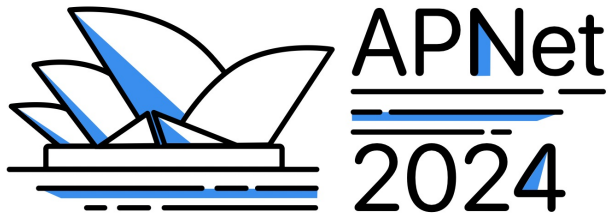


An Integrated Solution for High-efficiency In-band Network Telemetry

Xinxin Xiong , Yi Xie , Xiaochou Chen ✉ ,
Shaojie Zheng , Wenju Huang , Jiahao Feng

August 4 , 2024

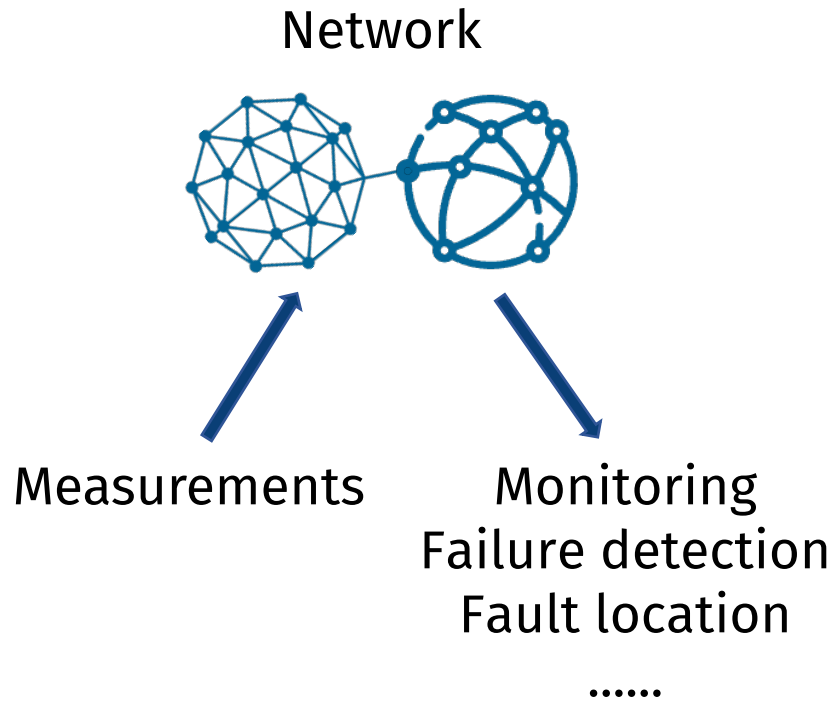
Network Measurement

New features of data center networks

- High speed
- Large scale
- Traffic unpredictable

Traditional network measurements **Focus on end to end**

- Active: PING, Traceroute **Additional traffic overhead**
- Passive: NetFlow, sFlow, IPFIX **Local state information**
- Hybrid: AM-PM



Software-defined network

Programmable data-plane

In-band Network Telemetry

In-band Network Telemetry (INT)

INT Specification (P4.org)

- Source: Insert INT instructions (Node ID, Hop latency, Ingress timestamp…) and metadata
- Transit: Insert metadata and forward packets
- Sink: Insert metadata and separate packets and form INT reports
- Collector: Aggregate telemetry data

Advantages:

- Packet level detection
- Accurate and fine-grained measurement
- Programmability

