Transport Diversity: Performance Routing (PfR) Design Guide

Cisco Validated Design I

February 11, 2008

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Text Part Number: OL-15864-01 (EDCS-610117)

Cisco Validated Design

The Cisco Validated Design Program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information visit www.cisco.com/go/validateddesigns.

Transport Diversity: Performance Routing (PfR) Design Guide

CONTENTS

Transport Diversity: Performance Routing (PfR)

Preface

Transport diversity is a general terminology used for selecting or preferring a network exit-point for end-user application traffic across network topologies that have a variety of characteristics. These characteristics include things like monetary cost, reliability or availability, availability of bandwidth, and latency.

One example of transport diversity is a branch office environment that has a primary path using Frame Relay and a backup or alternate path using basic rate ISDN. An example of why the concept of diversity is important is evident in Frame Relay outages that affected over 6,000 customers following a series of events that included a software upgrade of a Frame Relay switch. Enterprise customers who relied solely on Frame Relay for their branch office connectivity may have experienced outages lasting several hours or days. Enterprise customers who deployed branches with a primary link provisioned as Frame Relay and a backup link using basic rate ISDN were able to maintain branch office connectivity throughout the network failure. This WAN diversity is based on decision making based on link failure.

As the WAN technologies advance and mature, the concept of transport diversity also advances to include path selection over ‘always on’ links like Cable, DSL, wireless broadband, or satellite. Now, it is economically feasible to maintain dedicated multiple WAN transport links as there is no variable cost structure or dial-up delay as is the case with ASYNC or ISDN dial.

Performance Routing (PfR) then, is the general term used for features that take into account diverse WAN characteristics and make an informed-decision on the best path to reach a network or application, given multiple choices that may have varied performance characteristics. PfR by its nature takes into account the network performance, delay, loss, and link loading, where traditional routing protocols typically rely solely on cost (total bandwidth) once reachability, in that there is a neighbor relationship between routers, exists across a WAN link.

Interior gateway Protocols (IGPs), particularly Open Shortest Path First (OSPF), uses a simple single metric component, cost, which is based on the bandwidth of the link. Enhanced Interior Gateway Routing Protocol (EIGRP) is slightly more aware of the link characteristics in that it calculates a metric based on cumulative delay (delay is simply an arbitrary assigned value) and the minimum bandwidth value encountered between the source and destination. The only commonly used Exterior Gateway Protocol (EGP), Border Gateway Protocol (BGP), by default uses the number of hops (a hop being all routers within an autonomous system (AS), to determine the best path to the destination network address.

Corporate Headquarters:
Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134-1706 USA

With both IGP and EGP protocols, the concept of transport diversity means equal or unequal cost load-sharing through the use of the routing protocols such as Routing Information Protocol (RIP), EIGRP, or OSPF and through external BGP Multipath (maximum-paths *n*) to insert multiple routes for a destination network address into the routing table.

The concept of load sharing is often associated with the capabilities of a routing protocol, however the routing protocol only serves to inject more than one route into the IP routing table. Once routes are in the routing table, it is the function of the switching path; process, fast, or Cisco Express Forwarding (CEF) to actually accomplish a degree of load sharing or load balancing.

Load balancing is the term used to describe two or more links that are used equally between two sites. However, in order to accomplish an equitable distribution between the two links, per-packet load balancing is usually required to obtain this distribution when the number of flows are small. As an example, consider a file transfer using FTP. With such a single large flow between the two sites, fast or CEF switching uses only one of the links, as the switching path selects an exit based on the destination IP address for fast switching, or for CEF switching, on a per source and destination IP address basis. In either case, only one link is used unless CEF per-packet is enabled.

Tip In most cases, as the number of flows increase between two source and destination networks, so does the ability of any load sharing mechanism to more equally distribute packets across multiple links. Per-packet load sharing can address load sharing with a single or few flows, but at the cost of increasing the likelihood of packets arriving out of sequence, which introduces inefficiencies.

Complicating path selection is the overlay of logical interfaces, IPSec tunnels, for example, which means that path selection must be addressed inside the tunnel. The tunnel destination endpoint may also have multiple paths between source and destination. The *V3PN: Redundancy and Load Sharing Design Guide* (www.cisco.com/go/srnd) was written to assist the network manager in implementing IPSec encryption in the presence of multiple paths or dial-up connections to provide a higher degree of availability. As a general recommendation, load sharing inside the tunnel interface and configuring the tunnel with an affinity to a particular physical interface will provide the best results.

PfR is a technology used to improve on the capabilities of routers and routing protocols to make more granular and intelligent decisions on injecting routes into the routing table so application performance can be optimized to meet the needs of the end-user applications.

Technology Primer

As with any emerging technology, basic features and capabilities are initially implemented in Early Deployment (ED) releases of the Cisco IOS and supported on the most commonly used hardware platforms. As the technology is adopted, customer feedback is used to enhance the capability of the existing features and add new features as well as support additional product lines. Performance Routing (PfR) is no exception to this implementation life cycle.

PfR is Cisco's strategy for advanced route optimization. Optimized Edge Routing (OER) was designed to provide route optimization to destination IP prefixes. PfR leverages OER technology to provide application route optimization and other application services. In this document, references to OER should be in the context of a subset of the broader subject of PfR.

OER was initially targeted at addressing Internet and WAN reliability, addressing the issue where the routing protocol, typically BGP to an Internet service provider (ISP), provides network reachability vectors but does not address transient connectivity failures (brownouts) or offer load-sharing based on measured network performance. Additionally, routing protocols like BGP are not aware of the monetary

cost of links that may incur a per-byte or per-packet basis fee. Some links have both a fixed cost and a variable cost structure. In other words, there may be a monthly charge for the link and some additional charge per-byte or additional charges once some threshold (or usage tier) is reached.

Enterprise customers use the Internet extensively for electronic commerce and often the entire business model is based on sales of product through their Internet portal. The network managers wanted some means of controlling the exit point of their traffic to optimize the network performance for their users but without tools like OER, the solution was to purchase network connectivity from as many ISP networks as practical and hope that the best path to a user was through the ISP that offered the least number autonomous system (AS) paths. With OER, metrics like delay could also be used to determine the best path rather than only rely on the length of the AS path advertised by their respective ISPs.

Tip BGP chooses, by default, the best path based on the fewest AS between the source and destination. OER, on the other hand, can influence traffic based on reachability, delay, loss jitter, throughput, load, monetary cost, and even mean opinion score (MOS).

OER uses various Cisco IOS capabilities, such as NetFlow and IP SLA, to create these advanced metrics for best path selection to improve the user experience.

Design and Implementation Considerations

This section includes an overview of design and implementation considerations the network manager must consider when implementing OER.

General

In any OER implementation, a master controller (MC) and at least one border router (BR) must be configured. The MC commands and controls the BRs and maintains a central repository for the data collected by the BRs. BRs are in the user traffic switching path. BRs collect data from their NetFlow cache and the IP SLA probes they generate, provide a degree of aggregation of this information, and influence the packet switching path to manage user traffic. The MC communicates with the BRs over an authenticated TCP socket, but has no requirement for populating its own IP routing table with anything more than a route to reach the BRs.

Because OER is a path selection technology, there must be at least two external interfaces under the control of OER and at least one internal interface. There must be at least one BR configured. If there is only one BR configured, then both external interfaces are attached to the single BR. If more than one BR is configured, then the two or more external interfaces are configured across these BRs. External links, or exit points, are therefore owned by the BR; they may be logical (tunnel interfaces) or physical links.

The MC function can be collocated (configured) on the same router as the BR, or it can be a dedicated, standalone chassis. The MC is the decision maker. Typically, at a headend campus location, the MC is a standalone chassis while at branch locations the MC is collocated (configured) on the same chassis as the BR. As a general rule, the headend campus location manages more network prefixes and/or applications than a branch deployment and thus consumes more CPU and memory resources for the MC function. Therefore, it makes a good design practice to dedicate a chassis for the MC at the headend campus. The branch typically manages fewer network prefixes and/or applications and due to the costs associated with dedicating a chassis at each branch, the network manager can collocate the MC and BR on the same chassis.

Tip If there are two distinct BRs, only one is configured as the MC. If there are two external interfaces on one branch BR and a third external interface on a separate BR, the MC should be configured on the BR with the two external interfaces. This way, should the BR with the single exit fail, the surviving BR/MC has two functional exits to meet the requirement for at least one internal and two external exits.

Routing Protocol Specific Items

OER can learn prefixes dynamically through the traffic statistics from the NetFlow cache. Both TCP and non-TCP traffic can be learned based on highest throughput. Delay learning is limited to TCP-only traffic, but throughput can be calculated for non-TCP traffic. Network prefixes can be manually defined and learning need not be configured, or prefixes can both be learned dynamically and configured statically. In any of these use cases, a parent route is required to manage a network prefix or application. Parent routes are routes injected into the routing table by either eBGP or static routes which OER then augments with more specific routes (or uses policy-based routing (PBR)) to manage traffic across the external interfaces. Through an assumed definition, the parent routes must therefore be of equal cost and administrative distance so that more than one path for the parent route exists in the routing table of the border router at the same time.

OER learns prefixes that fall under a parent route, the least specific parent route is a default route (0/0), or more specific networks and masks may be configured. For example 10.0.0.0/8 could be used as a parent route. The learning of network prefixes that fall within the parent route is a function of NetFlow. NetFlow is enabled automatically by OER, however it does not appear in the running or startup configuration.

In the current implementation of OER, external BGP or static routes can serve as parent routes with external interfaces being point-to-point or multipoint interfaces (Ethernet) with a single next hop. In other words, multipoint GRE interfaces (as with a DMVPN configuration) that has multiple next hops reachable from the mGRE interface are not supported. Additionally, Ethernet interfaces with multiple next hops, which is a common BGP peering deployment topology, is not currently supported.

IP routing is not required on the MC, it simply must communicate with the BRs. The MC may be protected by firewall or access control lists. The MC and BRs communicate with each other on TCP port 3494 by default, but this is configurable. The MC listens on TCP port 3494 and the BRs initiate the TCP connection.

Details

By default OER manages external interfaces by priority of WAN performance (delay), then loading (utilization). This means, therefore, that one exit point may be more fully used than another, if that exit point exhibits lower (better) application latency (delay) than an other exit point.

Note OER is designed to optimize end-to-end application performance, not simply WAN load balancing. Historically, routing protocols were geared to load sharing across multiple links in hopes of providing better application performance, but load sharing is link (or hop) specific. OER can deduce end-to-end application performance and optimize the exit point to achieve optimal application performance across the internetwork.

In learn mode, delay (and reachability) is determined by observing TCP flows. Round trip delay is determined by the amount of delay observed in TCP flows during session setup; the TCP first two exchanges of the TCP three-way handshake. The client active open is a TCP SYN to the server. In response, the server replies with a SYN-ACK. This level of visibility into TCP flows is obtained by observing flows (through the NetFlow cache) traversing the border routers.

Tip When OER is configured in learn mode/passive, TCP flows must be observed by the border routers to manage prefixes. This means to test OER in a lab environment, some tool to generate actual TCP/UDP traffic and another to introduce delay, loss, etc,. is necessary to observe meaningful results.

Limitations

There are limitations that the network manager must be aware of in order to successfully implement OER. Cisco Express Forwarding (CEF) must be enabled on all border routers. Up to 10 border routers and a total of 20 external interfaces are supported per master controller. If using BGP as parent routes, the border routers must have external BGP neighbors on directly connected interfaces. That neighbor cannot be an iBGP neighbor, although the border router(s) must be iBGP neighbors with other routers in the network to advertise the exit point of the OER managed routes. Static routes are supported as parent routes.

Depending on the BGP configuration, the use of the **maximum-paths 2** command may be required to insert more than one BGP learn network prefix in the routing table. Also, the Cisco IOS hidden command, **bgp bestpath as-path multipath-relax** is also used for this same purpose. This feature is introduced by enhancement CSCea19918 - BGP: need to do multipath with different as-paths.

EIGRP/OSPF learned routes can satisfy the parent route requirement in future Cisco IOS Releases incorporating CSCsk39768 - PfR-EIGRP integration or CSCsm34644 PfR-OSPF integration.

NAT/pNAT compatibility has been added as of the Cisco IOS Release 12.4(15)T; however, NAT/pNAT is not tested in this design guide. The number of network prefixes able to be managed by OER is discussed in the Internet Content Server section.

The use of multipoint interfaces (mGRE) and multiple next hop addresses is not currently supported. The tracking number CSCsi69186 provides additional information regarding future release integration.

tracking number CSCsi69186 provides additional information on multipoint interfaces (mGRE) and multiple next hop addresses.

Passive and Active Monitoring

Passive monitoring is the act of OER gathering information on user packets assembled into flows by NetFlow. OER, when enabled, automatically enables NetFlow on the managed interfaces on the border routers. By aggregating this information on the border routers and periodically reporting the collected data to the master controller, the network prefixes and applications in use can automatically be learned. Additionally, attributes like throughput, reachability, loading, packet loss, and latency can be deduced from the collected flows.

Active monitoring is the act of generating IP SLA probes to generate test traffic for the purpose of obtaining information regarding the characteristics of the WAN links. Active probes can either be implicitly generated by OER when passive monitoring has identified destination hosts, or explicitly configured by the network manager in the OER configuration. An example of configuring an explicit IP SLA jitter probe is shown in Branch/SOHO VPN Deployment, page 64.

Reachability Must Be Verified

For OER to consider an exit interface as a candidate for traffic, reachability to the target network prefix must be verified. When OER is configured as passive mode (mode monitor passive), TCP flows must be present across the exit interface to learn the validity of reachability across the exit. Note that a parent route needs to be present to direct traffic for a target network out the external interfaces, in order to allow the NetFlow subsystem to identify the validity of reachability through the TCP flows. Given this, if there is no TCP traffic out an exit interface, no passive measurements are available to NetFlow/OER. Or, if there are long lived TCP flows, flows lasting longer than the OER monitor period, no TCP SYNc and TCP SYNc/ACK are seen during the monitor period. So in this case, traffic may be active, but because the TCP SYNc and TCP SYNc/ACK is not seen during the monitor period, no delay and reachability can be deduced from this long persistent flow.

Tip Passive monitoring of delay, loss, and reachability rely on OER observing the NetFlow reported TCP traffic over an exit interface. OER can learn prefixes based on throughput for non-TCP flows. Excluding VoIP, which is UDP-based, TCP-based applications represents the largest share of traffic on the Internet and most enterprise networks.

For OER to function optimally in passive monitor mode, more TCP flows equate to more data points for the master controller to analyze and manage. As the number of TCP flows increase, the database becomes more granular, meaning more delay, loss, and reachability information is available for a given network prefix.

Passive Mode Example illustrates the need for traffic to be observed by NetFlow over more than one exit interface when mode monitor passive is configured.

Sup720/RSP720 (Earl7) Limitations

Because of architectural limitations with the NetFlow implementation on the EARL 7 (PFC3)-based hardware present on supervisor engines of the 6500/7600 series, OER cannot determine performance (delay/loss/reachability) characteristics from passive monitoring of TCP flows. Passive throughput is supported by Earl7. Throughput is the calculation of the number of packets output from the external interfaces of the OER border routers over a unit of time, usually represented as a rate per second (as in megabits per second). Therefore, throughput is synonymous with using OER to manage for load sharing.

More information regarding this limitation is available in Internet Content Server section of this document.

Authentication

Communication between the master controller and border routers must be authenticated through a referenced key-chain and they share a like key-string. In the following configuration examples, assume that the border router and master controller, either collocated on the same chassis or on separate chassis, reference a key chain in their respective configuration files that share a like key-string. An example follows:

```
!
! Example of key chain with master controller
! and border router on the same device
!
interface Loopback0
 ip address 10.0.0.1 255.255.255.255
!
!
key chain BLUE
 key 10
   key-string 7 0035262F277034241D2E5B40
!
!
oer master
 !
 border 10.0.0.1 key-chain BLUE
!
oer border
 master 10.0.0.1 key-chain BLUE
!
end
```

Warning **OER authentication fails with a key string greater than 15 bytes. See CSCsd00633 for more details.**

Process Flow

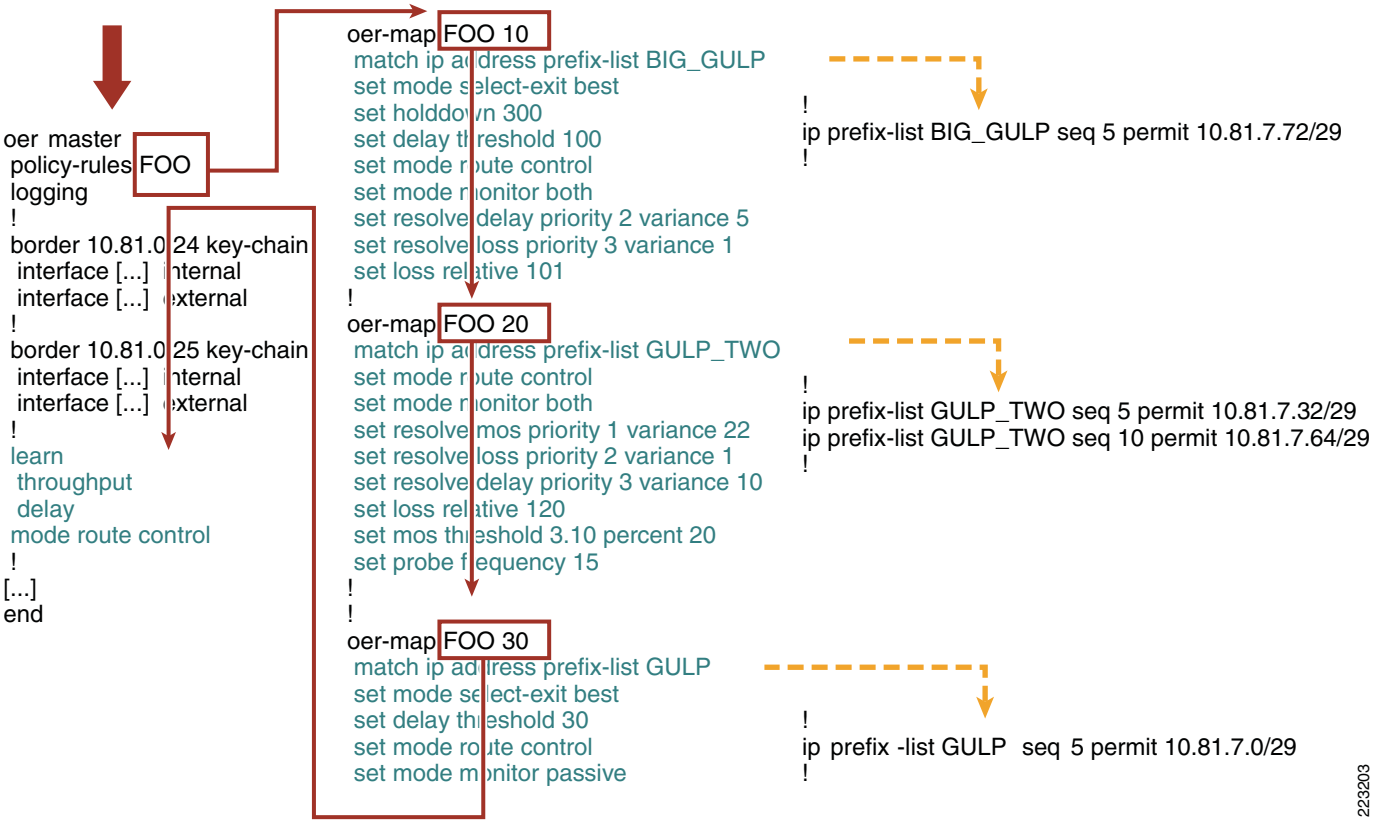
The OER configuration section in the Cisco IOS command line interface provides a means to be very granular in selecting the types of applications or network prefixes are targeted for performance routing. Additionally, the policy associated with each application or network prefix can differ from the default policy and is unique and specific to the network prefixes or applications identified in the configuration.

To better understand this process, one sample configuration is provided (see Figure 1) to demonstrate how a network manager may configure a master controller policy to identify four remote branch networks and apply different policies through three separate OER maps. Note that the second OER map identifies two network prefixes while the first and third identify a single prefix.

The OER master configuration section is parsed through the policy-rule reference to OER map FOO. The sequenced references to FOO are parsed and the distinct policy is associated with the selected addresses identified in the prefix-list. Upon completion of parsing the oer-maps, the remainder of the global OER master configuration is parsed. In this case, prefix learning (*learn* is referenced under the oer *master* construct) is configured, meaning that this OER master configuration will both identify network prefixes based on explicitly configured address as well as through learning prefixes based on traffic identified by NetFlow.

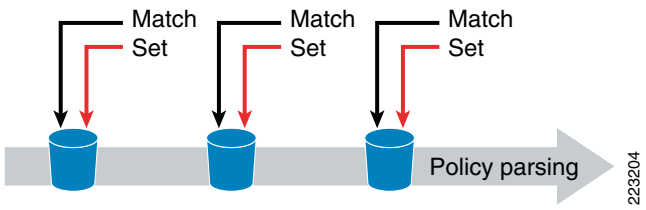
Figure 1 demonstrates this process flow.

Figure 1 Process Flow for OER Configuration



Looking at this graphically, An OER policy is analogous to a container or bucket to hold match and set statements, which are evaluated in order of the OER map sequence numbers and then when the OER map is completely evaluated, the global OER master configuration statements are evaluated. This is shown in Figure 2.

Figure 2 OER Policy



The policies in effect can be shown by using the **show oer master policy** command. Additionally, the **show oer master prefix** *n.n.n.n/n* **policy** command can be used to display the policy in effect for a particular prefix. An example of the output of this command is shown in Displaying the Policy for a Prefix, page 21.

Principles of Operation

This section examines the principles of operation for the OER sub-system within the Cisco IOS and how it interacts with other subsystems including the IP routing table, BGP, NetFlow, and IP SLA.

Routing Protocol Interaction

This section describes how OER functions in a basic configuration using passive monitoring and prefix learn mode. This is the simplest means to enable OER, and it relies on NetFlow data of TCP sessions to provide this function.

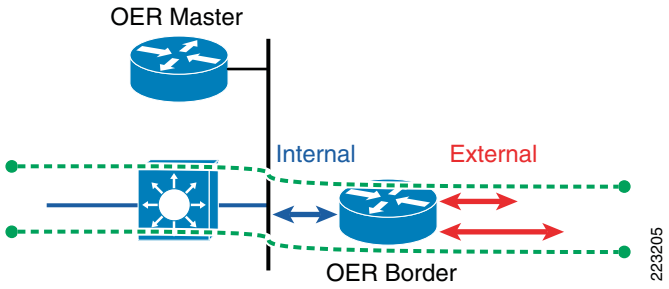
OER can use both static routes and BGP as the method to provide parent routes, each method is shown. OER is configured to control routes (route control mode), rather than simply to observe.

Static Routing

First, look at a simple configuration in Figure 3 where there is one OER border router with a single internal interface and two external interfaces. The OER master controller is shown as a separate router in this configuration, but it could also be also configured on the same chassis as the OER border router.

A single campus switch is shown in the topology (see Figure 3). Because both exits are on the same chassis, the Layer 3 campus switch routes all packets to the OER border router, allowing it to make the exit interface decision off its own IP routing table. In this example, OER is simply influencing static routes in the IP routing table and these statics do not need to be redistributed into an IGP as there is only one Layer 3 campus switch in the topology.

Figure 3 Static Routing



The principles of operation for OER in this topology are described as follows:

- Parent route, static routes with a destination of the external interfaces, are injected into the IP routing table as equal cost routes to the destination network(s). These routes are manually configured and present in the startup/running configuration.
- IP CEF switches user traffic (packets) using these equal cost parent routes out the OER external interfaces. CEF switching is enabled by default and is required for OER to function.
- NetFlow, enabled automatically and transparently by OER, captures the resulting flow data from packets using the exit points.
- The OER border router reports this learned flow data to the OER master for analysis.
- When the OER master controller detects traffic out-of-policy, it instructs the OER border router to inject a static route directly to the IP routing table. This directs out-of-policy traffic through a new path to reach the destination network.

- By default, OER injects the static into the IP routing table as a /24 network prefix. This length is configurable. However, the key point is that OER is influencing network traffic through a prefix with a longer mask than the parent route. For example, route control of /24 prefixes maybe sufficient for Internet load-sharing policies, but /24 is too short for branch office load-sharing if the branch has a /24 or longer subnet design.

The following output illustrates the relationship between the OER master prefix database and the IP routing table. There are two parent routes in the routing table; 10.0.0.0/8 and 64.102.0.0/16.

```
ip route 10.0.0.0 255.0.0.0 10.81.7.225 30 tag 300 name OER_parent
ip route 10.0.0.0 255.0.0.0 10.81.7.193 30 tag 300 name OER_parent
ip route 64.102.0.0 255.255.0.0 Tunnel200 tag 300 name OER_parent
ip route 64.102.0.0 255.255.0.0 Tunnel100 tag 300 name OER_parent


joeking-vpn-1811#show ip route static
     64.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
S       64.102.0.0/16 is directly connected, Tunnel200
                      is directly connected, Tunnel100
S       64.102.223.16/28 [1/0] via 192.168.2.1
     10.0.0.0/8 is variably subnetted, 4 subnets, 3 masks
S       10.0.0.0/8 [30/0] via 10.81.7.225
                   [30/0] via 10.81.7.193
```

This output identifies an OER prefix that is currently being controlled by OER and is in the state of INPOLICY. In this example, 10.16.151.0/24 is used.

```
joeking-vpn-1811#show oer master prefix learned
OER Prefix Statistics:
 Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
 P - Percentage below threshold, Jit - Jitter (ms),
 MOS - Mean Opinion Score
 Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
 E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
 U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
 # - Prefix monitor mode is Special, & - Blackholed Prefix
 % - Force Next-Hop, ^ - Prefix is denied

Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS
--------------------------------------------------------------------------------
...
joeking-vpn-1811#show oer master prefix learned | inc INPOLICY
64.102.16.0/24          INPOLICY        0 10.81.7.73      Tu100           STATIC
64.102.4.0/24           INPOLICY        0 10.81.7.73      Tu200           STATIC
64.102.6.0/24           INPOLICY        0 10.81.7.73      Tu100           STATIC
10.16.151.0/24          INPOLICY        0 10.81.7.73      Tu100           STATIC
64.102.31.0/24          INPOLICY        0 10.81.7.73      Tu100           STATIC
```

From the above display, the protocol specified is **static**, meaning a static route has been injected into the IP routing table. Reviewing the IP routing table:

```
joeking-vpn-1811#show ip route 10.16.151.0 255.255.255.0
Routing entry for 10.16.151.0/24
  Known via "static", distance 1, metric 0
  Tag 5000
  Routing Descriptor Blocks:
  * 10.81.7.225
      Route metric is 0, traffic share count is 1
      Route tag 5000
```

Note that the next hop is identified from the value specified for the parent route (next hop is IP address 10.81.7.225) and that its route tag is 5000. OER by default uses a route tag value of 5000.

Note OER injects static routes into the running configuration. They are not in the startup Cisco IOS configuration.

Looking at all static routes in the IP routing table, there are two OER parent routes, the route to 64.102.0.0/16 and 10.0.0.0/8. Note that in the first case, the next hop is specified by the logical interface name (Tunnel200 and Tunnel100) and for the second case, the next hop is specified by IP address.

```
joeking-vpn-1811#show ip route static
     64.0.0.0/8 is variably subnetted, 7 subnets, 3 masks
S       64.102.6.0/24 [1/0] via 0.0.0.0, Tunnel100
S       64.102.4.0/24 [1/0] via 0.0.0.0, Tunnel200
S       64.102.0.0/16 is directly connected, Tunnel200
                      is directly connected, Tunnel100
S       64.102.19.0/24 [1/0] via 0.0.0.0, Tunnel100
S       64.102.16.0/24 [1/0] via 0.0.0.0, Tunnel100
S       64.102.31.0/24 [1/0] via 0.0.0.0, Tunnel100
S       64.102.223.16/28 [1/0] via 192.168.2.1
     10.0.0.0/8 is variably subnetted, 5 subnets, 4 masks
S       10.0.0.0/8 [30/0] via 10.81.7.225
                   [30/0] via 10.81.7.193
S       10.16.151.0/24 [1/0] via 10.81.7.225
```

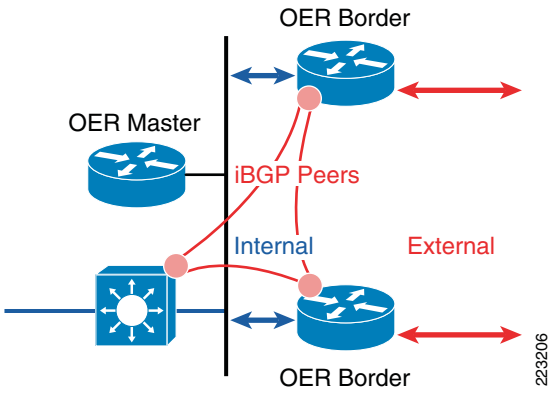
Now with this /24 route in the IP routing table, user traffic for 10.16.151.0 is directed out one of the two exits.

BGP Routing

Using BGP as a source for parent routes is also an option. From the principles of operation for OER description in the previous section, only one line item is changed. That item relates to OER injecting a static route in the routing table to influence the overall path selection. When BGP is configured, OER injects a network prefix and mask into the BGP table, not the IP routing table. In turn, these BGP routes are advertised to the other BGP routers and BGP routes are injected into the routing table through the BGP selection process.

In Figure 4, the topology is modified slightly to have two border routers, each with one external interface.

Figure 4 OER and BGP Routing



These OER border routers are external BGP (eBGP) peers with their respective ISP, while the Layer 3 campus switch and the two OER border routers are iBGP peers. Where in the previous example, the Layer 3 campus switch needed no dynamic routing protocol as all packets were forwarded to the single OER border router. Now the iBGP session between the Layer 3 campus switch and the two OER border routers is used to advertise the OER managed prefixes injected into the BGP table, not directly into the IP routing table, to influence a subset of the total traffic. The BGP routing process then scans the BGP table and inserts routes from the BGP table into the IP routing table of both OER border routers as well as into the Layer 3 campus switch.

Note In this topology, OER could use static parent routes and redistribute the OER static routes that are injected into the IP routing table of the OER border routers into some dynamic IGP routing protocol, like OSPF, RIP, or EIGRP. This example, however, shows the use of BGP as parent routes so that is the nature of the example.

To reiterate, the eBGP sessions provide OER Parent routes, their existence in the IP routing table, along with IP CEF, NetFlow and the reporting by the OER border routers to the OER master controller cause the master controller to direct the border router to inject routes into the BGP table. In turn, these entries in the BGP table are advertised to the configured iBGP peers, and then potentially injected into the IP routing table of these peers.

Internal BGP (iBGP) is therefore the means to influence path selection upstream from the OER border router. In this example, the upstream device(s) is the Layer 3 campus switch.

The following is a sample of a prefix (192.168.192.0/24) that is injected into the IP routing table through BGP.

```
vpn-jk2-3725-1#show oer master prefix
OER Prefix Statistics:
 Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
 P - Percentage below threshold, Jit - Jitter (ms),
 MOS - Mean Opinion Score
 Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
 E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
 U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
 # - Prefix monitor mode is Special, & - Blackholed Prefix
 % - Force Next-Hop, ^ - Prefix is denied
```

```
Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS
------------------------------------------------------------------------------
192.168.17.0/24         INPOLICY      @36 192.168.131.2   AT3/0.135       BGP

                              U        U        0        0        0        0
                              9        9        0        0        1        1
                              N        N
192.168.33.0/24         INPOLICY       83 192.168.131.1   AT2/0.235       BGP

                              U        U        0        0        0        0
                              4        4        0        0        1        1
                              N        N
192.168.193.0/24        HOLDDOWN       56 192.168.131.2   AT3/0.135       BGP

                              U        U        0        0        0        0
                              U        U        0        0        0        0
                              N        N
192.168.192.0/24        INPOLICY       46 192.168.131.1   AT2/0.235       BGP

                              U        U        0        0        0        0
                              U        U        0        0        0        1
                              N        N
```

This prefix is displayed from the BGP table:

```
vpn-jk2-3725-1# show ip bgp 192.168.192.0/24
BGP routing table entry for 192.168.192.0/24, version 228
Paths: (1 available, best #1, table Default-IP-Routing-Table, not advertised to
EBGP peer)
  Advertised to update-groups:
        1
  65001 65002, (injected path from 192.168.192.0/18)
    192.168.129.5 from 192.168.129.5 (192.168.191.1)
      Origin IGP, localpref 100, valid, external, best
      Community: no-export
vpn-jk2-3725-1#
```

Note OER injected routes remain local to this AS as they have a community value of no-export, meaning do not advertise this route to EBGP peers.

And, in this example, the route is also injected into the IP routing table by the BGP process:

```
vpn-jk2-3725-1#show ip route bgp
B    192.168.192.0/24 [20/0] via 192.168.129.5, 00:07:49
...

vpn-jk2-3725-1#show ip route 192.168.192.0 255.255.255.0
Routing entry for 192.168.192.0/24
  Known via "bgp 65030", distance 20, metric 0
  Tag 65001, type external
  Redistributing via eigrp 100
  Advertised by eigrp 100 route-map ELIMINATE_RIB_failure
  Last update from 192.168.129.5 00:06:29 ago
  Routing Descriptor Blocks:
  * 192.168.129.5, from 192.168.129.5, 00:06:29 ago
      Route metric is 0, traffic share count is 1
      AS Hops 2
      Route tag 65001
```

```
vpn-jk2-3725-1#
```

Now that the method of OER influencing traffic has been shown, the next section explores how OER then verifies and re-evaluates on an ongoing basis.

Operational Modes

Mode Monitor Passive

The border routers report traffic flows identified by NetFlow to the master controller. The average delay, of the flows, packet loss, and reachability along with the outbound throughput in terms of bits per second is determined for the destination IP prefixes observed in the NetFlow data.

Measurements of the TCP traffic flows is characterized by:

- Delay—Time between TCP SYNC and TCP SYNC/ACK in a TCP three-way handshake.
- Loss—TCP sequence numbers are tracked, loss can estimated when lower sequence numbers than the highest sequence number observed are seen.
- Reachability—Repeated TCP SYNCs without an accompanying TCP SYNC/ACK identify reachability failures.
- Throughput—Throughput is calculated from NetFlow and measured in bits per second (bps).

