

# New Integrated proof Method on Iterated Hash Structure and New Structures

Duo Lei

Department of Science, National University of Defense Technology,  
Changsha, China  
Duoduo1ei@gmail.com

**Abstract.** A secure hash structure in Random Oracle Model may not be a secure model in true design. In this paper, we give an integrated proof method on security proof of iterated hash structure. Based on the proof method, we can distinguish the security of Merkle-Damgård structure, wide-pipe hash, double-pipe hash and 3c hash and know the requirement of true design on compression function, and give a new recommend structure. At last, we give new hash structure, MAC structure, encryption model, which use same block cipher round function and key schedule algorithm, the security proofs on those structures are given.<sup>1</sup>

## 1 Introduction

Most of hash functions are iterated hash function and most of compression functions were iterated by Merkle-Damgård[18, 31] construction(noted M-D construction in this paper) with constant IV[47]. Since the MD5 and SHA1 were attacked by [9][57][58], more and more attentions had been paid on hash function.

### 1.1 Introduction

Generally, the security proof on hash structure are based on the Random Oracle Model, which make an assumption of compression function with Random Oracle Model. Some structures are secure in Random Oracle Model, which may be not a good structure in true design. For example, let M-D hash  $H^M : \{0, 1\}^{\kappa^*} \times \{0, 1\}^n \rightarrow \{0, 1\}^n$ ,  $z = H^M(m, IV)$ ,  $IV \in \{0, 1\}^n$ ,  $m \in \{0, 1\}^{\kappa^*}$ ,  $z \in \{0, 1\}^n$ , the compression function  $F : \{0, 1\}^\kappa \times \{0, 1\}^n \rightarrow \{0, 1\}^n$ ,  $x_h \in \{0, 1\}^n$ ,  $x_m \in \{0, 1\}^\kappa$ ,  $y \in \{0, 1\}^n$ ,  $y = F(x_m, x_h)$ , in hash iteration  $x_h$  is chaining value. The pseudo random properties of compression function can not prevent things happen like that: there may be exist message  $m_0 = \mathbf{m}_0 \parallel \dots \parallel \mathbf{m}_0$  with  $H^M(\mathbf{m}_0 \parallel \dots \parallel \mathbf{m}_0, x) \equiv z_0$ , for most  $x \in \{0, 1\}^n$  with height probability ( we call such property as cluster), if exist such value, then it means for any message  $m'$  the equation  $H^M(m_0 \parallel m', IV) = H^M(m_0, H^M(m', IV)) = z_0$  being hold. The probability of existence of  $m_0$  may be very small or the value  $m_0$  be too long to be used in

<sup>1</sup> Revised: July 31, 2006

practical attack, and also the complexity of finding such value is  $\mathcal{O}(t \cdot 2^n)$ ,  $t$  is some value of  $t \in [1, 2^n]$ . However, if a true design exist such value and the  $t$  is a small value, at least which means a failure of the hash function in theory level, more importantly, the cluster properties are influenced by both linear and nonlinear components of compression function.

The proves based on Random Oracle Model can not distinguish exist cluster or not, because, in Random Oracle Model, we have  $P_{Y|X_m=x_m} = \frac{1}{2^n}$ , in true design only permutation has such property, if the compression function is permutation then it exist inverse function, at least the hash will not immune to meet in middle attack on preimage[41]. The best selection of compression function should be one way permutation, the one way permutation is difficult to build in true design, we always select one way function, if we select a one way function as compression function then exist  $x_{m_0}$  with  $P_{Y|X_m=x_{m_0}} > \frac{1}{2^n}$ , that means repeat the message block  $m_0 = x_{m_0}$  may result in cluster, if exist cluster, we can append such value at end of any given message build collision.

In the attacks on hash function, the attackers can not assume that they are lucky enough to select some 'advantage value', by which, they are more easy to build an attack. But the designers should assume the attackers are lucky enough to select some advantage value for searching and build collision.

In original discussion about hash function based on sequence of games, the advantage of attacks are an average advantage of success, two summaries of that part were given by Victor[52] and Bellare and Rogaway[6]. In paper[52], the author given some historical remark about "Hybird arguments" and sequence of games.

Generally, the illustration of advantage is an average of success, most of proofs based on game-technology are based on assumption of that, the based function is pseudo random function or pseudo random permutation, the advantage can not tell us how many collision exist and what is the true complexity of attack when the attacker is lucky enough to select some 'advantage value' for search.

The M-D structure is not immune to extend attack, fix point attack and multi-collision attack, moreover, some slight weakness in compression function (like some special plaintexts can make collision) may result in failure of hash function, so some revised structures have been given, include wide-pipe hash and double-pipe hash, but the proofs were based on immune against known attack.

The main ideas of the recent attacks on hash functions are differential attack[8] and were known in block ciphers years ago, which means the attacks against block ciphers and hash functions are similar. The design criteria of block ciphers have received much attention and had an interesting framework and also block cipher cryptanalysis techniques were partially used against hash functions[7, 21]. More and more attentions have been paid on hash functions be possible to designed by the same technology as block ciphers with same principles and design criteria[7].

## 1.2 The Motivation on Security Proof

In this paper, we give a new evaluate method to evaluate the security of structure, the new method give the maximum advantage of success based on the conditional probability of whole structure, the relationship between conditional probability and the maximum advantage of success on attack can give the information about existence of collision. Since the distribution of conditional probability is a bijective transformation with hash function, if the maximum conditional probability is close to uniform distribution, we can assume the structure is secure, where the compression function is assumed as Black Box Model, the assumption of Black Box Model means the chosen plaintext attack and adaptive chosen plaintext with same complexity and the compression function is immune against those attack, and also the distribution of input and output of compression function is independent, so we prefer the compression function is designed as block cipher design principle.

If the Random Oracle Model analysis method considers the whole structure is Random Oracle or not, when the compression function is Random Oracle, our criteria considers, if the compression function is not a Random Oracle Model, whole structure is how far away from a Random Oracle Model.

In this paper, firstly, we give the definitions of conditional probability of compression function and that of whole structure. In fact the maximum of conditional probability is the existence of maximum collision and maximum preimage, if the upper bound of the conditional probability is given, then the upper bound of existence of collision and preimage are given, and also if the upper bound is near to uniform distribution, we can assume the structure is secure structure, and what is needed in true design are that compression function can be designed as black box model.

Then, we give some different definition of advantage or sequences of games, by which the maximum advantage of success of the preimage and collision attack are required, in fact, the maximum advantage has bound of that of conditional probability, so if the conditional probability is given, then the maximum advantage can be gotten, then we give close relation among the conditional probability, maximum advantage of attack and existence of collision and preimage, if the whole hash function can be seen as black box model, that means the security of structure against preimage and collision attack become clear.

After the revision of some basic definitions, we build a toy hash to give an intuitional illustration of weakness in M-D construction and it's conditional probability. In toy hash, there may exist cluster, if the compression function exist  $x_{m_0}$ , with  $P_{Y|X_m=x_{m_0}} > \frac{1}{2^n}$ , even though the compression function is a APN[21] function. We get conclusion of that, if the compression function  $y = F(x_{m_0}, x_h)$  is not a permutation, it is hard to give the proof of a M-D constructed hash function not existing cluster, for such properties are influenced by both linear and nonlinear components of compression function. The cluster is also illustrated by conditional probability, where if there exist cluster, then the conditional probability  $P_{z|M=m_0}(z_0)$  are very big, the worst condition is that value equals 1. Then we give a theorem about the graph and conditional probability. The clus-

ter property was discussed in [46, 41], in this paper, we give a more systematical discussion.

Based on previous discussion, we reanalysis the known hash structures, include M-D construction[31, 18], wide-Pipe hash[40], double-pipe hash[40], and 3C[25], the conclusions are that, if the compression functions has property of exist  $x_{m_0}$  with  $y = F(x_{m_0}, x_h)$  not a permutation, then cluster may exist in the previous three structured hash functions. But the upper bound of conditional probability of 3c structure is equals with that of compression function, which implies if the compression function is pseudo random function, 3c structure can be seen as pseudo random function, where we assume the compression function is black box model. Based on conditional probability of structures, we give an ideal model of hash structure called ideal-pipe hash, which has properties of that the upper bound of conditional probability is equals with that of compression function, the 3c structure can be seen as example of ideal-pipe hash.

After discussion about conditional properties of the structures, we give the maximum advantages of compression function and the iterated structure based on game-technology, the conclusions of that part are that: in view point of collision resistance, the M-D structure is secure structure, if the compression function  $y = F(x_m, x_h)$  with  $\forall x_{m_0}, y = F(x_{m_0}, x_h)$  being permutation, but not immune against meet in middle attack on preimage, when the some of that compression function  $y = F(x_{m_0}, x_h)$  are not permutation, in true design, the proof of not existing cluster on whole hash should be given. Let  $g(x)$  be not invertible black box model, if the the compression function  $y = F(x_m, x_h)$  with  $\forall x_{m_0}, y = F(x_{m_0}, x_h)$  being permutation, the wide-pipe hash is secure structure and the upper bound of conditional probability is better than 3c structure, if there exist  $x_{m_0}$  with  $y = F(x_{m_0}, x_h)$  is not permutation, more discussion should be given which may exist cluster same as M-D construction. The double-pipe hash is same as M-D construction, it is not better than wide-pipe hash with  $F : I_{2n} \times I_{2n} \rightarrow I_{2n}$ . The maximum advantages of 3c hash and ideal-pipe hash are also given, which are good structures for build hash, if only the compression function is black box model and close to pseudo random function.

In the previous discussion, we assume the compression function is black box model, which means the compression function is immune against chosen plaintext attack and adaptive chosen plaintext attack, and the distribution of outputs and inputs are independent, that properties are required in block cipher design and have plenty of results, so we prefer design a hash structure based on block cipher's structure and round function and key schedule algorithm. We find the Feistel structure is good structure of building hash. Then a new structure is given based on Feistel structure, the security proofs are also given. In fact, we given new structures that can build hash, block cipher(itself is block cipher), MAC and Block cipher encrypt model based on same compression function and key schedule algorithm.

And also we give a brief discussion about the relation of ROM, pseudo random function and conditional probability.

### 1.3 Motivation of Feistel Hash

The possibility of building dedicate hash function based on Feistel structure is discussed in this paper. Feistel structure is known as a good structure for building block cipher, fixed the left half  $n$  bits input with zero, output the right half byte of last round output, the Feistel structure become an  $n$ -bit to  $n$ -bit transformation instead of a  $2n$ -bit to  $2n$ -bit transformation, and is not invertible, we call it FL-structure(Feistel Like Structure). If the round function of such Feistel construction is selected same as design criteria of Feistel block cipher, the new construction inherent the almost all properties of Feistel block cipher except invertible, the security of Feistel block cipher has been studied long time and no weakness are founded in structure itself, we deem the new construction is a good way to build hash compression function. In this paper, we discussed the possibility of building whole hash function based on FL structure, and give a recommend model, which is named F-Hash, we give a proof of the security of such hash function and such hash function is immune against all known attacks, where most of the securities are based on the securities of based block cipher. Then with same round function and key schedule, we can also build a hash, MAC and block cipher encrypt mode called FBC mode. The figure illustration of Feistel compression function, F-Hash, F-MAC and FBC mode are given in Fig8, Fig9, Fig10 and Fig11, respectively. In this paper, the padding is padding zero at end of message, so we don't consider the padding, and the figure is drawn similar as the MAC Alred[17].

Luby and Rackoff[30] introduced a model that permits the assessment of the security of some block cipher constructions, in their discussion, only the high-level structure is considered, while the lower-level operations are replaced by random functions. Using such methods, Patarin[33–37] given the security proof of Feistel structure. Piret given proof of the round function with random permutation[38, 39], similar conclusions were also given in the paper of Vaudenay[55, 56]. But all the discussions were based on assumption of round functions are independent pseudo random functions. In this paper, we make a assumption of exist a Feistel block cipher is Black Box Model, then exist FL-Function with Black Box Model. The previous discussions paid more attention on condition of same key, the design of hash function requires pay same attention on the influence of different key encrypt the same plaintext.

The aim of discussions about security of Feistel structure were giving proof of no ways to distinguish the fix keyed cipher from random permutation, which were based on assumption that compression function is pseudo random function, but that is not hold in true design. And in Feistel compression function, the key schedule algorithm should be considered separately. At end of this paper, we give some discussion about the security of round function and key schedule algorithm, but this part is far more hard than design of block cipher. So the security of true F-Hash is need more discussion.

## 2 Basic Notation, Definition and Theorem

### 2.1 Basic Notation

The paper include the knowledge of block cipher, Hash function and complexity theory, we try to not change the original notation, then many notations are included in this paper.

- $I_n$  denotes the set  $\{0, 1\}^n$ .
- for  $a, b \in I_n$ ,  $a \| b \in I_{2n}$ ,  $|a| = n$ ;
- $k^{(1)}, k^{(2)}, \dots$  are round keys of key, the key schedule algorithm is denoted as  $\varphi(k)$ ,  $k^{(0)} = k$ ,  $k^{(i)} = \varphi(k^{(i-1)})$ ;
- $f : I_\kappa \times I_n \rightarrow I_n$ : round function of block cipher<sup>2</sup>,  $y = f(k^{(i)}, x)$ ,  $k^{(i)} \in I_\kappa$ ,  $x \in I_n$ , round function is also denoted  $f_{k^{(i)}}$ .
- Let  $\mathbf{m}^{(t)} \stackrel{def}{=} \mathbf{m} \| \dots \| \mathbf{m}$ , where  $|\mathbf{m}^{(t)}| = t \cdot 2^n$ ,  $\mathbf{m} \in I_n$ .
- message<sup>3</sup>  $m \in I_{n^*}$ ,  $\mathbf{m}_i$  is message block with  $\mathbf{m}_i \in I_n$ ,  $m = \mathbf{m}_* \| \mathbf{m}_{* - 1} \| \dots \| \mathbf{m}_1$ , then  $\mathbf{m}_* \subseteq m$ . A selected  $m$  is denoted  $m_i \in I_{n \cdot t}$ ,  $t \geq 1$ ;
- $\circ$  is the composition of function.
- $x', x, y', y \in I_n$ ,  $\Psi(f_k)(x' \| x) = y' \| y \stackrel{def}{\Leftrightarrow} \begin{cases} y' = x \\ y = x' \oplus f_k(x) \end{cases}$
- $(x)^R$  and  $(x)^L$ : the right and left  $n$  bits of binary sequence  $x$ , respectively;
- $\tilde{0}$ :  $n$ -bit binary, all bits are 0;
- $E^{Fe} : I_n \times I_{2n} \rightarrow I_{2n}$  is Feistel structured block cipher with round function  $f$ ,  $E^{Fe}(k, x' \| x) \triangleq \Psi^R(f)(x' \| x) = \Psi(f_{k^{(R)}}) \circ \Psi(f_{k^{(R-1)}}) \circ \dots \circ \Psi(f_{k^{(1)}})(x' \| x)$ ;
- $E^{Sp} : I_n \times I_n \rightarrow I_n$ , SPN structured block cipher with round function  $f$ ,  $E^{Sp}(k, x) = f_{k^{(R')}} \circ f_{k^{(R'-1)}} \circ \dots \circ f_{k^{(1)}}(x)$
- $F_c : I_n \times I_n \rightarrow I_n$  is Feistel-Like Structured function with round function  $f$ ,  $F_c(k, x) = (E^{Fe}(k, \tilde{0} \| x))^R$ ;
- $F_{c-1}$ : Feistel-Structured function with one round fewer than  $F_c$ ;
- $x, x', y, y', x_h, x_m \in I_n$ ;
- $X_m, X_h, M, X, K, Y, Z, \tilde{Z}$ : Random variables;
- $E^{-1}$ : The inverse of  $E$ , where  $E$  is a permutation;
- $y = F(x_m, x_h)$ : the hash compression function, where  $x_h$  is chaining value,  $x_m$  is message block;
- $z = H(m, x)$ : the hash function, where  $m$  is message and  $x$  is initial value;
- $z = H^M(m, x)$ : iterated hash with M-D construction, if  $m = \mathbf{m}_t \| \dots \| \mathbf{m}_1$ , then  $z = F(\mathbf{m}_t, F(\mathbf{m}_{t-1}, \dots F(\mathbf{m}_1, x) \dots))$  with compression function  $y = F(x_m, x_h)$ ;
- Message Padding: adding zero at the end of Message.
- $[1, n]$  be the set of  $\{1, 2, 3, \dots, n\}$ ;
- $z \triangleq y_{\{a_1, \dots, a_t\}}$  be binary with  $t$  bits and the  $i$ th bit of  $z$  is  $a_i$ th bit of  $y$ ;

<sup>2</sup> Since the paper has too many notations, to make things simple, we assume the key length  $\kappa = n$ , that is not required in inner iteration procedure, in last round iteration if is required, we can using padding to make the input with required length.

<sup>3</sup> When message block is used as key, the message block length be  $\kappa$ .

*Remark 1.* In iterated hash function  $H(m, x)$ , we consider the  $x \in I_n$ , because we can redefine a hash function  $H'(m, x') \stackrel{def}{=} H(m||x, IV) = H(m, F(x, IV))$ , in selected hash IV is constant, more discussion is given in following section.

Let  $G : I_\kappa \times I_l \rightarrow I_n, y = G(m, x)$  then

- $G(\cdot, x) : I_\kappa \rightarrow I_n, y = G_{m_0}(\cdot, x) \stackrel{def}{=} G(m_0, x)$
- $G(m, \cdot) : I_l \rightarrow I_n, y = G_{x_0}(m, \cdot) \stackrel{def}{=} G(m, x_0)$
- $\{(y, m, x)\} \stackrel{def}{=} \{(y, m, x) | m \in I_\kappa, x \in I_l, y \in I_n\}$ ;
- $\{x_0\} \stackrel{def}{=} \{x | x \in \{x\}, x = x_0\}$ ;
- $\Lambda_n \subset I^n$ ;
- $\{(y, m, x)\}^G \stackrel{def}{=} \{(y, m, x) | (y, m, x) \in \{(y, m, x)\}, G(m, x) = y\}$ ;
- $\{(y_0, m, x)\}^G \stackrel{def}{=} \{(y_0, m, x) | (y, m, x) \in \{(y, m, x)\}^G, y = y_0\}$ ;
- $\{(y, m, x_0)\}^G_{x_0 \in \Lambda} \stackrel{def}{=} \{(y, m, x) | (y, m, x_0) \in \{(y, m, x)\}^G, x_0 \in \Lambda\}$
- $\{\{(y_0, m, x)\}^G\}_{y_0 \in \Lambda} \stackrel{def}{=} \bigcup_{y_0 \in \Lambda} \{\{(y_0, m, x)\}^G\}$
- $\Omega \stackrel{def}{=} \{(y, m, x)\}, \Omega^G = \{(y, m, x)\}^G, \omega^G \in \Omega^G$ ;
- $\#\{\cdot\}$ : the count of  $\omega$ , where  $\omega \in \{\cdot\}$
- $T_G \stackrel{def}{=} \max_{y_0, x_0} \#\{(y_0, m, x_0)\}^G$ ,
- $S_G \stackrel{def}{=} \max_{y_0, m_0} \#\{(y_0, m_0, x)\}^G$ ,
- $R_G \stackrel{def}{=} \max_{y_0} \#\{(y_0, m, x)\}^G$ .

We make a assumption of  $\frac{1}{0} = 0$ .

## 2.2 Probability Theory

The notations of the probability in the paper are followed that of PHD paper of Christian Cachin[12].

A discrete random variable  $X$  is a mapping from the sample space  $\Omega$  to an alphabet  $\mathcal{X}$ .  $X$  assigns a value  $x \in \mathcal{X}$  to each elementary event in the  $\Omega$  and the probability distribution of  $X$  is the function

$$P_X : \mathcal{X} \rightarrow \mathfrak{R} : x \mapsto P_X(x) = P[X = x] = \sum_{\omega \in \Omega : X(\omega) = x} P[\omega]$$

$$P_X(x) = P[x \stackrel{\$}{\leftarrow} \mathcal{X}; \omega \leftarrow \Omega : X(\omega) = x]$$

If the conditioning event involves another random variable  $Y$  defined on the same sample space, the conditional probability distribution of  $Y$  given that  $X$  takes on a value  $x$  is:

$$P_{Y|X=x}(y) = \frac{P_{XY}(x, y)}{P_X(x)}$$

whenever  $P_X(x)$  is positive.

**Theorem 1 (Derived Probability).** Let function  $y = G(m, x)$ ,  $G : I_{n \cdot t} \times I_n \rightarrow I_n$ ,  $t \in \mathbf{N}$ , let the distributions of independent random variable  $M$  and  $X$  are  $P_X(x)$  and  $P_M(m)$ , let function  $\chi_{G(m,x)}(y)$  is defined as that

$$\chi_{G(m,x)}(y) \stackrel{\text{def}}{=} \begin{cases} 1 & y = G(m, x) \\ 0 & y \neq G(m, x) \end{cases}$$

the random variable  $Y$ 's distribution can be derived from  $X$  and  $M$  by:

$$\begin{aligned} P_Y(y) &\stackrel{\text{def}}{=} P_Y(y = G(M, X)) \\ &= \sum_{x \in I_n} \sum_{m \in I_{n \cdot t}} P_{XM}(x, m) \chi_{G(m,x)}(y) \\ &= \sum_{x \in I_n} \sum_{m \in I_{n \cdot t}} P_X(x) P_M(m) \chi_{G(m,x)}(y) \end{aligned}$$

we call the probability of  $Y$  is derived probability of  $M$  and  $X$ .

