

CISCO SYSTEMS



High Availability – Globally Resilient IP

Richard Shao 

Cisco IOS Technical Marketing Engineer

rshao@cisco.com

5 – 9's Network Availability

Cisco.com

- **99.999 Network Availability**

**downtime allowed
per year = ~ 5.26
minutes**

Objectives and Overview

Cisco.com

- **Introduces Cisco IOS features to increase network availability**
- **1 hour session**
- **Overviews what a feature does, how it works, and what deployment scenario examples are**
- **Configurations are not covered**

Downtime is Lost Money

Cisco.com

Survey of 52 top Service Providers*

The average downtime costs incurred in the past 12 months:

\$21.6 Million.

(Ranges from \$500,000 to \$298 million)

Equates to an average of \$130,000/hour

*Source: Sage Research, Aug. 2001

Survey of Enterprise Customers**

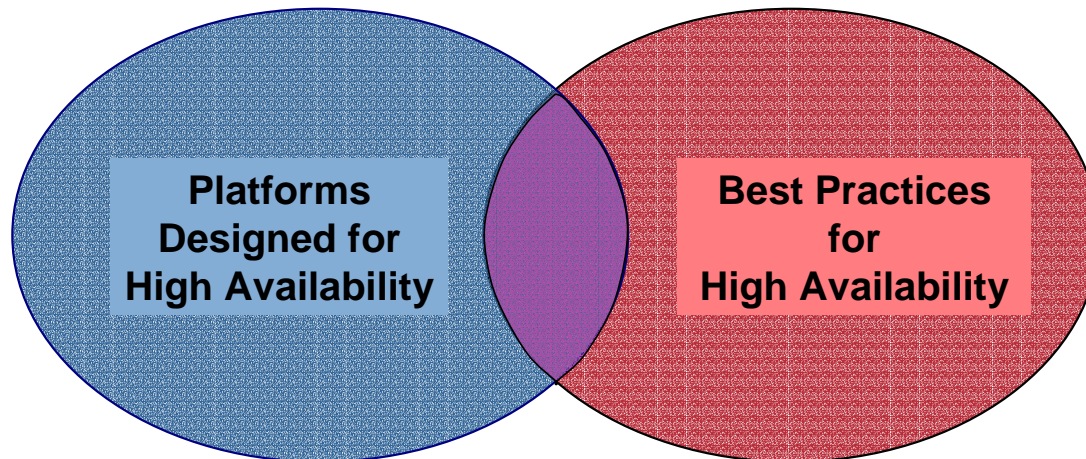
Energy:	\$2.8M/hr
Telecomm:	\$2M/hr
Financial:	\$1.5M/hr
Retail:	\$1M/hr
Chemicals:	\$704K/hr
Health Care:	\$636K/hr
Media:	\$340K/hr
AVERAGE:	\$1M/hr

** Source: Network Computing
(<http://www.networkcomputing.com>), March 5, 2001]

Traditional Industry Approaches to Achieve High Availability IP Networks

Cisco.com

Platform Design & Best Practices



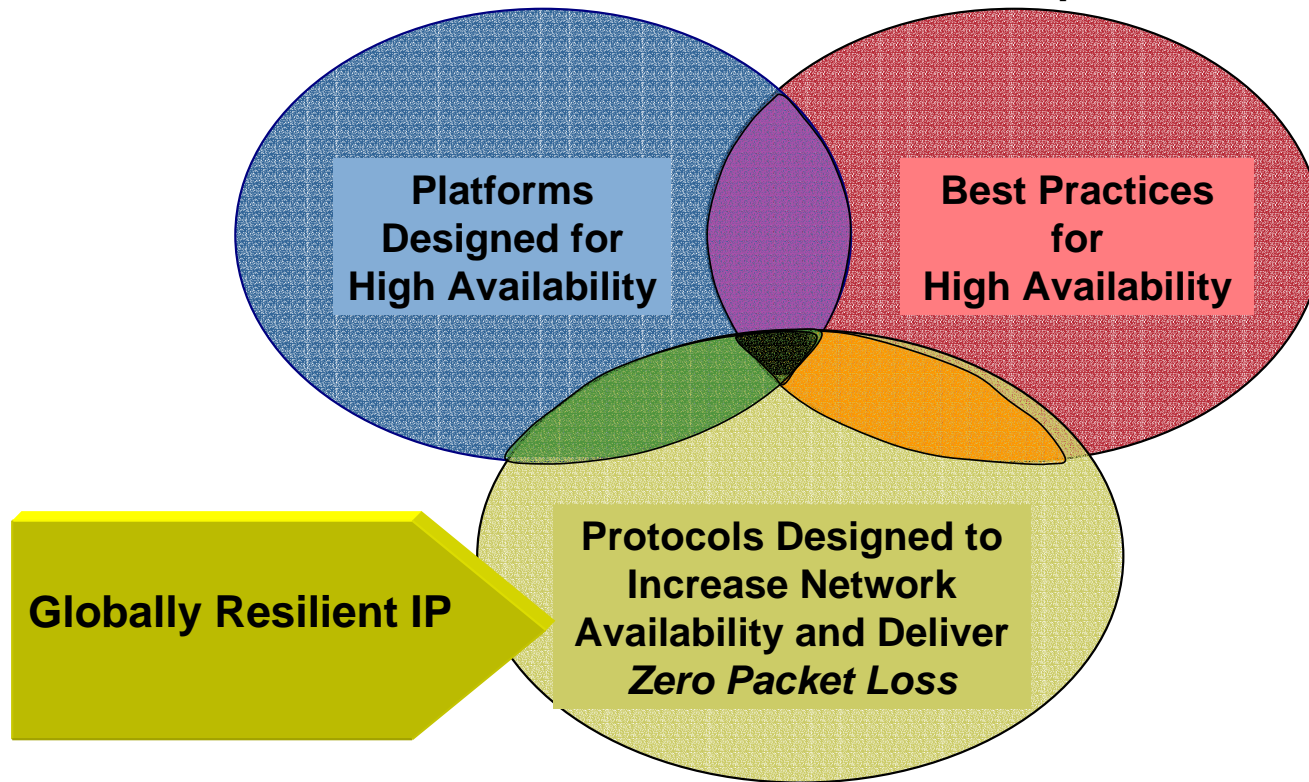
- ✓ • Redundant power & fans
- Dual CPUs & line cards
- NEBS-compliant platforms

- ✓ • Network design
- Operations excellence
- Training & Education

Where/How Can We Further Improve Network Availability ?

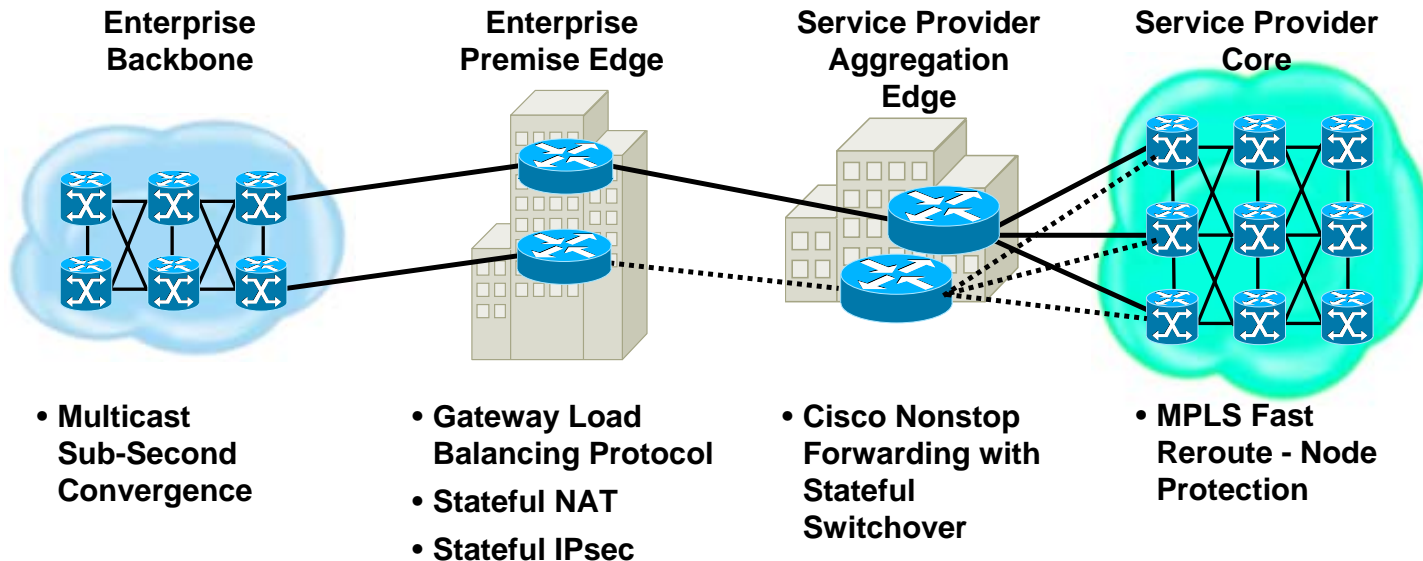
Cisco.com

To meet network availability expectations for real-time online services, IP networks require:



New Cisco IP Technology Solutions for Globally Resilient IP Networks

Cisco.com



Routing Protocol Convergence Enhancements

- IP Event Dampening
- BGP Convergence Optimization
- Incremental SPF Optimization

Enterprise Backbone / Premise Edge

Multicast Sub-Second Convergence
Gateway Load Balancing Protocol
Stateful NAT
Stateful IPSec

Multicast Sub Second Convergence

Cisco.com

- Really a series of enhancements.