Measurements of non-TCP traffic flows is characterized by throughput only.

Mode Monitor Active

In this mode, Cisco IOS IP service level agreements (SLAs) probes are generated by the border routers and transmitted at the configured probe frequency value. Active probes are created implicitly by OER; however, the network manager may also explicitly create active probes.

By default, an active probe is of the type of ICMP echo. If VoIP is to be characterized, the network manager may choose to explicitly configure an active probe. Following is an example from an oer-map using a traffic-class that matches VoIP streams.

```
set active-probe jitter 10.1.1.1 target-port 33033 codec g729a
!
set probe frequency 2
```

In this example, the target IP address configured in the explicit active probe, 10.1.1.1 in this example must be a Cisco router configured for **ip sla responder** command. Most IP hosts will respond to an ICMP echo, unless administratively disabled or prohibited, however to determine MOS, jitter and other characteristics associated with VoIP quality measurements, the capabilities and function of an IP SLA responder must be enabled.

It is possible, and practical, to use active monitoring for specific traffic-class or IP prefix address, identified through an OER-MAP referenced by a **policy-rules** statement for measuring VoIP traffic, while using a global configuration option defaulting to passive monitoring of all other traffic through the TCP flows.

Note An example of using both active monitoring for VoIP and passive monitoring for the remaining traffic flows is shown in Displaying the Policy for a Prefix, page 21.

Mode Monitor Both

Mode monitor both is the default value and combines the capabilities of passive and active monitoring. Up to five IP addresses are actively probed for each destination prefix learned through passive monitoring. By default, an IP SLA ICMP ECHO probe is automatically generated for the learned IP addresses.

By monitoring both actively and passively, additional data points regarding a network prefix can be obtained through two separate and distinct tools; NetFlow for passive measurements and IP SLA for the active measurements. However, the inclusion of active probing also has disadvantages. The ICMP ECHO requests that are generated by default constitute additional background traffic on the network. When used on the Internet, activating probing may not be desirable in that ICMP packets may be blocked or administratively prohibited and may be considered a threatening or abusive posture to the target hosts. Because of this, *mode monitor both* is best suited for use within the private internal network of the enterprise. Unlike *mode monitor fast*, which is described in a later section, active probing does not probe all exit points continuously. It probes only the current exit point provided the status is INPOLICY and probes are generated after the prefix timer value is exhausted.

To illustrate, prefix 192.168.33.0/34 is being monitored by both passive and active probing. In looking at the detail display of the prefix, several items bear notice:

- State of INPOLICY*—The asteric (*) indicates this prefix is uncontrolled by OER (the parent route controls routing) , but is currently inpolicy.
- The ‘at sign’ (@) on the Time Remaining value means the prefix is being actively probed. The numerical value is a countdown timer indicating when this state will expire.
- The latest statistics from the active probes, by the five individual IP addresses are shown, along with the corresponding values. Note that each of the five target IP addresses were attempted two times, and successfully completed each attempt. The sum of the delay values, along with the minimum, maximum, and derived average delay (Dly) is shown:

```
vpn4-3800-15#show oer master prefix 192.168.33.0/24 detail
Prefix: 192.168.33.0/24
   State: INPOLICY*   Time Remaining: @11
   Policy: Default

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
  *192.168.131.1   Gi0/1.651              5        5        0        0
   192.168.131.2   Gi0/1.652              6        6        0        0

   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min      Max      Dly
   echo     192.168.33.5        N     2     2       7       3        4       3
   echo     192.168.33.6        N     2     2      20       8       12      10
   echo     192.168.33.25       N     2     2       2       1        1       1
   echo     192.168.33.16       N     2     2       8       4        4       4
   echo     192.168.33.26       N     2     2       8       4        4       4

Prefix performance history records
 Current index 50, S_avg interval(min) 5, L_avg interval(min) 60

Age       Border          Interface       OOP/RteChg Reasons
Pas: DSum  Samples  DAvg  PktLoss  Unreach   Ebytes   Ibytes     Pkts    Flows
Act: Dsum Attempts  DAvg    Comps  Unreach   Jitter LoMOSCnt   MOSCnt
00:00:55  192.168.131.1   Gi0/1.651
       36        6     6        0        0    13268    18783      287       76
        0        0     0        0        0        N        N        N
00:02:11  192.168.131.1   Gi0/1.651
```

```
24          5     4          0          0    10472    15276       237        66
 0          0     0          0          0        N        N         N
```

In this example, also note that both passive and active data points are shown collectively in the output for the most recent data statistics as well as the performance history section. The performance history section can be used to see trends in the characteristics of a given prefix.

Mode Monitor Special

Mode monitor special is an alternate syntax to *mode monitor both* for the Cisco 6500 and Cisco 7600 series implementations. Active probing is enabled to accommodate the EARL 7 (PFC3) passive monitoring limitations described in a previous section.

Mode Monitor Fast

This feature was introduced in Cisco IOS Release 12.4(15)T as a key component to the Fast Reroute feature. This mode generates active probes through all exists continuously at the configured probe frequency. This differs from either *active* or *both* mode in that these modes only generate probes through alternate paths (exits) in the event the current path is out-of-policy. One way to describe this behavior is the OER subsystem quantifies the alternatives only when the current path is known to be deficient, where with Fast Reroute, the characteristics of the alternative paths are always known, allowing immediate use as required. If unreachable is determined to be out-of-policy for the current exit, the alternate exit is selected as the current exit, assuming the unreachable values for the alternate exit is in policy.

The unreachable threshold is calibrated in number of failed probes per million probe attempts. If the unreachable value is set to 1, a single probe fails on the current exit, an attempt is made to locate a alternate exit. However, if the alternate exits also have a single failed probe, they are not selected because they too are out-of-policy.

Warning **Setting the unreachable threshold to a value of 1 may cause an alternate exit to be out-of-policy in the event a transient error occurred in the past, but which has now cleared.**

The Fast Reroute feature, therefore, allows rerouting actions to be taken, at an interval approaching the configured probe frequency value. Probe frequency can now be set as low as 2 seconds if fast mode is configured. This allows re-routing at slightly more than the configured probe frequency value. While the Fast Reroute feature was not scale tested in this design guide, in an ideal deployment, rerouting can occur in as little as 3 seconds.

Note This feature may be best described as **continuous monitoring of alternate paths**, as opposed to **as required monitoring of alternate paths**.

The obvious drawback to this feature is the potential for adding additional network traffic overhead associated with the probes themselves and additional CPU resources to the OER border routers, the source of the active probes. Unless the prefix is deleted or in the default state, probes are generated.

The active probe results are used for out-of-policy and to control routing. Passive data collected is for information only, the throughput transmit and receive Kbps values (**show oer mast border detail**) are used for load balancing.

Network Prefix States

Default

A network prefix may be shown in the default state if it is manually configured or learned but has not been determined to be in or out-of-policy. Prefixes may revert back to default state if, for some reason, OER can no longer control the prefix. This may happen if all the exits are out-of-policy.

The default state means that the parent IP routes control the exit for this destination prefix. This would be the same behavior as if OER were not configured or shutdown.

Inpolicy

The prefix is *inpolicy*, which means that it meets the policy associated with this prefix or application. The prefix can be *inpolicy* and being controlled by OER, or *inpolicy** and not controlled by OER. The presence of the asteric (*) on the state attribute indicates the network is known to OER, but is under the control of the parent route. When no asteric (*) is present, the prefix is being controlled by OER. The state of *inpolicy* is considered to be a desirable state.

Out-of-Policy (OOP)

The prefix or application has been identified as failing to meet its respective policy. If traffic is identified as being out-of-policy, OER moves the traffic to an alternative exit to bring the traffic inpolicy or unmanages the traffic, allowing it to revert back to the default exits as determined by the parent routes in the IP routing table. If the traffic reverts back to the default state, OER will again cycle this traffic, like all other traffic on the network, in an attempt to optimize based on the configured or default OER policy. The state of out-of-policy is considered undesirable.

Holddown

The holddown state is enabled when a traffic class is initially controlled by OER. This holddown concept is applied to prevent churning or erratic behavior of OER managed routes from being injected and withdrawn from the IP routing table (and subsequently being redistributed by some IGP) or BGP tables.

Once a prefix has been changed, it enters holddown for the specified (holddown) period before it can be deemed in or out-of-policy. A network prefix can leave holddown state before the timer expires if the current exit point experiences an unreachable out-of-policy condition. All other out-of-policy conditions are ignored during holddown state.

Key Concepts

This section provides an overview to some of the terminology and concepts of OER

Variance

The concept of variance in OER is similar in implementation to the variance keyword that enables EIGRP to install multiple unequal cost routes in the local routing table. Variance is a means to specify a range in which two unequal values are considered similar enough to be treated as equal.

From the context of OER, variance is a percentage, from 1-100. If delay is set to an absolute value of 80ms and a 10 percent variance is configured, delay values from 80 to 88ms will be considered equal.

Path Selection

OER selects the best path based on:

- Excluding links currently overloaded (refer to Max BW from the output of the **show oer master border detail** command).
- Best performing link depending on configured priorities and their associated variance.

Granularity

Without sufficient granularity, meaning the number of flows and prefixes being learned or configured, OER cannot effectively do optimal load balancing. This is also true of CEF or fast switching when load balancing two equal cost paths in the IP routing table. CEF has an advantage over fast switching in that CEF load balances based on source and destination IP address, while fast switching only load balances based on destination IP address. However, OER has an advantage over CEF or fast switching in that it can load balance based on Layer 4 fields or ToS byte (DSCP) instead of simply network (destination) prefixes, to provide better visibility and granularity. OER also takes into consideration link utilization, where CEF does not. From a campus headend, granularity may not be an issue; however, it may be from the branch router.

Tip OER performs most effectively with the more flows and the resulting network prefixes it has observed. If the number of destination network prefixes are low, consider adding more granularity by monitoring applications instead of simply network prefix addresses.

Interval Period

The configured interval period value determines ***how often*** traffic is analyzed.

Monitor Period

The configured monitor period determines for ***how long*** traffic is measured before being reported by the border router to the master controller. This is the means of specifying the learning interval. The default is 5 minutes. Flows are aggregated on the border router during this interval. At the end of the interval, the top (***prefixes*** keyword value, subordinate to the **learn** command) prefixes based on throughput are reported to the master controller.

Loss

Packet loss is based on packets per million (PPM) regardless of how many hosts are involved, and loss is based on both passive and active monitoring; however, with active monitoring, loss is reported only for jitter probes. Loss is specified as a relative percentage or maximum number of packets.

Note If the fast re-route feature is implemented to support voice or video over IP and packet loss is one criteria desired to trigger the reroute, then an explicitly configured jitter probe is required.

Unreachable

Unreachable is based on flows per million (FPM). Unreachable hosts only apply to TCP sessions. Reachability failures are determined by TCP SYNCs without an accompanying TCP SYNC/ACK. Unreachable can either be an absolute maximum number or a relative percentage.

Feature Summary

Table 1 lists the features and the release train implemented.

Table 1 Implemented Features

Release	Feature
12.4(6)T	Voice Traffic Optimization
12.4(9)T	DSCP Monitoring BGP Inbound Optimization
12.4(11)T	Dynamically Learned Well Known Applications
12.4(15)T	Link Grouping, Fast Re-route, NAT/pNAT
12.2(33)SRB	OER on 7600
12.2(33)SXH	OER BR on 6500

Best Practices, Tips and Techniques

This section demonstrates useful commands, best practices and other tips and techniques to assist the network manager in deploying and maintaining OER in a production environment.

Load Interval and Bandwidth

To provide the most granular and accurate information to the master controller, configure the load-interval on internal and external interfaces on the border routers to the minimum value of 30 seconds. Additionally, the bandwidth statement on the interface should also be appropriately configured.

```
interface Tunnel100
 bandwidth 256
 load-interval 30
```

```
joeking-vpn-1811#show oer mast bor det
Border           Status   UP/DOWN             AuthFail  Version
10.81.7.73       ACTIVE   UP       00:32:08          0  2.0
 Tu200           EXTERNAL UP
 Tu100           EXTERNAL UP
 Vl1             INTERNAL UP

 External         Capacity      Max BW   BW Used    Load Status          Exit Id
 Interface         (kbps)       (kbps)    (kbps)    (%)
 ---------        --------      ------   -------- ------- ------           ------
 Tu200                 256         204         74      28 UP                    2
                                   192          0       0
 Tu100                 256         192         67      25 UP                    1
                                   192         91      35
!
```

Note The above display was captured while a file transfer was executing an FTP PUT through Tunnel 200 and a VoIP call was active on Tunnel 100. This accounts for the display showing bidirectional data on Tunnel 100, but primarily unidirectional data on Tunnel 200.

The Max BW value is derived from the default value of 75 percent or the configured value. For Tunnel 200 80 percent of 256K is 204K, and for Tunnel 100, the default value of 75 percent (which is not shown in the configuration) is represented as 192Kbps. This display was captured from a router using the following configuration:

```
oer master
 policy-rules BRANCH
 logging
 !
 border 10.81.7.73 key-chain GREEN
  interface Tunnel200 external
   max-xmit-utilization percentage 80
  interface Tunnel100 external
  interface Vlan1 internal
 !
```

The *max-xmit-utilization* value is to bound the path selection algorithm. Links that are currently overloaded (links that have loading that exceeds the maximum bandwidth value) are removed from consideration for selecting the best path.

Displaying the Policy for a Prefix

This command displays the policy in effect for a particular prefix:

```
vpn-jk2-3725-1#show oer master prefix 192.168.193.0/24 policy
Default Policy Settings:
  backoff 90 3000 300
  delay relative 50
  holddown 300
  periodic 180
  probe frequency 56
  mode route control
  mode monitor both
  mode select-exit best
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  resolve delay priority 11 variance 20
  resolve utilization priority 12 variance 20
 *tag 0
```

Note If the output of **show oer master prefix** command is null, then that prefix has not been learned or configured.

Active and Passive Combined

This example also demonstrates a configuration for OER Fast re-route. This feature is introduced in Cisco IOS Release 12.4(15)T:

```
!
hostname vpn-jk2-3725-1
!
! System image file is "flash:c3725-advipservicesk9-mz.124-15.T"
!
key chain GREEN
 key 10
   key-string 7 11283B263343595F500F0D03
!
```

```
!
oer master
 policy-rules ENTERPRISE_CAMPUS
 logging
 !
 border 192.168.131.1 key-chain GREEN
  interface ATM2/0.235 external
  interface FastEthernet0/1.100 internal
  interface FastEthernet0/1.102 internal
 !
 border 192.168.131.2 key-chain GREEN
  interface ATM3/0.135 external
  interface FastEthernet1/0.100 internal
  interface FastEthernet1/0.102 internal
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  expire after time 30
 backoff 90 3000 300
 mode route control
 mode select-exit best
 periodic 180
 !
!
!
oer border
 logging
 local FastEthernet0/1.100
 master 192.168.131.1 key-chain GREEN
!
!
ip access-list extended VOICE
 permit udp any 10.0.0.0 0.255.255.255 dscp ef
 permit udp any 10.0.0.0 0.255.255.255 dscp af41
 permit udp any 10.0.0.0 0.255.255.255 dscp cs5
!
! For each branch you would need one map entry (sequence no.) because we
! are manually configuring the probe destination IP address.
!
oer-map ENTERPRISE_CAMPUS 10
 match traffic-class access-list VOICE
 set holddown 300
 set delay threshold 150
 set mode route control
 set mode monitor fast
!
! The order in priority is jitter, delay then MOS
!
 set resolve jitter priority 1 variance 10
 set resolve delay priority 2 variance 10
 set resolve mos priority 10 variance 10
 set jitter threshold 15
 set mos threshold 4.00 percent 15
 set active-probe jitter 10.1.1.1 target-port 33033 codec g729a
!
```

Tip IP SLA responder must be configured on the target router at 10.1.1.1.

```
!
 set probe frequency 2
!
end
```

Solution Overview

This solution is comprised of the following deployment models:

- Internet Content Server, page 25
- WAN Hub: Dual MPLS Service Providers, page 50
- Branch/SOHO VPN Deployment, page 64
- Branch VPN Deployment with Cisco Wide Area Application Services (WAAS), page 76

The deployment models are described and documented each in their own section; however, there are some similarities across the sections.

- The Internet Content Server section focuses on master controller scalability.
- The WAN Hub: Dual MPLS Service Providers section focuses on border router scalability, but the master controller scalability findings are applicable to both deployments.
- The Branch VPN Deployment with Cisco Wide Area Application Services (WAAS) section builds on the topology and results described in theBranch/SOHO VPN Deployment section.
- In the Internet Content Server and the Branch VPN Deployment with Cisco Wide Area Application Services (WAAS) sections, a standby master controller configuration is tested and documented.

Internet Content Server

This represents an Internet edge deployment with two or more ISP links receiving full Internet routing table advertisements. The remote users are unknown individual user clients accessing web hosting servers. As the bulk of the traffic is from server-to-client, OER is used to control only routing to the Internet. The OER configuration deployed is simple passive monitoring of TCP traffic and dedicated chassis for the control function.

The goal is to obtain the scalability limits of managing large number of IP network prefixes to manage user traffic. Because of architectural limitations with the NetFlow implementation on the EARL 7[1] (PFC3)-based architectures, OER cannot deduce performance (delay, loss, reachability) characteristics from passive monitoring of TCP flows. Also many Internet hosts do not respond to active probes, the IP SLA ICMP echo probe. Because of these limitations, the PFC3-based architectures are not used as border routers in this section. There is a feature enhancement request, CSCsi59058, to add support for Internet path availability probing for load-balancing. This feature is targeted to support the PFC3-based architectures for Internet load sharing.

In the topology tested in this section, the Cisco 7200VXR series of routers are deployed as OER border routers. The master controller function is tested using the Cisco 7200VXR NPE-G2, Cisco 3845, and Cisco 7600-rsp720. An active/standby master controller configuration is also tested to demonstrate this function and to document a working configuration.

Design Requirements and Considerations

The Internet content server use case is the most common deployment scenario as this is the primary customer use case the OER technology was developed to address; optimization of large numbers of client devices sourced from several ISP connections. In terms of megabits per second, the bulk of the user traffic is from server to client. OER, therefore, is configured and addresses the path selection from server to client over two or more links to typically multiple ISPs.

The majority of the user traffic is TCP traffic, specifically HTTP (port 80) and SSL/HTTPS (port 443). The tested configuration uses two Cisco 7200VXR NPE-G2 as WAN edge routers terminating links to their respective ISPs. These are OER border routers in all test cases.

The MC function is tested using the Cisco 7200VXR NPE-G2, Cisco 3845, and Cisco 7600-rsp720. An active/standby MC configuration is also tested to demonstrate this function and to document a tested working configuration.

The objective of testing an Internet content server deployment is to determine what resource (memory or CPU utilization) is the limiting factor in scaling a dedicated master controller. In this deployment, it is assumed that the master controller is deployed on a separate chassis rather than collocated on a border router, because the goal is to scale the total number of prefixes being managed the resources consumed by the master controller function should not be limited or reserved in order to switch user packets or process other network functions like QoS, BGP peering, NAT, access-lists or Cisco IOS firewall.

1. EARL: Encoded Address Recognition Logic, describes the ASIC forwarding complex in a Catalyst switch. EARL 7 refers to the PFC3. The Supervisor 720 uses a PFC3.

It is important to note that the concept of performance routing is an optimization technique. In other words, it adds to or is an enhancement of the core function of the WAN aggregation role of switching packets to and from the Internet service providers and maintaining BGP peering sessions to send and receive network prefix advertisements. As such, using a dedicated chassis not only provides better opportunity to scale the performance routing function but also isolates it from the core function of WAN aggregation. It is a good design practice to dedicate a chassis for key functions where stability and isolation are important to the overall design. Using a dedicated chassis for the master controller function is analogous to using a dedicated route-reflector in large BGP deployment, a dedicated DLSw peer or TN3270 server.

Additionally, to provide design guidance and verification, the concept of implementing a standby master controller is demonstrated and tested in this section. The use of a standby master controller is useful to maintain the performance routing function in the event the primary master controller must be taken offline for service or experiences a hardware or software failure.

Scalability Considerations

July 2, 2007, a Cisco RTP campus Internet gateway (BGP peering with AS 7018 and AS 701) had 220,508 network prefixes in the BGP table using approximately 25MB of memory. If it is assumed that an Internet content server is receiving flows from 1 to 2% of these network prefixes during any given interval of time, managing 2,000 to 4,000 network prefixes is an expected requirement of the customer deployment. This snapshot is to provide context, as a point of reference, for the number of prefixes the typical enterprise customer may encounter. Selected content service providers may have more aggressive requirements.

OER, however, does not use the BGP routing table as the source of data to populate the master controller database, rather active flows from the NetFlow cache are used to populate the database. Actual user traffic, as cached by NetFlow, are used to determine what network prefixes are to be managed. The next section examines the basic configuration used in scale testing and how the configurable parameters influence prefix collection and retention.

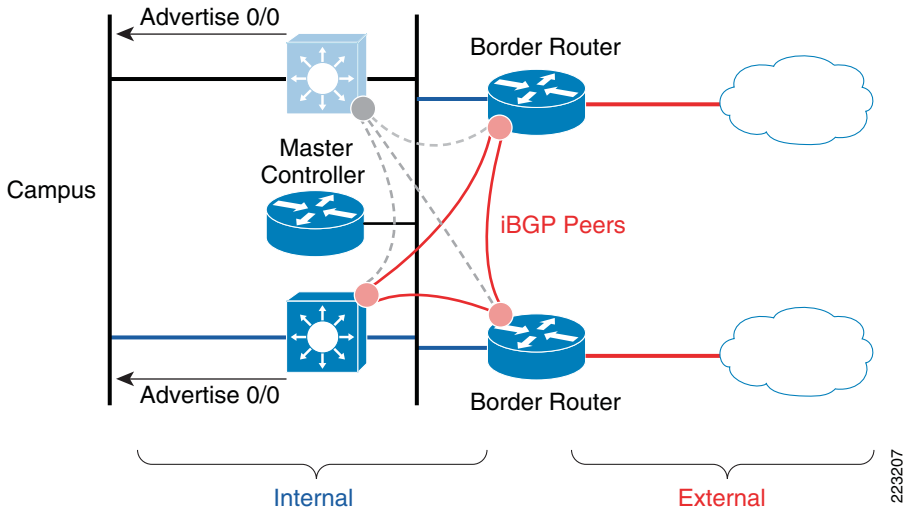
Prefix Management

This section describes how network prefixes are collected, aggregated, stored and reported between the border routers and the master controller. In these illustrations, the pictures of the plastic pails represents the collection and storage of network prefixes.

Underlying Routing

First the underlying routing configuration must be discussed. A very typical Internet edge deployment is shown in Figure 5.

Figure 5 Internet Edge Deployment Example



In Figure 5, there are two Layer 3 campus switches and two WAN edge routers. One of the Layer 3 campus switches is shown grey or subdued. In the lab test phase of this section, the second Layer 3 campus switch is not included in the topology, as this switch is deployed to provide redundancy. The testing was not meant to cover campus switching redundancy. However, for the purpose of explaining the underlying IP routing, assume that the subdued switch is advertising a default (0.0.0.0 / 0.0.0.0) route into the campus through some IGP routing protocol or is participating as a HSRP peer with the primary Layer 3 campus switch The primary Layer 3 campus switch is also advertising a default (0.0.0.0 / 0.0.0.0) route and/or is the active HSRP router.

A sample configuration to advertise a default route from both Layer 3 campus switches could be configured as shown below:

```
!
router eigrp 64
 redistribute static metric 9 5000 255 1 1500 route-map DFLT
 network 10.0.0.0
 no auto-summary
!
route-map DFLT permit 10
 match ip address DFLT_NET
!
ip access-list standard DFLT_NET
 permit 0.0.0.0
 deny   any
!
ip route 0.0.0.0 0.0.0.0 Null 0
!
end
```

This configuration provides availability for the application servers in the event of a Layer 3 campus switch failure. The default route can use the Null 0 interface as the next hop because more specific routing information is available through BGP routes. If no BGP learned route is available for the destination network, the packets are sent to the Null 0 interface, effectively dropping them.

iBGP Peering

The Layer 3 campus switches are iBGP peers with all WAN edge routers. The WAN edge routers are eBGP peers with their respective Internet service providers. The WAN edge routers are OER border routers.

In an OER deployment using only passive monitoring (*mode monitor passive*), the Layer 3 campus switches in this example must have two equal cost routes in the routing table to the destination IP network. By doing so, the NetFlow cache is populated by the traffic destined to the Internet through more than one OER external interface.

Tip This deployment does not use active probes. For OER to verify reachability for a destination network prefix, TCP traffic must be observed on more than one exit interface so OER has more than one exit with validated reachability to the target network prefix.

Assuming the eBGP routers receive the same network advertisement from more than one eBGP neighbor, the **maximum-paths iBGP 2** command inserts both entries in the routing table of the campus switch so traffic for a given prefix can be sent out both exits. If the **maximum-paths iBGP** command was not configured on the Layer 3 campus switches, only one of the routes to the destination network would be put in the routing table.

The following is a sample iBGP configuration from one of the Layer 3 campus switches:

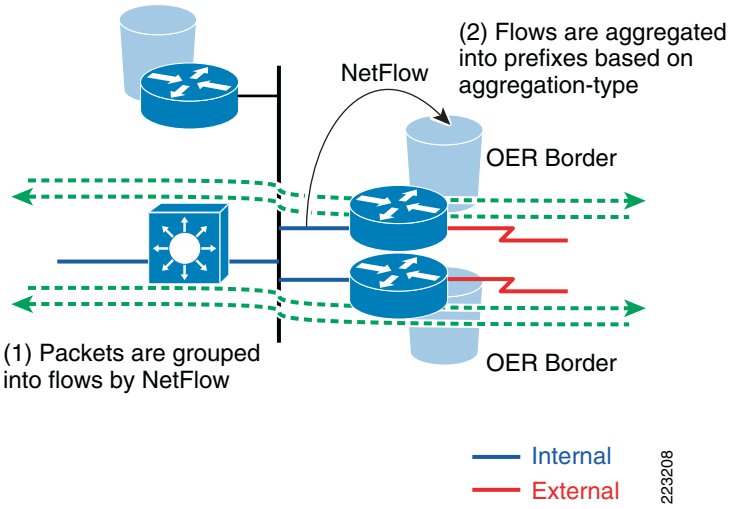
```
!
router bgp 65030
 no synchronization
 bgp log-neighbor-changes
 neighbor OER_Border peer-group
 neighbor OER_Border remote-as 65030
 neighbor OER_Border update-source Loopback0
 neighbor 192.168.130.1 peer-group OER_Border
 neighbor 192.168.130.2 peer-group OER_Border
! neighbor 192.168.130.99 peer-group OER_Border # Second Layer 3 campus switch
 maximum-paths ibgp 2
 no auto-summary
!
```

Note The Layer 3 campus switches and the WAN routers must advertise the external WAN links IP addresses and the iBGP peering addresses (loopback interfaces) are used in the previous example) by means of some common IGP.

Packet Aggregation into Flows

To simplify the topology for clarity, focus on the two border routers and external links, the single dedicated master controller and the Layer 3 campus switch. The illustrations of the containers in Figure 6 represent the collection and storage of network prefixes.

Figure 6 packet Aggregation



Packets leaving the application server destined for the client workstations on the Internet are grouped into flows by NetFlow. NetFlow is enabled automatically by OER on the internal and external interfaces of the border routers as specified in the master controller configuration.

Because learn mode is configured on the master controller, the border routers learns network prefixes from the NetFlow cache and break this learning step into one minute intervals specified by the **monitor-period** keyword.

The border routers summarize or aggregate flows into memory based on the value specified by the **aggregation-type** keyword. In this example and in scale testing, prefixes are summarized on a Classless Inter-Domain Routing (CIDR) length of /29. In dotted-decimal notation this is represented as a mask of 255.255.255.248. For reference, a chart of CIDR to dotted-decimal notation conversions is included in Appendix, page 123.

```
!
oer master
 logging
 !
 border 192.168.131.97 key-chain BLUE
  interface GigabitEthernet0/1.108 internal
  interface Serial 0/1  external
 !
 border 192.168.131.98 key-chain BLUE
  interface GigabitEthernet0/1.108 internal
  interface Serial 0/1  external
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  aggregation-type prefix-length 29
 mode route control
```

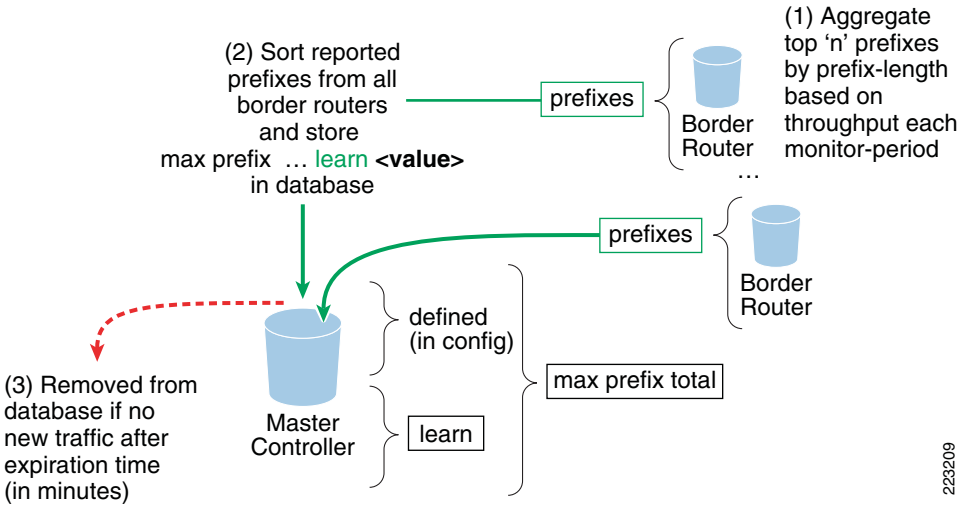
```
 mode select-exit best
 !

END
```

The **periodic-interval** value of 0 indicates the border routers immediately begin relearning prefixes after the **monitor-period** has expired and the collection of prefixes have been reported to the master controller.

Reporting Flows to Master Controller

At the end of the prefix learning interval specified by **monitor-period**, the border routers each sort their respective collected prefixes by total throughput, and report the top prefixes to the master controller. The number of prefixes reported is limited by the value specified as the **prefixes** keyword subordinate to the **learn** command. In the previous configuration example, the value is 2500, meaning the top 2,500 prefixes based on total throughput observed during the **monitor-period** sent to the master controller from each border router. Figure 7 shows this as the first step in the process.

Figure 7 Reporting Prefixes to Master Controller



As the master controller now has received the observed flows from all the border routers, there are several other keywords that govern how these prefixes are managed. Refer to the following master controller configuration example:

```
!
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
!
  expire after time 6
  aggregation-type prefix-length 29
 max prefix total 5000 learn 2500
 mode route control
 mode monitor passive
 mode select-exit best
```

```
 periodic 180
!

end
```

The master controller must sort the collection of learned prefixes from the most current period from all border routers along with prefixes learned previously. The **max prefix** keyword determines the total number of prefixes stored by the master controller, as well as the maximum number of learned prefixes.

The other method a prefix may be managed is by reference in an oer-map through a prefix-list referenced from a match traffic-class statement. This would account for the difference between the total value and the learned value. The **expire after time** keyword is a means of removing prefixes from the collection if no new traffic is observed in the number of minutes defined on the keyword. In other words, it is a means of aging and removing older entries which are no longer active.

Summary

This section described the method by which traffic is collected into flows, and flows are sorted and reported to the master controller by the border routers. The prefixes are maintained in association with the learning border router and exit interfaces. The master controller sorts and manages the learned prefixes, aging stale entries, to provide currency of the managed prefixes.

Scalability and Performance Results

This section describes the scale testing performed by Cisco.

Performance Results Summary

The performance results for the three hardware platforms tested is shown in Table 2.

Table 2 Performance Testing Results

Platform	Limiting Factor	Number of Prefixes *max prefix total*
Cisco Route Switch Processor 720 RSP720-3C-GE *PowerPC MPC8548_E at 1200MHz*	Software *Note 1*	up to 2,500
Cisco 7206VXR (NPE-G2) *Motorola Freescale MPC7448 CPU at 1666Mhz*	CPU	up to 15,000
Cisco 3845 *256MB of memory*	Memory	up to 3,000

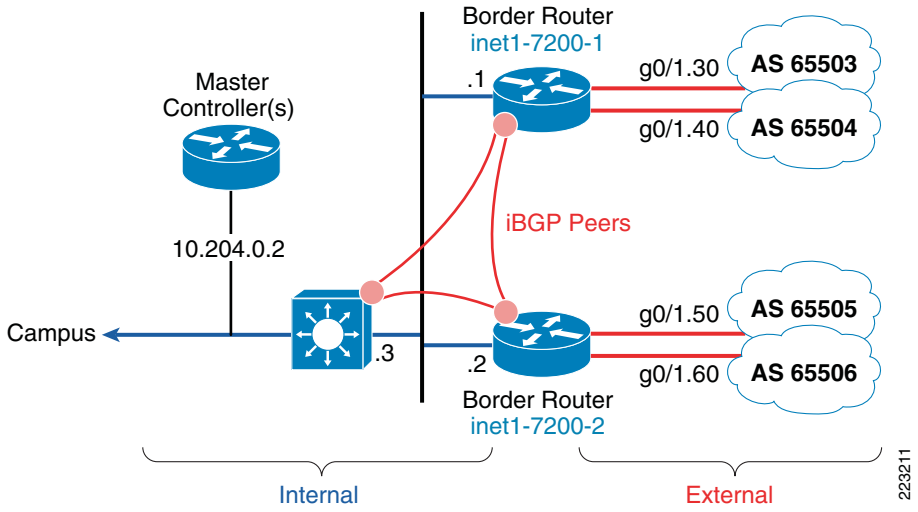
These maximum prefix recommendations assume a dedicated master controller configuration without other CPU intensive processes configured. The recommended number of prefixes is intended to be a conservative guideline. If a border router function is also configured with the master controller, and BGP is also configured, the maximum number of prefixes may be less than shown.

