A Note on the Security Framework of Two-key DbHtS MACs

Tingting Guo[1,2] and Peng Wang[1,2](✉)

[1] SKLOIS, Institute of Information Engineering, CAS
`guotingting@iie.ac.cn, w.rocking@gmail.com`
[2] School of Cyber Security, University of Chinese Academy of Sciences

Abstract. Double-block Hash-then-Sum (DbHtS) MACs are a class of MACs achieve beyond-birthday-bound (BBB) security, including SUM-ECBC, PMAC_Plus, 3kf9 and LightMAC_Plus etc. Recently, Shen et al. (CRYPTO 2021) proposed a security framework for two-key DbHtS MACs in the multi-user setting, stating that when the underlying blockcipher is ideal and the universal hash function is regular and almost universal, the two-key DbHtS MACs achieve $2n/3$-bit security. Unfortunately, the regular and universal properties can not guarantee the BBB security of two-key DbHtS MACs. We propose three counter-examples which are proved to be $2n/3$-bit secure in the multi-user setting by the framework, but can be broken with probability 1 using only $\mathcal{O}(2^{n/2})$ queries even in the single-user setting. We also point out the miscalculation in their proof leading to such a flaw. However, we haven't found attacks against 2k-SUM-ECBC, 2k-PMAC_Plus and 2k-LightMAC_Plus proved $2n/3$-bit security in their paper.

Keywords: MAC · DbHtS · Beyond-birthday-bound security · Multi-user security.

1 Introduction

Message Authentication Code (MAC). MAC is a symmetric-key crypto primitive to ensure integrity of messages. Most of their security proofs, including XCBC [4], PMAC [5,13], HMAC [2], and NMAC [2], follow the Hash-then-(Fixed-Input-Length) Function (HtF) framework:

$$\mathrm{HtF}[H, E](K_h, K, M) = E_K(H_{K_h}(M)).$$

When H is an almost universal (AU) hash function and E is a fixed-input-length PRF (often instantiated as a blockcipher), HtF is a variable-input-length PRF [16] with birthday bound security, i.e. they are secure up to $\mathcal{O}(2^{n/2})$ queries where n is the input size of E. However, birthday bound security is always not enough for modes of lightweight blockciphers, such as PRESENT [6], GIFT [1], etc. whose block size is 64-bit. In this case, the security is only 32-bit (i.e., security up to 2^{32} queries), which is practically vulnerable. Therefore researchers make great efforts to construct MACs with better security bounds.

Birthday-Birthday-Bound (BBB) MACs. Plenty of MACs with BBB security have been put forward, such as SUM-ECBC [17], PMAC_Plus [18], 3kf9 [19], LightMAC_Plus [11], and so on. Their primary proofs [17,18,19,11] gave $2n/3$-bit security (ignoring the maximum message length). At FSE 2019, Datta et al. showed that these MACs all follow the three-key Double-block Hash-then-Sum (DbHtS) framework [8]:

$$\mathrm{DbHtS}[H, E]\,(K_h, K_1, K_2, M) = E_{K_1}(H^1_{K_{h,1}}(M)) \oplus E_{K_2}(H^2_{K_{h,2}}(M)),$$

where M is the massage, $K_h = (K_{h,1}, K_{h,2})$ is the key for two universal hash functions H^1 and H^2, K_1, K_2 are keys for blockcipher E. In the following, we treat the blockcipher is as a Pseudorandom Permutation (PRP). E_{K_1} and E_{K_2} mean two independent PRPs. Datta et al. proved that when H is weak-cover-free and weak-block-wise universal, the three-key DbHtS is $2n/3$-bit secure. Later, Leurent et al. [10] showed the best attacks to them cost $\mathcal{O}(2^{3n/4})$ queries. Recently at EUROCRYPT 2020, Kim et al. [9] have proved the tight $3n/4$-bit security (ignoring the maximum message length) if H is only almost universal.

Datta et al. [8] also found that the two-key DbHtS, that is to say, $K_1 = K_2$ in the above framework, is still $2n/3$-bit security when H is cover-free, block-wise universal and colliding.

Two-key DbHtS in the multi-user setting. All the above MAC frameworks only considered single-user setting. In practice, the adversary can attack multiple users. For instance, MACs are core elements of real-world security protocols such as TLS, SSH, and IPsec, which are used by lots of websites with plenty of daily active users. However, by a generic reduction, all above BBB results degrade to (or even worse than) the birthday bound in the multi-user setting [14].

At CRYPTO 2021, Shen et al. [14] revisited the security of two-key DbHtS framework in the multi-user setting elaborately. Their framework (Theorem 1 in [14]) states that when the underlying blockcipher is an ideal cipher and the two independent universal hash functions are both regular and almost universal, the two-key DbHtS MACs, including 2k-SUM-ECBC, achieve $2n/3$-bit security. They adjusted the proof of the framework for adapting to 2k-PMAC_Plus and 2k-LightMAC_Plus based on two dependent universal hash functions, stating they achieve $2n/3$-bit security, too.

We summarise the above security frameworks for MACs in Table 1.

Our contributions. We show that *Theorem 1 in Shen et al.'s paper [14], giving the security of two-key DbHtS framework, has a critical flaw* by three counter-examples. According to their framework, these counter-examples are proved $2n/3$-bit security (ignoring the maximum message length) in the multi-user setting. However, they are all attacked successfully with only $\mathcal{O}(2^{n/2})$ queries even in the single-user setting. We also show clearly the miscalculation in their proof leading to such a flaw.

Table 1. Summary of security frameworks for MACs. n is the input size of E. 'SU' means single-user setting. 'MU' means multi-user setting. The security of MACs ignores the maximum message length.

