

Solace Message Routers and Cisco Ethernet Switches: Unified Infrastructure for Financial Services Middleware

What You Will Learn

The goal of zero latency in financial services has caused the creation of entirely new industries, products, and career paths as well as volumes of data, analysis, and opinion. The degree to which speed in market-data delivery, trading algorithms, and transactions influences profit cannot be overemphasized and is the motivation behind many financial firms' obsession with low latency. Every aspect of the data path, from messaging platform architecture to networking infrastructure, is being reviewed carefully in an attempt to gain competitive advantage.

Many IT departments have approached the task by disaggregating the elements and testing them in isolation with simple synthetic tests that vary dramatically from the production requirements. However, it is important that mission-critical systems be tested in combination with as many real-world behaviors as possible to understand how factors like network load, variable message size, and spikes in volume affect latency and jitter.

This document introduces the architecture and performance of an end-to-end middleware solution based on Solace message routers and Cisco® networking equipment. This all-hardware solution enables customers to deploy and manage middleware environments that support the ultra-low latency and high message rates they need with much lower cost and complexity than systems that introduce software into the equation.

The tests show consistent low latency, even with bursty traffic, variable message sizes, and high message volumes. This document discusses a comprehensive set of tests, but Table 1 shows an example of the results measured from publisher to subscriber using a Cisco Catalyst® 4900M Switch and a Solace 3260 Message Router.

Table 1. Sample Test Results Using a Cisco Catalyst 4900M Switch and a Solace 3260 Message Router

| Configuration 1a | Message Pattern 1 | Message Pattern 2 |
|-------------------------------------|-------------------|-------------------|
| Average Latency | 31 microseconds | 33 microseconds |
| 99 th Percentile Latency | 47 microseconds | 51 microseconds |
| Standard Deviation | 3.1 microseconds | 3.8 microseconds |

* Message Pattern 1 = 500,000 messages per second (msg/sec) continuous + 500,000 msg/sec bursts

* Message Pattern 2 = 1 million msg/sec continuous + 500,000 msg/sec bursts

Defining the Low-Latency Systems of Tomorrow

Delivering Ultra-Low Latency

A wide array of options is available to financial services companies that are seeking to be first to market with trading ideas or trade execution. Historically, end-to-end market-data latency was often 100 microseconds or more, with most of that latency in the applications and messaging layers, and with the network representing a very small percentage. Consequently, any “above the network” logic that can be optimized in hardware or consolidated to reduce transport latency has the potential to deliver the most significant performance improvements. Solace’s unique approach uses field-programmable gate arrays (FPGAs), application-specific integrated circuits (ASICs), and network processors to perform critical messaging and message processing in hardware and enables ultra-low, consistent latency, even at high volumes.

At the same time, many networking engineers and architects correctly observe that as improvements are achieved for messaging, the percentage of latency attributable to the network increases. With many algorithmic trading strategies, success or loss can be measured in microseconds, causing architects to aim for low end-to-end latency. Just as important is low variability of latency, or jitter. Cisco's Ethernet switches deliver the low latency at both Layers 2 and 3, allowing clients to select the architectures that best meet their business requirements. Moreover, the design of the switch includes custom ASICs that provide deep buffers that allow the switch to accommodate the inevitable bursty traffic encountered in trading and market-data environments. The testing in this document demonstrates how the Cisco Catalyst 4900M Switch and the Cisco Nexus® 5010 Switch reduce retransmissions and thus keep latency low and steady. Choosing network equipment with higher retransmission rates can have catastrophic consequences for application performance and risks noncompliance with regulatory requirements for fairness.

Unfortunately, many teams that build low-latency systems find that components that tested well in isolation using synthetic traffic patterns may deliver less optimal or less predictable results when stressed by periods of market volatility, when they are most needed. Market-data test methodologies are so well understood by the industry that some vendors would appear to have optimized their products for test suites rather than real-world circumstances.

It is essential to evaluate the components both in isolation and when working together as a system to help ensure the best competitive advantage in the end system. In the test results that follow, Solace and Cisco equipment demonstrates market-leading performance using a range of tests that simulate real-world conditions such as mixed message sizes and bursty market-data rates.

Enabling Consolidation

CIOs must find ways to do more with less without sacrificing their ability to quickly scale and shift direction, and the best way to reduce architectural complexity and total cost of ownership (TCO) is by consolidating multiple technologies onto a single platform. For middleware, consolidation has not been possible until now because no one technology could satisfy all middleware needs or provide the performance and scalability needed to support enterprise deployments.

Solace's Unified Messaging Platform consolidates all messaging functions into a single appliance with a common API and shared administration environment. Solace's hardware-based middleware can perform high-fanout, low-latency, guaranteed, Java Message Service (JMS) and WAN messaging, as well as content-based routing, transformation, and message caching. As a hardware-based platform, Solace's solution delivers exceptional performance and resilience with low cost of ownership.

For networking equipment, the expansion of capacity from 100 Mbps to 1 Gbps and now 10 Gbps enables a greater variety and volume of data to flow through a given device or data center. Cisco supports consolidation in this arena with the Cisco Catalyst and Cisco Nexus switches, which have been designed for the data center and deliver low, consistent latency even when carrying many different kinds of traffic. By using the intelligence of the network for context and collaboration, organizations can move with greater speed and agility.

Enabling Intelligent Architecture

One downside of running multiple purpose-specific networks and messaging systems is that financial institutions have had to sacrifice good network design principles to achieve the lowest possible latency. The combination of Solace message routers and Cisco switches enables a clean, scalable network design that consists of Layer 2 and 3 where appropriate while enabling extremely low, predictable latency.

Whether Solace's message routers are given their own data center subnet or share a subnet with other assets such as file servers, each department (equities, fixed income, finance, etc.) can have its own subnet for network connectivity without sacrificing messaging performance. This feature makes it easy to isolate traffic at the IP and messaging layers while gaining the advantages of a common infrastructure: simpler deployment and administration, greater visibility and control, and lower TCO. Such a system supports the connection of local and remote sites, along with the existence of a mixture of 10-Gbps, 1-Gbps, and 100-Mbps host interface speeds. The solution does not require any custom tuning of the network to achieve optimum performance, is highly scalable, and follows good network design principles.

High-Performance Market-Data Delivery Using TCP Unicast

Solace's message routers deliver information using unicast over client-specific TCP connections instead of multicast. Thus, subscriber-specific filtration is performed in the Solace message router as part of the delivery process, so clients are sent only the information they need. Bandwidth is not wasted on the delivery of unwanted messages, clients do not waste CPU cycles filtering and discarding unwanted messages, and interrupt coalescing can be turned off because clients can handle the inbound rate of this custom prefiltered data feed.

