Lecture 12-13 - Congestion Control ANS

1. What primary problem did the 1980s Internet face before TCP implemented congestion control?
A. Low link utilization due to conservative sending rates
B. Congestion collapse from hosts retransmitting aggressively after loss
C. Frequent routing loops caused by distance-vector instability
D. Packet corruption dominating over packet loss

ANS: B – Without congestion control, TCP senders kept retransmitting lost packets at high rates, flooding links and causing congestion collapse around 1986.

2. In the router-queue model from Lecture 12, what happens to average packet delay as load approaches 100%?
A. Delay stays constant while loss increases linearly
B. Delay decreases because queues drain faster
C. Delay grows rapidly even before hitting 100% load
D. Delay is unaffected by load in a FIFO queue

ANS: C – For bursty arrivals, average delay grows sharply as utilization approaches capacity, because packets spend longer waiting in the queue even before persistent loss.

3. Why is congestion control fundamentally described as a resource allocation problem?
A. It chooses which routing paths packets follow
B. It determines how bandwidth is divided among competing connections
C. It decides which packets must be encrypted for security
D. It configures router buffer sizes at startup

ANS: B – Bandwidth is a limited resource and each connection demands some share; congestion control decides how much rate each flow actually gets.

4. Which of the following is NOT listed as a design goal for a congestion control algorithm?
A. Avoid congestion by minimizing delay and loss
B. Maximize link utilization (efficiency)
C. Ensure fairness among connections
D. Guarantee zero packet loss under all conditions

ANS: D – The goals are avoiding congestion, efficiency, fairness, and practicality; zero loss is unrealistic and not a design goal.

5. Why is the "dynamic adjustment" approach especially suitable for the deployed Internet?

- A. It assumes routers coordinate globally using a central controller
- B. It requires no prior knowledge of available bandwidth or payment model
- C. It relies on per-flow reservations at every router
- D. It eliminates the need for flow control at receivers

ANS: B – Dynamic adjustment lets end hosts infer congestion from feedback without reservations or pricing, fitting the already-deployed best-effort Internet.

6. In the host-based dynamic adjustment approach, which three components must every sender implement?

A. Route discovery, header compression, and packet scheduling
B. Initial-rate discovery, congestion detection, and rate increase/decrease rules
C. Fair queuing, ECN marking, and bufferbloat detection
D. Error-correcting codes, flow labeling, and ACK pacing

ANS: B – The sketch decomposes the problem into discovering an initial safe rate, detecting congestion, and reacting by adjusting the rate up or down.

7. What signal does classic TCP use as its primary indication of congestion?
A. Explicit ECN bits set by routers
B. Growth of RTT beyond a threshold
C. Packet loss detected via missing ACKs or receiving multiple duplicate ACKs
D. Application-level throughput drops
ANS: C – TCP is loss-based: it infers congestion when packets are not acknowledged and time out or when multiple duplicate ACKs indicate loss.

8. Why can opening many parallel TCP connections be considered a cheating strategy?
A. Because TCP shares bandwidth per host rather than per connection, so multiple connections are treated as one
B. Because each connection receives its own share of bandwidth, so a user with many connections captures more total bandwidth than a user with one
C. Because routers automatically deprioritize packets from users with multiple simultaneous flows
D. Because duplicate ACK detection is disabled across parallel connections

ANS: B – TCP congestion control operates per connection; bandwidth is divided roughly equally among connections, so a host running N connections can receive up to N times the share of a host running one connection.

9. What is the main purpose of TCP slow start?
A. Keep the window small to minimize queuing delay
B. Quickly discover an approximate available bandwidth starting from a low sending rate
C. Guarantee fairness between flows of different RTTs
D. Prevent all packet loss during connection startup

ANS: B – Slow start begins with a conservative window and doubles it each RTT to rapidly approach a safe sending rate without prior bandwidth knowledge.

10. In slow start, what rate-adjustment behavior causes the window to grow exponentially?
A. Adding a fixed number of packets to CWND each RTT
B. Halving CWND on every ACK
C. Doubling CWND every RTT by adding one packet per ACK in that RTT
D. Keeping CWND constant and only adjusting the timeout

ANS: C – Each ACK increases CWND by one packet; over one RTT with CWND packets, this leads to roughly doubling CWND each RTT.

11. What distinguishes AIMD from AIAD, MIAD, and MIMD in the fairness analysis?
A. AIMD always keeps CWND fixed while the others change it
B. AIMD is the only scheme that converges to a fair allocation

C. AIMD guarantees zero loss while others do not
D. AIMD is the only scheme that can be implemented at routers

ANS: B –AIAD preserves absolute gaps, MIMD preserves ratios, MIAD amplifies unfairness, while AIMD shrinks the gap and converges to fairness.

