Tight Security for Key-Alternating Ciphers with Correlated Sub-Keys[⋆]

Stefano Tessaro and Xihu Zhang

University of Washington, Seattle, USA
{tessaro,xihu}@cs.washington.edu

Abstract. Substantial effort has been devoted to proving optimal bounds for the security of key-alternating ciphers with independent sub-keys in the random permutation model (e.g., Chen and Steinberger, EUROCRYPT '14; Hoang and Tessaro, CRYPTO '16). While common in the study of multi-round constructions, the assumption that sub-keys are truly independent is unrealistic since they are generally highly correlated and generated from shorter keys.
In this paper, we show the existence of non-trivial distributions of limited independence for which a t-round key-alternating cipher achieves optimal security. Our work is a natural continuation of the work of Chen et al. (CRYPTO '14), which considered the case of $t = 2$ when all sub-keys are identical. Here, we show that key-alternating ciphers remain secure for a large class of $(t-1)$-wise and $(t-2)$-wise independent distributions of sub-keys.
Our proofs proceed by generalizations of the so-called sum-capture theorem, which we prove using Fourier-analytic techniques.

Keywords: Provable Security, Key-alternating Ciphers

1 Introduction

Key-alternating ciphers (KACs) alternate the application of fixed, invertible, and key-independent permutations $P_1, \ldots, P_t$ on the n-bit strings with xor-ing $t+1$ n-bit sub-keys $s_0, s_1, \ldots, s_t$, i.e., the output of the KAC on input x and sub-keys $\vec{s} = (s_0, s_1, \ldots, s_t)$ is

$$\mathrm{KAC}_{\vec{s}}(x) = s_t + P_t(s_{t-1} + P_{t-1}(\cdots P_2(s_1 + P_1(s_0 + x))\cdots)) ,$$

where $+$ denotes the bit-wise xor. Several modern block cipher designs are KACs, including Substitution-Permutation Networks (SPNs) like AES [10], PRESENT [3] and LED [14].

Most theoretical analyses of KACs [13,4,22,18,6,9,16] have proven their security as a (strong) pseudorandom permutation in a model where the permutations $P_1, \ldots, P_t$ are randomly and independently chosen and can be queried by the adversary. Moreover, the sub-keys $\vec{s} = (s_0, s_1, \ldots, s_t)$ are also chosen *independently*.[1] These results show that the number of queries q (to the keyed construction, and the permutations) needed to break the construction is roughly $q = N^{t/(t+1)}$ (where $N = 2^n$), which has been shown to be optimal.

This paper: Security with correlated sub-keys. Real sub-keys are *not* independent. They are generated from a shorter key using a specific *key schedule* algorithm. However, scant progress has been made in identifying when such key schedules are secure, and independence assumptions are common even in cryptanalysis. In this paper, we therefore pose the following question:

[⋆] A preliminary version of this paper appears in the proceedings of ASIACRYPT 2021. This is the full version.

[1] In fact, Chen and Steinberger [6] already noted that their result holds in the case where the underlying subkeys are t-wise independent. The tight concrete bound proven by Hoang and Tessaro [16] also extends to the t-wise independent setting.

For which distributions of sub-keys can we still obtain optimal security against $q = N^{t/(t+1)}$ queries?

We note that this question was partially addressed by Dunkelman *et al.* [11] for $t = 1$ and later by Chen *et al.* [5], who proved such bounds for the case where $t = 2$ and the sub-keys satisfy the constraint $s_0 = s_1 = s_2$.[2] Here, we consider the extension of their work beyond three rounds.

We also stress that our goal is not to find practical key schedules that are comparable to those used in actual block cipher designs. Rather, we aim to broaden the understanding of correlated key schedules and to identify when they preserve optimal security. At present, even modest savings in randomness to generate the keys are not known for multi-round KACs.

Reducing key dependence for arbitrary rounds. As our first contribution, we show that for *any* t-round KAC with $t + 1$ subkeys, there are key schedules that merely depend on $t - 1$ independent and uniform keys that achieve $q = \Omega(N^{t/(t+1)})$ security. This generalizes the result for $t = 2$ proved by Chen et al. [5] to multi-round instantiations.

We give a general sufficient condition on key distributions for $\vec{s}$ that achieves optimal security. Specifically, our condition considers distributions where the $t + 1$ sub-keys $\vec{s}$ for the t-round KAC are a linear function of a vector $\vec{k}$ of $t - 1$ "master" keys, denoted as $\vec{s} = A\vec{k}$, in which we view each master key and subkey as an element of the field $\mathbb{F}_{2^n}$. The conditions to be met by the key schedules are:

1. Any $t - 2$ rows of A must form a matrix of rank $t - 2$.
2. For any t rows of A
   - the t rows must form a matrix of rank $t - 1$.
   - there exists a linear combination of the t rows that gives a zero vector and two neighboring rows with non-zero coefficients.

For example, a suitable and natural key schedule that satisfies our condition is where $\vec{s}$ is from the $(t - 1)$-wise independent distribution obtained by evaluating a random polynomial of degree $t - 2$ at $t + 1$ distinct points over $\mathbb{F}_{2^n}$. In fact, while our condition on key schedules is more restrictive than $(t-2)$-wise independence, it still permits simple key schedules for small rounds (e.g, $t = 3$ and $t = 4$) that do not require field multiplication, which may be considered an expensive operation, i.e., for $t = 3$, we show that one can set $\vec{s} = (k_0, k_0, k_1, k_1)$ to have $q = \Omega(N^{3/4})$. For $t = 4$, we set $\vec{s} = (k_0, k_1, k_2, k_0 + k_1, k_1 + k_2)$ to have $q = \Omega(N^{4/5})$.

Less independence for more rounds. Of course, we would like to save even more randomness. We make progress by saving n more bits for a sufficiently large number of rounds. Again, we give a general condition on distributions characterized by linear functions mapping $t - 2$ n-bit keys $\vec{k}$ to $t+1$ keys $\vec{s}$, i.e., $\vec{s} = A\vec{k}$. For any linear mapping A satisfying the property that each $t-2$ rows of A have rank $t-2$, our security proof shows, for $t > 5$, a bound that gives security strictly better than $q = \Omega(N^{(t-1)/t})$; for $t \geqslant 8$, we achieve $q = \Omega(N^{t/(t+1)})$ security. Again, one particular instantiation is obtained by evaluating a random polynomial of degree $t - 3$ at $t + 1$ distinct points over $\mathbb{F}_{2^n}$.

How far can we go? The end question is ultimately whether we can push our results even further. Ideally, it would be possible to use a single-key schedule (as in Chen *et al.*) for an arbitrary number of rounds. However, as we explain below, the classical approach to prove security for limited

[2] Actually, Chen *et al.* [5] also addressed reducing the number of keys and permutations in parallel. They showed that a 2-round KAC is secure against $q = \Omega(N^{2/3})$ queries when instantiated by a single permutation and a single key with a key schedule built over a linear orthomorphism.

independence is via so-called "sum-capture theorems" [2,23]. In this paper, we show that the sum-capture theorem necessary to study the trivial key schedule beyond two rounds is not true. Though this does *not* suggest that the resulting construction is insecure, improving beyond the results of this paper would require substantially new counting techniques (see Section 4.3).

Other related work. Another important aspect of theoretical analyses over KACs is reducing the number of random permutations used in the construction. Recently, Wu *et al.* [24] showed that a three-round KAC instantiated with four uniform and independent sub-keys and a single random permutation is secure against $q = \Omega(N^{3/4})$ adversarial queries. Dutta [12] considered minimizing the tweakable KAC by reducing the number of random permutations and proved the security of $q = \Omega(N^{2/3})$ for the 2-round tweakable KAC by Cogliati *et al.* [7] and 4-round tweakable KAC by Cogliati and Seurin [8].

1.1 Technical Overview

Our paper follows the well-established paradigm of proving the security of key-alternating ciphers based on the *expectation method* by Hoang and Tessaro [16] and generalizations of sum-capture theorems proposed by Chen *et al.* [5].

Chain-based analyses. The core of existing analyses proceeds by identifying a set of *bad* transcripts that contain *chains*: these are transcripts where the adversary has made direct queries to $P_1, P_2, \ldots, P_t$, and/or to the construction, which are linked together by the chosen subkeys. In the ideal world, such bad transcript would likely become inconsistent with the real world, i.e., the probability of obtaining a bad transcript from the real world is zero. Formally, we represent a transcript as $\tau = (\mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_t, \vec{k})$, where $\mathcal{Q}_E$ contains queries to the construction and $\mathcal{Q}_i$s are the queries to individual permutations. Further, $\vec{k}$ are the keys from which the actual sub-keys $\vec{s} = (s_0, s_1, \ldots, s_t)$ are generated. (Since our statements are independent of whether such queries occurred in the forward or backward direction, and of their order, we consider the transcript to be comprised of sets of input-output pairs.) We say that such a τ is *bad* if the sub-keys $(s_0, s_1, \ldots, s_t)$ admit some queries $(u_{t+1}, v_0) \in \mathcal{Q}_E, (u_1, v_1) \in \mathcal{Q}_1, \ldots, (u_t, v_t) \in \mathcal{Q}_t$ to constitute a chain, i.e., if there exists an index i such that for all $j \in \{0, \ldots, t\}$ with $j \neq i$, one has $v_j + u_{j+1} = s_j$, then we say they form the i-th type of chain. If the sub-keys $\vec{s}$ are independent and uniform, then the number of chains is at most $(t+1) \cdot q^{t+1}$ (by a simple union bound over all types of chain); thus, the probability that the transcript is bad is at most $O((t+1)q^{t+1}/N^t)$. (Note that every chain definition involved only t subkeys.)

Handling limited independence. The preceding argument does not hold if $\vec{s}$ is generated, for example, from $(t-1)$-wise independent and uniform n-bit keys, since we can expect (at best) to prove $O((t+1)q^{t+1}/N^{t-1})$. We resolve this by considering a generalized version of the sum-capture quantity that provides a tighter bound over the number of chains; namely, we define

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) := \left| \left\{ (v_0, (u_1, v_1), \ldots, (u_{t-1}, v_{t-1}), u_t) \in V_0 \times \mathcal{Q}_1 \times \ldots \times \mathcal{Q}_{t-1} \times U_t \ : \sum_{i=0}^{t-1} c_i (v_i + u_{i+1}) = 0 \right\} \right| , \quad (1)$$

where $V_0, U_t \subseteq \{0,1\}^n$ and the coefficients $\vec{c} = (c_0, \ldots, c_{t-1})$ are field elements of $\mathbb{F}_{2^n}$. A bound on this quantity can be used to non-trivially bound the number of chains, as long as the coefficients arising are compatible with the underlying method to generate the sub-keys and satisfy certain conditions. This in turn help us characterize of which distributions actually yield the desired optimal security.

Concretely, when the linear coefficients $\vec{c} = (c_0, \ldots, c_{t-1})$ satisfy the condition that there is an index $0 \leqslant \mathsf{idx} < t-1$ such that $c_{\mathsf{idx}} \neq 0$ and $c_{\mathsf{idx}+1} \neq 0$, we prove the tight bound $\mu_{\vec{c}} = \Theta(q^{t+1}/N)$ using Fourier analysis techniques.

Reducing key dependencies further. To obtain our results for constructions with sub-keys generated from $t-2$ independent and uniform keys, we need to upper bound an even more restrictive version of the preceding sum-capture quantity where two linear constraints are imposed, i.e.,

$$\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) := \left| \left\{ (v_0, (u_1, v_1), \ldots, (u_{t-1}, v_{t-1}), u_t) \in V_0 \times \mathcal{Q}_1 \times \ldots \times \mathcal{Q}_{t-1} \times U_t \ : \ \sum_{i=0}^{t-1} c_i (v_i + u_{i+1}) = 0 \, , \sum_{i=0}^{t-1} d_i (v_i + u_{i+1}) = 0 \right\} \right| . \quad (2)$$

For the 2-constraint case, we specifically examine the coefficients $\vec{c} = (c_0, \ldots, c_{t-1})$ and $\vec{d} = (d_0, \ldots, d_{t-1})$ that characterize the underlying sub-keys generated when the linear key schedule is $(t-2)$-wise independent and uniform. In this case, we show that

- For $t > 5$, the t-round KAC is secure against $q = \omega(N^{\frac{t-1}{t}})$ queries.
- For $t \geqslant 8$, the t-round KAC has tight security bound (i.e., $q = \Omega(N^{\frac{t}{t+1}})$).

Given that (2) is a natural generalization of its one-constraint counterpart, it is tempting to conclude that upper-bounding (2) is no harder than upper-bounding (1). However, the two constraints make the problem of upper-bounding (2) much harder. Moreover, the tightness of upper-bounding (1) crucially relies on a particular step, referred to as the "Cauchy-Schwartz trick" [2,23,5], which does not seem to apply here. We bypass this limitation by introducing a novel representation for the upper bound of (2) as the 2-norm of a matrix. In particular, one can interpret the Cauchy-Schwartz trick upper bound as a special case of the matrix norm bound where each matrix row and column contains at most one non-zero entry. Then, we use the matrix Frobenius norm, which is easier to compute for bounding the matrix 2-norm. Though our current technique proves tight security bound only for $t \geqslant 8$, we believe that the matrix 2-norm is the right characterization and one can extend the tightness result to $t \geqslant 4$ via a better tool to derive the 2-norm bound, since the usage of Frobenius norm is, in most cases, not tight[3].

While (2) remains a promising candidate for saving two keys, we show that for $t = 3$, i.e., for the 3-round KAC with identical subkey and independent permutations, the quantity of (2) is lower bounded by q^3/N with good probability. Hence, we need a sum-capture quantity with highly non-trivial characterizations or an alternative proof strategy for the 3-round KAC to obtain the desired $q = \Omega(N^{3/4})$ security bound.

[3] In fact, the Frobenius norm and 2-norm can have up to a $\sqrt{N}$ multiplicative gap for $N \times N$ matrix (e.g., the identity matrix), and we believe that a large gap exists in our Frobenius norm bound. However, improving the 2-norm bound requires a more in-depth understanding of our defined matrix for analyzing (2) than we currently have.

Good transcript analysis. Since we have bounded the probability of a transcript being bad, we move to understand the remaining transcripts that we consider to be good. We rely on the expectation method proposed by Hoang and Tessaro [16], which is a generalization of the H-coefficient method [6,21]. In the expectation method, the final security upper bound is

$$\mathsf{Security\ bound} \leqslant \mathbb{E}_{X_1}[g(X_1)] + \mathsf{Pr}[X_1 \text{ is bad}] ,$$

where X_1 is the random variable representing the transcript generated from the adversary interacting with the ideal world and $g : \mathcal{T} \to [0, +\infty)$ is a non-negative function such that $g(\tau)$ upper bounds the real-world ideal-world probability ratio of any good transcript τ. The goal is to find a function $g : \mathcal{T} \to [0, +\infty)$ that minimizes the value of $\mathbb{E}_{X_1}[g(X_1)]$.

It is tempting to believe that the sub-keys must be at least t-wise independent and uniform when applying the techniques in [16] to achieve the tight security bound for good transcripts. However, surprisingly, we show (in Section 5) that as long as the underlying sub-keys $\vec{s} = (s_0, \ldots, s_t)$ are $(t-2)$-wise independent and uniform, we can pick a non-negative function g so that

$$\mathbb{E}_{X_1}[g(X_1)] \leqslant O(q^{t+1}/N^t) .$$

Therefore, as long as the t-round KAC has a key schedule that gives $(t-2)$-wise independent and uniform subkeys, the good transcript analysis gives the optimal bound.

1.2 Paper Organization

In Section 2 we define basic notations and indistinguishability framework. We give the main theorems and show tight security for classes of t-round KAC in Section 3, and we analyze the sum capture quantity for upper-bounding the number of bad transcripts in Section 4. We then provide the analysis of good transcripts in Section 5 and conclude our theorem proofs in Section 6. Finally we provide conclusions and open problems in Section 7.

2 Preliminaries

Notations. For a finite set S, we write $x \xleftarrow{\$} S$ to denote that x receives a uniformly sampled value from S. For an algorithm A, we write $y \leftarrow A(x_1, \ldots; r)$ to denote that A takes $x_1, \ldots$ as inputs, runs with the randomness r and assigns the output to y. We let $y \xleftarrow{\$} A(x_1, \ldots,)$ be that A, given the inputs, is executed over a randomly chosen r, and the resulting value is assigned to y.

We use $\mathbb{F}_p$ to denote a finite field of size p. For any two elements $u, v \in \{0,1\}^n$, we use $\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$ to denote the inner product of u and v, where u_i, v_i are the i-th bit of u, v, respectively. For any number $1 \leqslant b \leqslant a$, we write $a^{(b)} = a(a-1)\cdots(a-b+1)$ and take $a^{(0)} = 1$ by convention. In all the following, for any two elements $u, v \in \{0,1\}^n$, we take $u + v$ and uv as the field addition and multiplication in $\mathbb{F}_{2^n}$, respectively, where $u + v$ is implemented as the bit-wise xor over $\{0,1\}^n$. For a fixed n, we write $N = 2^n$. For any vector u and matrix A, we write $u^\top$ and $A^\top$ as their transpose.

PRP security of block ciphers. We study the security of the KAC in the random permutation model. Let $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ be a blockcipher that is constructed over a set of independent, uniformly random permutations $\vec{P} = (P_1, P_2, \ldots, P_t)$. Let $\mathcal{A}$ be an adversary. the strong PRP advantage of $\mathcal{A}$ is defined as

$$\mathsf{Adv}_E^{\pm\mathsf{prp}}(\mathcal{A}) := \mathsf{Pr}[K \xleftarrow{\$} \mathcal{K} : \ \mathcal{A}^{E_K, \vec{P}} = 1] - \mathsf{Pr}[P_0 \xleftarrow{\$} \mathsf{Perm}(n) : \ \mathcal{A}^{P_0, \vec{P}} = 1] ,$$

where P_0 is a random permutation independent of $\vec{P}$ and "±" denotes that the adversary $\mathcal{A}$ can query the oracles in both a forward and backward direction.

Indistinguishability framework. We consider a computationally unbounded distinguisher $\mathcal{A}$ interacting with two systems, $\mathbf{S}_0$ and $\mathbf{S}_1$. The interaction between $\mathcal{A}$ and $\mathbf{S}_b$ (where $b \in \{0,1\}$) defines a transcript $\tau = ((u_1, v_1), \ldots, (u_q, v_q))$ that records the q pairs of queries/replies $\mathcal{A}$ made to/received from the system $\mathbf{S}_b$. Let X_b be the random variable representing the transcript. Then the goal is to upper bound the following statistical distance

$$\varDelta(X_0, X_1) = \sum_{\tau} \max\{0, \mathsf{Pr}[X_1 = \tau] - \mathsf{Pr}[X_0 = \tau]\} \ .$$

Formulating systems. We follow [19] to describe the system behavior of $\mathbf{S}$ by associating every possible transcript $\tau = ((u_1, v_1), \ldots, (u_q, v_q))$ with a value $\mathsf{p}_{\mathbf{S}}(\tau) \in [0,1]$. One can interpret $\mathsf{p}_{\mathbf{S}}(\tau)$ as the probability that, if the queries $u_1, \ldots, u_q$ in τ are asked sequentially, $\mathbf{S}$ will answer with $v_1, \ldots, v_q$, respectively. Note that $\mathsf{p}_{\mathbf{S}}(\cdot)$ is defined only by the underlying system $\mathbf{S}$ and is hence independent of any distinguisher. We also note that $\mathsf{p}_{\mathbf{S}}(\cdot)$ is not a probability distribution over the transcripts because the sum over all $\mathsf{p}_{\mathbf{S}}(\tau)$ does not necessarily give one.

Since the distinguisher is computationally unbounded, it is sufficient to consider deterministic distinguishers only. Fix any deterministic distinguisher $\mathcal{A}$, let X denote the transcript distribution of $\mathcal{A}$ interacting with $\mathbf{S}$. It then holds that $\mathsf{Pr}[X = \tau] \in \{0, \mathsf{p}_{\mathbf{S}}(\tau)\}$ for any τ because either $\mathcal{A}$ issues the queries $u_1, \ldots, u_q$ when given the answers $v_1, \ldots, v_q$, leading to $\mathsf{Pr}[X = \tau] = \mathsf{p}_{\mathbf{S}}(\tau)$, or it does not, resulting in $\mathsf{Pr}[X = \tau] = 0$.

