

Overview of NetFlow Monitoring Adapter

Martin žádník

3.11.2004

1 Abstract

Speed of communication among computers and other related devices grows very fast and computers are not able to process these data. Limiting factor is bus between hardware and software. New approaches must be introduced to deal with it. One of them could be implementation of desired functionality right in the card. COMBO6 card offers flexible solution for many applications, additional changes and features. One of these applications can be NetFlow(network monitoring) which demands a lot of data to be passed to software. Possible better solution is to aggregate data in hardware then send processed data via PCI bus to software.

2 Introduction

One of many applications suitable for hardware acceleration is to measure network traffic. Providing accurate information is basic for planning new networks, guaranteed bandwidth, detecting DoS attacks, billing and accounting.

This document contains main idea of architecture which implements network traffic measurement also known as NetFlow. This architecture is convenient for VHDL and synthesizable in FPGA.

Main blocks are:

- IBUF – Input Buffer,
- TSU – Timestamp Unit,
- HFE – Header Field Extractor,
- HASH – Hash Unit,
- CAM – unit which controls TCAM(ternary content address memory),
- MAN – management unit for CAM and SRAM,
- SRAM – unit which controls SSRAM(synchronous SRAM),
- FIFO – FIFO which stores unified header information before they are passed to SRAM,
- SW_FIFO – short time FIFO with expired records from SRAM. It allows software to read those records via PCI bus.

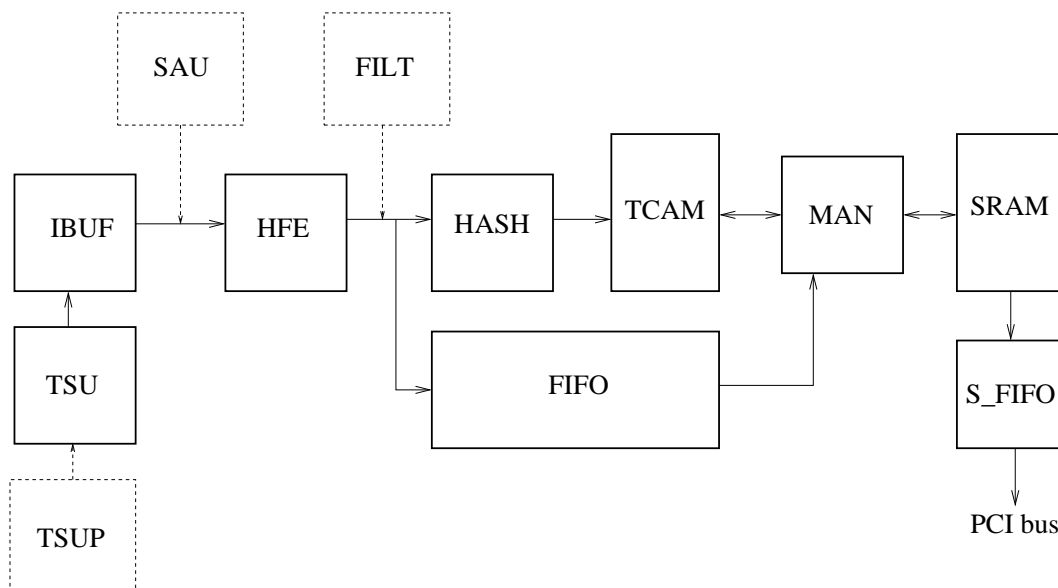


Figure 1: Main idea of architecture

Optional:

- TSUP – Precise Timestamp Unit on COMBO PTM card,
- FILT – Filtering Unit,
- SAU – Sampling Unit.

3 Content of Cards

Several different types of cards have been introduced during period of Libero-outer project.

There is one basic mother card called COMBO6 that has to be always used. It contains one FPGA(Virtex II v1000) surrounded by various chips as TCAM, 3x SSRAM, PLX, SDRAM etc. Its duty is to provide enough computing power for interface card and a good connection with PC via PCI bus.

There are several add-on interface cards for COMBO6 card, which provide flexible connection between various types of network interfaces and mother card. COMBO4-MTX with four copper Gbps interfaces is one of these cards. This card is equipped with FPGA(Virtex II v1000), 2xSSRAM (in future design), TCAM (in future design) etc.

This architecture of cards supports our design for NetFlow monitoring adapter.

4 Block Description

4.1 IBUF

The Input Buffer (IBUF) is used as a storage for incoming packets. Packets are received from GMII interface and only those one with correct CRC together with assigned packet timestamp are saved into the internal IBUF memory.

4.2 TSU

Implemented as 37 bits counter at frequency of 100MHz ($T=10\text{ns}$). For next processing only 32 most significant bits are considered so output of this unit is 32 bit long timestamp with precision of 320 ns ($2^5 \cdot 10\text{ns}$). This precision allows to distinguish two following packets and it is unique for 1300 s ($2^{37} \cdot 10\text{ns}$). This approach requires reading this register to software and to interpolate final values of timestamps with UNIX system time.

