

# Live in the Express Lane

Patrick Jahnke<sup>\*§</sup>, Vincent Riesop<sup>§</sup>, Pierre-Louis Roman<sup>†</sup>, Pavel Chuprikov<sup>†</sup>, and Patrick Eugster<sup>\* +▲</sup>

<sup>\*</sup>TU Darmstadt, <sup>§</sup>SAP, <sup>†</sup>Università della Svizzera italiana, <sup>▲</sup>Purdue University

# Crucial challenges of coordination tasks

---



- Transition in the last decade
  - Conceived as distributed (cloud-based) application
  - Data center (DC) interferences prevent predictability
  - Weak synchrony assumptions to guarantee safe execution
- Mitigation of interference in DCs to accelerate distributed systems (DSs)
  - Awareness of timely sensitive interactions
  - Tight upper bounds required
  - Foundation to increase performance of DS coordination tasks

# Related work



- Low latency [1, 2, 3, 4]
  - Generic approaches
  - 99<sup>th</sup> percentile
  - No process response time
- In network processing [5, 6]
  - Specific to particular service
  - Specific hardware required
  - No process response time

*Communication links are reliable but asynchronous* [7]

- [1] M. P. Grosvenor, M. Schwarzkopf, I. Gog, R. N. M. Watson, A. W. Moore, S. Hand, and J. Crowcroft. *Queues Don'T Matter when You Can JUMP Them!* In: USENIX NSDI. 2015, pp. 1–14.
- [2] B. Montazeri, Y. Li, M. Alizadeh, and J. K. Ousterhout. *Homa: A Receiver-Driven Low-Latency Transport Protocol Using Network Priorities*. In: ACM SIGCOMM. 2018, pp. 221–235.
- [3] J. Perry, A. Ousterhout, H. Balakrishnan, D. Shah, and H. Fugal. *Fastpass: A Centralized "Zero-Queue" Datacenter Network*. In: ACM SIGCOMM. 2014, pp. 307– 318.
- [4] G. Prekas, M. Kogias, and E. Bugnion. *ZygOS: Achieving Low Tail Latency for Microsecond-Scale Networked Tasks*. In: ACM SOSP. 2017, pp. 325–341.
- [5] H. T. Dang, D. Sciascia, M. Canini, F. Pedone, and R. Soulé. *NetPaxos: Consensus at Network Speed*. In: SIGCOMM SOSR. 2015, pp 5:1–5:7.
- [6] Z. István, D. Sidler, G. Alonso, and M. Vukolic. *Consensus in a Box: Inexpensive Coordination in Hardware*. In: USENIX NSDI. 2016, pp 425–438.
- [7] A. Basu, B. Charron-Bost, and S. Toueg. *Simulating Reliable Links with Unreliable Links in the Presence of Process Crashes*. In: WDAG. 1996, pp 105-122.

# Features of X-Lane

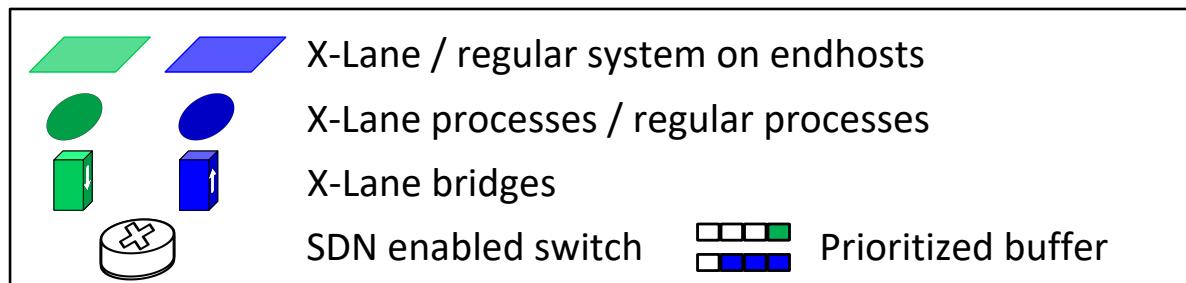
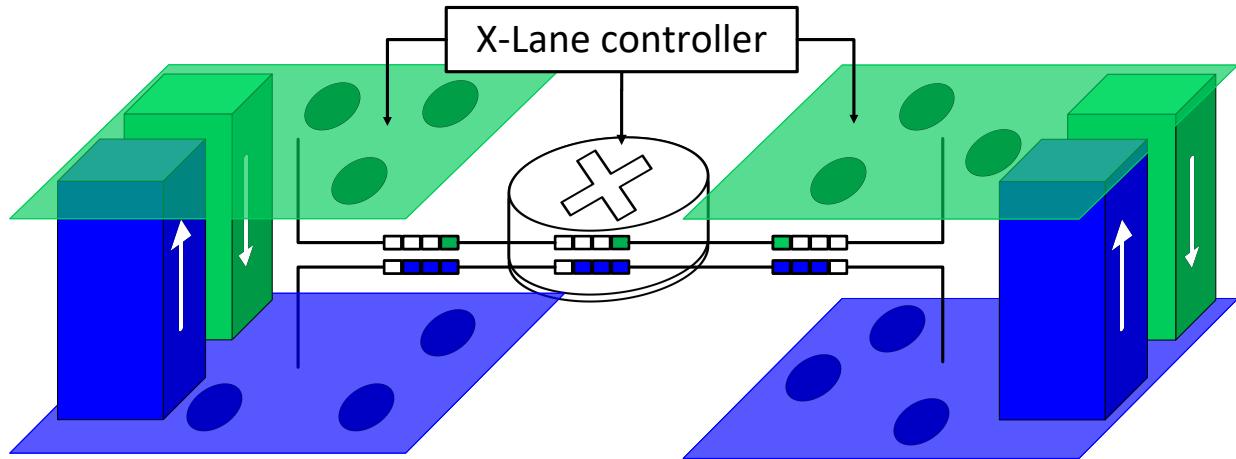
---



