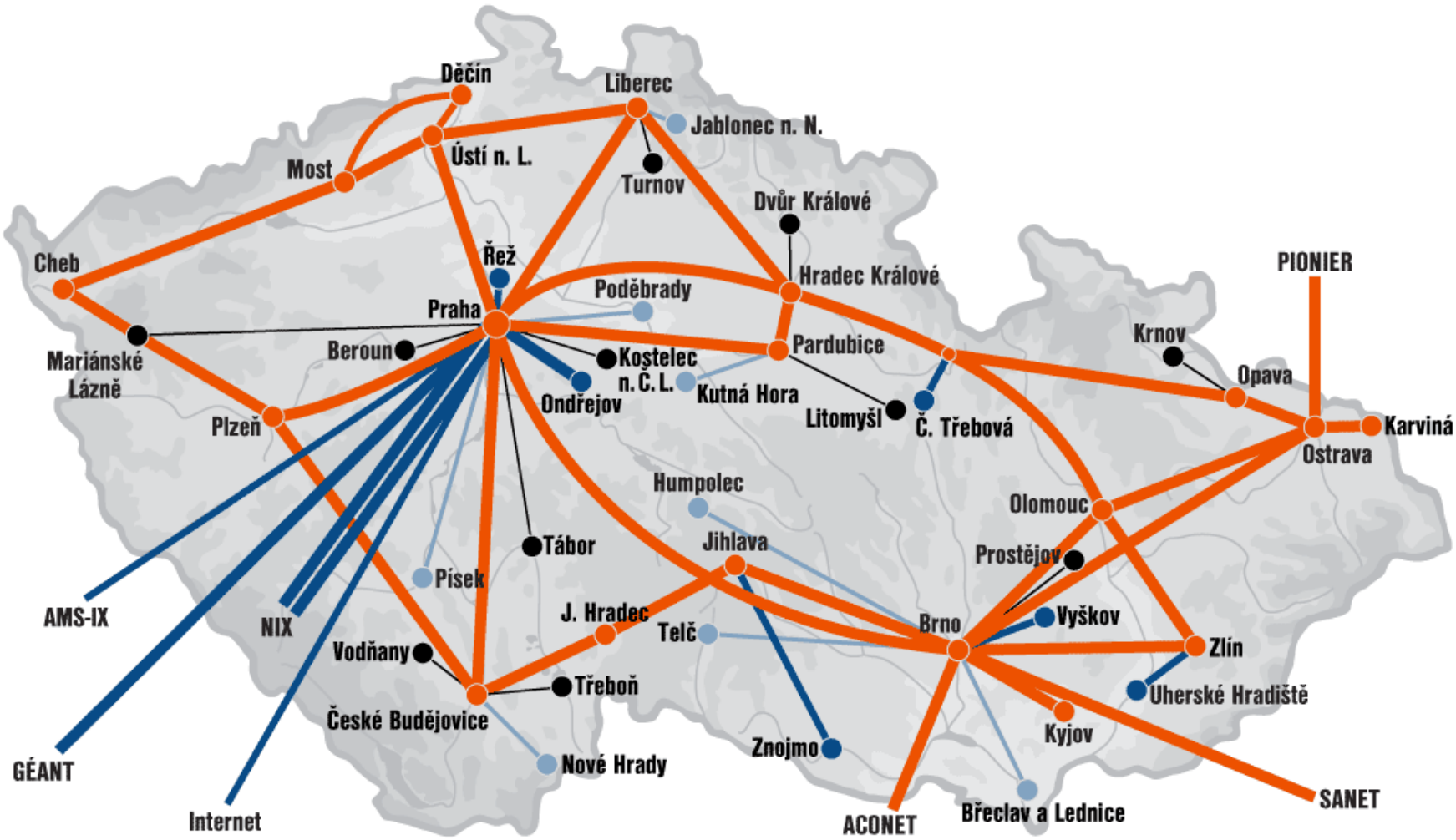
Flexible network monitoring at 100Gbps and beyond

Lukáš Kekely, Viktor Puš

{kekely,pus}@cesnet.cz

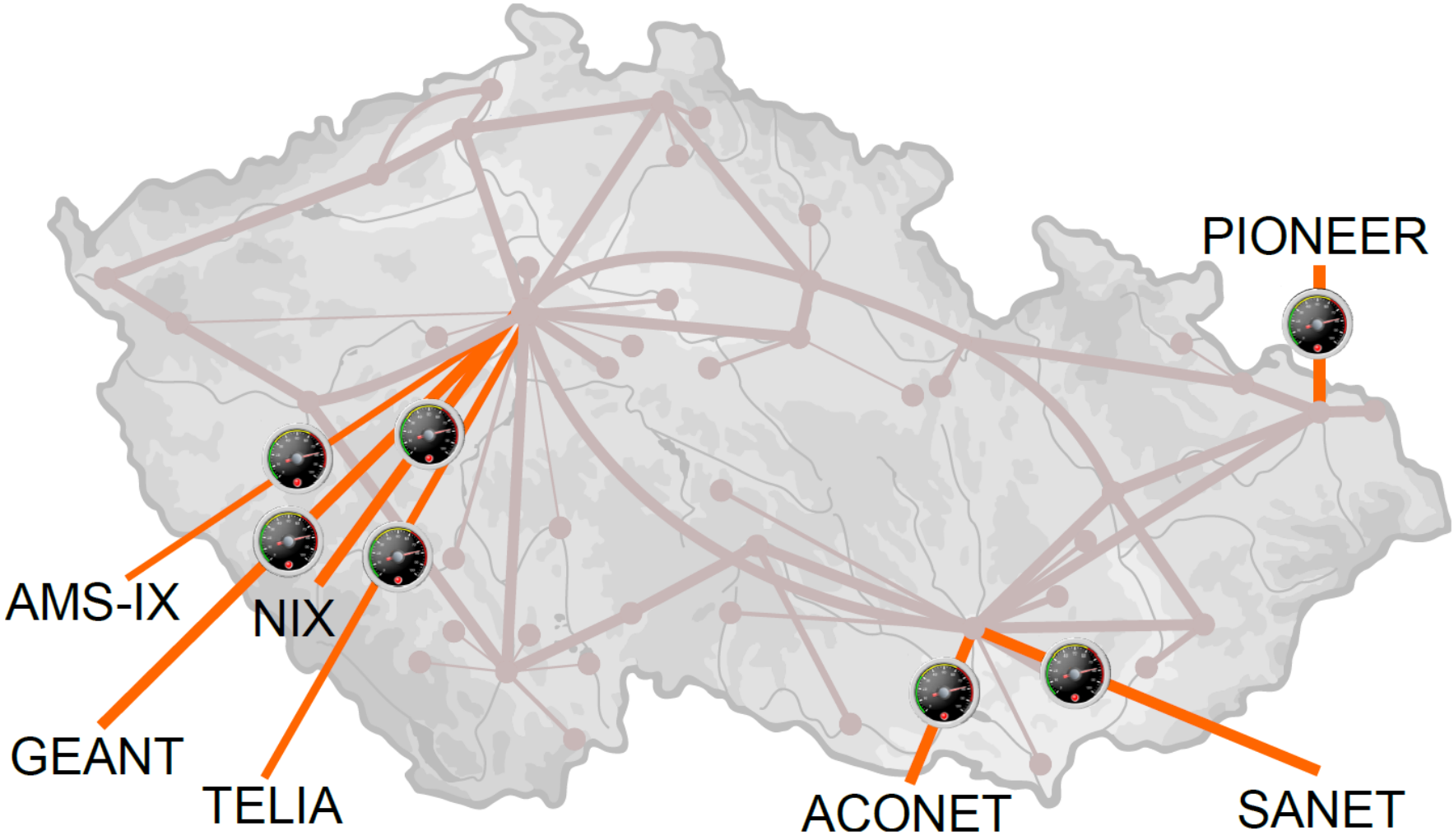
2nd SIG-PMV meeting
17th May 2017

- Czech NREN with over 400,000 connected users



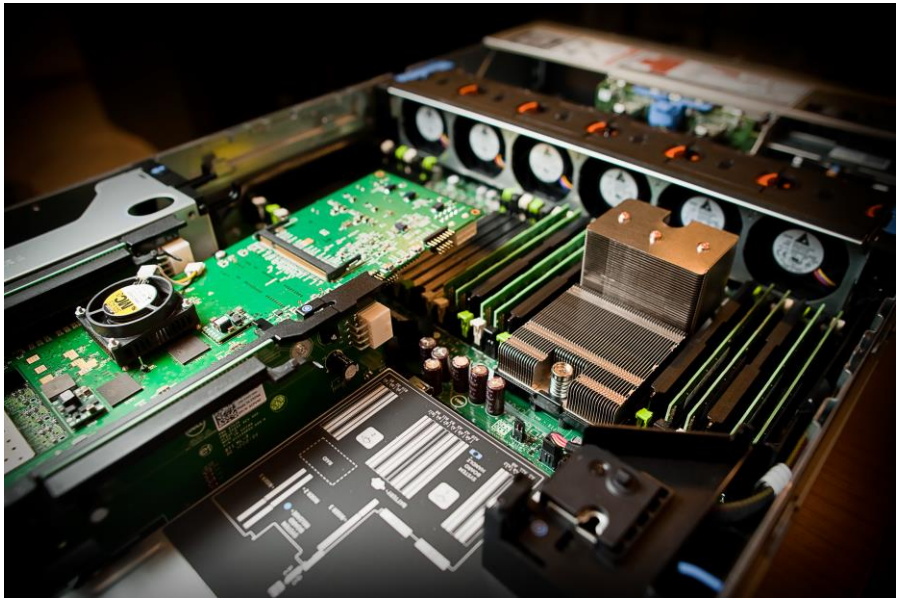
CESNET monitoring (Liberouter group)

