Load Test Report

Moscow Exchange Trading & Clearing Systems

30 March 2019

Contents

Testing objectives

1. To verify the trading and clearing systems operation under conditions of peak loading and an increased number of orders and trades. The trading systems of the following Moscow Exchange's markets were tested:

    a. The Equity & Bond Market;

    b. The FX Market;

    c. The Derivatives Market.

2. To estimate the time of order filling and data delivery from the trading and clearing systems at different load levels and software and hardware configurations.

3. To carry out a public testing of a new Equity & Bond market system with independent trading and clearing engines.

4. To allow third party software developers and brokers to test their systems and estimate the throughput capacity of communication channels to the exchange venues.

List of participants

In accordance with the Moscow Exchange IT committee's recommendation, the list of load testing participants will be published on exchange's website at moex.com.

Main results

The Equity & Bond Market trading and clearing system

The testing was carried out on the new version of the system with independent trading and clearing engines that is scheduled to go live in 2019. This new version is 100% backwards compatible with the current production version but requires separate client connections to be established for trading and clearing data and transactions. Server configuration used during the tests was the same as it was expected to be in production. No changes in hardware are planned.

The table below shows comparative performance in testing in 2017. The 'accepted transactions' term means all the incoming transactions that led either to order registration or to successful cancelation of an order.

	Transactions	Orders	Trades
Values reached (units), 2019	124 527 476	64 906 598	1 381 905
Values reached (units), 2017	108 550 150	60 733 370	1 735 735
Max processing rate for accepted transactions (units per sec), 2019	62796	31913	3289
Max processing rate for accepted transactions (units per sec), 2017	31500	16405	1356
Performance growth 2019 to 2017	**+99%**	**+98%**	

The graph below shows the frequency of transactions, orders and transactions by clients – testing participants. The maximum system performance level was not reached during the load test.
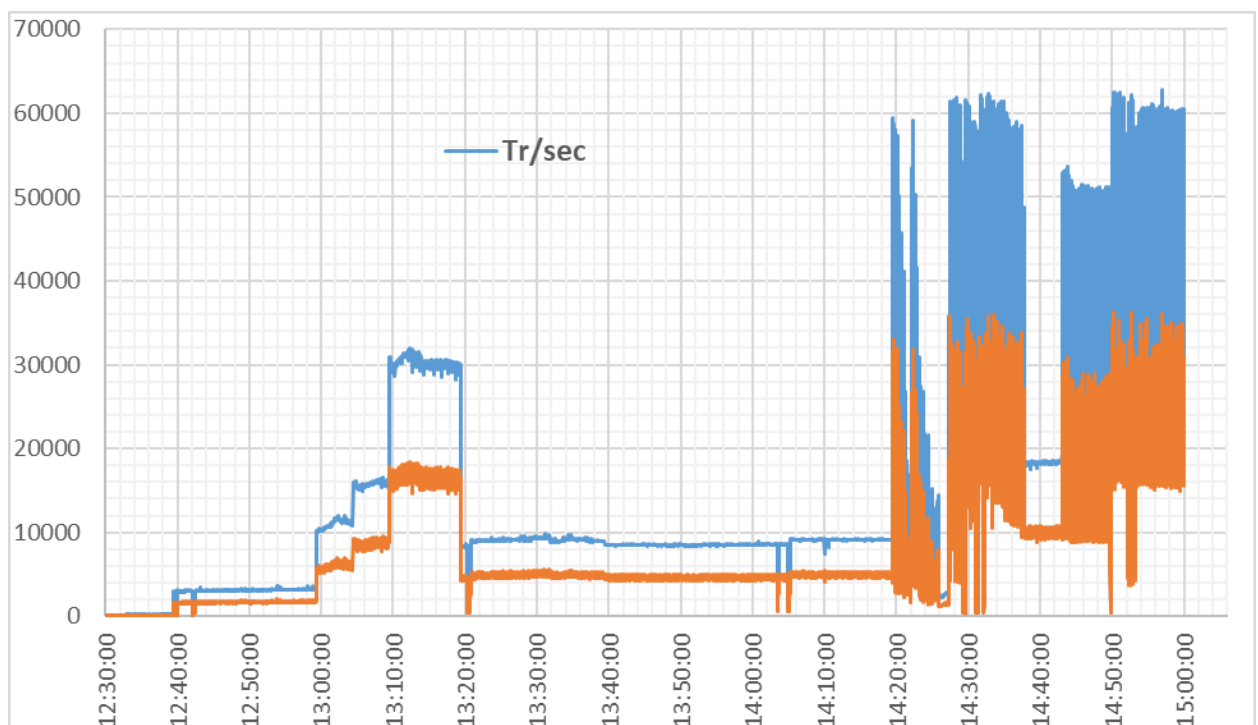
The share of client generated activity was 4.8% of all transactions in 2019, that is significally lower than their share in 2017 (9.3%).

