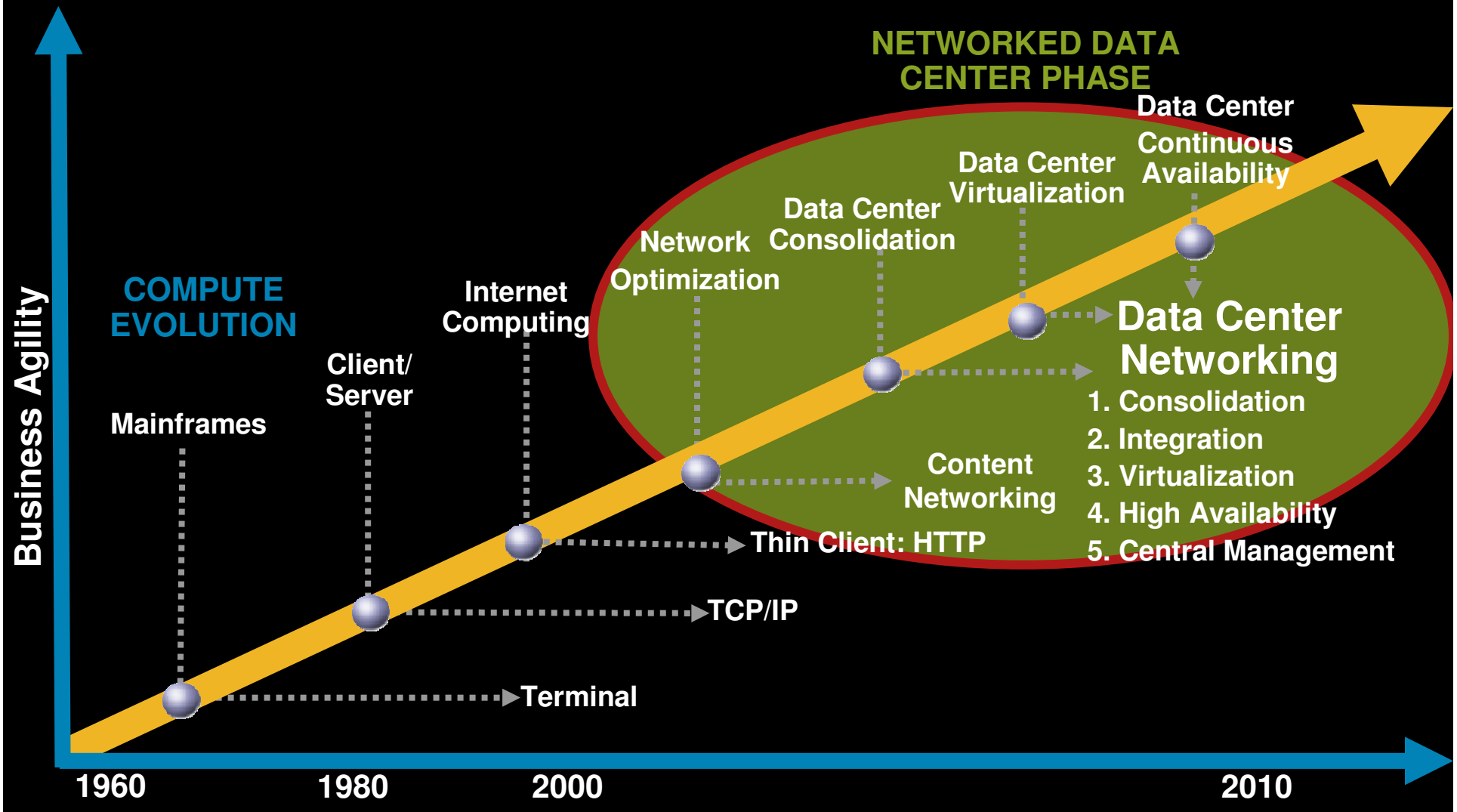
Cisco Data Centre Solutions

Business Continuance and Disaster Recovery

Višnja Milovanović, Systems Engineer

vmilovan@cisco.com

Data Center Evolution

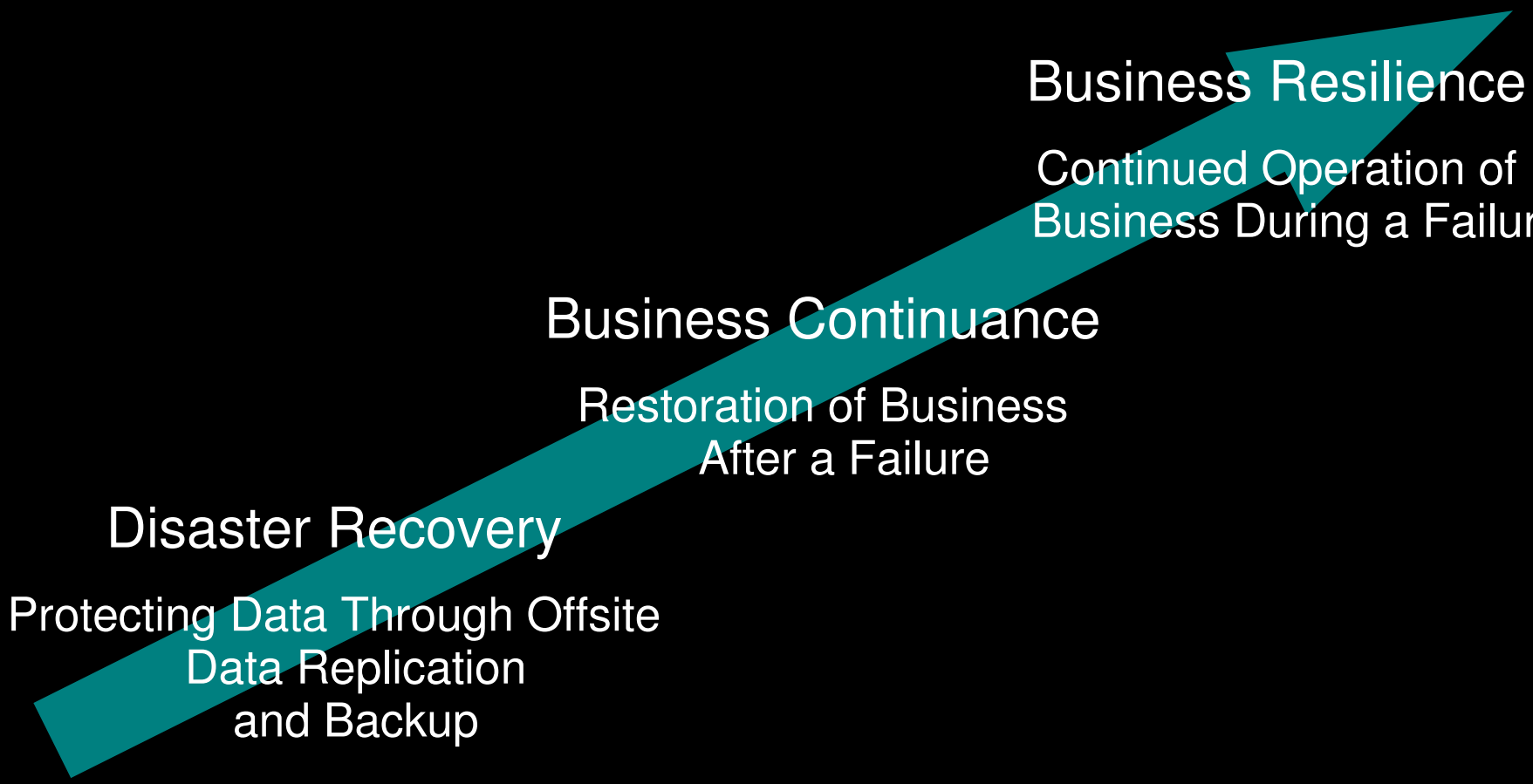


Agenda

- **Business Continuance and Disaster Recovery**
  - **Site selection techniques:**
    - **RHI Injection and IP SLA**
    - **DNS based Site Selection**
  - **Datacenter interconnect options**

Business Continuance and Disaster Recovery

What It Means for Business



Zero Down Time is the ultimate goal

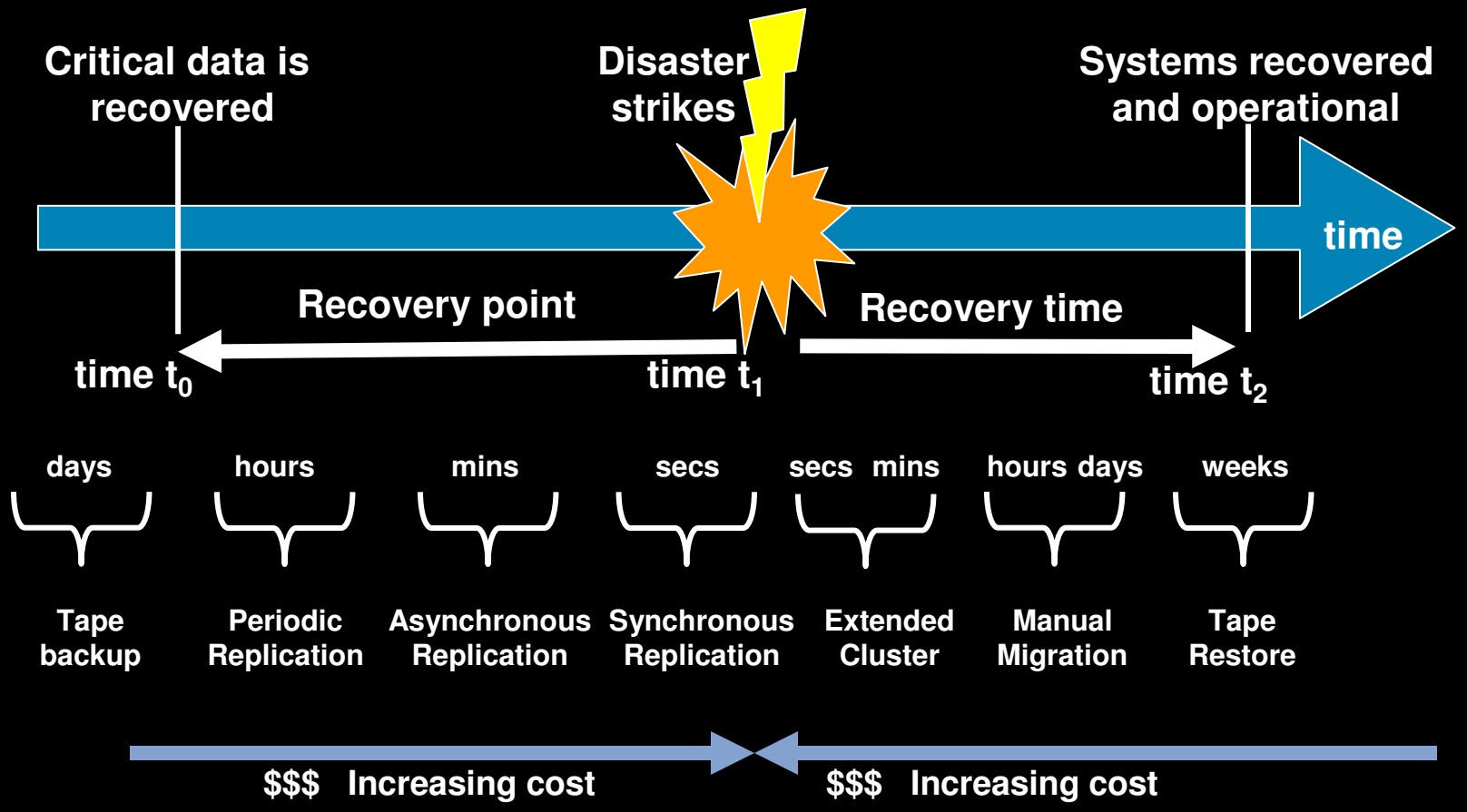
Why Resilient, Distributed Data Centers

- Required by disaster recovery, business continuance, and business resiliency
- Avoid single, concentrated data depository
- High availability of applications and data access
- Load balancing together with performance scalability
- Better response and optimal content routing: proximity to clients

