Security Proofs for Key-Alternating Ciphers with Non-Independent Round Permutations

Liqing Yu [*] Yusai Wu [†] Yu Yu [‡] Zhenfu Cao [§] Xiaolei Dong [¶]

Abstract

This work studies the key-alternating ciphers (KACs) whose round permutations are not necessarily independent. We revisit existing security proofs for key-alternating ciphers with a single permutation (KACSPs), and extend their method to an arbitrary number of rounds. In particular, we propose new techniques that can significantly simplify the proofs, and also remove two unnatural restrictions in the known security bound of 3-round KACSP (Wu et al., Asiacrypt 2020). With these techniques, we prove the first tight security bound for t-round KACSP, which was an open problem. We stress that our techniques apply to all variants of KACs with non-independent round permutations, as well as to the standard KACs.

[*] East China Normal University & Shanghai Qi Zhi Institute, China. Email: lqyups@126.com
[†] Shanghai Qi Zhi Institute, China. Email: yusaiwu@126.com
[‡] Shanghai Jiao Tong University & Shanghai Qi Zhi Institute, China. Email: yuyu@yuyu.hk
[§] East China Normal University, China. Email: zfcao@sei.ecnu.edu.cn
[¶] East China Normal University, China. Email: dong-xl@sei.ecnu.edu.cn

Contents

1 Introduction

The key-alternating ciphers (see Eqn. (1)) generalize the Even-Mansour construction [EM97] over multiple rounds. They can be viewed as abstract constructions of many substitutionpermutation network (SPN) block ciphers (e.g. AES [DR02]). In addition, there are various variants of the key-alternating ciphers.

This work only considers the case of independent round keys, and reducing their independence is a relatively parallel topic. That is, we are concerned with different variants of KACs on round permutations, while the round keys are always independent and random. For convenience, we simply use KAC to represent the standard KAC with independent permutations, and refer to all the other variants as *KAC-type constructions*. In particular, KACSP is a KAC-type construction in which all the round permutations are identical.

In a t-round KAC or KAC-type construction, the number of different round permutations, denoted t', is an important parameter. Clearly, we have $t' = t$ in the case of KAC and $t' = 1$ in the case of KACSP. When $t' < t$, it means that there are different rounds using the same permutation. For a given construction, we name the round permutations as follows. In particular, the name P_k will be assigned to each round permutation in order from round 1 to round t, where $k \in \{1, \ldots, t'\}$. For round i, we check if there exists $j < i$ such that round j uses the same permutation as round i. If so, we use the same name as the permutation in round j; otherwise, we use the name P_k, where $k \in \{1, \ldots, t'\}$ is the smallest integer not used in previous rounds. For simplicity, we sometimes only use the permutation names to denote a construction, such as $P_1P_2P_3$-construction (i.e. 3-round KAC), $P_1P_1P_1$-construction (i.e. 3-round KACSP), $P_1P_1P_2$-construction, etc.

We now give a more formal definition of KAC and KACSP constructions. Let $x \in \{0,1\}^n$ denote the plaintext, $\kappa_0, \kappa_1, \ldots, \kappa_t \in \{0,1\}^{n\times(t+1)}$ denote the $t+1$ round keys, and $P_1, \ldots, P_t$ denote the permutations over $\{0,1\}^n$, then the outputs of t-round KAC and t-round KACSP are computed as follows.

$$\mathsf{KAC}^{P_1,\ldots,P_t;\,\kappa_0,\kappa_1,\ldots,\kappa_t}(x) \stackrel{\text{def}}{=} \kappa_t \oplus P_t(\kappa_{t-1} \oplus P_{t-1}(\cdots P_2(\kappa_1 \oplus P_1(\kappa_0 \oplus x))\cdots)), \quad (1)$$

$$\mathsf{KACSP}^{P_1;\,\kappa_0,\kappa_1,\ldots,\kappa_t}(x) \stackrel{\text{def}}{=} \kappa_t \oplus P_1(\kappa_{t-1} \oplus P_1(\cdots P_1(\kappa_1 \oplus P_1(\kappa_0 \oplus x))\cdots)). \quad (2)$$

Related Works. Bogdanov et al. [Bog+12] were the first to study the provable security of t-round KAC (for $t \geq 2$), and showed that it is secure up to $\mathcal{O}(2^{\frac{2}{3}n})$ queries. On the other hand, they presented a simple distinguishing attack using $\mathcal{O}(2^{\frac{t}{t+1}n})$ queries, and conjectured that this attack cannot be improved intrinsically. Thus, their result is optimal for 2-round KAC. After a series of papers [Ste12; LPS12; CS14; HT16], the above conjecture was proved. Roughly, it says that unless $\Omega(2^{\frac{t}{t+1}n})$ queries are used, one cannot distinguish t-round KAC from a truly random permutation with non-negligible advantage, where the round permutations are public and random.

Another line of research focuses on the variants of KAC constructions, where round permutations and keys may not be independent of each other. [DKS12] was the first to study the minimalism of Even-Mansour cipher, and showed that several of its single-key variants could achieve the same level of security as it. Later, Chen et al. [Che+18] proved that a variant of 2-round KAC still enjoys security close to $\mathcal{O}(2^{\frac{2}{3}n})$ when only n-bit key and a single permutation are used. Next, [WYCD20] generalized Chen et al.'s technique and proved a tight security bound (with two unnatural restrictions) for 3-round KACSP. Recently, Tessaro and Zhang [TZ21] showed that $(t-2)$-wise independent round keys are sufficient for t-round KAC to achieve the tight security bound, where $t \geq 8$.

Our Contributions. This work focuses on the provable security of KAC or KAC-type constructions in random permutation model. Our main contribution is to prove the tight security bound $\mathcal{O}(2^{\frac{t}{t+1}n})$ for t-round KACSP.

We revisit the security proofs in [Che+18; WYCD20]. The idea of their proofs is not hard to understand, but the analysis is quite laborious. In particular, the security bound of [WYCD20] (see Thm. 5) has two unnatural restrictions, making the result far from elegant. The first is the existence of an error function $\zeta(\cdot)$, and the second is that it requires $28q_e^2/2^n \leq q_p \leq q_e/5$, where q_p and q_e denote the number of two types of queries made by the distinguisher respectively.

We propose new techniques that can significantly simplify proofs, thus making the security proofs of KAC-type constructions easier to understand and read. One of the key techniques is a general transformation, which reduces our task to bounding only one probability in the form of (9) (even for t-round constructions). Note that [WYCD20] needs to bound at least 3 such probabilities. We stress that the transformation is general and may also be used to simplify other security proofs. To increase the number of constructive methods, we introduce a new notion of *recycled-edge* which is different from the *shared-edge* used in [Che+18] and [WYCD20]. Roughly speaking, recycled-edge is to reuse existing permutation queries made by distinguisher to save resources, while shared-edge is to reuse the permutation queries generated on-the-fly. We point out that recycled-edge has the following features compared to shared-edge. First, the analysis of recycled-edge is easier, which is another important reason why our proof is simpler. Second, the recycled-edge has wider applicability and is less sensitive to constructions.

Moreover, we provide new ideas to remove the two unnatural restrictions in the security bound of [WYCD20]. For the first restriction, our approach is to consider the security proof in two disjoint cases, and provide separate proofs for each case. It should be pointed out here that these two proofs will be almost identical, except for slightly different calculations. For the second restriction, our approach is to increase the number of variables[1] so that we can better exploit the power of multivariate hypergeometric distribution used in the calculation. Our main finding here is that the improvements in security bound are largely influenced by computational rather than conceptual factors. This is a key to addressing the security bound of t-round KACSP. More details about our new techniques can be found in Section 3.

With the above new techniques, we first obtain a neat security bound for the 3-round KACSP (see Thm. 6), and discuss its proof in detail in Section 4. We then generalize the proof to the general t-round KACSP (see Thm. 11), using almost the same techniques. It should be emphasized that our proof techniques apply to KAC and all kinds of KAC-type constructions. For example, we also apply the proof techniques to other variants of 3-round KAC (see Thms. 17 and 18).

Organization. In Section 2, we introduce the notation and basic concepts used throughout this work. In Section 3, we outline the techniques used in security proofs of this work. In Section 4, we show an improved security bound of 3-round KACSP (compared to [WYCD20]) and also provide its proof in detail. In Section 5, we give the first tight security bound of t-round KACSP for any $t \geq 4$. In Section 6, we give some remarks on other variants of KACs.

2 Preliminaries

2.1 Notation

Let $N = 2^n$ and $\mathcal{P}_n$ be the set of all permutations over $\{0,1\}^n$. For a permutation $P \in \mathcal{P}_n$, we let P^{-1} denote its inverse permutation. If A is a finite set, then $|A|$ and $\overline{A}$ represent the cardinality and complement of A, respectively. Given a set of n-bit strings A and a fixed $k \in \{0,1\}^n$, we will use $A \oplus k$ to denote the set $\{a \oplus k : a \in A\}$. For a finite set S, let $x \leftarrow_\$ S$ denote the act of sampling uniformly from S and then assigning the value to x. The falling factorial is usually written by $(a)_b = a(a-1)\dots(a-b+1)$, where $1 \leq b \leq a$ are two integers. For a set of pairs $\mathcal{Q} = \{(x_1, y_1), \dots, (x_q, y_q)\}$, where x_i's (resp. y_i's) are distinct n-bit strings, and a permutation $P \in \mathcal{P}_n$, we say that P extends the set $\mathcal{Q}$, denoted as $P \downarrow \mathcal{Q}$, if $P(x_i) = y_i$ for $i = 1, 2, \dots, q$. In particular, we write $\mathsf{Dom}(\mathcal{Q}) := \{x_1, \dots, x_q\}$ (resp. $\mathsf{Ran}(\mathcal{Q}) := \{y_1, \dots, y_q\}$) as the domain (resp. range) of $\mathcal{Q}$.

[1] Each variable represents the number of new edges that can be saved by some constructive method, usually denoted by h_i in the proofs.

2.2 Random Permutation Model, Transcripts and Graph View

Random Permutation Model. This work studies the security of KAC or KAC-type constructions under the *random permutation model*. The model can be viewed as an enhanced version of black-box indistinguishability with additional access to the underlying permutations, making security analysis more operable.

Given a t-round KAC or KAC-type construction, the task of distinguisher $\mathcal{D}$ is to tell apart two worlds, the *real world* and the *ideal world*. In the real world, the distinguisher can interact with $t'+1$ oracles $(E_K, P_1, \ldots, P_{t'})$, where E_K is the t-round target cipher (denoted as E) computed based on t' independent random permutations $P_1, \ldots, P_{t'}$ and a key K. In the ideal world, there are also $t'+1$ oracles but the first oracle E_K is replaced by an independent random permutation P_0. That is, what interact with the distinguisher $\mathcal{D}$ are $t'+1$ independent random permutations $(P_0, P_1, \ldots, P_{t'})$. Furthermore, we allow the distinguisher to be adaptive and query each permutation oracle in both directions. We can then define the *super-pseudorandom permutation* (SPRP) advantage of distinguisher $\mathcal{D}$ on t-round E_K (with t' different permutations) as follows.

$$\mathbf{Adv}^{\mathsf{SPRP}}_{E,t}(\mathcal{D}) = \left|\Pr_{\substack{K \leftarrow_\$ \{0,1\}^{(t+1)n};\\ P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}}[\mathcal{D}^{E_K, P_1, \ldots, P_{t'}} = 1] - \Pr_{P_0, P_1, \ldots, P_{t'} \leftarrow_\$ \mathcal{P}_n}[\mathcal{D}^{P_0, P_1, \ldots, P_{t'}} = 1]\right|, \tag{3}$$

where all oracles can be queried bidirectionally. In particular, we refer to the queries on the first oracle (i.e. E_K or P_0) as *construction queries* and to the set formed by them and their answers as $\mathcal{Q}_0$. Similarly, the queries on the other t' oracles are called *permutation queries* and the resulting sets are denoted as $\mathcal{Q}_i$, where $i = 1, \ldots, t'$.

Transcripts. Formally, the interaction between $\mathcal{D}$ and $t'+1$ oracles can be represented by an ordered list of queries, which is often called *transcript*. Each query in the transcript is in the form of (i, b, u, v), where $i \in \{0, 1, \ldots, t'\}$ represents the oracle being queried, b indicates whether it is a forward query or backward query, u is the query value and v is the corresponding answer. We can assume *wlog* that the adversary $\mathcal{D}$ is deterministic and does not make redundant queries, since it is computationally unbounded. That means the output of $\mathcal{D}$ is entirely determined by its transcript, which can also be encoded (requiring a description of $\mathcal{D}$) into $t'+1$ unordered lists of queries.

In addition, we are more generous to the distinguisher $\mathcal{D}$ in the analysis, so that it will receive the actual key used in the real world (after all queries are done but before a decision is made). To maintain consistency, $\mathcal{D}$ would also receive a dummy key in the ideal world (even the key is not used). This modification is justified since it only increases the advantage of $\mathcal{D}$. From the perspective of $\mathcal{D}$, a transcript $\tau \in \mathcal{T}$ has the form of $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}, K)$, and can be rewritten as the following unordered lists.

$$\tau = \left\{\begin{array}{l} \mathcal{Q}_0 = \{(x_1, y_1), \ldots, (x_{q_e}, y_{q_e})\}, \\ \mathcal{Q}_1 = \{(u_{1,1}, v_{1,1}), \ldots, (u_{1,q_1}, v_{1,q_1})\}, \\ \quad \cdots, \\ \mathcal{Q}_{t'} = \{(u_{t',1}, v_{t',1}), \ldots, (u_{t',q_{t'}}, v_{t',q_{t'}}), \\ \ K = (\kappa_0, \ldots, \kappa_t) \end{array}\right\}, \tag{4}$$

where $y_j = E_K(x_j)$ or $y_j = P_0(x_j)$ (depending on which world) for all $j \in \{1, \ldots, q_e\}$ and $v_{i,j} = P_i(u_{i,j})$ for all $i \in \{1, \ldots, t'\}$ and $j \in \{1, \ldots, q_i\}$, and where $K \in \{0,1\}^{(t+1)n}$ is a $(t+1)n$-bit key.

Statistical Distance of Transcript Distributions. We already know that the output of $\mathcal{D}$ is a deterministic function on transcript. For any fixed distinguisher $\mathcal{D}$, its advantage is obviously bounded by the statistical distance of transcript distributions in two worlds. That is, it is usually to determine the upper bound of the value (3) as follows,

$$\begin{aligned}(3) \leq \|\mathcal{T}_{\mathsf{real}} - \mathcal{T}_{\mathsf{ideal}}\| &\overset{\mathsf{def}}{=} \frac{1}{2}\sum_{\tau} |\Pr[\mathcal{T}_{\mathsf{ideal}} = \tau] - \Pr[\mathcal{T}_{\mathsf{real}} = \tau]| \\ &= \sum_{\tau} \max\{0, \Pr[\mathcal{T}_{\mathsf{ideal}} = \tau] - \Pr[\mathcal{T}_{\mathsf{real}} = \tau]\},\end{aligned} \tag{5}$$

where $\|\cdot\|$ represents the statistical distance, and $\mathcal{T}_{\mathsf{real}}$ (resp. $\mathcal{T}_{\mathsf{ideal}}$) denotes the transcript random variable generated by the interaction of $\mathcal{D}$ with the real (resp. ideal) world. We let $\mathcal{T}$ denote the set of *attainable* transcripts τ such that $\Pr[\mathcal{T}_{\mathsf{ideal}} = \tau] > 0$. It is worth noting that although the set $\mathcal{T}$ depends on $\mathcal{D}$, the probabilities $\Pr[\mathcal{T}_{\mathsf{ideal}} = \tau]$ and $\Pr[\mathcal{T}_{\mathsf{real}} = \tau]$ (for any $\tau \in \mathcal{T}$) are independent of $\mathcal{D}$, since they are inherent properties of the two worlds. The task of bounding (5) is to figure out two (partial) distributions, of which the one for ideal world is simple and easy to deal with. Thus, the main effort in various proofs is essentially to study the random value $\mathcal{T}_{\mathsf{real}}$.

Crucial Probability in the Real World. The basis of studying $\mathcal{T}_{\mathsf{real}}$ is the probability $\Pr[\mathcal{T}_{\mathsf{real}} = \tau]$, which can be reduced to a conditional probability with intuitive meaning (see Eqn. (7)). For any fixed transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}, K) \in \mathcal{T}$, it has

$$\begin{aligned}\Pr[\mathcal{T}_{\mathsf{real}} = \tau] &= \Pr_{\substack{\kappa \leftarrow_\$ \{0,1\}^{(t+1)n}; \\ P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}}[E_\kappa \downarrow \mathcal{Q}_0 \wedge P_1 \downarrow \mathcal{Q}_1 \wedge \cdots \wedge P_{t'} \downarrow \mathcal{Q}_{t'} \wedge \kappa = K] \\ &= \Pr_{\substack{\kappa \leftarrow_\$ \{0,1\}^{(t+1)n}; \\ P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}}[P_1 \downarrow \mathcal{Q}_1 \wedge \cdots \wedge P_{t'} \downarrow \mathcal{Q}_{t'} \wedge \kappa = K] \quad (6) \\ &\quad \times \Pr_{P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_0 \mid P_1 \downarrow \mathcal{Q}_1 \wedge \cdots \wedge P_{t'} \downarrow \mathcal{Q}_{t'}] \quad (7)\end{aligned}$$

The central task of calculating $\Pr[\mathcal{T}_{\mathsf{real}} = \tau]$ is to evaluate Eqn. (7)[2], since the value of Eqn. (6) can be solved trivially for any KAC or KAC-type construction. In this work, we will use a *graph view* (basically taken from [CS14] and to be defined in next part), then Eqn. (7) can be interpreted as the probability that all the paths between x_j and y_j (where $(x_j, y_j) \in \mathcal{Q}_0$) are completed, when each random permutation P_i extending the corresponding set $\mathcal{Q}_i$.

Graph View. It is often more convenient to work with constructions and transcripts in a graph view. Here we take only the t-round KAC or KAC-type construction as an example, and other constructions are similar. For a given construction, all the information of transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}, K) \in \mathcal{T}$ can be encoded into a *round graph* $G(\tau)$. First, one can view each set $\mathcal{Q}_i$ as a bipartite graph with shores $\{0,1\}^n$ and containing q_i (resp. q_e, in the case of $\mathcal{Q}_0$) disjoint edges. To have maximum generality, we here keep the value of $K = (\kappa_0, \ldots, \kappa_t)$ in graph $G(\tau)$[3], where each mapping of XORing round key κ_i is viewed as a full bipartite graph (i.e. it contains 2^n disjoint edges).

More specifically, graph $G(\tau)$ contains $2(t+1)$ shores, each of which is identified with a copy of $\{0,1\}^n$. The $2(t+1)$ shores are indexed as $0, 1, 2, \ldots, 2t+1$. We use the ordered pair $\langle i, u \rangle$ to represent the string u in shore i, where $i \in \{0, 1, \ldots, 2t+1\}$ and $u \in \{0,1\}^n$. For convenience, we simply use u to denote a string if it is clear from the context which shore the u is in. In particular, the vertices in shore 0 and shore $2t+1$ are often called *plaintexts* and *ciphertexts*, respectively. More care should be taken when $t' < t$, as this means that the target construction uses the same permutation in different rounds. For any $i \neq j \in \{1, \ldots, t\}$ that round i and round j use the same permutation, the shores $2i-1$ and $2j-1$ are actually the same, and the shores $2i$ and $2j$ are also the same. That is, $\langle 2i-1, u \rangle = \langle 2j-1, u \rangle$ and $\langle 2i, v \rangle = \langle 2j, v \rangle$ for all $u, v \in \{0,1\}^n$.

We define the even-odd edges between shore $2i$ and shore $2i+1$ as $E_{(2i,2i+1)} := \{(v, v \oplus \kappa_i) : v \in \{0,1\}^n\}$ and call them *key-edges*, where $i \in \{0, \ldots, t\}$. The key-edges $E_{(2i,2i+1)}$ correspond to the step of XORing round key κ_i in the KAC or KAC-type construction, and form a perfect matching of bipartite graph.

For $i \in \{1, \ldots, t\}$, we use the odd-even edges between shore $2i-1$ and shore $2i$ to represent the queries made to the permutation in round i, and call them *permutation-edges*. Naturally, the term P_k-permutation-edge is used to indicate the round permutation associated with it, where $k \in \{1, \ldots, t'\}$. Based on the definition of strings above, more care should also be taken when $t' < t$. For any $i \neq j \in \{1, \ldots, t\}$ that round i and round j use the same permutation, the bipartite graph between the shore $2i-1$ and $2i$, and the bipartite graph between the shore $2j-1$ and $2j$ are the same one. More

[2] For t-round KAC, the technical lemma of [CS14] (see Lemma 1) solves exactly this probability when $|\mathcal{Q}_0| = 1$.

[3] Although this leads to a somewhat redundant notation, it is still relatively easy to understand. For a concrete example, you can refer to Figure 1 in Appendix C.1.

specifically, we define the permutation-edges between shore $2i-1$ and $2i$ as $E_{(2i-1,2i)} := \{\langle u, P_k, v\rangle : (u,v) \in \mathcal{Q}_k\}$[4] for $i = 1, \ldots, t$, where P_k $(1 \le k \le t')$ is the name of round permutation between shore $2i-1$ and $2i$ (see the naming in Section 1). That is, we distinguish strings and permutation-edges by the round permutation associated with them, rather than by shores.

In addition, we should keep in mind that there are implicit permutation-edges (i.e., $\{\langle x_i, \mathcal{Q}_0, y_i\rangle : (x_i, y_i) \in \mathcal{Q}_0\}$, although not drawn) directly from shore 0 to shore $2t+1$ according to the construction queries in $\mathcal{Q}_0$, i.e. these edges are from the plaintexts x_i's to the corresponding ciphertexts y_i's. Throughout this work, we use symbols related to x (e.g., x_i and x_i') and y (e.g., y_i and y_i') to denote plaintexts (i.e., strings in shore 0) and ciphertexts (i.e., strings in shore $2t+1$), respectively.

Basic Definitions about Graph. We say shore i is to the left of shore j if $i < j$, and view paths as oriented from left to right. For convenience, the index of the shore containing vertex u is written as $\mathsf{Sh}(u)$. A vertex u in a shore i is called *right-free*, if no edge connects u to any vertex in shore $i+1$. A vertex v in a shore j is called *left-free*, if no edge connects v to any vertex in shore $j-1$. Notice that right-free vertices and left-free vertices must be located on the odd and even shores, respectively.

We write $\mathsf{R}(u)$ for the rightmost vertex in the path of $G(\tau)$ starting at u, and $\mathsf{L}(v)$ for the leftmost vertex in the path of $G(\tau)$ ending at v. For any odd $i \in \{0, \ldots, 2t+1\}$ and $i < j \in \{0, \ldots, 2t+1\}$, we let U_{ij} denote the set of paths that starts at a left-free vertex in shore i and reaches a vertex in shore j. Similarly, for any $i < j \in \{0, \ldots, 2t+1\}$, we use Z_{ij} to denote the set of paths that starts at a vertex in shore i and reaches a vertex in shore j. That is, the only difference between Z_{ij} and U_{ij} is that the starting vertices on shore i in the former need not be left-free.

Path-Growing Procedure. In this work, we usually imagine the crucial probability (7) as connecting all x_j with y_j through a (probabilistic) *path-growing procedure*, where $(x_j, y_j) \in \mathcal{Q}_0$. Note that all the key-edges already exist, so we only need to generate edges from odd shores to the next shore. Given $G(\tau)$ and a vertex u, we define the following procedure to generate a path $(u, w_1, \ldots, w_r)$ from u.

Let $w_0 = u$. For i from 1 to r, if w_{i-1} is not right-free and adjacent to some vertex z in shore i, then let $w_i = z$; otherwise, sample u_i uniformly at random from all left-free vertices in shore i, and let $w_i = u_i$.

For convenience, we let $u \to v$ denote the event that u is connected to v through the above path-growing procedure and write $\Pr_G[u \to v] = \Pr_G[w_r = v]$, where v is a vertex in shore $\mathsf{Sh}(u) + r$. We are now ready to give the key lemma of [CS14] (adapted slightly to fit here) as follows.

Lemma 1 (Lemma 1 of [CS14]). *Given any $G(\tau)$ as described above, let u be any right-free vertex in shore 1 and v be any left-free vertex in shore $2t$, then it has*

$$\Pr_{G(\tau)}[u \to v] = \frac{1}{N} - \frac{1}{N}\sum_{\sigma}(-1)^{|\sigma|}\prod_{j=1}^{|\sigma|}\frac{|U_{i_{j-1}i_j}|}{N - |\mathcal{Q}_{(i_j-1)/2}|}. \tag{8}$$

where the sum is taken over all sequences $\sigma = (i_0, \ldots, i_s)$ with $1 = i_0 < \cdots < i_s = 2t+1$ (where $i_0, i_1, \ldots, i_s$ are required to be odd integers), and $|\sigma| = s$.

