

Active Probing of an E1 TDM Stream - A Measurement Facility -

TIK Report 217

Rainer Baumann, Ulrich Fiedler

Computer Engineering and Networks Laboratory
Swiss Federal Institute of Technology
ETH-Zentrum, Gloriastrasse 35
CH-8092 Zurich, Switzerland
{baumann,fiedler}@tik.ee.ethz.ch

11th August 2005

Abstract

Emerging Metro Ethernets create new opportunities to converge data and telephony services. However, to connect legacy telephony equipment, networks have to meet customary QoS requirements even at the presence of bursty data cross traffic. We show with active probing measurements that without appropriate mechanisms the QoS requirements cannot be met. Thus, mechanisms for prioritization are needed.

1 Introduction

Recent developments create new opportunities to converge data and telephony services. The emerging deployment of fiber fosters the proliferation of Metropolitan Gigabit Ethernets. Thus a major aspect of interest is to employ these networks to connect legacy private branch exchanges (PBX) and GSM base stations to the core telephony network. This protects investments in existing infrastructure and creates new revenues for network providers. However, to implement this convergence, these networks have to be configured in a way that customary QoS requirements for telephony, which are significantly more stringent than the ones for VoIP, can be met. This is a crucial problem given that the data cross traffic in these networks is bursty. This burstiness can cause excessive queuing delay and frame losses due to buffer overflow [13]. Moreover, the burstiness of data traffic in these networks is known to be self-similar. This means that there is very little smooth out when aggregating the traffic over time which in turn means that buffering has little effect in mitigating the burstiness. It was shown in simulation and theory that it is not reasonable to expect that TDM E1 telephony traffic, so called circuit emulation traffic, can be run on Metropolitan Gigabit Ethernets without some kind of prioritization [6].

For verifying these conclusions in real world networks we build a high accurate measurement facility and carry out dedicated measurements in a Metropolitan Gigabit Ethernet. The traces from these measurements are then evaluated regarding to criteria for circuit emulation traffic specified by the Metro Ethernet Forum (MEF). This evaluation shows violations of requirements. Especially the loss ratio on links with higher utilization is too high.

The rest of this paper is structured the following way. In the following section we present the measurement facility and set-up. In section three we present our results and findings and in the last section we will quickly review them.

2 Methods

In this section we first present a high accurate measurement facility for active probing running on of the shelf PCs with a precision of $< 3\mu s$. Then we point out some mathematical considerations concerning the credibility of obtained data. And finally, the measurement set-up will be explained.

For measuring the performance of communication networks, there are two fundamental approaches: passive and active measurement. By passive measurement we mean the approach of tracking the performance and behavior of packet streams simply by monitoring the traffic passing by the measurement point. Under active measurement we understand the injection of artificial probe traffic into the network, and the measurement of its characteristics at different points, typically back at the origin (round-trip end-to-end measurement), or at some terminating destination (one-way end-to-end).

Passive measurements are usually used to measure metrics pertaining to a certain network element, that is at-a-point metrics such as link throughput and packet size statistics. However from the application point of view, particularly real-time applications, end-to-end quality of service metrics are more important. For these the passive approach is inappropriate as the presence of traffic between the end points is not guaranteed.

The active approach relies on the capability to inject test packets into the network. As such it does create extra traffic but it provides explicit control on the feeding of packets. This includes control on the nature of traffic generation, the sampling techniques, the timing, frequency, scheduling, packet sizes and types (to emulate various applications), statistical quality, the path and function chosen to be monitored. Being active implies testing what you want, when you need it. Emulation of scenarios is easy and checking if Quality of Service (QoS) are met is relatively straightforward. Thus active measurement methods are typically used to obtain end-to-end statistics such as delay, loss and route availability.

Summarizing, workload and routing measurements typically utilize passive measurements while performance and topology measurements rely on active measurement methods. In spite of this general classification, we developed a measurement facility for E1 TDM traffic based on the active approach.

