# Accurate Estimation of the Full Differential Distribution for General Feistel Structures

Jiageng Chen[1⋆], Atsuko Miyaji[2,3,4⋆⋆], Chunhua Su[2⋆⋆⋆], and Je Sen Teh[5]

[1] Computer School, Central China Normal University, Wuhan 430079, China,
[2] School of Information Science,
Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
[3] Japan Science and Technology Agency (JST) CREST,
Kawaguchi Center Building 4–1–8, Honcho, Kawaguchi-shi,
Saitama, 332–0012 Japan,
[4] Graduate School of Engineering, Osaka University, Japan,
[5] Information Security Lab, Universiti Sains Malaysia, Malaysia..
chinkako@gmail.com, miyaji@jaist.ac.jp, chsu@jaist.ac.jp,
jesen_teh@hotmail.com

**Abstract.** Statistical cryptanalysis is one of the most powerful tools to analyze symmetric key cryptographic primitives such as block ciphers. One of these attacks, the differential attack has been demonstrated to break a wide range of block ciphers. Block cipher proposals previously obtain a rough estimate of their security margin against differential attacks by counting the number of active S-Box along a differential path. However this method does not take into account the complex clustering effect of multiple differential paths. Analysis under full differential distributions have been studied for some extremely lightweight block ciphers such as KATAN and SIMON, but is still unknown for ciphers with relatively large block sizes. In this paper, we provide a framework to accurately estimate the full differential distribution of General Feistel Structure (GFS) block ciphers with relatively large block sizes. This framework acts as a convenient tool for block cipher designers to determine the security margin of their ciphers against differential attacks. We describe our theoretical model and demonstrate its correctness by performing experimental verification on a toy GFS cipher. We then apply our framework to two concrete GFS ciphers, LBlock and TWINE to derive their full differential distribution by using super computer. Based on the results, we are able to attack 25 rounds of TWINE-128 using a distinguishing attack, which is comparable to the best attack to date. Besides that, we are able to depict a correlation between the hamming weight of an input differential characteristic and the complexity of the attack.

⋆ This study is partly supported by the National Natural Science Foundation of China under Grant 61302161.
⋆⋆ This study is partly supported by Grant-in-Aid for Scientific Research (C)(15K00183) and (15K00189).
⋆⋆⋆ This study is partly supported by JSPS KAKENHI 15K16005.

Based on the proposed framework, LBlock and TWINE have shown to have 178 and 208-bit security respectively.

**Keywords:** Differential attack, GFS , differential distribution , LBlock, TWINE

## 1 Introduction

Block ciphers have been playing an important role in information security to achieve confidentiality and integrity. Recently, block ciphers with lightweight designs start attracting research attention due to their wide range of potential applications such as RFID, wireless sensor networks and etcetera. These lightweight block ciphers usually have small block sizes which are less or equal to 64 bits and a smaller key size, filling in the gap where the traditional ciphers such as AES are not applicable anymore. The General Feistel Structure (GFS) is among one of the most popular designs that have received a lot of analysis. Recently proposed lightweight ciphers such as LBlock [22] and TWINE [21] belong to this design category.

Among all the methods to analyze block ciphers, differential attacks are one of the most powerful methods since its invention back in 1990 [5]. The attack is statistical in nature and its success relies on finding long differential paths with high probability. For a long time, one single ad hocfound path is usually used in the differential cryptanalysis. Thus the study of the differential path has not received much attention until recently. First in papers [8] and [9], multiple differential cryptanalysis was theoretically analyzed to show that the attacker generally has more power in building the differential distinguisher if he or she has more knowledge in the differential distribution. Later in paper [1], the author analyzed an extremely lightweight block cipher, KATAN32 by computing the whole differential distribution, and indeed it further increased the number of rounds that can be attacked compared to the previous results. The downside of using the whole differential distribution is that the attacker is unable to filter subkey bits, which may cause the complexity to increase. Thus there exists another branch of research focusing more on the key recovery phase and key relation such as related key attacks. Representative results include [6] and [19] which will not be addressed further in this paper since our focus is only the single key model. The full differential distribution can be computed if the block size is less than 32 bits, as shown in [1]. However, for ciphers with large block sizes, it is currently computationally infeasible

to construct the full distribution. Thus to a large extent, the method to derive an accurate full distribution remains unexploited.

From the provable security's point of view, it is desirable to derive a security bound on the number of rounds that is secure against differential attack. Currently for block ciphers with S-Box-based design, counting the number of active S-Box [18], which is the number of S-Box on the differential path, is the common way to evaluate the security. In the proposal of both LBlock and TWINE, the number of active S-Box multiplied by the largest differential probability of the S-Box is used to evaluate security margin. For more complicated designs which involves MixColumn operation as in AES, paper [17] provided a tight lower bound for the minimum number of active S-box for several GFS ciphers. Although counting the number of active S-Box may be a good approximation for one single path, the actual differential distribution involves complicated clustering effects which cannot be addressed by this model. Thus the security margin evaluated in this way may not be accurate, or in other words, the lower bound may be underestimated.

In this paper, we contribute mainly in two aspects. Firstly, we address the full differential distribution for GFS ciphers with relatively large block sizes by providing both theoretical and experimental frameworks. We partition the block according to the length of the S-Box input, which is the size of data blocks processed by these ciphers. Then we theoretically model the computation of the full differential distribution for any number of rounds and verify our evaluation by using a toy GFS cipher to show that the truncated differential distribution can be used to accurately evaluate the concrete differential distribution. Furthermore, due to the truncated differentials, the ability to store all the internal states allow us to perform quick computing of the distribution even for large rounds. By taking advantage of the supercomputer, we can perform the experiment to obtain full differential distributions for every input difference. As a result, our experiments have provided us with several new findings regarding the differential attack. Firstly, we discovered that input differences with relatively small hamming weights tend to lead to better distinguishers. Based on our framework, we evaluate two GFS ciphers LBlock and TWINE to derive the best differential attack so far. Especially for TWINE-128, we are able to obtain a comparable result by attacking 25 rounds. Also, we are able to provide the precise security margins against differential attacks for the full rounds of both LBlock and TWINE for the first time. This is by far the most accurate security proof for GFS designs to date.