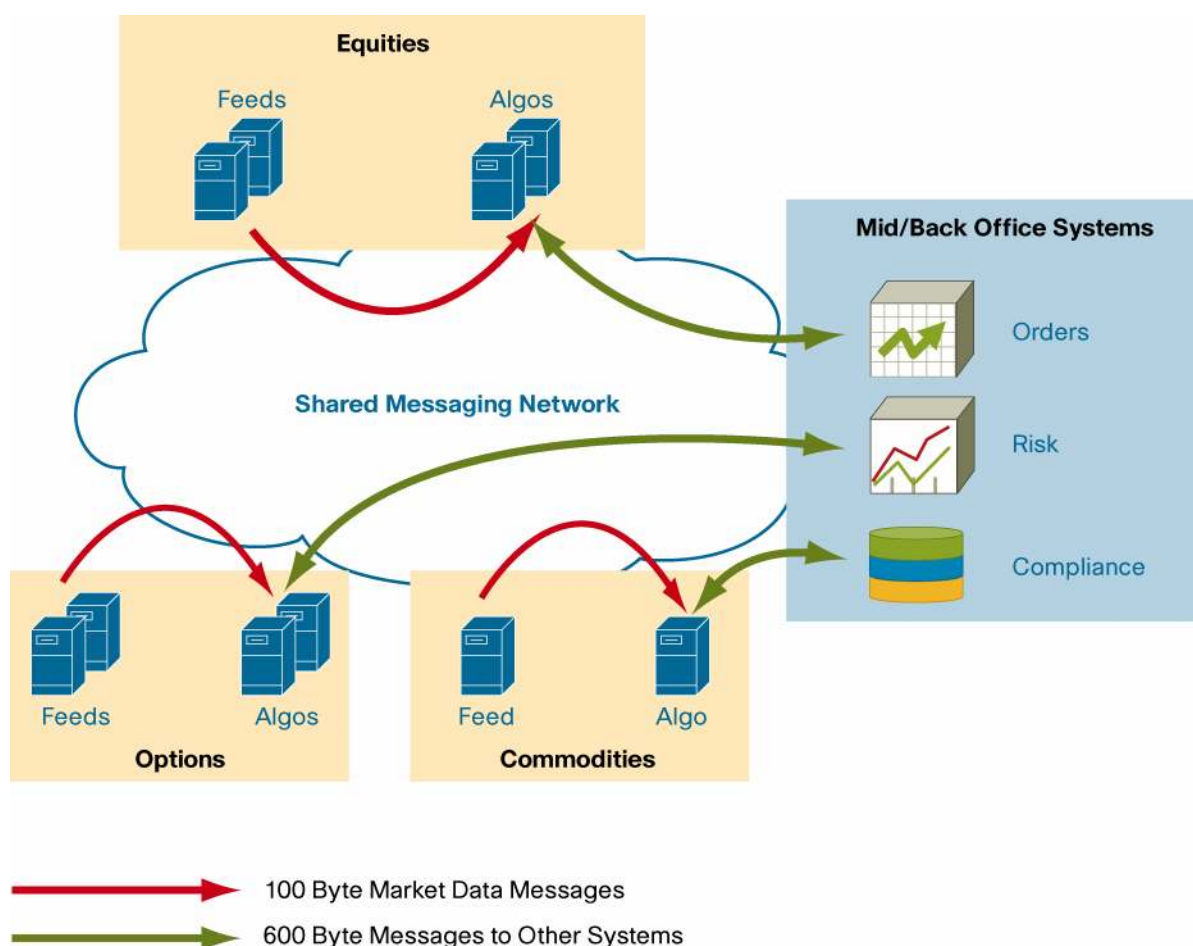
Solace's routing protocols are optimized for this unicast delivery, enabling highly efficient data distribution over both LANs and WANs. The use of TCP-based unicast also eliminates the hassle of administering multicast groups, the risk of high volumes of negative-acknowledgment traffic, and the potentially debilitating ripple effects of slow or misbehaving clients. It also provides granular visibility with client-specific metrics that let administrators address performance problems with speed and certainty. (For more information, please read [this document at Solace's website](#).)

Proving the Power of Cisco and Solace

To demonstrate the advantages of building a front-to-back financial services infrastructure using a combination of Cisco and Solace hardware, Cisco and Solace developed a reference architecture and ran a series of tests designed to represent a real-world scenario that is characterized by high message volume, mixed message sizes, and the kind of variability that characterizes periods of market volatility.

