networktest

Pocket Rocket:

Scaling Up Metro Ethernet Access Networks With the Cisco ME 6524

prepared for Cisco Systems

September 2006

TABLE OF CONTENTS

ILLUSTRATIONS

1 Executive Summary

Mention "Cisco Catalyst 6500 Series" to many network architects and they're likely to think big – big switch, big footprint, big network. But the rapid rise of Metro Ethernet networks also is driving a need for small but powerful access and aggregation devices.

Cisco Systems addresses that need with a new product family, the Cisco ME 6500 Series Ethernet Switches. With a form factor of 1.5 rack units (3.5 inches), the first two members of the ME series are both far smaller than other Cisco Catalyst 6500 Series switches. The ME-6524-GS-8S offers 32 gigabit Ethernet fiber interfaces, while the ME-6524-GT-8S combines 24 10/100/1000 copper interfaces plus eight fiber uplink ports. Power consumption is another area where the ME 6524 switches are smaller: Both models draw less than 400 watts.

Despite its small size, the ME 6524 is built around the same Supervisor 720 engine that powers all Cisco Catalyst 6500 Series switches, and it supports the full range of functions needed to build highly scalable Metro Ethernet access networks. The ME 6524 offers flexibility by supporting a full range of carrier Ethernet services on a single platform – including pseudowires, layer-2 and layer-3 MPLS VPNs, IPv4 and IPv6 routing, and highly scalable IP multicast. In addition, the ME 6524 carries MEF 9 (Ethernet UNI) and MEF 14 (hard QOS) certification for all relevant service definitions.

The ME 6524 also supports Embedded Event Manager (EEM), a policy-based device management framework that can monitor and manage virtually any aspect of switch behavior. EEM goes beyond conventional network management systems by providing in-depth monitoring and control onboard each switch. EEM is a key capability for managing and troubleshooting large carrier Ethernet networks – and for ensuring maximum uptime and revenue.

Cisco Systems commissioned Network Test, an independent benchmarking and network design consultancy, to validate the performance, scalability, and management of the ME 6524. Among the test highlights:

- The Cisco ME 6524 routes IPv4 traffic at more than 15 million packets per second, and routes IPv6 traffic at more than 11.7 million pps
- For IP multicast service, the ME 6524 replicates and forwards traffic at 12.5 million pps in both layer-2 and layer-3 configurations involving 20,000 mroutes
- For layer-2 VPNs involving 3,072 type 4 VLAN-mode pseudowires, the Cisco ME 6524 forwards traffic at more than 11.1 million pps. With type 5 port-mode pseudowires, the switch forwards traffic at more than 11.4 million pps
- Using fast reroute reoptimization after a link failure, the ME 6524 converged a network carrying 100 pseudowires in 58 milliseconds

- For layer-3 VPNs, the Cisco ME 6524 was tested with 230,400 routes distributed into 192 virtual routing and forwarding instances, while simultaneously forwarding traffic at nearly 12.5 million pps

- In all relevant performance tests, the ME 6524 not only forwarded traffic at high rates to large numbers of destinations, but also concurrently provided QoS classification and remarking; security access control via a 10,000-line ACL; and Netflow monitoring of all packets on all interfaces

- EEM running simple user-defined scripts can monitor the ME 6524 and respond to denial-of-service attacks, protecting switch resources

- The Cisco ME 6524 can learn 90,000 MAC addresses at gigabit line rate. It also dynamically limits and disables address learning on a per-VLAN basis

- After a link failure, Cisco's Flexlink redirected traffic for 90,000 MAC addresses in 336 milliseconds, demonstrating redundancy and fast layer-2 convergence for dual-PE network designs

- The Cisco ME 6524 increases uptime by allowing hot-swapping of key system components such as the fan tray and redundant power supplies

This report is organized as follows. This section introduces the ME 6524. Section 2 covers performance test methodology and test results. Section 3 discusses EEM for device monitoring and configuration management. Section 4 offers a conclusion.

2 Performance Tests

This section describes performance and scalability tests in the areas of baseline IPv4 and IPv6 forwarding; layer-2 VPN and layer-3 VPN handling; IP multicast support; and hot-swappability of system components.

Network Test used a 300-second duration for all tests. Five minutes is 10 times longer than the 30-second duration recommended in the industry-standard switch testing methodology, RFC 2889. Longer test durations are more stressful on switch/routers, and they're also better predictors of how devices behave when handling long-lived flows such as applications involving voice and video feeds.

For all performance tests, Network Test concurrently verified the ability of the Cisco ME 6500 Series Ethernet Switch to perform three essential control-plane tasks:

Services enabled	Description
QoS classification and remarking	Classify traffic using 500-line QOS ACL and re-mark all traffic, including re-marking MPLS EXP bits where relevant
ACLs	Apply access controls to 10,000 L2/L3/L4 entries
Netflow	Enable Cisco Netflow statistics monitoring on all traffic

For all three sets of services in all performance tests, Network Test confirmed that the ME 6524 correctly classified, blocked, and reported on traffic.

The key metric used in these tests is "throughput." The term has a very specific meaning in the context of device benchmarking. As defined in RFC 1242, throughput is the highest offered load at which zero packets are dropped. Maintaining zero packet loss with heavy traffic loads is far more difficult for the device under test.

2.1 IPv4 and IPv6 Scalability and Throughput

A fundamental requirement of any switch/router is the ability to forward traffic at high rates to a large number of dynamically learned networks.

To verify Cisco's performance claims for the Cisco ME 6500 Series Ethernet Switch, Network Test conducted separate tests of IPv4 and IPv6 scalability and throughput. In both sets of tests, we configured the ME 6524 to use 24 access ports (facing CE devices) and 8 uplink ports (facing P or nPE devices).

2.1.1 IPv4 Unicast Scalability and Throughput

For the IPv4 tests, we configured a Spirent TestCenter traffic generator/analyzer to establish an OSPF adjacency on each of the eight uplink ports. TestCenter then offered 30,000 external link-state advertisements (LSAs) on each adjacency, for a total of 240,000 networks.

The massive size of the routing table in this test is noteworthy. The IGP databases at even the largest service providers rarely if ever exceed 50,000 entries, compared with the 240,000 routes used in this test. In fact, the routing table for this test is roughly 48,000 entries larger than all the networks in the global Internet combined (192,000 routes, as of late August 2006). OSPF is also a more complex protocol than the Border Gateway Protocol (BGP) used in the global Internet, potentially placing a heavier processing burden on the switch/router.

Once the routes were propagated, TestCenter then offered traffic to all ports in a bidirectional "backbone" pattern, where traffic from all ports on the CE side of the switch

was destined for all 240,000 networks on all uplink ports, and vice-versa[1]. We offered test traffic using multiple frame sizes, in each case for a duration of 300 seconds.

Aggregate IPv4 throughput in this test topped out above 15 million frames per second when we offered 64-byte frames. We also measured throughput for multiple frame lengths, ranging from 64 bytes through 9,216-byte jumbo frames.

Table 1 below presents results of the IPv4 routing scalability and throughput tests in tabular form.

IPv4 Unicast Routing Scalability and Throughput (240,000 OSPFv2 routes, 300-second test duration)	
Frame length (bytes)	Aggregate throughput* (frames/second)
64	15,009,592
128	9,933,748
256	5,954,112
512	3,250,110
1,024	1,700,212
1,280	1,371,135
1,518	1,143,183
2,034	865,532
4,068	427,715
9,216	190,559

*Tested with 10,000-line ACL; QoS classification and re-marking; and Netflow monitoring on all packets

Table 1: IPv4 Unicast Routing Scalability and Throughput

2.1.2 IPv6 Unicast Scalability and Throughput

Large production networks have already been running IPv6 for years in the Asia/Pacific region, and IPv6 deployment is growing in other areas. A combination of factors is driving IPv6 adoption, including an ever-smaller supply of IPv4 addresses; a growing number of customers doing business in the Asia/Pacific region; and a U.S. federal government mandate to adopt IPv6.

