CMU 18–746 **Storage Systems** **24 Feb 2010**

Spring 2010 **Exam 1**

Name: ______________________________

Instructions

There are four (4) 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!

If you have several calculations leading to a single answer, please place a $\boxed{\text{box around your answer}}$.

Problem 1 : Short answer. [48 points]

(a) Ext-2 does not allow creation of hard links to directories. What is one problem that is avoided by disallowing such hard links?

(b) Arif is using a file system with write-ahead logging enabled. The system crashes immediately after he appends a megabyte of data to a file. After the system is rebooted and the log replayed, Arif notices that the file size is 1 megabyte larger, as he hoped, but that not all of the data he wrote during the append is in the file. What is the most likely cause of this problem?

(c) A disk has writes outstanding to the following cylinders: 2, 500, 7, 5, 1000, 3 and has just finished writing data to cylinder 600 and read data from cylinder 578 before that. List the order in which these writes (identified by cylinder number) will be serviced by each the following disk scheduling algorithms.

- FCFS:
- SSTF:
- SCAN (LOOK):
- C-SCAN (C-LOOK):

(d) When using a POSIX-based filesystem, how can an application quickly read byte 123,456?

(e) Suppose a disk can service 100 I/Os per second. Given a workload that issues requests at an exponentially distributed rate w/mean 20 I/Os per second, what is the utilization of the disk? What is the average time that a request spends in the disk queue?

(f) In a large datacenter with 3000 machines, 100 failures are observed over a period of 2000 days. What is the resulting MTBF?

Problem 2 : FSCK. [28 points]

Greg decides that he is going to implement his own version of myfsck for ext-2 during his spare time. Help Greg finish his implementation.

For this problem, recall the functionality of the four passes of myfsck:

- **Verify and fix directory pointers**: Verify for each directory: that the first directory entry is "." and self-references, and that the second directory entry is ".." and references its parent inode. If an error is found, correct the entry.
- **Insert unreferenced inodes into lost+found**: Check to make sure all allocated inodes are referenced in a directory entry somewhere. If an unreferenced inode is found, place it in the /lost+found directory.
- **Verify and fix inode link counts**: Count the number of directory entries that point to each inode (e.g., the number of hard links) and compare that to the inode link counter. If a discrepancy is found, update the inode link counter.
- **Verify the block allocation bitmap**: Walk the directory tree and verify that the block bitmap is correct. If a block that should (or should not) be marked in the bitmap is found, correct the bitmap.

(a) Greg first attempts to implement all of the passes of myfsck within a single depth-first directory traversal. Give one reason why this will not work.

(b) List the steps necessary to add an inode to lost+found, while keeping the filesystem consistent. Note the steps you list may include functionality from other passes of myfsck.

(c) Greg decides to implement pass 2 of myfsck by using the following algorithm:

```
for (i = 1; i < num_total_inodes; i++) {
  if (inode is allocated) {
    traverse directory tree to see if it is referenced
    if (!referenced \&\& inode's link count > 0) {
      add inode to lost+found
    }
  }
}
```

This algorithm may insert more items in lost+found than strictly necessary. Explain a scenario in which this could happen.

(d) Greg wants to make sure he understands link counts properly. He analyzes a directory (not the root directory) with 10 files and 2 subdirectories. What should its link count be?

Problem 3 : More short answer. [28 points]

(a) Figure 1 shows the results obtained by running a variant of the skippy algorithm. Label it with the rotational latency and head/cylinder switches. How would the graph change if the rotational speed of the disk were increase?

(b) A raw device interface, which bypasses the filesystem cache, was used when running the Skippy experiments to obtain the graphs given in the homework and in the exam. Why was this necessary?

(c) Joe has 10 disks available for his disk array, but needs help deciding whether to use RAID-5 or mirroring. Some of his considerations include: (1) no concerns about capacity (he doesn't expect to need even half of the total space), (2) a customer demand for tolerating at least one disk failure, (3) a workload consisting of reading and writing large files sequentially in their entirety. Make a recommendation and justify it.

Problem 4 : Olympic-themed trivia. [up to 2 bonus points]

(a) List one paper you have read so far for which Greg is an author.

(b) What is Greg's term for "poor performance" this semester (although, it never seems to change...)?

(c) Where are the Winter Olympics being held this year? Do you think the Olympics should have been moved to Pittsburgh, given all the snowfall here over the past few weeks?

(d) Was Garth happy or sad about the outcome of the U.S.A. vs. Canada ice hockey game? Why?

(e) Pick one instructor or TA and identify the winter olympic sport for which you think he or she would have the best chance of winning a gold medal. Why?

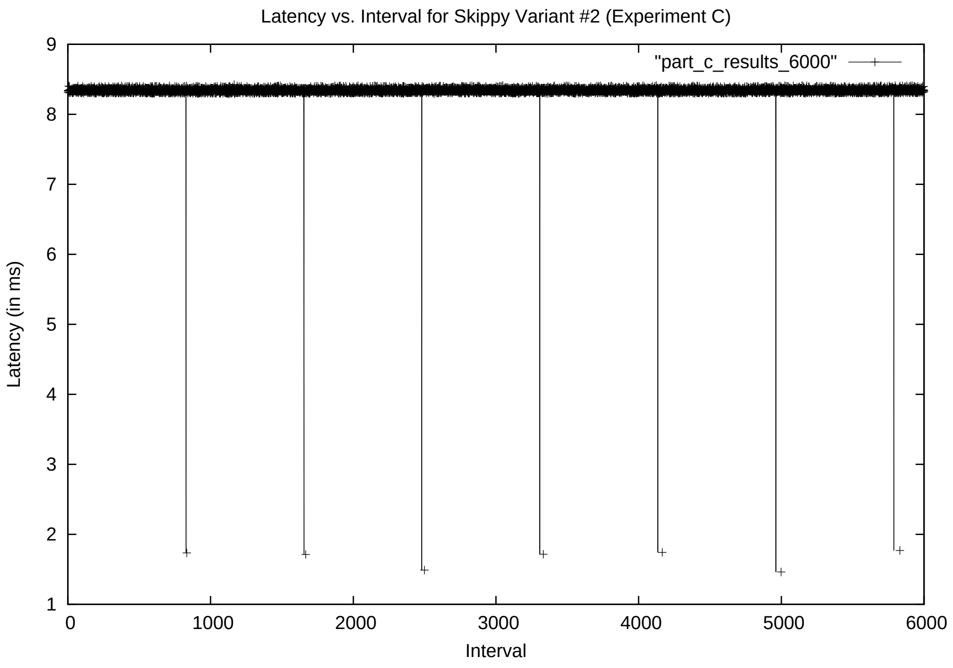

Figure 1: *Skippy results*