Examination
IK2218 Protocols and Principles of the Internet
EP2120 Internetworking

Date: 07 January 2016 at 08:00–12:00

a) ***No help material is allowed - You are not allowed to use dictionaries, books, or calculators!***
b) *You may answer questions in English or in Swedish.*
c) *Please answer each question on a separate page (not sheet).*
d) *Please write concise answers!*
e) *Put a mark in the table on the cover page for each question you have addressed.*
f) *The grading of the exam will be completed no later than 28 January 2016.*
g) *After grading, exams will be available for inspection online.*
h) *Deadline for written requests for grading review is 15 February 2016.*
i) *Contact person for IK2218 is Markus Hidell, phone 070-249 0252.*
j) *Contact person for EP2120 is György Dán, phone 08-790 4253.*

Important note!

Your grade is F in any of these two cases:

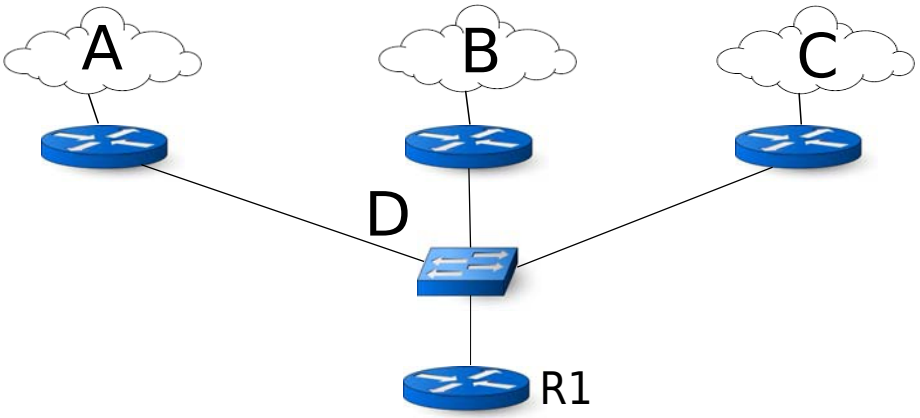
- **if you do not reach at least 10 (ten) points out of 20 for problems 1-4 or**
- **if you reach less than 30 points in total.**

We advise you to start with problems 1-4.

Part I (Problems 1-4)

1. IP and addressing (5p)

Consider the network below, a routed network in an organization's enterprise network. The organization built a core network connected to a central router R1, and connected their edge/access routers with (long-haul) switched Ethernet (network D). Router R1 connects the enterprise network to the Internet through the Internet Service Provider. The access routers are connected to a set of local offices (networks A to C). All networks use Ethernet on the link layer, and the routers use RIPv2 for routing.



Your task is to make an address allocation in the network by allocating an address block to networks A to C in the following way:

1. You need to use the address block 100.40.128.0/21 for address allocation.
2. The networks A–C require 500 hosts each. Create a minimal block for each local office A through C and for network D. Start with the lowest address for network A.
3. There are no unnumbered point-to-point links: all Ethernet networks have a corresponding IP sub-network and all nodes (routers and hosts) have an IP address on each of their network interfaces. All nodes need to be reachable from any other host.

Based on your address allocation, provide the answers to the following:

a) What is the longest prefix length that you could consider for the networks A–C? What is the corresponding netmask? (1p)
b) Give the network address of networks A to D in CIDR notation and their directed broadcast address! What could be a possible IP address for router R1's interface on network D? Motivate your answer. (1p)
c) Router R1 sends a datagram to destination address 224.0.0.9 on its Northern interface. How many hosts or routers will receive this datagram? (1p)
d) Explain the difference between the IPv4 strict source route and loose source route options. Explain the reasons why the usefulness of these IPv4 options is limited in large internetworks. (1 p)
e) When performing reassembly, IPv4 relies on four fields of the IPv4 header, the source IP, the destination IP, the protocol and the identification. Could it be that IPv4 erroneously reassembles fragments that do not belong to the same datagram? (1p)

SOLUTION:

a) The prefix length should be /23 or shorter, the corresponding netmask is 255.255.254.0.
b) With the prefix length of 23 bits, each network can accommodate up to 510 hosts.

