Fairness and Protection Behavior of Resilient Packet Ring Nodes Using Network Processors

Andreas Kirstädter[1], Axel Hof[1], Walter Meyer[2], Erwin Wolf[2]

[1]Siemens AG, Corporate Technology
Information and Communications
Otto-Hahn-Ring 6
81730 München, Germany
{Andreas.Kirstaedter, Axel.Hof}@siemens.com
[2]Siemens AG, ICN
Hofmannstr. 51
81379 München, Germany
{Walter.Meyer, Erwin.Wolf}@siemens.com

Abstract. *The Resilient Packet Ring IEEE 802.17 is an evolving standard for the construction of Local and Metropolitan Area Networks. The RPR protocol scales to the demands of future packet networks and includes sophisticated resilience mechanisms that allow the reduction of equipment costs. Network processors are a new opportunity for the implementation of network nodes offering a high flexibility and a reduced time to markets. This paper describes the implementation of a Resilient Packet Ring line card for a SDH/Sonet add-drop-multiplexer using the Motorola C-5 network processor. We show the novel system architecture of the ring node influenced by the use of the network processor. System simulations and field-trial measurements verify the performance of the implemented protection and fairness mechanisms. Even with the usage of a protection steering mechanism implemented on flexible network processor hardware we were able to achieve reconfiguration times well below 50 milliseconds.*

1. Introduction

The Resilient Packet Ring (RPR) is a draft standard to transport data traffic over ring-based media with link data rates scalable up to many gigabits per second in Local or Metropolitan Area Networks. The RPR standardization (IEEE 802.17) working group of the Institute of Electrical and Electronic Engineers (IEEE) started to work on the specification in December 2000 with the intention to create a new Media Access Control layer for RPR.

Two counter-rotating buffer-insertion rings build up an RPR [1,2], as shown in Figure 1. Adjacent nodes are interconnected via a (fiber) link pair. The link bit-rate of an RPR can take values in the range from 155 Mbit/s up to 10 Gbit/s [2].

Among many other deployment areas, RPR rings are especially attractive for the use within SDH/Sonet Add-Drop Multiplexers in Metropolitan Area networks. Here SDH/Sonet paths constitute the links between the RPR nodes.

The RPR line card described in this paper offers on the tributary-interface side the choice between 10/100 Mbps and 1 Gbps Ethernet. On the (SDH) ring side, either VC-4 paths or VC-4-4v paths can be supported. To achieve this flexibility a network processor (NP) was selected for the task of data processing [3]. The network processor C-5 from C-Port/Motorola proved to be appropriate for this kind of application [4].

In principle, several solutions exist for protecting RPR ring networks. These solutions differ in their protection speed and bandwidth efficiency. In our implementation we selected a steering mechanism for the ring protection and implemented it in the software of the NP and its controlling General Purpose Processor (GPP).

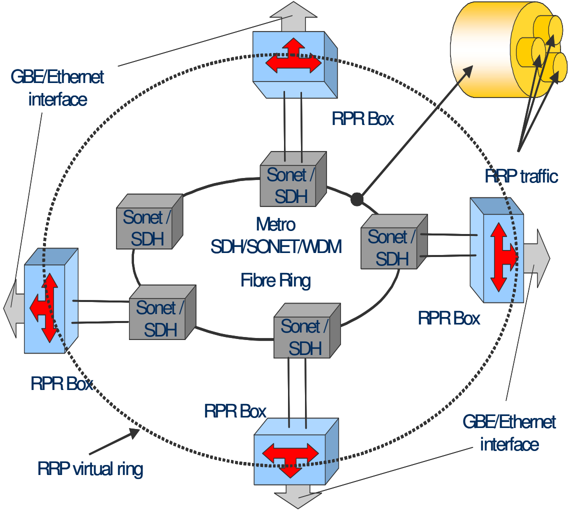

Fig. 1. RPR topology on the basis of SDH links

The rest of the paper is organized as follows: Chapter 2 describes the architecture of the RPR card and the SDH add-drop multiplexer it is connected to. Chapter 3 gives a small overview of the C-5 network processor and shows the main features of this processor. Ring protection and the steering mechanism are presented in chapter 4 together with the measurement results from a field trial with a lead customer. Additionally, we carried out some system simulations of the RPR ring. Chapter 5 describes the simulator and presents various results of system measurements and simulation.

