



DEUTSCHE BÖRSE  
GROUP

# Insights into trading system dynamics

## Deutsche Börse's T7<sup>®</sup>

July 2019



# Agenda

**2** Overview

**8** Partition-specific gateway

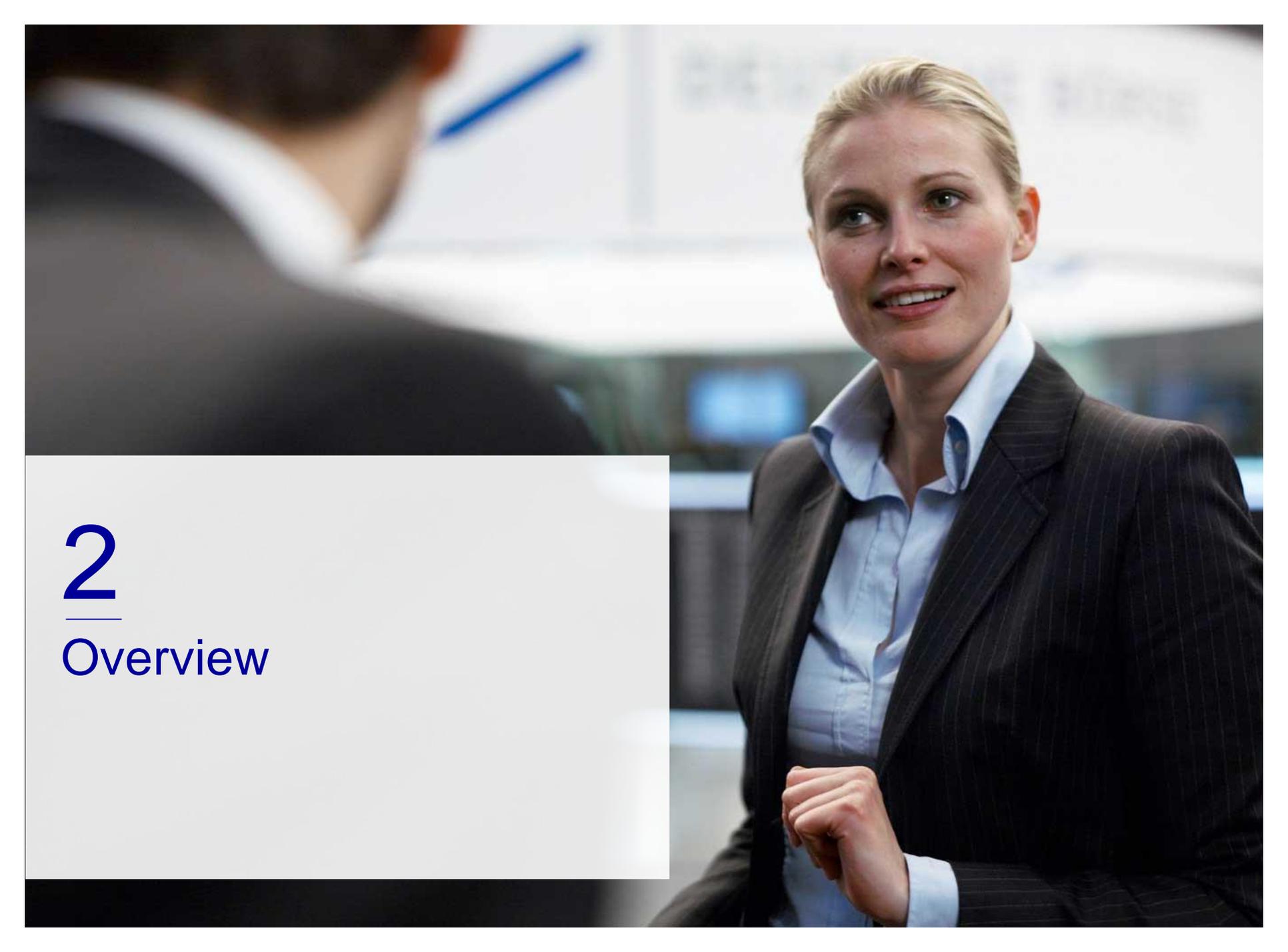
**14** Network access

**19** Latency aspects

**28** Market data

**38** What you need to be fast

**41** Appendix

A woman with blonde hair, wearing a dark pinstriped blazer over a light blue collared shirt, is smiling and looking towards a man whose back is to the camera. The man is wearing a dark suit jacket and a white shirt. They are in an office environment with blurred background elements like computer monitors and office furniture.

2

Overview

# T7<sup>®</sup> Technology Roadmap

Deutsche Börse is pursuing its technology roadmap to deliver innovative and superior trading technology.

## Recent developments

- HPT file service was launched (executions only: July 2018, all futures: March 2019).
- White Rabbit time synch service went live in Q3 2018.
- The number of Eurex low-frequency gateways was reduced from 6 to 4.
- T7 7.0 was released in December 2018 and T7 7.1 went live in May 2019 with further optimization of partition-specific gateways outbound processing.
- EOBI for all options was introduced February 2019.
- Hardware refresh in April/May 2019 targeted LF gateways.
- Passive Liquidity Protection (PLP) was introduced on 3<sup>rd</sup> June 2019 for German and French equity options.
- PLP replaced dual latency gateways in T7FX on 27<sup>th</sup> May 2019.
- Bulgaria Stock Exchange (XBUL) migrated to T7 on 24<sup>th</sup> June 2019.

## Outlook

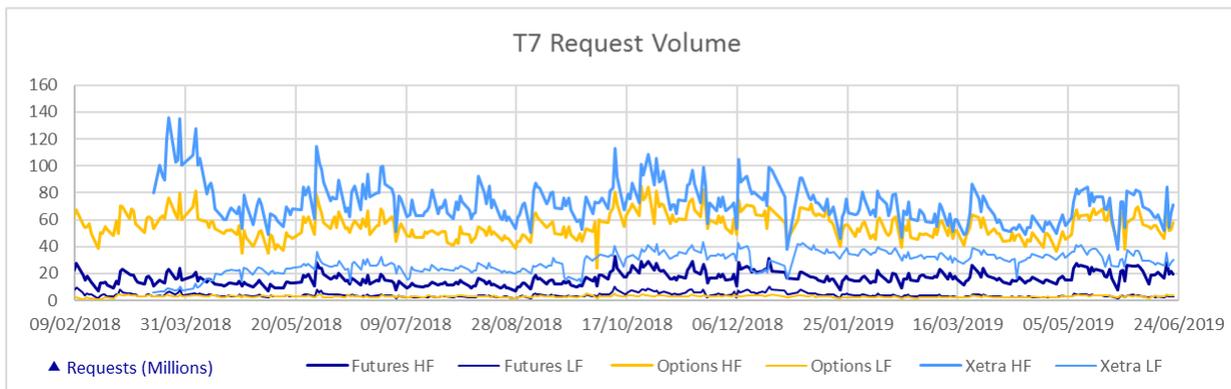
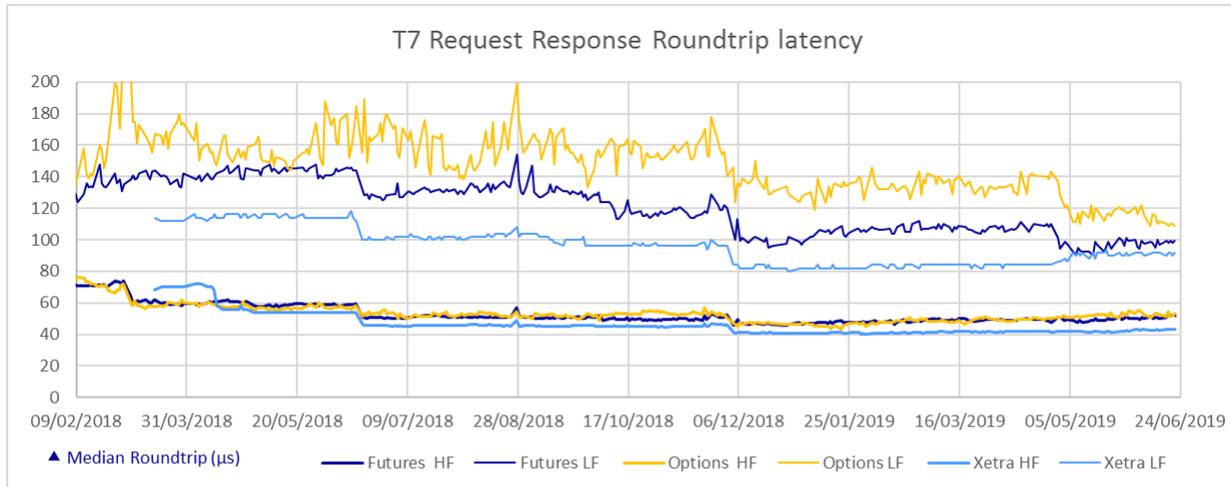
- After release 8.0 introduction the LF gateway requests will be routed via the PS gateway (see page 11 for details).
- In December 2018 partner exchanges of the Vienna market will migrate to T7 (XPRG, XBUD, XLJU, XZAG).
- In March 2020 the market Frankfurt (XFRA) will migrate to T7 on a separate instance.

For further details about T7 please visit our websites:

[www.eurexchange.com/t7](http://www.eurexchange.com/t7) and [www.xetra.com/xetra-en/technology/t7](http://www.xetra.com/xetra-en/technology/t7)

# Processed transactions and response times

## T7 request – response round-trip times

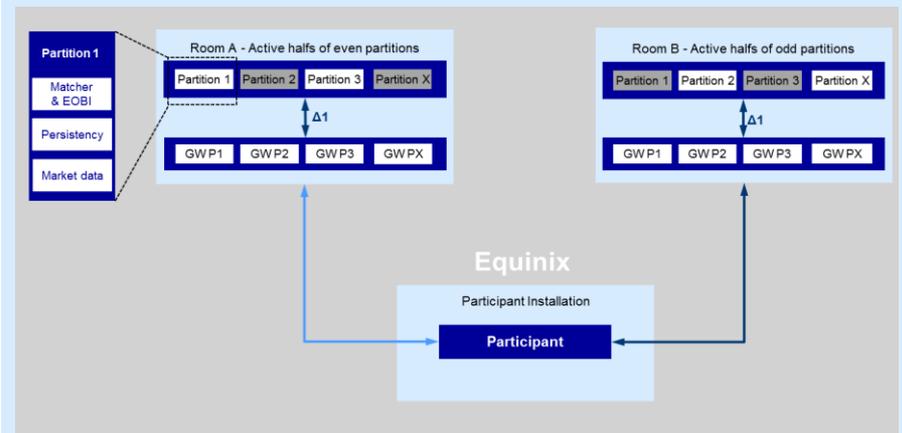


- Deutsche Börse has continuously invested in its trading system and is holding up transparency while providing a low latency trading venue.
- After introduction of PS gateways for Eurex in Q1/2018 and Xetra in April 2018 we continued to add functionality at the same time tuning our system further.
- Release 6.1 and release 7.0 further lowered the median latency for requests sent via PS gateways to below 50 µs (Eurex) and to around 40 µs for Xetra.
- Since release 7.1 tuning measures for the response path and a hardware refresh of the low-frequency gateways lowered higher percentile latencies.

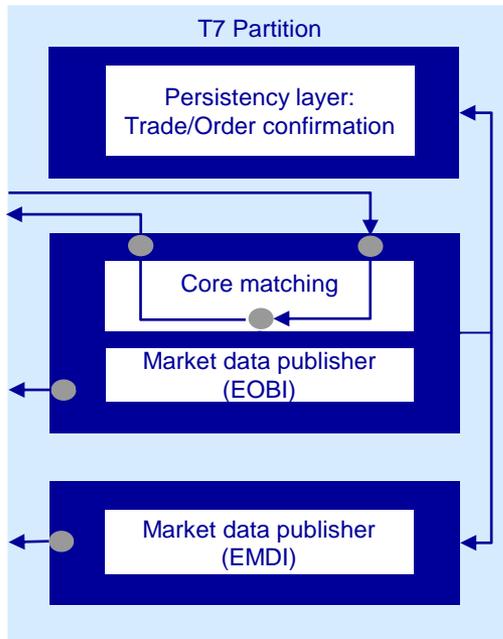
# Overview of T7

- T7 consists of partitions. 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 Eurex T7 and 10 Xetra T7 partitions.
- Separate partitions are used for markets Vienna (XVIE), EEX (XEEE), Malta (XMAL) and Bulgaria (XBUL).
- The reference data contains the mapping of products to partition IDs.
- With the introduction of partition-specific (PS) gateways there is a one-to-one mapping of active PS gateways to partitions. The default active PS gateways are located on the same side as the active matching engines.
- 4 low-frequency (LF) gateways allow access to all Eurex partitions and the separate EEX partition.
- 4 low-frequency gateways are shared between all cash markets (XETR, XMAL, XVIE, XBUL).

- Note that the active half of a partition and its partition-specific gateway is either on side A (for even partitions) or on side 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.



# 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 data to the EOBI Market data publisher and
- 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 responses and private broadcasts 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).

If during this phase several new orders/quotes transactions arrive at the core matching component the processing remains unchanged, i.e. no batching takes place.

The generation of market data other than EOBI (by the market data publisher), listener broadcasts and trade confirmations (by the persistency server) is done on separate servers. Hence the order of the resulting messages from these servers is not strictly deterministic.

# Middleware, network, hardware and OS overview

## **T7 uses state-of-the-art infrastructure components**

Intel Xeon E5-2667 v3 CPUs (Haswell) on all servers hosting core services (Matching engines, un-netted market data publishers)

Intel Xeon Gold 6144 CPU (Skylake) for partition-specific gateways.

Intel Xeon Gold 6148 CPU (Skylake) on non-performance critical servers.

We use the Red Hat Enterprise Linux operating system version and are currently migrating from 6.9 (real-time kernel) to version 7.6.

T7 internal communication between its core components is based on Confinity Low Latency Messaging using an Infiniband network in order to deliver the required speed, capacity and stability requirements.

## **T7 network access**

Deutsche Börse offers Trading Participants to connect via 10 Gbit/s cross connects to its T7 platform in the Equinix data centre.

The co-location offering uses Cisco 3548x switches. All cables are normalized to give an overall maximum deviation between any two cross connects of less than +/- 0.4 m (+/- 2 ns). Insight into network dynamics is offered by the High Precision Timestamp File service (see <https://datashop.deutsche-boerse.com/high-precision-timestamps>).

Participant facing interface cards on the gateways and market data publishers use Solarflare EnterpriseOnload wire order delivery API to bypass the kernel TCP stack and deliver messages in the same order received by the network card.