- Interference free environment
  - Bounded communication latency / jitter
  - Bounded processing latency / jitter
- Generic design
  - Supports multiple coordination protocols
  - Commodity and specialized hardware



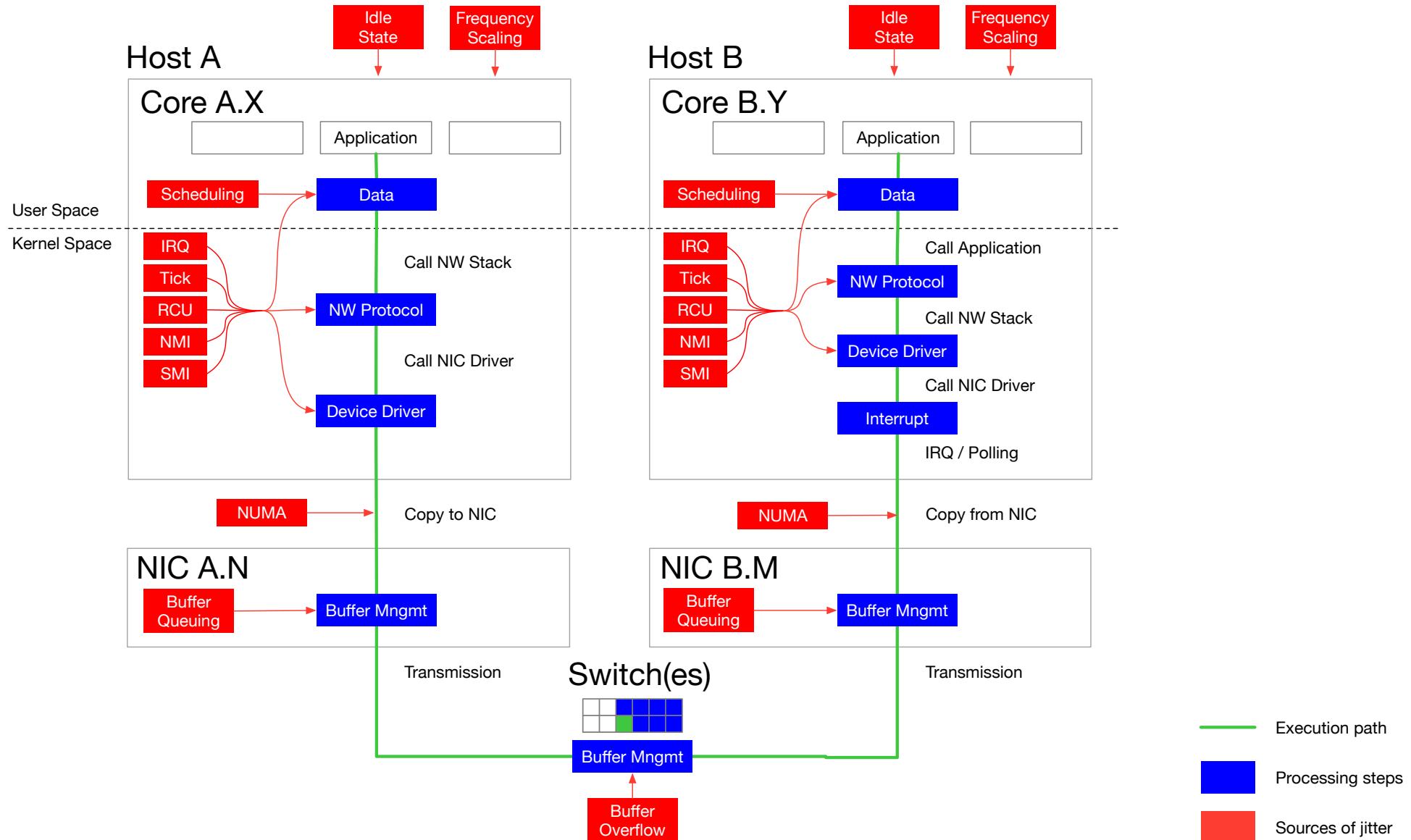
# Separation between X-Lane & regular system



