Data encryption using event-related brain signals

K.V.R. Ravi[1], R. Palaniappan[2], C. Eswaran[3] and S. Phon-Amnuaisuk[3]
[1]*School of Information & Communications Technology, Republic Polytechnic, 9 Woodlands Ave 9, 738964, Singapore.*
[2]*BioSignal Analysis Group, Dept. of Computing and Electronic Systems, Colchester, CO4 University of Essex, Vivenhoe Park, 3SQ,United Kingdom.*
[3]*Faculty of Information Technology, Multimedia University,63100 Cyberjaya, Malaysia.*
[1]*kvr_ravi@rp.sg;* [2]*rpalan@essex.ac.uk;* [3]*eswaran@mmu.edu.my*
[3]*somnuk.amnuaisuk@mmu.edu.my*

Abstract

A method based on event-related brain signal is used for data encryption. The idea is to shuffle the Huffman tree using an encryption key generated by electroencephalogram (EEG) signals recorded when the user perceives a common black and white line picture. As different persons have different thought processes, the generated key is unique to each individual and hence the encryption is robust to fraudulent attacks as compared to other encryption systems. Further, as Huffman tree is used to encode the data during encryption, the method achieves both compression and encryption. This pilot study has shown the huge potential of the method as it is impossible to be compromised.

1. Introduction

Biometric cryptosystems utilise physiological and behavioural characteristics to generate an encryption key that will allow conversion of plaintext (i.e. given data) into cipher text (i.e. encrypted data). Some of the biometrics that could be used for this purpose are face, fingerprint, hand geometry, iris, keystroke, signature, voice and odour [1]. In recent years, biometric methods based on electroencephalogram (EEG) signals have been proposed as a fraud resistant alternative to standard biometrics like fingerprints [2]. Generic cryptographic system is possession based, where possession of decrypting key is sufficient to decrypt the cipher text into plaintext. As most of the cryptographic keys are lengthy and in random order, they are normally stored on a computer or smart card and released through simple password authentication [1]. Biometric cryptosystems are obviously advantageous to generic cryptographic system as the encrypting key is more difficult to be stolen or compromised.

One of the major hurdles in biometric cryptosystems is the problem of obtaining a biometric encryption key that is secretive enough. For example, fingerprints are left everywhere and iris images can be captured by hidden cameras [3]. Further, most of the standard biometrics could be easily obtained through force (for example by threatening with a knife). However, the use of brain signals evoked during the perception of a common black and white line picture as proposed in this study circumvents this hurdle as it is practically impossible to be faked and the user could easily avoid concentrating on the picture during 'forced' situations. But the use of brain signals or electroencephalogram (EEG) in cryptography has not met with huge success simply because of the difficulty in obtaining unique key for each individual. In this study, we show that EEG in gamma band frequencies from specific channels could be utilized to obtain repeatable and unique encryption key. Further, the method introduces data compression as well (though this is not the main objective) as it is based on the use of the encryption key to shuffle leaf nodes of a Huffman tree obtained with Huffman data coding [4].

2. Basic principle of encryption method

Assume the word 'title' is to be encrypted. Using standard Huffman coding that assigns fewer bits (i.e. codewords) to alphabets with higher occurrences, we can get the tree as shown in Figure 1(a). If some of the leaf nodes (i.e. those with 2 children) were to be shuffled (swapped) using an encryption key, the tree would be different. For example, if an encryption key of 011 was used (where the bits represent the leaf nodes to be swapped; in this case children of L1 and L2 would be swapped), we would obtain the tree as in Figure 1(b). The original compressed data would be 1000101001, while the compressed and encrypted data would be 1011100010. If the original Huffman tree were to be used to decompress the data, we would get the incomprehensible 'tlttit_' where the underscore represents the last 1 bit of the data that would not have a matching codeword.

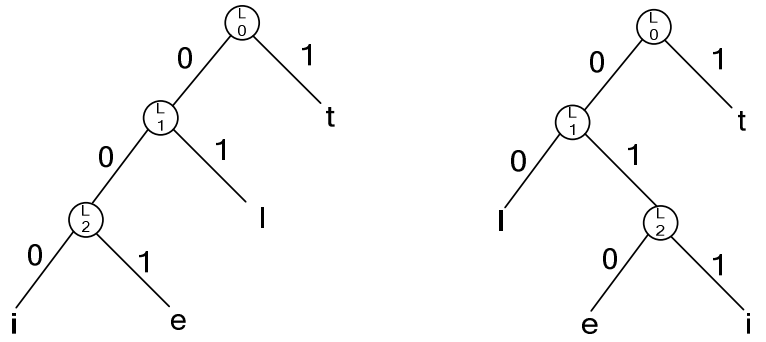

Figure. 1. (a) Original Huffman tree for the word 'title' (b) Shuffled Huffman tree (with leaf nodes L1 and L2 shuffled)

