CS339 Computer Networks Chapter 4 -- Network Layer

1. What Does Routers in Network Layer Do?

1. Routing: Determine the route from source to destination.
2. Forwarding: Move the packet from input to output.
3. Congestion Control: Drop packet.

2. IP Service is Simple

No connection state, unreliable.

So that the reliable service can be developed on the top of the network layer.

Also, IP requires almost nothing from lower layer, which makes it suitable for most of the connection patterns.

3. Forwarding Table

Use **32-bit binary number** to represent an `IP` address. To match the target `IP` address with certain entry, we obey the **longest prefix matching** rule.

4. Virtual Circuit

To implement the virtual circuit in `IP`, the router will give each link a `VC` number. These `VC` numbers will be locally unique and maintained unchanged. The packet will contain the `VC` number instead of destination `IP` address when going forward. Thus, the same packet will always follow the same route to the destination.

5. Router Architecture

The router has two parts: routing processor and switching fabric. Routing processor runs routing protocols, while the switching fabric directs the packets to the correct output port.

5.1 Three Types of Switching Fabric

1. Memory
2. Bus
3. Crossbar

5.2 Input Port Queuing

Head of the line blocking: When the packet at the head of the queue get blocked. The packets after it were able to be transmitted to some idle output port, but due to the head block, it cannot be transmitted. This is head of the line blocking.

The input port queuing may cause packet loss if the input buffer is full.

5.3 Output Port Queuing

This happens when the packet arrives at output port faster than the speed of sending out. When the output port is full, packet may be dropped.

6. Packet Scheduling

Just know two kinds of scheduling:

1. Fair scheduling
2. Weighted fair scheduling

7. Buffer Design

Usually the buffer size is designed to be

$$\mathrm{RTT} \times C$$

where `C` is the bandwidth of the link.

When there are `N` flows, the buffer size should be equal to

$$\frac{\mathrm{RTT} \times C}{\sqrt{N}}$$

8. `IP` Datagram Format

Some important fields are: version (4 or 6), time to live, destination `IP` address. Upper layer protocol field is important as well, `ICMP=1 TCP=6 UDP=17`. Another point is that `IP` header defines the length of the datagram, `UDP` defines this as well, but `TCP` **doesn't**. This means that `TCP` length is actually bounded by `IP` datagram. Similarly, we will see that the Ethernet also has a `MTU`, which requires **breaking big datagrams into different frames and then reassembling them**.

9. IP Addressing

Divide the `IP` addresses into 5 classes. Class A has prefix `0`, class B has prefix `10`, class C has prefix `110`, class D has prefix `1110`, class F has prefix `1111`. But class D and E are for multicast and reserved use respectively. Among A,B,C classes, there are certain addresses are used as private addresses. For class A, `10.0.0.0 - 10.255.255.255` are private addresses; for class B, `172.16.0.0 - 172.31.255.255` are private addresses; for class C, `192.168.0.0 - 192.168.255.255` are private addresses.

10. Subnet

Definition of Subnet: The hosts in the subnet can communicate with each other without sending information through the gateway router.

10.1 The Subnet Aims to Solve:

Problem: The IP addresses are two many for the users in subnets, but not enough for the routers/subnets.

Solution: Borrow some bits from the host field and combine it with the prefix to form a longer prefix field.

10.2 Subnet Mask

Use two kinds of representations, `xxx,xxx,xxx,xxx/xx` or use 1 to indicate subnet ID bits.

10.3 Subnet IP Address Assignment

Consider assign subnet ID part first: the number of department is fixed.

Consider assign host ID part first: the number of users in one department is fixed.

10.4 VLSM: Variable Length Subnet Mask

Namely you can decide the length of the subnet mask according to the practical use.

11. Supernet

11.1 CIDR: Classless InterDomain Routing

Problem: The number of subnets are too many for the router to remember. The routing table becomes too big.

Solution: Aggregate multiple entries with the same index into one big entry with shorter subnet mask. If there is a hole, write another entry for that hole explicitly. Since the IP address matching obeys the longest prefix matching principle, there will be no possibility for IP address in hole goes to other ports.

