

Quiz 3

You can ask the course staff any *specific* and *factual* questions you like *during* the quiz, and we will answer them! You are responsible only for reasoning, not knowing facts. Examples of factual questions include details of how an algorithm works, any formulas, and definitions of terms.

When answering essay questions, keep your answers as short as possible while covering all points.

1 Multiple Answer 1 point

A Low Earth Orbit (LEO) satellite internet company has noted that LEO networks offer lower latency between two points than terrestrial networks, but at a much lower bandwidth. Low latency is possible because radio signals (and light) travels faster in air than in glass. But radio communication is lower throughput than light-based fiber optic communication.

They argue that centralized deployments are easier to manage than CDN/edge deployments, and LEO networks will allow comparable latencies despite the centralization. Given below are some applications which traditionally use CDNs or edge computing. Which of these can benefit from this newer LEO-based model? You may assume both the central deployments and the end users are directly connected to the LEO network.

- ☐ Large software downloads
- ☐ Video-on-demand streaming services.
- ☐ AR services where the server renders the frames and streams them to the user. Latency matters here.
- ☐ High frequency trading applications that send a small amount of latency-critical data between cities.
- ☐ Multi-player games that only send key-strokes and mouse-movements between users, but latency matters.

Bob has started a server that listens for UDP packets on port 3555. It responds to every packet with the current timestamp. To do so, it has a high-precision atomic clock connected as a peripheral. The CPU can query it to get the correct current time.

He has also designed a client that people can use to talk to this server and synchronize their own clocks. The client notes the time, t_A , sends a request and waits for the response. It notes the time at which it received the response packet as t_B and the timestamp the server wrote in the packet as t_R . Note: t_A and t_B are noted using its own clock. t_R is from the server's reference clock.

As soon as it receives the packet, it resets its own clock's time as $t_R + (t_B - t_A) / 2$.

What assumption is Bob making when he does this?

Bob notices that the client's clock does not perfectly synchronize with the server's reference. What could be the reasons? Errors could come from many parts of the system (including the network and the two end-points). Name at least three different parts/layers that can cause the problem

Bonus: For an extra 0.5 point, name how he might determine which of the three causes you named is causing the most error.

 100%

1 min / 1000 max (word limit) Must use at least 1 word

Video conferencing

A hyperscaler X (i.e. a company like Google, Microsoft or Amazon with massive compute and network infrastructure) has a video conferencing application that lets two or more users talk in real-time over video and/or audio.

Help them make design choices

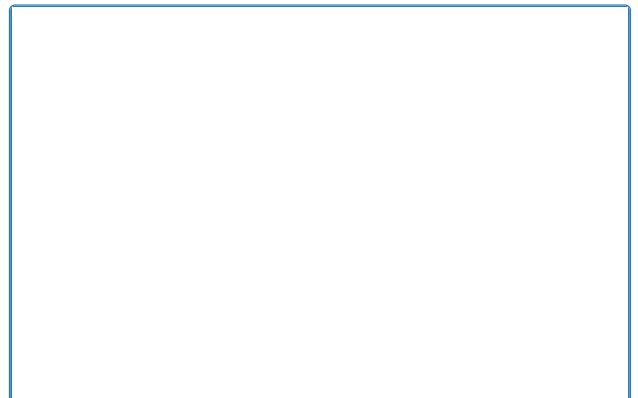
They notice that in some cases, their users experience very poor network quality on the public internet while X's internal networks have spare capacity available. They would like to let their users use their internal network. This way packets enter X's network as quickly as possible, and leave it as late as possible.

Suppose they have a method called PathMagic that can decide whether the default route is better or X's internal route is better. Two methods have been proposed to make packets flow on the desired path:

- 1) Sending BGP advertisements that modify how the users' ASes route packets according to PathMagic
- 2) If PathMagic decides to use the internal network, clients use an overlay network where users first send packets to IP addresses of servers within X's network, which then forward packets to the other users

After some debate the team decided that option 2 is clearly better even though it requires X to allocate server capacity. Why did they make this decision?

