Taylor & Francis
Taylor & Francis Group

Journal of Discrete Mathematical Sciences and Cryptography

ISSN: 0972-0529 (Print) 2169-0065 (Online) Journal homepage: http://www.tandfonline.com/loi/tdmc20

Low information leakage random padding scheme for block encryption

Chuan-Chi Wang , Min-Chih Kao & Yi-Shiung Yeh

To cite this article: Chuan-Chi Wang , Min-Chih Kao & Yi-Shiung Yeh (2008) Low information leakage random padding scheme for block encryption, Journal of Discrete Mathematical Sciences and Cryptography, 11:4, 385-391, DOI: 10.1080/09720529.2008.10698191

To link to this article: http://dx.doi.org/10.1080/09720529.2008.10698191

Low information leakage random padding scheme for block encryption

Chuan-Chi Wang [*]

Department of Computer Science and Information Engineering
Ching-Yun University
229, Chien-Hsin Rd., Jung-Li
Taiwan 320
R.O.C.

Min-Chih Kao [†]

Yi-Shiung Yeh [‡]

Department of Computer Science and Information Engineering
National Chiao-Tung University
1001, Ta-Hsueh Road
Hsinchu 30050
Taiwan
R.O.C.

Abstract

We propose a new random padding scheme for symmetric key block encryption. In the padding scheme, a padding string is key-dependent and almost random. Thus, the padding string causes extreme low information leakage to the adversary with ciphertext-only manner. The intention to collect plaintext-ciphertext pairs relating to the underlying secret key from padding strings becomes very difficult. We also show that with the padding scheme the simple CBC encryption mode becomes strong to defeat the padding oracle attacks.

Keywords and phrases : *Padding, block cipher, encryption mode, padding oracle attack.*

---

[*] *E-mail*: `wcc@cvu.edu.tw`

[†] *E-mail*: `gau.csie91g@csie.nctu.edu.tw`

[‡] *E-mail*: `ysyeh)@csie.nctu.edu.tw`

Journal of Discrete Mathematical Sciences & Cryptography
Vol. 11 (2008), No. 4, pp. 385–391

1. Introduction

Block cipher encryptions, such as AES [2] in encryption modes of simple *Electronic Code Book* (ECB) and *Cipher Block Chaining* (CBC) [3], require their input to be a multiple of the block size. Otherwise, a padding string will be appended to the plaintext to make it fitting the requirement. The padding string should be removed unambiguously at the time of decryption.

Several padding schemes are used as conventions. Some of them use simple and constant padding string. We call them the *Constant-Padding* (CP) schemes for instance as follows.

(1) CP 1: Pad with zero characters.
(2) CP2: Pad with zero characters and fill the last byte with the number of padding characters.
(3) CP3: Pad with bytes of the same value as the number of padding bytes. The method was recommended in PKCS#5 [7] and RFC2630 [8].
(4) CP4: Pad with 0x80 followed by zero bytes.

Constant-Padding schemes are easy for implementation but with the drawback of leaking vast of information about the plaintext of the padding string. The leaked information supports the adversary with high advantage to collect pairs of plaintext and ciphertext relating to the underling secret key. Besides, constant padding string is favorable for the padding oracle attacks proposed in [9] as discussed in Section 4. The strategy to reduce the possible information leakage is by random padding. That is, random values are included in the padding string. There are two *Random-Padding* (RP) schemes as below.

(1) RP1: Pad with randomly selected characters and fill the last byte with the number of padding bytes.
(2) RP2: Pad with randomly selected characters X and Y by the form of XY^n where $X \neq Y$ [4].

In this paper, we propose a new random padding scheme which causes very low information leakage about the padding string. The rest of paper is organized as follows. In the next section, we describe the information leakage of a padding string. In Section 3, we describe the proposed new random padding scheme. In Section 4, we discuss the security of the proposed scheme against the padding oracle attacks. Finally, some conclusions are given in Section 5.