2.1 Active Probing Facility

We developed a high precision active probing facility containing a measurement application based on a hard real-time system running on of the shelf PCs. In [14] was shown, that high accurate active probing is achievable with of the shelf PCs using highly specialized software.

In this section we first describe the underlining hard real-time system, then we have a look at the measurement application and finally we explain the mathematics used for analysis.

2.1.1 Underlining Hard Real-time System

The underlining hard real-time system mainly consists of two components.

1. A hard real-time operating system
2. A framework for hard real-time Ethernet

The hard real-time constraint is absolutely essential for the precision of the active probing facility. A hard real-time system embody guaranteed timing, cannot miss deadlines and must have bounded latencies while a soft real-time system is characterised by its ability to execute a task according to a desired time schedule on the average.

Today, there exists several different hard real-time systems as for example RTLinux [9] and RTAI [2]. For our active probing facility, we decided to use RTAI because it is Linux based, has very good performance and is completely freely available under the GPL2 [1].

On RTAI, the Linux operating system runs as the lowest priority task of a small real-time operating

system. Thus, Linux undergoes no changes to its operation from the standpoint of the user or the Linux kernel, except that it is permitted to execute only when there are no real-time tasks executing. RTAI provides the capability of running special real-time tasks and interrupt handlers that execute whenever needed, regardless of what Linux tasks may be running. The timer granularity of RTAI is limited by the frequency of the PCI 8254 which runs at $1193180Hz$ what gives a resolution of $0,838\mu s$. For communication between real-time and Linux tasks a device interface (RT-FIFO) or shared memory can be used (Figure 1).

RTAI is implemented in the following way. It inserts a hardware abstraction layer (HAL) under the standard Linux kernel. The HAL comprises of a structure of pointers to the interrupt vectors, and the interrupt enable/disable functions. Real-time tasks are loaded and executed as Linux kernel modules.

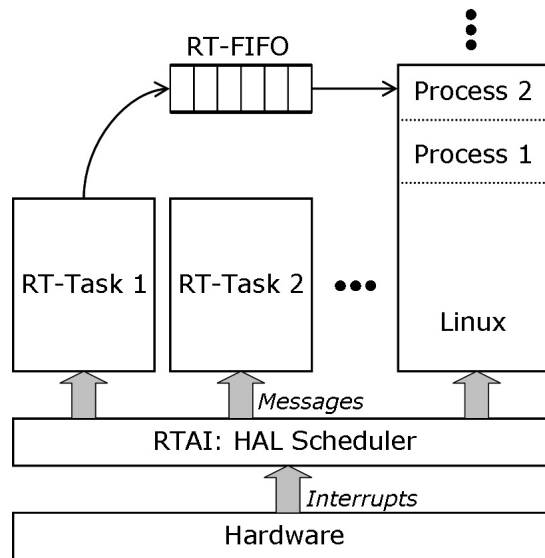


Figure 1: *Structure of the RT-System*

As framework for hard real-time Ethernet, we use RTnet (Hard Real-Time Networking for Linux/RTAI) [3]. RTnet provides features for protocols and tools for configuration, operation and diagnosis of Ethernet networks. It offers hard real-time via slightly modified Linux drivers which allows for atomic transmission time stamps based on the TSC processor register [14]. Concerning hardware, RTAI and RTnet support x86 and PowerPC platforms with a large amount of established Ethernet controllers.

2.1.2 Measurement Application

Based on the RTnet stack, we developed a measurement application for active probing (Figure 2). It transmits and archives packets with high accurate time stamps between two PCs. Since on full duplex Ethernet links, the load of each direction is independent, one has to measure each direction independently. Therefore it is no option to perform measurements based on a packet round trip between the two PCs.

The measurement application mainly consists of three parts.

- A Generator on the source PC
- A Sink on the destination PC
- A Archivist on the destination PC

The generator is implemented as a real-time task and transmits UDP segments with a configurable ratio and amount of payload. For measurement purpose, each packet carries a unique sequence number and a sender time stamp with its transmission time taken from the TSC processor register [14].

