P4.org Applications Working Group Update

Mukesh Hira (VMware)

Jeongkeun Lee (Barefoot Networks)

Introduction to the Applications Working Group

- Started in November 2017
- Charter and Goals
  - Forum for discussing P4 applications requiring interoperability across vendors
  - Development of interoperability specifications
  - Development of open source code and test cases
  - Identification of gaps if any with respect to P4 architecture/language/API

- Logistics
  - Bi-weekly Working Group Meetings
  - Mailing list: p4-apps@lists.p4.org (Subscribe at lists.p4.org)
  - Github repository: https://github.com/p4lang/p4-applications
    - Meeting Slides, Specifications, Source Code

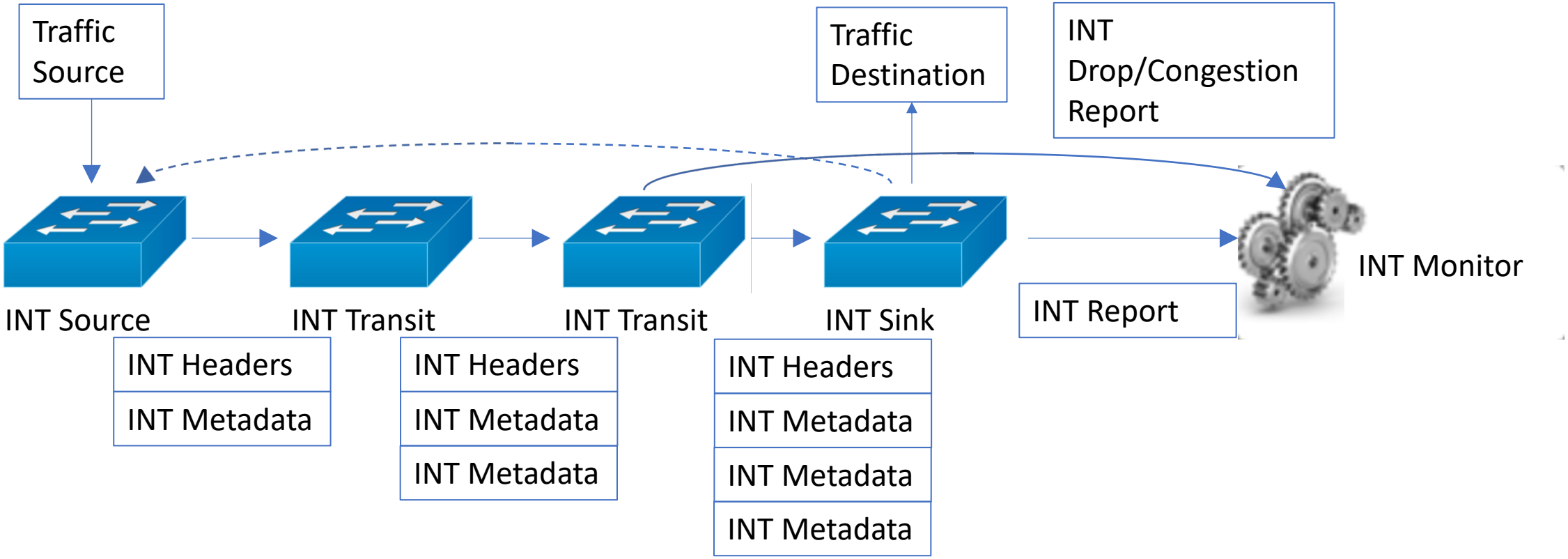
- Initial Focus: In-band Network Telemetry (INT)
- 30 Minute Time Slots provided at WG meetings for presenting other P4 applications

Working Group Participants

Mario Baldi, Gordon Brebner, Carmelo Cascone, Andy Fingerhut, Barak Gafni, Senthil Ganesan, Anoop Ghanwani, Robert Halstead, Jonghwan Hyun, Raja Jayakumar, Madhu Kashyap, Petr Kastovsky, Andy Keep, Xiaozhou Li, Alan Lo, Jonatas Marques, Tal Mizrahi, Gidi Navon, Brian O'Connor, Michael Orr, Heidi Ou, Shyam Parekh, Jianwen Pi, Albert Ross, Milad Sharif, Rajesh Sharma, Mickey Spiegel, Chris Sommers, Tom Tofigh, Ronald Vanderpol, Bapi Vinnakota, and others

