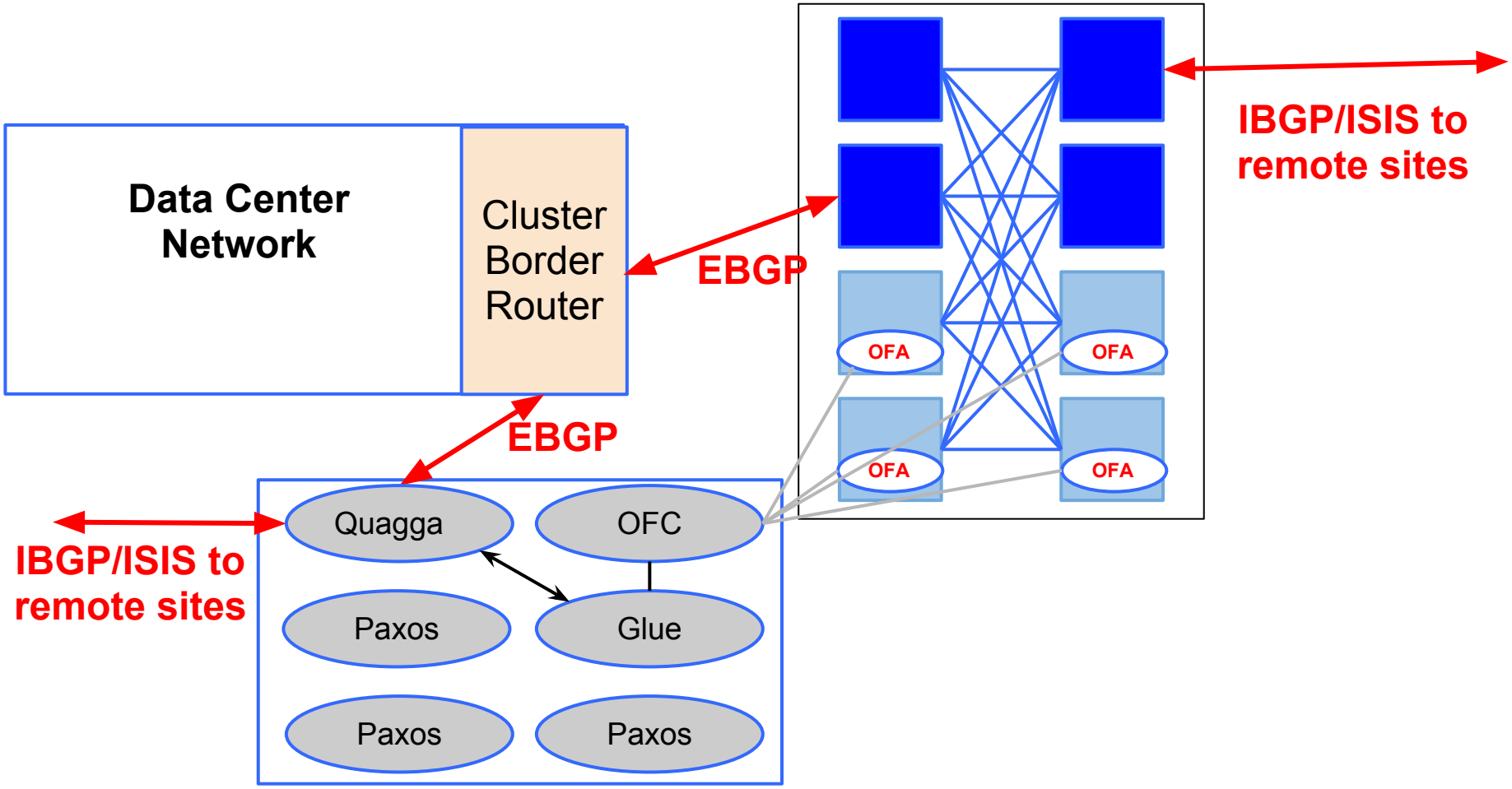
Mixed SDN Deployment

Google™



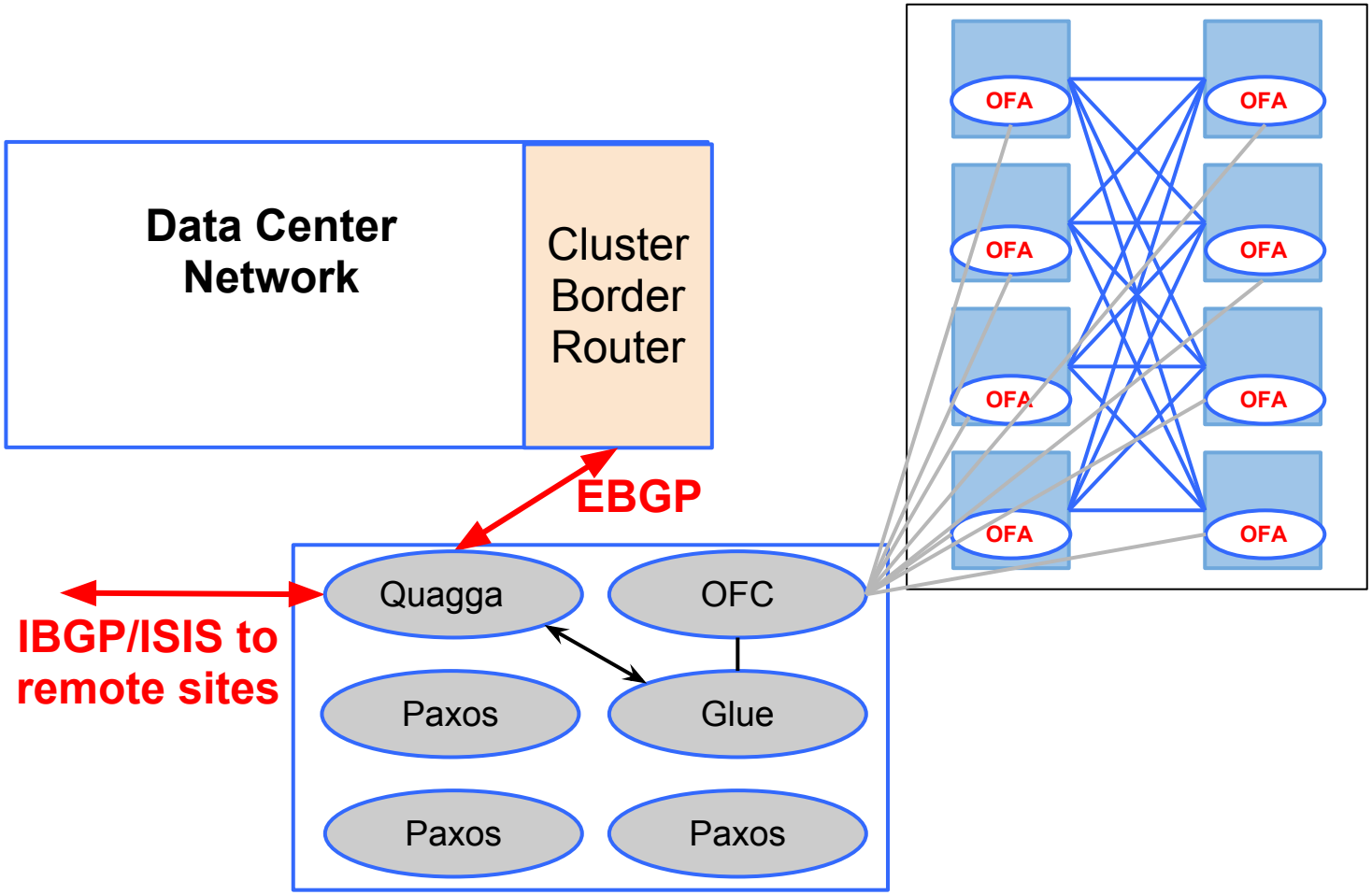
Mixed SDN Deployment

Google



Mixed SDN Deployment

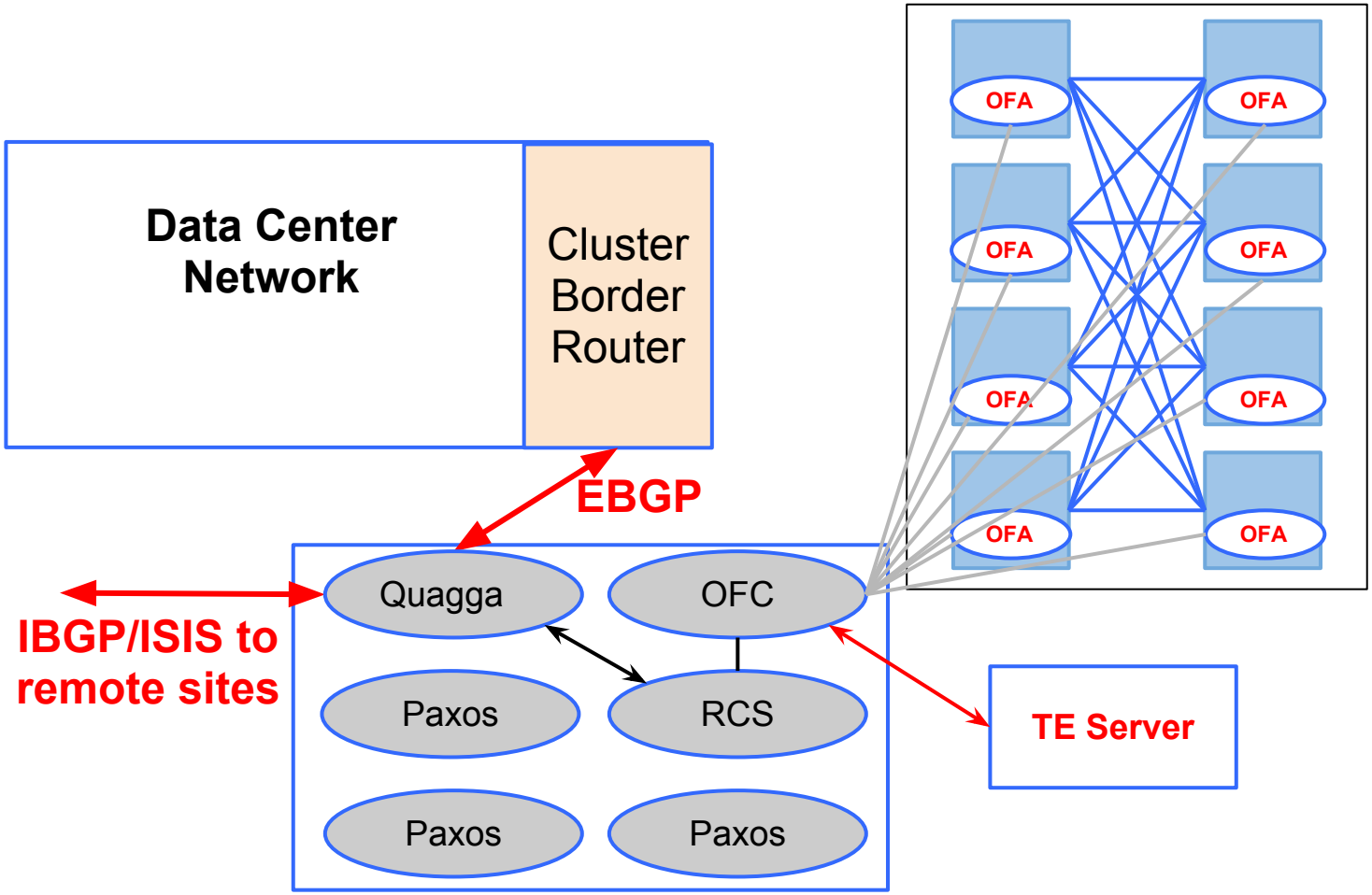
Google



- SDN site delivers full interoperability with legacy sites

Mixed SDN Deployment

Google

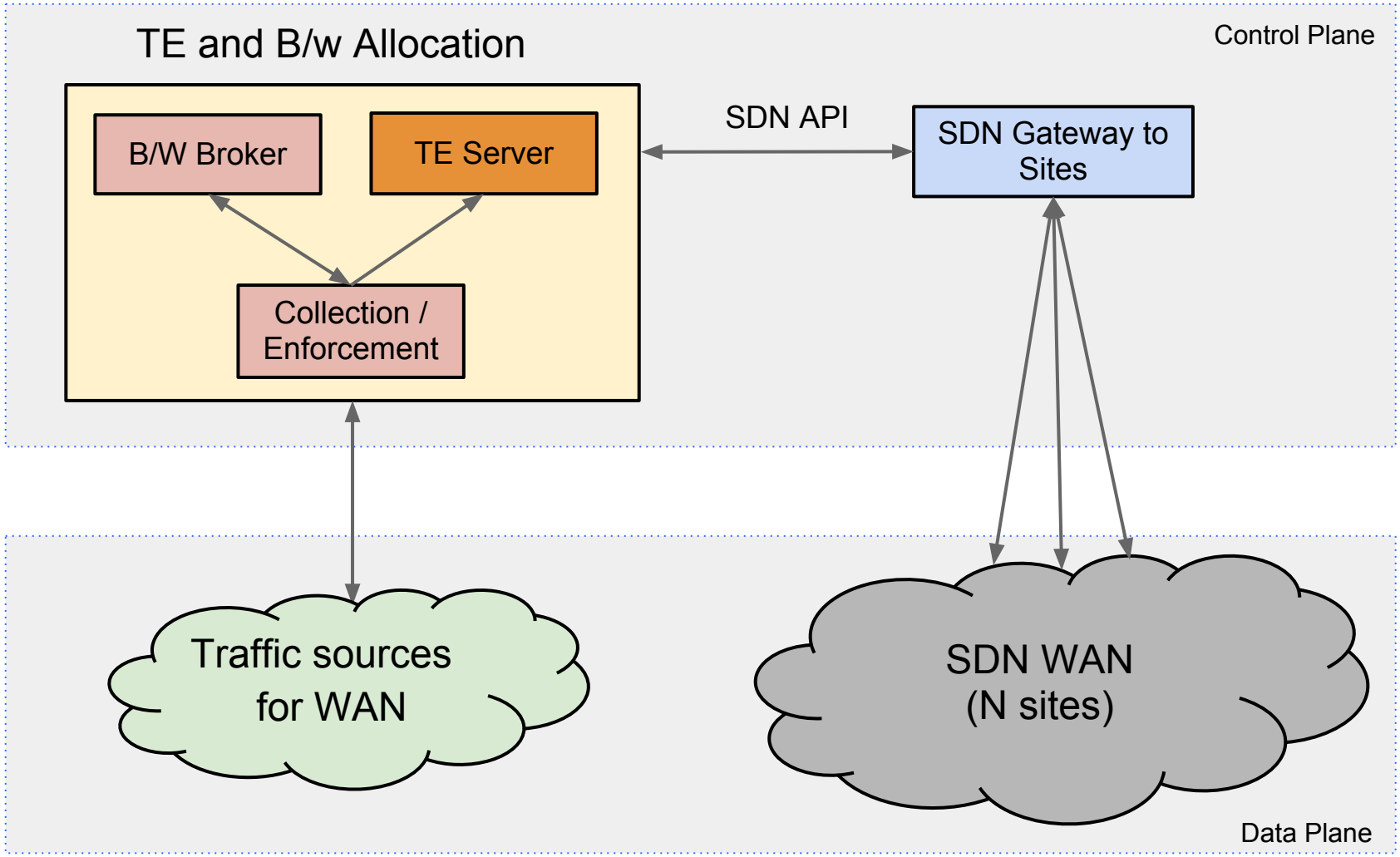


- Ready to introduce new functionality, e.g., TE

