White Rabbit

Maciej Lipiński

Hardware and Timing Section
Beam Controls Group
CERN

PERG
Institute of Electronic Systems
Warsaw University of Technology

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

- Accelerator’s control and timing
- Based on well-known technologies
- Open Hardware and Open Software
- International collaboration
- 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)
Two separate services (enhancements to Ethernet) provided by WR:

- High accuracy/precision synchronization
- Deterministic, reliable and low-latency Control Data delivery
**Time Distribution in White Rabbit Network**

- **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 UDP/IP, DeviceNet, and Ethernet)
- Synchronizes local clock with the master clock
- Link delay evaluated by measuring and exchanging packets tx/rx timestamps
Synchronous Ethernet (SyncE)

- 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
PTP is OK but ...

What are the issues... and ... how we address them

PTP-base syntonization $\Rightarrow$ SyncE

limited precision and resolution $\Rightarrow$ SyncE

unknown link asymmetry $\Rightarrow$ DDTMD phase detection

WR extension to PTP (WR PTP) for extra data exchange and logic
Extension to PTP (IEEE1588) – defined as PTP Profile
- Addresses PTP’s limitations
  (granularity, asymmetry, syntonization)
- Compatible with ”standard” PTP gear
- 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
- 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

Standardization goal: include WR into PTPv3

John Eidson:
“Why don’t you propose to include WR into PTPv3? You could do it in that way...”
Data distribution in White Rabbit

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

<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_{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/Multicast (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 $\rightarrow$ eRSTP or
  - use multiple paths $\rightarrow$ eLACP
Determinism and Low Latency

Control Data:
- 7th Class of Service (priority)

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

Careful diagnostics
White Rabbit Network Components

- **Components**
- **Applications**
- **Performance**
- **FIE and WR**
- **Summary**
- **Q&A**

The diagram illustrates the White Rabbit network components, including:

- Master
- Switches
- Nodes

Nodes include:

- Sensor
- Actuator
- Database
- Actuator
- Monitoring

Connections and data/timing information flow through the network.
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
- TBI/Serdes
- WR PTP Core
- 1-PPS timecode frequency
- control/status pins
- I²C
- External oscillators
- CLKREF, CLKDMTD, adjust
- source
- sink
- MAC I/F
- Pipelined WB Slave I/F
- timing I/F
- WR PTP Core
- WR Node: WR PTP Core
- Maciej Lipiński

- 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

Much more than accelerator control and timing system ...
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/Multicast of **Control Messages** within network(s)
WR at CERN

LHC Network

CCR (Prevessin)

Data Masters (WR Nodes)

LHC (node)
LIC (node)
AD (node)
REX (node)

Timing Masters

Time source

WR Network Backbone

ISOLDE

LIC Network (Meyrin)

Maciej Lipiński
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 (e.g. 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
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)

Institute for Nuclear Research of the Russian Academy of Sciences
- Moscow State University
- Irkutsk State University
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)

REF clock
PPS out

10 MHz

10 M in

WR Switch (slave 3)

UP0
REF clock
PPS out

5 km - long rolls of fiber G.652

WR Switch (slave 2)

DP0 UP0
REF clock
PPS out

WR Switch (slave 1)

DP0 UP0
REF clock
PPS out

IN

LeCroy WavePro 7300A oscilloscope
clock offset analysis

Agilent E5052 Signal Source Analyzer
phase noise analysis
WR time transfer performance: lab tests

Histogram of offsets between master and each slave:

- Slave 3 (C4): mean = -135.25 ps, sdev = 6.14 ps
- Master (C1): mean = 24.67 ps, sdev = 5.30 ps
- Slave 1 (C2): mean = 161.86 ps, sdev = 5.45 ps

Matlab plot of collected data:

Oscilloscope screenshot:

<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>
**WR Time Transfer Performance: CNGS Deployment**

- **System performance**
  - **Accuracy:** 0.517 ns
  - **Precision:** 0.119 ns (std. dev)

- **Duration:** 31 d, 7 h, 40 s (2.7 \times 10^6\) samples)

- **Timestamping reference PPS with 300 ps accuracy and 55 ps precision**
Out of $2.7 \times 10^6$ samples, 9 values of $\chi_{diff}$ [0.0003%] exceeded MTIE=1ns.
Future Internet Engineering redefines/improves 3-7 OSI Layers
uses newest solutions for 1-2 OSI Layers
visualizes

