CMU 18–746 **Storage Systems** **30 Nov 2006**

Fall 2006 **Exam 2**

Name: _____

Instructions

There are three 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. [49 points]

(a) Many in the storage industry view it as simply a matter of time before today's common storage buses (e.g., Serial ATA and parallel SCSI) and networks (e.g., FibreChannel) are replaced with Ethernet. Why are file system designers not concerned about such a change?

(b) Consider two 8-disk disk arrays using a RAID5 data distribution: one with 50GB disks and one with 500GB disks. Which one is more likely to suffer data loss? Explain. (Assume that the disks in both arrays have the same MTBF and performance characteristics, and that all other aspects of the hardware, software, and administration are identical.)

(c) In experimenting with a disk array that uses parity-based redundancy, you observe interesting performance results as you consider different block sizes for your database system. You know that your workload is almost entirely random writes of whatever block size you select; you also know that there is no concurrency (one request at a time from the database system to the disk array controller). In looking at per-request latency, from the database system's perspective, you observe: 20ms for a block size of 8KB, 20ms for 16KB, 20ms for 32KB, 10ms for 64KB, 20ms for 96KB, 10ms for 128KB. Explain why this is happening.

(d) Stratos wrote data to a file, closed it, and then sent email to tell Greg that it was there. Greg opened the file right away, but didn't see the new data and complained. Stratos patiently asked Greg to try again in a minute, and Greg saw the data when he did. What distributed file system are they using? Explain.

(e) Many enterprise environments use both hourly snapshots (maintained by each filer for the data that it stores) and nightly tape backups. Why do both?

(f) Block-based storage and file-based storage often get compared in terms of performance by direct measurement. Why does block-based storage often beat file-based storage, in simple data bandwidth measurements, when the block devices are "SCSI over FibreChannel" and the file servers are "NFS/TCP/IP over Ethernet"? (Assume that the two interconnects have the same bandwidth.)

(g) We know that deleting data from a file system does not actually remove all traces of its contents from the disks that stored it. Many have proposed using encryption to solve this problem. Identify two practical problems with doing so.

Problem 2 : Disk Striping. [14 points]

You are given a disk array with 4 disks. Assume that each disk has an average 10ms positioning time (including seek and rotational latency) and a transfer rate of 80MB/s. The array implements simple disk striping, but there is no redundancy.

(a) Assume a workload consisting of random 1MB requests. Compute the average request service time and the maximum array throughput for each of these stripe unit size options:

- 256KB.

- 1MB.

(b) What workload information, beyond that given above, would you need in order to make a good decision between these stripe unit size options? Explain.

Problem 3 : Decentralized file system design. [37 points]

(a) Consider a decentralized file system that has a single centralized file/metadata manager and a collection of data storage servers. Each client request in our hypothetical system involves one 1ms access to the metadata server (to acquire data location information) and one 20ms access to one of the data servers.

(a) What is the largest number of data servers that can be in the system before the metadata server will be a throughput bottleneck?

(b) With 10 data servers, what is the maximum throughput?

(c) Assume 10 data servers and 50 clients, each of which performs one request at a time and thinks for 5ms between the completion of one request and the generation of the next. At the maximum throughput from (b), what is the expected response time for a request?

(b) Consider a decentralized file system architecture as in 3(a) above again. The designers would like to reduce the metadata server bottleneck, allowing larger collections of data servers. To do so, they propose having clients cache data location information. Identify a problem that this introduces and one solution to solving it.

(c) Consider a decentralized file system architecture as in 3(a) above again. Assume now that data protection is provided by having a single parity server that stores parity for blocks that are striped across the data servers. Describe one approach to protecting the integrity of the parity from client crashes.

(d) Consider a decentralized file system architecture as in 3(a) above again. Assume that the data servers know nothing about file structure; they simply support reads and writes to fixed-size blocks (much like disks). Briefly describe an approach that would allow the metadata server to control which reads and writes the data servers will execute. (Any clients can send any access to any data server, but we want a way for the data servers to reject anything other than accesses deemed ”ok” by the metadata server.)

Problem 4 : Bonus Questions. [Maximum of three points points]

(a) Who won The Game (Michigan vs. Ohio State)?

(b) Was Greg happy about it?

(c) What year will Stratos graduate?

(d) If we get Timmy a PlayStation for the holidays, what game should we include with it?

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

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