- X-Lane isolated from “regular system”
- Prioritize X-Lane packets to prevent losses
- Process communication over bridges

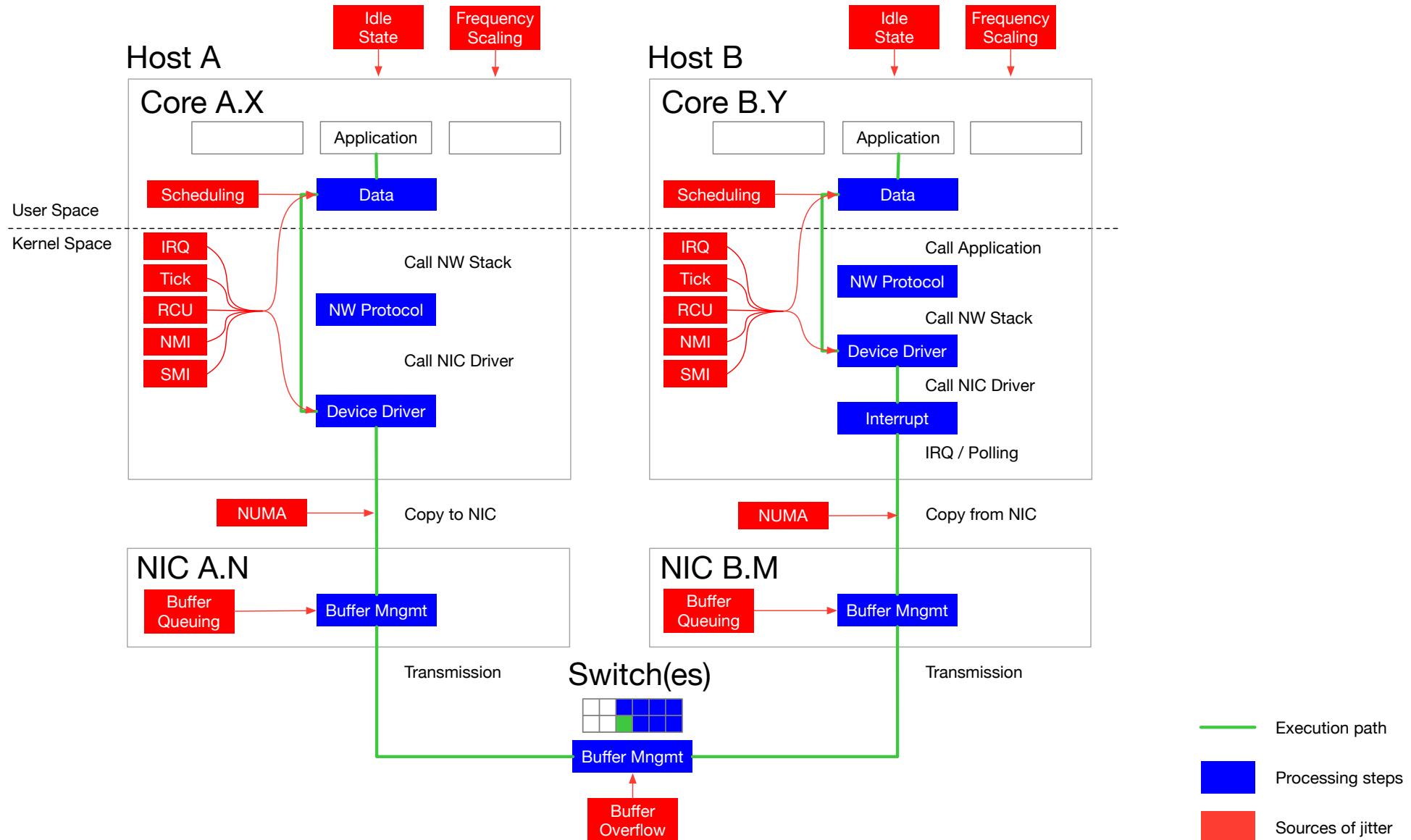


