Tight Multi-User Security Bound of DbHtS

Nilanjan Datta[1] and Avijit Dutta[1] and Mridul Nandi[2] and Suprita Talnikar[2]

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 that DbHtS is secure up to $2^{2n/3}$ queries in the multi-user setting independent of the number of users, where the double block hash function H consists of two independent n-bit keyed hash function $(\mathsf{H}_{K_h,1}, \mathsf{H}_{K_h,2})$ 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 key-reduced variants of DbHtS MACs including 2K-SUM-ECBC, 2K-PMAC_Plus and 2K-LightMAC_Plus without requiring any additional domain separation. Recently, Guo and Wang have shown three instantiations of DbHtS framework where each of their n-bit keyed hash functions is $O(2^{-n})$ universal and regular but the constructions are itself 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 DbHtS construction in the ideal-cipher model. As an instantiation, we show that Polyhash based DbHtS construction is multi-user secured up to $2^{3n/4}$ queries in the ideal-cipher model. Moreover, due to the result of 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 know paradigm of designing variable input length PRF, in which an input message of arbitrary length is hashed and then the hash value is encrypted through a PRF to obtain a short tag. Most of the popular well known MACs including CBC-MAC [3], PMAC [9], OMAC [19], LightMAC [23] are designed using the HtP paradigm. Although the method is simple, in particular being deterministic and stateless, yet the security bound of the 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, if we instantiate PMAC with AES-128, then it permits roughly 2^{48} queries (using $5\ell q^2/2^n$ [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 to use 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 success 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, 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, if the mode in which the 3-DES is used, provides only birthday bound security, then it does not give adequate 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 birthday bound security instead of replacing the cipher with a larger block size is a better option. [1]

Double-block Hash-then-Sum. In this line of research, 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 encryption of two n-bit hash values are 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 finally their encryptions are 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 in short DbHtS, defined as follows:

$$\mathsf{DbHtS}(M) \overset{\Delta}{=} \mathsf{E}_{K_1}(\Sigma) \oplus \mathsf{E}_{K_2}(\Theta), \quad (\Sigma, \Theta) \leftarrow \mathsf{H}_{K_h}(M),$$

where H_{K_h} 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 [2], 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. Until now, we have 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 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

[1] Note that there are no standard block ciphers of size higher than 128 bits.

[2] A family of keyed hash functions is said to be universal if for any distinct x and x', the probability that for a randomly sampled hash function makes a collision in their hash value is negligible.

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 bound 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 such 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 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 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 with the single-user security.

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 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 makes a single query and the number of users reaches q. Thus, a direct analysis is 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, as defined below,

$$\mathsf{Two\text{-}Keyed\text{-}DbHtS}(M) \overset{\Delta}{=} \mathsf{E}_K(\mathsf{H}_{K_h,1}(M)) \oplus \mathsf{E}_K(\mathsf{H}_{K_h,2}(M))$$

is secured 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 $\mathsf{H}_{K_h,1}$ and $\mathsf{H}_{K_h,2}$. In particular, they have shown that if each of the $\mathsf{H}_{K_h,1}$ and $\mathsf{H}_{K_h,2}$ are $O(2^{-n})$

[3] two-keyed stands for one hash key and one block cipher key.

regular and $O(2^{-n})$ universal [4], then the multi-user security bound of the two-keyed DbHtS is of the order

$$\frac{qp\ell}{2k+n} + \frac{q^3}{2^{2n}} + \frac{q^2p+qp^2}{2^{2k}},$$

where q is the total number of MAC queries across all u users, p is the total number of ideal cipher queries, ℓ is the maximum number of message blocks among all queries and n, k are the block size and the key size of the 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.

1.1 Issue of CRYPTO'21 Paper [32]

Two-Keyed-DbHtS framework has been proven to be multi-user secured upto $2^{2n/3}$ queries in the ideal-cipher model [32] under the assumption that the each of the underlying n-bit independent keyed hash functions are $O(2^{-n})$-universal and regular. As an instantiation of the framework, they have shown $2n/3$-bit multi-user security of 2K-SUM-ECBC, 2K-LightMAC_Plus and 2K-PMAC_Plus in the ideal-cipher model by proving the regular advantage and the universal advantage of the underlying hash functions upto $O(\ell/2^n)$, where ℓ is the maximum number of message blocks among all queries. However, the analysis of the regular and the universal advantage of the hash functions of the above three constructions have not been proven in the ideal-cipher model; instead the authors have shown these bounds in the standard model (where the adversary is not allowed to query the underlying block ciphers of the hash functions of all three constructions). We believe that to prove the security of the constructions in the ideal-cipher model for the block-cipher based DbH function, one needs to give a generalized definition of the universal and regular advantage in the ideal-cipher model and prove their security under that model only, which were missing in [32].

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

$$S(J) := \{(W_a^i, X_a^i) \in \{0,1\}^n \setminus \mathsf{Ran}(\Phi_j)^{(2|F(J)|)} : W_a^i \oplus X_a^i = T_a^i\}.$$

Then for all $(i, a) \in F(J)$, (U_a^i, V_a^i) is sampled from $S(J)$ and then they are set as the permutation output of Σ_a^i and Θ_a^i respectively and finally, they have shown a lower bound on the cardinality of the set $S(J)$ from Lemma 2. The fallacy is that in Lemma 2, they have proved that cardinality of the set

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

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

The third issue is that in a recent work, Guo and Wang [16] have came up with three instantiations of Two-Keyed-DbHtS paradigm, where they have shown that even if the

[4] A family of keyed hash function is said to be ϵ_1-regular if for any x and y, the probability that a randomly sampled hash function from the family maps x to y is ϵ_1; it is said to be ϵ_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 ϵ_2.

underlying two independent n-bit keyed hash functions of Two-Keyed-DbHtS are $O(2^{-n})$ regular and $O(2^{-n})$ universal, but plugging them up in the Two-Keyed-DbHtS framework does not yield $2n/3$-bit PRF security. In particular, they have shown a birthday bound distinguishing attack on each of the individual constructions. However, they have been unable to find any birthday bound attack on 2K-SUM-ECBC, 2K-PMAC_Plus and 2K-LightMAC_Plus. This leaves a room to quest for the properties of the underlying double block hash functions which makes Two-Keyed-DbHtS beyond birthday bound secure without requiring any additional domain separation technique.

1.2 Our Contribution

The contribution in this paper is twofold: in the first part of the paper we first define the notion of a **good** double-block hash function and prove that if the underlying $2n$-bit DbH function of the Two-Keyed-DbHtS construction is good such that it is the concatenation of two independent n-bit keyed hash functions, each of which is ϵ_{reg}-regular and ϵ_{univ}-universal, then the multi-user security bound of our construction in the ideal-cipher model is of the order

$$\frac{9q^{4/3}}{8\cdot 2^n}+\frac{3q^{8/3}}{2\cdot 2^{2n}}+\frac{q^2}{2^{2n}}+\frac{9q^{7/3}}{8\cdot 2^{2n}}+\frac{8q^4}{3\cdot 2^{3n}}+\frac{q}{2^n}+\frac{2u^2}{2^{k_h+k}}+\frac{2q^2}{2^{n+k}}$$
$$+\frac{2qp\epsilon_{\text{reg}}}{2^k}+\frac{q^2\epsilon_{\text{univ}}}{2^n}+\frac{2q^2\epsilon_{\text{reg}}}{2^{k_h}}+3q^{4/3}\epsilon_{\text{univ}}+\frac{q^2\epsilon_{\text{univ}}^2}{2}+\frac{2qp}{2^{n+k}}$$

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_h 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 in the second part of the paper that Polyhash based DbHtS construction is multi-user secured upto $2^{3n/4}$ queries in the ideal cipher model. The security proof of the construction crucially depends on a refined result of the 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 Gaëtan et al. [21] on the DbHtS paradigm with $2^{3n/4}$ queries, the multi-user security bound of our construction is tight.

Organization. We developed the required notations and security definitions of cryptographic primitives in Sect. 2. We depict the construction and its security result in Sect. 3 and in Sect. 4, we have presented the security proof of the construction. We instantiated 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 *block*. For a pair of blocks $x=(x_\ell,x_{\mathtt{r}})\in\{0,1\}^{2n}$, we write $\mathsf{left}(x)$ to denote x_ℓ and $\mathsf{right}(x)$ to denote $x_{\mathtt{r}}$. 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 $x,y\in\{0,1\}^n$ 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 $1\leq|x_l|\leq n$. For $x\in\{0,1\}^n$, where $x=x_{n-1}\|\ldots\|x_0$, $\mathsf{lsb}(x)$ 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 $x = x_{n-1}\| \dots \|x_0$ and returns an another binary string $x' = (x_{n-1}\| \dots \|b)$ where $\mathsf{lsb}(x)$ is fixed to bit b.

Given a finite set $\mathcal{S}$ and a random variable X, we write $X \leftarrow\!\!\$\, \mathcal{S}$ to denote that X is sampled uniformly at random from $\mathcal{S}$. We say that $X_1, X_2, \dots, X_q$ are with replacement (wr) sampled from $\mathcal{S}$, which we denote as $X_1, X_2, \dots X_q \leftarrow\!\!\$\, \mathcal{S}$, if for each $i \in [q]$, $X_i \leftarrow\!\!\$\, \mathcal{S}$. We also use this notion to denote that these random variables are sampled uniformly and independently from $\mathcal{S}$. For a finite subset $\mathcal{S}$ of $\mathbb{N}$, $\max \mathcal{S}$ denotes the maximum valued elements of $\mathcal{S}$. ϕ denotes the empty set. We write $\mathcal{S} \leftarrow \phi$ to denote that $\mathcal{S}$ is defined to be an empty set. We use the same notation $\Phi \leftarrow \phi$ to denote that the function Φ is undefined at every point of its domain. Moreover, the same notation $Y \leftarrow X$ is used to denote the assigment of variable X to Y.

The set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathsf{Func}(\mathcal{X}, \mathcal{Y})$. Similarly, the set of all permutations over $\mathcal{X}$ is represented by $\mathsf{Perm}(\mathcal{X})$. A function Φ is said to be a *block function* if it maps elements from an arbitrary domain to $\{0,1\}^n$. Set of all block functions with domain $\mathcal{X}$ is denoted as $\mathsf{Func}(\mathcal{X})$. [5] We call Φ to be a *double-block function* if it maps elements from an arbitrary set $\mathcal{X}$ to $(\{0,1\}^n)^2$. For a given double block function $\Phi : \mathcal{X} \to \{0,1\}^{2n}$, we write $\Phi_\ell : \mathcal{D} \to \{0,1\}^n$ such that for every $x \in \mathcal{X}$, $\Phi_\ell(x) = \mathsf{left}(\Phi(x))$. Similarly, we write $\Phi_{\mathtt{r}} : \mathcal{X} \to \{0,1\}^n$ such that for every $x \in \mathcal{X}$, $\Phi_{\mathtt{r}}(x) = \mathsf{right}(\Phi(x))$. For two block functions $\Phi_\ell : \mathcal{X} \to \{0,1\}^n$ and $\Phi_{\mathtt{r}} : \mathcal{X} \to \{0,1\}^n$, one can naturally define a double block function $\Phi : \mathcal{X} \to \{0,1\}^{2n}$ such that $\Phi(x) = (\Phi_\ell(x), \Phi_{\mathtt{r}}(x))$, which we write as $\Phi = (\Phi_\ell, \Phi_{\mathtt{r}})$. For a finite set $\mathcal{X}$ and an integer q, we write $\mathcal{X}^{(q)}$ to denote the set $\{(x_1, x_2, \dots, x_q) : x_i \in \mathcal{X}, x_i \neq x_j\}$. For integers $1 \leq b \leq a$, we write $\mathbf{P}(a,b)$ to denote $a(a-1)\dots(a-b+1)$, where $\mathbf{P}(a,0) = 1$ by convention. Therefore, $|\mathcal{X}^{(q)}| = \mathbf{P}(|\mathcal{X}|, q)$.