12. In the two-flow rate-adjustment graph, what does the "efficiency line" X + Y = C represent?
A. The set of allocations where one flow gets all bandwidth
B. Allocations where total used bandwidth equals link capacity (full utilization of the link)
C. Allocations with minimal delay and no loss
D. Allocations that are guaranteed fair

ANS: B – The efficiency line X + Y = C denotes full utilization of the link; points above are congested, below leave capacity unused.

13. What is a major benefit of Explicit Congestion Notification (ECN) over pure loss-based congestion signaling?

- A. It guarantees zero queueing delay on all links
- B. It lets routers warn senders about congestion before queues overflow and packets are dropped
- C. It removes the need for ACKs entirely
- D. It makes TCP throughput completely independent of RTT

ANS: B – ECN uses a header bit set by congested routers to signal impending congestion before queues are full, reducing delays and avoiding the confusion between corruption-induced loss and congestion-induced loss.

14. Why can opening many parallel TCP connections be considered a cheating strategy?
A. Because TCP shares bandwidth per host rather than per connection, so multiple connections are treated as one
B. Because each connection receives its own share of bandwidth, so a user with many connections captures more total bandwidth than a user with one
C. Because routers automatically deprioritize packets from users with multiple simultaneous flows
D. Because duplicate ACK detection is disabled across parallel connections

ANS: B – TCP congestion control operates per connection; bandwidth is divided roughly equally among connections, so a host running N connections can receive up to N times the share of a host running one connection.

15. In AIMD, what happens to the difference between two flows' rates when both decrease on loss?
A. The difference doubles
B. The difference stays the same
C. The difference halves
D. The difference becomes zero immediately

ANS: C – When both halve their rates, the difference becomes 0.5·(Y − X), so gaps shrink on each multiplicative decrease event.

16. In the event-driven TCP model, which events cause CWND updates?
- A. New ACKs only
- B. New ACKs and duplicate ACKs only
- C. New ACKs, duplicate ACKs, and timeout events
- D. Routing updates and ARP replies

ANS: C – The sender updates CWND when new data is acknowledged, when duplicate ACKs signal possible loss, and when a retransmission timer expires.

17. During event-driven slow start, why does CWND effectively double every RTT?
A. The sender doubles CWND whenever a timeout occurs
B. Each ACK both frees one packet in the window and increments CWND by one, allowing two packets to be sent per ACK
C. The receiver explicitly sends its desired window size each RTT
D. The router increases the rate field in packet headers

ANS: B – Each ACK lets the sender send one replacement packet and also increments CWND by 1, so across one RTT the number of outstanding packets doubles.

18. What role does SSTHRESH play in TCP congestion control?
A. It stores the maximum allowed RTT
B. It records the last observed loss rate
C. It remembers a "safe" window size where slow start should stop and AIMD should begin
D. It indicates the number of duplicate ACKs seen so far

ANS: C – SSTHRESH is set to about half the window at which loss occurred; slow start stops when CWND exceeds SSTHRESH, switching to additive increase.

19. In the fast-recovery example, what is the main problem with naive CWND halving after an isolated loss?
A. It causes persistent routing loops
B. It forces the sender to stop sending until the retransmission is fully acknowledged
C. It increases the number of corrupt packets
D. It makes duplicate ACKs impossible

ANS: B – After halving CWND, the sender’s window may not permit new transmissions, causing it to sit idle until the retransmitted packet is acknowledged.

20. What key idea underlies fast recovery in TCP Reno/New Reno?
A. Use ECN bits to avoid all packet loss
B. Grant temporary "credit" for each duplicate ACK to keep packets in flight
C. Reset CWND to 1 packet after every loss
D. Move congestion control into routers instead of hosts

ANS: B – Fast recovery interprets duplicate ACKs as evidence that packets have left the network, so it temporarily inflates CWND by one per duplicate ACK to keep sending.

21. In the combined TCP-with-congestion-control logic, what happens on three duplicate ACKs?

A. CWND is set to 1 packet and slow start restarts
B. SSTHRESH is set to CWND/2 and CWND is set to CWND/2 + 3, then the leftmost unacked packet is retransmitted
C. RTT estimation is reset and no retransmission is done
D. The receiver increases its advertised window

ANS: B – Three duplicate ACKs trigger fast retransmit and fast recovery: SSTHRESH ← CWND/2, CWND ← CWND/2 + 3, and the oldest unacked packet is resent.

