Connections between data reconciliation and generalized error-correcting codes

Ari Trachtenberg *trachten@bu.edu*
Mark Karpovsky* *markkar@bu.edu*
Lev Levitin *levitin@bu.edu*

Reliable Computing Lab
Boston University

Abstract

We consider the problem of multiset reconciliation, in which two physically separated multisets of data must be reconciled with a minimum amount of communication. Though motivated by applications to gossip protocols, this problem applies naturally to a wide variety of applications, such as resource discovery, synchronization of mobile data, and file comparison. We assume a general model in which one multiset is the image of the other multiset under any of an arbitrary set of transformations.

Under this model, we show that the problem of reconciliation is equivalent to variants of the problem of graph coloring and provide consequent upper and lower bounds on the communication complexity of reconciliation. More interestingly, we show by means of an explicit construction that the problem of reconciliation is equivalent to the problem of finding good error-correcting codes provided the set of transformations has two general properties. When this is not the case, we demonstrate with counterexamples that reconciliation requires more communication than error correction. This allows the application of the vast body of coding theory literature to the reconciliation problem. We show analogous results for the problem of multiset verification, in which we wish to determine whether two multisets are equal using a minimum amount of communication.

Finally, we provide a variety of examples that justify our general model of reconciliation. We analyze such problems as: calibration, in which multisets are cyclic shifts of one another; file synchronization, in which multisets encode files that differ by a prescribed number of edits; client-server reconciliation, in which one multiset is a contained in the other; reconciliation amidst page errors; and reconciliation of permuted multisets.

*The work of this author is supported by the NSF under grant MIP9630096