Small Stretch Problem of the DCT Scheme and How to Fix It

Yuchao Chen[1, 2] Tingting Guo[3] Lei Hu[4, 5] Lina Shang[6]
Shuping Mao[4, 5] Peng Wang[7]

[1]School of Cyber Science and Technology, Shandong University, Qingdao, China

[2]Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Shandong University, Jinan, China

[3]Research Center for Data Hub and Security, Zhejiang lab, Hangzhou, China

[4]Key Laboratory of Cyberspace Security Defense, IIE, CAS, Beijing, China

[5]School of Cyber Security, UCAS, Beijing, China

[6]Space Star Technology Co., Ltd, Beijing, China

[7]School of Cryptology, UCAS, Beijing, China

March 25, 2024

Outline

1. Overview
2. A DAE Scheme: DCT
3. Small Stretch Problem of DCT
4. Attacks on DCT with Small Stretch
5. How to Fix It: Robust DCT

Outline

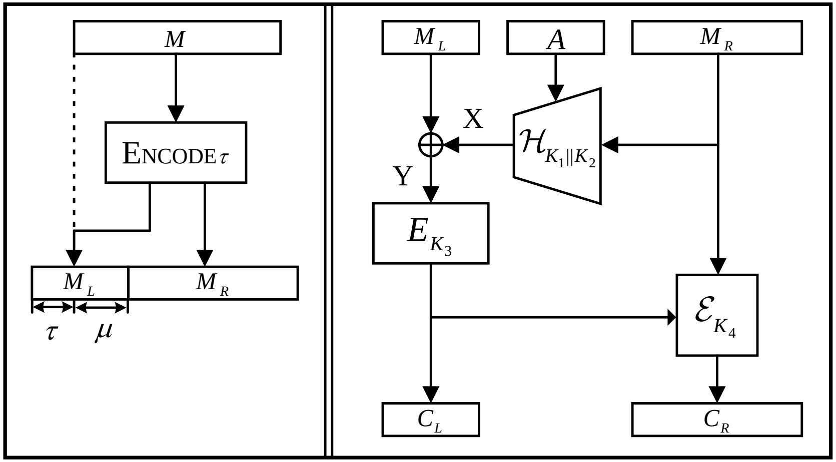
Overview

- We propose a systematic technique to linearize the BRW polynomial employed by the instantiation of DCT.
- We show that although DCT employs the BRW polynomial, it still suffers from a small stretch problem similar to that of GCM.
- We propose a variant of DCT named Robust DCT (RDCT) with minimal modification, and we prove the DAE security of RDCT.

Outline

A DAE Scheme: DCT

- Proposed by Forler et al. [FLLW16].
- Beyond-Birthday-Bound secure.
- Ensuring integrity by adding redundancy (**left**).



The instantiation of DCT employs a CTR-like encryption scheme as $\mathcal{E}_{K_4}$, a $2n$-bit permutation as E, and uses the BRW polynomial to instantiate $\mathcal{H}_{K_1\|K_2}$. ENCODE$_\tau$ encodes the τ-bit of zero into the message.

Outline

Small Stretch Problem of DCT

- Both GCM [MV04] and DCT employ a polynomial-based UHF.
- When the stretch length τ of DCT is small, using the linear modification technique proposed by Ferguson [Fer05], we can choose a special m-block message, and reduce the number of queries required by a successful forgery to $\mathcal{O}(2^\tau/m)$.
- Our attack efficiently balances space and time complexity but does not contradict the security bounds of DCT.

Linear Modification Technique

- The authentication function of GCM can be denoted as:

$$T := R \oplus \sum_{i=1}^{m} C_i H^i.$$

- When GCM uses a small truncated tag, the adversary can change the ciphertext by solving a system of linear equations to obtain potential successful modifications with higher probability.

Example

When GCM uses a 32-bit tag, and the adversary knows the ciphertext for a message consisting of 2^{17} blocks (about 2 MB), with Ferguson's technique, the probability of an adversary forging a 32-bit tag is 2^{-16} instead of 2^{-32}.

