

CSEE4119 Computer Networking HW4

Tong Wu, tw2906

Question 1

Part a

Section i

It is **not possible** to forward two packets at the same time in a shared bus, since the shared bus can only deliver 1 datagram to 1 output at the same time.

Section ii

It is **possible** to forward two packets to different output ports at the same time, since the crossbar has multiple links between input ports to output ports, as long as these two packets are coming from different input ports.

Section iii

It is **not possible** to forward two packets to the same output port at the same time by using crossbar, as said in section ii, the crossbar can only deliver packets from different input ports to different output ports.

Part b

Section i

Five packets in three input ports need to spend **5 time slots** to transfer all packets to output ports by switch fabric. As said in part a, section i, the switch fabric can only deliver 1 packet in each time slot.

Section ii

For the first time slot, X in the first input port is sent to the output port X, and Y in the second input port is sent to the output port Y.

For the second time slot, X in the second input port is sent to the output port X, and Y in the third input port is sent to the output port Y.

For the third time slot, Z in the third input port is sent to the output port Z.

OR:

For the first time slot, X in the first input port is sent to the output port X, and Y in the third input port is sent to the output port Y.

For the second time slot, Y in the second input port is sent to the output port Y, and Z in the third input port is sent to the output port Z.

For the third time slot, X in the second input port is sent to the output port X.

Both two deliver sequences use three time slots.

Part c

Section i

Value of r should be 4, according to the diagram in Facebook website [17], there are 4 inputs and outputs LC.

Value of n should be 16, where the number of I/O ports is 16.

Value of m should be 16.

Section ii

$$4 \times 16 \times 16 = 1024 \text{ GE}$$

Section iii

According to the paper [18], if the number of m is bigger than $2n-1$, then the network is non-blocking.

The value of m is 16, and value of n is 16. So in traditional Clos network, it is blocking.

Section iv

Facebook decrease the intermediary switch number to 12 [17], and reduce the input/output port to 6.

Section v

$$12 \times 4 \times 4 = 192 \text{ GE}$$

Section vi

According to the paper [18], if the number of m is bigger than $2n-1$, then the network is non-blocking.

The value of m is 12, and value of n is 6. So in traditional Clos network, it is non-blocking [17].

Section vii

Clos network using more small throughput switch in order to instead the large switch is because that the small scale switch is more cost-saving.

Section viii

$$128 \times 100 \text{ GE}$$

Section ix

A single-chip system allows for easier management and operations [19].

Section x

$$n = 2$$

$$r = 8$$

$$m = 8$$

INPUT SWITCHES:8, INTERMEDIARY SWITCHES: 8, OUTPUT SWITCHES:8

Part d

Section i

For RR, each round (time slot) the packets will be picked followed by the queue (1 → 2 → 3 → 1 → ...). The transmission order should be:

A → C → B → H → F → D → G → E

Section ii

The weight of class 1 and class 2 is 1, and class 3 is 2. So the allocation of bandwidth for three classes should be 1 : 1 : 2. The allocation of bandwidth should always obey the ratio.

Transfer timeline is written below.

The transfer timeline shows what each packet is doing in each important time stamp, including the event of start and finish the transfer, allocated bandwidth, transferred length units, remaining length units and time units.

The transmission order indicates the order of sequence that packets arrive at the destination.

Time 0:

The packet A starts transfer, since packet A is the only one that arrives at time 0.

Packet A will get full bandwidth, needs 8 time units to finish.

Transmission order: None

Time 5:

Packet B and C arrive and start transfer now.

Packet A will get $\frac{1}{4}$ of total bandwidth, transferred 5 length units, 3 length units to go, 12 time units to go.

Packet B will get $\frac{1}{2}$ of total bandwidth, 12 time units to finish.

Packet C will get $\frac{1}{4}$ of total bandwidth, 40 time units to finish.

Transmission order: None

Time 17:

Packet D, E, F, G arrives but queued. Packet A and B finished transmission. For the first input port, packet H does not arrive yet, so no packet will be sent at this time. For the third input port, packet D arrived, packet D starts transfer now.

Packet C will get $\frac{1}{3}$ of total bandwidth, transferred 3 length units, 7 length units to go, 21 time units to go.

Packet D will get $\frac{2}{3}$ of total bandwidth, 13.5 time units to finish.

Transmission order: A → B

Time 20:

Packet H arrives, transfer now.

Packet H will get $\frac{1}{4}$ of total bandwidth, needs 32 time units to finish.