Framework	Property of H	Property of E	Setting	Security
HtF	AU	PRF	SU	$n/2$ [16]
Three-key DbHtS	Weak-cover-free Weak-block-wise universal	PRP	SU	$2n/3$ [8]
Two-key DbHtS	Cover-free Block-wise universal Colliding	PRP	SU	$2n/3$ [8]
Three-key DbHtS	AU	PRP	SU	$3n/4$ [9]
Two-key DbHtS	Regular AU	Ideal cipher	MU	$2n/3$ [14]

2 Preliminaries

Notation. For a finite set $\mathcal{X}$, let $X \xleftarrow{\$} \mathcal{X}$ denote sampling X from $\mathcal{X}$ uniformly and randomly. Let $|\mathcal{X}|$ be the size of the set $\mathcal{X}$. For a domain $\mathcal{X}$ and a range $\mathcal{Y}$, let $\mathsf{Func}(\mathcal{X}, \mathcal{Y})$ denote the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$.

Multi-user Pseudorandom Function (PRF). Let $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ be a function. The game $\mathbf{G}_F^{\mathrm{prf}}(\mathscr{A})$ about adversary $\mathscr{A}$ is defined as follows.

1. Initialize $K_1, K_2, \ldots \xleftarrow{\$} \mathcal{K}$, $f_1, f_2, \ldots \xleftarrow{\$} \mathsf{Func}(\mathcal{X}, \mathcal{Y})$, and $b \xleftarrow{\$} \{0,1\}$;
2. $\mathscr{A}$ queries Eval function with (i, X) and get $\mathsf{Eval}(i, X)$, where $i \in \{1, 2, \ldots\}$, $X \in \mathcal{X}$, and
$$\mathsf{Eval}(i, X) = \begin{cases} F(K_i, X), & \text{if } b = 0, \\ f_i(X), & \text{if } b = 1; \end{cases}$$
3. $\mathscr{A}$ output b'.

Then the advantage of the adversary $\mathscr{A}$ against the multi-user PRF security of F is
$$\begin{aligned} \mathrm{Adv}_F^{\mathrm{prf}}(\mathscr{A}) &= |2\Pr[b' = b] - 1| \\ &= |\Pr[b' = 1 | b = 1] - \Pr[b' = 1 | b = 0]| \end{aligned}$$

The H-coefficient technique. When considering interactions between an adversary $\mathscr{A}$ and an abstract system $\mathbf{S}$ which answers $\mathscr{A}$'s queries, let X_i denote the query from $\mathscr{A}$ to $\mathbf{S}$ and Y_i denote the response of X_i from $\mathbf{S}$ to $\mathscr{A}$. Then the resulting interaction can be recorded with a transcript $\tau = ((X_1, Y_1), \ldots, (X_q, Y_q))$. Let $p_{\mathbf{S}}(\tau)$ denote the probability that $\mathbf{S}$ produces τ. In fact, $p_{\mathbf{S}}(\tau)$ is the description of $\mathbf{S}$ and independent of the adversary $\mathscr{A}$. Then we describe the H-coefficient technique [7,12]. Generically, it considers an adversary that aims at distinguishing a "real" system $\mathbf{S}_1$ from an"ideal" system $\mathbf{S}_0$. The interactions of the adversary with those two systems induce two transcript distributions D_1 and D_0 respectively. It is well known that the statistical distance $\mathsf{SD}(D_0, D_1)$ is an upper bound on the distinguishing advantage of $\mathscr{A}$.

Lemma 1. [7,12] *Suppose that the set of attainable transcripts for the ideal system can be partitioned into good and bad ones. If there exists* $\epsilon \geq 0$ *such that* $\frac{p_{S_1}(\tau)}{p_{S_0}(\tau)} \geq 1-\epsilon$ *for any good transcript* τ*, then*

$$\mathsf{SD}\left(D_0, D_1\right) \leq \epsilon + \Pr[D_0 \text{ is bad}].$$

Regular and AU. Let $H : \mathcal{K}_h \times \mathcal{X} \to \mathcal{Y}$ be a hash function where $\mathcal{K}_h$ is the key space, $\mathcal{X}$ is the domain and $\mathcal{Y}$ is the range. Hash function H^i is said to be ϵ_1-regular if for any $X \in \mathcal{X}$, $Y \in \mathcal{Y}$,

$$\Pr[K_h \overset{\$}{\leftarrow} \mathcal{K}_h : H_{K_h}(X) = Y] \leq \epsilon_1.$$

And hash function H is said to be ϵ_2-AU if for any two distinct strings $X, X^{'} \in \mathcal{X}$,

$$\Pr[K_h \overset{\$}{\leftarrow} \mathcal{K}_h : H_{K_h}(X) = H_{K_h}(X^{'})] \leq \epsilon_2.$$

3 BBB-security Framework in [14]

Let $\mathcal{M}$ be the message space and $\mathcal{K}_h \times \mathcal{K}$ be the key space. Let blockcipher $E : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ and $\mathcal{K} = \{0,1\}^k$. Let hash function $H : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n \times \{0,1\}^n$. The function H is consist of two n-bit hash functions H^1 and H^2, i,e., $H_{K_h}(M) = (H^1_{K_{h,1}}(M), H^2_{K_{h,2}}(M))$ where $K_h = (K_{h,1}, K_{h,2}) \in \mathcal{K}_{h,1} \times \mathcal{K}_{h,2}$ and $K_{h,1}, K_{h,2}$ are two independent keys. Then the two-key DbHtS framework in paper [14] (see Fig.1) is

$$\mathrm{DbHtS}[H, E]\left(K_h, K, M\right) = E_K\left(H^1_{K_{h,1}}(M)\right) \oplus E_K\left(H^2_{K_{h,2}}(M)\right).$$

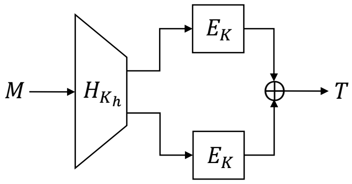

Fig. 1. The two-key DbHtS construction. Here H is a $2n$-bit hash function from $\mathcal{K}_h \times \mathcal{M}$ to $\{0,1\}^n \times \{0,1\}^n$, and E is a n-bit blockcipher from $\mathcal{K} \times \{0,1\}^n$ to $\{0,1\}^n$.