Outline

The Universal Hash Function of DCT

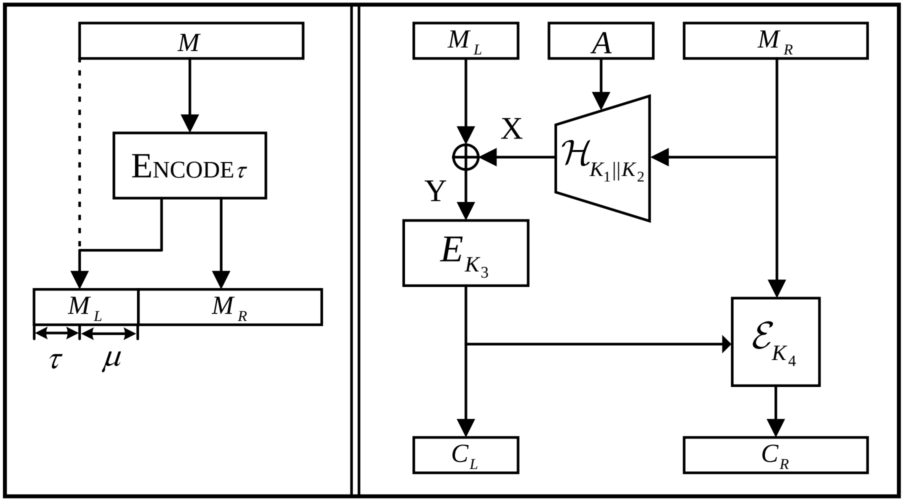
$$\mathcal{H}_{K_1\|K_2}(X_1, X_2) = KBRW_{K_1}(M)\|KBRW_{K_2}(M).$$

Definition 1 (KBRW polynomial)

Given an m-block message $M = (M_1, \cdots, M_m)$, $M_i \in \{0,1\}^n$, the polynomial $KBRW_K(M)$ is defined as follows:

$$\begin{aligned}
&KBRW_K(\varepsilon) = 0^n;\\
&KBRW_K(M_1) = M_1K;\\
&KBRW_K(M_1, M_2) = M_1K^2 \oplus M_2K;\\
&KBRW_K(M_1, M_2, M_3) = K^4 \oplus M_1K^3 \oplus M_2K^2 \oplus (M_1M_2 \oplus M_3)K;\\
&KBRW_K(M_1, \cdots, M_m) = KBRW_K(M_1, \cdots, M_{t-1})(K^t \oplus M_t) \oplus\\
&\qquad KBRW_K(M_{t+1}, \cdots, M_m) \textit{ if } t \le m < 2t \textit{ for } t = 2^i, i \ge 2.
\end{aligned}$$

Idea of Our Attacks



The forgery is successful if and only if:

$$MSB_\tau(M_L) = MSB_\tau(E^{-1}_{K_3}(C_L) \oplus KBRW_K(M_R \oplus C_R \oplus C'_R)) = 0^\tau.$$

So the forgery attack is reduced to the problem of looking for a modification string $D = C_R \oplus C'_R = M_R \oplus M'_R$ while keeping

$$MSB_\tau(KBRW_K(M)) = MSB_\tau(KBRW_K(M \oplus D)).$$

Linearizing KBRW with Special Length Message

Example: When $m = 3$

$$KBRW_K(M) = K^4 \oplus M_1 K^3 \oplus M_2 K^2 \oplus (M_1 M_2 \oplus M_3)K.$$

Let M_1 remain invariable ($D_1 = 0$), and only modify M_2 and M_3 by unknowns D_2 and D_3, respectively, so that

$$KBRW_K(M) \oplus KBRW_K(M \oplus D) = D_2 K^2 \oplus (M_1 D_2 \oplus D_3)K$$

is a linear function of K, where $D = (D_1, D_2, D_3)$.

