
networktest

**FlexNetwork Architecture Delivers
Higher Speed, Lower Downtime
With HP IRF Technology**

August 2011

HP IRF Performance Assessment

Executive Summary

HP commissioned Network Test to assess the performance of Intelligent Resilient Framework (IRF), a method of virtualizing data center switch fabrics for enhanced bandwidth and availability.

On multiple large-scale test beds, IRF clearly outperformed existing redundancy mechanisms such as the spanning tree protocol (STP) and the virtual routing redundancy protocol (VRRP).

Among the key findings of these tests:

- Using VMware's vMotion facility, average virtual machine migration time took around 43 seconds on a network running IRF, compared with around 70 seconds with rapid STP
- IRF virtually doubled network bandwidth compared with STP and VRRP, with much higher throughput rates regardless of frame size
- IRF converged around failed links, line cards, and systems vastly faster than existing redundancy mechanisms such as STP
- In the most extreme failover case, STP took 31 *seconds* to recover after a line card failure; IRF recovered from the same event in 2.5 *milliseconds*
- IRF converges around failed network components far faster than HP's 50-millisecond claim

This document briefly explains IRF concepts and benefits, and then describes procedures and results for IRF tests involving VMware vMotion; network bandwidth; and resiliency.

Introducing IRF

IRF consolidates multiple physical switches so that they appear to the rest of the network as a single logical entity. Up to nine switches can comprise this virtual fabric, which runs on HP's high-end switch/routers¹ and can encompass hundreds or thousands of gigabit and 10-gigabit Ethernet ports.

IRF offers advantages in terms of simpler network design; ease of management; disaster recovery; performance; and resiliency. A virtual fabric essentially "flattens" the data center from three layers into one or two using HP Virtual Connect technology, requiring fewer switches.

Device configuration and management also becomes simpler. Within an IRF domain, configuration of a single primary switch is all that's needed; the primary switch then distributes relevant configuration and protocol information to other switches in the IRF domain. IRF also supports an in-service software upgrade (ISSU) capability that allows individual switches to be taken offline for upgrades without affecting the rest of the virtual fabric.

For disaster recovery, switches within an IRF domain can be deployed across multiple data centers. According to HP, a single IRF domain can link switches up to 70 kilometers (43.5 miles) apart.

¹ IRF support is included at no cost on HP 12500, 9500, 7500, 58xx, and 55xx switches.

IRF improves performance and resiliency, as shown in test results described later in this report. A common characteristic of existing data center network designs is their inefficient redundancy mechanisms, such as STP or VRRP.

Both these protocols (along with modern versions of STP such as rapid STP and multiple STP) use an “active/passive” design, where only one pair of interfaces between switches forwards traffic, and all others remain idle until the active link fails. With active/passive mechanisms, half (or more) of all inter-switch links sit idle most of the time. Moreover, both STP and VRRP take a relatively long time to recover from link or component faults, typically on the order of seconds.

IRF uses an “active/active” design that enables switches to forward traffic on all ports, all the time. This frees up bandwidth, boosting performance for all applications. Data centers using virtualization benefit especially well from this design, since the additional bandwidth allows virtual machines to be moved faster between hypervisors. IRF’s active/active designs also reduce downtime when link, component, or system failures occur.

IRF Speeds VMware Performance

Over the past few years VMware’s vMotion capability has become the “killer app” for large-scale data centers and cloud computing. The ability to migrate virtual machines between physical hosts with zero downtime is a boon to network managers, but also a challenge. As data centers scale up in size and network managers use vMotion to migrate ever-larger numbers of virtual machines, network performance can become a bottleneck. This is an acute concern for disaster recovery and other high-availability applications, where rapid migration of virtual machines is essential.

Network Test and HP engineers constructed a large-scale test bed to compare vMotion performance using IRF and rapid spanning tree protocol (RSTP). With both mechanisms, the goal was to measure the time needed for vMotion migration of 128 virtual machines, each with 8 Gbytes of RAM, running Microsoft SQL Server on Windows Server 2008. Before each migration event, test engineers verified maximum memory usage on each VM, ensuring the most stressful possible load in terms of network utilization. The test migrated virtual machines between VMware ESXi hosts running on a total of 32 HP BL460 G7 blade servers.

In the RSTP case, the network used a typical active/passive design, with some ports forwarding traffic between access and core switches, and others in blocking state (see Figure 1). In this design, RSTP provides excellent loop prevention but also limits available bandwidth; note that half the inter-switch connections shown here are in blocking state.