In fact:

1.  $P_X(x) \geq 0, P_M(m) \geq 0 \Rightarrow P_Y(y) \geq 0$ ;
- 2.

$$\begin{aligned} &\sum_{y \in I_n} \sum_{x \in I_n} \sum_{m \in I_{n \cdot t}} P_X(x) P_M(m) \cdot \chi_{G(m,x)}(y) \\ &= \sum_{x \in I_n} \sum_{m \in I_{n \cdot t}} P_X(x) P_M(m) \cdot \sum_{y \in I_n} \chi_{G(m,x)}(y) \\ &= \sum_{x \in I_n} \sum_{m \in I_{n \cdot t}} P_X(x) P_M(m) 1 = 1 \Rightarrow \sum_{y \in I_n} P_Y(y) = 1 \end{aligned}$$

For any  $y \in I_n$ , if does not exist  $P_Y(y)$ , then we have  $P_Y(y) \stackrel{\text{def}}{=} 0$ .

**Theorem 2 (Conditional Probability).** The conditional probabilities are given as follows:

1.  $P_{Y|M=m_0}(y_0) = \sum_{x \in I_n} P_X(x) \chi_{G(y_0, m_0, x)}$ ;
2.  $P_{Y|X=x_0}(y_0) = \sum_{m \in I_{n \cdot t}} P_M(m) \chi_{G(y_0, m, x_0)}$ ;
3.  $P_{Y|X=x_0, M=m_0}(y_0) = \chi_{G(y_0, m_0, x_0)}$ ;
4.  $P_{\dot{Y}|M=m_0}(y_0) = \sum_{x \in I_n} \frac{1}{2^n} \chi_{G(y_0, m_0, x)}$ ;
5.  $P_{\dot{Y}|X=x_0}(y_0) = \sum_{m \in I_{n \cdot t}} \frac{1}{2^{n \cdot t}} \chi_{G(y_0, m, x_0)}$ ;
6.  $P_{\dot{Y}|X=x_0, M=m_0}(y_0) = \chi_{G(y_0, m_0, x_0)}$ ;

If  $X$  and  $M$  are uniformly distributed, which means  $P_M(m) = \frac{1}{2^{n \cdot t}}$ ,  $P_X(x) = \frac{1}{2^n}$ , when  $X$  and  $M$  are uniformly distributed, we use notation of  $\dot{P}_Y(y)$ , that is also hold in conditional probability.

**Theorem 3.** The relation between conditional probabilities and existence of such value is given as follows:

1.  $P_{\dot{Y}}(y_0) = \frac{\#\{(y_0, m, x)\}^G}{2^n \cdot 2^{n \cdot t}};$
2.  $P_{\dot{Y}|M=m_0}(y_0) = \frac{\#\{(y_0, m_0, x)\}^G}{2^n};$
3.  $P_{\dot{Y}|X=x_0}(y_0) = \frac{\#\{(y_0, m, x_0)\}^G}{2^{n \cdot t}};$
4.  $P_{\dot{Y}|X=x_0, M=m_0}(y_0) = \#\{(y_0, m_0, x_0)\}^G.$

**Corollary 1.**

$$\#\{(y_0, m, x)\}^G \leq \max\{\#\{(y_0, m_0, x)\}^G \cdot 2^{n \cdot t}, \#\{(y_0, m, x_0)\}^G \cdot 2^n\}$$

**2.3 Definitions Based on Sequence Games**

The notations about advantages are from paper[52] and paper[6].

*Remark 2.* The theorem3 give the relation between the conditional probability and existence of collision and preimage.

In describing probabilistic processes, we write:  $x \stackrel{\$}{\leftarrow} X$  to denote the action of assigning to the variable  $x$  a value sampled according to the distribution  $X$ . If  $S$  is a finite set, we simply write  $s \stackrel{\$}{\leftarrow} S$  to denote assignment to  $s$  of an element sampled from the uniform distribution on  $S$ . We shall write  $Pr[x_1 \stackrel{\$}{\leftarrow} X_1, x_2 \stackrel{\$}{\leftarrow} X_2(x_1), \dots, x_n \stackrel{\$}{\leftarrow} X_n(x_1, \dots, x_{n-1}) : \phi(x_1, \dots, x_n)]$  to denote the probability that when  $x_1$  is drawn from a certain distribution  $X_1$ , and  $x_2$  is drawn from a certain distribution  $X_2(x_1)$ , possibly depending on the particular choice of  $x_1$ , and so on, all the way to  $x_n$ , the predicate  $\phi(x_1, \dots, x_n)$  is true. We allow the predicate  $\phi$  to involve the execution of probabilistic algorithms.

**Lemma 1 (Difference Lemma[52]).** *Let  $A, B, F$  be events defined in some probability distribution, and suppose that  $A \wedge \neg F \Leftrightarrow B \wedge \neg F$ . Then  $|Pr[A] - Pr[B]| \leq Pr[F]$ .*

In this section we give some new definitions about attacks on hash function, the reasons are that:

1. If  $y = F(x_m, x_h)$  is block cipher, where the  $x_m$  is the key, the  $Adv_F^{Inv}(A) \stackrel{def}{=} Pr[y_0 \stackrel{\$}{\leftarrow} I_n; x_m \leftarrow I_n : F(x_m, \cdot) = y_0] = 1$ , since we can get  $x_h$  by  $x_h = F_{x_m}^{-1}(y_0)$ , if  $Adv_F^{Inv}(A) \stackrel{def}{=} Pr[y_0, x_{h_0} \stackrel{\$}{\leftarrow} I_n; x_m \leftarrow I_n : F(x_m, \cdot) = y_0]$ , then  $Adv_F^{Inv}(A) = \frac{1}{2^n}$ , we have to distinguish the two case.
2. We want to distinguish the difference between the existence of some value and the advantage of finding such value, more precisely, if block cipher  $y = F(x_m, x_h)$ , where  $x_m$  is the key, then  $Pr[y_0, x_{h_0} \stackrel{\$}{\leftarrow} I_n; x_m \leftarrow I_n : F(x_m, x_{h_0}) = y_0] = \frac{1}{2^n}$ ,  $Pr[y_0, x_{m_0} \stackrel{\$}{\leftarrow} I_n; x_h \leftarrow I_n : F(x_{m_0}, x_h) = y_0] = 1$ , but  $\#\{(y_0, x_m, x_{h_0})\}^F \approx \#\{(y_0, x_{m_0}, x_h)\}^F$ .

3. We want to distinguish the difference between experiments of select  $y \stackrel{\$}{\leftarrow} S$  of  $Adv(A) = Pr[y \stackrel{\$}{\leftarrow} S; x_m \leftarrow A : y = F(x_m, x_h)]$  and  $\max_{y_0} Pr[y_0 \in I_n, x_m \leftarrow A : y_0 = F(x_m, x_h)]$ .
4. The previous definition illustrate an average of success, we want to illustration the condition of that Adversary is lucky enough to success with maximum probability;
5. We also want to give a precise bound about the advantage of finding some value and existence of such value and also we want to give a more precise searching space or probability space. The paper[52] also suggest the games should be defined on a common probability space.

The definition:  $Adv_F^{Inv}(A) \stackrel{def}{=} Pr[y_0 \stackrel{\$}{\leftarrow} I_n; x_m, x_h \leftarrow I_n : F(x_m, \cdot) = y_0]$  can be given based on games, we also use games to define our objects and to describe our work, that is based on definition given in [26].

$Game(Inv, A, F)$

$y_0 \stackrel{\$}{\leftarrow} I_n$

$A(y_0) \rightarrow (x_m, x_h)$

$A$  wins if  $F(x_m, x_h) = y_0$ .

**Definition 1.** *The definitions about the maximum advantage of  $A$  in finding Preimage and Collision of function  $H$  and compression function  $F$  are as follows,  $Adv(q) \stackrel{def}{=} \max_q \{Adv(A)\}$ , where the maximum is taken over adversaries that ask at most  $q$  queries,  $q < \min\{2^{n-1}, 2^{\kappa-1}\}$ , write  $\tilde{Adv}(A) \stackrel{def}{=} \max\{Adv(A)\}$ , where the maximum is get the luckiest adversary's advantage, if  $F$  is invertible with  $F^{-1}$ , then  $A$  can ask queries of  $F$  and  $F^{-1}$ , the search space is the whole space.*

1. **(Fress Start) Pseudo Preimage Attack:**

$$\tilde{Adv}_F^{Pre}(A) = \max_{y_0} Pr[y_0 \in I_n; \omega \leftarrow A^F : \omega \in \{(y_0, x_m, x_h)\}^F]$$

$$\tilde{Adv}_H^{Pre}(A) = \max_{z_0} Pr[z_0 \in I_n; \omega \leftarrow A^{F,H} : \omega \in \{(z_0, m, x)\}^H]$$

2. **(Fixed Start) Preimage Attack:**

$$\tilde{Adv}_F^{FixP}(A) = \max_{y_0, x_{h_0}} Pr[y_0 \in I_n, x_{h_0} \in I_n; \omega \leftarrow A^F : \omega \in \{(y_0, x_m, x_{h_0})\}^F]$$

$$\tilde{Adv}_H^{FixP}(A) = \max_{y_0, x_0} Pr[y_0 \in I_n, x_0 \in I_n; \omega \leftarrow A^{F,H} : \omega \in \{(z_0, m, x_0)\}^H]$$

3. **(Free Start) Pseudo Collision Attack:**

$$\tilde{Adv}_F^{Coll}(A) = \max_{y_0} Pr[\omega, \omega' \leftarrow A^F : \omega, \omega' \in \sigma, \sigma \in \{\{(y_0, x_m, x_h)\}^F\}_{y_0 \in I_n}]$$

$$\tilde{Adv}_H^{Coll}(A) = \max_{y_0} Pr[\omega, \omega' \leftarrow A^{F,H} : \omega, \omega' \in \sigma, \sigma \in \{\{(z_0, m, x)\}^H\}_{z_0 \in I_n}]$$

4. **(Fixed Start) Collision Attack:**

$$\begin{aligned}\tilde{Adv}_F^{FixC}(A) &= \max_{y_0, x_{h_0}} Pr[x_{h_0} \in I_n; \omega, \omega' \leftarrow A^F : \\ &\quad \omega, \omega' \in \sigma, \sigma \in \{\{(y_0, x_m, x_{h_0})\}^F\}_{y_0 \in I_n}]\end{aligned}$$

$$\begin{aligned}\tilde{Adv}_H^{FixC}(A) &= \max_{y_0, x_0} Pr[x_0 \in I_n; \omega, \omega' \leftarrow A^{F,H} : \\ &\quad \omega, \omega' \in \sigma, \sigma \in \{\{(z_0, m, x_0)\}^H\}_{z_0 \in I_n}]\end{aligned}$$

The definitions can be also given as follows, an example is given:

$$\tilde{Adv}_F^{Pre}(A) = \max_{y_0} Pr[y_0 \in I_n; \omega \leftarrow A^F : \omega \in \{(y_0, x_m, x_h)\}^F]$$

$$\begin{aligned}Game(Adv_F^{Pre}(A), A, F, y_0) \\ A(y_0) \rightarrow (x_m, x_h) \\ A \text{ wins if } F(x_m, x_h) = y_0.\end{aligned}$$

$$\tilde{Adv}_F^{Pre}(q) = \max_{y_0} Pr[y_0 \in I_n; \omega \leftarrow A^F : \omega \in \{(y_0, x_m, x_h)\}^F]$$

$$\begin{aligned}Game(Adv_F^{Pre}(q), A, F, y_0) \\ \text{For } i = 1, \dots, t \text{ do :} \\ A(y_0) \rightarrow (x_{m_i}, x_{h_i}) \\ A \text{ wins if } \exists i \text{ st. } F(x_{m_i}, x_{h_i}) = y_0.\end{aligned}$$

**Definition 2 (Black Box Model).**  $G : I_\kappa \times I_n \rightarrow I_n$  is a Black Box Model, if the probabilities of success of Game0 and Game1 are same, and  $G$  is immune against those attacks. where  $q \leq 2^{-\frac{\kappa}{2}}$ :

$$\begin{aligned}Game0(A, F, y_0, q) \\ \text{For } i = 1, \dots, t \text{ do :} \\ A(y_0) \rightarrow (x_{m_i}, x_{h_i}) \\ A \text{ wins if } \exists i \text{ st. } F(x_{m_i}, x_{h_i}) = y_0. \\ Game1(A, F, y_0, q) \\ Q \leftarrow \emptyset \\ \text{For } i = 1, \dots, t \text{ do :} \\ A(y_0, Q) \rightarrow (x_{m_i}, x_{h_i}) \\ Q \leftarrow Q \cup (F(x_{m_i}, x_{h_i}), x_{m_i}, x_{h_i}) \\ A \text{ wins if } \exists i \text{ st. } F(x_{m_i}, x_{h_i}) = y_0.\end{aligned}$$

If no special statement is given, the Black Box Model means  $G(m, \cdot)$  and  $G(\cdot, x)$  are not invertible.

*Remark 3.* The definition is also same with  $y = G(x)$ .

**Definition 3 (Random Oracle Model).** A fixed-size Random Oracle is a function  $f : I_a \rightarrow I_b$ , chosen uniformly at random from the set of all such functions.

**Theorem 4.** For Hash function  $H$  and Hash compression function  $F$ :

$$\tilde{Adv}_F^{FixC}(A) \leq \tilde{Adv}_F^{Coll}(A)$$

$$\tilde{Adv}_H^{FixC}(A) \leq \tilde{Adv}_H^{Coll}(A)$$

$$\tilde{Adv}_H^{FixC}(A) \leq \tilde{Adv}_H^{Pre}(A)$$

**Theorem 5.** If  $y = F(x_m, x_h)$  is black box model, then

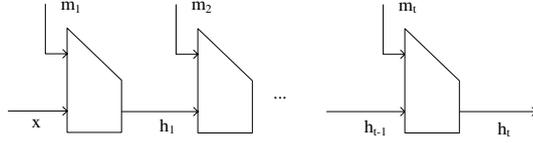
$$\tilde{Adv}_F^{FixP}(A) = \max_{y_0, x_{h_0}} P_{Y|X_h=x_{h_0}}(y_0).$$

## 2.4 The Definitions of Known Structures

Let  $F : I_\kappa \times I_n \rightarrow I_n$ ,  $y = F(x_m, x_h)$ ,  $x_h \in I_n$ ,  $y \in I_n$ ,  $G : I_n \rightarrow I_\omega$ ,  $\bar{y} = G(x_h)$ ,  $\bar{y} \in I_\omega$ ,  $m \in I_{\kappa,*}$ ,  $m = m_* \| \dots \| m_1$ , then the definition are given as follows:

**Definition 4 (M-D hash).** Let  $y = F(x_m, x_h)$  is a compression function of hash function  $H^M$ , the  $H^M : I_{\kappa,*} \times I_n \rightarrow I_n$  with M-D construction is defined as (figure illustration is given in Fig1):

$$z = H^M(m, x) \stackrel{def}{=} F(m_*, F(m_{* - 1}, \dots (F(m_1, x)) \dots))$$



**Fig. 1.** The M-D Hash

**Definition 5 (Wide-Pipe Hash).** Let  $y = F(x_m, x_h)$ , the wide-pipe hash[40]  $H^W : I_{\kappa,*} \times I_n \rightarrow I_\omega$  is defined as (figure illustration is given in Fig2):

$$\tilde{z} = H^W(m, x) \stackrel{def}{=} G(H^M(m_* \| \dots \| m_1, x))$$

where  $z = H^M(m, x)$ ,  $\tilde{z} = G(z)$ .

**Definition 6 (Double-Pipe hash[40]).** Let  $\tilde{F} : I_\kappa \times I_{2n} \rightarrow I_{2n}$ ,  $\tilde{G} : I_\kappa \times I_{2n} \rightarrow I_n$  the double-pipe hash is defined as  $H^D : I_{\kappa,*} \times I_{2n} \rightarrow I_n$ :

$$\begin{aligned} \tilde{z} &= H^D(m, x' \| x) \stackrel{def}{=} \tilde{G}(m_*, H^M(m_{* - 1} \| \dots \| m_1, x' \| x)) \\ &= \tilde{G}(m_*, \tilde{F}(m_{* - 1}, \dots \tilde{F}(m_1, x' \| x) \dots)) \end{aligned}$$

where  $x, x' \in I_n$ ,  $y' \| y = \tilde{F}(x_m, x'_h \| x_h)$ ,  $y, y' \in I_n$ ,  $m \in I_{\kappa,*}$ ,  $m = m_* \| \dots \| m_1$ ,  $z' \| z = \tilde{F}(m_{* - 1}, \dots \tilde{F}(m_1, x' \| x) \dots)$ ,  $\tilde{z} = G(m_*, z' \| z) = H^D(m, x' \| x)$ ,  $z, z' \in I_n$ ,  $\tilde{z} \in I_n$ .

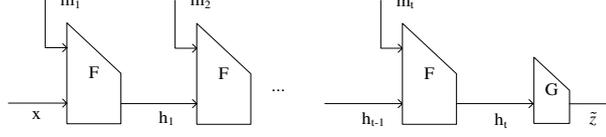


Fig. 2. The Wide-Pipe Hash

Remark 4. The Double-Pipe hash has some different from original design[40], we give a more general model, the original model is an example of that model.

Figure illustration is given in Fig3

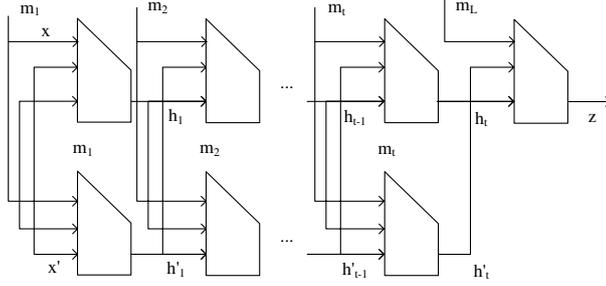


Fig. 3. The Double-Pipe Hash

Remark 5. In iterated hash function  $H(m, x)$ , we consider the  $x \in I_n$ , because we can redefine a hash function  $H'(m, x) \stackrel{def}{=} H(m \| x, IV) = H(m, F(x, IV))$ .

2.5 Definitions of Improved Structures

Definition 7 (Ideal-Pipe Hash). Let  $y = F(x_m, x_h)$ ,  $f : I_{\kappa.*} \times I_n \rightarrow I_n$ ,  $\bar{G} : I_n \times I_n \rightarrow I_n$ , the function  $F$  is hash compression function,  $\tilde{y} = f(m, x)$ ,  $\tilde{y} = \bar{G}(x_m, x_h)$ . the Ideal-Pipe hash structure is defined as  $H^I : I_{\kappa.*} \times I_n \rightarrow I_n$ :

$$H^I(m, x) \stackrel{def}{=} \bar{G}(f(m, x), H(m_* \| \dots \| m_1, x))$$

where  $z = F(m_*, \dots F(m_1, x) \dots)$ ,  $\tilde{z} = \bar{G}(f(m, x), z) = H^I(m, x)$ .

Remark 6. We give a improved M-D structure called Ideal-pipe hash, the motivation of this structure is to make the conditional probability  $P_{\tilde{Z}|M=m}^{\cdot}(z)$  approx  $2^{-n}$ , details are discussed in following section.

Similarly, we give a improved wide pipe hash, which were also given in paper [25] and that was before than me, so we also call it 3C as given in [25], although we give the structure separately.

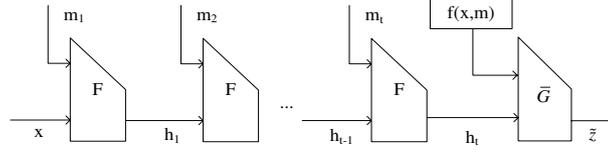


Fig. 4. The Ideal-Pipe Hash

**Definition 8 (3C Structure).** Let  $F : I_n \times I_n \rightarrow I_n$ , the improved wide pipe hash, 3c-hash structure [25]  $H^N : I_{n,*} \times I_n \rightarrow I_n$  is defined as (figure illustration is given in Fig5):

$$\tilde{z} = H^N(m, x) \stackrel{def}{=} F(h_* \oplus \dots \oplus h_1 \oplus h_0, H(m_* || \dots || m_1, x))$$

where  $x \in I_n$ ,  $y \in I_n$ ,  $y = F(x_m, x_h)$ ,  $m \in I_{\kappa,*}$ ,  $m = m_* || \dots || m_1$ ,  $h_i = F(m_i, h_{i-1})$ ,  $h_0 = x$ ,  $z = F(m_*, \dots F(m_1, x) \dots)$ ,  $\tilde{z} = H^N(m, x)$ .

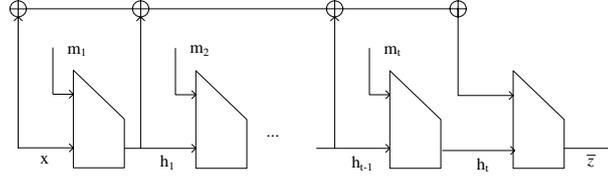


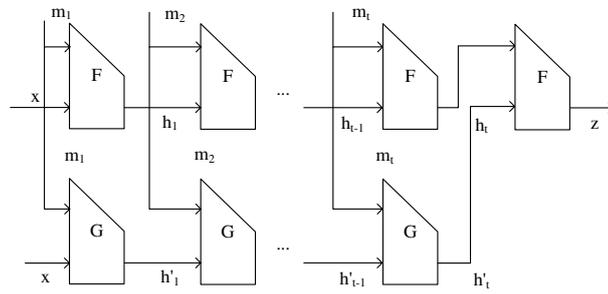
Fig. 5. 3C hash

The 3C hash can be considered as a improved structure of wide-pipe hash. In similar way, we can improve the structure of double-pipe hash, which is called improved double-pipe hash(The figure illustration is given in Fig.6).