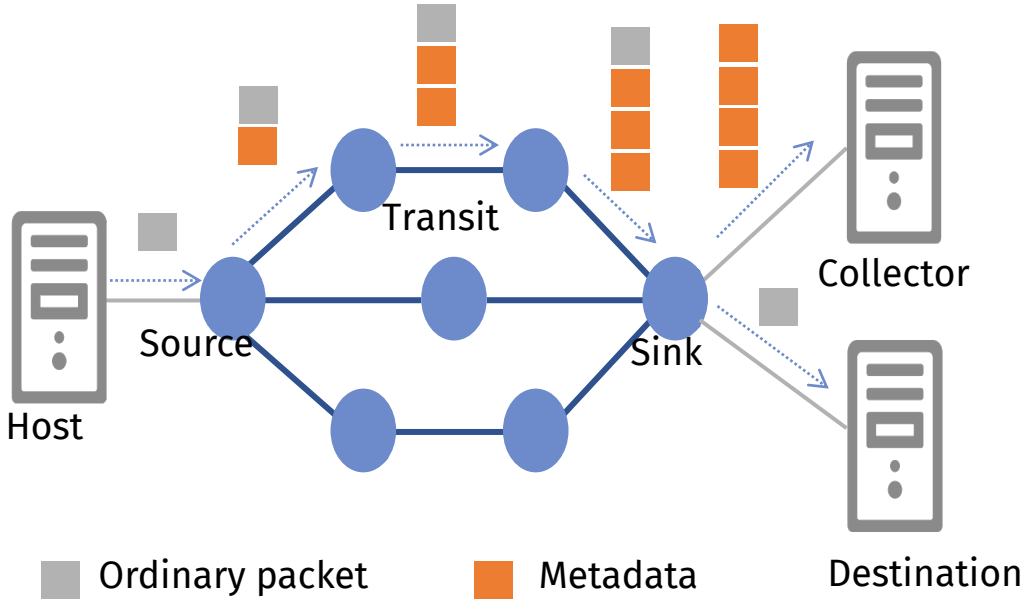
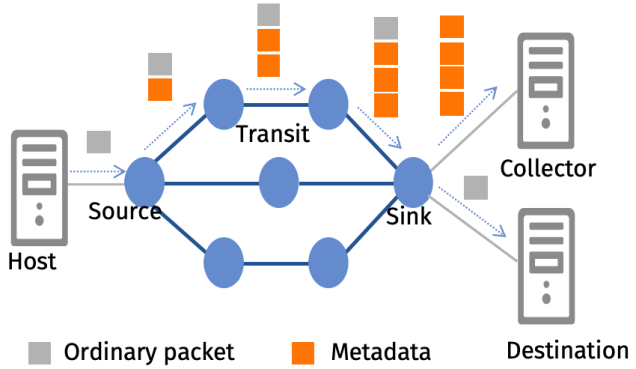
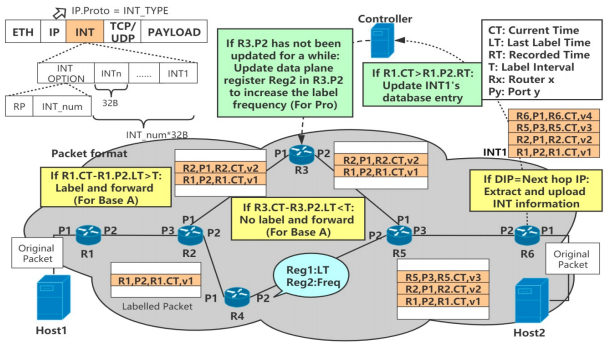
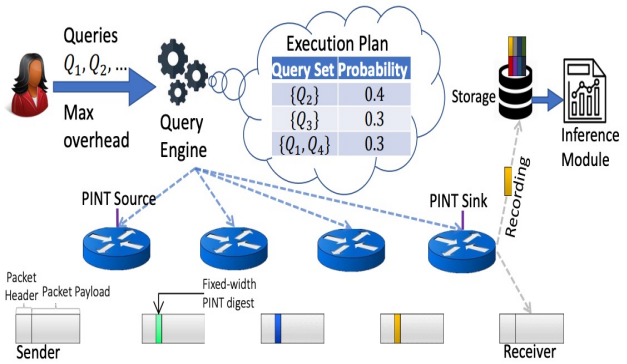


Figure 1: Overview of INT (MD-type)

Existing INT schemes have limitations







P4 INT (Specification 2020) INT-Label (INFOCOM 2021) PINT(SIGCOMM 2020)

Telemetry overhead

- Packet size increasing with the number of nodes

Slow report processing

- Rapid accumulation of telemetry reports may give rise to data lakes in the collector

Hardware-level acceleration

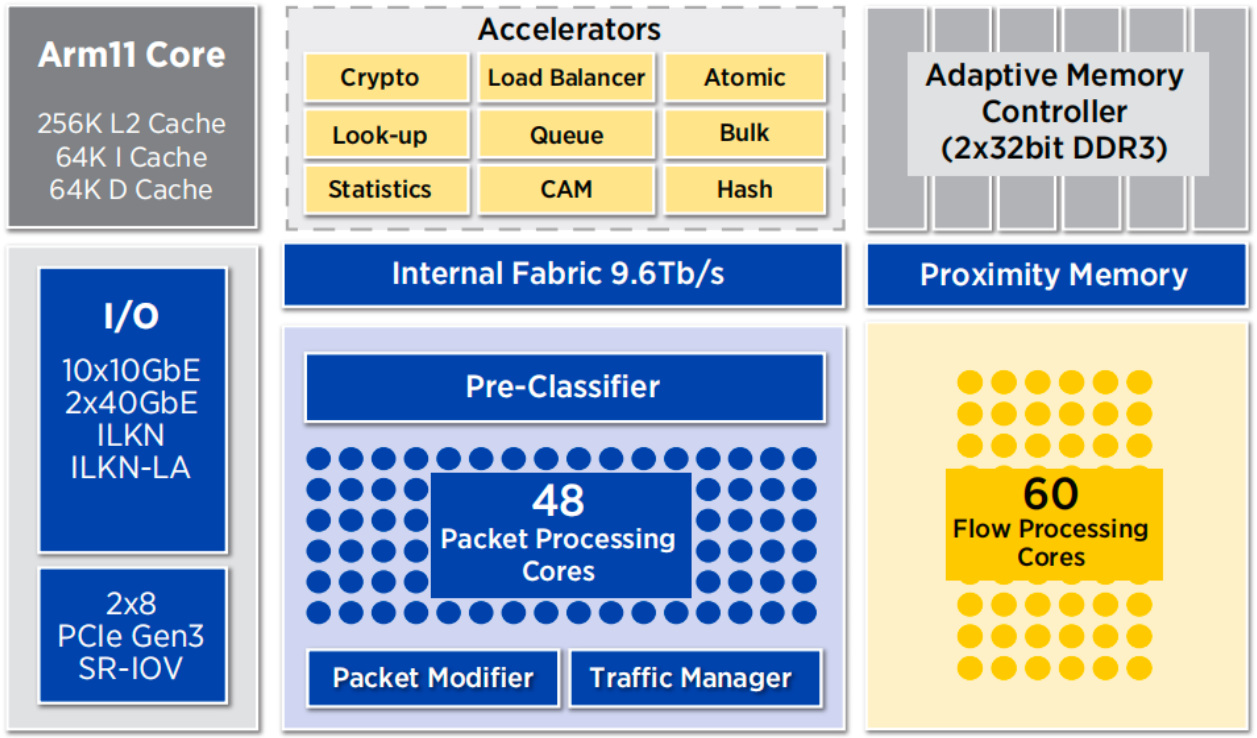
Offload CPU

→ **SmartNIC**

SmartNIC

Features

- Multi-threaded and multi-core stream processors (Processing packets in parallel)
- Programmability (P4, MicroC)
- Offloads packet processing from CPU



