THE HONG KONG POLYTECHNIC UNIVERSITY

Department of Computing

This is a open-book examination.

(COMP5311)

Internet Infrastructure and Protocols

17 December, 2004 3.5 hours

[Answer at most 7 questions in section A and both questions in section B.]

Section A: Please answer ANY SEVEN questions in this section [8 marks each, making up a total of 56 marks out of 100. Each part under a question, if any, carries an equal weight].

1. Besides using the Dynamic Host Configuration Protocol (DHCP), another way of configuring an IP address to a host is through self-configuration. In fact, the prefix 169.254.0.0/16 has been allocated for this purpose. This set of addresses is referred to as link-local addresses and they can only be used on the same link, i.e., without going through an IP router.

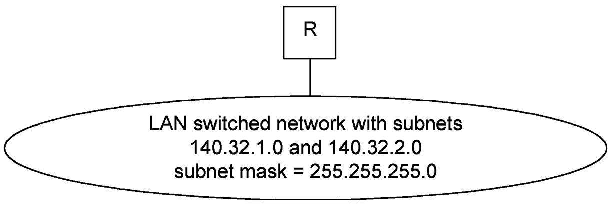
   A host can therefore self-configure a link-local address by itself. But before claiming the self-configured address, it sends out an ARP probe to make sure that the address is not in use. Let the selected address be 169.254.1.1.

   (a) (3 marks) In this ARP probe, what should be the target IP address and target MAC address and why?

   (b) (3 marks) In this ARP probe, what should be the source IP address and source MAC address and why?

   (c) (2 marks) After claiming the selected address successfully, the host is required to announce the claimed address by sending another ARP message. What should be the values of the four parameters in this ARP announcement and why?

2. Consider a switched LAN network configured with two virtual LANs, each of which is configured as a different IP subnet, as depicted below. A router R is connected to allow



   inter-virtual LAN (IP subnet) communications. Write down the routing tables for R, a host A on virtual LAN 1 (IP subnet 1), and a host B on virtual LAN 2 (IP subnet 2), so that the two virtual LANs can communicate with each other. Note that R's interface is configured with 2 IP addresses, say 140.32.1.1 and 140.32.2.1.

   (a) (3 marks) Router R

Destination	Subnet mask	Next-hop
...		

(b) (2.5 marks) Host A

Destination	Subnet mask	Next-hop
. . .		

(c) (2.5 marks) Host B

Destination	Subnet mask	Next-hop
. . .		

3. Consider host A sends IP packets to host B, and they are separated by a number of different IP networks. Assume that IP options are absent. Suppose we know that the path MTU for A to send packets to B is given by P bytes. If host A sends an IP datagram of size D bytes to host B, no fragmentation occurs if $P \geq D$.

   (a) (3 marks) If $P < D$, what is the minimum number of IP fragments as a result of IP fragmentation, in terms of P and D?

   (b) (3 marks) However, the actual number of fragments for this datagram can be larger than the minimum obtained from (a). Give an example to illustrate this claim.

   (c) (2 marks) Give one example that the actual number of fragmentation is equal to the minimum obtained from (a).

