INFORMATICA, 2000, Vol. 11, No. 4, 1

A New Redundancy Reducing Cipher

Min-Shiang HWANG
Department of Information Management, Chaoyang University of Technology
168, Gifeng E. Rd., Wufeng, Taichung County, Taiwan 413, R.O.C.
e-mail: mshwang@mail.cyut.edu.tw

Received: June 2000

Abstract. This paper discusses a known-plaintext attack on a redundancy reducing cipher method which is proposed by Wayner. We also propose an extension of Wayner's redundancy reducing cipher scheme so that the security will be improved greatly.

Key words: cryptography, data encryption standard (DES), Huffman coding.

1. Introduction

In order to protect encrypted data against statistical analysis, Shannon suggested that some of the redundancy of the language be removed before encryption (Huffman, 1952; Schneier, 1996). Data compression is one method for removing redundancy before encryption (Amsterdam, 1986).

Wayner (Wayner, 1988) proposed an efficient encrypting compression algorithm that is based on the well-known Huffman coding. He modified the Huffman coding so as to permit encryption to accompany compression. This approach reduces the redundancy of a file to near the theoretical minimum. Wayner intended to encipher and compress the plaintext simultaneously. His method does not require other encryption algorithms (i.e., the Data Encryption Standard (Schneier, 1996)) as a second pass to enhance security. In this paper, however, we show that his method cannot withstand a known-plaintext attack when the attacker knows a few pairs of (plaintext, ciphertext). We also propose a redundancy reducing cipher for extending Wayner's enciphering scheme so that the security will be improved greatly.

2. The Weakness of Wayner's Scheme

In Wayner's scheme, he pre-built a standard binary tree (Huffman tree) depending on the relative frequency of every character in the text. Since the standard tree is fixed in the system, the cryptanalyst can discover the tree by analyzing many different messages. To enhance the security, Wayner used the standard tree with a key to control how the variable length codes are determined. Instead of each branch in the tree being fixed as zero or one (binary), the numbers that control the branching are determined by the bits of the key.

Wayner proposed two methods for assigning a key to a tree. Method 1 assigns one bit for each level of a tree. Method 2 assigns one bit for each node of a tree. This bit is operated XOR (exclusive-or) to the addresses for each node on that level. Figs. 2, 2, and 2 illuminate the above mentioned method.

As just stated, the standard tree can be discovered by analyzing many different messages. The key of Method 1 in Wayner's scheme (Wayner, 1988) is easily discovered when a character that is in the longest path of a tree and its corresponding cipher-code is known. For example, if we know the standard tree (Fig. 2) and the cipher-code "0001" of "H" or "0000" of "A", the tree (Fig. 2) associated with the key "1001" can be derived.

The key of Method 2 in Wayner's scheme (Wayner, 1988) is also easily derived when a few characters and their corresponding cipher-codes are known. We let x be the number of nodes which have two leaves. We thus need to know only x pairs of (character, cipher-code), where the character is one of the leaf-characters of the x nodes. Since the tree is in binary, x is less than $n/2$, where n is the number of leaves in a tree. For example, whenever we know the standard tree (Fig. 2) and the cipher-code "0011" of "H" and "11" of "T", the tree (Fig. 2) associated with the key "10011" can be derived. There are six characters in the tree, but we need only to know two characters T (or R) and H (or A) and their corresponding cipher-code to derive the private key "10011".

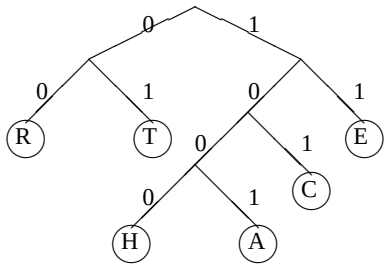

Fig. 1. A standard tree that codes "R" as "00", "T" as "01", "E" as "11", "C" as "101", "A" as "1001", and "H" as "1000".

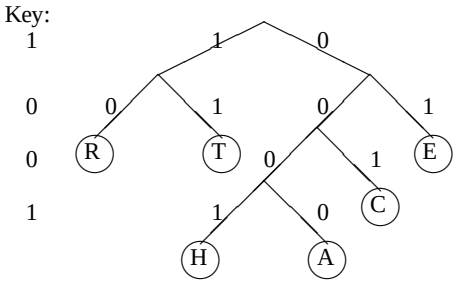

Fig. 2. An example of Method 1, where the key "1001" modifies the tree so "T" is coded as "11", "A" as "0000", etc.

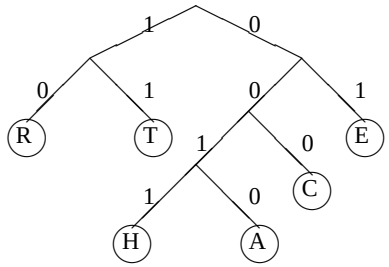

Fig. 3. An example of Method 2, where the key "10011" modifies the tree so "T" is coded as "11", ""E" as "01", "H" as "0011", etc.

3. Our Scheme

Shannon proposed two encryption techniques to protect against attacks based on statistical analysis: confusion and diffusion (Chang, 2000; Hwang, 1999; Schneier, 1996). Since Wayner only confuses the plaintext and does not include any process of diffusion, his approach is not secure when the analyst has knowledge of pairs of plaintext and corresponding cipher-code. To enhance the security, we suggest that the following two methods be combined with Wayner's method: confusion by operating an exclusive-or private key and Huffman codes, and diffusion by permuting the Huffman codes with a fixed period. In other words, the point of using Huffman coding or another data compression method is only to remove redundancy.

