THE HONG KONG POLYTECHNIC UNIVERSITY

Postgraduate Schemes

Subject Code : COMP526

Subject Title : Internetworking Protocols and Software

Session : Semester 1, 2003/2004

Date : 12 December, 2003 **Time**: 18:30 to 22:00

Time allowed : 3.5 hours **Subject Examiner(s)**: Dr. Rocky K. C. Chang

---

This question paper has a total of 9 pages.

Instructions to Candidates :

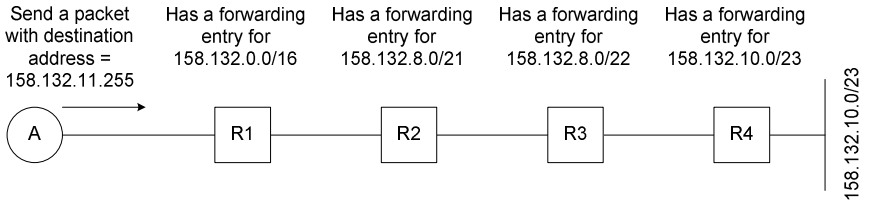
This is an open-book examination.

Available from Invigilator :

None

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. As depicted in the following diagram, host A sends an IP broadcast packet to a subnet 158.132.10.0/23. Assume that all routers are configured to drop any broadcast IP packet.
   (a) Describe the fate of this IP broadcast packet.



   (b) Modify the following IP forwarding algorithm to include dropping of IP broadcast packets.

```
D = Destination IP address in a datagram
Search each entry (network/subnet ID, subnet mask, next-hop) in a
decreasing order of prefix length
D1 = Subnet mask & D
if (D1 == Network/subnet ID)
   if next-hop is an interface
      deliver the datagram directly to destination (ARP D)
   else
      deliver the datagram to Next Hop (ARP the next-hop)
```

2. Write down the forwarding table of a host with a single network interface for the cases in (a) and (b). Please use the following forwarding table format:

Destination Address	Subnet Mask	Next-Hop Address

   If the next-hop is the destination, use the interface's IP address in the `Next-Hop Address`.
   (a) The host is configured with only a single link-local address, say 169.254.1.1.
   (b) In addition to 169.254.1.1, the host is also configured with a non-link-local address, say 158.132.1.1 on a subnet 158.132.1.0/24. There is a router in the subnet with IP address 158.132.1.10.
   (c) If a router receives a packet with the source address as a link-local address, would the router forward it, why or why not?

3. A router's forwarding table consists of 5 prefixes: 0.0.0.0 (default), 64.0.0.0, 160.0.0.0, 208.0.0.0, and 232.0.0.0.
   (a) A binary trie data structure is used for storing the prefixes in a forwarding table. Draw the resulting binary trie data structure for the forwarding table.
   (b) Apply the path compression method to the tree in (a) and draw the resulting tree.

4. The current Nagle algorithm can be described as follows:

   If a TCP sender has a small segment (less than MSS) to transmit and if any previous segment has not yet been acknowledged, do not transmit the small segment.

   An Internet draft proposed to modify the Nagle algorithm to

   *If a TCP sender has a small segment (less than MSS) to transmit and if any previous **small** segment has not yet been acknowledged, do not transmit the small segment.*

   (a) What is the potential advantage of this modified Nagle algorithm? Use a simple example to illustrate your answer.