Theorem 1 in [14]. Let E be modeled as an ideal cipher. Let H^1 and H^2 both satisfy ϵ_1-regular and ϵ_2-AU. Then Shen et al. [14] proved the security of two-key DbHtS in the multi-user setting as following, which is the core of their paper and they named it Theorem 1. For any adversary $\mathscr{A}$ that makes at most q evaluation queries and p ideal-cipher queries,

$$\begin{aligned}\mathrm{Adv}^{\mathrm{prf}}_{\mathrm{DbHtS}}(\mathscr{A}) \leq & \frac{2q}{2^k} + \frac{q(3q+p)(6q+2p)}{2^{2k}} + \frac{2qp\ell}{2^{n+k}} + \frac{2qp\epsilon_1}{2^k} + \frac{4qp}{2^{n+k}} \\ & + \frac{4q^2\epsilon_1}{2^k} + \frac{2q^2\ell\epsilon_1}{2^k} + 2q^3\left(\epsilon_1+\epsilon_2\right)^2 + \frac{8q^3\left(\epsilon_1+\epsilon_2\right)}{2^n} + \frac{6q^3}{2^{2n}}\end{aligned} \tag{1}$$

where ℓ is the maximal block length among these evaluation queries and assuming $p + q\ell \le 2^{n-1}$.

An overview of the proof of Theorem 1 in [14]. They proved Theorem 1 based on H-coefficient technique. Let $\mathbf{S}_1$ be "real" system and $\mathbf{S}_0$ be "ideal" system. For $b \in \{0,1\}$, system $\mathbf{S}_b$ performs the following procedure.

1. Initialize $(K_h^1, K_1), \ldots, (K_h^u, K_u) \overset{\$}{\leftarrow} \mathcal{K}_h \times \mathcal{K}$ if $b = 1$; otherwise, initialize $f_1, \ldots, f_u \overset{\$}{\leftarrow} \mathsf{Func}(\mathcal{M}, \{0,1\}^n)$;
2. If an adversary $\mathscr{A}$ queries Eval function with (i, M), where $i \in \{1, 2, \ldots\}$, $M \in \mathcal{M}$, return
$$\mathsf{Eval}(i, M) = \begin{cases} \mathrm{DbHtS}[H, E](K_h^i, K_i, M), & \text{if } b = 1, \\ f_i(M), & \text{if } b = 0; \end{cases}$$
3. If an adversary $\mathscr{A}$ queries Prim function with (J, X), where $J \in \mathcal{K}, X \in \{+,-\} \times \{0,1\}^n$, return
$$\mathsf{Prim}(J, X) = \begin{cases} E_J(x), & \text{if } X = \{+, x\}, \\ E_J^{-1}(y), & \text{if } X = \{-, y\}. \end{cases}$$

They called the query to Eval evaluation query and the query to Prim ideal-cipher query. For each query $T \leftarrow \mathsf{Eval}(i, M)$, they associated it with an entry $(eval, i, M, T)$. The query to Prim is similar to it. These two entries are included in transcript τ. Then they defined bad transcripts, including fourteen cases. If a transcript is not bad then they said it's good. Let D_1 and D_0 be the random variables for the transcript distributions in the system $\mathbf{S}_1$ and $\mathbf{S}_0$ respectively. They firstly bounded the probability that D_0 is bad as follows. Let Bad_i be the event that the i-th case of bad transcripts happens. They calculated the probability $\Pr[\mathrm{Bad}_1], \ldots, \Pr[\mathrm{Bad}_{14}]$ in sequence. After summing up, they got

$$\begin{aligned} \Pr\left[D_0 \text{ is bad }\right] \le& \sum_{i=1}^{14} \Pr\left[\mathrm{Bad}_i\right] \\ \le& \frac{2q}{2^k} + \frac{q(3q+p)(6q+2p)}{2^{2k}} + \frac{2qp\ell}{2^{k+n}} + \frac{2qp\epsilon_1}{2^k} + \frac{4qp}{2^{n+k}} \\ &+ \frac{4q^2\epsilon_1}{2^k} + \frac{2q^2\ell\epsilon_1}{2^k} + 2q^3\left(\epsilon_1 + \epsilon_2\right)^2 + \frac{8q^3\left(\epsilon_1 + \epsilon_2\right)}{2^n}. \end{aligned}$$

Besides, they proved the transcript ratio $\frac{\mathrm{p}_{\mathbf{S}_1}(\tau)}{\mathrm{p}_{\mathbf{S}_0}(\tau)} \ge 1 - \frac{6q^3}{2^{2n}}$ for any good transcript τ. Thus they concluded Theorem 1 in [14] by Lemma 1.

4 Counter-examples

We show that the regular and universal properties of hash function can not guarantee the BBB security of two-key DbHtS MACs. We construct three universal hash functions which satisfy the properties, leading to $2n/3$-bit security of two-key DbHtS MACs by Theorem 1 in [14]. But all the instantiations can be broken with probability 1 using only $\mathcal{O}(2^{n/2})$ queries even in the single-user setting.

