

Performance Considerations in FX Trading Platforms

In this article, Mike O'Hara, publisher of The Trading Mesh - talks to Ash Gawthorp of The Test People, Matt Barrett of Adaptive Consulting, Andy Phillips of LMAX Exchange and Dan Marcus of ParFX about how technology challenges - particularly around performance - are being addressed as both the retail and the professional FX trading sectors adopt new trading platforms.



Introduction

The Foreign Exchange (FX) market is the most heavily traded market in the world, averaging \$5.3 trillion per day in 2013¹. Although most of this takes place in the Institutional and Interbank markets, a significant portion - \$185 billion according to industry estimates - is traded by individuals and professional investors.

In both the retail and the wholesale FX sectors, user demands are changing as technology plays an ever greater role. From a retail perspective, traders are making more and more use of mobile and hand-held devices. And in the wholesale and Interbank market where electronification and the growth of high-speed algorithmic trading has had a significant impact, a number of market operators are now completely re-assessing their technology strategies and business models.

Performance of mobile trading apps

From a retail end-user perspective, there is no shortage of FX trading platforms that can now be accessed via mobile apps. Today, most retail FX brokers and trading venues offer some flavour of trading app, many of which include news, charts, position keeping, price updates and order entry/execution functionality. Often however, trying to trade on these apps can be a frustrating experience, as **Ash Gawthorp**, Technical Director of performance engineering specialist firm The Test People, explains

"Applications are often not written with performance on mobile networks in mind", he says.

"When it comes to mobile and desktop applications, the performance aspect of the trading app is not considered early enough in the development life cycle."

Ash Gawthorp, Technical Director at The Test People



"Developers often have high-spec desktop machines with fast network connections in to test systems, which positively skew their perception of actual end user performance. What works well on 1Gb Ethernet on a quad core workstation will not work anywhere near as well on a mobile data connection from half way around the world".

In Gawthorp's experience, when it comes to mobile and desktop applications, the performance aspect of the trading app is not considered early enough in the development life cycle.

¹Figures from the Bank for International Settlements.

“Non-functional requirements and performance targets should be baked in from the start, with regular architectural reviews undertaken and reviewed amongst peers to discuss messaging, UI design and reconnection strategy, mindful of the capabilities of the device and the network it is targeted for”, he says.

“Everyone has a responsibility for performance; it often either can't or shouldn't be 'fixed' by just one team, it requires collaboration between teams involved in architecture, development, test, infrastructure operations, networks and deployment. Everyone has a part to play and each team should understand the other team's challenges and requirements as often a performance challenge encountered in one team can be fixed or removed entirely by a team further upstream or downstream in the process.”, he continues.

Replicating the end-user experience

Gawthorp believes that platform providers need to take a different approach in order to gain a truer picture of the end-user experience, particularly where trading apps are rolled out internationally.

“They need to perform load testing in the cloud from locations around the world, providing a picture of how price latency increases with user load levels by location, additionally performing client driven performance tests on the actual client application than at a protocol level, so they are able to see the errors as observed by a user of the application. That is essentially the approach The Test People takes in order to get meaningful results”, he says.

One interesting observation, mobile data networks aren't just for mobile devices. Frequently for users in the Far East wired broadband is unavailable or prohibitively expensive, instead users adopt mobile data services for desktop internet connectivity - this creates a performance double-whammy of increased latency back to the client data centre often still in London or NY/NJ and increased latency from the user's ISP to the handset. To worsen the picture, data usage on mobile data contracts are often capped, an inefficient, bloated messaging protocol will further increase latency and chew through the client's data usage limits with every price update.

“We can then re-use these tests in a continuous framework in the live environment, monitoring key metrics such as price latency, login times and time to download the assets required for the application. This provides rich insights into how the system performs continuously, building visibility into patterns of poor performance and also providing the support teams with an early heads up of any issues”.

“All too often, client side performance and resiliency tests are performed as a one-off-activity or not at all conducted alongside server side scalability and load testing. This often shows that something worked and there wasn't a problem on a very short sample of data, often in a production-like environment but - crucially - not in the production over an elongated period of time”.

“Connections dropping is a big problem. On a mobile device in the real world it can happen all the time. So you need to be using the right middleware to communicate between your device and the server.”

Matt Barrett, Director at Adaptive Consulting



Dealing with dropped connections

Matt Barrett, Director at Adaptive Consulting, a firm specialising in front-office trading systems development, explains some of the additional technology challenges associated with mobile FX trading apps.

