Decorrelation over Infinite Domains: the Encrypted CBC-MAC Case

Serge Vaudenay

Swiss Federal Institute of Technology (EPFL)
Serge.Vaudenay@epfl.ch

Abstract Decorrelation theory has recently been proposed in order to address the security of block ciphers and other cryptographic primitives over a finite domain. We show here how to extend it to infinite domains, which can be used in the Message Authentication Code (MAC) case.
In 1994, Bellare, Kilian and Rogaway proved that CBC-MAC is secure when the input length is fixed. This has been extended by Petrank and Rackoff in 1997 with a variable length.
In this paper, we prove a result similar to Petrank and Rackoff's one by using decorrelation theory. This leads to a slightly improved result and a more compact proof.
This result is meant to be a general proving technique for security, which can be compared to the approach which was announced by Maurer at CRYPTO'99.

Decorrelation theory has recently been introduced. (See references [17] to [22].) Its first aim was to address provable security in the area of block ciphers in order to prove their security against differential [7] and linear cryptanalysis [10]. As a matter of fact, these techniques have also been used in order to prove Luby-Rackoff -like pseudorandomness results [9] in a way similar to Patarin's "coefficient H method" [14,15]. All previous cases however address random functions over a finite domain, which is not appropriate for MACs.

The CBC-MAC construction is well known in order to make Message Authentication Codes from a block cipher in Cipher Block Chaining mode. Namely, if C is a permutation defined on a block space $\{0,1\}^m$, for a message $x = (m_1,\ldots,m_\ell) \in (\{0,1\}^m)^\ell$ we define

$$\mathrm{MAC}(x) = C(C(\ldots C(m_1) + m_2 \ldots) + m_\ell).$$

The addition is traditionally the XOR operation but can be replaced by any group (or even quasigroup) law. In 1994, Bellare, Kilian and Rogaway proved that if C is a uniformly distributed random permutation, then for any integer ℓ and any distinguisher between MAC and a truly random function which is limited to d queries, the advantage is less than $3d^2\ell^2 2^{-m}$ [6]. This shows that no adaptive attack can forge a new valid $(x, \mathrm{MAC}(x))$ pair with a relevant probability unless the total number of known blocks $d\ell$ is within the order of $2^{\frac{m}{2}}$. This however holds when all messages have the fixed length ℓ. If the attacker is allowed to use messages with different length, it is easy to notice that for any message x and any block a the MAC of x concatenated with $a - \mathrm{MAC}(x)$ is

$$\mathrm{MAC}(x, a - \mathrm{MAC}(x)) = C(a)$$

which does not depend on x and allows to forge a new authenticated message by replacement of x.

In 1997, Petrank and Rackoff addressed the case of DMAC defined by

$$\mathrm{MAC}(x) = C_2(C_1(C_1(\ldots C_1(m_1) + m_2 \ldots) + m_\ell))$$

(see [16]). This type of construction does not mean any originality since it is already suggested by several standards [2,3,4]. Its security was however formally proved in [16] for the first time.

If we replace C_2 by $C_2 \circ C_1^{-1}$ we can obviously remove the last C_1 application. We can thus consider the MAC defined by

$$\mathrm{MAC}(x) = C_2(C_1(\ldots C_1(m_1) + m_2 \ldots) + m_\ell)$$

which we call the "encrypted CBC-MAC" in the sequel. In this paper we give a security proof which is different from [16] and with a slightly improved reduction. Our proof also happens to be more compact (it is less than 2-page long), thanks to use of the decorrelation theory tools. Our approach is also more general and can be applied to other schemes. In this way it can be compared to the information theoretic general approach which was announced by Maurer at CRYPTO'99 [12].

1 Prerequisite

1.1 Definitions and Notations

First of all, for any random function F from a set $\mathcal{M}_1$ to a set $\mathcal{M}_2$ and any integer d we associate the "d-wise distribution matrix" which is denoted $[F]^d$, defined in the matrix set $\mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}_2^d}$ by

$$[F]^d_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)} = \Pr[F(x_1) = y_1, \ldots, F(x_d) = y_d].$$

Given a metric structure D in $\mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}_2^d}$ we can define the distance between the matrices associated to two random functions F and G. This is the "d-wise decorrelation distance". If G is a random function uniformly distributed in the set of all functions from $\mathcal{M}_1$ to $\mathcal{M}_2$ (we let F^* denote such a function), this distance is called the "d-wise decorrelation bias of function F" and denoted $\mathrm{DecF}^d_D(F)$. When F is a permutation (which will usually be denoted C as for "Cipher") and G is a uniformly distributed permutation (denoted C^*) it is called the "d-wise decorrelation bias of permutation F" and denoted $\mathrm{DecP}^d_D(F)$. In previous results we used the metric structures defined by the norms denoted $||.||_2$ (see

[18]), $|||.|||_\infty$, $||.||_a$, $||.||_s$ (see [21]). These four norms are matrix norms, which means that they are norms on $\mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}_2^d}$ with the property that

$$||A \times B|| \leq ||A||.||B||.$$

This property leads to non-trivial inequalities which can shorten many treatments on the security of conventional cryptography.

