ETH zürich

Institut für Technische Informatik und Kommunikationsnetze

Department ITET
August 2020

Prof. Dr. Laurent Vanbever
R. Birkner, C. Busse-Grawitz, T. Bühler, T. Holterbach, R. Meier

Exam: Communication Networks

26 August 2020, 09:30–12:00, Room HIL F 15

General remarks:

- ▷ Write your **name** and your **ETH student number** below on this front page and **sign it**.
- ▷ Put your **legitimation card** on the top right corner of your desk. Make sure that the side containing your name and **student number** is visible.
- ▷ Check if you have received **all task sheets** (Pages **1 - 34**).
- ▷ Do **not separate** the **task sheets** as we collect the exams **only after you left** the room.

- ▷ Write your answers directly on the task sheets.
- ▷ **All answers fit within the allocated space and often in much less.**
- ▷ If you need more space, use the three extra sheets at the **end of the exam**. Indicate the **task** in the corresponding field.

- ▷ **Read each task completely before you start solving it**.
- ▷ **For the best mark, it is not required to score all points.**

- ▷ Please answer either in **English or German**.
- ▷ **Write clearly** in blue or black ink (not red) using a **pen**, not a pencil.
- ▷ **Cancel** invalid parts of your solutions **clearly**.
- ▷ At the end of the exam, **place the exam face up on the top left corner** of your desk. Then collect all your belongings and **exit the room** according to the given instructions.
- ▷ In case of close contact with the TAs (e.g., questions), please put your **face mask** on.

Special aids:

- ▷ All written materials (vocabulary books, lecture and lab scripts, exercises, etc.) are allowed.
- ▷ Using a calculator is allowed, but the use of electronic communication tools (mobile phone, computer, etc.) is strictly forbidden.

Family name: Student legi nr.:

First name: Signature:

Do not write in the table below (used by correctors only):

Task	Points	Sig.
Ethernet & IP	/28	
Intra-domain routing	/24	
Inter-domain routing	/43	
Reliable transport	/25	
Applications	/30	
Total	/150	

leave blank

Task 1: Ethernet & IP **28 Points**

leave blank

a) Warm-up **(6 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true	false	
☐	☐	While rare, collisions in a full-duplex switched Ethernet network can happen.
☐	☐	Using one common spanning tree for all the VLANs instead of one spanning tree per VLAN necessarily reduces the available network bandwidth.
☐	☐	Consider two hosts A and B connected to the same layer-2 switch. When A sends an IP packet addressed to B, the destination MAC address is necessarily the MAC address of B.
☐	☐	DHCP cannot run on top of TCP.
☐	☐	The TTL of an IP packet captured in a layer-2 network always corresponds to the initial TTL value used by the host that originated the packet.
☐	☐	The size of an IPv4 packet header is usually 20 bytes.

leave blank

b) Allocating IP addresses **(9 Points)**

Consider the 3-parts topology in Figure 1. At the top sits a private network with one layer-2 switch and three hosts. The layer-2 switch is connected to an Internet Service Provider (ISP) via an IP router that performs NAT. In the middle sits the ISP network which contains one layer-2 switch and three IP routers. Finally, at the bottom sits a small data-center network composed of three servers connected to the ISP network via one router.

Your goal is to allocate IP addresses to the various devices/interfaces in the topology to enable connectivity between all devices.

Note: We provide boxes in Figure 1 next to each device or interface in which you can write the IP addresses. **However**, not all of them need to be filled, i.e. some devices or interfaces do not need an IP address to enable end-to-end connectivity.

