XENOFLOW: BRINGING A L3 LOAD BALANCER CLOSER TO THE NETWORK

ARCS 2026

Max Schrötter, Sten Heimbrodt, Bettina Schnor
max.schroetter@uni-potsdam.de

Institute of Computer Science
University of Potsdam

24. March 2026

MOTIVATION

- Load balancers are a critical component of modern data centers
- Software-based load balancers suffer from **kernel network stack overhead**
- **eBPF-based** load balancers Facebook's Katran or **DPDK-based** load balancers dpvs improve performance but still do **not reach line rate** of current NICs
- NVIDIA advertises the NIC subsystem (100 Gbit/s – 400 Gbit/s) can "**accelerate data path at line rate**"

RQ: Can a load balancer be implemented using the BlueField DOCA Flow API ?
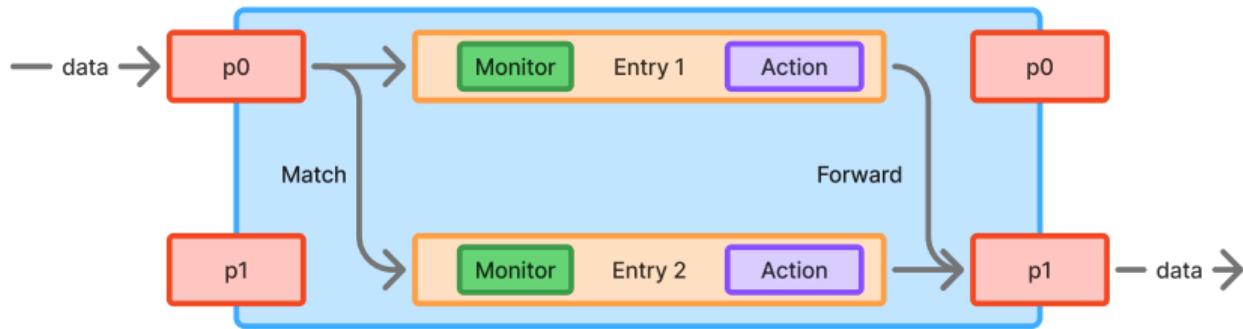
BLUEFIELD-3
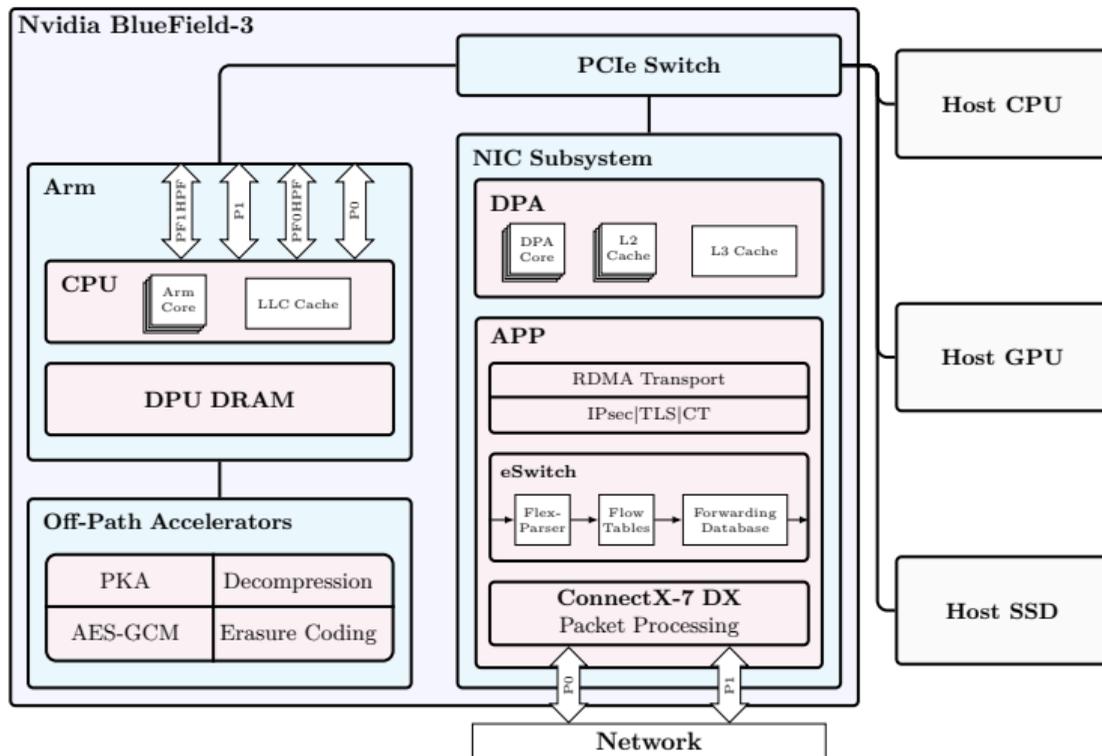
- **related work (Laconic)** only partially offloads flow processing, still **relying on BlueField's offpath ARM cores** for balancing incoming flows

XENOFLOW 0.1

- is a **prototype** to evaluate the capabilities of the BlueField-3
- single root pipe, two entries (backends),
- balances flows using **simple match** on the last bit of the source IP address

APP: "**Programmable 64-128 packet processor** for flexible networking"

RQ: Can XenoFlow achieve line rate on the BlueField-3 ?

