Examination
IK2218 Protocols and Principles of the Internet
EP2120 Internetworking

Date: 31 October 2013 at 08:00–13:00

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 21 November 2013.*
g) *After grading, EP2120 exams will be available for inspection at STEX (Q-building) and IK2218 exams will be available for inspection online.*
h) *Deadline for written requests for grading review is 5 December 2013.*
i) *Course responsible IK2218 is Peter Sjödin, phone 08-790 4255.*
j) *Course responsible EP2120 is György Dán, phone 08-790 4253.*

Important note!

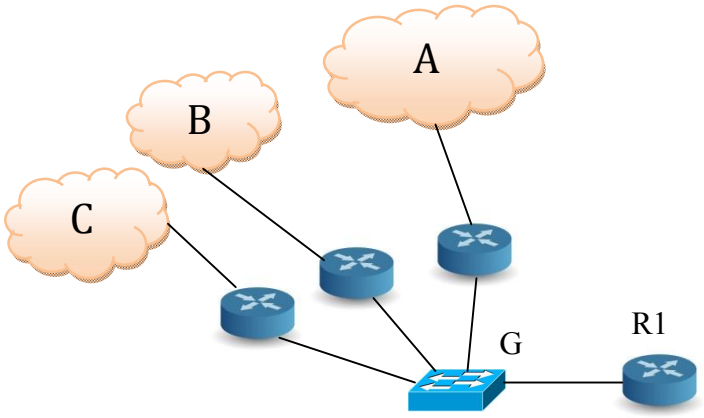
Your grade is F in any of these two cases:
- if you do not reach at least 10 (ten) points out of 20 for problems 1-4 or
- if you reach less than 30 points in total.

We advise you to start with problems 1-4.

Part I (Problems 1-4)

1. *IP and addressing (5p)*

Consider the IPv4 network shown in the figure below. Networks A, B, C, and G are switched Ethernet, router R1 connects the network to the Internet through an Internet Service Provider. Network A should be able to host 2000 hosts, while networks B and C should host 1000 hosts each.



a) What is the longest prefix length that you should consider for subnet A? What is the corresponding netmask? (1p)
b) Assume that you use the address block 172.31.0.0/20 for address allocation and the lowest possible address blocks for networks *A*, *B*, and *C*. Give the network address and the directed broadcast address of subnet *A* in CIDR notation! (1p)
c) What is the smallest block of addresses to accommodate all the hosts in subnet G? (1p)
d) Are the addresses allocated to networks *A, B, C*, and *G* routable on the public Internet? Do you need any particular functionality in R1 to be able to communicate from a host in network *A* with the rest of the Internet? If yes, what? Motivate your answer. (1p)
e) What is the difference between *link-local* and *unique-local* addresses in IPv6? How do they relate to private addresses used in IPv4? (1p)

SOLUTIONS:
a) The prefix length should be /21 or shorter, the corresponding netmask is 255.255.248.0.
b) The network address is 172.31.0.0/21, the directed broadcast 172.31.7.255.
c) You need 4 IP addresses + network+broadcast, the smallest block is /29 .
d) The addresses are not publicly routable, since they are private IP addresses. Thus, router R1 must have NAT functionality in order to enable communication between the hosts and the rest of the Internet .
e) IPv6 link-local addresses are valid only for addressing on a single link, while a unique local address can be used within a site very much like private addresses in IPv4. Every IPv6 enabled host has a link-local address, but they do not need to have a unique local address.

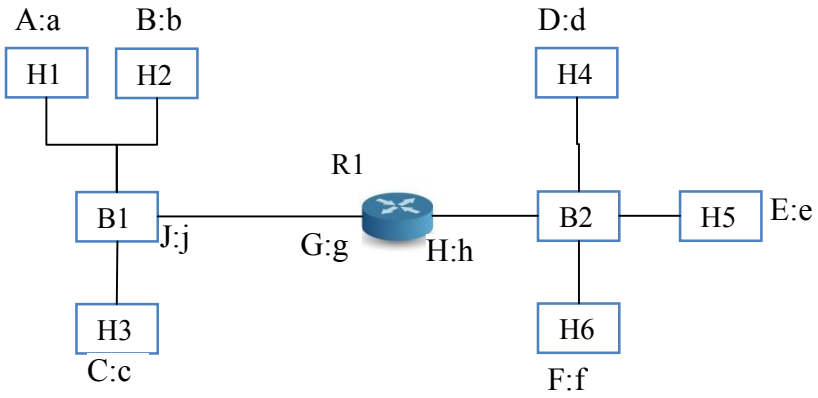
2. *Delivery and address resolution (5p)*