Bandwidth Broker and Traffic Engineering

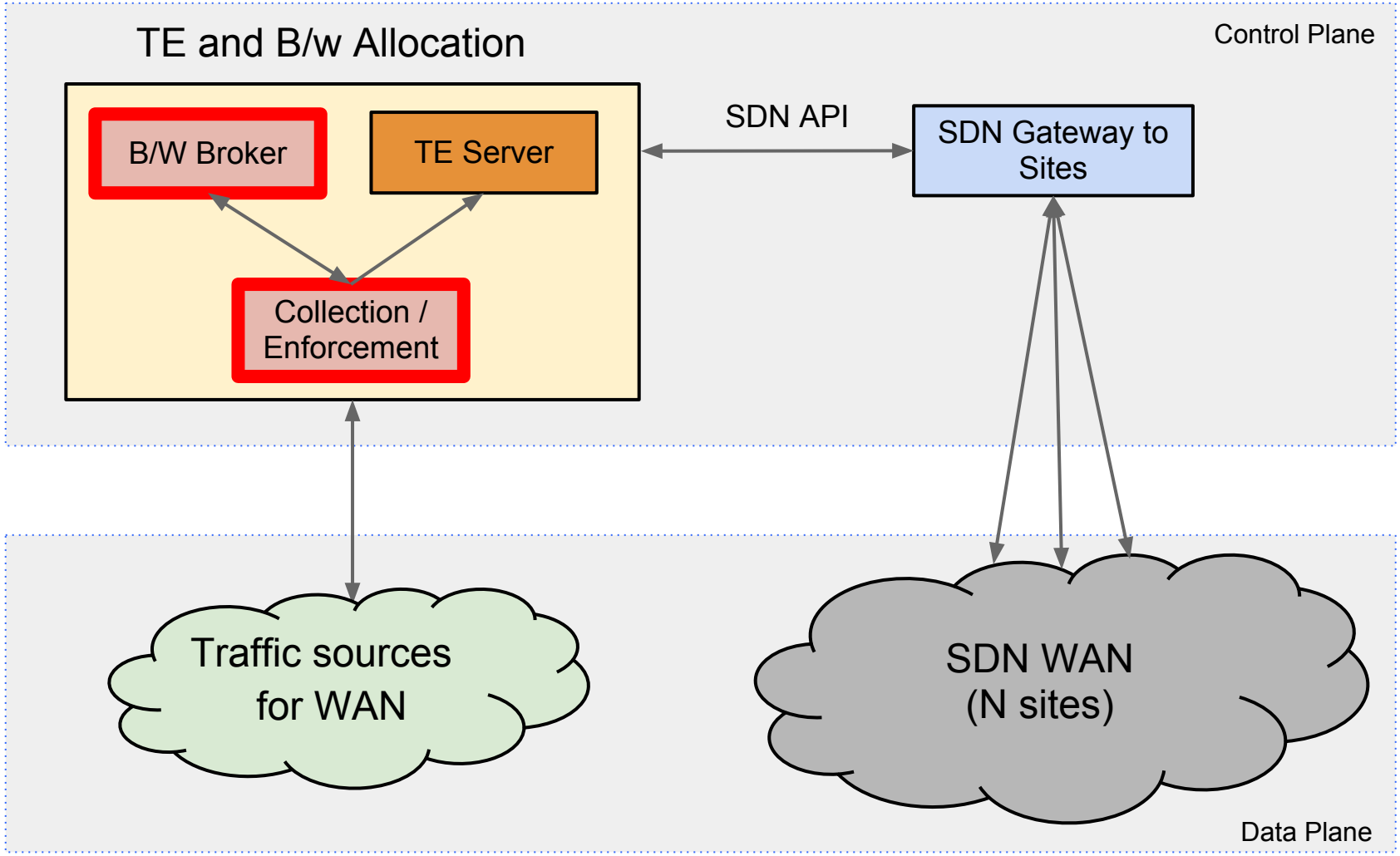
High Level Architecture

Google

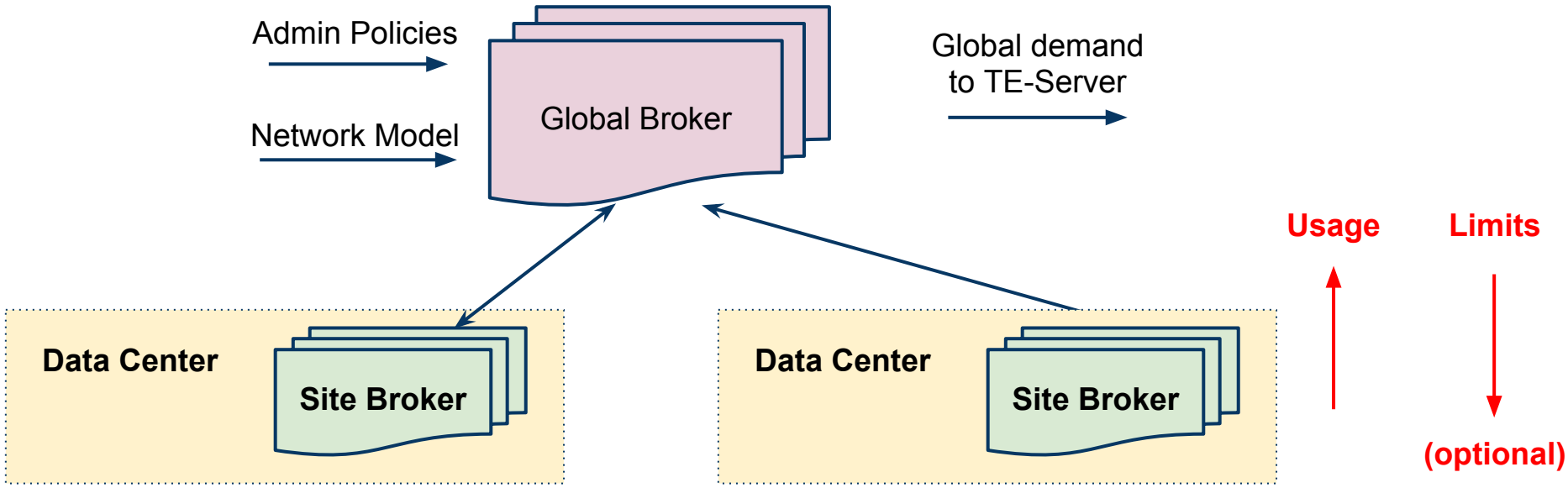


High Level Architecture

Google

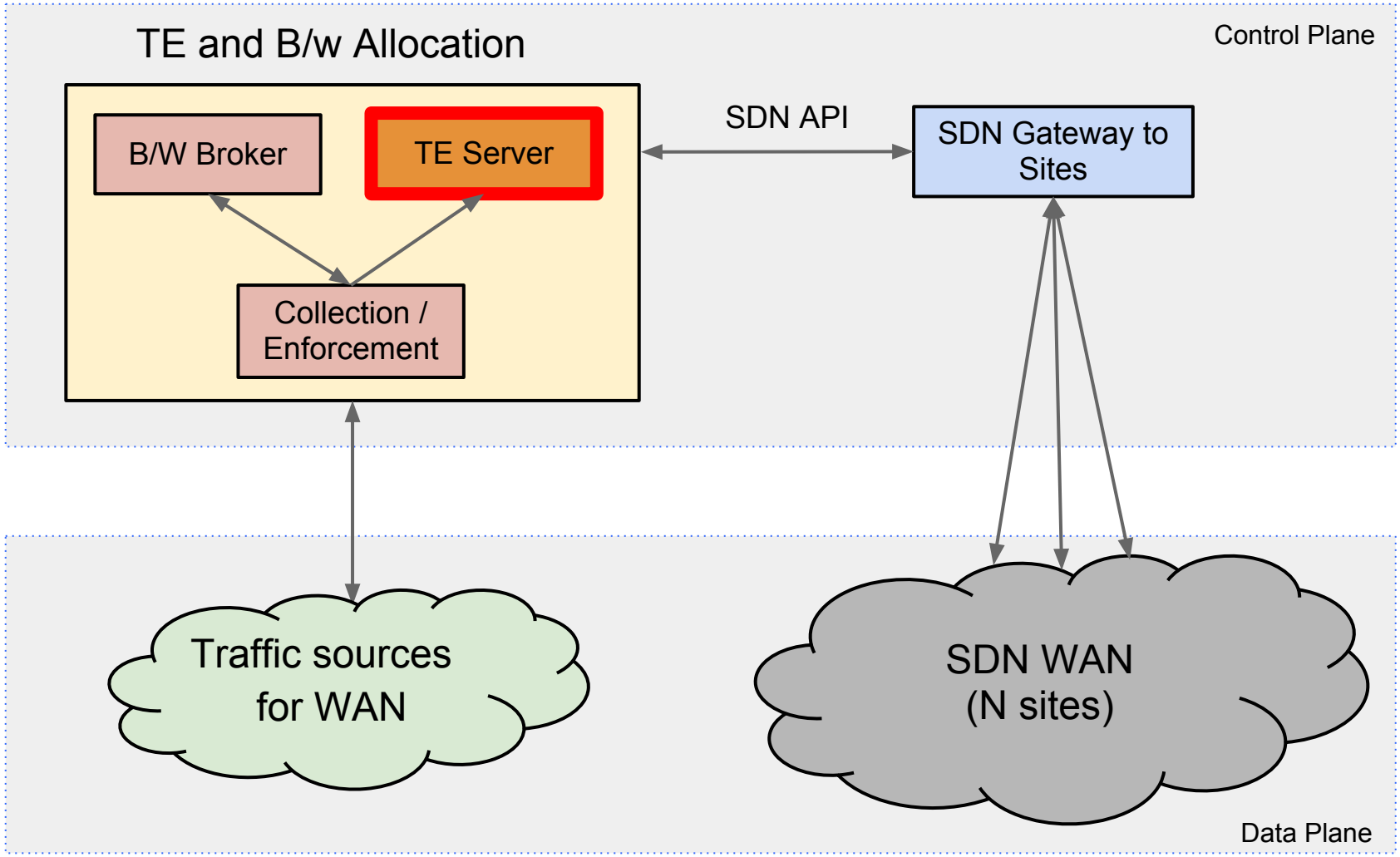


Bandwidth Broker Architecture

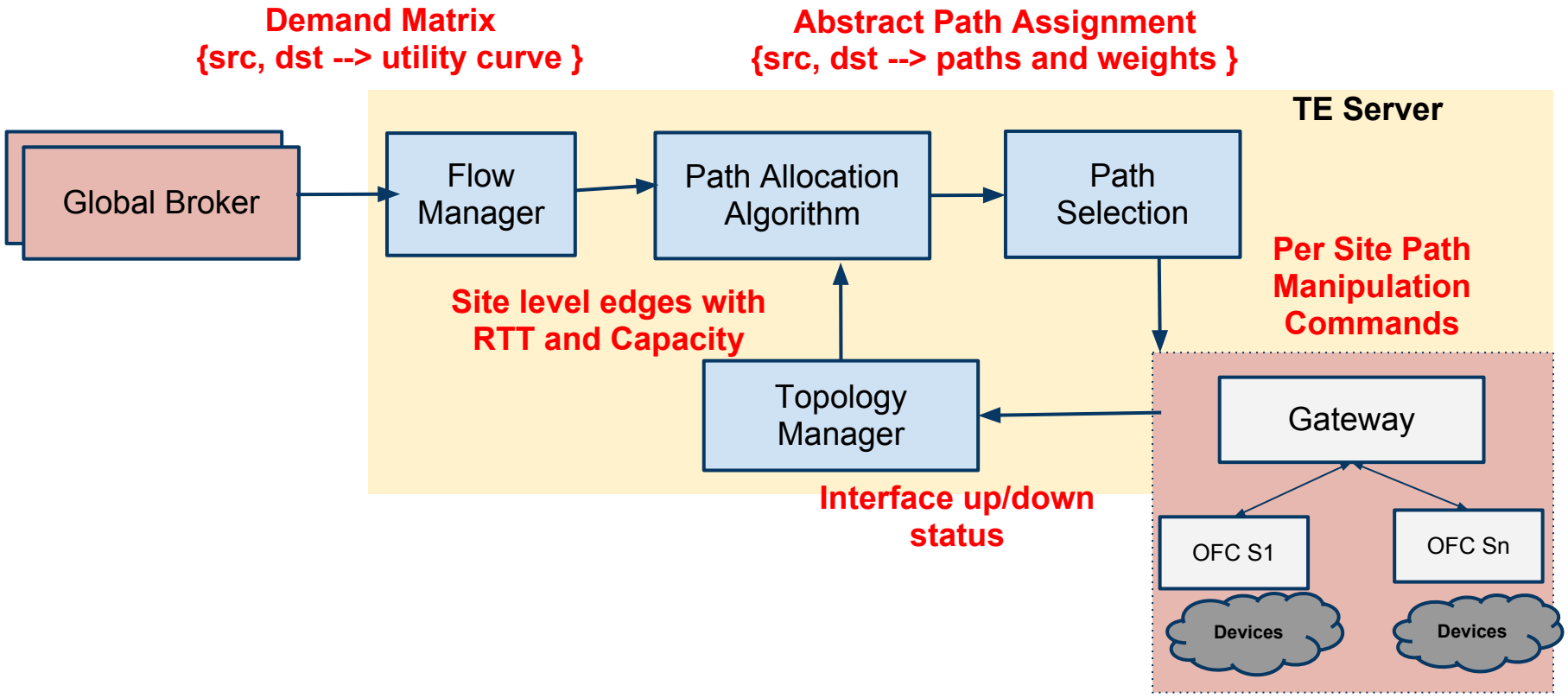


High Level Architecture

Google

