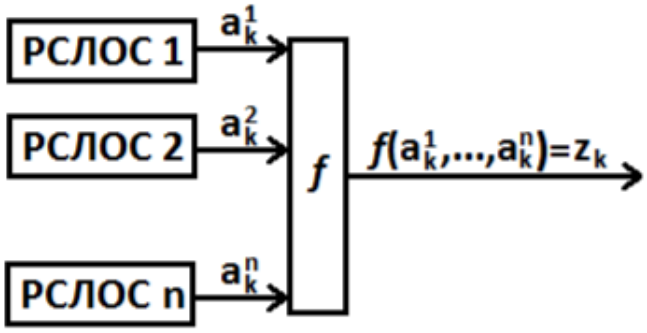
Корреляционная атака Чжэня и Фэня

1. Введение

Какими начальными знаниями мы обладаем для проведения атаки:

Нам известно число ЛРС, которые используются в комбинирующем генераторе, их полиномы обратной связи и комбинирующая булева функция. Найти надо начальные заполнения регистров сдвига.

Если у нас есть n регистров длины L бит каждый:



В таком случае, нам нужно перебрать $\prod_{i=1}^{n} 2^{L_i}$ начальных комбинаций бит, если хотим восстановить начальные заполнения полным перебором.

Пусть f - это комбинирующая функция от n переменных и величина

$$p_j = P(f(v_1, ..., v_n) = v_j) = 0.5 + \epsilon(\epsilon > 0) \tag{1}$$

- вероятность того, что цифра в гамме совпадет с цифрой выходной последовательности j-го ЛРС.

v_1	v_2	v_3	f
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Как уже упоминалось, для заданной функции f и каждого номера j можно вычислить p_j:

Для неё $p_1 = \frac{7}{8}$, $p_2 = \frac{5}{8}$, $p_3 = \frac{5}{8}$.

Мы будем рассматривать один регистр сдвига, поэтому зафиксируем номер j и для краткости будем писать не p_j, а p.

Каждый бит гаммы можно выразить через линейное соотношение битов начального заполнения ЛРС. Сделать это можно через порождающую матрицу (построенную на основе рекуррентного соотношения):

$$\begin{pmatrix} a_k \\ a_{k+1} \\ a_{k+2} \\ \vdots \\ a_{k+L-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ c_0 & c_1 & \dots & \dots & c_{L-1} \end{pmatrix} \begin{pmatrix} a_{k-1} \\ a_k \\ a_{k+1} \\ \vdots \\ a_{k+L-2} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ c_0 & c_1 & \dots & \dots & c_{L-1} \end{pmatrix}^k \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{L-1} \end{pmatrix}$$

Из столбцов данной матрицы можно построить одну большую матрицу (беря нужные столбцы из матриц нужной степени):

$$G = \begin{pmatrix} g_0^0 & g_0^1 & \dots & g_0^{N-1} \\ g_1^0 & g_1^1 & \dots & g_1^{N-1} \\ \vdots & \vdots & \dots & \vdots \\ g_{L-1}^0 & g_{L-1}^1 & \dots & g_{L-1}^{N-1} \end{pmatrix} \tag{2}$$

$$(a_0, ..., a_{L-1}, ..., a_{N-1}) = (a_0, ..., a_{L-1}) \cdot G \tag{3}$$

$$a_i = (a_0, ..., a_{L-1}) \cdot g^{(i)},\ g^{(i)} = (g_0^i, g_1^i, ..., g_{L-1}^i)^T \tag{4}$$

2. Первый шаг атаки

2.1. Уравнения проверки четности

Нас интересует возможность восстановить заполнение регистра по частям. Для этого нам нужно найти линейные соотношения, которые будут содержать только биты из первой части регистра. Для получения таких соотношений мы скомбинируем линейные соотношения для битов гаммы:

$$z_{i_1} \oplus ... \oplus z_{i_{t^{(1)}}} \sim g^{(i_1)} \oplus ... \oplus g^{(i_{t^{(1)}})} = (x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)}, \underbrace{0, ..., 0}_{L-k^{(1)}})^T, \tag{5}$$

где $(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)})$ - некий вектор, полученный при суммировании, разный для разных наборов $(g^{(i_1)}, ..., g^{(i_{t^{(1)}})})$.

Принцип формирования уравнений мы обсудим чуть позже.