Linearizing KBRW with Special Length Message

Example: When $m = 7$

$$
\begin{aligned}
KBRW_K(M) = K^8 \oplus M_1 K^7 \oplus M_2 K^6 \oplus (M_1 M_2 \oplus M_3) K^5 \\
\oplus (M_4 \oplus 1) K^4 \oplus (M_1 M_4 \oplus M_5) K^3 \oplus (M_2 M_4 \oplus M_6) K^2 \\
\oplus (M_1 M_2 M_4 \oplus M_3 M_4 \oplus M_5 M_6 \oplus M_7) K.
\end{aligned}
$$

Let $M_1 = 0$, M_2, M_3 and M_5 remain invariable ($D_2 = D_3 = D_5 = 0$), and only modify M_4, M_6 and M_7 by unknowns D_4, D_6 and D_7, respectively, so that

$$
\begin{aligned}
KBRW_K(M) \oplus KBRW_K(M \oplus D) = D_4 K^4 \oplus (M_2 D_4 \oplus D_6) K^2 \\
\oplus (M_3 D_4 \oplus M_5 D_6 \oplus D_7) K
\end{aligned}
$$

is a linear function of K.

Linearizing KBRW with Special Length Message

Assume the message length is $m = 2^u - 1$.

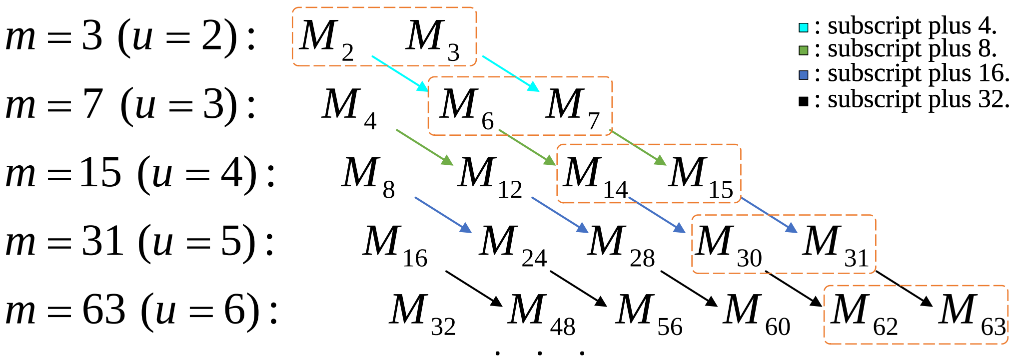
- $\mathcal{V}_0^u$ and $\mathcal{V}_1^u$ are sets of blocks that can be chosen arbitrarily and modified by unknowns;
- $\mathcal{A}_0^u$ and $\mathcal{A}_1^u$ are sets of blocks that can be chosen arbitrarily but not modified by unknowns;
- $\mathcal{F}_0^u$ and $\mathcal{F}_1^u$ are sets of blocks that are fixed as 0 and 1 respectively and not modified by unknowns.

Example: When $u = 2$

$$KBRW_K(M) = K^4 \oplus M_1 K^3 \oplus M_2 K^2 \oplus (M_1 M_2 \oplus M_3) K.$$

Let $\mathcal{V}_0^2 = \{M_2, M_3\}$ and $\mathcal{A}_0^2 = \{M_1\}$ (the value of M_1 should remain invariable in our forgery attacks ($D_1 = 0$)).

Linearizing KBRW with Special Length Message



Example: When $u = 3$, $t = 4$

$$\begin{aligned} KBRW_K(M) &= KBRW_K(M_1, M_2, M_3)(K^4 \oplus M_4) \oplus KBRW_K(M_5, M_6, M_7) \\ &= K^8 \oplus M_1 K^7 \oplus M_2 K^6 \oplus (M_1 M_2 \oplus M_3) K^5 \\ &\oplus (M_4 \oplus 1) K^4 \oplus (M_1 M_4 \oplus M_5) K^3 \oplus (M_2 M_4 \oplus M_6) K^2 \\ &\oplus (M_1 M_2 M_4 \oplus M_3 M_4 \oplus M_5 M_6 \oplus M_7) K. \end{aligned}$$

$\mathcal{V}_0^3 = \{M_{i+2^2} | M_i \in \mathcal{V}_0^2\} = \{M_6, M_7\}$.
Note that the term $(M_4 \oplus 1)K^4$, we choose $\mathcal{V}_1^3 = \{M_4\}$.

