

Cisco's virtual switch smashes throughput records

Exclusive tests show new Catalyst 6500 management blade sets records for recovery times and throughput

CLEAR CHOICE TEST

By David Newman,
Network World Lab Alliance

Virtualization, long a hot topic for servers, has entered the networking realm. With the introduction of a new management blade for its Catalyst 6500 switches, Cisco can make two switches look like one while dramatically reducing failover times in the process.

In an exclusive Clear Choice test of Cisco's new Virtual Switching System (VSS), *Network World* conducted its largest-ever benchmarks to date, using a mammoth test bed with 130 10G Ethernet interfaces.

The results were impressive: VSS not only delivers a 20fold improvement in failover times but also eliminates layer-2 and layer-3 redundancy protocols at the same time.

The performance numbers are even more startling: A VSS-enabled virtual switch moved a record 770 million frames per second in one test, and routed more than 5.6 billion unicast and multicast flows in another. Those numbers are exactly twice what a single physical Catalyst 6509 can do.

All links, all the time

To maximize uptime, network architects typically provision multiple links and devices at every layer of the network, using an alphabet soup of redundancy protocols to protect against downtime. These include rapid spanning tree protocol (RSTP), hot standby routing protocol (HSRP), and virtual router redundancy protocol (VRRP).

This approach works, but has multiple downsides. Chief among them is the "active-passive" model used

by most redundancy protocols, where one path carries traffic while the other sits idle until a failure occurs. Active-passive models use only 50% of available capacity, adding considerable capital expense.

Further, both HSRP and VRRP require three IP addresses per subnet, even though routers use only one address at a time. And while rapid spanning tree recovers from failures much faster than the original spanning tree, convergence times can still vary by several seconds, leading to erratic application performance. (Strictly speaking, spanning tree was intended only to prevent loops, but it's commonly used as a redundancy mechanism.)

There's one more downside to current redundant network designs: They create twice as many network elements to manage. Regardless of whether network



CLEAR CHOICE TEST

NETRESULTS

Product	Virtual Switching Supervisor 720-10G
Vendor	Cisco www.cisco.com
Price	Starts at \$31,500
Pros	Eliminates need for L2 and L3 redundancy protocols; doubles fabric capacity; reduces failover times; works with servers as well as switches.
Cons	Virtualization is proprietary; switch fabric is blocking in some configurations.
Score	4.625

SCORECARD

Action	Weight	Score
Features	25%	4.5
Performance	25%	4.0
Resiliency	25%	5.0
Management	25%	5.0
Total score		4.625

Scoring key: 5: Exceptional; 4: Very good; 3: Average; 2: Below average; 1: Subpar or not available.

managers use a command-line interface or an SNMP-based system for configuration management, any policy change needs to be made twice, once on each redundant component.

Introducing Virtual Switching

In contrast, Cisco's VSS uses an "active-active" model that retains the same amount of redundancy, but makes use of all available links and switch ports.

While many vendors support link aggregation (a means of combining multiple physical interfaces to appear as one logical interface), VSS is unique in its ability to virtualize the entire switch — including the switch fabric and all interfaces. Link aggregation and variations such as Nortel's Split Multi-Link Trunk (SMLT) do not create virtual switches, nor do they eliminate the need for layer-3 redundancy mechanisms such as HSRP or VRRP.

At the heart of VSS is the Virtual Switching Supervisor 720-10G, a management and switch fabric blade for Cisco Catalyst 6500 switches. VSS — which began shipping last month — requires two new supervisor cards, one in each physical chassis. The management blades create a virtual switch link (VSL), making both devices appear as one to the outside world: There's just one media access control and one IP address used, and both systems share a

common configuration file that covers all ports in both chassis.

