Research

Eurex NTA Timings

06 June 2013

Dennis Lohfert

Introduction

Eurex introduced a new trading platform that represents a radical departure from its previous platform based on OpenVMS (Eurex Version 14.0, the final release on this technology stack). It's the somewhat poorly named "New Trading Architecture ", also known as "NTA" (One has to wonder what the next "new" version is going to be called...).
NTA was introduced via a staggered transition, with the least important products, e.g. weather, property and inflation derivatives, migrating first.
This first wave was released on 3 December 2012, with several other migrations (such as single stocks options) following throughout the first quarter of 2013.

Finally, the first real benchmark products have arrived on NTA. As of 8 May 2013, we have some meaningful production statistics for some highly liquid futures such as FDAX and FESX. Benchmark fixed income instruments (FGBx) have not transitioned yet and are due to deploy on 10 June.

Performance Improvements

We have analyzed our first performance measurements with regard to latency of the Eurex matching network: Gateways, Matching Hosts and Market Data publishers.

An inspection of 4 weeks of trading gives the following values for the FESX, unconditional, across the entire exchange:

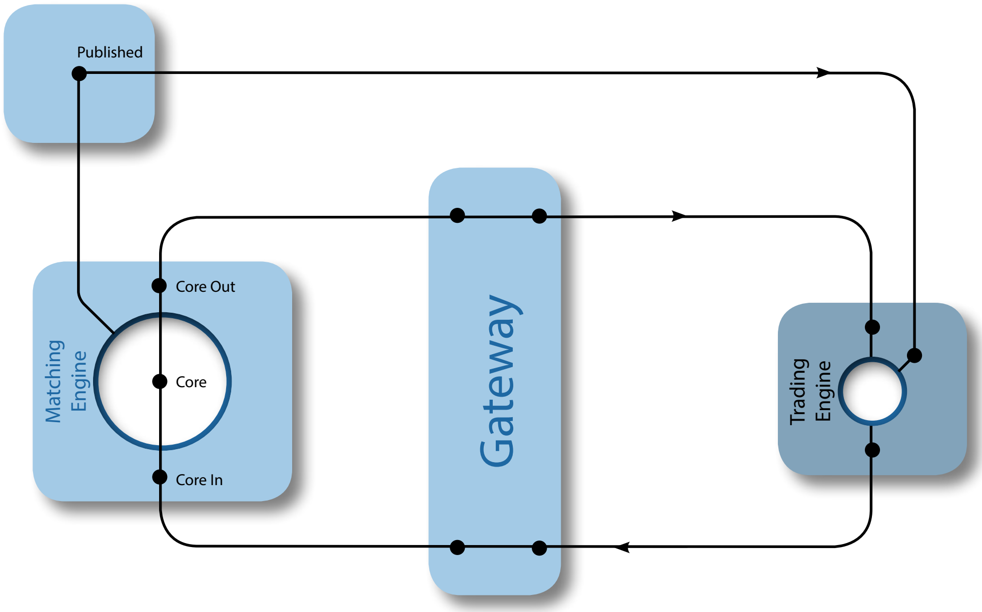
	Core RTT			Gateway RTT		
	Avg	50%	99%	Avg	50%	99%
Version 14.0	977.4	300.7	7,877.4	1,641.7	542.3	12,005.8
NTA Version 1.0	233.0	93.5	2,052.0	463.8	273.0	2,856.0
Ratio	4.2	3.2	3.8	3.5	2.0	4.2

timings are in microseconds

Depending on the criteria used for measurement, we can see that the processing times have been shortened by a factor of 3-4. This is a decent speed-up considering the complexity of the new system. As every system architect knows, increases of an order of magnitude are always hoped for, but almost never achieved. On top of this, while perhaps antiquated in terms of architecture, Eurex Version 14 was still considered a fast exchange by most participants, supporting an impressive daily throughput north of 100 million transactions, while its sister (Xetra, currently on Version 13) supports north of 200 million daily transactions. (Note that "transactions" in this document refers to messages, not to matched trades).

To analyze the various components we draw a network diagram of the various paths for dataflow between the exchange and our trading hosts.