2.1 Distinguishing Advantage

An adversary A is modeled as a randomized algorithm with access to some external oracle $\mathcal{O}$. Such an adversary is called an *oracle adversary*. An oracle $\mathcal{O}$ is an algorithm itself which could be a cryptographic scheme being analyzed. The interaction between A and $\mathcal{O}$, denoted by $\mathsf{A}^{\mathcal{O}}$, generates a transcript $\tau = \{(x_1, y_1), (x_2, y_2), \dots, (x_q, y_q)\}$, where $x_1, x_2, \dots, x_q$ are q queries of A to oracle $\mathcal{O}$ and $y_1, y_2, \dots, y_q$ be the corresponding responses, where $y_i = \mathcal{O}(x_i)$. We assume that A is **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 $\mathcal{O} \in \{\mathsf{F}, \mathsf{G}\}$, A outputs 1, which is denoted as $\mathsf{A}^{\mathcal{O}} \Rightarrow 1$. Such an adversary is called a *distinguisher* and the game is called a *distinguishing game*. The task of the distinguisher in a distinguishing game is to tell which of the two systems it has interacted with. The advantage of distinguisher A in distinguishing the random system F from G is defined as

$$\mathbf{Adv}_{\mathsf{G}}^{\mathsf{F}}(\mathsf{A}) \overset{\Delta}{=} |\ \Pr[\mathsf{A}^{\mathsf{F}} \Rightarrow 1] - \Pr[\mathsf{A}^{\mathsf{G}} \Rightarrow 1]\ |,$$

where the above probability is defined over the probability spaces of A and $\mathcal{O}$. The maximum advantage in distinguishing F from G is defined as the

$$\max_{\mathsf{A} \in \mathcal{A}} \mathbf{Adv}_{\mathsf{G}}^{\mathsf{F}}(\mathsf{A}),$$

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, $\mathbf{Adv}_{\mathsf{G}_1, \dots, \mathsf{G}_m}^{\mathsf{F}_1, \dots, \mathsf{F}_m}(\mathsf{A})$ denotes the advantage of A in distinguishing $(\mathsf{F}_1, \dots, \mathsf{F}_m)$ from $(\mathsf{G}_1, \dots, \mathsf{G}_m)$.

[5] When $\mathcal{X} = \{0,1\}^n$ then we write Func to denote $\mathsf{Func}(\{0,1\}^n)$.

2.2 Block Cipher

A block cipher $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \rightarrow \{0,1\}^n$ 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 $k \in \mathcal{K}$, $\mathsf{E}(k,\cdot)$ 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 $\mathsf{E}_k(x)$ to represent $\mathsf{E}(k,x)$. Let $\mathsf{BC}(\mathcal{K},\{0,1\}^n)$ denotes the set of all n-bit block ciphers with key space $\mathcal{K}$. We say that a block cipher E is an (q,ϵ,t)-secure strong pseudorandom permutation, if for all distinguishers A that makes total q forward and inverse queries with run time at most t, the following holds:

$$\mathbf{Adv}_{\Pi}^{\mathsf{E}}(\mathsf{A}) \stackrel{\Delta}{=} |\ \Pr[K \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{\mathsf{E}_K} \Rightarrow 1] - \Pr[\Pi \leftarrow_{\$} \mathsf{Perm} : \mathsf{A}^{\Pi} \Rightarrow 1]\ | \leq \epsilon.$$

2.3 PRF Security in Ideal Cipher Model

A *keyed function* with the key space $\mathcal{K}$, the domain $\mathcal{X}$ and the range $\mathcal{Y}$ is a function $\mathsf{F} : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$. We denote $\mathsf{F}(k,x)$ by $\mathsf{F}_k(x)$. A random function RF from $\mathcal{X}$ to $\mathcal{Y}$ is a uniform random variable over the set $\mathsf{Func}(\mathcal{X},\mathcal{Y})$, i.e., $\mathsf{RF} \leftarrow_{\$} \mathsf{Func}(\mathcal{X},\mathcal{Y})$. 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 $\mathsf{BC}(\mathcal{K},\{0,1\}^n)$). For simplicity, we write $\mathsf{F}_K^{\mathsf{E}}$ to denote F with a uniformly sampled block cipher $\mathsf{E} \leftarrow_{\$} \mathsf{BC}(\mathcal{K},\{0,1\}^n)$ with randomly sampled key $K \leftarrow_{\$} \mathcal{K}$. The distinguisher A is given access to either $(\mathsf{F}_K^{\mathsf{E}}, \mathsf{E}^{\pm})$ for $K \leftarrow_{\$} \mathcal{K}$ or $(\mathsf{RF}, \mathsf{E}^{\pm})$, where $\mathsf{E} \leftarrow_{\$} \mathsf{BC}(\mathcal{K},\{0,1\}^n)$ is a uniformly sampled n-bit block cipher such that A can make forward or inverse query to E, which is denoted as $\mathsf{E}^{\pm}$. We define the prf-advantage of A against a keyed function F in the ideal cipher model as

$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{PRF}}(\mathsf{A}) \stackrel{\Delta}{=} \mathbf{Adv}_{(\mathsf{RF},\mathsf{E}^{\pm})}^{(\mathsf{F}_K^{\mathsf{E}},\mathsf{E}^{\pm})}(\mathsf{A}).$$

We say F is (q,p,ϵ,t)-PRF if $\mathbf{Adv}_{\mathsf{F}}^{\mathrm{PRF}}(\mathsf{A}) \leq \epsilon$ for all adversaries A that makes q queries to F, p forward and inverse offline queries to E and runs for at most time t.

2.4 Multi-User PRF Security in Ideal Cipher Model

In the multi-user setting we assume there are u users, such that the i-th user executes $\mathsf{F}_{K_i}^{\mathsf{E}}$. Moreover, 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 make queries to the oracles in the form of (i,M) to the i-th user and obtains $T \leftarrow \mathsf{F}_{K_i}^{\mathsf{E}}(M)$. We call these queries as **construction queries**. For $i \in [u]$, we assume A makes q_i queries to the i-th oracle. We also assume that A make queries to the underlying block cipher E and its inverse with some chosen keys k^j. We call these queries as **primitive queries**. Let A chooses s many distinct block cipher keys $(k^1,\ldots,k^s)$ and makes p_j many primtive 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 $q = q_1 + \ldots + q_u$ be the total number of construction queries and $p = p_1 + \ldots + p_s$ be the total number of primitive queries to the block cipher E with total running time of A is 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 $(\mathsf{F}^{\mathsf{E}}, \mathsf{E}^{\pm})$ from $(\mathsf{RF}, \mathsf{E}^{\pm})$ in the multi-user seting, where $\mathsf{RF} \leftarrow_{\$} \mathsf{Func}(\mathcal{X},\mathcal{Y})$, is defined as

$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{mu\text{-}PRF}}(\mathsf{A}) \stackrel{\Delta}{=} \left|\Pr\left[\mathsf{A}^{((\mathsf{F}_{K_1}^{\mathsf{E}},\ldots,\mathsf{F}_{K_u}^{\mathsf{E}}),\mathsf{E}^{\pm})} \Rightarrow 1\right] - \Pr\left[\mathsf{A}^{((\mathsf{RF},\ldots,\mathsf{RF}),\mathsf{E}^{\pm})} \Rightarrow 1\right]\right|,$$

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

$$\mathbf{Adv}_{\mathsf{F}}^{\text{mu-PRF}}(u, q, p, \mathsf{t}) \overset{\Delta}{=} \max_{\mathsf{A}} \mathbf{Adv}_{\mathsf{F}}^{\text{mu-PRF}}(\mathsf{A}),$$

where the maximum is taken over all (u, q, p, t)-adversaries A. In this paper, we skip the time parameter of the adversary as we will assume throughout the paper that the adversary is computationally unbounded. This will render us to assume that the adversary is deterministic. When $u = 1$, then it renders to the single user distinguishing advantage.

2.5 Security of Keyed Hash Function

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

$$\Pr[K_h \leftarrow_{\$} \mathcal{K}_h : \mathsf{H}_{K_h}(x) \oplus \mathsf{H}_{K_h}(x') = \Delta] \leq \epsilon_{\text{axu}}.$$

Moreover, H is said to be an ϵ-universal hash function, if for any distinct $x, x' \in \mathcal{X}$,

$$\Pr[K_h \leftarrow_{\$} \mathcal{K}_h : \mathsf{H}_{K_h}(x) = \mathsf{H}_{K_h}(x')] \leq \epsilon_{\text{univ}}.$$

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

$$\Pr[K_h \leftarrow_{\$} \mathcal{K}_h : \mathsf{H}_{K_h}(x) = \Delta] \leq \epsilon_{\text{reg}}.$$

2.6 Mirror Theory

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

$$\lambda(\mathcal{L}) = \lambda_1 \oplus \lambda_2 \oplus \ldots \oplus \lambda_\ell.$$

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

1. $\mathcal{G}$ contains no isolated vertex, i.e., every vertex is incident with atleast one edge.
2. The vertex set $\mathcal{V}$ is partitiond into two disjoint parts denoted by $\mathcal{P}$ and $\mathcal{Q}$, where there is no edges within the vertex set in partite $\mathcal{P}$ or in partite $\mathcal{Q}$. All the edges are between the vertex in $\mathcal{P}$ and the vertex in $\mathcal{Q}$. We call such kind of graphs as *bipartite graphs.*
3. $\mathcal{G}$ contains no cycle.
4. $\lambda(\mathcal{L}) \neq 0^n$ for any path $\mathcal{L}$ in $\mathcal{G}$.

Any bipartite graph $\mathcal{G}$ satisfying the above properties will be called a **good graph**. Note that, a good bipartite graph $\mathcal{G}$ contains no cycle, where every edge connects a vertex in $\mathcal{P}$ to one in $\mathcal{Q}$. Therefore, $\mathcal{G}$ is decomposed into its connected components, all of which are trees; let

$$\mathcal{G} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \ldots \sqcup \mathcal{C}_\alpha \sqcup \mathcal{D}_1 \sqcup \mathcal{D}_2 \sqcup \ldots \sqcup \mathcal{D}_\beta$$

for some $\alpha, \beta \geq 0$, where $\mathcal{C}_i$ denotes a component of size greater than 2, and $\mathcal{D}_i$ denotes a componenent size of 2. We will write $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \ldots \sqcup \mathcal{C}_\alpha$ and $\mathcal{D} = \mathcal{D}_1 \sqcup \mathcal{D}_2 \sqcup \ldots \sqcup \mathcal{D}_\beta$.