**Outline of the paper.** Section 2 provides the theoretical model to compute the complete differential distribution for truncated GFS with bijective S-box design. Experiments on the toy model are also provided in this Section to verify the correctness of the model. In Section 3, concrete evaluations on LBlock and TWINE are provided. Lastly, we conclude our paper with some final statements.

## 2 Differential characteristic revisited

Since the proposal of differential attack in [5], methods to find long differential paths with high probability becomes the key to the success of the attack. Matsui in [13] first proposed a branch and bound algorithm to efficiently search the high probability linear and differential path for cipher DES. The algorithm applies the greedy strategy to find the best single path with the highest probability. Since then, researchers began to follow this strategy when searching for good property paths. As an extension of the differential attack, the multi-differential attack tries to take advantage of multiple differential paths to further increase the attacker's advantage when distinguishing from random distribution. Works [8] and [9] are two of the representative ones. For block ciphers with S-Box based design, researchers count the number of active S-Box as a criteria to measure the security margin against differential attack. It is well known [11] that there usually exists more than one path that can lead from the same input $\alpha$ to the output $\beta$, so that the probability of the corresponding path is actually bigger. Unfortunately, researchers usually do not consider this differential cluster or linear hull effect when searching good paths. [7] recently took advantage of the differential cluster to further improve the rounds of the differential paths.

Let's assume a block cipher $E$ is a markov cipher with $n$-bit block size and $r_f$ rounds in total. Previously, researchers try to identify one single $r < r_f$ round path $\alpha_0 \rightarrow \beta_{r-1}$ with high probability $Prob(\alpha_0 \rightarrow \beta_{r-1}) > 2^{-n}$, so that the attacker does not use up the entire message space. Usually, $r$ is far from the full rounds $r_f$ if the cipher is well designed. If we continue the search for more rounds, we will end up with a single path with a tiny probability much smaller than $2^{-n}$. On the other hand, if we assume all the differential paths are randomly distributed, for a full $r_f$-round cipher, the probability of any differential path $Prob(\alpha_0 \rightarrow \beta_{r_f-1})$ should be around $2^{-n}$. Obviously, there is a gap between the two results. From the differential cluster or linear hull effect, we make the following assumption.

**Lemma 1.** *For an r-round ideal Markov block cipher E, a single r-round differential path is defined as* $(\alpha_0 \to \beta_{r-1})_{single} = (\alpha_0, \gamma_{1,i_1}, \gamma_{2,i_2}, ..., \gamma_{r-2,i_{r-2}}, \beta_{r-1})$*, where* $I_t^{min} \le i_t \le I_t^{max}, 1 \le t \le r-2$*. Here* $I_t^{min}$ *and* $I_t^{max}$ *denote the smallest and largest differential values in round t respectively. Let's define its probability to be* $Prob((\alpha_0 \to \beta_{r-1})_{single}) = p_{i_1,i_2,...,i_{r-2}}$*. Then the total probability of differential path* $\alpha_0 \to \beta_{r-1}$ *can be computed by*

$$Prob(\alpha_0 \to \beta_{r-1}) = \sum_{i_1=I_1^{min}}^{I_1^{max}} \cdots \sum_{i_{r-2}=I_{r-2}^{min}}^{I_{r-2}^{max}} p_{i_1,i_2,...,i_{r-2}} \approx 2^{-n}$$

*which is approximately equal to* $2^{-n}$*. And we call*

$$CS_{(\alpha_0,\beta_{r-1})} = \sum_{i_1=I_1^{min}}^{I_1^{max}} \cdots \sum_{i_{r-2}=I_{r-2}^{min}}^{I_{r-2}^{max}} 1$$

*the corresponding cluster size* $CS_{(\alpha_0,\beta_{r-1})}$*.*

For large number of rounds $r$, we may assume $p_{i_1,i_2,...,i_{r-2}}$ to be tiny and have the relation $p_{i_1,i_2,...,i_{r-2}} \propto CS^{-1}_{(\alpha_0,\beta_{r-1})}$. As a result, the complexity to find the real probability of some specific path is related to the corresponding cluster size $CS_{(\alpha_0,\beta_{r-1})}$. As the number of rounds grow, cluster size becomes bigger which makes it more difficult to compute the real probability. Also notice that for real cipher, the probability varies for different paths and the cluster size is related to the input differential property. This relation will be discussed later in this paper. Next, we will discuss first how to theoretically evaluate the cluster size and the probability, and then efficiently compute the full clusters for GFS ciphers based on bijective S-Box design.

### 2.1 Theoretical Model to Evaluate the Cluster Size and Probability

General Feistel Structure (GFS) is one of the most popular and widely studied design strategies for constructing block ciphers. Recently in paper [20], the authors studied different permutations and derived the optimized ones for different parameter settings. Recently proposed lightweight block ciphers LBlock [22] and TWINE [21] belong to the GFS design.

In GFS, the plaintext is divided into $d$ subblocks $P = (x_0^0, x_1^0, ..., x_{d-1}^0)$, where $|x_j^i| = 2^{n/d}$ bits in length. The output of the $i$-th round is derived as follows:

$$(x_0^i, x_1^i, ..., x_{d-1}^i) \leftarrow \pi(x_0^{i-1}, F^{i-1}(x_0^{i-1}) \oplus x_1^{i-1}, ..., F^{i-1}(x_{d-2}^{i-1}) \oplus x_{d-1}^{i-1})$$

where $\pi$ is the permutation, and function $F : \{0,1\}^{n/d} \rightarrow \{0,1\}^{n/d}$ is the only non-linear function in GFS. For S-box based design with large subblock size $n/d$, usually MDS matrix is applied to provide further mixing within each subblock. However, in recent lightweight designs such as [22] and [21], $n/d$ is small in size (usually 4 bits), and $F$ is equivalent to a single S-Box. Figure 1 shows the GFS8 defined in [20] with two corresponding $F$ functions. For the simplicity, in this paper we will stick to the lightweight version of GFS without the application of MDS.

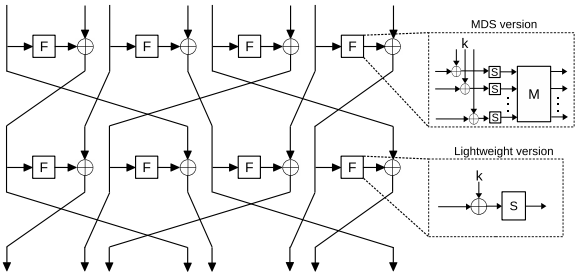


Fig. 1: GFS8[20]

Below are some definitions that will be used for the theoretical evaluation. From now on, we use symbol $\alpha^C$ and $\alpha^T$ to denote a concrete differential and a truncated differential respectively.