“Connections dropping is a big problem. On a mobile device in the real world it can happen all the time. So you need to be using the right middleware to communicate between your device and the server”, he says.

“You can't be polling for example, because if you drop the connection with the server, the costs associated with re-establishing the connection can be massive. A naïve platform

implementation will, on connection, re-acquire all the data it needs to have synchronized between the client and the server. That's far too expensive to do on a mobile device, when it might be happening almost constantly as you move around. So the client apps have to be smarter and you have to take a slightly more sophisticated approach, only sending what's changed rather than constantly informing about the current state of the world, which puts more pressure on your server and your infrastructure", says Barrett.

The introduction of multi-path TCP will go a long way towards addressing this challenge, according to Barrett.

"When fully deployed, this will allow a single TCP connection to move between different network interfaces (wifi, cellular) without the application using the connection to notice the difference. For FX trading apps on mobile devices this is obviously great because a user could start their session on wifi in the office, move out of range and have it seamlessly hop across to 4G, then move back to wifi when the user is back within range, all without any TCP reconnection and the associated re-synchronisation", explains Barrett.

Cross-platform development

Given that there are so many different devices now available for running FX trading apps, another key point to consider is the development effort on multiple client platforms, according to Barrett.

"There are recent advances in technology that allow the sharing of some level of code, which means that costs can also be shared across development for multiple platforms, whether that's thick client desktop development or IOS or Android mobile development", he says.

"Obviously the desktop is going to be significantly richer in functionality than the hand-held device. So reusing the backend services and the first tier of the client layer is something you can save a lot of money by doing, you just need to be aware that the user experience is going to differ, which isn't necessarily a bad thing. You shouldn't try to force 'one size fits all' across all your client platforms", he adds.

Satisfying the Professional and Institutional traders

In the Institutional (Funds, Asset Managers) and Interbank markets, where message traffic has increased exponentially in the last few years with the growth of high-speed electronic trading, trading venues face further technology challenges.

"We face the same sorts of issues that any large internet site faces, but we also have to try to make everything run as quickly as possible, in a few milliseconds."

Andy Phillips, Director of Technical Operations at LMAX Exchange



This is particularly the case for execution venues catering to clients from multiple market segments.

LMAX Exchange, the first regulated MTF (Multilateral Trading Facility) for FX, is uniquely positioned in the industry as it the only venue to service clients from all three market segments: Professional, Institutional and InterBank.

"We face the same sorts of issues that any large internet site faces", says **Andy Phillips**, Director of Technical Operations at LMAX Exchange.

"But we also have to try to make everything run as quickly as possible, in a few milliseconds. So we find ourselves in the position of having all of the challenges of a leading-edge exchange in terms of latency and quality of execution and thousands of orders per second flowing through our system, as well as trying to provide good quality of service for individual traders accessing us over the Internet".

Adopting an agile approach

To meet this challenge, central to the trade matching infrastructure of LMAX Exchange is the ground breaking 'Disruptor' concurrent programming framework, which was discussed in detail on The Trading Mesh back in August 2011 and which will no doubt be familiar to many technologists in the Java open source community. That said, Disruptor however is just one of several projects that has been open-sourced by LMAX Exchange over the years.

The development team use continuous integration and continuous delivery to release new software every two weeks. And the Agile process used for development is not just specific to building software. According to Phillips, it is used across the organization.

"We use the same Agile Kanban-type methods for running the operations teams, the network teams and so on, which means we're all speaking a common language, because there are occasions when you need to have an integrated approach across all departments", he says.

Dealing with growth in message volumes

Given the fact that end-users might be using a range of applications and different interfaces to access the market, how do venues ensure that the price users see on their screens or within their applications is the true, latest, current price in the market and that they are able to deal on those prices?

"That's a problem of any fast-moving market", says Phillips.

"At peak times, we can be processing 40,000 updates per second. So the sheer volume of price updates means that some customers are not able to deal on them because the market is moving too quickly. They might get the market data, make a decision to trade and send the order, but by the time the order reaches us, the market has moved. For people who are on latent connections, or who quite frankly can't deal with the volume of market data that we throw at them (which can be a function of latency and packet loss) as well the speed of their system, what looks to them like the slippage and rejects they would get trading on a last look liquidity are in fact just a consequence of the fact that it's a fast-moving market".