Definition 1. Let $\mathcal{E}_{\mathcal{G}}$ be a system of equations corresponding to a good biparite graph $\mathcal{G}$ (as defined above). An injective function $\Phi : \mathcal{P} \sqcup \mathcal{Q} \to \{0,1\}^n$, is said to be an *injective solution* to $\mathcal{E}_{\mathcal{G}}$ if $\Phi(P_i) \oplus \Phi(Q_j) = \lambda_{ij}$ for all $\{P_i, Q_j\} \in \mathcal{E}$.

We remark that if we assign any value to a vertex in P , then the labeled edges determine the values of all the other vertices in the component containing P, where the assignment is unique since $\mathcal{G}$ contains no cycle, and the values in the same part are all distinct as $\lambda(\mathcal{L}) \neq 0^n$ for any path $\mathcal{L}$. However, the number of possible assignments of distinct values to the vertices in $\mathcal{G}$ is $\mathbf{P}(2^n, |\mathcal{P}| + |\mathcal{Q}|)$. One might expect that when such an assignment is chosen uniformly at random, it would satisfy all the equations in $\mathcal{G}$ with probability 2^{-nq}, where q denotes the number of edges (i.e., equations) in $\mathcal{G}$. Indeed, we can prove that the number of solutions is closed to $\mathbf{P}(2^n, |\mathcal{P}| + |\mathcal{Q}|)/2^{nq}$ up to a certain error. Formally, we have the following result:

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

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

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

3 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 $\mathsf{H}^1 : \mathcal{K}_h \times \{0,1\}^* \to \{0,1\}^n$ and $\mathsf{H}^2 : \mathcal{K}_h \times \{0,1\}^* \to \{0,1\}^n$ be two keyed hash functions. Based on H^1 and H^2, we define the Double-block Hash or in short DbH function $\mathsf{H} : \mathcal{K}_h \times \mathcal{K}_h \times \{0,1\}^* \to \{0,1\}^{2n}$ as follows:

$$\mathsf{H}_{(L_1,L_2)}(M) = (\mathsf{H}^1_{L_1}(M), \mathsf{H}^2_{L_2}(M)). \tag{1}$$

We compose this DbH function with a very simple and efficient single-keyed xor function $\mathsf{XOR}_K(x, y) = \mathsf{E}_K(x) \oplus \mathsf{E}_K(y)$, 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:

$$\mathsf{C}_2[\mathsf{H}, \mathsf{E}](M) := \mathsf{XOR}_K(\mathsf{H}^1_{L_1}(M), \mathsf{H}^2_{L_2}(M)).$$

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 have been proven secure upto $2^{2n/3}$ queries in the standard model [13] with single user setting. In [32], all these three constructions have been proven secure upto $2^{2n/3}$ queries in the ideal-cipher model with multi-user setting. We would like to note here that as the xor function is not a PRF over two blocks, we can not apply the tradition *Hash-the-PRP* composition result directly to analyze the security of the two-keyed DbHtS. This says that we need a different type of composition result for the security analysis of the Two-Keyed-DbHtS construction in which we require some higher security properties from its underlying DbH function instead of having only the universal property or regular property.

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

- H^1 is a family of ϵ_{reg}-regular functions and ϵ_{univ}-universal functions.
- H^2 is a family of ϵ_{reg}-regular functions and ϵ_{univ}-universal functions.
- For every $M, M' \in \{0,1\}^*$, $\Pr[L_1 \leftarrow_{\$} \mathcal{K}_h, L_2 \leftarrow_{\$} \mathcal{K}_h : \mathsf{H}^1_{L_1}(M) = \mathsf{H}^2_{L_2}(M')] = 0$.

The first two condition says that the regular advantage and the universal advantage of both the hash functions should be negligible whereas the last condition says that the first hash output for any message cannot collide with the second hash output. Having defined the Two-Keyed-DbHtS construction, we now state and prove its security result. For the sake of brevity of the naming of the construction, we refer Two-Keyed-DbHtS construction as C_2.

Theorem 1. *Let $\mathcal{K}, \mathcal{K}_h$ and $\mathcal{M}$ be three non-empty finite sets. Let $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be an n-bit block cipher. Let $\mathsf{H}^1 : \mathcal{K}_h \times \{0,1\}^* \to \{0,1\}^n$ and $\mathsf{H}^2 : \mathcal{K}_h \times \{0,1\}^* \to \{0,1\}^n$ be two n-bit keyed hash functions such that each one of them is ϵ_{reg}-regular and ϵ_{univ}-universal. Let $\mathsf{H} : \mathcal{K}_h \times \mathcal{K}_h \times \{0,1\}^* \to \{0,1\}^{2n}$ be a* **good** *double-block hash function defined as above in Eqn. (1). Then, for any computationally unbounded distinguisher making a total of q construction queries across all u users and a total of p primitive queries to the block cipher E, can distinguish C_2 from an n-bit uniform random function by*

$$\begin{aligned}\mathbf{Adv}^{\text{mprf}}_{\mathsf{C}_2}(u,q,p,\ell) \quad \leq \quad & \frac{9q^{4/3}}{8\cdot 2^n} + \frac{3q^{8/3}}{2\cdot 2^{2n}} + \frac{q^2}{2^{2n}} + \frac{9q^{7/3}}{8\cdot 2^{2n}} + \frac{8q^4}{3\cdot 2^{3n}} + \frac{q}{2^n} + \frac{2u^2}{2^{k_h+k}} + \frac{2q^2}{2^{n+k}} \\ & + \frac{2qp\epsilon_{\text{reg}}}{2^k} + \frac{q^2\epsilon_{\text{univ}}}{2^n} + \frac{2q^2\epsilon_{\text{reg}}}{2^{k_h}} + 3q^{4/3}\epsilon_{\text{univ}} + \frac{q^2\epsilon^2_{\text{univ}}}{2} + \frac{2qp}{2^{n+k}}.\end{aligned}$$

4 Proof of Theorem 1

We consider an computationally unbounded non-trivial deterministic distinghisher A, that interacts with a pair of oracles in either the real world or in the ideal world, described as follows: in the real world, A is given access to u independent instances of C_2 construction, i.e., access to a tuple of u oracles $(\mathsf{C}_2[(L_1^i, L_2^i, K^i)])_{i\in[u]}$, where each (L_1^i, L_2^i) is independent to (L_1^j, L_2^j), K^i is independent to K^j and $\mathsf{E} \leftarrow_{\$} \mathsf{BC}(\mathcal{K}, \{0,1\}^n)$ be an ideal block cipher. Additionally, A has access to the oracle $\mathsf{E}^{\pm}$, underneath the construction C_2. In the ideal world, A is given access to a (i) tuple of u independent random functions $(\mathsf{RF}_1, \ldots, \mathsf{RF}_u)$, where each RF_i is the random function over $\{0,1\}^*$ to $\{0,1\}^n$ that can be equivalently described as a procedure that returns an n-bit uniform string on input of any arbitrary message, and (ii) and the oracle $\mathsf{E}^{\pm}$, where $\mathsf{E} \leftarrow_{\$} \mathsf{BC}(\mathcal{K}, \{0,1\}^n)$ is an ideal block cipher, sampled independent to the distribution of sequence of u independent random functions. In both the worlds, the first oracle is called the *construction oracle* and the latter one is called the *ideal-cipher oracle*. Using the ideal-cipher oracle, distinguisher A can evaluate any query x under its chosen key J. Query to the construction oracle is called the *construction query* and to that of the ideal-cipher oracle is called the *ideal-cipher query*. Note that, distinguisher A can make either *forward* ideal-cipher queries (i.e., it evaluates E with chosen key and input), or it can make *inverse* ideal-cipher queries (i.e., it evaluates E^{-1} with chosen key and input). The description of the ideal oracle is depicted in Fig. 4.1 and Fig. 4.2.

4.1 Description of the Ideal World

The ideal world is consisting of two phases: (i) online phase and (ii) 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 $\mathsf{Func}(\{0,1\}^*, \{0,1\}^n)$ that maps 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 k-bit key and n-bit input. In the online phase, when the distinguisher makes a-th construction query for the i-th user M_a^i, then it returns $T_a^i \leftarrow f_i(M_a^i)$. Similarly, if the distinguisher makes a forward (resp. inverse) primitive query with chosen block cipher key J and input x, then it returns $\mathsf{E}(J, x)$ (resp. $\mathsf{E}^{-1}(J, x)$). However, if any response of the construction queries happens to be all zero string, i.e., 0^n, then we set the bad flag BadT to 1 and abort the game.

```
Online Phase of $\mathcal{O}_{\text{ideal}}$
$\mathsf{E} \leftarrow\!\!\$\ \mathsf{BC}(\mathcal{K}, \{0,1\}^n)$;
Construction Query:
    On $a$-th query of $i$-th user $M_a^i$, return $T_a^i \leftarrow\!\!\$\ \{0,1\}^n$;
    if $\exists(i,a): T_a^i = \mathbf{0}$ then $\boxed{\textsf{Bad-Tag} \leftarrow 1}$, $\perp$;
Primitive Query:
    On $j$-th forward query with chosen key $J^j$ and input $u_\alpha^j$, return $v_\alpha^j \leftarrow \mathsf{E}_{J^j}(u_\alpha^j)$;
    On $j$-th backward query with chosen key $J^j$ and input $v_\alpha^j$, return $u_\alpha^j \leftarrow \mathsf{E}_{J^j}^{-1}(v_\alpha^j)$;
    $\mathsf{Dom}(\mathsf{E}_{J^j}) \leftarrow \mathsf{Dom}(\mathsf{E}_{J^j}) \cup \{u_\alpha^j\}$,  $\mathsf{Ran}(\mathsf{E}_{J^j}) \leftarrow \mathsf{Ran}(\mathsf{E}_{J^j}) \cup \{v_\alpha^j\}$;
```

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 $\perp$) and returns the remaining values of the transcript in any arbitrary manner. So, if the game aborts for some bad event, then we can surely assume that its previous bad events have not happened.

After the interaction is over, the offline phase begins. In this phase, we sample u pair of dummy hash keys $(L_1^i, L_2^i)_{i\in[u]} \leftarrow\!\!\$\ \mathcal{K}_h \times \mathcal{K}_h$ and a u many dummy block cipher keys $(K^i)_{i\in[u]} \leftarrow\!\!\$\ \mathcal{K}$, where we refer L_1^i (resp. L_2^i) as 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 (Σ_a^i, Θ_a^i) for all queries across u users, where we often refer $\Sigma_a^i \leftarrow \mathsf{H}^1_{L_1^i}(M_a^i)$ as the *left hash output* and $\Theta_a^i \leftarrow \mathsf{H}^2_{L_2^i}(M_a^i)$ 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 output or the right hash output for its a-th query collides with some chosen ideal-cipher key and one of its corresponding input of the forward ideal-cipher query, then we set the bad flag $\mathsf{Bad1}$ to 1 and abort the game.

For the i-th user, if the left hash outputs or the right hash outputs for two of its queries collides and the corresponding responses also collides each others (i.e., $\Sigma_a^i = \Sigma_b^i, T_a^i = T_b^i$), then we consider it to be a bad event. Similarly, for a pair of users i_1 and i_2, if their left hash output or the right hash output collides with each other and the corresponding responses also collides with each other, then we consider it to be a bad event. If at least one of the above bad events happen, then we set $\mathsf{Bad2}$ to 1 and abort the game. We also set another flag $\mathsf{Bad3}$ to 1 and abort the game if for the i-th user, the number of the pair of queries whose either left hash output or right hash output collides with each other, is at least $q_i^{2/3}$, where q_i is the number of queries that i-th user has made.