On the access side of Cisco's virtual switch, downstream devices still connect to both physical chassis, but a bonding technology called Multichassis EtherChannel (MEC) presents the virtual switch as one logical device. MEC links can use industry-standard 802.3ad link aggregation or Cisco's proprietary port aggregation protocol. Either way, MEC eliminates the need for spanning tree. All links within a MEC are active until a circuit or switch failure occurs, and then traffic continues to flow over the remaining links in the MEC.

Servers also can use MEC's link aggregation support, with no additional software needed. Multiple connections were already possible using "NIC teaming," but that's usually a proprietary, active/passive approach.

On the core side of Cisco's virtual switch, devices also use MEC connections to attach to the virtual switch. This eliminates the need for redundancy protocols such as HSRP or VRRP, and also reduces the number of routes advertised. As on the access side, traffic flows through the MEC in an "active/active" pattern until a failure, after which the MEC continues to operate with fewer elements.

The previous examples focused on distribution-layer switches, but VSL links work between any two Catalyst 6500 chassis. For example, virtual switching can be used at both core and distribution layers, or at the core, distribution and access layers. All attached devices would see one logical device wherever a virtual switch exists.

A VSL works only between two chassis, but it can support up to eight physical links. Multiple VSL links can be established using any combination of interfaces on the new supervisor card or Cisco's WS-6708 10G Ethernet line card. VSS also requires line cards in Cisco's 67xx series, such as the 6724 and 6748 10/100/1000 modules or the 6704 or 6708 10G Ethernet modules. Cisco says VSL control traffic uses less than 5% of a 10G Ethernet link, but we did not verify this.

At least for now, VSL traffic is proprietary. It isn't possible to set up a VSL between, say, a Cisco and Foundry switch.

A big swath of fabric

We assessed VSS performance with tests focused on fabric bandwidth and delay, failover times and unicast/multicast performance across a network backbone.

In the fabric tests we sought to answer two simple questions: How fast does VSS move frames, and how long does it hang on to each frame? The setup for this test was anything but simple. We attached Spirent TestCenter analyzer/generator modules to 130 10G Ethernet ports on two Catalyst 6509 chassis configured as one virtual switch.

These tests produced, by far, the highest throughput we've ever measured from a single (logical) device. When forwarding 64-byte frames, Cisco's virtual switch moved traffic at more than 770 million frames per second. We then ran the same test on a single switch, without virtualization, and measured throughput of 385 million frames per second — exactly half the result of the two fabrics combined in the virtual switch. These results prove there's no penalty for combining switch fabrics.

We also measured VSS throughput for 256-byte frames (close to the average Internet frame length) of 287 million frames per second and for 1,518-byte frames (until recently, the maximum in Ethernet, and still the top end on most production networks) of 53 million frames per second. With both frame sizes, throughput was exactly double that of the single-switch case.

The 1,518-byte frames per second number represents throughput of nearly 648Gbps. This is only around half the theoretical maximum rate possible with 130 10G Ethernet ports. The limiting factor is the Supervisor 720 switch fabric, which can't send line-rate traffic to all 66 10G ports in each fully loaded chassis. VSS doubles fabric capacity by combining two switches, but it doesn't extend the capacity of the fabric card in either physical switch.

We also measured delay for all three frame sizes. With a 10% intended load, Spirent TestCenter reported average delays ranging from 12 to 17 microseconds, both with and without virtual switching. These numbers are similar to those for other 10G switches we've tested, and far below the point where they'd affect performance of any application. Even the maximum delays of around 66 microsec with virtual switching again are too low to slow down any application, especially considering Internet round-trip delays often run into the tens of milliseconds.

