MJH: A Faster Alternative to MDC-2

Jooyoung Lee[1]* and Martijn Stam[2]

[1] Sejong University, Seoul, Korea, jlee05@sejong.ac.kr
[2] University of Bristol, Bristol, United Kingdom, martijn.stam@bristol.ac.uk

Abstract. In this paper, we introduce a new class of double-block-length hash functions. Using the ideal cipher model, we prove that these hash functions, dubbed MJH, are asymptotically collision resistant up to $O(2^{n(1-\epsilon)})$ query complexity for any $\epsilon > 0$ in the iteration, where n is the block size of the underlying blockcipher.
When based on n-bit key blockciphers, our construction, being of rate 1/2, provides better provable security than MDC-2, the only known construction of a rate-1/2 double-length hash function based on an n-bit key blockcipher with non-trivial provable security. Moreover, since key scheduling is performed only once per message block for MJH, our proposal significantly outperforms MDC-2 in efficiency.
When based on a $2n$-bit key blockcipher, we can use the extra n bits of key to increase the amount of payload accordingly. Thus we get a rate-1 hash function that is much faster than existing proposals, such as Tandem-DM with comparable provable security. This is the full version of [19].

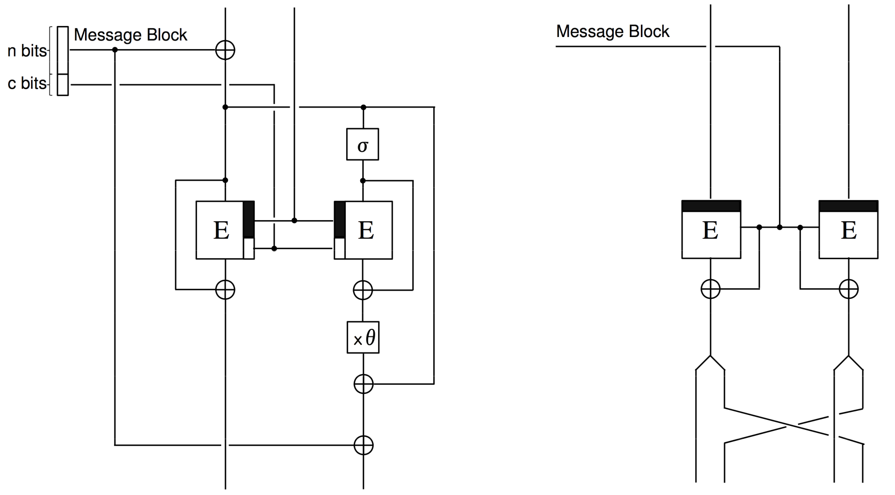
1 Introduction

A cryptographic hash function takes a message of arbitrary length, and returns a bit string of fixed length. The most common way of hashing variable length messages is to iterate a fixed-size compression function according to the Merkle-Damgård paradigm [7, 23]. The underlying compression function can either be constructed from scratch, or be built upon off-the-shelf cryptographic primitives such as blockciphers. In this work, we will focus on blockcipher-based construction as one of the widely studied approaches [1, 2, 10–13, 16, 25–33]. By instantiating a blockcipher-based construction with an extensively studied blockcipher, one can conveniently transfer the trust in the existing blockcipher to the hash function. This approach is particularly useful in highly constrained environments such as RFID systems, since a single implementation of a blockcipher can be used for both a blockcipher and a hash function. Compared to blockcipher-based hash functions, most dedicated hash functions require significant amounts of state and the operations in their designs are not always as hardware friendly [3].

One might want to construct a blockcipher-based hash function using a $2n$-to-n bit compression function that makes a certain number of calls to an n-bit blockcipher. However, such a function, called a *single-block-length* (SBL) compression function, might be vulnerable to collision attacks due to its short output length. For example, one could successfully mount a birthday attack on a compression function based on AES-128 using approximately 2^{64} queries. Based on DES, the threshold number of queries would be reduced to 2^{32}. This observation motivated substantial research on *double-block-length* (DBL) hash functions, where the output length is twice the block length of the underlying blockcipher(s).

An important distinction can be made on whether the underlying n-bit blockcipher has n-bit or $2n$-bit keys. Whereas for the latter scenario several proposals with good provable security are known, the construction of a double-length hash function based on an n-bit-key blockcipher remains elusive (see also the related works section). Currently the only known candidate providing both efficiency and a reasonable level of provable security is MDC-2 [6, 24], which makes two calls to an n-bit key blockcipher to compress a single message block (thus its rate, the ratio of message blocks hashed per blockcipher calls, equals 1/2). In 2007, 20 years after its original proposal, Steinberger

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2013R1A1A2007488).



(a) MJH compression function.

(b) MDC-2 compression function.

Fig. 1. The MJH compression function defined by an $(n+c)$-bit key blockcipher E $(c \geq 0)$, an involution σ and a constant $\theta \in \mathbb{F}_{2^n} \backslash \mathbb{F}_2$ and the MDC-2 compression function based on an n-bit key blockcipher E.

was the first to provide a non-trivial bound on the collision resistance of MDC-2 in the ideal cipher model [33]. In particular, he showed that an adversary asking fewer than $2^{3n/5-\epsilon}$ queries (for any fixed $\epsilon > 0$) has only a negligible chance of finding a collision. The best attack against MDC-2 however still requires $\Omega(2^{n-\log n})$ queries [15], leaving a considerable gap between the upper and lower bounds for its security.

Our contribution. We propose a new construction, dubbed MJH, that significantly outperforms MDC-2 both in terms of efficiency and what can currently be proven about it. Figures 1(a) and 1(b) depict our proposed compression function and the MDC-2 compression function, respectively. A formal definition of MJH will follow in Section 3.1. From a high level, we first construct a $2n$-bit to $2n$-bit function by concatenating the output of two parallel blockcipher calls run in Davies-Meyer mode. Here we use Hirose's trick [13] of an involution without fixed points to achieve implicit domain separation so we can use the same blockcipher for both strains. The resulting, reasonably random looking function is subsequently used as primitive for the JH construction [34], creating a $(3n+c)$-bit to $2n$-bit compression function and iterating it using the Merkle-Damgård transform to obtain a hash function.

Efficiency comparison. Per message block, the MJH construction makes two calls to a blockcipher with the same key. In addition, there are four n-bit xors (since the last two xors on the right strain can be merged) and the operations σ and θ. For the involution σ it suffices to toggle a single bit. The multiplication by the nonzero constant θ can be efficiently implemented by an n-bit wise shift (corresponding to multiplication by x in polynomial representation of the field) followed by a conditional xor of a mask (corresponding to reduction modulo the minimal polynomial).

MDC-2 on the other hand has two calls to a blockcipher with *distinct* keys, but it only needs two n-bit xors and a swap to complete evaluation of the compression function. Depending on platform and implementation (and without AES instruction set), AES-128 key scheduling takes up around 25% to 35% (of the cycles) of a single call to the blockcipher. This translates into a speedup of about 10% to 15% of MJH over MDC-2 (the extra key-schedule processing is not an issue here). We also note that due to recent related-key attacks [5] future blockciphers can be expected to have key scheduling algorithms that are even costlier than current ones; this would tilt the balance even further in our advantage.

Recently, Bos, Özen and Stam [4] provided a software benchmark for a large range of 256-bit blockcipher-based hash functions. With the underlying blockcipher instantiated with AES, MJH turned out to the fastest one among the existing schemes.

On hardware we obtain an advantage if the blockciphers are implemented twice (to increase throughput) since the key scheduling circuit can be shared. Our construction does need an extra feedforward in comparison to MDC-2 (2 vs. 1), which is a disadvantage posing memory constraints on the hardware circuit [3].

Security results. In the ideal cipher model, we prove that our MJH hash functions are collision resistant up to $O(2^{n(1-\epsilon)})$ query complexity for any $\epsilon > 0$ in the iteration, improving over the bound $O(2^{\frac{2n}{3}-\log n})$ given in [19]. Like prior works dealing with collision resistance in the iteration [22,32], in the proof we consider a graph whose nodes correspond to chaining variables and where edges are drawn whenever an adversary has asked the "correct" queries to connect two chaining variables.

As a first observation, note that queries to the blockcipher are paired by the involution σ (just as in Hirose's scheme); we will call such a pair a *query-response cycle.* A challenge presents itself in that any query-response cycle typically adds 2^{n+1} edges to the graph due to the JH structure. Our core observation is that *any pair* of query-response cycles can *only* be connected in four possible ways (the use of σ introduces the not-quite-uniqueness), which in turn allows us to put even more stringent bounds on the number of chains of connected query-response cycles. This suffices to bound the probability of the adversary connecting two largish components in the graph; bounding the probability of actually constructing a collision still involves a considerable amount of additional case analysis.

Asymptotically almost optimal, our bound is the best one known for rate-1/2 n-bit key blockcipher based DBL hash functions. However, we note that our security bound is rather theoretical, since a large constant is hidden in the query complexity $O(2^{n(1-\epsilon)})$. For this reason, for practical parameters, say $n = 128$, our result guarantees only 81-bit security, still far from optimal. Nevertheless, this is better than Steinberger's bound for MDC-2. See Table 2 in Section 3.6 for numerical comparison in detail.

Our analysis of MJH also opens up the possibility to get comparable bounds for JH itself. To the best of our knowledge, currently the best (published) bounds for JH appeared in [17] showing collision resistance up to roughly $2^{n/4}$ queries (where n is the internal JH state size), whereas a straightforward mapping of our bounds would get closer to $2^{n/2}$. Unfortunately, for JH itself dealing with inverse (permutation) queries creates difficulties that we could avoid for MJH (by internally using Davies-Meyer, which would be meaningless for JH).

Related work. While many DBL compression functions of rate-1 have been proposed, unfortunately it turned out that a large class do not provide security in terms of collision resistance and preimage resistance beyond that already offered by single-block-length constructions [10, 11, 14]. This holds true both for constructions based on blockciphers with n-bit keys or $2n$-bit keys.

In the latter category, Lucks recently proposed the first DBL hash function of rate-1 with (almost) optimal security in the iteration [22,25]. Later, an alternative rate-1 secure DBL *compression* function was given by Stam [32] and generalized by Lee and Steinberger [20]. However, both constructions use full finite field multiplications, significantly degrading their efficiency.

Classical DBL compression functions of rate below 1 include MDC-2, MDC-4, Tandem-DM and Abreast-DM [6,16]. We have already discussed MDC-2 and the remarkable results obtained by Steinberger [33]. MDC-4 also uses an n-bit key blockcipher, but it is twice as inefficient.

Both Tandem-DM and Abreast-DM [6,16] are rate-1/2 hash functions based on a blockcipher with $2n$-bit key. The main challenge providing a proof is the fact that the same blockcipher is called twice, but recently Lee, Stam, and Steinberger [21] proved the security of Tandem-DM (correcting [8]). As in the case of MDC-2, the security bound obtained is parameterized: optimizing the parameter gives collision resistance of Tandem-DM up to the birthday bound. The collision resistance of Abreast-DM was independently proved in [9] and [18].

While the design of hash functions based on $2n$-bit key blockciphers is considerably easier than that based on n-bit key blockciphers, the former—while more robust with respect to for instance preimage resistance—are typically less efficient even at the same "rate". Indeed, a blockcipher with $2n$-bit keys is required to provide $2n$-bit security as opposed to n-bit security for the smaller n-bit key blockcipher. For example, AES-256 consists of 14 rounds, 4 rounds more than AES-128. So as a first rough estimate (ignoring key scheduling) one expects AES-256 to be about 40% slower than AES-128. As an aside, AES-256 no longer guarantees a sufficient level of security due to the recent related-key attacks by Biryukov and Khovratovich [5].

2 Preliminaries

General notation. Let $\mathbb{F}_{2^n}$ denote a finite field of order 2^n. Throughout our work, we will identify $\mathbb{F}_{2^n}$ and $\{0,1\}^n$, assuming a fixed mapping between the two sets. For two bitstrings x and y, $x||y$ denotes the concatenation of x and y. For a bitstring $x \in \{0,1\}^{2n}$, x_L and x_R denote the unique n-bit strings such that $x = x_L||x_R$.

The ideal cipher model. For positive integers κ and n, let $BC(\kappa, n)$ be the set of all blockciphers with n-bit blocks and κ-bit keys. In the ideal cipher model, a (κ, n)-blockcipher E is chosen from $BC(\kappa, n)$ uniformly at random. It allows for two types of oracle queries $E(K, X)$ and $E^{-1}(K, Y)$ for $X, Y \in \{0,1\}^n$ and $K \in \{0,1\}^\kappa$. The response to an inverse query $E^{-1}(K, Y)$ is $X \in \{0,1\}^n$ such that $E(K, X) = Y$. Here, X, Y and K are called plaintext, ciphertext, and key, respectively. In this paper, we assume $\kappa = n + c$ for some $c \geq 0$.