The MTF infrastructure has tools to monitor how much data clients can consume and has the ability to send out less data to certain clients and tune things, according to Phillips.

"It's the exchange execution experience that is the big differentiator for us. We're finding that a lot of customers who come to us stay with us. There's a real 'stickability' from customers because they haven't experienced elsewhere the kind of execution that we can give them once we've matched the data rate we send to the rate they can consume safely", he says.

Can randomization prevent 'gaming'?

With this kind of growth in message volumes, some FX trading platforms are now introducing artificial delays into their matching processes, in a bid to 'even the playing field' and remove opportunities for high-speed gaming.

"We've got a pilot scheme starting with a number of buy side participants which include some who use high frequency trading strategies, they're all really interested in this model."

Dan Marcus, CEO of ParFX



Last year EBS introduced a process whereby incoming orders are batched and assigned a random delay of between one and three milliseconds before being matched. Thomson Reuters recently announced proposals along similar lines.

Another venue adopting a similar approach is the recently launched ParFX, which applies a random pause to every order entering its system.

Dan Marcus, CEO of ParFX, explains how this works.

"Every single message that comes into ParFX - whether that's an order submission, cancel or replace - is subject to a randomized pause for any period between twenty and eighty milliseconds. We use this period because it's meaningful enough to nullify anyone who is aiming to game, but meaningless enough for anyone who wants to actually trade spot FX for hedging purposes."

According to Marcus, this is not necessarily an anti-HFT measure, but it does bring the focus back from pure speed-based to value-based plays.

"We've got a pilot scheme starting with a number of buy-side participants which include some who use high frequency trading strategies, they're all really interested in this model", he says.

"Does that mean they're going to try to game us? Not necessarily. Generally they'll have a bunch of strategies, only some of which are based on speed. Others are based on relative value, which are perfect for our platform as they can operate in a benign environment that rewards intelligence."

Many would assume that introducing an artificial delay would lead to wider spreads, but Marcus says this has not happened.

"We thought that might be the case as the vision is focussed on firmness of liquidity, i.e. what you see is not a mirage by virtue of disruptive HFT activity or duplicated on multiple platforms. But due to the high quality nature of the participants and the increasing network effects as we onboard we've seen tight spreads as well as excellent depth", he says.

IRS technology

The technology behind the ParFX trading platform is based upon Trad-X, Tradition's interest rate swaps (IRS) platform, which is designed to handle very large volumes of orders.

"We produce well over a billion messages a day on that platform, at around 80,000 messages per second. The high message traffic is due to the implied engine used on our CLOB which is required as IRSs are relative value products", says Marcus.

"IRSs are a bit confusing in that they are low latency in relation to message traffic but traded on a comparatively low volume basis with large size. We took that technology and developed it as the basis for our FX platform, so we had a huge amount of headroom for message traffic. Where it got quite complicated was the messaging around the randomizer, because it generates more messages, i.e. message going into the randomizer, message going out, message hitting the platform. That was a challenge".

Conclusion

For the individual FX trader accessing the markets via a smartphone or the multinational bank executing billions of dollars in FX transactions every day, performance is crucial. Every single participant in the market needs accurate prices and quality executions with minimal slippage.

The real challenge is how to achieve optimum performance at every level throughout design, development, testing and deployment. And the interesting thing here is that it is not necessarily all about trying to make everything go faster, as can be witnessed by the ParFX approach (and that of other major FX platforms) of slowing the market down with the aim of improving execution performance for the majority of participants, or LMAX Exchange's approach of tuning the levels of data sent to certain clients.

For more information on the companies mentioned in this article, visit:

<http://www.thetestpeople.com>

<http://weareadaptive.com>

<http://www.lmax.com>

<https://www.parfx.com>

<http://www.thetradingmesh.com>



Financial Markets Insights from The Realization Group, is a series of interviews with thought leaders in financial and capital markets. The purpose of the series is to provide exclusive insights into industry developments, through in-depth conversations with C-level executives and key experts from banks, exchanges, vendors and other firms within the financial markets ecosystem. For more information, please visit www.financialmarketsinsights.com

The Realization Group is a full service marketing and business development services company specialising in the capital markets. Our team contains industry practitioners from both the trading and post trade disciplines and we have expertise equally in the on-exchange and OTC trading environments. We apply our comprehensive set of marketing programs and wide-ranging media and business networks to complement the business development requirements of our client organisations.