Join/Prune aggregation: We used to send one PIM packet per (S,G) or (*,G) entry after an Rendezvous Point failover. We now aggregate these into only a few PIM packets with multiple entries.

Triggered RPF checks following unicast convergence. Now, as soon as unicast is converged, it causes an instantaneous start of RPF checks (Old default was 5 seconds).

New PIM HELLO option: Added option which advertises the HOLDDTIME in milliseconds. This allows subsecond failover of Designated Router. This is Cisco proprietary.

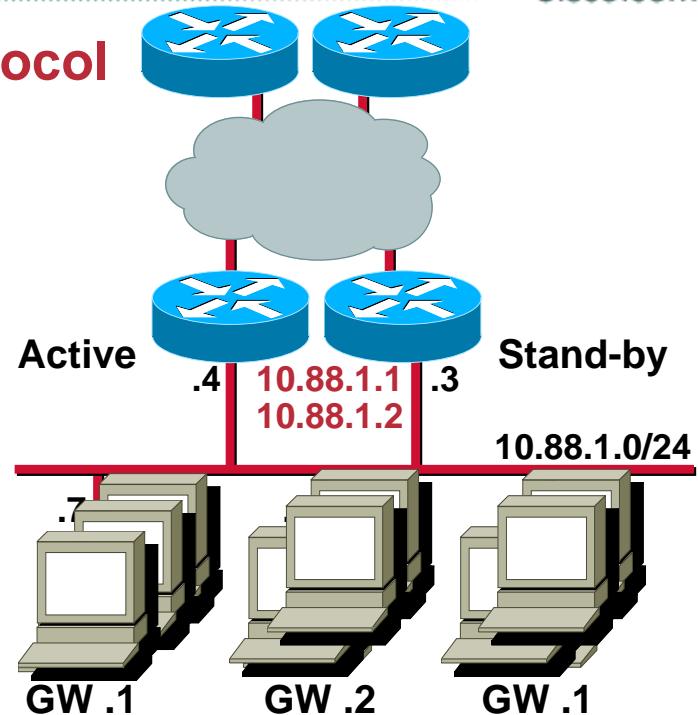
```
ip pim query-interval <interval> [msec]
```

Current HSRP Operation

Cisco.com

Cisco Hot Standby Router Protocol

- Provides redundant default gateway services
- Load balancing possible by configuring different default gateways

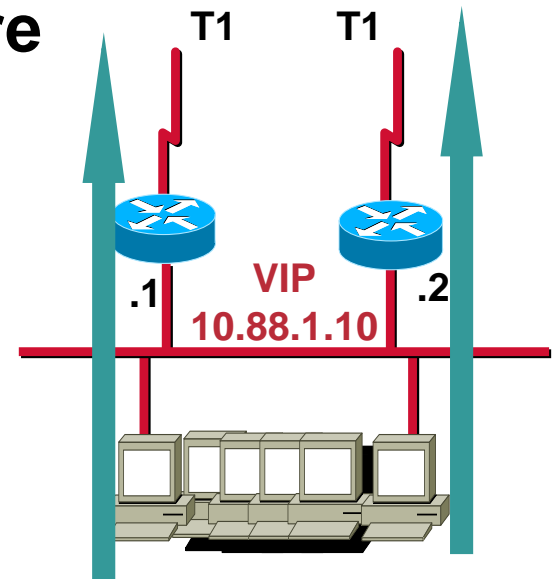


Problem: Either resource not fully utilized or extra administrative costs needed

Gateway Load-balancing Protocol

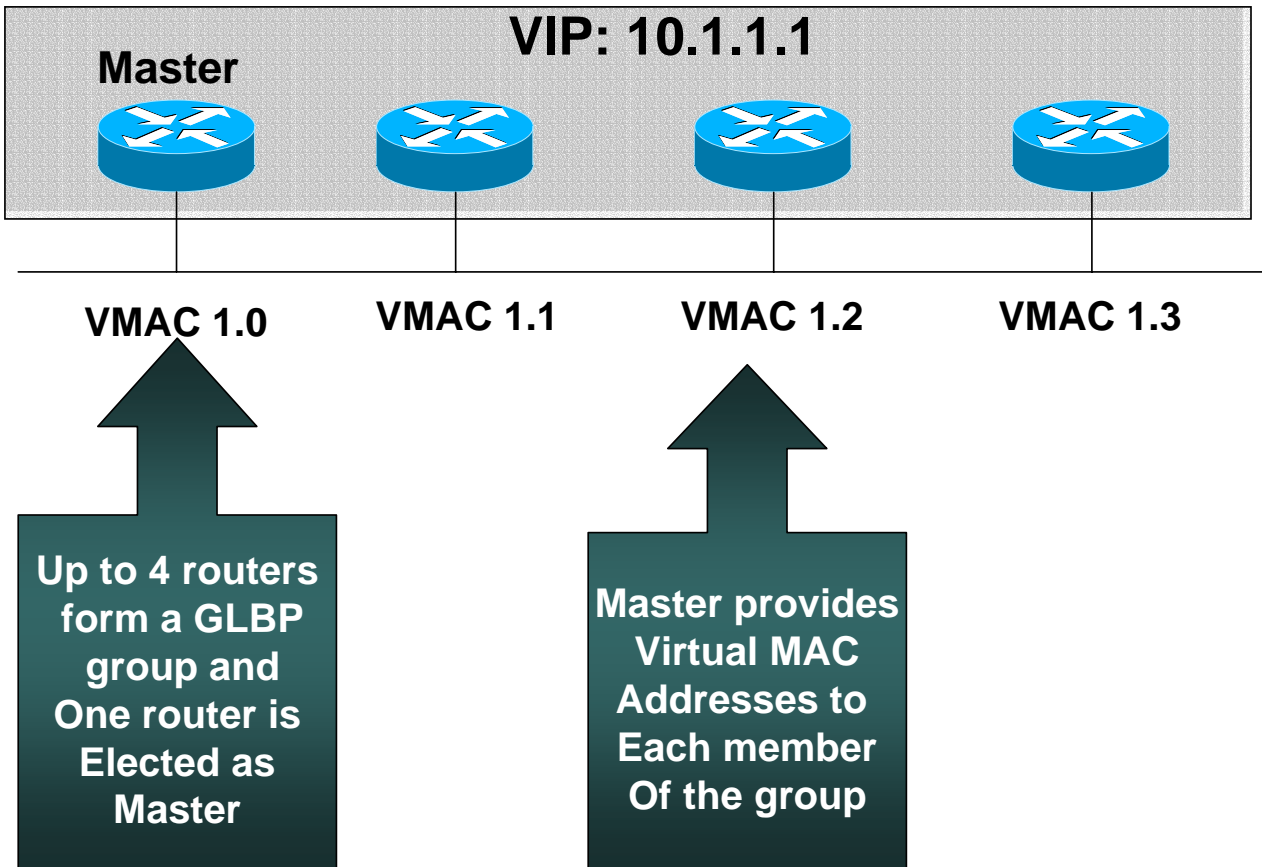
Cisco.com

- GLBP allows routers within an HSRP group to load-share connections
- Increases forwarding processing by number of routers in the GLBP group
- Increases upstream bandwidth
- Reduces administrative tasks



