ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 136

4th International Conference on Sensors, Mechatronics and Automation (ICSMA 2016)

A Novel Identification Approach to Encryption Mode of Block Cipher

Cheng Tan[1, a], Yifu Li [2,b] and Shan Yao[*2,c]

[1]Science and Technology on Communication Security Laboratory, Chengdu, Sichuan Province, China

[2] National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China

[a]ctan2007@163.com, [b]liyf@cert.org.cn, [c]yaoshan@cert.org.cn

Keywords: identification of encryption mode; encryption mode; block cipher; SVM technology

Abstract. Encryption mode describes how a block cipher operates in a cryptosystem and the implementation detail of cryptographic algorithm. In general, the encryption mode of a block cipher is unknown to a cryptanalyst. It plays a significant role in cryptanalysis work on block cipher to identify the encryption mode. In this paper, we propose a novel identification approach to encryption mode of block cipher only with ciphertext information available. Based on the idea of SVM technology, we construct an identification system of encryption mode of block cipher. We can identify the encryption mode with a high identification rate when the key and IV of testing ciphertext files are same as those of training ciphertext files. Even if the keys or IVs of training and testing ciphertext files are different, the identification result is still superior to that of randomly guessing. Furthermore, we find that it is more difficult to identify the encryption mode of AES than DES, Blowfish, 3DES and RC5.

Introduction

Cryptography is widely used for protecting people's private information from been obtained by middle attackers, which is the kernel of information security. Cryptanalysis is a subject for evaluating the security level of a cryptosystem. The main idea of cryptanalysis is that a cryptanalyst infers the plaintext or the key from ciphertext.

Due to some advantages, such as simple program, good diffusivity and so on, block cipher has been widely applied in area of information security. In this case, cryptanalysis on block cipher is always a hot spot of research. According to Kerckhoffs's assumption on cryptanalysis, a cryptanalyst has known which cryptographic algorithm is used and the implementation detail of cryptographic algorithm. For block cipher, the implementation detail of cryptographic algorithm is the encryption mode. Only when cryptographic algorithm and encryption mode are known to a cryptanalyst can he/she conduct cryptanalysis work more efficiently.

Some ways to identifying cryptographic algorithm of block cipher from ciphertext have been proposed in some previous work. Dileep and Sekhar have used bag-of-words approach for representing a ciphertext file by a document vector, and then the work of identifying cryptographic algorithm from ciphertext is converted into a task of document categorization [1]. S. Mishra and A. Bhattacharjya have combined block length/stream detection, entropy/ reoccurrence analysis and dictionary and decision tree based approach to propose a combined approach to analyze the pattern of ciphertext for AES, DES and Blowfish [2]. J. Chopra and S. Satav have used Naïve Bayesian and K-Nearest neighbor algorithms to classify data sets encrypted by AES, 3DES, and Rijndael [3]. And then the performance results show that there exist distinguished patterns in encrypted data. Sharif and Mansoor have considered eight kinds of pattern recognition techniques to identify DES, IDEA, AES, and RC2, and the simulation results show that using one key provides better classification than using different keys [4]. Besides, more ciphertext files will increase the identification rate. The identification study of the finalist block algorithms of AES contest has been done in some other work [5-7], in which genetic algorithm, graph theory, clustering algorithm, neural network algorithm and so on are applied.

In our previous work, an identification system of AES, Blowfish, 3DES, RC5 and DES based on support vector machine (SVM) classifier is established [8]. We considered 3 different situations, that

is, same key for training and testing ciphertext, different keys for training and testing ciphertext and one to one identification. In conclusion, the identification of these 5 block ciphers is easy to operate if keys are the same for training and testing ciphertext. Besides, we can still obtain a high identification rate for one to one identification associated with AES even if keys are different for training and testing ciphertext.

Now people can succeed in identifying some common block cipher through some current methods. The encryption mode describes how a block cipher operates in a cryptosystem, which means that the ciphertext also reflects the character of encryption mode of block cipher. Even if a cryptanalyst is aware of the cryptographic algorithm, it is a tough job for him/her to continue to make cryptanalysis without knowing the encryption mode operated by the block cipher. As far as we know, no public research about identifying encryption mode of block cipher is available.

Hence we propose a novel identification approach to encryption mode of block cipher in this paper. We concentrate on 4 kinds of encryption mode, namely, Electronic Code Book (ECB), Cipher Block Chaining (CBC), Cipher Feedback (CFB) and Output Feedback (OFB). Based on previous work [8], we continue to consider these 5 common block ciphers, namely, DES, AES, Blowfish, 3DES and RC5.

System Model

In this paper, we identify the encryption mode of block cipher with only ciphertext information available. Based on the idea of machine learning technique, we construct an identification system of encryption mode of block cipher.