12. DHCP

`DHCP` protocol is used to get `IP` address from server. `DHCP` protocol is broadcast from the client. When the host receives the `DHCP`, it replies with the its `IP` address and the allocated `IP` address.

13. NAT

NAT is used to solve the problem of IP address shortage. All the devices behind the NAT uses local `IP` addresses. NAT will use **one** public `IP` address and 16-bit port number to communicate with outside.

14. IPv6

Recall that `IPv4` has 4 byte address field (32 bits), `IPv6` has 16 byte address field and thus can make devices in the world have their own unique address.

14.1 IPv6 Header

The IPv6 header is larger than IPv4 header, but it contains less fields. It contains the version, traffic class, next header (upper level header) and hop limit (which is TTL in IPv4), source address and destination address. Since it contains less field, IPv6 is more efficient than IPv4.

15. VPN

VPN is easy. Just add an `IP` header to the original header, so that the packet is directed to some hop in distance. At that hop, the outer `IP` header is dropped and the real `IP` header will direct the packet to the destination.

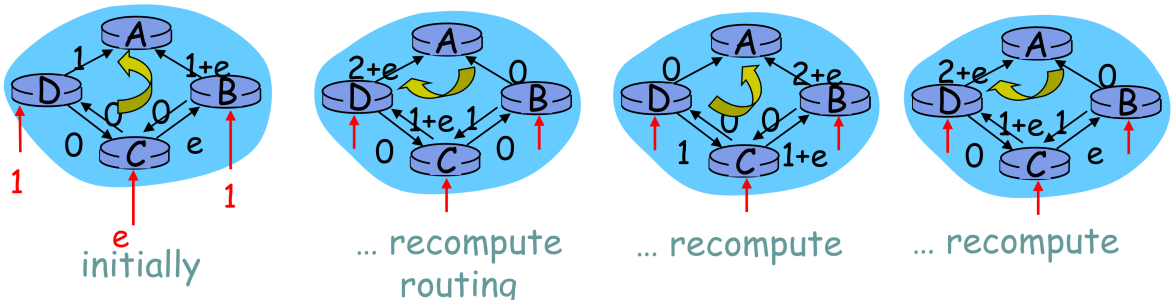
16. Link State Routing

Each router has the whole topology, and thus can use **Dijkstra algorithm** to find the shortest path in

$$\mathcal{O}(n \log n)$$

time.

Problem: The LSR may experience oscillation. That is, all the traffic oscillates between two links.



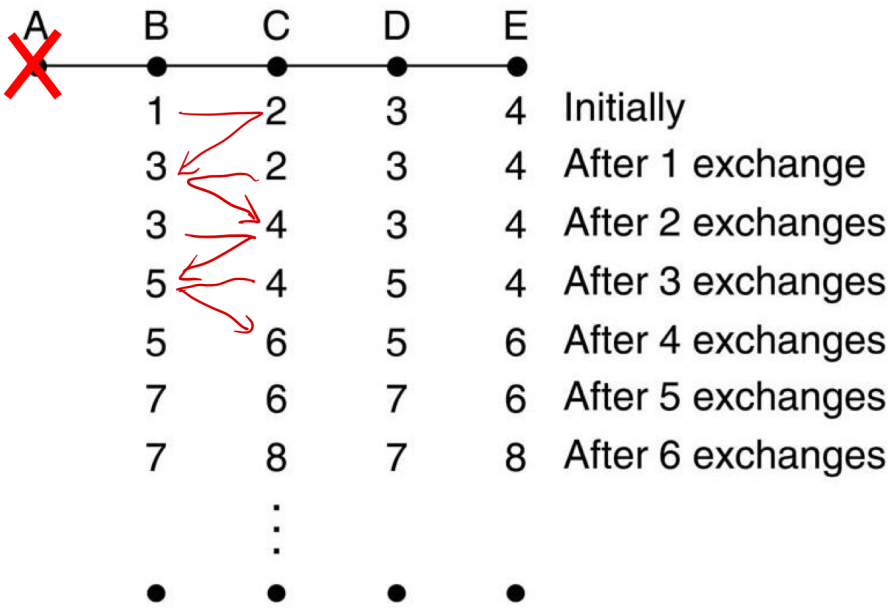
17. Distance Vector Routing

Each router only knows the information from its neighbor, and thus us dynamic programming to get the shortest distance. The recurrence relation is the **Bellman Ford Equation**. The algorithm runs in

$$\mathcal{O}(n)$$

time in total.