NFP-4000 Flow Processor Block Diagram

[5] Netronome Systems, Inc. 2020. *Netronome NFP-4000 Flow Processor*. https://d3ncevyc0dfnh8.cloudfront.net/media/documents/PB_NFP-4000-7-20.pdf

Our Integrated Solution for High-efficiency INT

- Leverage the programmability to design Stateful-INT (SF-INT) : node's register + store-and-forward actions
- Leverage a SmartNIC-equipped collector to process INT reports

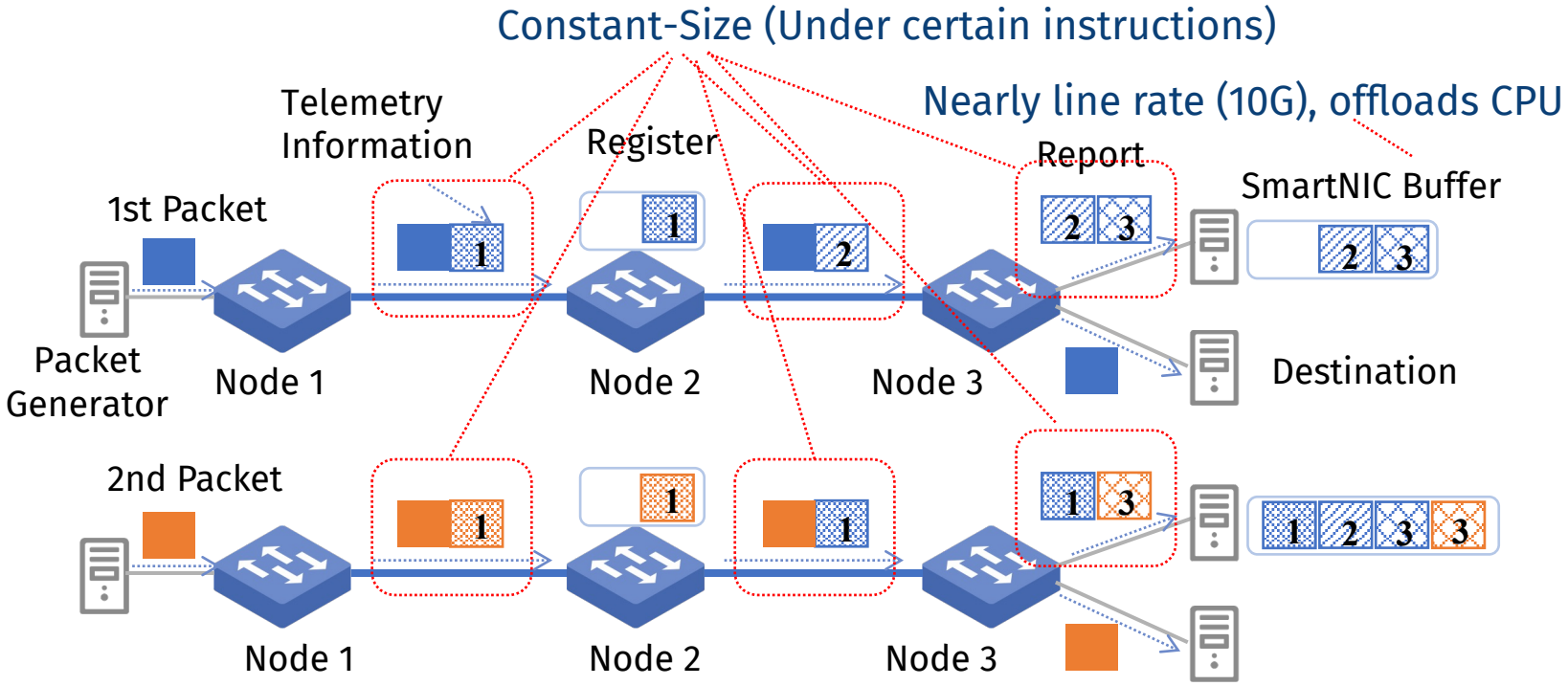


Figure 3: SF-INT Prototype: store-and-forward process

Table 1: SF-INT's Rules of store-and-forward actions

Rule	State		Action	
	Arrival packet	Register store	Forward packet	Register update
1	M(A)	H(S)	H(S)	L(A)
2	M(A)	L(S)	M(C)	H(A)
3	M(A)	E()	M(C)	H(A)
4	H(A)	H(S)	H(S)	H(A)
5	H(A)	L(S)	L(S)	L(A)
6	H(A)	E()	H(A)	E()
7	L(A)	H(S)	H(S)	L(A)
8	L(A)	L(S)	L(S)	L(A)
9	L(A)	E()	L(A)	E()

- E(): The register is empty.
- A: Metadata in the arrival packet
- S: Metadata stored in the node register
- C: Metadata collected at the current node
- Priorities in descending order: high (H), medium (M), and low (L).