You can assume PathMagic works equally well in both cases.



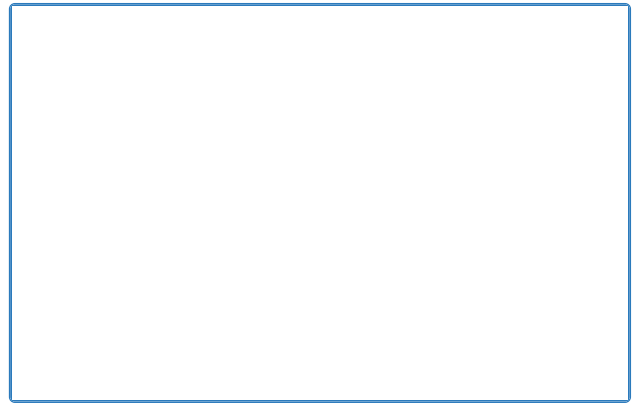
1 min / 500 max (word limit)

Must use at least 1 word

4

Essay 1 point

In a two-way video call, suppose the users are behind a NAT, is there a condition when direct peer-to-peer connection establishment (i.e. using the default route between the users) is practically impossible? If yes, give an example. If no, explain why not



1 min / 500 max (word limit)

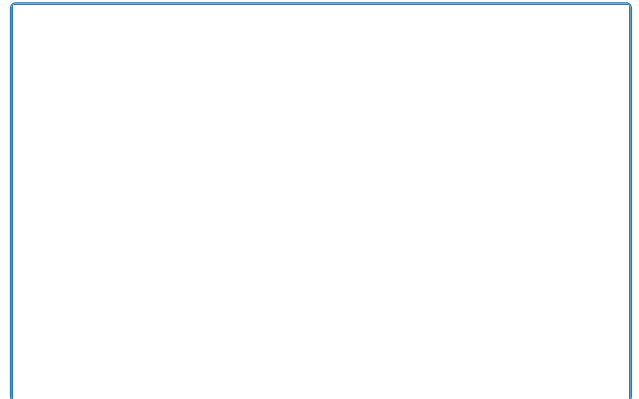
Must use at least 1 word

5

Essay 1 point

They tested a direct peer-to-peer video call over a dedicated 100 Mbit/s link (i.e. the users' send video/audio data directly to each other over a TCP connection). No other traffic is present, the available bandwidth is far more than the call requires, and when they run a ping command, the RTT between the users is only 10ms. However, they still experience a large end-to-end delay between when one person speaks and the other hears it. The delay remains consistently high over the entire call and does not improve with time. What could be causing this persistent latency?

Hint: it might have something to do with their using the AIMD congestion control algorithm.



1 min / 500 max (word limit)

Must use at least 1 word

Social media

We are designing a social media application as a web page. Consider the main feed page. It has the following components:

1) Visual page elements that are present in everyone's page: the logo, basic icons, and the overall page layout

2) A javascript program that makes subsequent HTTP requests to load posts.

3) The posts themselves, some of which may contain images and/or videos

4) Direct messages between users

6

Ordering 1 point

If the goal is to be the fastest to show the *entire* web page as quickly as possible, in which order should we make the browser load the objects?

Notes:

1) By "entire page" we mean, displaying the first 10 posts. Additional posts are loaded only if the user scrolls.

2) The browser can send multiple HTTP requests back-to-back within the same RTT. The server will respond with the objects in the order in which the requests were sent.

3) There is only one TCP connection

4) There is no CDN; All objects are served from the same server.

5) None of the objects are cached in the browser.

☐

⋮ The javascript program mentioned on the left

☐

⋮ The objects to display the posts

☐

⋮ The HTML file

☐

⋮ The logo, basic icons and other small objects needed for the overall page layout

7

Ordering 1 point

Rank the following objects by ease of serving from a CDN.