**Definition 9 (Improved Double-Pipe Hash).** Let  $F : I_{\kappa} \times I_n \rightarrow I_n$ ,  $G : I_{\kappa} \times I_n \rightarrow I_n$  the improved double-pipe structure is defined as  $H^{ND} : I_{\kappa,*} \times I_n \rightarrow I_n$ :

$$H^{ND}(m, x) \stackrel{def}{=} F(H^G(m_* || \dots || m_1, x), H^F(m_* || \dots || m_1, x))$$

where  $x \in I_n$ ,  $m \in I_{\kappa,*}$ ,  $m = m_* || \dots || m_1$ ,  $H^F = F(m_*, \dots F(m_1, x) \dots)$ ,  $H^G = G(m_*, \dots G(m_1, x) \dots)$ .



**Fig. 6.** The Improved Double-Pipe Hash



Part I

Illustration Based on Graph  
Theory



### 3 Conditional Probability Based on Grapy Theory

In this section, we will discuss the design principle of compression function to prevent the conditional probability  $P_{Z|M=m_i}(z)$ , where  $m_i = \mathbf{m}_i || \dots || \mathbf{m}_i, \mathbf{m}_i \in I_n$  of hash function  $z = H^M(m_i, x)$ , not increase when the message length increased. Let  $G$  is a directed graph, the notations about graph  $G$  are from [19]. The main conclusions of this section are Assumption1, which implies we can only give the proof of permutation does not existing cluster, and Theorem6, which give the maximum conditional probability of M-D hash.

**Definition 10.** We call digraph  $G_{\mathbf{m}}$  is derived from function  $y = F_{\mathbf{m}}(x_m, x_h), m \in I_n$ , if the graph  $G_{\mathbf{m}}$  is build in following way: the vertices of graph  $G_{\mathbf{m}}$  is all  $v_{x_h} \in I_n$ , the directed edges are build in following ways: for  $\forall x_{h_0} \in I_n$ , if we have  $x'_h = F(m, x_{h_0})$  then draw an edge from  $v_{x_{h_0}}$  and directed at  $v_{x'_h}$ , the procedure is stopped until all  $F(m, x_h), \forall x_h \in I_n$  are computed and draw an directed edge from  $v_{x_h}$  to  $v_{F(m, x_h)}$ .

- $V = V(G_{\mathbf{m}})$  is the set of vertices of  $G_{\mathbf{m}}$ ;
- $E = E(G_{\mathbf{m}})$  is the set of edges of  $G_{\mathbf{m}}$ .
- $|G|$ : The number of vertices of graph  $G$  is its order;
- $\|G\|$ : The number of edges of Graph  $G$ ;
- $v_{x_{h_0}}$ : The vertex of  $G_{\mathbf{m}}$ , where the value of vertex of  $v_{x_{h_0}}$  is  $x_{h_0}$ , we also use  $x_{h_0}$  to denote the vertex directly;
- $(v_{x_{h_1}} v_{x_{h_2}})$ : The edge of  $G_{\mathbf{m}}$ ,  $(v_{x_{h_1}} v_{x_{h_2}}) \in E(G_{\mathbf{m}})$ , we also use  $(x_{h_1} x_{h_2})$  to denote the directed edge, where  $x_{h_2} = F(\mathbf{m}, x_{h_1})$ ;
- the vertex  $v_{x_{h_0}}$  is *incident* with an edge  $e$  if  $v_{x_{h_0}} \in e$ ; then  $e$  is an edge at  $v_{x_{h_0}}$ ;
- The indegree, the outdegree and degree of a vertex at Graph  $G_{\mathbf{m}}$  are defined as follows

$$d_{G_{\mathbf{m}}}^I(v_{x_{h_0}}) = \#\{x_h | x_{h_0} = F(\mathbf{m}, x_h), x_h \in I_n\}$$

$$d_{G_{\mathbf{m}}}^O(v_{x_{h_0}}) = \#\{x_h | x_h = F(\mathbf{m}, x_{h_0}), x_h \in I_n\}$$

$$d_{G_{\mathbf{m}}}(v_{x_{h_0}}) = d_{G_{\mathbf{m}}}^I(v_{x_{h_0}}) + d_{G_{\mathbf{m}}}^O(v_{x_{h_0}})$$

- $\delta(G_{\mathbf{m}}) \stackrel{def}{=} \min\{d_{G_{\mathbf{m}}}(v_{x_{h_0}}) | v_{x_{h_0}} \in V(G_{\mathbf{m}})\}$ ,  $\Delta(G_{\mathbf{m}}) \stackrel{def}{=} \max\{d_{G_{\mathbf{m}}}(v_{x_{h_0}}) | v_{x_{h_0}} \in V(G_{\mathbf{m}})\}$ ;
- $\mathcal{H}_i$  is connective subgraph of  $G_{\mathbf{m}}$ , the number of connective subgraph included in  $G_{\mathbf{m}}$  is denoted  $\Upsilon$ ;  $C_i$  to illustrate the cycles in subgraph  $\mathcal{H}_i$  (In each subgraph only exist a cycle);  $T_{ij}, j \in \mathbf{N}$  to illustrate the trees in  $\mathcal{H}_i - E(C_i)$ , the number of tree in  $\mathcal{H}_i - E(C_i)$  is denoted  $\tau_i$ ;
- Two digraphs  $G1, G2$  are isomorphic (symbol  $\cong$ ) if there is a bijection  $\phi : V(G1) \rightarrow V(G2) : (xy) \in E(G1) \Leftrightarrow (\phi(x)\phi(y)) \in E(G2)$ ;
- If  $G'$  is any subgraph with  $G' \subseteq G$ ,  $G - G'$  is obtained from  $G$  by deleting all the vertices in  $V(G')$  and their incident edges.
- If  $x_{h_0} = F(\mathbf{m}, x_{h_0})$  then the Graph is a loop, we still call it a cycle.

### 3.1 A Toy Hash

To illustrate the ideal more precisely, an example of hash function with M-D structure is given. Let us consider the compression function  $F : \{0, 1\}^4 \times \{0, 1\}^3 \rightarrow \{0, 1\}^3$ , the chaining value set  $A = I_3$ , the message set  $B = I_4$ , the value in  $A$  and  $B$  is illustrated in hexadecimal notation with a dot (010 is denoted  $\dot{2}$ ). The Toy Hash  $H^T(m, x)$  is a iterated hash function with M-D construction and the compression function is  $y = F(x_m, x_h)$ . Table1 is the table illustration of  $y = F(x_m, x_h)$ , both the count of row and column from 0, the first row is input value  $x_h$  and the first column is input value  $x_m$ , the value of column 1 and row 1 means  $\dot{1} = F(\dot{0}, \dot{0})$ . Fig7 is the Graph illustrations of  $G_i, i \in [0, 15], \dot{i} \in I_4$ . The compression functions  $y = F(\cdot, x_h)$  are designed with different properties, to illustate the different properties of design principle of compression function, first sub-figure  $G_{\dot{0}}$  is  $F_{\dot{0}}(\cdot, x_h)$ , which is permutation,  $F_{\dot{E}}(\cdot, x_h) = F_{\dot{0}}(\cdot, x_h) \oplus x_h$ ,  $F_{\dot{F}}(\cdot, x_h) = F_{\dot{0}}(\cdot, x_h) \wedge x_h$ ,  $F_{\dot{8}}(\cdot, x_h) = F_{\dot{2}}(\cdot, x_h) \oplus \dot{3}$ ,  $F_{\dot{9}}(\cdot, x_h) = F_{\dot{2}}(\cdot, x_h) \oplus \dot{4}$ ,  $F_{\dot{A}}(\cdot, x_h) = F_{\dot{2}}(\cdot, x_h \oplus \dot{5})$ ,  $F_{\dot{B}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{7}$ ,  $F_{\dot{C}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{3}$ ,  $F_{\dot{D}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{5}$ ,  $F_{\dot{0}}(\cdot, x_h)$  and  $F_{\dot{2}}(\cdot, x_h)$  are APN Function,  $F_{\dot{3}}(\cdot, x_h)$  is almost a linear function,  $F_{\dot{C}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{3}$ ,  $F_{\dot{D}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{5}$ .

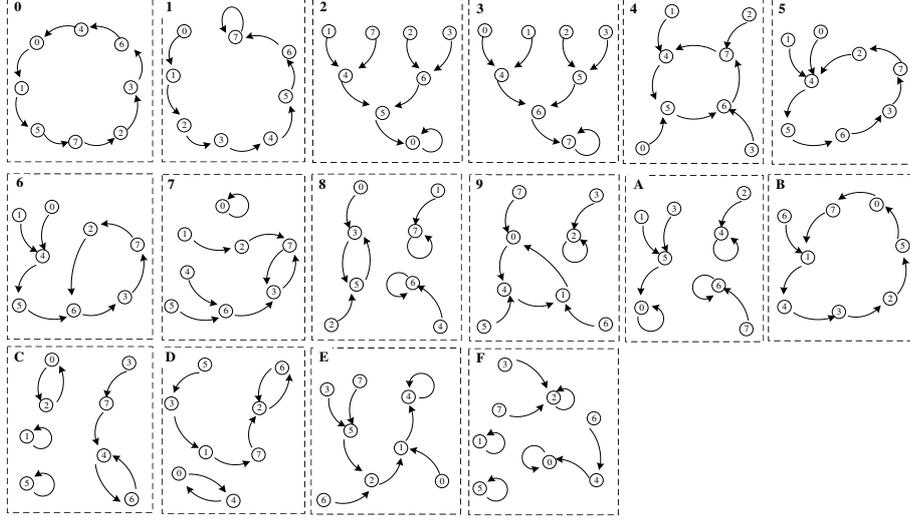
**Table 1.** The compression function of Toy Hash, where the row  $i$  is the output of  $F(i - 1, x_h)$ , the column  $j$  is output of  $F(x_m, j - 1)$ .

$x_m \backslash x_h$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$	$\dot{5}$	$\dot{6}$	$\dot{7}$
$\dot{0}$	$\dot{1}$	$\dot{5}$	$\dot{7}$	$\dot{2}$	$\dot{3}$	$\dot{6}$	$\dot{4}$	$\dot{0}$
$\dot{1}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$	$\dot{5}$	$\dot{6}$	$\dot{7}$	$\dot{7}$
$\dot{2}$	$\dot{0}$	$\dot{4}$	$\dot{6}$	$\dot{6}$	$\dot{5}$	$\dot{0}$	$\dot{5}$	$\dot{4}$
$\dot{3}$	$\dot{4}$	$\dot{4}$	$\dot{5}$	$\dot{5}$	$\dot{6}$	$\dot{6}$	$\dot{7}$	$\dot{6}$
$\dot{4}$	$\dot{5}$	$\dot{4}$	$\dot{7}$	$\dot{6}$	$\dot{5}$	$\dot{6}$	$\dot{7}$	$\dot{4}$
$\dot{5}$	$\dot{4}$	$\dot{4}$	$\dot{4}$	$\dot{7}$	$\dot{5}$	$\dot{6}$	$\dot{3}$	$\dot{2}$
$\dot{6}$	$\dot{4}$	$\dot{4}$	$\dot{6}$	$\dot{7}$	$\dot{5}$	$\dot{6}$	$\dot{3}$	$\dot{2}$
$\dot{7}$	$\dot{0}$	$\dot{2}$	$\dot{7}$	$\dot{7}$	$\dot{6}$	$\dot{6}$	$\dot{3}$	$\dot{3}$
$\dot{8}$	$\dot{3}$	$\dot{7}$	$\dot{5}$	$\dot{5}$	$\dot{6}$	$\dot{3}$	$\dot{6}$	$\dot{7}$
$\dot{9}$	$\dot{4}$	$\dot{0}$	$\dot{2}$	$\dot{2}$	$\dot{1}$	$\dot{4}$	$\dot{1}$	$\dot{0}$
$\dot{A}$	$\dot{0}$	$\dot{5}$	$\dot{4}$	$\dot{5}$	$\dot{4}$	$\dot{0}$	$\dot{6}$	$\dot{6}$
$\dot{B}$	$\dot{7}$	$\dot{4}$	$\dot{5}$	$\dot{2}$	$\dot{3}$	$\dot{0}$	$\dot{1}$	$\dot{1}$
$\dot{C}$	$\dot{2}$	$\dot{1}$	$\dot{7}$	$\dot{7}$	$\dot{6}$	$\dot{5}$	$\dot{4}$	$\dot{4}$
$\dot{D}$	$\dot{4}$	$\dot{7}$	$\dot{6}$	$\dot{1}$	$\dot{0}$	$\dot{3}$	$\dot{2}$	$\dot{2}$
$\dot{E}$	$\dot{1}$	$\dot{4}$	$\dot{1}$	$\dot{5}$	$\dot{4}$	$\dot{2}$	$\dot{2}$	$\dot{5}$
$\dot{F}$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{2}$	$\dot{0}$	$\dot{5}$	$\dot{4}$	$\dot{2}$

In Toy Hash we have following properties, where the attacks are build only by iterated a same message block again and again:

1. We can append message  $\dot{2}||\dot{2}||\dot{2}$  at the end of any message to build collision.

$$\forall m, m' \in I_{4^{**}}, H^T(\dot{2}^{(3)}||m, IV) = H^T(\dot{2}^{(3)}||m', IV) = \dot{0}$$



**Fig. 7.** The Compression Function of Toy Hash

2. Both Graph  $G_1$  and  $G_2$  have cluster at fixed vertex, but the minimum block length of message block, making collision by appending repeating of same message block, are different.

$$\forall m, m' \in I_{4,*}, H^T(\dot{1}^{(t)} \| m, IV) = H^T(\dot{1}^{(t)} \| m', IV), \forall t \geq 7$$

$$\forall m, m' \in I_{4,*}, H^T(\dot{2}^{(t)} \| m, IV) = H^T(\dot{2}^{(t)} \| m', IV), \forall t \geq 3$$

3.  $G_2$  and  $G_3$  are isomorphic, but the function  $y = F_2(\cdot, x_h)$  is APN and  $y = F_3(\cdot, x_h)$  is near a linear function.
4.  $P_{Y|M=\dot{3}}(y)$  and  $P_{Y|M=\dot{4}}(y)$  with same distribution, but the derived Graph  $G_2$  and  $G_3$  are far away from each other.
5.  $\max_y P_{Y|M=\dot{5}}(y) > \max_y P_{Y|M=\dot{2}}(y)$ , but collision is more easy to build in iteration procedure when select the message  $\dot{2} \| \dots \| \dot{2}$  or  $\dot{5} \| \dots \| \dot{5}$ .
6.  $F_{\dot{8}}(\cdot, x_h) = F_2(\cdot, x_h) \oplus \dot{3}$ ,  $F_{\dot{9}}(\cdot, x_h) = F_2(\cdot, x_h) \oplus \dot{4}$ ,  $F_{\dot{A}}(\cdot, x_h) = F_2(\cdot, x_h \oplus \dot{5})$  but the derived graphs are totally different.
7.  $F_{\dot{B}}(\cdot, x_h) = F_1(\cdot, x_h) \oplus \dot{7}$ , but in Graph in  $G_1$  exist a cluster, the properties of Graph  $G_{\dot{B}}$  is more like the properties of Graph  $G_0$ . but the derived graph is totally different.
8.  $F_{\dot{E}}(\cdot, x_h) = F_0(\cdot, x_h) \oplus x_h$ ,  $F_{\dot{F}}(\cdot, x_h) = F_0(\cdot, x_h) \wedge x_h$ , the function  $y = F_0(\cdot, x_h)$  is APN permutation, but the Graph properties of derived graphs of  $y = F_0(\cdot, x_h) \oplus x_h$  and  $y = F_0(\cdot, x_h) \wedge x_h$  are totally different. Although the collision is more easy to build in compression function  $F_0(\cdot, x_h) \wedge x_h$ , in iteration procedure, the collision of  $H^T$  is more easy to build by append the message  $\dot{E}^{(4)}$ .

$$9. F_{\dot{C}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{3}, F_{\dot{D}}(\cdot, x_h) = F_{\dot{1}}(\cdot, x_h) \oplus \dot{5},$$

$$\forall m \in I_{4,*}, H^T(\dot{1}^{(7)} \| m, IV) \in \{7\}$$

$$\forall m \in I_{4,*}, H^T(\dot{C}^{(2)} \| m, IV) \in \{\dot{0}, \dot{1}, \dot{2}, \dot{4}, \dot{5}, \dot{6}\}$$

$$\forall m \in I_{4,*}, H^T(\dot{D}^{(4)} \| m, IV) \in \{\dot{0}, \dot{2}, \dot{4}, \dot{6}\}$$

10. For  $G_{\dot{7}}$ , we have:

$$P_{Z|M=\dot{2}}(z) = \begin{cases} \frac{1}{4}, & z \in \{3, 6, 7\} \\ \frac{1}{8}, & z \in \{0, 2\} \\ 0, & z \in \{1, 4, 5\} \end{cases}$$

$$P_{Z|M=\dot{2}^{(2)}}(z) = \begin{cases} \frac{1}{6}, & z \in \{3\} \\ \frac{1}{6}, & z \in \{7\} \\ \frac{1}{6}, & z \in \{0\} \\ 0, & z \in \{1, 2, 4, 5, 6\} \end{cases}$$

$$P_{Z|M=\dot{2}^{(2)}}(z) = \begin{cases} \frac{1}{6}, & z \in \{7\} \\ \frac{1}{6}, & z \in \{3\} \\ \frac{1}{6}, & z \in \{0\} \\ 0, & z \in \{1, 2, 4, 5, 6\} \end{cases}$$

### 3.2 The Properties of Graph $G_{\mathbf{m}}$

Since the graph  $G_{\mathbf{m}}$  is build from Function  $y = F_{\mathbf{m}}(\cdot, x_h)$ , it has some properties that original directed graph does not have.

**Lemma 2.** *Graph  $G_m$  is build from  $y = F(m, x_h)$ :*

1. The degree of graph  $G_m$ ,  $|G_m| = 2^n$ ;
2. The edge of graph  $G_m$ ,  $\|G_m\| = 2^n$ ;
3.  $d_{G_m}^O(v_{x_h}) = 1; \delta(G_m) = 1$ ;
4.  $\Delta(G_m) = 2^n + 1$ .

*Proof.* 1.  $\#\{0, 1\}^n = 2^n \Rightarrow |G_{\mathbf{m}}| = 2^n$ ;

2.  $\forall x_{h_0} \in I_n$  exist and only exist one  $y_0 \in I_n$  with  $y_0 = F(\mathbf{m}, x_{h_0})$ , for  $y_0 \in I_n$ , that means for each  $x_{h_0}$  exist one directed edge from it in  $G_{\mathbf{m}}$ , there are  $2^n$  directed edge exist, we get  $\|G_{\mathbf{m}}\| = 2^n$ ;

3. for each  $x_{h_0} \in I_n$ , we can compute  $y_0 = F(\mathbf{m}, x_{h_0})$ , that means  $d_{G_{\mathbf{m}}}^O(x_{h_0}) = 1$ , if does not exist  $x'_{h_0}$  satisfy  $x_{h_0} = F(\mathbf{m}, x'_{h_0})$ , then  $d_{G_{\mathbf{m}}}(x_{h_0}) = 1$ .

4. When all edge point at same  $x_{h_0}$  include itself, then  $d_{G_{\mathbf{m}}}(x_{h_0}) = 2^n + 1$

**Lemma 3.** *Graph  $G_m$  is build from  $y = F(m, x_h)$ :*

1. If  $x_{h_0} \in C$ , exist  $x'_{h_0} \in C$  with  $x_{h_0} = F(m, x'_{h_0}) \in C$  and  $F(m, x_{h_0}) \in C$ .
2. all cycles  $C, C'$  in  $G_{\mathbf{m}}$  with  $C \cap C' = \emptyset$  or  $C = C'$ .
3. Exist and only exist one cycle  $C_i$  in each connective subgraph  $\mathcal{H}_i$ ;

4. Exist  $\Upsilon$  cycles  $C_1, \dots, C_\Upsilon$  in  $G_m \Leftrightarrow$  exist  $\Upsilon$  connective subgraph  $\mathcal{H}_1, \dots, \mathcal{H}_\Upsilon$  with  $\mathcal{H}_i \cap \mathcal{H}_j = \emptyset$  and  $G_m = \mathcal{H}_1 \cup \dots \cup \mathcal{H}_\Upsilon$ .
5. vertexes in each cycle compose a permutation;

*Proof.* Let  $|C| = t$ .