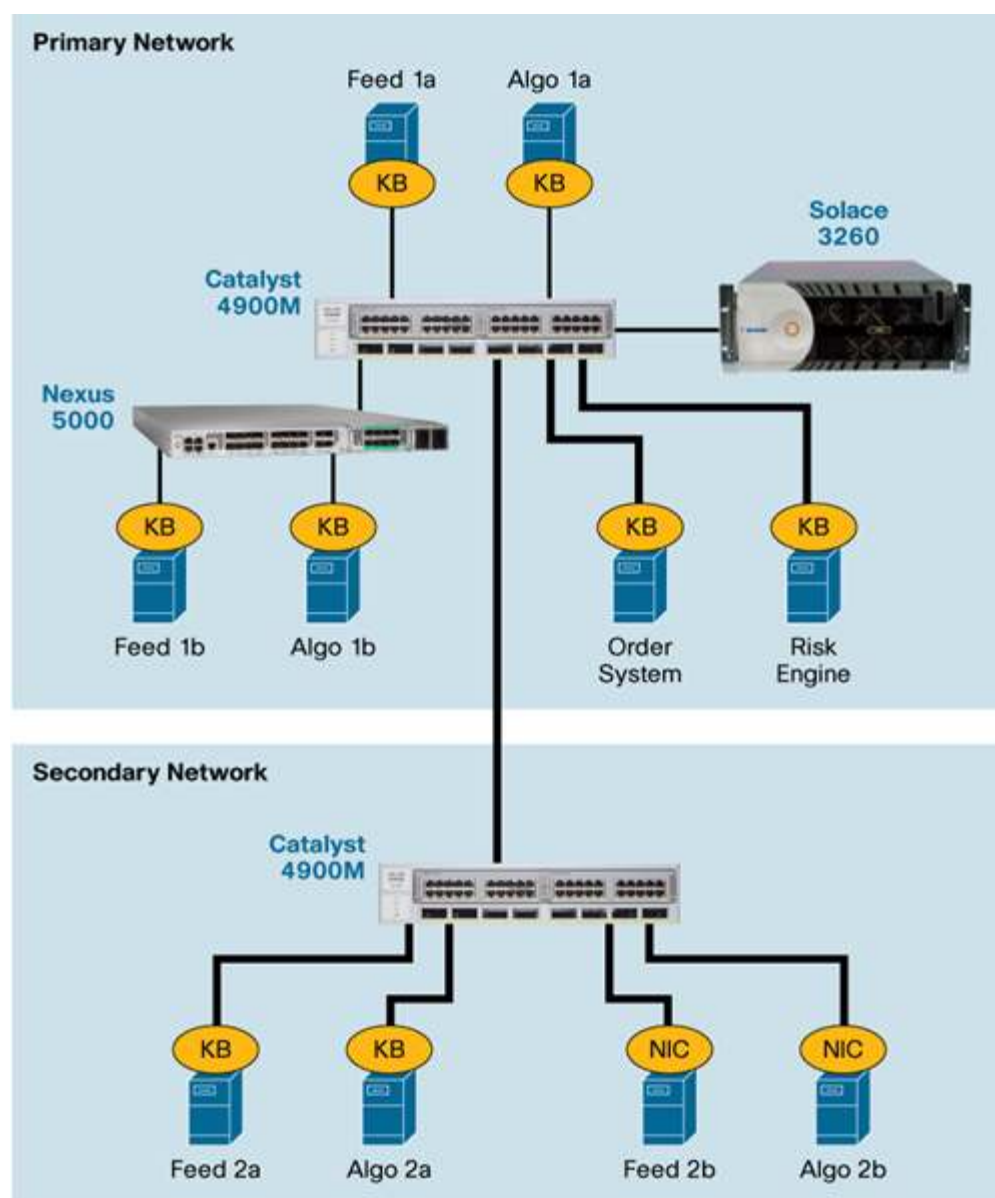
The architecture and test traffic simulates a high-frequency trading environment, with market data coming in from a variety of feeds, being distributed to algorithmic trading engines, and ultimately being sent to several mid- and back-office systems for risk analysis and trade execution.

To approximate real-world message and network traffic, the tests included a high degree of variability in many dimensions. All tests used mixed messages sizes to simulate a single consolidated information bus on a shared network infrastructure. All tests described in this document generated 67 percent market-data volume (100 bytes) and 33 percent mid- and back-office volume (600 bytes). Traffic volume was varied with the addition of 1-second bursts of additional data every other second (Figure 1).

Figure 1. Reference Architecture

System Architecture and Test Scenarios

Figure 2 shows the architecture, components, and physical topology of the test system. The test system uses four configurations spanning a primary network and a secondary network. In these tests, a Cisco Catalyst switch in the primary network has direct connections to the Solace message router, and back-office servers and other departments are on separate subnets. Each of the primary and secondary networks has two separate test configurations (a and b), as shown in the figures and detailed in the results section of this document.

Figure 2. Test System Architecture, Components, and Topology

Shared Messaging Infrastructure

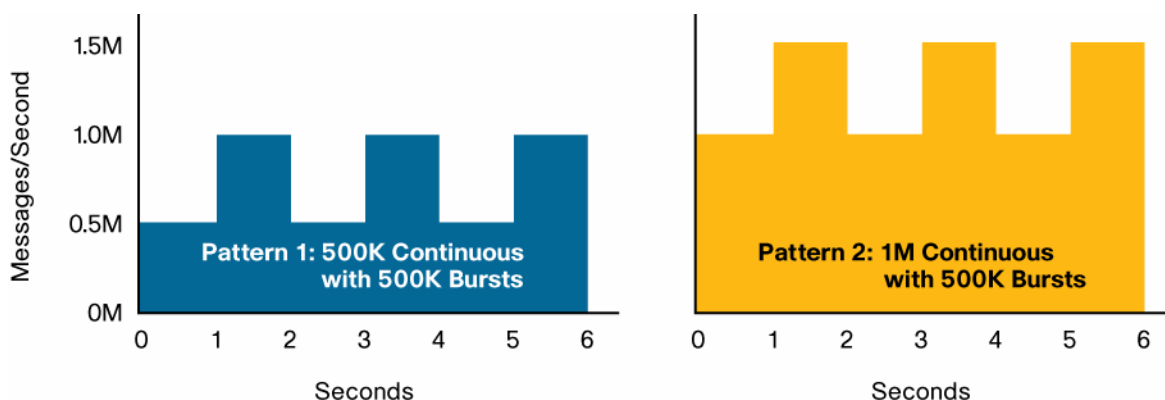
All applications shared a messaging backbone located in the primary network. In scenario 1a, the Solace message router was directly connected to a Cisco Catalyst 4900M low-latency switch. In scenario 1b, the feed and algorithm are connected to a Cisco Nexus 5000 Series Switch, which routes traffic at Layer 2 to the Cisco Catalyst 4900M in the same network. In the secondary network, clients and servers were connected to their own Cisco 4900M switches, which used Layer 3 routing to direct traffic to the primary network and in turn to the Solace message router.

In these tests, a single Solace message router is shared for message routing in both networks. This design is commonly selected because it produces all the performance improvements of hardware messaging while reducing costs by reusing routers across applications and message types. For extremely latency-sensitive applications, customers have the option of placing additional Solace message routers in the secondary network to reduce network hops and increase performance.

Bursty Traffic

In each configuration, two message pattern tests were run that have baseline volumes of 500,000 and 1 million messages per second with bursts of an additional 500,000 messages every other second: that is, in the first test the volume oscillated between 500,000 and 1 million messages per second each second, and in the second test the volume oscillated between 1 million and 1.5 million messages per second each second (Figure 3).

