INDIAN STATISTICAL INSTITUTE

M. TECH. (COMPUTER SCIENCE) DISSERTATION

An Distinguishing Attack on LDT

Author:
MANABENDRA GIRI
Roll No: CS-1604
Mtech in Computer Science

Supervisor:
DR. MRIDUL NANDI
Assistant Professor
Applied Statistics Unit

A thesis submitted in fulfillment of the requirements for the degree of Mtech in Computer Science

July 8, 2018

CERTIFICATE

This is to certify that the dissertation entitled **“An Distinguishing Attack On LDT"** submitted by **Manabendra Giri** to Indian Statistical Institute, Kolkata, in partial fulfillment for the award of the degree of **Master of Technology in Computer Science** is a bonafide record of work carried out by him under my supervision and guidance. The dissertation has fulfilled all the requirements as per the regulations of this institute and, in my opinion, has reached the standard needed for submission.

Dr. Mridul Nandi
Assistant Professor
Applied Statistics Unit
Indian Statistical Institute,
Kolkata-700108, INDIA.

Declaration of Authorship

I, MANABENDRA GIRI, declare that this thesis titled, "An Distinguishing Attack on LDT" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a Mtech degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“One individual may die for an idea, but that idea will, after his death, incarnate itself in a thousand lives ”

Netaji Subhas Chandra Bose

INDIAN STATISTICAL INSTITUTE

Abstract

Mtech in Computer Science

An Distinguishing Attack on LDT

by MANABENDRA GIRI

Length doublers are cryptographic functions that transform an n bit tweakable block cipher into an efficient and secure cipher that length-preservingly encrypts any string of length in [n,...,(2n-1)] . One of them, LDT by Chen et al.(ToSC 2017) is secure up to $2^{n/2}$ queries . Let m $\in$ [kn,(n-1)] for some constant $k < 1$ .They described a possibility of attack against LDT in approximately $2^{n-(m/2)}$ queries of size $n + m$ based on distinguishing a truncated permutation from random function . We here describe attack for $k = 2/3$ .. . .

Acknowledgements

I would like to show my highest gratitude to my dissertation supervisor Prof. Mridul Nandi for agreeing to guide me and for helping me to undertake work in the topic.

I would like to thank all the teachers of Indian Statistical Institute, my parents and family and all of my friends for their help and support.

Manabendra Giri
M.Tech in Computer Science
Roll CS-1604
Indian Statistical Institute,
Kolkata.

Contents

List of Figures

Chapter 1

Introduction

Block ciphers are keyed-deterministic functions that can encrypt bit strings of fixed length n into bit strings of the same length. Many applications, however, deal with arbitrary-length messages, hence block ciphers on their own are not sufficient. Variable-input-length (VIL) encryption is achieved by evaluating a block cipher iteratively in a mode of operation. Basic solutions such as CBC mode can only encrypt messages of size a multiple of n. To handle messages whose size is not a multiple of n bits, one can pad the data to an integral number of n-bit blocks.

Padding is, in many cases, an undesirable solution: message length is typically not preserved, which means the resulting ciphertext will always be larger or equal to the original plaintext. This makes the solution unsuitable for disk encryption (where the size of ciphertext and plaintext must remain the same as the sector size of the disk) and low-bandwidth network protocols (as an increase in ciphertext length results in more data to be transmitted).

A generic method of length preserving variable length encryption is cyphertext stealing . Informally , it encrypts the first $(l-1)$ blocks as is, but to encrypt the non-integral lth block, it is expanded first to n bits by scraping sufficiently many ciphertext bits from $(l-1)$-th block and gluing these to M_l . But this approach only works on modes of use for which ciphertext block can be decrypted independently of each other ; otherwise one cannot recover the ciphertext bits scrapped off of C_{l-1}.

In 2007, Ristenpart and Rogaway introduced length doublers as an elegant way of achieving variable-length encryption. A length doublers is a length-preserving-encryption on the set of bit strings of size between n and (2n-1) bits, where n is the block size of underlying primitive .

Length-preserving VIL encryption can then be achieved by gluing a VIL encryption scheme for integral data blocks with the doubler. Length doublers are suitable solutions for various authenticated encryption schemes that treat integral and fractional data separately.

Chen et al. considered the design of length doublers from tweakable block ciphers and introduced LDT which makes 2 two calls of tweakable block cipher and uses a pure mixing function (generalisation of swap).Without loss of generality , here we replace pure mix function by swap. LDT is secure up to $2^{n/2}$ queries . Let m $\in [kn, (n-1)]$ for some constant $k < 1$ .They described a possibility of attack against LDT in approximately $2^{n-(m/2)}$ queries of size $n + m$ based on distinguishing a truncated permutation from random function . We here describe a PRP distinguishing attack for $k = 2/3$.

Chapter 2

All Related definitions

2.1 Notations

- For two bit strings $X,\ Y\ \in \{0,1\}^*$, we let $X||Y$ or XY be their concatenation and $X \oplus Y$ be their bitwise exclusive or..

- We denote by $|X|$ the length of the string X .

- For a natural number n, we denote by $\{0,1\}^n$ the set of bit strings of size n.

- For natural numbers $m \leq n$ we define $\{0,1\}^{[m,\ldots,n]}$.

- For some finite set S, we denote by $s \overset{\$}{\leftarrow} S$ the uniformly random selection of s from S .

