Some insights into the details that matter for high-frequency trading!

Eurex Exchange's T7

November 2013

W. Eholzer – Head of Trading IT

Agenda

- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

Eurex Group Technology Roadmap

In 2006 Eurex Group initiated its Technology Roadmap to deliver innovative and superior technology.

This initiative includes:
- Provision of a portfolio of interfaces that meet the needs of different user groups
- High throughput and low latency of the trading system
- Delivery of functionality the market demands with shortened lead time
- Exceptional level of transparency & customer service

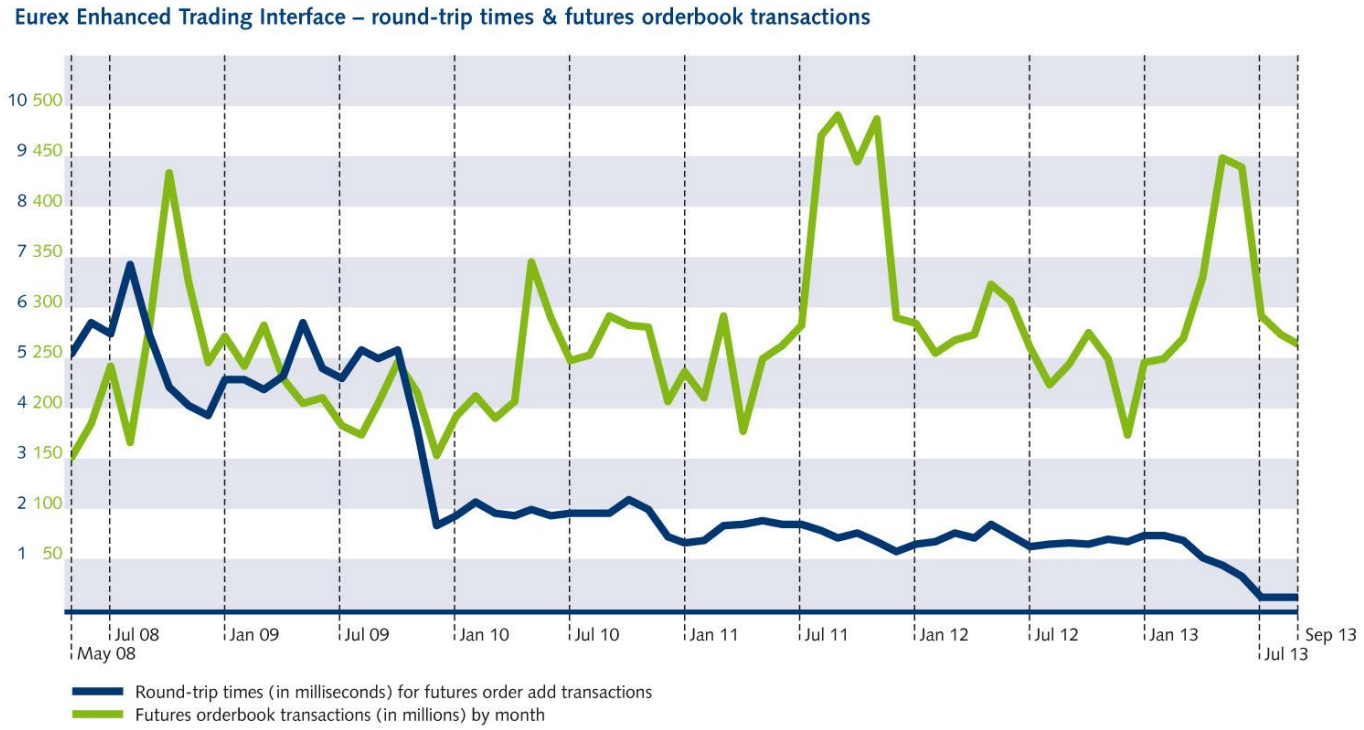
The Eurex Group Technology Roadmap will continue in 2013 on the trading layer with the launch of T7 release 2.0 on 25. November.

Eurex Exchange's T7 is based on Deutsche Börse Group's proprietary global trading architecture, which is also in use at the International Securities Exchange (ISE).

For further details about T7 please have a look at our web page: www.eurexchange.com/t7

This presentation aims at providing information about T7 to latency sensitive users.

Futures order add round-trip times

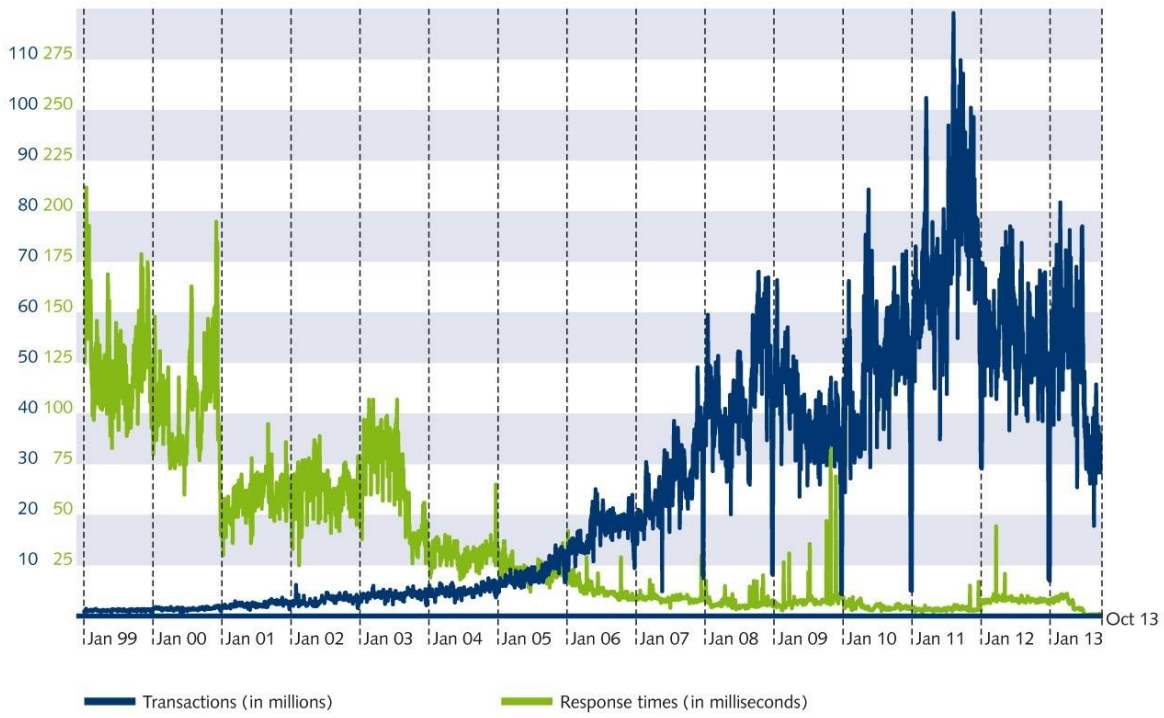


Futures order add latency has been significantly reduced to a level of a daily average of about 0.2 – 0.35ms for users of the Eurex Enhanced Trading Interface located in the Equinix data center in Frankfurt.

** From 1 July 2013 only T7 transactions have been considered.*

Processed transactions & response times

Number of processed transactions at Eurex Exchange & response times



Eurex Group has continuously invested in its trading system and has been able to reduce the processing time of technical transactions significantly although the daily load on the system has grown extremely*.

