

# 1 Approach

In this article, we name the MAC scheme adopted in Cost-Effective Tag Design[] CETD-MAC.

## 1.1 Specification of CETD-MAC

In this section, we depict the design details of CETD-MAC scheme. At the beginning we express notations required to understand CETD-MAC scheme.

**Notations** Let  $\{0,1\}^n$  be the set of all n-bit binary strings. The set of all binary string is expressed as  $\{0,1\}^*$ . For a string  $X \in \{0,1\}^n$ ,  $|X|$  is its length in bits, and  $|X|_l = \lceil |X|/l \rceil$  is the length of X in l-bit blocks. Let  $0^l$  and  $1^l$  denote bit strings of all zeros and all ones. For a bit string X and an integer l that  $|X| \geq l$ ,  $\text{msb}_l(X)$  denotes the most significant l bits(left most l bits) of X and  $\text{lsb}_l(X)$  for least significant l bits(right most l bits) of X. For two bit string X and Y, we denote  $X\|Y$  or  $XY$  as the their concatenation. For bit string X whose length in bits is multiple of integer l, we denote X parted into l-bit sub-strings as  $X = (X[1]X[2] \dots X[n])_l$ , where  $X[1], X[2], \dots, X[n] \in \{0,1\}^l$ . The number of bits in a string of X is denoted as  $\text{len}(X)$ .

The block cipher encryption of a string X with a secret key K is denoted as  $E_K(X)$ .  $E_K(X)$  expresses the String mapping of  $\{0,1\}^n \rightarrow \{0,1\}^n$  where n is the  $\text{len}(X)$  and  $\text{len}(\text{output})$ .

**CETD-MAC Scheme** CETD-MAC scheme can be expressed as  $\text{tag} = \text{CETD-MAC}(M, \text{nonce-input})$ . The input arguments of CETD-MAC are message M and a tuple named nonce-input. The tuple nonce-input is the concatenation of the memory address of M, a counter and a random number, denoted as  $\text{nonce-input} = (\text{address}\|\text{counter}\|\text{random})$ . The length of nonce-input,  $\text{len}(\text{nonce-input})$ , is identical to the length of input to block cipher  $E_K(X)$ . The output of CETD-MAC is named tag, whose length is optional. We use Sub-BLK-No to express the value of  $\text{len}(M)/\text{len}(\text{tag})$ . One preliminary of CETD-MAC is that Sub-BLK-No should be no less than 2 to assure that swapping stage can work and 0s will be concatenated to the leftmost of input message when Sub-BLK-No is not integer.

We follow the defination of CETD-MAC in [?].

- Message is splitted to sub-blocks each of whose length is tag length
- The first stage is several rounds of bit-segment swapping. For each round , two sub-blocks are randomly chosen. Introduce shuffle paremeter  $V[i]$
- The output sub-blocks from swapping stage, named X blocks, are sent to block-level rotate shifting stage. Each X block is rotate shifted for several bits. Introduce shift parameter  $S[i]$
- The output sub-blocks from shifting stage, named Y blocks, are XORed to form the final tag

## 1.2 Tag Collision Under Replay Attack

In this section, we depict the weaknesses we found in CETD-MAC under replay attack. These weaknesses in mechanism let the CETD-MAC to produce tags with high collision probability with some chosen inputs. The tags with highly collision probability ease the adversary's attack.

**Security Definitions of MAC Schemes under Replay Attacks** Our security definition of MAC schemes under replay attack follows the definition under chosen-message attack. An adversary is given access to a tag generation black-box(named oracle) and a message-tag pair verification oracle. The encryption key is maintained unchanged while the nonce value is updated for each calling of tag generation oracle no matter whether the value of input message collide with a value in old timestamp. The adversary can record the message-tag pairs in old time stamp to copies and replace a pair in new timestamp with a piece of copies. The forgery advantage  $F_{CETD-MAC}$  under replay attack is the probability that the adversary can get the verification oracle to accept a copy with old timestamp at a new time point.

Assume the copy with old timestamp from the adversary is expressed as  $(M1, T1)$  where M1 is the message and T1 is the corresponding tag, and the corresponding nonce is denoted as N1. When applying M1 to the tag generation at a new time point, we mark the corresponding tag T2 and nonce N2. If the adversary can succeed a replay attack, then T1 is identical to T2. Then  $F_{CETD-MAC}$  under replay attack can be denoted as a probability  $\Pr[T1 = T2 \mid \text{Message value is M1 \& N1, N2 are randomly generated}]$ .