HP IRF Performance Assessment

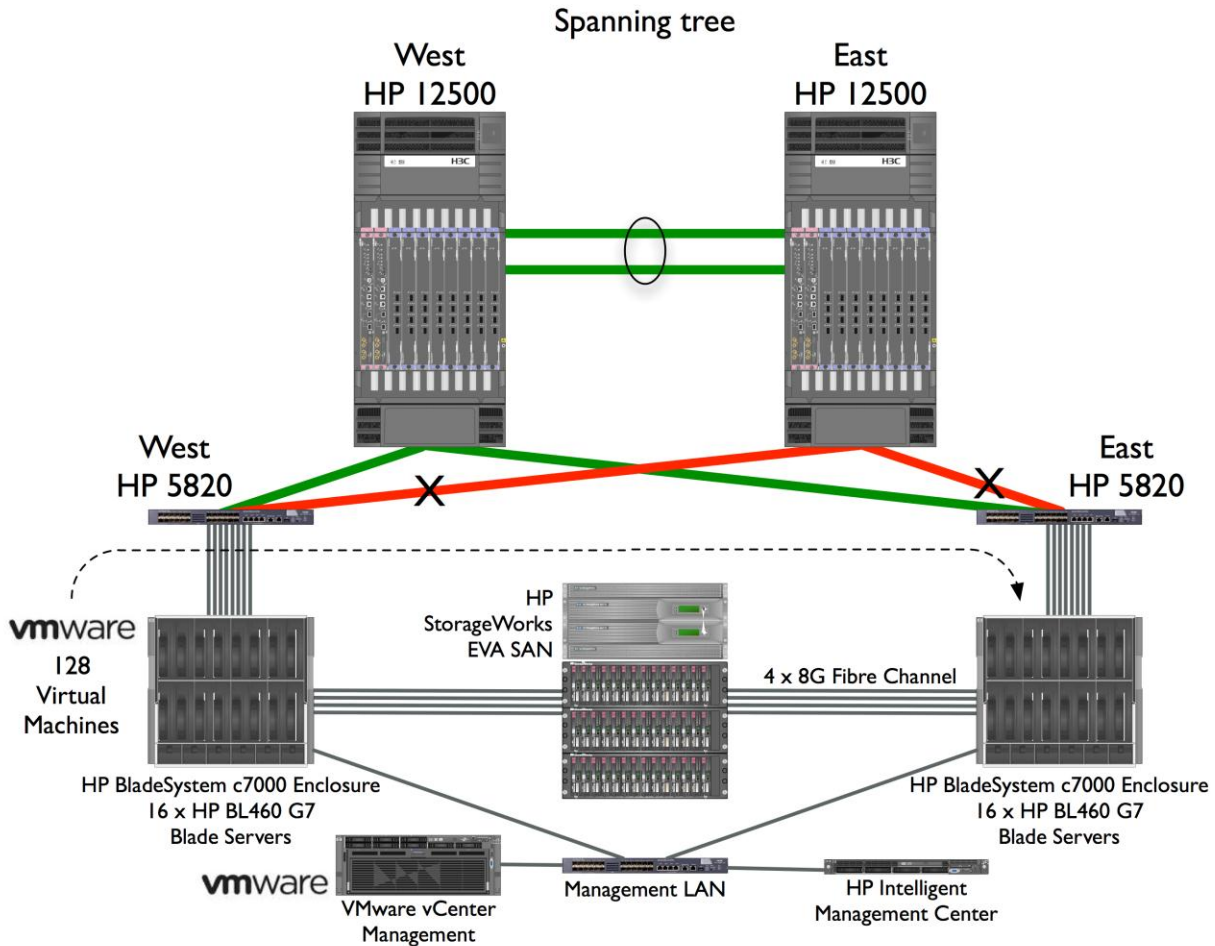


Figure 1: VMware vMotion with RSTP

By virtualizing the core switching infrastructure, IRF increases network capacity (see Figure 2). Here, all inter-switch ports are available, with no loss in redundancy compared with spanning tree. The core switches appear to the rest of the network as a single logical device. That converged device can forward traffic on all links with all attached access switches.

Moreover, IRF can be used together with the link aggregation control protocol (LACP) to add capacity to links between switches (or between switches and servers), again with all ports available all the time. This isn't possible with STP or RSTP since at least half (and possibly more) of all inter-switch links must remain in blocking state at all times.

