NetQoS
Performance First

WHITE PAPER

Time Is Money: Managing the Performance of Financial Trading Applications

By Ben Erwin, NetQoS, Inc.

With trillions of dollars traded annually on the NASDAQ alone, financial services companies are investing heavily in optimizing their electronic trading infrastructure and providing nearly instantaneous access to the latest market data. The widespread adoption of electronic trading has spurred many global IT initiatives to reduce application latency to extremely low levels. Put simply, a one-millisecond advantage can be worth millions. Acquiring and sustaining this advantage has created a fiercely competitive landscape, where financial services companies must keep a watchful eye on their trading infrastructure. Alongside competitive advantage, regulatory initiatives are also forcing trading services to provide market access in a timely manner. Now more than ever, financial services companies need the right solutions to manage the delivery of their trading applications.

The financial services industry has standardized on a communication protocol for financial trading applications called Financial Information eXchange (FIX). For the trade desk, managing application delivery requires an understanding of the FIX protocol; however, many financial services companies are not equipped with the right solutions to monitor and troubleshoot FIX applications. Solutions that provide visibility into network traffic, hop by hop latency, data center application performance, and retrospective analysis of FIX traffic can provide the needed insight. Without these solutions, financial services companies are left in the dark when they need to validate low-latency trade execution.

In this white paper, we define the issues and outline the capabilities required to effectively manage the delivery of electronic trading applications and order management systems.

Introduction

Trading activity between exchanges, brokerage firms, hedge funds, and other financial services companies is a core component of the world economy. With trillions of dollars traded annually on the NASDAQ alone, financial services companies are investing heavily in optimizing electronic trading applications and infrastructure, with the aim of providing nearly instantaneous access to the markets. Put simply, when buying or selling electronically, a one-millisecond performance advantage can be worth millions. The need to acquire and sustain this advantage has created a fiercely competitive landscape, where financial services companies must frequently upgrade and monitor their order management system (OMS) and IT infrastructure. At the same time, regulatory initiatives like the Regulation National Market System (Reg NMS) in the US and the European Union's Markets in Financial Instruments Direction (MiFID) are also forcing trading services to handle vastly increased volumes of market data and still provide order execution to all clients in a timely manner.

Issues surrounding government compliance and the need to maintain a competitive advantage are driving more trading firms to favor automated (or "algorithmic") trading applications over phone calls and human decision-makers. This trend is evident on the New York Stock Exchange, where half the trades in 2007 were managed by software applications. The industry has even standardized on a communication protocol for financial trading applications called Financial Information eXchange (FIX).

The widespread adoption of electronic trading has spurred many IT initiatives to reduce application latency to extremely low levels. The London Stock Exchange recently overhauled its trading platform to reduce latency from 110 milliseconds to 10 milliseconds. A US-based market exchange moved its data center across the country to shave latency in its order management system by 30 milliseconds. Hosting providers such as BT-Radianz and Savvis are getting into the race, providing buyers with direct market access (DMA) into data centers co-located with exchanges. DMA has grown in popularity over the last few years as more financial services companies show their willingness to invest heavily in low-latency architectures.

Now more than ever, financial services companies need the right solutions to manage the delivery of their financial trading applications. For the trade desk, managing the delivery of applications requires an understanding of the FIX protocol; however, many financial services companies are not equipped with the right solutions to monitor and troubleshoot FIX applications. Solutions that provide visibility into network traffic, hop-by-hop latency, data center application performance, and retrospective analysis of FIX traffic can provide insight into application performance. Without these solutions, financial services companies are left in the dark when they urgently need to validate low-latency order execution.

In this white paper, we define the issues involved and outline the capabilities required to effectively manage the performance of financial trading applications. Latency is a critical metric, but a broader understanding is needed. With any networked application, a number of factors can affect performance, and in an order management system, poor performance can easily impact the bottom line. In this paper, we discuss managing latency and performance in the fast-paced, constantly evolving sector of the financial services industry devoted to electronic trading. In this space, the ability to respond to the changes, keep ahead of competitors, and adhere to the rules is critical, and it requires the right tools. We'll begin by contextualizing our discussion, summarizing the key players in the industry and describing how they communicate.

