Linear Approximations of Addition Modulo 2^n-1

Chunfang Zhou, Xiutao Feng and Chuankun Wu

State Key Laboratory of Information Security, Institute of Software,
Chinese Academy of Sciences, China

FSE 2011

Outline

1. Motivation
2. Preliminaries
   - Linear Approximation and Its Correlation
   - Linear Approximations of Addition Modulo 2^n
3. Addition Modulo $2^n - 1$
   - Addition Modulo $2^n - 1$ with Two Inputs
   - Addition Modulo $2^n - 1$ with More Inputs
4. The Limit of $cor(1; 1^k)$
5. Conclusion

The Basic Problem That We Studied

Given an integer $n \geq 2$, consider the operation

$$y = x_1 + x_2 + \cdots + x_k \ mod\ 2^n - 1$$

where $1 \leq y, x_i \leq 2^n - 1$, $1 \leq i \leq k$.

Question: How can we approximate this function linearly and measure the linear approximation?

Why we study the problem?

- In ZUC
  - LFSR is defined on prime field $GF(2^{31} - 1)$
  - the feedback logic of the LFSR consist of "+" and "$\times$" on prime filed $GF(2^{31} - 1)$
  - the LFSR registers are range from 1 to 2^{31}-1
- In linear analysis, we should approximate the nonlinear part of the cipher by linear function.

Some basic definitions

n: a positive integer.
Z_{2^n}: $\{x|0 \leq x \leq 2^n - 1\}$.
Given an integer $x \in Z_{2^n}$, let

$$x = x^{(n-1)}x^{(n-2)} \dots x^{(0)} = \sum_{i=0}^{n-1} x^{(i)} 2^i$$

be the binary representation of x, where $x^{(i)} \in \{0, 1\}$.
For arbitrary two integers $w, x \in Z_{2^n}$, the inner product of w and x is defined as below

$$w \cdot x = \bigoplus_{i=0}^{n-1} w^{(i)} x^{(i)}.$$

The linear approximation

Definition 1

Let J be a nonempty subset of Z_{2^n}, k be a positive integer and f be a function from J^k to J. Given $k+1$ constants $u, w_1, \cdots, w_k \in Z_{2^n}$, the linear approximation of the function f associated with $u, w_1, \cdots, w_k$ is an approximate relation of the form

$$u \cdot f(x_1, \cdots, x_k) = \bigoplus_{i=1}^{k} w_i \cdot x_i,$$

and the $(k+1)$-tuple $(u, w_1, \cdots, w_k)$ is called a linear mask *of f.*

The correlation

Definition 2

The efficiency of the linear approximation is measured by its correlation, which is defined as below

$$\mathbf{cor}_f(u; w_1, \cdots, w_k) = 2\Pr(u \cdot f(x_1, \cdots, x_k) = \bigoplus_{i=1}^{k} w_i \cdot x_i) - 1,$$

where the probability is taken over uniformly distributed $x_1, \cdots, x_k$ *over* J.

Addition Modulo 2^n

Denote by $\boxplus$ the addition modulo 2^n, that is,

$$u = x_1 \boxplus x_2 = (x_1 + x_2) \mod 2^n.$$

Given the linear mask (u, w_1, w_2) of the addition $\boxplus$, we can derive a sequence $\underline{z} = z_{n-1} \cdots z_0$ as follows

$$z_i = u^{(i)}2^2 + w_1^{(i)}2 + w_2^{(i)}, \quad i = 0, 1, \cdots, n-1.$$

Transition matrix

Define

$$M_n(u, w_1, w_2) = \prod_{i=0}^{n-1} A_{z_i},$$

where A_j $(j = 0, 1, \cdots, 7)$ are constant matrices of size 2×2 and defined as follows

$$A_0 = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, A_1 = A_2 = -A_4 = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix},$$
$$-A_3 = A_5 = A_6 = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, A_7 = \frac{1}{4}\begin{pmatrix} 3 & -1 \\ 1 & -3 \end{pmatrix}.$$

For any given linear mask (u, w_1, w_2), let $M_n(u, w_1, w_2)$ be defined as above. Set $M_n(u, w_1, w_2) = (M_{i,j})_{0 \le i,j \le 1}$. Then we have

$$\begin{aligned} M_{i,j} = {} & \Pr(u \cdot (x_1 \boxplus x_2) = w_1 \cdot x_1 \oplus w_2 \cdot x_2 \wedge c_n = i \wedge c_0 = j) \\ & - \Pr(u \cdot (x_1 \boxplus x_2) \neq w_1 \cdot x_1 \oplus w_2 \cdot x_2 \wedge c_n = i \wedge c_0 = j), \end{aligned}$$

where c_0 is an initial carry bit, and c_n is the n-th carry bit of the addition x_1 and x_2 with the initial carry bit c_0.
By convention $c_0 = 0$, we have

$$\mathbf{cor}_{\boxplus}(u; w_1, w_2) = M_{0,0} + M_{1,0}.$$

Summarized as:

$$(u, w_1, w_2) \longrightarrow \underline{z} \longrightarrow M_n(u, w_1, w_2) \longrightarrow \mathbf{cor}_{\boxplus}(u; w_1, w_2)$$