Packet D will get $\frac{1}{2}$ of total bandwidth, transferred 2 length units, 7 length units to go, 14 time units to go.

Packet C will get $\frac{1}{4}$ of total bandwidth, transferred 4 length units, 6 length units to go, 24 time units to go.

Transmission order: A → B

Time 34:

Packet D finish transmission. Packet E starts transfer now.

Packet H will get $\frac{1}{4}$ of total bandwidth, transferred 3.5 length units, 4.5 length units to go, 18 time units to go.

Packet C will get $\frac{1}{4}$ of total bandwidth, transferred 7.5 length units, 2.5 length units to go, 10 time units to go.

Packet E will get $\frac{1}{2}$ of total bandwidth, 16 time units to finish.

Transmission order: A → B → D

Time 44:

Packet C finish transmission. Packet F starts transfer now.

Packet H will get $\frac{1}{4}$ of total bandwidth, transferred 6 length units, 2 length units to go, 8 time units to go.

Packet E will get $\frac{1}{2}$ of total bandwidth, transferred 5 length units, 3 length units to go, 6 time units to go.

Packet F will get $\frac{1}{4}$ of total bandwidth, 24 time units to finish.

Transmission order: A → B → D → C

Time 50:

Packet E finish transmission. No more packets from class 3.

Packet H will get $\frac{1}{2}$ of total bandwidth, transferred 7.5 length units, 0.5 length units to go, 1 time units to go.

Packet F will get $\frac{1}{2}$ of total bandwidth, transferred 1.5 length units, 4.5 length units to go, 9 time units to go.

Transmission order: A → B → D → C → E

Time 51:

Packet H finish transmission. No more packets from class 1.

Packet F will get full bandwidth, transferred 2 length units, 4 length units to go, 4 time units to go.

Transmission order: A → B → D → C → E → H

Time 55:

Packet F finish transmission. Packet G starts transfer now.

Packet G will get full bandwidth, 10 time units to finish.

Transmission order: A → B → D → C → E → H → F

Time 65:

Packet G finish transmission.

Transmission done.

Transmission order: A → B → D → C → E → H → F → G

Question 2

Part a

Section i

The original subnet prefix is 127.0.8.0/23. According to the binary conversion, 127.0.8.0 to binary is:

```
1 | 01111111 00000000 00001000 00000000
```

The subnet mask is 23, which means the first 23 bits are not allowed to change, the IPv4 address has 32 bits, so the subnet prefix becomes:

```
1 | 01111111 00000000 0000100X XXXXXXXX
```

Now the first subnet needs to contain 12 users. Using the /28 subnet mask, which can contain 14 users. So the subnet prefix should be: **127.0.8.0/28**

```
1 | 01111111 00000000 00001000 0000XXXX
```

Note that **/28 subnet contains 16 IP address**, the first IP address in the subnet is allocated as network IP address, and the last IP address is allocated as broadcast address, which are reserved for use. So 127.0.8.0/28 has **14 IP addresses are free to allocate**.

Section ii

The second subnet needs to contain 61 users, the upscale is 6, where $2^6 - 2 = 62$, so the subnet mask is $32 - 6 = 26$. The subnet prefix should be: 127.0.8.0/26

```
1 | 01111111 00000000 00001000 00XXXXXX
```

If this is the answer for a dedicate question then it is fine. However, there is a subnet already use the last 4 bits, so if the second subnet use 127.0.8.0/26, there will be several IP addresses conflict with the first subnet. By changing the 24th, 25th and 26th bit can avoid the conflict and under the original subnet.

By changing the 26th bit from 0 to 1:

```
1 | 01111111 00000000 00001000 01XXXXXX
```

There, the second subnet prefix becomes: **127.0.9.64/26**

Note that the **/26 subnet contains 64 IP addresses**. However, the first and the last IP address is reversed for network IP address and broadcast address. So the subnet prefix 127.0.9.64/26 has **62 IP addresses are free to allocate**.

Section iii

The third subnet needs to contain 61 users, the upscale is 6, where $2^6 - 2 = 62$, so the subnet mask is $32 - 6 = 26$.

Same problem as the previous section, so changing the 25th bit from 0 to 1:

```
1 | 01111111 00000000 00001000 10XXXXXX
```

Now the third subnet prefix should be: **127.0.8.128/26**

Note that the **/26 subnet contains 64 IP addresses**. However, the first and the last IP address is reversed for network IP address and broadcast address. So the subnet prefix **127.0.9.128/26 has 62 IP addresses are free to allocate**.

