Lossless Data Center Networks Congestion Control

Congestion control (CC) can help achieve the requirements of today's and future generation data center networks (DCNs): namely high bandwidth, ultra-low latency and stability. There are several state-of-the-art proposals but there is no optimal CC scheme yet. Categorizing the CC schemes into end-host control and in-network scheduling, we at Huawei Munich Research Center believe that in-network scheduling provides faster convergence of transmission rate towards the fair rate of the network as compared to end-host control. A novel in-network scheduling CC algorithm together with an optimal load balancing scheme has been designed for RoCEv2 (RDMA over Converged Ethernet version 2) transport. The proposed CC algorithm leverages in-network telemetry (INT) to estimate link load information and pace traffic accordingly, thereby, reducing congestion. It has optimal packet level scheduling and is traffic insensitive, thus fair to all flows (like that of Processor Sharing). The algorithm is also simple (in terms of simple operations on packet arrival) and can be deployed in the DCN switches. The goal of the algorithm is to minimize the FCT (flow completion time) of large flows and provide a bounded delay for short flows. The load balancing scheme built on top of the optimal scheduling scheme aims to optimize the overall network performance, achieving higher network utilization while guaranteeing the performance of individual flows. We intend to compare this algorithm by simulation in Omnet++ simulator with some state-of-the-art algorithms from which the algorithm derives its motivation.

Advisors

Patrick Krämer, Johannes Zerwas, Andreas Blenk
Siyu Tang (Huawei)