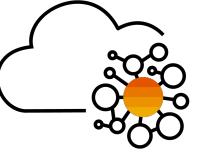
# Usual workflow for a packet



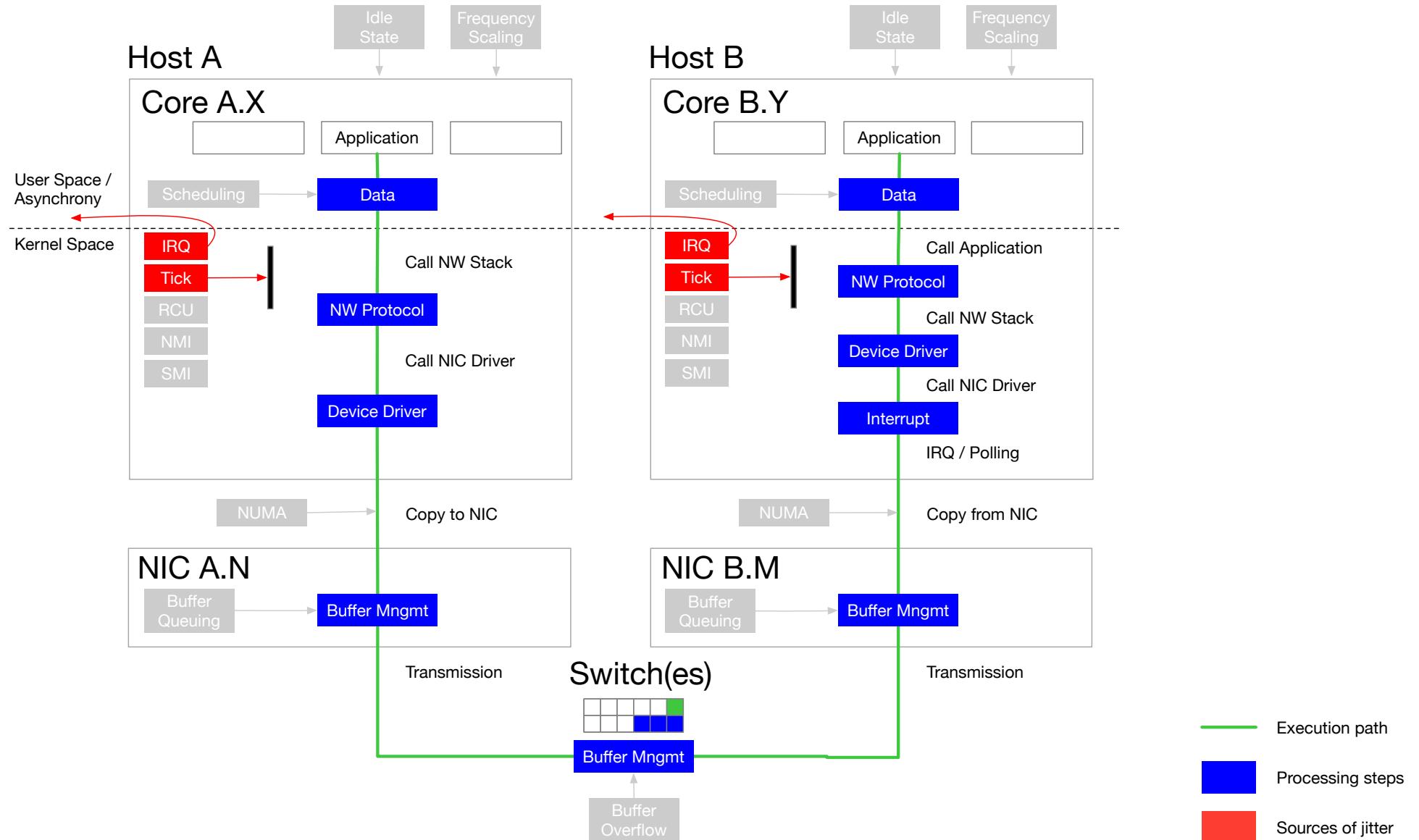


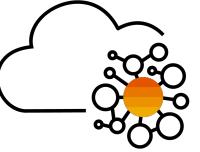