- We denote by $Func(n,m)$ the set of all functions from $\{0,1\}^n$ to $\{0,1\}^m$.

- For a natural number n and $X \in \{0,1\}^{[0,\ldots(n-1)]}$, we define a padding function

  $$pad(X) \ = \ X||10^{n-|X|-1}.$$

  As the function is injective, we can consider its inverse *unpad* that on input of a string of length n removes the rightmost string 10^* and outputs the remainder.

- $Perm(n)$ is the set of all permutation on $\{0,1\}^n$

- $Perm(t,n)$ is the set of all functions $\pi \ : \ \{0,1\}^t \times \{0,1\}^n \ \rightarrow \ \{0,1\}^n$ such that $\pi(T,.)$ is in $Perm(n)$ for all $T \ \in \ \{0,1\}^t$

- $VPerm([n...2n-1])$ the set of all functions ρ that are length-preserving and invertible. Note that a randomly drawn function $\rho \overset{\$}{\leftarrow} Vperm([n..2n1])$ is equivalent to n random permutations $\rho_i \overset{\$}{\leftarrow} Perm(i)$ for $i \ = \ n \ ,...,2n-1$ as $\rho(M) \ = \ \rho_{|M|}(M)$.

2.2 Related Theorems

Definition 2.1 Statistical Distance : *Let X and Y be two random variables taking values on a finite set S. We define statistical distance between two random variables by*

$$\begin{aligned} d_{stat}(X,Y) &:= \underset{T\subseteq S}{max}|Pr[X\in T]-Pr[Y\in T]| \\ &= \underset{T\subseteq S}{max}|Pr[X\notin T]-Pr[Y\notin T]| \end{aligned}$$

The statistical distance is also popularly known as information theoretic distance.

Definition 2.2 Computational Distance :*Let $\mathcal{A}()$ be a probabilistic algorithm which runs with an input $a\in S$ and giving output 0 or 1. Define, $\mathcal{A}$-distance between X and Y as follows; by*

$$d^{\mathcal{A}}(X,Y) := |Pr[\mathcal{A}(X)=1]-Pr[\mathcal{A}(Y)=1]|$$

Here, A(X) means the distribution of output of A(z) where z follows the distribution of X. Similarly for A(Y).

Theorem 2.1 *For any function $f:\mathbb{R}\to\mathbb{R}$, $d_{stat}(X,Y)\geq d_{stat}(f(X),f(Y))$*

Theorem 2.2 Chebyshev's Inequality:- *Let X (integrable) be a random variable with finite expected value m and finite non-zero variance σ^2. Then for any real number $k>0$,* $\Pr(|X-m|\geq k\sigma)\leq \frac{1}{k^2}$.

2.3 Tweakable Block Cipher

For $k,t,n\in\mathbb{N}$, a tweakable block cipher is a function $E:\{0,1\}^k\times\{0,1\}^t\times\{0,1\}^n\to\{0,1\}^n$ such that for fixed key $K\in\{0,1\}^k$ and tweak $T\in\{0,1\}^t$, $E_K(T,\ )=E(K,T,\ )$ is a permutation on $\{0,1\}^n$. We denote its inverse (for fixed key and tweak) by $E_K^{-1}(T,\ )=E^{-1}(K,T,\ )$. The key is usually a secret parameter; the tweak is a public parameter, and E_K^{-1} should behave independently for different tweaks.The security of a tweakable block cipher E is measured by considering a distinguisher $\mathcal{D}$ that is given two-sided access to either E_K for secret key $K\xleftarrow{\$}\{0,1\}^k$ or a random tweakable permutation $\pi\xleftarrow{\$}Perm(t,n)$ and its goal is to determine which oracle it is given access to:

$$\mathbf{Adv}_E^{sprp}(\mathcal{D}) = |\ \ Pr\Big[K\xleftarrow{\$}\{0,1\}^k;\mathcal{D}^{E_K,E_K^{-1}}=1\Big]\ -\ Pr\Big[\pi\xleftarrow{\$}Perm(t,n);\mathcal{D}^{\pi,\pi^{-1}}=1\Big]\ \ |$$

2.4 Length Doubler

For $k,n\in\mathbb{N}$, a length doubler is a function

$$\mathcal{E}:\{0,1\}^k\times\{0,1\}^{[n,\ldots,2n-1]}\ \to\ \{0,1\}^{[n,\ldots,2n-1]}$$

such that for fixed key $K\in\{0,1\}^k$, $\mathcal{E}_K(\ .)=\mathcal{E}(K,\ .)$ is a length preserving invertible function on $\{0,1\}^{[n,\ldots,2n-1]}$. We denote its inverse (for fixed key) by $\mathcal{E}_K^{-1}(\ .)=\mathcal{E}^{-1}(K,\ .)$. $\mathcal{E}$ should behave like a random permutation for every length input $m\in[n..2n-1]$. The security of a length doubler $\mathcal{E}$ is measured by considering a distinguisher $\mathcal{D}$ that is given two-sided access to either $\mathcal{E}_K$ for secret key $K\xleftarrow{\$}\{0,1\}^k$ or a random length preserving permutation $\rho\xleftarrow{\$}VPerm([n,...,2n-1])$ and its goal is to determine which oracle it is given access to:

$$\mathbf{Adv}_{\mathcal{E}}^{vsprp}(\mathcal{D}) =$$
$$|\ \ Pr\Big[K \overset{\$}{\leftarrow} \{0,1\}^k; \mathcal{D}^{\mathcal{E}_K,\mathcal{E}_K^{-1}} = 1\Big]\ -\ Pr\Big[\rho \overset{\$}{\leftarrow} VPerm([n,...,2n-1]); \mathcal{D}^{\rho,\rho^{-1}} = 1\Big]\ \ |$$