2.3 Two Useful Lemmas

The H-coefficient technique [CS14] is a very popular tool for bounding the statistical distance between two distributions (e.g. Eqn. (5)). Its core idea is to properly partition the set of attainable transcripts $\mathcal{T}$ into two disjoint sets, the good transcripts set $\mathcal{T}_1$ and the bad transcripts set $\mathcal{T}_2$. If for any $\tau \in \mathcal{T}_1$, we are able to obtain a lower bound (e.g. $1 - \varepsilon_1$) on the ratio $\Pr[\mathcal{T}_{\mathsf{real}} = \tau]/\Pr[\mathcal{T}_{\mathsf{ideal}} = \tau]$. And we can also obtain an upper bound (e.g. ε_2) on the value of $\Pr[\mathcal{T}_{\mathsf{ideal}} \in \mathcal{T}_2]$. The statistical distance is then bounded by $\varepsilon_1 + \varepsilon_2$. All of the above are formalized in the following lemma.

[4]Due to the uniqueness, we will interchangeably use the permutation-edge $\langle u, P_k, v\rangle$ and the input-output pair (u,v) under P_k.

Lemma 2 (H-Coefficient Technique, [CS14]). *Let E denote the target t-round KAC or KAC-type construction, and $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$ be the set of attainable transcripts. Assume that there exists a value $\varepsilon_1 > 0$ such that*

$$\frac{Pr[\mathcal{T}_{\mathsf{real}} = \tau]}{Pr[\mathcal{T}_{\mathsf{ideal}} = \tau]} \geq 1 - \varepsilon_1$$

holds for any $\tau \in \mathcal{T}_1$, and there exists a value $\varepsilon_2 > 0$ such that $Pr[\mathcal{T}_{\mathsf{ideal}} \in \mathcal{T}_2] \leq \varepsilon_2$. Then for any information-theoretic distinguisher $\mathcal{D}$, it has $\mathbf{Adv}^{\mathsf{SPRP}}_{E,t}(\mathcal{D}) \leq \varepsilon_1 + \varepsilon_2$.

To apply Lemma 2, the main task is usually to determine the value of ε_1. As we have argued in the previous section, it is essentially to calculate the crucial probability (7). The following lemma re-emphasizes this fact.

Lemma 3 (Lemma 2 of [Che+18]). *Let E denote the target t-round KAC or KAC-type construction, and $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}, K) \in \mathcal{T}$ be an attainable transcript, where K is the $(t+1)n$-bit key. We denote $p(\tau) = \Pr_{P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}[(E_K \downarrow \mathcal{Q}_0) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})]$, then*

$$\frac{Pr[\mathcal{T}_{\mathsf{real}} = \tau]}{Pr[\mathcal{T}_{\mathsf{ideal}} = \tau]} = (N)_{q_e} \cdot p(\tau).$$

3 Technical Overview

This section outlines the techniques used in security proofs of this work. We first review the known proof method, then propose a general transformation to simplify it, and finally give new proof strategies to further simplify security proofs and remove unnatural restrictions in the known result.

3.1 Proof Method of [Che+18]

The proof method for KAC-type constructions was originally proposed by Chen et al. [Che+18] in their analysis of the minimization of 2-round KAC. We note that [WYCD20] also follows this method and further refines it into an easy-to-use framework. Our approach is more closely inspired by that of [WYCD20] than by [Che+18].

At a high level, the proof method uses the H-coefficient technique (see Theorem 2), so the values of ε_1 and ε_2 need to be determined for good and bad transcripts, respectively. We focus here only on the main challenge, the value of ε_1, which is equivalent to the crucial probability (7) (see Lemma 3).

For a given construction and transcript (represented equivalently in graph view), we call a set of pairs of strings $A^{\equiv} = \{(\langle 0, a_1\rangle, \langle 2t+1, b_1\rangle), \ldots, (\langle 0, a_m\rangle, \langle 2t+1, b_m\rangle)\}$ a *uniform-structure-group*, if $\mathsf{Sh}(\mathsf{R}(a_1)) = \cdots = \mathsf{Sh}(\mathsf{R}(a_m)) < \mathsf{Sh}(\mathsf{L}(b_1)) = \cdots = \mathsf{Sh}(\mathsf{L}(b_m))$. Clearly, all pairs in $A^{\equiv}$ have a uniform structure in graph view, i.e., the numbers and locations of missing permutation-edges are the same for each pair of strings $(\langle 0, a_i\rangle, \langle 2t+1, b_i\rangle)$. We now give the general problem abstracted in [WYCD20], but slightly different to fit better here.

Definition 1 (Completing A Uniform-Structure-Group, [WYCD20]). *Consider a t-round KAC or KAC-type construction E, and fix arbitrarily an attainable transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}, K)$. Let $\mathcal{Q}_0^{\equiv} = \{(x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), \ldots, (x_{i_s}, y_{i_s})\} \subseteq \mathcal{Q}_0$ be a uniform-structure-group of plaintext-ciphertext pairs[5], then the problem is to evaluate the probability that $\mathcal{Q}_0^{\equiv}$ is completed (i.e. all plaintext-ciphertext pairs in $\mathcal{Q}_0^{\equiv}$ are connected), written as*

$$p_\tau(\mathcal{Q}_0^{\equiv}) = \Pr_{P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}[(E_K \downarrow \mathcal{Q}_0^{\equiv}) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})]. \quad (9)$$

For 3-round KACSP, [WYCD20] showed that the set $\mathcal{Q}_0$ can be divided into six disjoint uniform-structure-groups $\mathcal{Q}_{0,1}^{\equiv}, \mathcal{Q}_{0,2}^{\equiv}, \mathcal{Q}_{0,3}^{\equiv}, \mathcal{Q}_{0,4}^{\equiv}, \mathcal{Q}_{0,5}^{\equiv}, \mathcal{Q}_{0,6}^{\equiv}$, and the crucial probability (7) can be decomposed into six probabilities (in the form of (9)) associated with them. Then, all that remains is to find a good lower bound on the probability (9).

[5]Recall that x_i's and y_i's are by default in shore 0 and shore $2t+1$ respectively, so we use the simplified notation here.

It is shown in [WYCD20] that there exists a general framework for the task. To state it, we should first look at a useful concept called Core.

Definition 2 (Core, [WYCD20])**.** *For a complete path from x_j to y_j, we refer to the set of permutation-edges that make up the path as the* Core *of (x_j, y_j), and denote it as* $\mathsf{Core}(x_j, y_j)$. *That is,*

$$\mathsf{Core}(x_j, y_j) := \{\langle u, P_k, v\rangle : \langle u, P_k, v\rangle \textit{ is in the path from } x_j \textit{ to } y_j\}.$$

Similarly, when a uniform-structure-group $\mathcal{Q}_0^{\equiv}$ is completed, we can also define its Core *, i.e. the set of permutation-edges used to connect all plaintext-ciphertext pairs in $\mathcal{Q}_0^{\equiv}$, denoted as* $\mathsf{Core}(\mathcal{Q}_0^{\equiv})$. *That is,*

$$\mathsf{Core}(\mathcal{Q}_0^{\equiv}) := \bigcup_{(x_j,y_j)\in\mathcal{Q}_0^{\equiv}} \mathsf{Core}(x_j, y_j).$$

In order to illustrate the definition of Core more clearly, we also provide several concrete examples in Appendix B.

Note that the probability (9) is equivalent to counting all possible permutations $P_1, \dots, P_{t'}$ that complete $\mathcal{Q}_0^{\equiv}$ and also satisfy the known queries $\mathcal{Q}_1, \cdots, \mathcal{Q}_{t'}$. The idea of the general framework is to classify all such possible permutations $P_1, \dots, P_{t'}$, according to the number of new edges added to each round permutation (relative to the known $\mathcal{Q}_1, \cdots, \mathcal{Q}_{t'}$) in $\mathsf{Core}(\mathcal{Q}_0^{\equiv})$. Since the goal is to obtain a sufficiently large lower bound, a constructive approach can be used. In particular, for each sequence of the numbers of newly added edges in round permutations, we should construct as many permutations $P_1, \dots, P_{t'}$ as possible that complete $\mathcal{Q}_0^{\equiv}$ and satisfy these parameters. Summing up a sufficient number of sequences will give a desired lower bound.

More precisely, we let $\mathcal{P}_C = \{(P_1, \dots, P_{t'}) \in \mathcal{P}_n^{t'} : (E_K \downarrow \mathcal{Q}_0^{\equiv}) \wedge (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})\}$ denote the set of all permutations that complete $\mathcal{Q}_0^{\equiv}$ and extend respectively $\mathcal{Q}_1, \dots, \mathcal{Q}_{t'}$, and let $\mathcal{C} = \{\mathsf{Core}(\mathcal{Q}_0^{\equiv}) : \mathcal{Q}_0^{\equiv}$ *is completed by a sequence of round permutations* $(P_1, \dots, P_{t'}) \in \mathcal{P}_C\}$ denote the set of all possible Cores. For each $\widetilde{C} \in \mathcal{C}$, we can determine a tuple of numbers $(|\widetilde{C}_1|, |\widetilde{C}_2|, \dots, |\widetilde{C}_{t'}|)$, where $|\widetilde{C}_j|$ represents the number of *edges newly added* to $\mathcal{Q}_j$ in the $\widetilde{C}$. Then, we can give a more general form than the framework in [WYCD20] (i.e., setting $t' = 1$) as follows,

$$\begin{aligned}
(9) &= \Pr_{P_1,\dots,P_{t'} \leftarrow_\$ \mathcal{P}_n}[(E_K \downarrow \mathcal{Q}_0^{\equiv}) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})] \\
&= \frac{|\mathcal{P}_C|}{(N - |\mathcal{Q}_1|)! \times \cdots \times (N - |\mathcal{Q}_{t'}|)!} \\
&= \frac{\sum_{\widetilde{C}\in\mathcal{C}} \left|(P_1, \dots, P_{t'}) \in \mathcal{P}_C : \mathsf{Core}(\mathcal{Q}_0^{\equiv}) = \widetilde{C}\right|}{(N - |\mathcal{Q}_1|)! \times \cdots \times (N - |\mathcal{Q}_{t'}|)!} \\
&= \frac{\sum_{\widetilde{C}\in\mathcal{C}} \prod_{j=1}^{t'} (N - |\mathcal{Q}_j| - |\widetilde{C}_j|)!}{(N - |\mathcal{Q}_1|)! \times \cdots \times (N - |\mathcal{Q}_{t'}|)!} \\
&= \frac{\sum_{(m_1,m_2,\dots,m_{t'})} \left|\{\widetilde{C} \in \mathcal{C} : |\widetilde{C}_1| = m_1, \dots, |\widetilde{C}_{t'}| = m_{t'}\}\right| \times \prod_{j=1}^{t'} (N - |\mathcal{Q}_j| - m_j)!}{(N - |\mathcal{Q}_1|)! \times \cdots \times (N - |\mathcal{Q}_{t'}|)!} \\
&= \sum_{m_1} \cdots \sum_{m_{t'}} \frac{\left|\{\widetilde{C} \in \mathcal{C} : |\widetilde{C}_1| = m_1, \dots, |\widetilde{C}_{t'}| = m_{t'}\}\right|}{(N - |\mathcal{Q}_1|)_{m_1} \times \cdots \times (N - |\mathcal{Q}_{t'}|)_{m_{t'}}}.
\end{aligned} \tag{10}$$

As mentioned earlier, Eqn. (10) essentially turns the task into constructing as many Cores as possible for different tuples $(m_1, \dots, m_{t'})$, and then summing their results. In general, the framework can be carried out in three steps. The first step is to design a method that, for each given tuple $(m_1, \dots, m_{t'})$, ensures to generate Cores $\widetilde{C}$ satisfying $|\widetilde{C}_1| = m_1, \dots, |\widetilde{C}_{t'}| = m_{t'}$. The second step is then to count the possibilities that can be generated by the first step. And the third step is to perform a summation calculation, where a trick[6] of hypergeometric distribution (pioneered by [Che+18]) will be used.

[6] The terms arising from a (multivariate) hypergeometric distribution are introduced to help calculate a lower bound on the target probability, see Eqn. (30) for an example.

Note. It should be pointed out here that all proofs in this work are conducted under the guidance of this framework (i.e., Eqn. (10)). In particular, we showed that the key task of H-coefficient technique (i.e., Lemma 2) is to bound the probability (7) in the real world, which can then be reduced to bound the probabilities of the form (9). Therefore, the framework provides a *high-level intuition* that we can always accomplish the above task in three steps (for any KAC or KAC-type construction[7]): constructing Cores with specific cardinalities, counting the number of Cores and performing a summation calculation. When analyzing different constructions, such as the KACs (setting $t' = t$) and KACSPs (setting $t' = 1$), the subtle difference mainly lies in step 1, where the available constructive methods will be slightly different. In contrast, the detailed analysis and calculations in steps 2 and 3 are similar.

3.2 A General Transformation

We propose a general transformation to simplify the above proof method of [Che+18], such that only one probability (9) needs to be bounded. As we shall see, it does cut out a lot of tedious work and significantly simplify the proof. We apply this transformation to the security proofs of various constructions in this work.

For each pair (x_j, y_j), there are $r_j := \big(\mathsf{Sh}(\mathsf{L}(y_j)) - \mathsf{Sh}(\mathsf{R}(x_j)) + 1\big)/2$ undefined edges between x_j and y_j, where $r_j \in \{1, \ldots, t\}$ for a good transcript[8]. We call r_j the *actual distance* between x_j and y_j. We say that (x_i, y_i) is *farther* than (x_j, y_j) if $r_i > r_j$; or *closer* if $r_i < r_j$; or *equidistant*, otherwise. Clearly, all pairs in a uniform-structure-group are equidistant.

The idea of our general transformation is quite natural. First note that the set $\mathcal{Q}_0$ usually contains pairs with various actual distances, leading to the existence of multiple uniform-structure-groups. Just by intuition, the farther pair (x_i, y_i) feels more "hard" (conditionally, in fact) to connect than the closer pair (x_j, y_j), given the same available resources. After all, the former tends to consume more resources (e.g. new edges), so fewer edges can be freely defined. Assuming this argument holds, we can define a set $\widehat{\mathcal{Q}_0}$ satisfying $|\widehat{\mathcal{Q}_0}| = |\mathcal{Q}_0|$ and in which all pairs have the maximal actual distance t. That is, all the easier pairs in $\mathcal{Q}_0$ are replaced with the hardest ones, thus making $\widehat{\mathcal{Q}_0}$ itself a uniform-structure-group. Then, for the same known queries $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}$, it should have

$$\begin{aligned} \text{Eqn. (7)} &= \Pr_{P_1,\ldots,P_{t'} \leftarrow_{\$} \mathcal{P}_n}[(E_K \downarrow \mathcal{Q}_0) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})] \\ &\geq \Pr_{P_1,\ldots,P_{t'} \leftarrow_{\$} \mathcal{P}_n}[(E_K \downarrow \widehat{\mathcal{Q}_0}) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})] \stackrel{\text{def}}{=} p_\tau(\widehat{\mathcal{Q}_0}). \end{aligned} \tag{11}$$

Clearly, if we can obtain a good lower bound for $p_\tau(\widehat{\mathcal{Q}_0})$, it holds for the target crucial probability as well. The advantage of this treatment is that we only need to bound a single probability (9), namely $p_\tau(\widehat{\mathcal{Q}_0})$. Of course it comes at a price, so we need to keep the probability loss within an acceptable range. In short, this transformation can be seen as sacrificing a small amount of accuracy for great computational convenience.

All that remains is to find a method to transform closer pairs into farther ones, and make sure that they are less likely to be connected. We first point out that *the direct transformation* does not necessarily hold, although it is intuitively sound. Taking KAC as an example, we can know from the well-known Lemma 1 that the direct transformation does hold in the average case. However, it does not hold in the worst case, since counterexamples are not difficult to construct.

We next show that the direct transformation can be proved to hold, if a simple constraint is added on the replaced farther pairs. First of all, we say that a vertex u is connected to a vertex v *in the most wasteful way*[9], if all growing permutation-edges in the path are new (i.e. not defined before then) and

[7] In fact, the idea of this framework is quite general and it can be easily generalized to other constructions.

[8] The definition of good transcripts usually excludes the case where $r_j = 0$. Please note that we keep all key-edges in the graph view here for maximum generality.

[9] Intuitively, this kind of paths require the most new-edges and do not share any edges with other paths. In the words of [WYCD20], the most wasteful way actually means *sampling an exclusive element for each inner-node*. It had also been shown in [WYCD20] that such samples are easy to analyze. More concrete examples and analysis can be found in the security proofs, such as the Figure 1 and Appendix C.3.

each of them is used exactly once. Similarly, we can also connect a group of pairs of nodes in the most wasteful way, where all growing permutation-edges in these paths are new and each of them is used exactly once. The following is a *useful property*: for a given group of pairs, the number of new edges added to each round permutation P_j is fixed (denoted as m_j), among all possible paths generated in the most wasteful way. These numbers $m_1, \ldots, m_{t'}$ must be the maximum values (i.e. the number of missing edges between the group of pairs), determined by the construction and the number of pairs.

More formally, we give below the definition of the most wasteful-way (in the contex of plaintext-ciphertext pairs for ease of notation; other cases can be defined similarly).[10]

Definition 3 (The Most Wasteful Way)**.** *Consider a t-round KAC or KAC-type construction E, and fix arbitrarily the set of construction queries $\mathcal{Q}_0$ and the key K. Let $\mathcal{Q}'_k$ denote the set of all P_k-permutation-edges fixed so far, where $k = \{1, \ldots, t'\}$. Let $\tilde{\mathcal{Q}}_0 = \{(x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), \ldots, (x_{i_s}, y_{i_s})\} \subseteq \mathcal{Q}_0$ be a set of plaintext-ciphertext pairs to be connected, where $\mathsf{Sh}(\mathsf{R}(x_{i_j})) < \mathsf{Sh}(\mathsf{L}(y_{i_j}))$ for all $j \in \{1, \ldots, s\}$. We denote by m_k the total number of P_k-permutation-edges missing in the paths between all pairs in $\tilde{\mathcal{Q}}_0$ (given $\mathcal{Q}'_1, \ldots, \mathcal{Q}'_{t'}$), where $k = \{1, \ldots, t'\}$.*

Then, $\tilde{\mathcal{Q}}_0$ is said to be connected in the most wasteful way (with respect to $\mathcal{Q}'_1, \ldots, \mathcal{Q}'_{t'}$), if the Core *of the completed $\tilde{\mathcal{Q}}_0$ contains exactly m_k new P_k-permutation-edges compared to $\mathcal{Q}'_k$ for all $k \in \{1, \ldots, t'\}$.*

At this point, we are ready to describe our transformation from $\mathcal{Q}_0$ to $\widehat{\mathcal{Q}_0}$: *all pairs in $\mathcal{Q}_0$ whose actual distance is less than t are replaced with new pairs whose actual distance is equal to t, and it is required that these replaced new pairs must be connected in the most wasteful way.* The correctness of this transformation can be verified by repeatedly using the general Lemma 4, the proof of which is deferred to Appendix E.1.

Lemma 4 (The Closer The Easier)**.** *Consider a t-round ($t \geq 2$) KAC or KAC-type construction E, and fix arbitrarily the sets of known queries $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}$ and the key K.*

Let $A^{\equiv} = \{(x_1, y_1), \ldots, (x_s, y_s)\}$ be a uniform-structure-group of s plaintext-ciphertext pairs, where $\mathsf{Sh}(\mathsf{R}(x_1)) = \cdots = \mathsf{Sh}(\mathsf{R}(x_s)) = 3$ and $\mathsf{Sh}(\mathsf{L}(y_1)) = \cdots = \mathsf{Sh}(\mathsf{L}(y_s)) = 2t$. That is, the actual distance of each pair in $A^{\equiv}$ is $t - 1$.

Let $B^{\equiv} = \{(x'_1, y'_1), \ldots, (x'_s, y'_s)\}$ be a uniform-structure-group of s plaintext-ciphertext pairs, where $\mathsf{Sh}(\mathsf{R}(x'_1)) = \cdots = \mathsf{Sh}(\mathsf{R}(x'_s)) = 1$ and $\mathsf{Sh}(\mathsf{L}(y'_1)) = \cdots = \mathsf{Sh}(\mathsf{L}(y'_s)) = 2t$. That is, the actual distance of each pair in $B^{\equiv}$ is t.

Assume that $s \cdot t \leq |\mathcal{Q}_{i_2}|/2$ and $|U_{04}| \leq |\mathcal{Q}_{i_2}|/2$, where $\mathcal{Q}_{i_2}$ denotes the set of known queries to the second round permutation P_{i_2} (where $i_2 \in \{1, \ldots, t'\}$). If we both connect $A^{\equiv}$ and $B^{\equiv}$ in the most wasteful way, then the closer $A^{\equiv}$ is relatively easier. That is, for sufficiently large n, we have

$$\begin{aligned} &\Pr_{P_1, \ldots, P_{t'} \leftarrow_\$ \mathcal{P}_n}[(E_K \downarrow_w A^{\equiv}) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})] \\ \geq\ &\Pr_{P_1, \ldots, P_{t'} \leftarrow_\$ \mathcal{P}_n}[(E_K \downarrow_w B^{\equiv}) \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})], \end{aligned}$$

where $E_K \downarrow_w A^{\equiv}$ (resp. $E_K \downarrow_w B^{\equiv}$) denotes the event that $A^{\equiv}$ (resp. $B^{\equiv}$) is completed in the most wasteful way.

The Lemma 4 tells us that the closer pairs are easier to connect than the farther pairs, even if they are both in the wasteful way. Also note that the ordinary probability of connecting given pairs must be greater than when only the most wasteful way is allowed, since there may be other ways of connecting (e.g. reusing edges). Thus, our general transformation replaces the closer uniform-structure-group (whose connections are unrestricted) by a farther one that can only be connected in the most wasteful way, the connecting probability of course becoming smaller (i.e. Eqn. (11) holds). We should also stress that the assumptions $s \cdot t \leq |\mathcal{Q}_{i_2}|/2$ and $|U_{04}| \leq |\mathcal{Q}_{i_2}|/2$ are quite loose, and their only effect on the security proof is to add a few conditions to the definition of good transcripts. For convenience, we can simply ignore the assumptions, except that there is a negligible deviation in the value of ε_2. To see this more clearly, we first point out that the number of pairs that need to be replaced s is often much smaller than $|\mathcal{Q}_i|$ and the number of rounds t is a constant. In particular, the largest s encountered in

[10] It can be verified that the Examples 2 and 4 in Appendix B are both connected in the most wasteful way (we purposely assume $\mathcal{Q}_1 = \mathcal{Q}_2 = \emptyset$ over there to ensure that each permutation-edge fixed in the path(s) is new compared to $\mathcal{Q}_1$ and $\mathcal{Q}_2$).

the security proof for a t-round construction is $s = \mathcal{O}(|\mathcal{Q}_i|/N^{1/(t+1)})$. Second, since the expectation of $|U_{04}|$ is $|\mathcal{Q}_1| \cdot |\mathcal{Q}_{i_2}|/N$, the well-known Markov's inequality is sufficient to give a good upper bound on the probability $\Pr[|U_{04}| > |\mathcal{Q}_{i_2}|/2]$.

Finally, we illustrate how the general transformation can be applied in practical security proofs. The process is quite simple. Given a good transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}, K)$, we first partition the set $\mathcal{Q}_0$ into disjoint uniform-structure-groups, such as $\mathcal{Q}_{0,1}^{\equiv}, \ldots, \mathcal{Q}_{0,k}^{\equiv}$. Typically, there is only one uniform-structure-group, say $\mathcal{Q}_{0,k}^{\equiv}$, whose actual distance is t and $|\mathcal{Q}_{0,k}^{\equiv}| = |\mathcal{Q}_0| \cdot (1 - \mathcal{O}(\frac{1}{N^{t+1}}))$. That is, only about $s = \mathcal{O}(\frac{1}{N^{t+1}}) \cdot |\mathcal{Q}_0|$ plaintext-ciphertext pairs need to be replaced by the general transformation. We write *wlog* that $\mathcal{Q}_0 = \{(x_1, y_1), \ldots, (x_q, y_q)\}$ and $\mathcal{Q}_{0,k}^{\equiv} = \{(x_{s+1}, y_{s+1}), \ldots, (x_q, y_q)\}$. We first arbitrarily choose s right-free vertices $u_1, \ldots, u_s$ in the shore 1, and s left-free vertices $v_1, \ldots, v_s$ in the shore $2t$ (this always works since both s and $|\mathcal{Q}_i|$ are much smaller than N). Then, we define $(x_{q+i}, y_{q+i}) := (u_i \oplus \kappa_0, v_i \oplus \kappa_t)$ for $i = 1, \ldots, s$, and denote the set they form as $\mathcal{Q}_0^*$. Next, we set $\widehat{\mathcal{Q}_0} = \mathcal{Q}_{0,k}^{\equiv} \cup \mathcal{Q}_0^*$, i.e. $\widehat{\mathcal{Q}_0} = \{(x_{s+1}, y_{s+1}), \ldots, (x_q, y_q), (x_{q+1}, y_{q+1}), \ldots, (x_{q+s}, y_{q+s})\}$. It is easy to see that $\widehat{\mathcal{Q}_0}$ is indeed a uniform-structure-group with actual distance t. Please note that all the known queries $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}$ remain unchanged throughout. Also, don't forget that the last s pairs (i.e. $\mathcal{Q}_0^*$) must be connected in the most wasteful way. Lastly, the property of general transformation (see Eqn. (11)) allows us to focus only on the lower bound of the new probability

$$\begin{aligned} &\Pr_{P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \widehat{\mathcal{Q}_0} \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})] \\ &= \Pr_{P_1,\ldots,P_{t'} \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,k}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid (P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})], \end{aligned} \tag{12}$$

where $E_K \downarrow_w \mathcal{Q}_0^*$ denotes the event that the plaintext-ciphertext pairs in $\mathcal{Q}_0^*$ are connected in the most wasteful way.