The Merkle-Damgård transform. For convenience, we recall the Merkle-Damgård transform as it will be applied to our (double-block-length) construction. Let $\mathsf{pad} : \{0,1\}^* \to \bigcup_{i=1}^{\infty} \{0,1\}^{\kappa i}$ be an injective padding. With this padding scheme and a predetermined constant $IV \in \{0,1\}^{2n}$, the *Merkle-Damgård transform* produces a variable-input-length function $MD[F] : \{0,1\}^* \to \{0,1\}^{2n}$ from a fixed-input-length function $F : \{0,1\}^{2n} \times \{0,1\}^\kappa \to \{0,1\}^{2n}$. For $M \in \{0,1\}^*$ such that $|\mathsf{pad}(M)| = l\kappa$, $MD[F](M)$ is computed as follows.

```
Function $MD[F](M)$
  $u[0] \leftarrow IV$
  Break $\mathsf{pad}(M)$ into $\kappa$-bit blocks, $\mathsf{pad}(M) = M[1]||\ldots||M[l]$
  for $i \leftarrow 1$ to $l$ do
      $u[i] \leftarrow F(u[i-1], M[i])$
  return $u[l]$
```

Collision resistance. We review the definition of collision resistance *in the information-theoretic model.* Given a function $H = H[\mathcal{P}]$ and an IT adversary $\mathcal{A}$ both with oracle access to an ideal primitive $\mathcal{P}$, the collision resistance of H against $\mathcal{A}$ is estimated by the $\mathbf{Exp}_H^{\mathsf{coll}}(\mathcal{A})$ experiment.

```
Experiment $\mathbf{Exp}_H^{\mathsf{coll}}(\mathcal{A})$
  $\mathcal{A}^{\mathcal{P}}$ updates $\mathcal{Q}$
  if $\exists\ M \neq M'$ and $u$ such that $u = H_{\mathcal{Q}}(M) = H_{\mathcal{Q}}(M')$ then
      output 1
  else
      output 0
```

This experiment records every query-response pair that $\mathcal{A}$ obtains by oracle queries into a *query history* $\mathcal{Q}$. We write $u = H_{\mathcal{Q}}(M)$ if $\mathcal{Q}$ contains all the query-response pairs required to compute $u = H(M)$. At the end of the experiment, $\mathcal{A}$ would like to find two distinct evaluations yielding a collision. The *collision-finding advantage* of $\mathcal{A}$ is defined to be

$$\mathbf{Adv}_H^{\mathsf{coll}}(\mathcal{A}) = \mathbf{Pr}\left[\mathbf{Exp}_H^{\mathsf{coll}}(\mathcal{A}) = 1\right].$$

The probability is taken over the random choice of $\mathcal{P}$ and $\mathcal{A}$'s coins (if any). For $q > 0$, we define $\mathbf{Adv}_H^{\mathsf{coll}}(q)$ as the maximum of $\mathbf{Adv}_H^{\mathsf{coll}}(\mathcal{A})$ over all adversaries $\mathcal{A}$ making at most q queries.

3 The MJH Hash Functions and their Collision Resistance

3.1 The MJH Construction

For $c \geq 0$, let E be an $(n+c)$-bit key blockciphers in $BC(n+c, n)$. Let σ be an involution on $\{0,1\}^n$ with no fixed point, and let $\theta \neq 0, 1$ be a constant in $\mathbb{F}_{2^n}$. Then (σ, θ) defines a compression function $\{0,1\}^{2n} \times \{0,1\}^{n+c} \to \{0,1\}^{2n}$ by

$$\begin{aligned} G[\sigma,\theta] : \{0,1\}^{2n} \times \{0,1\}^{n+c} &\longrightarrow \{0,1\}^{2n} \\ (u_L || u_R, z || z') &\longmapsto (v_L || v_R), \end{aligned}$$

where $v_L = E(K, X) + X$ and $v_R = \theta\left(E(K, \sigma(X)) + \sigma(X)\right) + X + z$ for $X = u_L + z$ and $K = (u_R || z')$. $G[\sigma, \theta]$ is depicted in Figure 1(a). The compression function $G[\sigma, \theta]$ is fed to the Merkle-Damgård transform, yielding the *MJH hash function* $H[\sigma, \theta] = MD[G[\sigma, \theta]]$ associated with (σ, θ).

Since the padding is injective, we can simplify our collision analysis by assuming that the domain of a MJH hash function is $\bigcup_{i=1}^{\infty} \{0,1\}^{(n+c)i}$ (and ignore the padding scheme).

3.2 Query-Response Cycles and a Modified Adversary

Let $H = H[\sigma, \theta]$ be the MJH hash function defined by (σ, θ), and let $\mathcal{A}$ be an information-theoretic adversary with oracle access to E and E^{-1}. Note that $\mathcal{A}$ records a triple (X, K, Y) into the query history $\mathcal{Q}$ if $\mathcal{A}$ asks for $E(K, X)$ and gets back Y, or if it asks for $E^{-1}(K, Y)$ and gets back X. Since σ is an involution, it holds that $X_\sigma = \sigma(X)$ iff $X = \sigma(X_\sigma)$. In other words, σ induces a natural way to pair queries: for (X, K, Y) and (X_σ, K, Y_σ) in $\mathcal{Q}$ with $X_\sigma = \sigma(X)$ we call

$$\Delta = ((X, K, Y), (X_\sigma, K, Y_\sigma))$$

a *query-response cycle* (or simply a cycle) and the corresponding queries each other's *conjugates*.

We can now transform $\mathcal{A}$ into an adversary $\mathcal{B}$ that records its query history $\mathcal{Q}_\Delta$ in terms of query-response cycles as described in Figure 2. If $\mathcal{A}$ makes at most q queries, then the corresponding adversary $\mathcal{B}$ makes at most $2q$ queries, and records at most q query-response cycles. Since

$$\mathbf{Adv}_H^{\mathsf{coll}}(\mathcal{A}) \leq \mathbf{Adv}_H^{\mathsf{coll}}(\mathcal{B}),$$

```
Algorithm $\mathcal{B}^{E,E^{-1}}$
  $\mathcal{Q}_\Delta \leftarrow \emptyset$
  Run $\mathcal{A}$
  if $\mathcal{A}$ makes a fresh query for $E(K,X)$ then
      Make queries for $Y = E(K,X)$ and $Y_\sigma = E(K,X_\sigma)$ for $X_\sigma = \sigma(X)$
      $\mathcal{Q}_\Delta \leftarrow \mathcal{Q}_\Delta \cup \{\Delta\}$, where $\Delta = ((X,K,Y),(X_\sigma,K,Y_\sigma))$
      Return $Y$ to $\mathcal{A}$
  else if $\mathcal{A}$ makes a fresh query for $E^{-1}(K,Y)$ then
      Make queries for $X = E^{-1}(K,Y)$ and $Y_\sigma = E(K,X_\sigma)$ for $X_\sigma = \sigma(X)$
      $\mathcal{Q}_\Delta \leftarrow \mathcal{Q}_\Delta \cup \{\Delta\}$, where $\Delta = ((X,K,Y),(X_\sigma,K,Y_\sigma))$
      Return $X$ to $\mathcal{A}$
  else
      Return the response using query history $\mathcal{Q}_\Delta$
```

Fig. 2. The modified adversary $\mathcal{B}$. A query is called "fresh" if its response cannot be obtained from $\mathcal{B}$'s query history.

it suffices to consider the security of H against a modified adversary that records exactly q query-response cycles.

Henceforth, the i-th query-response cycle is denoted $\Delta^i = \big((X^i, K^i, Y^i), (X^i_\sigma, K^i, Y^i_\sigma)\big)$, where we assume that (X^i, K^i, Y^i) was obtained before $(X^i_\sigma, K^i, Y^i_\sigma)$. We will write $Z^i = X^i + Y^i$, $Z^i_\sigma = X^i_\sigma + Y^i_\sigma$ and $K^i = (K^i_1 || K^i_2)$ where $K^i_1 \in \{0,1\}^n$ and $K^i_2 \in \{0,1\}^c$. Additionally, we will write $N = 2^n$ and $N' = N - 2q$ assuming $N' > N/2$.

3.3 Graph Representation of the Adversary's Endeavors

Each query-response cycle $\Delta^i = \big((X^i, K^i, Y^i), (X^i_\sigma, K^i, Y^i_\sigma)\big) \in \mathcal{Q}_\Delta$ determines two types of evaluations

$$G[\sigma,\theta] : \big((X^i + z)||K^i_1, z||K^i_2\big) \longmapsto (X^i + Y^i)\,||\,(\theta(X^i_\sigma + Y^i_\sigma) + X^i + z), \tag{1}$$

and

$$G[\sigma,\theta] : \big((X^i_\sigma + z)||K^i_1, z||K^i_2\big) \longmapsto (X^i_\sigma + Y^i_\sigma)\,||\,(\theta(X^i + Y^i) + X^i_\sigma + z), \tag{2}$$

for $z \in \{0,1\}^n$.

We now define a directed graph $\mathcal{G}$ on $\{0,1\}^{2n}$ where the nodes correspond to chaining variables and edges are added depending on the evaluations of $G[\sigma,\theta]$ the adversary can make given the available query-response cycles. In other words, a directed edge from u to v labeled i is added to $\mathcal{G}$ when the i-th query-response cycle determines an evaluation $G[\sigma,\theta](u, z||z') = v$ for some $z \in \{0,1\}^n$ and $z' \in \{0,1\}^c$. Such a connection is denoted by $u \overset{i}{\to} v$. More specifically, we write $u \overset{i+}{\to} v$ (resp. $u \overset{i-}{\to} v$) when the evaluation of $G[\sigma,\theta]$ is obtained by (1) (resp. (2)). Note that an edge $u \overset{i+}{\to} v$ appearing in $\mathcal{G}$ implies $u_R = K^i_1$, $v_L = Z^i$ and $u_L + v_R = \theta Z^i_\sigma$. On the other hand, $u \overset{i-}{\to} v$ implies $u_R = K^i_1$, $v_L = Z^i_\sigma$ and $u_L + v_R = \theta Z^i$. This property will be frequently used in the following analysis.

A *path* $\mathcal{P}$ is a subgraph of $\mathcal{G}$ of the form

$$u[0] \overset{(i_1,\delta_1)}{\longrightarrow} u[1] \overset{(i_2,\delta_2)}{\longrightarrow} \cdots \overset{(i_{l-1},\delta_{l-1})}{\longrightarrow} u[l-1] \overset{(i_l,\delta_l)}{\longrightarrow} u[l],$$

where $l \geq 0$, $i_1, \ldots, i_l \in [1,q]$ and $\delta_1, \ldots, \delta_l \in \{+,-\}$.[3] The nonnegative integer l is called the *length* of $\mathcal{P}$, and denoted by $|\mathcal{P}|$. Node $u[0]$ is the *start node* of $\mathcal{P}$, denoted by $\mathsf{s}(\mathcal{P})$. Node $u[l]$ is the *end node* of $\mathcal{P}$, denoted by $\mathsf{e}(\mathcal{P})$.

[3] We allow a path that consists of a single node.

We say two paths $\mathcal{P} : u[0] \xrightarrow{(i_1,\delta_1)} \cdots \xrightarrow{(i_s,\delta_s)} u[s]$ and $\mathcal{P}' : u'[0] \xrightarrow{(i'_1,\delta'_1)} \cdots \xrightarrow{(i'_t,\delta'_t)} u'[t]$ are *colliding* if $\mathsf{e}(\mathcal{P}) = \mathsf{e}(\mathcal{P}')$ and $((i_1,\delta_1), \ldots, (i_s,\delta_s)) \neq ((i'_1,\delta'_1), \ldots, (i'_t,\delta'_t))$. Furthermore, the colliding paths $\mathcal{P}$ and $\mathcal{P}'$ are called *irreducible* if $(i_s,\delta_s) \neq (i'_t,\delta'_t)$.

3.4 Decomposing a Collision

Let Coll denote the event that $\mathcal{B}$ makes a collision of H. For a fixed parameter $l \geq 2$, we can decompose the event Coll as follows.

Proposition 1. *Let C_i, $i = 0, 1, 2, 3$, be events(configurations) defined as follows (see also Fig. 3).*

- C_0 : *Some Δ^i generates edges $u \xrightarrow{i+} v$ and $u' \xrightarrow{i-} v'$ such that $v = v'$.*
- C_1 : *Some Δ^i generates an edge $u \xrightarrow{i} v$ where*
  1. *$u = \mathsf{e}(\mathcal{P}_1)$ for some path $\mathcal{P}_1 \in \mathcal{G}$ such that either ($|\mathcal{P}_1| < l$ and $\mathsf{s}(\mathcal{P}_1) = IV$) or $|\mathcal{P}_1| = l$,*
  2. *$v = \mathsf{s}(\mathcal{P}_2)$ for some* irreducible *colliding paths $\mathcal{P}_2, \mathcal{P}_3 \in \mathcal{G}$ such that either ($|\mathcal{P}_2| + |\mathcal{P}_3| \leq l$ and $\mathsf{s}(\mathcal{P}_3) = IV$) or $|\mathcal{P}_2| + |\mathcal{P}_3| = l + 1$.*
- C_2 : *Some Δ^i generates edges $u \xrightarrow{i} v$ and $u' \xrightarrow{i} v'$ where*
  1. *$u = \mathsf{e}(\mathcal{P}_1)$ for some path $\mathcal{P}_1 \in \mathcal{G}$ such that either ($|\mathcal{P}_1| < l$ and $\mathsf{s}(\mathcal{P}_1) = IV$) or $|\mathcal{P}_1| = l$,*
  2. *$v = \mathsf{s}(\mathcal{P}_2)$ and $u' = \mathsf{e}(\mathcal{P}_2)$ for some path $\mathcal{P}_2 \in \mathcal{G}$ such that $|\mathcal{P}_2| < l$.*
- C_3 : *Some Δ^i generates edges $u \xrightarrow{i} v$ and $u' \xrightarrow{i} v'$ where*
  1. *$u = \mathsf{e}(\mathcal{P}_1)$ for some path $\mathcal{P}_1 \in \mathcal{G}$ such that either ($|\mathcal{P}_1| < l$ and $\mathsf{s}(\mathcal{P}_1) = IV$) or $|\mathcal{P}_1| = l$,*
  2. *$v = \mathsf{s}(\mathcal{P}_2)$ and $v' = \mathsf{s}(\mathcal{P}_3)$ for some* irreducible *colliding paths $\mathcal{P}_2, \mathcal{P}_3 \in \mathcal{G}$ such that $0 < |\mathcal{P}_2| + |\mathcal{P}_3| \leq l$ and $|\mathcal{P}_2| \leq |\mathcal{P}_3|$.*

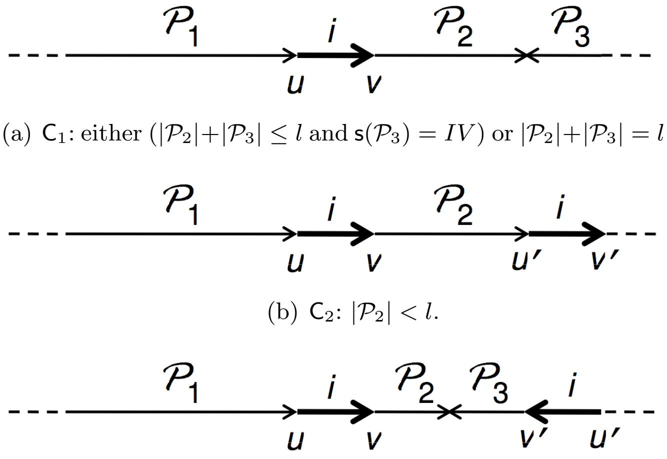
Then we have $\mathsf{Coll} \Longrightarrow \mathsf{C}_0 \vee \mathsf{C}_1 \vee \mathsf{C}_2 \vee \mathsf{C}_3$.