4. Recall that another approach to bridge the heterogeneous MTU values across different networks is to let a sender find the path MTU (PMTU), such that the packets will not be fragmented in the network. In other words, we need to design a PMTU discovery algorithm for this purpose. To this end, assume that a new error ICMP message is available to you, which is sent back to the source by a router when the packet is too large to be forwarded to the next hop and the received IP packet has its `DF` bit turned on, i.e. no IP fragmentation allowed. The ICMP message structure is given below, where `MTU` is the Maximum Transmission Unit of the next-hop link.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Code      |          Checksum             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              MTU                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Portion of the original packet that triggers          |
+                       this ICMP message                       +
|                                                               |
```

In point form, design such a PMTU discovery algorithm by using the new ICMP message. In each point, describe the steps taken by the algorithm.

5. Consider a host on LAN A that sends an IP packet to another host on LAN D in Fig. 1. The packet will traverse the $R2-R3$ tunnel and the $R1-R4$ tunnel. Assume that the original packet's don't-fragment (`DF`) bit is on, i.e., no IP fragmentation allowed, and that all the tunnel IP headers also have their `DF` bits on. Consider the following scenarios. Will the sending host receive an ICMP "fragmentation needed but don't-fragment bit set" error message under each scenario and why?

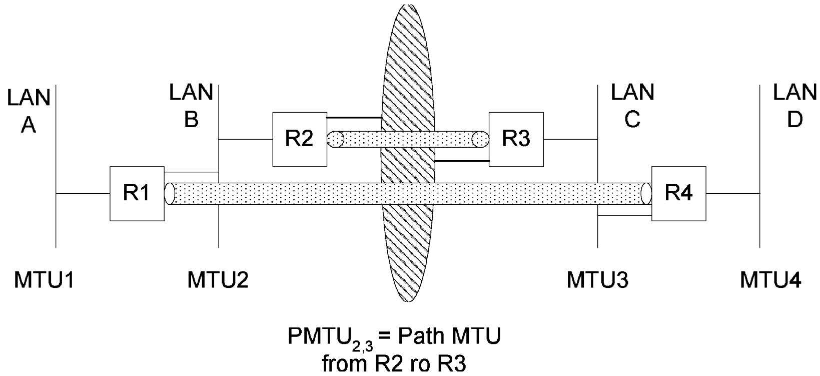


Figure 1: An IP network with 2 tunnels.

(a) (2 marks) The packet is too large to be forwarded to LAN B.

(b) (2 marks) The packet is too large to be forwarded to the network between $R2$ and $R3$.

(c) (2 marks) The packet is too large to be forwarded to LAN C.

(d) (2 marks) The packet is too large to be forwarded to LAN D.

6. Consider that a TCP connection is established between a sender S and a receiver R. The `MSS` is equal to 9KB (kilobytes). Assume that the Nagle's algorithm is turned on and the receiver will send an ACK *immediately* when either the data received is `2MSS` or is at least 35% of the receive buffer size. Otherwise, it has to wait for the firing of the delayed acknowledgment timer before it can send a new ACK.

Consider the following combinations of buffers and determine whether there will be a temporary throughput deadlock. Assume that the sender always has application data to send. Explain your answers.

(a) (2 marks) Send buffer = 8KB and receive buffer = 24KB

(b) (3 marks) Send buffer = 8KB and receive buffer = 16KB

(c) (3 marks) Send buffer = 16KB and receive buffer = 16KB

7. Consider the following traces of a TCP connection which experiences a sudden 3-second delay in the network, starting at around 16s. There are no packet loss events in the traces. There are two main traces: the data segments sent by the sender and the acknowledgments received by the sender.

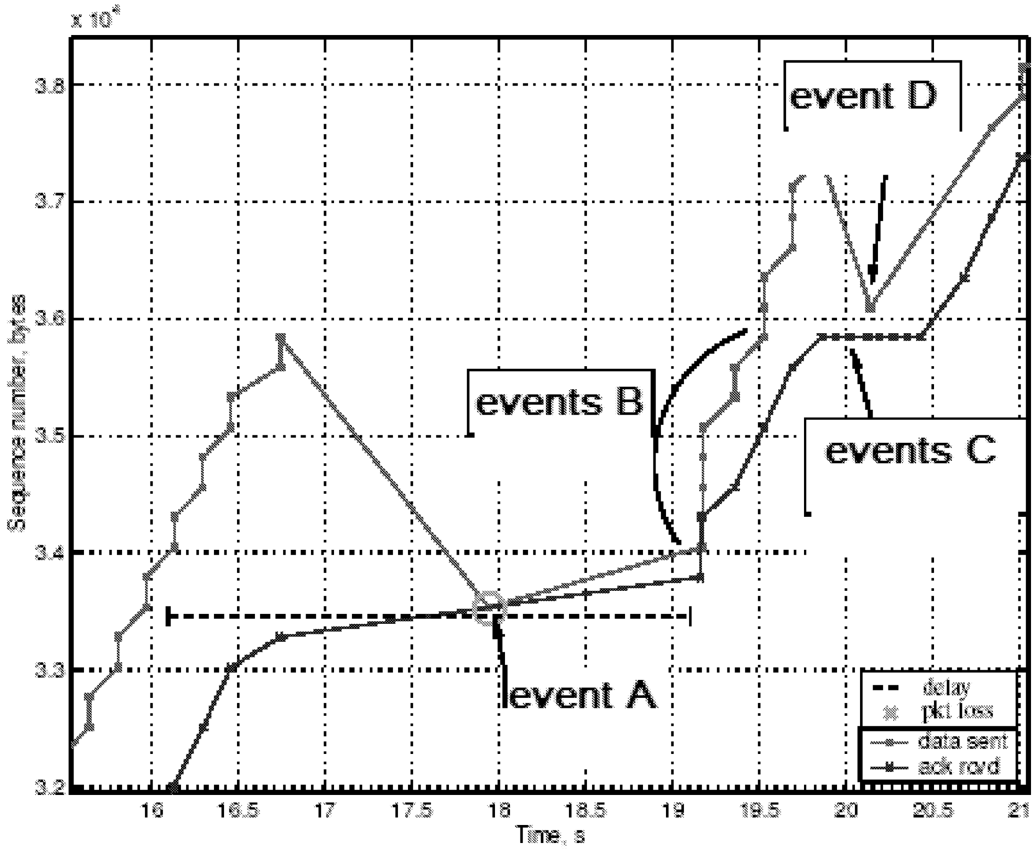


Figure 2: Spurious TCP retransmissions.

(a) (3 marks) Concisely describe the events A, B, C, and D in Fig. 2.

(b) (2 marks) From the traces, what is the receiver's acknowledgment strategy?

(c) (3 marks) Explain why the ACK value remains the same between 19.8s and 20.4s?

8. Consider the routed network in Fig. 3 in which the subnet masks for all subnets are given by 255.255.255.0.