# Usual workflow for a packet



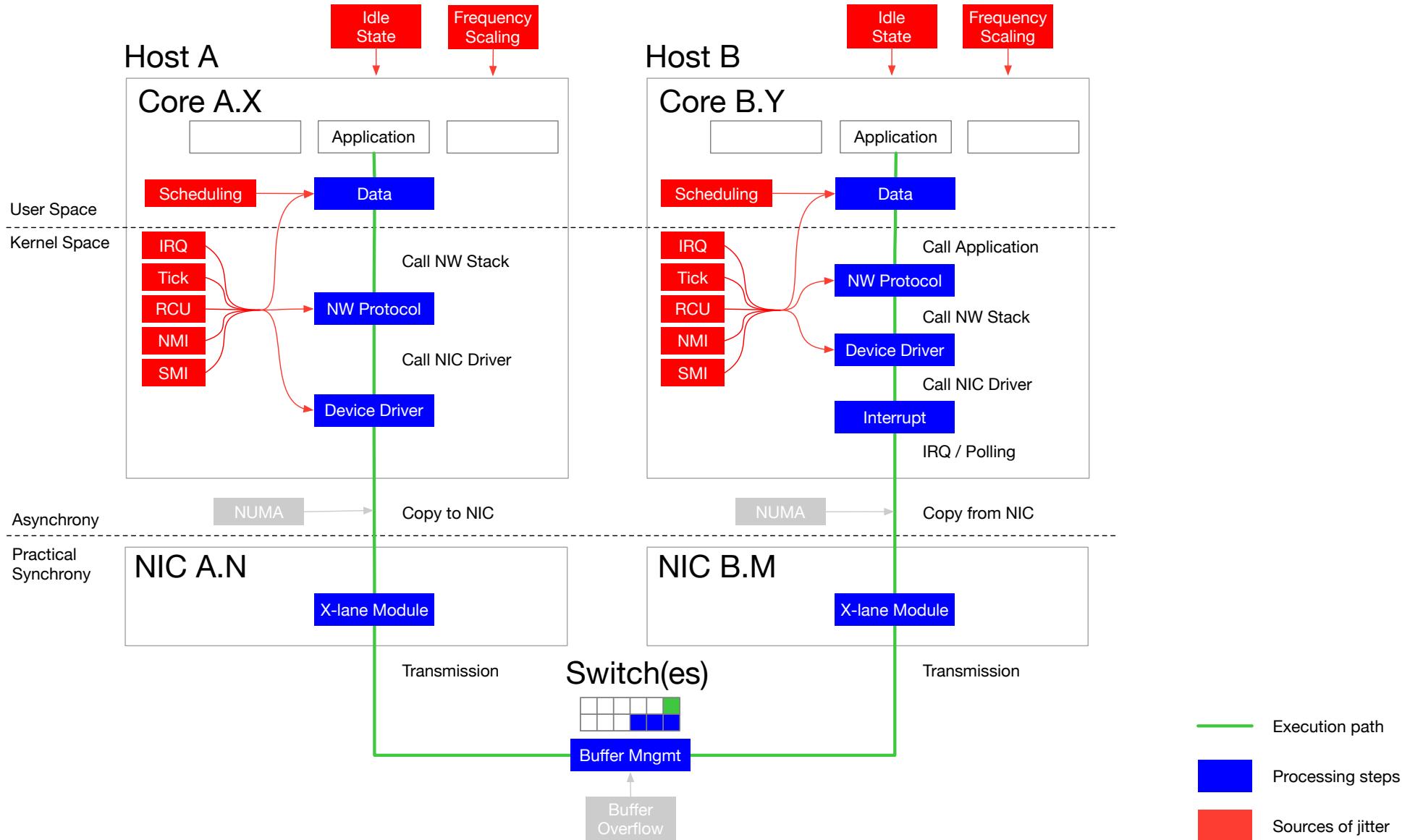


# X-Lane workflow for commodity hardware