According to the design of CETD-MAC, the tag can be expressed as  $Y[1] \oplus Y[2] \oplus \dots \oplus Y[x]$ , x is the number of input sub-blocks to CETD-MAC. If two tags T1 and T2 are identical, the following equation is met:  $Y_1[1] \oplus Y_1[2] \dots \oplus Y_1[x] \oplus Y_2[1] \oplus Y_2[2] \dots \oplus Y_2[x] = 0$ , where  $Y_1$  blocks produce T1 and  $Y_2$  blocks generate T2. This equation indicates that the behaviour of block-level rotate shifting stage effect the probability of tag collision directly. For easy understanding, we firstly analyze the case that shuffle stage does not work under replay attack, which means  $X_1[i]$  is identical to  $X_2[i]$  for any  $i \in [1, x]$ . The analysis of effect from shuffle stage on X blocks follows then.

As the nonce is generated by block cipher encryption  $E_K$ . Assume  $E_K$  performace like a ideal random value generator, for two distinct input nonce-input1 and nonce-input2, the corresponding nonce values N1 and N2 should be random and the probability that N1 is identical to N2 should be  $1/2^n$ , where n is  $\text{len}(N1)$ .

### 1.2.1 The Block-level Pattern Collision

Firstly, we treat  $Y1$  blocks and  $Y2$  blocks as two block sequences, which means in a Y block sequence, each block has a unique index. The definition of block sequence equality is expressed below:

**Definition 1.1.** *Block Sequence Equality: Two block sequences  $X1$  and  $X2$  are Block Sequence Equality if  $X1[i]$  is identical to  $X2[i]$  for any  $i \in [1, x]$ .*

The treatment of Y blocks as sequences indicate the first case leading to tag collision. Assume Y1 sequence( $Y1[1]||Y1[2] \dots Y1[i] \dots ||Y1[x]$ ) produces tag T1 and Y2 sequence( $Y2[1]||Y2[2] \dots Y2[i] \dots ||Y2[x]$ ) generates T2. If  $Y1[i]$  is identical to  $Y2[i]$  for any  $i \in [1, x]$ , then T1 is identical to T2.

In this section, we express the first weakness in the rotate shifting stage in CETD-MAC.

**Definition 1.2.** *If the a input block of rotate shifting stage  $X1[i]$  from  $X1$  sequence is identical to the block  $X2[i]$  with same index from  $X2$  sequence, the following two properties exist:*

- *The probability that corresponding output blocks of rotate shifting stage,  $Y1[i]$  and  $Y2[i]$ , are identical can be expressed with the length of pattern in  $X1[i]$ .*
- *The more identical X block pairs( $X1[i], X2[i]$ ) are formed by repeating pattern, the higher probability of identical block pairs in Y sequences ( $Y1[i], Y2[i]$ ) exist, then the higher probability of tag collision( $T1=T2$ )*

These two properties mean if the adversary conduct replay attack on the memory frame whose message part is formed with pattern, then the probability of successful forgery, expressed as  $F_{CETC-MAC}$ , is high.

### 1.2.2 The Set-level Pattern Collision

Secondly we treat Y1 blocks and Y2 blocks are two sets allowing identical elements. We give the definition of block set equality below:

**Definition 1.3.** *Block Set Equality: Two block sets  $X1$  and  $X2$  are Block Set Equality if the following properties are met:*

- *$X1$  and  $X2$  have same number of elements*
- *The number of distinct value of elements in  $X1$  is identical to the number in  $X2$*

If Y1 and Y2 sets are set-level identically, then T1 is identical to T2. Based on this fact, we find another weakness of shifting stage leading to tag collision under replay attack:

**Definition 1.4.** *For two identical block sequences  $X1$  and  $X2$ , if each block in  $X1$  can be formed by rotate shifting another element in  $X1$ , then the output sets Y1 and Y2 have high probability to form Block Set Equality.*

This weakness means the high succeed probability of replay attack on memory frames whose data part is consist of concatenated blocks shifted from a same base block.