On the receiver PC, there are two parts: the sink and the archivist. The sink is implemented as a real-time task and the archivist as common Linux task. For each packet, before destruction, the sink hands over some information to the archivist using an RT-FIFO. This information contains the sender time stamps and the sequence number from the packet as well as a receiver time stamp taken from the TSC processor register. The archivist is then responsible for writing the collected data to disk.

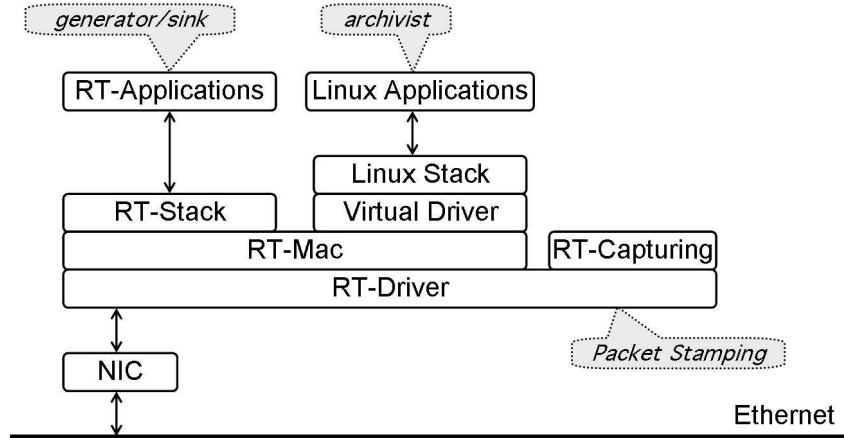


Figure 2: *RTNet and Application Stack*

2.1.3 Mathematical Analysis

In this section, we focus on the mathematical transformations of the collected data to meaningful network traces.

As described above the collected data contains the following per packet data.

- Sequence number
- Sender time stamp
- Receiver time stamp

The packet sequence number allows to detect drops and multifies. Based on the sender and receiver time stamps, the relevant delay variation (difference of delay) can be calculated the following way. For time periods e.g. of 5s, the offset and skew of the clocks form the two PCs are eliminated using the simple skew model [17]. We argue that it is enough to focus on the delay variation instead of the delay itself for measuring characteristics of high speed Ethernets. Since a delay series can be decomposed into delay variation and the minimal delay [15]. The minimal delay is constant and consists of the processing delay introduced by every hop and the transmission duration. In todays high speed Ethernets, the typical minimal processing delay of high-end switches and routers is highly constant and lies in the range of tens of microseconds [16] as well as the propagation and the transmission delay of gigabit Ethernet.

So this is no issue for reaching QoS requirements for VoIP [4, 5, 7] or circuit emulation [8].

2.1.4 Hardware and Performance of the Measurement Facility

As underlying hardware for our active probing measurement facility, we used IntelPentium III machines with 1GHz, 512MB ram and 2 3COM 3C595 Fast EtherLink network interface cards (one for remote control and the other as real-time interface).

For measuring the accuracy of our measurement facility, we directly connected a sender with a receiver over a crossover cable for 12 hours. According to an E1 TDM signal, the generator sent 8000packets/s with a payload of 32Bytes .

The resulting 3σ confidence interval for delay variation is $[0.00; 2.83]$ us. (Table 1)

Samples	$3,3 \cdot 10^8$
Avg. delay variation	1,30us
σ delay variation	0,51us
Max. delay variation	13,55us

Table 1: Accuracy of Measurement Infrastructure

2.2 Measurement Set-up

For getting a high statistical quality of measurement data, we performed active probing of an E1 stream over several weeks over various links in an exemplary Metropolitan Gigabit Ethernet. Corresponding to the E1 specifications [11], the generator was configured to transmits 8000 packets per second containing a payload of 32 Bytes.

As an exemplary Gigabit Metropolitan Ethernet, the backbone of ETH network was at our disposal. The ETH network spreads the whole city of Zurich for connecting its institutes, located all over the city. The architecture of the backbone follows the CISCO recommendations [10] using a three level hierarchy and a dual core (Figure 3). The access layer switches are Catalyst 4000 while the distribution and core layer switches are Catalyst 6500.