4.1 Counter-example 1

Our first counter-example is a universal hash function with fixed input length. Let function

$$H_{K_h}(M) = (H^1_{K_1}(M), H^2_{K_2}(M)) = (M \oplus K_1, M \oplus K_2),$$

where M is the message from massage space $\{0,1\}^n$, $K_h = (K_1, K_2)$ and $K_1, K_2 \xleftarrow{\$} \{0,1\}^n$. Let blockcipher $E_K : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$. Then the derived two-key DbHtS MAC is $F : \{0,1\}^{2n} \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ as

$$F[H,E](K_h, K, M) = E_K(H^1_{K_1}(M)) \oplus E_K(H^2_{K_2}(M)).$$

H^1 and H^2 are $\frac{1}{2^n}$-regular and 0-AU. It is easy to know that for any $M \in \{0,1\}^n, Y \in \{0,1\}^n$ and $i \in \{1,2\}$,

$$\Pr[K_i \xleftarrow{\$} \{0,1\}^n : M \oplus K_i = Y] \le \frac{1}{2^n}.$$

And for any two distinct strings $M, M^{'} \in \{0,1\}^n$ and $i \in \{1,2\}$,

$$\Pr[K_i \xleftarrow{\$} \{0,1\}^n : M \oplus K_i = M^{'} \oplus K_i] = 0.$$

So hash functions H^1 and H^2 are both $\frac{1}{2^n}$-regular and 0-AU.

$2n/3$-bit security. According to Theorem 1 [14], function F is secure up to $\mathcal{O}(2^{2n/3})$ evaluation queries assuming ideal-cipher queries is $\mathcal{O}(1)$ in the multi-user setting.

Attack with $\mathcal{O}(2^{n/2})$ query complexity. It is easy to know that for all keys in keyspace and messages in message space,

$$\begin{aligned} F[H,E](K_h, K, M \oplus K_1 \oplus K_2) &= E_K(M \oplus K_2) \oplus E_K(M \oplus K_1) \\ &= F[H,E](K_h, K, M). \end{aligned}$$

It means F has a period $s := K_1 \oplus K_2$. Based on this, there is an adversary $\mathscr{A}$ can distinguish F from random function f with only $\mathcal{O}(2^{n/2})$ evaluation queries as follows, which is contradictory to Theorem 1 [14].

1. $\mathscr{A}$ firstly makes $\mathcal{O}(2^{n/2})$ evaluation queries of distinct massages $M_1, M_2, \ldots$ chosen uniformly and randomly, and get $T_1, T_2, \ldots$;
2. $\mathscr{A}$ outputs 1 if it searches a message pair (M_i, M_j) for $M_i \ne M_j, M_i, M_j \in \{M_1, M_2, \ldots\}$ which makes (i) and (ii) hold.
   (i) $T_i = T_j$;
   (ii) After make another two evaluation queries with massages M' and $M' \oplus M_i \oplus M_j$ for $M' \notin \{M_i, M_j\}$, $\mathscr{A}$ gets two identical answers.
   Else, $\mathscr{A}$ outputs 0.

If the evaluation query is to F, one can expect on average that there exists one message pair (M_i, M_j) among $\mathcal{O}(2^{n/2})$ massages such that $M_i = M_j \oplus s$. Conditions (i) and (ii) in the second step of $\mathscr{A}$ filter out such pair. However, random function f has no period. If the evaluation query is to f, on average there exists one message pair (M_i, M_j) among $\mathcal{O}(2^{n/2})$ massages such that $T_i = T_j$. However, the probability of $f(M') = f(M' \oplus M_i \oplus M_j)$ for any $M' \notin \{M_i, M_j\}$ is only $1/2^n$. So $\mathscr{A}$ finds a pair (M_i, M_j) satisfying conditions (i) and (ii) with negligible probability. Thus $\mathscr{A}$ distinguish F from random function with probability $1 - 1/2^n$.

4.2 Counter-example 2

Compared with the first counter-example with fixed input length, our second counter-example can handle variable-length input. We construct two hash functions H^1 and H^2 dealing with messages from $(\{0,1\}^n)^*$:

$$H^i_{K_i}(M) = M[1] \oplus M[2]K_i \oplus M[3]K_i^2 \oplus \ldots \oplus M[m]K_i^{m-1} \oplus |M|K_i^m, i = 1, 2.$$

where $M = M[1] \parallel M[2] \parallel \ldots \parallel M[m]$ and every message block is n-bit. This example is a variant of PolyMAC [9].

H^1 and H^2 are $\frac{\ell}{2^n}$-regular and $\frac{\ell}{2^n}$-AU. Assume the maximal block length of all evaluation queries is ℓ. Any equation of at most ℓ degree has at most ℓ roots. So it is easy to know that for any $M \in (\{0,1\}^n)^*, Y \in \{0,1\}^n$ and $i \in \{1,2\}$,

$$\Pr[K_i \xleftarrow{\$} \{0,1\}^n : H^i_{K_i}(M) = Y] \leq \frac{\ell}{2^n}.$$

And for any two distinct strings $M, M^{'} \in (\{0,1\}^n)^*$ and $i \in \{1,2\}$,

$$\Pr[K_i \xleftarrow{\$} \{0,1\}^n : H^i_{K_i}(M) = H^i_{K_i}(M^{'})] \leq \frac{\ell}{2^n}.$$

It means H^1 and H^2 are both $\frac{\ell}{2^n}$-regular and $\frac{\ell}{2^n}$-AU.

$2n/3$-bit security. According to Theorem 1 [14], function F is secure up to $\mathcal{O}(2^{2n/3})$ evaluation queries assuming ideal-cipher queries is $\mathcal{O}(1)$ and $\ell = \mathcal{O}(1)$ in the multi-user setting.