Fig. 1 has showed the producing of training and testing ciphertext files. For a block cipher, we produce m classes of ciphertext files with respect to m kinds of encryption modes. For each class of ciphertext files, we divide them into two subclasses, namely A and B. The ciphertext files in each A are used for training the identification model, which are called as training ciphertext files. The ciphertext files in each B are divided into n groups averagely, and then we integrate each single group responding to each encryption mode to n groups of testing ciphertext files.

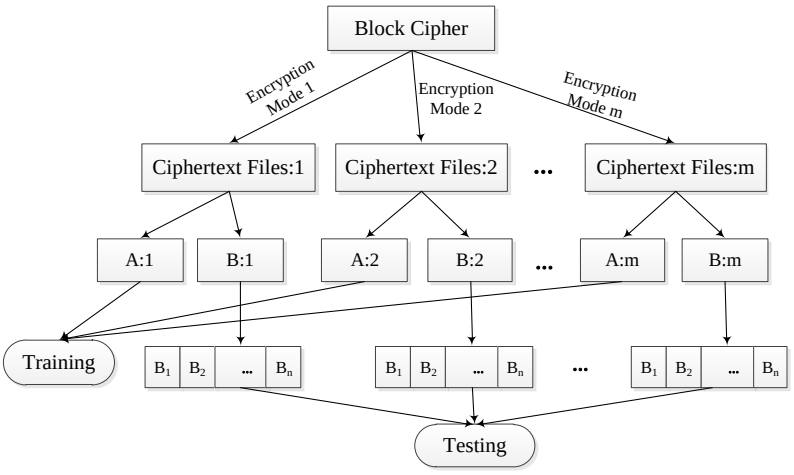

Figure 1. The producing of training and testing ciphertext files.

In Fig. 2, we present the complete implementation architecture of our identification system of encryption mode. For one block cipher, we produce some ciphertext files, consisting of training and testing ciphertext files. We set up a feature database integrating encryption modes of the block cipher through extracting the features of training ciphertext files. Combined with SVM algorithm, we construct an identification model for encryption mode. Extract the features from testing ciphertext files, and input them into the identification model. The work of identifying encryption mode from ciphertext files is mainly done by M SVM classifiers, and the counter calculates out classification results of these classifiers to identify the final encryption mode. Note that if there are m kinds of encryption modes to be identified, then M=m(m-1)/2. Compare these identification results of testing ciphertext files with the real encryption mode, and we obtain the identification rate of a group of

testing ciphertext files. Set a threshold value, and if the identification rate is less than the threshold, modify the rule for extracting ciphertext feature. Reidentify the encryption mode of all groups of ciphertext files until the rule for extracting ciphertext feature leads to an identification rate larger than the threshold. It is easily seen that the larger the threshold is set, the higher the possibility of modifying the rule is. Maybe one rule is more suitable for some groups of ciphertext files under a large threshold, but less suitable for other groups. Besides, we will probably exclude all available rules for extracting ciphertext feature with a large threshold. In this case, we have to set a suitable threshold which is not very large. We finally construct a complete identification system of encryption mode through adding feature databases of some other block ciphers.

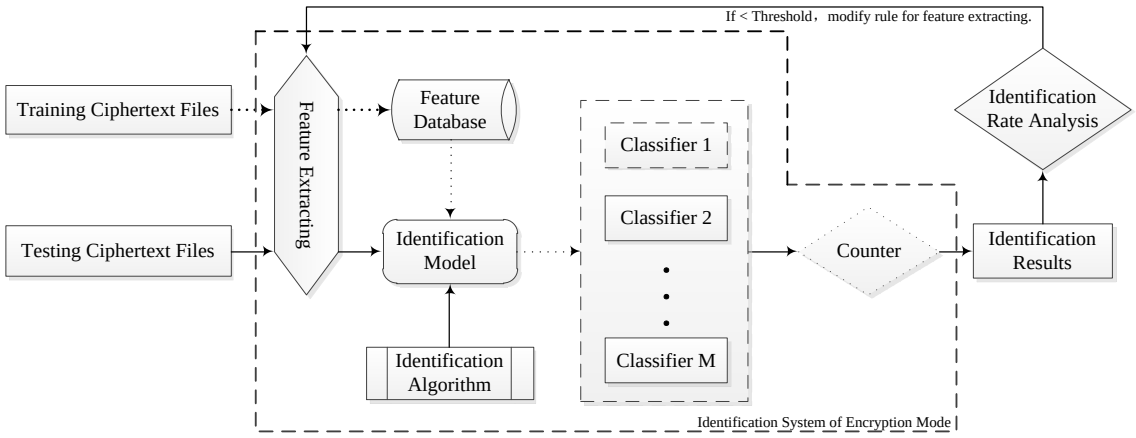

Figure 2. Implementation architecture of our identification system of encryption mode.

