Understanding an ultra-fast market through ultra-accurate time synchronization

Andreas Lohr and Sebastian Neusüß
Problem Statement

How to capture and timestamp all customer cross-connects in co-location?

- 500+ capture ports
- 60+ capture devices
- 4 data center modules
- PTP +/- 60ns jitter in our infrastructure at best
- Serialization time of order entry message = 120ns
- Goal of sub-10ns precision
- Distances too long for PPS over coax cables
What is White Rabbit?

- White Rabbit is a fully deterministic Ethernet-based network for **data transfer** and **time synchronization**
- Initially developed at CERN
- Provides sub-nanosecond accuracy and picoseconds precision of synchronization
- Tried and tested
- PTP over Synchronous Ethernet
- White Rabbit is the High Accuracy profile of the future standard IEEE1588-20XX
- Commercially available (except the Hello Kitty model)

Issue for us:
No native White Rabbit support in NICs and switches

Solution:
We use White Rabbit to distribute 1PPS
White Rabbit deployed hardware
Timestamping devices synchronized by 1PPS over White Rabbit
Time Synchronization

PTP Example

![Diagram showing PTP Example]

PTP t_3a > t_3n (three partitions)
Time Synchronization

White Rabbit Example

![Diagram showing time synchronization and White Rabbit example with t_3a and t_3d markers.](image)
PTP / White Rabbit

PTP

Parent Offset

50ns

Metawatch Sync Error

0.5ns

White Rabbit
White Rabbit Services for our Trading Customers

Launched two WR based services

1. High-Precision Timestamp (HPT) File (more in a moment)
2. High Precision Time Service (connect to our WR network)

Value 42.96ps
Mean -25.07ps
Min -48.34ps
Max 16.14ps
St Dev 20.54ps
White Rabbit Synchronized Timestamps

Latency from customer rack to t_3a (unadjusted, i.e. incl. cabling)

- **Mean**: 945.4, 945.2, 947.3, 947.5
- **Standard Deviation**: 1, 1.2, 1.1, 1.1
- **Minimum**: 942, 941, 944, 944
- **5%**: 944, 943, 946, 946
- **10%**: 944, 944, 946, 946
- **25%**: 945, 944, 947, 947
- **50%**: 945, 945, 947, 947
- **75%**: 946, 946, 948, 948
- **95%**: 947, 947, 949, 949
- **99%**: 948, 948, 950, 950
- **Maximum**: 949, 949, 951, 951
- **Range**: 7, 8, 7, 7

**Market**: eurex, xetra, eurex, xetra
I will be right after the break
Network Design

Order Entry and Market Data Networks separated

> 260 order entry lines individually 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
Network Designs

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 gray

*latencies are mid tap to mid tap on first bit
Why we need nanosecond precision and accuracy

Interarrival time at network entry \((t_{3a})\)

Interarrival time at \(t_{3a}\) arbitrarily small

⇒ Best precision and accuracy needed
Why we need nanosecond precision and accuracy

Measuring switch jitter

Order entry switch jitter

+/-5 ns*

⇒ Best precision and accuracy needed

*5th to 95th percentile: 9 ns
Why we need nanosecond precision and accuracy

Measuring switch queuing

Nanobursts lead to queuing depending on number of packets and packet size.
Why we need nanosecond precision and accuracy

Switch determinism

Overtaking probability drops to zero above 5 ns interarrival time

Interarrival time at t_3a/t_3a'

Overtaking Probability
Why we need nanosecond precision and accuracy

Interarrival time at network entry (t_3a) compared to gateway entry (t_3n)

Interarrival time at t_3a arbitrarily small
Interarrival time at t_3n at least serialization time of preceding packet (> 100 ns)
Network card resolution 8 ns

⇒ t_3a for precise interarrival time
High Precision Timestamp File

Provides network timestamps to trading participants
- Order Entry timestamps at network entry point t_3a
- Market Data timestamps at common network measurement point t_9d

Takes time synchronization and queuing effects out of the equation

Allows more precise judgement on competition ("How far behind am I?")

Allows back testing with highest timestamp quality

Is available as ‘executions only’ for all products and ‘all market data updates’ for futures and equities.
High Precision Timestamp File
Reaction times of trading participants

* Confidence Interval: 5th to 95th percentile
High Precision Timestamp File

FESX => FDAX Reaction time based on T7® times (t_9d to t_3n)*

Median network transport latency (~2736 ns)  
*Distribution of t_3n – t_9d – median (t_3n – t_3a) shown
High Precision Timestamp File

FESX => FDAX Reaction time based on T7® times \((t_{9d} \text{ to } t_{3n})^{*}\) (close up)

*Distribution of \((t_{3n} – t_{9d})\) – median \((t_{3n} – t_{3a})\) shown

Median network transport latency (~2736 ns)
High Precision Timestamp File
FESX => FDAX Reaction time based on network times (t_9d to t_3a)

Fastest reaction median ~ 2775 ns (39 ns net)

Gray distribution shows correction using median(t_3n-t_3a)

Median network transport latency (~2736 ns)
High Precision Timestamp File
FESX => FDAX Reaction time – now compared to half a year ago

Fastest reaction median Jan 2019: ~ 2795 ns (59 ns)*

Fastest reaction median ~ 2775 ns (39 ns net)

Gray distribution shows data from January 2019

Median network transport latency (~2736 ns)
High Precision Timestamp File

Reaction times of trading participants

<table>
<thead>
<tr>
<th>Participant</th>
<th>Access network</th>
<th>T7®</th>
</tr>
</thead>
<tbody>
<tr>
<td>t_1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t_11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t_3a</td>
<td>Request</td>
<td></td>
</tr>
<tr>
<td>t_9a</td>
<td>MarketData (EOBI)</td>
<td>912 +/- 2 ns*</td>
</tr>
<tr>
<td>t_9d</td>
<td></td>
<td>902 +/- 6 ns*</td>
</tr>
<tr>
<td>t_9</td>
<td>EOBI</td>
<td></td>
</tr>
</tbody>
</table>

~ 39 +/- 8 ns*

* Confidence Interval: 5th to 95th percentile
I will be right after the break
What could possibly go wrong?

https://commons.wikimedia.org/wiki/File:Hair pulling stress.jpg by stuartpilbrow, licensed under cc-by-sa-2.0
What could possibly go wrong
Tap to timestamp cable length

*latencies are mid tap to mid tap, but measurements (e.g. t_9d, t_3a) include tap to timestamping cable length
Exhibit 1

Let's just upgrade - what could possibly go wrong?

SOMEONE'S TICK-TO-TRADE IMPROVED BY 30 NANOSECONDS

count

one-way-latency in the lab [nanoseconds]
Exhibit 2

Mean
140 ps
-4.018 ns
-4.221 ns
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
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.