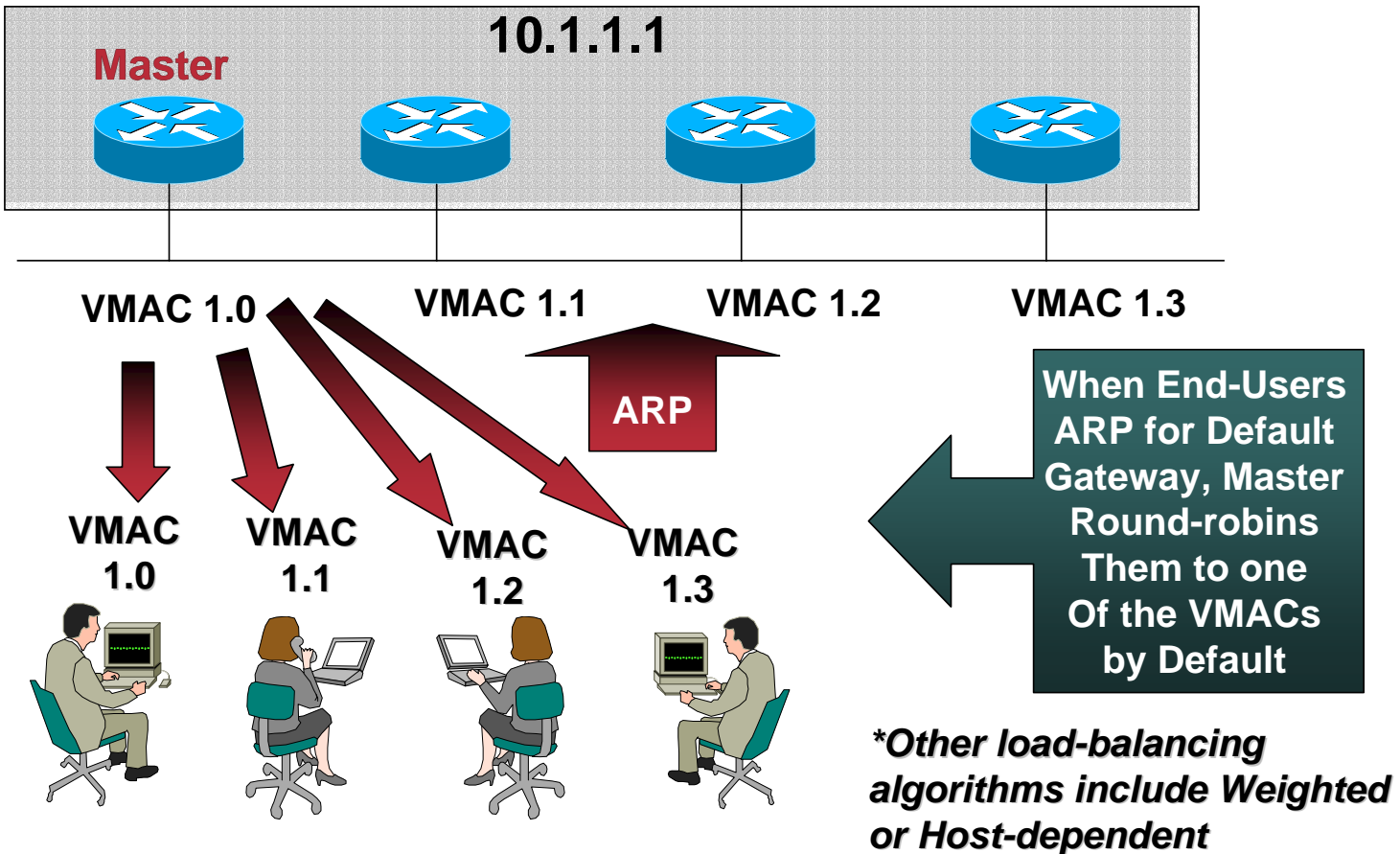
How GLBP Works

Cisco.com



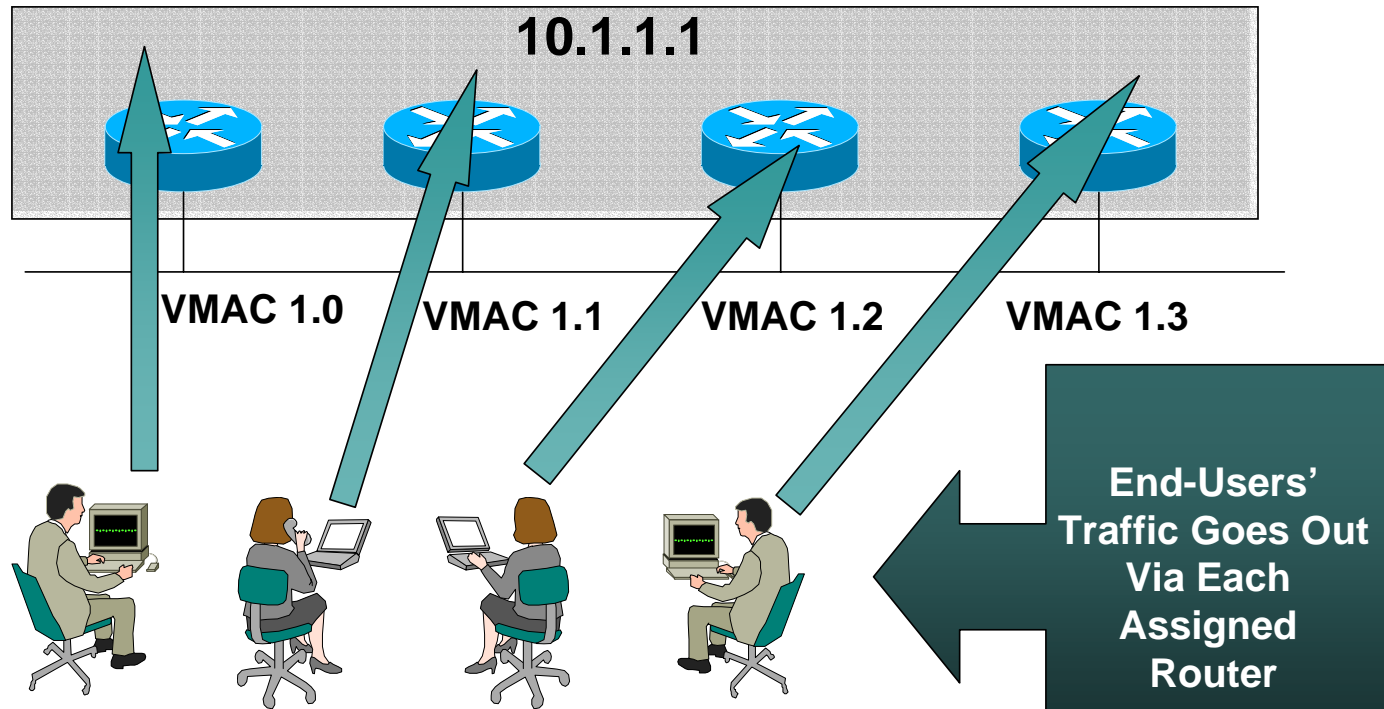
How GLBP Works

Cisco.com



How GLBP Works

Cisco.com



Automatically traffic load balancing without configuring different default gateways on the users' devices

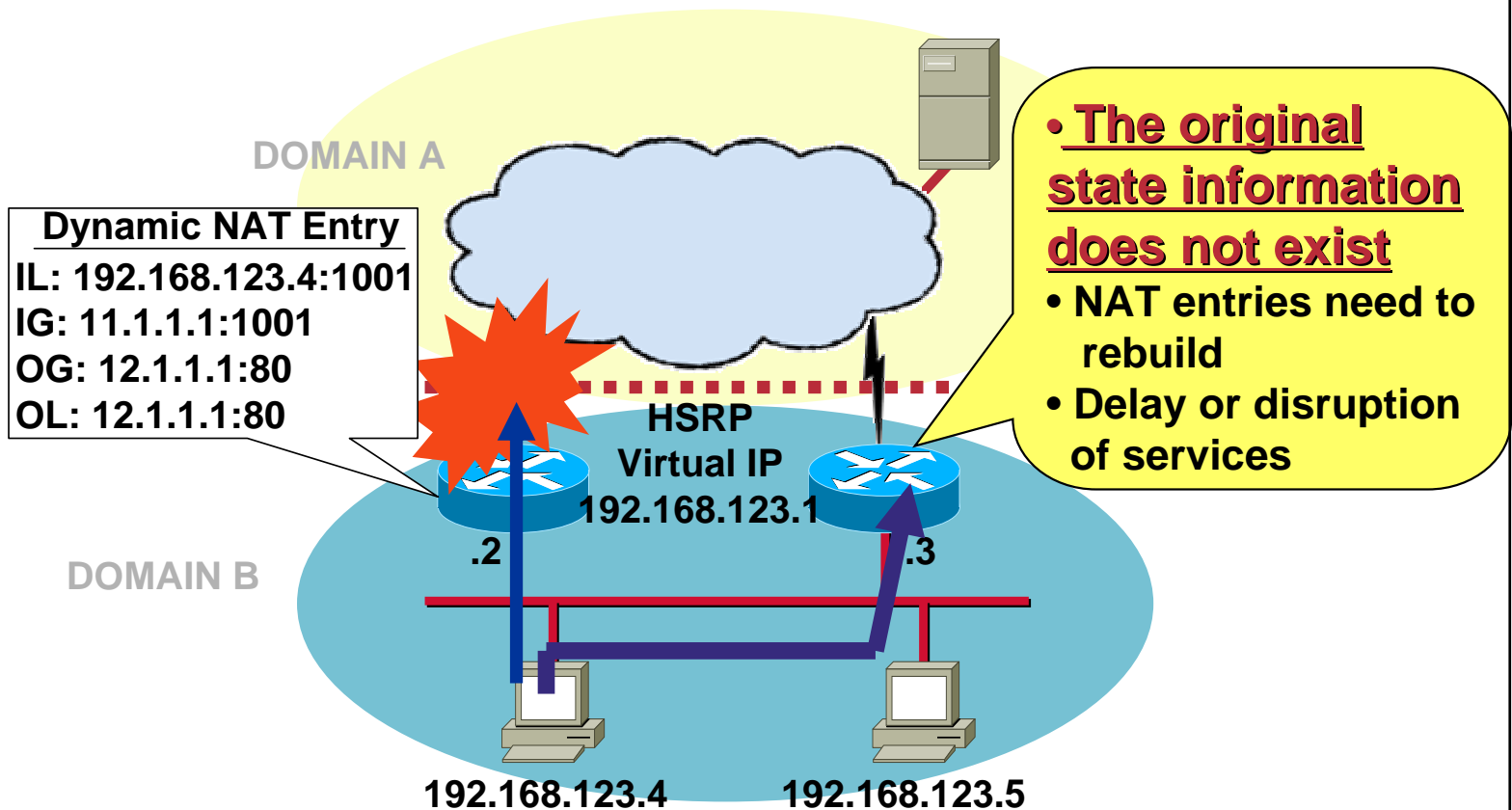
Questions to Ponder

Cisco.com

- **Do you consider your network restored:**
 - ...When a router is back on line?**
 - ...When the router is ready to serve the clients and applications?**
 - i.e. performing network address translation, encrypting/decrypting packets, and routing packets**