Proof. Suppose that the i-th query-response cycle Δ^i of $\mathcal{B}$ completes a collision of H on a certain node w. Then, for some positive integers s and t, there would exist irreducible colliding paths

$$\begin{aligned} \mathcal{P} :\quad & IV \xrightarrow{j_1} u[1] \xrightarrow{j_2} u[2] \xrightarrow{j_3} \cdots \xrightarrow{j_{s-1}} u[s-1] \xrightarrow{j_s} w, \\ \mathcal{P}' :\quad & IV \xrightarrow{j'_1} u'[1] \xrightarrow{j'_2} u'[2] \xrightarrow{j'_3} \cdots \xrightarrow{j'_{t-1}} u'[t-1] \xrightarrow{j'_t} w, \end{aligned}$$

such that $j_1, \ldots, j_s, j'_1, \ldots, j'_t \leq i$. We will say an i-labeled edge $u[\alpha] \xrightarrow{j_{\alpha+1}(=i)} u[\alpha+1]$ in $\mathcal{P}$ has *a tail of length l* if either ($\alpha \geq l$ and $j_{\alpha-l+1}, j_{\alpha-l+2}, \ldots, j_\alpha < i$) or ($\alpha < l$ and $j_1, j_2, \ldots, j_\alpha < i$). This type of i-labeled edges are similarly defined for $\mathcal{P}'$. Note that the i-labeled edge nearest IV (in any path) has a tail of length l.

Suppose that $\mathcal{P}$ contains an i-labeled edge. Among the i-labeled edges with a tail of length l in $\mathcal{P}$, choose the nearest one, say $u \xrightarrow{i} v$, to the colliding node w and look at the configuration around this edge. If there is another i-labeled edge between v and w, then we will see configuration C_2. Otherwise, we will see configuration either C_1 or C_3 around $u \xrightarrow{i} v$ without the occurrence of C_0. In particular, we can assume $|\mathcal{P}_2| \leq |\mathcal{P}_3|$ in C_3. Otherwise, we can look at the i-labeled edges with a tail of length l in $\mathcal{P}'$ and find the nearest one to the colliding node. □



(a) C_1: either ($|\mathcal{P}_2|+|\mathcal{P}_3| \le l$ and $\mathsf{s}(\mathcal{P}_3) = IV$) or $|\mathcal{P}_2|+|\mathcal{P}_3| = l+1$.

(b) C_2: $|\mathcal{P}_2| < l$.

(c) C_3: $0 < |\mathcal{P}_2| + |\mathcal{P}_3| \le l$ and $|\mathcal{P}_2| \le |\mathcal{P}_3|$.

Fig. 3. Configurations for C_1, C_2 and C_3. For each configuration, either ($|\mathcal{P}_1| < l$ and $\mathsf{s}(\mathcal{P}_1) = IV$) or $|\mathcal{P}_1| = l$.

Example 1 *Let $l = 2$. Then Coll is decomposed by D_i, $i = 0, \ldots, 15$, where*

$$
\begin{aligned}
\mathsf{D}_0 &: \exists\ u \overset{i+}{\rightarrow} v \overset{i-}{\leftarrow} w,\\
\mathsf{D}_1 &: \exists\ u \overset{i}{\rightarrow} IV \textit{ for some } u \in R,\\
\mathsf{D}_2 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\leftarrow} IV \textit{ for some } u \in R \textit{ and } j < i,\\
\mathsf{D}_3 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} IV \textit{ for some } u \in R \textit{ and } j < i,\\
\mathsf{D}_4 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\leftarrow} v' \overset{j'}{\leftarrow} IV \textit{ for some } u \in R \textit{ and } j, j' < i,\\
\mathsf{D}_5 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} w \overset{j'}{\leftarrow} IV \textit{ for some } u \in R \textit{ and } j, j' < i,\\
\mathsf{D}_6 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} w \overset{j'}{\rightarrow} IV \textit{ for some } u \in R \textit{ and } j, j' < i,\\
\mathsf{D}_7 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\leftarrow} v' \overset{j'}{\leftarrow} v'' \overset{j''}{\leftarrow} v''' \textit{ for some } u \in R \textit{ and } j, j', j'' < i,\\
\mathsf{D}_8 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} w \overset{j'}{\leftarrow} v' \overset{j''}{\leftarrow} v'' \textit{ for some } u \in R \textit{ and } j, j', j'' < i,\\
\mathsf{D}_9 &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} v' \overset{j'}{\rightarrow} w \overset{j''}{\leftarrow} v'' \textit{ for some } u \in R \textit{ and } j, j', j'' < i,\\
\mathsf{D}_{10} &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} v' \overset{j'}{\rightarrow} v'' \overset{j''}{\rightarrow} v''' \textit{ for some } u \in R \textit{ and } j, j', j'' < i,\\
\mathsf{D}_{11} &: \exists\ u \overset{i}{\rightarrow} v \overset{i}{\rightarrow} v' \textit{ for some } u \in R,\\
\mathsf{D}_{12} &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} v' \overset{i}{\rightarrow} v'' \textit{ for some } u \in R \textit{ and } j < i,\\
\mathsf{D}_{13} &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\leftarrow} v' \overset{i}{\leftarrow} v'' \textit{ for some } u \in R \textit{ and } j < i,\\
\mathsf{D}_{14} &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\leftarrow} v' \overset{j'}{\leftarrow} v'' \overset{i}{\leftarrow} v''' \textit{ for some } u \in R \textit{ and } j, j' < i,\\
\mathsf{D}_{15} &: \exists\ u \overset{i}{\rightarrow} v \overset{j}{\rightarrow} w \overset{j'}{\leftarrow} v' \overset{i}{\leftarrow} v'' \textit{ for some } u \in R \textit{ and } j, j' < i,
\end{aligned}
$$

and

$$
\begin{aligned}
R = \{&u \in \{0,1\}^{2n} : \exists\ u'' \overset{j'}{\rightarrow} u' \overset{j}{\rightarrow} u \textit{ for some } j, j' < i\}\\
&\cup \{u \in \{0,1\}^{2n} : \exists\ IV \overset{j}{\rightarrow} u \textit{ for some } j < i\} \cup \{IV\}.
\end{aligned}
$$

The events defined in Proposition 1 have been decomposed again as $\mathsf{C}_0 = \mathsf{D}_0$*,* $\mathsf{C}_1 = \mathsf{D}_1 \vee \cdots \vee \mathsf{D}_{10}$*,* $\mathsf{C}_2 = \mathsf{D}_{11} \vee \mathsf{D}_{12}$ *and* $\mathsf{C}_3 = \mathsf{D}_{13} \vee \mathsf{D}_{14} \vee \mathsf{D}_{15}$*, according to the path lengths.*

In order to upper bound the probability of the events C_1, C_2 and C_3 in the general case, we need to define certain auxiliary subsets of $\{0,1\}^{2n}$.

1. For $w \in \{0,1\}^{2n}$, $z_1, z_2 \in \{0,1\}^n$ and $k \geq 0$,

$$\begin{aligned}
\mathcal{U}_k^0(w) &= \{\mathsf{s}(\mathcal{P}) : \ \mathcal{P} \in \mathcal{G} \text{ such that } |\mathcal{P}| = k,\ \mathsf{e}(\mathcal{P}) = w\},\\
\mathcal{U}_k^1(z_1, z_2) &= \{\mathsf{e}(\mathcal{P}) : \ \mathcal{P} \in \mathcal{G} \text{ such that } |\mathcal{P}| = k,\ \mathsf{s}(\mathcal{P})_L = z_1,\ \mathsf{e}(\mathcal{P})_R = z_2\},\\
\mathcal{U}_k^2(z_1, z_2) &= \{\mathsf{e}(\mathcal{P}) : \ \mathcal{P} \in \mathcal{G} \text{ such that } |\mathcal{P}| = k,\ \theta\mathsf{s}(\mathcal{P})_L + \mathsf{e}(\mathcal{P})_L = z_1,\ \mathsf{e}(\mathcal{P})_R = z_2\}.
\end{aligned}$$

2. For $z, z_1, z_2 \in \{0,1\}^n$, k_1 and k_2 such that $k_1 + k_2 > 0$,

$$\begin{aligned}
\mathcal{V}_{k_1,k_2}^1(z) = \{\mathsf{e}(\mathcal{P}_1) : \ &\text{irreducible colliding paths } \mathcal{P}_1, \mathcal{P}_2 \in \mathcal{G}\\
&\text{such that } |\mathcal{P}_1| = k_1,\ |\mathcal{P}_2| = k_2,\ \mathsf{s}(\mathcal{P}_1)_L = \mathsf{s}(\mathcal{P}_2)_L,\ \mathsf{e}(\mathcal{P}_1)_R = \mathsf{e}(\mathcal{P}_2)_R = z\},
\end{aligned}$$

$$\begin{aligned}
\mathcal{V}_{k_1,k_2}^2(z_1, z_2) = \{\mathsf{e}(\mathcal{P}_1) : \ &\text{irreducible colliding paths } \mathcal{P}_1, \mathcal{P}_2 \in \mathcal{G}\\
&\text{such that } |\mathcal{P}_1| = k_1,\ |\mathcal{P}_2| = k_2,\ \mathsf{s}(\mathcal{P}_1)_R + \theta\mathsf{s}(\mathcal{P}_2)_L = z_1,\ \mathsf{e}(\mathcal{P}_1)_R = \mathsf{e}(\mathcal{P}_2)_R = z_2\}.
\end{aligned}$$

3. For $z \in \{0,1\}^n$ and k_1, $k_2 \geq 0$,

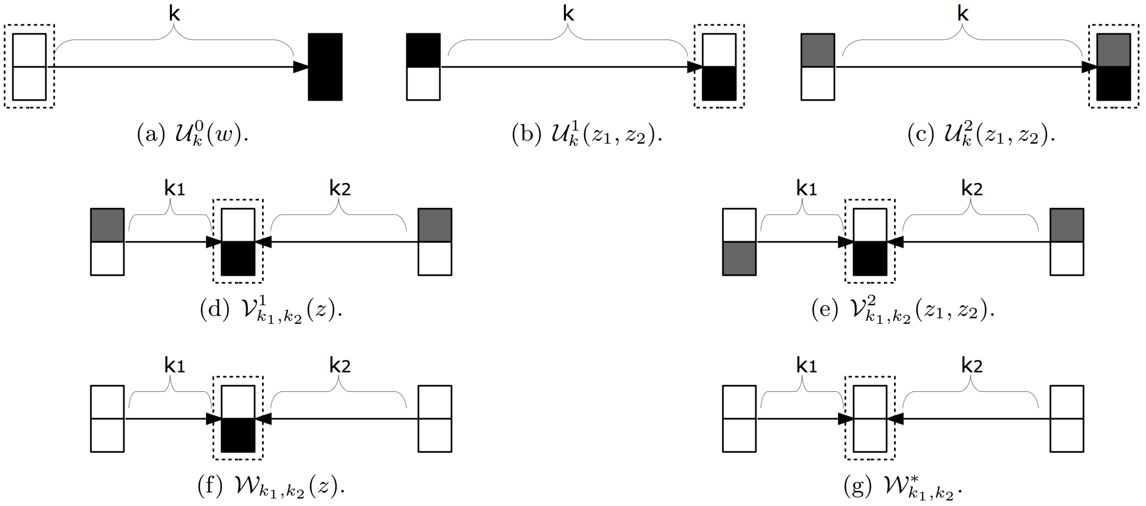
$$\begin{aligned}
\mathcal{W}_{k_1,k_2}(z) = \{\mathsf{e}(\mathcal{P}_1) : \ &\text{irreducible colliding paths } \mathcal{P}_1, \mathcal{P}_2 \in \mathcal{G}\\
&\text{such that } |\mathcal{P}_1| = k_1,\ |\mathcal{P}_2| = k_2,\ \mathsf{e}(\mathcal{P}_1)_R = \mathsf{e}(\mathcal{P}_2)_R = z\},
\end{aligned}$$

$$\mathcal{W}_{k_1,k_2}^* = \{\mathsf{e}(\mathcal{P}_1) : \ \text{irreducible colliding paths } \mathcal{P}_1, \mathcal{P}_2 \in \mathcal{G} \text{ such that } |\mathcal{P}_1| = k_1,\ |\mathcal{P}_2| = k_2\}.$$

By symmetry, we have $\mathcal{V}_{k_1,k_2}^1(z) = \mathcal{V}_{k_2,k_1}^1(z)$, $\mathcal{W}_{k_1,k_2}(z) = \mathcal{W}_{k_2,k_1}(z)$ and $\mathcal{W}_{k_1,k_2}^* = \mathcal{W}_{k_2,k_1}^*$ for any $z \in \{0,1\}^n$. Pictorial illustration of the auxiliary subsets is given in Fig 4. In terms of the auxiliary subsets, we can restate events C_1, C_2 and C_3 as follows.