(b) The Nagle algorithm can be implemented by

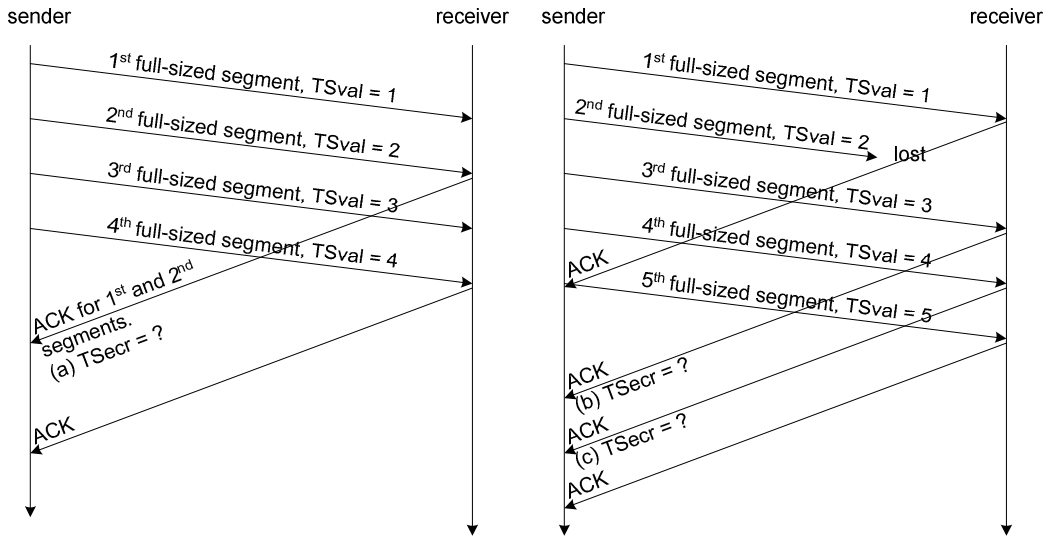
```
Before transmitting a small segment, check
while ()
  If (snd_una == snd_max)
    send the small segment if the send window permits
  Else
    Hold the small segment
```

Write a similar pseudo-code for implementing the modified Nagle algorithm (you may need a new state variable).

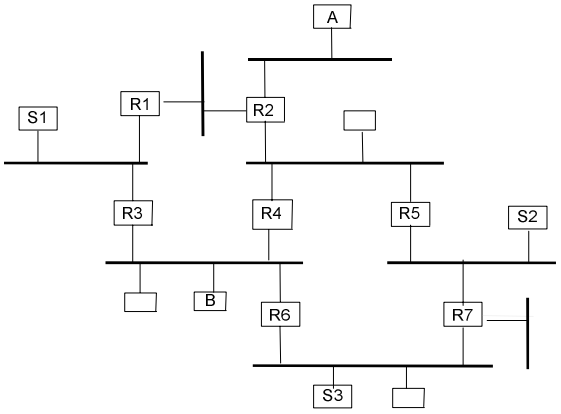
5. Many TCP implementations today support a timestamp option which, among other advantages, can provide a more accurate RTT estimate. For example, a sender sends out a data segment with a timestamp of 1 (in the `TSval` field), which is included as a TCP option, and the receiver echoes the segment's timestamp in the ACK (in the `TSecr` field). This is also done similarly in the other direction (from sender to receiver). For example,

```
Sender                                              Receiver
    <Data segment 1, TSval=1, TSecr=120> ------>
   <---- <ACK for segment 1, TSval=127, TSecr=1>
    <Data segment 2, TSval=5, TSecr=127> ------>
   <---- <ACK for segment 2, TSval=131, TSecr=5>
```

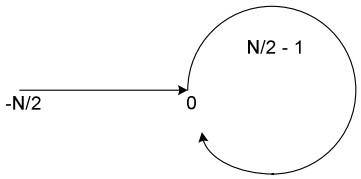
In order to minimize the state (the timestamp values) kept in the receiver, the receiver should retain at most one timestamp value at any time. Consider the following scenarios and what should be the value of `TSecr` for (a)-(c) in order to provide a more accurate RTT estimate? Please explain your answers.



6. In the current DNS, there are 13 identical DNS servers. In order provide highly resilient name service, all the servers are configured with the same IP address and a DNS client is supposed to send a DNS query to a nearest server, i.e., IP anycasting. Similarly, servers S1, S2, and S3 in the following diagram are configured with the same IP address *X*, and the routers use RIP for unicast routing. Show clearly how hosts A and B send packets to the nearest IP address *X*. You must consider all the possible default router configurations, if there are multiple routers to choose from.



7. Some link state routing protocol employs a lollipop sequence space as depicted below.



The linear part is for providing unique sequence numbers when a router boots up. Whenever a router receives a link state advertisement, it needs to compare two sequence numbers `x` and `y` and determine which one is newer (one in the advertisement and another in its database). Give the conditions where `y` is newer than `x` for the following three cases and give the reasons.

(a) $x < 0$
(b) $x > 0$ and $x < y$
(c) $x, y > 0$ and $x > y$

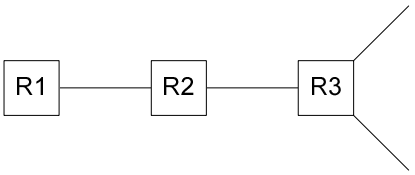
8. Consider a link state routing protocol that does not use periodic link state advertisement and aging mechanism. To make it simple, assume that the sequence number space is unwrapped and it starts with SN = 0, which is used as an initial number as well as when a router is booted up.