### 1.2.3 The Behaviour of X Blocks from Shuffling Stage

## 1.3 A Solution to Eliminate the Collision

In this section we provide an modification of original CETD-MAC. Our approach does not require additional component or information to the original CETD-MAC. We then prove that the optimized CETD-MAC can fix the weakness and protect input message from replay attack.

**Question** Using only shuffle and shift with random nonce, is it possible to ensure the range of tag to  $2^n$ ? If not, why?

## 2 The Collision Probability of Y Set

### 2.1 Preliminaries

**The Equality of Two Sets** Assume there are two sets, marked as X\_A and X\_B, having same NO. of elements and the number is M. Each element is given an index from 1 to M. The definition of Identical Index Equality is expressed here:

**Definition 2.1.** *Identical Index Equality(IIE): For any  $i \in [1, M]$ ,  $X_A[i] = X_B[i]$ .*

The definition of Cross Index Equality is expressed here:

**Definition 2.2.** *Cross Index Equality(CIE): For any  $i, j \in [1, M]$ ,  $X_A[i] = X_B[j]$ , where  $i \neq j$*

Any two sets that have same number of elements can be splitted into several sub-groups and each sub-groups contains several  $(X_A[i], X_B[i])$  pairs. For two specific sets, the number of sub-groups various from 1 to M. We give the definition of set level equality of two sets with same number of elements:

**Definition 2.3.** *Set Level Equality(SLE): Each sub-group have one of the two kinds of Equality.*

For two sets X\_A and X\_B, if there is no way of splitting that at least one sub-group is of CIE and there are at least one way of split that all sub-sets are of IIE, then such X\_A and X\_B are IIE only Sets. The definition of CIE only Sets is given in same way.

#### 2.1.1 Replay Attack and Y Set Collision

**The Message Sets in CETD Under Replay Attack** The Cost-Effective Tag Design can be expressed as the following procedure:

- The data is split into blocks. The block length is tag length. The block set is marked as D, and the No. of blocks is M.

- The content of blocks are shuffled. The output set of shuffle is marked as X.
- The content of each block is rotate shifted. The output set of rotate shift is marked as Y.
- All the blocks in Y is XORed to form the tag T.
- The control parameters is retrived from the segment of nonce N. The parameter set of shuffle is marked as S. The set for rotate shift is R.

In replay attack, the adversary replaces a data-tag pair on the memory with a pair copied from the same address at an old time point. That means for the two pairs at different time point, the two message block sets  $D_A$  and  $D_B$  and related tags  $T_a$  and  $T_b$  are identical respectively, while the nonce  $N_A$  and  $N_B$  are randomly generated and the equality is unpredictable if their generator is of high quality. The shifting bits parameter segment on the nonce  $R_A$  and  $R_B$ , are randomly generated. We can see that if the Y set for the old time point, marked as  $Y_A$ , is SLE to the Y set for new time point, marked as  $Y_B$ ,  $T_a = T_b$ . Based on the concept of set equality, the message block sets in the generation of  $T_a$  and verification of  $T_b$  meet the following properties if the replay attack succeeds:

- $D_A$  and  $D_B$  are IIE.
- If the shuffle stage does not work.  $X_A = D_A$ ,  $X_B = D_B$
- $Y_A$  and  $Y_B$  are SLE

In this scenario, the probability of a successful attack can be expressed as the equation 1.1:

**Definition 2.4.**  $Pr[Successful\ Replay\ Attack] = Pr[T_a = T_b \mid (D_A = D_B) \ \& \ (R_A \text{ and } R_B \text{ are random})] = Pr[Y_A = Y_B \mid (D_A = D_B) \ \& \ (R_A \text{ and } R_B \text{ are random})]$

*$D_A$  and  $D_B$  sets are of Identical Index Equality.  $Y_A$  and  $Y_B$  is of Set Level Equality.*

For easy understanding, we assume the shuffle stage does not work at first, which means  $D_A = X_A = D_B = X_B$ (IIE). Hence the Y set is the output of rotate shifting stage in CETD, we will analyze the properties of block rotate shifting and the cases of input sets that can result Y set collision(SLE).

### 2.1.2 Rotate Shifting Introduction

Unlike logical shifting and arithmetic shifting, rotate shifting behaves like whirling a wheel. The empty position in a message block shifted is filled by the bits shifted out. Figure 1 express the concept of rotate shift.