Problem: When a node quits the network, its neighbors will ask their neighbors to update their distance to this quitting node. The inter-dependent distance will grow to infinite at last. This is called **count-to-infinity problem.** This problem can be solved in several ways, which will be discussed later.



18. Hierarchical Routing

In current network worldwide, the routers are usually divided into regions called "autonomous systems (AS)". Thus, the routing algorithms also have **intra-AS** and **inter-AS** algorithms. Routers in the same AS runs the same protocols. For inter-AS cases, policy becomes the most significant factor and the traditional routing algorithms may not be applied to this case.

19. Real Routing Protocols

19.1 RIP

RIP implements DVR. It **limits the network to 15 hops**, and this solves the count-to-infinity problem.

19.2 OSPF

OSPF implements LSR. The point is the hierarchical OSPF, which is a multi-layer routing model. It uses OSPF in intra-AS condition and uses BGP in inter-AS condition.

19.3 BGP

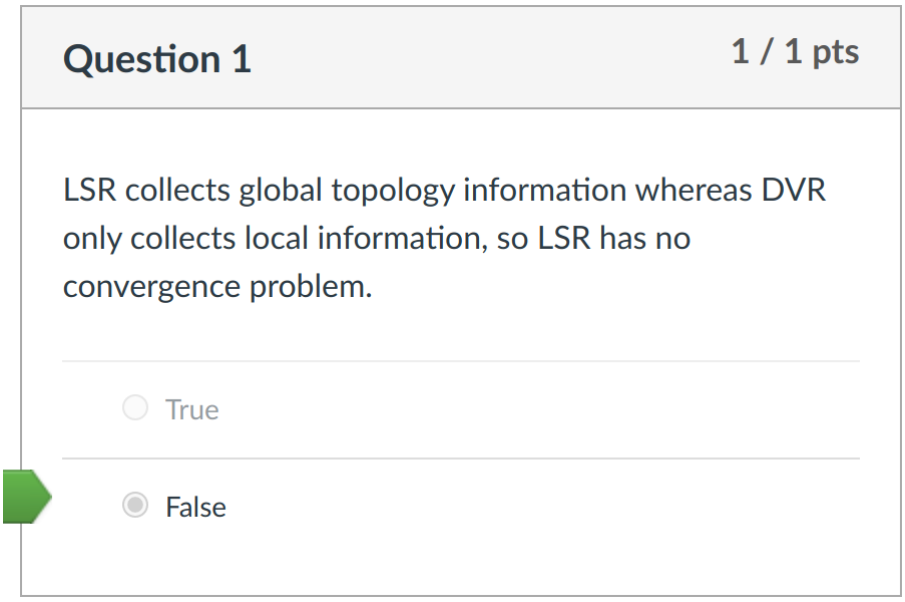
BGP implements DVR. It **adds the route information to the distance information** as well. So that when the count-to-infinity problem is about to happen, the router will not only judge the distance but also the last router corresponding to that distance. In this way, the inter-dependency can be detected and count-to-infinity problem can be avoided.