**Definition 1.** ***(Structure, Branch Weight, Hamming Weight, Cancel Weight).*** *Let $\alpha^{C,i} = (\alpha_0^{C,i}, \alpha_1^{C,i}, ..., \alpha_{d-1}^{C,i})$ denote the concrete differential states for each of the rounds $0 \leq i \leq N-1$. Function Trunc maps the concrete differential state to the truncated differential state: $\alpha^{T,i} = (\alpha_0^{T,i}, \alpha_1^{T,i}, ..., \alpha_{d-1}^{T,i}) \leftarrow Trunc(\alpha_0^{C,i}, \alpha_1^{C,i}, ..., \alpha_{d-1}^{C,i})$, where $\alpha_j^{T,i} = 1$ if $\alpha_j^{C,i} \neq 0$, and $\alpha_j^{T,i} = 0$ if $\alpha_j^{C,i} = 0$. We call*

$$(\alpha^{T,0}, \alpha^{T,1}, ..., \alpha^{T,r})$$

*a $r$-round truncated structure, or structure in short. We define the number of active S-Box of round $i$*

$$B_i = B_i(\alpha^{T,i}) = \alpha_0^{T,i} + \alpha_2^{T,i} + \cdots + \alpha_{d-2}^{T,i}$$

*to be the* ***Branch Weight*** *of the corresponding round. We define the* ***Hamming Weight*** *of the i-th round differential state to be*

$$H_i = H_i(\alpha^{T,i}) = \sum_{j=0}^{d-1} \alpha_j^{T,i}$$

*Finally, we define the* ***Canceling Weight*** $G_i$ *and* ***Non-Canceling Weight*** $W_i$ *for round i to be*

$$G_i = \alpha_0^{T,i} \wedge \alpha_1^{T,i} \wedge \neg\alpha_1^{T^{-1},i+1} + \cdots + \alpha_{d-2}^{T,i} \wedge \alpha_{d-1}^{T,i} \wedge \neg\alpha_{d-1}^{T^{-1},i+1}$$

$$W_i = \alpha_0^{T,i} \wedge \alpha_1^{T,i} \wedge \alpha_1^{T^{-1},i+1} + \cdots + \alpha_{d-2}^{T,i} \wedge \alpha_{d-1}^{T,i} \wedge \alpha_{d-1}^{T^{-1},i+1}$$

*where* $(\alpha_0^{T^{-1},i+1}, \alpha_1^{T^{-1},i+1}, ..., \alpha_{d-1}^{T^{-1},i+1}) \leftarrow \pi^{-1}(\alpha_0^{T,i+1}, \alpha_1^{T,i+1}, ..., \alpha_{d-1}^{T,i+1})$

$G_i$ counts the number of instances in round $i$ where $\alpha_j^{T,i} = \alpha_{j+1}^{T,i} = 1$ while $\alpha_{j+1}^{T^{-1},i+1} = 0$, and $W_i$ counts the number of instances in round $i$ where $\alpha_j^{T,i} = \alpha_{j+1}^{T,i} = \alpha_{j+1}^{T^{-1},i+1} = 1$. Now we are ready to have the following theorem:

**Lemma 2.** *Let* $\alpha_I^{C,0} \rightarrow \alpha_O^{C,r}$ *be a r-round concrete differential path with* $I \in \Omega_i$ *and* $O \in \Delta_o$. $\Omega_i$ *and* $\Delta_o$ *denotes the concrete differential set following the i-th input and o-th output truncated difference. Assume we have in total m structures which have the same truncated input and output* $\alpha_{\Omega_I}^{T,0}, \alpha_{\Delta_O}^{T,r}$ *while differing in the middle, we call m the truncated cluster size of truncated path* $(\alpha_{\Omega_I}^{T,0} \rightarrow \alpha_{\Delta_O}^{T,r})$. *The jth structure can be presented as follows* $(0 \leq j \leq m-1)$*:*

$$(\alpha_{\Omega_I}^{T,0}, \alpha^{T,1,j}, ..., \alpha^{T,r-1,j}, \alpha_{\Delta_O}^{T,r})$$

*Let's assume before proceeding round* $0 \leq i \leq r-1$ *in the jth structure, we have* $L_i^j$ *concrete differential paths which are resulted from input differential* $\alpha^{C,0}$. *Then after i-th round, the number of total paths generated from* $\alpha^{C,0}$ *becomes*

$$L_{i+1}^j = L_i^j \times R^{B_i^j} \times (2^{\frac{n}{d}} - 1)^{-G_i^j} \times (\frac{2^{\frac{n}{d}} - 1}{2^{\frac{n}{d}} - 2})^{-W_i^j}$$

*where R is the average branch number of the S-Box, and* $L_0^j = 1$ *(initially, there exists only one state). Then* $L_r^j$ *can be denoted as*

$$L_r^j = R^{\sum_{i=0}^{r-1} B_i^j} \cdot (2^{\frac{n}{d}} - 1)^{-\sum_{i=0}^{r-1} G_i^j} \cdot (\frac{2^{\frac{n}{d}} - 1}{2^{\frac{n}{d}} - 2})^{-\sum_{i=0}^{r-1} W_i^j}$$

*Proof.* For the $j$th structure $(\alpha_{\Omega_I}^{T,0}, \alpha^{T,1,j}, ..., \alpha^{T,r-1,j}, \alpha_{\Delta_O}^{T,r})$, we can easily compute parameters $B_i^j, H_i^j, W_i^j$ and $G_i^j$ for each round $i$. Assume before proceeding $i$-th round, we have $L_i^j$ concrete differential paths which are derived from the input differential $\alpha^{C,0}$ which follows the truncated form $\alpha^{T,0}$. Since there are $B_i^j$ active S-Box in this round, the increasing number of branches for each of the existed path can be computed as $R^{B_i^j}$. However, for each of the $G_i^j$ XOR operation, we know from the next round truncated pattern, the two input differences will be canceled out. The probability for this event to happen is $(2^{\frac{n}{d}}-1)^{-G_i^j}$. Also for each of the $W_i^j$ XOR operations, instead of probability 1, we need to exclude the cases where 0 may appear, thus the probability for this event to happen is $(2^{\frac{n}{d}}-1)^{-W_i^j}$. Since we need the concrete paths to follow the truncated pattern, only the paths that follow the truncated pattern can survive. As a result, we have $L_{i+1}^j = L_i^j \times R^{B_i^j} \times (2^{\frac{n}{d}}-1)^{-G_i^j} \times (2^{\frac{n}{d}}-1)^{-W_i^j}$ number of paths remaining. By computing this repeatedly, we can derive the total number of paths $L_r^j$ after $r$-th round. □

