E-channel traffic jams: Overcoming FX technology bottlenecks

Frances Maguire asks if the benefits of e-FX are still being held up by constraints in trading infrastructure and technology bottlenecks

The increase of automation and the growth of high frequency trading inevitably brings with it bottlenecks. The main bottlenecks are occurring at either side of the actual trade – both pre-trade and post-trade. With the increased use of algorithmic trading, banks and liquidity providers cannot get their prices out there fast enough and with ticket volumes spiralling out of control, the back office is struggling to keep up and confirm trades at the rate they are being traded. While netting will enable banks to improve the back office bottleneck, it is the fragmentation of eFX trading venues that will continue to challenge the eFX market.

Fragmented FX market

Unlike equities where banks can co-locate their servers in facilities at the stock exchanges, such a solution to network latency is not available in the fragmented FX market. Matt Meinel, global director of business development, at high performance messaging software provider 29West, says: "In the foreign exchange market banks have to put their servers in the main financial centres where the main network providers have facilities to do this or build data centres and hubs close to the public internet backbones. The closer you are to the highest speed channel, the better." Alongside Ultra Messaging for the Enterprise, 29West's messaging product suite for high speed, low latency, persistent, IP-based communications, 29West has developed Latency Busters Messaging, a fast, low latency streaming messaging system targeted specifically at market data applications with very high message rates.

Says Meinel: "RFQ's are always going to be slower than dealable streaming prices because they require multiple messages. The fastest way to execute is by sending out a live quote, the customer hits it and the deal is done with just two messages. A RFQ has a minimum of three messages and that is going to add latency to the transaction."

Matt Meinel
"The fastest way to execute is by sending out a live quote, the customer hits it and the deal is done with just two messages."

However, Meinel does believe that RFQs will continue to have a role to play amongst corporates trading infrequently where latency is not priority where the hedge they are doing overwhelms the execution price.

For the high-frequency FX trader, Meinel identifies at least two remaining bottlenecks. The first is the number of intermediate stops a quote might go through between the source and the algorithmic trader. He says: "If a bank is generating auto-quotes to FXall, which is then forwarding the quote to the client, then FXall adds latency in the middle. We see a lot of the banks spending a lot of money upgrading their own e-commerce site because if they can be faster the algorithmic traders are more likely to connect their algorithmic engines directly to the banks' sites. At this highly competitive end of the market, latency is a priority and can win or lose a deal."

The Internet

Another bottleneck that occurs at the technical level surrounds TCP/IP, the communications protocol on which the internet and most commercial networks run. The protocol has built in flow-control, which means that if the receiving computer is not keeping up with a high volume of messages, the protocol tells the sending machine to slow down the messages or stop sending them completely.

Meinel says: "Some of my FX customers turn actually turn this feature off and switch to UDP (User Datagram Protocol) with arrival order delivery instead of TCP's in-order delivery as they want the quotes to come as fast as possible and don't want flow control. The thinking is if they miss one, they will just get the next one, that is going to be more relevant than asking for the old quote to be re-transmitted, but this is not how TCP/IP was designed to work. This is something that players in the high-frequency trading are currently paying attention to and trying to address," he says.

