Meeting Minutes

CERN Timing Workshop 2

Date & Duration: 27.10.2008 10:30-18:00 and 28.10.2008 9:30-15:00

Attendees: **CERN (11):** Pablo Alvarez, Julian Lewis, Bruno Puccio, Javier Serrano, Ben Todd, Tomasz Włostowski, Ioan Kozsar, Alastair Bland, Stephen Page, Eugenia Hatziangeli, Jean Claude Bau

AAS (3): Patrick Loschmidt, , Georg Gaderer (2nd day)

GSI (5): Ralph Baer, Tibor Fleck, César Prados, Stefan Rauch, Matthias Thieme

IN2P3 (3): Dario Autiero, Jacques Marteau, Bruno Carlus

Ore (1): Nikolaus Kerö

INES (1): Hans Weibl

Cosylab (1): Igor Verstovsek

Absentees:

Location: CERN Meyrin Building 60 / CERN Prévessin Building 864

Written by: Pablo Alvarez, Patrick Loschmidt

The 2nd CERN Timing Workshop has been centred on a first proposal for the White Rabbit Protocol (WRP) functional specifications and demonstration of a prototype link with 1 ns time transfer accuracy by Tomasz Włostowski. Although the functional specifications are not considered definitive, they are a very good base for the next iteration. Work packages concerning designs, specifications and project funding are distributed to the different partners.

Workshop AM Agenda:

10:30	Introduction and status (5')	Javier Serrano (*CERN*)
10:40	Catalogue of existing solutions (35')	Nataša Simanić (*Austrian Academy of Sciences*)
11:25	White Rabbit: a proposal (35')	Tomasz Włostowski (*Warsaw University of Technology*)
12:30	End of first block	
14:00	General brainstorming, definition of future work packages	all
17:30	End of first day	
09:30	Timing theory and technologies	Javier Serrano, Pablo Alvarez (*CERN*)
14:00	Discussion about higher layer protocols Demonstration and Labs	all
16:00	End of second day	

Meeting Minutes

TOP 1 Introduction and Status

Introduction by Javier Serrano (CERN)

- Project started with Workshop in February
- Overview of the Project:
  - CERN wants to build two types of nodes (a cheap and “inaccurate” and an expensive, accurate one)
  - Legal issues have to be checked in any case, since patents could hinder open source as well
  - It should be possible to use the individual layers of the planned protocol stack independently of the others
  - The goal is to design the stack till layer 3 for control purposes but to leave it open for anybody to use it from there on
- Significant progress in WP2 and WP3
  WP2: complete user requirements delivered by Cosylab
  WP3: concept of sub-ns time transfer over fibre, preliminary spec for protocol and switch

TOP 2 Catalogue of Existing Solutions

Presentation by Nataša Simanić on State-of-the-Art on Ethernet Based Real-Time Protocols

- Ethernet POWERLINK and SERCOS III are the most "open" protocols
  - POWERLINK has inadequate asynchronous support
  - Low number of nodes for SERCOS III
  - Definitely at least study specification to get concepts and issues to be tackled in RT protocols
- EtherCAT missed possibility for open source hindered by licensing problems
  - Slave IP Cores are available, but not free
- Keep MODBUS TCP in mind
  - Concepts could be useful

Javier comments that Modbus is very used in general; Siemens normally uses Profinet and NI Ethercat. NI is also interested in WR. CERN was close to choose Powerlink, but it came out that it does not fit very well to CERN needs (asynchronous traffic and hard real-time requirements). Other investigated standard protocols had serious drawbacks as well.

TOP 3 White Rabbit: A Proposal

Presentation by Tomasz Włostowski on White Rabbit Protocol specification

- Presentation on White Rabbit Protocol
  High priority traffic has to be treated specially, as even a single frame must not be lost. Therefore, LT (Luby Transform) encoding is used based on distributing the payload into multiple frames (small MTU) and re-assembling the contents at the receiver side.
  In general the **evaluation of an appropriate failure model** is missing!
- Another important property of the WRP is the way how the latency of high priority traffic is handled. On-going low priority transmission is interrupted, yet keeping the data of the corrupted frame both at the sender and the receiver side. After completing

the high priority transmission sending the low priority frame previously interrupted is completed by merely transmitting all payload data not received already.