a) What is the purpose of the Address Resolution Protocol (ARP) in *IPv4* in the case of direct delivery? What is its purpose in the case of indirect delivery? (1p)

b) Consider a subnet in which there is a host with IP address *A* and MAC address *a*. The subnet uses Ethernet (802.3) at the link layer, which provides unreliable frame delivery. The only router connected to the subnet stored *A:a* in its ARP cache at time *t*. The router's ARP cache timeout is 300 seconds. At time *t*+30s the host is disconnected from the subnet, the router's forwarding table is unchanged. What happens if the router receives a datagram at time *t*+120 s with destination IP address A? (1p)

c) If a router fails to forward a datagram to the next hop, it usually sends an ICMP destination host unreachable or an ICMP destination network unreachable message to the sender. Name two conditions under which the router does not send an ICMP error message! Briefly explain the reason why an ICMP message is not sent under these two conditions! (1p)

Consider the following *IPv4* network consisting of 2 bridges and 1 router. Hosts H1 to H6 have one interface each. B1 and B2 are learning bridges. R1 is a router with an appropriate routing table. All ARP caches and the bridges' learning tables are empty. Assume that ARP snooping is used.



d) Identify the subnets in the figure! Observe that bridge B1 has a MAC address and an IP address. What purpose do these addresses serve for? (1p)

e) A process on Host H5 sends 100 bytes via UDP to a process on host H1. Show the contents of the learning tables and the ARP caches after the packet has been delivered. Assume that the process on Host H5 knows the IP address of Host H1. (1p)

SOLUTIONS:

a) In the case of direct deliver ARP is used to find the MAC address of the destination host, wheras in the case of indirect delivery its purpose is to find the MAC address of the next hop, i.e., a router.

b) The router will forward the datagram to host A. It will do so until the entry in its ARP cache expires. In particular, it will not send a destination host unreachable ICMP error message to the source IP address.

c) The router will not send an ICMP error message to the sender if:

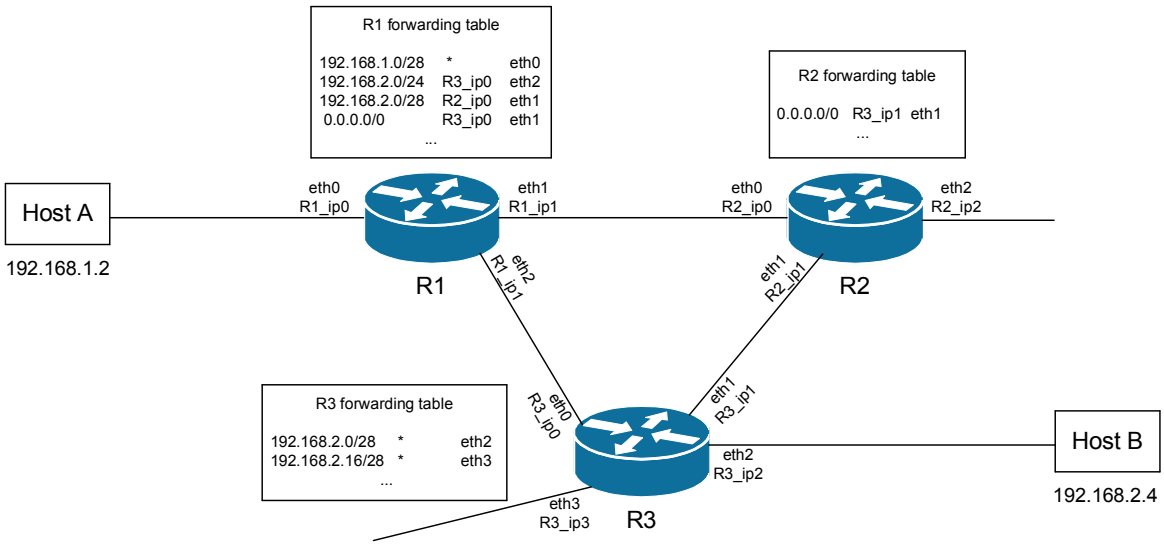
(i) the datagram contains an ICMP error message