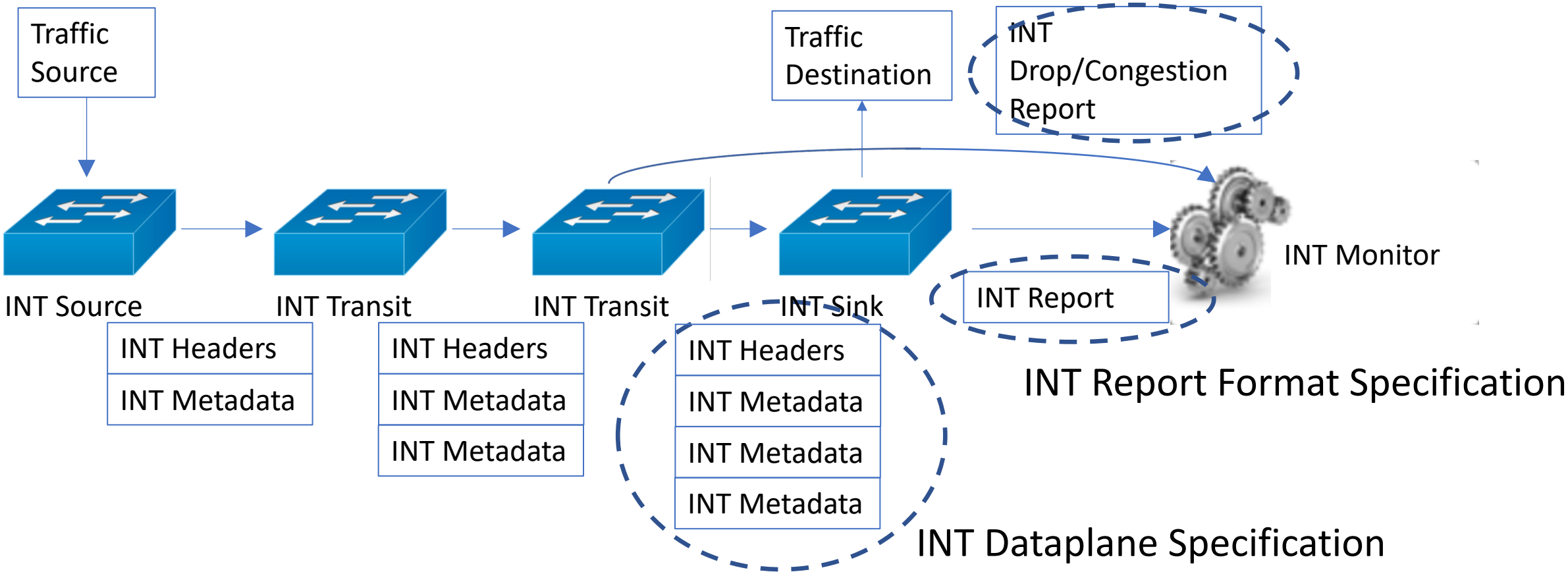
Thank you for your contributions

Quick Introduction to INT



- Real-time per-packet visibility into your network
  - What path did a packet take?
  - What network state did it experience at each hop?
    - Latency, Queue Occupancy, …

Working Group Accomplishments so far …



- Version 1.0 Specifications released in April 2018
- Version 1.0 Reference Code Development in progress

Version 1.0 INT Specifications Overview

- Dataplane Specification
  - INT header format is defined, INT header location is flexible
    - INT over Native Layer 4 with DSCP/Probe Marker indicator
    - INT over VXLAN/Geneve encapsulation
  - Simplicity and Efficiency are key goals
    - Each switch parses instruction bitmap, pushes its metadata at top of stack
    - Extremely simple MTU handling with no fragmentation support
    - Support for checksum neutral updates
  - Transit Switch Reference Code based on PSA architecture

- Report Format Specification
  - Flexible format
  - Two modes of operation
    - Reports from INT sinks + Drop and congestion reports from transit switches
    - Post-card reports from every switch

Lesson1: Dataplane Telemetry as First Class Citizen

- Dataplane telemetry = visibility into dataplane events at every packet
  - More than INT (= metadata collection in data pkts)
  - More than random pkt sampling → event-driven reporting
  - More than stat reading → direct correlation of event with pkt + metadata

- Dataplane telemetry interacts with standard forwarding features
  - INT encap affects MTU and L4 checksum
  - Reporting packet drops caused by various dataplane events (congestion, ACL, MTU limit, TTL …)
  - Share dataplane resources → efficiency matters

Lesson2: Inherit Speed and Flexibility of P4

- Protocol standardization at the speed of S/W release cadence
  - Two releases (v0.5, v1.0) of two specs in 5 months
  - Significant header changes, deprecated one header
  - Rapid revisions, use version number for dependency matching
- Encompass various dataplane architectures
  - Spec supports the union, not only intersection
  - INT v1.0 accommodates specifics of 4 different architectures
  - Yet do not sacrifice efficiency

Lesson3: P4 as Formal Specification

- Spec = header + protocol behavior & semantics

- P4 header = canonical expression
- Parser/table/control unambiguously express protocol behavior and interaction with other features

- Speed up implementation of the spec
  - By Xilinx, Netscope, ONF, Netronome, Barefoot Networks and more to come

- INT src/transit/sink reference implementation + unit tests in github, under review

Apps Working Group Roadmap

- Telemetry Metadata Semantics
  - Units (e.g. Queue Occupancy being reported in bytes/cells)
  - Precise Definitions (e.g. What is the size of a cell)
- Report Triggering Semantics
  - When does a switch generate a report? (E.g., periodic, flow arrival, congestion)
  - Different semantics and events over switches and smartNICs
- Other applications, e.g., computational networking
  - Offload compute/storage application logic to network
  - Identify missing primitives in the language spec or PSA
  - Generic in-band encapsulation similar to INT
    - Pkt from app client → encap app data → in-network processing → decap → app client or server