CSCI-GA.2620-001 Midterm Solution

April 6, 2022

Questions

1. What was the primary goal of the Internet when it was designed? Why was this the goal? Also, give 2 examples of specific non-goals of the ARPANET when it was first designed? (4)

Sol: The primary design goal of the Internet was low-effort interconnectivity of existing networks. The idea was to broaden the reach of existing individual networks by making it easy to connect them together into a broader global network. 2 non-goals were security and predictable performance. (Other non-goals: multicast, accountability, etc.)

2. Why did the Internet pick packet switching over circuit switching? (2)

Sol: Packet switching allows us to interleave packets from different flows on the same link without having to explicitly allocate a link to a particular flow. It is also more resilient to switch or router failures because it is easier to route around a failed switch as long as there is some path from the sender to the receiver. In contrast, in circuit switching, the flow needs to negotiate a new path from the sender to the receiver if a switch along the path fails.

3. Give 2 solutions to the problem of IPv4 address space depletion. (2)

Sol: NATs and switching to IPv6.

4. What is the purpose of TCP's retransmission timer. Explain how the retransmission timeout is calculated. (4)

Sol: The purpose is to compute how long to wait before declaring that a packet is lost. At this point, the packet is retransmitted. Overestimating the retransmission time can lead to under-utilization, which underestimating the retransmission time can lead to spurious retransmissions, which if done often enough, can lead to a congestion collapse. The retransmission timeout is computed by combining a moving average estimate of the mean round time along with a moving average estimate of the variance of the round trip time.

5. What was the problem being solved by Mobile IP? (1)

Sol: The problem solved by Mobile IP was to allow users to move from one network (e.g., WiFi) to another (e.g., cellular) while maintaining the same IP address. This would allow the application to function despite user mobility without having to reestablish TCP and UDP connections.

6. What is the difference between multicast and anycast? (2)

Sol: In multicast, a message is sent from a single sender to multiple receivers. In anycast, a message is sent from a single sender to one of multiple possible receivers.

7. Give an example of how BGP can be used to prevent access to the Internet for certain groups of people? (2)

Sol: By advertising a non-existent route to a particular destination prefix (e.g., the Pakistan Telecom YouTube incident).

8. Give an example of how BGP can be used to protect a web server from an ongoing denial-of-service attack against it. (2)

Sol: By advertising a route to a particular destination prefix that goes through an AS (A) that is different from the destination prefix's AS. This way the AS A can bear the brunt of the DoS attack, scrub the attack traffic, and send only the legitimate-looking traffic to the actual destination.

9. Explain the MED attribute in BGP. (1)

Sol: The MED attribute is used by BGP so that one AS (A) can tell another AS (B) how A would like to receive traffic from B destined to a particular destination. This question is relevant when A and B connect at multiple locations and receiving traffic at one of the locations is beneficial to A relative to receiving it at a different location.

10. Explain briefly the working of the congestion avoidance and control algorithm described in Jacobson and Karel's paper. (4)

Sol: The algorithm increases the congestion window by 1 on every ACK (during slow start) and by 1/W on every ACK (during congestion avoidance). When there is a packet loss, the algorithm decreases its congestion window by a factor of 2. Hence, the name additive-increase-multiplicative-decrease.

11. How does BBR solve the congestion control problem? Why is this any better than Jacobson and Karel's approach? (4)

Sol: BBR solves the congestion control problem by maintaining windowed estimates of the round-trip time and bottleneck bandwidth. It uses these estimates to compute an estimate bandwidth-delay product (which it uses to determine its congestion window) and a sending rate, used to pace out its transmissions. BBR does not rely on packet loss to determine congestion, making it a more reliable way to detect and respond to congestion, compared with AIMD TCP.

12. Explain how XCP decouples utilization control from fairness control. Why does XCP use different control strategies for each? (4)

Sol: Utilization control only computes aggregate feedback to ensure that the link is fully utilized. The aggregate feedback depends only on the spare bandwidth and queue size. XCP uses a multiplicative decrease-multiplicative decrease controller for utilization control. For fairness control, the aggregate feedback from utilization control is apportioned across the different flows based on the round-trip times and window sizes present in their headers. XCP uses additive-increase-multiplicative-decrease for fairness control because this is guaranteed to converge to fairness (cf. theoretical work by Chiu and Jain).

13. How does WFQ enforce fairness on a packet-to-packet basis? (2)

Sol: WFQ computes the finish times of packets according to an idealized fair bit-by-bit round-robin schedule and schedules packets in order of these finish times to approximate this bit-by-bit round-robin schedule as closely as possible.