Proposition 2. *Let* C_1^**,* C_2^* *and* C_3^* *be events defined as follows.*

- C_1^* *: Some* Δ^i *generates* $u \xrightarrow{i} v$ *where*
  1. *either* $u \in \mathcal{W}_{0,l}(K_1^i)$ *or* $u \in \mathcal{U}_k^1(IV_L, K_1^i)$ *for some* $0 \leq k < l$*,*
  2. $v \in \mathcal{U}_{k_1}^0(w)$ *where* $k_1 \geq 0$ *and*
     (a) *either* $w \in \mathcal{U}_{k_2}^1(IV_L, z)$ *for some* $z(= w_R) \in \{0,1\}^n$ *and* $k_2 \geq 0$ *such that* $k_1 + k_2 \leq l$ *or,*
     (b) $w \in \mathcal{W}_{k_1,k_2}^*$ *for some* $k_2 \geq 0$ *such that* $k_1 + k_2 = l + 1$*.*
- C_2^* *: Some* Δ^i *generates* $u \xrightarrow{i} v$ *where*
  1. *either* $u \in \mathcal{W}_{0,l}(K_1^i)$ *or* $u \in \mathcal{U}_k^1(IV_L, K_1^i)$ *for some* $0 \leq k < l$*,*
  2. $v \in \mathcal{U}_k^0(w)$ *where* $0 \leq k < l$ *and* $w = z || K_1^i$ *for some* $z(= u'_L) \in \{0,1\}^n$*.*
- C_3^* *: Some* Δ^i *generates* $u \xrightarrow{i} v$ *where*
  1. *either* $u \in \mathcal{W}_{0,l}(K_1^i)$ *or* $u \in \mathcal{U}_k^1(IV_L, K_1^i)$ *for some* $0 \leq k < l$*,*
  2. $v \in \mathcal{U}_{k_1}^0(w)$ *where* $k_1 \geq 0$ *and*



(a) $\mathcal{U}_k^0(w)$. (b) $\mathcal{U}_k^1(z_1, z_2)$. (c) $\mathcal{U}_k^2(z_1, z_2)$.

(d) $\mathcal{V}_{k_1,k_2}^1(z)$. (e) $\mathcal{V}_{k_1,k_2}^2(z_1, z_2)$.

(f) $\mathcal{W}_{k_1,k_2}(z)$. (g) $\mathcal{W}_{k_1,k_2}^*$.

Fig. 4. Auxiliary subsets. An upper half box (resp. a lower half box) represents the left half (resp. the right half) of a node in $\{0,1\}^{2n}$. Each black box represents a certain fixed value in it. A pair of grey boxes are related by a certain equation. A dotted box represents a node that each auxiliary subset counts.

(a) either $w \in \mathcal{V}_{k_1,k_2}^1(z)$ for some $z \in \{0,1\}^n$ and $k_1 \le k_2$ such that $0 < k_1 + k_2 \le l$ or,

(b) $w \in \mathcal{V}_{k_1,k_2}^2(u_L, z)$ for some $z \in \{0,1\}^n$ and $k_1 \le k_2$ such that $0 < k_1 + k_2 \le l$.

Then we have $\mathsf{C}_1 \Longrightarrow \mathsf{C}_1^*$, $\mathsf{C}_2 \Longrightarrow \mathsf{C}_2^*$ and $\mathsf{C}_3 \Longrightarrow \mathsf{C}_3^*$.

Proof. If an i-labeled edge completes a configuration C_α, $\alpha = 1, 2, 3$, with $|\mathcal{P}_1| = l$, then we have $u \in \mathcal{W}_{0,l}(K_1^i)$. If $|\mathcal{P}_1| = k < l$ and $\mathsf{s}(\mathcal{P}_1) = IV$, then we have $u \in \mathcal{U}_k^1(IV_L, K_1^i)$.

Suppose that two i-labeled edges complete configuration C_3 with $|\mathcal{P}_2| = k_1$ and $|\mathcal{P}_3| = k_2$. Let w denote the colliding node of $\mathcal{P}_2$ and $\mathcal{P}_3$. (Namely, $w = \mathsf{e}(\mathcal{P}_2) = \mathsf{e}(\mathcal{P}_3)$.) If the two i-labeled edges have the same sign, then for some possible value $z \in \{0,1\}^n$ for w_R, we have $w \in \mathcal{V}_{k_1,k_2}^1(z)$ and $v \in \mathcal{U}_{k_1}^0(w)$.

If the two i-labeled edges have opposite signs, then we have $w = \mathsf{e}(\mathcal{P}_1) = \mathsf{e}(\mathcal{P}_2) \in \mathcal{V}_{k_1,k_2}^2(u_L, z)$ for some possible value $z \in \{0,1\}^n$ for w_R since $u_L + v_R = \theta v'_L$ or $v_R + \theta v'_L = u_L$. Again, we have $v \in \mathcal{U}_{k_1}^0(w)$, proving $\mathsf{C}_3 \Longrightarrow \mathsf{C}_3^*$. Implications $\mathsf{C}_1 \Longrightarrow \mathsf{C}_1^*$ and $\mathsf{C}_2 \Longrightarrow \mathsf{C}_2^*$ are proved similarly. □

Since

$$\mathbf{Pr}[\mathsf{C}_0] = \mathbf{Pr}[Z^i = Z_\sigma^i \text{ for some } 1 \le i \le q] \le \frac{q}{N'},$$

we have

$$\mathbf{Pr}\left[\mathsf{Coll}\right] \le \mathbf{Pr}\left[\mathsf{C}_0\right] + \mathbf{Pr}\left[\mathsf{C}_1^*\right] + \mathbf{Pr}\left[\mathsf{C}_2^*\right] + \mathbf{Pr}\left[\mathsf{C}_3^*\right] \le \frac{q}{N'} + \mathbf{Pr}\left[\mathsf{C}_1^*\right] + \mathbf{Pr}\left[\mathsf{C}_2^*\right] + \mathbf{Pr}\left[\mathsf{C}_3^*\right]. \qquad (3)$$

We have the following proposition, which is easy to prove, but useful throughout the security analysis.

Proposition 3. *For fixed $u, v \in \{0,1\}^{2n}$, the probability that the i-th query-response cycle $\Delta^i = \big((X^i, K^i, Y^i), (X_\sigma^i, K^i, Y_\sigma^i)\big)$ such that $K_1^i = u_R$ generates an edge $u \xrightarrow{i} v$ is at most $2/(N')^2$.*

By Propositions 2, 3 and inequality (3), the only remaining problem for the security proof is to upper bound the size of the auxiliary subsets, which is equivalent to upper bounding the number of possible start/end nodes for a certain "collision-completing edge".

3.5 Bounding the start/end nodes of collision-completing edges

Proposition 4. *For an integer $a > 1$, let $\mathsf{Mult}(a)$ denote the event that a multiset $\{Z^i, Z^i_\sigma : 1 \leq i \leq q\}$ has an element of multiplicity $> a$. Then we have*

$$\mathbf{Pr}\left[\mathsf{Mult}(a)\right] \leq N\left(\frac{2eq}{(a+1)N'}\right)^{a+1}. \tag{4}$$

Furthermore, for any $k \geq 0$ and $w \in \{0,1\}^{2n}$, event $\neg\mathsf{Mult}(a)$ implies $|\mathcal{U}^0_k(w)| \leq a^k$.

Proof. Fix $z \in \{0,1\}^n$. The probability that z has multiplicity greater than a in the multiset $\{Z^i, Z^i_\sigma : 1 \leq i \leq q\}$ is at most

$$\binom{2q}{a+1}\left(\frac{1}{N'}\right)^{a+1} \leq \left(\frac{2eq}{(a+1)N'}\right)^{a+1}.$$

Inequality (4) follows since z is chosen from a set of size N.

If $u \in \mathcal{U}^0_k(w)$ for some w, then there would exist a path

$$\mathcal{P}: \quad u \overset{j_1}{\to} u[1] \overset{j_2}{\to} u[2] \overset{j_3}{\to} \cdots \overset{j_{k-1}}{\to} u[k-1] \overset{j_k}{\to} w.$$

Without the occurrence of $\mathsf{Mult}(a)$, we have at most a possible nodes for position $u[k-1]$. Given a node for position $u[k-1]$, we also have at most a possible nodes for position $u[k-2]$. In this way, we see that there are at most a^k possible nodes for position u. □

Proposition 5. *For $\alpha \in \{1,2\}$, $k \geq 0$ and $M > 0$, let $\mathsf{U}^\alpha_k(M)$ denote the event that $|\mathcal{U}^\alpha_k(z_1,z_2)| > M$ for some $z_1, z_2 \in \{0,1\}^n$. Then for $\alpha \in \{1,2\}$ and $d = 2m+1 > a > 0$, we have*

$$\mathbf{Pr}\left[\mathsf{U}^\alpha_k(d^k) \wedge \bigwedge_{\substack{\alpha=1,2\\0\leq j<k}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a)\right] \leq N^2\left(\frac{2eS_1(a,d,k)qN}{(m+1)(N')^2}\right)^{m+1}, \tag{5}$$

where

$$S_1(a,d,k) = \frac{d^k - a^k}{d-a} + 2\sum_{0\leq s<t<k} a^{k-t-1}d^{t-s-1}.$$

Proof. Since $\mathcal{U}^1_0(z_1,z_2) = \{z_1||z_2\}$ and $\mathcal{U}^2_0(z_1,z_2) = \{z_1/(1+\theta)||z_2\}$ for any z_1, $z_2 \in \{0,1\}^n$, we have

$$\mathbf{Pr}\left[\mathsf{U}^1_0(1)\right] = \mathbf{Pr}\left[\mathsf{U}^2_0(1)\right] = 0.$$

Therefore the proposition holds for $k = 0$. Now we will assume event

$$\bigwedge_{\substack{\alpha=1,2\\0\leq j<k}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a)$$

for $k \geq 1$ and estimate the probability of $\mathsf{U}^1_k(d^k)$ on this condition. The probability of $\mathsf{U}^2_k(d^k)$ can be analyzed similarly. Our estimation consists of the following three steps.

Step 1. We fix z_1, $z_2 \in \{0,1\}^n$, and then upper bound the probability, denoted p, that Δ^i increases $|\mathcal{U}_k^1(z_1,z_2)|$ (at least by one). First, we estimate the probability, denoted p_1, that Δ^i completes a path

$$\mathcal{P}: \quad u = u[0] \to u[1] \to u[2] \to \cdots \to u[k-1] \to u[k] = v$$

such that $u_L = z_1$ and $v_R = z_2$ by contributing a *single* edge in $\mathcal{P}$. Let $u[s] \xrightarrow{i} u[s+1]$ be the single edge. If we fix $z^* \in \{0,1\}^n$ as the value of v_L, then $u[s+1]$ should be one of the nodes in $\mathcal{U}_{k-s-1}^0(z^*||z_2)$. (Note that $v = z^*||z_2$.) On the other hand, $u[s]$ should be one of the nodes in $\mathcal{U}_s^1(z_1, K_1^i)$. Since $|\mathcal{U}_s^1(z_1,K_1^i)| \le d^s$ and $|\mathcal{U}_{k-s-1}^0(z^*||z_2)| \le a^{k-s-1}$, and by Proposition 3, we have

$$p_1 \le \sum_{0\le s<k} \left(N \cdot d^s \cdot a^{k-s-1} \cdot \frac{2}{(N')^2} \right) = \frac{2a^{k-1}N}{(N')^2} \sum_{s=0}^{k-1} \left(\frac{d}{a}\right)^s = \frac{2N}{(N')^2} \cdot \frac{d^k - a^k}{d-a}.$$

Next, we estimate the probability, denoted p_2, that Δ^i completes a path $\mathcal{P}$

$$\mathcal{P}: \quad u = u[0] \to u[1] \to u[2] \to \cdots \to u[k-1] \to u[k] = v$$

such that $u_L = z_1$ and $v_R = z_2$ by contributing at least two edges in $\mathcal{P}$ (assuming $k \ge 2$). Let $u[s] \xrightarrow{i} u[s+1]$ and $u[t] \xrightarrow{i} u[t+1]$ denote the last two i-labeled edges appearing in $\mathcal{P}$, where $0 \le s < t \le k-1$. We will upper bound the number of possible values for $u[t]$ and $u[t+1]$, and use Proposition 3.