Attack with $\mathcal{O}(2^{n/2})$ query complexity. Fix any arbitrary string

$$M_{fix} := M[2] \| M[3] \| \ldots \| M[m] \in (\{0,1\}^n)^{m-1},$$

where $2 \leq m \leq \ell = O(1)$. Let

$$K_i^{'} := M[2]K_i \oplus M[3]K_i^2 \oplus \ldots M[m]K_i^{m-1} \oplus nmK_i^m, i = 1, 2.$$

Then it is easy to obtain for any keys in key space and $M[1] \in \{0,1\}^n$,

$$\begin{aligned}
&F[H,E](K_h, K, (M[1] \oplus K_1^{'} \oplus K_2^{'}) \parallel M_{fix}) \\
=&E_K(M[1] \oplus K_2^{'}) \oplus E_K(M[1] \oplus K_1^{'}) \\
=&F[H,E](K_h, K, M[1] \parallel M_{fix}).
\end{aligned}$$

It means F has a period $s := (K_1^{'} \oplus K_2^{'}) \parallel 0^{n(m-1)}$ for any $M \in \{0,1\}^n \times \{M_{fix}\}$. Based on this, there is an adversary $\mathscr{A}$ who can distinguish F from random function f with only $\mathcal{O}(2^{n/2})$ evaluation queries as follows, which is contradictory to Theorem 1 [14].

1. $\mathscr{A}$ firstly makes $\mathcal{O}(2^{n/2})$ evaluation queries with distinct massages $M_1 \parallel M_{fix}$, $M_2 \parallel M_{fix}, \dots$ where $M_1, M_2, \dots \xleftarrow{\$} \{0,1\}^n$, and get $T_1, T_2, \dots$;
2. $\mathscr{A}$ outputs 1 if it searches a pair (M_i, M_j) for $M_i \neq M_j, M_i, M_j \in \{M_1, M_2, \dots\}$ which makes (i) and (ii) hold.
   (i) $T_i = T_j$;
   (ii) After make another two evaluation queries with massages $M' \parallel M_{fix}$ and $(M' \oplus M_i \oplus M_j) \parallel M_{fix}$ for $M' \notin \{M_i, M_j\}$, $\mathscr{A}$ gets two identical answers.

   Else, $\mathscr{A}$ outputs 0.

The same as counter-example 1, $\mathscr{A}$ distinguishes F from f with probability almost 1.

4.3 Counter-example 3

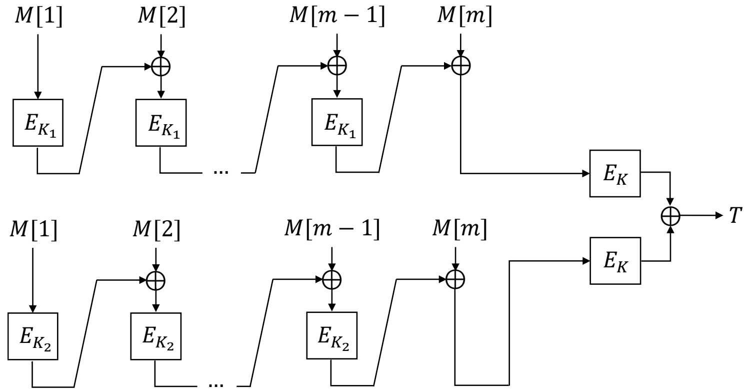

Fig. 2. The variant of 2k-SUM-ECBC. K_1, K_2, K_3 are three independent keys in $\{0,1\}^n$. E is a n-bit blockcipher from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}^n$.

.

Unlike counter-examples 1 and 2, the third counter-example is based on blockciphers. Let $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ be a blockcipher with key $K \in \{0,1\}^n$. The two n-bit hash functions used in this function are two CBC MACs without the last blockcipher call, which we name $\mathrm{CBC}^{'}$. They are keyed with two independent keys K_1 and K_2 respectively. And they deal with at least two message blocks respectively. For a message $M = M[1] \parallel M[2] \parallel \dots \parallel M[m]$ where every message block is n-bit and $m \geq 2$, the $\mathrm{CBC}^{'}$ algorithm $\mathrm{CBC}^{'}[E](K, M)$ is defined as Y_m, where

$$\begin{aligned} &Y_1 = M[1], \\ &Y_j = E_K(Y_{j-1}) \oplus M[j], j = 2, \ldots, m. \end{aligned}$$

Let $K_h = (K_1, K_2)$. Then we define the function as

$$F[\mathrm{CBC}^{'}[E], E](K_h, K, M) = E_K(\mathrm{CBC}^{'}[E](K_1, M)) \oplus E_K(\mathrm{CBC}^{'}[E](K_2, M)).$$

F (see Fig.2) can be seen as a variant of 2k-SUM-ECBC [14].

$\mathrm{CBC}^{'}[E]$ is $\left(\frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}\right)$-regular and $\left(\frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}\right)$-AU. For any two different message $M, M^{'} \in (\{0,1\}^n)^*$ with at most ℓ blocks and the blockcipher E_K being a random permutation, Ballare et al. [3] showed that for $i \in \{1, 2\}$,

$$\Pr[E_K(\mathrm{CBC}^{'}[E](K_i, M)) = E_K(\mathrm{CBC}^{'}[E](K_i, M^{'}))] \leq \frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}.$$

Blockcipher E_K is a permutation. So

$$\Pr[\mathrm{CBC}^{'}[E](K_i, M) = \mathrm{CBC}^{'}[E](K_i, M^{'})] \leq \frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}.$$

Thus for ideal blockcipher E we get

$$\Pr[K_i \xleftarrow{\$} \{0,1\}^n : \mathrm{CBC}^{'}[E](K_i, M) = \mathrm{CBC}^{'}[E](K_i, M^{'})] \leq \frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}.$$