We can refer that the result of rotate shifting a message block depends on the value of block and the bits shifted. If two message blocks whose value are

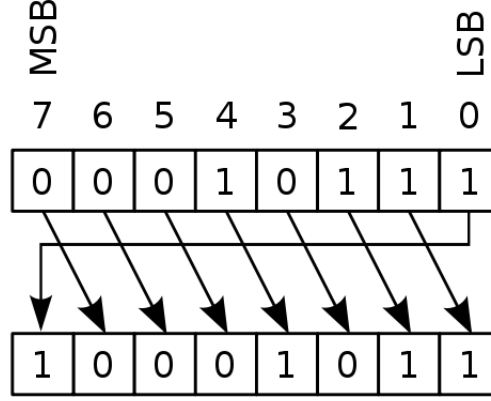


Figure 1: The Concept of Rotate Shifting(Right)

identical are shifted with distinct bits, the result blocks may be identical. That means when shifting bits fixed, the mapping from X blocks to the shifted result blocks is not injection. This case is expressed in Figure 2(a).

For two distinct X blocks, however, their result blocks may be identical when their shifting bits are distinct. This case is expressed in Figure 2(b). When the shifting bits of two message blocks are identical, the equality of result blocks is same as the equality of their input blocks.

## 2.2 Rotate Shifting and Y Set Collision

Assume the shuffle stage does not work, then  $D\_A = X\_A = D\_B = X\_B$ , which means the following properties exist in  $X\_A$  and  $X\_B$ :

- $X\_A[i] = X\_B[i]$  for all  $i \in [1, M]$ , M is the number of elements
- The equality of elements in a X set is uncertain.

Then  $X\_A$  and  $X\_B$  are of Identical Index Equality. All the analysis in this section is based on the assumption that shuffle stage does not work.

If such  $X\_A$  and  $X\_B$  result two Y sets of Set Level Equality, the related tag  $T_a = T_b$ .

Figure 3 and Figure 4 express the examples of X sets that lead to SLE Y sets with distinct R sets. In this paper, we analyze the cases of the input set pair ( $X\_A$  and  $X\_B$ ) of shifting stage in CETD under replay attack that can lead to Y set pair collision(SLE).

### 2.2.1 Case of X Sets Resulting Y Set Collision

If  $R\_A$  and  $R\_B$  are IIE, then for any two identical set  $X\_A$  and  $X\_B$ ,  $Y\_A$  and  $Y\_B$  are IIE. If there is at least one pair of blocks in  $R\_A$  and  $R\_B$ , marked as  $(R\_A[i], R\_B[i])$ , is distinct, then equality of  $Y\_A$  and  $Y\_B$  is uncertain.