The ME 6524 supports a long list of features service providers will need in making the transition to IPv6. The switch supports the IPv6 versions of major routing protocols, including MP-BGP, OSPFv3, and RIPng, and also offers other key IPv6 features such as support for IPv4-to-IPv6 tunneling and ICMPv6. The ME 6524 is equally adept at handling IPv4, IPv6, or any combination of the two versions of IP traffic.

[1] As defined in the industry-standard switch testing methodology, RFC 2889, this pattern is officially known as a "partial mesh multiple device" topology.

To verify Cisco's IPv6 performance and scalability claims, Network Test used a test bed similar to that of the IPv4 tests, again involving eight uplink and 24 access ports.

For this test, Network Test used OSPFv3 to advertise routes to 118,000 IPv6 networks, or 14,750 per uplink interface. Note that this is a smaller number than the 240,000 routes advertised in the IPv4 tests. IPv6 addresses are four times larger than IPv4 addresses, and accordingly consume more memory in routing tables. Even so, an OSPFv3 database with 118,000 entries is enormous by current standards, certainly far larger than any OSPF network in production today.

The ME 6524 moved IPv6 traffic at more than 11.7 million fps when handling 78-byte frames. Network Test measured throughput for a range of frame lengths, going from 78 bytes (the minimum for IPv6 frames possible in the Spirent TestCenter instrument with signature fields enabled) through 9,216-byte jumbo frames.

Table 2 below presents results of the IPv6 tests in tabular form.

IPv6 Unicast Routing Scalability and Throughput (118,000 OSPFv3 routes, 300-second test duration)	
Frame length (bytes)	Aggregate throughput* (frames/second)
78	11,720,831
128	9,965,234
256	5,942,222
512	3,242,502
1,024	1,676,237
1,280	1,363,103
1,518	1,137,828
2,034	856,863
4,068	424,411
9,216	187,851

*Tested with 10,000-line ACL; QoS classification and re-marking; and Netflow monitoring on all packets

Table 2: IPv6 Unicast Routing Scalability and Throughput

2.2 Pseudowire Scalability, Throughput, and Resiliency

Delivering Ethernet over MPLS offers a number of advantages for service providers. For starters, Ethernet is more cost-effective than legacy Sonet/SDH due to lower component costs. At the same time, enterprise customers are already familiar with the technology, since it's "just Ethernet."

Aggregating multiple customers' networks into a single nPE router reduces the complexity of the meshed pseudowire network topology, significantly boosting scalability. The ME 6524 is well suited to uPE service.

Network Test assessed the scalability, throughput, and fault tolerance of the ME 6524 as a uPE device with three tests of Ethernet over MPLS (EoMPLS). In the first test, we set up more than 3,000 type 4 VLAN-mode pseudowires and measured throughput. Then we ran a similar test with type 5 port-mode pseudowires, again measuring throughput. Finally, we measured pseudowire convergence time—the interval needed to reroute traffic upon failure of a primary link.

To allow rapid provisioning of large numbers of pseudowires, recent versions of IOS for the Cisco Catalyst 6500 support a built-in interpreter for tool command language (TCL), the well-known scripting language. For this project, Network Test used the TCL interpreter to automate pseudowire creation, greatly simplifying switch configuration.

2.2.1 Type 4 VLAN-Mode Pseudowire Scalability and Throughput

In the context of Metro Ethernet access networks, network designers can assign an IEEE 802.1q VLAN tag to a customer network and then associate that tag with a unique pseudowire ID across the MPLS network. For PE devices supporting this method, the key questions are: How many pseudowires are supported, and what is the throughput when moving traffic over a large number of pseudowires?

To answer these questions, Network Test constructed a test bed in which the ME 6524 moves traffic across 3,072 pseudowires. Figure 2 below illustrates the type 4 VLAN-mode pseudowire test bed.

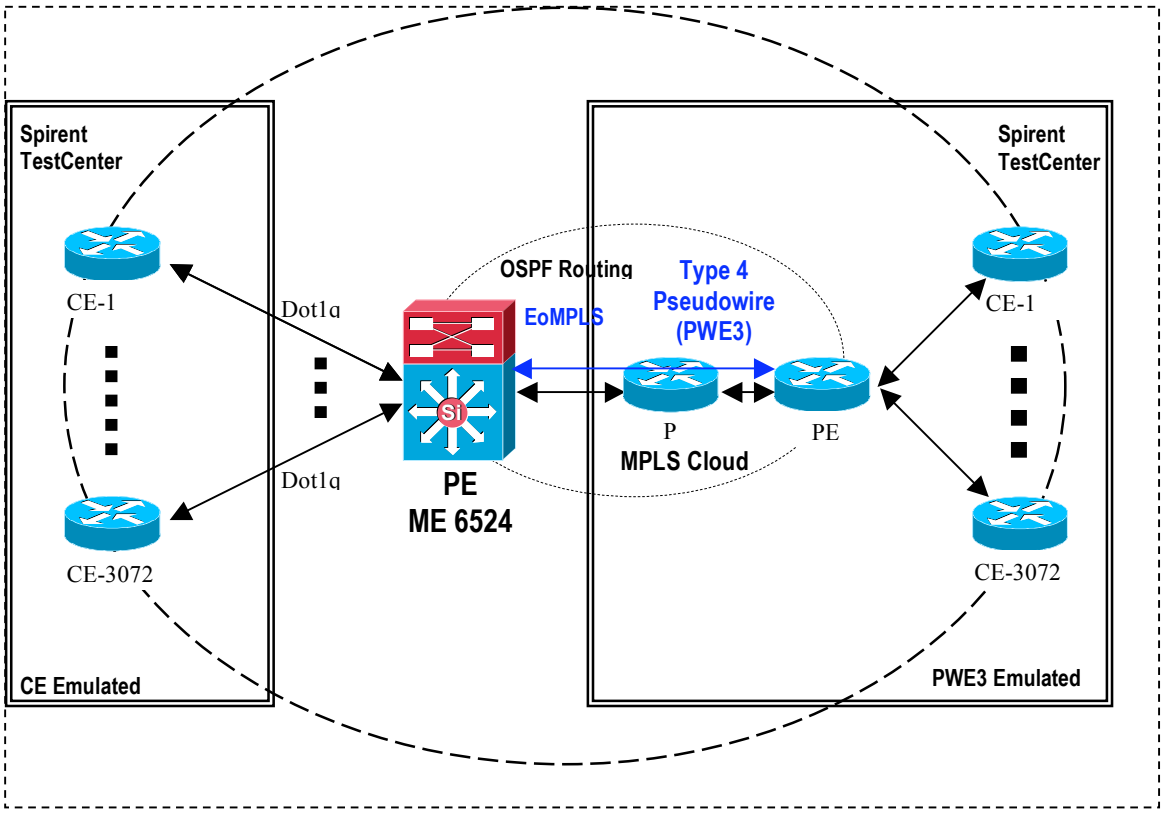

Figure 2: The Type 4 VLAN-Mode Pseudowire Test Bed

In this test, we define 128 subinterfaces, each with a unique VLAN ID, on each of 24 access ports, for a total of 3,072 unique customer networks. On the uplink side, we use LDP between the switch under test and the emulated PE devices to distribute MPLS labels. We also used OSPF to distribute routes from PE to emulated PE. Each of the 8 uplink interfaces supports 384 pseudowires, again for a total of 3,072.

As in most other tests for this project, we measured throughput for a variety of frame lengths; however, in this test, VLAN tagging affected the frame lengths of test traffic. For minimum-length tests, we configured TestCenter to offer 68-byte frames to each CE-facing interface (the 64-byte minimum length-Ethernet frame, plus 4 bytes for a VLAN tag). We offered 80-byte frames from the emulated P device to account for the additional Ethernet header and MPLS labels involved.

The ME 6524 achieved throughput of more than 11.1 million frames per second in this test. Table 3 below presents results from the type 4 VLAN-mode pseudowire scalability and throughput tests in tabular form.