NAT And HSRP

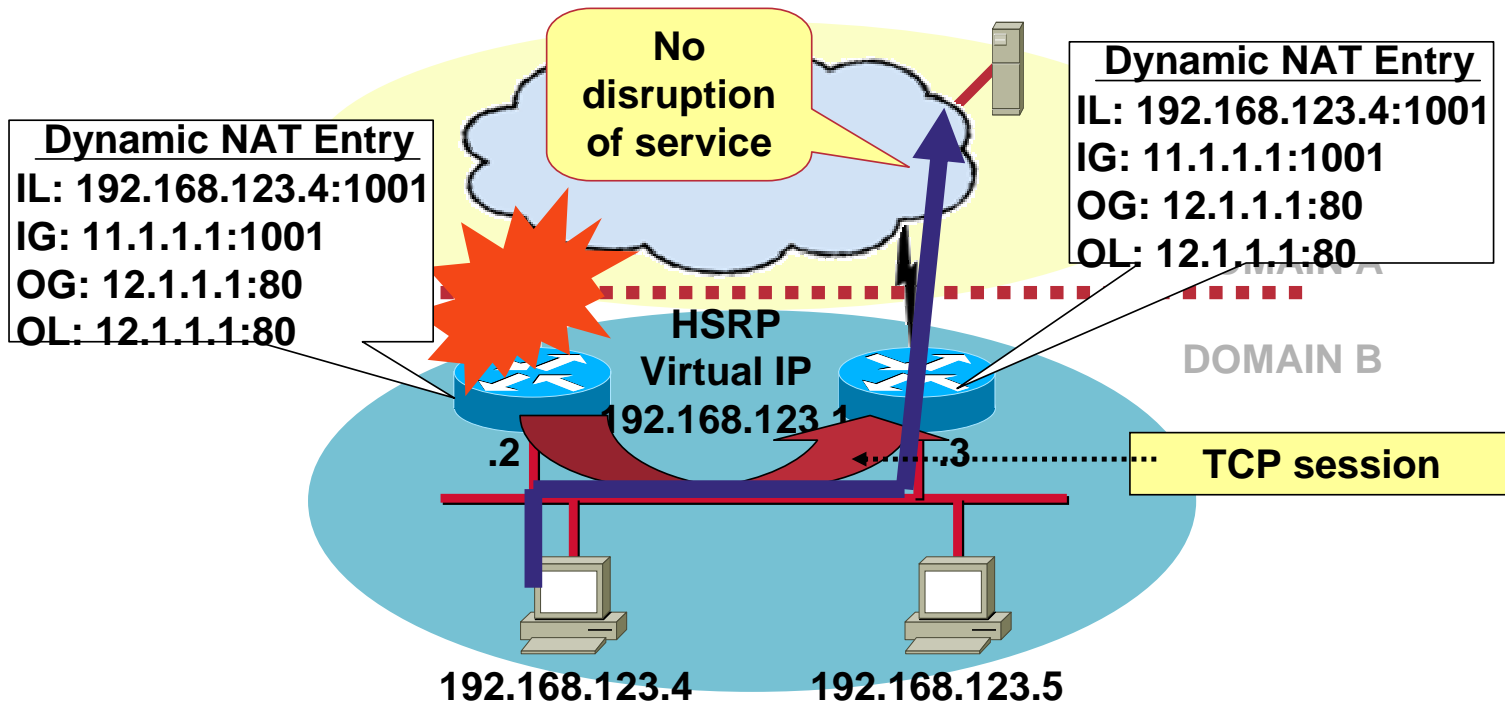
Cisco.com



Stateful NAT And HSRP

Cisco.com

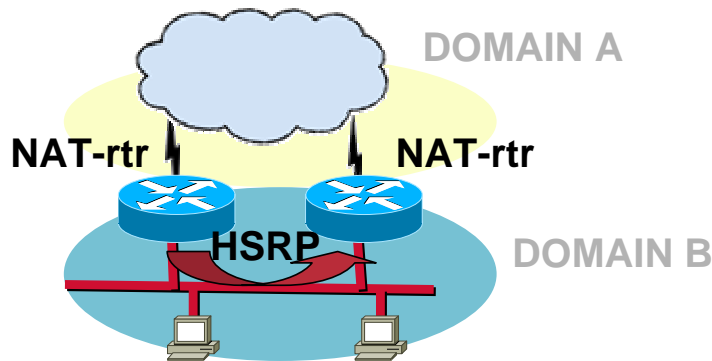
- Stateful NAT allows two (or more) routers that are providing redundancy with a network to share state information concerning existing NAT sessions



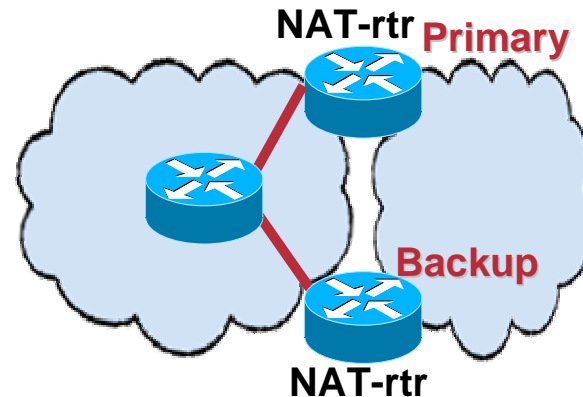
Stateful NAT: Deployment Variations

Cisco.com

- **Stateful NAT with HSRP – HSRP Mode:**
- Utilizing HSRP to elect primary and backup election
- Typical in the access layer



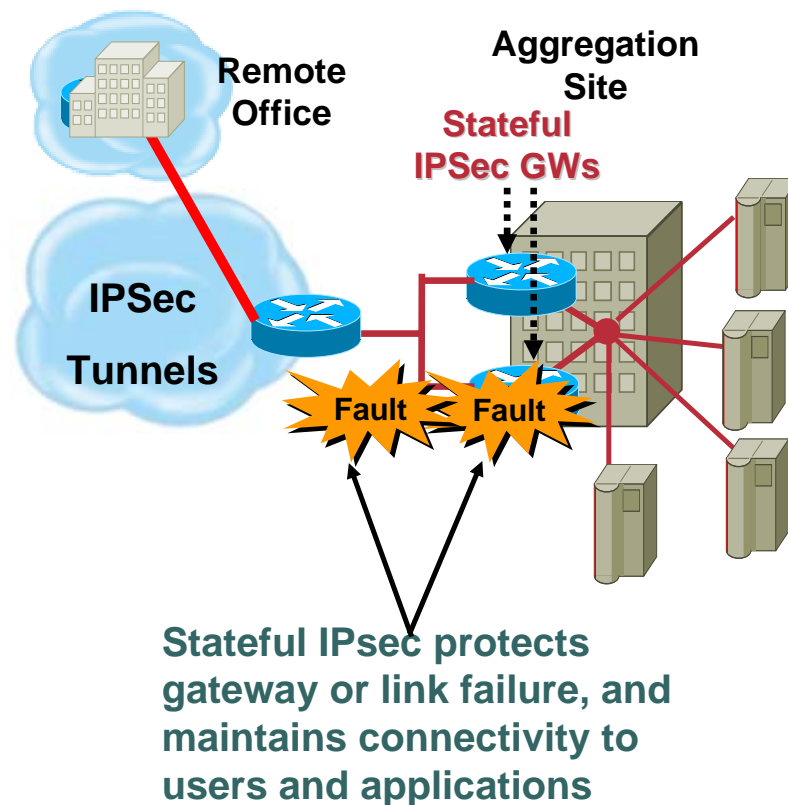
- **Stateful NAT without HSRP – Primary/Backup Mode:**
- The primary and backup state is manually configured
- Typical in the edge of the networks