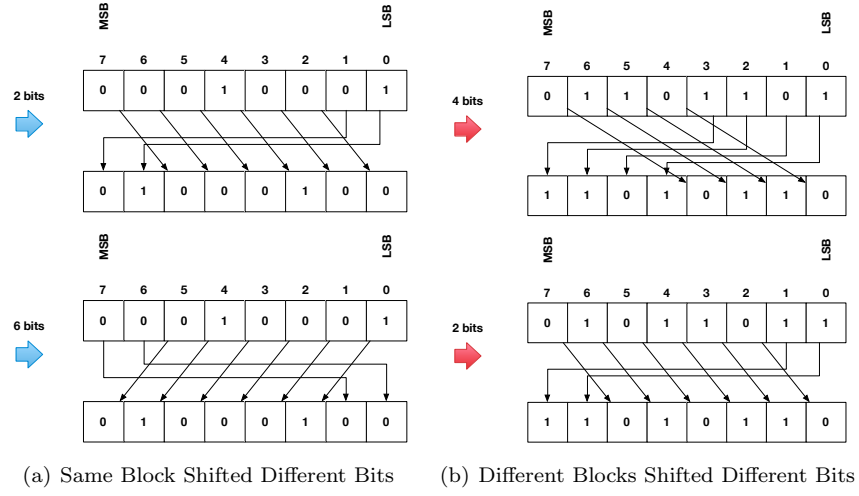


Figure 2: The Examples of Y Block Collision

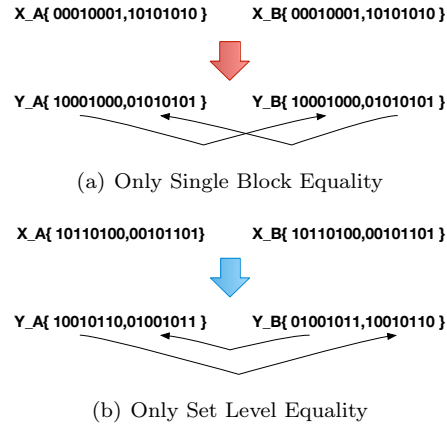


Figure 3: X Set Pairs with Only One Type of Y Equality

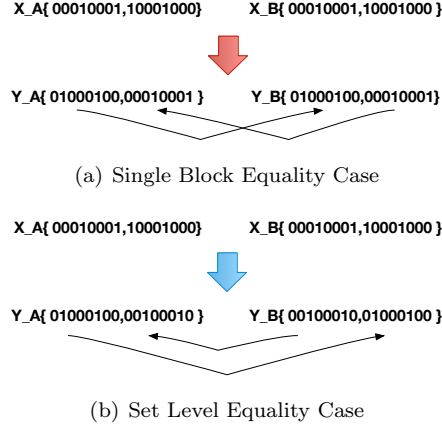


Figure 4: A X Set Pair with Two Types of Y Equality

Assume the number of distinct block pairs in  $R\_A$  and  $R\_B$  is  $R_d$ . For these  $R$  block pairs, their related  $X\_A$  and  $X\_B$  blocks formed a sub-group, marked as  $X\_A_s$  and  $X\_B_s$ , and this sub-group is IIE. The related sub-group in  $Y\_A$  is marked as  $Y\_A_s$  and the one in  $Y\_B$  for  $Y\_B_s$ .

In this part, we discuss the equality case of  $Y\_A_s$  and  $Y\_B_s$ .

**Identical Index Equality Case** If  $Y\_A_s$  and  $Y\_B_s$  are IIE, then for all  $i \in [1, R_d]$  the following two properties are met:

- $Y\_A_s[i] = Y\_B_s[i]$
- $R\_A_s[i] \neq Y\_B_s[i]$
- $X\_A_s[i] = X\_B_s[i]$

Figure 2(a) provides a instance of IIE Y sets. The identical block pairs  $(X\_A_s[i], X\_B_s[i])$  that lead to IIE  $Y\_A_s$  and  $Y\_B_s$  have the property in Theorem 1.1:

**Theorem 2.1.** Assume  $X\_A[i]$  and  $X\_B[i]$  are two identical block from two sub-group  $X\_A_s$  and  $X\_B_s$  with same index  $i$ .  $X\_A[i]$  and  $X\_B[i]$  have same number of bits  $N=2^n$ . The result of rotate shifting  $X\_A[i]$  and  $X\_B[i]$  with distinct shifting bit parameter  $R\_A[i]$  and  $R\_B[i]$  are marked as  $Y\_A[i]$  and  $Y\_B[i]$ .

Then  $Y\_A[i]$  and  $Y\_B[i]$  can be identical only when  $X\_A[i]$  and  $X\_B[i]$  are formed by repeating a binary pattern  $P$ , which is a binary segment. The length of this pattern  $P$  can be expressed as:

$$P\_L = 2^p, p \in [0, n-1]$$

Base on Theorem 1.1, we got the following corollaries:

**Corollary 2.2.** If a pattern is not formed by a pattern with short length, we call it a distinct pattern of length  $P\_L$ . The No. of distinct patterns with length  $P\_L = 2^p$  is  $2^{2^p} - 2^{2^{p-1}}$



**Corollary 2.3.** *Assume the length of a block is  $N=2^n$ , then the No. of all distinct patterns with all possible length is  $2^{N/2}$*

**Cross Index Equality Case** If  $Y\_A_s$  and  $Y\_B_s$  are CIE, then for all distinct  $(i,j)$  that  $i,j \in [1, R_d]$  the following two properties are met:

- $Y\_A_s[i] = Y\_B_s[j]$
- $R\_A_s[i] \neq Y\_B_s[i]$
- $X\_A_s[i] = X\_B_s[i]$

Figure 2(b) provides an instance of CIE Y sets. The blocks in  $X\_A_s$  and  $X\_B_s$  that lead to CIE  $Y\_A_s$  and  $Y\_B_s$  have the property in Theorem 1.2:

**Theorem 2.4.** *If  $Y\_A_s$  and  $Y\_B_s$  are CIE, each element in  $X\_A_s$  can be formed by rotate shifting another element in  $X\_A_s$ .*

If each element in  $X\_A_s$  is formed by a pattern can be formed by rotate shifting another element in  $X\_B_s$  then the related  $Y\_A_s$  and  $Y\_B_s$  can be either IIE or CIE or SLE. The proof of Theorem 1.1 and Theorem 1.2 can be refered in appendix.

