Linear Approximating to Integer Addition

Li An-Ping

Beijing 100085, P.R. China
apli0001@sina.com

Abstract

The integer addition is often applied in ciphers as a cryptographic means. In this paper we will present some results about the linear approximating for the integer addition.

Keywords : Linear approximating, bias, integer addition.

1. Preliminary

For the undecided effect of carry operations in the integer addition, it is often used as a cryptographic means in some ciphers, for instance, in the candidate ciphers of eSTREAM (The ECRYPT Stream Cipher Project) some of them employed the combination of the integer addition, XOR and rotations as main cryptographic transformation. Therefore, it is significant to know the effect of the integer addition in cryptography. J. Wallen [3] provided an algorithm for computing the correlation of linear approximation of addition modulo 2^n. In this paper, we will show some explicit results about the linear approximating to the integer addition.

2. Some basic results

In this paper, the symbol $\oplus$ as usually stands for XOR operation. Suppose z is a binary segment of length n, denoted by $z[i]$ the *i-th* bit, and let $s_1(z) = \sum_i z[i]$, $s_0(z) = n - s_1(z)$, and $d(z) = s_0(z) - s_1(z)$, that is, $s_0(z)$ and $s_1(z)$ are the numbers of the bit "0" and bit "1" in z respectively, and $d(z)$ is the bias of them. Let x and y be two integers of length n bits, denoted by $L(x, y) = (x + y) \oplus (x \oplus y)$, $\tilde{L}(x, y) = (1 + x + y) \oplus (x \oplus y)$, and define

$$\begin{aligned} D_i &= (2^{2n} - 2\sum_{x,y} L(x, y)[i]) / 2^{2n}, \\ \tilde{D}_i &= (2^{2n} - 2\sum_{x,y} \tilde{L}(x, y)[i]) / 2^{2n}, \\ D &= \sum_{x,y} d(L(x, y)) / n \cdot 2^{2n}. \end{aligned} \tag{2.1}$$

We have the following result

Proposition 1

$$D_i = 1/2^i, \quad 0 \le i < n, \tag{2.2}$$

$$\tilde{D}_i = -1/2^i, \quad 0 \le i < n, \tag{2.3}$$

$$D = \frac{2}{n} \cdot \left(1 - \frac{1}{2^n}\right). \tag{2.4}$$

Proof. It is easy to check directly $D_0 = 1$, so we assume that $i > 0$. For an integer z, denoted by z_i the integer formed by the segment of z from bit 0 to bit i. Let $w = L(x, y)$, denoted by N_i the number of $L(x, y)$ with $w[i] = 1$. It is easy to know that $w[i] = 0$ if and

only if $x_{i-1} + y_{i-1} < 2^i$, so

$$N_i = (\sum_{1 \le k < 2^i} k) \cdot 2^{(n-i)} = (2^{2n} - 2^{2n-i})/2. \tag{2.5}$$

Hence,

$$D_i = (2^{2n} - 2 \cdot N_i)/2^{2n} = 1/2^i.$$

The proof of (1.3) is similar to the above, so omitted. It is easy to know that

$$D = (\sum_{0 \le i < n} D_i)/n, \tag{2.6}$$

so,

$$D = (\sum_{0 \le i < n} 2^{-i})/n = \frac{2}{n} \cdot \left(1 - \frac{1}{2^n}\right).$$

□

From Proposition 1, we have seen that $x + y$ is still some like $x \oplus y$ on the bits in statistics, especially for the first bits, though there are undecided carry operations in the addition. In other words, the probability $(x+y)[i] = (x \oplus y)[i]$ has notable advantage when i is small, e.g. $i = 0,1,2,\cdots,etc$. In the following, we will show a more general result on the linear approximating to the integer addition.

Suppose that z is a integer variable over the domain Ω, denoted by $\delta(z) = \bigoplus_i z[i]$ and define

$$\Delta_z = (|\Omega| - 2\sum_{z \in \Omega} \delta(z))/|\Omega|. \tag{2.7}$$

Moreover, for a constant integer c, denoted by $C = \{\, i \mid c[i] = 1 \,\}$. Suppose the $C = \{\, i_k \,\}_1^s$, $i_1 > i_2 > \cdots > i_s$, as usual, $|C|$ represents the cardinality of set C, i.e. $|C| = s$, and define $\|C\| = \sum_{1 \le i \le s} (-1)^{k-1} i_k$.