Network	Network address	Broadcast address

A	100.40.128.0/23	100.40.129.255
B	100.40.130.0/23	100.40.131.255
C	100.40.132.0/23	100.40.133.255
D	100.40.134.0/29	172.30.134.7

Router R1's IP address could be any IP address in the block allocated to network D, except the network address and the directed broadcast address.

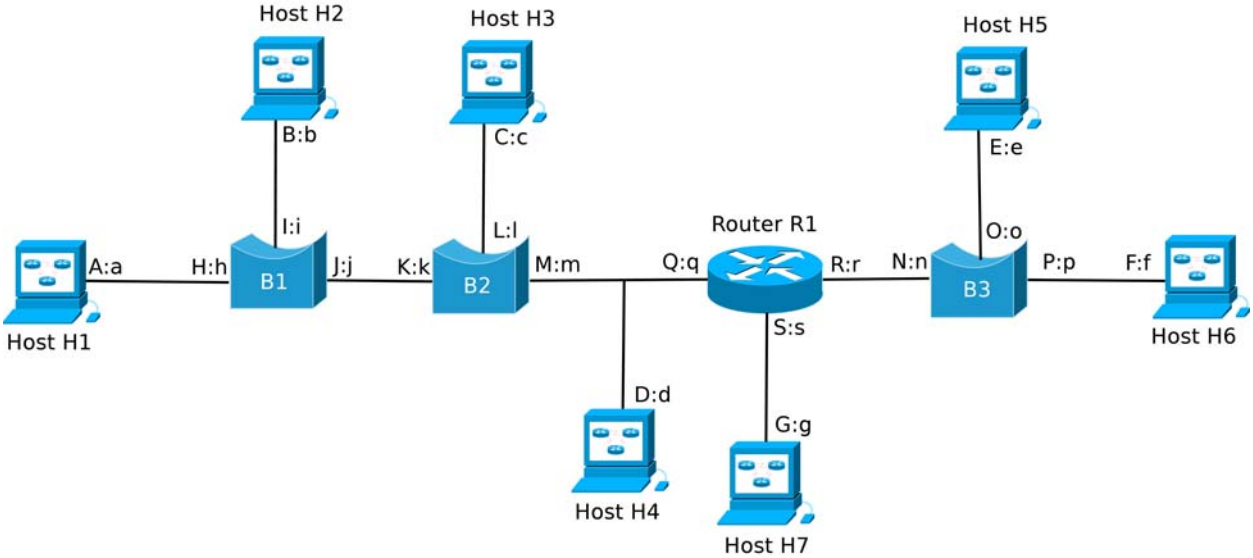
c) Three (3) routers would receive the datagram. The destination address is the all-rip2-routers multicast address.

d) Strict source route specifies all routers that are visited by the datagram. Loose source route specifies a sequence of routers to be visited. There are two main reasons. First, the size of the IPv4 option is limited to 40 bytes, hence only 9 IP addresses fit. Second, for source routing one needs to know the network topology, including router addresses. This information is unlikely to be available. Third, due to security issues, source routing is disabled in most routers (organizations).

e) Yes and no. If the value of the ID field is not repeated by the sender during the maximum datagram lifetime for the same source IP/destination IP/protocol combination then no. Otherwise it could happen. See Rfc 6864 for a thorough discussion.

2. Delivery and address resolution (5p)

Consider an IPv4 network consisting of 7 hosts, 3 bridges and 1 router shown in the figure. Hosts H1 to H7 have one interface and a configured forwarding table each. B1 to B3 are learning bridges. R1 is a router with appropriate routing tables. Logical (IP) addresses are represented by capital letters, physical (MAC) addresses are represented by small letters. All ARP caches and the bridges' learning tables are empty.



a) How does a host determine whether a destination (given with its IP address) is located on the local subnet? When is direct delivery used to deliver a datagram in an IP network? (1p)

b) Identify the subnets of the network. Which of the physical (MAC) and logical (IP) addresses are not needed? (1p)

c) A process on Host H6 sends 150 bytes via UDP to a process on Host H1. Using the

notation in the figure, show the contents of the learning tables and of the ARP caches after the datagram has been delivered. Assume that the process on Host H6 knows the IP address of Host H1, and that ARP snooping (passive ARP learning) is used. (1p)

d) After the process on host H1 received the message sent in c), it replies to the process on host H6 with 100 bytes sent using UDP. Show the new contents of the ARP caches and the learning tables. Assume that ARP snooping (passive ARP learning) is used. (1p)

e) After the process on host H6 received the message sent in d), a process on Host H2 sends a message with 200 bytes via UDP to Host H5. Using the notation in the figure, show the new contents of the ARP caches and of the learning tables after the datagram has been delivered. Assume that Host H2 knows the IP address of Host H5 and that ARP snooping is used. (1p)

