WHITERABBIT - A NOVEL, HIGH PRECISION TIMING SYSTEM

M. Kreider, R. Baer, T. Fleck, C. Prados (GSI, Darmstadt)
E. Garcia Cota, J. Serrano, T. Wlostowski (CERN, Geneva)

Abstract

The WhiteRabbit timing network is a deterministic field bus, based on synchronous GBit Ethernet and the Precision Time Protocol (PTP). The WR protocol was designed to provide precise timing and event distribution for high end real-time systems and was therefore chosen as the timing basis for the new GSI FAIR accelerator facility. With precise phase measurement to compensate for signal propagation delay, a timing accuracy down to sub-nanosecond range is feasible. To achieve necessary determinism and robustness (packet loss of 10^{-12}), an OSI layer two Forward Error Correction and Quality of Service protocol have been introduced to the concept. Special switches wield the WR protocol, while being transparent to normal Ethernet traffic. Switch hardware is currently under development at CERN and will be a mixed FPGA/CPU solution. Working prototype cards have been introduced at the 3rd WR Workshop at CERN in 2009, demonstrating phase measurement and PTP capabilities. The presentation will contain detail on technical concepts, current project status, as well as future areas of application will be part of the discussion.

INTRODUCTION

Purpose

WhiteRabbit was designed to provide very accurate clock synchronisation to a facility and control its machines with equal precision. Any event sent to a physical machine causes a certain action to be executed at a given absolute time.

The goal here is to the know the exact link delay to destination in advance, so each outgoing event can be sent out early enough to arrive on time.

In order to achieve that, certain unpredictable factors to the response time have to be addressed. One is packet loss due to data corruption on the physical medium, the other factor is collisions resulting from packet switching in the network.

NETWORK LAYOUT

WR utilises GigaBit Ethernet on fiber or copper links. Fiber links have an advantage here, because copper tranceivers and their channel encoding logic are more complex and often show a non-determinstic behavior. Optical links enable a higher measurement accuracy on link delay.

The topology of WR system may take any non-meshed form, since time synchronisation must be unidirectional. If the network is indeed meshed, a Spanning-Tree algorithm must be used to avoid loops in time distribution.

GSI/FAIR is planning to employ a Tree Topology with a GPS receiver as UTC timing reference at the source. Below come several layers of switches, fanning timing out to endpoints throughout the facility.

WR uses special switches and endpoints to wield its protocol. Current design of the WR switch has one uplink and sixteen downlinks, each has a second physical port for redundancy. GSI/FAIR is planning a system with roughly two thousand timing receivers

Making extensive use of commercially available ethernet basic components lowers costs for WR switches and endpoints. It will be possible to integrate non-White rabbit nodes into the network. WR is compatible with PTP devices and can time sync these nodes. However, PTP nodes can only be synchronized with reduced accuracy, since they lack the special hardware for high precision phase measurement.

General purpose ethernet nodes could also be connected to the network. While being compatible with basic functions, WhiteRabbit design does not support full Ethernet standard at the time.

TECHNOLOGY

Synchronous Ethernet - SyncE

SyncE describes the special case of IEEE 802.3 ethernet standard where the recovered RX clock from its master is used as its own TX clock, making the whole system synchronous. 8b/10b channel encoding is used to make RX clock recovery from the incoming RX data signal possible. This adjustment is done in hardware and is the basis for the PTP fine measurements.

Phase Measurement - Aliasing and DPLL

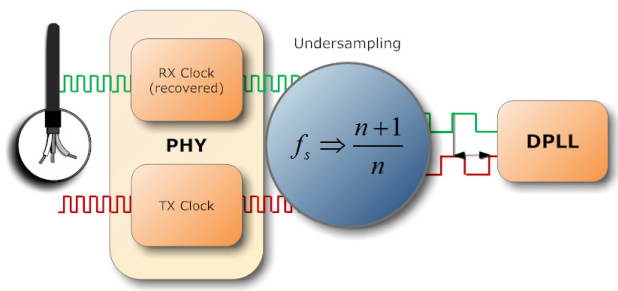

Figure 1: Aliasing and Phasemeasurement

After SyncE has adjusted the PTP clients frequency to the masters, the PTP can now measure the time difference

and lag between nodes. With WR, this is aided by hardware doing the precise measurement on the clocks phase difference, bringing timing accuracy from 8ns to a theoretical value of 32ps. In order to get highest precision, the clocks frequency would have to be in optimal range of the PLL.

The endpoint achieves this by undersampling both clocks with a frequency very close to their own. Assuming mid term stability of the oscillator, this produces low alias frequencies which lie in the optimum measuring range and still possess the proportional phase shift of the original while jitter is greatly reduced .

Time synchronisation - PTP

PTP addresses the basic problem of clock synchronisation when message lag and local time difference are unknown. A handshake between master and client is initialised, all incoming and outgoing messages are timestamped. After two messages and four time values, it is possible for the master to calculate link delay and difference of localtime. The master then communicates the correctional value down to the client which adjusts its own clock.

WhiteRabbit uses an extended version of the IEEE 1588 Precision Time Protocol. Here, synchronisation direction is fixed, hardware phase measurement increases accuracy and assymetry in link delay is taken into account.

This assymmetry is a result of chromatic dispersion coming from wavelength multiplexing in the medium. A single fiber is used for both RX and TX, employing two different wavelengths to differ between incoming and outgoing messages. Light propagation in a an optical fiber is a function of its wavelength, so RX and TX will differ in propagation delay. [1]
Choosing fiber type with a nearly equal dampening for each wavelength helps balancing signal strength and therefore maximum range.

