

ABSTRACT

Due to dramatic growth of data storage and transfer for most enterprises nowadays, it is necessary to have larger memories to store numbers of binary generated from everyday communication and documentary. Especially, for large or multinational enterprises where various parts of digital communication and documentary are duplicated and redundant, many data compression techniques have been used to reduce the storage requirements by compressing the data binaries.

In this research, a simple preprocessing technique for improving compression performance of text data in tabular form is introduced. Data sharing through cloud computing in the field of supply chain is selected as the illustration. Sets of duplicated data are preprocessed based on analyses of static-data ratio and number of receivers in order to generate a single compressed file, while different decoding overheads are encoded and separately distributed to particular receivers in parallel.

The results show that compression performance can be improved with some trade-offs for decoding overheads. Static-data ratio and size of duplicated static data are found to be directly proportional to the compression performance with some benefits to data confidentiality, while number of receivers is related to the efficient use of encoding tokens.