Data Center TCP (DCTCP) Evaluation and Analysis

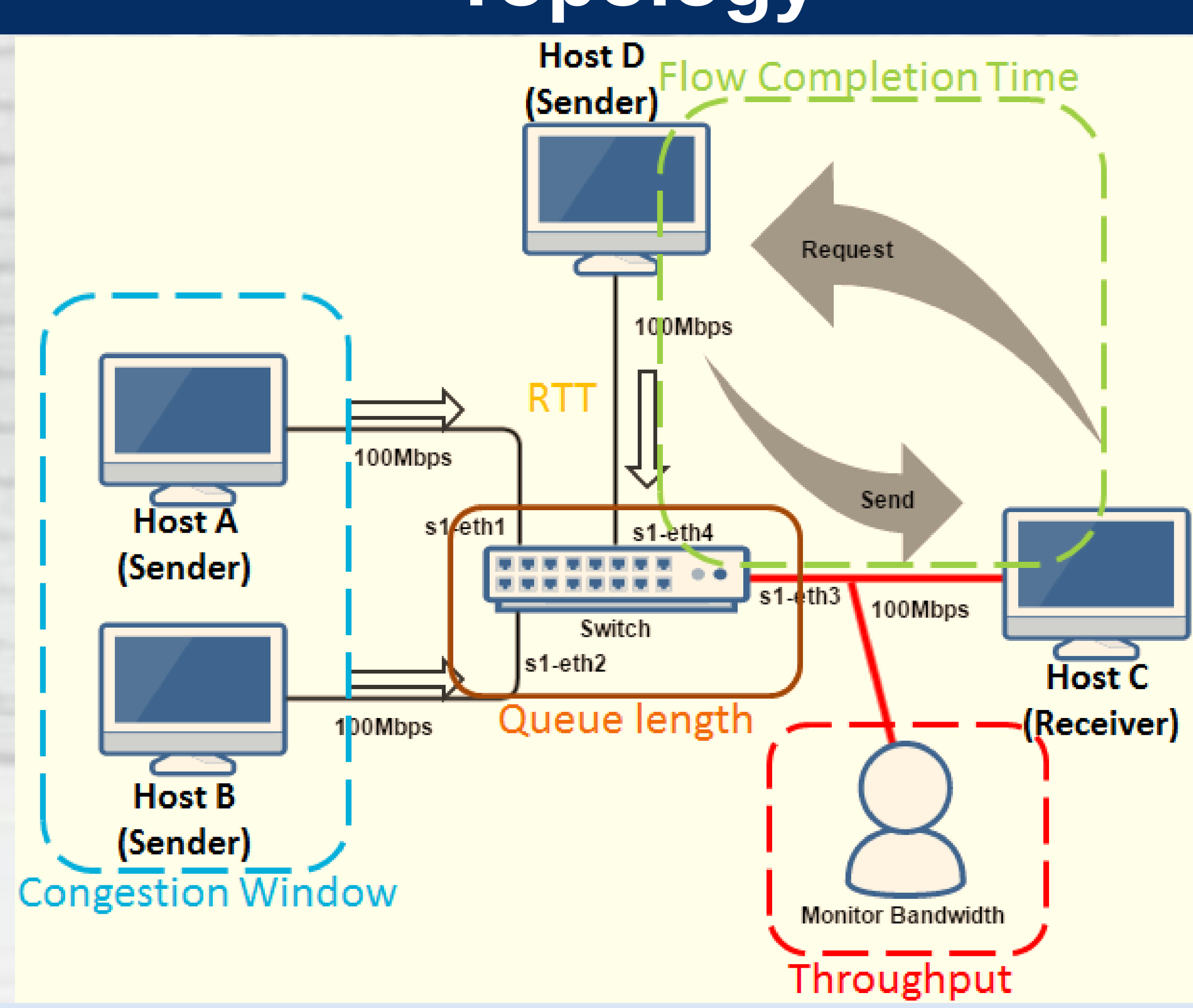
Lin Yiu Yeung, Lam Tsz Ho, Tam Chun Ho, Hung Faan Chung
Group: KAI 2
Advised by Prof. Kai CHEN
Computer Engineering Program

Abstract

Data Center is an essential networking component nowadays. However, more than 99.9% of the traffic are packets constructed with TCP and it is not specifically designed for the data center environment, leading to some buffer pressure. A customized protocol for data center environment, Data Center TCP (DCTCP), has been deployed and thus tried to alleviate the impairment.

The performances of TCP and DCTCP have been compared by using different parameters. Throughout the FYP, a new aspect and direction have been used to study DCTCP, marking threshold K. It is a significant parameter affecting the performance. And, the potential hidden problem that has not been mentioned was found.

Topology



In Mininet(Virtual Network), a star topology with four hosts connecting to the same switch has been built. Host C is the receiver and host A, B and D as the senders to generate traffic. The bottleneck is the link between the switch and C. Different scenarios have been tested by using this setup.

