Abstract

The explosion in the volumes of data being stored online has resulted in distributed storage systems transitioning to erasure coding based schemes. Yet, the codes being deployed in practice are fairly short. In this work, we address what we view as the main coding theoretic barrier to deploying longer codes in storage: at large lengths, failures are not independent and correlated failures are inevitable. This motivates designing codes that allow quick data recovery even after large correlated failures, and which have efficient encoding and decoding. We propose that code design for distributed storage be viewed as a two step process. The first step is choose a *topology* of the code, which incorporates knowledge about the correlate d failures that need to be handled, and ensures local recovery from such failures. In the second step one specifies a code with the chosen topology by choosing coefficients from a finite field $\mathbb{F}_q$. In this step, one tries to balance reliability (which is better over larger fields) with encoding and decoding efficiency (which is better over smaller fields). This work initiates an in-depth study of this reliability/efficiency tradeoff. We consider the field-size needed for achieving *maximal recoverability*: the strongest reliability possible with a given topology. We propose a family of topologies called grid-like topologies which unify a number of topologies considered both in theory and practice, and prove the following results about codes for such topologies:

- The first super-polynomial lower bound on the field size needed for achieving maximal recoverability in a simple grid-like topology. To our knowledge, there was no super-linear lower bound known before, for any topology.
- A combinatorial characterization of erasure patterns correctable by Maximally Recoverable codes for a topology which corresponds to tensoring MDS codes with a parity check code. This topology is used in practice (for instance see [MLRH14]). We conjecture a similar characterization for Maximally Recoverable codes instantiating arbitrary tensor product topologies.