2. RPR Card and SDH Add-Drop Multiplexer

The SDH/Sonet add-drop multiplexer is a multi-service system that is configured in a rack with multiple flavors of line cards. A SDH/Sonet back plane provides the interworking among the cards across a switch fabric. A control processor card manages the operation of the system. The line cards run with OC-3, OC-12 and OC-48.

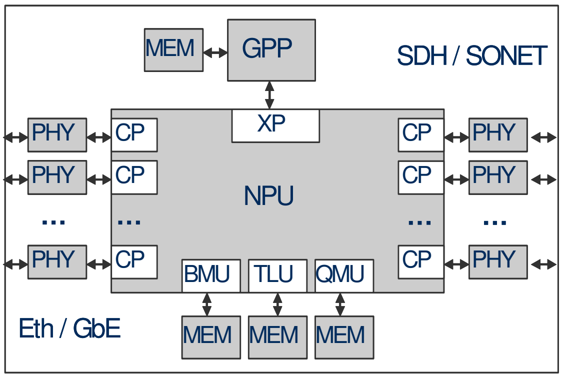

Fig. 2. Architecture of the RPR line card

As shown in Figure 2 the RPR card mainly consists of the NP, a GPP, and some interface and memory chips. The GPP controls the network processor and consists of a Power-PC processor connected via a PCI bridge to the network processor. The GPP takes care about the generation of the routing table, the alarm handling and bandwidth reservation. Via the PCI bridge the GPP can access a part of the data memory of the NP, e.g. for downloading routing tables to the NP or reading of some statistical data. A control processor core (XP) within the NP handles the access of the GPP to the data memory on the NP chip.

The RPR implementation supports stream and best effort traffic. A fairness algorithm on top of the ring guaranties the reserved bandwidth for the stream traffic and distributes the remaining bandwidth between the ring nodes in a fair manner for the best effort traffic. An input rate control at the tributary Ethernet interfaces regulates the throughput of added packets on the ring. A feedback mechanism of the fairness algorithm influences the settings of the input rate control and can back press the Ethernet packets in case of ring congestion.

3. Network Processor Architecture

The C-5 network processor from Motorola contains 16 parallel channel processors (CP). Each of them consists of a RISC core together with a Serial Data Processor (SDP) for the bit and byte processing [5]. Additionally to the XP block mentioned above, there are also four other special-purpose units on the C-5 for the buffering (BMU), queuing (QMU), table lookups (TLU), interconnection to a switch fabric (FP), as shown in Figure 2.

The 16 parallel channel processors (CP) are ordered into four clusters of four processors each. The four processors in one cluster can run the same application and share an instruction memory of 24 kByte that also can be subdivided so that each CP gets a dedicated 6kByte sub-array. Three independent data buses (Figure 3) provide internal communication paths between the different internal processors

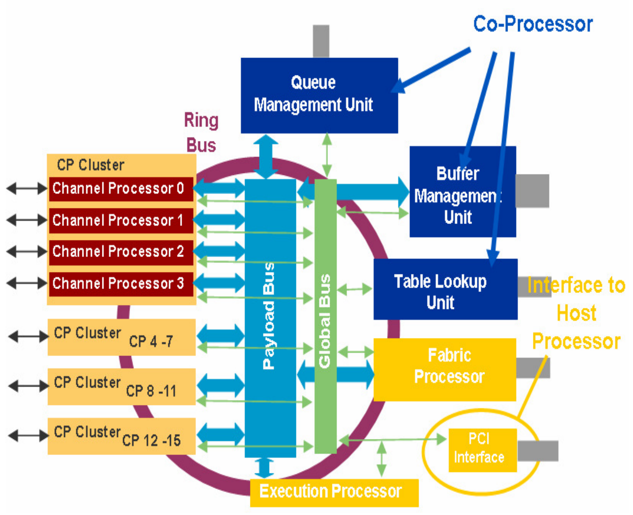

Fig. 3. C-5 network-processor architecture

Each of the sixteen CPs contains a Reduced Instruction Set Computer (RISC) Core controlling cell and packet processing in its channel via the execution of a MIPS ^{TM} 1 instruction set (excluding multiply, divide, floating point).