Type 4 VLAN-Mode Pseudowire Scalability and Throughput (3,072 pseudowires, 300-second test duration)	
Frame length (bytes)	Aggregate throughput* (frames/second)
68	11,136,538
128	9,528,185
256	5,887,980
512	3,205,048
1,024	1,700,212
1,280	1,371,135
1,518	1,143,183
2,034	865,532
4,068	427,715
9,212	190,641

*Tested with 10,000-line ACL; QoS classification and re-marking; and Netflow monitoring on all packets

Table 3: Type 4 VLAN-Mode Pseudowire Scalability and Throughput

2.2.2 Type 5 Port-Mode Pseudowire Scalability and Throughput

While VLAN-to-pseudowire ID mapping is convenient, there are cases where network designers simply want to send raw Ethernet frames across an MPLS VPN, so that each customer essentially sees the metro Ethernet network as its own private broadcast domain. While type 5 port-mode (untagged) pseudowire designs are simpler than type 4 VLAN-mode designs, they raise the same questions about PE scalability and throughput.

Network Test assessed the ME 6524 for type 5 port-mode pseudowire service using a configuration similar to that for the type 4 VLAN-mode benchmarks. Figure 3 below illustrates the type 5 port-mode pseudowire test bed.

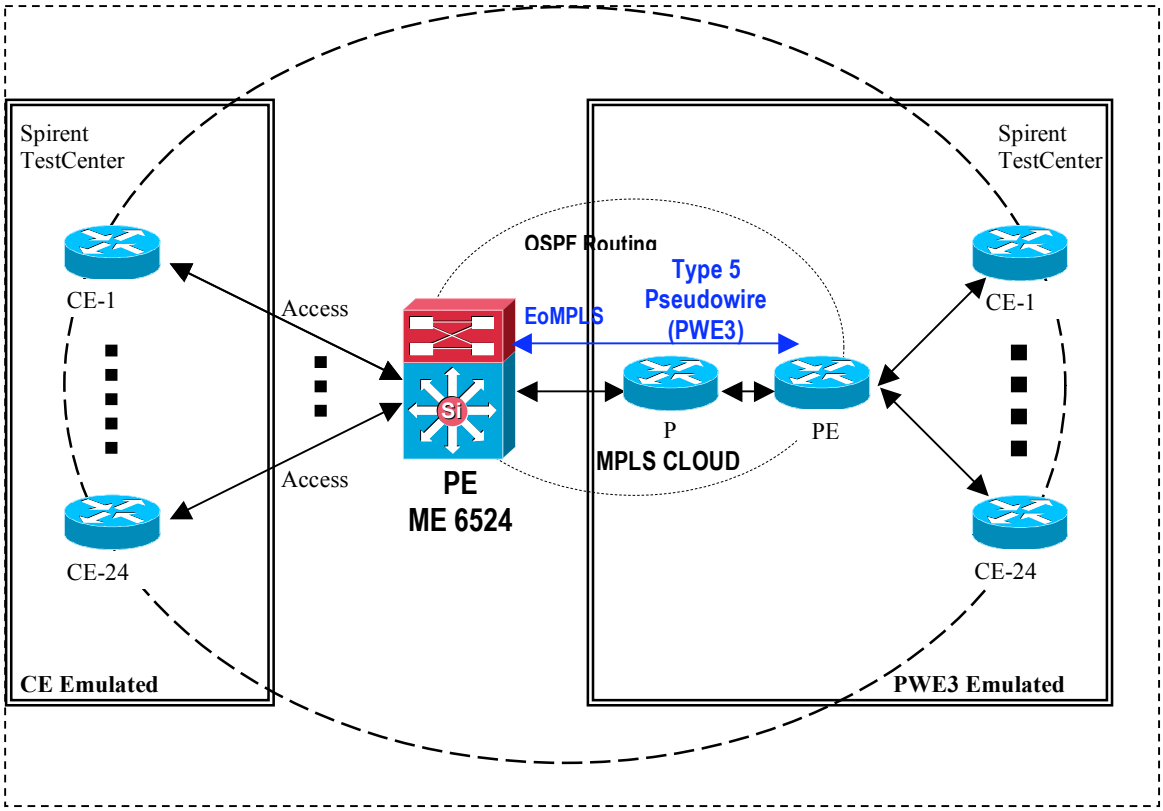

Figure 3: The Type 5 Port-Mode Pseudowire Test Bed

There are three major differences between this test and the previous one. First, the TestCenter traffic generator offers untagged frames to the CE-facing interfaces of the device under test. Second, because the 24 downlink ports are not divided into subinterfaces, only 24 pseudowires are used in this test. Third, on the uplink side, the device sets up type 5 port-mode pseudowires rather than type 4 VLAN-mode, indicating the Ethernet traffic is untagged.

The ME 6524 achieved throughput of more than 11.4 million frames per second in this test. Table 4 below presents results from the type 5 port-mode pseudowire scalability and throughput tests in tabular form.

Type 5 Port-Mode Pseudowire Scalability and Throughput (24 pseudowires, 300-second test duration)	
Frame length (bytes)	Aggregate throughput* (frames/second)
64	11,429,145
128	9,528,185
256	5,887,980
512	3,205,048
1,024	1,700,212
1,280	1,371,135
1,518	1,143,183
2,034	865,532
4,068	427,715
9,216	190,641

*Tested with 10,000-line ACL; QoS classification and re-marking; and Netflow monitoring on all packets

Table 4: Type 5 Port-Mode Pseudowire Scalability and Throughput

2.2.3 Pseudowire Failover and Convergence

High availability ranks above high performance and high scalability for service providers for an obvious reason: A down circuit is not a revenue-producing circuit. To help ensure maximum uptime, the ME 6524 supports fast reroute, an IETF technology for rapid redirection around failed links and nodes in MPLS and IP networks[2]. With fast reroute, large numbers of pseudowires can quickly be moved onto alternative paths with minimal disruption for customer traffic.

To determine pseudowire failover time, Network Test forced a failure on a protected link carrying 100 pseudowires, and measured convergence time needed for the pseudowires to be set up on a second circuit.

Figure 4 below illustrates the pseudowire failover test bed. Using RSVP-TE, we set up 100 pseudowires between R1 (the ME 6524 under test) and R2 across a primary label-switched path (LSP), labeled tunnel 1. A second LSP through R3 (tunnel 2) provided a secondary path across the test bed.

[2] RFC 4090 describes extensions to RSVP-TE for fast reroute in MPLS networks. The IETF's routing area working group (rtgwg) is defining similar fast reroute mechanisms for IP networks.