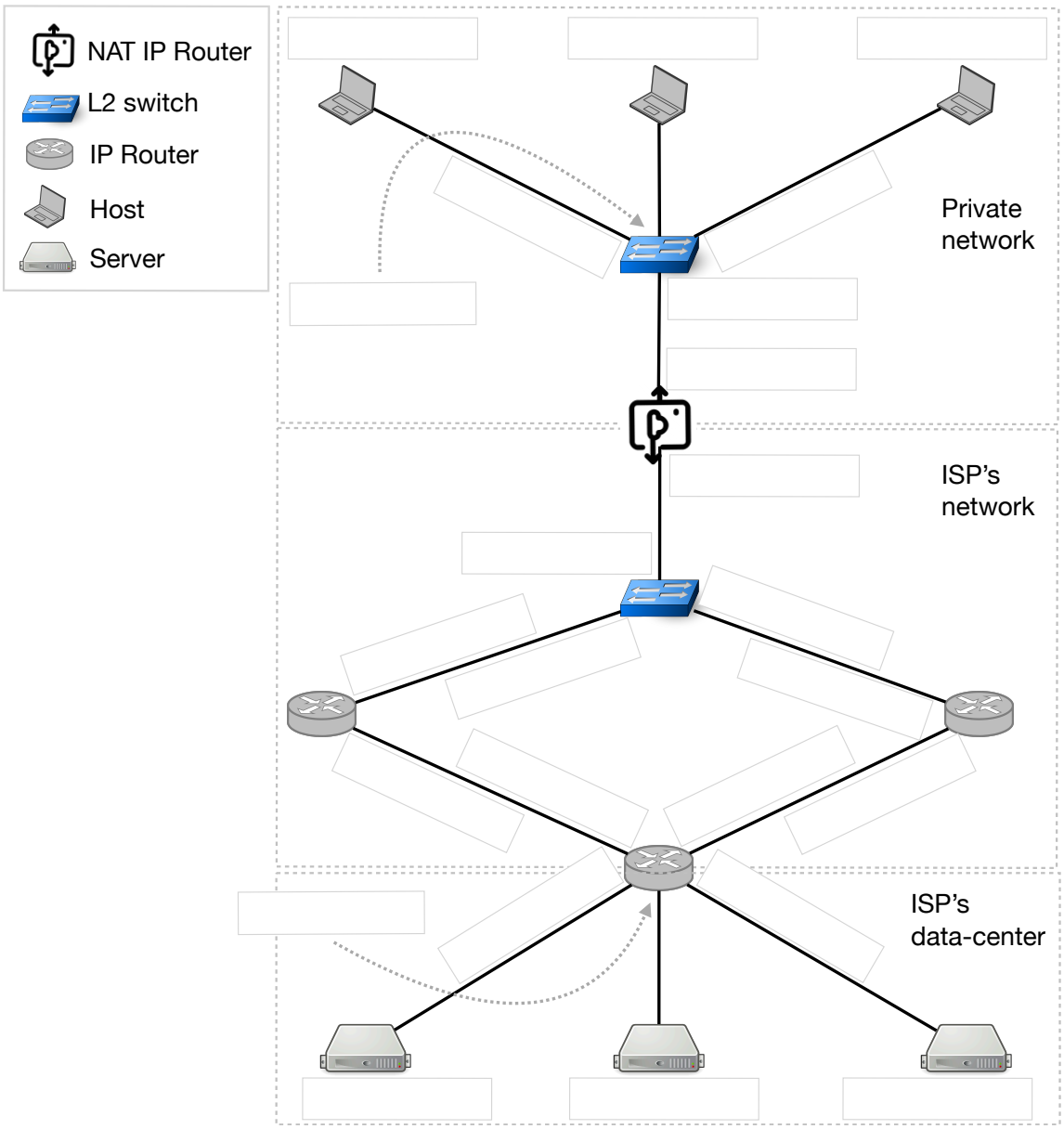

Figure 1: A network with L2 switches, IP routers, a NAT IP router, three hosts and three servers.

(i) Write down possible IPv4 addresses (including the subnets) that you could configure in the private network. Recall that not all the boxes necessarily require an IP address.

(2 Points)

(ii) Assuming that the ISP network uses IP addresses in the `1.0.0.0/8` subnet, write down possible IP addresses (including the subnets) that you could configure in the ISP network.

(2 Points)

(iii) Write down possible IP addresses that the data-center could use knowing that:

- The ISP can only allocate IP addresses in its data-center that are in the following list (not all the listed IP addresses need to be used).

1. `2.0.0.1/8`
2. `2.0.0.2/8`
3. `2.0.0.3/8`
4. `2.0.0.4/8`
5. `2001:0DB8:AC10:FE01::1/64`
6. `2001:0DB8:AC10:FE01::2/64`
7. `2001:0DB8:AC10:FE01::3/64`
8. `2001:0DB8:AC10:FE01::4/64`
9. `3.0.167.1/20`
10. `3.0.178.1/21`
11. `3.0.181.1/17`
12. `3.0.188.1/21`
13. `2001:45AB:E559:AA34::1/64`
14. `2001:45AB:E559:AA34::2/64`
15. `4.0.0.1/24`
16. `4.0.0.2/24`
17. `2001:19F4:AA55:A345::1/64`
18. `2001:19F4:AA55:A345::2/64`

- The servers must be reachable from outside the data-center using IPv4 but, within the data-center, the servers should be able to reach each other using IPv6. You may therefore need to configure two IP addresses (one IPv4 and one IPv6 address) per device or interface.

If there is not enough space in a box to write an IP address, you can write its number in the list above, e.g. 3 for `2.0.0.3/8` and 5 for `2001:0DB8:AC10:FE01::1/64`.

(5 Points)

leave blank

c) Reverse engineering a L2 network **(13 Points)**

You started to work as a network engineer in a company and your first task is to operate the L2 network. Unfortunately, the previous network engineer forgot to share with you all the details about how forwarding is set up in the network. You only know:

- The topology of the network, which is depicted in Figure 2;
- The 7 hosts connected to the network, together with their MAC addresses;
- The content of the forwarding table for switches A, B, D and F;
- That a switch broadcasts a packet if no entry in the forwarding table matches that packet;
- That the Spanning Tree Protocol is running with a unary cost on each link;
- That the root of the spanning tree is (currently) **unknown**;
- That there is no VLAN;
- The connectivity is working fine: hosts can communicate with each other, without losses.

With this in mind, you decide to reverse engineer how the network works.