# 8

## Partition-specific Gateways



# Partition-specific gateway

## Implementation

The partition-specific gateway uses the same ETI protocol as all ETI gateways (current low-frequency and previous high-frequency).

### **Session setup**

All high-frequency sessions are eligible to connect to a partition-specific gateway.

A session may only connect to a single gateway at any given point in time.

There is a maximum number of sessions per participant allowed to login to a single PS gateway at any given point in time. This limit is currently set to 80.

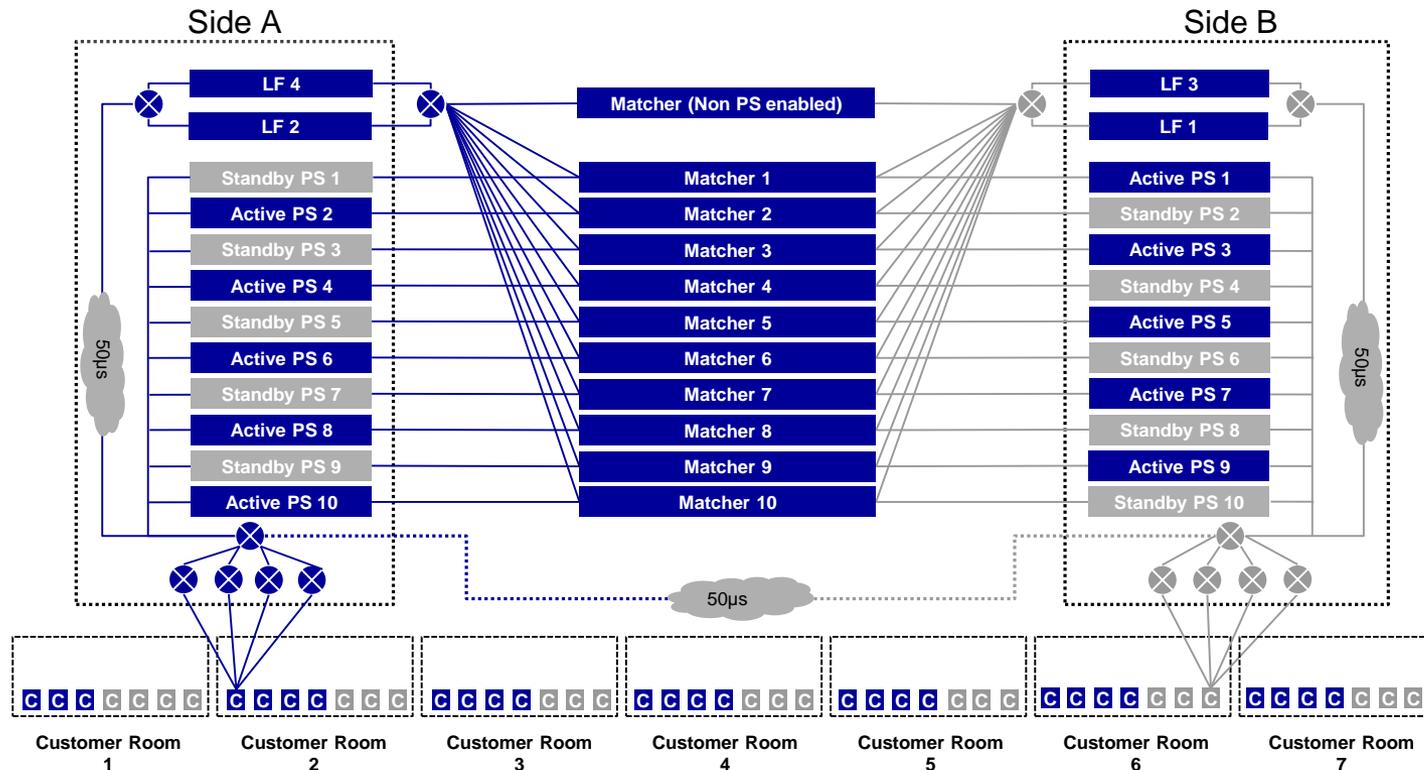
### **Connection**

Since Release 7.1 sessions directly connect to the PS/LF gateways.

You may send a session logon to the standby PS gateway to test network connectivity. Those logons will be rejected with the appropriate error code (refer to the respective ETI manual for details).

# Partition-specific gateway

## Topology



The active partition-specific gateway resides on the same side as the active matching engine per default. There is a network link between side A and B via the distribution layer switches with a one way latency of 50 µs. This guarantees that all partition-specific gateways are reachable via a single line in case of a failure.

Note that PS gateways are available only for Xetra and Eurex, whereas other markets offer access via low-frequency gateways only.

The network link to LF gateways is around 50 µs slower than the access to PS gateways via 10 Gbit/s lines.

# Partition-specific gateway

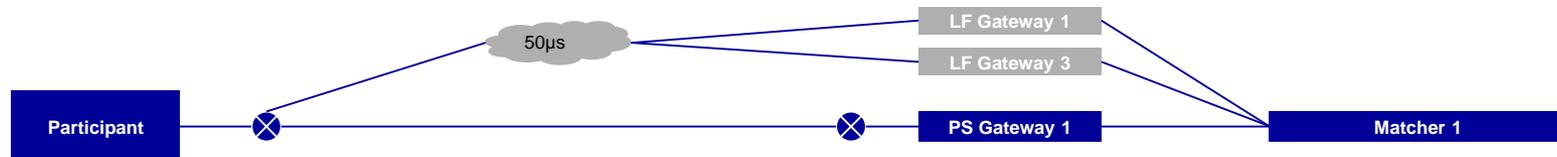
## Release 8.0 – routing of LF requests via PS gateways

After introduction of T7 release 8.0 we will route all LF gateway requests via PS gateways (if PS gateways are active for the corresponding market).

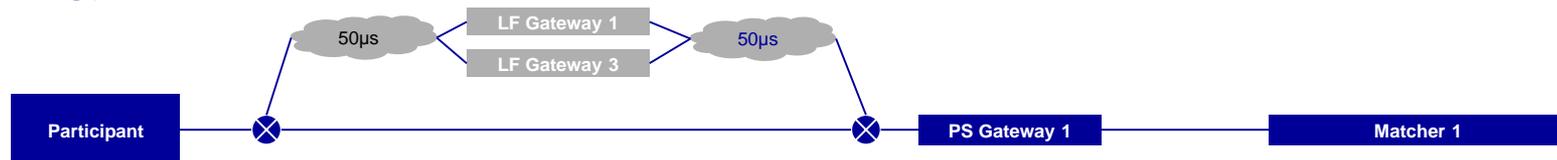
This will make overtaking between the different gateway types impossible and allows us to publish the gateway in timestamp in market data for all market data updates, i.e. the matching engine in timestamp will be replaced by the PS gateway in timestamp in market data.

The exact date of the migration will be communicated later.

### Topology R7.1



### Topology R8.0



# Partition-specific gateway

## Failover

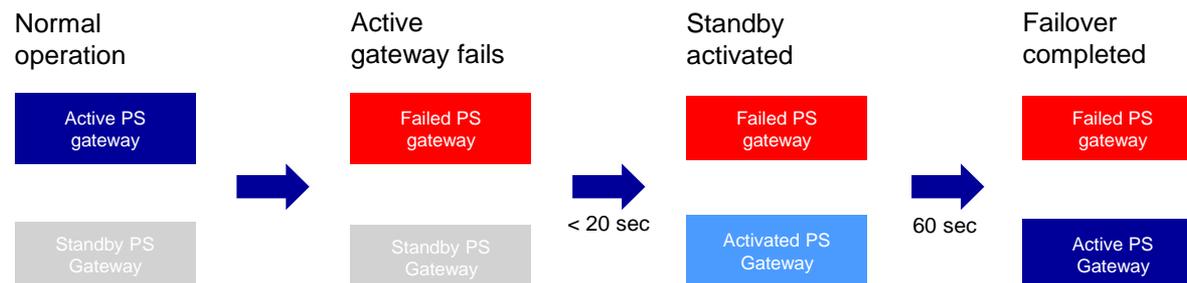
When a failure of a partition-specific gateway is detected, all sessions logged in via that gateway will be logged out and their non persistent orders and quotes are deleted.

Subsequently the standby PS gateway will be activated and allow session logins.

There will be an activation phase during which no order management via the activated PS gateway will be possible. This is currently set to 60 seconds to allow participants some time to evaluate the situation and re-login. Make sure you adhere to the limit of outstanding logins (50 per business unit, 10 per session, see next slide).

After the activation phase an ETI service availability broadcast will be sent to the connected sessions and order management service will be available.

### Schematic partition-specific gateway failover



## Throttle and session limits

**In order to protect the trading system, T7 has several measures in place to ensure that its most vital components are not harmed by a malfunctioning client application. Therefore transaction limits are imposed on T7 sessions.**

All ETI sessions (HF and LF) are available with throttle values of 150 messages/sec or 50 messages/sec. Furthermore LF sessions that cannot enter orders/quotes but can only receive trade and listener broadcasts are available (at a reduced price).

All ETI session types have an assigned disconnect limit of

- 450 for sessions with a throttle value of 150 messages/sec, i.e. a session will be disconnected in case of more than 450 consecutive rejects due to exceeding the transaction limit (throttle).
- 150 for sessions with a throttle value of 50 messages/sec, i.e. a session will be disconnected in case of more than 150 consecutive rejects due to exceeding the transaction limit (throttle).

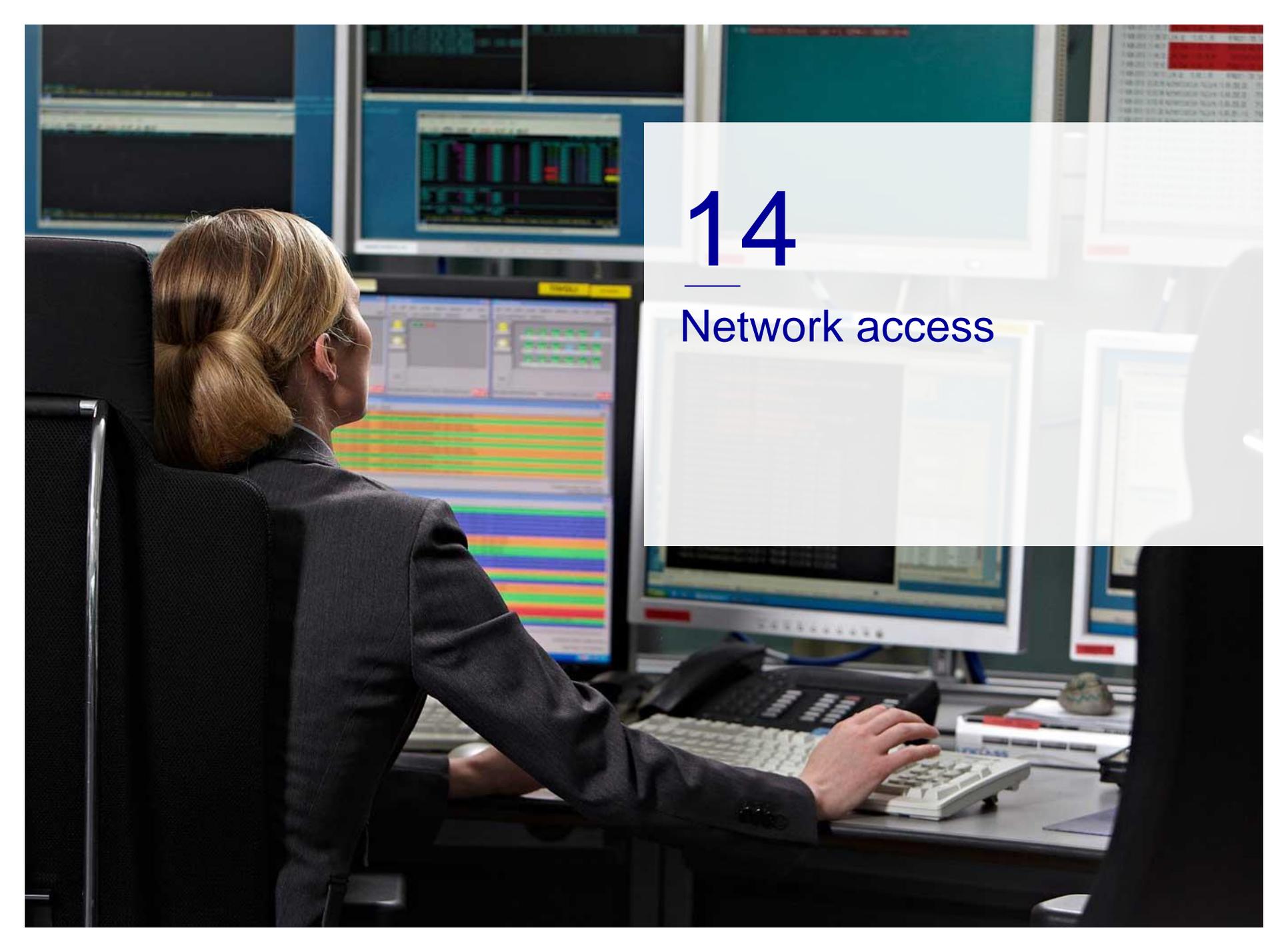
Please note that in case the disaster recover facility is used, all ETI sessions will have a throttle limit of 30 messages per second. For both limits, all technical transactions are counted using a sliding window.

The number of ETI sessions which can be ordered is limited. Currently, up to 80 sessions can be ordered. If more than 80 sessions are required please get in touch with your Technical Key Account Manager.

There is also a limit on maximum number of sessions per Participant and partition that can connect to a partition-specific gateway concurrently. This limit is currently configured to 80 sessions but subject to review.

On July 1<sup>st</sup> 2019 we introduced a limit on maximum number of outstanding session and trader login requests possible per business unit and per session at any given point in time. This limit is set to 50 on business unit level, 10 on session level. We recommend a synchronous login procedure, where a login request is sent on a session only after the previous login has been responded to. Please refer to the Incident Handling Guide for details.

The number of order entry cross connects in colocation that may be used concurrently on a single day is limited to 32.



