LETTER

Efficient Subversion of Symmetric Encryption with Random Initialization Vector

Joonsang BAEK[†a)], ***Nonmember and*** **Ilsun YOU**[††b)], ***Member***

SUMMARY This paper presents an efficient subverted symmetric encryption scheme, which outputs a random initialization vector (IV). Compared with the available scheme of the same kind in the literature, our attack provides a saboteur (big brother) with much faster recovery of a key used in a victim's symmetric encryption scheme. Our result implies that care must be taken when a symmetric encryption scheme with a random IV such as randomized CBC is deployed.

key words: *subversion, symmetric encryption, random IV*

1. Introduction

Motivation. Since Edward Snowden revealed that the US and UK governments made a great deal of effort to subvert widely deployed cryptographic systems [4], *algorithm substitution attack (ASA)* [1] on cryptographic schemes has been being actively explored [2], [3], [5]. ASA, proposed by Bellare, Paterson and Rogaway (BPR), refers to an activity whereby a *saboteur* [5] (or a *big brother* [1], [6]) replaces an original implementation of a cryptographic scheme with a subverted one. This subverted scheme can leak partial or entire information about the secret key or message.

Note that the notion of ASA is very similar to that of "kleptographic" attack in the early literature [6]. Note also that Degabriele, Farshim and Poettering recently refined and improved BPR's security notions related to ASA [3].

In this paper, we focus on BPR's IV-replacement attack whereby a random IV, which is present as a ciphertext component of a symmetric encryption scheme such as randomized CBC, is replaced by a concocted IV that will make it possible for a saboteur to recover the key used in the symmetric encryption scheme. More precisely, the saboteur replaces the victim's symmetric encryption algorithm with a malicious one that outputs a ciphertext to encrypt the victim's secret key under his subversion key using a symmetric encryption scheme of his choice. This ciphertext is disguised as a random IV of the victim's symmetric encryption scheme. By decrypting the ciphertext disguised as the random IV with his subversion key, the saboteur will be able to obtain the victim's secret key and be in full control of the cryptosystem.

Manuscript received October 27, 2015.
Manuscript revised January 4, 2016.
Manuscript publicized January 14, 2016.
[†]The author is with Khalifa University of Science, Technology and Research, UAE.
[††]The author is with the Soonchunhyang University, Korea.
a) E-mail: joon.baek@kustar.ac.ae
b) E-mail: ilsunu@gmail.com (Corresponding author)
DOI: 10.1587/transinf.2015EDL8224

However, there are some technical difficulties in realizing IV-replacement attack, which were not addressed fully in BPR's work [1]. That is, the saboteur needs to maintain a state, which is to "remember" *one* particular subverted ciphertext that contains disguised IV that encrypts the victim's symmetric key. Since maintaining such a state (remembering one ciphertext among many) is not always practical, if not impossible. BPR also proposed the "stateless" subverted scheme in which the saboteur does not need to maintain a state. However, the problem of this approach is that the saboteur needs to collect more than 896 ciphertexts (if 128-bit key is used) to recover the victim's key. Our aim is to revisit BPR's subverted symmetric encryption scheme and improve its efficiency.

Our Contributions. We show that the aforementioned IV-replacement attack can be performed very efficiently, more efficiently than those presented in BPR's paper [1]. The idea behind our improvement is that we propose a "semi-stateful" subversion technique that *every* one of two consecutive ciphertexts contains subliminal information that can lead to the leakage of a victim's key. Basically, our attack is stateful but in our improved attack, the saboteur does not need to "remember" one particular ciphertext unlike the stateful attack presented by BPR neither does he need to collect as many as (at least) 896 ciphertexts, which is necessary in BPR's stateless attack [1]. Our attack requires the saboteur to have at most two to six ciphertexts to perform subversion sucessfully.

2. Preliminaries

In this section, we review several basic notions that will be used throughout this paper. First, we review the formal definitions of a symmetric encryption scheme, followed by pseudorandom function.