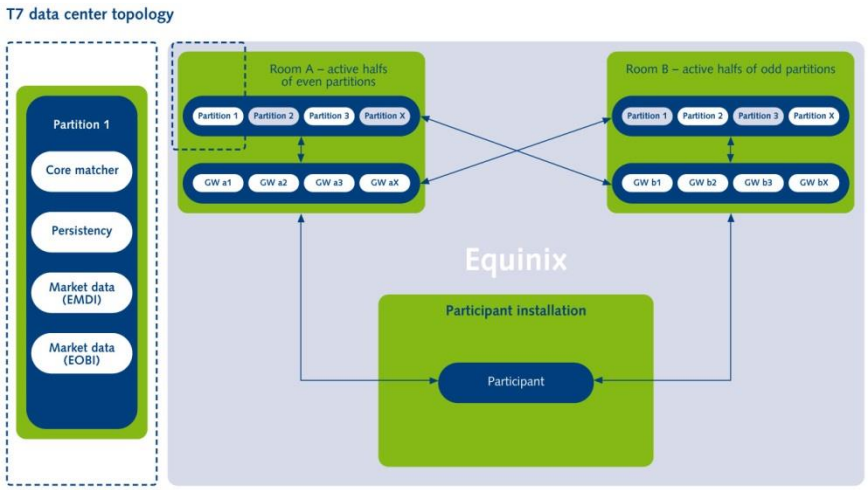
** From 1 July 2013 only T7 transactions have been considered.*

Agenda

- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

Overview of T7

- T7 consists of partitions. Here, a partition is a failure domain in charge of matching, persisting and producing market data for a subset of products*. Each T7 partition is distributed over two rooms in the Equinix data centre.
- There are 10 T7 partitions in charge of futures and options trading available on Eurex Exchange. A separate additional partition is planned to be used for European Energy Exchange products.
- The high-frequency Eurex Enhanced Trading Interface gateways in the Equinix facility provide the fastest access to T7 **.
- There are 16 such gateways in the Equinix data centre shared by all Trading Participants of Eurex Exchange. Note that it currently is not planned to change the number of high-frequency Eurex Enhanced Trading Interface gateways.
- The reference data contains the mapping of products to partition IDs. The physical location of the high-frequency Eurex Enhanced Trading Interface gateways in the Equinix data centre relative to the room where a matching engine resides has no impact to the order latency.
- Note that normally the active half of a partition is either in room A (for even partitions) or in room B (for odd partitions).
- Only in case of the failure of a matching engine or a market data publisher, the active half of the service will shift to the other room.



** Note that the product to partition mapping is expected to typically change on a time scale of at least weeks/month. Even in case of e.g. a hardware failure, the mapping will not change intra-day.*

*** The additional 6 low-frequency gateways have a base round-trip latency that is about 100 usec higher than high-frequency gateway and a higher variance in processing times.*

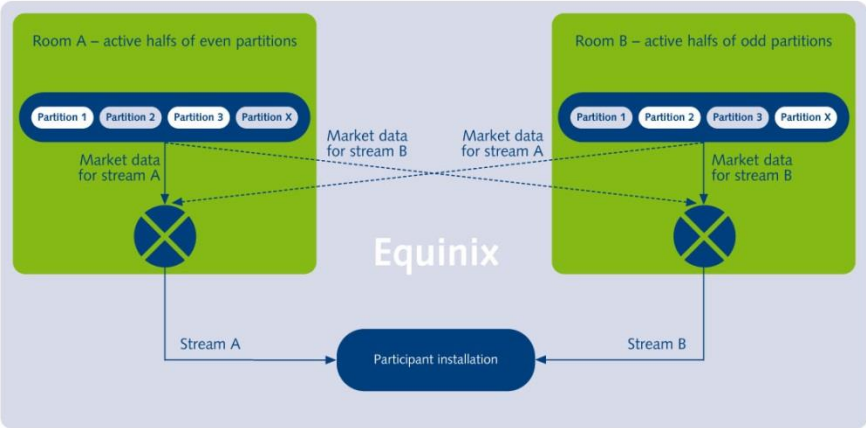
Middleware, network, hardware and OS overview

T7 uses state-of-the-art infrastructure components:

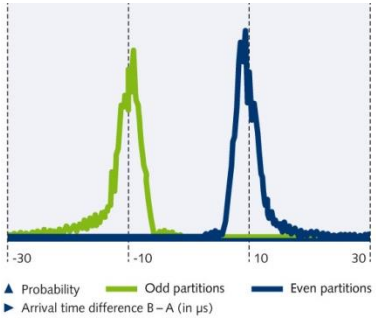
- The hardware used for the core components of T7 is
  - HP Gen8 DL380 servers with Intel Sandy Bridge CPUs (E5-2690) for core components (matching engines, market data publishers, gateways,…).
  - Participant facing interface cards on the gateways and market data publishers use Solarflare OpenOnload technology to bypass the kernel TCP stack.
- The used operating system is Red Hat Linux 6.3 with real-time kernel on all core components.
- On the network layer Eurex Exchange offers Trading Participants to connect via 10 GBit cross connects to its platform in the Equinix data centre.
- Internally, all core components are connected via an Infiniband network in order to provide lowest possible latency.
- T7 is based on IBM WebSphere MQ Low Latency Messaging in order to deliver the required speed, capacity and stability requirements.

Market data distribution

Enhanced Market Data Interface – latency impact for market data

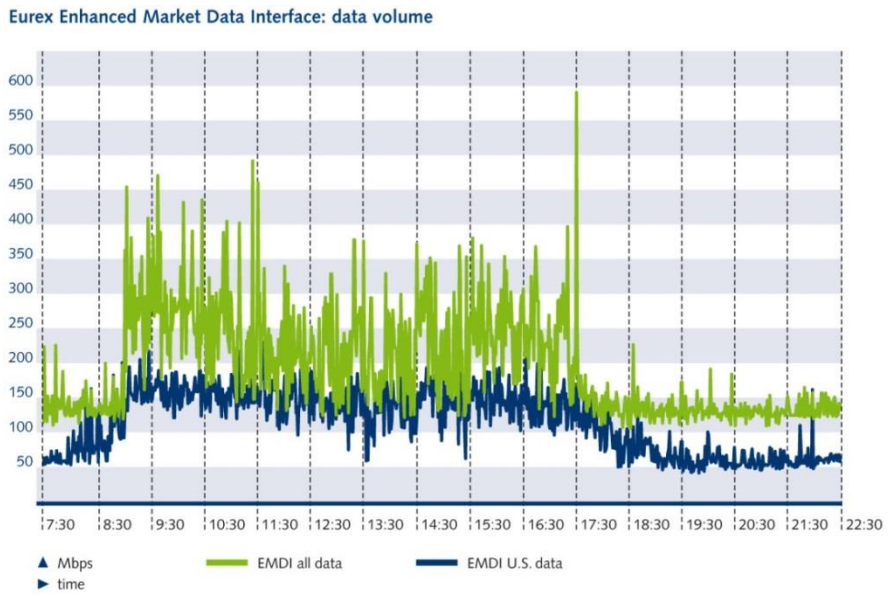


- Please note that for products assigned to an even partition, market data is published first on the A stream and then on the B stream whereas, for products assigned to an odd partition market data is published first on the B stream and then on the A stream.
- The partition ID is contained in the UDP datagram header of the order book incremental messages and can be used for filtering on UDP datagram level.
- Furthermore, a UDP datagram on the Eurex EMDI order book delta or snapshot channel will only contain data of exactly one product.



- The average latency difference between the A and the B feed is about 10usec (see diagram to the left).
- Eurex Exchange provides a csv file on a daily basis with the minute-by-minute network latency (minimum, average, maximum 99 percent) for the A and B streams of the Eurex Enhanced Market Data Interface for non-co-location access points. This information can help you determine whether you or Eurex Exchange had an issue causing market data latency.