- 7 metering points guarding the perimeter @ **40/100 Gbps**

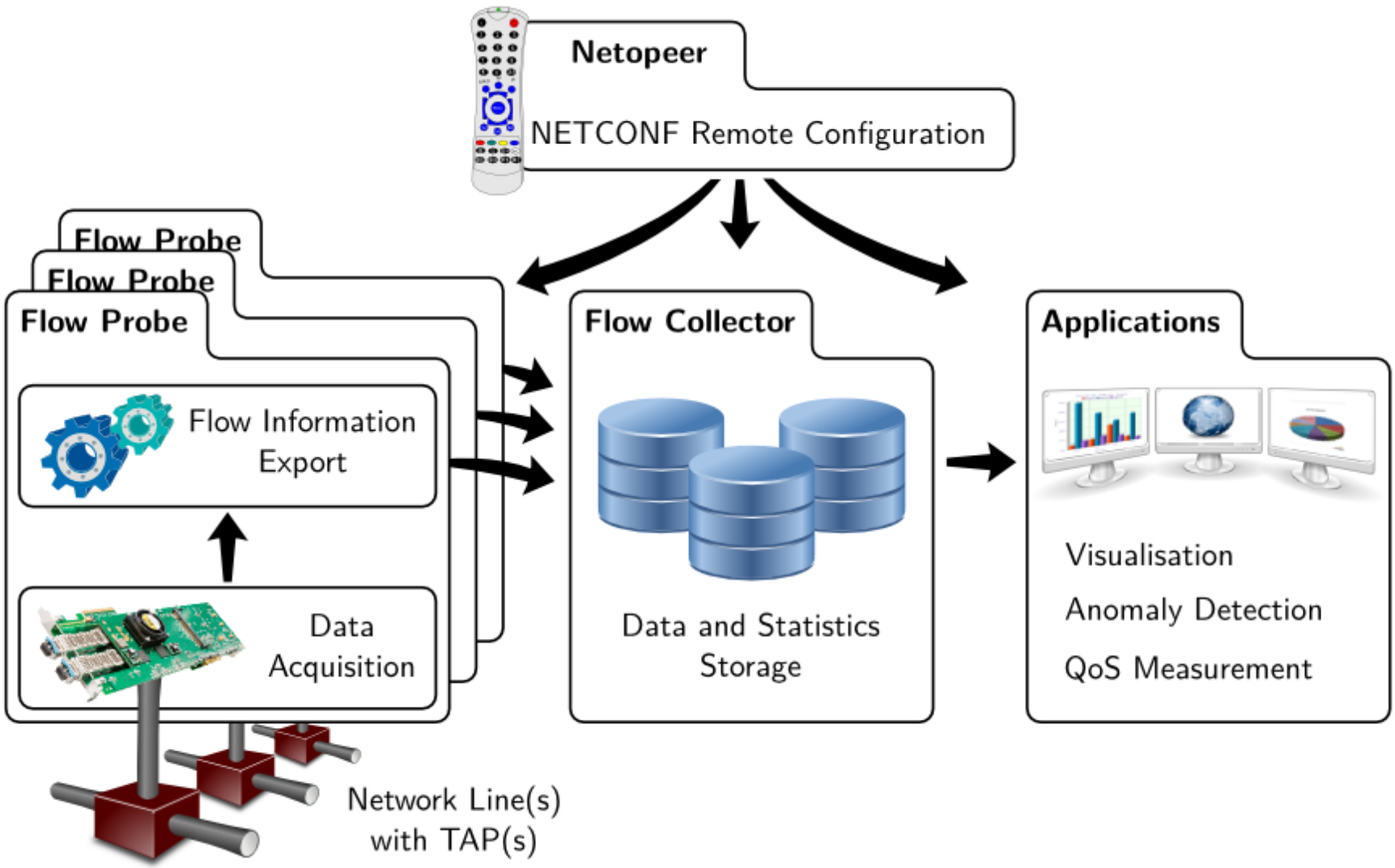


Monitoring point

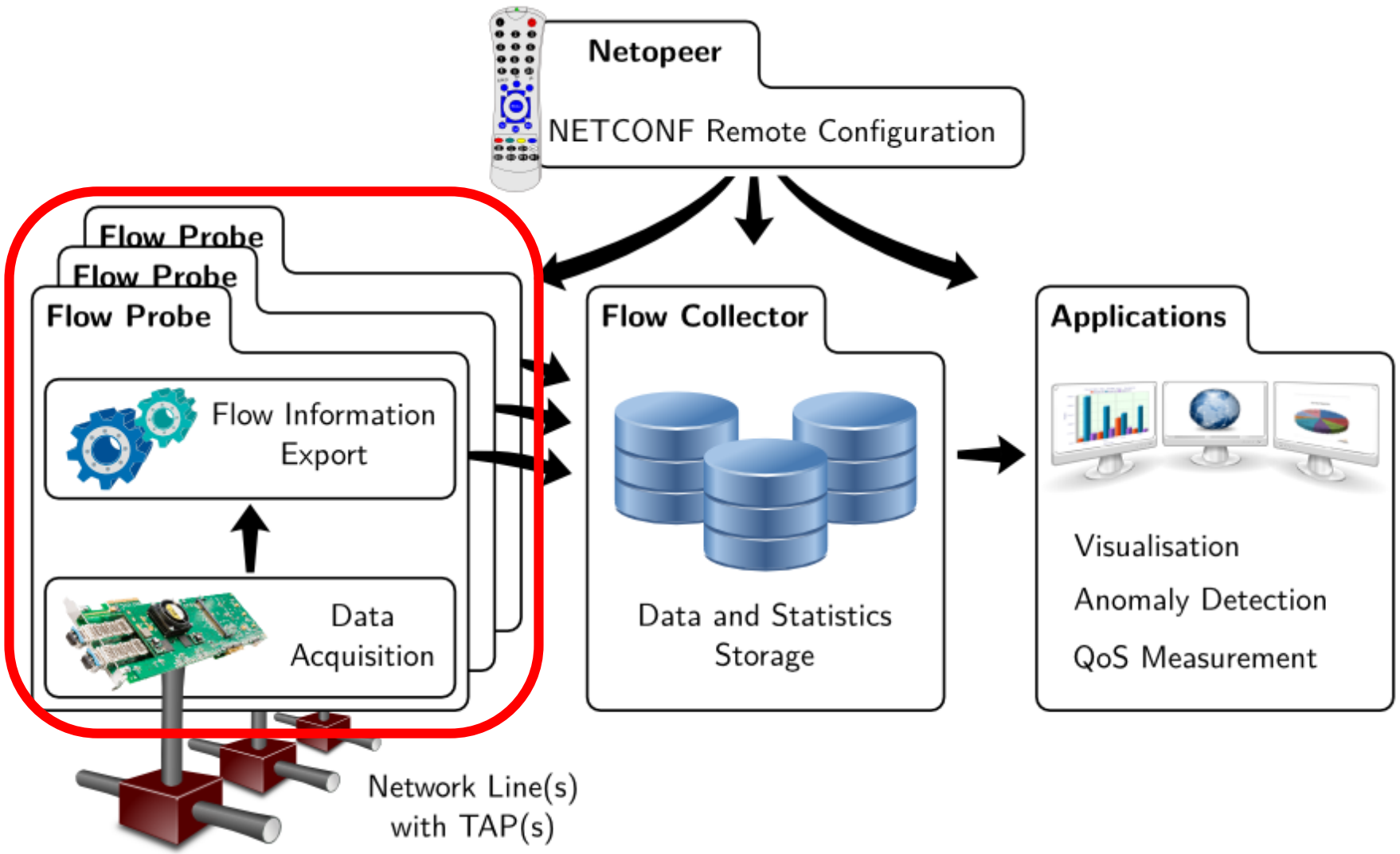
- TAPed network link
- commodity Linux server(s)
  - production and testing
- FPGA accelerated NICs