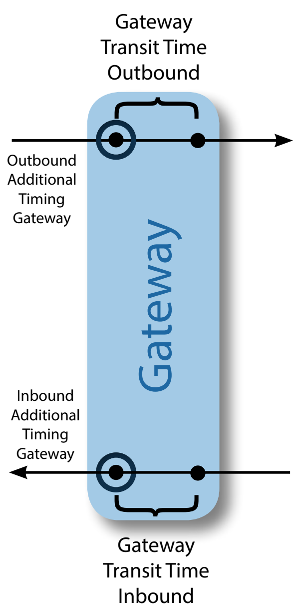
Figure 01: Simplified Exchange Message Route



We can directly observe the timing along the edges of the loop by inspecting message traffic.

In this simplified view, we are not paying attention to the various elements of redundancy that exist inside the exchange (such as A and B failover clusters) or the plethora of components that are at work out side of the matching engine (persistence systems, failover devices, clearing interfaces, etc.)

Figure 02: Gateway Timing



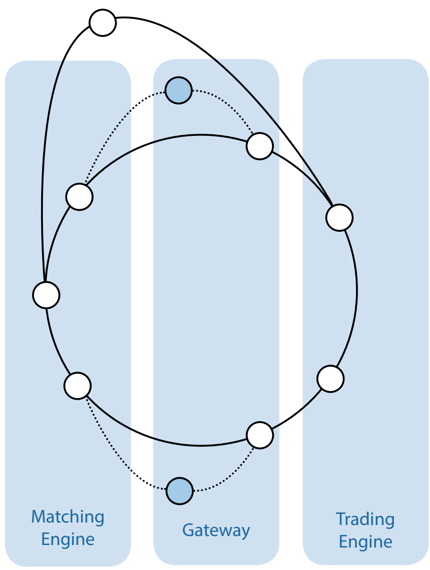
In addition to the timing points that are directly observable, there are two additional measuring points on the far side of the gateway (the side facing the matching engines).

These additional timing points give direct insight on how loaded the gateways are. These timings are not available publicly but they give an important indication of which Gateways are currently particularly busy.

However, we can reverse-engineer the transit times for this by modelling them as four unknown edges along the left side of loop (whereas the sum of the entire semi-loop is known).

Figure 03 shows the graph model we use to solve for the unknown timing elements in the latency graph.

Figure 03: Graph Model

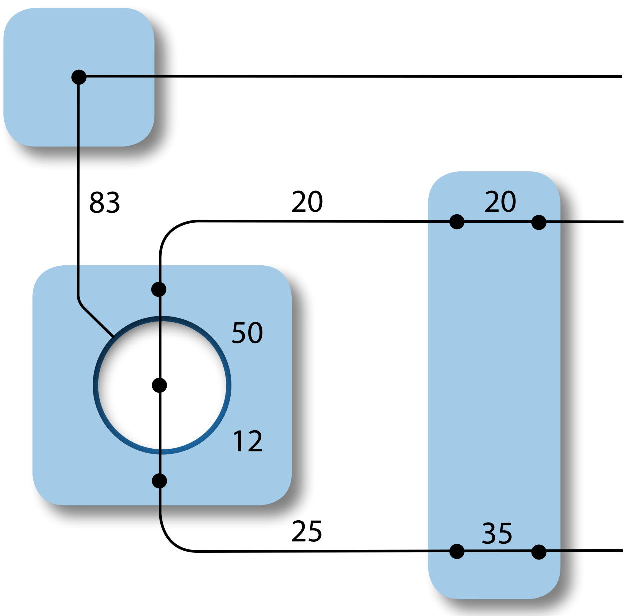


From a trader's perspective, the Gateway is the bottleneck of the entire process (the real bottleneck is of course the matching engine, but we have no control about what happens beyond the Gateway, so from our perspective the gateway is the de-facto bottleneck). The choice of gateway is important as sending an order to a Gateway that is busy can make the difference between getting to the matching engine in time or not (more on this later).

When analyzing network latencies in a system, it is important to establish a baseline latency in the idle state. We can look at various percentiles in the distribution or we can look at minimum transit times along the edges. Usually the median is a good choice as it will resemble a "typical" transaction, but the percentile (of which the median is a specific case) is not a linear operator. Averages however are linear, so when we compare the averages along edges of the graph, we at least get the same result as for the entire route, which aids intuition.

As we look at the base latencies along each segment of the loop, we can approximate a bare minimum time required to get in (and back out).