Let $\mathcal{T}$ be the set of transcripts τ that has $\mathsf{Pr}[X_1 = \tau] > 0$. Further, noting that $\mathsf{Pr}[X_0 = \tau] = \mathsf{p}_{\mathbf{S}_0}(\tau)$ if $\tau \in \mathcal{T}$, we can rewrite the statistical distance as

$$\varDelta(X_0, X_1) = \sum_{\tau} \max\{0, \mathsf{p}_{\mathbf{S}_1}(\tau) - \mathsf{p}_{\mathbf{S}_0}(\tau)\} = \sum_{\tau} \mathsf{p}_{S_1}(\tau) \cdot \max\left\{0, 1 - \frac{\mathsf{p}_{\mathbf{S}_0}(\tau)}{\mathsf{p}_{\mathbf{S}_1}(\tau)}\right\} \ .$$

The expectation method. We now review the expectation method proposed by [16], which was developed based on the H-coefficient method [6,21]. In the H-coefficient method, the set of transcripts $\mathcal{T}$ is partitioned into $\mathcal{T}_{\mathsf{good}}$ and $\mathcal{T}_{\mathsf{bad}}$ so that for any $\tau \in \mathcal{T}_{\mathsf{good}}$, $\mathsf{p}_{\mathbf{S}_0}(\tau)/\mathsf{p}_{\mathbf{S}_1}(\tau) \geqslant 1 - \varepsilon$ for some carefully chosen parameter ε. Then, an upper bound of the advantage directly follows, i.e.,

$$\varDelta(X_0, X_1) \leqslant \varepsilon + \mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] \ .$$

However, instead of giving a uniform bound over all good transcripts, we can associate each τ with a non-negative value $g(\tau)$ so that $\mathsf{p}_{\mathbf{S}_0}(\tau)/\mathsf{p}_{\mathbf{S}_1}(\tau) \geqslant 1 - g(\tau)$ for every $\tau \in \mathcal{T}_{\mathsf{good}}$. Hence, we can instead derive the upper bound as

$$\varDelta(X_0, X_1) \leqslant \sum_{\tau \in \mathcal{T}_{\mathsf{good}}} \mathsf{p}_{\mathbf{S}_1}(\tau) \cdot g(\tau) + \sum_{\tau \in \mathcal{T}_{\mathsf{bad}}} \mathsf{p}_{\mathbf{S}_1}(\tau) \leqslant \mathbb{E}_{X_1}[g(X_1)] + \sum_{\tau \in \mathcal{T}_{\mathsf{bad}}} \mathsf{p}_{\mathbf{S}_1}(\tau) \ ,$$

where we take the expectation over all $\tau \in \mathcal{T}$ by the fact that $g(\cdot)$ is non-negative. Therefore, we have the following lemma.

Lemma 1 (The expectation method). *If there exists a partition of* $\mathcal{T} = \mathcal{T}_{\mathsf{good}} \sqcup \mathcal{T}_{\mathsf{bad}}$ *and a function* $g : \mathcal{T} \to [0, +\infty)$ *such that for any* $\tau \in \mathcal{T}_{\mathsf{good}}$, $\mathsf{p}_{\mathbf{S}_0}(\tau)/\mathsf{p}_{\mathbf{S}_1}(\tau) \geqslant 1 - g(\tau)$, *then*

$$\varDelta(X_0, X_1) \leqslant \mathbb{E}_{X_1}[g(X_1)] + \mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] \ .$$

3 Main Results

We consider the PRP security of a t-round KAC that is built on t random permutations $\vec{P} = (P_1, \ldots, P_t)$ over $\{0,1\}^n$ and $t+1$ subkeys $(s_0, \ldots, s_t)$ where $s_i \in \{0,1\}^n$. When given input $M \in \{0,1\}^n$, the t-round KAC outputs

$$s_t + P_t(s_{t-1} + P_{t-1}(\cdots P_1(s_0 + M)\cdots)) \; .$$

The sub-keys are generated from the master key denoted as $(k_0, \ldots, k_w)$, where k_i are sampled from $\{0,1\}^n$ uniformly and independently. Therefore, the length of the master key is $(w+1)n$ bits. Here, we consider only linear key schedule algorithms, which can be represented as a matrix A over $\mathbb{F}_{2^n}$. We define the column vectors $\vec{s} = (s_0, \ldots, s_t)^\top$ and $\vec{k} = (k_0, \ldots, k_w)^\top$, where we naturally take each n-bit string as an element in $\mathbb{F}_{2^n}$ and use $\vec{s} \leftarrow A\vec{k}$ to denote the key-scheduling process.

The case of A being an identity matrix of size $(t+1) \times (t+1)$ has been well studied, i.e., it was proved in [6,16] that, when the subkeys $s_0, \ldots, s_t$ are independent and uniform and the permutations $P_1, \ldots, P_t$ are independent, any adversary needs at least $q = \Omega(N^{t/(t+1)})$ queries to achieve a constant distinguishing advantage. We now consider the case where the permutations are independent but the sub-keys are correlated and generated via linear key schedules from $t-1$ independent n-bit keys (considered Theorem 1) or $t-2$ independent n-bit keys (Theorem 2).

We starts by providing a security bound of a t-round KAC for a class of key schedules that generate $t+1$ sub-keys from $t-1$ independent keys.

Theorem 1. *For the t-round* KAC *constructed over t random permutations $\vec{P} = (P_1, \ldots, P_t)$, let the key of KAC be $\vec{k} = (k_0, k_1, \ldots, k_{t-2})^\top$, where k_is are independently uniformly sampled from $\mathbb{F}_{2^n}$. Let sub-keys $\vec{s} = (s_0, s_1, \ldots, s_t)^\top$ be derived by $\vec{s} \leftarrow A\vec{k}$, where A is a $(t+1) \times (t-1)$ matrix over $\mathbb{F}_{2^n}$, with the rows denoted as $A_0, \ldots, A_t$, such that*

1. *Any $t-2$ rows of A forms a matrix of rank $t-2$.*
2. *For any $I \subseteq \{0, \ldots, t\}$, $|I| = t$, then the row vectors $(A_\ell)_{\ell \in I}$ satisfy that*
   - *$(A_\ell)_{\ell \in I}$ forms a matrix of rank $t-1$.*
   - *there exist values $(c_\ell)_{\ell \in I}$ such that $\sum_{\ell \in I} c_\ell A_\ell = \vec{0}$, and there are two indices $\mathsf{idx}_1, \mathsf{idx}_2 \in I$ satisfying $\mathsf{idx}_1 - \mathsf{idx}_2 \in \{1, t\}$ and $c_{\mathsf{idx}_1}, c_{\mathsf{idx}_2}$ that are both non-zero.*

Then, for any adversary $\mathcal{A}$ that issues at most q queries to KAC*, $P_1, \ldots, P_t$, where $9(t+2)n \leqslant q \leqslant N/4$,*

$$\mathsf{Adv}^{\pm\mathsf{prp}}_{\mathsf{KAC}}(\mathcal{A}) \leqslant (t^2 + t + 1) \cdot \frac{4q^{t+1}}{N^t} + 3(t+1)\sqrt{\frac{q^{2t-1}(t+2)n}{N^{2t-2}}} \; .$$

First, we give a key schedule that gives $(t-1)$-wise independent and uniform sub-keys for t-round KACs with any round t.

Corollary 1. *For $t < 2^n$, pick distinct elements $\alpha_0, \ldots, \alpha_t \in \mathbb{F}_{2^n}$ and let subkey $s_i = F(\alpha_i)$, where $F(X) = \sum_{j=0}^{t-2} k_j \cdot X^j$. Then, an adversary needs $\Omega(N^{t/(t+1)})$ queries to achieve a constant distinguishing advantage.*

Corollary 1 directly follows from the fact that A is a Vandermonde matrix, so every $t-1$ rows of A forms a full-rank sub-matrix. Hence, any t rows of A are linear dependent, with the coefficients $(c_\ell)_{\ell \in I}$ satisfying $c_\ell \neq 0$ for all ℓ.

Table 1. $q = \Omega(N^{\lambda})$ for a constant security bound in Theorem 2.

t	3	4	5	6	7	8	9	10	$\cdots$
$\lambda = \log_N q$	0.571	0.720	0.800	0.842	0.870	0.889	0.9	0.909	$\cdots$
$t/(t+1)$	0.750	0.800	0.833	0.857	0.875	0.889	0.9	0.909	$\cdots$

Note that by letting $t = 2$ in Corollary 1, our result implies the optimal security bound of a 2-round KAC with identical subkeys and independent permutations proven by Chen *et al.* [5].

Though it is implied in the theorem statement that we need the subkeys to be $(t-2)$-wise independent and uniform, for a small round t we can still obtain some simple key schedules that achieve the optimal bound for q but do not require any field multiplication operations, which may be considered expensive in key-scheduling.

Corollary 2. *Let there be a 3-round* KAC *with key schedule*

$$\vec{s} = (k_0, k_0, k_1, k_1)$$

where k_0, k_1 are two independently uniform n-bit keys; an adversary then needs $\Omega(N^{3/4})$ queries to achieve a constant distinguishing advantage.

Corollary 3. *Let there be a 4-round* KAC *with key schedule*

$$\vec{s} = (k_0, k_1, k_2, k_0 + k_1, k_1 + k_2)$$

where k_0, k_1, k_2 are three independently uniform n-bit keys; An adversary then needs $\Omega(N^{4/5})$ queries to achieve a constant distinguishing advantage.

One can check that the sub-keys in Corollary 2 (or Corollary 3) are 1-wise (or pairwise) independent and uniform and any t rows form a sub-matrix of rank $t-1$ with the coefficients $(c_\ell)_{\ell \in I}$ satisfying the given conditions via Gaussian elimination.

As Theorem 1 gives a tight bound for all t, one may optimistically expect that similar results can be proven with ease when saving one more key. However, for the t-round KAC with sub-keys generated from $t-2$ keys, we can make only partial progress by proving the following theorem, which implies tight security only for $t \geqslant 8$.

Theorem 2. *For the t-round* KAC *constructed over t random permutations $\vec{P} = (P_1, \ldots, P_t)$, let the key of KAC be $\vec{k} = (k_0, k_1, \ldots, k_{t-3})^\top$, where k_is are independently and uniformly sampled from $\mathbb{F}_{2^n}$. Let sub-keys $\vec{s} = (s_0, s_1, \ldots, s_t)^\top$ be derived by $\vec{s} \leftarrow A\vec{k}$, where A is a $(t+1) \times (t-2)$ matrix over $\mathbb{F}_{2^n}$ such that any $t-2$ rows of A form a matrix of rank $t-2$. Then, for any adversary $\mathcal{A}$ that issues at most $(t+2)nN^{2/3} \leqslant q \leqslant N/4$ queries to* $\mathsf{KAC}, P_1, \ldots, P_t$,

$$\mathsf{Adv}^{\pm\mathsf{prp}}_{\mathsf{KAC}}(\mathcal{A}) \leqslant (t^2 + 2t) \cdot \frac{(5q)^{t+1}}{N^t} + (t+1)^2 \cdot \frac{(3q)^{2t-2.5}}{N^{2t-4}} \,.$$

Table 1 summarizes the order of q that causes the security bound to be $\Omega(1)$. We observe that Theorem 2 does not initially give good bound for $t \leqslant 7$. From $t \geqslant 5$, the bound starts improving relative to $q = \Omega(N^{(t-1)/t})$, which is obtained by instantiating a $(t-1)$-round KAC from the provided $t-2$ keys and applying Theorem 1. When $t \geqslant 8$, the bound achieves the optimal $q = \Omega(N^{t/(t+1)})$. The tightness results for $t \leqslant 7$ are left open.

A feasible instantiation of Theorem 2 is to let the sub-keys be the evaluations at $t+1$ distinct points of a degree $t-3$ polynomial. Then, the following corollary holds.

Corollary 4. *For $8 \leqslant t < 2^n$, pick distinct elements $\alpha_0, \ldots, \alpha_t \in \mathbb{F}_{2^n}$ and let subkey $s_i = F(\alpha_i)$, where $F(X) = \sum_{j=0}^{t-3} k_j \cdot X^j$. Then, an adversary needs $\Omega(N^{t/(t+1)})$ queries to achieve a constant distinguishing advantage.*

Proof framework. We use the expectation method (i.e., Lemma 1) to prove both theorems. Given the query record $\vec{\mathcal{Q}} = (\mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_t)$, we generously allow the adversary $\mathcal{A}$ to see the key $\vec{k}$ after making all queries. Therefore, we let the transcript $\tau = (\vec{\mathcal{Q}}, \vec{k})$ by attaching $\vec{k}$ to the end of $\vec{\mathcal{Q}}$. In the ideal world, we sample and attach a dummy key $\vec{k}$ to $\vec{\mathcal{Q}}$. Here, we define the set of bad transcripts for a t-round KAC.

Definition 1 (Bad transcripts). *For a t-round* KAC*, we say a transcript $\tau = (\vec{\mathcal{Q}}, \vec{k})$ is bad if*

$$\vec{k} \in \mathsf{Badkey}_{\vec{\mathcal{Q}}} = \bigcup_{i=0}^{t} \mathsf{Badkey}_{\vec{\mathcal{Q}},i} \,,$$

where for every i,

$$\begin{aligned}\mathsf{Badkey}_{\vec{\mathcal{Q}},i} := \{\vec{k} :\ & \vec{s} \leftarrow \mathsf{KeySchedule}(\vec{k}), \textit{there exists } (u_{t+1}, v_0) \in \mathcal{Q}_E, (u_1, v_1) \in \mathcal{Q}_1, \ldots, (u_t, v_t) \in \mathcal{Q}_t \\ & \textit{s.t. for all } 0 \leqslant j \leqslant t,\ j \neq i,\ v_j + s_j = u_{j+1}\} \,.\end{aligned}$$

Otherwise, we say that τ is good. We use $\mathcal{T}_{\mathsf{good}}$ to denote the set of all good transcripts and $\mathcal{T}_{\mathsf{bad}}$ to denote the set of all bad transcripts. Hence, $\mathcal{T} = \mathcal{T}_{\mathsf{good}} \sqcup \mathcal{T}_{\mathsf{bad}}$.

Then, we break the analysis into the bad and the good transcript cases. We use the generalized sum-capture quantity in Section 4 as an upper bound for the bad transcripts. We analyze the good transcripts in Section 5. Section 6 presents the final proof of theorems.

More fine-grained security. In the preceding theorems, we use q to be the uniform upper bound over all query types. However, we note that our proof technique also provides bounds when the number of cipher queries q_e and the number of permutation queries q_p are separated. We provide the bounds in Appendix A for both theorems but omit the proofs, which are essentially the same as proving the case for $q_e = q_p$ but involve more case discussions.

4 Generalized Sum Capture Quantity for KAC

In [5] Chen *et al.* considered minimizing the 2-round KAC, where they proved a variant of "sum-capture" results [2,15,1,17,23]. The results are often stated that, for a randomly chosen set A of size q, the quantity

$$\mu(A) := \max_{\substack{X,Y \subseteq \mathbb{Z}_2^n \\ |X|=|Y|=q}} |\{(a, x, y) \in A \times X \times Y : a = x + y\}| \tag{3}$$

is close to its expected value q^3/N (when A, X, Y are all chosen at random) with high probability. In the 2-round KAC with an identical key schedule, the sum-capture quantity is defined as

$$\mu(\mathcal{Q}) := \max_{\substack{X,Y \subseteq \mathbb{Z}_2^n \\ |X|=|Y|=q}} |\{(x, (u, v), y) \in X \times \mathcal{Q} \times Y : x + u = v + y\}| \,, \tag{4}$$

where one can view the query transcript $\mathcal{Q}$ derived from the interaction of an adversary $\mathcal{A}$ with the permutation to be equivalent to the set A in (3) defined by $A = \{u + v \mid (u, v) \in \mathcal{Q}\}$.

However, both (3) and (4) consider only a single random permutation with a single linear constraint. To generalize the sum-capture quantity so that we can handle the KAC that saves more keys, we consider the sum-capture quantity that involves $(t-1)$ independently random permutations and $r \in \{1, 2\}$ linear constraints over $\mathbb{F}_{2^n}$ for the t-round KAC with a linear key schedule.

For the $r = 1$ case, we prove the tight bounds of a sum-capture quantity for any choice of linear constraint, leading to a feasible set of key schedules that saves two keys for an arbitrary t-round KAC with tight security. However, as we increase the number of constraints to $r = 2$, the problem becomes more complicated, and we lack a sufficiently sophisticated technique to yield a tight bound or handle arbitrary linear constraints. We can prove only a loose upper bound for the linear-constraints that characterize the underlying sub-keys as being $(t-2)$-wise independent, leading to a partial result for saving three keys of t-round KAC.

Fourier Analysis. To prove the bounds, we rely on the tool of Fourier analysis. In this part we define some notations for the Fourier analysis over $\{0, 1\}^m$. Given a function $f : \{0, 1\}^m \to \mathbb{R}$, the Fourier coefficient of f with $\alpha \in \{0, 1\}^m$ is defined as

$$\hat{f}(\alpha) := \frac{1}{2^m} \sum_{x \in \{0,1\}^m} f(x)(-1)^{\langle \alpha, x \rangle} .$$

Then, we have

$$f(x) = \sum_{\alpha \in \{0,1\}^m} \hat{f}(\alpha)(-1)^{\langle \alpha, x \rangle} . \tag{5}$$

For any set $S \subseteq \{0, 1\}^m$, we let $\mathbb{1}_S : \{0, 1\}^m \to \{0, 1\}$ be the 0/1 indicator function of S. Then, the following properties hold for $\mathbb{1}_S$:

$$\widehat{\mathbb{1}_S}(0) = \frac{|S|}{2^m} = \sum_{\alpha \in \{0,1\}^m} \widehat{\mathbb{1}_S}(\alpha)^2 , \tag{6}$$

$$\forall \alpha \in \{0, 1\}^m : \ |\widehat{\mathbb{1}_S}(\alpha)| \leqslant \widehat{\mathbb{1}_S}(0) = \frac{|S|}{2^m} \ . \tag{7}$$

4.1 1-constraint Sum Capture Quantity

We associate a 1-constraint sum-capture quantity with a vector of coefficients $\vec{c} = (c_0, c_1, \ldots, c_{t-1})$ as follows:

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) :=$$
$$\left|\left\{(v_0, (u_1, v_1), \ldots, (u_{t-1}, v_{t-1}), u_t) \in V_0 \times \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_{t-1} \times U_t \ : \ \sum_{j=0}^{t-1} c_j(v_j + u_{j+1}) = 0\right\}\right| .$$

Lemma 2. *Let $t \geqslant 2$. Let $P_1, \ldots, P_{t-1}$ be $t - 1$ independent uniformly random permutations of $\{0, 1\}^n$, and let $\mathcal{A}$ be a probabilistic algorithm that makes adaptive queries to $P_1, \ldots, P_{t-1}$. Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}$ be the query transcripts of $P_1, \ldots, P_{t-1}$ interacting with $\mathcal{A}$. Let $\vec{c} = (c_0, \ldots, c_{t-1})$ be* *any coefficients so that there exists an index* $0 \leqslant \mathsf{idx} < t-1$ *satisfying* $c_{\mathsf{idx}} \neq 0$ *and* $c_{\mathsf{idx}+1} \neq 0$. *Then, for any* $\mathcal{A}$ *that makes at most* q *queries to each permutations,*

$$\mathsf{Pr}_{P_1,\ldots,P_{t-1}}\Bigg[\exists V_0, U_t \subseteq \mathbb{F}_{2^n}, |V_0| = |U_t| = q, \\ \mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \geqslant \frac{3q^{t+1}}{N} + 3q^{t-1/2}\sqrt{(t+2)n}\Bigg] \leqslant \frac{2t}{N^t} \ .$$

We let $\Phi(\mathcal{Q}_i) := \max_{\alpha\neq 0,\beta\neq 0} N^2|\widehat{\mathbb{1}_{Q_i}}(\alpha,\beta)|$ for the query records $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}$. To show Lemma 2, we first rely on the following Lemma 3, which states the upper bound in terms of the $\Phi(\mathcal{Q}_i)$ we previously defined. Then, we apply Lemma 4 by Chen *et al.* [5], which provides an upper bound for the $\Phi(\mathcal{Q}_i)$ term to conclude the proof.

Lemma 3. *Fix any* $\vec{c} = (c_0, \ldots, c_{t-1})$ *such that* $c_{\mathsf{idx}} \neq 0$ *and* $c_{\mathsf{idx}+1} \neq 0$ *for some index* $0 \leqslant \mathsf{idx} < t-1$*, then for any subsets* V_0, U_t *with* $|V_0| = |U_t| = q$*,*

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \leqslant \frac{q^{t+1}}{N} + q^{t-1}\Phi(\mathcal{Q}_{\mathsf{idx}+1}) \ .$$

Proof. The first step is to write $\mu_{\vec{c}}$ as a sum over indicator functions; we then will perform a Fourier transform over each indicator function. Although the summation will be over many terms and Fourier coefficients, the key point is that we can eliminate most summation terms and simplify the equality to sum over only a single Fourier coefficient term.

Here, we sum over the indicator functions.

