IOP Conf. Series: Materials Science and Engineering **322** (2018) 062011 doi:10.1088/1757-899X/322/6/062011

File compression and encryption based on LLS and arithmetic coding

Changzhi Yu[1], Hengjian Li[2,*], Xiyu Wang[1]

[1]School of Information Science and Engineering, University of Jinan, Jinan, China
[2] School of Information Science and Engineering, University of Jinan, Jinan, China

*Corresponding author e-mail: ise_lihj@ujn.edu.cn

Abstract. e propose a file compression model based on arithmetic coding. Firstly, the original symbols, to be encoded, are input to the encoder one by one, we produce a set of chaotic sequences by using the Logistic and sine chaos system(LLS), and the values of this chaotic sequences are randomly modified the Upper and lower limits of current symbols probability. In order to achieve the purpose of encryption, we modify the upper and lower limits of all character probabilities when encoding each symbols. Experimental results show that the proposed model can achieve the purpose of data encryption while achieving almost the same compression efficiency as the arithmetic coding.

1. Introduction

With the progressive of social science and the rapidly development of multimedia technology, more and more information are transmitted on the network, and the network has gradually becoming the main way to get information, In order to transmit the information more quickly and make better use of the network traffic, then the encrypted ciphertext needs to be transmitted, the space occupied by the ciphertext is as small as possible. According to different data types and reconstruction quality requirements, the compression algorithms are also different. For example, video compression, image compression [1, 2], speech compression and so on, thesealgorithms can be divided into two categories, one is loss compression [3], and the other is lossless compression.

Lossless compression is required for the compression of text data. At present, there are many algorithms for text compression. For example, predictive coding [4, 5], LZ coding, run-length coding, LZW and other algorithms are prominent text search algorithms. LZW search algorithm needs to establish a dictionary, which can improve the search speed. The most widely used text encryption tools are WinRAR and WinZip. Chaos system is widely used in encryption, mainly taking into account the characteristics of the sequence stream generated by the chaotic system, suitable for encryption system key stream. Utilize the chaotic system's characteristics: initial condition sensitivity, internal randomness, non-regular order, non-correlation, the generated chaotic sequence can be used as an encrypted sequence.

At present, the most common chaotic encryption systems are as follow: Logistic chaotic map, sine map, Lorenz chaotic map and so on. Arithmetic coding as a kind of entropy coding has higher compression efficiency than Huffman coding. Due to AC's good compression performance, it had become the mainstream of lossless compression and been widely used in various compression standards such as H.264 [6], H.265 [7] and JPEG. For the security of arithmetic coding, adaptive arithmetic coding has a better encryption effect [8] because of the relatively random coding interval

generated in the coding process, however, it can't resist selective plaintext Therefore, in order to improve the security of traditional arithmetic coding, and we proposed a compressed encryption scheme based on arithmetic coding.

The rest of this paper is organized as follows: section2 we present a brief literature review. The proposed model is described in section3. Experimental results and security analyses are given in section4. Section5, we conclude this paper.

2. Literature Review

In this section, we present a brief literature review on our proposed model. After Shannon first put forward the theory of information, Elias proposed the basic idea of arithmetic coding. In 1987, Witten et al [9] pointed out that arithmetic coding has better compression efficiency than Huffman coding and applies arithmetic coding to data compression.

The main idea of Arithmetic coding model is: starting from the entire symbol sequence, we represent the symbol sequence by a true subset of [0, 1). Every time a symbol is encoded, the encoding interval is reduced once, the extent of each reduction depends on the current join the probability of the source symbol of each new interval so that each can only represent a period of information. Some papers [10, 11] states that arithmetic coding has become a viable entropy coding. The independent probability model and coding module composed of arithmetic coding is more flexible than other entropy coding, which also makes it possible to encrypt arithmetic code while compressing.

For arithmetic coding, there are two kinds of probability and statistics model: static model and adaptive model. The static model means that the probability of the symbol to be encoded remains the same in the encoding process. The adaptive model means that the probability of encoding the symbol changes dynamically along with the encoding process.