It means $\mathrm{CBC}^{'}$ is $\left(\frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}\right)$-AU. Let $M = X[1] \parallel (X[2] \oplus Y) \parallel Z \in (\{0,1\}^n)^* \times \{0,1\}^n \times \{0,1\}^n$ and $M^{'} = 0^n \parallel Z \in \{0,1\}^n \times \{0,1\}^n$. Then

$$\begin{aligned} &\Pr[K_i \xleftarrow{\$} \{0,1\}^n : \mathrm{CBC}^{'}[E](K_i, X[1] \parallel X[2]) = Y] \\ =&\Pr[K_i \xleftarrow{\$} \{0,1\}^n : \mathrm{CBC}^{'}[E](K_i, M) = \mathrm{CBC}^{'}[E](K_i, M^{'})] \\ \leq&\frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}. \end{aligned}$$

So $\mathrm{CBC}^{'}$ is $\left(\frac{2\ell}{2^n} + \frac{16\ell^4}{2^{2n}}\right)$-regular.

$2n/3$-bit security. According to Theorem 1 [14], function F is secure up to $\mathcal{O}(2^{2n/3})$ evaluation queries assuming no ideal-cipher queries and $\ell = \mathcal{O}(1)$ in the multi-user setting.

Attack with $\mathcal{O}(2^{n/2})$ query complexity. Fix any arbitrary string $M_{fix} \in (\{0,1\}^n)^{m-1}$ where $2 \leq m \leq \ell = O(1)$. Let

$$s' = \mathrm{CBC}^{'}[E](K_1, M_{fix} \parallel 0^n) \oplus \mathrm{CBC}^{'}[E](K_2, M_{fix} \parallel 0^n)).$$

Then it is easy to obtain for any keys in key space and $M[m] \in \{0,1\}^n$,

$$\begin{aligned}
&F[\mathrm{CBC}^{'}[E], E](K_h, K, M_{fix} \parallel (M[m] \oplus s'))\\
=&E_K(\mathrm{CBC}^{'}[E](K_2, M_{fix} \parallel 0^n) \oplus M[m]) \oplus\\
&E_K(\mathrm{CBC}^{'}[E](K_1, M_{fix} \parallel 0^n) \oplus M[m])\\
=&E_K(\mathrm{CBC}^{'}[E](K_2, M_{fix} \parallel M[m])) \oplus E_K(\mathrm{CBC}^{'}[E](K_1, M_{fix} \parallel M[m]))\\
=&F[\mathrm{CBC}^{'}[E], E](K_h, K, M_{fix} \parallel M[m]).
\end{aligned}$$

It means F has a period $s := 0^{n(m-1)} \parallel s'$ for any $M \in \{M_{fix}\} \times \{0,1\}^n$. So there is an adversary $\mathscr{A}$ distinguishes F from random function with only $\mathcal{O}(2^{n/2})$ evaluation queries when considering single user similar as counter-example 2.

5 The Flaw of the Proof of Theorem 1 in [14]

In section 3, we have shown the procedure of how Shen et al. [14] proved Theorem 1 based on H-coefficient technique. However, we find they make a critical flaw when they were calculating $\Pr[\mathrm{Bad}_9]$ in their proof, which leads to our counter-examples.

Assume there are u users and the adversary make q_i evaluation queries to the i-th user in all. Let $(eval, i, M_a^i, T_a^i)$ be the entry obtained when the adversary makes the a-th query to user i. During the computation of entry $(eval, i, M_a^i, T_a^i)$, let Σ_a^i and Λ_a^i be the internal outputs of hash function H, namely $\Sigma_a^i = H^1_{K_{h,1}}\left(M_a^i\right)$ and $\Lambda_a^i = H^2_{K_{h,2}}\left(M_a^i\right)$ respectively, and denote by U_a^i and V_a^i the outputs of blockcipher E with inputs Σ_a^i and Λ_a^i respectively, namely $U_a^i = E\left(K_i, \Sigma_a^i\right)$ and $V_a^i = E\left(K_i, \Lambda_a^i\right)$ respectively. The adversary also obtain the entries of ideal-cipher queries. After all queries, they further give it: (i) the keys $\left(K_h^i, K_i\right)$ where $K_h^i = \left(K_{h,1}^i, K_{h,2}^i\right)$ and (ii) the internal values U_a^i and V_a^i. In the "ideal" system $\mathbf{S}_0$, they instead give the adversary truly random strings $\left(K_h^i, K_i\right) \xleftarrow{\$} \mathcal{K}_h \times \mathcal{K}$, independent of its queries. In addition, they give the adversary dummy values U_a^i and V_a^i computed by an simulation oracle. These additional information can only help the adversary. Thus a transcript consists of the revealed keys $\left(K_h^i, K_i\right)$, the internal values U_a^i and V_a^i, the ideal-cipher queries and evaluation queries.

The ninth bad event is

"There is an entry $(eval, i, M_a^i, T_a^i)$ *such that either* $\Sigma_a^i = \Sigma_b^i$ *or* $\Sigma_a^i = \Lambda_b^i$*, and either* $\Lambda_a^i = \Lambda_b^i$ *or* $\Lambda_a^i = \Sigma_b^i$ *for some entry* $(eval, i, M_a^i, T_a^i)$*."*

They defined this event bad for the reason that the appearance of such bad event is easy used to distinguish systems $\mathbf{S}_1$ and $\mathbf{S}_0$. We call the event of either $\Sigma_a^i = \Sigma_b^i$ or $\Sigma_a^i = \Lambda_b^i$ as event 1, and the event of either $\Lambda_a^i = \Lambda_b^i$ or $\Lambda_a^i = \Sigma_b^i$ as event 2. Then we can regard the simultaneous events 1 and 2 as one of the following 4 events:

- Event 3: $\Sigma_a^i = \Sigma_b^i \wedge \Lambda_a^i = \Lambda_b^i$;

- Event 4: $\Sigma_a^i = \Sigma_b^i \wedge \Lambda_a^i = \Sigma_b^i$;
- Event 5: $\Sigma_a^i = \Lambda_b^i \wedge \Lambda_a^i = \Lambda_b^i$;
- Event 6: $\Sigma_a^i = \Lambda_b^i \wedge \Lambda_a^i = \Sigma_b^i$.