2.5 LDT Doubler

Let $k, n \in \mathbb{N}$. Let $E : \{0,1\}^k \times \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a tweakable block cipher. Our length doubler $\mathcal{E} = LDT[E]$ with key space $\{0,1\}^{2k}$ and state $\{0,1\}^{[n..(2n-1)]}$ is given in Figure 1. Note that the decryption function is very similar to the encryption function and can be defined as

$$LDT[E]^{-1}_{K_1,K_2} = LDT[E^{-1}]_{K_2,K_1}$$

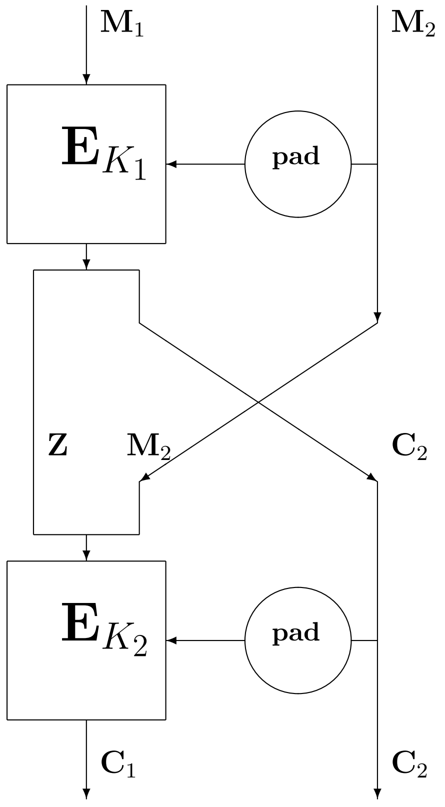

Figure 2.1: Encryption of length doubler LDT, with E a tweakable block cipher

Here $|\mathrm{M}_1| = |\mathrm{C}_1| = n\ >\ m = |\mathrm{M}_2| = |\mathrm{C}_2|$ and without loss of generality , we replace pure mix function by swap.

.

```
Algorithm 1 LDT[E,mix] encryption
Input : $(K_1, K_2) \in \{0,1\}^{2k}$,  $M_1||M_2 \in \{0,1\}^{[n,...,(2n-1)]}$  with $|M_1| = n$ and $|M_2| = s$
Output : $C_1||C_2 \in \{0,1\}^{n+s}$
process :
    1. $Z||U \leftarrow E_{K_1}(pad(M_2), M_1)$ with $|Z| = n - s$
    2. $V||C_2 \leftarrow mix(U, M_2)$
    3. $C_1 \leftarrow E_{K_2}(pad(C_2), Z||V)$
    4. return  $C_1||C_2$
```

Figure 2.2: LDT encryption algorithm

2.6 Mix function

Let U, V, M, C are all of s bit for m$\leq$ s $\leq$n and m,n,s$\in \mathbb{N}$. If $V||C = mix(U, M)$ then we define mix$_L$(U,M) = V = left half of its evalution and mix$_R$(U,M) = C = right half of its evalution . In above algorithm $mix : \cup_{s=m}^{n}(\{0,1\}^s)^2 \rightarrow \cup_{s=m}^{n}(\{0,1\}^s)^2$ be a length preserving permutation with following properties:

- mix$_L$(U,.) is a permutation for all U$\in \{0,1\}^s$
- mix$_R$(.,M) is a permutation for all M$\in \{0,1\}^s$

2.7 Truncation and Truncated Permutation

Let Truncation function **Trunc**$_{n,m}$: $\{0,1\}^n \rightarrow \{0,1\}^m$ be defined by the mapping $(x_n, x_{n-1}, ..., x_1)$ $\mapsto$ $(x_m, x_{m-1}, ..., x_m)$. The "**Truncated Permutation Family**" is defined by the composition **Trunc**$_{n,m} \circ \pi$, where π is a permutation on $\{0,1\}^n$, choosen uniformly at random.

Chapter 3

Previous Work

Chen et al studied the security of LDT length doubler. They declared the lower bound and upper bound stated below.

Theorem 3.1 *Let $k, n \in \mathbb{N}$. Let $E\{0,1\}^k \times \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a tweakable block cipher. Consider $\mathcal{E} = LDT[E, mix]$ is of Algorithm 1. For any distinguisher $\mathcal{D}$ making at most q queries,there exist distinguishers $\mathcal{D}_1$ and $\mathcal{D}_2$ with the same query complexity such that*

$$\mathbf{Adv}_{LDT}^{vsprp}(\mathcal{D}) \;\leq\; \mathbf{Adv}_{E}^{sprp}(\mathcal{D}_1) \;+\; \mathbf{Adv}_{E}^{sprp}(\mathcal{D}_2) \;+\; \frac{q(q-1)}{2^n}$$

Let m $\in [kn, (n-1)]$ for some constant $k < 1$ .They described a possibility of attack against LDT in approximately $2^{n-(m/2)}$ queries of size $n + m$ based on distinguishing a truncated permutation from random function .

