DEUTSCHE BÖRSE GROUP

Open Day 2018

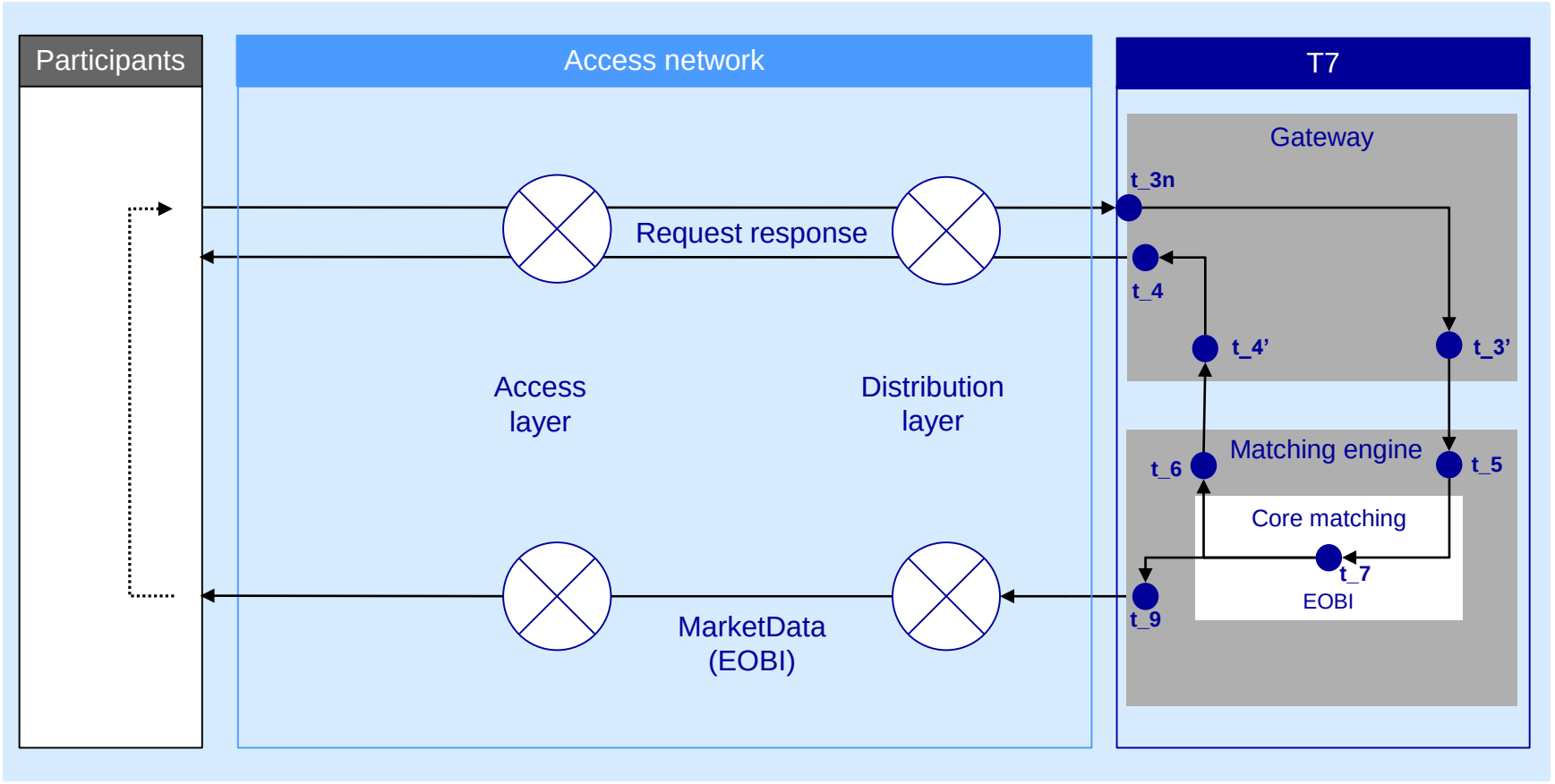
T7® infrastructure and latency

Andreas Lohr and Sebastian Neusüß

27 September 2018

Contents

T7® topology



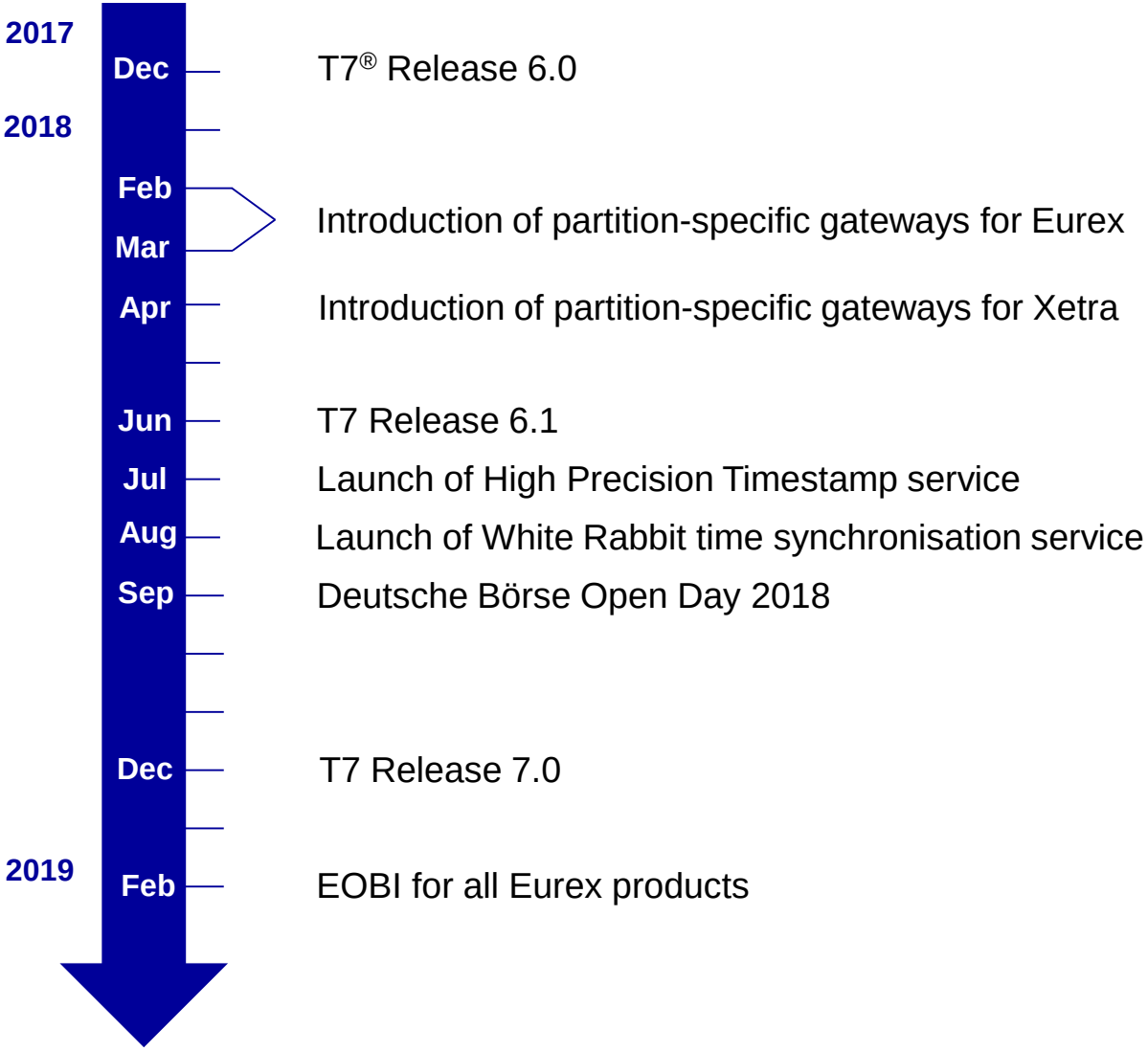
- Timestamps provided in T7 API (in real time) in dark blue (t_3n: taken by network card, other: application level)
- Cisco 3548X switches operating in cut-through mode.

3

Developments since last Open Day

Developments since last Open Day

Timeline

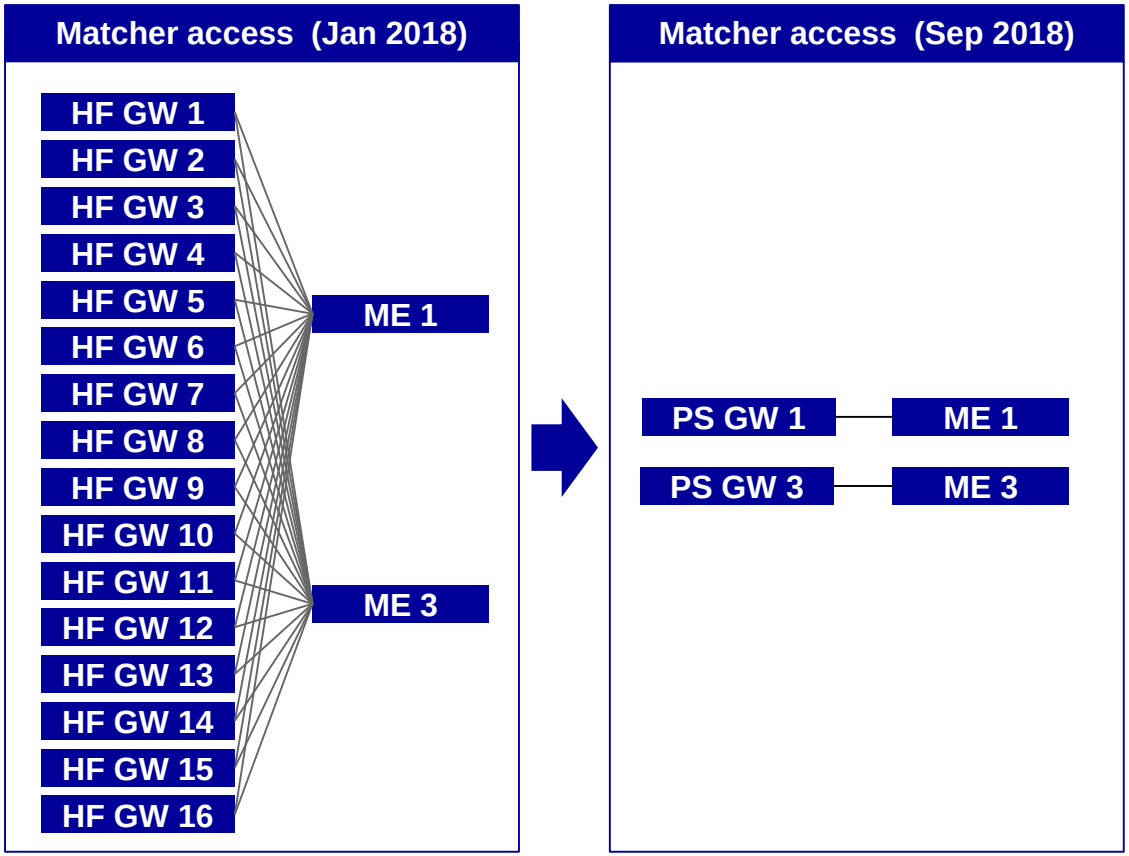


Developments since last Open Day

Partition-specific (PS) gateway – motivation revisited

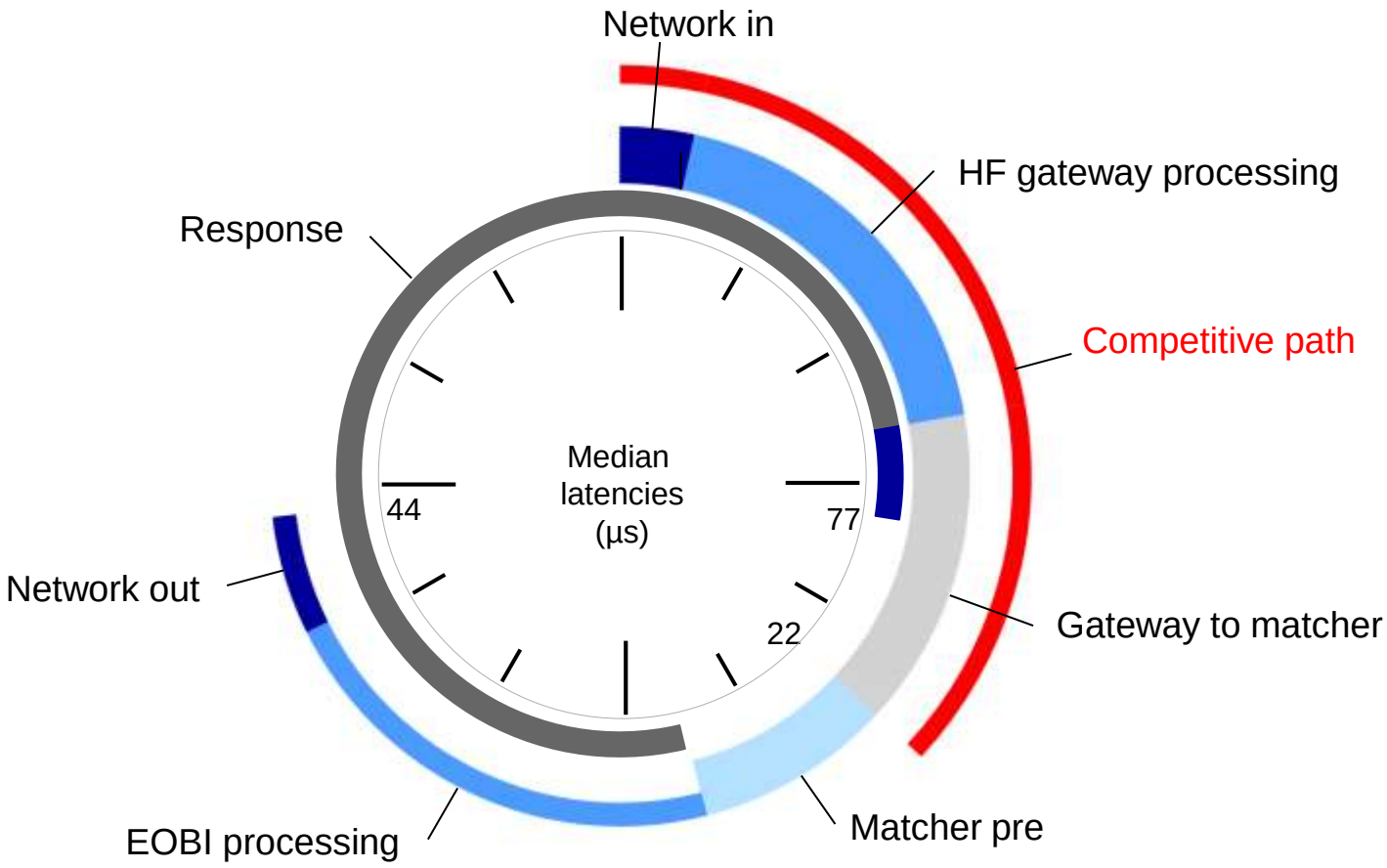
PS gateway was introduced as a single point of (low latency) entry to the exchange in order to …

- enhance determinism.
- simplify set-up.
- reduce multiplicity.



Developments since last Open Day

The micro clock – January 2018

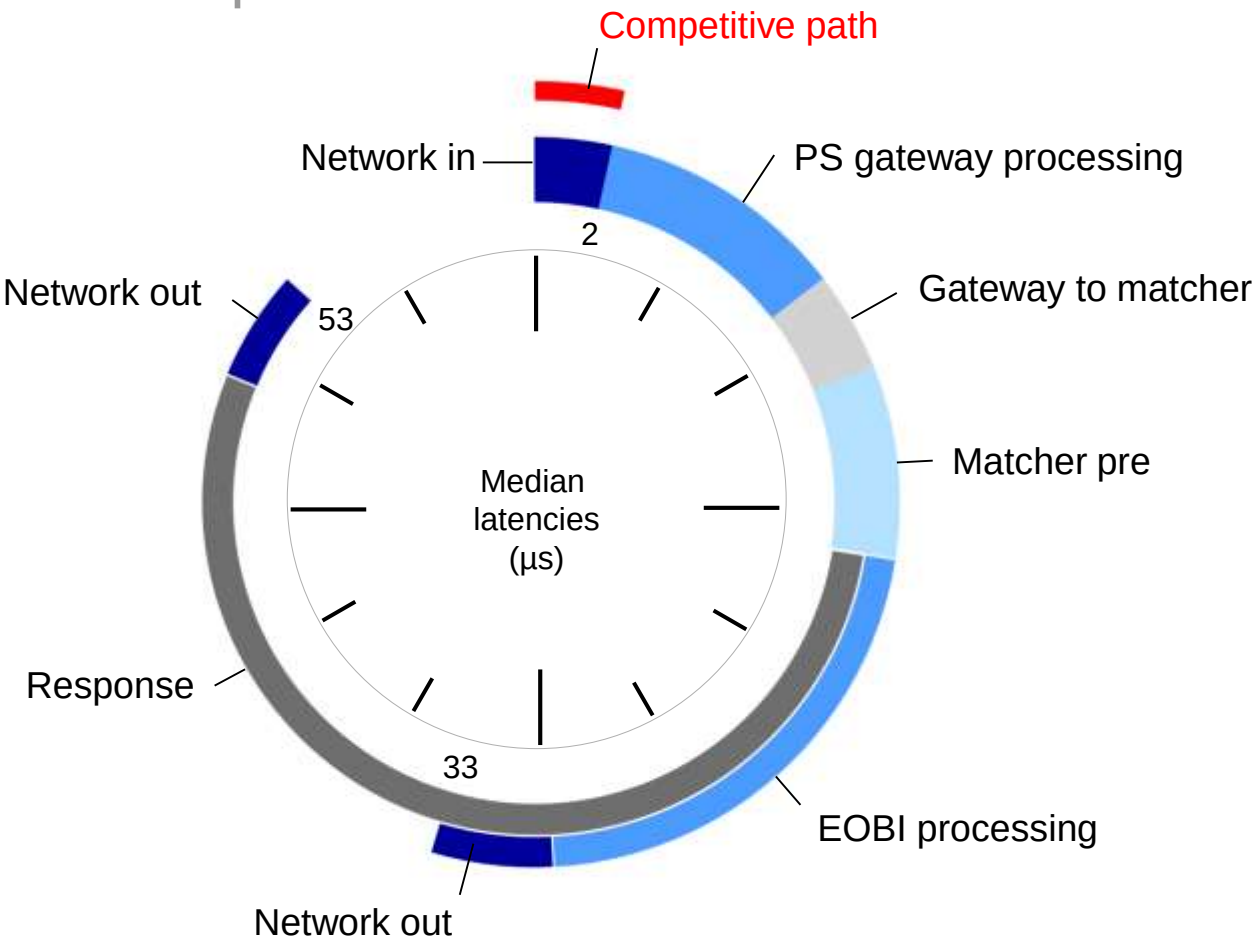


Competitive path of 22 µs with multiple µs variance

16 high-frequency gateways

Developments since last Open Day

The micro clock – September 2018



Competitive path shortened from 22 to sub 2 µs with ns level variance

Overall latency reduced by 25 per cent, public first principle untouched

Developments since last Open Day

T7® latency composition

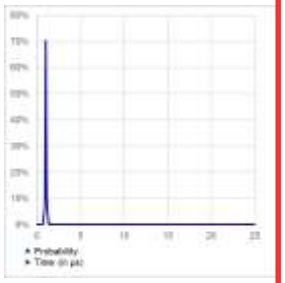
The charts below show a comparison of latencies for Eurex futures sent via HF/PS gateways.

