Technical Brief

The BIG-IP System and Message Assurance for Low Latency Financial Information eXchange (FIX)

The FIX protocol used by financial institutions to facilitate trading requires fast, reliable, and consistent response times. In fact, the industry pushes for performance that is so fast and reliable that the FIX environment brings new meaning to high speed and availability. F5 BIG-IP products ensure high availability, improve performance, and enable flexible authentication for financial services organizations implementing the FIX standard.

by Don MacVittie
Technical Marketing Manager

Contents

Introduction

In trading, microseconds and secure authentication both matter. In the time squandered by device latency, trades can go from big money to big bust. More importantly, traders can go from content with a service to unhappy about even small losses caused by slow response times. Furthermore, those traders and the financial institutions that serve them need to recognize that users of high-frequency trading systems are also authorized users. Should a miscreant gain access, the integrity of the entire organization is at risk.

The IT departments of financial institutions are faced with the dual tasks of securing access to these systems and optimizing their performance to a degree far beyond what "optimized" normally implies. The tolerance for latency in a trading environment is well below the average latency accepted in most other industries.

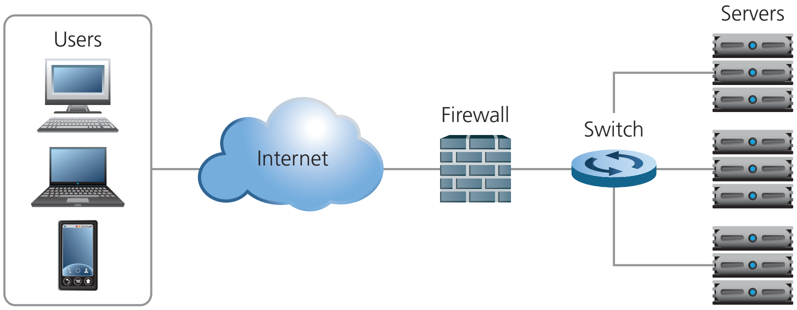

Figure 1: Simplified view of a standard trading network

In the high-volume environment of trading platforms, providing security and availability without introducing latency is nirvana. If the security solution replaces existing, latency-introducing devices, that's even better, because shaving milliseconds and even microseconds from trading times is the definition of success for IT.

Existing infrastructures for retail trading are highly optimized to deliver trades at the lowest possible latency. Security and availability are intrinsic to this architecture, but products supporting these functions are chosen based on their introduction of latency to trading response time. The trick is to improve upon performance that is already very fast while still managing both security and high availability.

F5 BIG-IP Provides Flexibility in Application Deployments

"F5 BIG-IP provides flexibility in application deployments and traffic manipulation with iRules. It provides the biggest value add to the environment."

–IT Professional, Global 500 Financial Services Company

Source: TechValidate
TVID: 7AB-202-B29

Low Latency Trading Issues

In trading, high availability, security, and scalability matter. The users are remote enough that performance is important but is not the only consideration. The need to guarantee service while building the user base brings increased security concerns, which must be balanced with high performance for customer trades. High-speed logging takes on an even greater importance in an environment with a significant number of users and active servers at any given moment.

Conveniently, the latency requirements for retail trading are high enough for IT to implement security, redundancy, server address protection, and authentication, and low enough to benefit from high-speed switching based on content. The issues that IT must address are speed, availability, security, and manageability. While speed cannot be compromised, the other items are not less important.

Rapid Intelligent Switching

Providing the ability to rapidly switch traffic away from a server that is experiencing performance problems or goes offline is an absolute must in an environment with heavy traffic from users who expect the services to be available. Fast switching between client and server, with rapid communication based upon TCP header information, is a requirement. Switching to a new server on failure, however, is slow and fragile in a traditional, fixed-configuration environment. FIX processing could be simplified by rapid, automated switching from a downed server to an available server. This requires not only an advanced switching mechanism, but one with a nearly real-time view of the health of a server or application.