Theorem 3.2 *Let $k, n \in \mathbb{N}$. Let $E\{0,1\}^k \times \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a tweakable block cipher.mix : $\cup_{s=m}^{n}(\{0,1\}^s)^2 \rightarrow \cup_{s=m}^{n}(\{0,1\}^s)^2$ be a mixing function. Consider $\mathcal{E} = LDT[E, mix]$ is of Algorithm 1. Let m $\in$ [$kn, (n-1)$] for some constant $k < 1$. Then there exists a distinguisher $\mathcal{D}$ making at most q queries, such that*

$$\mathbf{Adv}_{LDT}^{vsprp}(\mathcal{D}) \;\geq\; c_1\frac{q^2}{2^{2n-m}} - c_2\frac{q^2}{2^{2n+m}}$$

for some constant c_1, c_2

For increasing m, the 1st term becomes larger whereas the 2nd term becomes negliigible. For m=n-1, the bound is tight.

In original paper; type of attack, adversary algorithm and details analysis are not present. Our work is to give all this things for some certain range.

Chapter 4

Our contribution in short

We want to distinguish **LDT** from a random length preserving invertible function taking input from n-bit to (2n-1)-bit strings. We take m such that n≤(n+m)≤(2n-1) . Adversary do atmost q encryption queries only of distinct (n+m) bit strings keeping last m-bit same (length and value both) for all queries, say 00...0 [last m-bit of strings]. Due to construction of **LDT**, if 1st n-bit is ignored, it is simply a permutation on n-bit and then truncation from n-bit to m-bit strings .So we have advantage at least distinguishing advantage of truncated permutation from random string.

Here we show a PRP distinguishing attack between LDT and length preserving function. Using the reply of encryption queries (q) more than $1 + 50 \cdot 2^{(n-\frac{m}{2})}$ based on messages of size $(n+m)$ at least $4 + \frac{5}{3} \cdot n$, we distinguish LDT with at least 0.98 probability. Here n is block size.

Let X is total number of pairwise collision on reply of q queries. Let indexed 0 denote the oracle for truncated permutation on without replacement on n bits; and indexed 1 for random selection with replacement on m bits . Let m_i and σ_i be mean and standard deviation of X when oracle$_i$ is used. For $a > 10(1+\sqrt{2})$, $q > 1 + a\sqrt{2}\frac{N-1}{\sqrt{M-1}}$, $N = 2^n$ and $M = 2^m$; we get a threshold T such that

$$\mathrm{m}_0 + t\sigma_0 \ < \ \mathrm{T} \ < \mathrm{m}_1 - t\sigma_1.$$

with probability more than 0.98

Algorithm 2 Adversary

```
Input : Response of q encryption queries;     Last m bits of query messages are all 0
Output : 0/1
process :
    1. $X$ is total number of pairwise collision on last m bits.
    2. If $X > T$ return 1 otherwise return 0
       We will describe later the value of q, m and T
```

Figure 4.1: Adversary algorithm

$\boxed{\textbf{Advantage of adversary}}$:- Adv(q)=
$= |Pr[\textbf{Adv(random)} \implies 1] - Pr[\textbf{Adv(tr-Permut)} \implies 1]|$
$=|Pr[\textbf{X(random)} > T] - Pr[\textbf{X(tr-Permut)} > T]|$
$=|Pr[\textbf{X(random)} > T] + Pr[\textbf{X(tr-Permut)} < T] - 1|$

And $Pr[\textbf{X(random)} > T] + Pr[\textbf{X(tr-Permut)} < T] - 1$
$\geq Pr[\ |\textbf{X(random)} - \mathrm{m}_1| < t\sigma_1] + Pr[\ |\textbf{X(tr-Permut)} - \mathrm{m}_0| < t\sigma_0] - 1$
. if $\mathrm{m}_0 + t\sigma_0 < \mathrm{T} < \mathrm{m}_1 - t\sigma_1$
$\geq 2(1 - \frac{1}{t^2}) - 1 = 0.98$ if $t = 10$ (by Chebyshev's Inequality).

Chapter 5

Our work in details

Gilloba and Gueron studied truncated permutation and gave a tight bound. If $0 \leq m < n$ and $q > 1$. Then given a budget of q queries and truncation from n bit to m bit, Advantage of truncated permutation dishtinguishing from random function is

$$\mathbf{Adv}_{n,m}(q) = \Theta\left(\min\left\{1, \frac{q^2}{2^n}, \frac{q\sqrt{2^m}}{2^n}\right\}\right)$$

Then $\mathbf{Adv}_{n,m}(q) = 1 \implies 1 \leq \frac{q\sqrt{2^m}}{2^n} < \frac{q^2}{2^n}$. Therefore q should be at least order of $2^{n-\frac{m}{2}}$ ($q > 2^{\frac{n}{2}} = 2^{n-\frac{n}{2}}$, $q > 2^{n-\frac{m}{2}}$).

5.1 Truncate/Random String

Consider a set $\mathcal{N}$ of size N and a set $\mathcal{M}$ of size M. Let $trun$ be given $\frac{N}{M}$-regular function from $\mathcal{N}$ to $\mathcal{M}$. $PERM(N)$ = set of all permutations on $\mathcal{N}$. And $FUNC(N, M)$ = set of all functions from $\mathcal{N}$ to $\mathcal{M}$. Let $C_2^{(1)}, C_2^{(2)}, ..., C_2^{(q)}$ be values from $\mathcal{M}$ either choosen randomly or choosen a permutation π from $PERM(N)$ and apply $trun$ function on it. Let $w = (C_2^{(1)}, C_2^{(2)}, ..., C_2^{(q)})$ be transcript and lies in $(\mathcal{M})^q$, collection of all possible transcript.