Dark blue are recent figures, light blue are from January 2018.

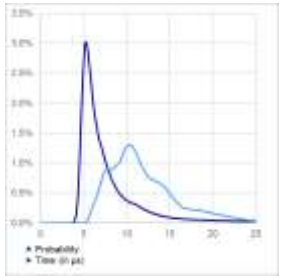
Most noticeable is the reduction of gateway processing and internal network times.

Request and market data (EOBI)

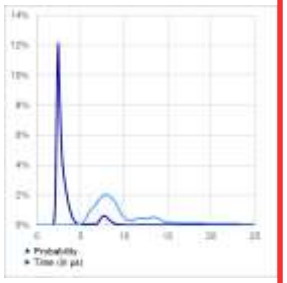
Network request (t_3a to t_3n)



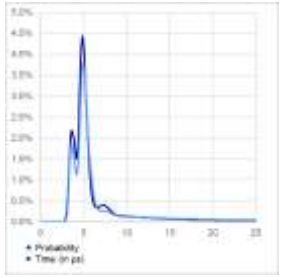
Gateway request (t_3n to t_3')



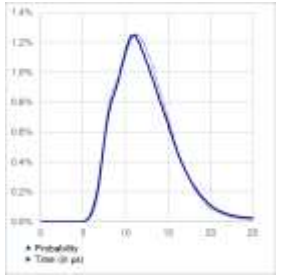
Gateway to matcher (t_3' to t_5)



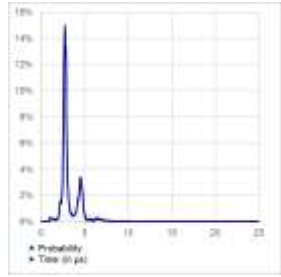
Matcher pre (t_5 to t_7)



EOBI processing (t_7 to t_9)

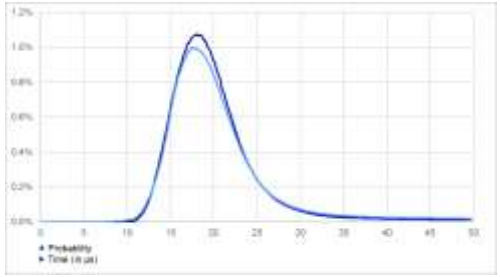


Network EOBI (t_9 to t_9a)

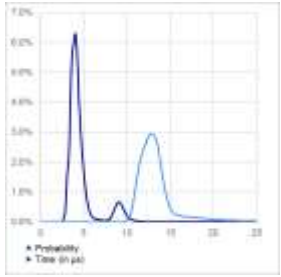


Response path

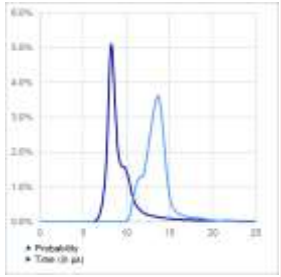
Matcher processing (response) (t_7 to t_6)



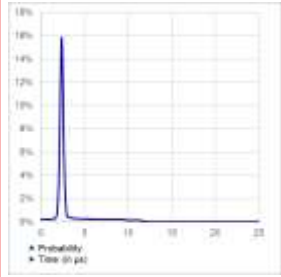
Matcher to gateway (t_6 to t_4')



Gateway response (t_4' to t_4)



Network (t_4 to t_4a)



Developments since last Open Day

Multiplicity

Latency jitter on parallel inbound paths had incentivised multiplicity 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 for high-frequency gateways (2) and introduction of PS gateways as a single (low-latency) point of entry (3) led to a reduction of multiplicity.

Ratio of sent vs executed IOCs in Eurex benchmark futures

Developments since last Open Day
Problem solved?

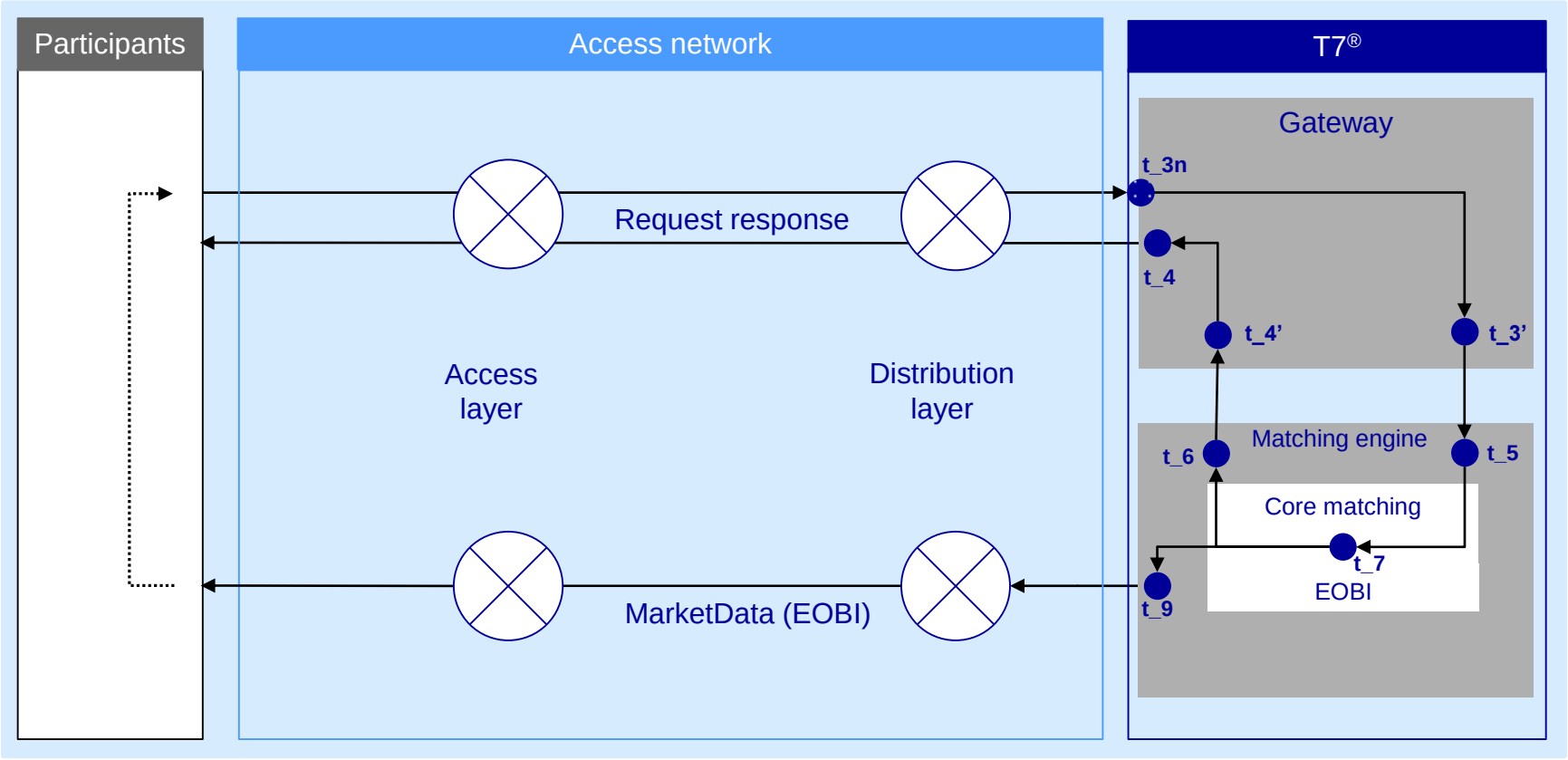
- Fair and equal access?
- State of the race to zero?
- Competition: does the winner take all?
- Can you even measure all this?

11

Network dynamics

Network dynamics

Simplified topology

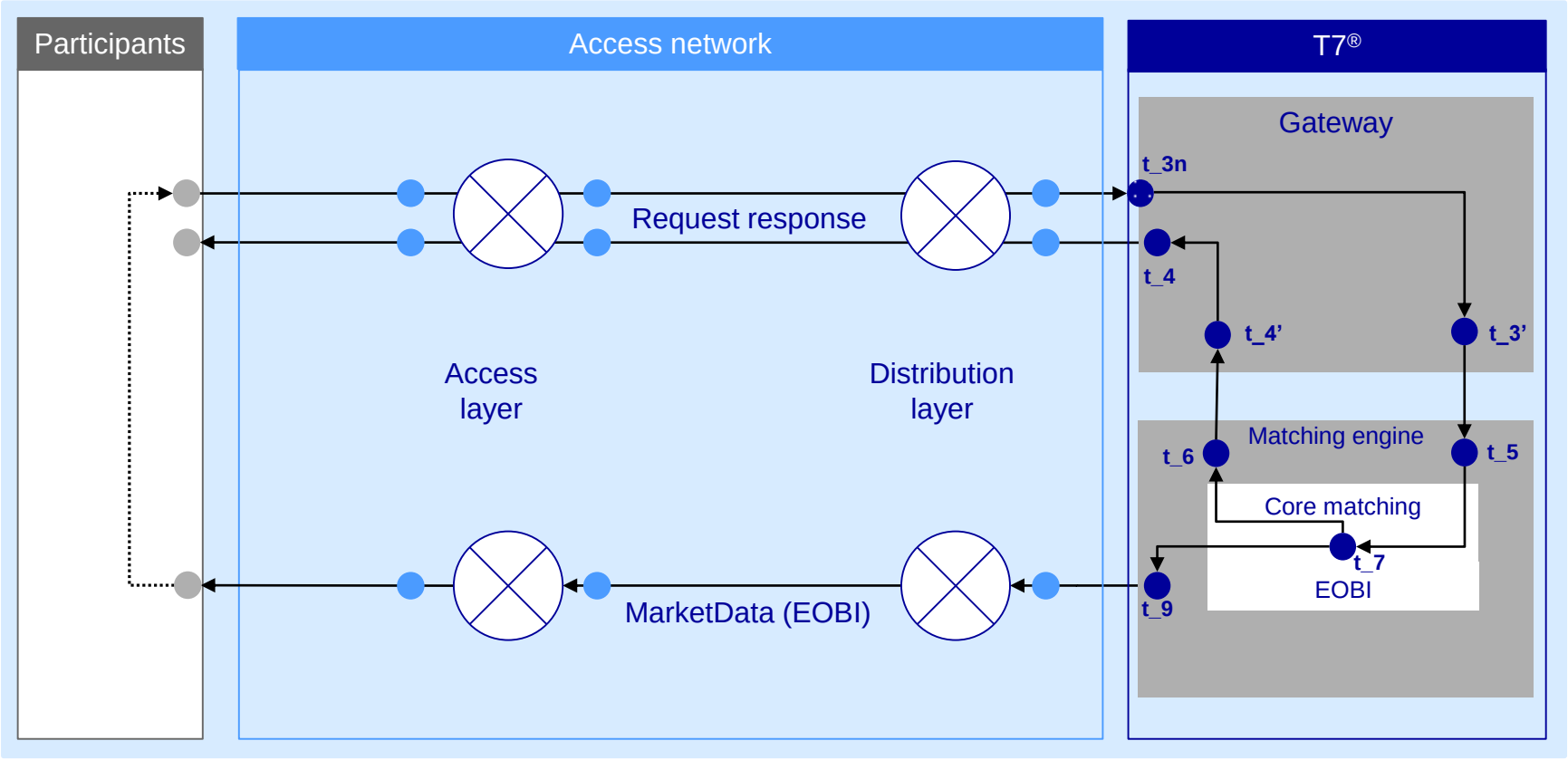


● Timestamps provided in T7 API (in real time) in dark blue (t_3n: taken by network card, other: application level)

⊗ Cisco 3548X switches operating in cut-through mode.

Network dynamics

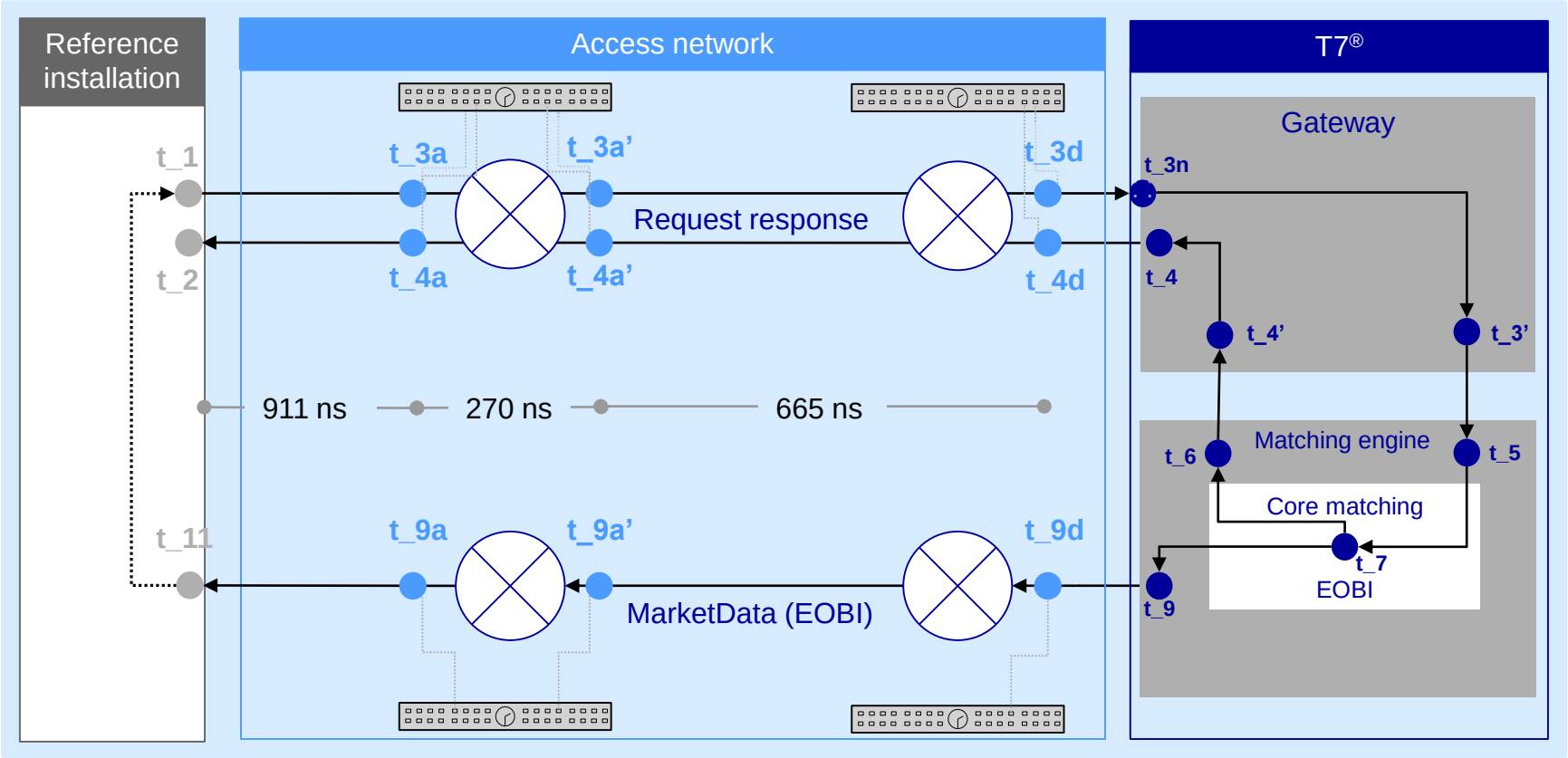
Taps



- Timestamps provided in T7 API (in real time) in dark blue (t_3n: taken by network card, other: application level)
- Network taps shown in light blue
- Timestamps possibly taken by participants shown in grey

Network dynamics

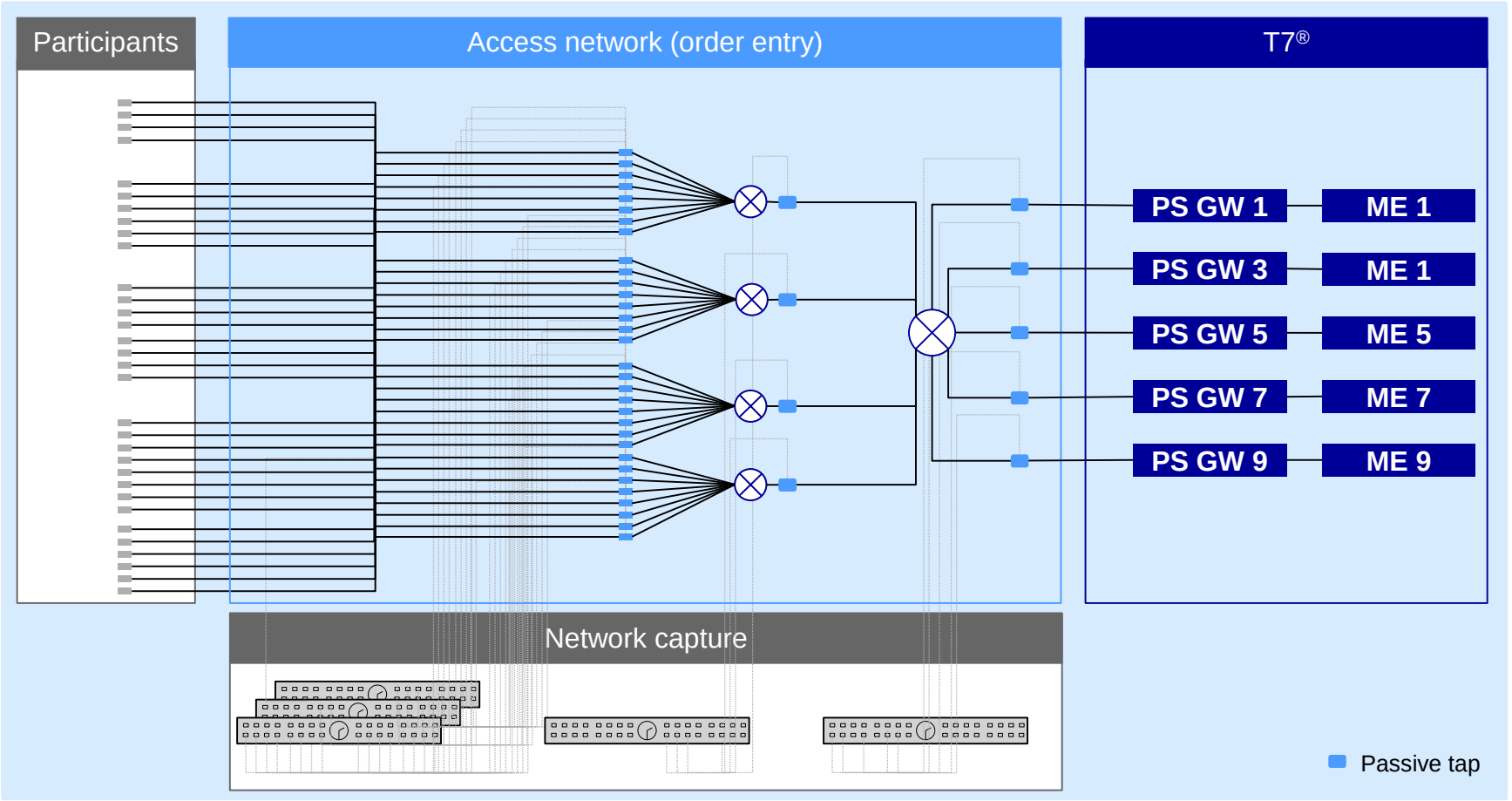
Timestamps



- Timestamps provided in T7 API (in real time) in dark blue (t_3n: taken by network card, other: application level)
- Network timestamps taken using taps and timestamping switches (Metamako)
- Timestamps possibly taken by participants shown in grey