2. Information leakage of a padding string

A block cipher with good pseudo random property makes an adversary, who doesn't know the secret key, difficult to guess the corresponding plaintext for a given ciphertext. Ideally, all possible plaintexts are distributed uniformly. That is, given n bits ciphertext, the success probability to guess the corresponding n bits plaintext would be 2^{-n}. However, when the plaintext distribution space is not uniform, the adversary may get higher advantage to guess the plaintext. In the extreme case that the plaintext is a known constant, the probability to guess the plaintext is obviously 1. The information of plaintext is useful for analyzing the underling block cipher. Even the block cipher is strong to defeat known-plaintext attack, it would be better to hide all information of plaintext for reducing the probability of attack that is unknown currently. Here, we define the information leakage of the plaintext corresponding to a given ciphertext. The definition can be used to evaluate the information leakage of a padding string. For simplicity, the plaintext and ciphertext mentioned below, and at the rest of the paper, are assumed to be computed by an ideal block cipher.

Definition 1 (Entropy [6]). Let X be a random variable which takes on a finite set of values x_i, with $1 \leqq i \leqq n$, and has probability distribution $p_i = p(X - x_i)$. The entropy of X is:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \,. \tag{1}$$

Definition 2 (Information leakage). The random variable X, the same as in Definition 1, has information leakage, said $L(X)$, as follows. ($|X|$ stands for the average length of X)

$$L(X) = |X| - H(X) = \log_2 n + \sum_{i=1}^{n} p_i \log_2 p_i \,. \tag{2}$$

Definition 3 (Information leakage of a plaintext). Given an s-bit ciphertext C, the corresponding plaintext P is an element in the set $\{0, \ldots, 2^s - 1\}$ as the secret key is chosen at random. Let X be a random variable on the set $\{x_i \,|\, 0 \leqq x_i \leqq 2^s - 1\}$ with the probability distribution $p_i = p(X = x_i) =$ (the probability that x_i is the plaintext of C from the viewpoint of an adversary A). The information leakage of P to the adversary A, denoted as $L_A(P) = L(X)$.

According to the above definitions, we can get the following results. To guess an s-bit plaintext P for which all the 2^s possible elements have the same probability, i.e., $1/2^s$, from the viewpoint of the adversary, the information leakage of P is 0. However, if the plaintext is determined, the information leakage is s. So that it is reasonable to believe that the plaintext with low information leakage is more difficult to be guessed than the one with high information leakage, assuming that the two plaintexts have equal length.

When estimating the information leakage of a padding string, we consider the situation that length of padding string is known by the adversary. This is reasonable in the case that the adversary is able to observe information of a communication system. For example, an encrypted message may be decrypted and then be saved in a file. Although the adversary is not authorized to open this file, he/she may be able to read its attributes. Thus the attribute of file size may leak the length of the plain message to the adversary. Then, the length of the padding string can be derived out. According to the above definitions of information leakage, the known padding schemes mentioned in Section 1 have information leakage as listed in Table 1. For simplicity, we calculate in the case that the block cipher has 64-bit data block, the padding is byte-oriented, and the encryption mode is simple ECB mode.

Table 1

Information leakage of known padding schemes

$L_A(P)$	CP1	CP2	CP3	CP4	RP1	RP2
(pad = 1 byte)	8	8	8	8	8	NA
(pad = 2 bytes)	16	16	16	16	8	~0
(pad = 3 bytes)	24	24	24	24	8	8
(pad = 4 bytes)	32	32	32	32	8	16
(pad = 5 bytes)	40	40	40	40	8	24
(pad = 6 bytes)	48	48	48	48	8	32
(pad = 7 bytes)	56	56	56	56	8	40
(pad = 8 bytes)	64	64	64	64	8	48

3. The proposed random padding scheme

The proposed padding scheme uses a secret value as a mark word to construct a padding string. Let M be the mark word. A padding string PS is constructed as follows. (The notation $\|$ denotes concatenation.)

$$PS = M\|r_1\|r_2\|r_3\|\ldots, \text{ where } r_i \text{ is a random word and } r_i \neq M. \quad (3)$$

The mark word M is used as a distinguishable symbol for unambiguously removing the padding string from the plaintext. It can be an extended part of the secret key shared by the message sender and receiver. Let the length of a word be w bits. The padding string $PS = M\|r_1\|r_2\|\ldots\|r_t$ has information leakage, according to Definition 3, as follows.

$$L_A(PS) = t(w - \log_2(2^w - 1)). \quad (4)$$

Consider the same arguments as described in Section 2. That is, the word size is a byte, and the block size is 64 bits. The information leakage of the proposed random padding scheme is shown in Table 2.