**Theorem 1.** *Assume we have $2^N$ concrete input differentials having the same truncated input difference, and the average single path probability for the truncated structure is $P_{ave}^{\sum_{i=0}^{r-1} B_i^j}$. Let the counter $X^j$ denote the number of hits for any concrete output differences following the same output truncated difference $\alpha_{\Omega_I}^{T,r}$ in the $j$-th structure. Then*

$$X^j_{\alpha_{\Omega_I}^{C,0}, \alpha_{\Delta_O}^{C,r}} \sim \mathcal{B}(2^N \cdot L_r^j, \quad (2^{n/d}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j}) \approx$$

$$\mathcal{N}\Bigg(2^N \cdot L_r^j \cdot (2^{\frac{n}{d}}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j}, \quad 2^N \cdot L_r^j$$
$$\cdot (2^{\frac{n}{d}}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j} \cdot (1-(2^{\frac{n}{d}}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j})\Bigg)$$

*Denote random variable $P^j = \frac{1}{2^N} \cdot X^j$ be the probability for the concrete path $\alpha_{\Omega_I}^{C,0} \to \alpha_{\Delta_O}^{C,r}$, and let $\Gamma_j^r = \frac{(2^{\frac{n}{d}}-1)^{-\sum_{i=0}^{r-1}(G_i^j+W_i^j)-H_r}}{(2^{\frac{n}{d}}-2)^{-\sum_{i=0}^{r-1} W_i^j}}$, then*

$$P^j_{(\alpha_{\Omega_I}^{C,0} \to \alpha_{\Delta_O}^{C,r})} \sim \mathcal{N}\Bigg(\Gamma_j^r, \quad (\Gamma_j^r \cdot (1-(2^{n/d}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i}))/2^N\Bigg)$$

*where $P_{ave}$ is the average differential probability of the S-Box. $2^N$ should satisfy the condition*

$$\frac{10}{\Gamma_j^r} \leq 2^N \leq (2^{\frac{n}{d}}-1)^{H_0}$$

*Proof.* Since the truncated output difference has hamming weight $H_r$, the concrete differential space is $(2^{n/d}-1)^{H_r}$ (excluding the 0 case). For any $\alpha_{\Delta_{O,j}}^{C,r} \in \{0,1\}^{log(2^{n/d}-1)^{H_r}}$, the probability that it gets hit by the $2^N L_i^j$ paths $x$ times follows the binomial distribution $\mathcal{B}(2^N \cdot L_r^j, \quad (2^{n/d}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j})$. Since $2^N L_i^j$ is large, we can approximate it by normal distribution as shown above. To derive its probability distribution, we only need to divide by the number of total pairs $2^N$. After extending $L_r^j$ as above, branch number $R$ is canceled by $P_{ave}$ since for any $S$-Box, $R \cdot P_{ave} = 1$. Replace with $\Gamma_j^r$ we derive the result. Notice that the mean of the distribution is not affected by the number of input pairs $2^N$.

$$P^j_{(\alpha_{\Omega_I}^{C,0} \to \alpha_{\Delta_O}^{C,r})} \sim$$

$$\mathcal{N}\Bigg(L_r^j \cdot (2^{\frac{n}{d}}-1)^{-H_r^j} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j}, \quad (L_r^j \cdot (2^{\frac{n}{d}}-1)^{-H_r^j} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j} \cdot (1-(2^{\frac{n}{d}}-1)^{-H_r^j} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i^j}))/2^N\Bigg) =$$

$$\mathcal{N}\Bigg((R\cdot P_{ave})^{\sum_{i=0}^{r-1} B_i} \cdot \Gamma_j^r, \quad ((R\cdot P_{ave})^{\sum_{i=0}^{r-1} B_i} \cdot \Gamma_j^r \cdot (1-(2^{n/d}-1)^{-H_r^j} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i}))/2^N\Bigg) =$$

$$\mathcal{N}\Bigg(\Gamma_j^r, \quad (\Gamma_j^r \cdot (1-(2^{n/d}-1)^{-H_r^j} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i}))/2^N\Bigg)$$

The maximum number of pairs $2^N$ is upper bounded by the input hamming weight $H_0$, while the lower bound can be derived from the condition of good approximation of binomial distribution by using normal distribution, which requires $np > 10$ for binomial distribution $\mathcal{B}(n,p)$. □

**Corollary 1.** *The distribution of probability* $(\alpha_{\Omega_I}^{C,0} \to \alpha_{\Delta_O}^{C,r})$ *after considering the entire truncated cluster with size* $m$ *has the following distribution.*

$$P_{(\alpha_{\Omega_I}^{C,0} \to \alpha_{\Delta_O}^{C,r})} \sim \mathcal{N}\Bigg(\sum_{j=0}^{m-1} \Gamma_j^r, \quad \sum_{j=0}^{m-1} \Gamma_j^r/2^N\Bigg)$$

Corollary 1 is straightforward by taking the truncated cluster into consideration. Notice that for large number of rounds, $(1-(2^{n/d}-1)^{-H_r} \cdot P_{ave}^{\sum_{i=0}^{r-1} B_i})$ can be approximated to be one, and thus the distribution can be simplified as stated.

Since for any S-Box, we know that $R \cdot P_{ave} = 1$, thus the expect value will converge to some stable value $\sum \Gamma$ as the number of rounds become large. Actually, we can see that as the number of rounds becomes large, the probability of the paths tends to gather around the mean.

### 2.2 Experimental Verification

The evaluation of the probability for the concrete differential cluster is the key to the attack. Thus it is necessary to verify the correctness of the probability calculation, especially, the mean ($\Gamma$) of the probability distribution in Corollary 1. Our experiment has the following settings.