(ii) the destination address is a multicast or a broadcast address

(iii) datagram sent as link layer broadcast,
(iv) the datagram's source address is 0.0.0.0.
The reason is
(i) to avoid loops
(ii-iii) to avoid broadcast storms, and
(iv) the lack of sender information.
d) Two subnets (left and right of router), bridges do not need IP and MAC address, but if a bridge has an IP/MAC address then it can be remotely managed.
e)
H1: G-g
H2: G-g
H3: G-g
H4: E-e
H5: H-h
H6: E-e
R1: E-e, A-a
B1: g-East, a-North
B2: e-East, h-West

3. *IP forwarding (5p)*

a) Consider a part of an IPv4 network as shown in the figure below. The MTU for every link is 1500 bytes. Host A sends an IP datagram of size 950 bytes with TTL = 64 to Host B. Host B successfully receives the packet.
What will be the TTL value of the IP header when Host B receives the datagram? Will any other field of the IP header have a different value when Host B receives it (compared to as sent by Host A)? If yes, which field(s)? (1p)



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

Destination	Next hop	Flag	Interface
10.16.0.0/16	-	U	m0
172.18.64.0/18	-	U	m1

192.168.138.0/24	-	U	m2
172.18.65.124/32	10.16.0.1	UGH	m0
192.168.19.0/24	172.18.65.173	UG	m1
10.17.0.0/16	10.16.0.245	UG	m0
0.0.0.0/0	192.168.138.3	UG	m2
192.168.19.5/32	192.168.138.3	UGH	m2

b) 10.17.213.72 (1p)
c) 192.168.138.8 (1p)
d) 192.168.19.4 (1p)
e) 130.235.15.67 (1p)

SOLUTIONS

a) TTL = 61. The IP Checksum will be different due to the change in the TTL value.
b) 10.16.0.245 on m0
c) Direct delivery on m2
d) 172.18.65.173 on m1
e) 192.168.138.3 on m2

4. *TCP (5p)*

Consider two hosts, A and B, connected by a network running IPv4. The capacity of all links is 100Mbps and the round trip time is 400ms. A process P_A on host A would like to transmit 30000 bytes to a process P_B on host B using TCP. The path MTU is known to be 1540 bytes. The receiving host has a receiver window size limit of 6000 bytes, which it advertises during connection establishment. The sender uses a value of 65535 for *sshthresh* for congestion control. Delayed acknowledgements are not used. The receiver can process the data as fast as they arrive.

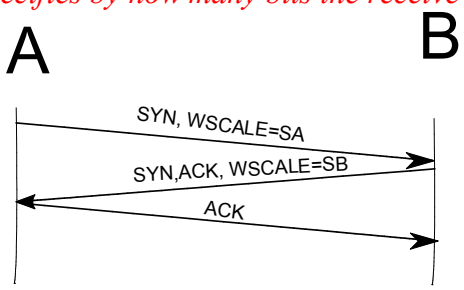
a. What is bandwidth delay product of the channel? How big should the receiver window size be in order to be able to fully utilize the channel (not considering congestion control)? (1 p)
b. Can TCP in general support a receiver window size sufficiently big to fully utilize the channel? If yes, use a drawing to illustrate the steps that TCP would follow to be able to fully utilize the channel: what information would be provided, when and where. Explain how the information would be used. (1p)
c. Consider that at time *t* the connection is started; the active open is performed by A. How much time does it take to transmit the data from A to B 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. The initial congestion window size is 3xMSS. (2 p)
d. RFC5681 describes two different algorithms for updating the congestion window during the congestion avoidance phase. According to one algorithm the CWND can be increased by MSS every time the number of acknowledged bytes reaches CWND. According to the other algorithm upon every ACK the sender uses the update rule: CWND += MSS*MSS/CWND.

