List of Figures

Figures	*Title*	*page*
1.1	Philosophy of Data center	2
1.2	Industries relying on the Data center	3
1.3	Importance of Latency	4
1.4	Example of Aggregation/Partition technique [Wilson et al., 2011]	5
1.5	The problem of Incast in Aggregation/Partition	6
1.6	Global Data center traffic in 2020 [Cisco, 2015]	6
1.7	Queuing of mice flows	7
1.8	Multipath topology in Data center	9
1.9	An example of flow and packet-level multipath routing	10
2.1	Bisectional width of network	14
2.2	An example of 2 and 3-tier topologies	16
2.3	An example of Tree-based topology	17
2.4	An example of k=4 Fat-Tree topology	18
2.5	Example of n=4, Bcube1 topology [Guo et al., 2009]	19
2.6	Example of n=4, DCell topology [Guo et al., 2008]	20
2.7	An example of VL2 topology	20
2.8	Example of c-Through topology [Wang et al., 2010]	21
2.9	Example of Helios topology [Farrington et al., 2010]	22
2.10	DCTCP congestion control	25
2.11	HULL architecture [Alizadeh et al., 2012]	25
2.12	HULL topology	26
2.13	Gamma-correction function [Vamanan et al., 2012]	27
2.14	Resource reservation-based scheme (PDQ)	28
2.15	Disadvantages of fair share policy for delay sensitive traffic	28
2.16	Dynamic mapping of 4 flows to 2 queue and its advantage	29
2.17	Max-min timeslot allocation [Perry et al., 2014]	29
2.18	Priority based scheduling/dropping in pFabric	30
2.19	On-path data aggregation [Mai et al., 2014]	31
2.20	Base RTT estimation in DX	31
2.21	QJUMP design	32
2.22	Key components of NDP [Handley et al., 2017]	33
2.23	Path change in dynamic load-balancing scheme	35
2.24	Flow and Packet-level load-balancing	35
2.25	Source and Switch based load-balancing	36
2.26	Flow placement in Hedera with GFF	37
2.27	CONGA load-balancing	38
2.28	Cross-layer protocol stack of DeTail	39
2.29	Problem of load oblivious load-balancing in ECMP	39
2.30	MPTCP architecture	40
2.31	load-balancing in MPTCP	40
2.32	Presto architecture	41

2.33 Packet aggregation using GRO [He et al., 2015] 41
2.34 Hermes overview [Zhang et al., 2017] 43
2.35 Working of Flowtune 44

3.1 3-tier data center topology 48
3.2 2-tier simulation topology for TAP 51
3.3 3-tier simulation topology for TAP 52
3.4 2-tier: Average flow completion time 53
3.5 2-tier: Total timeout events 53
3.6 2-tier: Total flow duration 54
3.7 2-tier: Average throughput of large flows 54
3.8 CDF of 99.5th percentile flow completion time of 2-tier data center 55
3.9 3-tier: Average flow completion time 56
3.10 3-tier: Total timeout events 56
3.11 3-tier: Total flow duration 57
3.12 3-tier: Average throughput of large flows 57
3.13 CDF of 99.5th percentile flow completion time of 3-tier data center 58
3.14 Simulation topology for TP-SRSF 59
3.15 Average flow completion time for TP-SRSF 59
3.16 Total timeout events for TP-SRSF 59
3.17 Total flow duration for TP-SRSF 60
3.18 Average throughput of large flows for TP-SRSF 60
3.19 Effect of flow preemption 61
3.20 CDF of 99.5th percentile flow completion time of TP-SRSF 61

4.1 Data center topology with oversubscription 64
4.2 Average flow completion time Vs Packet size 65
4.3 Number of packets dropped Vs Packet size 65
4.4 Packet merge and split at ToR switch 68
4.5 AFCT of Tahoe 72
4.6 Number of packets dropped of Tahoe 73
4.7 AFCT of Reno 74
4.8 Number of packets dropped in Reno 75
4.9 AFCT of Newreno 76
4.10 Number of dropped packets in Newreno 77
4.11 AFCT of Sack 77
4.12 Number of dropped packets in Sack 78

5.1 Fat-Tree Data center topology 82
5.2 Flow collision in faulty network 82
5.3 Network after re-routing of flow F2 83
5.4 An example of naive path selection scheme 83
5.5 An example of FlowFurl Algorithm 85
5.6 Topology used for simulations 91
5.7 Flow distribution with size at different load 92
5.8 Mean flow latency 93
5.9 Tail flow latency 93
5.10 Average throughput of flows 94
5.11 Total number of packets dropped 94
5.12 Number of re-routeing events 95
5.13 Deadline achieve 95
5.14 Mean flow latency normalized to ECMP 96

5.15 Tail flow latency normalized to ECMP 96
5.16 Number of re-routeing events normalized to FlowFurl-Basic 97
5.17 Deadline achieve 97
5.18 Toy topology with only two path between end hosts 98
5.19 Behavior of flow-level switching 98

A.1 Re-ordering problem 104
A.2 Proposed algorithm for subflow segment estimation 105
A.3 Simulation Topology for 2 subflows 108
A.4 Simulation Topology for 3 subflows 109
A.5 Total number of re-transmissions 110
A.6 Total number of re-ordered segments 110
A.7 Average waiting time per segment 111
A.8 Total number of re-transmissions 112
A.9 Total number of re-ordered segments 112
A.10 Average waiting time per segment 113