In-Network Compression for Accelerating IoT Analytics at Scale

HOTI2023

- Rafael Oliveira
- Ada Gavrilovska

Georgia Tech

IoT Background 2m

- IoT Applications
- Devices
    - Telemetry.
    - Sensor-Actuator
- Pub-Sub Model
    - Message sample

Georgia Tech

In-Network Accelerator Background 1m

- Where is it placed?
- Types of in-network accelerators
- Generic Type

In-Network Accelerator Challenge 1m

- Transferring data from NIC to Host CPU

- Message Types and Challenges
  - Modified: small (values only). Yields high message rate but low Gbps
  - Unmodified: Lots of redundant - JSON characters. Yields high Gbps but low message rate.

In-Network Compression Opportunity 1m

- Post Pipeline Opportunity (post critical path)

- Insight
  - Batch small messages
  - Compress and Batch bigger messages

In-Network Compression Challenge 1m

- General Compression algorithm
  - Resource hungry – takes resources from critical path
  - Yields subpar compression rate

Comprex
In-Network Compression for Accelerating IoT Analytics at Scale

Georgia Tech

Insights Driving Comprex 1m

- Post critical-path
  - Relaxed or no Real-Time requirements
- Device Message Behavior
  - Once deployed, device will send same message format
- Redundancy
  - Information to be transferred might already be present in the HOST
- Batching

Comprex: General Structure 1m

- Overview Image

Comprex: Compression 2m

- Overview image with Compression Engine, the Topic Table and the Host Comp.Diff side.

- Comp.Diff
  - Comp.Diff detects Static and Dynamic Regions and sends them to the CE
  - Reconstructs message with static regions
- CE:
  - Checks to see if Static Region has changed
  - Removes static regions from message

Comprex: Compression 2m

- CE animated example of how it works

Compression 1m

- Relationship of stored messages and N and the impact on lossiness
- CE Check.Diff
  - CE parallel checks on static region allows N to grow bigger with no impact on lossiness

Micro-Bench 1m

- Decompression throughput vs N

Comprex: Batching 1m

- How it works:
  - Message sizes, batch size
  - Encoding
  - Execution

Evaluation 1m

- Comprex vs Snappy

Evaluation 1m

- Comprex impact on throughput
- Comprex no-negative-impact on latency

Summary .5m

- Comprex: as IoT-Specific compression that yields high data transfer throughput without impacting latency.

Thank You.