One-Key Compression Function Based MAC with BBB Security

Avijit Dutta, Mridul Nandi, Goutam Paul

Indian Statistical Institute, Kolkata 700 108, India.
avirocks.dutta13@gmail.com, mridul.nandi@gmail.com,
goutam.paul@isical.ac.in

Abstract. Gaži et al. [CRYPTO 2014] analyzed the NI-MAC construction proposed by An and Bellare [CRYPTO 1999] and gave a tight birthday-bound of $O(lq^2/2^n)$, as an improvement over the previous bound of $O(l^2q^2/2^n)$. In this paper, we design a simple extension of NI-MAC, called NI$^+$-MAC, and prove that it has $O(q^2l^4/2^{2n})$ security bound. Our construction not only lifts the security of NI-MAC beyond birthday, it also reduces the number of keys from 2 (NI uses 2 independent keys) to 1. Before this work, Yasuda had proposed [FSE 2008] a single fixed-keyed compression function based BBB-secure MAC that uses an extra tweak. However, our proposed construction NI$^+$ does not require any extra tweak and thereby has reduced the state size compared to Yasuda's proposal [FSE 2008]. Further, the security proof of Yasuda's construction is straight-forward, as tweakable functions are replaced by uniform independent random functions. On the other hand, our proof technique is completely different and uses the structure graph based analysis introduced by Bellare et al. [CRYPTO 2005].

Keywords: Beyond Birthday, MAC, NI, Structure-Graph.

1 Introduction

In symmetric key paradigm, MAC (Message Authentication Code) is used for preserving message integrity and message origin authentication. The design of a MAC should not only consider achieving security, but also target attaining efficiency. In the literature, three different approaches of designing a MAC exists: (a) universal hash function based MAC, a popular example of which is UMAC [8], (b) a compression function based MAC, like NMAC [2], HMAC [2], NI [1] etc. (c) Block cipher based MAC, such as CBC MAC [4], PMAC [9], OMAC [17]. etc.

Most of the popular MACs are block cipher based MACs, but each one of them suffers from the same problem - security is guaranteed up to the *birthday bound*. When the block length of the underlying block cipher is 128-bit, then birthday bound does not seem to be a problem, as we are guaranteed to have 64 bits of security which is well acceptable for many practical applications. But when we deal with 64-bit block cipher (e.g HIGHT [16], PRESENT [10]) as used in many light weight crypto devices (e.g RFID, smartcard) then birthday bound problem becomes the main bottleneck.

Related Work on NMAC and HMAC. NMAC and its variant HMAC [2] is the first re-keying compression function based MAC where a key is appended to a message and then the appended message is hashed using Merkle-Damgård technique. It has been standardized in [23]. and has become popular and widely used in many network protocols like SSH, IPSec, TLS etc. Bellare et al. in [2] proves that NMAC is a secure PRF based on the assumption (i) $\mathtt{f}$ is a secure PRF and (ii) $\mathtt{Casc}^{\mathtt{f}}$ is a WCR (weakly collision resistant). HMAC when instantiated with MD4 or SHA-1, both of them play the role of $\mathtt{Casc}^{\mathtt{f}}$ then they have been found not to satisfy the WCR property [35, 36] and hence the security of HMAC [2] stands void. To restore the PRF security of NMAC, Bellare in [6] investigates the proof and drops assumption (ii). Koblitz and Menezes in [22] criticizes the way [6] discusses the practical implication of their result against uniform and non-uniform reductions used in the proof.

Dodis et.al in [12] investigates the indifferentiable property of HMAC from a keyed random oracle. In a recent line of researches, generic attack against iterated hash based MAC are being investigated [30, 31, 29, 24]. More recently, Gaži et. al in [14] showed a tight bound on NMAC. There is also a recent result [15] on the generic security analysis of NMAC and HMAC with input whitening.

Yausda in [38] had proposed a novel way of iterating a compression function dedicated for the use of MAC which is more efficient than standard HMAC to process data much faster. In [40] Yasuda has showed that classical sandwiched construction with Merkle-Damgård iteration based hashing provides a secure MAC which is an alternative for HMAC, useful in situation where the message size is small and high performance is required. A new secret-prefix MAC based on hash functions is presented in [43] which is similar to HMAC but does not require the second key.

U.Maurer et. al in [26] has presented a MAC construction namely PDI, that tranforms any FIL MAC to AIL MAC and investigated the tradeoff between the efficiency of MAC and the tightness of its security reduction. In [27] construction of AIL MAC from a FIL MAC with a single key was presented which is better than NI [1].

Related Work on Beyond birthday Secure MAC.
Block Cipher Based BBB MAC. In recent researches, many MAC constructions have been proposed with security beyond the birthday barrier without degrading the performance. The first attempt was made in ISO 9797-1 [3] without security proof. But Algorithm 4 of ISO 9797-1 was attacked by Joux et al. [20] that falsified the security bound. Algorithm 6 of ISO 9797-1 was proven to be secure against $O(2^{2n/3})$ queries with restrictions on the message length [44]. In [44] Yasuda also presented SUM-ECBC, a 4-key rate-1/2 construction with beyond birthday bound security. In 2011, Yasuda improved the number of keys and rate over SUM-ECBC and proposed a 3-key rate-1 PMAC_Plus construction [45] with beyond birthday security. In 2012, Zhang et al. [48] proposed a 3key version of f9 MAC (3kf9) that achieves BBB security.

There is also another deterministic MAC mode provides security beyond the birthday bound. Given an n-bit to n-bit fixed-key blockcipher with MAC security ϵ against q queries, Dodis et al. [13] have designed a variable-length MAC achieving $O(\epsilon q \text{poly}(n))$ MAC security. However, this design requires even longer keys and more block cipher invocations. By parity method, Bellare et al. present MACRX [3] with BBB security, conditioned on the input parameters are random and distinct. In [18], Jaulmes et al. proposed a randomized MAC that provides BBB security based on the ideal model (or possibly based on tweakable block cipher). Another BBB secure randomized construction called generic enhanced hash then MAC has been proposed in [28] by Minematsu. Recently Datta et al. in [11] unify PMAC_Plus and 3kf9 in one key setting with beyond birthday security.

Compression Function based BBB MAC. Besides the block cipher based BBB MAC constructions, Yasuda in [39] proposed a compression function based MAC construction - Multi-lane HMAC, that acheievs BBB security. In [42] Yasuda presented a double pipe mode operation (Lucks Construction [25]) for constructing AIL MAC from a FIL MAC that acheives BBB security. This work is further extended to provide full security in [46]. In [41] Yasuda has proposed a fixed single keyed compression function based cascaded MAC in a tweakable setting where the tweaks are some distinct masking keys of b bits. Thus for a l blocks message, one needs to compute l many different masks where the masks are generated from a single mask Δ_0 using the field multiplication. The security of the scheme has been proved to be $O(lq^2/2^{2n})$. Further improvement on [41] is followed in [47].

Related Work on fixed-key MAC. An et al.in [1] proposed a fixed-keyed compression function based MAC called NI-MAC. The construction of NI-MAC is similar to that of NMAC [2], the only difference is that NI-MAC uses two independent keyed compression functions f_1, f_2. The motivation of designing NI was to avoid constant re-keying on multi-block messages in NMAC and to allow for a security proof starting by the standard switch from a PRF to a random function, followed by information-theoretic analysis.

We mention here that the security proof technique for re-keying compression function based MAC is completely different from that of fixed-keyed compression function based MAC. The security of the former scheme is proved using reduction argument, whereas that of the latter is proved by replacing the fixed-keyed compression function with a random function.

Gaži et al.in [14] revisited the proof of NI-MAC and gave a tight birthday bound of $O(\frac{lq^2}{2^n})$, a better bound than earlier $O(\frac{l^2q^2}{2^n})$.

Our Contributions. The main disadvantage of the scheme of [41] is that one needs to store the masking key Δ_0. Thus, from the hardware point of view, it is infeasible to use the scheme in low-buffer and light-weight crypto devices, which are the basic target for achieving BBB security of a cryptographic scheme. Moreover, the proof technique of [41] is straight forward, as tweakable keyed functions are replaced by independent uniform random functions. In this paper, we propose NI^+, a non-tweakable single-keyed, rate-$b/(b+n)$, compression function based MAC that achieves beyond-birthday security, where b is the block length and n is the number of output bits. Moreover, our scheme is better than [41] in terms of required state size. Since our scheme does not use any extra tweak, our security proof technique is completely different than [41]. For our proof, we use the structure graph analysis of [5] and consider more bad events.

We mention here that NI^+ is an extension of NI-MAC, and it not only lifts the security of NI beyond birthday (Sect. 4), but also reduces the number of required keys from two (NI uses two independent keys) to one. In this context, we have shown that keeping the original structure of NI-MAC, with Θ being the sum of all intermediate chaining variables and Σ being the last block output (Σ, Θ defined in Sect. 3), cannot achieve BBB security.