Network dynamics

Order entry



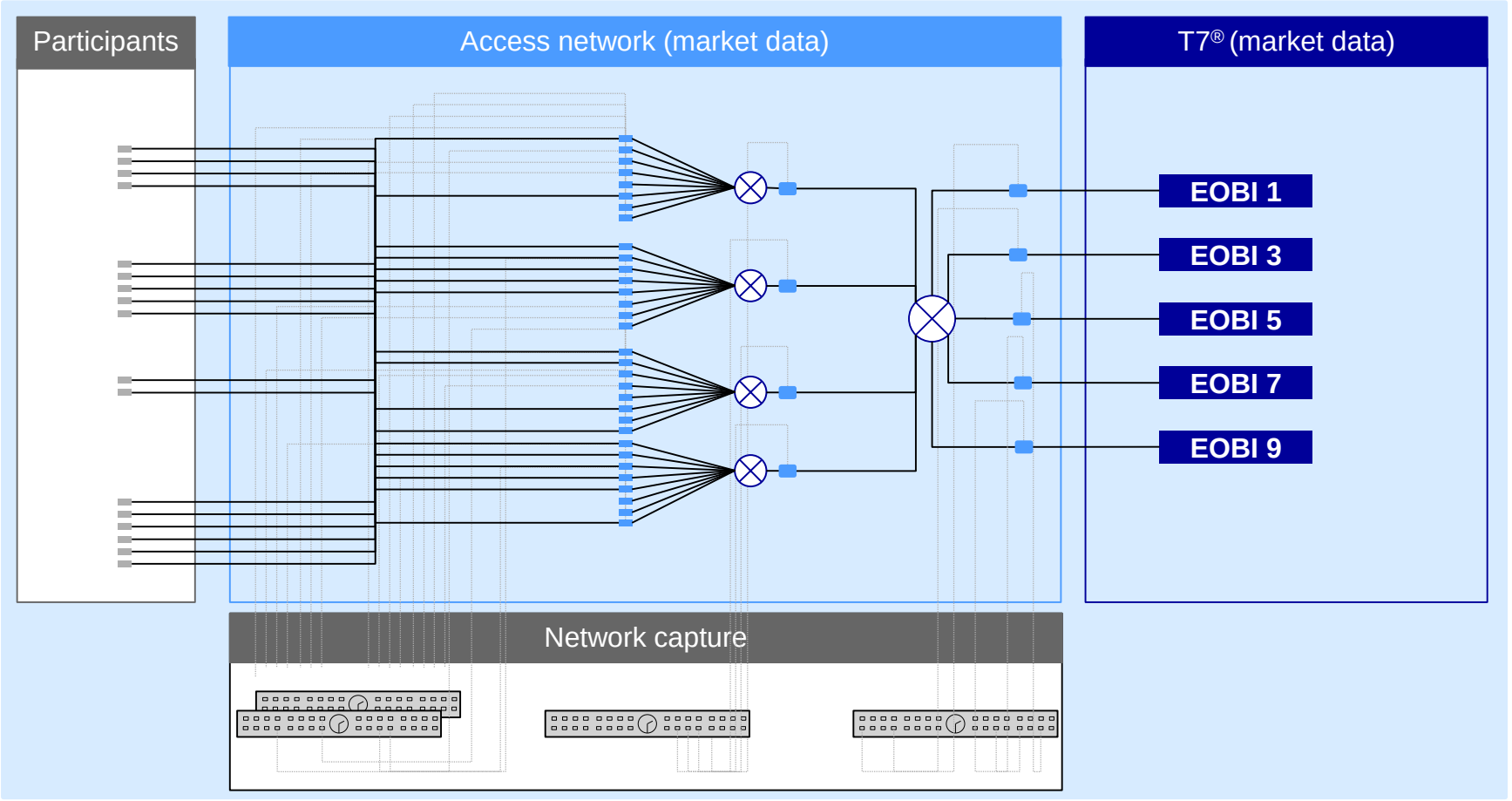
> 260 order entry lines captured (> 500 capture ports)

Identical set-up regardless of participant room location and assigned access switch

Only one side of one Market (Eurex) is shown for simplicity, a fifth switch will be added on 1 October 2018.

Network dynamics

Market data



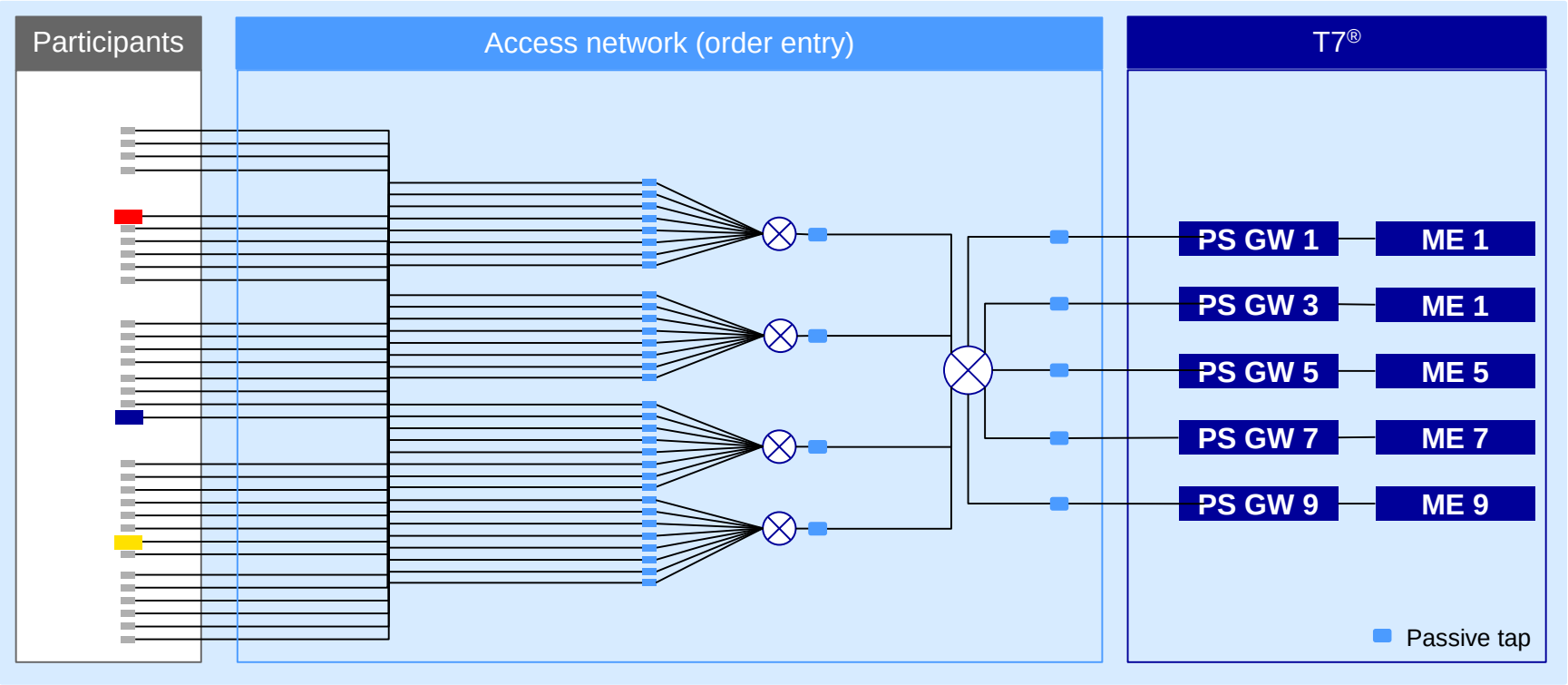
One market data access line per switch captured permanently – others configurable

Identical set-up regardless of participant room location and assigned access switch (differences < +/- 5 ns)

Only one side of one market (Eurex) is shown for simplicity.

Network dynamics

Order entry

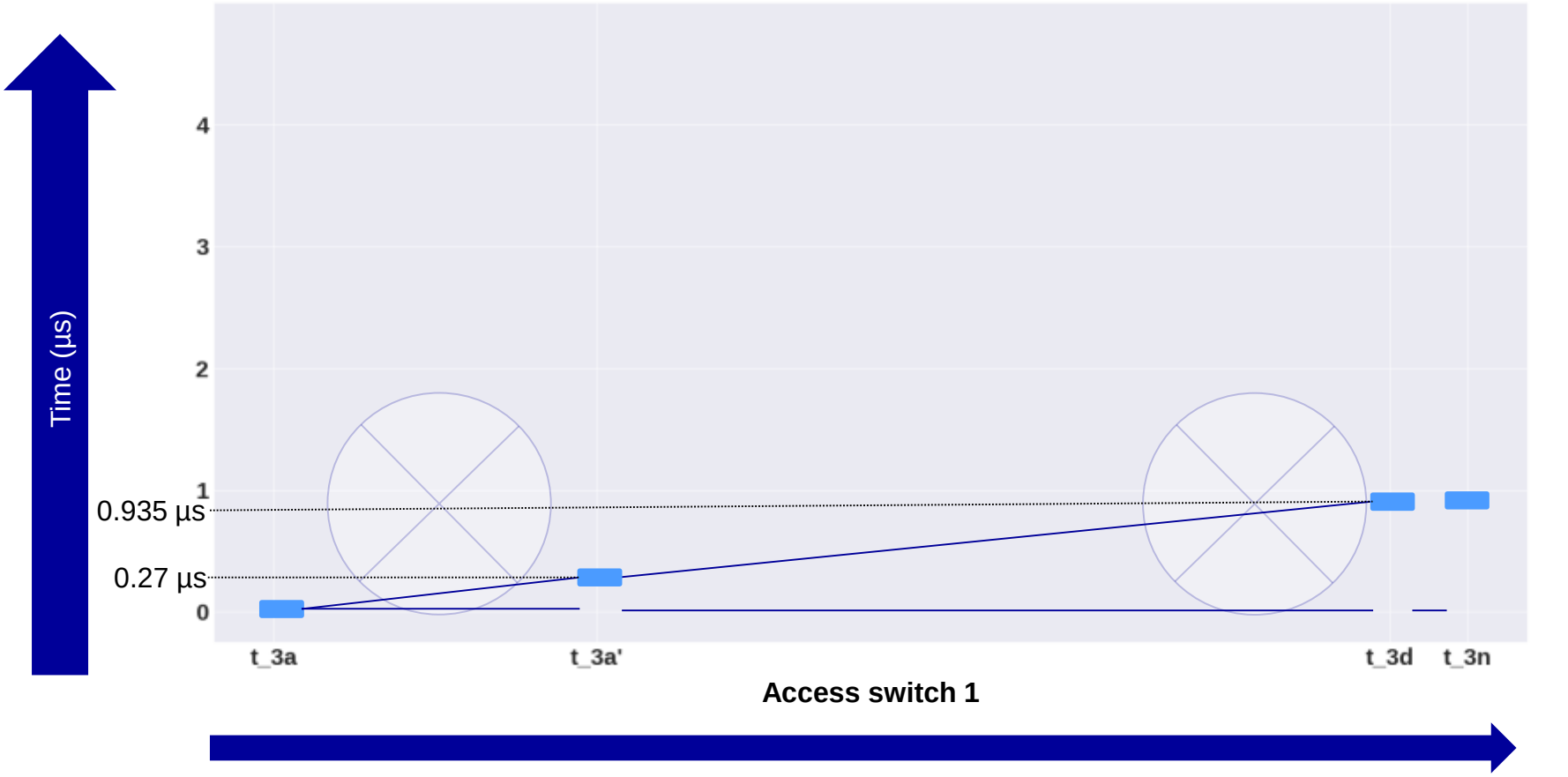


In highly competitive trading conditions network serialisation time leads to sizable delays in the network.

The gateway entry timestamp (t_3n) does include the queueing times and, therefore, cannot answer the question “How much too late was I?” in these scenarios.

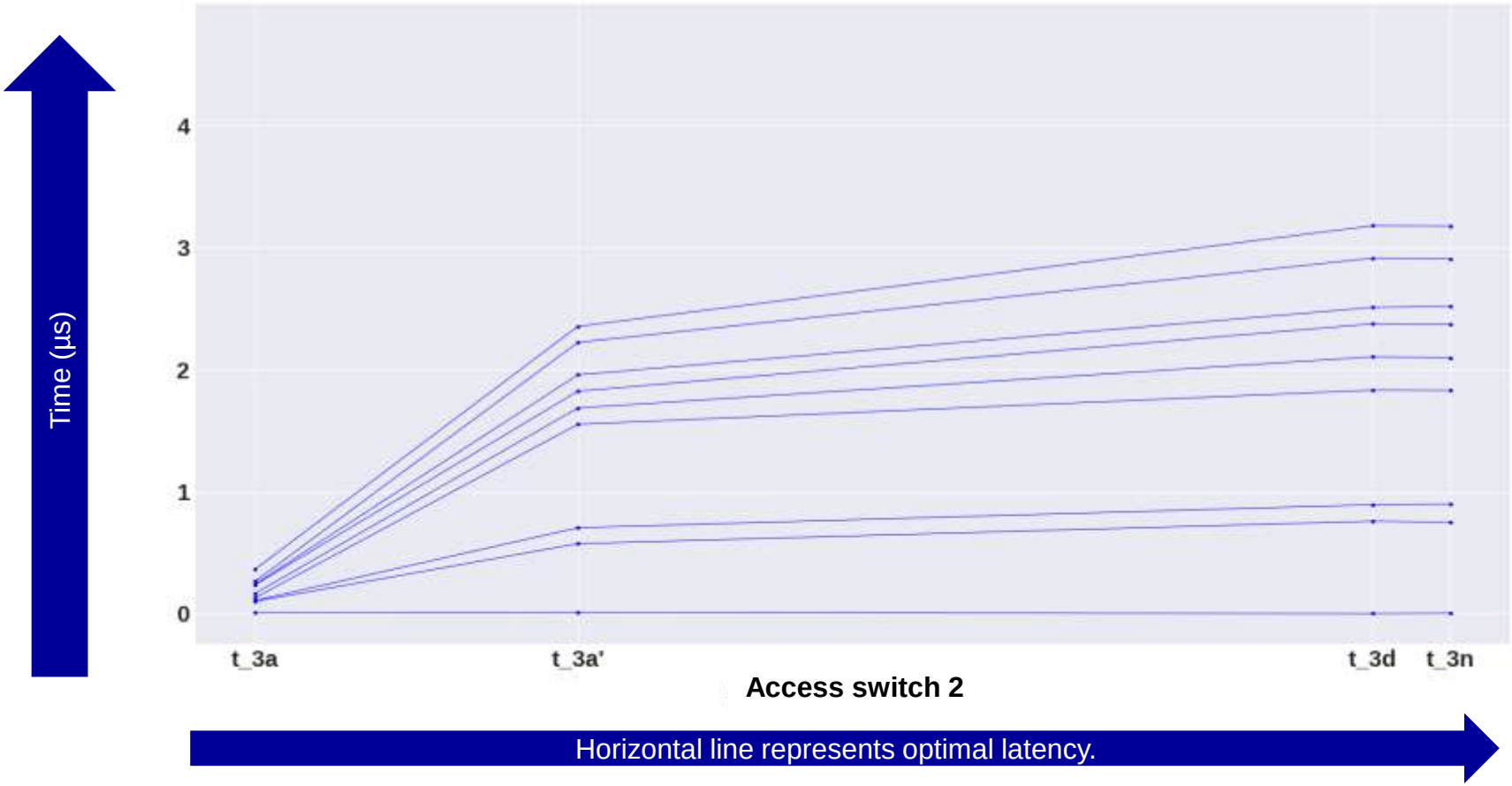
Network dynamics

Order entry – network timing diagram



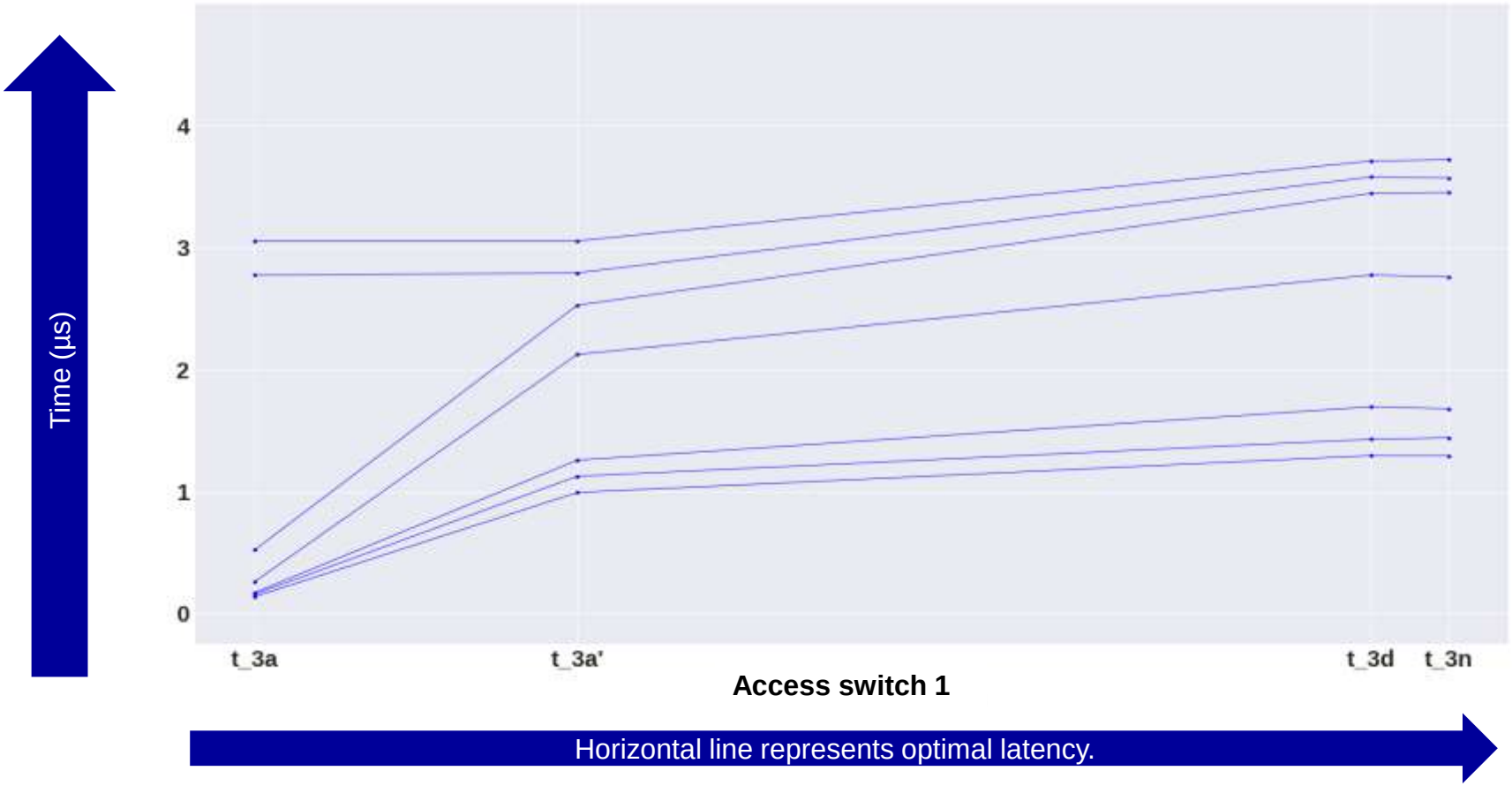
Network dynamics

Competitive burst (FESX shown)