1.  $d_{G_m}^O(x_h) = 1, d_C(x_h) = 2, |C| = t, \sum_{v_{x_h} \in C} d_C^I(v_{x_h}) = \sum_{v_{x_h} \in C} d_C^O(v_{x_h}) \Rightarrow \sum_{v_{x_h} \in C} d_C^O(v_{x_h}) \leq t, \sum_{v_{x_h} \in C} d_C(v_{x_h}) = 2t \Rightarrow \sum_{v_{x_h} \in C} d_C^I(v_{x_h}) = t \Rightarrow d_C^I(v_{x_h}) = 1, d_C^O(v_{x_h}) = 1$
2. From item 1 we have  $\forall v_{x_h} \in C, F(m, x_h) \in C, \forall v_{x_h} \in C \cap C' \Rightarrow F(m, x_h) \in C \cap C'$ , since  $C$  and  $C'$  have finite vertexes, we have  $C = C'$  or  $C \cap C' = \emptyset$ ;
3. Since  $\mathcal{H}_i$  is connective subgraph,  $\mathcal{H}_i \cap \mathcal{H}_j = \emptyset, i \neq j$ , we have  $\forall v_{x_h} \in \mathcal{H}_i, F(m, x_h) \in \mathcal{H}_i$ ; let  $v_{x_{h_0}} \in \mathcal{H}_i, x_{h_1} = F(m, x_{h_0})$ , then  $v(x_{h_1}) \in \mathcal{H}_i$ , let  $x_{h_2} = F(m, x_{h_1})$  then  $x_{h_2} \in \mathcal{H}_i$ , since  $H_i$  is limited subgraph, there must be exist  $x_{h_t}$  with  $x_{h_t} = x_{h_i}, i \leq t$ , we find the cycle. If exist  $C_i, C'_i \in \mathcal{H}_i$ , then there must exist  $x_{h_0}, x_{h'_0} \in \mathcal{H}_i$ , but  $x_{h_0} \notin C$  and  $x_{h'_0} \notin C'$ , with  $F(m, x_{h_0}) \in C$  and  $F(m, x_{h'_0}) \in C'$ , if  $x_{h_0} = x_{h'_0}$ , then get conflict, if we then find value  $x_{h_t}$  with  $x_{h_{t-1}} = F(m, x_{h_t}), x_{h'_{t-1}} = F(m, x_{h_t})$ , since  $G_1$  is limited graph, we get conflict. We get in one connected subgraph only exist one cycle.
4. Direct result of item 3.
5.  $\forall x_{h_0} \in C, F(m, x_{h_0}) \in C$  and exist  $x'_{h_0} \in C$  with  $x_{h_0} = F(m, x'_{h_0})$ .

### 3.3 The Relation Between $G_m$ and M-D Hash

From the toy hash we can make such conclusion:

**Assumption 1** Let Graph  $G_m$  is derived from  $y = F(m, x_h), m \in I_n$ :

1. The paths of Graph  $G_m$  are influenced by both linear components and non-linear components of compression function  $y = F(\cdot, x_h)$ ;
2. The cycle number and the length of cycle of Graph  $G_m$  are influenced by both linear components and nonlinear components of compression function  $y = F(\cdot, x_h)$ ;
3. The connectivity of Graph  $G_m$  is influenced by both linear components and nonlinear components of compression function  $y = F(\cdot, x_h)$ ;
4. If the vertexes of Graph  $G_m$  are not distinguished, then degree of Graph  $G_m$  is not influenced by linear components of compression function  $y = F(\cdot, x_h)$ ;
5. If the Graph  $G_m - v(x_{h_0})$  is a tree, by append message  $m_0^{(t)}$  at the end of messages  $m, m' \in I_{n,*}$ , we can build collision.
6.  $y = F(\cdot, x_h)$  is designed as pseudo random function, which can not prevent the cluster in iteration procedure.
7.  $y = F(\cdot, x_h)$  is a pseudo random permutation, then there is no cluster in iteration procedure.
8. If the two arrow from  $x_{h_1}$  and  $x_{h_2}$  point at same value, we know we find a collision with  $F(m, x_{h_1}) = F(m, x_{h_2})$ .

9. If no arrow pointed at some point  $x_{h_0}$  means  $\forall x_h \in I_n, F(m_0, x_h) \neq x_{h_0}$ .

**Theorem 6.** In M-D hash  $z = H^M(m, x)$ ,  $\forall m \in I_n$  and digraph  $G_m$  exist  $L \in \mathbf{N}$  such that

$$P_{\dot{Z}|M=m^{(t)}}(z) = 0, \forall i \in [1, \Upsilon], z \notin C_i, t \geq L$$

$$\frac{1}{2^n} \leq P_{\dot{Z}|M=m^{(t)}}(z) \leq \frac{|H_i| - |C_i|}{2^n}, \exists i \in [1, \Upsilon], z \in C_i, C_i \subseteq H_i, t \geq L$$

$$\max_z P_{\dot{Z}|M=m^{(t)}}(z) = \max \frac{|T_{ij}|}{2^n}, \forall t \geq L, j \in [1, \tau_i], i \in [1, \Upsilon].$$

*Proof.* 1. If  $z \notin C_i$ , which implies after several step of iteration, does not exist  $x_h \in I_n$  with  $H(m^{(t)}, x_h) = z$ ;  
 2. If  $z_0 \in C_i$  then exist  $x_{h_0} \in C_i$  with  $z_0 = F(m, x_{h_0}) \Rightarrow P_{\dot{Z}|M=m^{(t)}}(z_0) \geq \frac{1}{2^n}$   
 3. Since if  $z_0$  is a root of  $T_{ij}$ , exists  $t_0$  with  $\forall x_h \in T_{ij}, H(m^{(t_0)}, x_h) = z_0$ , so we have the conclusion. □

## Part II

# Security Proof of Structures



## 4 Conditional Probability of the Structures

### 4.1 Conditional Probability of Known Structures

**Theorem 7 (Conditional Probability of M-D Hash).** *Let  $y = F(x_m, x_h)$ ,  $z = H^M(m, x)$ ,  $x \in I_n$ ,  $m = m_* \dots m_1 \in I_{*n}$ ,  $m_i \in I_{\kappa}$ ,  $* \leq t$ ,  $m$  and  $x$  are independent from each other and the distribution of  $y$  is independent from  $x_m$  and  $x_h$  then:*

1.  $P_{\dot{Z}|M=m}(z) \leq \frac{(S_F)^t}{2^n}$ ;
2.  $P_{\dot{Z}|X=x}(z) \leq \frac{T_F}{2^\kappa}$ .

*Proof.* The Proof is given by deduction theory.

1. When  $t = 1$ :

$$\begin{aligned} P_{\dot{Z}|M=m}(z) &\leq \max_{m_0, z_0} \sum_{x \in I_n} P_X(x) \chi_{F(m_0, x)}(z_0) \\ &= \max_{m_0, z_0} \sum_{i \in [1, 2^n]} 2^{-n} \#\{(z_0, m_0, x_i)\}^F \leq 2^{-n} S_F \end{aligned}$$

Suppose  $t < l$  the inequality is true, when  $t = l$ :

$$\begin{aligned} P_{\dot{Z}|M=m}(z) &= P_{\dot{Z}|M=\mathbf{m}_l|m'}(z) \\ &= \sum_{x \in I_n} P_X(x) \chi_{F(\mathbf{m}_l, H^M(m', x))}(z) \\ &= \sum_{x \in I_n} \sum_{u \in I_n} P_X(x) \chi_{F(\mathbf{m}_l, u)}(z) \chi_{H^M(m', x)}(u) \\ &= \sum_{u \in I_n} \chi_{F(\mathbf{m}_l, u)}(z) \sum_{x \in I_n} P_X(x) \chi_{H^M(m', x)}(u) \\ &= \sum_{u \in I_n} \chi_{F(\mathbf{m}_l, u)}(z) P_{\dot{U}|M'=m'}(u) \\ &\leq 2^{-n} S_F^{l-1} \sum_{u \in I_n} \chi_{F(\mathbf{m}_l, u)}(z) \\ &\leq 2^{-n} S_F^{l-1} S_F = 2^{-n} S_F^l \end{aligned}$$

2. When  $t = 1$

$$\begin{aligned} P_{\dot{Z}|X=x}(z) &\leq \max_{x_0, z_0} \sum_m P_M(m) \chi_{F(m, x_0)}(z_0) \\ &= \max_{x_0, z_0} \sum_i 2^{-\kappa} \#\{(m_i, x_0, z_0)\}^F \leq 2^{-\kappa} T_F \end{aligned}$$

When  $t > 1$ :

$$\begin{aligned}
P_{\tilde{Z}|X=x}(z) &= \sum_{m \in \cup_{i=1}^t I_{\kappa \cdot i}} P_M(m) P_{\tilde{Z}|M=m, X=x}(z) \\
&= \sum_{\mathbf{m}_l \in I_{\kappa}} \sum_{m' \in \cup_{i=1}^{t-1} I_{\kappa \cdot i}} P_{M'}(m') P_{M_l}(\mathbf{m}_l) \\
&\quad P_{\tilde{Z}|M=m, X=x} P((z = F(\mathbf{m}_l, u), u = H^M(m', x))) \\
&= \sum_{\mathbf{m}_l \in I_{\kappa}} \sum_{m' \in \cup_{i=1}^{t-1} I_{\kappa \cdot i}} \sum_{u \in I_{\kappa}} \\
&\quad P_{M'}(m') P_{M_l}(\mathbf{m}_l) \chi_{F(\mathbf{m}_l, u)}(z) \chi_{H^M(m', x)}(u) \\
&= \sum_{u \in I_{\kappa}} \sum_{\mathbf{m}_l \in I_{\kappa}} P_{M_l}(\mathbf{m}_l) \chi_{F(\mathbf{m}_l, u)}(z) \\
&\quad \sum_{m' \in \cup_{i=1}^{t-1} I_{\kappa \cdot i}} P_{M'}(m') \chi_{H^M(m', x)}(u) \\
&= \sum_{u \in I_{\kappa}} P_{\tilde{Z}|U=u}(z) P_{U|X=x}(u) \\
&\leq 2^{-\kappa} T_F \sum_{u \in I_{\kappa}} P_{U|X=x}(z) = 2^{-\kappa} T_F
\end{aligned}$$

From induction principle we get the conclusions.  $\square$

*Remark 7.* The proof require the compression function with property of  $y = F(x_m, x_h)$  with  $P_{Y|X_m=x_m}(y) = P_Y(y)$ ,  $P_{Y|X_h=x_h}(y) = P_Y(y)$ , that are also required in design of block cipher, so we prefer design hash compression function based on block cipher design principle.

**Theorem 8 (Conditional Probability of Wide Pipe Hash).** *Let  $z = H^M(m, x)$ ,  $\tilde{z} = H^W(m, x) = G(z)$ ,  $x \in I_n$ ,  $m = m_* \dots m_1 \in I_{* \cdot n}$ ,  $m_i \in I_{\kappa}$ ,  $* \leq t$ ,  $m$  and  $x$  are independent and distribution of  $y$  is independent from that of  $x_m$  and  $x_h$ :*

1.  $P_{\tilde{Z}|M=m}(\tilde{z}) \leq \frac{(S_F)^t S_G}{2^n}$ ;
2.  $P_{\tilde{Z}|X=x}(\tilde{z}) \leq \frac{T_F S_G}{2^{\kappa}}$ .

*Proof.* –  $\forall t \geq 1$ :

$$\begin{aligned}
P_{\tilde{Z}|M=m}(\tilde{z}) &= \sum_{x \in I_n} P_X(x) P_{\tilde{Z}|M=m, X=x}(\tilde{z} = G(z), z = H^M(m, x)) \\
&= \sum_{x, z \in I_n} P_X(x) \chi_{G(z)}(\tilde{z}) \chi_{H^M(m, x)}(z) \\
&= \sum_{z \in I_n} \chi_{G(z)}(\tilde{z}) \sum_{x \in I_n} P_X(x) \chi_{H^M(m, x)}(z)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{z \in I_n} \chi_{G(z)}(\tilde{z}) P_{\tilde{Z}|M=m}(z) \\
&\leq \max_z P_{\tilde{Z}|M=m}(z) \sum_{z \in I_n} \chi_{G(z)}(\tilde{z}) \\
&\leq \frac{(S_F)^t S_G}{2^n}
\end{aligned}$$

–  $\forall t \geq 1$ :

$$\begin{aligned}
P_{\tilde{Z}|X=x}(\tilde{z}) &= \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\tilde{Z}|X=x, M=m}(\tilde{z} = G(z), z = H^M(m, x)) \\
&= \sum_{z \in I_n} \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\tilde{Z}|Z=z}(\tilde{z}) P_{\tilde{Z}|M=m, X=x}(z = H^M(m, x)) \\
&= \sum_{z \in I_n} \chi_{G(z)}(\tilde{z}) \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\tilde{Z}|M=m, X=x}(z = H^M(m, x)) \\
&= \sum_{z \in I_n} \chi_{G(z)}(\tilde{z}) P_{\tilde{Z}|X=x}(z) \\
&\leq \max_z P_{\tilde{Z}|X=x}(z) \sum_{z \in I_n} \chi_{G(z)}(\tilde{z}) \\
&\leq \frac{T_F S_G}{2^\kappa}
\end{aligned}$$

□

**Theorem 9 (Conditional Probability of Double Pipe Hash).** *Let  $z' \| z = H^M(m, x' \| x)$ ,  $\tilde{z} = H^D(m, x' \| x) = G(z)$ ,  $m = m_* \dots m_1 \in I_{* \cdot n}$ ,  $m_i \in I_\kappa$ ,  $i \leq *, * \leq t$ ,  $m$  and  $x$  are independent from each other and distribution of  $y$  is independent from that of  $x_m$  and  $x_h$  then:*

1.  $P_{\tilde{Z}|M=m}(\tilde{z}) \leq \frac{(S_F)^t S_G}{2^{2n}}$ ;
2.  $P_{\tilde{Z}|X=x}(\tilde{z}) \leq \frac{T_G}{2^\kappa}$ .

*Proof.* –  $\forall t \geq 2$ :

$$\begin{aligned}
&P_{\tilde{Z}|M=m}(\tilde{z}) \\
&= \sum_{x' \| x \in I_{2n}} P_{X' \| X}(x' \| x) \\
&\quad (P_{\tilde{Z}|M=\mathbf{m}_* \| m', X' \| X=x' \| x}(\tilde{z} = G(\mathbf{m}_*, z' \| z), z' \| z = H^M(m', x' \| x))) \\
&= \sum_{x' \| x, z' \| z \in I_{2n}} P_{X' \| X}(x' \| x) \chi_{G(\mathbf{m}_*, z' \| z)}(\tilde{z}) \chi_{H^M(m', x' \| x)}(z' \| z) \\
&= \sum_{z' \| z \in I_{2n}} \chi_{G(\mathbf{m}_*, z' \| z)}(\tilde{z}) \sum_{x' \| x \in I_{2n}} P_{X' \| X}(x' \| x) \chi_{H^M(m', x' \| x)}(z' \| z)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{z' \| z \in I_{2n}} \chi_{G(\mathbf{m}_*, z' \| z)}(\tilde{z}) P_{\dot{Z}|M=m'}(z' \| z) \\
&\leq \max_{z' \| z} P_{\dot{Z}|M=m'}(z' \| z) \sum_{z' \| z \in I_{2n}} \chi_{G(\mathbf{m}_*, z' \| z)}(\tilde{z}) \\
&\leq \frac{(S_{\tilde{F}})^t S_G}{2^{2n}}
\end{aligned}$$

–  $\forall t \geq 2$ :

$$\begin{aligned}
&P_{\dot{Z}|X' \| X=x' \| x}(\tilde{z}) \\
&= \sum_{\mathbf{m}_* \| m' \in \cup_{i=1}^t I_{n \cdot i}} P_M(\mathbf{m}_* \| m') \\
&P_{\dot{Z}|X' \| X=x' \| x, M=\mathbf{m}_* \| m'}(\tilde{z} = G(\mathbf{m}_*, z' \| z), z = H^M(m', x' \| x)) \\
&= \sum_{z' \| z \in I_{2n}} \sum_{\mathbf{m}_* \| m' \in \cup_{i=1}^t I_{n \cdot i}} P_M(\mathbf{m}_* \| m') \\
&P_{\dot{Z}|Z' \| Z=z' \| z, M_*=\mathbf{m}_*}(\tilde{z} = G(\mathbf{m}_*, z' \| z)) \\
&P_{\dot{Z}|M'=m', X' \| X=x' \| x}(z = H^M(m', x' \| x)) \\
&= \sum_{z' \| z \in I_{2n}, \mathbf{m}_* \in I_n} \chi_{G(\mathbf{m}_*, z' \| z)}(\tilde{z}) \\
&\quad \sum_{m' \in \cup_{i=1}^{t-1} I_{n \cdot i}} P_{M'}(m') P_{Z' \| Z|M'=m', X' \| X=x' \| x}(z' \| = H^M(m', x' \| x)) \\
&= \sum_{z' \| z \in I_{2n}} P_{\dot{Z}|Z' \| Z=z' \| z}(\tilde{z}) P_{Z' \| Z|X' \| X=x' \| x}(z' \| z) \\
&= \max_{z' \| z \in I_{2n}} P_{\dot{Z}|Z' \| Z=z' \| z}(\tilde{z}) \sum_{z' \| z \in I_{2n}} P_{Z' \| Z|X' \| X=x' \| x}(z' \| z) \\
&\leq \frac{T_G}{2^\kappa}
\end{aligned}$$

□

**Corollary 2.**  $\forall y$ , if  $y = F_{x_m}(\cdot, x_h)$  is permutation, then for  $z = H^M(m, x)$ ,  $\tilde{z} = H^W(m, x)$ :

1.  $P_{Z|M=m}(z) = 2^{-n}$ ;
2.  $P_{\tilde{Z}|M=m}(\tilde{z}) = S_G 2^{-n}$ ;

## 4.2 Conditional Probability of Improved Structures

**Theorem 10 (Ideal-Pipe Hash).**  $\tilde{y} = f(m, x)$ ,  $\tilde{y} = \bar{G}(f(m, x), x_h)$ ,  $y = F(x_m, x_h)$ , for  $z = H^M(m, x)$ ,  $\tilde{z} = H^I(f(m, x), z)$ ,  $m = m_t \| \dots \| m_1 \in I_{\kappa \cdot t}$ ,  $m$  and  $x$  are independent from each other, if  $f(m, x)$  is independent from  $x, m$  and  $z$  then:

1.  $P_{\tilde{Z}|M=m}(\tilde{z}) \leq \frac{S_f T_G}{2^\kappa};$
2.  $P_{\tilde{Z}|X=x}(\tilde{z}) \leq \frac{T_f T_G}{2^n}.$

*Proof.*  $\forall t \geq 1:$

$$\begin{aligned}
P_{\tilde{Z}|M=m}(\tilde{z}) &= P_{\tilde{Z}|M=m}(\tilde{z} = \bar{G}(u, z), u = f(m, x), z = H^M(m, x)) \\
&= \sum_{x, u, z \in I_n} P_X(x) \chi_{\bar{G}(z, u)}(\tilde{z}) \chi_{H^M}(z, m, x) \chi_{f(m, x)}(u) \\
&\quad u \text{ and } z \text{ are independent, and } P_X(x) = 2^{-n} \\
&= \sum_{u, z \in I_n} \chi_{\bar{G}(z, u)}(\tilde{z}) \sum_{x \in I_n} P_X(x) \chi_{H^M}(z, m, x) \\
&\quad \sum_{x \in I_n} P_X(x) \chi_{f(m, x)}(u) \\
&= \sum_{u, z \in I_n} \chi_{\bar{G}(z, u)}(\tilde{z}) P_{\dot{U}|M=m}(u) P_{\dot{Z}|M=m}(z) \\
&\leq \max_{u_0} P_{\dot{U}|M=m}(u_0) 2^n \sum_z P_{\dot{Z}|M=m}(z) \sum_u 2^{-n} \chi_{\bar{G}(z, u)}(\tilde{z}) \\
&= \max_{u_0} P_{\dot{U}|M=m}(u_0) 2^n \sum_z P_{\dot{Z}|M=m}(z) P_{\dot{Z}|Z=z}(\tilde{z}) \\
&\leq \max_{u_0} P_{\dot{U}|M=m}(u_0) \max_{z_0, \tilde{z}_0} 2^n P_{\dot{Z}|Z=z_0}(\tilde{z}_0) \sum_z P_{\dot{Z}|M=m}(z) \\
&= \max_{u_0} P_{\dot{U}|M=m}(u_0) \leq \frac{S_f T_G}{2^\kappa}
\end{aligned}$$

$\forall t \geq 1:$

$$\begin{aligned}
P_{\tilde{Z}|X=x}(\tilde{z}) &= P_{\tilde{Z}|X=x}(\tilde{z} = \bar{G}(u, z), u = f(m, x), z = H^M(m, x)) \\
&= \sum_{u, z \in I_n} \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\dot{Z}|\dot{U}=u, \dot{Z}=z}(\tilde{z}) \\
&\quad P_{\dot{U}, \dot{Z}|M=m, X=x}(u = f(m, x), z = H^M(m, x)) \\
&\quad \text{Since } P_M(x) = 2^{-\sum_i i \cdot n} \text{ and } u, z \text{ are independent} \\
&= \sum_{u, z \in I_n} \chi_{\bar{G}(z, u)}(\tilde{z}) \\
&\quad \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\dot{U}|M=m, X=x}(u = f(m, x)) \\
&\quad \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\dot{Z}|M=m, X=x}(z = H^M(m, x)) \\
&= \sum_{u, z \in I_n} \chi_{\bar{G}(z, u)}(\tilde{z}) P_{\dot{U}|X=x}(u) P_{\dot{Z}|X=x}(z)
\end{aligned}$$

$$\begin{aligned}
&\leq \max_{u_0} P_{\dot{U}|X=x}(u_0) 2^n \sum_z P_{\dot{Z}|X=x}(z) \sum_u 2^{-n} P_{\dot{Z}|U=u, Z=z}(\tilde{z}) \\
&= \max_{u_0} P_{\dot{U}|X=x}(u_0) 2^n \sum_z P_{\dot{Z}|X=x}(z) P_{\dot{Z}|Z=z}(\tilde{z}) \\
&\leq \max_{u_0} P_{\dot{U}|X=x}(u_0) \max_{z_0, \tilde{z}_0} 2^n P_{\dot{Z}|Z=z_0}(\tilde{z}_0) \sum_z P_{\dot{Z}|X=x}(z) \\
&= \max_{u_0} P_{\dot{U}|X=x}(u_0) \leq \frac{T_f T_G}{2^n}
\end{aligned}$$

□

**Theorem 11.**  $z = H^m(m, x)$ ,  $z \in I_n, \tilde{z} = H^N(m, x)$ ,  $m = m_t \parallel \dots \parallel m_1 \in I_{\kappa \cdot t}$ ,  $h_i = F(m_i, h_{i-1})$ ,  $h_0 = x$  and  $m_1, \dots, m_t$  are independent from each other, if  $f(m, x) = h_* \oplus \dots \oplus h_1 \oplus h_0$  is independent from  $x$  and  $m$  then:

1.  $P_{\dot{Z}|M=m}(\tilde{z}) \leq \frac{S_f T_G}{2^n}$ ;
2.  $P_{\dot{Z}|X=x}(z) \leq \frac{T_f T_G}{2^n}$ .