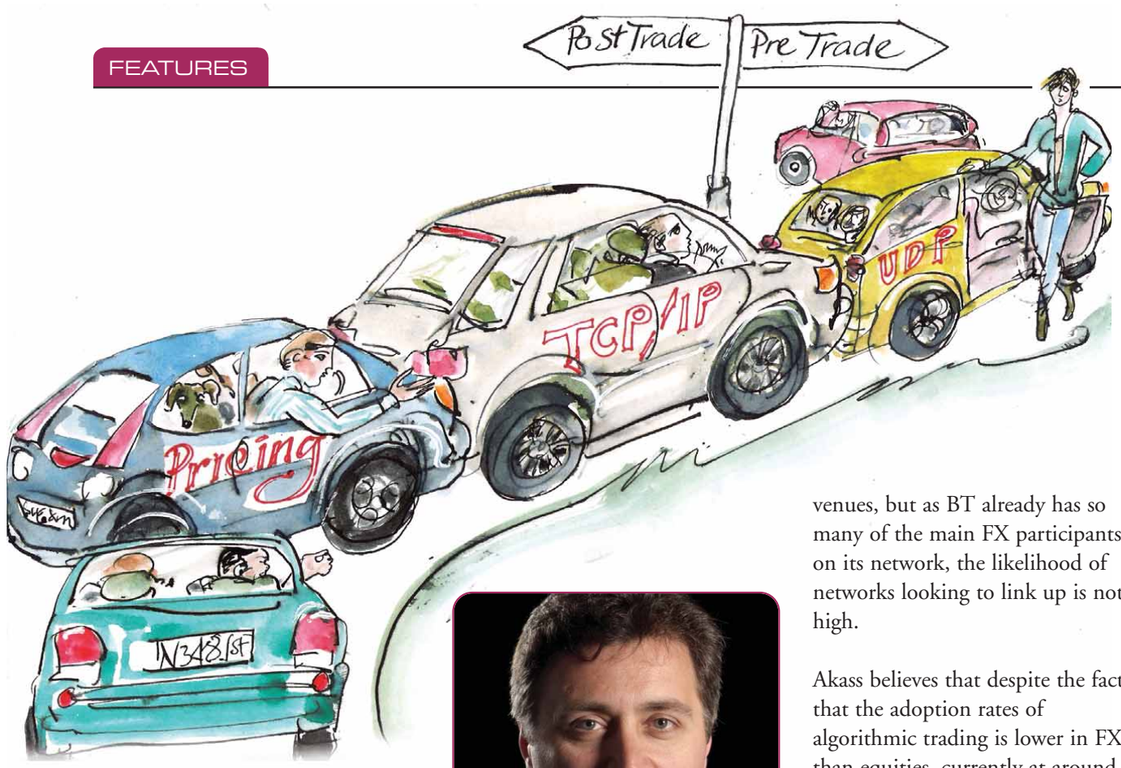
BT Global Financial Services provides shared market infrastructure and manages the service end to end, which includes scaling up the links from the different trading venues and providers of services out to the client.

Mark Akass, CTO at BT Global Financial Services, says: "The managed service offering takes care of the capacity, scaling and bottleneck issues. This becomes our problem and to deal with this we ensure that we provision capacity at every stage to cope with peak market traffic flow."

"This differentiates us from the internet in that we are monitoring the market volume and the trends in the peaks. We look at the peak one-minute volume on the worst trading day or the worst month and anticipate how that is growing in a 6-9 month rolling window. We are constantly upgrading infrastructure capacity to stay ahead of that curve."

Not having enough capacity can also cause latency as the congestion can slow down the speed of information. A high capacity will ensure that the operation's throughput is kept high as well. For those doing latency arbitrage on pricing, latency is clearly critical. The managed service offering looks after the client connectivity, the connectivity to the portals and management of the shared infrastructure on a global basis.

In FX liquidity is fragmented so access is needed to all the different



venues. BT already offers connectivity to more than 70 FX trading applications to enable faster access to market for participants and also faster time to market for FX service providers. An alternative is the internet but clearly the problems with the internet are capacity, scaling and security.

Akass adds: "The internet tends to work fairly well within countries but it tends not to work as well internationally where performance can be very unpredictable over distances which opens time-sensitive trading open to risk. Clearly also the environment is not as secure as a closed network."

Unlike equities, FX trading is more distributed. Without the equivalent of stock exchanges it is not possible to use the same-proximity type solutions where servers can sit close to each other.

Akass says: There are interesting discussions going on at the moment as to whether there will be greater consolidation in the eFX industry with regional hubs forming, or more exchange-like venues setting up. Instead of having many single and multi-bank portals there could potentially be convergence between participants so, for example, orders could be routed through a small number of players that dominate on a geographical basis. It may or not be something that the market ends up driving towards."

Mark Akass
"The internet tends to work fairly well within countries but it tends not to work as well internationally where performance can be very unpredictable over distances which opens time-sensitive trading open to risk."

At present users connected to BT can provide connectivity beyond their connection point into other venues, but as BT already has so many of the main FX participants on its network, the likelihood of networks looking to link up is not high.

Akass believes that despite the fact that the adoption rates of algorithmic trading is lower in FX than equities, currently at around 15%, the fragmentation of the FX markets is causing much less of a problem as they can be connected through a shared market infrastructure. "In terms of network technology, the current levels of algorithmic trading in FX, do not present a challenge. We can quite comfortably scale to cope with growth at the current rate."