3. The Proposed Encryption Compression Scheme

We find that just in a limited or discontinuous range that chaotic system appears chaotic behavior [13]. The security of low-dimensional chaotic system is not high, it is easily deciphered by attackers, and the randomness of chaotic sequences output by low-dimensional chaotic system is not good enough. In order to solve such problems, Yu et al and Frunzete et al proposed LSS chaotic system which is constructed by logistic chaotic map and sine map, respectively [14]. The definition of LSS chaotic system can be described by the Eq. (3)

$$x_{n+1} = A_{LS}(r, x_n) = (L(r, x_n) + S((4-r), x_n)) \bmod 1 = (r x_n (1 - x_n) + (4-r)\sin(\pi x_n)/4) \bmod 1 \quad (3)$$

Where parameter $r \in (0,4]$.

We can find that the LSS system makes the system with a wider range of parameters in the chaotic state. When one of its seed maps is out of chaotic range, LSS system can still have excellent chaotic behaviors.

Our proposed encryption algorithm's steps is shown in Figure 1: First reading the encrypted and compressed file, starting from the first character, we input the coded characters to the encoder in turn. Through the statistical model to update the frequency of each emerging symbol, through the probability of statistical modules, we update the symbols the cumulative frequency of each symbol dynamically. Finally, changing cumulative frequency of each symbol by LSS chaotic system. As shown as Figure 2.

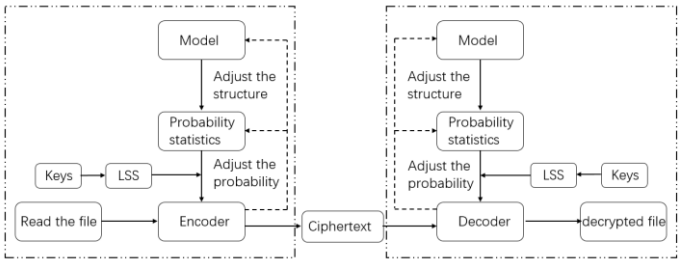

Fig.1 The proposed encryption algorithm

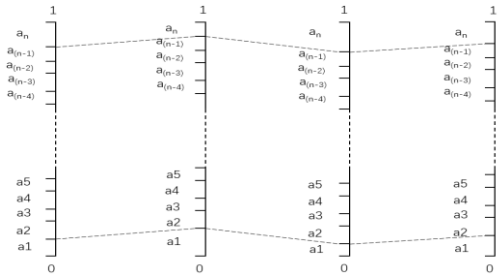

Fig.2 Changes probability range randomly

As shown in Figure 2, we change the cumulative frequency of each symbol according to the chaotic value of the corresponding chaotic sequence. Although the upper and lower limits of the symbols, which is to be encoded, have been changed, the probabilities have not alter, so the compression efficiency of the arithmetic coding does not decrease. It should be noted that the probability of first and last symbols in the probability model will change, but this error is negligible relative to most of the characters to be encoded.

4. Experimental Results and Analysis.

Our proposed model is implemented using Matlab 2016a. To demonstrate our findings various 8-bit grayscale images like Lena, Peppers camera and Baboon, each of size 256×256, are used.

4.1. ity analysis

The model, we proposed, is a data compression then encryption algorithm based on arithmetic coding. Therefore, the security of the model is very important. We calculate the key spaces of our proposed model. When encrypt and compress the data, we need to input two sets of keys, each with a precision of 10^{-15}. Therefore, the key space is: $10^{15} \times 4 \times 10^{15} \approx 2^{95}$. Therefore, we can get the conclusion that our model has better security strength.

4.2. Encrypted-compressed text file

In our experiment, we selected four files, which size are: 40960 bytes, 66048 bytes, 236032 bytes and 846336 bytes. When encrypting a file, the key, we used, are 0.1, 3.4567. At the same time, in order to verify the difference between the compression efficiency of our proposed compression model and the compression efficiency of the arithmetic coding model, we compress the above four test files using the existing mature integer arithmetic coding. The final results are shown in Table 2:

Table 2. Compression efficiency

ORIGINAL FILE (bytes)	COMPRESSED file (bytes)	ENCRYPTE-COMP FILE (bytes)	COMPRESSION EFFICIENCY	ENCRYPTE-COMP EFFICIENCY
40960	25727	25732	37.1900%	37.1777%
66048	44005	44016	33.3742%	33.3576%
236032	169177	169284	28.3245%	28.2792%
846336	625466	625504	26.0972%	26.0927%

From Table 2, we can see that the proposed compression coding model based on arithmetic coding has a small difference between the compression efficiency and the integer arithmetic coding. Therefore, our model can achieve good compression efficiency and better Encryption effect.

