15-440/15-640: Homework 4

Due: December 6, 2016 10:30am

Name:
Andrew ID:

1 RAID (16 points)

Raj has designed a new database system which he wants run on a RAID setup.

1. Which RAID configuration (from among the basic 5) should Raj use if he wants to optimize the database system for workloads consisting of small random reads and writes and why? **[4 points]**

> RAID 5 is the correct configuration to use since it optimized for small le writes. Due to rotated parity, we can utilize multiple disks in parallel and remove the bottleneck on the single parity disk and small writes will result in keeping all disks evenly busy.
> RAID 0 is also accepted as question does not ask for reliability.

2. Using a 10 disk RAID setup and each disk is capable of average throughput of 100 MB / sec (read and write) and an average latency of 10 ms (again, for both, reads and writes), what is the latency and throughput of random reads and writes of Raj's RAID scheme? **[8 points]**

> A 10 disk RAID setup implied 9 data disks and 1 parity disk per stripe.
> For a RAID 5 configuration (as in part 1),
> Random read throughput = 10 * throughput of 1 disk = 10 * 100 = 1 GB / sec,
> Random write throughput = 10 / 4 * throughput of 1 disk = 2.5 * 100 = 250 MB / sec, since 4 I/Os have to be performed per write.
> Read latency = latency of reading 1 disk = 10 ms,
> Write latency = latency of reading data disk and parity disk followed by writing data disk and parity disk = 20 ms.

3. What is the mean time to data loss of the setup in part 2 if the mean time to failure of a disk is 100 years and there are 10000 disk arrays of 10 disks each (assume no rebuild)? **[4 points]**

> $MTTDL_{system} = \frac{MTBF_{set}}{\text{number of sets}}$
> $MTBF_{set} = \frac{MTBF_{disk}}{10} + \frac{MTBF_{disk}}{9}$
> Final answer = $\frac{\frac{100}{10} + \frac{100}{9}}{10000} = 0.00211 years = 18.5 hrs$

2 Virtual Machines (18 Points)

1. In no more than 5 sentences, describe the difference between Type I (Hypervisor) and Type II (Hosted) VMMs. **[4 points]**

Type I VMMs run directly on the system hardware and are also known as bare- metal hypervisors. They provide higher performance, availability, and security than Type II VMMs. eg: Citrix XenServer Type II VMMs run on a host operating system such as windows, Linux etc. It uses a guest operating system which runs on top of the hypervisor. eg: VMWare workstation.

2. In virtualization, can the hypervisor (or VMM) allocate and assign more than the actual physical resources it has available at its disposal (memory, processors) to individuals actual Virtual Machines (VMs)? Please explain. **[4 points]**

YES. Virtualization allows you to overcommit resources and provide more virtual resources than actual physical resources, for example by time multiplexing the same set of CPU/processors across multiple VMs or using memory ballooning to use less or more memory. (either CPU or memory example is fine)

3. When does inflation of memory ballooning happen? And what problem might it cause? **[4 points]**

It happens when the hypervisor is running out of physical memory and needs to reclaim memory from guest OS. Although ballooning gets hypervisor more memory, the targeted guest OS may suffer from performance decay caused by reduced memory space and subsequently more paging out.

4. You are a data center engineer. Explain two (2) advantages and one (1) disadvantage of using system virtualization in the data center to your manager. **[6 points]**

Advantages:
Portability: can move VMs around, consolidate them, save energy
Isolation: Fault isolation, so if one VM fails the system does not come down; Performance isolation, one VM cannot easily hog the resources of other VMs
Encapsulation: All state can be captured cleanly, hence allows snapshots, cloning, etc.

Disadvantages:
Performance: Inherently, since Virtualization adds another layer of indirection there will always be some extra overhead.
Higher Risk of Hardware Faults: If there are multiple VMs on the same physical machine and the actual hardware has a fault, (hard drive crash, network cable unplugged) it affects all of the VMs at once.
More complicated: Since VMs can move around you now have to worry about where a particular VM is located, so adds some extra management overhead.