In "real" system $\mathbf{S}_1$, event 4 or 5 leads to $T_a^i = 0^n$; event 3 or 6 leads to $T_a^i = T_b^i$. However in "ideal" system $\mathbf{S}_0$ these happen with negligible probability by the randomness of random function f_i. And in "real" system $\mathbf{S}_1$, $U_a^i = E\left(K_i, \Sigma_a^i\right)$ and $V_a^i = E\left(K_i\right)$. However in "ideal" system $\mathbf{S}_0$, $(U_a^i, V_a^i) = \perp$ by the simulation oracle defined by them. Thus it is easy distinguish these two systems.

When calculating $\Pr[\text{Bad}_9]$ in "ideal" system $\mathbf{S}_0$, Shen et al. [14] regarded that the event 1 is independent from event 2 when $K_{h,1}^i, K_{h,2}^i$ are independent from each other. So by H^1, H^2 are both ϵ_1-regular and ϵ_2-AU, they thought the probability of event 1 (resp. event 2) is at most $\epsilon_1 + \epsilon_2$. Note that for each user, there are at most q_i^2 pairs of (a, b). So they summed among u users and got

$$\Pr[\text{Bad}_9] \leq \Sigma_{i=1}^u q_i^2 (\epsilon_1 + \epsilon_2)^2 \leq q^2 (\epsilon_1 + \epsilon_2)^2.$$

In fact, even if $K_{h,1}^i, K_{h,2}^i$ are independent of each other, the event 1 and event 2 may not be independent, which has been shown in counter-examples 1-3. We regard the ninth event as the union set of events 3,4,5 and 6. Event 3 holds with probability at most ϵ_2^2 by the assumption that H^1 and H^2 are ϵ_2-AU. Event 4 holds with probability at most $\epsilon_1\epsilon_2$ by the assumption that H^1 is ϵ_2-AU and H^2 is ϵ_1-regular. Event 5 holds with probability at most $\epsilon_1\epsilon_2$ by the assumption that H^1 is ϵ_1-regular and H^2 is ϵ_2-AU. For event 6,

$$\begin{aligned} &\Pr[K_{h,1}^i \xleftarrow{\$} \mathcal{K}_{h,1}, K_{h,2}^i \xleftarrow{\$} \mathcal{K}_{h,2} : \Sigma_a^i = \Lambda_b^i \wedge \Lambda_a^i = \Sigma_b^i] \\ =&\Pr[K_{h,1}^i \xleftarrow{\$} \mathcal{K}_{h,1}, K_{h,2}^i \xleftarrow{\$} \mathcal{K}_{h,2} : \Sigma_a^i = \Lambda_b^i | \Lambda_a^i = \Sigma_b^i] \\ &\cdot \Pr[K_{h,1}^i \xleftarrow{\$} \mathcal{K}_{h,2}, K_{h,2}^i \xleftarrow{\$} \mathcal{K}_{h,1} : \Lambda_a^i = \Sigma_b^i] \\ \leq& \epsilon_3\epsilon_1 \end{aligned}$$

by the assumption that H^2 is ϵ_1-regular and let

$$\epsilon_3 = \Pr[K_{h,1}^i \xleftarrow{\$} \mathcal{K}_{h,1}, K_{h,2}^i \xleftarrow{\$} \mathcal{K}_{h,2} : \Sigma_a^i = \Lambda_b^i | \Lambda_a^i = \Sigma_b^i].$$

So we sum among u users and got

$$\Pr[\text{Bad}_9] \leq \Sigma_{i=1}^u q_i^2 (\epsilon_2^2 + 2\epsilon_1\epsilon_2 + \epsilon_3\epsilon_1) \leq q^2 (\epsilon_2^2 + 2\epsilon_1\epsilon_2 + \epsilon_3\epsilon_1).$$

For counter-examples 1-3, it is easy to get $\epsilon_3 = 1$. So for these cases, $\Pr[\text{Bad}_9] \leq q^2(\epsilon_2^2 + 2\epsilon_1\epsilon_2 + \epsilon_1)$. If we substitute our $\Pr[\text{Bad}_9]$ for that in paper [14], we get the security of proofs of counter-examples 1-3 should be up to $\mathcal{O}(2^{n/2})$ evaluation queries assuming ideal-cipher queries are $\mathcal{O}(1)$ and the maximal block length of all evaluation queries is $\mathcal{O}(1)$, which is consistent with attacks.

6 Conclusion

In this paper, we point out a flaw of the security framework for two-key DbHtS in the multi-user setting proposed by Shen et al. [14] by constructing three counter-examples. We also analyze how the flaw happens in their proof. This is due to the fact that the authors overlooked the dependence of $\varSigma_a^i = \varLambda_b^i$ and $\varLambda_a^i = \varSigma_b^i$ in the proof of Theorem 1 [14]. In their paper, they also stated 2k-SUM-ECBC, 2k-PMAC_Plus, and 2k-LightMAC_Plus all achieve $2n/3$-bit security. For 2k-SUM-ECBC based on two independent CBC MACs, the probability ϵ_3 is about $\frac{1}{2^n}$. So if we substitute our $\Pr[\mathrm{Bad}_9]$ for that in paper [14], 2k-SUM-ECBC still achieves $2n/3$ security. The two universal hash functions of 2k-PMAC_Plus or 2k-LightMAC_Plus are dependent, they adjusted the concrete proof of these two MACs from the framework. We haven't found attacks against these three MACs.