TCP Impairment in Data Center

- Incast (Flows aggregation exhausting buffer size),
- Queue Build Up (Large Flows occupying the queue)
- Buffer Pressure (Not be able to handle traffic burst when buffer is exhausted)

Congestion Control in TCP / DCTCP

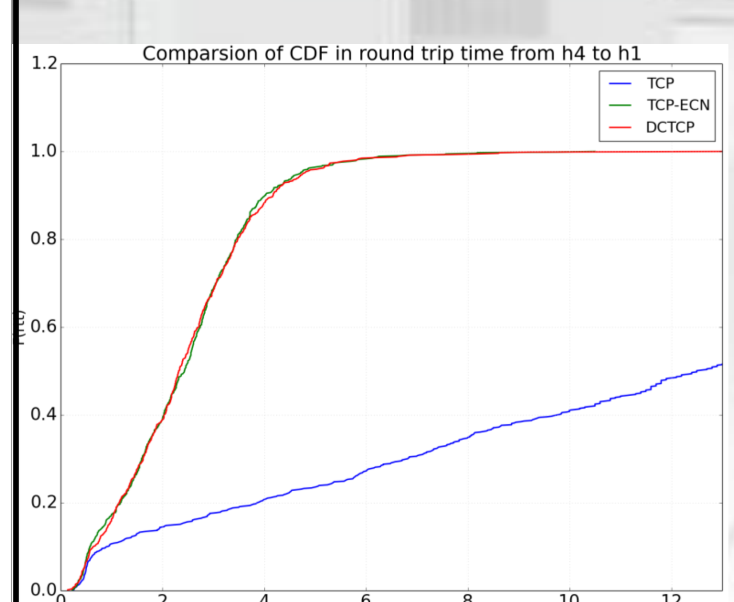
Receiver:

Under TCP, the buffer does nothing but drops the incoming packets. TCP with **Explicit Congestion Notification (ECN)**, notifies the sender once the buffer exceeds a certain threshold K with increasing probability. Under DCTCP, it leverages the ECN and marks every packet which exceeding the K.

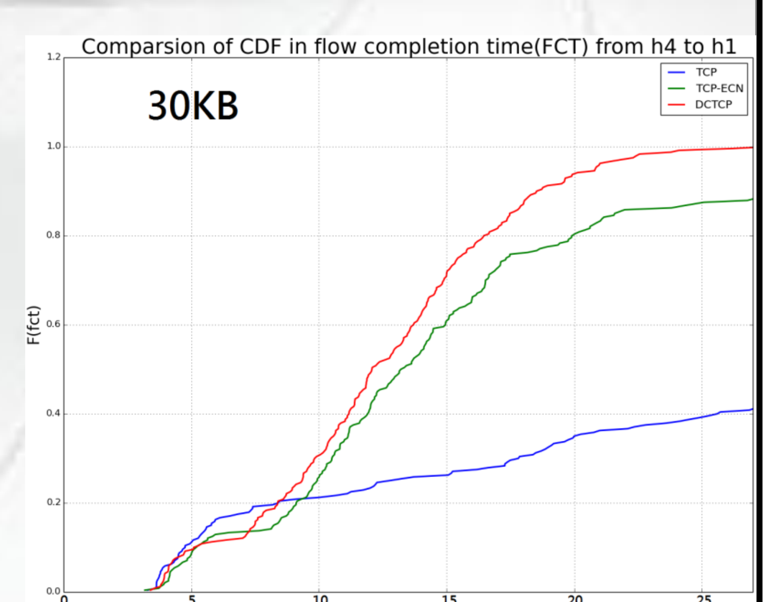
Sender:

Under TCP Reno scheme, the congestion window reduces to 1 when timeout or ½ the original value when 3 ACKs is received. In DCTCP protocol, the congestion window is just slightly reduced and it is based on *cwnd = cwnd x [1 - (α/ 2)]* , *where* α is the estimation of the packet being marked in the next coming session.

