CESNET and INVEA-TECH demonstrate 100 Gbps transfers over PCIe with a single FPGA

Introduction

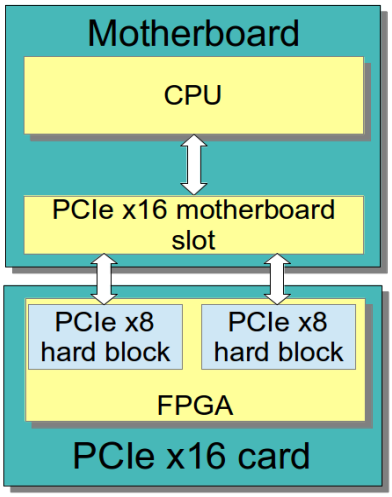
With the rising adoption of 100G Ethernet standard, there are growing opportunities for various hardware appliances supporting this throughput. A considerable part of this landscape can make use of a combination of FPGA and x86 CPU, which offers a good tradeoff in terms of programmability, performance and ease of use.

Current FPGAs support PCIe hard macros up to gen3 x8. Given the data rate of 8 Gbps per PCIe gen3 lane, the theoretical throughput of a single PCIe interface supported by any current FPGA is 64 Gbps - not enough for 100 Gbps applications. Real throughput is even lower due to the PCIe protocol overhead. The question here is: Can we get to 100 Gbps with current FPGAs?

There is an option to include a PCIe switch chip in the x8+x8=x16 configuration on the card. The switch chip will, however, have a power consumption somewhere around 6 Watts, not to speak about the complicated PCB and increased BoM. There is a much more elegant way to 100 Gbps: PCIe bifurcation.

Bifurcation is the way to go

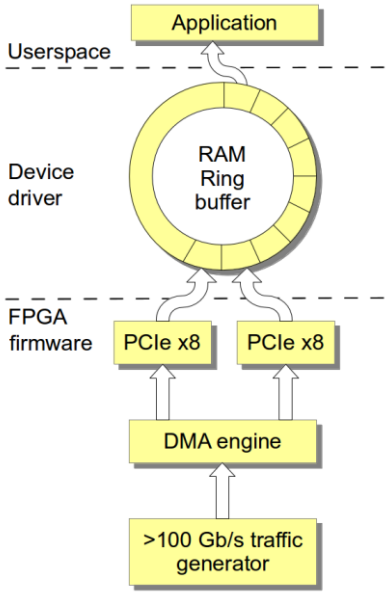
This concept was introduced by Intel's Core CPUs a couple of years ago, but the necessary support from the motherboard vendors was often neglected. With this situation slowly changing, it is now possible to make use of bifurcation to build a 100 Gbps system with a single FPGA and without the need for PCIe switch chip on-board. With bifurcation, single physical PCIe x16 slot can be configured at boot time to function as two electrical and logical PCIe x8 interfaces. The operating system running on the CPU sees two logical PCIe devices, but this can be easily disguised by the device driver so that the user applications see a single (yet very fast) application layer interface.