Table 2

Information leakage of the proposed random padding scheme

$L_A(PS)$	Pad = 1 byte	Pad = 2 bytes	Pad = 3 bytes	Pad = 4 bytes	Pad = 5 bytes	Pad = 6 bytes	Pad = 7 bytes	Pad = 8 bytes
	0	0.0056	0.0113	0.0169	0.0226	0.0282	0.0339	0.0452

To the best of our knowledge, the proposed scheme has smallest information leakage as shown in Table 2 under the situation that length of padding string is known.

4. The security against the padding oracle attacks

Vaudenay [9] introduced the notion of padding oracle attacks on CBC mode encryption with CBC-PAD padding scheme. This attack assumes that a padding oracle which receives a ciphertext and then answers whether or not the corresponding plaintext is correctly padded. By querying to the oracle with appropriate modified initial vector and the ciphertext, an attacker is possible to invert the underlying block cipher. Several CBC mode encryptions in well-known products and standards are shown potentially vulnerable to this attack [1, 5].

Specifically, the attacker sends a CBC mode ciphertext block (C) including the *initial vector* (IV) to the padding oracle. If the answer indicates that the padding string P is correct. Consider the formula: $IV \oplus D_k(C) = P$ where $D_k(\,)$ is the decryption function. That is, $D_k(C) = IV \oplus P$. Since IV is known by the attacker. Thus, if the attacker also knows the padding string P, then the plaintext of C $(= D_k(C))$ is discovered. Therefore, simple CBC mode encryption with the padding schemes that use constant or regular padding string is vulnerable to this attack.

The simple CBC mode encryption with the proposed random padding scheme is strong for defeating the padding oracle attacks under ciphertext-only manner. Assume that a padding oracle answers that the padding string is correct or not by checking whether the mark word exists or not. By applying a padding oracle attack, the attacker can control IV to confirm that $IV \oplus D_k(C) = M$ where M is the mark word. But he/she can't know the exact value of M. That is, he/she can't get the information of M by ciphertext-only manner because the mark word can be key-dependent and is hidden from the eavesdropper. Therefore, the plaintext of C will not be derived out.

5. Conclusions

As described in this text, the proposed random padding scheme has very low information leakage. The appended padding string is almost random and hidden from the eavesdropper. The recently famous padding oracle attacks are extremely difficult to be applied on the simple CBC mode encryption with the proposed padding scheme by ciphertext-only manner.

References

[1] A. K. L. Yau, K. G. Paterson and C. J. Mitchell, Padding Oracle attacks on CBC-mode encryption with secret and random Ivs, *FSE 2005*, pp. 299–319.

[2] Federal Information Processing Standards Publication FIPS PUB 197, *Advanced Encryption Standard (AES)*, U.S. Department of Commerce/National Institute of Standards and Technology, 26 November 2001.

[3] Federal Information Processing Standard FIPS PUB 81, *DES Modes of Operation*, U.S. Department of Commerce/National Institute of Standards and Technology, 2 December 1980.

[4] J. Black and H. Urtubia, Side-channel attacks on symmetric encryption schemes: the case for authenticated encryption, in *Proceedings of 11th USENIX Security Symposium*, San Francisco, 2002, pp. 327–338.

[5] K. G. Paterson and A. Yau, Padding Oracle attacks on the ISO CBC mode padding standard, in *Topics in Cryptology CT-RSA 2004*, T. Okamoto (editor), Vol. 2964 of *Lecture Notes in Computer Science*, Springer-Verlag, 2004, p. 305–323.

[6] N. Smart, *Cryptography: An Introduction*, McGraw-Hill, 2003.

[7] PKCS #5, *Password-Based Encryption Standard*, RSA Laboratories, Version 2.0, March 1999.

[8] RFC 2630, *Cryptographic Message Syntax*, R. Housley, June 1999.

[9] S. Vaudenay, Security flaws induced by CBC padding - Applications to SSL, IPSEC, WTLS…, in *Advances in Cryptology – Eurocrypt'02*, *Lecture Notes in Computer Science*, Vol. 2332, 2002, Springer-Verlag, pp. 534–545.

Received August, 2006