Buy Sides and Sell Sides

The financial services industry comprises two key groups: firms on the "buy side" and those on the "sell side." Buy sides are people or institutions that use market services; sell-side institutions provide market services. Buy sides typically connect into multiple sell sides, expecting lightning-fast trade execution and the most up-to-date market data. Examples of buy-side firms include hedge funds, proprietary trading shops ("prop shops"), pension funds, and mutual funds. Examples of sell sides include brokerage firms, electronic communications networks (ECNs), and exchanges. You're probably familiar with buy-side institutions, but those on the sell side may be more obscure. An exchange (such as the NYSE or NASDAQ) is a marketplace for traders to buy and sell securities. Some buy sides will pay for direct market access, but, with the exception of large hedge funds, many buy side firms cannot afford it. The high cost of direct access provides opportunities for ECNs, brokerage firms, and hosting providers, who leverage economies of scale to provide affordable DMA into exchanges. An ECN is a trading network such as Bloomberg's TradeBook or Instinet that matches buy and sell orders at specified prices. The example in Figure 1 shows communications between buy sides and sell sides.

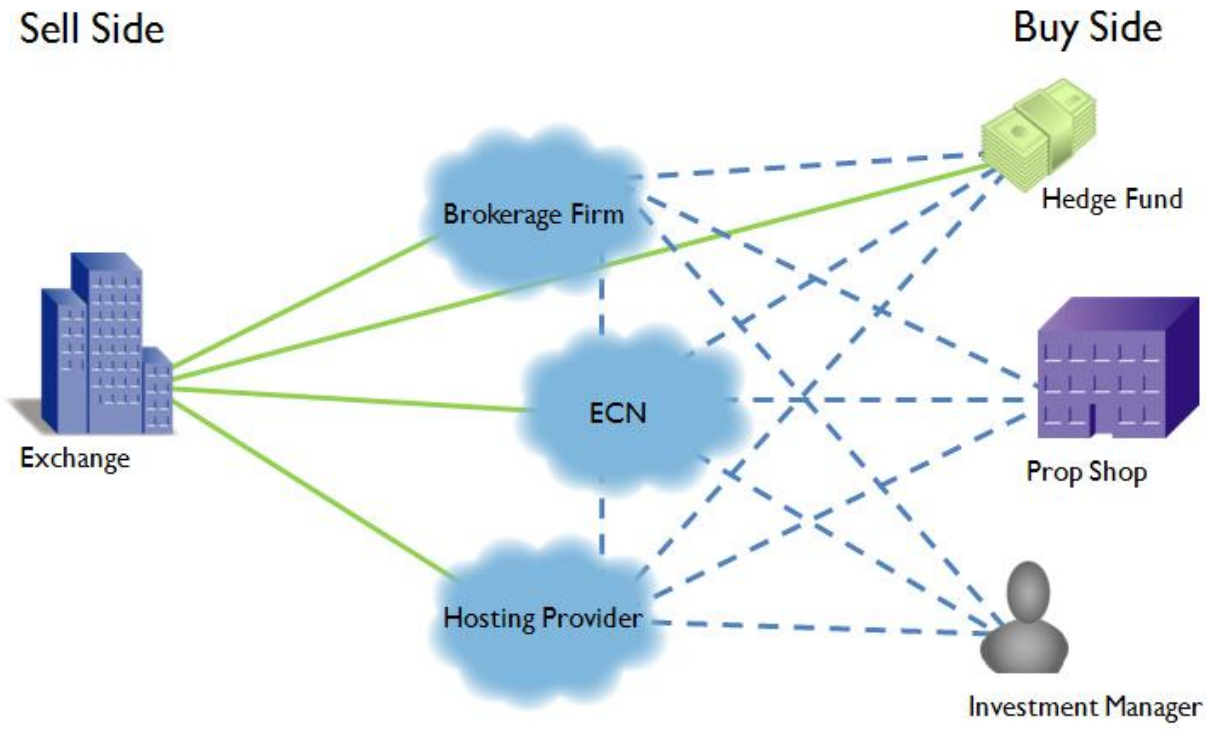


Figure 1 – Buy- and sell-side connectivity across a trading network

Given the financial implications of timely trade execution, buy sides will select the sell-side providers with the fastest market access and/or guaranteed trade execution times. Therefore, sell sides (and buy sides paying for DMA) must keep a very close eye on the latency of their application services to ensure consistent service delivery and competitive advantage. The application services responsible for order execution leverage the FIX protocol; therefore, knowledge of FIX is important for understanding how quickly trades are being executed and where problems exist.

The FIX Protocol

The FIX protocol first surfaced in 1992 as an electronic communications method to be used between the financial services companies Fidelity Investments and Salomon Brothers. FIX has undergone several revisions since its introduction, the latest being FIX v5.0, released at the end of 2006, and lately it has also seen explosive adoption rates along with the rise of algorithmic trading, investments in low-latency architectures, and the advent of regulatory compliance initiatives like RegNMS and MiFID. The FIX standard has several benefits, including real-time trade execution and vendor neutrality. FIX applications can also scale to high volumes of trade activity while achieving less than 10-millisecond delays. As a communication protocol, FIX is free, open, and widely used by software engineers who develop financial trading applications. It can be thought of as a common electronic communication platform between financial services companies all over the world.

The FIX protocol standard is simple but verbose. Each FIX message contains information on the sending and receiving computer, ticker symbol, order type, number of shares, order time, and other relevant information to the trade. Each piece of information in a FIX message is represented as a series of tag-value pairs, as shown in Figure 2 below.

```
8=FIX.4.2|9=143|35=D|34=76|49=SGLM|56=SBI|
34=92|50=George|57=Joe|52=20000926 04:39:59
