Tight Multi-User Security Bound of DbHtS
(Long Paper)

Nilanjan Datta¹ and Avijit Dutta¹ and Mridul Nandi² and Suprita Talnikar²

Institute for Advancing Intelligence, TCG-CREST, Kolkata
Indian Statistical Institute, Kolkata.
nilanjan.datta@tcgcrest.org,avirocks.dutta13@gmail.com,mridul.nandi@gmail.com,
suprita45@gmail.com

Abstract. In CRYPTO’21, Shen et al. have proved in the ideal cipher model that Two-Keyed_DbHtS construction is secure up to $2^{2n/3}$ queries in the multi-user setting independent of the number of users, where the underlying double-block hash function $H$ of the Two-Keyed_DbHtS construction is realized as the concatenation of two independent $n$-bit keyed hash functions $(H_{K_1}, H_{K_2})$ such that each of the $n$-bit keyed hash function is $O(2^{-n})$ universal and regular. They have also demonstrated the applicability of their result to the key-reduced variants of DbHtS MACs, including 2K-SUM-ECBC, 2K-PMAC_Plus and 2K-LightMAC_Plus without requiring domain separation technique and proved $2n/3$-bit multi-user security of these constructions in the ideal cipher model. Recently, Guo and Wang have invalidated the security claim of Shen et al.’s result by exhibiting three constructions, which are the instantiations of the Two-Keyed_DbHtS framework, such that each of their $n$-bit keyed hash functions being $O(2^{-n})$ universal and regular, while the constructions themselves are secure only up to the birthday bound. In this work, we show a sufficient condition on the underlying Double-block Hash ($DbH$) function, under which we prove $3n/4$-bit multi-user security of the Two-Keyed_DbHtS construction in the ideal-cipher model. As an instantiation, we show that two-keyed Polyhash-based DbHtS construction is multi-user secure up to $2^{3n/4}$ queries in the ideal-cipher model. Furthermore, due to the generic attack on DbHtS constructions by Gaëtan et al. in CRYPTO’18, our derived bound for the construction is tight.

Keywords: DbHtS · PRF · Polyhash · H-Coefficient Technique · Mirror Theory.

1 Introduction

Hash-then-PRF [33] (or HtP) is a well-known paradigm for designing variable input-length PRFs, in which an input message of arbitrary length is hashed and the hash value is encrypted through a PRF to obtain a short tag. Most popular MACs including the CBC-MAC [3], PMAC [9], OMAC [19] and LightMAC [23] are designed using the HtP paradigm. Although the method is simple, in particular being deterministic and stateless, the security of MACs following the HtP paradigm is capped at the birthday bound due to the collision probability of the hash function. Birthday bound-secure constructions are acceptable in practice when any of these MACs are instantiated with a block cipher of moderately large block size. For example, instantiating PMAC with AES-128 permits roughly $2^{48}$ queries (using $56q^2/2^n$ [30] bound) when the longest message size is $2^{16}$ blocks, and the success probability of breaking the scheme is restricted to $2^{-10}$. However, the same construction becomes vulnerable if instantiated with some lightweight (smaller block size) block ciphers, whose number has grown tremendously in recent years, e.g. PRESENT [10], GIFT [1], LED [15], etc. For example, PMAC, when instantiated with the PRESENT block cipher (a
64-bit block cipher), permits only about $2^{16}$ queries when the longest message size is $2^{16}$ blocks, and the probability of breaking the scheme is $2^{-10}$. Therefore, it becomes risky to use birthday bound-secure constructions instantiated with lightweight block ciphers. In fact, in a large number of financial sectors, web browsers still widely use 64-bit block ciphers 3-DES instead of AES in their legacy applications with backward compatibility feature, as using the latter in corporate mainframe computers is more expensive. However, it does not give adequate security if the mode in which 3-DES is used provides only birthday bound security, and hence a beyond birthday secure mode solves the issue. Although many secure practical applications use the standard AES-128, which provides 64-bit security in a birthday bound-secure mode, which is adequate for the current technology, it may not remain so in the near future. In such a situation, using a mode with beyond the birthday bound security instead of replacing the cipher with a larger block size is a better option.\footnote{Note that there are no standard block ciphers of size higher than 128 bits.}

**Double-Block Hash-then-Sum.** Many studies tried to tweak the HtP design paradigm to obtain beyond the birthday bound secure MACs; while they possess a similar structural design, the internal state of the hash function is doubled and the two $n$-bit hash values are first encrypted and then xored together to produce the output. In [35], Yasuda proposed a beyond the birthday bound secure deterministic MAC called SUM-ECBC, a rate-1/2 sequential mode of construction with four block cipher keys. Followed by this work, Yasuda [36] came up with another deterministic MAC called PMAC_Plus, but unlike SUM-ECBC, PMAC_Plus is a rate-1 parallel mode of construction with three block cipher keys. Zhang et al. [37] proposed another rate-1 beyond the birthday bound secure deterministic MAC called 3kf9 with three block cipher keys. In [29], Naito proposed LightMAC_Plus, a rate $(1 - s/n)$ parallel mode of operation, where $s$ is the size of the block counter. The structural design of all these constructions first applies a $2n$-bit hash function on the message, then the two $n$-bit output values are encrypted and xored together to produce the tag, where $n$ is the block size of the block cipher. Moreover, all of them also give $2n/3$-bit security. In FSE 2019, Datta et al. [13] proposed a generic design paradigm dubbed as the double-block hash-then-sum or DbHtS, defined as follows:

$$\mathrm{DbHtS}(M) \triangleq E_{K_1}(\Sigma) \oplus E_{K_2}(\Theta), \; (\Sigma, \Theta) \leftarrow H_{K_3}(M),$$

where $H_{K_3}$ is a double-block hash function that maps an arbitrary-length string to a $2n$-bit string. Within this unified framework, they revisited the security proof of existing DbHtS constructions, including PolyMAC [20], SUM-ECBC [35], PMAC_Plus [36], 3kf9 [37] and LightMAC_Plus [29] and also their two-keyed versions [13] and confirmed that all the constructions are secure up to $2^{2n/3}$ queries when they are instantiated with an $n$-bit block cipher.

In CRYPTO 2018, Gaëtan et al. [21] proposed a generic attack on all these constructions using $2^{3n/4}$ (short message) queries, leaving a gap between the upper and the lower bounds for the provable security of DbHtS constructions. Recently, Kim et al. [20] have improved the bound of DbHtS constructions from $2^{2n/3}$ to $2^{3n/4}$. They have shown that if the underlying $2n$-bit hash function is the concatenation of two independent $n$-bit-universal hash functions\footnote{A family of keyed hash functions is said to be universal if for any distinct $x$ and $x'$, the probability of a collision in their hash values for a randomly sampled hash function from the family is negligible.}, then the resulting DbHtS paradigm is secure up to $2^{3n/4}$ queries. They have also improved the security bound of PMAC_Plus, 3kf9 and LightMAC_Plus from $2^{2n/3}$ to $2^{3n/4}$ and hence closed the gap between the upper and the lower bounds of the provable security of DbHtS constructions.

**Multi-user security of DbHtS.** We have so far discussed the security bounds of DbHtS constructions in which adversaries are given access to some keyed oracles for a single unknown randomly sampled key. Such a model is known as the single-user
security model, i.e., when the adversary interacts with one specific machine in which the cryptographic algorithm is deployed and tries to compromise its security. However, in practice, cryptographic algorithms are usually deployed in more than one machine. For example, AES-GCM [24, 25] is now widely used in the TLS protocol to protect web traffic and is currently used by billions of users daily. Thus, the security of DbHtS constructions in the multi-key setting is worth investigating; in other words, we ask to what extent the number of users will affect the security of DbHtS constructions, where adversaries are successful if they compromise the security of one out of many users. That means the adversary’s winning condition is a disjunction of single-key winning conditions.

The notion of multi-user (mu) security was introduced by Biham [7] in symmetric cryptanalysis and by Bellare, Boldyreva, and Micali [2] in the context of public-key encryption. In the multi-user setting, attackers have access to multiple machines such that a particular cryptographic algorithm $F$ is deployed in each machine with independent secret keys. An attacker can adaptively distribute its queries across multiple machines with independent keys. Multi-user security considers attackers that succeed in compromising the security of at least one machine, among others.

Multi-user security for block ciphers is different from multi-user security for modes. In the single-key setting, the best attacks against block cipher such as AES do not improve with increased data complexity. However, in the multi-key environment, they do, as first observed by Biham [7] and later refined as a time-memory-data trade-off by Biryukov et al. [8]. These results demonstrate how one can take advantage of the fact that recovering a block cipher key out of a large group of keys is much easier than targeting a specific key. The same observation can be applied to any deterministic symmetric-key algorithm, as done for MACs by Chatterjee et al. [12]. A more general result guarantees that the multi-user advantage of an adversary for a cryptographic algorithm is at most $u$ times its single user advantage. Therefore, for any cryptographic algorithm, a multi-user security bound involving a factor $u$ is easily established using a hybrid argument that shows the upper bound of the adversarial success probability to be roughly $u$ times its single-user security advantage. Bellare and Tackmann [5] first formalized a multi-user secure authenticated encryption scheme and also analyzed countermeasures against multi-key attacks in the context of TLS 1.3. However, they derived a security bound that also contained the factor $u$. Such a bound implies a significant security drop of the construction when the number of users is large, and in fact, this is precisely the situation faced in large-scale deployments of AES-GCM such as TLS.

As evident from [4, 5, 11, 17, 18, 22, 28], it is a challenging problem to study the security degradation of cryptographic primitives with the number of users, even when its security is known in the single-user setting. Studies of multi-user security of MACs are somewhat scarce in the literature except for the work of Chatterjee et al. [12], and a very recent work of Andrew et al. [27], and Bellare et al. [4]. The first two consider a generic reduction for MACs, in which the security of the primitive in the multi-user setting is derived by multiplying the number of users $u$ by the single-user security advantage. Bellare and Tackmann [5] first formalized a multi-user secure authenticated encryption scheme and also analyzed countermeasures against multi-key attacks in the context of TLS 1.3. However, they derived a security bound that also contained the factor $u$. Such a bound implies a significant security drop of the construction when the number of users is large, and in fact, this is precisely the situation faced in large-scale deployments of AES-GCM such as TLS.

In CRYPTO’21, Shen et al. [32] have analyzed the security of DbHtS in the multi-user setting. It is worth noting here that by applying the generic reduction from the single-user to the multi-user setting, the security bound of DbHtS would have capped at worse than the birthday bound, i.e. $uq^{4/3}/2^n$, when each user made a single query and the number of users reached $q$. Thus, a direct analysis was needed for deriving the multi-user bound of the construction. Shen et al. [32] have shown that in the multi-user setting, the two-keyed $^3$ DbHtS paradigm,

$$\TwoKeyedDbHtS(M) \triangleq E_K(H_{K_1}(M)) \oplus E_K(H_{K_2}(M)),$$

$^3$two-keyed stands for one hash key and one block cipher key.
is secure up to $2^{2n/3}$ queries in the ideal-cipher model when the $2n$-bit double-block hash function is the concatenation of two independent $n$-bit keyed hash functions $H_{K_{\alpha,1}}$ and $H_{K_{\alpha,2}}$. In particular, they have shown that if both $H_{K_{\alpha,1}}$ and $H_{K_{\alpha,2}}$ are $O(2^{-n})$-regular and $O(2^{-n})$-universal, then the multi-user security bound of the two-keyed DbHtS is of the order of
\[
\frac{qp\ell}{2k + n} + \frac{q^3}{2^{2n}} + \frac{q^2p + qp^2}{2^{2k}},
\]
where $q$ is the total number of MAC queries across all $u$ users, $p$ is the total number of ideal-cipher queries, $\ell$ is the maximum number of message blocks among all queries, and $n, k$ are the block size and the number of blocks per block cipher respectively. Note that the above bound is independent of the number of users $u$, which can be adaptively chosen by the adversary and grows as large as $q$. Besides this result, Shen et al. have also shown that $2K$-SUM-ECBC [13], $2K$-PMAC\_Plus [13] and $2K$-LightMAC\_Plus [13] are all secure roughly up to $2^{2n/3}$ queries (including all MAC and ideal-cipher queries) in the multi-user setting independent of the number of users, where these constructions do not employ domain separation techniques.