## 5 The Advantage of Hash Function

### 5.1 The Advantage of Compression Function

**Theorem 12 (Pseudo Preimage Attack).** Let  $y = F(x_m, x_h)$  is Black Box Model then:

$$\tilde{Adv}_F^{Pre}(q) \leq 2q \cdot \max\left\{\frac{T_F}{2^\kappa}, \frac{S_F}{2^n}\right\}.$$

$$\tilde{Adv}_F^{Pre}(A) = \max\left\{\frac{T_F}{2^\kappa}, \frac{S_F}{2^n}\right\}.$$

$$\tilde{Adv}_F^{Pre}(A) = \max_{x_m, x_h} \max_y \{P_{\dot{Y}|X_h=x_h}(y), P_{\dot{Y}|X_m=x_m}(y)\}.$$

*Proof.*  $F(x_m, x_h)$  is Black Box Model, the only way to get preimage is exhaustive search. The exhaustive search has following ways:

- For given  $y_0, x_{h_0}$  searching  $x_m$  with  $y = F(x_m, x_{h_0})$

Game( $A, F, q, y_0, x_{h_0}$ )

For  $i = 1, \dots, q$  do :

$A(y_0, x_{h_0}) \rightarrow (x_{m_i})$

$A$  wins if  $\exists i$  st.  $F(x_{m_i}, x_{h_0}) = y_0$ .

the success probability of  $i$ th time selection  $p_{[C_i]}$  is[3]:

$$p_{[C_i]} = \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^{\kappa-i+1}}$$

The success probability of  $q$  times trying is:

$$\begin{aligned} p_{[C_1 \vee \dots \vee C_t]} &= \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa} + \dots + \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa - q + 1} \\ &\leq q \cdot \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa - q} \leq q \cdot \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa - 2^{\kappa-1}} \\ p_{[C_1 \vee \dots \vee C_t]} &= \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa} + \dots + \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa - q + 1} \\ &\geq q \cdot \frac{\#\{(y_0, x_m, x_{h_0})\}^F}{2^\kappa} \end{aligned}$$

We get the maximum success probability is  $2q \cdot \frac{T_F}{2^\kappa}$ .

- For given  $y_0, x_{m_0}$  searching  $x_h$ , we get the maximum success probability is  $2q \cdot \frac{S_F}{2^n}$ .

$Game(A, F, q, y_0, x_{m_0})$

For  $i = 1, \dots, q$  do :

$A(y_0, x_{m_0}) \rightarrow (x_{h_i})$

$A$  wins if  $\exists i$  st.  $F(x_{m_0}, x_{h_i}) = y_0$ .

- For given  $y_0$ , randomly searching  $x_h$  and  $x_m$ , the maximum success probability is  $2q \cdot \frac{R_F}{2^\kappa 2^n}$ .

$Game(A, F, q, y_0)$

For  $i = 1, \dots, q$  do :

$A(y_0) \rightarrow (x_{m_i}, x_{h_i})$

$A$  wins if  $\exists i$  st.  $F(x_{m_i}, x_{h_i}) = y_0$ .

□

**Theorem 13 (Pseudo Collision Attack).** *Let  $y = F(x_m, x_h)$  is Black Box Model then:*

$$\tilde{Adv}_F^{Coll}(q) \leq q(q-1) \cdot \max\left\{\frac{T_F-1}{2^\kappa}, \frac{S_F-1}{2^n}, \frac{R_F-1}{2^\kappa 2^n}\right\}.$$

$$\tilde{Adv}_F^{Pre}(A) = \max\left\{\frac{T_F-1}{2^\kappa-1}, \frac{S_F-1}{2^n-1}, \frac{R_F-1}{2^\kappa 2^n-1}\right\}.$$

*Proof.* The collision can be get only by exhaustive search.

- The fastest way to search for collision is the way based on birthday paradox. For random selected  $x_{h_0}$  searching  $x_{m_1}, x_{m_2}, \dots, x_{m_t}$  finding collision of  $F(x_{m_i}, x_{h_0}) = F(x_{m_j}, x_{h_0})$ ,

$Game(A, F, q, x_{h_0})$

For  $i = 1, \dots, q$  do :

$A() \rightarrow (x_{m_i}), (x_{m_i} \neq x_{m_j}, j < i)$

$A$  wins if  $\exists i, j$  st.  $F(x_{m_i}, x_{h_0}) = F(x_{m_j}, x_{h_0})y_0$ .

let  $C_i$  be the event that the  $i$ -th selection make collision with one of the previous ones. Then  $p_{[C_i]}$  is at most[3]:

$$p[C_i] \leq \frac{i(T_F-1)}{2^\kappa - i}$$

Then the maximum success probability of  $q$  times trying is:

$$\begin{aligned} p[C_1 \vee \dots \vee C_q] &\leq p[C_1] + \dots + p[C_q] \\ &\leq \sum_{i=0}^{q-1} \frac{i(T_F-1)}{2^\kappa - i} \leq \frac{q \cdot (q-1)(T_F-1)}{2 \cdot 2^\kappa - q} \\ &\leq \frac{q \cdot (q-1)(T_F-1)}{2 \cdot 2^{\kappa-1}} = q(q-1) \frac{(T_F-1)}{2^\kappa} \end{aligned}$$

– similar as item 1, we get for selected  $x_m$  the complexity is  $q(q-1) \frac{(S_F-1)}{2^n}$ ;  
*Game*( $A, F, q, x_{m_0}$ )

For  $i = 1, \dots, q$  do :

$A() \rightarrow (x_{h_i}), (x_{h_i} \neq x_{h_j}, j < i)$

$A$  wins if  $\exists i, j$  st.  $F(x_{m_0}, x_{h_i}) = F(x_{m_0}, x_{h_j})y_0$ .

– similar as item 1, searching  $x_m, x_h$ , the complexity is  $q(q-1) \frac{(R_F-1)}{2^\kappa 2^n}$ .

*Game*( $A, F, q$ )

For  $i = 1, \dots, q$  do :

$A() \rightarrow (x_{m_i}, x_{h_i}), ((x_{m_i}, x_{h_i}) \neq (x_{m_i}, x_{h_j}), j < i)$

$A$  wins if  $\exists i, j$  st.  $F(x_{m_0}, x_{h_i}) = F(x_{m_0}, x_{h_j})y_0$ .

□

**Theorem 14 (Fix Start Preimage Attack).** Let  $y = F(x_m, x_h)$  then:

1. If  $y = F(x_m, \cdot)$  is invertible then:

$$\tilde{Adv}_F^{FixP}(A) = 1.$$

2. If  $y = F(x_m, \cdot)$  is Black Box Model then

$$\tilde{Adv}_F^{FixP}(q) \leq 2q \frac{T_F}{2^\kappa}$$

$$\tilde{Adv}_F^{FixP}(A) = \frac{T_F}{2^\kappa}$$

*Proof.*

1. For  $x_{h_0}, y_0$ , compute  $F_{x_{h_0}}^{-1}(y, \cdot)$ , let  $x_m = F_{x_{h_0}}^{-1}(y, \cdot) \Rightarrow \tilde{Adv}_F^{FixP}(A) = 1$ .

2. For given  $x_{h_0}$ , there are two ways to search the preimage of  $y_0$ :

– Select  $x_{h_0}$ , search  $x_m$  satisfy  $y_0 = F(x_m, x_{h_0})$ ,

*Game*( $A, F, q, y_0, x_{h_0}$ )

For  $i = 1, \dots, q$  do :

$A(y_0, x_{h_0}) \rightarrow (x_{m_i})$

$A$  wins if  $\exists i$  st.  $F(x_{m_i}, x_{h_0}) = y_0$ .

in  $q$  times trying the maximum success probability is  $2q \frac{T_F}{2^\kappa}$ .

- For given  $y_0$ , select  $x_{m_i}$  get  $x_{h_i}$ , satisfying  $y_0 = F(x_{m_i}, x_{h_i})$ , then check  $x_{h_i} = x_{h_0}$  being satisfied or not,  
 $Game(A, F, q, y_0, x_{h_0})$   
For  $i = 1, \dots, q$  do :  
 $A(y_0) \rightarrow (x_{m_i})$   
 $x_{h_i} \leftarrow F_{x_{m_i}}^{-1}(\cdot, y_0)$   
A wins if  $\exists i$  st.  $x_{h_0} = x_{h_i}$ .  
for random selected  $x_{m_i}$ , the maximum probability of  $x_{h_i} = x_{h_0}$  is:  
 $p = P_{Y|X_h=x_{h_0}}(y_0) \leq \frac{T_h}{2^\kappa}$ .  $\square$

**Theorem 15 (Fix Start Collision Attack).** Let  $y = F(x_m, x_h)$ :

1. If  $y = F(x_m, \cdot)$  is invertible then:

$$\tilde{Adv}_F^{FixC}(q) = \begin{cases} 1 & T_F > 1 \\ 0 & \text{else} \end{cases};$$

2. If  $y = F(x_m, \cdot)$  is black box model then:

$$\tilde{Adv}_F^{FixC}(q) \leq q(q-1) \frac{T_F - 1}{2^\kappa}.$$

$$\tilde{Adv}_F^{FixC}(A) = \frac{T_F - 1}{2^\kappa - 1}.$$

*Proof.*

1. For  $x_{h_0}$ , random select  $y_0$ , get  $x_m$  with  $y_0 = F(x_m, x_{h_0})$ , if  $T_F > 1$  exist collision, else no collision.
2. There are two ways to attack:
  - The maximum success probability of random select  $x_{m_1}, \dots, x_{m_t}$  getting  $y = F(x_{h_0}, x_{m_i})$ , check  $F(x_{h_0}, x_{m_i}) = F(x_{h_0}, x_{m_j})$  being satisfied or not, same as proof of Theorem13, the minimum requirement of computation is  $q(q-1) \frac{T_F - 1}{2^\kappa}$ ;  
 $Game(A, F, q, x_{h_0})$   
For  $i = 1, \dots, q$  do :  
 $A() \rightarrow (x_{m_i}), (x_{m_i} \neq x_{m_j}, j < i)$   
A wins if  $\exists i, j$  st.  $F(x_{m_i}, x_{h_0}) = F(x_{m_j}, x_{h_0})y_0$ .
  - For given  $y_0$  and  $x_{h_0}$ , the success probability of finding collision with  $F(x_{h_0}, x_{m_i}) = F(x_{h_0}, x_{m_j})$  is same as finding the preimage of  $y_0 = F(x_{h_0}, x_m)$ ,  
 $Game(A, F, q, y_0, x_{h_0})$   
For  $i = 1, \dots, q$  do :  
 $A(y_0) \rightarrow (x_{m_i})$   
 $x_{h_i} \leftarrow F_{x_{m_i}}^{-1}(\cdot, y_0)$   
A wins if  $\exists i, j$  st.  $x_{h_0} = x_{h_i}, x_{h_0} = x_{h_j}$ ,  
the success probability is smaller than  $q \frac{T_F - 1}{2^\kappa}$ .  $\square$

## 5.2 Advantage of Known Structures

**Lemma 4 (Conditional Probability of M-D Hash).** *Let  $y = F(x_m, x_h)$ ,  $z = H^M(m, x)$ ,  $x \in I_n$ ,  $m = m_* \dots m_1 \in I_{* \cdot \kappa}$ ,  $m_i \in I_\kappa$ ,  $* \leq t$ ,  $m$  and  $x$  are independent from each other then:*

1.  $P_{\dot{Z}|M=m}(z) \leq \frac{(S_F)^t}{2^n}$ ;
2.  $P_{\dot{Z}|X=x}(z) \leq \frac{T_F}{2^\kappa}$

*Remark 8.* The bound item 2 is tight, the bound item 1 should be discussed in true design.

**Theorem 16.** *If  $z = H^M(m, x)$ ,  $y = F(x_m, x_h)$ ,  $m = m_* \dots m_1 \in I_{* \cdot \kappa}$ ,  $* \leq t$ ,  $IV$  is the initial value, if*

1.  $\max_m \max_z P_{\dot{Z}|M=m}(z) \leq \frac{S_1}{2^n}$ ;
2.  $\max_m \max_z P_{\dot{Z}|M=m^{(t)}}(z) \leq \frac{S_2}{2^n}$ ;
3.  $\max_x \max_z P_{\dot{Z}|X=x}(z) \leq \frac{T_1}{2^\kappa}$ .

then<sup>4</sup>:

1. If  $y = F(x_m, x_h)$  is Black Box Model:
  - (a)  $\tilde{Adv}_{H^M}^{Pre}(q) \leq \max\{2q \frac{S_1}{2^n}, 2q \frac{T_1}{2^\kappa}, q \frac{S_2}{2^n}\}$
  - (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
  - (c)  $\tilde{Adv}_{H^M}^{FixP}(q) \leq \max\{2q^2 \frac{S_1^2}{2^{2n}}, 2q \frac{T_1}{2^\kappa}, q \frac{S_2}{2^n}\}$
  - (d)  $\tilde{Adv}_{H^M}^{FixC}(q) \leq \max\{q^2(q-1) \frac{S_1^2}{2^{2n}}, q(q-1) \frac{T_1}{2^\kappa}, q \frac{S_2}{2^n}\}$
2. If  $y = F(x_m, \cdot)$  is invertible then
  - (a)  $\tilde{Adv}_{H^M}^{Pre}(q) = 1$
  - (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
  - (c)  $\tilde{Adv}_{H^M}^{FixP}(q) = 1$
  - (d)  $\tilde{Adv}_{H^M}^{FixC}(q) = 1$
3. If  $y = F(x_m, \cdot)$  is black box model then
  - (a)  $\tilde{Adv}_{H^M}^{Pre}(q) = 1$
  - (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
  - (c)  $\tilde{Adv}_{H^M}^{FixP}(q) \leq \max\{2q^2 \frac{S_1^2}{2^{2n}}, q(q-1) \frac{T_1}{2^\kappa}, q \frac{S_2}{2^n}\}$
  - (d)  $\tilde{Adv}_{H^M}^{FixC}(q) \leq \max\{q^2(q-1) \frac{S_1^2}{2^{2n}}, q(q-1) \frac{T_1}{2^\kappa}, q \frac{S_2}{2^n}\}$

*Proof.*

1. If  $F$  is Black Box Model: since the function  $y = F(x_m, x_h)$  is black box model, then the hash  $H^M(m, x)$  can be seen as black box model.

---

<sup>4</sup> Where we did pre-computation of  $H^M(m_0^{(l)}, IV)$ ,  $l \geq 1$ , until get  $l'$  and  $l''$  with  $H^M(m_0^{(l')}, IV) = H^M(m_0^{(l'')}, IV)$ , for a selected  $m_0$ .

- (a) For any selected
- $x_0 \in I_n$
- ,

**Game0**( $A, F, q, z_0, x_0$ )For  $i = 1, \dots, q$  do : $A(z_0, x_0) \rightarrow (m_i)$ A wins if  $\exists i$  st.  $z_0 = H^M(m_i, x_0)$ .The success probability in  $q$  time trying is:  $2q \frac{T_1}{2^\kappa}$ .**Game1**( $A, F, q, z_0, m_0$ )For  $i = 1, \dots, q$  do : $A(z_0, m_0) \rightarrow (x_i)$ A wins if  $\exists i$  st.  $z_0 = H^M(m_0, x_i)$ .

The success probability in  $q$  time trying is:  $2q \frac{S_1}{2^n}$ . if exist  $z_0$  with  $P_{Z|M=m_0}(z_0)$  get the maximum conditional probability, then the  $m_0$  be a large message, and one time trying of of Game1 be very big, and also  $m_0$  may be very hard to find, the complexity of finding such value is  $\mathcal{O}(2^n)$ , but if  $s_1 \geq 2^{\frac{n}{2}}$  and  $|m| \leq \frac{n}{2} \cdot \kappa$ , the attack is still possible.

**Game2**( $A, F, z_0, x_0, \mathbf{m}_0$ ) $z_1 \leftarrow x_0; Q = \emptyset;$ For  $i = 1, \dots, t$  do : $z_1 \leftarrow F(\mathbf{m}_0, z_1)$  $Q \leftarrow Q \cup z_1$ A wins if  $z_0 = z_1$  or  $z_1 \in Q$ Return  $t$ .

The Game2 can be used to find any preimage, for different  $\mathbf{m}_0$ , the  $t$  is different, the best way is select  $\mathbf{m}_0$  with the minimum  $t$ . If the  $\mathbf{m}_0$  is the value with  $P_{Z|M=\mathbf{m}_0^{(T)}}(z_0) = S_2 2^{-n}$ , then the success probability of finding preimage  $z_0$  with game2 is  $q S_2 2^{-n}$ , then the success probability is  $q S_2 2^{-n}$ .

- (b) Since
- $H^M(\mathbf{m}_2 || \mathbf{m}_1, x) = H^M(\mathbf{m}_2, H^M(\mathbf{m}_1, x))$
- , then we find the collision.

- (c) For
- $IV \in I_n$
- ,

**Game0**( $A, F, q, z_0, IV$ )For  $i = 1, \dots, q$  do : $A(z_0, IV) \rightarrow (m_i)$ A wins if  $\exists i$  st.  $z_0 = H^M(m_i, IV)$ .The success probability in  $q$  time trying is:  $2q \frac{T_1}{2^\kappa}$ .**Game1**( $A, F, q, z_0, m_0, IV$ ) $Q \rightarrow \emptyset;$ For  $i = 1, \dots, q$  do : $A(z_0, m_0) \rightarrow (x_i, m_i)$  $Q \rightarrow Q \cup H(m_i, IV)$ A wins if  $\exists i$  st.  $z_0 = H^M(m_0, x_i) \wedge x_i \in Q$ .

The success equals finding two preimage, so we have, in  $q$  time trying, the probability is  $:2q \frac{S_1}{2^n} 2q \frac{S_1}{2^n}$ .

**Game2**( $A, F, z_0, IV, \mathbf{m}_0$ ) $z_1 \leftarrow IV; Q = \emptyset;$ For  $i = 1, \dots, t$  do :

$$z_1 \leftarrow F(\mathbf{m}_0, z_1)$$

$$Q \leftarrow Q \cup z_1$$

*A wins if  $z_0 = z_1$  or  $z_1 \in Q$*

*Return  $t$ .*

The Game2 can be used to find any preimage, for different  $\mathbf{m}_0$ , the  $t$  is different, the best way is select  $\mathbf{m}_0$  with the minimum  $t$ . If the  $\mathbf{m}_0$  is the value with  $P_{Z|M=\mathbf{m}_0^{(t)}}(z_0) = S_2 2^{-n}$ , then the success probability of finding preimage  $z_0$  with game2 is  $qS_2 2^{-n}$ , then the success probability is  $qS_2 2^{-n}$ .

(d) For  $IV \in I_n$ ,

**Game0**( $A, F, q, IV$ )

*For  $i = 1, \dots, q$  do :*

$$A(Q, IV) \rightarrow (m_i)$$

*A wins if  $\exists i, j$  st.  $H^M(m_i, IV) = H^M(m_j, IV)$ .*

The success probability in  $q$  time trying is:  $q(q-1)\frac{T_1}{2^r}$ .

**Game1**( $A, F, q, m_0, IV$ )

$$Q \rightarrow \emptyset;$$

*For  $i = 1, \dots, q$  do :*

$$A(z_0, m_0) \rightarrow (x_i, m_i)$$

$$Q \rightarrow Q \cup H(m_i, IV)$$

*A wins if  $\exists i, j$  st.  $H^M(m_0, x_i) = H^M(m_0, x_j) \wedge (x_i \in Q \wedge x_j \in Q)$ .*

The success equals finding preimage and collision, so we have, in  $q$  time trying, the probability is  $:q\frac{S_1}{2^n}(q-1)q\frac{S_1}{2^n}$ .

**Game2**( $A, F, z_0, IV, \mathbf{m}_0$ )

$$z_1 \leftarrow IV; Q = \emptyset;$$

*For  $i = 1, \dots, t$  do :*

$$z_1 \leftarrow F(\mathbf{m}_0, z_1)$$

$$Q \leftarrow Q \cup z_1$$

*A wins if  $z_0 = z_1$  or  $z_1 \in Q$*

*Return  $t$ .*

Game2 also can be used to find the collision, the success probability is  $qS_2 2^{-n}$ .