|11=1|21=2|55=AMZN|54=1|38=5000|40=1|59=1|
10=044
```

Figure 2 - Sample FIX message

The challenges for financial institutions when managing the performance of FIX applications include minimizing server response time, responding to FIX protocol errors, and ensuring that trades are going through as requested. FIX applications must very rapidly interpret a high and variable volumes of FIX messages, receive and process market data updates, and summarize thousands of these messages every minute, turning them into actionable information. IT organizations that manage trading networks and FIX applications need summarized performance reports, latency details, actionable alerts, and retrospective analysis capabilities to properly manage financial trading applications.

However, monitoring FIX applications alone is not enough to manage the delivery of trading services. These applications are primarily responsible for trade execution, but monitoring other communications, like the critically important market data feeds that provide updated pricing and news that affects the markets, is also a necessity. In the next section of this paper, we describe all of the relevant IT components required for a complete view into trading application delivery.

IT Technologies for Managing Trading Application Delivery

Managing the performance of financial trading applications requires visibility on multiple levels. Trading networks are a complex mesh of front- and back-office end points, and measuring latency between these end points can present a challenge. In addition, financial services companies may house thousands of servers in multiple data centers that are responsible for processing trades and sending instantaneous responses. They require highly scalable monitoring solutions to handle the trade traffic loads at these server clusters. Finally, with thousands of FIX messages traversing the network every minute, monitoring trade execution can be an overwhelming task. Without deploying additional network infrastructure all over the globe to collect and analyze the data, how can financial services companies measure the delivery of their trading applications?

The answer is a multi-pronged strategy that includes source-to-destination monitoring, data center performance monitoring, and FIX forensic analysis.

Source-to-Destination Monitoring

Trading applications responsible for supplying the latest market updates communicate via one-way multicast blasts to multiple customers who are requesting the latest market prices. Determining hop-by-hop latency across a complex network mesh can be an expensive and time-consuming process without the appropriate monitoring capabilities. Cisco's IP SLA (Internet Protocol Service Level Agreement) technology embedded in Cisco routers is a quick method for determining hop-by-hop, one-way latency. IP SLA is an integrated component of Cisco IOS that can be leveraged in almost every Cisco routing environment to execute synthetic application transactions between source and destination end points. Once a transaction has completed, IP SLA measures the latency of the transactions and makes the data available to third-party reporting solutions. An IP SLA-compatible reporting solution can extract these measurements from Cisco routers to provide actionable alerts and detailed analysis. See Figure 3 below for an example of the types of metrics you can gather using IP SLA testing.