3.3 New Proof Strategies

Although we are guided by the proof method of [Che+18], the low-level proof strategies are quite different.

We introduce a new notion of *recycled-edge*, while [WYCD20] only uses the *shared-edge*. Intuitively, our use of a recycled-edge means that an edge is recycled from the known queries (i.e. from $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}$) to build the path, so that one less new edge is added. Thus, recycled-edges serve the same purpose as shared-edges, i.e. to reduce the use of new edges when growing paths (relative to the most wasteful way). The difference between them is that the former reuses known edges, while the later reuses the newly added edges. We point out that recycled-edge has the following features compared to shared-edge. First, the analysis of recycled-edges is easier because each of recycled-edge involves only one path, whereas each shared-edge involves multiple paths. Second, the recycled-edge is less sensitive to the construction, and its analysis is relatively uniform in different constructions. In particular, it exists in the KAC construction where edges cannot be shared as in [WYCD20].

We provide new ideas to remove the two unnatural restrictions in the security bound of [WYCD20] (i.e., Thm. 5). The first restriction is the existence of an error term $\zeta(q_e)$, making it impossible to obtain a uniform bound for all q_e's. To get a good bound, [WYCD20] needs to choose an appropriate c for different values of q_e. In particular, it is unnatural that their bound does not converge to 0 as the number of queries q_e decreases to 0. Our observation is that this problem may be due to the nature of the hypergeometric distribution, whose variance is not a monotonic function. This leads to the fact that the tail bound obtained by Chebyshev's inequality (see Lemma 16) is also not monotonic, and thus only works well for part of the q_e's, e.g. $q_e = \omega(N^{1/2})$. A natural solution is to give a different proof for the range of $q_e = \mathcal{O}(N^{1/2})$. But one thing to note here is that we need to get a beyond-birthday-bound (i.e. $\mathcal{O}(N^{1/2+\epsilon})$-bound for $\epsilon > 0$), so that the bound is negligible for all $q_e = \mathcal{O}(N^{1/2})$. We found that the proof for $q_e = \omega(N^{1/2})$ can be adapted to the case of $q_e = \mathcal{O}(N^{1/2})$ just by modifying several constants defined in the proof (e.g., the values of M and M_0 in Section 4). Therefore, the security proofs in this work usually consider two cases, one is large $q_e = \omega(N^{1/2})$ and the other is small $q_e = \mathcal{O}(N^{1/2})$. Their proofs are almost identical except for slightly different calculations.

The second restriction is that it requires $q_p \le q_e/5$, where q_p and q_e are the number of permutation queries and construction queries respectively. This is an unnatural limitation on the access ability of distinguisher. After a lot of effort and calculation, we found that under the proof method of [Che+18], the main factor affecting the final security bound is the number of *variables*. Each variable is used to represent the number of new edges reduced in a Core (relative to the most wasteful way), and is denoted by h_i in our proofs. That is, more variables usually means a more accurate bound. It is important to note here that each variable actually corresponds to a constructive method of reducing new edges, and the results generated by these different methods are required to be disjoint. On the other hand, there seems to be an upper bound on the number of constructive methods of reducing new edges. Therefore, a big challenge is to perform a fine-grained analysis that allows us to find an appropriate number of variables to meet both requirements (i.e., accuracy and feasibility).

4 Improved Security Bound of $P_1P_1P_1$-Construction

4.1 Comparison of the Results

Known Result. Wu et al. [WYCD20] were the first to prove a tight security bound for the $P_1P_1P_1$-construction, and their proof was quite laborious.

Theorem 5 ($P_1P_1P_1$-Construction, Theorem 1 of [WYCD20])**.** *Consider the $P_1P_1P_1$-construction. Assume that $n \ge 32$ is sufficiently large, $\frac{28(q_e)^2}{N} \le q_p \le \frac{q_e}{5}$ and $2q_p + 5q_e \le \frac{N}{2}$, then for any $6 \le c \le \frac{N^{1/2}}{8}$, the following upper bound holds:*

$$\mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_1P_1}(\mathcal{D}) \le 98c \cdot \left(\frac{q_e}{N^{3/4}}\right) + 10c^2 \cdot \left(\frac{q_e}{N}\right) + \zeta(q_e), \tag{13}$$

where $\zeta(q_e) = \begin{cases} \frac{32}{c^2}, & \text{for } q_e \le \frac{c}{6}N^{1/2} \\ \frac{9N}{q_e^2}, & \text{for } q_e \ge \frac{7c}{6}N^{1/2} \end{cases}$ and $\mathcal{D}$ can be any distinguisher making q_e construction queries and q_p permutation queries.

It can be seen that the above security bound has two unnatural restrictions. The first is the error term $\zeta(q_e)$, where the entire range of q_e cannot be covered by a single value c. In particular, this term is non-negligible for small values of q_e, such as $q_e = \mathcal{O}(N^{1/2})$, making the security bound quite counter-intuitive. The second is the requirement on q_e and q_p, that is, $28(q_e)^2/N \le q_p \le q_e/5$, which is not a reasonable limit on the ability of distinguisher.

Our Result. Using the general transformation and new proof strategies outlined in Section 3, we obtain a neat security bound for the $P_1P_1P_1$-construction and the proof is much simpler.

Theorem 6 ($P_1P_1P_1$-Construction, Improved Bound)**.** *Consider the $P_1P_1P_1$-construction. For any distinguisher $\mathcal{D}$ making q_e construction queries and q_p permutation queries, the following upper bound holds:*

$$\mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_1P_1}(\mathcal{D}) \le \begin{cases} \frac{69q}{N^{3/4}} + \frac{125q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{78q}{N} + \frac{32N}{q^2}, & \text{for } q = \omega(N^{1/2}) \\ \frac{12q}{N^{7/10}} + \frac{125q^2}{N^{7/5}} + \frac{135q}{N^{3/4}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{32}{N^{1/10}}, & \text{for } q = \mathcal{O}(N^{1/2}) \end{cases}$$

where $q := \max\{q_e, q_p\}$.

In contrast to Theorem 5, our bound does give a negligible bound for all $q = \mathcal{O}(N^{1/2})$ (which is better than $\mathcal{O}(N^{2/3})$-bound but slightly worse than $\mathcal{O}(N^{3/4})$-bound), and has no restriction on the values of q_e and q_p. In fact, the bound for $q = \mathcal{O}(N^{1/2})$ can be easily improved to $\mathcal{O}(N^{3/4-\epsilon})$-bound for any $\epsilon > 0$, by modifying M to $q/N^{1/2-\epsilon}$ and M_0 to $q/N^{1/4+\epsilon}$, where M and M_0 are two constants to be defined in the proof. Even if we focus only on the large $q = \omega(N^{1/2})$, our bound is better than Eqn. (13) (for which the optimal $c = 6$ is set). Most importantly, the proof of Theorem 6 is simpler and can be found in Section 4.2.

Remarks. It should be pointed out that the tightness of our bound is with respect to attacks achieving constant probability, i.e., an adversary needs $q = \Omega(N^{3/4})$ queries to distinguish $P_1P_1P_1$-construction from random with a high advantage. The curve of our bound (i.e., roughly $(q^4/N^3)^{\frac{1}{4}}$) is not as sharp as the tigher bound (i.e., roughly q^4/N^3) achieved in the study of KACs (e.g., [HT16]).

We here show that the exact threshold of the two bounds in Theorem 6 can be determined. In fact, there are values of q that satisfy both bounds (for these q's, we can choose the better one at the time of use). More specifically, the first bound holds for all $q >= N^{1/2+\epsilon}$ for any $\epsilon > 0$, and the second bound holds for all $q \leq N^{11/20}/2$. Thus, any value in the interval $[N^{1/2+\epsilon}, N^{11/20}/2]$ (e.g., $N^{0.53}$) can be safely chosen as the threshold.

The main reason that leads us to discuss two cases is the Eqn. (35), where the magnitudes of MN and q^2 need to be compared. For more details, please refer to the calculation below Eqn. (34), which shows the analysis for all $q \geq N^{1/2+\epsilon}$. If we set $M = \frac{q}{N^{9/20}}$ here, then it can be verified that the second bound holds for $q \leq N^{11/20}/2$.

4.2 Proof of Theorem 6

As discussed in Section 3, we will consider two disjoint cases separately to remove the first restriction, namely the case $q = \omega(N^{1/2})$ and the case $q = \mathcal{O}(N^{1/2})$. For each case, the proof is guided by the proof method of [Che+18], thus using the H-coefficient technique (see Lemma 2) at a high level. Following the technique, we define the sets of good and bad transcripts, and then determine the values of ε_1 and ε_2, respectively. When calculating the value of ε_1, we apply the general transformation (see Eqn. (11)) so that only a single probability need to be considered. Finally, we address this single probability using the general framework (see Eqn. (10)) combined with our new proof strategies.

Preparatory Work. First, we point out the simple fact that for every distinguisher $\mathcal{D}$ that makes q_e construction queries and q_p permutation queries, there exists a $\mathcal{D}'$ making q construction queries and q permutation queries with at least the same distinguishing advantage, where $q = \max\{q_e, q_p\}$. We can just let $\mathcal{D}'$ simulate the queries of $\mathcal{D}$, and then perform additional $q - q_e$ construction queries and $q - q_p$ permutation queries, which obviously increases its advantage. For computational convenience, we consider the distinguisher $\mathcal{D}'$ that makes q construction queries and q permutation queries in the analysis. That is, for each attainable transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$, it has $|\mathcal{Q}_0| = |\mathcal{Q}_1| = q$.

To illustrate the key probability (7) of a good transcript, we can assume that there is no path of length 7 starting from $x_i \in \mathsf{Dom}(\mathcal{Q}_0)$ in shore 0 or ending at $y_i \in \mathsf{Ran}(\mathcal{Q}_0)$ in shore 7 (otherwise, it would be a bad transcript). Then, as in [WYCD20], the set $\mathcal{Q}_0$ can be partitioned into the following 6 uniform-structure-groups.

- Denote WLOG that $\mathcal{Q}_{0,1}^{\equiv} = \{(x_1, y_1), \ldots, (x_{\alpha_2}, y_{\alpha_2})\} \subset \mathcal{Q}_0$, where $\mathsf{Sh}(\mathsf{R}(x_i)) = 5$ and $\mathsf{Sh}(\mathsf{L}(y_i)) = 6$ for $i = 1, \ldots, \alpha_2$. That is, the actual distance of $\mathcal{Q}_{0,1}^{\equiv}$ is 1 and $|\mathcal{Q}_{0,1}^{\equiv}| = \alpha_2$. We also denote by $\mathsf{R}(\mathcal{Q}_{0,1}^{\equiv}) = \{\mathsf{R}(x_i) : i = 1, \ldots, \alpha_2\}, \mathsf{L}(\mathcal{Q}_{0,1}^{\equiv}) = \{\mathsf{L}(y_i) : i = 1, \ldots, \alpha_2\}$.
- Denote WLOG that $\mathcal{Q}_{0,2}^{\equiv} = \{(x_{\alpha_2+1}, y_{\alpha_2+1}), \ldots, (x_{\alpha_2+\beta_2}, y_{\alpha_2+\beta_2})\} \subset \mathcal{Q}_0$, where $\mathsf{Sh}(\mathsf{R}(x_i)) = 1$ and $\mathsf{Sh}(\mathsf{L}(y_i)) = 2$ for $i = \alpha_2 + 1, \ldots, \alpha_2 + \beta_2$. That is, the actual distance of $\mathcal{Q}_{0,2}^{\equiv}$ is 1 and $|\mathcal{Q}_{0,2}^{\equiv}| = \beta_2$. We also denote by $\mathsf{R}(\mathcal{Q}_{0,2}^{\equiv}) = \{\mathsf{R}(x_i) : i = \alpha_2 + 1, \ldots, \alpha_2 + \beta_2\}, \mathsf{L}(\mathcal{Q}_{0,2}^{\equiv}) = \{\mathsf{L}(y_i) : i = \alpha_2 + 1, \ldots, \alpha_2 + \beta_2\}$.
- Denote WLOG that $\mathcal{Q}_{0,3}^{\equiv} = \{(x_{\alpha_2+\beta_2+1}, y_{\alpha_2+\beta_2+1}), \ldots, (x_{\delta_2}, y_{\delta_2})\} \subset \mathcal{Q}_0$, where $\mathsf{Sh}(\mathsf{R}(x_i)) = 3$ and $\mathsf{Sh}(\mathsf{L}(y_i)) = 4$ for $i = \alpha_2 + \beta_2 + 1, \ldots, \delta_2$. That is, the actual distance of $\mathcal{Q}_{0,3}^{\equiv}$ is 1 and $|\mathcal{Q}_{0,3}^{\equiv}| := \gamma_2 = \delta_2 - \alpha_2 - \beta_2$. We also denote by $\mathsf{R}(\mathcal{Q}_{0,3}^{\equiv}) = \{\mathsf{R}(x_i) : i = \alpha_2 + \beta_2 + 1, \ldots, \delta_2\}, \mathsf{L}(\mathcal{Q}_{0,3}^{\equiv}) = \{\mathsf{L}(y_i) : i = \alpha_2 + \beta_2 + 1, \ldots, \delta_2\}$.
- Denote WLOG that $\mathcal{Q}_{0,4}^{\equiv} = \{(x_{\delta_2+1}, y_{\delta_2+1}), \ldots, (x_{\delta_2+\alpha_1}, y_{\delta_2+\alpha_1})\} \subset \mathcal{Q}_0$, where $\mathsf{Sh}(\mathsf{R}(x_i)) = 3$ and $\mathsf{Sh}(\mathsf{L}(y_i)) = 6$ for $i = \delta_2 + 1, \ldots, \delta_2 + \alpha_1$. That is, the actual distance of $\mathcal{Q}_{0,4}^{\equiv}$ is 2 and $|\mathcal{Q}_{0,4}^{\equiv}| = \alpha_1$.

We also denote by $\mathsf{R}(\mathcal{Q}^{\equiv}_{0,4}) = \{\mathsf{R}(x_i) : i = \delta_2+1,\ldots,\delta_2+\alpha_1\}, \mathsf{L}(\mathcal{Q}^{\equiv}_{0,4}) = \{\mathsf{L}(y_i) : i = \delta_2+1,\ldots,\delta_2+\alpha_1\}$.

- Denote WLOG that $\mathcal{Q}^{\equiv}_{0,5} = \{(x_{\delta_2+\alpha_1+1}, y_{\delta_2+\alpha_1+1}),\ldots,(x_{\delta_2+\delta_1}, y_{\delta_2+\delta_1})\} \subset \mathcal{Q}_0$, where $\mathsf{Sh}(\mathsf{R}(x_i)) = 1$ and $\mathsf{Sh}(\mathsf{L}(y_i)) = 4$ for $i = \delta_2+\alpha_1+1,\ldots,\delta_2+\delta_1$. That is, the actual distance of $\mathcal{Q}^{\equiv}_{0,5}$ is 2 and $|\mathcal{Q}^{\equiv}_{0,5}| := \beta_1 = \delta_1 - \alpha_1$. We also denote by $\mathsf{R}(\mathcal{Q}^{\equiv}_{0,5}) = \{\mathsf{R}(x_i) : i = \delta_2+\alpha_1+1,\ldots,\delta_2+\delta_1\}, \mathsf{L}(\mathcal{Q}^{\equiv}_{0,5}) = \{\mathsf{L}(y_i) : i = \delta_2+\alpha_1+1,\ldots,\delta_2+\delta_1\}$.
- Denote WLOG that $\mathcal{Q}^{\equiv}_{0,6} = \{(x_{\delta_2+\delta_1+1}, y_{\delta_2+\delta_1+1}),\ldots,(x_q, y_q)\} \subset \mathcal{Q}_0$, where $\mathsf{Sh}(\mathsf{R}(x_i)) = 1$ and $\mathsf{Sh}(\mathsf{L}(y_i)) = 6$ for $i = \delta_2+\delta_1+1,\ldots,q$. That is, the actual distance of $\mathcal{Q}^{\equiv}_{0,6}$ is 3 and $|\mathcal{Q}^{\equiv}_{0,6}| = \delta_0 = q-\delta_1-\delta_2$. We also denote by $\mathsf{R}(\mathcal{Q}^{\equiv}_{0,6}) = \{\mathsf{R}(x_i) : i = \delta_2+\delta_1+1,\ldots,q\}, \mathsf{L}(\mathcal{Q}^{\equiv}_{0,6}) = \{\mathsf{L}(y_i) : i = \delta_2+\delta_1+1,\ldots,q\}$.

It is easy to see that the crucial probability

$$(7) = \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_0 \mid P_1 \downarrow \mathcal{Q}_1] = \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[\bigwedge_{j=1}^{6} E_K \downarrow \mathcal{Q}^{\equiv}_{0,j} \mid P_1 \downarrow \mathcal{Q}_1]. \tag{14}$$

In [WYCD20], the probability (14) was decomposed into several conditional probabilities, which were quite cumbersome to analyze.

Applying General Transformation. We use the general transformation (see Eqn. (11)) here to reduce the task to bounding only one probability. The basic idea is to replace the uniform-structure-groups whose actual distance is less than 3 (i.e. $\mathcal{Q}^{\equiv}_{0,1}, \mathcal{Q}^{\equiv}_{0,2}, \mathcal{Q}^{\equiv}_{0,3}, \mathcal{Q}^{\equiv}_{0,4}, \mathcal{Q}^{\equiv}_{0,5}$) with a new uniform-structure-group whose actual distance is 3, and make the connecting probability smaller.

First note that when $q = \mathcal{O}(N^{3/4})$, the expectation of $\alpha_2, \beta_2, \gamma_2$ is $q^3/N^2 = \mathcal{O}(q/N^{1/2})$, and the expectation of α_1, β_1 is $q^2/N = \mathcal{O}(q/N^{1/4})$. Then, we denote $s = \delta_1+\delta_2 = \alpha_1+\beta_1+\alpha_2+\beta_2+\gamma_2 = \mathcal{O}(q/N^{1/4})$ as the number of pairs to be replaced. As discussed in Section 3, we take arbitrarily s vertices in shore 0 from the set $\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_0$ and denote them as $x_{q+1},\ldots,x_{q+s}$. We also take arbitrarily s vertices in shore $2t+1$ from the set $\{0,1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \setminus \mathsf{Ran}(\mathcal{Q}_1) \oplus \kappa_3$ and denote them as $y_{q+1},\ldots,y_{q+s}$. Then, we define the new uniform-structure-group $\mathcal{Q}^*_0 := \{(x_i, y_i) : i = q+1,\ldots,q+s\}$ and set $\widehat{\mathcal{Q}_0} := \mathcal{Q}^{\equiv}_{0,6} \cup \mathcal{Q}^*_0$, where the pairs in $\mathcal{Q}^*_0$ must be connected in the most wasteful way. Using Lemma 4 several times, we can know that

$$\begin{aligned}(14) &= \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[\bigwedge_{j=1}^{6} E_K \downarrow \mathcal{Q}^{\equiv}_{0,j} \mid P_1 \downarrow \mathcal{Q}_1] \\ &\geq \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \widehat{\mathcal{Q}_0} \mid P_1 \downarrow \mathcal{Q}_1] \\ &= \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}^{\equiv}_{0,6} \wedge E_K \downarrow_w \mathcal{Q}^*_0 \mid P_1 \downarrow \mathcal{Q}_1]. \end{aligned} \tag{15}$$

4.2.1 Case 1: $q = \omega(N^{1/2})$

We mainly focus on the large values of $q = \omega(N^{1/2})$, and the other case of $q = \mathcal{O}(N^{1/2})$ is similar. Let $M = \frac{q}{N^{1/2}}$ and $M_0 = \frac{q}{N^{1/4}}$. We first give the definition of good and bad transcripts.

Definition 4 (Bad and Good Transcripts, $P_1P_1P_1$-Construction). *For an attainable transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$, we say that τ is bad if $K \in \bigcup_{i=1}^{5} \mathsf{BadK}_i$; otherwise τ is good. The definitions of BadK_i are shown below:*

$K \in \mathsf{BadK}_1 \Leftrightarrow$ *there exists a path of length 7 starting from a vertex $x_i \in \mathsf{Dom}(\mathcal{Q}_0)$ in shore 0 or ending at a vertex $y_i \in \mathsf{Ran}(\mathcal{Q}_0)$ in shore 7*

$K \in \mathsf{BadK}_2 \Leftrightarrow \alpha_2 > M \vee \beta_2 > M \vee \gamma_2 > M \vee \alpha_1 > M_0 \vee \beta_1 > M_0$

$K \in \mathsf{BadK}_3 \Leftrightarrow$ $\mathsf{Dom}(\mathcal{Q}_1), \mathsf{R}(\mathcal{Q}^{\equiv}_{0,1}), \mathsf{R}(\mathcal{Q}^{\equiv}_{0,2}), \mathsf{R}(\mathcal{Q}^{\equiv}_{0,3})$ *are not pairwise disjoint* $\vee$ $\mathsf{Ran}(\mathcal{Q}_1), \mathsf{L}(\mathcal{Q}^{\equiv}_{0,1}), \mathsf{L}(\mathcal{Q}^{\equiv}_{0,2}), \mathsf{L}(\mathcal{Q}^{\equiv}_{0,3})$ *are not pairwise disjoint*

$$K \in \mathsf{BadK}_4 \;\Leftrightarrow\; \begin{cases} \;\;|\{x \in \mathsf{Dom}(\mathcal{Q}_0) : x \oplus \kappa_0 \oplus \kappa_1 \textit{ is not left-free}\}| > M_0 \\ \vee\, |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1) \cap (\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3)| > M_0 \\ \vee\, |\{y \in \mathsf{Ran}(\mathcal{Q}_0) : y \oplus \kappa_3 \oplus \kappa_2 \textit{ is not right-free}\}| > M_0 \\ \vee\, |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_2) \cap (\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0)| > M_0 \\ \vee\, |\{x \in \mathsf{Dom}(\mathcal{Q}_0) : x \oplus \kappa_0 \oplus \kappa_2 \textit{ is not left-free}\}| > M_0 \\ \vee\, |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2) \cap (\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3)| > M_0 \\ \vee\, |\{y \in \mathsf{Ran}(\mathcal{Q}_0) : y \oplus \kappa_3 \oplus \kappa_1 \textit{ is not right-free}\}| > M_0 \\ \vee\, |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_1) \cap (\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0)| > M_0 \end{cases}$$

$$K \in \mathsf{BadK}_5 \;\Leftrightarrow\; |U_{05}| > M_0 \vee |U_{27}| > M_0.$$

We can determine the value of $\varepsilon_2 = \frac{12q}{N^{3/4}} + \frac{3q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5}$ from the following lemma, the proof of which can be found in Appendix E.2.

Lemma 7 (Bad Transcripts, $q = \omega(N^{1/2})$). *For any given $\mathcal{Q}_0, \mathcal{Q}_1$ such that $|\mathcal{Q}_0| = |\mathcal{Q}_1| = q$, we have*

$$\Pr_{K \leftarrow_\$ \{0,1\}^{4n}}[\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \textit{ is bad}] \le \frac{12q}{N^{3/4}} + \frac{3q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5}.$$

The following lemma gives a lower bound on Eqn. (15) for any good transcript.

Lemma 8 (Good Transcripts, $q = \omega(N^{1/2})$). *Fix arbitrarily a good transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$ as defined in Definition 4. Let $\mathcal{Q}_{0,6}^{\equiv}$ and $\mathcal{Q}_0^*$ be as described in Eqn. (15), then we have*

$$\Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,6}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1] \ge \frac{1}{(N)_q} \times \Big(1 - \frac{57q}{N^{3/4}} - \frac{122q^2}{N^{3/2}} - \frac{78q}{N} - \frac{32N}{q^2}\Big). \quad (16)$$

Before giving the proof of Lemma 8, we first show how to obtain the final security bound from the above two lemmas. First note that (16) is also a lower bound on the crucial probability (7), i.e. $p(\tau)$ in Lemma 3 when $t = 3, t' = 1$. Then it is not difficult to determine the value of $\varepsilon_1 = \frac{57q}{N^{3/4}} + \frac{122q^2}{N^{3/2}} + \frac{78q}{N} + \frac{32N}{q^2}$. According to the H-coefficient technique (see Lemma 2), we can obtain

$$\begin{aligned} \mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_1P_1}(\mathcal{D}) &\le \varepsilon_1 + \varepsilon_2 \\ &= \frac{12q}{N^{3/4}} + \frac{3q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{57q}{N^{3/4}} + \frac{122q^2}{N^{3/2}} + \frac{78q}{N} + \frac{32N}{q^2} \\ &= \frac{69q}{N^{3/4}} + \frac{125q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{78q}{N} + \frac{32N}{q^2}, \end{aligned} \quad (17)$$

which is the result of large $q = \omega(N^{1/2})$ in Theorem 6.

Proof of Lemma 8. Let $\mathcal{Q}_0^{\equiv} = \widehat{\mathcal{Q}_0} := \mathcal{Q}_{0,6}^{\equiv} \cup \mathcal{Q}_0^*$ and $t = 3, t' = 1$, then the target probability is exactly an instantiation of the general problem (9). We apply the general framework (10) to bound it, so roughly in three steps.

