



Latency Round Table

The routes from L ...

...are paved with good technology. These days, it takes a lot longer to say the L-word, latency, than it does to pass a message from one end of a trading network to the other. But that doesn't mean we can sit back and leave it to the competition to shave another microsecond off the round trip to the market and back.

Working with key industry figures, Automated Trader has convened a round table on the enduring issue that is latency. Are there common practices emerging in the industry? What can we learn from each other's management of latency? Could there even be a case for moving this issue from the competitive to the cooperative space?

One idea that has gained traction in the recent past is latency standardisation: would a latency standard help with measurement and benchmarking? And if so, how? That's where we begin today. Our round-table participants are gathered from across the industry, and if you would like to join them, you will find that these pages are only the beginning: the discussion continues at www.automatedtrader.net, where there is the facility to add your own contribution.

We hope not only that you will find this round-table instructive, but also that you will go online and join it. See you at www.automatedtrader.net very soon.

Round-table participants:

- Kevin J Houston, *Chairman, Rapid Addition*
- Emmanuel Carjat, *CEO of Atrium Network.*
- Hirander Misra, *COO of Chi-X Europe*
- Frederic Ponzo, *MD, NET2S*
- Scott Ignall, *CTO, Lightspeed Financial*
- John Oddie, *CTO, Celoxica*
- Petrina Steele, *Vice President of Business Development in Europe, Equinix*

- Shawn Melamed, *CEO, Correlix*
- Rob Ciampa, *VP Product Management, Tervela*
- Ben Newton, *WebSphere Front Office, EMEA Technical Sales*
- Mark Akass, *CTO, Global Banking & Financial Markets, BT*
- Alasdair Moore, *Director, Fixnetix*

Do you consider that an industry standard for latency would be useful?

Kevin J Houston: Yes, at the moment, building low latency systems is way harder and more expensive than it has to be, because each builder has to measure each components performance to determine how well it performs, the difference are vast yet all vendors claim to be the best. One "high performance, low latency" FIX engine we know of is lucky to handle 2,000 msgs/sec another 100,000 msg/sec yet there is no way of comparing them other than performing a Proof of Concept with each.

Emmanuel Carjat: An industry standard for latency is inevitable. Given there is so much investment being made, there needs to be a way of "measuring" the success or identifying how much further you need to go. However the challenge starts with: who is measuring "latency" and what are they measuring? Historically, connectivity providers have been a source of time-stamps.

As the market move towards finer granularity of measurement (ie microseconds instead of milliseconds), connectivity providers have an opportunity to continue providing this service, using highly accurate time stamps.

Hirander Misra: Yes, it is important to compare apples with apples as far as latency is concerned to ensure consistency. An independent standard to benchmark against will greatly assist this process and distinguish real performance from marketing hype.

Frederic Ponzo: Indeed, it would be. However they are some basic impracticalities that get in the way. Latency measurement across multiple sites is notoriously hard to do, because of simple things such as clock synchronization, and harder aspects such a deploying measurement kits in somebody else's data centre.

Transparency of the latency in trading venues, or network links, will rapidly mean that clients (banks and funds managers alike) will start asking for SLAs on latency, ▶



Kevin J Houston



Emmanuel Carjat



Hirander Misra



Frederic Ponzo

Latency Round Table - The routes from L ...



Scott Ignall

which obviously creates a negative incentive on the providers. Also, latency is a relative quantity; it is about how fastER or slowER you are against your competitors. A standard benchmark might disclose too much information in the eyes of some players.

Scott Ignall: Not particularly. The standard is defined by the customer and he or she will seek out the providers that can offer the lowest possible latency. A significant reason why we have excelled in this space is due to the speed of our automated trading interface and market data products. It is our belief that execution times over 1 millisecond roundtrip (from our clients to the market and back) is really not "low" latency - our clients would certainly agree.



John Oddie

John Oddie: A number of companies are building businesses based on low latency trading, some are vendors of tools, others service vendors like Celoxica and of course the companies that perform automated trading devote significant dollars to achieving the lowest possible latency in their trading applications. Standards will probably emerge as a result of these investments as particular companies achieve dominance. As yet I see no concerted move to open standards in this space.

Petrina Steele: Yes, an industry standard for latency would be highly useful. However, there are almost too many components to understand - as there is currently no generic standard for measuring, it makes latency almost impossible to measure and benchmark. Even the likes of Corvil, SeaGate, Correlix and TS-A, who are highly focused on the latency arena, define their measurements differently.

Shawn Melamed: In high frequency trading, low latency is a critical factor to maximize profitability. Given the complexity of the entire

chain of trade events from price discovery to trade execution, there are many points where latency needs to be managed and it is often outside the trader's control. By setting an industry standard for latency, traders would be able to pinpoint and identify latency bottlenecks to achieve best execution.

Rob Ciampa: Yes, we believe there should be latency standards. And, the reality is there should be many - not just one. An industry standard benchmark for latency would be useful for out-of-the box evaluation of messaging performance, but this is only a start.

Ben Newton: We'd need a set of standards, each representing a consistent piece of the total picture. If we could state exactly what each component absolutely needed to do then we could begin to make proper comparisons. A balance will need to be struck between concise standards and the F1 rule book which stretches into 100's of pages. Without careful thought a standard could be either unusable or meaningless. A well thought out set of standards would be universally helpful.