Business continuance solutions

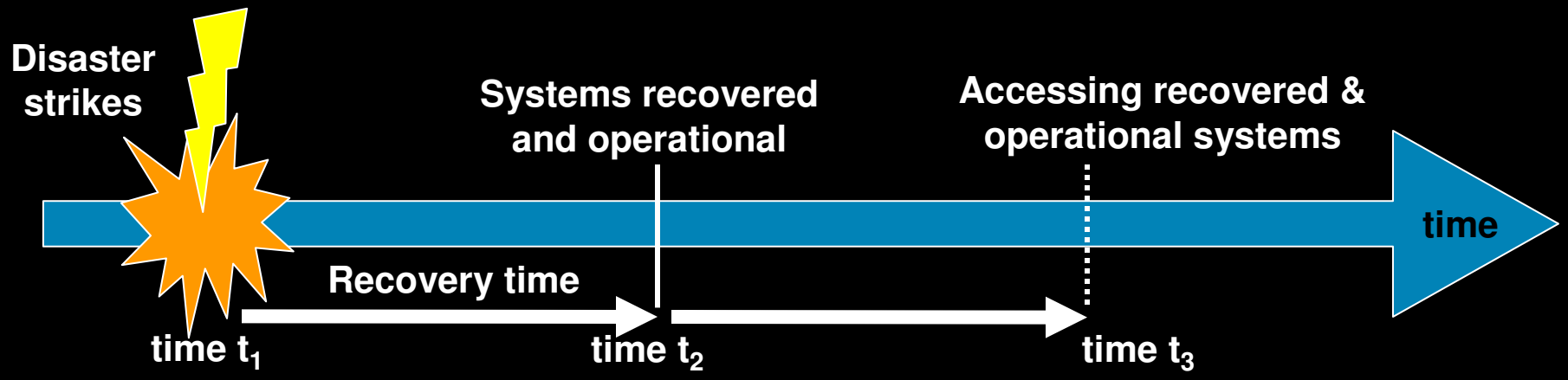
- Motivation for BC/DR solutions
  - Protect operation of enterprise or organization during unexpected event (‘disaster’)
  - Sometimes enforced by law or other regulation (Basel II etc.)
- What have to be done to create BC/DR policy
  - Identification of Critical Applications
  - Distance between disaster and recovery zone
  - Mode of operation (active-active, active-stanby)
  - Tolerable Application down time
- What parameters should BC/DR policy have
  - Must be measurable
  - RTO, RPO, RAO

Disaster recovery parameters:
Recovery Time Objective and Recovery Point Objective



- How current or fresh is the data after recovery?
- How quickly can systems and data be recovered?

Disaster recovery parameters:
Recovery Access Objective (RAO)



(t_2)→ Recovery Time Objective
($t_3 – t_2$)→ Recovery Access Objective

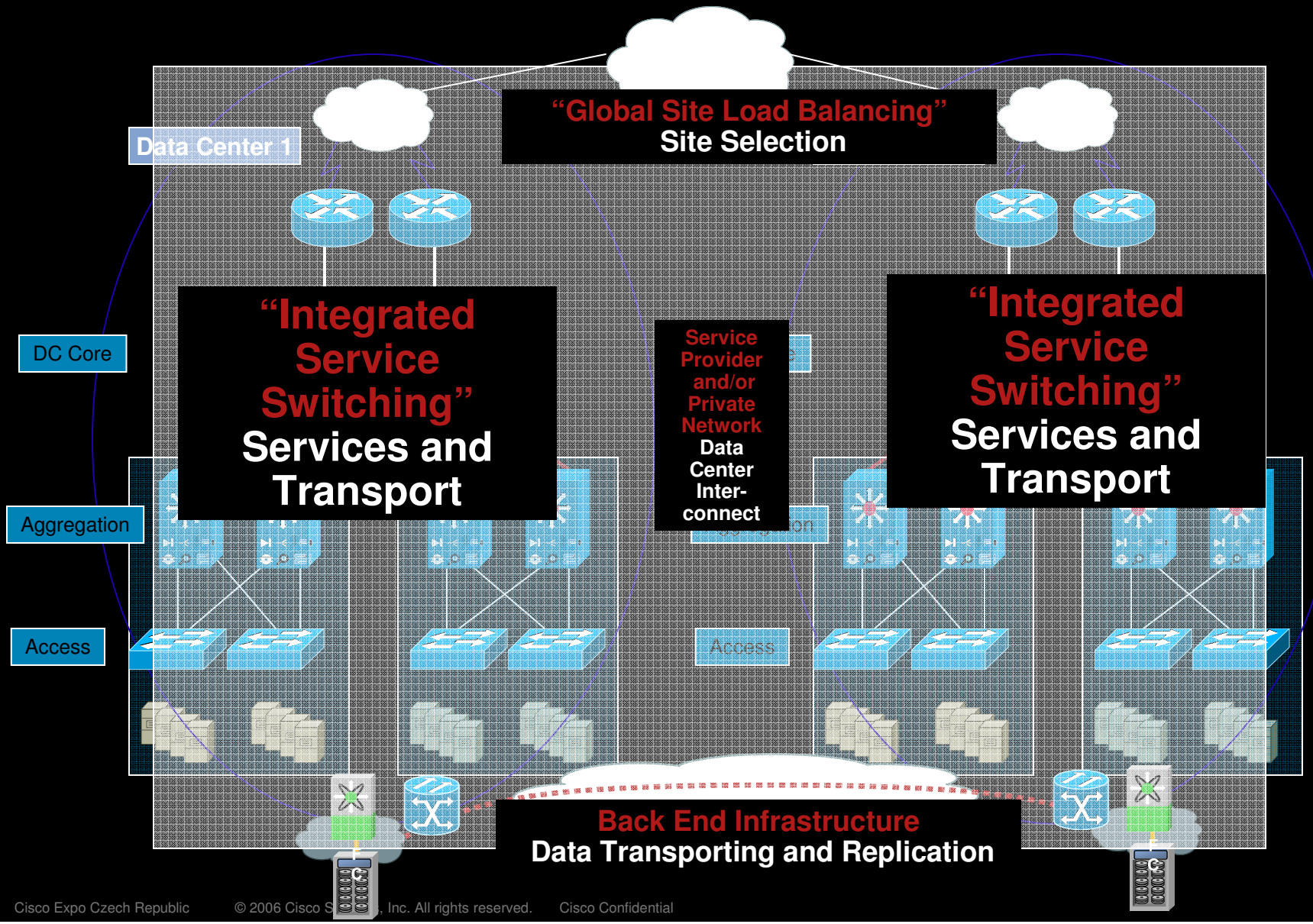
Time taken by network to converge and provide a path for clients to access the applications and data
Note: RAO can be +ve or –ve w.r.t. Recovery time (t2)

Disaster Recovery Site Considerations

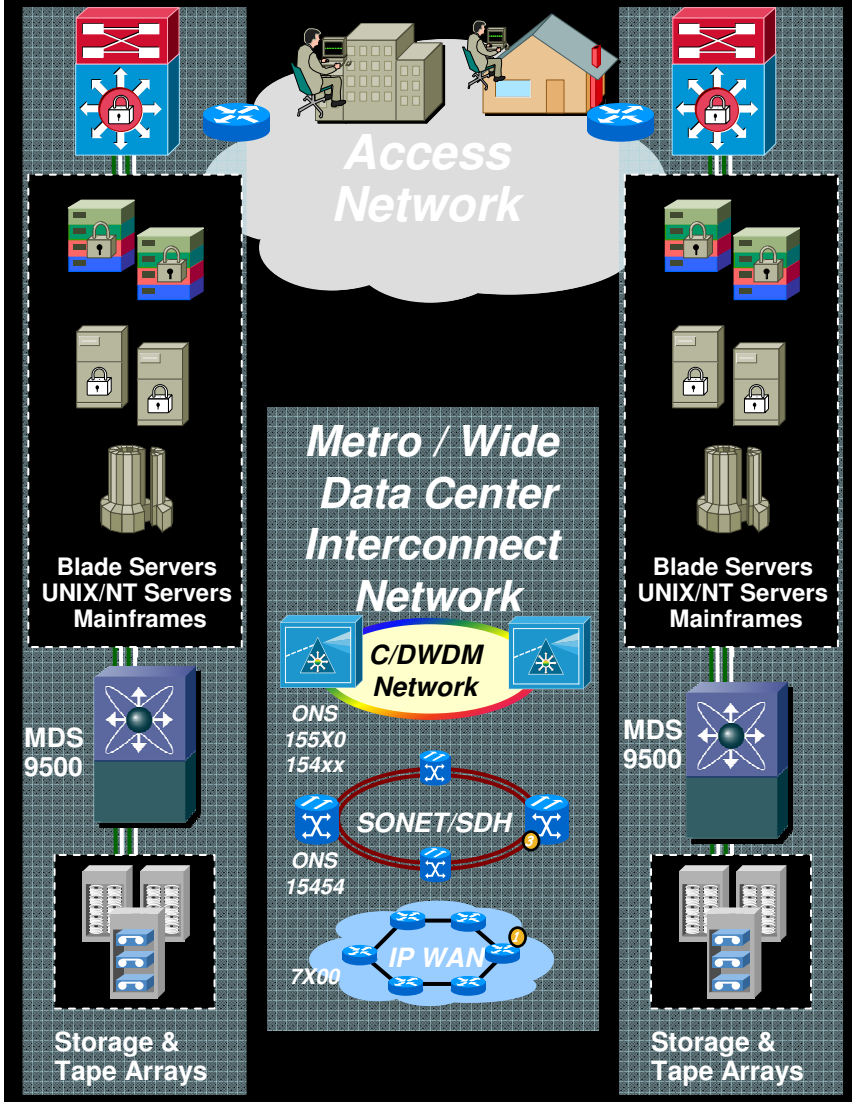
	Attributes	Recovery Time	Cost	Data Replication	RTO RPO
Cold / Cool Standby Site	▪Manual Failover ▪Pre-Configured devices 'on' or 'off'	24 hours to 5 days		▪Tape, optical media, via point-to-point	★
Warm Standby Site	▪Manual Failover ▪Can function as testing data center until needed for disaster recovery	30 min to 8 hours		▪High Speed Connection ▪ Periodic Replication	★★
Hot Standby Site (Active / Standby)	▪Fully operational ▪Little or no human intervention	1 min to 20 min		▪Asynchronous data replication. ▪Active data at only one site	★ ★★
Hot Standby Site (Active / Active - Hybrid)	▪ One or more data centers active at the same time. ▪ Per Application basis	Business Continuance		▪ Synchronous Real time replication ▪DB Locking mechanisms	★ ★★ ★★

Source for "Recovery Time": ANSI TIA-942 BICSI standard

Data Center Resiliency Components – Architecting, Resilient Distributed Data Centers



Business Continuance / Disaster Recovery
Logical solution components



Front end: Site selection
- pointing users to operational site
- *DNS based solutions*
- *solutions based on routing protocols (RHI)*
- *HTTP redirection*