3 Byzantine Fault Tolerance (16 Points)

In this question, we'll explore the links between replication for fail-stop failure resilience and replication for byzantine failure resilience. Recall that Paxos—an algorithm for fault tolerant replication under non-byzantine failures—requires $2f+1$ replicas to handle f failures. BFT, on the other hand, requires $3f+1$.

1. Why is it sufficient in Paxos to use a majority vote among $2f + 1$ nodes to ensure consistency? (In other words, what property of a majority are we relying on?). **[4 points]**

> The key property is that any two majorities intersect in at least one node. Thus, any decision reached by one majority *must* be observed by any other majority, ensuring that the value chosen will propagate even if the other nodes didn't see it.

2. Prove by providing a contradicting example that normal Paxos using 3 replicas cannot handle a byzantine fault in an asynchronous network. Use three replicas A, B, and C. For a proof, we want you to sketch a series of communication in which two clients observe an inconsistent result (which violates the requirement of a consistent answer from the system) when there's a single byzantine ("evil") node among the replica set. **[8 points]**

> This answer comes from lecture. The key observation for this answer is that the byzantine node can make different statements to different observers. Let B be the evil node:
>
> ```
> C1 --value 1-->  {A, B},  pkt lost to C
>                  {A, B} decide to accept value 1.
>
> C2 --value 2-->  {B, C},  pkt lost to A
>                  B pretends to have never seen a value before
>                  thus, {B, C} decide to accept value 2
> ```
>
> Any client asking A, B for the value will hear 1, but any client asking B, C for the value will hear 2. Both clients will think they got the right answer because an f+1 majority agreed on the answer they observe.

3. Why can BFT succeed with using $3f + 1$ replicas and requiring a consistent answer from $2f + 1$ of the nodes? (2 sentences) **[4 points]**

> In 3f+1 nodes, every majority of $2f + 1$ nodes must contain a majority of "good" nodes (f+1, since there can only be f bad nodes). Thus, the good guys overrule the bad guys.

4 Security Protocols (18 points)

Yuraj and Srini wants to discuss the final exam questions through email. They decided to encrypt their emails to avoid being attacked by genius students like you. The first thing they have to do is to agree on a secret key. A TA suggests that they can use Diffie-Hellman key exchange protocol.

1. Suppose they have agreed on g = 23 and p = 5 (g and p are public). Now Yuraj picks his secret number 6 and Srini picks his secret number 8. What should Yuraj sends to Srini? What should Srini sends to Yuraj? And what is secret key they agree on? **[6 points]**

> Yuraj sends to Srini: g^a mod p = 23^6 mod 5 = 4
> Srini sends to Yuraj: g^b mod p = 23^8 mod 5 = 1
> The secret key they agree on: 1^6 mod 5 (4^8 mod 5) = 1

2. Assuming they use cryptographically secure primes, why can the final exam questions still be stolen by an attacker? Give an example of an attack using the parameters from part 1. **[8 points]**

> Use Man-In-The-Middle (MITM) attack. Suppose I can intercept the network traffic. When I see A = g^a mod p = 4 from Yuraj to Srini, instead of let it go through, I can choose my secret number c = 7 and send g^{ac} mod p = A^c mod p = 4^7 mod 5 = 4. Similarly, when I see B = g^b mod p = 1 from Srini to Yuraj, I will replace it with g^{bc} mod p = B^c mod p = 1^7 mod 5 = 1. Now I agree with both of them on a secret key g^{abc} mod p = 1. I can decrypt whatever they send and they still think they are talking to each other! (Actually if you swap p and g you should get different secret keys. We intended to use g = 5 and p = 23 but made a mistake in the question.)