Figure 3 - Latency measurements from Cisco IP SLA

While latency is the most critical measurement for monitoring FIX application delivery, visibility into all of the traffic flows across a trading network is also necessary. Regulatory initiatives like RegNMS and MiFID have increased mandatory communications between financial services companies, resulting in very high message volume across network links. And trading networks are not always dedicated to transporting market data; they sometimes serve as a path for other applications critical to a financial services enterprise. Traffic from business-critical applications, such as VoIP and email, may be found alongside trading applications on some network links and should be closely monitored as well. With a finite amount of network bandwidth, keeping an eye on how these

applications consume network resources and comparing their consumption with that of trading applications is important for maintaining the highest performance levels. Like Cisco IP SLA, Cisco NetFlow can be leveraged on a Cisco network infrastructure to provide details of traffic flows and traffic composition. NetFlow can export detailed records of every application transaction across the network to a third-party NetFlow collector. Paired with a scalable collection infrastructure, Cisco NetFlow is an effective, straightforward way to achieve visibility into all of the applications that are sharing the trading network.

Data Center Performance Monitoring

A data center can create a latency bottleneck in a trading network. Financial services data centers are filled with complex, multi-tier server environments responsible for responding to order requests in sub-millisecond timeframes. Because these servers are critical to the delivery of trading services, installing third-party applications on them to monitor their performance is not advisable due to overhead, load, and security concerns. Rather, a less intrusive, passive approach to monitoring latency to and from servers via the switching infrastructure is preferred. Figure 4 below provides a schematic architecture of such a nonintrusive approach, which monitors server traffic on a shared switch. In addition, to manage the overall delivery of trading applications proactively, performance monitoring solutions must be able to measure the latency of application flows and provide actionable alerts of performance degradations. These tools should also provide analyses that aid capacity planning and enable overall application performance to be assessed from the data center.