1. We design a toy version of GFS cipher. It has 32-bit block size with 8 4-bit subblocks. TWINE's S-Box is applied and we apply the optimal block shuffle No.2 for $k = 8$ from [20] as the permutation layer to guarantee good diffusion property. It can be seen as a smaller block size version of TWINE.
2. We target 7 rounds differential path and choose the truncated input difference $\alpha_{\Omega_I}^{T,0}$ and output difference $\alpha_{\Delta O}^{T,7}$, such that the concrete differential cluster size evaluated by the theoretical model is close to but less than $2^{30}$ so that we can practically collect enough sample data.
3. We compute $10^4$ differential paths with randomly generated input and output concrete differences $\alpha_{\Omega_I}^{C,0}$ and $\alpha_{\Omega_O}^{C,7}$. The probability $Prob(\alpha_{\Omega_I}^{C,0} \rightarrow \alpha_{\Omega_O}^{C,7})$ is computed by considering every possible differential path from $\alpha_{\Omega_I}^{C,0}$ to $\alpha_{\Omega_O}^{C,7}$.

Even for 7 rounds, the computational cost is high when trying to find all the paths connecting some specific input and output difference $\alpha_{\Omega_I}^{C,0}$ and $\alpha_{\Omega_O}^{C,7}$. We apply the meet-in-the-middle approach when searching the path probability. First, we split the 7 rounds into two, 3 rounds + 4 rounds. Then starting from $\alpha_{\Omega_I}^{C,0}$, we compute every differential path till the middle point and save them in a hash table along with the corresponding probabilities. Then starting from $\alpha_{\Omega_O}^{C,7}$, we compute backwards for all the differential paths, and match the ones in the hash table. Once we find a match, update the total probability.

As a result, the computational cost is reduced from computing 7 rounds to computing the longer half, which is 4 rounds. The bottleneck is the memory storage, which is bounded by the hamming weight of the truncated difference in the matching round. The experimental results are summarized in Figure 2. From the figure, it shows that the mean of the

probability distribution is evaluated very accurately. The experimental mean is $2^{-31.9984}$ while the theoretical value is $2^{-31.9958}$. From the left figure, the histogram confirms the normal distribution of the probability. For this particular case, the normal approximation becomes rather accurate when the number of input pairs reaches around $2^N \approx 2^{37.4042}$. And this value also satisfies the condition in Theorem 1, which again confirms the accuracy of our model.

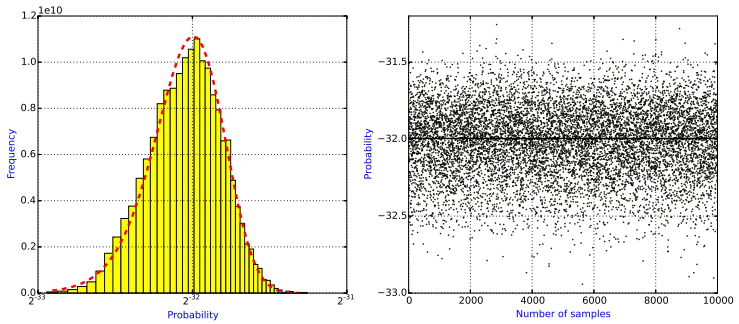


Fig. 2: Experimental result for Toy cipher

## 3 Statistical Distinguisher and some observations for LBlock and TWINE

It it well known that when there are only two distributions to distinguish from, hypothesis testing based on Neyman-Pearson lemma [14] provides us with the most powerful test. [4] first provided a former analysis on how to build an optimal distinguisher between two sources, specifically one from random distribution and one from a real cipher distribution as in our context. They further derived the complexity to distinguish in the form of number of observable outputs or the input queries regarding the block cipher analysis based on the log-likelihood ratio statistics. Several following papers such as [9] and [1] take advantage of this distinguisher framework, and after combining with order statistics techniques addressed in [3], they were able to accurately evaluate the successful probability of a key recovery of the attack. Also, they were able to apply not only the traditional differential attack but also multiple, truncated and impossible differential attacks. The relation between a good statistical distinguisher

and the number of rounds we can attack is pretty much straightforward. What may not seem to be trivial is the complexity of the key recovery, which will rely on the format of the output differential. However, it is known that if we use multiple differential outputs, the distinguisher behaves better and since we are especially interested in the extent to which we can distinguish theoretically for large rounds of GFS, we omit the key recovery discussion in this paper. We rearrange the core theorems from [4] that will be used in our evaluation as follows.

**Theorem 2 ([4]).** *Considering that $Z_1, Z_2, ...$ is a sequence of iid random variables of distribution $\boldsymbol{D}$ and that $D_0$ and $D_1$ share the same support, the log-likelihood ratio statistic follows normal distribution,*

$$Pr[\frac{LLR(Z^n) - n\mu}{\sigma\sqrt{n}} < t] \xrightarrow{n\to\infty} \Phi(t)$$

*where $\mu = \mu_j$ with $\mu_0 = D(D_0||D_1)$, $\mu_1 = -D(D_1||D_0)$ and $\sigma_j^2 = \sum_{z\in\mathcal{Z}} Pr_{D_j}[z](log\frac{Pr_{D_0}[z]}{Pr_{D_1}[z]}) - \mu_j^2$ for $j \in \{0, 1\}$. And*

$$LLR(Z^n) = \sum_{a\in\mathcal{Z}} N(a|Z^n) log\frac{Pr_{D_0}[a]}{Pr_{D_1}[a]}$$

*Denote $v$ to be the number of samples need to distinguish between $D_0$ and $D_1$, then*

$$v = \frac{4 \cdot \Phi^{-1}(P_e)^2}{\sum_{z\in\mathcal{Z}} \frac{(Pr_{D_0}[z] - Pr_{D_1}[z])^2}{Pr_{D_1}[z]}}$$

*where $P_e$ is the error probability, and $D$ denotes the Kullback-Leibler distance*

$$D(D_0||D_1) = \sum_{z\in\mathcal{Z}} Pr_{D_0}[z] log\frac{Pr_{D_0}[z]}{Pr_{D_1}[z]}$$

Here we assume $D_1$ has the uniform distribution, then $Pr_{D_1}[z] = 2^{-n}$ for $\forall z \in \{0, 1\}^n$. From Corollary 1, we know that $Pr_{D_0}[z_i]$ follows different normal distributions. We know that the mean of the distribution is the unbiased point estimator for $Pr_{D_0}[z_i]$. Thus by replacing $Pr_{D_0}[z_i]$ with the corresponding mean derived by Corollary 1, we are able to compute the required number of samples $v$ in order to distinguish.