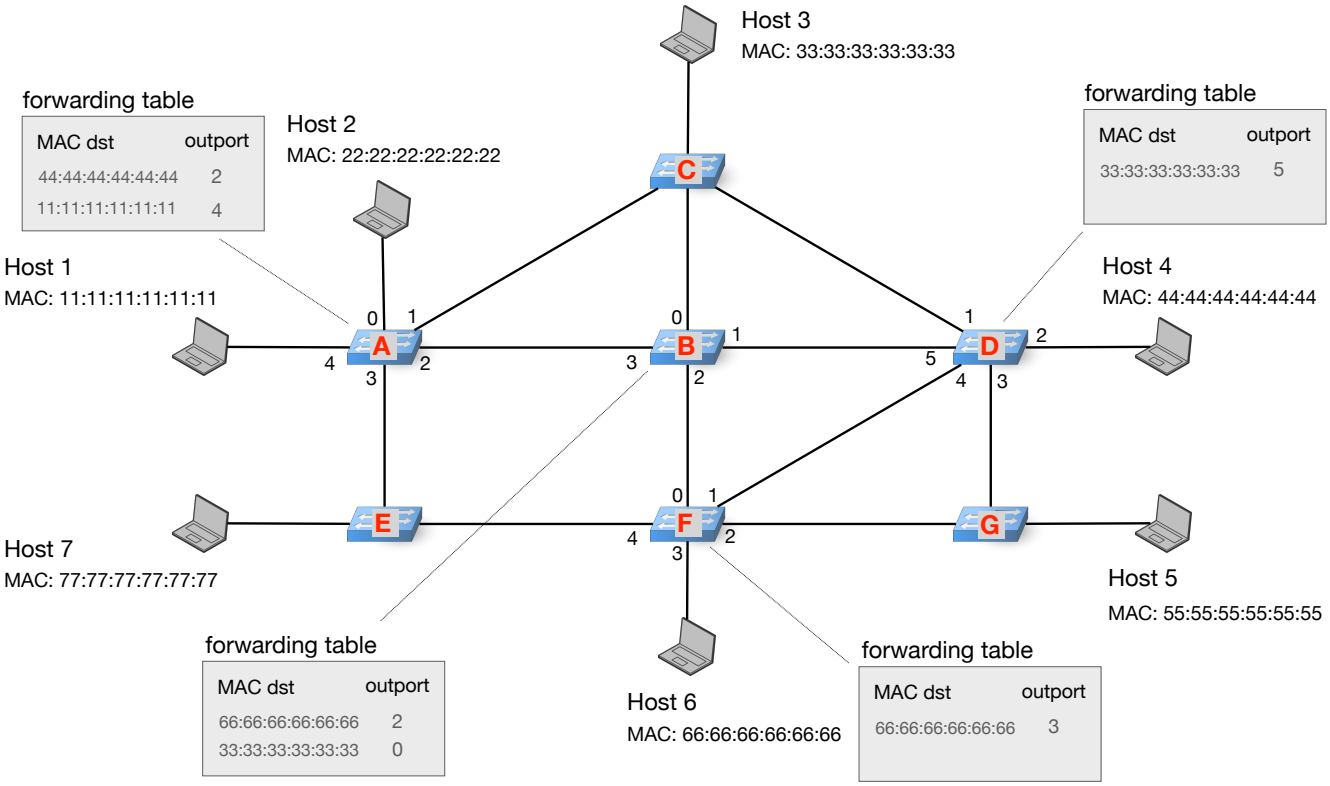

Figure 2: The L2 network with 7 switches and 7 hosts.

(i) Derive the links that are deactivated by the Spanning Tree Protocol, the links that are activated (i.e., used by the traffic), and the links for which we cannot determine whether they are activated or deactivated (we say they are undetermined). Name a link based on the two adjacent switches, e.g., the link between switch A and B is named AB.

Note: You can ignore the links connecting a switch and a host, as those ones are necessarily activated. (4 Points)

Activated links: _____

Deactivated links: _____

Undetermined links: _____

(ii) Derive the switch(es) that could be the root of the spanning tree. Explain your reasoning. (2 Points)

(iii) A colleague decides to help you and records all the packets traversing a link in the network. Figure 3 shows the recorded packets. Recall that there is no packet loss and the trace shows all the packets traversing the link.

Pkt ID	Source MAC	Dest. MAC	Protocol	Content
1	4e:d7:aa:9a:56:89	01:80:C2:00:00:00	BPDU	the # of hops to reach the root is 2
2	22:22:22:22:22:22	FF:FF:FF:FF:FF:FF	ARP	who-has 1.0.0.6 tell 1.0.0.2
3	66:66:66:66:66:66	11:11:11:11:11:11	ICMP	src_ip=1.0.0.6 dst_ip=1.0.0.1 Echo request
4	44:44:44:44:44:44	FF:FF:FF:FF:FF:FF	ARP	who-has 1.0.0.7 tell 1.0.0.4
5	77:77:77:77:77:77	44:44:44:44:44:44	ARP	1.0.0.7 is at 77:77:77:77:77:77
6	55:55:55:55:55:55	FF:FF:FF:FF:FF:FF	ARP	who-has 1.0.0.3 tell 1.0.0.5

Figure 3: Packets recorded on a link in the L2 network.

Unfortunately, your colleague has forgotten on which link he has recorded the packets. With the information you have derived in the previous questions as well as the information you can find by looking at the traffic trace, derive the link on which the trace has been recorded.

Hint: The link is connecting two switches.

Use the following table to explain your reasoning. For each packet, write the information that you can derive in the table. If a packet does not bring any additional information enabling you to narrow down the link, simply leave the corresponding line empty.

(7 Points)

Pkt ID	Information derived
1	
2	
3	
4	
5	
6	

Write down below the link on which the packets were recorded:

leave blank

Task 2: Intra-domain routing **24 Points**

leave blank

a) Warm-up **(5 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true ☐ false ☐ Decreasing by one the weight of a single link alongside the shortest path between two nodes will not change the all-pair shortest paths computed by Dijkstra's algorithm.

true ☐ false ☐ Dijkstra's algorithm iterates over all the nodes until all of them are added to the set of explored nodes. In the last iteration, when handling the last remaining node, the found shortest paths never change.

true ☐ false ☐ If `A-B-C` is *not* the shortest path for A to reach C, then `A-B-C-D-E` *cannot* be the shortest path for A to reach E.

true ☐ false ☐ In any network with 10 nodes and unary link cost, reducing the infinity value from 16 down to 8 would not change the forwarding state when using a distance-vector routing protocol.

true ☐ false ☐ The hierarchical forwarding table discussed in the lecture as a way to reduce router convergence time is a classical example of how using hierarchical structures help increase scalability.

leave blank

b) Analyzing ping output **(6 Points)**

Figure 4 shows a network with four IP routers and various link delays. In this network you run the ping application from a host connected to router A (which has IP `1.1.1.2`) towards a host connected to router D with IP `1.1.1.10`. Figure 5 shows the output of the ping command. You can assume that the only source of delay in the network comes from the link delays indicated in the figure. All routers and hosts react immediately and there is e.g., no queuing delay or other source of noise present in the network. The network runs OSPF to find the shortest paths and each link has a bidirectional weight assigned.

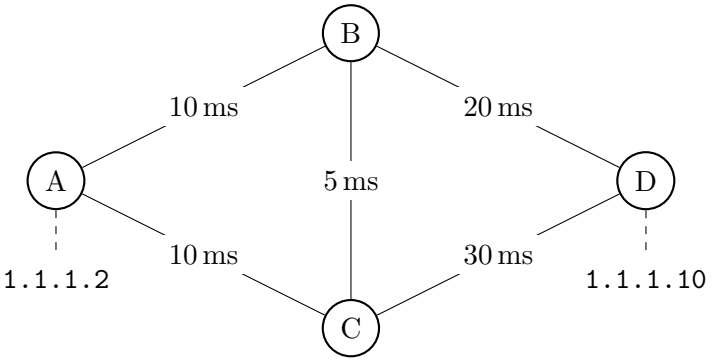

Figure 4: A network with four routers and different link delays in ms.

```
Host connected to router A # ping 1.1.1.10
64 bytes from 1.1.1.10: icmp_seq=0 time=90ms
64 bytes from 1.1.1.10: icmp_seq=1 time=90ms
64 bytes from 1.1.1.10: icmp_seq=2 time=90ms
64 bytes from 1.1.1.10: icmp_seq=3 time=90ms
64 bytes from 1.1.1.10: icmp_seq=4 time=75ms
64 bytes from 1.1.1.10: icmp_seq=5 time=60ms
64 bytes from 1.1.1.10: icmp_seq=6 time=60ms
64 bytes from 1.1.1.10: icmp_seq=7 time=60ms
                    ***
64 bytes from 1.1.1.10: icmp_seq=35 TTL expired in transit
64 bytes from 1.1.1.10: icmp_seq=36 TTL expired in transit
64 bytes from 1.1.1.10: icmp_seq=37 TTL expired in transit
                    ***
64 bytes from 1.1.1.10: icmp_seq=52 request timed out
64 bytes from 1.1.1.10: icmp_seq=53 request timed out
64 bytes from 1.1.1.10: icmp_seq=54 request timed out
                    ***
```

Figure 5: The output of a ping from a host connected to router A towards a host connected to router D.