Введем некоторые обозначения: всего таких соотношений получится $\Omega^{(1)}$ штук, а среднее количество бит гаммы z_i, участвующих в уравнении, равно $t^{(1)}$ (в оригинальной статье данные обозначения вводятся немного иначе, но для связанности повествования нам стоит ввести их так).

Теперь, когда мы набрали линейных соотношений для интересующей нас части регистра, мы должны как-то выяснить, какое начальное состояние - верное.

Любое из линейных соотношений можно переписать вот так:

$$a_{i_1} \oplus ... \oplus a_{i_{t^{(1)}}} = \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i \tag{6}$$

Перепишем (6) как

$$z_{i_1} \oplus ... \oplus z_{i_{t^{(1)}}} = \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i \oplus \sum_{j=1}^{t^{(1)}} e_{i_j} \tag{7}$$

где z_j - биты гаммы, а $e_{i_j} = a_{i_j} \oplus z_{i_j}$, $(j = i_1, ..., i_{t^{(1)}})$ - случайный шум со следующим распределением: $P(e_j = 0) = 0.5 + \epsilon$ и $P(e_j = 1) = 0.5 - \epsilon$.

Предположим, что предполагаемое начальное состояние - это вектор $(a_0^{'}, ..., a_{k^{(1)}-1}^{'})$. Перепишем равенство (7) следующим образом:

$$z_{i_1} \oplus ... \oplus z_{i_{t^{(1)}}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i^{'} = \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} (a_i \oplus a_i^{'}) \oplus \sum_{j=1}^{t^{(1)}} e_{i_j} \tag{8}$$

Введем обозначения

$$\Delta(i_1, ..., i_{t^{(1)}}) = \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} (a_i \oplus a_i^{'}) \oplus \sum_{j=1}^{t^{(1)}} e_{i_j} \tag{9}$$

Если начальное состояние угадано правильно, то

$$\Delta(i_1, ..., i_{t^{(1)}}) = \sum_{j=1}^{t^{(1)}} e_{i_j} \tag{10}$$

Если каждую из ошибок e_{i_j} мы принимаем как случайную величину, то вероятность суммы этих случайных величин мы можем рассчитать, используя лемму Мацуи (лемму о набегании знаков):

$$q^{(1)} = P(\sum_{j=1}^{t^{(1)}} e_{i_j} = 0) = 0.5 + 2^{t^{(1)}-1} \epsilon^{t^{(1)}} \tag{11}$$

Если начальное состояние угадано правильно, то величина $\sum_{i=1}^{\Omega^{(1)}} (\Delta(i_1, ..., i_{t_1}) \oplus 1)$ (мы делаем приписку $\oplus 1$, чтобы инвертировать каждое событие e_{i_j}, потому что

для нас благоприятным является ноль, а не единица) имеет биномиальное распределение с параметрами $(\Omega^{(1)}, q^{(1)})$, а если неправильно, то с параметрами $(\Omega^{(1)}, \frac{1}{2})$ (потому что подмешивается случайная величина $\sum_{i=0}^{k^{(1)}-1} x_i^{(1)}(a_i \oplus a_i^{'})$, которая и двигает вероятность к $\frac{1}{2}$).

Теперь, когда у нас есть величина $q^{(1)}$, поговорим о формировании уравнений проверки четности. Нас интересуют такие уравнения проверки четности, которые имеют наибольшее значение $q^{(1)}$, потому что они с большей вероятностью верны. Из этого следует, что мы должны формировать уравнения из минимально возможного количества бит гаммы, чтобы $t^{(1)}$ было минимально.

Мы поступим следующим образом: мы имеем для каждого бита гаммы линейную комбинацию начальных бит:

$$(x_0^{(1)}, x_1^{(1)}, ..., x_{L-1}^{(1)})$$

а нам нужно получить уравнение (5). Для этого мы итеративно будем обнулять группы слагаемых, начиная с правого конца. Допустим, что нам нужно найти все уравнения проверки четности, которые сформированы не более чем 4-мя битами гаммы. Для этого будем рассматривать соотношение как:

$$(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)}, \underbrace{x^{(1)}, ..., x^{(1)}}_{\text{2-ая группа}}, \underbrace{x^{(1)}, ..., x^{(1)}}_{\text{1-ая группа}})$$