Time is adjusted sequentially down the layers of switches and nodes. Further consideration for the link delay model are slowly changing characteristics of the physical medium, caused by temperature, moisture and aging effects.

Encoding - Forward Error Correction

The goal for WR is an event loss of 10^{-12}. Normal TCP protocol for example handles the problem of data loss by re-requesting the damaged packet. In WR, there is no time for this backup mechanism. Forward Error Correction algorithms are a class of encoding that can introduce enough redundancy to the data that chances of a packet being irrecontructably lost are minimimal. An event stream always consists of several packets, and the packet header is additionally secured with a CRC check. Individual packets themselves may be lost or corrupted, the event must reach its destination nevertheless. A detailed description and analysis of suitable algorithms and effectiveness is available here [5].

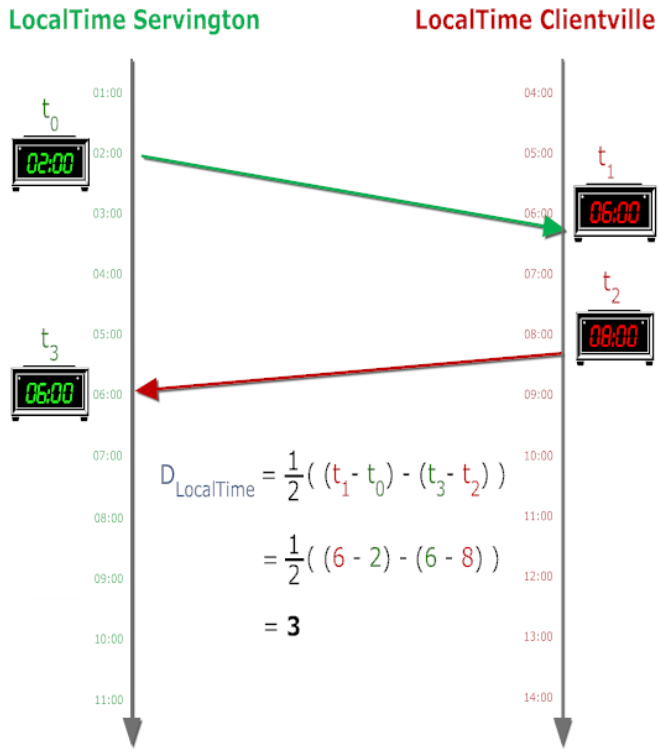

Figure 2: Simplified PTP Delay Calculations

Packet Switching - QoS

In order to guarantee absolute maximum lag time, it is necessary to prefer time critical packets to others. When a switch has more than one output packet for a port at a time, a second arriving packet must be treated differently depending on priority. Standard Priority packets , SP, can be queued if buffer space allows, else they are dropped. SP packets are not time critical and therefore re-requesting a dropped packet is possible.

Any arrival of an HP packet will cause a currently sent SP packet to be fragmented and resumed when HP traffic is over. This makes the 64 cycles maximum delay when crossing a switch possible.

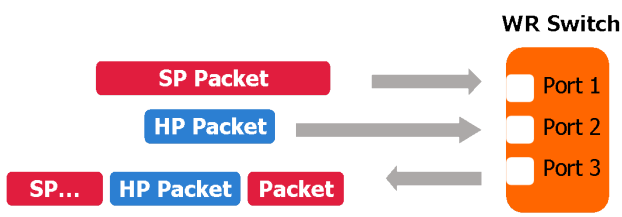

Figure 3: HP packet preempting SP packet

OPEN HARDWARE PROJECT

When thinking about the implementation of WhiteRabbit, care has been taken from the beginning not to use commercial components that come with royalty fees. At the same time, WR needed protection against possible lawsuits for suggested patent infringements or similar.

WhiteRabbit Hardware and Software is completely open and documentation and sources are available at `http://www.ohwr.org/`

CONCLUSION

Working Point-To-Point time synchronisation has first been shown at the WR workshop in 2009. Since then, switch hardware was under continous development and a first WR switch prototype with switching capabilities will be ready by the end of 2010.

Running side by side with hardware development, WR protocol specs were expanded and improved. Timing Receiver boards are also currently under development and will be made in various form factors. First planned are PCIe and VME boards to accomodate a timing receiver in many existing systems, first prototypes are expected early 2011.

WhiteRabbit is a timing system for the future. GSI is planning to control all its current and new FAIR machines over WhiteRabbit, with the only exception of the HF generators.

A fully deployed system at the FAIR facility is to be expected by 2016.

REFERENCES

[1] P.P.M. Jansweijer, H.Z. Peek, “Measuring propagation delay over a 1.25 Gbps bidirectional data link”, ETR 2010-01, Amsterdam, Netherlands, May 2010

[2] P. Moreira et al., “White Rabbit: Sub-Nanosecond Timing Distribution over Ethernet”, ISPCS 2009, Brescia, Italy, Oct 2009

[3] J. Serrano et al., “THE WHITE RABBIT PROJECT”, TUC4 ICALEPS2009, Kobe, Japan, Oct 2009

[4] WR Switch Specifications `http://www.ohwr.org/`

[5] C. Prados Boda, T. Fleck, “ FEC in Deterministic Control Systems over Gigabit Ethernet”, THPL011 PCaPAC2010, Saskatoon, Canada, Oct 2010