Note The tested software version (122-33.SRB) has a configuration limit of 2,500 prefixes. However, even without this absolute limit, the RSP720 will encounter the same CPU constraint as the NPE-G2; albeit at a lower number of prefixes due to the lower clock rate of the RSP720 CPU.

Topology

Figure 8 illustrates the topology being tested. The four autonomous systems (ASs) contain the client addresses and the campus contains the Chariot/IXIA endpoint representing the application server. The campus switch is a Layer 3 switch with iBGP peering to the OER border routers, which are also eBGP peers with the Internet service providers. The OER master controller is on a VLAN attached to the campus switch.

Figure 8 Topology for Performance Testing



Traffic Profile

The traffic profile and testbed for this test consisted of 1,000 real routers. Each router is allocated a /24 address space. Each router initiates 20 flows; 10 HTTP flows with a DSCP value of BE, 10 HTTP flows with DSCP value of AF21. The flows are generated by Chariot running on an IXIA chassis. In order to simulate up to 20,000 network prefixes, the OER master configuration is set to aggregate on a /29 boundary. The IP addresses from the /24 address space were allowed in /29 increments. In other words, if the remote address space is 10.10.10.0/24, flows are generated from individual IP addresses of 10.10.10.1, 10.10.10.9, 10.10.10.17, and so on, up through the 20th /29th block of addresses.

Software Release

The tested software releases are based on the Cisco IOS 124-15.T1 train for the Cisco 7200VXR NPE-G2 and Cisco ISR 3845 and 122-33.SRB for the Cisco 7600-rsp720.

Tested Configuration

The master controller configuration tested is based on the following configuration:

```
learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes {number}
  expire after time {number}
  aggregation-type prefix-length 29
 max prefix total {number} learn {number}
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
```

This configuration is a very basic and straight forward passive learning configuration suitable for an Internet content server deployment.

Cisco 7200VXR NPE-G2 as Master Controller

The Cisco 7200VXR NPE-G2 when deployed as a dedicated master controller is able to manage over 15,000 network prefixes; however, the CPU is 100 percent used by the master controller process for over 6 seconds when the border routers report learned prefixes to the master controller.

This issue is tracked by the following bug ID:

CSCsk48862 - CPU HOG while learning if very large number of prefixes

The CPUHOG messages also appear with 10,000 network prefixes in the database; however, the message may not appear as frequently as with 15,000 prefixes.

Warning **%SYS-3-CPUHOG warning messages are displayed if a process does not relinquish control of the processor for more than 2 seconds. The CPUHOG warning is reported every 2 seconds until the process exits. The negative impact of this event is other processes may not obtain a timely share of the CPU resources. For example, routing protocol and HSRP hellos may not be sent, causing neighbor relationships to drop.**

CPU Characteristics

In the following output from **show proc cpu history** command, the reported one-minute CPU values that corresponded with the CPUHOG syslog messages correspond with the 35 to 36 minute data points.

The **show proc cpu hist** command displays the following output:

```
    he1-7200-4   12:41:51 PM Tuesday Oct 2 2007 EDT

             1
        6998997896979079999696858686999868977926578798743521111
        8537221338331033823046863732789671372709650524966083132111111
    100  *           *  *           **             *
```

```
 90  *****  * * ** *** * * * * **** **  *    * *
 80  ***** ** * ** *** * * * * **** ** **   ** **
 70 ****************** * * * * *********   ******
 60 ***********************************  ********  *
 50 ***********************************  ********* *
 40 ***********************************  ***********
 30 ************************************************
 20 ************************************************
 10 ##################################*##########*****
   0....5....1....1....2....2....3....3....4....4....5....5....6
             0    5    0    5    0    5    0    5    0    5    0
               CPU% per minute (last 60 minutes)
              * = maximum CPU%   # = average CPU%
```

```
Oct  2 12:06:05.563 EDT: %SYS-3-CPUHOG: Task is running for (2004)msecs, more than
(2000)msecs (8/8),process = OER Master Controller.
-Traceback= 0x25F47B0 0x25F5114 0x25F651C 0x263CC94 0x263DA3C 0x263A8D0 0x25DDC18
0x25DE1EC 0x25DE3C8 0x263BC98 0x25E7CB0 0x6F2240
Oct  2 12:06:07.567 EDT: %SYS-3-CPUHOG: Task is running for (4004)msecs, more than
(2000)msecs (9/8),process = OER Master Controller.
-Traceback= 0x25F4AE4 0x25F5114 0x25F6568 0x263CEB8 0x263DA3C 0x263A8D0 0x25DDC18
0x25DE1EC 0x25DE3C8 0x263BC98 0x25E7CB0 0x6F2240
Oct  2 12:06:09.567 EDT: %SYS-3-CPUHOG: Task is running for (6004)msecs, more than
(2000)msecs (9/8),process = OER Master Controller.
-Traceback= 0x25F4AF8 0x25F5114 0x25F6568 0x263CEB8 0x263DA3C 0x263A8D0 0x25DDC18
0x25DE1EC 0x25DE3C8 0x263BC98 0x25E7CB0 0x6F2240
```

Note that the tabular CPU report of the 35 to 36 minute observation is 73 percent and 77 percent maximum CPU with the average CPU busy value reported at approximately 10 percent.

What occurred can be characterized by burst of processing associated with sorting and managing the database of network prefixes to rank them by total throughput so as to discard those prefixes that fall outside the total learned prefix parameter. Note that the master controller is configured with the **max prefix total** <*n*> *learn* <*n*> command, where the **learn** value is the maximum number of learned prefixes that can be stored in the master controller database.

As a design best practice, several items can be addressed to avoid any issues with the high CPU required to sort and manage large numbers of network prefixes by the master controller. They include:

- When managing large (over 5,000 on a 7200VXR NPE-G2) numbers of prefixes, use a dedicated master controller.
- Use static routes instead of a routing protocol on the master controller to eliminate the need for processing routing protocol hello and updates.
- If using a standby master controller, as shown later in this section, configure the HSRP hello and dead interval values sufficiently high to prevent loss of HSRP adjacency during the CPUHOG reported period.

Using HSRP standby timers at 10 seconds for the hello and a dead interval of 3 to 4 times the hello interval (**standby 0 timers 10 31**) is a recommended starting value in this deployment. The hello value can be adjusted up or down, with the dead interval at 3 to 4 times the hello value, based on the customer deployment experience.

Memory Characteristics

In testing, approximately 532MB of memory was shown as free with 15,000 network prefixes under management. There is approximately 745M of memory shown as free at the start of the test. Approximately 213M of memory is used for 15,000 prefixes. As such, memory utilization is not a constraint. The 7200VXR NPE-G2 has 1GB memory installed:

```
-===== he1-7200-4: show mem summ Tue Oct  2 11:42:17 2007 =====-
show mem summ
                Head    Total(b)     Used(b)     Free(b)   Lowest(b)  Largest(b)
Processor    61EC200   820067260    38018500   782048760   630679684   719403876
      I/O   38000000    67108864     9252500    57856364    57843856    57793660
Transient   37000000    16777216       14996    16762220    16301732    16761208


-===== he1-7200-4: show mem summ Tue Oct  2 12:39:41 2007 =====-
show mem summ
                Head    Total(b)     Used(b)     Free(b)   Lowest(b)  Largest(b)
Processor    61EC200   820066380   261546732   558519648   555552280   551630836
      I/O   38000000    67108864     9259980    57848884    57843792    57793660
Transient   37000000    16777216       14996    16762220    16301732    16761208
```

System Hardware

The software version tested is Cisco IOS Release 12.4(15)T1 on a Cisco 7206VXR (NPE-G2) processor (MPC7448 CPU at 1666Mhz) with 917504K/65536K bytes of memory.

http://harry/solutions/vpn/results/ngwan30/oer/mc/7200-g2/15k/

Tested Configuration

```
!
hostname he1-7200-4
!
boot-start-marker
boot system flash disk2:c7200p-adventerprisek9-mz.124-15.T1
boot system flash disk2:
boot-end-marker
!
!
oer master
 no keepalive
 !
 border 192.168.33.1 key-chain NGWAN
  interface GigabitEthernet0/1.40 external
  interface GigabitEthernet0/1.30 external
  interface GigabitEthernet0/2 internal
 !
 border 192.168.33.2 key-chain NGWAN
  interface GigabitEthernet0/1.60 external
  interface GigabitEthernet0/1.50 external
  interface GigabitEthernet0/2 internal
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  ! expire after time 65000
  aggregation-type prefix-length 29
 ! max prefix total 100000 learn 100000
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
 !
```

```
!
!
buffers tune automatic
!
!
interface GigabitEthernet0/2
 description GigabitEthernet0/2-Inside
 ip address 10.204.0.2 255.252.0.0
 hold-queue 4096 in
 hold-queue 4096 out
!
end
```

Note The configuration statements of **expire after time** and **max prefix total** are listed a comments in the above configuration. The values shown were used during testing, but are values which facilitate scale testing and may not be suitable for every customer deployment.

Cisco RSP720 as Master Controller

The scalability of this configuration is not limited by either CPU or memory utilization of the SUP720 with 2,500 network prefixes actively managed. The OER implementation in 12.2(33)SRB is limited by the software configuration to manage a maximum of 2,500 learned prefixes.

CPU and Memory Characteristics

This test was executed for a 30-minute period. At the conclusion of the test, the CPU history was recorded and the output is shown below:

```
show proc cpu hist Tue Oct  2 15:59:59 2007

    11111    1  111111   11111112
    33420997708733145379908655499 1 1 11  1 119111 1 1  11 1
100
 90
 80
 70
 60
 50
 40
 30                              *
 20                 *     **** **
 10 ****************************             *
   0....5....1....1....2....2....3....3....4....4....5....5....
             0    5    0    5    0    5    0    5    0    5
               CPU% per minute (last 60 minutes)
              * = maximum CPU%   # = average CPU%
```

The results are such that the CPU utilization of the RSP720 is not a limiting factor when managing 2,500 network prefixes. At the conclusion of the test, the processor memory utilization was captured and over 600MB of memory is available with 2,500 prefixes in the master controller database.

-===== he3-7600-1: show mem summ Tue Oct 2 15:59:55 2007 =====-

```
show mem summ
                Head    Total(b)     Used(b)     Free(b)   Lowest(b)  Largest(b)
Processor    E8C5B44   762553532   143289236   619264296   615942236   615923308
      I/O   3C000000    67108864    23805276    43303588    43303588    43301884
```

System Hardware

The Cisco Route Switch Processor 720 (RSP720-3C-GE) was tested as an OER master controller. The border routers are Cisco 7200VXR NPE-G2. The chassis is a CISCO7606 (M8500) with 851968K/65536K bytes of memory. The software image is version 12.2(33)SRB.

http://harry/solutions/vpn/results/ngwan30/oer/mc/rsp720/5000/

Tested Configuration

The following master controller configuration is used on the Cisco RSP720 during the performance test

```
!
hostname he3-7600-1
!
boot-start-marker
boot system flash sup-bootdisk:c7600rsp72043-adventerprisek9-mz.122-33.SRB.bin
boot-end-marker
!
```

```
key chain NGWAN
 key 10
  key-string cisco
!
!
oer master
 no keepalive
 !
 border 192.168.33.1 key-chain NGWAN
  interface GigabitEthernet0/1.40 external
  interface GigabitEthernet0/1.30 external
  interface GigabitEthernet0/2 internal
 !
 border 192.168.33.2 key-chain NGWAN
  interface GigabitEthernet0/1.60 external
  interface GigabitEthernet0/1.50 external
  interface GigabitEthernet0/2 internal
 !
learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
 expire after time 65000
  aggregation-type prefix-length 29
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
end
```

Cisco 3845 as Master Controller

The Cisco 3845 ISR router is tested as an OER master controller. The border routers are Cisco 7200VXR NPE-G2.

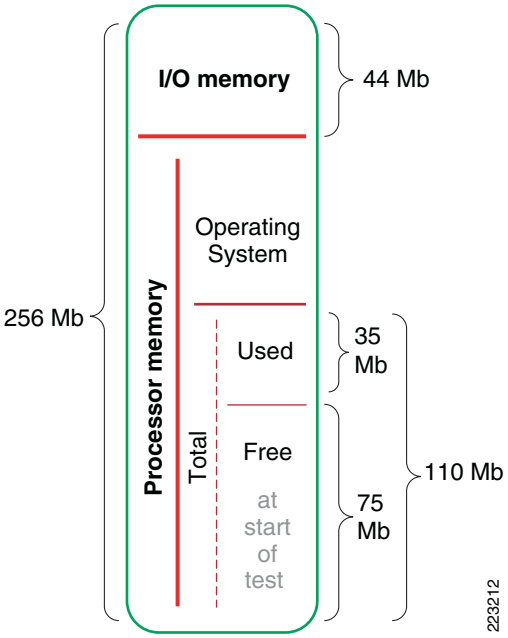
The scalability of this configuration is limited by the available memory of the Cisco 3845 ISR. The CPU utilization is not a limiting factor. The CPU busy as reported by **show proc cpu** command did not exceed 10 percent for either the one-minute or five-minute values during the test.

The installed memory for the test is 256Mb. When deploying the Cisco 3845 ISR as a dedicated master controller, the recommended number of managed prefixes should not exceed 3,000.

Memory Allocation

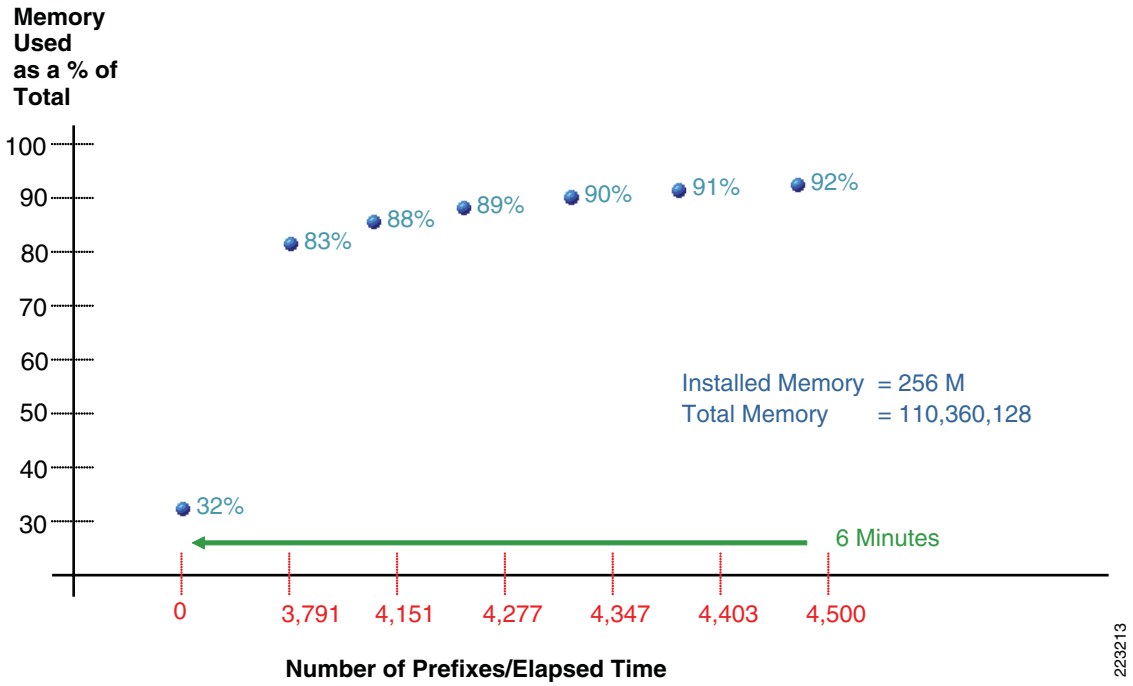
To illustrate the memory configuration of the tested Cisco 3845 ISR, the memory allocation map in Figure 9 depicts how memory was allocated at the beginning of the test.

Figure 9 Memory Allocation



At the beginning of the test, the amount of used memory as a percent of total memory as approximately 32 percent, calculated by dividing used of 35MB / 110Mb or 31.8 percent. As prefixes are learned by the border routers, the unallocated free memory is consumed as shown in Figure 10.

Figure 10 Memory Consumption as Prefixes are Learned



As the chart in Figure 10 illustrates, after 6 minutes, 4.500 prefixes are learned and 92 percent of the total memory is required to store these learned prefixes.

System Hardware

The Cisco IOS imaged tested was 12.4(15)T1 and the Cisco 3845 hardware configuration included a total of 256Mb of installed memory.

```
System image file is "flash:c3845-adventerprisek9-mz.124-15.T1.bin"
Cisco 3845 (revision 1.0) with 219136K/43008K bytes of memory.
Processor board ID FTX0939A135
20 FastEthernet interfaces
2 Gigabit Ethernet interfaces
2 Virtual Private Network (VPN) Modules
DRAM configuration is 64 bits wide with parity enabled.
479K bytes of NVRAM.
62720K bytes of ATA System CompactFlash (Read/Write)
```

Tested Configuration

```
!
hostname he1-3800-2
!
!
oer master
 no keepalive
 !
 border 192.168.33.1 key-chain NGWAN
  interface GigabitEthernet0/1.40 external
  interface GigabitEthernet0/1.30 external
```

```
 interface GigabitEthernet0/2 internal
!
border 192.168.33.2 key-chain NGWAN
 interface GigabitEthernet0/1.60 external
 interface GigabitEthernet0/1.50 external
 interface GigabitEthernet0/2 internal
!
learn
 throughput
 delay
 periodic-interval 0
 monitor-period 1
 prefixes 2500
 expire after time 65000
 aggregation-type prefix-length 29
max prefix total 100000 learn 100000
mode route control
mode monitor passive
mode select-exit best
periodic 180
!
buffers small permanent 1500
buffers small max-free 2000
buffers small min-free 450
buffers middle permanent 1000
buffers middle max-free 1500
buffers middle min-free 300
buffers big permanent 1000
buffers big max-free 1500
buffers big min-free 300
!
end
```

Troubleshooting

The following commands provide useful output for managing and monitoring the deployment. Commands beginning with **show oer border** are executed on a border router, commands which begin with **show oer master**, are master controller-related commands.

show oer border passive learn

```
#show oer border pass learn
 OER Border Learn Configuration :
    State is enabled
    Measurement type: throughput and delay, Duration: 1 min
    Aggregation type: prefix-length, Prefix length: 29
    No port protocol config
```

show oer border routes bgp

```
#show oer border routes bgp
BGP table version is 72407, local router ID is 192.168.130.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
OER Flags: C - Controlled, X - Excluded, E - Exact, N - Non-exact, I - Injected

   Network          Next Hop        OER    LocPrf Weight Path
*> 192.168.48.0/20  192.168.129.33  CE              0 65001 65005 i
*> 192.168.129.24/29
                    192.168.129.33  CEI             0 65001 i
```

show oer border

```
#show oer border
OER BR 192.168.131.98 ACTIVE, MC 192.168.131.126 UP/DOWN: UP 1w2d,
  Auth Failures: 0
  Conn Status: SUCCESS, PORT: 3949
  Version: 2.1  MC Version: 2.1
  Exits
  Gi0/1.108       INTERNAL
  Gi0/1.652       EXTERNAL
  Gi0/1.653       EXTERNAL
```

show oer master border

```
#show oer master border
Border           Status   UP/DOWN             AuthFail  Version
192.168.131.98   ACTIVE   UP       1w2d          0  2.1
192.168.131.97   ACTIVE   UP       1w2d          0  2.1
```

show oer master policy

```
#show oer master policy
Default Policy Settings:
  backoff 90 270 90
```

```
  delay relative 50
  holddown 300
  periodic 180
  probe frequency 56
  mode route control
  mode monitor both
  mode select-exit best
  loss threshold 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  resolve loss priority 1 variance 10
  resolve delay priority 11 variance 20
  resolve utilization priority 12 variance 20
```

show oer master prefix learned

```
#show oer master prefix learned
OER Prefix Statistics:
 Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
 P - Percentage below threshold, Jit - Jitter (ms),
 MOS - Mean Opinion Score
 Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
 E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
 U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
 # - Prefix monitor mode is Special, & - Blackholed Prefix
 % - Force Next-Hop, ^ - Prefix is denied

Prefix                  State     Time Curr BR         CurrI/F         Protocol
                     PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                     ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                     ActSJit  ActPMOS  ActSLos  ActLLos
--------------------------------------------------------------------------------
192.168.195.8/29        HOLDDOWN      310 192.168.131.97  Gi0/1.651       BGP
                               U        U        0        0        0        0
                               N        N        N        N        0        1
                               N        N
```

show oer master

```
#show oer master
OER state: ENABLED and ACTIVE
  Conn Status: SUCCESS, PORT: 3949
  Version: 2.1
  Number of Border routers: 2
  Number of Exits: 4
  Number of monitored prefixes: 1 (max 100000)
  Max prefixes: total 100000 learn 2500
  Prefix count: total 1, learn 0, cfg 1

Border           Status   UP/DOWN             AuthFail  Version
192.168.131.98   ACTIVE   UP       00:00:53          0  2.1
192.168.131.97   ACTIVE   UP       00:00:53          0  2.1

Global Settings:
  max-range-utilization percent 20 recv 20
  mode route metric bgp local-pref 5000
  mode route metric static tag 5000
  trace probe delay 1000
  logging
```

```
Default Policy Settings:
  backoff 90 270 90
  delay relative 50
  holddown 300
  periodic 180
  probe frequency 56
  mode route control
  mode monitor passive
  mode select-exit best
  loss threshold 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  resolve loss priority 1 variance 10
  resolve delay priority 11 variance 20
  resolve utilization priority 12 variance 20

Learn Settings:
  current state : STARTED
  time remaining in current state : 96 seconds
  throughput
  delay
  no inside bgp
  no protocol
  monitor-period 1
  periodic-interval 0
  aggregation-type prefix-length 29
  prefixes 2500
  expire after time 6
```

Standby Master Controller

To provide availability of the master controller function, a standby (backup) master controller can be implemented using a Hot Standby Router Protocol (HSRP) virtual IP address as the destination IP address configured in the border routers to identify the master controller. The implementation of OER is such that the master controller listens on TCP port 3949 (by default, the port number is configurable) and the border routers initiate the TCP connection.

If the border routers lose contact with the master controller, the border routers remove OER controlled routes and fallback to parent route control. The border routers will continuously initiate a connection to the master controller IP address until a session is established.

To implement a standby master controller, do the following:

Step 1 Configure the OER master controller configuration identically on both primary and backup.

Step 2 Implement HSRP on the the two master controllers.

Step 3 Configure the border routers to use the HSRP virtual IP address to identify the master controller.

Note The standby master controller functions in a stateless failover mode. However, standalone (dedicated) master controllers are not in the data plane switching path, so user traffic is not disrupted; it is simply not optimized during failover and recovery.

The general guideline is that the standby master controller is effectively online and operational beginning with the monitor period following the failure. So effectively, with a one minute monitor-period, prefixes can be managed within 2 to 3 minutes following a failure.

More specific implementation details are described in the following section.

Operational Overview

In this configuration sample, the master controller is configured to instruct the border routers to learn traffic for a one minute interval (**monitor-period 1**) and initiate subsequent monitoring (**periodic-interval 0**) immediately.

The learned prefixes is maintained by the master controller for the period specified by the **expire after time** command value. Arguably, maintaining learned prefixes for long periods of time with little or no activity is of questionable value in that Internet content server deployment have very transient clients. The number of prefixes learned by the border routers in one interval is specified by the **prefixes** command under the learn statement in the master controller configuration. All prefixes learned are sorted by throughput and the highest (top) *'n'* throughput prefixes are reported to the master controller.

OER Keepalive Disabled

If the OER keepalive is disabled (**no keepalive** configured under the **oer master** section), the state of the TCP session triggers the failover to the standby master controller.

At the failure of the primary master controller, the standby master controller becomes the active HSRP router and takes over control of the virtual IP address. The border routers have an open TCP session with the virtual address on TCP port 3949. When the standby master controller receives a TCP packet destined for port 3949 (for which it is listening, but has no active session) from the border router, it sends a TCP

RST (Reset) in response. The border routers log the fact the master controller is down, and immediately attempt to re-establish a TCP session to the virtual address on port 3949. In testing, the border routers were able to re-establish communication with the (backup) master controller within approximately 5 seconds after receiving the TCP RST.

Network prefixes which are currently under control of an OER border router (static or BGP) are removed from the IP routing or BGP table when the master controller is deemed down.

The standby master controller does not have any knowledge of previously learned prefixes, but it will begin accumulating the learned prefixes for every monitor period interval subsequent to the failure of the primary master controller.

Tip If testing this failure condition, administratively shutting down the primary master controller interface will not demonstrate a true failure condition. Cisco IOS sends a TCP RST (reset) for the active TCP connection on that interface to the border routers. The border routers in response tear down their TCP connection and retry approximately 5 seconds later to establish a connection to the master controller IP address.

The use of performance routing techniques is a WAN optimization. Optimization is not perfect. OER, through the nature of its management and control of traffic, routinely reverts back to the default routing table behavior if a prefix is out-of-policy (OOP). When a prefix is initially managed, it is placed in HOLDDOWN state for a minimum of 5 minutes and only exits holddown prior to the minimum time, if a blackout or brownout causes an unreachable condition.

The fact that the standby master controller does not have a database populated with historical data from the failed primary master controller does not mean that the capabilities of performance routing are totally negated following a stateless recovery. There is a period of a few minutes where routing reverts to the default nature while the fallback is completed.

OER Keepalive Enabled

If OER keepalives are enabled, these OER specific keepalives are also used to detect a communication failure between the master controller and the border routers. The border router can detect a failure of the master controller after three keepalives are unanswered and deem the master controller down. Once the master controller is down, either because of the TCP RST message received from the standby master controller or the loss of three keepalives, the next course of action by the border router is the same—attempt to establish a new TCP session with the master controller.

In testing a keepalive value of one second was configured and the master controller was declared down 3 seconds after the link was failed between border routers and master controller. The default keepalive value is 5 seconds. The hold-time (dead interval) is 3 times the configured keepalive value.

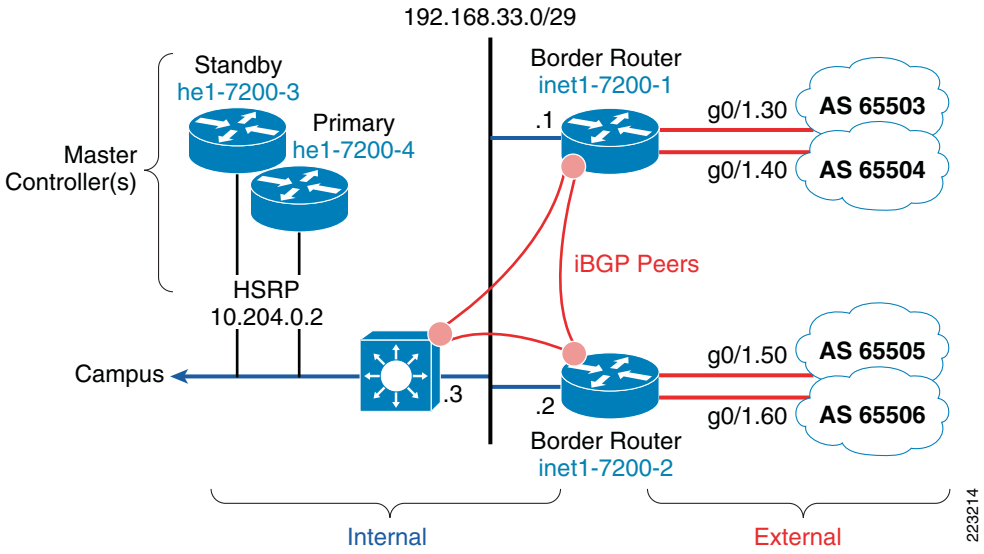
The use of OER keepalive is most applicable to verifying communications with a single master controller. As the standby master controller sends a TCP RST following assuming control of the virtual IP address, the use of the OER keepalive is not required as a recovery mechanism.

Tip OER keepalive can be disabled in this configuration. By doing so, it also eliminates the overhead associated with sending and responding to keepalives and can be an aid in scaling.

Topology

The tested topology is the Internet content server deployment scenario. Two border routers are implemented, each with two external (exit) links. A primary and standby master controller are implemented as shown in Figure 11.

Figure 11 Internet Content Server Deployment Example



A single Layer 3 campus switch is implemented in this test deployment. In most customer implementations, multiple Layer 3 campus switches are deployed for increased availability.

Authentication

While not shown in every configuration example, authentication of border routers and master controllers is accomplished by the use of a key-chain, a sample configuration is shown:

```
!
key chain NGWAN
 key 10
   key-string cisco
!
```

The key-string value of **cisco** is a lab standard and not recommended for a customer deployment.

Master Controller Configuration

In the test environment, both primary and standby master controller are configured identically as far as the **oer master** section of the configuration. This configuration section is shown as follows:

```
oer master
 no keepalive
 !
 border 192.168.33.1 key-chain NGWAN
```

```
  interface GigabitEthernet0/2 internal
  interface GigabitEthernet0/1.30 external
  interface GigabitEthernet0/1.40 external
 !
 border 192.168.33.2 key-chain NGWAN
  interface GigabitEthernet0/2 internal
  interface GigabitEthernet0/1.50 external
  interface GigabitEthernet0/1.60 external
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
!
! See notes below regarding the commented configuration entries which follow
!
  ! expire after time 65000
! aggregation-type prefix-length 29
  ! max prefix total 100000 learn 100000
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
```

The **expire after time** value of 65000 in the configuration is configured as part of the scale test to determine the upper bounds of the number of concurrent managed prefixes. This was, therefore, an arbitrary value, not necessarily representing a best practice for an Internet content server deployment. In a customer deployment, a recommended value of 4 minutes may be more appropriate, given the transient nature of the typical client base. With a 4-minute value, if the prefix is not relearned after 4 minutes, it is removed from the database. Therefore, this is an aging technique to maintain only current entries in the database.

The **aggregation-type prefix-length** value of 29 (/29) is a value that was applicable to the test lab address space assignment. The default value is 24 (/24) and this value is used by most Internet deployments. A value suitable for production deployment depends on the traffic distribution, an aggregation value of /23, /22 or less may be required.

The **max prefix total** value should be configured based on the memory and CPU performance characteristics of the master controller in use. The Performance Results Summary, page 31 provides recommended values.

Primary Master Controller HSRP interface

The interface used as the target IP address by the border routers for the primary master controller is shown in the following example. Note the standby IP address is 10.204.0.2 and this address is used in the border router configuration rather than the actual interface IP address of the primary master controller.

```
interface GigabitEthernet0/2
 ip address 10.204.0.4 255.252.0.0
 no ip redirects
 no ip unreachables
...
 standby 0 ip 10.204.0.2
 standby 0 timers 3 9
 standby 0 priority 110
 standby 0 preempt
 hold-queue 4096 in
```

```
 hold-queue 4096 out
!
```

The configured standby priority of 110 is higher than the default value of 100, making this router the preferred active HSRP router.

The default value for HSRP timers is 3 second hello and 10 second dead (**standby timers 3 10***)* the value shown was used to populate the value as output in the interface configuration. Given this is a stateless failover configuration and these master controllers are not in the data path plane (they are not switching any packets, only managing the master controller database) there is no appreciable value of aggressively tuning the HSRP timers to any value lower than the default values. In fact, increasing the HSRP hello and dead interval times is recommended so that a HSRP failover due to CPU busy on the primary master controller (a false positive switchover) is avoided. HSRP failover and recovery within 30 seconds to one minute will be sufficient for most deployments.

The hold-queue values of 4096 for both input and output is recommended as all packets on these dedicated master controller routers is process switched. Buffer tuning is also advised to eliminate buffer misses and failures.

Standby Master Controller HSRP Interface

The standby master controller HSRP interface uses the default HSRP priority of 100 and has not been configured to preempt, otherwise, it is identical to the primary master controller configuration.

```
!
interface GigabitEthernet0/2
 ip address 10.204.0.3 255.252.0.0
. . .

 standby 0 ip 10.204.0.2
 standby 0 timers 3 9
 hold-queue 4096 in
 hold-queue 4096 out
!
```

Border Router Configuration

Both border routers are identically configured as shown below:

```
!
oer border
 logging
 local GigabitEthernet0/2
 master 10.204.0.2 key-chain NGWAN
!
```

Summary

Implementing a standby master controller is a simple and effective means of providing stateless failover for the master controller function for large scale Internet content server deployments. An additional advantage to doing so is that OER keepalives can be disabled since the standby master controller can notify the border routers indirectly through a TCP RST message that the primary master controller has failed and the standby master controller has taken control of the virtual IP address.

WAN Hub: Dual MPLS Service Providers

This deployment is based on an unencrypted traffic profile where voice and data are supported from dual-central site locations to branch offices. It represents a large enterprise deployment that is migrating away from an ATM/Frame Relay WAN to dual MPLS service providers.

The OER border router function is implemented on a Cisco 7600 series for one service provider and the second service provider is supported by a Cisco 6500 series. This demonstrates the suitability of either platform as a WAN edge device. The OER master controller function is implemented on a Cisco 7200VXR chassis. This implementation demonstrates the use of the *mode monitor special*, where active probing is always enabled to provide reachability information for path rerouting.

The goal is to demonstrate a working deployment of over 1,000 branch routers. Each branch router is dual attached, one access link to each MPLS service provider. The access links are Ethernet handoff with a separate physical interface to each service provider. The enterprise branch router is configured to shape the uplink at a T1 data rate while the individual service provider will provide a similar QoS configuration and data rate on the downlink to the branch. The campus to service provider routing protocol is eBGP while the branch locations use static routes. OER can be deployed with both static or eBGP routes as the control or parent routes.

Latency, loss, and jitter are introduced artificially over time to degrade the quality of one service provider verses the other on a service provider (and optionally on a per-branch) basis. This would reflect a deployment where some branches encounter a better path through one service provider while other branches favor the other service provider.

Design Requirements and Caveats

The Cisco Catalyst 6500 Supervisor 32 (Sup32) is not a supported configuration for the OER Border Router functionality. The Supervisor 32 is more aligned at wiring closet deployments. The Sup32 is a "classic" supervisor that connects to the 32 Gbps bus. The Supervisor engine 720 (SUP720) connects to the 720-Gbps crossbar switch fabric as does the Route Switch Processor 720 (RSP 720)

The 6500 and 7600 chassis deployed in the border router role in this test have a Supervisor Engine 720 (SUP720) . The Route Switch Processor 720 (RSP 720) was not tested.