(a) Consider that Rj receives a link state advertisement, which is originated by Ri. And the link state in Rj's database is identified by SN = `m` and the one in the received advertisement is identified by SN = `n`. Fill in the blank.

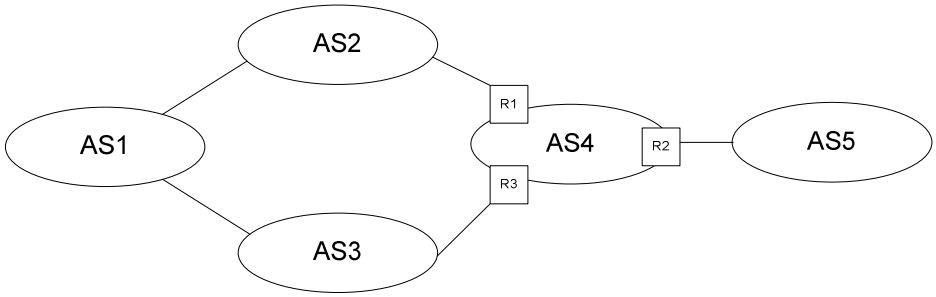
```
If (n ≤ m) ignore and drop the link state advertisement.
Else {
   If j ≠ i,
      replace the link state in the database by the received advertisement,
      and send a copy to all other links except the one that receives the
      advertisement.
   If j = i,

      _____________________________________________________________________
}
```

(b) Consider the following example where all the link states are identified by SN = 0. Moreover, link (R2, R3) goes down and then link (R1, R2) goes down, i.e., R2 is considered down. After link (R2, R3) is restored, R2 and R3 exchange their link states in their databases (bringing up adjacency). After exchanging the link states, use the rules in (a) to show whether R2 and R3 have the same view of (R2→R1).



9. Consider the following AS topology where each link represents a BGP connection and R1-3 are BGP speakers.



- AS1 is advertising 180.180.1.0/24 to both AS2 and AS3.
- AS2 is advertising both 180.180.1.0/24 and 180.180.0.0/16 into AS4.
- AS3 is advertising 180.180.1.0/24 into AS4.

When R1 receives both 180.180.1.0/24 and 180.180.0.0/16 prefixes, it will attach the route for 180.180.1.0/24 with a NO_EXPORT path attribute, and set its LOCAL-PREF (local preference) attribute to a high value, and then propagates this through AS4. The NO-EXPORT attribute is to notify other BGP speakers in the same AS not to further announce the route to other ASes.

(a) If AS5 sends a packet to 180.180.1.0/24 in AS1, what is the forwarding path for this packet? Specify the forwarding path inside AS4 as well.

(b) Repeat (a) when the link (BGP connection) between AS1 and AS2 is broken.

10. IP fragmentation may create problems for packet filtering performed by firewalls. When an IP datagram is fragmented, only the first fragment (indicated by an offset field of 0) contains the transport layer information and other fragments don't.

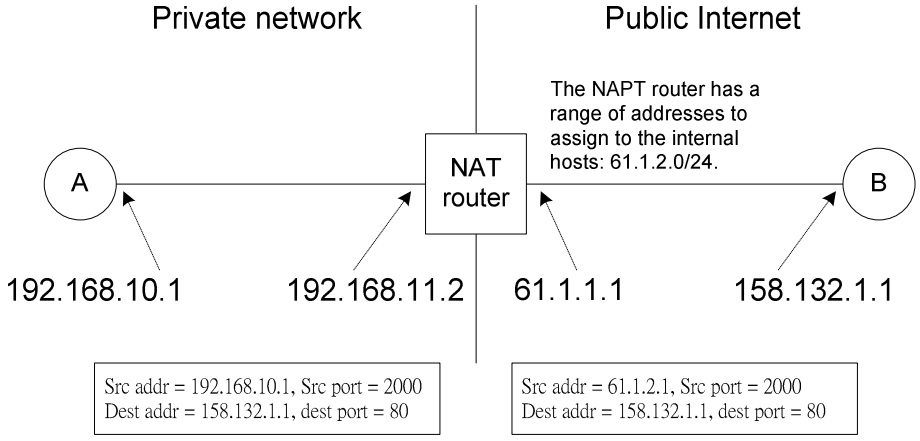
(a) Describe the problem with fragmented IP datagrams using the filtering rule below.