Теперь мы выделим из всех линейных соотношений (пока что для одного бита гаммы) те, у которых 1-ая группа занулена. К ним же прибавим комбинации из двух бит гаммы, которые образуют линейное соотношение с зануленой 1-ой группой бит:

$$(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)}, \underbrace{x^{(1)}, ..., x^{(1)}}_{\text{2-ая группа}}, \underbrace{0, ..., 0}_{\text{1-ая группа}})$$

Теперь из полученных соотношений мы выделяем те, для которых занулена и вторая группа, а также формируем соотношения из 4-х бит, для которых 2-ая группа также занулена:

$$(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)}, \underbrace{0, ..., 0}_{\text{2-ая группа}}, \underbrace{0, ..., 0}_{\text{1-ая группа}})$$

В результате мы получили необходимые линейные соотношения из одного, двух, трех (если во второй группе встретится пара линейных комбинаций, включающие 1 и 2 различных бита выходной гаммы) и четырех бит выходной гаммы.

Вероятно, можно найти более оптимальный алгоритм поиска линейных соотношений в источнике [11].

Нам потребуются те варианты линейных соотношений, у которых разная часть $(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)})$. Если встретятся несколько одинаковых линейных соотношений с разными результатами $z_{i_1} \oplus ... \oplus z_{i_{t_1}}$, то выбираем итоговый результат по принципу

большинства. То есть разных линейных соотношений у нас получится не более чем $2^{k^{(1)}}$ штук.

2.2. Критерий правильности начального состояния

Перед тем как продолжить, нам нужно разобраться, в чем суть всех дальнейших магических пасов руками. На данный момент мы имеем линейные соотношения, которые показывают с некоторой вероятностью $q^{(1)} > 0.5$, чему будет равно некоторое линейное соотношение битов начального заполнения. Мы хотим проверить все возможные начальные заполнения, подставив их в найденные линейные соотношения и проверив, сходятся ли результаты. Если две одинаковых величины проксорить, то получим ноль, поэтому для соотношения с линейными коэффициентами $(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)})$ равенство $z_{i_1} \oplus ... \oplus z_{i_{t^{(1)}}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i = 0$ должно выполниться. если рассматривать $z_{i_1} \oplus ... \oplus z_{i_{t_1}}$ как функцию $f(x)$, то можно переписать равенство как $f(x) \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i = 0$ при совпадении значений и $f(x) \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i = 1$ при их различии.

$f(x)$ является булевой функцией, а значит её можно выразить как взвешенную сумму всех возможных линейных приближений.

Конкретно нас интересует разложение на линейные приближения, найденное с помощью преобразования Уолша-Адамара, потому что его поэффициенты рассчитываются следующим образом:

$$W_f(\overline{u}) = \frac{1}{2^n} \sum_{x \in V_n} (-1)^{f(x) \oplus <\overline{x}, \overline{u}>}$$

Если рассчитать линейное приближение

$$\delta(f(x) \oplus < \overline{x}, \overline{u} >) = 1 - 2P(f(x) \oplus < \overline{x}, \overline{u} >= 1),$$

то коэффициенты Уолша-Адамара можно переписать как

$$W_f(\overline{u}) = 1 - \frac{1}{2^{n-1}} || f(x) \oplus < \overline{x}, \overline{u} > ||,$$

где $||...||$ - вес. Данная интерпретация коэффициента показывает близость исследуемой функции $f(x)$ к линейной функции $< \overline{x}, \overline{u} >$.

Но у нас довольно интересная ситуация - нам известны не все значения для функции $f(x)$. Мы можем воспользоваться тем фактом, что если в разложении Фурье раскладывать по этому же базису не $f(x)$, а функцию $(-1)^{f(x)}$, то мы получим коэффициенты Уолша-Адамара. То есть мы можем недостающие значения функции $(-1)^{f(x)}$ заменить на 0, и применять непосредственно к ним быстрое преобразование Фурье, получив таким образом преобразование Уолша-Адамара.

Теперь мы можем приступить к разработке критерия правильности начального состояния.

Для фиксированного вектора $(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)})$ определим вспомогательную функцию

$$h(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)}) = \sum_{(i_1,...,i_{t_1})\in(6)} (-1)^{z_{i_1}\oplus...\oplus z_{i_{t_1}}}$$

(суммируем по наборам индексов бит из гаммы $(i_1, ..., i_{t_1})$ из множества $\Omega^{(1)}$).