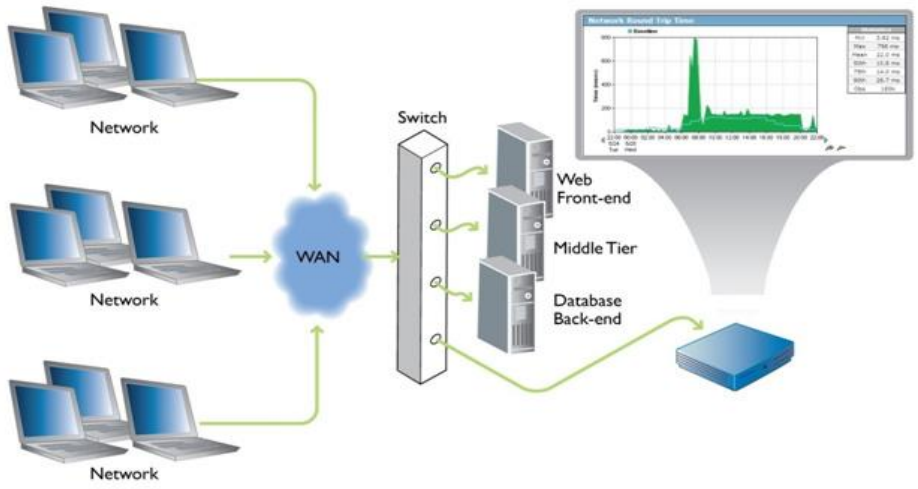
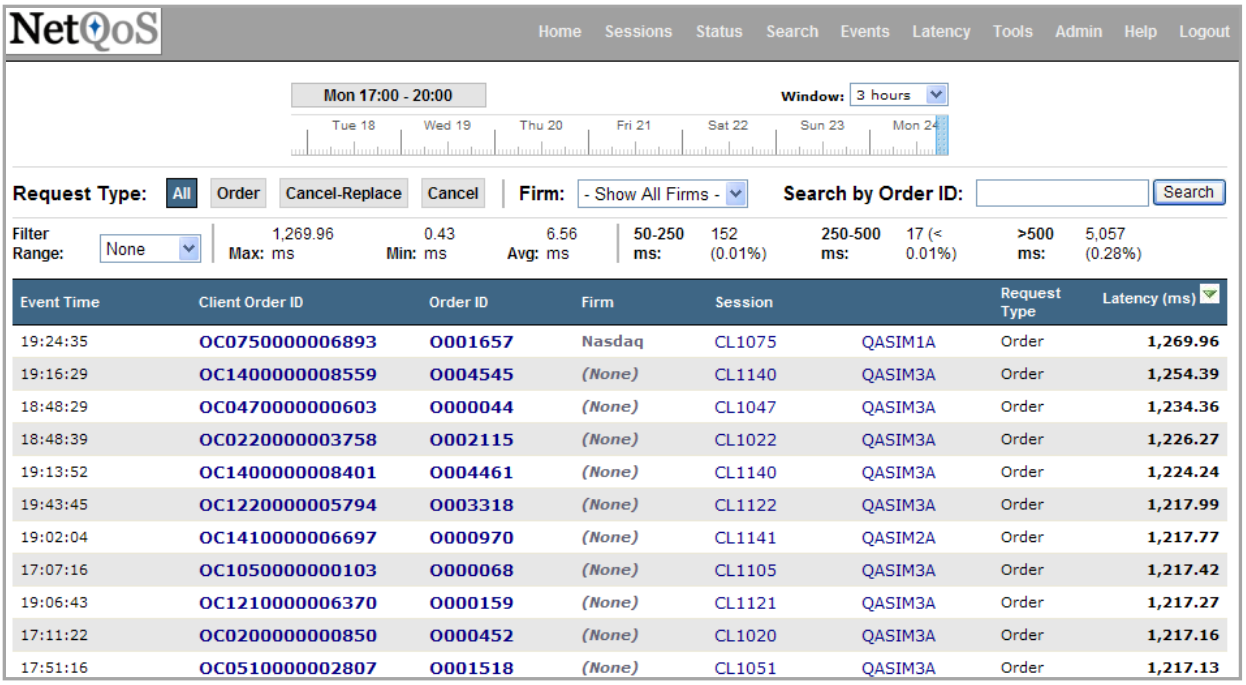


Figure 4 - Passive monitoring of transactions between data center servers

FIX Forensics

As previously discussed, DMA connections are exploding in popularity, as buy sides continue to invest heavily in the lowest-latency connections into exchanges. Each of these DMA connections transports FIX application traffic between buy sides and sell sides to execute trades. Financial services companies need FIX-based monitoring solutions that can passively monitor FIX transactions and record them for forensic analysis. A FIX monitoring solution should also be intelligent enough to match FIX requests and responses by presenting the data in a conversation format between clients and servers. This level of visibility allows both Trade Desk and IT personnel to immediately diagnose performance outliers and continue supporting millisecond trade execution times [Figure 5]. Actionable alerts and notifications are extremely valuable for the FIX trade desk to help minimize mean time to repair.



Event Time	Client Order ID	Order ID	Firm	Session		Request Type	Latency (ms)
19:24:35	OC0750000006893	O001657	Nasdaq	CL1075	QASIM1A	Order	1,269.96
19:16:29	OC1400000008559	O004545	(None)	CL1140	QASIM3A	Order	1,254.39
18:48:29	OC0470000000603	O000044	(None)	CL1047	QASIM3A	Order	1,234.36
18:48:39	OC0220000003758	O002115	(None)	CL1022	QASIM3A	Order	1,226.27
19:13:52	OC1400000008401	O004461	(None)	CL1140	QASIM3A	Order	1,224.24
19:43:45	OC1220000005794	O003318	(None)	CL1122	QASIM3A	Order	1,217.99
19:02:04	OC1410000006697	O000970	(None)	CL1141	QASIM2A	Order	1,217.77
17:07:16	OC1050000000103	O000068	(None)	CL1105	QASIM3A	Order	1,217.42
19:06:43	OC1210000006370	O000159	(None)	CL1121	QASIM3A	Order	1,217.27
17:11:22	OC0200000000850	O000452	(None)	CL1020	QASIM3A	Order	1,217.16
17:51:16	OC0510000002807	O001518	(None)	CL1051	QASIM3A	Order	1,217.13