If $z^* \in \{0,1\}^n$ is chosen as the value of v_L, then we have $u[t+1] \in \mathcal{U}_{k-t-1}^0(z^*||z_2)$. Suppose that we have chosen specific values u_L^* and u_R^* for $u[t+1]_L$ and $u[t+1]_R$, respectively. If the two i-labeled edges have the same sign, then we have $u[t] \in \mathcal{U}_{t-s-1}^1(u_L^*||K_1^i)$ since $u[s+1]_L = u_L^*$. Otherwise, we have $u[t] \in \mathcal{U}_{t-s-1}^2(u_R^*||K_1^i)$. Since $|\mathcal{U}_{k-t-1}^0(z^*||z_2)| \le a^{k-t-1}$ and $|\mathcal{U}_{t-s-1}^1(u_L^*||K_1^i)|$, $|\mathcal{U}_{t-s-1}^2(u_R^*||K_1^i)| \le d^{t-s-1}$, we have

$$p_2 \le \sum_{0\le s<t<k} \left(N \cdot a^{k-t-1} \cdot 2 \cdot d^{t-s-1} \cdot \frac{2}{(N')^2} \right) = \frac{4N}{(N')^2} \sum_{0\le s<t<k} a^{k-t-1} d^{t-s-1},$$

and hence

$$p = p_1 + p_2 \le \frac{2N}{(N')^2} \left(\frac{d^k - a^k}{d-a} + 2 \sum_{0\le s<t<k} a^{k-t-1} d^{t-s-1} \right) = \frac{2S_1(a,d,k)N}{(N')^2}.$$

Step 2. For any z_1, $z_2 \in \{0,1\}^n$, the number of cycles that increase $|\mathcal{U}_k^1(z_1,z_2)|$ is not greater than m except with probability

$$\begin{aligned} P &\le N^2 \binom{q}{m+1} p^{m+1} \le N^2 \binom{q}{m+1} \left(\frac{2S_1(a,d,k)N}{(N')^2} \right)^{m+1} \\ &\le N^2 \left(\frac{eq}{m+1} \right)^{m+1} \left(\frac{2S_1(a,d,k)N}{(N')^2} \right)^{m+1} \le N^2 \left(\frac{2eS_1(a,d,k)qN}{(m+1)(N')^2} \right)^{m+1}. \end{aligned} \tag{6}$$

Step 3. We estimate the maximum increment that a single cycle Δ^i can make in $|\mathcal{U}_k^1(z_1,z_2)|$ for fixed values z_1 and z_2. Any node that is newly added to $\mathcal{U}_k^1(z_1,z_2)$ by Δ^i is associated with a path $\mathcal{P}$ such that $|\mathcal{P}| = k$, $\mathsf{s}(\mathcal{P})_L = z_1$ and $\mathsf{e}(\mathcal{P})_R = z_2$. This path (not necessarily unique) consists of edges labeled $\le i$ and certainly contains at least one i-labeled edge. Among the end nodes added by Δ^i, the number of nodes whose corresponding path has $u[s] \to u[s+1]$ as the nearest i-labeled edge to the end node would be at most $2d^{k-s-1}$ since such nodes should be contained in either $\mathcal{U}^1_{k-s-1}(Z^i, z_2)$ or $\mathcal{U}^1_{k-s-1}(Z^i_\sigma, z_2)$. Therefore, the total increment made by Δ^i is bounded by

$$2(1+d+\cdots+d^{k-1}) = 2\left(\frac{d^k-1}{d-1}\right) = \frac{d^k-1}{m}.$$

From Step 2 and Step 3, we conclude that

$$|\mathcal{U}^1_k(z_1, z_2)| \le \frac{d^k-1}{m} \cdot m \le d^k,$$

except with probability P. By (6), we obtain inequality (5). □

Proposition 6. *For $k_2 \ge k_1$ such that $k = k_1 + k_2 > 0$ and $M > 0$, define the following events.*

- $\mathsf{V}^1_{k_1,k_2}(M) : |\mathcal{V}^1_{k_1,k_2}(z)| > M$ *for some* $z \in \{0,1\}^n$.
- $\mathsf{V}^2_{k_1,k_2}(M) : |\mathcal{V}^2_{k_1,k_2}(z_1, z_2)| > M$ *for some* $z_1, z_2 \in \{0,1\}^n$.

Then for $\alpha \in \{1,2\}$ and integers $d = 2m+1 > a > 0$, we have

$$\mathbf{Pr}\left[\mathsf{V}^1_{k_1,k_2}(d^{k_2}) \wedge \bigwedge_{\substack{j_1<k_1, j_2<k_2\\ j_1 \le j_2, j_1+j_2>0}} \neg\mathsf{V}^1_{j_1,j_2}(d^{j_2}) \wedge \bigwedge_{\substack{\alpha=1,2\\ 0\le j<k_2}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right] \le N\left(\frac{2eS_2(a,d,k_1,k_2)qN}{(m+1)(N')^2}\right)^{m+1}, \quad (7)$$

$$\mathbf{Pr}\left[\mathsf{V}^2_{k_1,k_2}(d^{k_2}) \wedge \bigwedge_{\substack{j_1<k_1, j_2<k_2\\ j_1 \le j_2, j_1+j_2>0}} \neg\mathsf{V}^1_{j_1,j_2}(d^{j_2}) \wedge \bigwedge_{\substack{\alpha=1,2\\ 0\le j<k_2}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right] \le N^2\left(\frac{2eS_2(a,d,k_1,k_2)qN}{(m+1)(N')^2}\right)^{m+1}, \quad (8)$$

where

$$S_2(a,d,k_1,k_2) = \left(k_1 a^{k_1-1} d^{k_2} + a^{k_1} \cdot \frac{d^{k_2}-a^{k_2}}{d-a}\right) + 2\sum_{0\le s<t<k_2} a^{k_2-t-1} d^{t-s-1} + 2\left(\frac{a^{k_1}-1}{a-1}\right)\left(\frac{d^{k_2}-1}{d-1}\right).$$

Proof. We begin with assuming event

$$\bigwedge_{\substack{j_1<k_1, j_2<k_2\\ j_1 \le j_2, j_1+j_2>0}} \neg\mathsf{V}^1_{j_1,j_2}(d^{j_2}) \wedge \bigwedge_{\substack{\alpha=1,2\\ 0\le j<k_2}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0.$$

Here we give a proof only for $\alpha = 2$, while the other case can be proved similarly. Our proof consists of the following three steps as the previous proposition.

Step 1. We fix $z_1, z_2 \in \{0,1\}^n$, and then upper bound the probability, denoted p, that Δ^i increases $|\mathcal{V}^2_{k_1,k_2}(z_1,z_2)|$ (at least by one). First, we estimate the probability, denoted p_1, that Δ^i completes a pair of irreducible colliding paths

$$\begin{aligned} \mathcal{P}_1: &\quad u = u[0] \to u[1] \to u[2] \to \cdots \to u[k_1-1] \to u[k_1] = w, \\ \mathcal{P}_2: &\quad v = v[0] \to v[1] \to v[2] \to \cdots \to v[k_2-1] \to v[k_2] = w \end{aligned} \tag{9}$$

such that $u_R + \theta v_L = z_1$ and $w_R = z_2$, by contributing a single edge in $\mathcal{P}_1 \cup \mathcal{P}_2$. Suppose that the single edge $u[s] \xrightarrow{i} u[s+1]$ is contained in $\mathcal{P}_1$. If we fix $z^* \in \{0,1\}^n$ as the value of $u[s]_L$, then $u[s] = z^*||K_1^i$ and u should be one of the nodes in $\mathcal{U}^0_s(z^*||K_1^i)$. Once u is fixed, we have $w \in \mathcal{U}^1_{k_2}((u_R + z_1)/\theta, z_2)$ and $u[s+1] \in \mathcal{U}^0_{k_1-s-1}(w)$. Therefore the number of possible values for $(u[s], u[s+1])$ is at most $N \cdot a^s \cdot d^{k_2} \cdot a^{k_1-s-1}$.

For the case where the single edge $v[t] \xrightarrow{i} v[t+1]$ is contained in $\mathcal{P}_2$, we fix $z^* \in \{0,1\}^n$ as the value of w_L. Then we can determine $w = z^*||z_2$. Once w is fixed, we have $v[t+1] \in \mathcal{U}^0_{k_2-t-1}(w)$ and $u \in \mathcal{U}^0_{k_1}(w)$. Each possible value of u determines $v_L = (u_R + z_1)/\theta$. Since $v[t] \in \mathcal{U}^1_t(v_L, K_1^i)$, the number of possible values for $(v[t], v[t+1])$ is at most $N \cdot a^{k_2-t-1} \cdot a^{k_1} \cdot d^t$. Therefore we have

$$\begin{aligned} p_1 &\le \sum_{0 \le s < k_1} \left(N \cdot a^s \cdot d^{k_2} \cdot a^{k_1-s-1} \cdot \frac{2}{(N')^2} \right) + \sum_{0 \le t < k_2} \left(N \cdot a^{k_2-t-1} \cdot a^{k_1} \cdot d^t \cdot \frac{2}{(N')^2} \right) \\ &= \frac{2N}{(N')^2} \left(k_1 a^{k_1-1} d^{k_2} + a^{k_1} \cdot \frac{d^{k_2} - a^{k_2}}{d-a} \right). \end{aligned}$$

Next, we estimate the probability, denoted p_2, that Δ^i completes a pair of irreducible colliding paths of form (9) by contributing either at least two edges in $\mathcal{P}_1$ or at least two edges in $\mathcal{P}_2$. The estimation of p_2 is similar to Proposition 5, so we have

$$p_2 \le \frac{4N}{(N')^2} \sum_{0 \le s < t < k_2} a^{k_2-t-1} d^{t-s-1}.$$

Finally, we estimate the probability, denoted p_3, that Δ^i completes a pair of irreducible colliding paths of form (9) by contributing exactly two edges, one from $\mathcal{P}_1$ and the other from $\mathcal{P}_2$. Let $u[s] \xrightarrow{i} u[s+1]$ in $\mathcal{P}_1$ and $v[t] \xrightarrow{i} v[t+1]$ in $\mathcal{P}_2$ denote the two edges, respectively. Without the occurrence of C_0, either $s+1 < k_1$ or $t+1 < k_2$. If $k_1 - s - 1 \le k_2 - t - 1$, then we upper bound the number of possible values for $u[s]$ and $u[s+1]$. Otherwise, we can apply the same argument to $v[t]$ and $v[t+1]$.

For the case where the two i-labeled edge have the same sign, fix $z^* \in \{0,1\}^n$ as the value of $u[s]_L$. Then $u[s]$ is determined by $u[s] = z^*||K_1^i$. Furthermore, we have $u[s+1] \in \mathcal{U}^0_{k_1-s-1}(w)$ for some $w \in \mathcal{V}^1_{k_1-s-1,k_2-t-1}(z_2)$ since $u[s+1]_L = v[t+1]_L$. Therefore the probability p_3^+ that two edges of the same sign, one from $\mathcal{P}_1$ and the other from $\mathcal{P}_2$, complete configuration (9) is at most

$$\begin{aligned} p_3^+ &\le \sum_{\substack{0 \le s < k_1 \\ 0 \le t < k_2 \\ k_1 - s \le k_2 - t}} \left(N \cdot a^{k_1-s-1} \cdot d^{k_2-t-1} \cdot \frac{2}{(N')^2} \right) \\ &\le \frac{2N}{(N')^2} \left(\sum_{0 \le s < k_1} a^{k_1-s-1} \right) \left(\sum_{0 \le t < k_2} d^{k_2-t-1} \right) = \frac{2N}{(N')^2} \left(\frac{a^{k_1} - 1}{a-1} \right) \left(\frac{d^{k_2} - 1}{d-1} \right). \end{aligned}$$

For the case where the two i-labeled edge have opposite signs, fix $z^* \in \{0,1\}^n$ as the value of w_L. Then we have $w = z^*||z_2$, $u[s+1] \in \mathcal{U}^0_{k_1-s-1}(w)$ and $v[t+1] \in \mathcal{U}^0_{k_2-t-1}(w)$. So we have

a^{k_1-s-1} possible candidates for $u[s+1]_L$, and a^{k_2-t-1} possible candidates for $v[t+1]_L$. Since the probability that either $(Z^i, Z^i_\sigma) = (u[s+1]_L, v[t+1]_L)$ or $(Z^i, Z^i_\sigma) = (v[t+1]_L, u[s+1]_L)$ is at most $2/(N')^2$ for each $(u[s+1]_L, v[t+1]_L)$, the probability p_3^- that two edges of opposite signs complete configuration (9) is estimated by

$$p_3^- \leq \frac{2N}{(N')^2}\left(\sum_{0\leq s<k_1} a^{k_1-s-1}\right)\left(\sum_{0\leq t<k_2} a^{k_2-t-1}\right) = \frac{2N}{(N')^2}\left(\frac{a^{k_1}-1}{a-1}\right)\left(\frac{a^{k_2}-1}{d-1}\right).$$

To summarize, we have

$$\begin{aligned} p &= p_1 + p_2 + p_3 = p_1 + p_2 + p_3^+ + p_3^- \\ &\leq \frac{2N}{(N')^2}\left(k_1 a^{k_1-1} d^{k_2} + a^{k_1}\cdot\frac{d^{k_2}-a^{k_2}}{d-a}\right) + \frac{4N}{(N')^2}\sum_{0\leq s<t<k_2} a^{k_2-t-1} d^{t-s-1} \\ &+ \frac{4N}{(N')^2}\left(\frac{a^{k_1}-1}{a-1}\right)\left(\frac{d^{k_2}-1}{d-1}\right) = \frac{2S_2(a,d,k_1,k_2)N}{(N')^2}. \end{aligned}$$