Packet Formats of SF-INT

Compatible with IPv6 packets

- Utilize the hop-by-hop option header in IPv6

SF-INT Packet

- Design INT-Option header
- Design SF-INT header and metadata stack

SF-INT Report

- Insert headers and metadata that stripped from the SF-INT packet

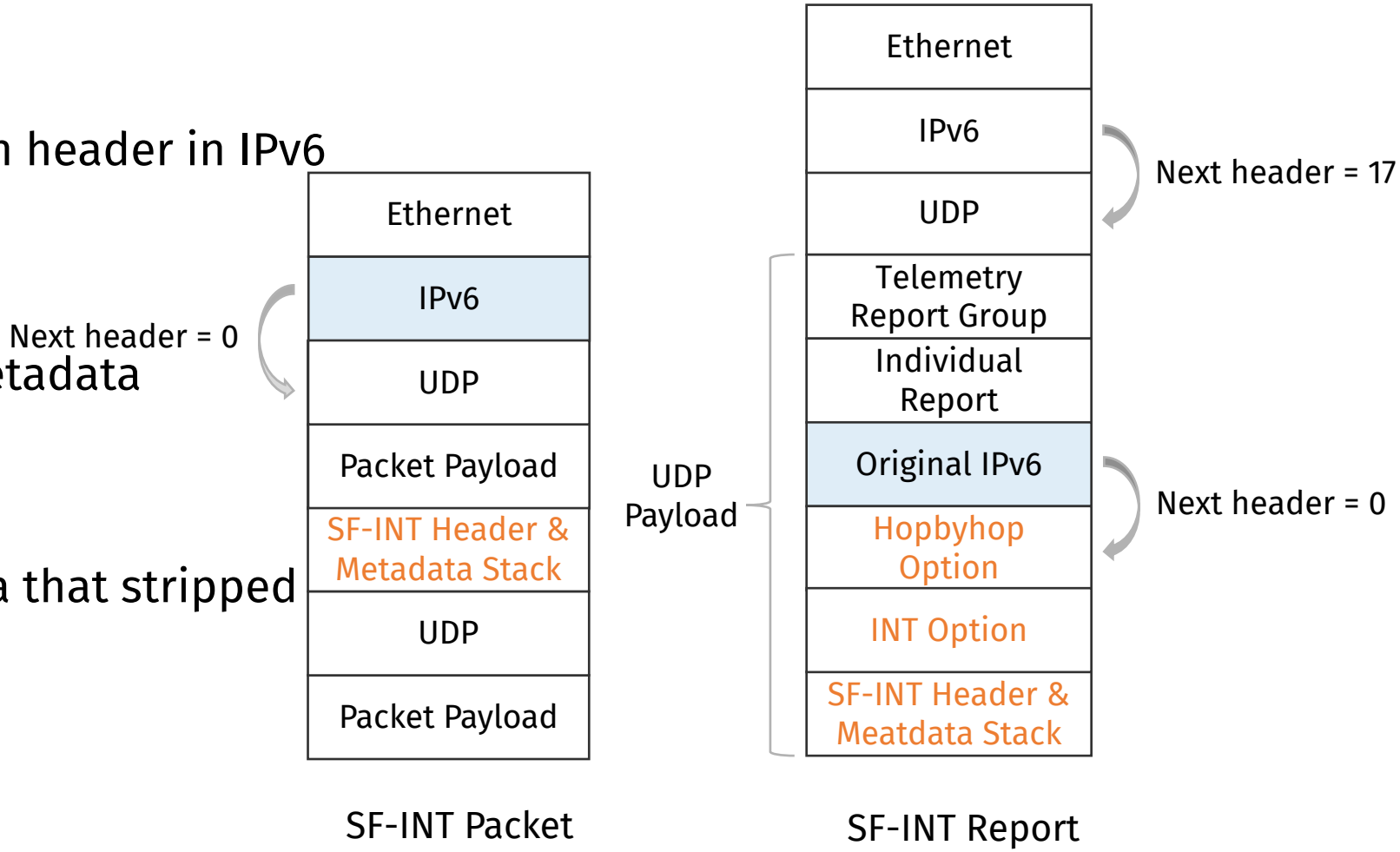


Figure 4: Packet Formats of SF-INT

Design of SF-INT Header

- ***C*** **:** SF-INT packet flag **U :** Represent the emergency mode (Compatible with standard INT)
- ***ST*** **:** Priority, High (H), Medium (M), and Low(L)
- ***packetID*** **:** A marker for packets group