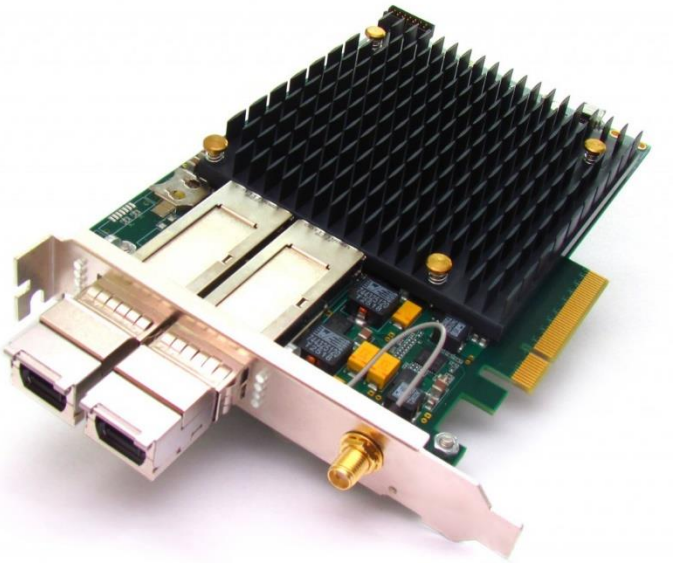
Monitoring overview

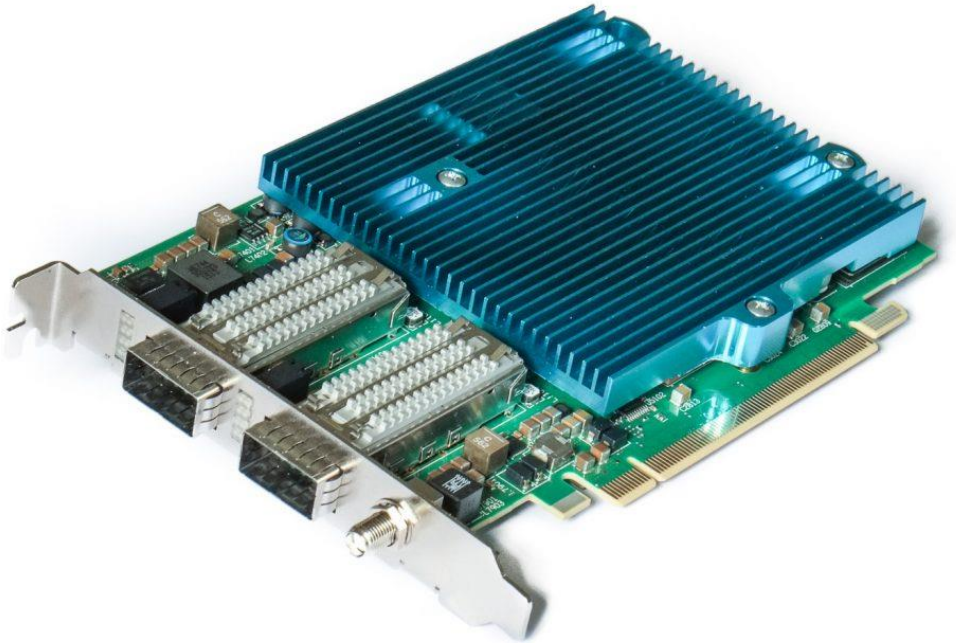


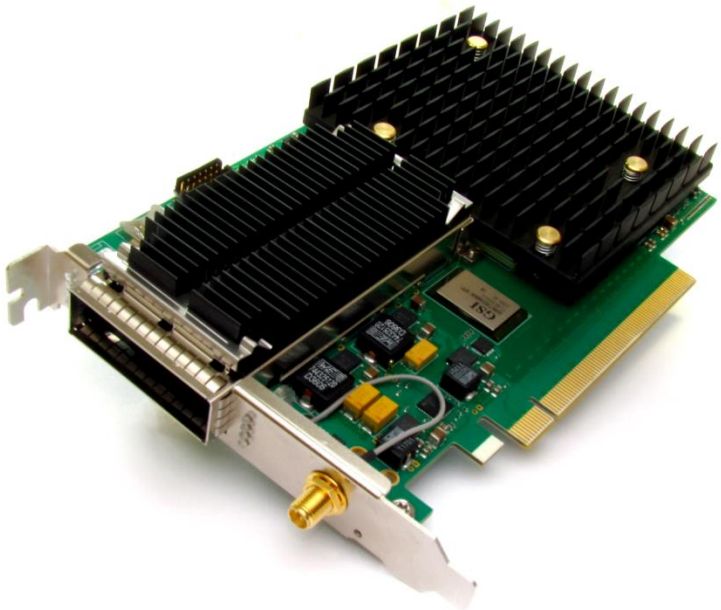
Monitoring overview

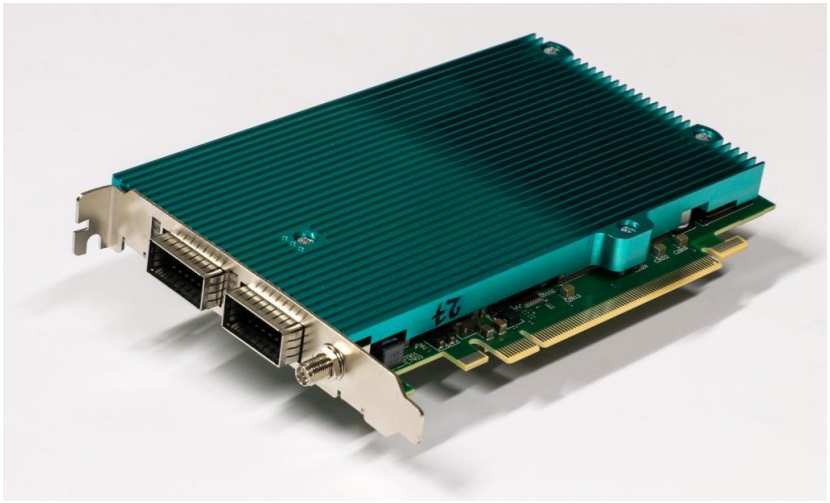


Family of accelerated NICs



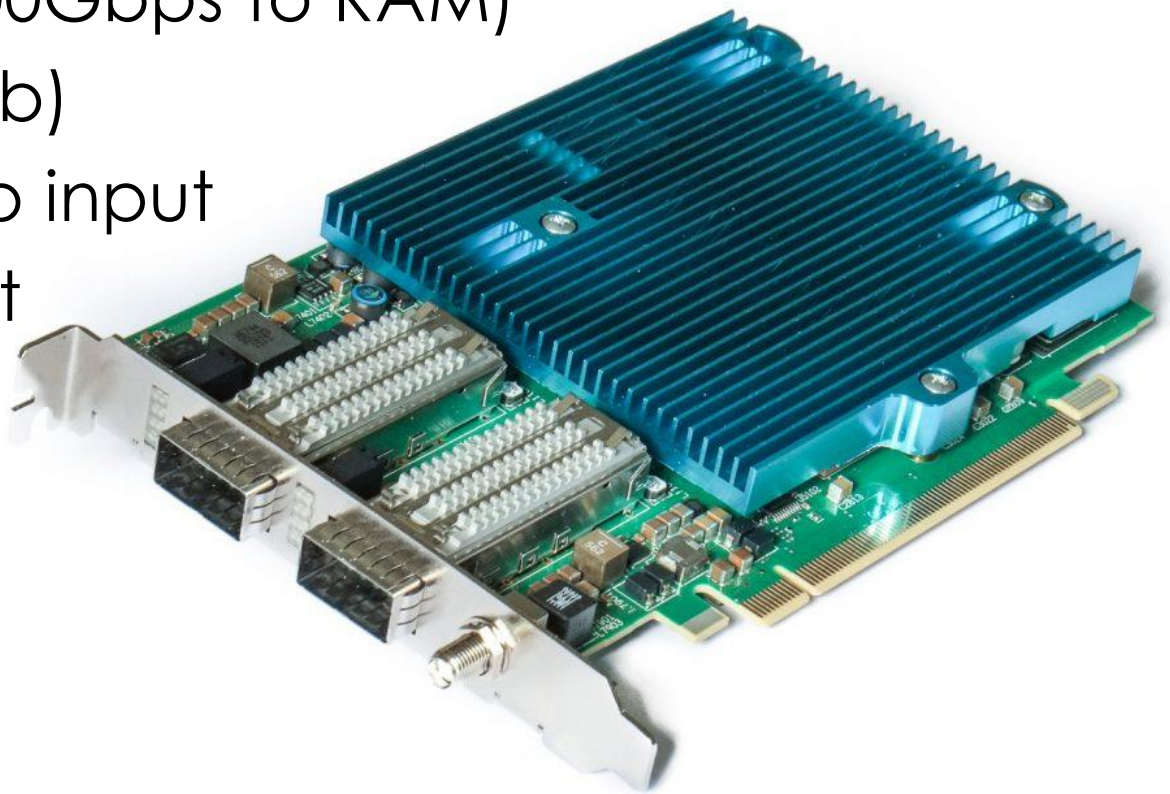






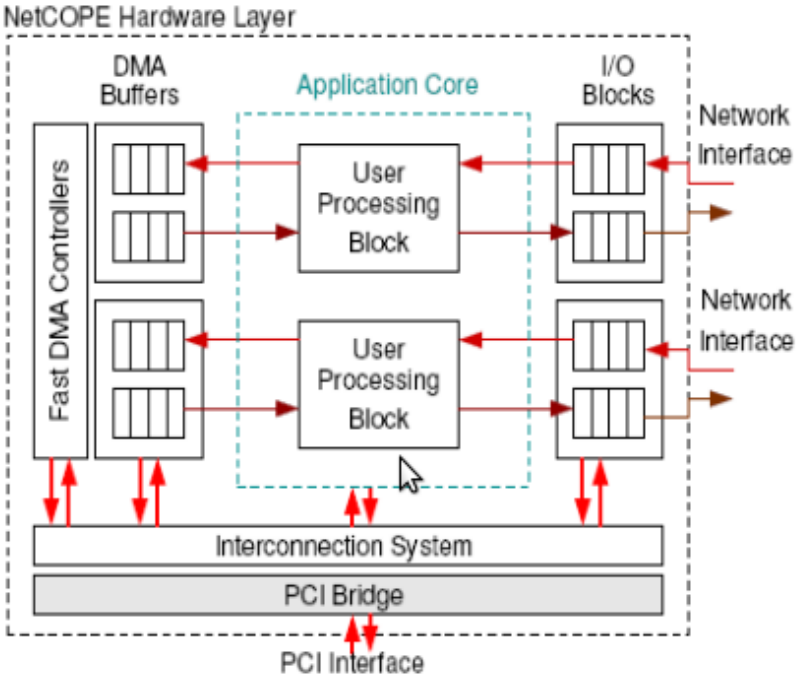
NFB-100G2Q

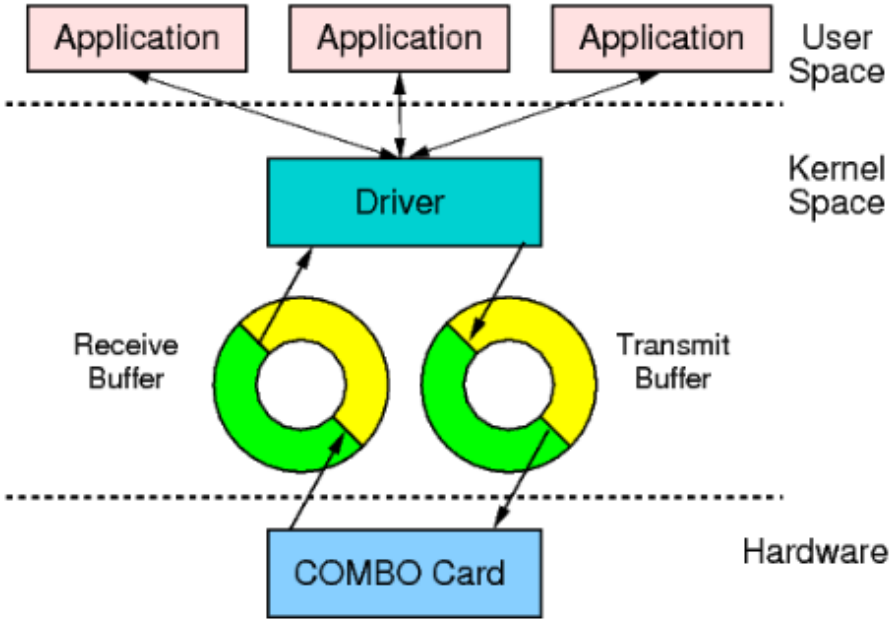
- ***Virtex7 H580T FPGA***
- 2x QSFP28 transciever cage
  - 100GE or 4x 10GE