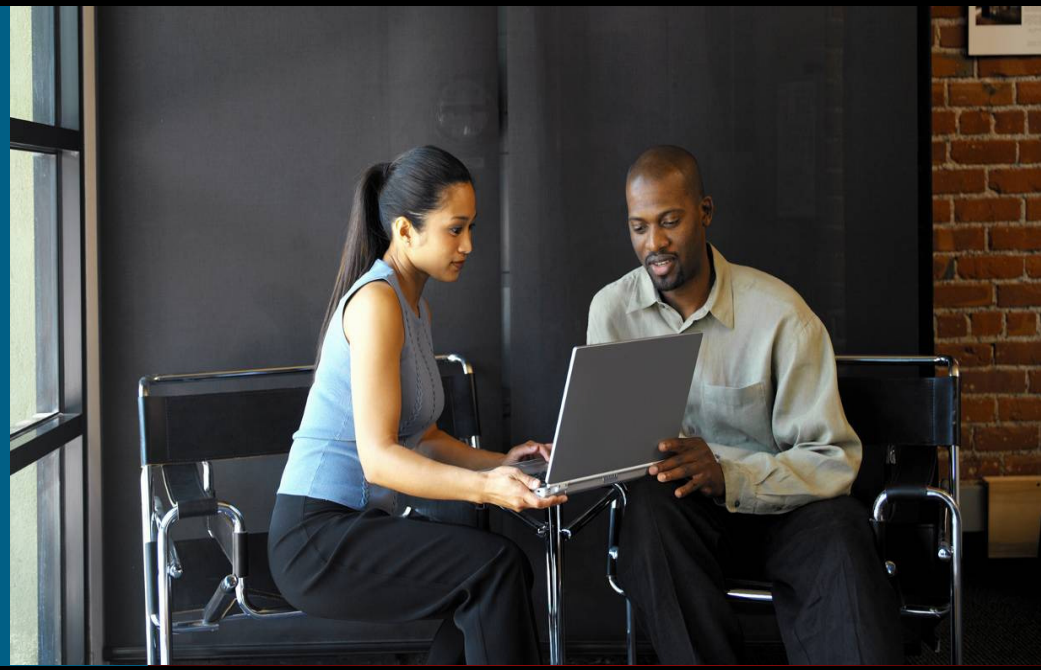
Application: Content switching
- selecting the appropriate server to perform requested operation
- load balancing
- load and health monitoring

Back end: Data replication and inter datacenter transport
- ensuring data availability in case of disaster or failure
- storage solutions for array based mirroring
- optical solutions (DWDM, CWDM, SDH)
- FCIP

Business Continuance and Disaster Recovery

Site selection techniques

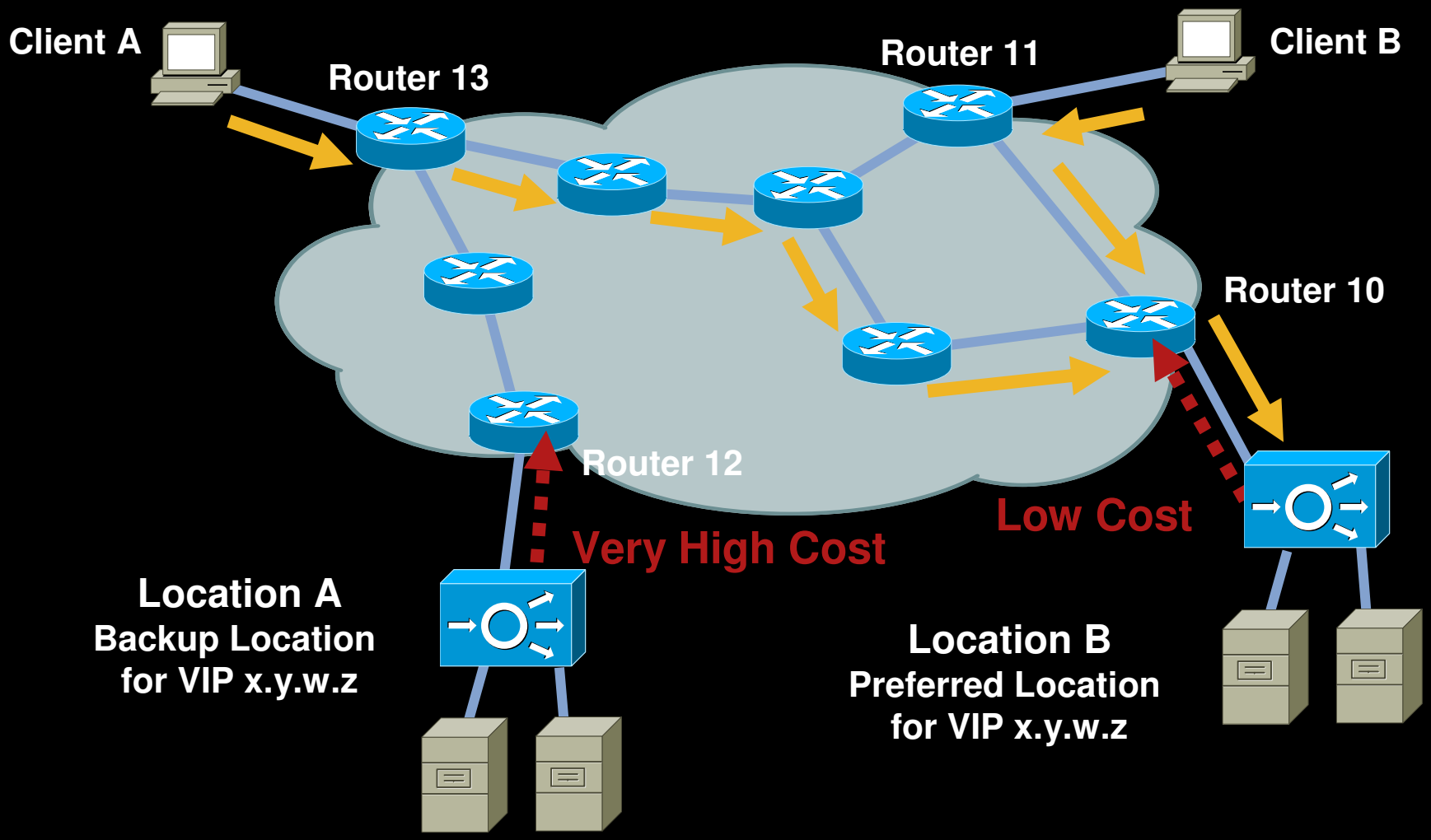
Route Health Injection



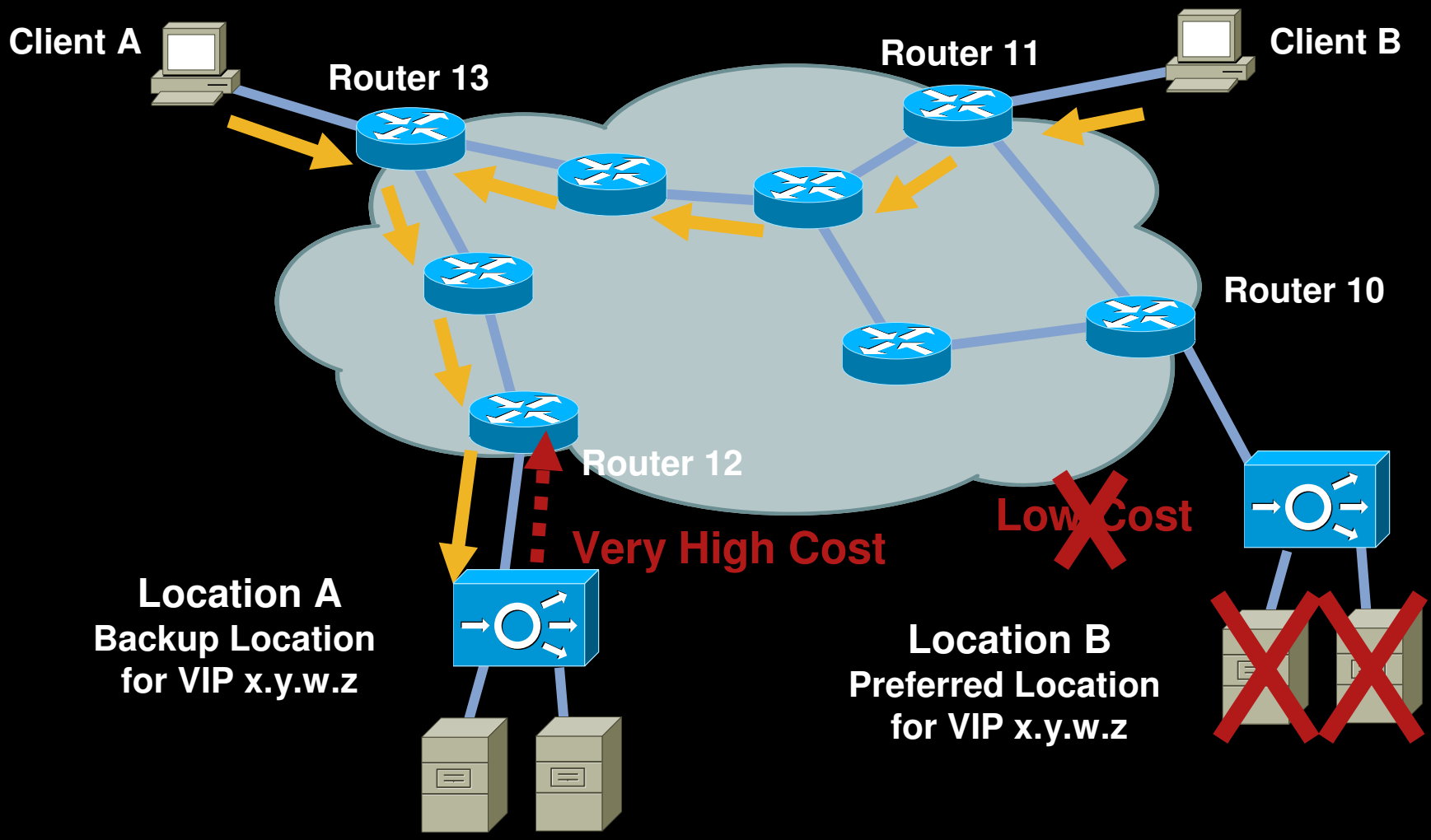
RHI: The Idea

- Server and application health monitoring provided by local Server Load Balancers
- SLB can advertise or withdraw VIP address to upstream routing devices depending on the availability of the local server farm
- Same VIP addresses can be advertised from multiple Data Centers
- Relying on L3 routing protocols for route propagating and content request routing
- Disaster Recovery provided by network convergence