⋮ Direct messages between two people

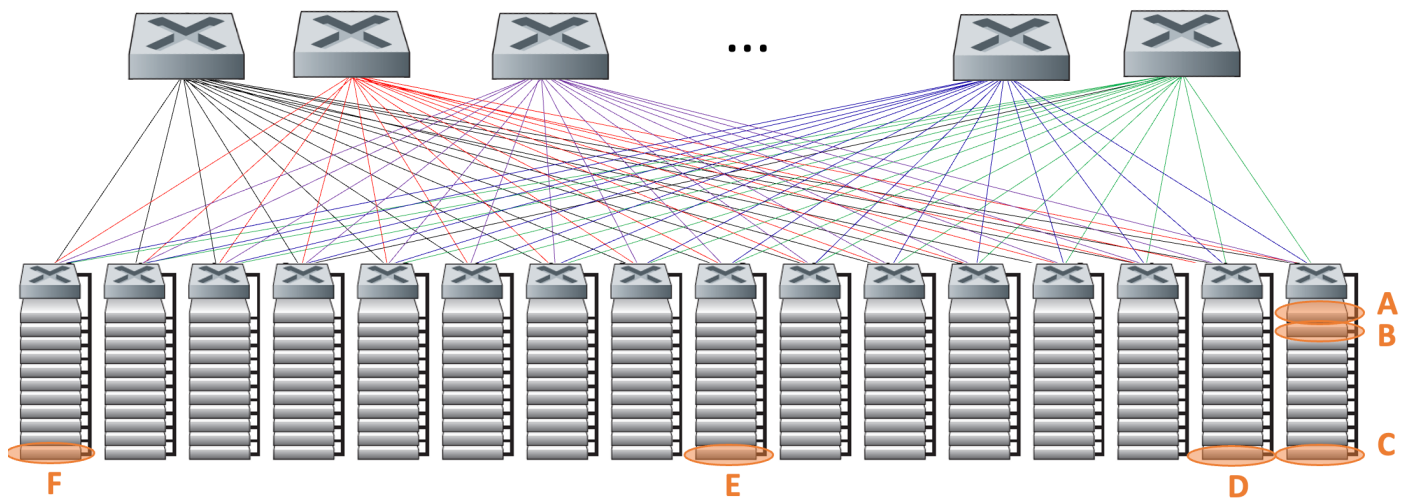
⋮ Videos and images for the viral social media posts

⋮ Logo and icons

⋮ Videos and images for social media posts viewed only by 100-1000 people

8

Categorization 1 point



[Credit: Image adapted from Kurose and Ross' slides]

Shown above is a datacenter topology. Rank servers **B**, **C**, **D**, **E**, and **F** by their distance from **A**. Servers that you expect to have lower latency and higher throughput to A should be considered *closer*.

To rank them in Canvas, assign each server to a **distance category (1–5)**, with 1 being closest. Servers at the same distance should be placed in the same category. As a result, some of the higher-numbered categories may remain empty. If all servers are at different distances, each category will contain exactly one server.

Distance 1

Distance 2

Distance 3

Distance 4

Distance 5

Possible answers

⋮ B

⋮ C

⋮ E

⋮ F

⋮ D

9 Essay 0 points

Spare question for including any nuances/explanations you like. Most students will not need to use this field

10 Essay 1 point



Alice has built a new wireless physical layer link. She has integrated it with IP and when she sends individual UDP+IP packets, they are received as expected. She increases the send rate and the link is able to support 1000 full-sized (1.5 KByte) IP packets per second. 5% of the packets are lost due to physical-layer reasons (she does not want to change this).

Excited, she tries to download a large file over this link over HTTP. However, the throughput is extremely low. Much lower than 950 packets/second you'd expect due to a 5% loss rate. She suspects there is some cross-layer phenomenon to blame.

What do you think is going on? How can she fix it?

1 min / 500 max (word limit) Must use at least 1 word

