Paper Evaluation, Hedera: Dynamic Flow Scheduling for Data Center Network

Jiani Jiang <jianij@kth.se>

1. Paper summary

Problems of aggregate bandwidth demands to clusters of millions of hosts limited by single point of failure and oversubscription of high-level links are called for deliberation by authors. The paper focuses on utilizing aggregate network recourses (paths or ports) to deliver a better bisection bandwidth using dynamic flow scheduling method which is called Hedera. To maximize bisection bandwidth, authors proposed multi-rooted fat-tree with detecting large flows by switches, estimating natural demands and installing paths on switches by demand estimator and scheduler. For the scheduling algorithm, Global First Fit and Simulated Annealing are compared to ECMP and Simulated Annealing is the best one that delivers near-optimal bisection bandwidth for a wide range of communication patterns with tens of milliseconds runtime for a large number of iterations. This idea is evaluated by well-founded experiments of physical testbed and simulated data center networks which both positively proofed authors' idea.

2. Top 3 contributions

The authors achieve high degree of available path diversity and more equal-cost paths between any given source and destination host pair. And it can deliver bandwidth gains with effective cost. Meanwhile, there is no modifications to end-host network stacks or operation systems.

3. Problems

1.Since there is only one centralized central scheduler, when the network load increases, it may be hard to guarantee all flows will be accommodated in an acceptable latency. 2. The runtime of scheduler is tens of milliseconds, however, the scheduling frequency is 5s limited by the OpenFlow NetFPGA implementation. Is there a better way to speed up the frequency to achieve a better bisection bandwidth?