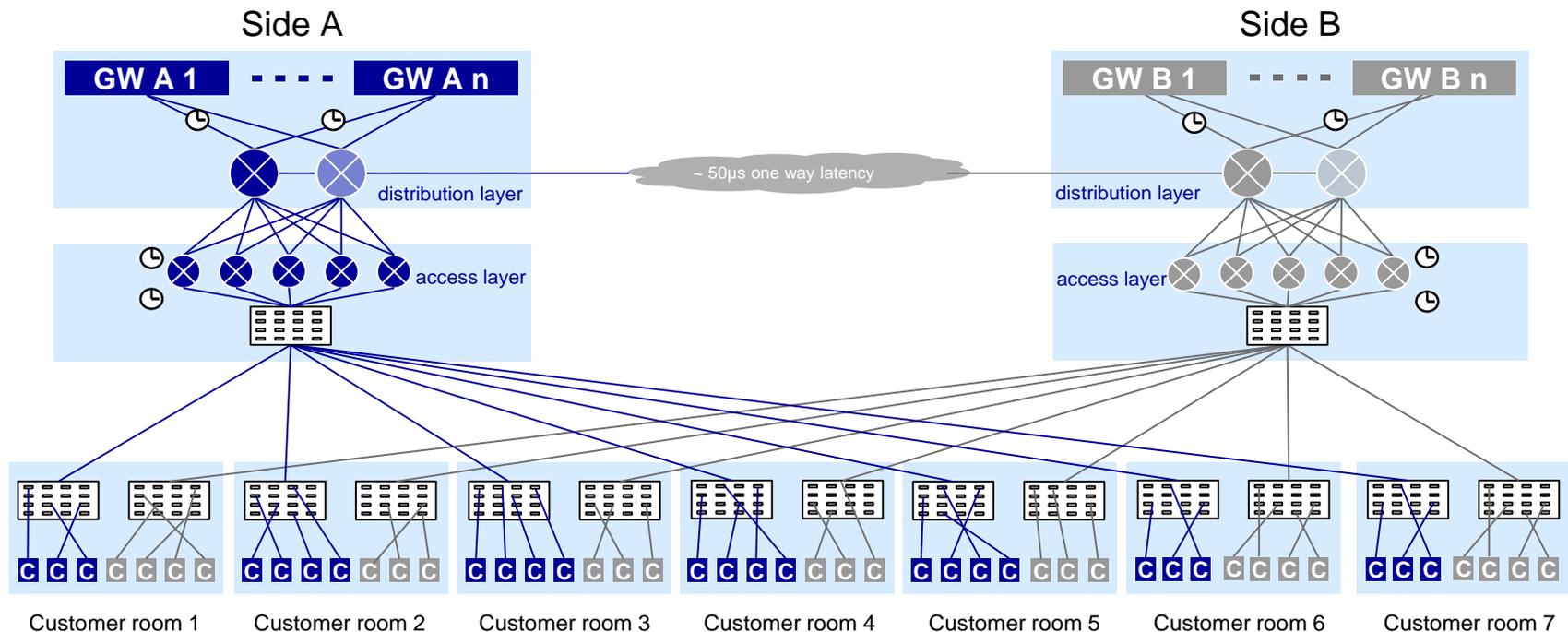
# 14

## Network access

# Network access

## Co-Location Topology

- There are separate networks for order entry and market data with centrally located access layer and distribution layer switches (diagram below shows Eurex order entry network as an example).
- Customers may connect to any access layer switch from any of the 7 co-located rooms:
  - Eurex order entry: 5 per side, Xetra order entry: 2 per side
  - Eurex market data: 4 per side, Xetra market data: 2 per side
- 2 distribution layer switches per gateway room per market
- The network link between side A and B order entry distribution layer switches has a one way latency of  $\sim 50\mu\text{s}$ .



# Network access

## Latency aspects

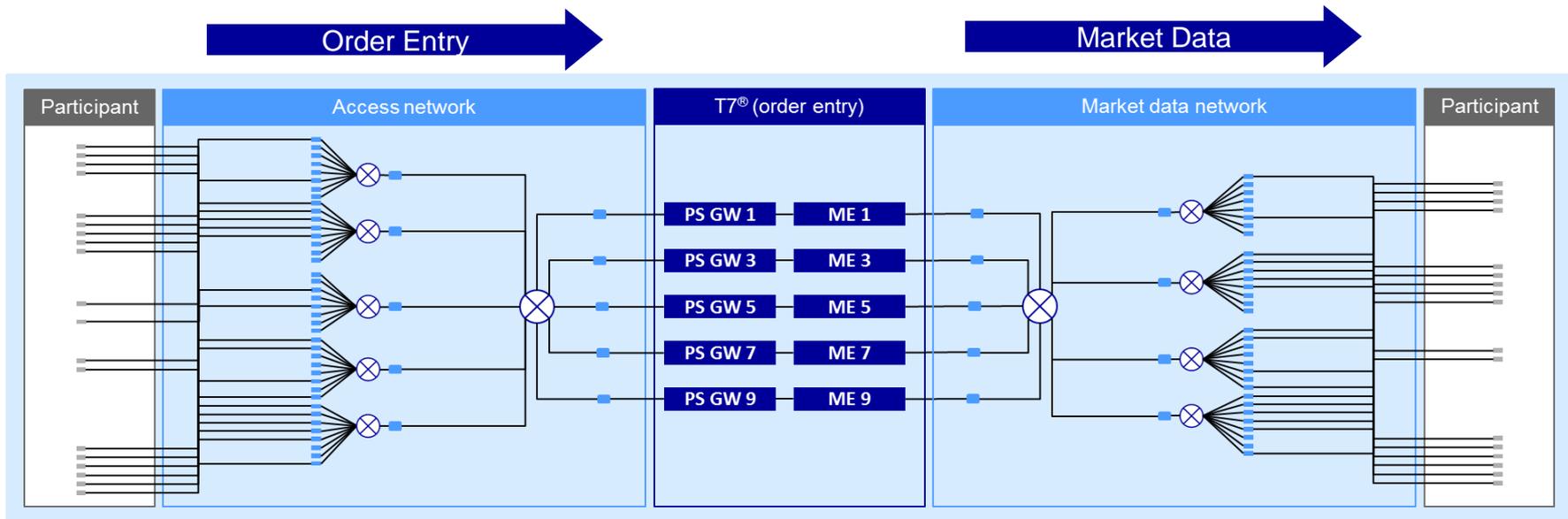
With the introduction of first in first out partition specific gateways the latency aspects of the network have moved into focus.

We are constantly measuring the latency profile of our co-location infrastructure using taps (■ below) and aggregation switches that use hardware assisted timestamping on ingress port.

The important factors are fairness (i.e. each connection is treated equally) and determinism – both on order entry and on market data dissemination.

Extra care has been taken to ensure the best time synchronization between these timestamping devices. We use a combination of white rabbit and pps achieving a time synch better than +/- 1 ns. The timestamping switches have a jitter of +/- 2 ns.

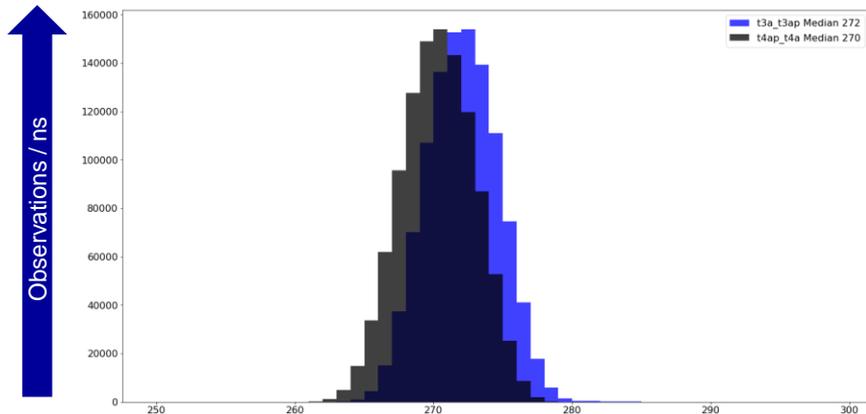
A detailed discussion of this can be found in our recent presentation 'Understanding an ultra-fast market through ultra-accurate time synchronization' to be found at [https://www.eurexchange.com/exchange-en/technology/high-frequency\\_trading](https://www.eurexchange.com/exchange-en/technology/high-frequency_trading)



\*Co-Location Topology - Eurex Side B – Odd Partitions shown as example

# Network access

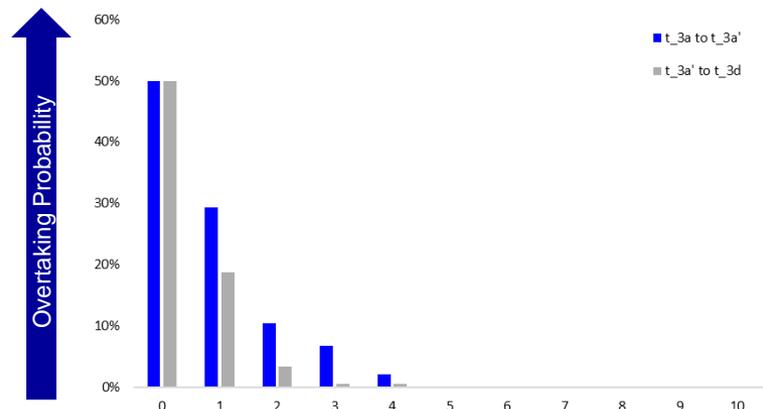
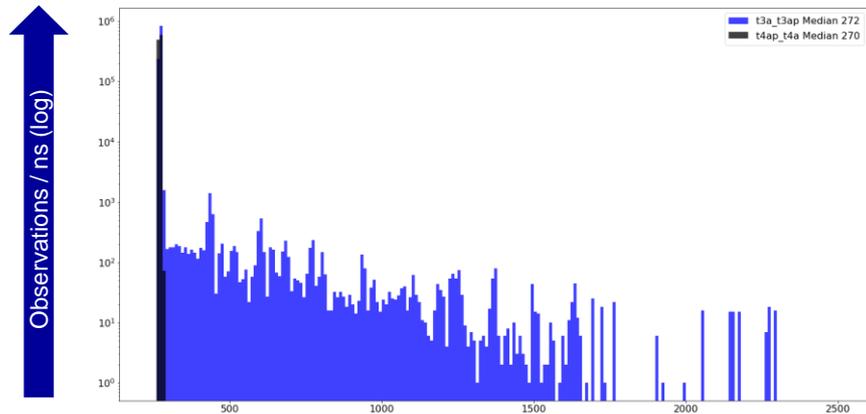
## Order entry latency



The base latency and latency jitter is identical for all access layer switches within the measurement precision. The spectrum is very tight with a confidence interval (5<sup>th</sup> percentile to 95<sup>th</sup> percentile) of 10 ns (top chart). 4 ns of that is attributable to the jitter of the timestamping device.

Note that queues occur in bursts, because multiple lines funnel into a single uplink per switch.

The chart on the bottom right shows the overtaking probability on the access layer switch between a first message in a burst and consecutive messages.

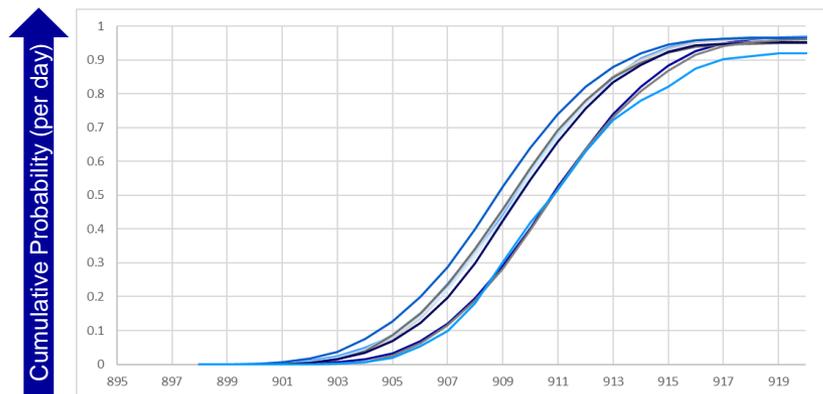
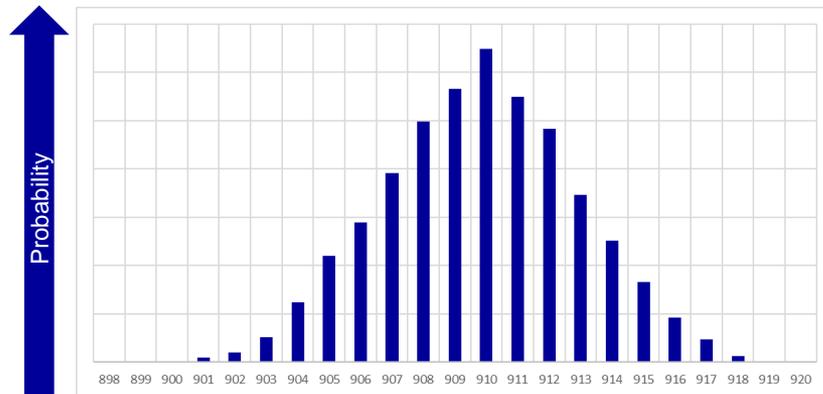


Switch latency (ns)

Interarrival time at t\_3a/t\_3a'

# Network access

## Market data dissemination latency



The market data network has a funnel in – fan out topology. It funnels in data from different market data disseminators (on the distribution layer switch) and fans it out via multiple access layer switches.

We took extra care to ensure a fair and deterministic network. The data volume flowing from the distribution layer to the access layer switches is identical for each switch (fully loaded). The switch models were chosen to minimize jitter between different ports and to provide a deterministic latency profile.

The latency through the network is dominated by the jitter of the switches (top left chart, 5<sup>th</sup> to 95<sup>th</sup> percentile: 13 ns,  $t_{9d}$  to  $t_{3a}$  latency for FESX EOBI shown).

There is also a dynamic effect, which is caused by the (varying) multicast subscribers on the access switches. The bottom left chart shows the cumulative latency distribution for different days over a period of two weeks in June 2019. Note that the median latency did vary by around +/- 2 ns from day to day.

The funnel in on the distribution layer switch can lead to queues when multiple market data disseminators send packets at the same time. Queues lead to bigger tails of the latency distribution, e.g. for FESX the 95 percentile is between 920 and 1100 ns, the 99<sup>th</sup> percentiles can go up to around 1600 ns at times.

