Understanding the *Security of Symmetric Encryption against Mass Surveillance* (Type-1 Final Project)

Prakhar Sinha, Áine Keenan, and Julia Heiler

University of California, Davis

1 Introduction

The academic paper, "Security of Symmetric Encryption against Mass Surveillance", by Mihir Bellare, Kenneth G. Paterson, and Phillip Rogaway, focuses on the Algorithm Substitution Attack (ASA), which replaces an encryption scheme with an NSA-authored alternative. The subverted scheme should reveal the plaintext to NSA, referred to as "big brother", but the ciphertexts are indistinguishable to users with the real encryption scheme. The subverted scheme reveals the plaintext by allowing for the recovery of the key K or message M. Therefore, by observing the ciphertexts from the subverted scheme, big brother is able to break privacy. The security notion for privacy will be the ability for an adversary to distinguish the real encryption of the message versus the encryption of zero bits of a length equal to the length of the message. Overall, big brother can break privacy for a large amount of ciphertexts by mass surveillance using ASAs. To best prevent these attacks, the paper discusses that stateful, deterministic encryption should be used and is highly recommended.

An encryption scheme Π is defined by a key Space K, encryption algorithm E, and eecryption algorithm D. A subversion is another scheme $\widetilde{\Pi}$, which is defined by another key Space $\widetilde{K}$, a nonempty finite set, encryption algorithm $\widetilde{E}$, and decryption algorithm $\widetilde{D}$. Instead of employing E, $\widetilde{E}$ is chosen to be used, allowing for big brother to potentially learn information about the underlying plaintexts. $\widetilde{E}$ also needs to decrypt correctly under $\widetilde{D}$, achieving the original plaintext.

The paper denotes two important notions of security: detection security and surveillance security. Detection security requires that subverted ciphertexts by big brother, from $\widetilde{E}$, are indistinguishable from real ciphertexts, by E, even if user keys K are known. However, big brother's key $\widetilde{K}$ is not known. Surveillance security requires that big brother cannot distinguish subverted ciphertexts from real ciphertexts, even with their key, but not user keys. Notably, the paper only looks at schemes that are correct. Correctness is defined for a scheme $\Pi = (K, E, D)$ as for all vectors of messages, each message is encrypted by E with their corresponding associated data to make a vector of ciphertexts $\mathbf{C}$. Once $\mathbf{C}$ is decrypted, to get a vector of messages, each message is equal to the original message. Additionally, the paper stresses the importance of stateful encryption schemes as it relates to ASAs. Stateless schemes, for both the encryption and decryption, are defined by the second component of E/D consisting of ε (empty string).

Big brother's subverted encryption scheme is defined as below, with $\widetilde{K}$ being a finite non-empty set. $\widetilde{E}$ is the encryption algorithm, which has the input of the big brother's key $\widetilde{K}$, the real key K, the message M, Associated Data A, the current state, σ and public information identifiers i for the user like IP address. $\widetilde{E}$ outputs a ciphertext $\mathbf{C}$, and an updated state, σ'.

$$\widetilde{E}(K, \widetilde{K}, M, A, \sigma, i) \rightarrow \mathbf{C}, \sigma'$$

$\widetilde{D}$ is the plaintext recovery, which has an input of $\widetilde{K}$, $\mathbf{C}$ = a collection of ciphertexts, A, i, and outputs M = a vector of messages corresponding to each ciphertext.

$$\widetilde{D}(K, \mathbf{C}, A, i) \rightarrow M$$

It is important that for all subverted schemes, $\widetilde{D}$ is decryptable and outputs the correct plaintext; otherwise, it is easy for the user to detect it is a subverted scheme. Therefore, we define that $(\widetilde{K} \times K, \widetilde{E}, \widetilde{D})$ based on $\widetilde{\Pi}$ and Π is a correct encryption scheme, where $\widetilde{D}(K, K, A, \mathbf{C}, \text{state})$ equals $D(K, A, \mathbf{C}, \text{state})$. Therefore, encryption is based on $\widetilde{K}$ and K, but D is decryptable with just the real key K.

Big brother does not want their ASA to be detected. Detectability is the ability for normal users, who know their secret key but not big brother's master key, to tell if the ciphertext is from the real or subverted

encryption scheme. The condition for detectability is a detection test by users, U, which has an advantage defined by the ability to distinguish a ciphertext encrypted by the real encryption scheme or the subverted encryption scheme. The detection test is given the keys of the user but it is not given a big brother master key. The advantage is quantified by two times the probability that U will correctly identify if it was the subverted encryption scheme or the real encryption scheme.

$$\mathbf{Adv}^{\text{det}}_{\Pi,\widetilde{\Pi}}(\mathcal{U}) = 2\Pr[\text{DETECT}^{\mathcal{U}}_{\Pi,\Pi_e} \Rightarrow \text{true}] - 1$$

If the advantage here is low, then the subverted encryption scheme and real encryption scheme are undetectable, a success for the adversary. After undetectability, the adversary is super successful if they are able to recover the plaintext from the subverted ciphertext. Even if the advantage is large, many users may not know what to look for to detect subverted encryption schemes, making it hard to utilize U.