RQ: Can XenoFlow achieve line rate?

RQ: Does XenoFlow drop packets?

29 Bytes UPD payload
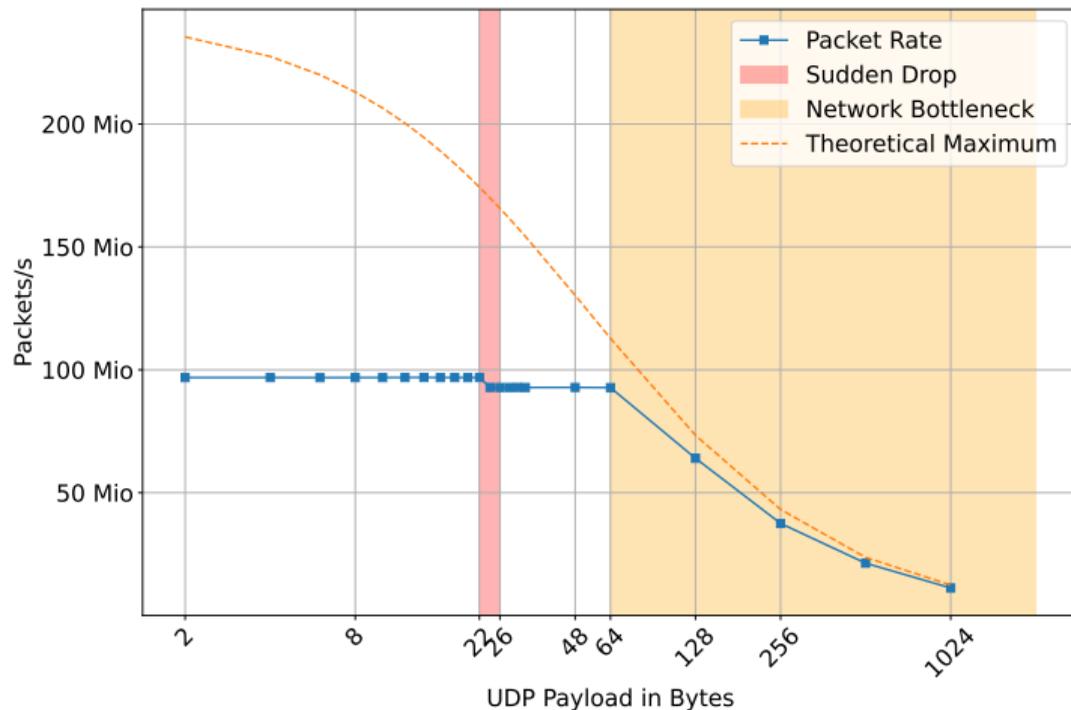
Compute Components

Resource	Host	ARM	DPA
CPU	2 × Intel Xeon Silver 4514Y	ARMv8.2+ A78 Hercules	RISC-V RV64IMAC(B)-USM
Cores	2 × 16	16	16
Threads	2 × 32	32	256
Frequency	3.4 GHz	2.0 GHz[1]	1.8 GHz
RAM	128 GB	32 GB	
L1D Cache	2 × 16 × 48 KB	16 × 64 KB	256 × 1 KB
L1I Cache	2 × 16 × 32 KB	16 × 64 KB	8 KB
L2 Cache	2 × 16 × 2 MB	16 × 512 KB	1.5 MB
L3 Cache	2 × 30 MB	16 MB	3 MB

APP: "**Programmable 64-128 packet processor** for flexible networking"

RQ: How does XenoFlow affect the latency?

- XenoFlow adds $5\,\mu s$[1] latency compared to the baseline without a load balancer
- comparable eBPF-based load balancers adds a latency of $22\,\mu s$[1]
- **XenoFlow achieves a load balancer latency reduction of 77%**