Encryption Mode Identification and Performance Analysis

We in this paper consider 5 block ciphers, DES, AES, Blowfish, 3DES and RC5, for encryption mode identification. The block length of AES is 128 bit, while the block length of other 4 ciphers is 64 bit. Hence the length of the initialization vector (IV) for AES is 128 bit, and 64 bit for other 4 ciphers. As is known to all, the length of key for DES is 64 bit. For simplicity, the length of key for AES, Blowfish, 3DES and RC5 is set to be 128 bit in the following experiments uniformly.

Note that we have constructed an identification system of encryption mode, in which the feature databases and rule for extracting ciphertext feature stay fixed. Next we will use the identification system of encryption mode to conduct some experiments.

For each block cipher, we produce 180 extra testing ciphertext files under each encryption mode. These testing ciphertext files are equally divided into 9 groups, that is to say, 20 testing ciphertext files under one encryption mode are to be identified in each of 9 tests. Next we will give experiment results and performance analysis in different situations.

Situation 1

In this situation, the key of testing ciphertext files is same as that of training ciphertext files. Additionally, the IV of testing ciphertext files keeps the same with that of training ciphertext files under CBC, CFB and OFB mode.

Table 1. Identification results among ECB, CBC, CFB and OFB (Same key and IV, in %)

Cipher \ Size (KB)	500	100	20	4
DES	77.64(9.75)	79.31(8.99)	77.92(9.80)	67.5(13.78)
AES	68.19(13.65)	66.39(15.31)	61.81(17.09)	58.33(17.56)
Blowfish	78.06(9.31)	79.72(8.38)	77.92(9.94)	65.97(15.18)
3DES	80.97(8.38)	78.89(9.34)	75.69(9.75)	67.22(14.47)
RC5	79.17(9.58)	78.75(8.71)	75.69(10.44)	65.28(14.83)

Table 1 shows the identification results of encryption mode among ECB, CBC, CFB and OFB. We have considered 4 different sizes of ciphertext file for DES, AES, Blowfish, 3DES and RC5, namely 500 KB, 100 KB, 20 KB and 4KB. The data like A(B) in the table represents that A% is the average identification rate in 9 tests, and B% is the standard deviation value of 9 identification rates.

According to Table 1, we find that if the size of ciphertext file is larger than 20 KB, the identification rate can reach a relatively high value. These identification rates for ciphertext file with a size of 500 KB, 100KB and 20 KB don't seem to be much different from each other, and the identification rate decreases significantly when the size of ciphertext file turns to 4 KB. Compared to other 4 block ciphers, the identification rate of encryption mode for AES is lower, and the standard deviation value appears to be larger. The block length of AES is larger than other 4 block ciphers, which leads to a fact that there exist little prominent feature in AES ciphertext file even under ECB mode. As a consequence, the encryption mode of AES is harder to identify than other 4 block ciphers.

Firstly, ECB is the simplest of all existing encryption modes. Secondly, the key streams of training and testing ciphertext files stay the same for the reason that keys and IVs of training and testing ciphertext files are the same under OFB mode. Hence the identification of encryption mode under ECB and OFB mode is easier than that under CBC and CFB mode. Next we will consider CBC and CFB mode separately to verify our analysis.

Situation 2

In situation 2, the key and IV of testing ciphertext files are still same as those of training ciphertext files. Differently, we study on the identification of encryption mode under a combination of ECB, CBC/CFB and OFB mode. For simplicity, we only consider DES and AES from now on.

Table 2. Identification results among ECB, CBC and OFB (Same key and IV, in %)

Cipher \ Size (KB)	500	100	20	4
DES	96.11(3.00)	99.26(0.88)	96.48(2.94)	77.22(13.92)
AES	82.78(10.77)	80.55(11.12)	80.19(10.91)	66.67(13.64)

Table 3. Identification results among ECB, CFB and OFB (Same key and IV, in %)

Cipher \ Size (KB)	500	100	20	4
DES	98.15(1.76)	98.52(1.76)	96.67(3.12)	77.78(11.73)
AES	82.04(11.54)	81.67(10.38)	78.33(11.12)	67.41(14.07)

From Table 2 and Table 3, we see that the identification rates of encryption mode are much higher than those in Table I. Besides, lower standard deviation values represent that the identification results in Table 2 and Table 3 are more reliable. However, the standard deviation values of at least 10% for AES tell that it is still hard work to identify the encryption mode of AES with a stable identification rate. As for DES, the standard deviation values are so low that we can always identify the encryption mode of DES with a high identification rate if the size of ciphertext file is larger than 4 KB. According to Table 1, Table 2 and Table 3, we can infer that it is very hard to distinguish the feature

of ciphertext file under CBC mode from that under CFB mode. In this case, the identification results turn to be better when we consider CBC and CFB mode separately.