In the following table we compare the different parameters along with their security bound of known BBB secure MACs. We write BC to denote block cipher based MAC, CF_{rk} denotes re-keying compression function based MAC (e.g HMAC), CF_{fk} denotes fixed-keyed compression function based MAC (e.g NI), Rate $\triangleq \frac{b}{rs}$, where b-size of message block, s-total input size of the function without the key part and r is the total number of function calls to process a single message block.

Construction	Type	# Keys	Rate	Security Bound	State size (#bits)
SUM-ECBC [44]	BC	4	1/2	$O(l^3q^3/2^{2n})$	$2n$
PMAC_Plus [45]	BC	3	1	$O(l^3q^3/2^{2n})$	$4n$
3kf9 [48]	BC	3	1	$O(l^3q^3/2^{2n})$	$2n$
1kf9 [11]	BC	1	1	$O(q^3l^4/2^{2n})$	$2n$
1k_PMAC+ [11]	BC	1	1	$O(q^3l^4/2^{2n})$	$4n$
L-Lane ($L = 2$) HMAC [39]	CF_{rk}	3	1/2	$O(q^2/2^{2n})$	$2n$
1-pass mode [41]	CF_{fk}	1	1	$O(lq^2/2^{2n})$	$(2b + 2n)$
NI^+[This paper]	CF_{fk}	1	$b/(b + n)$	$O(q^2l^4/2^{2n})$	$(b + 2n)$

2 Preliminaries

In this section, we briefly discuss the notations and definitions used in this paper. We also state some existing basic results.

2.1 Notation and Definitions

We denote $|S|$ as the cardinality of set S. Let $x \xleftarrow{\$} S$ denote that x is chosen uniformly at random from S. $[n]$ denotes the set of integers $\{1, 2, \ldots, n\}$. $(s)_{|n}$ denotes the last n bit substring of b bit string s.

Let M be a binary string over $\{0, 1\}$. Length of M in bits is denoted by $|M|$. When $|M| \mod b \neq 0$, we pad 10^d to M to make $|M| \mod b = 0$ where $d = n - 1 - |M| \mod b$ and b denotes the block length of M. $M_1||M_2|| \ldots ||M_l$ denotes the partition of message M after M is being padded, where each $M_i \in \{0, 1\}^b$ and l denotes the number of blocks of M. ℓ denotes the maximum number of blocks in a message. By a q-set or a q-tuple $x := (x_i : i \in I)$ for an index set I, we mean a set or a tuple of size q. When all elements $x_i's$ are distinct we write $x \in \mathsf{dist}_q$.

RANDOM FUNCTIONS. Let $Func(A, B)$ denote the set of all functions from A to B. A **random function** $\mathtt{F}$ is a function which is chosen from $Func(A, B)$ following some distribution, not necessarily uniform. In particular, a function ρ_n is said to be a uniform random function, if ρ_n is chosen uniformly at random from the set of all functions from a specified finite domain $\mathcal{D}$ to $\{0,1\}^n$. Throughout the paper we fix a positive integer n.

We will specify a uniform random function by performing *lazy sampling*. In lazy sampling, initially the function ρ is undefined at every point of its domain. We maintain a set $\mathrm{Dom}(\rho)$ that grows dynamically to keep the record of already defined domain points of ρ. $\mathrm{Dom}(\rho)$ is initialized to be empty. If $x \notin \mathrm{Dom}(\rho)$ then we will choose $y \xleftarrow{\$} \{0,1\}^n$ and add x in $\mathrm{Dom}(\rho)$. In this regard, x is said to be *fresh*. On the other hand, if $x \in \mathrm{Dom}(\rho)$ (i.e $x = x'$) then $y \leftarrow f(x')$. In this regard x is said to be *covered*.

2.2 Security Definitions

We consider that an adversary $\mathcal{A}$ is an oracle algorithm with access to its oracle $\mathcal{O}(\cdot)$ and outputs either 1 or 0. Accordingly, we write $\mathcal{A}^{\mathcal{O}(\cdot)} = 1$ or 0. The resource of $\mathcal{A}$ is measured in terms of the time complexity t which takes into account the time it takes to interacts with its oracle $\mathcal{O}(\cdot)$ and the time for its internal computations, query complexity q takes into account the number of queries asked to the oracle by the adversary, data complexity ℓ takes into account the maximum number of blocks in each query.

PSEUDO-RANDOM FUNCTION. We define **distinguishing advantage** of an oracle algorithm $\mathcal{A}$ for distinguishing two random functions $\mathtt{F}$ from $\mathtt{G}$ as

$$\mathbf{Adv}_{\mathcal{A}}(\mathtt{F}\ ;\ \mathtt{G}) := \Pr[\mathcal{A}^{\mathtt{F}} = 1] - \Pr[\mathcal{A}^{\mathtt{G}} = 1].$$

We define prf-advantage of $\mathcal{A}$ for an n-bit construction $\mathtt{F}$ by

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathtt{F}}(\mathcal{A}) := \mathbf{Adv}_{\mathcal{A}}(\mathtt{F}\ ;\ \rho_n).$$

We call $\mathcal{A}$ a (q, ℓ, t)-distinguisher if it makes at most q queries with at most ℓ-blocks in each query and runs in time at most t. We write $\mathbf{Adv}^{\mathrm{prf}}_{\mathtt{F}}(q, \ell, t) = \max_{\mathcal{A}} \mathbf{Adv}^{\mathrm{prf}}_{\mathtt{F}}(\mathcal{A})$ where maximum is taken over all (q, ℓ, t)-distinguisher $\mathcal{A}$. In an information theoretic situation we also ignore the time parameter t. We call a keyed construction $\mathtt{F}$ is (q, ℓ, ϵ)-prf if $\mathbf{Adv}^{\mathrm{prf}}_{\mathtt{F}}(q, \ell) \leq \epsilon$. Informally, if ϵ is negligible then $\mathtt{F}$ is said to be a secure PRF.

COLLISION-FREE AND COVER-FREE. Now we define some other information-theoretic security advantages (in which there is no presence of an adversary). Let $\mathtt{H}$ be a random function which outputs two n bit blocks, denoted by $(\Sigma, \Theta) \in (\{0,1\}^n)^2$. For a q-tuple of distinct messages $\mathcal{M} = (M^1, \ldots, M^q)$, we write $\mathtt{H}(M^i) = (\Sigma^i, \Theta^i)$. For a q-tuple of pairs $(\Sigma^i, \Theta^i)_i$, we say that

1. A tuple $(\Sigma^i, \Theta^i)_i$ is **collided** if $\exists i, j \in [q]$ such that $\Sigma^i = \Sigma^j$ and $\Theta^i = \Theta^j$ for some $j \neq i$. Otherwise the tuple is said to be **collision-free**.
2. A tuple $(\Sigma^i, \Theta^i)_i$ is **covered** if $\exists i, j \in [q]$ such that $\Sigma^i = (M^j_\alpha)_n$ and $\Theta^i = Y^j_{\alpha-1}$ where $\alpha \in [l_i]$ or $\alpha \in [l_j]$ and j could be equal to i, M^j_α denotes the

α^{th} block of j^{th} message M^j and $Y^j_{\alpha-1} \in \{0,1\}^n$. Otherwise the tuple is said to be **cover-free**.

Definition 1. *We define* (q,ℓ)-collision advantage *and* (q,ℓ)-cover-free advantage *as*

$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{coll}}(q,\ell) = \max_{M \in \mathit{dist}_q} Pr[(\Sigma_i, \Theta_i)_i \text{ is not collision-free}].$$
$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{cf}}(q,\ell) = \max_{M \in \mathit{dist}_q} Pr[(\Sigma_i, \Theta_i)_i \text{ is not cover-free}].$$

Clearly, $\mathbf{Adv}_{\mathsf{F}}^{\mathrm{coll}}(q,\ell) \le \frac{q^2}{2}\mathbf{Adv}_{\mathsf{F}}^{\mathrm{coll}}(2,\ell)$. Similarly, $\mathbf{Adv}_{\mathsf{F}}^{\mathrm{cf}}(q,\ell) \le \frac{q^2}{2}\mathbf{Adv}_{\mathsf{F}}^{\mathrm{cf}}(2,\ell)$. So it would be sufficient to concentrate on a pair of messages while bounding collision free or cover-free advantages. We say that a construction F is (q,ℓ,ϵ)-xxx if $\mathbf{Adv}_{\mathsf{F}}^{\mathrm{xxx}}(q,\ell) \le \epsilon$ where xxx denotes either **collision-free** or **cover-free**.

2.3 Structure Graphs

In this section, we briefly revisit the structure graph analysis [5, 14].

Consider a cascaded construction with a function f, where f is a uniform random function, that works on a message $M = M_1||M_2||\ldots||M_l$ of length l blocks as follows:

$$Y_0 = \mathbf{0}, \text{ and } Y_i = f(Y_{i-1}, M_i) \text{ for } i = 1, \ldots, l.$$

Informally, for a set of any two fixed distinct messages $\mathcal{M} = \{M^1, M^2\}$ and a uniformly chosen random function f, we construct the structure graph $\mathcal{G}^f(\mathcal{M})$ with $\{0,1\}^n$ as the set of nodes as follows. We follow the computations for M^1 followed by those of M^2 by creating nodes labelled by the values y_i of the intermediate chaining variables Y_i with the edge (y_i, y_{i+1}) labelled by the block M_{i+1}. In this process, if we arrive at a vertex already labelled, while not following an existing edge, we call this event an f-collision. The sequence of alternating vertices and edges corresponding to the computations for a message M^j is called an M^j-walk, denoted by W_j. A more formal discussion on structure graph appears in Appendix A.