Direction	Source address	Destination address	Protocol	Source port	Destination port	Action
In	External	Internal	TCP	> 1023	≤ 1023	Permit
Any	Any	Any	Any	Any	Any	Deny

(b) Propose a simple solution to solve the problem (e.g., add a new filtering rule). Comment on the security of your solution.

Section B: Please answer ALL questions in this section [Each question carries 22 marks, making up a total of 44 marks.]

11. Private IP addresses (e.g., 192.168/16) have been used by many small and medium size companies to connect their networks to the public Internet. As a result, traditional Network Address Translation (NAT) routers are placed on the network border to translate the private addresses (and possibly ports) in the outgoing packets to public addresses, as depicted in the following diagram. The return packets will be subject to similar translation before delivering to 192.168.10.1.

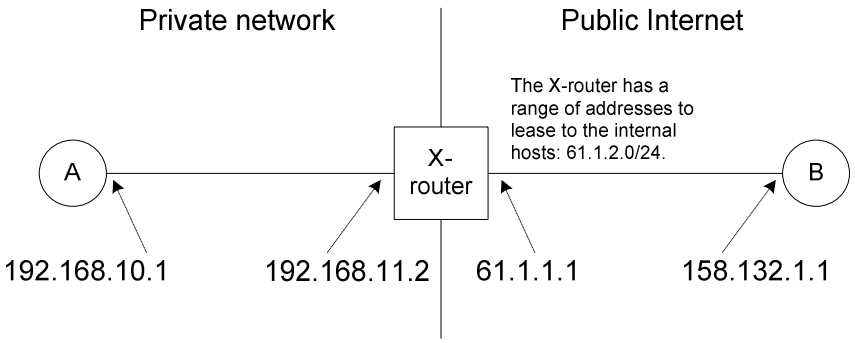


(a) (1 mark) Besides the address translation, what other information in an outgoing packet will be subject to modification by the NAT router?

(b) (1 mark) Is the TCP connection between A and B still considered end-to-end? Explain why or why not.

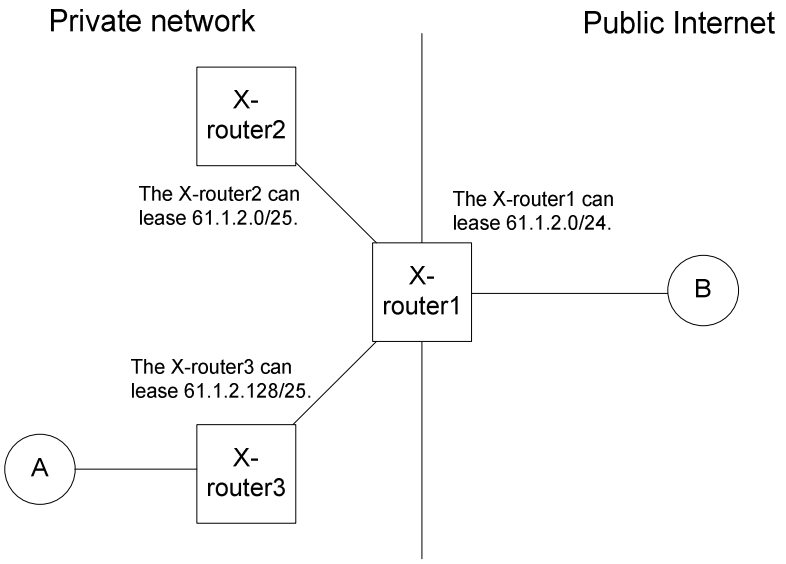
Although NAT is a very popular solution to connecting private IP networks to the public Internet, it suffers from a number of problems. As a result, some developers proposed another approach, which we simply call *X-IP*. This approach can be best explained by a similar example as above, in which

- A first negotiates with the X-router to obtain a public IP address (and possibly a port number).
- After granting the address, the X-router creates a binding of A's private address and the leased public address.
- Upon receiving the address information, A may send packets to B by first tunneling the original packet to the X-router which then detunnels the packet and sends it out.
- The returned packets are also tunneled back to A by the X-router.
- When the connection to the public Internet is no longer needed, A will send a message to the X-router to return the leased address.



