INTERNATIONAL JOURNAL OF
INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Hamming Distance Based Substitution Cipher For Coding Optimization

Prof. Atul S. Joshi
Associate Professor,
Department of Electronics and Telecommunication Engineering,
Sipna College of Engineering and Technology,
S.G.B.Amravati University,
Amravati (Maharastra State), India

Dr. Prashant R. Deshmukh
Professor & Head of Department of Computer Science and Engineering,
Sipna College of Engineering and Technology,
S.G.B.Amravati University,
Amravati (Maharastra State), India

Abstract:
The traditional way of data transmission is to first compress the data & then encrypt it. The proposed algorithm suggests a reversing the order of these steps without loss of information. Polygram substitution method is adopted for encryption. & hamming distance of codeword is taken into account for compression techniques. Bit stream is chopped into either Four or eight bits each, and a minimized code is found for each of the chunks by first encrypting it & then compressing.

The proposed algorithm can be used for various file formats such as images, videos and text. A full cycle for the proposed algorithm is to compress a file, encrypt it, decrypt it, and finally decompress it back identically to the original file.

Key words: *Compression, Decompression, Substitution, Encryption, Decryption.*

Introduction

Remarkable new development that advances state of art in the field of data Communication, provide variety of services suitable for data transfers. Estimation of the power is based on the number of bits needed to represent the data sets. Maximization of the overall system performance in high speed data transmission network can be achieved by appropriate encoding scheme which further optimize the storage requirement. Concept based on information theory has now come of age and proved its value by providing sophisticated methods capable of producing high compression ratio, widely used in image, video & audio standards.

In data communication, presentation layer establishes concrete transfer syntax for data type and handle pass through of services from session to application layer [1]. Examples of these services are text compression, data encryption, terminal handling and file transfer. The proposed work examines the more challenging part of the transfer syntax for nondeterministic data and develop improved encoding algorithm by combining the number of matching patterns together based on Hamming distance of the code vectors as shown in figure no.1 The algorithm looks ahead for the encryption of the compressed data. Proposed algorithm reduces the transmission bandwidth & achieves higher fidelity without additional encoding complexity.