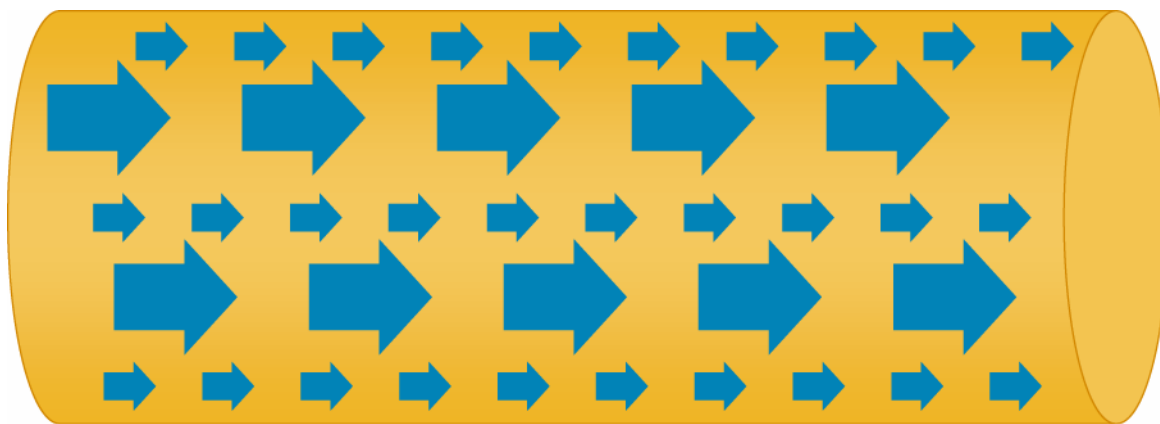
Figure 3. Message Patterns



Mixed Message Sizes

All tests were designed to show how one consolidated infrastructure could meet the needs of front-, middle-, and back-office traffic. For each test, two-thirds of the messages were 100 bytes in size to simulate market-data messages, and the other third were 600 bytes in size to simulate middle- or back-office traffic such as orders, risk management updates, and compliance messages (Figure 4).

Figure 4. Mixed Message Sizes

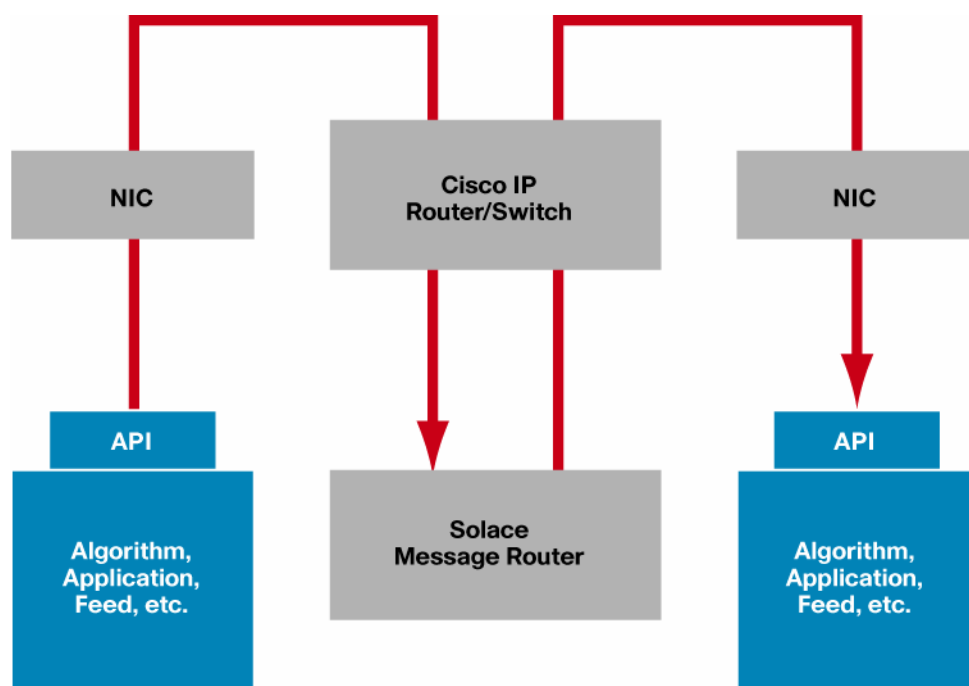


All messages were sent individually using reliable messaging. In reality, some of the middle- and back-office messages would be sent using guaranteed messaging. The same Solace equipment that was used in this test can also route 135,000 guaranteed messages per second without affecting reliable messaging throughput or latency.

API-to-API Measurements

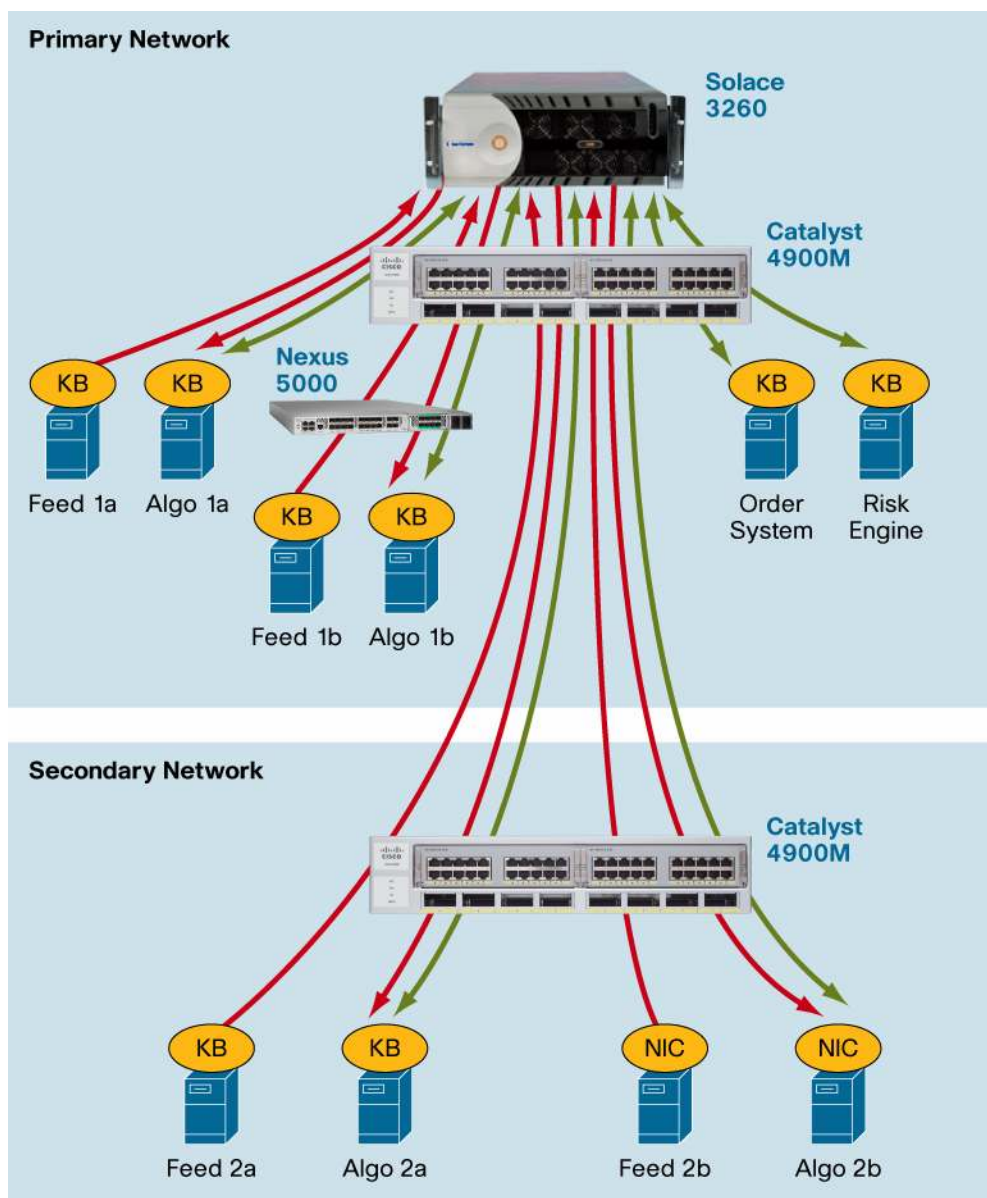
All latency measurement was performed from the messaging API on the feed to the messaging API on the algorithmic engine (Figure 5). This approach provides a more realistic representation of performance than measuring just wire-to-wire latency or time through only the messaging bus. All results include everything you should expect in a production deployment of a Solace messaging backbone and Cisco network.