We will introduce some parameters related to w.

- Let $\mathrm{I}_{(i,j)}$ be idicator variable gives 1 if $C_2^{(i)} = C_2^{(j)}$; otherwise gives 0. Then $\{\ \mathrm{I}_e : e \in \mathcal{I}\}$ be the collections of all indicator variables.
- Let $X = \sum\limits_{e \in \mathcal{I}} \mathrm{I}_e$. So X is total number of pairwise collision on reply of q queries.
- Let indexed 0 denote the oracle for truncated permutation on without replacement on $\mathcal{N}$; and indexed 1 for random selection with replacement on $\mathcal{M}$.
- Let m_i and σ_i be mean and standard deviation of X when oracle$_i$ is used.
- Let p_i is mean of I_e when oracle$_i$ is used.

5.2 Main Calculations

Let $r = \frac{M-1}{N-1}$, $\frac{1}{c \cdot t} > 1 + \sqrt{2}$ where c and t are constants; B and $|\mathcal{I}|$ is defined in chapter 4.4. We also use formulas given in chapter 4.4 to calculate mean and variance .

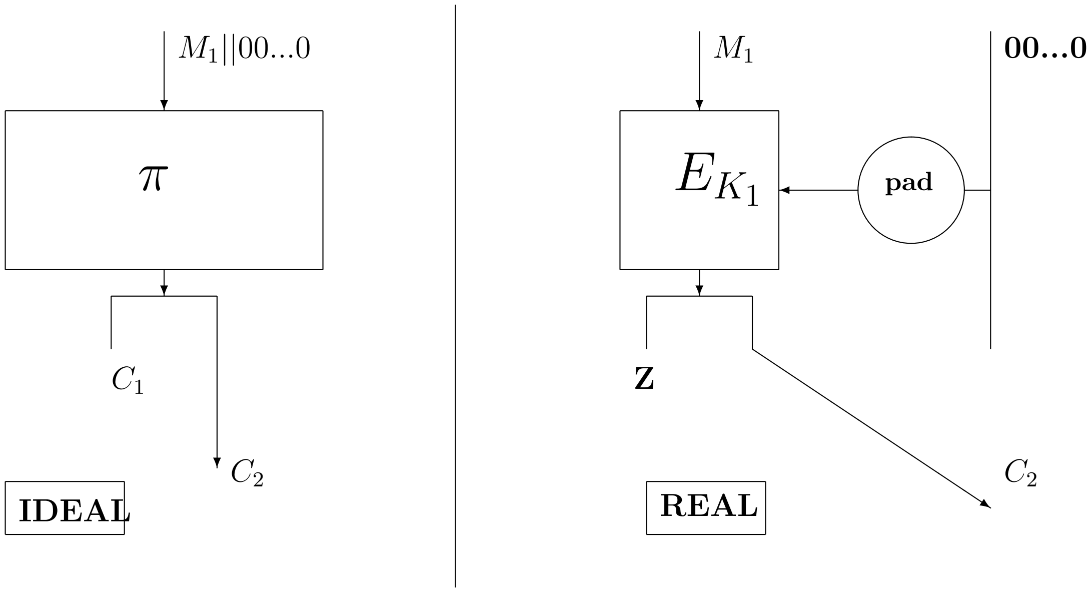

Figure 5.1: A part of LDT

- $p_1 = \frac{1}{M}$, $p_0 = \frac{N-M}{M(N-1)} \ = \ \frac{1}{M}(1-r) \ < \ \frac{1}{M}$
- Ideal World
  - **Mean :-** $m_1 = |\mathcal{I}|\frac{1}{M}$
  - **Variance :-** $Var_1[X] = |\mathcal{I}|\frac{1}{M}\frac{M-1}{M} \leq \left[\frac{|\mathcal{I}|}{M}\cdot c\cdot r\right]^2$ if $|\mathcal{I}|(cr)^2 > \ M-1$
- Real World
  - **Mean :-** $m_0 \ = \ |\mathcal{I}|\cdot\frac{1}{M}\left[1-r\right] \ < \ m_1$
  - So we want to find **T** such that m$_0$ + $t\sigma_0$ < T < m$_1$ − $t\sigma_1$.
  - **Variance :-** $Var_0[X] \ \leq \ |\mathcal{I}|\cdot\mathrm{Var}_0[\mathrm{I}_e] \ + \ 2\cdot 0 \ + \ 2B\cdot\mathrm{Cov}_0[\ \mathrm{I}_{(i,j)}\ ,\ \mathrm{I}_{(k,l)}\ ]$;
    . i,j,k,l are all distinct

$$\leq \ |\mathcal{I}|\frac{1}{M}\cdot(1-r)\frac{N}{M}r + 2\cdot 0 + 2B\cdot\frac{5}{M^3}$$
$$= \ \frac{|\mathcal{I}|}{M^2}\cdot r(1-r)N\cdot 1 + 2B\cdot\frac{5}{M^3}$$
$$\leq \ \frac{|\mathcal{I}|}{M^2}\cdot(M-1)\cdot|\mathcal{I}|(cr)^2\frac{1}{M-1} + |\mathcal{I}|^2\cdot\frac{5}{M^3}$$
$$\leq \ \left[\frac{|\mathcal{I}|}{M}\cdot c\cdot r\right]^2\cdot\left[1+\frac{5}{Mr^2c^2}\right] \ \leq \ 2\cdot\left[\frac{|\mathcal{I}|}{M}\cdot c\cdot r\right]^2 \ \text{ if } \ 5 \ < \ Mr^2c^2$$