14. Give a reason why WFQ is preferable to XCP. (2)

Sol: WFQ can deal with misbehaving sources (e.g., sources that do not follow the designated end-host protocol) better than XCP.

15. Contrast eBGP and iBGP. (2)

Sol: eBGP is used by different autonomous systems to exchange reachability information between themselves, i.e., what sequence of ASes a packet must follow to reach a particular destination. iBGP is used to propagate information learned from eBGP to other border routers and internal routers within an AS so that all routers within an AS can route to a particular destination.

16. What are the Gao-Rexford conditions for stable Internet routing? Why are these conditions preferable to having a global registry of routes? (4)

Sol: The Gao-Rexford conditions are a set of constraints that can be imposed by ASes locally on their route import preferences (e.g., prefer a route through a customer compared to a route through a peer or a provider). This is preferable to a global registry of routes because the Gao-Rexford conditions can be applied by each AS locally without an AS having to reveal its routing policies to anyone else.

17. What does "stability" mean in the context of XCP paper. How does it differ from stability in the context of the Gao-Rexford paper? (3)

Sol: Stability in the context of the XCP paper means that the queue sizes, input rate, and output rate eventually settle down to nearly constant values without oscillations. Stability in the context of the Gao-Rexford paper means that the routes settle down to a constant sequence of ASes without oscillations between different sequences of ASes.

18. What is the programming abstraction exposed by SDX? (2)

Sol: SDX exposes a virtual layer 2 switch that connects an AS (A) to all other ASes in the IXP that A has a peering relationship with.

19. Give 2 examples of use cases that are easier to achieve using SDX relative to BGP. (2)

Sol: Application-specific peering and inbound traffic engineering.

20. How does Espresso ensure fast reaction to local network events? (2)

Sol: Espresso "fails static" to the last known good route. This way, local failures don't have to be propagated all the way to the central controller for slower feedback.

21. How does DCTCP leverage ECN to control congestion? (3)

Sol: DCTCP reacts to the extent of congestion using the fraction of ECN marks observed in a window of time to decide how to proportionally reduce its congestion window. This is unlike TCP's coarser reaction to congestion where it reacts merely to the presence or absence of congestion.

22. How did DCTCP implement its ECN marking using features already available in switches in 2010?(3)

Sol: DCTCP used the RED algorithm to implement its ECN marking. It did so by setting the lower and upper thresholds of the RED algorithm to the same value: the value at which packets should be marked with the ECN bit.

23. Why does scale-out perform better than scale-up in the context of VL2 in terms of (1) fault tolerance and (2) agility? (2)

Sol: Scale-out topologies offer us a way to build a network with full bisection bandwidth and no oversubscription. This allows us the freedom to place servers wherever we want in the network because the network offers uniformly high server-to-server capacity regardless of the server's physical location in the network. Scale-out topologies also improve fault tolerance by increasing the number of switches at every layer. Hence, if a single switch fails, n-1 other switches can still be used for routing, offering better fault tolerance properties than the 1:1 model of scale-up topologies.

24. When a new flow starts, how does Ethane enforce fine-grained policies for this flow across the network? (2)

Sol: Ethane uses the central controller to determine the policy for this flow and then installs rules in the switches along the path of that flow, where the path is determined by the policy.

25. What is the purpose of the POL-ETH policy language in the Ethane system? Give 2 examples of policies that POL-ETH can represent. Feel free to make up your own syntax if you do not remember the details of the POL-ETH syntax. (4)

Sol: The POL-ETH language's goal is to easily express policies over flows within the network. Any examples similar to Figure 4 of `https://cs.nyu.edu/~anirudh/CSCI-GA.2620-001/papers/ethane.pdf`

26. How does OpenFlow allow a network operator to isolate research traffic from production traffic? (3)

Sol: OpenFlow allows a network operator to create separate slices for research traffic and production traffic, potentially assigning one a higher priority than the other or rate limiting traffic from the research traffic slide or assigning them distinct VLAN IDs.

27. Why does VL2 load balance at the level of flows rather than packets? (2)

Sol: Balancing at the level of packets can cause reordering of packets at the receiver, which can cause retransmissions and lead to reductions in the congestion window.

28. How does VL2 work around the limitations of ECMP to achieve good load balancing? (3)

Sol: VL2 uses VLB to make use of multiple anycast addresses to scale beyond the group size of ECMP.

29. How many servers can a 2-layer full-bisection bandwidth network with k-port switches support? What about a 3-layer network? (4)

Sol: $k^2/2$ and $k^3/4$

30. In VL2, where is the mapping from application addresses to location-specific addresses implemented?(3)

Sol: On the end hosts.