On Fault-Tolerant Data Placement in Storage Networks

Today, it is usual for companies to face the permanent growth of relevant data by means of storage consolidation, i.e. the storage of the data on (perhaps different sized storage systems) inside a single storage area network (SAN) by regarding that a) every access to the data blocks can always be answered quickly and efficiently and furthermore that b) access to the data is ensured permanently under any circumstances. To guarantee this, data is cut into peaces (data blocks) which are then distributed over the connected storage subsystems in the SAN as even as possible what leads to access parallelisms. Importantly, as the failure probability of the connected systems increase with their number, fault-tolerant data placement strategies are hardly required.

In this thesis, we investigate the efficient fault-tolerant placement of homogeneous data blocks in static and dynamic storage (area) networks We introduce a couple of fault-tolerant strategies for an efficient data allocation in a SAN while fault-tolerance is guaranteed by introducing redundancy into the system. These strategies can be distinguished with respect to their required storage overhead as well as their capability to cope with scalable and/or heterogeneous capacities of the given storage subsystems. We show by detailed analysis that randomized algorithms are the better choice to provide an efficient data distribution in scalable and heterogeneous storage environments whereas for homogeneous storage systems we rather present novel erasure resilient codes, called Read-Write Codes, that, in contrast to usual erasure codes, like RAID and Reed-Solomon codes, offer advanced data modification properties concerning the stored data blocks.