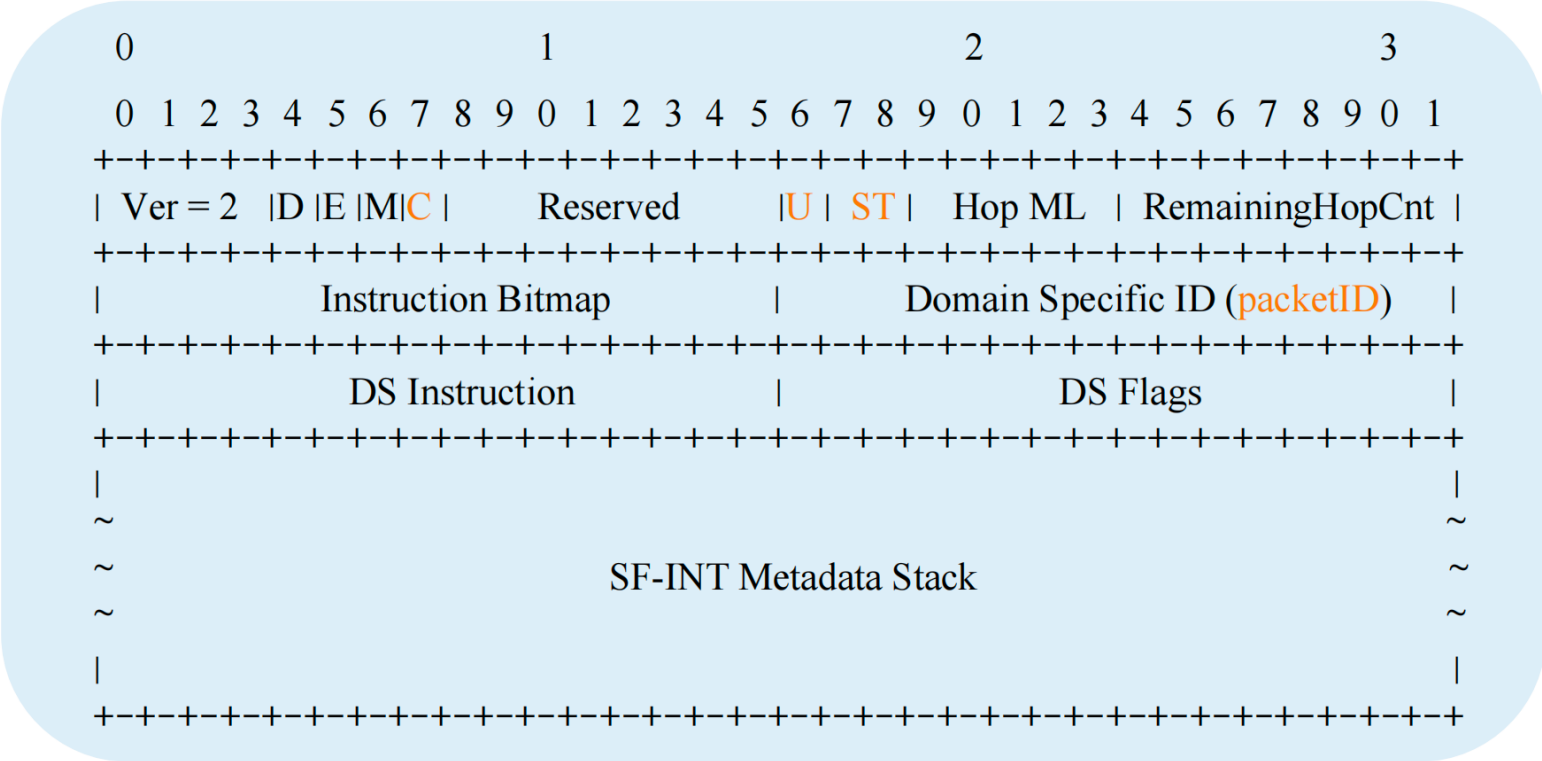


Figure 2: SF-INT header and metadata stack

SmartNIC Processes Reports

Processing Flow

- P4 module:

  Parse SF-INT reports

- Micro-C module:

  Extract telemetry data

- Host module:

  Read telemetry data

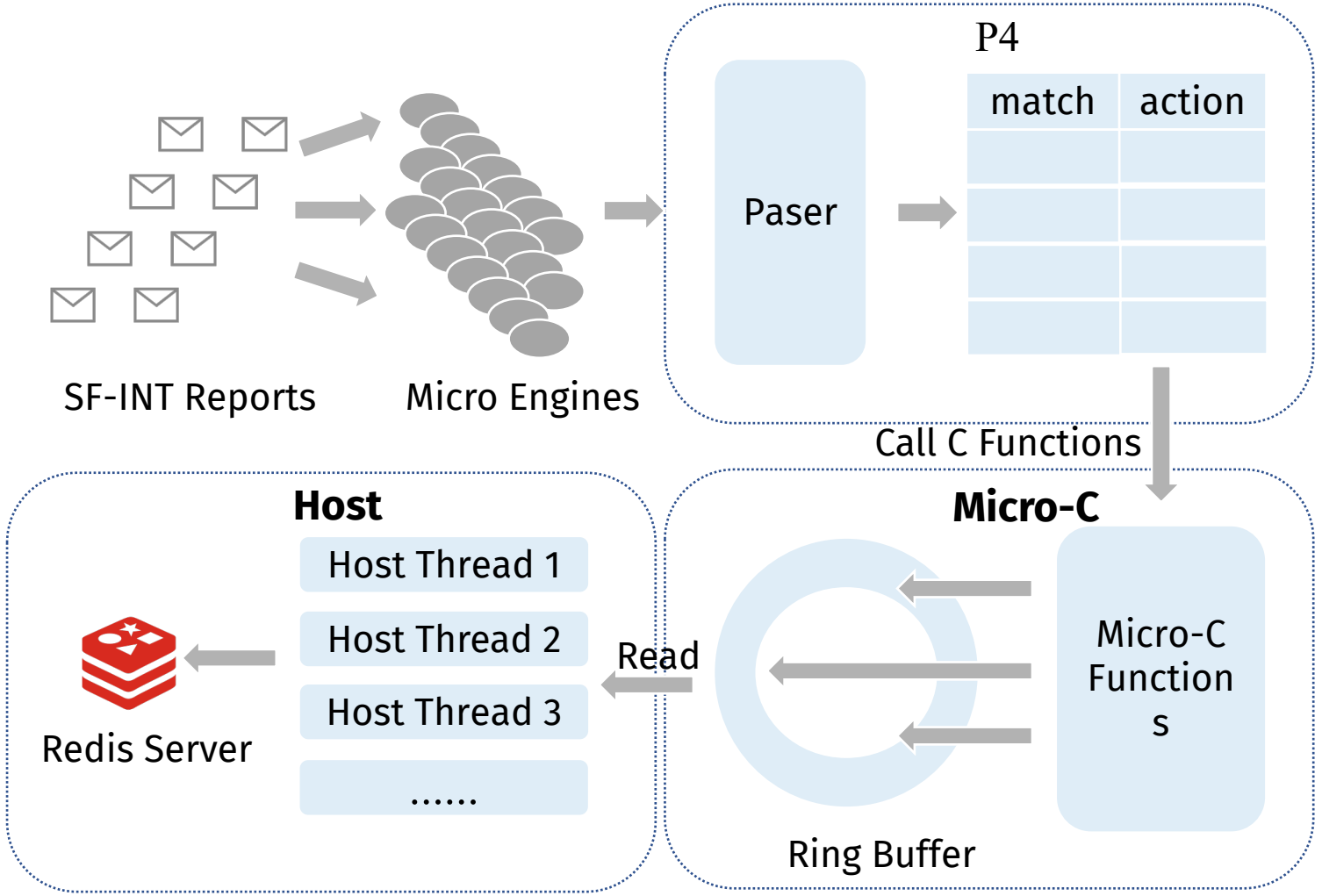


Figure 5: SmartNIC's Processing of SF-INT Reports

Build a Real Testbed

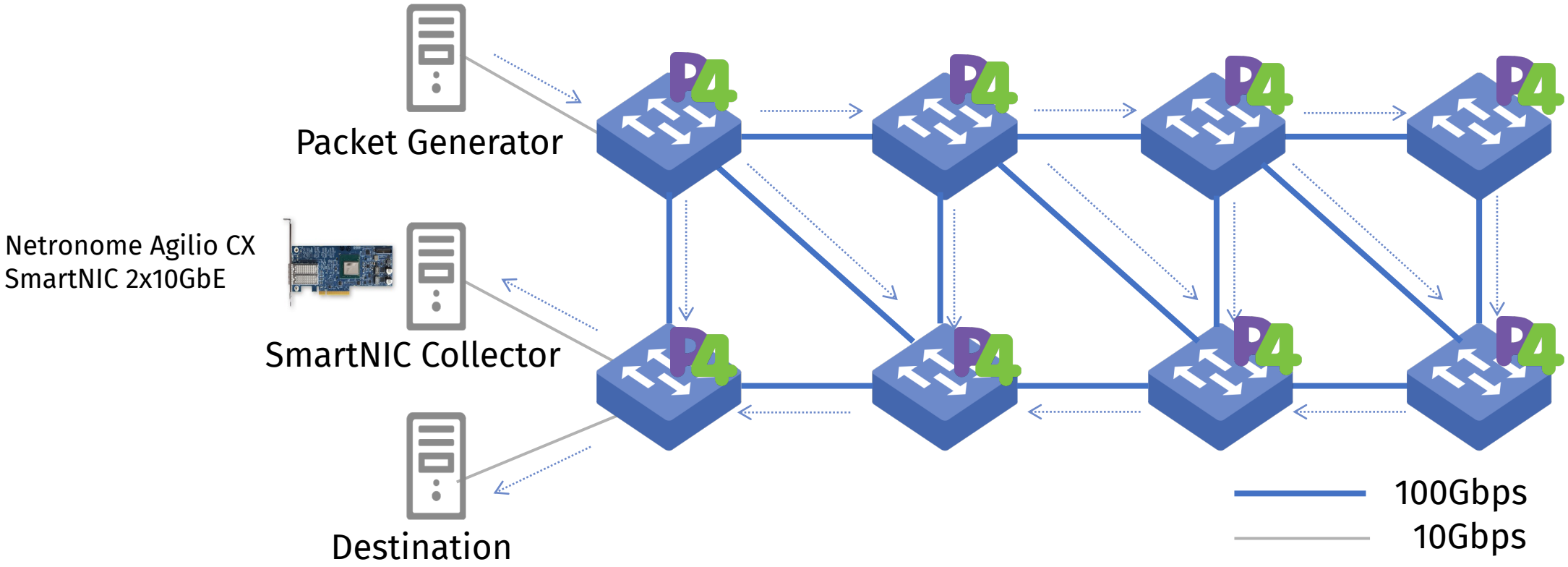


Figure 6: Testbed Topology

Intel Tofino P4 programmable switches(2 pipelines)

Ubuntu: 24 CPU Cores, 128G Memory

Low Overhead

Compared with INT, SF-INT packets Keep constant size during the whole path

- 5 instructions

  Node ID, Hop latency, Ingress timestamp, Egress timestamp, Level 2 Ingress & Egress interface ID
- 8 network nodes