Finally, we set the flag $\mathsf{Bad4}$ to 1 if at least one of the following events hold: (a) for the i-th user, two left hash outputs collides and their corresponding right hash output also collides, or (b) for the i-th user there exists a tuple of four query indices a, b, c, d such that either (i) $\Sigma_a^i = \Sigma_b^i, \Theta_b^i = \Theta_c^i, \Sigma_c^i = \Sigma_d^i$ holds or (ii) $\Theta_a^i = \Theta_b^i, \Sigma_b^i = \Sigma_c^i, \Theta_c^i = \Theta_d^i$ holds. We

```
Offline Phase of $\mathcal{O}_{\text{ideal}}$
1 :  $(L_1^i, L_2^i)_{i\in[u]} \leftarrow\$\, \mathcal{K}_h \times \mathcal{K}_h;\quad (K^i)_{i\in[u]} \leftarrow\$\, \mathcal{K};$
2 :  if $\exists \mathsf{b} \in \{1,2\}$ and $i_1, i_2 \in [u]$ such that $K^{i_1} = K^{i_2} \wedge L_b^{i_1} = L_b^{i_2};$
3 :  then $\boxed{\mathsf{BadK} \leftarrow 1}$, $\perp;$
4 :  $\forall i \in [u], \forall a \in [q_i] : \quad (\Sigma_a^i, \Theta_a^i) \leftarrow (\mathsf{H}^1_{L_1^i}(M_a^i), \mathsf{H}^2_{L_2^i}(M_a^i));$
5 :  if one of the following holds:
       $(a)\ \exists i \in [u], j \in [s], u[0]_\alpha^j \in \mathsf{Dom}(\mathsf{E}_{J^j}),$ such that $K^i = J^j \ \wedge\ \Sigma_a^i = u[0]_\alpha^j;$
       $(b)\ \exists i \in [u], j \in [s], u[1]_\alpha^j \in \mathsf{Dom}(\mathsf{E}_{J^j}),$ such that $K^i = J^j \ \wedge\ \Theta_a^i = u[1]_\alpha^j;$
6 :  then $\boxed{\mathsf{Bad1} \leftarrow 1}$, $\perp;$
7 :  if one of the following holds:
       $(a)\ \exists i \in [u], a, b \in [q_i],$ such that $\Sigma_a^i = \Sigma_b^i \ \wedge\ T_a^i = T_b^i;$
       $(b)\ \exists i_1, i_2 \in [u], a \in [q_{i_1}], b \in [q_{i_2}],$ such that $K^{i_1} = K^{i_2} \ \wedge\ \Sigma_a^{i_1} = \Sigma_b^{i_2};$
       $(c)\ \exists i \in [u], a, b \in [q_i],$ such that $\Theta_a^{i_1} = \Theta_b^{i_1} \ \wedge\ T_a^{i_1} = T_b^{i_1};$
       $(d)\ \exists i_1, i_2 \in [u], a \in [q_{i_1}], b \in [q_{i_2}],$ such that $K^{i_1} = K^{i_2} \ \wedge\ \Theta_a^{i_1} = \Theta_b^{i_2};$
8 :  then $\boxed{\mathsf{Bad2} \leftarrow 1}$, $\perp;$
9 :  if one of the following holds:
       $(a)\ \exists i \in [u],$ such that $\left|\left\{(a,b) : \Sigma_a^i = \Sigma_b^i\right\}\right| \geq q_i^{2/3};$
       $(b)\ \exists i \in [u],$ such that $\left|\left\{(a,b) : \Theta_a^i = \Theta_b^i\right\}\right| \geq q_i^{2/3};$
10 : then $\boxed{\mathsf{Bad3} \leftarrow 1}$, $\perp;$
11 : if one of the following holds:
       $(a)\ \exists i \in [u], a, b \in [q_i]$ such that $\Sigma_a^i = \Sigma_b^i \ \wedge\ \Theta_a^i = \Theta_b^i;$
       $(b)\ \exists i \in [u], a, b, c, d \in [q_i]$ such that $\Sigma_a^i = \Sigma_b^i \ \wedge\ \Theta_b^i = \Theta_c^i \ \wedge\ \Sigma_c^i = \Sigma_d^i;$
       $(c)\ \exists i \in [u], a, b, c, d \in [q_i]$ such that $\Theta_a^i = \Theta_b^i \ \wedge\ \Sigma_b^i = \Sigma_c^i \ \wedge\ \Theta_c^i = \Theta_d^i;$
12 : then $\boxed{\mathsf{Bad4} \leftarrow 1}$, $\perp;$
13 : go to subroutine 4.3;
```

Figure 4.2: Offline 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 $\perp$) and returns the remaining values of the transcript in any arbitrary manner. So, if the game aborts for some bad event, then we can surely assume that its previous bad events have not happened.

would like to note here that as the DbH function H is **good**, it implies that Σ_a^i will not collide with Θ_b^i and hence this event will not be taken into account.

If the game is not aborted at this stage, then it follows that the none of the bad events have happened. It is to be noted that all the query-response belongs to exactly one of the sets $\mathcal{Q}^=$ or $\mathcal{Q}^{\neq}$ as defined in line-13 and line-14 of Fig. 4.2, where $\mathcal{Q}^=$ be 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 have collided with any ideal-cipher queries and $\mathcal{Q}^{\neq}$ be the set of all queries across all users such that the block cipher key of the i-th user has not collided with any ideal-cipher key. We also define two additional sets: $\mathcal{I}^=$ and $\mathcal{I}^{\neq}$ for $\mathcal{Q}^=$ and $\mathcal{Q}^{\neq}$, where $\mathcal{I}^=$ (resp. $\mathcal{I}^{\neq}$) is the set of all i such that $(i, \star) \in \mathcal{Q}^=$ (resp. $(i, \star) \in \mathcal{Q}^{\neq}$). We partition $\mathcal{I}^=$ into r non-empty equivalent classes $\mathcal{I}_1^=, \mathcal{I}_2^=, \ldots, \mathcal{I}_r^=$ based on the relation that i-th user key K^i collides with J^j if and only if $i \in \mathcal{I}_j^=$. Similarly, we

Offline Phase of $\mathcal{O}_{\text{ideal}}$, Sampling Phase

```
1 :  $\mathcal{Q}^{=} := \{(i,a) \in [u] \times [q_i] : \exists j \in [s], K^i = J^j, \Sigma^i_a \notin \mathsf{Dom}(\mathsf{E}_{J^j}), \Theta^i_a \notin \mathsf{Dom}(\mathsf{E}_{J^j})\};$
2 :  $\mathcal{I}^{=} := \{i \in [u] : (i,\star) \in \mathcal{Q}^{=}\} = \mathcal{I}^{=}_1 \sqcup \mathcal{I}^{=}_2 \sqcup \ldots \sqcup \mathcal{I}^{=}_r;$      // $i \in \mathcal{I}^{=}_j \Leftrightarrow K^i = J^j$
3 :  $\forall j \in [r] : \widetilde{\Sigma^j} = \bigcup_{i \in \mathcal{I}^{=}_j} \{(\Sigma^i_1, \Sigma^i_2, \ldots, \Sigma^i_{q_i})\}, \ \widetilde{\Theta^j} = \bigcup_{i \in \mathcal{I}^{=}_j} \{(\Theta^i_1, \Theta^i_2, \ldots, \Theta^i_{q_i})\};$
4 :  $\forall j \in [r]$ do the following steps:
5 :      $\forall i \in \mathcal{I}^{=}_j$ let $\Sigma^i_a$ be not fresh in $(\Sigma^i_1, \Sigma^i_2, \ldots, \Sigma^i_{q_i});$
6 :      if $\Sigma^i_a \notin \mathsf{Dom}(\mathsf{E}_{J^j})$, then $\Psi(\Sigma^i_a) \leftarrow Z^i_{1,a} \leftarrow\$ \{0,1\}^n \setminus \mathsf{Ran}(\mathsf{E}_{J^j}), \ Z^i_{2,a} \leftarrow Z^i_{1,a} \oplus T^i_a;$
7 :      else $Z^i_{1,a} \leftarrow \Psi(\Sigma^i_a), \ Z^i_{2,a} \leftarrow Z^i_{1,a} \oplus T^i_a;$
8 :      if $Z^i_{2,a} \in \mathsf{Ran}(\mathsf{E}_{J^j})$ then $\boxed{\mathsf{Bad\text{-}Samp} \leftarrow 1}, \ \perp;$
9 :      else $\mathsf{Dom}(\mathsf{E}_{J^j}) \leftarrow \mathsf{Dom}(\mathsf{E}_{J^j}) \cup \{(\Sigma^i_a, \Theta^i_a)\}, \ \mathsf{Ran}(\mathsf{E}_{J^j}) \leftarrow \mathsf{Ran}(\mathsf{E}_{J^j}) \cup \{(Z^i_a, Z^i_a \oplus T^i_a)\};$
10 : $\forall (i,a) \in \mathcal{Q}^{=} : \ \Psi(\Sigma^i_a) \leftarrow Z^i_{1,a}, \ \Psi(\Theta^i_a) \leftarrow Z^i_{2,a};$
11 : $\mathcal{Q}^{\neq} := \{(i,a) \in [u] \times [q_i] : \forall j \in [s], K^i \neq J^j\};$
12 : $\mathcal{I}^{\neq} := \{i \in [u] : (i,\star) \in \mathcal{Q}^{\neq}\} = \mathcal{I}^{\neq}_1 \sqcup \mathcal{I}^{\neq}_2 \sqcup \ldots \sqcup \mathcal{I}^{\neq}_{r'};$      // $i \in \mathcal{I}^{\neq}_j \Leftrightarrow K^i = K^j$
13 : $\forall j \in [r'] : \ f_j :=$ distinct number of elements in the tuple $\widetilde{\Sigma_j} \cup \widetilde{\Theta_j};$
14 : $\forall j \in [r'] : \ (Z^i_{1,a}, Z^i_{2,a})_{i \in \mathcal{I}^{\neq}_j, a \in [q_i]} \leftarrow\$ \mathcal{S}_j := \{(Q^i_a, R^i_a)_{i \in \mathcal{I}^{\neq}_j, a \in [q_i]} \in (\{0,1\}^n)^{(f_j)} : \ Q^i_a \oplus R^i_a = T^i_a\};$
15 : $\forall j \in [r'] :$ do the following steps:
16 :     $\mathsf{Dom}(\mathsf{E}_J) \leftarrow \mathsf{Dom}(\mathsf{E}_J) \cup \{(\Sigma^i_a, \Theta^i_a) : i \in \mathcal{I}^{\neq}_j, a \in [q_i]\};$
17 :     $\mathsf{Ran}(\mathsf{E}_J) \leftarrow \mathsf{Ran}(\mathsf{E}_J) \cup \{(Z^i_{1,a}, Z^i_{2,a}) : i \in \mathcal{I}^{\neq}_j, a \in [q_i]\};$
18 : $\forall (i,a) \in \mathcal{Q}^{\neq} : \Psi(\Sigma^i_a) \leftarrow Z^i_{1,a}, \Psi(\Theta^i_a) \leftarrow Z^i_{2,a};$
19 : return $(\Sigma^i_a, \Theta^i_a, Z^i_{1,a}, Z^i_{2,a})_{(i,a) \in [u] \times [q_i]};$
```