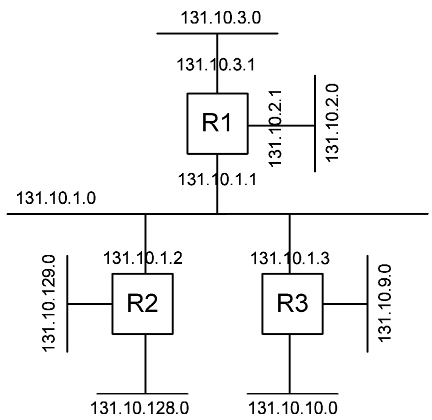


Figure 3: A routed network.

(a) (3 marks) Fill in the next-hop information in $R1$'s forwarding table in Table 1.

Table 1: $R1$'s forwarding table.

Destinations	Next hop	Subnet mask
131.10.1.0		255.255.255.0
131.10.2.0		255.255.255.0
131.10.3.0		255.255.255.0
131.10.128.0		255.255.255.0
131.10.129.0		255.255.255.0
131.10.10.0		255.255.255.0
131.10.9.0		255.255.255.0

(b) (3 marks) "Compress" the forwarding table above by choosing new subnet masks, i.e., reducing the number of forwarding table entry. If there are multiple possible subnet masks to choose from, select the one that has the fewest number of '1's.

(c) (2 marks) If you would like to add a new subnet 131.10.127.0 to the network. Which router would this subnet be connected to and why?

9. Consider the IP network in Fig. 4 in which the routers use RIP-I to share the routing information *with split horizon and poisonous reverse.* Use a hop count of 16 to represent infinity. Assuming that the routing protocol has reached the steady state, write down the distance vectors sent by $R2$ and $R4$ on subnet 4 (use the table format in Table 2). Explain your answers briefly.

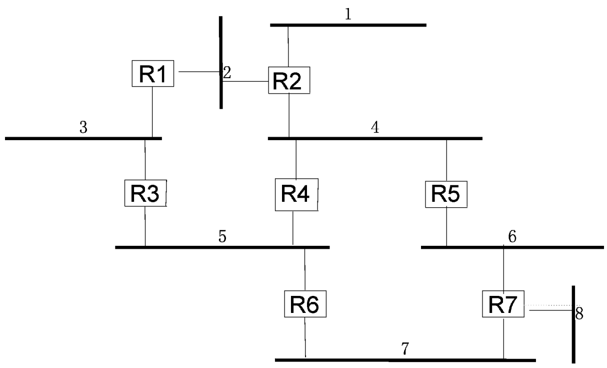


Figure 4: An IP network running the RIP-I protocol.

Table 2: Distance vectors sent by $R2$ and $R4$ on subnet 4.

Destinations (Subnet no.)	Number of hops
1	
2	
3	
4	(not included)
5	
6	
7	
8	

Section B: Please answer both questions in this section [The two questions do not carry the same marks, and they make up a total of 44 marks.]

---

10. (28 marks) Consider the first 5 TCP segments when a Web client establishes a TCP connection with a Web server. Table 3 summarizes the segment information, and please pay attention to the following points.

- Let C_0 and S_0 be the initial sequence numbers for the client and server, respectively.
- $C \to S$ denotes that the segment is sent from the Web client to the Web server, and correspondingly for $S \to C$.
- Assume that the Web server uses W_C as its send window size (because $W_C <$ `cwnd`).
- HTTP request and HTTP reply are the application data sent in the TCP segments' payloads.
- Assume that the HTTP reply's size is larger than $3W_C$.

Table 3: The first 5 TCP segments.

No.	Segment	Sequence Number	Acknowledgment Number	Advertised Window (in bytes)	Segment Type	Payload Length (in bytes)
1	$C \to S$	C_0	0	W_C	SYN	0
2	$S \to C$	S_0	?	–	SYN/ACK	0
3	$C \to S$	C_1	?	W_C	HTTP request	X_1
4	$S \to C$	S_1	?	–	Pure ACK	0
5	$S \to C$	S_1	?	–	HTTP reply	W_C

(a) (4 marks) Fill in the missing acknowledgment values in Table 3.

(b) (8 marks) After the first 5 segments, the client sends two pure ACKs (segments 6 and 7) to the server which are given in Table 4. Assume that the two ACKs arrive at the server in the same order as they are transmitted. Segments 6 and 7 trigger the server to respond with segments 8 and 9, respectively. Fill in the missing items in the table and offer some explanation to your answers.

(c) (8 marks) Now we consider that the two pure ACKs are reordered. That is, segment 7 arrives at the server before segment 6. Therefore, in this case segments **7 and 6** trigger the server to respond with segments 8 and 9, respectively. Fill in the missing

Table 4: The client sends 2 more pure ACKs.

No.	Segment	Sequence Number	Acknowledgment Number	Advertised Window (in bytes)	Segment Type	Payload Length (in bytes)
6	$C \to S$	C_2	$S_1 + W_C$	W_C	Pure ACK	0
7	$C \to S$	C_2	$S_1 + 2W_C$	W_C	Pure ACK	0
8	$S \to C$	?	?	–	HTTP reply	?
9	$S \to C$	?	?	–	HTTP reply	?