Remark 1. In their paper [13], Datta et al. named the two-keyed variants of SUM-ECBC, PMAC\_Plus and LightMAC\_Plus as $2K$-SUM-ECBC, $2K$-PMAC\_Plus and $2K$-LightMAC\_Plus respectively, where for each of these constructions, the domain separation technique ensured disjointness of the set of values of $\Sigma$ and $\Theta$. However, in [32], Shen et al. considered the same constructions but without any domain separation, and refer to them using the same names. Henceforth, we shall implicitly mean the non domain-separated variants only (unless otherwise stated) when referring to the two-keyed constructions $2K$-SUM-ECBC, $2K$-PMAC\_Plus and $2K$-LightMAC\_Plus.

1.1 Issue with the CRYPTO’21 Paper [32]

In this section, we discuss three issues with [32]. The first two issues examine flaws in the security analysis of the construction and the last issue points out a flawed security claim of the construction. We begin by identifying the first issue. The Two-Keyed-DbHtS framework was proven to be multi-user secure up to $2^{2n/3}$ queries in the ideal-cipher model [32] under the assumption that each of the underlying $n$-bit independent keyed hash functions is $O(2^{-n})$-universal and regular. As an instantiation of the framework, [32] showed $2n/3$-bit multi-user security of $2K$-SUM-ECBC, $2K$-LightMAC\_Plus and $2K$-PMAC\_Plus in the ideal-cipher model. In the security proof of these instantiated constructions, they only bounded the regular and the universal advantages of the corresponding hash functions (i.e., the DbH of $2K$-SUM-ECBC, $2K$-LightMAC\_Plus and $2K$-PMAC\_Plus) up to $O(\ell/2^n)$, where $\ell$ is the maximum number of message blocks amongst all queries. However, the regular and universal advantages of the underlying double block hash functions of the above three constructions were not proven in the ideal-cipher model; instead, the authors bounded them in the standard model, where the adversary is not allowed to query the underlying block ciphers of the corresponding hash functions. In other words, considering the example of $2K$-LightMAC\_Plus, while bounding the probability of the event $\Sigma = \Sigma_j$ (where $\Sigma_i = \Sigma_j \Rightarrow Y_1 \oplus Y_2 \oplus \ldots \oplus Y_j = Y_1 \oplus Y_2 \oplus \ldots \oplus Y_j$ and $Y_a = \text{E}_K(M_a || (a))$, the authors have simply assumed that at least one of variables $Y$ in the above equation will be fresh, thus providing sufficient entropy for bounding the event. However, the authors have miserably missed the fact that existence of such a variable $Y$ may not always be guaranteed in the ideal-cipher model. For example, suppose an adversary makes the following three forward primitive queries:

\[4\text{A family of keyed hash function is said to be $c_1$-regular if for any $x$ and $y$, the probability that a randomly sampled hash function from the family maps $x$ to $y$ is $c_1$; it is said to be $c_2$-universal if for any distinct $x, x'$, the probability that a randomly sampled hash function from the family yields a collision on the pair $(x, x')$ is $c_2$.}
1. forward query with \( (x||1)_s \) and obtains \( y_1 \)
2. forward query with \( (x'||1)_s \) and obtains \( y_2 \)
3. forward query with \( (x''||2)_s \) and obtains \( y_3 \)

Let us assume that the (albeit probabilistic) event \( y_1 \oplus y_2 \oplus y_3 = 0 \) occurs. Suppose the adversary makes two more queries: the first, a construction query with \( (x) \) and the second, a construction query with \( (x''||x''') \). Then, one cannot find any fresh variable \( Y \) in the following equations:

\[
Y_1^1 = Y_1^2 \oplus Y_2^2.
\]

Therefore, to prove the security of such block cipher-based DbHtS constructions in the ideal-cipher model, one needs to consider the fact that the regular or universal advantage of the underlying double block hash functions must be bounded under the assumption that the adversary makes primitive queries to the underlying block cipher. We therefore believe that to prove the security of the constructions in the ideal-cipher model for the block cipher-based DbH function, one needs to provide a generalized definition of the universal and regular advantages in the ideal-cipher model and prove their security under this model, which was missing in [32].

The second issue is regarding the good transcript analysis of the Two-Keyed-DbHtS construction. In Fig. 4 of [32], the authors have identified the set of \( (i, a) \in [u] \times [q_i] \), which they denoted as \( F(J) \), such that both \( \Sigma_a \) and \( \Theta_a \) are fresh. They have also defined a set \( S(J) \),

\[
S(J) := \{(W_a^i, X_a^i) \in \{0, 1\}^n \setminus \text{Ran}(\Phi_J)^{2^{|F(J)|}} : W_a^i \oplus X_a^i = T_a^i \}.
\]

Then for all \((i, a) \in F(J), (U_a^i, V_a^i)\) is sampled from \( S(J) \) and is set as the permutation output of \( \Sigma_a \) and \( \Theta_a \), respectively. Finally, they have provided a lower bound on the cardinality of the set \( S(J) \) from Lemma 2. Noting that Lemma 2 proves the cardinality of the set

\[
S := \{(U_i, V_i) \in \{0, 1\}^{2q} : U_i \oplus V_i = T_i \}
\]

to be at least \( 2^n(2^n - 1) \ldots (2^n - 2q + 1)/2^{nq} \cdot (1 - 6q^3/2^{2n}) \), which is used to obtain a lower bound on \(|S(J)|\), reveals a fallacy as the two sets \( S \) and \( S(J) \) are not isomorphic to each other.

The third issue is regarding the flawed security claim of the Two-Keyed-DbHtS construction in [32]. In Theorem 1 of [32], Shen et al. show that when the underlying double block hash function of the Two-Keyed-DbHtS construction is the concatenation of two independent \( n \)-bit keyed hash functions such that each is \( O(2^{-n}) \)-universal and \( O(2^{-n}) \)-regular, Two-Keyed-DbHtS achieves \( 2n/3 \)-bit multi-user security in the ideal-cipher model. In a recent work by Guo and Wang [16], the authors came up with three concrete constructions that are instantiations of the Two-Keyed-DbHtS paradigm such that the underlying double block hash function of each of the three constructions is the concatenation of two independent \( n \)-bit keyed hash functions. Guo and Wang also show that each of the \( n \)-bit hash functions for these three constructions meets the \( O(2^{-n}) \)-universal and \( O(2^{-n}) \)-regular advantages. However, the constructions have a birthday bound distinguishing attack. As a consequence, the security bound of Two-Keyed-DbHtS as proven in Theorem 1 of [32] stands flawed. We would like to mention here that the attack holds only for those instances of Two-Keyed-DbHtS where the underlying DbH is the concatenation of two independent \( n \)-bit hash functions and it does not have any domain separation. In fact, authors of [16] were not able to show any birthday bound attack on 2K-PMAC_Plus and 2K-LightMAC_Plus as the underlying DbH function of these two constructions is not the concatenation of two independent \( n \)-bit keyed hash functions. However, it is to be noted that as the double block hash function for 2K-SUM-ECBC is the concatenation of two independent \( n \)-bit CBC functions, the attack of [16] holds for it.
1.2 Our Contribution

In this paper we prove that the Two-Keyed-DbHtS construction is multi-user secure up to $2^{3n/4}$ queries in the ideal-cipher model. To prove it, we first define the notion of a good double-block hash function, which informally means that the concatenation of two independent $n$-bit keyed hash functions is “good” if each has negligible universal and regular advantages, and the probability that the outputs of two hash function colliding for any pair of messages $M,M'$ is zero. Then, we prove that if the underlying $2n$-bit DbH function of the Two-Keyed-DbHtS construction is good, such that each of the $n$-bit keyed hash functions is $\epsilon_{\text{reg}}$-regular and $\epsilon_{\text{univ}}$-universal, then the multi-user security of our construction in the ideal-cipher model is of the order

$$\frac{9q^{4/3}}{8 \cdot 2^n} + \frac{3q^{8/3}}{2 \cdot 2^n} + \frac{q^2}{2^{2n}} + \frac{9q^{7/3}}{8 \cdot 2^{2n}} + \frac{8q^4}{3 \cdot 2^{3n}} + \frac{q}{2^n} + \frac{2u^2}{2k_{\text{reg}} + k} + \frac{2\sqrt{2}}{2n+k} + \frac{q^2}{2^{2n+k}} + \frac{2pq_{\text{reg}}}{2^n} + \frac{q^2\epsilon_{\text{univ}}}{2^n} + \frac{2q^2\epsilon_{\text{univ}}}{2^n} + \frac{q^2\epsilon_{\text{univ}}}{2^n} + \frac{2q^2\epsilon_{\text{univ}}}{2^n} + \frac{2\epsilon_{\text{univ}}}{2^{n+k}} + \frac{2q^2\epsilon_{\text{univ}}}{2^n},$$

where $q$ is the total number of MAC queries across all $u$ users, $p$ is the total number of ideal-cipher queries, $n$ is the block size of the block cipher, $k_{\text{reg}}$ is the size of the hash key and $k$ is the key size of the block cipher of the construction. As an instantiation of the Two-Keyed-DbHtS framework, we have proved that $C_2[PH-DbH,E]$, the PolyHash-based Two-Keyed-DbHtS construction which was proposed in [13] and proven to be secure up to $2^{3n/4}$ queries in the single-user setting, is multi-user secure up to $2^{3n/4}$ queries in the ideal-cipher model. The security proof of the construction crucially depends on a refined result of mirror theory over an abelian group $\{(0,1)^n, \oplus\}$, where one systematically estimates the number of solutions to a system of equations to prove the security of the finalization function of the construction up to $2^{3n/4}$ queries. Due to the attack result of Leurent et al. [21] on the DbHtS paradigm with $2^{3n/4}$ queries, the multi-user security bound of our construction is tight.

Organization. We have developed the required notations and security definitions of cryptographic primitives in Sect. 2. We demonstrate the construction and present its security bound in Sect. 3 and in Sect. 4, we prove the security of the construction. We instantiate the framework along with its security result in Sect. 5.

2 Preliminaries

General Notations: For a positive integer $q$, $[q]$ denotes the set $\{1, \ldots, q\}$, and for two natural numbers $q_1, q_2$ such that $q_2 > q_1$, $[q_1,q_2]$ denotes the set $\{q_1, \ldots, q_2\}$. For a fixed positive integer $n$, we write $\{0,1\}^n$ to denote the set of all binary strings of length $n$ and $\{0,1\}^* = \cup_{i\geq 0} \{0,1\}^i$ to denote the set of all binary strings with arbitrary finite length. We refer to the elements of $\{0,1\}^n$ as blocks. For a pair of blocks $x = (x_1, x_2) \in \{0,1\}^{2n}$, we write left$(x)$ to denote $x_1$ and right$(x)$ to denote $x_2$. For any element $x \in \{0,1\}^*$, $|x|$ denotes the number of bits in $x$ and for $x, y \in \{0,1\}^*$, $x \| y$ denotes the concatenation of $x$ followed by $y$. We denote the bitwise xor operation of $x, y \in \{0,1\}^*$ by $x \oplus y$. We parse $x \in \{0,1\}^*$ as $x = x_1 \| x_2 \| \ldots \| x_l$, where for each $i = 1, \ldots, l - 1$, $x_i$ is a block and $1 \leq |x_i| \leq n$. For $x \in \{0,1\}^n$, $x_0 \ldots x_{n-1} \| x_0$, lsb$(x)$ denotes the least significant bit $x_0$ of $x$. For a given bit $b$, fix$b$ is a function from $\{0,1\}^n$ to $\{0,1\}^n$ that takes an $n$-bit binary string $x = x_{n-1} \ldots \| x_0$ and returns another binary string $x' = (x_{n-1} \ldots \| b)$, where lsb$(x)$ is fixed to bit $b$. Given a tuple $\overline{x} = (x_1, x_2, \ldots, x_q)$ of $n$-bit binary string, we say that an element $x_i$ of the tuple $\overline{x}$ is non-fresh if there exists at least one $j \neq i$ such that $x_i = x_j$. Otherwise, we call that element $x_i$ is fresh.