Post trade challenges

According to Nick Dyne, founder and CEO of Logicscope the prevalent bottlenecks in eFX occur at either side of execution – pre-trade and post-trade. He says: "The front office has dominated the eFX business. The focus has been on building faster and slicker trading applications with less time spent on the integration and the plumbing. In terms of the post-trade area, very few are properly connected for post-trade notifications, which greatly impacts the speed at which trading positions are up dated, and creates unnecessary bottlenecks in the settlement of trades through data re-entry and error correction."

In terms of pre-trade pricing and RFQs, Dyne says that the lack of synchronisation in the price feeds coming from the different venues into the banks is causing errors in the processing of quotes. By building a server within the bank, granularity can be improved and prices blended into a single price stream before hitting the traders' screens. "If built correctly, this significantly reduces the risk of misapplying rates and reduces the number of concurrent rates required by pricing engines," he adds.

Options pricing is perhaps the hardest bottleneck for the FX industry to resolve as the more complex the instrument is, the harder it is to automate. Furthermore as soon as an option becomes vanilla enough to be automated, traders look for something more exotic. Says Dyne: "Electronic trading of an option is essentially a mathematical dialogue regarding the counterparties' views on the volatility, it is never going to be easy to standardise in order to automate due to the complexity of the multi-leg transaction and the subsequent delivery of the ticket."

He adds that even when two counterparties are trying to connect using the same specifications for messaging, it still takes several days to map each field, for each leg, of each strategy. "The mid and back office bottleneck depends very much on the software's ability to handle the different strategies," he says.

Nick Dyne
"In terms of the post-trade area, very few are properly connected for post-trade notifications, which greatly impacts the speed at which trading positions are up dated, and creates unnecessary bottlenecks in the settlement of trades through data re-entry and error correction."

Logicscope is also integrating real time market data and delivering deal notification and STP solutions, over either a distributed architecture or from a post trade feed with a client adapter, into banks and proprietary traders' back office systems.

TradeSTP, which enables real-time post-trade deal notification data to be distributed automatically from multiple e-trading venues to customers' back office environments, has been recently installed by leading FX trading boutique Dsquare. By fully automating post-trade deal notifications, Dsquare has been able to gain significant market and operational efficiency.

Automated post-trade notifications will also be available shortly on the BT Radianz Shared Market Infrastructure following a recent agreement to offer hosted access to TradeSTP.

Says Dyne: "The FX market has embraced electronic trading wholeheartedly in terms of pre-trade information and deal execution. The challenge has been to make the post-trade environment equally efficient."