Solution:

a) The host performs a bitwise AND of its IP address and its netmask, then a bitwise AND of the destination IP address and its netmask. If the two results are equal, then the destination host is on the local subnet. Direct delivery is used if the destination host is on the same physical network as the host that tries to send the datagram.

b) Subnet 1: A, B, C,D, Q. Subnet 2: E, F, R. Subnet 3: G,S. Bridges B1, B2 and B3 do not need an IP and a MAC address.

c) Contents of the ARP caches are as follows.

H1: Q-q
H2: Q-q
H3: Q-q
H4: Q-q
H5: F-f
H6: R-r
H7: -
B1: q-east, a-west
B2: q-east, a-west
B3: f-east, r-west
R1: F-f, A-a

d) No new content, the current ARP caches and learning tables have enough information to deliver the message.

e) New entries are as follows.

H5: R-r
B1: b-north
B2: b-west
B3: e-north
R1: B:b, E:e

3. IP forwarding (5p)

A router has the IPv4 forwarding table shown below. Determine the next-hop address and the outgoing interface for the datagrams arriving to the router with destination addresses as given in points (a) – (e).

Destination	***Next-hop***	***Flag***	***Interface***
20.114.108.128/27	–	U	m0
130.78.24.0/21	–	U	m1
133.24.96.0/19	–	U	m2
178.131.192.0/18	130.78.30.140	G	m1

Destination	*Next-hop*	*Flag*	*Interface*
153.167.43.208/32	20.114.108.148	UGH	m0
51.77.236.0/25	133.24.119.17	UG	m2
149.72.175.128/32	133.24.117.139	UGH	m2
0.0.0.0/0	130.78.28.221	UG	m1

a) 20.114.108.143 (1p)
b) 153.167.43.208 (1p)
c) 178.131.192.223 (1p)
d) 255.255.255.255 (broadcast address) (1p)
e) 107.255.191.176 (1p)

Solution:
a) direct delivery on m0
b) 20.114.108.148 on m0
c) discarded, because the link is not up (U-flag not set)
d) discarded, because broadcasts are not forwarded
e) 130.78.28.221 on m1 (default route)

4. TCP (5p)

a) In TCP the receiver can control the sending rate through the advertised receive window. What happens when the receiver advertises a receive window of 0 bytes? How does the sender know when to send data? Consider in your answer that segments can get lost. (1p)

Consider two hosts, A and B, connected by a network running IPv4. The capacity of all links is 10Mbps and the round trip time is 400ms. The path MTU is known to be 1540 bytes. TCP on the receiving host has a receiver window size limit of 10500 bytes, which it advertises during connection establishment. The sender uses a value of 65535 for *sshthresh* for congestion control. Delayed acknowledgements (two full sized segments) are used with a maximum delay of 200ms. The receiver can process the data as fast as they arrive.

b) What is the maximum throughput that can be achieved using a TCP connection given these parameters? How big should the receiver window size be in order to be able to fully utilize the channel (not considering congestion control)? (1 p)

A process P_A on host A would like to transmit 19500 bytes to a process P_B on host B using TCP. The active open is performed by A, and since one of the segments is lost during connection establishment, the initial congestion window size is 1xMSS.