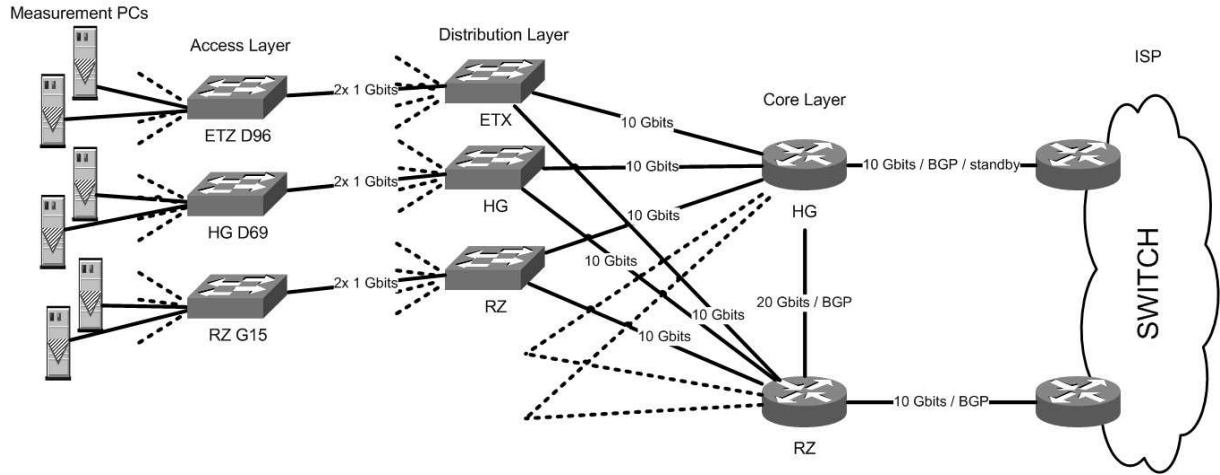


Figure 3: Architecture of measured network (ETH Zurich)

In common networks the average utilization rates (AUR) is about 1% [12]. For getting a diversified picture we have chosen three different locations with various AURs (HG: similar AUR as in common networks. ETX: lowest AUR in network. RZ: highest AUR in network) (Table 2). A measurement serie was carried out for each pair and direction of them.

Location	ADLL Utilization [%]	DCLL Utilization [%]
	<i>up/down</i>	<i>up/down</i>
ETX	0,3/0,5	0,2/0,5
HG	0,2/0,2	0,5/1,5
RZ	0,4/1,0	3,1/5,1

Table 2: Average Utilization

(ADLL: Access-Distribution Layer Link / DCLL: Distribution-Core Layer Link)

3 Results

For analyzing our measurement traces, we use the criteria specified by the Metro Ethernet Forum. The measurement results are presented in relation to these criteria and show that some requirements are violated.

For circuit emulation traffic, the Metro Ethernet Forum specifies strict criteria concerning quality of service [8]. This criteria include jitter, delay and losses as well as the summation of them (equation 1). The complete list of requirements is specified in table 3.

$$FER = JVR + DVR + LR \quad (1)$$

Abbreviations: JVR: Jitter Violation Ratio / DVR: Delay Violation Ratio /
LR: Loss Ratio / FER: Frame Error Ratio

Metric	QoS Requirement
Jitter	10ms
Delay	25ms
Losses	$< 8.75 \cdot 10^{-7}$
FER	$< 8.75 \cdot 10^{-7}$

Table 3: Circuit emulation QoS Requirements

As described earlier in section 2.2, we conduct measurements over six different connections for about 10^{11} packets. The recorded traces from these measurements are analyzed regarding to the requirements specified above. The results from this analysis are summarized in figure 4. It stands out that the frame error ratio is violated over the two connections to the RZ which has the highest load in the network. This violation is mainly due to the high packet loss ratio. These losses can be explained by heavy data burst accumulating on the DCLL to the computer center.