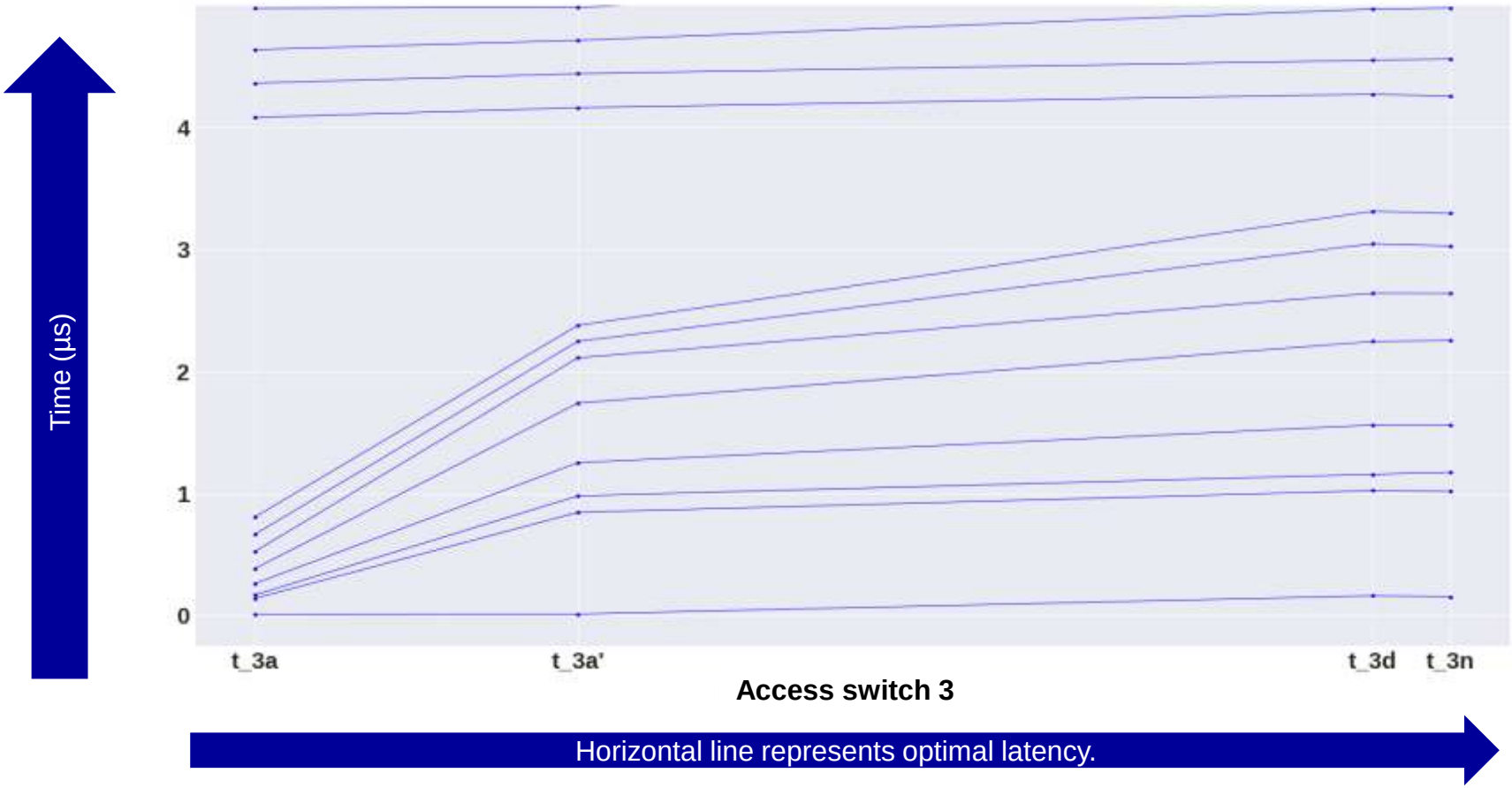
Network dynamics

Competitive burst (FESX shown)



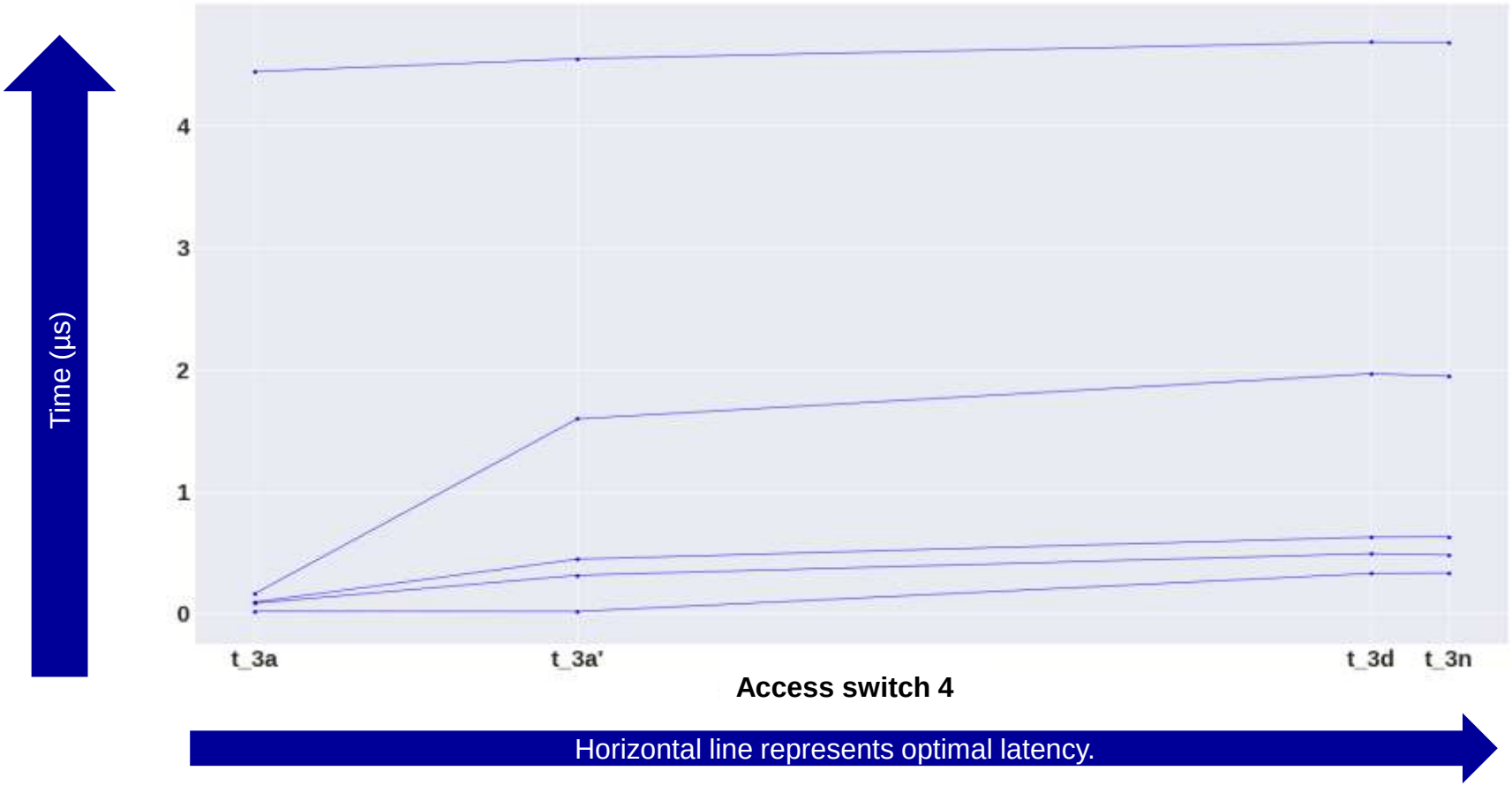
Network dynamics

Competitive burst (FESX shown)



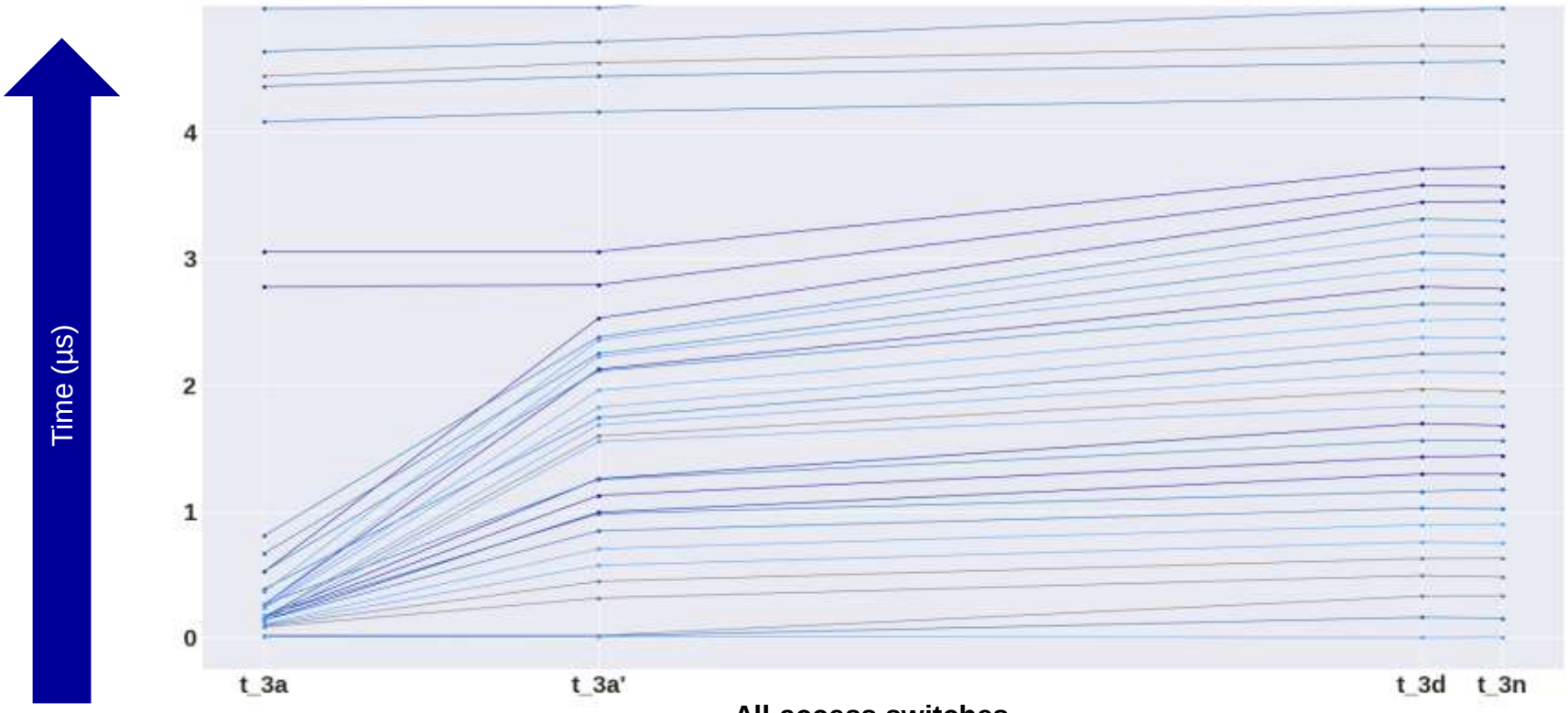
Network dynamics

Competitive burst (FESX shown)



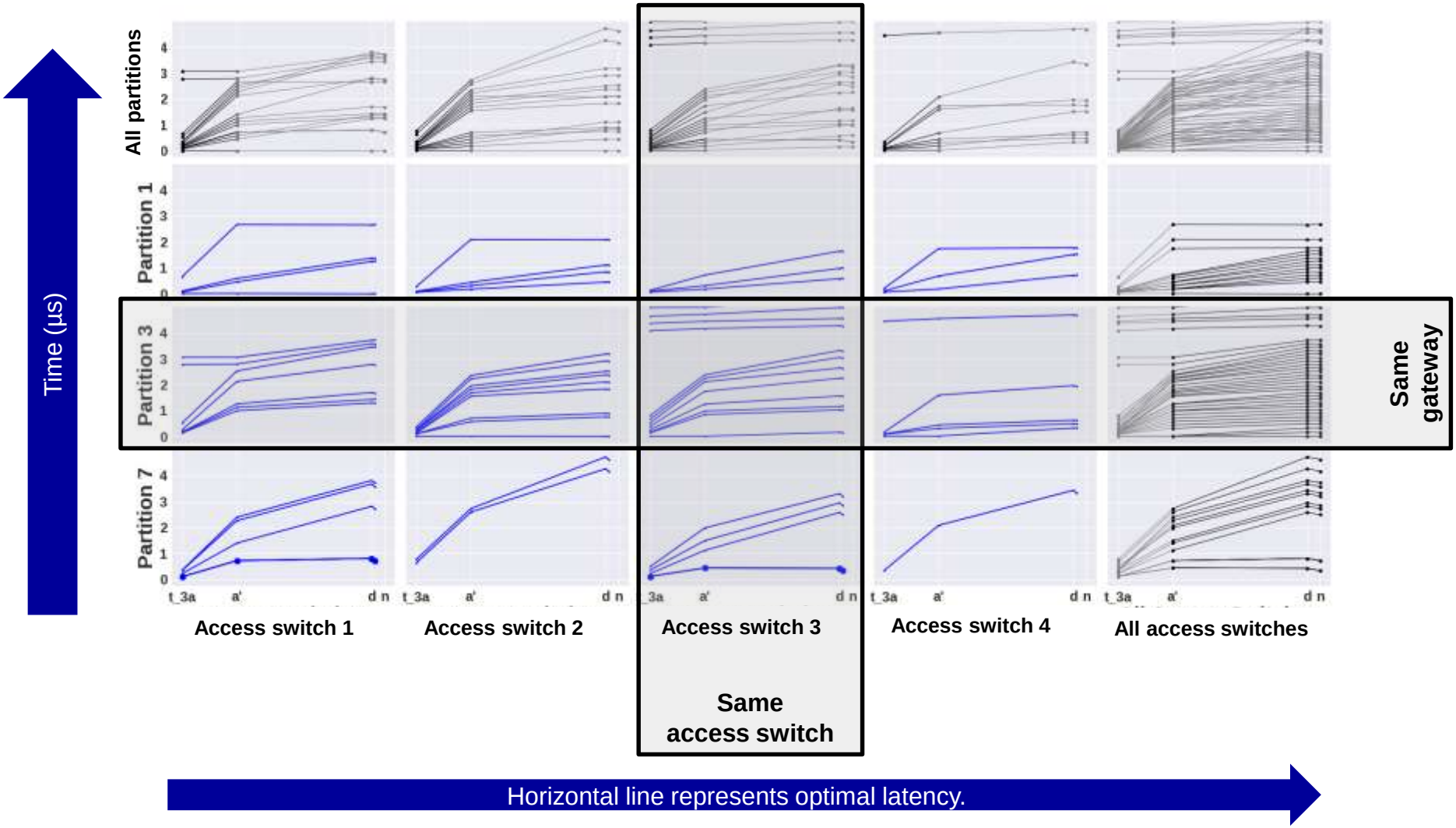
Network dynamics

Competitive burst (FESX shown)



Network dynamics

Eurex B side (FESX, FDAX and OESX)