Figure 5. API-to-API Measurements



Test Environment and Scenarios

Figure 6 shows the logical topology of the network, including a breakdown of subnets. The red arrows signify the flow of market-data messages (100 bytes) between each feed and algorithm pair, and the green arrows show traffic from the algorithms to the back-office applications (600 bytes).

Figure 6. Logical Network Topology

Two traffic volumes were tested, and latency was measured for three configurations. The first test used steady traffic of 500,000 messages per second, with 1-second bursts of an additional 500,000 messages per second every other second. The second test was run with steady traffic of 1 million messages per second, with 1-second bursts of an additional 500,000 messages every other second.

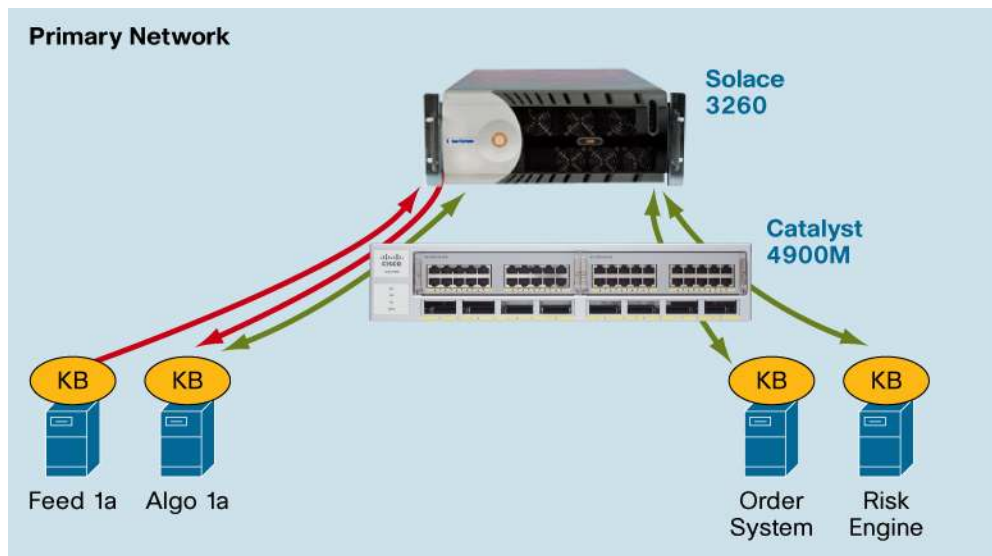
In both tests, two-thirds of the traffic consisted of 100-byte messages to simulate market data, and the other third were 600 bytes to simulate back-office traffic such as orders being sent for execution or risk assessment.

Most client applications used kernel bypass (KB in the figures) technology to accelerate TCP processing on the hosts. Kernel bypass is a technique that can be used in conjunction with either a TCP offload engine or a special-purpose network interface card (NIC) to eliminate operating system delays, resulting in improved TCP performance at high volumes. For comparison, test 2b shows the performance of a standard software TCP stack with a standard NIC and no kernel bypass.

Results of Configuration 1a

In the first configuration (Figure 7), feed 1a, algorithmic engine 1a, and the Solace 3260 Message Router were directly connected to the same Cisco Catalyst 4900M Switch using Layer 2 switching for all traffic. The feed and algorithmic engine servers used TCP kernel bypass technology installed to accelerate network processing at the application hosts. This configuration, with the message router directly connected to the switch, represents the most common approach to achieving the lowest possible latency since the number of network hops is reduced and hardware middleware is keeping the latency of high-volume traffic low and predictable.

Figure 7. Configuration 1a



Note that in message pattern 2, despite a 50 percent increase in the message volume and network load, average latency increased only 2 microseconds, and standard deviation increased only 700 nanoseconds. With 67 percent of the messages being 100 bytes in size and 33 percent being 600 bytes in size, message pattern 2 was peaking above 3.2 Gbps of throughput. This is not close to the limits of the network or messaging equipment, but is a significant load. One million market-data messages plus 500,000 back-office events per second exceeds most trading room capabilities today, and the test results demonstrate much lower latency and standard deviation than would be expected in most currently deployed systems (Table 2).