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \\ = \sum_{v_0}\sum_{u_1,v_1}\cdots\sum_{u_{t-1},v_{t-1}}\sum_{u_t} \mathbb{1}_{V_0}(v_0)\mathbb{1}_{\mathcal{Q}_1}(u_1,v_1)\cdots\mathbb{1}_{\mathcal{Q}_{t-1}}(u_{t-1},v_{t-1})\cdot\mathbb{1}_{U_t}(u_t)\cdot\mathbb{1}_{\mathsf{Eq}}\left(0, \sum_{j=0}^{t-1} c_j(v_j + u_{j+1})\right) \ ,$$

where $\mathbb{1}_{\mathsf{Eq}}(x,y)$ is the equality indicator function so that $\mathbb{1}_{\mathsf{Eq}}(x,y) = 1$ if and only if $x = y$. Note that for the equality indicator function, we can perform a Fourier transformation to get

$$\mathbb{1}_{\mathsf{Eq}}(x,y) = \sum_{\alpha,\beta}\widehat{\mathbb{1}_{\mathsf{Eq}}}(\alpha,\beta)\cdot(-1)^{\langle\alpha,x\rangle+\langle\beta,y\rangle} = \frac{1}{N}\cdot\sum_{\alpha}(-1)^{\langle\alpha,x+y\rangle},$$

where we use the fact that

$$\widehat{\mathbb{1}_{\mathsf{Eq}}}(\alpha,\beta) = \begin{cases} 1/N & \text{if } \alpha = \beta \\ 0 & o.w. \end{cases} \ .$$

We expand each indicator function using a Fourier transform and continue the calculation.

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \\ = \sum_{\substack{v_0,u_1,v_1,\cdots \\ u_{t-1},v_{t-1},u_t}} \left(\sum_{\beta_0}\widehat{\mathbb{1}_{V_0}}(\beta_0)(-1)^{\langle\beta_0,v_0\rangle}\right)\cdot\left(\sum_{\alpha_1,\beta_1}\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha_1,\beta_1)(-1)^{\langle\alpha_1,u_1\rangle+\langle\beta_1,v_1\rangle}\right)\cdots$$

$$
\cdot \left(\sum_{\alpha_{t-1},\beta_{t-1}} \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\alpha_{t-1},\beta_{t-1})(-1)^{\langle \alpha_{t-1},u_{t-1}\rangle + \langle \beta_{t-1},v_{t-1}\rangle} \right)
$$

$$
\cdot \left(\sum_{\alpha_t} \widehat{\mathbb{1}_{U_t}}(\alpha_t)(-1)^{\langle \alpha_t,u_t\rangle} \right) \cdot \frac{1}{N} \left(\sum_{\gamma} (-1)^{\langle \gamma, \sum_{j=0}^{t-1} c_j(v_j+u_{j+1})\rangle} \right) .
$$

Here, notice that all Fourier coefficients depend only on the variables αs, βs and γ, so we can expand the multiplication and change the order of summation to obtain

$$
\begin{aligned}
&\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \\
&= \frac{1}{N} \cdot \sum_{\beta_0} \sum_{\alpha_1,\beta_1} \cdots \sum_{\alpha_{t-1},\beta_{t-1}} \sum_{\alpha_t} \sum_{\gamma} \widehat{\mathbb{1}_{V_0}}(\beta_0) \widehat{\mathbb{1}_{Q_1}}(\alpha_1,\beta_1) \cdots \widehat{\mathbb{1}_{Q_{t-1}}}(\alpha_{t-1},\beta_{t-1}) \widehat{\mathbb{1}_{U_t}}(\alpha_t) \\
&\quad \cdot \sum_{v_0} \sum_{u_1,v_1} \cdots \sum_{u_{t-1},v_{t-1}} \sum_{u_t} (-1)^{\langle \beta_0,v_0\rangle} (-1)^{\langle \alpha_1,u_1\rangle + \langle \beta_1,v_1\rangle} \ldots (-1)^{\langle \alpha_{t-1},u_{t-1}\rangle + \langle \beta_{t-1},v_{t-1}\rangle} \\
&\quad \cdot (-1)^{\langle \alpha_t,u_t\rangle} \cdot (-1)^{\langle \gamma, \sum_{j=0}^{t-1} c_j(v_j+u_{j+1})\rangle} \\
&= \frac{1}{N} \cdot \sum_{\beta_0} \sum_{\alpha_1,\beta_1} \cdots \sum_{\alpha_{t-1},\beta_{t-1}} \sum_{\alpha_t} \sum_{\gamma} \widehat{\mathbb{1}_{V_0}}(\beta_0) \widehat{\mathbb{1}_{Q_1}}(\alpha_1,\beta_1) \cdots \widehat{\mathbb{1}_{Q_{t-1}}}(\alpha_{t-1},\beta_{t-1}) \\
&\quad \cdot \widehat{\mathbb{1}_{U_t}}(\alpha_t) \cdot \left(\sum_{v_0} (-1)^{\langle \beta_0,v_0\rangle + \langle \gamma, c_0v_0\rangle} \right) \cdot \left(\sum_{u_1} (-1)^{\langle \alpha_1,u_1\rangle + \langle \gamma, c_0u_1\rangle} \right) \\
&\quad \left(\sum_{v_1} (-1)^{\langle \beta_1,v_1\rangle + \langle \gamma, c_1v_1\rangle} \right) \cdots \left(\sum_{u_t} (-1)^{\langle \alpha_t,u_t\rangle + \langle \gamma, c_{t-1}u_t\rangle} \right) .
\end{aligned}
$$

The last equality is simply grouping the inner products that share the same u, v terms. Note that the field multiplication of $c \cdot x$ can be represented as a matrix A_c[4] that applies to an n-dimensional vector x over $\mathbb{F}_2$. If $c = 0$, then $A_c = O$, where we use O to denote an all zero matrix; otherwise, A_c is a full-rank matrix. Taking the summation over the v_0 term as an example, we rewrite the $\langle \gamma, c_0v_0\rangle$ term as $\langle \gamma, c_0v_0\rangle = \gamma^\top A_{c_0} v_0 = (A_{c_0}^\top \gamma)^\top v_0 = \langle A_{c_0}^\top \gamma, v_0\rangle$, where $A_{c_0}^\top$ is the transpose of A_{c_0}. So we get

$$
\begin{aligned}
&\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \\
&= \frac{1}{N} \cdot \sum_{\beta_0} \sum_{\alpha_1,\beta_1} \cdots \sum_{\alpha_{t-1},\beta_{t-1}} \sum_{\alpha_t} \sum_{\gamma} \widehat{\mathbb{1}_{V_0}}(\beta_0) \widehat{\mathbb{1}_{Q_1}}(\alpha_1,\beta_1) \cdots \widehat{\mathbb{1}_{Q_{t-1}}}(\alpha_{t-1},\beta_{t-1}) \cdot \widehat{\mathbb{1}_{U_t}}(\alpha_t) \\
&\cdot \left(\sum_{v_0} (-1)^{\langle \beta_0 + A_{c_0}^\top \gamma, v_0\rangle} \right) \cdot \left(\sum_{u_1} (-1)^{\langle \alpha_1 + A_{c_0}^\top \gamma, u_1\rangle} \right) \left(\sum_{v_1} (-1)^{\langle \beta_1 + A_{c_1}^\top \gamma, v_1\rangle} \right) \\
&\cdots \left(\sum_{u_{t-1}} (-1)^{\langle \alpha_{t-1} + A_{c_{t-2}}^\top \gamma, u_{t-1}\rangle} \right) \left(\sum_{v_{t-1}} (-1)^{\langle \beta_{t-1} + A_{c_{t-1}}^\top \gamma, v_{t-1}\rangle} \right) \left(\sum_{u_t} (-1)^{\langle \alpha_t + A_{c_{t-1}}^\top \gamma, u_t\rangle} \right) .
\end{aligned}
$$

[4] Since we are taking the natural field interpretation over $\{0,1\}^n$, where the field addition is the bit-wise xor operation, we define the i-th column of A_c as the n-dimension vector representation of field element $c \cdot \nu_i$, where ν_i is the field element that has the corresponding representation to be a basis vector with the i-th position being one and the rest positions being zero.

It is known that $\sum_{x \in \{0,1\}^n} (-1)^{\langle \alpha, x \rangle} = N$ if and only if $\alpha = 0$; otherwise, it equals zero. We are interested only in the case where the Fourier coefficients give a non-zero summation. Interestingly, we can obtain the following equalities that are written in terms of γ, i.e.,

$$\forall i \in \{0, \dots, t-1\} : \ \alpha_{i+1} = \beta_i = A_{c_i}^\top \gamma \ ,$$

where A_{c_i} is full-rank if $c_i \neq 0$ or $A_{c_i} = O$ if $c_i = 0$. Hence, the equality calculation can be greatly simplified as

$$\begin{aligned} \mu_{\vec{c}}(V_0, \mathcal{Q}_1, \dots, \mathcal{Q}_{t-1}, U_t) &= N^{2t-1} \sum_{\gamma} \widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma) \widehat{\mathbb{1}_{Q_1}}(A_{c_0}^\top \gamma, A_{c_1}^\top \gamma) \cdots \widehat{\mathbb{1}_{Q_{t-1}}}(A_{c_{t-2}}^\top \gamma, A_{c_{t-1}}^\top \gamma) \widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma) \\ &= \frac{q^{t+1}}{N} + N^{2t-1} \sum_{\gamma \neq 0} \widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma) \widehat{\mathbb{1}_{Q_1}}(A_{c_0}^\top \gamma, A_{c_1}^\top \gamma) \cdots \widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma) \\ &\leqslant \frac{q^{t+1}}{N} + N^{2t-1} \sum_{\gamma \neq 0} |\widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{Q_1}}(A_{c_0}^\top \gamma, A_{c_1}^\top \gamma)| \cdots |\widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma)| \ . \end{aligned}$$

We let

$$\begin{aligned} \mathsf{left} &:= \text{min of } i \text{ such that } c_i \neq 0 \\ \mathsf{right} &:= \text{max of } i \text{ such that } c_i \neq 0 \ . \end{aligned}$$

To proceed, we need a case discussion of $(\mathsf{left}, \mathsf{right})$. Here, we consider the case of $\mathsf{left} = 0$ and $\mathsf{right} = t-1$ (i.e., $c_0 \neq 0$ and $c_{t-1} \neq 0$). The other cases yield the same upper bound, and we describe them in appendix B.1 for completeness.

Therefore, we obtain

$$\begin{aligned} &\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \dots, \mathcal{Q}_{t-1}, U_t) \\ &\leqslant \frac{q^{t+1}}{N} + N^{2t-1} \sum_{\gamma \neq 0} |\widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{Q_1}}(A_{c_0}^\top \gamma, A_{c_1}^\top \gamma)| \cdots |\widehat{\mathbb{1}_{Q_{t-1}}}(A_{c_{t-2}}^\top \gamma, A_{c_{t-1}}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma)| \\ &\leqslant \frac{q^{t+1}}{N} + N^{2t-3} \sum_{\gamma \neq 0} |\widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma)| \cdot \left(\frac{q}{N^2}\right)^{t-2} \cdot \Phi(\mathcal{Q}_{\mathsf{idx}+1}) \cdot |\widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma)| \\ &= \frac{q^{t+1}}{N} + q^{t-2} N \Phi(\mathcal{Q}_{\mathsf{idx}+1}) \cdot \sum_{\gamma \neq 0} |\widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma)| \leqslant \frac{q^{t+1}}{N} + q^{t-1} \Phi(\mathcal{Q}_{\mathsf{idx}+1}) \ . \end{aligned} \tag{8}$$

Given the condition that $c_{\mathsf{idx}} \neq 0$ and $c_{\mathsf{idx}+1} \neq 0$, we have $N^2 |\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{idx}+1}}}(A_{c_{\mathsf{idx}}} \gamma, A_{c_{\mathsf{idx}+1}} \gamma)| \leqslant \Phi(\mathcal{Q}_{\mathsf{idx}+1})$ for any $\gamma \neq 0$. We also used (7) that, for any α, β, $|\widehat{\mathbb{1}_{\mathcal{Q}_i}}(\alpha, \beta)| \leqslant q/N^2$. We note that the last step of inequality holds because by (6) we have $\sum_\gamma \widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma)^2 = \sum_\gamma \widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma)^2 = q/N$, so we can apply the Cauchy-Schwartz inequality to obtain the result. This exact inequality step ensures the tight bound and was referred to as the *Cauchy-Schwartz trick* in [2,23,5].

Therefore, we proved Lemma 3. □

The remaining step is to upper bound $\Phi(\mathcal{Q}_{\mathsf{idx}+1})$. Here, we apply the following lemma, which has essentially the same proof of Lemma 6 proved by Chen *et al.* in [5], with the only adjustment of changing their parameter δ to $\delta = \sqrt{(12 \ln N)/q}$.

Lemma 4. *Assume that $9(t+2)n \leqslant q \leqslant N/2$. Fix an adversary making q queries to a random permutation P. Let Q denote the transcript of interaction of $\mathcal{A}$ with P. Then,*

$$\mathsf{Pr}_{P,\omega}\left[\Phi(Q) \geqslant \frac{2q^2}{N} + 3\sqrt{(t+2)nq}\right] \leqslant \frac{2}{N^t} \ ,$$

where the probability is taken over the random permutation P and the random coins ω used by $\mathcal{A}$.

Plugging in the inequality, we get

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \leqslant \frac{q^{t+1}}{N} + q^{t-1}\Phi(Q_{\mathsf{idx}+1}) \leqslant \frac{q^{t+1}}{N} + q^{t-1}\left(\frac{2q^2}{N} + 3\sqrt{(t+2)nq}\right)$$
$$= \frac{3q^{t+1}}{N} + 3q^{t-1/2}\sqrt{(t+2)n}$$

with probability at least $1 - \frac{2t}{N^t}$. Hence, we proved Lemma 2.

Tightness of Lemma 2. We examine the tightness of 1-constraint sum-capture quantities in two aspects. The first is, given that $\vec{c} = (c_0, \ldots, c_{t-1})$ where there are two neighboring c_i, c_{i+1} so that $c_i \neq 0, c_{i+1} \neq 0$ whether or not the upper bound is tight.

We now give the following proposition, showing that if there exist neighboring coefficients $c_i \neq 0$ and $c_{i+1} \neq 0$, then for moderately large q (e.g. $q > N^{2/3}$), $\mu_{\vec{c}} \geqslant q^{t+1}/2N$ with high probability. We defer the detailed proof to Appendix B.2.

Proposition 1. *Let q be any positive integer of power of two. Fix any $\vec{c} = (c_0, \ldots, c_{t-1})$ such that there exists an index $0 \leqslant i < t-1$ satisfying $c_i \neq 0$ and $c_{i+1} \neq 0$. Then, there is an explicit algorithm $\mathcal{A}$ that makes at most q queries to each of $P_1, \ldots, P_{t-1}$, and $V_0, U_t \subseteq \mathbb{F}_{2^n}$ that have $|V_0| = |U_t| = q$ so that*

$$\mathsf{Pr}\left[\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \geqslant \frac{q^{t+1}}{2N}\right] \geqslant 1 - \frac{N}{q} \cdot e^{-q^2/8N} \ .$$

The second proposition, complementary to Proposition 1, states that if $\vec{c} = (c_0, \ldots, c_{t-1})$ satisfies that if for any $0 \leqslant i < t-1$, either $c_i = 0$ or $c_{i+1} = 0$, then $\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t)$ can achieve up to q^t, which is larger than q^{t+1}/N. We leave the proof to Appendix B.3.

Proposition 2. *Let q be any positive integer of power of two. Fix any $\vec{c} = (c_0, \ldots, c_{t-1})$ such that for any $0 \leqslant i < t-1$, either $c_i = 0$ or $c_{i+1} = 0$. Then, there is an explicit algorithm $\mathcal{A}$ that makes at most q queries to each of $P_1, \ldots, P_{t-1}$, and $V_0, U_t \subseteq \mathbb{F}_{2^n}$ that have $|V_0| = |U_t| = q$, so that*

$$\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \geqslant q^t \ .$$

4.2 2-constraint Sum-Capture Quantity

We now consider the sum-capture quantity for which the number of constraints $r = 2$. We associate the 2-constraint sum-capture quantity with two vectors of coefficients, $\vec{c} = (c_0, c_1, \ldots, c_{t-1})$ and $\vec{d} = (d_0, d_1, \ldots, d_{t-1})$, as

$$\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) := \\ \left| \left\{ (v_0, (u_1, v_1), \ldots, (u_{t-1}, v_{t-1}), u_t) \in V_0 \times \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_{t-1} \times U_t \ : \right. \right. \\ \left. \left. \sum_{j=0}^{t-1} c_j(v_j + u_{j+1}) = 0, \ \sum_{j=0}^{t-1} d_j(v_j + u_{j+1}) = 0 \right\} \right| \ . \quad (9)$$

Though the 2-constraint sum-capture quantity is a natural generalization of the 1-constraint case, we note that adding only one more constraint makes proving the tightest upper bound of (9) much harder. Here, we give bounds only over the sum-capture quantity with a specific class of coefficients $\vec{c}, \vec{d}$ that can be derived from the $(t-2)$-wise independently uniform sub-keys. We obtain a bound that gives the tightest KAC security for $t \geqslant 8$. However, for $t < 5$, our 2-constraint upper bound is even worse than a reduction-based bound (see Appendix B.4). While it is interesting to investigate whether our bound can be improved, for $t = 3$, in particular, we show that the preceding sum-capture quantity is lower-bounded by $\Omega(q^3/N)$ and hence cannot be used to prove $q = \Omega(N^{3/4})$ for the 3-round KAC with identical sub-keys.

We prove upper bounds for the class of linear constraint coefficients $\vec{c} = (c_0, \ldots, c_{t-1}), \vec{d} = (d_0, \ldots, d_{t-1})$ with the property that $c_0 = d_{t-1} = 1$, $c_{t-1} = d_0 = 0$, and for all $i \in \{1, \ldots, t-2\}$, $c_i \neq 0, d_i \neq 0$, and for all $i, j \in \{1, \ldots, t-2\}$ such that $i \neq j$, $c_i d_i^{-1} \neq c_j d_j^{-1}$. We justify that $\vec{c}, \vec{d}$ corresponds to the linear key schedule from $t-2$ independent keys that gives $(t-2)$-wise independently uniform sub-keys.

Justification. We use $s_0, \ldots, s_{t-1}$ to denote the sub-keys. Given that the sub-keys are generated linearly from $t-2$ independent keys and are $(t-2)$-wise independently uniform, the middle $t-2$ sub-keys $s_1, \ldots, s_{t-2}$ uniquely fix the original master keys; hence, the first subkey s_0 and the last sub-keys s_{t-1} can be uniquely determined as a linear combination of $s_1, \ldots, s_{t-2}$, i.e.,

$$s_0 = \sum_{i=1}^{t-2} c_i s_i \ , \quad s_{t-1} = \sum_{i=1}^{t-2} d_i s_i \ .$$

Note that all c_i, d_i should be non-zero because otherwise we could obtain a linear combination $t-2$ sub-keys that sum to zero, breaking the $(t-2)$-wise independence. Further, we show by contradiction that if there exists i, j such that $i \neq j$ and $c_i d_i^{-1} = c_j d_j^{-1}$, then we pick the set of sub-keys $\{s_0, s_{t-1}\} \cup \{s_k \mid 1 \leqslant k \leqslant t-2 \wedge k \notin \{i, j\}\}$ and have

$$s_0 + c_i d_i^{-1} s_{t-1} = \sum_{k \notin \{0,i,j,t\}} (c_i d_i^{-1} d_k + c_k) s_k \ ,$$

which is a linear dependence among $t-2$ sub-keys. Thus, all $c_i d_i^{-1}$ must be distinct.

Then, we have the following lemma for the 2-constraint sum-capture quantity.

Lemma 5. *Let $t \geqslant 3$. Let $P_1, \ldots, P_{t-1}$ be $t-1$ independent uniformly random permutations of $\{0,1\}^n$, and let $\mathcal{A}$ be a probabilistic algorithm that makes adaptive queries to $P_1, \ldots, P_{t-1}$. Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}$ be the query transcripts of $P_1, \ldots, P_{t-1}$ interacting with $\mathcal{A}$. Let coefficients $\vec{c}, \vec{d}$ be defined as above. Then, for any $\mathcal{A}$ that makes at most $q \geqslant (t+2)nN^{2/3}$ queries to each permutation,*

$$\mathsf{Pr}_{P_1,\ldots,P_{t-1}}\Bigg[\exists V_0, U_t \subseteq \mathbb{F}_{2^n}, |V_0| = |U_t| = q,$$

$$\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \geqslant \frac{q^{t+1}}{N^2} + t \cdot \frac{(3q)^{2t-3}}{N^{t-2}} + \frac{(3q)^{2t-2.5}}{N^{t-2}}\Bigg] \leqslant \frac{2t}{N^t} \ .$$

Discussion. Note that when $t \geqslant 5$, the security bound starts to improve over the $t-1$ round KAC bound $q = \Omega(N^{\frac{t-1}{t}})$. For $t \geqslant 8$, the security bound achieves optimal security of $q = \Omega(N^{\frac{t}{t+1}})$.

As in the 1-constraint case, we prove an upper bound of $\mu_{\vec{c},\vec{d}}$ conditioning on $\Phi(\mathcal{Q}_i)$ being small for all i.