Market data volume

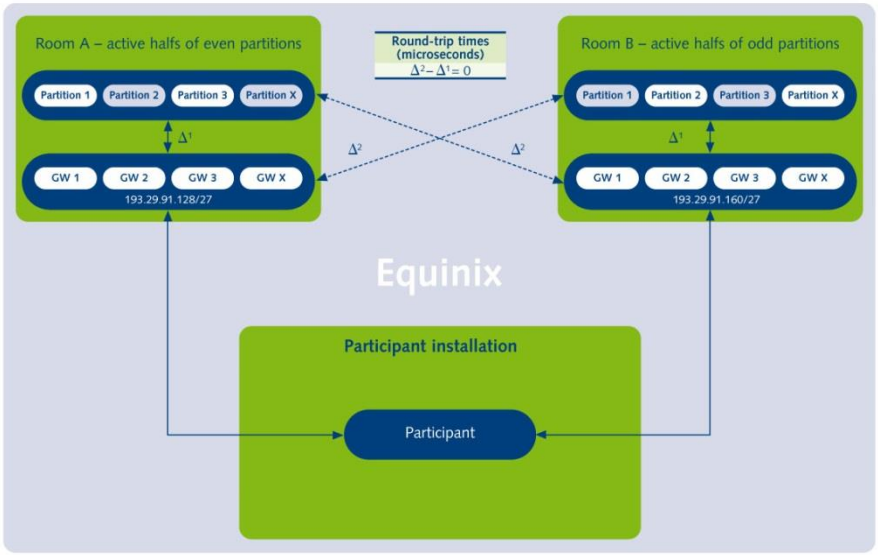


- The provided data shows one data point per minute for 26 August 2013.
- Each data point equals the maximum bandwidth produced on a 1ms scale by the complete B stream in kbps.
- Eurex Enhanced Market Data Interface peak data volume can be a significantly higher on high volume trading days. Hence participants that want to receive data for all Eurex Exchange's products or U.S. only products with less than 1ms queuing delays need to use a connection with a bandwidth of more than 500 Mbps (all products) or 200 Mbps (for U.S. only products), respectively.

- Latency sensitive Trading Participants are advised to use two 10 GE connections (one for each market data stream) in co-location to receive market data.
- Please note that access to the new Eurex Enhanced Order Book Interface (an order-by-order feed for benchmark futures) with T7 2.0 in November will be available via 10 GE connections in co-location only.

Orders/quotes – optimal access

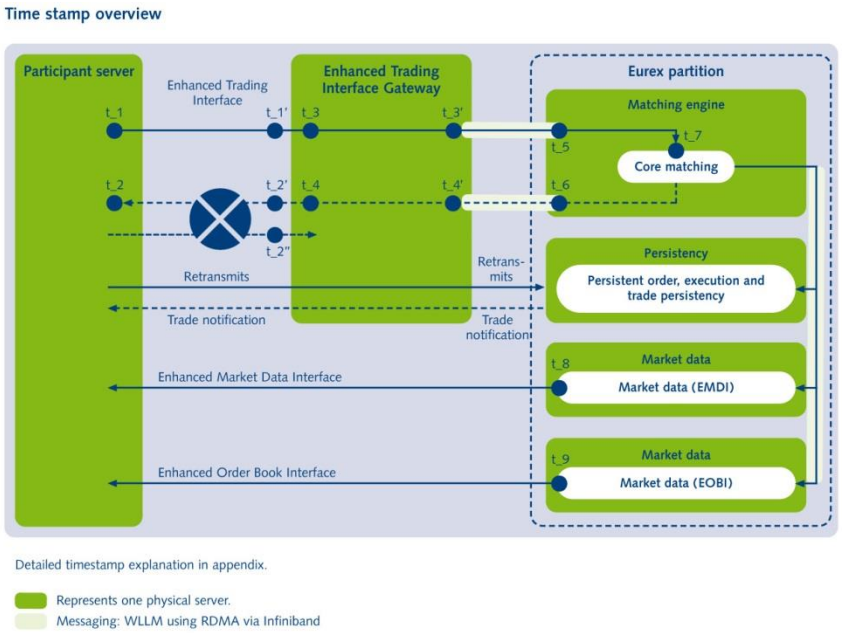
Eurex Enhanced Trading Interface – latency impact for orders



- Fastest access to T7 is provided via the high-frequency Eurex Enhanced Trading Interface gateways in Equinix.
- For optimal routing of orders, the room in which the gateways and matching engines are located is no longer relevant.
- The T7 gateways duties include all validations that do not need the knowledge of the order book or market state.
- To achieve lowest possible latency it is recommended to use the short order layout if possible - this saves about 7 usec gateway processing time on the way in compared to "normal" lean orders.

- Daily statistics about private "last mile" performance between the high-frequency gateways and Participant servers as well as best in class numbers (per location and system wide) are currently provided for the 10GBit connections to T7 (this will be extended to cover all Eurex ETI sessions soon). Eurex Exchange expects that a good daily average TCP/IP round-trip will be less than 6 usec for 10 GBit connections.

T7 topology

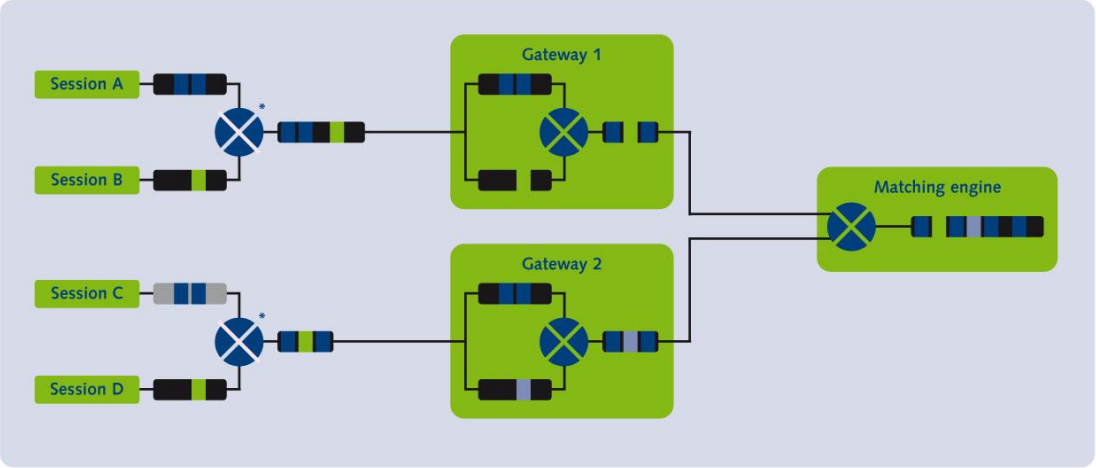


- Matching engine:
  - order book maintenance & execution
  - creation of direct responses as well as execution messages all for passive orders/quotes
  - creation of listener broadcast for standard orders
- Persistency: the persistency component is in charge of the following functions
  - persistent order storage,
  - trade/execution history,
  - transaction history for standard orders
- Market Data (Eurex EMDI):
  - creation of order book delta messages
  - creation of order book snapshot messages
- Market Data (Eurex EOBI):
  - creation of order book messages
  - creation of order book snapshot messages

- PTP based synchronization of clocks using hardware support is used for high-frequency gateways, matching engines and market data servers in production (and also in simulation). Hence time stamps on these servers can be used to analyse one way transport times.
- t_3' and t_4' will be made available in T7 2.0 in all Eurex Enhanced Trading Interface responses.