2. If  $x_m = F^{-1}(y, x_h)$  then: The conclusions can be get by the direct computation, since  $x_m = F^{-1}(y, x_h)$ .

3. we make assumption of  $x_h$  can be gotten from  $x_h = F^{-1}(x_{m_0}, y)$ , (that is the worst condition).

(a) Can be gotten from direct computation;

(b) Can be gotten from direct computation;

(c) For  $IV \in I_n$ ,

**Game0**( $A, F, q, z_0, IV$ )

$Q \rightarrow \emptyset;$  *For  $i = 1, \dots, q$  do :*

$$A(z_0, IV) \rightarrow (m_i, \mathbf{m}_i)$$

$$Q \rightarrow F^{-1}(\mathbf{m}_i, z_0)$$

*A wins if  $\exists i$  st.  $H^M(m_i, IV) \in Q$ .*

The success probability in  $q$  time trying is:  $q(q-1)\frac{T_1}{2^r}$ .

**Game1**( $A, F, q, z_0, m_0, IV$ )

$Q \rightarrow \emptyset$ ;  
 For  $i = 1, \dots, q$  do :  
      $A(z_0, m_0) \rightarrow (x_i, m_i)$   
      $Q \rightarrow Q \cup H(m_i, IV)$   
 A wins if  $\exists i$  st.  $z_0 = H^M(m_0, x_i) \wedge x_i \in Q$ .

The success equals finding two preimage, so we have, in  $q$  time trying, the probability is  $:2q \frac{S_1}{2^n} 2q \frac{S_1}{2^n}$ .

**Game2**( $A, F, z_0, IV, \mathbf{m}_0$ )  
 $z_1 \leftarrow IV; Q = \emptyset$ ;  
 For  $i = 1, \dots, t$  do :  
      $z_1 \leftarrow F(\mathbf{m}_0, z_1)$   
      $Q \leftarrow Q \cup z_1$   
 A wins if  $z_0 = z_1$  or  $z_1 \in Q$   
 Return  $t$ .

The Game2 can be used to find any preimage, for different  $\mathbf{m}_0$ , the  $t$  is different, the best way is select  $\mathbf{m}_0$  with the minimum  $t$ . If the  $\mathbf{m}_0$  is the value with  $P_{Z|M=\mathbf{m}_0^{(T)}}(z_0) = S_2 2^{-n}$ , then the success probability of finding preimage  $z_0$  with game2 is  $q S_2 2^{-n}$ , then the success probability is  $q S_2 2^{-n}$ .

(d) For  $IV \in I_n$ ,

**Game0**( $A, F, q, IV$ )  
 For  $i = 1, \dots, q$  do :  
      $A(Q, IV) \rightarrow (m_i)$   
 A wins if  $\exists i, j$  st.  $H^M(m_i, IV) = H^M(m_j, IV)$ .

The success probability in  $q$  time trying is  $:q(q-1) \frac{T_1}{2^n}$ .

**Game1**( $A, F, q, m_0, IV$ )  
 $Q \rightarrow \emptyset$ ;  
 For  $i = 1, \dots, q$  do :  
      $A(z_0, m_0) \rightarrow (x_i, m_i)$   
      $Q \rightarrow Q \cup H(m_i, IV)$   
 A wins if  $\exists i, j$  st.  $H^M(m_0, x_i) = H^M(m_0, x_j) \wedge (x_i \in Q \wedge x_j \in Q)$ .

The success equals finding preimage and collision, so we have, in  $q$  time trying, the probability is  $:q \frac{S_1}{2^n} (q-1) q \frac{S_1}{2^n}$ .

**Game2**( $A, F, z_0, IV, \mathbf{m}_0$ )  
 $z_1 \leftarrow IV; Q = \emptyset$ ;  
 For  $i = 1, \dots, t$  do :  
      $z_1 \leftarrow F(\mathbf{m}_0, z_1)$   
      $Q \leftarrow Q \cup z_1$   
 A wins if  $z_0 = z_1$  or  $z_1 \in Q$   
 Return  $t$ .

Game2 also can be used to find the collision, the success probability is  $q S_2 2^{-n}$ .

□

**Corollary 3.** *If  $z = H^M(m, x)$ ,  $y = F(x_m, x_h)$ ,  $m = m_* \dots m_1 \in I_{*, \kappa}$ ,  $* \leq t$ ,  $IV$  is the initial value, If  $\forall x_{m_0} \in$ ,  $y = F(x_{m_0}, x_h)$  is permutation,  $y = F(x_m, \cdot)$  is black box model then:*

1.  $\tilde{Adv}_{H^M}^{FixP}(q) \leq q(q-1) \frac{T_F}{2^\kappa}$
2.  $\tilde{Adv}_{H^M}^{FixC}(q) \leq q(q-1) \frac{T_F}{2^\kappa}$

**Theorem 17 (Wide-Pipe Hash).** *Let  $G$  is Black Box Model and not invertible,  $z = H^M(m, x)$ ,  $\tilde{z} = H^W(m, x) = G(z)$ ,  $x$  and  $m_1, \dots, m_t$  are independent from each other, if*

1.  $\max_m \max_z P_{\tilde{Z}|M=m}(z) \leq \frac{S_1}{2^n}$ ;
2.  $\max_m \max_z P_{\tilde{Z}|M=m^{(t)}}(z) \leq \frac{S_2}{2^n}$ ;
3.  $\max_x \max_z P_{\tilde{Z}|X=x}(z) \leq \frac{T_1}{2^\kappa}$ .

then<sup>5</sup>:

1. *If  $y = F(x_m, x_h)$  is Black Box Model or  $y = F(x_m, \cdot)$  is black box model :*
  - (a)  $\tilde{Adv}_{H^M}^{Pre}(q) \leq \max\{2q \frac{S_G S_1}{2^\omega}, 2q \frac{S_G T_1}{2^\omega}, q \frac{S_G S_2}{2^\omega}\}$
  - (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
  - (c)  $\tilde{Adv}_{H^M}^{FixP}(q) \leq \max\{2q^2 \frac{S_G S_1^2}{2^{2\omega}}, 2q \frac{S_G T_1}{2^\omega}, q \frac{S_G S_2}{2^\omega}\}$
  - (d)  $\tilde{Adv}_{H^M}^{FixC}(q) \leq \max\{q^2(q-1) \frac{S_G S_1^2}{2^{2\omega}}, q(q-1) \frac{S_G T_1}{2^\omega}, q \frac{S_G S_2}{2^\omega}\}$
2. *If  $y = F(x_m, \cdot)$  is invertible then*
  - (a)  $\tilde{Adv}_{H^M}^{Pre}(q) = 1$
  - (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
  - (c)  $\tilde{Adv}_{H^M}^{FixP}(q) = 1$
  - (d)  $\tilde{Adv}_{H^M}^{FixC}(q) = 1$

**Corollary 4.** *Let  $G$  is Black Box Model and not invertible,  $z = H^M(m, x)$ ,  $\tilde{z} = H^W(m, x) = G(z)$ ,  $x$  and  $m_1, \dots, m_t$  are independent from each other, If  $\forall x_{m_0} \in$ ,  $y = F(x_{m_0}, x_h)$  is permutation,  $y = F(x_m, \cdot)$  is black box model then:*

1.  $\tilde{Adv}_{H^M}^{FixP}(q) \leq 2q \frac{T_F}{2^\omega}$
2.  $\tilde{Adv}_{H^M}^{FixC}(q) \leq q(q-1) \frac{T_F}{2^\omega}$

### 5.3 The Advantages of Improved Structures

**Theorem 18 (Ideal-Pipe Hash).**  *$\tilde{y} = f(m, x)$ ,  $\tilde{y} = \tilde{G}(f(m, x), x_h)$ ,  $y = F(x_m, x_h)$ , for  $z = H^M(m, x)$ ,  $\tilde{z} = H^I(f(m, x), z)$ ,  $m = m_t || \dots || m_1 \in I_{\kappa, t}$ ,  $m$  and  $x$  are independent from each other, if  $f(m, x)$  is independent from  $x, m$  and  $z$  then:*

<sup>5</sup> Where we did pre-computation of  $H^M(m_0^{(l)}, IV)$ ,  $l \geq 1$ , until get  $l'$  and  $l''$  with  $H^M(m_0^{(l')}, IV) = H^M(m_0^{(l'')}, IV)$ , for a selected  $m_0$ .

1. If  $y = F(x_m, x_h)$  is Black Box Model or  $y = F(x_m, \cdot)$  is black box model then:

- (a)  $\tilde{Adv}_{H^M}^{Pre}(q) \leq \max\{2q \frac{T_f T_{\bar{G}}}{2^\kappa}, q \frac{S_f T_{\bar{G}}}{2^n}\}$
- (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = \max\{q(q-1) \frac{T_f T_{\bar{G}}}{2^\kappa}, q \frac{S_f T_{\bar{G}}}{2^n}\}$
- (c)  $\tilde{Adv}_{H^M}^{FixP}(q) \leq \max\{2q \frac{T_f T_{\bar{G}}}{2^\kappa}, q \frac{S_f T_{\bar{G}}}{2^n}\}$
- (d)  $\tilde{Adv}_{H^M}^{FixC}(q) \leq \max\{q(q-1) \frac{T_E}{2^\kappa}, q \frac{S_E}{2^n}\}$

2. If  $y = F(x_m, \cdot)$  is invertible then

- (a)  $\tilde{Adv}_{H^M}^{Pre}(q) = 1$
- (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
- (c)  $\tilde{Adv}_{H^M}^{FixP}(q) = 1$
- (d)  $\tilde{Adv}_{H^M}^{FixC}(q) = 1$

**Theorem 19 (3C Hash).** Let  $z = H^m(m, x)$ ,  $z \in I_n, \tilde{z} = H^N(m, x)$ ,  $m = m_t \parallel \dots \parallel m_1 \in I_{\kappa \cdot t}$ ,  $h_i = F(m_i, h_{i-1})$ ,  $h_0 = x$  and  $m_1, \dots, m_t$  are independent from each other, if  $f(m, x) = h_* \oplus \dots \oplus h_1 \oplus h_0$  is independent from  $x$  and  $m$  then:

1. If  $y = F(x_m, x_h)$  is Black Box Model or  $y = F(x_m, \cdot)$  is black box model then:

- (a)  $\tilde{Adv}_{H^M}^{Pre}(q) \leq \max\{2q \frac{T_f T_{\bar{G}}}{2^\kappa}, q \frac{S_f T_{\bar{G}}}{2^n}\}$
- (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = \max\{q(q-1) \frac{T_f T_{\bar{G}}}{2^\kappa}, q \frac{S_f T_{\bar{G}}}{2^n}\}$
- (c)  $\tilde{Adv}_{H^M}^{FixP}(q) \leq \max\{2q \frac{T_f T_{\bar{G}}}{2^\kappa}, q \frac{S_f T_{\bar{G}}}{2^n}\}$
- (d)  $\tilde{Adv}_{H^M}^{FixC}(q) \leq \max\{q(q-1) \frac{T_E}{2^\kappa}, q \frac{S_E}{2^n}\}$

2. If  $y = F(x_m, \cdot)$  is invertible then

- (a)  $\tilde{Adv}_{H^M}^{Pre}(q) = 1$
- (b)  $\tilde{Adv}_{H^M}^{Coll}(q) = 1$
- (c)  $\tilde{Adv}_{H^M}^{FixP}(q) = 1$
- (d)  $\tilde{Adv}_{H^M}^{FixC}(q) = 1$



## Part III

# The Feistel Hash Constructions



## 6 The Feistel Constructions

A Feistel structure is a general way of constructing block ciphers from simple functions. The original idea was used in the block cipher, invented by Horst Feistel[22]. The security of the Feistel structure is not obvious, but analysis of DES[23] has shown that it is a good way to construct ciphers. And some new ciphers based on Feistel structure of SPN function have been discussed recently and no weakness is found in Feistel structure itself.

### 6.1 The Feistel Compression Function

Let  $\Psi(f_k)(x' \| x) = y' \| y \stackrel{def}{\Leftrightarrow} \begin{cases} y' = x \\ y = x' \oplus f_k(x) \end{cases}$ ,  $x', x, y', y, k \in I_n$ , the  $R$  round

Feistel structured block cipher is  $E^{Fe} : I_n \times I_{2n} \rightarrow I_{2n}$ ,  $E^{Fe}(k, x' \| x) \triangleq \Psi(f_{k(R)}) \circ \Psi(f_{k(R-1)}) \circ \dots \circ \Psi(f_{k(1)})(x' \| x)$ .

A  $R'$  round SPN structured block cipher is defined as  $E^{Sp} : I_n \times I_n \rightarrow I_n$ ,  $E^{Sp}(k, x) \triangleq f_{k(R')} \circ f_{k(R'-1)} \circ \dots \circ f_{k(1)}(x)$ , with same compression function and key schedule as Feistel compression function.

**Definition 11 (FL Structure).** Let function  $F_c : I_n \times I_n \rightarrow I_n$ , if  $F_c(k, x) = (\Psi(f_{k(R)}) \circ \dots \circ \Psi(f_{k(2)})(x \| f_{k(1)}(x)))^R$ , then we call function  $F_c$  is FL-Function (Feistel Like Structured function) and the structure of such function is called FL-Structure(Feistel Like Structure).

In fact, the FL-structure is a Feistel structure with left half bits input of first round is all zero and output the only right half bites of last round output. Fig8 gives the structure of Feistel construction and Fl-construction.

**Definition 12 (Feistel Compression Function).** IF the function  $F_c$  is used as compression function of iterated hash with format  $y = F_c(x_m, x_h)$ , where  $x_h$  is chaining value, we call such function is Feistel Compression Function.

### 6.2 Feistel Constructions

For M-D hash  $z = H^M(m, x)$ , may exist some  $m$  with  $P_{Z|M=m}(z_0) = 1$ , which means  $H^M(m \| m', IV)$  is constant, although if exist such  $m$ , it is still difficult to find and the  $|m|$  may be too big to be used as message, we still have to prohibit it. In this subsection we give a new failure tolerant structure of M-D hash with Feistel compression function.

**Definition 13 (F-Hash).** Let  $H^F : I_{n.*} \times I_n \rightarrow I_n$ , F-Hash  $\tilde{z} = H^F(m, x)$ ,  $m \in I_{n.t}$  is defined as:

$$\begin{aligned} h_0 &= x \\ h_i &= F_c(m_i, h_{i-1}), & (i = 1, \dots, t) \\ h'_i &= F_{c-1}(m_i, h_{i-1}), & (i = 1, \dots, t) \\ \tilde{z} &= E^{Sp}(\bigoplus_{i=1}^t h'_i, h_t) \end{aligned}$$

In fact, we have  $h'_i || h_i = E^{Fe}(m_i, \tilde{0} || h_{i-1})$ , ( $i = 1, \dots, t$ ). The figure illustration of F-Hash is given in Fig9, the security is discussed in next section.

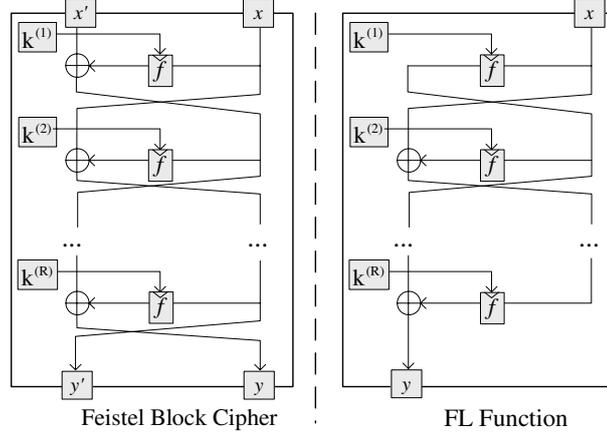


Fig. 8. Contrast Between Feistel Structure and FL-Structure

**Definition 14 (F1-MAC).** Let  $M^{F1} : I_n \times I_{n \cdot t} \times I_n \rightarrow I_n$ , F-MAC is defined as  $\tilde{z} = M^{F1}(k, m, x)$ ,  $m \in I_{n \cdot t}$  :

$$\begin{aligned} h_0 &= x \\ h'_i || h_i &= E^{Fe}(k, m_i || h_{i-1}), \quad (i = 1, \dots, t) \\ \tilde{z} &= E^{Sp}\left(\bigoplus_{i=1}^t h'_i, h_t\right) \end{aligned}$$

**Definition 15 (F2-MAC).** Let  $M^{F2} : I_n \times I_{n \cdot t} \times I_n \rightarrow I_n$ , F-MAC is defined as  $\tilde{z} = M^{F2}(k, m, x)$ ,  $m \in I_{n \cdot t}$  :

$$\begin{aligned} h_0 &= x \\ h'_i || h_i &= E^{Fe}(k, m_i || h_{i-1}), \quad (i = 1, \dots, t) \\ \tilde{z} &= E^{Sp}(h'_t, h_t) \end{aligned}$$

The advantage of F-MAC are that:

- The F-MAC can prohibit the weak key( Exist  $m_0, k_0$  and  $z_0$  make the probability with  $P_{\tilde{Z}|M=m_0, K=k_0}(z_0)$  near to 1, where  $z_0 = M^F(k_0, m_0, x)$ , like M-D hash);
- In some protocol, the  $x$  can be selected as group key(each group with same key);
- Both  $k$  and  $x$  can be seen as key.

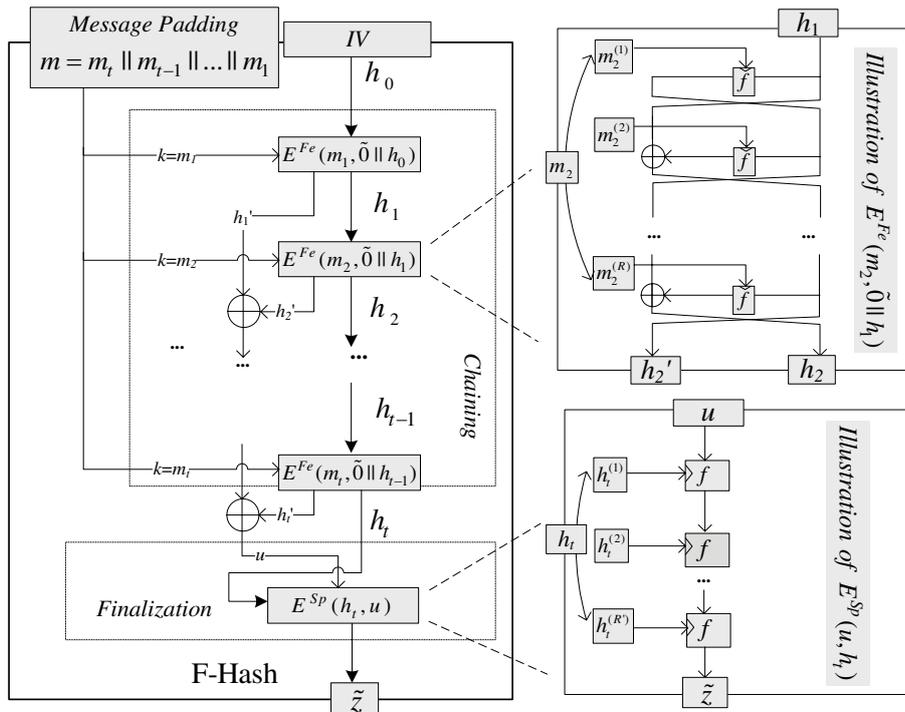


Fig. 9. F-Hash Function

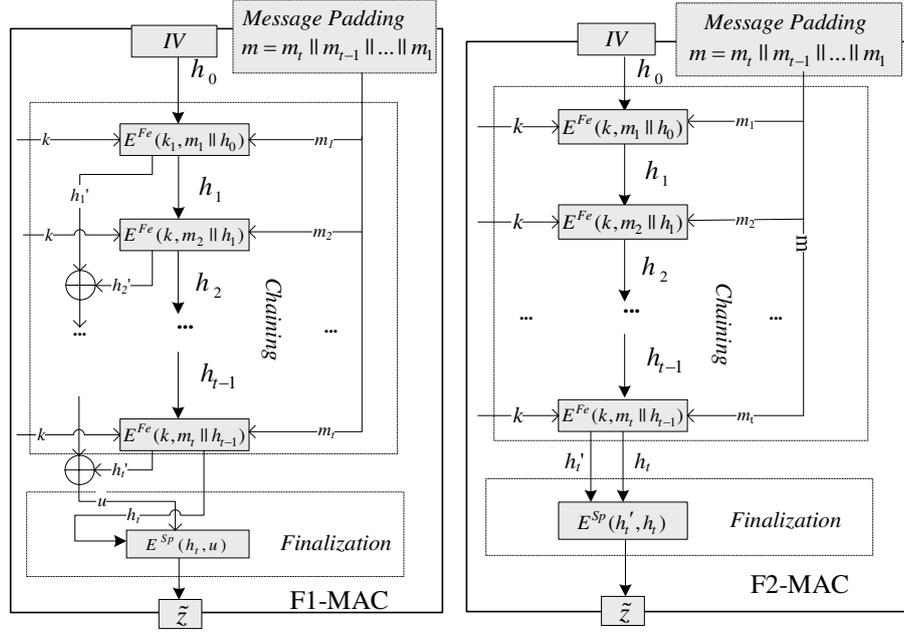


Fig. 10. F1-MAC and F2-MAC

**Definition 16 (FBC Mode).** Let  $E^{FBC} : I_n \times I_{n-t} \times I_n \rightarrow I_{n \cdot (t+1)}$ ,  $c = E_{FBC}(k, m, x)$  is defined as:

$$\begin{aligned} h_0 &= x \\ c_i \parallel h_i &= E^{Fe}(k, m_i \parallel h_{i-1}), \quad (i = 1, \dots, t) \\ c &= h_t \parallel c_t \parallel c_{t-1} \parallel \dots \parallel c_1 \end{aligned}$$

In FBC mode the ciphertext of Previous encryption is added to subsequent one encryption, which making the following ciphertext appears "randomly", that mode is similar as CBC mode, which inherent all advantage of CBC mode and have some improvements on it.