(i) Assign weights to each link such that you achieve a forwarding state in which the ping command returns a time output of 90 ms as it is the case at the beginning of Figure 5, e.g. `icmp_seq=0`. (1 Point)

Weight link `A-B` _____

Weight link `A-C` _____

Weight link `B-C` _____

Weight link `B-D` _____

Weight link `C-D` _____

(ii) What happens at `icmp_seq=4` and `icmp_seq=5` in Figure 5? How is it possible that we observe a delay of 75 ms and then 60 ms? (2 Points)

(iii) Starting from `icmp_seq=35`, you observe a strange behavior: your pings no longer succeed. The TTL of the requests expire in transit and, starting from `icmp_seq=52`, the requests just time out. You suspect that someone has introduced static routes in the network, bypassing OSPF. Based on this assumption, describe which path your packets could take and give an example of a static route which could induce each problem. Note that the static route does not need to be the same for both problems. (3 Points)

TTL expired in transit:

Path used: _____

_____

with static route for prefix _____ from router _____ towards router _____

Request timed out:

Path used: _____

_____

with static route for prefix _____ from router _____ towards router _____

c) Reverse Dijkstra (13 Points)

leave blank

The network engineer at your company just retired and you have to take over. Unfortunately, it is unclear how the current network looks like. All you know is that it consists of 10 nodes (Figure 6). In addition, you know that there is at most one link between two nodes and that each link has a non-negative weight. However, you neither know which links exist nor the weights configured on these links.

(i) To figure out the links and the corresponding weights, you look at an output of Dijkstra's algorithm performed from node **U**. Table 1 shows the entire output of the algorithm. For each iteration, the table indicates the shortest path found so far towards each other node (starting from node **U**). The algorithm follows the one discussed in the lecture. If after one iteration there are multiple nodes with an equally-shortest path, the algorithm continues with the node which comes first in the alphabet.

On Figure 6, add all the links with their corresponding weight that you can identify based on the output from Dijkstra's algorithm (Table 1). (8 Points)

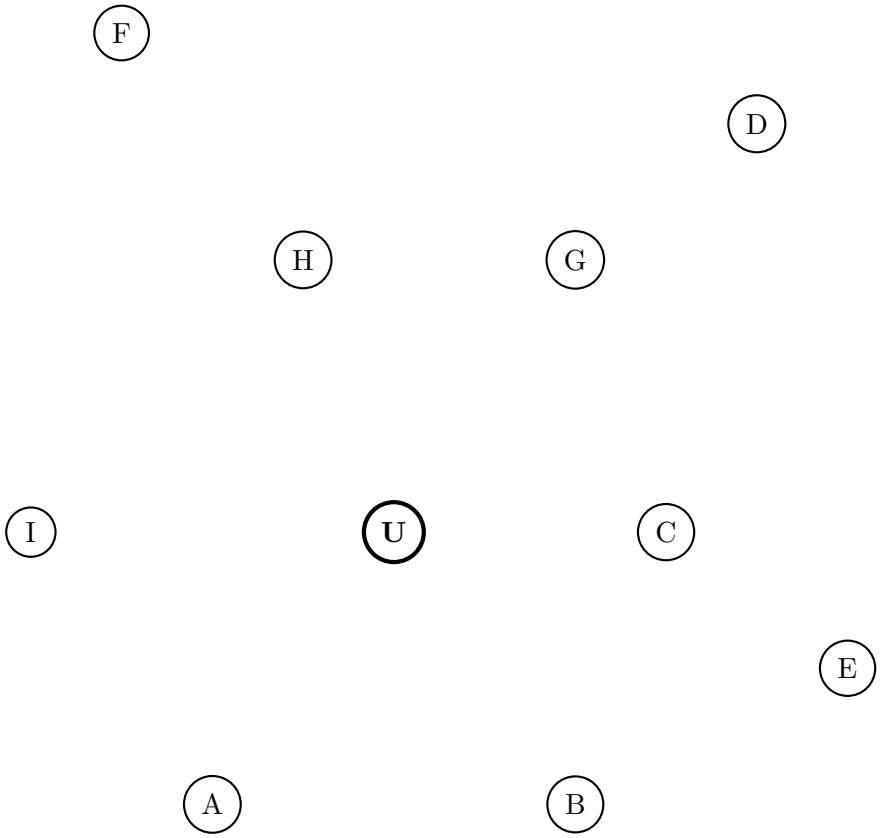

Figure 6: A network consisting of 10 nodes with unknown links and link weights.

#	**U**	A	B	C	D	E	F	G	H	I
1	0	2	3	1	-	-	-	10	-	11
2	0	2	2	1	8	-	-	10	-	11
3	0	2	2	1	8	-	-	10	-	11
4	0	2	2	1	8	100	-	10	-	11
5	0	2	2	1	8	9	15	10	-	11
6	0	2	2	1	8	9	15	10	-	11
7	0	2	2	1	8	9	13	10	14	11
8	0	2	2	1	8	9	12	10	14	11
9	0	2	2	1	8	9	12	10	13	11
10	0	2	2	1	8	9	12	10	13	11

Table 1: For each iteration (1 to 10) the table shows the shortest path found by Dijkstra's algorithm performed on node **U** towards all other nodes.

(ii) After analyzing the output from Dijkstra's algorithm, you are unsure if you really found all links in the network.