Step 2. For any $z_1, z_2 \in \{0,1\}^n$, the number of cycles that increase $|\mathcal{V}^1_{k_1,k_2}(z_1,z_2)|$ is not greater than m except with probability

$$\begin{aligned} P &\leq N^2\binom{q}{m+1}p^{m+1} \\ &\leq N^2\binom{q}{m+1}\left(\frac{2S_2(a,d,k_1,k_2)N}{(N')^2}\right)^{m+1} \leq N^2\left(\frac{2eS_2(a,d,k_1,k_2)qN}{(m+1)(N')^2}\right)^{m+1}. \end{aligned} \tag{10}$$

Step 3. With the same analysis as Proposition 5, the maximum increment that a single cycle $\varDelta^i$ can make in $|\mathcal{V}^1_{k_1,k_2}(z_1,z_2)|$ for fixed values z_1 and z_2 is upper bounded by

$$2\left(1 + d + \cdots + d^{k_2-1}\right) \leq 2\left(\frac{d^{k_2}-1}{d-1}\right) \leq \frac{d^{k_2}}{m}.$$

From Step 2 and Step 3, we conclude that $|\mathcal{V}^1_{k_1,k_2}(z_1,z_2)| \leq (d^{k_2}/m)\cdot m = d^{k_2}$ except with probability P. By (10), we obtain inequality (8). □

Proposition 7. *For $k_2 \geq k_1 \geq 0$ and $M > 0$, let $\mathsf{W}_{k_1,k_2}(M)$ denote the event that $|\mathcal{W}_{k_1,k_2}(z)| > M$ for some $z \in \{0,1\}^n$. For $2q/N \leq c < 1$ and integers $d = 2m+1 > a > 0$, we have*

$$\begin{aligned} \mathbf{Pr}\Bigg[\mathsf{W}_{k_1,k_2}(c^k N) \wedge \bigwedge_{\substack{j_1\leq k_1, j_2\leq k_2\\ j_1\leq j_2, j_1+j_2<k}} \neg\mathsf{W}_{j_1,j_2}(c^{j_1+j_2}N) \\ \wedge \bigwedge_{\substack{j_1<k_1, j_2<k_2\\ j_1\leq j_2, j_1+j_2>0}} \neg\mathsf{V}^1_{j_1,j_2}(d^{j_2}) \wedge \bigwedge_{\substack{\alpha=1,2\\ 0\leq j<k_2}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\Bigg] \\ \leq N\left(\frac{2ed^{k_2}S_3(a,k_1,k_2)qN}{mc(N')^2} + \frac{4ed^{k_2}S_4(a,d,k_1,k_2)q}{mc^k(N')^2}\right)^{\frac{mc^k N}{d^{k_2}}}, \end{aligned} \tag{11}$$

where $k = k_1 + k_2$*,*

$$S_3(a, k_1, k_2) = \left(\frac{a^{k_1} - 1}{a - 1} + \frac{a^{k_2} - 1}{a - 1}\right),$$
$$S_4(a, d, k_1, k_2) = \sum_{0 \le s < t < k_2} a^{k_2 - t - 1} d^{t - s - 1} + \left(\frac{a^{k_1} - 1}{a - 1}\right)\left(\frac{d^{k_2} - 1}{d - 1}\right).$$

Proof. It is easy to prove

$$\mathbf{Pr}\left[\mathsf{W}_{0,0}(N)\right] = \mathbf{Pr}\left[\mathsf{W}_{0,1}(2q)\right] = 0.$$

Therefore we assume event

$$\bigwedge_{\substack{j_1 \le k_1, j_2 \le k_2 \\ j_1 \le j_2, j_1 + j_2 < k}} \neg \mathsf{W}_{j_1, j_2}(c^{j_1 + j_2} N) \wedge \bigwedge_{\substack{j_1 < k_1, j_2 < k_2 \\ j_1 \le j_2, j_1 + j_2 > 0}} \neg \mathsf{V}^1_{j_1, j_2}(d^{j_2}) \wedge \bigwedge_{\substack{\alpha = 1,2 \\ 0 \le j < k_2}} \neg \mathsf{U}^\alpha_j(d^j) \wedge \neg \mathsf{Mult}(a) \wedge \neg \mathsf{C}_0,$$

for $k > 1$ and follow the argument used in the previous propositions.

Step 1. We fix $z \in \{0,1\}^n$, and then upper bound the probability, denoted p, that $\varDelta^i$ increases $|\mathcal{W}_{k_1,k_2}(z)|$ (at least by one). First, we estimate the probability, denoted p_1, that $\varDelta^i$ completes a pair of irreducible colliding paths

$$\begin{aligned} \mathcal{P}_1 : \quad & u = u[0] \to u[1] \to u[2] \to \cdots \to u[k_1 - 1] \to u[k_1] = w, \\ \mathcal{P}_2 : \quad & v = v[0] \to v[1] \to v[2] \to \cdots \to v[k_2 - 1] \to v[k_2] = w \end{aligned} \tag{12}$$

such that $w_R = z$, by contributing a single edge in $\mathcal{P}_1 \cup \mathcal{P}_2$. If the single edge $u[s] \xrightarrow{i} u[s+1]$ is contained in $\mathcal{P}_1$, then we have $u[s] \in \mathcal{W}_{0,s}(K_1^i)$ and $u[s+1] \in \mathcal{U}^0_{k_1 - s - 1}(w)$ for some $w \in \mathcal{W}_{k_1 - s - 1, k_2}(z)$. Applying a similar argument to the case where the single edge is contained in $\mathcal{P}_2$, we have

$$\begin{aligned} p_1 &\le \sum_{0 \le s < k_1} \left(c^s N \cdot c^{k - s - 1} N \cdot a^{k_1 - s - 1} \cdot \frac{2}{(N')^2}\right) + \sum_{0 \le t < k_2} \left(c^t N \cdot c^{k - t - 1} N \cdot a^{k_2 - t - 1} \cdot \frac{2}{(N')^2}\right) \\ &= \frac{2c^{k-1} N^2}{(N')^2} \left(\frac{a^{k_1} - 1}{a - 1} + \frac{a^{k_2} - 1}{a - 1}\right) = \frac{2 S_3(a, k_1, k_2) c^{k-1} N^2}{(N')^2}. \end{aligned}$$

Next, the probability that $\varDelta^i$ completes a pair of irreducible colliding paths of form (12) by contributing either at least two edges in $\mathcal{P}_1$ or at least two edges in $\mathcal{P}_2$ is estimated by

$$p_2 \le \frac{4N}{(N')^2} \sum_{0 \le s < t < k_2} a^{k_2 - t - 1} d^{t - s - 1}.$$

using the same analysis as Proposition 6. Also, the probability that $\varDelta^i$ completes a pair of irreducible colliding paths of form (12) by contributing exactly two edges, one from $\mathcal{P}_1$ and the other from $\mathcal{P}_2$ is estimated by

$$p_3 \le \frac{4N}{(N')^2} \left(\frac{a^{k_1} - 1}{a - 1}\right)\left(\frac{d^{k_2} - 1}{d - 1}\right).$$

To summarize, we have

$$p = p_1 + p_2 + p_3 \le \frac{2 S_3(a, k_1, k_2) c^{k-1} N^2}{(N')^2} + \frac{4 S_4(a, d, k_1, k_2) N}{(N')^2}.$$

Step 2. For any $z \in \{0,1\}^n$, the number of cycles that increase $|\mathcal{W}_{k_1,k_2}(z)|$ is not greater than $mc^k N/d^{k_2}$ except with probability

$$\begin{aligned}
P &\le N\binom{q}{mc^k N/d^{k_2}} p^{\frac{mc^k N}{d^{k_2}}} \\
&\le N\binom{q}{mc^k N/d^{k_2}} \left(\frac{2S_3(a,k_1,k_2)c^{k-1}N^2}{(N')^2} + \frac{4S_4(a,d,k_1,k_2)N}{(N')^2}\right)^{\frac{mc^k N}{d^{k_2}}} \\
&\le N\left(\frac{2ed^{k_2}S_3(a,k_1,k_2)qN}{mc(N')^2} + \frac{4ed^{k_2}S_4(a,d,k_1,k_2)q}{mc^k(N')^2}\right)^{\frac{mc^k N}{d^{k_2}}} .
\end{aligned} \tag{13}$$

Step 3. With the same analysis as Proposition 5, the maximum increment that a single cycle $\mathit{\Delta}^i$ can make in $|\mathcal{W}_{k_1,k_2}(z)|$ for a fixed z is upper bounded by

$$2\left(1 + d + \cdots + d^{k_2-1}\right) \le 2\left(\frac{d^{k_2}-1}{d-1}\right) \le \frac{d^{k_2}}{m}.$$

From Step 2 and Step 3, we conclude that $|\mathcal{W}_{k_1,k_2}(z)| \le (d^{k_2}/m)\cdot(mc^k N/d^{k_2}) = c^k N$ except with probability P. By (13), we obtain inequality (11). □

Proposition 8. *For $k_2 \ge k_1 \ge 0$ and $M > 0$, let $\mathsf{W}^*_{k_1,k_2}(M)$ denote the event that $|\mathcal{W}^*_{k_1,k_2}| > M$. For $2q/N \le c < 1$ and integers $d = 2m+1 > a > 0$, we have*

$$\begin{aligned}
\mathbf{Pr}\Bigg[\mathsf{W}^*_{k_1,k_2}\left(\frac{d^{k_2}M}{m}\right) &\wedge \bigwedge_{\substack{j_1\le k_1, j_2\le k_2 \\ j_1\le j_2, j_1+j_2<k}} \neg\mathsf{W}_{j_1,j_2}(c^{j_1+j_2}N) \wedge \bigwedge_{\substack{j_1< k_1, j_2< k_2 \\ j_1\le j_2, j_1+j_2>0}} \neg\mathsf{V}^1_{j_1,j_2}(d^{j_2}) \\
&\wedge \bigwedge_{\substack{\alpha=1,2 \\ 0\le j<k_2}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\Bigg] \le \frac{2qN^2}{M(N')^2}\left(S_3(a,k_1,k_2)c^{k-1}N + 2S_4(a,d,k_1,k_2)\right).
\end{aligned}$$

Proof. This proof uses Markov's inequality, so we begin with definition of certain random variables. For each $(z,i) \in \{0,1\}^n \times \{1,\dots,q\}$, we define a random variable $X_{z,i}$, where $X_{z,i}(\mathcal{Q}_\mathit{\Delta}) = 1$ if

1. there exist a pair of irreducible colliding paths
$$\begin{aligned}
\mathcal{P}_1: &\quad u[0] \xrightarrow{s_1} u[1] \xrightarrow{s_2} \cdots \xrightarrow{s_{k_1-1}} u[k_1-1] \xrightarrow{s_{k_1}} w, \\
\mathcal{P}_2: &\quad v[0] \xrightarrow{t_1} v[1] \xrightarrow{t_2} \cdots \xrightarrow{t_{k_2-1}} v[k_2-1] \xrightarrow{t_{k_2}} w,
\end{aligned} \tag{14}$$
such that $w_R = z$ and $\max\{s_1,\dots,s_{k_1},t_1,\dots,t_{k_2}\} = i$,
2. the following predicate is true:
$$\mathsf{D} = \bigwedge_{\substack{j_1\le k_1, j_2\le k_2 \\ j_1\le j_2, j_1+j_2<k}} \neg\mathsf{W}_{j_1,j_2}(c^{j_1+j_2}N) \wedge \bigwedge_{\substack{j_1< k_1, j_2< k_2 \\ j_1\le j_2, j_1+j_2>0}} \neg\mathsf{V}^1_{j_1,j_2}(d^{j_2}) \wedge \bigwedge_{\substack{\alpha=1,2 \\ 0\le j<k_2}} \neg\mathsf{U}^\alpha_j(d^j) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0.$$

We also define

$$X = \sum_{\substack{z\in\{0,1\}^n \\ 1\le i\le q}} X_{z,i}.$$

First, we upper bound the expected value $E(X)$. We begin with upper bounding $E(X_{z,i})$ for a fixed pair (z,i). In fact, $E(X_{z,i})$ is upper bounded by the probability that the i-th cycle $\mathit{\Delta}^i$

completes a pair of of irreducible colliding paths of form (14) such that $w_R = z$ under the condition D. The analysis of this probability is exactly the same as Proposition 7. Namely,

$$E(X_{z,i}) \leq \frac{2S_3(a,k_1,k_2)c^{k-1}N^2 + 4S_4(a,d,k_1,k_2)N}{(N')^2}.$$

Therefore, we have

$$E(X) = \sum_{\substack{z\in\{0,1\}^n \\ 1\leq i\leq q}} E(X_{z,i}) \leq Nq\left(\frac{2S_3(a,k_1,k_2)c^{k-1}N^2 + 4S_4(a,d,k_1,k_2)N}{(N')^2}\right).$$

By Markov's inequality, we have

$$\mathbf{Pr}\left[E(X) \geq M\right] \leq \frac{Nq}{M}\left(\frac{2S_3(a,k_1,k_2)c^{k-1}N^2 + 4S_4(a,d,k_1,k_2)N}{(N')^2}\right).$$

Since each (z,i) increases the size of $\mathsf{W}^*_{k_1,k_2}$ at most by

$$2(1+d+\cdots+d^{k_2-1}) \leq 2\left(\frac{d^{k_2}-1}{d-1}\right) \leq \frac{d^{k_2}}{m},$$

we have

$$\mathbf{Pr}\left[\mathsf{W}^*_{k_1,k_2}\left(\frac{d^{k_2}M}{m}\right) \wedge \mathsf{D}\right] \leq \frac{2qN^2}{M(N')^2}\left(S_3(a,k_1,k_2)c^{k-1}N + 2S_4(a,d,k_1,k_2)\right). \qquad \square$$