Latency (ns)

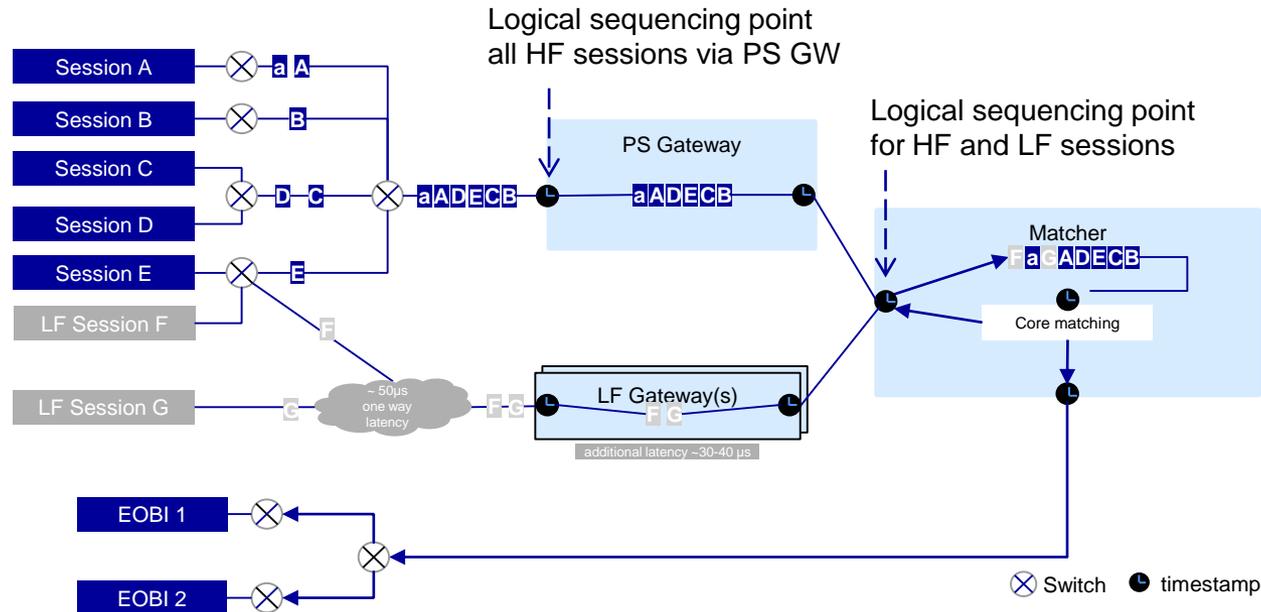
# 19

## Latency aspects



# Inbound message sequencing

## PS gateway architecture



### Inbound sequencing inside the T7 system takes place

- on the network in front of the PS trading gateway,
- in the matcher for messages from PS and LF gateways.

LF gateways have a 75 μs higher median latency ( $t_{5} - t_{3a}$ ).

Requests sent to LF gateways will be routed via PS gateways with introduction of Release 8.0.

### Inbound ordering is preserved

- within all messages from one PS gateway,
- between the messages sent from each LF/PS gateway to one matching engine (=partition).

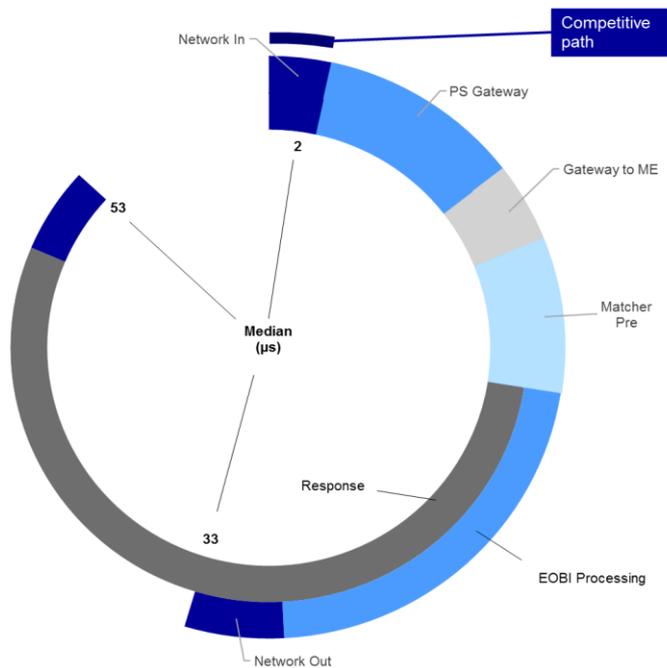
# T7 Latency Composition

The diagrams below compare the composition of latency for Eurex futures transactions for T7 release 7.1 and 6.1. The full circle represents 60  $\mu$ s.

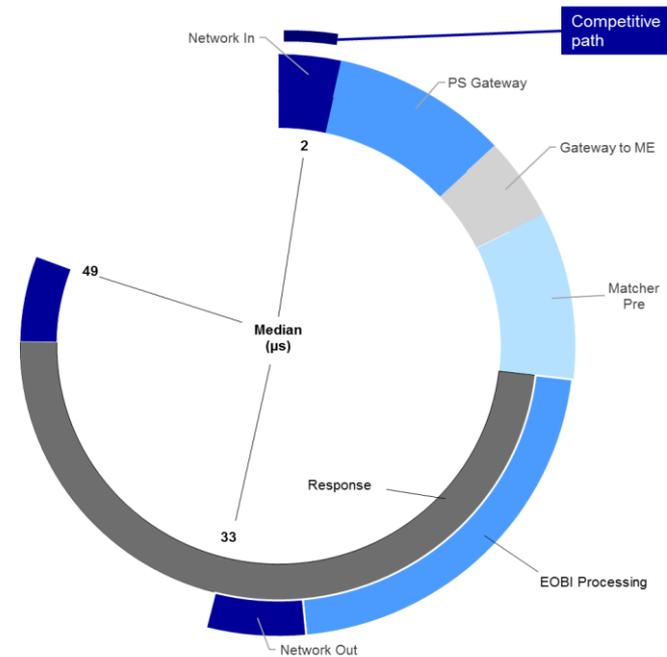
The latency in the response path has dropped by 4 microseconds with introduction of T7 release 7.0.

Note the latency difference between order book updates sent on the public path via EOBI and the private response.

Median latencies Release 6.1



Median latencies Release 7.1



# T7 Latency Composition

The below charts show a comparison of latencies since the last update (June 2018) for Eurex futures sent via PS gateways.

Tuning measures for the PS gateway lead to significantly reduced processing times for responses, while introduction of pre trade risk limits led to increased matcher processing times.

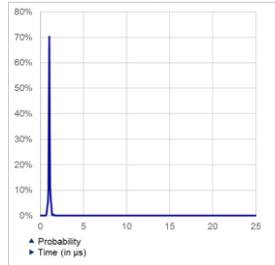
Other latencies are stable since the introduction of PS gateways.

Network response and market data include TCP/UDP stack on the server.

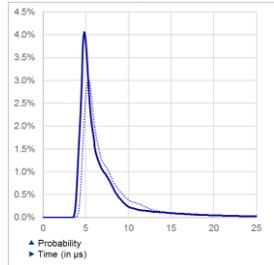
Dotted lines represent data from 19 June 2018 and solid lines data from 18 June 2019.

Request path and market data (EOBI)

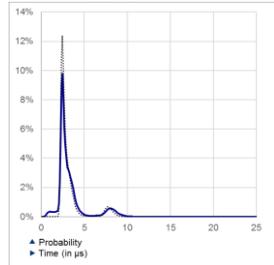
Network request (t<sub>3n</sub> – t<sub>3a</sub>)



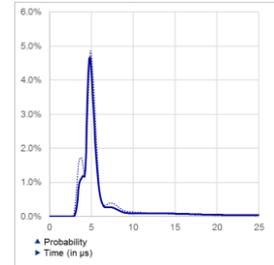
Gateway request (t<sub>3'</sub> – t<sub>3n</sub>)



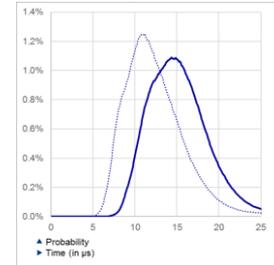
Gateway to matcher (t<sub>5</sub> – t<sub>3'</sub>)



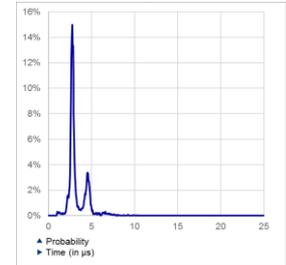
Matcher pre (t<sub>7</sub> – t<sub>5</sub>)



EOBI processing (t<sub>9</sub> – t<sub>7</sub>)

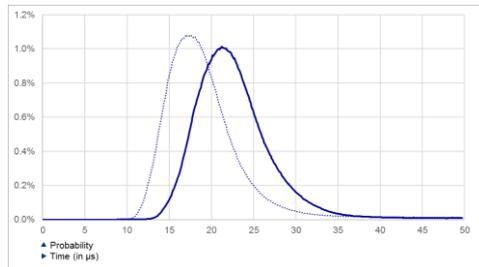


Network EOBI (t<sub>9a</sub> – t<sub>9</sub>)

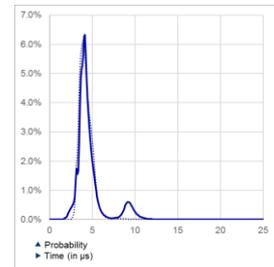


Response path

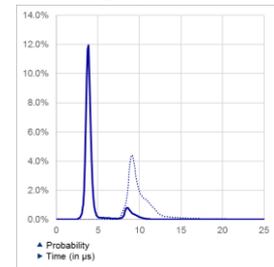
Matcher processing (response) (t<sub>6</sub> – t<sub>7</sub>)



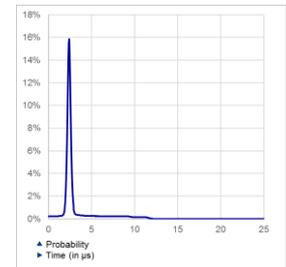
Matcher to gateway (t<sub>4'</sub> – t<sub>6</sub>)



Gateway response (t<sub>4</sub> – t<sub>4'</sub>)



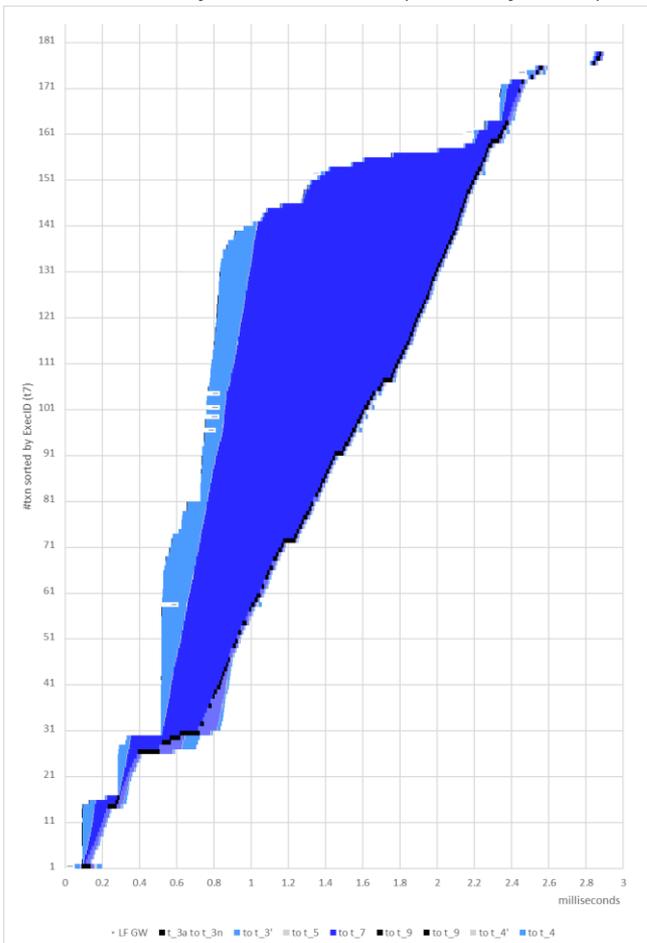
Network (t<sub>4a</sub> – t<sub>4</sub>)



# T7 Latency

## Microburst dynamics

Microburst dynamics FESX (20<sup>th</sup> May 2019)



During micro-bursts, the input into the trading system may be greater than the throughput capabilities. This in turn causes queuing which results in higher latencies.

Higher latencies cause risk (i.e. it takes longer to place/pull an order).

T7 provides real-time performance insights by providing up to seven timestamps with each response and key timestamps with every market data update.

The chart on the left shows the paths

- Co-location 2.0 access layer to t<sub>3(n)</sub> to
- Matching engine in (t<sub>5</sub>) to
- Start matching (t<sub>7</sub>) to
- EOBI SendingTime (t<sub>9</sub>) [where available] to
- Matching engine out (t<sub>6</sub>) to
- Gateway response received (t<sub>4'</sub>) to
- ETI SendingTime (t<sub>4</sub>).

Typical throughput rates (1/ms) are 8000 at t<sub>3n</sub>, ~220 at t<sub>3'</sub>, 100 at t<sub>7</sub> and 150 to 220 at t<sub>4</sub>.

EOBI send times are usually well before the gateway send time of responses.

Note that base latency for requests entered via LF Gateways is ~ 75  $\mu$ s higher (~50 $\mu$ s network + ~25  $\mu$ s LF processing and internal network). Nonetheless, in micro-bursts there is a possibility of requests sent via LF to overtake those that are queuing in the PS Gateway. This usually happens after the first 20-25 PS Gateway requests have been processed by the matching engine.

In order to mitigate this and guarantee first in/first out across gateway types requests sent via low-frequency sessions will be routed via the PS Gateways to the Matching Engine with T7 release 8.0.

# PS gateway - Motivation revisited

## Multiplicity

Latency jitter on parallel inbound paths had incentivized multiplicity (i.e. usage of multiple parallel paths) to reduce latency. This led to higher system load at busy times and thus created higher, less predictable latencies.

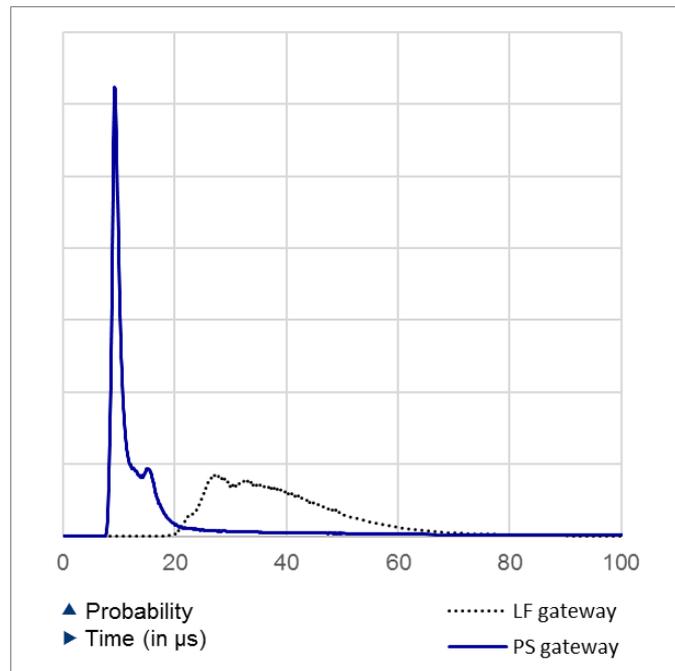
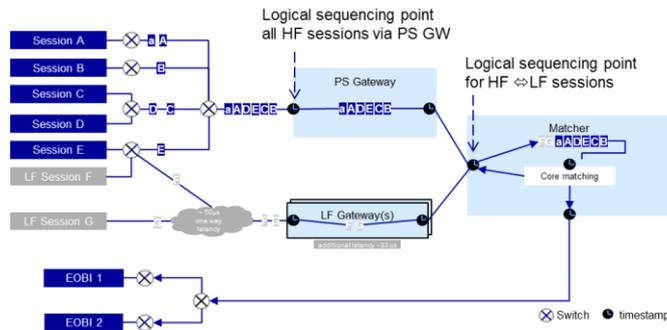
The introduction of a more deterministic network infrastructure (1), first-in-first-out (FIFO) processing of high-frequency gateways (2) and the recent migration to PS gateways as a single (low-latency) point of entry (3) led to a sizable reduction of multiplicity.

