

How to make \$100 Million in a Millisecond – Fighting latency and data loss

Within securities markets, trading firms are constantly competing to get the most valuable information and to apply that information in a way that optimizes the prices at which securities are bought and sold. Under these ideal conditions, enormous trading profits can be realized.

Trading orders are often placed electronically, based on trading decisions made using a variety of data inputs such as news feeds, bid and ask quotes, and trading volume history. These data inputs are sent by information and content providers and transferred to feed handlers and trading engines at trading firms via various computer networks. Trades are also placed electronically and using networks.

In recent years, news feeds have become quite sophisticated. Much of the financial news is distributed in a compact, electronically readable form (such as an xml file) that can be quickly evaluated by software rather than a human reader - often in a millisecond or less. Examples are Reuters' Data Feed Direct and Dow Jones' Elementalized News Feed.

If a trading firm sets out to optimize the entire trading process, it becomes important to manage the input data faster and more appropriately than the other firms. However, certain questions arise at the outset:

- How can I get the news information faster?
- How can I make decisions faster?
- How can I make better decisions?
- How can I get my transaction orders to the market faster?
- How can I get my order confirmations faster?

If you can figure out the answers to these questions more quickly than your competitors, then you might be able to gain as many as a millisecond advantage in the time it takes to process input data and place a transaction order. This could mean the difference between placing a successful order and reaping the trading profits - or having a competitor preempt you.

By one estimate published in Information Week, effective management of data can make a difference of \$100 million per year, or roughly \$400,000 per trading day, for a major brokerage firm. Conversely, if one of your competitors has an implementation that is one millisecond faster, that competitor could make an extra \$100 million a year - most likely at your expense or than of another competitor.

Obstacles to Automated Trading

A first step toward optimizing an automated trading system is to have a way of measuring possible obstructions that can interfere with the overall process of evaluating input data and placing transaction orders. Two main culprits are latency and message loss.

Two distinctly different types of latency affect the automated trading process – application latency and network latency. Application latency refers to the time it takes a

trading engine to assess the data that it has at any given moment, placing a transaction order that is optimum for that moment. Application latency can be minimized by writing algorithms and software that execute quickly, and running that software on the fastest possible computer platforms. But, this may come at a price; some complex trading programs may require the crunching and interpretation of a vast amount of data.

On the other hand, network latency can be minimized using a number of methodologies. One of these is co-location – locating the decision-making trading engine or computer in direct physical proximity to an exchange's server, or as close as possible to the source of a news feed. Much can be done in this area, but there are practical and external limits to this approach. For example, a news service may be located in New York, but an options exchange may be located in Chicago; alternately, the news may affect stocks that trade only in Sydney.

Sometimes a certain amount of latency is inherent to the way a network is laid out. For example, a news feed coming from another location has fixed minimum latency because of the physical distance, but it can have additional latency that varies as a function of the amount of traffic. News feeds are, by design, uni-directional multicast streams. They have to pass through many network elements, including switches and routers, on route to subscribers. These switches and routers can, on occasion, be overwhelmed by traffic even in a split second, causing undesirable queuing of data and - in extreme cases - loss of parts of the message; this network phenomenon is also known as a microburst.

Configuring a network to minimize latency is a critical goal, but is not always under the complete control of the automated trading system. Where external variables can increase latency, it is important to be able to know with great precision just how much market latency exists, for example, in an incoming data stream. It is also important for a news provider to know how much latency can occur between itself and its customers in order to be able to accurately characterize the service that it offers.

How to Track Latency and Message Loss

The most effective way to keep track of latency and message loss is to install a latency and message loss management system. These systems typically work by deploying suitable monitoring and recording probes at various points along the data path. The capabilities of these systems vary from vendor to vendor. However, there are some key features to keep an eye out for. Look for solutions that can capture and store all network data over relatively long periods of time – days, week, even months. Then, be sure that the probes can be accessed from a central server through an out-of-band network so that data comparisons can be performed among the various recording probes located throughout the network. This is the key to monitoring and measuring latency and message loss.

In addition, be sure that the probes have an accurate time-sensing mechanism so that discrepancies between clocks on the probes are greatly reduced, if not eliminated. Often times NTP (Network Time Protocol) is sufficient, but for sub-millisecond accuracy (down

to 50 microseconds) a GPS (Global Positioning System) receiver installed on each probe may be required. In this case, be sure that your latency management system supports that. IEEE1588 is yet another mechanism used to provide clocking information to probes. It is perhaps the most difficult scheme to implement because it requires that the network is entirely IEEE1588-compliant. This means that all network elements must support IEEE1588, which may be difficult to achieve.