information for the server's responses for this case. To address this question, consider the following TCP sender's behavior.

A TCP endpoint checks the ACK based on Eq. (1). If the ACK is acceptable, it will update its send window by setting $\texttt{SND.UNA} = \texttt{SEG.ACK}$. Otherwise, if $\texttt{SEG.ACK} < \texttt{SND.UNA}$, i.e., a duplicate ACK, it will ignore it (we don't consider the fast retransmission here). Moreover, if $\texttt{SEG.ACK} > \texttt{SND.NXT}$, it will drop the segment and send back a pure ACK whose value is specified in the second part of Eq. (2).

$$\texttt{SND.UNA} < \texttt{SEG.ACK} \leq \texttt{SND.NXT}. \tag{1}$$

$$\texttt{SEG.SEQ} = \texttt{SND.NXT},\ \texttt{SEG.ACK} = \texttt{RCV.NXT}. \tag{2}$$

(d) (8 marks) Based on the above, discuss how the Web client can use the 2 pure ACKs to differentiate the following 4 packet reordering cases:

- No packet reordering in both the forward path (client to server) and backward path (server to client).
- Packet reordering only in the forward path
- Packet reordering only in the backward path
- Packet reordering in both the forward path and backward path.

11. (16 marks) Fig. 5 depicts the architecture of an Ethernet-based access network. Briefly,

- The subscriber line is an individual physical connection to each customer network.
- The aggregation network performs aggregation and concentration of customer traffic. Here the aggregation network is Ethernet, which is a broadcast medium.
- The Ethernet access node (EAN) connects the aggregation network and the subscriber lines, which is a layer-2 switch with additional features.
- The access router connects the access network to the Internet.

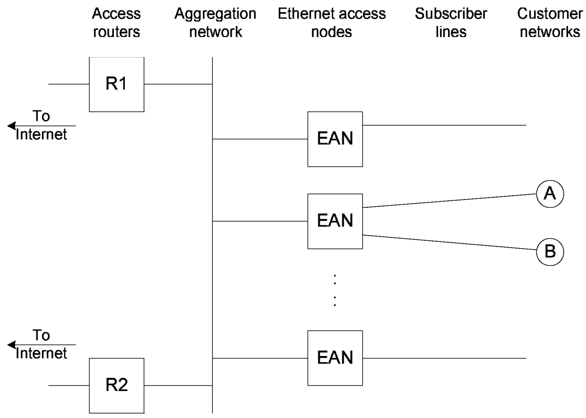


Figure 5: An Ethernet-based access network architecture.

Consider two customer networks A and B in the figure, which, to make it simple, are simply 2 Ethernet hosts situated in 2 different homes. Due to the shortage of IPv4 addresses, it is inefficient to allocate 2 different IP subnets to A and B. Therefore, their assigned IP addresses belong to the same IP subnet. As a result, if the EAN is simply a Ethernet switch, A can send traffic directly to B and vice versa, which obviously present a serious problem to security and privacy.

Therefore, the traffic for these 2 hosts must be "separated." To solve this problem, one solution is to enforce all traffic generated from A, for example, to go to an access router first, regardless of the actual destination. In the rest of this question, we assume that A's access router (and default router) is $R1$, and B's access router (and default router) is $R2$. Moreover, each EAN knows and remembers the access router information for the IP addresses assigned to its customer networks.

(a) (4 marks) To implement the solution, the EAN has to implement the proxy ARP function. For example, if A arps for B's MAC address, EAN intercepts the message and discovers that the target IP address is in the same subnet of A. It then replies with an ARP reply back to A. What is the target MAC address recorded in the ARP reply? Explain your answer.

(b) (4 marks) Trace the path where the packet was sent from A to B after A is equipped with the target MAC address learnt from ARP in part (a).

(c) (4 marks) Besides the proxy ARP function, what else does an EAN need to perform when receiving an unicast Ethernet frame from the subscriber lines and why?

(d) (4 marks) Normally a router is equipped with the ICMP redirect function. However, the access routers Fig. 5 have to switch off the ICMP redirect function. Why?