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

partition $\mathcal{I}^{\neq}$ into s equivalent classes based on the equivalent relation $i \sim j$ if and only if $K^i = K^j$. Now, for the j-th equivalent class of $\mathcal{I}^{=}$, we consider the tuple

$$\widetilde{\Sigma}_j := \bigcup_{i \in \mathcal{I}^{=}_j} \{(\Sigma^i_1, \Sigma^i_2, \ldots, \Sigma^i_{q_i})\}, \quad \widetilde{\Theta}_j := \bigcup_{i \in \mathcal{I}^{=}_j} \{(\Theta^i_1, \Theta^i_2, \ldots, \Theta^i_{q_i})\}.$$

Note that due to the event in line number $7.(b)$ (resp. $7.(d)$) of Fig. 4.2, we have $\Sigma^{i_1}_a \neq \Sigma^{i_2}_b$ (resp. $\Theta^{i_1}_a \neq \Theta^{i_2}_b$) for $i_1, i_2 \in \mathcal{I}^{=}_j$ and $a \in [q_{i_1}], b \in [q_{i_2}]$. Now, if Σ^i_a is not fresh in the tuple $(\Sigma^i_1, \Sigma^i_2, \ldots, \Sigma^i_{q_i})$ for some $(i,a) \in \mathcal{I}^{=}_j \times [q_i]$ and the output of Σ^i_a has not been sampled yet, then we sample the its output $Z^i_{1,a}$ from outside the range of E_{J^j} and set the output of Θ^i_a as the xor of Z^i_a and T^i_a (see line-6 of Fig. 4.3). Otherwise, the oracle sets the output of Σ^i_a to the already defined element and adjusts the output of the other hash vaue accordingly (see line-7 of Fig. 4.3). Note that in the latter case, the oracle does not sample the output. In the above said adjustment, if the output of Θ^i_a happens to collide with any previously sampled output, then we set flag $\mathsf{Bad\text{-}Samp}$ to 1 and abort the game (see line-8 of Fig. 4.3) and aborts the game. Note that, this event cannot hold for the real oracle, as Θ^i_a is fresh in $(\Theta^i_1, \Theta^i_2, \ldots, \Theta^i_{q_i})$ for $i \in \mathcal{I}^{=}_j$ and $a \in [q_i]$. If the above flag is not set to 1, then the sampling for the output of Σ^i_a, where $(i,a) \in \mathcal{Q}^{=}$ preserves the permutation compatibility. Moreover, for all other $(i,a) \in \mathcal{Q}^{\neq}$, we sample $Z^i_{1,a}$ and $Z^i_{2,a}$ such that $Z^i_{1,a} \oplus Z^i_{2,a} = T^i_a$.

4.2 Attack Transcript

We summarize 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 $\tau_c = \tau_c^1 \cup \tau_c^2 \cup \ldots \cup \tau_c^u$, where $\tau_c^i = \{(M_1^i, T_1^i), \ldots, (M_{q_i}^i, T_{q_i}^i)\}$ 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 $\tau_p = \tau_p^1 \cup \tau_p^2 \cup \ldots \cup \tau_p^s$, 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 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 the interaction but has not output yet the decision bit. In the real world, we reveal all the keys (L_1^i, L_2^i, K^i) for all u instances, which are used in the construction. In the ideal world, we sample them uniformly at random from their respective key space and reveal them to the distinguisher. Once the keys are revealed to the distinguisher, A can compute $(\Sigma_a^i, \Theta_a^i, \Psi(\Sigma_a^i), \Psi(\Theta_a^i))$, where the function Ψ, defined for the ideal world, is given in Fig. 4.3, whereas for the real world, we define Ψ as follows:

$$\Psi(\Sigma_a^i) = \mathsf{E}_{K^i}(\Sigma_a^i), \quad \Psi(\Theta_a^i) = \mathsf{E}_{K^i}(\Theta_a^i).$$

Therefore, each transcript τ_i^c is now modified to include the corresponding intermediate input-output values for the i-th instance of the construction. Thus,

$$\tau_c^i = \{(M_1^i, T_1^i, \Sigma_1^i, \Theta_1^i, \Psi(\Sigma_1^i), \Psi(\Theta_1^i)), \ldots, (M_{q_i}^i, T_{q_i}^i, \Sigma_{q_i}^i, \Theta_{q_i}^i, \Psi(\Sigma_{q_i}^i), \Psi(\Theta_{q_i}^i))\}.$$

In all the following, the complete construction query transcript is

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

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

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

$$\frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} \overset{\Delta}{=} \frac{\Pr[\mathsf{X}_{\mathrm{re}} = \tau]}{\Pr[\mathsf{X}_{\mathrm{id}} = \tau]} \geq 1 - \epsilon_{\mathrm{ratio}},$$

and there exists $\epsilon_{\mathrm{bad}} \geq 0$ *such that* $\Pr[\mathsf{X}_{\mathrm{id}} \in \mathsf{BadT}] \leq \epsilon_{\mathrm{bad}}$*. Then,*

$$\mathbf{Adv}_{\Pi}^{\mathrm{mprf}}(\mathsf{A}) \leq \epsilon_{\mathrm{ratio}} + \epsilon_{\mathrm{bad}}. \tag{2}$$

Therefore, to prove the security of the construction using the H-coefficient technique, we need to identify the set of bad transcripts and upper bounding their probability in the ideal world. Then we lower bound 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 upper bound the probability of bad transcripts in the ideal world and lower bounding the ratio of 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 $i \in [u]$, $\mathsf{C}_2[\mathsf{E}, (L_1^i, L_2^i), K^i] \mapsto \tau_c^i$ denotes the followings:

1. $\Sigma_a^i = (\mathsf{H}^1_{L_1^i}(M_a^i)), \Theta_a^i = (\mathsf{H}^2_{L_2^i}(M_a^i))$
2. $\mathsf{E}_{K^i}(\Sigma_a^i) = \Psi(\Sigma_a^i), \mathsf{E}_{K^i}(\Theta_a^i) = \Psi(\Theta_a^i)$
3. $\mathsf{E}_{K^i}(\Sigma_a^i) \oplus \mathsf{E}_{K^i}(\Theta_a^i) = T_a^i$

4.3 Bounding the Probability of Bad Transcripts

We call a transcript $\tau = (\tau_c, \tau_p)$ to be **bad** if at least one of the flags, during the generation of the transcript in Fig. 4.1 and Fig. 4.2, is set to 1. Recall that $\mathsf{BadT} \subseteq \Theta$ is the set of all attainable bad transcripts and $\mathsf{GoodT} = \Theta \setminus \mathsf{BadT}$ be 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 Θ_{b} be defined as above. Then*

$$\begin{aligned}\Pr[\mathsf{X}_{\mathrm{id}} \in \mathsf{BadT}] \quad \leq \quad & \frac{q}{2^n} + \frac{2u^2}{2^{k_h+k}} + \frac{2qp\epsilon_{\mathrm{reg}}}{2^k} + \frac{q^2\epsilon_{\mathrm{univ}}}{2^n} + \frac{2q^2\epsilon_{\mathrm{reg}}}{2^{k_h}} + 3q^{4/3}\epsilon_{\mathrm{univ}} \\ & + \frac{q^2\epsilon_{\mathrm{univ}}^2}{2} + \frac{2qp}{2^{n+k}} + \frac{2q^2}{2^{n+k}}.\end{aligned}$$

Proof. Using the union bound, we write

$$\Pr[\mathsf{X}_{\mathrm{id}} \in \mathsf{BadT}] \quad \leq \quad \Pr[\mathsf{Bad\text{-}Tag}] + \Pr[\mathsf{BadK}] + \sum_{i=1}^{4} \Pr[\mathsf{Badi}] + \Pr[\mathsf{Bad\text{-}Samp}]. \tag{3}$$

We individually bound each bad events and then use Eqn. (3) to derive the result. In the subsequent analysis, we assume that $|\mathcal{K}_h| = k_h$ and $|\mathcal{K}| = k$.

4.3.1 Bounding Event BadT

▷ Bounding B.13: For a fixed choice of indices, the probability of the event can be bounded to $1/2^n$ as the outputs of the construction quries are sampled uniformly and independently from other random variables. Therefore, by summing over all possible choices of indices, we have

$$\Pr[\mathsf{BadT}] \leq \frac{q}{2^n}. \tag{4}$$

4.3.2 Bounding Event BadK

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

$$\Pr[\mathsf{BadK}] \leq \frac{2u^2}{2^{k_h+k}}. \tag{5}$$

4.3.3 Bounding Event Bad1

▷ Bounding B.11: For a fixed choice of indices, $\Sigma_a^i = u[0]_\alpha^j$ is bounded by the regular advantage of the hash function $\mathsf{H}^1_{L_1^i}$. As the hash key L_1^i is independent over the block cipher key K^i, we have

$$\begin{aligned}\Pr[\mathsf{B.11}] &\leq \sum_{\substack{i\in[u]\\ a\in[q_i]}} \sum_{\substack{j\in[s]\\ \alpha\in[p_j]}} \Pr[K^i = J^j]\cdot\Pr[\Sigma_a^i = u[0]_\alpha^j] \\ &= \sum_{\substack{i\in[u]\\ a\in[q_i]}} \sum_{\substack{j\in[s]\\ \alpha,\beta\in[p_j]}} \epsilon_{\mathrm{reg}}\cdot\frac{1}{2^k} \overset{(1)}{\leq} \frac{qp\epsilon_{\mathrm{reg}}}{2^k}, \end{aligned} \tag{6}$$

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

▷ Bounding B.12: With an identical argument, one can show that the probability of the event B.12 can be bounded by $\frac{qp\epsilon_{\mathrm{reg}}}{2^k}$, i.e.,

$$\Pr[\mathsf{B.12}] \leq \frac{qp\epsilon_{\mathrm{reg}}}{2^k}. \tag{7}$$

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

$$\Pr[\mathsf{Bad1}] = \Pr[\mathsf{B.11}\vee\mathsf{B.12}] \leq \frac{2qp\epsilon_{\mathrm{reg}}}{2^k}. \tag{8}$$

4.3.4 Bounding Event Bad2

▷ Bounding B.21: For a fixed choice of indices, we analyze the probability of the event:

$$\Sigma_a^i = \Sigma_b^i \wedge T_a^i = T_b^i.$$

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

$$\Pr[\mathsf{B.21}] \leq \sum_{i\in[u],\ a,b\in[q_i]} \Pr[\Sigma_a^i = \Sigma_b^i \wedge T_a^i = T_b^i] \leq \frac{q^2\epsilon_{\mathrm{univ}}}{2^{n+1}}. \tag{9}$$