Master Controller is aware of the limitations of the Catalyst 6500 and Cisco 7600 chassis using an EARL 7 (PFC3) The Supervisor 720 uses a PFC3. The use of the *mode monitor special* is required. If *mode monitor both* is configured, the configuration is modified to *mode monitor special* when the master controller detects a border router which is PFC3-based. In this configuration, the master controller directs the border routers to measure throughput using NetFlow. Active probes are run continuously on the border routers to measure delay, loss, and reachability.

Tip The output from **show oer master prefix** command lists an '#' next to the prefix with **mode monitor special**.

Scalability Considerations

The scalability testing is not intended to address master controller scalability. The master controller scalability testing is discussed in Internet Content Server, page 25. The goal of testing the Cisco 6500 and 7600 chassis with the Supervisor engine 720 (SUP720) is to demonstrate the suitability of implementing these chassis in an environment where active probes are not considered detrimental, as is

the case with the Internet Content Server deployment. The use of active probes over the Internet can be construed as a denial-of-service attack or at the least, not practical, in that many Internet hosts are administratively prohibited from responding to an ICMP echo request. The WAN in the tested topology is dual Layer 3 MPLS service providers. Because MPLS provides a virtual private network (VPN), the use of active probes is considered normal or tolerated by the branch routers and hosts as they are all under the administrative control of a single network domain.

One consideration that should be made is that generating active probes continuously may burden the CPU of the border routers to the point it is detrimental to the forwarding ability of these WAN edge routers. The goal therefore of this testing is twofold; demonstrate a working configuration and verify the continuous active probing is not detrimental to the functioning of the network.

Scalability and Performance Results

This section describes the scale testing performed.

Performance Results Summary

The results of this testing shows a working OER configuration with *mode monitor special*, which means that the border routers must generate active probes to the collection of workstations or devices supported by the 1,000 branch routers for the purposes of verifying reachability.

In the load-sharing section, the number of active probes is plotted as well as the CPU utilization of the border routers, the 7600 and 6500 chassis. These results demonstrate that active probing for 1,000 branch routers is within performance capabilities of the Supervisor 720 where eBGP peering with an MPLS service provider is also enabled.

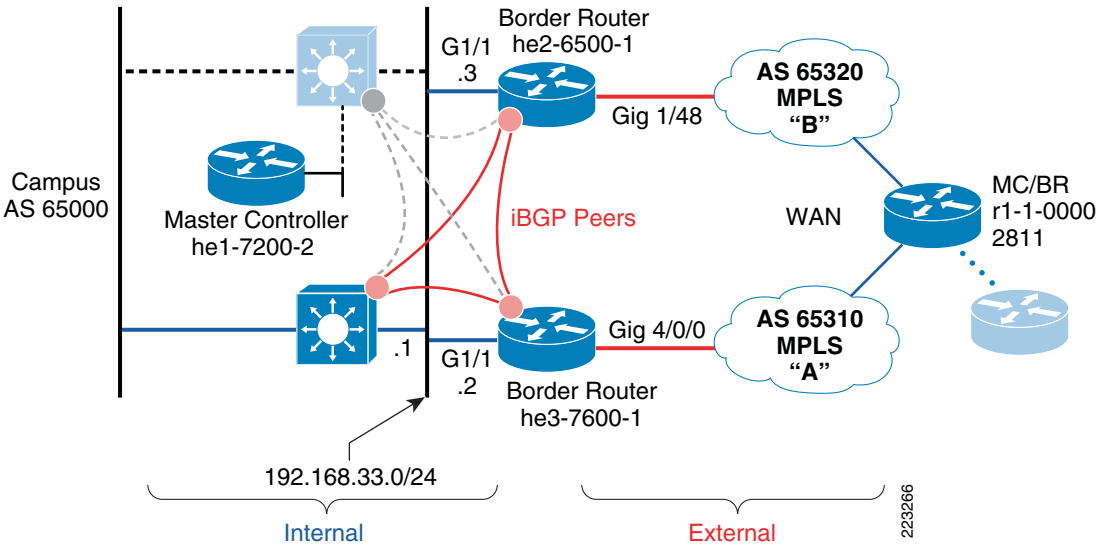
In the latency optimization section, the border routers are shown to learn and report on 2,500 prefixes every interval and report these learned prefixes to the master controller. The master controller then directs the border routers to insert routes in the BGP table to direct user traffic over the MPLS provider offering the lowest latency.

Topology

Figure 12 illustrates the topology being tested. The MPLS service providers each are assigned their own distinct autonomous system (AS) numbers (65310 and 65320 respectively) and are external BGP peers with the headend campus. The headend campus is assigned AS number 65000. One OER border router is a Cisco 6500 and the second border router is a Cisco 7600. A dedicated master controller is implemented at the headend campus. One representative branch router is shown, which is a collocated border router and master controller.

The campus switch is an iBGP peer with each border router. In a typical enterprise deployment, two campus switches would be implemented for high availability. This test deployment is not a campus high availability test; therefore, the subdued icon in the Figure 12 represents where this switch would be deployed in a actual implementation. It was not implemented in the lab test environment.

Figure 12 Scalability Testing Topology



Traffic Profile

The traffic profile for this test consisted of 1,000 real branches. Each branch is allocated a /24 address space. Each branch is connected to both MPLS service providers. An enterprise mix of VoIP and data traffic is used in the testing.

Software Release

Table 3 summarizes the Cisco IOS images used in testing.

Table 3

Hardware Platform	Cisco IOS Release (Image)
Cisco **7200VXR** (Master Controller)	c7200p-adventerprisek9-mz.124-15.T1
Cisco **CISCO7606** WS-SUP720-3BXL (Border Router)	c7600s72033-adventerprisek9-mz.122-33.SRB
Cisco **WS-C6506-E** WS-SUP720-3BXL (Border Router)	s72033-adventerprisek9-mz.122-33.SXH
Cisco **2811** ISR (Branch Router)	c2800nm-adventerprisek9-mz.124-9.T2

Load Sharing Performance Results

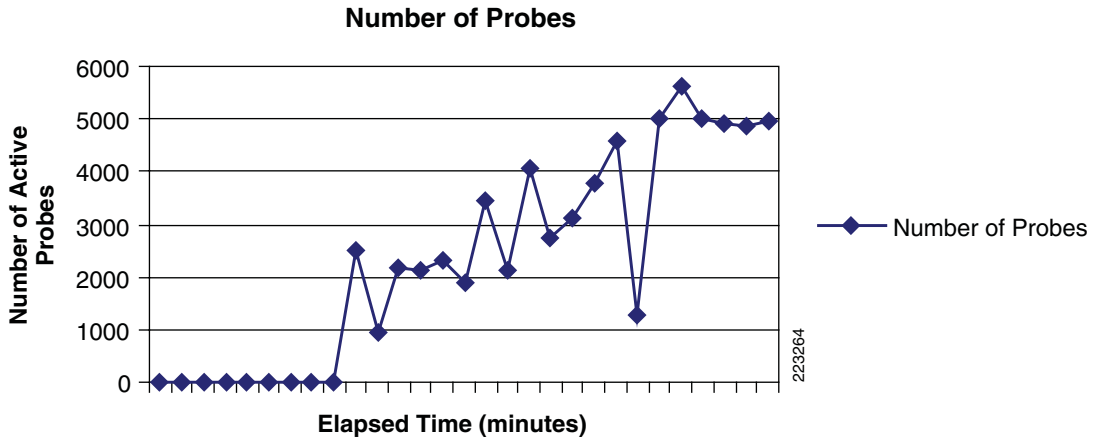
The results in this section demonstrate that active probing for 1,000 branch routers is within performance capabilities of the Supervisor 720 while also supporting other overhead activities like BGP peering. A load sharing test is used as the baseline. Latency optimization is addressed in the second test.

Active Probes per Minute

While there are 1,000 branch routers in the test, the number of active probes will increase above the total number of branch once traffic is learned from these branches through the NetFlow cache. Each branch has a /24 network address and the OER aggregation is configured on a /29 boundary. Because each branch contains simulated client workstations with distinct IP addresses, the number of probes is far greater than the number of branch routers.

Figure 13 plots the number of active probes throughout the test.

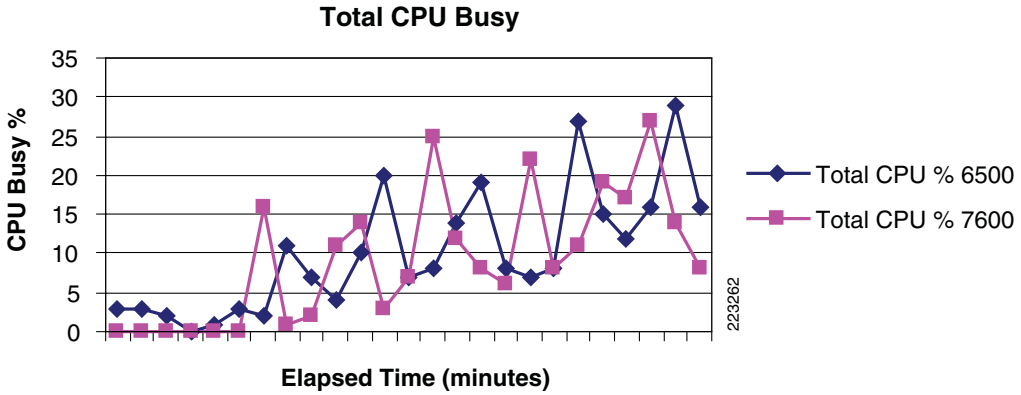
Figure 13 Load-Sharing Performance Results—Active Probes Per-Minute



Border Router CPU Utilization

For both border routers, the total CPU busy of the Supervisor 720 is plotted. Because the Cisco 6500 and 7600 are distributed systems, the main CPU is not invoked for switching every individual packet. Both border routers are also internal and external BGP peers. Note that the main CPU busy increases in a similar manner over the elapsed time of the test, and is generally consistent with the number of active probes. Because OER influences traffic in the BGP environment by inserting routes into the BGP table and advertising these changes to the iBGP peers, there will also be additional BGP workload. However, the CPU busy remains at or below 30 percent for both border routers.

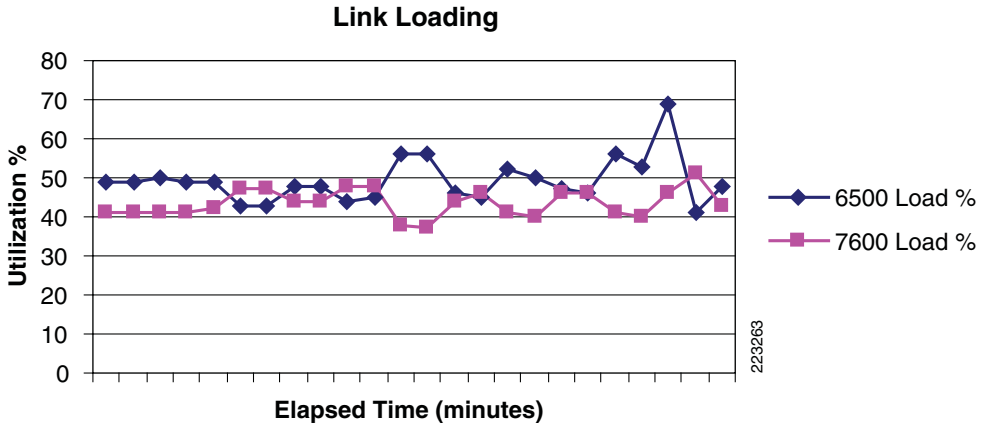
Figure 14 Load-Sharing Performance Results—Border Router CPU Utilization



Load Balancing Utilization Percentage

As proof that OER is functioning in a load-balancing mode, the external link loading of the two border routers is shown in Figure 14.

Figure 15 Load-Sharing Performance Results—Load-Balancing Utilization



Latency Optimization Performance Results

This section describes the performance results from the latency optimization test. Many customers are implementing dual-cloud MPLS networks. A Layer 3 MPLS network is deployed using two different MPLS service providers to provide greater availability in the event one of the service providers have a catastrophic network outage that disrupts the service offering to all branch offices.

As is often the case, as branches are implemented with connectivity over two MPLS providers, the customer discovers that one service provider network exhibits markedly different delay characteristics compared to the alternate service provider. Because of this, the desire is to use the service provider with the lowest delay as much as practical when both service providers are fully functional. This series of tests are designed to illustrate how OER can manage traffic over time to favor the better performing MPLS service provider.

Note Traffic may be routed asymmetrically, both the branch routers and the hub routers must be similarly configured and have OER enabled to optimize traffic flow in both directions.

The following sub-sections show the number of OER learned prefixes on the master controller during the hour long test. Also, the aggregate input and output traffic in megabits per-second is shown for each MPLS service provider.

Test Results Summary

The results in this section are from a controlled test with a duration of one hour. The test traffic is initiated and 5 minutes from the beginning of the test OER is enabled. At the beginning of the test, delay in MPLS network provider ‘B’ is minimal, 1 minute per-second (ms). The delay in MPLS provider ‘A’ is random between 30 and 50 ms throughout the hour-long test. After an elapsed time of 15 minutes, the delay in MPLS provider ‘B’ is increased to a random amount ranging from 80 to 120ms. In other words, the delay in MPLS provider ‘A’ is typical of an actual service provider and the delay for MPLS provider ‘B’ starts out as extremely low (ideal) and then is changed to much worse than MPLS provider ‘A’.

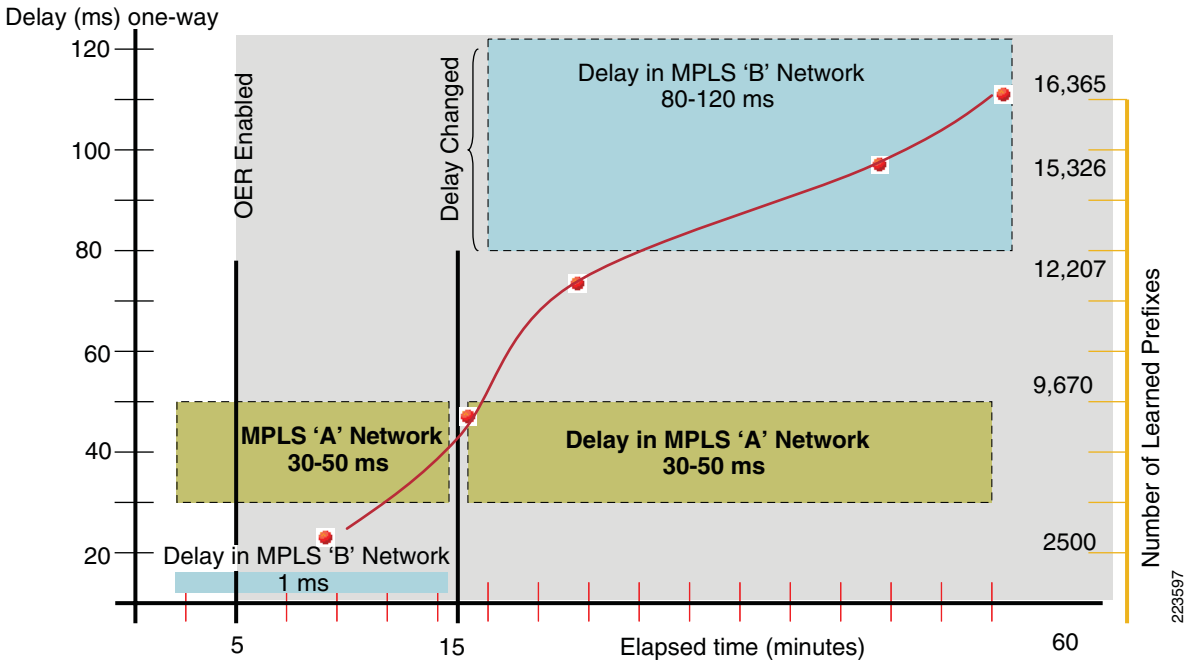
The test results demonstrate that the overall user traffic migrates over time to the MPLS provider offering the lowest delay of the two providers.

Number of Learned Prefixes

Figure 16 demonstrates the actual delay of the two MPLS service providers as it relates to the elapsed time of the test. Also shown in Figure 16 is the plot of network prefixes learned by the master controller. These prefixes are learned, rather than manually configured, and as such, the number of prefixes learned in any one interval is bounded by the prefixes 2500 learn mode sub-command. After the first interval 2,500 prefixes are learned, the value increases during subsequent intervals. The interval is set by the **monitor-period 2** command, meaning two minutes per learn interval. The aggregation (**aggregation-type prefix-length 29**) is configured with a /29 prefix length. Each remote branch is allocated a /24 address on the inside VLAN. There are approximately 1,000 real branch routers in this test.

Over the duration of the test, 16,365 network prefixes are learned. Figure 16 shows delay of the two MPLS providers as the primary Y-axis and Number of Learned Prefixes as the alternate Y-axix. Elapsed time is the X-axis.

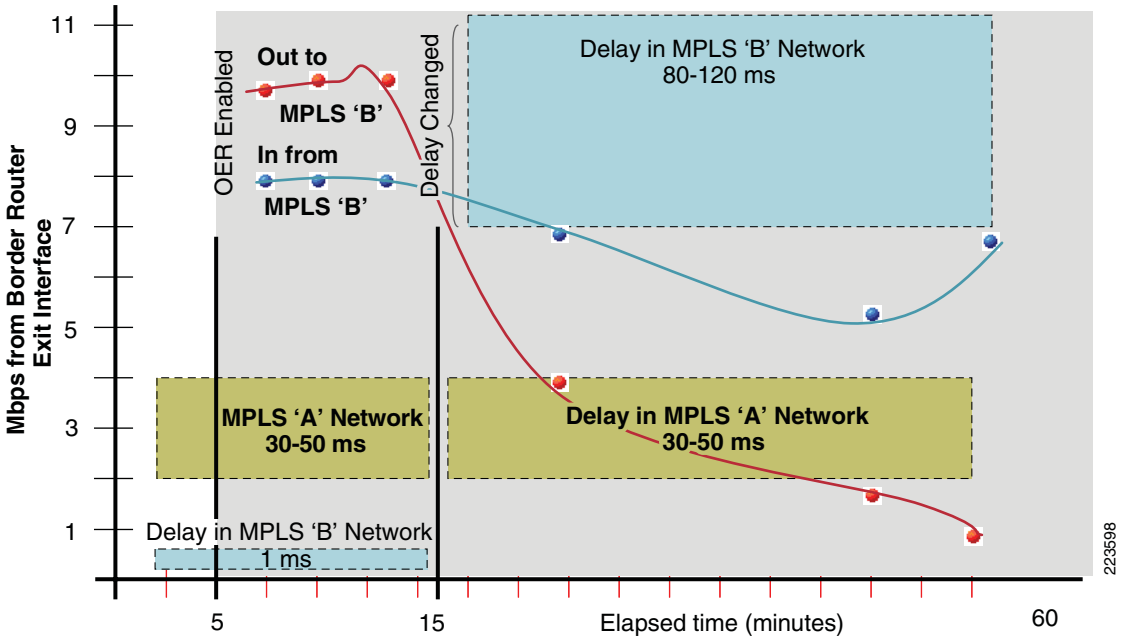
Figure 16 Latency Optimization Performance Results



Network Traffic Pattern for MPLS Provider 'B'

The chart in Figure 17 plots the input and output in megabits per-second for the exit interface of the Cisco Catalyst 6500 chassis that is connected to MPLS provider 'B'. As the delay characteristics of MPLS provider 'B' is changed during the test, the amount of both input and output traffic serviced by this border router decreases. In other words, the amount of network traffic to and from MPLS provider 'B' decreases as the latency of the provider increases. This is clearly illustrated inFigure 17.

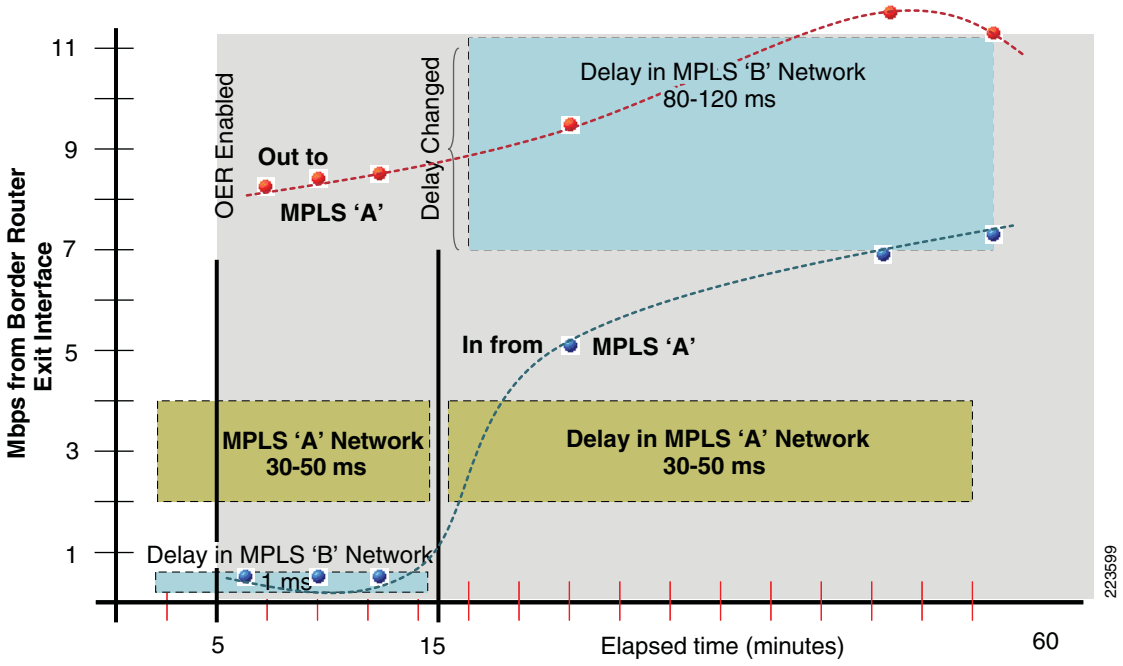
Figure 17 Network Traffic Pattern for MPLS Provider 'B'



Network Traffic Pattern for MPLS Provider 'A'

Now looking at the Cisco 7600 border router that is attached to MPLS provider 'A' (see Figure 18), clearly as the delay of MPLS provider 'B' worsens in relation to the consistent 30 to 50ms delay of MPLS provider 'A. This border router's exit interface encounters an increased amount of both input and output traffic.

Figure 18 Network Traffic Pattern for MPLS Provider 'A'



Recall that on both charts (Figure 17 and Figure 18), the amount of traffic input to the headend campus border routers is a function of the routing and OER configuration of the branch routers. Also, the amount of traffic output to each MPLS providers, at the beginning of the test, is governed by the IP routing table, and because the network prefixes are learned, the ideal configuration provides at least two equal cost routes in the campus core routing table so both providers are equally loaded. Prior to OER being enabled, approximately 8 to 9 Mbps of traffic exits the headend campus for each MPLS provider.

Tested Configuration

This section contains the sample configuration file used in the test environment. This example is a reasonable illustration for a customer use case; however, it is not intended to be used verbatim on all possible network topologies.

Master Controller

The master controller configuration is shown in both the load-balance configuration and the latency optimization configuration.

```
!
hostname he1-7200-2
!
!
key chain NGWAN
 key 10
   key-string cisco
!
!
```

Load Balance Configuration

```
    !
    oer master
     ! shutdown
     max-range-utilization percent 5
     ! no keepalive
     !
     border 192.168.33.2 key-chain NGWAN
      interface GigabitEthernet1/1 internal
      interface GigabitEthernet4/0/0 external
       max-xmit-utilization percentage 50
     !
     border 192.168.33.3 key-chain NGWAN
      interface GigabitEthernet1/1 internal
      interface GigabitEthernet1/48 external
       max-xmit-utilization percentage 50
     !
     learn
      throughput
      protocol tcp
      protocol udp
      periodic-interval 0
      monitor-period 2
      prefixes 2500
      expire after time 65000
      aggregation-type prefix-length 29
     max prefix total 100000 learn 100000
     mode route control
     mode select-exit best
     periodic 180
     resolve range priority 1
     resolve utilization priority 2 variance 10
    !
    !
    end
```

Latency (Delay) Optimization

```
    !
    oer master
```

```
 ! shutdown
 no max-range-utilization
 ! no keepalive
 !
 border 192.168.33.2 key-chain NGWAN
  interface GigabitEthernet4/0/0 external
  interface GigabitEthernet1/1 internal
 !
 border 192.168.33.3 key-chain NGWAN
  interface GigabitEthernet1/48 external
  interface GigabitEthernet1/1 internal
 !
 learn
  throughput
  protocol tcp
  protocol udp
  periodic-interval 0
  monitor-period 2
  prefixes 2500
  expire after time 65000
  aggregation-type prefix-length 29
 max prefix total 100000 learn 100000
 delay threshold 100
 jitter threshold 9
 mode route control
 mode select-exit best
 periodic 180
 resolve delay priority 1 variance 10
 resolve jitter priority 2 variance 10
 no resolve utilization
!
!
end
```

Border Router (6500)

This border router configuration is used in both test methods and is implemented on the Cisco Catalyst 6500 chassis.

```
!
hostname he2-6500-1
!
key chain NGWAN
 key 10
   key-string cisco
!
oer border
 logging
 local GigabitEthernet1/1
 master 10.208.0.2 key-chain NGWAN
 active-probe address source interface GigabitEthernet1/48
!
interface GigabitEthernet1/1
 description To Campus GigabitEthernet1/1
 ip address 192.168.33.3 255.255.255.0
!
interface GigabitEthernet1/48
 description Inside
 bandwidth 20000
 ip address 192.168.188.3 255.255.255.248
!
router bgp 65000
```

```
 no synchronization
 bgp log-neighbor-changes
 bgp bestpath as-path multipath-relax
 redistribute connected
 neighbor 192.168.33.1 remote-as 65000
 neighbor 192.168.33.1 next-hop-self
 neighbor 192.168.33.2 remote-as 65000
 neighbor 192.168.188.1 remote-as 65320
 maximum-paths 8
 maximum-paths ibgp  4
 no auto-summary
!
```

Border Router (7600)

This border router configuration is used in both test methods and is implemented on the Cisco 7600 chassis.

```
!
hostname he3-7600-1
!
key chain NGWAN
 key 10
  key-string cisco
!
oer border
 logging
 local GigabitEthernet1/1
 master 10.208.0.2 key-chain NGWAN
 active-probe address source interface GigabitEthernet4/0/0
!
interface GigabitEthernet1/1
 description To Campus GigabitEthernet1/1
 ip address 192.168.33.2 255.255.255.0
!
interface GigabitEthernet4/0/0
 bandwidth 20000
 ip address 192.168.160.114 255.255.255.248
!
router bgp 65000
 no synchronization
 bgp log-neighbor-changes
 neighbor 192.168.33.1 remote-as 65000
 neighbor 192.168.33.1 next-hop-self
 neighbor 192.168.33.3 remote-as 65000
 neighbor 192.168.160.113 remote-as 65310
 no auto-summary
!
end
```

Branch Router

This is a branch router configuration representative of one of the 1,000 branch routers in this test.

```
!
hostname r1-1-0000
!
!
key chain NGWAN
```

```
 key 10
  key-string cisco
!
!
oer master
 logging
 !
 border 10.192.0.1 key-chain NGWAN
  interface FastEthernet0/0.3300 external
  interface FastEthernet0/0.2200 external
  interface FastEthernet0/1 internal
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  aggregation-type prefix-length 29
 delay threshold 100
 mode route control
 mode select-exit best
 periodic 180
 !
!
oer border
 logging
 local FastEthernet0/1
 master 10.192.0.1 key-chain NGWAN
!
interface FastEthernet0/0
 description do1tQ Trunk
  . . .
 service-policy output PER_CLASS_3mb
!
interface FastEthernet0/0.2200
 description MPLS A
 encapsulation dot1Q 2200
 ip address 192.168.144.2 255.255.255.248
!
interface FastEthernet0/0.3300
 description MPLS-B
 encapsulation dot1Q 3300
 ip address 192.168.180.2 255.255.255.252
!
end
```

Summary

This deployment of 1,000 branch routers demonstrates how OER can be implemented on both the branch routers and headend campus routers in a dual MPLS provider WAN topology to implement performance routing across the service provider with the best delay characteristics. These results demonstrate the use of the *mode monitor special* option, where active probing is always enabled to provide reachability information for path rerouting when Cisco 7600 and 6500 chassis are implemented as border routers.

The use of OER in this deployment is a risk mitigating technique, as the user traffic will take the WAN offering the best performance from two or more possible service providers. This is implemented without continual monitoring and reporting by network management tools and manual adjustments to the routing

tables by the network operations staff. The cost to implement is minimal, with this 1,000 branch deployment only requiring the installation of a dedicated master controller chassis on the headend campus.

Branch/SOHO VPN Deployment

This deployment topology consists of a single router supporting a branch office with two tunnels to the campus hub location. The IPSec encrypted point-to-point GRE tunnels/physical access links traverse a WAN that have separate and distinct latency characteristics on the two paths. OER is used to implement performance routing on both the branch to hub and hub to branch paths. VoIP is the primary target application for optimization in the test results, however other applications, including mission critical TN3270 sessions, are also present.

The goal is to demonstrate that the key performance metrics (latency, jitter) of a VoIP stream are influenced as OER implements policy routing of the application to optimize the quality of the voice call. Latency, and the resulting jitter, are present at the start of the test, degrading the quality of one path verses the other. During the test, the latency parameters are changed, showing how OER tracks these changes and adapts to the new environment.

Chariot/IXIA is used to quantify the characteristics of the VoIP G.729 RTP streams through graphs showing latency and jitter of the individual Real Time Protocol (RTP) streams transporting the VoIP call. The graphs show data points captured for all calls over the 30-minute test.

The branch topology in this test is a single branch router running both MC and border router functions. Both active and passive monitoring is configured to provide network performance metrics to the control function.

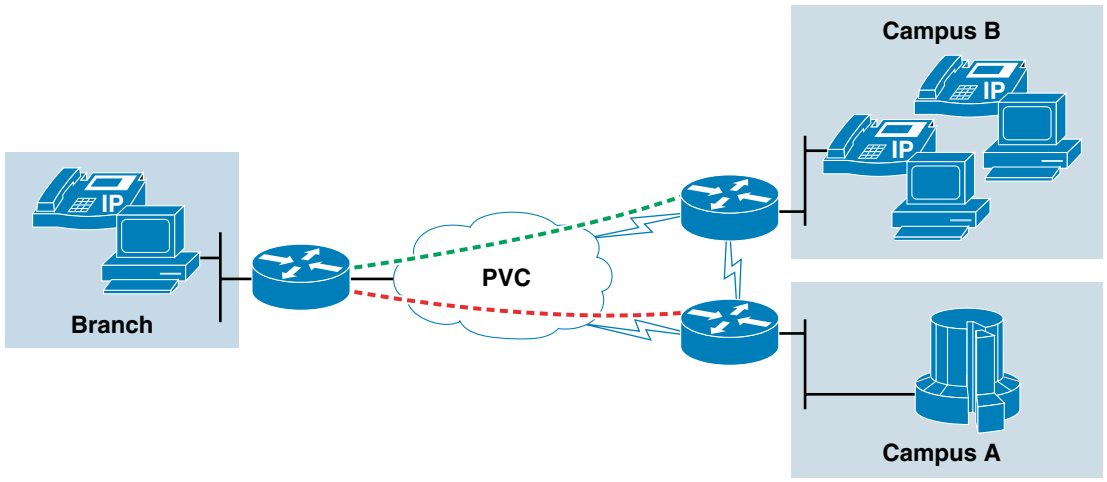
Design Requirements and Considerations

The focus of this section is targeted at a branch or small office/home office deployment. In this deployment scenario, cost of both the equipment and access links, as well as the number of remote offices involved in the typical deployment often mandate a design using a single router and physical access circuit. However, the design is also applicable to multiple access circuits to separate WAN providers that may introduce different degrees of latency, loss, and jitter.

Regardless if the branch is served by one or two access circuits, this section demonstrates how an application, VoIP in this example, can be optimized by forwarding the application over the logical WAN link with the lowest latency.

Even with a single router and access circuit, it may be advantageous for an application to use one logical tunnel verses the other. Assume the topology in Figure 19.

Figure 19 Small Office/Home Office Deployment Example



In this example shown in Figure 19, Campus A supports the organization's mainframe and other supporting servers. The Campus B location houses the majority of the office space and the IP phones. The two campus locations are connected by one or more private WAN links. Because of this, from the perspective of the branch router, both logical WAN link can be used for voice and data traffic. If both WAN links have equal cost and both are inserted into the IP routing table (as would be the case if equal cost default routes are the only routes advertised to the branch), some of the VoIP traffic will traverse the link to Campus A to reach Campus B.

If a default route is advertised from the Campus A hub and more specific subnets are advertised from Campus B, the VoIP traffic always traverses the link to Campus B, unless that logical link is down due to a failure condition. In that event, the Campus A link is the surviving link and all traffic uses the single link.

However, these methods of providing availability to the branch are solely based on availability and network reachability and not the performance characteristics of the individual links. This section demonstrates how OER can be implemented on the branch router to re-route VoIP traffic, based on the delay characteristic data obtained from IP SLA probles. In this example, an oer-map is configured to identify, by the ToS byte of the UDP packets, VoIP as an application and use an explicitly configured active (IP SLA) probe to determine the characteristics of the WAN links.

VoIP is then routed out the link which provides the best end-user experience for this application.

Design Limitations

OER must have a parent route for managing an application or network prefix. At this time, BGP and static routes fulfill this parent route requirement. Software development is underway to implement EIGRP/OSPF routes as parent routes. This enhancement is tracked by the CSCsk39768—PfR-EIGRP integration and CSCsm34644—PfR-OSPF integration.

OER is not supported on DMVPN network topologies at this time. DMVPN can be defined on mGRE interfaces that are either point-to-point or point-to-multipoint. For DMVPN network to be supported, CSCsk39768—PfR-EIGRP integration and CSCsm34644—PfR-OSPF integration is required as well as the software enhancement CSCsi69186—PfR-DMVPN/mGRE integration.