   Consider the TCP throughput approximation $B(p) \approx \frac{1}{RTT}\sqrt{\frac{3}{2bp}}$, where p is the packet loss probability and b is the delayed acknowledgement parameter, i.e., the receiver sends an ACK after every b segments. Consider the above two algorithms and explain why and how the delayed acknowledgement parameter b can affect the average TCP throughput. Does the dependence on b apply to both algorithms? (1p)

SOLUTION:

a. The bandwidth delay product is $100*10^6*0.4s=40Mbit=5MB$. This is the receiver window size that one would need.

b. The Window scale option can be used to increase the receive window size above 65535 bytes. The use of the window scale option is agreed upon during connection establishment, and each party sends its window scale factor to the other party. The scale factor is between 0 and 14, and specifies by how many bits the receive window has to be shifted.



*c. The exchange of the first two segments related to connection establishment takes 1 RTT, data can be sent with the subsequent segment at t+RTT. The sender then sends 3 segments, worth 4500 bytes and waits for an ACK. It receives 1 ACK per sent segment at t+2*RTT. It then increases CWND by 1 MSS for each ACK, thus CWND=6xMSS, which exceeds the RWND (=6000bytes, 4MSS). The sender from now on sends 6000 bytes (4 segments) at a time, until it finishes sending all data. The data is in total 30000 bytes, which is 20xMSS, so the sender will send 16 segments in 4 rounds, and then 1 segment. Thus, the total time needed for the transmission is 1 RTT (conn.est)+6 RTT (data trans.)=7RTT=2.8s.*

*d. If one uses the first algorithm then the throughput would not depend on the delayed ACK parameter. If one uses the first algorithm then under delayed ACKS with parameter b then the receiver receives CWND/MSS/b number of ACKs for every CWND number of bytes sent. Thus, during one round trip time the CWND is increased by approximately CWND/(MSS*b)*MSS*MSS/CWND=MSS/b bytes.*

Part II (Problems 5-12)

5. *Fragmentation and UDP (5p)*

a) Hosts A and B are connected by an IPv4 internetwork. An application on Host A transmits 2597 bytes of data via UDP to Host B. The UDP header is 8 bytes long, there are no IP options used. The path between A and B goes through two networks connected by a router: the MTU of the first network is 1500 bytes and the MTU of the second network is 1000 bytes.
How many IP fragments arrive at Host B? Give the segment sizes, the fragmentation offset and the more fragments (MF) bit of all fragments. (3p)

b) Assume that the first fragment gets delayed in the intermediate router. It arrives 0.5s before the reassembly timer expires. How does this affect UDP and the application process in host B? (1p)

c) The router connecting the two networks has a software bug, and due to the software bug the router's memory that stores the source IP address of the datagrams to be forwarded gets

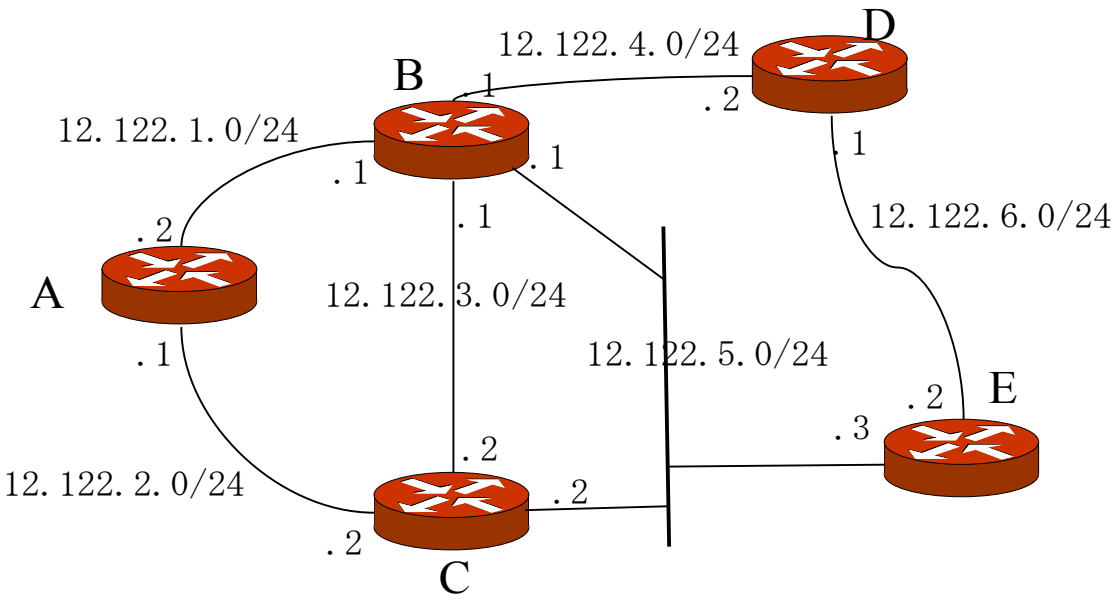
overwritten by the value 0xEFA5B4C. This happens after the router checks the IP checksum and before it computes the new checksum.
Does this bug affect the number of fragments received by Host B? Does it affect the time it takes for the process on Host B to receive the data from UDP? (1p)