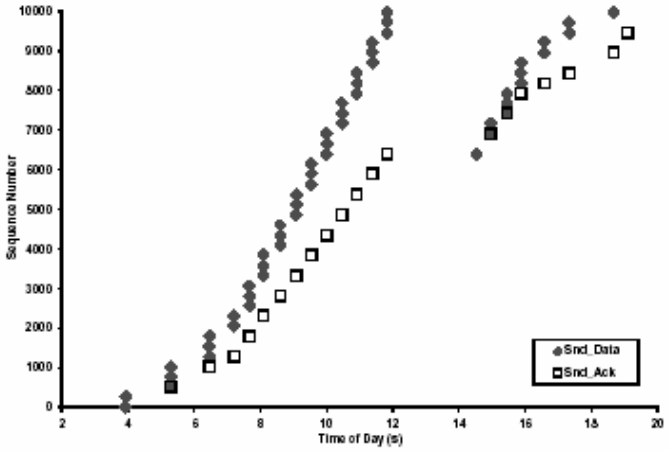
(c) (2 marks) Draw the IP packet sent out by A to the X-router (use the same port numbers as before).
(d) (2 marks) Is the TCP connection between A and B still considered end-to-end? Explain why or why not.
(e) (2 marks) How does the X-router verify that a tunneled packet received from an internal host is a valid one?
(f) (2 marks) How does the X-router verify that a packet received from the public Internet is a valid one?
(g) (2 marks) If the packets from B to A are fragmented some where in the public Internet, where will the packet reassembly be performed and why? You must consider the general case of leased address and port.
(h) Special attention must be paid to the processing of ICMP messages. Describe the correct actions taken by the X-router for the following cases. Assume that a public IP address is already assigned to A. You must consider the general case of leased address and port.
   i. (1 mark) A tunneled ICMP echo request message is received from the private network.
   ii. (1 mark) An ICMP echo reply message is received from the public Internet.
   iii. (1 mark) An ICMP echo request message is received from the public Internet.
   iv. (1 mark) An ICMP error message is received from the public Internet.
   v. (1 mark) A tunneled ICMP error message is received from the private network.
(i) (5 marks) Moreover, X-IP supports a hierarchical distribution of public IP addresses, as depicted in the following diagram (for simplicity, we do not show the addresses).
   - Unlike the previous case, here X-router2 and X-router3 are dedicated the responsibility of leasing a block of public IP addresses.
   - An internal host may contact the nearest X-router for the lease of public IP addresses.

   Describe step-by-step how host A, after receiving a public IP address, sends out a packet to B, and how a returned packet is sent back to A.



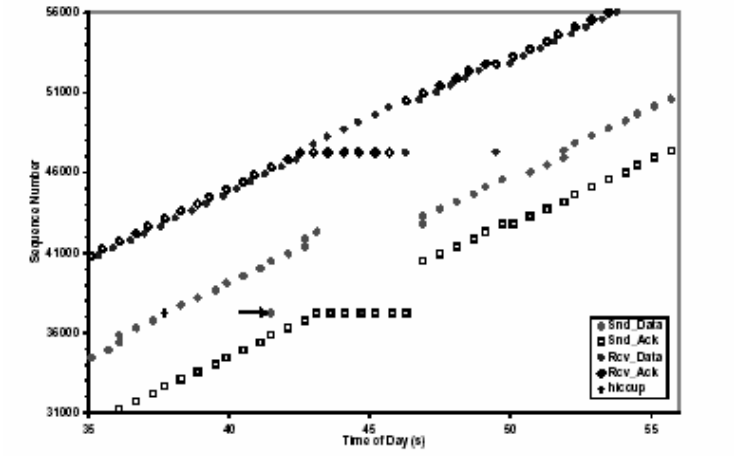
12. In this question, we consider the *retransmission ambiguity problem* in TCP, i.e., a TCP sender's inability to distinguish an ACK for the original transmission of a segment from the ACK for its retransmission. It turns out that the retransmission ambiguity problem could significantly degrade TCP performance.