Section iv

The last subnet needs to contain 64 users, the upscale is 7, where $2^7 - 2 = 124$, so the subnet mask is $32 - 7 = 25$.

```
1 | 01111111 00000000 00001000 0XXXXXXX
```

Same problem as the previous section, so change the 24th bit from 0 to 1:

```
1 | 01111111 00000000 00001001 00XXXXXX
```

Now the last subnet prefix should be: **127.0.9.0/25**

Note that the **/25 subnet contains 128 IP addresses**. However, the first and the last IP address is reversed for network IP address and broadcast address. So the subnet prefix **127.0.9.0/25 has 126 IP addresses are free to allocate**.

Part b

For the key change from IPv4 to IPv6:

The IPv6 is not allowed to fragmentation at the router side, while IPv4 needs to do fragmentation at destination or at routers [1].

Part c

The reason of IPv6 is not allowed to fragmentation at the router side is that, one is the fragmentation cost performance of routers, but more important, IPv6 can fragment at sender, and each fragment has to go through every router and the router needs to store information of extension header [1]. That's why fragmentation of IPv6 cannot be done in router.

Part d

For another key change from IPv4 to IPv6:

IPv6 header does not contains checksum field, while IPv4 support.

Part e

The reason of IPv6 header does not contain checksum field is that, the whole-packet link layer checksumming is provided in protocols, such as Ethernet, and upper layer protocols (TCP, UDP) also use checksum, which are sufficient and makes IPv6 is no need to contain checksum [2].

Part f

Although the IPv6 flag day is defined at February 1st, 2030 [3], but for the personal thought, it is hard to be applied and IPv4 and IPv6 may have to continue to coexist for a long time.

IPv4 is too widely used, which makes a one-size-fits-all switchover approach confusing. A reasonable approach would be to gradually allow IPv6 to dominate and gradually abandon IPv4, but in the meantime, IPv4 support should not be abandoned. Although the thought is perfect, it is more likely that IPv6 flag day may be far in the future, and it is even possible that IPv6 routing will have to be permanently compatible with IPv4.

Part g

RFC 801

Part h

1 January 1983 [4]

Part i

NCP [4]

Part j

TCP/IP flag day influence a small area. The TCP/IP flag day only influence the ARPANET [5], which was established by ARPA and DOD [6]. So the change affected a relatively small community, and And if an error occurs, the problem can be easily found and quickly fixed.

Part k

Dual stack means that devices can run IPv4 and IPv6 parallel (at the same time), it allows hosts able to reach both IPv4 and IPv6 content [7].

Part l

IPv4/IPv6 tunnelling allows an existing IPv4 routing infrastructure to carry IPv6 traffic [8], connects an isolated IPv6 site through IPv4 network [9].

Part m

The dual stack approach needs all devices in the network has been assigned IPv4 and IPv6 addresses, if one or more devices are IPv6/IPv4 only, then the dual stack approach is not working [10].

The tunnelling approach can deal the problem presented by dual stack, which using tunnel provides media to transfer packets between IPv6 and IPv4. However, frequently using tunnelling in network will increase the complexity and cost, and tunnelling has to be manually configured [11]. Dual stack approach can deal with this problem.

Part n

An IPv6 address has 128 bits, the first 64 bits are network ID (more specific, first 48 bits form global routing prefix, last 16 bits form subnet [13]), the remaining 64 bits are host ID [12], so when ISP assigning IPv6 addresses, the addresses is more likely continuous. From this point, the longest prefix matching can take more advantage on IPv6, since the longer the prefix that matched, means it may in a subnet or the neighbouring interface, makes the network propagation faster. However, the long address of IPv6 cause more performance costing to read the routing table.

Question 3

Part a

iBGP protocol.

Part b

AS2 has no physical link with AS4, and 1d will put an entry in its forwarding table once 1d learns about x. If AS2 wants to communicate with AS4, it should send packets to the gateway router 2a, then 2a sends packets to 1b using eBGP protocol, then 1b sends packets to 1d, which has forwarding table to AS4. Router 1d needs to send packets to AS1's gateway router with AS3, which is 1c, so according to the least cost path, router 1d should choose l1 as the path.

Part c

In BGP route selection, the local preference is the first element of consideration, so if AS1 assigns a higher local preference to routes learned from AS3 than AS2, then AS1 will choose routes passing through AS3. In this case, l1 will be chosen since it has the least AS-PATH.