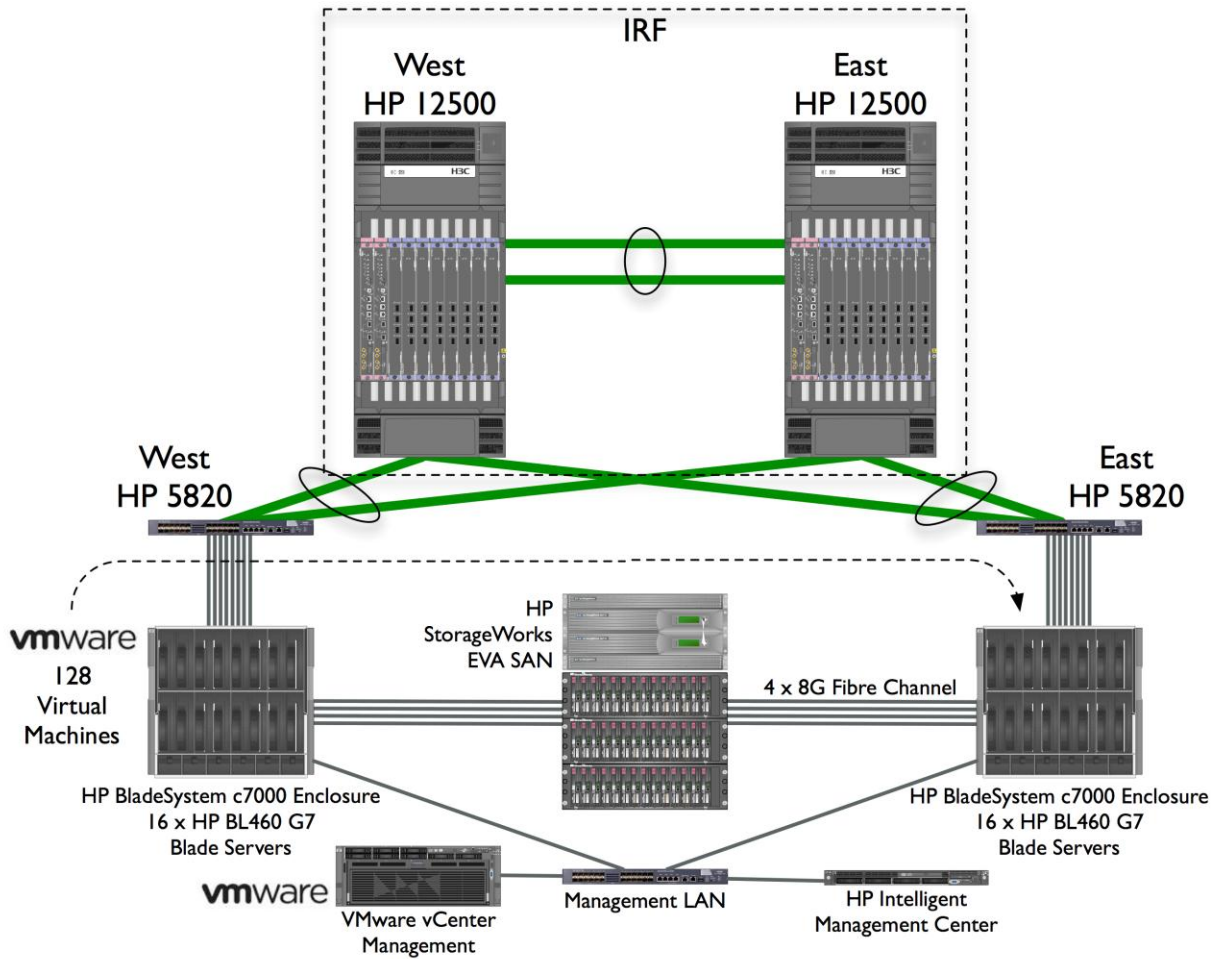


Figure 2: VMware vMotion with IRF

For both RSTP and IRF scenarios, engineers used custom-developed scripts to trigger vMotion migration of 128 virtual machines. In all three trials run, **IRF clearly outperformed RSTP in terms of average vMotion migration time** (see Figure 3). On average, migrations over RSTP took around 70 seconds to complete, while vMotion over IRF took 43 seconds.

HP IRF Performance Assessment

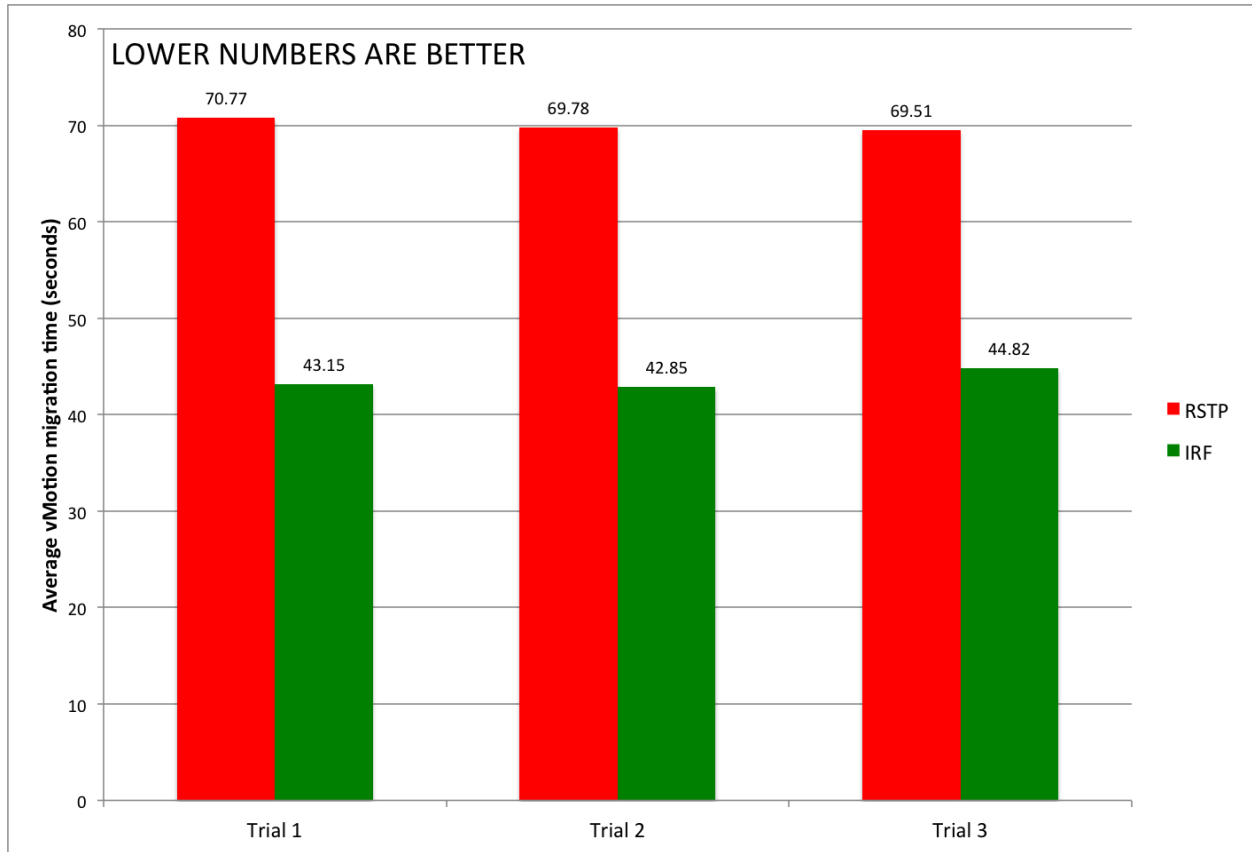


Figure 3: Average vMotion times over RSTP and IRF

Boosting Bandwidth: IRF Speeds Throughput

It's important to note that actual vMotion times depend on many variables, including server and storage resources, virtual machine configuration, and VMware vSphere parameters. As with all application performance tuning, results can vary from site to site. To determine a more basic metric – raw network capacity – Network Test ran additional tests to compare bandwidth available with and without IRF.

Network Test conducted throughput tests comparing IRF with STP at layer 2 and VRRP at layer 3.²

Figure 4 below shows the test beds used to compare throughput in the STP and IRF test cases. While both tests involve the same 12500 and 5820 switches, note that half the ports between them are in blocking state in the STP test cases (seen in the left side of Figure 4).