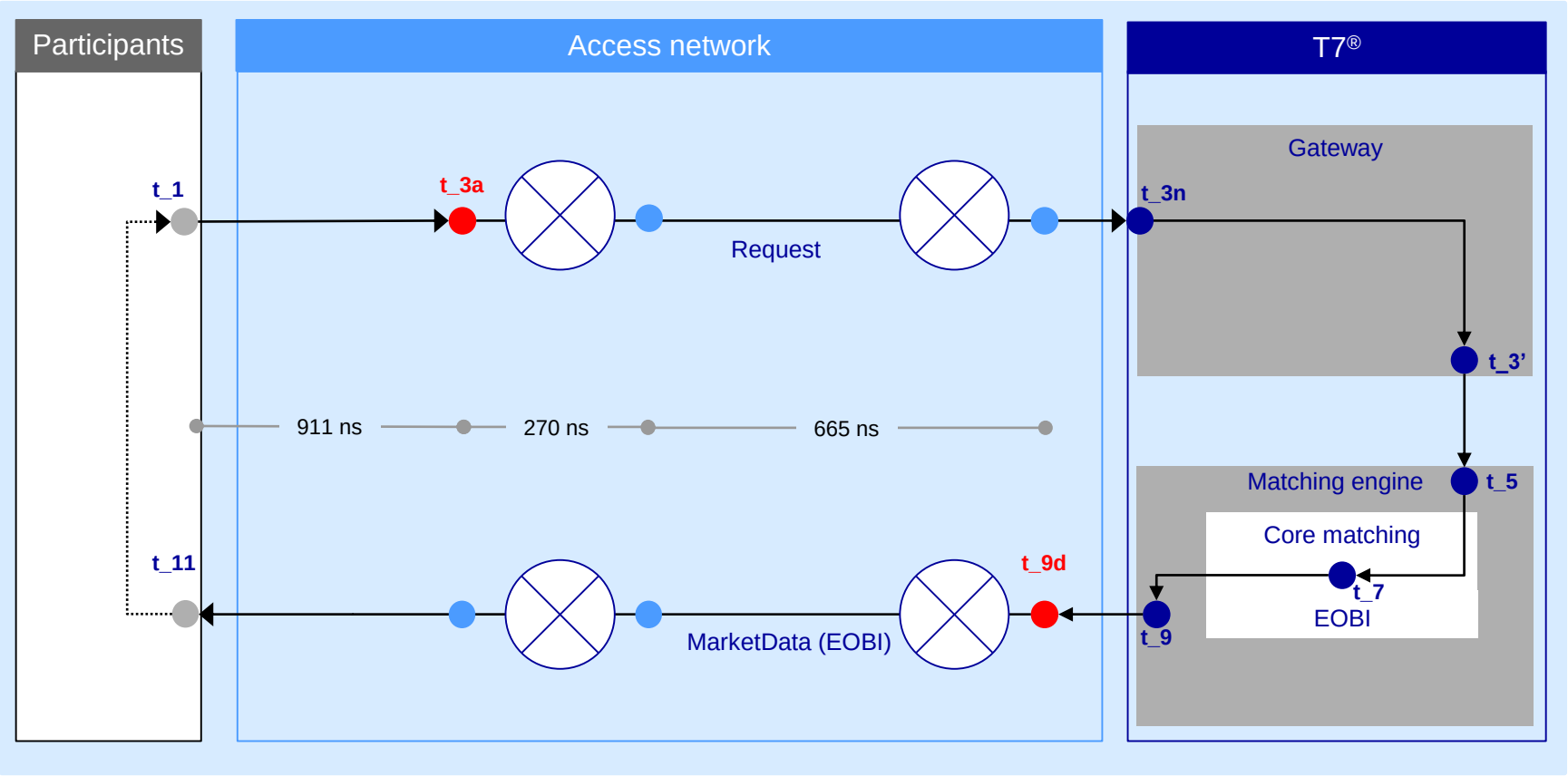


25

High Precision Timestamp File

High Precision Timestamp File

Contains network times t_9d and t_3a for all trades

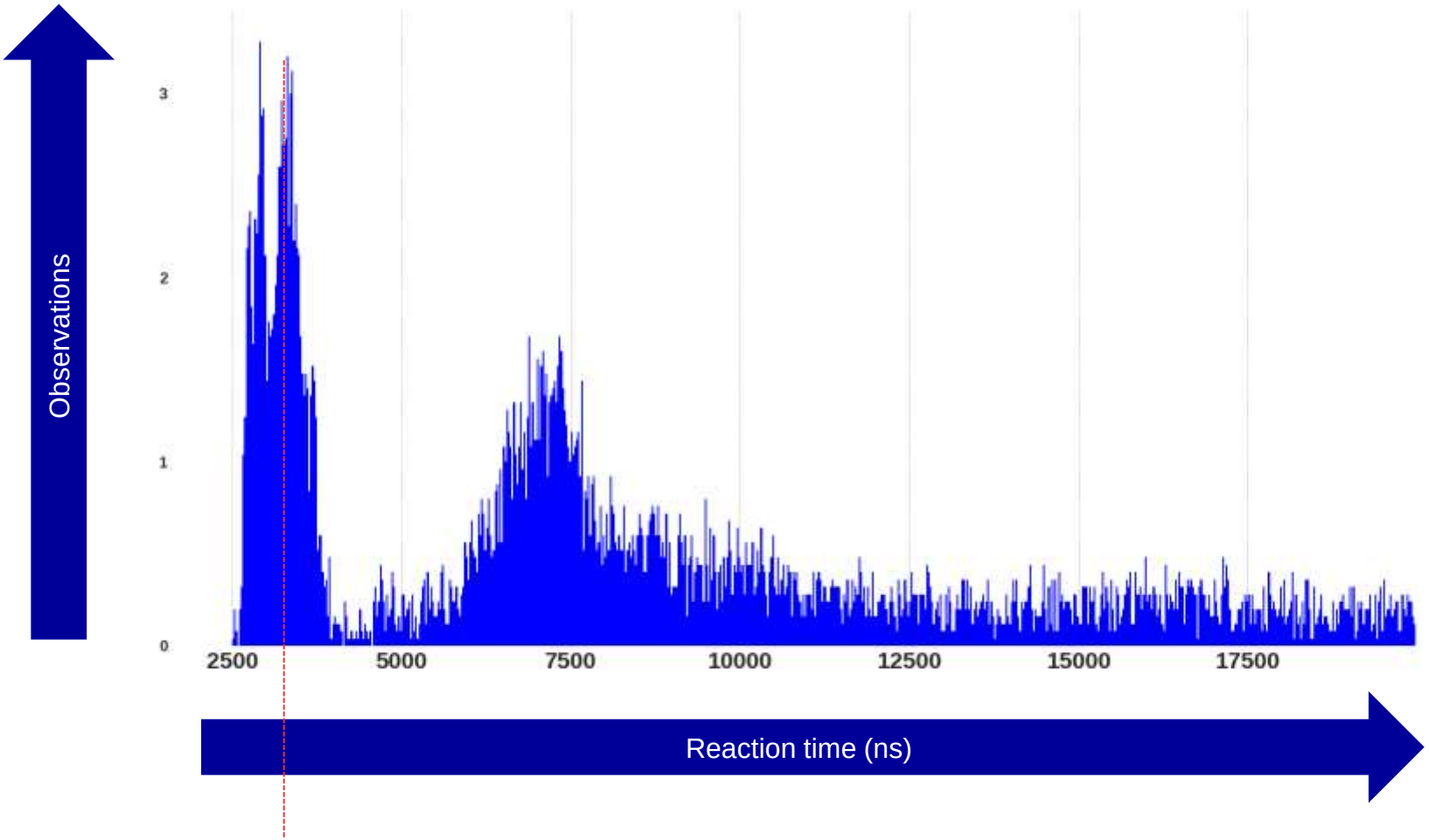


Use case: the signal generated by T7 leads to reactions by multiple trading participants. HPT allows to calculate reaction time differences with higher precision.

http://datashop.deutsche-boerse.com/High-precision-timestamps

High Precision Timestamp File

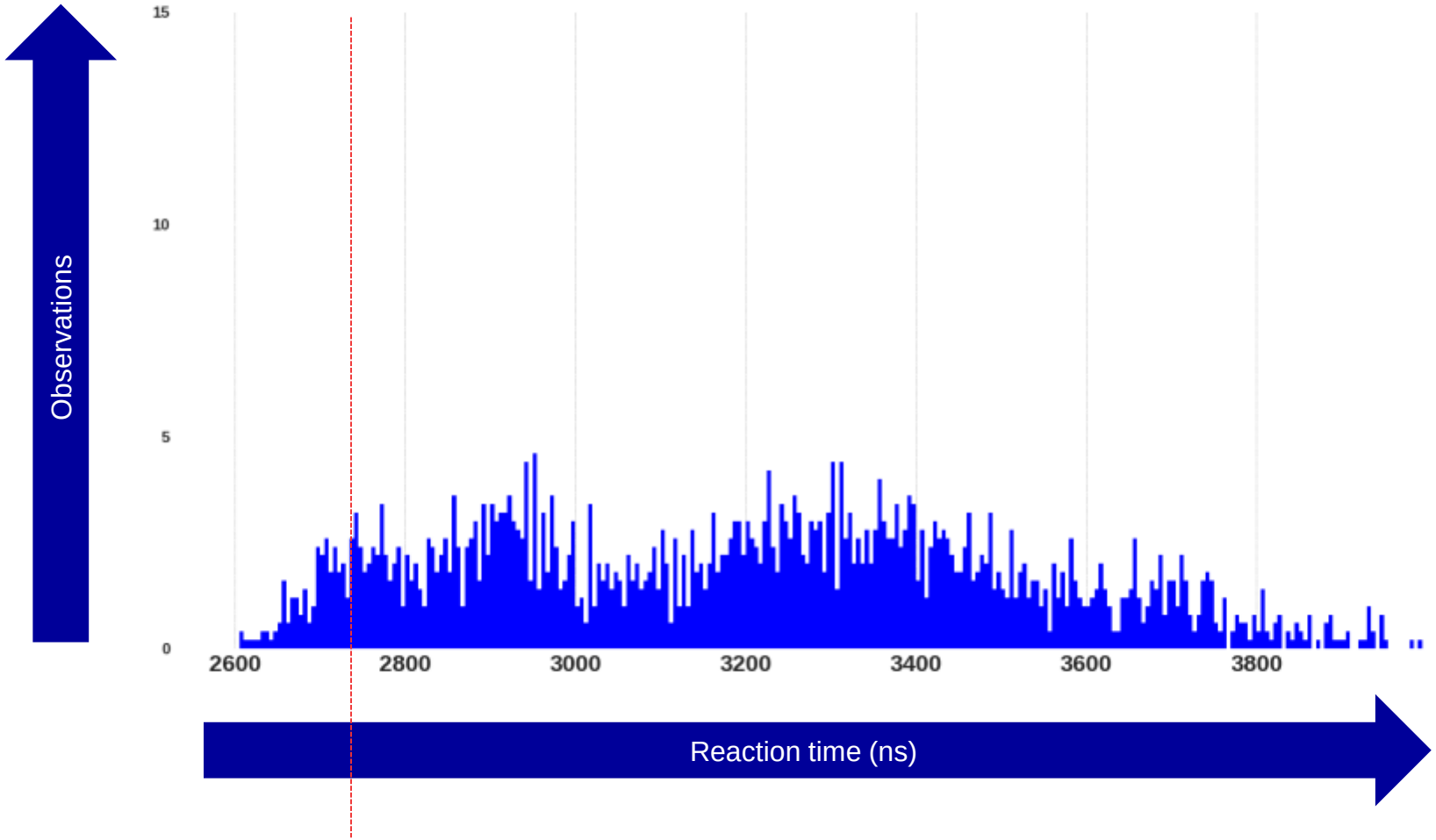
Reaction time based on T7® times (t_9 to t_3n)*



*Distribution of t_3n – t_9 – median (t_9d – t_9) – median (t_3n – t_3a) shown

High Precision Timestamp File

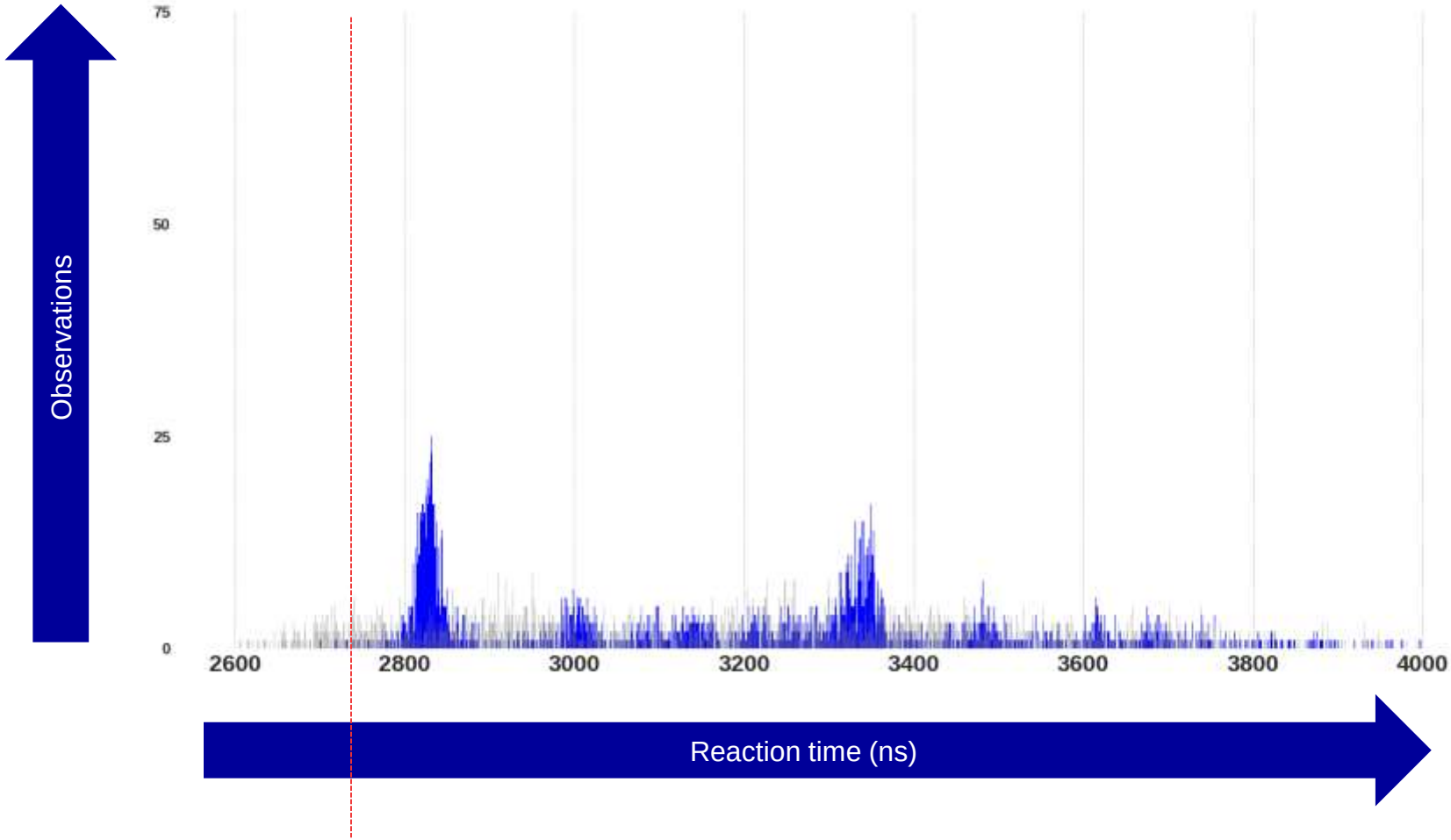
Reaction time based on T7® times (t_9 to t_3n)* (close up)



*Distribution of t_3n – t_9 – median (t_9d – t_9) – median (t_3n – t_3a) shown

High Precision Timestamp File

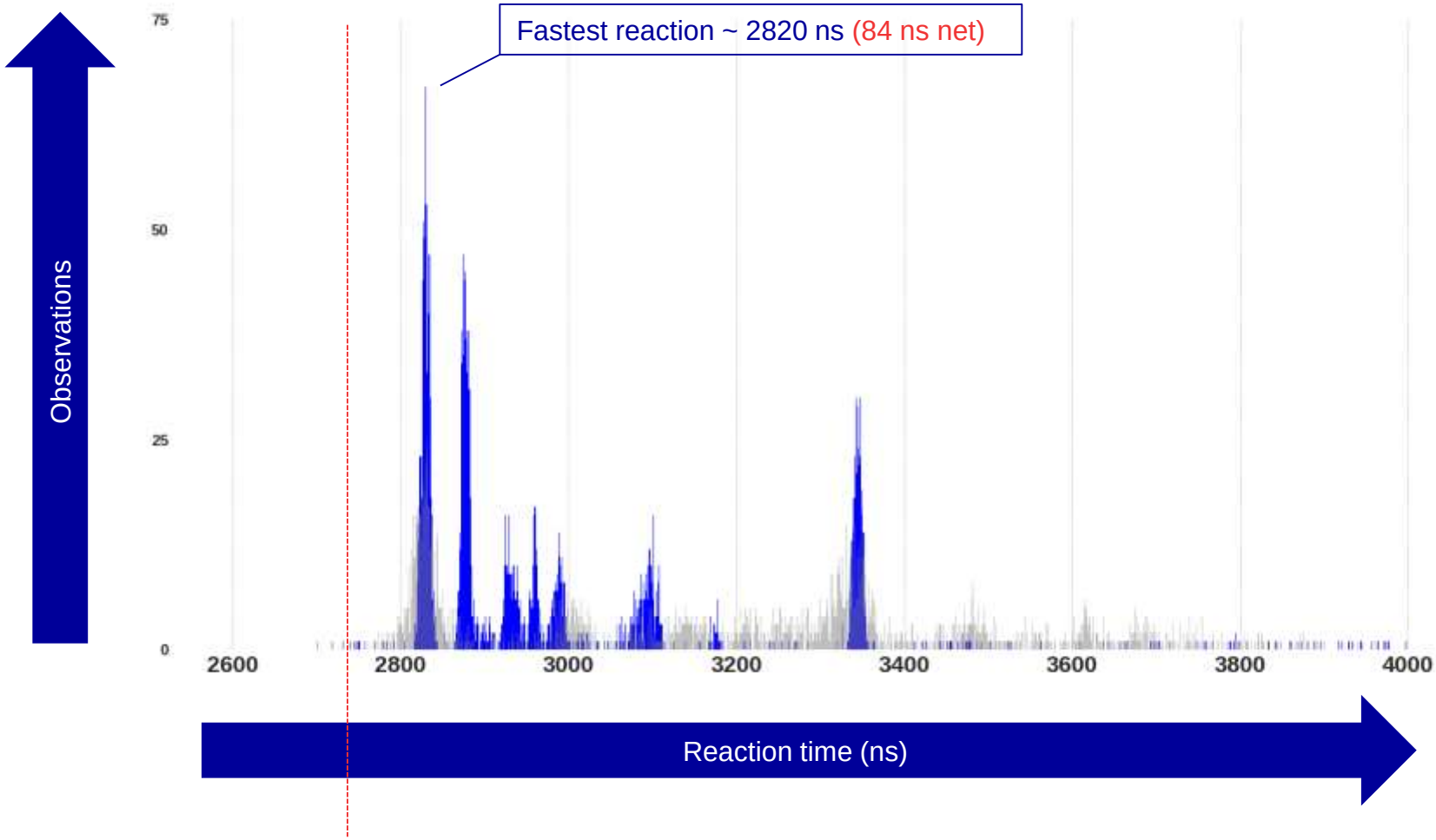
Reaction time based on T7® times (time synch corrected)*



*Distribution of t_3n – t_9 – rolling median [(t_9d – t_9) – (t_3n – t_3a)] shown

High Precision Timestamp File

Reaction time based on t_9d and t_3a (HPT)



*Distribution of t_3a – t_9d

31

T7® time synchronisation

T7® time synchronisation

Our PTP network can synchronise clocks down to +/- 50 ns in the best case.