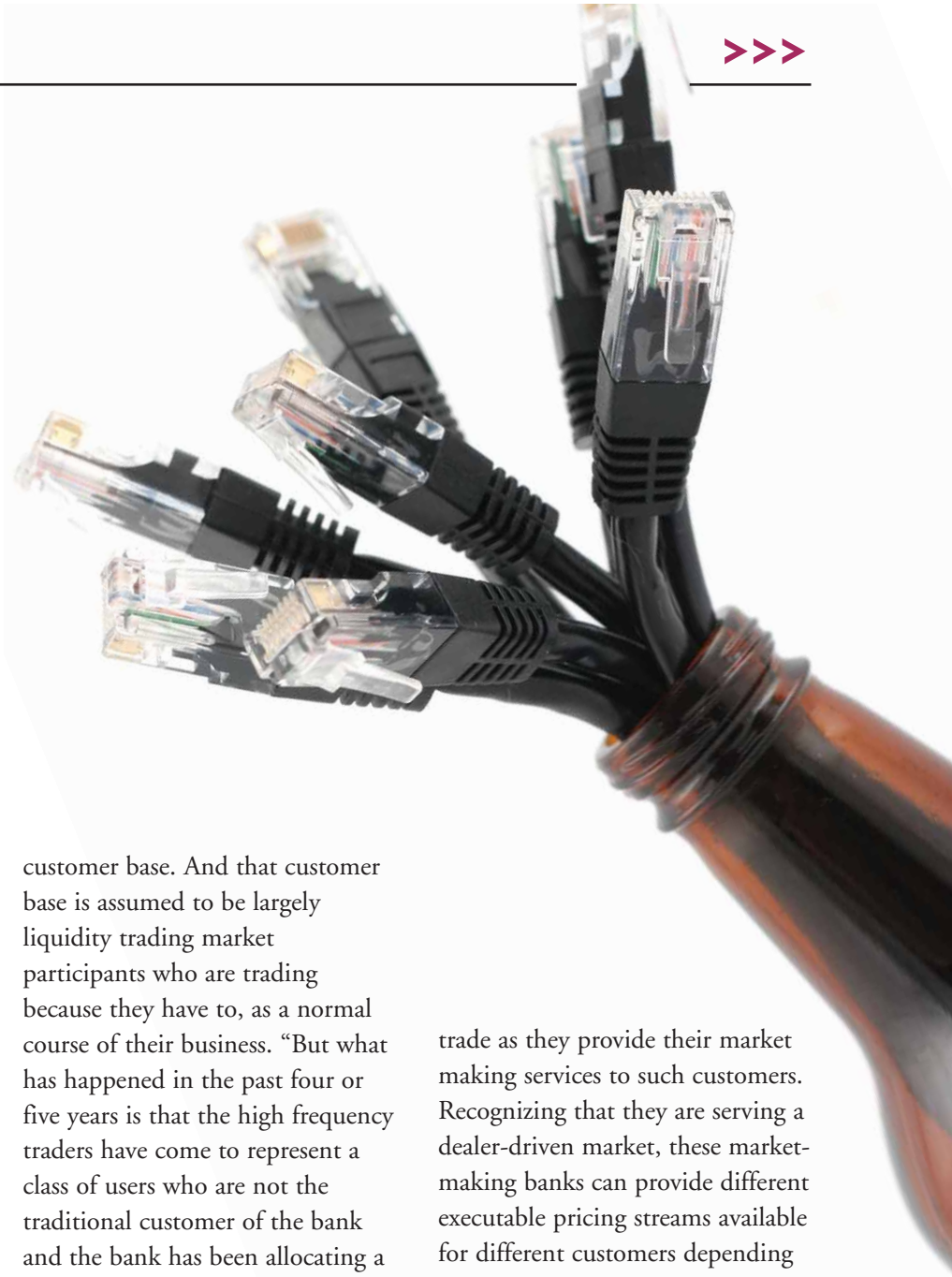
High frequency trading issues

Gil Mandelzis, CEO, Traiana agrees that the biggest remaining bottleneck in the delivery of eFX prime brokerage is the sheer volume of tickets generated by the front office and being passed through to the back office. He says: "The advent of API-based black box trading and the growth of small ticket retail traders, means funds are trading increasingly smaller amounts at higher frequencies. This presents a real capacity problem and cost pressure for the FX prime broker as the sheer number of tickets increases exponentially. In order to deliver quality client service prime brokers must find a solution to these constraints."

He adds that Traiana has responded to calls from the major banks with a ticket netting service called NetLink, which helps manage the increasing number of tickets by netting them at points during the day. NetLink is now live with major FX prime brokers and is in process of being rolled-out to the industry.

Volumes have gone up significantly and high-frequency traders are generating 10-100 times the number of tickets that previous customers would have done, so at that high level, low-cost

Gil Mandelzis
"The advent of API-based black box trading and the growth of small ticket retail traders, means funds are trading increasingly smaller amounts at higher frequencies."



automation and no-touch processing are needed. Latency is death to any provider or taker of prices. To a high frequency trader any built-in latency, because of software, hardware or network design implications, would be a certain barrier to entry into today's FX market, where servers are needed to be co-located at data centres, and as close to counterparties as possible.

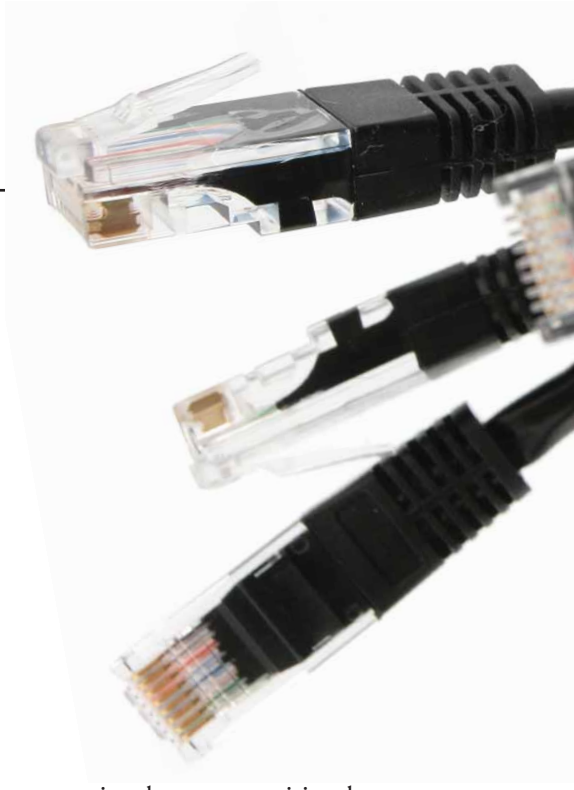
For this reason, Harpal Sandhu, CEO of Integral Development, says that the real issue that is now coming to light has to do with a bi-product of the new entrants into the electronic FX market – the high-frequency traders. He says: "This has caused a mis-matching of counterparties with each other in the electronic space, and the problems that arise from that."