The first step is to generate Cores **with specific numbers of new edges.** We will use four variables (denoted as h_1, h_2, h_3, h_4) to obtain a sufficiently accurate security bound, so four constructive methods of reducing new edges are needed.

The first method we use is called *recycled-edge-based method*, which exploits recycled-edges to reduce a specified number of new edges when building paths. Intuitively, when we construct a path connecting plaintext-ciphertext pair (x_i, y_i) with an actual distance of 3, the choice of the permutation-edge between shore 3 and 4 is quite free and can be "recycled" from the known edges in $\mathcal{Q}_1$ for use. Thus, we can construct the path with one less new edge. Furthermore, most of the known edges in

$\mathcal{Q}_1$ (about the proportion of $1 - \mathcal{O}(1/N^{1/4})$) can be used as recycled-edges. More details about the recycled-edge-based method can be found in Appendix C.1.

The other three methods we use are *shared-edge-based methods*, each of which exploits a different type of shared-edges to reduce a specified number of new edges when building paths. Intuitively, we consider two plaintext-ciphertext pairs together and let them share exactly 1 permutation edge. The two paths can then be connected with one less new edge than the most wasteful way. In particular, this work only considers shared-edges of this type, each of which saves 1 new edge for 2 paths. To distinguish, we refer to a shared-edge as (i,j)*-shared-edge*, where i and j represent the rounds that the shared-edge lies in two paths respectively. Note that the positions of the two paths are interchangeable, so (i,j)-shared-edges and (j,i)-shared-edges are essentially the same type. More details about the shared-edge-based methods can be found in Appendix C.2.

Recalling the Eqn. (15), our task is to connect the q pairs of $\widehat{\mathcal{Q}_0} = \mathcal{Q}_{0,6}^{\equiv} \cup \mathcal{Q}_0^*$ using a specified number of new edges, where $\mathcal{Q}_0^*$ is connected in the most wasteful way. Let h_1, h_2, h_3, h_4 be four integer variables in the interval $[0, M]$, where $M = \frac{q}{N^{1/2}}$ is a constant determined by q. We combine the recycled-edge-based method, the shared-edge-based methods and the most wasteful way to accomplish the task in five steps.

1. Select h_1 distinct pairs from $\mathcal{Q}_{0,6}^{\equiv}$, and connect each of these pairs using the recycled-edge-based method.

2. Apart from the h_1 pairs selected in Step 1, select $2h_2$ appropriate pairs from $\mathcal{Q}_{0,6}^{\equiv}$, and connect these pairs using the $(1,2)$-shared-edge-based method.

3. Apart from the $h_1 + 2h_2$ pairs selected in Steps 1 and 2, select $2h_3$ appropriate pairs from $\mathcal{Q}_{0,6}^{\equiv}$, and connect these pairs using the $(1,3)$-shared-edge-based method.

4. Apart from the $h_1 + 2h_2 + 2h_3$ pairs selected in Steps 1–3, select $2h_4$ appropriate pairs from $\mathcal{Q}_{0,6}^{\equiv}$, and connect these pairs using the $(2,3)$-shared-edge-based method.

5. Connect the remaining $\delta_0 - h_1 - \sum_{i=2}^{4} 2h_i$ pairs in $\mathcal{Q}_{0,6}^{\equiv}$ and the s pairs in $\mathcal{Q}_0^*$ in the most wasteful way.

Clearly, the above procedure must generate a $\mathsf{Core}(\widehat{\mathcal{Q}_0})$ containing exactly $3q - \sum_{i=1}^{4} h_i$ new edges, and all the pairs of $\mathcal{Q}_0^*$ are connected in the most wasteful way.

As mentioned in Section 3, the main factor affecting the final security bound is the number of variables. A simple explanation is that more variables make the multivariate hypergeometric distribution used in the calculations more tunable. That is why we define four variables h_1, h_2, h_3, h_4 here (i.e., to improve the accuracy), and it can be verified that these four methods necessarily produce different types of paths (i.e., to ensure the plausibility). Note that even considering only the shared-edge-based methods, our strategy is simpler than [WYCD20]. In particular, a single selection operation of theirs may generate three different types of shared-edges, whereas each of our selection operations will only generate shared-edges of the same type.

The second step is to evaluate the number of Cores that can be generated in the first step. According to the above procedure of connecting q plaintext-ciphertext pairs of $\widehat{\mathcal{Q}_0}$, we determine the number of possibilities for each step as follows. In the following, $\mathsf{RC}_i(j)$ denotes the *Range (set) of all possible Candidate values* for the to-be-assigned nodes in shore j (according to the constructive method used in Step i).

1. Since $|\mathcal{Q}_{0,6}^{\equiv}| = \delta_0$, it has $\binom{\delta_0}{h_1}$ possibilities to select h_1 distinct pairs from $\mathcal{Q}_{0,6}^{\equiv}$. After the h_1 pairs are chosen, we use the recycled-edge-based method to connect them by first determining a set $\mathsf{RC}_1(3)$ (the analysis of which is referred to the $\mathsf{RC}(3)$ in Appendix C.1) and choosing h_1 different u's from it, and then assigning one u to each pair. In total, the possibilities of Step 1 is at least $\binom{\delta_0}{h_1} \cdot (|\mathsf{RC}_1(3)|)_{h_1}$. (See Eqn.(20) for the definition of a subset of $\mathsf{RC}_1(3)$).

2. For simplicity, we can define a set of plaintext-ciphertext pairs $Z \subset \mathcal{Q}_{0,6}^{\equiv}$ (see Eqn. (18) for the definition of Z), so that the $2(h_2+h_3+h_4)$ distinct pairs in Step 2–4 can all be selected from Z. Then in Step 2, we have $\binom{|Z|}{h_2} \cdot \binom{|Z|-h_2}{h_2}$ possibilities to sequentially select h_2 distinct pairs from Z twice, where the first (resp. second) selected h_2 pairs will be constructed as the upper-paths (resp. lower-paths)[11] in the $(1,2)$-shared-edge-based method. We then use the $(1,2)$-shared-edge-based method to connect these $2h_2$ pairs. According to the discussion in Appendix C.2, the core task of $(1,2)$-shared-edge-based method is to determine two sets denoted by $\mathsf{RC}_2(2)$ and $\mathsf{RC}_2(4)$. By simple counting, the possibilities of Step 2 is at least $\frac{(|Z|)_{2h_2}}{h_2!} \cdot (|\mathsf{RC}_2(2)|)_{h_2} \cdot (|\mathsf{RC}_2(4)|)_{h_2}$, where $\frac{(|Z|)_{2h_2}}{h_2!} = \binom{|Z|}{h_2} \cdot \binom{|Z|-h_2}{h_2} \cdot h_2!$. (See Eqns.(21) and (22) for the definitions of subsets of $\mathsf{RC}_2(2)$ and $\mathsf{RC}_2(4)$; the constraints on $\mathsf{RC}_2(2)$ and $\mathsf{RC}_2(4)$ can be obtained by analyzing the $\widetilde{w}_{11}$ and $\widetilde{w}_{12}$ in Appendix D.1 respectively.)

3. For Step 3, we can select $2h_3$ distinct pairs from Z after removing the $2h_2$ pairs chosen in Step 2. Then, we have $\binom{|Z|-2h_2}{h_3} \cdot \binom{|Z|-2h_2-h_3}{h_3}$ possibilities to sequentially select h_3 distinct pairs from the rest of Z twice (similar to Step 2, the first and second selected h_3 pairs will play different roles). After the $2h_3$ pairs are chosen, we use the $(1,3)$-shared-edge-based method to connect them. According to an analysis similar to that in Appendix C.2, the core task of $(1,3)$-shared-edge-based method is also to determine two sets denoted by $\mathsf{RC}_3(4)$ and $\mathsf{RC}_3(2)$. By simple counting, the possibilities of Step 3 is at least $\frac{(|Z|-2h_2)_{2h_3}}{h_3!} \cdot (|\mathsf{RC}_3(2)|)_{h_3} \cdot (|\mathsf{RC}_3(4)|)_{h_3}$. (See Eqns.(23) and (24) for the definitions of subsets of $\mathsf{RC}_3(4)$ and $\mathsf{RC}_3(2)$; the constraints on $\mathsf{RC}_3(2)$ and $\mathsf{RC}_3(4)$ can be obtained by analyzing the $\widetilde{w}_{22}$ and $\widetilde{w}_{21}$ in Appendix D.2 respectively.)

4. For Step 4, we can select $2h_4$ distinct pairs from Z after removing the $2(h_2+h_3)$ pairs chosen in Step 2 and Step 3. Then, we have $\binom{|Z|-2h_2-2h_3}{h_4} \cdot \binom{|Z|-2h_2-2h_3-h_4}{h_4}$ possibilities to sequentially select h_4 distinct pairs from the rest of Z twice (similar to Step 2, the first and second selected h_4 pairs will play different roles). After the $2h_4$ pairs are chosen, we use the $(2,3)$-shared-edge-based method to connect them. According to an analysis similar to that in Appendix C.2, the core task of $(2,3)$-shared-edge-based method is to determine two sets denoted by $\mathsf{RC}_4(4)$ and $\mathsf{RC}_4(2)$. By simple counting, the possibilities of Step 4 is at least $\frac{(|Z|-2h_2-2h_3)_{2h_4}}{h_4!} \cdot (|\mathsf{RC}_4(2)|)_{h_4} \cdot (|\mathsf{RC}_4(4)|)_{h_4}$. (See Eqns.(25) and (26) for the definitions of subsets of $\mathsf{RC}_4(4)$ and $\mathsf{RC}_4(2)$; the constraints on $\mathsf{RC}_4(2)$ and $\mathsf{RC}_4(4)$ can be obtained by analyzing the $\widetilde{w}_{32}$ and $\widetilde{w}_{31}$ in Appendix D.3 respectively.)

5. Step 5 is to connect the remaining $(\delta_0 - h_1 - \sum_{i=2}^{4} 2h_i)$ pairs in $\mathcal{Q}_{0,6}^{\equiv}$ and the s pairs in $\mathcal{Q}_0^*$ in the most wasteful way. According to the analysis in Appendix C.3, we can determine a set $\mathsf{RC}_5(2)$ and choose $(\delta_0 - h_1 - \sum_{i=2}^{4} 2h_i) + s = q - h_1 - \sum_{i=2}^{4} 2h_i$ different $w_{3,2}$'s from it, and assign one $w_{3,2}$ to each pair; then determine a set $\mathsf{RC}_5(4)$ and choose $q - h_1 - \sum_{i=2}^{4} 2h_i$ different $w_{3,4}$'s from it, and then assign one $w_{3,4}$ to each pair. In total, the possibilities of Step 5 is at least $(|\mathsf{RC}_5(2)|)_{q-h_1-\sum_{i=2}^{4} 2h_i} \cdot (|\mathsf{RC}_5(4)|)_{q-h_1-\sum_{i=2}^{4} 2h_i}$. (See Eqns.(27) and (28) for the definitions of subsets of $\mathsf{RC}_5(2)$ and $\mathsf{RC}_5(4)$.)

All that's left is to give a lower bound on the cardinality for Z and each $\mathsf{RC}_j(i)$ mentioned above. Let Λ_1 denote the set of h_1 pairs selected from $\mathcal{Q}_{0,6}^{\equiv}$ in Step 1. We first give the definition of set Z below[12], and denote by $|Z| = q_0$.

$$\begin{aligned} Z := \{(x_i, y_i) \in \mathcal{Q}_{0,6}^{\equiv} \setminus \Lambda_1 : \ & x_i \notin \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \oplus \kappa_3 \wedge x_i \notin \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \oplus \kappa_3 \\ & \wedge y_i \notin \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \oplus \kappa_3 \wedge y_i \notin \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \oplus \kappa_3 \wedge x_i \oplus \kappa_0 \oplus \kappa_1 \text{ is left-free} \\ & \wedge x_i \oplus \kappa_0 \oplus \kappa_2 \text{ is left-free} \wedge y_i \oplus \kappa_1 \oplus \kappa_3 \text{ is right-free} \wedge y_i \oplus \kappa_2 \oplus \kappa_3 \text{ is right-free}\} \end{aligned} \tag{18}$$

[11] In Figure 1, the paths between (x_2, y_2) and (x_2', y_2') are called the *upper-path* and *lower-path*, respectively.

[12] See Appendix D for an analysis of the constraints on Z, which are the sum of constraints from the three shared-edge-based methods.

From the BadK_4 in Defn. 4, we can know that

$$q_0 = |Z| \geq \delta_0 - h_1 - 8M_0. \tag{19}$$

Based on the analysis in Appendix C.1–C.3, we proceed to lower-bound the cardinality of each $\mathsf{RC}_j(i)$ as follows.

$$\begin{aligned} |\mathsf{RC}_1(3)| &\geq \left|\mathsf{Dom}(\mathcal{Q}_1) \setminus S_1 \setminus S_2\right| \\ &\geq q - 2M_0, \end{aligned} \tag{20}$$

since $|S_1| = |U_{05}| \leq M_0, |S_2| = |U_{27}| \leq M_0$ hold in any good transcript (see BadK_5 in Defn. 4).

$$\begin{aligned} |\mathsf{RC}_2(2)| &\geq \left|\{0,1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus V \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus U \oplus \kappa_1 \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus U \oplus \kappa_2\right| \\ &\geq \left|\{0,1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Ran}(\mathcal{Q}_1) \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_1 \right. \\ &\qquad \left. \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_2\right| - 3 \cdot (2h_1) \\ &\geq N - 6q - 6h_1, \end{aligned} \tag{21}$$

where U (resp. V) denotes the domain (resp. range) of all P_1-input-output-pairs fixed so far (i.e., after Step 1) and $3 \cdot (2h_1) = 6h_1$ is the maximum number[13] of new values generated by Step 1 that fall within the constraints of $\mathsf{RC}_2(2)$. This is exactly the consequence of updating U, V discussed in Appendix C.2. Due to the similarity, we directly give the remaining lower bounds without explanation.

$$\begin{aligned} |\mathsf{RC}_2(4)| &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus U \oplus \kappa_2 \setminus V\right| \\ &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_1)\right| - 2(2h_1 + 5h_2) \\ &\geq N - 4q - 4h_1 - 10h_2, \end{aligned} \tag{22}$$

$$\begin{aligned} |\mathsf{RC}_3(4)| &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus U \oplus \kappa_2 \setminus V\right| \\ &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_1)\right| - 2(2h_1 + 5h_2) \\ &\geq N - 4q - 4h_1 - 10h_2, \end{aligned} \tag{23}$$

$$\begin{aligned} |\mathsf{RC}_3(2)| &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus U \oplus \kappa_1 \setminus V\right| \\ &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_1)\right| \\ &\qquad - 2(2h_1 + 5h_2 + 5h_3) \\ &\geq N - 4q - 4h_1 - 10h_2 - 10h_3, \end{aligned} \tag{24}$$

$$\begin{aligned} |\mathsf{RC}_4(4)| &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus U \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus V \right. \\ &\qquad \left. \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_1 \oplus \kappa_2 \setminus V \oplus \kappa_1 \oplus \kappa_2\right| \\ &\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Ran}(\mathcal{Q}_1) \right. \\ &\qquad \left. \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_1 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_1) \oplus \kappa_1 \oplus \kappa_2\right| - 3(2h_1 + 5h_2 + 5h_3) \\ &\geq N - 6q - 6h_1 - 15h_2 - 15h_3, \end{aligned} \tag{25}$$

[13]Note that Step 1 will generate $2h_1$ new permutation-edges, so there will be $2h_1$ new elements added to U and V respectively (compared to $\mathsf{Dom}(\mathcal{Q}_1)$ and $\mathsf{Ran}(\mathcal{Q}_1)$). It can be seen that there are only three constraints related to U and V in Eqn. (21), $6h_1$ is obviously the maximum number of changes. We need to point out that this is actually an overestimation. For example, newly added permutation-edges in Step 1 of the form $\langle x_i \oplus \kappa_0, P_1, * \rangle$ cause the set $U \oplus \kappa_1$ to add new elements (i.e., $x_i \oplus \kappa_0 \oplus \kappa_1$) which are already included in $\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1$. A finer analysis could provide more accurate results, but this simplified treatment is sufficient here since we are not seeking to optimize the constant coefficients in security bounds. Also, we use this easily verifiable overestimation in the evaluation of Eqns. (22) – (27) below.

$$|\mathsf{RC}_4(2)| \geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus U \oplus \kappa_1 \setminus V\right| \tag{26}$$
$$\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_1)\right|$$
$$- 2(2h_1 + 5h_2 + 5h_3 + 5h_4)$$
$$\geq N - 4q - 4h_1 - 10h_2 - 10h_3 - 10h_4,$$

$$|\mathsf{RC}_5(2)| \geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus U \oplus \kappa_1 \setminus V\right| \tag{27}$$
$$\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_1 \setminus \mathsf{Ran}(\mathcal{Q}_1)\right|$$
$$- 2(2h_1 + 5h_2 + 5h_3 + 5h_4)$$
$$\geq N - 4q - 4h_1 - 10h_2 - 10h_3 - 10h_4,$$

$$|\mathsf{RC}_5(4)| \geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus U \oplus \kappa_2 \setminus V\right| \tag{28}$$
$$\geq \left|\{0,1\}^n \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus \mathsf{Dom}(\mathcal{Q}_1) \oplus \kappa_2 \setminus \mathsf{Ran}(\mathcal{Q}_1)\right|$$
$$- 2(2h_1 + 5h_2 + 5h_3 + 5h_4) \underbrace{-(q - h_1 - 2h_2 - 2h_3 - 2h_4)}_{\text{See footnote}^{14}}$$
$$\geq N - 5q - 3h_1 - 8h_2 - 8h_3 - 8h_4.$$

Let $\#\mathsf{Cores}_i$ denote the number of $\mathsf{Cores}(\widehat{\mathcal{Q}_0})$ containing exactly i new edges (relative to $\mathcal{Q}_1$). Combining all the above, we finally obtain that

$$\begin{aligned}
&\#\mathsf{Cores}_{3q-\sum_{i=1}^4 h_i} \\
&\geq \binom{\delta_0}{h_1} \cdot (|\mathsf{RC}_1(3)|)_{h_1} \cdot \frac{(|Z|)_{2h_2+2h_3+2h_4}}{h_2! \cdot h_3! \cdot h_4!} \cdot (|\mathsf{RC}_2(2)|)_{h_2} \cdot (|\mathsf{RC}_2(4)|)_{h_2} \cdot (|\mathsf{RC}_3(4)|)_{h_3} \cdot (|\mathsf{RC}_3(2)|)_{h_3} \\
&\quad \times (|\mathsf{RC}_4(4)|)_{h_4} \cdot (|\mathsf{RC}_4(2)|)_{h_4} \cdot (|\mathsf{RC}_5(2)|)_{q-h_1-2h_2-2h_3-2h_4} \cdot (|\mathsf{RC}_5(4)|)_{q-h_1-2h_2-2h_3-2h_4} \\
&\left.\begin{aligned}
&\geq \frac{(\delta_0)_{h_1}(q-2M_0)_{h_1}}{h_1!} \cdot \frac{(q_0)_{2h_2+2h_3+2h_4}}{h_2! \cdot h_3! \cdot h_4!} \cdot (N-6q-6h_1)_{h_2} \cdot (N-4q-4h_1-10h_2)_{h_2} \\
&\quad \cdot (N-4q-4h_1-10h_2)_{h_3} \cdot (N-4q-4h_1-10h_2-10h_3)_{h_3} \\
&\quad \cdot (N-6q-6h_1-15h_2-15h_3)_{h_4} \cdot (N-4q-4h_1-10h_2-10h_3-10h_4)_{h_4} \\
&\quad \cdot (N-4q-4h_1-10h_2-10h_3-10h_4)_{q-h_1-2h_2-2h_3-2h_4} \\
&\quad \cdot (N-5q-3h_1-8h_2-8h_3-8h_4)_{q-h_1-2h_2-2h_3-2h_4}.
\end{aligned}\right\}
\end{aligned} \tag{29}$$

The third step is to perform the summation calculation. Since the lower bound on $\#\mathsf{Cores}_{3q-\sum_{i=1}^4 h_i}$ is known, we are now ready to calculate the final result. From the Eqns. (10) and (29), we have

$$\begin{aligned}
(15) =& \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,6}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1] \\
\geq& \sum_{0 \leq h_1, \ldots, h_4 \leq M} \frac{\#\mathsf{Cores}_{3q-\sum_{i=1}^4 h_i}}{(N-q)_{3q-\sum_{i=1}^4 h_i}} \\
=& \frac{1}{(N)_q} \times \sum_{0 \leq h_1, \ldots, h_4 \leq M} \frac{(\delta_0)_{h_1} \cdot (q-2M_0)_{h_1}}{h_1!} \times \frac{(q_0)_{\sum_{i=2}^4 2h_i}}{\prod_{i=2}^4 h_i!}
\end{aligned}$$

[14] Based on the analysis in Appendix C.3, after we use the elements in $\mathsf{RC}_5(2)$ to assign the $(q - h_1 - 2h_2 - 2h_3 - 2h_4)$ nodes $w_{3,2}$'s, there will be $(q - h_1 - 2h_2 - 2h_3 - 2h_4)$ new permutation-edges of the form $\langle x_i \oplus \kappa_0, P_1, w_{3,2}\rangle$ generated. It can be seen that the elements newly added to set U (all in the form of $x_i \oplus \kappa_0$) are already included in the set $\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0$. Therefore, although there are two constraints related to U, V in Eqn. (28), here we only need to subtract the value $(q - h_1 - 2h_2 - 2h_3 - 2h_4)$.

$$
\begin{aligned}
&\times (N-6q-6h_1)_{h_2} \cdot (N-4q-4h_1-10h_2)_{h_2} \\
&\times (N-4q-4h_1-10h_2)_{h_3} \cdot (N-4q-4h_1-10h_2-10h_3)_{h_3} \\
&\times (N-6q-6h_1-15h_2-15h_3)_{h_4} \cdot (N-4q-4h_1-10h_2-10h_3-10h_4)_{h_4} \\
&\times \frac{(N)_q (N-4q-4h_1-\sum_{j=2}^4 10h_j)_{q-h_1-\sum_{i=2}^4 2h_i} (N-5q-3h_1-\sum_{j=2}^4 8h_j)_{q-h_1-\sum_{i=2}^4 2h_i}}{(N-q)_{3q-\sum_{i=1}^4 h_i}} \\
&\times \underbrace{\frac{\prod_{i=1}^4 h_i! \cdot (N)_q}{(q)_{\sum_{i=1}^4 h_i} \cdot \prod_{i=1}^4 (q)_{h_i} \cdot (N-4q)_{q-\sum_{i=1}^4 h_i}} \cdot \mathsf{MHyp}_{N,q,q,q,q}(h_1,\ldots,h_4)}_{=1 \text{ (see Definition 6)}} \qquad (30)
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{(N)_q} \times \sum_{0\le h_1,\ldots,h_4 \le M} \mathsf{MHyp}_{N,q,q,q,q}(h_1,\ldots,h_4) \qquad (31)\\
&\times \frac{(\delta_0)_{h_1} \cdot (q-2M_0)_{h_1} \cdot (q_0)_{\sum_{i=2}^4 2h_i}}{\prod_{i=1}^4 h_i!} \times \frac{\prod_{i=1}^4 h_i!}{(q)_{\sum_{i=1}^4 h_i} \prod_{i=1}^4 (q)_{h_i}} \qquad (32)\\
&\times \frac{(N)_q (N-4q-4h_1-\sum_{i=2}^4 10h_j)_q (N-5q-3h_1-\sum_{i=2}^4 8h_j)_q (N)_q}{(N-q)_{3q}(N-4q)_q} \qquad (33)\\
&\left.\begin{aligned}
&\times \frac{(N-6q-6h_1-\sum_{j=2}^3 15h_j)_{h_2+h_4} (N-4q-4h_1-\sum_{j=2}^4 10h_j)_{h_2+2h_3+h_4}}{(N-5q-\sum_{j=2}^4 2h_j)_{h_1+\sum_{i=2}^4 2h_i} (N-6q+h_1)_{h_1+\sum_{i=2}^4 2h_i}} \\
&\cdot (N-5q+\sum_{j=1}^4 h_j)_{\sum_{i=1}^4 h_i} \cdot (N-4q+\sum_{j=1}^4 h_j)_{\sum_{i=1}^4 h_i}.
\end{aligned}\right\} \qquad (34)
\end{aligned}
$$

We next calculate the lower bound for each of the Eqns. (31)–(34) as follows.

$$
\begin{aligned}
(31) \ge& \frac{1}{(N)_q} \times \left(1 - \sum_{i=1}^4 \sum_{h_i > M} \mathsf{MHyp}_{N,q,q,q,q}(h_1,\ldots,h_4)\right) \\
=& \frac{1}{(N)_q} \times \left(1 - \sum_{i=1}^4 \sum_{h_i > M} \mathsf{Hyp}_{N,q,q}(h_i)\right) \\
\ge& \frac{1}{(N)_q} \times \left(1 - \frac{4q^2(N-q)^2}{(MN-q^2)^2(N-1)}\right) \qquad (35)\\
\ge& \frac{1}{(N)_q} \times \left(1 - \frac{4N^2}{(q-N^{1/2})^2(N-1)}\right) \\
\ge& \frac{1}{(N)_q} \times \left(1 - \frac{32N}{q^2}\right),
\end{aligned}
$$

where we use Lemma 16 for the second inequality, and the third inequality holds by substituting $M = \frac{q}{N^{1/2}}$, and the last inequality uses the fact that $q = \omega(N^{1/2}), N^{1/2} \le \frac{q}{2}, N-1 \ge \frac{N}{2}$ for sufficiently large n.