➢ INT packet size grows with node

➢ SF-INT packet size keeps constant

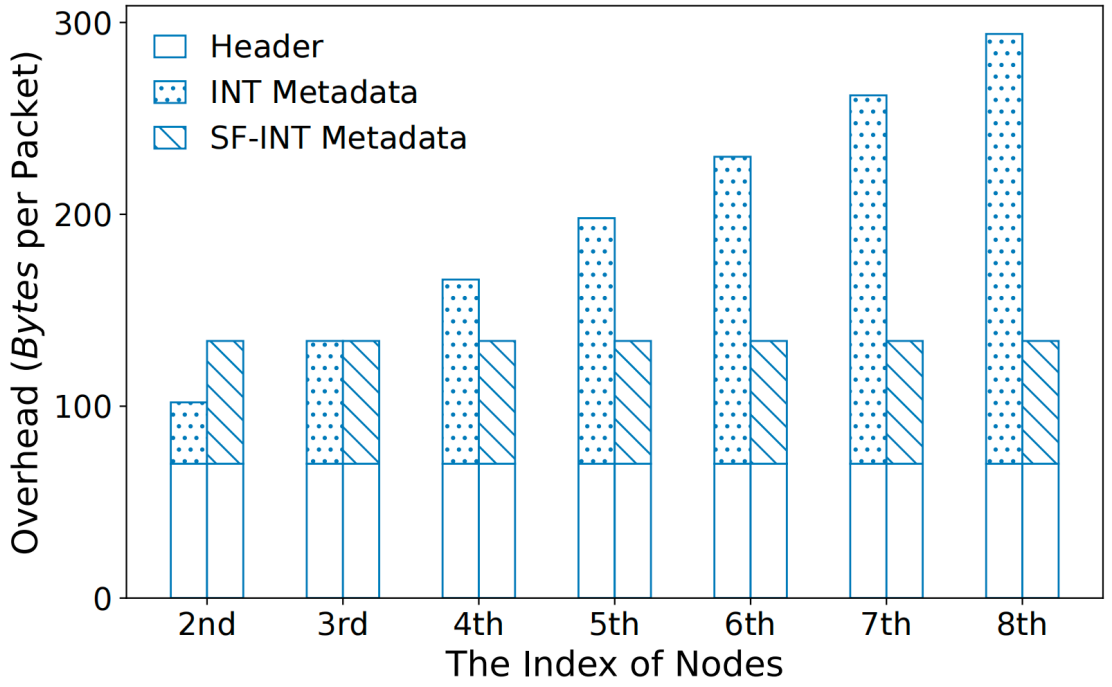


Figure 7: Overhead comparison of SF-INT and INT

High Throughput and Slight Latency

Throughput

- Maximum 9.13 Gbps (under 10G link)
- Process the generated telemetry reports over a 10G link nearly the line rate

