White Rabbit: a next generation synchronization and control technology for large distributed systems

Maciej Lipiński

Hardware and Timing Section
Beam Controls Group
CERN

PERG
Institute of Electronic Systems
Warsaw University of Technology

Future Internet Engineering
Video-conference
23rd November 2012
What is White Rabbit?

- Accelerator’s control and timing
- International collaboration
- Based on well-known technologies
- Open Hardware and Open Software
- Main features:
  - transparent, **high-accuracy** synchronization
  - low-latency, **deterministic** data delivery
  - designed for **high reliability**
What is White Rabbit?

- Accelerator’s control and timing
- International collaboration
- Based on well-known technologies
- Open Hardware and Open Software
- Main features:
  - transparent, **high-accuracy** synchronization
  - low-latency, **deterministic** data delivery
  - designed for **high reliability**

![Ethernet]

+ synchronism
+ determinism
White Rabbit – enhanced Ethernet

- Few thousands nodes
- Fiber medium
- Up to 10 km fiber links
- Bandwidth: 1 Gbps
- WR Switch: 18 ports
- Non-WR Devices
- Ethernet features (VLAN) & protocols (SNMP)
White Rabbit – enhanced Ethernet

Two separate services (enhancements to Ethernet) provided by WR:

- High accuracy/precision synchronization
- Deterministic, reliable and low-latency Control Data delivery
Why White Rabbit?

- Renovation of CERN General Machine Timing (GMT)
- GMT is great but...:
  - **RS-422**, 500kbps
  - **Unidirectional** communication
  - Separate network required
  - **Custom design**, complicated maintenance
- White Rabbit is meant to solve these problems
Why White Rabbit?

- Renovation of CERN General Machine Timing (GMT)
  - GMT is great but...:
    - **RS-422**, 500kbps
    - Unidirectional communication
    - Separate network required
    - Custom design, complicated maintenance
  - White Rabbit is meant to solve these problems
Why White Rabbit?

- Renovation of CERN General Machine Timing (GMT)
- GMT is great but...:
  - **RS-422**, 500kbps
  - **Unidirectional communication**
  - Separate network required
  - **Custom design, complicated maintenance**
- White Rabbit is meant to solve these problems
High accuracy/precision synchronization

Deterministic, reliable and low-latency Control Data delivery
Synchronization with **sub-ns** accuracy and **ps** precision

Combination of

- Precision Time Protocol (**PTP**) synchronization
- Synchronous Ethernet (**SyncE**) syntonization (L2)
- Digital Dual-Mixer Time Difference (**DDMTD**) phase detection
Precision Time Protocol (IEEE1588)

- Packet-based synchronization protocol (mapping over different physical medium)
- Synchronizes local clock with the master clock
- Link delay evaluated by measuring and exchanging packets tx/rx timestamps
All network devices use the same physical layer clock

Clock is encoded in the Ethernet carrier and recovered by the receiver chip (PHY).
DDMTD: Phase tracking

- PTP limitation: timestamping granularity
- Solution: take advantage of SyncE and measure phase shift
WR PTP

- Extension to PTP (IEEE1588) – defined as PTP Profile
- Addresses PTP’s limitations (granularity, asymmetry, syntonization)
- Compatible with “standard” PTP gear
- Ongoing standardization effort
- Lab & field-tested for sub-ns synchronization
Extension to PTP (IEEE1588) – defined as PTP Profile
Addresses PTP’s limitations (granularity, asymmetry, syntonization)
Compatible with ”standard” PTP gear
Ongoing standardization effort
Lab & field-tested for sub-ns synchronization

According to ISPCS Plug Fest results ...

... White Rabbit is the most accurate PTP implementation in the world!
WR PTP Standardization effort

- We want to standardize WR PTP
WR PTP Standardization effort

- We want to standardize WR PTP
- Many possibilities
  - Profile (ITU-T, IEEE, ...)
  - AVB gen 2
  - Consortium
WR PTP Standardization effort

- **We want to standardize WR PTP**
- Many possibilities
  - Profile (ITU-T, IEEE, ...)
  - AVB gen 2
  - Consortium
- **WR Standardization Group**
  - John Eidson
  - ITU-T/IEEE people
  - Companies
WR PTP Standardization effort

- We want to standardize WR PTP
- Many possibilities
  - Profile (ITU-T, IEEE, ...)
  - AVB gen 2
  - Consortium
- WR Standardization Group
  - John Eidson
  - ITU-T/IEEE people
  - Companies

John Eidson:
“Why don’t you propose to include WR into PTPv3? You could do it in that way...”
WR PTP Standardization effort

- **We want to standardize WR PTP**
- Many possibilities
  - Profile (ITU-T, IEEE, ...)
  - AVB gen 2
  - Consortium
- WR Standardization Group
  - John Eidson
  - ITU-T/IEEE people
  - Companies
- ISPCS2012:
  - PTP will be opened for revision
  - WR PTP proposed to be included in PTPv3

---

John Eidson:

“Why don’t you propose to include WR into PTPv3? You could do it in that way...”
WR PTP Standardization effort