Given two random functions F and G from $\mathcal{M}_1$ to $\mathcal{M}_2$ we call "distinguisher between F and G" any oracle Turing machine $\mathcal{A}^O$ which can send $\mathcal{M}_1$-element queries to the oracle O and receive $\mathcal{M}_2$-element responses, and which finally outputs 0 or 1. In particular the Turing machine can be probabilistic. In the following, the number of queries to the oracle will be limited to d. The distributions on F and G induces a distribution on $\mathcal{A}^F$ and $\mathcal{A}^G$, thus we can compute the probability that these probabilistic Turing machines output 1. The advantage for distinguishing F from G is

$$\mathrm{Adv}_{\mathcal{A}}(F,G) = \Pr\left[\mathcal{A}^F \to 1\right] - \Pr\left[\mathcal{A}^G \to 1\right].$$

For any class of distinguishers Cl we will denote

$$\mathrm{Adv}_{\mathrm{Cl}}(F,G) = \max_{\mathcal{A}\in\mathrm{Cl}} \mathrm{Adv}_{\mathcal{A}}(F,G).$$

We notice that if $\mathcal{A}$ is a distinguisher, we can always define a complementary distinguisher $\bar{\mathcal{A}} = 1 - \mathcal{A}$ which gives the opposite output. There is no need for investigating the minimum advantage when the class is closed under the complement (which is the case of the above class) since

$$\mathrm{Adv}_{\bar{\mathcal{A}}}(F,G) = -\mathrm{Adv}_{\mathcal{A}}(F,G).$$

We consider the class Cl_a^d of all (adaptive) distinguishers limited to d queries.

1.2 Properties

The d-wise distribution matrices have the property that if F and G are independent random functions, F from $\mathcal{M}_2$ to $\mathcal{M}_3$ and G from $\mathcal{M}_1$ to $\mathcal{M}_2$, then

$$[F \circ G]^d = [G]^d \times [F]^d.$$

Thus, if we are using a matrix norm $||.||$, we obtain

$$\mathrm{DecF}^d_{||.||}(F \circ G) \leq \mathrm{DecF}^d_{||.||}(F).\mathrm{DecF}^d_{||.||}(G).$$

and the same for permutations.

The $||.||_a$ norm defined in [21] has the quite interesting property that it characterizes the best advantage of a distinguisher in Cl_a^d.

Lemma 1 ([21]). *For any random functions F and G we have*

$$||[F]^d - [G]^d||_a = 2.\mathrm{Adv}_{\mathrm{Cl}_a^d}(F,G).$$

In this paper, we will use the $||.||_a$ norm only and omit it in the notations.

Finally we recall the following lemma.

Lemma 2 ([21]). *Let d be an integer, $F_1,\ldots,F_r$ be r random function oracles, and $C_1,\ldots,C_s$ be s random permutation oracles. We let Ω be a deterministic oracle Turing machine which can access to the previous oracles and an input tape x. It defines a random function $G(x)=\Omega(F_1,\ldots,F_r,C_1,\ldots,C_s)(x)$. We assume that Ω is such that the number of queries to F_i is limited to some integer a_i, and the number of queries to C_j is limited to b_j in total for any $i=1,\ldots,r$ and any $j=1,\ldots,s$. We let the F_i^* (resp. C_j^*) be independent uniformly distributed random functions (resp. permutations) on the same range than F_i (resp. C_j) and we let $G^*=\Omega(F_1^*,\ldots,F_r^*,C_1^*,\ldots,C_s^*)$. We have*

$$\mathrm{DecF}^d(G)\leq\sum_{i=1}^{r}\mathrm{DecF}^{a_i d}(F_i)+\sum_{j=1}^{s}\mathrm{DecP}^{b_j d}(C_j)+\mathrm{DecF}^d(G^*).$$

This lemma actually separates the problem of studying the decorrelation bias of a construction scheme into the problem of studying the decorrelation biases of its internal functions F_i and C_j and studying the decorrelation bias of an idealized version G^*.

1.3 The Coefficient H Method

Patarin introduced the "coefficient H method" which enables to make pseudo-randomness proofs more systematic. In the decorrelation theory setting, this method can be formalized by the following lemma.

Lemma 3 ([22]). *Let d be an integer. Let F be a random function from a set $\mathcal{M}_1$ to a set $\mathcal{M}_2$. We let $\mathcal{X}$ be the subset of $\mathcal{M}_1^d$ of all $(x_1,\ldots,x_d)$ with pairwise different entries. We let F^* be a uniformly distributed random function from $\mathcal{M}_1$ to $\mathcal{M}_2$. We assume there exist a subset $\mathcal{Y}\subseteq\mathcal{M}_2^d$ and two positive numbers ϵ_1 and ϵ_2 such that*

- $|\mathcal{Y}|(\#\mathcal{M}_2)^{-d}\geq 1-\epsilon_1$
- $\forall x\in\mathcal{X}\quad\forall y\in\mathcal{Y}\quad[F]^d_{x,y}\geq(1-\epsilon_2)(\#\mathcal{M}_2)^{-d}$.