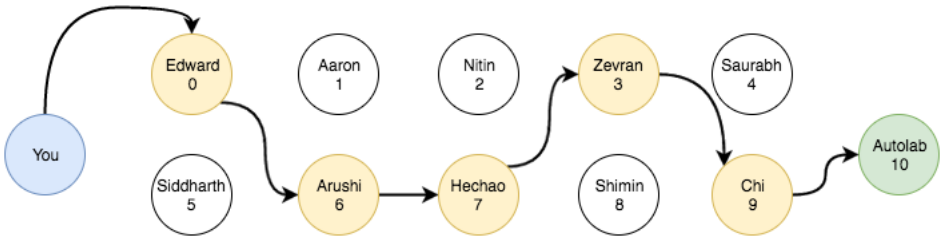
3. What is the fundamental problem of this protocol (i.e. what security property is missing)? What different protocol or variation in this protocol can protect the final exam?? **[4 points]**

> The problem is that DH protocol itself does not provide authentication. One solution is to use authentication protocol along with this key exchange protocol. Specifically, both Yuraj and Srini choose asymmetric key pairs usable for digital signatures. Yuraj signs whatever he sends to Srini, and Srini verifies that signature (using Yuraj's public key) after receiving the email and vice versa. Any email with invalid signature should be discard. Another solution is to use public key cryptography instead so that they do not need the shared secret key.

5 Anonymous Routing (12 points)

Recall that Tor, or "The Onion Router" (`https://www.torproject.org/`) is a decentralized system that allows people to anonymously browse the Internet. After learning about that, 440 TAs decided to set up a Tor network for students to submit homework on autolab. Each TA has an ID with him / her which works as the IP.

1. In the Tor circuit below, what is the packet that you send out? What is the packet that Hechao sees? Suppose that each packet is in the form [Next hop TA ID, data], M is your original message, K_i is the public key of TA i and E(K_i, m) denotes *encrypted data m using key* K_i. **[4 points]**

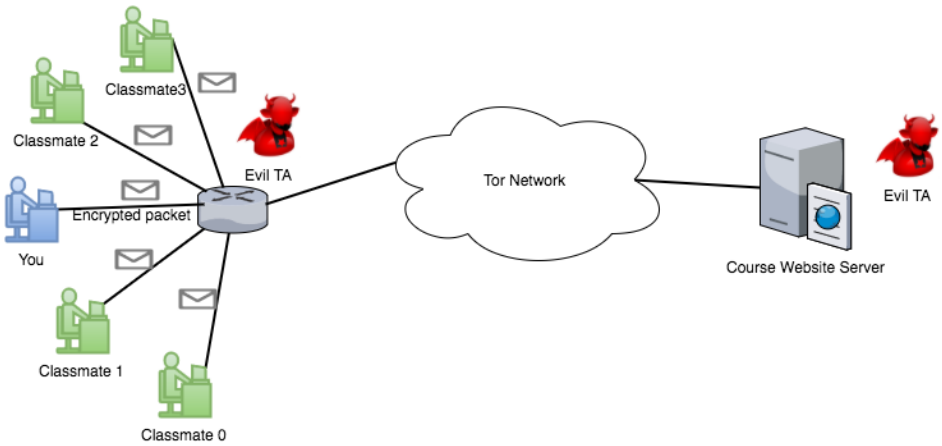


> Packet sent out: $[0, E(K_0, [6, E(K_6, [7, E(K_7, [3, E(K_3, [9, E(K_9, [10, M])])])])])]$
> Packet Hechao sees: $[7, E(K_7, [3, E(K_3, [9, E(K_9, [10, M])])])]$

2. Suppose you use HTTP protocol in which your data is not encrypted. In the Tor circuit above, who may see your user name and password and why? **[4 points]**

Chi. Because the last hop of Tor is unencrypted.

3. As you may know, Tor does not provide perfect anonymity. It is vulnerable to an attack called "traffic analysis". Suppose you are asked to submit an anonymous feedback through the course website. You decide to use the Tor network to access the page. However, there is an evil TA who controls both the router right before the Tor entry point and the website server. Describe how can the TA distinguish you from your classmates.[**4 points**]



On the server side, once receiving a feedback, inject a traffic pattern into the TCP connection where it received the feedback. Then obtain the flow data from the entry router. Compute the correlation coefficient between the server to exit Tor node traffic and entry router to students traffic.