Packet buffering and queuing in the C-5 is handled as follows: The payload of the incoming packet is stored in the external memory, which is controlled by the Buffer Management Unit (BMU). The BMU controls the storage of the payload and returns a descriptor of the memory block for the payload storage to the CP. After the lookup at the Table Lookup Unit (TLU) the CP sends the descriptor of the payload buffer to the Queue Management Unit (QMU) and enqueues it into the queue of the transmitting CP.

For the programming and configuration of the special entities like the BMU, QMU and TLU exists a library of service functions [6], which is part of the C-Ware Software Toolset (CST). The XP, being the only processor with no linkage to the data path, controls the operation of the other processors and downloads the configuration onto them and the special units. During runtime the XP generates control messages or table entries within the TLU.

4. Ring Protection measurement and field trial

As mentioned in the introduction several alternatives exist for the protection of RPR rings. A pure protection on the SDH level - below the RPR protocol - is surely the fastest way but occupies a lot of protection bandwidth and does not cover failures of the packet node or on the Ethernet level.

The IEEE 802.17 protocol itself will support wrapping and steering for ring protection, which allows the spatial re-use of bandwidth.

The faster alternative is wrapping being less bandwidth effective due to the wrapping loops. Wrapping occurs locally and requires two nodes to perform protection switching. As shown in Figure 4 the two nodes neighbouring the failed span have to loop the traffic onto the other ring. The dashed line is the original traffic flow, whereas the solid line symbolizes the protection path. Fast wrapping generates the lowest packet loss on the cost of higher bandwidth consumption.

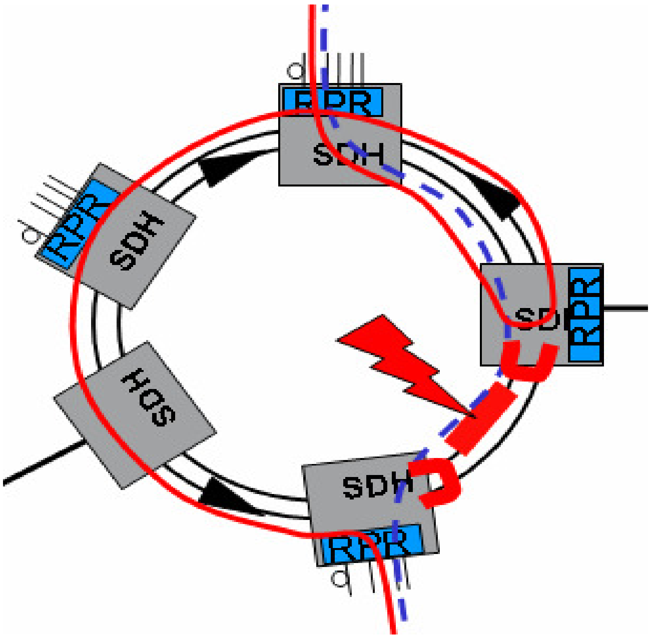

Fig. 4. Wrapping for ring protection

Steering reacts to the failure by modifying the routing tables in all nodes. Therefore it is more bandwidth efficient but also slower due to the messaging and the generation of the tables. In the RPR layer the two neighboring nodes would signal to other nodes span-status changes via control messages carried on opposite ring. Instead of wrapping the ring each node then independently reroutes the traffic it is sourcing onto the ring using the updated topology (see Figure 5).

The RPR foresees a recovery time of 50 milliseconds in event of fibre/node failure on the ring. Steering will be the lowest common denominator when both steering and wrapping nodes are on the ring.

For the wrapping mechanism the outgoing node has to store a high number of packets in case of a switch back to the original link after the failure recovery to avoid packet disorder. The original link is shorter then the protection link and therefore the transmission from the incoming node to the outgoing node includes more hops. The total number of stored packets in a C-5 NP is limited to 16000. The number of packet descriptors is needed for the proper operation of the fairness algorithm. Due to the limited number of packet descriptors we selected a steering mechanism and implemented it in the GPP and NP software.