Proposition 2 Suppose that c is a constant integer, denoted by $L_c = L(x,y)\,\&\,c$ and $\tilde{L}_c = \tilde{L}(x,y)\,\&\,c$, we have

$$\begin{aligned} \Delta_{L_c} &= 1/2^{\|C\|}, \\ \Delta_{\tilde{L}_c} &= (-1)^{|C|}/2^{\|C\|}, \end{aligned} \tag{2.8}$$

Proof. We prove the formula (2.8) by the induction on $|C|$. By the Proposition 1, we have

known the formula (2.8) is true when $s=1$. Now assume it is true in the case s. We consider the first i_{s+1} bits, denoted by N_0 and N_1 the numbers of the pairs (x,y) of i_{s+1}-bits integers such that $x+y<2^{i_{s+1}}$ and $x+y\geq 2^{i_{s+1}}$, we have known that which are equal the numbers of the pairs (x,y) with that $L(x,y)[i_{s+1}]=0$ and 1 respectively and

$$N_0=(2^{i_{s+1}}+1)2^{i_{s+1}-1},\quad N_1=(2^{i_{s+1}}-1)2^{i_{s+1}-1}. \tag{2.9}$$

Denote $j_k=i_k-i_{s+1}, 1\leq k\leq s$, and $C'=\{\ j_k\ \}_1^s$, then we apply the induction on the set C', that is, on the segments of integers beginning the i_{s+1}-th bit, we have

$$\Delta_{L_{c'}}=1/2^{\|C'\|},$$
$$\Delta_{\tilde{L}_{c'}}=(-1)^s/2^{\|C'\|},$$

So,

$$\begin{aligned}
\Delta_{L_c}&=2^{-2\cdot i_{s+1}}\cdot\{N_0\cdot(1+\Delta_{L_{c'}}/2)+N_1\cdot(1+(-1)^{s-1}\Delta_{L_{c'}}/2)\}\\
&\quad-2^{-2\cdot i_{s+1}}\cdot\{N_0\cdot(1-\Delta_{L_{c'}}/2)+N_1\cdot(1-(-1)^{s-1}\Delta_{L_{c'}}/2)\}\\
&=2^{-2\cdot i_{s+1}}\cdot(N_0+(-1)^{s-1}N_1)\cdot\Delta_{L_{c'}}\\
&=((1+(-1)^{s-1})/2+2^{-i_{s+1}}\cdot(1+(-1)^s)/2)\cdot\Delta_{L_{c'}}\\
&=2^{-(1+(-1)^s)i_{s+1}/2}\cdot\Delta_{L_{c'}}\\
&=2^{-(\sum\limits_{1\leq k\leq s}(-1)^k j_k+(1+(-1)^s)i_{s+1}/2)}\\
&=2^{-\sum\limits_{1\leq k\leq s}(-1)^{k-1}i_k+(1-(-1)^s)i_{s+1}/2-(1+(-1)^s)i_{s+1}/2}\\
&=2^{-\sum\limits_{1\leq k\leq s}(-1)^{k-1}i_k+(1-(-1)^s)i_{s+1}/2-(1+(-1)^s)i_{s+1}/2}\\
&=2^{-\sum\limits_{1\leq k\leq s+1}(-1)^{k-1}i_k}.
\end{aligned}$$

The proof for the second formula of (2.8) is similar to the above, so omitted. □

In the some real cases, it is possible to come into the partial cases, that is, in $L(x,y)=(x+y)\oplus(x\oplus y)$ the variable $y=a$ is a constant. Let c be a constant, denoted by $L_c=L(x,a)\&c$, we define

$$\phi_1(a,c)=\sum_{0\leq x<2^n}\delta(L_c),\quad \phi_0(a,c)=2^n-\phi_1(a,c). \tag{2.10}$$

Moreover, let $\Phi(a,c)=(\phi_0(a,c),\phi_1(a,c))$. We will simply write them as Φ, ϕ_1 and ϕ_0

if the parameters are clear from the context, In the following we will mainly calculate the $\Phi(a,c)$.

Suppose that $A=\{i \mid a[i]=1\}$ and $C=\{i \mid c[i]=1\}$, that is, A and C are the sets of the positions of bits "1" of the constants a and c respectively. Without loss the generality, we assume that

$$A \cup C = A_1 C_1 A_2 C_2 \cdots A_s C_s, \tag{2.11}$$

which is the arrangement of A and C in the order from small to large, where $A=\bigcup_{1\le i\le s} A_i, C=\bigcup_{1\le i\le s} C_i$. In this paper we restrict the case $A\cap C=\varnothing$. Denoted by a_k the segment of the integer a formed by A_1 to A_k. Let α_i be the smallest element of A_i but $\alpha_1=0,\ \alpha_{s+1}=n$, and $n_i=\alpha_{i+1}-\alpha_i,\ 1\le i\le s$, $\chi(A_i)=2^{-\alpha_i}\sum_{t\in A_i} 2^t$. Suppose $C_i=\{x_i\}_0^k$, $x_0<x_1<\cdots<x_k$, we define

$$\partial(C_i)=\sum_{0\le j\le k}(-1)^j 2^{-x_j},\ \ \tau(C_i)=2^{\alpha_{i+1}}\cdot\partial(C_i). \tag{2.12}$$