From 14.20-15.00, the transaction generator emited transactions with variable frequences. The average speed of transaction flow was more than than 30,000 tran/sec, up to five times higher than the actual production activity peaks (up to 6,000 tran/sec and 99.9% of one-second intervals with activity less than 2,300 tran/sec).

The peak frequency was generated in the repetitive intervals with durations of approximatily 1-2 seconds. Unlike the permanent peak load, this scenario allows to measure recovery time needed for some components of the trading system whise ability to function normally was interrupted by peak load. A stress nature of the load testing was remained.

### The FX Market trading and clearing system

The testing was carried out on the new version of the system with several independent clearing engines that is scheduled to go live in 20 April 2019. This new version is 100% backwards compatible with the current production. Server configuration used during the tests was similar to that expected to be in production.

The table below shows comparative performance between the current production version and the release candidate. The 'accepted transactions' term means all the incoming transactions that led either to order registration or to successful cancelation of an order.

	Transactions	Orders	Trades
Values reached (units), 2019	129 787 106	70 068 396	886128
Values reached (units), 2017	135 906 705	72 683 460	1 858 811
Max processing rate for accepted transactions (units per sec), 2019	69924	34442	1185
Max processing rate for accepted transactions (units per sec), production configuration	46500	24000	
Performance improvement, 2019 / 2017, %	**+50**%	**+43%**	

The graph below shows the frequency of transactions, orders and transactions by clients – testing participants. The maximum system performance level was not reached during the load test.
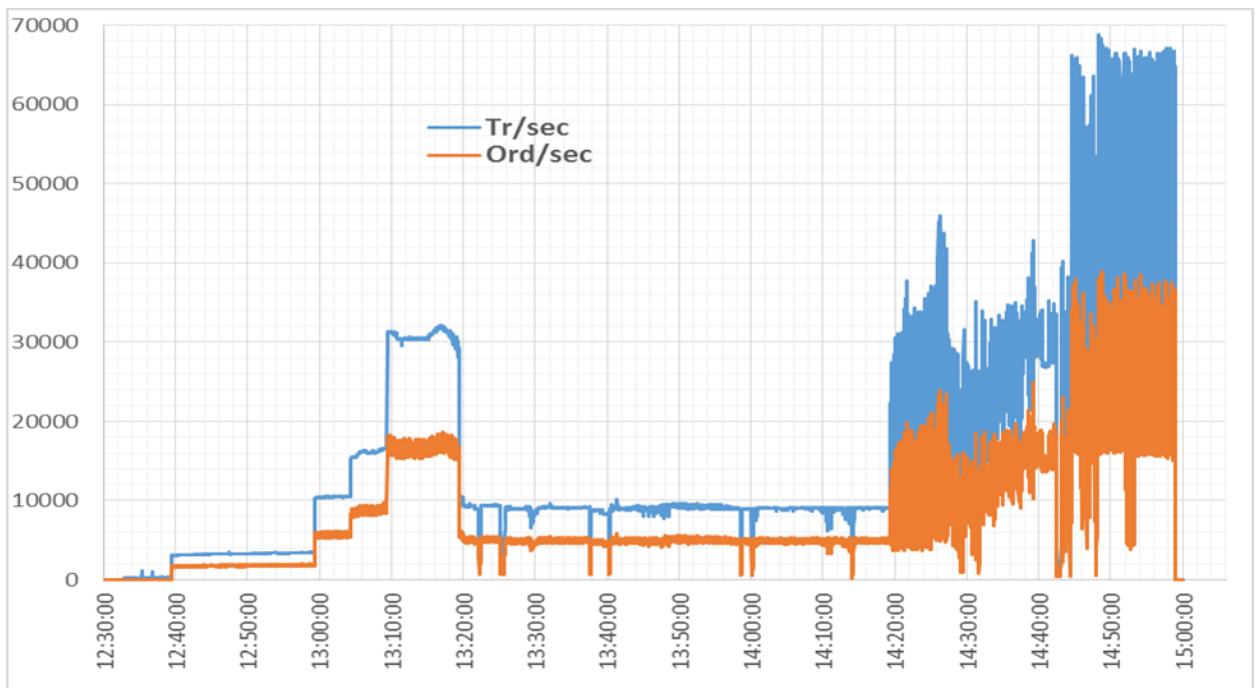
The share of client generated activity was 5.4% of all transactions that is significally lower than the client share in 2017 (8.2%).

From 14.20 to 15.00, the transaction generator emited transactions with variable frequences. The average speed of transaction flow was more than 30,000 tran/sec, up to five times higher than the actual production activity peaks (up to 6,000 tran/sec and 99.9% of one-second intervals with activity less than 3,000 tran/sec).

The peak frequency was generated in the repetitive intervals with durations of approximatily 1-2 seconds. Unlike the permanent peak load, this scenario allows to measure recovery time needed for some components of the trading system whose ability to function normally was interrupted by peak load. A stress nature of the load testing was remained.