### 3.1 Efficient algorithm to compute $D_0$

Deriving the full distribution $D_0$ is a practical issue. For GFS with 4-bit nibble and 64-bit block size, the truncated differential domain is shrunk down to $2^{16}$. However, the computational cost will still grow exponentially as the number of rounds grows. Fortunately, we can store all the $2^{16}$ differential states for each of the rounds, which makes the computational cost grow linearly regarding the number of rounds. This will dramatically speed up the computing for $D_0$ regarding large number of rounds.

**Algorithm 1** Searching $D_0$ for all input and output truncated differences

**Input:** Input truncated difference $\alpha^{T,0}$.
**Output:** Full distribution of $D_0$ given $\alpha^{T,0}$.
**procedure** DIST_SEARCH($r \leftarrow 0, \alpha^{T,0}$)
    $M = \{(s_i, p_i) | 0 \leq i \leq 2^{n/d}\} \leftarrow \emptyset$
    Append $(\alpha^{T,0}, 1.0)$ to $M$.
    **while** $r! = N - 1$ **do**
        $M_{out} \leftarrow M$
        **for** $\forall (s_i, p_i) \in M$ **do**
            // Given $s_i, p_i$, round function returns all the possible output diff and probabilities
            $\{(o_0, p_0^{'}), ..., (o_{t-1}, p_{t-1}^{'})\} \leftarrow round(s_i, p_i)$
            **for** $\forall (o_i, p_i^{'})$ **do**
                **if** $o_i \in M_{out}$ **then**
                $p_i \leftarrow p_i + p_i^{'}$
                **else**
                Append $(o_i, p_i^{'})$ to $M_{out}$
        $M \leftarrow M_{out}$
    Output $(s_i, p_i) \in M, 0 \leq i \leq 2^{n/d}$

For GFS with 4-bit sub-blocks and 64-bit block size, after around 7 rounds, $M$ will include every truncated internal state. We apply the GMP library [10] when computing the probability so that we do not lose precision. However, as the number of rounds grow, the bias becomes miniscule, requiring large amounts of memory to store the precision. When we reach some large rounds, we cannot produce accurate result due to the memory limit. The algorithm is still very efficient considering that we need to perform the search for not only one but all the $2^{n/d} - 1$ input difference $\alpha^{T,0}$. The following experimental results show the number of rounds we have achieved with full precision as well as some rounds where precision was lost partially.

### 3.2 Observations on LBlock and TWINE

LBlock is a 32-round, 64-bit block cipher with Feistel structure proposed by Wenling Wu et al. in [22]. In each round after the left 32-bit side goes through a non-linear function F, it is XOR-ed with the right side that has performed an 8-bit left cyclic shift. TWINE is also a 64-bit block cipher with GFS structure proposed by Tomoyasu Suzaki, etc in [21]. Different from LBlock, it supports 80 and 128 bits key length which both have the same 36 rounds. The F function of LBlock and round operation of TWINE are shown in Figure 3 and Figure 4.

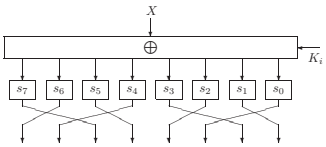


Fig. 3: F function for LBlock

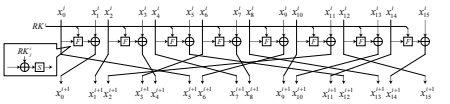

Fig. 4: One round for TWINE

In [21], the authors already identified that both ciphers are very similar to each other regarding the Feistel structure and the permutation layer. This is also our motivation to study these two ciphers, first to compare the security margins and secondly, obtain the observataions for the behavior of GFS.

As we have pointed out, our framework can be used to exploit all the distributions under our theoretical model. In order to get a close look at the strength and weakness of the various differential paths given different input differences, we need to perform Algorithm 1 for all the $2^{16} - 1$ input differences for different number of rounds. Figure 5 and Figure 6 show the experimental results of how many samples are required in order to distinguish the cipher from a uniformly distributed random source. Particularly, for each of the input differences (hamming weight), we consider all the possible output differences to derive the corresponding distinguisher. The experiment was performed on supercomputer Cray XC30 with 700 CPU cores (Intel Xeon E5-2690v3 2.6GHz (Haswell)) running in parallel for around three days.

Both figures share some similarities which provide us with an insight into the properties of other GFS with bijective S-Box design. It also provides us with strategies on how to perform efficient cryptanalysis. Firstly,

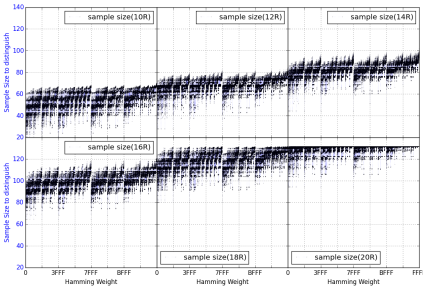


Fig. 5: Distinguisher for LBlock

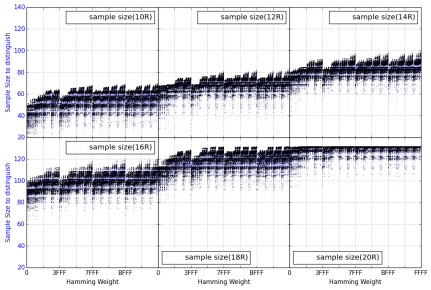


Fig. 6: Distinguisher for TWINE

within the same number of rounds, we notice that the distinguisher will perform better as the hamming weight of the input differences decrements. Considering many previous researchers such as [16] favor the input difference with small hamming weight, this result seems to be straightforward. However, previous results did not consider the clustering effect where many small paths could eventually lead to a better cluster. Here we clarify this situation by showing that input differences with large hamming weight tend to have better randomization property with respect to the differential distribution, thus an attacker should focus on searching the paths with small input hamming weight.