Because of these limitations, this section implements encrypted tunnels using IPSec encrypted GRE tunnels. GRE keepalives can be enabled in this configuration, and as such, static routes referencing the tunnel interface may be used for parent routes. Because lost GRE keepalives is used to bring down

(interface state as UP/DOWN) the GRE interface, static routes in the routing table are removed when the GRE keepalives are lost and the interface transitions to an UP/DOWN state. Static routes referencing the next hop IP address of the point-to-point GRE tunnel are also supported and would also be removed from the routing table in the event the interface transitions to an UP/DOWN state.

If encryption is not required, OER can also be implemented on a broad range of physical interfaces. The test results presented in this section are also applicable to environments without encryption.

Delay is used as a metric to deduce VoIP quality rather than MOS score. As part of this testing, an issue was identified when tracking MOS. For tracking purposes, the defect *CSCsl10489 (OER ignores MOS score of jitter probe when inpolicy timer expires)* can be reviewed for more information.

Scalability Considerations

The purpose of this deployment example is to illustrate the positive influence OER asserts on VoIP quality from the branch router perspective. As the number managed applications and network prefixes are well within the memory limitations documented in the Internet Content Server section, there are no concern regarding scalability of this deployment for a branch router.

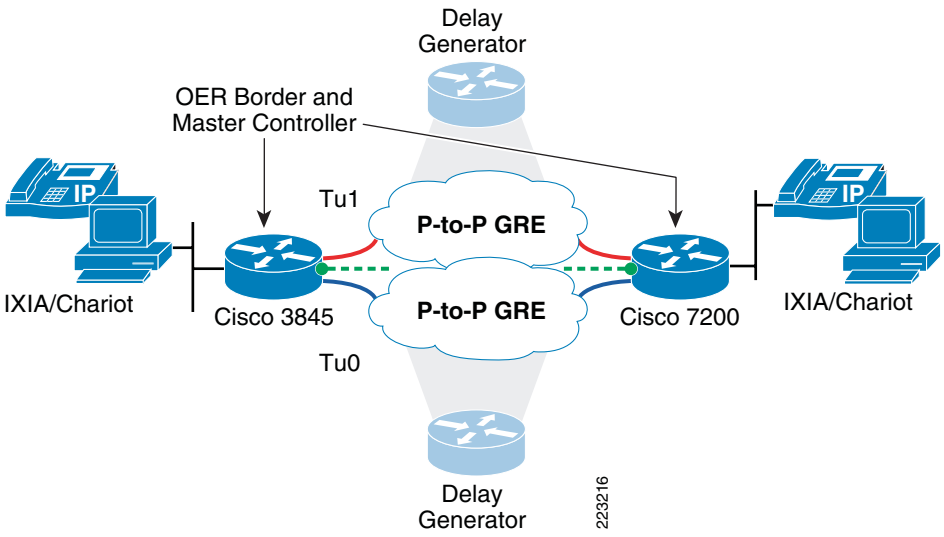
Because EIGRP/OSPF and DMVPN interfaces are currently not supported, this deployment model is most likely associated with a site-to-site model rather than a hub and spoke or large scale IPSec aggregation deployment model. When OER can be implemented on hub routers with large numbers of remote branches in a DMVPN cloud, headend scale issues need to be quantified. Currently this is outside of the scope of this design guide.

Topology

The test topology shown in Figure 20 is a branch and campus location with two IPSec encrypted point-to-point GRE tunnels traversing some underlying WAN topology. Because the physical WAN topology is hidden or masked by the crypto tunnels, it is not relevant for the purposes of this testing. In this test, a simulated Metro Ethernet WAN is used. QoS is enabled on the physical interfaces. It is assumed that the QoS policy is configured to support all VoIP calls in the priority queue of a single physical path.

In the test topology (see Figure 20), Tunnel 0 is always transported over a single path, and Tunnel 1 is always transported over the second path. Each path has a traffic delay generation tool inserted in the path to emulate the delay associated with an actual WAN deployment. For the purpose of demonstrating the functionality of OER, varying magnitudes of delay is introduced at differing times during the test to simulate the effects of one WAN link experiencing degraded performance verses the other link. Changes in delay is indicative of some shared WAN transport like a Layer 2 (pseudowire) or Layer 3 MPLS service or Internet transport. Encrypting traffic over the Internet or MPLS service provider is a common deployment scenario as described in the *Voice and Video Enabled IPsec VPN (V3PN) SRND*, available at www.cisco.com/go/srnd.

Figure 20 Campus/Branch Location with P-to-P GRE



Delay Generation

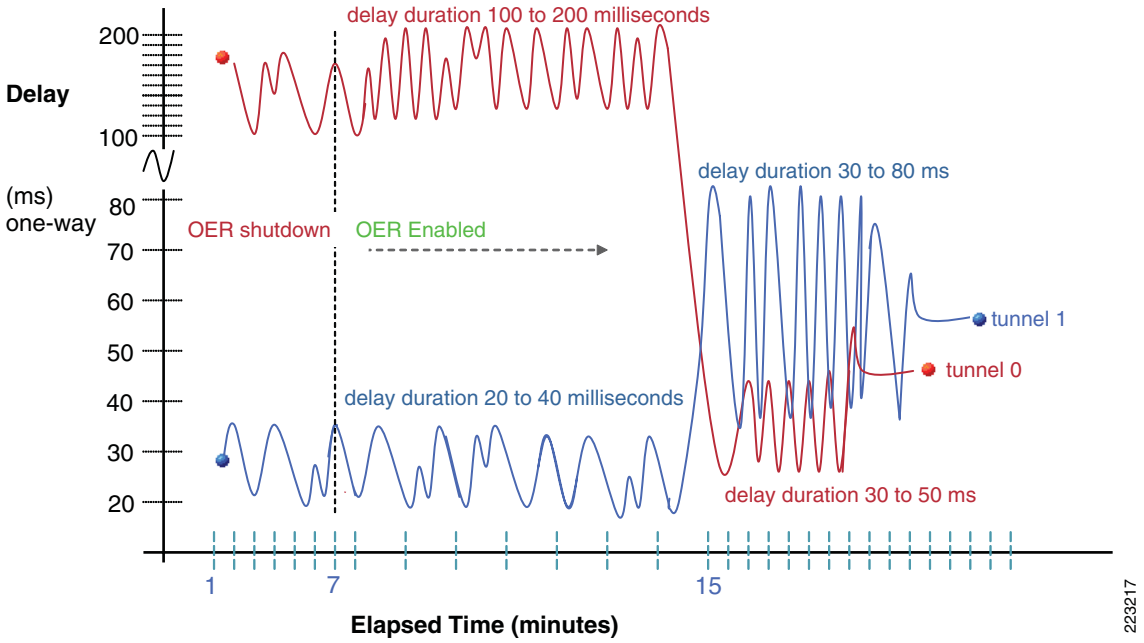
In this test, delay is introduced on the WAN link associated with the respective tunnels between the branch and headend. The OER master controller is in a shutdown state at the beginning of the test. Delay on Tunnel 0 is random for each packet, ranging between 100 and 200 milliseconds with packets kept in sequence. Delay for tunnel 1 is between 20 and 40 milliseconds for each packet with the packet sequence maintained.

At an elapsed time of 7 minutes from the beginning of the test, the OER master controller is enabled (**no shutdown**). At 15 minutes elapsed time, the delay on both tunnels is changed. Tunnel 0 is changed to 30-50ms (an improvement in delay) and Tunnel 1 is changed to 30-80ms (a degradation in delay) , with packets kept in sequence.

Note When a random range of delay is introduced with packets kept in sequence, the average value for delay tends toward the upper configured value. Introducing random delay also introduces jitter, as jitter is simply the difference between the theoretical inter packet interval and the actual observed. If random delay is introduced and packets are allowed to be re-ordered, VoIP packets will incur loss, sometimes approaching 10% or more, as the jitter buffer of the receiving VoIP simulator or device may discard packets that arrive too late. In this case, loss is introduced by the nature of the application, rather than by the network devices transporting the VoIP stream.

This delay associated with the respective tunnels is illustrated in Figure 21.

Figure 21 Delay Generation



In the following section, the impact of enabling OER, and as it accumulates statistics through active probing, the resulting managing of VoIP as the target application, is apparent in the test results.

In the next section, OER is implemented on the links with these delay parameters, and it is demonstrated that the VoIP RTP streams will track with the links experiencing the lowest delay.

VoIP Quality Verification

This branch test included 17 pairs of RTP (VoIP) streams between the branch router and the campus headend router. Two paths are available (two exit points are required for OER to function) and each link and the associated IPSec/GRE tunnel has QoS applied to the physical interface. Data traffic is also present in the test, with this traffic being managed by the ***learn*** configuration in the global OER master configuration.

Because OER requires parent routes in the routing table, there are two equal cost static routes in the routing table. Load sharing is based on CEF switching on a per source/destination IP address during the period OER is shutdown or when OER has the application in the DEFAULT state.

In reviewing the following charts, OER is enabled only after the Chariot test has initialized. Also the delay parameters on both tunnels goes through a transition at the mid-point of the 30 minute test.

http://harry/solutions/vpn/results/ngwan30/oer/branch/1branch2exits/chariot-BHonly-3072kb-branch0.html

One-Way Delay

OER views delay in the context of a round trip delay and thresholds (**set delay threshold**) values are specified as round trip. From the prospect of VoIP quality, delay is normally shown as one-way values. VoIP simulation tools, like IXIA/Chariot, report delay for each call leg, meaning one-way delay. OER implementations where OER is configured both on the branch and headend, will over time, tend toward symmetrical routing, if the goal is to manage for delay and the paths differ sufficiently to have a clearly better path between the two. There may be transition periods where the routing is asymmetrical, or if the two links have similar characteristics there may also be asymmetrical routing in an ongoing basis.

In the chart shown in Figure 22, several points are to be noted. First, because there are two equal cost paths in the routing table, some VoIP streams use Tunnel 0 and some use Tunnel 1. This is a result of CEF load sharing over the two paths. Because OER is not enabled at the start of the test, and given the delay characteristics vary considerably between the two paths, one set of VoIP calls have a delay approaching the 193ms legend on the chart and the remaining calls have a delay plotted slightly above the 33ms legend. As such, call quality and usability varies depending on the tunnel transporting the call.

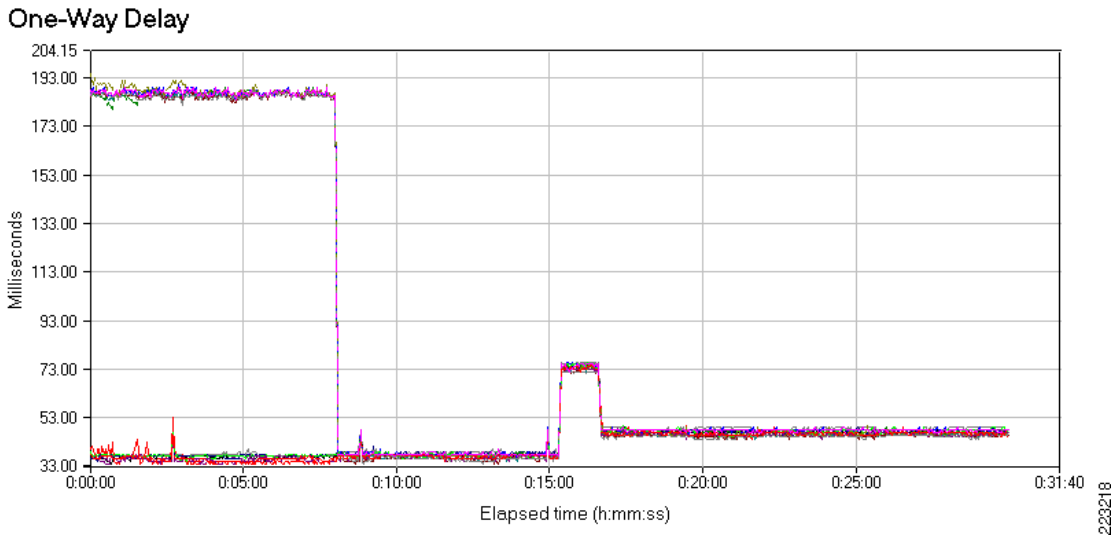
The goal for delay is the generally accepted ITU target of 150 ms one-way, while ideally less than 100ms is most desirable. The delay generator is configured to introduce delay exceeding the 150ms target while the best link has a delay below the 100ms value.

The chart in Figure 22 shows that once OER is enabled on the branch (also on the headend, but this illustration is branch centric), the RTP streams which are experiencing the excessive delay are managed by OER such that they are routed over the exit with the minimal delay. It takes approximately one minute from the time OER is enabled until the network characteristics can be quantified by the active probes and acted upon. In this case, the VoIP is identified by an access-list match on DSCP CS5 (ToS byte 0xA0, decimal 160) for UDP packets between any source and destination addresses.

In the Internet Content Server section, OER managed IP network prefixes by inserting routes in the BGP table, resulting in IP routing table entries. In this configuration, OER is managing the applications through policy routing the application data packets. Policy routing need not be configured manually by the network manager, OER automatically applies or removes policy routing as required.

Figure 22 shows the delay experienced by the RTP streams as OER is enabled, the path changes, and the delay is manually changed at the mid-point of the 30-minute test.

Figure 22 One-Way Delay



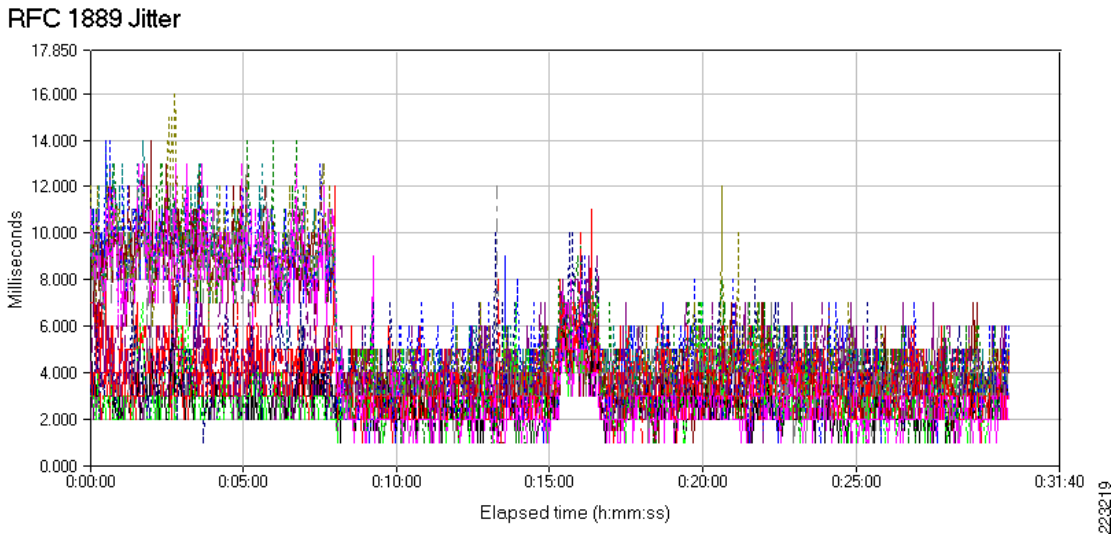
In the period between 8 minutes and 15 minutes, all VoIP streams exhibit a delay characteristic slightly above the 33ms mark. At 15 minute elapsed time, the delay generator is reconfigured to simulate a change in the network characteristics of the WAN links. Now the delay on Tunnel 0 is reduced and Tunnel 1 is increased slightly. Effectively Tunnel 0 is now a better path than Tunnel 1, however both tunnels are below the goal of 100ms.

In Figure 22, note that because OER is actively managing the VoIP traffic for all pairs through Tunnel 0, the delay of all VoIP pairs increases along with the change in delay of the Tunnel 0 path. However, because OER is configured to actively monitor using an explicitly configured jitter probe, the path is changed, through a policy route change, to use Tunnel 1 as the preferred exit.

Jitter

This section includes the Chariot generated RFC 1889 Jitter graph that corresponds with the delay graph in the previous section. The time when OER is enabled, as well as the OER path changes and when the delay characteristics of the two tunnels is changed, is indicated on the graph in Figure 23.

Figure 23 Jitter



In this OER test, jitter is not manually configured to be a policy priority in the oer-map and jitter is not a default priority. Only delay and utilization are configured by default. The output from the **show oer master policy** command is shown:

```
#show oer master policy BRANCH
...
 *resolve delay priority 1 variance 10
 *resolve utilization priority 12 variance 20
```

However, in most networks as delay increases, jitter also increases. It is apparent, from the above RFC 1889 Jitter chart, that managing for delay also positively influences jitter for the RTP streams.

Note In lab testing using a Frame Relay WAN, it was found that if jitter is less than 8ms, voice quality is acceptable as it is in a similarly configured real-world deployment. The Chariot test tool was used to report the jitter information.

As Figure 23 shows, as the VoIP streams are managed for delay, jitter is also brought within the 8ms target.

Configuration Examples

This section contains the sample configuration file used in the test environment. This example is a reasonable illustration for a customer use case, however it is not intended to be used verbatim on all possible network topologies. Note that comments and clarifications regarding the config are included in-line with the configuration.

```
!
oer master
 policy-rules BRANCH
 ! shutdown
 ! no keepalive
 logging
 !
 border 10.192.0.1 key-chain NGWAN
  interface GigabitEthernet0/1 internal
  interface Tunnel1 external
  interface Tunnel0 external
 !
```

Tip The data traffic (TN3270, SMTP, FTP, HTTP, etc.) is not selected by the oer-map and; it is learned and selected for management by the global section of the oer master configuration, shown below which.

```
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  aggregation-type prefix-length 29
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
```

Note As this is a single router deployment example, the OER master controller and border router are configured on the same router.

```
oer border
 logging
 local GigabitEthernet0/1
 master 10.192.0.1 key-chain NGWAN
!
```

Note Parent IP routes are required. In this case they are to the tunnel interfaces, and as these tunnels are IPSec encrypted GRE tunnels, GRE keepalives are configured to remove the parent routes from the routing table in the event of a path failure.

```
ip route 0.0.0.0 0.0.0.0 Tunnel0
ip route 0.0.0.0 0.0.0.0 Tunnel1
!
```

In the following example, VoIP is identified as CS5, equal to IP Precedence 5:

```
ip access-list extended VOICE
 permit udp any any dscp cs5
!
```

Note The VoIP application is referenced by means of the voice access-list. This example controls the application using active probles, and an explicit active probe, a jitter probe, is shown configured. Delay has the highest priority for selecting the best exit.

```
!
oer-map BRANCH 10
 match traffic-class access-list VOICE
 set delay threshold 250
 set mode route control
 set mode monitor active
 set resolve delay priority 1 variance 10
 set active-probe jitter 10.204.0.1 target-port 33333 codec g729a
 set probe frequency 30
!
end
```

Troubleshooting

This section includes some show commands and other examples that may be useful as a reference for a customer deploying a similar configuration on their network.

show oer master appl detail

This section shows the command output of the **show oer master appl detail** command for the VoIP RTP application. This application is identified by the oer-map referencing the voice access-list in the traffic-class reference. In the access-list configuration, VoIP is matched by IP Precedence of 5 (DSCP CS5) for UDP packets between any source and destination IP addresses. This is one simple method of identifying VoIP, but any access-list that selects this or other applications is applicable.

The following access-list reference is used in testing:

```
!
ip access-list extended VOICE
 permit udp any any dscp cs5
!
oer-map BRANCH 10
 match traffic-class access-list VOICE
 ...
!
```

There are several items in this display that bear review and clarification, including:

- DSCP value of 160 is a reference to the 8 bits of the ToS byte shown in decimal, where 160 is 0xA0 or binary 10100000 which is IP Precedence 5.
- Protocol 17 is UDP.
- The current state is HOLDDOWN, the previous state was DEFAULT*, this state change is due to the delay differences between the two exits.
- The time remaining in HOLDDOWN is 297 seconds. This display is shown 3 seconds after transitioning to this state; by default, HOLDDOWN duration is 300 seconds or five minutes.
- The current exit point is Tunnel 1, with an active short and long term delay of 44ms; Tunnel 0 has a reported active short and long term delay of 198ms.
- Because an explicitly configured active probe, a jitter probe (**set active-probe jitter**), is configured, OER can obtain and store jitter and MOS data as well. However, only delay and utilization are configured in the policy. Jitter and MOS are not used in this test to determine the exit selection. Jitter is shown at 4 and 8ms, where Tunnel 1 has the lowest jitter as well as the lowest delay.

```
show oer master appl detail
Prefix: 0.0.0.0/0  Protocol: 17  Port: [1, 65535] [1, 65535] DSCP: 160
   State: HOLDDOWN    Time Remaining: 297
   Policy: 10

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
  *10.192.0.1      Tu1                     0        0       44       44
   10.192.0.1      Tu0                     0        0      198      198

   Most recent voice data per exit
   Border          Interface         ActSJit  ActPMOS  ActSLos  ActLLos
  *10.192.0.1      Tu1                     4        0        0        0
   10.192.0.1      Tu0                     8        0        0        0

   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min     Max     Dly
   jitter   10.204.0.1      33333     2   200    8991      24      64      44


   Latest Active Voice Stats on Current Exit:
   Type     Target          TPort     Codec Attem Comps  JitSum      MOS
   jitter   10.204.0.1      33333     g729a     2   200     993     4.06


... [output truncated] ...
```

When managing either prefix or application data, use the **show oer master prefix** or **show oer master appl** commands to better understand the data points OER is using to make decisions regarding network prefixes or applications.

Syslog File

The following syslog output illustrates the logging messages generated by the OER master controller during the testing. These events occur:

- The master controller goes active (**no shutdown** command issued in master controller configuration mode)

- The VoIP application is recognized through the **policy-rules BRANCH** oer-map reference in the configuration.
- Because **set mode monitor active** command is configured along with an explicit probe, all exits are probed.
- The best exit based on delay is Tunnel1.
- Forwarding of the VoIP traffic through policy routing is enabled to Tunnel 1
- The application state for this application traffic is in HOLDDOWN for (300 seconds) five minutes.

```
Oct 25 16:35:35.721 EDT: %OER_MC-5-NOTICE: MC Active
he1-3800-2#
Oct 25 16:35:35.721 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: PDP start timer = 51 secs, prefix state = DEFAULT*
Oct 25 16:35:35.733 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: Prefix timeout, state DEFAULT*
Oct 25 16:35:35.737 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: APC Attempting to probe all exits
he1-3800-2#
Oct 25 16:36:05.937 EDT: %OER_MC-5-NOTICE: Prefix Learning STARTED
Oct 25 16:36:05.937 EDT: %OER_BR-5-NOTICE: Prefix Learning STARTED
he1-3800-2#
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: Prefix timeout, state DEFAULT*
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: PDP choose exit, prefix state = DEFAULT*, 0
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: Check ACT REL unreachable: unreachable 0, policy
50%, notify FALSE
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: Check ACT ABS delay: delay 34, policy 250, notify
FALSE
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: Best exit is 10.192.0.1 Tu1, based on delay
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: Start FWD on new exit, br = 10.192.0.1, i/f =
Tu1, nexthop 0.0.0.0, seq 18663, proto 8, exact TRUE
Oct 25 16:36:27.137 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: PDP start timer = 15 secs, prefix state = CHOOSE
Oct 25 16:36:27.337 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: prefix_status 0 received, br = 10.192.0.1 i/f = Tu1
Oct 25 16:36:27.337 EDT: OER MC APPL0.0.0.0/0 Prot 17 Port [1, 65535]
[1, 65535] ToS 160: PDP start timer = 300 secs, prefix state = HOLDDOWN
he1-3800-2#
Oct 25 16:36:27.337 EDT: %OER_MC-5-NOTICE: Route changed Appl Prefix
0.0.0.0/0 cs5  17 [1, 65535] [1, 65535], BR 10.192.0.1, i/f Tu1,
Reason None, OOP Reason Timer Expired
```

Policy Routing of Application(s)

When OER is policy routing application data, the **show route-map dynamic detail** command may be issued to show matches on this dynamically created route-map:

```
he1-3800-1#show route-map dynamic detail
route-map OER-11/01/07-19:28:11.592-5-OER, permit, sequence 0, identifier 1734834188
  Match clauses:
    ip address (access-lists): oer#1
      Extended IP access list oer#1
          1073741823 permit udp any any dscp cs5
  Set clauses:
```

```
    interface Tunnel3
    ip next-hop 10.56.12.1
  Policy routing matches: 0 packets, 0 bytes
Current active dynamic routemaps = 1
```

The route-map and policy routing statements do not show in the running configuration file.

Summary

This section demonstrates the ability of OER to manage applications, VoIP for example, by measuring the characteristics of available links and dynamically re-routing the application to optimize the managed application. In the case of VoIP, selecting an available exit with the lowest delay also reduces jitter, which can improve the audio fidelity of a VoIP call. Reducing delay for a VoIP call improves the usability of the call by reducing the walkie-talkie effect. The walkie-talkie effect occurs where there is high delay on a VoIP call, leading to the likelihood the parties on the call will talk at the same time.

Branch VPN Deployment with Cisco Wide Area Application Services (WAAS)

This deployment builds on the Branch/SOHO VPN Deployment deployment model with the addition of the Cisco WAAS at the branch location. Because of the costs associated with including the Cisco WAAS and the low cost, high bandwidth broadband access links available to the home, a SOHO deployment is not addressed. Therefore, this is targeted at a mid-range to large branch deployment with dual-access routers and links.

The primary objective is to demonstrate OER and WAAS compatibility. A secondary objective is to demonstrate a primary/standby MC configuration. As this test includes dual access routers and links, one access router is a primary MC and BR with a single IPSec/GRE tunnel. The second access router is a standby MC and BR with two IPSec/GRE tunnels. During the test, the primary MC router is failed and it is demonstrated the standby MC can resume OER functions over the remaining two exits.

This deployment topology consists of a dual router supporting a branch office, each with two tunnels to the campus hub location. The IPSec encrypted point-to-point GRE tunnels/physical access links traverse a WAN, which may have separate and distinct latency characteristics on the paths OER is used to implement performance routing on both the branch-to-hub and hub-to-branch paths..

The goal of this section is to build upon the topology in the Branch/SOHO VPN deployment by:

- Increasing availability by adding a second branch router.
- Demonstrating an active/standby master controller on dual border routers.
- Adding Wide Area Application Services (WAAS)
- Demonstrating WAAS survivability and recovery from a single branch router failure.

WAAS implements a set of caching, compression, and TCP message optimization techniques to reduce WAN link utilization and decrease application latency. This improves the overall application response for branch users accessing centralized data center applications over the WAN.

As in Branch/SOHO VPN Deployment, VoIP is the primary target application for optimization by OER; however, now WAAS application acceleration for the mission critical TN3270 sessions is examined. The basic OER configuration from the Branch/SOHO VPN Deployment section is used in this topology as well.

Chariot/IXIA is used to generate the VoIP G.729 Real Time Protocol (RTP) streams and also other TCP-based applications, including TN3270.

This section demonstrates a functional OER and WAAS branch implementation that can maintain branch availability and optimization in the event one of the two branch routers or access links fail or is taken out of service for maintenance.

Design Requirements and Considerations

General Topology

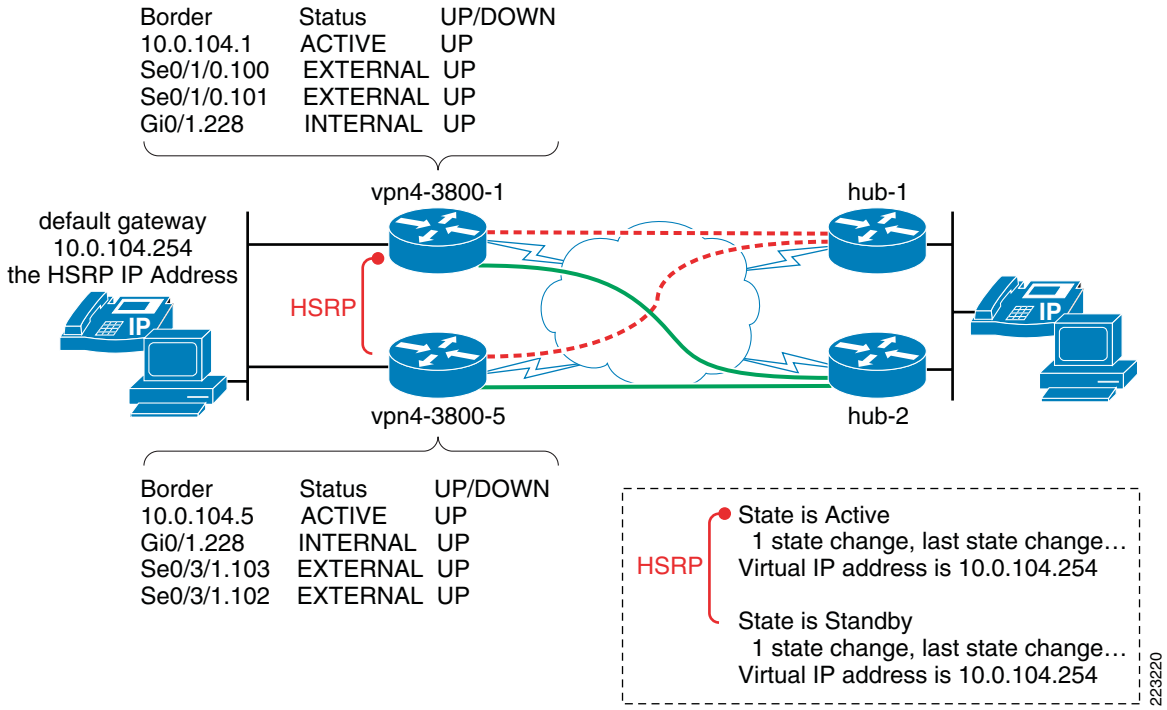
This branch topology provides additional network components from the topology deployed in the Branch/SOHO VPN Deployment section to demonstrate a branch OER deployment with multiple routers, each with more than one external exit.

In Chapter 9 of the *V3PN: Redundancy and Load Sharing Design Guide*, a similar network topology is shown without the capabilities of PfR. Chapter 9 describes a large branch (Frame Relay/Broadband Load Sharing and Backup) deployment based on an actual enterprise customer large retail-store network. One goal of this section is to include two WAN attached routers in the branch network, each with at least one physical circuit. The topology demonstrates two logical interfaces on each of the physical circuits. These logical interfaces could be Frame Relay PVCs, a Metro Ethernet deployment with two logical interfaces (dot1Q trunks) or two crypto tunnels. In any case, the logical interfaces from each branch router is terminated across two headend hub routers so that both branch routers can remain functional in the event a single hub fails or is taken out-of-service for maintenance.

Because a second goal of this section is to demonstrate stateless failover of the OER master controller function in the event a branch router is out of service, two logical interfaces on each branch router is also a requirement to meet the need of OER to have at least two external exits to remain functional.

A sample branch topology is shown in Figure 24.

Figure 24 Sample Branch Topology



In this example, each branch router is an OER border router. One router is the active master controller and the second branch router is the standby master controller. The router that is the HSRP active router is the primary master controller and the HSRP standby router is the standby master controller. The same master controller configuration is configured on each branch router.

The user workstations, IP phones, and other devices on the branch LAN use the HSRP IP address as their default gateway.

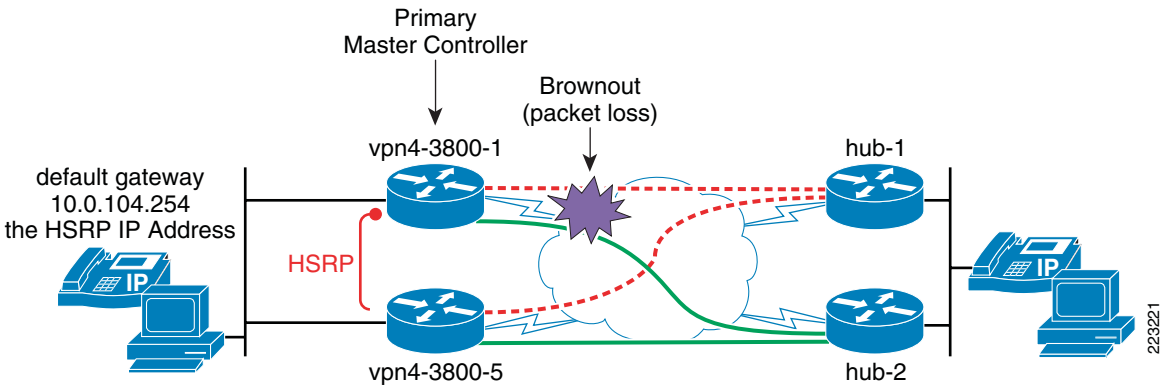
Failure Situation

In the topology shown in Figure 25, there is a possibility that the access-link to the router with the active HSRP state (the router controlling the virtual IP address) suffers a transient or intermittent failure condition leading to packet loss but not loss of carrier on the circuit. With HSRP, it is possible to decrement the HSRP priority based on a link transitioning to a down state. Using the **standby preempt** and **standby track** commands, control of the HSRP virtual address can be moved from one router to another.

If the line protocol of a tracked interface goes down, the HSRP priority is decremented. Another HSRP router with higher priority can become the active router if that router has **standby preempt** enabled. However, using an OER configuration on the branch routers, traffic can be routed out interfaces connected to the standby HSRP router, without configuring standby tracking of an interface and without encountering a hard link down situation.

Figure 25 illustrates a partial failure of the access circuit of the WAN link and the associated sub-interfaces (or logical interfaces) of the active HSRP router.

Figure 25 Access Circuit Failure Example



One example of a brownout condition could be high error rates on the circuit; CRC errors , for example, that causes packet loss of application traffic, but not a total link failure triggered by the loss of the Layer 2 keepalive packets. This condition is often manifested as a link that repeatedly transitions from up to down state, and then restarts. During periods where the link is up, application throughput is poor, with very slow response time and connection resets.