The Derivatives Market trading and clearing system

The testing was carried out on the SPECTRA system version 6.2 used in production since 2 March 2019 on the servers of the main DSP data center.

The peculiarity of the given test scenario was a capacity test of trading system. The test was needed to obtain data on quality (such as speed and resilience) of the trading system in conditions of speedy retail business growth and significant number of clients' account registration. This direction is a priority for the market and it will be supported by online clients' registration functionality that is scheduled to go live in August 2019.

Over 4,000,000 client accounts were added to the trading system before load testing start with new positions added on those accounts during the load test.

The capacity test was done successfully that shows readiness of the trading system to function in conditions of significant growth of registered client accounts.

The ratio and load profile of transactions submitted over the different protocols was similar to the production.

The order-to-trade ratio in the testing was also near the production value. 267 million transactions were sent and 7.1 million trades were executed during the testing. The peak performance was 167,000 transaction/sec with the stable trading system performance at average speed of around 130,000 transaction/sec.

	Transactions	Orders	Trades
Values reached (units), 2017	294 790 508	159 317 680	12 653 800
Values reached (units), 2019	267 744 181	135 000 000	7 100 000
Max processing rate (units per sec), 2017	128 000	-	-
Max processing rate (units per sec), 2019	167 000	-	-

During the testing, we carried out the scheduled intraday clearing session. Despite large volumes of orders and trades, clearing was performed as usual within the established time frames.

The graphs below shows a transaction load on the Derivatives Market trading system. Clients participating in testing generated 2.58% of the transactions.
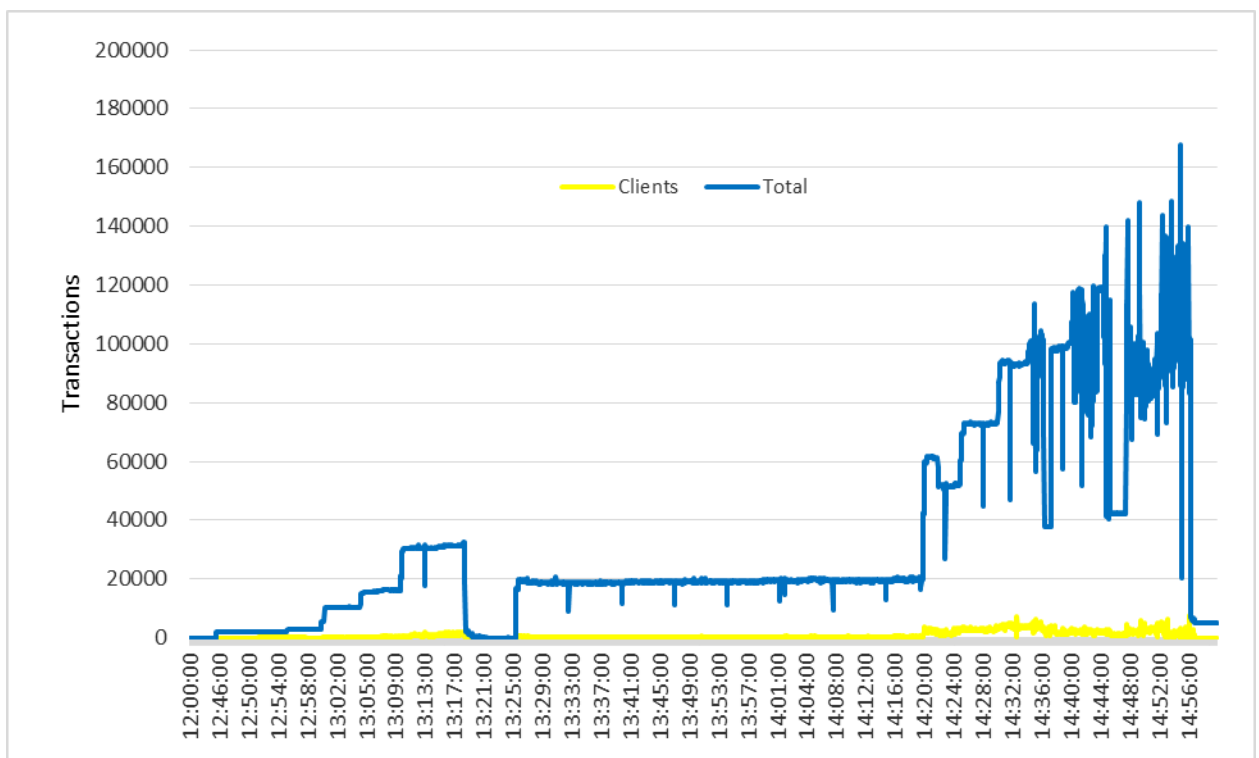
ASTS Gateways

Equity market ASTS gateways as well as gateways of the FX market installed at both DSP and M1 data centers were functioning correctly.