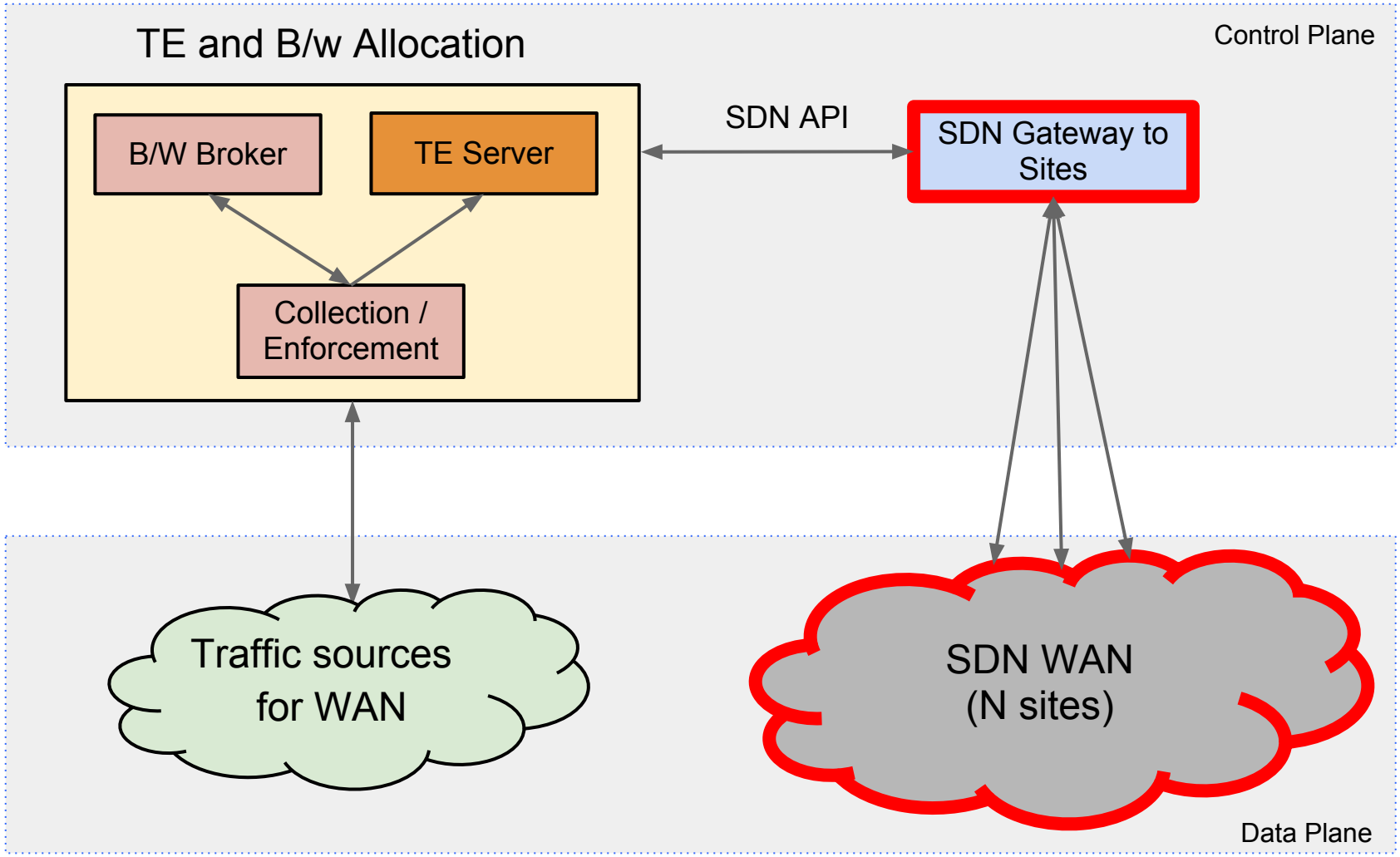


TE Server Architecture

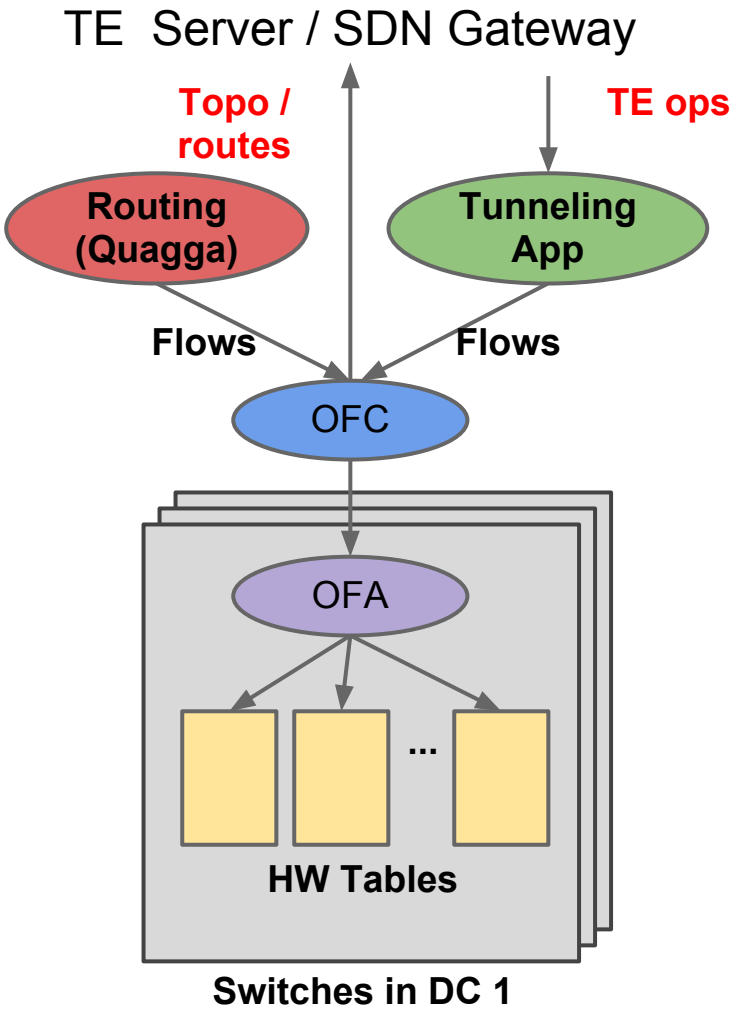


High Level Architecture

Google

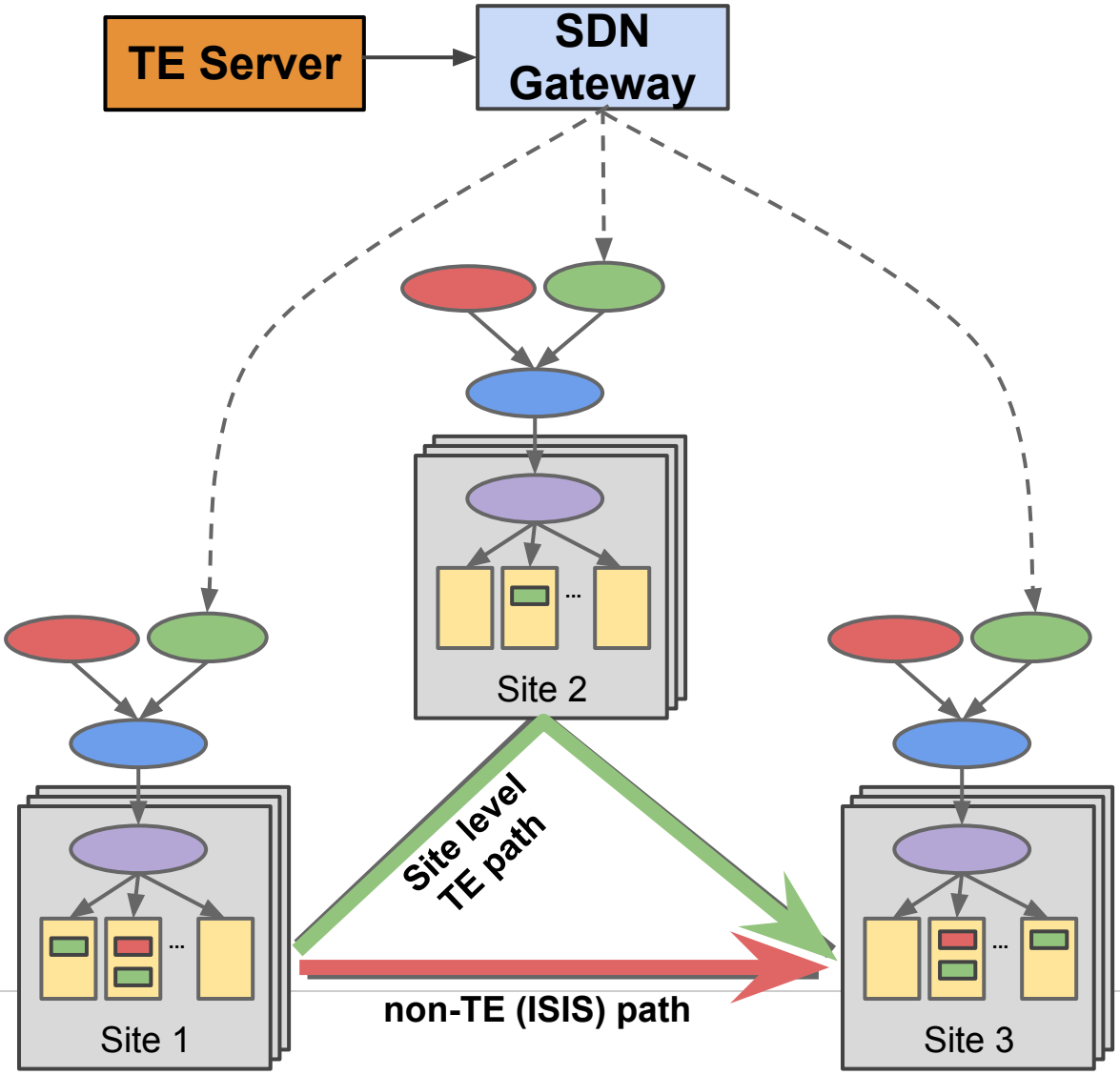


Controller Architecture

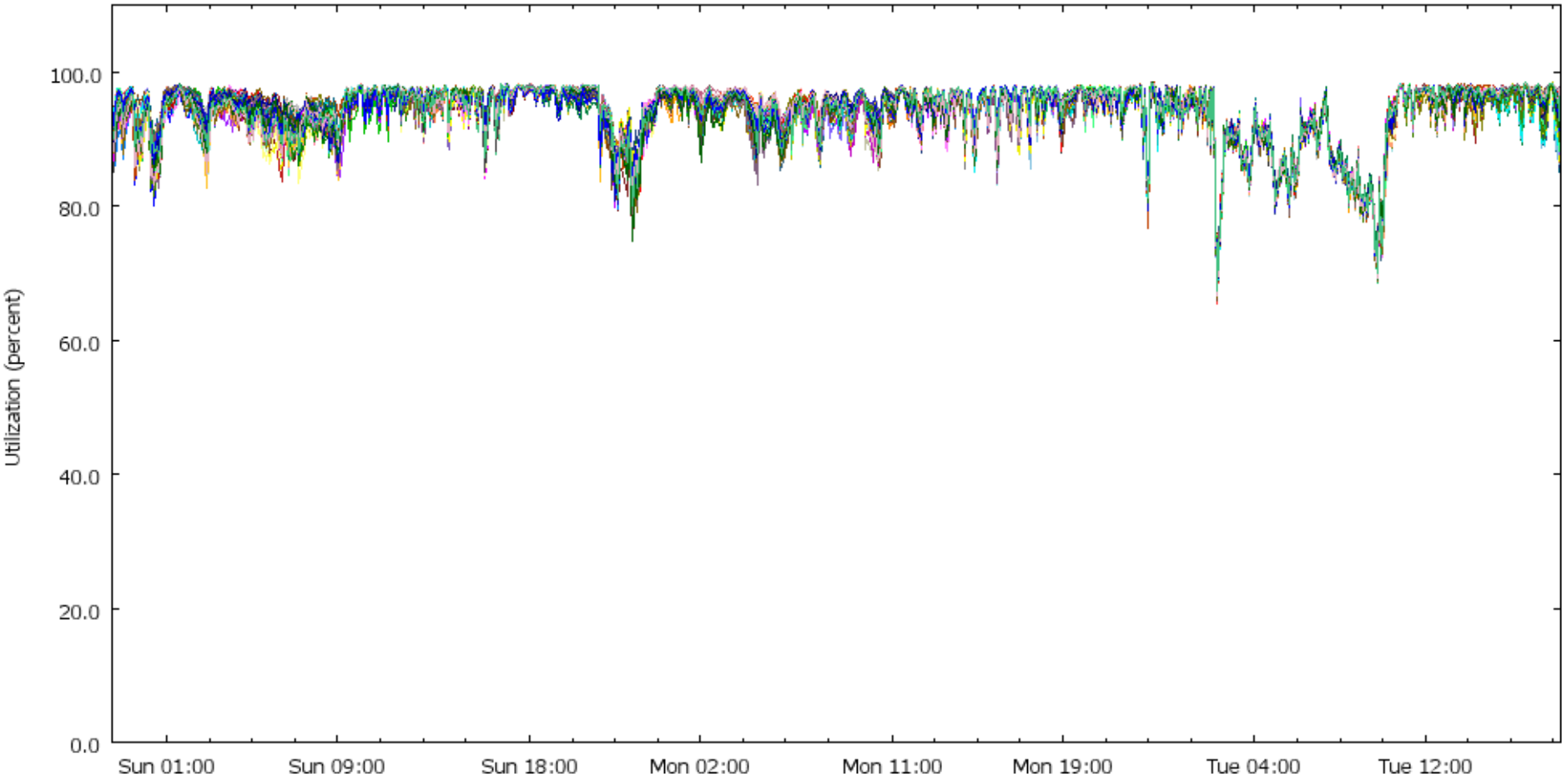


Controller Architecture

Google

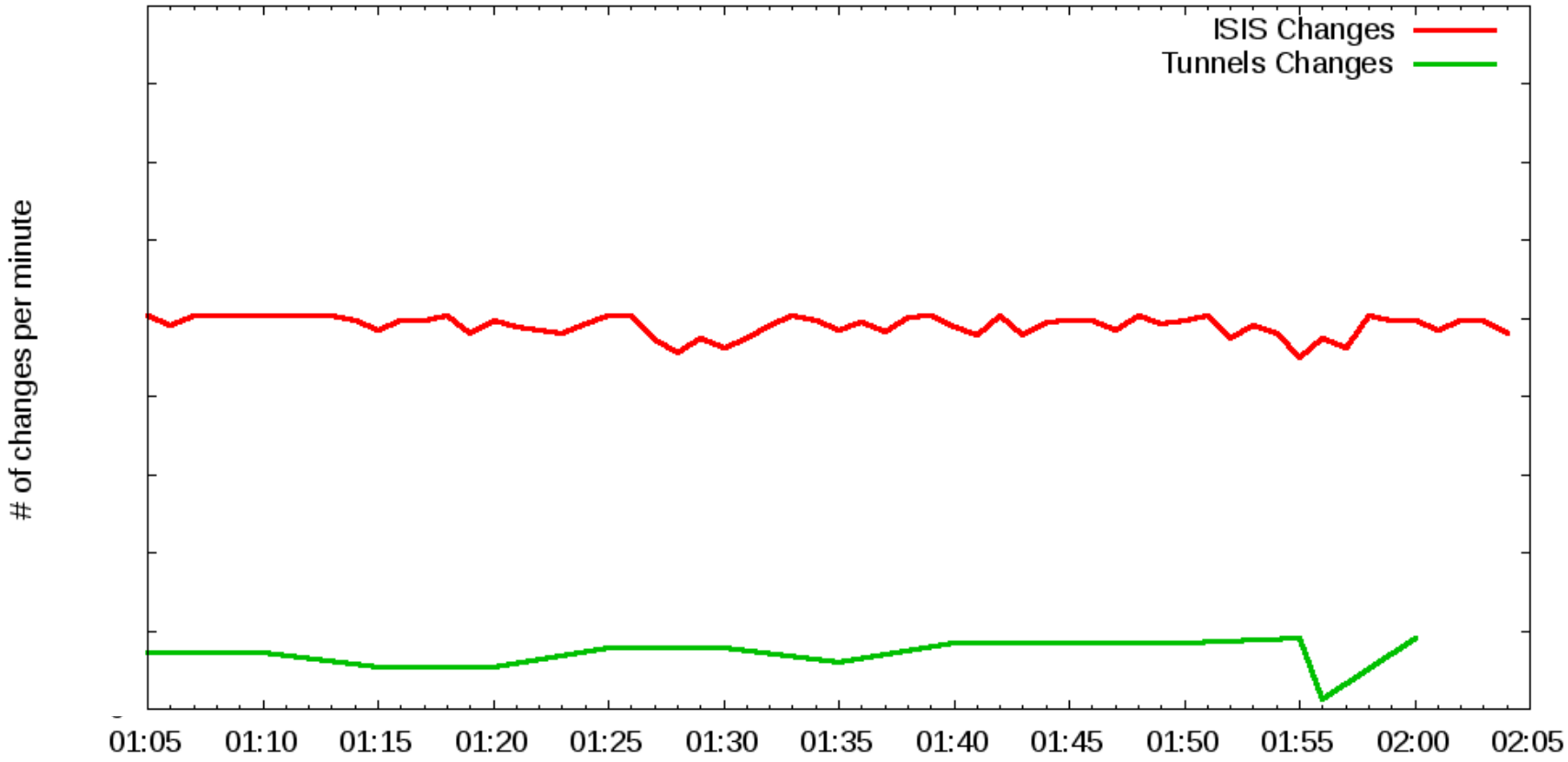


Sample Utilization



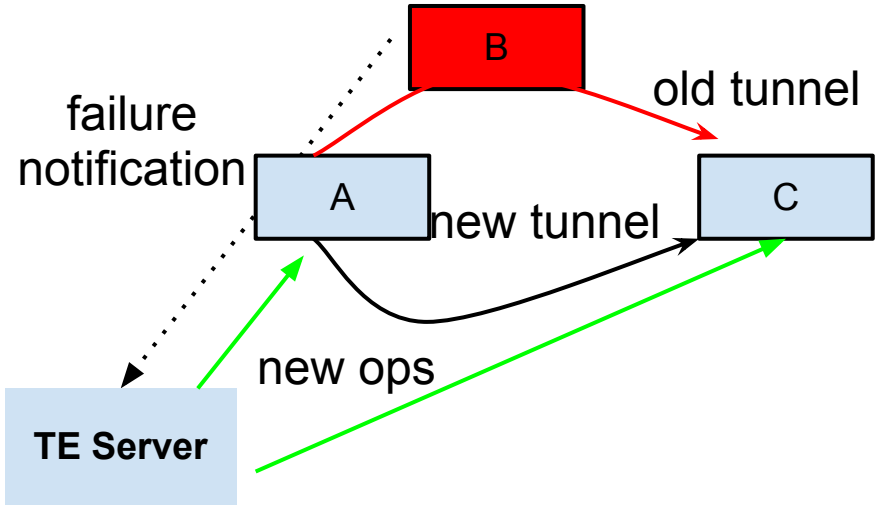