Part d

According to the BGP route selection, the local preference is the first consideration, which is equal for AS2 and AS3. Shortest AS-PATH is the second consideration, where both l1 and l2 will pass one AS, which is also the same. The closest NEXT-HOP is the third consideration, while the question content does not indicate the intra-domain cost. In this case, l may change from l1 to l2 if 1d → 1b has less intra-domain cost than 1d → 1a.

Part e

If AS5 is added between AS2 and AS4, then the AS-PATH passing through AS2 will change from AS2 to AS2, AS4, which will pass through two ASes. However, the route passing through AS1 remains one AS-PATH, so l1 will be chosen.

Part f

Because the IGP can tell the router how to access the next hop, but iBGP cannot, so that's why NEXT-HOP must be reachable without using BGP, but can via IGP.

Part g

Because Bellman-Ford equation considers both movement cost to the neighbour router, and the further cost that this neighbour to the final destination [12], which will ensure that the selected router will not be re-selected by the algorithm. In this case, the routing loop will not be created if the link costs are static.

Part h

Because the Bellman-Ford equation needs to know all metrics of each route, in eBGP, it is not possible to get all information that BF needs at one time. In this case, eBGP uses calculating AS-PATH to determine the distance of each routes and prevent loop.

Part i

For iBGP, split horizon rule indicates that the update sent by one iBGP neighbour should not sent to another iBGP neighbour [13].

Question 4

Part a

TATA will not carry the traffic from Los Nettos destined to NYSERNET. The reason is that, Los Nettos and NYSERNET has a peering link, but TATA is Tier-1 router, which is Los Nettos' provider's provider. According to the prefer-customer routing, AS should always prefer peers than providers.

Part b

Los Nettos will not carry traffic from TATA destined to NTT, since the NYSERNET prefers to Los Nettos, but not its provider NTT. Then, Los Nettos will choose AT&T since it has less AS-PATH to NTT. Finally, AT&T transmit to NTT.

Part c

NYSERNET will not carry traffic from TATA destined to NTT, since TATA and NTT are peer relations, and TATA will prefer peer transmission according to prefer-customer routing policy.

Part d

TATA will not carry traffic from NYSERNET to NTT, since both TATA and NTT are NYSERNET's provider, but TATA has more AS-PATH than directly transmit to NTT.

Part e

In order to route all traffics destined to level 3 via TATA if it is exist, and use NTT as a backup, the router should configure:

1. Add TATA's (loopback) IP address to NYSERNET's BGP router configuration.
2. Active TATA BGP route.
3. Configure an additional path as an backup by using `bgp additional-paths select backup` [14].
4. Add NTT's (loopback) IP address to NYSERNET's BGP router backup path configuration [14].

Part f

In order to set NYSERNET prefers to receive traffic via TATA when sender does not has a strong preference, but will let the sender choose if sender has a strong preference, the router should configure:

1. Add TATA's (loopback) IP address to NYSERNET's BGP router configuration.
2. Active BGP route
3. Configure the TATA BGP route local preference as 1 `set protocol bgp local-preference 1`, in order to use TATA first if sender has no preference.
4. On the other router side, set local preference of BGP route to NYSERNET bigger than 1 if it has strong preference.

Part g

Prefix	AS Path(s) of Best Route from AS1
1.0.0.0/8	
2.0.0.0/8	2
3.0.0.0/8	3
4.0.0.0/8	4
5.0.0.0/8	3,5
6.0.0.0/8	2,5
7.0.0.0/8	3,7 OR 4,7
8.0.0.0/8	2,6,8
9.0.0.0/8	2,6,9
10.0.0.0/8	2,6,10
11.0.0.0/8	3,7,11 OR 4,7,11
12.0.0.0/8	3,7,12 OR 4,7,12

Part h

AS10 can receive a route to 5.0.0.0/8 from AS6, since AS6 is sharing a provider's advertisement to its customer, which is not obey the no-valley routing policy.

Part i

AS5 cannot receive a route to 7.0.0.0/8 from AS6, since if so, AS6 is sharing a peer's advertisement to its provider, which is obey the no-valley routing policy.

Part j

AS6 cannot receive a route to 4.0.0.0/8 from AS7, since if so, AS7 is sharing a provider's advertisement to its peer, which is obey the no-valley routing policy.

Question 5

Part a

Section i & Section ii

S2 Flow Table

Match	Action
ingress port=4, dest=10.3.0.*	forward(1)
ingress port=3, dest=10.1.0.*	forward(2)

Part b

Since the flow table does not clarify the action of host 4 sends a datagram to host 3, then it may drop the datagram, or sent to OpenFlow controller.