Standardization goal

WR PTP included into PTPv3
WR PTP Standardization effort

Standardization goal

WR PTP included into PTPv3
Data distribution in White Rabbit

- High accuracy/precision synchronization
- Deterministic, reliable and low-latency Control Data delivery
Data distribution in WR Control System

Control Message

Event

ID
Payload

Data Master

Switch

Switch
Data distribution in WR Control System

Control Message

Event

<table>
<thead>
<tr>
<th>Requirement</th>
<th>GSI</th>
<th>CERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max latency</td>
<td>100μs</td>
<td>1000μs</td>
</tr>
<tr>
<td>CM failure rate</td>
<td>$3.17 \cdot 10^{12}$</td>
<td>$3.17 \cdot 10^{11}$</td>
</tr>
<tr>
<td>CMs lost per year</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$d_{\text{max}}$ from DM</td>
<td>2km</td>
<td>10km</td>
</tr>
<tr>
<td>CM size</td>
<td>200-500 bytes</td>
<td>1200-5000 bytes</td>
</tr>
<tr>
<td>Accuracy</td>
<td>8ns</td>
<td>1μs to 2ns</td>
</tr>
</tbody>
</table>
Control Data

- Two types of data:
  - **Control Data** (High Priority, HP)
  - Standard Data (Best Effort)

- Characteristics of **Control Data**
  - Sent in Control Messages
  - Sent by Data Master(s)
  - Broadcast (one-to-alot)
  - Deterministic and low latency
  - Reliable delivery
Data Redundancy

- Re-transmission of Control Data not possible
- **Forward Error Correction** – additional transparent layer:
  - One Control Message encoded into N Ethernet frames,
  - Recovery of Control Message from any M (M<N) frames
- FEC can prevent data loss due to:
  - **bit error**
  - network reconfiguration
Topology Redundancy

- Standard Ethernet solution: Rapid/Multi Spanning Tree Protocol
- Reconfiguration time: \( \approx 1\text{s} \) (best: milliseconds)
- \( 1\text{s} = \approx 82,000 \) Ethernet Frames lost
- Solution:
  - take advantage of FEC
  - speed up (R/M)STP – >eRSTP or
  - use multiple paths – >eLACP
**Determinism and low latency**

- **Control Data:** 7\textsuperscript{th} Class of Service (priority)
- **WR Switch:**
  - Quality of Service: resource reservation
  - Upper bound latency by design: <10\textmu s
  - Cut-through
- Careful diagnostics
Determinism and low latency

- Control Data:
  7\textsuperscript{th} Class of Service (priority)

- WR Switch:
  - Quality of Service: resource reservation
  - Upper bound latency by design: < 10\textit{us}
  - Cut-through

- Careful diagnostics
Data Distribution summary

- Optional feature
- Openness enables everyone to verify the parameters
- Ongoing efforts (2012/2013)
- Commonalities with IEEE effort for 2\textsuperscript{nd} gen Audio Video Bridging
A White Rabbit network is composed of

- **WR Switches**
- **WR Nodes**
- **WR Timing Master**
- **Copper/Fiber links**

Same optical link used for transmission/reception.
White Rabbit Switch (V3)

- Central element of WR network
- Original design optimized for timing, designed from scratch
- 18 1000BASE-BX10 ports
- Capable of driving 10 km of SM fiber
- Open design (H/W and S/W)
WR Node: WR PTP Core

- Ethernet
- External PHY
- External oscillators
- EEPROM
- WR PTP Core
  - TBI/Serdes
  - CLKREF
  - CLKDMTD
  - adjust
  - 1-PPS
  - timecode
  - frequency
- MAC I/F
- Pipelined WB Slave I/F
- timing I/F
- control/status pins

Table of Contents:
- Introduction
- Time Distribution
- Data Distribution
- Components
- Applications
- Performance
- FIE and WR
- Summary
- Q&A
WR Node: SPEC board

Co-HT FMC-based Hardware Kit:
- FMCs (FPGA Mezzanine Cards) with ADCs, DACs, TDCs, fine delays, digital I/O
- Carrier boards in PCI-Express, VME and uTCA formats
- All carriers are equipped with a White Rabbit port
White Rabbit applications

- Control and timing system
- Field bus recommended at CERN
- Time Transfer
- RF distribution
- Distributed oscilloscope
- ...
WR at CERN

- 4 accelerator networks
- **Separate Data Master (DM) for each network**
- **LIC Data Master** communicates with other DMs and control devices in their networks
- Broadcast of **Control Messages** within network(s)
Ethernet Clock distribution a.k.a. Distributed DDS

Distributed Direct Digital Synthesis

- Replaces dozens of cables with a single fiber.
- Works over big distances without degrading signal quality.
- Can provide various clocks (TTC, RF, bunch clock) with a single, standardized link.
**Distributed oscilloscope**

- **Common clock in the entire network**: no skew between ADCs.
- **Ability to sample with different clocks via Distributed DDS.**
- **External triggers can be time tagged with a TDC and used to reconstruct the original time base in the operator’s PC.**
CERN Neutrinos to Gran Sasso (CNGS)
Investigation of neutrino oscillation
Time of Flight (ToF) measurement
Other White Rabbit applications