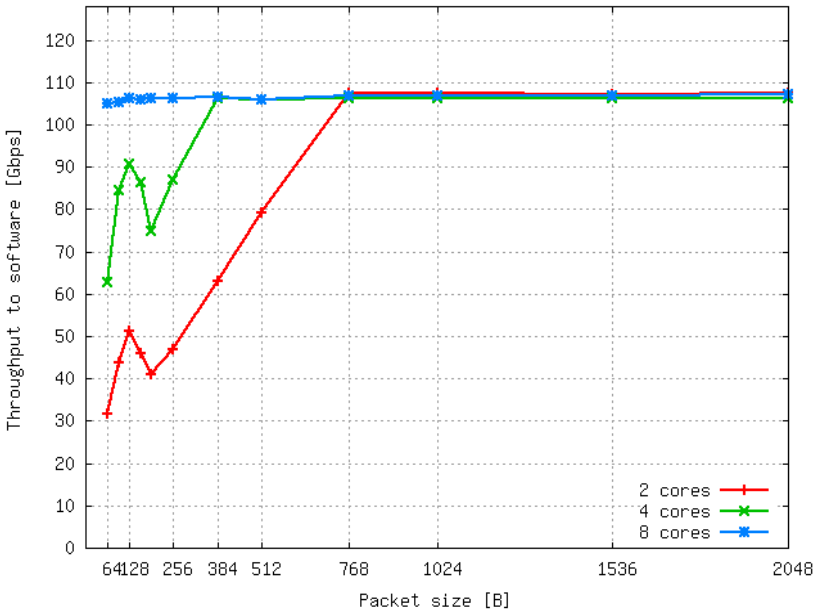


Demonstration

CESNET and INVEA-TECH have conducted a series of experiments to demonstrate the real benefits of PCIe bifurcation. The test setup included a custom FPGA card equipped with Xilinx Virtex-7 H580T. Two of the FPGA's PCIe x8 hard blocks were connected to a single PCIe x16 slot of the card. The FPGA firmware worked in a concert with Linux device drivers to transfer data to the buffer in the PC's RAM. In the bifurcation setup, both x8 interfaces were used in a round-robin manner to transfer data to a single buffer. Even though the card used has a 100G Ethernet interface via CFP2 optical module, a random packet data generator inside the FPGA was used to generate traffic even faster than 100 Gbps. The results are shown in the following graph:

INVEA-TECH, U Vodarny 2965/2, 616 00 Brno, Czech Republic
+ 420 511 205 250 | +420 602 647 684 | info@invea.com
www.invea.com

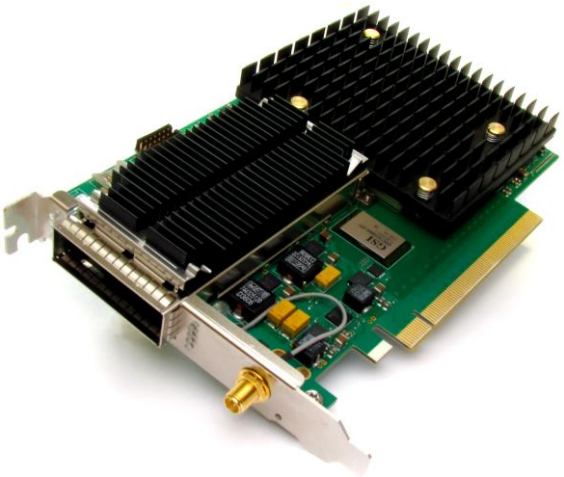




Achieved throughput is 107 Gbps and is computed from the payload of PCIe transaction layer, so that no other overhead is imposed by PCIe. The DMA Engine groups packets together, so that the packet length has no effect on the raw PCIe throughput. However, performance of the software application used is limited in the packets per second domain. Traffic distribution among multiple CPU cores (at least 8 in our setup) is needed to scale the processing to 100 Gbps.

Conclusion

CESNET and INVEA-TECH have demonstrated 100 Gbps transfers over PCIe with the use of bifurcation. The demonstration shows how to construct a packet capture solution with Xilinx FPGA. Data processing at the CPU quickly becomes a bottleneck, which can however be solved by a smart distribution of packets among multiple CPU cores. Further performance scaling is possible through offloading time-critical parts of the application into FPGA chip. The demonstration shows that PCIe bifurcation is a feature which allows to improve the throughput of FPGA to CPU transfers without the need for a discrete PCIe switch chip.



INVEA-TECH has already integrated this approach into their current portfolio of packet capture cards.

Acknowledgment

This work was supported by the project TA03010561 funded by the Technology Agency of the Czech Republic.

INVEA-TECH, U Vodarny 2965/2, 616 00 Brno, Czech Republic
+ 420 511 205 250 | +420 602 647 684 | info@invea.com
www.invea.com

About CESNET

CESNET, association of legal entities, was held in 1996 by all universities of the Czech Republic and the Czech Academy of Sciences. Its main goals are operation and development of the Czech NREN, research and development of advanced network technologies and applications, broadening of the public knowledge about the advanced networking topics. CESNET is a long-time Czech academic network operator and participant of corresponding international projects.

Department of Tools for Monitoring and Configuration (formerly the Liberouter project) focuses its research activities at hardware acceleration of network security and monitoring technologies as well as distributed storage and processing of monitoring data and network anomaly detection algorithms.

About INVEA-TECH

INVEA-TECH is a leading manufacturer and provider of high-performance network solutions. We offer complete portfolio of products for acceleration of traffic processing using FPGA technology. These products are ideal both for OEM use, development of hardware-accelerated applications as well as R&D and prototyping. Our network adapters with advanced features allow our customers and partners gain a competitive advantage in the world of high-speed networks.

INVEA-TECH was founded in 2007 with headquarter located in Brno, Czech Republic. Our team and technology emerged from Masaryk University, Brno University of Technology, CESNET association and Liberouter project.

INVEA-TECH, U Vodarny 2965/2, 616 00 Brno, Czech Republic
+ 420 511 205 250 | +420 602 647 684 | info@invea.com