- $\mathrm{m}_0 + t\sigma_0 \ \leq \ |\mathcal{I}|\frac{1}{M}(1-r+\sqrt{2}t\cdot c\cdot r)$
  $= \ |\mathcal{I}|\frac{1}{M}\big(1-(1-\sqrt{2}t\cdot c)r\big)$
- $\mathrm{m}_1 - t\sigma_1 \ \geq |\mathcal{I}|\frac{1}{M}(1-t\cdot c\cdot r) \ > \ m_0 + t\sigma_0$, if $(1-\sqrt{2}t\cdot c) \ > \ t\cdot c$
- If $q \geq 1+\frac{2}{c}\big(\frac{N}{\sqrt{M}}\big)$ and $\frac{5}{Mr^2c^2} \ < \ 1$ and $t = 10$
  - $|\mathcal{I}| \ = \ \frac{1}{2}q(q-1) \ > \ \frac{1}{2}(q-1)^2 \ > \ \frac{2N^2}{c^2M} \ > \ \frac{(N-1)^2}{c^2(M-1)} \ = \ (\frac{1}{rc})^2(M-1) \ > \ \frac{N^2}{c^2M}$
    $\implies |\mathcal{I}|\cdot(cr)^2\cdot\frac{1}{M-1} \ > \ 1$

- Adv(q) > 0.98
- $5 \; < \; Mr^2 \cdot c^2 \; = \; c^2 \cdot M(M-1)^2(N-1)^{-2}$
  $\iff \; 5 \cdot (N-1)^2 \; < \; c^2 \cdot M(M-1)^2$

- If n is block size, $c = \frac{2}{50}$ and $t = 10$; then the followings are sufficient for advantage of adversary > 0.98
  - no. of query $\geq 1 + 50 \cdot 2^{(n-\frac{m}{2})}$
  - length of query messages $\geq (4 + \frac{5}{3} \cdot n)$
  - T $= \frac{|\mathcal{I}|}{M}\big(1 - 0.417 \cdot \frac{M-1}{N-1}\big)$

5.3 Covariance

5.3.1 Results for covariance

For $(i,j) < (k,l)$ and $(i,j),(k,l) \in \mathcal{I}$,

- **In real world,** $\text{Cov}[\ \text{I}_{(i,j)},\text{I}_{(k,l)}\ ] \leq \begin{cases} \frac{5}{M^3} & \text{for} \quad \{i,j\} \cap \{k,l\} = \phi \\ 0 & \text{for} \quad \{i,j\} \cap \{k,l\} \neq \phi \end{cases}$
- **In ideal world,** $\text{Cov}[\ \text{I}_{(i,j)},\text{I}_{(k,l)}\ ] = 0$

5.3.2 Proof of results for covariance

Covariance of $Oracle_1$ (Ideal World)

- **If (i,j) $<$ (k,l); i,j,k,l are not all distinct**
  $Cov_1(I_{(i,j)}, I_{(j,l)}) = Cov_1(I_{(i,l)}, I_{(j,l)})$

  $= Cov_1(I_{(i,j)}, I_{(i,l)}) = Pr_1[C_2^{(i)} = C_2^{(j)} = C_2^{(l)}] - (p_1)^2$

  $= \; 1 \cdot \frac{1}{M} \cdot \frac{1}{M} - (\frac{1}{M})^2 \; = \; 0$
- **If (i,j) $<$ (k,l); i,j,k,l are all distinct**
  $Cov_1(I_{(i,j)}, I_{(k,l)}) = Pr_1[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)}] - (p_1)^2$

  $= \; 1 \cdot \frac{1}{M} \cdot 1 \cdot \frac{1}{M} - (\frac{1}{M})^2 \; = \; 0$

Covariance of $Oracle_0$ (Real World)

- **If (i,j) $<$ (k,l); i,j,k,l are not all distinct**
  $Cov_0(I_{(i,j)}, I_{(j,l)}) = Cov_0(I_{(i,l)}, I_{(j,l)})$

  $= Cov_0(I_{(i,j)}, I_{(i,l)}) = Pr_0[C_2^{(i)} = C_2^{(j)} = C_2^{(l)}] - (p_0)^2$

  $= \frac{N}{N} \cdot \frac{\frac{N}{M}-1}{N-1} \cdot \frac{\frac{N}{M}-2}{N-2} - \frac{(N-M)^2}{M^2(N-1)^2}$

