

# Tech Talk

## Co-Lo Networks

By Scott Caudell

Over the past decade, futures trading technology has undergone a major transformation. We've seen the speed of trading move from being measured in seconds in the pit to milliseconds on the screen and now microseconds with the advent of proximity access and strategy automation. Ten years ago co-location was not in anyone's vocabulary. Now it is almost a given in the high-speed trading community, and the search is on for new ways to maintain the so-called "infrastructure edge."

One answer is to focus on network latency. As more and more exchanges have moved to support high-speed access to their match engines, new opportunities have emerged for arbitrage between different market centers or different asset classes. For example, the emergence of BATS, an all-electronic exchange optimized for high-speed execution, accelerated the migration towards low latency exchange access across the entire U.S. equities marketplace, paving the way for a surge of interest in high-speed arbitrage between equity exchange-traded funds trading in New York and equity index futures trading in Chicago.

A deluge of telecommunication companies and network vendors are pouring resources into the financial space, making it possible to build extremely fast networks of co-located servers. Telecommunication carriers are now building and marketing new fiber paths between major financial centers that are specifically tailored to financial customers. This new landscape has created a renewed interest in cross-asset trading that links market centers in Chicago, New York, London, Frankfurt and Tokyo.

For this type of trading, knowing the fastest route between two markets can be just as important as having co-located servers at either end. Several years ago, when IntercontinentalExchange first began offering WTI crude oil futures in competition with the New York Mercantile Exchange, a number of high-velocity trading firms invested in high-speed connectivity to the ICE match engine in Atlanta in order to arbitrage between the two markets. Because of the way that ICE managed its technology, firms in Chicago had to

route market data and orders through Indiana, Ohio, Pennsylvania, Virginia, North Carolina and South Carolina prior to getting to the ICE match engine located in Atlanta. For a period of time, trading firms savvy enough to use a lesser-known path that went through Indiana, Kentucky and Tennessee could save 13 milliseconds on transit time. This was rendered moot when ICE chose to move its matching engine to Chicago, but while it lasted it was a massive advantage for firms engaged in this type of trading strategy.

All of the big four futures exchanges—CME, Eurex, ICE and NYSE Liffe—now offer high-speed access via co-location services. Singapore Exchange began offering this last year through a facility managed by SingTel, and several Japanese exchanges will be rolling out new co-location services this year. The expansion of co-location services has opened up a number of interesting arbitrage opportunities among these futures markets and also between these futures markets and related markets in other asset classes such as equities, equity options and foreign exchange.

Even precious metals come into play. Now that Eurex has listed a gold futures contract, there is a potential three-way arbitrage between Frankfurt, the old CBOT gold contract now traded on NYSE Liffe US, and the old Comex contract traded on CME's Globex. This same arbitrage could also be deployed on an even greater global scale with the gold futures traded on exchanges in Tokyo, Taipei, Dubai, Mumbai and Shanghai. Yet another example would be the abundance of currency spreading that goes on between the spot forex platforms based in New York and London and the CME and ICE forex futures in Chicago.

### Network Latency

While there are numerous firms that successfully leverage existing co-location solutions and proximity access, the challenge today lies in successfully connecting physically diverse trading locations. The latency encountered between two or more co-location facilities can

be a major factor in determining the success of cross-market trading.

Direct fiber paths between two proximity sites are rarely available. Rather than simply drawing a straight line from A to B and then laying fiber along that path, a telecommunications carrier first must find right-of-way passage, which literally means utilizing railway or expressway routes to run fiber. A carrier also considers how best to maximize population and communication density along the line. Consequently a connection from A to B can end up taking a very indirect path via the available transportation routes to hit as many cities and towns as possible.

The route between New York City and Chicago is a typical example. There is no direct fiber path between these two cities, at least not yet. Instead there are several indirect paths that zig-zag through various cities. One carrier might use a path that runs via Chicago-Detroit-Cleveland-Buffalo-Albany-Boston-New York, while another might use a path that runs Chicago-Cleveland-Pittsburgh-Philadelphia-Newark-New York.

Distance travelled is not the only factor in determining the latency of any particular path. Some of the cables are older and are not capable of running newer and faster optical equipment cost-effectively. The actual fiber itself can also be impacted by varying levels of attenuation depending on manufacturing standards, which means optical signals may need more amplification and regeneration on some routes than others.

As a result, there are significant variations in latency. On the New York to Chicago axis, latency can range from upwards of 30 milliseconds to below 20 milliseconds. On the New York to London axis, latency can range from over 100 milliseconds to below 80 milliseconds. In a world where trading firms are spending hundreds of thousands of dollars to shave just a few microseconds off their transit time, these variations are huge.

