ABSTRACT

Data center networks often use multi-rooted Clos topologies to provide a large number of equal cost paths between two hosts. Thus, load balancing traffic among the paths is important for high performance and low latency. However, it is well known that ECMP---the \textit{de facto} load balancing scheme---performs poorly in data center networks. The main culprit of ECMP’s problems is its congestion agnostic nature, which fundamentally limits its ability to deal with network dynamics.

We propose Expeditus, a novel distributed congestion-aware load balancing protocol for general 3-tier Clos networks. The complex 3-tier Clos topologies present significant scalability challenges that make a simple per-path feedback approach infeasible. Expeditus addresses the challenges by using simple local information collection, where a switch only monitors its egress and ingress link loads. It further employs a novel two-stage path selection mechanism to aggregate relevant information across switches and make path selection decisions. Testbed evaluation on Emulab and large-scale ns-3 simulations demonstrate that, Expeditus outperforms ECMP by up to 45\% in tail flow completion times (FCT) for mice flows, and by up to 38\% in mean FCT for elephant flows in 3-tier Clos networks.

This paper will be presented at ACM Symposium on Cloud Computing 2016, October 5-7, 2016, Santa Clara, California.

Supervisor: Dr Henry XU
Research Interests: Data center networking; SDN; NFV

All are welcome!