Example:

	Value in HW	Interpolation	Value in SW
TSU register	0x0000000A	none – just remember	12:30:00,000
TSU register	0x00008000	none – just remember	12:30:00,021
Start timestamp	0x00000800	$0x0000000A + 0x00000800 = 0x0000080A$	12:30:00,001

Table 1: Example of interpolation

4.3 HFE

The Header Field Extractor is intended for analyzing of input packets. It is a processor based on RISC architecture controlled by specific instruction set. HFE reads data of packet from Input Buffer, analyzes control information in its headers, extracts required fields from IP and TCP/UDP headers and assemble the unique key which designates each flow. This key consists of IP source address, IP destination address, source port, destination port, transport layer protocol, type of service (ToS). After processing each datagram HFE is also able to provide packet information for FIFO.

5 FIFO

The packet FIFO is used as a storage for information about incoming packets from HFE processor. Records are stored in this block until the control path (HASH->TCAM) is passed through. FIFO contains following records for every incoming

packet: number of bytes, timestamp, flags, key (NetFlow packet identification). These records are provided during update or creation of record in SSRAM.

5.1 HASH

Hash block implements a hash function (for example CRC). Input is six main fields of datagram. IP destination and source address, destination and source port, protocol, ToS. There is need to hash these fields because TCAM is configured as 32 768 entries x 68(64) bits wide. Probability that two different flows would map to the same entry is (supposing 200 000 flows in one second $(200000/2^{64})=1E-14$). Therefore this value is a unique identifier for every flow in reasonable time period (1 s).

5.2 CAM

CAM block consists of TCAM (Ternary Content Address Memory). This memory is configured to 32 768 records with 64 bits length. CAM driver tries to match a hash number in TCAM. If there is a record for this flow it returns pointer to SSRAM. If there is no record for this flow it creates new record and returns pointer to allocated memory.

CAM also obeys instructions given by MAN and frees expired records and then sends acknowledge back to MAN.

5.3 MAN

Management between TCAM and SSRAM is provided by MAN. Its basic function is to hold information about flows which are stored in TCAM and SSRAM and to add or free flows to assure enough free space for incoming flows.

Disposing of inactive flows is implemented as a 3 bit-field. There is a pointer which goes round and round this field and decrement value in each row. If it reads 010 value then record is inactive and has to be disposed (2^3-2 , that is precision 1/6 of set value). Speed of pointer circling depends on value given by software. When new flow is created or matched in TCAM value in appropriate row is updated to 111b.

MAN also contains information about occupation of SSRAM that tells what rows are empty. That also allows MAN to tell SRAM whether current operation is update, new record or delete. When SRAM requires to delete record which resides too long MAN manage this operation.

For reason of pipelined processing MAN has one reserved value in this bit-field called waiting for delete. It holds record whether flow entry was deleted in TCAM or not.

There is a register that holds number of records either in TCAM or SSRAM that solves the problem with of overflow with too many records in either TCAM or SSRAM. This register also influence mode of disposing records. That means if there are too many records stored in SSRAM then MAN tries to dispose records aggressively.

To sum up, MAN gives CAM and SRAM commands (for example delete record, update) and also listen to their commands. MAN should bound CAM and SRAM to seem like one unit.

6 SRAM

Purpose of this unit is to load and update data in SSRAM.

For every flow following record must be stored:

Name	Size	Description
Start timestamp	4 Bytes	When the flow has begun
End timestamp	4 Bytes	When the flow ended
Number of bytes	8 Bytes	Total number of bytes for this flow
Number of datagrams	4 Bytes	Total number of datagrams for this flow
Flags	1 Byte	Aggregated flags from packets
Source IP address	16 Bytes	Either IPv4 or IPv6 source address
Destination IP address	16 Bytes	Either IPv4 or IPv6 destination address
Source port	2 Bytes	Source port of transport protocol
Destination port	2 Bytes	Destination port of transport protocol
Transport layer protocol	1 Byte	Transport layer protocol
Type of Service	1 Byte	Type of Service field in IP packet header
Together	59 Bytes	This field is not part of record

Table 2: Structure of record

Possible extensions are available to 64 B (for example extension of timestamp).

During update Start timestamp is checked against active time register and disposed or stored again according to interval for expiration of active flow.

When record is stored end timestamp, number of packet, number of bytes and flags are updated. This format of data allows us to gather information for 1600 seconds ($T/(G/B/S)=2^{32}/(2^{30}/8/48)$) considering the worst case. It means that there is only one flow at speed of 1 Gbps assuming one packet comes every 500 ns. Because TSU provides only 32 bits with range of 1200s it has no sense to detect overflow of number of packets or number of bytes. They must be read before they overflow. There is another bad case when every packet creates

new flow. That is a problem especially for reading via PCI bus to computer because there is no compression of data. Possible solution is to discard data which cannot be read or implement some aggregations in hardware (see section below).

Symbol	Description
T	Timestamp
G	1 Gbps
B	Byte
S	Shortest length of packet(in Bytes)

Table 3: Description of used symbols

6.1 SW_FIFO

In future design will be possible to introduce another type of aggregation. For example to place another unit between SRAM and PCI bus instead of SW_FIFO which serves as a short time buffer. This unit would be able to process data according to various aggregation schemes. These schemes would be changeable according to software wish.

7 Conclusion

This document suggests main idea of hardware architecture which could be implemented in COMBO6 cards. As result there should be hardware acceleration card for NetFlow monitoring.