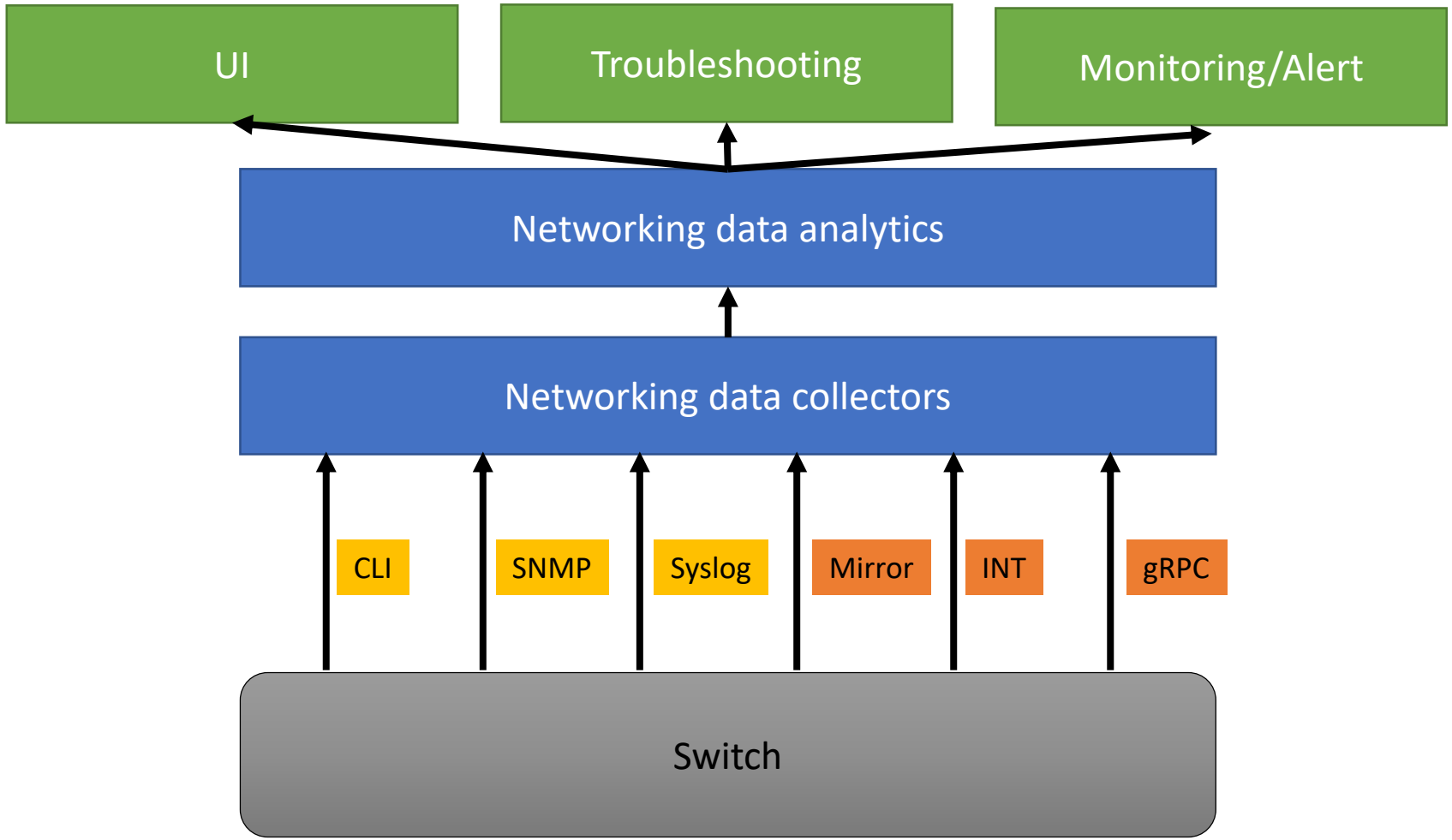
SONiC Network Telemetry from User Perspective