Given a finite set $S$ and a random variable $X$, we write $X \overset{\text{S}}{\rightarrow}$ to denote that $X$ is sampled uniformly at random from $S$. We say that $X_1, X_2, \ldots, X_q$ are sampled with replacement
A block cipher \( \Phi \) is denoted as \( \Phi \colon \mathcal{X} \to \{0,1\}^n \) with domain \( \mathcal{X} \). The set of all functions from an arbitrary domain to \( \{0,1\}^n \). The set of all block functions with domain \( \mathcal{X} \) is denoted as \( \text{Func}(\mathcal{X}) \). \(^5\) We call \( \Phi \) to be a \textit{double-block function} if it maps elements from an arbitrary domain to \( \{0,1\}^n \). Similarly, the set of all permutations over \( \mathcal{X} \) is represented by \( \text{Perm}(\mathcal{X}) \). A function \( \Phi \) is said to be a \textit{block function} if it maps elements from an arbitrary domain to \( \{0,1\}^n \). The set of all block functions with domain \( \mathcal{X} \) is denoted as \( \text{Func}(\mathcal{X}) \). \(^5\) We call \( \Phi \) to be a \textit{double-block function} if it maps elements from an arbitrary domain to \( \{0,1\}^n \). For a given double-block function \( \Phi : \mathcal{X} \to \{0,1\}^n \), we write \( \Phi_Y : D \to \{0,1\}^n \) such that for every \( x \in \mathcal{X} \), \( \Phi_Y(x) = \text{left}(\Phi(x)) \). Similarly, we write \( \Phi_X : \mathcal{X} \to \{0,1\}^n \) such that for every \( x \in \mathcal{X} \), \( \Phi_X(x) = \text{right}(\Phi(x)) \). For two block functions \( \Phi_1 : \mathcal{X} \to \{0,1\}^n \) and \( \Phi_2 : \mathcal{X} \to \{0,1\}^n \), one can naturally define a double-block function \( \Phi : \mathcal{X} \to \{0,1\}^n \) such that \( \Phi(x) = (\Phi_1(x), \Phi_2(x)) \), which we write as \( \Phi = (\Phi_1, \Phi_2) \). For a finite set \( \mathcal{X} \) and an integer \( q \), we write \( \mathcal{X}^{(q)} \) to denote the set \( \{(x_1, x_2, \ldots, x_q) : x_i \in \mathcal{X}, x_i \neq x_j\} \). For integers \( 1 \leq b \leq a \), we write \( P(a, b) \) to denote \( a(a - 1) \cdots (a - b + 1) \), where \( P(a, 0) = 1 \) by convention. Therefore, \( |\mathcal{X}^{(q)}| = P(|\mathcal{X}|, q) \).

2.1 Distinguishing Advantage

An adversary \( A \) is modeled as a randomized algorithm with access to an external oracle \( O \). Such an adversary is called an \textit{oracle adversary}. An oracle \( O \) is an algorithm that may be a cryptographic scheme being analyzed. The interaction between \( A \) and \( O \), denoted by \( A^O \), generates a transcript \( \tau = \{(x_1, y_1), (x_2, y_2), \ldots, (x_q, y_q)\} \), where \( x_1, x_2, \ldots, x_q \) are \( q \) queries of \( A \) to oracle \( O \) and \( y_1, y_2, \ldots, y_q \) be the corresponding responses, where \( y_i = O(x_i) \). We assume that \( A \) is \textit{adaptive}, which means that \( x_i \) is dependent on the previous \( i - 1 \) responses.

**Distinguishing Game.** Let \( F \) and \( G \) be two random systems and an adversary \( A \) is given oracle access to either of \( F \) or \( G \). After interaction with an oracle \( O \in \{F, G\} \), \( A \) outputs \( 1 \), which is denoted as \( A^O \Rightarrow 1 \). Such an adversary is called a \textit{distinguisher} and the game is called a \textit{distinguishing game}. The task of the distinguisher in a distinguishing game is to tell with which of the two systems it has interacted. The advantage of the distinguisher \( A \) in distinguishing the random system \( F \) from \( G \) is defined as

\[
\text{Adv}_F^G(A) \overset{\Delta}{=} \text{Pr}[A^F \Rightarrow 1] - \text{Pr}[A^G \Rightarrow 1],
\]

here the above probability is defined over the probability spaces of \( A \) and \( O \). The maximum advantage in distinguishing \( F \) from \( G \) is defined as

\[
\max_{A \in \mathcal{A}} \text{Adv}_F^G(A),
\]

where \( \mathcal{A} \) is the class of all possible distinguishers. One can easily generalize this setting when the distinguisher interacts with multiple oracles, which are separated by commas. For example, \( \text{Adv}_{F_1, \ldots, F_m}^{G_1, \ldots, G_m}(A) \) denotes the advantage of \( A \) in distinguishing \( (F_1, \ldots, F_m) \) from \( (G_1, \ldots, G_m) \).

2.2 Block Cipher

A block cipher \( E : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n \) is a function that takes a key \( k \in \mathcal{K} \) and an \( n \)-bit input data \( x \in \{0,1\}^n \) and produces an \( n \)-bit output \( y \) such that for each key \( k \in \mathcal{K} \), \( E(k, \cdot) \)

\(^5\)When \( \mathcal{X} = \{0,1\}^n \), we write \( \text{Func} \) to denote \( \text{Func}(\{0,1\}^n) \).
is a permutation over \( \{0,1\}^n \). \( \mathcal{K} \) is called the key space of the block cipher and \( \{0,1\}^n \) is its input-output space. In shorthand notation, we write \( E_k(x) \) to represent \( E(k,x) \).

Let \( BC(\mathcal{K}, \{0,1\}^n) \) denotes the set of all \( n \)-bit block ciphers with key space \( \mathcal{K} \). We say that a block cipher \( E \) is an \((q, \epsilon, t)\)-secure strong pseudorandom permutation, if for all distinguishers \( A \) that make a total of \( q \) forward and inverse queries with run time at most \( t \), the following holds:

\[
\text{Adv}^{\text{PRF}}_E(A) \stackrel{\Delta}{=} | \Pr[K \leftarrow \mathcal{K} : A^{E_K} \Rightarrow 1] - \Pr[\Pi \leftarrow \text{Perm} : A^{\Pi} \Rightarrow 1] | \leq \epsilon.
\]

### 2.3 PRF Security in the Ideal-Cipher Model

A **keyed function** with the key space \( \mathcal{K} \), domain \( \mathcal{X} \) and range \( \mathcal{Y} \) is a function \( F : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y} \). We denote \( F(k,x) \) by \( F_k(x) \). A random function \( \text{RF} \) from \( \mathcal{X} \) to \( \mathcal{Y} \) is a uniform random variable over the set \( \text{Func}(\mathcal{X}, \mathcal{Y}) \), i.e., \( \text{RF} \leftarrow \text{Func}(\mathcal{X}, \mathcal{Y}) \). We define the pseudorandom security of \( F \) under the ideal-cipher model. We assume that \( F \) makes internal calls to a publicly evaluated block cipher \( E \) with a randomly sampled block cipher key \( K \leftarrow \mathcal{K} \) (\( F \) can make calls to multiple block ciphers when all of them are independent and uniform over the set \( BC(\mathcal{K}, \{0,1\}^n) \)). For simplicity, we write \( F_K \) to denote \( F \) with a uniformly sampled block cipher \( E \leftarrow BC(\mathcal{K}, \{0,1\}^n) \), which is keyed by a randomly sampled \( K \leftarrow \mathcal{K} \). The distinguisher \( A \) is given access to either \( (F^F_K, E^\pm) \) for \( K \leftarrow \mathcal{K} \) or \( (RF, E^\pm) \), where \( E \leftarrow BC(\mathcal{K}, \{0,1\}^n) \) is a uniformly sampled \( n \)-bit block cipher such that \( A \) can make forward or inverse queries to \( E \), which is denoted as \( E^\pm \). We define the prf-advantage of \( A \) against a keyed function \( F \) in the ideal cipher model as

\[
\text{Adv}^{\text{PRF}}_F(A) \stackrel{\Delta}{=} \text{Adv}^{(F^F_K, E^\pm)}_{\text{RF}, E^\pm}(A).
\]

We say \( F \) is a \((q,p,\epsilon,t)\)-PRF if \( \text{Adv}^{\text{PRF}}_F(A) \leq \epsilon \) for all adversaries \( A \) that make \( q \) queries to \( F \), \( p \) forward and inverse offline queries to \( E \) and run for time at most \( t \).

### 2.4 Multi-User PRF Security in the Ideal-Cipher Model

We assume there are \( u \) users in the multi-user setting, such that the \( i \)-th user executes \( F^F_0 \). Furthermore, the \( i \)-th user key \( K_i \) is independent of the keys of all other users. An adversary \( A \) has access to all the \( u \) users as oracles. \( A \) makes queries to the oracles in the form of \((i,M)\) to the \( i \)-th user and obtains \( T \leftarrow F^F_0(M) \). We call these **construction queries**. For \( i \in [u] \), we assume \( A \) makes \( q_i \) queries to the \( i \)-th oracle. We also assume that \( A \) makes queries to the underlying block cipher \( E \) and its inverse with some chosen keys \( k^j \). We call these **primitive queries**. Suppose \( A \) chooses \( s \) distinct block cipher keys \((k^1, \ldots, k^s)\) and makes \( p_j \) primitive queries to the block cipher \( E \) with chosen keys \( k^j \) for \( 1 \leq j \leq s \). Let \( A \) be a \((u,q,p,t)\)-adversary against the PRF security of \( F \) for all \( u \) users such that \( q = q_1 + \ldots + q_u \) is the total number of construction queries and \( p = p_1 + \ldots + p_u \) is the total number of primitive queries to the block cipher \( E \) with the total running time \( A \) being at most \( t \). We assume that for any \( i \in [u] \), \( A \) does not repeat any construction query to the \( i \)-th user. Similarly, \( A \) does not repeat any primitive query for any chosen block cipher key \( k^j \) to the block cipher \( E \). The advantage of \( A \) in distinguishing \((F^F_K, F^F_{K_2}, \ldots, F^F_{K_u}, E^\pm)\) from \((RF_1, RF_2, \ldots, RF_u, E^\pm)\) in the multi-user setting, where \( RF_1, RF_2, \ldots, RF_u \leftarrow \text{Func}(\mathcal{X}, \mathcal{Y}) \) are \( u \) independent random functions, is defined as

\[
\text{Adv}^{\mu\text{-PRF}}_F(A) \stackrel{\Delta}{=} \left| \Pr \left[ A^{((F^F_K, \ldots, F^F_{K_u}), E^\pm)} \Rightarrow 1 \right] - \Pr \left[ A^{((RF_1, \ldots, RF_u), E^\pm)} \Rightarrow 1 \right] \right|,
\]

where the randomness is defined over \( K_1, \ldots, K_u \leftarrow \mathcal{K}, E \leftarrow BC(\mathcal{K}, \{0,1\}^n) \) and the randomness of the adversary (if any). We write

\[
\text{Adv}^{\mu\text{-PRF}}_F(u,q,p,t) \stackrel{\Delta}{=} \max_A \text{Adv}^{\mu\text{-PRF}}_F(A),
\]
where the maximum is over all \((u, q, p, t)\)-adversaries \(A\). In this paper, we skip the time parameter of the adversary as we shall assume that the adversary is computationally unbounded. This also leads to the assumption that the adversary is deterministic. When \(u = 1\), it makes \(\text{Adv}^\text{PRF}_u(Q, q, p, t)\) the single-user distinguishing advantage.