Lemma 1 If $s=1$, then

$$\phi_1(a,c)=a\cdot\tau(C). \tag{2.13}$$

Proof. It is clear that the number $\phi_1(a,c)$ is the number of $L(x,a)$ with odd carries in the positions of constant c. Suppose that $C=\{x_i\}_0^k,\ \ x_0<x_1<\cdots<x_k$, denoted by N_r the number of $L(x,a)$ with r carries in the positions of constant c. It is not difficult to know that the r carries must appear in the first r positions of C, i.e. $x_0<x_1<\cdots<x_{r-1}$, hence it is easy to have

$$\begin{aligned} N_r &= a\cdot(2^{(x_r-x_{r-1})}-1)\cdot 2^{n-x_r} \\ &= a\cdot(2^{n-x_{r-1}}-2^{n-x_r}). \end{aligned} \tag{2 .14}$$

Therefore,

$$\phi_1(a,c)=\sum_i N_{1+2i}=a\cdot\sum_{0\le 2i\le k}(2^{n-x_{2i}}-2^{n-x_{2i+1}})=a\cdot\tau(C). \tag{2.15}$$

□

For two vectors $v_1=(x,y), v_2=(u,w)$, we define

$$v_1 * v_2 = (xu + yw, xw + yu). \tag{2.16}$$

and for 2-vector (x, y), we define $T(x, y) = (y, x)$. Let $p(a,c) = \sum_{2^n - a \le x < 2^n} \delta(L_c)$, $q(a,c) = a - p(a,c)$, and $\Gamma(a,c) = (q(a,c), p(a,c))$. Denoted by $I = 2^n - 1$, and for the integers $k, 0 \le k < s$, let $I_k = 2^{\alpha_{k+1}} - 1$, $a_k = I_k \,\&\, a$, $\overline{a}_k = 2^{\alpha_{k+1}} - a_k$, and $c_k = I_k \,\&\, c$. Moreover, denoted by $\lambda = \sum_{1 \le i \le s} |C_i|$.

Lemma 2

$$\Gamma(a,c) = T^{\lambda}\left(\sum_{1 \le k \le s} \chi(A_{s-k+1}) \cdot \Phi(\overline{a}_{s-k}, c_{s-k})\right), \tag{2.17}$$

where it is assumed that $\Phi(\overline{a}_0, c_0) = (1, 0)$.

Proof. Let z be the integer of length n such that $z \equiv x + a \bmod(2^n)$, and $\overline{a} = 2^n - a$, hence it has $x = z + \overline{a} \bmod(2^n)$, so

$$L(a,x) = z \oplus (\overline{a} + z) \oplus a = z \oplus (\overline{a} + z) \oplus \overline{a} \oplus (a \oplus \overline{a}) = L(z, \overline{a}) \oplus (a \oplus \overline{a}),$$

Denoted by $I' = a \oplus \overline{a}$, suppose that $\alpha = \min\{t \mid t \in A\}$, then it is easy to know $I' = 2^n - 2^{\alpha+1}$. Thus it has,

$$L(a,x) \,\&\, c = (L(z, \overline{a}) \oplus I') \,\&\, c = (L(z, \overline{a}) \,\&\, c) \oplus c.$$

Hence,

$$p(a,c) = \sum_{0 \le z < a} \delta(L(z, \overline{a}) \,\&\, c) \oplus \delta(c). \tag{2.18}$$

Denoted by $\overline{p}[s] = \sum_{0 \le z < a} \delta(L(z, \overline{a}) \,\&\, c)$, $\overline{q}[s] = a - \overline{p}[s]$ and $\overline{\Gamma}[s] = (\overline{q}[s], \overline{p}[s])$, from (2.18) we know that in order to prove (2.17) it suffice to prove

$$\overline{\Gamma}(a,c) = \sum_{1 \le k < s} \chi(A_{s-k+1}) \cdot \Phi(\overline{a}_{s-k}, c_{s-k}). \tag{2.19}$$

It is easy to know that the bits of $L(z, \overline{a})$ excel over A_s will be "0" for the bits of $\overline{a}$ and $\overline{a} + z$ after A_s all are "1", so it follows that

$$L(z, \overline{a}) \,\&\, c = L(z, \overline{a}) \,\&\, c_{s-1}$$

We divide the a integers in the interval $[0, a)$ into $\chi(A_s) + 1$ classes $S_i, 0 \le i \le \chi(A_s)$, an

integer z belongs to S_i iff $z >> \alpha_s = \chi(A_s) - i$. It is clear that for $i = 1, 2, \cdots, \chi(A_s), |S_i| = 2^{\alpha_s}$, and $|S_0| = a_{s-1}$. Hence we have

$$\sum_{z \in S_i} \delta(L(z, \bar{a}) \& c) = \sum_{z \in S_i} \delta(L(z, \bar{a}) \& c_{s-1}) = \phi_1(\bar{a}_{s-1}, c_{s-1}), \quad for\ i > 0,$$
$$\sum_{z \in S_0} \delta(L(z, \bar{a}) \& c) = \sum_{0 \le z < a_{s-1}} \delta(L(z, \bar{a}_{s-1}) \& c_{s-1}) = \bar{p}[s-1]$$

.