Linearizing KBRW with Special Length Message

Theorem 2

For the KBRW polynomial, assume $m = 2^u - 1$, $u \geq 2$. *Let* $\mathcal{V}_0^2 = \{M_2, M_3\}$, $\mathcal{A}_0^2 = \{M_1\}$, *and initialize the remaining set to* $\emptyset$. *We can obtain the following recursions:*

$$
\begin{aligned}
\mathcal{V}_0^u &= \{M_{i+2^{u-1}} | M_i \in \mathcal{V}_0^{u-1}\}, \\
\mathcal{V}_1^u &= \{M_{2^{u-1}}\} \bigcup \{M_{i+2^{u-1}} | M_i \in \mathcal{V}_1^{u-1}\}, \\
\mathcal{A}_0^u &= \mathcal{V}_0^{u-1} \bigcup \{M_{i+2^{u-1}} | M_i \in \mathcal{A}_0^{u-1}\}, \\
\mathcal{A}_1^u &= \mathcal{V}_1^{u-1} \bigcup \{M_{i+2^{u-1}} | M_i \in \mathcal{A}_1^{u-1}\}, \\
\mathcal{F}_0^u &= \mathcal{F}_0^{u-1} \bigcup \{M_{i+2^{u-1}} | M_i \in \mathcal{F}_0^{u-1}\} \bigcup \mathcal{A}_0^{u-1}, \\
\mathcal{F}_1^u &= \mathcal{F}_1^{u-1} \bigcup \{M_{i+2^{u-1}} | M_i \in \mathcal{F}_1^{u-1}\} \bigcup \mathcal{A}_1^{u-1},
\end{aligned}
$$

where $i \in \mathbb{Z}^+$. *Then, after assigning the message blocks according to the recursions above,* $KBRW_K(M) \oplus KBRW_K(M \oplus D)$ *is a linear function of* K.

Linearizing KBRW with General Length Message I

For general m, we define six disjoint sets of message blocks as V_0^m, V_1^m, A_0^m, A_1^m, F_0^m and F_1^m.

Theorem 3

For the KBRW polynomial, assuming the message length is m, $t \leq m < 2t, t = 2^u, u \geq 2$. *Let* $V_0^1 = \{M_1\}$, $V_0^2 = \{M_1, M_2\}$, $V_0^3 = \{M_2, M_3\}$, $A_0^3 = \{M_1\}$ *and initialize the remaining set to* $\emptyset$. *We can obtain the following recursions when* $m \geq 4$:

$$A_0^m = V_0^{t-1} \bigcup \{M_{i+t} | M_i \in A_0^{m-t}\},$$

$$A_1^m = V_1^{t-1} \bigcup \{M_{i+t} | M_i \in A_1^{m-t}\},$$

$$F_0^m = F_0^{t-1} \bigcup \{M_{i+t} | M_i \in F_0^{m-t}\} \bigcup A_0^{t-1},$$

$$F_1^m = F_1^{t-1} \bigcup \{M_{i+t} | M_i \in F_1^{m-t}\} \bigcup A_1^{t-1}.$$

Linearizing KBRW with General Length Message II

Theorem 3

Furthermore, we can obtain the following recursions when $m \geq 7$:

$$V_0^m = \begin{cases} \{M_{i+t} | M_i \in V_0^{m-t}\} \bigcup \{M_t\}, & m < \frac{3t}{2} \\ \{M_{i+t} | M_i \in V_0^{m-t}\}, & otherwise \end{cases}$$

$$V_1^m = \begin{cases} \{M_{i+t} | M_i \in V_1^{m-t}\}, & m < \frac{3t}{2} \\ \{M_{i+t} | M_i \in V_1^{m-t}\} \bigcup \{M_t\}, & otherwise, \end{cases}$$

where $i \in \mathbb{Z}^+$. Then, after assigning the message blocks according to the above recursions, $KBRW_K(M) \oplus KBRW_K(M \oplus D)$ is a linear function of K.