Orders/quotes – inbound message sequencing

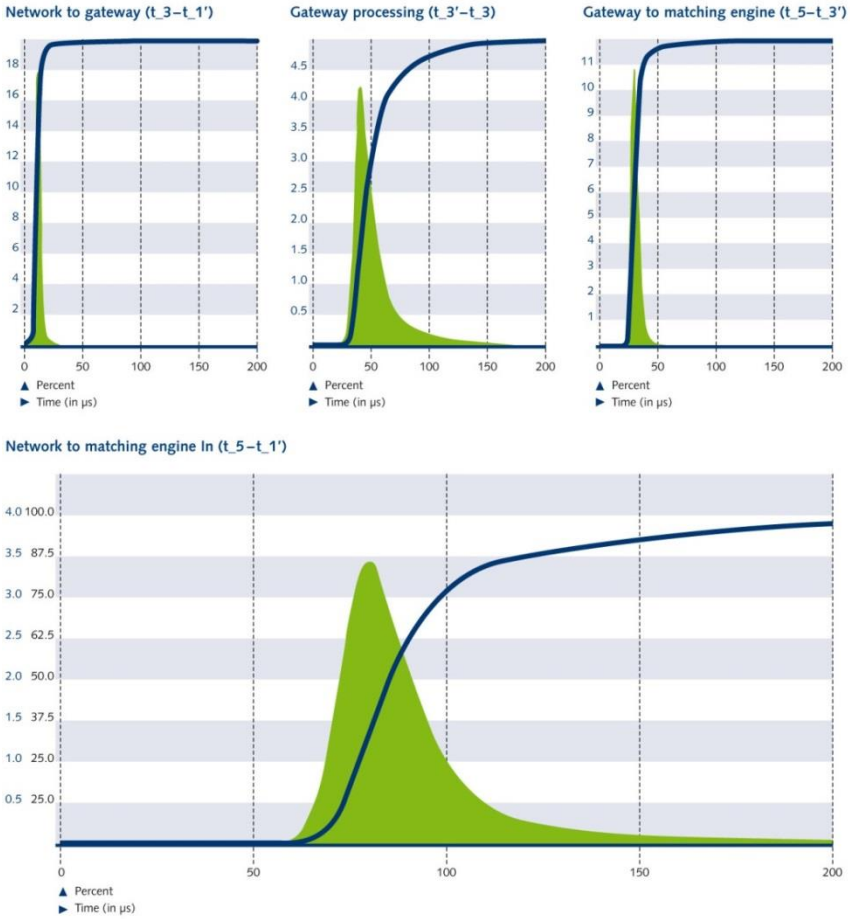
- For many participants understanding how inbound messages are forwarded through the Eurex infrastructure and between which points the sequencing is preserved is vital. The following diagram shows where inbound sequencing happens.



Inbound sequencing inside the T7 system takes place
- In the gateway for messages of all sessions connected routed to one matching engine (=partition).
- In the matcher for messages of all sessions.

Inbound ordering is preserved
- Within the messages of a session routed to one matching engine (=partition).
- Between the messages sent from one gateway to one matching engine (=partition).

Orders/quotes – inbound latency profile



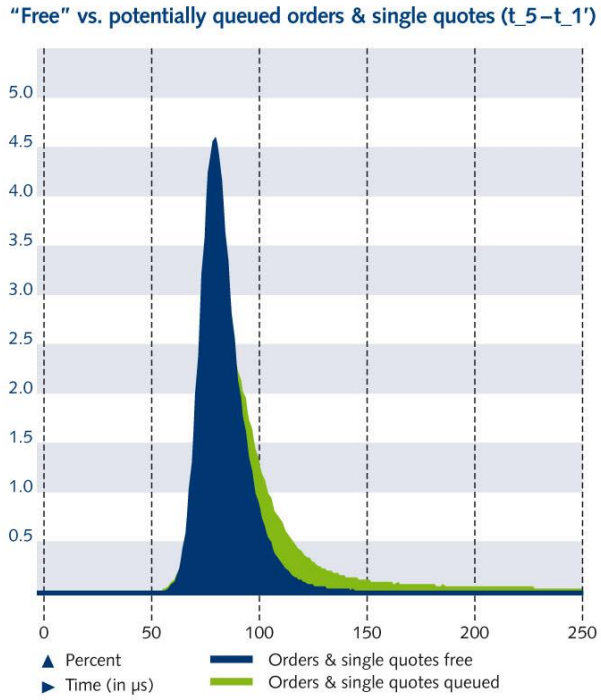
- The top graphs on the left show the latency patterns of three different parts of our trading infrastructure (network tap to gateway in (user space), gateway applications space processing, gateway out (application space) to matching engine in (application space) an order/quote passes through before time-priority is assigned.
- We can see that total variance is dominated by the variance of the gateway applications space processing

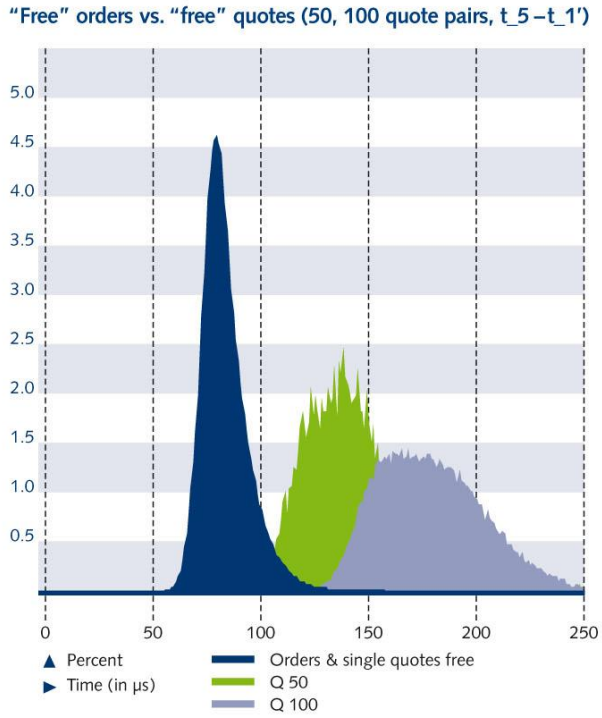
- The bottom graph shows the aggregate inbound latency of all three parts leading an order/quote passes through before time-priority is assigned on the matching engine. The gateway part is (by far) most relevant for the total variance on the inbound path of a transaction.

Orders/quotes – inbound latency variance

Inbound latency variance has three main sources:

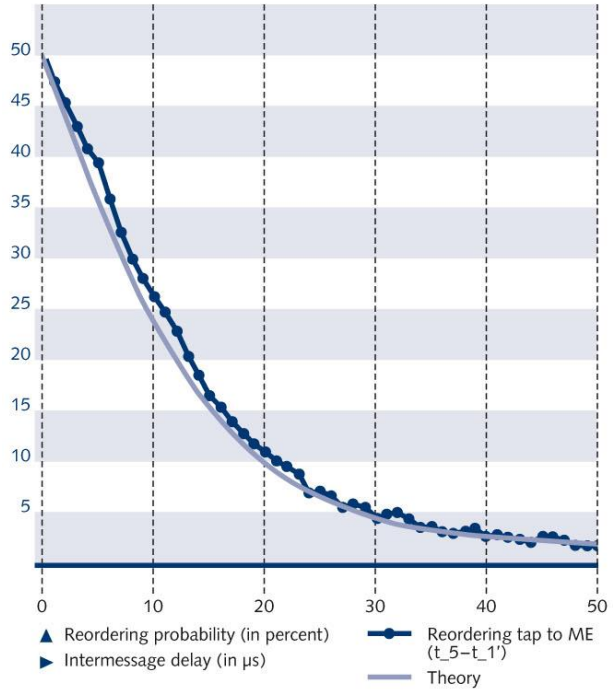
- Statistical effects (always present)
- Queuing/Overloading effects (e.g. 'microbursts', input rate > processing rate)
- Data dependent latency differences (e.g. quotes with 100 quote pairs vs. single quote)





Orders/quotes – inbound latency variance implications

- As outlined, the matching sequence may be different from the incoming sequence because of the statistical distribution of latency figures.
- The graph below shows the 'theoretically expected' probability of reordering of two consecutive –but unrelated- transactions compared to the actual observed reordering.
- The expectation matches the data quite well, indicating that the latency of two messages is indeed statistically unrelated.



