Exercise 1

1 Disk Failures

Read the following paper:

> B. Schroeder and G. A. Gibson, "Understanding disk failure rates: What does an MTTF of 1,000,000 hours mean to you?," *ACM Transactions on Storage*, vol. 3, no. 3, 2007.
> http://doi.acm.org/10.1145/1288783.1288785

Write a short summary the main findings for yourself and be prepared to present it briefly (5 min).

2 Replication and Erasure Coding

Suppose a storage system *M* stores a payload of size S. The storage system consists of d disks, each of which has size $s \ll S$. There are enough disks so that the data can be stored replicated with redundancy $r = \frac{ds}{S} > 1$. (For simplicity, assume r is an integer and S is a multiple of s.)

An *m-out-of-n erasure code* is a method to split some data into m parts, and to encode the parts into n fragments (each one of the same size as a part), such that the data can be recovered from any m of the n fragments.

a) Suppose the data is stored using r-fold replication. Given the failure rate of a disk λ_{Disk}, derive the MTTF or the failure rate of *M*.

b) A typical observed annual replacement rate for disks is 3%, which corresponds to an MTTF of 292'000h. Assume hence that $MTTF_{Disk} = 500'000h$, and let *M* store $S = 20$TB on 600 disks of capacity $s = 100$GB each.

Compare the MTTF of the following coding schemes:

i. The disks are divided into 3 equal-sized pools, and the data is replicated on to the three pools.

ii. Three disks are each taken together to form one logical volume with 3-fold replication, and the data is stored on the 200 logical volumes.

iii. A $(20, 10)$ MDS erasure code is used to create a logical volume across every group of 20 disks (tolerating the failure of 10 disks inside each group of 20), and the data is stored on the 30 logical volumes.