3.6 Putting the Pieces Together

Theorem 9. *Let $l \geq 2$, $d = 2m+1 > a > 0$, $M > 0$ and $2q/N \leq c < 1$. Then for any modified adversary $\mathcal{B}$ that records q query-response cycles,*

$$\begin{aligned}
\mathbf{Adv}^{\mathsf{coll}}_H(\mathcal{B}) \leq{} & \frac{2qN^2}{M(N')^2} \sum_{\substack{0\leq k_1\leq k_2 \\ k_1+k_2=l+1}} \left(S_3(a,k_1,k_2)c^l N + 2S_4(a,d,k_1,k_2)\right) \\
& + N \sum_{\substack{0\leq k_1\leq k_2 \\ k_1+k_2\leq l}} \left(\frac{2ed^{k_2}S_3(a,k_1,k_2)qN}{mc(N')^2} + \frac{4ed^{k_2}S_4(a,d,k_1,k_2)q}{mc^{k_1+k_2}(N')^2}\right)^{\frac{mc^{k_1+k_2}N}{d^{k_2}}} \\
& + (N+N^2) \sum_{\substack{0\leq k_1\leq k_2 \\ 0<k_1+k_2\leq l}} \left(\frac{2eS_2(a,d,k_1,k_2)qN}{(m+1)(N')^2}\right)^{m+1} + 2N^2 \sum_{0\leq k<l} \left(\frac{2eS_1(a,d,k)qN}{(m+1)(N')^2}\right)^{m+1} \\
& + N\left(\frac{2eq}{(a+1)N'}\right)^{a+1} + \frac{q}{N'} \\
& + q\left(c^l N + \frac{d^l-1}{d-1}\right)\left(N \sum_{0\leq k_1+k_2\leq l} a^{k_1}d^{k_2} + \frac{M}{m}\sum_{k_1=0}^{l+1} a^{k_1}d^{\max\{k_1,l+1-k_1\}}\right)\frac{2}{(N')^2} \\
& + q\left(c^l N + \frac{d^l-1}{d-1}\right)\left(\frac{a^l-1}{a-1}\right)\frac{2N}{(N')^2} + q\left(c^l N + \frac{d^l-1}{d-1}\right)\left(\sum_{\substack{0\leq k_1\leq k_2 \\ 0<k_1+k_2\leq l}} a^{k_1}d^{k_2}\right)\frac{4N}{(N')^2} \\
& := \epsilon(N,q,M,c,l,a,m).
\end{aligned}$$

Proof. Let

$$\begin{aligned}\mathsf{Ex} = \bigvee_{\substack{0\le k_1\le k_2\\ k_1+k_2=l+1}} \mathsf{W}^*_{k_1,k_2}\left(\frac{d^{k_2}M}{m}\right) \vee \bigvee_{\substack{0\le k_1\le k_2\\ k_1+k_2\le l}} \mathsf{W}_{k_1,k_2}(c^{k_1+k_2}N) \vee \bigvee_{\substack{0\le k_1\le k_2\\ 0<k_1+k_2\le l}} \mathsf{V}^1_{k_1,k_2}(d^{k_2})\\ \vee \bigvee_{\substack{0\le k_1\le k_2\\ 0<k_1+k_2\le l}} \mathsf{V}^2_{k_1,k_2}(d^{k_2}) \vee \bigvee_{\substack{\alpha=1,2\\ 0\le k<l}} \mathsf{U}^\alpha_k(d^k) \vee \mathsf{Mult}(a) \vee \mathsf{C}_0.\end{aligned}$$

Then we have

$$\begin{aligned}\mathbf{Pr}\left[\mathsf{Ex}\right] \le{}& \frac{2qN^2}{M(N')^2}\sum_{\substack{0\le k_1\le k_2\\ k_1+k_2=l+1}}\Big(S_3(a,k_1,k_2)c^{k_1+k_2-1}N + 2S_4(a,d,k_1,k_2)\Big)\\ &+ N\sum_{\substack{0\le k_1\le k_2\\ k_1+k_2\le l}}\left(\frac{8ed^{k_2}S_3(a,k_1,k_2)q}{mcN} + \frac{8ed^{k_2}S_4(a,d,k_1,k_2)q}{mc^{k_1+k_2}N'N}\right)^{\frac{mc^{k_1+k_2}N}{d^{k_2}}}\\ &+ (N+N^2)\sum_{\substack{0\le k_1\le k_2\\ 0<k_1+k_2\le l}}\left(\frac{4eS_2(a,d,k_1,k_2)q}{(m+1)N'}\right)^{m+1} + \sum_{\substack{\alpha=1,2\\ 0\le k<l}} N^2\left(\frac{4eS_1(a,d,k)q}{(m+1)N'}\right)^{m+1}\\ &+ N\left(\frac{2eq}{(a+1)N'}\right)^{a+1} + \frac{q}{N'} := \epsilon_0.\end{aligned}$$

Example 2 *Let $l = 2$. Then we decompose* Ex *as follows.*

$$\begin{aligned}&\mathsf{Ex} \Rightarrow \mathsf{C}_0 \vee \mathsf{Mult}(a) \vee \big(\mathsf{U}^1_1(d) \wedge \neg\mathsf{Mult}(a)\big) \vee \big(\mathsf{U}^2_1(d) \wedge \neg\mathsf{Mult}(a)\big)\\ &\quad\vee \left(\mathsf{V}^1_{0,1}(d) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right) \vee \left(\mathsf{V}^1_{0,2}(d^2) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right)\\ &\quad\vee \left(\mathsf{V}^1_{1,1}(d) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right) \vee \left(\mathsf{V}^2_{0,1}(d) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right)\\ &\quad\vee \left(\mathsf{V}^2_{0,2}(d^2) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right) \vee \left(\mathsf{V}^2_{1,1}(d) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right)\\ &\quad\vee \left(\mathsf{W}_{0,2}(c^2N) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right) \vee \left(\mathsf{W}_{1,1}(c^2N) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right)\\ &\quad\vee \left(\mathsf{W}^*_{0,3}\left(\frac{d^3M}{m}\right) \wedge \neg\mathsf{W}_{0,2}(c^2N) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right)\\ &\quad\vee \left(\mathsf{W}^*_{1,2}\left(\frac{d^2M}{m}\right) \wedge \neg\mathsf{W}_{1,1}(c^2N) \wedge \neg\mathsf{W}_{0,2}(c^2N) \wedge \bigwedge_{\alpha=1,2}\neg\mathsf{U}^\alpha_1(d) \wedge \neg\mathsf{Mult}(a) \wedge \neg\mathsf{C}_0\right),\end{aligned}$$

where we are omitting predicates $\mathsf{U}^1_0(1)$, $\mathsf{U}^2_0(1)$, $\mathsf{W}_{0,0}(N)$ *and* $\mathsf{W}_{0,1}(2q)$ *since*

$$\mathbf{Pr}\left[\mathsf{U}^1_0(1)\right] = \mathbf{Pr}\left[\mathsf{U}^2_0(1)\right] = \mathbf{Pr}\left[\mathsf{W}_{0,0}(N)\right] = \mathbf{Pr}\left[\mathsf{W}_{0,1}(2q)\right] = 0.$$

Now we can upper bound the probability of each term in the righthand side of the above implication by using Proposition 5, 6, 7 and 8.

Back to the proof, since $\mathsf{C}_0 \Longrightarrow \mathsf{Ex}$, we have

$$\mathbf{Pr}\left[\mathsf{Coll}\right] \le \mathbf{Pr}\left[\mathsf{C}_0 \vee \mathsf{C}_1^* \vee \mathsf{C}_2^* \vee \mathsf{C}_3^*\right] \le \mathbf{Pr}\left[\mathsf{Ex}\right] + \mathbf{Pr}\left[\mathsf{C}_1^* \wedge \neg\mathsf{Ex}\right] + \mathbf{Pr}\left[\mathsf{C}_2^* \wedge \neg\mathsf{Ex}\right] + \mathbf{Pr}\left[\mathsf{C}_3^* \wedge \neg\mathsf{Ex}\right]. \quad (15)$$

In order to upper bound $\mathbf{Pr}\left[\mathsf{C}_1^* \wedge \neg\mathsf{Ex}\right]$, we fix $1 \le i \le q$, and count the number of possible nodes u satisfying the first condition appearing in the description of event C_1^* in Proposition 2 and the number of possible nodes v satisfying the second condition. Without the occurrence of event Ex, they are upper bounded by

$$L_u = c^l N + \sum_{0 \le k < l} d^k = c^l N + \frac{d^l - 1}{d - 1},$$

and

$$\begin{aligned} L_v &= N \sum_{0 \le k_1 + k_2 \le l} a^{k_1} d^{k_2} + \sum_{k_1 = 0}^{l+1} \frac{a^{k_1} d^{\max\{k_1, l+1-k_1\}} M}{m} \\ &= N \sum_{0 \le k_1 + k_2 \le l} a^{k_1} d^{k_2} + \frac{M}{m} \sum_{k_1 = 0}^{l+1} a^{k_1} d^{\max\{k_1, l+1-k_1\}}, \end{aligned}$$

respectively. For a fixed pair of u and v, the probability that the i-th query-response cycle Δ^i generates an edge from u to v is at most $2/(N')^2$ by Proposition 3. Therefore we have

$$\begin{aligned} \mathbf{Pr}\left[\mathsf{C}_1^* \wedge \neg\mathsf{Ex}\right] &\le q \cdot L_u \cdot L_v \cdot \frac{2}{(N')^2} \\ &= q \left(c^l N + \frac{d^l - 1}{d - 1} \right) \left(N \sum_{0 \le k_1 + k_2 \le l} a^{k_1} d^{k_2} + \frac{M}{m} \sum_{k_1 = 0}^{l+1} a^{k_1} d^{\max\{k_1, l+1-k_1\}} \right) \frac{2}{(N')^2} \\ &:= \epsilon_1. \end{aligned}$$

Similarly, we can prove

$$\begin{aligned} \mathbf{Pr}\left[\mathsf{C}_2^* \wedge \neg\mathsf{Ex}\right] &\le q \left(c^l N + \sum_{0 \le k < l} d^k \right) \left(N \sum_{0 \le k < l} a^k \right) \frac{2}{(N')^2} \\ &= q \left(c^l N + \frac{d^l - 1}{d - 1} \right) \left(\frac{a^l - 1}{a - 1} \right) \frac{2N}{(N')^2} := \epsilon_2, \\ \mathbf{Pr}\left[\mathsf{C}_3^* \wedge \neg\mathsf{Ex}\right] &\le q \left(c^l N + \sum_{0 \le k < l} d^k \right) \left(2N \sum_{\substack{0 \le k_1 \le k_2 \\ 0 < k_1 + k_2 \le l}} a^{k_1} d^{k_2} \right) \frac{2}{(N')^2} \\ &= q \left(c^l N + \sum_{0 \le k < l} d^k \right) \left(\sum_{\substack{0 \le k_1 \le k_2 \\ 0 < k_1 + k_2 \le l}} a^{k_1} d^{k_2} \right) \frac{4N}{(N')^2} := \epsilon_3 \end{aligned}$$

Therefore, by inequality (15), we have

$$\mathbf{Adv}_H^{\mathsf{coll}}(\mathcal{B}) \le \epsilon_0 + \epsilon_1 + \epsilon_2 + \epsilon_3,$$

which completes the proof. □

Example 3 *Let $l = 3$, $d = 2m + 1 > a > 0$, $M > 0$ and $2q/N \leq c < 1$. Then by Theorem 9 and using the constants S_1, S_2, S_3, S_4 from Table 1, we have*

$$\begin{aligned}
\mathbf{Adv}_H^{\mathsf{coll}}(q) \leq{}& \frac{2qN^2}{M(N')^2}\left(\left(a^3+2a^2+4a+5\right)c^3N+4d^2+(4a+8)d+2a^2+6a+10\right) \\
&+N\left(\frac{2edqN}{mc(N')^2}\right)^{\frac{mcN}{d}}+N\left(\frac{2ed^2(a+1)qN}{mc(N')^2}+\frac{4ed^2q}{mc^2(N')^2}\right)^{\frac{mc^2N}{d^2}} \\
&+N\left(\frac{2ed^3(a^2+a+1)qN}{mc(N')^2}+\frac{4ed^3(d+a+1)q}{mc^3(N')^2}\right)^{\frac{mc^3N}{d^3}}+N\left(\frac{4edqN}{mc(N')^2}+\frac{4edq}{mc^2(N')^2}\right)^{\frac{mc^2N}{d}} \\
&+N\left(\frac{2ed^2(a+2)qN}{mc(N')^2}+\frac{4ed^2(d+2)q}{mc^3(N')^2}\right)^{\frac{mc^3N}{d^2}}+(N+N^2)\left(\frac{2eqN}{(m+1)(N')^2}\right)^{m+1} \\
&+2(N+N^2)\left(\frac{2e(d+a+2)qN}{(m+1)(N')^2}\right)^{m+1}+(N+N^2)\left(\frac{2e(d^2+(a+2)d+a^2+2a+2)qN}{(m+1)(N')^2}\right)^{m+1} \\
&+(N+N^2)\left(\frac{2e(d^2+(a+2)d+a^2+4)qN}{(m+1)(N')^2}\right)^{m+1} \\
&+2N^2\left(\frac{2eqN}{(m+1)(N')^2}\right)^{m+1}+2N^2\left(\frac{2e(d+a+2)qN}{(m+1)(N')^2}\right)^{m+1}+N\left(\frac{2eq}{(a+1)N'}\right)^{a+1}+\frac{q}{N'} \\
&+q\left(c^3N+d^2+d+1\right)\left(d^3+(a+1)d^2+(a^2+a+1)d+(a^3+a^2+a+1)\right)\frac{2N}{(N')^2} \\
&+q\left(c^3N+d^2+d+1\right)\left((a^4+1)d^4+(a^3+a)d^3+a^2d^2\right)\frac{2M}{m(N')^2} \\
&+q\left(c^3N+d^2+d+1\right)\left(a^2+a+1\right)\frac{2N}{(N')^2} \\
&+q\left(c^3N+d^2+d+1\right)\left(d^3+(a+1)d^2+(a+1)d\right)\frac{4N}{(N')^2}.
\end{aligned}$$

Let $N = 2^{128}$. Given the query complexity q, we can optimize the upper bound by carefully choosing the parameters a, m, M and c. As a result, we can obtain Table 2 comparing the collision resistance of MDC-2 and MJH.