RHI: Implementation

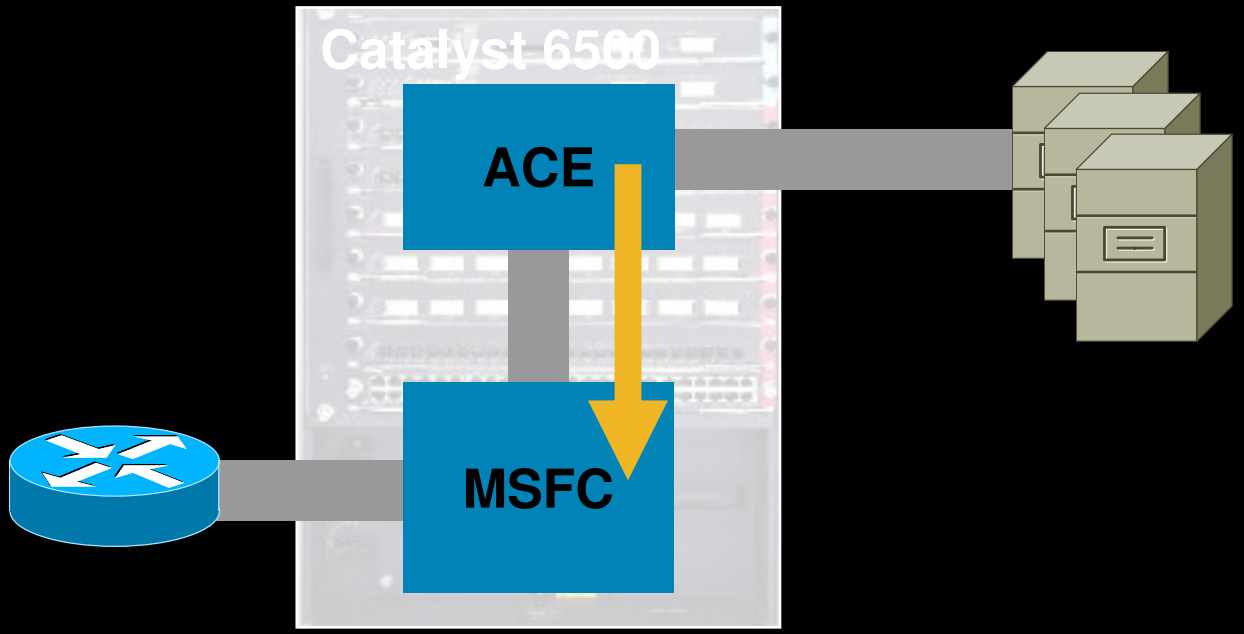


RHI: Implementation (Cont.)

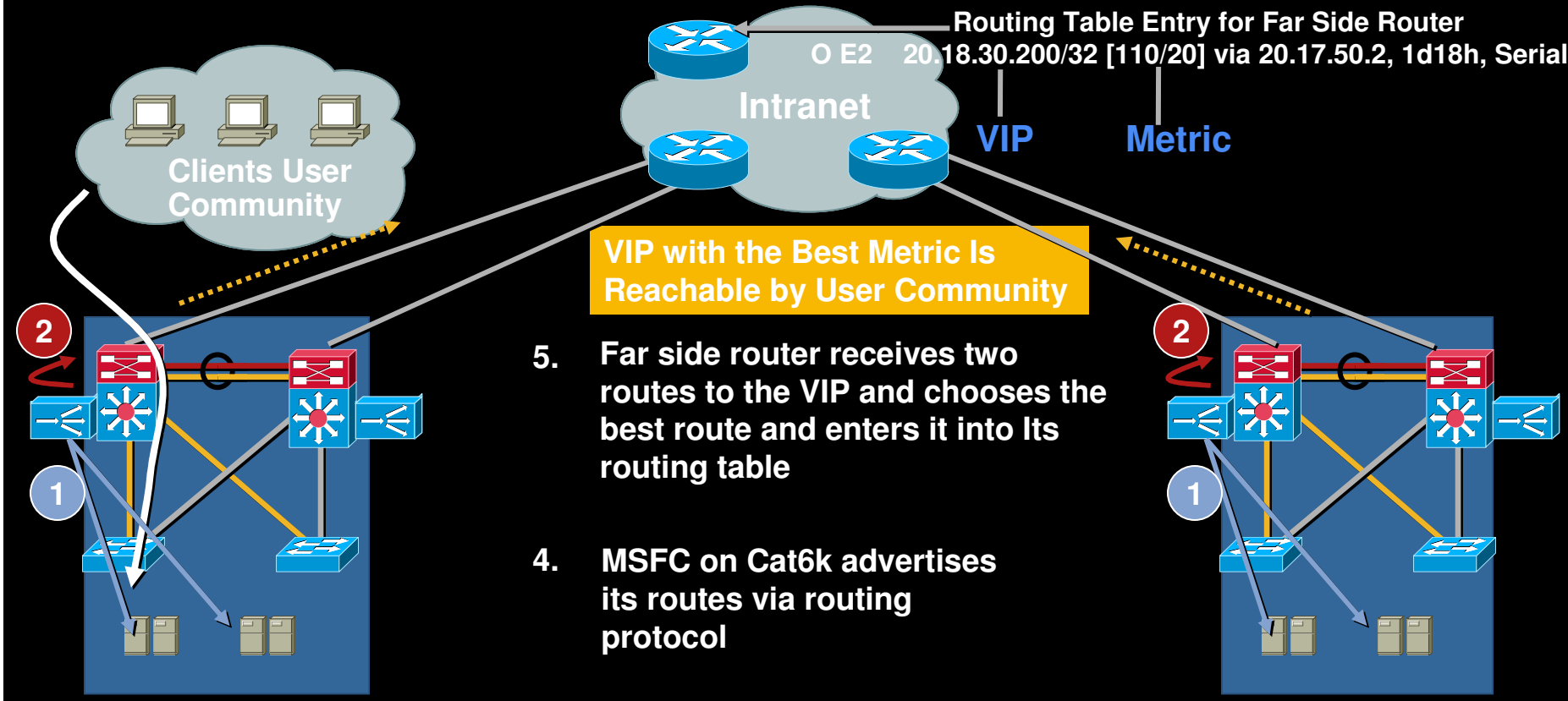


RHI: ACE/CSM + MSFC

- Application Control Engine (ACE) and Content Switching Module (CSM) can be configured to “inject” a 32-bit host route as a static route in the MSFC routing table
- The ACE or CSM injects or remove the route based on the health of the load balanced servers (checked with L3-7 probes or inband health monitoring)



Site selection
Server aware routing - RHI solution



1. ACE Probes Server Farm
2. ACE Sends an Advertise Message to MSFC on Cat6k if at least one server is active

3. MSFC on Cat6k determines the VLAN ID and adds the VIP and the VLAN ID to its routing table for available VIPs

4. MSFC on Cat6k advertises its routes via routing protocol

5. Far side router receives two routes to the VIP and chooses the best route and enters it into Its routing table

1. ACE Probes Server Farm
2. ACE sends an advertise message to MSFC on cat6k if at least one server is active

Advantages of the RHI Approach

- Supports legacy application and does not rely on a DNS infrastructure
- Very good reconvergence time, especially in Intranets where L3 protocols can be fine tuned appropriately
- Protocol-independent: works with any application
- Robust protocols and proven features

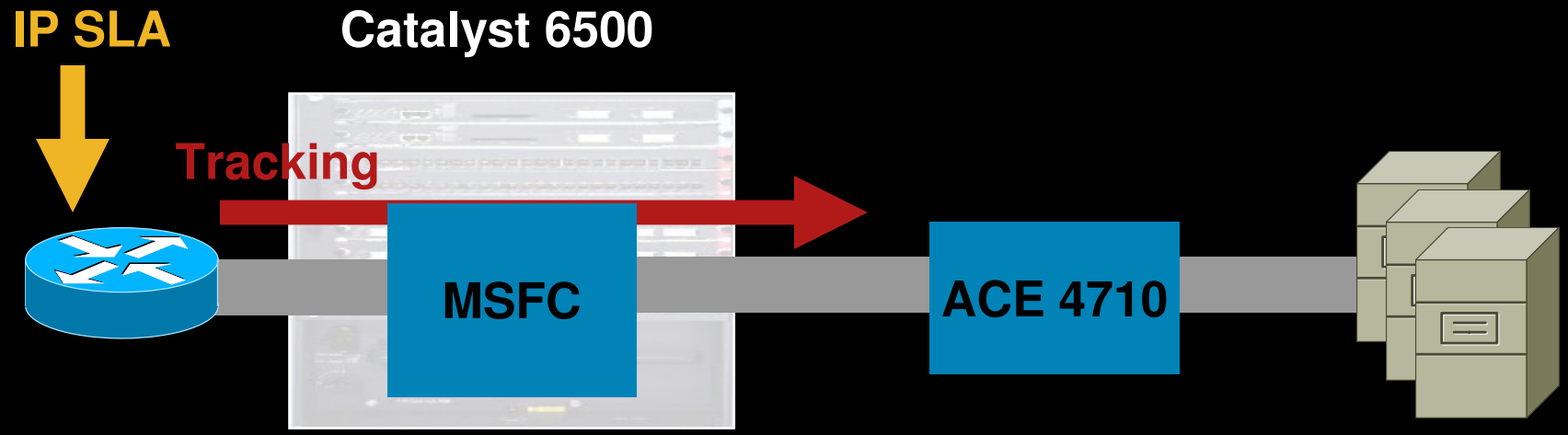
IP SLA

IP SLA: The Idea

- Upstream router of the Load Balancer can install a static route to the VIP
- Health of the VIP can be monitored via ICMP, TCP or HTTP GET keepalives by the router
- Server and application health monitoring provided by local Server Load Balancers
- Same VIP addresses can be advertised from multiple Data Centers
- Relying on L3 routing protocols for route propagating and content request routing
- Disaster Recovery provided by network convergence

IP SLA: Implementation

- Upstream router can be configured to “inject” a 32-bit host route as a static route in the routing table using IP SLA/Tracking
- The router injects or removes the route based on the health of the back-end servers (checked with ICMP, TCP or HTTP GET)



Advantages of the IP SLA Approach

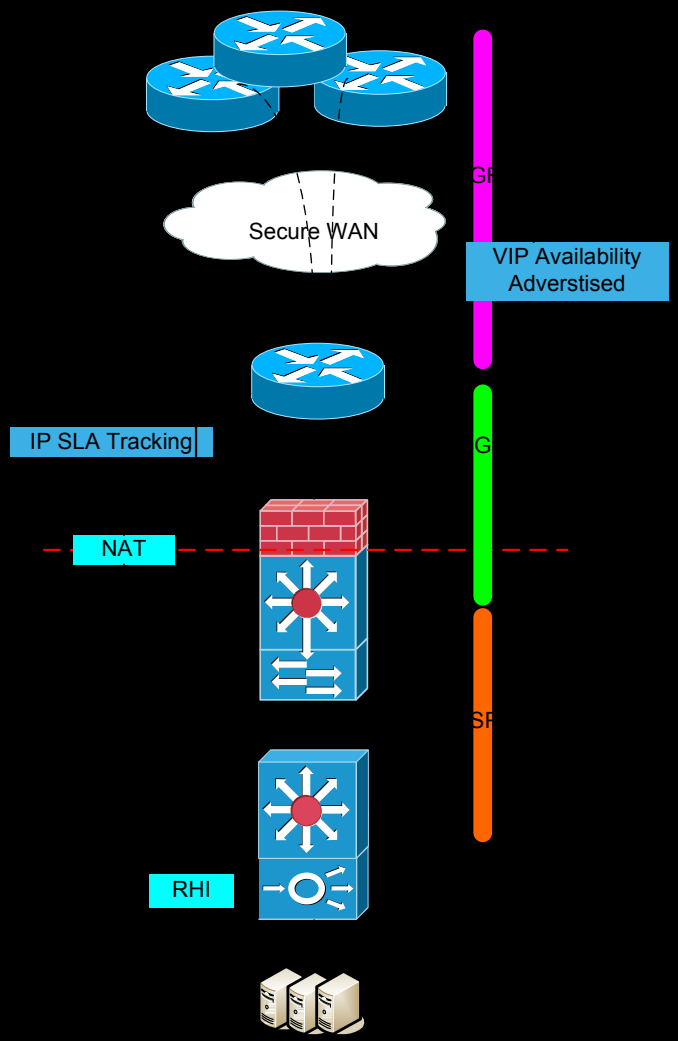
- It can track NATed VIP
- Segmentation for security and load-balancing functions. No need to turn on inspections on the distribution devices and/or ACE
- Routing protocol and environment tuning can account for very fast convergence during failure conditions
- This design can be used during application migration where VIP addresses cannot be changed