Figure 5 - FIX capture of performance outliers

Conclusion

Maintaining visibility and measuring latency throughout the trading network are must-haves for ensuring consistent service delivery for end users. Managing the delivery of trading applications can be a difficult process without a comprehensive monitoring solution in place. Latency bottlenecks are extremely costly and can occur between any source and destination in the global client network, or within the data center infrastructure itself. The technologies currently available for gaining visibility into end-to-end performance and measuring FIX server latency include:

- Cisco IP SLA and NetFlow, for hop-by-hop source and destination monitoring across multicast networks. IP SLA provides latency metrics between end points, while NetFlow provides visibility into all application flows.

- Passive data center performance monitoring of the trading server infrastructure. Without the hassle and risks of deploying software probes on servers, passive monitoring solutions can provide actionable information on application performance between servers in the data center and out to clients.

- FIX forensics monitoring of trading applications. This type of monitoring provides application-level, detailed information about every FIX session between buy sides and sell sides. A FIX monitoring solution should passively monitor trading applications for FIX server response times and record every transaction for any necessary trade verification or order troubleshooting.

Using the best-fit solution is a critical factor in maintaining a competitive edge in a cutthroat environment and retaining the trust of electronic trading clients. In addition to ensuring execution quality and minimizing latency, the right toolset will provide valuable assistance in infrastructure capacity planning and network management.

About the NetQoS Performance Center

The NetQoS Performance Center is a single web-based portal that delivers global visibility into the entire network infrastructure for the insight to resolve performance issues, troubleshoot infrastructure problems, perform capacity planning, plan for and validate the effects of change, and track Service Level Agreements (SLAs). The NetQoS Performance Center provides an integrated view of critical performance data delivered by NetQoS' best in class products—SuperAgent® for end-to-end performance data, ReporterAnalyzer™ for traffic analysis, and NetVoyant® for device performance (SNMP Polling) data. With this visibility, you will make more informed decisions based on precise network infrastructure usage data, improve staff efficiency, and resolve performance issues rapidly. You can also mitigate risks and validate the impact of planned changes such as VoIP deployments, MPLS migrations, WAN optimization, QoS policy implementation, and application rollouts. The NetQoS Performance Center is fast and easy to deploy, and scales to the largest networks.

About NetQoS

NetQoS is the fastest growing network performance management products and services provider. NetQoS has enabled hundreds of the world's largest organizations to take a Performance First approach to network management—the new vanguard in ensuring optimal application delivery across the WAN. By focusing on the performance of key applications running over the network and identifying where there is opportunity for improvement, IT organizations can make more informed infrastructure investments and resolve problems that impact the business. Today, NetQoS is the only vendor that can provide global visibility for the world's largest enterprises into all key metrics necessary to take a Performance First management approach. More information is available at **www.netqos.com**.

NetQoS Global Headquarters

5001 Plaza On The Lake

Austin, TX, 78746 United States

Phone: 512.407.9443

Toll-Free: 877.835.9575

Fax: 512.407.8629

NetQoS EMEA

1650 Arlington Business Park

Theale Reading, RG7 4SA United Kingdom

Phone: + 44 (0) 118 929 8032

Fax: + 44 (0) 118 929 8033

NetQoS APAC

NetQoS Singapore Representative Office

Level 21, Centennial Tower

3 Temasek Ave., Singapore 039190

Phone: + 65 6549 7476

Fax: + 65 6549 7001

Website: www.netqos.com

E-mail: sales@netqos.com

20080521