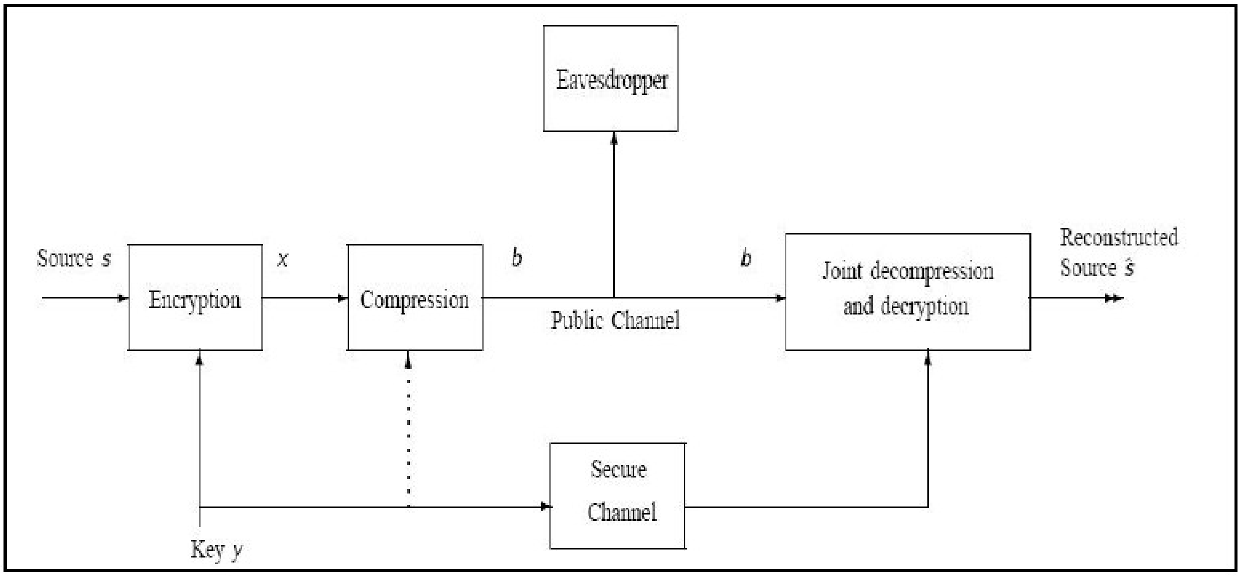

Figure 1: Proposed encryption & compression

Problem of finding out the frequently occurring appropriate dictionary and bit mask patterns which need the knowledge of the entire data before compression [2] is tackled by the said algorithm.

The proposed method is evaluated with various reversible variable length encoding techniques which improves coding efficiency at the expenses of exponential complexity .However this work perform well in overall efficiency even for large amount of data without system complexity.

Review Of Existing Approaches & Discussion

Wavelet predictive algorithm results well for relatively small data set. If the high degree of image compression is to be achieved then wavelet algorithm closely approximates the original data sets leaving only small residual values. However the predictive algorithm would be no useful for the text compression because there is no underlining deterministic process in natural language text stream that can approximate the function [3].

Huffman procedure proposed by Wolfe and Chanin [4] creates optimal code for set of symbol and probability subject to constraints that symbol be coded at one time. It is very effective as both the frequency and probability occurrence of the source symbol are taken into account. But since tree progressively spares results in lengthy search procedure for locating the symbol.

Daniel Hillel Schonberg [5] presents practical distributed source code that provide framework for compression of encrypted data. Since encryption masks the source code, traditional compression algorithm is ineffective. Through the use of distributed source coding techniques, compression of encrypted data is possible. However it requires special framework for distributed source coding prior to compression & encryption.

M. A. Haleem, K.P. Subbalakshmi & R. Chandramouli [6] proposed a Joint encryption & compression scheme. It reduces complexity of compression process & at the same time use cryptographic principles to ensure the security. However proposed scheme required fundamentals of distributed source coding which makes it complex.

Dr.V.K.Govindan & B.S.Shajee [7] Mohan proposed better encoding scheme which offers higher compression ratio & better security towards all possible ways of attack. This algorithm compression transform the text into some intermediate form which can be compressed with a better efficiency and which exploits the natural redundancy of the language in making this transformation. However selection of appropriate and intelligent dictionary is required prior to encoding.

Dictionary based code encoding techniques suggested by H. Lekatasas and J. Henkel [8] provide good compression efficiency as well as fast decompression mechanism. The basic idea is to take advantage of commonly occurring instruction sequence by using a dictionary & repeatedly occurrence are replaced by codeword that point to index of dictionary. However it does not compress the mismatch patterns.

Compression using bit mask tries to incorporate maximum bit changes using mask patterns without adding significant cost such that compression ratio is improved. The compressed code along with any uncompressed ones is composed serially to generate the final compressed program code. Larger bit mask patterns generate more matching patterns. However doing so may not generate better compression. A longer bit mask patterns is associated with higher cost, hence applying more bit mask not always beneficial.

Major challenge of the data communication is that it should ensure encoding that reduces total number of bits to conserve communication bandwidth and further it should transformed data into non intelligible format at the sending end to protect the network resources against unauthorized disclosure, modification, utilization or destruction.

Proposed algorithm creates number of matching patterns based on Hamming distance to compress the bit stream & compressed patterns are again compressed by using parallel mechanism at the output stage. Due to parallel mechanism used for encoding which required only few levels for encoding any type of data, results into increased efficiency for the encoder and required simple hardware construction & increases speed of operation. It generates the output in encrypted form based on Polygram substitution & maintains the network security & suitable for any type of input data.

