Optimal Linear Functional-Repair Regenerating Storage Codes

Henk D.L. Hollmann

University of Tartu, Estonia

(In part, joint work with Junming Ke and Ago-Erik Riet)

The amount of data in need of storage keeps increasing at an astonishing rate. Therefore it is of paramount importance to develop methods to store all this data in a reliable, efficient, and economically feasible way.

In modern storage systems, the data storage is handled by a *Distributed Storage System (DSS)*. Typically, a DSS stores data in encoded form accross a number of storage units, often referred to as storage nodes. So a DSS employs *redundancy* to be able to handle the occasional loss of a storage node. Often, a DSS simply employs *replication*. But many modern DSS employ non-trivial erasure codes (for example based on *Reed-Solomon* codes), especially to store "*cold*" data (data that remains unchanged, for example for archiving purposes, but sometimes even for "*lukewarm*" data (data that needs only occasional updating). An obvious requirement is that at any time, the original data can be reconstructed from the data stored by the storage nodes.

Typically, to repair a lost storage node, the DSS contacts a small number of *helper*— nodes, downloads a small amount of data from each of them, and uses this data to compute a new data block that is then stored on a replacement node. In the simplest form of repair, the data on a lost node is reconstructed exactly; this repair regime is commonly referred to as *exact repair*. A more interesting and potentially more efficient repair regime is *functional repair*, where the lost data is not necessarily reconstructed exactly, but still data integrity is maintained over time (this will be explained in detail during the talk).

In this talk, we first introduce *Regenerating Codes (RGC's)*, a commonly used storage code model introduced in a famous paper by Dimakis et.al. The performance of RGC's is described by the so-called *cut-set bound*. We will present a new simple derivation of this bound that demonstrates some properties of codes that meet this bound (called *optimal* RGC) that seem to have gone unnoticed before.

Interestingly, it is known that every point on the cut-set bound can be realized by a RGC, but with the exception of two special points called the *Minimum Storage Regenerating or MSR* and the *Minimum Bandwidth Regenerating or MBR* points, codes realizing such points are necessarily functional repair codes; moreover, except for MSR and MBR codes, these codes require huge fields for their construction, and come without any (useful) repair method. In fact, with one single exception, no optimal *explicit* codes, that is, codes that have a simple description with an efficient repair algorithm, are known that realize a boundary point different from the MSR or MBR point.

In our talk, we mainly discuss *linear* storage codes, where the data symbols are elements of a finite field and where all operations (encoding, data recovery, repair) are linear over that field. Such codes can be described "basis-free", that is, in terms of sub-spaces of the ambient vector space formed by the data vectors. Inspired by the newly derived properties of optimal RGC's, we then introduce the notion of (r, α)-regular configurations of subspaces, and show how to use these configurations to construct optimal linear RGC's also outside the MSR and MBR point. These codes are simple to describe and come with efficient repair, but are very large; but sometimes it is possible to find small *sub-codes* of these codes. We will discuss two such examples, one producing this exceptional code mentioned before and one of a new code. Both codes are based on vector space partitions; the vector space partition of the new code turns out to form a (non-linear) representation of $\mathrm{PG}(2,8)$.

References

[1] H.D.L. Hollmann, W. Poh, *Characterizations and construction methods for linear functional-repair storage codes*, Proceedings 2013 IEEE International Symposium on Information Theory, ISIT-2013, Istanbul, Turkey, 7–12 July 2013, pp. 336–340.

[2] H.D.L. Hollmann, *A family of optimal linear functional-repair storage codes*, Entropy 27, no. 4: 376 (2025). https://doi.org/10.3390/e27040376.

[3] H.D.L. Hollmann, J. Ke, A.-E. Riet, *An optimal binary linear functional-repair storage code with efficient repair related to* $\mathrm{PG}(2,8)$, Des. Codes Cryptogr. (2025). https://doi.org/10.1007/s10623-025-01682-0.