Comprehensive Exam: Networks: Answers (60 points) Closed Book: Fall 2007

Prof. David R. Cheriton

October 31, 2007

Note: these answers are longer and more details than expected for passing the exam. We expect answers to demonstrate a secure understanding of the core focus of each question, not provide detailed/complete explanations.

1. (15 points total) *End to end*

   (a) (7 points) Define the so-called "end-to-end principle" as applied to the Internet and provide an example of how it has been applied.

   Answer: The canonical example is file transfer over the network (think FTP) - although each router along the path from file source to destination could perform a variety of checksums (for instance, to validate packet integrity or that no router along the path has accidentally flipped a bit), this does not protect against disk read/write errors. As such, it is necessary for there to be an "end-to-end" checksum of the file after it has been written to disk, so all of the checksums in the intermediary stages are duplicating effort. On the other hand, if file transfer takes a long time and errors in the medium are common, it may take excessively long to transmit successfully - thus incremental checksums may be merited as a performance optimization.

   (b) (5 points) Britney Spears goes into rehab, blaming her troubles on the presence of "middleboxes" such as web caches, load-balancers, firewalls, VPN gateways in the Internet which she claims compromise the end-to-end principle. Describe to what degree she is right (if any) and wrong, if so. (You can predicate your comments on assumptions of what these various boxes do, if you are not familiar with their operation.)

Answer: Web caches, when implemented correctly, are simply a performance optimization and thus do not violate e2e

Loadbalancers interfere with e2e to some extent by making it impossible for sources to determine which destination their traffic will arrive at. However, because in typical use cases they are controlled by one of the endpoints, it would be hard to claim that this is too flagrant a violation of the spirit of the e2e argument if one considers the larger definition of “endpoint” as content source rather than simply the host at the end of the line

Firewalls, when controlled by a user (think windows firewall) behave from an e2e perspective much like loadbalancers - it’s perhaps a technical violation but because it’s controlled by the “real end” of the communications channel who could choose a different behavior if desirable, this seems reasonable. When implemented to keep users in check, however, as in many corporate environments, they absolutely violate e2e because they can prevent users who happen to know their traffic is safe from communicating if the firewall decides for some reason that the traffic is not safe.

VPN gateways seem to violate e2e in the sense that they maintain state about connections and drop packets which do not authenticate. However, from an end-device standpoint (when implemented well) they should simply seem like one of many hops through the network, and it would be difficult to claim that an end-device is harmed by dropping packets which would not authenticate upon arrival anyway - this is akin to an ethernet device dropping packets whose CRC does not validate. The maintenance of state is more troubling, since it suggests that should the VPN gateway go away and come back the existing end-to-end connection would be forever terminated - but the short-lived nature of most information flow today suggests that a web browser would simply retry after the VPN recovered.

NAT boxes recomputing the TCP checksum are really a violation of E2E, if you view TCP as providing end-to-end reliability.

(c) (3 points) There is a known major risk to high winds coming up when fighting forest fires in steep terrain. In the summer of 1994 in Glenwood Springs, Colo., 13 firefighter died tragically when the wind came in a steep canyon in which they were fighting a forest fire (mirroring a similar trajedy in 1949). In the book ”Fire on the Mountain,” John N. Maclean documents how the firefighters did not receive a revised weather report because of "bureaucratic bungling" at the Bureau of Lands and Mines. How would an "end-to-end" firefighter operate?

Answer: An end-to-end firefighter would explicitly ask for periodic weather reports and decide for himself whether or not it was safe to proceed. Such a weather report would clearly need to include a timestamp so that the firefighter could determine whether the information was current enough. He/she would get out of the area immediately at any time a reliable up-to-date weather report was not available.

2. (15 points total) *Transport Protocol Design*

   (a) (8 points) Describe for each of: a) slow start b) fast retransmit c) AIMD (additive-increase-multiplicative-decrease) what it is, how it works and why it is compelling to include in TCP. (If you do not recognize these terms, describe how TCP supports congestion control in the Internet.)