The notion of being secure against ASAs is that even with $\widetilde{K}$, big brother, $\mathcal{B}$, cannot tell ciphertexts encrypted under the real scheme from ciphertexts encrypted under the subverted scheme. The surveillance advantage looks at whether the big brother adversary can distinguish ciphertexts encrypted under the real encryption scheme from ciphertexts encrypted under the subverted encryption scheme. The formula is:

$$\mathbf{Adv}^{\text{srv}}_{\Pi,\widetilde{\Pi}}(\mathcal{B}) = 2\Pr[\text{SURV}^{\mathcal{B}}_{\Pi,\Pi_e} \Rightarrow \text{true}] - 1$$

However, unlike the condition for undetectability, the surveillance advantage adversary knows the big brother's master key, but does not know the keys of the user. Therefore, it uses the master key to determine if the ciphertext is encrypted under the real scheme or the subverted scheme. To be secure against surveillance, the advantage must be small for all real encryption schemes, subverted encryption schemes, and all big brother adversaries.

2 Mounting ASAs

With these notions of detectability and surveillance, we can build ASAs that can attack different symmetric encryption schemes successfully. The schemes fail in the sense that the user key can be extracted by an ASA from subverting ciphertext and the scheme remains undetectable. The two attacks described include *IV-replacement attacks* and *the biased-ciphertext attack*. Both of these attacks have the ability to have devestating consequences on the modern internet security infastrcuture, namely SSL/TLS, IPSec, and SSH.

2.1 IV-replacement attacks

Stateless encryption schemes that insert a random nonce into the ciphertext are vulnerable to this attack. For this attack, we will consider a randomized, stateless scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ that writes $C \leftarrow \mathcal{E}(K, M, A; IV)$. These attacks deal with *surfacing an IV*. A scheme surfaces its IV if there exists an efficient algorithm $\mathcal{X}$ such that $\mathcal{X}(\mathcal{E}(K, M, A; IV)) = IV$. Plainly, the function $\mathcal{X}$ can recover the IV from a given ciphertext. There are two different versions of these attacks.

2.1.1 Stateful Attack

There are a few straight forward steps involved in a stateful IV-replacement attack. We consider the subverted encryption algorithm $\widetilde{\Pi} = (\widetilde{\mathcal{K}}, \widetilde{\mathcal{E}}, \widetilde{\mathcal{D}})$. Let the bit length of the IV and the key be n and assume we have blockcipher $E : \widetilde{\mathcal{K}} \times \{0,1\}^n \rightarrow \{0,1\}^n$ with block length n.

1. The state, σ, of $\widetilde{\mathcal{E}}$ is initialized to 0.
2. The IV is selected. If σ is 0, the IV is set to an encryption of the key K. Formally: $IV \leftarrow E(\widetilde{\mathcal{K}}, K)$. Else the IV is set randomly. It is presumed that this is to avoid repetition. The IV can only be set to a specific value, i.e. the encryption of the key K once. If it were to be repeated, this would be very suspicious.
3. A user, i, requests encryption of message $\mathbf{M}[1]$ under associated data $\mathbf{A}[1]$ with $\sigma = 0$. The ciphertext $\mathbf{C}[1] = \widetilde{\mathcal{E}}(\widetilde{K}, K, \mathbf{M}[1], \mathbf{A}[1], 0, i)$ is returned.

4. Big brother has the ciphertext encrypted under key $\widetilde{\mathcal{K}}$. With this information, they can decrypt the ciphertext and extract key K. With the key, all future ciphertexts are compromised.

The subversion stays undetectable given E is a secure PRF/PRP, the subverted IV will look no different then a random one.

The idea of a system reset is brought up later. Say the user resets their system, then the state, σ will be reset to 0 and the IV will once again send the same subverted IV. This would look highly suspicious that the scheme always sends the same IV upon system reset. As such, instead of subverting the IV whenever $\sigma = 0$ in the same way, the big brother could incorporate a randomized symmetric encryption scheme such that the IV would be subverted in a way that looks random. This way, the attack needn't rely on $\sigma = 0$. It could attack given any message and the sensitive ciphertext information would continuously be leaked.