Parent Routes

OER must have parent routes; in this case, static routes configured on both branch routers to satisfy the parent route requirement. Because of this, both branch routers have two static routes for the default network, each static route identifies the interface corresponding to the OER external link.

```
vpn4-3800-1#show run | inc ip route
ip route 0.0.0.0 0.0.0.0 Serial0/1/0.101 name OER_Parent
ip route 0.0.0.0 0.0.0.0 Serial0/1/0.100 name OER_Parent


vpn4-3800-5#show run | inc ip route
ip route 0.0.0.0 0.0.0.0 Serial0/3/1.102 name OER_Parent
ip route 0.0.0.0 0.0.0.0 Serial0/3/1.103 name OER_Parent
```

Tip The next hop IP address associated with the interface could also be used. If the IP address of Serial0/1/0.101 is 10.0.65.5 with a /30 mask, 10.0.65.6 could be substituted for Serial0/1/0.101 in the previous example.

There is no requirement to implement a dynamic routing protocol on the serial interfaces or on the LAN interfaces. The campus headend routers would be configured with similar static parent routes, identifying the LAN subnet(s) at the branch. These parent routes would then be redistributed into a dynamic routing protocol. The branch has no redistribution requirement because devices on the branch LAN use the HSRP virtual address as their gateway address and no routers or Layer 3 switches are present in this branch topology.

Note The branch routers have no requirement to identify each other through static routes. OER policy-based routing PBR addresses the fact WAN exits exist on both routers.

Recovery

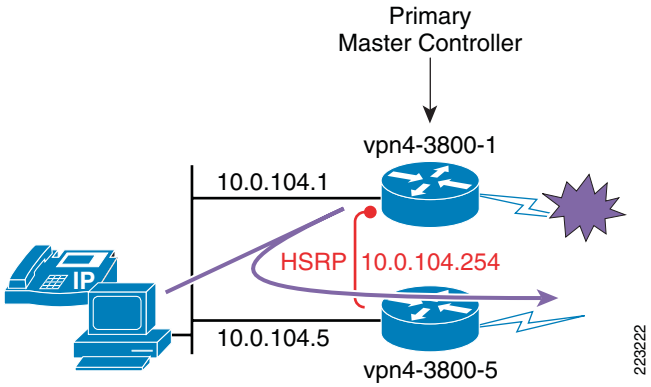
The tested configuration in this section demonstrates the use of OER application-aware routing based on policy-based routing where the exits are attached to two separate branch routers that are known to the devices on the LAN by the HSRP virtual address.

In this situation the following points are relevant:

- The brown out packet loss condition occurs on the HSRP active router,
- The OER configuration detects the packet loss, or other configured metric, like delay, throughput, etc.
- OER identifies and selects an exit on the HSRP standby router as the preferred exit for the application traffic
- The LAN attached devices continue to send all packets to the HSRP virtual address which is managed by the router experiencing the brown out condition.

Figure 26 illustrates these points.

Figure 26 Recovery



OER policy routes user traffic on both WAN routers to accommodate this path selection.

Policy Routing

Following the identification of the brownout condition on the HSRP active router, OER has selected an exit on the HSRP standby router as the preferred path. The following output shows that Serial interface 0/3/1.102 on border router 10.0.104.5 is the current exit.

```
vpn4-3800-1#show oer master appl
OER Prefix Statistics:
 Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
 P - Percentage below threshold, Jit - Jitter (ms),
 MOS - Mean Opinion Score
 Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
 E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
 U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
 # - Prefix monitor mode is Special, & - Blackholed Prefix
 % - Force Next-Hop, ^ - Prefix is denied

Prefix                Prot Port [src][dst]/ApplId        DSCP Source Prefix
                        State     Time Curr BR         CurrI/F         Protoco
l
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS
------------------------------------------------------------------------------
0.0.0.0/0             udp [1, 65535] [1, 65535]          cs5 0.0.0.0/0
                      INPOLICY        0 10.0.104.5       Se0/3/1.102     PBR

                            N        N        N        N        N        N
                           10       10        0        0        N        N
                            0        0
```

Because application traffic continues to be forwarded to the HSRP active router, OER policy routes the application on both WAN routers. Packets received by the HSRP active router (HSRP address 10.0.104.254) , are policy routed out of the same interface they were received on (GigabitEthernet0/1.228), to the second WAN router, the OER border router at 10.0.104.5.

```
vpn4-3800-1#show route-map dynamic app
Application - OER
  Number of active routemaps = 1

vpn4-3800-1#show route-map dynamic
route-map OER-11/05/07-15:35:45.231-7-OER, permit,
  sequence 0, identifier 174975
4340
  Match clauses:
    ip address (access-lists): oer#1
  Set clauses:
    ip next-hop 10.0.104.5
    interface GigabitEthernet0/1.228
  Policy routing matches: 0 packets, 0 bytes
 Current active dynamic routemaps = 1
```

Now looking at the OER border router 10.0.104.5, these application packets are policy routed out WAN interface Serial0/3/1.102:

```
vpn4-3800-5#show route-map dynamic
route-map OER-11/05/07-15:41:21.354-9-OER, permit,
 sequence 0, identifier 173730
```

```
0244
  Match clauses:
    ip address (access-lists): oer#1
  Set clauses:
    interface Serial0/3/1.102
  Policy routing matches: 0 packets, 0 bytes
 Current active dynamic routemaps = 1
```

The result of this configuration provides a higher degree of intelligence in path selection when OER is enabled than can be achieved if a dynamic routing protocol and HSRP **standby preempt** and **standby track** commands are used alone.

Design Limitations

The design limitations from Branch/SOHO VPN Deployment section are also applicable in this section and are not repeated in the interest of brevity.

During concept development of the topology example in the Branch/SOHO VPN Deployment section, it was determined that OER was not able to bring external interfaces into an UP (active) state following a link flap. To address this, *CSCsl20658 OER: BR IS INACTIVE State on MC after LINK-3-UPDOWN on Frame encap link* was opened to track this problem.

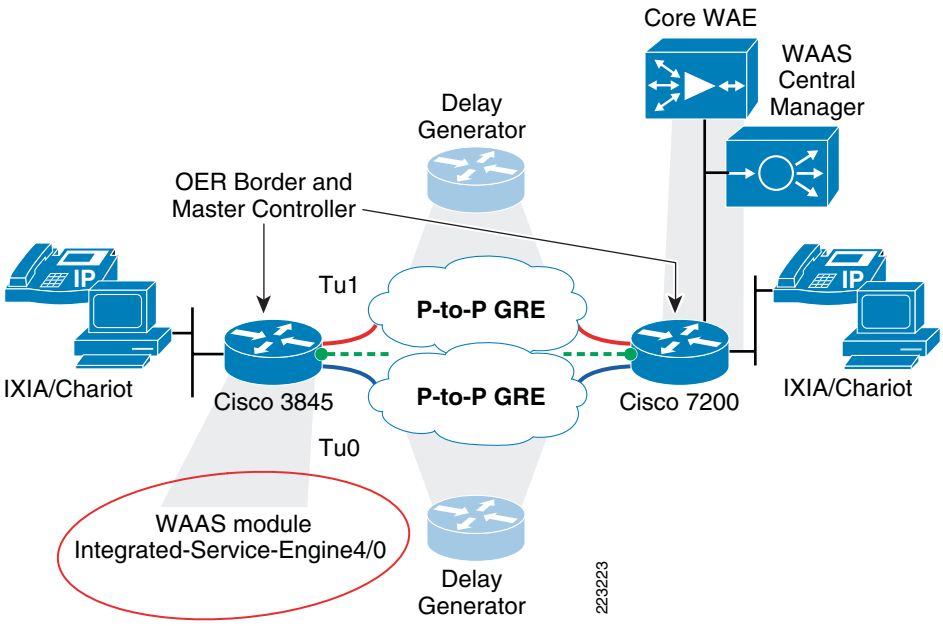
Topology with WAAS Network Module

The first step in migrating the branch topology to the ultimate goal of adding a second branch router and WAAS is to implement WAAS on an Integrated Service Engine (WAAS module) in slot 4/0 of the existing 3845 branch router. The NME-WAE is configured as inside so that it precedes the OER function. In this setup, WCCP is configured with IP forwarding for both the edge and core WAEs. GRE return is another alternative that allows the placement of the core WAE in the same subnet as the server farm. The OER configuration from the Branch/SOHO VPN Deployment section remains unchanged.

In this configuration, OER is configured at both the branch and the headend. If OER selects a different WAN interface on the return path for the branch client, it is possible that there would be asymmetric routing. However, this setup will preserve WAAS optimizations performed on the application traffic.

Figure 27 shows the topology with the addition of the WAAS module in the ISR 3845.

Figure 27 OER Configuration with WAAS



The same methodology is used from the testing in the Branch/SOHO VPN Deployment section, A delay generator is in the path of both tunnels. The delay on Tunnel 0 is random for each packet, ranging between 100 and 200 milliseconds with packets kept in sequence. Delay for Tunnel 1 is between 20 and 40 milliseconds, for each packet with the packet sequence maintained.

At 15 minutes elapsed time, the delay on both tunnels is changed. Tunnel 0 is changed to 30-50ms and Tunnel 1 is changed to 30-80ms, with packets kept in sequence.

Test Results

In the Branch/SOHO VPN Deployment section, VoIP is the application examined to illustrate the benefits of OER making path selections based on delay. Because WAAS is targeted at TCP optimization, this section focuses on how the WAAS implementation is beneficial to mission critical HTTP applications.

Figure 28 illustrates the benefit of both OER and OER with the additional of WAAS.

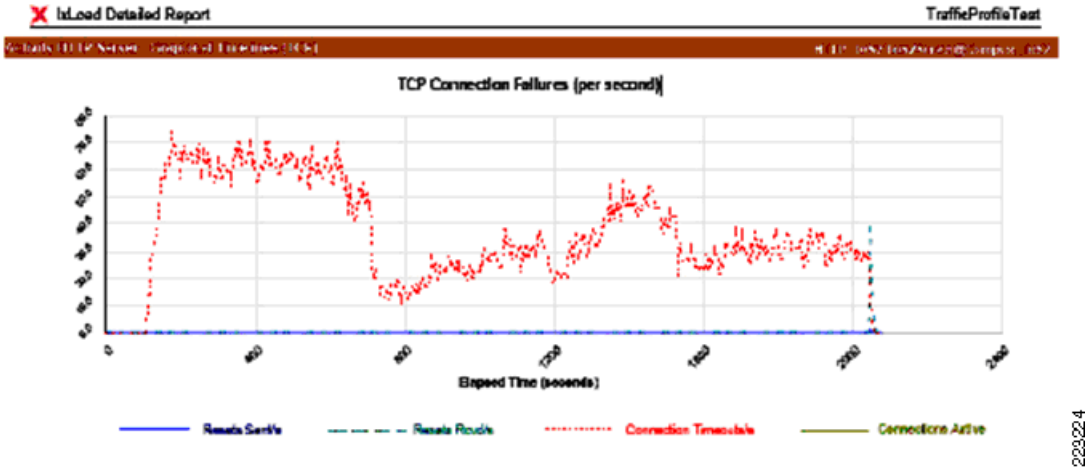
TCP Connection Failures

The first test is run without WAAS enabled. The TCP connection failures per second are charted across the duration of the test.

Results of OER Alone

A chart of TCP connection failure rates per second of HTTP mission critical traffic (DSCP AF21) is shown in Figure 28.

Figure 28 Results of OER



Note the following regarding the connection failure rate from the chart in Figure 28:

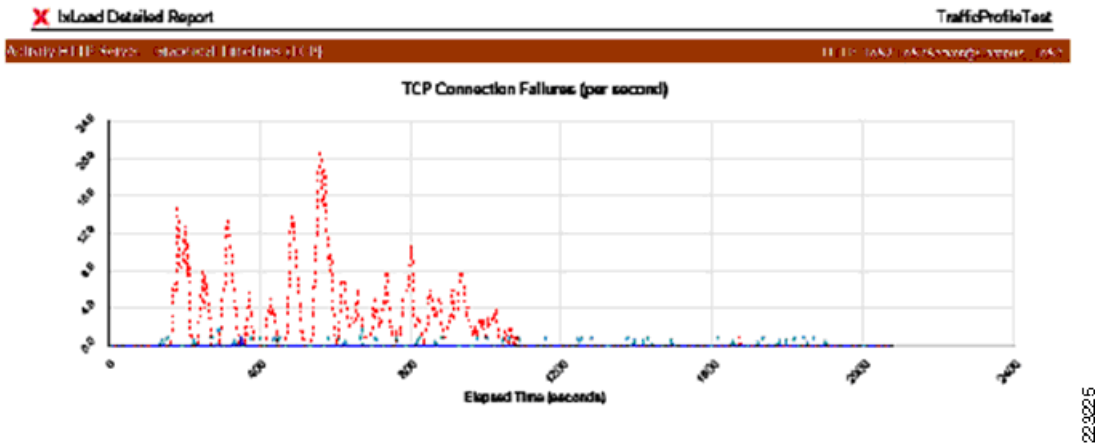
- The rate (per second) tracks the high delay present on tunnel 0 initially.
- The rate decreases dramatically around the 800 second mark as OER begins managing the destination network prefixes.
- The rate again increases as the delay changes.
- OER again decreases the rate as it reacts to the delay change.

The chart in Figure 28 shows that data applications exhibit improved performance as OER routes traffic through the external interface with the lowest delay characteristics. In the last section, it was demonstrated that VoIP latency and jitter are positively influenced as well.

Results of OER and WAAS

In the second test, the delay characteristics are the same, OER is initiated 7 minutes into the test as was done previously, but now WAAS is active throughout the entire test run. The chart in Figure 29 shows the TCP connection failure rates with WAAS enabled along with OER.

Figure 29 Results of OER and WAAS



Note the following from the chart in Figure 29:

- The failure rate with WAAS enabled is below 20 per second before OER can manage the network prefix to optimize the path based on delay.
- Without WAAS and OER, the corresponding failure rate was in the 60 to 70 connection failures per second.
- With OER but without WAAS, the failure rate shown in the previous charge was in steady state, observed in the 20 to 30 failures per second.
- With both OER and WAAS enabled, the failure rate decreases (approaching zero).

From the charts shown in Figure 28 and Figure 29, note that OER and WAAS are complementary technologies that improve both VoIP and data applications for the end-user at a branch office location.

Scalability Considerations

This section discusses issues the network manager should consider regarding branch scalability.

Policy-Based Routing

The use of dynamic route-maps is the means by which OER can implement Policy-Based Routing (PBR) to route application traffic out the appropriate exit. This feature was introduced in Cisco IOS version 12.4(2)T and is documented as a new feature enhancement titled *OER Application-Aware Routing: PBR feature,* at the following URL:

http://www.cisco.com/en/US/products/ps6441/products_feature_guide09186a00804d679f.html

One key point from the feature description relates to the use of PBR, that Cisco Express Forwarding (CEF) must be enabled on all participating devices.

Beginning in Cisco IOS Release 12.0, PBR is supported in CEF switching path. No special configuration is required to enable CEF-switched PBR. It is on by default.

In Cisco IOS Releases prior to 12.0, PBR is process switched by default and fast switching can optionally be enabled. Some network managers therefore may have experienced high CPU consumption associated with process switching PBR which may no longer an issue in later releases of Cisco IOS.

Router CPU Consumption

The WAAS testing in this document is targeted at demonstrating compatibility between WAAS and OER on a branch router deployment. This document is not intended to be a WAAS design guide nor to include exhaustive performance guidelines associated with a WAAS implementation. However, the network manager must be aware that the increased link efficiency and end-user application performance benefits of adding WAAS in the deployment incurs some additional CPU resources.

This table compares the branch router megabits per second (Mbps), packet per second (pps), and main CPU utilization with the test from the previous section. The increased application throughput is evident by a higher pps rate, and as a result there is a corresponding increase of main CPU utilization of the 3845 branch router in the tests. (See Table 4.)

Table 4 Router CPU Consumption

Test	Main CPU Busy %	WAN Interface Mbps IN	WAN Interface PPS IN	WAN Interface Mbps OUT	WAN Interface PPS OUT
Baseline	12%	9.2	1,782	3.8	1,578
With WAAS	30%	10	2.054	1.8	2,046

Most network managers would find the increase in CPU a very cost effective trade-off for the benefits associated with including the WAAS network module to support a higher user transaction rate. The network manager should take this into consideration to properly size the branch router.

Configuration Example—Single Branch Router with WAAS module

This section contains the relevant portions of the configuration file used in the test environment. This example is a reasonable illustration for a customer use case; however, it is not intended to be used verbatim on all possible network topologies.

Comments and clarifications regarding the configurations are included in-line with the configuration.

The Cisco IOS Release c3845-adventerprisek9-mz.124-11.T3.bin was used.

http://harry/solutions/vpn/results/ngwan30/oer/branch/1branch2exits/waas/peek-.s.he1-3800-2

```
!
hostname he1-3800-2
!
ip wccp 61
ip wccp 62
ip cef
ip tcp path-mtu-discovery
!
key chain NGWAN
 key 10
  key-string cisco
!
!
oer master
 policy-rules BRANCH
 ! shutdown
 no keepalive
 logging
 !
 border 10.192.0.1 key-chain NGWAN
  interface GigabitEthernet0/1 internal
  interface Tunnel1 external
  interface Tunnel0 external
 !
```

Tip The oer-map branch selects VoIP packets (UDP with DSCP of CS5), all remaining packets are identified by the learn mode.

```
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  aggregation-type prefix-length 29

 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
oer border
 logging
 local GigabitEthernet0/1
 master 10.192.0.1 key-chain NGWAN
!
!
```

Tip AES 256 is used for both IKE and IPSec.

```
crypto isakmp policy 10
 encr aes 256
 authentication pre-share
 group 2
crypto isakmp key bigsecret address 192.168.31.253
crypto isakmp key bigsecret address 192.168.31.252
crypto isakmp keepalive 10
!
!
crypto ipsec transform-set AES_SHA_TUNNEL esp-aes 256 esp-sha-hmac
no crypto ipsec nat-transparency udp-encaps
!
crypto map static-map 100 ipsec-isakmp
 set peer 192.168.31.252
 set transform-set AES_SHA_TUNNEL
 match address tun0
crypto map static-map 101 ipsec-isakmp
 set peer 192.168.31.253
 set transform-set AES_SHA_TUNNEL
 match address tun1
!
```

Tip No dynamic routing protocol is enabled on the GRE interfaces; however, GRE keepalives can be used in this configuration to transition the links to a down state in the event of a connection failure.

```
interface Tunnel0
 description Tunnel0
 bandwidth 10240
 ip address 10.56.1.0 255.255.252.0
 keepalive 10 3
 load-interval 30
 qos pre-classify
 tunnel source 192.168.0.2
 tunnel destination 192.168.31.252
 crypto map static-map
!
interface Tunnel1
 description Tunnel1
 bandwidth 10240
 ip address 10.56.5.0 255.255.252.0
 keepalive 10 3
 load-interval 30
 qos pre-classify
 tunnel source 192.168.0.146
 tunnel destination 192.168.31.253
 crypto map static-map
!
interface GigabitEthernet0/0.2200
 description Primary WAN
 encapsulation dot1Q 2200
 ip address 192.168.0.2 255.255.255.252
 service-policy output PER_CLASS_3mb
!
interface GigabitEthernet0/0.3300
 description Secondary WAN
 encapsulation dot1Q 3300
```

```
 ip address 192.168.0.146 255.255.255.252
 service-policy output PER_CLASS_3mb
!
interface GigabitEthernet0/1
 description GigabitEthernet0/1 Inside LAN interface
 ip address 10.192.0.132 255.255.255.192 secondary
 ip address 10.192.0.129 255.255.255.192 secondary
 ip address 10.192.0.1 255.255.255.128
 ip wccp 61 redirect in
 ip wccp 62 redirect out
 load-interval 30
 duplex full
 speed 100
 media-type rj45
 no keepalive
!
interface GigabitEthernet0/1.3302
 description <<** subinterface for NME-WAE **>>
 encapsulation dot1Q 3302
!
!
interface Integrated-Service-Engine4/0
 description <<** WAAS NME-WAE **>>
 ip address 192.168.0.153 255.255.255.252
 no ip redirects
 no ip unreachables
 ip wccp redirect exclude in
 service-module ip address 192.168.0.154 255.255.255.252
 service-module ip default-gateway 192.168.0.153
 no keepalive
!
!
ip route 0.0.0.0 0.0.0.0 Tunnel0
ip route 0.0.0.0 0.0.0.0 Tunnel1
ip route 10.0.0.0 255.0.0.0 Tunnel0
ip route 10.0.0.0 255.0.0.0 Tunnel1
ip route 10.204.0.4 255.255.255.255 Tunnel1
ip route 10.208.0.1 255.255.255.255 Tunnel0
ip route 192.168.0.0 255.255.0.0 192.168.0.1
ip route 192.168.0.0 255.255.0.0 192.168.0.145
ip route 192.168.0.154 255.255.255.255 Integrated-Service-Engine4/0
ip route 192.168.31.252 255.255.255.255 192.168.0.1
ip route 192.168.31.253 255.255.255.255 192.168.0.145
ip route 192.168.31.254 255.255.255.255 192.168.0.145
!
ip access-list extended VOICE
 permit udp any any dscp cs5
!
ip access-list extended tun0
 permit gre host 192.168.0.2 host 192.168.31.252
ip access-list extended tun1
 permit gre host 192.168.0.146 host 192.168.31.253
!
!
oer-map BRANCH 10
 match ip address access-list VOICE
 set delay threshold 250
 set mode route control
 set mode monitor active
 set resolve delay priority 1 variance 10
 set jitter threshold 11
 set mos threshold 4.06 percent 20
 set active-probe jitter 10.204.0.1 target-port 33333 codec g729a
 set probe frequency 30
```

```
!
end
```

Dual Branch Router with WAAS Appliance

In this section a second branch router is added, the OER master controller is implemented on both branch routers, which are OER border routers, to provide availability of the master controller should a single branch router fail. Because each branch router has two exit interfaces, even with a branch router failure, there remain two exists which OER can manage.

The WAE appliance was selected over the network module as the edge WAE or medium to large scale branches with dual routers. In this configuration, a Cisco WAE-512 appliance was used, although the sizing selection of the edge WAE hardware platform will vary based on a number of criteria such as disk size and memory needed for caching, simultaneous TCP connections established between WAE peers, and available WAN bandwidth. Refer to the *Cisco Wide Area Engine Data Sheet* at the following URL for a list of available hardware platforms:

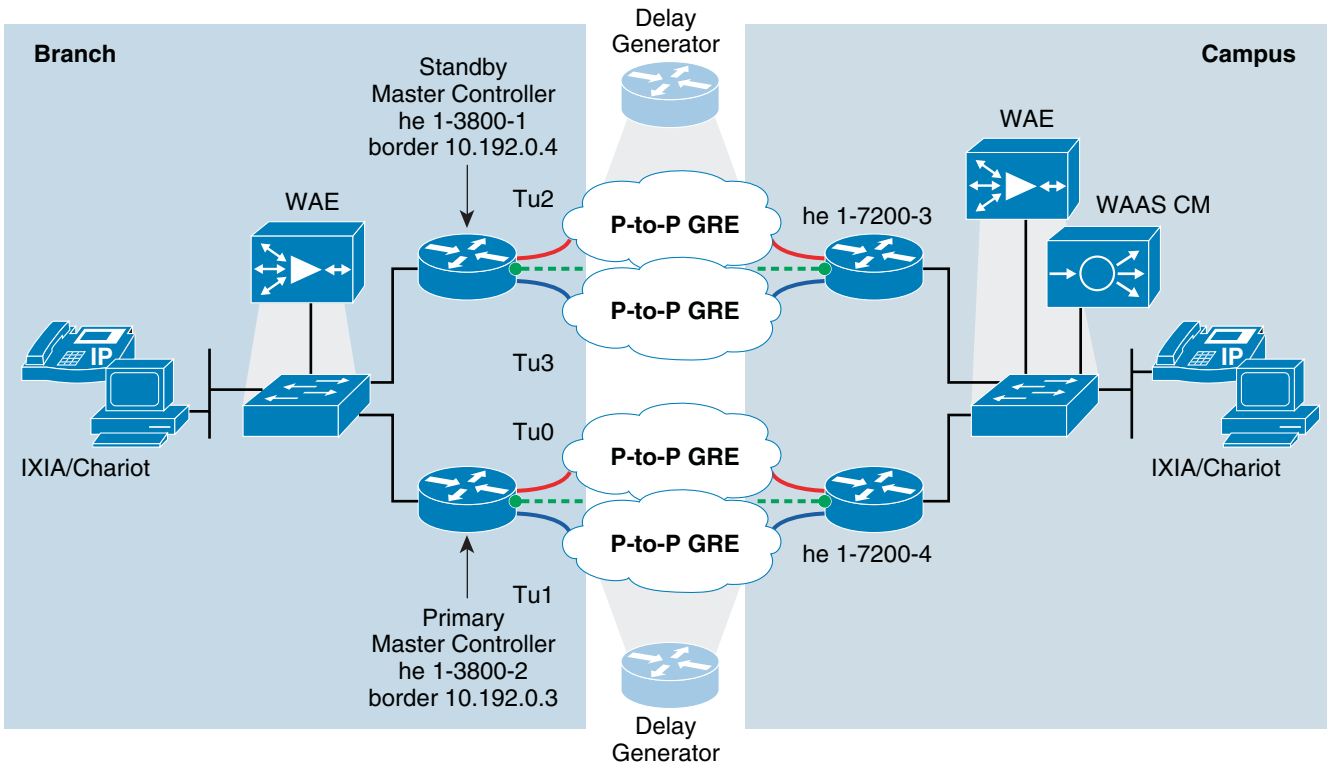
http://www.cisco.com/en/US/partner/prod/collateral/contnetw/ps5680/ps6474/product_data_sheet0900aecd80329e39_ps6870_Products_Data_Sheet.html

Topology Including WAAS Appliance

The branch topology now includes a second branch router and WAAS running on a WAE appliance. The OER policy and oer-map configuration from the previous section remains unchanged; however, both routers can function as master controller. There are two OER border routers, each with two exits.

Figure 30 shows the reference topology for this section.

Figure 30 Branch Topology with WAAS Appliance



In the topology shown in Figure 30, there is no redundant WAE appliance at either the branch or campus. The configurations of the campus OER border and master controller are not shown; but they are implemented in a similar manner to the branch deployment.

Test Results

The goal of this test is to demonstrate OER and WAAS compatibility. A secondary objective is to demonstrate an active/backup MC configuration. During the test, the primary MC router has failed and it is demonstrated the backup MC can resume OER functions over the remaining two exits.

Branch WAAS Compressions Ratios

To demonstrate that the WAAS appliance was functioning during the test, the output from the transport flow optimization (TFO) statistics is used to demonstrate that WAAS is optimizing the TCP traffic intercepted by the WAAS devices.

The user traffic generated in this test is from IXIA, rather than actual user clients. The branch compression ratios for both Web (HTTP) and File Transfers (FTP-DATA) are shown below:

```
ESE-EDGE-WAE2#sho stat tfo sav Web
Application           Inbound                Outbound
                      ---------------------- ----------------------
Web
    Bytes Savings                2539338782              86728150
    Packets Savings                 1804003               1766649
    Compression Ratio                49.7:1                 4.8:1



ESE-EDGE-WAE2#sho stat tfo sav File-Transfer

Application           Inbound                Outbound
                      ---------------------- ----------------------
File-Transfer
    Bytes Savings                   9538920             869827203
    Packets Savings                  305696                611506
    Compression Ratio                 3.2:1                 40.6:1
```

From this output, the tested topology demonstrated that TCP traffic is being optimized by the WAAS appliance.

These results are not claimed to be typical or expected results from a customer deployment; however, they are meant to demonstrate that the deployment of WAAS plus OER is a merging of complementary technologies to provide an enhanced user experience and better WAN utilization than can be achieved without these services enabled.

Tip The effective capacity reported by the edge WAE for FTP (31% of traffic) and WEB (68% of traffic) is 33.3X, including both applications in total. Effective capacity is an estimate of the WAN bandwidth that is required to transport unoptimized traffic in the same amount of time.

OER Master State Change

The second objective is to demonstrate a primary/standby MC configuration. Provided here is an example of a failure of the branch switch port connecting to the primary master controller router. The switch port is administratively shutdown to simulate the port failure. Following that failure the standby MC resumes OER functions over the remaining two exits.

Table 5 shows the command output of a **show oer master** command for the primary master controller and the standby master controller. The command is issued to each of the two routers every thirty seconds during the test execution. The command output is shown for both routers before and after the connectivity failure. The relevant states are highlighted in Table 5.

Table 5 OER Master State Change

Primary Master Controller - he1-3800-2	Standby Master Controller - he1-3800-1
Controlling both border routers, HSRP address Local	**HSRP address Standby, BR is UP with primary MC**

+===== Commands Iteration number: 52 =====+
-====== he1-3800-2: show oer master Thu Nov 15 10:32:02 2007
======-
show oer master
OER state: ENABLED and ACTIVE
  Conn Status: SUCCESS, PORT: 3949
  Version: 2.0
  Number of Border routers: 2
  Number of Exits: 4
  Number of monitored prefixes: 11 (max 5000)
  Max prefixes: total 5000 learn 2500
  Prefix count: total 11, learn 7, cfg 4
 	
Border Status UP/DOWN AuthFail
Version
10.192.0.4 ACTIVE UP 00:21:50 0 2.0
10.192.0.3 ACTIVE UP 00:21:50 0 2.0
	+===== Commands Iteration number: 38 =====+
	-====== he1-3800-1: show oer master Thu Nov 15 10:25:14 2007
	======-
	show oer master
	OER state: ENABLED and INACTIVE
	  Conn Status: SUCCESS, PORT: 3949
	  Version: 2.0
	  Number of Border routers: 2
	  Number of Exits: 4
	  Number of monitored prefixes: 4 (max 5000)
	  Max prefixes: total 5000 learn 2500
	  Prefix count: total 4, learn 0, cfg 4
	
	Border Status UP/DOWN AuthFail
	Version
	10.192.0.3 INACTIVE DOWN 0 0.0
	10.192.0.4 INACTIVE UP 17:43:43 0 2.0

Connection Failure, HSRP address is UNKNOWN	**HSRP address is Local, BR is UP with Standby (this) MC**

+===== Commands Iteration number: 53 =====+
-====== he1-3800-2: show oer master Thu Nov 15 10:32:33 2007
======-
show oer master
OER state: ENABLED and INACTIVE
  Conn Status: SUCCESS, PORT: 3949
  Version: 2.0
  Number of Border routers: 2
  Number of Exits: 4
  Number of monitored prefixes: 4 (max 5000)
  Max prefixes: total 5000 learn 2500
  Prefix count: total 4, learn 0, cfg 4
 	
Border Status UP/DOWN AuthFail
Version
10.192.0.4 INACTIVE DOWN 0 2.0
10.192.0.3 INACTIVE DOWN 0 2.0
	+===== Commands Iteration number: 39 =====+
	-====== he1-3800-1: show oer master Thu Nov 15 10:25:45 2007
	======-
	show oer master
	OER state: ENABLED and ACTIVE
	  Conn Status: SUCCESS, PORT: 3949
	  Version: 2.0
	  Number of Border routers: 2
	  Number of Exits: 4
	  Number of monitored prefixes: 4 (max 5000)
	  Max prefixes: total 5000 learn 2500
	  Prefix count: total 4, learn 0, cfg 4
	
	Border Status UP/DOWN AuthFail
	Version
	10.192.0.3 INACTIVE DOWN 0 0.0
	10.192.0.4 ACTIVE UP 00:00:14 0 2.0

The upper left quadrant is the primary master controller and one border router in normal operation, the master controller is enabled and active with both border routers active and up.

The upper right quadrant is the standby master controller and the second border router in normal operations. This master controller is enabled and inactive because no border routers can connect to it as it does not own the HSRP virtual address. This border router is connected to the primary master controller at this time. The state is shown as inactive and up.

The lower left quadrant is the primary master controller after switch port failure. Because the master controller is not administratively down, is is shown as enabled, but all other status is inactive as it does not own the HSRP virtual address. Also note it has flushed all learned prefixes and only configured prefixes are known.

The lower right quadrant is the standby master controller following the failure of the primary master controller/border router. At this time, the standby is the functioning master controller for the border router configured on this chassis. Two external exists are available so performance routing is functioning on the remaining links.

Syslog Output

The syslog output from both the primary and standby master controller is shown to illustrate the state changes before and after the switch port failure.

Syslog from Primary Master Controller (and Border Router)

In this failure situation, the switch port to the primary master controller is shutdown to simulate a LAN connectivity failure. The syslog file output is shown to illustrate the fact the primary master controller will uncontrolled the managed prefixes and become idle in regards to any OER functions.