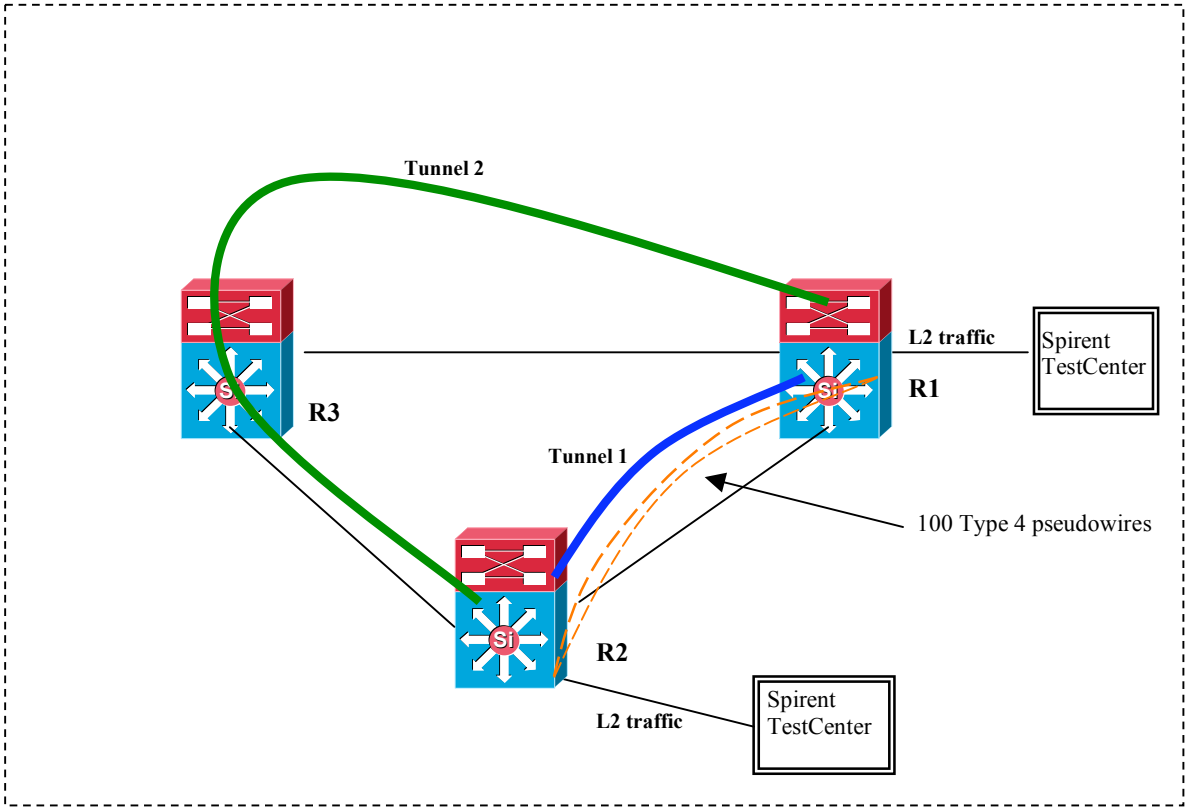

Figure 4: The Pseudowire Failover Test Bed

We offered test traffic across all pseudowires and then physically disconnected the link between R1 and R2. Since we offered traffic at a rate of 100,000 frames per second, each lost frame represented 10 microseconds of convergence time.

In our tests, the ME 6524 using fast reroute successfully reconverged all pseudowires within 58 milliseconds. Table 5 below summarizes results from this test.

Pseudowire Failover and Convergence Time	
Configured pseudowires	100
Convergence time from tunnel 1 to tunnel 2 (milliseconds)	58.01

Table 5: Pseudowire Failover and Convergence Time

2.3 MPLS VPN Scalability and Throughput

Layer-3 virtual private networks (VPNs) rank among the most widely deployed uses of MPLS technology. Layer-3 VPNs allow service providers to give each customer a unique IP address space and routing table, even though multiple customers' networks attach to the same router. While this eliminates the need to provision one router per customer, it also places scalability burdens on the provider edge (PE) device.

To assess the suitability of the ME 6524 for MPLS VPN service, Network Test configured the switch to support 192 virtual routing and forwarding (VRF) instances and 230,400 routes. This means 192 customers could each use the same addressing (for example, the same net-10 private address space), while the switch handled routing tables that collectively were larger than the global Internet's full routing table (192,000 routes as of late August 2006, compared with 230,400 routes used in this test).

Figure 5 below illustrates the MPLS VPN test bed. We used 24 downlink (CE-facing) and eight uplink (P-facing) interfaces, while Spirent TestCenter emulated CE, P, and PE devices. There were 8 subinterfaces on each access port, making a total of 192 virtual routing and forwarding (VRF) instances on the router under test.

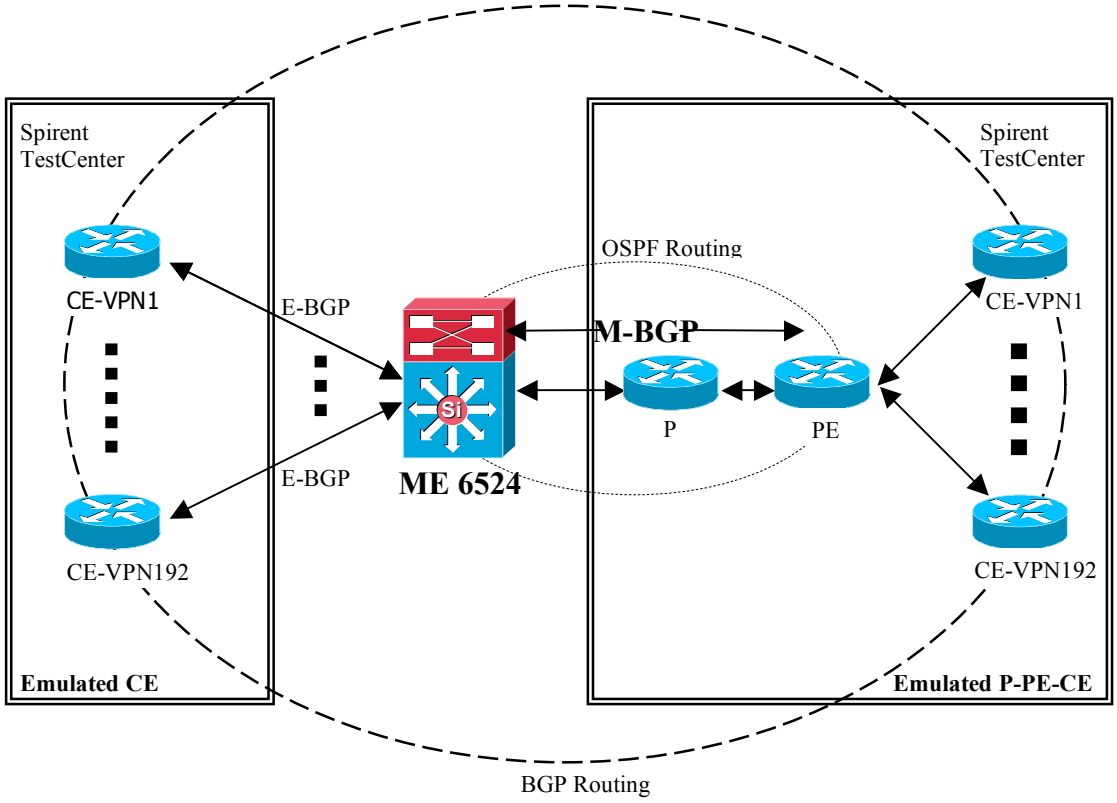

Figure 5: The MPLS VPN Test Bed

The router under test uses OSPF for connectivity with the emulated PE, and M-BGP for connectivity with emulated P and PE devices. Some 95 percent of VPN routes originate from the emulated P-PE-CE side, and the remaining 5 percent from the CE side. This simulates the common scenario where a customer site advertises a few networks and communicates with a far larger number of networks.

Once the routes were propagated, TestCenter then offered traffic to all ports in a bidirectional "backbone" pattern, where traffic from all ports on the CE side of the device under test was destined for all networks on all uplink ports, and vice-versa[3]. We offered test traffic using multiple frame sizes, in each case for a duration of 300 seconds.

The ME 6524 achieved throughput of nearly 12.5 million frames per second in this test. Table 6 below presents results from the MPLS VPN scalability and throughput tests in tabular form.

[3] As defined in the industry-standard switch testing methodology, RFC 2889, this pattern is officially known as a "partial mesh multiple device" topology.

MPLS VPN Scalability and Throughput (192 VRF instances, 230,400 routes, 300-second test duration)	
Frame length (bytes)	Aggregate throughput* (frames/second)
64	12,488,251
128	8,909,819
256	5,649,236
512	3,238,209
1,024	1,696,991
1,280	1,369,027
1,518	1,141,717
2,034	864,694
4,068	427,509
9,216	190,601

*Tested with 10,000-line ACL; QoS classification and re-marking; and Netflow monitoring on all packets

Table 6: MPLS VPN Scalability and Throughput

2.4 Carrier Ethernet Services

Service providers and network equipment manufacturers are extending Ethernet's reach, finding it an ideal technology for building DSL or FTTH access and aggregation networks. With metro Ethernet devices such as the ME 6524, it's possible to give each customer a private broadcast domain; provide per-port and per-VLAN service classification; and recover rapidly from link failures.