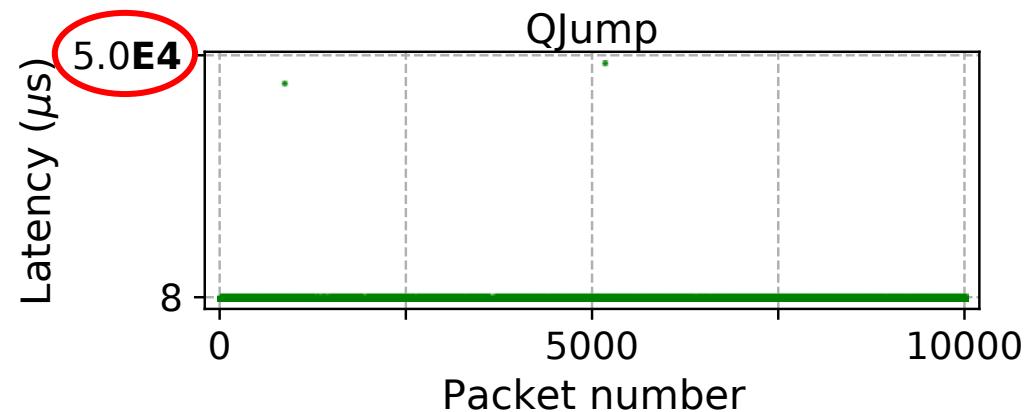
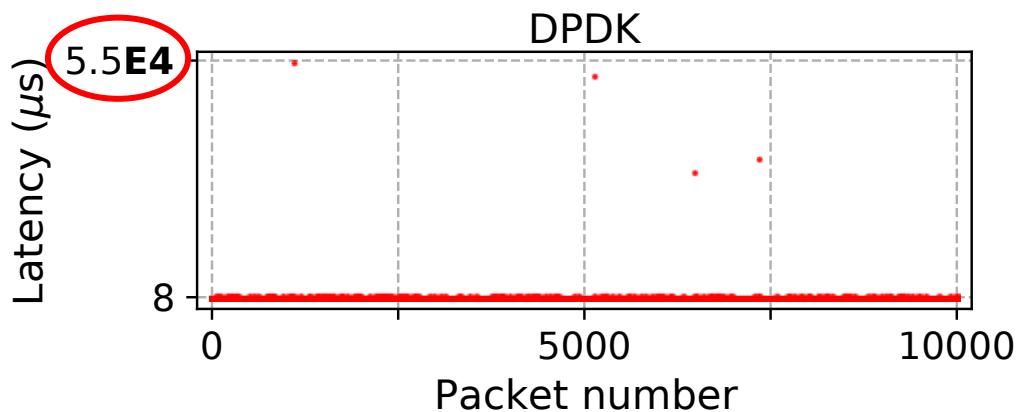
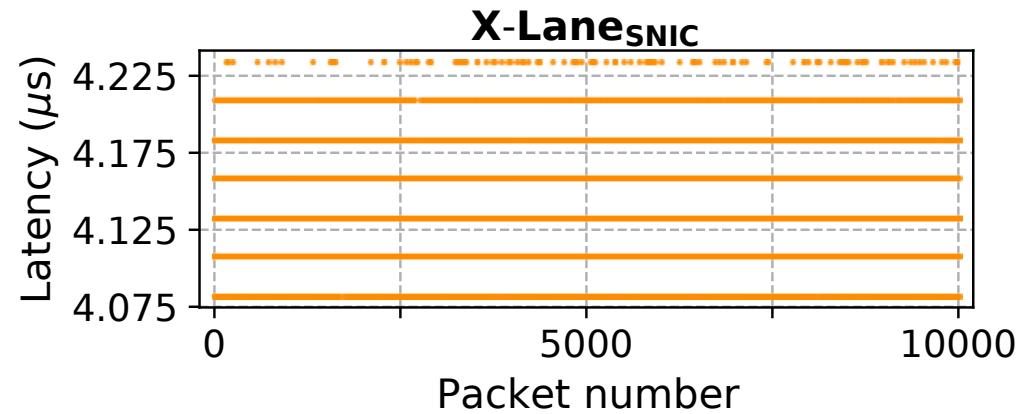
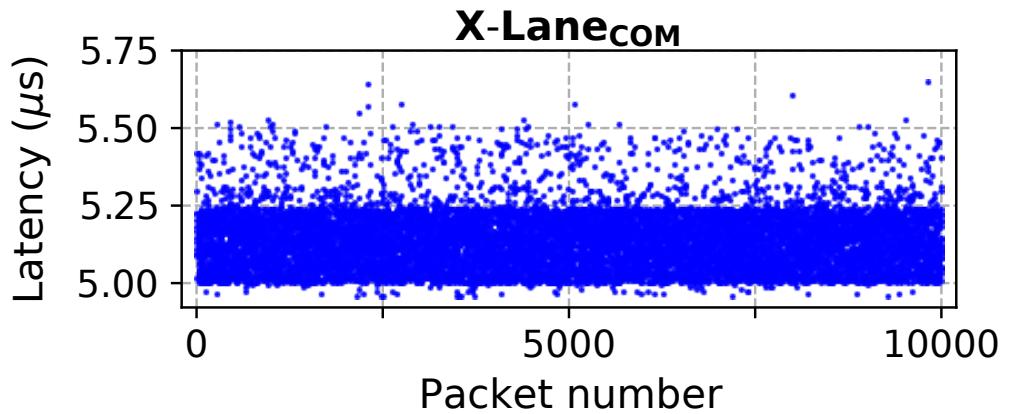