- PCIe x16 gen3 (100Gbps to RAM)
- 3x QDRIIIe (3x72Mb)
- precise timestamp input
- Intel DPDK support



NetCOPE platform

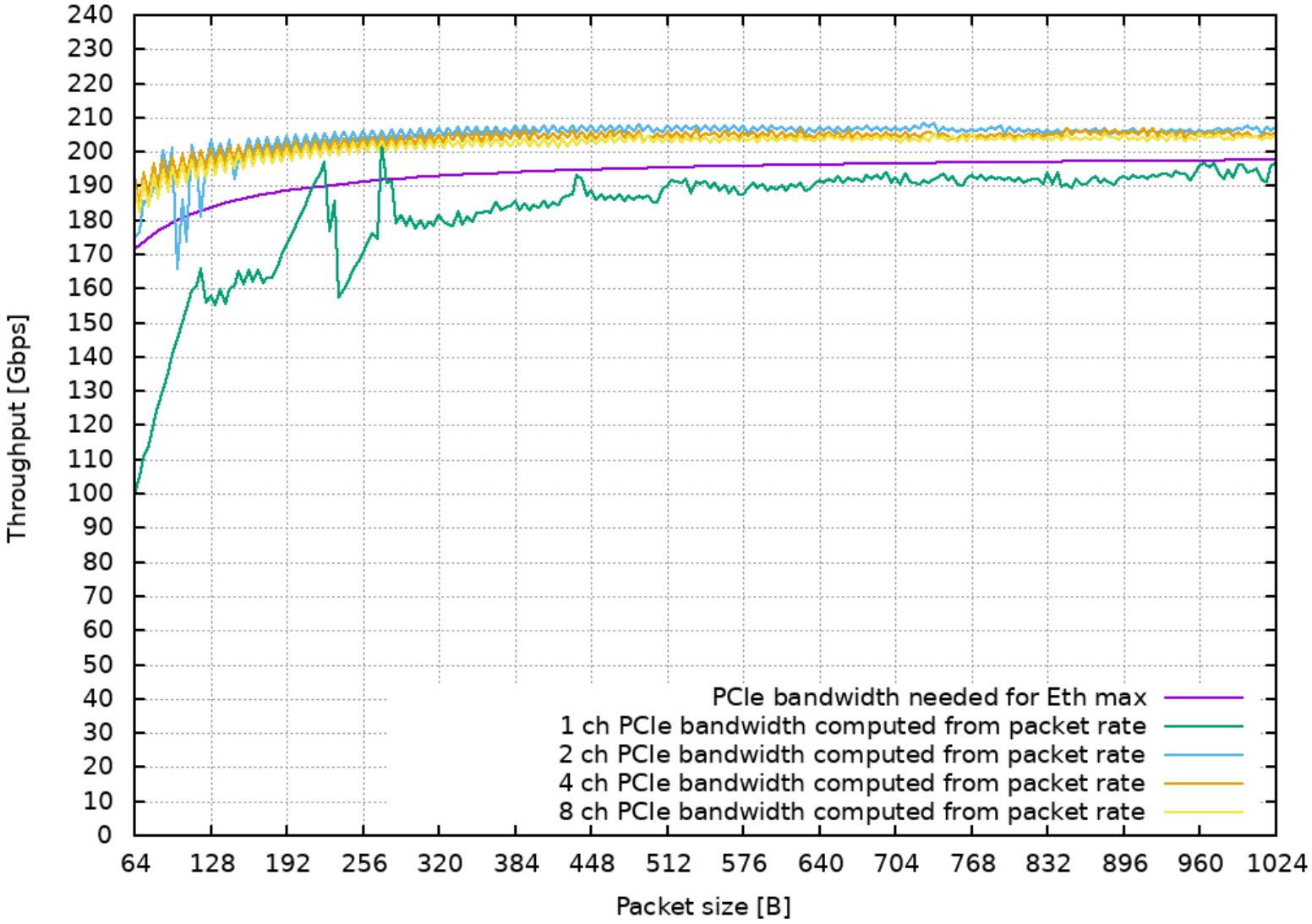
- rapid development of network applications on our NICs
  - multi-card support (porting) made easy
  - commonly usable IP cores (network modules, parsers …)
  - generic data transfer protocol towards used accelerators
- **fast DMA** transfers of packets into host memory