- the CBC mode requires the communication sides hold the initial value  $x$ , the FBC mode does not requirement of that.
- when the  $x$  is randomized, all the ciphertext be randomized.
- the fixed initial value  $x$  can be used as authentication.
- the encryption and decryption with same structure with different key schedule.

The drawback of that mode is the ciphertext has one more block than the plaintext.

**Theorem 20.** Every key recovery attack on FBC mode can be convert to known plaintext key recovery attack on the Feistel block cipher.

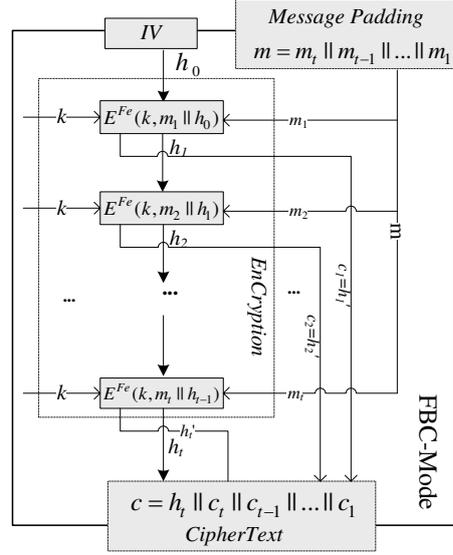


Fig. 11. FBC Encrypt Mode

### 6.3 Security Proof of the Constructions

The securities of F-Hash, F-MAC are partly based on security of Feistel Block cipher, we give following assumption for  $E^{Fe}$ .

#### Secure of Feistel Compression Function

**Assumption 2** For  $E^{Fe}$  and  $E^{Sp}$ :

- For each constant key, the encryption procedure and decryption procedure can be seen as invertible Black Box Model.
- The best way to find the weak key ( $E^{Fe}(k, x || x) = E^{Fe}(k', x' || x)$ ) of  $E^{Fe}$  and  $E^{Sp}$  is exhaustive key search attack based on birthday paradox;
- No weakness are found in  $E^{Fe}$  and  $E^{Sp}$ ;

**Assumption 3** Let the key schedule algorithm of  $E^{Fe}$  is  $\psi(k)$  with  $k^{(i)} = \psi(k^{(i-1)})$ ,  $k^{(0)} = k$ , if the  $E^{Fe}$  satisfy Assumption2, then the  $E^{Fe}$  with key schedule algorithm  $k^{(i)} \triangleq \psi(k^{(i-1)}) \oplus k$ ,  $k^{(0)} = k$  still satisfy Assumption2.

*Remark 9.* In the prove of security of Feistel structure[33–36], the compression function is assumed as pseudo random function, then  $f_{\psi(k) \oplus k}$  is still a pseudo random function. The Assumption3 implies the key schedule algorithm  $\psi(k)$  has properties of does not exit  $i, j$  with  $\psi(k_{\{j\}}^{(i)}) \equiv k_{\{j\}}$ .

**Theorem 21.** Let for  $E^{Fe}$ , the round function  $f$  with format  $f(k^{(i)}, x) = f(x \oplus k^{(i)})$ , key schedule algorithm  $\varphi(k)$  is not a linear transformation, if there exist  $E^{Fe}$  satisfy Assumption1, then excising  $\tilde{F}_c$  that can be seen as a not invertible Black Box Model.

*Proof.* Let  $E^{Fe}$  with rounds  $r$ , and round function  $f(k^{(i)}, x) = f(x \oplus k^{(i)})$  then

$$E^{Fe}(k_0, x' \| x)^R = \tilde{F}_c(x', x \oplus x') \oplus x' \quad (1)$$

the key schedule algorithm of  $\tilde{F}_c$  is  $x'^{(i)} = \psi(k_0^{(i)}) \oplus x'$ , where the key schedule algorithm of  $E^{Fe}$  is  $\psi(k)$ , the prove of Eq.1 is given as bellow.

When  $r = 1$  :

$$\begin{aligned} (\Psi(f_{k_0^{(1)}})(x' \| x))^R &= x' \oplus f(x \oplus \psi(k_0^{(1)})) \\ &= \tilde{0} \oplus f((x \oplus x') \oplus (\psi(k_0^{(0)}) \oplus x')) \oplus x' \\ &= (\Psi(f_{k_0^{(1)} \oplus x'})(\tilde{0} \| x \oplus x'))^R \oplus x' \end{aligned}$$

Let assume  $r < k$  the equation is true then:

$$\begin{aligned} &(\Psi(f_{k_0^{(r)}} \circ \dots \circ f_{k_0^{(1)}})(x' \| x))^R \\ &= (\Psi(f_{k_0^{(r-2)} \oplus x'} \circ \dots \circ f_{k_0^{(1)} \oplus x'}) (\tilde{0} \| x \oplus x'))^R \oplus x' \\ &\quad \oplus f((\Psi(f_{k_0^{(r-1)} \oplus x'} \circ \dots \circ f_{k_0^{(1)} \oplus x'}) (\tilde{0} \| x \oplus x'))^R \oplus x' \oplus k_0^{(r)}) \\ &= (\Psi(f_{k_0^{(r-2)} \oplus x'} \circ \dots \circ f_{k_0^{(1)} \oplus x'}) (\tilde{0} \| x \oplus x'))^R \\ &\quad \oplus f((\Psi(f_{k_0^{(r-1)} \oplus x'} \circ \dots \circ f_{k_0^{(1)} \oplus x'}) (\tilde{0} \| x \oplus x'))^R \oplus (x' \oplus k_0^{(r)})) \oplus x' \\ &= (\Psi(f_{k_0^{(r)} \oplus x'} \circ \dots \circ f_{k_0^{(1)} \oplus x'}) (\tilde{0} \| x \oplus x'))^R \oplus x' \end{aligned}$$

And also we have

$$F_c(k, x) = \tilde{E}^{Fe}(k, x'_0 \| x \oplus x'_0)^R \oplus x'_0 \quad (2)$$

where the key schedule algorithm of  $\tilde{E}^{Fe}$  is  $x'_0{}^{(i)} = \psi(k^{(i)}) \oplus x'_0$ . From Eq.1, we know if  $E^{Fe}$  satisfy item 1 of Assumption2, then  $\tilde{F}_c$  is a Black Box Model. From Eq.2, if exist  $\tilde{E}^{Fe}$  satisfy Assumption2, then  $F_c$  is Black Box Model. If for each key,  $y' \| y = E^{Fe}(x' \| x)$  is invertible Black Box Model, then it is no advantage to select  $y'$  make the  $x' = \tilde{0}$ , so  $F_c$  is not invertible Black Box Model.  $\square$

Theorem21 implies existing Feistel compression function with Black Box Model, when existing  $E^{Fe}$  satisfy assumption, 2.

**Lemma 5.** For  $F_c$  and  $E^{Fe}$ :

$$- T_1 \triangleq \max_{x_{m_0}, y_0} \#\{(y_0, x_{m_0}, x_h)\}^{F_c} \equiv \max_{x_{m_0}, y_0} \#\{(y' \| y_0, x_{m_0}, \tilde{0} \| x_h)\}^{E^{Fe}};$$

$$\begin{aligned}
- T_2 &\triangleq \max_{x_{h_0}, y_0} \#\{(y_0, x_m, x_{h_0})\}^{F_c} \equiv \max_{x_{h_0}, y_0} \#\{(y' \| y_0, x_m, \tilde{0} \| x_{h_0})\}^{E^{F_e}}; \\
- T_3 &\triangleq \max_{y_0} \#\{(y_0, x_m, x_h)\}^{F_c} \equiv \max_{y_0} \#\{(y' \| y_0, x_m, \tilde{0} \| x_h)\}^{E^{F_e}}.
\end{aligned}$$

Let

$$\begin{aligned}
- S_1 &\triangleq \max_{x_{m_0}, y_0} \#\{(y_0, x_{m_0}, x_h)\}^{F_{c-1}}; \\
- S_2 &\triangleq \max_{x_{h_0}, y_0} \#\{(y_0, x_m, x_{h_0})\}^{F_{c-1}}; \\
- S_3 &\triangleq \max_{y_0} \#\{(y'_0 \| y_0, x_m, x'_{h_0} \| x_{h_0})\}^{E^{F_e}}.
\end{aligned}$$

### The Security of F-Hash

**Theorem 22.** For F-Hash  $\tilde{z} = H^F(m, x)$ ,  $x \in I_n$ ,  $m = m_* \| \dots \| m_0$ ,  $\tilde{z} = E^{Sp}(u, z)$ ,  $z = H^M(m, x) = h_*$ ,  $h_0 = x$ ,  $O_h(m, x) \triangleq u = \bigoplus_{i=1}^t h'_i$ ,  $h_i = F_c(m_i, h_{i-1})$ ,  $h'_i = F_{c-1}(m_i, h_{i-1})$ ,  $m_i$  are independent from each other,  $x$  and  $m$  are independent and uniformly distributed in  $I_n$  and  $\bigcup_{i=1}^t I_{n-i}$ , respectively, if  $u = O_h(m, x)$  and  $z = H^M(m, x)$  are independent,  $h'_i, i \in [1, *]$  are independent from each other then :

- 1:  $P_{\tilde{Z}|M=m}(z) \leq 2^{-n} T_1^{\frac{|m|}{n}}$ ;
- 2:  $P_{\tilde{Z}|X=x}(z) \leq 2^{-n} T_2$ ;
- 3:  $P_{\tilde{Z}|M=m}(\tilde{z}) \leq 2^{-n} S_1$ ;
- 4:  $P_{\tilde{Z}|X=x}(\tilde{z}) \leq 2^{-n} S_2$ ;

*Proof.* The Proof is given by deduction theory.

1. When  $t = 1$ :

$$\begin{aligned}
P_{\tilde{Z}|M=m}(z) &\leq \max_{m_0, z_0} \sum_{x \in I_n} P_X(x) \chi_{F_c(m_0, x)}(z_0) \\
&= \max_{m_0, z_0} \sum_{i \in [1, 2^n]} 2^{-n} \#\{(m_0, x_i, z_0)\}^{F_c} \leq 2^{-n} T_1
\end{aligned}$$

Suppose  $t < l$  the inequality is true, when  $t = l$ :

$$\begin{aligned}
P_{\tilde{Z}|M=m}(z) &= P_{\tilde{Z}|M=\mathbf{m}_l \| m'}(z) \\
&= \sum_{x \in I_n} P_X(x) \chi_{F_c(\mathbf{m}_l, H^M(m, x))}(z) \\
&= \sum_{x \in I_n} \sum_{u \in I_n} \frac{1}{2^n} \cdot \chi_{F_c(\mathbf{m}_l, u)}(z) \cdot \chi_{H^M(m', x)}(u)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{u \in I_n} \sum_{x \in I_n} \frac{1}{2^n} \cdot \chi_{F_c(\mathbf{m}_l, u)}(z) \cdot \chi_{H^M(m', x)}(u) \\
&= \sum_{u \in I_n} (\chi_{F_c(\mathbf{m}_l, u)}(z) \cdot \sum_{x \in I_n} \frac{1}{2^n} \chi_{H^M(m', x)}(u)) \\
&= \sum_{u \in I_n} \chi_{F_c(\mathbf{m}_l, u)}(z) \cdot P_{\dot{U}|M'=m'}(u) \\
&\leq 2^{-n} T_1^{l-1} \sum_{u \in I_n} \chi_{F_c(\mathbf{m}_l, u)}(z) \\
&\leq 2^{-n} T_1^{l-1} T_1 = 2^{-n} T_1^l
\end{aligned}$$

2. When  $t = 1$

$$\begin{aligned}
P_{\dot{Z}|X=x}(z) &\leq \max_{x_0, z_0} \sum_m P_M(m) \chi_{F_c(m, x_0)}(z_0) \\
&= \max_{x_0, z_0} \sum_i 2^{-n} \#\{(\mathbf{m}_i, x_0, z_0)\}^{F_c} \leq 2^{-n} T_2
\end{aligned}$$

When  $t > 1$ :

$$\begin{aligned}
P_{\dot{Z}|X=x}(z) &= \sum_{m \in \cup_{i=1}^l I_{n,i}} P_M(m) P_{\dot{Z}|M=m, X=x}(z) \\
&= \sum_{\mathbf{m}_l \in I_n} \sum_{m' \in \cup_{i=1}^{l-1} I_{n,i}} P_{M'}(m') P_{M_l}(\mathbf{m}_l) \\
&\quad P_{\dot{Z}|M=m, X=x}(P(z = F_c(\mathbf{m}_l, H^M(m', x)))) \\
&= \sum_{\mathbf{m}_l \in I_n} \sum_{m' \in \cup_{i=1}^{l-1} I_{n,i}} \sum_{u \in I_n} \\
&\quad P_{M'}(m') P_{M_l}(\mathbf{m}_l) \cdot \chi_{F_c(\mathbf{m}_l, u)}(z) \cdot \chi_{H^M(m', x)}(u) \\
&= \sum_{u \in I_n} \sum_{\mathbf{m}_l \in I_n} P_{M_l}(\mathbf{m}_l) \cdot \chi_{F_c(\mathbf{m}_l, u)}(z) \\
&\quad \sum_{m' \in \cup_{i=1}^{l-1} I_{n,i}} P_{M'}(m') \cdot \chi_{H^M(m', x)}(u) \\
&= \sum_{u \in I_n} P_{\dot{Z}|U=u}(z) P_{\dot{U}|X=x}(u) \\
&\leq 2^{-n} T_2 \sum_{u \in I_n} P_{\dot{U}|X=x}(z) = 2^{-n} T_2
\end{aligned}$$

3.  $\forall t \geq 1$ :

$$P_{\tilde{Z}|M=m}(\tilde{z}) = P_{\tilde{Z}|M=m}(\tilde{z} = E^{Sp}(u, z), u = O_h(m, x), z = H^M(m, x))$$

$$\begin{aligned}
&= \sum_{x,u,z \in I_n} P_X(x) \chi_{E^{Sp}(z,u)}(\tilde{z}) \chi_{H^M}(z,m,x) \chi_{O_h(m,x)}(u) \\
&\quad u \text{ and } z \text{ are independent, and } P_X(x) = 2^{-n} \\
&= \sum_{u,z \in I_n} \chi_{E^{Sp}(z,u)}(\tilde{z}) \sum_{x \in I_n} P_X(x) \chi_{H^M}(z,m,x) \\
&\quad \sum_{x \in I_n} P_X(x) \chi_{O_h(m,x)}(u) \\
&= \sum_{u,z \in I_n} \chi_{E^{Sp}(z,u)}(\tilde{z}) P_{\dot{U}|M=m}(u) P_{\dot{Z}|M=m}(z) \\
&\leq \max_{u_0} P_{\dot{U}|M=m}(u_0) 2^n \sum_z P_{\dot{Z}|M=m}(z) \sum_u 2^{-n} \chi_{E^{Sp}(z,u)}(\tilde{z}) \\
&= \max_{u_0} P_{\dot{U}|M=m}(u_0) 2^n \sum_z P_{\dot{Z}|M=m}(z) P_{\tilde{Z}|Z=z}(\tilde{z}) \\
&\leq \max_{u_0} P_{\dot{U}|M=m}(u_0) \max_{z_0, \tilde{z}_0} 2^n P_{\tilde{Z}|Z=z_0}(\tilde{z}_0) \sum_z P_{\dot{Z}|M=m}(z) \\
&= \max_{u_0} P_{\dot{U}|M=m}(u_0)
\end{aligned}$$

$$\begin{aligned}
P_{\dot{U}|M=m}(u) &= \sum_{x \in I_n} P_X(x) P_{U|M=m, X=x}(u = h'_1 \oplus \bigoplus_{i=2}^t h'_i) \\
&= \sum_{x \in I_n} P_X(x) P_{U|M=m', \|m_1, X=x}(u = v \oplus h'_1, v = \bigoplus_{i=2}^t h'_i) \\
&= \sum_{x \in I_n} \sum_{v \in I_n} P_X(x) P_{UV|M=m', \|m_1, X=x}(u = v \oplus h'_1, v = \bigoplus_{i=2}^t h'_i) \\
&= \sum_{x \in I_n} \sum_{v \in I_n} P_X(x) P_{U|M_1=\mathbf{m}_1, X=x}(u = h'_1 \oplus v) \\
&\quad P_{V|M'=m', X=x}(v = \bigoplus_{i=2}^t h'_i) \\
&= \sum_{v \in I_n} P_{U|M_1=\mathbf{m}_1}(u = h'_1 \oplus v) P_{V|M'=m'}(v = \bigoplus_{i=2}^t h'_i) \\
&= \max P_{U|M_1=\mathbf{m}_1}(u) \sum_v P_{V|M'=m'}(v = \bigoplus_{i=2}^t h'_i) \leq \frac{S_1}{2^n}
\end{aligned}$$

4.  $\forall t \geq 1$ :

$$\begin{aligned}
P_{\tilde{Z}|X=x}(\tilde{z}) &= P_{\tilde{Z}|X=x}(\tilde{z} = E^{Sp}(u,z), u = O_h(m,x), z = H^M(m,x)) \\
&= \sum_{u,z \in I_n} \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\tilde{Z}|\dot{U}=u, Z=z}(\tilde{z})
\end{aligned}$$

$$\begin{aligned}
& P_{\dot{U}, \dot{Z}|M=m, X=x}(u = O_h(m, x), z = H^M(m, x)) \\
& \text{Since } P_M(x) = 2^{-\sum_i^t i \cdot n} \text{ and } u, z \text{ are independent} \\
& = \sum_{u, z \in I_n} \chi_{E^{Sp}(z, u)}(\tilde{z}) \\
& \quad \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\dot{U}|M=m, X=x}(u = O_h(m, x)) \\
& \quad \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{\dot{Z}|M=m, X=x}(z = H^M(m, x)) \\
& = \sum_{u, z \in I_n} \chi_{E^{Sp}(z, u)}(\tilde{z}) P_{\dot{U}|X=x}(u) P_{\dot{Z}|X=x}(z) \\
& \leq \max_{u_0} P_{\dot{U}|X=x}(u_0) 2^n \sum_z P_{\dot{Z}|X=x}(z) \sum_u 2^{-n} P_{\tilde{Z}|U=u, Z=z}(\tilde{z}) \\
& = \max_{u_0} P_{\dot{U}|X=x}(u_0) 2^n \sum_z P_{\dot{Z}|X=x}(z) P_{\tilde{Z}|Z=z}(\tilde{z}) \\
& \leq \max_{u_0} P_{\dot{U}|X=x}(u_0) \max_{z_0, \tilde{z}_0} 2^n P_{\tilde{Z}|Z=z_0}(\tilde{z}_0) \sum_z P_{\dot{Z}|X=x}(z) \\
& = \max_{u_0} P_{\dot{U}|X=x}(u_0)
\end{aligned}$$

$$\begin{aligned}
P_{\dot{U}|X=x}(u) & = \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{U|M=m, X=x}(u = h'_1 \oplus \bigoplus_{i=2}^t h'_i) \\
& = \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{U|M=m', \|m_1, X=x}(u = v \oplus h'_1, v = \bigoplus_{i=2}^t h'_i) \\
& = \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} P_M(m) P_{UV|M=m', \|m_1, X=x}(u = v \oplus h'_1, V = \bigoplus_{i=2}^t h'_i) \\
& = \sum_{m \in \cup_{i=1}^t I_{n \cdot i}} \sum_{v \in I_n} P_M(m) P_{U|M_1=\mathbf{m}_1, X=x}(u = h'_1 \oplus v) \\
& \quad P_{V|M'=m', X=x}(v = \bigoplus_{i=2}^t h'_i) \\
& = \sum_{v \in I_n} \sum_{\mathbf{m}_t \in I_n} P_{M_t}(\mathbf{m}_t) P_{U|M_1=\mathbf{m}_1, X=x}(u = h'_1 \oplus v) \\
& \quad \sum_{m' \in \cup_{i=1}^{t-1} I_{n \cdot i}} P_{V|M'=m', X=x}(v = \bigoplus_{i=2}^t h'_i) \\
& = \sum_{v \in I_n} P_{U|X=x}(u = h'_1 \oplus v) P_{V|X=x}(v = \bigoplus_{i=2}^t h'_i)
\end{aligned}$$

$$= \max_{u_0} P_{U|X=x}(u_0) \sum_v P_{V|X=x}(v = \bigoplus_{i=2}^t h'_i) \leq \frac{S_2}{2^n}$$

□

*Remark 10.* In the proof of Theorem22, we use the assumption of  $h'_i, i \in [1, *]$  being independent from each other, which seems is pretty strong assumption, in fact if the compression function is Black Box Model, which means  $h'_{i-1}$  give no information about  $h'_i$ , if the compression function is build with pseudo random function then the assumption can be happen in true design.

