Recitation 10 — DCTCP

Motivation

- Datacenters: owned and operated by a single company. Traffic types tend to follow specific patterns (e.g., partition/aggregate)
- Different traffic types: latency-sensitive and throughput-sensitive
- Problems in datacenters: incast, queue buildup, buffer pressure

Main idea

- DCTCP uses ECN as an early notification to reduce the window size based on the fraction of marked packets
- Lets DCTCP keep queues mostly empty, allowing for space to handle partition-aggregate traffic patterns
- Goals: high burst tolerance, low latency, high throughput

DCTCP Algorithm

- Mark packets if queue length > K on packet arrival. Marks are reflected in ACKs.
- Sender maintains an estimate (a) of fraction of packets marked. a = 0 → no congestion; a = 1 → high congestion.
- New window size = current window size * (1 - a/2)

Why does it work?

- Reacts earlier to congestion than TCP
- Marks based on instantaneous queue lengths (faster feedback to sources)
- Reacts in proportion to congestion, not just its presence
- Keeps queues small
- Enough room in queues to absorb bursts