## 2.3 The Probability of Y Set Collision

Hence the shift bit parameter set  $R\_A$  and  $R\_B$  are randomly generated, the probability of Y set collision for each input case is determined by the number of specific combinations of  $R\_A$  and  $R\_B$ . The computation of  $\Pr[Y\_A = Y\_B]$  is based on this idea. The following properties are met by any cases of D sets and R sets in replay attack:

- $D\_A[i] = D\_B[i]$  for all  $i \in [1, M]$
- If  $R\_A[i] = R\_B[i]$  and  $X\_A[i] = X\_B[i]$ , then  $Y\_A[i] = Y\_B[i]$ .

### 2.3.1 Identical Index Equality Case

When  $R\_A[i]$  and  $R\_B[i]$  are distinct for some  $i \in [1, M]$ , the  $Y\_A[i]$  and  $Y\_B[i]$  blocks can form an IIE sub-group if each  $X\_A[i]$  and  $X\_B[i]$  block is formed by a pattern.

For two identical X blocks  $X\_A[i]$  and  $X\_B[i]$ , the probability that  $Y\_A[i] = Y\_B[i]$  when  $R\_A[i]$  and  $R\_B[i]$  are randomly generated is marked as  $\Pr[Y \text{ block collision}]$ .

$\Pr[Y \text{ block collision}]$  is expressed in Theorem 1.3:

**Theorem 2.5.** *If the No. of bits of each block in  $X\_A[i]$ - $X\_B[i]$  pair is  $N=2^n$ , the pattern length  $P_l=2^p$  where  $p \in [0, n-1]$ . The pattern contains no internal sub-pattern, then :*

$$\Pr[Y \text{ block collision}] = 1/2^p$$

### 2.3.2 Cross Index Equality Case

Assume two identical sets  $X\_A$  and  $X\_B$  satisfy the following properties:

- Each element can be formed by rotate shifting any other element in the set
- None of the elements is formed by pattern.

As each element in the  $X$  set can be formed by rotate shifting a base block, we call such  $X$  set Same Base Block(SBB)  $X$  sets, short for SBB  $X$  sets. If none of the element is formed by a pattern ,we call such  $X$  set SBB only sets.

From Corollary 1.3 we know that for a block that is  $N$  bits long, the No. of possible values that is formed by a pattern is  $2^{N/2}$ , which also means the No. of distinct values that is not formed by a pattern is  $2^{N/2}$ . Assume  $D\_A$  and  $D\_B$  are randomly generated, the probability of generating SBB only  $D$  set is expressed:

**Definition 2.5.**  $Pr[SBB\ only\ D(X)] =$

$$2^{n/2} * N^{M-1} / (2^N)^M$$

Using SBB only  $X\_A$  and  $X\_B$ , the probability of CIE of  $Y\_A$  and  $Y\_B$  is marked as  $Pr[CIE\ Y\ Sets]$  and expressed as the following way:

**Definition 2.6.**  $Pr[CIE\ Y\ Sets] =$

$$\prod_{i=1, j=1}^M Pr[Y\_A[i] = Y\_B[j]]$$

where  $M$  is the No. of elements in a set,  $i, j \in [1, M]$  and  $i \neq j$

For two randomly generated  $R\_A$  and  $R\_B$ , the probability that  $Y\_A$  and  $Y\_B$  are CIE when  $X\_A$  and  $X\_B$  are SBB only is expressed in the following theorem:

**Theorem 2.6.** Assume  $X\_A$  and  $X\_B$  are SBB only,  $R\_A$  and  $R\_B$  are randomly generated, then:

$Pr[CIE\ Y\ Sets]:$

$$\left( \sum_{K=1}^{\min(N, M)} \binom{N}{K} * (M! / v_1! v_2! v_3! \dots v_k!)^2 \right) / (N^M)^2$$

### 2.3.3 The Intersection Case

If the blocks in  $X$  set is formed by rotate shifting a base block and this base block is formed by a pattern, then such  $X$  set pair  $X\_A$  and  $X\_B$  can result either IIE or CIE  $Y$  set pairs.