20. Broadcast Routing

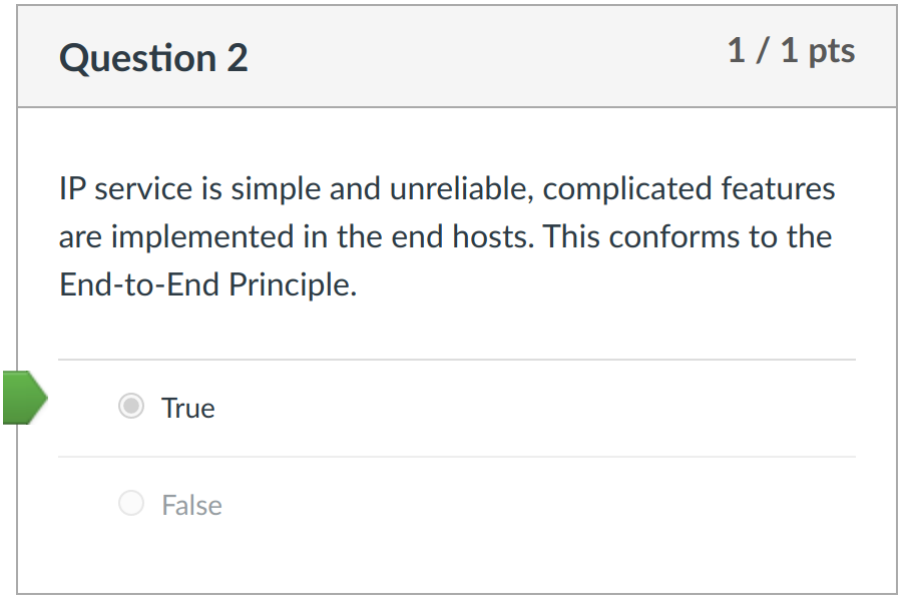
Broadcast routing is inefficient and most of IPv4 routers do **not** support broadcasr/multicast routing.

- Flooding: Send packets to all neighbors. May cost broadcast storm/cycles.
- Controlled Flooding: Use TTL to control the flooding range.
- Spanning Tree: No duplicated packet.

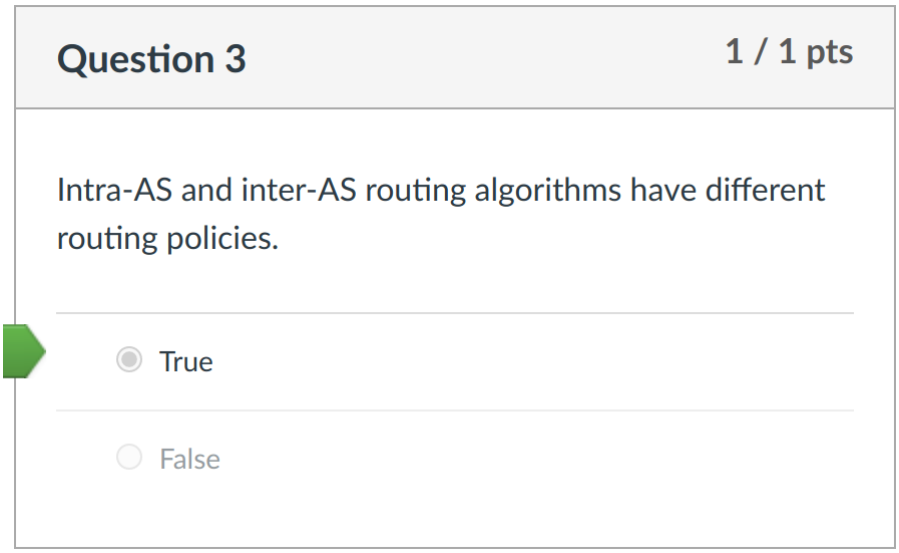
Homework and Quiz Review



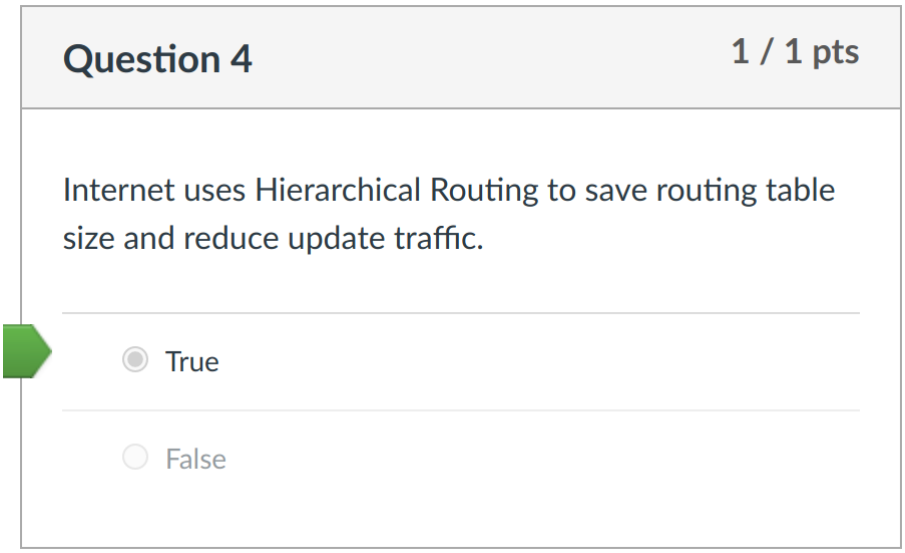
Explain: LSR will have oscillations.