During the normal operation of the Equity and FX markets systems, the clearing gateways will only be available at the main DSP data center. The clearing engine located at the backup facility works in a slave mode and will only serve clearing gateways in case of the main data center failure and migration of the trading engine to the backup facility. This scenario was not included into the load test program.

The clearing gateways were functioning correctly at a constant transaction flow rate of up to 45,000 transactions per second. When this threshold was being exceeded, there were delays in clearing data refreshes at gateways as compared to the main clearing engine. In real trading, the peak frequencies of more than 45,000 transactions per second happen only in short bursts, so the expected delays will not exceed 5 milliseconds.

The FIX gateways configuration for both FX and Equity markets was similar to its production version. The gateways functioned normally across the whole range of transaction frequencies.

SPECTRA Gateways

During the load testing, no deviations from the normal gateways performance was seen. The FIX/Twime gateways configuration was similar to its production version. The gateways functioned normally across the whole range of transaction frequencies. The DR scenario was not included into load test program.

ASTS remote Gateways

The ASTS trading platform employs gateways operating in regional technical centers.

To ensure proper operation of gateway servers on each market, at least 8 Mbit/s of network channel bandwidth is required for every gateway software instance of every market.

When the constant transaction frequency exceeded 15,000 transactions per second, some remote gateways started delaying data tables if hardware specifications and/or data channel throughput were not sufficient to sustain increased transaction load. The project of updating technical facilities in regional data centers will allow fixing that inconsistency. .

Latency for transactions

Equity & Bond Market and FX Market trading systems

Transaction generators at the Exchange side were using Linux version of the embedded ASTS Bridge (libmtesrl.so library) or FIX protocol. These generators had been set up on a server connected to the trading network. Latency data for the FIX protocol for Equity and FX Markets was collected using the network monitoring system based on Corvil solution and customized for ASTS FIX messages.

Below is the screen that shows statistic data on replies to FIX transactions for the Equity and FX Markets. Every point on the graph shows the median latencies for approximately 70,000 incoming FIX transactions.
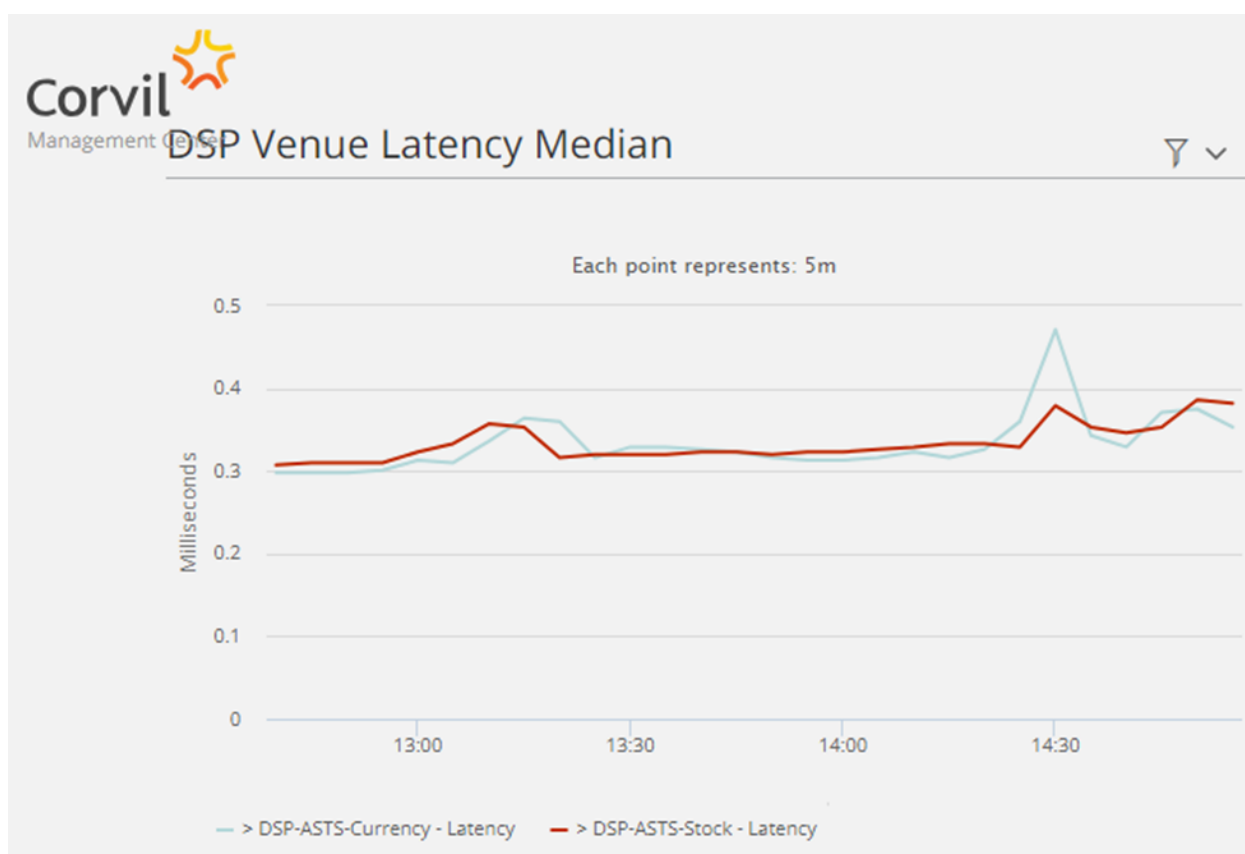
Within the range from 1,000 to 60,000 transactions per second, the average reply time was changing from 300 mks to 375 mks. The peak latency of 470 mks at 14:35 was induced by abnormal frequencies of trades from the load generator (over 1,200 per second).