- Presentation on the White Rabbit Switch Concept
  It's the central component of the planned network and will be done by Tomasz till mid of next year.
- Question about having to wait for all HP packets in a group to decode a frame (Patrick).
  This behaviour is based on the fact that presently timing actions are validated at the beginning of a millisecond. Therefore, there is no need for sending frames in a single packet. (Javier)
- Nick suggests that sending this information in a single frame could save some extra BW. Javier remarks that the BW is largely covered by WR. Tomasz explains that spreading the HP information over a timing slot will decrease the probability of loosing information.
- Possible security issues when connecting a non-WR device to a WR network. (Alastair)

TOP 4 General Brainstorming

Presentation by Patrick Loschmidt on possible White Rabbit issues

First Slide

- The discussion starts with the necessity of preemption. WRP specifies a time-triggered approach, so why not just allow non-RT traffic in a special slot, like PowerLink?
  Nick argues that the notion of time is basic to the WRP. Javier mentions that if there is a node that is not compliant, PowerLink stops working (in the sense of not guaranteeing worst case latencies). Patrick argues this is not a problem because the traffic can be reordered in the switch. Pablo comments that preemption does not allow a deterministic BW.

  Nick asks if the HP traffic is event triggered. Javier confirms this is the current state for the present GMT (General Machine Timing). Julian asks if millisecond frames will be broadcast in the WRP. Javier answers that this is not necessary. The millisecond boundaries are regenerated in the slaves using PTP. The HP packets will be placed in the right position to ensure their validation.

  Tomasz argues that using preemption possibly increases the BW. Nick agrees with preemption but believes that fragmentation is very tricky to be implemented. Tomasz thinks it is not such a problem and relies on TCP. Javier says that if it turns out to be a problem we can just abandon it. Preemption is an option to Layer 2.
- Igor asks if WR will be also used as a field bus and if so, how. Javier answers that in WorldFip and Powerlink the master schedules the traffic in a fixed round robin. WorldFip is used at CERN because its hardware is based in a rad-hard chip, and also because the standard guarantees determinism in latencies. If WR is used as a field bus there should not be any non-WR devices connected.
  Julian comments that allowing non-WR devices may be a good help for monitoring. Ralph indicates that is a good idea but is a bit sceptic about the price paid for non-WR devices support.
- Javier indicates that there may be an overflow of SP frames. Pablo asks about the switch behaviour when several slaves send SP frames at the same time. Javier says we can decide that the switch will just send the first one arriving. Pablo comments about a possible BW loss when several slaves send non-SP frames simultaneously. Hans indicates that TCP can handle the packet loss. Pablo indicates that if preemption is enabled, the HP packets BW will be reduced if the SP traffic increases. Hans says that buffer overflow will indicate if there is a problem.

Second Slide

- Patrick asks why broadcasting HP frames to the whole network when multicasting could be used as well. Further, there is the possibility of removing HP traffic coming from the slaves.
  Javier replies it is necessary to mimic WorldFip behaviour. In WF there are no collisions because the master requests the HP frame. Patrick indicates that the switch behaviour should be specified. Ralph comments on marking the delayed HP packet. Tomasz pointed out the possibility of reducing the BW to ports. Dario asked if it is possible to turn SP frames into HP frame to increase accuracy.
- Patrick asked how it is guaranteed that PTP will have enough BW. Pablo suggested the possibility adding a phase measurement field to HP frames in order to reduce jitter between WR devices.

Summary of the discussion

- WRP is meant as a generic layer 2 protocol, which works transparently for upper layers, based on event-triggered pre-emption That was one important issue clarified, together with the fact that all other issues should be shifted to higher protocol layers
- The protocol should use multicast transmission instead of broadcast in order to avoid congestion on node to node communication links. (Patrick)
- Handling of collisions on upstream high priority traffic is not yet specified (first-come-first-serve?) (Javier)
- The LT coding does not secure the header and is therefore problematic; there should be thoughts about an FEC for the whole frame. Again, an appropriate error model is not yet available to define the best correction algorithm.
- High priority upstream traffic will be organized by a higher layer master, the layer 2 protocol itself won't handle anything, although not fully symmetric. (The symmetry is violated by the fact, that HP downstream can be delivered in any case, while HP upstream might produce collisions)
  This fact should be mentioned in the specification (all)
- Number of necessary types of traffic is still open
- Re-assembly of corrupted low priority frames is questioned in favour leaving packet re-transmission to higher layers (e.g. TCP).

TOP 5 Fibre Delay Compensation Demonstration

Tomasz sets up a laboratory demonstration where two VIRTEX5 prototype boards communicate with each other using a bidirectional SFP fiber transceiver and the VIRTEX5's RocketIO™ GTP transceivers. Tomasz shows how his design is able to compensate cable delays with accuracy better than 1 ns for cable lengths from 1 km up to 5 km. The fiber and the cards where heated with a fan without showing an appreciable change of phase between both cards.

TOP 6 Discussion on Further Work Packages

WR specs - layer 2

Ralph considers that Tomasz should continue and expresses its felicitations to Tomasz. There is a global agreement on this point.