The equation above can be written as

$$\bar{\Gamma}[s] = \chi(A_s)\Phi(\bar{a}_{s-1}, c_{s-1}) + \bar{\Gamma}[s-1] \tag{2.20}$$

So, the equation (2.19) will be followed by the induction. □

For each integer k , $1 \le k \le s$, denoted by $d_k = \chi(A_k) \cdot \tau(C_k)$, $\zeta_k = (2^{n_k} - d_k, d_k)$, Moreover, let $p(a_k)$ and $q(a_k)$ be defined as in the Lemma 2 and denoted by $e_k = \tau(C_k) \cdot (q(a_{k-1}) - p(a_{k-1}))$, $\sigma_k = (-e_k, e_k)$, and assuming that $\sigma_0 = (1, 0), \sigma_1 = (0, 0)$, and $\zeta_{s+1} = (1, 0)$, then we have

Proposition 3 If $A \cap C = \varnothing$,

$$\Phi(a, c) = \sum_{0 \le k \le s} (\sigma_k \prod_{k < i \le s+1} \zeta_i), \tag{2.21}$$

Proof. We will apply the induction on s . From Lemma 1, we can know that the Proposition 3 is true for $s = 1.$ Consider the integers of length α_s , by the induction we have

$$(\phi_0[s-1], \phi_1[s-1]) = \sum_{0 \le k \le s-1} (\sigma_k \prod_{k < i \le s} \zeta_i).$$

In the case s , it is clear that there are a_{s-1} integers that will carry in the position α_s , which are $2^{\alpha_s} - a_{s-1}, 2^{\alpha_s} - a_{s-1} + 1, \cdots, 2^{\alpha_s} - 1$. Suppose that in these a_{s-1} integers there are $p[s-1]$ ones in $\phi_1[s-1]$ and $q[s-1]$ ones in $\phi_0[s-1]$ respectively. Consider the last segments of integers from the α_s-bit to the end, let a_k and c_k be defined as the above and $\tilde{a}_k = a - a_k$, $\tilde{c}_k = c - c_k$, and denoted by N_0 and N_1 the numbers of the segments $L(x, \tilde{a}_{s-1}) \& \tilde{c}_{s-1}$ with even and odd bit ”1” respectively, by Lemma 1, we know

$$N_1 = \chi(A_s)\cdot\tau(C_s) = d_s, \quad N_0 = 2^{n_s} - d_s .$$

Let $N_1' = (\chi(A_s)+1)\cdot\tau(C_s), \quad N_0' = 2^{n_s} - N_1'$, it has,

$$\begin{aligned}\phi_1[s] &= (\phi_0[s-1]-q[s-1])\cdot N_1 + (\phi_1[s-1]-p[s-1])\cdot N_0 + N_1' q[s-1] + N_0' p[s-1] \\ &= \phi_0[s-1]\cdot N_1 + \phi_1[s-1]\cdot N_0 + (q[s-1]-p[s-1])\tau(C_s) \\ &= \mathrm{Im}(\sum_{0\le k\le s}(\sigma_k \prod_{k<i\le s+1}\zeta_i)).\end{aligned}$$

Where $\mathrm{Im}(x,y) = y.$ □

As a example, the case $s = 2$,

$$\phi_1(a,c) = d_1\cdot 2^{n_2} + d_2\cdot 2^{n_1} - 2d_1\cdot d_2 + (-1)^{|C_1|} a_1\cdot\tau(C_2) .$$

The case $A\cap C \neq \varnothing$ may be treated in a similar way but will require a little more consideration to set $A_i \cap C_i = B_i$ and as the first step to calculate in the case of a block ABC, the detail discussion is omitted.

3. Conclusion

We hope that these results presented here will be useful in the designs and cryptanalysis of ciphers in the future, the results of Proposition 1 were once appeared in the paper [1] and [2].

Reference

[1] Li An-Ping, Linear approximating for the cipher Salsa20, available at http://www.ecrypt.eu.org/stream/papersdir/056.pdf

[2] Li An-Ping, Linear approximating for the cipher Salsa20 (II), available at http://www.ecrypt.eu.org/stream/papersdir/067.pdf

[3] J. Wallen, Linear Approximations of Addition Modulo 2^n, *Fast Software Encryption FSE'2003*, LNCS v. 288, pp. 261 – 273, Springer-Verlag, 2003.