Different customer motives

Many of the banks that participate in providing prices in e-FX do so given their historical bias of providing a service to their customer base. And that customer base is assumed to be largely liquidity trading market participants who are trading because they have to, as a normal course of their business. "But what has happened in the past four or five years is that the high frequency traders have come to represent a class of users who are not the traditional customer of the bank and the bank has been allocating a product to this particular group, that really is not appropriate and has proved disadvantageous to the bank as they are informed traders, using various algorithms to profit from arbitraging the markets. This mis-matching has led to a great deal of losses on the part of the banks, and a great deal of artificial profitability on the part of the algorithmic traders," says Sandhu.

It has only been in the past six months that there has been a realisation that the market making banks can, in fact, take into account their customers' motives to trade as they provide their market making services to such customers. Recognizing that they are serving a dealer-driven market, these market-making banks can provide different executable pricing streams available for different customers depending on their customers' motives to trade. And this is how they can operate a reasonably profitable business again.

Sandhu says that the biggest problem banks and liquidity providers are facing is that much of the existing electronic infrastructure available does not allow the flexibility to customise who should see who in the marketplace, and under what context they should see them – whether anonymous or not, a central limit order book, a streaming price or a request for



quotes. Most of the systems available are 'one size fits all'.

The upshot of this mis-match is that the volume of tickets have spiralled out of control but the real issue is that if the liquidity provider (market makers) was actually making the same profitability margin on the increased volume then the technology issues would not be a problem – it would just be a case of investing in more infrastructure.

The problem is that margins have gone down significantly as volumes have soared and if the bank did not give such an off market price it would most likely not have so many tickets.

Harpal Sandhu
"the solution to the problem is not fixing the computers to process more tickets but fixing the appropriateness of the price stream and the venues on which counterparties meet."

Says Sandhu: "So the solution to the problem is not fixing the computers to process more tickets but fixing the appropriateness of the price stream and the venues on which counterparties meet. Taking advantage of the fact that one size does not fit all. Once the liquidity providers engage in the right sort of business, the ticket numbers may either go down while the profits will increase."

"Banks need to step back and look at the cause and effect, and not allow the tail to wag the dog. It may well not be the fault of the systems but the positioning in the market."

Having zero latency, high performance, low-cost, and increased automations are still important but they are not necessarily the heart of the problem. Sandhu says that for successful banks the question is not about the technology that is leading to the bottlenecks and prevents active participation in the market. For them, the real question is how to offer the right pricing stream to the right customer through the right channel for the right reasons with the utmost strategic and business agility."

Integral developed a solution to this problem two years ago, FX Grid, which is only now coming into widespread use. All 14 of the top FX banks have adopted the platform along with 100's of taker institutions. FX Grid is still a shared transactional network, but liquidity providers choose who they will trade with and users can choose who they call. This gives all the advantages of having a shared environment and being able to talk to all customers without having to build proprietary infrastructures, but gives the trading participants the visibility and the control required, so it allows them to recognize the opportunities that exist when 'one size does not fit all'.

FX Grid eliminates a lot of the business risks by allowing the banks the ability of giving different executable pricing streams to different customers based on those customers' needs and motives to trade, and has given banks the ability to change, in real time, who they are trading with and under what platform they are dealing on.

Says Sandhu: "It has essentially given them back control over what they are doing in the e-commerce space, transformed the way they view the risks of the business, and enabled them to look more closely at business strategies and business performance."

The eFX industry, it seems, has a choice to make – either keep up the arms race, and build even faster technology, or step back and reassess what exactly they are trying to achieve. Going back to basics could be simply a case of looking at the business case. If generating more business is causing costly bottlenecks, less can actually be more.