Definition 1: Let $(\mathsf{G}, \mathsf{E}, \mathsf{D})$ be a symmetric encryption scheme. The key generation algorithm G takes 1^n (n denotes the security parameter) as input and outputs a key k. Taking a key k and a message m as input, the encryption algorithm generates a ciphertext c. We denote this operation by $c \leftarrow \mathsf{E}(k, m)$. Taking a key k, and a ciphertext as input, the decryption algorithm outputs a message pair m or $\perp$ (reject symbol). We denote this operation by m (or $\perp$) $\leftarrow \mathsf{D}(k, c, s)$.

Definition 2: Let $F : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a length-preserving keyed function. F is a pseudorandom function if

for all probabilistic polynomial-time distinguishers $\mathcal{D}$, there exists a negligible function $\epsilon(n)$ such that $\left|\Pr[\mathcal{D}^{F(k,\cdot)}(1^n) = 1] - \Pr[\mathcal{D}^{f(\cdot)}(1^n) = 1]\right| \leq \epsilon(n)$, where the first probability is taken over uniform choice of k and the randomness used in $\mathcal{D}$, and the second probability is taken over uniform choice of f from the set of all functions mapping n-bit strings to n-bit strings and the randomness used in $\mathcal{D}$.

Now, we review the definition of a subverted symmetric encryption scheme [1], which is based on the generic symmetric encryption scheme defined in Definition 1.

Definition 3: Let $(\mathsf{G}, \mathsf{E}, \mathsf{D})$ be a symmetric encryption scheme. Let $(\overline{\mathsf{G}}, \overline{\mathsf{E}}, \overline{\mathsf{D}})$ be a subverted symmetric encryption scheme. The subversion key generation algorithm $\overline{\mathsf{G}}$ takes 1^n (n denotes the security parameter) as input and outputs a subversion key $\tilde{k}$. Taking a subversion key $\tilde{k}$, a key k generated by G, a message m and a state σ as input, the encryption algorithm generates a subverted ciphertext c and a new state σ'. We denote this operation by $(c, \sigma') \leftarrow \overline{\mathsf{E}}(\tilde{k}, k, m, \sigma)$. Taking a subversion key $\tilde{k}$ and a key k generated by G, a ciphertext c and a state σ as input, the decryption algorithm outputs a message/state pair (m, σ') or $\perp$ (reject). We denote this operation by $(m, \sigma')(\text{ or } \perp) \leftarrow \overline{\mathsf{D}}(\tilde{k}, k, c, \sigma)$.

Note that in the above definition, the subverted symmetric encryption scheme is defined as stateful in general, i.e., either of σ and σ' is non-empty string. If both σ and σ' are empty strings in both encryption and decryption algorithms, the scheme is stateless.

Intuitively, a basic security requirement for the subverted symmetric encryption scheme from the point of the saboteur would be that a ciphertext generated from the subverted scheme is not distinguishable from the one generated from the regular symmetric encryption scheme. The following definition called "undetectability" captures this intuition [1].

Definition 4: Let $(\overline{\mathsf{G}}, \overline{\mathsf{E}}, \overline{\mathsf{D}})$ be a subverted symmetric encryption scheme. Let $\mathcal{A}$ be an adversary. Consider the following game $\mathsf{Detect}_{\mathcal{A}}(n)$:

1. A uniform subversion key $\tilde{k}$ is generated by $\overline{\mathsf{G}}$.
2. A uniform bit $b \in \{0, 1\}$ is chosen.
3. $\mathcal{A}$ is given oracle access to $\mathsf{KEY}(\cdot)$ and $\mathsf{Enc}(\cdot)$, which are described as follows.
   - $\mathsf{KEY}(\cdot)$: On input i, this oracle returns a uniform key k_i generated by G. (Note that i is an identity associated with a key k.)
   - $\mathsf{Enc}(\cdot, \cdot)$: On input (i, m), this oracle computes $(c, \sigma_i') \leftarrow \overline{\mathsf{E}}(\tilde{k}, k_i, m, \sigma_i)$ and returns c if $b = 0$ and returns $c \leftarrow \mathsf{E}(k_i, m)$ otherwise.
4. $\mathcal{A}$ returns its guess b'.
5. The game outputs 1 if $b' = b$, and 0 otherwise.