Lemma 6. *Fix $\vec{c}, \vec{d}$ defined as in Lemma 5. Then, conditioning on $\Phi(\mathcal{Q}_i) \leqslant 9q^2/N$ for all $1 \leqslant i \leqslant t-1$, it holds that for any subsets $V_0, U_t \subseteq \mathbb{F}_{2^n}$ with $|V_0| = |U_t| = q$,*

$$\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \leqslant \frac{q^{t+1}}{N^2} + t \cdot \frac{(3q)^{2t-3}}{N^{t-2}} + \frac{(3q)^{2t-2.5}}{N^{t-2}} \ .$$

Proof. We first expand the quantity as sums over indicator function, and perform Fourier expansion, grouping inner products as we did for the 1-constraint case. The following shows the calculation result, and we left the details in Appendix B.5.

$$\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) = N^{2t-2} \sum_{\alpha,\beta} \widehat{\mathbb{1}_{V_0}}(\theta_0)\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\theta_0, \theta_1)\widehat{\mathbb{1}_{\mathcal{Q}_2}}(\theta_1, \theta_2) \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, \theta_{t-1})\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})$$

where

$$\theta_0 = \alpha, \quad \theta_{t-1} = \beta,$$
$$\forall i \in \{1, \ldots, t-2\} : \ \theta_i = A_{c_i}^\top \alpha + A_{d_i}^\top \beta \ .$$

We write $\mathsf{Coeff} = \{\theta_0, \theta_1, \ldots, \theta_{t-1}\}$. Here, we partition the summation into three cases and discuss the set of (α, β) assignments that falls into each case.

1. At least two θs in Coeff are zero.
2. Exactly one θ in Coeff is zero.
3. No θs in Coeff is zero.

The following claim shows that if case one happens, then all coefficients θ are zero.

Claim 1 *If two θs in Coeff are zero, then $\alpha = \beta = 0$.*

Proof. If $\theta_0 = \alpha = \beta = \theta_{t-1} = 0$, then the claim is trivial. If $\alpha = \theta_0 = \theta_i = 0$ for some i with $1 \leqslant i \leqslant t-2$, then given that $\theta_i = A_{c_i}^\top \alpha + A_{d_i}^\top \beta = A_{d_i}^\top \beta$ and $A_{d_i}^\top$ is full-rank (because $d_i \neq 0$), we can infer that $\beta = 0$. Similarly, we can infer $\alpha = 0$ if $\beta = \theta_{t-1} = \theta_i = 0$ for some i with $1 \leqslant i \leqslant t-2$. Now, if $\theta_i = \theta_j = 0$ for some i, j such that $1 \leqslant i, j \leqslant t-2$ and $i \neq j$, the choice of (α, β) must satisfy

$$\begin{cases} A_{c_i}^\top \alpha + A_{d_i}^\top \beta = 0 \\ A_{c_j}^\top \alpha + A_{d_j}^\top \beta = 0 \end{cases}$$

implying that $A^\top_{d^{-1}_{i+1}c_{i+1}}\alpha = (A^\top_{d_{i+1}})^{-1}A^\top_{c_{i+1}}\alpha = \beta = (A^\top_{d_{j+1}})^{-1}A^\top_{c_{j+1}}\alpha = A^\top_{d^{-1}_{j+1}c_{j+1}}\alpha$. Hence,

$$\left(A^\top_{d^{-1}_{i+1}c_{i+1}} + A^\top_{d^{-1}_{j+1}c_{j+1}}\right)\alpha = \left(A^\top_{d^{-1}_{i+1}c_{i+1}+d^{-1}_{j+1}c_{j+1}}\right)\alpha = 0 \; .$$

Here, α can be non-zero only if $d^{-1}_{i+1}c_{i+1} = d^{-1}_{j+1}c_{j+1}$. However, as we explained in the $(t-2)$-wise independently uniform property of sub-keys, this is not possible. □

Let μ_1, μ_2, μ_3 correspond to summation over (α, β) for case one, two, three, respectively.

Proposition 3.

$$\mu_1 = \frac{q^{t+1}}{N^2}$$

Proof. Since case one happens only when $\alpha = \beta = 0$, we have $\theta_i = 0$ for all i. Therefore, a direct calculation using the fact that $\widehat{\mathbb{1}_{V_0}}(0) = \widehat{\mathbb{1}_{U_t}}(0) = q/N$ and $\widehat{\mathbb{1}_{\mathcal{Q}_i}}(0,0) = q/N^2$ proves the bound. □

Proposition 4.

$$\mu_2 \leqslant \frac{t \cdot (3q)^{2t-3}}{N^{t-2}}$$

Since the proof of Proposition 4 can be derived via a moderate change to the proof of the 1-constraint sum-capture quantity upper bound (i.e., Lemma 2), we left the complete proof to Appendix B.6.

Proposition 5.

$$\mu_3 \leqslant \frac{(3q)^{2t-2.5}}{N^{t-2}}$$

Proof (of Proposition 5). We define an $N \times N$ matrix M with each entry labeled by $(\alpha, \beta) \in \mathbb{F}_{2^n} \times \mathbb{F}_{2^n}$ so that

$$M_{\alpha,\beta} = \begin{cases} 0 & \text{if some } \theta \in \mathsf{Coeff} \text{ is } 0 \\ \widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha, A^\top_{c_1}\alpha + A^\top_{d_1}\beta) \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(A^\top_{c_{t-2}}\alpha + A^\top_{d_{t-2}}\beta, \beta) & o.w. \end{cases}$$

Note that M is a $2^n \times 2^n$ matrix. We also define the column vectors $\vec{v}$, $\vec{u}$ with each entry labeled by $\alpha \in \mathbb{F}_{2^n}$ so that $\vec{v}_\alpha = \widehat{\mathbb{1}_{V_0}(\alpha)}$ and $\vec{u}_\alpha = \widehat{\mathbb{1}_{U_t}(\alpha)}$. Therefore, we can write μ_3 as

$$\mu_3 = N^{2t-2} \sum_{\alpha,\beta \,\mid\, M_{\alpha,\beta} \neq 0} \widehat{\mathbb{1}_{V_0}(\alpha)} \cdot M_{\alpha,\beta} \cdot \widehat{\mathbb{1}_{U_t}(\beta)} = N^{2t-2}\vec{v}^\top M \vec{u} \; .$$

Noting that the equivalent definition of the matrix 2-norm is

$$\|M\|_2 := \sup_{\|x\|_2=1} \|Mx\|_2 = \sup_{\|x\|_2=1, \|y\|_2=1} y^\top M x \; ,$$

we can use the matrix norm as the upper bound of μ_3, that is,

$$\mu_3 = N^{2t-2} \cdot \vec{v}^\top M \vec{u} \leqslant N^{2t-2} \|\vec{v}\|_2 \|M\|_2 \|\vec{u}\|_2 \; .$$

By (6), we can infer that $\|\vec{v}\|_2 = \sqrt{\sum_\alpha v_\alpha^2} = \sqrt{\sum_\alpha \widehat{\mathbb{1}_{V_0}}(\alpha)^2} = \sqrt{q/N}$ and $\|\vec{u}\|_2 = \sqrt{q/N}$. We also use the fact that $\|M\|_2 \leqslant \|M\|_F$, where $\|M\|_F = \sqrt{\sum_{i,j} M_{i,j}^2}$ is the Frobenius norm. Then, we have

$$\mu_3 \leqslant N^{2t-2} \cdot \sqrt{\frac{q}{N}}\, \|M\|_2 \sqrt{\frac{q}{N}} \leqslant qN^{2t-3}\, \|M\|_F = qN^{2t-3} \sqrt{\sum_{\alpha,\beta} M_{\alpha,\beta}^2}\ ,$$

where

$$\begin{aligned}
\sum_{\alpha,\beta} M_{\alpha,\beta}^2 &= \sum_{\alpha,\beta \,\mid\, M_{\alpha,\beta} \neq 0} \widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha, A_{c_1}^\top \alpha + A_{d_1}^\top \beta)^2 \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(A_{c_{t-2}}^\top \alpha + A_{d_{t-2}}^\top \beta, \beta)^2 \\
&\leqslant \sum_{\alpha,\beta \,\mid\, M_{\alpha,\beta} \neq 0} \widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha, A_{c_1}^\top \alpha + A_{d_1}^\top \beta)^2 \cdot \frac{(3q)^{4(t-2)}}{N^{6(t-2)}} \\
&\leqslant \frac{(3q)^{4(t-2)}}{N^{6(t-2)}} \sum_{\alpha,\beta} \widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha, A_{c_1}^\top \alpha + A_{d_1}^\top \beta)^2 = \frac{(3q)^{4(t-2)}}{N^{6(t-2)}} \cdot \frac{q}{N^2} \leqslant \frac{(3q)^{4t-7}}{N^{6t-10}}\ .
\end{aligned}$$

Therefore, we get

$$\mu_3 \leqslant qN^{2t-3} \cdot \frac{(3q)^{2t-3.5}}{N^{3t-5}} \leqslant \frac{(3q)^{2t-2.5}}{N^{t-2}}\ .$$

□

Putting all propositions together, we have

$$\mu_{\vec{c},\vec{d}} = \mu_1 + \mu_2 + \mu_3 \leqslant \frac{q^{t+1}}{N^2} + t \cdot \frac{(3q)^{2t-3}}{N^{t-2}} + \frac{(3q)^{2t-2.5}}{N^{t-2}}\ .$$

□

4.3 Tightness of 2-constraint Sum-Capture Quantity for 3-round KAC

A natural question is whether the upper bound of the 2-constraint sum-capture quantity can be improved to give a tight security bound for a t-round KAC when $t < 7$. In particular, the most interesting case is to prove a tight security bound $q = \Omega(N^{3/4})$ for a 3-round KAC with identical sub-keys that corresponds to the instantiation in Corollary 4 when $t = 3$. However, for the 3-round KAC with an identical key schedule, we prove that it is impossible to show the conjectured optimal security bound via upper-bounding the sum-capture quantity, because the sum-capture quantity is lower-bounded by $\Omega(q^3/N)$ with high probability, giving $\mu_{\vec{c}}/N = \Omega(q^3/N^2)$ instead of the desired q^4/N^3. The sum-capture quantity lower bound for a 3-round identical sub-key KAC directly follows from Proposition 6 with $c_1 = d_1 = 1$. We left the proof of proposition to Appendix B.7.

Proposition 6. *Let q be any positive integer of power of two. Let $t = 2$ and fix $\vec{c} = (1, c_1, 0)$, $\vec{d} = (0, d_1, 1)$, where c_1, d_1 are non-zero. Then, there exists an explicit algorithm $\mathcal{A}$ that makes at most q queries to each of P_1, P_2 and $V_0, U_3 \subseteq \mathbb{F}_{2^n}$ that have $|V_0| = |U_3| = q$, so that*

$$\mathsf{Pr}[\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \mathcal{Q}_2, U_3) \geqslant q^3/2N] \geqslant 1 - \frac{N}{q} \cdot e^{-q^2/8N}\ .$$

Though Proposition 6 gives a lower bound of $\Omega(q^3/N)$ for the sum-capture quantity $\mu_{\vec{c},\vec{d}}$, it does not immediately imply a distinguishing attack against the 3-round KAC. This is because the number of bad keys generated by our constructed $\mathcal{A}$ is at most q, so we have $\mathsf{Pr}[\vec{k} \in \mathsf{Badkey}] \leqslant q/N$; The reason of $\mu_{\vec{c},\vec{d}}$ being too large is that a bad key may be counted multiple times in the sum capture quantity. Therefore, we cannot prove the optimal $q = \Omega(N^{3/4})$ bound for a 3-round KAC with identical sub-keys if overcounting cannot be eliminated.

5 Good Transcript Analysis

We next obtain upper bounds of $1 - \mathsf{ps}_0(\tau)/\mathsf{ps}_1(\tau)$ for each $\tau \in \mathcal{T}_{\mathsf{good}}$ in the following lemma.

Lemma 7. *If the t-round KAC is instantiated with a key schedule that gives $(t-2)$-wise independently uniform sub-keys, then there exists a function $g : \mathcal{T} \to [0, +\infty)$ so that for any $\tau = (\vec{\mathcal{Q}}, \vec{k}) \in \mathcal{T}_{\mathsf{good}}$,*

$$1 - \frac{\mathsf{ps}_0(\tau)}{\mathsf{ps}_1(\tau)} \leqslant g(\tau) ,$$

and for any query records $\vec{\mathcal{Q}}$,

$$\mathbb{E}_{\vec{k}}\left[g(\vec{\mathcal{Q}}, \vec{k})\right] \leqslant \frac{t^2(4q)^{t+1}}{N^t} .$$

To obtain the desired function $g(\cdot)$, we must first understand the ratio $\mathsf{ps}_0(\tau)/\mathsf{ps}_1(\tau)$. Given the transcript $\tau = (\vec{\mathcal{Q}}, \vec{k})$ where $\vec{\mathcal{Q}} = (\mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_t)$, we write $E \downarrow \mathcal{Q}_E$ to denote that the real-world cipher construction E is consistent with the recorded query $\mathcal{Q}_E$; that is, for each $(x, y) \in \mathcal{Q}_E$, it holds that $E(x) = y$. Similarly, we write $P_i \downarrow \mathcal{Q}_i$ to denote that the permutation P_i is consistent with the recorded query $\mathcal{Q}_i$. Then, following [5,16], one can derive that

$$\frac{\mathsf{ps}_0(\vec{\mathcal{Q}}, \vec{k})}{\mathsf{ps}_1(\vec{\mathcal{Q}}, \vec{k})} = N^{(|\mathcal{Q}_E|)} \cdot \mathsf{Pr}[E_{\vec{k}} \downarrow \mathcal{Q}_E \ \mid \ P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t] , \tag{10}$$

where $N^{(|\mathcal{Q}_E|)} = N(N-1)\cdots(N - |\mathcal{Q}_E| + 1)$. We provide a proof of (10) in Appendix C.1.

To analyze the probability term on the RHS, we take the following graph view for KAC, which was originally introduced by Chen and Steinberger in [6].

5.1 Graph Definition and a Useful Lemma

Let G be a graph that consists of vertices divisible into $m+1$ layers $L_0, \ldots, L_m$ such that each layer contains exactly N vertices and edges that can be partitioned into m sets $\vec{E} = (E_{(0,1)}, E_{(1,2)}, \ldots, E_{(m-1,m)})$ such that $E_{(i,i+1)}$ forms a partial (but possibly perfect) matching from L_i to L_{i+1}.

We say a vertex $u \in L_i$, where $i < m$, is right-free if no edge connects u to any vertex in L_{i+1}. Analogously, we say a vertex $v \in L_j$, where $j > 0$, is left-free if no edge connects v to any vertex in L_{j-1}.

For any vertex $u \in L_0$, we define the following probabilistic procedure that generates a path $(w_0, w_1, \ldots w_m)$ from u to a vertex in L_m.

- Let $w_0 = u$.

- For i from 1 to m, if w_{i-1} is not right-free and connects to some vertex $w' \in L_i$, then let $w_i = w'$; otherwise, let w_i be uniformly sampled from all left-free vertices in L_i.

We write $\mathsf{Pr}[u \to v]$ to denote the probability that the path $(u, w_1, \ldots, w_m)$ satisfies $w_m = v$. In particular, we are interested in the pair of (u, v) where u is right-free and v is left-free.

For the layered graph G, we let $\mathcal{U}_G(a, b)$, where $a \leqslant b$, be the set of paths that starts at a left-free vertex in L_a and reaches a vertex in L_b. We note that the path in $\mathcal{U}_G(a, b)$ does not necessarily end in L_b. We write $U_G(a, b) = |\mathcal{U}_G(a, b)|$. Note that $U_G(a, a)$ denotes the total number of left-free vertices in L_a.

Given any $\sigma = ((i_0, i_1), (i_1, i_2), \ldots, (i_{|\sigma|-1}, i_{|\sigma|}))$ in which $i_0 < i_1 < \cdots < i_{|\sigma|}$, we say σ is an interesting (a, b)-segment partitions with regard to the index set $\mathcal{I} \subseteq \{0, \ldots, m\}$ if $i_0 = a, i_{|\sigma|} = b$ and for all $0 < j < |\sigma|$ we have $i_j \in \mathcal{I}$. We use $\mathcal{B}_{\mathcal{I}}(a, b)$ to denote the set that contains all interesting (a, b)-segment partition of the set $\mathcal{I}$. Given a layered graph G, we define the interesting indices of G as

$$\mathcal{I}(G) := \{i \in \{0, 1, 2, \ldots, m\} \mid U_G(i, i) > 0\} .$$

We are then ready to state the following lemma, a slightly different variant of the lemma proved by Chen and Steinberger in [6] but with essentially the same proof. We include a version of the proof in Appendix C.2 for completeness.

Lemma 8. *For any graph G defined as above, and any $u \in L_0, v \in L_m$ such that u is right-free and v is left-free, it holds that*

$$\mathsf{Pr}[u \to v] = \frac{1}{N} - \frac{1}{N} \sum_{\sigma \in \mathcal{B}_{\mathcal{I}(G)}(0,m)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_G(i_{h-1}, i_h)}{U_G(i_h, i_h)} .$$

5.2 Graph View of KAC

The KAC can also be interpreted in the graph view. Given a transcript $\tau = (\vec{\mathcal{Q}}, \vec{k})$, where $\vec{\mathcal{Q}} = (\mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_t)$ and sub-keys $\vec{s} = (s_0, \ldots, s_t)$ are generated from the key $\vec{k}$, we define $E_{(2i,2i+1)} := \{(v, v + s_i) \mid v \in L_{2i}\}$ for $i \in \{0, \ldots, t\}$. That is, L_{2i} and L_{2i+1} are connected by the "sub-key edges;" this corresponds to the step of xoring the sub-key s_i in the KAC execution. For $i \in \{1, \ldots, t\}$, we let $E_{(2i-1,2i)} := \{(u, v) \mid (u, v) \in \mathcal{Q}_i\}$. This corresponds to the queries made to the permutation P_i. Now, we note that the interesting indices for KAC can only be a subset of $\{0, 2, 4, \ldots, 2t\}$.

For fixed query records $\vec{\mathcal{Q}}$, let $Z_{\vec{s}}(a, b)$, where $a \leqslant b$, be the total number of paths that connects a vertex in L_a and a vertex in L_b when $\vec{s}$ being the sub-keys. Note that the paths do not necessarily start at L_a or end at L_b. For the ℓ-th cipher query (x_ℓ, y_ℓ), let $\alpha_\ell[\vec{s}]$ denote the largest possible index of the layer that is reachable from x_ℓ when $\vec{s}$ being the sub-keys. Let $\beta_\ell[\vec{s}]$ denote the smallest index of the layer than is reachable from y_ℓ. In the good key case, we always have $\alpha_\ell[\vec{s}] < \beta_\ell[\vec{s}]$.

Now, to bound the probability $\mathsf{Pr}[E \downarrow \mathcal{Q}_E \mid P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t]$, we analyze the following experiment, which can be divided into $|\mathcal{Q}_E|$ stages.

1. Initially, G_0 is defined according to the given transcript $\tau = (\vec{\mathcal{Q}}, \vec{k})$.
2. For ℓ from 1 to $|\mathcal{Q}_E|$, given that $G_{\ell-1}$ is defined, the probabilistic path-generating process is run for the ℓ-th query $(x_\ell, y_\ell) \in \mathcal{Q}_E$ over the graph $G_{\ell-1}$, from vertex $x_\ell \in L_0$.
   – If the generated path from x_ℓ does not arrive at y_ℓ, the experiment outputs 0 and aborts.
   – Otherwise, we first set $G_\ell = G_{\ell-1}$ and then remove all vertices on the path of (x_ℓ, y_ℓ) from G_ℓ. The new graph G_ℓ will have $N - \ell$ vertices in each layer.

3. If $G_{|\mathcal{Q}_E|}$ is successfully defined, the experiment outputs 1.

So, we have

$$\frac{\mathsf{ps}_0(\vec{\mathcal{Q}}, \vec{k})}{\mathsf{ps}_1(\vec{\mathcal{Q}}, \vec{k})} = N^{(|\mathcal{Q}_E|)}\mathsf{Pr}[\mathsf{Exp}(\tau) = 1] = N^{(|\mathcal{Q}_E|)} \prod_{\ell=1}^{|\mathcal{Q}_E|} \mathsf{Pr}[x_\ell \to y_\ell \mid G_{\ell-1}] \,.$$

We are now ready to state the core lemma that defines the function $g(\vec{\mathcal{Q}}, \vec{k})$ and prove it using Lemma 8.

Lemma 9. *For any query records $\vec{\mathcal{Q}}$ with $q \leqslant N/4$ and key $\vec{k}$ such that the transcript $\tau = (\vec{\mathcal{Q}}, \vec{k}) \in \mathcal{T}_{\mathsf{good}}$,*

$$\frac{\mathsf{ps}_0(\vec{\mathcal{Q}}, \vec{k})}{\mathsf{ps}_1(\vec{\mathcal{Q}}, \vec{k})} \geqslant 1 - \sum_{\ell=1}^{q} \sum_{1 \leqslant a \leqslant b \leqslant t} \mathsf{R}_{2a-1,2b,\ell}[\vec{s}] \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N - 2q} \,,$$

where the set of interesting indices $\mathcal{I}$ of the segment partition set $\mathcal{B}_{\mathcal{I}}$ is defined as $\mathcal{I} = \{0, 2, \ldots, 2t\}$, and $\mathsf{R}_{a,b,\ell}[\vec{s}] := \mathbb{1}(\alpha_\ell[\vec{s}] \geqslant a, \beta_\ell[\vec{s}] \leqslant b)$.