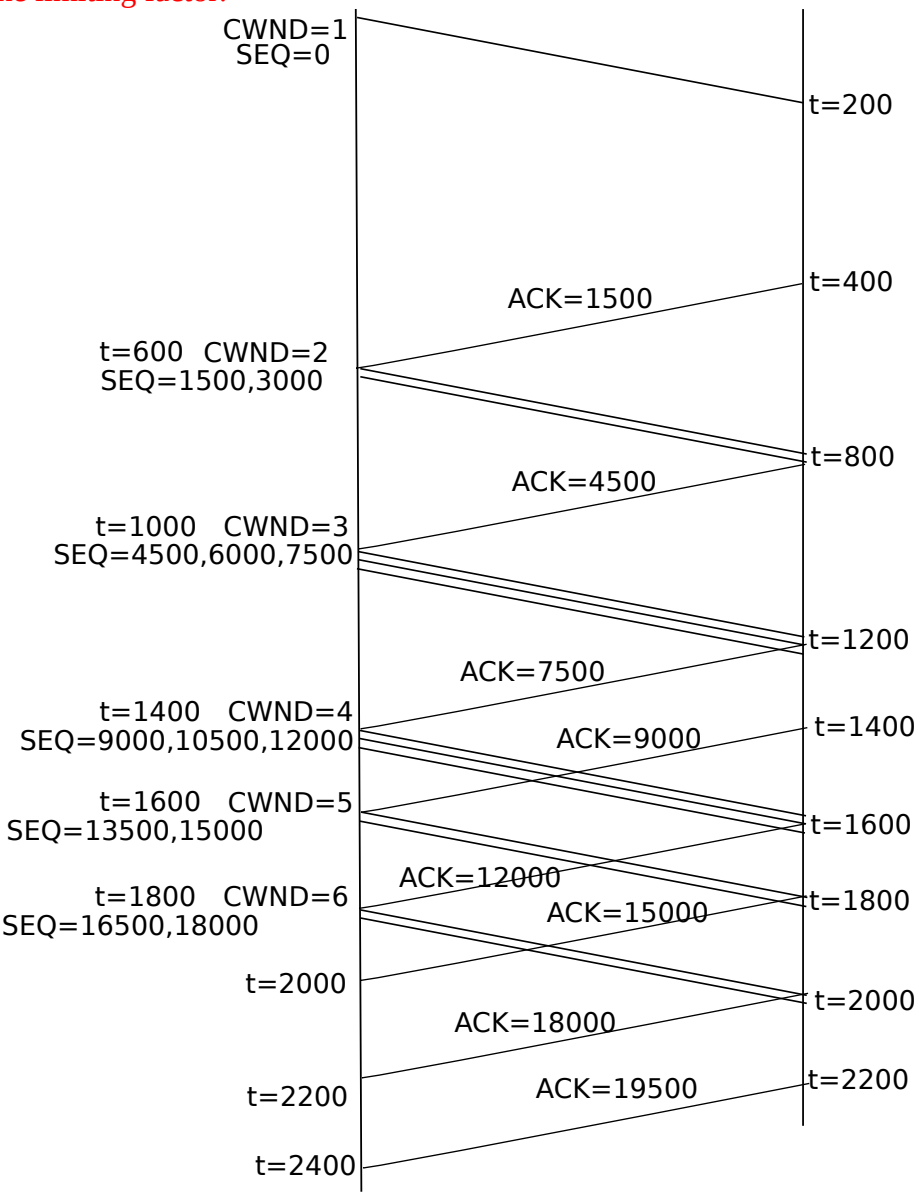
c) How much time does it take to transmit the data from A to B, not including the connection establishment, until the last ACK is received by A? You can ignore the transmission times of the packets, but you should consider the impact of congestion and flow control. If a delayed ACK is to be sent at a time instant when a new segment arrives, the delayed ACK is sent first. Support your solution with a drawing of the segments sent, including the CWND, time sent, ACKed data, etc. (3p)

SOLUTION:
a) When the receiver advertises a receive window size of 0 bytes, the sender cannot send more data. The sender must wait for an advertised window size greater than 0 to be able to send data again. A potential problem is that if the segment carrying the positive advertised window

size is lost, the sender would not be able to continue sending data. To avoid this problem, the sender use a timer called persistence timer, and sends a window probe when the timer expires.

b. The sender can send 10500bytes per RTT, hence the throughput is 10500B/0.4s=10500*8b/0.4s=210000bits/s. The bandwidth delay product is $10*10^6$*0.4s=4Mbit. This is the receiver window size that one would need to be able to fully utilize the channel.

c. The MSS=1500bytes. In the figure time 0 corresponds to the first data segment sent, connection establishment is not shown. CWND is measured in terms of MSS, SEQ corresponds to the first byte of the segment, ACK is the next byte expected. In total the transmission takes 2400ms. Observe that starting from t=1600, CWND>=RWND, and thus RWND is the limiting factor.