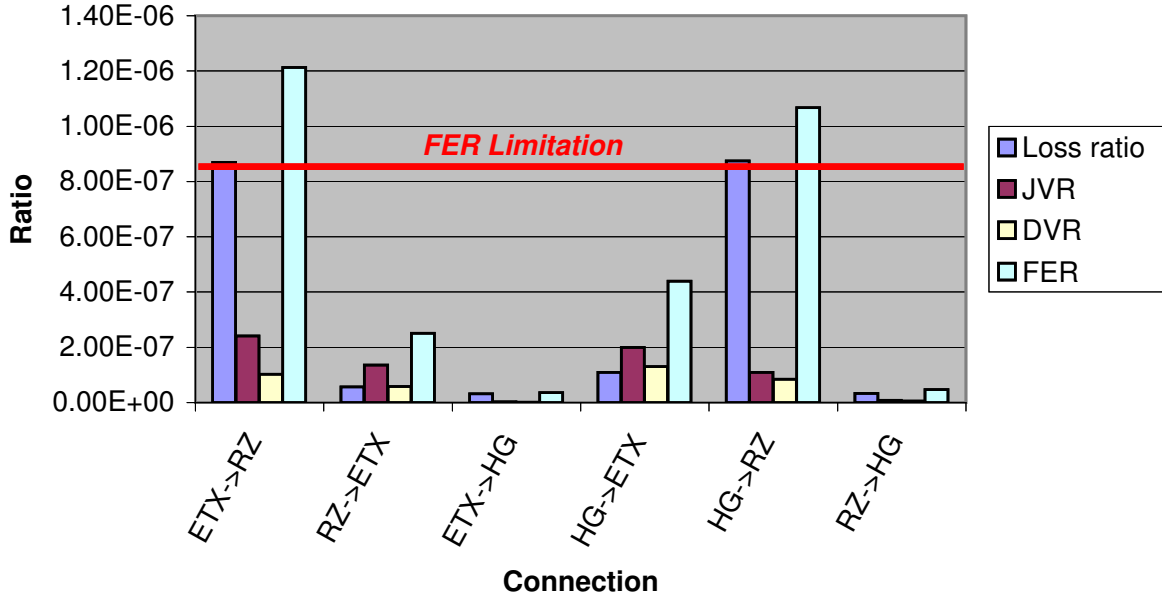


Figure 4: Summarized measurement results

For the complete results refer to Appendix A.

4 Conclusions

In this report, we describe a high accurate active probing facility which we use for measuring characteristics of circuit emulation traffic in a exemplary Metropolitan Gigabit Ethernet. These measurements show that the loss ratio exceeds the allowed boundary. This finding correspond to our conclusion in [6] which investigated the same issue in theory and simulation. Concluding we can state that a mechanisms for prioritization is needed for circuit emulation.

5 Acknowledgments

We would like to specially thank Derk Valenkamp and his team from the Informatikdienste ETH Zurich for getting us access to the network and their great support. Additional acknowledge goes to our head of the research group Prof. Dr. Bernhard Plattner, our industry partner Siemens Schweiz and KTI/CTI the Innovation Promotion Agency of the Swiss Confederation for fending this project.