Then we have $\mathrm{DecF}^d(F)\leq 2\epsilon_1+2\epsilon_2$.

This lemma intuitively means that if $[F]^d_{x,y}$ is close to $[F^*]^d_{x,y}$ for all x and almost all y, then the decorrelation bias of F is small. It is quite straightforward with techniques inspired by Patarin [14,15] and Maurer [11].

As an illustration, Lemma 3 can be used in order to prove the famous Luby-Rackoff Theorem easily as shown in Appendix.

Theorem 4 (Luby-Rackoff 1986 [9]). *Let F_1^*,F_2^*,F_3^* be three independent random functions on $\{0,1\}^{\frac{m}{2}}$ with uniform distribution. We have*

$$\mathrm{DecF}^d(\Psi(F_1^*,F_2^*,F_3^*))\leq 2d^2.2^{-\frac{m}{2}}$$
$$\mathrm{DecP}^d(\Psi(F_1^*,F_2^*,F_3^*))\leq 2d^2.2^{-\frac{m}{2}}.$$

The results hold for Feistel schemes defined from any (quasi)group operation.[1]

2 Decorrelation Biases of Functions over an Infinite Domain

In order to define decorrelation biases of MACs, we need to address the problem of having infinite sets. Let for instance F be a random function defined from $\mathcal{M}_1^*$ to $\mathcal{M}_2$ ($\mathcal{M}_1^*$ is the set of all finite sequences with entries in $\mathcal{M}_1$). We define the $[F]^{q_1,\ldots,q_d}$ matrix with rows defined on $\mathcal{M}_1^{q_1} \times \ldots \times \mathcal{M}_1^{q_d}$ and columns defined on $\mathcal{M}_2^d$. Next we define $\mathrm{DecF}^{q_1,\ldots,q_d}(F)$ as the distance between $[F]^{q_1,\ldots,q_d}$ and $[F^*]^{q_1,\ldots,q_d}$, where F^* has a uniform distribution. Additionally, we can define

$$\mathrm{DecF}^{d,q}(F) = \max_{q_1+\ldots+q_d=q} \mathrm{DecF}^{q_1,\ldots,q_d}(F).$$

We can easily check that all previous results remain valid for these definitions, namely:

- The best advantage of a distinguisher limited to d (adaptively) chosen queries with a total length of q blocks between F and F^* is $\frac{1}{2}\mathrm{DecF}^{d,q}(F)$.
- As in Lemma 2, if $G = \Omega(F_1, \ldots, F_r, F'_1, \ldots, F'_s)$ uses functions F_i and F'_j on fixed input length, but with occurrence numbers of $a_i\ell$ and b_j respectively where ℓ is the length of the input of G, we have

$$\mathrm{DecF}^{d,q}(G) \leq \sum_{i=1}^{r} \mathrm{DecF}^{a_i q}(F_i) + \sum_{j=1}^{s} \mathrm{DecF}^{b_j d}(F'_j) + \mathrm{DecF}^{d,q}(G^*).$$

  We can use permutations C_i and C'_j as well and have DecP instead of DecF, or even mixtures of functions and permutations.
- Lemma 3 still holds with $\mathrm{DecF}^{d,q}$ instead of DecF^d and $\mathcal{X}$ equal to the set of $(x_1, \ldots, x_d)$ with total length q.

3 Security of MAC

Message Authentication Codes (MAC) are functions which map any binary string onto a fixed length value[2] with a secret key. In this paper, we consider functions defined on the set $(\{0,1\}^m)^*$ of finite sequences of m-bit integers.[3] For

[1] Here $\Psi(F_1^*, F_2^*, F_3^*)$ is the standard notation for a Feistel cipher with three rounds and round functions F_1^*, F_2^*, F_3^*.

[2] More precisely, the MAC is the output of the function, but we will improperly call the function a MAC.

[3] Note that arbitrary bit strings do not always have an integral number of blocks. For this we must use a padding scheme like the Merkle-Damgård [8,13] one in order to transform an arbitrary string into a string with an integral number of blocks. In this paper we prove the security for padded messages which induces the security for the whole scheme with the padding scheme.

instance, given a block cipher Enc_K which is a permutation on $\{0,1\}^m$ defined from a secret key K, we consider the CBC-MAC construction defined by

$$\mathrm{MAC}_K(m_1,\ldots,m_\ell) = \mathrm{Enc}_K(\mathrm{Enc}_K(\ldots\mathrm{Enc}_K(m_1)+m_2\ldots)+m_\ell).$$

Since the secret key K is unknown by the opponent and chosen at random by the legitimate user, we can consider equivalently $C = \mathrm{Enc}_K$ as a random permutation with a given publicly known distribution, and the MAC itself as a random function.