2.1.2 Stateless Attack

A stateless attack is immune to the system reset problem described earlier. In this attack, we'll describe a few new parameters: k is the key length of Π (the encryption scheme), $v = \lceil \log_2(k) \rceil$ and $E : \widetilde{\mathcal{K}} \times \{0,1\}^n \rightarrow \{0,1\}^n$ is a blockcipher where n is the length of the IV. The subversion is a triple, $(\widetilde{\mathcal{K}}, \widetilde{\mathcal{E}}, \widetilde{\mathcal{D}})$ that works as follows:

1. $\widetilde{\mathcal{E}}(\widetilde{K}, K, M, A, i) \rightarrow C$
   a) The first step is to generate a random index, ℓ, from the range $[1...k]$.
   b) Next, we generate a random string, R, of length $n - v - 1$.
   c) Compute the IV such that $IV \leftarrow E(\widetilde{K}, K[\ell] + \ell + R)$ (here + means concatenation of strings). We are concatenating ℓth bit of K, ℓ and R which is just a random string.
   d) Finally we encrypt K, M, A and the IV and return the ciphertext, C.
2. $\widetilde{\mathcal{D}}(\widetilde{K}, \mathbf{C}, \mathbf{A}, i) \rightarrow M$
   a) For each ciphertext block, $\mathbf{C}[j]$, extract the ℓth bit of the key as mentioned in the subverted encryption algorithm.
   b) The decryption algorithm maintains state and stores all the bits of key it's extracted. Given enough ciphertexts, the key will be able to be reconstructed. Thus, the all future messages are now compromised.

As stated earlier, the advantage of this stateless algorithm is that a system reset would not led to the detection of the algorithm.

2.2 The Biased-Ciphertext Attack

The above method is quite strong but it is the case that there exist more than a few encrypted-IV schemes that do not surface their IV, CBS2, IACBC and XCBC$ are notable examples. In these cases another attack can be used that applies to *any* randomized and stateless encryption scheme, as long as it used a minimal amount of randomness. It is also resistant to the system reset condition was brought up earlier.

The main feature of this argument is the idea of manipulating the randomness employed by a scheme to leak a key of length k after k encryptions. Furthermore, the bias that is created by this subverted PRF should be undetectable, even when the user is in possession of the key k.

There are a few points of discussion surrounding this attacks. For one, this attack is a *no-chosen message attack*. This means that as long as k messages are encrypted, big brother will win and the key will be recovered. In accordance with formal security definitions, big brother's advantage is almost 1.

Say we have a new scheme with $\mathcal{E} \leftarrow (K, M, A; \delta)$. $delta$ represents the random coins that will be used in the encryption. For biased-ciphertext attack to work, the scheme that it attacks must be *coin injective*. What does this mean? For an scheme to be coin injective, different values of δ must produce different encryptions for each message and associated data. For example, schemes that surface their IV are coin injective as well as

ones that use a random nonce. It is necessary for a scheme that falls victim to this attack to be coin injective because the attack manipulates the random coins to recover the key K. With this out of the way, it is time to break down the attack:

1. $\widetilde{\mathcal{E}}(\widetilde{K}, K, M, A, \sigma, i) \rightarrow C$
   a) The first step of the encryption algorithm is to determine j. j corrusponds to the index of the key that will be leaked. This is represented by the line $j \leftarrow \sigma \mod |K|; j \leftarrow j + 1$
   b) Compute the function g: $g(\cdot) \leftarrow \mathcal{E}(K, M, A; \cdot)\|\sigma\|i$. What does the function g do? g has the values K, M, A, j, σ and i embedded in the function. On input coins, δ, the function returns $\mathcal{E}(K, M, A; \delta)\|\sigma\|i$. This function is important as, in the next step, it is used to calculate the subverted random coins.
   c) We next "randomly" select a random coin here: $\delta \twoheadleftarrow S^{F(\widetilde{K},\cdot),g(\cdot)}(K[j], D)$. In this line we are randomly selecting a coin from a fixed pool coins such that any coin chosen will leak information about the key. The set we are choosing from is the set $\{\delta \in D : F(\widetilde{K}, g(\delta)) = K[j]$. All keys chosen from this set will leak $K[j]$.
   d) Finally, encrypt the message, associated data, and subverted random coin under the key, K and the normal encryption algorithm and return C. This ciphertext hides information that will be used by $\widetilde{\mathcal{D}}$ to leak $K[j]$.
2. $\widetilde{\mathcal{D}}(\widetilde{K}, \mathbf{C}, \mathbf{A}, i) \rightarrow \mathbf{M}$
   a) The decryption algorithm functions similarly to the one from the stateless IV-replacement attack, constructing the key K bit by bit until it has all the pieces. For each block of the ciphertext, $\mathbf{C}[j]$, extract the bits of the key using the PRF from $\widetilde{\mathcal{E}}$.
   b) The algorithm maintains and updates it's state until it has all the pieces of K. After that, as in the previous examples, the key K has been retreived and all future messages are comprimised. Big brother has won.