Protecting the Customer Edge: Stateful IPsec Tunneling

Cisco.com

- **TCP connection formed from Active to each standby router**
- **State Synchronization Protocol (SSP) is used to transfer state**
- **State info includes sequence number counters, IKE Session keys, Security Association attributes, such as cipher, authentication and compression algorithms.**



Stateful IPsec Deployment Example

Cisco.com

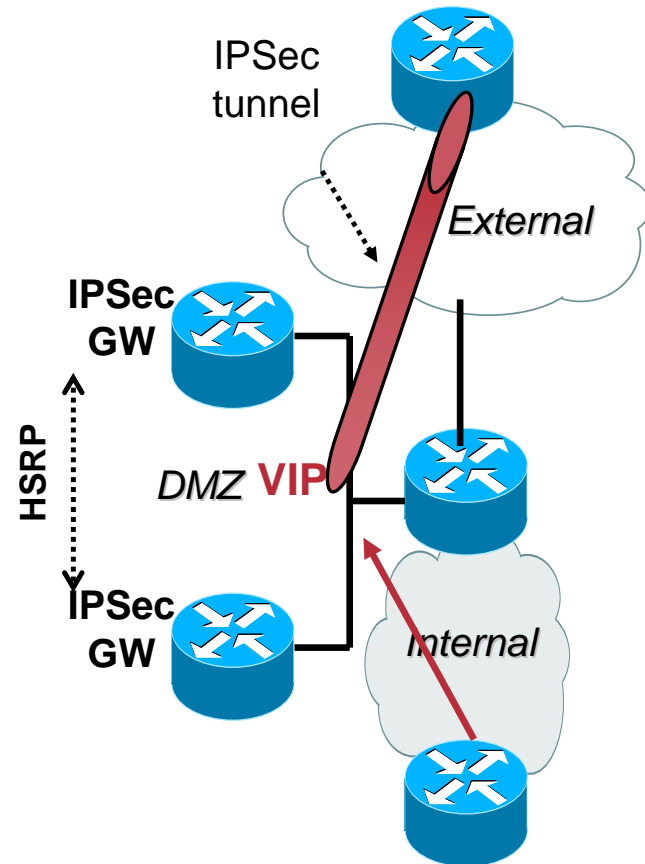
- **Used in conjunction with HSRP**

HSRP Virtual IP is used as source/destination for IPsec tunnels

HSRP active/standby election is used to determine a stateful IPsec peer status

- **Routing manipulation may be needed**

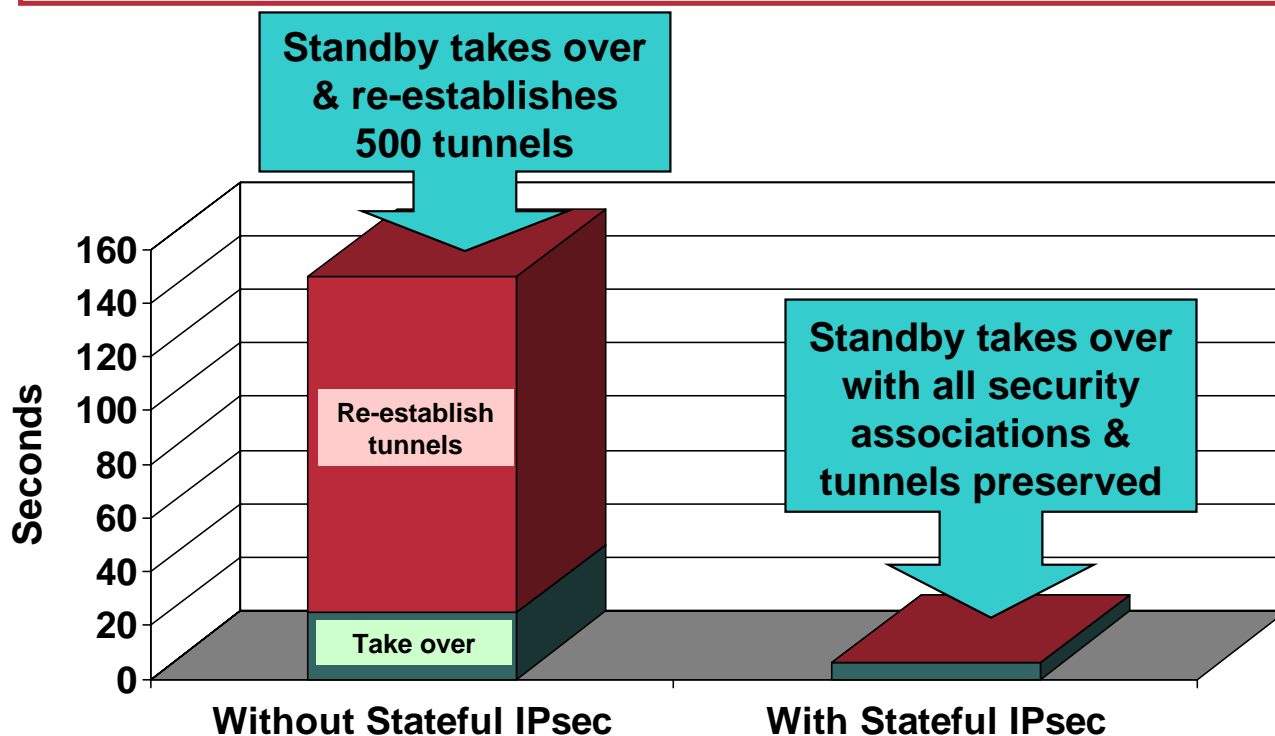
i.e. traffic from internal to external needs to send to the VIP



Resilient IP Services: Stateful IPsec

Cisco.com

Stateful IPsec substantially reduces time & preserves tunnels during first-hop takeover



Enterprise / Service Provider Edge

Nonstop Forwarding with Stateful Switchover

Stateful Switchover/ Non-stop Forwarding Overview

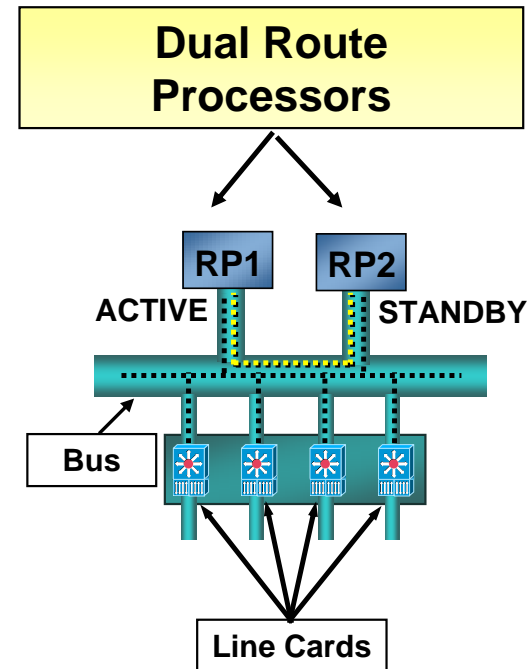
Cisco.com

- **Stateful Switchover (SSO):**
Allows stateful maintenance of Layer 2 protocols after a route processor switchover

Supports PPP, ATM, Frame Relay,
Cisco HDLC today

Will support MPLS, multilink PPP,
other protocols in future

- **Non-Stop Forwarding (NSF):**
Uses CEF or dCEF table to continue forwarding packets during a switchover, and as routing table is being rebuilt

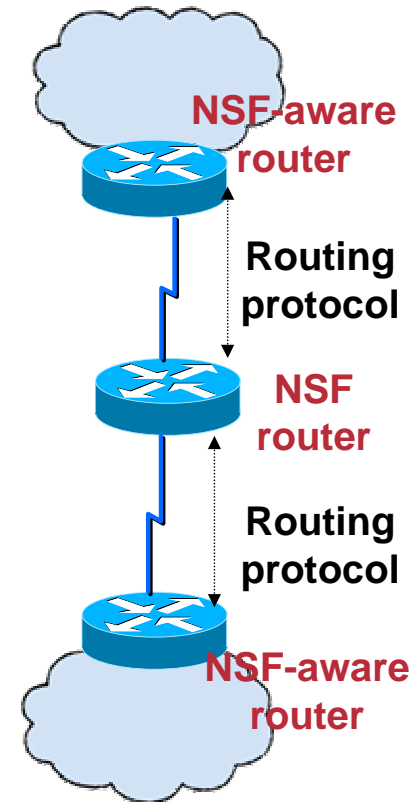


Enhancements to Routing Protocols

Cisco.com

- Even though router continues to forward packets, there is still a problem
- *Neighbor routers must know* that an NSF router can still forward packets
- Enhancements to OSPF, BGP and ISIS designed to prevent route-flapping

