

ABSTRACT

The purpose of this research is how to get lossless compression encryption in millisecond of completion. Some notable research challenges are about storing, transferring and safety of deoxyribonucleic Acid (DNA) order. Although pattern matching for text compression has been observed for a few years and many publications are available in literatures, there is still space to enhance the effectiveness in terms of both compression & encryption. Human beings are always fond of acquiring more and more information in least possible time and space. Nowadays sending (power and so on) of DNA/RNA/protein order especially over wireless network is very common.

The DNA database size increases greatly, changing from millions to billions annually. Therefore, for storing, searching the DNA database needed a systematic lossless compression and encryption algorithm for safe transmission. In the field of Bioinformatics the storing and transmission of DNA is very important with respect to compression rate, ratio and encryption point of view. The DNA order needs greater space for storing & more time for encryption causing much loss of time in sending of information.

The recurrence of the DNA short pattern are the highest qualities in biological orders. The offered compression algorithm will be based on combinations of REPEAT, REVERSE, GENETIC PALINDROME & PALINDROME. Another offered compression & selection encryption algorithm have modified Huffman's and RSA algorithm. This algorithm is based on searching exact repetitions, substring substitution by corresponding ASCII code and producing library file, as an outcome there is cumulating of data facts. In this method the data is safe by using the ASCII code for information interchange value and producing the library file which act as a signature. The Huffman's algorithm is used on the output in the first stage of the repetition method and also including the change of Huffman's tree level & node position for encryption.

It can give the safety of facts, the sudden coded value is essential for decoding by using only certain coded value allotted by encoded time. This offered method safeguards the sequence by applying ASCII symbol, it is user friendly. This type of security is provided in tier one. In tier two selective encryption techniques are used for higher quality safety.

Form the information point of view the most demanding question nowadays is about the safety of information during transmission. The selection encryption process seems to provide security and this technique is applied on compressed data or in the library file or in both. . The fractional part of a message is encrypted in the selective encryption method, leaving the remaining part unchanged. This is top most important with respect to the selective encryption

system. The offered selection encryption makes smaller the computational alternatives of this data. The safety of this is ensured by signature that depends on ASCII code & progressive library file acting as a key. As an outcome of that systematize lossless compression technique, data structure, to store effectively, access secure communication and search the greatly sized data sets are essential. These days DNA/RNA sequence with a complex structure that stores facts of different types at the same time are in common use. The operating time is very less and it depends on the input file size. The assessment of encryption system depends on its rate of motion and levels of safety it gives. The operating time of this algorithm is minimum, needed minute memory and can be facilely used. The mass request is for the need of minimum place for storing and low computational price, so, systematized algorithm is needed for compression encryption.

The compression minimizes the file size and encryption makes certain the safety of a particular file which is to be sent over some uncertain network like the internet. In this age of information sharing and transferring of data have increased to a great extent. Generally the information exchange is done using open narrow ways by making it unsafe to interception. On the other hand, effective information retrieval is needed to quickly discover the relevant information from this huge mass of facts using ready to be used materials.

For that purpose make greater, stronger, more complete six compression algorithms for making shorter greatly sized collections of DNA orders and two selection encryption of modified Huffman's & RSA are presented. When a user searches for any order for an organism, an encrypted compressed order DNA sequence can be sent from source to user. The encrypt compressed the DNA sequences then can be decrypted & decompressed at the client end producing lower transmission time over the internet.

The experimental results show that our compression-encryption algorithm is in competition with the best algorithms and is almost the fastest among all views when the number of pattern is not very greatly sized. As an outcome of that, this algorithm is desired for general string matching applications. These data structures and algorithms can be used in several situations and experimentally show that they can successfully make an attempt to be placed over with other techniques commonly used in those fields (of knowledge). This work, therefore, is greatly economical and has market potential

This algorithm also experiments on benchmark data and equivalent length of artificial sequences. By applying modified Huffman's technique the rate & ratio is lowered. It also makes a comparison of the compression technique with published results and selection encryption with RSA algorithms.