Slight Latency

- Average latency of 9.3 us , 99% of latency values below the average

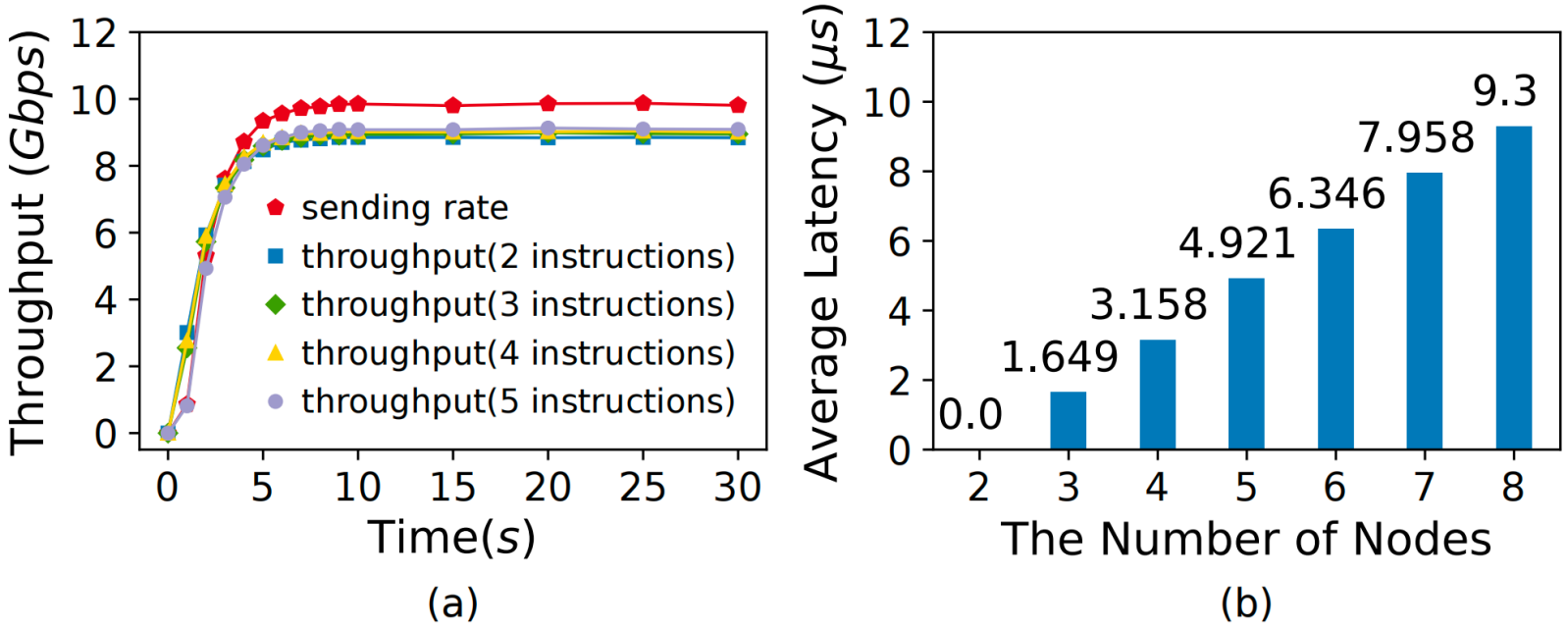


Figure 8: Perfomrance of our integrated solution

SmartNIC Processes Reports Efficiently

- 7.13 Mpps in the case of 2 telemetry instructions (Maximum rate)
- Do not decrease with the number of nodes
- Compare output rate with the Input rate (Almost the same)

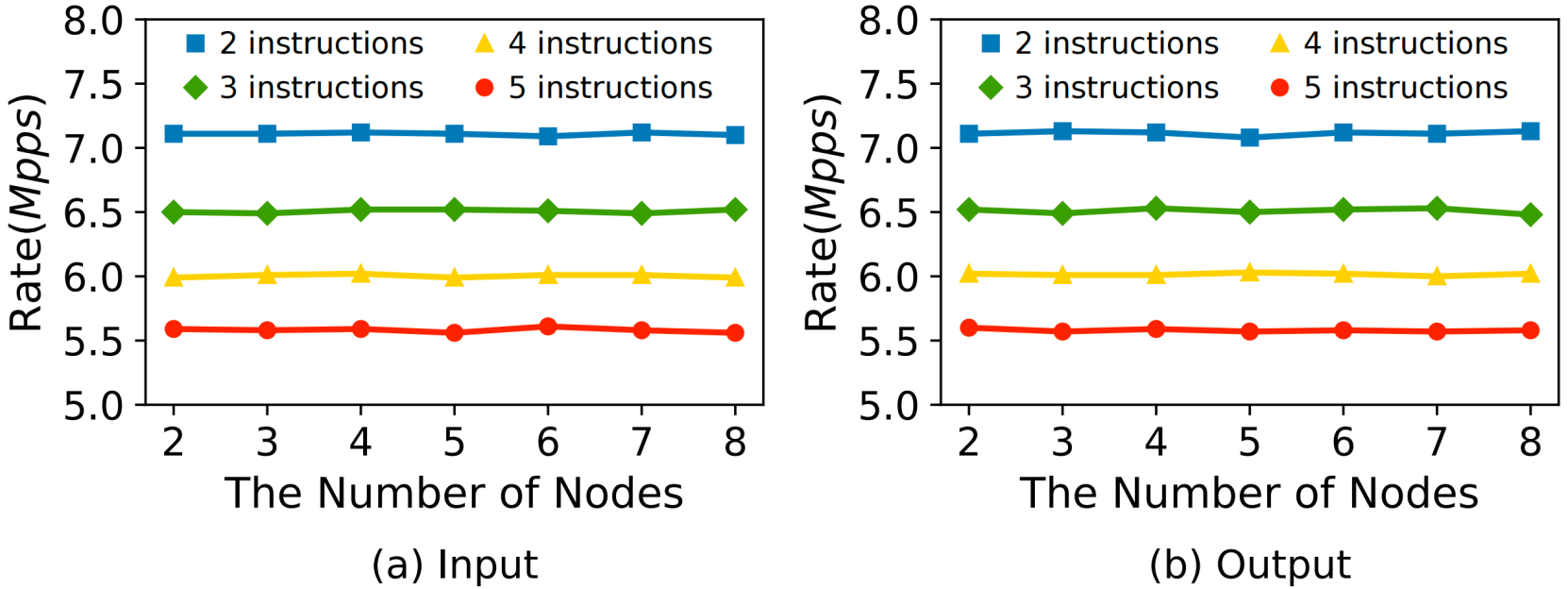


Figure 9: Processing rate of SmartNIC

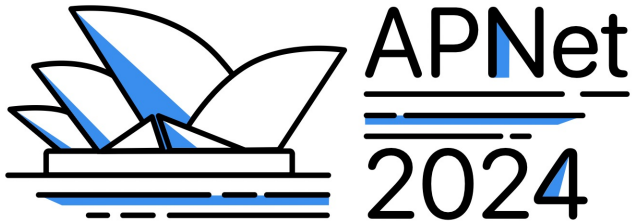
Conclusion and Future Work

Conclusion

- Low overhead SF-INT -- Keeping a constant packet size
- High-performance SmartNIC-equipped collector -- Nearly the line rate
- At the cost of slight delay
- A helpful attempt to promote the application of INT in the next-generation network

Future Work

- Compare INT, PINT, INT-Label under different application scenarios
- Couple with the existing sampling INT schemes
- Introduce 40G SmartNICs
- Enhance the ability to handling exceptions



Thank you !

An Integrated Solution for High-efficiency In-band Network Telemetry

Speaker：Xinxin Xiong

Email: holoxiong@stu.xmu.edu.cn