Proof. For the ℓ-th cipher query (x_ℓ, y_ℓ) given the graph support $G_{\ell-1}$, we define a graph G from $G_{\ell-1}$ that removes all layers L_i for $i < \alpha_\ell[\vec{s}]$ and L_j for $j > \beta_\ell[\vec{s}]$. Thus, in the graph G, we start at a right-free vertex $u \in L_0$ and target a left-free vertex $v \in L_m$, letting us apply Lemma 8.

$$\begin{aligned}
&\mathsf{Pr}_G[(x_\ell \to y_\ell) \mid G_{\ell-1}] \\
&= \frac{1}{N - \ell + 1}\left(1 - \sum_{\sigma \in \mathcal{B}_{\mathcal{I}(G)}(0,m)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_G(i_{h-1}, i_h)}{U_G(i_h, i_h)}\right) \\
&= \frac{1}{N - \ell + 1}\left(1 + \frac{U_G(0,m)}{U_G(m,m)} - \sum_{\sigma \in \mathcal{B}_{\mathcal{I}(G)}(0,m),\ |\sigma| \geqslant 2} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_G(i_{h-1}, i_h)}{U_G(i_h, i_h)}\right) \\
&\geqslant \frac{1}{N - \ell + 1}\left(1 - \sum_{\sigma \in \mathcal{B}_{\mathcal{I}(G)}(0,m),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{U_G(i_{h-1}, i_h)}{U_G(i_h, i_h)}\right) \\
&\geqslant \frac{1}{N - \ell + 1}\left(1 - \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(\alpha_\ell[\vec{s}],\beta_\ell[\vec{s}]),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N - 2q}\right) \,. \qquad (11)
\end{aligned}$$

Now we consider only the case where the lower bound (11) $\geqslant 0$ for all ℓ. Otherwise, Lemma 9 becomes trivially true. Hence, we have

$$\begin{aligned}
\frac{\mathsf{ps}_0(\vec{\mathcal{Q}}, \vec{k})}{\mathsf{ps}_1(\vec{\mathcal{Q}}, \vec{k})} &\geqslant \prod_{\ell=1}^{q}\left(1 - \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(\alpha_\ell[\vec{s}],\beta_\ell[\vec{s}]),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N - 2q}\right) \\
&\geqslant 1 - \sum_{\ell=1}^{q} \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(\alpha_\ell[\vec{s}],\beta_\ell[\vec{s}]),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N - 2q} \qquad (12)
\end{aligned}$$

$$\geqslant 1 - \sum_{\ell=1}^{q} \sum_{1 \leqslant a \leqslant b \leqslant t} \mathsf{R}_{2a-1,2b,\ell}[\vec{s}] \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N-2q} \ , \tag{13}$$

where (12) holds since $(1-a)(1-b) \geqslant 1-a-b$ for any $a, b \geqslant 0$ and (13) holds because the indicator function R is non-negative and satisfies $\mathsf{R}_{\alpha[\vec{s}],\beta[\vec{s}],\ell}[\vec{s}] = 1$. We note that (13) is the exact quantity we pick for $1 - g(\vec{\mathcal{Q}}, \vec{k})$. □

Lemma 10. *If $q \leqslant N/4$, then*

$$\mathbb{E}_{\vec{k}} \left(\sum_{\ell=1}^{q} \sum_{1 \leqslant a \leqslant b \leqslant t} \mathsf{R}_{2a-1,2b,\ell}[\vec{s}] \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N-2q} \right) \leqslant t^2 \cdot \frac{(4q)^{t+1}}{N^t} \ .$$

Proof. By the sum of expectation and noting that none of $\sigma \in \mathcal{B}_{\mathcal{I}}(2a-1, 2b)$ would have $|\sigma| \geqslant 2$ if $a = b$, we have

$$\begin{aligned} &\mathbb{E}_{\vec{k}} \left(\sum_{\ell=1}^{q} \sum_{1 \leqslant a \leqslant b \leqslant t} \mathsf{R}_{2a-1,2b,\ell}[\vec{s}] \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma| \geqslant 2} \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N-2q} \right) \\ &= \sum_{\ell=1}^{q} \sum_{1 \leqslant a < b \leqslant t} \sum_{\sigma \in \mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma| \geqslant 2} \mathbb{E}_{\vec{s}} \left(\mathsf{R}_{2a-1,2b,\ell}[\vec{s}] \cdot \prod_{h=1}^{|\sigma|} \frac{Z_{\vec{s}}(i_{h-1}, i_h)}{N-2q} \right) . \end{aligned}$$

Hence, it is sufficient to derive bounds for each (a, b, σ). Note that for each a, b, the random variable $R_{2a-1,2b,j}[\vec{s}]$ depends only on the sub-keys $s_0, \ldots, s_{a-2}, s_{b+1}, \ldots, s_t$, which are $(a-2+1)+(t-(b+1)+1) = t-b+a-1$ sub-keys in total.

Next, given a fixed $\sigma = ((i_0, i_1), (i_1, i_2), \ldots, (i_{|\sigma|-1}, i_{|\sigma|}))$, we analyze the key dependency for each $Z_{\vec{s}}(i_{h-1}, i_h)$.

1. For (i_0, i_1), given that $i_0 = 2a-1$ is odd and i_1 is even, $Z_{\vec{s}}(i_0, i_1)$ depends on $(i_1 - i_0 - 1)/2$ subkeys between L_{i_0} and L_{i_1}.
2. For any (i_{h-1}, i_h) where $h > 1$, given i_{h-1} is an even number, implying that $L_{i_{h-1}}$ and $L_{i_{h-1}+1}$ are connected by “key-edges”, which always form a perfect matching regardless of the sub-key choice, then, the equality $Z_s(i_{h-1}, i_h) = Z_s(i_{h-1}+1, i_h)$ always holds. And we can see that $Z_s(i_{h-1}+1, i_h)$ depends only on $(i_h - i_{h-1} - 2)/2$ sub-keys.

Also note that the sets of dependent sub-keys for $Z_s(i_{h-1}, i_h)$ and $\mathsf{R}_{2a-1,2b,j}[s]$ are disjoint. After fixing (a, b, σ), putting the results together shows that the total number of sub-keys that each expectation term depends on is at most

$$\begin{aligned} \#\text{dependent subkeys} &= (t-b+a-1) + \frac{i_1 - i_0 - 1}{2} + \sum_{h=2}^{|\sigma|} \left(\frac{i_h - i_{h-1}}{2} - 1 \right) \\ &= (t-b+a-1) + \frac{\sum_{h=1}^{|\sigma|}(i_h - i_{h-1}) - 1}{2} - (|\sigma| - 1) \\ &= t - b + a - 1 + \frac{2b - 2a}{2} - |\sigma| + 1 \\ &= t - |\sigma| \leqslant t - 2 \ , \end{aligned}$$

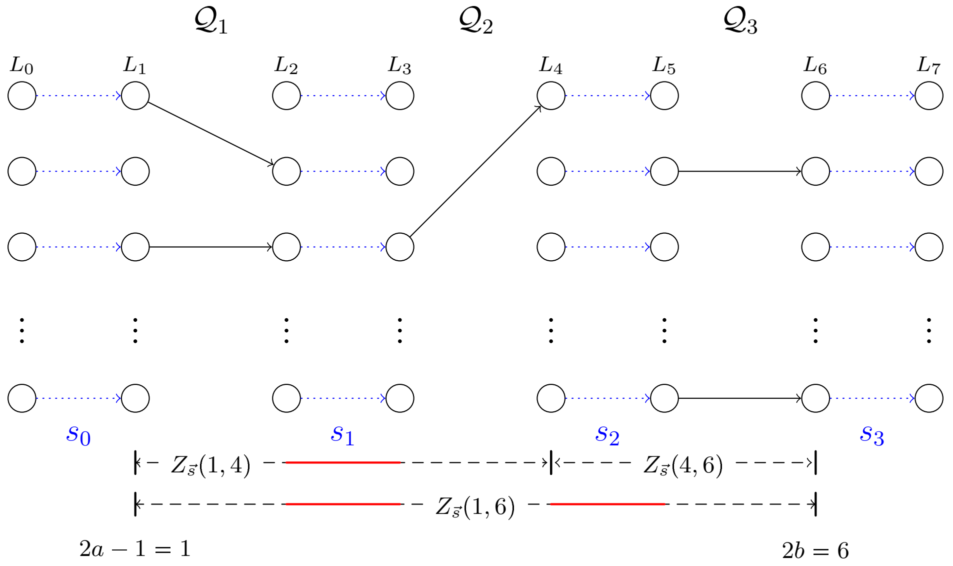

Fig. 1. A 3-round KAC with fixed query records $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3$. The sub-keys $\vec{s} = (s_0, \ldots, s_3)$ are random and to be sampled. The solid red line indicates that the $Z_{\vec{s}}(\mathsf{left}, \mathsf{right})$, which counts the number of paths from L_{left} to L_{right}, depends on the corresponding sub-keys. Consider $2a-1 = 1, 2b = 6$; then, $\mathsf{R}_{1,6,\ell}[\vec{s}] = 1$ and depends on $(a-1)+(3-b) = 0$ sub-keys because any s_0 allows x_ℓ from L_0 to reach L_1 and y_ℓ from L_7 to L_6. For $\sigma = ((1,6))$, the value of $Z_{\vec{s}}(1,6)$ depends on two sub-keys s_1, s_2. However, if the σ is further paritioned into $((1,4),(4,6))$, then $Z_{\vec{s}}(1,4)$ depends on s_1 but $Z_{\vec{s}}(4,6)$ does not depend on any sub-keys, because $Z_{\vec{s}}(4,6) = Z_{\vec{s}}(5,6) = |\mathcal{Q}_3|$.

where we observe that a summation term of (a, b, σ) depends on fewer sub-keys if the size of σ is larger (see Figure 1 for a specific case illustration). Because our construction ensures that any $t-2$ sub-keys are independently and uniformly distributed, the random variables in each expectation terms are mutually independent; hence, we can break the terms into

$$\begin{aligned}
&\mathbb{E}_{\vec{k}}\left(\sum_{\ell=1}^{q}\sum_{1\leqslant a\leqslant b\leqslant t}\mathsf{R}_{2a-1,2b,\ell}[\vec{s}]\sum_{\sigma\in\mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma|\geqslant 2}\ \prod_{h=1}^{|\sigma|}\frac{Z_{\vec{s}}(i_{h-1},i_h)}{N-2q}\right)\\
&\leqslant\sum_{\ell=1}^{q}\sum_{1\leqslant a<b\leqslant t}\ \sum_{\sigma\in\mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma|\geqslant 2}\mathbb{E}_{\vec{s}}\left(\mathsf{R}_{2a-1,2b,\ell}[\vec{s}]\cdot\prod_{h=1}^{|\sigma|}\frac{2Z_{\vec{s}}(i_{h-1},i_h)}{N}\right)\\
&=\sum_{\ell=1}^{q}\sum_{1\leqslant a<b\leqslant t}\ \sum_{\sigma\in\mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma|\geqslant 2}\mathbb{E}_{\vec{s}}\left(\mathsf{R}_{2a-1,2b,\ell}[\vec{s}]\right)\cdot\prod_{h=1}^{|\sigma|}\mathbb{E}_{s}\left(\frac{2Z_{\vec{s}}(i_{h-1},i_h)}{N}\right) &&(14)\\
&\leqslant\sum_{\ell=1}^{q}\sum_{1\leqslant a<b\leqslant t}\left(\frac{q}{N}\right)^{t-b+a-1}\sum_{\substack{\sigma\in\mathcal{B}_{\mathcal{I}}(2a-1,2b),\\ |\sigma|\geqslant 2}}\left(\frac{2q}{N}\right)^{(i_1-i_0+1)/2}\prod_{h=2}^{|\sigma|}\left(\frac{2q}{N}\right)^{(i_h-i_{h-1})/2} &&(15)\\
&=\sum_{\ell=1}^{q}\sum_{1\leqslant a<b\leqslant t}\left(\frac{q}{N}\right)^{t-b+a-1}\cdot\sum_{\sigma\in\mathcal{B}_{\mathcal{I}}(2a-1,2b),\ |\sigma|\geqslant 2}\left(\frac{2q}{N}\right)^{(2b-(2a-1)+1)/2}
\end{aligned}$$

$$\leqslant \sum_{\ell=1}^{q} \sum_{1\leqslant a<b\leqslant t} \left(\frac{q}{N}\right)^{t-b+a-1} \cdot \left(\frac{4q}{N}\right)^{b-a+1} \leqslant t^2 \cdot \frac{(4q)^{t+1}}{N^t} . \tag{16}$$

In the preceding calculation, (14) holds since the sub-keys are $(t-2)$-wise independent. The first "q/N" term of (15) comes from moving the $\mathbb{E}_{\vec{s}}\left(\mathsf{R}_{2a-1,2b,\ell}[\vec{s}]\right)$, and inside the summation the "$2q/N$" terms are the direct calculation upper bound of $\mathbb{E}_s\left(2Z_{\vec{s}}(i_{h-1}, i_h)/N\right)$ for each (i_{h-1}, i_h). Finally, the first inequality of (16) holds because the size of $\mathcal{B}_{\mathcal{I}}(2a-1, 2b)$ is upper-bounded by 2^{b-a}, which is absorbed into the "$2q/N$" term, yielding a "$4q/N$" term. □

6 Concluding the Proof

6.1 Proof of Theorem 1

Proof. We partition the set of transcripts $\mathcal{T} = \mathcal{T}_{\mathsf{good}} \sqcup \mathcal{T}_{\mathsf{bad}}$ according to Definition 1. By applying Lemma 1, we have $\varDelta(X_0, X_1) \leqslant \mathbb{E}_{X_1}[g(X_1)] + \mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}]$. We start with bounding the $\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}]$.

Claim.

$$\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] \leqslant (t+1) \cdot \frac{3q^{t+1}}{N^t} + 3(t+1)\sqrt{\frac{q^{2t-1}(t+2)n}{N^{2t-2}}} + \frac{t(t+1)}{N^t} .$$

Proof (of claim). We note that in the system $\mathbf{S}_1$, the set of bad keys $\mathsf{Badkey}_{\vec{\mathcal{Q}}}$ is defined only by the query records $\vec{\mathcal{Q}} = (\mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_t)$. Therefore, we have

$$\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] \leqslant \mathsf{Pr}_{\vec{\mathcal{Q}}}\left[|\mathsf{Badkey}_{\mathcal{Q}}| > C\right] + \frac{C}{N^{t-1}} .$$

To get the size bound for $\mathsf{Badkey}_{\mathcal{Q}}$, we compute the size of $\mathsf{Badkey}_{\mathcal{Q},i}$ for $0 \leqslant i \leqslant t$. Then, we have

$$\begin{aligned}
|\mathsf{Badkey}_{\vec{\mathcal{Q}},0}| &\leqslant \mu_{\vec{c}_0}(V_1, \mathcal{Q}_2, \mathcal{Q}_3, \ldots, \mathcal{Q}_{t-1}, \mathcal{Q}_t, U_{t+1}) \\
|\mathsf{Badkey}_{\vec{\mathcal{Q}},1}| &\leqslant \mu_{\vec{c}_1}(V_2, \mathcal{Q}_3, \mathcal{Q}_4, \ldots, \mathcal{Q}_t, \mathcal{Q}_E, U_1) \\
&\vdots \\
|\mathsf{Badkey}_{\vec{\mathcal{Q}},i}| &\leqslant \mu_{\vec{c}_i}(V_{i+1}, \mathcal{Q}_{i+2}, \mathcal{Q}_{i+3}, \ldots, \mathcal{Q}_t, \mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_{i-1}, U_i) \\
&\vdots \\
|\mathsf{Badkey}_{\vec{\mathcal{Q}},t-1}| &\leqslant \mu_{\vec{c}_{t-1}}(V_t, \mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-2}, U_{t-1}) \\
|\mathsf{Badkey}_{\vec{\mathcal{Q}},t}| &\leqslant \mu_{\vec{c}_t}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) .
\end{aligned}$$

where the linear coefficient tuples $\vec{c}_i$ are given by condition 2 of Theorem 1 so that two neighboring coefficients are non-zero and

$$\begin{aligned}
\forall i \in \{1, \ldots, t\} :\ U_i &= \{u \mid \exists v :\ (u,v) \in \mathcal{Q}_i\}, & V_i &= \{v \mid \exists u :\ (u,v) \in \mathcal{Q}_i\} \\
U_{t+1} &= \{u \mid \exists v :\ (u,v) \in \mathcal{Q}_E\}, & V_0 &= \{v \mid \exists u :\ (u,v) \in \mathcal{Q}_E\} .
\end{aligned}$$

The size of $\mathsf{Badkey}_{\vec{\mathcal{Q}},i}$ is bounded by $\mu_{\vec{c}_i}$ because any key $\vec{k} \in \mathsf{Badkey}_{\vec{\mathcal{Q}},i}$ is uniquely mapped to the sub-keys $(s_0, \ldots, s_{i-1}, s_{i+1}, s_t)$ since the linear mapping has rank $t-1$ (stated in condition 2 of Theorem 1).

We now can apply Lemma 2 to upper bound $\mathsf{Badkey}_{\vec{\mathcal{Q}},i}$ with high probability. For every i, by letting $C_i = \frac{3q^{t+1}}{N} + 3q^{t-1/2}\sqrt{(t+2)n}$, we obtain that $\mathsf{Pr}_{\mathcal{Q}}[|\mathsf{Badkey}_{\mathcal{Q},i}| > C_i] \leqslant \frac{2}{N^t}$. Therefore, setting $C = \sum_{i=0}^{t} C_i$, we have

$$\begin{aligned}\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] &\leqslant \mathsf{Pr}_{\vec{\mathcal{Q}}}\left[|\mathsf{Badkey}_{\mathcal{Q}}| > C\right] + \frac{C}{N^{t-1}} \\ &\leqslant \sum_{i=0}^{t} \mathsf{Pr}_{\vec{\mathcal{Q}}}\left[|\mathsf{Badkey}_{\mathcal{Q}_i}| > C_i\right] + \frac{C}{N^{t-1}} \\ &\leqslant \frac{2t(t+1)}{N^t} + (t+1)\cdot\frac{3q^{t+1}}{N^t} + 3(t+1)\cdot\frac{q^{t-1/2}\sqrt{(t+2)n}}{N^{t-1}} \ .\end{aligned}$$

Hence, we proved the claim. □

The next step is to pick a function g and upper bound $\mathbb{E}_{X_1}[g(X_1)]$. Note that by condition 1 of Theorem 1, any $t-2$ rows of key schedule matrix A have rank $t-2$, implying that any $t-2$ sub-keys are independent and uniform. Therefore, we can apply Lemma 7 and obtain a function g. Noting that because X_1 is sampled from the ideal world, $\vec{k}$ is sampled independently of $\vec{\mathcal{Q}}$. So we have

$$\mathbb{E}_{X_1}[g(X_1)] = \mathbb{E}_{\vec{\mathcal{Q}}}\mathbb{E}_{\vec{k}}[g(\vec{\mathcal{Q}},\vec{k})] \leqslant \mathbb{E}_{\vec{\mathcal{Q}}}\left[\frac{t^2(4q)^{t+1}}{N^t}\right] = \frac{t^2(4q)^{t+1}}{N^t} \ .$$

Then, by summing the two quantities and numerical simplifications, the theorem follows. □

6.2 Proof of Theorem 2

Proof. We again use Definition 1 for the bad transcripts. By applying Lemma 1, we obtain that $\Delta(X_0, X_1) \leqslant \mathbb{E}_{X_1}[g(X_1)] + \mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}]$, where we start with bounding the term $\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}]$. Again, we have $\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] \leqslant \mathsf{Pr}_{\vec{\mathcal{Q}}}\left[|\mathsf{Badkey}_{\mathcal{Q}}| > C\right] + \frac{C}{N^{t-2}}$. To get the size bound for $\mathsf{Badkey}_{\mathcal{Q}}$, we compute the size of $\mathsf{Badkey}_{\mathcal{Q},i}$ for $0 \leqslant i \leqslant t$. Then, we have

$$\begin{aligned}|\mathsf{Badkey}_{\vec{\mathcal{Q}},0}| &\leqslant \mu_{\vec{c}_0,\vec{d}_0}(V_1, \mathcal{Q}_2, \mathcal{Q}_3, \ldots, \mathcal{Q}_{t-1}, \mathcal{Q}_t, U_{t+1}) \\ |\mathsf{Badkey}_{\vec{\mathcal{Q}},1}| &\leqslant \mu_{\vec{c}_1,\vec{d}_1}(V_2, \mathcal{Q}_3, \mathcal{Q}_4, \ldots, \mathcal{Q}_t, \mathcal{Q}_E, U_1) \\ &\vdots \\ |\mathsf{Badkey}_{\vec{\mathcal{Q}},i}| &\leqslant \mu_{\vec{c}_i,\vec{d}_i}(V_{i+1}, \mathcal{Q}_{i+2}, \mathcal{Q}_{i+3}, \ldots, \mathcal{Q}_t, \mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_{i-1}, U_i) \\ &\vdots \\ |\mathsf{Badkey}_{\vec{\mathcal{Q}},t-1}| &\leqslant \mu_{\vec{c}_{t-1},\vec{d}_{t-1}}(V_t, \mathcal{Q}_E, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-2}, U_{t-1}) \\ |\mathsf{Badkey}_{\vec{\mathcal{Q}},t}| &\leqslant \mu_{\vec{c}_t,\vec{d}_t}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \ ,\end{aligned}$$

where

$$\begin{aligned}\forall i \in \{1,\ldots,t\}:\ U_i &= \{u \mid \exists v:\ (u,v) \in \mathcal{Q}_i\}, & V_i &= \{v \mid \exists u:\ (u,v) \in \mathcal{Q}_i\} \\ U_{t+1} &= \{u \mid \exists v:\ (u,v) \in \mathcal{Q}_E\}, & V_0 &= \{v \mid \exists u:\ (u,v) \in \mathcal{Q}_E\} \ .\end{aligned}$$