Assume the  $R\_A$  and  $R\_B$  are randomly generated, the probability that  $Y\_A$  and  $Y\_B$  are SLE is expressed in the following theorem:

**Theorem 2.7.** Assume  $X_A$  and  $X_B$  are both SBB and IPL,  $R_A$  and  $R_B$  are randomly generated, then:

$Pr[SLE \ Y \ Sets] =$

$$\left( \sum_{K=1}^{\min(N,M)} \binom{N}{K} * (M! / v_1! v_2! v_3! \dots v_k!)^2 \right) / (N^M)^2$$

The proof of theorems is expressed in the appendix.

## 2.4 A General Case of Sets in CETD

**The general Case of D sets** During the replay attack, the  $D_A$  and  $D_B$  sets are IIE. Any D set can be split into sub-groups of blocks in the following way:

$$D = \{\{IIEonly\}, \{CIEonly\}, \{Intersection\}, \{Noregularity\}\} \quad (1)$$

The No. of the elements in these sub-groups is marked as  $M_{IIE}$ ,  $M_{CIE}$ ,  $M_{inter}$ ,  $M_{non}$ . The following equations are met:

$$M_{IIE}, M_{CIE}, M_{inter}, M_{non} \in [0, M] \quad (2)$$

$$M_{IIE} + M_{CIE} + M_{inter} + M_{non} = M \quad (3)$$

**The General Case of X Sets** In general case, the shuffle stage can change the split of blocks. For a D set that contains regularity, shuffle can reduce the No. of regularity sub-groups and the No. of elements in each group. The properties of no regularity sub-group is analyzed in the next section.

## A Proof of Pattern Introduction

### A.1 Proof of Theorem 1.1

This part proves that if two identical block  $X_A[i]$  and  $X_B[i]$  are shifted different bits and the result blocks remain identical,  $X_A[i]$  is formed by pattern. The pattern length and  $\delta = |R_A[i] - R_B[i]|$  has such correlation:

- If  $\Delta = P \cdot L$  then  $Y_A[i] = Y_B[i]$  where  $P \cdot L$  the length of pattern

Assume the length of  $X_A[i]$  is  $N$  bits, where  $N = 2^n$ . When  $X_A[i]$  is formed by a pattern whose length is  $P \cdot L = 2^p$  bits, then the domain of  $Y_A[i]$  has  $P \cdot L$  distinct values. There are  $N / P \cdot L$  distinct  $R_A[i]$  values that shift  $X_A[i]$  to a  $Y_A[i]$ .

## A.2 Proof of Theorem 1.2

If two distinct X blocks result two identical Y block. Assume the shift bits parameter blocks are  $R\_A[i]$  and  $R\_B[i]$ . Then  $X\_A[i]$  can be formed by rotate shifting  $Y\_A[i]$  for  $(N-R\_A[i]) \bmod N$  bits.  $X\_B[i]$  can be formed for  $(N-R\_B[i]) \bmod N$  bits.  $X\_B[i]$  can be formed by rotate shifting  $X\_A[i]$  for  $(R\_A[i] + N - R\_B[i]) \bmod N$  bits. Theorem 1.2 proved.

## B Proof of Probability Computation

### B.1 Proof of Theorem 1.3

For identical each block pair  $X\_A[i]$  and  $X\_B[i]$ , the No. of combination of  $R\_A[i]$  and  $R\_B[i] = N*N$ , where  $N$  is the length a X block. If  $X\_A[i]$  is formed by pattern, then  $Y\_A[i]=Y\_B[i]$  if  $|R\_A[i]-R\_B[i]| = P\_L * K$ , where  $P\_L$  is the pattern length and  $K$  is a positive integer. Then for two random  $R\_A[i]$  and  $R\_B[i]$ ,  $\Pr[Y\_A[i]=Y\_B[i] \mid X\_A[i] = X\_B[i] \ \& \ \text{pattern} = P\_L]$  (shoft for  $\Pr[Y\_A[i] = Y\_B[i]]$ ) can be expressed in the following way:

- $\Pr[Y\_A[i]=Y\_B[i]] = N * (N/P\_L)/(N*N) = 1/P\_L$

### B.2 Proof of Theorem 1.4

From Theorem 1.1 we can see that if a X block is formed by a pattern, then the No. of distinct values in the range of Y block is  $P\_L$ . While the domain of a R block has  $N$  distinct values, then for a block X, there are  $N/P\_L$  distinct R values that lead to a Y value.