Percentiles	Tap to ME (t_5–t_1')	Tap to GW (t_3–t_1')	GW (t_3'–t_3)	Gw to ME (t_5–t_3')
10%	74	8	35	26
25%	79	9	39	27
50%	86	11	45	30
75%	98	12	56	32
90%	121	14	77	35
Confidence intervals				
25–75%	19	3	17	5
10–90%	47	6	42	9

Orders/quotes – detailed performance data for futures

Our transparency

- For the top 10 futures products, daily statistics about the matching engine processing times as well as Eurex Enhanced Transaction Interface gateway processing times, are provided via the 'Member Section' on the Eurex website.

10 September 2013

Product	Product ID	Matching engine Round-trip times (in ms)			Enhanced Trading Interface Round-trip times (in ms)		
		Average	Median	99th percent	Average	Median	99th percent
EURO STOXX 50® Index Futures	FESX	0.156	0.068	1.320	0.328	0.205	1.900
STOXX® Europe 50 Index Futures	FSTX	0.062	0.060	0.132	0.197	0.184	0.475
DAX® Futures	FDAX®	0.088	0.063	0.560	0.231	0.197	0.900
MDAX® Futures	F2MX	0.061	0.061	0.101	0.184	0.178	0.301
SMI® Futures	FSMI	0.063	0.058	0.168	0.190	0.178	0.412
Euro-Bund Futures	FGBL	0.105	0.064	0.720	0.258	0.201	1.140
Euro-Bobl Futures	FGBM	0.099	0.061	0.820	0.248	0.196	1.180
Euro-Schatz Futures	FGBS	0.068	0.060	0.239	0.203	0.188	0.540
Euro-Buxl® Futures	FGBX	0.061	0.058	0.158	0.191	0.179	0.405
Long-Term Euro-BTP Futures	FBTP	0.073	0.062	0.248	0.209	0.192	0.500
Euro-OAT Futures	FOAT	0.071	0.060	0.274	0.213	0.190	0.640

Throttle limits

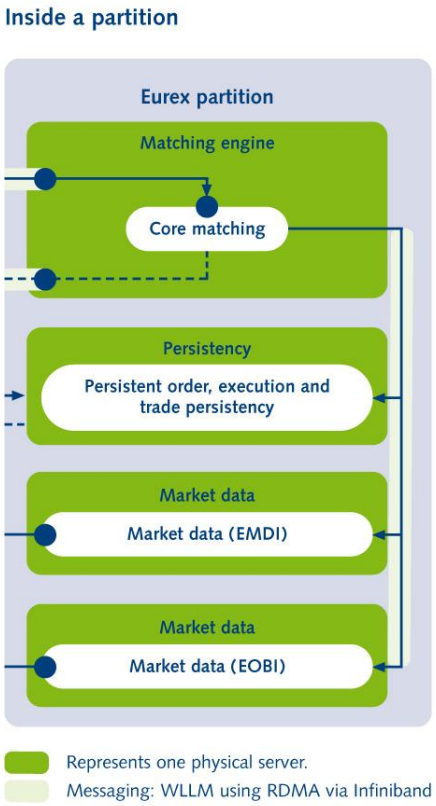
In order to protect its trading system, Eurex Exchange has several measures in place to ensure that its most vital components are not harmed by a malfunctioning client application. In particular Eurex Exchange uses the following throttles:

- All Eurex ETI sessions (HF and LF) are available with throttle values of 150 txns/s or 50 txns/s.
- Furthermore, since T7 1.1 LF sessions are available that can not enter orders/quotes but can only receive trade and listener broadcasts (at a reduced price).
- All Eurex ETI sessions type have an assigned disconnection limit of
  - 450, i.e. in case of more than 300 consecutive rejects due to exceeding the transaction limit (throttle) for sessions with a throttle value of 150 txn/sec
  - 150, i.e. in case of more than 100 consecutive rejects due to exceeding the transaction limit (throttle) for sessions with a throttle value of 50 txn/sec.
- For both limits all technical transactions are counted in a sliding window.

Agenda

- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

Processing inside a partition



- Orders/quotes entered for a specific product are sent by the gateway server to the respective matching engine (residing in a partition).
- The matching priority is assigned when the orders/quotes are read into the matching engine.
- The core matching component works as follows:
  - when an order/quote arrives, it is functionally processed (e.g. put in the book or matched).
  - handover of all data resulting from the (atomic) processing of the incoming order/quote to the market data and persistency components in the partition.
  - resulting messages for all orders/quotes are sent out in the following order:
    - direct response to the order/quote entered (for persistent as well as for non-persistent orders and quotes)
    - fast execution information for booked orders/quotes (in case of a match)
- In case that during this phase several new orders/quotes transactions arrive at the core matching component the processing remains unchanged, i.e. no batching takes place any more!
- The generation of market data (by the market data distributers) and trade confirmations (by the persistency server) are done on the respective servers. Hence the order of the resulting messages from these servers is not deterministic.

Agenda

- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

Eurex Enhanced Order Book Interface

General characteristics

- Provides the entire visible order book by publishing information on each individual order and quote-side, executions as well as state information in real-time and un-netted manner during continuous trading.
- Is available for selected benchmark futures.
- Eurex EOBI is only available via 10Gbit network connections.
- Each order and quote can be uniquely identified by the combination of
  - Instrument identifier, side and priority timestamp.
- No information about synthetic prices is provided.
- Trade statistics are available in snapshot messages only.
- Entire order book is published via the incremental stream when the order book is opened and via snapshots periodically.
- Fixed-length binary messages without any data compression are used (as in Eurex ETI).
- Snapshots and incremental market-data messages are delivered via separate channels (out-of-band).
- An execution summary message is provided to enable fast trading decisions.

Eurex Enhanced Order Book Interface

Expected number of orders

- Total number of open orders stored in the order book for benchmark futures is typically well below 10.000.
- The following table shows the maximum number of open orders at the start of continuous trading as well as the intra-day maximum of open orders for the top 6 futures during the period 1. August – 30. September 2013:

Maximum number of open orders

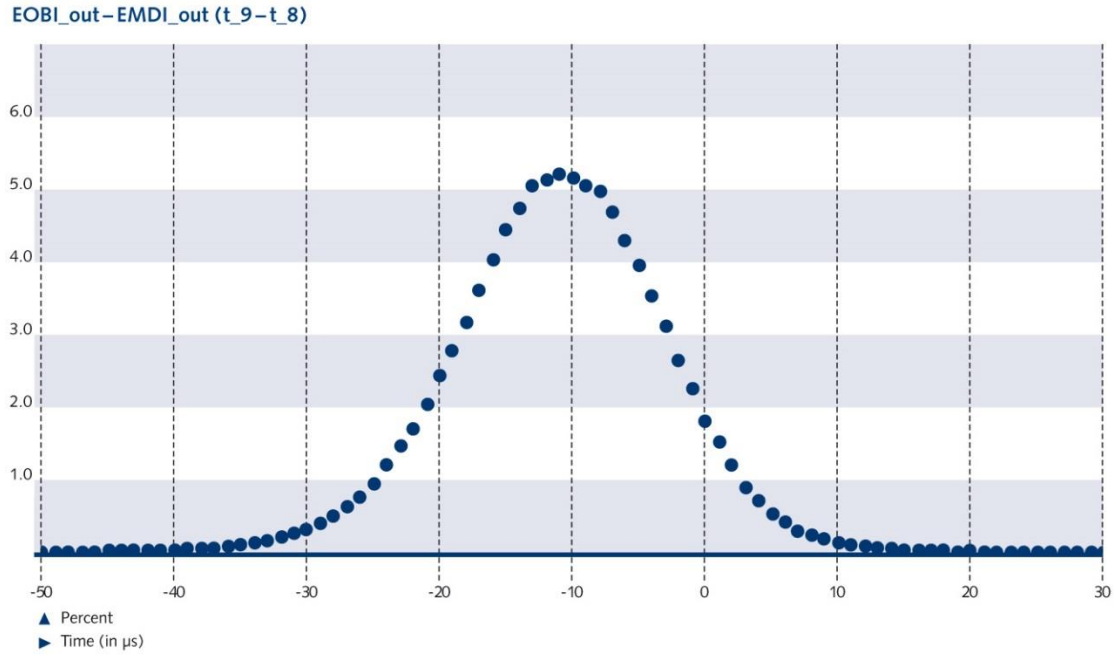
Product	Product ID	Start of trading	Intraday
EURO STOXX 50® Index Futures	FESX	7,300	8,000
DAX® Futures	FDAX®	1,600	2,500
SMI® Futures	FSMI	200	500
Euro-Bund Futures	FGBL	5,800	6,800
Euro-Bobl Futures	FGBM	3,600	4,200
Euro-Schatz Futures	FGBS	4,000	4,400

- The expected frequency with which snapshots of the order book are published depends on the number of open orders of the products of the respective feed instance (= multicast + port combination).
- The initial configuration of the system is planned to ensure that snapshots are provided at least once every 10sec.

Eurex Enhanced Order Book Interface

Latency characteristics: order book updates without match

- Although the design does not guarantee that the market data provided for each incremental Eurex EOBI update is faster than Eurex EMDI market data, lab data indicates that this is the case in most (>93%) instances.

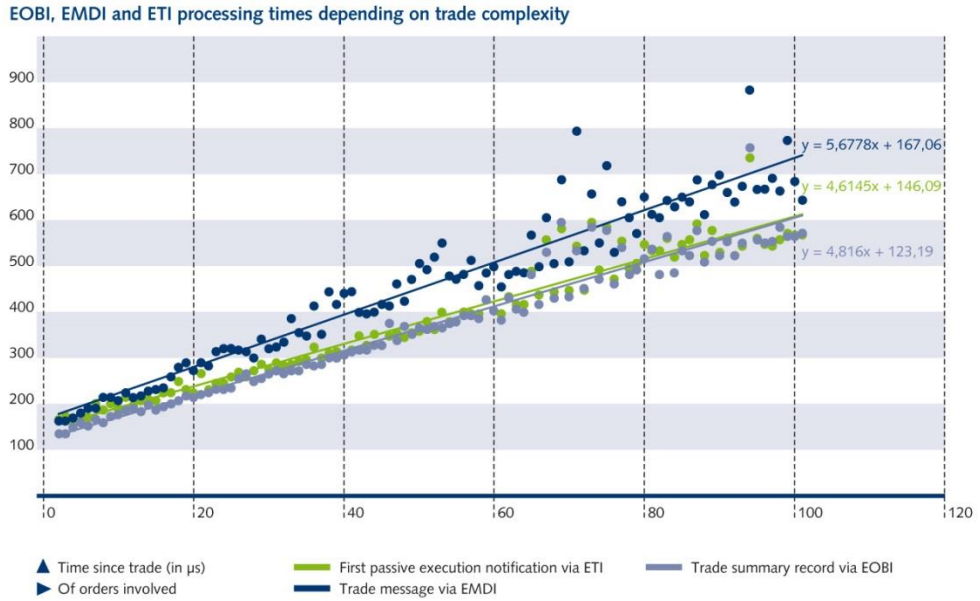


- The diagram on the left shows the probability distribution of t_9 - t_8, i.e. EOBI_out versus EMDI_out, for futures order transactions (without match).

Eurex Enhanced Order Book Interface

Latency characteristics: order book updates with match

- Although the design does not guarantee that execution related data provided for each incremental Eurex EOBI update is faster than via Eurex ETI or Eurex EMDI, lab data indicates that this is typically the case:



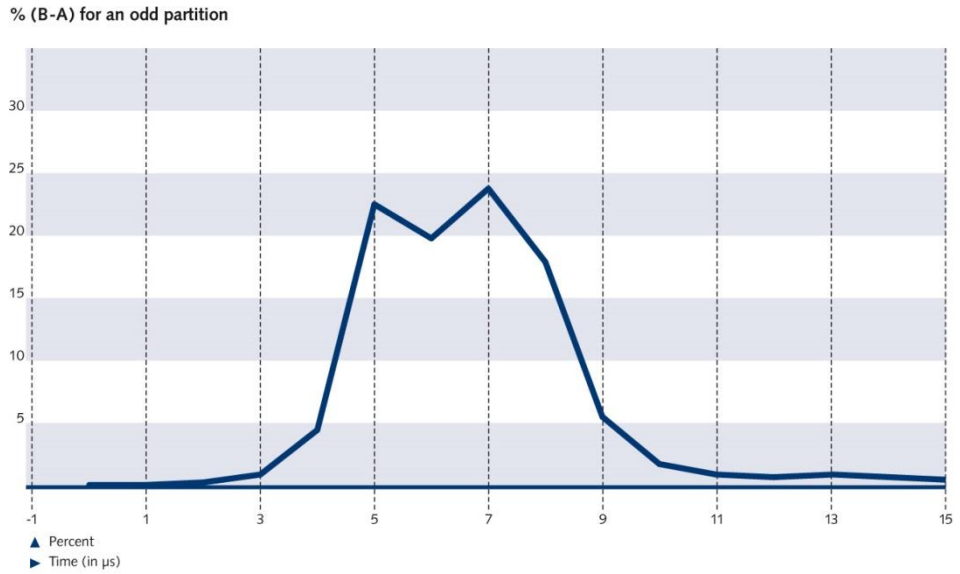
- This diagram displays the dependency of the median latency on the complexity of a match for Eurex ETI (t_4-t_7), Eurex EMDI (t_8-t_7) and Eurex EOBI (t_9-t_7). Note that for Eurex ETI we display the gateway-out time of the first passive notification and for Eurex EOBI the time-out of the UDP datagram containing the execution summary record.

Eurex Enhanced Order Book Interface

Latency characteristics: A versus B feed

- The incremental messages of the Eurex Enhanced Order Book interface are published on the 'A-feed' prior to being published on the 'B-feed' for all even partitions (and on the 'B feed' prior to the 'A feed' for odd partitions).