22. What is a major benefit of Explicit Congestion Notification (ECN) over pure loss-based congestion signaling?
A. It guarantees zero queueing delay on all links
B. It lets routers warn senders about congestion before queues overflow and packets are dropped
C. It removes the need for ACKs entirely
D. It makes TCP throughput completely independent of RTT

ANS: B – ECN uses a header bit set by congested routers to signal impending congestion before queues are full, reducing delays and avoiding the confusion between corruption-induced loss and congestion-induced loss.

23. What simplifying assumption is made in the TCP throughput model about loss events?
A. Losses are random and independent of window size
B. Exactly one packet is lost whenever CWND exceeds a fixed Wmax
C. Loss always happens in bursts of multiple packets
D. Loss is caused only by link corruption, never congestion

ANS: B – The model assumes a single connection with fixed RTT and bandwidth where one packet is lost when CWND reaches Wmax, detected by duplicate ACKs.

24. Under the derived TCP throughput model, how does average window size relate to Wmax?
A. It equals Wmax
B. It equals Wmax/2
C. It equals 3/4 · Wmax
D. It equals 2 · Wmax

ANS: C – With a linear increase from Wmax/2 up to Wmax then a drop, the average window over each sawtooth is 3/4 · Wmax.

25. What key qualitative conclusion does the TCP throughput equation draw about flows with different RTTs?
A. Throughput is independent of RTT, so all flows are RTT-fair
B. Throughput is inversely proportional to RTT, so shorter RTT flows get more bandwidth
C. Throughput is proportional to RTT squared
D. RTT only affects slow start, not steady-state throughput

ANS: B – The equation implies throughput ∝ 1/RTT, so flows with smaller RTTs achieve higher steady-state rates, reducing fairness across RTTs.

26. Why do short TCP connections often suffer higher latency than long ones?
A. They always traverse more congested paths
B. They spend most of their lifetime in slow start and may see timeouts instead of fast retransmit due to duplicate ACKs
C. Routers deprioritize short flows by default
D. They cannot use acknowledgments for reliability

ANS: B – Short flows send few packets, often staying in slow start and lacking enough packets for duplicate ACK-based loss detection, so losses lead to costly timeouts.

27. What problem does bufferbloat create in TCP networks?
A. Extremely low utilization because queues are too small
B. Huge queues that add latency before loss signals arrive
C. Inability to distinguish ACKs from data packets
D. Excessive corruption due to noisy links

ANS: B – Large buffers let TCP queues grow very long; by the time loss occurs, packets have already experienced large queueing delays for everyone.

28. Which property of fair queuing is highlighted as a benefit over simple FIFO queues?
A. It guarantees zero congestion on all links
B. It isolates flows so cheaters and RTT differences have less impact on others
C. It removes the need for congestion control at end hosts
D. It is simpler to implement than FIFO

ANS: B – Fair queuing provides per-flow isolation so a misbehaving or high-RTT flow cannot as easily starve others, although it does not eliminate congestion.

29. What is the core idea behind Explicit Congestion Notification (ECN)?
A. Routers silently drop packets earlier than usual
B. Routers set a bit in packet headers to warn of impending congestion before loss
C. Senders mark packets to request higher priority
D. Receivers estimate congestion by measuring jitter

ANS: B – ECN lets routers mark packets with a congestion bit instead of dropping them, so endpoints can reduce rates before queues overflow.

30. Why are congestion control and reliability described as "intertwined" in classic TCP?
A. Both functions are implemented in routers only
B. Congestion windows, ACKs, and timeouts are used simultaneously for rate control and detecting loss
C. Reliability is implemented at the application layer but depends on router ECN
D. Congestion control requires forward error correction instead of retransmission

ANS: B – TCP reuses ACKs and timeouts both to ensure reliability and to drive congestion window adjustments, so modifying one aspect often affects the other.

31. In the max-min fairness example with total capacity C = 10 and demands A = 8, B = 6, C = 2, what is the correct max-min fair allocation?
A. A = 5, B = 3, C = 2
B. A = 4, B = 4, C = 2

C. A = 6, B = 2, C = 2
D. A = 8, B = 1, C = 1

ANS: B – Max-min fairness satisfies low-demand flows first (C gets its full 2), then splits the remaining 8 equally between A and B (4 each), since neither can be fully satisfied.

32. What does a TCP receiver do when it receives 2 identical duplicate packets, both with sequence number 50?
A. Send one packet with ACK number 50.
B. Send one packet with ACK number 51.
C. Send two packets with ACK number 50.
D. Send two packets with ACK number 51

ANS: D – the ACK number is the next expected sequence number, so after receiving packet 50, the receiver expects packet 51 next.