  $= \frac{(N-M)}{M^2(N-1)} \cdot \frac{(M-1)(-N)}{(N-1)(N-2)} \; \leq \; 0$

- **If (i,j) $<$ (k,l); i,j,k,l are all distinct**

$$Cov_0(I_{(i,j)}, I_{(k,l)}) = Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)}] - (p_0)^2$$

$$= Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} = C_2^{(k)}] \cdot Pr_0[C_2^{(j)} = C_2^{(k)}]$$

$$+ Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} \neq C_2^{(k)}] \cdot Pr_0[C_2^{(j)} \neq C_2^{(k)}] \; -(p_0)^2$$

$$\leq \; Pr_0[C_2^{(i)} = C_2^{(j)} = C_2^{(k)} = C_2^{(l)}]$$

$$+ Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} \neq C_2^{(k)}] - (p_0)^2$$

$$\leq 2\tfrac{1}{M^3} + \tfrac{3}{M^3} \;\leq\; \tfrac{5}{M^3} \qquad \text{(probability is calculated below)}$$

$$\underline{Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} \neq C_2^{(k)}]}$$

$$Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} \neq C_2^{(k)}]$$

$$= \tfrac{N}{N} \cdot \tfrac{\frac{N}{M}-1}{N-1} \cdot \tfrac{N-\frac{N}{M}}{N-2} \cdot \tfrac{\frac{N}{M}-1}{N-3}$$

$$= \tfrac{N(N-M)^2(M-1)}{M^3(N-2)(N-3)(N-1)}$$

$$= \tfrac{(N-M)^2}{M^3(N-2)(N-3)(N-1)^2}(M+3)(N-2)(N-3) - N(N-5M)$$

$$= p_0^2 \tfrac{1}{M}(M+3)$$

$$= p_0^2 + 3 \cdot \tfrac{1}{M} \cdot p_0^2$$

$$= p_0^2 + 3 \cdot \tfrac{1}{M^3}$$

$$\underline{Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} = C_2^{(k)}]}$$

$$Pr_0[C_2^{(i)} = C_2^{(j)}; C_2^{(k)} = C_2^{(l)} | C_2^{(j)} = C_2^{(k)}]$$

$$= Pr_0[C_2^{(i)} = C_2^{(j)} = C_2^{(k)} = C_2^{(l)}] \;=\; 1 \cdot \tfrac{\frac{N}{M}-1}{N-1} \cdot \tfrac{\frac{N}{M}-2}{N-2} \cdot \tfrac{\frac{N}{M}-3}{N-3}$$

$$= \tfrac{1}{M} \cdot p_0^2 \cdot \tfrac{N-2M}{N-2} \cdot \tfrac{N-3M}{N-3} \cdot \tfrac{N-1}{N-M}$$

$$\leq \; \tfrac{1}{M} \cdot p_0^2 \cdot 1 \cdot 1 \tfrac{N-1}{N} \cdot \tfrac{N}{N-M} \;\; \leq \; \tfrac{1}{M} \cdot p_0^2 \cdot 1 \cdot 1 \cdot 1 \cdot \tfrac{2N}{2N-N}$$

$$\leq \; \tfrac{1}{M} \tfrac{1}{M^2} \cdot 2 \;=\; 2\tfrac{1}{M^3}$$

5.4 Indexed set and Indicator variable

1. **Indexed Set :**Let $\mathcal{I} = \{(i,j) : 1 \le i < j \le q\}$ be indexed set.
   (a) $|\mathcal{I}| \;=\; \frac{1}{2}q(q-1)$
   (b) Now i < j < l $\Leftrightarrow$ (i,j) < (j,l) ; (i,l) < (j,l) ; (i,j) < (i,l)
   (c) $|\{(i,j,l) : 1 \le i < j < l \le q\}| = \sum_{l=3}^{q}\sum_{j=2}^{l-1}(j-1)$
   $= \sum_{l=3}^{q}\sum_{j=1}^{l-2}(j) = \sum_{l=3}^{q}\frac{1}{2}\cdot(l-2)(l-1) = \sum_{l=1}^{q-2}\frac{1}{2}\cdot l(l+1)$
   $= \frac{1}{2}\cdot\frac{(q-2)(q-1)(2q-3)}{6} + \frac{(q-2)(q-1)}{2} = \frac{1}{12}\cdot(q-2)(q-1)[(2q-3)+3] = \frac{1}{6}\cdot(q-2)(q-1)q$
   (d) (i,j) < (k,l) means either i < j and i < k < l or i = k and i < j < l
   So (i,j) < (k,l) means
   either i < j and i < k < l and i,j,k,l are all different
   or j = k and i < j < l
   or j = l and i < k < l
   or i = k and i < j < l
   (e) $|\{$((i,j),(j,l)) $\in \mathcal{I}\times\mathcal{I}$: (i,j) < (j,l)$\}| = |\{(i,j,l) : 1 \le i < j < l \le q\}| = \frac{1}{6}\cdot(q-2)(q-1)q$
   $|\{$((i,l),(k,l)) $\in \mathcal{I}\times\mathcal{I}$: (i,l)<(k,l)$\}| = |\{(i,k,l) : 1 \le i < k < l \le q\}| = \frac{1}{6}\cdot(q-2)(q-1)q$
   $|\{$((i,j),(i,l)) $\in \mathcal{I}\times\mathcal{I}$: (i,j) < (i,l)$\}| = |\{(i,j,l) : 1 \le i < j < l \le q\}| = \frac{1}{6}\cdot(q-2)(q-1)q$
   (f) $B \;=\; |\{$((i,j),(k,l)) $\in \mathcal{I}\times\mathcal{I}$: (i,j) < (k,l) and i,j,k,l are all distinct $\}|$
   $=\frac{1}{2}\cdot|\mathcal{I}|(|\mathcal{I}|-1) - 3\cdot\frac{1}{6}\cdot(q-2)(q-1)q$
   $=\frac{1}{4}\cdot|\mathcal{I}|\cdot[\;2\cdot|\mathcal{I}| - 2 - 4\cdot(q-2)]$
   $=\frac{1}{4}\cdot|\mathcal{I}|\cdot[\;q^2 - q - 2 - 4q + 8]$
   $=\frac{1}{4}\cdot|\mathcal{I}|\cdot[\;q^2 - 5q + 6]$
   $= \frac{1}{8}\cdot q(q-1)(q-2)(q-3) \;<\; \frac{1}{2}\cdot|\mathcal{I}|^2$