SOLUTIONS:
a) The data to be transmitted includes the UDP header, in total 2605 bytes.
Fragment 1: 976 bytes IP payload, offset=0, MF=1
Fragment 2: 504 bytes IP payload, offset=122, MF=1
Fragment 3: 976 bytes IP payload, offset=185, MF=1
Fragment 4: 149 bytes IP payload, offset=307, MF=0

b) Reassembly is done at the IP layer. Therefore, UDP and the application will only experience some delay.

c) It does not affect the number of fragments received. If the UDP checksum is used then UDP will drop the segment (thus it will never be delivered). If the UDP checksum is not used then the error does not influence the time.

6. *Routing (5p)*



In the IPv4 network shown in the figure all routers A-E run RIPv2 and all link metrics are 1. The addresses of the IPv4 networks and the associated interface addresses are given in the figure. Note that the letters A-E do *not* denote addresses. Assume an initial state for all routers, where only the addresses of the directly connected networks are present in the routing tables. The destinations in the network are the /24 prefixes. Assume also that all RIP implementations support Equal-cost-multi-path (ECMP). All routers implement split-horizon and poison reverse.
Express routes as 'destination, metric, next-hop'. If the destination is a directly connected network, the route is given as 'destination, metric, -'.
a. What is the initial routing state of B? (1p)

b. Assume that router B starts by sending a RIP response to its neighbours. What is the routing state of C after it has received the initial distance-vector from B? (1p)

c. Assume that the second event that happens in the network is that router A sends RIP responses to its neighbours. Which RIP response messages does A send, and which distance-vectors do they contain? You should indicate the source and the destination address of each RIP message, on which interface it is sent out (and to where) and which distance-vectors (destination-metric tuples) are contained in each message. (1p)

d. What are the routing states of B and C after they have received the distance-vector from A in the previous step (c)? (1p)

e. Assume that the third event that happens in the network is the failure of the link between B and D. Briefly describe what happens in the network assuming that router E is the first one sending RIP responses to its neighbours after the link failure. What is the routing state of B after it has received the distance-vector from E? (1p)

SOLUTION

a)

12.122.1.0/24, 1, - #east interface
12.122.4.0/24, 1, - #north interface
12.122.5.0/24, 1, - #south-east interface
12.122.3.0/24, 1, - #south interface

b)

12.122.2.0/24, 1, - # west interface
12.122.3.0/24, 1, - #north interface
12.122.5.0/24, 1, - #east interface
12.122.1.0/24, 2, 12.122.3.1
12.122.1.0/24, 2, 12.122.5.1
12.122.4.0/24, 2, 12.122.3.1
12.122.4.0/24, 2, 12.122.5.1

c)

On the north interface, A sends a RIP response message with src address 12.122.1.2 and destination address 224.0.0.9. Alternatively, if the link between A and B is point-to-point, the destination address may be 12.122.1.1. The distance-vector of this message using split-horizon with poison reverse is:
12.122.2.0/24, 1
12.122.3.0/24, 16
12.122.4.0/24, 16
12.122.5.0/24, 16
(12.122.1.0/24, 16 # RIP implementations may announce this network but it is not necessary since all connected routers have this as a directly connected network: it is accepted both to have this route and to omit it)

On the south interface, A sends a RIP response message with src address 12.122.2.1 and destination address 224.0.0.9. Alternatively, if the link between A and C is point-to-point, the destination address may be 12.122.2.2. The distance-vector of this message using split-horizon with poison reverse is:
12.122.1.0/24, 1
12.122.3.0/24, 2
12.122.4.0/24, 2