Derivatives Market trading system

The internal Spectra monitoring system, Corvil solution and log files for client transactions have been used to measure the Derivatives Market latency.
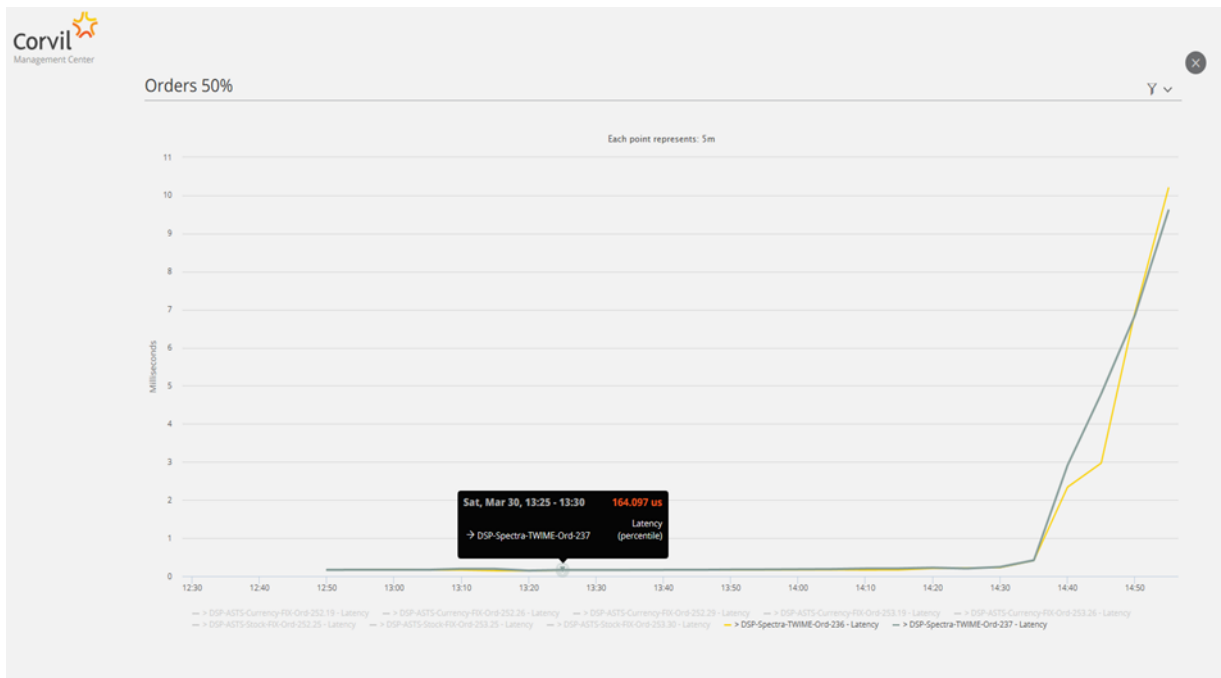

Within the range from 1,000 to 100,000 transactions per second, the average RTT for TWIME gate was changing from 150 mks to 200 mks. However, RTT's growth for the frequencies of over 100,000 was induced by the rising load on the trading system core.

FIX/FAST UDP multicast marketdata of the Equity & Bond Market and FX Market

Equity and FX markets FAST servers configuration was the same as in production. The servers operated correctly at all load levels.

Statistical data on new order messages from the trading system (OLR feed, 279=0) relative to FIX protocol order accepted messages (Execution Report / 150=0) has been collected using the Corvil equipment. This data is being constantly collected during the normal trading.