Ex: BGP Graceful Restart Capability



Relevant IETF Internet Drafts Authored/Coauthored by Cisco

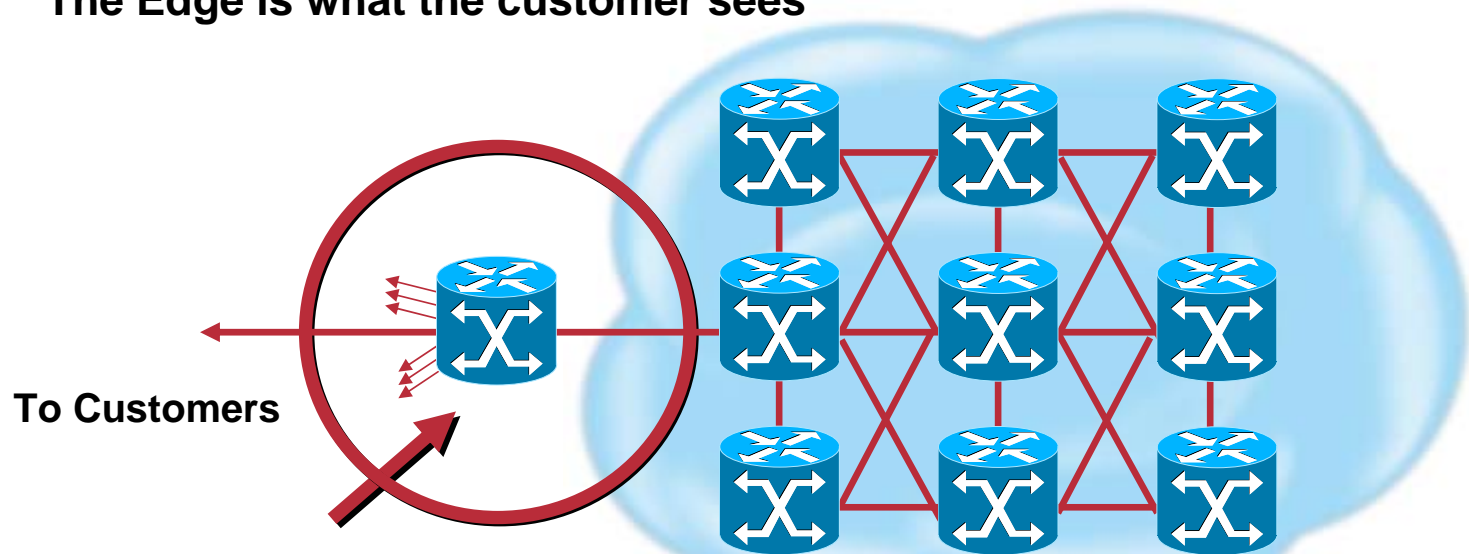
Cisco.com

- **Cisco Nonstop Forwarding – BGP Graceful Restart**
<http://www.ietf.org/internet-drafts/draft-ietf-idr-restart-03.txt>
- **Cisco Nonstop Forwarding – ISIS Restart**
<http://www.ietf.org/internet-drafts/draft-shand-isis-restart-01.txt>
- **Cisco Nonstop Forwarding – OSPF Restart**
<http://www.ietf.org/internet-drafts/draft-nguyen-ospf-lls-00.txt>
<http://www.ietf.org/internet-drafts/draft-nguyen-ospf-restart-00.txt>
<http://www.ietf.org/internet-drafts/draft-nguyen-ospf-oob-resync-00.txt>

Nonstop Forwarding and Stateful Switchover are Invaluable at the Edge

Cisco.com

The Core is redundant enough to disguise failures
The Edge is a Single Point of Failure
The Edge is what the customer sees



**Failures here may affect
thousands of customers
or remote offices**

**** Features will work in the core**

Sample Recovery Times

Cisco.com

	Single RP	RPR	RPR+	SSO/NSF
Cisco 7500	4:50	2:06	0:30	0:06–0:08
Cisco 10000	2:45	0:26	0:14	0:01.63
Cisco 12000	2:32	1:20	0:08	<u>0:00</u>
Cisco 6500/7600	4:50	3:00	0:30	TBD
Link Flaps?	Yes	Yes	No	No
Route Flaps?	Yes	Yes	Yes	No

Service Provider Core

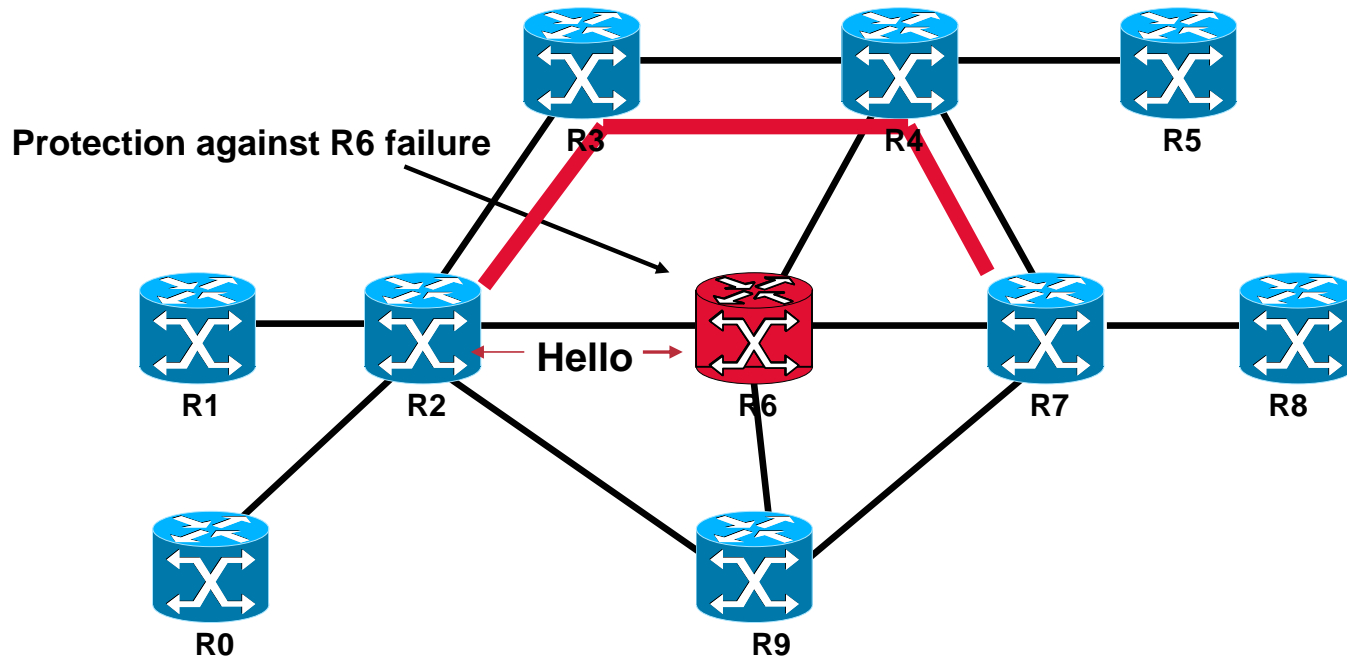
MPLS Fast Reroute – Node Protection

MPLS Traffic Engineering Fast Reroute

– Node Protection

Cisco.com

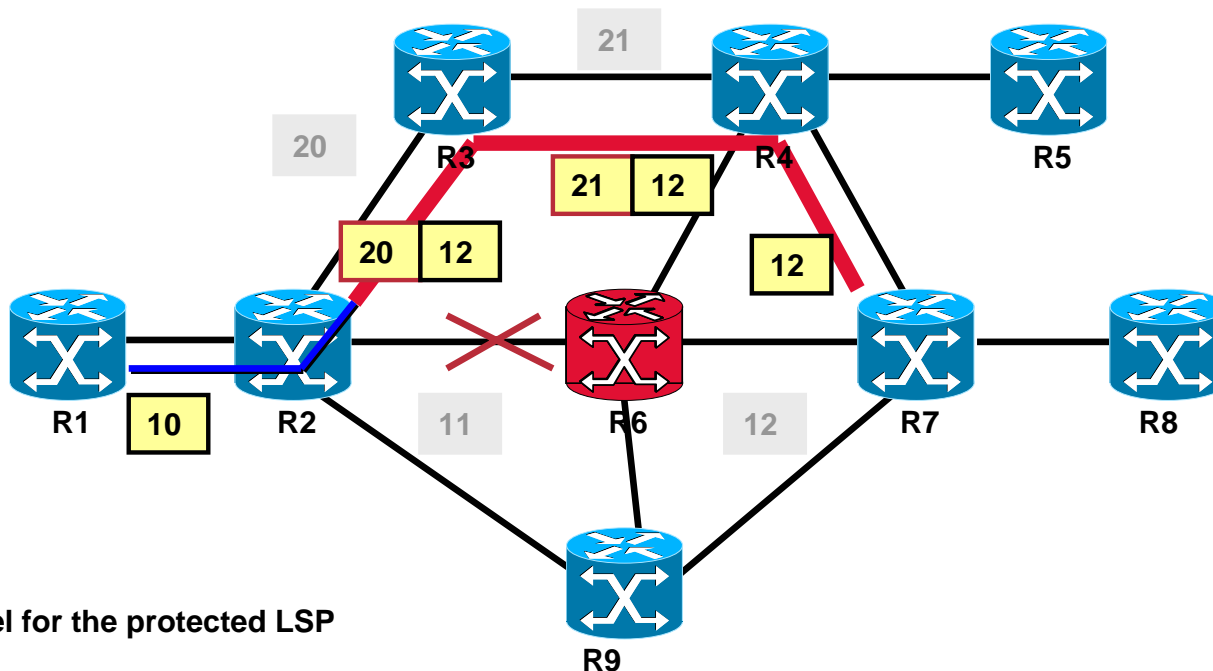
- Node protection allows to configure a back-up tunnel to the next-next-hop. This allows to protect against link AND node failure



RSVP HELLO extension to monitor availability of NHOP router

MPLS TE FRR – Backup Labels

Cisco.com



x Label for the protected LSP

- The Point of Local Repair learns the label to use from the Record Route Object object carried in the Resv message when the reroutable LSP is first established – Global Label Allocation must be used.