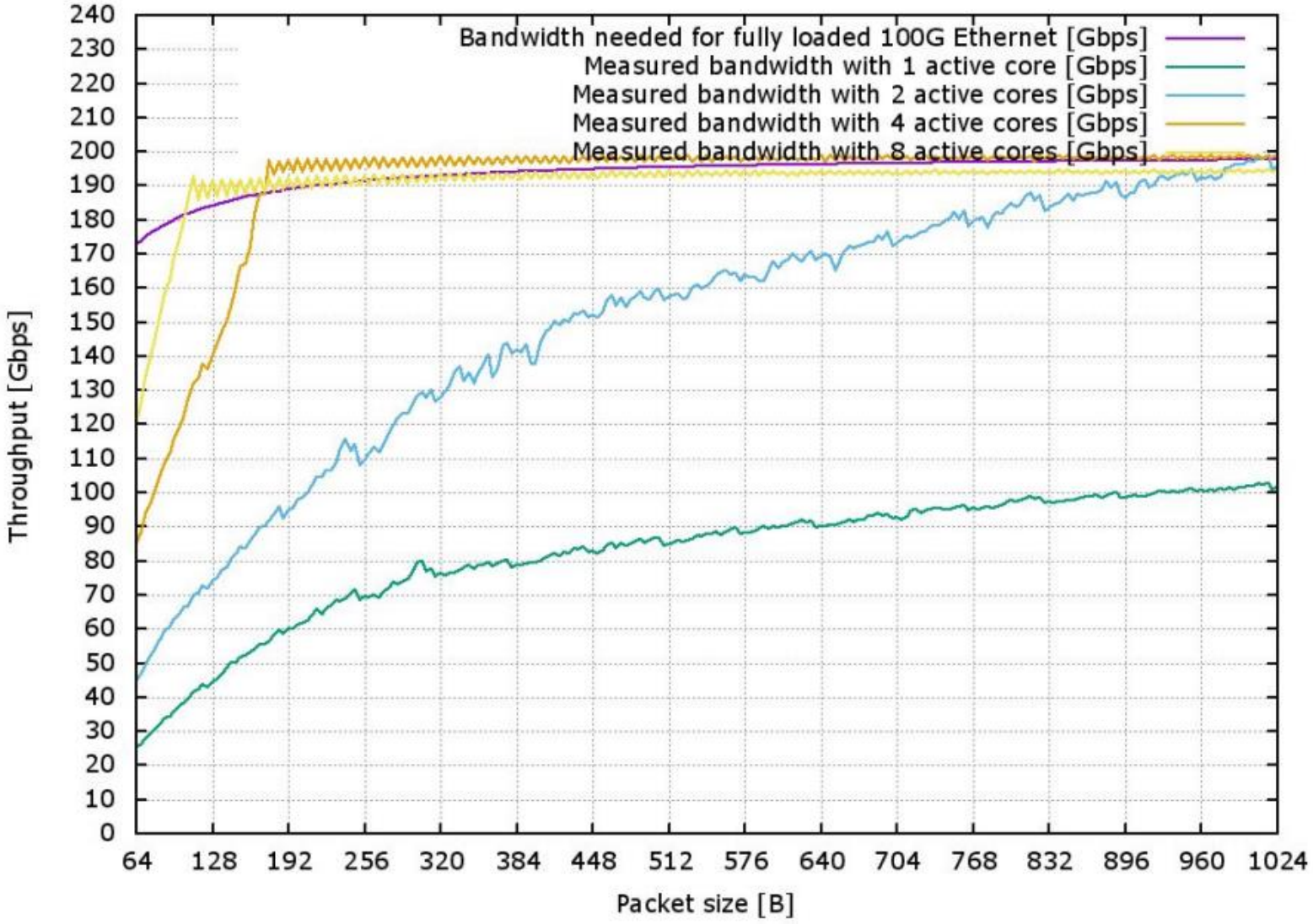
DMA bus-master: proprietary SZE2

- the fastest DMAs available – full-duplex **100GE line-rate**



DMA bus-master: Intel DPDK

- DPDK **performance record** set in April



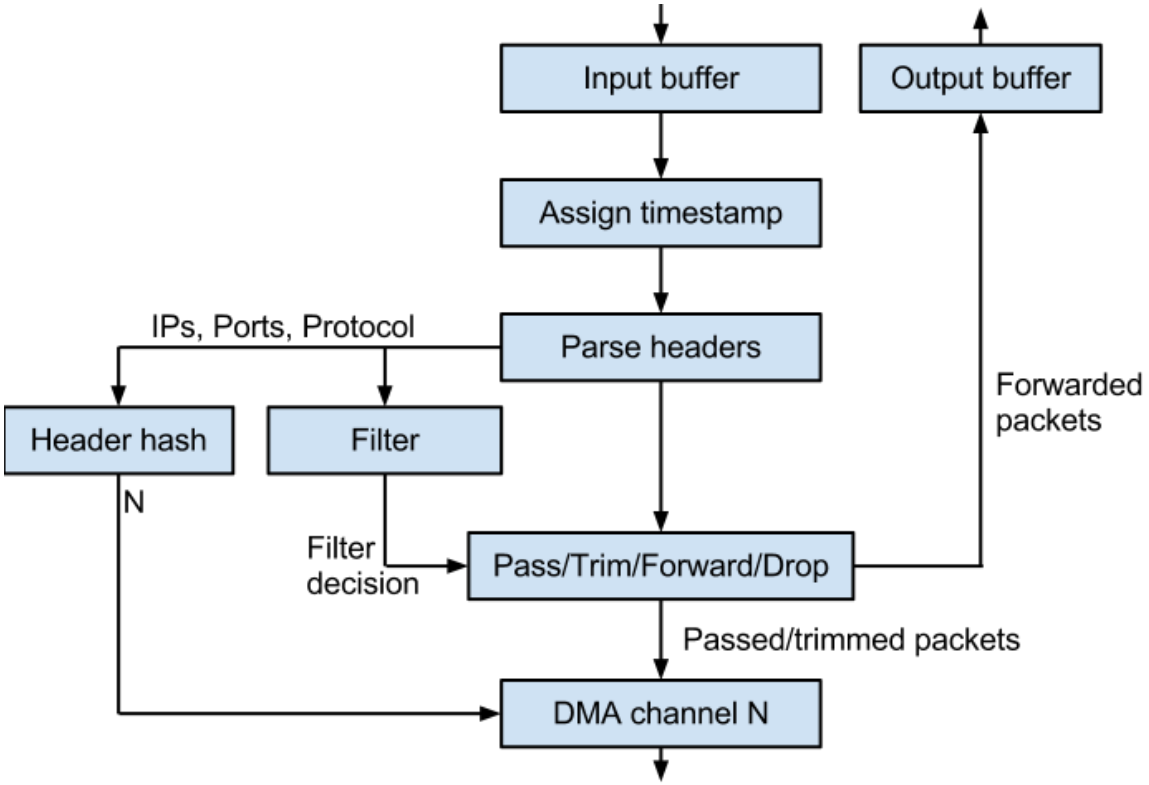
P4 language

- high-level language for description of packet processing
  - protocol stack independent header parsing of incoming packets
  - decision making and related actions (match-action tables)
  - modification and assembly of outgoing packets

- development of unique **P4-to-VHDL** translator (generator)
  - parsing & de-parsing done; match-action underway
- live demonstration today at **P4 Workshop @ Stanford**
  - P4 generated 100GE In-Band Network Telemetry (INT) sink
  - delay heatmap of the whole network visualized as a result

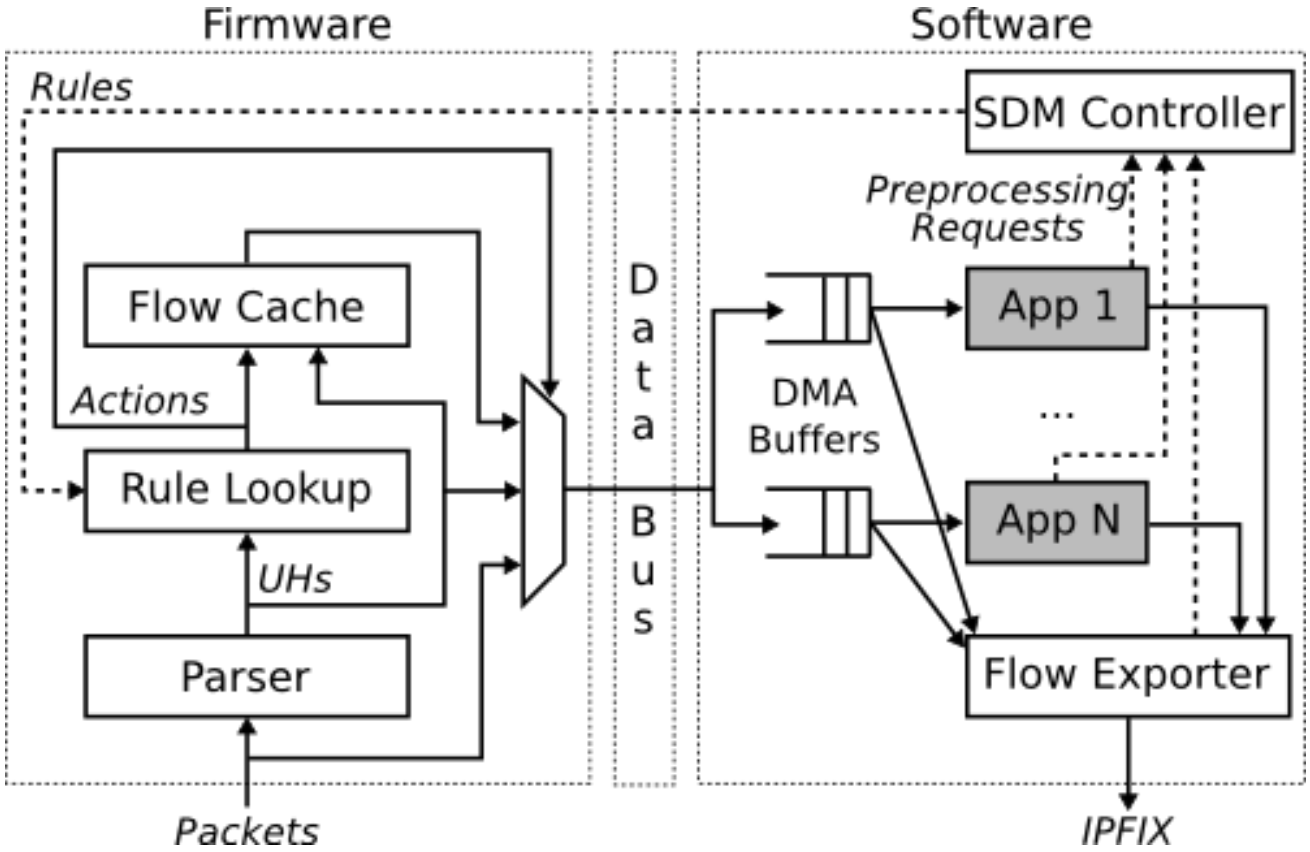
Hardware accelerated NIC (HaNIC)

