# Hardware Accelerated Flow Measurement of 100 Gb Ethernet

Viktor Puš, Petr Velan, Lukáš Kekely, Jan Kořenek
CESNET, z. s. p. o
Zikova 4, 160 00 Prague, Czech Republic
Email: pus,petr.velan,kekely,korenek@cesnet.cz

Pavel Minařík
INVEA-TECH a. s.
U Vodárny 2965/2, 616 00 Brno, Czech Republic
Email: minarik@invea.com

***Abstract*—This demo demonstrates results of a joint research project of CESNET and INVEA-TECH focused on 100 GbE network flow monitoring using FPGA. It shows, to the best of our knowledge, the first flow monitoring setup capable of handling fully saturated 100 G Ethernet line. We present COMBO-CG card that provides accurate timestamps for high-resolution traffic monitoring. The card is complemented by fast DMA engine and optimized Linux drivers which were designed and implemented to achieve 100 Gbps data transfers through PCIe bus with low CPU utilization. Network traffic can be distributed among multiple CPU cores based on configurable hash functions. Our flow exporter is able to fully utilize available CPU cores to provide wire-speed performance for processing of the 100 Gbps traffic. The demo will show complete 100 G flow monitoring setup – from packet generator to flow collector.**

## I. Technology Overview

**100 Gigabit Ethernet (100 GbE)** was first defined by the IEEE 802.3ba standard ratified in 2010 and currently is the fastest existing standard of Ethernet for computer networks. As its name suggests, it enables transmission of frames at rate of 100 Gbps, which translates to more than 148 millions frames per second. That this means that a new frame is transferred every 6.7 ns. The 100 GbE standard encompasses a number of different physical layer specifications.

Our **COMBO-CG card** (shown in Figure. 1) is the first PCI Express adapter card to support 100 Gb Ethernet technology world-wide. This hardware accelerator with one 100 GbE port uses fast PCI Express bus that allows it to achieve very high throughput of data transfers between the card and memory of the host computer. This combination makes it suitable for the fastest backbone networks and high-throughput data centers.

Figure 1. Front view of COMBO-CG card

The heart of the card is an field-programmable gate array (FPGA) chip. Unlike other computing devices, such as fixed application-specific integrated circuits (ASICs) or programmable CPUs, FPGAs allow to change their internal *structure* by programming their firmware. To achieve required throughput and high processing performance, the FPGA firmware is usually designed as a deep processing pipeline, which enables to utilize FPGA's inherent massive parallelism. In the case of 100 Gbps traffic processing, 512 bits wide pipeline at 200 MHz provides a sufficient throughput.

High-speed packet capture requires specialized drivers such as the Data Plane Development Kit (DPDK) [1]. We have developed Linux device drivers with near to zero CPU overhead, tools for card management, and a zero-copy library for high-speed data transfers between the card and the host memory. The development framework for COMBO-CG card also specifies generic interface to optional traffic processing engine in FPGA. It allows to extend the functionality of a basic NIC to support packet filtering, parsing and distributing among multiple CPU cores.

**FlowMon Exporter** is a flow measurement software developed by INVEA-TECH. It is highly optimized for parallel processing which allows to utilize full processing power of current multi-core CPU architectures. The network traffic is processed in multiple threads. Due to unique functionality of the FPGA firmware, packets that belong to the same flow are passed to the same CPU core. Communication among CPU cores is significantly reduced, which leads to high processing speed.

INVEA-TECH also develops **FlowMon Collector**, which is a powerful flow collector capable of processing up to 200 000 flows per second.

We have presented network flow monitoring using two 80 G FPGA cards in our previous work [2]. However, the theoretical throughput of the cards was significantly limited by a PCIe bus. We are able to achieve the same theoretical throughput using single 100 G card utilizing two logical PCIe endpoints. In this demo we show that it is indeed possible.

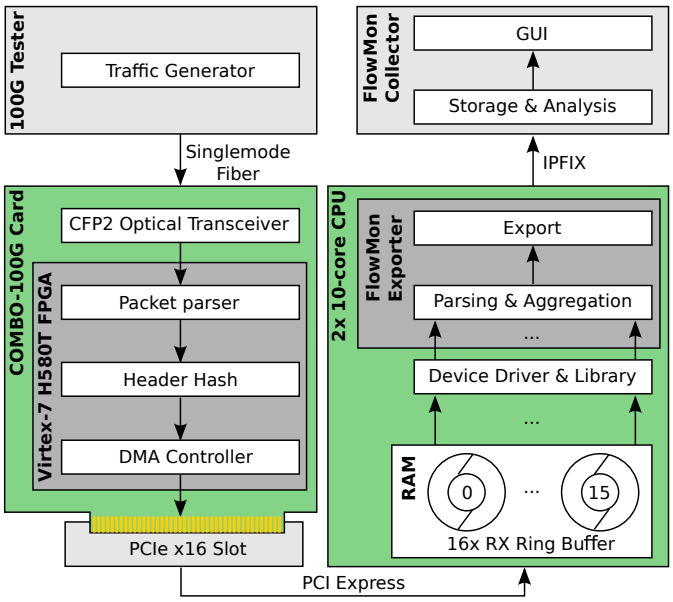


Figure 2. Illustration of demo system architecture

## II. Demo description

The demo presents our 100 G flow measurement setup. We present the whole process from packet generator to flow collector. Especially, we:

- show reception of 100 Gbps traffic at full wire-speed,
- demonstrate the hash based packet distribution to multiple cores,
- process all received packets with FlowMon Exporter,
- collect exported data on dedicated FlowMon Collector.

For the demo, we use the architecture shown in Figure 2. Network traffic is generated by the Spirent TestCenter hardware generator and sent over singlemode fiber cable to COMBO-CG card. The optical signal is converted to electrical by the CFP2 transceiver module. The FPGA decodes the packets, performs all of the steps required by the Ethernet standard (e.g., CRC check) and provides physical layer statistics about the network line, such as numbers of correct/malformed packets, protocol errors and buffer overflows. The demo will display such low-level statistics in real time.

Our 100G-capable FPGA packet parser [3] performs analysis of the packets and header field extraction. It includes support for Ethernet, VLAN, MPLS, IPv4, IPv6 (including extension headers), TCP and UDP protocols. Configurable set of packet header fields are then used to compute a hash function. Four bits of the hash are used to determine the target RX queue (one of 16 available). The DMA Controller uploads the packets into the correct RX queue, which is realized as a ring buffer in host RAM. Consistent mapping of flows to RX queues is maintained by using the hash function computed over the flow ID fields. Therefore, flows are statistically evenly distributed among RX queues. The demo shows number of packets and bytes received in each RX queue to demonstrate the distribution and raw PCI Express throughput.

FlowMon Exporter software parses the packets in sixteen independent threads. Each thread reads data from one RX queue. It is important to use flow aggregation keys that are a superset of keys used for hashing in the FPGA to ensure that packets aggregated to the same flow belong to the same RX queue. The aggregation of packets to flows is performed in a separate thread for each input. Therefore, the exporter can utilize up to $16 * 2 + 1 = 33$ threads. The last thread is used for exporting complete flows to flow collector. The exporter provides internal statistics of processed packets, flows and CPU utilization. We use these statistics in combination with export to flow collector to demonstrate the throughput of the exporter and the load of the system.

FlowMon Collector stores the flow records and provides means to analyze and visualize the data through the GUI. The data are processed in 5 minute intervals, which prevents the results to be shown in real time. However, we can use the collector to determine total number of flows, packets and bytes. Moreover, it allows to verify that the exported packets are correctly aggregated to the flows. Using the collector GUI, we demonstrate that the received traffic mix matches the setup of the Spirent packet generator.

We use two packet generator configurations for the demonstration. The first configuration tests the throughput of our solution on 64 B long packets, which is the worst case for any packet processing application. We choose UDP packets with variable IP addresses to generate $2^{19}$ concurrent flows. This number of flows is our estimation for a fully loaded 100 G network line. The second configuration attempts to simulate more realistic network traffic mix. We use a combination of VLAN, MPLS, IPv4, IPv6, TCP, UDP, ICMP and ICMPv6 packets with variable packet lengths and flow intensities to demonstrate the behavior of our system under non-uniform conditions.

The monitoring setup runs on a Dell R730 server equipped with two E5-2660 v3 CPUs, 64 GB DDR4 RAM and COMBO-CG FPGA-based card. The complete 100 G flow measurement probe fits into single 2U server. This setup provides FlowMon Exporter with 20 CPU cores which can be fully utilized for packet reception, processing, flow management and export.

## Acknowledgement

This material is based upon work supported by the project TA03010561 funded by the Technology Agency of the Czech Republic and the "CESNET Large Infrastructure" project LM2010005 funded by the Ministry of Education, Youth and Sports of the Czech Republic.

## References

[1] Intel, "Data plane development kit," http://dpdk.org/.

[2] P. Velan and V. Puš, "High-Density Network Flow Monitoring," in *Proceedings of the 2015 IEEE International Symposium on Integrated Network Management, IM 2015*, 2015, to appear.

[3] V. Puš, L. Kekely, and J. Kořenek, "Design methodology of configurable high performance packet parser for fpga," in *17th IEEE Symposium on Design and Diagnostics of Electronic Circuits and Systems*. Warszaw, Poland: IEEE Computer Society, 2014, pp. 189–194.