$$
\begin{aligned}
(32) =& \frac{(\delta_0)_{h_1} \cdot (q-2M_0)_{h_1} \cdot (q_0)_{\sum_{i=2}^4 2h_i}}{(q)_{\sum_{i=1}^4 h_i} \prod_{i=1}^4 (q)_{h_i}} \\
\ge& \frac{(\delta_0)_{h_1}(q-2M_0)_{h_1}}{(q)_{h_1}(q)_{h_1}} \cdot \frac{(q_0)_{2h_2}}{(q)_{h_2}(q)_{h_2}} \cdot \frac{(q_0-2h_2)_{2h_2}}{(q)_{h_2}(q)_{h_2}} \cdot \frac{(q_0-2h_2-2h_3)_{2h_3}}{(q)_{h_3}(q)_{h_3}} \\
\ge& \prod_{i=0}^{h_1-1} \left(1 - \frac{\delta_1+\delta_2}{q-i}\right) \cdot \left(1 - \frac{2M_0}{q-i}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\times \prod_{i=2}^{4}\prod_{j=0}^{h_i-1}\Big(1-\frac{\sum_{l=2}^{i-1}2h_l+h_1+8M_0+\delta_1+\delta_2}{q-j}\Big)\cdot\Big(1-\frac{\sum_{l=2}^{i-1}2h_l+h_i+h_1+8M_0+\delta_1+\delta_2}{q-j}\Big)\\
\geq&1-\frac{h_1(\delta_1+\delta_2+2M_0)}{q-\sum_{i=1}^{4}h_i}-\frac{(\sum_{i=2}^{4}h_i)(2h_1+16M_0+2\delta_1+2\delta_2)}{q-\sum_{i=1}^{4}h_i}-\frac{\sum_{i=2}^{4}h_i(\sum_{j=2}^{i-1}4h_j+h_i)}{q-\sum_{i=1}^{4}h_i}\\
\geq&1-\frac{M(3M_0+M)}{q-\sum_{i=1}^{4}h_i}-\frac{3M(4M+18M_0)}{q-\sum_{i=1}^{4}h_i}-\frac{15M^2}{q-\sum_{i=1}^{4}h_i}\\
\geq&1-\frac{57MM_0}{q}-\frac{28M^2}{q}\\
=&1-\frac{57q}{N^{3/4}}-\frac{28q}{N},
\end{aligned}
$$

where the second inequality hold due to $\delta_0=q-(\delta_1+\delta_2)$ and $q_0=\delta_0-h_1-8M_0$, the forth and last inequalities hold due to the fact that $h_1,\ldots,h_4\leq M,\delta_2\leq M,\delta_1\leq M_0,\sum_{i=1}^{4}h_i\leq\frac{q}{2}$, and the last equality holds by substituting $M=\frac{q}{N^{1/2}},M_0=\frac{q}{N^{1/4}}$.

$$
\begin{aligned}
(33)=&\frac{(N)_q(N-4q-4h_1-\sum_{i=2}^{4}10h_j)_q(N-5q-3h_1-\sum_{i=2}^{4}8h_j)_q(N)_q}{(N-q)_q(N-2q)_q(N-3q)_q(N-4q)_q}\\
=&\frac{(N)_q(N-4q)_q(N-5q)_q(N)_q}{(N-q)_q(N-2q)_q(N-2q)_q(N-4q)_q}\\
&\times\frac{(N-4q-4h_1-\sum_{i=2}^{4}10h_j)_q(N-5q-3h_1-\sum_{i=2}^{4}8h_j)_q}{(N-4q)_q(N-5q)_q}\\
=&\prod_{i=0}^{q-1}\Big(\underbrace{1+\frac{(N-i)^3q+o\big((N-i)^3q\big)}{(N-i-q)(N-i-2q)(N-i-3q)(N-i-4q)}}_{>1}\Big)\\
&\times\prod_{i=0}^{q-1}\Big(1-\frac{4h_1+\sum_{i=2}^{4}10h_j}{N-i-4q}\Big)\cdot\prod_{i=0}^{q-1}\Big(1-\frac{3h_1+\sum_{i=2}^{4}8h_j}{N-i-5q}\Big)\\
\geq&\Big(1-\frac{34Mq}{N-5q}\Big)\cdot\Big(1-\frac{27Mq}{N-6q}\Big)\\
\geq&1-\frac{122Mq}{N}\\
=&1-\frac{122q^2}{N^{3/2}},
\end{aligned}
$$

where the first and second inequalities hold due to the fact that $h_1,\ldots,h_4\leq M$ and $5q<6q<\frac{N}{2}$, and the last equality holds by substituting $M=\frac{q}{N^{1/2}}$.

$$
\begin{aligned}
(34)\geq&\underbrace{\frac{(N-5q+\sum_{j=1}^{4}h_j)_{\sum_{i=1}^{4}h_i}}{(N-5q-3h_1-\sum_{j=2}^{4}8h_j)_{\sum_{i=1}^{4}h_i}}\cdot\frac{(N-4q-4h_1-\sum_{j=2}^{4}10h_j)_{\sum_{i=2}^{4}h_i}}{(N-5q-4h_1-\sum_{j=2}^{4}9h_j)_{\sum_{i=2}^{4}h_i}}}_{>1}\\
&\times\underbrace{\frac{(N-4q+\sum_{j=1}^{4}h_j)_{\sum_{i=1}^{4}h_i}}{(N-6q-2h_1-7h_2-6h_3-7h_4)_{\sum_{i=2}^{4}h_i}}\cdot\frac{(N-4q-4h_1-\sum_{j=2}^{4}11h_j)_{h_3}}{(N-6q-3h_1-8h_2-7h_3-8h_4)_{h_3}}}_{>1}\\
&\times\frac{(N-6q-6h_1)_{h_2}(N-6q-6h_1-15h_2-15h_3)_{h_4}}{(N-6q-2h_1-\sum_{j=2}^{4}6h_j)_{h_2+h_4}}\\
\geq&\prod_{i=0}^{h_2-1}\Big(1-\frac{4h_1}{N-i-6q-2h_1-\sum_{j=2}^{4}6h_j}\Big)\cdot\prod_{i=0}^{h_4-1}\Big(1-\frac{3h_1+9h_3+9h_3}{N-i-6q-3h_1-\sum_{j=2}^{4}6h_j}\Big)
\end{aligned}
$$

$$\geq 1 - \frac{50M^2}{N}$$
$$\geq 1 - \frac{50q}{N}$$

where the third inequalities hold due to the fact that $h_1, \ldots, h_4 \leq M$ and $6q + 3h_1 + \sum_{j=2}^{4} 6h_j < \frac{N}{2}$, and the last equality holds by substituting $M = \frac{q}{N^{1/2}}$.

Putting all pieces together, we obtain that

$$\begin{aligned}(15) =& \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,6}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1] \\ \geq& (31) \times (32) \times (33) \times (34) \\ \geq& \frac{1}{(N)_q} \times \left(1 - \frac{57q}{N^{3/4}} - \frac{122q^2}{N^{3/2}} - \frac{78q}{N} - \frac{32N}{q^2}\right),\end{aligned}$$

which completes the proof. □

4.2.2 Case 2: $q = \mathcal{O}(N^{1/2})$

The entire proof is almost the same as in the case $q = \omega(N^{1/2})$, except for a slight modification to the calculations related to M and M_0. As mentioned before, for any positive $\epsilon > 0$, if we set $M = q/N^{1/2-\epsilon}$ and $M_0 = q/N^{1/4+\epsilon}$, then we can get a $\mathcal{O}(N^{3/4-\epsilon})$-bound.

For simplicity, we here set $M = \frac{q}{N^{9/20}}$ and $M_0 = \frac{q}{N^{3/10}}$, i.e. $\epsilon = \frac{1}{20}$. We omit the details of proof and only list the following two technical lemmas.

Lemma 9 (Bad Transcripts, $q = \mathcal{O}(N^{1/2})$). *For any given $\mathcal{Q}_0, \mathcal{Q}_1$ such that $|\mathcal{Q}_0| = |\mathcal{Q}_1| = q$, we have*

$$\Pr_{K \leftarrow_\$ \{0,1\}^{4n}}[\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \text{ is bad}] \leq \frac{12q}{N^{7/10}} + \frac{3q^2}{N^{7/5}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5}.$$

Lemma 10 (Good Transcripts, $q = \mathcal{O}(N^{1/2})$). *Fix arbitrarily a good transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$ as defined in Definition 4. Let $\mathcal{Q}_{0,6}^{\equiv}$ and $\mathcal{Q}_0^*$ be as described in Eqn. (15), then we have*

$$\Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,6}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1] \geq \frac{1}{(N)_q} \times \left(1 - \frac{122q^2}{N^{7/5}} - \frac{135q}{N^{3/4}} - \frac{32}{N^{1/10}}\right).$$

According to the H-coefficient technique (see Lemma 2), we can obtain

$$\begin{aligned}\mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_1P_1}(\mathcal{D}) &\leq \varepsilon_1 + \varepsilon_2 \\ &= \frac{12q}{N^{7/10}} + \frac{3q^2}{N^{7/5}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{122q^2}{N^{7/5}} + \frac{135q}{N^{3/4}} + \frac{32}{N^{1/10}} \\ &= \frac{12q}{N^{7/10}} + \frac{125q^2}{N^{7/5}} + \frac{135q}{N^{3/4}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{32}{N^{1/10}},\end{aligned}$$

which is the result of small $q = \mathcal{O}(N^{1/2})$ in Theorem 6.

5 Tight Security Bound of t-Round KACSP

In this section, we generalize the proof of 3-round KACSP to the general t-round KACSP. The proof idea is basically the same, except the notation is heavier.

Theorem 11 (t-Round KACSP). *Consider the t-round KACSP (where $t \geq 4$), denoted as $P_1^{(t)}$-construction. For any distinguisher $\mathcal{D}$ making q_e construction queries and q_p permutation queries, the following upper bound holds:*

$$\mathbf{Adv}^{\mathsf{SPRP}}_{P_1^{(t)}}(\mathcal{D}) \leq \begin{cases} \frac{27t^4q}{N^{t/(t+1)}} + \frac{15t^5q^2}{N^{2t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t} + \frac{4t^2N}{q^2}, & \text{for } q = \omega(N^{1/2}) \\ \frac{4tq}{N^{7/10}} + \frac{15t^5q^2}{N^{7/5}} + \frac{q^{t-1}}{N^{7(t-1)/10}} + \frac{22t^4q}{N^{3/4}} + \frac{tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t} + \frac{4t^2}{N^{1/10}}, & \text{for } q = \mathcal{O}(N^{1/2}) \end{cases}$$

where $q := \max\{q_e, q_p\}$.

Note that the value of $t = \mathcal{O}(1)$ is a constant. Therefore, the above bound does show that unless $\mathcal{D}$ makes $q = \Omega(N^{t/(t+1)})$ queries, its advantage of distinguishing $P_1^{(t)}$ from a truly random permutation is negligible (for sufficiently large n). In other words, t-round KACSP has the same security level as the t-round KAC.

Proof of Theorem 11. As discussed in Section 4.2, we also consider that the distinguisher makes q construction queries and q permutation queries in the analysis. That is, for each attainable transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$, it has $|\mathcal{Q}_0| = |\mathcal{Q}_1| = q$. Furthermore, we let $\mathsf{AD}_{t-i} \subset \mathcal{Q}_0$ denote the set of pairs $(x_i, y_i) \in \mathcal{Q}_0$ whose actual distance is i, where $i = 1, \ldots, t$. We also let $\delta_i := |\mathsf{AD}_i|$. For convenience, we simply use $\mathcal{Q}_{0,t}^{\equiv}$ to denote AD_0 since it is a uniform-structure-group.

Applying General Transformation. First of all, we also use the general transformation (see Eqn. (11)) here to reduce the task to bounding only one probability. The basic idea is to replace the uniform-structure-groups whose actual distance is less than t with a new uniform-structure-group whose actual distance is t, and make the connecting probability smaller.

Note that the expectation of δ_i is $\mathcal{O}(q/N^{i/(t+1)})$ and we can *wlog* assume that $q = \mathcal{O}(N^{t/(t+1)})$ (otherwise the security bound is invalid). Then, we denote $s = \sum_{i=1}^{t-1} \delta_i = \mathcal{O}(q/N^{1/(t+1)})$ as the number of pairs to be replaced. As discussed in Section 3, it is easy to construct a new uniform-structure-group $\mathcal{Q}_0^* := \{(x_i, y_i) : i = q+1, \ldots, q+s\}$ and set $\widehat{\mathcal{Q}_0} := \mathcal{Q}_{0,t}^{\equiv} \cup \mathcal{Q}_0^*$, where the pairs in $\mathcal{Q}_0^*$ must be connected in the most wasteful way. Using Lemma 4 several times, we can know that the crucial probability

$$
\begin{aligned}
(7) \geq\ & \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \widehat{\mathcal{Q}_0} \mid P_1 \downarrow \mathcal{Q}_1] \\
=\ & \Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,t}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1].
\end{aligned} \tag{36}
$$

Thus, Eqn. (36) becomes the target probability for which we need a lower bound.

5.1 Case 1: $q = \omega(N^{1/2})$

As in Section 4.2, we mainly focus on the large values of $q = \omega(N^{1/2})$, and the other case of $q = \mathcal{O}(N^{1/2})$ is similar. We also first give the definition of good and bad transcripts.

Let $\mathsf{R}_{t-1} = \{\mathsf{R}(x_i) : (x_i, y_i) \in \mathsf{AD}_{t-1}\}$ and $\mathsf{L}_{t-1} = \{\mathsf{L}(y_i) : (x_i, y_i) \in \mathsf{AD}_{t-1}\}$ denote the set of all rightmost and leftmost vertices of the pairs whose actual distance is 1, respectively. Next, we define $t-1$ constants $M_j = \frac{q}{N^{j/(t+1)}}$ related to the value of q, where $j = 1, 2, \ldots, t-1$.

Definition 5 (Bad and Good Transcripts, $P_1^{(t)}$-Construction). *For an attainable transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$, we say that τ is bad if $K \in \bigcup_{i=1}^{5} \mathsf{BadK}_i$; otherwise τ is good. The definitions of BadK_i are shown below:*

$K \in \mathsf{BadK}_1 \Leftrightarrow$ *there exists a path of length $2t+1$ starting from a vertex $x_i \in \mathsf{Dom}(\mathcal{Q}_0)$ in shore 0 or ending at a vertex $y_i \in \mathsf{Ran}(\mathcal{Q}_0)$ in shore $2t+1$*

$K \in \mathsf{BadK}_2 \Leftrightarrow \delta_i > M_i$ *where $i = 1, 2, \ldots, t-1$*

$K \in \mathsf{BadK}_3 \Leftrightarrow |\mathsf{R}_{t-1} \cup \mathsf{Dom}(\mathcal{Q}_1)| < \delta_{t-1} + q \vee |\mathsf{L}_{t-1} \cup \mathsf{Ran}(\mathcal{Q}_1)| < \delta_{t-1} + q$

$$
K \in \mathsf{BadK}_4 \Leftrightarrow \begin{cases}
\bigvee_{i=1}^{t-1} |\{x \in \mathsf{Dom}(\mathcal{Q}_0) : x \oplus \kappa_0 \oplus \kappa_i \text{ is not left-free}\}| > M_1 \\
\bigvee_{i=1}^{t-1} |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_i) \cap (\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_t)| > M_1 \\
\bigvee_{i=1}^{t-1} |\{y \in \mathsf{Ran}(\mathcal{Q}_0) : y \oplus \kappa_3 \oplus \kappa_i \text{ is not right-free}\}| > M_1 \\
\bigvee_{i=1}^{t-1} |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_i) \cap (\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0)| > M_1
\end{cases}
$$

$K \in \mathsf{BadK}_5 \Leftrightarrow |U_{05}| > M_1 \vee |U_{27}| > M_1.$

We can determine the value of $\varepsilon_2 = \frac{5tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t}$ from the following lemma, the proof of which can be found in Appendix E.3.

Lemma 12 (Bad Transcripts, $q = \omega(N^{1/2})$). *For any given $\mathcal{Q}_0, \mathcal{Q}_1$ such that $|\mathcal{Q}_0| = |\mathcal{Q}_1| = q$, we have*

$$\Pr_{K \leftarrow_\$ \{0,1\}^{(t+1)n}}[\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \text{ is bad}] \leq \frac{5tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t}.$$

The following lemma gives a lower bound on Eqn. (36) for any good transcript.

Lemma 13 (Good Transcripts, $q = \omega(N^{1/2})$). *Fix arbitrarily a good transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$ as defined in Definition 5. Let $\mathcal{Q}_{0,t}^{\equiv}$ and $\mathcal{Q}_0^*$ be as described in Eqn. (36), then we have*

$$\Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,t}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1] \geq \frac{1}{(N)_q} \times \left(1 - \frac{22t^4q}{N^{t/(t+1)}} - \frac{15t^5q^2}{N^{2t/(t+1)}} - \frac{4t^2N}{q^2}\right). \quad (37)$$

The proof of Lemma 13 is deferred to Appendix E.4. We next show how to obtain the final security bound from the above two lemmas. First note that (37) is also a lower bound on the crucial probability (7), i.e. $p(\tau)$ in Lemma 3 when $t' = 1$. Then it is not difficult to determine the value of ε_1. According to the H-coefficient technique (see Lemma 2), we can obtain

$$\begin{aligned} \mathbf{Adv}^{\mathsf{SPRP}}_{(P_1)^t}(\mathcal{D}) &\leq \varepsilon_1 + \varepsilon_2 \\ &= \frac{5tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t} + \frac{22t^4q}{N^{t/(t+1)}} + \frac{15t^5q^2}{N^{2t/(t+1)}} + \frac{4t^2N}{q^2} \\ &\leq \frac{27t^4q}{N^{t/(t+1)}} + \frac{15t^5q^2}{N^{2t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t} + \frac{4t^2N}{q^2}, \end{aligned} \quad (38)$$

which is the result of large $q = \omega(N^{1/2})$ in Theorem 11. □

5.2 Case 2: $q = \mathcal{O}(N^{1/2})$

The entire proof is almost the same as in the case $q = \omega(N^{1/2})$, except for a slight modification to the calculations related to M_1 and M_{t-1} and we here set $M_1 = \frac{q}{N^{3/10}}$ and $M_{t-1} = \frac{q}{N^{9/20}}$. We omit the details of proof and only list the following two technical lemmas.

Lemma 14 (Bad Transcripts, $q = \mathcal{O}(N^{1/2})$). *For any given $\mathcal{Q}_0, \mathcal{Q}_1$ such that $|\mathcal{Q}_0| = |\mathcal{Q}_1| = q$, we have*

$$\Pr_{K \leftarrow_\$ \{0,1\}^{(t+1)n}}[\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \text{ is bad}] \leq \frac{4tq}{N^{7/10}} + \frac{q^{t-1}}{N^{7(t-1)/10}} + \frac{tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t}.$$

Lemma 15 (Good Transcripts, $q = \mathcal{O}(N^{1/2})$). *Fix arbitrarily a good transcript $\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \in \mathcal{T}$ as defined in Definition 5. Let $\mathcal{Q}_{0,t}^{\equiv}$ and $\mathcal{Q}_0^*$ be as described in Eqn. (36), then we have*

$$\Pr_{P_1 \leftarrow_\$ \mathcal{P}_n}[E_K \downarrow \mathcal{Q}_{0,t}^{\equiv} \wedge E_K \downarrow_w \mathcal{Q}_0^* \mid P_1 \downarrow \mathcal{Q}_1] \geq \frac{1}{(N)_q} \times \left(1 - \frac{15t^5q^2}{N^{7/5}} - \frac{22t^4q}{N^{3/4}} - \frac{4t^2}{N^{1/10}}\right).$$

According to the H-coefficient technique (see Lemma 2), we can obtain

$$\begin{aligned} \mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_1P_1}(\mathcal{D}) &\leq \varepsilon_1 + \varepsilon_2 \\ &= \frac{4tq}{N^{7/10}} + \frac{q^{t-1}}{N^{7(t-1)/10}} + \frac{tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t} + \frac{15t^5q^2}{N^{7/5}} + \frac{22t^4q}{N^{3/4}} + \frac{4t^2}{N^{1/10}} \\ &= \frac{4tq}{N^{7/10}} + \frac{15t^5q^2}{N^{7/5}} + \frac{q^{t-1}}{N^{7(t-1)/10}} + \frac{22t^4q}{N^{3/4}} + \frac{tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t} + \frac{4t^2}{N^{1/10}}, \end{aligned}$$

which is the result of small $q = \mathcal{O}(N^{1/2})$ in Theorem 6.

6 Remarks On Other Variants of KACs

Our proof technology in this work applies to various KAC-type constructions as well as the standard KAC construction. Our general transformation also works and the proof idea is similar. The core task is to find enough constructive methods of reducing new edges, so that the final security bound is sufficiently accurate.

We also find that the more rounds means more methods, so it seems easier to find enough methods in constructions with more rounds. This is somewhat counter-intuitive. It might be interesting to figure out whether this phenomenon is an artifact of the proof technology, or because larger constructions inherently have more security redundancy.

Acknowledgements. We would like to thank the anonymous reviewers of TCC 2023 for their valuable comments. Yu Yu is supported by the National Natural Science Foundation of China (Grant Nos. 62125204 and 92270201), the National Key Research and Development Program of China (Grant No. 2018YFA0704701), and the Major Program of Guangdong Basic and Applied Research (Grant No. 2019B030302008). Yu Yu also acknowledges the support from the XPLORER PRIZE. This work is supported in part by the National Key Research and Development Program of China (Grant No. 2022YFB2701400) and in part by the National Natural Science Foundation of China (Grant No. 62132005, 62172162).

References

[Bog+12] 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 *Advances in Cryptology - EUROCRYPT 2012*, pages 45–62. 2012. https://doi.org/10.1007/978-3-642-29011-4_5. (Cited on p. 1)

[Che+18] Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. "Minimizing the Two-Round Even-Mansour Cipher". *J. Cryptology*, (4):1064–1119, 2018. https://doi.org/10.1007/s00145-018-9295-y. (Cited on pp. 1, 2, 6, 7, 8, 10, 11, 12)

[CS14] Shan Chen and John P. Steinberger. "Tight Security Bounds for Key-Alternating Ciphers". In *Advances in Cryptology - EUROCRYPT 2014*, pages 327–350. Springer, 2014. https://doi.org/10.1007/978-3-642-55220-5_19. (Cited on pp. 1, 4, 5, 6)

[DKS12] Orr Dunkelman, Nathan Keller, and Adi Shamir. "Minimalism in Cryptography: The Even-Mansour Scheme Revisited". In *Advances in Cryptology - EUROCRYPT 2012*, pages 336–354. 2012. https://doi.org/10.1007/978-3-642-29011-4_21. (Cited on p. 1)

[DR02] Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard*. https://doi.org/10.1007/978-3-662-04722-4. Springer, 2002. ISBN: 3-540-42580-2 (cited on p. 1).

[EM97] Shimon Even and Yishay Mansour. "A Construction of a Cipher from a Single Pseudorandom Permutation". *J. Cryptology*, (3):151–162, 1997. https://doi.org/10.1007/s001459900025. (Cited on p. 1)

[HT16] Viet Tung Hoang and Stefano Tessaro. "Key-Alternating Ciphers and Key-Length Extension: Exact Bounds and Multi-user Security". In *Advances in Cryptology - CRYPTO 2016, Part I*, pages 3–32. 2016. https://doi.org/10.1007/978-3-662-53018-4_1. (Cited on pp. 1, 12)

[LPS12] Rodolphe Lampe, Jacques Patarin, and Yannick Seurin. "An Asymptotically Tight Security Analysis of the Iterated Even-Mansour Cipher". In *Advances in Cryptology - ASIACRYPT 2012*, pages 278–295. 2012. https://doi.org/10.1007/978-3-642-34961-4_18. (Cited on p. 1)

[Ste12] John P. Steinberger. "Improved Security Bounds for Key-Alternating Ciphers via Hellinger Distance". *IACR Cryptology ePrint Archive*, ()481, 2012. http://eprint.iacr.org/2012/481. (Cited on p. 1)

[TZ21] Stefano Tessaro and Xihu Zhang. "Tight Security for Key-Alternating Ciphers with Correlated Sub-keys". In *Advances in Cryptology - ASIACRYPT 2021, Part III*, pages 435–464. Springer, 2021. https://doi.org/10.1007/978-3-030-92078-4_15. (Cited on p. 1)

[WYCD20] Yusai Wu, Liqing Yu, Zhenfu Cao, and Xiaolei Dong. "Tight Security Analysis of 3-Round Key-Alternating Cipher with a Single Permutation". In *Advances in Cryptology - ASIACRYPT 2020, Part I*, pages 662–693. Springer, 2020. https://doi.org/10.1007/978-3-030-64837-4_22. (Cited on pp. 1, 2, 6, 7, 8, 10, 11, 12, 13, 15, 25, 26)

A Basic Tail Inequality

Definition 6. *The hypergeometric distribution and multivariate hypergeometric distribution we used are shown below,*

$$\mathsf{Hyp}_{A,a,H}(h) \stackrel{\text{def}}{=} \frac{\binom{H}{h}\binom{A-H}{a-h}}{\binom{A}{a}} = \frac{(a)_h(H)_h(A-H)_{a-h}}{h!(A)_a}, \tag{39}$$

$$\mathsf{MHyp}_{A,a,H_1,\ldots,H_k}(h_1,\ldots,h_k) \stackrel{\text{def}}{=} \frac{\prod_{i=1}^k \binom{H_i}{h_i} \cdot \binom{A-\sum_{j=1}^k H_j}{a-\sum_{j=1}^k h_j}}{\binom{A}{a}}, \text{ where } k \geq 2. \tag{40}$$

Lemma 16 (Chebyshev's Inequality, Lemma 7 of [WYCD20])**.** *Let $X \sim \mathsf{Hyp}_{A,a,H}$ be a hypergeometric distribution random variable, then for any $\lambda > 0$ we have*

$$\Pr[X > \lambda] \leq \frac{aH(A-H)(A-H)}{(\lambda A - aH)^2(A-1)}.$$

B Examples of Core

For simplicity, we let all permutations be defined on the set $\{0,1,2,3,4,5\}$ and all round keys in the following constructions are set to 0 (since Core only relates to permutation-edges).