Yongfeng Liu

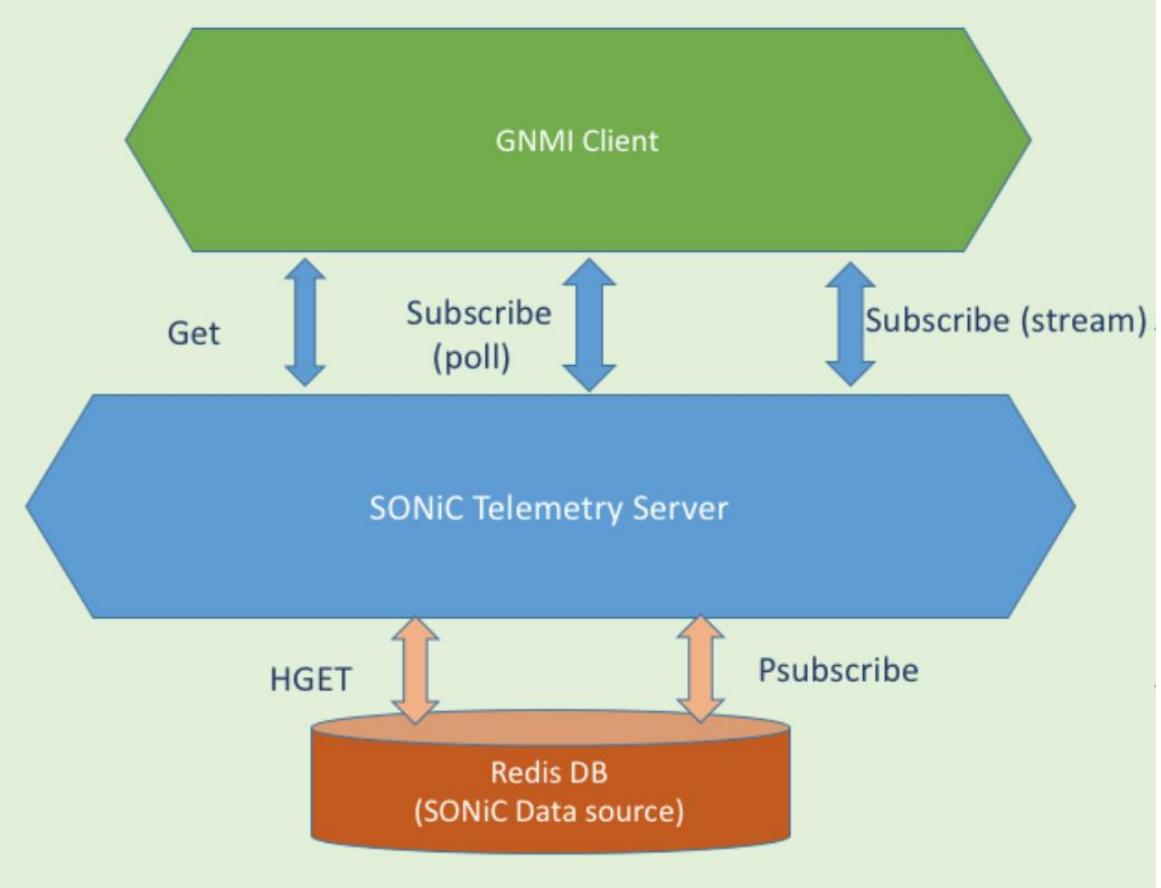
Network Telemetry Overview



Network telemetry on both Control plane and Data plane

- Switch/System level telemetry
  - Switch level ASIC status and statistics
    - Port/Queue counters
    - Buffer utilization/TAM snapshot
    - etc
  - Streamed to external collector from control plane.
- Packet/Flow level telemetry
  - Packet/Flow level event and information
    - Flow tracking event (generate event based on TCP SYN/FIN packet)
    - Per flow Latency/path change/congestion event
    - Packet drop event with detailed reason reported with flow info
  - Sent to external monitor/analyzer from data plane