Recent competition in the ultra low latency space has raised the ratio of traded vs. executed IOC orders again (see below chart). Note however that the recent increase is not driven by multiplicity – we rather see reactions on many more market data events than before.

**Ratio of sent versus executed IOC orders in Eurex benchmark products**



# PS Gateway versus LF Gateways



Using 10 Gbit/s cross connects in co-location and access via partition-specific ETI gateways in the Equinix facility provides the fastest way for order and quote management in T7.

low-frequency gateways on the other hand allow access to all partitions of a market via a single session.

The base latency of the path to the matcher is more than  $75 \mu\text{s}$  higher for LF gateways,

- $50 \mu\text{s}$  because of a slower network access from customer to the LF gateway (see top left and page 20),
- $25 \mu\text{s}$  because of slower processing of requests by the gateway and a slightly longer path to the matching engine (see bottom left)

In case of microbursts however the number of requests being sent to the PS gateway may be significantly higher than the throughput capabilities.

In those cases requests sent via a lightly loaded LF gateways may overtake those that queue in the PS gateway. This currently happens at around queue position 20-25, i.e. when requests queue behind at least 20-25 messages in the PS gateway.

The graph on the left hand side shows the latency distribution from gateway to matching engine, from  $t3_n$  for PS gateway,  $t3$  for LF gateway to  $t5$  (for partition 3).

Note that with Release 8.0 all ETI requests via LF Gateways will be routed through the PS Gateway and overtaking will no longer be possible.

# Latency

## Comparison of access types

The table below gives an overview of current latency figures of the T7 trading system. All times given are in microseconds.

Network timestamps (t[3489][aa'd]) are synchronized using PPS and white rabbit. The time synch quality between these timestamps is thus <<5ns. Other T7 timestamps are subject to jitter of up to +/- 50 ns, so that t3d\_t3n and t4\_t4d have to be read with a grain of salt.

The latency difference of the inbound path (t\_3a to t\_5) of LF vs PS Gateways is 78 µs for Eurex, 74 µs for Xetra.

Market	Gateway Type	percentile	t3a	t3a'	t3a'	t3d	t3d_t3n	t3a_t3n	t3n_t5	t4_t4d	t4d_t4ap	t4a'	t4a	t4_t4a	t3a_t4a
Eurex	PS	50%	0.27	0.65	0.22	1.15	10.02	1.31	0.68	0.27	2.26	49.3			
	LF	50%					54.83	34.65				42.87	64.8		
	Delta LF - PS	50%					53.7	24.6				40.61	174.2		
							-- 78.3 --								186.4
Xetra	PS	50%	0.27	0.65	0.24	1.16	9.69	1.30	0.68	0.27	2.24	211.3			
	LF	50%					53.14	31.36				43.08	42.5		
	Delta LF - PS	50%					52.0	21.7				40.84	46.9		
							-- 73.7 --								

Eurex	PS	25-75%	0.004	0.004	0.074	0.1	3.7	0.3	0.004	0.004	0.30	21.1			
	LF	25-75%					4.1	9.2				2.37	26.4		
Xetra	PS	25-75%	0.004	0.003	0.052	0.1	2.8	0.271	0.004	0.004	0.27	12.6			
	LF	25-75%					4.5	12.6				1.92	37.1		

# Orders/quotes – detailed performance data

## Our transparency

For the top 15 futures products, daily statistics about the matching engine processing times as well as gateway processing times are provided via the 'Member Section' on Eurex Exchange's website. The ETI round-trip times are calculated based on  $t_4 - t_3$  (gateway SendingTime – gateway application start).

Matcher processing times increased since the introduction of pre trade risk limits. The increase was balanced by a faster gateway processing so that the medians stayed stable since mid 2018. Note that latencies increased significantly for FGBX and FBTP since they were migrated to the FGBL partition in preparation for the introduction of IPS. They will be moved back to their original partitions after release 8.0 goes live.

The table below additionally contains latency figures for DAX equities. All data displayed below refers to 18 Jun 2019.

Product	Product	Matching engine Round-trip times (in $\mu$ s)			Enhanced Trading Interface Round-trip times (all GWs, $t_4 - t_3$ in $\mu$ s)			Enhanced Trading Interface Round-trip times (PS GWs, $t_4 - t_{3n}$ in $\mu$ s)		
		Average	Median	99th percent	Average	Median	99th percent	Average	Median	99th percent
EURO STOXX 50® Index Futures	FESX	88	30	940	165	57	1680	121	52	1250
STOXX® Europe 50 Index Futures	FSTX	73	31	900	132	51	1360	107	52	1250
DAX® Futures	FDAX	66	29	700	112	58	920	100	52	1000
Mini-DAX® Futures	FDXM	49	30	390	86	58	580	75	54	480
MDAX® Futures	F2MX	53	28	760	113	49	1320	84	48	1250
SMI® Futures	FSMI	37	28	213	77	51	479	60	49	254
Euro-Bund Futures	FGBL	119	32	1260	213	85	1660	159	58	1500
Euro-Bobl Futures	FGBM	77	30	840	119	63	920	108	52	1000
Euro-Schatz Futures	FGBS	59	31	580	100	66	580	107	60	1000
Euro-Buxl® Futures	FGBX	73	41	540	221	99	1620	109	72	750
Long-Term Euro-BTP Futures	FBTP	89	37	640	211	120	1420	127	66	1000
Euro-OAT Futures	FOAT	65	32	560	97	59	700	91	58	750
EURO STOXX® Banks Futures	FESB	70	31	780	147	56	1400	102	52	1250
VSTOXX® Futures	FVS	62	32	860	134	60	1360	106	56	1250
STOXX® Europe 600 Index Futures	FXXP	55	29	740	126	51	1360	84	48	1000
DAX® Equities		43	31	258	60	48	272	60	48	272

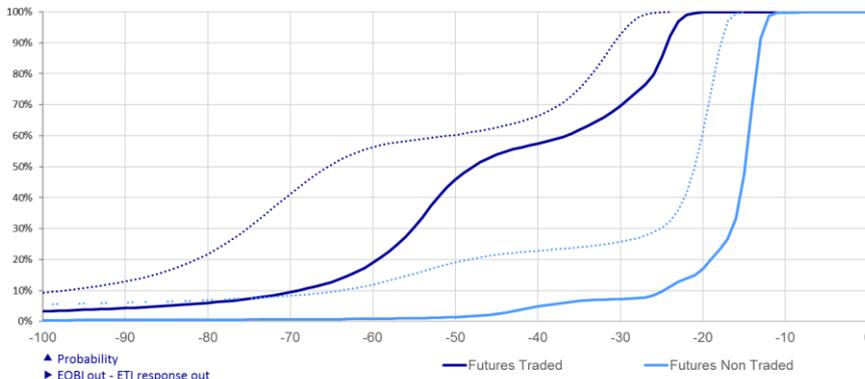
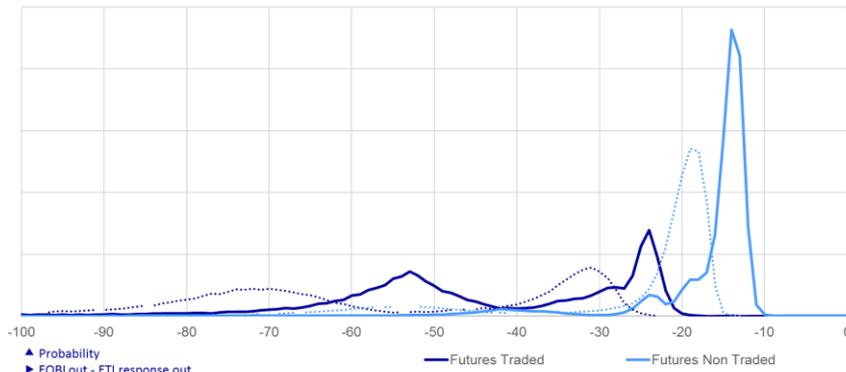
A black and white photograph of a financial market data board. The board consists of a grid of small, illuminated rectangular panels, each displaying a number. The numbers are arranged in rows and columns, with some numbers appearing to be in the process of being updated or highlighted. In the foreground, a person's hand is visible, resting on a surface, possibly a desk or a control panel. The overall scene is dimly lit, with the primary light source being the glowing numbers on the board.

28

Market data

# Trading system dynamics

## Latency characteristics of EOBI versus ETI - Futures



T7 is designed to publish order book updates first on its public data feed.

The diagram shows the time difference distribution between public and private data in microseconds (EOBI first datagram vs ETI responses,  $t_9 - t_4$ ).

The data is a production sample from 13 June 2019 while dashed lines represent data from mid 2018.

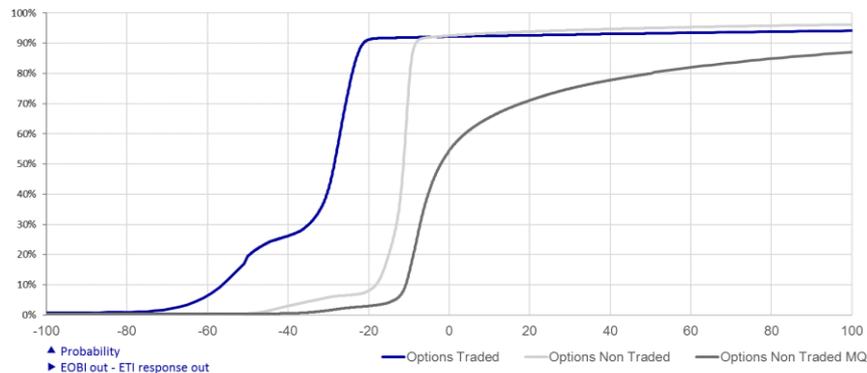
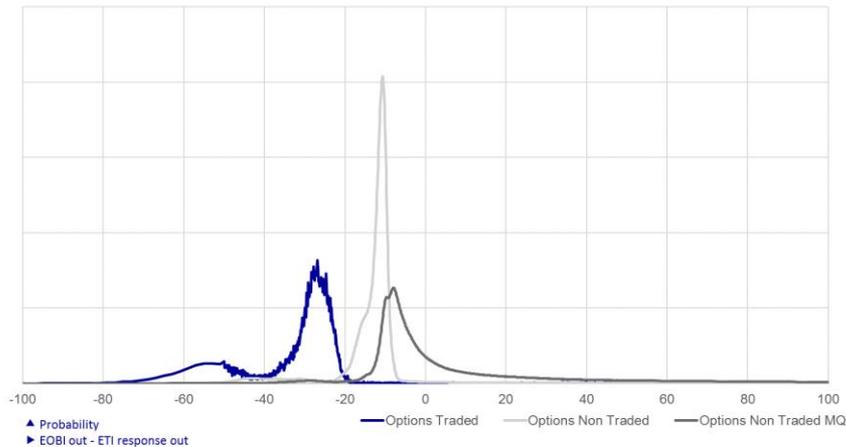
Tuning measures in the PS Gateway outbound path reduced the response latency by 5  $\mu\text{s}$  and limited the number of very slow ETI responses. Note that the second mode in latency is due to the slower LF gateway path.

EOBI market data is still 15  $\mu\text{s}$  faster than the ETI response for order book updates and 25  $\mu\text{s}$  faster for executions.

The first EOBI datagram was faster in more than 99.9 percent of the cases compared to the ETI response and also the first passive ETI book order notification for simple transactions.

# Trading system dynamics

## Latency characteristics of EOBI versus ETI - Options



In Q1 2019 EOBI was enabled for all option instruments.

The data is a production sample from 13 June 2019.

We distinguish between orders leading to a trade, Mass Quotes with more than one quote pair and orders and single quote updates.

Trades and single updates to the orderbook are received first on EOBI in over 90% of the cases with a median latency advantage of 15  $\mu$ s (orders) and 30  $\mu$ s (trades).

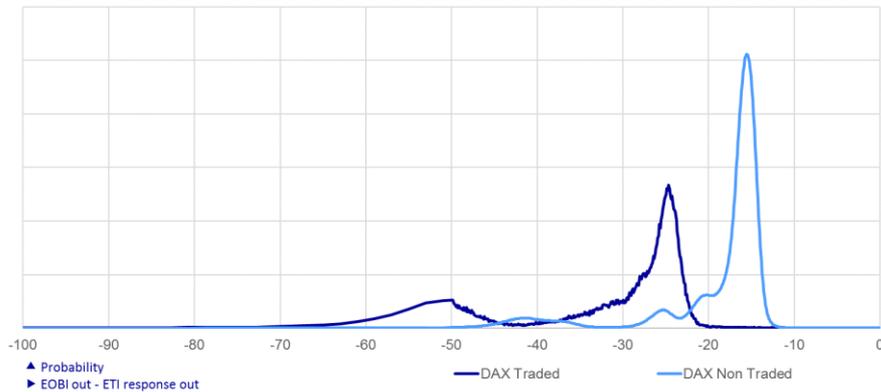
There are two main reason for EOBI delays:

- 1) The transaction is delayed by preceding messages (queues)
- 2) The transaction causes a market maker protection with many quote deletions

The latency profile for Mass Quotes is dominated by larger mass quotes, where the EOBI publisher has to broadcast each quote, leading to longer delays and queues in the EOBI path, while the ETI path only deals with a simple mass quote ack.