Serialisation time of order entry message in 10 Gbit network ~120 ns

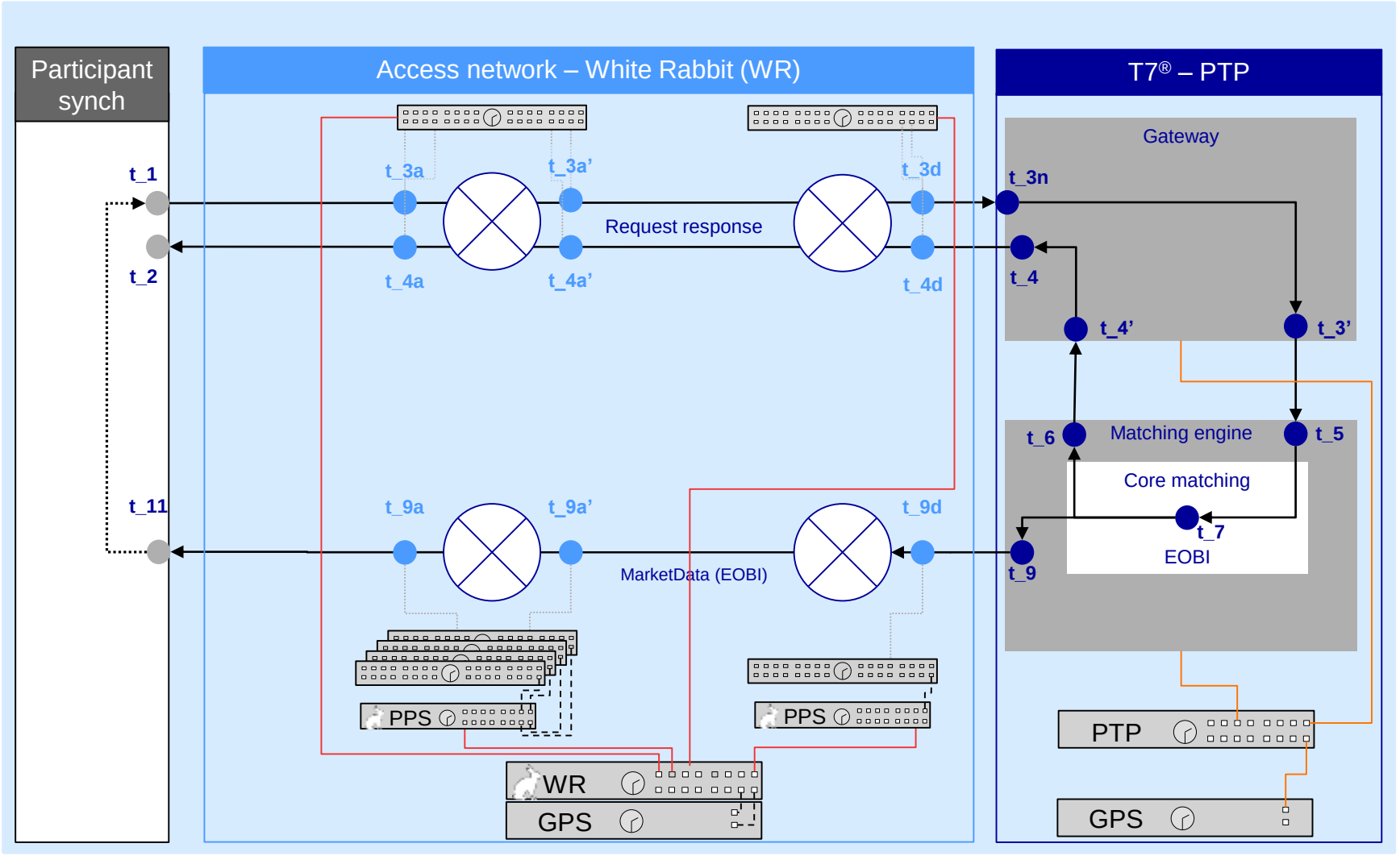
Time delta between two messages in the network often < 10 ns

➔ PTP not good enough to answer questions like:

- “Who came first?”
- “Is the network as deterministic as you claim?”

White Rabbit (WR) to the rescue

T7® time synchronisation

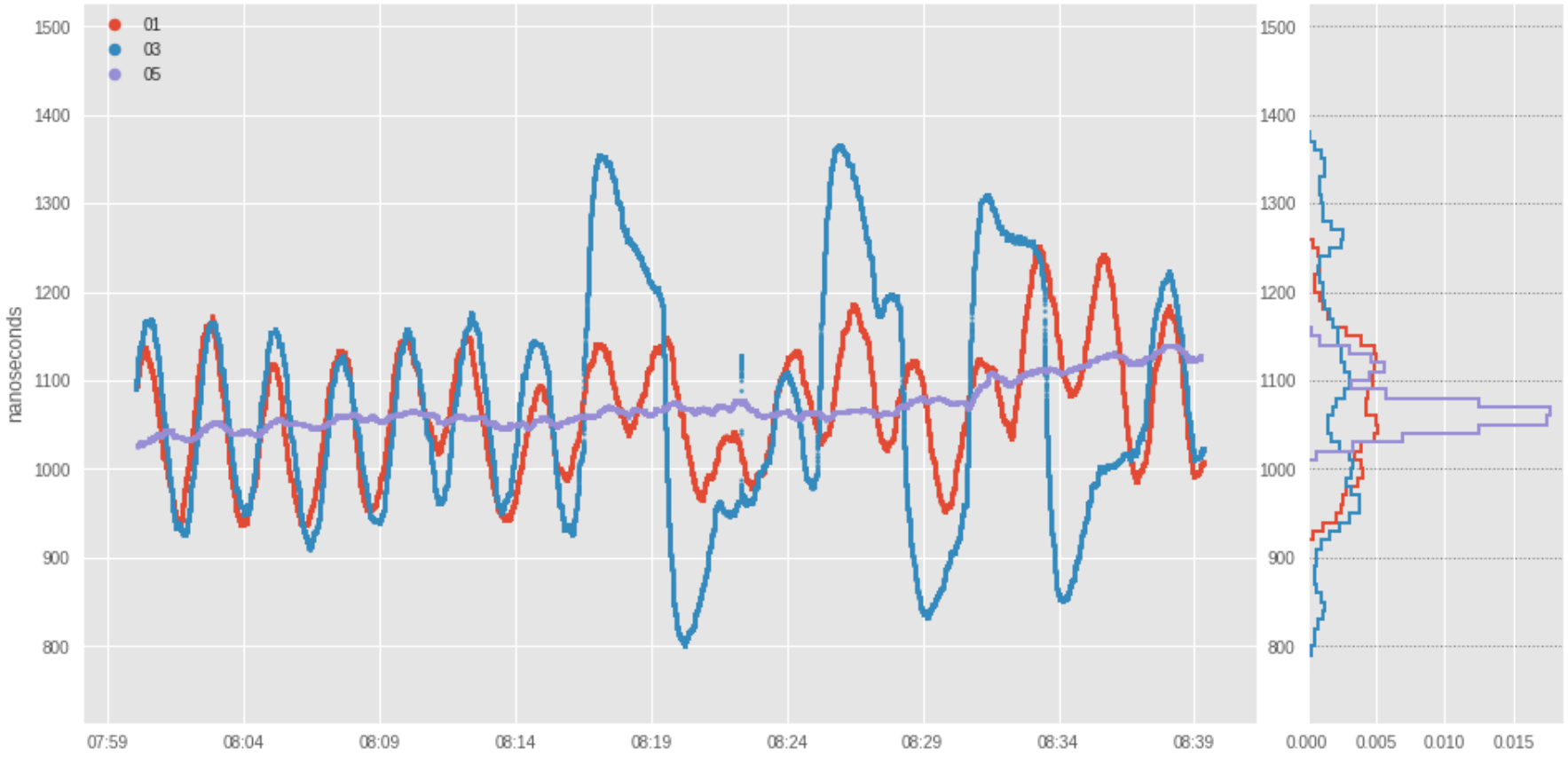


PTP shown in dark blue, WR in light blue for monitoring and HPT file service

T7® time synchronisation

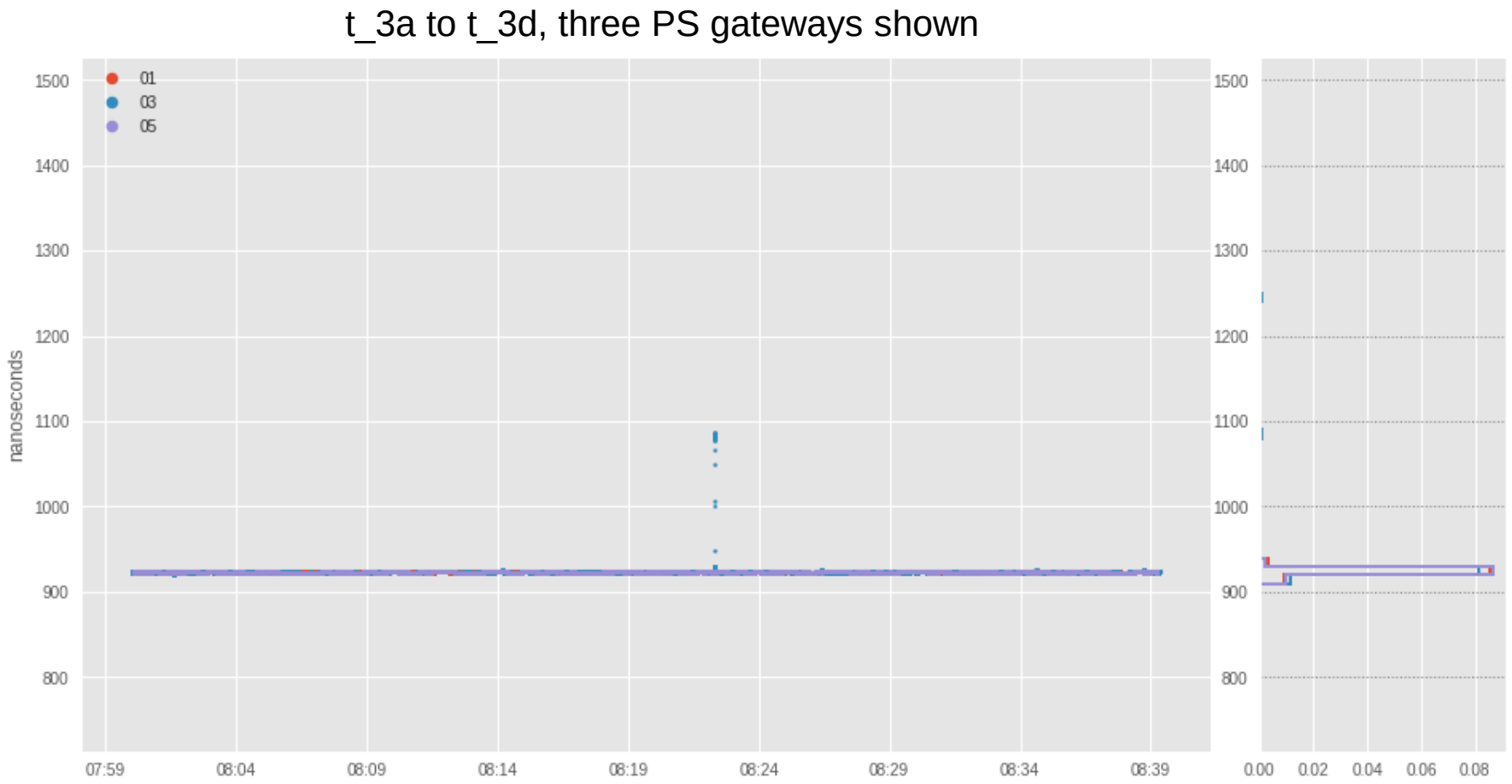
t_3a to t_3n (WR to PTP)

t_3a to t_3n, three PS gateways shown



T7® time synchronisation

t_3a to t_3d (WR to WR)



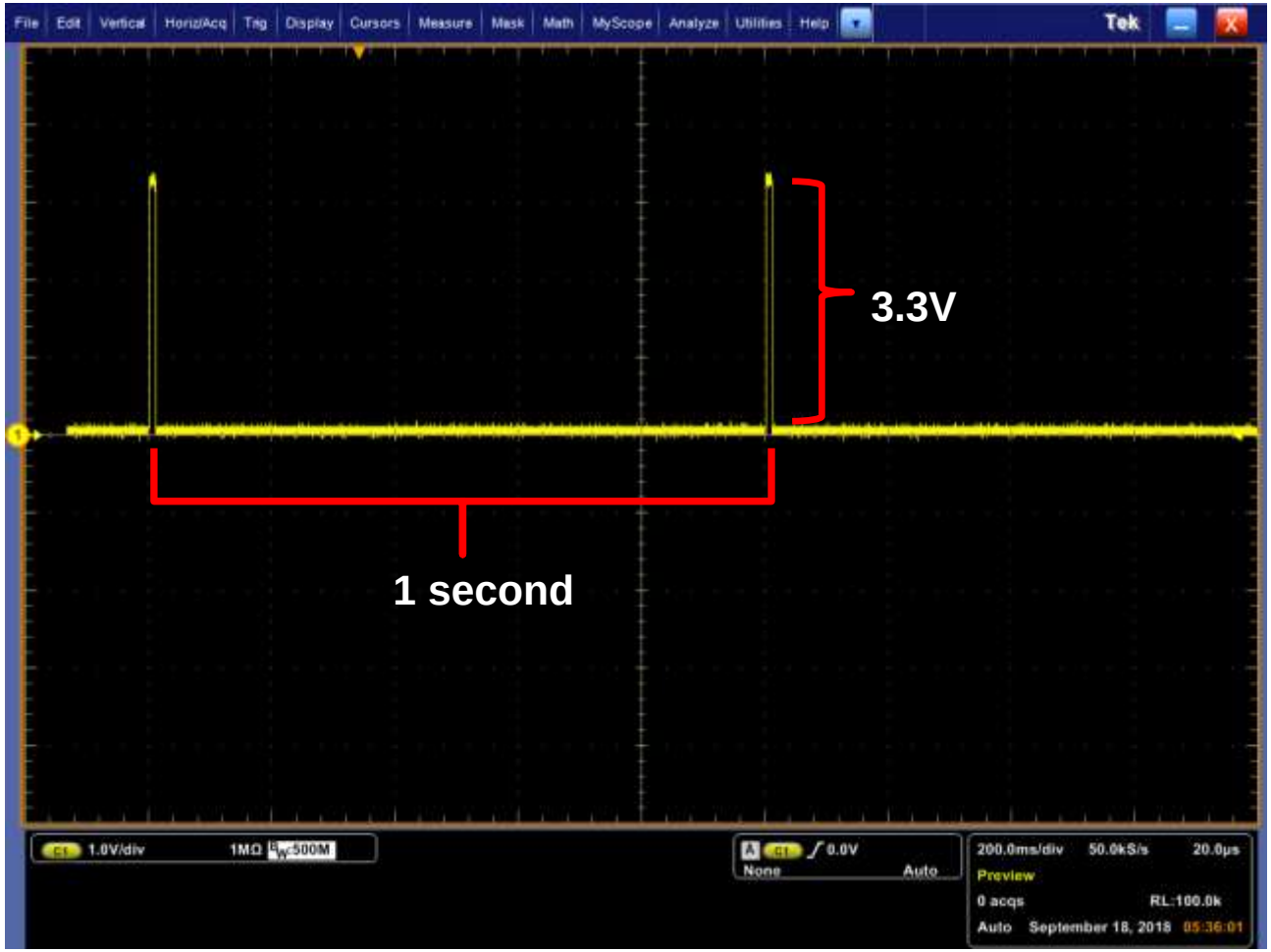
T7® time synchronisation

1PPS over White Rabbit

- White Rabbit was initially developed at CERN.
- White Rabbit is a fully deterministic Ethernet-based network for general-purpose **data transfer** and **time synchronisation**.
- It provides sub-nanosecond accuracy and picoseconds precision of synchronisation.
- There is no native White Rabbit support in standard NICs and switches.
- We use White Rabbit to distribute **1PPS.**

