

Name: _____

Instructions

There are three (3) questions on the exam. You may find questions that could have several answers and require an explanation or a justification. As we've said, many answers in storage systems are "It depends!". In these cases, we are more interested in your justification, so make sure you're clear. Good luck!

Problem 1 : Short answer. [48 points]

- (a) In addition to logging writes to NVRAM, modern disk arrays use a variety of other integrity mechanisms. Explain why each of the three mechanisms below is used (1 sentence for each explanation);
- ECC on memory and bus

 - Scrubbing data blocks on the disk

 - Checksum on each data block
- (b) Many disk arrays compute parity across disks to protect against disk failures. What makes this approach more difficult in multi-server distributed file systems seeking to protect against file server failure?

- (c) Your friendly TA claims to have developed a new backup algorithm that backs up your system during idle time (desktop systems have plenty of idle time, so you realize that the idea is not bad). The algorithm is very simple: each time the computer is idle, the algorithm copies raw blocks from your disk to a secondary backup disk (ala physical backup). What is one serious problem with this seemingly simple algorithm?
- (d) Raja decided to sell his Audi car to make room for a large storage system he wants to build inside his garage. He's hoping to make big money by using it to store high-demand movies and distribute them to his neighbors for a small fee. He is very picky about designing a robust system and wants to get every little detail right. In particular, he's considering whether to buy 100 servers from a single vendor (e.g., Dell), or from 10 different vendors. Give one argument in favor of each option.
- (e) Many organizations that store highly-classified data require an elaborate process for disposing of old disks. The process sometimes involves burning the disks or overwriting them hundreds of times. Give one reason why such companies might not choose to simply encrypt the data and throw away the key when they want to get rid of the disks?

- (f) Give two reasons why callbacks are not used for client cache consistency in the World Wide Web.
- (g) The NFS file system was originally conceived to allow clients to use a file server in a completely transparent manner – that is, users and application programs should not notice that a remote server is involved at all. Briefly identify two ways in which it fails to do so.
- (h) Imagine a sensor monitoring program that opens a file and periodically updates it with the latest values from the sensors. If the file is stored in AFS, why would someone reading that file every hour never see any changes?

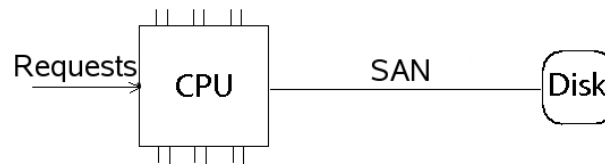


Figure 1: A simple file server.

Problem 2 : System analysis. [28 points]

- (a) Company MiniMiniPC builds simple file servers as shown in Figure 1. Assume that each request needs some CPU processing and some I/O from the disk. Measurements have shown that a request in this system needs, on average, 10 nanoseconds of CPU time (CPU processing), 10 microseconds of SAN network time (propagation delay), and 10 milliseconds of disk time (reading and writing data blocks).
- Assuming a closed-loop system with zero think-time and no concurrency (i.e., only one client can send requests), what is the maximum sustainable throughput of this system (in requests/sec)
 - What is the average response time for a request under these assumptions?

Hint: Don't try to remember formulas. Use intuition here. However, still be complete in your answer.

(b) The company hires you as a consultant to improve the system. They are considering one of two options, and they need your help in picking the best one:

- Buy 1GB memory and place it in front of the disk (hits incur zero-cost, misses have to go to disk) OR
- Upgrade the disk with one that is twice as fast

(They cannot do both, since they really want to save money)

Do you have enough information to make a sound decision? If so, make the decision and evaluate the new throughput and response time of the system. If not, make some assumptions about the information you don't have, make the decision and evaluate the new throughput and response time.

Show all your work for full credit.

Problem 3 : Designing for disasters. [24 points]

This question will allow you to make some real-life assumptions and compromises on what disaster prevention and recovery technique to use for your data. There is no single good answer here, rather you will need to explain your assumptions.

You are hired as a consultant by three different businesses. The first, is a university that deals with STUDENTS and their data. The second is a BANK that deals with financial information. The third is an ONLINE RETAILER that sells books online. You are to design a storage system for each of these businesses. Armed with the knowledge you learned in this class you feel ready.

- (a) Fill in Table 1. Data loss cost refers to the cost of losing hours of work (e.g., if I prepare this exam and the computer crashes before I save, I just lost 3 hours worth of work) and is measured in dollars per hour. Outage cost refers to the cost of losing access to the data and is measured in dollars/hour.

	Data loss penalty \$/hr	Data outage penalty \$/hr
STUDENT	200	0
BANK	50M	10M
ONLINE RETAILER	500K	5M

Table 1: How much is data worth?

You need to make up these costs, however provide a reasonable explanation for your estimates below.

(b) Next, we look at some techniques to protect against data loss and outages. Fill in Table 2 and Table 3. Hints and assumptions are given in the tables' captions.

Protection technique	Data loss (seconds)Reason for your decision.....
Synchronous remote mirroring	0	Data is sent instantaneously to the remote center, hence there is no data loss
Async remote mirroring with 1 GB buffer	100	In the worst case the whole buffer can be full
Backup every day on local tape	24 * 60 * 60	A whole day's worth of data can be lost

Table 2: Protecting against data loss

Note: Synchronous remote mirroring refers to writing to the local storage system and a remote replica "at the same time", as is done with RAID 1 in disk array systems. In the case of async remote mirroring, data is not immediately sent to a secondary server, but it is first buffered (a write doesn't wait for the secondary to be updated). Assume that 1GB can hold roughly 100 seconds worth of data.

Protection technique	Recovery time (seconds)Reason for your decision.....
Failover to secondary	0	Flip a switch and the secondary takes over
RAID reconstruction at primary	3 hours	Assuming 100MB/s reconstruction bandwidth

Table 3: Protecting against outages

Assume both primary and secondary systems use a RAID-5 system and the outage is in the form of a disk in the RAID array failing. The companies are extra careful and prohibit access until the RAID is fully repaired. Assume the RAID box is holding approximately 1TB of data. Make a reasonable guess as to how long it takes to reconstruct.

(c) Now, make a decision on what protection and outage mechanisms you will use for each of the businesses. Assume that the original purchase costs of the various options are all equal. Your choices should go in Table 4.

	Data loss mechanism	Data outage mechanism
STUDENT	Local backup on tape	RAID reconstruction
BANK	Sync mirroring	Failover
ONLINE RETAILER	Async mirroring	Failover

Table 4: Making the final decision

Problem 4 : Extra credit. [up to 2 bonus points]

(a) What is Greg's favorite football team?

(b) Which TA is taller?

(c) What is Greg's standard word for "data loss" and other disastrous circumstances, this term?

(d) What was the most interesting thing you learned this term (in this class)?

This page intentionally left blank in case you need scratch space.