Metro Ethernet services impose some new requirements on access devices. While basic switch benchmarks such as MAC address table capacity still apply, it's also desirable to restrict or disable address learning on a per-VLAN basis. This ensures that no one customer's network can overwhelm a switch by flooding it with too many addresses. In addition, network redundancy and rapid failover are bedrock requirements in any sound network design, and switches must provide appropriate mechanisms to meet these requirements. Service providers also expect switches to support the emerging set of carrier Ethernet services defined by the Metro Ethernet Forum (MEF).

The ME 6524 is specifically intended for metro Ethernet service. It carries MEF 9 (Ethernet UNI) and MEF 14 (hard QOS) certifications for all relevant service definitions, including Ethernet private line (EPL), Ethernet virtual private line (EVPL), and Ethernet LAN (E-LAN) services. It also supports the following features for metro Ethernet service:

- Support for up to 96,000 dynamically learned MAC addresses
- IEEE 802.1q tunneling for customer VLAN transparency
- Layer 2 Protocol Tunneling (L2PT) support for layer-2 service transparency policers to provide CIR/PIR functions
- Classification and marking at the UNI interface, performed on a per-port, per-VLAN, or per-class of service basis
- Disabling or limiting MAC address learning on a per-VLAN basis
- Flexlink failover with fast reroute

In its testing of carrier Ethernet services on the ME 6524, Network Test focused on four areas: address learning capacity and learning rate; per-VLAN address limiting; per-VLAN address disabling; and failover using Cisco Flexlink and fast reroute.

2.4.1 MAC Address Learning Capacity and Rate

A fundamental benchmark for any Ethernet switch is its address learning capacity – the number of MAC addresses the device can learn and store in its layer-2 forwarding database. As described in RFC 2889, address learning capacity is determined by offering addresses to a switch and verifying that the switch does not broadcast, or "flood", any of the addresses. A separate but related metric is address learning rate, which determines how quickly a switch can populate its MAC address table.

Cisco claims the ME 6524 can support 96,000 dynamically learned MAC addresses. In our lab environment, Network Test verified that the ME 6524 can learn 90,000 MAC addresses. The difference can be explained by the non-random address pattern we used in testing. In production networks, MAC addresses are likely to be randomly distributed. Hence learning of all 96,000 addresses is more likely than in our test lab environment.

Table 7 below presents results from that test.

	Total	stc4 01- 1	stc4 01- 2	stc4 01- 5	stc4 01- 6	stc4 01- 8	stc4 02- 1	stc4 02- 2
Tx Signature Frames	90000	0	30000	0	30000	0	30000	0
Rx Signature Frames	90000	30000	0	30000	0	30000	0	0

Table 7: MAC Address Learning Capacity

In this test, we transmit 30,000 64-byte frames from TestCenter ports 1-2, 1-6, and 2-1, destined to addresses already learned on TestCenter ports 1-1, 1-5, and 1-8. A seventh

monitor interface, TestCenter port 2-2, does not receive any traffic, proving that no broadcasting or "flooding" of traffic took place.

To determine MAC address learning rate, we first ran the capacity test using an intended load of 30,000 frames per second on each port. After allowing all address timers to expire, we then reran the test at gigabit line rate, or 1,488,095 frames per second using 64-byte frames. The ME 6524 again learned all 90,000 addresses without flooding. Therefore, we concluded that the switch is capable of MAC address learning at line rate.

2.4.2 Disabling MAC Address Learning Per VLAN

There are situations where it may be desirable to disable MAC address learning for a given port or a given VLAN. For example, when provisioning a point-to-point layer-2 VPN to a customer, there is no need for dynamic address learning. In such situations, it is desirable to disable dynamic MAC address learning on a per-customer basis.

The ME 6524 allows disabling of MAC address learning both on a per-port and a per-VLAN basis. Network Test validated this capability by rerunning the same address capacity test as above, but this time with address learning disabled for the VLAN to which all interfaces belonged.

Table 8 below presents results from tests with address learning disabled. This time, note that the switch floods all 90,000 addresses to all ports – proving that address learning has not occurred.

	Total	stc4 01- 1	stc4 01- 2	stc4 01- 5	stc4 01- 6	stc4 01- 8	stc4 02- 1	stc4 02- 2
Tx Signature Frames	90000	0	30000	0	30000	0	30000	0
Rx Signature Frames	360000	90000	0	90000	0	90000	0	90000

Table 8: Disabling MAC Address Learning

2.4.3 Limiting MAC Address Learning Per VLAN

The ME 6524 also can limit the number of MAC addresses learned on a given VLAN, providing a useful method for service providers to restrict the number of MAC addresses allocated to a given customer. Address learning limitation also prevents a virus or worm at one customer's site from overrunning the switch's address table with rogue addresses, thus protecting all customers' traffic.

To verify the address limiting feature, Network Test configured the switch to learn 15,000 addresses on a given VLAN defined on one subinterface. We then configured Spirent TestCenter to offer 30,000 addresses destined to that switch port, with the expected outcome that 15,000 of the addresses would be flooded to the monitor port.

That is exactly what happened. Table 9 below presents results from the test. The switch correctly limited learning to 15,000 addresses (as received on TestCenter port 1-02), and flooded the other 15,000 frames (as seen on TestCenter port 1-03).

	Total	stc4 01- 1	stc4 01- 2	stc4 01- 3
Tx Signature Frames	30000	0	30000	0
Rx Signature Frames	30000	15000	0	15000

Table 9: MAC Address Limiting

2.4.4 Failover with Flexlink

Fast convergence after link or interface failure is a requirement in any network, and metro Ethernet networks are no exception. Consider the case of hub-and-spoke Metro Ethernet access network where a uPE router experiences a link failure on one of two redundant links to nPE routers.

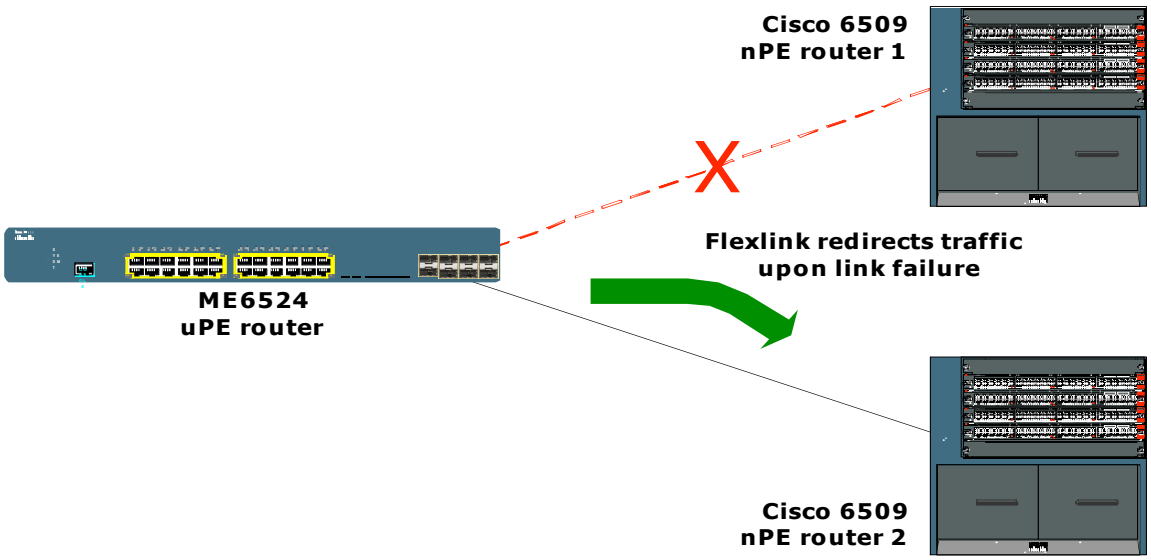

Figure 6: Rerouting Around Link Failure

There are multiple redundancy technologies available to handle this situation, but each carries tradeoffs. The obvious candidate is spanning tree protocol, but spanning trees are limited in size and require support on every switch in the network. Technologies like gateway load balancing protocol (GLBP) and hot standby routing protocol (HSRP) effectively route around failures, but these are layer-3 technologies and may not be appropriate for the layer-2 deployment described here.

