I^{2}MP — An architecture for hardware supported high-precision traffic measurement*

Detlef Sass and Sascha Junghans

University of Stuttgart
Institute of Communication Networks and Computer Engineering
{sass,junghans}@ikr.uni-stuttgart.de

Abstract. Accurate measurement data are crucial for correct understanding of traffic. Ever increasing bandwidth pushes the requirements for measurement equipment beyond pure SW/PC solutions. We developed a scalable architecture of a hardware-supported, passive measurement platform, the I^{2}MP, for high-precision traffic measurement for Gigabit Ethernet with flexible filter possibilities and intelligent selection of the cutting area within a frame. In this paper, this platform will be presented.

1 Introduction

The increasing predominance of the Internet protocol (IP) and emerging new services and applications change network traffic significantly. Also the users' behaviour change drastically due to the increasing availability of broadband access rates and a mostly time independent tariffs. These changes affect not only the access networks but also the metro and core networks. So, up-to-date and correct understanding of the traffic is crucial for traffic engineering and builds a basis for a more precise traffic characterisation.

One method for achieving this, is passive traffic measurement in many relevant network locations. However, a key issue of measurement is accuracy and reliability [1].

In particular, reliable and high-precision accuracy of time is hardly realisable with pure, off-the-shelf PC solutions due to the unpredictable time behaviour of the system at time constraints of current LAN/WAN link speeds and the resulting I/O requirements. For example, the minimal inter-arrival time in Gigabit Ethernet is 672 ns, which is beyond the scheduling time-scale for multi-tasking operating systems, like Linux. Specific hardware (HW) components are often used for resolving these limitations [3].

In the following, we will present our own developed, hardware-supported passive measurement platform, the I^{2}MP (IKR Internet Measurement Platform), where the performance critical and time critical tasks are realised in hardware, enabling a high-precision and high-performance traffic capturing.

* This work was partly funded within the EU IST Project NOBEL (FP6-506760)

The I^{2}MP is equipped currently with optical or electrical Ethernet interfaces for links speeds up to Gigabit Ethernet (GE). It provides a number of flexible filtering possibilities on arbitrary areas within a frame with individual definitions of how much data is captured for each filter.

2 Architecture of the I^{2}MP

In the following, we will give a description of the overall architecture, the functional aspects of the main units with its principal functions and the realisation approaches for the core components.

2.1 General Description

The overall architecture of the I^{2}MP is depicted in Fig. 1. The platform consists of two parts, the HW-unit and the PC-unit. The HW-unit is equipped with two GE network interfaces, one for capturing traffic and the second for communicating with the PC-unit for control and configuration tasks and the actual transmission of the captured measurement data. The measurement data consists of the captured frame areas with its meta data, called data record each, which are collected and assembled to data containers before transfered from the HW-unit to PC-unit.

The HW-unit is developed on the universal hardware platform (UHP) of the IKR [4] based on Field Programmable Gate Arrays. It carries the modular network interfaces on daughter boards, currently up to GE. On the HW-unit, the performance critical and time critical functions are realised. These are the time stamping, filtering and cutting of different frame areas according the filter, assembly of the data records and its transmission.

The main function of the PC-unit is the reliable processing of the received data containers, comprising the reception of data containers, its disassembly and storage on hard disks in an appropriate format. Apart from this, the PC-unit controls and configures the HW-unit.

This architecture separates the tasks with specific requirements to certain units, communicating over standard interfaces, and allows a flexible way for optimizing each unit. It also enables a scalable system where multiple HW-units or PC-units can be used for increasing the number of actively measured links at

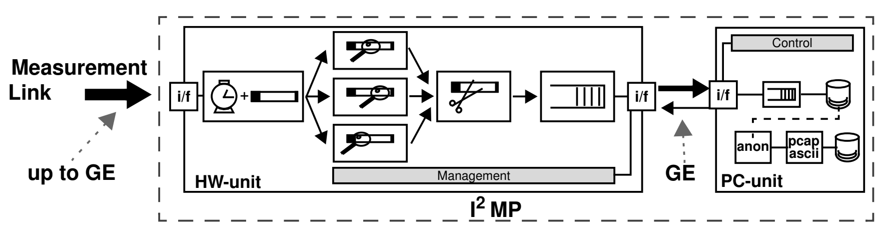

Fig. 1. Architecture of I^{2}MP

the same time or to increase the processing/storage capacity, respectively. The concept of transmitting the collected data to the PC not individually but assembled as data containers decrease significantly the critical burden of interrupt handling and context-switching on the PC side to a manageable level.