The subverted encryption scheme is said to be undetectable if there exists a negligible function $\epsilon(n)$ such that $\Pr[\mathsf{Detect}_{\mathcal{A}}(n)] \leq \frac{1}{2} + \epsilon(n)$.

3. The Proposed Subverted Symmetric Encryption Scheme

We now describe our subverted symmetric encryption scheme $(\overline{\mathsf{G}}, \overline{\mathsf{E}}, \overline{\mathsf{D}})$, which is to conduct IV-replacement attack. Let $(\mathsf{G}, \mathsf{E}, \mathsf{D})$ be a symmetric encryption scheme with security parameter $n \in \mathbb{Z}^+$. We assume that the scheme $(\mathsf{G}, \mathsf{E}, \mathsf{D})$ is stateless and "surfaces" IV, meaning that IV is explicitly present as one of input parameters and there exists an algorithm $\mathcal{S}$ that efficiently recovers IV, i.e. $\mathcal{S}(\mathsf{E}(k, m, IV)) = IV$. Let E be an encryption algorithm of a block cipher whose key, input and output lengths are the same as n. We assume that $|IV| = n$ for the sake of simplicity (as done in [1]) and that $j \in \mathbb{Z}^+ \cup \{0\}$ is a *global counter*, initialized to 0. We also assume that a state σ is initialized to `NULL` (a null string). The description of each sub-algorithms is as follows.

- $\tilde{k} \leftarrow \overline{\mathsf{G}}(1^{n'})$: On input $1^{n'}$ such that $n' = n$, $\overline{\mathsf{G}}$ chooses a uniform $\tilde{k} \in \{0, 1\}^n$.
- $(c_j, \sigma) \leftarrow \overline{\mathsf{E}}(\tilde{k}, k, m, \sigma)$: On input a subversion key $\tilde{k}$ and a key k of E, both of which are selected at random, a counter j, a message m and a state σ, $\overline{\mathsf{E}}$ works as follows.

  If $j = 0 \bmod 2$, do the following:

  Choose a uniform $IV_j \in \{0, 1\}^n$.
  Compute $c_j \leftarrow \mathsf{E}(k, m, IV_j)$.
  Set $\tau \leftarrow IV_j$.
  Return c_j.

  If $j = 1 \bmod 2$, do the following:

  Set $\iota \leftarrow k \oplus \tau$.
  Compute $IV_j \leftarrow E(\tilde{k}, \iota)$.
  Compute $c_j \leftarrow \mathsf{E}(k, m, IV_j)$.
  Return c_j.

  Set $j \leftarrow j + 1$ and $\sigma \leftarrow (j, \tau)$.
- $\overline{\mathsf{D}}(\tilde{k}, c_l, c_{l+1})$ for some $l \in \mathbb{Z}^+ \cup \{0\}$: On input a subversion key $\tilde{k}$, two consecutive ciphertexts c_l and c_{l+1}, $\overline{\mathsf{D}}$ recovers IV_l and IV_{l+1} from c_l and c_{l+1} respectively, computes $\iota \leftarrow E^{-1}(\tilde{k}, IV_{l+1})$ and $k \leftarrow \iota \oplus IV_l$ and outputs $\mathsf{D}(k, c_{l+1})$.