For the FX Market, the average time of publication in the FAST feeds is shown in the screen below for the primary (225) and secondary (228) publishing lines.
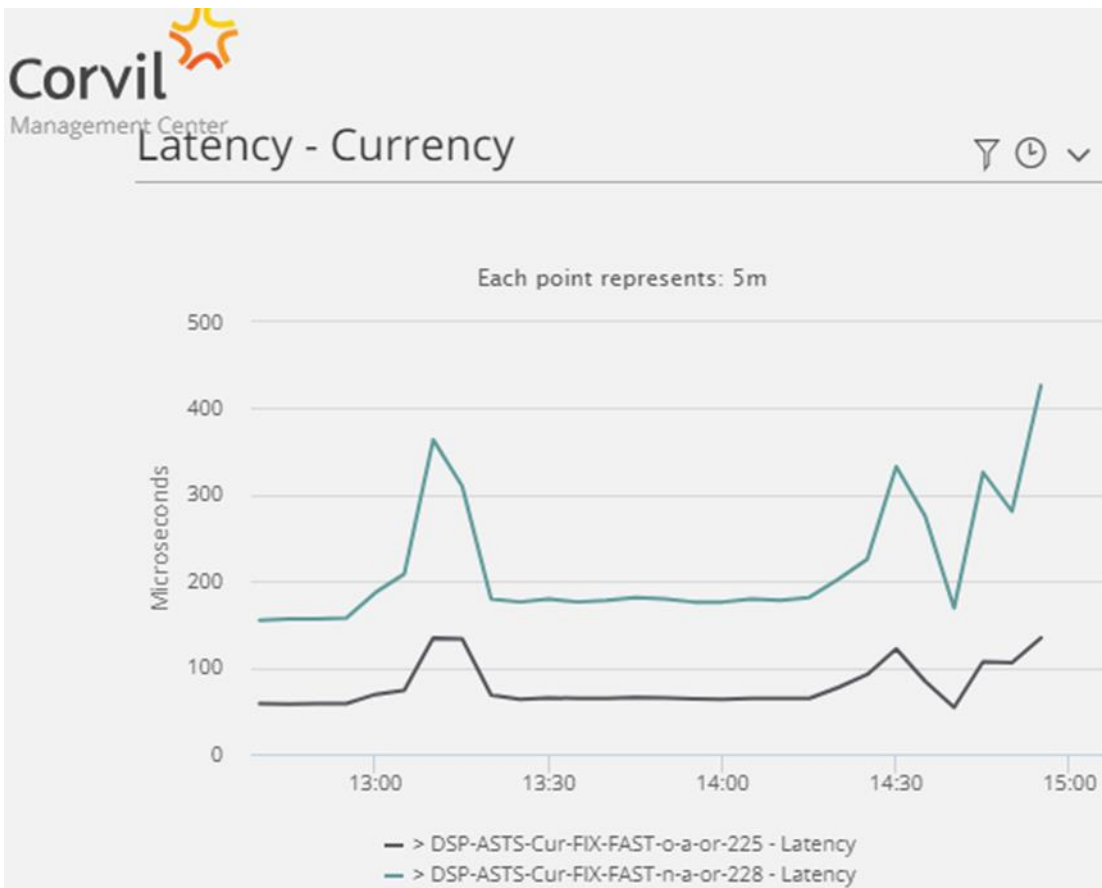
At maximum transaction frequencies, the average relative publication times increase to 150 and 450 microseconds, respectively.