A slightly different solution is obtained if one assumes that the delayed ACK at t=1800 is sent before the next segment arrives. That solution is of course correct if it is consistent otherwise (i.e., follows slow-start, rwnd, cwnd, etc).

Part II (Problems 5-11)

5. Fragmentation and UDP (5p)

a) An application wants to transmit 1900 bytes of data via UDP from host A to host B. The UDP header is 8 bytes long, there are no IP options used, the network layer protocol is IPv4. The path consists of two local area networks connected by a router: the MTU of the first network is 1300 bytes and the MTU of the second network is 800 bytes. The path MTU is not known. How many IPv4 fragments arrive at host B if the DF flag is not set? Give the segment sizes, the fragmentation offset and the more fragments (MF) bit of all fragments. (3p)

b) How many IPv4 fragments would arrive at host B if path MTU discovery had been used? Give the segment sizes, the fragmentation offset and the more fragments (MF) bit of all fragments. (1p)

c) In IPv4 reassembly is only done at the destination host. Name two reasons for not doing reassembly in the routers. (1p)

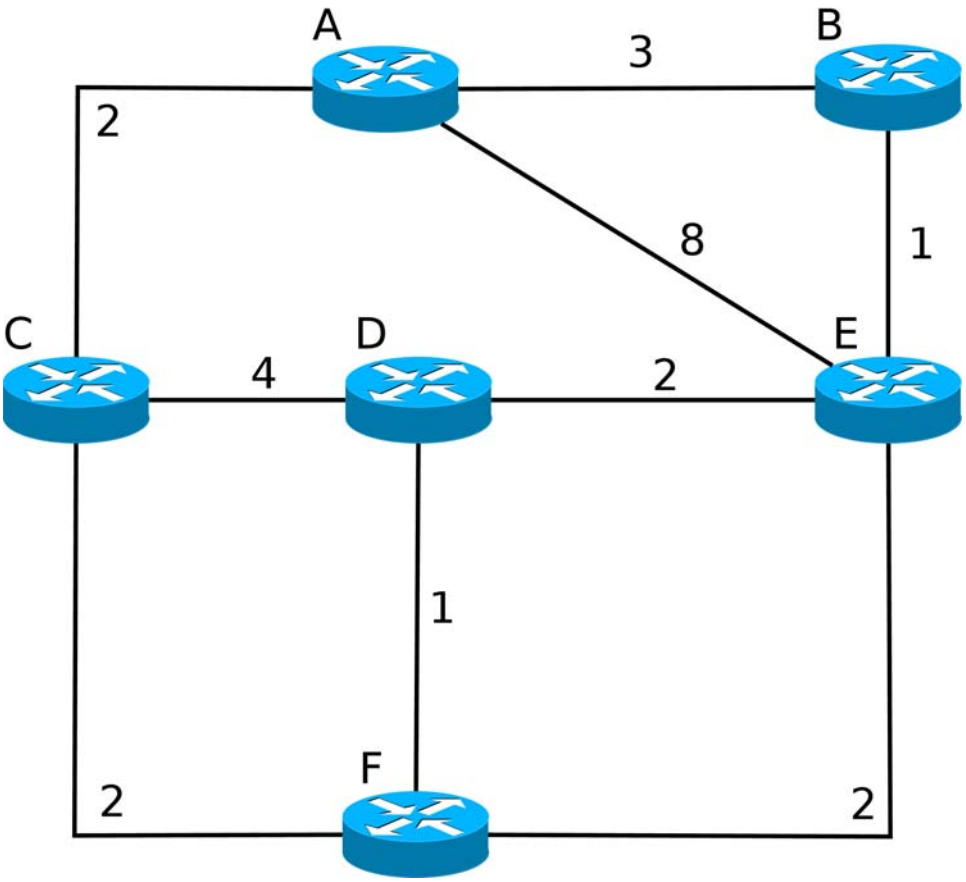
a) In total 1900+8=1908 bytes of IP payload have to be transmitted. The IP payload on the first link can be 1280 bytes. The IP payload on the second link can be 780 bytes, but since 780 is not divisible by 8, it will have to be 776 bytes.
Upon leaving host A, the IPv4 payload, offset and MF values are (all in bytes):
1) 1280, 1, 0
2) 628, 0, 1280
Upon forwarding by the router, the IPv4 payload, offset and MF values are (all in bytes):
1) 776, 1, 0
2) 504, 1, 776
3) 628, 0, 1280

b) With path MTU discovery the IPv4 payload, offset and MF values are (all in bytes):
1) 776, 1, 0
2) 776, 1, 776
3) 356, 0, 1552

c) Reassembly is a complex operation, and would increase the complexity of routers. Fragments do not have to traverse the same path in the network. Link MTUs on a path can be different and a packet could be fragmented and reassembled several times, which would be a waste of resources.

