

1135-15-2012

Ryan Gabrys* (ryan.gabrys@navy.mil), 4573 Monroe Ave., San Diego, CA 92115. *Reconciling Similar Sets of Data.*

Suppose we have some data (a file for instance) on Host A and some related data on Host B, and it is desired that Host B obtains all the information on Host A. One naive approach would be for Host A to simply transmit all of its data to Host B; however, if the sets are similar, this approach could be wasting valuable network resources. Another approach is to compute hashes on the data and then compare hashes to iteratively determine the difference. With this approach, many rounds of communication are required. In adverse network conditions, the more rounds of communication required, the greater the stress placed on already limited network resources. As a result, we are interested schemes for synchronizing data that require a small number of communication rounds and have small communication overhead.

In this work, we consider the problem of synchronizing two sets of data where the size of the symmetric difference between the sets is small and, in addition, the elements in the symmetric difference are related through the Hamming distance metric. Upper and lower bounds are derived on the minimum amount of information exchange. Furthermore, explicit encoding and decoding algorithms are provided for many cases. (Received September 25, 2017)