# Trading system dynamics

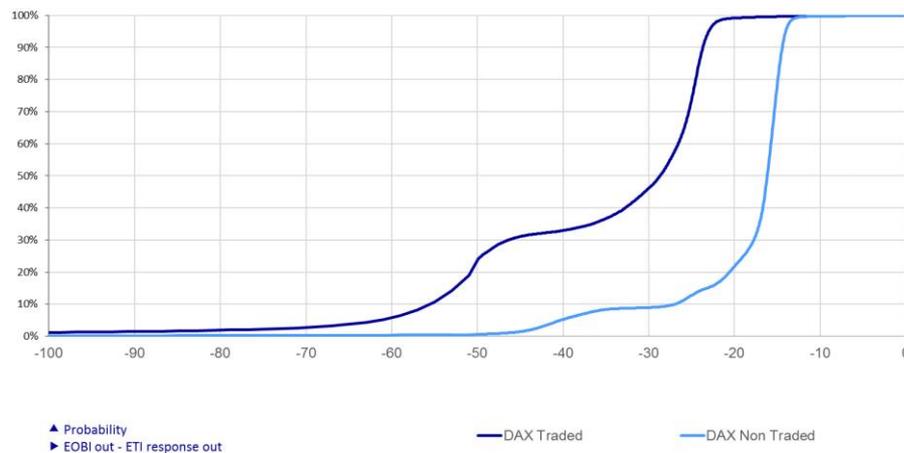
## Latency characteristics of EOBI versus ETI - Xetra



The diagram shows the time difference distribution between public and private data in microseconds for XETRA DAX products (EOBI first datagram vs ETI responses,  $t_9 - t_4$ ).

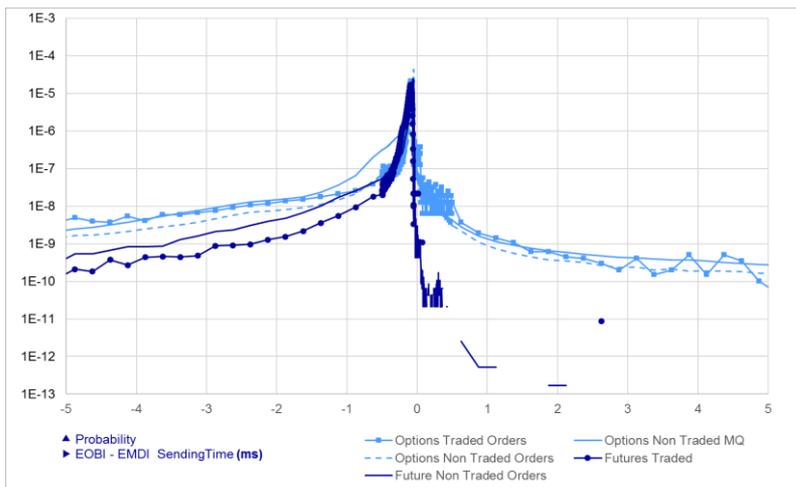
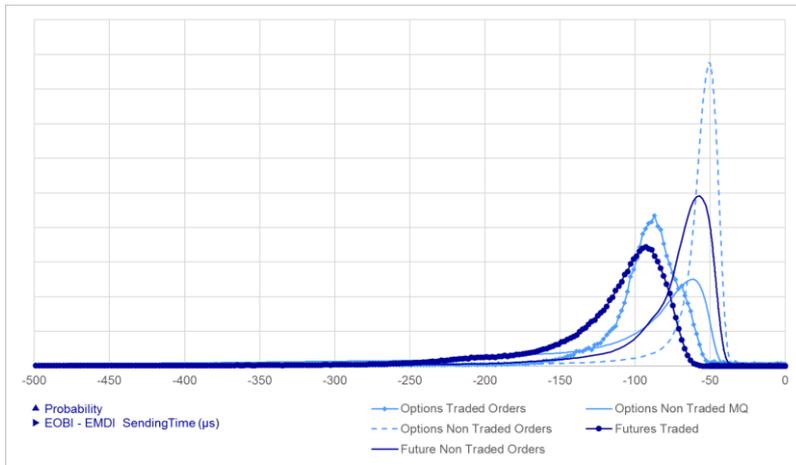
The data is a production sample from 13 June 2019.

The latency distribution is very similar to the Eurex futures, simple orderbook updates are received on EOBI 15  $\mu$ s faster, whereas trades are usually 25  $\mu$ s faster on EOBI.



# Trading system dynamics

## Latency characteristics of EOBI versus EMDI



Market data updates provided via EOBI are almost always faster than EMDI updates.

The diagrams show the distribution of  $t_9$  minus  $t_8$ , i.e. EOBI first datagram versus EMDI sending time for 18<sup>th</sup> June 2019 (negative values: EOBI is faster).

The lower diagram is a log plot of the same distribution (the x-axis scale is in milliseconds). It shows that EMDI is significantly slower especially for mass quotes.

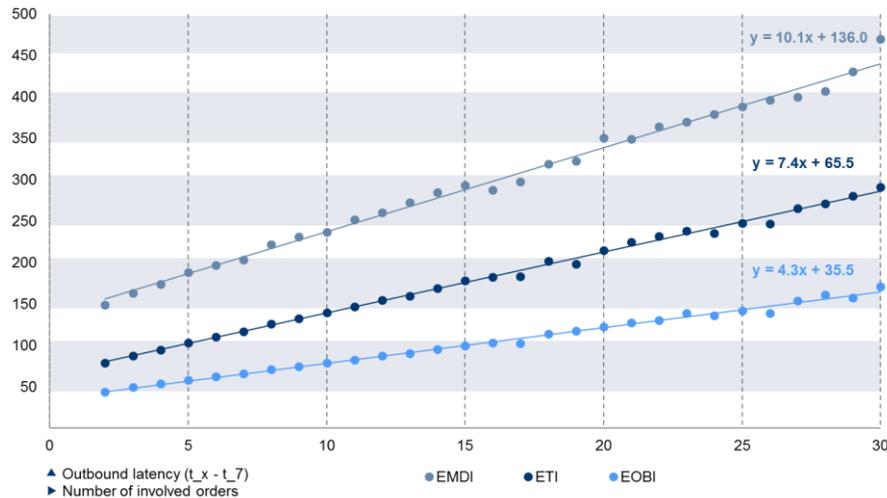
At the same time there are some events where option market data is slower on EOBI. The reason is that EOBI slows down when there is a massive amount of individual updates. The table below shows the delay at various percentiles in  $\mu s$ .

Percentile	Options Traded Orders	Options Non Traded MQ	Options Non Traded Orders	Futures Traded	Future Non Traded Orders
0.1	-12625	-11375	-10625	-3625	-5625
1	-5375	-3625	-2875	-499	-1125
10	-177	-499	-131	-183	-145
25	-107	-241	-63	-131	-87
50	-89	-101	-53	-103	-65
75	-77	-67	-49	-87	-55
90	-65	-55	-45	-77	-50
99	181	109	28	-65	-42
99.9	25375	20625	15375	-59	-38
EOBI faster	98.4%	98.6%	98.8%	100.0%	100.0%

# Trading system dynamics

## Latency characteristics of ETI versus EOBI versus EMDI

Outbound latency for FESX trades (13 Jun 2019)



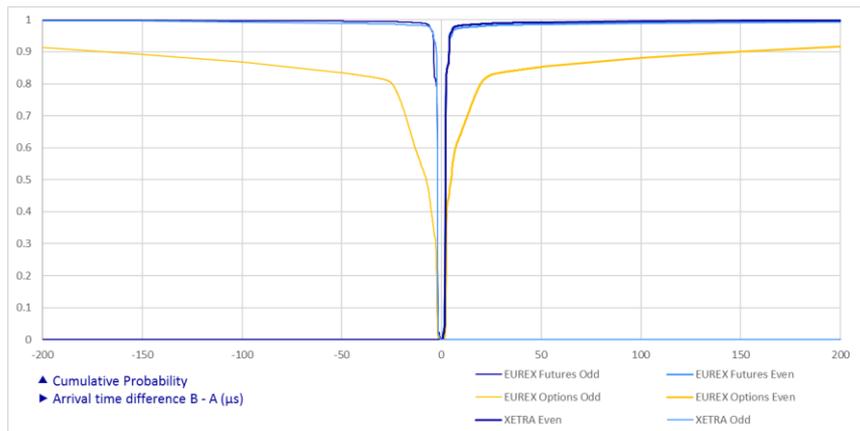
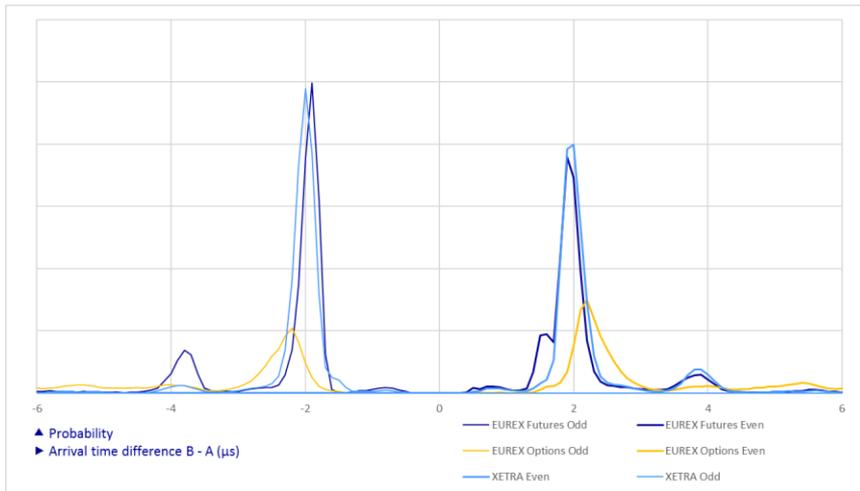
This diagram displays the dependency of the median latency on the complexity of a trade for ETI (t<sub>4</sub>-t<sub>7</sub>), EMDI (t<sub>8</sub>-t<sub>7</sub>) and EOBI (t<sub>9</sub>-t<sub>7</sub>). Note that for ETI we display the gateway sending time of the first passive notification and for EOBI the sending time of the UDP datagram containing the Execution Summary message.

In about 99.99% of all trades, we disseminate order book data on EOBI first (even true for larger trades).

ETI response latency has slightly increased since the migration to PS gateways.

# Market data distribution

## EOBI latency difference of primary and secondary feed



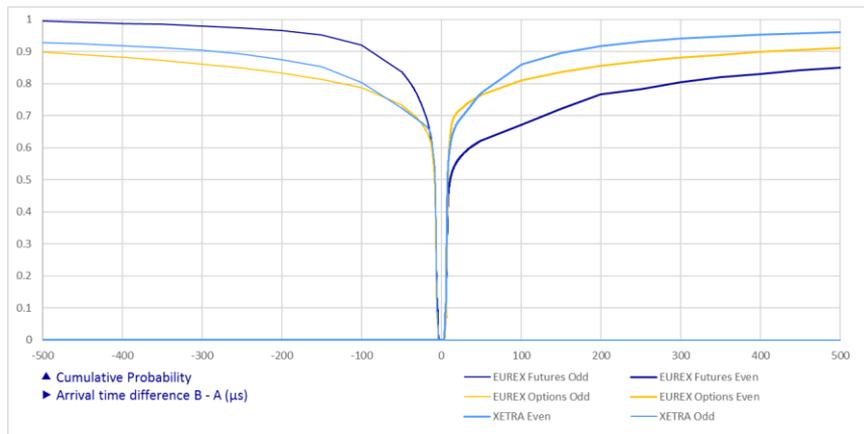
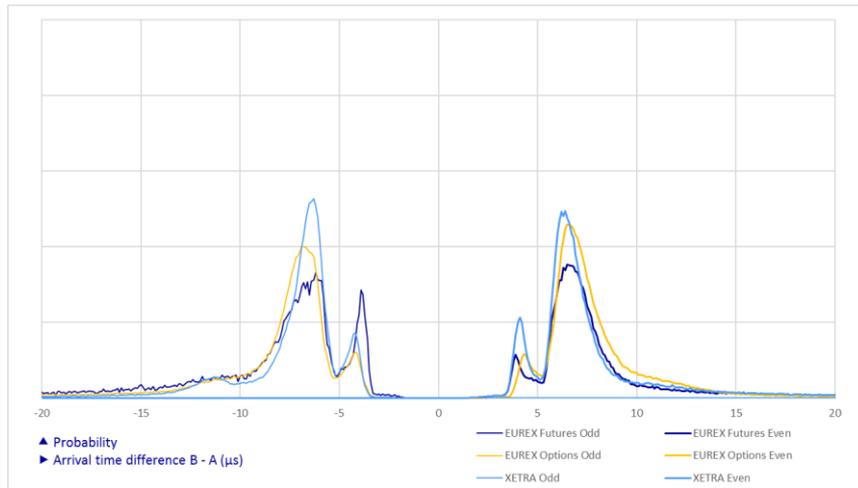
Please note that for products assigned to an even partition, market data is published first on the A stream and then on the B stream whereas, for products assigned to an odd partition market data is published first on the B stream and then on the A stream. The partition ID / product ID is contained in the UDP datagram header of the order book incremental messages and can be used for filtering on UDP datagram level for EMDI / EOBI. Furthermore, a UDP datagram on the T7 EMDI / EOBI order book delta or snapshot channel will only contain data of exactly one product (e.g. EURO STOXX 50® Index Futures).

The data for both primary and secondary streams is disseminated by the same server using two ports connected to the two sides of the network. The process sends the data first on the primary interface. After all datagrams of a transaction are sent it starts sending on the secondary interface.

The median latency difference between the A and the B EOBI incremental feed is about 2 µs for futures and cash products. For options the median is slightly higher and there are far more outliers (i.e. much slower secondary feed) as the datagrams are published on the secondary feed only after all datagrams of a transaction are sent on the primary feed. As a single transaction may contain many datagrams this time varies by complexity of the transaction, i.e. a mass quote with 200 single quote updates will lead to a higher delay than a single order entry.

# Market data distribution

## EMDI latency difference of primary and secondary feed



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 / product ID is contained in the UDP datagram header of the order book incremental messages and can be used for filtering on UDP datagram level for EMDI / EOBI.

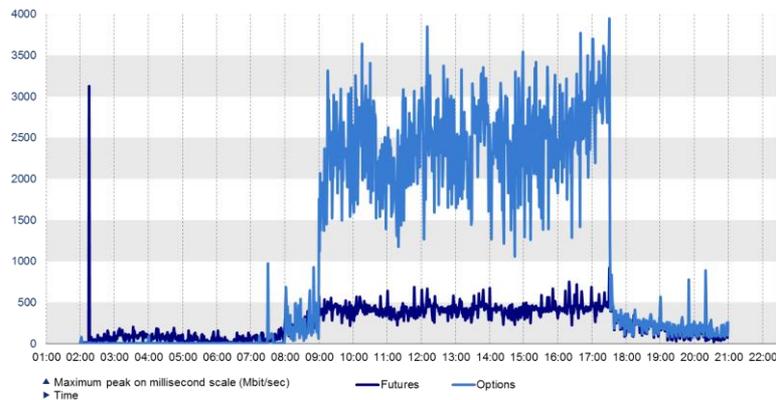
Furthermore, a UDP datagram on the T7 EMDI / EOBI order book delta or snapshot channel contains data of exactly one product (e.g. EURO STOXX 50® Index Futures).