The IRF configuration makes use of all inter-switch links (see in the right of the figure), with the two 12500 switches appearing to the network as a single entity. The test bed for VRRP was similar to that shown here, except that Network Test used 20 gigabit Ethernet connections between each traffic generator and the 5820s to increase emulated host count and ensure more uniform distribution of

² Network Test did not compare IRF and rapid spanning tree protocol [RSTP] throughput because RSTP results would be identical to those with STP in this particular configuration.

traffic across VRRP connections. Engineers configured the 12500 switches in “bridge extended” mode, which improves performance by allocating additional memory to switching processes.

The IRF configuration uses the link aggregation control protocol (LACP) on connections between the 12500 and 5820 switches. From the perspective of the 5820 access switches, the link-aggregated connection to the core appears to be a single virtual connection. This allows for interesting network designs in disaster-recovery scenarios, for example with IRF and link aggregation connecting switches in different physical locations.

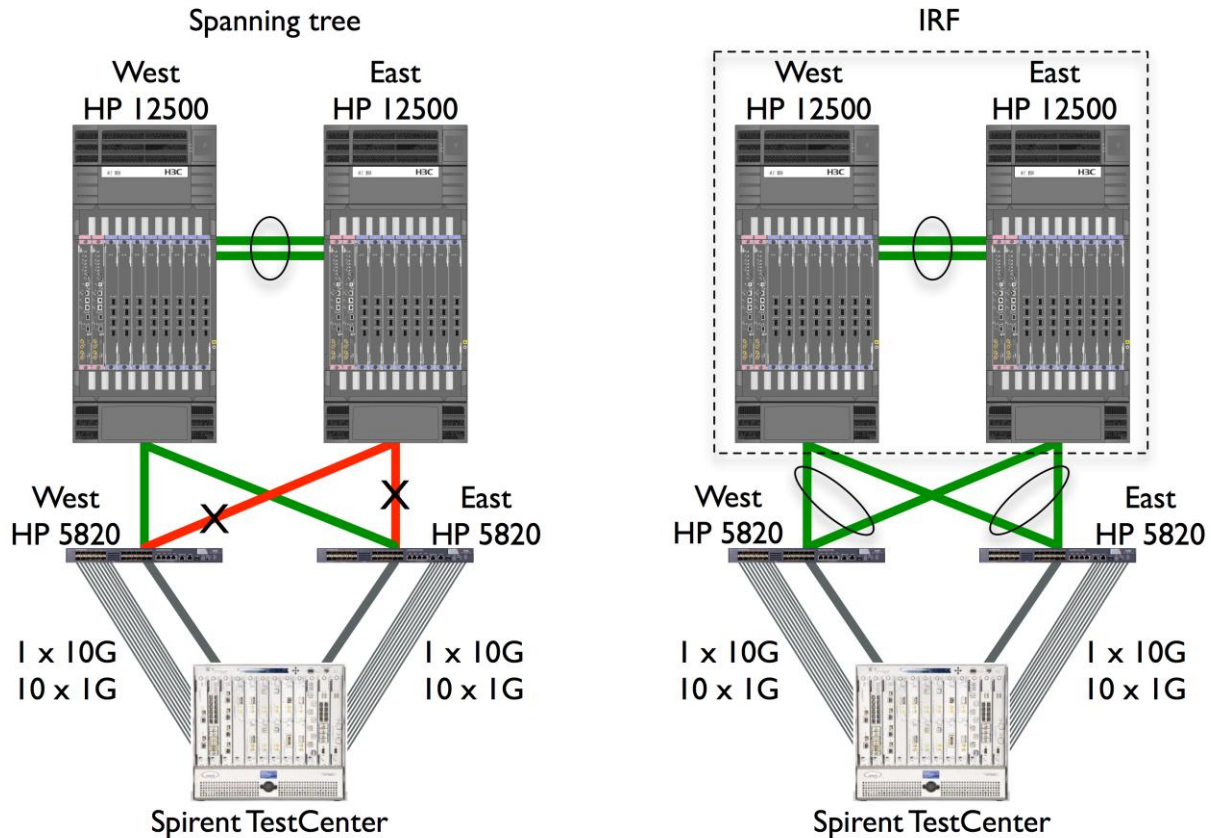


Figure 4: Spanning tree and IRF test beds

Network Test followed the procedures described in [RFCs 2544](#) and [2889](#) to determine system throughput. Engineers configured the Spirent TestCenter traffic generator/analyzer to offer bidirectional traffic in an “east/west” direction, meaning all frames on the “west” side of the test bed were destined for the “east” side and vice-versa. The aggregate load offered to each side was equivalent to 20 Gbit/s of traffic, equal to the theoretical maximum capacity of the access-core links.

To measure throughput, engineers offer traffic at varying loads, each for a 60-second duration, to determine the highest rate at which the switches would forward all frames with zero frame loss. As defined in [RFC 1242](#), this is the throughput rate.

HP IRF Performance Assessment

Engineers repeated these tests with various frame sizes ranging from 64 bytes (the minimum in Ethernet) through 1,518 bytes (the nominal maximum in Ethernet) through 2,176 bytes (often seen in data centers that use Fibre Channel for storage) through 9,216 (the nonstandard but still widely used jumbo frames common in data centers).

Figure 5 below presents throughput results, expressing the throughput rate as a percentage of the theoretical maximum rate. **For all frame sizes, IRF nearly doubled channel capacity**, delivering near line-rate throughput while the active/passive solutions delivered throughput of only 50 percent of channel capacity.