Cisco offers an alternative with Flexlink, a layer-2 availability feature that provides rapid cutover around failed links or interfaces. Flexlink is local to the switch on which it's configured; unlike spanning tree, it does not need to be configured on each participating switch. Moreover, Flexlink rapidly points all affected MAC addresses to the secondary port upon failure of a primary port.

To measure failover time with Flexlink, Network Test constructed a test bed comprised of three ME 6524es, linked in a fashion similar to Figure 6 above. We loaded the switch representing the uPE device with 90,000 MAC addresses and then offered 64-byte frames to all addresses; thus, any link failure would force Flexlink to migrate all addresses to a secondary port. Once traffic generation began we physically removed a cable on the primary link, forcing a failover to a secondary link. We derived failover time from frame loss statistics.

Table 10 below shows results from the Flexlink failover tests. Note that frames are received on both the primary and secondary nPE devices (before and after the link failure, respectively). We obtain failover time by subtracting the sum of the received frames from the total transmitted frames, and dividing the result by the offered rate.

Flexlink Failover Time	
Tx frames	3,674,061
Sum of Rx frames	3,614,811
Difference (frame loss)	59,250
Tx rate (fps)	176,190
Failover time (milliseconds)	336.2

Table 10: Flexlink Failover Time

2.5 IP Multicast Performance

Applications such as IPTV and video on demand require network devices to replicate and forward a large number of multicast flows at high rates. The ME 6524 supports efficient distribution of multicast applications with a distributed replication architecture.

The ME 6524 supports all major IP multicast protocols, including internet group management protocol (IGMP), protocol-independent multicast (PIM), and PIM-source specific multicast (PIM-SSM). SSM reduces operational complexity by eliminating the need to configure rendezvous points. SSM requires IGMPv3, which is supported on the ME 6524. For set-top boxes that only support IGMPv2, the switch can translate IGMPv2 messages into PIM-SSM (S,G) addresses.

For pure layer-2 network designs, the switch also supports IGMP snooping and layer 2 multicast replication.

Network Test assessed multicast replication performance of the ME 6524 in layer-2 and layer-3 multicast scenarios. In both cases, we performed two tests: One with a single uplink port and 200 multicast sources, and again with eight uplink ports and 25 multicast sources. The first case represents an IPTV provider using a large number of broadcast servers to send content to subscribers. The second case represents a service provider offering video from multiple content suppliers, each with multiple video servers.

In all cases, Network Test used a total of 20,000 mroutes to assess device scalability. This number could represent hundreds of thousands of subscribers, all attached via a single ME 6524.

2.5.1 Layer-2 Multicast: 1 Uplink, 20,000 Mroutes

In this test, we configured the ME 6524 as a layer-2 device with 24 receiver ports on the access side, and one uplink port connected to multicast sources. We configured Spirent TestCenter to generate traffic to 20,000 unique (s,g) mroutes.

We configured TestCenter to transmit IGMP joins for 100 groups on each access port, and then to transmit multicast traffic via the switch's uplink port to all 100 groups on each access port. The total number of replicated ports in this test was 24.

This topology represents a subscriber base of nearly half a million subscribers. With 200 sources sending traffic to 100 groups, the total number of mroutes was 20,000. For each of the 20,000 mroutes, there is one subscriber on each of 24 access ports, making 480,000 multicast receivers.

Network Test assessed multicast replication performance by determining throughput for frame sizes ranging from 64 through 9,216 bytes. In all cases, the test duration was 300 seconds.

The ME 6524 achieved throughput of 12.5 million frames per second in this configuration. Table 11 below presents throughput results for all frame lengths.

Layer-2 Multicast Throughput, 1 Uplink (20,000 mroutes, 24 replicated ports, 300-second test duration)	
Frame length (bytes)	Aggregate throughput (frames/second)
64	12,500,000
128	7,094,295
256	3,706,564
512	1,897,533
1,024	964,940
1,280	772,847
1,518	651,890
2,034	488,122
4,068	243,640
9,216	107,839

Table 11: Layer-2 Multicast Throughput, 1 Uplink

2.5.2 Layer-2 Multicast: 8 Uplink, 20,000 Mroutes

For this test, we used all eight uplink ports of the ME 6524 and 24 receiver ports on the access side, again in a layer-2 device configuration. Using Spirent TestCenter, we generated traffic to 20,000 unique (s,g) mroutes.

We configured TestCenter to transmit IGMP joins for 100 groups for each group of three access ports, making a total of 800 groups joined in all. We then used TestCenter to transmit multicast traffic on each uplink port to all 100 groups on each group of three access ports. Unlike the previous test, where traffic was replicated to all 24 access ports, the total number of (s,g) replicated ports this time was three. In other words, the switch replicated each multicast frame from each source three times.
Network Test assessed multicast replication performance by determining throughput for frame sizes ranging from 64 through 9,216 bytes. In all cases, the test duration was 300 seconds.

The ME 6524 achieved throughput of 12.5 million frames per second in this configuration. Table 12 below presents throughput results for all frame lengths.

Layer-2 Multicast Throughput, 8 Uplinks (20,000 mroutes, 3 replicated ports, 300-second test duration)	
Frame length (bytes)	Aggregate throughput (frames/second)
64	12,500,000
128	7,094,295
256	3,683,398
512	1,897,533
1,024	964,940
1,280	772,847
1,518	651,890
2,034	488,122
4,068	243,640
9,216	107,839

Table 12: Layer-2 Multicast Throughput, 8 Uplinks

2.5.3 Layer-3 Multicast: 1 Uplink, 20,000 Mroutes

Our first layer-3 multicast test is similar to the first layer-2 test, with one key difference: In this test, the switch acts as an IP multicast router, running PIM-sparse mode (PIM-SM) on all interfaces.

As before, the topology for this test involves 24 receiver ports on the access side and one uplink port connected to multicast sources. We configured Spirent TestCenter to generate traffic to 20,000 unique (s,g) mroutes.

We configured TestCenter to transmit IGMP joins for 100 groups on each access port, and then to transmit multicast traffic via the switch's uplink port to all 100 groups on each access port. The total number of replicated ports in this test was 24.

This topology represents a subscriber base of nearly half a million subscribers. With 200 sources sending traffic to 100 groups, the total number of mroutes was 20,000. For each of the 20,000 mroutes, there is one subscriber on each of 24 access ports, making 480,000 multicast receivers.

Network Test assessed multicast replication performance by determining throughput for frame sizes ranging from 64 through 9,216 bytes. In all cases, the test duration was 300 seconds.

The ME 6524 achieved throughput of nearly 9.3 million frames per second in this test. Table 13 below presents throughput results for all frame lengths.

Layer-3 Multicast Throughput, 1 Uplink (20,000 mroutes, 24 replicated ports, 300-second test duration)	
Frame length (bytes)	Aggregate throughput (frames/second)
64	9,284,333
128	7,036,060
256	3,730,763
512	1,897,533
1,024	962,348
1,280	772,847
1,518	651,890
2,034	488,122
4,068	243,640
9,216	107,839

Table 13: Layer-2 Multicast Throughput, 1 Uplink

2.5.4 Layer-3 Multicast: 8 Uplinks, 20,000 Mroutes

Network Test also ran a test case with 8 uplinks sending traffic from 25 multicast sources apiece, and the switch functioning as an IP multicast router.

The test bed topology is identical to that of the layer-2 multicast test with 8 uplinks (see above). As before, TestCenter transmitted IGMP joins for 100 groups to each group of three access ports, for 800 groups joined in total. Then TestCenter transmitted multicast traffic to all 800 groups. For any given (s,g) mroute, the total number of replicated ports in this test was three.

Network Test assessed multicast replication performance by determining throughput for frame sizes ranging from 64 through 9,216 bytes. In all cases, the test duration was 300 seconds.

The ME 6524 achieved throughput of nearly 9.6 million frames per second in this test. This is higher than the layer-3 multicast results with one uplink port due to the lesser amount of replication involved (three ports, vs. 24 ports in the earlier test). Table 14 below presents throughput results for all frame lengths.