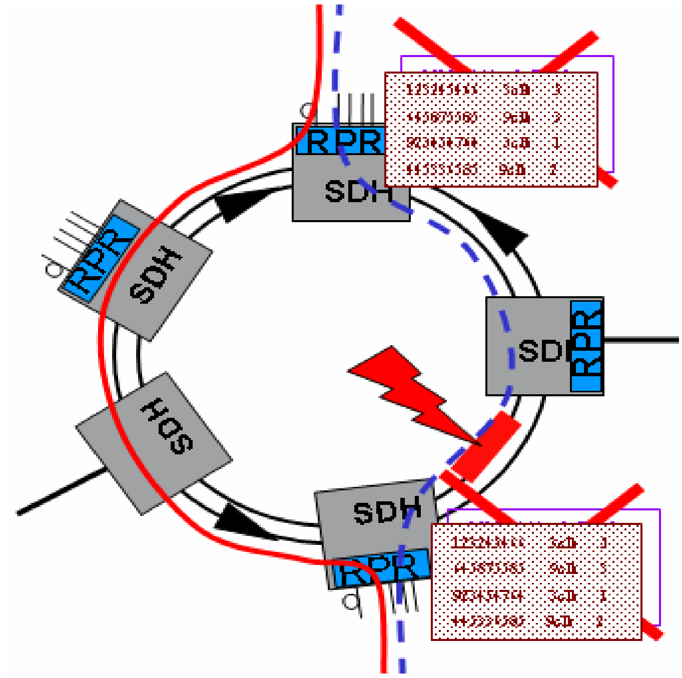

Fig. 5. Steering for ring protection

Link failures are detected by SDH alarming in the SDH overhead or frame. The physical layer device (PHY in Fig. 2) for SDH analyzes the SDH overhead and the SDH framer within the C-5 NP controls the frame errors. The GPP collects all failure alarms and generates an alarm message. The nodes neighboring the failure inform all other nodes via alarm messaging on the RPR level. To reduce the transmission time of the alarm messages, the control packets got the highest priority (above stream traffic).

In the GPP of each single node the routing tables are recalculated according the collected routing information in the received alarm message and then downloaded into the NP via the PCI Bridge. Due to the limited instruction memory in the XP in the C-5 NP the routing generation is part of the GPP software. In [7] the authors describe a solution for the speed up of the interconnection between a NP and GPP, which will lead to a shorter rerouting time. For the C-5e a faster rerouting would be feasible by the inclusion of the routing table generation code into the XP, since the C-5e has twice the instruction memory capacity.

System measurements with different SDH failures verified ring protection times well within 50 milliseconds. Table 1 presents the results for the failure insertion and the failure removal on a 1200 km ring with 12 nodes. In all cases of failures removal the protection switching time stays below 20 milliseconds. All error detections have no integration time to keep the delay as low as possible.

Table 1. L2 protection times for different SDH faults

Failure	L2 Protection Switching time (msecs.)	
	Failure Insertion	Failure Removal
LOS	44	15
AU4-AIS	44	15
UNEQ	20	15
LOM	20	15

The Loss of Signal (LOS) alarm is raised when the synchronous signal (STM-N) level drops below the threshold at which a BER of 1 in 10^3 is predicted. This could be due to a cable cut, excessive attenuation of the signal, or equipment fault. The LOS state will be cleared as soon as two consecutive framing patterns are received and no new LOS condition is detected.

The Alarm Indication Signal (AIS) for STS-3c (AU4) is an all-ONES characteristic or adapted information signal. It is generated to replace the normal traffic signal when it contains a defect condition in order to prevent consequential downstream failures being declared or alarms being raised.

The Loss of Multi-frame (LOM) state occurs when the incorrect H4 values for 8 frames indicate lost alignment.

The unequipped (UNEQ) alarm is raised when z consecutive frames contain the all-ZEROS activation pattern in the unequipped overhead.