Proposed Algorithm

- Plaintext bit steam is consider as a blocks.
- Each block consist of 4 bits
- For the block of plaintext, Key is generated of 4 bits.
- Key is compared with plaintext block.
- According to plaintext, bits of the Key gets change.
- Code book is form based on the Hamming Distance of the KEY. In the codebook all entries are Hamming distance of 2 from the Key. These entries are dictionaries by 3 bits right from 001 to 110.
- From dictionary 3bits are assigning to encrypted data. Thus data gets compressed.

Compression Prior To Encryption & Encryption Prior To Compression

The behavior of fixed-block-length codes and fixed-delay codes can be quite different in contexts where the message to be communicated is revealed to the encoder gradually as time progresses rather than being known all at once. It is assume that information arises as a stream generated in real time at the source. The encoded bit stream is also assumed to be transported at a steady rate. The acceptable end-to-end delay is determined by the application and can often be much larger than the natural granularity of the information being communicated .The end-to-end delay perspective here is common in the networking community. This is different from cases in which information arises in large bursts with each burst needing to be received by the destination before the next burst even becomes available at the source. In the traditional compression-first approach because of the lack of side-information compression can be modeled by having relevant error exponent with end-to-end delay [9]

For encryption prior to compression approach [10] the secret key is used at a rate of log2 |S| bits per source symbol to generate uniform virtual side-information. From the encoded data bits, the eavesdropper learns nothing about the source symbols. Marginal distributions for both the encrypted data and the secret key are uniform. This means that nothing higher than the fixed-block-length error exponent for source coding can be achieved with respect to end-to-end delay if the encryption-first architecture is adopted .In practical terms, this means that if both the end-to-end delay and acceptable probability of symbol error are constrained then the approach of encryption followed by compression can advantageous to use.

Conclusion

Proposed work is applicable for network applications involving Public transmission facilities such as PSTN, telegraph network, multimedia networking and text conferencing. Simple & parallel mechanisms is going to generate the number of matching patterns in proposed method and provide encoding and encrypted output for large unknown data string. Hence may be suitable for performance improvement of concrete transfer syntax in data communication.

Reference

1. Gred E. Keiser , " Local area network " , Tata Mc Graw Hill Edition , 1997, pp 443-497
2. Seon-Won Seong & P. Mishra , " Bitmask based code compression for embedded system " , IEEE transaction on computer aided design of integrated circuit & system , vol. 27 , No. 4 , April 2008 , pp 673-685
3. S. Shani, B.C. Vemuri , F. Chenc Kapoor, " State of art image compression algorithm" , October 30, 1997.
4. Wolf & A. Chanin , " Executing compressed program of embedded RISC architecture" , in poc. Int. symp. Micro, 1992,pp 81-91
5. Daniel Hillel Schonberg , " Practical Distributed Source Coding & its application to the compression of Encrypted data" , Technical Report No. UCB/EECS-2007-93, July 2007
6. M.A. Haleem , K.P. Subbalakshmi , R. Chandramouli , Joint encryption & compression of correlated sources", EURASIP Journal on Information Security , Jan. 2007
7. Dr. V.K. Govindan , B.S. Shajee Mohan , " IDBE – An intelligent Dictionary Based Encoding Algorithm for text data compression for high speed Data transmission ", Proceeding of International conference on Intelligent signal processing , Feb 2004
8. H.Lekats , J. Henkel , " Design of one cycle decompression hardware " , in pocd. Des. Conf. 2002 , pp 34-39
9. R. L. Dobrushin, "An asymptotic bound for the probability error of Information transmission through a channel without memory using the feedback," Problem Kibernetiki, vol. 8, pp. 161–168, 1962.
10. C.-P. Wu and C.-C. J. Kuo, "Efficient multimedia encryption via entropy codec design," in Security and Watermarking of Multimedia Contents III, vol. 4314 of Proceedings of SPIE,pp.128–138, San Jose, Calif, USA, January 2001.