Could there be an additional link starting from node **U** which you could not identify based on the output from Dijkstra? If you think that is possible, give an example (link between node **U** and node ...) and indicate in which range the weight of this link could be. Otherwise, explain why this is not possible. (2 Points)

(iii) Could there be an additional link starting from node C which you could not identify based on the output from Dijkstra? If you think that is possible, give an example (link between node C and node ...) and indicate in which range the weight of this link could be. Otherwise, explain why this is not possible. (3 Points)

leave blank

Task 3: Inter-domain routing **43 Points**

leave blank

a) Warm-up **(6 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true	false	
☐	☐	To influence how traffic enters their network, network operators need to adapt their outbound BGP policies.
☐	☐	By tweaking the Multiple Exit Discriminator (MED) attribute, network operators can make their transit traffic leave via the closest egress point.
☐	☐	Consider a network which learns a single BGP route for `23.0.0.0/8` (an external prefix) and that this route has the highest possible local-preference. Then, any IP traffic matching `23.0.0.0/8` will necessarily be forwarded according to that route.
☐	☐	During a BGP oscillation, traffic will keep shifting from one path to another but will not necessarily be dropped.
☐	☐	RPKI relies on a database that stores the list of all the ASes that are allowed to propagate an advertisement for a given IP prefix.
☐	☐	Assuming RPKI is fully deployed (Internet-wide and for all existing prefixes), it is impossible for an attacker to perform a more-specific prefix hijack.

leave blank

b) Internet Routing Project **(10 Points)**

Put yourself in the shoes of a CommNet TA during the Internet routing project.

You are running the Internet topology depicted in Figure 7. It consists of two regions with 6 ASes each that are connected through an IXP, and two tier-1 ASes (AS 10 and AS 20). Single-headed plain arrows point from providers to their customers (e.g., AS 14 is the provider of AS 16), while double-headed dashed arrows connect peers (e.g., AS 23 and AS 24 are peers). The connections through the IXP can be regarded as direct peer connections (e.g., AS 12 and AS 25 are peers). On the AS path, the IXP is not visible. A correctly configured AS applies the default selection and exportation BGP policies based on their customers, peers and providers.

The internal topology of AS 24 is shown on the right. R_1 is connected to its two providers (AS 21 and AS 22), R_2 is connected to the peer (AS 23), and R_4 is connected to its two customers (AS 25 and AS 26).

Every AS announces a /8 prefix corresponding to its AS number (e.g., AS 13 advertises the prefix `13.0.0.0/8`). In between ASes, prefixes within `179.0.0.0/8` are used, which are not distributed any further.

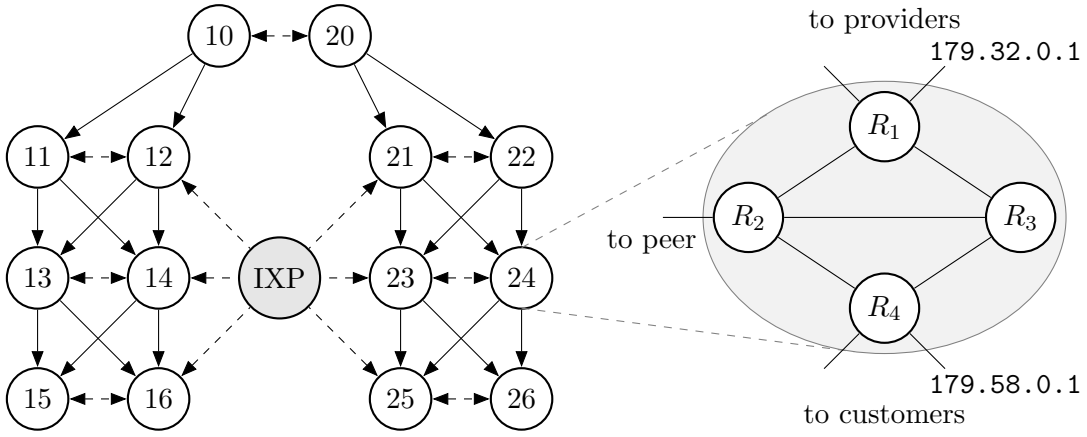

Figure 7: The topology of your mini-Internet routing project consisting of two regions connected through an IXP and two tier-1 ASes.

(i) The group of AS 12 contacts you because they are surprised to see that they are not using the direct path to AS 11, but go via AS 14 as shown in the looking glass below:

```
BGP table version is 296, local router ID is 12.151.0.1, vrf id 0
Default local pref 100, local AS 12
Status codes:  * valid, > best, i internal

    Network          Next Hop    LocPrf  Path
    ...              ...            ...  ...
*> i11.0.0.0/8       12.154.0.1     150  14 16 23 21 20 10 11 i
*                    179.1.91.1     100  11 i
    ...              ...            ...  ...
```

Did AS 12 make a mistake while implementing the standard BGP policies? Explain why or why not. (1 Point)

(ii) Along the path (`14 16 23 21 20 10 11`) some groups misconfigured their network. List all ASes that made a mistake and explain what they are doing wrong. (3 Points)

(iii) The group of AS 24 contacts you as they cannot ping any destination in their provider's network (AS 22) from router R_3 despite having an entry for `22.0.0.0/8` in the looking glass of R_3:

```
BGP table version is 296, local router ID is 24.153.0.1, vrf id 0
Default local pref 100, local AS 24
Status codes:  * valid, > best, i internal

    Network          Next Hop    LocPrf  Path
    ...              ...            ...  ...
*> i22.0.0.0/8       179.32.0.1      50  22 i
    ...              ...            ...  ...
*>  26.0.0.0/8       179.58.0.1     150  26 i
   END OF LOOKING GLASS
```

Explain what the problem is and how one can solve it.

(3 Points)

(iv) The group of AS 25 wants to know how they can make their provider AS 23 send the traffic to `25.0.0.0/8` via AS 24 instead of using the direct link. How can they achieve that? What is the disadvantage of your chosen approach?

(3 Points)

leave blank

c) Detecting the cheater **(7 Points)**

Consider the BGP network in Figure 8. Single-headed plain arrows point from providers to their customers (AS A is the provider of AS C), while double-headed dashed arrows connect peers (AS C and AS D are peers). Each AS is made up of a single BGP router and applies the default selection and exportation BGP policies based on their customers, peers and providers.

AS E is the only AS to originate the prefix `44.44.0.0/16`, which it advertises to its two providers. AS B announces a default-route (`0.0.0.0/0`) over BGP to AS D, since it is its only provider.