---
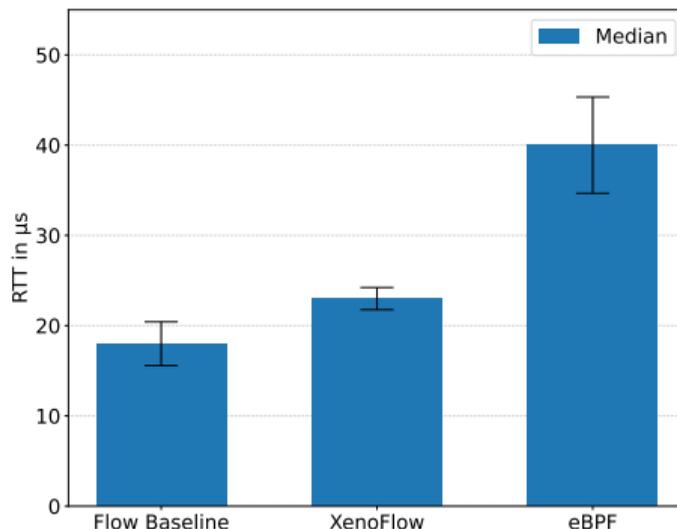
[1]Median of 900 measurements

RQ: Does the latency change with increasing load?
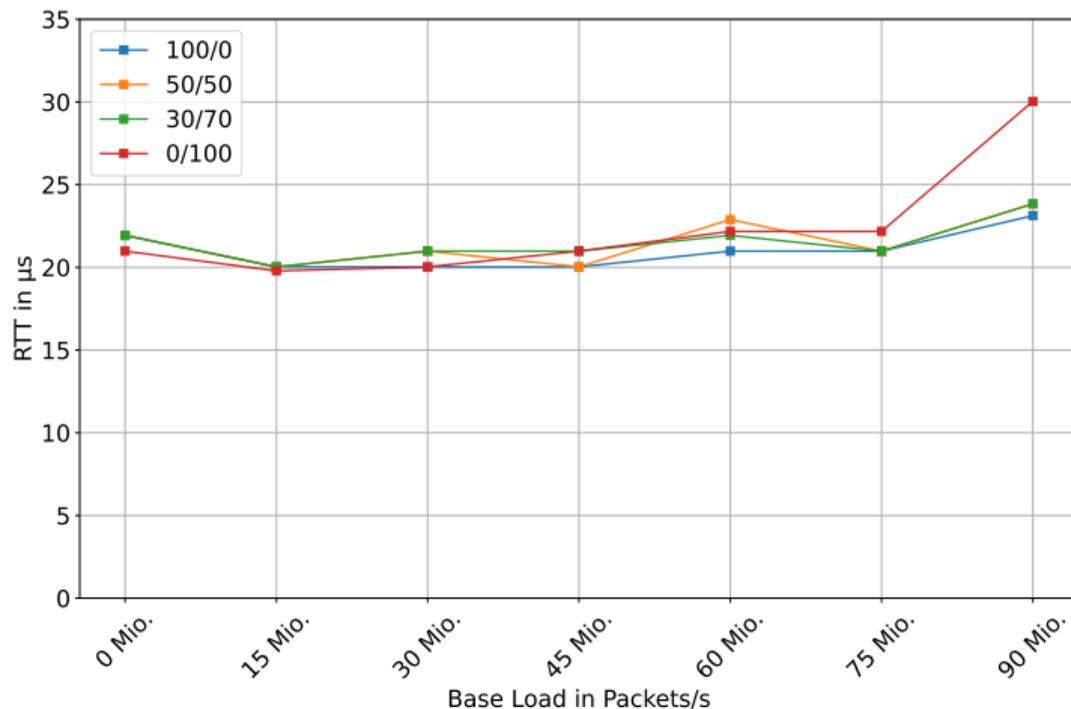
RQ: Can a load balancer be implemented using the BlueField DOCA Flow?

- Per Connection Consistency (PCC) is a key requirement for load balancers
- Approaches like Maglev (Google) use a two part strategy:
  - i) use consistent hashing to assign flows to backends
  - ii) use a flow table to ensure PCC for existing flows

- DOCA Flow has a median entry update latency of $249\,\mu s$
- DOCA Flow supports $8,192$ entries per pipe with a maximum of 505 pipes
  → $4,136,960$ entries in total

RQ: Can a load balancer be implemented using the BlueField DOCA Flow?

- Per Connection Consistency (PCC) is a key requirement for load balancers
- Approaches like Maglev (Google) use a two part strategy:
    i) use consistent hashing to assign flows to backends
    ii) use a flow table to ensure PCC for existing flows

- DOCA Flow has a median entry update latency of $249\,\mu s$
- DOCA Flow supports $8,192$ entries per pipe with a maximum of 505 pipes
  $\rightarrow 4,136,960$ entries in total

i) feasible　　　　　　**ii) needs further research**

CONCLUSION & FUTURE WORK

XenoFlow

- XenoFlow shows the potential[2] of the BlueField-3
- XenoFlow achieves **near-line rate** for packets > 64 bytes
- XenoFlow has a 77% **lower load balancer latency** compared to eBPF-based load balancers

- Extend XenoFlow and evaluate strategies to achieve PCC for existing flows
- Since recent DOCA Flow versions (v3.0.0) multiple algorithms are supported for the hash pipe.

---

[2] despite limited visibility into NVIDIAs closed-source hardware architecture and evolving documentation