Part c

Section i & Section ii

S2 Flow Table

Match	Action
src=10.1.0.*, dest=10.2.0.3, dest_port=80	forward(3)
src=10.1.0.*, dest=10.2.0.4, dest_port=80	forward(4)

Part d

Using the priority to disambiguate overlapping patterns. In this case, set the second rule as higher priority to make sure the traffic of h3 will use the second rule.

Part e

Section i

Doing routing in software can decrease the cost in hardware, and can easily mitigate between DCs and networks [15].

Section ii

Centralized perspective of routing is easy to access the status information [16].

References

- [1] IPv6 Fragmentation Header. (2020, May 20). *GeeksforGeeks* . <https://www.geeksforgeeks.org/ipv6-fragmentation-header/>
- [2] Internet checksum. (2022). In *Wikipedia* . https://en.wikipedia.org/w/index.php?title=Internet_checksum&oldid=1125166220
- [3] *IPv4 flag day* . (n.d.). IPv4 Flag Day. Retrieved 4 December 2022, from <http://ipv4flagday.net/>
- [4] *NCP/TCP transition plan* (Request for Comments RFC 801). (1981). Internet Engineering Task Force. <https://doi.org/10.17487/RFC0801>
- [5] Flag day (computing). (2022). In *Wikipedia* . [https://en.wikipedia.org/w/index.php?title=Flag_day_\(computing\)&oldid=1122755616](https://en.wikipedia.org/w/index.php?title=Flag_day_(computing)&oldid=1122755616)
- [6] ARPANET. (2022). In *Wikipedia* . <https://en.wikipedia.org/w/index.php?title=ARPANET&oldid=1125330914>
- [7] *Dual Stack Network* . (2010). CISCO. https://www.cisco.com/c/dam/en_us/solutions/industries/docs/gov/IPV6at_a_glance_c45-625859.pdf
- [8] *IBM Documentation* . (2022, November 15). <https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/aix/7.2?topic=6-ipv6-tunneling>
- [9] *IPv6 over IPv4 tunneling Feature Overview and Configuration Guide* . (n.d.). Allied Telesis. https://www.alliedtelesis.com/sites/default/files/documents/configuration-guides/ipv6_over_ipv4_tunneling_feature_overview_guide.pdf
- [10] Doyle, J. (2009, June 4). *The Dual Stack Dilemma* . Network World. <https://www.networkworld.com/article/2235990/the-dual-stack-dilemma.html>
- [11] *Using Tunneling to Transition to IPv6* . (n.d.). Retrieved 4 December 2022, from <http://www.globalknowledge.com/us-en/resources/resource-library/articles/using-tunneling-to-transition-to-ipv6/>
- [12] Fumie Costen. (n.d.). *EEEN30024 Data Networking 2021-22 1st Semester. Theory 7: Routing algorithms* . The University of Manchester.
- [13] BGP Loop prevention. (2015, February 26). *Network Study* . <http://network.jecool.net/bgp-loop-prevention/>
- [14] *IP Routing: BGP Configuration Guide – BGP Diverse Path Using a Diverse-Path Route Reflector [Cisco ASR 1000 Series Aggregation Services Routers]* . (n.d.). Cisco. Retrieved 6 December 2022, from https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_bgp/configuration/xr-16/irg-xr-16-book/irg-diverse-path.html
- [15] *How will virtual router software benefit the networking industry?* | *TechTarget* . (n.d.). Networking. Retrieved 6 December 2022, from <https://www.techtarget.com/searchnetworking/answer/How-will-virtual-router-software-benefit-the-networking-industry>
- [16] Patel, A. (2020, June 14). Difference between Centralized Routing and Distributed Routing. *Propatel* . <https://www.propatel.com/centralized-routing-distributed-routing/>

- [17] User, I. (2015, February 11). Introducing “6-pack”: The first open hardware modular switch. *Engineering at Meta* . <https://engineering.fb.com/2015/02/11/product/on-engineering/introducing-6-pack-the-first-open-hardware-modular-switch/>
- [18] Clos, C. (1953). A study of non-blocking switching networks. *The Bell System Technical Journal* , 32 (2), 406–424. <https://doi.org/10.1002/j.1538-7305.1953.tb01433.x>
- [19] Gunning, M. (2019, March 14). Reinventing Facebook’s data center network. *Engineering at Meta* . <https://engineering.fb.com/2019/03/14/data-center-engineering/f16-minipack/>