Benefits of Aggregation

Google



Convergence under Failures

Google



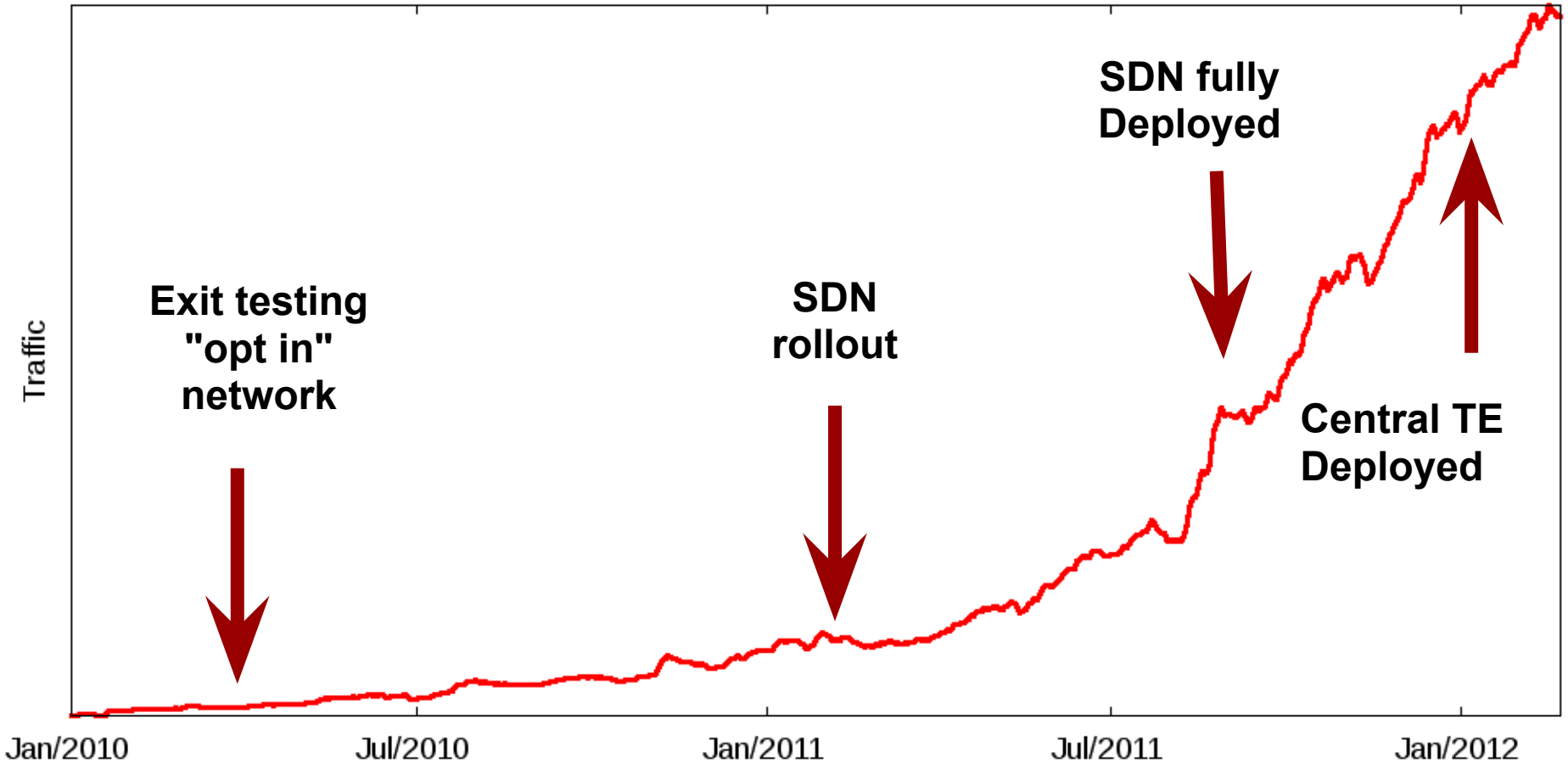
no-TE: traffic drop ~ 9 sec
with-TE: traffic drop ~ 1 sec

Without TE: Failure detection and convergence is slower:

- Delay 'inside' TE << timers for detecting and communicating failures (in ISIS)
- Fast failover may be milliseconds, but not guaranteed to be either accurate or "good"