The purpose of MACs is to authenticate messages. Namely, the legitimate authenticator provides $\mathrm{MAC}(x)$ is order to authenticate a message x. Saying that a MAC is (d,q,p)-secure means that for any opponent who can use the legitimate authenticator as an oracle for at most $d-1$ chosen messages $x_1,\ldots,x_{d-1}$ and issue an (x_d,c) pair such that $x_d \neq x_i$ for any i and that the total length of $x_1,\ldots,x_d$ is of q m-bit blocks, the probability that $c = \mathrm{MAC}(x_d)$ is less than p. This is the security against adaptive existential forgery attacks.

We notice that if MAC is such that $\mathrm{DecF}^{d,q}(\mathrm{MAC}) = \epsilon$, then it is a $(d,q,2^{-m}+\frac{\epsilon}{2})$-secure MAC. Namely, for any opponent we can make a distinguisher who just query the forged x_d and check whether the output is c or not. Since the advantage must be less than $\frac{\epsilon}{2}$, the probability of success of the opponent must be less than $\frac{\epsilon}{2}$ plus the probability of success against a truly random function, which is 2^{-m}. Hence we use $\mathrm{DecF}^{d,q}(\mathrm{MAC})$ upper bounds as security evidences.

For instance, we can consider the Bellare-Kilian-Rogaway result which works with a fixed input length ℓ.

Theorem 5 (Bellare-Kilian-Rogaway 1994 [6]). *For any fixed integer ℓ, we consider the function* MAC *defined on ℓ m-bit blocks from a uniformly distributed random function F^* as follows.*

$$\mathrm{MAC}(m_1,\ldots,m_\ell) = F^*(F^*(\ldots F^*(m_1)+m_2\ldots)+m_\ell).$$

For any d we have $\mathrm{DecF}^d(\mathrm{MAC}) \leq 6d^2\ell^2 2^{-m}$. This holds for any (quasi)group addition.

Here is another result which is quite similar to the An-Bellare result [5].

Theorem 6 ([22]). *Let F_1 and F_2 be two independent random functions from $\{0,1\}^{b+m}$ to $\{0,1\}^b$. For any ℓ and any $(m_1,\ldots,m_\ell) \in (\{0,1\}^m)^\ell$ we define*

$$\mathrm{MAC}(m_1,\ldots,m_\ell) = F_2(F_1(\ldots F_1(F_1(0,m_1),m_2)\ldots,m_\ell),\ell)$$

where 0 means a b-bit zero string, and ℓ means an m-bit string which represents the ℓ value. Considering distinguishers limited to d queries and a total length of qm bits we have

$$\mathrm{DecF}^{d,q} \leq \mathrm{DecF}^q(F_1) + \mathrm{DecF}^d(F_2) + q(q-1)2^{-m}.$$

Finally, here is the Petrank-Rackoff [16] result.

Theorem 7 (Petrank-Rackoff [16]). *Let C_1 and C_2 be two independent random permutations on $\{0,1\}^m$ with the same distribution C. For any ℓ and any $(m_1,\ldots,m_\ell) \in (\{0,1\}^m)^\ell$ we define*

$$\mathrm{MAC}(m_1,\ldots,m_\ell) = C_2(C_1(C_1(\ldots C_1(C_1(m_1)+m_2)\ldots+m_{\ell-1})+m_\ell)).$$

Considering adaptive distinguishers limited to d queries and a total length of qm bits we have

$$\mathrm{DecF}^{d,q}(\mathrm{MAC}) \leq 2\mathrm{DecP}^q(C) + 4q^2 2^{-m}.$$

The result holds for any (quasi)group addition.

4 Encrypted CBC-MAC

Here is our main result.

Theorem 8. *Let C_1 and C_2 be two independent random permutations over $\{0,1\}^m$. For any ℓ and any $(m_1,\ldots,m_\ell) \in (\{0,1\}^m)^\ell$ we define*

$$\mathrm{MAC}(m_1,\ldots,m_\ell) = C_2(C_1(\ldots C_1(C_1(m_1)+m_2)\ldots+m_{\ell-1})+m_\ell).$$

Considering adaptive distinguishers limited to d queries and a total length of qm bits we have

$$\begin{aligned}\mathrm{DecF}^{d,q}(\mathrm{MAC}) \leq\ & \mathrm{DecP}^{q-d}(C_1) + \mathrm{DecP}^d(C_2) \\ & + d(d-1)2^{-m} + q(q+1)(1+q2^{-m})2^{-m}.\end{aligned}$$

The result holds for any (quasi)group addition.

This result is slightly better than Theorem 7.

Proof. Lemma 2 reduces to the case where C_1 and C_2 are independent uniformly distributed random permutations.