1. Consider the $P_1P_1P_1$-construction and assume that initially no P_1-permutation-edges are fixed (i.e., $\mathcal{Q}_1 = \emptyset$). If we connect $x_1 = 1$ to $y_1 = 2$ as follows,

$$x_1 = 1 \xrightarrow{P_1} 2, 2 \xrightarrow{P_1} 1, 1 \xrightarrow{P_1} 2 = y_1,$$

then $\mathsf{Core}(x_1, y_1) = \{\langle 1, P_1, 2\rangle, \langle 2, P_1, 1\rangle\}$ and $|\mathsf{Core}(x_1, y_1)| = 2$. Note that the P_1-permutation-edge $\langle 1, P_1, 2\rangle$ is used twice in the path.

2. Consider the $P_1P_1P_2$-construction and assume that initially no P_1-permutation-edges and P_2-permutation-edges are fixed (i.e., $\mathcal{Q}_1 = \mathcal{Q}_2 = \emptyset$). If we connect $x_2 = 1$ to $y_2 = 2$ as follows,

$$x_2 = 1 \xrightarrow{P_1} 2, 2 \xrightarrow{P_1} 1, 1 \xrightarrow{P_2} 2 = y_2,$$

then $\mathsf{Core}(x_2, y_2) = \left\{ {}^{\langle 1,P_1,2\rangle,}_{\langle 2,P_1,1\rangle,} \langle 1, P_2, 2\rangle \right\}$ and $|\mathsf{Core}(x_2, y_2)| = 2 + 1 = 3$.

3. Consider the $P_1P_1P_1$-construction and assume that initially no P_1-permutation-edges are fixed (i.e., $\mathcal{Q}_1 = \emptyset$). If we connect $x_3 = 1$ to $y_3 = 4$ and $x_4 = 3$ to $y_4 = 1$ as follows,

$$x_3 = 1 \xrightarrow{P_1} 2, 2 \xrightarrow{P_1} 3, 3 \xrightarrow{P_1} 4 = y_3,$$

$$x_4 = 3 \xrightarrow{P_1} 4, 4 \xrightarrow{P_1} 5, 5 \xrightarrow{P_1} 1 = y_4,$$

then $\mathsf{Core}(\{(x_3,y_3),(x_4,y_4)\}) = \{\langle 1,P_1,2\rangle, \langle 2,P_1,3\rangle, \langle 3,P_1,4\rangle, \langle 4,P_1,5\rangle, \langle 5,P_1,1\rangle\}$ and $|\mathsf{Core}(\{(x_3,y_3),(x_4,y_4)\})| = 5$. Note that the permutation-edge $\langle 3,P_1,4\rangle$ is used in both paths, that is, it is *shared* by the two paths.

4. Consider the $P_1P_1P_2$-construction and assume that initially no P_1-permutation-edges and P_2-permutation-edges are fixed (i.e., $\mathcal{Q}_1 = \mathcal{Q}_2 = \emptyset$). If we connect $x_5 = 1$ to $y_5 = 4$ and $x_6 = 3$ to $y_6 = 1$ as follows,

$$x_5 = 1 \xrightarrow{P_1} 2, 2 \xrightarrow{P_1} 3, 3 \xrightarrow{P_2} 4 = y_5,$$
$$x_6 = 3 \xrightarrow{P_1} 4, 4 \xrightarrow{P_1} 5, 5 \xrightarrow{P_2} 1 = y_6,$$

then $\mathsf{Core}(\{(x_5,y_5),(x_6,y_6)\}) = \left\{\begin{smallmatrix}\langle 1,P_1,2\rangle, & \\ \langle 2,P_1,3\rangle, & \langle 3,P_2,4\rangle, \\ \langle 3,P_1,4\rangle, & \langle 5,P_2,1\rangle \\ \langle 4,P_1,5\rangle, & \end{smallmatrix}\right\}$ and $|\mathsf{Core}(\{(x_5,y_5),(x_6,y_6)\})| = 4+2 = 6$.

C Typical Methods of Building Paths

C.1 Recycle-Edge-Based Methods

Taking the $P_1P_1P_1$-construction as an example, we illustrate what a recycled-edge is in Fig. 1 and also show how our method chooses recycled-edges to build paths. Given a pair (x_1, y_1) with actual distance 3, we can connect x_1 to y_1 with only 2 new edges if the second permutation-edge is recycled from the known edges in $\mathcal{Q}_1$. It therefore reduces one new edge compared to the most wasteful way, which always consumes 3 new edges. Note that this is the only case of recycled-edge, as it is not possible to construct recycled-edges for the first and third round permutations.

We now describe how our recycled-edge-based method connects a pair, say (x_1, y_1), whose actual distance is 3. A simple observation is that once the recycled-edge is chosen, then entire path from x_1 to y_1 is determined. Thus, our approach is to select a suitable vertex u from $\mathsf{Dom}(\mathcal{Q}_1)$, and then determine all other vertices accordingly. Similar to [WYCD20], we here use $\mathsf{RC}(3)$ to denote the *Range (set) of all possible Candidate values* for u, where 3 represents the shore where u is located. That is, the core task of recycled-edge-based method is to determine the set $\mathsf{RC}(3)$. As shown in Fig. 1, we can know that $\mathsf{RC}(3) = \mathsf{Dom}(\mathcal{Q}_1) \setminus S_1 \setminus S_2$.

Note that unlike the shared-edge-based method used in [WYCD20], the path constructed by our recycled-edge-based method does not share any new edges with other paths. This property makes the recycled-edge-based method very easy to analyze and can be directly generalized to connect multiple pairs. For example, to connect h pairs of (x_i, y_i)'s (each with an actual distance of 3) using the recycled-edge-based method, we can simply select h different u_i's from $\mathsf{RC}(3)$ and assign one u_i to each pair as in the one-path case. Then, the connection of all h pairs only needs $2h$ new edges, which reduces h edges compared to the most wasteful way (which needs $3h$ new edges).

C.2 Shared-Edge-Based Methods

The shared-edge-based methods have been studied in detail in [WYCD20]. We here only illustrate the $(1,2)$-shared-edge in $P_1P_1P_1$-construction as an example, and the general (i,j)-shared-edges can be analyzed in a similar way.

As shown in Fig. 1, each $(1,2)$-shared-edge involves two pairs, say (x_2, y_2) and (x'_2, y'_2), where the vertex of path (x'_2, y'_2) on shore 3 is exactly $x_2 \oplus \kappa_0$. Then, the edge from $x_2 \oplus \kappa_0$ to $w_{2,2}$ will be used twice in the two paths, i.e., it is a $(1,2)$-shared-edge. As a result, one less new edge will be used to connect the two paths than in the most wasteful way.

Using the $(1,2)$-shared-edge-based method to connect pairs (x_2, y_2) and (x'_2, y'_2) needs to select two suitable vertices $w_{2,2}$ and $w_{2,4}$. Similar to the recycled-edge-based method, we let $\mathsf{RC}(2)$ and

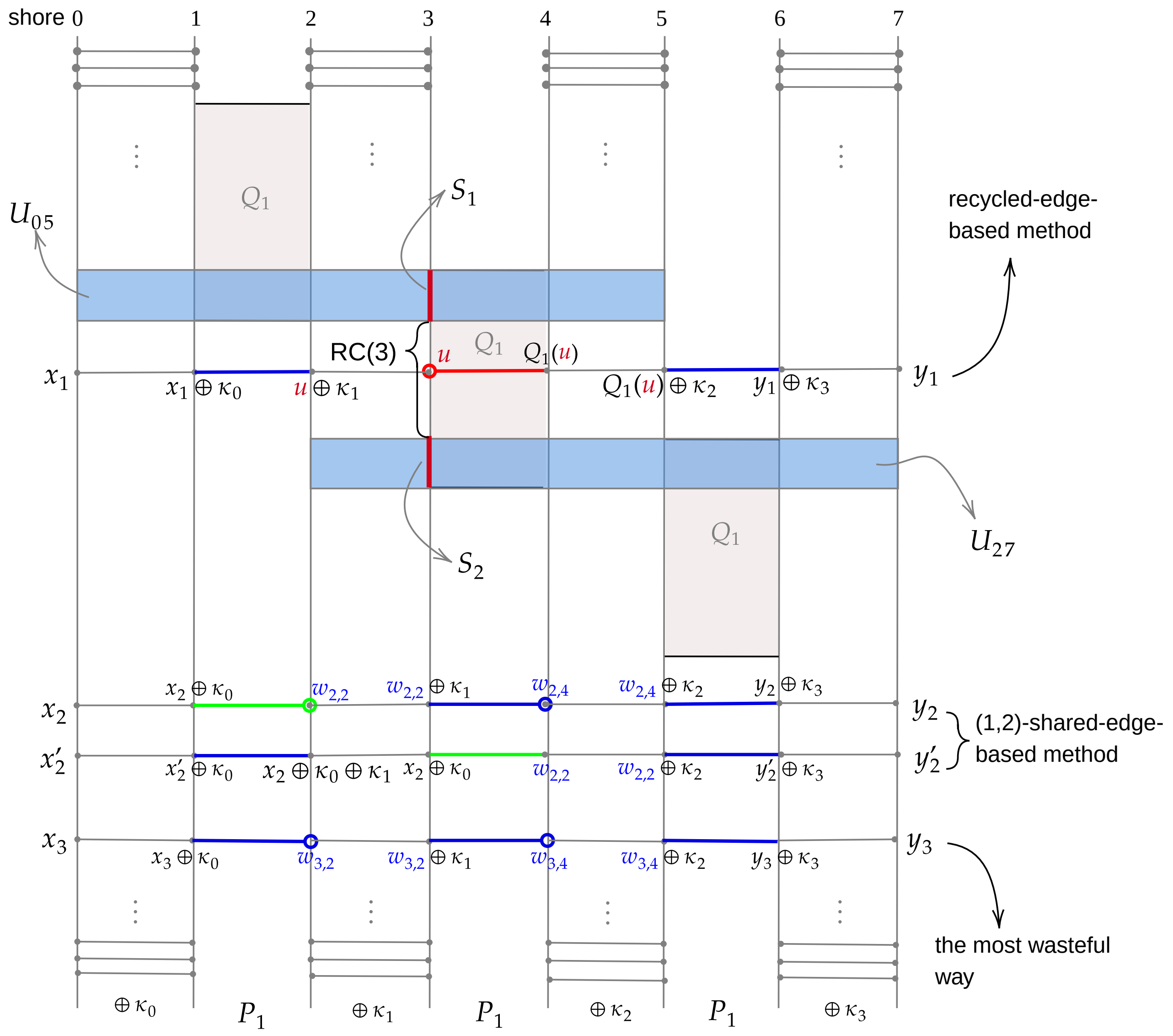

Figure 1: Illustration of the recycled-edge-based method, the $(1, 2)$-shared-edge-based method and the most wasteful way in $P_1P_1P_1$-construction. We use the gray rectangle to represent the set $\mathcal{Q}_1$, and the two light blue rectangles to represent the sets U_{05} and U_{27} respectively. The intersection of U_{05} and $\mathcal{Q}_1$ (resp. U_{27} and $\mathcal{Q}_1$) at shore 3 is denoted by $S1$ (resp. $S2$). In particular, we use black edges for the key-edges, red edge for the recycled-edge, green edges for the $(1, 2)$-shared-edge, and blue edges for the newly added permutation-edges that will never be reused. The pair of (x_1, y_1) is connected using the recycled-edge-based method, whose task is essentially to choose the vertex u (marked by a circle). The pairs of (x_2, y_2) and (x'_2, y'_2) are connected using the $(1, 2)$-shared-edge-based method, whose task is essentially to choose the vertices $w_{2,2}$ and $w_{2,4}$ (marked by circles). The pair of (x_3, y_3) is connected in the most wasteful way, whose task is essentially to choose the vertices $w_{3,2}$ and $w_{3,4}$ (marked by circles).

$\mathsf{RC}(4)$ denote the *Range (set) of all possible Candidate values* for $w_{2,2}$ and $w_{2,4}$, respectively. Then, the core task of the $(1, 2)$-shared-edge-based method is to determine the sets $\mathsf{RC}(2)$ and $\mathsf{RC}(4)$.

By analyzing the $w_{2,2}$ in Figure 1, we can know that $\mathsf{RC}(2) \supseteq \{0, 1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus V \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus U \oplus \kappa_1 \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus U \oplus \kappa_2$, where U (resp. V) denotes the domain (resp. range) of all permutation queries fixed so far. In this definition, the U, V conditions ensure that the shared-edge from $x_2 \oplus \kappa_0$ to $w_{2,2}$ can be defined without incompatibility, and the $\mathsf{Dom}(\mathcal{Q}_0), \mathsf{Ran}(\mathcal{Q}_0)$ conditions ensure that it does not cause additional shared-edges. By a similar analysis on the $w_{2,4}$ in Figure 1, we can know $\mathsf{RC}(4) \supseteq \{0,1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus V \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus U \oplus \kappa_2$, where U (resp. V) denotes the domain (resp. range) of all permutation queries fixed so far.

What needs to be pointed out here is that, although we use the same symbols U, V in $\mathsf{RC}(2)$ and $\mathsf{RC}(4)$ they are not necessarily the same. The reason is that by definition they are not actually static, but are dynamically updated as new permutation-edges are defined during the process of connecting paths. This makes dertermining the cardinality of each RC often a very careful task.

C.3 The Most Wasteful Way

As shown in Fig. 1, using the most wasteful way to connect (x_3, y_3) needs to select two suitable vertices $w_{3,2}$ and $w_{3,4}$. Similar to other methods, we let $\mathsf{RC}(2)$ and $\mathsf{RC}(4)$ denote the *Range (set) of all possible Candidate values* for $w_{3,2}$ and $w_{3,4}$, respectively. Then, the core task of the most wasteful way is to determine the sets $\mathsf{RC}(2)$ and $\mathsf{RC}(4)$.

By definition, the path constructed by the most wasteful way does not share any edges with other paths. Like the recycled-edge-based method, the most wasteful way is easy to analyze and can be directly generalized to connect multiple pairs. Here, we can know that $\mathsf{RC}(2) \supseteq \{0,1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus V \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \setminus U \oplus \kappa_1$, where U (resp. V) denote the domain (resp. range) of all permutation queries fixed so far. In this definition, the U, V conditions ensure that the defined permutation edge from $x_3 \oplus \kappa_0$ to $w_{3,2}$ is new, while the $\mathsf{Dom}(\mathcal{Q}_0), \mathsf{Ran}(\mathcal{Q}_0)$ conditions ensure that it does not cause shared-edges. Similarly, we can know that $\mathsf{RC}(4) \supseteq \{0,1\}^n \setminus \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \setminus V \setminus \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2 \setminus U \oplus \kappa_2$, where U (resp. V) denotes the domain (resp. range) of all permutation queries fixed so far.

Note that although the definition of $\mathsf{RC}(4)$ looks the same as in the $(1,2)$-shared-edge-based method, in practice the sets U, V would be very different.

D Illustrative Analysis of Constraints on Z

This section shows an analysis of the properties that the elements in $Z \subset \mathcal{Q}_{0,6}^{\equiv}$ should satisfy (see Eqn. (18)). In particular, we require each pair chosen from Z can be constructed as the upper-paths and lower-paths in $(1,2)$-shared-edge-based method, $(1,3)$-shared-edge-based method and $(2,3)$-shared-edge-based method. Thus, the constraints on Z are the sum of constraints from these three methods.

D.1 Constraints from $(1,2)$-shared-edge-based method

The two paths completed using the $(1,2)$-shared-edge-based method consist of 5 permutation-edges and have the following form.

$$\widetilde{x}_1 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{w}_{11}, \qquad \widetilde{w}_{11} \oplus \kappa_1 \xrightarrow{P_1} \widetilde{w}_{12}, \qquad \widetilde{w}_{12} \oplus \kappa_2 \xrightarrow{P_1} \widetilde{y}_1 \oplus \kappa_3,$$
$$\widetilde{x}'_1 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{x}_1 \oplus \kappa_0 \oplus \kappa_1, \qquad \widetilde{x}_1 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{w}_{11}, \qquad \widetilde{w}_{11} \oplus \kappa_2 \xrightarrow{P_1} \widetilde{y}'_1 \oplus \kappa_3,$$

It can be seen that the value $\widetilde{x}_1 \oplus \kappa_0 \oplus \kappa_1$ should be assigned to the image of $\widetilde{x}'_1 \oplus \kappa_0$. That is, $\widetilde{x}_1 \oplus \kappa_0 \oplus \kappa_1$ should be left-free and $\widetilde{x}_1 \oplus \kappa_0 \oplus \kappa_1 \notin \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3$. Then, we obtain two constraints on Z, i.e., $x_i \oplus \kappa_0 \oplus \kappa_1$ should be left-free and $x_i \notin \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1 \oplus \kappa_3$ in Eqn. (18).

D.2 Constraints from $(1,3)$-shared-edge-based method

The two paths completed using the $(1,3)$-shared-edge-based method consist of 5 permutation-edges and have the following form.

$$\widetilde{x}_2 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{y}'_2 \oplus \kappa_3, \qquad \widetilde{y}'_2 \oplus \kappa_3 \oplus \kappa_1 \xrightarrow{P_1} \widetilde{w}_{21}, \qquad \widetilde{w}_{21} \oplus \kappa_2 \xrightarrow{P_1} \widetilde{y}_2 \oplus \kappa_3,$$
$$\widetilde{x}'_2 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{w}_{22}, \qquad \widetilde{w}_{22} \oplus \kappa_1 \xrightarrow{P_1} \widetilde{x}_2 \oplus \kappa_0 \oplus \kappa_2, \qquad \widetilde{x}_2 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{y}'_2 \oplus \kappa_3,$$

It can be seen that the value $\widetilde{y}'_2 \oplus \kappa_3$ should be assigned to the image of $\widetilde{x}_2 \oplus \kappa_0$. That is, $\widetilde{x}_2 \oplus \kappa_0 \oplus \kappa_2$ should be left-free, $\widetilde{x}_2 \oplus \kappa_0 \oplus \kappa_2 \notin \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3$, $\widetilde{y}'_2 \oplus \kappa_3 \oplus \kappa_1$ should be right-free, $\widetilde{y}'_2 \oplus \kappa_3 \oplus \kappa_1 \notin$

$\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0$. Then, we obtain four constraints on Z, i.e., $x_i \oplus \kappa_0 \oplus \kappa_2$ should be left-free, $y_i \oplus \kappa_3 \oplus \kappa_1$ should be right-free, $x_i \notin \mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_0 \oplus \kappa_2$ and $y_i \notin \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_3 \oplus \kappa_1$ in Eqn. (18).

D.3 Constraints from $(2,3)$-shared-edge-based method

The two paths completed using the $(2,3)$-shared-edge-based method consist of 5 permutation-edges and have the following form.

$$\widetilde{x}_3 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{w}_{31} \oplus \kappa_2 \oplus \kappa_1, \quad \widetilde{w}_{31} \oplus \kappa_2 \xrightarrow{P_1} \widetilde{y}'_3 \oplus \kappa_3, \quad \widetilde{y}'_3 \oplus \kappa_3 \oplus \kappa_2 \xrightarrow{P_1} \widetilde{y}_3 \oplus \kappa_3,$$
$$\widetilde{x}'_3 \oplus \kappa_0 \xrightarrow{P_1} \widetilde{w}_{32}, \quad \widetilde{w}_{32} \oplus \kappa_1 \xrightarrow{P_1} \widetilde{w}_{31}, \quad \widetilde{w}_{31} \oplus \kappa_2 \xrightarrow{P_1} \widetilde{y}'_3 \oplus \kappa_3,$$

It can be seen that the value $\widetilde{y}'_3 \oplus \kappa_3 \oplus \kappa_2$ should be assigned to the image of $\widetilde{y}_3 \oplus \kappa_3$. That is, $\widetilde{y}'_3 \oplus \kappa_3 \oplus \kappa_2$ should be right-free, $\widetilde{y}'_3 \oplus \kappa_3 \oplus \kappa_2 \notin \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0$. Then, we obtain two constraints on Z, i.e., $y_i \oplus \kappa_3 \oplus \kappa_2$ should be right-free and $y_i \notin \mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_3 \oplus \kappa_2$ in Eqn. (18).

E Omitted Proofs

E.1 Proof of Lemma 4

Proof. We let M_A (resp. M_B) denote the number of permutations $P_1, \ldots, P_{t'}$ that satisfy $E_K \downarrow_w A^{\equiv}$ (resp. $E_K \downarrow_w B^{\equiv}$) conditioned on $(P_1 \downarrow \mathcal{Q}_1) \wedge \cdots \wedge (P_{t'} \downarrow \mathcal{Q}_{t'})$. Clearly, it is equivalent to prove that

$$M_B \leq M_A,$$

since the probability space is the same in both target probabilities.

Since $A^{\equiv}$ is completed in the most wasteful way, each $\mathsf{Core}_w(A^{\equiv})$ contains the same number of newly added edges to each round permutation (see the useful property and Defn. 3). We let m_i denote the number of edges newly added to round permutation P_i (relative to $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}$), where $i = 1, \ldots, t'$. Thus, whenever $A^{\equiv}$ is completed in the most wasteful way, the number of edges that can be freely defined in each round permutation is fixed, i.e. $N - |\mathcal{Q}_i| - m_i$, where $i = 1, \ldots, t'$. If let $\#\mathsf{Core}_w(A^{\equiv})$ denote the number of different Cores for the event $E_K \downarrow_w A^{\equiv}$ to occur, then we have

$$M_A = \#\mathsf{Core}_w(A^{\equiv}) \times \prod_{i=1}^{t'} (N - |\mathcal{Q}_i| - m_i)! \tag{41}$$

Similarly, $B^{\equiv}$ is also completed in the most wasteful way, so each $\mathsf{Core}_w(B^{\equiv})$ contains the same number of newly added edges to each round permutation. The only difference from $A^{\equiv}$ is that the first round permutation (i.e. P_1) here will add s more new-edges. In this case, for each possible $\mathsf{Core}_w(B^{\equiv})$ generated in the most wasteful way, the numbers of edges that can be freely defined in each round permutation are $N - |\mathcal{Q}_1| - m_1 - s, N - |\mathcal{Q}_2| - m_2, \ldots, N - |\mathcal{Q}_{t'}| - m_{t'}$, respectively. If we use $\#\mathsf{Core}_w(B^{\equiv})$ to denote the number of different Cores for the event $E_K \downarrow_w B^{\equiv}$ to occur, then we have

$$\begin{aligned} M_B &= \#\mathsf{Core}_w(B^{\equiv}) \times (N - |\mathcal{Q}_1| - m_1 - s)! \times \prod_{i=2}^{t'} (N - |\mathcal{Q}_i| - m_i)! \\ &= \#\mathsf{Core}_w(B^{\equiv})_{0,\ldots,3} \times \#\mathsf{Core}_w(B^{\equiv})_{3,\ldots,2t+1|0,\ldots,3} \times \frac{\prod_{i=1}^{t'} (N - |\mathcal{Q}_i| - m_i)!}{(N - |\mathcal{Q}_1| - m_1)_s}, \end{aligned} \tag{42}$$

where $\#\mathsf{Core}_w(B^{\equiv})_{0,\ldots,3}$ represents the number of possible paths connecting the shore 0 to 3 of $B^{\equiv}$ in the most wasteful way, and $\#\mathsf{Core}_w(B^{\equiv})_{3,\ldots,2t+1|0,\ldots,3}$ represents the number of possible paths connecting the shore 3 to $2t+1$ of $B^{\equiv}$ in the most wasteful way under the condition that the shore 0 to 3 are connected in the most wasteful way (i.e., the s nodes in shore 2 of $B^{\equiv}$ are fixed).