Table 2. Results of Test 1a

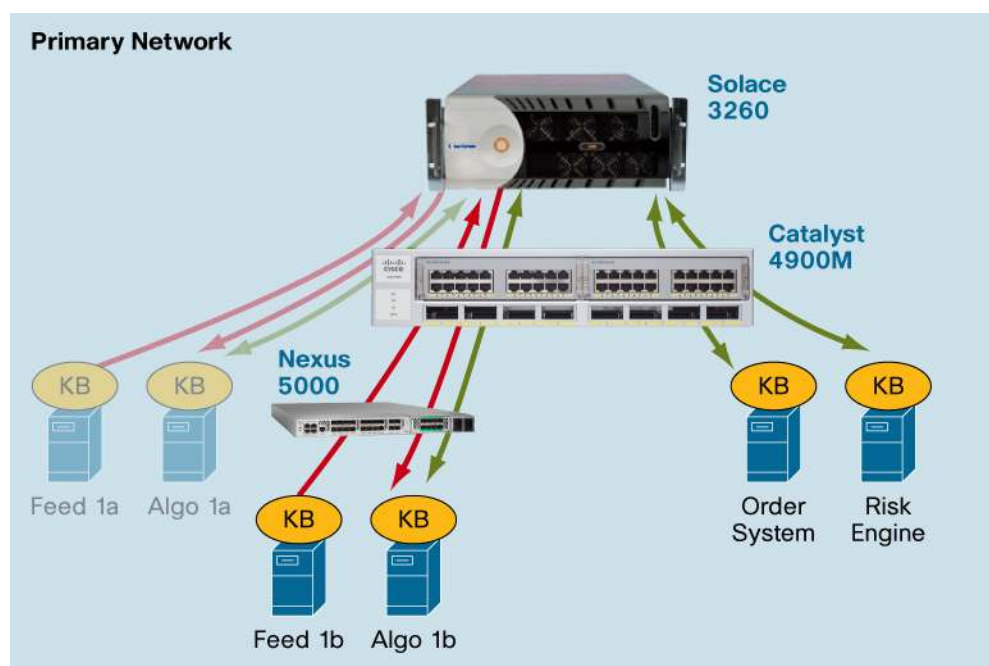
| Configuration 1a | Message Pattern 1 | Message Pattern 2 |
|-------------------------------------|-------------------|-------------------|
| Average Latency | 31 microseconds | 33 microseconds |
| 99 th Percentile Latency | 47 microseconds | 51 microseconds |
| Standard Deviation | 3.1 microseconds | 3.8 microseconds |

* Message Pattern 1 = 500,000 msg/sec continuous + 500,000 msg/sec bursts

* Message Pattern 2 = 1 million msg/sec continuous + 500,000 msg/sec bursts

Results of Configuration 1b: Layer 2 with Cisco Nexus Switch

The second configuration (Figure 8) introduces a Cisco Nexus 5000 Series Switch, providing additional 10 Gigabit Ethernet ports to the primary network. Feed 1b and algorithm 1b are connected to the Solace message router through the Cisco Nexus 5000 Series Switch and the Cisco Catalyst 4900M Switch entirely using Layer 2. Both the feed and algorithm devices are using kernel bypass, as was true in configuration 1a.

Figure 8. Configuration 1b

The results show a modest increase in absolute latency and little effect on jitter despite the introduction of a second switch and the high volumes tested (Table 3).

Table 3. Results of Test 1b

| Configuration 1b | Message Pattern 1 | Message Pattern 2 |
|-------------------------------------|-------------------|-------------------|
| Average Latency | 37 microseconds | 40 microseconds |
| 99 th Percentile Latency | 54 microseconds | 57 microseconds |
| Standard Deviation | 3.4 microseconds | 3.9 microseconds |

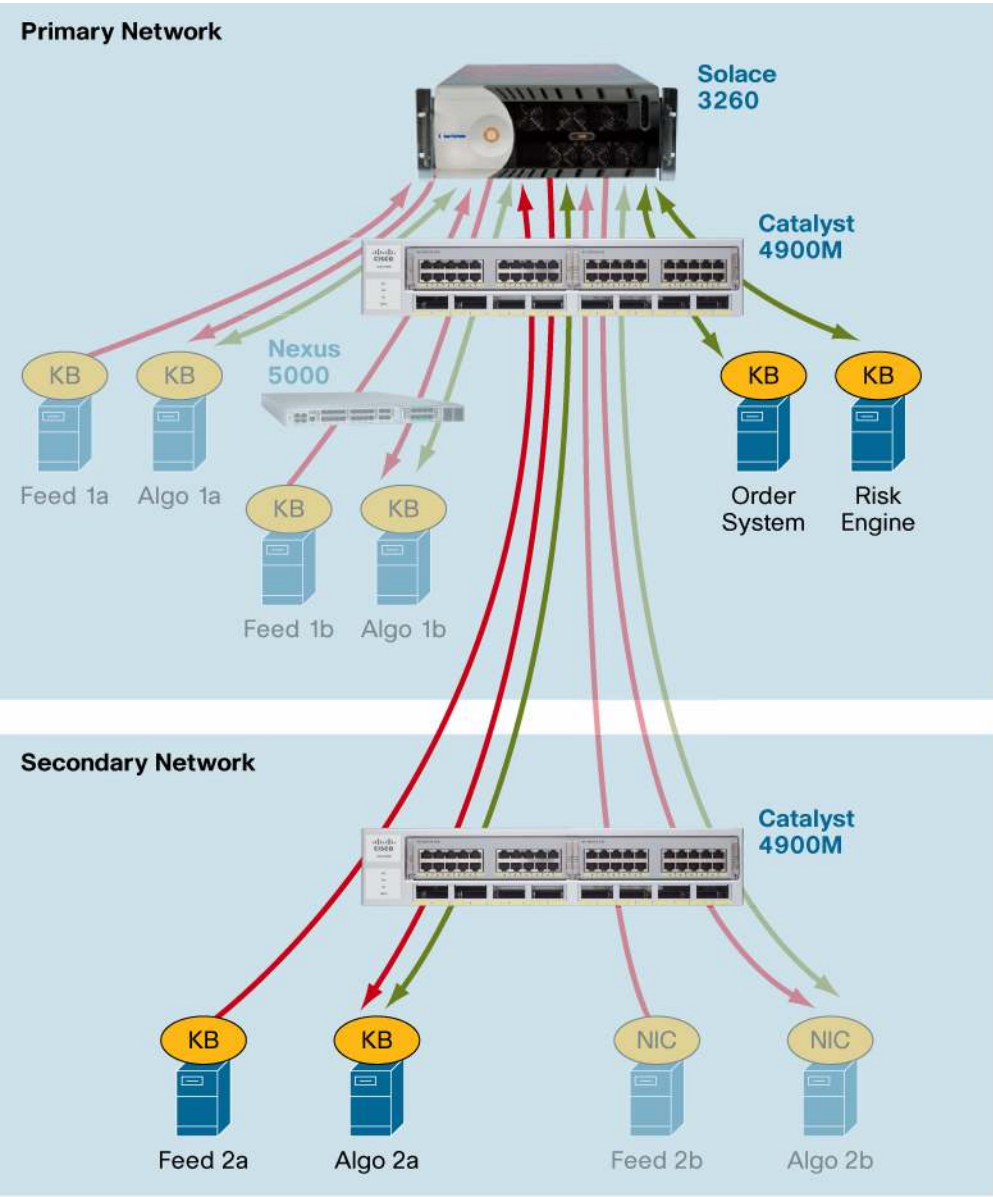
* Message Pattern 1 = 500,000 msg/sec continuous + 500,000 msg/sec bursts

* Message Pattern 2 = 1 million msg/sec continuous + 500,000 msg/sec bursts

Results of Configuration 2a: Layer 3 with Kernel Bypass

The third configuration (Figure 9) is the first of two in the secondary network. Feed 2a and algorithm 2a are connected to the Solace message router through the Cisco Catalyst 4900M Switches in the secondary and primary networks using Layer 3 routing. Both the feed and algorithm devices use kernel bypass, as in the first two configurations. The difference is that the Cisco router in configuration 2a connects using Layer 3, whereas the Cisco switch in configurations 1a and 1b was connected using Layer 2.