- accelerated **packet capture** solution with extra features
  - flow-aware (hash-based) traffic distribution
  - packet filtering/classification – IP prefixes, ports, protocol …
  - bi-directional flows, sampling, trimming, headers

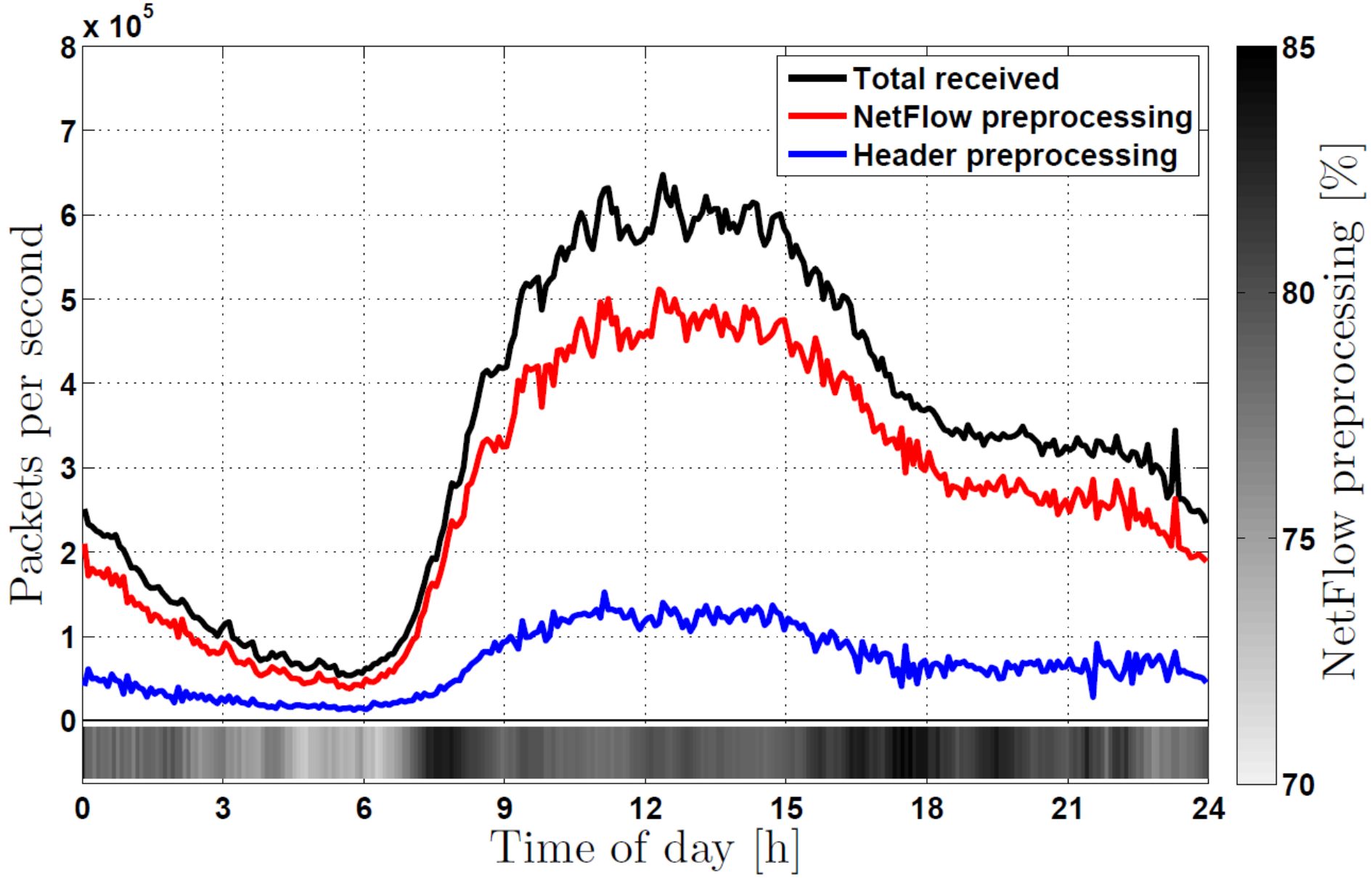


Software Defined Monitoring (SDM)

- new concept of hardware accelerated **flow monitoring**
  - extensible application-specific processor for stateful flow processing
  - SW applications can offload processing of bulk traffic to HW
  - aimed to enable high-speed application layer monitoring



Software Defined Monitoring (SDM)

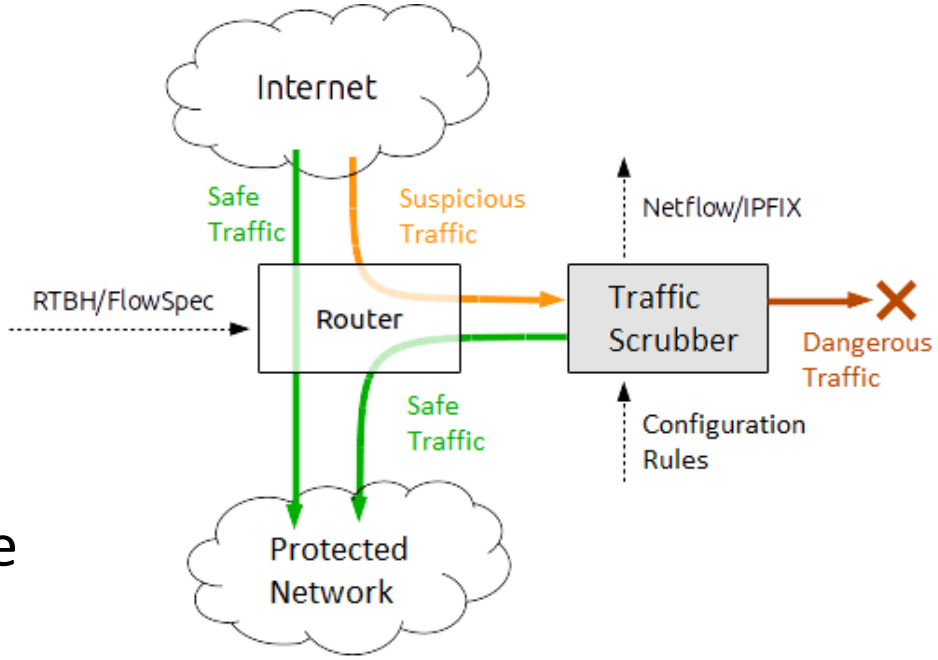


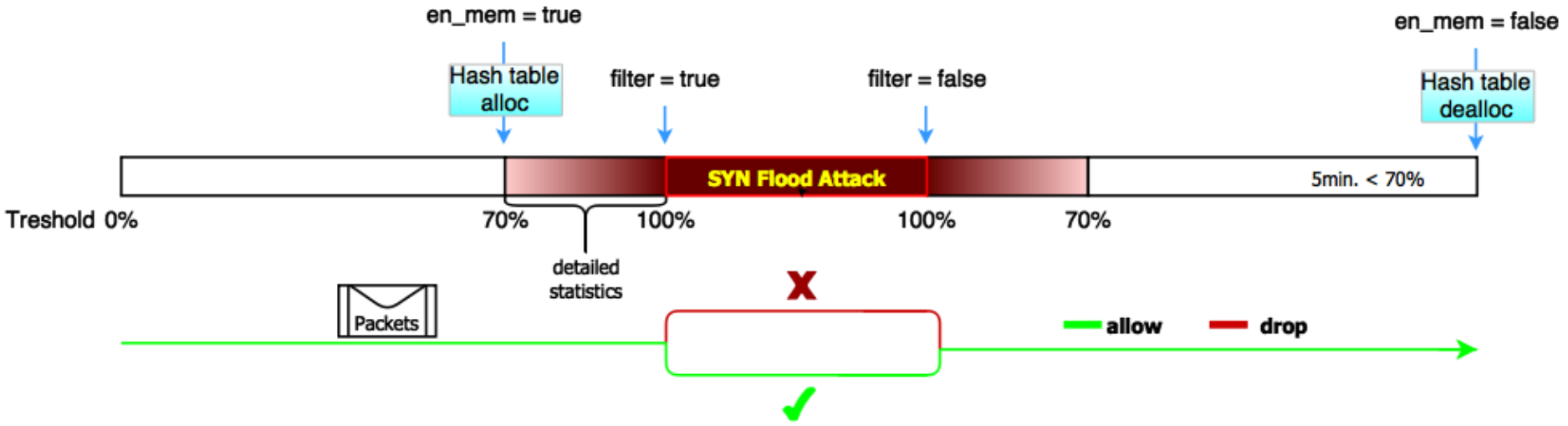
Flow exporter

- we use **FlowMonExp** from our partner Flowmon Technologies
  - highly optimized implementation (hugepages, NUMA aware …)
  - comfigurable management of flow cache records
  - flexible architecture supporting user defined plugins
  - input – PCAP, DPDK, our SZE2 format, preprocessed packets
  - processing – DNS & HTTP analyzers, Heartbleed detector
  - export – CSV, NetFlow, IPFIX