Layer-3 Multicast Throughput, 1 Uplink (20,000 mroutes, 24 replicated ports, 300-second test duration)	
Frame length (bytes)	Aggregate throughput (frames/second)
64	9,642,427
128	7,054,674
256	3,706,564
512	1,897,533
1,024	964,940
1,280	772,847
1,518	651,890
2,034	488,122
4,068	243,640
9,216	107,839

Table 14: Layer-2 Multicast Throughput, 1 Uplink

2.6 Hot-Swappable Fan Tray and Power Supplies

To help ensure maximum uptime, the ME 6524 supports hot-swapping of its fan tray and redundant power supplies. Cisco says temporary removal of a power supply or fan tray will not affect device operation.

To verify that claim, Network Test removed and reinserted a fan tray and one of two power supplies while running the IPv6 throughput tests. After each change in device state, the IOS CLI generated a syslog alert about the event. However, there was no packet loss or any other discernable effect on device performance.

3 Introducing Embedded Event Manager (EEM)

Given the strong correlation between uptime and revenue, service providers understandably consider network management at least as important as performance and scalability. It's critically important to monitor of the health of all elements within the network, and take corrective action as needed.

With Embedded Event Manager (EEM) running on Cisco ME 6500 Series Ethernet Switches, network managers can both monitor and modify network elements as needed. This policy-based framework allows switch administrators to load scripts (their own, or supplied by Cisco) that define what steps should be taken in response to virtually any aspect of switch behavior.

For example, EEM can reconfigure a switch to limit data rates when CPU utilization rises above a given threshold. EEM also can generate security warnings, alerting administrators when flow count exceeds normal levels and dynamically altering switch configuration in response.

EEM complements and extends existing network management systems (NMSs) such as SNMP. EEM offers numerous built-in SNMP functions, such as calling and manipulating object identifiers (OIDs).

But EEM also goes beyond SNMP in several ways. Because EEM runs onboard each Cisco ME 6500 Series Ethernet Switch, it does not require a centralized management server – a boon in situations where communications links fail, leaving the centralized server unable to "see" devices in the network. For high-frequency events such as DoS attacks, an onboard system is preferable to constant polling by a centralized system (assuming high polling rates are even possible). And EEM does not require predefined management information base (MIB) objects. Virtually any action an administrator can take from the IOS command line can be automated through EEM.

3.1 The EEM Architecture

Figure 1 illustrates the major components of the EEM architecture including event detectors, the embedded event manager server, and the policy director. Event detectors determine when certain actions occur within Cisco IOS software and notify the embedded event manager server. The policy director registers with the embedded event manager server to receive events, and it also implements policy actions. Once the policy director registers, the policy's actions take place whenever a defined event occurs.

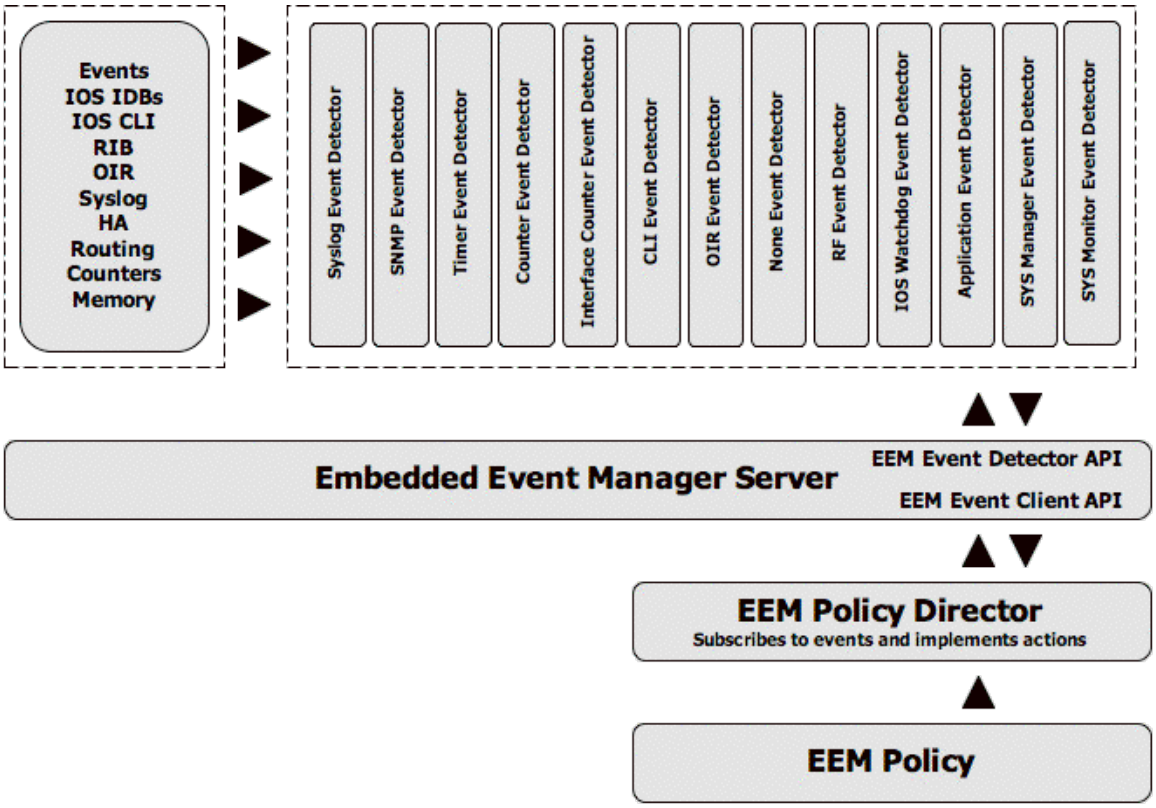

Figure 1: EEM Architecture View

In the software version assessed by Network Test[4], EEM supports 13 event detectors. These cover virtually anything that Cisco IOS software detects, such as interface counters, syslog and SNMP traps, system timers, and many more indicators of resource utilization. Again, EEM can "see" virtually anything Cisco IOS can.

Users can define policies in two ways: applets and TCL scripts. Applets offer a simple way to create and register policies from the switch CLI. While TCL scripts offer greater functionality and customization, users may find applets are preferable for getting started with EEM and adding basic management functions.

Creating an applet is a simple matter of entering commands at the CLI. The following sequence creates an applet to log any attempt to enter configuration mode:

```
ME 6524#conf t
ME 6524(config)#event manager applet my_applet
ME 6524(config-applet)#event cli pattern "conf t" sync no skip no
ME 6524(config-applet)#action 1.0 syslog msg "Configuration by
authorized personnel ONLY"
ME 6524(config-applet)#end
#
```

[4] The ME 6524 devices in this project ran IOS version 12.2(18)ZU. Network Test verified this version of IOS was available from the CCO Web site at test time.

Using the CLI event detector, "my_applet" will now generate a syslog entry for each access attempt to the switch's configuration mode. Note that entering applet commands is identical to performing any configuration task in IOS.

TCL scripts leverage the true power and flexibility of EEM. EEM-enabled versions of IOS on Cisco Catalyst 6500 Series switches include an interpreter for Tool Control Language (TCL), a simple and powerful scripting language.

Two types of TCL scripts can be found on the switch: Cisco supplied scripts (also known as mandatory scripts) and those loaded on the switch that are written by the user (user scripts). EEM places some restrictions on resource consumption for user-written scripts; this safeguard prevents a runaway user script from, for example, consuming all available CPU cycles. Mandatory scripts are more powerful, in that they can affect virtually all aspects of switch behavior. They are enabled by default on system startup but can be disabled from the switch CLI.