Switch level telemetry via gRPC

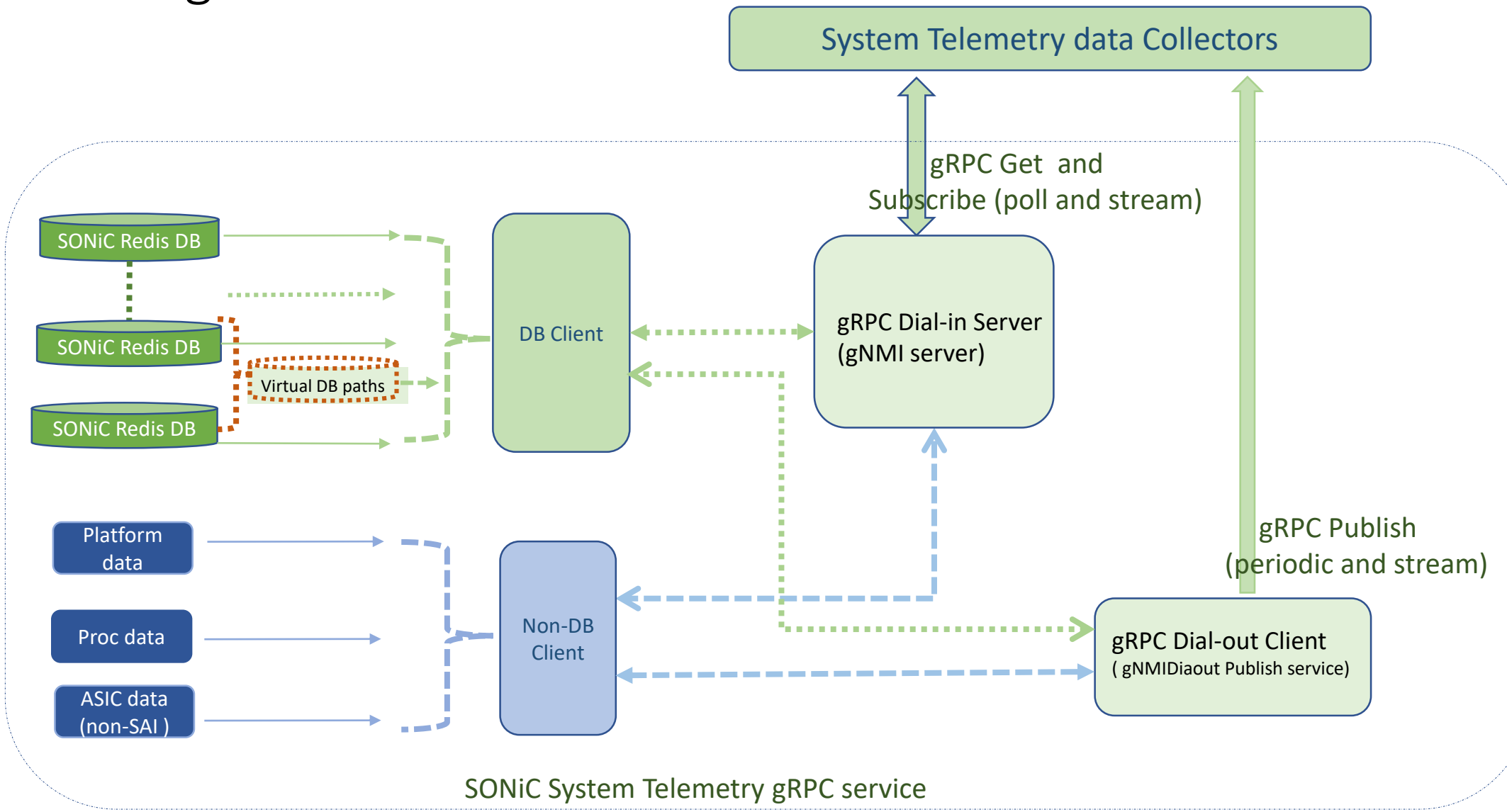


Efficient collection stack

Change only mode

Push instead of poll

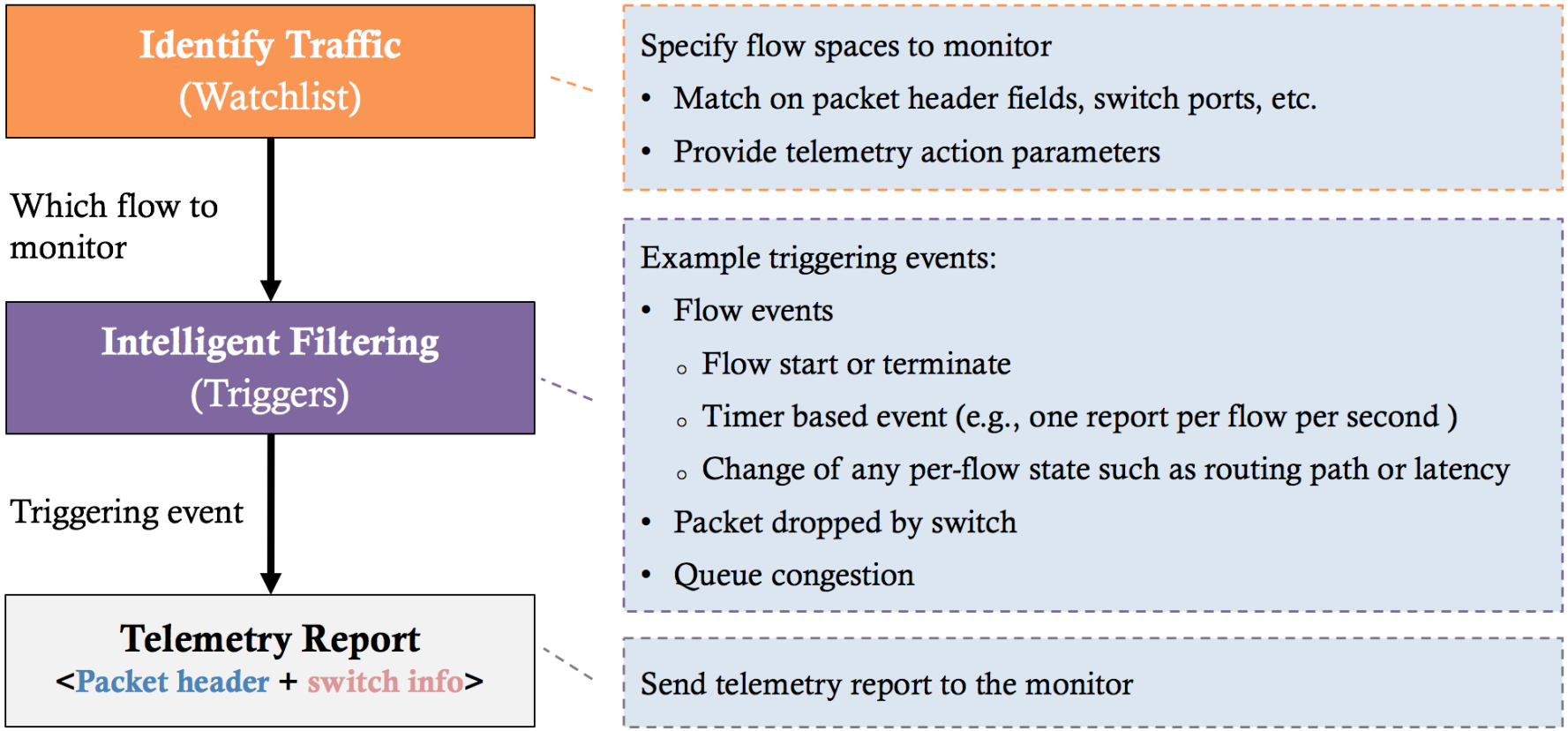
SONiC gRPC Architecture



Challenge of flow/packet level telemetry

- Lack of production deployment and operation experience.
- Flow/packet level information process requires heavy ASIC resources.
- Metrix of supporting feature requires flexibility and programmability of switch ASIC on varies aspects.
- Needs minimum side effect to packet forwarding functions.
- Inter-operability between different ASIC vendors.

Data plane telemetry workflow



Telemetry Identification Header

- Header format
  - Option 1: Probe marker
  - Option 2: DSCP
- Rule table
  - Option 1: dedicated watch list
  - Option 2: reuse existing ACL
- Action
  - Option 1: apply on original packet
  - Option 2: apply on copied packet
- Policy
  - Option 1: insert header for each packet of matched flow
  - Option 2: insert header with sampled rate of matched flow

Telemetry Instruction Header

- Header format
  - Header version
  - Instruction bit map
- Instruction bitmap could be and not limited to:
  - Switch ID
  - Ingress/egress port ID
  - Hop latency/Ingress timestamp/Egress timestamp
  - Queue ID + Queue occupancy/congestion status
  - TX utilization on egress port