3. Using event-related EEG to generate the biometric encryption key

Forty event-related EEG signals were extracted from 10 subjects. One-second measurements were taken from 61 electrodes placed on the subject's scalp, which were sampled at 256 Hz. The electrode positions were located using extension of the standard sites (Standard Electrode Position Nomenclature, American Encephalographic Association). The event-related EEG signals were extracted from subjects while being exposed to a single stimulus, which were pictures of objects chosen from the Snodgrass and Vanderwart picture set. These pictures are common black and white line drawings like an airplane, a banana, a ball, etc. executed according to a set of rules that provide consistency of pictorial representation. The pictures have been standardized on variables of central relevance to memory and cognitive processing. These pictures represent different concrete objects that are easily named, i.e. they have definite verbal labels.

The EEG signals were filtered using Elliptic finite impulse response band-pass filter from 30- 50 Hz as it was shown earlier in [2] that EEG in this frequency range is unique to each individual. Energy of the filtered EEG was obtained and divided with the total energy from all the channels. Next, these values were normalized to obtain zero mean and standard deviation of 1.0. Positive normalized values were converted to binary digit 1 and negative values to binary digit 0. The procedure for generating an encryption key from one channel is shown in Figure 2. Table 1 shows the encryption keys from 61 channels for a subject.

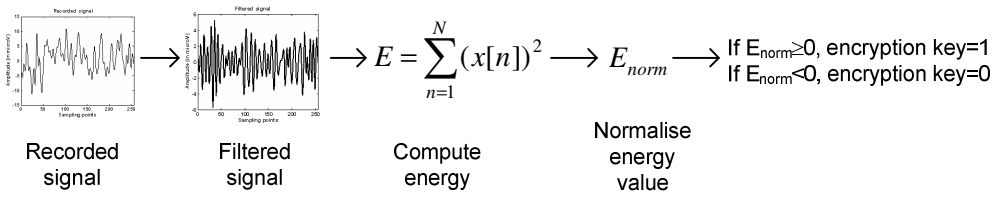

Figure. 2. Encryption key generation (for a channel)

Table 1. Encryption keys for a subject (from 61 channels)

0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 1 0 0 1 0 0 0

Since there were a total of 61 channels, up to 61 leaf nodes in the Huffman tree could be labeled randomly with the channels. This labeling would be stored for each data to be encrypted. If the Huffman tree had less than 61 leaf nodes, then a smaller number of channels equal to the number of leaf nodes would be used. In the study here, since we had 26 alphabets and four randomly introduced characters: \ ] ^ _ and space, we had a total of 31 symbols and hence 30 leaf nodes with corresponding 30 channels. The four randomly introduced characters and space serve to increase the encryption strength. The 30 channels were chosen randomly. These 30 channels, represented as bits 1 or 0, would be the encryption keys and will be used to determine if the corresponding leaf nodes should be swapped or not, respectively.

4. Data compression and encryption

Huffman tree would be generated based on the data and then the data would be encrypted with the shuffled Huffman tree. Only the encrypted data and original Huffman tree (with channel labels) would be stored for decryption. For decryption, the procedures up to obtaining the EEG encryption key would be the same. The encryption keys would be used to convert the original Huffman tree back to the shuffled Huffman tree using the labels inherent in the original Huffman tree and the data would be then decrypted. Figures 3 and 4 shows block diagrams of the encryption and decryption procedures.