12.122.5.0/24, 2
(12.122.2.0/24, 16 # Same comment as above)
d)
Routing state of B:
12.122.1.0/24, 1, - # west interface
12.122.3.0/24, 1, - #south interface
12.122.4.0/24, 1, - #east interface
12.122.5.0/24, 1, - #south-east interface
12.122.2.0/24, 2, 12.122.1.2

Routing state of C:
12.122.2.0/24, 1, - # west interface
12.122.3.0/24, 1, - #north interface
12.122.5.0/24, 1, - #east interface
12.122.1.0/24, 2, 12.122.3.1
12.122.1.0/24, 2, 12.122.5.1
12.122.4.0/24, 2, 12.122.3.1
12.122.4.0/24, 2, 12.122.5.1
12.122.1.0/24, 2, 12.122.2.1
e)

Router D sets the distance to networks 12.122.4.0/24, 12.122.1.0/24, 12.122.3.0/24, 12.122.5.0/24 to infinity (16). Router B sets the distance to network 12.122.4.0/24 to infinity (16). None of the other routers updates their state. Router E sends RIP responses to its neighbours, the routing state of B becomes:

Routing state of B:
12.122.1.0/24, 1, - # west interface
12.122.3.0/24, 1, - #south interface
12.122.4.0/24, 16, - #east interface
12.122.5.0/24, 1, - #south-east interface
12.122.2.0/24, 2, 12.122.1.2
12.122.6.0/24, 2, 12.122.5.3

7. *Electronic Mail (3p)*

In regular mail communication, there are several entities involved. The textbook describes three entities (or agents): User Agent, Mail Transfer Agent, and Mail Access Agent. Suppose that Alice and Bob are two users in different organizations. Alice sends a mail to Bob. Explain the steps in transferring the mail from Alice to Bob. For each step:

- Describe the agent(s) involved.
- Give an example of an email protocol that could be used for the communication.
- Explain which part is client and which part is server in the communication.

Solution

- Step 1: From Alice's mail client (UA, client) to her outgoing mail server (MTA, server). SMTP.
- Step 2: From Alice's outgoing mail server (MTA, client) to Bob's incoming mail server (MTA, server). SMTP.

- Step 3: From Bob's incoming mail server (MAA, server) to Bob's mail client (MAA or UA, client). POP or IMAP.

8. *Web (5p)*

You have a web site that you want to make more attractive in order to get more visits. You come up with the following reward scheme: when a user has visited your web site 25 times, a message "Congratulations Alice!" is displayed (assuming that the user has registered at your site with the name "Alice"). The user is then offered to participate in a lottery. The lottery is simple: the user is presented with a number of lottery tickets and draws one by clicking on it, and then a message is displayed whether the user won or not.

In order to build your web site this way, you need to use certain tools: You use cookies, dynamic documents and active documents. For each of these tools, explain in general terms for what purpose you would use the tool in order to implement the desired functionality.

Hint: you only need to describe the web site in general terms, at the level of detail that has been covered during lectures. You do not need to describe individual documents at any level of detail. Neither should you go into any details about the implementation of the lottery, what language you would use, etc. Of course, your solution should be entirely web-based, taking place in the browser.

Solution

First, you need to use cookies. A user needs to register in order to give her name, and a cookie is then created and sent back to the client. In the server, you should maintain a database that counts how many times each user has visited the site, where you use the cookies to identify users.

Furthermore, you would use dynamic documents (executed on the server) to decide whether to invite the user to the lottery and to generate the congratulations message with the user's name. Finally, the actual lottery would be an active document, executing in the browser.

9. *DNS (6p)*

Study the following excerpt from a DNS zone file.

```
$TTL        86400
@           IN    SOA    wagstaff.org. hostmaster.wagstaff.org.(
                         2013103101
                         10800
                         3600
                         604800
                         3600 )

                         IN  NS         ns.wagstaff.org.
                         IN  NS         ns2.kornblow.org.
                         IN  MX   20    mx.wagstaff.org.
ns                       IN  A          192.0.1.5
                         IN  A          192.0.33.4
ns2.kornblow.org.        IN  A          198.16.12.2
server                   IN  A          192.0.1.38
```

```
                        IN  AAAA      2001:6b01::bffd
mx                      IN  CNAME     server
```

a) Which is the domain? Answer with the Fully Qualified Domain Name (FQDN). (1 p)

b) Suppose that a client sends a request to this name server, asking for the address(es) of the host "ns". What is the answer? (1 p)

c) The response (like all DNS responses) contains three sections: "ANSWER", "AUTHORITY", and "ADDITIONAL". Explains what these three sections would contain in this case. (2 p)

d) Suppose that you are connected to a network in another organization, which has a local DNS server. You are curious to know whether there have been any recent accesses (within a few seconds) from the organization to the site in question. Explain how you could find that out by sending DNS requests, for instance using a tool such as "dig". Describe the DNS requests you would send, and explain what you would learn from the responses. (2 p)