Both applets and TCL scripts can perform key network management tasks, including but not limited to the following:

- Execute an IOS CLI command and receive the result
- Send a CNS event
- Log a message to syslog
- Send an email or system page
- Increment or decrement an EEM counter
- Force a failover to the SSO standby supervisor
- Request system information
- Invoke another EEM policy to be started
- Reload the switch
- Send an SNMP trap with custom data
- Publish an application-specific EEM event
- Run a TCL Script

3.2 EEM Tests: Detecting Link Failure

Network Test verified EEM functionality with tests involving link failure; high CPU utilization; and abnormally high flow count.

In the link failure test, we configured EEM to send an email message to a network administrator upon failure of a link in the ME 6524. Using EEM commands, the switch would detect the link had failed, gather interface statistics, perform time-division reflectometry (TDR) on the link to show the approximate location of the fault, and send an email message with all this information to the network manager.

We forced the link failure by physically disconnecting a cable from the switch chassis. Within five seconds, this email message arrived:

```
To: noc@networktest.com
Cc:
Subject: Interface GigabitEthernet1/1 has gone down on 6524-POD4

Port        Align-Err    FCS-Err   Xmit-Err    Rcv-Err UnderSize OutDiscards
Gi1/1               0          0          0          0         0           0

Port      Single-Col Multi-Col  Late-Col Excess-Col Carri-Sen     Runs    Giants
Gi1/1              0         0         0          0         0         0         0

Port       SQETest-Err Deferred-Tx IntMacTx-Err IntMacRx-Err Symbol-Err
Gi1/1                0           0            0            0          1

TDR test last run on: August 14 15:39:43
Interface Speed Pair Cable length        Distance to fault   Channel Pair status
--------- ----- ---- ------------------- ------------------- ------- ------------
Gi1/1     1000  1-2  44   +/- 20 m       N/A                 Pair A  Terminated
                3-4  45   +/- 20 m       N/A                 Pair B  Terminated
                5-6  42   +/- 20 m       N/A                 Pair C  Terminated
                7-8  42   +/- 20 m       N/A                 Pair D  Terminated
```

As configured through EEM, the email message displays counter interface statistics and the results of TDR analysis of the cable. The TDR analysis estimates the cable length and distance to the cable fault (not applicable in this case, since all four pairs of wires are terminated, also shown in the TDR display).

This test illustrated EEM's ability to help diagnose physical-layer problems. As shown in the email output, EEM quickly gives network managers the information needed to identify and correct cable faults and other link problems.

3.3 EEM Tests: Controlling CPU Utilization

A common form of denial-of-service attack overwhelms a switch with frames that require processing by the switch's CPU. Once CPU utilization rises high enough, the switch may become sluggish or completely unresponsive. EEM can help avoid this form of DoS attack by monitoring CPU utilization and dynamically applying rate-limiting policies as needed.

To verify this feature, Network Test loaded an EEM applet that monitors CPU utilization and changes interface policies in response.

We began by verifying that no rate-limiting policy was in place at the start of the test. Then we offered the switch a stream of frames in which the IP time-to-live (TTL) value was set to 1. The ME 6524 decremented the TTL to 0 for each frame, rapidly sending CPU utilization above 90 percent. This, in turn, triggered the application of a rate-limiting policy on the interface receiving the attack frames. The switch's CPU utilization quickly fell to normal levels.

3.4 EEM Tests: Netflow Monitoring

To get a sense of the full power of EEM and TCL scripting, we ran another test that uses Cisco Netflow to monitor average and peak flow count, logs abnormally high numbers of flows, and dynamically modifies switch configuration in response.

Flow count monitoring (which is one form of what's known as "anomaly detection" in network security circles) is an effective method of detecting denial-of-service attacks and other unauthorized activity.

This EEM test uses a TCL script that periodically checks Netflow output on multiple interfaces to determine the average number of flows handled by the Cisco ME 6500 Series Ethernet Switch. If the number of flows exceeds a user-defined threshold, the script prints a syslog message and stops assessing average flow count.

This particular EEM script goes beyond passive monitoring. If it detects an abnormally flow count on any interface, it applies a QoS ACL to classify traffic and apply a policer, effectively shutting down a DoS attack or other unauthorized activity.

To test EEM's ability to detect Cisco Netflow events, we combined the Spirent TestCenter traffic generator with Spirent ThreatEx, an attack generation and analysis tool, to offer three classes of traffic. First, TestCenter offered a moderate number of background flows of "normal" traffic – not enough to trip the EEM Netflow script's threshold.

Second, ThreatEx offered 500 flows of "SQL Slammer," a well-known worm attack that propagates rapidly from host to host, from networks directly attached to the switch. Finally, ThreatEx also offered 500 additional flows of SQL Slammer traffic from spoofed IP addresses. We used two classes of attack traffic to emulate attacks from within the network (simulating a compromised host on a local subnet) and from external sources, including spoofed sources.

For attacks from within the network, the script would dynamically assign a QoS ACL to the affected interface, rate-limiting traffic on that interface. For attacks from spoofed sources, Cisco's engineers used another IOS feature, unicast Reverse Path Filtering (uRPF), that performs reverse lookups on source addresses and drops any packets whose next hop is not on the ingress interface[5]. In this way, the switch protected the network from both inside and outside attack.

Network Test observed the switch configuration both before and after the test. Once the attacks had been launched, we observed three responses. First, the EEM script generated a syslog entry noting the abnormally high flow count. Second, for attacks sourced from the local network, the EEM script dynamically installed a new QoS ACL on the source

[5] uRPF was first described in RFC 2827. The IETF labels the same document as "Best Current Practice 38" for its description of the security benefits of ingress filtering.

interface to mitigate the attack. Third, uRFP succeeded in blocking attack traffic from spoofed IP source addresses.

For both types of attack traffic, the switch had no advance known of the exact source address the attack traffic would use. We ran the tests multiple times using different source addresses unknown to the switch or to Cisco's engineers. In each test, the EEM script correctly identified the source address in its syslog message.

Our tests verified that EEM offers a simple, dynamic, and powerful method for device management. EEM allows the Cisco ME 6500 Series Ethernet Switch to become an active rather than a passive participant in network management, responding dynamically as network and system conditions require.

4 Conclusion

Network Test successfully validated Cisco's claims for the ME 6524 for use in Metro Ethernet service. The switch scales to provide thousands of pseudowires, reroutes quickly around failed links, and supports massive unicast and multicast routing tables. Because the ME 6524 uses the same Supervisor 720 engine as larger members of the Cisco Catalyst 6500 switch family, it can forward traffic at rates of up to 15 million fps. Further, Embedded Event Manager (EEM) automates the monitoring (and dynamic response, if needed) of virtually any aspect of switch behavior.

The ME 6524 is fully certified for metro Ethernet service. The Metro Ethernet Forum has issued its MEF 9 (Ethernet UNI) and MEF 14 (hard QOS) certifications for all relevant service definitions, including Ethernet private line (EPL), Ethernet virtual private line (EVPL), and Ethernet LAN (E-LAN) services.

The switches has already won two industry awards. At Globalcomm 2006, the ME 6524 won the Excellence in Technology award in the backbone/edge category. At Interop Tokyo 2006, was the winner for Infrastructure Building Products (Middle Range).

networktest

About Network Test

Network Test is an independent benchmarking and network design consultancy in Westlake Village, California, founded in 1999. We test networking equipment and live networks. Our clients include equipment manufacturers, large enterprises, service providers, industry consortia, and trade publications. Independence and rigorous adherence to standards are the hallmarks of Network Test's benchmarks. Whenever clients need to measure performance, scalability, interoperability, or protocol conformance, they can rely on Network Test for unbiased, in-depth analysis.

For more information:

Network Test Inc.
31324 Via Colinas, Suite 113
Westlake Village, CA 91362-6761
+1-818-889-0011 voice
+1-818-865-0033
info@networktest.com

Acknowledgements

Network Test gratefully acknowledges the support of Spirent Communications, which supplied engineering expertise as well as the Spirent TestCenter traffic generator/analyzer and Spirent ThreatEx attack generator for this project. Spirent test engineer Mark Hall provided configuration and troubleshooting assistance for the Spirent TestCenter application used in these tests, and Phil Trainor configured the ThreatEx appliance with attacks used in the EEM tests.