3.7 Asymptotic Result

For $N = 2^n$ and $l \geq 2$, let $q = N^{\frac{l}{l+1}}/\log N$, $a = l$, $m = 2l + 1$, $d = 2m + 1 = 4l + 3$, $M = N$ and $c = 1/N^{\frac{1}{l+1}}$. By Theorem 9, we have

$$\mathbf{Adv}_H^{\mathsf{coll}}(\mathcal{B}) \leq \epsilon(N, q, M, c, l, a, d) = \epsilon(N, N^{\frac{l}{l+1}}/\log N, N, 1/N^{\frac{1}{l+1}}, l, l, 2l+1).$$

If we regard the parameter l as a constant, then each term of the above upper bound is analyzed as follows.

1.

$$\frac{2qN^2}{M(N')^2}\sum_{\substack{0\leq k_1\leq k_2\\ k_1+k_2=l+1}}\Big(S_3(a,k_1,k_2)c^lN+2S_4(a,d,k_1,k_2)\Big) \leq C_1c^lq+\frac{C_2q}{N}=\frac{C_1}{\log N}+\frac{C_2}{N^{\frac{1}{l+1}}\log N}.$$

Table 1. Constants S_1, S_2, S_3, S_4.

$S_1(a,d,0) = 0$	$S_3(a,0,0) = 0$	$S_4(a,d,0,0) = 0$
$S_1(a,d,1) = 1$	$S_3(a,0,1) = 1$	$S_4(a,d,0,1) = 0$
$S_1(a,d,2) = d + a + 2$	$S_3(a,0,2) = a + 1$	$S_4(a,d,0,2) = 1$
$S_2(a,d,0,1) = 1$	$S_3(a,1,1) = 2$	$S_4(a,d,1,1) = 1$
$S_2(a,d,0,2) = d + a + 2$	$S_3(a,0,3) = a^2 + a + 1$	$S_4(a,d,0,3) = d + a + 1$
$S_2(a,d,1,1) = d + a + 2$	$S_3(a,1,2) = a + 2$	$S_4(a,d,1,2) = d + 2$
$S_2(a,d,0,3) = d^2+(a+2)d+a^2+2a+2$	$S_3(a,0,4) = a^3 + a^2 + a + 1$	$S_4(a,d,0,4) = d^2+(a+1)d+a^2+a+1$
$S_2(a,d,1,2) = d^2 + (a+2)d + a^2 + 4$	$S_3(a,1,3) = a^2 + a + 2$	$S_4(a,d,1,3) = d^2 + 2d + a + 2$
	$S_3(a,2,2) = 2a + 2$	$S_4(a,d,2,2) = (a+1)d + a + 2$

Table 2. Best known upper bounds on $\mathbf{Adv}^{\mathsf{coll}}_{\text{MDC-2}}(q)$ and $\mathbf{Adv}^{\mathsf{coll}}_{\text{MJH}}(q)$ for $n = 128$. The last column contains the parameters used to obtain the upper bounds for MJH.

q	$\mathbf{Adv}^{\mathsf{coll}}_{\text{MDC-2}}(q) \leq$	$\mathbf{Adv}^{\mathsf{coll}}_{\text{MJH}}(q) \leq$	(a, m, M, c)
2^{64}	7.57×10^{-7}	2.94×10^{-12}	$(2, 5, 2^{123.30}, 2^{-38.70})$
$2^{74.91}$	$1/2$	7.35×10^{-8}	$(2, 6, 2^{123.40}, 2^{-37.64})$
$2^{81.85}$	1	$1/2$	$(2, 6, 2^{123.20}, 2^{-32.38})$

2.

$$N \sum_{\substack{0\le k_1\le k_2\\ k_1+k_2\le l}} \left(\frac{2ed^{k_2}S_3(a,k_1,k_2)qN}{mc(N')^2}+\frac{4ed^{k_2}S_4(a,d,k_1,k_2)q}{mc^{k_1+k_2}(N')^2}\right)^{\frac{mc^{k_1+k_2}N}{d^{k_2}}}$$

$$\le C_3N\left(\frac{C_4q}{cN}+\frac{C_5q}{c^lN^2}\right)^{C_6c^lN} \le C_3N\left(\frac{C_4}{\log N}+\frac{C_5}{N^{\frac{2}{l+1}}\log N}\right)^{C_6N^{\frac{1}{l+1}}}.$$

3.

$$(N+N^2)\sum_{\substack{0\le k_1\le k_2\\ 0<k_1+k_2\le l}}\left(\frac{2eS_2(a,d,k_1,k_2)qN}{(m+1)(N')^2}\right)^{m+1}+2N^2\sum_{0\le k<l}\left(\frac{2eS_1(a,d,k)qN}{(m+1)(N')^2}\right)^{m+1}$$

$$\le C_7N^2\left(\frac{C_8q}{N}\right)^{2l+2}\le\frac{C_9}{(\log N)^{2l+2}}.$$

4.

$$N\left(\frac{2eq}{(a+1)N'}\right)^{a+1}+\frac{q}{N'}\le\frac{C_{10}}{(\log N)^{l+1}}+\frac{C_{11}}{N^{\frac{1}{l+1}}\log N}.$$

5.

$$q\left(c^lN+\frac{d^l-1}{d-1}\right)\left(N\sum_{0\le k_1+k_2\le l}a^{k_1}d^{k_2}+\frac{M}{m}\sum_{k_1=0}^{l+1}a^{k_1}d^{\max\{k_1,l+1-k_1\}}\right)\frac{2}{(N')^2}$$

$$+\,q\left(c^lN+\frac{d^l-1}{d-1}\right)\left(\frac{a^l-1}{a-1}\right)\frac{2N}{(N')^2}+q\left(c^lN+\frac{d^l-1}{d-1}\right)\left(\sum_{\substack{0\le k_1\le k_2\\ 0<k_1+k_2\le l}}a^{k_1}d^{k_2}\right)\frac{4N}{(N')^2}$$

$$\le\frac{C_{12}}{\log N}+\frac{C_{13}}{N^{\frac{1}{l+1}}\log N}.$$

Here C_i, $i=1,\dots,13$, are certain constants represented as functions of l. It is easy to show that each term converges to zero as N goes to infinity. Given $\epsilon>0$, we can take the parameter l so that inequality $N^{1-\epsilon}<N^{\frac{l}{l+1}}/\log N$ holds for sufficiently large N. Therefore, we can conclude that the MJH hash function is asymptotically collision resistant up to $O(N^{1-\epsilon})$ query complexity for any $\epsilon>0$.

References

[1] J. Black, M. Cochran and T. Shrimpton. On the impossibility of highly-efficient blockcipher-based hash functions. Eurocrypt 2005, LNCS 3494, pp. 526–541, Springer, Heidelberg (2005).

[2] J. Black, P. Rogaway and T. Shrimpton. Black-box analysis of the block-cipher-based hash-function construction from PGV. Crypto 2002, LNCS 2442, pp. 320–325, Springer, Heidelberg (2002).

[3] A. Bogdanov, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw and Y. Seurin. Hash functions and RFID tags: mind the gap. CHES 2008, LNCS 5154, pp. 283–299, Springer, Heidelberg (2008).

[4] J. W. Bos, O. Özen and M. Stam: Efficient hashing using the AES instruction set. CHES 2011, LNCS 6917, pp. 507–522. Springer, Heidelberg (2011)

[5] A. Biryukov and D. Khovratovich. Related-key cryptanalysis of the full AES-192 and AES-256. Asiacrypt 2009, LNCS 5912, pp. 1–18, Springer, Heidelberg (2009).

[6] B. Brachtl, D. Coppersmith, M. Heyden, S. Matyas, C. Meyer, J. Oseas, S. Pilpel and M. Schilling. Data authentication using modification detection codes based on a public one-way encryption function. US Patent #4,908,861. Awarded March 13, 1990 (filed August 28, 1987).

[7] I. Damgård. A design principle for hash functions. Crypto 1989, LNCS 435, pp. 416–427, Springer, Heidelberg (1990).

[8] E. Fleischmann, M. Gorski and S. Lucks. On the security of Tandem-DM. FSE 2009, LNCS 5665, pp. 85–105, Springer, Heidelberg (2009).

[9] E. Fleischmann, M. Gorski and S. Lucks. Security of cyclic double block length hash functions. Cryptography and Coding 2009, LNCS 5921, pp. 153–175, Springer, Heidelberg (2009).

[10] M. Hattori, S. Hirose and S. Yoshida. Analysis of double block length hash functions. IMA 2003, LNCS 2898, pp. 290–302, Springer, Heidelberg (2003).

[11] S. Hirose. A security analysis of double-block-length hash functions with the rate 1. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E89-A, NO. 10, pp. 2575–2582 (2006).

[12] S. Hirose. Provably secure double-block-length hash functions in a black-box model. ICISC 2004, LNCS 3506, pp. 330–342, Springer, Heidelberg (2005).

[13] S. Hirose. Some plausible construction of double-block-length hash functions. FSE 2006, LNCS 4047, pp. 210–225, Springer, Heidelberg (2006).

[14] L. R. Knudsen, J. L. Massey and B. Preneel. Attacks on fast double block length hash functions. Journal of Cryptology, Vol. 11, NO. 1, pp. 59–72 (1998).

[15] L. R. Knudsen, F. Mendel, C. Rechberger and S. S. Thomsen. Cryptanalysis of MDC-2. Eurocrypt 2009, LNCS 5479, pp. 106–120, Springer, Heidelberg (2009).

[16] X. Lai and J. L. Massey. Hash function based on block ciphers. Eurocrypt 1992, LNCS 658, pp. 55–70, Springer, Heidelberg (1993).

[17] J. Lee and D. Hong. Collision Resistance of the JH Hash Function. IEEE Transactions on Information Theory 58(3), pp. 1992–1995 (2012).

[18] J. Lee and D. Kwon: The security of Abreast-DM in the ideal cipher model. IEICE Transactions 94-A(1), pp. 104–109 (2011).

[19] J. Lee and M. Stam. MJH: A Faster Alternative to MDC-2. CT-RSA 2011, LNCS 6558, pp. 213–236, Springer, Heidelberg (2011).

[20] J. Lee and J. Steinberger. Multi-property-preserving domain extension using polynomial-based modes of operation. Eurocrypt 2010, LNCS 6110, pp. 573–596, Springer, Heidelberg (2010).

[21] J. Lee, M. Stam and J. Steinberger. The collision security of Tandem-DM in the ideal cipher model. Crypto 2011, LNCS 6841, pp. 561–577, Springer, Heidelberg (2011)

[22] S. Lucks. A collision-resistant rate-1 double-block-length hash function. Symmetric Cryptography, Dagstuhl Seminar Proceedings 07021 (2007).

[23] R. Merkle. One way hash functions and DES. Crypto 1989, LNCS 435, pp. 428–446, Springer, Heidelberg (1990).

[24] C. Meyer and M. Schilling. Chargement securise d'un programma avec code de detection de manipulation (1987).

[25] O. Özen and M. Stam. Another glance at double-length hashing. Cryptography and Coding 2009, LNCS 5921, pp. 176–201, Springer, Heidelberg (2009).

[26] B. Preneel, R. Govaerts and J. Vandewalle. Hash functions based on block ciphers: A synthetic approach. Crypto 1993, LNCS 773, pp. 368–378, Springer, Heidelberg (1994).

[27] T. Ristenpart and T. Shrimpton. How to build a hash function from any collision-resistant function. Asiacrypt 2007, LNCS 4833, pp. 147–163, Springer, Heidelberg (2007).

[28] P. Rogaway and J. Steinberger. Constructing cryptographic hash functions from fixed-key blockciphers. Crypto 2008, LNCS 5157, pp. 433–450, Springer, Heidelberg (2008).

[29] P. Rogaway and J. Steinberger. Security/efficiency tradeoffs for permuation-based hashing. Eurocrypt 2008, LNCS 4965, pp. 220–236, Springer, Heidelberg (2008).

[30] T. Shrimpton and M. Stam. Building a collision-resistant function from non-compressing primitives. ICALP 2008, LNCS 5126, pp. 643–654, Springer, Heidelberg (2008).

[31] M. Stam. Beyond uniformity: Security/efficiency tradeoffs for compression functions. Crypto 2008, LNCS 5157, pp. 397–412, Springer, Heidelberg (2008).

[32] M. Stam. Blockcipher based hashing revisited. FSE 2009, LNCS 5665, pp. 67–83, Springer, Heidelberg (2009).

[33] J. Steinberger. The collision intractability of MDC-2 in the ideal-cipher model. Eurocrypt 2007, LNCS 4515, pp. 34–51, Springer, Heidelberg (2008).

[34] H. Wu. The Hash Function JH. Submission to NIST (2008).
http://www3.ntu.edu.sg/home/wuhj/research/jh/index.html