If a X block is not formed by a pattern, then for each distinct R value, there is a distinct Y value. That means for a given X set that none of the blocks is formed by a pattern, each R set will lead to a distinct Y set. The map between R set and Y set is bijection.

When each R block is randomly generated, the possible combination of R sets is  $N^M$ . That means for a given X set, the No. of possible Y set is  $N^M$ .

Assume  $X\_A$  and  $X\_B$  are IIE and each block can be formed by rotate shifting another block in the set. On the other hand, none of blocks is formed by a pattern. Then the sub-group of blocks in  $X\_A$  and  $X\_B$  can form CIE with specific  $R\_A$  and  $R\_B$  set.

Assume set  $Y\_A$  and  $Y\_B$  are SLE, then  $Y\_B$  is a permutation of  $Y\_A$ . This concept can be modeled in the following way:

- Assume the  $M$  elements in  $Y\_A$  contain  $K$  distinct values. The No. of the elements that have each value are marked as  $v_1, v_2 \dots v_k$ .
- IF the elements in  $Y\_B$  are a permutation of the elements in  $Y\_A$ , then  $Y\_A$  and  $Y\_B$  are SLE.

Based on the basic concept in Combinatorics, the No. of a permutation of  $Y\_A$  that contains  $K$  distinct value can be expressed in the following equation:

**Definition B.1.** *No. of Permutation with K distinct Values:*

$$\binom{M}{v1} * \binom{M-v1}{v2} * \binom{M-v1-v2}{v3} \cdots * \binom{vk}{vk} = M!/v1!v2!v3! \cdots vk!$$

As any one of the elements in X set can be formed by rotate shifting another element, then the No. of distinct values of M elements, K, various from 1 to M. If Y\_A has K distinct values, the No. of possible combination of these K elements in Y\_A can be expressed as Com\_Y\_A:

**Definition B.2.** *Com\_Y\_A:*

$$M!/v1!v2!v3! \cdots vk!$$

For each case of Y\_A in Com\_Y\_A, the No. of Y\_B to form SLE is also Com\_Y\_A. Then if Y\_A has K distinct values, Pr[CIE Y sets] can be expressed as:

**Definition B.3.** *Pr[CIE Y Sets]:*

$$\binom{N}{K} * (M!/v1!v2!v3! \cdots vk!)^2 / (N^M)^2$$

If R\_A and R\_B are randomly generated, then the value of K various from 1 to M. We do th sum to get the expression in Theorem 1.4

**Definition B.4.** *Pr[CIE Y Sets]:*

$$\left( \sum_{K=1}^{\min(N,M)} \binom{N}{K} * (M!/v1!v2!v3! \cdots vk!)^2 \right) / (N^M)^2$$

### B.3 Proof of Theorem 1.5

If the element contains both the properties of CIE and IIE, then the following properties:

- For two distinct sets R\_A and R\_B, Y\_A and Y\_B can be IIE
- For each value of a Y block, there are N/P\_L distinct R block that can shift X to this Y. That means for a distinct Y set, the No. of R sets is not one.
- For two distinct sets R\_A and R\_B, Y\_A and Y\_B can be CIE

Base on the Theorem 1.1, if the X block is formed by pattern, then the No. of values in the range of Y is P\_L, base on theorem 1.4, the Pr[SLE Y sets] can be expressed as:

**Definition B.5.** *Pr[SLE Y Sets]:*

$$\binom{P_L}{K} * (M!/v1!v2!v3! \cdots vk!)^2 / (P_L^M)^2$$

If  $R_A$  and  $R_B$  are randomly generated, then the value of  $K$  varies from 1 to  $M$ . We do the sum to get the expression in Theorem 1.5

**Definition B.6.** *Pr[SLE Y Sets]:*

$$\left( \sum_{K=1}^{\min(P, L, M)} \binom{P-L}{K} * (M! / v_1! v_2! v_3! \dots v_k!)^2 \right) / (P-L)^M)^2$$

## C Experiments and Results