Server Protection Meets High Availability

Implementing network address translation (NAT) is imperative in trading simply because the customers are dispersed and accessing the system over the public Internet. NAT protects servers directly on the Internet by making the NAT proxy visible to the world but the servers behind it only visible through it. Originally, the point was to preserve public IP addresses because providers charged extra for them. Today, the proxy provides another layer of protection and abstraction for critical servers. With high-speed NAT, the server responding to requests can be changed quickly without impacting the IP address used by clients, which is a huge benefit in a trading environment with a lot of connections from a lot of customers.

When combining NAT with high availability, the server infrastructure becomes more fault-tolerant.

Authentication

A relatively large number of valid users coming via the public Internet creates a need for authentication. That authentication should occur before the connection ever reaches the servers hosting trading applications. Otherwise, attackers are already on the network before they are rejected, leaving them a window of opportunity to gather valuable information about the systems and applications they are trying to target. If they are rejected before they reach those targets, however, there is no way for them to perform simple fingerprinting, let alone more sophisticated snooping or attacks. Thus authentication needs to occur at the edge. It also needs to utilize existing authentication, authorization, and access (AAA) tools. If authentication is not tied to centralized AAA mechanisms, it becomes more difficult to customize and maintain AAA policies and procedures, adding another device that must be updated when users are added, removed, or have rights change.

SSL Offloading

Since retail trading is often performed over the public Internet, SSL offloading is also a valuable tool for improving server response time. Shifting CPU-intensive SSL encryption operations onto a specifically designed hardware device saves CPU cycles that can be utilized by the application to speed performance. SSL encryption may also simply be faster, depending upon the latency between the encryption device and the server—typically very low in a retail trading environment.

High-Speed or Highly Available FIX Switching

Switching connections, based upon FIX content, to connect users with their dedicated server is necessary in a high-speed retail environment. Being able to pick out the Sender Comp ID header and connect it to the server previously assigned—or the replacement newly assigned after a server outage—provides for high performance while ensuring the connectivity the user expects. The problems with such switching are speed and flexibility: finding an implementation fast enough to maintain performance in the face of heavy usage during the normal trading day and operating at a level of flexibility that allows for rapid replacement of a downed server in the event of a failure. A system that parses layer 4 data and then uses hardware to switch after first connection is the best solution, introducing a small amount of latency when a connection starts up and using hardware switching thereafter.

If high availability is a larger concern than a small amount of latency, the ultimate expression of high availability is the ability to parse the Sender Comp ID and use it to point at a pair (or more) of servers that are arranged as an active/standby set, eliminating the need to redirect IP traffic in the event of a system failure. If server 1 is active, it handles the request. If not, server 2 handles the request. This arrangement, establishing a primary and a secondary for each and every valid Target Comp ID, also allows for in-place upgrades. Simply upgrade server 2 first, and then take down server 1 (causing failover to the upgraded server 2) to upgrade server 1.

Logging at the Speed of Trading

When monitoring overall performance or tracking down a connection problem, the more information available, the better. But logging can interfere with the performance of the network, and the higher the level of logging, the greater the chance of interference. Trading organizations require a system that can log at sufficiently high speed without introducing latency or latency jitter into the overall network flow. Such a solution also needs the ability to report into a centralized location for easier log management and correlation of events.

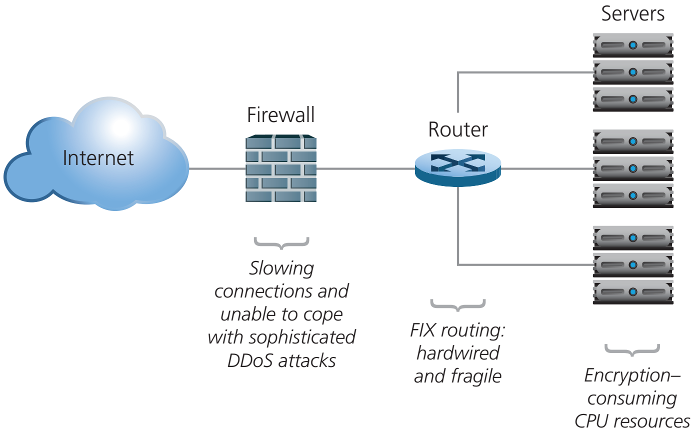
Servers
Firewall
Router
Internet
Slowing connections and unable to cope with sophisticated DDoS attacks
FIX routing: hardwired and fragile
Encryption–consuming CPU resources

Figure 2: Today's operational environment is difficult to manage, with multiple traffic-slowing points.

