

“These days trading floors may have disappeared, but the prize still goes to the swift and nimble in the market”

VIJAY KEDIA



▲ Vijay Kedia, President and CEO of FlexTrade Systems, Inc., founded the company in 1996 when he designed and built the first “broker-neutral” algorithmic trading system, which is the recipient of several industry awards and has a trusted global presence at more than 120 institutions around the world trading equities, options, futures and FX.

Lifting the veil on latency

Vijay Kedia, President and CEO of FlexTrade Systems, Inc.

While seemingly a new area of focus in the media, latency always has been a challenging issue for trading firms. From the earliest days of stock market trading to the present, those with the fastest access to information (in the past they were the people who stood closest to the specialists on the trading floor) have held the advantage.

These days trading floors may have disappeared, but the prize still goes to the swift and nimble in the market. And as electronic trading has taken hold, firms with the fastest network and performance across all of the links in the trading chain will be at the head of the queue, able to react speedily to changing market conditions. Once measured in seconds, latency is now calculated in microseconds, and as markets continue to be volatile, speed of thought and execution remain paramount for many firms.

But what really is latency? And what factors influence it?

In the world of electronic trading, latency is essentially the delay between receiving knowledge of a change in the market and acting upon it. It is during this time that information travels through both software and hardware and across communication links, each element along the path contributing several factors that

introduce a measurable delay before a message reaches its destination. It is the impact of these factors that makes latency an unavoidable, foregone conclusion.

Latency factors

The primary goal, however, is to reduce latency while making things happen quickly at all levels – but it's not only how fast a firm can access data, but also how rapidly it can react to that data. To accomplish this task, one must understand that there are numerous factors that contribute to latency, be they internal to the firm or external. These factors include the trading firm's communications infrastructure, trading system, trading strategy, frequency of trading and trading volume.

It is only when all of these factors are considered that a trading firm will be able to gain a true insight into its latency figures. In order to do this a firm must measure latency across all components, between the trading system and the exchange, from end-to-end (and back again), to accurately track the speed of its trades and to benchmark itself against its peers.

Just how important microseconds are in the trade cycle also depends on a firm's trading strategy, and no two strategies are the same.

Arbitrage traders, for example, will place much more emphasis on speed than, say, a firm with a strategy that is playing out over hours, days, or months. Firms that are sensitive to latency must optimise every component in the chain to stay at the head of the queue. They have to ask whether the data feed they access is the fastest available and how many links there are in the chain between the data provider and themselves. Physical distance is also a concern; there is no point getting data all the way from Asia if you are running a trading engine in New York.

Maintain balance

A balance has to be struck between functionality and pure latency. For example, a strategy running a simple algorithm based on a single stock will put much less demand on trading systems than one that tracks every stock in the S&P 500 and the correlations in movements among them. The latter strategy will require a significant number of computations, which means some latency will inevitably be sacrificed in order to gain the calculations required.

Another balance has to be achieved between latency and access to liquidity. Firms seeking liquidity across multiple venues, including exchanges, ECNs, MTFs and dark pools, cannot locate their trading systems next to all of these venues – thus the advantages of co-location (the modern day equivalent of standing next to the specialist on the trading floor) will be lost.

What to do next?

Remember that hardware and connectivity matters. No matter how quick and efficient the system promises to be, if the entire path to market is not properly evaluated, the perceived gains within the software will be negated by saturated market data bandwidth or overloaded switches on the way to the exchange. Also keep in mind that latency numbers given in vendor marketing materials often have implicit assumptions which do not accurately reflect realistic trading scenarios

and should always be taken with a grain of salt. Without a full understanding of how the metrics were generated, the numbers are just as useful as lottery numbers contained inside a fortune cookie.

A trader, therefore, must ask many questions of his trading systems vendor in order to understand the published latency figures. A system may deliver 200 microseconds latency, but under what conditions? Latency can vary based on the number of transactions undertaken and the time of day a trade is made. Low latency may occur when a firm is trading, say, 1000 times a day, but if it trades one million times a day the burden on the same trading system will be greater, and greater delays will be introduced.

Remember: No one size fits all

While some elements of latency are external to the vendor, such as the communications network, source of the market data feed, and distance to an exchange, vendors should make every effort to ensure that their platforms are designed to introduce the least amount of latency for the complexity required by their individual clients. Vendor architectures should provide rapid processing of market data and the accompanying processing using that data. As noted above, it is not only how fast data is transmitted, it is just as significant to process and act quickly upon the data. This, combined with an internally-developed FIX engine, will ensure the fastest overall performance. This customized, modular approach will enable clients to prioritise components and manage the level of latency they are willing to accept.

In this vein, trading firms should know that a reputable vendor will provide a realistic setting for them to test all of their trading strategies. This would be accomplished on a hosted platform in which a trader can implement his or her strategy and benchmark latency for all scenarios, thus generating different but relevant performance figures from the system. ■