For the Equities Market, the average time of publication in the FAST feeds is shown in the screen below for the primary (227) and secondary (229) publishing lines.
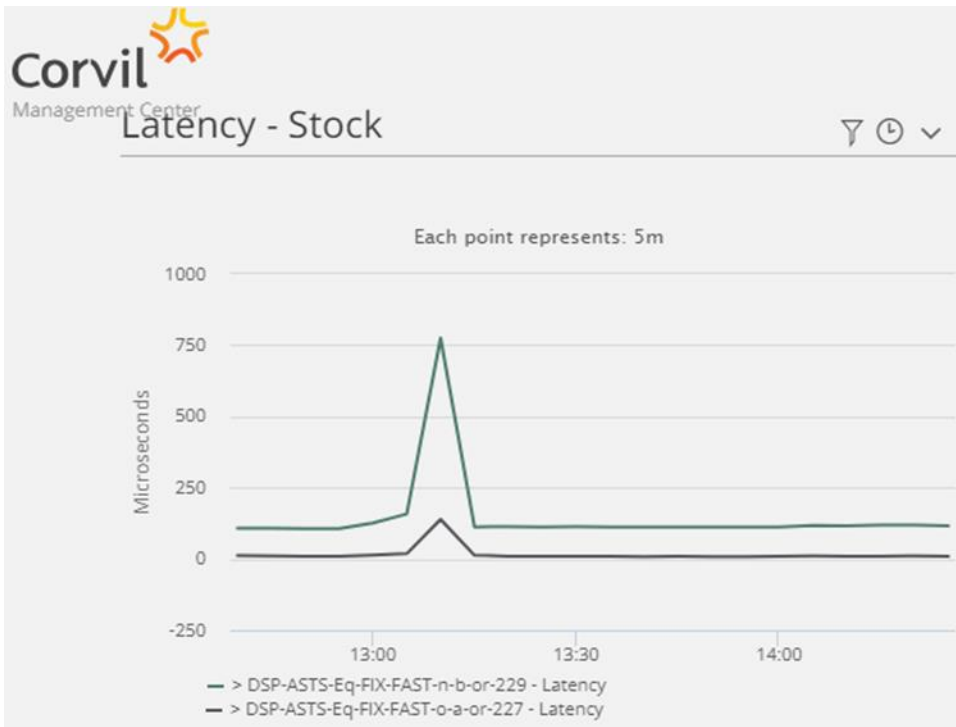
FAST publication latency increased significantly as transaction frequencies were approaching peak values.
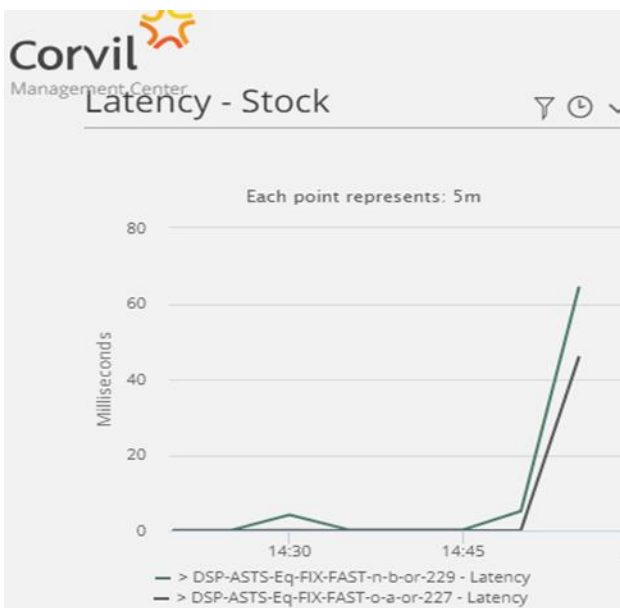
After the production launch of the new Equities Market trading and clearing systems, adjustment procedures for FAST servers, similar to those applied on the FX Market, will be needed to exclude abnormal growth of publishing latencies. Median publishing latencies within the range of transaction frequencies of up to 30,000 transactions per second are shown in the table below.

Server	Market	FIX Execution Report => FAST update delay, mks
Line1 (*227)	Equity	10
Line1 (*225)	FX	65
Line2 (*229)	Equity	112
Line2 (*228)	FX	180

UDP multicast traffic for the FX market reached the following values in each of copies A and B:

Feed	FX market, Mbps	Equity market, Mbps
Active orders (OLR)	18	14
Market statistics (MSR)	27	20
other feeds	2	2

Participants connecting to the service via data distribution channels are recommended to carefully select their subscriptions to data feeds and consider the network bandwidth as the total combined traffic of the two FAST lines of the FX Market and Equity& Bond Market in copies A and B may reach 500 Mbps.

In the production environment, the short FAST traffic bursts would most probably correspond to the network requirements stated above.

Recommendations given at http://www.moex.com/a1160 are applicable to each FAST line.
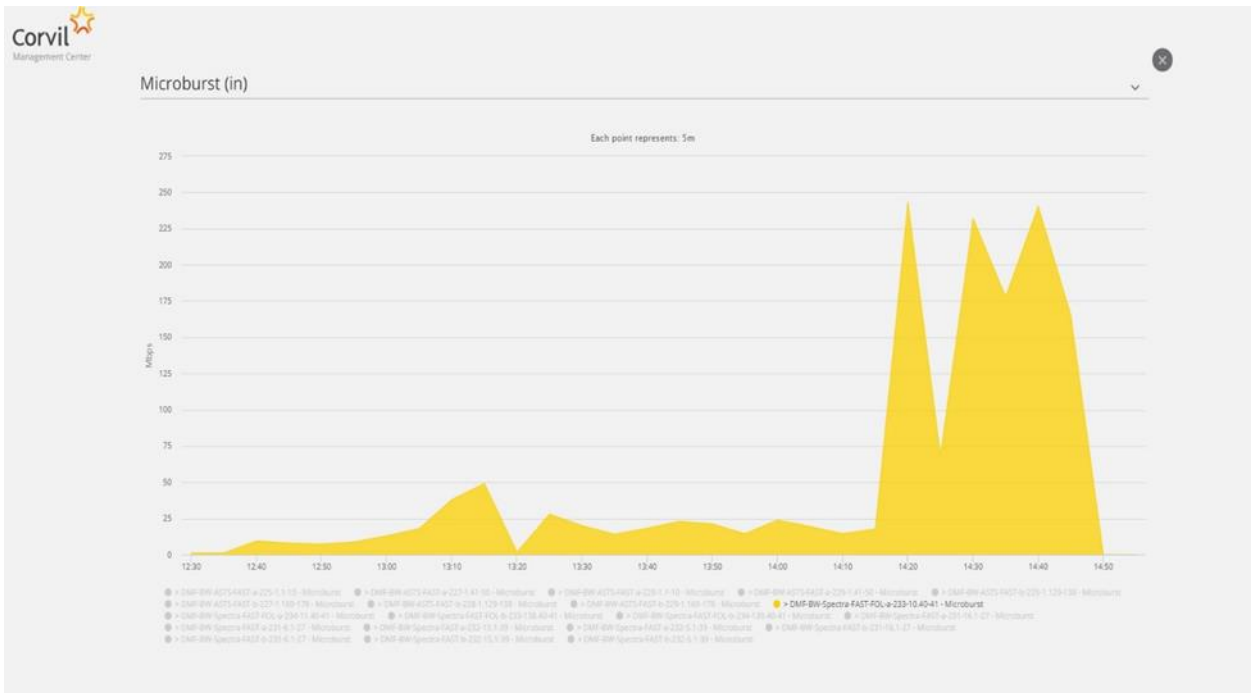
FAST UDP multicast marketdata servers of the Derivatives Market