### 2.5 Security of a Keyed Hash Function

Let \(K_h\) and \(\mathcal{X}\) be two non-empty finite sets. A keyed function \(H : K_h \times \mathcal{X} \to \{0,1\}^n\) is \(\epsilon\)-almost-xor universal (axu) if for any distinct \(x, x' \in \mathcal{X}\) and for any \(\Delta \in \{0,1\}^n\),

\[
\Pr[K_h \leftarrow K_h : H_{K_h}(x) \oplus H_{K_h}(x') = \Delta] \leq \epsilon_{\text{axu}}.
\]

Moreover, \(H\) is an \(\epsilon\)-universal hash function if for any distinct \(x, x' \in \mathcal{X}\),

\[
\Pr[K_h \leftarrow K_h : H_{K_h}(x) = H_{K_h}(x')] \leq \epsilon_{\text{univ}}.
\]

A keyed hash function is said to be \(\epsilon\)-regular if for any \(x \in \mathcal{X}\) and for any \(\Delta \in \{0,1\}^n\),

\[
\Pr[K_h \leftarrow K_h : H_{K_h}(x) = \Delta] \leq \epsilon_{\text{reg}}.
\]

### 2.6 Mirror Theory

Mirror theory is a collection of combinatorial results that give a lower bound on the number of solutions to a system of bivariate affine equations \(\mathbb{E}\) over an abelian group \((\{0,1\}^n, \oplus)\). We represent a system of equations by a simple graph \(G = (V, \mathcal{E})\) containing no loops or multiple edges, where each vertex denotes an \(n\)-bit unknown (for a fixed \(n\)), and we connect vertices \(P\) and \(Q\) with an edge labeled \(\lambda \in \{0,1\}^n\) if \(P \oplus Q = \lambda \in \mathcal{E}\). For a path \(\mathcal{L} = P_1 \xrightarrow{\lambda_1} P_2 \xrightarrow{\lambda_2} \ldots \xrightarrow{\lambda_i} P_t\) in the graph \(G\), we define the label of the path

\[
\lambda(\mathcal{L}) = \lambda_1 \oplus \lambda_2 \oplus \ldots \oplus \lambda_i.
\]

In this work, we focus on a graph \(G = (V, \mathcal{E})\) with certain properties as listed below:

1. \(G\) contains no isolated vertex, i.e., every vertex is incident with at least one edge.
2. The vertex set \(V\) is partitioned into two disjoint sets denoted by \(P\) and \(Q\), where there are no edges within the vertex set in partition \(P\) or in partition \(Q\). All edges connect a vertex in \(P\) to a vertex in \(Q\). We call such graphs bipartition graphs.
3. \(G\) contains no cycle.
4. \(\lambda(\mathcal{L}) \neq 0^n\) for any path \(\mathcal{L}\) in \(G\).

Any bipartition graph \(G\) satisfying the above properties shall be called a good graph. Note that a good bipartition graph \(G\) contains no cycle. Therefore, \(G\) can be decomposed into its connected components, all of which are trees; let

\[
G = C_1 \sqcup C_2 \sqcup \ldots \sqcup C_\alpha \sqcup D_1 \sqcup D_2 \sqcup \ldots \sqcup D_\beta
\]

for some \(\alpha, \beta \geq 0\), where \(C_i\) denotes a component of size greater than 2, and \(D_i\) denotes a component size of 2. We write \(C = C_1 \sqcup C_2 \sqcup \ldots \sqcup C_\alpha\) and \(D = D_1 \sqcup D_2 \sqcup \ldots \sqcup D_\beta\).

**Definition 1.** Let \(\mathcal{E}_G\) be a system of equations induced by a good bipartite graph \(G\). An injective function \(\Phi : P \sqcup Q \to \{0,1\}^n\) is said to be an injective solution to \(\mathcal{E}_G\) if \(\Phi(P_i) \oplus \Phi(Q_j) = \lambda_{ij}\) for all \(\{P_i, Q_j\} \in \mathcal{E}\).
We remark that assigning any value to a vertex in $P$ allows the labeled edges to uniquely determine the values of all the other vertices in the component containing $P$, since $G$ contains no cycle. The values in the same component are all distinct as $\lambda(L) \neq 0^n$ for any path $L$. The number of possible assignments of distinct values to the vertices in $G$ is $\mathbf{P}(2^n, |P| + |Q|)$. One may expect that when such an assignment is chosen uniformly at random, it would satisfy all the equations in $G$ with probability $2^{-q}$, where $q$ denotes the number of edges (i.e., equations) in $G$. Indeed, we can prove that the number of solutions is closed to $\mathbf{P}(2^n, |P| + |Q|)/2^{2n}$, up to a certain error. Formally, we have the following result:

**Lemma 1.** Let $G$ be a good bipartition graph, and let $q$ and $q^c$ denote the number of edges of $G$ and $C$, respectively. Let $v$ be the number of vertices of $G$. If $q < 2^n/8$, then the number of solutions to $G$, denoted $h(G)$, satisfies

\[
\frac{h(G)2^{nq}}{\mathbf{P}(2^n, v)} \geq \left( 1 - \frac{9(q^c)^2}{8 \cdot 2^n} - \frac{3q^cq^2}{2 \cdot 2^{2n}} - \frac{q^2}{2^{2n}} - \frac{9(q^c)^2q}{8 \cdot 2^{2n}} - \frac{8q^4}{3 \cdot 2^{3n}} \right).
\]

We refer the reader to [20] for a proof of the lemma.

### 3 The Two-Keyed DbHtS Construction

In this section, we describe the Two-Keyed Double-block Hash-then-Sum or in short, Two-Keyed-DbHtS construction to build a beyond birthday bound secure variable input length PRF. Let $H^1 : \mathcal{K}_h \times \{0, 1\}^* \rightarrow \{0, 1\}^n$ and $H^2 : \mathcal{K}_h \times \{0, 1\}^* \rightarrow \{0, 1\}^n$ be two keyed hash functions. Based on $H^1$ and $H^2$, we define the Double-block Hash or in short DbH function $H : \mathcal{K}_h \times \mathcal{K}_h \times \{0, 1\}^* \rightarrow \{0, 1\}^{2n}$ as follows:

\[
H_{(L_1, L_2)}(M) = (H_{L_1}^1(M), H_{L_2}^2(M)).
\]  

We compose this DbH function with a very simple and efficient single-keyed xor function $\text{XOR}_K(x, y) = E_K(x) \oplus E_K(y)$, where $E_K$ is an $n$-bit block cipher and the block cipher key $K$ is independent over the hash key $(L_1, L_2)$, to realize the two-keyed-DbHtS construction as follows:

\[
C_2[H, E]_{(L_1, L_2, K)}(M) := \text{XOR}_K(H_{L_1}^1(M), H_{L_2}^2(M)).
\]

We use the name **Two-Keyed-DbHtS** construction, as we count the hash key as one key and the xor function requiring one key, which is independent of the hash key. Most of the beyond birthday bound secure variable input length PRFs like 2K-SUM-ECBC, 2K-PMAC+Plus, 2K-LightMAC+Plus are specific instantiations of the Two-Keyed-DbHtS paradigm. These constructions (with domain separation technique) have been proven secure up to $2^{2n/3}$ queries in the standard model [13] for a single-user setting. In [32], all these three constructions (without domain separation technique) have been proven secure up to $2^{2n/3}$ queries in the ideal-cipher model for a multi-user setting. We note here that as the xor function is not a PRF over two blocks, we cannot apply the tradition *Hash-the-PRP* composition result directly to analyze the security of the two-keyed DbHtS. Thus, we need a different type of composition result for the security analysis of the Two-Keyed-DbHtS construction that utilizes higher security properties of its underlying DbH function instead of having only the universal or regular property.

**Definition 2.** Let $H^1 : \mathcal{K}_h \times \{0, 1\}^* \rightarrow \{0, 1\}^n$ and $H^2 : \mathcal{K}_h \times \{0, 1\}^* \rightarrow \{0, 1\}^n$ be two $n$-bit keyed hash functions. We say that the double-block hash function $H : \mathcal{K}_h \times \mathcal{K}_h \times \{0, 1\}^* \rightarrow \{0, 1\}^{2n}$ defined in Eqn. (1) is **good** if it satisfies the following conditions:

- $H^1$ is a family of $\epsilon_{\text{reg}}$-regular and $\epsilon_{\text{univ}}$-universal functions.
The ideal world consists of two phases: (i) the online and (ii) the offline phase. Before the game begins, we sample $u$ independent functions $f_1, f_2, \ldots, f_u$ uniformly at random from the set of all functions $\text{Func}(\{0,1\}^*, \{0,1\}^n)$ that map an arbitrary-length string to an $n$-bit string. We also sample an $n$-bit block cipher $E$ from the set of all block ciphers with a $k$-bit key and an $n$-bit input. In the online phase, when the distinguisher makes the $i$-th construction query for the $i$-th user $M'_i$, to the construction oracle, it returns $T'_i \leftarrow f_i(M'_i)$. Similarly, if the distinguisher makes a forward (resp. inverse) primitive
query with a chosen block cipher key $J$ and an input $x$ to the ideal cipher oracle, it returns $E(J,x)$ (resp. $E^{-1}(J,x)$). However, if any response of the construction queries is an all-zero string $0^n$, then the bad flag Bad-Tag is set to $1$ and the game is aborted.

**Online Phase of $O_{\text{ideal}}$**

1. $E \leftarrow \text{BC}(K, \{0,1\}^n)$;

**Construction Query:**

2. On $a$-th query of $i$-th user $M_i^a$, return $T_i^a \leftarrow \{0,1\}^n$;

3. if $\exists(i,a) : T_i^a = 0$ then $\text{Bad-Tag} \leftarrow 1$. ⊥;

**Primitive Query:**

4. On $j$-th forward query with chosen key $J^f$ and input $u_i^j$, return $v_i^j \leftarrow E_{J^f}(u_i^j)$;

5. On $j$-th backward query with chosen key $J^b$ and input $v_i^j$, return $u_i^j \leftarrow E_{J^b}^{-1}(v_i^j)$;

6. $\text{Dom}(E_{J^f}) \leftarrow \text{Dom}(E_{J^b}) \cup \{u_i^j\}$, $\text{Ran}(E_{J^f}) \leftarrow \text{Ran}(E_{J^b}) \cup \{v_i^j\}$;

**Figure 4.1:** Online Phase of the Ideal oracle $\$$. Boxed statements denote bad events. Whenever a bad event is set to 1, the ideal oracle immediately aborts (denoted as ⊥) and returns the remaining values of the transcript in an arbitrary manner. So, if the game aborts for some bad event, then its previous bad events must not have occurred.

After this interaction is over, the offline phase begins. In this phase, we sample $u$ pairs of dummy hash keys $(L_1^a, L_2^a)_{a=1}^u \leftarrow K_h \times K_h$ and $u$ dummy block cipher keys $(K^i)_{a=1}^u \leftarrow K$, where $L_1^a$ (resp. $L_2^a$) is the left (resp. right) hash key for the $i$-th user and $K^i$ is its block cipher key. If the block cipher key and a left (resp. right) hash key of the $i_1$-th user collides with the block cipher key and left (resp. right) hash key of the $i_2$-th user, then we set the flag BadK to 1 and abort the game. If the game is not aborted, then we can compute a pair of $2n$-bit hash values $(\Sigma_i^a, \Theta_i^a)$ for all queries across $u$ users, where we often refer to $\Sigma_i^a \leftarrow \mathcal{H}_{L_1^a}(M_i^a)$ as the left hash output and to $\Theta_i^a \leftarrow \mathcal{H}_{L_2^a}^{-1}(M_i^a)$ as the right hash output for the $a$-th query of the $i$-th user.

Now, if the block cipher key of the $i$-th user and the left hash or right hash output for its $a$-th query collides with some chosen ideal cipher key and one of the corresponding inputs of the forward ideal cipher query, then we set the bad flag Bad1 to 1 and abort the game.

For the $i$-th user, if the left or right hash outputs for two of its queries collide and the corresponding responses also collide with each other (i.e., $\Sigma_i^a = \Sigma_i^b, T_i^a = T_i^b$, $J_i^a = J_i^b$), then we consider it to be a bad event. Similarly, for a pair of users $i_1$ and $i_2$, if their left or right hash outputs collide with each other, then we again consider it to be a bad event. If at least one of the above bad events occurs, we set Bad2 to 1 and abort the game. We also set another flag Bad3 to 1 and abort the game if for the $i$-th user, the number of the pairs of queries whose either left or right hash outputs collide with each other is at least $q_i^{\delta/3}$, where $q_i$ is the number of queries made by the $i$-th user.

Finally, we set the flag Bad4 to 1 if at least one of the following events holds: (a) for the $i$-th user, two left hash outputs collide and their corresponding right hash outputs also collide, or (b) for the $i$-th user, there exists a tuple of four query indices $a,b,c,d$ such that either (i) $\Sigma_i^a = \Sigma_i^b, \Theta_i^a = \Theta_i^b, \Sigma_i^c = \Sigma_i^d$ holds or (ii) $\Theta_i^a = \Theta_i^b, \Sigma_i^c = \Sigma_i^d, \Theta_i^a = \Theta_i^b$ holds. As the DbH function $\mathcal{H}$ is good, $\Sigma_i^a$ cannot collide with $\Theta_i^b$. It is also to be noted here that as the hash function is good, i.e., the hash outputs of two hash functions never collide, it immediately rules out the attack of [16].