Case Study

Challenge 1

- Customer requirements
  - – All inbound traffic to the Data Center needs to be NATed and any selective outbound traffic also needs to be NATed for specific hosts
  - – 3 Routing protocols and scheme
    - EIGRP for WAN
    - BGP for WAN and Core Switches
    - OSPF private networks
  - – The VIPs must be advertised out dynamically to the branch for reachability to the applications

Diagram



Solution: IP SLA

- Proposed Solution
  - ACE deployed inside the FW with RHI enabled
  - IP SLA / Tracking on WAN edge router

Consideration

- The inside private network routing protocol is OSPF. All hosts and networks sitting in the inside that needs to be accessed by the branches will be statically NATed at the firewall
- When RHI is enabled, static routes will be redistributed into OSPF and then in turn OSPF networks will be redistributed in BGP. This is how the branch sites are aware as to how to get to the VIPs and/or inside hosts
- Use of distribute-lists, ACLs and route-maps to filter routes to prevent any routing loops, since we are doing mutual redistribution between protocols

Implementation

- ACE is used as the local SLB device
- ACE is configured with RHI to inject the application VIP address into local MSFC routing table
- OSPF will redistribute the /32 VIP address to Data Center Core switches
- FW is configured to NAT the application VIP address to a static outside IP address
- WAN edge router will have IP SLA object tracking configure to monitor the health of the NATed IP
- The edge router will inject the /32 NATed IP into EIGRP routing domain if the VIP is inservice

IP SLA: Pros and Cons

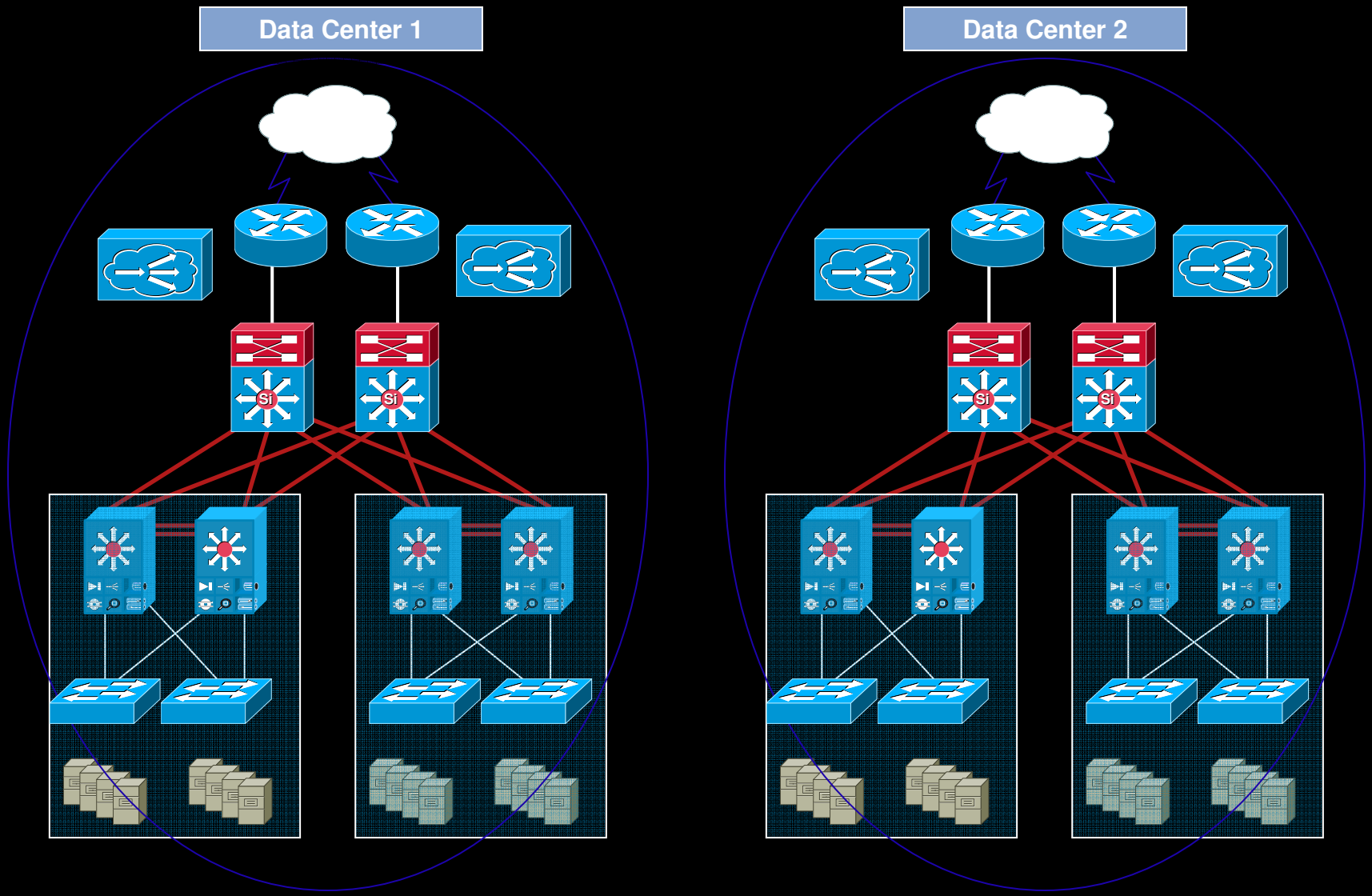
- Pros:
  - Providing NATed VIP route injection into routing table
  - Security for the VIP is being handled by the Firewall, instead of Load Balancer
- Cons:
  - If a new application with the same VIP with different port need to be added, the IP SLA could not track the VIP availability at the port level
  - It could be administratively challenging when the number of applications require NATed VIPs grow

DNS Based Site Selection

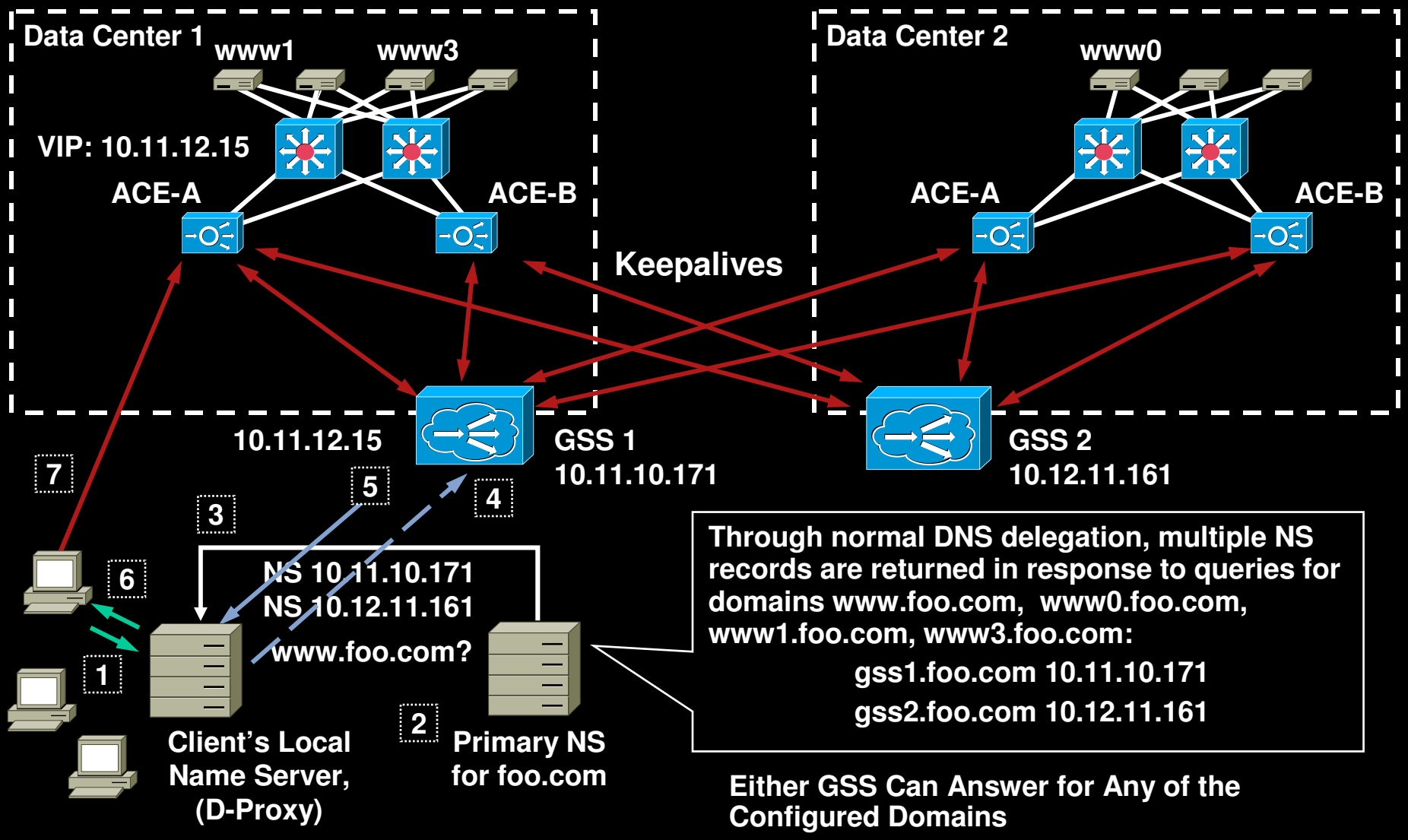
DNS-Based Solution: Global Site Selector

- The GSS operates at the DNS control plane, as authoritative name server for Load Balanced domains
- Communicates with ACE/CSS/CSM or servers that are located in the Data Centers
- Provides DNS replies based on one or more of the following:
  - Source IP of the requester, network topology
  - Destination domain (can be wildcarded)
  - Configured methods (orders and weights)
  - Proximity (from requesting D-proxy to the data center)
  - Health and load of the data centers
- Multiple GSSs can be clustered to provide redundancy