Let $\mathcal{G}(\mathcal{M})$ denote the set of all structure graphs corresponding to the set of messages $\mathcal{M}$ (by varying f over a function family). For a fixed graph $G \in \mathcal{G}(\mathcal{M})$, let $fColl(G)$ denote the set of all f-collisions in G. We state the following known results.

Proposition 1. *[14, Lemma 2] For a fixed graph G,* $\Pr_f[\mathcal{G}^f(\mathcal{M}) = G] \le 2^{-n|fColl(G)|}$.

Proposition 2. *[14, Lemma 3]* $\Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : |fColl(G)| \ge 2] \le \frac{4\tau^4}{2^{2n}}$*, where τ is the total number of blocks of the messages in $\mathcal{M}$.*

It is to be noted that for CBC-MAC analysis [5], $f(\alpha,\beta)$ is taken as $\pi(\alpha \oplus \beta)$ and for the NI-MAC analysis [14], $f(\alpha,\beta)$ is taken as $\rho(\alpha||\beta)$, where π is a random permutation over n bits and ρ is a random function from $b+n$ bits to n bits, where b is the message block-length and n is the length of the chaining variable as well as the tag.

3 Proposed Construction of NI^+ for Beyond-Birthday Secure MAC

We present the schematic diagram of NI^+ in Fig. 3.1 followed by the description in Algorithm 1. Let $f_{K_1} : \{0,1\}^{b+n} \to \{0,1\}^n$ be a keyed function from $b+n$ bits

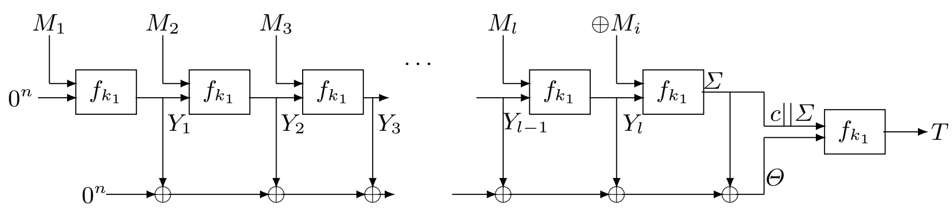

Fig. 3.1: Construction of NI^+ MAC

Input: $f_{K_1} : K_1 \xleftarrow{\$} \mathcal{K}$, $M \leftarrow \{0,1\}^*$, $c \leftarrow 10^{b-n-1}$
Output: $T \in \{0,1\}^n$
$M_1||M_2||\ldots M_l \leftarrow M||10^*$; //$l$ is the number of message blocks in M
$Z \leftarrow 0^n$; $Y \leftarrow 0^n$;
for $i = 1$ *to* l **do**
  $Y \leftarrow f_{K_1}(M_i, Y)$; $Z \leftarrow Z \oplus Y$;
end
$CS \leftarrow \oplus_{i=1}^{l} M_i$;
$Y \leftarrow f_{K_1}(CS, Y)$; $Z \leftarrow Z \oplus Y$;
$\Sigma \leftarrow Y$; $\Theta \leftarrow Z$;
$T \leftarrow f_{K_1}(c||\Sigma, \Theta)$;
Return T;

Algorithm 1: Algorithm for NI^+ MAC

to n bits where $b > n$. Recall that b refers to the block length of a message block and n refers to the output length in bits. Let $M \in \{0,1\}^{bl}$. So we can write $M = (M_1, M_2, \ldots, M_l)$ where each $M_i \in \{0,1\}^b$. We define a checksum block $CS = \oplus_{i=1}^{l} M_i$. We denote $\mathbf{Casc}^{\mathbf{f_{K_1}}}(M) := f_{K_1}(\ldots(f_{K_1}(f_{K_1}(0, M_1), M_2), \ldots, M_l)$. Output of $\mathbf{Casc}^{\mathbf{f_{K_1}}}(M)$ and the checksum block CS is passed through the same function f_{K_1} and the output is denoted as Σ. We obtain Θ by xoring all the intermediate chaining values (i.e $\oplus_{i=1}^{l} Y_i \oplus \Sigma$). We concatenate a fixed $b-n$ bit string $c = 10^{b-n-1}$ with the $2n$ bit string $\Sigma||\Theta$ to match the input size of f_{K_1} and then the entire concatenated b bit string (i.e $c||\Sigma||\Theta$) is passed through f_{K_1} and finally outputs the tag T. We sometimes denote CS by M_{l+1}.
Note that, NI^+ is similar to that of NI upto $\mathbf{Casc}^{\mathbf{f_{K_1}}}(M)$ except the following differences.

(a) In NI construction, b-bit encoding of $|M|$ and the last message block output Y_l is passed through a different keyed compression function f_{K_2}. In NI^+, we substitute the b-bit length encoding by the checksum block CS. Moreover, CS and Y_l is passed through the same keyed compression function.
(b) NI is a two fixed-keyed compression function based MAC. NI^+ is a single fixed-keyed compression function based MAC.
(c) NI provides only birthday bound ($lq^2/2^n$) security. NI^+ provides beyond birthday bound security ($q^2l^4/2^{2n}$).

3.1 Design Rationale

We mention here that beyond birthday security is not possible to achieve if we just keep the original structure of NI-MAC and output Σ as the last block output (i.e $\Sigma = f_{K_2}(|M|, Y_l)$) and Θ as the sum of all intermediate chaining variables (i.e $\Theta = \oplus_{i=1}^{l} Y_i \oplus \Sigma$). This is justified by the following attack.

Let us assume that the adversary $\mathcal{A}$ makes q many queries of fixed number of blocks l where the second message block is different in each query. The probability of $Y_2^i = Y_2^j$ for $1 \leq i \neq j \leq q$ is $\frac{1}{2^n}$. Given that the event $Y_2^i = Y_2^j$ occurs, $\Sigma^i = \Sigma^j$ and $\Theta^i = \Theta^j$ would be a trivial event which implies the collision in output. Therefore, for any adversary the collision probability would become $\frac{q^2}{2^n}$. Note that we keep all the queried message length same. To resist this attack, we introduce a checksum block which is processed through the same function after all the message blocks are processed. We mention two important properties of checksum :[1]
(i) Difference in a single block of two distinct messages makes the different checksum value.
(ii) Difference in at least two blocks of two distinct messages may equalize the checksum value.
Now due to property (i) the above attack cannot make a trivial match in Σ . Moreover, due to property (ii) if differences in two message blocks of two distinct messages (a^{th} block and b^{th} block), ($M_a^i \neq M_a^j$) and ($M_b^i \neq M_b^j$) makes the checksum value equal, we are still guaranteed to obtain two output blocks for which Θ^i and Θ^j will not have a trivial match.

4 Security Analysis of NI^+-MAC

Gaži et. al in [14] have shown that the advantage of NI-MAC is bounded above by $\frac{q^2}{2^n}\left(l + \frac{64l^4}{2^n}\right)$. In this section we analyse the advantage of our construction NI^+-MAC and show that the advantage of NI^+-MAC achieves beyond birthday bound security; better than that of NI-MAC. Thus we have the following theorem.

[1] All the two properties are followed from Hamming distance of checksum is 2.

Theorem 1. *Let $f : \{0,1\}^k \times \{0,1\}^b \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a (ϵ, t, q) secure PRF. Then NI^+ be a $\left(\epsilon^{'}, t^{'}, q, l\right)$ secure PRF, where*

$$\epsilon^{'} \leq \epsilon + \frac{q}{2^n} + \frac{2q^2l^2}{2^{2n}} + \frac{4q^2l^4}{2^{2n}},$$

such that $t = t^{'} + \tilde{O}\,(lq)$.