The difference between addition modulo 2^n and $2^n - 1$

There are several differences between addition modulo 2^n and $2^n - 1$:

- the range of inputs and output
  $[0, 2^n - 1]$ vs. $[1, 2^n - 1]$
- the probability of the input bits equal to 1
  $\frac{1}{2}$ vs. $\frac{2^{n-1}}{2^n-1}$
- the probability of the input bits equal to 0
  $\frac{1}{2}$ vs. $\frac{2^{n-1}-1}{2^n-1}$
- the carry of the most important position
  be discarded vs. be added to the least important position of the result

Denote $x_1 + x_2 \mod 2^n - 1$ by $x_1 \hat{\boxplus} x_2$.

$$x_1 \hat{\boxplus} x_2 = \begin{cases} x_1 + x_2 \mod 2^n & \text{if } 0 < x_1 + x_2 < 2^n \\ x_1 + x_2 + 1 \mod 2^n & \text{if } x_1 + x_2 \geq 2^n \end{cases}$$

It is difficult to calculate the correlation directly, we consider counting the pairs of (x_1, x_2) which satisfy the linear approximation:

$$\begin{aligned}
(u, w_1, w_2) &\rightarrow \underline{z} \rightarrow M_n(u, w_1, w_2) \\
&\rightarrow \begin{cases} M_{0,0} \rightarrow \sharp\{(x_1, x_2) | \text{satisfy the LA}, 0 \leq x_1 + x_2 < 2^n\} \\ M_{1,1} \rightarrow \sharp\{(x_1, x_2) | \text{satisfy the LA}, x_1 + x_2 + 1 \geq 2^n\} \end{cases} \\
&\rightarrow \begin{cases} M_{0,0} \rightarrow \sharp\{(x_1, x_2) | \text{satisfy the LA}, 0 < x_1 + x_2 < 2^n\} \\ M_{1,1} \rightarrow \sharp\{(x_1, x_2) | \text{satisfy the LA}, x_1 + x_2 \geq 2^n\} \end{cases} \\
&\rightarrow \mathbf{cor}_{\hat{\boxplus}}(u; w_1, w_2)
\end{aligned}$$

The formula for the correlation

Due to the similarity and the slight difference between these two operations, we can drive an exact formula for $\mathbf{cor}(u; w_1, w_2)$:

$$\mathbf{cor}(u; w_1, w_2) = \frac{2^{2n}(M_{0,0} + M_{1,1}) + 2^n \cdot c + 1}{(2^n - 1)^2},$$

where

$$c = \begin{cases} -3, & \text{if } u = w_1 = w_2 \text{ and } w_{\mathrm{H}}(w_2) \text{ is even}, \\ 1, & \text{if } u \neq w_1 = w_2 \text{ and } w_{\mathrm{H}}(w_2) \text{ is odd}, \\ 0, & \text{if } u, w_1 \text{ and } w_2 \text{ are pairwise different}, \\ -1, & \text{otherwise}, \end{cases}$$

and $w_{\mathrm{H}}(w_2)$ denotes the hamming weight of w_2 in the binary representation.

The formula for the correlation

The correlation of linear approximation of addition modulo $2^n - 1$ with more inputs can be computed recursively:

$$\begin{aligned}&\mathbf{cor}(u; w_1, \cdots, w_k)\\ &\quad = \frac{2^n-1}{2^n} \sum_{w=0}^{2^n-1} \mathbf{cor}(w; w_1, \cdots, w_{k-1})\mathbf{cor}(u; w, w_k).\end{aligned}$$

$\mathbf{cor}(1; 1^k)$

In this section, we will discuss the limit of $\mathbf{cor}(u; \underbrace{u, \cdots, u}_{k})$ for some integer $k \geq 2$ and $w_H(u) = 1$ when n goes to infinity. By the property:

$$\mathbf{cor}(u; w_1, \cdots, w_k) = \mathbf{cor}(u \lll l; w_1 \lll l, \cdots, w_k \lll l).$$

So it is enough to study $\mathbf{cor}(1; \underbrace{1, \cdots, 1}_{k})$. For simplicity, we denote it by $\mathbf{cor}(1; 1^k)$.