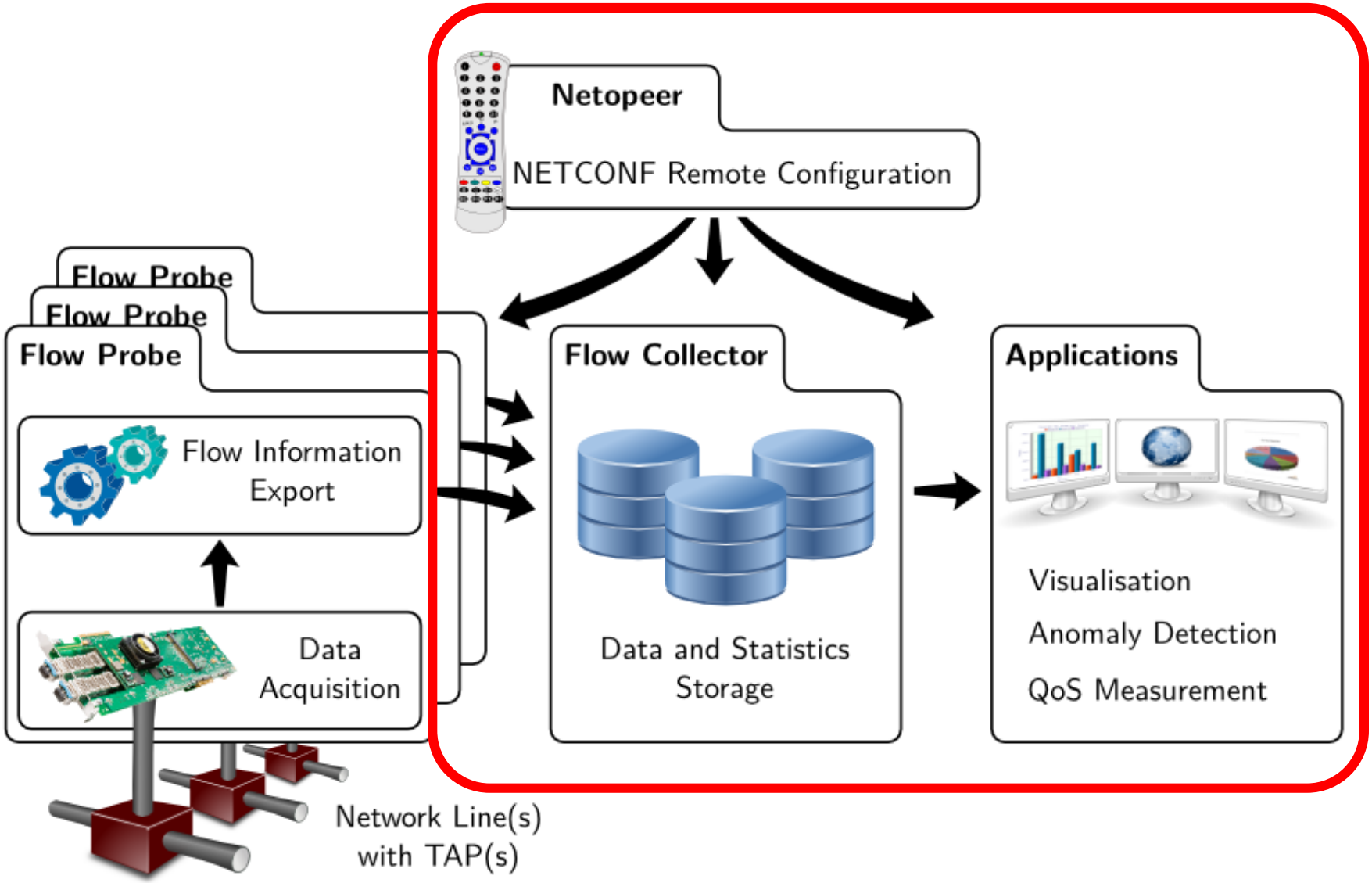
DDoS scrubber

- separate DDoS packets from legitimate traffic
  - HaNIC firmware with extra features (rate limit, VLAN tag)
  - measurement of statistics and mitigation of detected attacks
  - 100 Gbps (10x10GE) prototype already deployed in network

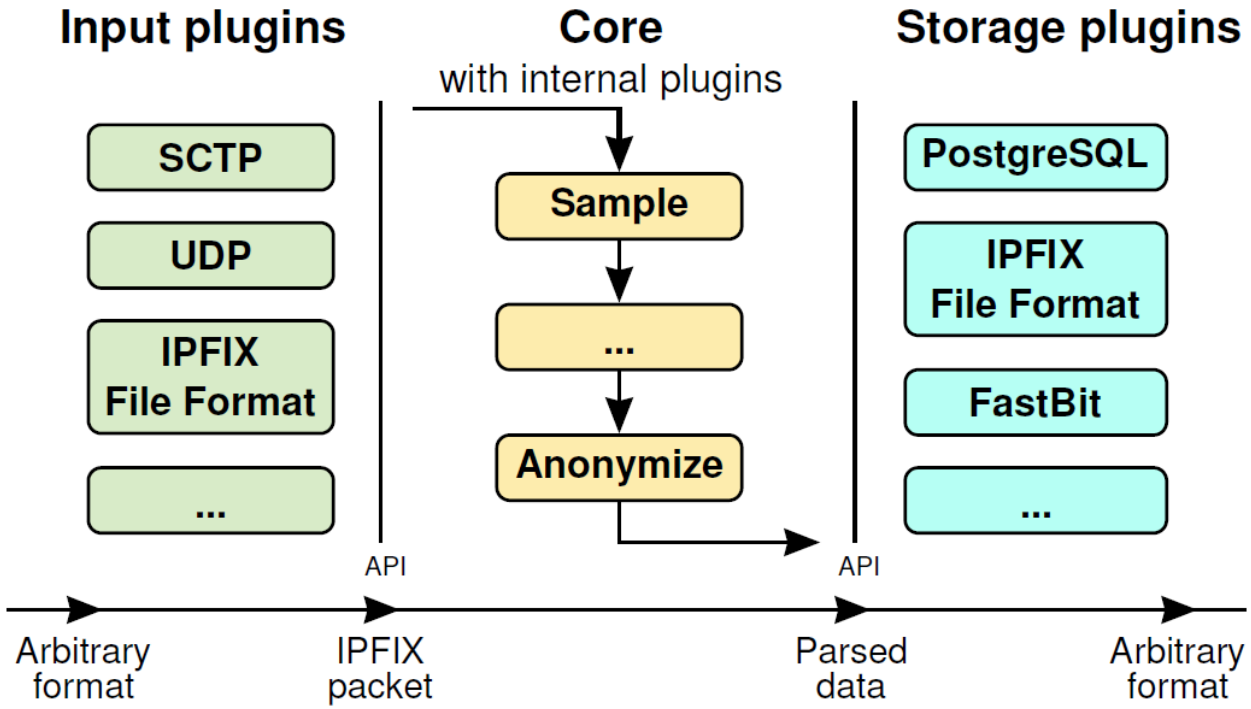




Monitoring overview



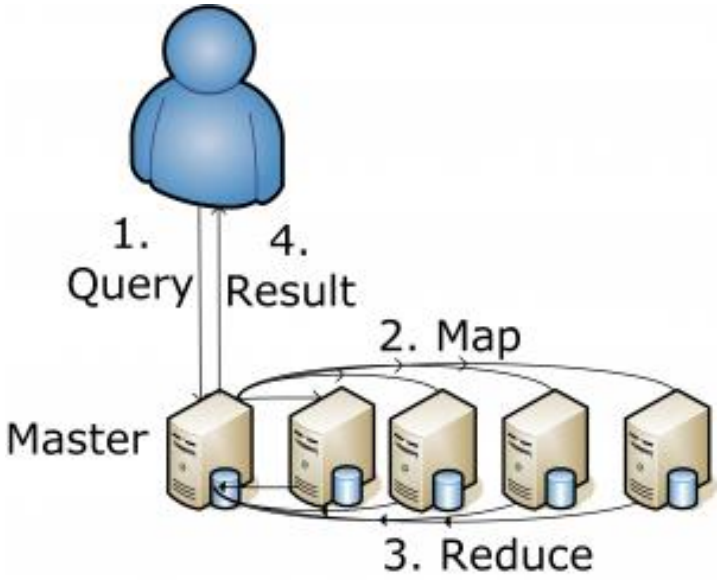
- collector **fully supporting IPFIX** including enterprise elements
  - include tools for subsequent data processing and mediation
  - high-performance **sufficient for 100GE** environment
  - extensible by various plugins (input, intermediate, storage)



