PADERBORN UNIVERSITY

Advanced Networked Systems SS24

Exercise 3: Data Center Networking

1 Basic Concepts

Question 1: Explain the concept of full bisection bandwidth.

Question 2: Brief explain how routing is done on a fat-tree network.

Question 3: What is ternary content addressable memory (TCAM) and how does it work?

Question 4: What is the load balancing granularity of a fat-tree network? What strategies can we use to improve load balancing on fat-tree networks?

Question 5: Compare the advantages and disadvantages of layer 2 and layer 3 addressing for data center networks on the following aspects: plug-and-play, scalability, switch state, and support for VM migration.

Question 6: What is the TCP incast problem?

Question 7: Explain the general idea of explicit congestion notification (ECN).

Question 8: How does DCTCP leverage ECN for data center congestion control?

Question 9: Why is BBR not suitable for data center networks?

Question 10: How does TIMELY adjusts the congestion window?

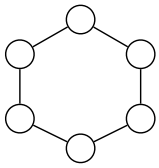
Question 11: What hardware modifications are needed to support TIMELY?

Question 12: How does Swift address host congestion?

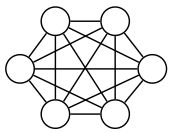
Question 13: What is extreme incast and how does Swift address it?

2 Bisection Bandwidth

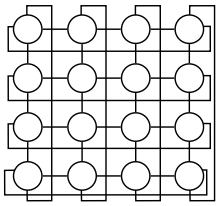
Given the following network topologies where the number of nodes is given by n. All links are assumed identical, with bandwidth of 1 Gbps.



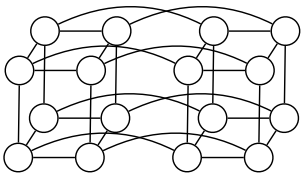
(1) Ring



(2) Clique



(3) 2D torus



(4) Hypercube

Question 1: What is the bisection bandwidth of each of these topologies?

3 Fat-tree Network Properties

Assume we build a fat-tree network with 64-port switches.

Question 1: What is the maximum number of servers can the fat-tree network connect?

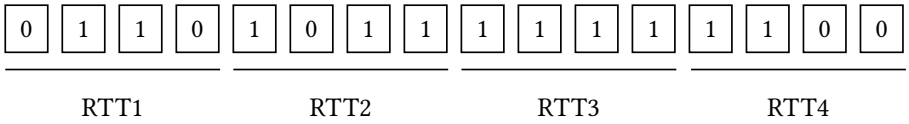
Question 2: How many switches are needed to build such a topology?

Question 3: How many equal-length paths are there between a pair of servers not in the same pod?

Question 4: Assume we are going to build a 2-rooted tree with the same switch specifications. How many switches are needed to interconnect the same set of servers?

4 Data Center TCP

We know DCTCP uses the rate of ECN markings in every RTT to estimate the congestion level with a smoothed moving average (assume the weight given to history is 0.5). We are given the following sequence of ECN markings in four RTTs when using DCTCP.



Question 1: What is the expected congestion window size at the end of every RTT in relation to the starting congestion window size before RTT1?