6. *Routing (5p)*

Consider the internetwork shown in the figure below. A – F represent 6 nodes in the internetwork and numbers along the links represent the cost of sending a datagram from one node to another along a particular link.



a) Use Dijkstra's algorithm to find the shortest paths from node A to all other nodes in the internetwork. Your solution should show path and cost for each best route and also the steps taken in the execution of the algorithm. (3p)

b) How can we be sure that Dijkstra's algorithm really finds the shortest path to all nodes? Briefly argue why the algorithm must be correct. (1p)

d) Is Dijkstra's algorithm used by Link-State routing protocols or by Distance-Vector routing protocols? What are the key differences between LS-routing protocols and DV-routing protocols in terms of information availability at the routers and in terms of information exchange between the routers? (1p)

Solution:

a) Start with a permanent set P containing the shortest path to A, and a tentative set T containing the neighbors of A together with the known least cost paths from A to each node. At each step, add the node with the least cost from T to P, and add the neighbors of the added node to T.

P	**T**	**D_B (Path)**	**D_C (Path)**	**D_D (Path)**	**D_E (Path)**	**D_F (Path)**
{A}	{B:3, C:2, E:8}	3 (A-B)	2 (A-C)	–	8 (A-E)	–
{A, C}	{B:3, E:8, D:6, F:4}	3 (A-B)	**2 (A-C)**	6 (A-C-D)	8 (A-E)	4 (A-C-F)
{A, C, B}	{E:4, D:6, F:4}	**3 (A-B)**	**2 (A-C)**	6 (A-C-D)	4 (A-B-E)	4 (A-C-F)

{A, C, B, E}	{D:6, F:4}	**3 (A-B)**	**2 (A-C)**	6 (A-C-D)	**4 (A-B-E)**	4 (A-C-F)
{A, C, B, E, F}	{D:5}	**3 (A-B)**	**2 (A-C)**	5 (A-C-F-D)	**4 (A-B-E)**	**4 (A-C-F)**
{A, C, B, E, F, D}	{}	**3 (A-B)**	**2 (A-C)**	**5 (A-C-F-D)**	**4 (A-B-E)**	**4 (A-C-F)**

b) Let us call the nodes in the permanent set visited, the rest unvisited. Initially, the source node A is included in the permanent set P, and is thus visited, the rest of the nodes are unvisited. In this state the algorithm is correct, i.e., the computed distance to each visited node is the least cost path.
We can use induction to show correctness. Assume, that after n-1 iterations the algorithm is correct. In the next iteration, we choose the univisited node X that has least distance among all unvisited nodes. This distance is minimal for all paths to X that go via visited nodes. If there was a shorter path that includes unvisited nodes, the for the first unvisited node Y on that path we would have dist[Y]<dist[X], and Y would have to be the unvisited node with least distance, a contradiction. Thus, the distance of X must be the shortest path distance.

c) Dijkstra's algorithm used by LS-routing protocols. LS-routing algorithms require each node to have complete knowledge of the network topology and of the costs of the edges. DV-routing algorithms are distributed, as nodes share information only with their immediate neighbors, perform calculations and distribute the results to their neighbors.

7. Electronic Mail (6 p)

In your incoming mailbox, you find an email with the following header lines:

```
To: bargains@special-deals.com
Subject: Special price for you
```

Clearly, the email is not addressed to you, but it still appears in your mailbox.

a) Explain how it possible with the Internet email protocols to deliver an email to someone without having that person's email address in the email header. (2 p)
b) An email message is delivered in several steps. Consider the case when Alice uses an email application on her computer to send an email to Bob. There are two email servers involved. Describe the steps from that Alice sends the email message from her computer until Bob reads the message with an email application on his computer. For each step, explain which protocol is used and which the communicating parties are: who is client, and who is server. (2 p)
c) In each step, the party acting as client needs to know the domain name for the server it should contact. Explain for each step how the client gets the domain name for the server. (2 p)