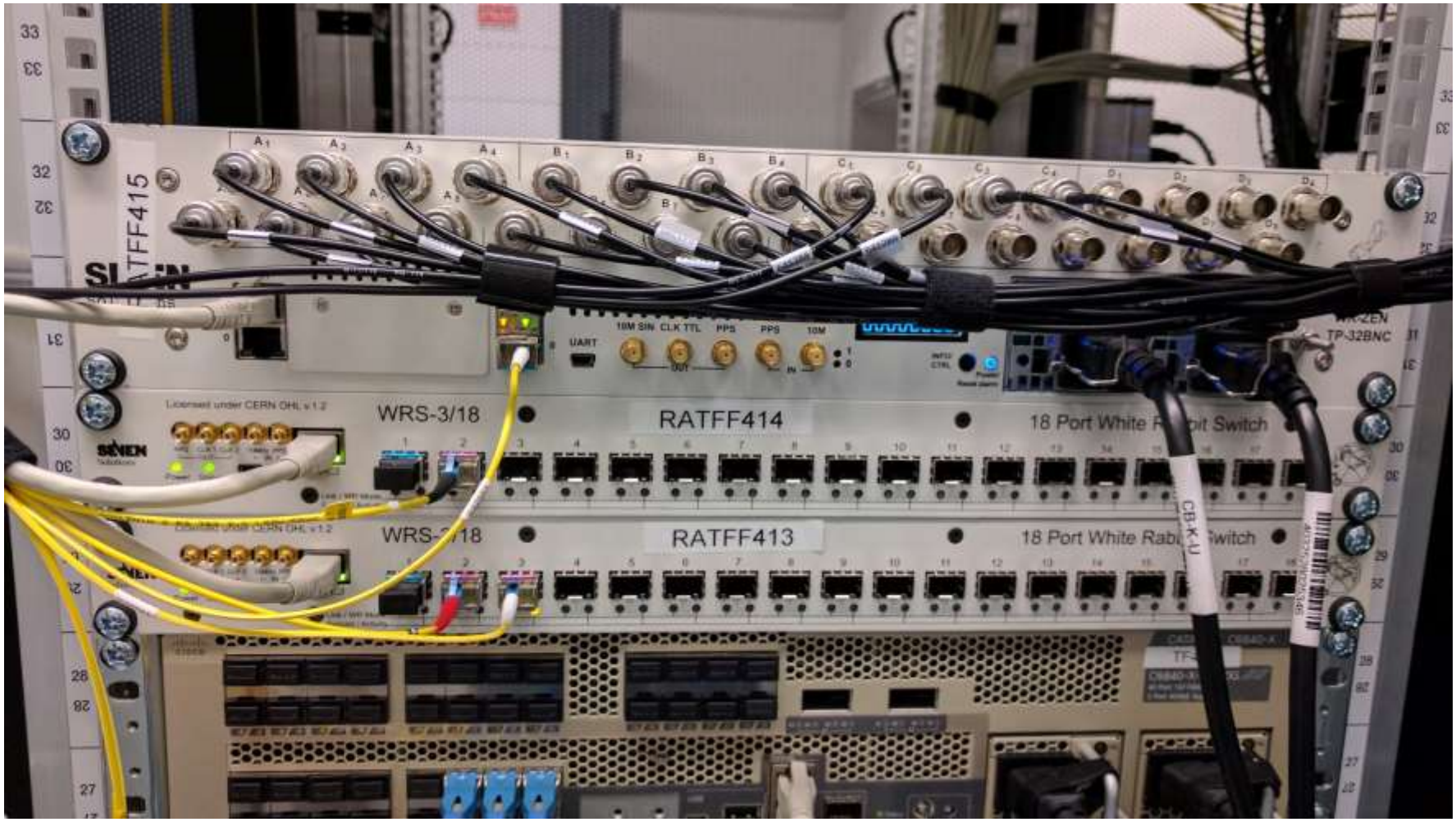
T7® time synchronisation

1PPS over White Rabbit



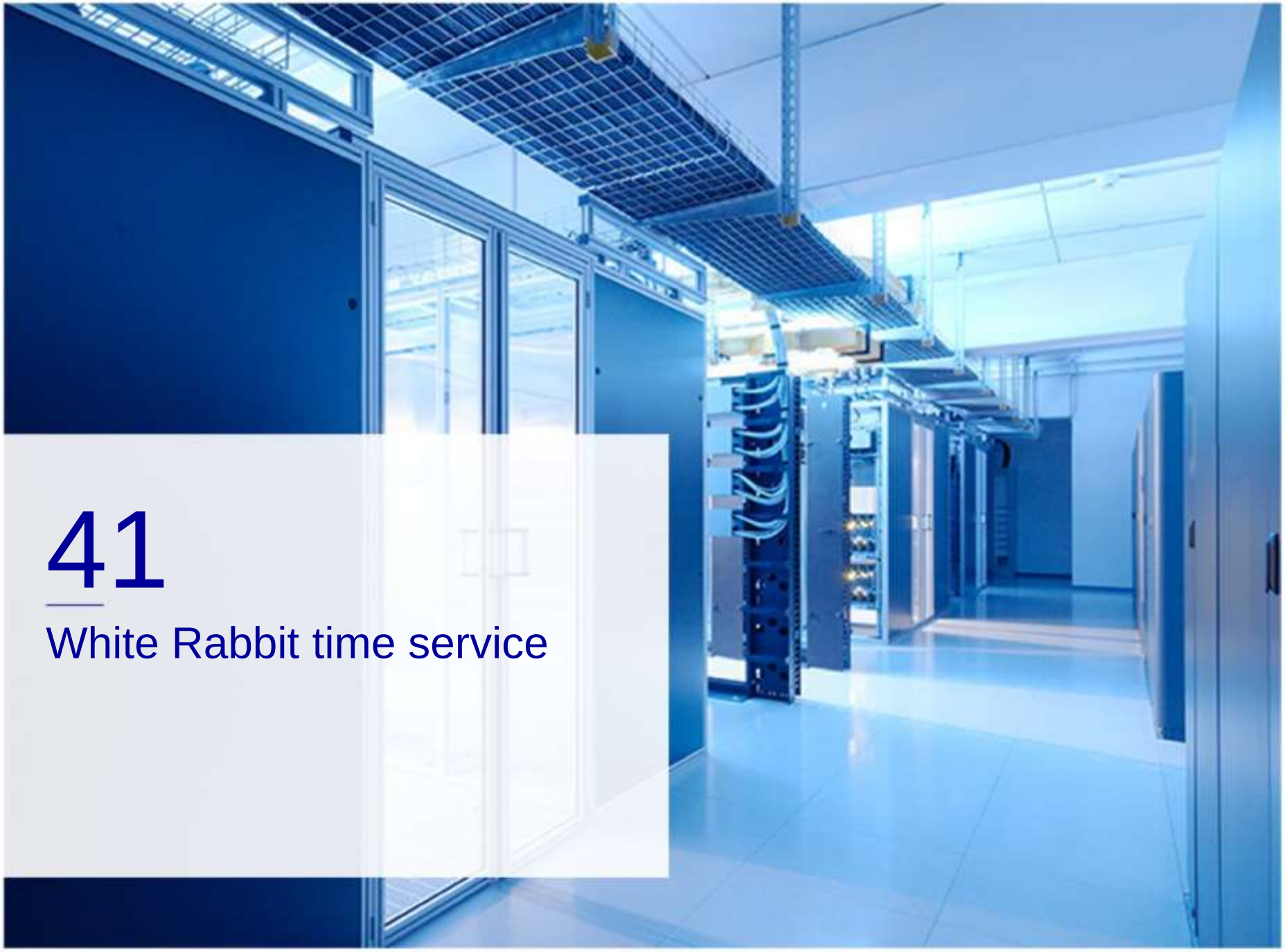
T7® time synchronisation

Timestamping devices synchronised by WR and PPS



T7® time synchronisation

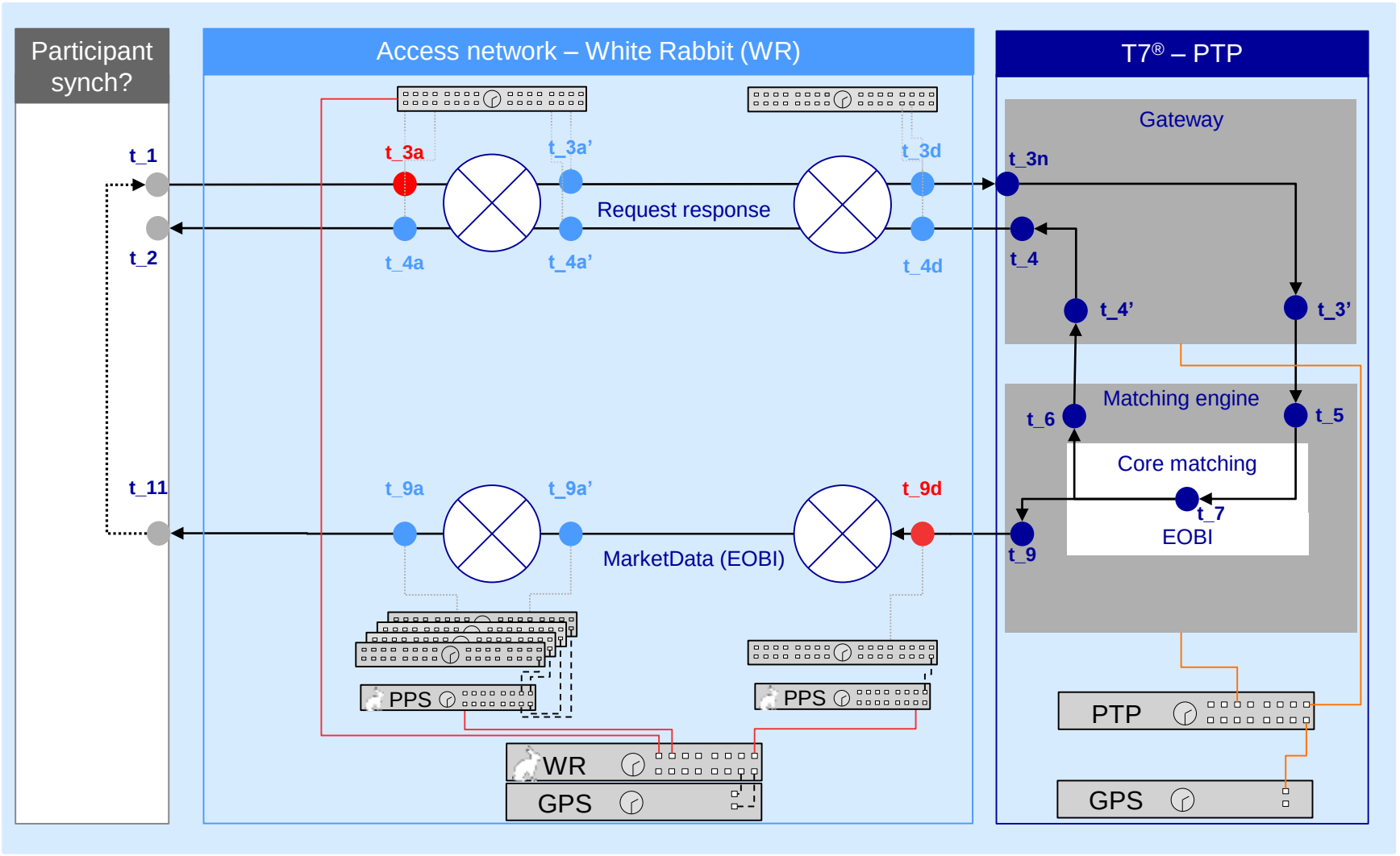
Timestamping devices synchronised by PPS



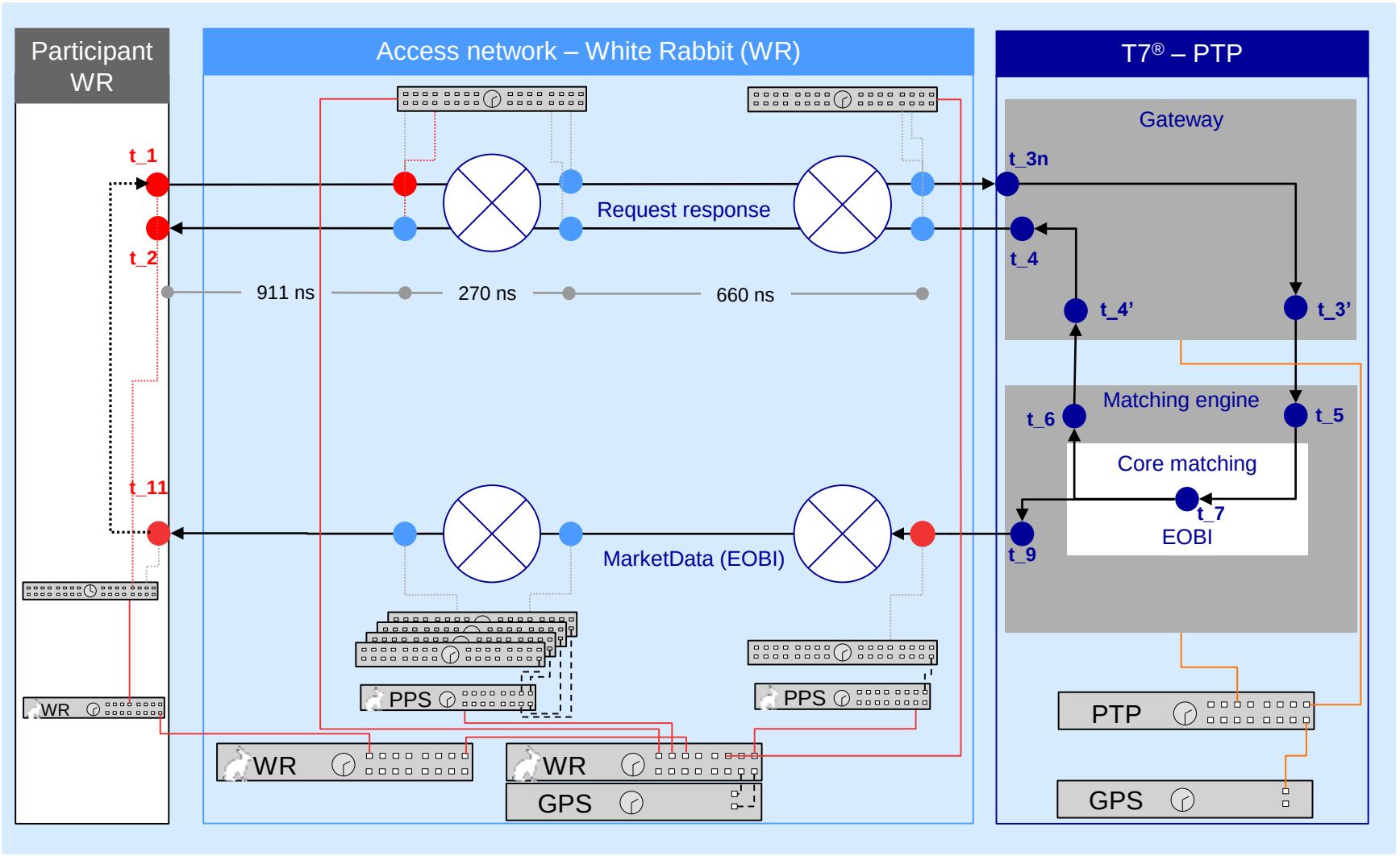
41

White Rabbit time service

White Rabbit time service

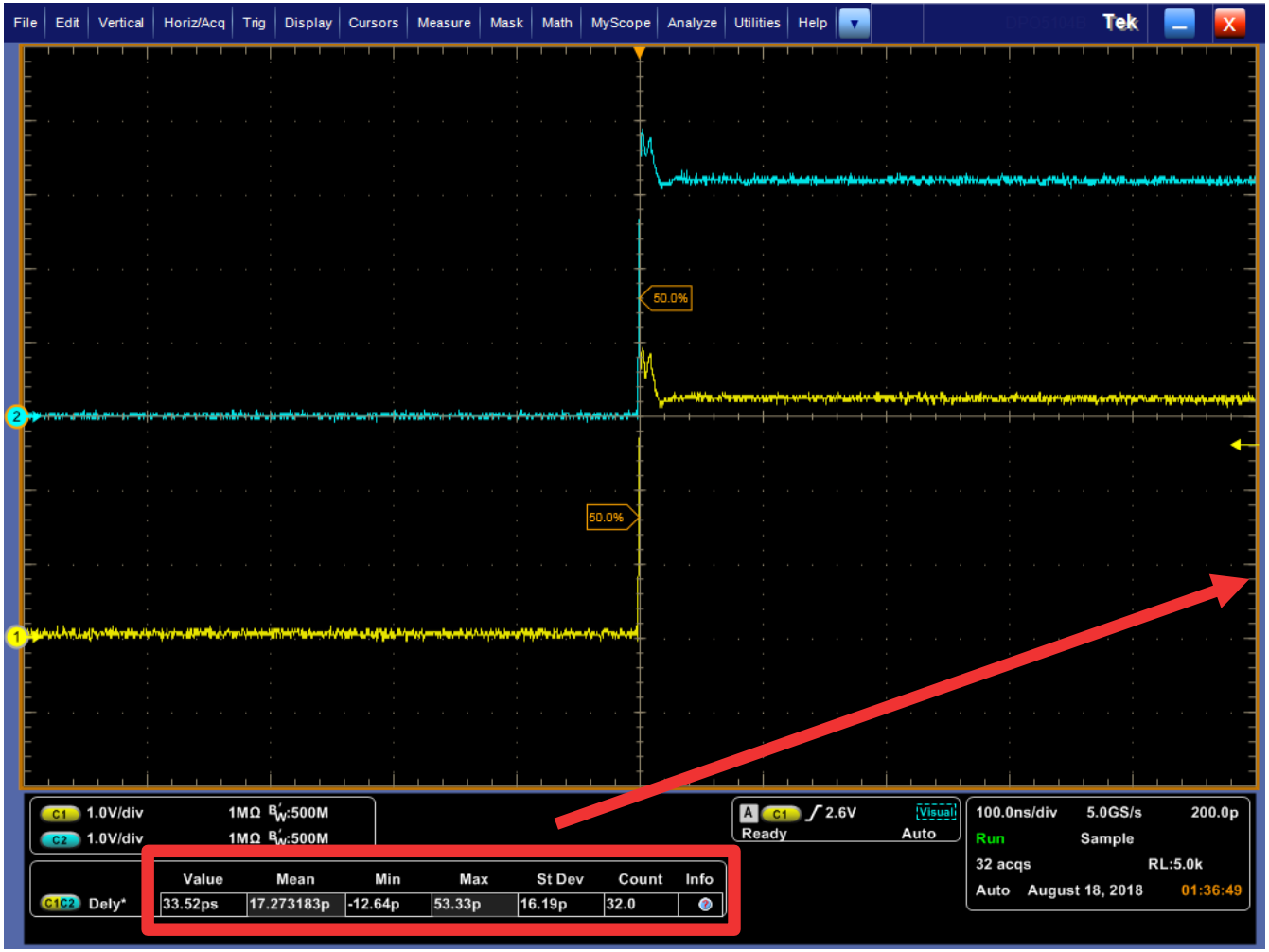


White Rabbit time service



White Rabbit time service

Two client switches connected to WR service



Value	33.52ps
Mean	17.27ps
Min	-12.64ps
Max	53.33ps
St Dev	16.19ps

White Rabbit time service

Pilot project

- Test the feasibility of this technology
- First phase running until end of 2018
- Free of charge during this time
- Available in co-location only
- Order in member section, requests and configuration

White Rabbit time service

What do you need?

White Rabbit (WR) is an open source and open hardware. It is commercially available via:
https://www.ohwr.org/projects/white-rabbit/wiki/wrcompanies

Most likely you want to transfer WR to 1PPS and connect to PPS input of your NIC or device. Vendors offer suitable products for this purpose.