Events to note include:

- HSRP state goes to Init and Standby.
- WCCP client service is lost.
- The master controller on this router is changed to the DOWN/Inactive state.
- Prefix learning is disabled and the applications are no longer controlled.

```
Nov 15 10:25:11.577 EST: %HSRP-5-STATECHANGE: FastEthernet1/1 Grp 2 state Active -> Speak
Nov 15 10:25:12.569 EST: %LINK-3-UPDOWN: Interface GigabitEthernet0/1, changed state to down
Nov 15 10:25:12.569 EST: %HSRP-5-STATECHANGE: GigabitEthernet0/1.1100 Grp 1 state Active -> Init
Nov 15 10:25:13.569 EST: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet0/1, changed
state to down
Nov 15 10:25:21.577 EST: %HSRP-5-STATECHANGE: FastEthernet1/1 Grp 2 state Speak -> Standby
Nov 15 10:25:24.185 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
Couldn't find the best exit
Nov 15 10:25:24.185 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
Couldn't choose exit in prefix timeout
Nov 15 10:25:36.089 EST: %WCCP-1-SERVICELOST: Service 61 lost on WCCP client 172.26.183.211
Nov 15 10:25:36.089 EST: %WCCP-1-SERVICELOST: Service 62 lost on WCCP client 172.26.183.211
Nov 15 10:25:51.837 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.56.4.0/29, No status in prefix timeout
Nov 15 10:25:59.517 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
Nov 15 10:26:28.701 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.56.12.0/29, No status in prefix timeout
Nov 15 10:26:59.429 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:27:29.125 EST: %OER_MC-5-NOTICE: Passive Unreachable OOP Appl Prefix 0.0.0.0/0 6 [1, 65535]
[1, 65535], No passive data on BR 10.192.0.3, intf Tu1
Nov 15 10:27:37.316 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 17 [1, 65535] [1, 65535],
Couldn't find the best exit
Nov 15 10:27:37.316 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 17 [1, 65535] [1, 65535],
Couldn't choose exit in prefix timeout
Nov 15 10:28:28.528 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.56.0.0/29, No status in prefix timeout
Nov 15 10:28:36.720 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.56.8.0/29, No status in prefix timeout
Nov 15 10:29:00.276 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:29:43.808 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.204.0.64/29, No status in prefix timeout
Nov 15 10:29:50.976 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.204.0.56/29, No status in prefix timeout
Nov 15 10:30:00.200 EST: %OER_MC-5-NOTICE: Uncontrol Prefix 10.204.0.48/29, No status in prefix timeout
Nov 15 10:31:01.132 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:31:19.052 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
Couldn't find the best exit
Nov 15 10:31:19.052 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
Couldn't choose exit in prefix timeout
Nov 15 10:32:19.999 EST: %OER_BR-5-NOTICE: MC 10.192.0.1 DOWN
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 DOWN
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Gi0/1 Unverified
```

```
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Tu2 Unverified
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Tu3 Unverified
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear exit, BR 10.192.0.4 i/f Tu2
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear exit, BR 10.192.0.4 i/f Tu3
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535], pbr
br-topology changed
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535], pbr
br-topology changed
Nov 15 10:32:27.443 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 17 [1, 65535] [1, 65535],
pbr br-topology changed
Nov 15 10:32:27.447 EST: %OER_MC-5-NOTICE: BR 10.192.0.3 DOWN
Nov 15 10:32:27.447 EST: %OER_MC-5-NOTICE: BR 10.192.0.3 IF Gi0/1.1100 Unverified
Nov 15 10:32:27.447 EST: %OER_MC-5-NOTICE: BR 10.192.0.3 IF Tu0 Unverified
Nov 15 10:32:27.447 EST: %OER_MC-5-NOTICE: BR 10.192.0.3 IF Tu1 Unverified
Nov 15 10:32:27.447 EST: %OER_MC-5-NOTICE: MC Inactive
Nov 15 10:32:27.451 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Exit down, BR 10.192.0.4 i/f Tu2
Nov 15 10:32:27.451 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Exit down, BR 10.192.0.4 i/f Tu3
Nov 15 10:32:27.451 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Exit down, BR 10.192.0.3 i/f Tu0
Nov 15 10:32:27.451 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Exit down, BR 10.192.0.3 i/f Tu1
Nov 15 10:32:27.451 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear Application all
Nov 15 10:32:27.451 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear prefix all
Nov 15 10:32:27.455 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear Application all
Nov 15 10:32:27.455 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear prefix all
Nov 15 10:32:27.455 EST: %OER_MC-5-NOTICE: Prefix Learning DISABLED
Nov 15 10:32:27.455 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
uncontrol non-optimized  prefix
Nov 15 10:32:27.455 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
uncontrol non-optimized  prefix
Nov 15 10:32:27.455 EST: %OER_MC-5-NOTICE: Uncontrol Appl Prefix 0.0.0.0/0 17 [1, 65535] [1, 65535],
uncontrol non-optimized  prefix
```

Syslog from Standby Master Controller (and Border Router)

The following is selected output from the syslog file on the standby master controller. Events to note include:

- HSRP state goes to Active.
- Master controller at 10.192.0.1 goes down, this is the session with the primary master controller.
- Master controller at 10.192.0.1 goes UP and MC is shown as Active, this standby MC receives a TCP session from the border router collocated on this router.
- Active absolute delay is shown out-of-policy for the VoIP and Data applications configured in the oer-map.
- The route is changed (policy routing is enabled) for these applications to select the best exit base on the criteria in the appropriate policy.

```
Nov 15 10:24:58.108 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
Nov 15 10:24:58.548 EST: %OER_BR-5-NOTICE: Prefix Learning STARTED
Nov 15 10:25:11.577 EST: %HSRP-5-STATECHANGE: FastEthernet1/1 Grp 2 state Standby -> Active
Nov 15 10:25:19.033 EST: %HSRP-5-STATECHANGE: GigabitEthernet0/1 Grp 1 state Standby -> Active
Nov 15 10:25:21.545 EST: %OER_BR-5-NOTICE: MC 10.192.0.1 DOWN
Nov 15 10:25:21.549 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
Nov 15 10:25:26.481 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 DOWN
Nov 15 10:25:26.481 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Gi0/1 Unverified
Nov 15 10:25:26.481 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Tu2 Unverified
Nov 15 10:25:26.481 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Tu3 Unverified
Nov 15 10:25:26.481 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear exit, BR 10.192.0.4 i/f Tu2
Nov 15 10:25:26.481 EST: %OER_MC-5-NOTICE: Uncontrol prefixes, Clear exit, BR 10.192.0.4 i/f Tu3
Nov 15 10:25:31.601 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 UP
Nov 15 10:25:31.605 EST: %OER_BR-5-NOTICE: MC 10.192.0.1 UP
Nov 15 10:25:31.905 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Gi0/1 UP
Nov 15 10:25:32.105 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Tu2 UP
Nov 15 10:25:32.105 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 Active
Nov 15 10:25:32.105 EST: %OER_MC-5-NOTICE: BR 10.192.0.4 IF Tu3 UP
Nov 15 10:25:32.105 EST: %OER_MC-5-NOTICE: MC Active
Nov 15 10:26:02.309 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:26:02.309 EST: %OER_BR-5-NOTICE: Prefix Learning STARTED
Nov 15 10:26:23.813 EST: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 0.0.0.0/0 6 [1, 65535] [1,
65535], delay 137, BR 10.192.0.4, i/f Tu2
```

```
Nov 15 10:26:23.813 EST: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 0.0.0.0/0 17 [1, 65535] [1,
65535], delay 140, BR 10.192.0.4, i/f Tu2
Nov 15 10:26:24.013 EST: %OER_MC-5-NOTICE: Route changed Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
BR 10.192.0.4, i/f Tu2, Reason None, OOP Reason Timer Expired
Nov 15 10:26:24.013 EST: %OER_MC-5-NOTICE: Route changed Appl Prefix 0.0.0.0/0 6 [1, 65535] [1, 65535],
BR 10.192.0.4, i/f Tu3, Reason None, OOP Reason Timer Expired
Nov 15 10:26:24.013 EST: %OER_MC-5-NOTICE: Route changed Appl Prefix 0.0.0.0/0 17 [1, 65535] [1,
65535], BR 10.192.0.4, i/f Tu2, Reason None, OOP Reason Timer Expired
Nov 15 10:27:03.110 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
Nov 15 10:27:03.310 EST: %OER_MC-5-NOTICE: Prefix Learning WRITING DATA
Nov 15 10:27:03.310 EST: %OER_MC-4-WARNING: Not monitoring prefixes on BR 10.192.0.3
Nov 15 10:27:03.750 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:27:03.750 EST: %OER_BR-5-NOTICE: Prefix Learning STARTED
Nov 15 10:27:11.942 EST: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 0.0.0.0/0 17 [1, 65535] [1,
65535], delay 136, BR 10.192.0.4, i/f Tu2
Nov 15 10:27:30.110 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.204.0.56/29, BR 10.192.0.4,
i/f Tu2
Nov 15 10:27:30.110 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.204.0.64/29, BR 10.192.0.4,
i/f Tu2
Nov 15 10:27:45.110 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.56.12.0/29, BR 10.192.0.4, i/f
Tu3
Nov 15 10:27:45.110 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.56.4.0/29, BR 10.192.0.4, i/f
Tu2
Nov 15 10:27:45.110 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.56.8.0/29, BR 10.192.0.4, i/f
Tu2
Nov 15 10:27:45.110 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.56.0.0/29, BR 10.192.0.4, i/f
Tu2
Nov 15 10:28:00.070 EST: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 0.0.0.0/0 17 [1, 65535] [1,
65535], delay 138, BR 10.192.0.4, i/f Tu2
Nov 15 10:28:04.111 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
Nov 15 10:28:04.311 EST: %OER_MC-5-NOTICE: Prefix Learning WRITING DATA
Nov 15 10:28:04.679 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:28:04.679 EST: %OER_BR-5-NOTICE: Prefix Learning STARTED
Nov 15 10:28:30.111 EST: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.204.0.48/29, BR 10.192.0.4,
i/f Tu2
Nov 15 10:28:48.199 EST: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 0.0.0.0/0 17 [1, 65535] [1,
65535], delay 138, BR 10.192.0.4, i/f Tu2
Nov 15 10:29:05.111 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
Nov 15 10:29:05.311 EST: %OER_MC-5-NOTICE: Prefix Learning WRITING DATA
Nov 15 10:29:05.611 EST: %OER_MC-5-NOTICE: Prefix Learning STARTED
Nov 15 10:29:05.611 EST: %OER_BR-5-NOTICE: Prefix Learning STARTED
Nov 15 10:29:36.332 EST: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 0.0.0.0/0 17 [1, 65535] [1,
65535], delay 138, BR 10.192.0.4, i/f Tu2
Nov 15 10:30:00.112 EST: %OER_MC-5-NOTICE: Route changed Prefix 10.56.8.0/29, BR 10.192.0.4, i/f Tu3,
Reason Non-OER, OOP Reason None
Nov 15 10:30:06.112 EST: %OER_BR-5-NOTICE: Prefix Learning STOPPED
```

Configuration Example—Dual Branch Routers with WAAS Appliance

This section provides sample configuration files from the branch routers and WAAS appliances in the test topology.

Primary Master Controller and Border Router

The system image file is flash:c3845-adventerprisek9-mz.124-11.T3.bin.

```
!
version 12.4
service timestamps debug datetime msec localtime show-timezone
service timestamps log datetime msec localtime show-timezone
no service password-encryption
service compress-config
!
hostname he1-3800-2
!
boot-start-marker
boot-end-marker
```

```
!
!
clock timezone EST -5
clock summer-time EDT recurring
clock calendar-valid
ip wccp 61
ip wccp 62
ip cef
ip tcp path-mtu-discovery
!
!
no ip dhcp use vrf connected
!
ip dhcp pool CPE-CLIENT
   network 10.18.1.0 255.255.255.0
   default-router 10.18.1.1
   dns-server 10.16.1.9
   domain-name tp.com
   option 150 ip 10.16.1.10
   netbios-name-server 10.16.1.9
!
!
key chain NGWAN
 key 10
  key-string cisco
!
!
!
oer master
 policy-rules BRANCH
 ! shutdown
 no keepalive
 logging
 !
 border 10.192.0.3 key-chain NGWAN
  interface GigabitEthernet0/1.1100 internal
  interface Tunnel0 external
  interface Tunnel1 external
 !
 border 10.192.0.4 key-chain NGWAN
  interface GigabitEthernet0/1 internal
  interface Tunnel2 external
  interface Tunnel3 external
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
  aggregation-type prefix-length 29
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
oer border
 logging
 local GigabitEthernet0/1.1100
 master 10.192.0.1 key-chain NGWAN
!
!
archive
 log config
```

```
  hidekeys
!
class-map match-any GOLD
 match ip dscp cs3
 match ip dscp cs6
class-map match-any TRANSACTIONAL_DATA
 match ip dscp af21
class-map match-any NETWORK_MANAGEMENT
 match ip dscp cs2
class-map match-any SILVER
 match ip dscp cs2
class-map match-any REAL_TIME
 match ip dscp cs5
 match ip dscp ef
 match ip dscp af41
class-map match-any STREAMING_VIDEO
 match ip dscp cs4
class-map match-any BEST_EFFORT
 match ip dscp default
 match ip dscp cs1
class-map match-any BULK_DATA
 match ip dscp af11
class-map match-any CALL-SETUP
 match ip dscp af31
 match ip dscp cs3
!
!
policy-map PER_CLASS_3mb
 class REAL_TIME
   police 1075000 conform-action transmit  exceed-action transmit  violate-action transmit
 class GOLD
  shape average 460800
 class SILVER
  shape average 768000
 class class-default
  shape average 1536000
policy-map INGRESS
 class REAL_TIME
  set ip dscp cs5
 class CALL-SETUP
  set ip dscp cs5
 class STREAMING_VIDEO
  set ip dscp cs2
 class TRANSACTIONAL_DATA
  set ip dscp cs3
 class NETWORK_MANAGEMENT
  set ip dscp cs3
 class BULK_DATA
  set ip dscp af21
!
crypto isakmp policy 10
 encr aes 256
 authentication pre-share
 group 2
crypto isakmp key bigsecret address 192.168.31.253
crypto isakmp key bigsecret address 192.168.31.252
crypto isakmp keepalive 10
!
crypto ipsec transform-set AES_SHA_TUNNEL esp-aes 256 esp-sha-hmac
no crypto ipsec nat-transparency udp-encaps
!
crypto map static-map 100 ipsec-isakmp
 set peer 192.168.31.252
 set security-association replay disable
```

```
 set transform-set AES_SHA_TUNNEL
 match address tun0
crypto map static-map 101 ipsec-isakmp
 set peer 192.168.31.253
 set security-association replay disable
 set transform-set AES_SHA_TUNNEL
 match address tun1
!
buffers small permanent 1500
buffers small max-free 2000
buffers small min-free 450
buffers middle permanent 1000
buffers middle max-free 1500
buffers middle min-free 300
buffers big permanent 1000
buffers big max-free 1500
buffers big min-free 300
!
interface Tunnel0
 description Tunnel0
 bandwidth 100000
 ip address 10.56.1.0 255.255.252.0
 load-interval 30
 qos pre-classify
 keepalive 10 3
 tunnel source 192.168.0.2
 tunnel destination 192.168.31.252
 crypto map static-map
!
interface Tunnel1
 description Tunnel1
 bandwidth 100000
 ip address 10.56.9.0 255.255.252.0
 load-interval 30
 qos pre-classify
 keepalive 10 3
 tunnel source 192.168.0.146
 tunnel destination 192.168.31.254
 crypto map static-map
!
!
interface GigabitEthernet0/0.2200
 description Primary WAN
 encapsulation dot1Q 2200
 ip address 192.168.0.2 255.255.255.252
 ip wccp redirect exclude in
 service-policy output PER_CLASS_3mb
!
interface GigabitEthernet0/0.3300
 description Secondary WAN
 encapsulation dot1Q 3300
 ip address 192.168.0.146 255.255.255.252
 ip wccp redirect exclude in
 service-policy output PER_CLASS_3mb
!
interface GigabitEthernet0/1
 description GigabitEthernet0/1
 no ip address
 load-interval 30
 duplex full
 speed 100
 media-type rj45
 service-policy input INGRESS
!
```

```
interface GigabitEthernet0/1.1100
 description ** LAN interface **
 encapsulation dot1Q 1100 native
 ip address 10.192.0.132 255.255.255.192 secondary
 ip address 10.192.0.3 255.255.255.128
 ip wccp 61 redirect in
 ip wccp 62 redirect out
 no keepalive
 standby 1 ip 10.192.0.1
 standby 1 ip 10.192.0.129 secondary
 standby 1 priority 110
 standby 1 preempt
!
!
interface FastEthernet1/1
 description ** WAE Interface **
 ip address 192.168.0.156 255.255.255.248
 ip wccp redirect exclude in
 duplex auto
 speed auto
 standby 2 ip 192.168.0.153
 standby 2 priority 110
 standby 2 preempt
 standby 2 track GigabitEthernet0/1.1100 20
!
!
ip route 0.0.0.0 0.0.0.0 Tunnel0
ip route 0.0.0.0 0.0.0.0 Tunnel1
ip route 10.0.0.0 255.0.0.0 Tunnel0
ip route 10.0.0.0 255.0.0.0 Tunnel1
ip route 10.208.0.1 255.255.255.255 Tunnel0
ip route 10.208.0.1 255.255.255.255 Tunnel1
!
ip route 192.168.31.252 255.255.255.255 192.168.0.1
ip route 192.168.31.254 255.255.255.255 192.168.0.145
!
!
ip access-list extended DATA
 permit tcp any any
 permit tcp any any dscp af21
ip access-list extended VOICE
 permit udp any any dscp cs5
ip access-list extended tun0
 permit gre host 192.168.0.2 host 192.168.31.252
ip access-list extended tun1
 permit gre host 192.168.0.146 host 192.168.31.254
!
ip sla responder
!
oer-map BRANCH 10
 match ip address access-list VOICE
 set delay threshold 100
 set mode route control
 set mode monitor active
 set resolve delay priority 1 variance 10
 set jitter threshold 11
 set mos threshold 4.06 percent 20
 set active-probe jitter 10.204.0.1 target-port 33333 codec g729a
 set probe frequency 30
!
oer-map BRANCH 20
 match ip address access-list DATA
 set delay threshold 100
 set mode route control
```

```
 set mode monitor both
 set resolve delay priority 2 variance 10
 set active-probe udp-echo 10.204.0.1 target-port 22222
 set probe frequency 30
!
end
```

Standby Master Controller and Border Router

The system image file is flash:c3825-adventerprisek9-mz.124-11.T3.

```
version 12.4
service timestamps debug datetime msec localtime show-timezone
service timestamps log datetime msec localtime show-timezone
no service password-encryption
service compress-config
!
hostname he1-3800-1
!
boot-start-marker
boot system flash flash:c3825-adventerprisek9-mz.124-11.T3
boot-end-marker
!
!
clock timezone EST -5
clock summer-time EDT recurring
clock calendar-valid
no network-clock-participate slot 1
ip wccp 61
ip wccp 62
ip cef
ip tcp path-mtu-discovery
!
!
key chain NGWAN
 key 10
  key-string cisco
!
oer master
 policy-rules BRANCH
 no keepalive
 logging
 !
 border 10.192.0.4 key-chain NGWAN
  interface GigabitEthernet0/1 internal
  interface Tunnel2 external
  interface Tunnel3 external
 !
 border 10.192.0.3 key-chain NGWAN
  interface Tunnel0 external
  interface Tunnel1 external
  interface GigabitEthernet0/1.1100 internal
 !
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  prefixes 2500
 expire after time 65000
```

```
  aggregation-type prefix-length 29
 mode route control
 mode monitor passive
 mode select-exit best
 periodic 180
!
oer border
 logging
 local GigabitEthernet0/1
 master 10.192.0.1 key-chain NGWAN
!
!
class-map match-any GOLD
 match ip dscp cs3
 match ip dscp cs6
class-map match-any TRANSACTIONAL_DATA
 match ip dscp af21
class-map match-any NETWORK_MANAGEMENT
 match ip dscp cs2
class-map match-any SILVER
 match ip dscp cs2
class-map match-any REAL_TIME
 match ip dscp cs5
 match ip dscp ef
 match ip dscp af41
class-map match-any STREAMING_VIDEO
 match ip dscp cs4
class-map match-any BEST_EFFORT
 match ip dscp default
 match ip dscp cs1
class-map match-any BULK_DATA
 match ip dscp af11
class-map match-any CALL-SETUP
 match ip dscp af31
 match ip dscp cs3
!
!
policy-map PER_CLASS_3mb
 class REAL_TIME
   police 1075000 conform-action transmit  exceed-action transmit  violate-action transmit
 class GOLD
  shape average 460800
 class SILVER
  shape average 768000
 class class-default
  shape average 1536000
policy-map INGRESS
 class REAL_TIME
  set ip dscp cs5
 class CALL-SETUP
  set ip dscp cs5
 class STREAMING_VIDEO
  set ip dscp cs2
 class TRANSACTIONAL_DATA
  set ip dscp cs3
 class NETWORK_MANAGEMENT
  set ip dscp cs3
 class BULK_DATA
  set ip dscp af21
!
!
!
crypto isakmp policy 10
 encr aes 256
```

```
 authentication pre-share
 group 2
crypto isakmp key bigsecret address 192.168.31.254
crypto isakmp key bigsecret address 192.168.31.255
crypto isakmp keepalive 10
!
!
crypto ipsec transform-set AES_SHA_TUNNEL esp-aes 256 esp-sha-hmac
no crypto ipsec nat-transparency udp-encaps
!
crypto map static-map 100 ipsec-isakmp
 set peer 192.168.31.254
 set security-association replay disable
 set transform-set AES_SHA_TUNNEL
 match address tun2
crypto map static-map 101 ipsec-isakmp
 set peer 192.168.31.255
 set security-association replay disable
 set transform-set AES_SHA_TUNNEL
 match address tun3
!
buffers small permanent 1500
buffers small max-free 2000
buffers small min-free 450
buffers middle permanent 1000
buffers middle max-free 1500
buffers middle min-free 300
buffers big permanent 1000
buffers big max-free 1500
buffers big min-free 300
!
!
interface Tunnel2
 description Tunnel2
 bandwidth 100000
 ip address 10.56.5.0 255.255.252.0
 load-interval 30
 qos pre-classify
 keepalive 10 3
 tunnel source 192.168.1.2
 tunnel destination 192.168.31.253
 crypto map static-map
!
interface Tunnel3
 description Tunnel3
 bandwidth 100000
 ip address 10.56.13.0 255.255.252.0
 load-interval 30
 qos pre-classify
 keepalive 10 3
 tunnel source 192.168.0.150
 tunnel destination 192.168.31.255
 crypto map static-map
!
interface GigabitEthernet0/0.2201
 description Primary WAN
 encapsulation dot1Q 2201
 ip address 192.168.1.2 255.255.255.252
 ip wccp redirect exclude in
 service-policy output PER_CLASS_3mb
!
interface GigabitEthernet0/0.3301
 description Secondary WAN
 encapsulation dot1Q 3301
```

```
 ip address 192.168.0.150 255.255.255.252
 ip wccp redirect exclude in
 service-policy output PER_CLASS_3mb
!
interface GigabitEthernet0/1
 description GigabitEthernet0/1
 ip address 10.192.0.131 255.255.255.192 secondary
 ip address 10.192.0.4 255.255.255.128
 ip wccp 61 redirect in
 ip wccp 62 redirect out
 load-interval 30
 duplex full
 speed 100
 media-type rj45
 no keepalive
 standby 1 ip 10.192.0.1
 standby 1 ip 10.192.0.129 secondary
 standby 1 preempt
 service-policy input INGRESS
!
interface FastEthernet1/1
 description EDGE WAE
 ip address 192.168.0.155 255.255.255.248
 ip wccp redirect exclude in
 load-interval 30
 duplex auto
 speed auto
 standby 2 ip 192.168.0.153
 standby 2 preempt
 standby 2 track GigabitEthernet0/1 20
!
ip route 0.0.0.0 0.0.0.0 Tunnel2
ip route 0.0.0.0 0.0.0.0 Tunnel3
ip route 10.0.0.0 255.0.0.0 Tunnel2
ip route 10.0.0.0 255.0.0.0 Tunnel3
ip route 10.208.0.1 255.255.255.255 Tunnel2
ip route 10.208.0.1 255.255.255.255 Tunnel3
!
ip route 192.168.31.253 255.255.255.255 192.168.1.1
ip route 192.168.31.255 255.255.255.255 192.168.0.149
!
!
!
ip access-list extended DATA
 permit tcp any any
 permit tcp any any dscp af21
ip access-list extended VOICE
 permit udp any any dscp cs5
ip access-list extended tun2
 permit gre host 192.168.1.2 host 192.168.31.253
ip access-list extended tun3
 permit gre host 192.168.0.150 host 192.168.31.255
!
ip sla responder

!
oer-map BRANCH 10
 match ip address access-list VOICE
 set delay threshold 100
 set mode route control
 set mode monitor active
 set resolve delay priority 1 variance 10
 set jitter threshold 11
 set mos threshold 4.06 percent 20
```

```
 set active-probe jitter 10.204.0.1 target-port 33333 codec g729a
 set probe frequency 30
!
oer-map BRANCH 20
 match ip address access-list DATA
 set delay threshold 100
 set mode route control
 set mode monitor both
 set resolve delay priority 2 variance 10
 set active-probe udp-echo 10.204.0.1 target-port 22222
 set probe frequency 30
!
control-plane
!

end
```

Branch WAAS Appliance

```
ESE-EDGE-WAE2#sho run
! WAAS version 4.0.13 (build b12 Aug  9 2007)
!
device mode application-accelerator
!
!
hostname ESE-EDGE-WAE2
!
!
clock timezone EST5EDT -5 0
!
!
ip domain-name ese.cisco.com
!
!
!
primary-interface GigabitEthernet 1/0
!
!
!
interface GigabitEthernet 1/0
 description ** WAE Primary Interface **
 ip address 192.168.0.154 255.255.255.248
 exit
interface GigabitEthernet 2/0
 description ** Management FLASH Address **
 ip address 172.26.183.211 255.255.252.0
 exit
!
interface InlineGroup 1/0
 no inline vlan all
 shutdown
 exit
interface InlineGroup 1/1
 no inline vlan all
 failover timeout 3
 shutdown
 exit
!
!
ip default-gateway 192.168.0.153
```

```
!
no auto-register enable
!
! ip path-mtu-discovery is disabled in WAAS by default
!
!
ip route 0.0.0.0 0.0.0.0 192.168.0.153
!
!
ntp server 10.204.0.6
!
!
wccp router-list 1 10.192.0.3 10.192.0.4
wccp tcp-promiscuous router-list-num 1
wccp version 2
!
!
!
username admin password 1 Mw2MFqqw3ZH02
username admin privilege 15
username admin print-admin-password 1 A00B9194BEDB81FEAAD3B435B51404EE 5C800F13A
3CE86ED2540DD4E7331E9A2
!
!
!
!
authentication login local enable primary
authentication configuration local enable primary
!
inetd enable ftp
!
!
!
!
central-manager address 10.204.0.6
cms enable
!
!
!
tfo tcp optimized-mss 1280
tfo tcp original-mss 1280
!
!
no adapter epm enable
!
!
policy-engine application
   name Authentication
   name Backup
   name CAD
   name Call-Management
   name Conferencing
   name Console
   name Content-Management
   name Directory-Services
   name Email-and-Messaging
   name Enterprise-Applications
   name File-System
   name File-Transfer
   name Instant-Messaging
   name Name-Services
   name P2P
   name Printing
   name Remote-Desktop
```

```
   name Replication
   name SQL
   name SSH
   name Storage
   name Streaming
...etc...
```

Campus WAAS Appliance

```
ESE-CORE-WAE#sho run
! WAAS version 4.0.13 (build b12 Aug  9 2007)
!
device mode application-accelerator
!
!
hostname ESE-CORE-WAE
!
!
primary-interface GigabitEthernet 1/0
!
interface GigabitEthernet 1/0
 description <<** CORE WAE PRIMARY INTERFACE **>
 ip address 10.208.0.2 255.252.0.0
 exit
interface GigabitEthernet 2/0
 description <<** ESE FLASH MANAGEMENT ADDRESS **>>
 ip address 172.26.183.208 255.255.252.0
 exit
!
!
!
ip default-gateway 10.208.0.1
!
no auto-register enable
!
! ip path-mtu-discovery is disabled in WAAS by default
!
!
ntp server 10.204.0.6
!
!
wccp router-list 1 10.204.0.3 10.204.0.4
wccp tcp-promiscuous router-list-num 1
wccp version 2
!
!
!
username admin password 1 Mw2MFqqw3ZH02
username admin privilege 15
username admin print-admin-password 1 A00B9194BEDB81FEAAD3B435B51404EE 5C800F13A
3CE86ED2540DD4E7331E9A2
!
!
authentication login local enable primary
authentication configuration local enable primary
!
!
central-manager address 10.204.0.6
cms enable
!
tfo tcp optimized-send-buffer 512
tfo tcp optimized-receive-buffer 512
```

```
!
!
no adapter epm enable
!
!
policy-engine application
   name Authentication
   name Backup
   name CAD
   name Call-Management
   name Conferencing
   name Console
   name Content-Management
   name Directory-Services
   name Email-and-Messaging
   name Enterprise-Applications
   name File-System
   name File-Transfer
   name Instant-Messaging
   name Name-Services

...etc...
```

Campus WAAS Central Manager

```
ESE-WAAS-CM#sho run
! WAAS version 4.0.13 (build b12 Aug  9 2007)
!
device mode central-manager
!
!
hostname ESE-WAAS-CM
!
primary-interface GigabitEthernet 1/0
!
!
interface GigabitEthernet 1/0
 description <<** WAAS CM PRIMARY INTERFACE **>>
 ip address 10.204.0.6 255.252.0.0
 exit
interface GigabitEthernet 2/0
 description <<** ESE FLASH MANAGEMENT ADDRESS **>>
 ip address 172.26.183.207 255.255.252.0
 exit
!
!
ip default-gateway 10.204.0.1
!
! ip path-mtu-discovery is disabled in WAAS by default
!
!
ip route 172.26.0.0 255.255.0.0 172.26.180.1
ip route 64.102.0.0 255.255.0.0 172.26.180.1
!
!
ntp server 172.26.176.1
!
!
username admin password 1 Mw2MFqqw3ZH02
username admin privilege 15
username admin print-admin-password 1 A00B9194BEDB81FEAAD3B435B51404EE
5C800F13A3CE86ED2540DD4E7331E9A2
```

```
!
!
authentication login local enable primary
authentication configuration local enable primary
!
inetd enable ftp
!
!
cms enable
!
!
banner motd message "Welcome to ESE WAAS Central Manager.\nPlease contact May Konfong or
Aeisha Bright for any issues or problems.\n"
banner enable
!
! End of WAAS configuration
```

Troubleshooting

This section includes show commands and other examples that may be useful as a reference for a customer deploying a similar configuration on their network.

Application Monitoring with oer-maps

This test configuration included a separate oer-map for VoIP and a second instance for data:

```
!
ip access-list extended VOICE
 permit udp any any dscp cs5
!
oer-map BRANCH 10
 match ip address access-list VOICE
 set delay threshold 100
 set mode route control
 set mode monitor active
 set resolve delay priority 1 variance 10
 set jitter threshold 11
 set mos threshold 4.06 percent 20
 set active-probe jitter 10.204.0.1 target-port 33333 codec g729a
 set probe frequency 30
!
```

The data oer-map instance selects TCP traffic with a default DSCP value and also with AF21. In testing, AF21 represents the transactional-data QoS class. This transactional-data QoS class is considered mission critical to the enterprise network manager.

```
ip access-list extended DATA
 permit tcp any any
 permit tcp any any dscp af21
!
oer-map BRANCH 20
 match ip address access-list DATA
 set delay threshold 100
 set mode route control
 set mode monitor both
```

```
 set resolve delay priority 2 variance 10
 set active-probe udp-echo 10.204.0.1 target-port 22222
 set probe frequency 30
!
```

Given that between the two oer-maps and referenced access-lists, there are three applications identified for monitoring:

- **permit udp any any dscp cs5**
- **permit tcp any any**
- **permit tcp any any dscp af21**

Now by issuing the **show oer master appl detail** command, these three items are shown separately. These are shown in the output as follows:

- Prefix: 0.0.0.0/0 Protocol: 17 Port: [1, 65535] [1, 65535] DSCP: 160
- Prefix: 0.0.0.0/0 Protocol: 6 Port: [1, 65535] [1, 65535] DSCP: 0
- Prefix: 0.0.0.0/0 Protocol: 6 Port: [1, 65535] [1, 65535] DSCP: 72

For reference, the order of the above bullet items equates to the order of the extended IP access list items shown in the previous bullet list.

The command output is shown below. It is displayed from the primary master controller in normal operation; both border routers functioning with all exists available.

```
-===== he1-3800-2: show oer master appl detail Thu Nov 15 10:06:08 2007 =====-
show oer master appl detail
Prefix: 0.0.0.0/0  Protocol: 6  Port: [1, 65535] [1, 65535] DSCP: 72
   State: DEFAULT*    Time Remaining: 0
   Policy: 20

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
   10.192.0.3      Tu1                     0        0        1        1
   10.192.0.4      Tu3                     0        0        0        0
   10.192.0.4      Tu2                     0        0       82       82
   10.192.0.3      Tu0                     0        0      361      361

   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min     Max     Dly
   udp-echo 10.204.0.1      22222     2     2       3       1       2       1
 ...[output suppressed]


--------------------------------------------------------------------------------
Prefix: 0.0.0.0/0  Protocol: 6  Port: [1, 65535] [1, 65535] DSCP: 0
   State: DEFAULT*    Time Remaining: 0
   Policy: 20

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
   10.192.0.3      Tu1                     0        0        1        1
   10.192.0.4      Tu3                     0        0        1        1
   10.192.0.4      Tu2                     0        0       82       82
   10.192.0.3      Tu0                     0        0      365      365
```

```
   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min      Max      Dly
   udp-echo 10.204.0.1      22222     2     2       2       1        1        1

...[output suppressed]

--------------------------------------------------------------------------------
Prefix: 0.0.0.0/0  Protocol: 17  Port: [1, 65535] [1, 65535] DSCP: 160
   State: DEFAULT*    Time Remaining: 0
   Policy: 10

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
   10.192.0.3      Tu1                     0        0        1        1
   10.192.0.4      Tu3                     0        0        0        0
   10.192.0.4      Tu2                     0        0       84       84
   10.192.0.3      Tu0                     0        0      356      356

   Most recent voice data per exit
   Border          Interface         ActSJit  ActPMOS
   10.192.0.3      Tu1                     0        0
   10.192.0.4      Tu3                     0        0
   10.192.0.4      Tu2                     6        0
   10.192.0.3      Tu0                    10      100

   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min      Max      Dly
   jitter   10.204.0.1      33333     2   200     267       1        2        1
   jitter   10.204.0.1      33333     2   200     394       1        2        1


   Latest Active Voice Stats on Current Exit:
   Type     Target          TPort     Codec Attem Comps  JitSum      MOS
   jitter   10.204.0.1      33333     g729a     2   200       8     4.06
   jitter   10.204.0.1      33333     g729a     2   200      64     4.06

...[output suppressed]
```

Summary

This section demonstrates how to implement a branch location with performance routing configured in an active/standby mode. The deployment allows for performance routing to function in the event of a failure of one of the two branch routers. The key component of the test is to demonstrate that OER and WAAS are companion technologies and can be configured to complement each other to provide a better user experience and better performance over existing WAN links. OER can provide intelligent path selection based on the round-trip delay characteristics of the links, while WAAS can increase the effective capacity of the links without purchasing more WAN bandwidth.