Solution

a) An email has two parts: envelope (SMTP) and message. The To: and Cc: fields are in the headers of the message, while the actual receiver is specified by SMTP ("on the envelope"). The SMTP information is not visible to the receiver.

b) 1) Alice's email application (client) sends the email with SMTP to Alice's outgoing email server (server). 2) Alice's outgoing email server (client) sends the email with

SMTP to Bob's incoming email server (server). 3) Bob fetches the email from his incoming email server (server) with IMAP or POP using the application on his computer (client).

c) The domain name of Alice's outgoing email server is configured in her application. Alice's outgoing email server gets the domain name for Bob's incoming mail server from DNS as an "MX" record. Finally, Bob's email application is configured with the domain name for his incoming server.

8. Domain Name System (DNS) (9 p)

Consider the excerpt below from a DNS zone file.

a) What is the the domain name for the zone? Answer with a Fully Qualified Domain Name (FQDN). (1 p)
b) What is the name server for the zone? Answer with FQDN and IP address(es). If it has more than one IP address, specify all of them. (2 p)
c) You submit a question to the name server, asking for the address(es) of "punchy". What is the response? If the response includes more than one IP address, specify all of them. (1 p)
d) Suppose that you send a DNS request for the "A" record for "cheever.loophole.circus.mx." From which name server would you get the response? Answer with name (as an FQDN) and IP address(es). If it has more than one IP address, specify all of them. (2 p)
e) While you are somewhere on the Internet, you desperately need to know the IP address for the domain name "atom.circus.mx." The local name server at your current location is temporarily out of order, but you know that root server "i.root-servers.net." has IP address "192.36.148.17." You use "dig" to perform the following DNS request:

```
dig atom.circus.mx. @192.36.148.17
```

What will be the response? (1 p)
f) The administrator for the zone wants to set up an alternative name server as backup, in case the main name server should become unavailable. In DNS terminology, there should be a primary and a secondary name server. The secondary name server should be called "secondary.opera.mx." In order to bring the secondary name server into operation, the administrator performs three main steps. What are those steps? (2 p)

```
$ORIGIN circus.mx.
$TTL 86400
@               IN   SOA    ns.circus.mx. hostmaster.circus.mx.(
                             2016010701 10800 3600 604800 3600 )

                IN   NS     ns.circus.mx.
loophole        IN   NS     ns2.circus.mx.
www             IN   CNAME  server
ns              IN   A      192.99.1.5
                IN   AAAA   2001:6b0:2::32:bb
pirelli         IN   A      192.99.1.38
pirelli         IN   A      192.99.33.123
punchy          IN   CNAME  pirelli
server          IN   A      192.99.1.43
server          IN   A      192.99.33.123
goliath         IN   A      192.99.1.58
atom            IN   A      192.99.1.59
```

```
ns2.circus.mx.   IN   A      130.237.72.246
                 IN   AAAA   2001:6b0:1::246
```

Solution

a) "circus.mx." Remember the dot "." at the end!

b) "ns.circus.mx.", with addresses 192.99.1.5 and 2001:6b0:2::32:bb.

c) "punchy" is an alias (CNAME) for "pirelli", with IP addresses 192.99.1.38 and 192.99.33.123.

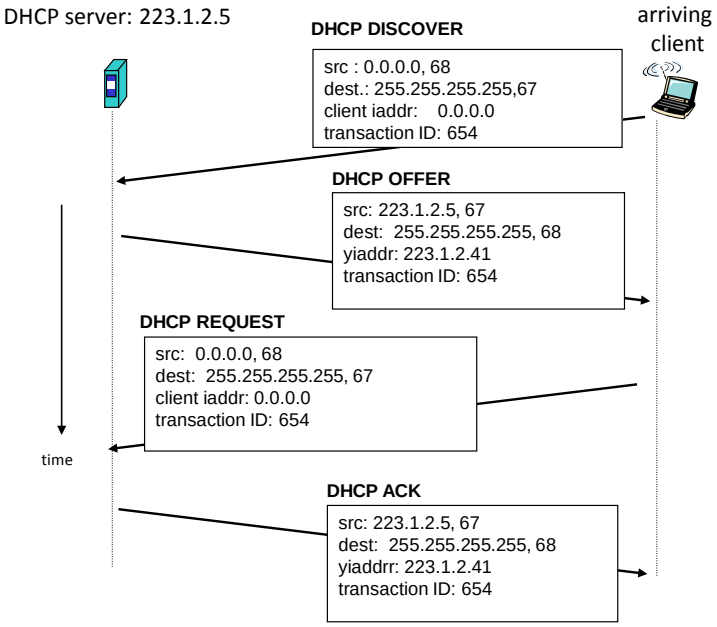
d) The response comes from "ns2.circus.mx.", which has IP addresses 130.237.72.246 and 2001:6b0:1::246.