AS D decides to cheat to lower its transit costs and configures a static route for `0.0.0.0/0` pointing to AS C. AS D indeed knows that AS C has providers on its own and therefore can reach all the destinations, even if those destinations are not advertised over BGP.

In the following, whenever there are multiple routes available *for the same prefix*, consider that static routes are preferred over BGP routes.

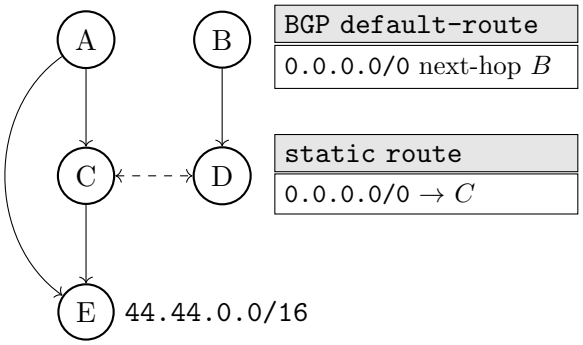

Figure 8: A simple BGP network consisting of five ASes.

(i) Which routes in AS D's routing table could be used for traffic destined to `44.44.0.0/16`? List them in order of preference (from most to least preferred). State the type of route (e.g., BGP), its prefix and the next-hop (e.g., AS C). (3 Points)

(ii) Since AS D has a history of using static routes to cheat BGP, AS C is suspicious of AS D. How can AS C confirm its suspicion and detect the static route?

Note: AS C can observe all traffic on its links and the only traffic present in the network is for `44.44.0.0/16`. (2 Points)

(iii) Assume that the link between AS C and AS E fails, what path does traffic from AS D to `44.44.0.0/16` take? Is it policy-compliant? Explain why or why not.

Note: The setting is still the same and AS C has not taken any action in reaction to the failure or AS D cheating. (2 Points)

d) BGP and IGP: Very creative! **(11 Points)**

leave blank

Consider the AS in Figure 9 with three border routers (A, B, F) and three internal routers (C, D, E). All three border routers receive a route announcement for the prefix `13.0.0.0/8` from their eBGP neighbors (not depicted), which they distribute internally. The iBGP sessions are depicted by double-headed dashed arrows (e.g., router A and F maintain an iBGP session). All routers follow the standard BGP decision process. The three border routers have `next-hop-self` configured on all iBGP sessions.

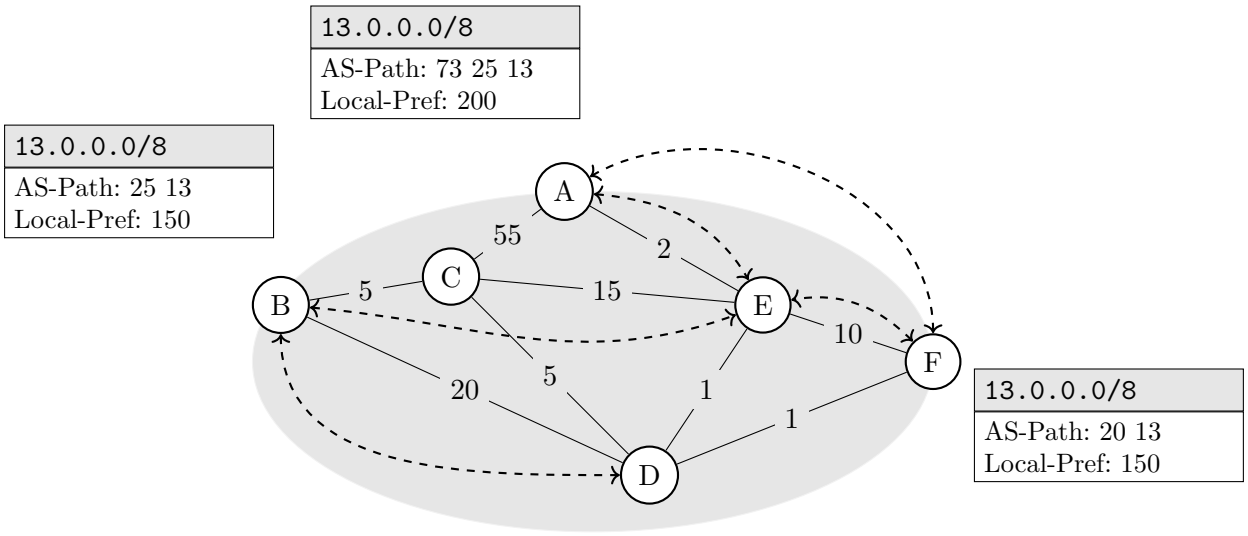

Figure 9: A simple BGP network **not** forming an iBGP full-mesh.

(i) For every router, list *(i)* the BGP next-hop, *(ii)* the path taken by the traffic and *(iii)* indicate whether the router's traffic can actually reach the destination. If the next-hop is external, put EXT. If there is no next-hop, put NO.

Note: We provide the entry for router C, which does not receive a route as it does not have any iBGP session, and the entry for router A. (3 Points)

Router	BGP next-hop	Path taken by the traffic	Reachable?
A	EXT	A → EXT	Yes
B			
C	NO	C → ∅	No
D			
E			
F			

(ii) Assume the eBGP session of router A fails and consequently, **the external route of A is not available anymore**. List for every router *(i)* the BGP next-hop, *(ii)* the path taken by the traffic and *(iii)* indicate whether the router's traffic can reach the destination. If the next-hop is external, put EXT. If there is no next-hop, put NO. (3 Points)

Router	BGP next-hop	Path taken by the traffic	Reachable?
A			
B			
C			
D			
E			
F			

(iii) The network operator reacted and **added a new iBGP session between routers B and C**. The failure still persists, i.e., the external route of A is not available. List for every router *(i)* the BGP next-hop, *(ii)* the path taken by the traffic and *(iii)* indicate whether the router's traffic can reach the destination. If the next-hop is external, put EXT. If there is no next-hop, put NO. (3 Points)

Router	BGP next-hop	Path taken by the traffic	Reachable?
A			
B			
C			
D			
E			
F			

(iv) Route announcements received over an iBGP session are not redistributed over another iBGP session. What is the reason for that? (2 Points)

leave blank

e) BGP Security **(9 Points)**

Consider the Internet topology consisting of 9 Autonomous Systems (ASes) in Figure 10. Single-headed plain arrows point from providers to their customers (AS A is the provider of AS D) while double-headed dashed arrows connect peers (AS A and AS B are peers). Each AS is made up of a single BGP router and applies the default selection and exportation BGP policies based on their customers, peers and providers.

