NVMe-oF SAN Telemetry with Programmable Switches

Surendra Anubolu, Maggie Gao and Mohan Kalkunte
Broadcom Inc
Fabric Solution Key Parameters

Port Speed
Scale
Cost

Operations
SLAs
Monitoring
NVMe and PCIe bandwidth is picking up pace!

PCIe Gen4 NVMe drives are available today!

PCIe Gen5 NVMe is expected within 2 years

Flash Array performance is increasing

New Non Volatile memory architectures

Bandwidth (GB/s) x4 PCIe
Ethernet as Fabric for Storage

Ethernet continues to grow as an unified Fabric

- 40X increase over 10 years
  – Faster than Moore’s Law!

- Large Ecosystem

- Economies of Scale
  – Cost per Terabit bandwidth is falling

- Management tools
  - TCP/IP

- Unification
Fabric Solution Key Parameters
- Ethernet As Fabric

- Port Speed
- Scale
- Cost

- Operations
- SLAs
- Monitoring
Network Operations

Failures occur
- Need tools to isolate the problem and resolve them


SLA
- Uptime
- IOPs
- Latency
- …

Deploy
- Plan
- Provision
- Access

Monitor
- SLA perf
- Errors
- Hotspots
Storage Centric Analytics

- Where are my storage devices?
- Can I track IOPs by Connection? Host? Storage Server?

- Do I have hotspots?
- Are SLAs met?
Event Driven Streaming Telemetry

Switch tracks
- internal state (queues, buffers)
- Flows

Triggers
- Can set Threshold Crossing Alarms
  - Latency > N microseconds
- Timer driven periodic report

Deduplication of reports
- Reduce load on collector

Steaming Telemetry
- Protobuf encoding
Programmable Switches for NVMe-oF

- Event Driven Streaming Telemetry
- Switch Attribute for Fabric monitoring and Telemetry
- Stateful Flow based Tracking
- Programmable Pipeline
- Deep Parsing
- User programmable Instrumentation Capabilities
- 12.8Tbps capacity per instance
- Flexible Tile architecture
  - Compiler driven dataplane
- Stateful tracking
  - Tracking at full packet rate
- Streaming Telemetry
Test Results

IOStat(/proc/diskstats) vs Trident 4 Switch Telemetry

Switch measurements matches reference IOStat
Summary

- Ethernet as Fabric presents a matching growth in connection speed and Switch Scale for Storage
- Trident 4 can Parse and Track NVMe-oF packets at high speeds
- Trident 4 stateful capabilities enable native storage metric reports
- Trident 4 offers native support for Rich Streaming Telemetry
Thank You!