Proof. Let $\mathcal{A}$ be a adaptive PRF-adversary against NI^+ running in time t and asking at most q queries, each of length at most ℓ blocks. NI^+ uses a single keyed function f. Now if we replace f by a uniformly distributed random function r such that $r \xleftarrow{\$} Func(\{0,1\}^b \times \{0,1\}^n, \{0,1\}^n)$ and call the resulting construction NI_r^+, then using the standard reduction from information theoretic setting to complexity theoretic setting we have,

$$\mathbf{Adv}^{\mathbf{prf}}_{\mathrm{NI}^+} \leq \epsilon + \mathbf{Adv}^{\mathbf{prf}}_{\mathrm{NI}_r^+}.$$

Therefore to prove Theorem 1, we only need to prove

$$\mathbf{Adv}^{\mathbf{prf}}_{\mathrm{NI}_r^+} \leq \tfrac{q}{2^n} + \tfrac{2q^2l^2}{2^{2n}} + \tfrac{4q^2l^4}{2^{2n}}.$$

Consider the following Game as shown in Algorithm 2 where the adversary $\mathcal{A}$ queries to oracle O with distinct messages M^i and obtains the response T^i. Note that Game G_0 truly simulates a uniform random function and G_1 simulates the actual construction NI_r^+. Therefore using the fundamental lemma of game-playing technique [7], we have the following:

$$\begin{aligned}\mathbf{Adv}^{\mathbf{prf}}_{\mathrm{NI}_r^+} &= \left|\Pr[\mathcal{A}^{G_1} = 1] - \Pr[\mathcal{A}^{G_0} = 1]\right| \\ &\leq \Pr[\mathcal{A}^{G_1} \text{ sets } \mathbf{badsigma} \vee \mathcal{A}^{G_1} \text{ sets } \mathbf{bad}] \\ &\leq \Pr[\mathcal{A}^{G_1} \text{ sets } \mathbf{badsigma}] + \Pr[\mathcal{A}^{G_1} \text{ sets } \mathbf{bad}]. \end{aligned} \tag{1}$$

Therefore, we evaluate now the probability $\Pr[\mathcal{A}^{G_1}$ sets $\mathbf{bad}]$. To evaluate this, let us define a double block function $\mathcal{H}_f(M) := (\Sigma, \Theta)$ with respect to a uniform random function f. Recall that the tuple $\mathcal{H}_f(M^i) := (\Sigma^i, \Theta^i)_i$, $\forall i \in [q]$ is said to be collision-free if $\forall i$, either $\Sigma^i \neq \Sigma^j$ or $\Theta^i \neq \Theta^j$ or both $\forall j \in [i-1]$. Similarly, the tuple $(\Sigma^i, \Theta^i)_i$ is said to be cover-free if $\forall i$, either $\Sigma^i \neq (M^j_\alpha)_{|n}$ or $\Theta^i \neq Y^j_{\alpha-1}$ or both $\forall j \in [i]$. Therefore, it is then easy to see that,

$$\begin{aligned}\Pr[\mathcal{A}^{G_1} \text{ sets } \mathbf{bad}] &\leq \mathbf{Adv}^{\mathrm{coll}}_{\mathtt{H}}(q,\ell) + \mathbf{Adv}^{\mathrm{cf}}_{\mathtt{H}}(q,\ell) \\ &\leq \frac{q^2}{2}(\mathbf{Adv}^{\mathrm{coll}}_{\mathtt{H}}(2,\ell) + \mathbf{Adv}^{\mathrm{cf}}_{\mathtt{H}}(2,\ell)). \end{aligned} \tag{2}$$

Now we state the following four lemmas, proof of which is deferred until next section. The first two lemmas (i.e Lemma 1 and 2) bounds the collision-free advantage and the last two lemmas (Lemma 3 and 4) bounds the cover-free advantage of function $H_f(\cdot)$.

Lemma 1. *For any two distinct messages M^i and M^j, each of length at most ℓ blocks,*

$$\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 0] \leq \tfrac{1}{2^{2n}}.$$

$initialize:$ **badsigma, bad** $\leftarrow false$;
On the j^{th} query M^j;
$M_1^j||M_2^j||\dots M_l^j \leftarrow M^j||10^* \leftarrow$ **Partition**(M^j), $Y_0 = 0$;
for $i = 1$ to l ;
  if $((M_i^j, Y_{i-1}^j) \in Dom(f))$ $Y_i \leftarrow f(M_i^j, Y_{i-1}^j)$;
  Else $Y_i^j \leftarrow \{0,1\}^n$;
  $f(M_i^j, Y_{i-1}^j) \leftarrow Y_i^j$;
  $Dom(f) \leftarrow Dom(f) \cup \{M_i^j, Y_{i-1}^j\}$;
if $((\oplus_{i=1}^l M_i^j, Y_l^j) \in Dom(f))$ $Y_{l+1}^j \leftarrow f(\oplus_{i=1}^l M_i^j, Y_l^j)$;
Else $Y_{l+1}^j \leftarrow \{0,1\}^n$;
$f(\oplus_{i=1}^l M_i^j, Y_l^j) \leftarrow Y_{l+1}^j$;
$Dom(f) \leftarrow Dom(f) \cup \{\oplus_{i=1}^l M_i^j, Y_l^j\}$;
$\Sigma^j \leftarrow Y_{l+1}^j$, $\Theta^j \leftarrow \oplus_{i=1}^{l+1} Y_i^j$;
if $(\Sigma^j = 0)$ **badsigma** $\leftarrow true$;
$T^j \xleftarrow{\$} \{0,1\}^n$;
if $((\Sigma^j, \Theta^j) = (\Sigma^i, \Theta^i)$ for some $i \in \{1, 2, \dots, j-1\}$, or $(c||\Sigma^j, \Theta^j) = (M_s^*, Y_{s-1}^*)$ such that $s \in [l_i + 1]$ or $s \in [l_j + 1], * \in \{i, j\})$;
  if $(\neg$**bad**$)$;
    Coll$(i, j) \leftarrow true$, **bad** $\leftarrow true$;
      if $((\Sigma^j, \Theta^j) = (\Sigma^i, \Theta^i))$ $\boxed{T^j \leftarrow f(\Sigma^i, \Theta^i)}$;
      Else $\boxed{T^j \leftarrow f(M_s^*, Y_{s-1}^*)}$;
Return T^j;

Algorithm 2: Game G_0 is without boxed statement and G_1 is with boxed statement.

Lemma 2. *For any two distinct messages M^i and M^j, each of length at most ℓ blocks,*

$$\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}.$$

Lemma 3. *For any two distinct messages M^i and M^j, each of length at most ℓ blocks, and a particular n bit constant x,*

$$\Pr[\Sigma^i = x \wedge \Theta^i = Y_s^t \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}.$$

Lemma 4. *For any two distinct messages M^i and M^j, each of length at most ℓ blocks, and a particular n bit constant x,*

$$\Pr[\Sigma^i = x \wedge \Theta^i = Y_s^t \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}.$$

Resume the proof of Theorem 1: Now we have all the materials to prove Theorem 1 which is given in the following.

We have the following results,

$$\mathbf{Adv}^{\mathrm{coll}}_{\mathtt{H}}(2,\ell) \le \frac{2\ell^2}{2^{2n}} + \frac{4\ell^4}{2^{2n}}. \quad \text{[From Lemma 1 and 2].} \tag{3}$$

$$\mathbf{Adv}^{\mathrm{cf}}_{\mathtt{H}}(2,\ell) \le \frac{2\ell^2}{2^{2n}} + \frac{4\ell^4}{2^{2n}}. \quad \text{[From Lemma 3 and 4].} \tag{4}$$

Substituting Equation (3) and (4) into Equation (2) we obtain

$$\Pr[\mathcal{A}^{G_1} \text{ sets } \mathbf{bad}] \le \frac{2q^2l^2}{2^{2n}} + \frac{4q^2l^4}{2^{2n}}.$$

Moreover it is easy to see that $\Pr[\mathcal{A}^{G_1}$ sets **badsigma**$] \le \frac{q}{2^n}$. Therefore, substituting these two probability expressions back to Equation (1) will give

$$\mathbf{Adv}^{\mathbf{prf}}_{\mathrm{NI}^+_r} \le \frac{q}{2^n} + \frac{2q^2l^2}{2^{2n}} + \frac{4q^2l^4}{2^{2n}}. \qquad \square$$

4.1 Proof of $\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 0] \le \frac{1}{2^{2n}}$.

In this section we will prove the following lemma.

Lemma 1. For any two distinct messages M^i and M^j, each of length at most ℓ blocks,

$$\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 0] \le \frac{1}{2^{2n}}.$$

Proof. We prove the lemma using the structure graph. After fixing two distinct messages M^i and M^j and choosing a function f uniformly at random from the set of all functions over $\{0,1\}^b \times \{0,1\}^n \to \{0,1\}^n$, we analyse the structure graph $G := \mathcal{G}^f(M^i, M^j)$. In particular, we analyse the probability of the event $\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j$ in view of the number of collisions $|fColl(G)| = 0$ occurred in the corresponding structure graph G. Let W denotes the event $\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j$.