   Answer:

   Slow start - when a TCP first starts, increase the window faster than it would normally increase (multiplicatively instead of additively) - this allows a TCP session to rapidly fill the pipe in cases where there is no congestion, and reduces the amount of time "wasted" well below the potential bandwidth especially for short connections (of which there are many on the modern internet)

   Fast retransmit: If a single packet is dropped (which can be sensed by receipt of the following packet), trigger the sender to resend that packet by re-acking the packet before it. The sender resends the last unack'ed packet when receiving duplicate acks. If the majority of drops are single-packet drops and reordering is uncommon (which is a reasonable assumption), this provides much more rapid recovery from packet loss than waiting for timeouts.

   AIMD: Increase the number of packets in flight slowly (additively) and reduce it quickly (multiplicatively). For each window transmitted, increase window size by 1. For each packet which is dropped, cut the window size in half. This allows TCP to always eventually fill the pipe (optimize bandwidth) while also backing away quickly when there is contention for bandwidth.

(b) (7 points) TCP performs a 3-way handshake on connection setup and connection teardown. Describe the purpose of each, how it accomplishes that, and what bad things could happen if you went to less mechanism, e.g. 2-way message exchange.

Answer:

Connection setup - SYN, SYN/ACK, ACK. Needs to match up sequence numbers (specified in SYN and SYN/ACK) and also reserve resources to deal with the connection on each side. A two-way handshake can allow a connection to be setup by an old spurious SYN packet in the network. A two-way handshake for connection setup can also allow denial-of-service attacks (it's easy to open lots of connections from arbitrary addresses by synthesizing packets, if you don't need to be able to receive anything at those addresses).

connection teardown - FIN, FIN/ACK, ACK(?). Bidirectional communication implies that each side must announce"no more data to send", and they both also need to hear confirmation. If the confirmation is dropped, then the FIN or FIN/ACK can be resent without harm, but if the FIN portion is dropped the other side of the connection is going to be sitting waiting for data forever. Thus 3 messages are necessary.

3. (15 points total) *Network Routing*

(a) (7 points) Describe how an IP router handles an incoming packet, focusing how it determines which port to send it out (if any) and what changes it makes to the packet (if any).

Answer: IP router receives a packet because the packet is destined for the ethernet address of one of its network interfaces. It then uses the destination IP address encoded in the packet (as well as, possibly, other information including IP source address and incoming interface) to perform a lookup in its local routing table to choose a next-hop ethernet address and outgoing interface, looking for the shortest matching prefix. The routing table entry typically specifies an interface/port for point-to-point links, or, for multi-point links, an interface or port plus information on how to find the destination party (such as by using ARP to find the MAC for some particular IP address). Default-free routers that find no matching entry in their forwarding table may generate an ICMP "destination unreachable" packet and send it to the original packet's source address. It then rewrites the ethernet header so that the packet appears to be sourced from the outgoing interface and destined for the next-hop address specified. It also decrements the TTL of the packet (discarding if too low) as an easy and guaranteed way of preventing routing loops. An IP router typically looks up the destination IP address in the packet's IP header in its forwarding table,

In parallel, a router typically verifies the IP checksum of the packet, checks various IP header options (source-routing, record-route, etc), checks the packet TTL field, and decrements it. Perhaps the router also interprets the TOS field in the packet to place it in different output queues.

More complex router policies can examine arbitrary packet state (TCP ports being a simple and common example) and keep state from one packet to the next.

(b) (8 points) Compare and contrast the three basic routing techniques, namely: flooding, distance-vector and link-state.

Answer: Flooding is highly reliable and simple but has poor scaling properties.

Distance-vector protocols are relatively simple but can take a long time to converge and may not scale that well to large networks either. It can be difficult to drop a route from the routing table.