Explain: We discussed in point 2. IP protocol is simple, giving the upper layer more freedom.



Explain: The inter-AS usually has more policy restrictions.



Explain: Hierarchical routing can handle large scale network routing problem by dividing layers.

Question 5

1 / 1 pts

CIDR is designed to deal with the shortage of IPv4 addresses, whereas VLSM is designed to solve the problem of routing table explosion.

- True
- False

Explain: Exchange them. CIDR is superneting and VLSM is subneting.

Question 6

1 / 1 pts

OSPF uses LSR and Hierarchical Routing to find the least-cost path, whereas BGP uses DVR, together with local policies, to find a good path instead of best path, with explicit AS path to avoid count-to-infinity problem.

- True
- False

Explain: Indeed, use the definition.

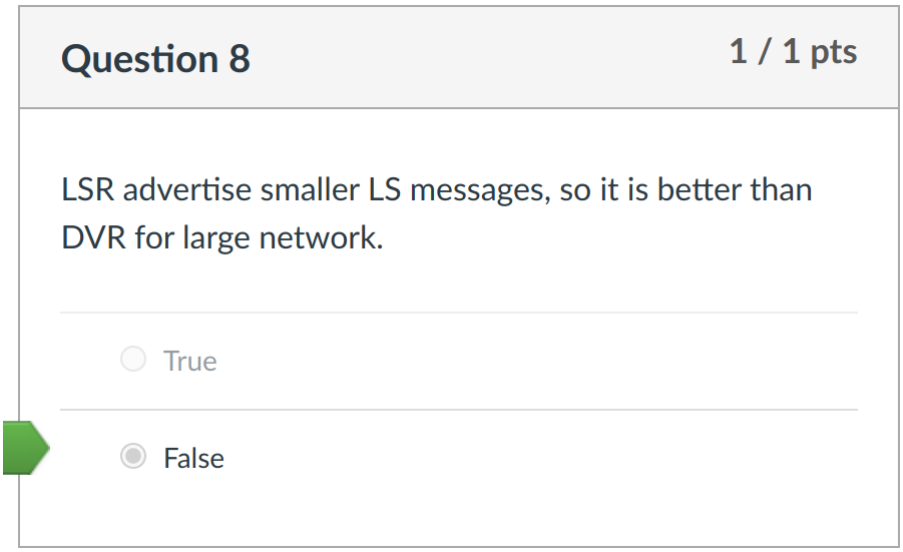
Question 7

1 / 1 pts

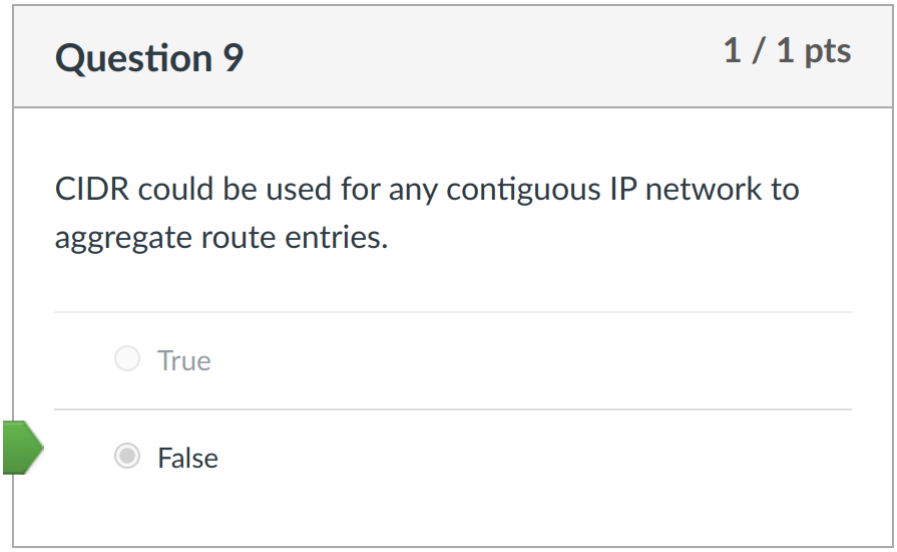
IPv6 is more efficient than IPv4 because IPv6 header is smaller than IPv4 header.