- open-source in C++ - https://github.com/CESNET/ipfixcol/

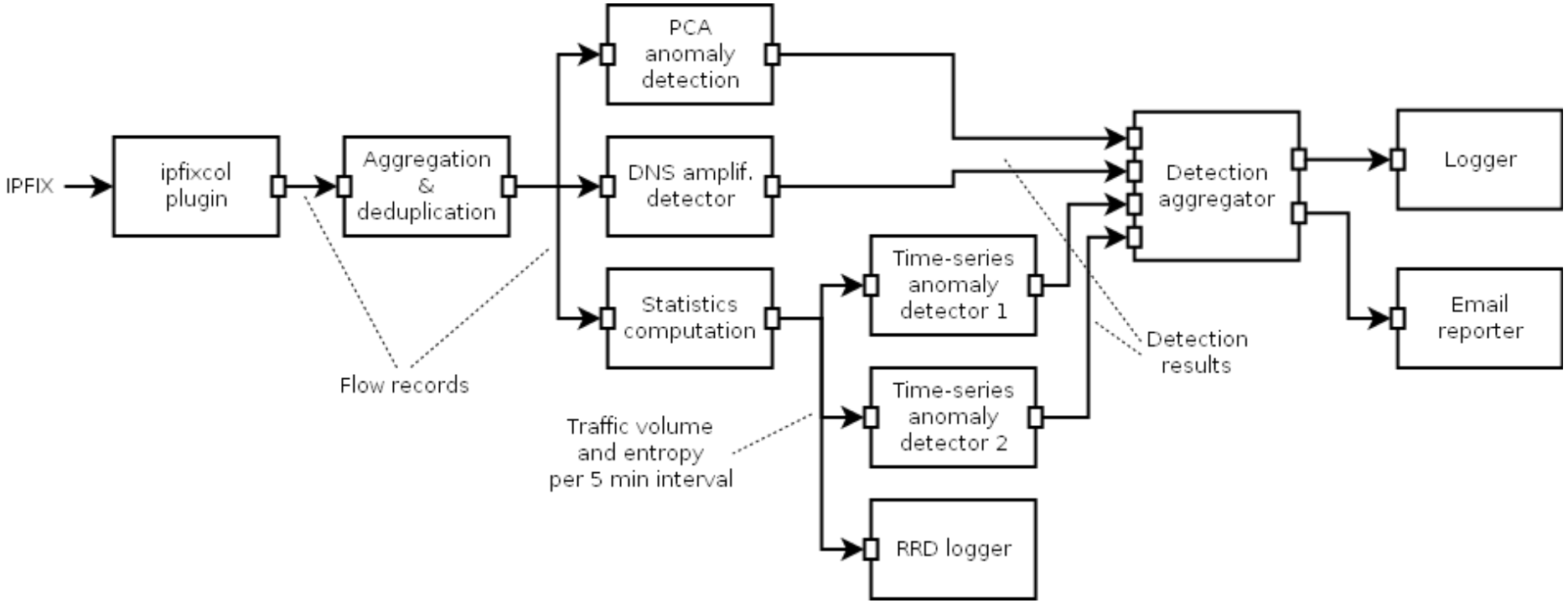
- **distributed** flow-based collector in development
  - master-slaves and proxy architecture
  - based on **IPFIXcol** to store and distribute data

  - **fdistdump** to execute queries on slaves



NEtwork MEasurements Analysis (NEMEA)

- framework for automated **real-time analysis** of flow data
  - build as a user-defined collection of various modules
  - TRAP + UniRec = high-performance and easy distribution
  - detected threads reported to CERTS/CSIRT systems



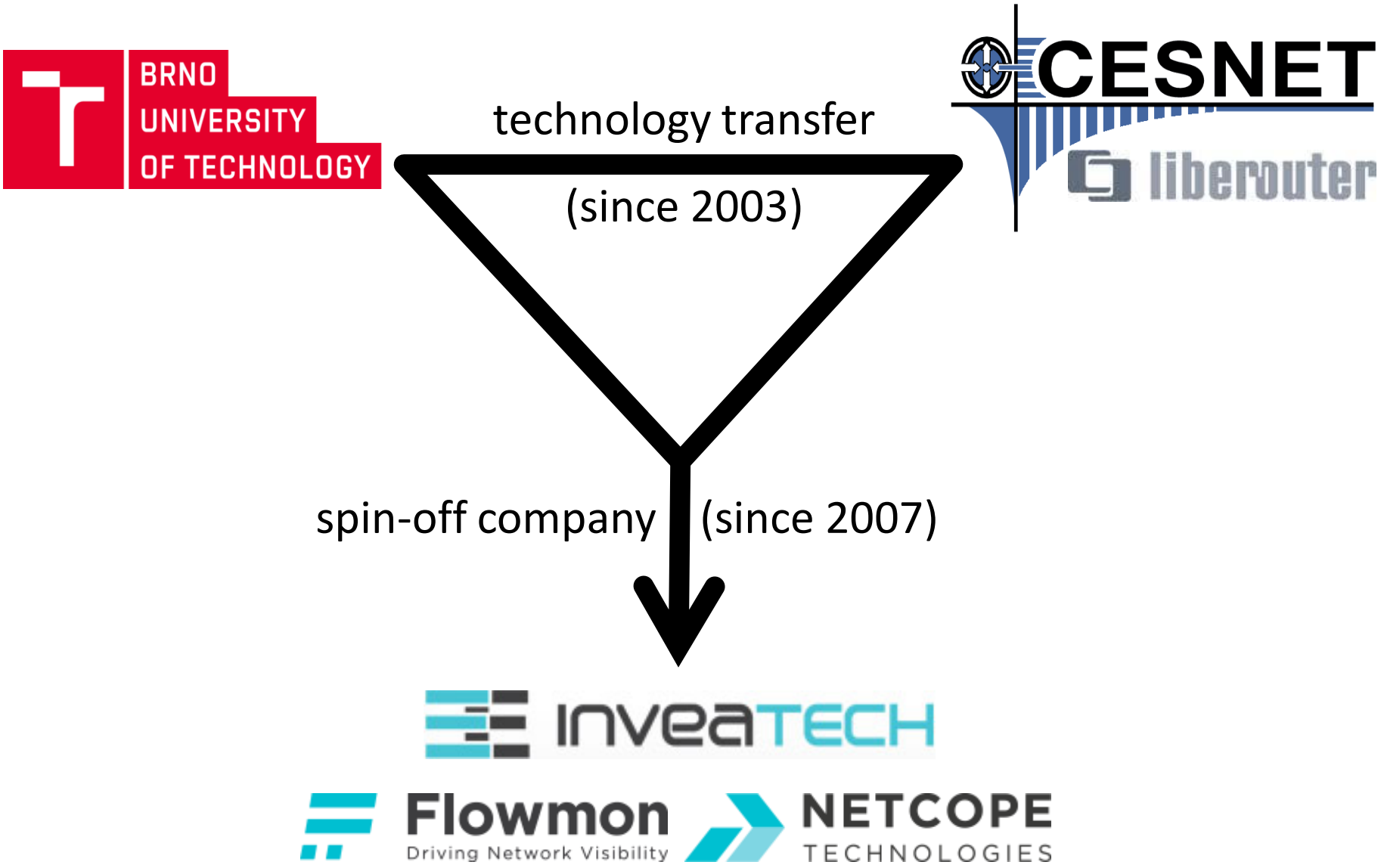
- open-source - https://github.com/CESNET/NEMEA

NETCONF and YANG

- development of tools for **full remote control** of our devices
  - in **cooperation with IETF**'s NETCONF & NETMOD groups
- **libyang -** YANG parser and validator with API in C
- **libnetconf -** NETCONF protocol implementation for Linux
  - generic client-server communication API written in C
  - device data modeling - v1 uses XML, v2 uses YANG
- **Netopeer -** set of applications with NETCONF protocol
  - implementations of server, clients (webGUI or CLI) and more

https://github.com/CESNET/{libyang,libnetconf,libnetconf2,netopeer}

Cooperation (National)



Cooperation (National)

- **Best Cooperation of the Year**
  - project TA03010561: *Distributed System for Complex Monitoring of High-Speed Networks*

- highest national research award **Czech Head**, **Industry award**
  - world's first 100 Gbps Ethernet interface card

Cooperation (International)

- University of Twente, **DACS** group
  - network monitoring and intrusion detection

- University of Cambridge, **NetOS** group
  - packet classification/filtering and dynamic reconfiguration

- part of GÉANT network and projects
  - PROTECTIVE, Firewall on Demand

Cooperation (International)

- **BEBA** (BEhavioural BAsed forwarding) H2020 EU project
  - finished last week with **"Excellent"** rating

Summary

- direct access to a lot of high-speed network data
- high-performance production and test monitoring probes
  - reconfigurable FPGA acceleration cards and extensible SW
- collection, analysis and storage of flow data
  - flexible and modifiable open-source tools
  - large database of collected IPFIX flow records
- close connections with university and industry environment
  - years of experience with national and EU research projects

We are open to new cooperation possibilities!

Thank you for your attention!

More info:

- *https://www.liberouter.org/*
- *@liberouter*
- *kekely@cesnet.cz*