Solving These Problems

While all of these problems are complex, and that complexity is exacerbated by the environment in which trading takes place, they all can be solved. F5® BIG-IP® Application Delivery Controllers (ADCs) represent solutions for each problem that are suitable to trading and readily implemented. The BIG-IP platform also offers other solutions that are suitable for high speed environments and improve overall security and performance.

A BIG-IP ADC acts as a full proxy and termination point for both ingress and egress application connections, providing unique access to external users, internal services, and all parts of IP connectivity, including application availability, bidirectional address translation, authentication and security services, DNS services for IPv6, management of AAA records, and DNS translations (DNS64) for both IPv4 and IPv6 responses.

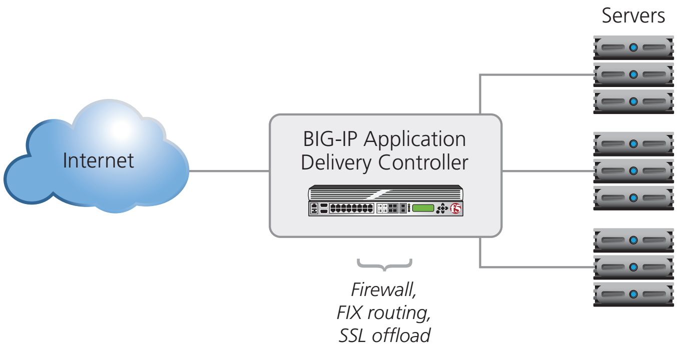

Figure 3: A FIX environment with a BIG-IP ADC in place to provide security, authentication, SSL offload, and traffic switching from one controlled, centralized location.

Fast Layer 4 Switching

For layer 4 switching, F5® BIG-IP® Local Traffic Manager™ (LTM) acts as a forwarding half-proxy, collecting just enough TCP header information to determine where the connection needs to go, then entering the connection into a hardware connection table and becoming a hardware switch from that point on, merely forwarding requests to the designated server. This allows switching decisions based on TCP packet information to be made in accelerated hardware for maximum performance. After session initiation, the latency introduced by the all-hardware solution is minimal, even by trading standards.

NAT and IPv6

The BIG-IP platform, as a full proxy, can implement NAT for all servers behind it, if desired. The ability to perform NAT goes hand-in-hand with the ability to translate between IPv4 and IPv6 addresses, providing the option of IPv6 connectivity without bulk upgrades to the systems that serve client requests. By protecting critical systems with NAT, a BIG-IP ADC braces the architecture to fend off attackers at the edge of the network, and with IPv6 translation, trading platforms are positioned for the future.

Data Center Firewall

An added benefit of implementing BIG-IP LTM at the edge of the trading network is that BIG-IP LTM is an ICSA Certified firewall. The ability to fend off distributed denial-of-service (DDoS) attacks and lock down ports without the need for a separate firewall reduces both complexity and latency in the network. Traditional firewalls are not holding up to modern attack vectors, and devices like BIG-IP LTM are rapidly taking their place to protect critical applications. Removing the firewall not only eliminates one latency introduction point between users and trading applications, it also reduces management time needed to keep multiple, disparate systems up to date.

Authentication

An effective data center firewall utilizes an organization’s existing AAA architecture to filter out users with incorrect credentials. By stopping such users at the edge of the network and routing them to a completely different set of servers to allow them to correct their errors, BIG-IP LTM ensures that a hacker attempting to get into critical trading systems never even gets close. The server notifying the user that the request has been rejected can be on a completely different subnet from critical trading applications, leaving attackers frustrated. BIG-IP LTM can connect to all major AAA protocols, supporting RADIUS, LDAP, and Active Directory, along with an array of two-factor authentication schemes, so no matter what directory services are deployed, BIG-IP LTM can utilize them to protect trading applications at the edge of the network.