- The typical latency difference between the A and B feed is shown in the following diagram (lab data):



Agenda

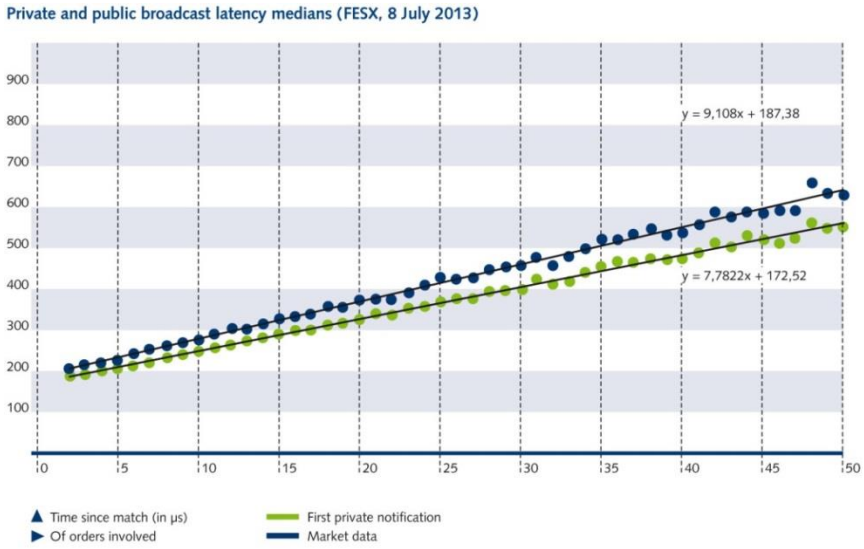
- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

Trading system dynamics

Latency of executions

The diagram on the right shows FESX®-Futures production data from 8 July 2013 on the relationship between trade complexity and broadcast latency

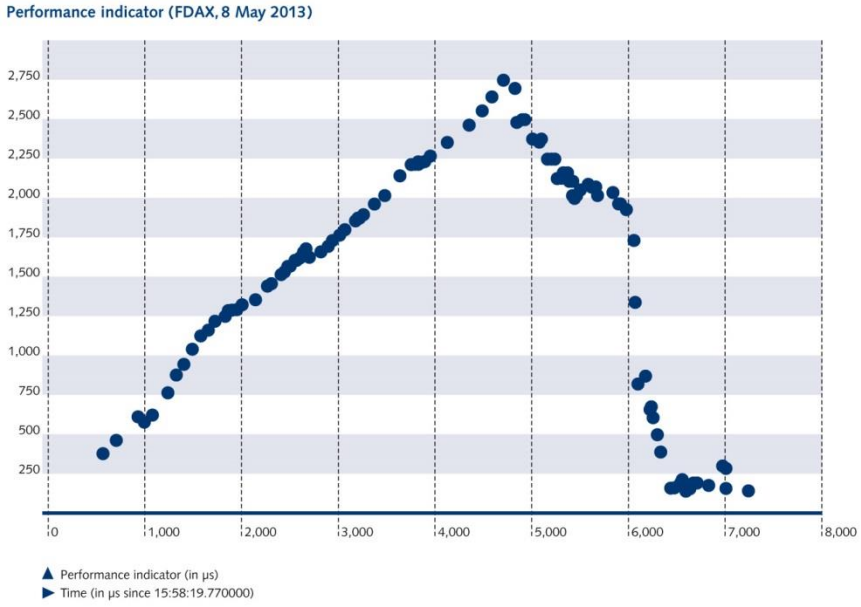
- for the latency profile of Eurex EMDI trade information versus the trade complexity – median of (t_8-t_7).
- for latency profile on the first passive execution information sent out for a trade –median of (t_4-t_7).
- The median (average) difference in time between two consecutive passive execution messages leaving the HF GWs is about 5 usec (8.4 usec).



Trading system dynamics

Matching engine and market data performance

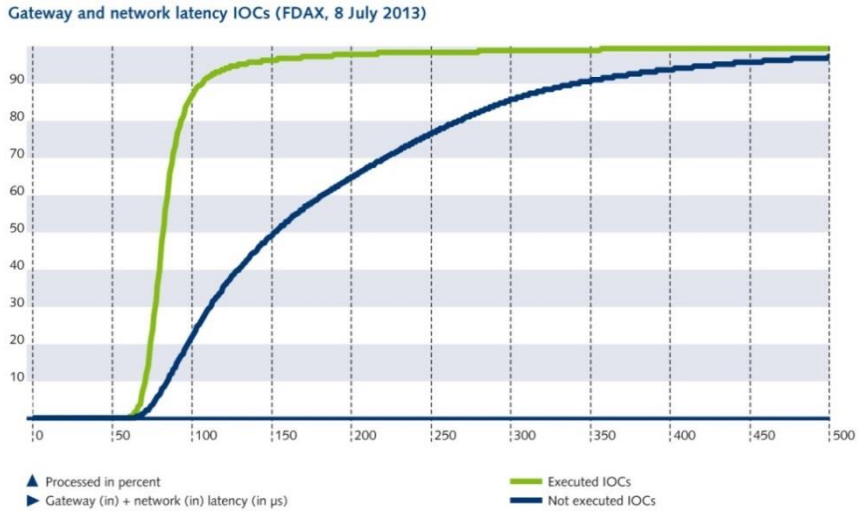
- During micro-bursts, our matching engines typically experience a temporary performance degradation.
- Unpredictable latencies cause risk (i.e. it takes longer to pull an order).
- To reduce unpleasant surprises, Eurex Exchange provides real-time insight in the matching and market data performance by providing the performance indicator (t_8-t_5) in the EMDI UDP packet header.
- This graph depicts the performance indicator (t_8-t_5), measured as the time needed by the matching engine to take an order/quote from the wire, process it, forward the resulting data to the market data server and put the market data incremental message on the wire for an example where a micro-burst caused a latency spike in DAX®-Futures on 8 May 2013.

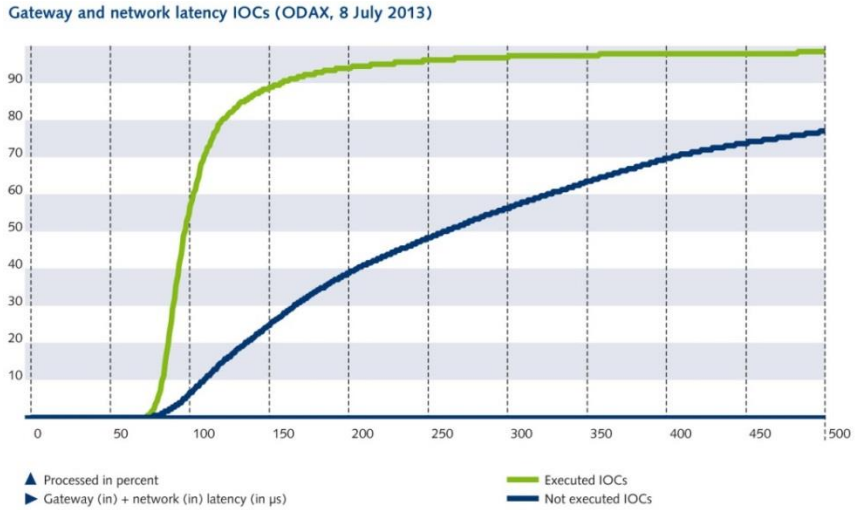


Trading system dynamics

Inbound gateway & network latency

- Congestion on the gateways and network between the gateways and matching engines can cause significant latency.
- Predictable latency can be found e.g. for executed IOCs (which typically do not have to queue behind other order transactions).
- The diagram shows the time of executed versus non-executed IOC orders in DAX®-Futures/DAX®-Options on 8 July spend on the way in from the inbound side of the gateway to the matching engine, i.e. t_5-t_3.
- The graph shows that for executed IOC orders the latency profile is very close to an ideal step function and hence very predictable.