Using Lemma 3, let $\mathcal{Y}$ be the set of all $y = (y_1,\ldots,y_d)$ with different y_is. We thus have

$$\epsilon_1 = 1 - \frac{2^{md}}{2^m(2^m-1)\ldots(2^m-d+1)} \leq \frac{d(d-1)}{2}2^{-m}.$$

Now for any collection of $x_i = (m_{i,1},\ldots,m_{i,q_i})$ we let

$$U_{i,j} = C_1(\ldots C_1(C_1(m_{i,1})+m_{i,2})\ldots+m_{i,j-1})+m_j.$$

We consider the event E that all U_{i,q_i} are pairwise different. We have

$$\begin{aligned}[\mathrm{MAC}]^{q_1,\ldots,q_d}_{x,y} &\geq \Pr[\mathrm{MAC}(x_i)=y_i; i=1,\ldots,d \text{ and } E] \\ &= \Pr[\mathrm{MAC}(x_i)=y_i; i=1,\ldots,d/E]\Pr[E] \\ &= \frac{1}{2^m(2^m-1)\ldots(2^m-d+1)}\Pr[E] \\ &\geq 2^{-md}(1-\Pr[\bar{E}])\end{aligned}$$

therefore we can take $\epsilon_2 = \Pr[\bar{E}] = \Pr[\exists i < r; U_{i,q_i} = U_{r,q_r}]$.

The remaining part of the proof consists of upper bounding ϵ_2 by $\frac{q(q+1)}{2}(1+q2^{-m})2^{-m}$ and applying Lemma 3.

We call a collision an event $U_{i,j} = U_{r,s}$. This collision is trivial if we have $(m_{i,1}, \ldots, m_{i,j}) = (m_{r,1}, \ldots, m_{r,s})$ and non-trivial otherwise. Let Inv be the event that $C_1(U_{i,j}) = 0$ for some i, j, and let Coll be the event that we have a non-trivial collision. We can easily show that the $\bar{E}$ event is included in Inv$\cup$Coll: if $U_{i,q_i} = U_{r,q_r}$, then either $m_{i,q_i} \neq m_{r,q_r}$ and it is a non-trivial collision, or it reduces to $U_{i,q_i-1} = U_{r,q_r-1}$ and we can iterate... Thus $\epsilon_2 \leq \Pr[\text{Inv}] + \Pr[\text{Coll}]$.

The probability that any adaptive attack against C_1 finds a preimage of 0 after $q - d$ queries is obviously less than $\frac{q}{2^m - q}$. Thus $\Pr[\text{Inv}] \leq \frac{q}{2^m - q}$.

We let $\mathcal{U}$ be the set of all $U_{i,j}$-indices, which means the set of all (i, j) such that $1 \leq i \leq d$ and $1 \leq j \leq q_i$. For $A \subseteq \mathcal{U}$ we let $c(A)$ be

$$c(A) = \{(i, j); \exists (r, s) \in A \quad i = r \text{ and } j \leq s\}.$$

Thus $c(A)$ is the set the indices of all $U_{i,j}$ which are required in order to compute all $U_{r,s}$ values for $(r, s) \in A$. We define an ordering on $2^{\mathcal{U}}$ by

$$A \leq B \iff c(A) \subseteq c(B).$$

We let $\mathcal{I}$ be the set of all indices pairs of potential non-trivial collisions $U_{i,j} = U_{r,s}$, namely the set of all pairs $\{(i, j), (r, s)\}$ of $\mathcal{U}$-elements such that $(m_{i,1}, \ldots, m_{i,j}) \neq (m_{r,1}, \ldots, m_{r,s})$. For any i, j, r, s such that $\{(i, j), (r, s)\} \in \mathcal{I}$ we let $\text{Coll}_{i,j,r,s}$ be the event of the collision $U_{i,j} = U_{r,s}$ (which is necessarily non-trivial since $\{(i, j), (r, s)\} \in \mathcal{I}$), and we let $\text{MinColl}_{i,j,r,s}$ be the complementary in $\text{Coll}_{i,j,r,s}$ of the union of all $\text{Coll}_{i',j',r',s'}$ for $\{(i', j'), (r', s')\} \in \mathcal{I}$ and $\{(i', j'), (r', s')\} < \{(i, j), (r, s)\}$, i.e. the event $U_{i,j} = U_{r,s}$ with no prior non-trivial collision. We easily notice that

$$\text{Coll} = \bigcup_{\{(i,j),(r,s)\} \in \mathcal{I}} \text{MinColl}_{i,j,r,s}.$$

We have at most $\frac{q(q-1)}{2}$ terms in $\mathcal{I}$. Hence

$$\Pr[\text{Coll}] \leq \frac{q(q-1)}{2} \max_{\{(i,j),(r,s)\} \in \mathcal{I}} \Pr[\text{MinColl}_{i,j,r,s}].$$

