Summary: Authentication and Misuse-Resistant Encryption of Key Dependent Data

1 Introduction

While there exist a wide variety of security schemes that provide both encryption and authentication, most security notions make no guarantees about security when the output of an encryption scheme is dependent on the key. In practical circumstances, this can happen when a key is used to encrypt data which includes the key itself such as encrypting an entire disk. In the provable security world, most notions of security make no guarantees when encrypting key dependent data. Given an IND-CPA secure encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, we can create an new IND-CPA encryption scheme $\bar{\Pi} = (\mathcal{K}, \bar{\mathcal{E}}, \bar{\mathcal{D}})$ that is completely insecure. For example, $\bar{\mathcal{E}}_k(M) = 1 \parallel K$ if $M = K$ and $\bar{\mathcal{E}}_k(M) = 0 \parallel \mathcal{E}_k(M)$ otherwise. The paper by Bellare and Keelveedhi (BK) makes three major contributions: first, it further extends the work of Black, Bellare and Shrimpton (BRS) [2], Backes [4] and Green [5] extending the KDM game to encryption schemes with headers and random or deterministic IVs. Second, it shows that key dependent headers and deterministic encryption schemes are impossible to secure and provides a way to turn any other secure authenticated-encryption scheme into a KDM-secure one. Finally, it provides results regarding misuse tolerance, such as when an adversary gains control of the PRNG. This summary covers all but the latter section.

2 Definitions

To test an encryption scheme against attacks with key-dependent data, a new notion of security is required. The IND-KDM, or indistinguishability under key dependent data is a game that requires the adversary to provide a function $g(K)$. $g(K)$ takes as input the key (as well as any other parameters provided by the adversary) and may be any function that maps $K \rightarrow \{0,1\}^{|t|}$ for some constant t. The output of g is then encrypted using $C \leftarrow \mathcal{E}(K, IV, H, g(K))$ and returned to the adversary. The IND-KDM advantage is defined as

$$\mathbf{Adv}_{\Pi}^{\text{ae}-\text{kdm}}(A) = Pr\left[\mathcal{E}(g(k)) : A \rightarrow 1\right] - Pr\left[\mathcal{E}(0^{|g(K)|}) : A \rightarrow 1\right]$$

This is an difficult notion of security to satisfy; BRS and BK both claim there is no known encryption scheme that satisfies these properties. A few restrictions, though, prevent trivial wins. First, A can communicate additional parameters to g, but g cannot access the memory of A to, say, write the key. Second, the length preserving aspect of g is required to prevent

the following attack: $g_i(K) = 0$ if $K[i] = 0$ (the i-th bit of K is 0), $g_i(K) = 00$ if $K[i] = 1$. Finally, the decryption oracle provides only verification ($\perp$ or 1) instead of the message to prevent recovery via replay.

All of the encryption schemes discussed below are Authenticated Encryption ones. This differs from notions discussed in class in that we not only provide $\mathcal{E}$ with a key K and a message M, but we also explicitly specify an IV and a header He where appropriate. Unlike the message, He is not encrypted (that is, it's not necessary to retain privacy), but the integrity header *must* be retained: $\mathcal{E}(K, IV, He, \mathcal{D}(K, IV, He, M)) = 1$ but $\mathcal{E}(K, IV, He, \mathcal{D}(K, IV, Z, M)) = \perp$ for all $Z \neq He$. In the context of the KDM game, in addition to providing a function for the message, the adversary provides a function for the header if that is key dependent as well.

All of the attacks and proofs below generalize to the case where there are multiple keys. For the sake of notational clarity and brevity, we treat each case as if there is only a single key.

3 Impossibility Results

3.1 A Deterministic Encryption Scheme is KDM-Insecure

BRS claimed that an attack that finds a function such that the inner product of a vector V (provided by the adversary) and the ciphertext is equal to $K[i]$ could be used to break a deterministic scheme, but did not provide details on how such a function would be found. BK takes a slightly different approach. It seeks to find a C such that $H(S, C) = K[i]$ for a pairwise independent family of hash functions $H : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}$. Lemma 3.1 establishes that for random S and l trials, $Pr\left[H(S, R) = T\right] \leq 1/l$ for random ciphertext R and $T \in 0, 1$. By encrypting $4k$ distinct messages and testing $H(S, C) = K[i]$ for random S per bit, the probability that $H(S, C) \neq K[i] \leq 1/4k$ and by a union bound we fail to recover the key with probability $1/4$. The authors also prove that a slight modification of the attack provided by BRS

$$H(S, C) = ((S[1]C[1] + S[2]C[2] + ...S[n]C[n]\%2) + S[n + 1])\%2$$

for $H(S, C) = \{0,1\}^{n+1} \times \{0,1\}^n \rightarrow \{0,1\}$ is pairwise independent and can be proven to succeed with the same probability for $l = 8k$ (that is, it is as good as BK's construction by a factor of two)

3.2 A Key Dependent Header is KDM-Insecure

The paper also proves that it is impossible to have key-dependent data in the header and maintain security. This is the case for both deterministic and randomized algorithms (IV is omitted below). It is important to note that the adversary's header function h doesn't encrypt $h(K)$, but it doesn't return the value to the adversary. The result is simply 'noted' by the encryption scheme as to be used for authentication. However, by using the fact that

we return $\perp$ for any header that doesn't match, we can break the encryption scheme as follows: construct a function $h_i(K) = K[i]$ and $g(K) = 0$. We can recover the i-th bit of the key by $(IV, C) \leftarrow \mathcal{E}(h_i, g)$; $M = \mathcal{D}(IV, g, C)$. If $M \neq \perp$, we know that $K[i] = 0$ and if $M = \perp$ then $K[i] = 1$. By querying for each bit of K, we can then recover the full key.