2. **Indicator variable :**If I_e be indicator variable and $\mathbb{E}[\mathrm{I}_e] = p$ and $e, e_1, e_2 \in \mathcal{I}$ then
   (a) $\mathrm{Var}[\mathrm{I}_e] = p - p^2$
   (b) $\mathbb{E}[\sum_{e}\mathrm{I}_e\,] = |\mathcal{I}|\cdot\mathbb{E}[\mathrm{I}_e] = |\mathcal{I}|\cdot p$, where $e \;\in\; \mathcal{I}$
   (c) $\mathrm{Var}[\,\sum_{e\,\in\,\mathcal{I}}\mathrm{I}_e\,] = \;|\mathcal{I}|\cdot\mathrm{Var}[\mathrm{I}_e] \;+\; 2\cdot\Big(\sum_{e_1<e_2}\mathrm{Cov}[\;\mathrm{I}_{e_1},\mathrm{I}_{e_2}\;]\Big)$
   $= \;|\mathcal{I}|\cdot(p-p^2) \;+\; 2\cdot\Big(\sum_{e_1<e_2}\mathrm{Cov}[\;\mathrm{I}_{e_1},\mathrm{I}_{e_2}\;]\Big)$, where $e_1 < e_2$.
   (d) $Cov(I_{(i,j)}, I_{(j,l)}) = Pr[C_2^{(i)} = C_2^{(j)}, C_2^{(j)} = C_2^{(l)}] - p^2 = Pr[C_2^{(i)} = C_2^{(j)} = C_2^{(l)}] - p^2$
   Similarly $Cov(I_{(i,j)}, I_{(j,l)}) = Cov(I_{(i,l)}, I_{(j,l)}) = Cov(I_{(i,j)}, I_{(i,l)})$
   $= Pr[C_2^{(i)} = C_2^{(j)} = C_2^{(l)}] - p^2$

5.5 Helping Calculations

1. **constant:** c and t are constants such that $c > 0$, $t > 1$, $t\cdot c \;<\; \frac{1}{1+\sqrt{2}}$. t has impact on advantage and c has impact on query. . Consider a sequence $\{k_n\}$ where $(n\cdot k_n - 1)^2 \le 2\cdot n^2 \;<\; (n\cdot k_n)^2$ and $(n\cdot k_n) \in \mathbb{N}$. Then $\sqrt{2} \;<\; k_n \le \sqrt{2} + \frac{1}{n}$. Then we can replace $t\cdot c$ by suitable $\frac{1}{1+k_n}$. Example, for $n = 2$ we have $k_2 = \frac{3}{2}$ and $c = \frac{3}{5}\cdot\frac{1}{t}$; for $n = 31$ we have $k_{31} = \frac{44}{31}$ and $c = \frac{31}{75}\cdot\frac{1}{t}$.

2. Let $q \geq 1 + \frac{2}{c}\left(\frac{N}{\sqrt{M}}\right)$, and $r = \frac{M-1}{N-1}$

3. $r(1-r)N \ = \ \frac{N-M}{N-1} \cdot \frac{N}{N-1} \cdot (M-1) \ < \ M-1$

4. $M \ \leq \ \frac{1}{2}N$; and let $N \ \geq \ 10$, then
   $N(M-1)(N-1) \ - \ (N-2)(N-3)M$
   $< \ N(M-1)N \ - \ (N-2)(N-3)M$
   $< N^2(M-1) \ - \ (N^2-5N)M$
   $= \ N(5M-N) \ \leq \ N(\frac{5}{2}N-N) \ = \ \frac{3}{2}N^2$
   $\leq \ 3N(N-5) \ \leq 3(N-2)(N-3)$
   So, $N(M-1)(N-1) \ < \ (M+3)(N-2)(N-3)$

Bibliography

[1] Yu Long Chen, Atul Luykx, Bart Menninkand, Bart Preneel : Efficient Length Doubling From Tweakable Block Cipher

[2] Mridul Nandi : A Simple and Unified Method of Proving Indistinguishability (Extended Abstract)

[3] Shoni Gilboa and Shay Gueron. The Advantage of Truncated Permutations. arXiv:1610.02518v1 [math.CO] 8 Oct 2016

[4] Kazuhiko Minematsu and Tetsu Iwata. Tweak-Length Extension for Tweakable Blockciphers. https://eprint.iacr.org/2015/888.pdf

[5] Srimanta Bhattacharya and Mridul Nandi. A note on the chi-square method: A tool for proving cryptographic security. Cryptography and Communications September 2018, Volume 10, Issue 5, pp 935957. https://link.springer.com/article/10.1007/s12095-017-0276-z

[6] Jacques Patarin. The "Coefficients H" Technique. International Workshop on Selected Areas in Cryptography/ SAC 2008: Selected Areas in Cryptography pp 328-345. https://link.springer.com/chapter/10.1007/978-3-642-04159-4_21