**Theorem 23.** Let  $F$ -Hash  $\tilde{z} = H^F(m, x)$  satisfying the assumptions and notations of Theorem22,  $F_c$  and  $E^{Sp}$  are Black Box Model then:

- $\tilde{Adv}_{H^F}^{F_i x P}(q) \leq q \cdot 2^{-n} \max\{2^{-n} S_1, 2^{-n} S_2\};$
- $\tilde{Adv}_{H^F}^{F_i x C}(q) \leq q^2 \cdot 2^{-n} \max\{2^{-n} S_1^2, 2^{-n} S_2^2\}.$

*Proof.* The proof needs  $H^F$  is Black Box Model:

- 1: Since the recovery of  $E^{Sp^{-1}}(\tilde{z}_0)$  is cipher only attack, and the  $F_c$  is not invertible, so the  $H^F$  is not invertible.
- 2: For  $x_0$  if exist  $H^F(m_t, x_0)$  give information about selection of  $m'_t$ , from  $E^{Sp}$  is Black Box Model, we get  $H^M(m_t, x_0)$  gives information about selection of  $m'_t$ , from  $H^M(m_t, x_0)$  is finite iteration, and  $m'_t$  has finite block, we get exist  $\mathfrak{m}_i \subseteq m_t$  and  $\mathfrak{m}'_j \subseteq m'_t$  with  $\mathfrak{m}'_j$  is influenced by  $\mathfrak{m}_i$ . but  $F_c$  is Black Box Model and that is imposible. □

**Security about F-MAC and FBC Mode** The security of F1-MAC, F2-MAC and FBC mode can be discussed similar as F-Hash, since the condition probability of  $F - Hash$  is given, we can give the security prove of the MACs and FBC mode. The security of F2-MAC can also be discussed similar as CBC-MAC[1] and the security of FBC mode is similar as that of CBC mode[3], and prohibit the attack based on fixed IV[3]. More precise discussion and true attacks should be based on the assumption of round function  $f$  and key schedule algorithm  $\psi$ , this paper only give the proof of security of the structure.

#### 6.4 Discussion

**The Value of T1, T2, S1, S2** Let  $g : I_{2n} \rightarrow I_{2n}, y || y' = g(x || x')$  is a random permutation, then we have[3]:

$$P_{Y'|X'=x_0}(y = g(x_0)) 2^{-2n}$$

then  $y = (g(x'_0 || x))^R$  is random function, let  $\mathfrak{f}(x_0, x) \triangleq (g(x'_0 || x))^R$ :

$$P_{Y^R|X=x_0}(y = \mathfrak{f}(x'_0 || x_0)) = 2^{-n}$$

then we have[3]

$$P_{Y^R|X_1=x_1, X_2=x_2}(y = \mathbf{f}(x'_0||x_1), y = \mathbf{f}(x'_0||x_2)) = \begin{cases} 2^{-2n} & x_1 \neq x_2 \\ 2^{-n} & x_1 = x_2 \end{cases}$$

In block cipher  $E^{Fe}$ , for each fixed key, if we can not distinguish the  $E^{Fe}$  from Pseudo random permutation, then we have

$$P(T_1 = k) = 2^{-k \cdot n} 2^n, \quad P(S_1 = k) = 2^{-k \cdot n} 2^n, \quad k \in \mathbf{N}$$

If the  $F_c$  is selected as Equation1, then we have

$$P(T_2 = k) = 2^{-k \cdot n} 2^n, \quad P(S_2 = k) = 2^{-k \cdot n} 2^n, \quad k \in \mathbf{N}$$

If for each  $x'_0||x_0$ , we can not distinguish  $E^{Fe}(k, x'_0||x_0)$  from random function then we have:

$$P(T_2 = k) = 2^{-k \cdot n} 2^n, \quad P(S_2 = k) = 2^{-k \cdot n} 2^n, \quad k \in \mathbf{N}$$

**Round function and Key Schedule Algorithm** In the prove of Theorem21, we find the  $x'$  can be moved into key schedule algorithm and the whole discussion we assume the round function  $f$  is permutation. The most common design of round function with permutation is SPN structure. The SP structure is used in Feistel structure can result in linear part can be moved into previous rotund or posterior round[29], so we prefer the round function with SPS(SBox-Linear part-Sbox) structure.

The key schedule algorithm  $\psi$  is assumed as not a linear transformation, we prefer the key schedule algorithm itself is pseudo random function, which has been discussed in PHD paper of Rijmen[49].

## 6.5 Attacks on F-Hash

**Multi Collision[27]** Suppose the multi collision is possible, for each inner collision  $H^M(m_{i+1}, H^M(m_i||\dots||m_1, x_0)) = H^M(m'_{i+1}, H^M(m'_i||\dots||m'_1, x_0))$ ,  $i \in [1, t]$ , if the inner collision can make true collision requires  $O_h(m, IV) = O_h(m', IV)$ , that is not always hold when the inner collision is occur. In fact that will happen with high probability when  $|m_i| = n$ .

**Extension Attack[47]** If the extension collision is possible, when exist inner collision  $H^M(m, x_0) = H^M(m', x_0)$ , the extension should be with  $O_h(m''||m, IV) = O_h(m''||m', IV)$ , the complexity of finding  $O_h(m''||m, IV) = O_h(m''||m', IV)$  is  $\mathcal{O}(2^{\sqrt{n}})$ , when the collision is final collision  $H^F(m, x) = H^F(m', x)$ , not a inner collision, the extension attack is impossible.

**Fixed Point Attack** The requirement on success of fixed point attack is similar as that multi collision attack, which requires  $O_h(m, IV) = O_h(m', IV)$  and the fixed block length should be  $|m_i| = n$ .

## 7 Random Oracle and Conditional Probability

The random oracle model has been introduced by Bellare and Rogaway as a "paradigm for designing efficient protocols"[2]. It assumes that all parties, including the adversary, have access to a public, truly random hash function  $H$ . This model has been proven extremely useful for designing simple, efficient and highly practical solutions for many problems[15]. From a theoretical perspective, it is clear that a security proof in the random oracle model is only a heuristic indication of the security of the system when instantiated with a particular hash function. In fact, many recent "separation" results[5, 13, 14, 20, 20, 32, 15] illustrated various cryptographic systems secure in the random oracle model but completely insecure for any concrete instantiation of the random oracle.

$x \xleftarrow{\$} A$  mean selecting a random value from  $A$ ,  $Func(D, R)$  be the family of all random functions of  $D$  to  $R$ ,  $Perm(D)$ , the family of all random permutations on  $D$ ,  $f \xleftarrow{\$} Func(D, R)$ ,  $\pi \xleftarrow{\$} Perm(D, R)$ , the probability is taken over a random choice of  $f$  from  $Func(I_l, I_L)$ , meaning that we have executed the operation  $f \xleftarrow{\$} Func(I_l, I_L)$  with properties of[3]:

1. If Fix  $x$  and  $y: P[f(x) = y] = 2^{-L}$ ;
2. If Fix  $x_1, x_2$  and  $y_1, y_2$ , and  $x_1 \neq x_2$  then:
  - (a)  $P[f(x_1) = y_1 | f(x_2) = y_2] = 2^{-L}$
  - (b)  $P[f(x_1) = y_1] = 2^{-L}$
  - (c)  $P[f_1(x_1) = y_1 | f_2(x_1) = y_1] = 2^{-L}$
3. If Fix  $x$  and  $y: P[\pi(x) = y] = 2^{-L}$
4. If Fix  $x_1, x_2$  and  $y_1, y_2$ , and  $x_1 \neq x_2$  then:

$$P[\pi(x_1) = y_1 | \pi(x_2) = y_2] = \begin{cases} \frac{1}{2^L - 1}, & y_1 \neq y_2 \\ 0, & y_1 = y_2 \end{cases}$$

$$P[\pi(x_1) = y_1] = 2^{-L}$$

$$P[\pi_1(x_1) = y_1 | \pi_2(x_1) = y_1] = 2^{-L}$$

If the  $f$  or  $\pi$  is true design function or permutation, then the properties of permutation is still hold, but for the compression function will not hold, it become:

1. exist  $x_0$  and  $y_0: P[f(x) = y] = \frac{T}{2^L}$ , where  $T > 1$ ;
2. If Fix  $x_1, x_2$  and  $y_1, y_2$ , and  $x_1 \neq x_2$  then:
  - (a)  $P[f(x_1) = y_1 | f(x_2) = y_2] = \frac{T}{2^L}$
  - (b)  $P[f(x_1) = y_1] = \frac{T}{2^L}$
  - (c)  $P[f_1(x_1) = y_1 | f_2(x_1) = y_1] = \frac{T}{2^L}$ .

Let  $H^M$  be a hash function with M-D construction, in random oracle model, the compression function  $F : I_n \times I_n \rightarrow I_n$  with  $S_F = 1$ , we have  $P_{\dot{Z}|M=m} \leq (S_F)^t 2^{-n} = 2^{-n}$ , then  $P_{\dot{Z}|M=m} = 2^{-n}$ , but in true design, if the  $F(\cdot, x_h)$  is not a permutation, we have  $S_F > 1$ , so in true design there may be exist a cluster, but the proof based on random oracle model, can not find the cluster.

Let function family  $\mathcal{F}(I_\kappa, I_n) : I_\kappa \times I_n \rightarrow I_n$ , for each  $x_m \in I_\kappa$ , define a function  $F_{x_m}(\cdot, x_h) : I_n \rightarrow I_n \in \mathcal{F}(I_\kappa, I_n)$ , if the function family  $\mathcal{F}(I_\kappa, I_n)$  is a set of pseudo random function, then we have:  $\forall F_{x_m}(\cdot, x_h) \in \mathcal{F}(I_\kappa, I_n)$  and  $\forall x, y \in I_n$  with  $P(F_{x_m}(\cdot, x) = y) = 2^{-n}$ , but for selected  $F_{x_m}(\cdot, x_h)$ , if we get  $x_2 = F_{x_m}(\cdot, x_1)$ , then  $P(x_2 = F_{x_m}(\cdot, x_2)) \neq 2^{-n}$ , which implies that even the compression function  $F(\cdot, x_h)$  is a random oracle model, we asking oracle  $x_m, x_{h_1}$  get  $x_{h_2} = F(x_m, x_{h_1})$ , then asking oracle  $x_m, x_{h_2}$ , then this model is not a random oracle model. From the analysis, the hash  $z = H^M(m, x)$  is not a random oracle model, if we can select the message, even the compression function set is pseudo random function, the conditional probability  $P_{\tilde{Z}|M=m} \leq S_F^t 2^{-n}$  just implies the output of  $H^M$  may not be uniformly distributed, no matter the compression function is Pseudo random function.

But, in 3C hash, Ideal-Pipe hash, we have  $P_{\tilde{Z}|M=m}(\tilde{z}) \leq S_F 2^{-n}$  implies that if the compression function is pseudo random function,  $S_F = 1 \Rightarrow P_{\tilde{Z}|M=m}(\tilde{z}) = 2^{-n}$ , then the hash function is a pseudo random function. So we have if the structure is secure based on conditional probability, then it be secure in random oracle model, but the structure is secure in random oracle model, may not be a secure structure in conditional probability, if use such structure, more discussions are required.

## 8 Conclusion

The securities of structures are only illustrated depend on conditional probability and maximum advantage, the security of ideal pipe hash against the specific attacks which have been illustrated in some known structures or known design are not given, we will fulfill this part.

**Acknowledgments** The paper include many areas in cryptography and we may not understand some of the reference papers very well, and may be have some misunderstanding, we hope any comments. And also we want to give thanks to all comments that were given, we will fulfill this part latter.

## References

1. M.Bellare, K.Pietrzak, and P.Rogaway, Improved Security Analyses for CBC MACs, In Advances in Cryptology Crypto 2005, LNCS 3621, pp.527-545, 2005.
2. M. Bellare and P. Rogaway, Random oracles are practical : a paradigm for designing efficient protocols. Proceedings of the First Annual Conference on Computer and Communications Security, ACM, 1993.
3. M.Bellare and P.Rogaway, Introduction to Modern Cryptography.
4. M.Bellare, R.Canetti, and H.Krawczyk. Keying hash functions for message authentication, In Advances in CryptologyCRYPTO'96, LNCS 1109, pp.1-15.
5. M. Bellare, A.Boldyreva and A.Palacio. An Uninstantiable Random-Oracle-Model Scheme for a Hybrid-Encryption Problem, In Advances in CryptologyE-CRYPTO'2004, LNCS 3027, pp.171-188.

6. M.Bellare and P.Rogaway, Code-Based Game-Playing Proofs and the Security of Triple Encryption, <http://eprint.iacr.org/2004/331.pdf>.
7. E.Biham. Recent advances in hash functions-the way to go. Presented at ECRYPT Conference on Hash Functions (Cracow, June 2005), see <http://www.ecrypt.eu.org/stvl/hfw/Biham.ps>.
8. E.Biham and A. Shamir, Differential cryptanalysis of DES-like cryptosystems, *Journal of Cryptology*, Vol.4, No.1, pp.3-72, 1991.
9. E.Biham and R.Chen. Near-Collisions of SHA-0, In *Advances in Cryptology CRYPTO'2004*, LNCS 3152, pp.290-305, 2004.
10. J.Black, P.Rogaway, and T.Shrimpton, "Black-box analysis of the block-cipher-based hashfunction constructions from PGV". In *Advances in Cryptology - CRYPTO'02*, volume 2442 of *Lecture Notes in Computer Science*. Springer-Verlag, 2002. pp.320-335.
11. J.Black, P.Rogaway, A Block-Cipher Mode of Operation for Parallelizable Message Authentication, In *Advances in Cryptology C Eurocrypt'02*, LNCS 2332, pp.384C397.
12. C.Chchin. Entropy Measures and Uncoditional Security in Cryptography, PHD thesis.
13. R. Canetti, O. Goldreich and S. Halevi, The random oracle methodology, revisited, *STOC98*, ACM, 1998.
14. R.Canetti, O.Goldreich and S.Halevi. On the random oracle methodology as applied to Length-Restricted Signature Schemes. In *Proceedings of Theory of Cryptology Conference*, pp. 40C57, 2004.
15. J.S.Coron, Y.Dodis, C.Malinaud, and P.Puniya. Merkle-damgard revisited: How to construct a Hash Function, In *Advances in CryptologyCRYPTO'05*, LNCS 3621, pp.430-448.
16. J.Daemen and V.Rijmen: *The Design of Rijndael: AES The Advanced Encryption Standard*. Springer, 2002.
17. J.Daemen and V. Rijmen, "A new MAC Construction Alred and a Specific Instance Alpha-MAC," , *Fast Software Encryption 2005*, LNCS H. Gilbert, H. Handschuh, Eds., Springer-Verlag, to appear.
18. I.Damgård. A design principle for hash functions. In G. Brassard, editor, *Advances in Cryptology-CRYPTO' 89*, volume 435 of *Lecture Notes in Computer Science*. Springer-Verlag, 1990.
19. R.Diestel, "Graph Theory", Springer-Verlag Heidelberg, New York 1997,2000,2005
20. Y. Dodis, R. Oliveira, K. Pietrzak, On the Generic Insecurity of the Full Domain Hash, *Advances in Cryptology - CRYPTO*, August 2005.
21. Ecrypt Consortium. Ongoing Research Areas in Symmetric Cryptography, January 2005. Available at URL <https://www.cosic.esat.kuleuven.ac.be/nessie/deliverables/D.STVL.3-2.1.pdf>.
22. H.Feistel. Cryptography and Computer Privacy. *Scientific American*.
23. FIPS 46-3: Data Encryption Standard. In National Institute of Standards and Technology, Oct. 1999.
24. D.Feng, W.Wu :Block Cipher Analysis and Design.
25. P.Gauravaram, W.Millan, J. Gonzalez Neito and E. Dawson: 3C-A Provably Secure Pseudorandom Function and Message Authentication Code. A New mode of operation for Cryptographic Hash Function. The preliminary draft version of this work is available at [eprint-2005/390](http://eprint-2005/390) .
26. D. Hong, B. Preneel, and S. Lee, Higher Order Universal One-Way Hash Functions, *ASIACRYPT 2004*, LNCS 3329, pp. 201C213, 2004.

27. A.Joux, Multicollisions in iterated Hash functions. Application to cascaded constructions. Proceedings Crypto 2004, Springer-Verlag LNCS 3152, pp.306-316, 2004.
28. P.Junod and S.Vaudenay, FOX : a New Family of Block Ciphers, Selected Areas in Cryptography-SAC 2004,LNCS 2595, pp.131-146
29. D.Lei, L.Chao, F. Keqin. New Observation On Camellia. Selected Area in Cryptography, SAC 2005, LNCS 3897, pp51C64, 2006.
30. M.Luby and C. Rackoff, How to construct pseudorandom permutations from pseudorandom functions. SIAM Journal on Computing, Vol. 17, No. 2 (1988) pp. 373C386.
31. R.C.Merkle, One Way Hash Functions and DES, In G. Brassard, editor, Advances in Cryptology-CRYPTO' 89, volume 435 of Lecture Notes in Computer Science. Springer-Verlag, pp.428-446, 1990.
32. J.B.Nielsen. Separating Random Oracle Proofs from Complexity Theoretic Proofs: The Non-Committing Encryption Case. In Advances in Cryptology - Crypto 2002, PP.111 -126
33. J.Patarin, Feistel Schemes with Six (or More) Rounds, Fast Software Encryption 1998, pp.103-121.
34. J.Patarin. Luby-Rackoff 7 Rounds are Enough for  $2n^{(1-\varepsilon)}$  Security. CRYPTO'03, Springer, LNCS 2729, pp.513-529.
35. J.Patarin, Security of Random Feistel Schemes with 5 or more rounds. CRYPTO '04, LNCS 3152, pp.106-122, Springer.
36. J.Patarin, Generic Attacks on Feistel Schemes, Available from the author.
37. J.Patarin, Security of Random Feistel Schemes with 5 or more rounds, Available from the author.
38. G.Piret, Luby-Rackoff Revisited: On the Use of Permutations as Inner Functions of a Feistel Scheme, Designs, Codes and Cryptography, 39, pp.233C245, 2006
39. G.Piret, Block Ciphers: Security Proofs, Cryptanalysis, Design, and Fault Attacks, PHD, 2005.
40. S.Lucks: A Failure-Friendly Design Principle for Hash Functions, ASIACRYPT 2005, LNCS 3788, pp. 474C494, 2005.
41. X.Lai and J. L. Massey: Hash functions based on block ciphers. In Advances in Cryptology Eurocrypt'92, Lecture Notes in Computer Science, Vol. 658. Springer-Verlag, Berlin Heidelberg New York (1993) 55-70. 228(5): 15-23.
42. A.Joux. Multi-collisions in iterated hash functions, application to cascaded constructions. Crypto 04, LNCS 3152, 306C316.
43. C.H.Meyer and S.M.Matyas. Cryptography: a New Dimension in Data Security. Wiley & Sons, 1982.
44. B.Preneel: The State of Cryptographic Hash Functions. In Lectures on Data Security, Lecture Notes in Computer Science, Vol. 1561. Springer-Verlag, Berlin Heidelberg New York (1999) 158-182.
45. B.Preneel, R.Govaerts, and J.Vandewalle, " Hash functions based on block ciphers," , In Advances in Cryptology -CRYPTO'93, Lecture Notes in Computer Science,pages 368-378. Springer-Verlag, 1994.
46. B. Preneel. Analysis and design of cryptographic hash functions. PhD thesis, Katholieke Universiteit Leuven, 1993.
47. B.Preneel, V.Rijmen, A.Bosselaers: Recent Developments in the Design of Conventional Cryptographic Algorithms. In State of the Art and Evolution of Computer Security and Industrial Cryptography. Lecture Notes in Computer Science, Vol 1528. Springer-Verlag, Berlin Heidelberg New York(1998) 106-131.

48. M.O.Rabin. Digitalized Signatures. In R. A. Demillo, D. P. Dopkin, A. K. Jones, and R. J. Lipton, editors, Foundations of Secure Computation, pages 155-166, New York, 1978. Academic Press.
49. V.Rijmen, Cryptanalysis and design of iterated block ciphers, Katholieke Universiteit Leuven, Belgium, 9 October 1997
50. B.V.Rompay, Analysis and design of cryptographic hash functions, MAC algorithms and block cipher, K. U. Leuven, Juni 2004.
51. P.Rogaway and T. Shrimpton. Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision-Resistance. FSE 2004, LNCS 3017, 371-388.
52. V.Shoup, Sequences of games: a tool for taming complexity in security proofs, <http://eprint.iacr.org/2004/332.pdf>
53. C.E.Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol.27,pp. 379-423,1948.
54. C.E.Shannon, Communication Theory of Secrecy Systems, Bell System Technical Journal, Vol 28:pp.656-715, 1949.
55. S.Vaudenay, On the Lai-Massey scheme. In K. Lam, T. Okamoto, and C. Xing, editors, Advances in Cryptology - ASIACRYPT'99, volume 1716 of Lecture Notes in Computer Science, pp. 8-19. Springer-Verlag, 2000.
56. S.Vaudenay, Decorrelation: A Theory For Block Cipher security. Journal of Cryptology, 16(4):pp.249-286, 2003.
57. X.Wang, H.Yu, How to Break MD5 and Other Hash Functions, EUROCRYPT'2005, Springer-Verlag, LNCS 3494, pp19-35, 2005.
58. X.Wang, X.Lai, D.Feng and H.Yu., Cryptanalysis of the Hash Functions MD4 and RIPEMD, EUROCRYPT 2005, Springer-Verlag,LNCS 3494, pp1-18, 2005.
59. A.F.Webster and S. E. Tavares. On the design of S-boxes. Advances in Cryptology-CRYPTO'85 Lecture Notes in Computer Science 218, pp.523-534.