If the game is not aborted at this stage, then it follows that none of the bad events have occurred. All the query-response pairs belong to exactly one of the sets $Q^=$ or $Q^\neq$ as...
Whenever a bad event is set to occur across all users such that the block cipher key of the user key aborts for some bad event, then we may assume that the previous bad events have not occurred.

\[ \exists i \in [u], j \in [s], u[0]_{a_i} \in \text{Dom}(E_{j,i}), \text{ such that } K_i^j = J^j \land \Sigma_i^a = u[0]_{a_i}; \]

if one of the following holds:

(a) \( \exists i \in [u], j \in [s], u[1]_{a_i} \in \text{Dom}(E_{j,i}), \text{ such that } K_i^j = J^j \land \Theta_i^a = u[1]_{a_i}; \)

\[ \text{ then } \text{ if } \text{Bad1} \leftarrow 1 \text{; } \]

if one of the following holds:

(a) \( \exists i \in [u], a, b \in [q_i], \text{ such that } \Sigma_i^a = \Sigma_i^b \land T_i^a = T_i^b; \)

(b) \( \exists i_1, i_2 \in [u], a \in [q_{i_1}], b \in [q_{i_2}], \text{ such that } K_{i_1}^{i_2} = K^j \land \Sigma_{i_1}^a = \Sigma_{i_2}^b; \)

(c) \( \exists i \in [u], a, b \in [q_i], \text{ such that } \Theta_i^a = \Theta_i^b \land T_i^a = T_i^b; \)

(d) \( \exists i_1, i_2 \in [u], a \in [q_{i_1}], b \in [q_{i_2}], \text{ such that } K_{i_1}^{i_2} = K^j \land \Theta_i^a = \Theta_i^b; \)

\[ \text{ then } \text{Bad2} \leftarrow 1 \text{; } \]

if one of the following holds:

(a) \( \exists \in [u], \text{ such that } \left\{|(a, b) : \Sigma_i^a = \Sigma_i^b\} \geq q_i^{2/3}; \right. \)

(b) \( \exists i \in [u], \text{ such that } \left\{|(a, b) : \Theta_i^a = \Theta_i^b\} \geq q_i^{2/3}; \right. \)

\[ \text{ then } \text{Bad3} \leftarrow 1 \text{; } \]

if one of the following holds:

(a) \( \exists i \in [u], a, b \in [q_i] \text{ such that } \Sigma_i^a = \Sigma_i^b \land \Theta_i^a = \Theta_i^b; \)

(b) \( \exists i \in [u], a, b, c, d \in [q_i] \text{ such that } \Sigma_i^a = \Sigma_i^b \land \Theta_i^a = \Theta_i^b \land \Sigma_i^c = \Sigma_i^d; \)

(c) \( \exists i \in [u], a, b, c, d \in [q_i] \text{ such that } \Theta_i^a = \Theta_i^b \land \Sigma_i^a = \Sigma_i^b \land \Theta_i^c = \Theta_i^d; \)

\[ \text{ then } \text{Bad4} \leftarrow 1 \text{; } \]

go to subroutine 4.3;

**Figure 4.2:** Offline Phase of the Ideal oracle \( \mathcal{O} \): Boxed statements denote bad events. Whenever a bad event is set to 1, the ideal oracle immediately aborts (denoted as \( \perp \)) and returns the remaining values of the transcript in an arbitrary manner. So, if the game aborts for some bad event, then we may assume that the previous bad events have not occurred.

defined in lines 13 and 14 of Fig. 4.2, where \( \mathcal{Q}^n \) is the set of all queries across all users such that the block cipher key of the \( i \)-th user collides with an ideal cipher key, but none of its hash outputs collide with any ideal cipher query, and \( \mathcal{Q}^f \) is the set of all queries across all users such that the block cipher key of the \( i \)-th user does not collide with any ideal cipher key. We also define two additional sets: \( \mathcal{I}^n \) and \( \mathcal{I}^f \) for \( \mathcal{Q}^n \) and \( \mathcal{Q}^f \), where \( \mathcal{I}^n \) (resp. \( \mathcal{I}^f \)) is the set of all \( i \) such that \((i, \star) \in \mathcal{Q}^n\) (resp. \((i, \star) \in \mathcal{Q}^f\)). We partition \( \mathcal{I}^n \) into \( r \) non-empty equivalence classes \( \mathcal{I}^n_1, \mathcal{I}^n_2, \ldots, \mathcal{I}^n_r \) based on the relation that the \( i \)-th user key \( K^i \) collides with \( J^j \) if and only if \( i \in \mathcal{I}^n_j \). Similarly, we partition \( \mathcal{I}^f \) into \( s \) equivalence classes based on the equivalence relation \( i \sim j \) if and only if \( K^i = K^j \). Now, for the \( j \)-th equivalence class of \( \mathcal{I}^n \), we consider the tuple

\[ \widetilde{\Sigma}_j := \bigcup_{i \in \mathcal{I}^n_j} \{ (\Sigma_i^1, \Sigma_i^2, \ldots, \Sigma_i^s) \}, \quad \widetilde{\Theta}_j := \bigcup_{i \in \mathcal{I}^n_j} \{ (\Theta_i^1, \Theta_i^2, \ldots, \Theta_i^s) \}. \]
#### Offline Phase of $O_{ideal}$. Sampling Phase

1. $Q^\ast := \{(i, a) \in [u] \times [q] : \exists j \in [s], K^i = J^j, \sum_{a}^i \notin \text{Dom}(E_{j,i}), \Theta_a^i \notin \text{Dom}(E_{j,i})\}$;
2. $I^\ast := \{i \in [u] : (i, *) \in Q^\ast = I_{T}^I \cup I_{T}^2 \cup \ldots \cup I_{T}^n ; \text{ / / } i \in I^\ast \Rightarrow K^i = J^j\}$;
3. $\forall j \in [r] : \Theta^j = \bigcup_{i \in I^j} \{(\sum^1_i, \sum^2_i, \ldots, \sum^n_i)\}, \overline{\Theta}^j = \bigcup_{i \in I^j} \{\{(\Theta^1_i, \Theta^2_i, \ldots, \Theta^n_i)\} \}$;
4. $\forall j \in [r]$ do the following steps:
   5. $\forall a \in I_j \text{ / / } a \text{ not fresh in } (\sum_i^1, \sum_i^2, \ldots, \sum_i^n);$
   6. if $\sum_i^a \notin \text{Dom}(E_{j,i})$, then $\Psi_j(\sum_i^a) \leftarrow Z_{1,a}^i \leftarrow \{0, 1\}^n \setminus \text{Ran}(E_{j,i}), Z_{2,a}^i \leftarrow Z_{1,a}^i \oplus T_{a}^i$;
   7. else $Z_{1,a}^i \leftarrow \Psi_j(\sum_i^a), Z_{2,a}^i \leftarrow Z_{1,a}^i \oplus T_{a}^i$;
   8. if $Z_{2,a}^i \in \text{Ran}(E_{j,i})$ then $\text{Bad-Samp} \leftarrow 1$;
9. else $\text{Dom}(E_{j,i}) \leftarrow \text{Dom}(E_{j,i}) \cup \{\sum_i^a, \Theta_a^i\}, \text{Ran}(E_{j,i}) \leftarrow \text{Ran}(E_{j,i}) \cup \{Z_{a}^i, Z_{a}^i \oplus T_{a}^i\}$;
10. Set $\Psi_j(\sum_i^a) \leftarrow Z_{1,a}^i, \Psi_j(\Theta^i_a) \leftarrow Z_{2,a}^i, \forall i \in I^j, a \in [q]$;
11. $Q^\ast := \{(i, a) \in [u] \times [q] : \forall j \in [s], K^i \neq J^j\}$;
12. $I^\ast := \{i \in [u] : (i, *) \in Q^\ast = I_{T}^I \cup I_{T}^2 \cup \ldots \cup I_{T}^n ; \text{ / / } i \in I^\ast \Rightarrow K^i = J^j\}$;
13. $\forall j \in [r^\prime] : f_j := \text{ distinct number of elements in the tuple } \overline{\Theta}^j$;
14. $\forall j \in [r^\prime]$:
   15. do the following steps:
   16. $\text{Dom}(E_{j,i}) \leftarrow \text{Dom}(E_{j,i}) \cup \{\sum_i^a, \Theta_a^i\} : i \in I^j, a \in [q]$;
   17. $\text{Ran}(E_{j,i}) \leftarrow \text{Ran}(E_{j,i}) \cup \{Z_{a}^i, Z_{a}^i \oplus T_{a}^i\} : i \in I^j, a \in [q]$;
   18. Set $\Psi_j(\sum_i^a) \leftarrow Z_{1,a}^i, \Psi_j(\Theta_a^i) \leftarrow Z_{2,a}^i, \forall i \in I^j, a \in [q]$;
19. return $(\sum_i^a, \Theta_a^i, Z_{1,a}^i, Z_{2,a}^i)_{(i,a)\in[u]\times[q]}$.

Figure 4.3: Offline Phase of the Ideal oracle $\$, where we sample the output of the hash values.

Note that due to the event in line number 7.(b) (resp. 7.(d)) of Fig. 4.2, we have $\sum_i^a \neq \sum_i^b$ (resp. $\Theta_a^i \neq \Theta_b^i$) for $i_1, i_2 \in I^j$, and $a \in [q]$. If $\sum_i^a$ is not fresh in the tuple $(\sum_i^1, \sum_i^2, \ldots, \sum_i^n)$ for some $(i, a) \in I^j \times [q]$ and the output of $\sum_i^a$ has not been sampled yet, then we sample the its output $Z_{1,a}^i$ from outside the range of $E_{j,i}$ and set the output of $\Theta_a^i$ as the xor of $Z_{a}^i$ and $T_{a}^i$ (see line 6 of Fig. 4.3). Otherwise, we set the output of $\sum_i^a$ to the already defined element and adjust the output of the other hash value accordingly (see line 7 of Fig. 4.3). Note that in the latter case, we do not sample the output. In the above adjustment, if the output of $\Theta_a^i$ happens to collide with any previously sampled output, then we set flag badsamp to 1 and abort the game (see line 8 of Fig. 4.3) and abort the game. Note that this event cannot hold for the real oracle, as $\Theta_a^i$ is fresh in $(\Theta_j^1, \Theta_j^2, \ldots, \Theta_j^n)$ for $i \in I^j$ and $a \in [q]$. If the above flag is not set to 1, then the sampling for the output of $\sum_i^a$, where $(i, a) \in Q^\ast$, preserves permutation compatibility. Finally, for all other $(i, a) \in Q^\ast$, we sample $Z_{1,a}^i$ and $Z_{2,a}^i$ such that $Z_{1,a}^i \oplus Z_{2,a}^i = T_{a}^i$.

#### 4.2 Attack Transcript

We summarize here, the interaction between the distinguisher and the challenger in a transcript. The set of all construction queries for $u$ instances are summarized in a transcript $\tau_\epsilon = \tau_1^i \cup \tau_2^i \cup \ldots \cup \tau_u^i$, where $\tau_a^i = \{(M_1^i, T_1^i), \ldots, (M_a^i, T_a^i)\}$ denotes the query-response.
transcript generated from the $i$-th instance of the construction. Moreover, we assume that $A$ has chosen $s$ distinct ideal cipher keys $J^1, \ldots, J^s$ such that it makes $p_j$ ideal cipher queries to the block cipher with the chosen key $J^j$. We summarize the ideal cipher queries in a transcript $\tau_p = \tau_p^1 \cup \tau_p^2 \cup \ldots \cup \tau_p^j$, where $\tau_p^j = \{(u_1^j, v_1^j), \ldots, (u_{p_j}^j, v_{p_j}^j), J^j\}$ denotes the transcript of the ideal cipher queries when the chosen ideal cipher key is $J^j$. We assume that $A$ makes $q_i$ construction queries for the $i$-th instance and $p_j$ ideal cipher queries (including forward and inverse queries) with chosen ideal cipher key $J^j$. We also assume that the total number of construction queries across $u$ instances is $q$, i.e., $q = (q_1 + \ldots + q_u)$ and the total number of ideal cipher queries is $p = (p_1 + \ldots + p_s)$. Since $A$ is non-trivial, none of the transcripts contain any duplicate elements.

