Exercise

A Content Service Provider (CSP) provides a service to its customers through a server battery (server farm) consisting of K servers.
The servers, while operating, are subject to failures. For this reason, the CSP has additional M backup servers: whenever one of the servers in the server farm fails, it goes for repair and is immediately replaced with a backup server, if there is one available.
During the time intervals when there are K servers in the server farm, the system provides the service normally ("Normal State").
However, the CSP considers it appropriate to suspend the provision of the service during the time intervals when the number of servers in the server farm is less than K; during these intervals the system is in "Failure State" and the servers, as they are not operational, are not subject to failures.
As soon as it is repaired, a server becomes available as a backup unit and, if the system is in "Failure State", it immediately becomes part of the server farm, so that the service is reactivated.
For simplicity, suppose that the components of a server subject to failures are the processor, the main memory, the hard drive.
Particularly, let the time-to-failure of a processor be an exponentially distributed random variable with mean 1/fp, the time-to-failure of the main memory be an exponentially distributed random variable with mean 1/fm, the time-to-failure of the hard drive be an exponentially distributed random variable with mean 1/fd.

Suppose, moreover, that the repair times of servers are exponentially distributed random variables with mean $1/\mu$.
After introducing any further needed hypotheses,

a) model the system with a CTMC, draw the related transition rate diagram and discuss the ergodicity;

b) evaluate, as a function of the steady-state probabilities for the markov chain defined (suppose they have been calculated):

- the fraction of the time when the service is suspended (Unavailability);
- The Mean Time To Failure (MTTF) of the system, meaning the mean time that elapses from an instant when the service is reactivated (the system enters the normal state) to the next instant when the service is suspended (the system enters the failure state);
- the frequency with which the system enters the failure state.