Foundations of Information Systems

Winter Semester 2025–26, Exercise 8

For discussion on Wednesday, January 7, 2026

1. Soft and hard links are often used for convenience purposes when structuring file data. To avoid unexpected behavior, it is important to understand their behavior, in particular the conceptual differences between soft and hard links. Answer each of the following questions for soft and for hard links separately.

   (a) Suppose you set a link named `linktofile` to a file named `fileA`. Then you delete `fileA`. What happens? Is the data gone? If not, how do you access it?

   (b) Suppose you set a link named `linktofile` to a file named `fileA`. Then you delete `linktofile`. What happens? Is the data gone? If not, how do you access it?

   (c) Can you determine if a file is being linked to? How? (There is no need to describe precise Unix/Windows commands, a conceptual answer suffices.)

   (d) Suppose you write a program to copy (a part of a) file system to a backup location and to restore the data from backup if necessary. What may go wrong if your data includes hard or soft linked files?

   (e) You are about to modify a file that is being linked to. Because you are unsure about the modification, you want to keep an old copy around. How to you proceed? What do you have to be aware of?

   (f) Can you link to a directory? What can possibly go wrong?

   (g) What is faster – file access via a hard link or via a soft link?

2. In RAID-level 5, data is lost if two disk drives fail within the time it takes to rebuild a failed disk.

   (a) Suppose p is the probability that a single disk fails within the rebuild time. Assuming that disk failures are statistically independent, show that for a RAID 5 array with n discs, the probability of total failure within that time is given by

   $$1 - \left((1-p)^n + n\,p\,(1-p)^{n-1}\right).$$

   Hint: Binomial distribution.

(b) Show that the expression from part (a) is approximately equal to

$$n\,(n-1)\,p^2\,.$$

Comment on this result.

(c) A different way to look at reliability is the rate of *unrecoverable read errors* (URE), which is defined as the the probability of a single bit error during read. A commonly cited value for consumer drives is $\mathrm{URE} = 10^{-14}$. Assuming that single-bit errors are statistically independent, compute the probability that a RAID array of n disks with capacity c can be fully recovered when one disk fails.

(d) Using the formula from (c), what is the chance of complete recovery for a RAID 5 array consisting of four consumer-grade disks with 2 TB each, upon failure of one of the disks? Comment on the result.

(e) What is the write speed of a RAID 5 array relative to the write speed of a single disk (i) when sequentially writing large files and (ii) when randomly updating single disk blocks?