For $\{(i, j), (r, s)\} \in \mathcal{I}$, let us consider the $\text{MinColl}_{i,j,r,s}$ event. We assume without loss of generality that $s \leq j$. Since we have no prior collision we must have $m_{i,j} \neq m_{r,s}$. Furthermore we must have $U_{i,j-1} \neq U_{r,s-1}$ because C_1 is a permutation (otherwise $C_1(U_{i,j-1}) + m_{i,j}$ cannot be equal to $C_1(U_{r,s-1}) + m_{r,s}$) and $j > 1$, and we need to consider the event

$$C_1(U_{i,j-1}) + m_{i,j} = U_{r,s}.$$

If we have a collision $U_{i,j-1} = U_{i',j'}$ with $(i, j - 1) \neq (i', j')$ and $(i', j') \in c(i, j, r, s)$, it must be trivial (otherwise the initial collision is not minimal) which means $j' = j-1$ and $i' = r \neq i$ and $(m_{i,1}, \ldots, m_{i,j-1}) = (m_{r,1}, \ldots, m_{r,j-1})$. If $s < j$ we have $U_{i,j} = U_{r,s}$ and $U_{r,s} = U_{i,s}$ thus $U_{i,j} = U_{i,s}$ which is non-trivial, which contradicts the minimality of the initial collision. Thus we must have $s = j$, but the trivial collision $U_{i,j-1} = U_{r,j-1}$ then contradicts $U_{i,j-1} \neq U_{r,s-1}$. Therefore $U_{i,j-1}$ is equal to no $U_{i',j'}$ for $(i',j') \in c(i,j,r,s) \backslash \{(i,j-1)\}$. This implies that the marginal distribution of $C_1(U_{i,j-1})$ with the knowledge of all previous $U_{i',j'}$ is uniform among a set of at least $2^m - q + 1$ elements. Hence $\Pr[\mathrm{MinColl}_{i,j,r,s}] \leq \frac{1}{2^m - q}$.

Finally we obtain

$$\epsilon_2 \leq \frac{q}{2^m - q} + \frac{q(q-1)}{2} \times \frac{1}{2^m - q} \leq \frac{q(q+1)}{2}(1 + q2^{-m})2^{-m}.$$

Applying Lemma 3 now completes the proof. □

5 Extensions

In our result we notice that since $d \leq q$, the bound is small until q reaches the order of $2^{\frac{m}{2}}$. This result is tight since usual collision attacks can break our construction within this complexity. Actually, we can query $2^{\frac{m}{2}}$ two-block messages until we get a collision $\mathrm{MAC}(m_1, m_2) = \mathrm{MAC}(m_1', m_2')$ then query $c = \mathrm{MAC}(m_1, m_2, m_3)$ and output a forged authenticated message $((m_1', m_2', m_3), c)$. We have $d = 2^{\frac{m}{2}} + 2$ and $q = 2.2^{\frac{m}{2}} + 6$ and $p \approx 1 - e^{-1}$.

We may think that since we have an m-bit MAC and a security of $2^{\frac{m}{2}}$ uses we have an efficiency loss in term of storage. We can improve this construction by shrinking the MAC on $\frac{m}{2}$ bits as suggested in most of standards. More precisely, let F be a random function from $\{0,1\}^m$ to $\{0,1\}^b$. We can define

$$\mathrm{MAC}(m_1, \ldots, m_\ell) = F(C(\ldots C(C(m_1) + m_2) \ldots + m_{\ell-1}) + m_\ell)$$

and we have

$$\mathrm{DecF}^{d,q}(\mathrm{MAC}) \leq \mathrm{DecP}^q(C) + \mathrm{DecF}^d(F) + q(q+1)(1 + q2^{-m})2^{-m}.$$

(In the proof of Theorem 8, we take $\mathcal{Y}$ equal to the full set so that $\epsilon_1 = 0$.)

If we now want to shorten the two keys, we can replace the independent C and F random functions by dependent ones. Let $||[CF]^q - [C_0F_0]^q||_a$ denote the decorrelation distance between the (C,F) pair and a pair (C_0, F_0) of independent random functions such that C_0 (resp. F_0) has the same distribution than C (resp. F). This is half of the best advantage for distinguishing them from q queries. We should still consider $\mathrm{DecP}^{q-d}(C)$ and $\mathrm{DecF}^d(F)$. So, even if C and F are dependent, we still have the following result.

Theorem 9. *Let C and C_0 be two identically distributed random permutations on $\{0,1\}^m$ and let F and F_0 be two identically distributed random functions from $\{0,1\}^m$ to $\{0,1\}^b$. We assume that C_0 and F_0 are independent. For any ℓ and any $(m_1, \ldots, m_\ell) \in (\{0,1\}^m)^\ell$ we define*

$$\mathrm{MAC}(m_1, \ldots, m_\ell) = F(C(\ldots C(C(m_1) + m_2) \ldots + m_{\ell-1}) + m_\ell).$$

Considering adaptive distinguishers limited to d queries and a total length of qm bits we have

$$\mathrm{DecF}^{d,q}(\mathrm{MAC}) \leq |||[CF]^q - [C_0F_0]^q|||_a + \mathrm{DecP}^{q-d}(C) + \mathrm{DecF}^d(F) + q(q+1)(1+q2^{-m})2^{-m}.$$

The result holds for any (quasi)group addition.