e) The root server responds with the name servers for the "mx." top-level domain. The response contains the domain names for the name servers in the AUTHORITY section, and their IP addresses as "glue records" in the ADDITIONAL section.

f) 1) Add NS and address records for "secondary.opera.mx." in the zone file, just like "ns.circus.mx."; 2) inform the delegating zone ("mx.") about the new name server; 3) and configure the primary and secondary name servers to do zone transfers, so that any updates to the zone file in the primary server are automatically propagated to the secondary.

9. Autoconfiguration (5 p)

Consider the following DHCP message exchange, where DHCP client arrives and requests an IP address from the DCHP server.



Answer the following questions:

a) Why can't the client use the IP address offered in the DHCP OFFER as source address when it sends the DHCP REQUEST back to the server? (2 p)
b) When (at what point in the message sequence above) can the client start using the IP address offered by the server? (1 p)
c) Why is the DHCP REQUEST sent to the limited broadcast address instead of directly to the DHCP server? (2 p)

Solution

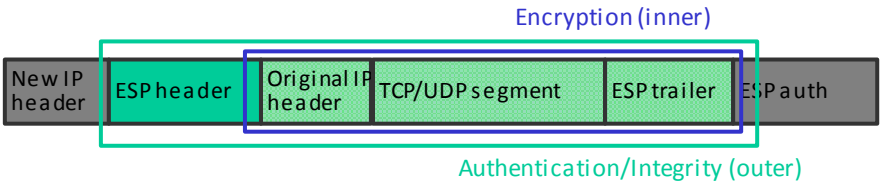
a) The IP address is not ready for use at that stage. The client may have received several offers from different DHCP servers and has selected one of them. Before it can be used, the client needs to inform the DHCP servers about its selection, which is done with the DCHP REQUEST.
b) After it has received the DHCP ACK from the server (the fourth message above).
c) By sending the DHCP REQUEST to limited broadcast it will reach all the DHCP servers so that other servers than the selected one get the information that their offers were not selected.

10. IPsec (5 p)

a) Draw an IP packet where IPsec ESP (Encapsulated Security Payload) is used in *tunnel* mode for both encryption and authentication. You don't need to show any header fields, just headers/trailers and payload. Mark the parts of the IP packet that are encrypted and the parts that are authenticated. (2p)
b) The ESP encapsulated IP packet arrives to the destination. Briefly describe how the destination determines what cryptographic algorithm to use to decrypt the packet? (3p).

Solution

a) IPsec ESP in tunnel mode:



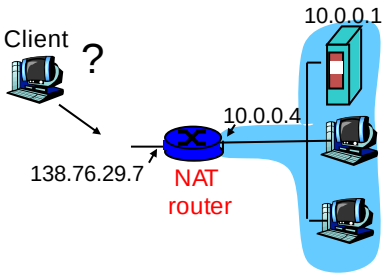
b) The receiver needs to lookup the correct SA (Security Association) in the security association database, and the SA will contain information about the cryptographic algorithm. The SA lookup is based on the {SPI, destination IP address, flags} retrieved from the ESP encapsulated IP packet.

11. NAT—Network Address Translation (5 p)

a) Explain and illustrate the NAT traversal problem. (2 p)
b) Explain and illustrate how Skype solves the NAT traversal problem. (3 p)

Solution

a) See the figure below. A server on the inside of the NAT has a private address and there is no address translation in the NAT. A client on the outside cannot connect to the server's private address since the private address cannot be used on the public Internet.



b) Skype uses a skype relay node which is located on the public Internet. Assume that an external skype client wants to connect to a skype client behind a NAT. The skype client behind the NAT will initiate a connection to the relay node. The external skype client will connect to the same relay node, which will be the bridge between the two connections. See figure below.