GSS Placement



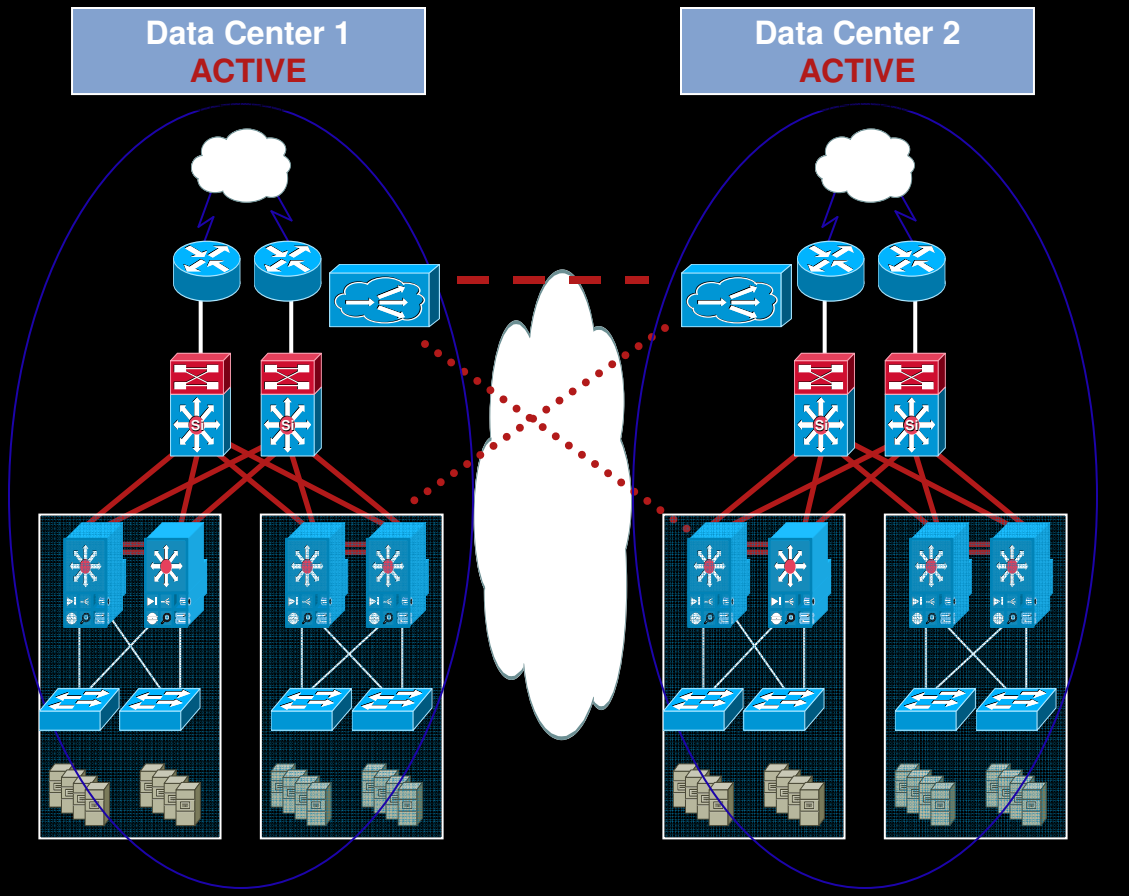
GSS Deployment Details



GSS Keepalives Challenges

GSS-to-GSS and GSS-to-VIP Keepalives

- Challenge 1
  - When links to ISPs in DC1 fail, the keepalives may flow through internal network. GSS still sees that answers are ‘up’. This will create a black Hole for clients who gets name-to-address resolution of DC 1 VIP
- Challenge 2
  - Keepalives will need to traverse through perimeter Firewall to reach the VIPs



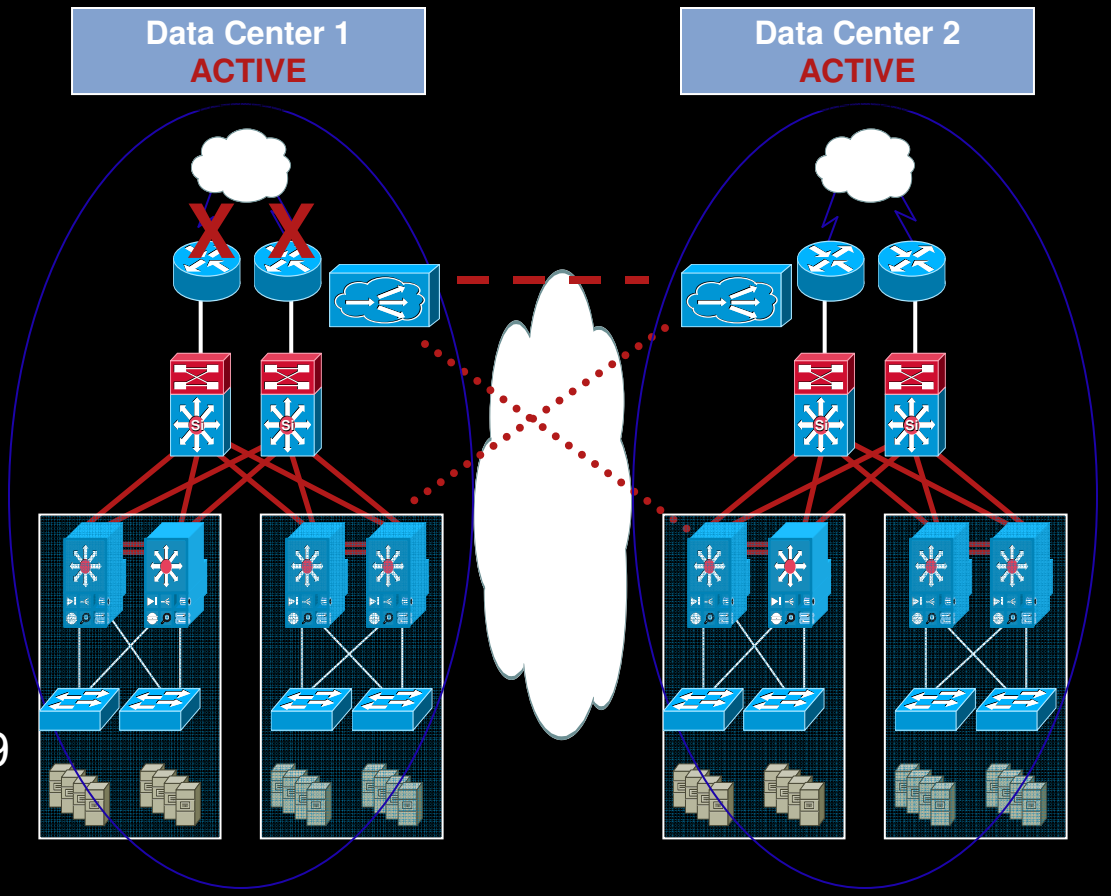
GSS Keepalives Considerations

GSS-to-GSS and GSS-to-VIP Keepalives

- Solution 1
  - – Tie the health of VIP in DC1 to the availability of ISP link using scripts or 'VIP dependency' command
- Solution 2
  - – Firewall will be configured to allow :

    UDP - 1304, 2000, 5002

    TCP - 2001-2009, 3001-3009



GSS DNS Rules

Defines How to Respond to DNS Query Requests as Follows:

- Requests arriving from a certain D-proxy
- Asking for a certain hosted domain
- Use this answer group
- With this balance method to choose the best answer

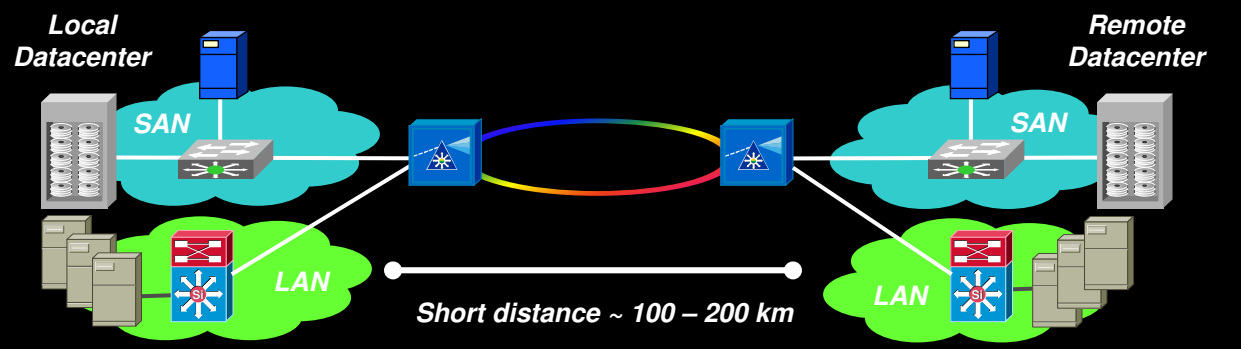
Business Continuance and Disaster Recovery

Datacenter interconnect options

Datacenter interconnect options

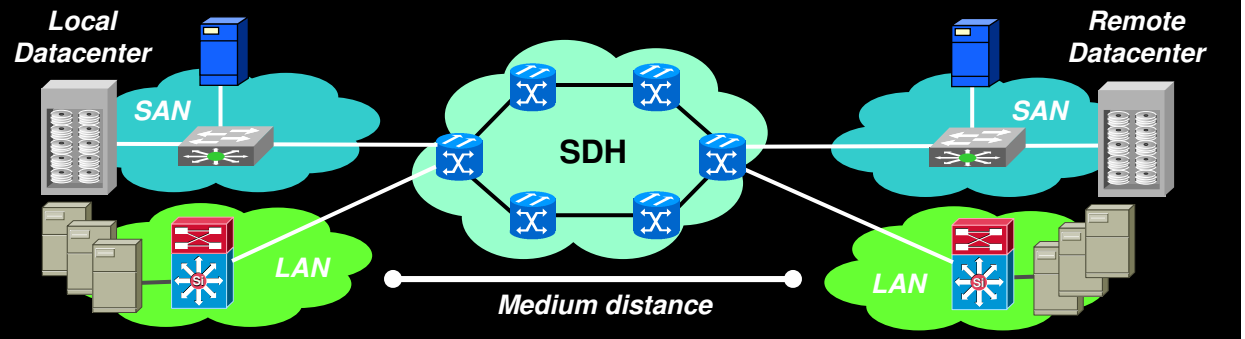
DWDM/CWDM

- most often short distance
- dark fiber must be available
- dedicated channels for LAN, SAN and other signals



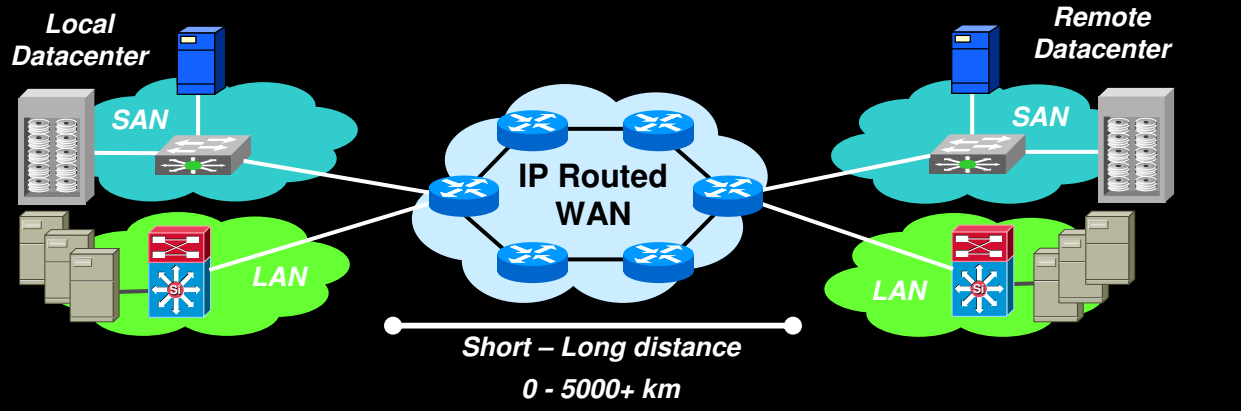
SONET/SDH

- most often short – intermediate distance
- dark fiber not avail. – distance, cost, exhaust
- links may be shared
- EoSDH and FCoSDH