The potential for error in this attack is when the reconstructed key is not equal to the actual key K. However, the likelihood of this is very small. The scenario in which there is the lowest probability of error is when there is a minimal amount of randomness involved in the original encryption scheme. The only thing left to prove is that this subversion is undetectable. **Lemma 1** claims the following (let $p = Pr[\delta = \widetilde{\delta}]$), $g : D \rightarrow R$, $d = |D|$):

1. If g is injective, then $p = (1 - 2^{-d})/d$
2. If g is k-regular, then $p = (1 - 2^{-d}/k)/d$

This lemma is then used to prove **theorem 3**. The theorem is slightly more complicated than this but, in essence, as long as Π draws its coins from a large sample space, and the PRF is secure, the subverted encryption scheme is undetectable.

3 Defeating ASAs

In order to effectively counter Algorithm-Substitution Attacks (ASAs), as detailed above, the focus must shift to designing encryption schemes that inherently resist such attacks. Given the vulnerabilities highlighted, it is evident that any viable scheme must be both deterministic and stateful. However, these characteristics alone are insufficient. Traditional security properties such as privacy and authenticity, which assume that adversaries never are able to possess the key K, fail to provide adequate protection in scenarios where the subverted encryption algorithm does possess the key. Encryption schemes typically rely on randomness to achieve security, but this very randomness can be exploited by ASAs. Randomized and stateless schemes are vulnerable because they allow adversaries to manipulate the encryption process and introduce undetectable biases. In contrast, deterministic and stateful encryption schemes offer a more robust defense. Consequently, ensuring resistance to ASAs necessitates reliance on specific combinatorial properties of the encryption scheme. Deterministic encryption generates the same ciphertext for a given plaintext and key every time it is used. This predictability eliminates the variability that attackers could exploit to insert hidden channels or biases. Stateful encryption maintains context across multiple encryption operations. One crucial property

identified is the uniqueness of ciphertexts, which plays a pivotal role in achieving surveillance security. By incorporating state information such as counters or nonces, stateful encryption ensures that each ciphertext is unique, even for the same plaintext and key, thereby thwarting attempts to insert undetectable subversions.

To achieve surveillance-resistant encryption, one would need to construct a symmetric encryption scheme that embodies the principles of determinism, statefulness, as well as the quality of producing unique ciphertexts. Two primary approaches are highlighted: permutation-based encryption (PBE) and transforming nonce-based schemes into stateful, deterministic ones. PBE leverages permutations to ensure that each encryption operation is unique. By applying a permutation to the input data before encryption, the scheme can achieve uniqueness and resist ASAs. Additionally, incorporating authentication mechanisms will result in any attempted tampering or subversion to become easily detectable. Nonce-based encryption schemes, which use a unique nonce for each encryption operation, can be adapted to stateful deterministic schemes. By treating a nonce as a stateful counter, the scheme ensures that each ciphertext is unique and verifiable. This transformation maintains the efficiency and security properties of the original-nonce based scheme while enhancing resistance to ASAs. This uniqueness guarantees that even if an adversary attempts to introduce a subverted algorithm, the resulting ciphertext will either match the expected output or be immediately detectable due to decryption failures.

Theorem 4. *Let* $\Pi = (K, E, D)$ *be a unique ciphertext symmetric encryption scheme. Let* $\widetilde{\Pi} = (\widetilde{K}, \widetilde{E}, \widetilde{D})$ *be a subversion of* Π *that obeys the decryptability condition relative to* Π. *Let B be an adversary. Then* $Adv_{srv}^{\Pi,\widetilde{\Pi}}(B) = 0$.

Furthermore, Theorem 4 formalizes the notion that a unique-ciphertext scheme cannot be subverted without violating the decryptability condition. The theorem asserts that if Π is a unique ciphertext symmetric encryption scheme, then any subversion Π' that adheres to the decryptability condition relative to Π will result in a surveillance advantage of zero for any adversary B. In essence, this means that big brother cannot distinguish between ciphertexts produced by the real encryption scheme and those produced by the subverted scheme, ensuring robust surveillance security.

In summary, designing encryption schemes that resist ASAs requires a combination of deterministic, stateful operations, and the principle of unique ciphertexts. These characteristics certify that each encryption operation produces a distinct and verifiable output, preventing adversaries from introducing undetectable subversions. By utilizing these principles, it is possible to construct encryption schemes that provide strong protection against mass surveillance and sophisticated cyber attacks.