This theorem clearly separates the security issues induced by the probabilistic dependence between C and F, the C algorithm, the F algorithm, and the MAC scheme.

As an example we can use

$$C(x) = \mathrm{DES}_K(x) \quad \text{and} \quad F(x) = \mathrm{Trunc}(\mathrm{DES}_{K+c}(x))$$

for a given constant c, and where Trunc truncates a 64-bit string onto its first half and DES is the Data Encryption Standard [1]. We get a MAC on $b = 32$ bits with a single 56-bit key and block of $m = 64$ bits. We obtain

$$\mathrm{DecF}^{d,q}(\mathrm{MAC}) \leq f(q) + q(q+1)(1+q2^{-64})2^{-64}$$

where $f(q)$ is the sum of the best advantages for distinguishing

- $(\mathrm{DES}_K, \mathrm{Trunc} \circ \mathrm{DES}_{K+c})$ from $(\mathrm{DES}_{K_1}, \mathrm{Trunc} \circ \mathrm{DES}_{K_2})$
- DES from C^*
- $\mathrm{Trunc} \circ \mathrm{DES}$ from F^*

within a total number of query blocks less than q. Let $q = \theta 2^{\frac{m}{2}}$ (which is a limit of 32θGB of queries). The advantage of any distinguisher is less than $\frac{f(q)+\theta^2}{2}$ thus the probability of success of any adaptive existential forgery attack is less than $2^{-32} + \frac{f(q)+\theta^2}{2}$. Let us conjecture that $f\left(\frac{2^{32}}{10}\right) \leq 2^{-7}$. If we authenticate less than 3GB, the probability of success of the best attack is less than 1%.

The Advanced Encryption Standard will soon provide better security with $m = 128$.

It shall however be outlined that this example is a little misleading since we do not assume any computational bound on the distinguisher which can thus perform an exhaustive search. This means that the conjecture is wrong. We can still modify the result and the computational model by limiting the time complexity to t. All reductions in this paper introduce simulators (like for instance a simulator for the MAC given an oracle for DES) which induce a small time complexity overhead which is often denoted $O(1)$. As a result we obtain that for the maximal time complexity t such that $f\left(\frac{2^{32}}{10}\right) \leq 2^{-7}$, the probability of success of any attack which is limited to a complexity of $t - O(1)$ is less than 1% after having authenticated 3GB.

6 Conclusion

We have shown that the regular CBC-MAC construction provides a secure MAC when the output is encrypted. The security analysis suggests that if m is the block length of the underlying block cipher, then we should not use the MAC construction on more than $2^{\frac{m}{2}}$ blocks in total.

In order to fit to the security, we can even reduce the MAC length down to $\frac{m}{2}$ bits, and shorten the key with extra security hypothesis. This enables to prove the security of existing standards.

These results are quite similar than the Petrank-Rackoff ones. Our technique based on decorrelation theory is however quite systematic and can be applied to most of current MAC constructions with compact proofs.

Finally, we believe that these techniques will contribute to making systematic proof analysis of cryptographic schemes and ultimately lead to some automatic security validation procedures.

References

1. Data Encryption Standard. *Federal Information Processing Standard Publication 46*, U. S. National Bureau of Standards, 1977.
2. ANSI X9.9. American National Standard - Financial Institution Message Authentication (Wholesale). ASC X9 Secretariat - American Bankers Association, 1986.
3. ISO 8731-2. Banking - Approved Algorithms for Message Authentication - Part 2: Message Authenticator Algorithm. International Organization for Standardization, Geneva, Switzerland, 1992.
4. *RACE Project*, Lectures Notes in Computer Science 1005, Springer-Verlag, 1995. .
5. J. H. An, M. Bellare. Constructing VIL-MACs from FIL-MACs: Message Authentication under Weakened Assumptions. In *Advances in Cryptology CRYPTO'99*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 1666, pp. 252–269, Springer-Verlag, 1999.
6. M. Bellare, J. Kilian, P. Rogaway. The Security of Cipher Block Chaining. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 839, pp. 341–358, Springer-Verlag, 1994.
7. E. Biham, A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
8. I. B. Damgård. A Design Principle for Hash Functions. In *Advances in Cryptology CRYPTO'89*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 435, pp. 416–427, Springer-Verlag, 1990.
9. M. Luby, C. Rackoff. How to Construct Pseudorandom Permutations from Pseudorandom Functions. *SIAM Journal on Computing*, vol. 17, pp. 373–386, 1988.
10. M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 839, pp. 1–11, Springer-Verlag, 1994.
11. U. M. Maurer. A Simplified and Generalized Treatment of Luby-Rackoff Pseudorandom permutation generators. In *Advances in Cryptology EUROCRYPT'92*, Balatonfüred, Hungary, Lectures Notes in Computer Science 658, pp. 239–255, Springer-Verlag, 1993.