Since any $t-2$ rows of the key schedule matrix A have rank $t-2$, the size of $\mathsf{Badkey}_{\vec{\mathcal{Q}},i}$ is bounded by $\mu_{\vec{c}_i,\vec{d}_i}$, and we can always choose the sets of linear coefficients $\vec{c}_i = (c_{i,0}, \ldots, c_{i,t-1})$ and $\vec{d}_i = (d_{i,0}, \ldots, d_{i,t-1})$ that satisfy the requirements of Lemma 5.

se can now apply Lemma 5 to upper bound $\mathsf{Badkey}_{\vec{\mathcal{Q}},i}$ with high probability. For every i, by letting $C_i = \frac{q^{t+1}}{N^2} + t \cdot \frac{(3q)^{2t-3}}{N^{t-2}} + \frac{(3q)^{2t-2.5}}{N^{t-2}}$, we obtain that $\mathsf{Pr}_{\mathcal{Q}}[|\mathsf{Badkey}_{\mathcal{Q},i}| > C_i] \leqslant \frac{2t}{N^t}$. Therefore, setting $C = \sum_{i=0}^{t} C_i$, we directly obtain the result

$$\mathsf{Pr}[X_1 \in \mathcal{T}_{\mathsf{bad}}] \leqslant (t+1) \cdot \frac{q^{t+1} + 2t}{N^t} + t(t+1) \cdot \frac{(3q)^{2t-3}}{N^{2t-4}} + (t+1)\frac{(3q)^{2t-2.5}}{N^{2t-4}} .$$

For the good transcripts, noting that any $t-2$ rows of key schedule matrix A have rank $t-2$, implying that any subset of $t-2$ sub-keys is independent and uniform. Therefore, we apply Lemma 7 and obtain $\mathbb{E}_{X_1}[g(X_1)] \leqslant \frac{t^2(4q)^{t+1}}{N^t}$. By summing the two quantities and some numerical simplification, the theorem follows. □

7 Conclusion and Open Problems

In this paper, we provided key schedules of limited independence for t-round key-alternating ciphers achieving tight security. We proved that the t-round key-alternating cipher remains tightly secure for a class of $(t-1)$-wise independent sub-key distributions and, when $t \geqslant 8$, for $(t-2)$-wise sub-key distributions.

Though our result does not extends to $(t-2)$-wise independent sub-key distributions for $3 \leqslant t \leqslant 7$, we expect that a tighter analysis of the matrix 2-norm for the sum-capture quantity should prove the tightness result for $4 \leqslant t \leqslant 7$. Also, given that we resolved the good key analysis for $(t-2)$-wise independent sub-keys, it would be interesting to investigate new methods for bounding the bad keys and proving tight security of 3-round KACs with identical key schedules. Another rich research vein involves the study of whether the tightness result holds for $(t-3)$-wise distributions or beyond.

Acknowledgements

We thank the anonymous reviewers for sharing many helpful suggestions that improved this paper. Stefano Tessaro and Xihu Zhang were partially supported by NSF grants CNS-1930117 (CAREER), CNS1926324, CNS-2026774, a Sloan Research Fellowship, and a JP Morgan Faculty Award.

References

1. Noga Alon, Tali Kaufman, Michael Krivelevich, and Dana Ron. Testing triangle-freeness in general graphs. In *17th SODA*, pages 279–288. ACM-SIAM, January 2006.
2. László Babai. The fourier transform and equations over finite abelian groups: An introduction to the method of trigonometric sums. Lecture notes, 1989. Available at http://people.cs.uchicago.edu/~laci/reu02/fourier.pdf.
3. 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 Pascal Paillier and Ingrid Verbauwhede, editors, *CHES 2007*, volume 4727 of *LNCS*, pages 450–466. Springer, Heidelberg, September 2007.
4. Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, François-Xavier Standaert, John P. Steinberger, and Elmar Tischhauser. Key-alternating ciphers in a provable setting: Encryption using a small number of public permutations - (extended abstract). In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 45–62. Springer, Heidelberg, April 2012.

5. Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the two-round Even-Mansour cipher. In Juan A. Garay and Rosario Gennaro, editors, *CRYPTO 2014, Part I*, volume 8616 of *LNCS*, pages 39–56. Springer, Heidelberg, August 2014.
6. Shan Chen and John P. Steinberger. Tight security bounds for key-alternating ciphers. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 327–350. Springer, Heidelberg, May 2014.
7. Benoit Cogliati, Rodolphe Lampe, and Yannick Seurin. Tweaking Even-Mansour ciphers. In Rosario Gennaro and Matthew J. B. Robshaw, editors, *CRYPTO 2015, Part I*, volume 9215 of *LNCS*, pages 189–208. Springer, Heidelberg, August 2015.
8. Benoît Cogliati and Yannick Seurin. Beyond-birthday-bound security for tweakable Even-Mansour ciphers with linear tweak and key mixing. In Tetsu Iwata and Jung Hee Cheon, editors, *ASIACRYPT 2015, Part II*, volume 9453 of *LNCS*, pages 134–158. Springer, Heidelberg, November / December 2015.
9. Benoit Cogliati and Yannick Seurin. On the provable security of the iterated Even-Mansour cipher against related-key and chosen-key attacks. In Elisabeth Oswald and Marc Fischlin, editors, *EUROCRYPT 2015, Part I*, volume 9056 of *LNCS*, pages 584–613. Springer, Heidelberg, April 2015.
10. Joan Daemen and Vincent Rijmen. *The design of Rijndael*, volume 2. Springer, 2002.
11. Orr Dunkelman, Nathan Keller, and Adi Shamir. Minimalism in cryptography: The Even-Mansour scheme revisited. In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 336–354. Springer, Heidelberg, April 2012.
12. Avijit Dutta. Minimizing the two-round tweakable Even-Mansour cipher. In Shiho Moriai and Huaxiong Wang, editors, *ASIACRYPT 2020, Part I*, volume 12491 of *LNCS*, pages 601–629. Springer, Heidelberg, December 2020.
13. Shimon Even and Yishay Mansour. A construction of a cipher from a single pseudorandom permutation. *Journal of Cryptology*, 10(3):151–162, June 1997.
14. Jian Guo, Thomas Peyrin, Axel Poschmann, and Matthew J. B. Robshaw. The LED block cipher. In Bart Preneel and Tsuyoshi Takagi, editors, *CHES 2011*, volume 6917 of *LNCS*, pages 326–341. Springer, Heidelberg, September / October 2011.
15. Thomas P. Hayes. A large-deviation inequality for vector-valued martingales, 2003. Available at https://www.cs.unm.edu/~hayes/papers/VectorAzuma/VectorAzuma20050726.pdf.
16. Viet Tung Hoang and Stefano Tessaro. Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In Matthew Robshaw and Jonathan Katz, editors, *CRYPTO 2016, Part I*, volume 9814 of *LNCS*, pages 3–32. Springer, Heidelberg, August 2016.
17. Eike Kiltz, Krzysztof Pietrzak, and Mario Szegedy. Digital signatures with minimal overhead from indifferentiable random invertible functions. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part I*, volume 8042 of *LNCS*, pages 571–588. Springer, Heidelberg, August 2013.
18. Rodolphe Lampe, Jacques Patarin, and Yannick Seurin. An asymptotically tight security analysis of the iterated Even-Mansour cipher. In Xiaoyun Wang and Kazue Sako, editors, *ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 278–295. Springer, Heidelberg, December 2012.
19. Ueli M. Maurer. Indistinguishability of random systems. In Lars R. Knudsen, editor, *EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 110–132. Springer, Heidelberg, April / May 2002.
20. Alessandro Panconesi and Aravind Srinivasan. Fast randomized algorithms for distributed edge coloring (extended abstract). In Norman C. Hutchinson, editor, *11th ACM PODC*, pages 251–262. ACM, August 1992.
21. Jacques Patarin. The "coefficients H" technique (invited talk). In Roberto Maria Avanzi, Liam Keliher, and Francesco Sica, editors, *SAC 2008*, volume 5381 of *LNCS*, pages 328–345. Springer, Heidelberg, August 2009.
22. John Steinberger. Improved security bounds for key-alternating ciphers via hellinger distance. Cryptology ePrint Archive, Report 2012/481, 2012. http://eprint.iacr.org/2012/481.
23. John P. Steinberger. Counting solutions to additive equations in random sets. *CoRR*, abs/1309.5582, 2013.
24. Yusai Wu, Liqing Yu, Zhenfu Cao, and Xiaolei Dong. Tight security analysis of 3-round key-alternating cipher with a single permutation. In Shiho Moriai and Huaxiong Wang, editors, *ASIACRYPT 2020, Part I*, volume 12491 of *LNCS*, pages 662–693. Springer, Heidelberg, December 2020.

Appendix

A More Fine-grained Security of Theorems

A.1 Theorem 1

Here we provide a version of Theorem 1 with the number of construction queries q_e and the number of permutation queries q_p being separated. We only provide a proof sketch for the theorem and lemmas, since the techniques are essentially the same as we were proving for the case of $q_e = q_p$.

Theorem 3. *For the t-round KAC constructed over t random permutations $\vec{P} = (P_1, \ldots, P_t)$, let the key of KAC be $\vec{k} = (k_0, k_1, \ldots, k_{t-2})^\top$ in which k_i's are independently uniformly sampled from $\mathbb{F}_{2^n}$. Let subkeys $\vec{s} = (s_0, s_1, \ldots, s_t)^\top$ be derived by $\vec{s} \leftarrow A\vec{k}$ in which A is a $(t+1) \times (t-1)$ matrix over $\mathbb{F}_{2^n}$, with the rows denoted as $A_0, \ldots, A_t$, such that*

1. *Any $t-2$ rows of A forms a matrix of rank $t-2$.*
2. *For any $I \subseteq \{0, \ldots, t\}$, $|I| = t$, then the row vectors $(A_\ell)_{\ell \in I}$ satisfy that*
   - *$(A_\ell)_{\ell \in I}$ forms a matrix of rank $t-1$.*
   - *there exists values $(c_\ell)_{\ell \in I}$ such that $\sum_{\ell \in I} c_\ell A_\ell = \vec{0}$ and there are two indices $\mathsf{idx}_1, \mathsf{idx}_2 \in I$ satisfying $\mathsf{idx}_1 - \mathsf{idx}_2 \in \{1, t\}$ and $c_{\mathsf{idx}_1}, c_{\mathsf{idx}_2}$ are both non-zero.*

Then for any adversary $\mathcal{A}$ that issues at most q_e queries to KAC, and at most q_p queries to $P_1, \ldots, P_t$, in which $9(t+2)n \leqslant q_e, q_p \leqslant N/4$

a. if $q_e \leqslant q_p$, then

$$\mathsf{Adv}^{\pm\mathsf{prp}}_{\mathsf{KAC}[\vec{P}]}(\mathcal{A}) \leqslant (t+1) \cdot \frac{4q_e^{1/2} q_p^{t+1/2}}{N^t} + 3(t+1)\sqrt{\frac{q_e q_p^{2t-2}}{N^{2t-2}} \cdot (nt^2 + 2nt)} + t^2 \cdot \frac{4q_e \cdot (4q_p)^t}{N^t} .$$

b. if $q_e \geqslant q_p$, then

$$\mathsf{Adv}^{\pm\mathsf{prp}}_{\mathsf{KAC}[\vec{P}]}(\mathcal{A}) \leqslant (t+1) \cdot \frac{4q_e^2 q_p^{t-1}}{N^t} + 3(t+1)\sqrt{\frac{q_e^2 q_p^{2t-3}}{N^{2t-2}} \cdot (nt^2 + 2nt)} + t^2 \cdot \frac{4q_e \cdot (4q_p)^t}{N^t} .$$

We note that both cases come from the same proof but are manipulated in favor of a unified representation for the class of defined KACs. When $q_e = q_p$, both cases provide tight bound. However, we stress that when $q_e \neq q_p$, both bounds are not necessarily the best possible bound. To prove a better bound, one may want to further focus on a specific key schedule instead of a class of key schedule algorithms.

Proof (sketch). We still follow the definition of bad transcripts as we were proving Theorem 1. The good transcript analysis follows the proof of Lemma 7 identically and contributes the term $t^2 \cdot \frac{(4q_e)\cdot(4q_p)^t}{N^t}$.

Here we state a more fine-grained version of sum capture theorem, with all queries are separated. We note that the proof is identical to the proof of Lemma 2.

Lemma 11. *Let $t \geqslant 2$. Let $P_1, \ldots, P_{t-1}$ be $t-1$ independent random permutations of $\{0,1\}^n$, and let $\mathcal{A}$ be a probabilistic algorithm that makes adaptive queries to $P_1, \ldots, P_{t-1}$. Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}$ be the query transcripts of $P_1, \ldots, P_{t-1}$ interacting with $\mathcal{A}$. Let $\vec{c} = (c_0, \ldots, c_{t-1})$ be the coefficients so that there is an index $\mathsf{idx} \in \{0, \ldots, t-2\}$ satisfying $c_{\mathsf{idx}} \neq 0$ and $c_{\mathsf{idx}+1} \neq 0$. Let left be the smallest index so that $c_{\mathsf{left}} \neq 0$ and let right be the largest index such that $c_{\mathsf{right}} \neq 0$. Let $q_0, \ldots, q_t$ be positive integers such that $9(t+2)n \leqslant q_i \leqslant N/2$ for every $0 \leqslant i \leqslant t$, then for any $\mathcal{A}$ that makes at most q_i queries to P_i,*

$$\mathsf{Pr}_{P_1,\ldots,P_{t-1}}\Bigg[\exists V_0, U_t \subseteq \mathbb{F}_{2^n}, |V_0| = q_0, |U_t| = q_t, \mu_{\vec{c}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \geqslant \frac{q_0 q_1 \cdots q_{t-1} q_t}{N} + \\ + \left(\sqrt{q_{\mathsf{left}} q_{\mathsf{right}+1}} \cdot \prod_{i \in \{0,\ldots,t\} \setminus \{\mathsf{left}, \mathsf{right}+1, \mathsf{idx}+1\}} q_i\right) \cdot \left(\frac{2q_{\mathsf{idx}+1}^2}{N} + 3\sqrt{(t+2)nq_{\mathsf{idx}+1}}\right)\Bigg] \leqslant \frac{2t}{N^t} \,.$$

We again upper bound the size of bad key sets by upper bounding their corresponding sum capture quantity. Note that when applying Lemma 11, we have exactly one quantity in $\{q_0, \ldots, q_t\}$ being q_e and the rest quantities being q_p.

When $q_e \leqslant q_p$, the above sum capture quantity is maximized if we have $q_{\mathsf{left}} = q_e$. When $q_e \geqslant q_p$, we break the second additive terms of above quantity into two terms, i.e., the term that multiplied by $2q_{\mathsf{idx}+1}^2/N$ and the term multiplied by $3\sqrt{(t+2)nq_{\mathsf{idx}+1}}$. Note that the first term obtains maximum value if $q_{\mathsf{idx}+1} = q_e$ and the second term obtains maximum value if $q_i = q_e$ for some $i \in \{0, \ldots, t\} \setminus \{\mathsf{left}, \mathsf{right}+1, \mathsf{idx}+1\}$. □

A.2 Theorem 2

Here we only provide a bound for the case of $q_e \leqslant q_p$, which is a reasonable assumption.

Theorem 4. *For the t-round KAC constructed over t random permutations $\vec{P} = (P_1, \ldots, P_t)$, let the key of KAC be $\vec{k} = (k_0, k_1, \ldots, k_{t-3})^\top$ in which k_i's are independently and uniformly sampled from $\mathbb{F}_{2^n}$. Let subkeys $\vec{s} = (s_0, s_1, \ldots, s_t)^\top$ be derived by $\vec{s} = A\vec{k}$ in which A is a $(t+1) \times (t-2)$ matrix over $\mathbb{F}_{2^n}$ such that any $t-2$ rows of A forms a matrix of rank $t-2$. Then for any adversary $\mathcal{A}$ that issues at most q_e queries to KAC and q_p queries to $P_1, \ldots, P_t$,*

a. if $(t+2)nN^{2/3} \leqslant q_p \leqslant N/4$, then

$$\mathsf{Adv}^{\pm\mathsf{prp}}_{\mathsf{KAC}[\vec{P}]}(\mathcal{A}) \leqslant (t+1)^2 \cdot \frac{4q_e \cdot (4q_p)^t}{N^t} + (t+1)^2 \cdot \frac{\sqrt{q_e} \cdot (3q_p)^{2t-3}}{N^{2t-4}} \,.$$

b. if $9(t+2)n \leqslant q_p \leqslant (t+2)nN^{2/3}$,

$$\mathsf{Adv}^{\pm\mathsf{prp}}_{\mathsf{KAC}[\vec{P}]}(\mathcal{A}) \leqslant (t+1)^2 \cdot \frac{4q_e \cdot (4q_p)^t}{N^t} + t(t+1)\sqrt{\frac{q_e q_p^5}{N^{t+1}}} \cdot \left(5(t+2)^2 n^2 \cdot \sqrt{\frac{q_p}{N}}\right)^{t-3} \\ + (t+1)\sqrt{\frac{q_e q_p^2}{N^{t-2}}} \cdot \left(5(t+2)^2 n^2 \cdot \sqrt{\frac{q_p}{N}}\right)^{t-2} \,.$$

We state the following sum capture lemma for proving the above theorem.

Lemma 12. *Let $t \geqslant 2$. Let $P_1, \ldots, P_{t-1}$ be $t-1$ independent random permutations of $\{0,1\}^n$, and let $\mathcal{A}$ be a probabilistic algorithm that makes adaptive queries to $P_1, \ldots, P_{t-1}$. Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}$ be the query transcripts of $P_1, \ldots, P_{t-1}$ interacting with $\mathcal{A}$. Let $\vec{c}, \vec{d}$ be the coefficients as in Lemma 5. Let $q_0, \ldots, q_t$ be positive integers exactly one of $\{q_0, \ldots, q_t\}$ equals q_e and the rest equal q_p, where $9(t+2)n \leqslant q_p \leqslant N/2$ and $q_e \leqslant q_p$, then for any $\mathcal{A}$ that makes at most q_i queries to P_i,*

$$\mathsf{Pr}_{P_1,\ldots,P_{t-1}}\left[\exists V_0, U_t \subseteq \mathbb{F}_{2^n}, |V_0| = q_0, |U_t| = q_t, \mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \geqslant \frac{q_e q_p^t}{N^2} + \right.$$
$$\left. + t \cdot \frac{q_e^{1/2} q_p^{5/2}}{N} \cdot \left(\frac{2q_p^2}{N} + 3\sqrt{(t+2)nq_p}\right)^{t-3} + q_e^{1/2} q_p \left(\frac{2q_p^2}{N} + 3\sqrt{(t+2)nq_p}\right)^{t-2}\right] \leqslant \frac{2t}{N^t}\ .$$

Proof (sketch). The general proof paradigm follows the proof of Lemma 6. We let the i-th case to be that $q_i = q_e$ and $q_j = q_p$ for all $j \neq i$ and proceed with the proof as in Lemma 6. However, we note that the worst bound comes from the case when $q_0 = q_e$ or $q_t = q_e$.

We again perform Fourier transform and simplification. After obtaining the summation over Fourier terms, we break the term μ into three cases and show the following propositions. We omit the proofs of propositions since most of them are case discussions and they are essentially following their counterpart in Lemma 6 but with some minor numerical manipulations.

Proposition 7.

$$\mu_1 = \frac{q_e q_p^t}{N^2}\ .$$

Proposition 8. *Conditioning on $q_e \leqslant q_p$ and all P_i satisfying $\Phi(P_i) \leqslant 2q_i^2/N + 3\sqrt{(t+2)nq_i}$,*

$$\mu_2 \leqslant t \cdot \frac{q_e^{1/2} q_p^{5/2}}{N} \cdot \left(\frac{2q_p^2}{N} + 3\sqrt{(t+2)nq_p}\right)^{t-3}\ .$$

Proposition 9. *Conditioning on $q_e \leqslant q_p$ and all P_i satisfying $\Phi(P_i) \leqslant 2q_i^2/N + 3\sqrt{(t+2)nq_i}$,*

$$\mu_3 \leqslant q_e^{1/2} q_p \left(\frac{2q_p^2}{N} + 3\sqrt{(t+2)nq_p}\right)^{t-2}\ .$$

B Omitted Proofs for Sum Capture Quantities

B.1 Missing Cases Proof of Lemma 3

In this part we provide the complete case discussions for the proof of Lemma 3. Recall that

$$\mathsf{left} := \text{min of } i \text{ such that } c_i \neq 0$$
$$\mathsf{right} := \text{max of } i \text{ such that } c_i \neq 0\ ,$$

then we have $\mathsf{left} \leqslant \mathsf{idx}$ and $\mathsf{right} \geqslant \mathsf{idx}+1$ given $c_{\mathsf{idx}} \neq 0$ and $c_{\mathsf{idx}+1} \neq 0$.