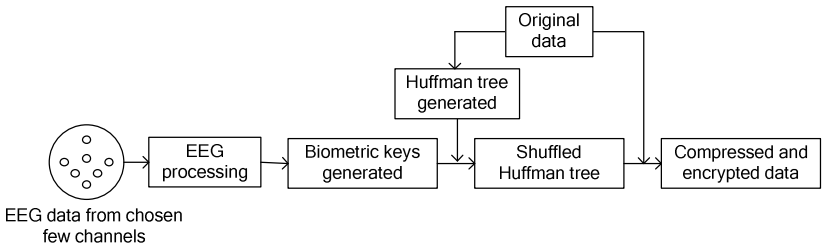

Figure 3. Encryption steps

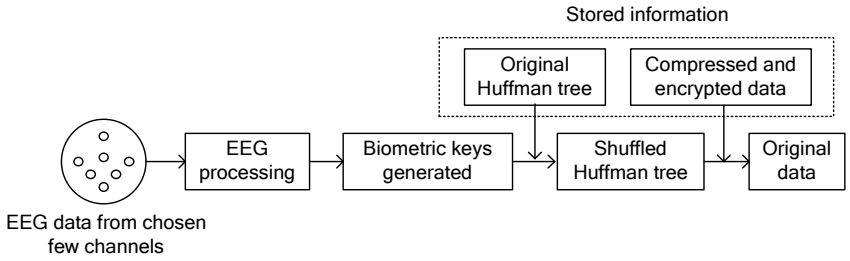

Figure 4. Decryption steps

5. Results and discussion

Table 2 shows the true positive (TP) decryption (correct decryption by the genuine subject) and true negative (TN) decryption (correct decryption by the impostor subjects). The results shown are TP and TN percentage values using the pangram (holalphabetic): 'The quick brown fox jumps over the lazy dog\]^_'. This sentence was encrypted with the proposed method and decrypted using 40 different event-related EEG extracted at different times from subject 1. The number of times that the sentence was correctly decrypted (exactly) was stored (this would represent TP value for subject 1). This was then repeated for this sentence but using 360 event-related EEG extracted at different times from subjects other than subject 1. Again, the number of times that the sentence was correctly decrypted was stored (this would represent TN value for subject 1). This procedure is then repeated for all the rest of the subjects. From Table 2, it could be easily seen that the TP ranges from 82.05-100% and the TN is below 27.22% for all the subjects with several subjects giving very small TN values.

Table 2. TP and TN results Encryption keys for a subject (from 61 channels)

Subject	TP(%)	TN(%)
1	84.62	19.44
2	92.31	23.61
3	100	23.06
4	82.05	27.22
5	89.74	1.94
6	92.31	1.67
7	82.05	0.28
8	92.31	2.22
9	97.44	23.06
10	87.18	23.61
Average	90.0±6.1	14.61±11.42

6. Conclusion

A novel method of data encryption using event-related EEG has been proposed. EEG from gamma band spectral range from specific channels was used to generate the biometric encryption key. This key was then used to shuffle leaf nodes in the Huffman tree generated from the data with Huffman coding, thereby changing the codewords that would be necessary during decoding. The complete key generation process takes slightly more than one second (therefore comparable to any other bio-encryption method) and combined with the simple but effective Huffman tree shuffling/de-shuffling, the whole procedure is fast enough for system implementation. Experimental study with biometric encryption key generated by 10 subjects gave good TP rates over 40 trials conducted at different times. Though the TN rates are somewhat high for system implementation, this is a pilot study and we believe that further work will improve the situation. The possible sources of error are that the subject does not concentrate on looking at the picture and due to noise in recording the EEG data. With this lower error rate, the proposed data encryption using EEG is especially advantageous over other biometrics because of its robustness to withstand fraudulent attacks. Further, data is not only encrypted but compressed as well. The only major disadvantage of the method lies in the cumbersome EEG data recording using wet electrodes but the current significant advances in dry electrode design would allow EEG to be recorded using a simple cap/hat. We hope that this study will stimulate and encourage further exploration on this rather neglected but promising biometric cryptosystem.

Acknowledgement

The authors thank the late Prof. Henri Begleiter at the Neurodynamics Laboratory at the State University of New York Health Centre at Brooklyn, USA who generated the raw VEP data and Mr. Paul Conlon, of Sasco Hill Research, USA for sending the data.

References

[1] Uludag, U., Pankanti, S., Prabhakar, S., and Jain A.K. “Biometric cryptosystems: Issues and challenges,” Proceedings of the IEEE, vol. 92, no.6, pp.948-960, June 2004.

[2] Ravi, K.V.R., and Palaniappan, R. “Neural network classification of late gamma band electroencephalogram features,” Soft Computing, vol. 10, no.2, pp. 163-169, 2006.

[3] Hao, F., Anderson, R., and Daugman, J. “Combining cryptography with biometrics effectively,” Computer Laboratory Technical Report, University of Cambridge, no. 640, July 2005.

[4] Wang, C.E, “Cryptography in Data Compression”, CodeBreakers Journal, vol. 1, no. 1, 2006. Available:http://www.codebreakers-ournal.com/index.php/CodeBreakersJournal/article/viewFile/9/10