White Rabbit
improves 2 OSI Layer (i.e. GbE)
high accuracy synchronization
determinism
reliability
hardware-supports
Future Internet Engineering
- uses cutting edge Layer 2 equipment (e.g.: PIONIER)
- global scale: thousands of km with millions of nodes
- application: mass, public

White Rabbit
- requires White Rabbit Layer 2 equipment
- large scale: tens of km with thousands of nodes
- application: dedicated, isolated, well-controlled
Summary

- **White Rabbit**
  - Thousands 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
Extras
6 accelerators
LHC: 27km perimeter
Thousands of devices to be controlled and synchronized
A huge real-time distributed system
**Events** – points in time at which actions are triggered

Each event is identified by an ID
Devices are subscribed to events
Each device "knows" what to do on particular event
Each event (ID) has a trigger time associated
A set of events is sent as a single **Control Message (CM)**
CM is broadcast to all the end devices (nodes)
Granularity Window:

- Controller-input to node-output (i.e. pulse)
- Maximum bound latency guaranteed by the system
- Processing and network latency included
Time Distribution in White Rabbit

- Synchronization with **sub-nano-second** accuracy over fiber
- Combination of
  - Precision Time Protocol (**PTP**) synchronization
  - Synchronous Ethernet (**SyncE**) synchronization
  - Digital Dual-Mixer Time Difference (**DDMTD**) phase detection

**WR Link:**

![Diagram of WR Master and WR Slave with offset and delay markings]
Fine Delay Measurement
Link Delay Model

\[
\begin{align*}
\text{delay}_{ms} &= \Delta_{txm} + \delta_{ms} + \Delta_{rxs} \\
\text{delay}_{sm} &= \Delta_{txs} + \delta_{sm} + \Delta_{rxm}
\end{align*}
\]

Relative Delay Coefficient ($\alpha$) for 1000BASE-BX10 over a Single-mode Optical Fiber

\[
\delta_{ms} = (1 + \alpha) \delta_{sm}
\]
Fixed Delays

Asymmetry sources: circuit SFP Serdes fiber
Link Asymmetry

\[ \mu = \frac{(t_4 - t_1) - (t_3 - t_2)}{2} \]

\[ \text{delay}_{ms} = \mu + \text{asymmetry} \]

\[ \text{offset}_{ms} = t_2 - (t_1 + \text{delay}_{ms}) \]
**Link Delay Model: fiber optic solution**

\[
\begin{align*}
\text{Announce} & \quad t_1 \\
\text{Sync} & \quad t_2 \\
\text{Follow_Up} & \quad t_3 \\
\text{Delay_Req} & \quad t_4 \\
\text{Delay_Resp} & \\
\end{align*}
\]

**Master time**  \quad **Slave time**

**Relative Delay Coefficient:**

\[
\alpha
\]

**Fixed Delays:**

\[
\begin{align*}
\Delta_{\text{txs}} & \quad \Delta_{\text{txm}} \\
\Delta_{\text{rxs}} & \quad \Delta_{\text{rxm}} \\
\end{align*}
\]

**Delay\text{\_ms} = \mu + \text{asymmetry}**

**Offset\text{\_ms} = t_2 - (t_1 + \text{delay\text{\_ms}})**

---

**Solution for Ethernet over a Single-mode Optical Fiber**

\[
\text{asymmetry} = \Delta_{txm} + \Delta_{rxs} - \frac{\Delta - \alpha\mu + \alpha\Delta}{2 + \alpha}
\]
White Rabbit extension to PTP (WRPTP)

- WR-peers recognition
- Calibration (fixed delays measurement)
- Exchange of WR-data
- Support of redundancy
- Mapping over IEEE802.3/Ethernet
WR-peer recognition and WR-data exchange
**WR Link Setup**

- **WR Announce MSGs**
- **Announce TLV**

**PTP Node**

- **PTP Master**
- **PTP Uncalibrated**
- **PTP Slave**

**WR Finite State Machine**

- Frequency locking
- Calibration
- Exchange of WR-parameters
- WR Signaling Messages
WR time transfer performance: temperature tests

CERN Control System

Maciej Lipiński

White Rabbit 54/38

Histogram of skew

sdev=36ps
sdev=55ps

Climatic Chamber

clock offset analysis

LeCroy WavePro 7300A oscilloscope

Single Mode Fiber (G.652.B type)

WR Switch (grandmaster)

WR Switch (slave)

10m
5km
11km