Figure 04: Bare Minimum Transit Times



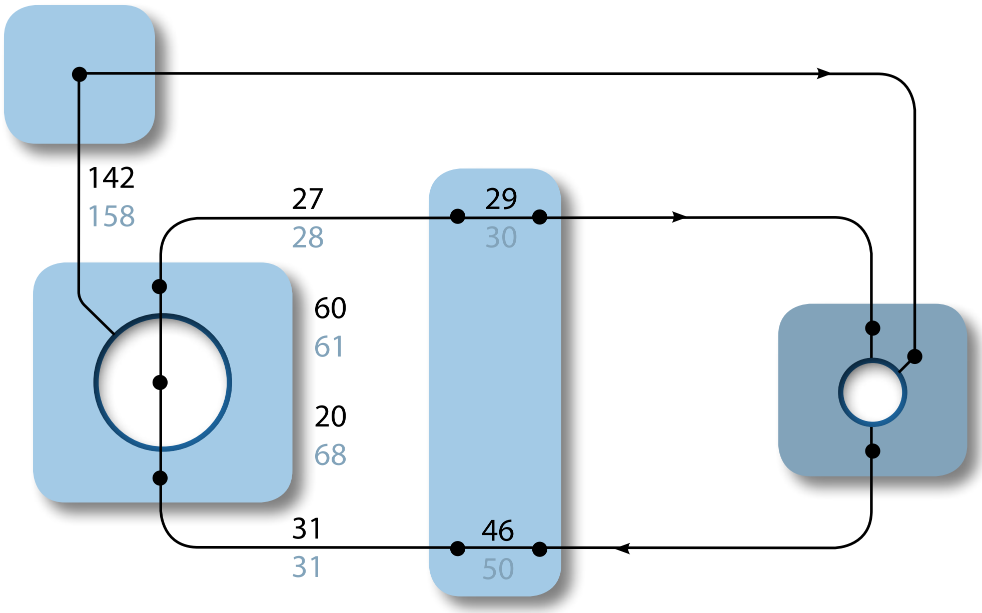
For inbound messages:
35+25+12 = 72 µs
For outbound messages:
50+20+20 = 90 µs

Note that the 83µs number for market data publication is for orderbook updates only. If a trade is generated, the outbound messages take at least twice as long (for trades that involve many orders this can be many hundreds of microseconds).

Essentially we are looking at a bare minimum of, say, 160 microseconds. Now the reality is, one will almost never see this number because the system just does not offer this degree of responsiveness at every point all the time.

Of course, this inspection is no more than a cursory glance, unconditional on any specific events such as trades being executed or the orderbook shifting significantly. And as all high-frequency traders will tell you, the times when "interesting" things happen is what actually really matters. Having established the baseline above, we can start looking at other latency numbers. Particularly, the distribution of messages passing through the gateways into the core is interesting, as this is ultimately the deciding factor for many opportunities.

Figure 05: Median and Average Latencies On Exchange Side



One of the important factors for how suitable an exchange is for high-frequency trading is how quickly we can act on new information. In particular, how quickly we can react to new significant information (which is when everybody wants to act and the nature of competition really starts to kick in).

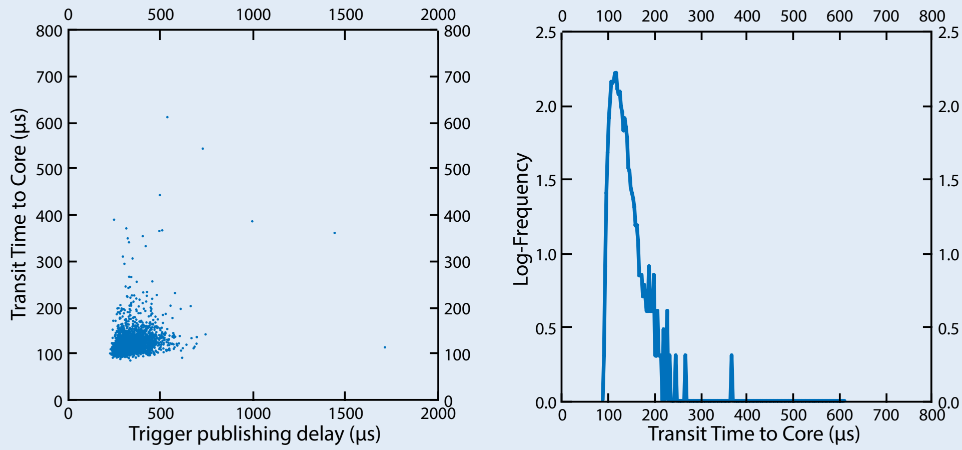
To that end, it is helpful to examine two points:

- The distributions of waiting times going into the matching engine (shown in Figure 06 on the right hand side)
- The distribution of waiting times at the Gateways (shown in Figure 07)

Beginning with the matching engine, there is a fat right-tail. What matters is when this fat right-tail occurs. If we correlate the queueing time to how long the trigger event took to process in the matching engine (an indication of trade complexity), we arrive at the essence of low-latency trading:
If information arriving is really interesting, everyone will start sending orders (building up a queue for everyone but the fastest).