By the above recursive formula of correlation of linear approximation of addition modulo $2^n - 1$ with k inputs, $\mathbf{cor}(1; 1^k)$ can be split into summations of the product of correlations of addition modulo $2^n - 1$ with two inputs.

$$\mathbf{cor}(1; 1^k) = \sum_{u_1 \in J} \sum_{u_2 \in J} \cdots \sum_{u_{k-2} \in J} \prod_{j=1}^{k-1} \mathbf{cor}(u_{j-1}; u_j, 1),$$

where $J = \{1, 2, \cdots, 2^n - 1\}$, $u_0 = u_{k-1} = 1$.

More Properties of Transfer Matrix

For linear mask $(u, 1, w)$, write $M_n(u, 1, w)$ as M simply. It is easy to see that $z_0 \in \{1, 3, 5, 7\}$ and $z_i \in \{0, 2, 4, 6\}$, $1 \leq i \leq n - 1$.

There are some facts on A_i, $0 \leq i \leq 7$.

1. $A_0 A_i = \frac{1}{2} A_i$, for $\forall\ i \in \{1, 2, 3, 4, 5, 6\}$;
2. $A_i A_0 = A_i$ if $i \in \{1, 2, 4\}$ and $A_i A_0 = \frac{1}{2} A_i$ if $i \in \{3, 5, 6\}$;
3. $A_i A_j = 0$, $i \in \{1, 2, 4\}$ *and* $j \in \{1, 2, 3, 4, 5, 6\}$;
4. $A_1 A_7 = A_2 A_7 = -A_4 A_7 = A_6$.

The necessary and sufficient condition of $\mathbf{Tr}(M) \neq 0$

By these properties, we can derive the necessary and sufficient condition of $\mathbf{Tr}(M) \neq 0$.
Finally we give an upper bound of $|cor(u; 1, w)|$. For any given integer $x \in Z_{2^n}$, define

$$J_x = \{x \oplus 2^i | 1 \leq i \leq LNB(x \oplus 1)\}.$$

$LNB(x)$ denotes the least position where 1 appears in the binary representation of x if $x \neq 0$, and $LNB(0) = n - 1$.
For any integers $u, w \in Z_{2^n}$, if $w \notin J_u$, then

$$|cor(u; 1, w)| < \frac{3}{2^n - 1}.$$

By stripping the correlations equal to zero or trend to zero when n goes to infinity, we get the following lemma:

For any integer $k \geq 3$, if $\lim_{n\to\infty} \mathbf{cor}(1; 1^k)$ exists, then

$$\lim_{n\to\infty} \mathbf{cor}(1; 1^k) = \lim_{n\to\infty} \sum_{u_1 \in J_1} \sum_{u_2 \in J_{u_1}} \cdots \sum_{u_{k-2} \in J_{u_{k-3}}} \prod_{j=1}^{k-1} \mathbf{cor}(u_{j-1}; u_j, 1),$$

where $u_0 = u_{k-1} = 1$.

The correlation can be divided into two parts and the second part can be limited by a const:

$$\mathbf{cor}(u; w_1, w_2) = \mathbf{Tr}(M_n(u, w_1, w_2)) + \frac{\delta(u, w_1, w_2)}{2^n - 1},$$

we can further strip $\frac{\delta(u_{j-1}, u_j, 1)}{2^n-1}$ from $\mathbf{cor}(u_{j-1}; u_j, 1)$, $j = 2, 3, \cdots, k-1$. Then finally we can get the following conclusion.

For any integer $k \geq 3$, if $\lim\limits_{n\to\infty} \mathbf{cor}(1; 1^k)$ exists, then

$$\lim_{n\to\infty} \mathbf{cor}(1; 1^k) =$$

$$\lim_{n\to\infty} \sum_{u_1 \in J_1} \sum_{u_2 \in J_{u_1}} \cdots \sum_{u_{k-1} \in J_{u_{k-2}}} \prod_{j=1}^{k-1} \mathbf{Tr}(M_n(u_{j-1}, u_j, 1)),$$

where $u_0 = u_{k-1} = 1$.

The general case of the limit

- the case of k is an even integer
  For any even positive integer k, the set of $\{u_0, u_1, \cdots, u_{k-1}\}$ satisfy the conditions of summation is an empty set, so we have $\lim_{n\to\infty} \mathbf{cor}(1; 1^k) = 0$.
- the case of k is an odd integer
  For any odd positive integer k, the set of $\{u_0, u_1, \cdots, u_{k-1}\}$ satisfy the conditions of summation is not an empty set, we could prove that if $\lim_{n\to\infty} \mathbf{cor}(1; 1^k)$ exists, then

$$|\lim_{n\to\infty} \mathbf{cor}(1; 1^k)| \geq \frac{1}{3} 2^{-(k-3)}$$

We discuss properties of linear approximations of addition modulo $2^n - 1$.

- For the case when two inputs are involved, an exact formula is given.
- For the case when more than two inputs are involved, an iterative formula is given.
- For the special linear approximation with all masks being equal to 1, we discuss the limit of their correlations when n goes to infinity.

Thanks for your attention!