Before evaluating the upper bounds of $\#\mathsf{Core}_w(B^{\equiv})_{0,\ldots,3}$ and $\#\mathsf{Core}_w(B^{\equiv})_{3,\ldots,2t+1|0,\ldots,3}$, let us justify the decomposition of $\#\mathsf{Core}_w(B^{\equiv})$ into these two terms. To construct a $\mathsf{Core}_w(B^{\equiv})$, it is equivalent to assigning values to the $s \cdot t$ nodes on the shores $2, 4, \ldots, 2t$ of the s paths. Note that in each $\mathsf{Core}_w(B^{\equiv})$, all permutation-edges are distinct and new compared to $\mathcal{Q}_1, \ldots, \mathcal{Q}_{t'}$. That means every assignment of a node adds exactly one new permutation-edge. This regular resource change with assignment is useful for counting. Another useful property is that s nodes in the same shore have the same requirements for the values to be assigned, so they can be processed together in a uniform manner. In particular, when we perform arbitrary permutation on the s values in any shore i of a $\mathsf{Core}_w(B^{\equiv})$, the result is still a $\mathsf{Core}_w(B^{\equiv})$. Therefore, we can evaluate the value $\#\mathsf{Core}_w(B^{\equiv})$ by calculating the number of possible assignments shore by shore.

First, it is not hard to verify that

$$\#\mathsf{Core}_w(B^{\equiv})_{3,\ldots,2t+1|0,\ldots,3} \leq \#\mathsf{Core}_w(A^{\equiv}), \tag{43}$$

since each path connecting the shore 3 to $2t+1$ of $B^{\equiv}$ in the most wasteful way can be slightly modified to be a path that completes $A^{\equiv}$ in the most wasteful way. In fact, connecting the shore 3 to $2t+1$ of $B^{\equiv}$ in the most wasteful way is essentially the same as the task of completing $A^{\equiv}$ in the most wasteful way. That is, both are to connect a uniform-structure-group with actual distance $t-1$ in the most wasteful way. The former has even fewer resources available, since s more P_1-permutation-edges are fixed in it.

Next, we give an upper bound on the value of $\#\mathsf{Core}_w(B^{\equiv})_{0,\ldots,3}$. Since the s edges of the 1-st round permutation are required to be new and not cause any edge reuse, the most basic constraint is that the s vertices next to $x'_1, \ldots, x'_s$ cannot be from the set $\mathsf{Ran}(\mathcal{Q}_1) \cup \mathsf{Dom}(\mathcal{Q}_{i_2}) \oplus \kappa_1$ (there are, of course, other constraints). Therefore,

$$\begin{aligned}
\#\mathsf{Core}_w(B^{\equiv})_{0,\ldots,3} &\leq (N - |\mathsf{Ran}(\mathcal{Q}_1) \cup \mathsf{Dom}(\mathcal{Q}_{i_2}) \oplus \kappa_1|)_s \\
&\leq (N - \underbrace{|\mathsf{Ran}(\mathcal{Q}_1)|}_{=|\mathcal{Q}_1|} - \underbrace{|\mathsf{Dom}(\mathcal{Q}_{i_2}) \oplus \kappa_1|}_{=|\mathcal{Q}_{i_2}|} + \underbrace{|\mathsf{Ran}(\mathcal{Q}_1) \oplus \kappa_1 \cap \mathsf{Dom}(\mathcal{Q}_{i_2})|}_{\leq |U_{04}| \leq |\mathcal{Q}_{i_2}|/2})_s \\
&\leq (N - |\mathcal{Q}_1| - |\mathcal{Q}_{i_2}|/2)_s.
\end{aligned} \tag{44}$$

Combining the Eqns. (41)–(44), we have

$$\begin{aligned}
M_B &\leq \#\mathsf{Core}_w(A^{\equiv}) \times \prod_{i=1}^{t'} (N - |\mathcal{Q}_i| - m_i)! \times \frac{\#\mathsf{Core}_w(B^{\equiv})_{0,\ldots,3}}{(N - |\mathcal{Q}_1| - m_1)_s} \\
&\leq M_A \times \frac{(N - |\mathcal{Q}_1| - |\mathcal{Q}_{i_2}|/2)_s}{(N - |\mathcal{Q}_1| - m_1)_s} \\
&\leq M_A,
\end{aligned}$$

where the last inequality uses the fact that $m_1 \leq s \cdot t \leq |\mathcal{Q}_{i_2}|/2$. □

E.2 Proof of Lemma 7

Proof. For the given sets $\mathcal{Q}_0, \mathcal{Q}_1$, we calculate the probability $\Pr[K \in \mathsf{BadK}_i]$ for $i = 1, \ldots, 4$, and then obtain the final result by union bound.

(1) We can rewrite the definition of BadK_1 as follows,

$$\begin{aligned}
K \in \mathsf{BadK}_1 \;\Leftrightarrow\quad & \exists (x,y) \in \mathcal{Q}_0,\ \exists (u_1,v_1),(u_2,v_2),(u_3,v_3), \in \mathcal{Q}_1 \\
& \text{such that } \kappa_0 = x \oplus u_1 \vee \kappa_1 = v_1 \oplus u_2 \vee \kappa_2 = v_2 \oplus u_3 \\
\text{or } & \exists (x,y) \in \mathcal{Q}_0,\ \exists (u_1,v_1),(u_2,v_2),(u_3,v_3) \in \mathcal{Q}_1 \\
& \text{such that } \kappa_3 = y \oplus v_3 \vee \kappa_2 = v_2 \oplus u_3 \vee \kappa_1 = v_1 \oplus u_2,
\end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_1] \leq \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N^3} + \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N^3} \leq \frac{2q^4}{N^3}.$$

(2) We first decompose the event $K \in \mathsf{BadK}_2$ into 5 sub-events, namely

$$K \in \mathsf{BadK}_{2,1} \Leftrightarrow \alpha_2 > M, \quad K \in \mathsf{BadK}_{2,2} \Leftrightarrow \beta_2 > M, \quad K \in \mathsf{BadK}_{2,3} \Leftrightarrow \gamma_2 > M,$$
$$K \in \mathsf{BadK}_{2,4} \Leftrightarrow \alpha_1 > M_0, \quad K \in \mathsf{BadK}_{2,5} \Leftrightarrow \beta_1 > M_0.$$

(a) For the probability $\Pr[K \in \mathsf{BadK}_{2,1}]$, we can rewrite the event $\mathsf{BadK}_{2,1}$ as

$$\begin{aligned} K \in \mathsf{BadK}_{2,1} &\Leftrightarrow \alpha_2 > M \\ &\Leftrightarrow \big|\{(x_1, y_1), (u_1, v_1), (u_2, v_2) \in \mathcal{Q}_0 \times \mathcal{Q}_1 \times \mathcal{Q}_1 : \\ &\qquad \kappa_0 = x \oplus u_1 \wedge \kappa_1 = v_1 \oplus u_2\}\big| > M, \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_{2,1}] \le \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N^2 \cdot M} = \frac{q^2}{N^{3/2}}.$$

(b) Similarly, we can know that

$$\Pr[K \in \mathsf{BadK}_{2,2}] \le \frac{q^2}{N^{3/2}} \text{ and } \Pr[K \in \mathsf{BadK}_{2,3}] \le \frac{q^2}{N^{3/2}}.$$

(c) For the probability $\Pr[K \in \mathsf{BadK}_{2,3}]$, we can rewrite the event $\mathsf{BadK}_{2,3}$ as

$$\begin{aligned} K \in \mathsf{BadK}_{2,4} &\Leftrightarrow \alpha_1 > M_0 \\ &\Leftrightarrow |\{(x_1, y_1), (u_1, v_1) \in \mathcal{Q}_0 \times \mathcal{Q}_1 : \kappa_0 = x \oplus u_1\}| > M_0, \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_{2,4}] \le \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1|}{N \cdot M_0} = \frac{q}{N^{3/4}}.$$

(d) Similarly, we can know that

$$\Pr[K \in \mathsf{BadK}_{2,5}] \le \frac{q}{N^{3/4}}.$$

Taken together, we finally obtain

$$\Pr[K \in \mathsf{BadK}_2] \le \sum_{j=1}^{5} \Pr[K \in \mathsf{BadK}_{2,j}] \le \frac{3q^2}{N^{3/2}} + \frac{2q}{N^{3/4}}.$$

(3) The event $K \in \mathsf{BadK}_3$ can be divided into the following 12 sub-events, namely

$$\begin{aligned}
K \in \mathsf{BadK}_{3,1} &\Leftrightarrow \mathsf{Dom}(\mathcal{Q}_1) \text{ is not disjoint from } \mathsf{R}(\mathcal{Q}_{0,1}^{\equiv}), \\
K \in \mathsf{BadK}_{3,2} &\Leftrightarrow \mathsf{Dom}(\mathcal{Q}_1) \text{ is not disjoint from } \mathsf{R}(\mathcal{Q}_{0,2}^{\equiv}), \\
K \in \mathsf{BadK}_{3,3} &\Leftrightarrow \mathsf{Dom}(\mathcal{Q}_1) \text{ is not disjoint from } \mathsf{R}(\mathcal{Q}_{0,3}^{\equiv}), \\
K \in \mathsf{BadK}_{3,4} &\Leftrightarrow \mathsf{R}(\mathcal{Q}_{0,1}^{\equiv}) \text{ is not disjoint from } \mathsf{R}(\mathcal{Q}_{0,2}^{\equiv}), \\
K \in \mathsf{BadK}_{3,5} &\Leftrightarrow \mathsf{R}(\mathcal{Q}_{0,1}^{\equiv}) \text{ is not disjoint from } \mathsf{R}(\mathcal{Q}_{0,3}^{\equiv}), \\
K \in \mathsf{BadK}_{3,6} &\Leftrightarrow \mathsf{R}(\mathcal{Q}_{0,2}^{\equiv}) \text{ is not disjoint from } \mathsf{R}(\mathcal{Q}_{0,3}^{\equiv}), \\
K \in \mathsf{BadK}_{3,7} &\Leftrightarrow \mathsf{Ran}(\mathcal{Q}_1) \text{ is not disjoint from } \mathsf{L}(\mathcal{Q}_{0,1}^{\equiv}), \\
K \in \mathsf{BadK}_{3,8} &\Leftrightarrow \mathsf{Ran}(\mathcal{Q}_1) \text{ is not disjoint from } \mathsf{L}(\mathcal{Q}_{0,2}^{\equiv}), \\
K \in \mathsf{BadK}_{3,9} &\Leftrightarrow \mathsf{Ran}(\mathcal{Q}_1) \text{ is not disjoint from } \mathsf{L}(\mathcal{Q}_{0,3}^{\equiv}), \\
K \in \mathsf{BadK}_{3,10} &\Leftrightarrow \mathsf{L}(\mathcal{Q}_{0,1}^{\equiv}) \text{ is not disjoint from } \mathsf{L}(\mathcal{Q}_{0,2}^{\equiv}), \\
K \in \mathsf{BadK}_{3,11} &\Leftrightarrow \mathsf{L}(\mathcal{Q}_{0,1}^{\equiv}) \text{ is not disjoint from } \mathsf{L}(\mathcal{Q}_{0,3}^{\equiv}),
\end{aligned}$$

$$K \in \mathsf{BadK}_{3,12} \;\Leftrightarrow\; \mathsf{L}(\mathcal{Q}_{\overline{0,2}}^{\equiv}) \text{ is not disjoint from } \mathsf{L}(\mathcal{Q}_{\overline{0,3}}^{\equiv}).$$

We only show here how to upper-bound the probability $\Pr[K \in \mathsf{BadK}_{3,1}]$, the other cases are very similar. For this, we first rewrite the event $K \in \mathsf{BadK}_{3,1}$ as

$$\begin{aligned} K \in \mathsf{BadK}_{3,1} \;\Leftrightarrow\; & \exists (x,y) \in \mathcal{Q}_0,\ \exists (u_1,v_1),(u_2,v_2),(u_1',v_1') \in \mathcal{Q}_1, \\ & \text{such that } \kappa_0 = x \oplus u_1 \vee \kappa_1 = v_1 \oplus u_2 \vee \kappa_2 = v_2 \oplus u_1', \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_{3,1}] \leq \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N^3} \leq \frac{q^4}{N^3}.$$

Similarly, we can also upper-bound the other probabilities as follows,

$$\begin{aligned} \Pr[K \in \mathsf{BadK}_{3,j}] &\leq \frac{q^4}{N^3}, \text{ for } j = 1,2,3,7,8,9 \\ \text{and } \Pr[K \in \mathsf{BadK}_{3,\ell}] &\leq \frac{q^6}{N^5}, \text{ for } \ell = 4,5,6,10,11,12 \end{aligned}$$

Taken together, we finally obtain

$$\Pr[K \in \mathsf{BadK}_3] \leq \sum_{j=1}^{12} \Pr[K \in \mathsf{BadK}_{3,j}] \leq \frac{6q^4}{N^3} + \frac{6q^6}{N^5}.$$

(4) Let S be the set of n-bit stings that are right-free, and T be the set of n-bit stings that are left-free, then $|S| = |T| = N - q$. We can rewrite the definition of BadK_4 as follows,

$$\begin{aligned} K \in \mathsf{BadK}_{4,1} &\Leftrightarrow |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1) \backslash T| > M_0 \Leftrightarrow \left|(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1) \cap \overline{T}\right| > M_0, \\ K \in \mathsf{BadK}_{4,2} &\Leftrightarrow |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2) \backslash T| > M_0 \Leftrightarrow \left|(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2) \cap \overline{T}\right| > M_0, \\ K \in \mathsf{BadK}_{4,3} &\Leftrightarrow |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_1) \backslash S| > M_0 \Leftrightarrow \left|(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_1) \cap \overline{S}\right| > M_0, \\ K \in \mathsf{BadK}_{4,4} &\Leftrightarrow |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_2) \backslash S| > M_0 \Leftrightarrow \left|(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_2) \cap \overline{S}\right| > M_0, \\ K \in \mathsf{BadK}_{4,5} &\Leftrightarrow |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1) \cap (\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3)| > M_0, \\ K \in \mathsf{BadK}_{4,6} &\Leftrightarrow |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_2) \cap (\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3)| > M_0, \\ K \in \mathsf{BadK}_{4,7} &\Leftrightarrow |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_1) \cap (\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0)| > M_0, \\ K \in \mathsf{BadK}_{4,8} &\Leftrightarrow |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3 \oplus \kappa_2) \cap (\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0)| > M_0. \end{aligned}$$

Then it has

$$\begin{aligned} \Pr[K \in \mathsf{BadK}_{4,1}] &\leq \frac{|\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1| \cdot \left|\overline{T}\right|}{M_0} \times \frac{1}{N} \leq \frac{q \cdot q}{\frac{q}{N^{1/4}}} \times \frac{1}{N} = \frac{q}{N^{3/4}}, \\ \Pr[K \in \mathsf{BadK}_{4,5}] &\leq \frac{|\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_1| \cdot |\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_3|}{M_0} \times \frac{1}{N} \leq \frac{q \cdot q}{\frac{q}{N^{1/4}}} \times \frac{1}{N} = \frac{q}{N^{3/4}}. \end{aligned}$$

Similarly, we can also upper-bound the other probabilities as follows,

$$\begin{aligned} \Pr[K \in \mathsf{BadK}_{4,j}] &\leq \frac{q}{N^{3/4}}, j = 2,3,4, \\ \Pr[K \in \mathsf{BadK}_{4,\ell}] &\leq \frac{q}{N^{3/4}}, \ell = 6,7,8. \end{aligned}$$

Taken together, we finally obtain

$$\Pr[K \in \mathsf{BadK}_4] \leq \sum_{i=1}^{8} \Pr[K \in \mathsf{BadK}_{4,i}] \leq \frac{8q}{N^{3/4}}.$$

(5) We can rewrite the definition of BadK_5 as follows,

$$K \in \mathsf{BadK}_5 \;\Leftrightarrow\; \begin{aligned} &|\{(u_1,v_1),(u_2,v_2) \in \mathcal{Q}_1 \times \mathcal{Q}_1 : \kappa_1 = v_1 \oplus u_2\}| > M_0 \\ \text{or } &|\{(u_2,v_2),(u_3,v_3) \in \mathcal{Q}_1 \times \mathcal{Q}_1 : \kappa_2 = v_2 \oplus u_3\}| > M_0, \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_5] \le \frac{|\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N \cdot M_0} + \frac{|\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N \cdot M_0} = \frac{2q}{N^{3/4}}.$$

Combining all the above, we obtain the final result

$$\begin{aligned} &\Pr_{K \leftarrow_\$ \{0,1\}^{4n}}[\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \text{ is bad}] \\ &= \Pr[K \in \bigcup_{1 \le i \le 5} \mathsf{BadK}_i] \\ &\le \sum_{i=1}^{5} \Pr[K \in \mathsf{BadK}_i] \\ &\le \frac{2q^4}{N^3} + \frac{3q^2}{N^{3/2}} + \frac{2q}{N^{3/4}} + \frac{6q^4}{N^3} + \frac{6q^6}{N^5} + \frac{8q}{N^{3/4}} + \frac{2q}{N^{3/4}} \\ &= \frac{12q}{N^{3/4}} + \frac{3q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5}. \end{aligned}$$

□

E.3 Proof of Lemma 12

Proof. For the given sets $\mathcal{Q}_0, \mathcal{Q}_1$, we calculate the probability $\Pr[K \in \mathsf{BadK}_i]$ for $i = 1, \ldots, 5$, and then obtain the final result by union bound.

(1) We can rewrite the definition of BadK_1 as follows,

$$K \in \mathsf{BadK}_1 \;\Leftrightarrow\; \begin{aligned} &\exists (x,y) \in \mathcal{Q}_0,\ \exists (u_1,v_1),(u_2,v_2),\ldots,(u_t,v_t) \in \mathcal{Q}_1 \\ &\text{such that } \kappa_0 = x \oplus u_1 \vee \kappa_i = v_i \oplus u_{i+1}, \text{ where } i = 1,2,\ldots,t-1 \\ \text{or } &\exists (x,y) \in \mathcal{Q}_0,\ \exists (u_1,v_1),(u_2,v_2),\ldots,(u_t,v_t) \in \mathcal{Q}_1 \\ &\text{such that } \kappa_t = y \oplus v_t \vee \kappa_i = v_i \oplus u_{i+1}, \text{ where } i = 1,2,\ldots,t-1, \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_1] \le \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1|^t}{N^t} + \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1|^t}{N^t} \le \frac{2q^{t+1}}{N^t}.$$

(2) We first decompose the event $K \in \mathsf{BadK}_2$ into $t-1$ sub-events, namely

$$K \in \mathsf{BadK}_{2,i} \;\Leftrightarrow\; \delta_i > M_i, \text{ where } i = 1,2,\ldots,t-1.$$

For the probability $\Pr[K \in \mathsf{BadK}_{2,i}]$, we can rewrite the event $\mathsf{BadK}_{2,i}$ as

$$\begin{aligned} K \in \mathsf{BadK}_{2,i} &\Leftrightarrow \delta_i > M_i \\ &\Leftrightarrow \big|\{(x,y),(u_1,v_1),\ldots,(u_i,v_i) \in \mathcal{Q}_0 \times \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_1 : \\ &\qquad \kappa_0 = x \oplus u_1 \wedge \kappa_j = v_j \oplus u_{j+1}, j = 1,2,\ldots,i-1.\}\big| > M_i, \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_{2,i}] \le \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1|^i}{N^i \cdot M_i} = \left(\frac{q}{N^{t/(t+1)}}\right)^i.$$

Taken together, we finally obtain

$$\Pr[K \in \mathsf{BadK}_2] \le \sum_{i=1}^{t-1} \Pr[K \in \mathsf{BadK}_{2,i}] \le \sum_{i=1}^{t-1} \left(\frac{q}{N^{t/(t+1)}}\right)^i \le \frac{(t-1)q}{N^{t/(t+1)}}.$$

(3) The event $K \in \mathsf{BadK}_3$ can be divided into $2 \cdot \binom{t+1}{2}$ sub-events, We only show here to upper-bound the probability of one case $\mathsf{BadK}_{3,1}$, the other cases are very similar. For this, we first rewrite the event $K \in \mathsf{BadK}_{3,1}$ as

$$\begin{aligned} K \in \mathsf{BadK}_{3,1} \Leftrightarrow\ & \exists(x,y) \in \mathcal{Q}_0,\ \exists(u_1,v_1),\ldots,(u_{t-1},v_{t-1}),(u_1',v_1') \in \mathcal{Q}_1, \\ & \text{such that } \kappa_0 = x \oplus u_1 \vee \kappa_i = v_i \oplus u_{i+1} \vee \kappa_{t-1} = v_{t-1} \oplus u_1', \\ & \text{where } i = 1,2,\ldots,t-2, \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_{3,1}] \le \frac{|\mathcal{Q}_0| \cdot |\mathcal{Q}_1|^t}{N^t} \le \frac{q^{t+1}}{N^t}.$$

Then, we finally obtain

$$\Pr[K \in \mathsf{BadK}_3] \le 2 \cdot \binom{t+1}{2} \times \Pr[K \in \mathsf{BadK}_{3,1}] \le \frac{(t^2+t)q^{t+1}}{N^t}.$$

(4) Let S be the set of n-bit stings that are right-free, and T be the set of n-bit stings that are left-free, then $|S| = |T| = N - q$. We can rewrite the definition of BadK_4 as follows,

$$\begin{aligned} K \in \mathsf{BadK}_{4,1} &\Leftrightarrow \bigvee_{i=1}^{t-1} |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_i) \backslash T| > M_1 \\ &\Leftrightarrow \bigvee_{i=1}^{t-1} \left|(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_i) \cap \overline{T}\right| > M_1, \\ K \in \mathsf{BadK}_{4,2} &\Leftrightarrow \bigvee_{i=1}^{t-1} |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_t \oplus \kappa_i) \backslash S| > M_1 \\ &\Leftrightarrow \bigvee_{i=1}^{t-1} \left|(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_t \oplus \kappa_i) \cap \overline{S}\right| > M_1, \\ K \in \mathsf{BadK}_{4,3} &\Leftrightarrow \bigvee_{i=1}^{t-1} |(\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_i) \cap (\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_t)| > M_1, \\ K \in \mathsf{BadK}_{4,4} &\Leftrightarrow \bigvee_{i=1}^{t-1} |(\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_t \oplus \kappa_i) \cap (\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0)| > M_1. \end{aligned}$$

Then it has

$$\begin{aligned} \Pr[K \in \mathsf{BadK}_{4,1}] &\le \bigcup_{i=1}^{t-1} \frac{|\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_i| \cdot |\overline{T}|}{M_1} \times \frac{1}{N} \le (t-1) \times \frac{q \cdot q}{\frac{q}{N^{1/(t+1)}}} \times \frac{1}{N} = \frac{(t-1)q}{N^{t/(t+1)}}, \\ \Pr[K \in \mathsf{BadK}_{4,3}] &\le \bigcup_{i=1}^{t-1} \frac{|\mathsf{Dom}(\mathcal{Q}_0) \oplus \kappa_0 \oplus \kappa_i| \cdot |\mathsf{Ran}(\mathcal{Q}_0) \oplus \kappa_t|}{M_1} \times \frac{1}{N} \\ &\le (t-1) \times \frac{q \cdot q}{\frac{q}{N^{1/(t+1)}}} \times \frac{1}{N} = \frac{(t-1)q}{N^{t/(t+1)}}. \end{aligned}$$

Similarly, we can also upper-bound the other probabilities as follows,

$$\Pr[K \in \mathsf{BadK}_{4,2}] \le \frac{(t-1)q}{N^{t/(t+1)}},\ \Pr[K \in \mathsf{BadK}_{4,4}] \le \frac{(t-1)q}{N^{t/(t+1)}}.$$

Then, we finally obtain

$$\Pr[K \in \mathsf{BadK}_4] = \sum_{i=1}^{4} \Pr[K \in \mathsf{BadK}_{4,i}] \le \frac{4(t-1)q}{N^{t/(t+1)}}.$$

(5) We can rewrite the definition of BadK_4 as follows,

$$\begin{aligned} K \in \mathsf{BadK}_5 \ \Leftrightarrow \quad & |\{(u_1,v_1),(u_2,v_2) \in \mathcal{Q}_1 \times \mathcal{Q}_1 : \kappa_1 = v_1 \oplus u_2\}| > M_1 \\ \text{or} \ & |\{(u_2,v_2),(u_3,v_3) \in \mathcal{Q}_1 \times \mathcal{Q}_1 : \kappa_2 = v_2 \oplus u_3\}| > M_1. \end{aligned}$$

then it has

$$\Pr[K \in \mathsf{BadK}_5] \le \frac{|\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N \cdot M_1} + \frac{|\mathcal{Q}_1| \cdot |\mathcal{Q}_1|}{N \cdot M_1} = \frac{2q}{N^{t/t+1}}.$$

Combining all the above, we obtain the final result

$$\begin{aligned} & \Pr_{K \leftarrow_\$ \{0,1\}^{(t+1)n}}[\tau = (\mathcal{Q}_0, \mathcal{Q}_1, K) \text{ is bad}] \\ &= \Pr[K \in \bigcup_{1 \le i \le 5} \mathsf{BadK}_i] \\ &\le \sum_{i=1}^{5} \Pr[K \in \mathsf{BadK}_i] \\ &\le \frac{2q^{t+1}}{N^t} + \frac{(t-1)q}{N^{t/(t+1)}} + \frac{(t^2+t)q^{t+1}}{N^t} + \frac{4(t-1)q}{N^{t/(t+1)}} + \frac{2q}{N^{t/t+1}} \\ &= \frac{(5t-3)q}{N^{t/(t+1)}} + \frac{(t^2+t+2)q^{t+1}}{N^t} \\ &\le \frac{5tq}{N^{t/(t+1)}} + \frac{2t^2q^{t+1}}{N^t}. \end{aligned}$$

□

E.4 Proof of Lemma 13

Proof. Let $\mathcal{Q}_0^{\equiv} = \widehat{\mathcal{Q}_0} := \mathcal{Q}_{0,t}^{\equiv} \cup \mathcal{Q}_0^*, t = t$ and $t' = 1$, then the target probability is exactly an instantiation of the general problem (9). We also apply the general framework (10) to bound it, so roughly in three steps.