References

- [1] GNU Lesser General Public License 2.
- [2] RTAI, Real-Time Application Interface.
- [3] RTnet, Hard Real-Time Networking for Linux/RTAI.
- [4] Control of Talker Echo. ITU-T Recommendation G.131, August 1996.
- [5] One-Way Transmission Time. ITU-T Recommendation G.114, February 1996.
- [6] Rainer Baumann and Ulrich Fiedler. Why QoS Will Be Needed in Metro Ethernet. In *Lecture Notes in Computer Science*, volume 3552, pages 379–381. January 2005.
- [7] G.H. Petit J. Janssen D. de Vleeschauwer, F. Poppe. Quality bounds for packetized voice transport. Alcatel Telecommunication Review, January 2000.
- [8] METRO Ethernet Forum. MEF 3, Circuit Emulation Service Definitions, Framework and Requirements in Metro Ethernet Networks. Technical Specification MEF, April 2004.
- [9] FSMLabs. RTLinux, Real-Time Linux.
- [10] CISCO Systems Inc. Hierarchical campus design-at-a-glance, 2004.
- [11] International Telecommunication Union. ITU-T G.X.
- [12] ANDREW ODLYZKO. The internet and other networks: Utilization rates and their implications. AT&T Labs, September 1998.
- [13] K. Park and W. Willinger. Self-similar network traffic: An Overview. In *Self-Similar Network Traffic and Performance Evaluation*, chapter 1. Wiley-Interscience, NY, 2000.
- [14] A. Pasztor and D. Veitch. A precision infrastructure for active probing. In *Proceedings of the PAM*.
- [15] A. Pasztor and D. Veitch. The scope of end-to-end probing methods, 2002.
- [16] Rolf McClellan. Cisco Catalyst 6500 independent performance evaluation, 1999.
- [17] Darryl Veitch, Satish Babu, and Attila Pasztor. Robust synchronization of software clocks across the internet. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 219–232, 2004. ISBN 1-58113-821-0.

A Detailed Results

From	To	Samples	Loss ratio	DVR	JVR	PRR	FER
ETX	RZ	$1,04 \cdot 10^{10}$	$8.69 \cdot 10^{-7}$	$2.41 \cdot 10^{-7}$	$1.03 \cdot 10^{-7}$	$5.47 \cdot 10^{-6}$	$1.11 \cdot 10^{-6}$
RZ	ETX	$1,04 \cdot 10^{10}$	$5.63 \cdot 10^{-8}$	$1.36 \cdot 10^{-7}$	$5.80 \cdot 10^{-8}$	$4.19 \cdot 10^{-7}$	$1.92 \cdot 10^{-7}$
ETX	HG	$2,03 \cdot 10^{10}$	$3.18 \cdot 10^{-8}$	$2.57 \cdot 10^{-9}$	$1.55 \cdot 10^{-9}$	$2.14 \cdot 10^{-7}$	$3.43 \cdot 10^{-8}$
HG	ETX	$2,03 \cdot 10^{10}$	$1.09 \cdot 10^{-7}$	$2.00 \cdot 10^{-7}$	$1.30 \cdot 10^{-7}$	$5.92 \cdot 10^{-7}$	$5.03 \cdot 10^{-7}$
HG	RZ	$1,19 \cdot 10^{10}$	$8.74 \cdot 10^{-7}$	$1.10 \cdot 10^{-7}$	$8.40 \cdot 10^{-8}$	$1.39 \cdot 10^{-5}$	$9.84 \cdot 10^{-7}$
RZ	HG	$1,19 \cdot 10^{10}$	$3.37 \cdot 10^{-8}$	$7.66 \cdot 10^{-9}$	$5.42 \cdot 10^{-9}$	$1.01 \cdot 10^{-7}$	$4.13 \cdot 10^{-8}$

Table 4: Detailed Measurement Results (JVR: Jitter Violation Ratio / DVR: Delay Violation Ratio / PRR: Packet Reorder Ratio / FER: Frame Error Ratio)

B IP Configuration

Domain	IP	Interface	Set-up 1	Set-up 2	Set-up 3
RZ	192.168.106.2	S	cop_s2	cop_s2	
RZ	192.168.106.3	S-RT	cop_s2	cop_s2	
RZ	192.168.106.4	R	cop_r2	cop_r2	
RZ	192.168.106.5	R-RT	cop_r2	cop_r2	
HG	192.168.106.18	S		cop_s1	cop_s1
HG	192.168.106.19	S-RT		cop_s1	cop_s1
HG	192.168.106.20	R		cop_r1	cop_r1
HG	192.168.106.21	R-RT		cop_r1	cop_r1
ETX	192.168.106.34	S	cop_s1		cop_s2
ETX	192.168.106.35	S-RT	cop_s1		cop_s2
ETX	192.168.106.36	R	cop_r1		cop_r2
ETX	192.168.106.37	R-RT	cop_r1		cop_r2

Table 5: IP Configuration for measurement