Attacking the Instantiation of DCT

Generic steps:

1. Select a particular message M to query the encryption of DCT and obtain the corresponding ciphertext $C_L \| C_R$.
2. Determine the value of each message block and modification block according to Theorem 2, to make $KBRW_K(M_R) \oplus KBRW_K(M_R \oplus D)$ a linear function of K. Then calculate a set of solutions $\mathcal{D}$ satisfying

$$MSB_u(KBRW_K(M_R) \oplus KBRW_K(M_R \oplus D)) = 0^u,$$

   where $u \leq \tau$.
3. Select a D from $\mathcal{D}$ and query the decryption of DCT with $C_L \| (C_R \oplus D)$. Repeat the step until passing the decryption verification. After about $2^{\tau - u}$ queries, we obtain a successful forgery.

Outline

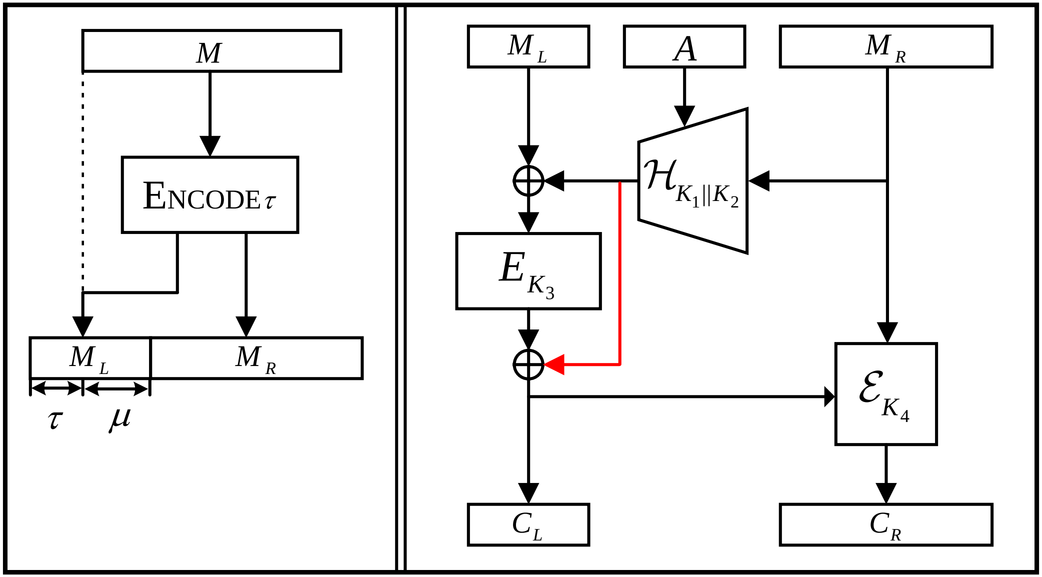


Figure 1: The ENCODE$_\tau$ process (**left**) and the encryption process of RDCT (**right**).

Encryption (resp. decryption) of RDCT **will lead to a random output**.

- The modification forms a tweakable blockcipher $\widetilde{E}$ based on $\mathcal{H}_{K_1\|K_2}$ and E_{K_3}:

$$\begin{aligned}&\widetilde{E}_{K_1,K_2,K_3}((A,M_R),M_L)\\&\quad:= E_{K_3}\left(M_L \oplus \mathcal{H}_{K_1\|K_2}(A,M_R)\right) \oplus \mathcal{H}_{K_1\|K_2}(A,M_R)\,.\end{aligned}$$

- The idea is similar to the paper by Ashur et al. [ADL17], which introduces minor tweaks, such as an additional XOR, to obtain a tweakable blockcipher.
- The core of RDCT is an instantiation of UIV construction [DK22].

Security

Lemma 1 (Confidentiality Advantage of RDCT)

Let $\widetilde{\Pi} = \mathrm{RDCT}_{\mathcal{H},E,\Pi_1,\Pi_2}$. Let $\mathbf{A}$ be a $\textsc{detPriv}$ adversary on $\widetilde{\Pi}$ that submits at most q_e encryption queries of at most m blocks in total and runs in time at most t. Then

$$\mathbf{Adv}_{\widetilde{\Pi}}^{\textsc{detPriv}}(\mathbf{A}) \leq 3q_e^2\epsilon + \frac{q_e(q_e-1)}{2^{2n+1}} + \mathbf{Adv}_E^{\mathrm{PRP}}(q_e, \mathcal{O}(t+q_e)) + \mathbf{Adv}_{\Pi_1}^{\mathrm{IVE}}(q_e, m, \mathcal{O}(t)).$$