# X-Lane workflow with smartNICs

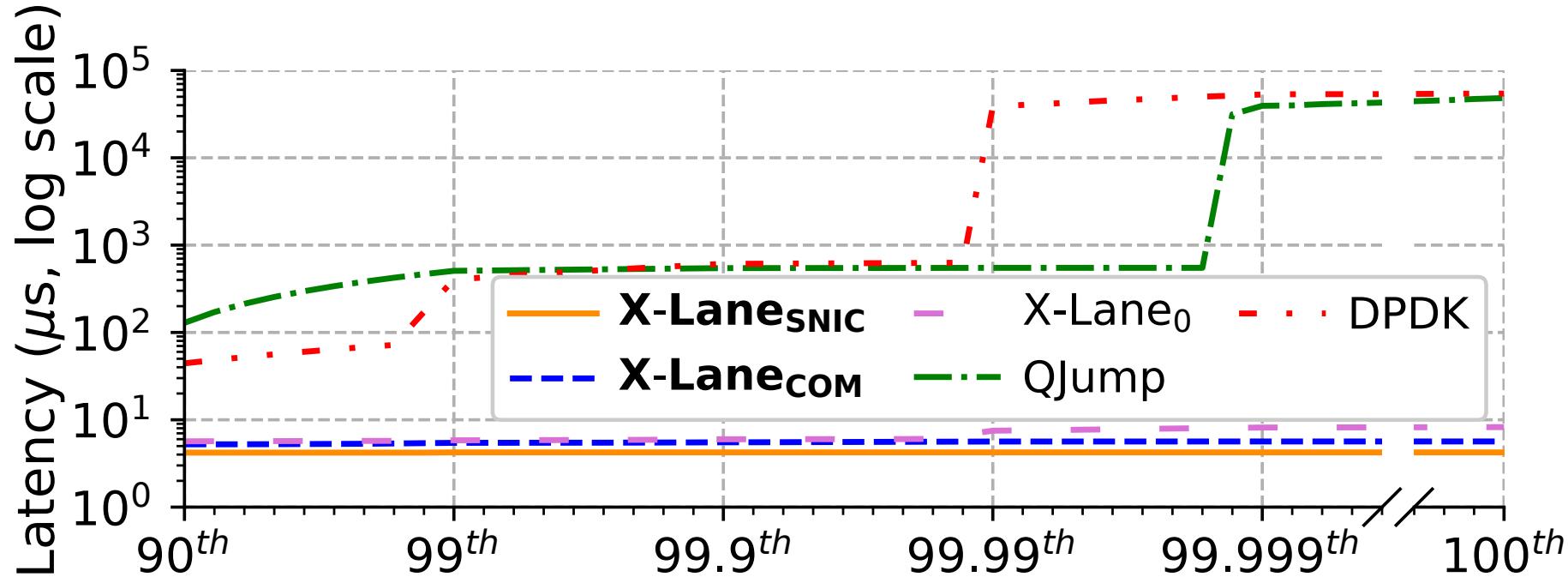


# Latency and jitter for DPDK, QJump and X-Lane

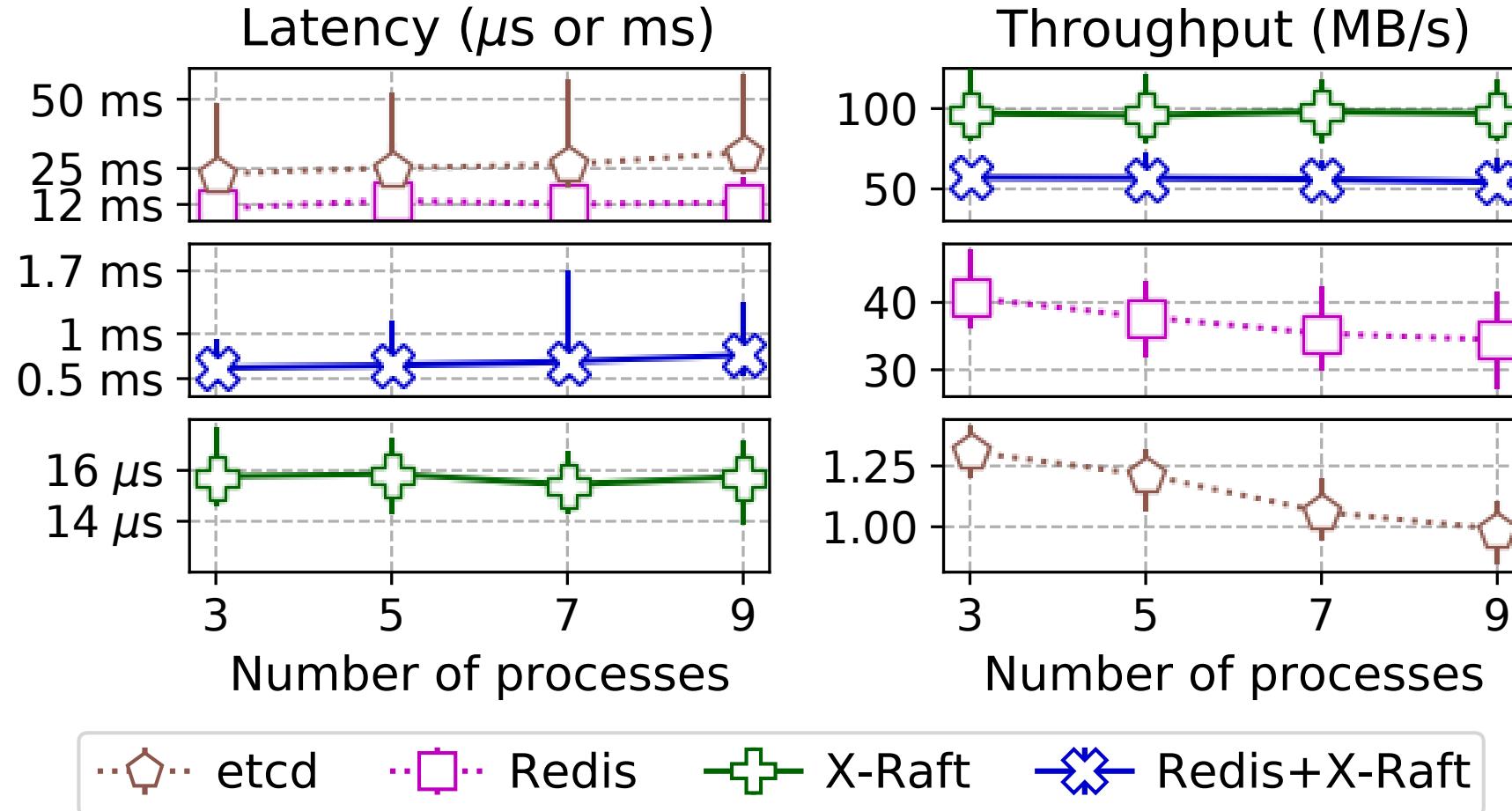
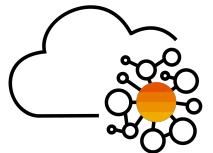




# Tail latency over 21 days



# Latency and throughput results for Raft