We modify the experiment by releasing internal information to $A$ after it has finished its interaction but has not yet output the decision bit. In the real world, we reveal all the keys $(L_1', L_2', K')$ for all $u$ instances used in the construction. In the ideal world, we sample them uniformly at random from their respective key spaces and reveal them to the distinguisher. Once the keys are revealed to the distinguisher, $A$ can compute $(\Sigma_{\Theta}^i, \Theta_j(\Sigma_{\Theta}^j), \Psi_j(\Theta_{\Psi}^j))$, where $i \in I_0^-$ or $i \in I_0^+$ and the function $\Psi_j$ defined for the ideal world is given in Fig. 4.3. On the other hand, for the real world, we define $\Psi_j$ as follows:

$$\Psi_j(\Sigma_{\Theta}^i) = E_K(\Sigma_{\Theta}^i), \quad \Psi_j(\Theta_{\Psi}) = E_K(\Theta_{\Psi})$$

for $i \in I_0^-$ or $i \in I_0^+$. Therefore, each transcript $\tau^c$, where $i \in I_0^-$ or $i \in I_0^+$, is now modified to include the corresponding intermediate input-output values for the $i$-th instance of the construction.

Thus,

$$\tau^c = \{(M_{\theta_1}, T_{\theta_1}, \Sigma_{\Theta}^i, \Psi_j(\Sigma_{\Theta}^i), \Psi_j(\Theta_{\Psi}^j)), \ldots, (M_{\theta_s}, T_{\theta_s}, \Sigma_{\Theta}^i, \Psi_j(\Sigma_{\Theta}^i), \Psi_j(\Theta_{\Psi}^j))\}.$$ 

In all the following, the complete construction query transcript is

$$\tau_c = \bigcup_{i=1}^{u} \tau^c_i$$

and the overall transcript is $\tau = \tau_c \cup \tau_p$. The modified experiment only makes the distinguisher more powerful and hence the distinguishing advantage of $A$ in this experiment is no less than its distinguishing advantage in the former. Let $X_{\Theta}$ denote the random variable that takes a transcript $\tau$ realized in the real world. Similarly, $X_{\Theta}$ denotes the random variable that takes a transcript $\tau$ realized in the ideal world. The probability of realizing a transcript $\tau$ in the ideal (resp. real) world is called the ideal (resp. real) interpolation probability. A transcript $\tau$ is said to be attainable with respect to $A$ if its ideal interpolation probability is non-zero, and $\Theta$ denotes the set of all such attainable transcripts. Following these notations, we now state the main theorem of the H-coefficient technique [31]:

**Theorem 2 (H-Coefficient Technique).** Let $\Theta = \text{GoodT} \cup \text{BadT}$ be a partition of the set of attainable transcripts. Suppose there exists $\epsilon_{\text{ratio}} \geq 0$ such that for any $\tau = (\tau_c, \tau_p) \in \text{GoodT}$,

$$\frac{p_{\text{ce}}(\tau)}{p_{\text{id}}(\tau)} \geq \frac{\Pr[X_{\Theta} = \tau]}{\Pr[X_{\Theta} = \tau]} \geq 1 - \epsilon_{\text{ratio}}.$$ 

and there exists $\epsilon_{\text{bad}} \geq 0$ such that $\Pr[X_{\Theta} \in \text{BadT}] \leq \epsilon_{\text{bad}}$. Then

$$\text{Adv}_{\text{ce}}^{\text{mprf}}(A) \leq \epsilon_{\text{ratio}} + \epsilon_{\text{bad}}.$$  

(2)

Therefore, to prove the security of the construction using the H-coefficient technique, we need to identify the set of bad transcripts and compute an upper bound for their probability
in the ideal world. Then we find a lower bound for the ratio of the real to ideal interpolation probability for a good transcript. We have already identified the bad transcripts in Fig. 4.1 and Fig. 4.2. Therefore, it only remains to bound the probability of bad transcripts in the ideal world and provide a lower bound for the ratio of the real to ideal interpolation probability for a good transcript. Having explained the H-coefficient technique in the view of our construction, it follows that for each $i \in [u]$, $C_2[H, E_{(L_1^i, L_2^i, K^i)}] \rightarrow \tau^i$ denotes the following:

1. $\Sigma^i_a = (H^1_{L_1^i}(M^i_a)), \Theta^i_a = (H^2_{L_2^i}(M^i_a)),$
2. $E_{K^i}(\Sigma^i_a) = \Psi(\Sigma^i_a), E_{K^i}(\Theta^i_a) = \Psi(\Theta^i_a)$, and
3. $E_{K^i}(\Sigma^i_a) \oplus E_{K^i}(\Theta^i_a) = \tau^i_c$.

4.3 Bounding the Probability of Bad Transcripts

We call a transcript $\tau = (\tau_c, \tau_p)$ bad if at least one of the flags is set to 1 during the generation of the transcript in Fig. 4.1 and Fig. 4.2. Recall that $BadT \subseteq \Theta$ is the set of all attainable bad transcripts and $GoodT = \Theta \setminus BadT$ is the set of all attainable good transcripts. We bound the probability of bad transcripts in the ideal world as follows.

**Lemma 2.** Let $\tau = (\tau_c, \tau_p)$ be any attainable transcript. Let $X_{id}$ and $BadT$ be defined as above. Then

$$\Pr[X_{id} \in BadT] \leq \frac{q}{2^u} + \frac{2u^2}{2k_h + k} + \frac{2q\epsilon_{reg}}{2k} + \frac{q^2\epsilon_{univ}}{2n} + \frac{2q^2\epsilon_{reg}}{2k_h} + 3q^{4/3}\epsilon_{univ} + \frac{q^2\epsilon_{univ}}{2} + \frac{2q\epsilon_{reg}}{2n + k} + \frac{2q^2}{2n + k}.$$  

**Proof.** By abusing the notation, we refer the bad events by their corresponding flag variables as defined in Fig. 4.1, Fig. 4.2 and Fig. 4.3. That is we use Bad-Tag to refer to that event for which Bad-Tag flag has been set to 1. In other words, we say that the event Bad-Tag holds if and only if Bad-Tag flag has been set to 1. Using the union bound, we write

$$\Pr[X_{id} \in BadT] \leq \Pr[Bad-Tag] + \Pr[BadK] + \sum_{i=1}^4 \Pr[Badi | BadK] + \Pr[Bad-Samp | BadK].$$

We individually bound each bad event and then use Eqn. (3) to derive the result. In the subsequent analysis, we assume that $|K_h| = k_h$ and $|K| = k$.

4.3.1 Bounding Event Bad-Tag

For a fixed choice of indices, the probability of the event can be bound by $1/2^n$ as the outputs of the construction queries are sampled uniformly and independently of other random variables. Therefore, by summing over all possible choices of indices, we have

$$\Pr[Bad-Tag] \leq \frac{q}{2^n}. \quad (4)$$

4.3.2 Bounding Event BadK

For a fixed choice of indices, the probability of the event can be bound by $1/2^{k_h + k}$ as the event $K^{i_1} = K^{i_2}$ is independent of $L_1^{i_1} = L_1^{i_2}$ for each $b \in \{1, 2\}$. Therefore, summing over all possible choices of indices, we have

$$\Pr[BadK] \leq \frac{2u^2}{2k_h + k}. \quad (5)$$
4.3.3 Bounding Event Bad1 | BadK

We say that the event Bad1 | BadK holds if either of the events defined in line 5.(a) or in line 5.(b) of Fig. 4.2 holds. We refer to the event defined in line 5.(a) as B.11 and refer to the event defined in line 5.(b) as B.12

\[ \text{B.11 | BadK: For a fixed choice of indices, } \sum_a = u[0] \text{ is bound by the regular advantage of the hash function } H^{1}_{L_1}. \text{ As the hash key } L_1 \text{ is independent of the block cipher key } K, \text{ we have} \]

\[
\Pr[B.11 \mid \text{BadK}] \leq \sum_{i \in [s]} \sum_{j \in [s]} \Pr[K^i = J^j] \cdot \Pr[\sum_a = u[0]]
\]

\[
= \sum_{i \in [s]} \sum_{j \in [s]} \epsilon_{reg} \cdot \frac{1}{2^k} \leq \frac{q \epsilon_{reg}}{2^k}, \tag{6}
\]

where (1) holds due to the fact that \((q_1 \ldots + q_a) = q\) and \((p_1^2 \ldots + p_a^2) \leq p^2\).

\[ \text{B.12 | BadK: With an identical argument, one can show that the probability of the event B.12 can be bounded by } \frac{q \epsilon_{reg}}{2^k}, \text{i.e.,} \]

\[
\Pr[B.12 \mid \text{BadK}] \leq \frac{q \epsilon_{reg}}{2^k}. \tag{7}
\]

Therefore, by combining Eqn. (6) and Eqn. (7), we have

\[
\Pr[\text{Bad1} \mid \text{BadK}] = \Pr[B.11 \mid \text{BadK} \lor B.12 \mid \text{BadK}] \leq \frac{2q \epsilon_{reg}}{2^k}. \tag{8}
\]

4.3.4 Bounding Event Bad2 | BadK

We say that the event Bad2 | BadK holds if either of the events defined in line 7.(a) or in line 7.(b) or line 7.(c) or in line 7.(d) of Fig. 4.2 holds. We refer to the event defined in line 7.(a) as B.21, in line 7.(b) as B.22, in line 7.(c) as B.23 and finally in line 7.(d) as B.24

\[ \text{B.21 | BadK: For a fixed choice of indices, we analyze the probability of the event} \]

\[ \Sigma_a = \Sigma_b \land T_a = T_b, \]

Due to independence of the hash key \(L_1\) and \(T_a\), the probability of this joint event can be bound by the universal property of the \(H^1\) hash function and the randomness of \(T_a\). Therefore,

\[
\Pr[B.21 \mid \text{BadK}] \leq \sum_{i \in [s], a, b \in [q]} \Pr[\Sigma_a = \Sigma_b \land T_a = T_b] \leq \frac{q^2 \epsilon_{univ}}{2^{n+1}}. \tag{9}
\]

\[ \text{B.22 | BadK: We bound the event given BadK, i.e. even if the block cipher keys for users } i_1 \text{ and } i_2 \text{ collide, their corresponding hash keys, i.e., } L_1^{i_1} \text{ and } L_2^{i_2} \text{ do not collide. Given this event, for a fixed choice of indices, we bound } \Sigma_a = \Sigma_b \text{ using the regular property of the hash function } H^1 \text{ with the randomness of the hash key } L_1^{i_1}. \text{ Moreover, the first event is independent of the second event and can thus be bound exactly by } 2^{-kh}. \text{ Therefore,} \]

\[
\Pr[B.22 \mid \text{BadK}] \leq \sum_{i_1, i_2 \in [s]} \epsilon_{reg} \cdot \frac{1}{2^k} \leq \frac{q^2 \epsilon_{reg}}{2^k}. \tag{10}
\]
We say that the event \( \text{BadK} \) is identical to bounding \( B.21 \) and \( B.22 \) respectively. Hence,

\[
\Pr[B.23 \mid \text{BadK}] \leq \frac{q^2 \epsilon_{\text{univ}}}{2^n+1}, \quad \Pr[B.24 \mid \text{BadK}] \leq \frac{q^2 \epsilon_{\text{reg}}}{2^k_n}.
\]

Therefore, by combining Eqn. (9)-Eqn. (11),

\[
\Pr[\text{Bad2} \mid \text{BadK}] \leq \Pr[B.21 \mid \text{BadK}] + \Pr[B.22 \mid \text{BadK}] + \Pr[B.23 \mid \text{BadK}] + \Pr[B.24 \mid \text{BadK}] \\
\leq \frac{q^2 \epsilon_{\text{univ}}}{2^n} + \frac{2q^2 \epsilon_{\text{reg}}}{2^k_n}.
\]

### 4.3.5 Bounding Event \( \text{Bad3} \) \( \mid \) \( \text{BadK} \)

We say that the event \( \text{Bad3} \mid \text{BadK} \) holds if either of the events defined in line 9.(a) or in line 9.(b) of Fig. 4.2 holds. We refer to the event defined in line 9.(a) as \( B.31 \) and in line 9.(b) as \( B.32 \).