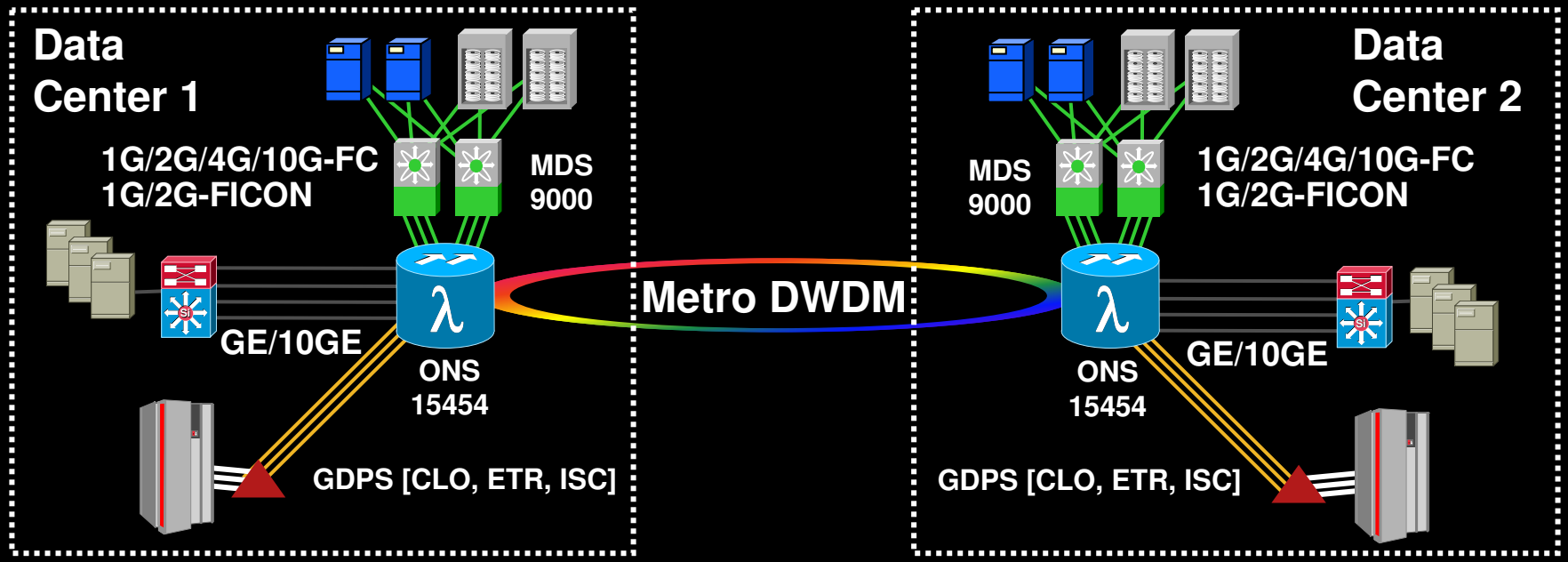


IP, IP/MPLS, Metro Ethernet

- short – long distance
- dark fiber not available
- links may be shared
- FCIP for FC and/or FICON

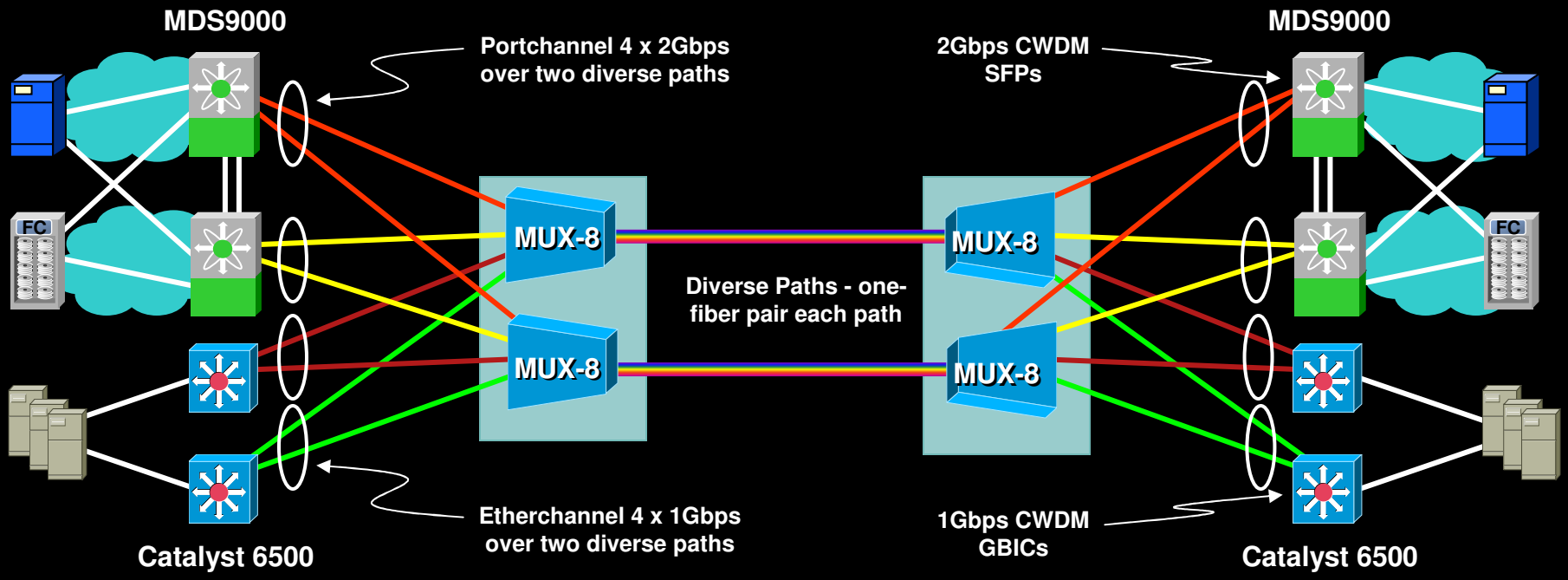


DR solution with transponder based DWDM
Cisco ONS 15454 MSTP



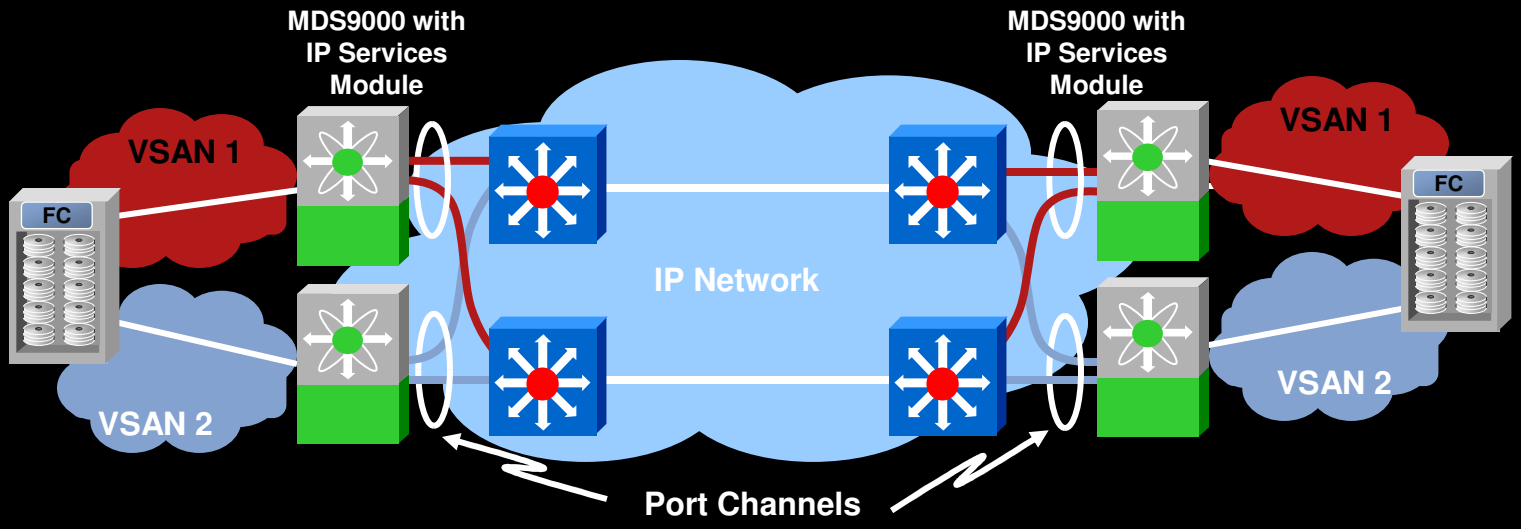
- Support of many different channel types: GE, 10GE, FC/FICON (1/2/4/10G), SDH (STM-1/4/16/64/256), ESCON, IBM solution specific interfaces (CLO, ETR, ISC), video interfaces, 2R transparent signal etc.
- Cost-effectively aggregates data and storage services into 2.5 or 10 Gbps lambda
- End-to-end Cisco Storage + IP over DWDM with VSAN support
- Buffer-to-buffer credits for distance extension
- Optical performance monitoring and comprehensive protocol (payload) monitoring
- Certified by major system/storage vendors (incl. IBM GDPS certification)
- Suitable for enterprise, regional and SP networks (including long-haul)

DR solution with integrated WDM optics
Cisco CWDM and DWDM passive filters and pluggables



- **Uses colored interfaces (GBICs, SFPs, XENPAKs) in CWDM or DWDM wavelength grid plugged directly in communication devices (ethernet or FC switches) and passive DWDM or CWDM filters**
- **Lower cost than transponder based system but less functionality**
- **Can be combined with Cisco ONS 15454 MSTP solution**

DR solution based on IP or IP/MPLS
SAN extension with FCIP

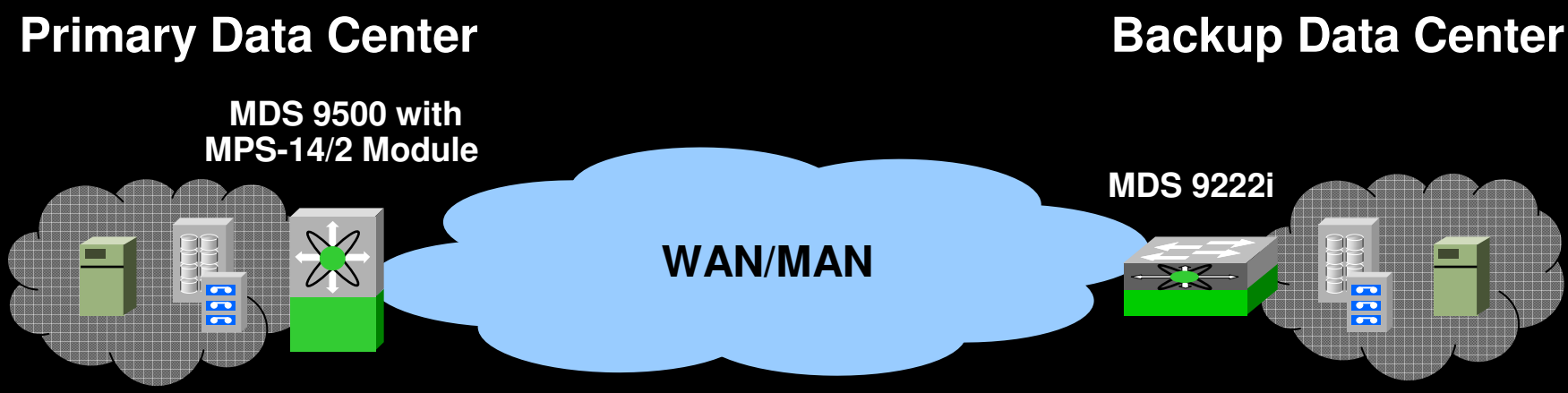


- Data network uses natural IP connectivity
- SAN extension uses FCIP

  FCIP enhancements can be used – shaping, compression, encryption, QoS marking, Inter VSAN routing, write and tape acceleration etc.
- Some other connection may be tunneled over IP or IP/MPLS using VoIP, TDMoIP, AToM etc.
- May be combined with optical technologies (like WDM) to increase bandwidth and reliability