Recently, Shen et al. refined their paper [15] because of what we have found in this paper. Their new framework for two-key DbHtS is not universal. Because the $2n/3$-bit security of two-key DbHtS MACs doesn't only come from the regular and AU properties of the hash functions any more. To be specific, they added two variables to capture the probabilities of two subcases '$\varSigma_a^i = \varLambda_b^i \wedge \varLambda_a^i = \varSigma_b^i$' and '$\varSigma_a^i = \varLambda_b^i \wedge \varLambda_a^i = \varSigma_c^i$', the values of which will be clear until in the analysis of concrete MAC. In fact, theses two added subcases have been included in the cover-free property of H by Datta et al. [8] when they considered the framework of two-key DbHtS in the single user setting.

References

1. Banik, S., Pandey, S.K., Peyrin, T., Sasaki, Y., Sim, S.M., Todo, Y.: GIFT: A small present - towards reaching the limit of lightweight encryption. In: CHES 2017. vol. 10529, pp. 321–345. Springer (2017), https://doi.org/10.1007/978-3-319-66787-4_16 1
2. Bellare, M.: New proofs for NMAC and HMAC: security without collision-resistance. In: CRYPTO 2006, Proceedings. LNCS, vol. 4117, pp. 602–619. Springer (2006), https://doi.org/10.1007/11818175_36 1
3. Bellare, M., Pietrzak, K., Rogaway, P.: Improved security analyses for CBC MACs. In: CRYPTO 2005. vol. 3621, pp. 527–545. Springer (2005), https://doi.org/10.1007/11535218_32 9
4. Black, J., Rogaway, P.: CBC MACs for arbitrary-length messages: The three-key constructions. In: CRYPTO 2000, Proceedings. LNCS, vol. 1880, pp. 197–215. Springer (2000), https://doi.org/10.1007/3-540-44598-6_12 1
5. Black, J., Rogaway, P.: A block-cipher mode of operation for parallelizable message authentication. In: EUROCRYPT 2002, Proceedings. LNCS, vol. 2332, pp. 384–397. Springer (2002), https://doi.org/10.1007/3-540-46035-7_25 1
6. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: an ultra-lightweight block cipher. In: CHES 2007. vol. 4727, pp. 450–466. Springer (2007), https://doi.org/10.1007/978-3-540-74735-2_31 1

7. Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. In: EUROCRYPT 2014. vol. 8441, pp. 327–350. Springer (2014), https://doi.org/10.1007/978-3-642-55220-5_19 3, 4
8. Datta, N., Dutta, A., Nandi, M., Paul, G.: Double-block hash-then-sum: A paradigm for constructing BBB secure PRF. IACR Trans. Symmetric Cryptol. **2018**(3), 36–92 (2018), https://doi.org/10.13154/tosc.v2018.i3.36-92 2, 3, 12
9. Kim, S., Lee, B., Lee, J.: Tight security bounds for Double-Block Hash-then-Sum MACs. In: EUROCRYPT 2020. vol. 12105, pp. 435–465. Springer (2020), https://doi.org/10.1007/978-3-030-45721-1_16 2, 3, 7
10. Leurent, G., Nandi, M., Sibleyras, F.: Generic attacks against beyond-birthday-bound macs. In: CRYPTO 2018. vol. 10991, pp. 306–336. Springer (2018), https://doi.org/10.1007/978-3-319-96884-1_11 2
11. Naito, Y.: Blockcipher-based macs: Beyond the birthday bound without message length. In: ASIACRYPT 2017. vol. 10626, pp. 446–470. Springer (2017), https://doi.org/10.1007/978-3-319-70700-6_16 2
12. Patarin, J.: The "coefficients H" technique. In: SAC 2008. vol. 5381, pp. 328–345. Springer (2008), https://doi.org/10.1007/978-3-642-04159-4_21 3, 4
13. Rogaway, P.: Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In: ASIACRYPT 2004, Proceedings. LNCS, vol. 3329, pp. 16–31. Springer (2004), https://doi.org/10.1007/978-3-540-30539-2_2 1
14. Shen, Y., Wang, L., Gu, D., Weng, J.: Revisiting the security of DbHtS MACs: Beyond-birthday-bound in the multi-user setting. In: CRYPTO 2021. vol. 12827, pp. 309–336. Springer (2021), https://doi.org/10.1007/978-3-030-84252-9_11 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
15. Shen, Y., Wang, L., Weng, J.: Revisiting the security of DbHtS MACs: Beyond-birthday-bound in the multi-user setting. IACR Cryptol. ePrint Arch. p. 1523 (2020), https://eprint.iacr.org/2020/1523 12
16. Shoup, V.: Sequences of games: a tool for taming complexity in security proofs. IACR Cryptol. ePrint Arch. p. 332 (2004), http://eprint.iacr.org/2004/332 1, 3
17. Yasuda, K.: The sum of CBC MACs is a secure PRF. In: CT-RSA 2010. vol. 5985, pp. 366–381. Springer (2010), https://doi.org/10.1007/978-3-642-11925-5_25 2
18. Yasuda, K.: A new variant of PMAC: beyond the birthday bound. In: CRYPTO 2011. vol. 6841, pp. 596–609. Springer (2011), https://doi.org/10.1007/978-3-642-22792-9_34 2
19. Zhang, L., Wu, W., Sui, H., Wang, P.: 3kf9: Enhancing 3GPP-MAC beyond the birthday bound. In: ASIACRYPT 2012. vol. 7658, pp. 296–312. Springer (2012), https://doi.org/10.1007/978-3-642-34961-4_19 2