implementations



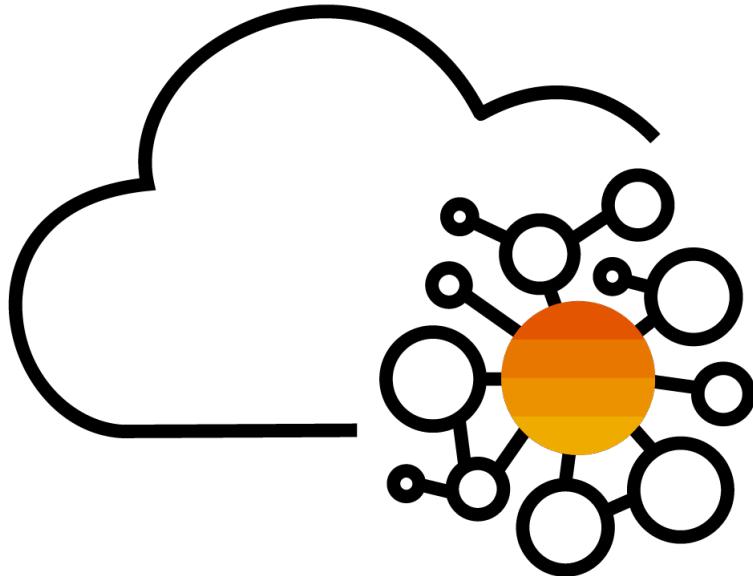
# Conclusion

---



- Low latency and jitter for coordination interaction
- X-Lane isolated from regular system
- Generic system design
- Commodity software / hardware and smartNIC support

**Interested?**



**Questions?**

**Thank YOU!**