Clustering network control in data centers

taking local decisions instantaneously

Background

- Amount of data to be processed grows explosively
- Stand alone servers are not enough anymore
- Data centers based on commodity hardware
- Conventional network protocols not always fit

A common data center

- 1000's of servers, 48 servers per rack
- High bandwidth (1/10 Gbps), Low latency (<100 usec)
- 3-4 tier multiple root tree-like topology (grows horizontally)
- Work-flow pattern, depends on DC purpose:
  - Scatter-gather (MapReduce, search)
  - One-to-one transfers (backup, storage, etc.)
- Similar purpose workers tend to be localized

Data center traffic characteristics

- Most (~80%) flows are small in size (<10KB)
- About 50% of flows last < 100ms
- Flow inter-arrival time at TOR ~10-100 ms
- Per-packet arrival is an ON/OFF process
- Intra-rack communication varies between 50 and 75% of all traffic
- Link utilization: edge/aggregate links – almost never hotspots; ~20% core links – hotspots at least 50% of time

Transport problems in DC

- In-cast
  - Many-to-one communication pattern
  - Synchronized retransmissions
- Queue buildup, increased latency
- Shared switch memory
- High RTO_{min} (200 ms)
- Coarse-grained OS timers
- Non-optimal flow placement

Solutions in literature

- Decreased RTO_{min}
- Use of high resolution OS timers
- Desynchronized RTO timeout
- DCTCP: ECN-based adjustable cwnd factor
- ICTCP: rate control at the receiver
- XCP: rate control at the routers

Solutions in literature (cont'd)

- Onix: distributed control plane for DCs
- Hedera: global flow scheduling
- MicroTE: scheduling with per-second granularity
- ElasticTree: energy saving global control
- PortLand: scalable layer 2 addressing/routing/forwarding
- VL2: layer 2 virtualization
- FatTree: scalable data center topology

Our approach

- Rationale: Global control in DC is necessary, but fine-grained decisions are hardly possible
- Hypothesis: Clustering the topology and making decisions locally allows for fast and accurate control
- Use cases:
  - Energy savings
  - Congestion avoidance
  - Load balancing

Q&A

Thank you!
Questions, thoughts, suggestions?