Link-state protocols propagate full knowledge of all links in the network and can converge faster than distance-vector protocols.

4. (15 points total) *Ethernet*

(a) (5 points) Describe how CSMA-CD (carrier-sensor multiple access-collision detection) provides good throughput with minimal delay in a 10 Mbps Ethernet, being as quantitative as you can. Recall that a minimum Ethernet packet is 64 bytes so a minimum-size packet is roughly 50 microseconds and the speed of light is one foot per nanosecond (but the signal propagates at closer to 50 percent of that speed.)

Answer: When the Ethernet segment is not in use, as determined by listening or sensing the network, CSMA-CD stations start transmitting packets right away, avoiding any latency associated with network scheduling. However, if two stations transmit at the same time, a collision occurs and is detected, allowing the stations to quickly retry rather than relying on higher-level recovery. Any node that observes a collision "jams" the network to announce this fact. Ethernet requires a minimum 64-byte packet size to ensure that the collision is observed by all nodes. At 10Mbps, 64 bytes take 51.2 microseconds to transmit, and with signal propagation of about c/2, travels quite far, 7.5km. If P is the worst-case propagation from one of the network to the other. In the worst-case, another station starts transmission at time P after the first transmitting station starts, so the latter only detects after 2P. This means that, the network diameter can be roughly at most 3km for the protocol to work correctly.

After a collision, senders wait for a random period of time, and try again. Senders use an exponential backoff algorithm, and eventually drop their packets, to avoid livelock.

(b) (5 points) Hillary Clinton, after hearing how successful CSMA-CD was in the original Ethernet, proposes to legislate its use as "universal healthcare" for all networks, including high-speed, low-speed, wireless, satellite, etc. Describe the issues that CSM-CD runs into as you change the speed of the network and go to other technologies such as wireless, switched, etc., again being quantitative.

Answer: CSMA-CD explicitly assumes that everyone on the network can notice when senders collide, which is valid on a wired system, but not for wireless where two nodes may send at the same time and be unable to hear one another but a node in between could see collisions. High-speed and long-distance networks such as gigabit, and satellite, have a significant network diameter measured in transmission bytes. This translates into a very long minimum packet size needed to ensure collision detection. For instance, gigabit with 64-byte minimum-size packets can have a physical diameter of 75 meters. To use gigabit within a 7.5-km city requires a minimum 6400 byte packet size, which would likely waste a lot of bandwidth with small packets and lots of padding. Similarly, satellite networks have very large distances that make it prohibitively expensive to detect collisions after the fact.

With wireless networks, another problem comes up, namely the fact that not all wireless nodes are guaranteed to be reachable from each other, and there isn't necessarily any symmetry. Two far-away nodes on opposite ends of a node in the middle can be both transmitting at the same time, not hearing each other, and more over, the middle node may not be able to communicate with anyone at all if those other nodes don't hear the middle node's transmissions and back off.

(c) (5 points) Describe how an Ethernet packet manages to find the destination port to which it is addressed in a switched Ethernet network. That is, does the Ethernet switch run a layer 2 routing protocol or what. Also, point out 3 key challenges in the way this facility is implemented.

Answer: Ethernet switches map ethernet addresses to output ports. When a packet is received on port P, the switch associates its source address with port P. Any subsequent messages destined for that address are sent out port P. If a packet arrives for an address not in the list, the switch sends it out every port, i.e. flooding. Thus, messages are always delivered and, if one assumes most communication is bidirectional, the switch broadcasts only when absolutely necessary. 1) Routing loops are a real danger when switches broadcast packets. 2) Devices which move from port to port (for instance, when a computer is moved) cause packets to be sent in the wrong direction. 3) If the switch is plugged into another switch there may be many many addresses associated with each port which suggests a potential memory allocation issue.

The MAC address table mechanism is also very insecure, allowing anyone to capture packets destined for other hosts and break their communication by changing MAC table entries.

The End