In this task, the routers break ties using the AS number of the neighbor: in case multiple routes are equally good, the router selects the route of the neighbor with the lowest AS number (in alphabetical order; e.g., a route from AS A is preferred over AS B in case of a tie).

AS I is the origin of prefix `33.33.0.0/16` and advertises it to its neighbors. Independently of what the external advertisements are, AS I ***always*** prefers its internal route to reach any IP destination in `33.33.0.0/16`.

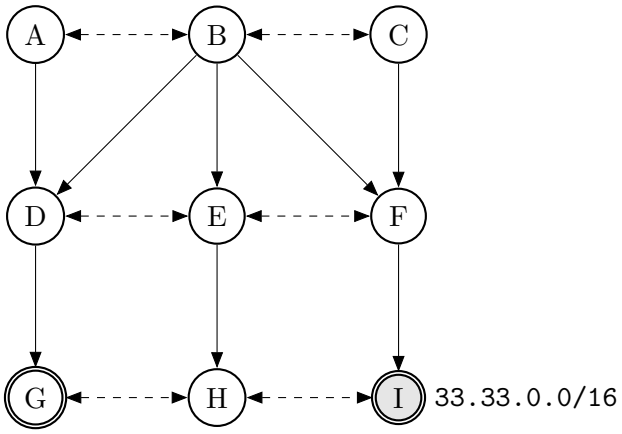

Figure 10: An Internet topology of 9 ASes in which AS I announces a prefix and AS G tries to hijack it.

(i) AS G wants to hijack the traffic going to AS I for `33.33.0.0/16`. It starts advertising the exact same prefix with itself, AS G, as origin. From which ASes is it able to hijack the traffic? (2 Points)

(ii) The ASes notice the hijack and, as a counter-measure, deploy Resource Public Key Infrastructure (RPKI) Internet-wide. After that, from which ASes is the attacker able to hijack the traffic by still advertising the exact same prefix with itself as origin? (1 Point)

(iii) RPKI has a flaw. What is the problem of RPKI? How can AS G hijack the prefix `33.33.0.0/16` despite RPKI? From which ASes is AS G able to hijack the traffic?
(3 Points)

(iv) In response, the ASes switch to BGPSec (Secure BGP). Explain what security it provides and how AS E can detect that the announcement from AS G has a forged AS path.
(3 Points)

leave blank

Task 4: Reliable transport **25 Points**

leave blank

a) Warm-up **(6 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true ☐ false ☐ Consider a 10 Gbps transit link which sees exactly 2 ongoing TCP connections c_1 and c_2 with equal throughput (i.e., 5 Gbps each). c_2 suddenly stops, leaving c_1 as the only flow traversing the link. The throughput of c_1 might not necessarily increase after that event.

true ☐ false ☐ Consider again a 10 Gbps transit link which sees exactly 2 ongoing TCP connections c_1 and c_2. Assuming the RTT of c_1 is twice as low as the one of c_2, c_1 throughput will be twice as high.

true ☐ false ☐ Given that the TCP ports are defined using 16 bits, the maximum number of TCP connections a host can maintain simultaneously is $2^{16} = 65536$.

true ☐ false ☐ Consider a TCP endpoint that has sent exactly 3 segments to another one. The maximum number of duplicate acknowledgments that the TCP endpoint can have received in return is 2.

true ☐ false ☐ TCP endpoints negotiate which congestion control algorithms to use when they establish a connection.

true ☐ false ☐ Increasing the size of the router queues is not sufficient to prevent congestion collapse.

b) How close do you need to be? **(4 Points)**

leave blank

Consider a "TCP-like" sender S and a receiver R which communicate using a directly-connected 10 Gbps link. R's receiver window is 10 Mbytes and you can assume that it is always smaller than S's sender window. Unlike normal TCP, R acknowledges every single bit individually.

Give an equation that describes the throughput achieved between S and R as a function of the round-trip-time `RTT`.

Hint: There are at least two cases to consider.

leave blank

c) Riding the TCP roller coaster (15 Points)

In this question you must compute the evolution of the size of the congestion window (*cwnd*) used by a TCP sender together with the segments it sends as it receives acknowledgments and undergoes timeouts.

The sender implements the congestion control algorithm given below (Algorithm 1). This algorithm is similar to the one seen in the lecture; it includes a slow start, congestion avoidance, and fast recovery phases.

Algorithm 1 TCP Congestion Control Algorithm

```
1:  cwnd ← 1; mss ← 1000 bytes
2:  sstresh ← ∞
3:  while true do
4:      upon receiving a new ACK
5:          if (cwnd < sstresh) then                          ▷ Slow Start
6:              cwnd ← cwnd + mss
7:          else                                              ▷ Congestion Avoidance
8:              cwnd ← int( cwnd + mss / int(cwnd/mss) )      ▷ int(x) returns the integer part of x
9:          dup_ack ← 0

10:     upon timeout
11:         sstresh ← cwnd/2                                  ▷ Multiplicative Decrease
12:         cwnd ← mss

13:     upon receiving a duplicate ACK
14:         dup_ack ← dup_ack + 1
15:         if (dup_ack ≥ 3) then                             ▷ Fast Recovery
16:             sstresh ← cwnd/2
17:             cwnd ← sstresh
18:             retransmit_packet()
```

As shown in Algorithm 1, the Maximum Segment Size (MSS) is 1000 bytes and the computation of *cwnd* are rounded using the integer part. Recall that the integer part $\mathbf{int}(x)$ returns the integer that has the largest value less than or equal to the absolute value of x (e.g., $\mathbf{int}(2.4) = 2$, $\mathbf{int}(2.8) = 2$, $\mathbf{int}(3) = 3$)

Use the following page to indicate the evolution of the size of the congestion window and the segments sent at each step assuming that:

- the TCP connection had a *cwnd* of 10000 bytes and just had a timeout;
- the receiver acknowledges the next expected byte;
- the receiver has acknowledged all data up to bytes 14999 (the next expected byte is 15000);
- the sender *only* sends full-sized packets (i.e., packets of 1000 bytes);
- the sender always has data to send and sends all outgoing packets before processing the next ack.

