15440/15640 Distributed Systems Homework 4

Q1. 3-Tier Web Scaling (20 Points)

Han and Yuvraj have built out a 3-tier web application, consisting of 3 layers: (Layer-1) a frontend, (Layer-2) an application server, and (Layer-3) a backend consisting of a SQL database and a DFS implementing AFS. Their application is a video-hosting site called TooYoob that is beginning to pick up a lot of traffic, which has slowed down their application drastically. Their frontend handles client requests directly and their application server computes video recommendations based on an internal algorithm. The application server handles nearly all algorithm computation in-memory for efficiency. The SQL database holds data about user accounts and video statistics, while the DFS contains the videos themselves. They're having a difficult time on some system design choices.

Q1.1 (5 Points)

Which layer should they allocate the most RAM to, and why (state any assumptions you make)?

Q1.2 (5 Points)

Currently, their application has one server running the front-end, and they are looking to scale out, however this is rather costly. They are trying to minimize the number of servers to scale out while improving the load on the frontend. A single frontend server can process 100 jobs per second for every 300 jobs arriving per second. Assuming the arrival rate stays constant and jobs are assigned equally by the load balancer, how many more servers should be added to bring the average load to 0.5 or below?

Q1.3 (5 Points)

After scaling out the front-end, Han and Yuvraj realize that their application is still running slow, likely due to slow database processing. While maintaining strict consistency on the database contents, should they scale up or scale out this component? Explain your answer.

Q1.4 (5 Points)

Give an alternative method to fix their database issue if they are unable to scale up or scale out the database itself, and explain how it would improve processing.

Q2. Diffie-Hellman Key Exchange (22 Points)

Will and Jack want to discuss top secret info through email. They decide to encrypt their emails and use the Diffie-Hellman key exchange protocol to first agree on a secret key.

Q2.1 (8 Points)

Suppose they have agreed on g = 7 and p = 23 (g and p are public). Will picks his secret number 9 and Jack picks his secret number 4. What should Will send to Jack? What should Jack send to Will? And what secret key do they agree on? We are using small numbers for ease of calculation, since in reality these numbers would be quite large.

Q2.2 (8 Points)

Assuming they use cryptographically secure primes, why can the top secret info still be stolen by an attacker? Give an example of an attack that can steal the top secret info.

Q2.3 (6 Points)

What is the fundamental problem of this protocol (i.e. what security property is missing)? What different protocol or variation in this protocol can protect the top secret info?

Q3. MAC (12 Points)

Yuvraj and Rashmi need to communicate to determine which TA is going to grade the next homework. They have a shared secret key, K_{Profs}, that allows them to create unforgettable message authentication codes (MAC) so that Rashmi can verify that Yuvraj did in fact create any message that is received. Rashmi and Yuvraj have a simple protocol: Rashmi sends a “Who grades HWX?” message to Yuvraj in plain text, and Yuvraj replies with one of two messages: M1 = MAC K_{Profs} (“Emma”), or M2 = MAC K_{Profs} (“Eunice”). When Rashmi receives either M1 or M2, she verifies the MAC using K_{Profs} and knows who will grade the next homework.

Q3.1 (6 Points)

This protocol is insecure. A malicious TA on a router between Rashmi and Yuvraj might be able to avoid ever having to grade a homework! In one sentence, describe the attack.

Q3.2 (6 Points)

What simple change to the above protocol could defend against this attack?

Q4. Virtualization (20 Points)

One day, after receiving a bill from AWS, Ethel decides that she is going to start her own cloud computing company. She wants to give her users the most freedom possible in their requests and wants to make sure that they always get the best possible service. Please help Ethel decide how she should deal with specific requests (possible answers: containers, VM, and Non virtual machines).

Q4.1 (5 Points)

A 15-410 student sends their kernel project, which assumes it has control of the memory management, to Ethel and asks her to test it out. What kind of virtualization would she provide and why?

Q4.2 (5 Points)

James McShadyguy mails Ethel an envelope that says “Run this in a container on the same machine as someone doing important bank information”. Ethel does not trust Mr. McShadyguy and wants to first test the new updated code to ensure it is well behaved since it is from a third party. What kind of virtualization would she provide and why?

Q4.3 (5 Points)

Vang Ogh, a good friend of Ethel's, is hoping to buy NFTs. Unfortunately, Vang keeps getting scammed when the sites turn out to be fake! As a result he wrote a program that takes the URL of an NFT site and determines whether the site is a scam. He wants to be able to send Ethel URLs and quickly hear a result. What kind of virtualization would she provide and why?

Q4.4 (5 Points)

Ethel decides early on to only have type 2 VMs. What types of requests would warrant her investing in type 1 VMs?