We first bound the event \( B.31 \mid \text{BadK} \). For a fixed choice of indices, we define an indicator random variable \( \mathbb{I}_{i,a,b} \) which takes the value 1 if \( \Sigma_a = \Sigma_b \), and 0 otherwise. Let \( \mathbb{I}^i = \sum_{a,b} \mathbb{I}_{i,a,b} \). By linearity of expectation,

\[
E[\mathbb{I}^i] = \sum_{a,b} E[\mathbb{I}_{i,a,b}] = \sum_{a,b} \Pr[\Sigma_a = \Sigma_b] \leq \frac{q^2 \epsilon_{\text{univ}}}{2^n}.
\]

Now,

\[
\Pr[B.31 \mid \text{BadK}] \leq \sum_{i \in [u]} \Pr[\{(a,b) \in [q_i^2] : \Sigma_a = \Sigma_b\}] \geq \frac{q_i^{2/3}}{q_i}
\]

\[
= \sum_{i=1}^{u} \Pr[\mathbb{I}^i \leq \frac{q_i^{2/3}}{q_i}] \leq \sum_{i=1}^{u} \frac{q_i^2 \epsilon_{\text{univ}}}{2q_i^{2/3}} \leq \frac{q^{4/3} \epsilon_{\text{univ}}}{2},
\]

where (1) holds due to the Markov inequality.

Similar to \( B.31 \mid \text{BadK} \), we bound \( B.32 \mid \text{BadK} \) as follows:

\[
\Pr[B.32 \mid \text{BadK}] \leq \frac{q^{4/3} \epsilon_{\text{univ}}}{2}.
\]

Therefore, by combining Eqn. (13) and Eqn. (14), we have

\[
\Pr[\text{Bad3} \mid \text{BadK}] = \Pr[B.31 \mid \text{BadK} \lor B.32 \mid \text{BadK}] \leq \frac{q^{4/3} \epsilon_{\text{univ}}}{2}.
\]

### 4.3.6 Bounding Event \( \text{Bad4} \) \( \mid \) \( \text{BadK} \)

We say that the event \( \text{Bad4} \mid \text{BadK} \) holds if either of the events defined in line 11.(a) or in line 11.(b) or in line 11.(c) of Fig. 4.2 holds. We refer to the event defined in line 11.(a) as \( B.41 \), line 11.(b) as \( B.42 \) and in line 11.(c) as \( B.43 \).

Due to independence of the hash key \( L_1^i \) and \( L_2^i \), for a fixed choice of indices, the probability of this joint event can be bound by the universal property of the individual hash functions \( H^1 \) and \( H^2 \). Therefore, varying over all possible choices of indices, we have

\[
\Pr[B.41 \mid \text{BadK}] \leq \sum_{i \in [u]} \Pr[\Sigma_a = \Sigma_b \land \Theta_a^i = \Theta_b^i] \leq \sum_{i \in [u]} \Pr[\Sigma_a = \Sigma_b] \cdot \Pr[\Theta_a^i = \Theta_b^i] \leq \frac{q^2 \epsilon_{\text{univ}}^2}{2}.
\]
> **Bounding B.42 | BadK and B.43 | BadK:** We first bound the event B.42 | BadK. We bound this event given B.31. This results in the fact that for a fixed $i \in [u]$, the number of quadruples $(a, b, c, d)$ such that $\Sigma^i_a = \Sigma^i_b$, $\Sigma^i_c = \Sigma^i_d$ holds is at most $q_i^{4/3}$. For a fixed choice of such quadruples, the event $\Theta^i_k = \Theta^i_j$ holds with probability at most $\epsilon_{\text{univ}}$ due to the universal property of the hash function $H^2$. Therefore,

$$\Pr[B.42 \mid B.31 \land \text{BadK}] \leq \sum_{i \in [u]} q_i^{4/3} \epsilon_{\text{univ}} \leq q^{4/3} \epsilon_{\text{univ}}. \quad (17)$$

Similar to B.42, we bound B.43 as follows:

$$\Pr[B.43 \mid B.31 \land \text{BadK}] \leq q^{4/3} \epsilon_{\text{univ}}. \quad (18)$$

By combining Eqn. (16), Eqn. (17) and Eqn. (18), we have

$$\Pr[\text{Bad4} \mid \text{BadK}] \leq \frac{q^2 \epsilon_{\text{univ}}^2}{2} + 2q^{4/3} \epsilon_{\text{univ}}. \quad (19)$$

### 4.3.7 Bounding Event Bad-Samp | BadK

We consider bounding this event as a union of several events, namely for a fixed $i \in [u], j \in [s]$ and $a \in [q_i]$, we define

$$\text{BS}_{i,j,a} \triangleq K^i = J^j \land Z^i_a \oplus T^i_a \in \text{Ran}(E_j).$$

Then we say that the event Bad-Samp | BadK holds if there exists an $i \in [u]$ and $j \in [s]$ such that $\text{BS}_{i,j,a}$ holds, where $Z^i_a \leftarrow \{0,1\}^n \setminus \text{Ran}(E_j)$. We first fix an index $j \in [s]$, which determines $\mathcal{I}_j^a$, an index $i \in \mathcal{I}_j^a$ and $a \in [q_i]$. For this choice of indices, the probability that $K^i = J^j \land Z^i_a \oplus T^i_a \in \text{Ran}(E_j)$ holds is at most $2^{-(k+n) \cdot (p_j + q_j)}$. This is due to the fact that the cardinality of $\text{Ran}(E_j)$ is bounded above by $(p_j + q_j)$, where $q_j$ is the number of tuples $(\Sigma^i_a, \Theta^i_a)_{i \in \mathcal{I}_j^a, a \in [q_i]}$ which have been added into the set $\text{Dom}(E_j)$ such that $K^i = J^j$. Moreover, as the event $K^i = J^j$ is independent of $Z^i_a \oplus T^i_a \in \text{Ran}(E_j)$, by taking the union bound, we have

$$\Pr[\text{Bad-Samp}] \leq \sum_{j=1}^{s} \sum_{i \in \mathcal{I}_j^a} \sum_{a \in [q_i]} \frac{1}{2^k} \cdot \frac{p_j + q_j}{2^n - (p_j + q_j)} \leq \frac{2q^i p + 2q^i q}{2^{n+k}}. \quad (20)$$

Note that the number of choices for $(i, a)$ is at most $q$ and the number of choices for $j$ is $s$. Thus, summing over all possible choices of $(i, j, a)$ and by assuming $(p_j + q_j) \leq 2^{n-1}$ and $\sum_{j=1}^{s} (p_j + q_j) \leq (p + q)$, we have the result.

Finally, the result follows by combining Eqn. (4)-Eqn. (20). 

### 4.4 Analysis of Good Transcripts

In this section, we compute a lower bound for the ratio of the real to ideal interpolation probability for a good transcript. We first consider the set of transcripts $\mathcal{Q}^+$. For each $j \in [s]$ and for each $i \in \mathcal{I}_j^a$, we consider the sequence

$$\tilde{\Sigma}^i := (\Sigma^i_1, \Sigma^i_2, \ldots, \Sigma^i_{q_i}), \tilde{\Theta}^i := (\Theta^i_1, \Theta^i_2, \ldots, \Theta^i_{q_i}).$$

From this sequence, we construct a bipartite graph $G_i$, where the nodes in one partition represent values $\Sigma^i_a$ and the nodes in other, $\Theta^i_a$; an edge connects the nodes $\Sigma^i_a$ and $\Theta^i_b$. If $\Sigma^i_a = \Sigma^i_b$, then we merge the corresponding nodes into a single node, and similarly for
Let \( \Theta_0 = \Theta_1 \). This leads us to break the graph into \( w_i \) components. As the transcript is good, it is easy to see that each component is acyclic (otherwise, B.41 would have been satisfied) and contains a path of length at most 3 (otherwise either B.42 or B.43 would have been satisfied). Let \( v_i \) be the total number of nodes of the graph \( G_i \). Similar to \( \mathcal{Q}^f \), we consider \( \mathcal{Q}^\tau \). For each \( j \in [r'] \) and for each \( i \in I^f_j \), consider the sequence

\[
\tilde{\Theta}^i := (\Theta^i_1, \Theta^i_2, \ldots, \Theta^i_{q_i}).
\]

Similar to \( G_i \), we construct a bipartite graph \( H_i \), one of whose partitions represents the nodes corresponding to \( \Sigma^i_a \) and the other, the nodes corresponding to \( \Theta^i_a \); an edge connects the nodes corresponding to \( \Sigma^i_a \) and \( \Theta^i_a \). If two nodes represent the same values, we merge them into a single node. Let \( w'_i \) be the number of components of \( H_i \) and \( v'_i \) be the total number of vertices. Then for a good transcript \( \tau = (\tau_c, \tau_p) \), realizing \( \tau \) is almost as likely in the real world as in the ideal world:

**Lemma 3 (Good Lemma).** Let \( \tau = (\tau_c, \tau_p) \in \text{GoodT} \) be a good transcript. Let \( X_{re} \) and \( X_{id} \) be defined as above. Then

\[
\Pr[X_{re} = \tau] \geq 1 - \frac{9q^{4/3}}{8 \cdot 2^2} - \frac{q^2}{2^2} - \frac{9q^{7/3}}{8 \cdot 2^2} - \frac{8q^4}{3 \cdot 2^3}.
\]

**Proof.** We are now ready to calculate the real interpolation probability. For this, we must bound the total number of input-output pairs on which the block cipher \( E \) with different keys is executed. As the transcript releases the \( 2k_h \)-bit hash keys and the \( k \)-bit block cipher key for each user, it contributes to a term \( 2^{-(2k_h+k)} \) in the real interpolation probability calculation. Now, for each \( j \in [r] \), the block cipher \( E \) with key \( J^j \) is evaluated on a total of

\[
p_j + \sum_{i \in I^f_j} v_i
\]

input-output pairs. For the remaining ideal cipher keys, with which none of the users' block cipher keys have collided, we have \( p_j \) input-output pairs, which are fixed due to the evaluation of the block cipher with those ideal cipher keys. Moreover, for each \( j \in [r'] \), the block cipher \( E \) is evaluated on a total of \( \sum_{i \in I^f_j} v'_i \) input-output pairs with key \( K^j \). Summarizing the above,

\[
\Pr[X_{re} = \tau] = \prod_{i=1}^u \frac{1}{2^{2k_h+k}} \left( \prod_{j=1}^{r} p_j \frac{1}{p_j + \sum_{i \in I^f_j} v_i} \right) \prod_{j \in |s| \setminus [r]} \frac{1}{p_j} \left( \prod_{j=1}^{r'} \frac{1}{p_j} \right) \left( \prod_{i \in I^f_j} \frac{1}{p_j} \sum_{i \in I^f_j} v'_i \right)
\]

**Ideal Interpolation Probability:** The term \( \prod_{i=1}^u 2^{-nq_i} \), which is contributed to the ideal interpolation probability due to the sampling of responses of the adversarial query, samples \( 2k_h \)-bit hash keys and \( k \)-bit block cipher keys for all \( u \) users. For each \( j \in [r] \), and for each \( i \in I^f_j \), we construct the graph \( G_i \) as defined above. It is easy to see that for each \( j \in [r] \) and for each \( i \in I^f_j \), the graph \( G_i \) is good. Next, for each \( j \in [r] \) and for each \( i \in I^f_j \), we sample the value of a node for each component of the graph \( G_i \). Hence, for \( j \in [r] \), the total number of sample points is

\[
p_j + \sum_{i \in I^f_j} w_i.
\]

Moreover, for each \( j \in [s] \setminus [r] \), the total number of sample points is \( p_j \). Subsequently, we consider the set of transcripts \( \mathcal{Q}^\tau \). For each \( j \in [r'] \), and for each \( i \in I^f_j \), we construct the
graph $H_i$ as defined above, and compute the set $S_j$ for each $j \in [r']$ as defined in line 14 of Fig. 4.3 (which is defined as the number of tuples $(Q'_i, R'_i)$ such that $Q'_i \oplus R'_i = T'_i$ for all $i \in I_j^d$ and for all $a \in [q_i]$). In summary,

$$\Pr[X_{id} = \tau] = \prod_{i=1}^{u} \frac{1}{2^{nq_i}} \cdot \prod_{j=1}^{r} \frac{1}{2^{2k_h} \tau^k} \left( \prod_{j=1}^{r} \frac{1}{P(2^n, p_j + \sum_{i \in I_j^d} w_i)} \right) \cdot \prod_{j \in |\sigma| \setminus |r|} \frac{1}{P(2^n, p_j)} \left( \prod_{i=1}^{r} \frac{1}{|S_j|} \right).$$