Case a. Let us consider M^i is not a prefix of M^j and M^j is not a prefix of M^i. Let p be the longest common prefix (lcp) of M^i and M^j. That means $M^i_{p+1} \ne M^j_{p+1}$ and $M^i_\alpha = M^j_\alpha$ where $1 \le \alpha \le p$. Therefore, $Y^i_\alpha = Y^j_\alpha$ and $Y^i_\beta \ne Y^j_\beta$ where $p+1 \le \beta \le \min\{l_i, l_j\}$ as $|fColl(G)| = 0$. Moreover, if $l_i > l_j$ then all Y^i_β would have been distinct as $|fColl(G)| = 0$ where $l_j + 1 \le \beta \le l_i$. Note that, it is also true that $Y^i_{l_i} \ne Y^j_{l_j}$. Therefore, we have,

$$\Pr[W \wedge |fColl(G)| = 0] = \Pr[\Theta^i = \Theta^j \wedge |fColl(G)| = 0 | \Sigma^i = \Sigma^j] \cdot \Pr[\Sigma^i = \Sigma^j]$$

It is obvious that $\Pr[\Sigma^i = \Sigma^j] \le \frac{1}{2^n - 2\ell}$ and the event $\Theta^i = \Theta^j \wedge |fColl(G)| = 0$ conditioned on the event $\Sigma^i = \Sigma^j$ implies a non trivial equation on $\boldsymbol{Y}$ as we will obtain some Y^i_{p+1} and Y^j_{p+1} for which $\Theta^i \oplus \Theta^j = 0$ would become non-trivial. Thus, $\Pr[\Theta^i = \Theta^j \wedge |fColl(G)| = 0 | \Sigma^i = \Sigma^j] \le \frac{1}{2^n}$. Therefore,

$$\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 0] \le \frac{1}{2^{2n}}.$$

Case b. Let us consider that either of the two messages is a prefix of other (w.l.o.g M^j is a prefix of M^i). Since $l_i > l_j$ therefore, $p = l_j$. Since $|fColl(G)| = 0$, $Y^i_{p+1}, \ldots Y^i_{l_i}$ are all distinct with each other and with $Y^i_1, \ldots, Y^i_{l_j}$. This implies that $Y^i_{l_i} \neq Y^j_{l_j}$ as depicted in Fig. 4.1. Therefore, the probability of $\Theta^i = \Theta^j \wedge |fColl(G)| = 0$ conditioned on the event $\Sigma^i = \Sigma^j$ will be $O(1/2^n)$ as we will obtain two random variables $Y^i_{l_i}$ and $Y^j_{l_j}$ for which $\Theta^i \oplus \Theta^j = 0$ would become non-trivial. Moreover, $\Pr[\Sigma^i = \Sigma^j] \leq \frac{1}{2^n}$. Therefore again,

$$\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}.$$

□

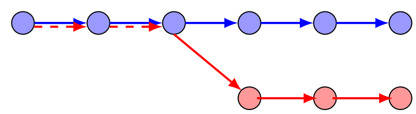

Fig. 4.1: Structure Graph of accident 0

4.2 Proof of $\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}$.

In this section we will prove the following lemma.

Lemma 2. For any two distinct messages M^i and M^j, each of length at most ℓ blocks,

$$\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}.$$

Proof. Again we prove the lemma using the structure graph analysis. We fix two distinct messages M^i and M^j and a uniformly chosen function f from the set of all functions over $\{0,1\}^b \times \{0,1\}^n \to \{0,1\}^n$. Then we analyze the structure graph $G := G^f(M^i, M^j)$. In particular, here we will analyze the probability of the event $\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j$ in view of number of collisions $|fColl(G)| = 1$ occurred in the corresponding structure graph G. Let W denotes the event $\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j$.

We analyze the probability of the above event in two subcases. a) We consider all those structure graphs G having $|fColl(G)| = 1$ with respect to $\mathcal{M} := \{M^i, M^j\}$ where none of W_i and W_j, where W_i and W_j is the walk of message M^i and M^j in structure graph G, contains a loop. It essentially implies that W_i and W_j are path which are denoted as P_i and P_j respectively. b) We consider all those structure graphs where either of W_i or W_j contains a loop.

Let $\mathcal{G}$ denote the set of all structure graphs G with $|fColl(G)| = 1$. Without loss of generality, let l_i and l_j be the lengths of the messages M^i and M^j respectively, with $l_i \geq l_j$. Let $\mathcal{G}_1 \subset \mathcal{G}$ be the set of all structure graphs such that the M^i, M^j-path does not contain any loop. The $\mathcal{G}_2 = \mathcal{G} \setminus \mathcal{G}_1$ is the set of the remaining structure graphs.

a) Analysis of $\mathcal{G}_1$. It is to be noted that if M^j is a proper prefix of M^i then $|\mathcal{G}_1| = 0$, as in that case $|fColl(G)| = 0$. So without loss of generality, lets assume that M^j is not a prefix of M^i. Suppose the first p blocks constitute the longest common prefix of M^i and M^j. Therefore, $Y^i_\alpha = Y^j_\alpha$, $1 \le \alpha \le p$. As number of collision is 1 therefore, let the colliding pair is $(Y^i_{\beta_i}, Y^j_{\beta_j})$, where $p+1 \le \beta_i \le l_i, p+1 \le \beta_j \le l_j$.

Case 1. Let $\beta_i = \beta_j = p+1$ and $l_i = l_j$ and after the collision $Y^i_\beta = Y^j_\beta$, for $p+2 \le \beta \le l_i$. In this case, it is clear that checksum block of i^{th} message CS^i and checksum block of j^{th} message CS^j would not be equal due to property (i) stated in Section 3.1 and hence, even if $Y^i_{l_i} = Y^j_{l_j}$, the event $\Sigma^i = \Sigma^j$ would not be trivial. Therefore, even though $\Pr[\Theta^i = \Theta^j | \Sigma^i = \Sigma^j \wedge |fColl(G)| = 1] = 1$, but the required randomness will be obtained from the following two equations : (i) $Y^i_{p+1} \oplus Y^j_{p+1} = 0$, (ii) $\Sigma^i \oplus \Sigma^j = 0$ such that the rank of the system of equations is 2.

Case 2. Let $\beta_i = \beta_j = p+1$ and $l_i = l_j$ and after the collision $Y^i_\beta \ne Y^j_\beta$, for $p+2 \le \beta \le l_i$. Then we will always obtain Y^i_k and $Y^j_{k'}$ such that $\Theta^i = \Theta^j$ is non-trivial for some k, k'. Therefore in this case we have,

$$\Pr[\Theta^i = \Theta^j \wedge |fColl(G)| = 1 | \Sigma^i = \Sigma^j] \le \frac{1}{2^n}.$$

Case 3. Let $\beta_i = \beta_j = p+2$ and $l_i = l_j$ and $Y^i_\beta = Y^j_\beta$, for $p+3 \le \beta \le l_i$, then $\Theta^i = \Theta^j$ would imply $Y^i_{p+1} = Y^j_{p+1}$; creates one more collision which violates the condition that the structure graph has only one collision.

Therefore, in general, we assume that the colliding pair is $(Y^i_{\beta_i}, Y^j_{\beta_j})$, where $p+1 \le \beta_i \le l_i, p+1 \le \beta_j \le l_j$. Since the number of collision allowed in G is 1, after the collision point either P_i and P_j follow the same path or they will get bifurcated right from the collision point and will never meet again. If P_i and P_j follows the same path, then for Case 1 we have shown that we can ensure to get the probability $O(1/2^n)$. If not, then except Case 3 where $\beta_i = \beta_j = p+2$, we will obtain two random variables Y^i_k and $Y^j_{k'}$ such that equation $\Theta^i \oplus \Theta^j = 0$ becomes non-trivial. If P_i and P_j gets bifurcated right after the collision point, then the equality of Θ becomes non-trivial for two random variables Y^i_{p+1} and Y^j_{p+1} as depicted in Fig. 4.2 and Fig. 4.3. Note that it is easy to follow that we will always obtain two such random variables.

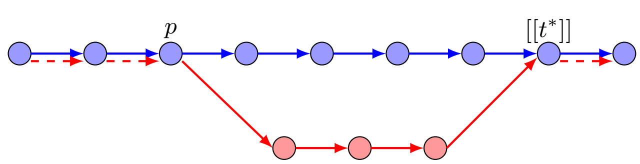

Fig. 4.2: Structure Graph of type G_1

Case 4. Finally, if $\beta_i = l_i$ and $\beta_j = l_j$ then one can easily find out two random variables from the set $\{Y^i_{p+1}, \ldots, Y^i_{l_i-1}\} \cup \{Y^j_{p+1}, \ldots, Y^j_{l_j-1}\}$ such that the equation on Θ becomes non-trivial.

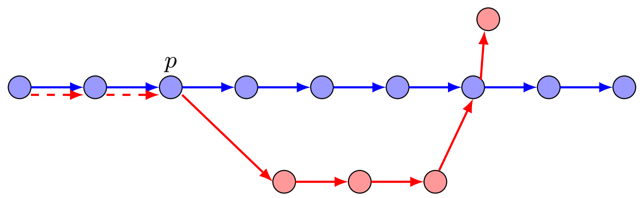

Fig. 4.3: Structure Graph of type G_1; M^i path has no loop

Since the structure graph involving only one collision is determined by the collision point, $|\mathcal{G}_1| \leq l^2$ and for each of this graph we have seen that the probability of desired event is $O(1/2^{2n})$. Therefore,

$\Pr[W \wedge |fColl(G)| = 1] = \sum_{G \in \mathcal{G}_1} \Pr[W \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}$.

b) Analysis of $\boldsymbol{\mathcal{G}_2}$. Recall that $\mathcal{G}_2$ is the set of all structure graphs with respect to $\mathcal{M}$ such that the number of collision is 1 and containing a loop. Without loss of generality we assume that W_i contains a loop. That means $Y^i_\alpha = Y^i_{\alpha+c}$ for $c \geq 1$. Here c denotes the loop size. Note that, the loop actually creates a collision and therefore, neither (i) W_j or W_i makes another different loop, nor (ii) W_j collides with W_i as in both of the cases number of collisions will increase to 2. Thus, the only possibilities are either (1) W_j completely lies on W_i, or (2) W_j could follow W_i but after a point W_j and W_i gets bifurcated and never meets. We will analyze the probability of the event $\Sigma^i = \Sigma^j \wedge \Theta^i = \Theta^j \wedge |fColl(G)| = 1$ separately for each of the above cases.