Если вектор $(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)})$ не соответствует ни одному из $\Omega^{(1)}$ уравнений проверки четности, то $h(x_0^{(1)}, x_1^{(1)}, ..., x_{k^{(1)}-1}^{(1)}) = 0$.

Тем самым мы определили функцию $h : GF(2)^{k^{(1)}} \to \mathbb{R}$.

Данную функцию мы соберем из наших уравнений проверок четности и составим для нее таблицу истинности (с учетом того, что там могут быть значения '-1', '0' и '1').

Его можно вычислить, используя быстрое преобразование Уолша-Адамара (FWT). Пусть векторы $u, x \in GF(2)^{k^{(1)}}$, тогда

$$u \cdot x = \sum_{i=0}^{k^{(1)}-1} u_i x_i.$$

Рассмотрим преобразование Уолша-Адамара первого рода для функции h:

$$H(u) = \sum_{x\in GF(2)^{k^{(1)}}} h(x)(-1)^{u\cdot x} = \sum_{\Omega^{(1)}} (-1)^{z_{i_1}\oplus...\oplus z_{i_{t_1}}\oplus\sum_{i=0}^{k^{(1)}-1} x_i^{(1)}u_i} = \Omega_0^{(1)} - \Omega_1^{(1)}$$

Здесь

$$\Omega_0^{(1)} = |\{(i_1, ..., i_{t^{(1)}}) \in (6) : z_{i_1} \oplus ... \oplus z_{i_{t^{(1)}}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)}u_i = 0\}|$$

$$\Omega_1^{(1)} = |\{(i_1, ..., i_{t^{(1)}}) \in (6) : z_{i_1} \oplus ... \oplus z_{i_{t^{(1)}}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)}u_i = 1\}|$$

$$\Omega_0^{(1)} + \Omega_1^{(1)} = \Omega^{(1)}$$

То есть фактически функция $h(x)$ показывает разницу между количеством уравнений проверок четности, для которых подошло данное заполнение, и количеством уравнений проверок четности, к которым данное начальное состояние не прошло.

Пусть вектор u равен $(a_0^{'}, ..., a_{k^{(1)}-1}^{'})$, тогда по формулам (8) и (9) можно получить следующее:

$$\sum_{(i_1,...,i_{t^{(1)}})\in\Omega^{(1)}} (\Delta(i_1, ..., i_{t_1}) \oplus 1) = \sum_{(i_1,...,i_{t^{(1)}})\in\Omega^{(1)}} \left(z_{i_1} \oplus ... \oplus z_{i_{t_1}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)}a_i^{'} \oplus 1 \right) = \Omega_0^{(1)}$$

Заметим, что

$$\frac{H(u)+\Omega^{(1)}}{2}=\frac{\Omega_0^{(1)}-\Omega_1^{(1)}+\Omega_0^{(1)}+\Omega_1^{(1)}}{2}=\Omega_0^{(1)}$$

Таким образом, мы получили равенство:

$$\sum_{i=1}^{\Omega^{(1)}}(\Delta(i_1,...,i_{t_1})\oplus 1)=\frac{H(u)+\Omega^{(1)}}{2}$$

Мы будем использовать пороговое значение $T^{(1)}$ как границу для принятия решения, то есть, если