From the above discussion, we see that Wayner's scheme cannot withstand a known-plaintext attack when the attacker knows a few pairs of (plaintext, ciphertext). We now give an extended scheme which improves Wayner's enciphering scheme for withstanding this type of attack. We also pre-build a standard binary tree (Huffman tree) depending on the relative frequency of every characher in the text. We also use Wayner's two methods for assigning a secret key to a tree except for that the key is a variable but not constant. In other words, all characters are encrypted with a different secret key. We assume that Alice wants to send an encrypted message to Bob. The enciphering procedure is as follows.

1. Alice and Bob agree on a binary tree.
2. Alice generates a secret key (K_1) in length of levels of the tree.
3. Alice distributes the secret key to Bob by secret channels.
4. Alice encrypts the first characher (H_1) with K_1 by using the same method as Wayner's scheme.
5. Alice changes the secret key by computing ($K_2 = K_1 \oplus H_1$). If the length of H_1 is less than that of K_1, append binary "0" to H_1 such that the length of H_1 is equal to that of K_1.

6. Alice repeats the above two steps to encrypt the second, third, $\cdots$, and the last character and change the secret key by computing $K_{i+1} = K_i \oplus H_i$.
7. Alice sends the encrypted message to Bob by public channels.

Whenever Bob receives the encrypted message (C), he decrypts it as follows.

1. Bob decrypts the first characher H_1 from C with K_1 by using the same method as Wayner's scheme.
2. Bob changes the secret key by computing ($K_2 = K_1 \oplus H_1$). If the length of H_1 is less than that of K_1, append binary "0" to H_1 such that the length of H_1 is equal to that of K_1.
3. Bob repeats the above two steps to decrypt the second, third, $\cdots$, and the last character and change the secret key by computing $K_{i+1} = K_i \oplus H_i$.

An example of our scheme to encrypt the message "TEACHER" with a standard tree (in Fig. 2) and a secret key "1001" is shown in Table 1. Therefore, the encrypted message of "TEACHER" is "11 00 1000 001 1010 01 01".

Table 1
An example of our scheme

i	K_i	H_i	C_i
1	1001	T(01)	11
2	1101	E(11)	00
3	0001	A(1001)	1000
4	1000	C(101)	001
5	0010	H(1000)	1010
6	1010	E(11)	01
7	0110	R(00)	01

The above enciphering and deciphering procedures are extended from Method 1 in Wayner's scheme. Method 2 in Wayner's scheme also can apply to the above procedures just by modifying the step of changing secret key.

Our extended scheme not only retains the advantages of Wayner's scheme but also enhances the security.

4. Conclusions

We have shown that Wayner's scheme cannot withstand a known-plaintext attack when an attacker knows a few pairs of (plaintext, ciphertext). We also have proposed an extended scheme which is a slight modification of the Wayner's scheme. The proposed scheme not only retains the advantages of Wayner's scheme which reduces the redundancy of a file to near the theoretical minimum, but also enhances the security.

Acknowledgements

The authors wish to thank many anonymous referees for their suggestions to improve this paper. Part of this research was supported by the National Science Council, Twain, R.O.C., under contract no. NSC89-2213-E-324-053.

References

Amsterdam, J. (1986). Data compression with huffman coding. *Byte*, **11**, 98.
Chang, C.C., M.S. Hwang and T.S. Chen (1988). A new encryption algorithm for image cryptosystems. *Journal of Systems and Software* (to appear).
Huffman, D. (1952). A method for the construction of minimum redundancy codes. *Proc. Inst. Radio Engrs*, **40**, 1098–1101.
Hwang, M.S. (1999). A new dynamic cryptographic key generation scheme in a hierarchy. *Nordic Journal of Computing*, **6**, 363–371.
Scheier, B. (1996). *Applied Cryptography*, 2nd Edition. John Wiley & Sons.
Wayner, P. (1988). A redundancy reducing cipher. *Cryptologia*, **12**, 107–112.

M.-S. Hwang received the B.S. in Electronic Engineering from National Taipei Institute of Technology, Taipei, Taiwan, Republic of China, in 1980; the M.S. in Industrial Engineering from National Tsing Hua University, Taiwan, in 1988; and the Ph.D. in Computer and Information Science from National Chiao Tung University, Taiwan, in 1995. He also studied Applied Mathematics at National Cheng Kung University, Taiwan, from 1984-1986. Dr. Hwang passed the National Higher Examination in field "Electronic Engineer" in 1988. He also passed the National Telecommunication Special Examination in field "Information Engineering", qualified as advanced technician the first class in 1990. From 1988 to 1991, he was the leader of the Computer Center at Telecommunication Laboratories (TL), Ministry of Transportation and Communications, ROC. He was also a project leader for research in computer security at TL in July 1990. He obtained the 1997, 1998, and 1999 Distinguished Research Awards of the National Science Council of the Republic of China. He is currently a professor and chairman of the Department of Information Management, Chaoyang University of Technology, Taiwan, ROC. He is a member of IEEE, ACM, and Chinese Information Security Association. His current research interests include database and data security, cryptography, image compression, and mobile communications.