12. U. M. Maurer. Information-Theoretic Cryptography. Invited lecture. In *Advances in Cryptology CRYPTO'99*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 1666, pp. 47–64, Springer-Verlag, 1999.
13. R. C. Merkle. One way Hash Functions and DES. In *Advances in Cryptology CRYPTO'89*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 435, pp. 416–427, Springer-Verlag, 1990.
14. J. Patarin. *Etude des Générateurs de Permutations Basés sur le Schéma du D.E.S.*, Thèse de Doctorat de l'Université de Paris 6, 1991.
15. J. Patarin. How to Construct Pseudorandom and Super Pseudorandom Permutations from One Single Pseudorandom Function. In *Advances in Cryptology EUROCRYPT'92*, Balatonfüred, Hungary, Lectures Notes in Computer Science 658, pp. 256–266, Springer-Verlag, 1993.
16. E. Petrank, C. Rackoff. CBC MAC for Real-Time Data Sources. *Journal of Cryptology*, vol. 13, pp. 315–338, 2000.
17. S. Vaudenay. Provable Security for Block Ciphers by Decorrelation. Invited talk. In *STACS 98*, Paris, France, Lectures Notes in Computer Science 1373, pp. 249–275, Springer-Verlag, 1998. Full Paper: technical report LIENS-98-8, Ecole Normale Supérieure, 1998. (`ftp://ftp.ens.fr/pub/reports/liens/`)
18. S. Vaudenay. Feistel Ciphers with L_2-Decorrelation. In *Selected Areas in Cryptography*, Kingston, Ontario, Canada, Lectures Notes in Computer Science 1556, pp. 1–14, Springer-Verlag, 1999.
19. S. Vaudenay. Resistance Against General Iterated Attacks. In *Advances in Cryptology EUROCRYPT'99*, Prague, Czech Republic, Lectures Notes in Computer Science 1592, pp. 255–271, Springer-Verlag, 1999.
20. S. Vaudenay. On the Lai-Massey Scheme. In *Advances in Cryptology ASIACRYPT'99*, Singapore, Lectures Notes in Computer Science 1716, pp. 8–19, Springer-Verlag, 2000.
21. S. Vaudenay. Adaptive-Attack Norm for Decorrelation and Super-Pseudorandomness. In *Selected Areas in Cryptography*, Kingston, Ontario, Canada, Lectures Notes in Computer Science 1758, pp. 49–61, Springer-Verlag, 2000.
22. S. Vaudenay. On Provable Security for Conventional Cryptography. Invited talk. (To appear in the proceedings of ICISC' 99, LNCS, Springer-Verlag.)

A Proof of Theorem 4

Following the Feistel scheme $F = \Psi(F_1^*, F_2^*, F_3^*)$, we let

$$\begin{aligned} x_i &= (z_i^0, z_i^1) \\ z_i^2 &= z_i^0 + F_1^*(z_i^1) \\ y_i &= (z_i^4, z_i^3) \end{aligned}$$

We let E be the event $z_i^3 = z_i^1 + F_2^*(z_i^2)$ and $z_i^4 = z_i^2 + F_3^*(z_i^3)$ for $i = 1, \ldots, d$. We thus have $[F]^d_{x,y} = \Pr[E]$. We now define

$$\mathcal{Y} = \left\{ (y_1, \ldots, y_d); \forall i < j \quad z_i^3 \neq z_j^3 \right\}.$$

We can easily check that $\mathcal{Y}$ fulfill the requirements of Lemma 3. Firstly we have

$$|\mathcal{Y}| \geq \left(1 - \frac{d(d-1)}{2} 2^{-\frac{m}{2}} \right) 2^{md}$$

thus we let $\epsilon_1 = \frac{d(d-1)}{2} 2^{-\frac{m}{2}}$. Second, for $y \in \mathcal{Y}$ and any x (with pairwise different entries), we need to consider $[F]^d_{x,y}$. Let E^2 be the event that all z_i^2s are pairwise different over the distribution of F_1^*. We have

$$[F]^d_{x,y} \geq \Pr[E/E^2]\Pr[E^2].$$

For computing $\Pr[E/E^2]$ we know that z_i^3s are pairwise different, as for the z_i^2s. Hence $\Pr[E/E^2] = 2^{-md}$. It is then straightforward that $\Pr[E^2] \geq 1 - \frac{d(d-1)}{2} 2^{-\frac{m}{2}}$ which is $1 - \epsilon_2$. We thus obtain from Lemma 3 that $\mathrm{DecF}^d(F) \leq 2d(d-1)2^{-\frac{m}{2}}$. From Lemma 3 it is straightforward that $\mathrm{DecF}^d(C^*) \leq d(d-1)2^{-m}$. We thus obtain $\mathrm{DecP}^d(F) \leq 2d^2 2^{-\frac{m}{2}}$ for $d \leq 2^{1+\frac{m}{2}}$. Since DecF is always less than 2, it also holds for larger d. $\square$