▷ Bounding B.22: We bound the event conditioned on $\overline{\mathsf{BadK}}$, i.e., even if the block cipher keys for user i_1 and i_2 collides, their corresponding hash keys, i.e., $L_1^{i_1}$ and $L_2^{i_2}$ do not collide. Conditioned on this event, for a fixed choice of indices, we bound $\Sigma_a^{i_1} = \Sigma_b^{i_2}$ using the regular property of the hash function H^1 with the randomness of the hash key $L_1^{i_1}$. Moreover, the first event is independent of the second event and thus the first event can be bounded exactly to 2^{-k_h}. Therefore, we have

$$\Pr[\mathsf{B.22} \mid \overline{\mathsf{BadK}}] \leq \sum_{\substack{i_1,i_2\in[u]\\ a\in[q_{i_1}],b\in[q_{i_2}]}} \epsilon_{\mathrm{reg}}\cdot\frac{1}{2^{k_h}} \leq \frac{q^2\epsilon_{\mathrm{reg}}}{2^{k_h}}. \tag{10}$$

▷ Bounding B.23 and B.24: Bounding B.23 and B.24 is identical to that of event B.21 and B.22 respectively. Hence, we have

$$\Pr[\mathsf{B.23}] \leq \frac{q^2\epsilon_{\mathrm{univ}}}{2^{n+1}}, \qquad \Pr[\mathsf{B.24} \mid \overline{\mathsf{BadK}}] \leq \frac{q^2\epsilon_{\mathrm{reg}}}{2^{k_h}}. \tag{11}$$

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

$$\Pr[\mathsf{Bad2}] \leq \Pr[\mathsf{B.21}\vee\mathsf{B.23}] + \Pr[\mathsf{B.22}\mid\overline{\mathsf{BadK}}] + \Pr[\mathsf{B.24}\mid\overline{\mathsf{BadK}}] \leq \frac{q^2\epsilon_{\mathrm{univ}}}{2^n} + \frac{2q^2\epsilon_{\mathrm{reg}}}{2^{k_h}}. \tag{12}$$

4.3.5 Bounding Event Bad3

$\triangleright$ BOUNDING $\mathsf{B.31}$ and $\mathsf{B.32}$: We first bound the event $\mathsf{B.31}$. For a fixed choice of indices, we define an indicator random variable $\mathbb{I}^i_{a,b}$ which takes the value 1 if $\Sigma^i_a = \Sigma^i_b$, and 0 otherwise. Let $\mathbb{I}^i = \sum_{a,b} \mathbb{I}^i_{a,b}$. Then, by the linearity of expectation, we have

$$\mathbf{E}[\mathbb{I}^i] = \sum_{a,b} \mathbf{E}[\mathbb{I}^i_{a,b}] = \sum_{a,b} \Pr[\Sigma^i_a = \Sigma^i_b] \leq \frac{q_i^2 \epsilon_{\text{univ}}}{2}.$$

Now, we have

$$\begin{aligned}\Pr[\mathsf{B.31}] &\leq \sum_{i \in [u]} \Pr[|\{(a,b) \in [q_i]^2 : \Sigma^i_a = \Sigma^i_b\}| \geq q_i^{2/3}] \\ &= \sum_{i=1}^{u} \Pr[\mathbb{I}^i \geq q_i^{2/3}] \overset{(1)}{=} \sum_{i=1}^{u} \frac{q_i^2 \epsilon_{\text{univ}}}{2q_i^{2/3}} \leq \frac{q^{4/3}\epsilon_{\text{univ}}}{2}, \end{aligned} \tag{13}$$

where (1) holds due to the Markov inequality.
Similar to $\mathsf{B.31}$, we bound $\mathsf{B.32}$ as follows:

$$\Pr[\mathsf{B.32}] \leq \frac{q^{4/3}\epsilon_{\text{univ}}}{2}, \tag{14}$$

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

$$\Pr[\mathsf{Bad3}] = \Pr[\mathsf{B.31} \vee \mathsf{B.32}] \leq q^{4/3}\epsilon_{\text{univ}}. \tag{15}$$

4.3.6 Bounding Event Bad4

$\triangleright$ BOUNDING $\mathsf{B.41}$: Due to the independence of the hash key L^i_1 and L^i_2, for a fixed choice of indices the probability of this joint event can be bounded by the universal property of the individual hash functions H^1 and H^2. Therefore, by varying over all possible choices of indices, we have

$$\begin{aligned}\Pr[\mathsf{B.41}] &\leq \sum_{\substack{i \in [u] \\ a,b \in [q_i]}} \Pr[\Sigma^i_a = \Sigma^i_b \wedge \Theta^i_a = \Theta^i_b] = \sum_{\substack{i \in [u] \\ a,b \in [q_i]}} \Pr[\Sigma^i_a = \Sigma^i_b] \cdot \Pr[\Theta^i_a = \Theta^i_b] \\ &\leq \frac{q^2 \epsilon^2_{\text{univ}}}{2}. \end{aligned} \tag{16}$$

$\triangleright$ BOUNDING $\mathsf{B.42}$ and $\mathsf{B.43}$: We first bound the event $\mathsf{B.42}$. We bound this event conditioned on $\overline{\mathsf{B.31}}$. This results to the fact that for a fixed $i \in [u]$, 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 $q_i^{4/3}$. Now, for a fixed choice of such quadruples, the event that $\Theta^i_b = \Theta^i_c$ holds with probability at most ϵ_{univ} due to the universal property of the hash function H^2. Therefore, we have

$$\Pr[\mathsf{B.42} \mid \overline{\mathsf{B.31}}] \leq \sum_{i \in [u]} q_i^{4/3} \epsilon_{\text{univ}} \leq q^{4/3}\epsilon_{\text{univ}}. \tag{17}$$

Similar to $\mathsf{B.42}$, we bound $\mathsf{B.43}$ as follows:

$$\Pr[\mathsf{B.43} \mid \overline{\mathsf{B.31}}] \leq q^{4/3}\epsilon_{\text{univ}}. \tag{18}$$

By combining Eqn. (16)-Eqn. (18), we have

$$\Pr[\mathsf{Bad4}] \leq \frac{q^2 \epsilon^2_{\text{univ}}}{2} + 2q^{4/3}\epsilon_{\text{univ}}. \tag{19}$$

4.3.7 Bounding Event Bad-Samp

The event holds if there exists an $i \in [u]$ and $j \in [s]$ such that the event (i) $K^i = J^j$ and (ii) $Z_a^i \oplus T_a^i \in \mathsf{Ran}(\mathsf{E}_{J^j})$ holds for some $a \in [q_i]$, where $Z_a^i \leftarrow_{\$} \{0,1\}^n \setminus \mathsf{Ran}(\mathsf{E}_{J^j})$. We first fix an index $j \in [s]$, which determines $\mathcal{I}_j^=$. We fix an index $i \in \mathcal{I}_j^=$ and $a \in [q_i]$. For such a fixed choice of indices, the probability that $Z_{1,a}^i \oplus T_a^i \in \mathsf{Ran}(\mathsf{E}_{J^j})$ holds is at most $2^{-n} \cdot (p_j + q)$ as the cardinality of $\mathsf{Ran}(\mathsf{E}_{J^j})$ is upper bounded by $(p_j + q)$. Now, note that the event (i) is independent over event (ii). Therefore, we bound the probability of the event as

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

Note that the number of choices for (i, a) is at most q and the number of choices for j is s. Thus, by summing over all possible choices of (i, j, a) and by assuming $p_j \leq p$ and $p \leq 2^{n-1}$, we have the result.
Finally, the result follows by combining Eqn. (4)-Eqn. (20). □

4.4 Analysis of Good Transcripts

In this section, we lower bound 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 $i \in \mathcal{I}_j^=$, we consider the sequence

$$\widetilde{\Sigma}^i := (\Sigma_1^i, \Sigma_2^i, \ldots, \Sigma_{q_i}^i), \widetilde{\Theta}^i := (\Theta_1^i, \Theta_2^i, \ldots, \Theta_{q_i}^i).$$

From this sequence, we construct a bipartite graph G_i where the nodes in one partite represents the value Σ_a^i and the nodes in other partite represents Θ_a^i. We put an edge between the node Σ_a^i and Θ_a^i. Note that, if $\Sigma_a^i = \Sigma_b^i$, then we merge the correponding nodes into a single node. Similarly, if $\Theta_a^i = \Theta_b^i$, then we merge the corresponding nodes into a single node. This leads us to break the graph into w_i many components. As the transcript is good, it is easy to see that each component is acyclic and each of them contains a path of length at most 3. Let v_i be the total number of nodes of the graph G_i. Similar to $\mathcal{Q}^=$, we consider $\mathcal{Q}^{\neq}$. For each $j \in [r']$ and for each $i \in \mathcal{I}_j^{\neq}$, we consider the sequence

$$\widetilde{\Sigma}^i := (\Sigma_1^i, \Sigma_2^i, \ldots, \Sigma_{q_i}^i), \widetilde{\Theta}^i := (\Theta_1^i, \Theta_2^i, \ldots, \Theta_{q_i}^i).$$

Similar to the construction of G_i, we construct a bipartite graph H_i, whose one partite represents the nodes corresponding to Σ_a^i and the other partite represents the nodes correspond to Θ_a^i. We put an edge between the nodes corresponding to Σ_a^i and Θ_a^i. If two nodes represents the same values then we merge them together into a single node. Let w_i' be the number of components of H_i and v_i' be its total number of vertices. Now, we state that for a good transcript $\tau = (\tau_c, \tau_p)$, realizing τ is almost as likely in the real world as in the ideal world. More formally,

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

$$\frac{\Pr[\mathsf{X}_{\mathrm{re}} = \tau]}{\Pr[\mathsf{X}_{\mathrm{id}} = \tau]} \geq 1 - \frac{9q^{4/3}}{8 \cdot 2^n} - \frac{3q^{8/3}}{2 \cdot 2^{2n}} - \frac{q^2}{2^{2n}} - \frac{9q^{7/3}}{8 \cdot 2^{2n}} - \frac{8q^4}{3 \cdot 2^{3n}}.$$

Proof. Having set up the background, we are now ready to calculate the real interpolation probability. In order to calculate it, we are basically required to bound count the total number of input-output pairs on which the block cipher E, with different keys, has been executed. As the transcript releases $2k_h$-bit hash keys and the k-bit block cipher key for each user, it contributes to $2^{-(2k_h+k)}$ term 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

$$p_j + \sum_{i \in \mathcal{I}_j^{=}} v_i$$

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 $j \in [r']$, block cipher E is evalued on a total of $\sum_{i \in \mathcal{I}_j^{\neq}} v'_i$ input-output pairs with key K^j. Summarizing above, we have

$$\Pr[\mathsf{X}_{\text{re}} = \tau] = \prod_{i=1}^{u} \frac{1}{2^{2k_h+k}} \cdot \left(\prod_{j=1}^{r} \frac{1}{\mathbf{P}(2^n, p_j + \sum_{i \in \mathcal{I}_j^{=}} v_i)} \right) \cdot \prod_{j \in [s] \setminus [r]} \frac{1}{\mathbf{P}(2^n, p_j)} \cdot \left(\prod_{j=1}^{r'} \frac{1}{\mathbf{P}(2^n, \sum_{i \in \mathcal{I}_j^{\neq}} v'_i)} \right). \tag{21}$$