- True
- False

Explain: No, IPv6 has larger header but less fields. Less fields make it efficient.



Explain: No, for a large network, LSR needs to remember a huge map, which is not memory friendly. Obviously, DVR(DP) is better.



Explain: You need to have a whole group of entries that covers all the possibilities of one prefix. If you have `01` (last two digits) and `10` (last two digits) with same prefix. You cannot aggregate them together since `00` and `11` does not exist.

Question 10

1 / 1 pts

Internet uses packet switching for the routers to store and forward the packets. The router's buffer is usually designed big enough to store the incoming packets, but when network congestion happens, the buffer can be full and the packets can be dropped and lost.

- (●) True
- ( ) False

Explain: Definitely.

Question 11

1 / 1 pts

NAT uses port number to multiplex IP address, therefore violates the independence layering principle.

- (●) True
- ( ) False

Explain: Yes, NAT is a way without better choice. It violates the layering principle.

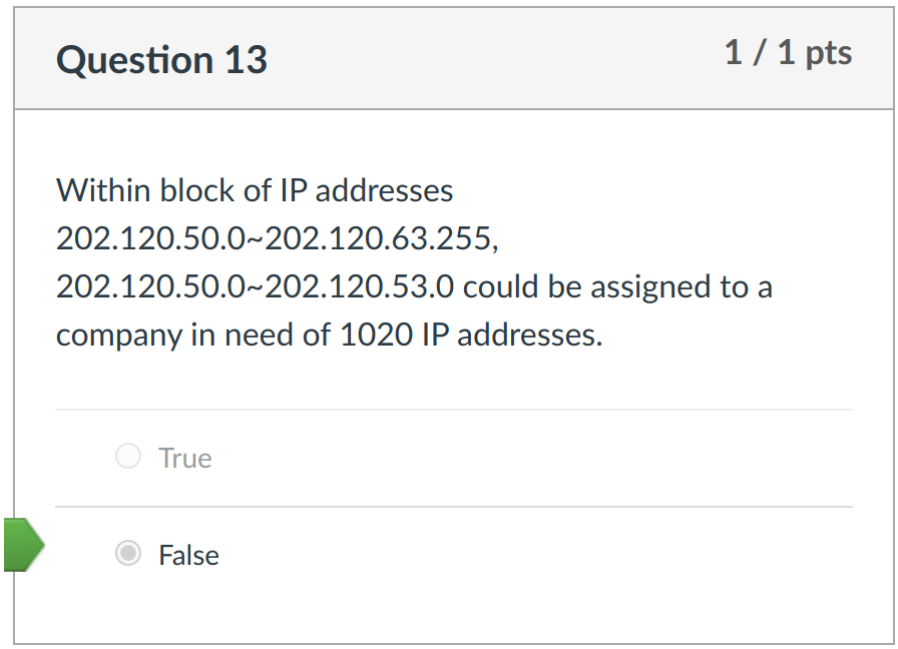
Question 12

1 / 1 pts

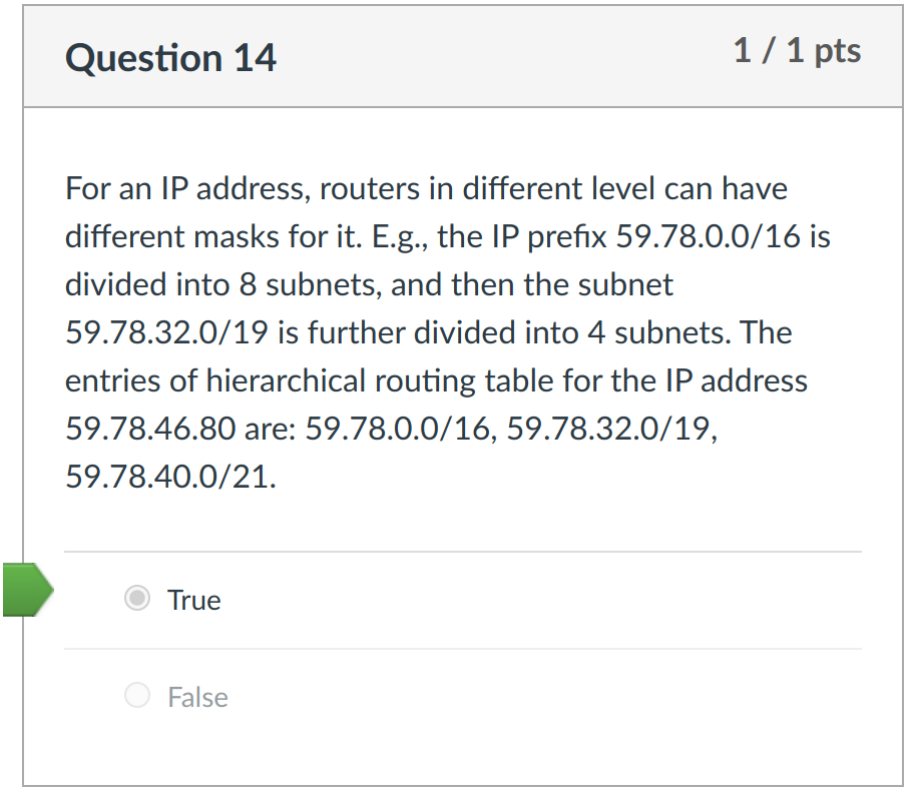
Tracert uses ICMP to get "TTL expired" error-reporting message from each hop to extract their IP addresses and evaluate the RTT.

- (●) True
- ( ) False

Explain: That is exactly how `traceroute` does its job.



Explain: 50~53 cannot be aggregated into one prefix, 52~55 can. So it is better to allocate 52~55.



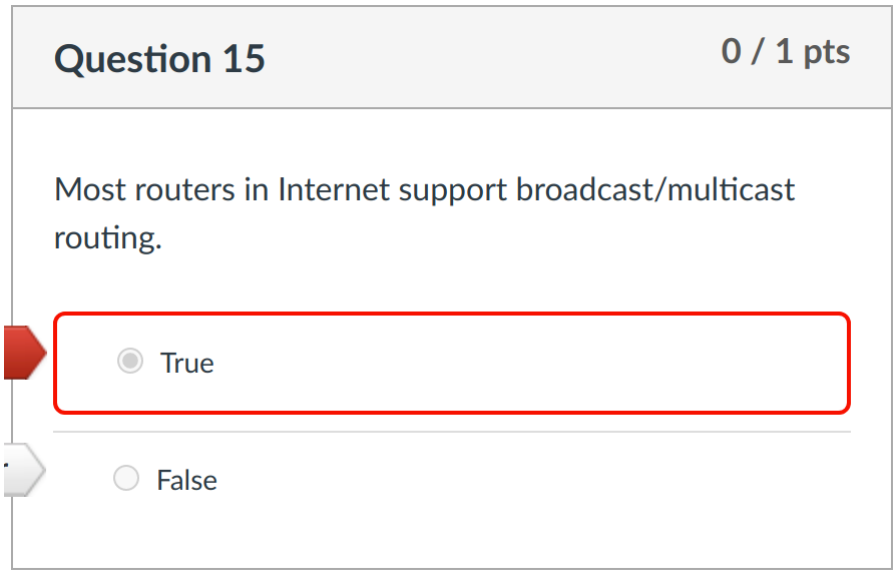
Explain: First division:

`59.78.0.0 ~ 59.78.32.0 ~ ... ~ 59.78.224.0`

Second division:

`59.78.32.0 ~ 59.78.40.0 ~ 59.78.48.0 ~ 59.78.56.0`

Thus, the `IP` address for `59.78.46.80`, goes to second entry for first division, second entry for second division. This gives the result.



Explain: Most of the routers use IPv4, which does not support broadcast/multicast.