- **Case 2.** $\mathsf{left} > 0, \mathsf{right} < t-1$.

$$
\begin{aligned}
&\mu(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) - \frac{q^{t+1}}{N} \\
&\leqslant N^{2t-1} \sum_{\gamma \neq 0} |\widehat{\mathbb{1}_{V_0}}(A_{c_0}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{Q_1}}(A_{c_0}^\top \gamma, A_{c_1}^\top \gamma)| \cdots |\widehat{\mathbb{1}_{Q_{t-1}}}(A_{c_{t-2}}^\top \gamma, A_{c_{t-1}}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{U_t}}(A_{c_{t-1}}^\top \gamma)| \\
&\leqslant N^{2t-3} \sum_{\gamma \neq 0} \frac{q}{N} \cdot \left(\frac{q}{N^2}\right)^{\mathsf{left}-1} \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{left}}}}(0, A_{c_{\mathsf{left}}}^\top \gamma)| \cdot \left(\frac{q}{N^2}\right)^{\mathsf{right}-\mathsf{left}-1} \cdot \varPhi(\mathcal{Q}_{\mathsf{idx}+1}) \\
&\quad \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(A_{c_{\mathsf{right}}}^\top \gamma, 0)| \cdot \left(\frac{q}{N^2}\right)^{t-2-\mathsf{right}} \cdot \frac{q}{N} \\
&= q^{t-2} N^3 \varPhi(\mathcal{Q}_{\mathsf{idx}+1}) \cdot \sum_{\gamma \neq 0} |\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{left}}}}(0, A_{c_{\mathsf{left}}}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(A_{c_{\mathsf{right}}}^\top \gamma, 0)|
\end{aligned}
$$

To this point, it is unclear how to directly apply the Cauchy-Schwartz trick, however, a few additional steps can help us get to similar calculations in Case 1.

We let $V' = \{v \mid \exists u : \ (u, v) \in \mathcal{Q}_{\mathsf{left}}\}$, $U' = \{u \mid \exists v : \ (u, v) \in \mathcal{Q}_{\mathsf{right}+1}\}$.

Claim. For any $\alpha \in \mathbb{F}_{2^n}$,

$$
\widehat{\mathbb{1}_{V'}}(\alpha) = \frac{1}{N} \widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{left}}}}(0, \alpha) , \quad \widehat{\mathbb{1}_{U'}}(\alpha) = \frac{1}{N} \widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(\alpha, 0) .
$$

Proof (of the claim). Because the proof for both equalities are similar, we only prove the first equality.

By the definition of Fourier transform, we have

$$
\begin{aligned}
\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{left}}}}(0, \alpha) &= \frac{1}{N^2} \sum_{u, v \in \mathbb{F}_{2^n}} \mathbb{1}_{\mathcal{Q}_{\mathsf{left}}}(u, v) (-1)^{\langle 0, u \rangle + \langle \alpha, v \rangle} \\
&= \frac{1}{N^2} \sum_{u, v \in \mathbb{F}_{2^n}} \mathbb{1}_{\mathcal{Q}_{\mathsf{left}}}(u, v) (-1)^{\langle \alpha, v \rangle} \\
&= \frac{1}{N} \cdot \frac{1}{N} \sum_{v \in \mathbb{F}_{2^n}} \mathbb{1}_{V'}(v) (-1)^{\langle \alpha, v \rangle} = \frac{1}{N} \widehat{\mathbb{1}_{V'}}(\alpha)
\end{aligned}
$$

□

Therefore, by applying the claim, we have

$$
\begin{aligned}
&\mu(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) - \frac{q^{t+1}}{N} \\
&\leqslant q^{t-2} N \varPhi(\mathcal{Q}_{\mathsf{idx}+1}) \cdot \sum_{\gamma} |\widehat{\mathbb{1}_{U'}}(A_{c_{\mathsf{left}}}^\top \gamma)| \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(A_{c_{\mathsf{right}}}^\top \gamma, 0)| \\
&\leqslant q^{t-1} \varPhi(\mathcal{Q}_{\mathsf{idx}+1})
\end{aligned}
$$

- **Case 3.** $\mathsf{left} = 0, \mathsf{right} < t - 1$.

In this case we have

$$
\mu(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) - \frac{q^{t+1}}{N}
$$

$$\leqslant N^{2t-3}\sum_{\gamma}|\widehat{\mathbb{1}_{V_0}}(A_{c_0}^{\top}\gamma)|\cdot\left(\frac{q}{N^2}\right)^{\mathsf{right}-1}\cdot\Phi(\mathcal{Q}_{\mathsf{idx}+1})\cdot|\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(A_{c_{\mathsf{right}}}^{\top}\gamma,0)|\cdot\left(\frac{q}{N^2}\right)^{t-2-\mathsf{right}}\cdot\frac{q}{N}$$
$$= q^{t-2}N^2\Phi(\mathcal{Q}_{\mathsf{idx}+1})\cdot\sum_{\gamma}|\widehat{\mathbb{1}_{V_0}}(A_{c_0}^{\top}\gamma)|\cdot|\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(A_{c_{\mathsf{right}}}^{\top}\gamma,0)|$$

Let $U' = \{u \mid \exists v :\ (u,v) \in \mathcal{Q}_{\mathsf{right}+1}\}$, then we have $\widehat{\mathbb{1}_{U'}}(\alpha) = \frac{1}{N}\widehat{\mathbb{1}_{\mathcal{Q}_{\mathsf{right}+1}}}(\alpha,0)$ for any α. Hence we get

$$\begin{aligned}&\mu(V_0,\mathcal{Q}_1,\ldots,\mathcal{Q}_{t-1},U_t) - \frac{q^{t+1}}{N}\\ &\leqslant q^{t-2}N\Phi(\mathcal{Q}_{\mathsf{idx}+1})\cdot\sum_{\gamma}|\widehat{\mathbb{1}_{V_0}}(A_{c_0}^{\top}\gamma)|\cdot|\widehat{\mathbb{1}_{\mathcal{U}'}}(A_{c_{\mathsf{right}}}^{\top}\gamma)|\\ &\leqslant q^{t-1}\Phi(\mathcal{Q}_{\mathsf{idx}+1})\end{aligned}$$

- **Case 4.** $\mathsf{left} > 0, \mathsf{right} = t-1$.

The proof for **Case 4** is a symmetric case of **Case 3**.

B.2 Proof of Proposition 1

Proof. Let

$$\ell := \text{min of } j \text{ such that } c_j \neq 0\ .$$

Therefore we have $\ell \leqslant i$. We define $\mathcal{A}$ to be the following algorithm.

1. $\mathcal{A}$ queries P_ℓ in the backward direction at points $c_\ell^{-1}\cdot X$ for all $0 \leqslant X < q$ (or sets V_0 be the corresponding queried points if $\ell = 0$)
2. $\mathcal{A}$ queries P_{i+1} in the forward direction at points $c_i^{-1}\cdot X$ for all $0 \leqslant X < q$.
3. $\mathcal{A}$ make q arbitrary queries to each P_j for $j \notin \{\ell, i+1\}$.

then we are left to count the number of tuples $(v_0,(u_1,v_1),\ldots,(u_{t-1},v_{t-1}),u_t)$ that satisfies $\sum_{j=0}^{t-1}c_j(v_j+u_{j+1}) = 0$. Here we consider the case that $\ell > 0$ as similar analysis applies to $\ell = 0$. The equality constraint can be rearranged as

$$c_{t-1}u_t + \sum_{j\in\{\ell+1,\ldots,t-1\},j\neq i+1}(c_{j-1}u_j + c_jv_j) = c_\ell v_\ell + c_iu_{i+1} + c_{i+1}v_{i+1} \tag{17}$$

in which we drop all terms $c_i = 0$ for $i < \ell$ per definition of ℓ.

Now we leave $(u_\ell, v_\ell) \in \mathcal{Q}_\ell$ and $(u_{i+1}, v_{i+1}) \in \mathcal{Q}_{i+1}$ undecided and pick other tuples arbitrarily, giving us q^{t-1} choices. This corresponds to exactly the LHS of (17) being fixed and the RHS of (17) being undecided. Then note that by the specification of $\mathcal{A}$, we have $0 \leqslant c_iu_{i+1} < q$ and $0 \leqslant c_\ell v_\ell < q$, so to let (17) hold, we need to pick (u_{i+1}, v_{i+1}) so that the higher $n - \log q$ bits of $c_{i+1}v_{i+1}$ equals the LHS.

To analyze the number of feasible v_{i+1}'s, note that if we group v_{i+1}'s by their higher $n - \log q$ bits of $c_{i+1}v_{i+1}$ into N/q groups, then one can show that in expectation each group receives q^2/N of v_{i+1}'s. For moderately large q, if v_{i+1}'s are sampled with replacement, then a direct Chernoff bound can show that every group has the number of v_{i+1}'s that are close to q^2/N with overwhelming probability. However, given v_{i+1}'s are sampled without replacement, we use the following lemma by Panconesi and Srinivasan in [20], where they observed that Chernoff bound also holds for negatively correlated variables.

Lemma 13 **([20]).** *Let $X_1, \dots, X_m$ be random variables taking values in $\{0,1\}$ such that for any set $S \subseteq \{1, \dots, m\}$, $\mathsf{Pr}[\bigwedge_{i \in S} X_i = 1] \leqslant \prod_{i \in S} \mathsf{Pr}[X_i = 1]$, let $X = \sum_{i=1}^{m} X_i$ and let $\eta = \mathbb{E}[X]$, then for any $0 < \delta < 1$*

$$\mathsf{Pr}[X < (1-\delta)\eta] \leqslant e^{-\delta^2 \eta/2} .$$

Note that in our case, we fix an $r \in \{0,1\}^{n - \log q}$ and let $X_j = 1$ if the higher $n - \log q$ bits of the j-th sampled v_{i+1} has its $c_{i+1}v_{i+1}$ equalled r, otherwise $X_j = 0$. Then we have

$$\mathsf{Pr}[\bigwedge_{i \in S} X_i = 1] = q^{(|S|)}/N^{(|S|)} \leqslant (q/N)^{|S|} = \prod_{i \in S} \mathsf{Pr}[X_i = 1] .$$

Applying Lemma 13 with $\delta = 1/2$, we have $\mathsf{Pr}[X < q^2/2N] \leqslant e^{-q^2/8N}$. Finally taking a union bound over all r's, we proved that every group of v_{i+1}'s has at least $q^2/2N$ elements with probability at least $1 - (N/q) \cdot e^{-q^2/8N}$.

Hence, for the q^{t-1} choices of tuples, we can pick at least $q^2/2N$ of (u_{i+1}, v_{i+1}) that cancels the higher $n - \log q$ bits of the LHS value of (17). The remaining value is in $\{0, \dots, q-1\}$ and hence can be cancelled to zero by a unique choice of v_ℓ. □

B.3 Proof of Proposition 2

Proof. For any i that has $c_i \neq 0$, we let $\mathcal{A}$ query P_i in the backward direction at points $c_i^{-1} \cdot X$ for all $0 \leqslant X < q$ (or we set V_0 be the corresponding queried points if $i = 0$) and we let $\mathcal{A}$ query P_{i+1} in the forward direction at points $c_i^{-1} \cdot X$ for all $0 \leqslant X < q$ (or we set U_t be the corresponding queried points if $i = t-1$). Finally $\mathcal{A}$ makes arbitrary q queries to any permutations that were never queried (or pick arbitrary q points for V_0, U_t if the points were not set). We note that each permutation receives exactly q queries because for every $1 \leqslant i \leqslant t-1$, either $c_{i-1} = 0$ or $c_i = 0$, so the permutation P_i can only receive either q forward queries or q backward queries.

If all $c_i = 0$, then obviously $\mu_{\vec{c}}(V_0, \mathcal{Q}_1, \dots, \mathcal{Q}_{t-1}, U_t) = q^{t+1}$. Otherwise, we can rewrite the linear constraint as

$$\sum_i c_i(v_i + u_{i+1}) = \sum_{i\colon c_i \neq 0} (c_i v_i + c_i u_{i+1}) = 0$$

in which for every i satisfying $c_i \neq 0$, we have $0 \leqslant c_i v_i < q$ and $0 \leqslant c_i u_{i+1} < q$. So, we can fix an index j such that $c_j \neq 0$ and leave the tuple (u_j, v_j) undecided (or v_0 undecided if $j = 0$). Then we have q^t choices over the other t tuples, and the equality constraint $\sum_{i\colon c_i \neq 0}(c_i v_i + c_i u_{i+1}) = 0$ uniquely fix a feasible v_j so that $0 \leqslant c_j v_j = \sum_{i \neq j\colon c_i \neq 0}(c_i v_i + c_i u_{i+1}) + c_j u_{j+1} < q$. Hence we have q^t feasible choice of tuples in total, giving a lower bound for the sum capture quantity.

B.4 A Reduction-based Bound for 2-constraint Sum Capture Quantity

Proposition 10. *With the same setting as in Lemma 5, for any $\mathcal{A}$ that makes at most $q \geqslant (t+2)nN^{2/3}$ queries to each permutations,*

$$\mathsf{Pr}_{P_1,\dots,P_{t-1}}\left[\exists V_0, U_t \subseteq \mathbb{F}_{2^n}, |V_0| = |U_t| = q, \mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \dots, \mathcal{Q}_{t-1}, U_t) \geqslant \frac{6q^t}{N}\right] \leqslant \frac{2(t-1)}{N^{t-1}} .$$

Proof (of Proposition 10). Given the 2-constraint sum capture quantity with the specific coefficient $\vec{c}, \vec{d}$, We show that one can use 1-constraint sum capture quantity to derive an upper bound for $\mu_{\vec{c},\vec{d}}$. Let $U_{t-1} := \{u \mid \exists v :\ (u, v) \in \mathcal{Q}_{t-1}\}$ and $c' = (c_0, \ldots, c_{t-2})$, then we show that

$$\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \leqslant \mu_{\vec{c'}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-2}, U_{t-1}) \ .$$

Given a fix tuple $(v_0, (u_1, v_1), \ldots, (u_{t-2}, v_{t-2}), (u_{t-1}, v_{t-1}), u_t) \in V_0 \times \mathcal{Q}_1 \times \cdots , \mathcal{Q}_{t-1} \times U_t$ satisfying $\sum_{j=0}^{t-1} c_j(v_j + u_{j+1}) = 0$ and $\sum_{j=0}^{t-1} d_j(v_j + u_{j+1}) = 0$, one can map the tuple to $(v_0, (u_1, v_1), \ldots, (u_{t-2}, v_{t-2}), u_{t-1}) \in V_0 \times \mathcal{Q}_1 \times \cdots , \mathcal{Q}_{t-2} \times U_{t-1}$ satisfying $\sum_{j=0}^{t-2} c_j(v_j + u_{j+1}) = 0$ by dropping the term v_{t-1} and u_t (note that $c_{t-1} = 0$). This mapping is injective because v_{t-1} can be deterministically derived given $u_{t-1} \in U_{t-1}$ and $\mathcal{Q}_{t-1}$, and u_t can be calculated deterministically via $d_{t-1}(u_t + v_{t-1}) = \sum_{j=0}^{t-2} d_j(v_j + u_{j+1})$ given $d_1, d_2, \ldots, d_{t-1}$ are all non-zero. Hence the inequality holds.

Then the proposition follows by applying Lemma 2 to upper bound $\mu_{\vec{c'}}$ and noting that $q \geqslant (t+2)nN^{2/3}$. □

B.5 Omitted Calculation of $\mu_{\vec{c},\vec{d}}$ in Lemma 6

Note that $c_0 = d_{t-1} = 1$ and $c_{t-1} = d_0 = 0$, we can write $\mu_{\vec{c},\vec{d}}$ as

$$\begin{aligned}
&\mu_{\vec{c},\vec{d}}(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) \\
&= \sum_{v_0, u_1, v_1, \ldots, u_{t-1}, v_{t-1}, u_t} \mathbb{1}_{V_0}(v_0) \mathbb{1}_{\mathcal{Q}_1}(u_1, v_1) \cdots \mathbb{1}_{\mathcal{Q}_{t-1}}(u_{t-1}, v_{t-1}) \mathbb{1}_{U_t}(u_t) \\
&\quad \cdot \mathbb{1}_{\mathsf{eq}}\left(v_0 + u_1, \sum_{i=1}^{t-2} c_i(v_i + u_{i+1})\right) \mathbb{1}_{\mathsf{eq}}\left(u_t + v_{t-1}, \sum_{i=1}^{t-2} d_i(v_i + u_{i+1})\right) \\
&= \sum_{v_0, u_1, v_1, \ldots, u_{t-1}, v_{t-1}, u_t} \left(\sum_{\beta_0} \widehat{\mathbb{1}_{V_0}}(\beta_0)(-1)^{\langle \beta_0, v_0 \rangle}\right) \left(\sum_{\alpha_1, \beta_1} \widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha_1, \beta_1)(-1)^{\langle \alpha_1, u_1 \rangle + \langle \beta_1, v_1 \rangle}\right) \\
&\quad \cdots \left(\sum_{\alpha_{t-1}, \beta_{t-1}} \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\alpha_{t-1}, \beta_{t-1})(-1)^{\langle \alpha_{t-1}, u_{t-1} \rangle + \langle \beta_{t-1}, v_{t-1} \rangle}\right) \left(\sum_{\alpha_t} \widehat{\mathbb{1}_{U_t}}(\alpha_t)(-1)^{\langle \alpha_t, u_t \rangle}\right) \\
&\quad \cdot \left(\frac{1}{N} \sum_{\gamma} (-1)^{\langle \gamma, v_0 + u_1 + \sum_{i=1}^{t-2} c_i(v_i + u_{i+1}) \rangle}\right) \left(\frac{1}{N} \sum_{\delta} (-1)^{\langle \delta, u_t + v_{t-1} + \sum_{i=1}^{t-2} d_i(v_i + u_{i+1}) \rangle}\right) \\
&= \frac{1}{N^2} \sum_{\beta_0, \alpha_1, \beta_1, \ldots, \alpha_{t-1}, \beta_{t-1}, \alpha_t} \widehat{\mathbb{1}_{V_0}}(\beta_0) \widehat{\mathbb{1}_{\mathcal{Q}_1}}(\alpha_1, \beta_1) \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\alpha_{t-1}, \beta_{t-1}) \widehat{\mathbb{1}_{U_t}}(\alpha_t) \\
&\quad \cdot \left(\sum_{v_0} (-1)^{\langle \beta_0, v_0 \rangle + \langle \gamma, v_0 \rangle}\right) \left(\sum_{u_1} (-1)^{\langle \alpha_1, u_1 \rangle + \langle \gamma, v_1 \rangle}\right) \left(\sum_{v_1} (-1)^{\langle \beta_1, v_1 \rangle + \langle \gamma, c_1 v_1 \rangle + \langle \delta, d_1 v_1 \rangle}\right) \\
&\quad \cdots \left(\sum_{u_{t-1}} (-1)^{\langle \alpha_{t-1}, u_{t-1} \rangle + \langle \gamma, c_{t-2} u_{t-1} \rangle + \langle \delta, d_{t-2} u_{t-1} \rangle}\right) \left(\sum_{v_{t-1}} (-1)^{\langle \beta_{t-1}, v_{t-1} \rangle + \langle \delta, v_{t-1} \rangle}\right) \\
&\quad \cdot \left(\sum_{u_t} (-1)^{\langle \alpha_t, u_t \rangle + \langle \delta, u_t \rangle}\right)
\end{aligned}$$

Here the set of Fourier coefficients that give non-zero sum can be represented in terms of (γ, δ) as

$$\begin{cases} \alpha_1 = \beta_0 = \gamma \\ \alpha_2 = \beta_1 = A_{c_1}^\top \gamma + A_{d_1}^\top \delta \\ \qquad\vdots \\ \alpha_{i+1} = \beta_i = A_{c_i}^\top \gamma + A_{d_i}^\top \delta \\ \qquad\vdots \\ \alpha_{t-1} = \beta_{t-2} = A_{c_{t-2}}^\top \gamma + A_{d_{t-2}}^\top \delta \\ \alpha_t = \beta_{t-1} = \delta \end{cases} .$$

So we obtain

$$\mu(V_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t-1}, U_t) = N^{2t-2} \sum_{\gamma,\delta} \widehat{\mathbb{1}_{V_0}}(\gamma)\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\gamma, A_{c_1}^\top \gamma + A_{d_1}^\top \delta) \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(A_{c_{t-2}}^\top \gamma + A_{d_{t-2}}^\top \delta, \delta)\widehat{\mathbb{1}_{U_t}}(\delta) .$$

B.6 Proof of Proposition 4

Proof. We let $\mu_{2,i}$ be the sum over all (α, β) that leads to only $\theta_i = 0$, so we have $\mu_2 = \sum_{i=0}^{t-1} \mu_{2,i}$. Since exactly one θ_i is zero for some $0 \leqslant i \leqslant t-1$, we break the indices i into two cases.