Simplex single mode cable (LC connector for SFP, SC connector for patch panel)
BIDI SFPs
https://www.ohwr.org/projects/white-rabbit/wiki/SFP

White Rabbit time service

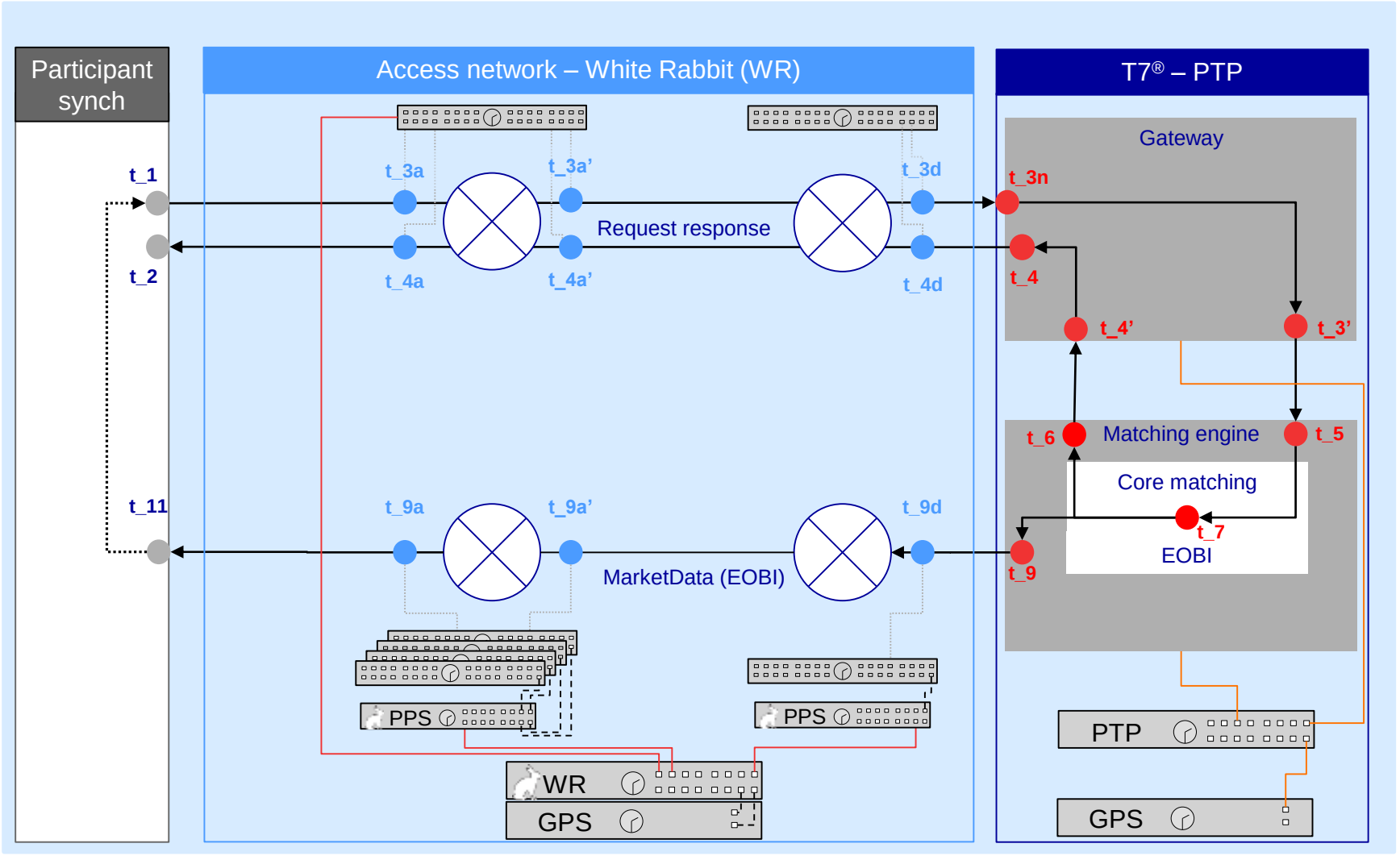
Goody bag

48

Outlook

Outlook

What about time synchronisation on servers?



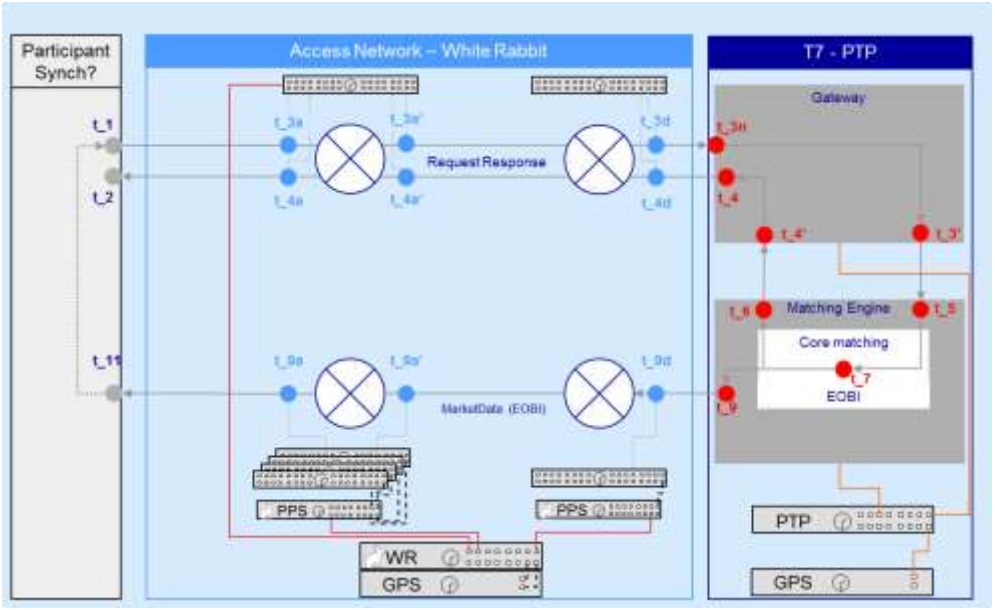
Outlook

What about time synchronisation on servers?

- We use White Rabbit (WR) to distribute 1PPS.
- There is no native WR support in servers or network devices we use.

Possible solutions:

- Network interface cards with higher time resolution
- Replace PTP distribution switches
- Hybrid WR/PTP
- Something completely different



Outlook

And now for something completely different

The New York Times

"Time split to the nanosecond is precisely what Wall Street wants."

- HUYGENS algorithm to synchronise nodes in a network
- Initially developed at Stanford University
- Now commercially developed by start-up Tick Tock Networks, Inc.

Very different from NTP or PTP:

- Exploits network effects, each server probes 10-20 others
- Coded probes
- Machine learning; support vector machines

https://www.nytimes.com/2018/06/29/technology/computer-networks-speed-nasdaq.html

Outlook

HUYGENS by Tick Tock Networks, Inc.

- We have conducted an initial test with three nodes.
- All components worked right out of the box.
- The figures look promising.

Exploiting a natural network effect for scalable, fine-grained clock synchronisation
https://www.usenix.org/conference/nsdi18/presentation/geng

Thank you for your attention.

Contact
Sebastian Neusüß
Andreas Lohr
E-mail monitoring@deutsche-boerse.com
Phone +49-(0) 69-2 11-1 86 86

27 September 2018

Disclaimer

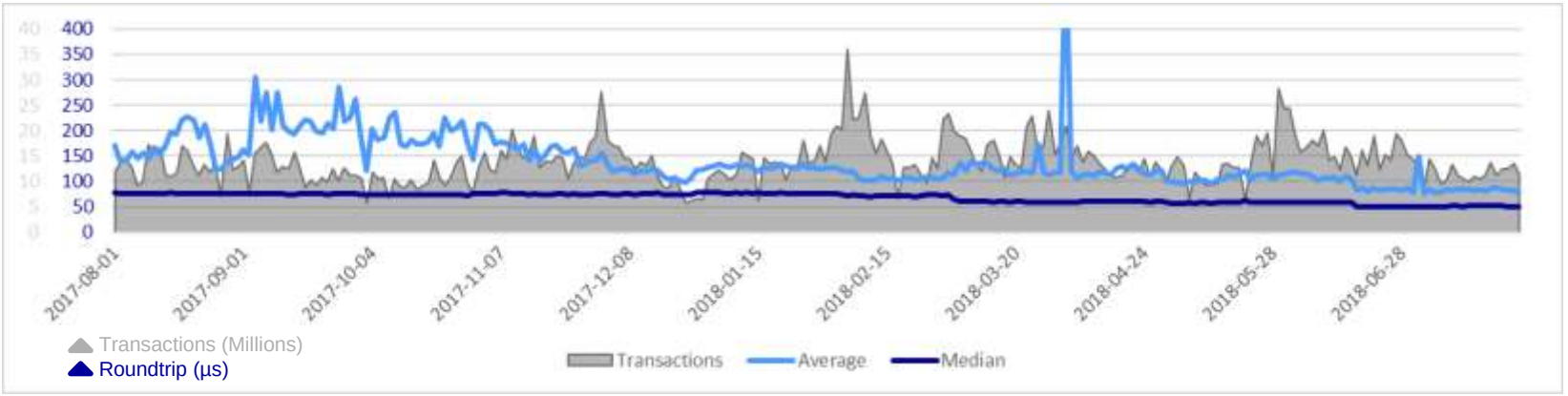
Deutsche Börse AG opens up international capital markets for its customers. Its product and service portfolio covers the entire process chain – from pre-IPO services and the admission of securities, through securities and derivatives trading through the settlement of transactions and the provision of market information to the development and operation of electronic trading, clearing and settlement systems. With its process-oriented business model, Deutsche Börse increases the efficiency of capital markets. Committed employees are the key factor for innovation and further growth: without them, Deutsche Börse Group would not have developed into one of the most modern exchange organisations in the world. More than 5,000 employees work for the Group – a dynamic, motivated and international team.

PS gateway

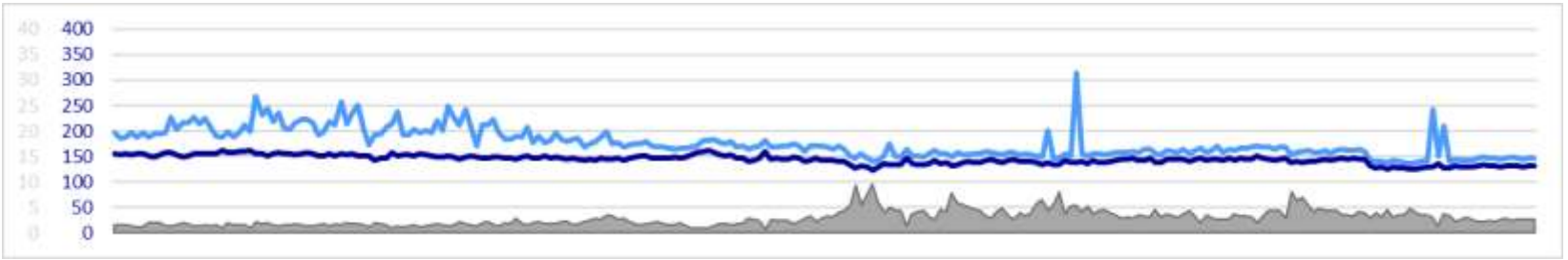
Latency (Eurex Futures)

The introduction of PS gateways allowed us to optimise the architecture and reduce the latency for high frequency sessions further.
The ratio of transactions sent via low frequency sessions almost doubled.

High frequency



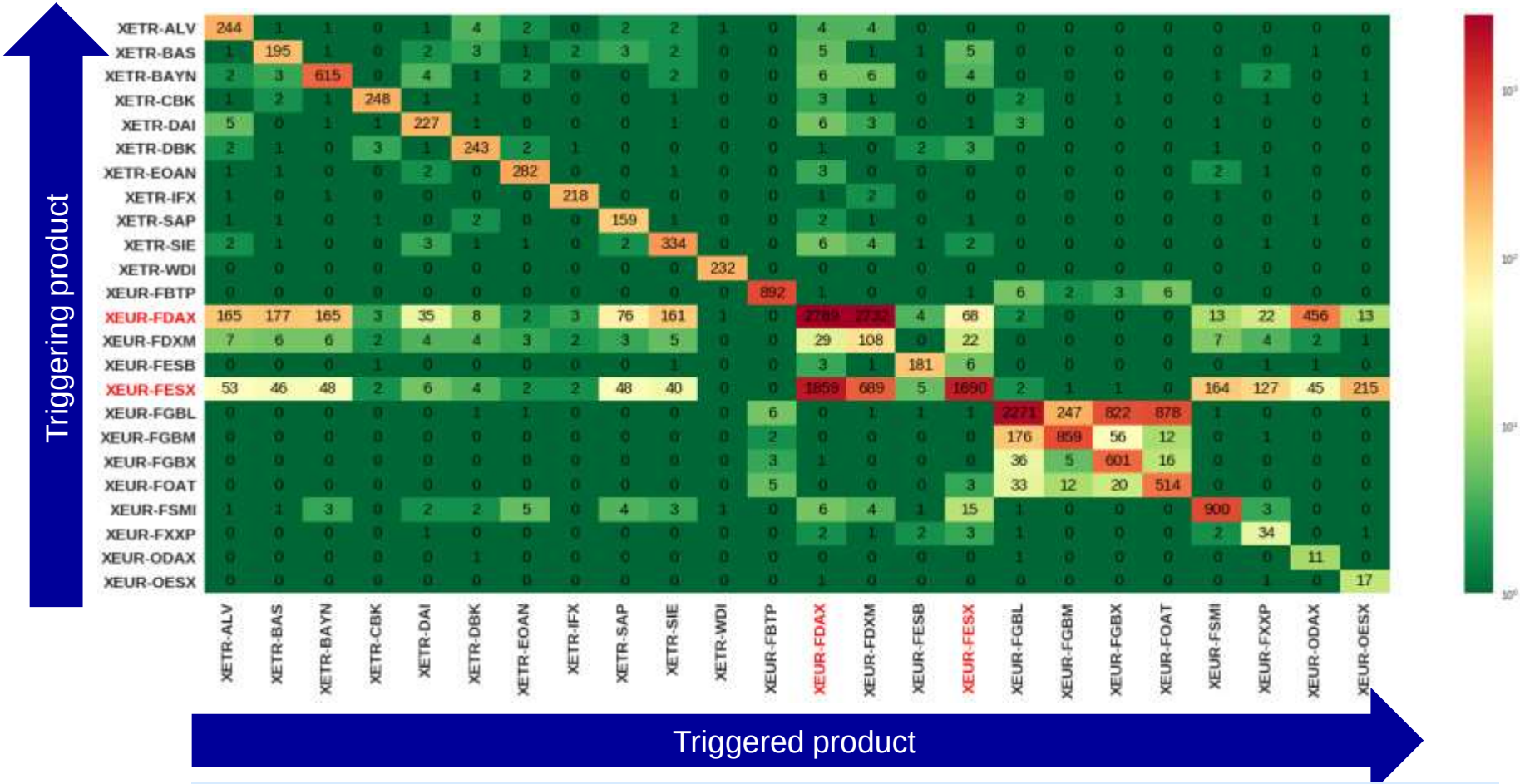
Low frequency



LF ratio

High Precision Timestamp (HPT) file service

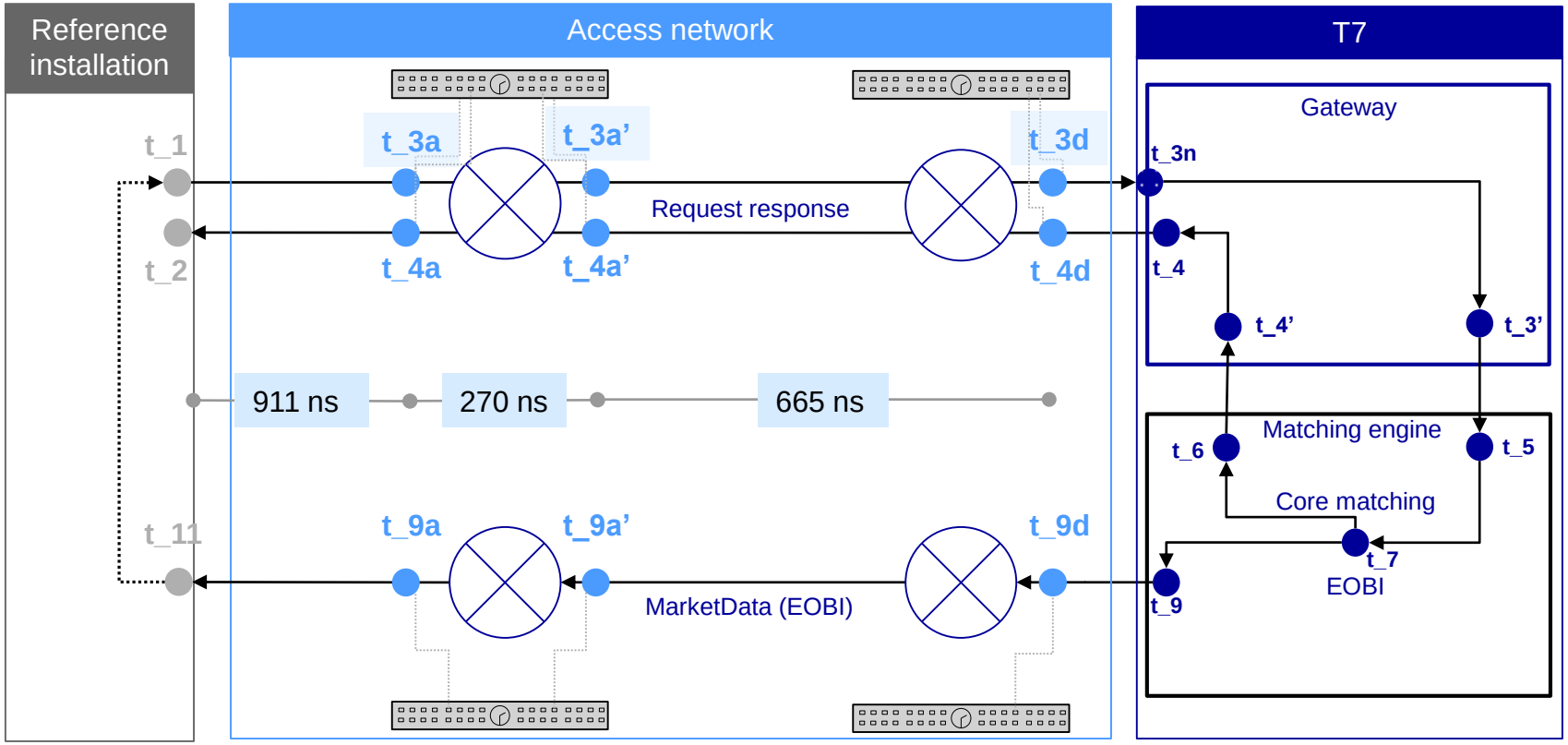
Fast correlations



"Fast correlation matrix" between different Eurex and Xetra instruments

Shown are the number of trades that followed another trade within five microseconds.

T7® timestamps



- Timestamps provided in T7 API (in real time) in dark blue (t_3n: taken by network card, other: application level)
- Network timestamps taken using taps and timestamping switches (Metamako)
- Timestamps possibly taken by participants shown in grey

Useful (?) units

1 *m*s	= 1 thousandth of a second
1 *μ*s	= 1 millionth of a second
1 *n*s	= 1 billionth of a second
1 *μ*min	= 60 *μ*s
1 *μ*h	= 3.6 *m*s
1 *n*century	~ *π s*

Distance light travels in

air	**fibre**
300 km	200 km
300 m	200 m
30 cm	20 cm
18 km	12 km
1080 km	720 km
1 Million km	