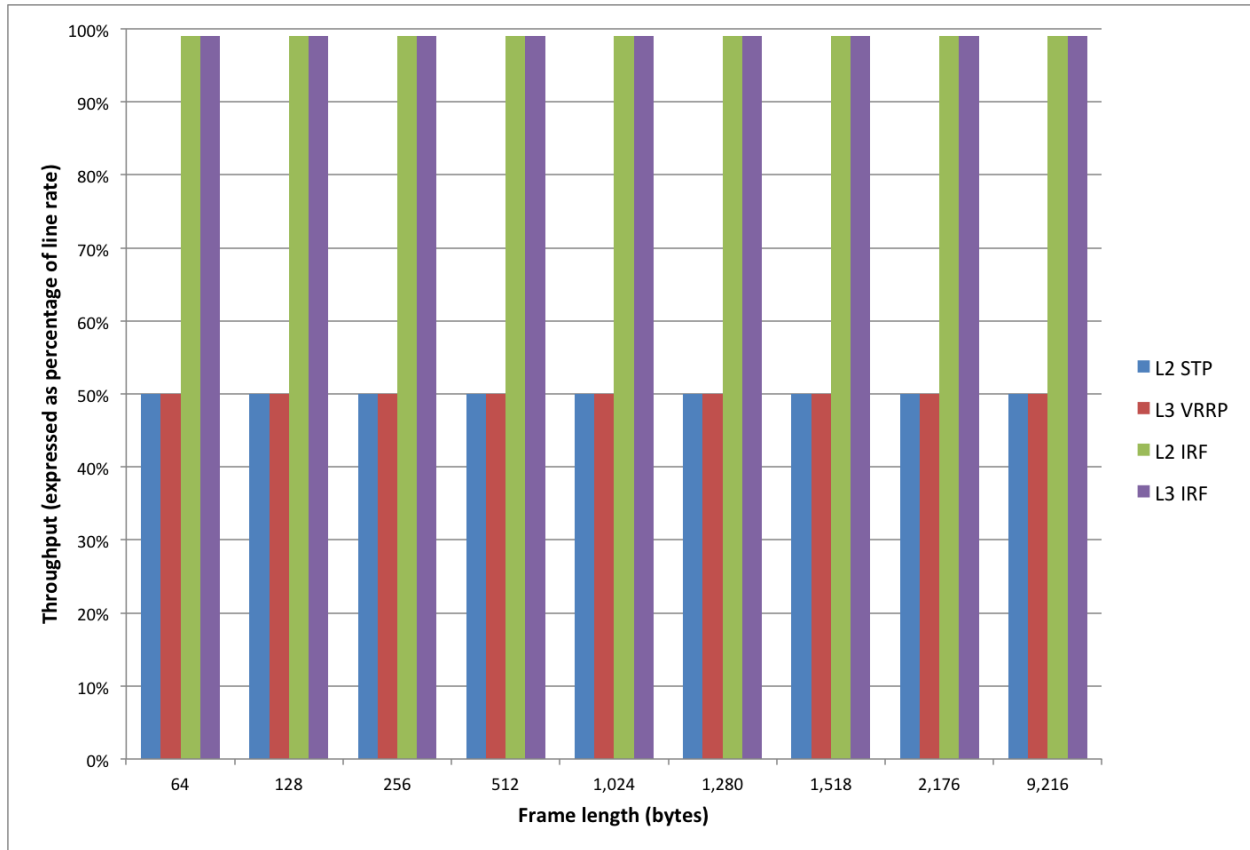


Figure 5: Throughput as a percentage of the theoretical maximum

Note also that **bandwidth utilization nearly doubles with IRF in every test case**, regardless of frame length. **This validates IRF’s ability to deliver far more network bandwidth, which in turn speeds performance for all applications, regardless of traffic profile.**

The throughput figures presented in Figure 5 are given as percentages of theoretical line rate. As a unit of measurement, throughput itself is actually a rate and not a percentage. Figure 6 below presents the same data from the throughput tests, this time with throughput expressed in frames per second for each test case. **Regardless of how it’s expressed, IRF provides nearly double the network bandwidth as other layer-2 and layer-3 resiliency mechanisms.**