4 Proof Results

Due to the results above we omit any reference to the header and assume any IV is chosen randomly by the encryption scheme. Additionally, the bounds presented in the paper are adjusted for $w = 1$ keys.

The impossibility results achieved by BRS and BK hint at the problem and a possible solution in the IND-KDM game. Namely, if g is privy to everything that $\mathcal{E}$ knows, the adversary can recover the key. By hashing the key with some randomness and using that as the key, the function g cannot mount an attack in the same fashion it does in Section 3 for deterministic encryption schemes because g will not receive the random IV under which $g(K)$ is to be encrypted until after the encryption has taken place, preventing the attack seen in section 3.1.

BRS proved that the construction $\mathcal{E}(K, IV, M) = IV \parallel (H(K \parallel IV) \oplus M)$ is approximately birthday bound (on the size of the keyspace) secure. The BK scheme, however, takes any existing encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ that is secure for key independent data (under e.g., IND-CPA) and turns it into a scheme that is IND-KDM secure. The scheme $\bar{\Pi}$ is

$$\bar{\mathcal{E}}(K, M) = \mathcal{E}(H(K \parallel IV), M) \rightarrow (IV, C); \bar{\mathcal{D}}(K, IV, C) = \mathcal{D}(H(K \parallel IV), C) \rightarrow M$$

where H is a hash function $\{0,1\}^k \rightarrow \{0,1\}^k$ that is represented in the proof with a random oracle.

We must do a reduction on Π. First, we distinguish between hash queries made by adversary A directly (call this HASH) and those made by $\mathcal{E}$,$\mathcal{D}$ and g (call this IHASH). The main challenges are to show that a) the probability of a hash collision via HASH or IHASH is unconditionally low given that g cannot compute C, b) the probability of A or g computing C is dependent on the security of the underlying encryption scheme (for example $\mathcal{E}(K, M) = M$ leads to an easy recovery of HASH($IV \parallel K$) c) these two properties ensure that A will unlikely call HASH($IV \parallel K$) and thus recover the encryption key.

Thereom 4.1 bounds the IND-KDM advantage of an adversary D constructed from an AE adversary A making q_e encryption queries, q_d decryption queries and q_h HASH queries and a random IV of r bits.

$$\mathbf{Adv}^{\mathrm{ae-kdm}}_{\bar{\Pi}}(A) \leq (24q_e^2 + 2q_d) \cdot \mathbf{Adv}^{\mathrm{ae}}_{\Pi}(D) + \frac{8q_e q_h}{|K|} + \frac{q_e(q_e + 2q_h)}{2^r}$$

The proof sets up a hybrid game parameterized by k where the encryption oracle returns real encryptions of $g(K)$ if query $i < k$, toggles a bit b specifying whether to return real

encryptions or random bits on $i = k$, and returns random encryptions on query $i > k$. The point of this construction is to assume A makes its critical query $H(K \parallel IV_k)$ at $i = k$. The reduction cannot answer this query and at $i > k$, nothing prevents the function $g(K)$ from calling $H(K \parallel IV_k)$ once IV_k is known. It is not the case that A makes this query before $i = k$, because it does not know the value of IV_k and it is only the case that it can make the query (via HASH) after calling $\bar{\mathcal{E}}$ and recovering K from the k previous queries. This sets up a game where calling $H(K \parallel IV_k)$ sets BAD and can be removed at the cost of BAD by the Fundamental Theorem of Game Playing.

The bound above is complex and spans 7 games (and the authors concede it is not tight) due to a number of technical issues: to extract the security of the underlying scheme, another hybrid is created to output real encryptions for only the k-th query. In turn, a third hybrid bridges between the other two by giving real encryptions for $p < i < k$. Finally, the proof must also handle decryption queries, which it does in a similar fashion.

5 Conclusion

This paper presents an interesting variation of traditional security models that have applications of minimizing problems when encryption schemes are used improperly. It provides impossibility results and a particularly unique proof that utilizes hybrid games to perform a reduction. Future avenues for exploration include tightening the bound above as well as determining additional constructions on the authenticated encryption model to secure key dependent headers.

References

[1] Mihir1 Bellare, Sriram Keelveedhi, Authenticated and Misuse-Resistant Encryption of Key-Dependent Data, Cryptology ePrint Archive, Report 2011/269 (final version of preliminary paper presented at CRYPTO 2011), 2001

[2] J. Black, P. Rogaway, and T. Shrimpton. Encryption-scheme security in the presence of key-dependent messages. In K. Nyberg and H. M. Heys, editors, SAC 2002, volume 2595 of LNCS, pages 6275. Springer, Aug. 2003.

[3] B. Applebaum. Key-dependent message security: Generic amplification and completeness. In K. G. Paterson, editor, EUROCRYPT 2011, volume 6632 of LNCS, pages 527546. Springer, May 2011.

[4] M. Backes, M. D urmuth, and D. Unruh. OAEP is secure under key-dependent messages. In J. Pieprzyk, editor, ASIACRYPT 2008, volume 5350 of LNCS, pages 506523. Springer, Dec. 2008.

[5] M. Green and S. Hohenberger. CPA and CCA-secure encryption systems that are not 2-circular secure. Cryptology ePrint Archive, Report 2010/144, 2010. http://eprint.iacr.org/.