Figure 9. Configuration 2a



The results show no significant performance difference between Layer 2 and Layer 3 connectivity in terms of latency or messaging volume (Table 4).

Table 4. Results of Test 2a

| Configuration 2a | Message Pattern 1 | Message Pattern 2 |
|-------------------------------------|-------------------|-------------------|
| Average Latency | 37 microseconds | 40 microseconds |
| 99 th Percentile Latency | 54 microseconds | 57 microseconds |
| Standard Deviation | 3.4 microseconds | 3.9 microseconds |

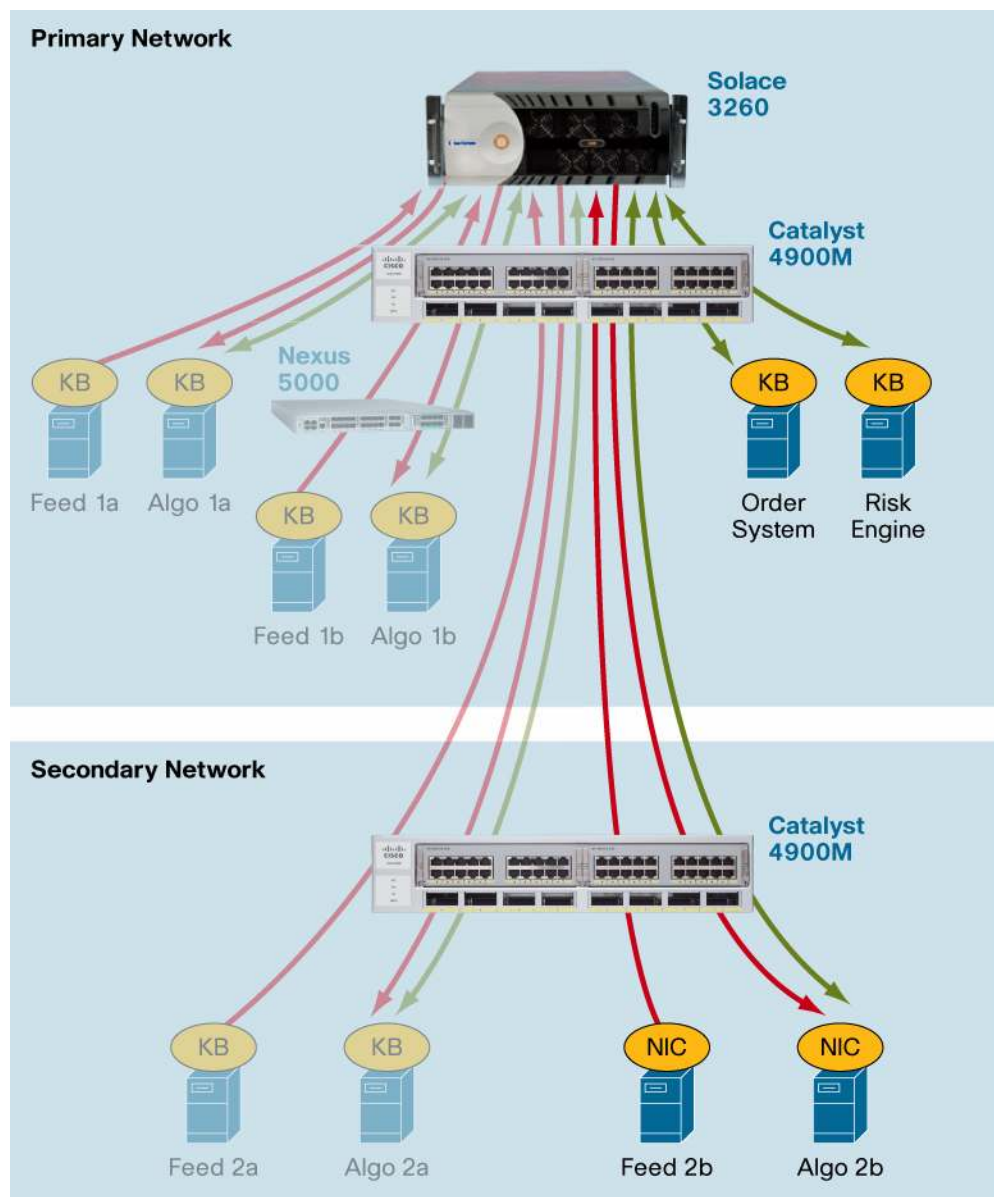
* Message Pattern 1 = 500,000 msg/sec continuous + 500,000 msg/sec bursts

* Message Pattern 2 = 1 million msg/sec continuous + 500,000 msg/sec bursts

Results of Configuration 2b: Layer 3 with NIC

The fourth configuration (Figure 10), using feed and algorithm pair 2b, is identical to configuration 2a except that the feed and algorithmic devices were configured without client-side TCP acceleration. The NIC is still a 10 Gigabit Ethernet interface card, but the TCP stack is running in software with normal OS kernel interactions.

Figure 10. Configuration 2b



The results show that running kernel bypass technology in test 2a reduced latency by 16 to 17 microseconds compared to the results in test 2b (Table 5).

Table 5. Results of Test 2b

| Configuration 2b | Message Pattern 1 | Message Pattern 2 |
|-------------------------------------|-------------------|-------------------|
| Average Latency | 54 microseconds | 56 microseconds |
| 99 th Percentile Latency | 71 microseconds | 71 microseconds |
| Standard Deviation | 6.5 microseconds | 6.2 microseconds |

* Message Pattern 1 = 500,000 msg/sec continuous + 500,000 msg/sec bursts

* Message Pattern 2 = 1 million msg/sec continuous + 500,000 msg/sec bursts

Conclusion

These test results demonstrate that an architecture consisting of hardware Solace message routers and Cisco switches deliver system latency and throughput that exceeds the requirements of the most demanding latency-sensitive financial services environments. This combination delivered consistently low latency and low jitter through both Layers 2 and 3 connectivity, even with bursty traffic, variable message sizes, and high message rates. Further reductions can be achieved using kernel bypass for TCP acceleration on the client application machines. Testing with traffic patterns that mimic real-world environments at times of market volatility helps ensure that the system will function in the production environment with the same performance characteristics that are documented in the evaluation environment.