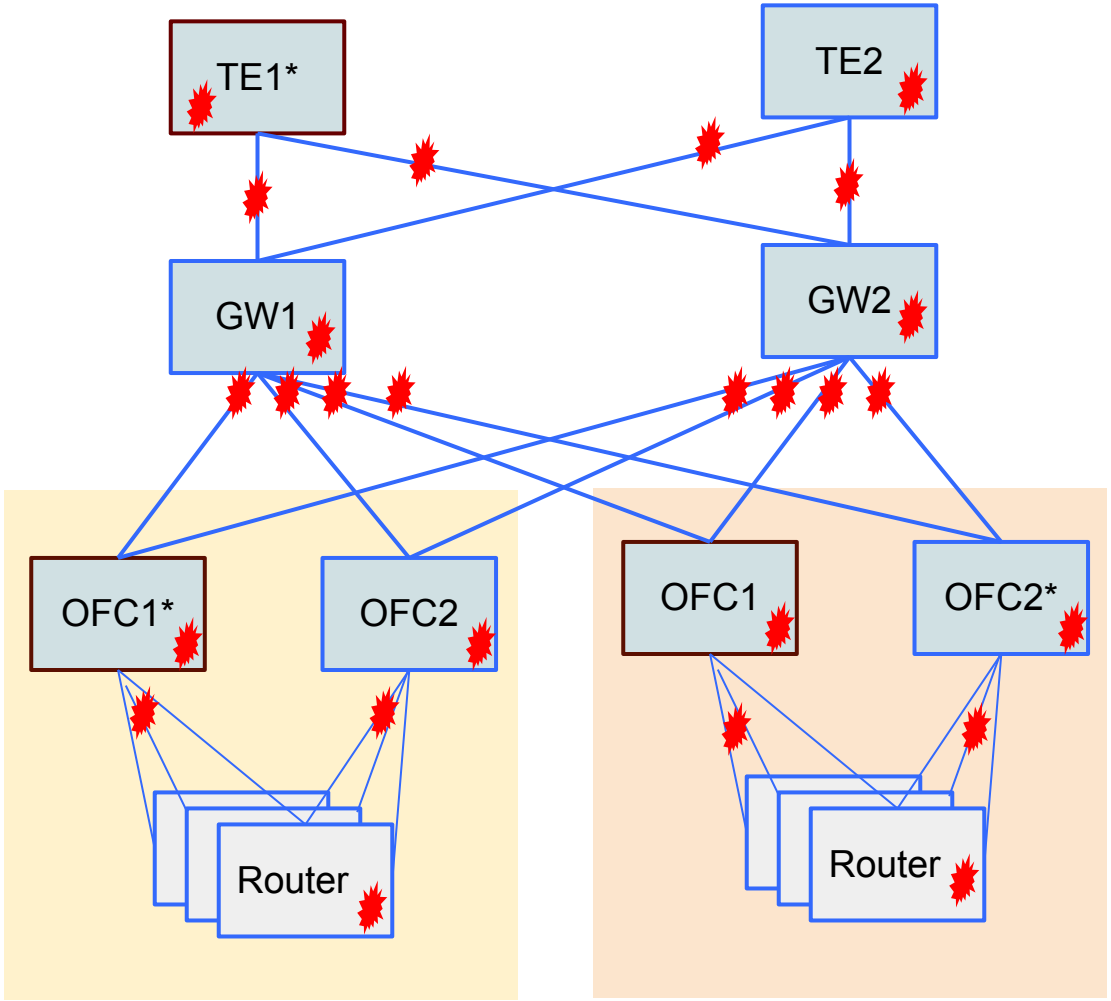
G-Scale WAN History

Google



Range of Failure Scenarios

Google



Potential failure condition

* indicates mastership

Trust but Verify: Consistency Checks

Google

TE View	OFC View	Is Valid	Comment
Clean	Clean	yes	Normal operation.
Clean	Dirty	no	OFC remains dirty forever
Clean	Missing	no	OFC will forever miss entry
Dirty	Dirty	yes	Both think Op failed
Dirty	Clean	yes	Op succeeded but response not yet received by TE
Dirty	Missing	yes	Op issued but not received by OFC
Missing	Clean	no	OFC has extra entry, and will remain like that
Missing	Dirty	no	(same as above)

Implications for ISPs

Google

- Dramatically reduce the cost of WAN deployment
  - Cheaper per bps in both CapEx and OpEx
  - Less overprovisioning for same SLAs
- Differentiator for end customers
  - Less cost for same BW or more BW for same cost
- Possible to deploy incrementally in pre-existing network
  - Leveraging known techniques for delivering any new functionality

Conclusions

Google

- Dramatic growth in WAN bandwidth requirements
  - Every 10x, something breaks
  - Existing software/hardware architectures make it impractical to deliver cheap bandwidth globally
- Software Defined Networking enables
  - Separation of hardware from software
  - Efficient logically centralized control/management
  - *Innovation and flexibility*
- Deployment experience with Google's global SDN production WAN
  - It's real and it works
  - This is just the beginning...

Thank you!