Situation 3

In situation 3, we continue to identify 4 kinds of modes, ECB, CBC, CFB and OFB. The key of testing ciphertext files is same as that of training ciphertext files, while the IV of testing ciphertext files is different with that of training ciphertext files.

Table 4. Identification results among ECB, CBC, CFB and OFB (Same key, different IVs, in %)

Size (KB) / Cipher	500	100	20	4
DES	50.83(4.33)	49.31(2.73)	51.11(3.77)	39.31(7.71)
AES	35.42(8.88)	35.14(6.11)	36.94(7.73)	29.17(8.71)

Table 4 has given the identification results of encryption mode in situation 3. For DES, the identification rate is about 50% when the size of ciphertext file is larger than 4 KB, which reflects that the change of IV leads to a decrease on identification rate. When the IV of testing ciphertext files is different with that of training ciphertext files, the ciphertext files of OFB mode are hard to be identified. For AES, the identification rate is about 35% when the size of ciphertext file is larger than 4 KB, which is still higher than a rate of 25% for randomly guessing.

Situation 4

In situation 4, we further try to consider different keys and IVs for training and testing ciphertext files. The identification results of encryption mode are showed in Table 5.

Table 5. Identification results among ECB, CBC, CFB and OFB (Different keys and IVs, in %)

Size (KB) / Cipher	500	100	20	4
DES	48.19(2.87)	50.14(4.44)	46.81(4.20)	34.86(7.22)
AES	32.22(7.01)	34.03(6.02)	33.75(5.56)	24.17(5.23)

Compare Table 4 with Table 5, if the IVs of training and testing ciphertext files are different, the condition that the keys of training and testing ciphertext files are the same or not only causes a tiny effect on the final identification rate.

Summary

Through analyzing the results in 4 situations, we find that if the keys and IVs of training and testing ciphertext files are the same, we can identify the encryption mode from ciphertext with a high identification rate, especially for identification among 3 kinds of encryption modes, thus ECB, CBC/CFB and OFB mode. However, the difference of the keys or IVs of training and testing ciphertext files causes a violent decreasing of the identification rate.

Conclusions

It is a novel idea to identify the encryption mode of block cipher from ciphertext. Actually, it is necessary for a cryptanalyst to be aware of the cryptographic algorithm and encryption mode if he/she decides to make cryptanalysis on a block cipher. To our best knowledge, there exists no public research about identification methods on encryption mode of block cipher, although a number of ways of identifying cryptographic algorithm have been proposed by some researchers. In this case, we introduce an approach to identify the encryption mode of block cipher only with ciphertext information. When the key and IV of training ciphertext files are same as those of testing ciphertext files, the identification results perform to our satisfaction. The only drawback is that our identification system of encryption mode behaves not very well when different keys and IVs for training and testing ciphertext files, which needs to be improved in future work. Moreover, we will continue to study on identifying of more other encryption modes of block cipher.

Acknowledgements

This work was financially supported by Foundation of Science and Technology on Communication Security Laboratory (9140C110602150C11053).

References

[1] A. D. Dileep and C. C. Sekhar, "Identification of block ciphers using support vector machines," Proc. International Joint Conference on Neural Networks. Vancouver, BC, Canada, 2006, pp. 2696-2701.

[2] S. Mishra and A. Bhattacharjya, "Pattern analysis of cipher text: A combined approach," Proc. International Conference on Recent Trends in Information Technology, 2013, pp. 393-398.

[3] J. Chopra and S. Satav, "Impact of encryption techniques on classification algorithm for privacy preservation of data," International Journal of Innovative Research in Science, Engineering and Technology, vol. 2, no. 10, pp. 5398-5402, October 2013.

[4] S. O. Sharif and S. P. Mansoor, "Performance evaluation of classifiers used for identification of encryption algorithms," ACEEE Int. J. on Network Security, vol. 2, no. 4, pp. 42-45, Oct. 2011.

[5] R. H. Torres, G. A. Oliveira, et al. Identification of Keys and Cryptographic Algorithms Using Genetic Algorithm and Graph Theory[J]. IEEE LATIN AMERICA TRANSACTIONS, VOL. 9, NO. 2, APRIL 2011:178-183.

[6] W. A. R. de Souza, L. A. V. de Carvalho and J. A. M. Xexéo. Identification of N Block Ciphers. IEEE LATIN AMERICA TRANSACTIONS, VOL. 9, NO. 2, APRIL 2011:184-191.

[7] Vina M. Lomte, Archana D. Shinde. Review of a New Distinguishing Attack Using Block Cipher with a Neural Network. International Journal of Science and Research, VOL. 3, No. 8, August 2014:733-736.

[8] C. Tan and Q. Y. Ji, "An approach to identifying crypto-graphic algorithm from ciphertext," Proc. International Conference on Communication Software and Networks, Beijing, China, 2016, pp. 19-23.