(22)

**Calculation of the ratio:** By plugging in the value of $|S_j|$ from Lemma 1 into Eqn. (22) and then taking the ratio of Eqn. (21) to Eqn. (22), we have

$$P(\tau) = \prod_{i=1}^{u} \frac{1}{2^{nq_i}} \cdot \prod_{j=1}^{r} \frac{P(2^n, p_j + \sum_{i \in I_j^d} w_i)}{P(2^n, p_j) \cdot \sum_{i \in I_j^d} v_i} \cdot \prod_{j=1}^{r'} \frac{1}{|S_j|} \cdot \prod_{j=1}^{r'} \left( \prod_{i=1}^{r} \frac{P(2^n, \sum_{i \in I_j^d} v_i)}{P(2^n, \sum_{i \in I_j^d} v_i') \cdot 2^{n \sum_{i \in I_j^d} (v_i' - w_i')}} \right)$$

$$= \prod_{i=1}^{u} \frac{1}{2^{nq_i}} \cdot \prod_{j=1}^{r} \frac{P(2^n, p_j - \sum_{i \in I_j^d} w_i, \sum_{i \in I_j^d} (v_i - w_i))}{P(2^n, \sum_{i \in I_j^d} v_i) \cdot \sum_{i \in I_j^d} v_i'} \cdot \prod_{j=1}^{r'} \left( \prod_{i=1}^{r} \frac{P(2^n, \sum_{i \in I_j^d} v_i)}{P(2^n, \sum_{i \in I_j^d} v_i') \cdot 2^{n \sum_{i \in I_j^d} (v_i' - w_i')}} \right)$$

$$= \prod_{j=1}^{r} \frac{P(2^n, p_j - \sum_{i \in I_j^d} w_i, \sum_{i \in I_j^d} (v_i - w_i))}{P(2^n, \sum_{i \in I_j^d} v_i) \cdot \sum_{i \in I_j^d} v_i'} \cdot \prod_{j=1}^{r'} \left( \prod_{i=1}^{r} \frac{P(2^n, \sum_{i \in I_j^d} v_i)}{P(2^n, \sum_{i \in I_j^d} v_i') \cdot 2^{n \sum_{i \in I_j^d} (v_i' - w_i')}} \right)$$

$$\geq \left( 1 - \sum_{j=1}^{r'} \epsilon_j \right)^{\geq 1} \geq \left( 1 - \sum_{j=1}^{r'} \left( \frac{9q_i^2}{8 \cdot 2n} + \frac{3q_i^2}{2 \cdot 2m} + \frac{q_i^2}{2m} + \frac{9q_i^2}{8 \cdot 2n} + \frac{8q_i^4}{3 \cdot 2m} \right) \right)$$

$$\geq \left( 1 - \sum_{j=1}^{r'} \sum_{i \in I_j^d} \left( \frac{9q_i^{4/3}}{8 \cdot 2n} + \frac{3q_i^{8/3}}{2 \cdot 2m} + \frac{q_i^2}{2m} + \frac{9q_i^{7/3}}{8 \cdot 2n} + \frac{8q_i^4}{3 \cdot 2m} \right) \right)$$

(1)

$$\geq 1 - \left( \frac{9q_i^{4/3}}{8 \cdot 2n} + \frac{3q_i^{8/3}}{2 \cdot 2m} + \frac{q_i^2}{2m} + \frac{9q_i^{7/3}}{8 \cdot 2n} + \frac{8q_i^4}{3 \cdot 2m} \right)$$

where (1) holds due to the fact that $q_i^2 \leq q_i^{4/3}$ for all $i \in I_j^d$ such that $j \in [r']$. Note that for each $j \in [r]$, $\sum_{i \in I_j^d} (v_i - w_i)$ denotes the total number of edges in the graph $\bigcup_{i \in I_j^d} G_i$, which is $\sum_{i \in I_j^d} q_i$. Similarly, for each $j \in [r']$, $\sum_{i \in I_j^d} (v_i' - w_i')$ denotes the total number of edges in the graph $\bigcup_{i \in I_j^d} H_i$, which is $\sum_{i \in I_j^d} q_i$. 
5 Tight Security Bound of Two-Keyed Polyhash based DbHtS Construction

Two-keyed Polyhash-based DbHtS construction $C_2[\text{PH-DbH}, E]$, as proposed in [13], is the instantiation of the Two-Keyed-DbHtS framework which is build on the Polyhash based double block hash function $\text{PH-DbH}$. In [13], the PRF security of $C_2[\text{PH-DbH}, E]$ has been proven to be roughly in the order of $q^2 \ell^2 / 2^{2n}$ in the single-user setting. In this section we improve its bound up to $2^{3n/4}$ queries in the multi-user setting. Moreover, the proof is based on the ideal cipher model. Before going to the security proof of the construction, we first revisit to the two-keyed Polyhash-based DbHtS construction.

PolyHash [14, 6, 34] is a very efficient algebraic hash function. For a fixed natural number $n$, it first samples an $n$-bit key $L$ uniformly at random from $\{0, 1\}^n$. To apply this function on a message $M \in \{0, 1\}^*$, we first apply an injective padding function $10^n$ (i.e. append a bit 1 followed by a minimum number of zeroes to the message $M$ so that the total number of bits in the padded message becomes a multiple of $n$). Let the padded message be $M^* = |M_1||M_2|\ldots||M_l$, where $l$ is the number of $n$-bit blocks in it. Then, we define the PolyHash function as follows:

$$\text{PH}_L(M^*) \triangleq M_1 \cdot L_l \oplus M_2 \cdot L_{l-1} \oplus \ldots \oplus M_l \cdot L,$$

where $l$ is the number of blocks of $M$ and the multiplications are defined in the field $GF(2^n)$. Then Polyhash [26] is $\ell/2^n$-regular, $\ell/2^n$-axu and $\ell/2^n$-universal, as shown in the following lemma, where $\ell$ is the maximum number of message blocks (the proof of the lemma is related to a result on the number of distinct roots of a polynomial):

**Lemma 4.** Let $\text{PH}$ be the PolyHash function as defined above. Then $\text{PH}$ is $\ell/2^n$-regular, $\ell/2^n$-almost-xor universal and $\ell/2^n$-uniiversal.

From Lemma 4, a simple corollary immediately follows:

**Corollary 1.** Let $\text{fix}_b(\text{PH})$ be the variant of the Polyhash function in which the least significant bit of the $n$-bit output of the function is fixed to bit $b$. Then, $\text{fix}_b(\text{PH})$ is a $2\ell/2^n$-regular, $2\ell/2^n$-almost-xor universal and $2\ell/2^n$-universal hash function.

We now define the Polyhash-based double-block hash function, (PH-DbH function):

$$\text{PH-DbH}_{(L_1, L_2)}(M) \triangleq \left( \begin{array}{c} \text{fix}_0(\text{PH}_{L_1}(M)) \\ \text{fix}_1(\text{PH}_{L_2}(M)) \end{array} \right). \quad (23)$$

Thus, two independent instances of the Polyhash function keyed with two independent keys $L_1$ and $L_2$ are applied separately to a message $M$, and the least significant bit of their output is chopped and prepended with bits 0 and 1 respectively. The two-keyed Polyhash-based DbHtS construction can now be defined directly from the Two-Keyed-DbHtS construction as follows: encrypt $\text{fix}_0(\text{PH}_{L_1}(M))$ and $\text{fix}_1(\text{PH}_{L_2}(M))$ through a block cipher $E_K$ and xor the result together to produce the output. An algorithmic description of the construction is shown in Fig. 5.1.

Clearly, the PH-DbH function is a good double-block hash function as the individual hash functions $H^1$ and $H^2$ are both $2\ell/2^n$-regular and universal. Furthermore, for a randomly chosen pair of keys $L_1, L_2$, and for any pair of messages $M, M' \in \{0, 1\}^*$,

$$\Pr[\text{fix}_0(\text{PH}_{L_1}(M)) = \text{fix}_1(\text{PH}_{L_2}(M'))] = 0.$$

Therefore, combining the Corollary 1 with Theorem 1, we derive the following security of the two-keyed PolyHash-based DbHtS construction $C_2[\text{PH-DbH}, E]$. 

C2[PH-DbH,E](K1,K2,K)(M) =  
1: \Sigma = \text{fix}_0(\text{PH}_K(M));
2: \Theta = \text{fix}_1(\text{PH}_K(M));
3: T = E_K(\Sigma) \oplus E_K(\Theta);
return T;

\text{PH}_L(M) =  
1: M_1 \parallel \ldots \parallel M_{\ell^*} \rightarrow M||10^*;
2: Y = M_1 \cdot L^* \parallel M_2 \cdot L^{*-1} \parallel \ldots \parallel M_{\ell} \cdot L; 
return Y;

\textbf{Figure 5.1:} The two-keyed Polyhash-based DbHtS construction \(C_2[\text{PH-DbH, E}]\) with PH-DbH as the underlying double-block hash function. \(M_1||M_2||\ldots||M_{\ell^*} \rightarrow M||10^*\) denotes the parsing of message \(M||10^*\) into \(n\) bit strings.

\textbf{Theorem 3.} Let \(K\) be a non-empty finite set. Let \(E : \mathcal{K} \times \{0,1\}^n \rightarrow \{0,1\}^n\) be an \(n\)-bit block cipher and \(\text{PH-DbH} : (\{0,1\}^n \times \{0,1\}^n) \times \{0,1\}^* \rightarrow (\{0,1\}^n)^2\) be the PolyHash-based double-block hash function as defined above. Then any computationally unbound distinguisher making a total of \(q\) construction queries across all \(u\) users such that each queried message is at most \(\ell\) blocks long with \(\ell \leq 2^{n-2}\) and a total of \(p\) primitive queries to the block cipher \(E\) can distinguish \(C_2[\text{PH-DbH, E}]\) from an \(n\)-bit uniform random function with advantage

\[
\text{Adv}^\text{mpf}_{C_2[\text{PH-DbH, E}]}(u, q, p, \ell) \leq \frac{9q^4}{8 \cdot 2^n} + \frac{3q^{8/3}}{2 \cdot 2^{2n}} + \frac{q^2}{2^{2n}} + \frac{9q^{7/3}}{8 \cdot 2^{2n}} + \frac{q^2}{2^{2n}} + \frac{2q}{n^2} + \frac{2^n}{2^{2n+k}} + \frac{4q^2\ell}{2^{2n+k}} + \frac{4q^2\ell}{2^{2n+k}} + \frac{8q^{1/3}\ell}{2^n} + \frac{4q^2\ell^2}{2^{2n}} + \frac{2q^2}{2^{2n+k}} + \frac{2^2}{2^{2n+k}}.
\]

\textbf{Remark 2.} We would like to mention that the definition of the Polyhash function used in this paper is different from that used in [16]. Nevertheless, one can also establish the \(3n/4\)-bit multi-user security of the two-keyed PolyHash-based DbHtS construction with the Polyhash function used in [16].

\section{Conclusion and Future Problems}

In this paper, we have shown that the Two-Keyed-DbHtS construction is multi-user secured up to \(2^{3n/4}\) queries in the ideal-cipher model. As an instantiation of the result, we have shown that Polyhash-based DbHtS provides \(3n/4\)-bit multi-user security in the ideal-cipher model. Combining it with the generic result on the attack complexity of the DbHtS construction makes the bound tight. However, we cannot apply this result to analyze the security of 2K-SUM-ECBC, 2K-PMAC_Plus and 2K-LightMAC_Plus, as their underlying DbH functions are based on block ciphers, and our proof technique does not support their security analysis in the ideal-cipher model. This is because the underlying DbH function of these constructions is build on the top of block ciphers. We believe that proving \(3n/4\)-bit security of the DbHtS construction based on block cipher-based double-block hash functions needs a careful study.

\section*{References}