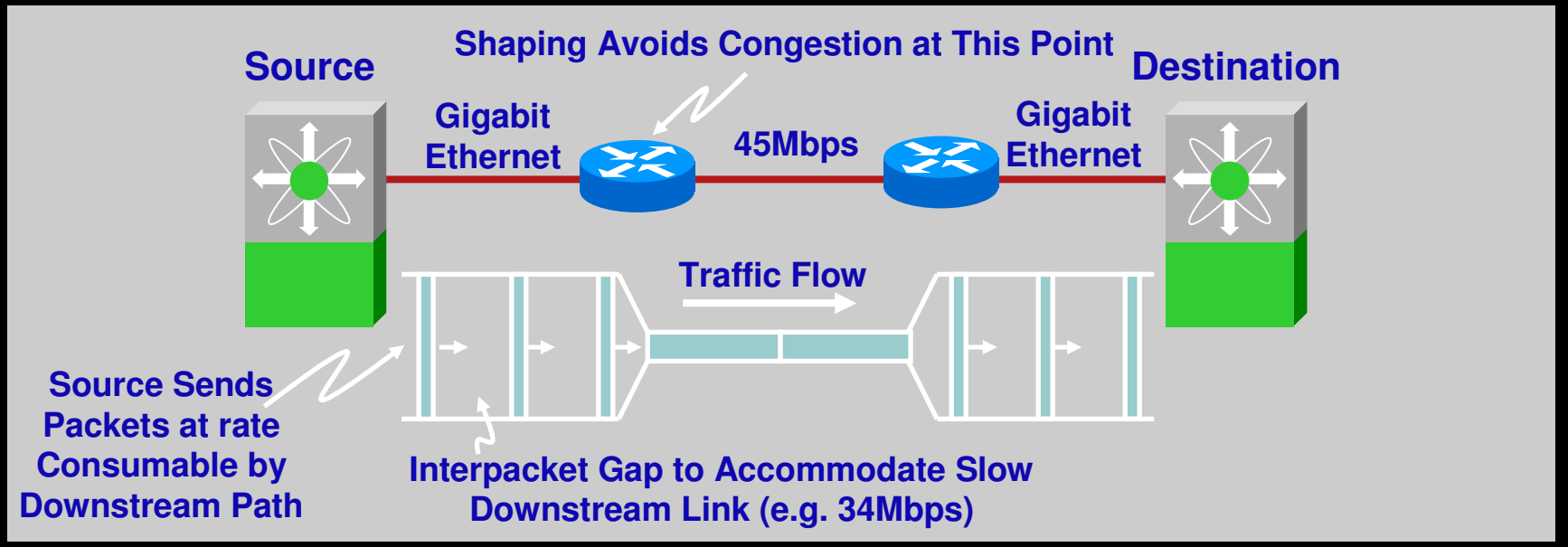
Cisco MDS 9000 FCIP Implementation

Comprehensive SAN Extension Solution over IP



Application Performance	**WAN Bandwidth Utilization**	**Security**	**Traffic Management**
Tape & Write Acceleration	Hardware Compression	FCIP Encryption and FC-SP Auth.	IVR, QoS, TCP Tuning, IPv6
Application Availability	**Application Tuning**	**Cost Reduction**	**Solution Management**
Inter-VSAN Routing	SAN Extension Toolkit (SET)	VSAN-Enabled Consolidation	Multiprotocol Fabric Manager

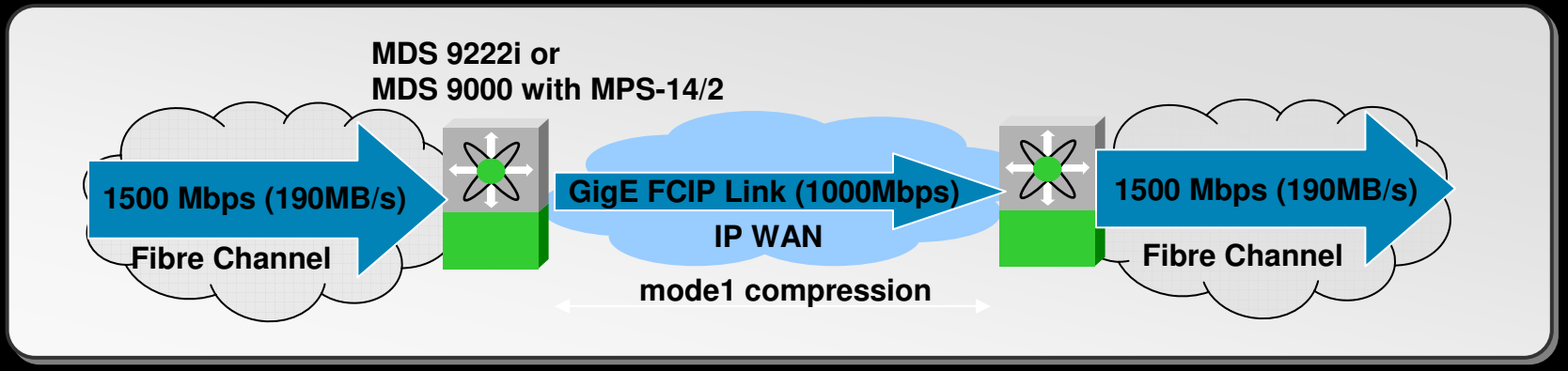
FCIP TCP Packet Shaping
Matching the available bandwidth



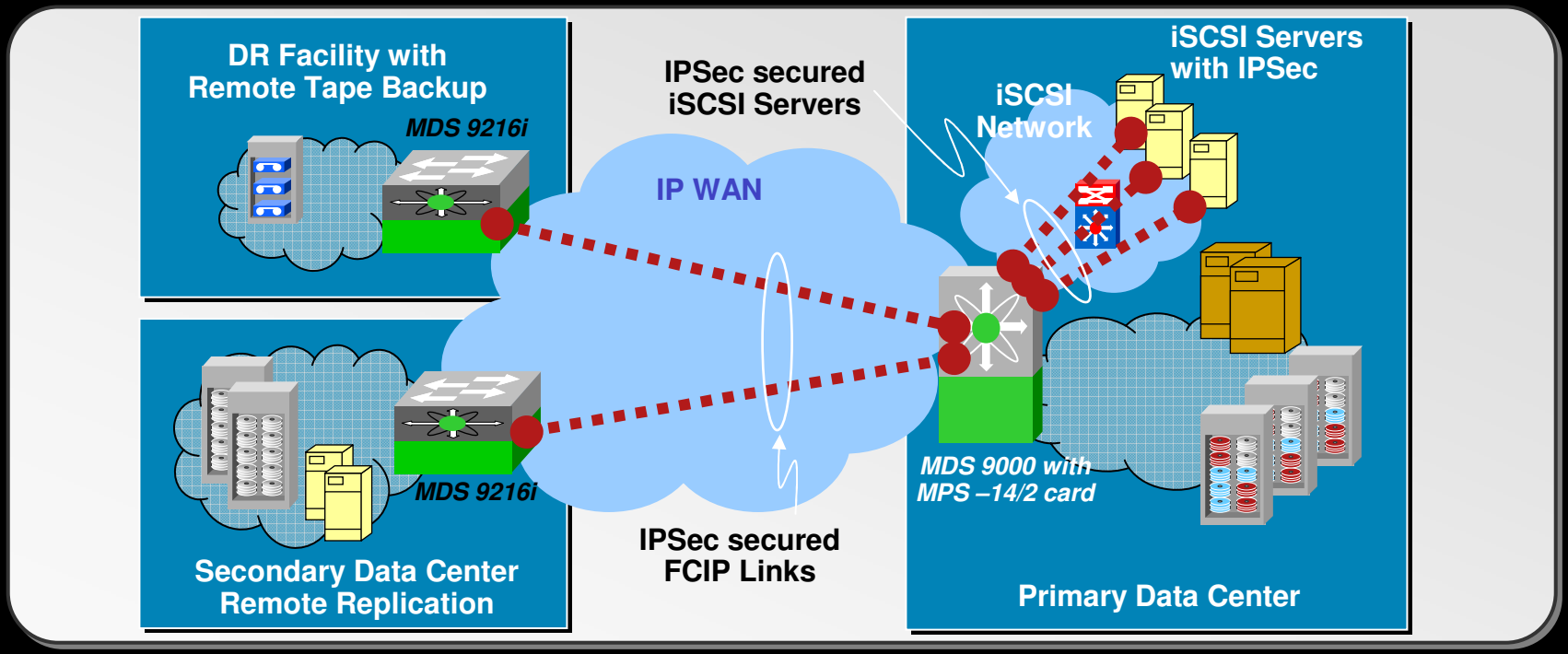
- Shaper sends at a rate consumable by the downstream path
  - Immediately sends at “minimum-bandwidth” rate (avoids early stages of traditional slow start)
  - Ramps up to “maximum-bandwidth” rate (using usual slow start and congestion avoidance methods)
- Requirements for shaper to engage:
  - Min-available-bandwidth > 1/20 max-bandwidth
  - SACK (Selective Ack) must be enabled

FCIP throughput optimization:
Integrated FCIP Compression

- Compression lowers WAN costs - more throughput with less bandwidth
- MPS-14/2 card and MDS 9222i offers Hardware Compression
  - Up to 190MB/s of Fibre Channel throughput over single GigE
- Compression Ratio depends on data stream
- Three Compression Modes - choose appropriate Mode for WAN Link
  - Mode1: WAN up to 1000Mbps – compression up to 9:1
  - Mode2: WAN up to 25Mbps – compression up to 30:1
  - Mode3: WAN up to 10Mbps – compression up to 33:1
- Frame batching for modes 2 and 3 – more compressed FC frames into one ethernet frame

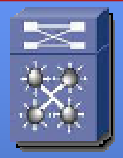
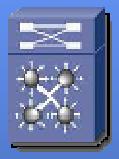
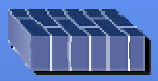


Securing Storage over Distance:
Wire Rate IPSec Encryption



- Ensures the Integrity and Confidentiality of Enterprise Data over FCIP
- Hardware-based GigE wire rate performance with latency ~ 10µs per packet
- Standards-based IPSec Encryption - implements RFC 2402 to 2410, & 2412
  - IKE for protocol/algorithm negotiation and key generation
  - IPSec ESP encapsulation with optional authentication and replay protection
  - Encryption: AES (128 or 256 bit key), DES (56 bit), 3DES (168 bit)
- Support for Digital Certificates

Cisco Data Center Product Families

Data Center Switching	Data Center Switching	Data Center Security	Storage	Application Network Services	Optical
					
Catalyst 6500 Series Catalyst 4948 Top-of-Rack Catalyst Blade Server Switches	Nexus 7000 Nexus 5000 Nexus 2000 Nexus 1000	Firewall Services Module Intrusion Detection Module CSA Server Security Agent ACE WAF	MDS 9500 Storage Directors MDS 91xx/90xx Fabric Switches MDS Blade Server Switches Storage Service Modules	Wide-Area Appliction Services ACE XML ACE SLB, SSL Termination, Application Acceleration GSS	Cisco ONS 15454 Cisco ONS 15216 CWDM SFP/GBIC XENPAK/X2/XFP

Data Center Management

Q and A