Network Wide

**IP Event Dampening
BGP Convergence Optimization
Incremental SPF**

IP Event Dampening - Concept

Cisco.com

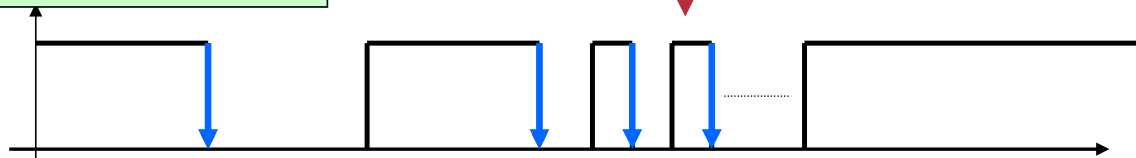
- **Takes the concept of BGP route-dampening and applies it as a “generalized dampening” library for all protocols**
applying penalty to a flapping interface
- **Concept is: If you have a link that is flapping, and causing constant route reconvergence, it is better for network stability to remove that link from the routing domain altogether**
- **Allows alternate routes to be used until the flapping link can be investigated or has proven itself to be trustworthy**

IP Event Dampening – Algorithm Illustration

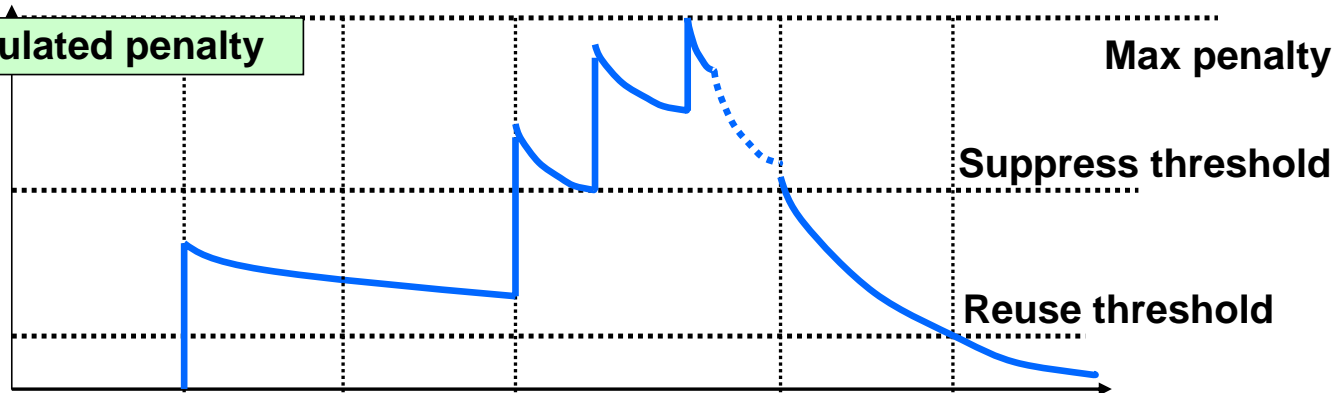
Cisco.com

Routing churn avoided

Actual interface state



Accumulated penalty

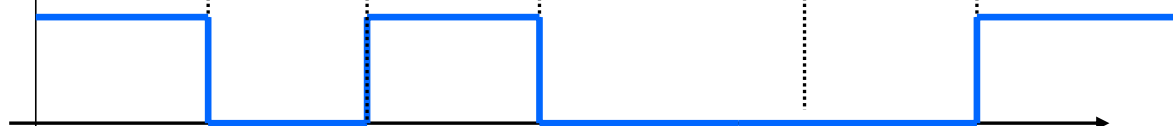


Max penalty

Suppress threshold

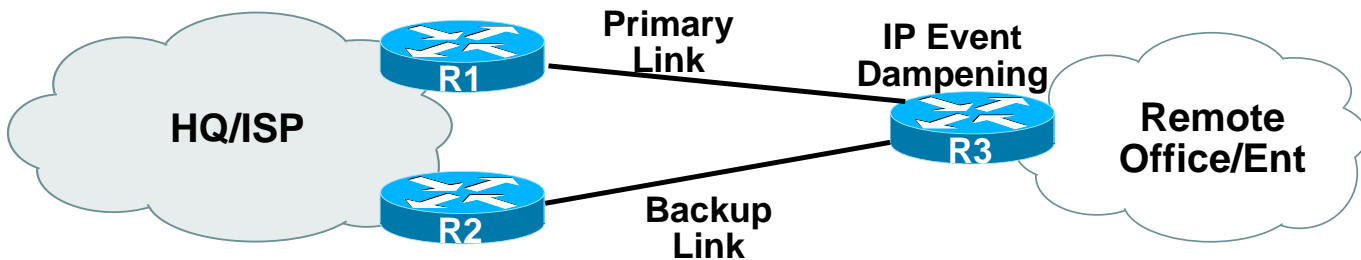
Reuse threshold

Perceived interface state by routing protocols

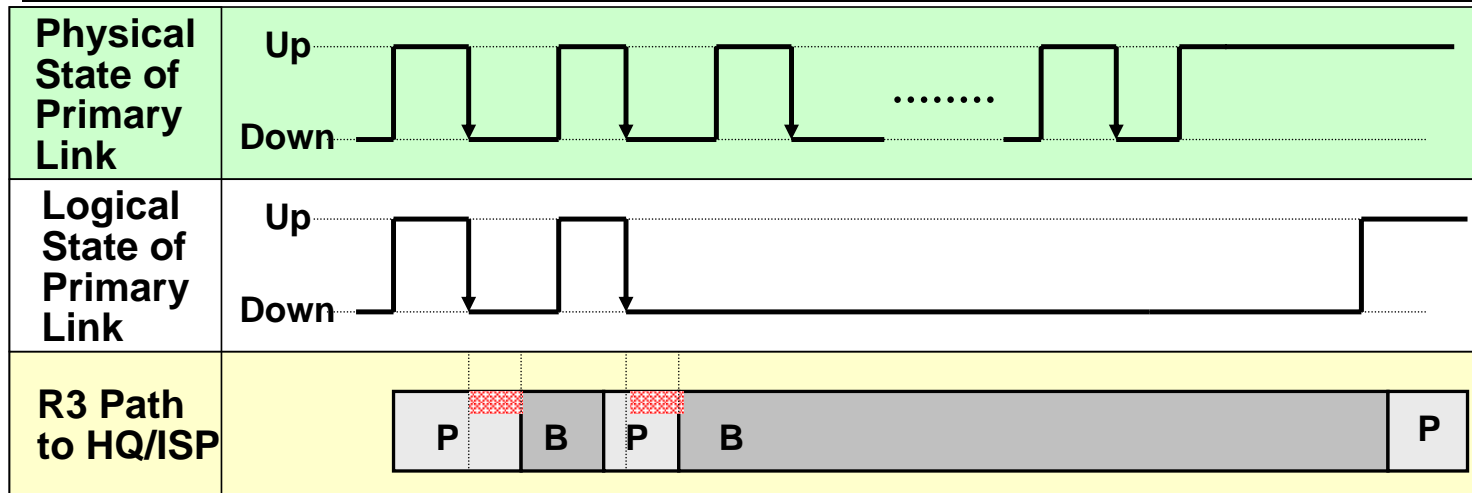


IP Event Dampening – Deployment

Cisco.com



IP Event Dampening absorbs link flapping effects on routing protocols



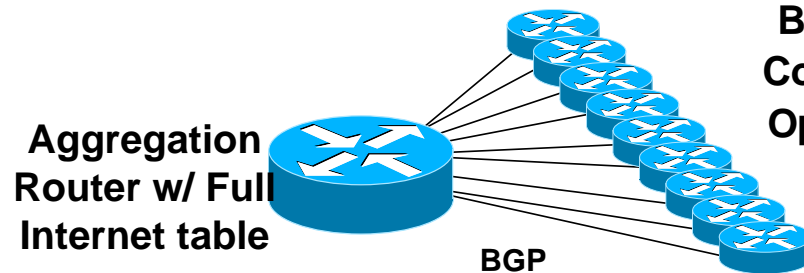
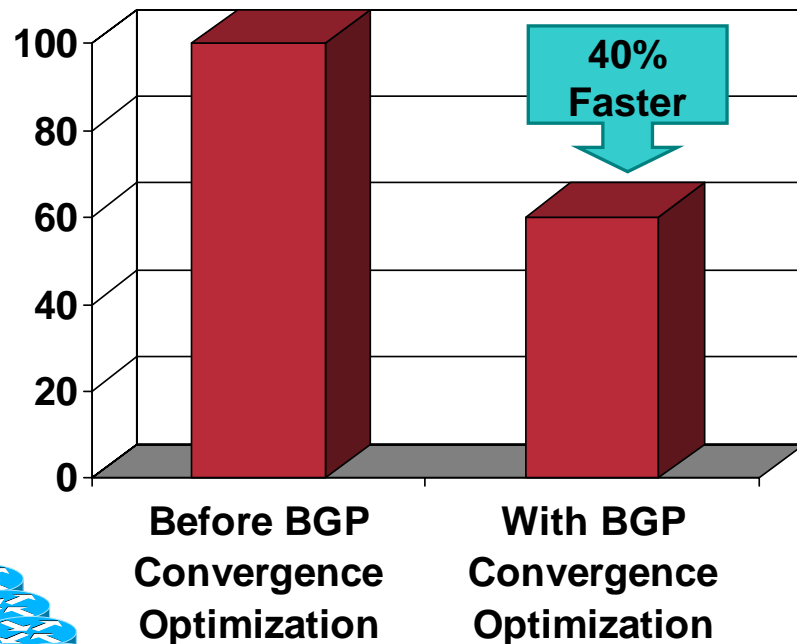
Duration of packet loss

BGP Convergence Optimization

Cisco.com

- **A series of enhancements:**

- *BGP Update Packing:*
- *Memory backoff algorithm*
- *TCP MTU Path Discovery*



Link State SPF Computation

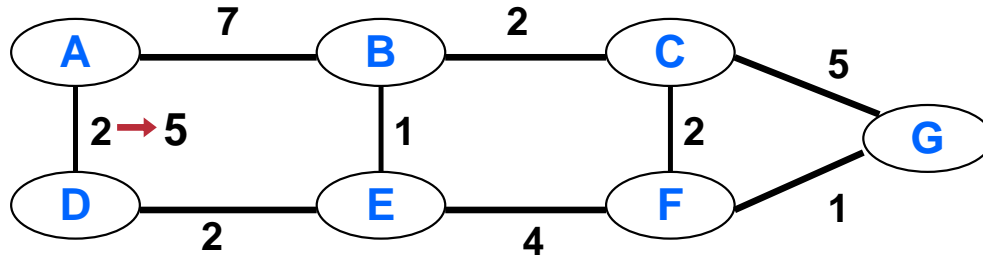
Cisco.com

- **Some changes effect only a partial of the Shortest Path Tree (SPT)**
- **Some changes do not effect the SPT at all**
- **Thus, it is unnecessary to run “Full” SPF computation when there is a topology change, or to run SPF at all when receiving a new LSA**

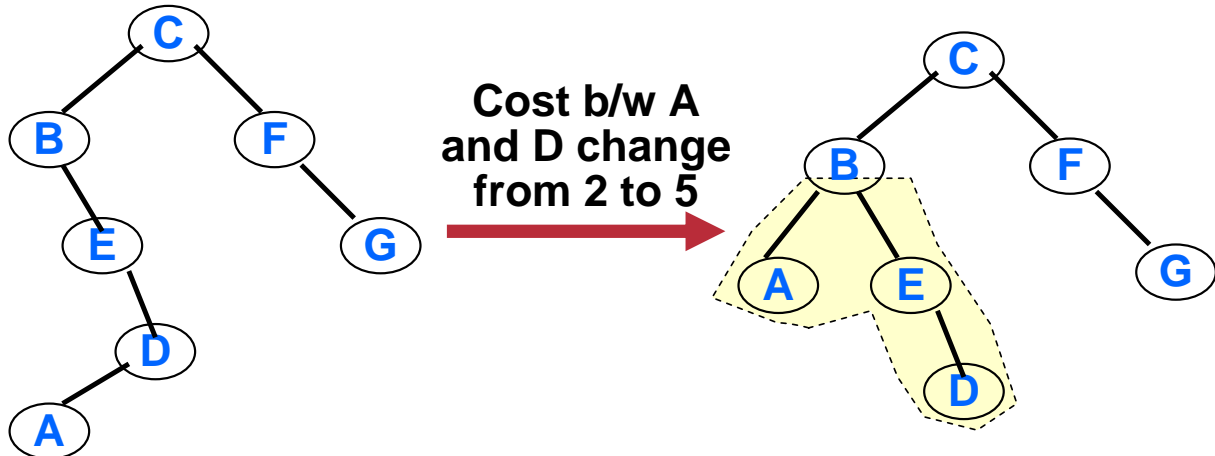
Shortest Path Tree

Cisco.com

Routing