The Derivatives Market FAST servers configuration was the same as in production. The servers operated correctly at all load levels.

Statistical data on new order messages from the trading system (Full order log) relative to TWIME protocol order accepted messages has been collected using the Corvil equipment. This data is being constantly collected during the normal trading.

The average time of publication in the FAST feeds is shown in the screen below (the intermission on the graph was caused by the intermediate clearing session).

Median publishing latencies at transaction frequency of up to 40,000 transactions per second were 10-20 mks; at greater transaction frequency, they grew to 50-60 mks.

Latency Median- Feed A

UDP multicast traffic for the Derivatives Market reached 40 mbps in each of copies A and B with bursts of up to 250 mbps.

Bandwidth (in)

The required network bandwidth for clients who wish to use the FAST service to receive ORDERS-LOG is minimum 100 Mbit/sec per feed. To receive two feeds, FEED A and FEED B, or data from more than one market, the 1-10 GBit/sec bandwidth is recommended.

Exchange network and colocation network parameters

The network monitoring system within the exchange perimeter and within the colocation zone perimeter (including the colocation zone core switches) showed that the network parameters had no deviation from normal work parameters. No retransmissions, packet loss or network latency growth were detected.

Subsystem for real time monitoring of the trading system parameters and market activity

Monitoring for the maintenance divisions and technical support. The monitoring facilities operated well and provided data visualization in graphic form. Message signals were produced in accordance with the established criteria; data was collected to the monitoring database without fails. Operation of the monitoring system did not influence the facility performance.

Corvil network monitoring system. During the testing, the Corvil (www.corvil.com) equipment and software, which had been adapted to analyze trading systems network traffic for all markets has been actively used.

The network timing monitoring system proved to be tolerant to the stress load:

- it did not fail;

- events have been detected correctly with no distortion;

- statistics were shown in real time;

- after the end of the testing, the standard web interface has allowed to successfully export tens of gigabytes of the collected data for further analysis.

Monitoring system overload in real trading is impossible.

Index server, market maker server, MOEX web site

All the listed systems have operated correctly, no failures or performance problems have been noted.

Test of clock synchronization over PTP (precision time protocol) at stress load

For the both data centers, an infrastructure for synchronizing system clocks over a high precision PTP protocol is deployed. During the tests, its stability at unusually high network loads on the Exchange infrastructure has been checked.

Precision checks for clocks on the network devices have shown that the time deviation has been no more than 500 nanoseconds and that is considered to be the excellent result. No failures of synchronization on the network devices or servers have been noted.

Conclusions

The Equity & Bond Market, the FX Market

1. Performance of the new Equities Mmarket system with independent trading and clearing engines is 98% higher compared to the current production version.

2. Performance of the new FX Market system with several independent clearing partitions is 45% higher compared to the current production version.

3. No failures of the system components caused by program errors or overload during this testing have been registered.

Derivatives Market

1. SPECTRA's performance is sufficient to meet demands of participants even at peak loads with respect to order processing and market data distribution.

2. The bandwidth requirements for customers who use Plaza 2 protocol remain unchanged compared to the previous year:

    a. At least 4 Mbps is required for stable operation of client bridges/terminals per each software instance.

    b. At least 10 Mbps for a bridge in case of using feeds with a full order/trade log (FORTS_ORDLOG_REPL/FORTS_DEALS_REPL).

3. We strongly recommend to check not just the available network bandwidth, but also the quality of communication channels out of the Exchange as well. Channels with high packet losses gravely affect latency and may lead to significant delays in data availability.

4. The required network bandwidth for clients who wish to use the FAST service to receive ORDERS-LOG is minimum 100 Mbit/sec per feed. To receive two feeds, FEED A and FEED B, or data from more than one market, the 1-10 GBit/sec bandwidth is recommended.