Lemma 2 (Integrity Advantage of RDCT)

Let $\widetilde{\Pi} = \mathrm{RDCT}_{\mathcal{H},E,\Pi_1,\Pi_2}$. Let $\mathbf{A}$ be a $\textsc{detAuth}$ adversary on $\widetilde{\Pi}$ that submits at most q_e encryption queries and q_d decryption queries of at most m blocks in total, and runs in time at most t. Then

$$\mathbf{Adv}_{\widetilde{\Pi}}^{\textsc{detAuth}}(\mathbf{A}) \leq 3q^2\epsilon + \frac{q(q-1)}{2^{2n+1}} + \frac{q_d}{2^\tau} + \mathbf{Adv}_E^{\mathrm{SPRP}}(q, \mathcal{O}(t+q)),$$

where $q = q_e + q_d$.

Theorem 4 (DAE Advantage of RDCT)

Let $\widetilde{\Pi} = \mathrm{RDCT}_{\mathcal{H},E,\Pi_1,\Pi_2}$. *Let* $\mathbf{A}$ *be a* DAE *adversary on* $\widetilde{\Pi}$ *that asks at most* q_e *encryption queries and* q_d *decryption queries of at most* m *blocks in total and runs in time at most* t. *Then,* $\mathbf{Adv}_{\widetilde{\Pi}}^{\mathrm{DAE}}(\mathbf{A})$ *is upper bounded by*

$$\mathbf{Adv}_{\widetilde{\Pi}}^{\mathrm{DAE}}(\mathbf{A}) \leq 6q^2\epsilon + \frac{q^2}{2^{2n}} + \frac{q_d}{2^\tau} + 2\mathbf{Adv}_E^{\mathrm{SPRP}}(q, \mathcal{O}(t+q)) + \mathbf{Adv}_{\Pi_1}^{\mathrm{IVE}}(q_e, m, \mathcal{O}(t)),$$

where $q = q_e + q_d$.

Comparison

- When DCT is implemented using the BRW polynomial with a bound of $\epsilon = \mathcal{O}(\frac{m^2}{2^{2n}})$ [FLLW16], the provable bounds of DCT are $\mathcal{O}(\frac{q^2 m^2}{2^{2n}} + \frac{qm^2}{2^\tau})$.
- Let $u + v = \tau$, when the adversary makes $q = \mathcal{O}(2^v)$ decryption queries of $m = \mathcal{O}(2^{u+2})$ blocks, $\frac{qm^2}{2^\tau} > 1$.
- The security of DCT depends on the length of the query. However, **the security of RDCT is not affected by it**.

Comparison

Scheme	Provable security	Query complexity		Query length	Ref.
		Encryption	Decryption		
GCM	$\mathcal{O}(\frac{q^2m^2}{2^n} + \frac{qm}{2^\tau})$	1	$2^{\tau-u}$	2^{u+1}	[Fer05]
DCT	$\mathcal{O}(\frac{q^2m^2}{2^{2n}} + \frac{qm^2}{2^\tau})$	1	$2^{\tau-u}$	$2^{u+2}-3$	Sect. 5.4
RDCT	$\mathcal{O}(\frac{q^2m^2}{2^{2n}} + \frac{q}{2^\tau-q})$	0	2^τ	1	Sect. 6

* n: size of the message block, m: maximum number of blocks of a query, q: number of queries, τ: number of bits in the GCM tag or the redundancy of DCT and RDCT, u: user-selected parameter, $2 \le u \le \tau$. The query length is the input length of the underlying UHF.

Summary

- We show that although DCT employs the BRW polynomial to instantiate its UHF, it still suffers from a small stretch problem similar to that of GCM.
- We propose a variant of DCT named Robust DCT (RDCT) with minimal modification, which has a better security bound.

Thanks for Your Attention!

chenyuchao@mail.sdu.edu.cn for any question!