Although high-velocity trading is concentrated in the major financial centers where the largest liquidity pools reside, there's a growing

amount of interest in emerging markets. This creates additional technological challenges. Finding the fastest path to these markets means grappling with vast differences in telecommunication standards and quality, and in some cases it requires dealing with a veritable rat's nest of submarine cable systems at varying degrees of age, performance and stability.

An important distinction is the difference between the terms "proximity hosting" and co-location. In the context of the financial industry, proximity hosting is a term used to describe hosting of trading applications on servers in facilities that are sponsored or supplied by an exchange. One of the main attractions of proximity hosting is that it does not require the use of telecom carriers between the exchange and the trading firm, eliminating a major variable in performance and greatly reducing the transit time for order messages.

Co-location is a broader term widely used prior to these advancements in the financial sector. Co-location facilities, or "telco hotels" as they were originally called, are carrier or third party owned facilities that originally offered off-site hosting capabilities as well as internet and carrier peering capabilities. These mainly catered only to the needs of large companies like Google or Yahoo and their massive server hosting needs. Following the dot com bust, these co-location facilities needed to find a new way to fill space and cover costs. This is one of the factors driving the sales and marketing focus on proximity hosting.

It is also important to distinguish between facilities that are managed by the exchanges themselves and facilities that are managed by neutral third-party providers such as Colt, DRT, Equinix, KVH and NTT. For instance, Liffe manages its own proximity location while Eurex relies on Colt and Equinix. Using third-party providers typically gives trading firms more control over the type of connections used and the equipment in the facility, and in some cases greater freedom in connecting to two exchanges through the same facility.

## The Cost/Benefit Equation

Not all market participants need to devote such attention to latency, of course. The ones

that do are primarily the trading firms that use the speed of their systems as a major component of their trading strategies. For these firms, the primary factor in determining profitability is getting to market first.

Even for these firms, it is important to weigh the cost of technology against the potential revenue a trading strategy could generate. There are too many firms that focus on expensive hardware and infrastructure improvements that yield only single digit microsecond improvements, but simultaneously ignore millisecond level improvements in software and logic enhancements. More and more it will become important for firms to leverage products and services that measure end-to-end latency across their entire technology stack, not just the network."

Even today, despite the immense amount of resources that has been invested in high-velocity trading, many firms would still be hard pressed to answer how long their trading applications take to run their algorithms and generate an order message. Occasionally, a trading group may have been lucky enough to experience a brief "blue ocean" or "golden age" in trading, but they will likely have a skewed view of cost-benefit because they are not able to identify what created that edge. It may have been a strategy that took them there, or perhaps it was a momentary stride past their competitors in technology.

Whatever the case may be, the truly insightful groups discover that despite those relatively brief advantages, it is the total picture and not simply the "best connection" that wins.

## Managing Capacity

Another key factor in managing financial networks today is the importance of capacity. In many cases, high-velocity trading strategies are at their most profitable when the markets are the most volatile. But these periods are also the most difficult to manage in terms of capacity. Spikes of data at sub-second or even sub 100 milliseconds intervals known as micro-bursts can overwhelm a system's capacity to process market data and add latency to the trading process.

This can be an infrastructure-wide manage-

ment issue and tends to be ignored given the amount of time and effort usually required to analyze, manage, and adapt to these events. Some of the more advanced firms today are beginning to understand these factors and what a large factor they can play in performance management, but there is still a large gap between companies that can effectively integrate these tasks into their overall approach and those that are still in the dark. This issue is all the more important when a trading firm is building a system that requires high-speed processing of market data from several market centers that are physically remote from each other.

This problem will only get worse as exchanges increase the resolution of their market data. Years ago, Eurex used a 100 millisecond pulse to manage bandwidth. Last year it moved towards real time non-pulsed feeds, vastly increasing the pressure on all of its users to increase capacity. A similar issue will arise as CME moves to offering 10 levels of depth in its order book for futures. Most high-velocity trading firms will need to receive every iota of that market data.

In response to these challenges, a host of companies have emerged to provide services to high-velocity trading firms. Some high-frequency trading firms build their entire IT infrastructure themselves, but as this sector of the futures industry has grown, more and more companies have stepped in to supply some part of that infrastructure at a lower cost. That may be as simple as providing rack space at a co-location facility or as complex as managing a global connectivity network.

As high-velocity trading expands globally, only a few firms will have the resources to maintain long term relationships with multiple co-location facilities around the world and build out international long-haul circuits to connect these facilities into one high-speed network. This is where providers like BT Radianz, GuavaTech, Options Technology, Quanthouse, 7Ticks, Savvis, and others can tie together all of these facilities.

**Scott Caudell** is chief technology officer at 7Ticks, a managed services firm that provides IT infrastructure services to trading firms. The firm designs, builds and manages networks for sophisticated trading firms.