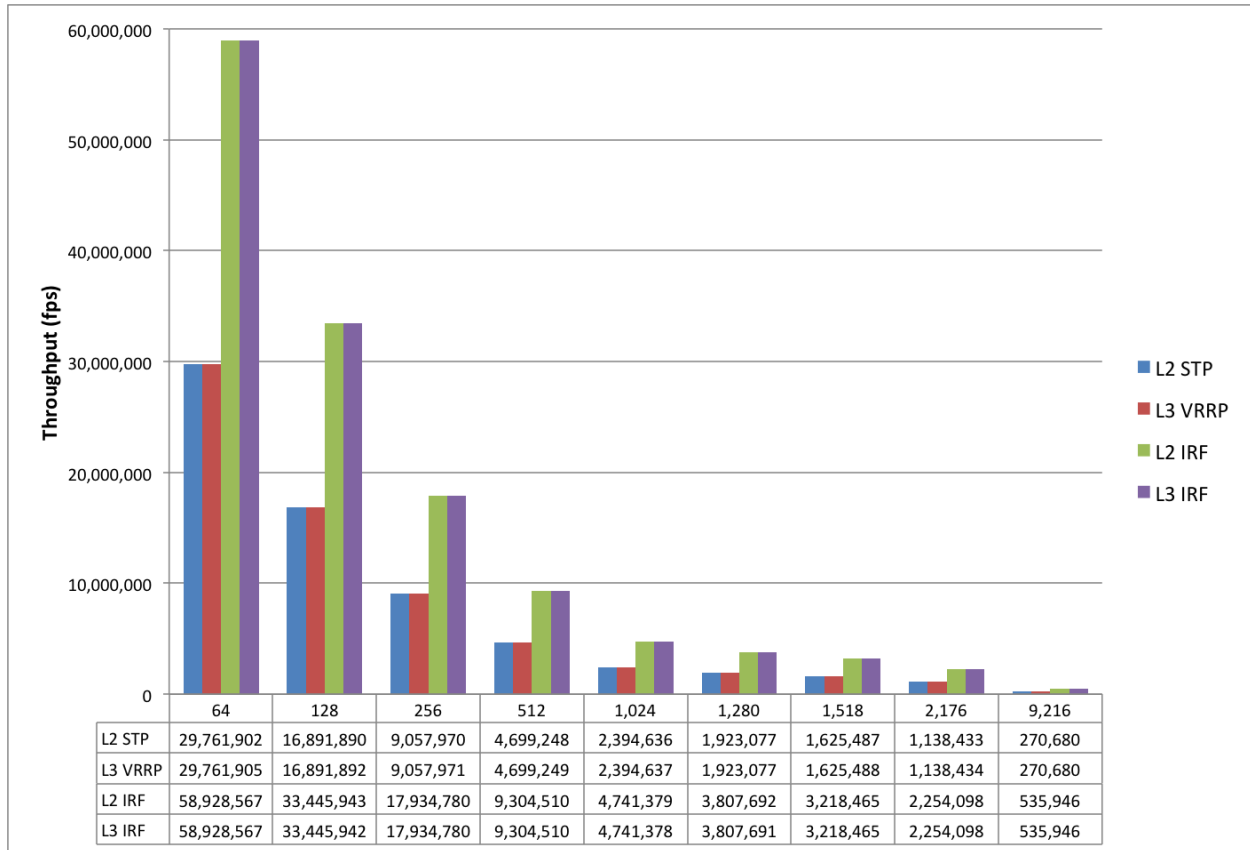


Figure 6: Comparing STP, VRRP, and IRF throughput

Faster Failovers: IRF Improves Resiliency

While *high performance* is certainly important, *high availability* is an even more critical requirement in enterprise networking. This is especially true in the data center, where even a small amount of downtime can mean significant revenue loss and other disruptions. IRF aims to improve network uptime by recovering from link or component failures far faster than mechanisms such as STP, RSTP, or VRRP.

Networks running STP can take between 30-60 seconds to converge following a single link or component failure. Rapid spanning tree and VRRP are newer and faster, but convergence time can still be significant. HP claims IRF will converge in 50 milliseconds.

To assess that claim, Network Test used the same test bed as in the STP and IRF throughput tests (see Figure 4, again). This time, engineers intentionally caused a failure, and then derived convergence time by examining frame loss as reported by Spirent TestCenter. Engineers tested three failure modes:

- **Link failure:** With traffic active, engineers disconnected the link between the east 12500 and east 5820 switches, forcing traffic to be rerouted onto an alternative path

HP IRF Performance Assessment

- **Card failure:** With traffic active, engineers pulled an active line card from the east 12500, forcing traffic to be rerouted
- **System failure:** With traffic active, engineers cut power to the east 12500, forcing traffic to be rerouted through the west 12500

In all cases, Spirent TestCenter offered bidirectional streams of 64-byte frames at exactly 50 percent of line rate throughout the test. This is the most stressful non-overload condition possible; in theory, the system under test is never congested, even during component failure. Thus, any frames dropped during this test were a result of, and only of, path re-computation following a component failure.

Figure 7 below compares convergence times for conventional STP with IRF configured for layer-2 operation. **For all failure modes, IRF converges vastly faster than STP. Further, IRF converges far faster than HP's 50-ms claim in all cases.** In fact, the differences between STP and layer-2 IRF are so large that they cannot be compared on the same scale as with other failover mechanisms.

STP convergence times in this test are, if anything, lower than those typically seen in production networks. In this test, with only two sets of interfaces transitioning between forwarding and blocking states, convergence occurred relatively quickly; in production networks with more ports and switches, STP convergence times typically run on the order of 45 to 60 seconds. IRF convergence times in production may be higher as well, although by a far smaller amount than with STP.