Note in the algorithm $\overline{\mathsf{E}}$ that the j-th random IV_j, where $j = 0 \bmod 2$, is XORed with the subversion key $\tilde{k}$ and is provided as input to the block cipher E. This procedure essentially randomizes E. Note also that while the underlying encryption scheme $(\mathsf{G}, \mathsf{E}, \mathsf{D})$ is stateless, its subversion $(\overline{\mathsf{G}}, \overline{\mathsf{E}}, \overline{\mathsf{D}})$ is stateful. However, different from the stateful subverted scheme presented in [1], our scheme does not require the saboteur to get *one particular* subverted ciphertext to recover the symmetric key k. (Obtaining the first subverted ciphertext can be difficult in practice due to an transmission error, delay, loss and etc.) Again, different from the stateless subverted symmetric scheme presented in [1], ours

does not require the saboteur to collect as many as 896 subverted ciphertexts, which is equivalent to the size of 128-bit symmetric key multiplied by the log value of it. In our subverted scheme, the saboteur needs only a small number of consecutive subverted ciphertexts to recover the key of the underlying symmetric scheme, as shown below.

Theorem 1: Assume that pairs of consecutitve ciphertexts are selected at random by a saboteur. Then the decryption algorithm $\overline{\mathsf{D}}$ will decrypt a ciphertext correctly with probability $1-(1/2)^t$, where t denotes the number of pairs of ciphertexts.

Proof 1: For a given pair of consecutive subverted ciphertexts, depending on whether the first of them is j-th ciphertext such that $j \bmod 2 = 0$ or not, the saboteur can succeed in decrypting them. This happens with probability 1/2 assuming that a pair of ciphertexts is selected at random. When decryption fails, the saboteur picks another pair of consecutive ciphertexts at random independently and tries to decrypt them. This process continues until the saboteur decrypt a pair of ciphertexts correctly. Note that at each trial (of selecting a pair of ciphertexts), the saboteur's success probability follows geometric distribution. Hence, the probability that less than or equal to t trials ($t \geq 1$) are necessary for the saboteur to succeed is $(1-1/2)^0(1/2)+(1-1/2)^1(1/2)+\cdots+(1-1/2)^{t-1}(1/2) = 1-(1-1/2)^t = 1-(1/2)^t$.

Note that even for small t, the saboteur can successfully decrypt the pair of consecutive subverted ciphertexts and hence obtain the key for the scheme $(\mathsf{G},\mathsf{E},\mathsf{D})$ with a high probability. Thus, we call our scheme "*semi-stateful*" meaning it is essentially stateful but it is quite close to stateless scheme, which only needs the small number of subverted ciphertexts to perform IV-replacement attack.

3.1 Security Analysis

In this section, we show that our subverted symmetric scheme presented in the previous section satisfies the undetectability property (Definition 4).

Theorem 2: Assuming that the block cipher E is a pseudorandom function, our subverted symmetric encryption scheme is undetectable.

Proof 2: Let $\mathcal{A}$ be an adversary of the undetectability game for the scheme $(\overline{\mathsf{G}},\overline{\mathsf{E}},\overline{\mathsf{D}})$. We construct an adversary $\mathcal{D}$ for pseudorandom function $E(\tilde{k},\cdot)$ that uses $\mathcal{A}$ as a subroutine as follows.

1. Choose a uniform key $k_i \in \{0,1\}^n$.
2. Set $j \leftarrow 0$, $\tau \leftarrow \perp$ and $\sigma \leftarrow (j,\tau)$.
3. Choose a bit $b \in \{0,1\}$ uniformly at random.
4. On receiving (i,m) from $\mathcal{A}$ (i is a key identity), do the following:

   Parse σ as (j,τ).
   If $b = 0$, do the following:

   If $\tau \neq \perp$, set $IV_j \leftarrow \tau$. Else pick a uniform $IV_j \in \{0,1\}^{|IV|}$ ($n = |IV|$) and set $\tau \leftarrow IV_j$.
   Compute $c_j \leftarrow \mathsf{E}(k_i, m, IV_j)$ and return c_j.

   Else query $\tau \oplus k_i$ to its oracle to get $IV_j \leftarrow O(\tau \oplus k_i)$, compute $c_j \leftarrow \mathsf{E}(k_i, m, IV_j)$ and return c_j.
   Set $j \leftarrow j+1$.