Case 1. Let us assume $W_i = Y^i_1||\ldots||Y^i_{\alpha-1}||(Y^i_\alpha||\ldots||Y^i_{\alpha+c-1})^k||Y^i_{\alpha+c+1}||\ldots||Y^i_{l_i}$ and $W_j = Y^j_1||\ldots||Y^j_{\alpha-1}||(Y^j_\alpha||\ldots||Y^j_{\alpha+c-1})^{k'}||Y^j_{\alpha+c+1}||\ldots||Y^j_{l_j}$ where $k' \geq 0$. Now we have the following cases:

Case 1.1. As W_j lies on W_i, it is easy to see that if $k' = 0$ then W_j be a subsequence of $Y^i_1||\ldots||Y^i_{\alpha-1}$ and therefore one can ensures the non-triviality of equation $\Theta^i = \Theta^j$ which holds with probability $\frac{1}{2^n}$. Moreover, $Y^i_{l_i} \neq Y^j_{l_j}$ and thus $\Sigma^i = \Sigma^j$ also holds with probability $\frac{1}{2^n}$ and therefore $\Pr[W \wedge |fColl(G)| = 1] \leq \frac{1}{2^{2n}}$.

Case 1.2. If $k' \geq 1$, then it is obvious that $Y^j_1||\ldots||Y^j_{\alpha-1} = Y^i_1||\ldots||Y^i_{\alpha-1}$. Now if we assume that the length of the tail of W_i (i.e $Y^i_{\alpha+c+1}||\ldots||Y^i_{l_i}$) is same as that of W_j then it must have been the case that $k \neq k'$ and without loss of generality we can assume that $k > k'$. Since $Y^i_{l_i} = Y^j_{l_j}$, depending on the equality of CS^i and CS^j we have $\Pr[\Sigma^i = \Sigma^j|\ \ |fColl(G)| = 1] = 1$. Therefore,

$$\begin{aligned}\Pr[W \wedge |fColl(G)| = 1] = \Pr[&\Theta^i = \Theta^j|\Sigma^i = \Sigma^j \wedge |fColl(G) = 1] \\ &\cdot\ \Pr[\Sigma^i = \Sigma^j|\ \ |fColl(G)| = 1] \cdot \Pr[|fColl(G) = 1]\end{aligned}$$

As $k > k'$ therefore, it is obvious to see that there must be at least two random variables Y^i_s and $Y^j_{s'}$ for which $\Theta^i = \Theta^j$ would become non-trivial as depicted in Fig. 4.4. Thus in the above equation, $\Pr[\Theta^i = \Theta^j|\Sigma^i = \Sigma^j \wedge |fColl(G) = 1] \leq$

$\frac{1}{2^n}$ and $\Pr[|fColl(G)| = 1] \leq \frac{1}{2^n}$. Therefore, $\Pr[W \wedge |fColl(G)| = 1] \leq \frac{1}{2^{2n}}$.

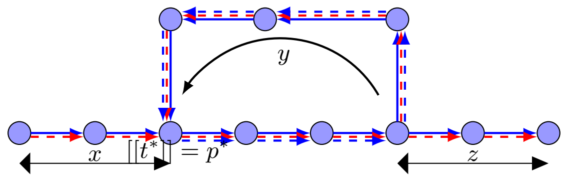

Fig. 4.4: Structure Graph of type G_2

Moreover, if we assume that the tail length of W_i and W_j are not same (w.l.o.g $tail(W_i) > tail(W_j)$) then we have either $k = k'$ or $k \neq k'$. The case of $k = k'$ has already been taken care of. If $k \neq k'$ then $Y^i_{l_i} \neq Y^j_{l_j}$ and therefore, $\Theta^i \oplus \Theta^j = 0$ would become non-trivial for the random variable $Y^i_{l_i}$ and $Y^j_{l_j}$. Moreover, $\Pr[\Sigma^i = \Sigma^j] \leq \frac{1}{2^n}$. Thus,

$$\Pr[W \wedge |fColl(G)| = 1] \leq \frac{1}{2^{2n}}.$$

Case 2. In this case W_j bifurcates from W_i right after some point X. This condition necessarily implies that $Y^i_{l_i} \neq Y^j_{l_j}$. Now it is to be noted that if W_j completely lies on W_i (as in $head(W_i) = head(W_j)$ and $k = k'$) and bifurcates right from the point $X = Y^i_{l_i-1}$, then $\Theta^i = \Theta^j$ would imply $Y^i_{l_i} = Y^j_{l_j}$, introduces one more collision and hence the number of collision would increase. Therefore, even if $head(W_i) = head(W_j)$ either $k \neq k'$ or W_j must get bifurcated from W_i from some earlier point of $Y^i_{l_i-1}$. In both of these cases one should obtain at least two random variables (*either from portion of loop or from portion of tail*) Y^i_s and $Y^i_{s'}$ for some s and s' that ensures the non-triviality of equation on Θ as depicted in Fig. 4.5.

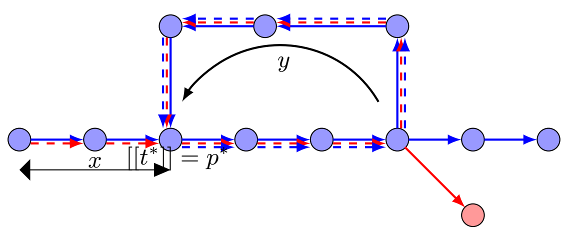

Fig. 4.5: Structure Graph of type G_2; M^i, M^j both path contain a loop

Moreover as $Y^i_{l_i} \neq Y^j_{l_j}$ this ensures that $\Pr[\Sigma^i = \Sigma^j] \leq \frac{1}{2^n}$. Hence, $\Pr[W \wedge |fColl(G)| = 1] \leq \frac{1}{2^{2n}}$.

Note that in all of these cases we have seen that the probability of the desired event becomes $\frac{1}{2^{2n}}$ and $|\mathcal{G}_2| \leq l^2$ as the structure graph G is completely determined by the size of the loop which is formed by choosing any two Y values in $\binom{l}{2}$ ways. Therefore,

$\Pr[W \wedge |fColl(G)| = 1] = \sum_{G \in \mathcal{G}_2} \Pr[W \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}$. $\square$

4.3 Proof of $\Pr[\Sigma^i = x \wedge \Theta^i = Y_s^t \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}$.

In this section we will prove the following lemma.

Lemma 3. For any two distinct messages M^i and M^j, each of length at most ℓ blocks, and a particular n bit constant x,

$$\Pr[\Sigma^i = x \wedge \Theta^i = Y_s^t \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}.$$

Proof. We again prove the lemma using structure graph where we fix two distinct messages M^i and M^j and a function f uniformly at random from the set of all functions over $\{0,1\}^b \times \{0,1\}^n \to \{0,1\}^n$, we construct the structure graph $G := \mathcal{G}^f(M^i, M^j)$ such that there is no accidental collision (i.e. f-collision) in G. Then we analyze the probability of the event denoted by $W := \Sigma^i = x \wedge \Theta^i = Y_s^t$ in view of the number of f-collisions $|fColl(G)| = 0$ occurred in the corresponding structure graph G where x is any non-zero n bit constant.

We analyse the probability of the event $W \wedge |fColl(G)| = 0$ in two separate subcases when (a) None of M^i, M^j is a prefix of each other and (b) either of M^i, M^j is a prefix of the other.

Case a. Let us consider M^i is not a prefix of M^j and M^j is not a prefix of M^i. Let p be the longest common prefix (lcp) of M^i and M^j. Therefore, $Y_\alpha^i = Y_\alpha^j$ where $1 \leq \alpha \leq p$ and $Y_\beta^i \neq Y_\beta^j$ where $p+1 \leq \beta \leq \min\{l_i, l_j\}$ as $|fColl(G)| = 0$. Moreover, if $l_i > l_j$ then all Y_β^i would have been distinct as $|fColl(G)| = 0$ where $l_j + 1 \leq \beta \leq l_i$. Note that, it is also true that $Y_{l_i}^i \neq Y_{l_j}^j$. Therefore, we have the following set of equations:

$$Y_{l_i+1}^i = x, \tag{5}$$
$$Y_1^i \oplus Y_2^i \oplus \ldots \oplus Y_{l_i+1}^i + Y_t^s = 0, \tag{6}$$

where s could be either i or j and $t \in [l_i+1]$ or $t \in [l_j+1]$. For each of these cases one can easily check that the above system of equation has rank 2. Therefore, $\Pr[W \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}$.