4.3. Encrypted-compressed image file

For combining image compression with image encryption, traditional algorithm is to compress and encrypt the process separately and this method has many shortcoming such as large amount of computation, slow speed and low flexibility. In order to overcome this problems, it is very important

to achieve compression and encryption simultaneously. And the algorithm, we designed, based on arithmetic encoding file compression encryption algorithm can not only achieve the lossless compression of documents but also make it possible to compress and encrypt at the same time, the features of our proposed model just to meet the image compression requirements.

The Lena (256 × 256) is used as our test image, the initial encryption password, we used, is 0.1, 3.4567, the following are the original image, enter the wrong key = 3.456700000000001 to decrypt the image, and the correct encryption of the image. What needs to be clarified here is that when the ciphertext image is incorrectly decoded, due to the change of the probability model, the coding range may occur the crossover between the upper limit and the lower limit. As a result, the decoder can't continue to decode. In order to achieve a comparable size to the original image, we manually zero and display the ciphertext that we need to decrypt. We decrypt all non-decryptable ciphertexts to 0 and display them. Finally, we show the results as Figure3.

(a) (b) (c) (d)

Fig.3 (a) The original file. (b) The encrypted image. (c) The decrypted file with a correct security key. (c) The decrypted file with an incorrect security key

5. Conclusion

Based on LSS system and arithmetic coding, an efficient image compression–encryption scheme is investigated, which possesses the compression ability of the arithmetic coding and accomplishes both image compression and image encryption simultaneously. The main idea of our proposed model is dynamically changed the upper and lower limits of the symbol probability model according to the random numbers which is generated by LSS chaotic system.

References

[1] DENG Jia-xian, DENG Hai-tao. "An image joint compression-encryption algorithm based on adaptive arithmetic coding", Chin Phys B, 2013, 22(9): 094202-1-094202-6.
[2] DENG Jia-xian, REN Yu-li. "Image joint compression encryption algorithm based on improved zero-tree coding", Acta Photonica Sinica, 2013, pp-121-126.
[3] WU J Z, WANG Y J, DING L P, et al. "Improving performance of network covert timing channel through Huffman coding", Mathematical and Computer Modelling, 2012, pp-69-79.
[4] TE Ri-gen, LI Xiong-fei, LI Jun. "Document compression coding scheme based on integer data", Journal of Jilin University, 2016, pp-228-234.
[5] ZIV J, LEMPEL A. "A universal algorithm for sequential data compression", IEEE Transactions on Information Theory, 1977, pp-337-343.
[6] Lorenz E N. "Deterministic Nonperiodic Flow", Journal of Atmospheric Sciences, 1963, pp-130-141.
[7] Gao M, Wang Q, Zhao D, et al. "Arithmetic coding using hierarchical dependency context model for H.264/AVC video coding", Multimedia Tools & Applications, 2016, pp-7351-7370.
[8] Cheng Chen. "High-Throughput Binary Arithmetic Encoder Architecture for CABAC in H.265/HEVC", IEEE Beijing Section. 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT) Proceedings. IEEE Beijing Section. 2016:3.
[9] WITTEN I H, CLEA R Y J G. "On the privacy afforded by adaptive text compression", Computer Security, 1988, pp-397-408.
[10] Ian H.Witten, Radford M.Neal, John G.Cleary, "Arithmetic coding for data compression", computing practices, 1987, pp-520-540.

[11] RISSANEN J, LANGDON G G. “Arithmetic Coding”, IBM Journal of Research and Development, 1979, pp-149-162.
[12] WITTEN I H, NEAL R M, CLEA R Y J G. “Arithmetic coding for data compression”, Communications of the ACM, 1987, pp- 520-540.
[13] C. Li, S. Li, M. Asim, J. Nunez, G. Alvarez, G. Chen. Comput. “On the security defects of an image encryption scheme”, ImageVis. 2009, pp-1371–1381.
[14] L. Yu, J.P. Barbot, G. Zheng, H. Sun. “Toeplitz-structured chaotic sensing matrix for compressive sensing”, International Symposium on Communication Systems Networks and Digital Signal Processing.2010, pp-229-233.