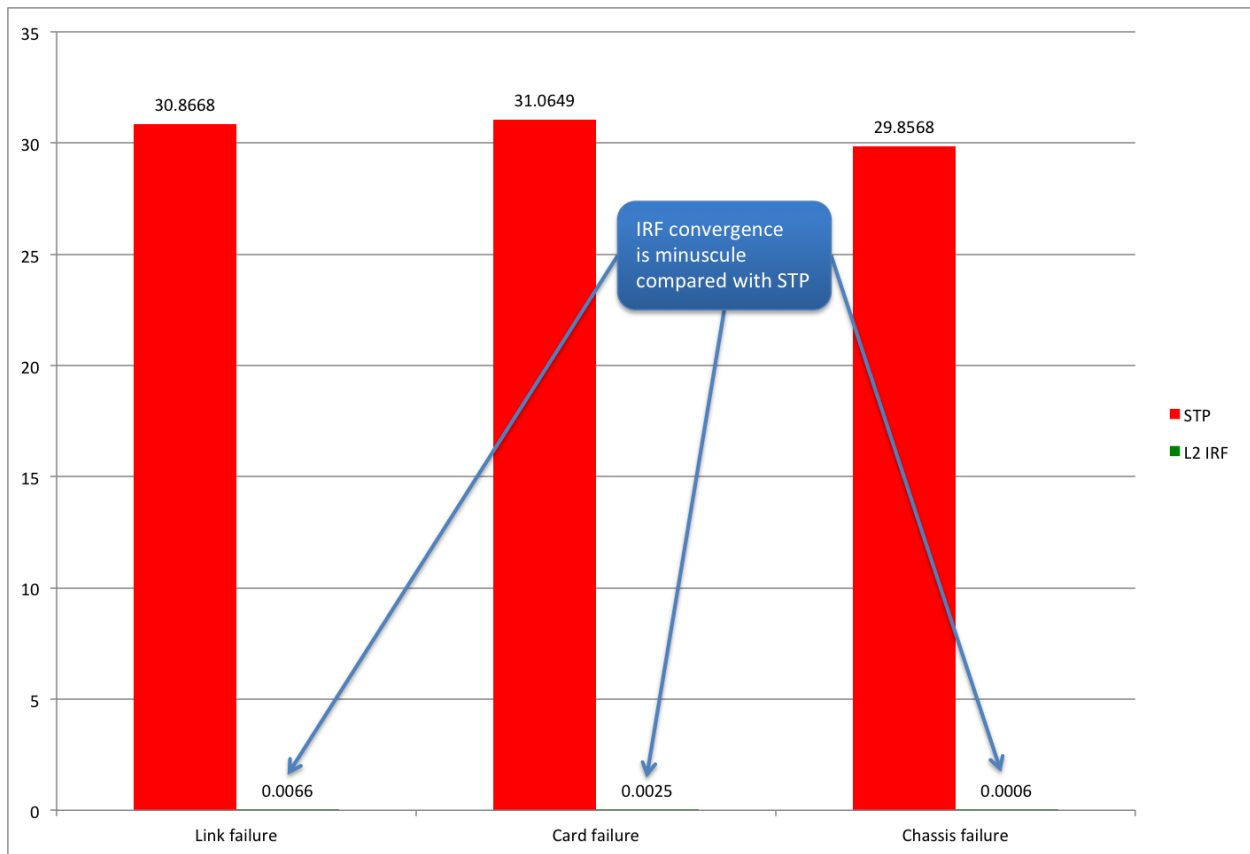


Figure 7: STP vs. IRF convergence times

Network Test also compared IRF with rapid spanning tree, the newer and faster mechanism described in IEEE specification 802.1w. **While RSTP converges much faster than STP, it's still no match for IRF in recovering from network outages** (see Figure 8).

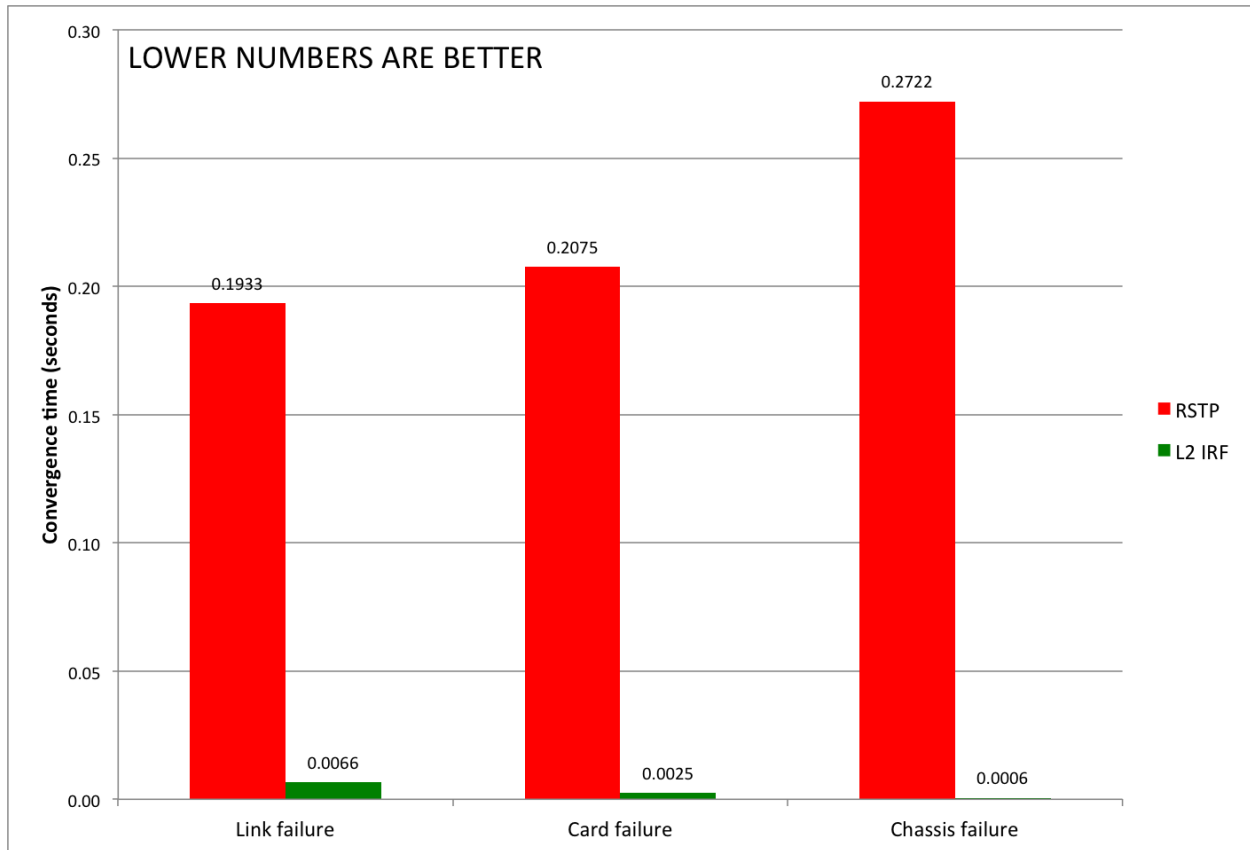


Figure 8: RSTP vs. IRF convergence times

As with STP, the convergence times measured for RSTP were substantially lower than those typically seen on production networks, perhaps due to the small number of links involved. In production settings, RSTP convergence often takes between 1 and 3 seconds following a link or component failure.