Future applications:
- GSI
Other White Rabbit applications

Future applications:

- GSI
- HiSCORE: Gamma&Cosmic-Ray experiment (Tunka, Siberia)
Other White Rabbit applications

Future applications:
- GSI
- HiSCORE: Gamma&Cosmic-Ray experiment (Tunka, Siberia)
- The Large High Altitude Air Shower Observatory (China)
Other White Rabbit applications

- **Future applications:**
  - GSI
  - HiSCORE: Gamma & Cosmic-Ray experiment (Tunka, Siberia)
  - The Large High Altitude Air Shower Observatory (China)

- **Potential applications:**
  - **SuperGPS through optical networks**
Other White Rabbit applications

Future applications:
- GSI
- HiSCORE: Gamma&Cosmic-Ray experiment (Tunka, Siberia)
- The Large High Altitude Air Shower Observatory (China)

Potential applications:
- SuperGPS through optical networks
- Cherenkov Telescope Array
Other White Rabbit applications

- **Future applications:**
  - GSI
  - HiSCORE: Gamma&Cosmic-Ray experiment (Tunka, Siberia)
  - The Large High Altitude Air Shower Observatory (China)

- **Potential applications:**
  - SuperGPS through optical networks
  - Cherenkov Telescope Array
  - ITER
Other White Rabbit applications

Future applications:
- GSI
- HiSCORE: Gamma&Cosmic-Ray experiment (Tunka, Siberia)
- The Large High Altitude Air Shower Observatory (China)

Potential applications:
- SuperGPS through optical networks
- Cherenkov Telescope Array
- ITER
- European deep-sea research infrastructure (KM3NET)
WR time transfer performance: lab tests

stable oscillator

Symmetricom CS4200 Cesium beam clock

WR Switch (master)

10 M in

Ref clock

PPS out

DP0

WR Switch (slave 1)

DP0 UP0

Ref clock

PPS out

WR Switch (slave 2)

UP0

Ref clock

PPS out

WR Switch (slave 3)

UP0

Ref clock

PPS out

10 MHz

5 km - long rolls of fiber G.652

hot-air gun

IN

Agilent E5052 Signal Source Analyzer

PSD

LeCroy WavePro 7300A oscilloscope

clock offset analysis

phase noise analysis
WR time transfer performance: lab tests

Histogram of offsets between master and each slave:

- **Slave 3 (C4)**
  - Mean: -135.25 ps
  - Standard Deviation: 6.14 ps

- **Slave 2 (C3)**
  - Mean: 24.67 ps
  - Standard Deviation: 5.30 ps

- **Slave 1 (C2)**
  - Mean: 161.86 ps
  - Standard Deviation: 5.45 ps

Table of collected data:

<table>
<thead>
<tr>
<th>Measure</th>
<th>P1: skew(C1,C2)</th>
<th>P2: skew(C1,C3)</th>
<th>P3: skew(C1,C4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>176 ps</td>
<td>42 ps</td>
<td>-117 ps</td>
</tr>
<tr>
<td>mean</td>
<td>161.86 ps</td>
<td>24.67 ps</td>
<td>-135.25 ps</td>
</tr>
<tr>
<td>min</td>
<td>141 ps</td>
<td>3 ps</td>
<td>-846 ps</td>
</tr>
<tr>
<td>max</td>
<td>183 ps</td>
<td>48 ps</td>
<td>-109 ps</td>
</tr>
<tr>
<td>sdev</td>
<td>5.45 ps</td>
<td>5.30 ps</td>
<td>6.14 ps</td>
</tr>
<tr>
<td>num</td>
<td>59.764e+3</td>
<td>59.764e+3</td>
<td>59.757e+3</td>
</tr>
</tbody>
</table>
Duration: 31 d, 7 h, 40 s (2.7 * 10^6 samples)
WR Nodes with TDC used
Measurement includes inaccuracy of TDC
Timestamping reference PPS
Accuracy: 0.517 ns
Precision: 0.119 ns (std. dev)
Out of $2.7 \times 10^6$ samples, 9 values of $x_{\text{diff}} [0.0003\%]$ exceeded MTIE=1ns.
Future Internet Engineering

- redefines/improves 3-7 OSI Layers
- uses cutting edge 1-2 OSI Layers (Ethernet)
- virtualizes

White Rabbit

- improves 2 OSI Layer (i.e. GbE)
- brings into Ethernet Networks:
  - high accuracy synchronization
  - determinism
  - reliability
- provides hardware-support
Future Internet Engineering

- Uses cutting edge Layer 2 equipment (PIONIER)
- Large scale: used globally with millions of nodes
- Application: mass scale, public

White Rabbit

- Uses White Rabbit Layer 2 equipment
- Large scale: tens of km with thousands of nodes
- Application: dedicated, isolated, well-controlled
Summary

- **White Rabbit**
  - 2000 nodes
  - < 1ns accuracy
  - determinism and reliability
  - tested up to 10km

- FIE and WR are complementary
- FIE is general-purpose and global-scale technology
- WR is specialized-purposed and large-scale technology
- WR improves the technology that FIE uses
Questions and answers