Secondly, this trend remains the same for different number of rounds, with the total number of pairs required to distinguish increasing as the number of rounds grows. This makes sense according to the Markov cipher model [15], which has been used to model modern block ciphers. Notice that for both LBlock and TWINE, starting from round 18, the number of pairs tends to converge to some threshold. This is due to the insufficient precision used in the GMP library. We expect that the original trend will persist no matter the number of rounds if we have enough memory space to store $2^{16}$ elements with large enough precision. In the current setting, we set the precision to be 10000 bits, which gives us a good balance between the precision of the results, and the experiment speed. Notice that even for 20 rounds, the results for the low hamming weight are still accurate and usable.

**Distinguishing Attack.** Now we give distinguishing attacks for LBlock and TWINE assuming the usage of the full code book. We have previously shown that input differences with small hamming weight tends to have better distinguishability. For any truncated input difference $\alpha^{T,0}$, the total

number of differential pairs that conform to the input differential $\alpha^{T,0}$ is $2^{63+4\times HW(\alpha^{T,0})}$, where $HW(\alpha^{T,0})$ denotes the hamming weight of $\alpha^{T,0}$. If the number of pairs $v$ in order to distinguish derived from the statistical framework is smaller than $2^{63+4\times HW(\alpha^{T,0})}$, then we are able to launch the distinguisher attack immediately. However, for larger rounds such as 18 rounds, the experimental result indicates that the input differential with the best distinguishing effect requires more pairs than the total amount that the cipher can provide. Therefore, instead of taking advantage of only one input difference, we can consider multiple input differences. One straightforward way is to store $2^{16}$ counters for each of the input difference, and we extend the distribution domain from $2^{16}$ to maximum $2^{32}$ counters. Let $v_i$ denote the number of pairs required for input difference $\alpha_i^{T,0}$, then the number of pairs $v_{0...i}$ to distinguish can be computed as follows:

$$v_{0...i} = (\sum_{x=0}^{i} \frac{1}{v_x})^{-1}$$

This equation can be derived directly from Theorem 2. Notice that we will proceed with the input difference with small hamming weight first, thus $v$ is sorted in ascending order based on hamming weight in order to provide which input difference to use first. In order to check the success of the attack, we need to be sure that

$$v_{0...i} < \sum_{x=0}^{i} 2^{63+4\times HW(\alpha_i^{T,0})}$$

For our distinguishing attack, the computational cost is the cost of the summing the counters, which requires $\sum_{x=0}^{i} 2^{63+4\times HW(\alpha_i^{T,0})}$ memory accesses. Under the conservative estimation that one memory access is equivalent to one round operation cost, which was also used in paper [12], the computational cost can be estimated as $\frac{1}{R} \times \sum_{x=0}^{i} 2^{63+4\times HW(\alpha_i^{T,0})}$ $R$-round computation, where $R$ is the number of total rounds to attack.

Although for larger number rounds we currently do not have the accurate distribution for all the input differences due to the computational limitations, the input differences with small hamming weight are still accurate. Therefore, we can take advantage of this accurate region to launch the attack. For 21 rounds of LBlock, if we take the first $2^{11}$ input differences sorted according to $v_i$, then $v_{0...2^{11}} \approx 2^{97.69}$ which is less than the total available pairs $2^{100.67}$. This means we can actually perform the distinguishing attack as long as we have enough computing resources. The time complexity here is thus $2^{93.3}$ 21 rounds LBlock encryptions. TWINE

behaves almost exactly the same as LBlock for the first 21 rounds. By applying our framework, we can provide an accurate security bound for different number of rounds. For example, a 21-round LBlock will theoretically fail to achieve the security level that we claim if we set the key size to be larger than 94 bits.

Next we summarize the security margin for both LBlock and TWINE regarding the distinguishing attack. Notice that we choose the distinguishing attack to bound the security since it is usually considered to be weaker than key recovery attack. So from a designer's point of view, we have to set the security parameter (key size) to be conservative in order to resist as many attacks as possible. Due to the limitation of computational resources, we can only derive the accurate values up to 21 rounds for both LBlock and TWINE accordingly. However, after observing the first 21 rounds for both LBlock and TWINE, the increase of the computational cost is log-linear with respect to the number of rounds. Thus the trend can be well extrapolated by using the least square methods. Figure 7 and 8 demonstrate the security level for full rounds of LBlock and TWINE, where the dotted line is the prediction while the solid line is the experimental results. Our analysis shows that if both ciphers use 80-bit key setting, then number of rounds considered to be secure is around 19. However, since TWINE also support 128-bit key, in order to satisfy the corresponding security, we will need at least 25 rounds. We notice that in [2], they can achieve 25-rounds key recovery attack for TWINE-128 by using MitM and impossible differential attack. By using truncated differential technique, however, they can only attack 23-rounds using dedicated techniques. Our result complements theirs by revealing a general pattern after an in-depth analysis of the differential distinguisher. From the differential characteristic's point of view, although Table 3 in [2] demonstrates several paths that are better than evaluated using active S-Box, they still cannot achieve more than 16 rounds for TWINE.

From the provable security's point of view, both full rounds LBlock and TWINE are secure, and our analysis can provide the accurate security margin which is around 178 bits and 208 bits for LBlock and TWINE respectively. The reason TWINE is more secure in this sense is that it has 4 more rounds than LBlock, and they are equivalently secure against differential attack if given the same number of rounds.

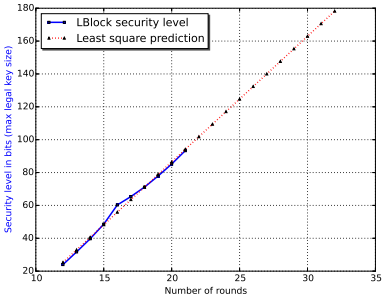


Fig. 7: Security level for LBlock

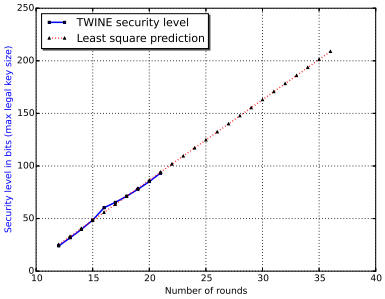


Fig. 8: Security level for TWINE

## 4 Conclusion

In this paper, we revisit the security of GFS with S-Box design regarding differential cryptanalysis. We evaluate the differential trails taking the full cluster into consideration by providing both theoretical and experimental results for the full distribution in truncated form. Our framework provides a solution for ciphers with relatively large block size to derive the full differential distribution. As a concrete application, we evaluate LBlock and TWINE to demonstrate the relationship between the hamming weight of the input difference and complexity of the attack. For TWINE-128, our attack can achieve 25 rounds, which is comparable to the best attacks up to date. More importantly, our framework enables us to compute the accurate security bound on full rounds LBlock and TWINE. As far as we know, this is the first achievement on security proof with exact security margin provided. This framework can be utilized by future cipher proposals to determine the minimum security margin of their designs.