The final test compared VRRP and IRF convergence times, with the HP 12500 and HP 5280 both configured in layer-3 mode. In this case, both VRRP and IRF present a single IP address to other devices in the network, and this address migrates to a secondary system when a failure occurs.

Here again, **IRF easily outpaced VRRP when recomputing paths after a network failure** (see Figure 9). VRRP took between 1.9 and 2.2 seconds to converge, compared with times in the single milliseconds or less for IRF.

HP IRF Performance Assessment

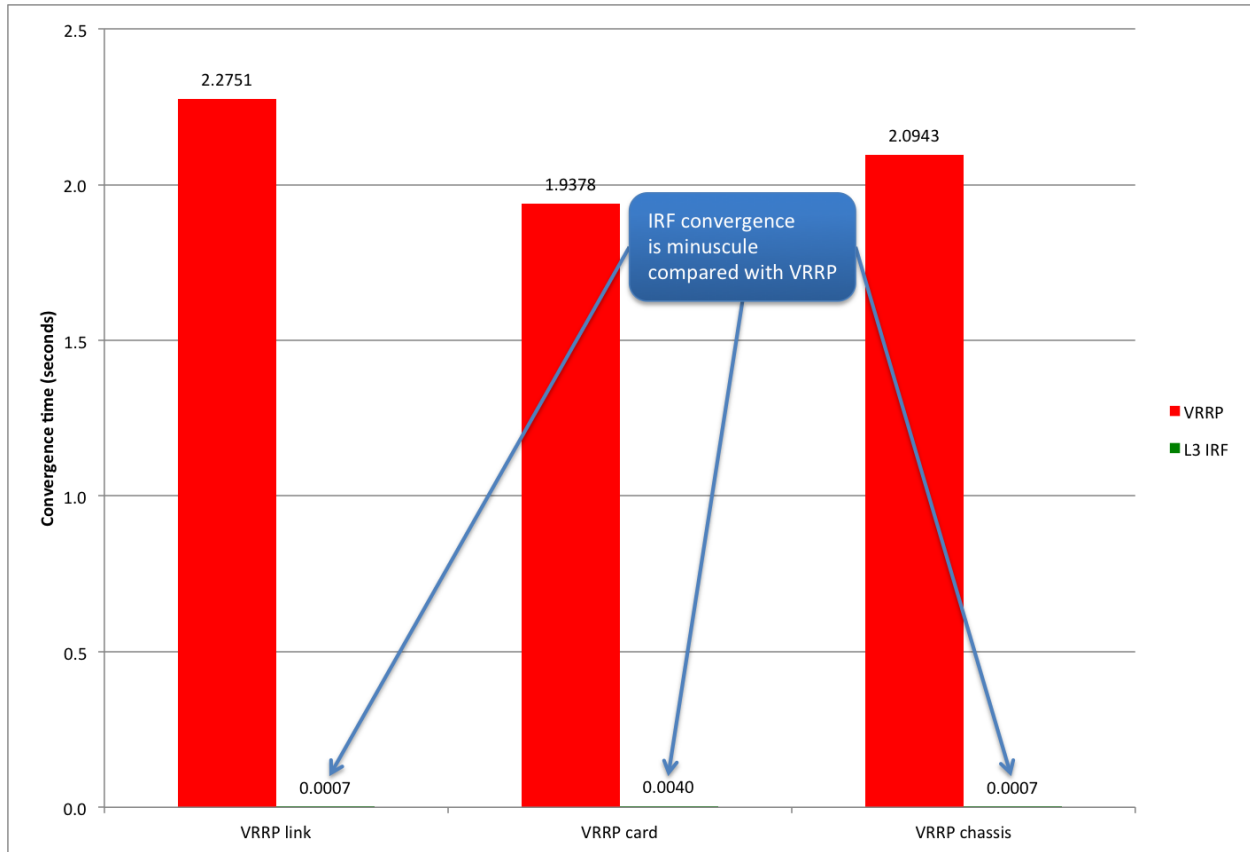


Figure 9: VRRP vs. IRF convergence times

Conclusion

These test results validate IRF’s benefits in the areas of network design, performance, and reliability. IRF simplifies network architectures in campus networks and data centers by combining multiple physical switches and presenting them as a single logical fabric to the rest of the network. This approach results in far faster transfer times for virtual machines using VMware vMotion. Performance testing also shows that IRF nearly doubles available bandwidth by virtue of its “active/active” design, compared with “active/passive” designs that tie up switch ports for redundancy. And the results also show huge improvements in convergence times following network failures, both in layer-2 and layer-3 modes, enhancing reliability and improving application performance.

Appendix A: About Network Test

Network Test is an independent third-party test lab and engineering services consultancy. Our core competencies are performance, security, and conformance assessment of networking equipment and live networks. Our clients include equipment manufacturers, large enterprises, service providers, industry consortia, and trade publications.

Appendix B: Software Releases Tested

This appendix describes the software versions used on the test bed. Testing was conducted in July and August 2011 at HP's facilities in Littleton, MA, and Cupertino, CA, USA.

Component	Version
HP 12500	5.20, Release 1335P03
HP 5820	5.20, Release 1211
VMware vSphere	4.1 Update 1
Spirent TestCenter	3.62.0686.0000

Appendix C: Disclaimer

Network Test Inc. has made every attempt to ensure that all test procedures were conducted with the utmost precision and accuracy, but acknowledges that errors do occur. Network Test Inc. shall not be held liable for damages that may result from the use of information contained in this document. All trademarks mentioned in this document are property of their respective owners.



Version 2011082200. Copyright 2011 Network Test Inc. All rights reserved.

Network Test Inc.

31324 Via Colinas, Suite 113
Westlake Village, CA 91362-6761
USA
+1-818-889-0011
<http://networktest.com>
info@networktest.com