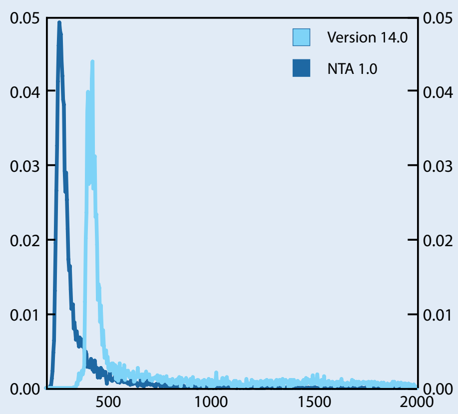
We can see that there is some impact (again, sampled unconditionally). The lower rising edge of the cloud is sure sign of correlated impact and something we need to consider when making trading decisions: how long will we potentially be queueing and is this delay going to cause risk (think: legging risk on inter-market spreads, for example)?

Figure 06: Queueing on Core



Looking at the overall responsiveness of the system, we can see that latency (as measured by Gateway RTT) has been improved by at least a factor of 2.0 (as measured by median) or a factor of 3.5 (as measured by the mean). We refer back to the table at the beginning of this note. The distribution of Gateway RTTs is also shown in Figure 07 below.

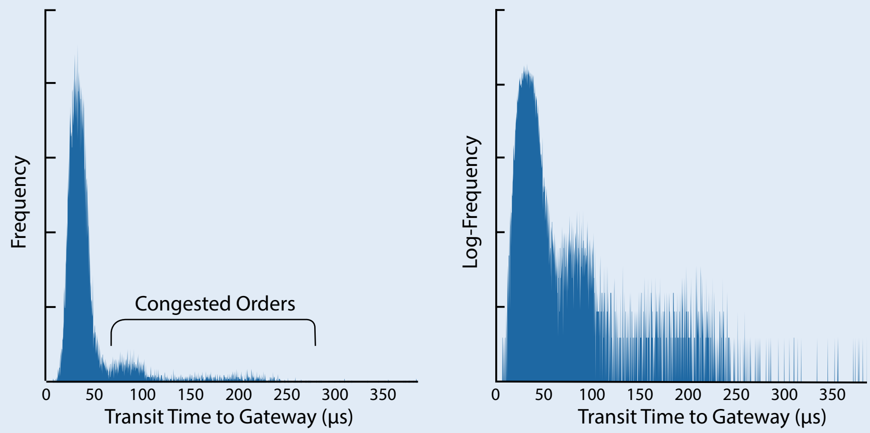
Figure 07: Responsiveness of Exchange System



The matching engine is one component, and it seems that the new edition is, in fact, very predictable. What about the gateways? From the perspective of the trader, where the bottle neck is. Eurex has tried to cater to different needs by splitting the gateways into High-Frequency Gateways and Low-Frequency Gateways (22 in total: 16 HF and 6 LF. However, Eurex is currently looking at possibly reducing the number of HF Gateways).

Here we see a significant queueing effect (see Figure 08 below). At times, there are orders that congest up to 10 times the average waiting time. This is the part that every high frequency trader complains about because it means that our orders had to wait before slipping into the Gateway. Note that these are unconditional waiting times. When things get interesting, the queue can get very long indeed and that is when latency matters most (unfortunately we cannot show these here, for obvious reasons).

Figure 08: Queueing on Gateway



Summary

The new architecture is significantly faster than the old one (by about a factor of 4). More to the point, Deutsche Börse likely has to room to optimize. Over the next year, I think its reasonable to expect them to reduce the Version 1.0 timings by 25% (they are already talking about an exfpected 20% reduction in Version 1.1, which will be hitting rack shelves in July).

How much difference this will actually make to P/L of course depends first and foremost on the trading style and what adjustment traders make to their platform. Users of commercial 3rd party solutions to conduct trading will have more problems as these tend to be slower and so the gap widens. For the market place it is good news overall however, as faster response times mean a better handle on risk (execution uncertainty), which in turn reduces the price for liquidity.