The first step is to generate Core**s with specific numbers of new edges.** We will use $k = \frac{t^2-t}{2} + 1$ variables (denoted as $h_1, \ldots, h_k$) to obtain a sufficiently accurate security bound, so k constructive methods of reducing new edges are needed. Similar to the proof of Theorem 6, the first method is the recycled-edge-based method, and the other $k-1$ methods are shared-edge-based methods, each of which exploits a different type of shared-edges.

Recalling the Eqn. (36) , our task is to connect the q pairs of $\widehat{\mathcal{Q}_0} = \mathcal{Q}_{0,t}^{\equiv} \cup \mathcal{Q}_0^*$ using a specified number of new edges, where $\mathcal{Q}_0^*$ is connected in the most wasteful way. Let $h_1, \ldots, h_k$ be k integer variables in the interval $[0, M_{t-1}]$, where $M_{t-1} = \frac{q}{N^{(t-1)/(t+1)}}$. We combine the recycled-edge-based method, the shared-edge-based methods and the most wasteful way to accomplish the task in three steps.

1. Select h_1 different plaintext-ciphertext pairs from $\mathcal{Q}_{0,t}^{\equiv}$, and then use the recycled-edge-based method to grow paths from these h_1 chosen plaintexts to shore 4;[15] then proceed to grow these h_1 paths to their corresponding ciphertexts in the most wasteful way.
2. Apart from the h_1 pairs selected in Step 1, select $2h_2, 2h_3, \cdots, 2h_k$ appropriate distinct pairs in turn from $\mathcal{Q}_{0,t}^{\equiv}$, and connect the particular $2h_\ell$ pairs using the $op(\ell)$-shared-edge-based method for $\ell = 2, \ldots, k$, where $op(\cdot) : \{2, \ldots, k\} \mapsto \{1, \ldots, t\} \times \{1, \ldots, t\}$ is a mapping defined as follows. First sort all the $k-1$ tuples in set $\{(i,j) : 1 \le i \le t-1, i < j \le t\}$ alphabetically, and number them in order from 2 to k. Then we define $op(\ell)$ as the tuple (i,j) with index ℓ.
3. Connect the remaining $\delta_0 - h_1 - \sum_{i=2}^{k} 2h_i$ pairs in $\mathcal{Q}_{0,t}^{\equiv}$ and the s pairs in $\mathcal{Q}_0^*$ in the most wasteful way.

Clearly, the above procedure must generate a $\mathsf{Core}(\widehat{\mathcal{Q}_0})$ containing $t \cdot q - \sum_{i=1}^{k} h_i$ new edges, and all the pairs of $\mathcal{Q}_0^*$ are connected in the most wasteful way.

The second step is to evaluate the number of Cores **that can be generated in the first step.** By analysis similar to in the proof of Theorem 6, we can determine the lower bound on the value of $\#\mathsf{Cores}_{t \cdot q - \sum_{i=1}^{k} h_i}$. As the process is routine and nothing particularly interesting or new, we omit the details here.

$$
\begin{aligned}
&\#\mathsf{Cores}_{t \cdot q - \sum_{i=1}^{k} h_i} \\
\ge\ & \frac{(\delta_0)_{h_1} \cdot (q - 2M_1)_{h_1}}{h_1!} \cdot \frac{(q_0)_{\sum_{i=2}^{k} 2h_i}}{\prod_{i=2}^{k} h_i!} \cdot \prod_{j=0}^{t-4} \Big(N - 4q - (2j+1)h_1 \Big)_{h_1} \\
& \times \prod_{i=2}^{k} \prod_{j=0}^{2t-5} \left(N - 6q - (3t-6)h_1 - \sum_{l=2}^{i-1} 3(2t-4)h_l - 3jh_i \right)_{h_i} \\
& \times \left(N - 4q - (2t-4)h_1 - \sum_{i=2}^{k} (4t-8)h_i \right)_{q - h_1 - \sum_{i=2}^{k} 2h_i} \\
& \times \prod_{\ell=0}^{t-3} \left(N - (5+2\ell)q - (2t-5-2\ell)h_1 - \sum_{j=2}^{k} (4t-10-4\ell)h_j \right)_{q - h_1 - \sum_{i=2}^{k} 2h_i} .
\end{aligned} \tag{45}
$$

The third step is to perform the summation calculation. Since the lower bound on $\#\mathsf{Cores}_{3q - \sum_{i=1}^{k} h_i}$ is known, we are now ready to calculate the final result. From the Eqns. (10) and (45), we have

$$
\begin{aligned}
(36) \ge & \sum_{0 \le h_1, \ldots, h_k \le M_{t-1}} \frac{\#\mathsf{Cores}_{tq - \sum_{i=1}^{k} h_i}}{(N-q)_{tq - \sum_{i=1}^{k} h_i}} \\
= & \frac{1}{(N)_q} \times \sum_{0 \le h_1, \ldots, h_k \le M_{t-1}} \frac{(\delta_0)_{h_1} \cdot (q - 2M_1)_{h_1}}{h_1!} \cdot \frac{(q_0)_{\sum_{i=2}^{k} 2h_i}}{\prod_{i=2}^{k} h_i!} \cdot \prod_{j=0}^{t-4} \Big(N - 4q - (2j+1)h_1 \Big)_{h_1} \\
& \times \prod_{i=2}^{k} \prod_{j=0}^{2t-5} \left(N - 6q - (3t-6)h_1 - \sum_{l=2}^{i-1} 3(2t-4)h_l - 3jh_i \right)_{h_i} \\
& \times \left(N - 4q - (2t-4)h_1 - \sum_{i=2}^{k} (4t-8)h_i \right)_{q - h_1 - \sum_{i=2}^{k} 2h_i} \\
& \times \frac{(N)_q \prod_{\ell=0}^{t-3} \Big(N - (5+2\ell)q - (2t-5-2\ell)h_1 - \sum_{j=2}^{k} (4t-10-4\ell)h_j \Big)_{q - h_1 - \sum_{i=2}^{k} 2h_i}}{\Big(N - q \Big)_{tq - \sum_{i=1}^{k} h_i}}
\end{aligned}
$$

[15] That is, we use a recycled-edge between shore 3 and 4 for each path, which is the same as in the proof of Theorem 6.

$$\times \underbrace{\frac{\prod_{i=1}^{k} h_i! \cdot (N)_q}{(q)_{\sum_{i=1}^{k} h_i} \cdot \prod_{i=1}^{k} (q)_{h_i} \cdot (N-kq)_{q-\sum_{i=1}^{k} h_i}} \cdot \mathsf{MHyp}_{N,q,q,\ldots,q}(h_1,\ldots,h_k)}_{=1 \text{(see Defn. 6)}}$$

$$= \frac{1}{(N)_q} \times \sum_{0 \le h_1,\ldots,h_k \le M_{t-1}} \mathsf{MHyp}_{N,q,q,\ldots,q}(h_1,\ldots,h_k) \tag{46}$$

$$\times \frac{(\delta_0)_{h_1} \cdot (q-2M_1)_{h_1} \cdot (q_0)_{\sum_{i=2}^{k} 2h_i}}{h_1! \cdot \prod_{i=2}^{k} h_i!} \times \frac{\prod_{i=1}^{k} h_i!}{(q)_{\sum_{i=1}^{k} h_i} \prod_{i=1}^{k} (q)_{h_i}} \tag{47}$$

$$\left.\begin{aligned} &\times \frac{(N)_q \prod_{\ell=0}^{t-3} \left(N-(5+2\ell)q-(2t-5-2\ell)h_1-\sum_{j=2}^{k}(4t-10-4\ell)h_j\right)_q (N)_q}{\left(N-q\right)_{tq} \left(N-kq\right)_q} \\ &\cdot \left(N-4q-(2t-4)h_1-\sum_{i=2}^{k}(4t-8)h_i\right)_q \end{aligned}\right\} \tag{48}$$

$$\left.\begin{aligned} &\times \frac{\prod_{j=0}^{t-4} \left(N-4q-(2j+1)h_1\right)_{h_1}}{\left(N-5q-(2t-3)h_1-\sum_{i=2}^{k}(4t-6)h_i\right)_{h_1+\sum_{i=2}^{k} 2h_i}} \\ &\cdot \frac{\prod_{i=2}^{k} \prod_{j=0}^{2t-5} \left(N-6q-(3t-6)h_1-\sum_{l=2}^{i-1} 3(2t-4)h_l-3jh_i\right)_{h_i}}{\prod_{\ell=0}^{t-3} \left(N-(6+2\ell)q-(2t-4-2\ell)h_1-\sum_{j=2}^{k}(4t-8-4\ell)h_j\right)_{h_1+\sum_{i=2}^{k} 2h_i}} \\ &\cdot \left(N-(t+1)q+\sum_{i=1}^{k} h_i\right)_{\sum_{i=1}^{k} h_i} \left(N-(k+1)q+\sum_{i=1}^{k} h_i\right)_{\sum_{i=1}^{k} h_i}. \end{aligned}\right\} \tag{49}$$

We next calculate the lower bound for each of the Eqns. (46)–(49) as follows.

$$\begin{aligned} (46) \ge& \frac{1}{(N)_q} \times \left(1-\sum_{i=1}^{k} \sum_{h_i > M_{t-1}} \mathsf{MHyp}_{N,q,q,\ldots,q}(h_1,\ldots,h_k)\right) \\ =& \frac{1}{(N)_q} \times \left(1-\sum_{i=1}^{k} \sum_{h_i > M_{t-1}} \mathsf{Hyp}_{N,q,q}(h_i)\right) \\ \ge& \frac{1}{(N)_q} \times \left(1-\frac{kq^2(N-q)^2}{[M_{t-1}N-q^2]^2(N-1)}\right) \\ \ge& \frac{1}{(N)_q} \times \left(1-\frac{kN^2}{[q-N^{2/(t+1)}]^2(N-1)}\right) \\ \ge& \frac{1}{(N)_q} \times \left(1-\frac{8kN}{q^2}\right), \end{aligned}$$

where we use Lemma 16 for the second inequality, and the third inequality holds by substituting $M_{t-1} = \frac{q}{N^{(t-1)/(t+1)}}$, and the last inequality uses the fact that $q = \omega(N^{1/2})$, $N^{2/(t+1)} \le \frac{q}{2}$, $N-1 \ge \frac{N}{2}$ for sufficiently large n.

$$\begin{aligned} (47) =& \frac{(\delta_0)_{h_1} \cdot (q-2M_1)_{h_1} \cdot (q_0)_{\sum_{i=2}^{k} 2h_i}}{(q)_{\sum_{i=1}^{k} h_i} \prod_{i=1}^{k} (q)_{h_i}} \\ \ge& \frac{(\delta_0)_{h_1} \cdot (q-2M_1)_{h_1}}{(q)_{h_1} \cdot (q)_{h_1}} \cdot \prod_{i=2}^{k} \frac{(q_0-\sum_{j=2}^{i-1} 2h_j)_{2h_i}}{(q)_{h_i} \cdot (q)_{h_i}} \\ \ge& \prod_{i=0}^{h_1-1} \left(1-\frac{\delta_1+\cdots+\delta_{t-1}}{q-i}\right) \cdot \left(1-\frac{2M_1}{q-i}\right) \end{aligned}$$

$$\times \prod_{\ell=2}^{k}\prod_{i=0}^{h_\ell-1}\left(1-\frac{\delta_1+\cdots+\delta_{t-1}+h_1+4(t-1)M_1+h_\ell+\sum_{j=2}^{\ell-1}2h_j}{q-i}\right)^2$$
$$\geq 1-\frac{(t+1)M_1M_{t-1}}{q-\sum_{i=1}^{k}h_i}-\frac{10(k-1)(t-1)M_1M_{t-1}+4(k-1)^2M_{t-1}^2}{q-\sum_{i=1}^{k}h_i}$$
$$\geq 1-\frac{2(t+1)(10k-9)M_1M_{t-1}}{q}-\frac{8(k-1)^2M_{t-1}^2}{q}$$
$$=1-\frac{2(t+1)(10k-9)q}{N^{t/(t+1)}}-\frac{8(k-1)^2q}{N^{(2t-2)/(t+1)}},$$

where the second inequality holds due to the fact that $\delta_0 = q-\delta_1-\cdots-\delta_{t-1}$, $q_0=\delta_0-h_1-4(t-1)M_1$, the third inequality holds for the fact that $h_1,\ldots,h_k \leq M_{t-1}, \delta_1,\delta_2,\ldots,\delta_{t-1} \leq M_1$, the forth inequality holds due to the fact that $\sum_{i=1}^{k}h_i \leq \frac{q}{2}$, and the second equality holds by substituting $M_1 = \frac{q}{N^{1/(t+1)}}, M_{t-1}=\frac{q}{N^{(t-1)/(t+1)}}$.

$$(48) = \frac{(N)_q(N-4q)_q\prod_{\ell=0}^{t-3}\Big(N-(5+2\ell)q\Big)_q(N)_q}{\prod_{j=1}^{t}(N-jq)_q(N-kq)_q}\cdot\frac{\Big(N-4q-(2t-4)h_1-\sum_{i=2}^{k}(4t-8)h_i\Big)_q}{(N-4q)_q}$$
$$\cdot\prod_{\ell=0}^{t-3}\frac{\Big(N-(5+2\ell)q-(2t-5-2\ell)h_1-\sum_{j=2}^{k}(4t-10-4\ell)h_j\Big)_q}{\Big(N-(5+2\ell)q\Big)_q}$$
$$=\prod_{i=0}^{q-1}\underbrace{\left(1+\frac{q(N-i)^t+o(q(N-i)^t)}{\prod_{j=1}^{t}(N-i-jq)(N-i-kq)}\right)}_{>1}\cdot\prod_{i=1}^{q-1}\left(1-\frac{(2t-4)h_1+\sum_{i=2}^{k}(4t-8)h_i}{N-i-4q}\right)$$
$$\cdot\prod_{\ell=0}^{t-3}\prod_{i=1}^{q-1}\left(1-\frac{(2t-5-2\ell)h_1+\sum_{j=2}^{k}(4t-10-4\ell)h_j}{N-i-\ell q}\right)$$
$$\geq 1-\frac{(2t-4)qM_{t-1}+(4t-8)(k-1)qM_{t-1}}{N-(t-2)q}-\frac{(2t-5)(k-2)qM_{t-1}+(4t-10)(k-2)(k-1)qM_{t-1}}{N-(t-2)q}$$
$$\geq 1-\frac{\big((4t-10)k^2-(6t-17)k-(2t-6)\big)qM_{t-1}}{N-(t-2)q}$$
$$\geq 1-\frac{2\big((4t-10)k^2-(6t-17)k-(2t-6)\big)q^2}{N^{2t/(t+1)}},$$

where the firsrt inequality holds due to the fact that $h_1,\ldots,h_k \leq M_{t-1}$, and the second inequality holds by substituting $2(t-2)q \leq \frac{N}{2}, M_{t-1}=\frac{q}{N^{(t-1)/(t+1)}}$.

$$(49) = \frac{\prod_{j=0}^{t-4}\Big(N-4q-(2j+1)h_1\Big)_{h_1}}{\Big(N-5q-(2t-3)h_1-\sum_{i=2}^{k}(4t-6)h_i\Big)_{h_1+\sum_{i=2}^{k}2h_i}}$$
$$\times\frac{\prod_{i=2}^{k}\prod_{j=0}^{2t-5}\Big(N-6q-(3t-6)h_1-\sum_{l=2}^{i-1}3(2t-4)h_l-3jh_i\Big)_{h_i}}{\prod_{\ell=0}^{t-3}\Big(N-(6+2\ell)q-(2t-4-2\ell)h_1-\sum_{j=2}^{k}(4t-8-4\ell)h_j\Big)_{h_1+\sum_{i=2}^{k}2h_i}}$$
$$\times\Big(N-(t+1)q+\sum_{i=1}^{k}h_i\Big)_{\sum_{i=1}^{k}h_i}\Big(N-(k+1)q+\sum_{i=1}^{k}h_i\Big)_{\sum_{i=1}^{k}h_i}$$
$$=\frac{\prod_{j=0}^{t-4}\Big(N-4q-(2j+1)h_1\Big)_{h_1}}{\prod_{\ell=0}^{t-3}\Big(N-(6+2\ell)q-(2t-4-2\ell)h_1-\sum_{j=2}^{k}(4t-8-4\ell)h_j\Big)_{h_1}}$$

$$
\begin{aligned}
&\times \frac{\left(N-(t+1)q+\sum_{i=1}^{k}h_i\right)_{h_1}\left(N-(k+1)q+\sum_{i=1}^{k}h_i\right)_{h_1}}{\left(N-5q-(2t-3)h_1-\sum_{i=2}^{k}(4t-6)h_i\right)_{h_1}}\\
&\times \frac{\prod_{i=2}^{k}\prod_{j=0}^{2t-5}\left(N-6q-(3t-6)h_1-\sum_{l=2}^{i-1}3(2t-4)h_l-3jh_i\right)_{h_i}}{\prod_{\ell=0}^{t-3}\left(N-(6+2\ell)q-(2t-3-2\ell)h_1-\sum_{j=2}^{k}(4t-8-4\ell)h_j\right)_{\sum_{i=2}^{k}2h_i}}\\
&\times \frac{\left(N-(t+1)q+\sum_{i=1}^{k}h_i\right)_{\sum_{i=2}^{k}h_i}\left(N-(k+1)q+\sum_{i=1}^{k}h_i\right)_{\sum_{i=2}^{k}h_i}}{\left(N-5q-(2t-2)h_1-\sum_{i=2}^{k}(4t-6)h_i\right)_{\sum_{i=2}^{k}2h_i}}\\
\geq&\prod_{i=0}^{h_1-1}\underbrace{\left(1+\frac{\frac{1}{2}(t^2-7t+16)qN^{(t-1)h_1}+o\left(qN^{(t-1)h_1}\right)}{(N-5q)_{h_1}\prod_{\ell=0}^{t-3}\left(N-(6+2\ell)q\right)_{\sum_{i=2}^{k}h_1}}\right)}_{>1}\\
&\times\prod_{i=2}^{k}\prod_{j=0}^{2h_i-1}\underbrace{\left(1+\prod_{\ell=0}^{t-3}\frac{2\ell q-(t-2\ell+3)h_1+\sum_{j=2}^{k}(4t-10-4\ell)h_j-\sum_{l=2}^{i-1}3(2t-4)h_l-6\ell h_i}{N-j-(6+2\ell)q-(2t-3-2\ell)h_1-\sum_{j=2}^{k}(4t-10-4\ell)h_j}\right)}_{>1}\\
&\times\prod_{i=2}^{k}\prod_{j=0}^{h_i-1}\left(1-\frac{(t-4)q+(2t-1)h_1+\sum_{i=2}^{k}(4t-5)h_i}{N-j-5q-(2t-2)h_1-\sum_{i=2}^{k}(4t-6)h_i}\right)\\
&\times\prod_{i=2}^{k}\prod_{j=0}^{h_i-1}\left(1-\frac{(k-4)q+(2t-1)h_1+\sum_{i=2}^{k}(4t-4)h_i}{N-j-5q-(2t-2)h_1-\sum_{i=2}^{k}(4t-5)h_i}\right)\\
\geq&\prod_{i=2}^{k}\left(1-\frac{(k-4)qh_i+(2t-1)h_1h_i+h_i\sum_{i=2}^{k}(4t-4)h_i}{N-h_i-5q-(2t-2)h_1-\sum_{i=2}^{k}(4t-5)h_i}\right)^2\\
\geq&1-2\frac{(k-1)(k-4)qh_i+(k-1)(2t-1)h_1h_i+(k-1)h_i\sum_{i=2}^{k}(4t-4)h_i}{N-h_i-5q-(2t-2)h_1-\sum_{i=2}^{k}(4t-5)h_i}\\
\geq&1-4\frac{(k-1)(k-4)qM_{t-1}+(k-1)(2t-1)M_{t-1}^2+4(k-1)(t-1)M_{t-1}^2}{N}\\
\geq&1-\frac{4(k-1)(k-4)q^2}{N^{2t/(t+1)}}-\frac{4(k-1)(6t-5)q^2}{N^{(3t-1)/(t+1)}}\\
\geq&1-\frac{4(k-1)(k-4)q^2}{N^{2t/(t+1)}}-\frac{4(k-1)(6t-5)q^2}{N^{2t/(t+1)}}.
\end{aligned}
$$

where the forth inequality holds for the fact that $h_1,\ldots,h_k\leq M_{t-1},\sum_{i=1}^{k}h_i\leq\frac{q}{2}$ and $h_i+5q+(2t-2)h_1+\sum_{i=2}^{k}(4t-5)h_i\leq q$, and the last equality holds by substituting $M_{t-1}=\frac{q}{N^{(t-1)/(t+1)}}$.

Putting all pieces together, we obtain that

$$
\begin{aligned}
(36)=&\Pr_{P_1\leftarrow_\$\mathcal{P}_n}[E_K\downarrow\mathcal{Q}_{0,t}^{\equiv}\wedge E_K\downarrow_w\mathcal{Q}_0^*\mid P_1\downarrow\mathcal{Q}_1]\\
\geq&(46)\times(47)\times(48)\times(49)\\
\geq&\frac{1}{(N)_q}\times\Bigg(1-\frac{8kN}{q^2}-\frac{2(t+1)(10k-9)q}{N^{t/(t+1)}}-\frac{8(k-1)^2q}{N^{(2t-2)/(t+1)}}\\
&-\frac{2\left((4t-10)k^2-(6t-17)k-(2t-6)\right)q^2}{N^{2t/(t+1)}}\\
&-\frac{4(k-1)(k-4)q^2}{N^{2t/(t+1)}}-\frac{4(k-1)(6t-5)q^2}{N^{2t/(t+1)}}\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \frac{1}{(N)_q} \times \left(1 - \frac{(10t^3+10t^2)q}{N^{t/(t+1)}} - \frac{(2t^5+t^4+12t^3)q^2}{N^{2t/(t+1)}} - \frac{2t^4 q}{N^{(2t-2)/(t+1)}} - \frac{4t^2 N}{q^2}\right) \\
&\geq \frac{1}{(N)_q} \times \left(1 - \frac{22t^4 q}{N^{t/(t+1)}} - \frac{15t^5 q^2}{N^{2t/(t+1)}} - \frac{4t^2 N}{q^2}\right),
\end{aligned}
$$

where $k = \frac{t^2-t}{2} + 1 \geq \frac{t^2}{2} (t \geq 4)$, then the proof completes. □

F More Applications

To illustrate the generality of our proof technique, we also apply it to other variants of 3-round KAC. Except for a slight difference in the definition of bad transcripts, the proofs and calculations are very similar to 3-round KACSP, so we omit the details.

Theorem 17 ($P_1P_1P_2$-Construction)**.** *Consider the $P_1P_1P_2$-construction. For any distinguisher $\mathcal{D}$ making q_e construction queries and q_p queries to each permutation, the following upper bound holds:*

$$
\mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_1P_2}(\mathcal{D}) \leq \begin{cases} \frac{10q}{N^{3/4}} + \frac{13q^2}{N^{3/2}} + \frac{12q^4}{N^3} + \frac{2q^6}{N^5} + \frac{4q}{N} + \frac{16N}{q^2}, & \text{for } q = \omega(N^{1/2}) \\ \frac{5q}{N^{7/10}} + \frac{13q^2}{N^{7/5}} + \frac{9q}{N^{3/4}} + \frac{12q^4}{N^3} + \frac{2q^6}{N^5} + \frac{16}{N^{1/10}}, & \text{for } q = \mathcal{O}(N^{1/2}) \end{cases},
$$

where $q := \max\{q_e, q_p\}$.

Theorem 18 ($P_1P_2P_1$-Construction)**.** *Consider the $P_1P_2P_1$-construction. For any distinguisher $\mathcal{D}$ making q_e construction queries and q_p queries to each permutation, the following upper bound holds:*

$$
\mathbf{Adv}^{\mathsf{SPRP}}_{P_1P_2P_1}(\mathcal{D}) \leq \begin{cases} \frac{7q}{N^{3/4}} + \frac{21q^2}{N^{3/2}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{4q}{N} + \frac{16N}{q^2}, & \text{for } q = \omega(N^{1/2}) \\ \frac{4q}{N^{7/10}} + \frac{21q^2}{N^{7/5}} + \frac{9q}{N^{3/4}} + \frac{8q^4}{N^3} + \frac{6q^6}{N^5} + \frac{16}{N^{1/10}}, & \text{for } q = \mathcal{O}(N^{1/2}) \end{cases},
$$

where $q := \max\{q_e, q_p\}$.