The data for both primary and secondary streams is disseminated by the same server using two ports connected to the two sides of the network. The process sends the data first on the primary interface. After all datagrams of a transaction are sent it starts sending on the secondary interface.

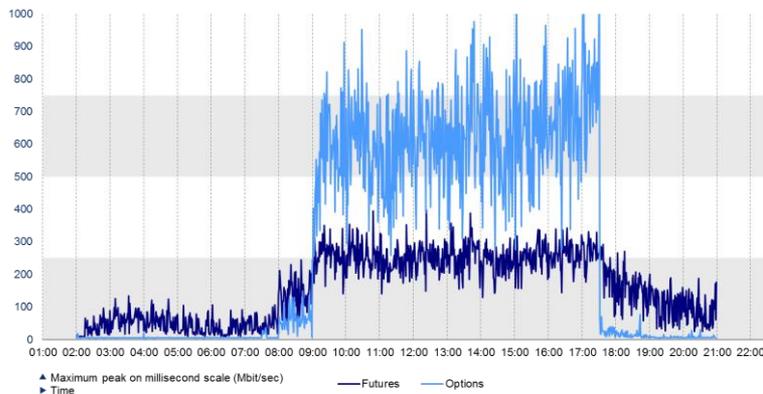
The median latency difference between the A and the B EMDI incremental feed is about 6 to 8 μs for futures and cash products. For options the median is slightly higher and there are far more outliers (i.e. much slower secondary feed) as the datagrams are published on the secondary feed only after all datagrams of a transaction are sent on the primary feed. As a single transaction may contain many datagrams this time varies by complexity of the transaction, i.e. a mass quote with 200 single quote updates will lead to a higher delay than a single order entry.

# Eurex: Market data volume

## Eurex EOBI



## Eurex EMDI



Each data point equals the maximum bandwidth produced on a 1 millisecond scale by the incremental A stream in Mbit/s.

The provided data shows one data point per minute for 16 May 2019.

Enhanced Order Book Interface (EOBI) peak volume is significantly higher than price level aggregated data volume EMDI. EOBI market data is therefore currently only available to Trading Participants using 10 Gbit/s connections.

The EOBI for options incremental data stream peaks around 4 Gbit/s on millisecond level, while the futures stream peaks at 1 Gbit/s outside market opening.

Participants that want to receive data for Eurex Exchange's products on EMDI with less than 1 ms queuing delays need to use a connection with a bandwidth of more than 1 Gbit/s (options) or 400 Mbit/s (futures) respectively.

Trading Participants are advised to use two 10 Gbit/s connections (one for each market data stream) in co-location to receive market data.

# Xetra: Market data volume

Each data point equals the maximum bandwidth produced on a 1 millisecond scale by the incremental B stream in Mbit/s.

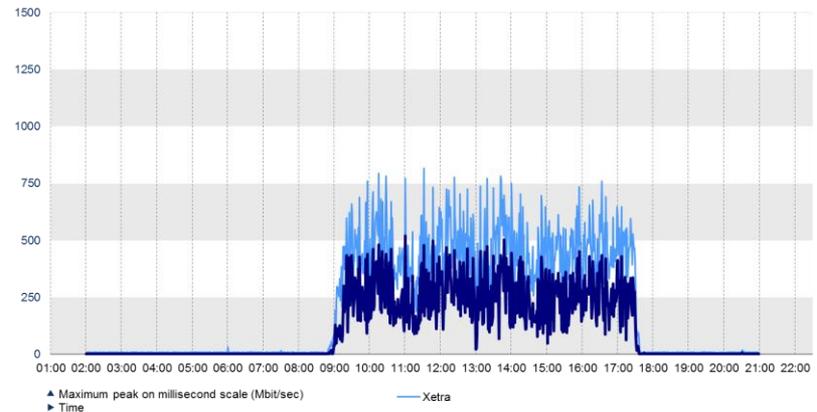
The provided data shows one data point per minute for 16 May 2019.

Market Data peak data volume can be a significantly higher on high volume trading days. Hence Participants that want to receive data with less than 1 ms queuing delays need to use a connection with a bandwidth of at least 800 Mbit/s (EMDI, All products) or 500 Mbit/s (EMDI, DAX<sup>®</sup> equities only).

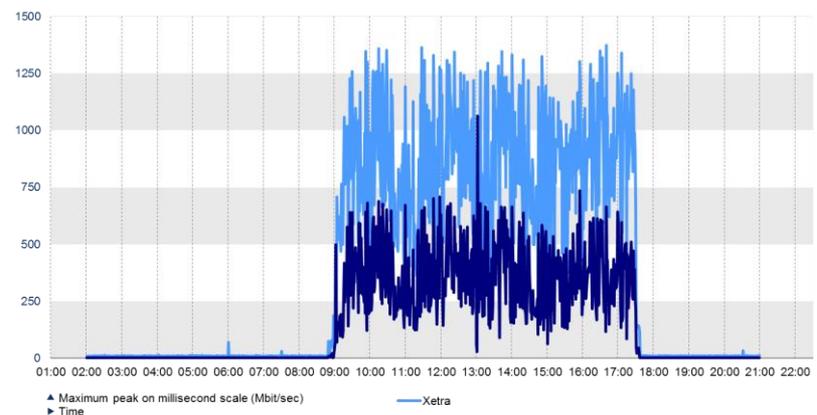
Enhanced Order Book Interface market data is currently only available to Trading Participants using 10 Gbit/s connections.

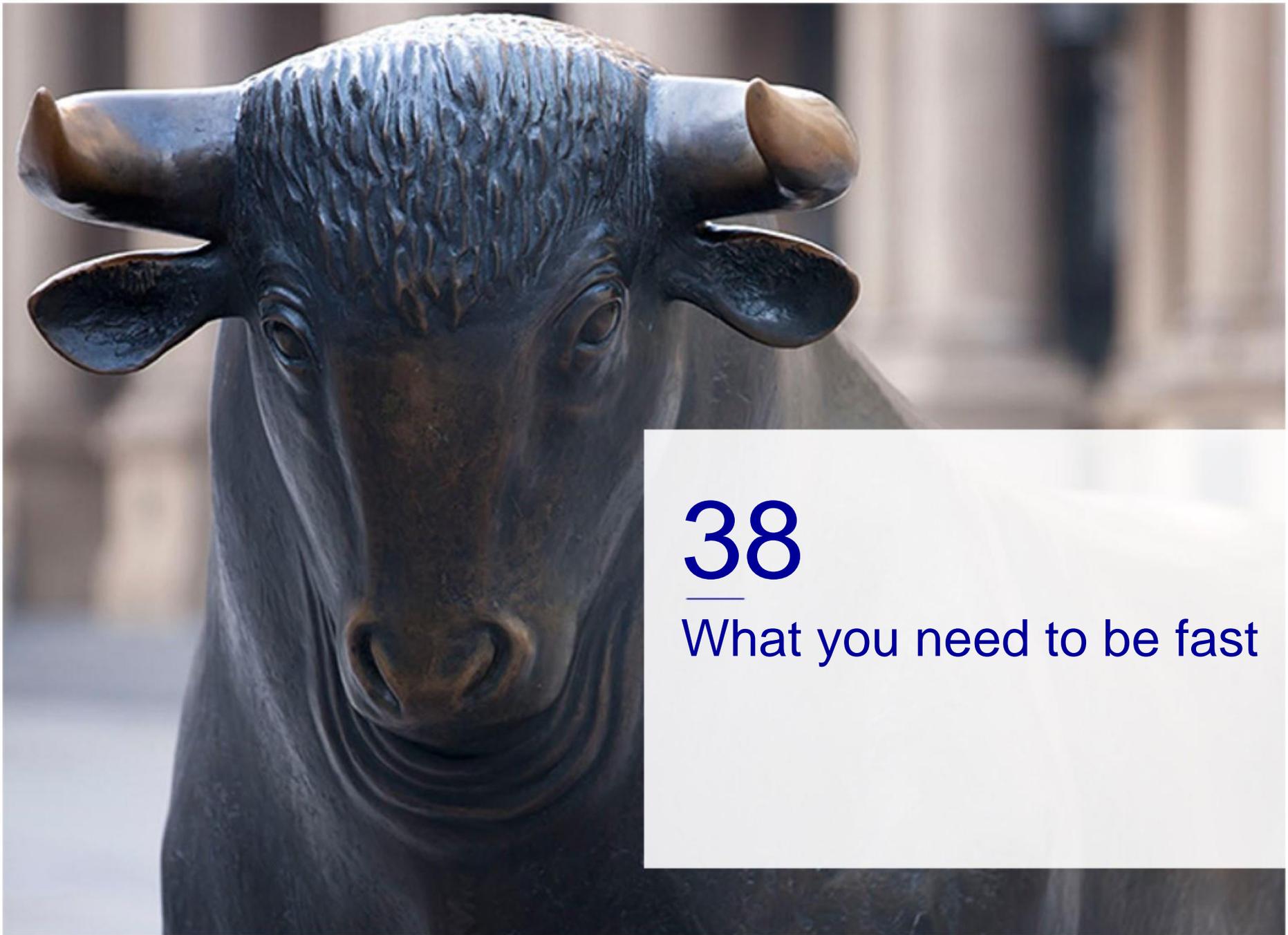
Trading Participant are advised to take two cross connects (one for each market data stream) in co-location to receive market data.

### Xetra EMDI



### Xetra EOBI





38

What you need to be fast

# What you need to be fast...

## A few recommendations to achieve the low latency

Use the Equinix co-location facility to be close to Deutsche Börse T7.

Use state-of-the-art switches (if any) and only have at most one hop between the exchange network and your server.

Use good network interface cards and TCP/IP acceleration, e.g. a kernel-by-pass library.

Use at least two 10 Gbit/s (cross-) connections in co-location for EOBI market data and two 10 Gbit/s connections for T7 ETI.

Use PS gateways and make sure you use the cross connect on the same side as the gateway you are connecting to (compare time-to-live values in the IP header in the responses from both sides).

Measure and analyze your own timestamps (e.g. the reaction time as recommended on the next slide).

Use state of the art time synchronization, i.e. GPS clocks and a high quality time distribution. The PTP signal you can get from us has a quality of +/- 50 ns. For our network timestamps we use the White Rabbit protocol and PPS breakouts. You can connect to our white rabbit time service providing you a time synchronization quality of 1-2 ns max, see <https://www.deutsche-boerse.com/dbg-en/products-services/ps-technology/ps-connectivity-services/ps-connectivity-services-time-services>

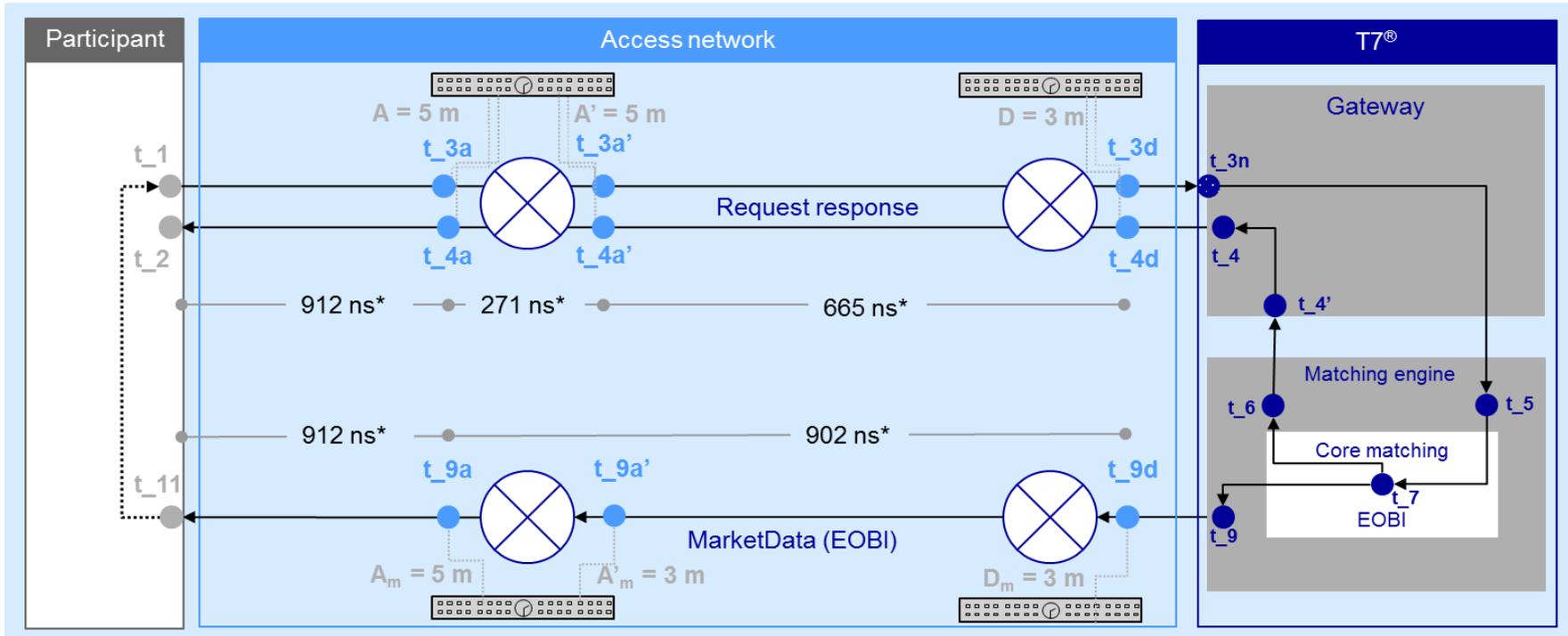
We provide highly accurate network timestamps of all orders leading to a market data update via the high precision timestamp file service, see <https://datashop.deutsche-boerse.com/high-precision-timestamps>.

Use the EOBI Execution Summary for fast trading decisions and position keeping (passive executions). For a consistent order book, all incremental updates following the Execution Summary should always be processed.

Trade notifications need to be processed to create safety. We recommend to use either a low-frequency ETI session or a FIX trade capture drop copy to confirm the fast execution information provided by the execution reports via high-frequency sessions.

# What you need to be fast...

## Participant reaction time measurement



Measure the time between market data reception ( $t_{10}/t_{11}$ ) and your reaction ( $t_1$ ), take note of aggressor in timestamps ( $t_{3n}$ ) of executions.