<u>Ideal Interpolation Probability:</u> First of all, we have $\prod_{i=1}^{u} 2^{-nq_i}$ term which is contributed to the ideal interpolation probability due to the sampling of the response of the adversarial query. Followed by this, we identify that it samples $2k_h$-bit hash keys and k-bit block cipher key for all u users. Now, for each $j \in [r]$, and for each $i \in \mathcal{I}_j^{=}$, we do the following: we construct the graph G_i as defined above, where recall that v_i is the number of vertices of graph G_i across of all w_i components. It is easy to see that for each $j \in [r]$ and for each $i \in \mathcal{I}_j^{=}$, the graph G_i *good*. Having set up the graph G_i, for each $j \in [r]$ and for each $i \in \mathcal{I}_j^{=}$, we sample the value of a node for each of the component of the graph G_i. Hence, for $j \in [r]$, the total number of sampled points is

$$p_j + \sum_{i \in \mathcal{I}_j^{=}} w_i.$$

Moreover, for each $j \in [s] \setminus [r]$, the total number of sample points is p_j. Now, we consider the set of transcripts $\mathcal{Q}^{\neq}$. For each $j \in [r']$, and for each $i \in \mathcal{I}_j^{\neq}$, we do the following: we construct the graph H_i as defined above, where recall that v'_i is the number of vertices of graph H_i across all w'_i components. Having set up the graph H_i for each $i \in \mathcal{I}_j^{\neq}$, we compute the set $\mathcal{S}_j$ for each $j \in [r']$, as defined in line-14 of Fig. 4.3, which is defined as the number of tuples (Q^i_a, R^i_a) such that $Q^i_a \oplus R^i_a = T^i_a$ for all $i \in \mathcal{I}_j^{\neq}$ and for all $a \in [q_i]$. Summarizing above, we have

$$\Pr[\mathsf{X}_{\text{id}} = \tau] = \prod_{i=1}^{u} \frac{1}{2^{nq_i}} \cdot \prod_{i=1}^{u} \frac{1}{2^{2k_h+k}} \cdot \left(\prod_{j=1}^{r} \frac{1}{\mathbf{P}(2^n, p_j + \sum_{i \in \mathcal{I}_j^{=}} w_i)} \right) \cdot \prod_{j \in [s] \setminus [r]} \frac{1}{\mathbf{P}(2^n, p_j)} \cdot \left(\prod_{j=1}^{r'} \frac{1}{|\mathcal{S}_j|} \right). \tag{22}$$

<u>Calculation of the ratio:</u> By plugging-in the value of $|\mathcal{S}_j|$ from Lemma 1 into Eqn. (22)

and then by taking the ratio of Eqn. (21) to Eqn. (22), we have

$$
\begin{aligned}
\mathsf{p}(\tau) &= \prod_{i=1}^{u} 2^{nq_i} \cdot \prod_{j=1}^{r} \frac{\mathbf{P}(2^n, p_j + \sum\limits_{i\in\mathcal{I}_j^{=}} w_i)}{\mathbf{P}(2^n, p_j + \sum\limits_{i\in\mathcal{I}_j^{=}} v_i)} \cdot \prod_{j=1}^{r'} \frac{|\mathcal{S}_j|}{\mathbf{P}(2^n, \sum\limits_{i\in\mathcal{I}_j^{\neq}} v'_i)} \\
&= \prod_{i=1}^{u} 2^{nq_i} \cdot \prod_{j=1}^{r} \frac{1}{\mathbf{P}(2^n - p_j - \sum\limits_{i\in\mathcal{I}_j^{=}} w_i, \sum\limits_{i\in\mathcal{I}_j^{=}} (v_i - w_i))} \cdot \prod_{j=1}^{r'} \frac{\mathbf{P}(2^n, \sum\limits_{i\in\mathcal{I}_j^{\neq}} v'_i) \cdot \Big(1 - \epsilon_j\Big)}{\mathbf{P}(2^n, \sum\limits_{i\in\mathcal{I}_j^{\neq}} v'_i) \cdot 2^{n \sum_{i\in\mathcal{I}_j^{\neq}} (v'_i - w'_i)}} \\
&= \prod_{i=1}^{u} 2^{nq_i} \cdot \prod_{j=1}^{r} \frac{1}{\mathbf{P}(2^n - p_j - \sum\limits_{i\in\mathcal{I}_j^{=}} w_i, \sum\limits_{i\in\mathcal{I}_j^{=}} (v_i - w_i))} \cdot \prod_{j=1}^{r'} \frac{1}{2^{n \sum_{i\in\mathcal{I}_j^{\neq}} (v'_i - w'_i)}} \cdot \prod_{j=1}^{r'} \Big(1 - \epsilon_j\Big) \\
&= \underbrace{\prod_{j=1}^{r} \frac{2^{n \sum_{i\in\mathcal{I}_j^{=}} q_i}}{\mathbf{P}(2^n - p_j - \sum\limits_{i\in\mathcal{I}_j^{=}} w_i, \sum\limits_{i\in\mathcal{I}_j^{=}} (v_i - w_i))}}_{\geq 1} \cdot \underbrace{\prod_{j=1}^{r'} \frac{2^{n \sum_{i\in\mathcal{I}_j^{\neq}} q_i}}{2^{n \sum_{i\in\mathcal{I}_j^{\neq}} (v'_i - w'_i)}}}_{\geq 1} \cdot \prod_{j=1}^{r'} \Big(1 - \epsilon_j\Big) \\
&\geq \Big(1 - \sum_{j=1}^{r'} \epsilon_j\Big) \geq 1 - \sum_{j=1}^{r'} \sum_{i\in\mathcal{I}_j^{\neq}} \left(\frac{9(q^{\mathsf{c}})_i^2}{8 \cdot 2^n} + \frac{3q_i^{\mathsf{c}} q_i^2}{2 \cdot 2^{2n}} + \frac{q_i^2}{2^{2n}} + \frac{9(q^{\mathsf{c}})_i^2 q_i}{8 \cdot 2^{2n}} + \frac{8q_i^4}{3 \cdot 2^{3n}} \right) \\
&\overset{(1)}{\geq} 1 - \sum_{j=1}^{r'} \sum_{i\in\mathcal{I}_j^{\neq}} \left(\frac{9q_i^{4/3}}{8 \cdot 2^n} + \frac{3q_i^{8/3}}{2 \cdot 2^{2n}} + \frac{q_i^2}{2^{2n}} + \frac{9q_i^{7/3}}{8 \cdot 2^{2n}} + \frac{8q_i^4}{3 \cdot 2^{3n}} \right) \\
&\geq 1 - \left(\frac{9q^{4/3}}{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{8q^4}{3 \cdot 2^{3n}} \right),
\end{aligned}
$$

where (1) holds due to the fact that $q_i^{\mathsf{c}} \leq q_i^{2/3}$ for all $i \in \mathcal{I}_j^{\neq}$ such that $j \in [r']$. Note that, for each $j \in [r]$, $\sum_{i\in\mathcal{I}_j^{=}} (v_i - w_i)$ denotes the total number of edges in the graph $\bigcup_{i\in\mathcal{I}_j^{=}} G_i$, which is nothing but $\sum_{i\in\mathcal{I}_j^{=}} q_i$. Similarly, for each $j \in [r']$, $\sum_{i\in\mathcal{I}_j^{\neq}} (v'_i - w'_i)$ denotes the total number of edges in the graph $\bigcup_{i\in\mathcal{I}_j^{\neq}} H_i$, which is nothing but $\sum_{i\in\mathcal{I}_j^{\neq}} q_i$.

5 Instantiation of Two-Keyed DbHtS with PolyHash

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^*, i.e., append a bit 1 followed by minimum number of zeroes to the message M so that the total number of bits in the padded message becomes mutiple 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:

$$\mathsf{PH}(L, M^*) \overset{\Delta}{=} M_1 \cdot L^l \oplus M_2 \cdot L^{l-1} \oplus \ldots \oplus M_l \cdot L,$$

where l is the number of blocks of M. Note that, if the size of the message M is of multiple of n, then we do not apply the padding function and apply the Polyhash function on the message M itself as follows:

$$\mathsf{PH}(L, M) = M_1 \cdot L^l \oplus M_2 \cdot L^{l-1} \oplus \ldots \oplus M_l \cdot L.$$

Note that, in both the equations, the multiplications are defined in the field $\mathrm{GF}(2^n)$. In the following, we state that Polyhash [26] is $\ell/2^n$-regular, $\ell/2^n$-axu and $\ell/2^n$-universal, where ℓ is the maximum number of message blocks. Proof of the lemma is related to the result on the number of distinct roots of a polynomial, given as below:

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

Proof. For regular advantage of the hash function, $\mathsf{PH}(L, M) = \Delta$ is a polynomial of L of constant term Δ and its degree is bounded by ℓ. $\mathsf{PH}(L, M) \oplus \Delta$ is a non-zero polynomial and hence we have, $\epsilon_{\text{reg}} = \ell/2^n$, where ℓ is the maximum number of message blocks among all q messages, as the maximum number of roots for the polynomial $\mathsf{PH}(L, M) \oplus \Delta$ is ℓ. Moreover, for any two distinct messages M and M', $\mathsf{PH}(L, M) \oplus \mathsf{PH}(L, M') \oplus \Delta$ is a non-zero polynomial of L with degree at most ℓ, and hence the maximum number of roots this polynomial can have is ℓ. Therefore, the almost-xor-universal advantage of PH is $\ell/2^n$. □

From Lemma 4, a simple corollary immediately follows:

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

Having defined the Polyhash function, we now define the Polyhash based double block hash function, in short $\mathsf{PH\text{-}DbH}$ function, as follows:

$$\mathsf{PH\text{-}DbH}(L_1, L_2, M) \triangleq \Big(\underbrace{\mathsf{fix}_0(\mathsf{PH}(L_1, M))}_{\mathsf{H}^1_{L_1}}, \underbrace{\mathsf{fix}_1(\mathsf{PH}(L_2, M))}_{\mathsf{H}^2_{L_2}} \Big). \qquad (23)$$

We take two independent instances of Polyhash function, keyed with two independent keys L_1 and L_2, apply them separately on a message M and then chop the least significant bit of their output to prepend a bit 0 and 1 respectively. Having defined the Polyhash based DbH function, it is now straight forward to define the Polyhash based DbHtS construction directly from the $\mathsf{Two\text{-}Keyed\text{-}DbHtS}$ construction as follows: encrypt $\mathsf{fix}_0(\mathsf{PH}(K_1, M))$ and $\mathsf{fix}_1(\mathsf{PH}(K_2, M))$ 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

```
$\mathsf{PH\text{-}DbHtS}(K_1, K_2, K, M)$
    $\Sigma = \mathsf{fix}_0(\mathsf{PH}(K_1, M))$;
    $\Theta = \mathsf{fix}_1(\mathsf{PH}(K_2, M))$;
    $T = \mathsf{E}_K(\Sigma) \oplus \mathsf{E}_K(\Theta)$;
return $T$;
```

```
$\mathsf{PH}(L, M)$
    $M_1 \| \ldots \| M_\ell \xleftarrow{n} M \| 10^*$;
    $Y = M_1 \cdot L^\ell \oplus M_2 \cdot L^{\ell-1} \oplus \cdots \oplus M_\ell \cdot L$;
return $Y$;
```

Figure 5.1: $\mathsf{PH\text{-}DbHtS}$ construction with $\mathsf{PH\text{-}DbH}$ as the underlying double block hash function. $M_1 \| M_2 \| \ldots \| M_\ell \xleftarrow{n} M \| 10^*$ denotes the parsing of message $M \| 10^*$ into n bit strings.