## References

1. Albrecht, M., Leander, G.: An all-in-one approach to differential cryptanalysis for small block ciphers. In: Knudsen, L., Wu, H. (eds.) Selected Areas in Cryptography, Lecture Notes in Computer Science, vol. 7707, pp. 1–15. Springer Berlin Heidelberg (2013)
2. Alex Biryukov, P.D., Perrin, L.: Differential analysis and meet-in-the-middle attack against round-reduced twine. Cryptology ePrint Archive, Report 2015/240 (2015)
3. Aydin Selcuk, A., Bicak, A.: On probability of success in linear and differential cryptanalysis. In: Cimato, S., Persiano, G., Galdi, C. (eds.) Security in Communication Networks, Lecture Notes in Computer Science, vol. 2576, pp. 174–185. Springer Berlin Heidelberg (2003)

4. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In: Lee, P. (ed.) Advances in Cryptology - ASIACRYPT 2004, Lecture Notes in Computer Science, vol. 3329, pp. 432–450. Springer Berlin Heidelberg (2004)
5. Biham, E., Shamir, A.: Differential cryptanalysis of des-like cryptosystems. In: Menezes, A., Vanstone, S. (eds.) Advances in Cryptology-CRYPT0' 90, Lecture Notes in Computer Science, vol. 537, pp. 2–21. Springer Berlin Heidelberg (1991)
6. Biryukov, A., Nikolifa, I.: Automatic search for related-key differential characteristics in byte-oriented block ciphers: Application to aes, camellia, khazad and others. In: Gilbert, H. (ed.) Advances in Cryptology - EUROCRYPT 2010, Lecture Notes in Computer Science, vol. 6110, pp. 322–344. Springer Berlin Heidelberg (2010)
7. Biryukov, A., Roy, A., Velichkov, V.: Differential analysis of block ciphers simon and speck. In: International Workshop on Fast Software Encryption-FSE (2014)
8. Blondeau, C., Gérard, B.: Multiple differential cryptanalysis: Theory and practice. In: Fast Software Encryption. pp. 35–54. Springer (2011)
9. Blondeau, C., Gerard, B., Nyberg, K.: Multiple differential cryptanalysis using llr and statistics. In: Visconti, I., De Prisco, R. (eds.) Security and Cryptography for Networks, Lecture Notes in Computer Science, vol. 7485, pp. 343–360. Springer Berlin Heidelberg (2012)
10. Granlund, T., et al.: The gnu multiple precision arithmetic library. TMG Datakonsult, Boston, MA, USA 2(2) (1996)
11. Knudsen, L.R., Robshaw, M.: The block cipher companion. Springer Science & Business Media (2011)
12. Lu, J., Yap, W.S., Wei, Y.: Weak keys of the full misty1 block cipher for related-key differential cryptanalysis. In: Dawson, E. (ed.) Topics in Cryptology – CT-RSA 2013, Lecture Notes in Computer Science, vol. 7779, pp. 389–404. Springer Berlin Heidelberg (2013)
13. Matsui, M.: On correlation between the order of s-boxes and the strength of des. In: De Santis, A. (ed.) Advances in Cryptology — EUROCRYPT'94, Lecture Notes in Computer Science, vol. 950, pp. 366–375. Springer Berlin Heidelberg (1995)
14. Neyman, J., Pearson, E.S.: On the problem of the most efficient tests of statistical hypotheses. Springer (1992)
15. O'Connor, L., Golić, J.: A unified markov approach to differential and linear cryptanalysis. In: Pieprzyk, J., Safavi-Naini, R. (eds.) Advances in Cryptology — ASIACRYPT'94, Lecture Notes in Computer Science, vol. 917, pp. 385–397. Springer Berlin Heidelberg (1995)
16. Özen, O., Varıcı, K., Tezcan, C., Kocair, Ç.: Lightweight block ciphers revisited: Cryptanalysis of reduced round present and hight. In: Information security and privacy. pp. 90–107. Springer (2009)
17. Shibutani, K.: On the diffusion of generalized feistel structures regarding differential and linear cryptanalysis. In: Biryukov, A., Gong, G., Stinson, D. (eds.) Selected Areas in Cryptography, Lecture Notes in Computer Science, vol. 6544, pp. 211–228. Springer Berlin Heidelberg (2011)
18. Shirai, T., Shibutani, K., Akishita, T., Moriai, S., Iwata, T.: The 128-bit blockcipher clefia (extended abstract). In: Biryukov, A. (ed.) Fast Software Encryption, Lecture Notes in Computer Science, vol. 4593, pp. 181–195. Springer Berlin Heidelberg (2007)
19. Sun, S., Hu, L., Wang, P., Qiao, K., Ma, X., Song, L.: Automatic security evaluation and (related-key) differential characteristic search: Application to simon, present, lblock, des(l) and other bit-oriented block ciphers. In: Sarkar, P., Iwata, T. (eds.) Advances in Cryptology - ASIACRYPT 2014, Lecture Notes in Computer Science, vol. 8873, pp. 158–178. Springer Berlin Heidelberg (2014)

20. Suzaki, T., Minematsu, K.: Improving the generalized feistel. In: Hong, S., Iwata, T. (eds.) Fast Software Encryption, Lecture Notes in Computer Science, vol. 6147, pp. 19–39. Springer Berlin Heidelberg (2010)
21. Suzaki, T., Minematsu, K., Morioka, S., Kobayashi, E.: TWINE: A lightweight block cipher for multiple platforms. In: Knudsen, L., Wu, H. (eds.) Selected Areas in Cryptography, Lecture Notes in Computer Science, vol. 7707, pp. 339–354. Springer Berlin Heidelberg (2013)
22. Wu, W., Zhang, L.: Lblock: A lightweight block cipher. In: Lopez, J., Tsudik, G. (eds.) Applied Cryptography and Network Security. Lecture Notes in Computer Science, vol. 6715, pp. 327–344. Springer Berlin Heidelberg (2011)