$t_{3a}$  and  $t_{9d}$  are available via the high precision timestamp file service, see <https://datashop.deutsche-boerse.com/high-precision-timestamps>  
Take our white rabbit signal to compare your timestamps with ours with ns accuracy <https://www.deutsche-boerse.com/dbg-en/products-services/ps-technology/ps-connectivity-services/ps-connectivity-services-time-services#tab-1556860-1556864>

41

Appendix



# Enhanced Order Book Interface

## Exceptions to fast trading decisions based on the Execution Summary

The Execution Summary at the beginning of an EOBI packet can be used for fast trading decisions (and passive position keeping) in the majority of cases. However there are certain scenarios where this may lead to a wrong perception of the order book on client side. This includes cases where:

The RequestTime field is not set:

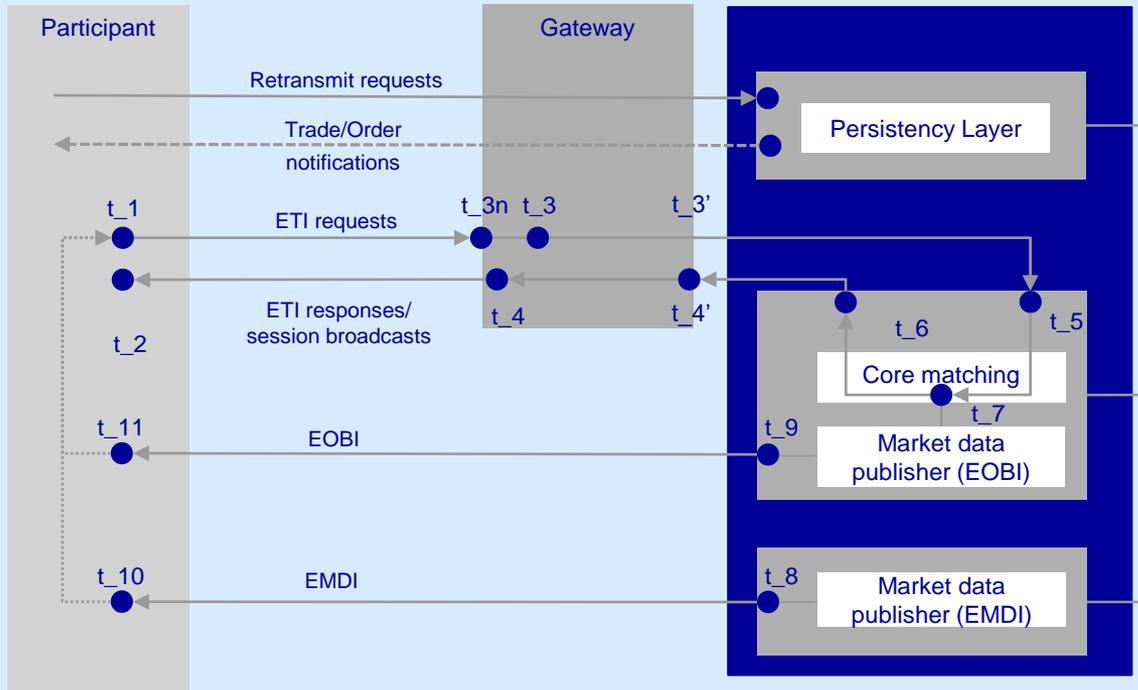
This is for example in case of more than one market order being triggered by an incoming order/quote. In such a scenario, there will be one Execution Summary sent for each market order. The Execution Summary for the first market order will be at the beginning of the packet but the Execution Summaries for the other market orders will follow in the same or in the next packet(s) before CompletionIndicator is set to 1.

The Implied flag is set:

In case of synthetic matching, the LastQty shows the total matched quantity that result from all involved instruments' order books. At the same time only the instrument of the aggressing order is given. In order to have correct order books, Participants have to process all incremental updates following the Execution Summary. There is a potential shortcut in this case which is explained on the next two slides.

In equity markets the execution of hidden quantity of Iceberg orders is currently not reflected in the hiddenQty field of the EOBI Execution summary. This field will be used with introduction of T7 release 8.0.

# T7<sup>®</sup> topology



## Matching engine:

- order book maintenance & execution
- creation of direct responses as well as execution messages for passive orders/quotes
- creation of EOBI order book messages
- creation of EOBI order book snapshot messages

## Persistency:

- persistent order storage
- trade/execution history
- transaction history for standard orders
- creation of listener broadcast for standard orders

## Market Data (EMDI):

- creation of order book delta messages
- creation of order book snapshot messages

## Description of timestamps

t\_1, t\_2 can be taken by a participant (e.g. via a network capture) when a request/ response is read from/written to the network.

t\_3n time taken by the ETI gateway when the first bit of a request arrives on the PS gateway NIC; contained in (private) ETI response for transactions sent via PS gateways.  
Consecutive messages via the same session may be assigned to the same t\_3n in rare cases.

t\_3 time taken by the ETI gateway application when a request is read from the socket on the Participant's side of the gateway; contained in (private) ETI response for transactions sent via LF gateways.

t\_4' time taken by the ETI gateway when a response/ notification is received by the ETI gateway from the matching engine; contained in (private) ETI response/ notification.

t\_4 time taken by the ETI gateway when a response/ notification is written to the socket on the Participant's side of the gateway; contained in (private) ETI response/ notification.

t\_5, t\_6 time taken by the matching engine when a request/response is read/written; contained in (private) ETI response.

t\_7 time at which the matching engine starts maintaining the order book

t\_8 time taken by EMDI publisher just before the first respective UDP datagram is written to the UDP socket.

t\_9 time taken by EOBI publisher just before the first respective UDP datagram is written to the UDP socket.

t\_10, t\_11 can be taken by a Participant (e.g. via a network capture) when a UDP datagram is read from the UDP socket.

t\_<x>a, t\_<x>a', t\_<x>d: time taken by network capture devices in the access and distribution layers.

# T7 timestamp reference

## The timestamps t\_3,...,t\_9 are available via the following fields:

t_3, t_3n:	Tag	5979	("RequestTime")	in the T7 ETI Response in the T7 EMDI Depth Incremental message, in case a trade is reported in the T7 EOBI Execution Summary message
t_4':	Tag	7765	("ResponseIn")	in the T7 ETI Response (from the matching engine)
	Tag	25043	("NotificationIn")	in the T7 ETI Notification (from the matching engine)
t_4:	Tag	52	("SendingTime")	in the T7 ETI Response and Notification
t_5:	Tag	21002	("TrdRegTSTimeIn")	in the T7 ETI Response (from the matching engine)
	Tag	21002	("TrdRegTSTimeIn")	in the T7 EOBI Order Add, Order Modify, Order Modify Same Priority and Order Delete messages, to be replaced by RequestTime with T7 release 8.0 (R8.0)
	Tag	28820	("AggressorTimestamp")	in the T7 EMDI Depth Incremental message, in case a trade is reported in the T7 EOBI Execution Summary message, to be replaced by RequestTime in R8.0
t_6:	Tag	21003	("TrdRegTSTimeOut")	in the T7 ETI Response and Notification (from the matching engine)
t_7:	Tag	17	("ExecID")	in the T7 ETI Response (from the matching engine) in the T7 EOBI Execution Summary message
	Tag	273	("MDEntryTime")	in the T7 EMDI Depth Incremental message
	Tag	21008	("TrdRegTSTimePriority")	in the T7 EOBI Order Add and Order Modify messages
	Tag	60	("TransactTime")	in the T7 EOBI Order Modify Same Priority and Order Delete messages
t_8:	no Tag		("SendingTime")	in the T7 EMDI UDP packet header
t_9:	Tag	60	("TransactTime")	in the T7 EOBI packet header
(t_8-t_5):	no Tag		("PerformanceIndicator")	in the T7 EMDI UDP packet header of the T7 EMDI Depth Incremental stream

### Notes on timestamps:

All timestamps provided are 8 byte integers (in nanoseconds after Unix epoch).

The PerformanceIndicator is a 4 byte integer (in nanoseconds).

The Network timestamps (t\_?a, t\_?a', and t\_?d) are not available in any field.



# Thank you for your attention

## Contact

**Sebastian Neusüß, Andreas Lohr & Manfred Sand**

E-Mail [monitoring@deutsche-boerse.com](mailto:monitoring@deutsche-boerse.com)

For updates refer to [http://www.eurexchange.com/exchange-en/technology/high-frequency\\_trading](http://www.eurexchange.com/exchange-en/technology/high-frequency_trading)

<http://www.xetra.com/xetra-en/technology/t7/publications>

© Deutsche Börse 2019

Deutsche Börse AG (DBAG), Clearstream Banking AG (Clearstream), Eurex Frankfurt AG, Eurex Clearing AG (Eurex Clearing) as well as Eurex Bonds GmbH (Eurex Bonds) and Eurex Repo GmbH (Eurex Repo) are corporate entities and are registered under German law. Eurex Zürich AG is a corporate entity and is registered under Swiss law. Clearstream Banking S.A. is a corporate entity and is registered under Luxembourg law. U.S. Exchange Holdings, Inc. is a corporate entity and is registered under U.S. American law. Deutsche Boerse Asia Holding Pte. Ltd., Eurex Clearing Asia Pte. Ltd. and Eurex Exchange Asia Pte. Ltd are corporate entities and are registered under Singapore law. Eurex Frankfurt AG (Eurex) is the administrating and operating institution of Eurex Deutschland. Eurex Deutschland and Eurex Zürich AG are in the following referred to as the "Eurex Exchanges".

All intellectual property, proprietary and other rights and interests in this publication and the subject matter hereof (other than certain trademarks and service marks listed below) are owned by DBAG and its affiliates and subsidiaries including, without limitation, all patent, registered design, copyright, trademark and service mark rights. While reasonable care has been taken in the preparation of this publication to provide details that are accurate and not misleading at the time of publication DBAG, Clearstream, Eurex, Eurex Clearing, Eurex Bonds, Eurex Repo as well as the Eurex Exchanges and their respective servants and agents (a) do not make any representations or warranties regarding the information contained herein, whether express or implied, including without limitation any implied warranty of merchantability or fitness for a particular purpose or any warranty with respect to the accuracy, correctness, quality, completeness or timeliness of such information, and (b) shall not be responsible or liable for any third party's use of any information contained herein under any circumstances, including, without limitation, in connection with actual trading or otherwise or for any errors or omissions contained in this publication.

This publication is published for information purposes only and shall not constitute investment advice respectively does not constitute an offer, solicitation or recommendation to acquire or dispose of any investment or to engage in any other transaction. This publication is not intended for solicitation purposes but only for use as general information. All descriptions, examples and calculations contained in this publication are for illustrative purposes only.

Eurex and Eurex Clearing offer services directly to members of the Eurex exchanges respectively to clearing members of Eurex Clearing. Those who desire to trade any products available on the Eurex market or who desire to offer and sell any such products to others or who desire to possess a clearing license of Eurex Clearing in order to participate in the clearing process provided by Eurex Clearing, should consider legal and regulatory requirements of those jurisdictions relevant to them, as well as the risks associated with such products, before doing so.

Eurex derivatives are currently not available for offer, sale or trading in the United States or by United States persons (other than EURO STOXX 50® Index Futures, EURO STOXX 50® ex Financials Index Futures, EURO STOXX® Select Dividend 30 Index Futures, EURO STOXX® Index Futures, EURO STOXX® Large/Mid/Small Index Futures, STOXX® Europe 50 Index Futures, STOXX® Europe 600 Index Futures, STOXX® Europe 600 Banks/Industrial Goods & Services/Insurance/Media/Travel & Leisure/Utilities Futures, STOXX® Europe Large/Mid/Small 200 Index Futures, Dow Jones Global Titans 50 IndexSM Futures (EUR & USD), DAX®/Mini-DAX®/MDAX®/TecDAX® Futures, SMIM® Futures, SLI Swiss Leader Index® Futures, MSCI World (FMWO, FMWP, FMWN)/Europe (FMEU, FMEP)/ Europe Value/Europe Growth/Emerging Markets (FMEM, FMEF, FMEN)/Emerging Markets Latin America/Emerging Markets EMEA/Emerging Markets Asia/China Free/India/Japan/Malaysia/South Africa/Thailand/AC Asia Pacific ex Japan Index Futures, TA-25 Index Futures, Daily Futures on TAIEX Futures, VSTOXX® Futures, Gold and Silver Futures as well as Eurex FX, property and interest rate derivatives).

#### Trademarks and Service Marks

Buxl®, DAX®, DivDAX®, eb.rexx®, Eurex®, Eurex Bonds®, Eurex Repo®, Eurex Strategy WizardSM, Euro GC Pooling®, FDAX®, FWB®, GC Pooling®, GCPI®, MDAX®, ODAX®, SDAX®, TecDAX®, USD GC Pooling®, VDAX®, VDAX-NEW® and Xetra® are registered trademarks of DBAG. All MSCI indexes are service marks and the exclusive property of MSCI Barra. ATX®, ATX® five, CECE® and RDX® are registered trademarks of Vienna Stock Exchange AG. IPD® UK Annual All Property Index is a registered trademark of Investment Property Databank Ltd. IPD and has been licensed for the use by Eurex for derivatives. SLI®, SMI® and SMIM® are registered trademarks of SIX Swiss Exchange AG. The STOXX® indexes, the data included therein and the trademarks used in the index names are the intellectual property of STOXX Limited and/or its licensors. Eurex derivatives based on the STOXX® indexes are in no way sponsored, endorsed, sold or promoted by STOXX and its licensors and neither STOXX nor its licensors shall have any liability with respect thereto. Dow Jones is a service mark of Dow Jones & Company, Inc. All derivatives based on these indexes are not sponsored, endorsed, sold or promoted by Dow Jones & Company, Inc. Dow Jones & Company, Inc. does not make any representation regarding the advisability of trading or of investing in such products. Bloomberg Commodity IndexSM and any related sub-indexes are service marks of Bloomberg L.P. All references to London Gold and Silver Fixing prices are used with the permission of The London Gold Market Fixing Limited as well as The London Silver Market Fixing Limited, which for the avoidance of doubt has no involvement with and accepts no responsibility whatsoever for the underlying product to which the Fixing prices may be referenced. PCS® and Property Claim Services® are registered trademarks of ISO Services, Inc. Korea Exchange, KRX, KOSPI and KOSPI 200 are registered trademarks of Korea Exchange Inc. Taiwan Futures Exchange and TAIFEX are registered trademarks of Taiwan Futures Exchange Corporation. Taiwan Stock Exchange, TWSE and TAIEX are the registered trademarks of Taiwan Stock Exchange Corporation. BSE and SENSEX are trademarks/service marks of Bombay Stock Exchange (BSE) and all rights accruing from the same, statutory or otherwise, wholly vest with BSE. Any violation of the above would constitute an offence under the laws of India and international treaties governing the same. The names of other companies and third party products may be trademarks or service marks of their respective owners.