Mark Akass: Some form of reference model would help in like-for-like comparisons and benchmarks. Otherwise there's a lot of claims made for performance that are hard to verify or assess. It's unlikely to be an enforceable standard, particularly if it adds materially to the costs. It might also be very useful to have agreed time source standards to provide a common reference point and allow data sourced from different elements to be integrated appropriately.

Alasdair Moore: No. A latency standard whilst clearly a desirable benchmark to evaluate vendors, is highly improbable. There are areas where this could more readily be applied such as hardware, where there are fewer variables. However not only is latency non static the

question is what are you measuring from where to where, what format, what environment and what throughput? Each vendor whether that be hardware, software or network provider would all have to agree to common metrics which is impossible.

What do you see as the biggest contributors to latency?*

Houstoun: That will depend, but once you optimise your components often you are left with a system where the box to box IP hop is the dearest piece of the chain.

Ponzo: Most financial institutions have tuned their networks or taken advantage of proximity hosting solutions. The network, on average, contributes to just 13% of a system's overall latency. This is now the second smallest contributor, below messaging (at just 2%). The new latency culprits are the applications at 65%, and firewalls at 20%; which together contribute to 85% of overall latency today.

Our research shows that the fight against latency in both the middleware and network layers is pretty much a won battle. The industry must now do battle with their applications if they are to finally win the latency war.

Ignall: The first stumbling block is usually network architecture that is not properly tuned for the traffic. The next one is usually application logic with too much complex processing. Less is more in the low latency business.

Oddie: Again this depends very much on the problem one is trying to solve, but obviously proximity, network bandwidth, message transports, message rates, hardware, operating systems and application architecture all contribute. What Celoxica have tried to do is to break the problems we solve into logical business components that can be

highly parallelised so that we can configure our solutions such that each component in the chain gets no more data than it can handle. All of our solutions can be configured so that they are thread-safe without context switching.

Steele: The biggest contributor to latency is simply a lack of real understanding of components: specifically their usage levels and complicated architectures.

Melamed: Trading infrastructure is a complex environment and the smallest change from new applications and technology upgrades to new trading strategies can impact the entire trade execution and market data delivery process. Since order executions traverse multiple trading environments and networks, having real-time inter-party monitoring and analysis is the only way to identify the biggest latency contributors.

Ciampa: There are so many systems involved in the modern data architecture that latency is often created by the interdependency of the different components. This makes latency very difficult to measure and the results of testing it don't always manifest as quantifiable latency. The performance lag can masquerade as application issues like lost trades, price slippage et cetera.

Burstiness and volatility both accentuate the weaknesses of software-based systems. A data architecture must be both application-aware and network-aware in order to overcome the congenital deficiencies of software in a legacy environment. When one system fails it can create a domino effect and take down other parts of the architecture.

Newton: The biggest single contributor to latency that I have seen is from messaging. Especially for older systems this can contribute 30%-40% of the end to end time. ▶



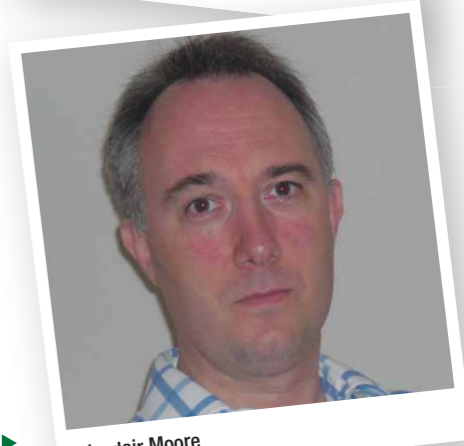
Ben Newton



Rob Ciampa



Mark Akass



Alasdair Moore

There are many techniques to speeding this up, but they all roughly fall in two camps, tune what you've got or buy something faster. We have seen a combination of the two proving most successful. For obvious reasons the fastest systems have ditched information access through disk based databases in favour of in memory databases like SolidDB, where lightning fast access and coherent caches have cut many latency headaches. Database access can drop to 20 micro seconds with SolidDB, multiple orders of magnitude faster than a traditional remote database query.

Akass: Global WAN latencies can be 10s and 100s of milliseconds. Within a metro area the network latency falls to sub-millisecond on a specialist tuned network (e.g. Radianz Ultra Access on BT's network). So for low-latency

trading you must first of all be able to route orders in the local market geography, e.g by proximity hosting with, your servers are in the same city with fast links to the venues (for example, by using BT's Radianz Proximity Solution), or in the same data centre as a specific venue.

After that, the next material latency issues focus on the application layer, such as the performance of the exchange's trading engine. Historically those were also in the 10s and 100s of milliseconds, but the best are now claiming sub-millisecond performance. For traders seeking absolute lowest-latency performance, the other material factors are the speed of processing the market data coming in, the performance of the computing stack for the algorithm and the performance of the smart order-routing tool.

A lot of this is now very well understood and fast becoming an entry qualification for being considered as being a provider in this market. But the application challenge is inevitably more complex and more variable.

Moore: For Fixnetix there are a number of areas of latency mitigation depending on whether we are solving for data delivery or trading connectivity. However in most cases the largest contributor to latency are the customers own requirements which define what is possible.

For more questions and answers on latency, and to make your own contribution to the debate, go to www.automatedtrader.net.

We're solving the needs of today's financial institutions by offering a combination of low-latency, high throughput market data solutions with flexible delivery options

We reduce the cost without sacrificing the service for hundreds of customers worldwide

www.activfinancial.com
sales@activfinancial.com
+1 212 599 1600
+44 203 178 3606

activ
FINANCIAL

ULTRA LOW-LATENCY MARKET DATA