Topology



Shortest
Path Tree
from
node C
view



Incremental SPF – Concept

Cisco.com

- **Incremental SPF (iSPF) allows routers to intelligently figure out where the impact of the change is in the SPT and then only re-computes the effected nodes to update the SPT**
- **As a result, it reduces convergence time by reducing SPF processing time**
- **The amount of convergence time and CPU cycles saved depend on how many nodes that Dijkstra algorithm would need to examine with and without iSPF**

The general rule of thumb is the further away you are from the change the more convergence time is saved

Incremental SPF – Configuration and Deployment

Cisco.com

OSPF Configuration

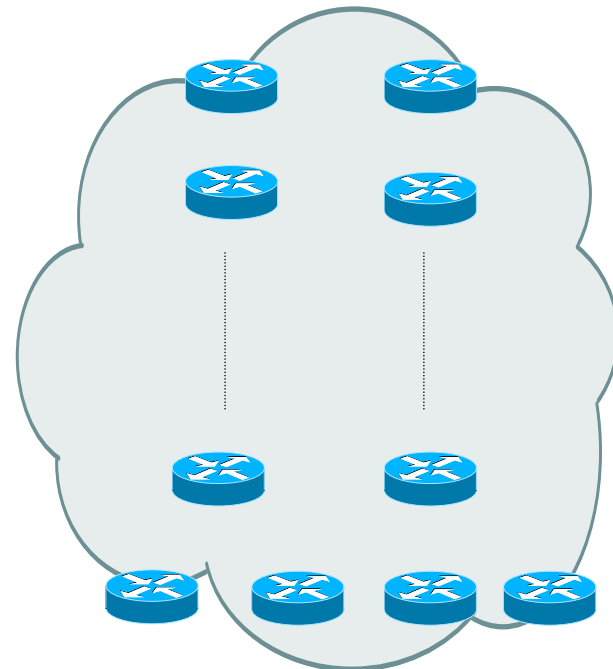
```
router ospf 1  
  [no] incremental-spf
```

ISIS Configuration

```
router isis  
  incremental-spf [level-1|level-2|level-1-2] [<1-100>]
```

Final parameter [<1-100>] is number of full Dijkstra runs which will be performed before incremental runs begin

Ideal for routing area/domain with large number of nodes and/or stub links



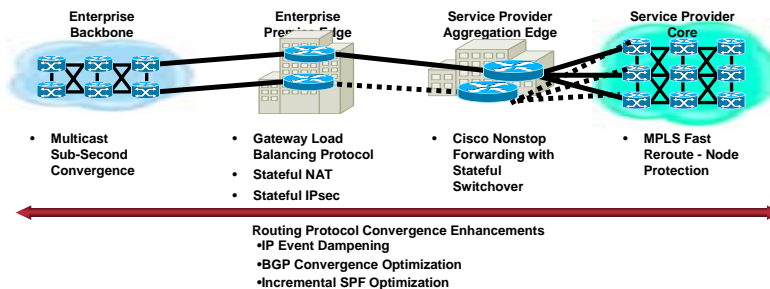
Conclusion

What Is GRIP?

Cisco.com



- **GRIP is a toolbox**
- **To gain benefit, use the right tool for the job**
- **Must be used in combination with good design and operational best practices**



Why GRIP is different

Cisco.com

- “[Cisco] offers the market the most encompassing effort to date...industry technology leadership” – Current Analysis
- Grip distinguishes itself by being a network-wide effort to reduce downtime
- Cisco can provide help at all levels
 - Technology to resolve specific issues
 - Design and deployment advice
 - Operational best practice experience

For More Information

Cisco.com

- **Grip Homepage**

<http://www.cisco.com/go/grip>

- **Mailing List**

`grip-info-ext@cisco.com`