High-Speed SSL Offload

SSL offload is difficult to implement in most environments, but when BIG-IP LTM is positioned as a strategic point of control at the edge of the network, it can terminate SSL connections coming in and offload encryption by encrypting traffic on the way out. If encryption inside of the network is required, BIG-IP LTM can terminate the SSL connection both coming in and going out, or it can just pass through the encrypted data. The ability to offload encryption does not apply in the "encrypted everywhere" scenario, of course, because the server had to encrypt the data before placing it on the network. The worthy benefit of terminating SSL is to enable switching based upon data in the packets (see below).

SSL offload and acceleration by BIG-IP LTM removes all the bottlenecks—including concurrent users, bulk throughput, and new transactions per second, supporting certificates up to 4096 bits—for secure, wire-speed processing. Though the cost in CPU cycles of processing 2048-bit keys is seven times the cost of processing 1024-bit keys, 2048-bit keys are the standard in financial services today. Offloading this burden to BIG-IP LTM's high-speed SSL processing hardware frees all of those CPU cycles for application processing.

FIX Switching

Switching based upon the Comp Sender ID is implemented easily in BIG-IP systems with F5® iRules®. A user-created iRule for this purpose is maintained on F5® DevCentral,™ our online community for BIG-IP system users. The iRule parses the Sender ID out of the incoming connection and then directs it to a predefined "pool" of servers, which could be a single or multiple servers. In the multiple server scenario, priorities can be set to identify one server as the primary and another as the standby. As long as the primary is active, all connections will come to that server. When the primary is down for any reason, all connections will go to the secondary. This hot standby implementation enables instant switching—as soon as BIG-IP LTM monitoring notices the primary is down, the secondary starts receiving messages. This multiple server scenario takes a tiny bit longer to transmit a packet, but provides for high availability. Both scenarios are easily configured in the BIG-IP LTM user interface, so the choice depends upon the needs of a given trading application.

When BIG-IP LTM is used for FIX switching, the system becomes more reliable and less fragile. Adding a new customer gateway no longer involves manual configuration of IP address maps, but rather creation of a public IP address and a pool of servers that service it. The same is true with unexpected system failures:

F5 BIG-IP Provides F5 for Trading

"Using F5 we distribute trading application traffic based on session header, which enabled to us to reduce servers dramatically."

–Chief Technology Officer, Medium Enterprise Financial Services Company

Source: TechValidate
TVID: 55B-899-B62

The secondary server in the pool takes over immediately, connections to the failed server receive instantaneous reset commands (as opposed to waiting for a TCP timeout), and customers continue with their activities, nearly uninterrupted.

High-Speed Logging

BIG-IP LTM has the ability to pass TCP or UDP log traffic at extremely high rates. Support for both local and external logging enables the use of industry-specific toolsets to analyze log data and detect malfeasance. Writing logs at the rate demanded by traffic and user-defined logging levels, without impacting performance, allows highly granular logging, which can be sent to log aggregation or analysis tools while connections keep flowing.

Mobile Security

If mobile retail trading may be allowed in the future, F5 technologies not only secure access but enable IP geolocation and consideration of device type (mobile, desktop) when making "allow/do not allow" decisions with AAA. A trader from Cleveland could be rejected if her mobile login attempt suddenly comes from Columbia, for example, even if the credentials were correct. Trades coming in from blacklisted geographies can also be automatically rejected and directed to a server informing the user that the trading platform legally cannot allow them to trade.

Conclusion

Trading environments require comprehensive solutions that solve multiple problems without introducing multiple sets of latency. The F5 BIG-IP platform offers solutions to today's most difficult technology trading problems while preparing the network for the needs and speeds of tomorrow.

Traders and the business owners of a trading platform want better performance, high security, more than "five nines" reliability, and quick recovery from hardware failures. While the order of importance for these business goals varies depending upon the trading platform and its customers, all apply to any given trading environment. F5 answers these demands with a unified platform customizable to a given platform's business needs. While AAA and NAT services and data center firewall functionality protect the network from intrusion and denial of service, layer 4 and FIX-based switching speed access and increase availability—either with wire-speed switching or with FIX-based traffic direction. SSL offload and load balancing improve performance, while high-speed logging helps track and diagnose problems with the network. All of these functions can be delivered from a single point of latency—a BIG-IP ADC—that will grow with the needs of the network.

F5 Networks, Inc. 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447 www.f5.com

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com