Higher layer protocol specs

Julian points out that basically anyone should be able to use the layer 2 protocol without constraints (WR should not impose restrictions on upper layers). Javier suggests arriving to an agreement. There should be an agreement on L2 and L3. This is the biggest work package and everybody should participate. Julian points out that WorldFip and GMT cannot

coexist. Javier thinks that that may not be the case. Alastair comments about consoles and WR. Javier indicates that this should be discussed in an AB/CO Technical Committee.

OHR (Open Hardware Repository)

Javier mentions that the OHR will probably be working by the end of the year. The OHR will be a useful tool for reviewing and sharing projects. Patrick asks about quality ensuring mechanisms and reminds the case of OpenCores, where not all cores reach a minimum of quality. Javier argues that quality will improve by use. Concerning licensing, Javier indicates that the designer clearly specifies the license used, normally OHL. Tomasz comments that RGPL could probably be more interesting, as it isn't forcing publishing entire file sources. Javier's major concern is not to be sued for making public his own inventions, something that actually seems to happen. Julian mentions the possibility of someone using the code commercially, but Javier considers it as irrelevant.

Nick shows concern about possible problems during the production of the PCBs. Pablo asks about the possibility of using IP cores in an open source project. Javier replies that open source does not exclude this possibility. The user should care about obtaining the suitable tools.

Julian asks about access to the repository and how intruders will be rejected. Javier says that anybody can have read access to the repository. Uploading material to the repository will be possible after a formal request.

Slave

Alastair commented that using power over Ethernet could be very interesting for stand alone slaves.

Platforms

Alastair asks about platforms, which will be supported. Javier mentions PCI, PCI express, VME. The switch could be based on µTCA. Tomasz prefers a standalone switch.

Alistair talks about the possibility of installing WR in the LHC as a replacement of WF. Although WF is based on copper, there can be some synergies between both standards. Javier replies that WF is rad-hard, while WR is not supposed to be rad-hard.

When distributing new work packages 2 and 11 (higher layer protocol specs and system requirements and architecture specification) Patrick says that he does not know about accelerators requirements. Igor offers to share his knowledge.

Simulation

Patrick asks about doing a traffic simulation. Pablo replies that the reliability of the system should be based on a fully deterministic protocol. He shows his worries about use of preemption and the collisions treatment. Patrick argues that simulation could reduce the final cost of the project. Javier admitted his limited experience with traffic simulations and agreed on doing one, although Tomasz comments that, as the switch functionality is based on HDL and C code, the network simulation is not a prerequisite to the switch development.

General

Patrick commented that he needs system requirements, including architecture.

Ralph asked for some figures, for financial considerations Javier has to discuss about money and expects to get some budget for this project. Ralph responds that they just need some prototypes for the moment. Patrick suggests that European FP programs could provide up to 2-3 million Euros. There are deadlines in April and November.

Alastair comments that if a network driver is developed we should be protected from someone misusing a slave and overloading the network.

Hans offers his help in the switch test.

IN2P3 is currently working on a µTCA-based solution.

Extensive integration testing (Package 9) remains unassigned.

TOP 7 Future Work Packages

1. WR specs (layer 2). This should include a study of reliability/robustness. CERN.
2. Higher layer protocol specs. Discussion on the drivers as well. Cosylab/CERN/GSI.
3. OHR. Cosylab.
4. VHDL/Verilog choice and guidelines. CERN/AAS/IN2P3.
5. PCB design standards. Choice of a standard tool? Oregano/CERN.
6. Do a master. CERN/GSI.
7. Do a switch. CERN/(IN2P3 if µTCA).
8. Do a slave. IN2P3/AAS.
9. Extensive integration testing.
10. Planning and management. AAS/CERN.
11. System requirements including architecture specification. Cosylab/CERN.
12. FP7 paperwork. CERN.
13. Network simulation. AAS.

IN2P3 seems to already have somehow a working slave for the scenario of fine delay compensation, although the details are unclear due to the missing specification and the fact that no documentation is available.

TOP 8 Timing Theory and Technologies

Presentation by Javier Serrano: Timing system objectives, phase deviation, phase noise, PLLs.

Presentation by Pablo Alvarez: Interesting summary on how to measure time intervals and phase shifts. Several hints to related work on FPGA based measurement using delay lines.

TOP 9 Visit of the Labs and Individual Discussions

- Demonstration of the fibre delay compensation system by Tomasz
  The system works automatically with different fibre lengths and is able to synchronize the PPS absolute to ±400ps. There are still some stability issues and problems with the integrated GTPs of Xilinx. The design itself is of questionable code quality, but well thought engineers works.
- Phase noise measurements with CERNs instruments