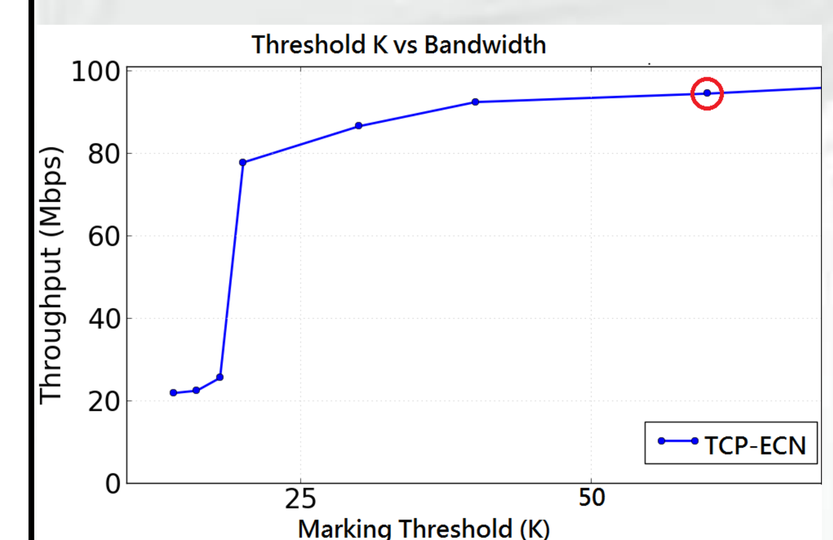
Result



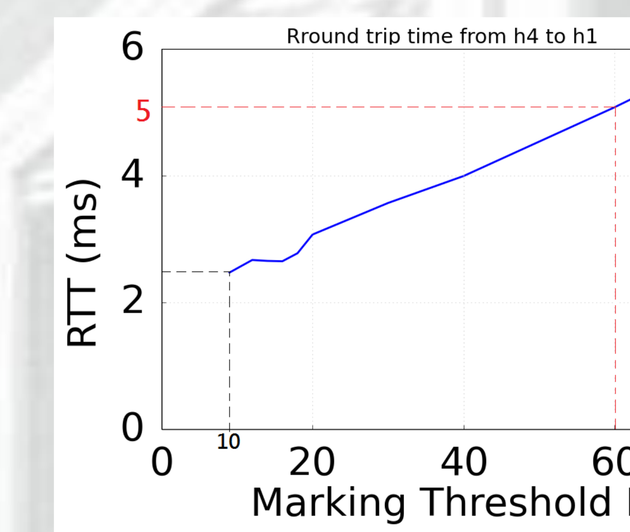
CDF of RTT



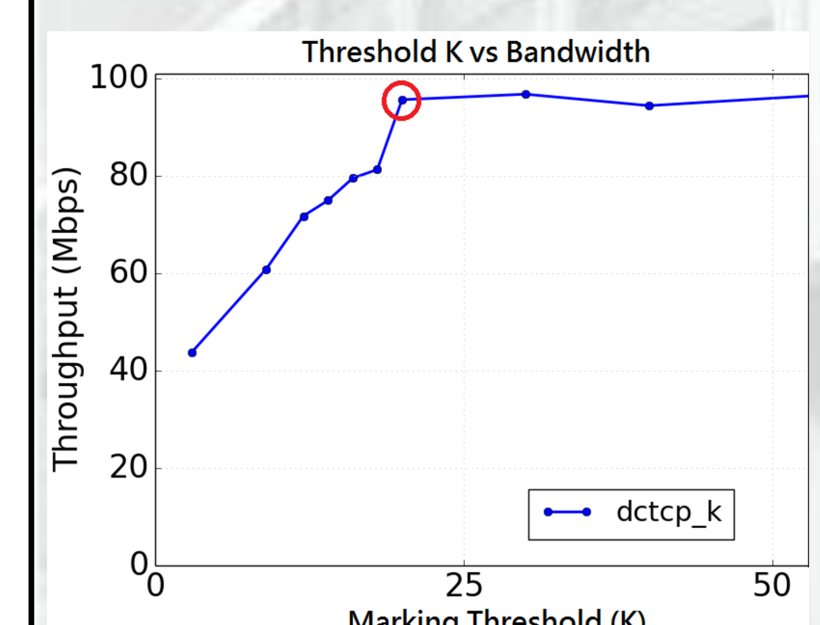
CDF of FCT (30KB)



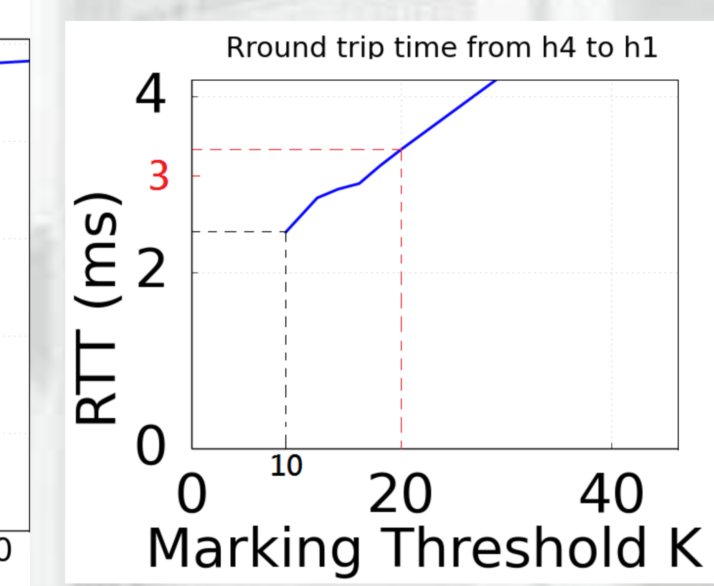
Throughput vs K (ECN)



RTT vs K (ECN)



Throughput vs K (DCTCP)



RTT vs K (DCTCP)

Conclusion

In our project, the results showed that DCTCP perform more outstanding than TCP, in terms of round-trip time, flow-completion time and the queue length. DCTCP has the advantages that the paper claimed, like: smaller latency for short flows, higher throughput for large flows and shorter queue length. However, the performance of DCTCP is dominated by the marking threshold K.

The original design of DCTCP suggested that the method of choosing the marking threshold K should be “K > (C x RTT) / 7”, where C is the bottleneck link capacity. But, the suggested models do not suitable for our experiment. More studies have to be done on this aspect.