After the system integration followed a field trial at a lead customer side. Delay measurements in a 12 nodes ring in Austria (Vienna, Salzburg, Klagenfurt) make up the main part of the trial. For the delay and packet loss measurements we used frame-sizes according to RFC 2544. We observed no packet loss and delays between 6.33 and 6.98 milliseconds depending on the frame size.

5. Simulations during System Development

The C-5 tool environment includes a cycle-accurate simulator together with a performance analyzer. Via this tool we made a first rough estimation of the workload on the network processor and were able to trace internal components of the C-5 like the buses and special units like the QMU.

Additionally to the cycle-accurate simulations the overall system behavior had to be verified. The cycle-accurate simulations deliver a very detailed picture of the internal operation of the C-5 running the RPR protocol. But for exact statements on the protocol behavior itself a separate simulator had to be developed since the system of several RPR nodes had to be observed for larger time intervals. Running these simulations with the cycle-accurate simulator was not possible due to the CPU time requirements: The maximum number of packets that could be observed during reasonable CPU times is around 100 to 1000. This short time frame was not sufficient to check the system behavior of a complete RPR ring concerning fairness and delay behavior.

The system simulator is programmed in C++; the libraries of CNCL (Communication Networks Class Library [8]) were used. The simulator is event based and is built in a very modular manner. The protocol and also the simulator are specially adapted to the behavior of the C-5 NP. To simulate different scenarios it is possible to use different sources with different distributions of packet length and destination addresses. Meters can be attached to points of interest in the investigated network to accomplish the behavior and performance investigation of the protocol by collecting data while the simulation is running.

As an example the following drawings show the priority handling in an eight-node topology where first node #4 at time=0sec sources 100 Mbps of low-priority traffic onto the ring (link capacity: 150 Mbps) for forwarding further downstream towards node #6 (see Figure 6). At the time=0.5sec node #5 sources 100 Mbps of high-priority traffic also destined to node #6.

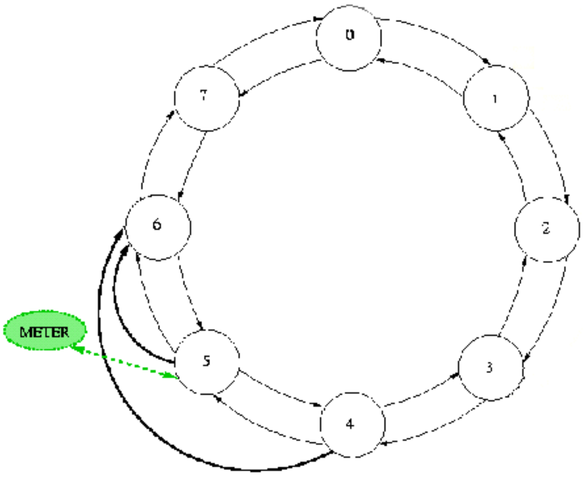

Fig. 6. Eight node network with traffic meter

For the simulations we used different traffic sources. The destination addresses are fix or have a uniform distribution. The packet length is fix or distributed negative exponential or like a so called bath tube (30% 64 bytes, 60% 1540 bytes). The sources are switched on and off after a duty cycle of 0.5sec. In this example the destination address is fix and the source are sending with a constant bit rate.

Figure 7 shows the resulting throughput in the form of forwarded low-priority traffic originated by node #4. Figure 8 shows the amount of high-priority traffic sourced by node #5.

When the high priority traffic is switched on the throughput of the low-priority traffic is reduced to remaining ring capacity of 50 Mbps