Agenda

- Achievements
- Topology and system
- Inside a partition
- Trading system dynamics
- Eurex Enhanced Order Book Interface
- Infrastructure
- Outlook
- Appendix

What you need to be fast…

A few recommendations to achieve the lowest possible latency:

- Use the Equinix co-location facility to be close to Eurex Exchange's T7.
- Try the two different high-frequency gateways assigned to each Eurex Enhanced Trading Interface session to see which delivers the better performance for your strategy (try it out and compare your time stamps as well as p&l for different days).
- Trade messages need to be processed to create safety (only the trade messages contain legally binding information about a trade!). Therefore, we recommend to use either a low-frequency Eurex Enhanced Trading Interface session or a FIX trade capture drop copy to confirm the fast execution information provided by the execution reports via high-frequency sessions*.
- Use two 10 GE (cross-) connections for the Eurex Enhanced Market Data Interface and two 10 GE (cross-) connections for the Eurex Enhanced Transaction Interface.
- Use a state-of-the-art switches and only have at most one hop between the exchange network and your server.
- Use good NIC cards or TCP/IP acceleration, e.g. a Linux kernel-by-pass library.
- Take advantage of the exchange provided time service and sync your clocks to ours via PTP.

** Please see the Eurex ETI manual for more details and best practices in handling of the "optimistic" Eurex Enhanced Trading Interface execution reports.*

Agenda

- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

Outlook

The next scheduled updates to the T7 system are planned already:

- 25 November 2013 (T7 2.0): several enhancements including
  - new pre-trade limits on the number of open orders for a single product,
  - t_3' and t_4' will be made available in the Eurex Enhanced Trading Interface responses,
  - further performance improvements to reduce the variance of the high-frequency Eurex Enhanced Trading Interface gateways,
  - a new market data interface for benchmark futures – Eurex Enhanced Order Book Interface – an order-by-order market data feed for benchmark futures*.

Maybe you also want to collect other information Eurex provides in the context of high-frequency trading: www.eurexchange.com > **High-frequency trading**

In case you are an exchange Participant and want to be notified (via e-mail) about a new version of this presentation, let us know by sending an e-mail to hft@eurexchange.com.

Finally, Eurex Exchange will continue to investigate possibilities of extending the transparency with respect to latency figures – in case you have any suggestion – please get in touch with us!

** for the draft specifications see www.eurexchange.com **> T7 > system documentation > release 2.0**. For details visit Eurex Circular 70/2013.*

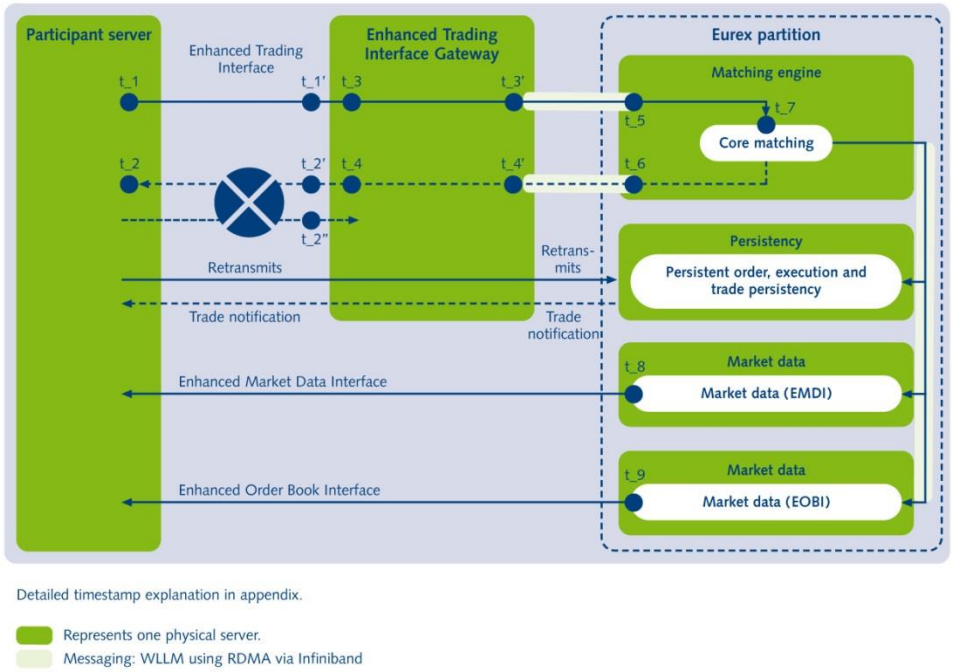
Thank you for your attention!

Agenda

- Achievements
- Topology and system
- Inside a partition
- Eurex Enhanced Order Book Interface
- Trading system dynamics
- Infrastructure
- Outlook
- Appendix

T7 topology & time stamps

Time stamp overview



Detailed timestamp explanation in appendix.

Represents one physical server.
Messaging: WLLM using RDMA via Infiniband

- PTP based synchronization of clocks using hardware support is used for high-frequency gateways, matching engines and market data servers in production (and also in simulation). Hence time stamps on these servers can be used to analyze one way transport times.
- t_3' and t_4' will be made available in T7 release 2.0 in all Eurex Enhanced Trading Interface responses.

Description of time stamps

Definition

- t_1,t_2: can be taken by the application when the request/response is read from/written to the socket.
- t_1',t_2' taken by a (passive) tap just in front of the High-frequency Gateway when the request/response passed through.
- t_3,t_4: taken by the Eurex Enhanced Trading Interface gateway when request/ response is read from/written to the socket on the Participant´s side of the gateway; contained in (private) Eurex Enhanced Trading Interface response.
- t_3',t_4': taken by the Eurex Enhanced Trading Interface gateway when request/ response is read from/written to the socket to/from the matching engine.
- t_5, t_6: taken by the matching engine when request/response is read/written; contained in (private) Eurex Enhanced Trading Interface response (note that consecutive t_5 timestamps are expected to differ by at least 5usec).
- t_1': time when a TCP packet sent by a Participant server passes a tap just in front of the Eurex high frequency gateway.
- t_2': time when a TCP packet sent by a Eurex gateway passes a tap just in front of the Eurex high frequency gateway.
- t_2'': time when the TCP acknowledge form a participant server of a TCP packet sent by the Eurex gateway to the participant server passes a tap just in front of a Eurex high frequency gateway.
- t_8: time taken by EMDI publisher just before the first respective UPD datagram is written to the UDP socket.
- t_9: time taken by EOBI publisher just before the first respective UPD datagram is written to the UDP socket.

T7 time stamp reference

The timestamps t_3,…,t_8 are available via the following fields:

•	t_3:	Tag	5979	("RequestTime")	in the Eurex ETI Response
•	t_4:	Tag	52	("SendingTime")	in the Eurex ETI Response
•	t_5:	Tag	21002	("TrdRegTSTimeIn")	in the Eurex ETI Response
•	t_6:	Tag	21003	("TrdRegTSTimeOut")	in the Eurex ETI Response
•	t_7:	Tag Tag	17 273	("ExecID") ("MDEntryTime")	in the Eurex ETI Response and in the Eurex EMDI Order Book Depth Incremental
•	t_0:	Tag	28820	("AggressorTimestamp")	in the Eurex EMDI Order Book Depth Incremental, in case a trade is reported
•	t_8:	no Tag		("SendingTime")	in the Eurex EMDI UDP packet header
•	t_9:	Tag	60	(„TransactTime")	in the Eurex EOBI packet header
•	(t_8-t_5):	no Tag		("PerformanceIndicator ")	in the Eurex EMDI UDP packet header of the Eurex EMDI Order Book Depth Incremental stream.

Notes on time stamps:

- All timestamps provided are 8 byte integers (in nanoseconds after Unix epoch).
- The PerformanceIndicator is a 4 byte integer (in nanoseconds as well).