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

$$\Pr[\mathsf{fix}_0(\mathsf{PH}(L_1, M)) = \mathsf{fix}_1(\mathsf{PH}(L, M'))] = 0.$$

Therefore, by combining the Corollary 1 with Theorem 1, we derive the following security result of Polyhash based DbHtS construction. For the sake of brevity, we write Π to denote the PH-DbHtS construction.

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

$$\begin{aligned}\mathbf{Adv}^{\mathrm{mprf}}_{\Pi}(u,q,p,\ell) \quad \leq \quad & \frac{9q^{4/3}}{8 \cdot 2^n} + \frac{3q^{8/3}}{2 \cdot 2^{2n}} + \frac{q^2}{2^{2n}} + \frac{9q^{7/3}}{8 \cdot 2^{2n}} + \frac{8q^4}{3 \cdot 2^{3n}} + \frac{q}{2^n} + \frac{2u^2}{2^{n+k}} + \frac{4qp\ell}{2^{n+k}} \\ & + \frac{4q^2\ell}{2^{2n}} + \frac{4q^2\ell}{2^{n+k}} + \frac{8q^{4/3}\ell}{2^n} + \frac{4q^2\ell^2}{2^{2n}} + \frac{2qp}{2^{n+k}} + \frac{2q^2}{2^{n+k}}.\end{aligned}$$

Remark 1. We would like to mention that the definition of the Polyhash function used in this paper is different from that of used in [16]. Nevertheless, one can establish the $3n/4$-bit multi-user security of the Polyhash based DbHtS construction with the Polyhash function used in [16].

6 Conclusion and Future Problems

In this paper, we have shown that Two-Keyed-DbHtS construction is multi-user secured upto $2^{3n/4}$ queries in the ideal-cipher model. As an instantiation of the result, we have shown that Polyhash based DbHtS construction provides $3n/4$-bit multi-user security in the ideal-cipher model. Combining it with the generic result on the attack complexity on DbHtS construction makes the bound tight. However, we cannot apply the result, proven in this paper, to analyse the security of 2K-SUM-ECBC, 2K-PMAC_Plus and 2K-LightMAC_Plus, as the underlying DbH function of these three constructions are based on block cipher and the current result does not support the security analysis of these constructions in the ideal-cipher model as their underlying DbH function 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 function needs a careful study as the problem is quite a non-trivial to solve.

References

[1] Subhadeep Banik, Sumit Kumar Pandey, Thomas Peyrin, Yu Sasaki, Siang Meng Sim, and Yosuke Todo. GIFT: A small present - towards reaching the limit of lightweight encryption. In Wieland Fischer and Naofumi Homma, editors, *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, volume 10529 of *Lecture Notes in Computer Science*, pages 321–345. Springer, 2017.

[2] Mihir Bellare, Alexandra Boldyreva, and Silvio Micali. Public-key encryption in a multi-user setting: Security proofs and improvements. In Bart Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000, International Conference on the Theory*

and Application of Cryptographic Techniques, Bruges, Belgium, May 14-18, 2000, Proceeding, volume 1807 of *Lecture Notes in Computer Science*, pages 259–274. Springer, 2000.

[3] Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of the cipher block chaining message authentication code. *J. Comput. Syst. Sci.*, 61(3):362–399, 2000.

[4] Mihir Bellare, Ted Krovetz, and Phillip Rogaway. Luby-rackoff backwards: Increasing security by making block ciphers non-invertible. In *EUROCRYPT '98, Proceeding.*, pages 266–280, 1998.

[5] Mihir Bellare and Björn Tackmann. The multi-user security of authenticated encryption: AES-GCM in TLS 1.3. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, volume 9814 of *Lecture Notes in Computer Science*, pages 247–276. Springer, 2016.

[6] Jürgen Bierbrauer, Thomas Johansson, Gregory Kabatianskii, and Ben J. M. Smeets. On families of hash functions via geometric codes and concatenation. In *Advances in Cryptology - CRYPTO '93, 13th Annual International Cryptology Conference, Santa Barbara, California, USA, August 22-26, 1993, Proceedings*, pages 331–342, 1993.

[7] Eli Biham. How to decrypt or even substitute des-encrypted messages in 2^{28} steps. *Inf. Process. Lett.*, 84(3):117–124, 2002.

[8] Alex Biryukov, Sourav Mukhopadhyay, and Palash Sarkar. Improved time-memory trade-offs with multiple data. In Bart Preneel and Stafford E. Tavares, editors, *Selected Areas in Cryptography, 12th International Workshop, SAC 2005, Kingston, ON, Canada, August 11-12, 2005, Revised Selected Papers*, volume 3897 of *Lecture Notes in Computer Science*, pages 110–127. Springer, 2005.

[9] John Black and Phillip Rogaway. A block-cipher mode of operation for parallelizable message authentication. In *EUROCRYPT 2002*, pages 384–397, 2002.

[10] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe. PRESENT: an ultra-lightweight block cipher. In *CHES 2007, Proceedings*, pages 450–466, 2007.

[11] Priyanka Bose, Viet Tung Hoang, and Stefano Tessaro. Revisiting AES-GCM-SIV: multi-user security, faster key derivation, and better bounds. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part I*, pages 468–499, 2018.

[12] Sanjit Chatterjee, Alfred Menezes, and Palash Sarkar. Another look at tightness. In Ali Miri and Serge Vaudenay, editors, *Selected Areas in Cryptography - 18th International Workshop, SAC 2011, Toronto, ON, Canada, August 11-12, 2011, Revised Selected Papers*, volume 7118 of *Lecture Notes in Computer Science*, pages 293–319. Springer, 2011.

[13] Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Goutam Paul. Double-block hash-then-sum: A paradigm for constructing bbb secure prf. *IACR Transactions on Symmetric Cryptology*, 2018(3):36–92, 2018.

[14] Bert den Boer. A simple and key-economical unconditional authentication scheme. *Journal of Computer Security*, 2:65–72, 1993.

[15] Jian Guo, Thomas Peyrin, Axel Poschmann, and Matthew J. B. Robshaw. The LED block cipher. *IACR Cryptology ePrint Archive*, 2012:600, 2012.

[16] Tingting Guo and Peng Wang. A note on the security framework of two-key dbhts macs. Cryptology ePrint Archive, Report 2022/375, 2022.

[17] Viet Tung Hoang and Stefano Tessaro. Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, pages 3–32, 2016.

[18] Viet Tung Hoang and Stefano Tessaro. The multi-user security of double encryption. In *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part II*, pages 381–411, 2017.

[19] Tetsu Iwata and Kaoru Kurosawa. OMAC: one-key CBC MAC. In *Fast Software Encryption, 2003*, pages 129–153, 2003.

[20] Seongkwang Kim, ByeongHak Lee, and Jooyoung Lee. Tight security bounds for double-block hash-then-sum macs. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part I*, volume 12105 of *Lecture Notes in Computer Science*, pages 435–465. Springer, 2020.

[21] Gaëtan Leurent, Mridul Nandi, and Ferdinand Sibleyras. Generic attacks against beyond-birthday-bound macs. In Hovav Shacham and Alexandra Boldyreva, editors, *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, volume 10991 of *Lecture Notes in Computer Science*, pages 306–336. Springer, 2018.

[22] Atul Luykx, Bart Mennink, and Kenneth G. Paterson. Analyzing multi-key security degradation. In Tsuyoshi Takagi and Thomas Peyrin, editors, *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part II*, volume 10625 of *Lecture Notes in Computer Science*, pages 575–605. Springer, 2017.

[23] Atul Luykx, Bart Preneel, Elmar Tischhauser, and Kan Yasuda. A MAC mode for lightweight block ciphers. In *Fast Software Encryption - 23rd International Conference, FSE 2016, Bochum, Germany, March 20-23, 2016, Revised Selected Papers*, pages 43–59, 2016.

[24] David A. McGrew and John Viega. The security and performance of the galois/counter mode (GCM) of operation. In Anne Canteaut and Kapalee Viswanathan, editors, *Progress in Cryptology - INDOCRYPT 2004, 5th International Conference on Cryptology in India, Chennai, India, December 20-22, 2004, Proceedings*, volume 3348 of *Lecture Notes in Computer Science*, pages 343–355. Springer, 2004.

[25] M.Dworkin. Recommendation for block cipher modes of operation: Galois/counter mode (gcm) and gmac, 2007.

[26] Kazuhiko Minematsu and Tetsu Iwata. Building blockcipher from tweakable blockcipher: Extending FSE 2009 proposal. In *Cryptography and Coding - 13th IMA International Conference, IMACC 2011, Oxford, UK, December 12-15, 2011. Proceedings*, pages 391–412, 2011.

[27] Andrew Morgan, Rafael Pass, and Elaine Shi. On the adaptive security of macs and prfs. In Shiho Moriai and Huaxiong Wang, editors, *Advances in Cryptology - ASIACRYPT 2020 - 26th International Conference on the Theory and Application of Cryptology and Information Security, Daejeon, South Korea, December 7-11, 2020, Proceedings, Part I*, volume 12491 of *Lecture Notes in Computer Science*, pages 724–753. Springer, 2020.

[28] Nicky Mouha and Atul Luykx. Multi-key security: The even-mansour construction revisited. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, volume 9215 of *Lecture Notes in Computer Science*, pages 209–223. Springer, 2015.

[29] Yusuke Naito. Blockcipher-based macs: Beyond the birthday bound without message length. In Tsuyoshi Takagi and Thomas Peyrin, editors, *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part III*, volume 10626 of *Lecture Notes in Computer Science*, pages 446–470. Springer, 2017.

[30] Mridul Nandi. Birthday attack on dual ewcdm. Cryptology ePrint Archive, Report 2017/579, 2017. https://eprint.iacr.org/2017/579.

[31] Jacques Patarin. The "coefficients h" technique. In *Selected Areas in Cryptography, 15th International Workshop, SAC 2008, Sackville, New Brunswick, Canada, August 14-15, Revised Selected Papers*, pages 328–345, 2008.

[32] Yaobin Shen, Lei Wang, Dawu Gu, and Jian Weng. Revisiting the security of dbhts macs: Beyond-birthday-bound in the multi-user setting. In Tal Malkin and Chris Peikert, editors, *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part III*, volume 12827 of *Lecture Notes in Computer Science*, pages 309–336. Springer, 2021.

[33] Victor Shoup. A composition theorem for universal one-way hash functions. In Bart Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000, International Conference on the Theory and Application of Cryptographic Techniques, Bruges, Belgium, May 14-18, 2000, Proceeding*, volume 1807 of *Lecture Notes in Computer Science*, pages 445–452. Springer, 2000.

[34] Richard Taylor. An integrity check value algorithm for stream ciphers. In *Advances in Cryptology - CRYPTO '93, 13th Annual International Cryptology Conference, Santa Barbara, California, USA, August 22-26, 1993, Proceedings*, pages 40–48, 1993.

[35] Kan Yasuda. The sum of CBC macs is a secure PRF. In *CT-RSA 2010*, pages 366–381, 2010.

[36] Kan Yasuda. A new variant of PMAC: beyond the birthday bound. In *CRYPTO 2011*, pages 596–609, 2011.

[37] Liting Zhang, Wenling Wu, Han Sui, and Peng Wang. 3kf9: Enhancing 3gpp-mac beyond the birthday bound. In *ASIACRYPT 2012*, pages 296–312, 2012.