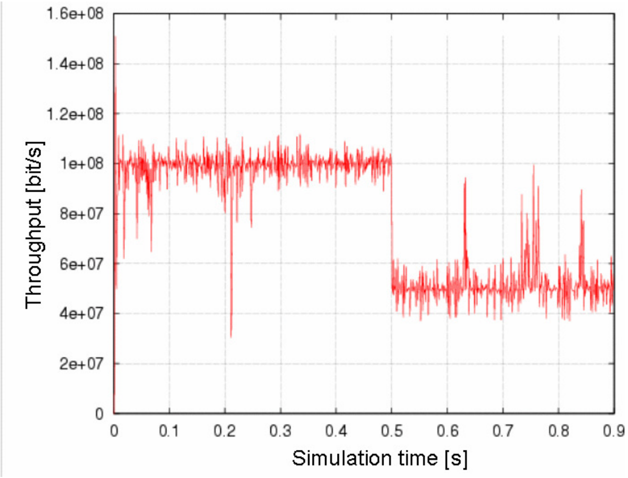

Fig. 7. Throughput from node #4 over time

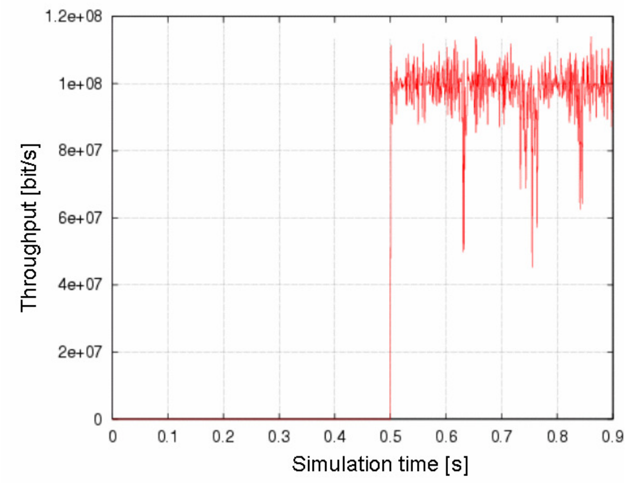

Fig. 8. Throughput from node #5 over time

As it can be seen from the diagrams above the ring fairness protocol preserves the strict priority between high and low-priority traffic. Exactly the same result was also measured in the experimental setup.

We repeated comparable simulations with different traffic sources and different ring topologies. During the system test we verified the fair distribution of the ring capacity with similar test scenarios.

In conjunction with the steering mechanism the fairness algorithm guarantees a proper behaviour also in case of ring protection. The additional sourced packets in case of ring protection is controlled and the input of low-priority traffic reduced to keep the guaranties of the high-priority traffic.

6. Conclusion

During the system development phase it was very helpful to have both the cycle-accurate and the system-level simulators at hand. Especially the system-level simulations delivered important details on the operation and optimization of protocol features that otherwise could not be verified in advance.

The system tests verified the fast protection switching with the usage of SDH alarms despite the steering mechanism. The generation of the rerouted tables in the GPP and the PCI transfer into the NP showed to be sufficiently fast without the need any additional special hardware. Additionally the fairness algorithm guaranties the appropriate subdivision of the link bandwidth between the single flows even in protection state.

The field trial provided us with long-time measurements and asserted the smooth system behavior of the RPR line card. Since some months the system is delivered to customers.

7. References

[1] IEEE 802.17 Resilient Packet Ring Working Group Website, http://www.ieee802.org/rprsg/.

[2] H.R. van As, “Overview of the Evolving Standard IEEE 802.17 Resilient Packet Ring,” 7th European Conference on Networks & Optical Communications (NOC), Darmstadt, Germany, June 18-21, 2002.

[3] T. Wolf, “Design of an Instruction Set for Modular Network Processors,” IBM Research Report, RC 21865, October 27, 2000

[4] N. Shah, “Understanding Network Processors,” Master’s Thesis, Dept of Electrical Engineering and Computer Science, Univ. of California, Berkeley, 2001

[5] C-5e Network Processor Architecture Guide Silicon Revision A0, Motorola, http://e-www.motorola.com/brdata/PDFDB/docs/C-5EC3EARCH-RM.pdf

[6] C-5e Application Documentation, Motorola, http://e-www.motorola.com/webapp/sps/site/prod_summary.jsp?code=C-5E#applications

[7] F.T. Hady, T. Bock, “Platform Level Support for High Throughput Edge Applications: The Twin Cities Prototype,” IEEE Network Magazin, July/August 2003

[8] RWTH Aachen, http://www.comnets.rwth-aachen.de/doc/cncl/index.html