Another way to track latency and message loss is to look for a way to be notified when an unacceptable amount of latency or packet loss occurs. The device should have user-settable thresholds and a versatile alarm system. The system should have a way of logging all instances of unacceptable latency or packet loss, as well as having the capability to run scripts. For example, if it becomes clear that data is arriving several milliseconds later than normal, the script could send an interrupt to the trading program, either to stop placing orders, or to put more weight on transaction data – quotes and volume – rather than relying solely on the news feed, which may have reached a competitors' automatic trading system first. This would help to avert a potential scenario in which a trading engine is making trading decisions based on stale information.

Finally, deploy a system that can watch for unusual latency or packet loss events that occur only intermittently. Sometimes it just does not work to wait for an alarm trigger and then attempt to start a capture to see what is going on. The ability to capture all of the packets may be the only way to detect these intermittent events. Since problems with latency and packet loss occur more often during periods of high traffic load, be sure that your latency management system can keep up with the full network line speed without dropping any packets itself.

Summary

Automated trading can produce substantial profits, but it depends critically on the ability of your system to assess the status of the dual enemies – latency and packet loss. A right decision based on old data becomes not only the wrong decision, but also a potentially disastrous one.

If you are a news feed provider, investing in equipment that can really tell you how quickly and reliably your news feeds are getting to your subscribers is critically important. It is perhaps one clear way that you can demonstrate your value over that of a competing news service. If you are an automated trading firm, investing in equipment that can tell you exactly how old a piece of data is, and promptly communicate with your trading software to make the appropriate adjustments that can mean the difference between making a trading profit or not.