5. Continue answering $\mathcal{A}$'s queries until $\mathcal{A}$ returns its guess $b' \in \{0,1\}$. If $b' = b$ return 1.

We first show that $\Pr[D^{E(\tilde{k},\cdot)}(1^n) = 1] = \Pr[\mathbf{Game}_{\mathcal{A}}^{Detect}(n) = 1]$: Note that when $O(\cdot) = E(\tilde{k},\cdot)$, $\mathcal{D}$ perfectly simulates $\mathcal{A}$'s environment including the interaction between $\mathcal{A}$ and the encryption oracle. (Note that IV's are created alternatively by choosing a uniform string of appropriate length and by encrypting the previous random IV XORed with the symmetric key k_i in this case.)

However, when $O(\cdot) = e(\cdot)$, where e is a function chosen uniformly at random from the family of functions mapping n-bit strings to n-bit strings (denoted $e \leftarrow Rand_n$), each of ciphertexts that $\mathcal{A}$ receives from the encryption oracle is identically distributed. (Their IV's are all uniform random.) Hence, $\mathcal{D}$ does not get any advantage through $\mathcal{A}$ in distinguishing subverted ciphertexts from original ones. Thus, $\Pr[\mathcal{D}^e(1^n) = 1] = \frac{1}{2}$. Then, we have $\left|\Pr_{\tilde{k}\leftarrow\{0,1\}^n}[\mathcal{D}^{E(\tilde{k},\cdot)}(1^n) = 1] - \Pr_{e\leftarrow Rand_n}[\mathcal{D}^{e(\cdot)}(1^n) = 1]\right| = \left|\Pr[\mathbf{Game}_{\mathcal{A}}^{Detect}(n) = 1] - \frac{1}{2}\right|$.

Since $\left|\Pr_{\tilde{k}\leftarrow\{0,1\}^n}[\mathcal{D}^{E(\tilde{k},\cdot)}(1^n) = 1] - \Pr_{e\leftarrow Rand_n}[\mathcal{D}^{e(\cdot)}(1^n) = 1]\right| \leq \epsilon(n)$ by assumption, we have $\Pr[\mathbf{Game}_{\mathcal{A}}^{Detect}(n) = 1] \leq \frac{1}{2} + \epsilon(n)$.

4. Concluding Remarks

Our result shows that subversion on symmetric encryption schemes with random IV can be performed very efficiently. Hence, care must be taken to implement and deploy such schemes as a form of software. It is imperative that proper verification and validation of this kind of software be conducted.

Acknowledgments

This work was partially supported by the Soonchunhyang University Research Fund.

References

[1] M. Bellare, K.G. Paterson, and P. Rogaway, "Security of symmetric encryption against mass surveillance," Proc. Crypto '14, LNCS 8616, pp.1–19, Springer, 2014.
[2] M. Bellare and V.T. Hoang, "Resisting randomness subversion: Fast deterministic and hedged public-key encryption in the standard model," Proc. Eurocrypt '15, LNCS 9057, pp.627–656, Springer, 2015.
[3] J.P. Degabriele, P. Farshim, and B. Poettering, "A more cautious approach to security against mass surveillance," Proc. FSE '15, LNCS 9054, pp.579–598, Springer, 2015.
[4] The Guardian, "Revealed: How US and UK spy agencies defeat internet privacy and security," Sept. 2013, Available at http://www.theguardian.com/world/2013/sep/05/

nsa-gchq-encryption-codes-security

[5] B. Schneier, M. Fredrikson, T. Kohno, and T. Ristenpart, "Surreptitiously weakening cryptographic systems," Cryptology ePrint Archive, 2015/097.

[6] A. Young and M. Yung, "Kleptography: Using cryptography against cryptography," Proc. Eurocrypt '97, LNCS 1233, pp.62–73, Springer, 1997.