Solution

a) "wagstaff.org." *period at the end (.) is mandatory – otherwise it is not an FQDN*

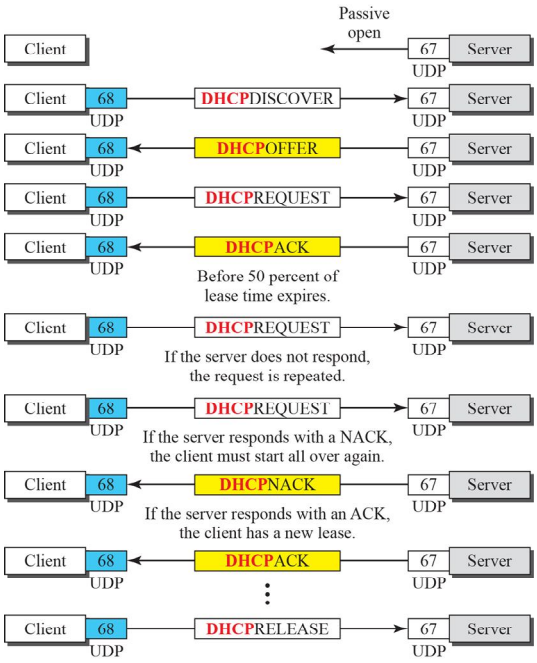
b) Two "A" records: 192.0.1.5 and 192.0.33.4

c) The ANSWER section contains the two "A" records. The AUTHORITY section contains the names of the name servers for the domain: "`ns.wagstaff.org.`" and "`ns2.kornblow.org.`". The ADDITIONAL section contains the IP addresses of the nameservers (three addresses in total, in this case).

d) You could start by sending a request to your local server for the nameserver of wagstaff.org, ie "dig ns wagstaff.org". In the response from the local nameserver, which will come from its cache, you will note the TTL and the name of the nameserver. Next you send the same question but explicitly directed to wagstaff.org's name server: "dig @ ns.wagstaff.org ns wagstaff.org". You will note the TTL from the response (which comes directly from the zone file). By comparing the two TTL values, you can determine how long the entry has been cached in the local server.

10. *Autoconfiguration (5p)*

Consider the figure below, illustrating a DHCP packet exchange between a client and a DHCP server. Assume that the IP address of the DHCP server is 136.15.15.23.



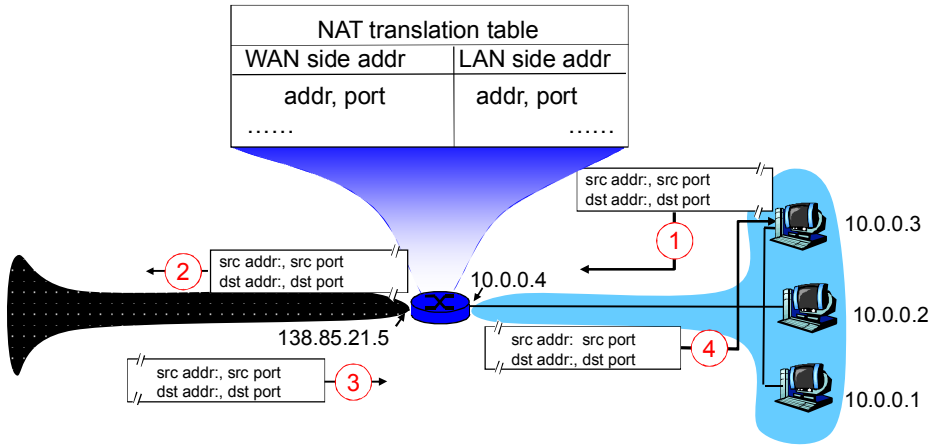
a) Which is the first message in which the client will find the IP address provided by the server and at what point in the message exchange is the client allowed start using the IP address it gets from the server? (2p)
b) In IPv6 stateless autoconfiguration, the client can create its own global unicast address instead of requesting it from a DHCP server. Briefly explain how that procedure works. From where will the client get the information it needs in order to create the global unicast address? What protocol is used in this procedure? If the client does not have a global unicast address, can it still communicate? If so, what type of address will the client then use? (3p)