Telemetry event detection

- Listed telemetry event from user perspective
  - TCP flow status reflecting establish and termination event.
  - Path change
  - Latency
  - Queue occupancy
  - Congestion status
  - TX utilization
  - Drop with reason code

Telemetry Report

- Report format
  - UDP/ERSPAN encapsulation
  - Composed with filled telemetry metadata
  - Along with truncated packet header
- Report handling
  - INT vs. Postcard
    - INT: Event detect on sink node
      - Stack metadata according to instruction hop by hop
      - Drop events needs special handling
    - Postcard: Event detect on each node
      - Trigger report as a mirrored packet on each hop
      - Analyzer needs additional steps to correlate mirrored packet of same flow acress multiple hops.
  - Options to periodically trigger report;
  - Options to suppress report without losing effective telemetry info;
  - Options to force report for debug purpose;

References

- on gRPC
  - https://github.com/jipanyang/sonic-telemetry/blob/master/doc/grpc_telemetry.md
- on DTEL (data plane telemetry)
  - https://github.com/p4lang/p4-applications/blob/master/docs/INT.pdf
  - https://github.com/p4lang/p4-applications/blob/master/docs/telemetry_report.pdf
  - https://github.com/CiscoDevNet/iOAM
  - https://tools.ietf.org/html/draft-ietf-ippm-ioam-data-02