Improving the Reliability and Repairability of Storage Systems Using Two-Dimensional Erasure Coding*

Kyumars Sheykh Esmaili

Nanyang Technological University

639798 Singapore

kyumarss@ntu.edu.sg

In the past few years, Erasure Codes have been embraced by distributed storage systems –including Microsoft Azure, Google File System, and Facebook's HDFS-RAID– as an alternative for replication, since they provide high reliability for low overheads.

Traditional erasure codes have mostly been designed to optimize the performance of communication-centric applications, and are not necessarily amenable to the needs of storage systems. One of such desirable properties is efficient replenishment of lost redundancy (a.k.a repair) following the failure of some of the storage nodes. To that end, there has been tremendous interest in both coding theory and storage systems research communities to build new erasure codes with good repairability properties.

In this paper we explore an engineering approach and build a two-dimensional coding scheme: a traditional erasure code is first applied on individual data objects, followed by the creation of RAID-4 like XOR parity over erasure encoded pieces of different objects, creating cross-object redundancy (CORE). This results in high fault tolerance (provided by the traditional code) and cheap repairs (provided by the parity code). The approach is simple, and based on mature techniques, yet it achieves very good repairability. We accordingly build the *CORE storage primitive*, integrate it into HDFS-RAID (Facebook's erasure code supporting extension of Apache HDFS) and benchmark it over a wide range of system configurations to demonstrate its efficacy.

In the first half of my talk, I will briefly explain the concept of Erasure Codes. In the second half, I will focus on the issue of repair inefficiency and after a quick review of the existing proposals, I will present CORE and highlight its benefits.

---

*The original paper:

K. S. Esmaili, L. Pamies-Juarez, and A. Datta, *CORE: Cross-Object Redundancy for Efficient Data Repair in Cloud Storage Systems*, in the Proceedings of the IEEE International Conference on Big Data, Santa Clara, October 2013.