SOLUTION

a) The client will see the IP address already in the DHCPOFFER, but it cannot use the address until after the DHCPACK in message four from the server.
b) The client will first generate an IPv6 link local address (LLA) based on the link local prefix and its MAC address (or a random number). The LLA can be used of communication on the local subnet. The client will then use ICMPv6 to communicate with routers on the subnet to figure out global prefix and subnet prefix. It can then combine its LLA with these prefixes to create a global unicast address.

11. *IP Gateways—NATs and Firewalls (5p)*

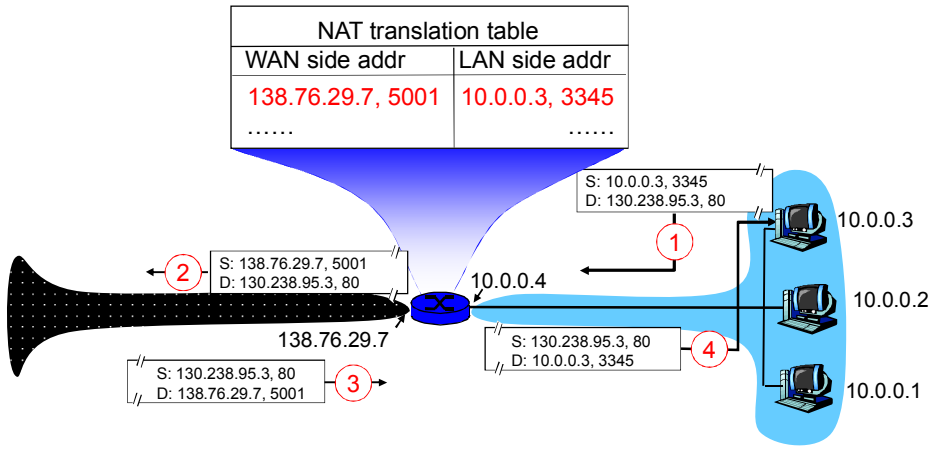
a) Consider the figure below. Assume that host 10.0.0.3 on a private network (10.0.0.0/24) sends an HTTP request through its NAT box to a web server (port 80) on address 130.238.95.3 and that this web server answers with an HTTP response back to the host. Fill in source address, source port, destination address, and destination port in the IP packets 1-4 in the figure. Also, fill in the NAT table as it will look when the four packets have been exchanged. (2p)



b) Explain what an application level gateway is and how it is different from a packet filter firewall. You answer should cover how the two are different in the way they operate and in what type of security features they can support. (3p)

SOLUTION

a)



b) An application level gateway splices and relays two application-specific connections. This should be compared to per packet operation performed by packet filters. The application level gateway can support high-level user-to-gateway authentication and it uses simpler filtering rules than for arbitrary TCP/IP traffic.

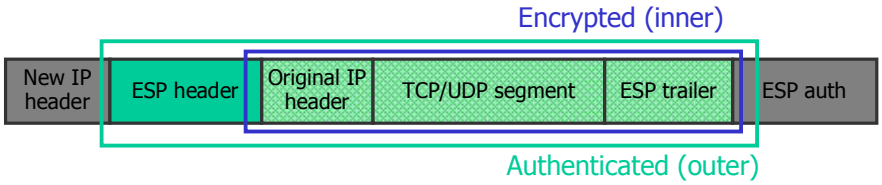
12. *IPsec and IKE (6p)*

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) An ESP encapsulated IP packet arrives to the destination. Briefly describe how the destination determines what cryptographic algorithm to use to decrypt the packet? (2p)
c) IKE is divided into two different phases. Briefly explain the two phases and why there are two phases. (2p)

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.
c) In phase 1, mutual authentication is done and IKE session (IKE SA) keys are established. In phase 2, one or more child SAs (IPsec SAs) are set up. The reason for having two phases is that mutual authentication is expensive. With the two phases, IPsec SAs can be renewed without repeating the mutual authentication.