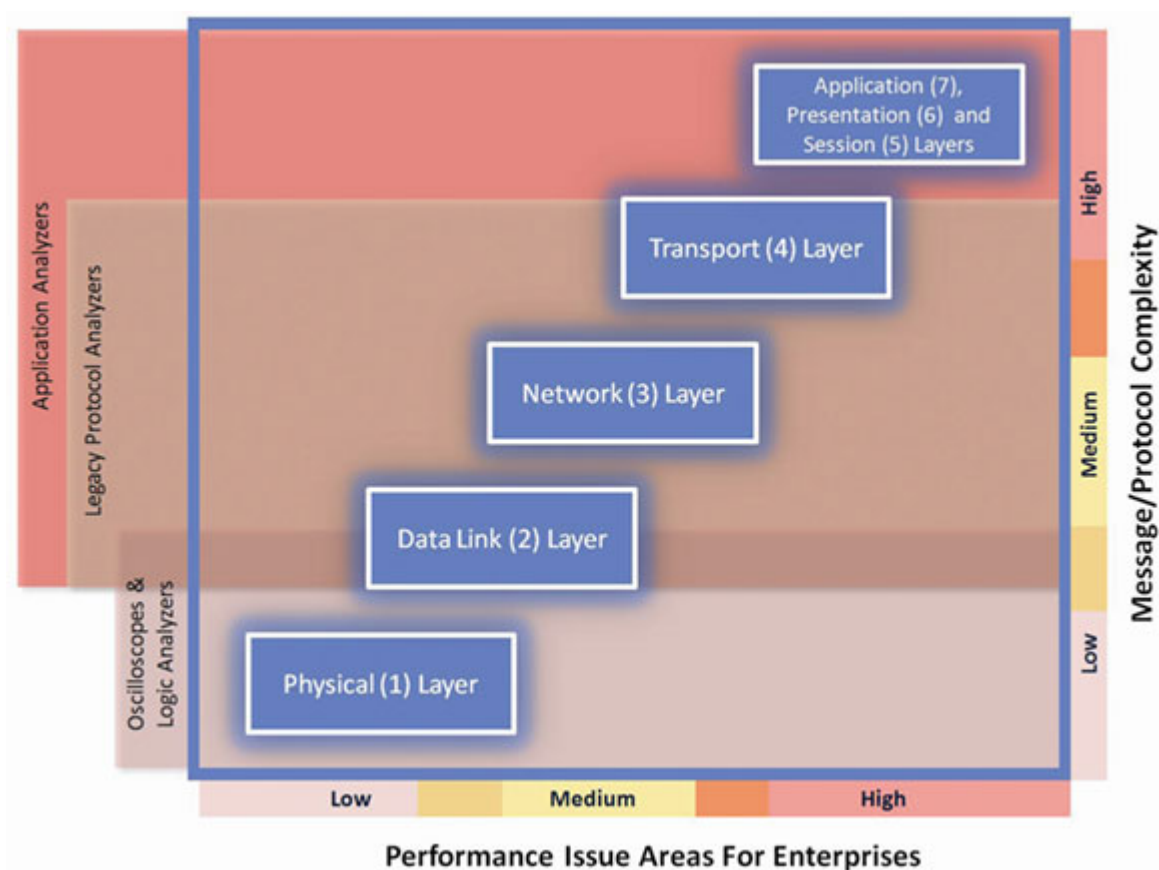
Optimize network performance

Today's management and monitoring tools can help prevent costly downtime.

by Steve Wong

The character of an enterprise network, as well as the sophistication of available network-analysis tools, has changed radically since the first use of Ethernet. Today, companies use the Internet for disseminating information and for e-commerce, affecting the amount of resources a corporation devotes to its network, as well as the importance the organization places on keeping its network running properly.

Along with the growth of network usage, there has been a corresponding increase in the availability of network-analysis and management tools. This is seen in the fragmented nature of the network-management industry; freeware analyzers, enterprise-class appliances and everything that falls in between are each jockeying for a place in the IT manager's tool kit.



An organization should ensure that the network analyzer it has chosen can see what is happening on most of the seven layers of the OSI network model.

Yet, enterprise networks are more complex today than ever, and the data these networks transport has expanded to include peer to peer, voice over IP (VoIP), storage, Internet and a myriad of other applications. The mean-time-to-resolution measurement (how long is necessary to fix a networking problem once it has occurred), a key statistic used by some organizations, has remained essentially unchanged, however.

There are online cost calculators that help estimate what an hour's worth of network downtime might cost an organization. Typical considerations include lost productivity, missed sales opportunities and the possibility of losing an important customer account.

One estimate puts the cost of downtime at about \$42,000 an hour for an average large business. Thus, if a business' network availability was 99 percent for the entire year, it would still have experienced three days of downtime; this works out to a cost of more than \$3 million.

In order to prevent these problems from occurring in the first place, corporate management should decide what specific tools are required and how much of the IT budget should be spent on equipment such as network tools. For example, an organization might budget 3 percent of the cost of its network infrastructure to tools meant for network management. The amount of gear required would depend specifically on how elaborate a network it has, and how critical the performance is on each network segment.

To make intelligent choices, organizations need to look for certain qualities in the equipment they purchase: ease of use, application-layer awareness, scalability and long-term capture capability.

Look for a network analyzer that has an intuitive user interface that highlights network problems in a way that requires little expertise to recognize. Color-coded icons, for example, might indicate the severity of a condition that might adversely affect network performance or cause loss of data.

The ability to issue alerts via e-mail or pager, or to run scripts to fix a problem is also an important feature.

An organization should be sure the network analyzer can see what is happening on most of the seven layers of the OSI network model, and particularly at the application layer. Research has consistently shown that application failures are one of the key causes of network outages; yet problems at this layer are also more difficult to troubleshoot because of the protocol complexity found at this layer.

Legacy products from just a few years ago are typically not application aware and lack the capability to see and troubleshoot what is today a critically important area.

If an organization has only a simple network with a few segments, a network analyzer program installed on a portable computer may be all it needs. On the other hand, an organization with many network segments, especially with some in remote locations, may require an analyzer that has a distributed architecture with agents that can run unattended and be accessed and controlled from a single central console.

If the network has mission-critical segments, an organization may want to consider a network-capture appliance that is able to capture and store network data over periods of days or months, without losing any packets. This is useful for diagnosing problems that are intermittent in nature. It can also help to spot times when network utilization spikes during hours when no one is watching.

A few years ago, the use of a protocol analyzer was considered an IT best practice. Today, protocol analyzers often are used in conjunction with a network-capture appliance.

If an organization is planning to use the network for voice or video transmission, it should look for a feature that will display quality metrics such as mean opinion score and jitter/latency values to help characterize VoIP quality. Features such as the ability to listen into a VoIP call are also key capabilities available in some high-end solutions.

Another useful capability is the ability to perform service-level agreement tests. These allow an organization to determine whether a problem resides in its own network, or whether it results from a provider that is not delivering the quality of service that it is paying for.

Even though corporate networks are more extensive and more complex than ever before, an organization can keep its network in good health more easily today. An investment in network-monitoring and management solutions can reduce the organization's total network operating cost and the risks that might otherwise arise from poor network performance.