Troubleshooting

This section illustrates common misconceptions and debugging examples intended to assist the network manager with problem determination during deployment of OER in the network.

DMVPN and EIGRP Integration

The example in this section illustrates what can occur when OER is implemented on a mGRE (DMVPN) interface using static routes as the parent routes.

The router output is from a teleworker deployment with a Cisco 1811 router attached to a broadband WAN. With current OER capabilities, OER must have two static routes as parent routes in this topology. In this example, the underlying IPSec/EIGRP connection to one peer went down, causing a connectivity failure. The nature of the connection failure to the headend is not known, nor relevant in this case. From the perspective of the branch router, the EIGRP neighbor to Tun100 was a lost adjacency, but the Tunnel interface is still UP/UP because there is no GRE keepalive support on the mGRE (tunnel protect) interface. The parent routes for OER are static routes, not EIGRP learned routes.

This situation may potentially cause traffic to be blackholed even with mGRE point-to-point interfaces as there is no GRE keepalive to bring the tunnel interface in a UP/DOWN state.

During this state there is a connectivity failure for the IP phone and PC on the VLAN segment of the SOHO router in the example.

This is an example of the parent route:

```
S       64.102.0.0/16 is directly connected, Tunnel200
                      is directly connected, Tunnel100
```

While there are two tunnels, there is only one EIGRP neighbor active:

```
joeking-vpn-1811#show ip eigrp neighbors
IP-EIGRP neighbors for process 64
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
1   10.81.7.193             Tu200             13 2d00h      73   528  0  24657
```

In looking at the crypto socket, Tunnel 200 is open and Tunnel 100 is closed:

```
show crypto sockets
Number of Crypto Socket connections 2

 Tu200 Peers (local/remote): 192.168.2.12/64.102.223.25
       Local Ident  (addr/mask/port/prot): (192.168.2.12/255.255.255.255/0/47)
       Remote Ident (addr/mask/port/prot): (64.102.223.25/255.255.255.255/0/47)
       IPSec Profile: "DMVPN_IPSEC_PROFILE_2"
       Socket State: Open
       Client: "TUNNEL SEC" (Client State: Active)
 Tu100 Peers (local/remote): 192.168.2.12/64.102.223.24
       Local Ident  (addr/mask/port/prot): (192.168.2.12/255.255.255.255/0/47)
       Remote Ident (addr/mask/port/prot): (64.102.223.24/255.255.255.255/0/47)
       IPSec Profile: "DMVPN_IPSEC_PROFILE"
       Socket State: Closed
       Client: "TUNNEL SEC" (Client State: Active)
```

By manually clearing both the IKE and IPSec security associations (SA), the failed (Closed), crypto socket is restarted and connectivity is restored:

```
joeking-vpn-1811#clear cry isa
joeking-vpn-1811#clear cry sa
```

Following the clear commands, both EIGRP neighbors are active. Note that for Tunnel 200, the neighbor Uptime is over two days, while the Tunnel100, the failed tunnel, is now up for 32 seconds.

```
joeking-vpn-1811#show ip ei neighbors
IP-EIGRP neighbors for process 64
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
0   10.81.7.225             Tu100             13 00:00:32   81   528  0  5687
1   10.81.7.193             Tu200             13 2d00h      73   528  0  24657
```

At this point, full connectivity to the SOHO router has been restored.

Cisco IOS implementation of these enhancements is required to address successful implementation of OER in this topology. They are:

- CSCsi69186—PfR-DMVPN / mGRE integration
- CSCsk39768—PfR-EIGRP integration
- CSCsm34644—PfR-OSPF integration

Note that even though OER can detect a black hole situation for packets, not all network prefixes in this example are under OER control. If prefixes are not under active control or in the DEFAULT state, they encounter an outage for at least some period of time by using static routes with an mGRE interface.

Routing Changes Outside of OER Control

In the event OER reports a discovered exit for a network prefix and then also reports a route change for that network with a reason code of non-OER, this is an indication that equal cost routes are in the routing table with OER observing the same destination network prefix on two exits.

```
Apr 20 14:43:27.794 edt: %OER_MC-5-NOTICE: Discovered Exit for Prefix 64.102.0.0/16, BR
10.81.7.73, i/f Tu100
Apr 20 14:43:27.794 edt: %OER_MC-5-NOTICE: Route changed Prefix 64.102.0.0/16, BR
10.81.7.73, i/f Tu200, Reason Non-OER, OOP Reason None
```

These messages are not an indication of a problem, rather, simply an indication of learned traffic flows within the network.

OER Probes and External Interfaces

OER must have at least two external interfaces to function. When active probing is enabled, with or without an explicity configured probe, OER generates a probe sourced from all the external interfaces in order to determine the characteristics of the paths.

Note OER generates probes based on the configured external interfaces and the appropriate next hop IP addresses to force a probe out all the applicable external interfaces. OER does this independently of the global IP routing table.

The following is an example and the proof that OER sources probes and sends them out all external interfaces.

The following sample configuration is implemented in a topology with two border routers, each of which have two exit interfaces. That then means OER has four exits from which to generate probes.

```
!
oer-map PPM 10
 match traffic-class prefix-list PPM
 set delay threshold 110
 set mode route control
 set mode monitor fast
 set resolve jitter priority 1 variance 25
set resolve delay priority 2 variance 20
set jitter threshold 8
set active-probe jitter 192.168.49.1 target-port 20222 codec g729a
set probe frequency 2
!
```

In the above configuration, the probe target IP address is 192.168.49.1. This IP address is the configured loopback address on the remote router.

```
vpn4-3800-13#sh ip int brief | inc Loopback|Interface
Interface                  IP-Address      OK? Method Status
Protocol
Loopback0                  192.168.49.1    YES manual up up
```

This remote router has received probes sourced from the four external of the border routers. This can be demonstrated by displaying the status of the IP SLA responder on the remote router. There are four unique IP address as sources of the IP SLA probe:

```
vpn4-3800-13#show ip sla responder
IP SLAs Responder is: Enabled
Number of control message received: 153113 Number of errors: 0 Recent sources:
        192.168.193.26 [09:43:30.350 EDT Tue Aug 14 2007]
        192.168.129.6 [09:43:30.350 EDT Tue Aug 14 2007]
        192.168.193.2 [09:43:30.138 EDT Tue Aug 14 2007]
        192.168.129.34 [09:43:30.138 EDT Tue Aug 14 2007]
        192.168.193.26 [09:43:28.350 EDT Tue Aug 14 2007]
```

From the previous display, there are two items to note. First, there are four distinct IP addresses represented with a timestamp indicating they were received within the same second; actually a few hundred milliseconds apart. Second, the fifth item displayed, is approximately two seconds earlier than the first four entries.

Given the configured probe frequency is 2 seconds and *mode monitor fast* is configured, the behavior of that configuration is confirmed to be the four external links are probed continuously at the stated frequency. We should expect to see four probes, one from each exit, every two seconds.

Passive Monitoring Caveats

In the event passive monitoring is configured and no passive measurements (network prefixes derived from user traffic) are shown on the master controller, there are several valid reasons why this is normal behavior:

- No user traffic, passive monitoring relies on the NetFlow cache containing references to user traffic.
- No TCP traffic, passive delay, loss and reachability rely on TCP traffic. Passive delay and reachability are determined by TCP SYN and TCP SYN/ACK messages.

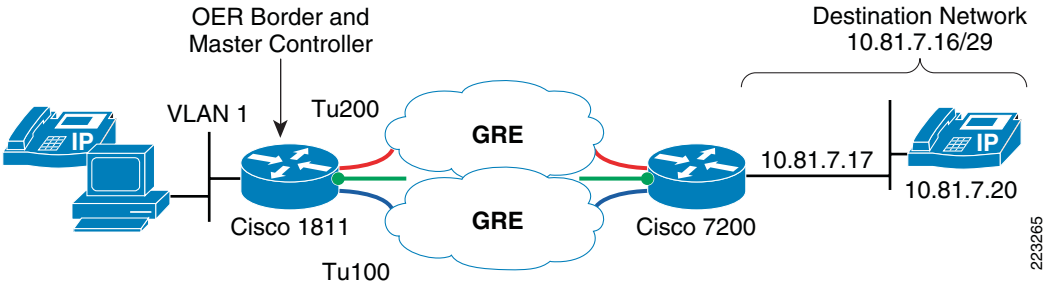
- Long lived TCP flows, while populate the NetFlow cache, it is the TCP session establishment that generates the TCP SYN and SYN/ACK. An FTP file transfer that runs for several hours, a very large file on a very low speed link is an example of a log lived TCP flow.

For passive monitoring, repeated TCP SYNCs without any TCP SYNC/ACK in response triggers a reachability failure. Loss can be determined when lower sequence numbers are observed than the highest number seen. This is an indication of a retransmission.

Passive Mode Example

This is an example of OER configured in passive mode only, rather than using mode monitor active or mode monitor both. There is a common misconception that OER can manage traffic out multiple exits in passive mode when traffic is observed on only one exit. While it is true that OER learns the network prefixes and designate them as being INPOLICY, HOLDDOWN, etc., while in passive mode only, traffic must be observed on an exit in order for OER to select that exit. Passive mode does not generate its own probes (as does active or both) and therefore must see end-user traffic through the exit to glean values for delay and reachability. Reachability to the target network must be verified in order for OER to use an exit. The topology in Figure 31 demonstrates this behavior.

Figure 31 Passive Mode Example Topology



The router configuration and example show commands are taken from the 1811 router shown in Figure 31. It is configured with two exits, represented by the tunnel interfaces, 100 and 200, and the internal network is VLAN1. A partial configuration follows:

```
oer master
 !
 border 10.81.7.73 key-chain GREEN
  interface Tunnel200 external
  interface Tunnel100 external
  interface Vlan1 internal
 !
!
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  aggregation-type prefix-length 30
 delay threshold 45
 mode route control
 mode monitor passive
 !
!
```

```
ip route 10.81.7.16 255.255.255.248 Tunnel100 name TEST_passive
!
ip route 10.0.0.0 255.0.0.0 Tunnel100 30 tag 300 name OER_parent
ip route 10.0.0.0 255.0.0.0 Tunnel200 30 tag 300 name OER_parent
!
end
```

The destination network for this example is 10.81.7.16/29 and there are two IP hosts at the address 10.81.7.17 and 10.81.7.20. Because the OER configuration includes **aggregation-type prefix-length 30** OER can aggregate traffic on the /30 boundary and may inject routes for 10.81.7.16/30 and 10.81.7.20/30. Given the destination network is a /29 address and aggregation can occur on a /30 boundary, the destination address is for all practical purposes essentially the address space of the /29 address can be divided in half, with .16 through .19 on the lower half and .20 through .23 in the upper half.

TCP traffic is generated by a workstation on VLAN1 (the internal network), initiating an HTTP session to the IP phone web server (port 80) on 10.81.7.20 and a Telnet session (port 23) to the router interface at 10.81.7.17.

For the purpose of this example, the configuration example includes a single static route to the destination network 10.81.7.16/29. This effectively forces all traffic to 10.81.7.16/29 through Tunnel 100. Both TCP session use Tunnel 100 as the exit because of this configuration. In a live network, this IP route should not be included in the configuration; it is used here for illustration purposes only. Note that the configuration does include two equal cost parent routes for 10.0.0.0/8. This satisfies the parent route requirement of OER.

OER is aware of the two exits, because the reference to the border router lists Tunnel 100 and Tunnel 200 as exits. However, because of the 10.81.7.16/29 static route, NetFlow data for this destination is only observed on Tunnel 100. Proof that OER is aware of this traffic is shown by the following syslog messages:

```
Dec 10 15:57:20.219 est: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.81.7.20/30, BR
10.81.7.73, i/f Tu100
Dec 10 16:01:59.643 est: %OER_MC-5-NOTICE: Route changed Prefix 10.81.7.16/30, BR
10.81.7.73, i/f Tu100, Reason Monitor, OOP Reason Timer Expired
```

OER has seen TCP traffic and lists 10.81.7.20/30 as INPOLICY (but unmanaged) while 10.81.7.16/30 is managed through Tunnel 100 and is in HOLDDOWN state. This can be seen from the following show commands:

```
joeking-vpn-1811#show oer mast pre 10.81.7.20/30 | beg Prefix
Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS  ActSLos  ActLLos
-------------------------------------------------------------------------------
10.81.7.20/30          INPOLICY*        0 10.81.7.73      Tu100           U
                              U        U        0        0        0        0
                              N        N        N        N        0        0
                              N        N


joeking-vpn-1811#show oer mast pre 10.81.7.16/30 | beg Prefix
Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS  ActSLos  ActLLos
-------------------------------------------------------------------------------
10.81.7.16/30            HOLDDOWN      199 10.81.7.73      Tu100           STATIC
                              U        U        0        0        0        0
```

```
                                    N        N        N        N        0        0
                                    N        N
```

Looking closely at the network 10.81.7.16/30 which is managed and in HOLDDOWN state, passive absolute delay is identified as out-of-policy, but because the exit Tunnel 200 is not used, no output traffic is observed and delay cannot be characterized for that exit. It is shown as zero in the following output display:

```
Dec 10 16:06:43.226 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.16/30,
delay 104, BR 10.81.7.73, i/f Tu100

joeking-vpn-1811#show oer mast pre 10.81.7.16/30 det
Prefix: 10.81.7.16/30
   State: HOLDDOWN    Time Remaining: 27
   Policy: Default

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
  *10.81.7.73      Tu100                 104      104        0        0
   10.81.7.73      Tu200                   0        0        0        0

   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min      Max      Dly

Prefix performance history records
 Current index 1, S_avg interval(min) 5, L_avg interval(min) 60

Age       Border          Interface       OOP/RteChg Reasons
Pas: DSum  Samples  DAvg  PktLoss  Unreach   Ebytes   Ibytes     Pkts    Flows
Act: Dsum Attempts  DAvg    Comps  Unreach   Jitter LoMOSCnt   MOSCnt
00:00:31  10.81.7.73      Tu100
      104        1   104        0        0      523      965       21        4
        0        0     0        0        0        N        N        N
```

In the following syslog messages, note that OER has observed that both networks are out-of-policy and they are transitioned to **uncontrolled** for observation. However, because no traffic is exiting via Tunnel 200, no passive data points can be collected.

```
Dec 10 16:07:20.227 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.16/30,
delay 104, BR 10.81.7.73, i/f Tu100
Dec 10 16:07:20.227 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.20/30,
delay 96, BR 10.81.7.73, i/f Tu100
Dec 10 16:07:20.235 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.20/30,
delay 96, BR 10.81.7.73, i/f Tu100
Dec 10 16:07:20.239 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.20/30,
delay 96, BR 10.81.7.73, i/f Tu100
Dec 10 16:07:20.239 est: %OER_MC-5-NOTICE: Uncontrol Prefix 10.81.7.20/30, OOP, mode
select-exit good
Dec 10 16:07:41.427 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.16/30,
delay 104, BR 10.81.7.73, i/f Tu100
Dec 10 16:07:41.427 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.16/30,
delay 104, BR 10.81.7.73, i/f Tu100
Dec 10 16:07:41.427 est: %OER_MC-5-NOTICE: Uncontrol Prefix 10.81.7.16/30, OOP, mode
select-exit good

Dec 10 16:08:20.228 est: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.81.7.20/30, BR
10.81.7.73, i/f Tu100
```

In this example, OER is not able to optimize traffic for 10.81.7.16/29 out both exits. To correct this configuration issue, to demonstrate proper management using both exits, a second equal cost parent route is added:

```
ip route 10.81.7.16 255.255.255.248 Tunnel200 name TEST_passive
```

Clear all prefixes currently under management following this change:

```
  clear oer mast pre *
```

Now Tunnel 200 can be used as an exit, and it is shown being selected from the following syslog messages:

```
Dec 10 16:14:00.236 est: %OER_MC-5-NOTICE: Discovered Exit for Prefix 10.81.7.20/30, BR
10.81.7.73, i/f Tu100

Dec 10 16:16:00.243 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.20/30,
delay 112, BR 10.81.7.73, i/f Tu100
Dec 10 16:16:00.275 est: %OER_MC-5-NOTICE: Passive ABS Delay OOP Prefix 10.81.7.20/30,
delay 112, BR 10.81.7.73, i/f Tu100
Dec 10 16:16:00.275 est: %OER_MC-5-NOTICE: Route changed Prefix 10.81.7.20/30, BR
10.81.7.73, i/f Tu200, Reason Monitor, OOP Reason Delay
```

Tunnel 200 is used for 10.81.7.20/30 as it is shown in HOLDDOWN state with the current interface as Tu200 with a static route injected by the OER subsystem:

```
joeking-vpn-1811#show oer mast pre 10.81.7.20/30
OER Prefix Statistics:
 Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
 P - Percentage below threshold, Jit - Jitter (ms),
 MOS - Mean Opinion Score
 Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
 E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
 U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
 # - Prefix monitor mode is Special, & - Blackholed Prefix
 % - Force Next-Hop, ^ - Prefix is denied

Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS  ActSLos  ActLLos
------------------------------------------------------------------------------
10.81.7.20/30           HOLDDOWN      282 10.81.7.73      Tu200           STATIC
                              U        U        0        0        0        0
                              N        N        N        N        1        1
                              N        N
```

Over a period of time, both Tunnel 100 and Tunnel 200 are shown as having non-zero values for both passive short and long-term delay as shown in the following example:

```
joeking-vpn-1811#show oer mast pre  10.81.7.16/30 det
Prefix: 10.81.7.16/30
   State: HOLDDOWN    Time Remaining: 270
   Policy: Default

   Most recent data per exit
   Border          Interface         PasSDly  PasLDly  ActSDly  ActLDly
  *10.81.7.73      Tu200                 100      100        0        0
   10.81.7.73      Tu100                  98       98        0        0

   Latest Active Stats on Current Exit:
   Type     Target          TPort Attem Comps    DSum     Min     Max     Dly
```

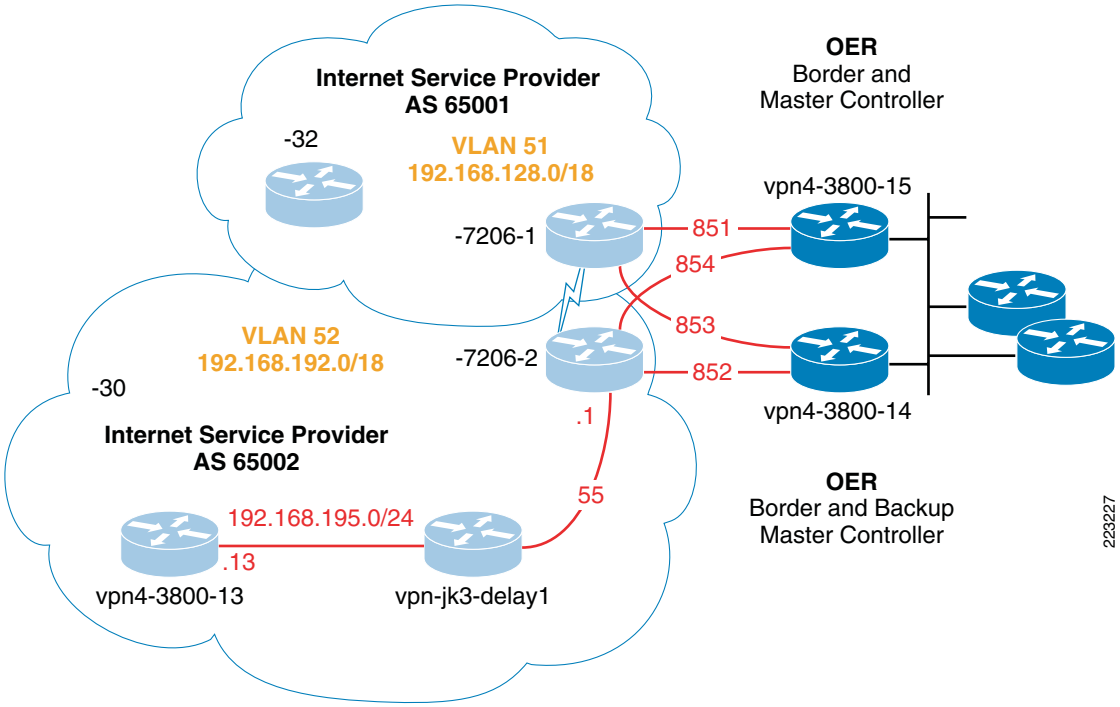
```
Prefix performance history records
 Current index 5, S_avg interval(min) 5, L_avg interval(min) 60

Age        Border          Interface       OOP/RteChg Reasons
Pas: DSum  Samples  DAvg  PktLoss  Unreach   Ebytes    Ibytes     Pkts    Flows
Act: Dsum Attempts  DAvg    Comps  Unreach   Jitter LoMOSCnt   MOSCnt
00:07:02  10.81.7.73      Tu100
       96        1    96        0        0     1306     1923       61        4
        0        0     0        0        0        N        N        N
00:09:42  10.81.7.73      Tu100            None/Discover
      100        1   100        0        0     1386     1963       64        4
        0        0     0        0        0        N        N        N
00:12:30  10.81.7.73      Tu200
       96        1    96        0        0     1426     1923       64        4
        0        0     0        0        0        N        N        N
00:14:30  10.81.7.73      Tu200            Delay/Monitor
      104        1   104        0        0     2487     2789      110        4
        0        0     0        0        0        N        N        N
00:16:07  10.81.7.73      Tu100
```

This demonstrates why equal-cost routes out two or more exits, when OER is configured for passive mode only, is important for proper functioning of performance routing.

Out-of-Policy (OOP) Example

This is an example of a network prefix that is out-of-policy for reason of absolute delay. The delay in the example is created by a test tool that is introducing a variable amount of delay in the range of 50 to 200ms for all packets. The topology is shown in Figure 32.

Figure 32 Out-of-Policy Example



The device named vpn4-3800-13 is configured to initiate a TCP connection to a host IP address in the campus core. The source IP address of this flow is 192.168.195.13. The device named vpn-jk3-delay1 is introducing to delay to this TCP session as well as the active probes (UDP ECHO) generated by the OER border routers.

Each border router (vpn4-3800-15 and vpn4-3800-14) are eBGP peers with both ISPs, AS 65001 and AS 65002; the numbers 651, 654, 652 and 653 are the VLANs IDs as well as the sub-interface instances.

The OER policy for this network prefix is shown below:

```
ip prefix-list PPM seq 10 permit 192.168.195.0/24


vpn4-3800-15#show oer mas pre 192.168.195.0/24 pol
* Overrides Default Policy Setting
oer-map PPM 10
  sequence no. 8444249301975040, provider id 1, provider priority 30
    host priority 0, policy priority 10, Session id 0
  match ip prefix-lists: PPM
  backoff 90 270 90
 *delay threshold 110           <-------- Key item for this example
  holddown 300
  periodic 180
  probe frequency 56
 *mode route control
 *mode monitor both
  mode select-exit best
  loss threshold 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  next-hop not set
  forwarding interface not set
  resolve loss priority 1 variance 10
  resolve delay priority 11 variance 20
  resolve utilization priority 12 variance 20
```

Looking at the current state of the prefix, it is INPOLICY*, meaning observed as in policy, but not being managed at the current time. The prefix has data for short and long-term delay by both passive and active measurements.

```
vpn4-3800-15#show oer mast pre

. . .


Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS  ActSLos  ActLLos
------------------------------------------------------------------------------
. . .
192.168.195.0/24       INPOLICY*      @34 192.168.131.98  Gi0/1.652       BGP
                             133      138        0        0        0     3636
                             176      126        0        0        1        1
                               N        N
vpn4-3800-15#
*Aug 10 11:20:10.072 edt: %OER_BR-5-NOTICE: Prefix Learning STOPPED
*Aug 10 11:20:10.548 edt: %OER_MC-5-NOTICE: Prefix Learning WRITING DATA
*Aug 10 11:20:10.552 edt: %OER_MC-5-NOTICE: Prefix Learning STARTED
*Aug 10 11:20:10.552 edt: %OER_BR-5-NOTICE: Prefix Learning STARTED
```

Note OER Master Prefix debugging is on for prefix 192.168.195.0/24.

```
vpn4-3800-15#
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Updated Target probe status for
Prefix 192.168.195.0/24 tgt 192.168.195.13 to TRUE
```

One border router generated two active probes and experienced an average delay of 116ms:

```
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Perf data point (bdr
192.168.131.97, if 9,nxtHop Default):  avg delay 116, loss 0, unreach 0, initiations 2,
completions 2, delay sum 232, ldelay max 160, ldelay min 72 jitter sum 0, mos low cnt 0
mos tot cnt 0
```

```
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Updated Target probe status for
Prefix 192.168.195.0/24 tgt 192.168.195.13 to TRUE
```

The second border router generated two active probes and reported average delay of 110ms:

```
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Perf data point (bdr
192.168.131.98, if 11, nxtHop Default):  avg delay 110, loss 0, unreach 0, initiations 2,
completions 2, delay sum 220, ldelay max 160, ldelay min 60 jitter sum 0, mos low cnt 0
mos tot cnt 0
```

The first border router now experiences average delay of 112ms:

```
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Updated Target probe status for
Prefix 192.168.195.0/24 tgt 192.168.195.13 to TRUE

*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Perf data point (bdr
192.168.131.97, if 10, nxtHop Default):  avg delay 112, loss 0, unreach 0, initiations 2,
completions 2, delay sum 224, ldelay max 160, ldelay min 64 jitter sum 0, mos low cnt 0
mos tot cnt 0
```

```
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: vpn4-3800-15#Updated Target probe
status for Prefix 192.168.195.0/24 tgt 192.168.195.13 to TRUE
```

The second border router again observed average delay of 114ms:

```
*Aug 10 11:20:30.807 edt: OER MC APC 192.168.195.0/24: Perf data point (bdr
192.168.131.98, if 12, nxtHop Default):  avg delay 114, loss 0, unreach 0, initiations 2,
completions 2, delay sum 228, ldelay max 156, ldelay min 72 jitter sum 0, mos low cnt 0
mos tot cnt 0
vpn4-3800-15#
```

```
*Aug 10 11:20:33.839 edt: OER MC PFX 192.168.195.0/24: Prefix timeout, state INPOLICY*
*Aug 10 11:20:33.839 edt: OER MC PFX 192.168.195.0/24: PDP choose exit, prefix state =
INPOLICY*, 0
*Aug 10 11:20:33.839 edt: OER MC PFX 192.168.195.0/24: Check ACT REL unreachable:
unreachable 0, policy 50%, notify FALSE
```

In this example, both the exists encountered delay that is higher than the absolute threshold, but because of the 20 variance configured for delay in the policy map, the exists are considered equal, so a new exit decision is make on a random basis.

```
*Aug 10 11:20:33.839 edt: OER MC PFX 192.168.195.0/24: Tie in exit selection,
192.168.131.97 Gi0/1.654 selected randomly
```

The OER subsystem injects an entry into the BGP table for the new exit point which was selected randomly.

```
*Aug 10 11:20:33.839 edt: OER MC PFX 192.168.195.0/24: Start FWD on new exit, br =
192.168.131.97, i/f = Gi0/1.654, nexthop 0.0.0.0, seq 2004, proto 4, exact TRUE


*Aug 10 11:20:33.839 edt: OER MC PFX 192.168.195.0/24: PDP start timer = 15 secs, prefix
state = CHOOSE
*Aug 10 11:20:34.039 edt: OER MC PFX 192.168.195.0/24: prefix_status 0 received, br =
192.168.131.97 i/f = Gi0/1.654
*Aug 10 11:20:34.039 edt: OER MC PFX 192.168.195.0/24: PDP start timer = 300 secs, prefix
state = HOLDDOWN
*Aug 10 11:20:34.039 edt: %OER_MC-5-NOTICE: Route changed Prefix 192.168.195.0/24, BR
192.168.131.97, i/f Gi0/1.654, Reason Delay, OOP Reason Delay

*Aug 10 11:20:34.039 edt: OER MC PFX 192.168.195.0/24: exclude_prefix status = 1
```

Now looking at the prefix, it is listed in HOLDDOWN because the IP route changed:

```
vpn4-3800-15#show oer mast pre
. . .

Prefix                  State     Time Curr BR         CurrI/F         Protocol
                      PasSDly  PasLDly   PasSUn   PasLUn  PasSLos  PasLLos
                      ActSDly  ActLDly   ActSUn   ActLUn      EBw      IBw
                      ActSJit  ActPMOS  ActSLos  ActLLos
------------------------------------------------------------------------------
. . .
192.168.195.0/24        HOLDDOWN      315 192.168.131.97  Gi0/1.654       BGP
                              U        U        0        0        0        0
                              U        U        0        0        1        1
                              N        N
```

In looking at the routing table for the primary router, we can see that a /24 advertisement has been entered from the BGP table:

```
vpn4-3800-15#show ip route bgp
B    192.168.195.0/24 [20/0] via 192.168.193.25, 00:05:51   <-------- route in RT from BGP
B    192.168.192.0/18 [20/0] via 192.168.193.25, 20:57:07
B    192.168.128.0/18 [20/0] via 192.168.129.5, 20:57:07
B    192.168.16.0/20 [20/0] via 192.168.129.5, 20:57:07
B    192.168.32.0/20 [20/0] via 192.168.129.5, 20:57:07
```

In this example, border router vpn4-3800-15 has the current exit for 192.168.195.0/24 is put into the BGP and routing table. In this example, the second BR (vpn4-3800-14) learns of the inclusion of 192.168.195.0/24 into the BGP table through iBGP from vpn4-3800-15, and due to the redistribution of BGP into EIGRP by vpn4-3800-15.

Therefore, from the perspective of vpn4-3800-14, the route for 192.168.195.0 is learned via iBGP and by EIGRP, because the EIGRP route is in the RT as an admin distance of 170, the iBGP route is not included in the routing table (RIB-failure indication) as an admin distance of 200 as the route already exists.

```
vpn4-3800-15#show ip bgp
BGP table version is 697, local router ID is 192.168.130.1
```

```
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
. . .
* i192.168.192.0/18 192.168.193.1            0    100      0 65002 i <--- Parent Route
*                   192.168.129.5                         0 65001 65002 i
*>                  192.168.193.25           0             0 65002 i
*> 192.168.195.0    192.168.193.25           0             0 65002 i



vpn4-3800-14#show ip bgp
BGP table version is 697, local router ID is 192.168.130.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
. . .
*> 192.168.192.0/18 192.168.193.1            0             0 65002 i
* i                 192.168.193.25           0    100      0 65002 i
*                   192.168.129.33                        0 65001 65002 i
r>i192.168.195.0    192.168.193.25           0   5000      0 65002 i
```

This example illustrates how OER measures and manages a prefix in an attempt to provide optimal routing to bring the prefix into the configured policy.

Appendix

References

- Load Balancing with CEF

  http://www.cisco.com/en/US/products/hw/modules/ps2033/prod_technical_reference09186a00800afeb7.html

- *How Does Unequal Cost Path Load Balancing (Variance) Work in IGRP and EIGRP?*

  http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a008009437d.shtml

Acknowledgements

The authors wish to thank the OER development team for their assistance and also to all our peers who have reviewed this document. Your help and insight is greatly appreciated.

Classless Inter-Domain Routing (CIDR) to Dotted Decimal Notation

Classless Inter-Domain Routing (CIDR) to Dotted Decimal Notation

CIDR	Dotted Decimal	Inverse Dotted Decimal
/1	128.0.0.0	127.255.255.255
/2	192.0.0.0	63.255.255.255
/3	224.0.0.0	31.255.255.255
/4	240.0.0.0	15.255.255.255
/5	248.0.0.0	7.255.255.255
/6	252.0.0.0	3.255.255.255
/7	254.0.0.0	1.255.255.255
/8	255.0.0.0	0.255.255.255
/9	255.128.0.0	0.127.255.255
/10	255.192.0.0	0.63.255.255
/11	255.224.0.0	0.31.255.255
/12	255.240.0.0	0.15.255.255
/13	255.248.0.0	0.7.255.255
/14	255.252.0.0	0.3.255.255
/15	255.254.0.0	0.1.255.255
/16	255.255.0.0	0.0.255.255
/17	255.255.128.0	0.0.127.255
/18	255.255.192.0	0.0.63.255
/19	255.255.224.0	0.0.31.255
/20	255.255.240.0	0.0.15.255
/21	255.255.248.0	0.0.7.255
/22	255.255.252.0	0.0.3.255
/23	255.255.254.0	0.0.1.255
/24	255.255.255.0	0.0.0.255
/25	255.255.255.128	0.0.0.127
/26	255.255.255.192	0.0.0.63
/27	255.255.255.224	0.0.0.31
/28	255.255.255.240	0.0.0.15
/29	255.255.255.248	0.0.0.7
/30	255.255.255.252	0.0.0.3
/31	255.255.255.254	0.0.0.1
/32	255.255.255.255	0.0.0.0

Reference Configuration for Load Balancing

Some customer deployments may desire only a load balancing function from an OER configuration. Much of this design guide focused on using OER to improve VoIP or data application performance. The following reference configuration has been used in customer networks to implement solely a load balancing function over the available exits.

```
oer master
 max-range-utilization percent 50
 resolve utilization priority 1 variance 5
 resolve range priority 2
 no resolve delay
!
border xx.xx.xx.xx
  interface Serial12/0 external
   max-xmit-utilization percent 90
  interface Serial13/0 external
   max-xmit-utilization percent 90
 border yy.yy.yy.yy
  interface Serial12/0 external
   max-xmit-utilization percent 90
  interface Serial13/0 external
   max-xmit-utilization percent 90
!
end
```

In the above configuration, **no periodic** does not appear in the command output as it is the Cisco IOS default. Periodic prefix evaluation is disabled as performance is not the objective, only load balancing.

Caveats

The following caveats and enhancements requests were identified during solution testing. They have been referenced and cited elsewere in this document and have been provide here collectively.

- CSCsd00633—OER Authentication fails with key string > 15 bytes
- CSCsi69186—PfR-DMVPN / mGRE integration
- CSCsk39768—PfR-EIGRP integration
- CSCsm34644—PfR-OSPF integration
- CSCsk48862—CPU HOG while learning if very large number of prefixes
- CSCsl10489—OER ignores MOS score of jitter probe when INPOLICY timer expires
- CSCsl20658—OER : BR IS INACTIVE State on MC after LINK-3-UPDOWN on Frame encap link