2.2 Realisation and Properties of Core Components

The time stamping mechanism in the HW-unit is the most critical component which determines the accuracy of the entire system. In the I^2MP, the time is represented by a 64-bit counter, incrementing at a frequency of 125 MHz. This results in a time resolution of 8 ns which corresponds with the byte-time of GE. Apart from an initial time offset, a finer time resolution would make no difference because in GE, the inter-arrival times are even multiples of byte-times, due to byte-alignment mechanisms based on so-called commas on the physical layer.

For filtering, a micro-programmed filter engine is used in the HW-unit. Here, a filter consists of several filter criterions and a cutting section. The latter defines the captured area of the frame, which can be an arbitrary area within the frame or in terms of the different protocol headers (e.g., Ethernet, IP, TCP/UDP,...). These criterions are defined flexibly by the pair (value, mask), where the mask define the relevant bits for matching. The actual pattern matching is performed by CAMs (content addressable memories).

On the PC-unit runs a stripped Linux distribution where only the needed services/applications are active. The function of the PC-unit is differentiated into online tasks, i.e., during a measurement session and offline tasks. The online tasks are mainly the reception, processing and storage of the data containers, which is realised by an in-house, thread-based software.

The offline tasks are processing of the raw trace to fulfill privacy constraints, like anonymisation of IP addresses, layer 4 port numbers and removing user payload, if any captured, and conversion into commonly used trace formats pcap [2] or ASCII.

The mapping of the original IP address to an anonymised IP address can be performed twofold: regarding the method of anonymisation and regarding for which address type. The methods are pure randomly, mappings are done in the order of the appearance of an address in the trace, or in a prefix preserving manner of the address. The latter allows a consistent mapping across multiple measurement sessions and preserves the network address structure. The method, which is to apply for an IP address, can be flexibly defined by a header field (source or destination address) individually or by belonging to a given network, (IP address, IP (sub)netmask).

2.3 Evaluation

The precise time resolution can be seen in Fig. 2, where the histogram of the measured inter-arrival time is plotted for an PC/SW-based solution and for the I^2MP. The frames are generated with identical inter-arrival times (iat) at

Gigabit Ethernet speed. The I^2MP measured correctly the iat of 4 μs, whereas the PC/SW-based solution reported different and large varying iat values.

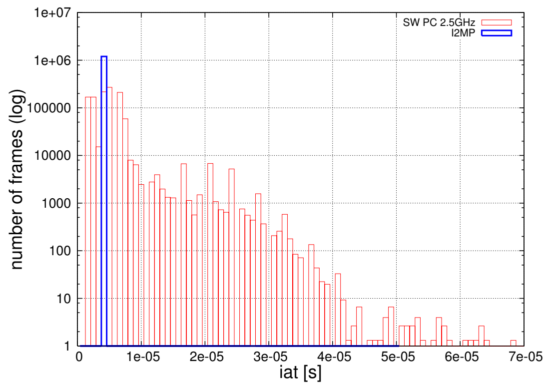

Fig. 2. Histogram of measured inter-arrival times

3 Summary and Outlook

We presented our own developed, scalable architecture of a passive measurement platform, the I^2MP. It is a high-performance, high-precision, hardware-supported platform for Gigabit Ethernet with flexible filter possibilities and intelligent selection of the cutting area within a frame.

This architecture separates the tasks with specific requirements to certain units and enables a scalable system.

Valuable measurement data is already captured by the I^2MP in a large dormitory network at the University of Stuttgart which used for the EU-IST project NOBEL. Subsequent measurement sessions are in preparation.

References

1. J. Cleary, S. Donnelly, I. Graham, A. McGregor, and M. Pearson. Design principles for accurate passive measurement. In *The First Passive and Active Measurement Workshop (PAM)*, pages 1–7, April 2000.
2. Van Jacobson, Craig Leres, and Steven McCanne. *pcap - Packet Capture library*, 2003. http://www.tcpdump.org.
3. J Micheel, S. Donnelly, and I. Graham. Precision timestamping of network packets. In *ACM Sigcomm Internet Measurement Workshop IMW*, Nov. 2001.
4. Institute of Communication Networks and Computer Engineering. The Universal Hardware Platform (UHP), 2005. http://www.ikr.uni-stuttgart.de/UHP.