Data Center TCP - Reproducing Key Findings

CS 656 Research Paper

Colin Howes
University of Waterloo
200 University Ave. W
Waterloo, ON N2L 3G1
chowes@uwaterloo.ca

ABSTRACT

Data centers supporting host a diverse range of heterogenous traffic with varying throughput and delay requirements. Transport layer protocols power data center traffic must be able to tolerate bursts of traffic, provide low latencies for time-sensitive traffic, all while maintaining high throughput for large data flows. Several modifications to the TCP protocol have been proposed to address shortcomings in conventional TCP congestion control algorithms, which relies on packet loss as a metric of network congestion. Data Center TCP, Incast TCP, Multipath TCP, and TCP BBR are discussed, and a replication of selected results related to DCTCP performance are presented.

1 INTRODUCTION

Modern distributed cloud applications rely heavily on the *partition/aggregate* design pattern, in which an application is broken in into hierarchical layers and time-sensitive requests at higher layers are divided and delegated to workers in the lower layers. Workers perform some component of a task and return a result to an aggregator, which is combined with results from other workers and passed back up through the hierarchy. A problem arises when workers simultaneously report results back to an aggregator, since this traffic must pass through a shared bottleneck, which results in high queueing delays for time-sensitive traffic.

2 IMPROVING TCP FOR DATACENTERS

2.1 Incast TCP

2.2 Multipath TCP

2.3 TCP BBR

2.4 Data Center TCP

Data center TCP (DCTCP) attempts to address the problem of latency in partition/aggregate traffic by reducing queue length without affecting throughput for large TCP flows.

3 REPRODUCING DCTCP RESULTS

3.0.1 Methods. Selected results from [1] were reproduced using the Mininet network emulator running on Ubuntu 12.04 with a patched version of the 3.2.18 Linux kernel patched to add in support for DCTCP. A custom utility was modified from the Mininet utilities repository to monitor queue length, and another utility was created to monitor bandwidth.

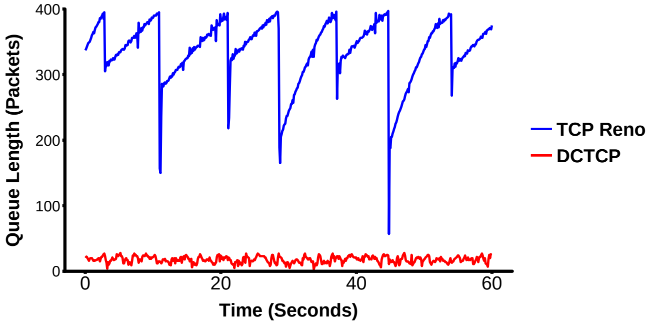

Figure 1: Comparison of queue length over time between DCTCP and TCP Reno with 2 flows

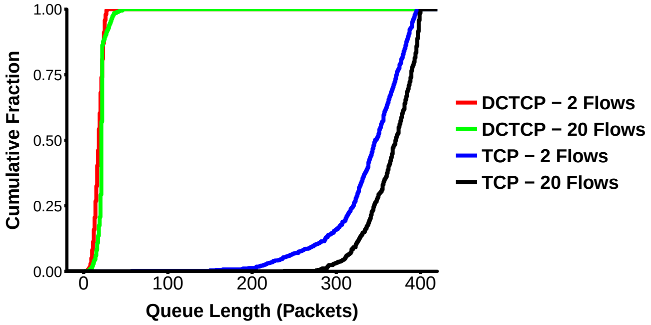

Figure 2: CDF of queue length for DCTCP and TCP Reno with 2 and 20 flows

3.0.2 Results.

3.0.3 Discussion.

3.0.4 Limitations.

4 CONCLUSIONS

REFERENCES

[1] Mohammad Alizadeh, Albert Greenberg, David A. Maltz, Jitendra Padhye, Parveen Patel, Balaji Prabhakar, Sudipta Sengupta, and Murari Sridharan. 2010. Data center tcp (dctcp). In *ACM SIGCOMM computer communication review*, Vol. 40. ACM, 63–74. http://dl.acm.org/citation.cfm?id=1851192

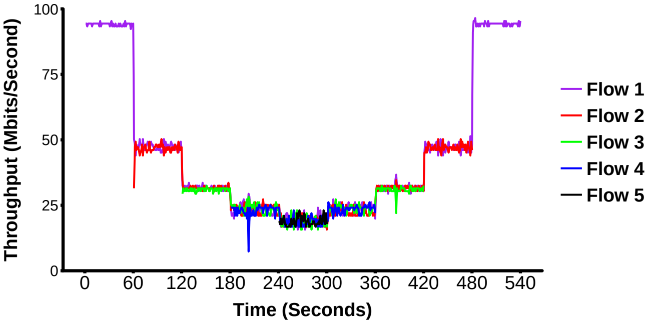

Figure 3: Convergence of 5 flows DCTCP

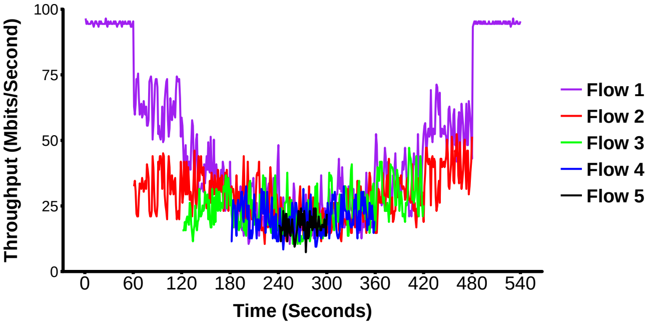

Figure 4: Convergence of 5 flows TCP Reno

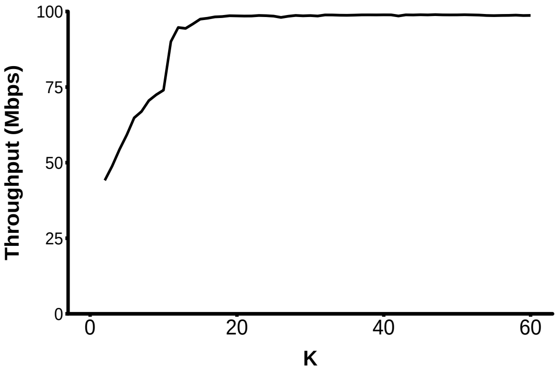

Figure 5: DCTCP throughput for varying values of K