    A specific problem arising from the retransmission ambiguity problem is known as *spurious timeouts*, i.e., timeouts that would not have occurred had the sender waited long enough in terms of time. A spurious timeout occurs when the RTT suddenly increases, to the extent that it exceeds the RTO, which is explained in the following diagram. SACK is not used in the example.
    - There are no packet losses, but some of the ACKs were delayed with the second batch of ACKs arriving at around 15 seconds.
    - A timeout occurs around 14.5 seconds, and a retransmission took place.
    - On the receipt of the first ACK after the timeout, the sender must interpret this ACK as acknowledging the retransmission, and that all other outstanding segments have also been lost.
    - Therefore, the sender retransmitted other segments in the entire window.



(a) (5 marks) How would the packet trace be different (please just explain and you do not need to produce the trace), if the sender is somehow able to tell that the first ACK after the timeout event actually comes from the original transmission?

Another problem arising from the retransmission ambiguity problem is known as *spurious fast retransmits*, i.e., fast retransmissions that would not have occurred had the sender waited long enough in terms of the number of duplicate ACKs. Spurious fast retransmits occur when packets get reordered beyond a threshold (normally set to 3) before reaching the receiver. An experiment was set up to simulate a packet reordering case, and the following diagram captures the packet trace result (you don't need to worry about the receiver-side states).



- At around 37.7 seconds, a segment was intentionally delayed (indicated by +). The next 6 segments were allowed to go through, and then followed by the delayed segment (indicated by the arrow).
- A fast retransmission occurs at around 45 seconds upon receiving the third duplicate ACK.

(b) (5 marks) How would the packet trace be different (please just explain and you do not need to produce the trace), if the sender is somehow able to tell that the first new ACK after the duplicate ACKs actually comes from the original transmission?

Since both spurious timeouts and spurious fast retransmits are caused by the retransmission ambiguity problem. A simple solution is to require every data segment and ACK to carry additional information, such that the sender is able to tell whether an ACK is for the original transmission or the retransmission. One way of implementing this is to use the TCP timestamp option. Please refer to Q5 in section A for the explanation of the option (you don't have to choose Q5 though).

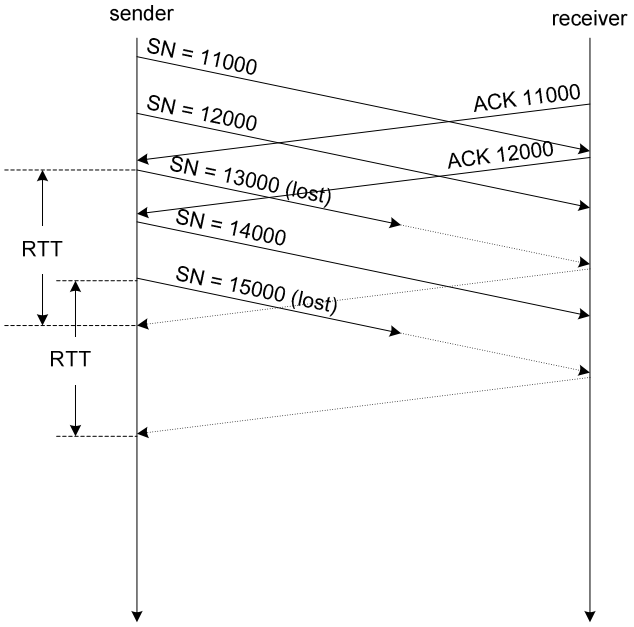
(c) (6 marks) Using the TCP timestamp option, a very smart MSc student proposed an algorithm to eliminating the retransmission ambiguity:
   - The sender always stores the timestamp of the **first** retransmission in a variable called `t_1st_rexmit` independent of whether it was triggered by timeout or fast retransmit.
   - When the first ACK that acknowledges the retransmission arrives, the sender compares the timestamp of that ACK with the value in `t_1st_rexmit`.

   Complete the remaining of the algorithm.

With the TCP timestamp options, one can also implement other more effective retransmission strategies for lost packets, such as the one below:

- When a duplicate ACK is received, check to see if the difference between the current time and the timestamp recorded for the relevant segment is greater than the timeout value. If it is, retransmit the segment without having to wait for 3 duplicate ACKs.
- When a nonduplicate ACK is received, if it is the first or second one after a retransmission, check to see if the time interval since the segment was sent is larger than the timeout value. If it is, retransmit the segment.

(d) (6 marks) Complete the missing information in the following diagram where each data segment sent is 1000 bytes long, and `cwnd` and other states for the TCP connection are not shown.



-- End of the examination paper --