We filled in the first two lines for you. Fill the other lines **using the same format**. For simplicity, we denote a TCP segment by the first byte it carries, e.g. `17000` indicates the segment carrying bytes [17000–17999].

Event	Size of $cwnd$ in bytes, *after* the event has taken place	First sequence number of *each* packet sent
Timeout	1000	15000
ACK 16000	2000	16000, 17000
ACK 17000		
ACK 18000		
ACK 19000		
ACK 20000		
ACK 21000		
ACK 22000		
ACK 23000		
ACK 24000		
ACK 24000		
ACK 24000		
ACK 24000		
ACK 30000		

leave blank

Task 5: Applications **30 Points**

leave blank

a) Warm-up **(7 Points)**

For the following true/false questions, check either *true*, *false* or nothing. For each question answered correctly, one point is added. For each question answered incorrectly, one point is removed. There is always one correct answer. This subtask gives at least 0 points.

true	false	
☐	☐	Each domain name has a unique IP address.
☐	☐	Subsequent DNS requests to `a.root-servers.net` (`198.41.0.4`) might reach different physical instances.
☐	☐	The content of an email does not affect the choice of the delimiter.
☐	☐	If you retrieve an email using the POP protocol, the mail server will delete the email.
☐	☐	Your mail client automatically puts a dot "." at the end of an email you send.
☐	☐	In video streaming, the client runs at least one HTTP GET request for each chunk.
☐	☐	Video streaming can use text-based encoding.

Your first job!

You have completed your studies at ETH successfully and got a job as a network operator. You find yourself in a meeting with your colleagues. Today's agenda contains two items which correspond to the first two subtasks; in the third subtask, you deal with an unexpected visit of the security team.

b) Agenda, item 1: Improve website loading time (8 Points)

leave blank

The analytics of our company's website indicate that a lot of users give up loading the website because it takes too long—you need to improve something.

(i) While you search through the website implementation for possible improvements, you make a shocking discovery: the webserver is still running HTTP 1.0! This means it supports neither pipelining nor persistent connections. Briefly explain how an upgrade to HTTP 1.1 could help improve the load time *(i)* generally, in terms of the number of RTTs needed to load the website; and *(ii)* more specifically if the content of the site does not fit on the client's screen. Finally, briefly describe the advantages that your Internet Service Provider (ISP) would see if you switched to HTTP 1.1. (4 Points)

General improvements:

Improvements for small screens:

Benefits for the network operator:

(ii) The website is hosted on a web server which fetches the data from several "backend" servers. One of these backend servers contains highly dynamic data used in some graphs. Bob mentions that this server tends to be overloaded by the amount of requests. He considers remedying this by caching the content on reverse proxies. He asks whether you think this will decrease the load on the backend server. If you think so, briefly explain why. If not, briefly explain an alternative solution. (2 Points)

(iii) Besides the graphical data, your website also displays a set of large static pictures which are duplicated on multiple backend servers. As a load-balancing policy, the web server simply round-robins each incoming request for a picture to a different replica—independently of its load. You remember that Netflix uses a more advanced mechanism for balancing the load between its content replicas. Briefly explain Netflix's approach.

(2 Points)

leave blank

c) Agenda, item 2: Improve video streaming (11 Points)

Your company's website is also used for streaming two marketing videos. To do so, your company relies on an HTTP-based architecture in which the client (running on the customer device) requests small chunks of the video, possibly at different bitrates, using subsequent HTTP requests. Unfortunately, it seems that this part too is up for improvement.

(i) A colleague explains that the web client has a button to adapt the resolution in case the user is not satisfied with the resolution choice picked automatically. For the first video, the website analytics indicate that users often increase the resolution and do not switch back to a lower one. However, for the second video, the opposite usually happens—users often decrease the resolution and stick with it. Explain a possible reason for this behavior and describe how you could improve the automated resolution choice. (2 Points)

Possible reason:

Suggested improvement:

(ii) A debate starts on how the client should predict the best resolution for the next chunk. Some people argue for basing the decision on the available capacity, while others argue for basing the decision on how filled the buffer is. Identify one argument for and against each side, and mention which solution Netflix claims to use. (5 Points)

Argument for capacity-based prediction:

Argument against capacity-based prediction:

Argument for buffer-based prediction:

Argument against buffer-based prediction:

Netflix's solution:

(iii) As for the pictures, the actual video chunks are stored on multiple backend servers and the webserver simply round-robins the request through them. You remember from the CommNet lecture that the client's browser downloads a `Manifest` file when it starts streaming a video which contains URLs for all chunks. At the moment, all URLs contain the same DNS domain which resolve to the IP address of the web server itself. For example, the URLs of the first two chunks for the first video are:

`https://domain.ch/video1/2s_480p/chunk_1.m4s`
`https://domain.ch/video1/2s_480p/chunk_2.m4s`

You wonder whether DNS might be useful for load-balancing by dynamically adapting the resolution of `domain.ch` to the IPs of the different backends. Name two potential drawbacks of such a solution. (2 Points)

Drawback 1:

Drawback 2:

(iv) Another option comes to your mind. To have even more flexibility, one could assign a distinct domain name to *each* video chunk and make them resolve to different backends. For example, the URLs of the first two chunks for the first video could be:

`https://chunk0001.domain.ch/video1/2s_480p/chunk_1.m4s`
`https://chunk0002.domain.ch/video1/2s_480p/chunk_2.m4s`

Explain two drawbacks of this approach. (2 Points)

Drawback 1:

Drawback 2:

leave blank

d) Agenda, unscheduled item 3: Spam issues **(4 Points)**

One of your colleagues storms into the meeting, reporting a massive spam mail campaign. The spam mail generators are using your company's domain for the sender address. This may damage the reputation of your company, so you think about using SPF (Sender Policy Framework). Explain the mechanism and one attack scenario which it does not protect against.

How SPF works:

Nonetheless possible attack:

Extra Sheet 1

In case you need more space, use the following pages. Make sure to always indicate the task to which the answer belongs (e.g., *3 d) (ii)*).

Task: _____

Extra Sheet 2

Task: _____

Task: _____

Extra Sheet 3

Task: _____

Task: _____