For More Information

For more information, please contact your Cisco or Solace salesperson to find out how this solution can help you improve the performance of your systems with remarkably low TCO.

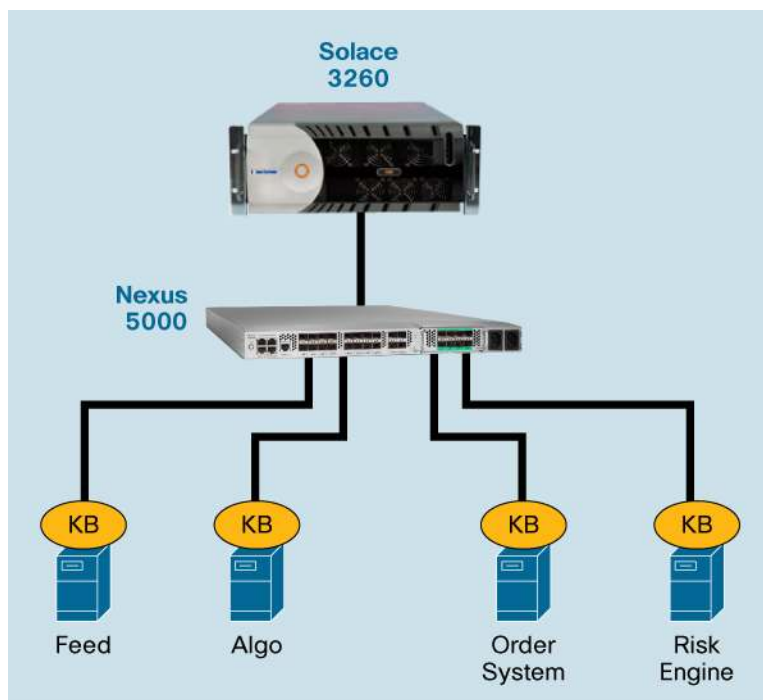
Appendix: Performance of a Direct Cisco Nexus 5000 Series Connection

For completeness, configuration 1a described earlier in this document was separately run with a Cisco Nexus 5000 Series Switch in place of the Cisco Catalyst 4900M as the device directly connected to all other test components.

System Architecture and Test Scenario

Figure 11 shows the logical topology of the network. The feed, algorithmic engine, back-office applications, and Solace 3260 Message Router were directly connected to the Cisco Nexus 5000 Series router using Layer 2 switching for all traffic. The application servers used TCP kernel bypass technology to accelerate TCP processing at the application hosts.

Figure 11. Configuration Using Cisco Nexus 5000 Series



The system was tested with the same parameters as those described earlier in this document. Two traffic volumes were tested: steady traffic of 500,000 messages per second, with 1-second bursts of an additional 500,000 messages per second every other second, and steady traffic of 1 million messages per second, with bursts of an additional 500,000 messages. In both tests, two-thirds of the traffic consisted of 100-byte messages to simulate market data, and the other third were 600 bytes to simulate back-office traffic such as orders being sent for execution or risk assessment. All client applications used kernel bypass technology to accelerate TCP processing. Kernel bypass is a technique that can be used in conjunction with either a TCP offload engine or a special-purpose NIC to eliminate OS interrupts, resulting in improved TCP performance at high volume.

Results

The system exhibited remarkably low latency, averaging 32 microseconds in message pattern 1 and 34 microseconds in message pattern 2. Note that despite a 50 percent increase in throughput, average latency increased just 2 microseconds, and standard deviation increased a mere 700 nanoseconds. With 67 percent of the messages being 100 bytes in size, and 33 percent being 600 bytes in size, test 2 was peaking above 3.2 Gbps throughput, which is not near the limits of the network or messaging equipment, but a significant load. One million

market-data messages plus 500,000 risk, trade, and other mid-back office events per second exceeds most trading room capabilities today, with much lower latency and standard deviation (Table 6).

Table 6. Results of Test Using Cisco Nexus 5000 Series

| Configuration Using Cisco Nexus 5000 Series | Message Pattern 1 | Message Pattern 2 |
|---------------------------------------------|-------------------|-------------------|
| Average Latency | 32 microseconds | 34 microseconds |
| 99 th Percentile Latency | 49 microseconds | 52 microseconds |
| Standard Deviation | 3.4 microseconds | 3.8 microseconds |

* Message Pattern 1 = 500,000 msg/sec continuous + 500,000 msg/sec bursts

* Message Pattern 2 = 1 million msg/sec continuous + 500,000 msg/sec bursts

Conclusion

These results show that an architecture consisting of Solace message routers and Cisco Nexus switches delivers latency and throughput that exceed the requirements of the most latency-sensitive financial services environments. This combination delivers consistently low latency and low jitter through, even with bursty traffic, variable message sizes, and high message rates.



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte Ltd
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

CCDE, CCENT, CCS, Cisco Eos, Cisco Unified Presence, Cisco IronPort, the Cisco logo, Cisco Nexus Connect, Cisco Prime, Cisco Security, Cisco ShieldPower, Cisco StadiumVision, Cisco TelePresence, Cisco Unified Computing System, Cisco WebEx, DCE, EPC Channels, EPC for Core, EPC Mini, EPCare (Design), EPC Ultra, EPC Video, EPC Video (Design), EPCare Broadband, and We come to the Human Network are trademarks. Changing the Way We Work, Live, Play and Learn, Cisco Capital, Cisco Capital (Design), Cisco Financial (Style), Cisco Store, EPC Card, and One Million Acts of Green are service marks, and Access Registered. Aironet, Aironet, AsyncOS, Bringing the Meeting to You, Catalyst, CCDA, CCDE, CCDE, CCDE, CCNA, CCNE, CCSE, CCSE, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Link, Cisco Nexus, Cisco Prime, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, Continuum, EtherFast, EtherSwitch, Event Center, Explorer, Flow Me Browsing, Galileo, IYX, IOS, iPhone, IronPort, the IronPort logo, iLearn Link, LightStream, Linksys, MeetingPlace, MeetingPlace Online Sound, MGX, Networkers, Networking Academy, PCNow, PDX, PowerKEY, PowerPanel, PowerTV, PowerTV (Design), PowerVu, Priams, ProConnect, ROSA, SenderBase, SMARTnet, Spectrum Expert, StackWise, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (09100)