Faster failovers

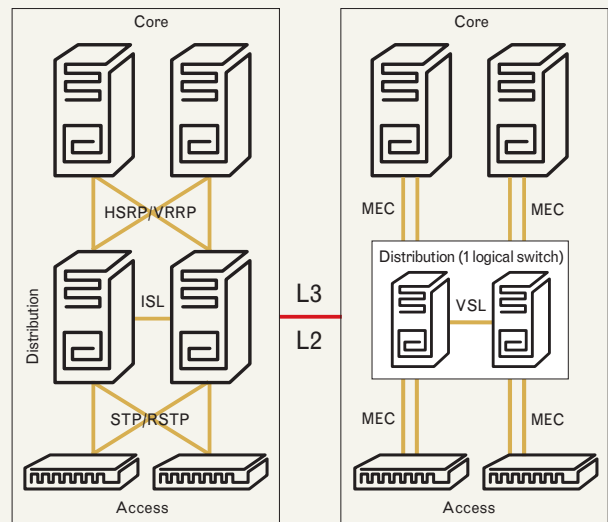
Our failover tests produced another record: The fastest recovery from a Layer 2/Layer 3 network failure we've ever measured.

We began these tests with a conventional setup: Rapid spanning tree at layer 2, HSRP at layer 3 and 16,000 hosts (emulated on Spirent TestCenter) sending traffic across redundant pairs of access, distribution and core switches. During the test, we cut off power to one of the distribution switches, forcing all redundancy mechanisms and routing protocols to reconverge. Recovery took 6.883 seconds in this setup.

Then we reran the same test two more times with VSS enabled. This time convergence occurred much faster. It took the network just 322 millisecond to converge with virtual switching on the distribution switches, and 341 millisecond to converge with virtual switching on the core and distribution switches. Both numbers represent better than 20fold improvements over the usual redundancy mechanisms.

Virtual switching reduces complexity, boosts capacity

The "active/passive" redundancy model in most network designs doubles port and link requirements (at left). The redundancy protocols that must be configured include spanning tree or rapid spanning tree (STP and RSTP) at layer 2 and hot standby routing protocol (HSRP) and virtual routing redundancy protocol (VRRP) at layer 3. In contrast, Cisco's virtual switching system, creates an "active/active" design by using all available link and switch ports (at right). Switches and routers still use multiple physical links to connect to a virtual switch for redundancy, but now these other devices use a Multichassis EtherChannel (MEC) in which all links are active until a failure occurs.



A bigger backbone

Our final tests measured backbone performance using a complex enterprise traffic pattern involving 176,000 unicast routes, more than 10,000 multicast routes, and more than 5.6 billion flows. We ran these tests with unicast traffic alone and a combination of unicast and multicast flows, and again compared results with and without VSS in place.

Just to keep things interesting, we ran all tests with a 10,000-entry access control list in place, and also configured switches to re-mark all packets' diff-serv code point (DSCP) fields. Re-marking DSCPs prevents users from unauthorized "promotion" of their packets to receive higher-priority treatment. In addition, we enabled NetFlow tracking for all test traffic.

Throughput in all the backbone cases was exactly double with virtual switching than without it. This was true for both unicast and mixed-class throughput tests, and also true regardless of whether we enabled virtual switching on distribution switches alone, or on both the core and distribution switches. These results clearly show the advantages of an "active/active" design over an "active/passive" one.

We measured delay as well as throughput in these tests. Ideally, we'd expect to see little difference between test cases with and without virtual switching, and between cases with virtual switching at one or two layers in the network. When it came to average delay,

CLEAR CHOICE TEST

that's pretty much how things looked. Delays across three pairs of physical switches ranged from around 26 to 90 microsec in all test cases, well below the point where applications would notice.

Maximum delays did vary somewhat with virtual switching enabled, but not by a margin that would affect application performance. Curiously, maximum delay increased the most for 256-byte frames, with fourfold increases over results without virtual switching. The actual amounts were always well less than 1

millisec, and also unlikely to affect application performance.

Cisco's VSS is a significant advancement in the state of the switching art. It dramatically improves availability with much faster recovery times, while simultaneously providing a big boost in bandwidth.

Newman is president of Network Test, an independent test lab in Westlake Village, Calif. He can be reached at dnewman@networktest.com.