Case 1. $\theta_i = 0$ for $1 \leqslant i \leqslant t-2$. In this case, it holds that $A_{c_i}^\top \alpha + A_{d_i}^\top \beta = 0$. Hence fixing a non-zero α gives us a unique non-zero β. Therefore, we have

$$\mu_{2,i} = N^{2t-2} \sum_{\alpha \neq 0} \widehat{\mathbb{1}_{V_0}}(\theta_0)\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\theta_0, \theta_1)\widehat{\mathbb{1}_{\mathcal{Q}_2}}(\theta_1, \theta_2) \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, \theta_{t-1})\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})$$

in which for $j \neq i$, $\theta_j \neq 0$ are uniquely fixed by α, and $\theta_i = 0$. Hence we have

$$\begin{aligned} \mu_{2,i} \leqslant & N^{2t-2} \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)||\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\theta_0, \theta_1)| \cdots |\widehat{\mathbb{1}_{\mathcal{Q}_{i-1}}}(\theta_{i-2}, \theta_{i-1})||\widehat{\mathbb{1}_{\mathcal{Q}_i}}(\theta_{i-1}, 0)| \\ & \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{i+1}}}(0, \theta_{i+1})||\widehat{\mathbb{1}_{\mathcal{Q}_{i+2}}}(\theta_{i+1}, \theta_{i+2})| \cdots |\widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, \theta_{t-1})||\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})| \\ \leqslant & N^4 \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)| \cdot \Phi(\mathcal{Q}_1) \cdots \Phi(\mathcal{Q}_{i-1}) \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_i}}(\theta_{i-1}, 0)| \\ & \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{i+1}}}(0, \theta_{i+1})| \cdot \Phi(\mathcal{Q}_{i+2}) \cdots \Phi(\mathcal{Q}_{t-1})| \cdot |\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})| \\ \leqslant & N^4 \cdot \left(\frac{9q^2}{N}\right)^{t-3} \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)||\widehat{\mathbb{1}_{\mathcal{Q}_i}}(\theta_{i-1}, 0)||\widehat{\mathbb{1}_{\mathcal{Q}_{i+1}}}(0, \theta_{i+1})||\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})| \\ \leqslant & \frac{(3q)^{2t-6}}{N^{t-7}} \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)| \frac{q}{N^2} \cdot \frac{q}{N^2} |\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})| \\ = & \frac{(3q)^{2t-4}}{9N^{t-3}} \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)| \cdot |\widehat{\mathbb{1}_{U_t}}(\theta_{t-1})| \leqslant \frac{(3q)^{2t-3}}{N^{t-2}} . \end{aligned}$$

Case 2. $\alpha = \theta_0 = 0$ or $\beta = \theta_{t-1} = 0$. As the two cases are very similar, we only give a proof for the case of $\beta = \theta_{t-1} = 0$.

$$\mu_{2,0} = N^{2t-2} \sum_{\alpha \neq 0} \widehat{\mathbb{1}_{V_0}}(\theta_0)\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\theta_0, \theta_1) \cdots \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, 0)\widehat{\mathbb{1}_{U_t}}(0)$$

in which for $j \neq t-1$, $\theta_j \neq 0$ are uniquely fixed by α, and $\theta_{t-1} = \beta = 0$. Hence we have

$$\begin{aligned}
\mu_{2,i} \leqslant & N^{2t-2} \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)||\widehat{\mathbb{1}_{\mathcal{Q}_1}}(\theta_0, \theta_1)| \cdots |\widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, 0)||\widehat{\mathbb{1}_{U_t}}(0)| \\
= & N^2 \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)| \cdot \varPhi(\mathcal{Q}_1) \cdots \varPhi(\mathcal{Q}_{t-2}) \cdot |\widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, 0)||\widehat{\mathbb{1}_{U_t}}(0)| \\
\leqslant & N^2 \cdot \left(\frac{9q^2}{N}\right)^{t-2} \cdot \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)||\widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, 0)| \cdot \frac{q}{N} \\
= & \frac{(3q)^{2t-3}}{3N^{t-3}} \cdot \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)||\widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\theta_{t-2}, 0)|
\end{aligned}$$

Now we let $U' = \{u \mid (u, v) \in \mathcal{Q}_{t-1}\}$, then we have $\widehat{\mathbb{1}_{U'}}(\delta) = \widehat{\mathbb{1}_{\mathcal{Q}_{t-1}}}(\delta, 0)/N$ for any $\delta \in \{0,1\}^n$. Therefore we have

$$\mu_{2,i} \leqslant \frac{(3q)^{2t-3}}{3N^{t-3}} \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_{V_0}}(\theta_0)||\widehat{\mathbb{1}_{U'}}(\theta_{t-2})| \cdot \frac{1}{N} \leqslant \frac{(3q)^{2t-3}}{N^{t-2}} \cdot \frac{q}{N} \leqslant \frac{(3q)^{2t-3}}{N^{t-2}} \, .$$

Finally by summing up all terms we obtain $\mu_2 \leqslant \sum_{i=0}^{t-1} \mu_{2,i} \leqslant \frac{t(3q)^{2t-3}}{N^{t-2}}$. □

B.7 Proof of Proposition 6

Proof. We design the algorithm $\mathcal{A}$ as the following:

1. $\mathcal{A}$ queries both V_0 at $c_1 \cdot X$ for all $0 \leqslant X < q$.
2. $\mathcal{A}$ queries both U_t at $d_1 \cdot X$ for all $0 \leqslant X < q$.
3. $\mathcal{A}$ queries P_1 at $c_1 \cdot X$ for all $0 \leqslant X < q$ in the forward direction.
4. $\mathcal{A}$ queries P_2 at $d_1 \cdot X$ for all $0 \leqslant X < q$ in the backward direction.

We note that the analysis is very similar to Proposition 1. Since $\mathcal{A}$ queries P_2 in the backward direction, we have the u_2's are sampled from $\{0,1\}^n$ without replacement. We group the u_2's by their higher $n - \log q$ bits into N/q groups, and show that for moderately large q, with overwhelming probability, each group has at least $q^2/2N$ elements.

We pick an arbitrary $r \in \{0,1\}^{n-\log q}$ and let $X_j = 1$ if the j-th sampled u_2 has the higher $n - \log q$ bits equal r, otherwise $X_j = 0$. Then one can show that for any $S \subseteq \{1, \ldots, q\}$,

$$\mathsf{Pr}[\bigwedge_{j \in S} x_j = 1] = q^{(|S|)}/N^{(|S|)} \leqslant (q/N)^{|S|} = \prod_{j \in S} \mathsf{Pr}[X_j = 1] \, .$$

Therefore, we can apply the Chernoff bound for negatively correlated variables (i.e., Lemma 13) and obtain that $\mathsf{Pr}[\sum_{j=1}^{q} X_j < \frac{q^2}{2N}] \leqslant e^{-q^2/8N}$. By a union bound over all N/q groups, we have every group maintaining at least $q^2/2N$ elements with probability at least $1 - (N/q) \cdot e^{-q^2/8N}$.

Now, we fix an arbitrary $(u_1, v_1) \in \mathcal{Q}_1$ out of q choices, then we have at least $q^2/2N$ choices of $(u_2, v_2) \in \mathcal{Q}_2$ such that $0 \leqslant v_1 + u_2 < q$. Therefore we can find a unique $v_0 \in \mathcal{V}_0$ and $u_3 \in U_3$ so that $v_0 + u_1 = v_1 + u_2 = v_2 + u_3$, concluding that the sum capture quantity is lower-bounded by $q^3/2N$. □

C Omitted Proofs for Good Transcript Analysis

C.1 Proof of inequality (10)

Proof. We note that $\tau = (\vec{\mathcal{Q}}, \vec{k}) \in \mathcal{T}$ is an attainable transcript. By definition, we can write the ratio as

$$\begin{aligned}\frac{\mathsf{ps}_0(\vec{\mathcal{Q}}, \vec{k})}{\mathsf{ps}_1(\vec{\mathcal{Q}}, \vec{k})} &= \frac{\mathsf{Pr}_{P_1,\ldots,P_t}[E \downarrow \mathcal{Q}_E, P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t]}{\mathsf{Pr}_{P_0,P_1,\ldots,P_t}[P_0 \downarrow \mathcal{Q}_E, P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t]} \\ &= \frac{\mathsf{Pr}_{P_1,\ldots,P_t}[E \downarrow \mathcal{Q}_E \mid P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t]}{\mathsf{Pr}_{P_0,P_1,\ldots,P_t}[P_0 \downarrow \mathcal{Q}_E \mid P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t]}\end{aligned}$$

in which E is the real-world cipher construction that depends on $P_1, \ldots, P_t$ and key $\vec{k}$. We note that the second equality directly follows from dividing both nominator and denominator by the probability $\mathsf{Pr}[P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t]$.

Next we compute the conditional probability term in the denominator, given P_0 is independent from the rest of permutations $P_1, \ldots, P_t$, we have

$$\mathsf{Pr}_{P_0,P_1,\ldots,P_t}[P_0 \downarrow \mathcal{Q}_E \mid P_1 \downarrow \mathcal{Q}_1, \ldots, P_t \downarrow \mathcal{Q}_t] = \prod_{i=0}^{|\mathcal{Q}_E|-1} \frac{1}{N-i} = \frac{1}{N^{(|\mathcal{Q}_E|)}}.$$

Hence we proved the equality. □

C.2 Proof of Lemma 8

We first restate the lemma.

Lemma 14. *For any graph G that consists of $m+1$ layers $L_0, \ldots, L_m$, with each adjacent layer forming a (partial) matching, we have that for any $u \in L_0, v \in L_m$ such that u is right-free and v is left-free,*

$$\mathsf{Pr}[u \to v] = \frac{1}{N} - \frac{1}{N} \sum_{\sigma \in \mathcal{B}_G(0,m)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_G(i_{h-1}, i_h)}{U_G(i_h, i_h)} ,$$

where $\mathcal{B}_G(0,m)$ contains all interesting $(0,m)$-segment partition of G.

Proof. Because all analyses will be performed over graph G, we simplify the notation by letting $\mathcal{U}_{i,j} := \mathcal{U}_G(i,j)$ and $U_{i,j} := U_G(i,j)$. We also define the set of indices between a and b as $\mathcal{I}(a,b) := \{i \in \mathcal{I}(G) \mid a \leqslant i \leqslant b\}$. To prove the lemma, we first need the following intermediate result.

Claim. For any $1 \leqslant i \leqslant j \leqslant m$ such that $i \in \mathcal{I}(G)$,

$$\mathsf{Pr}[w_i \in \mathcal{U}_G(i,j)] = \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 + \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}_G(y,i)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}\right)$$

Proof. We prove the claim by induction over i. We note that the base case $i = 1 \in \mathcal{I}(G)$ because $u \in L_0$ is right-free, implying that there exists a left-free vertex in L_1. It is trivial that $\mathsf{Pr}[w_1 \in \mathcal{U}_{1,j}] = U_{1,j}/U_{1,1}$.

For the inductive case, suppose that for all $k \in \mathcal{I}(G), k < i$ the claim holds. We note that for w_i, we have

$$\begin{aligned}\mathsf{Pr}[w_i \in \mathcal{U}_{i,j}] &= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 - \sum_{k\in\mathcal{I}(1,i-1)} \mathsf{Pr}[w_i \in \mathcal{U}_{k,i}]\right) \\ &= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 - \sum_{k\in\mathcal{I}(1,i-1)} \mathsf{Pr}[w_k \in \mathcal{U}_{k,i}]\right) . \end{aligned} \tag{18}$$

The first equality holds because, first, given w_i hits at the head of a path in L_i, it is impossible to have w_i belonging to any path starts earlier than L_i, while also noting that all events $w_i \in \mathcal{U}_{k,i}$ are disjoint. Then given w_i does not belong to any $\mathcal{U}_{k,i}$ for $k < i$, the random process samples a vertex from $U_{i,i}$ left-free vertices in L_i, with $U_{i,j}$ of them are the head of path that ends no earlier than layer L_j. The second equality is applying the fact that $\mathsf{Pr}[w_i \in \mathcal{U}_{k,i}] = \mathsf{Pr}[w_k \in \mathcal{U}_{k,i}]$.

Now we plug the equality of $\mathsf{Pr}[w_k \in \mathcal{U}_{k,i}]$ into (18) and proceed with the calculation.

$$\mathsf{Pr}[w_i \in \mathcal{U}_{i,j}]$$

$$= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 - \sum_{k\in\mathcal{I}(1,i-1)} \frac{U_{k,j}}{U_{k,k}} \left(1 + \sum_{y\in\mathcal{I}(1,k-1)} \sum_{\sigma\in\mathcal{B}(y,k)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}\right)\right) \tag{19}$$

$$= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 - \sum_{k\in\mathcal{I}(1,i-1)} \frac{U_{k,j}}{U_{k,k}} - \sum_{k\in\mathcal{I}(1,i-1)} \sum_{y\in\mathcal{I}(1,k-1)} \frac{U_{k,j}}{U_{k,k}} \sum_{\sigma\in\mathcal{B}(y,k)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}\right) \tag{20}$$

$$= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 - \sum_{k\in\mathcal{I}(1,i-1)} \frac{U_{k,j}}{U_{k,k}} - \sum_{y\in\mathcal{I}(1,i-2)} \sum_{k\in\mathcal{I}(y+1,i-1)} \frac{U_{k,j}}{U_{k,k}} \sum_{\sigma\in\mathcal{B}(y,k)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}\right) \tag{21}$$

$$= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 + \sum_{k\in\mathcal{I}(1,i-1)} (-1) \cdot \frac{U_{k,j}}{U_{k,k}} + \sum_{y\in\mathcal{I}(1,i-2)} \sum_{\sigma\in\mathcal{B}(y,i):\ |\sigma|\geqslant 2} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}\right) \tag{22}$$

$$= \frac{U_{i,j}}{U_{i,i}} \cdot \left(1 + \sum_{k\in\mathcal{I}(1,i-1)} \sum_{\sigma\in\mathcal{B}(k,i)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}\right) . \tag{23}$$

In the above calculation, (19) comes from directly expanding the $\mathsf{Pr}[w_k \in \mathcal{U}_{k,i}]$ terms. We switch the summation of k and y in (21). Note that we get (22) because the second k term is essentially enumerating a partition point for the interval $\{y, y+1, \ldots, i-1, i\}$, so we can merge the summation over k into σ, obtaining $\sigma \in \mathcal{B}(y, i)$ with $|\sigma| \geqslant 2$. Finally combining the remaining terms give us summation over σ without size requirement, as in (23).

Hence we completed the inductive step and proved the claim. □

To further proceed with the remaining proof, we need the following proposition and we will applying the following equality multiple times in later calculations.

Proposition 11. *For any $i \in \mathcal{I}(G)$,*

$$\frac{1}{U_{ii}} = \frac{1}{N} + \frac{1}{N} \sum_{x \in \mathcal{I}(0,i-1)} \frac{U_{x,i}}{U_{i,i}} \, .$$

Proof (of Proposition 11). We have $N = U_{i,i} + \sum_{x \in \mathcal{I}(0,i-1)} U_{x,i}$ because the total N paths that pass vertices in L_i start no later than L_i, and no paths start at L_x where $x \notin \mathcal{I}(G)$. Then we are done by dividing both sides by $N \cdot U_{i,i}$, which is non-zero when $i \in \mathcal{I}(G)$. □

We continue the calculation.

$$\begin{aligned}
&\mathsf{Pr}[w_i \in U_{i,j}] \\
&= \frac{U_{i,j}}{U_{i,i}} + \frac{U_{i,j}}{U_{i,i}} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} && (24) \\
&= \left(\frac{U_{i,j}}{N} + \frac{U_{i,j}}{N} \sum_{x \in \mathcal{I}(0,i-1)} \frac{U_{x,i}}{U_{i,i}} \right) && (25) \\
&\quad + \frac{U_{i,j}}{U_{i,i}} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \left(\frac{U_{i_0,i_1}}{N} + \sum_{x \in \mathcal{I}(0,i_0-1)} \frac{U_{i_0,i_1}}{N} \cdot \frac{U_{x,i_0}}{U_{i_0,i_0}} \right) \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} && (26) \\
&= \left(\frac{U_{i,j}}{N} + \frac{U_{i,j}}{N} \sum_{x \in \mathcal{I}(0,i-1)} \sum_{\sigma \in \mathcal{B}_G(x,i),\ |\sigma|=1} \frac{U_{i_0,i_1}}{U_{i_1,i_1}} \right) \\
&\quad + \frac{U_{i,j}}{U_{i,i}} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \cdot \frac{U_{i_0,i_1}}{N} \cdot \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} && (27) \\
&\quad + \frac{U_{i,j}}{U_{i,i}} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \sum_{x \in \mathcal{I}(0,i_0-1)} \frac{U_{i_0,i_1}}{U_{i_0,i_0}} \cdot \frac{U_{x,i_0}}{N} \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} && (28) \\
&= \left(\frac{U_{i,j}}{N} - \frac{U_{i,j}}{N} \sum_{x \in \mathcal{I}(0,i-1)} \sum_{\sigma \in \mathcal{B}_G(x,i),\ |\sigma|=1} (-1)^{|\sigma|} \frac{U_{i_0,i_1}}{U_{i_1,i_1}} \right) \\
&\quad + \frac{U_{i,j}}{U_{i,i}} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \cdot \frac{U_{i_0,i_1}}{N} \cdot \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} \\
&\quad + \frac{U_{i,j}}{U_{i,i}} \sum_{x \in \mathcal{I}(0,i-2)} \sum_{y \in \mathcal{I}(x+1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \frac{U_{i_0,i_1}}{U_{i_0,i_0}} \cdot \frac{U_{x,i_0}}{N} \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} && (29) \\
&= \left(\frac{U_{i,j}}{N} - \frac{U_{i,j}}{N} \sum_{x \in \mathcal{I}(0,i-1)} \sum_{\sigma \in \mathcal{B}_G(x,i),\ |\sigma|=1} (-1)^{|\sigma|} \frac{U_{i_0,i_1}}{U_{i_1,i_1}} \right) \\
&\quad + \frac{U_{i,j}}{U_{i,i}} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \cdot \frac{U_{i_0,i_1}}{N} \cdot \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}}
\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{U_{i,j}}{U_{i,i}} \sum_{x \in \mathcal{I}(0,i-2)} \sum_{\sigma \in \mathcal{B}(x,i)\colon\, |\sigma| \geqslant 2} (-1)^{|\sigma|} \frac{U_{x,i_0}}{N} \prod_{h=2}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_{h-1},i_{h-1}}} \qquad (30)\\
&= \left(\frac{U_{i,j}}{N} - \frac{U_{i,j}}{N} \sum_{x \in \mathcal{I}(0,i-1)} \sum_{\sigma \in \mathcal{B}_G(x,i),\ |\sigma|=1} (-1)^{|\sigma|} \frac{U_{i_0,i_1}}{U_{i_1,i_1}} \right)\\
&\quad + \frac{U_{i,j}}{N} \sum_{y \in \mathcal{I}(1,i-1)} \sum_{\sigma \in \mathcal{B}(y,i)} (-1)^{|\sigma|} \cdot \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_h,i_h}}\\
&\quad - \frac{U_{i,j}}{N} \sum_{x \in \mathcal{I}(0,i-2)} \sum_{\sigma \in \mathcal{B}(x,i)\colon\, |\sigma| \geqslant 2} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_h,i_h}} \qquad (31)\\
&= \frac{U_{i,j}}{N} - \frac{U_{i,j}}{N} \sum_{\sigma \in \mathcal{B}(0,i)} (-1)^{|\sigma|} \prod_{h=1}^{|\sigma|} \frac{U_{i_{h-1},i_h}}{U_{i_h,i_h}}
\end{aligned}
$$

In the above calculation, we get the term (25) by applying Proposition 11 to the first term in (24), and we get (26) by separating the (i_0, i_1) term in the multiplication out and applying Proposition 11. By splitting the term (26) we get terms (27) and (28). To get the term (29) from (28), we take out the summation over x and switch it with summation over y. Merging the terms giving us . In (31) we switch the denominator term $U_{i,i}$ outside the summation with the denominator term N inside the summation, causing the subscript i_{h-1} in the denominator shifted to i_h. And the terms for $x, y \geqslant 1$ are all canceled, finally giving us the desired equality.

If we slightly tweak the graph G by adding another layer $m+1$ with a single edge connecting v to L_{m+1}. Then we proved the lemma by computing $\mathsf{Pr}[u \to v] = \mathsf{Pr}[w_m \in \mathcal{U}_{m,m+1}]$ into which we plugging $U_{m,m+1} = 1$. □