Case b. W.l.o.g let us consider that M^j is a prefix of M^i. Since $l_i > l_j$ therefore, $p = l_j$. Since $|fColl(G)| = 0$, $Y_{p+1}^i, \ldots Y_{l_i}^i$ are all distinct with each other and with $Y_1^i, \ldots, Y_{l_j}^i$. This implies that $Y_{l_i}^i \neq Y_{l_j}^j$ as depicted in Fig. 4.1. Therefore, the set of equations (Equation (5) and (6)) has the full rank. Therefore, again we have, $\Pr[W \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}$.

Therefore, combining Case a and b we have,

$$\Pr[\Sigma^i = x \wedge \Theta^i = Y_t^s \wedge |fColl(G)| = 0] \leq \frac{1}{2^{2n}}$$

for any non-zero n bit constant x. $\square$

4.4 Proof of $\Pr[\Sigma^i = x \wedge \Theta^i = Y_s^t \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}$.

In this section we will prove the following lemma.

Lemma 4. For any two distinct messages M^i and M^j, each of length at most ℓ blocks, and a particular n bit constant x,

$$\Pr[\Sigma^i = x \wedge \Theta^i = Y_s^t \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}.$$

Proof. We prove this lemma using the structure graph analysis. The primary concern of this proof will be to analyze the probability of the event $W := \Sigma^i = x \wedge \Theta^i = Y_t^s$ in view of number of collisions $|fColl(G)| = 1$ occurred in the corresponding structure graph G where $G := G^f(M^i, M^j)$.

We analyze the probability of the event $W \wedge |fColl(G)| = 1$ in two separate subcases. a) When we consider all structure graphs G with respect to $\mathcal{M} := \{M^i, M^j\}$ such that $|fColl(G)| = 1$ and none of W_i and W_j, where W_i and W_j is the walk of message M^i and M^j in structure graph G, contains a loop. It implies that W_i and W_j are path which are denoted as P_i and P_j respectively. b) When we consider all those structure graphs where either W_i or W_j contains a loop.

As before, let $\mathcal{G}$ denote the set of all structure graphs G with $|fColl(G)| = 1$. Without loss of generality, let l_i and l_j be the lengths of the messages M^i and M^j respectively, with $l_i \geq l_j$. Let $\mathcal{G}_1 \subset \mathcal{G}$ be the set of all structure graphs such that the W_i, W_j does not contain any loop. The $\mathcal{G}_2 = \mathcal{G} \setminus \mathcal{G}_1$ is the set of the remaining structure graphs.

a) Analysis of $\mathcal{G}_1$. As before M^i or M^j could not be a prefix of each other. Let p be the *lcp* of M^i and M^j and let the colliding pair is $(Y^i_{\beta_i}, Y^j_{\beta_j})$, where $p+1 \leq \beta_i \leq l_i, p+1 \leq \beta_j \leq l_j$. In this case, it is easy to check that the following system of equations will have rank 2.

$$Y^i_{l_i+1} = x,$$
$$Y^i_1 \oplus Y^i_2 \oplus \ldots \oplus Y^i_{l_i+1} + Y^s_t = 0.$$

Therefore, we have $\Pr[W \wedge |fColl(G)| = 1] \leq \frac{1}{2^{2n}}$.
Note that $|\mathcal{G}_1|$ is l^2 as the graph is uniquely determined by the accident point and the set of messages $\mathcal{M}$. Therefore, $\Pr[W \wedge |fColl(G) = 1|] \leq \frac{l^2}{2^{2n}}$.

b) Analysis of $\mathcal{G}_2$. As before let us assume that W_i contains a loop of size c such that $Y^i_\alpha = Y^i_{\alpha+c}$ for $c \geq 1$. Since the loop creates a f-collision, neither (i) W_j or W_i makes another different loop, nor (ii) W_j collides with W_i as in both of the cases the number of collisions will increase to 2. Thus we have the following two possibilities.

(1) W_j coincides with W_i
(2) W_j could follow W_i but after a point W_i and W_j departs and never meets again.

We analyze the probability of the event $W \wedge |fColl(G)| = 1$ separately for each of the two above cases. In particular, in each of the following analysis our main concern will be to show the rank of the set of equations as defined earlier (i.e Equation (5) and (6)) to be 2, that is it achieves full rank in each of the following subcases.

Case 1. Let k denotes the number of iterations in the loop of W_i and k' be the number of iterations in the loop of W_j. Now irrespective of the value of k and k', the system of equations (Equation (5) and (6)) will have rank 2 and therefore, we can upper bound the probability of our desired event to $\frac{1}{2^{2n}}$. Note that $G \in \mathcal{G}_2$ is uniquely determined by the size of the loop c and hence, $|\mathcal{G}_2| \leq l^2$, Thus,

$$\Pr[W \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}. \qquad (7)$$

Case 2. The analysis for this case would be similar to Case 1. Here W_i and W_j bifurcates from a certain point say X and $l_i - X, l_j - X \neq 0$. Therefore, it is trivial to see that the set of equations (i.e Equation (5) and (6)) will have full rank. Again, as we have shown in the previous case that $|\mathcal{G}_2| \leq l^2$ and therefore Equation (7) will also hold in this case.

Combining the probability bound of the event $W \wedge |fColl(G)| = 1$ from each of the two above subcases, we have derived $\Pr[W \wedge |fColl(G)| = 1] \leq \frac{l^2}{2^{2n}}$. $\square$

5 Conclusion

In this paper, we have proposed a variant of NI-MAC, which we call as NI$^+$-MAC and have shown that NI$^+$-MAC achieves BBB security. We have also shown that keeping the original structure of NI as discussed in Section 3.1 cannot achieve BBB security. Moreover, our non-tweaked proposed construction is better than Yasuda's proposed single-fixed key compression function based MAC construction that uses an extra tweak. NI$^+$ is also efficient than NI-MAC in terms of number of keys and providing better security.

References

1. Jee Hea An and Mihir Bellare. Constructing vil-macsfrom fil-macs: Message authentication under weakened assumptions. In Wiener [37], pages 252–269.
2. Mihir Bellare, Ran Canetti, and Hugo Krawczyk. Keying hash functions for message authentication. In Neal Koblitz, editor, CRYPTO '96, volume 1109 of *LNCS*, pages 1–15. Springer, 1996.
3. Mihir Bellare, Oded Goldreich, and Hugo Krawczyk. Stateless evaluation of pseudorandom functions: Security beyond the birthday barrier. In Wiener [37], pages 270–287.
4. Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of cipher block chaining. In Yvo Desmedt, editor, CRYPTO '94, volume 839 of *LNCS*, pages 341–358. Springer, 1994.

5. Mihir Bellare, Krzysztof Pietrzak, and Phillip Rogaway. Improved security analyses for CBC macs. In Shoup [32], pages 527–545.
6. Mihir Bellare. New proofs for NMAC and HMAC: security without collision-resistance. In Cynthia Dwork, editor, CRYPTO 2006, volume 4117 of *LNCS*, pages 602–619. Springer, 2006.
7. Mihir Bellare and Phillip Rogaway. The security of triple encryption and a framework for code-based game-playing proofs. In Vaudenay [33], pages 409–426.
8. John Black, Shai Halevi, Hugo Krawczyk, Ted Krovetz, and Phillip Rogaway. UMAC: fast and secure message authentication. In Wiener [37], pages 216–233.
9. John Black and Phillip Rogaway. A block-cipher mode of operation for parallelizable message authentication. In Knudsen [21], pages 384–397.
10. Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe. PRESENT: an ultra-lightweight block cipher. In Pascal Paillier and Ingrid Verbauwhede, editors, CHES 2007, volume 4727 of *LNCS*, pages 450–466. Springer, 2007.
11. Nilanjan Datta, Avijit Dutta, Mridul Nandi, Goutam Paul, and Liting Zhang. One-key double-sum mac with beyond-birthday security. Cryptology ePrint Archive, Report 2015/958, 2015. http://eprint.iacr.org/.
12. Yevgeniy Dodis, Thomas Ristenpart, John P. Steinberger, and Stefano Tessaro. To hash or not to hash again? (in)differentiability results for H 2 and HMAC. In Reihaneh Safavi-Naini and Ran Canetti, editors, CRYPTO 2012, volume 7417 of *LNCS*, pages 348–366. Springer, 2012.
13. Yevgeniy Dodis and John P. Steinberger. Domain extension for macs beyond the birthday barrier. In Kenneth G. Paterson, editor, EUROCRYPT 2011, volume 6632 of *LNCS*, pages 323–342. Springer, 2011.
14. Peter Gazi, Krzysztof Pietrzak, and Michal Rybár. The exact prf-security of NMAC and HMAC. In Juan A. Garay and Rosario Gennaro, editors, CRYPTO 2014, volume 8616 of *LNCS*, pages 113–130. Springer, 2014.
15. Peter Gaži, Krzysztof Pietrzak, and Stefano Tessaro. Generic security of nmac and hmac with input whitening. Cryptology ePrint Archive, Report 2015/881, 2015. http://eprint.iacr.org/.
16. Deukjo Hong, Jaechul Sung, Seokhie Hong, Jongin Lim, Sangjin Lee, Bonseok Koo, Changhoon Lee, Donghoon Chang, Jaesang Lee, Kitae Jeong, Hyun Kim, Jongsung Kim, and Seongtaek Chee. HIGHT: A new block cipher suitable for low-resource device. In Louis Goubin and Mitsuru Matsui, editors, CHES 2006, volume 4249 of *LNCS*, pages 46–59. Springer, 2006.
17. Tetsu Iwata and Kaoru Kurosawa. OMAC: one-key CBC MAC. In Johansson [19], pages 129–153.
18. Éliane Jaulmes, Antoine Joux, and Frédéric Valette. On the security of randomized CBC-MAC beyond the birthday paradox limit: A new construction. In FSE, 2002, volume 2365 of *LNCS*, pages 237–251. Springer, 2002.
19. Thomas Johansson, editor. In FSE, 2003, volume 2887 of *LNCS*. Springer, 2003.
20. Antoine Joux, Guillaume Poupard, and Jacques Stern. New attacks against standardized macs. In Johansson [19], pages 170–181.
21. Lars R. Knudsen, editor. EUROCRYPT 2002, volume 2332 of *LNCS*. Springer, 2002.
22. Neal Koblitz and Alfred Menezes. Another look at hmac. *J. Mathematical Cryptology*, 7(3):225–251, 2013.
23. H. Krawczyk, M. Bellare, and R. Canetti. HMAC: Keyed-Hashing for Message Authentication. RFC 2104 (Informational), February 1997.

24. Gaëtan Leurent, Thomas Peyrin, and Lei Wang. New generic attacks against hash-based macs. In Kazue Sako and Palash Sarkar, editors, ASIACRYPT 2013, volume 8270 of *LNCS*, pages 1–20. Springer, 2013.
25. Stefan Lucks. A failure-friendly design principle for hash functions. In Bimal K. Roy, editor, ASIACRYPT 2005, volume 3788 of *LNCS*, pages 474–494. Springer, 2005.
26. Ueli M. Maurer and Johan Sjödin. Domain expansion of macs: Alternative uses of the FIL-MAC. In Nigel P. Smart, editor, *Cryptography and Coding, 2005*, volume 3796 of *LNCS*, pages 168–185. Springer, 2005.
27. Ueli M. Maurer and Johan Sjödin. Single-key ail-macs from any FIL-MAC. In Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung, editors, ICALP 2005, volume 3580 of *LNCS*, pages 472–484. Springer, 2005.
28. Kazuhiko Minematsu. How to thwart birthday attacks against macs via small randomness. In Seokhie Hong and Tetsu Iwata, editors, FSE, 2010, volume 6147 of *LNCS*, pages 230–249. Springer, 2010.
29. Yusuke Naito, Yu Sasaki, Lei Wang, and Kan Yasuda. Generic state-recovery and forgery attacks on chopmd-mac and on NMAC/HMAC. In Kazuo Sakiyama and Masayuki Terada, editors, IWSEC 2013, volume 8231 of *LNCS*, pages 83–98. Springer, 2013.
30. Thomas Peyrin, Yu Sasaki, and Lei Wang. Generic related-key attacks for HMAC. In Wang and Sako [34], pages 580–597.
31. Thomas Peyrin and Lei Wang. Generic universal forgery attack on iterative hash-based macs. In Phong Q. Nguyen and Elisabeth Oswald, editors, EUROCRYPT 2014, volume 8441 of *LNCS*, pages 147–164. Springer, 2014.
32. Victor Shoup, editor. CRYPTO 2005, volume 3621 of *LNCS*. Springer, 2005.
33. Serge Vaudenay, editor. EUROCRYPT 2006, volume 4004 of *LNCS*. Springer, 2006.
34. Xiaoyun Wang and Kazue Sako, editors. ASIACRYPT 2012, volume 7658 of *LNCS*. Springer, 2012.
35. Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full SHA-1. In Shoup [32], pages 17–36.
36. Xiaoyun Wang and Hongbo Yu. How to break MD5 and other hash functions. In Ronald Cramer, editor, EUROCRYPT 2005, volume 3494 of *LNCS*, pages 19–35. Springer, 2005.
37. Michael J. Wiener, editor. CRYPTO '99, volume 1666 of *LNCS*. Springer, 1999.
38. Kan Yasuda. Boosting merkle-damgård hashing for message authentication. In Kaoru Kurosawa, editor, ASIACRYPT 2007, volume 4833 of *LNCS*, pages 216–231. Springer, 2007.
39. Kan Yasuda. Multilane HMAC - security beyond the birthday limit. In K. Srinathan, C. Pandu Rangan, and Moti Yung, editors, INDOCRYPT 2007, volume 4859 of *LNCS*, pages 18–32. Springer, 2007.
40. Kan Yasuda. "sandwich" is indeed secure: How to authenticate a message with just one hashing. In Josef Pieprzyk, Hossein Ghodosi, and Ed Dawson, editors, ACISP 2007, volume 4586 of *LNCS*, pages 355–369. Springer, 2007.
41. Kan Yasuda. A one-pass mode of operation for deterministic message authentication- security beyond the birthday barrier. In Kaisa Nyberg, editor, FSE, 2008, volume 5086 of *LNCS*, pages 316–333. Springer, 2008.
42. Kan Yasuda. A double-piped mode of operation for macs, prfs and pros: Security beyond the birthday barrier. In Antoine Joux, editor, EUROCRYPT 2009, volume 5479 of *LNCS*, pages 242–259. Springer, 2009.

43. Kan Yasuda. HMAC without the "second" key. In Pierangela Samarati, Moti Yung, Fabio Martinelli, and Claudio Agostino Ardagna, editors, ISC 2009, volume 5735 of *LNCS*, pages 443–458. Springer, 2009.
44. Kan Yasuda. The sum of CBC macs is a secure PRF. In Josef Pieprzyk, editor, CT-RSA 2010, volume 5985 of *LNCS*, pages 366–381. Springer, 2010.
45. Kan Yasuda. A new variant of PMAC: beyond the birthday bound. In Phillip Rogaway, editor, CRYPTO 2011, volume 6841 of *LNCS*, pages 596–609. Springer, 2011.
46. Kan Yasuda. On the full MAC security of a double-piped mode of operation. *IEICE Transactions*, 94-A(1):84–91, 2011.
47. Kan Yasuda. A parallelizable prf-based MAC algorithm: Well beyond the birthday bound. *IEICE Transactions*, 96-A(1):237–241, 2013.
48. Liting Zhang, Wenling Wu, Han Sui, and Peng Wang. 3kf9: Enhancing 3gpp-mac beyond the birthday bound. In Wang and Sako [34], pages 296–312.

A Formal Discussion on Structure Graph

Let for two distinct messages M^1 and M^2 of l_1 and l_2 blocks respectively, where

$$M^1 = M_1^1||M_2^1||\dots||M_{l_1}^1 \text{ and } M^2 = M_1^2||M_2^2||\dots||M_{l_2}^2,$$

and the corresponding Y-values be given by

$$y_0^1, y_1^1, y_2^1, \dots, y_{l_1}^1 \text{ and } y_0^2, y_1^2, y_2^2, \dots, y_{l_2}^2$$

respectively. Let $\tau = l_1 + l_2$. We use the notation M_i where $1 \le i \le \tau$ to refer to the block M_i^1, when $i \le l_1$, otherwise refer to the block $M_{i-l_1}^2$. Similarly, let Y_i to refer to $\mathbf{0}$ when $i = 0$; Y_i^1, when $1 \le i \le l_1$; and $Y_{i-l_1}^2$, when $l_1 + 1 \le i \le \tau$. Now, we give a few definitions.

Definition 2. *We define two mappings* $[[\cdot]]$ *and* $[[\cdot]]'$ *on* $\{0, \dots, \tau\}$ *as follows:*
(1) $[[i]] \triangleq \min\{j : Y_i = Y_j\}$, *and*
(2) $[[i']] \triangleq [[i]]$ *for* $i \ne l_1$ *except that* $[[l_1]]' = 0$.

Definition 3. *For any fixed* f *and any two distinct messages* $\mathcal{M} = \{M^1, M^2\}$, *we define the structure graph* $\mathcal{G}^f(\mathcal{M})$ *as follows:*
$\mathcal{G}^f(\mathcal{M}) \triangleq (V, E, L)$, *where* $V = \{[[i]] : 0 \le i \le \tau\}$, $E = \{([[i-1]]', [[i]]) : 1 \le i \le \tau\}$, *and* $L((u, v)) = \{M_i : [[i-1]]' = u \text{ and } [[i]] = v\}$ *is an edge-labeling.*

Definition 4. *For the computation of* M^1, *the sequence* 0, $([[0]]', [[1]])$, $[[1]]$, $([[1]]', [[2]])$, $\dots$, $[[l_1]]$ *of alternating vertices and edges is called an* M^1*-walk. (An* M^2*-walk is defined analogously).*

Let (V_i, E_i, L_i) be the graph obtained after processing only the first i out of τ blocks of $\mathcal{M}$. We define a collision event as follows.

Definition 5. $(i, [[i]])$ *is an* f*-collision if* $[[i]] < i$ *and* $M_i \notin L_{i-1}([[i-1]]', [[i]])$.

Note that the last condition on M_i implies that collision occurred due to parallel edges with the same message label is not considered.