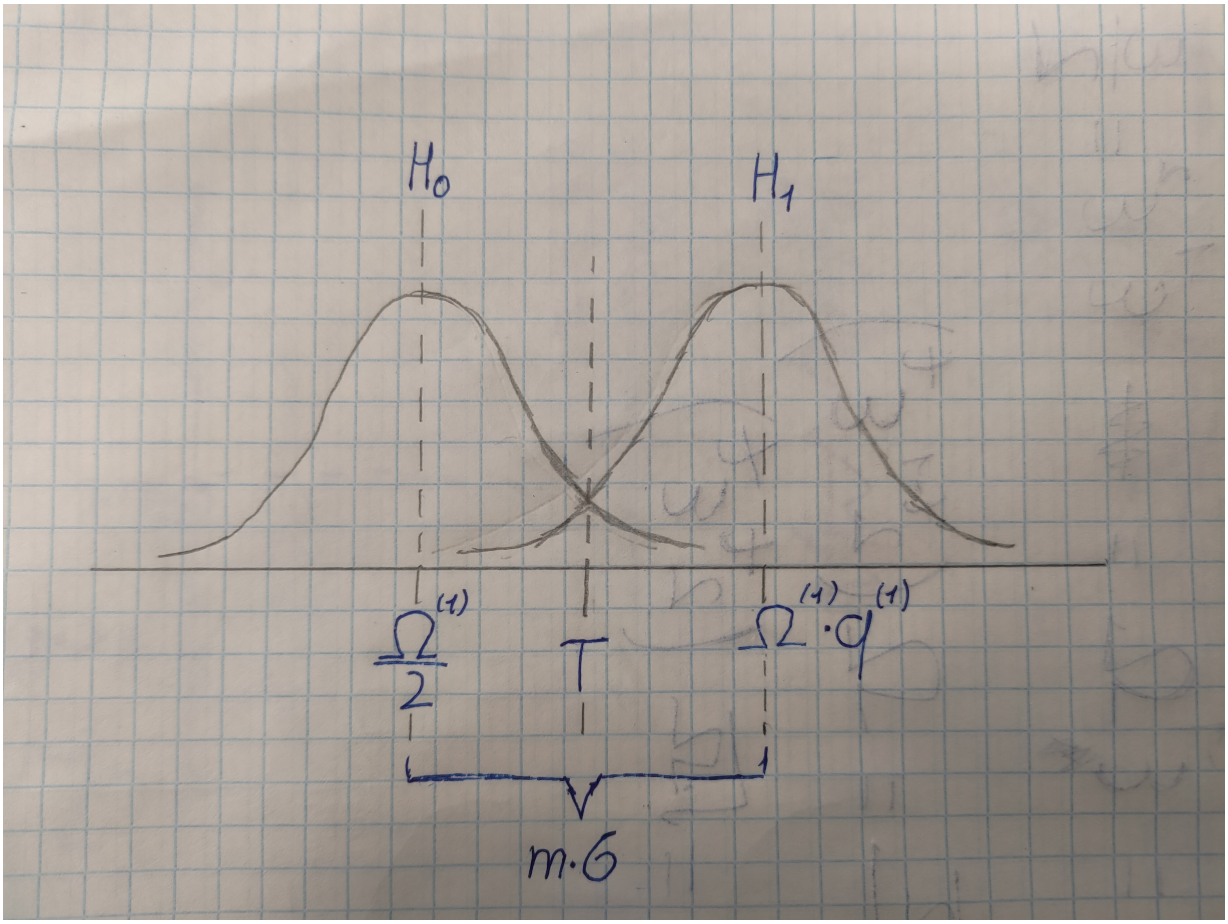
$$\frac{H((a_0^{'},...,a_{k^{(1)}-1}))+\Omega^{(1)}}{2}\geq T^{(1)}, \tag{12}$$

то мы считаем, что правильно угадали начальное состояние $(a_0^{'},...,a_{k^{(1)}-1})$. Ниже рассказано, как именно мы определяем значение порога $T^{(1)}$.

2.3. Пороговое значение

Нашей задачей является различение двух нормальных распределений (на самом деле распределения биномиальные (геометрические), но при большом количестве испытаний они стремятся к нормальным) с мат. ожиданиями $\Omega^{(1)}\cdot q^{(1)}$, $\frac{\Omega^{(1)}}{2}$ и с среднеквадратическим отклонением $\sigma=\sqrt{D}=\sqrt{\frac{\Omega^{(1)}}{4}}=\frac{\sqrt{\Omega^{(1)}}}{2}$ Для обоих случаев (для второго распределения $\sigma=\sqrt{\Omega^{(1)}q^{(1)}(1-q^{(1)})}\approx\frac{\sqrt{\Omega^{(1)}}}{2}$). Для этого нам нужно найти границу T, по которой мы будем принимать решения принятии или отвержении рассматриваемых гипотез.

Для лучшего понимания ситуации изобразим все графически:



где m-некоторый параметр, который в идеале должен равняться 6-ти (3 сигмы с каждой стороны, тогда по правилу трех сигм распределения почти не пересекутся). Выразим расстояние между центрами нормальных распределений:

$$\Omega^{(1)} q^{(1)} - \frac{\Omega^{(1)}}{2} =$$

$$= \Omega^{(1)}(0.5 + 2^{t^{(1)}-1}\epsilon^{t^{(1)}}) - \frac{\Omega^{(1)}}{2} =$$

$$= \Omega^{(1)} 2^{t^{(1)}-1}\epsilon^{t^{(1)}}$$

По построению данный данный промежуток равен $m\sigma$:

$$m\sigma = m\frac{\sqrt{\Omega^{(1)}}}{2} = \Omega^{(1)} 2^{t^{(1)}-1}\epsilon^{t^{(1)}}$$

$$m = \sqrt{\Omega^{(1)}} 2^{t^{(1)}}\epsilon^{t^{(1)}}$$

Исходя из данного выражения мы сможем сделать выводы либо о значении m (хорошо ли мы различим гипотезы при зафиксированном $\Omega^{(1)}$), либо о достаточности $\Omega^{(1)}$ при хорошо выбранном m (достаточно ли мы набрали уравнений для того, чтобы хорошо различать гипотезы).

Допустим, нас устраивает выбор m и $\Omega^{(1)}$. Границу T нужно выбрать в середине отрезка, поэтому

$$T = \frac{\Omega^{(1)}}{2} + \frac{m\sigma}{2} = \frac{\Omega^{(1)}}{2} + \Omega^{(1)} 2^{t^{(1)}-2} \epsilon^{t^{(1)}}$$

2.4. Второй и последующие шаги

После того как мы закончили первый шаг, переходим ко второму, на котором мы можем определить следующие $k^{(2)}$ битов начального состояния, используя гамму и уже восстановленную часть (для каждого из состояний $(a_0, ..., a_{k^{(1)}-1})$ мы будем пытаться восстановить вторую часть, считая уже найденную часть верной). Аналогично первому шагу мы перебираем все возможные наборы из t_2 столбцов матрицы G и находим те, для которых выполняются равенства следующего вида:

$$g^{(j_1)} \oplus ... \oplus g^{(j_{t^{(2)}})} = (x_0^{(2)}, ..., x_{k^{(1)}-1}^{(2)}, x_{k^{(1)}}^{(2)}, ..., x_{k^{(1)}+k^{(2)}-1}^{(2)}, \underbrace{0, ..., 0}_{L-k^{(1)}-k^{(2)}})^T, \tag{13}$$

где $\left(x_0^{(2)}, ..., x_{k^{(1)}-1}^{(2)}, x_{k^{(1)}}^{(2)}, ..., x_{k^{(1)}+k^{(2)}-1}^{(2)}\right)$ - вектор, полученный при суммировании.

Уравнения проверки четности на этом шаге будут выглядеть следующим образом:

$$z_{j_1} \oplus ... \oplus z_{j_{t_2}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{'(2)} a_i = \sum_{i=k^{(1)}}^{k^{(1)}+k^{(2)}-1} x_i^{(2)} a_i', \tag{14}$$

где $\left(a_{k^{(1)}}', ..., a_{k^{(1)}+k^{(2)}-1}'\right)$ - предполагаемое начальное состояние, а $a_{j_1} \oplus ... \oplus a_{j_{t_2}} = a_0 x_0^{'(1)} \oplus ... \oplus a_{k^{(1)}} x_{k^{(1)}-1}^{'(1)}$.

Пусть всего будет $\Omega^{(2)}$ уравнений. Преобразуем их и получим $\Omega^{(2)}$ равенств следующего вида:

$$\Delta(j_1, ..., j_{t^{(2)}}) = z_{j_1} \oplus ... \oplus z_{j_{t^{(2)}}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(2)} a_i \oplus \sum_{i=k^{(1)}}^{k^{(1)}+k^{(2)}-1} x_i^{(2)} a_i' =$$

$$= \sum_{i=k^{(1)}}^{k^{(1)}+k^{(2)}-1} \left(a_i \oplus a_i'\right) \oplus \sum_{j=1}^{t_2} e_{i_j}.$$

Как и на первом шаге, определим вспомогательную функцию

$$h^{(2)}\left(x_{k^{(1)}}^{(2)}, ..., x_{k^{(1)}+k^{(2)}-1}^{(2)}\right) = \sum_{(j_1,...,j_{t_1}) \in (13)} (-1)^{z_{j_1} \oplus ... \oplus z_{j_{t_2}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{'(1)} a_i}.$$

Если вектор $\left(x^{(2)}_{k^{(1)}}, ..., x^{(2)}_{k^{(1)}+k^{(2)}-1}\right)$ не соответствует ни одному из уравнений проверки четности, то $h^{(2)} = 0$ в этой точке.

Используем быстрое преобразование Уолша-Адамара, чтобы вычислить значение

$$\sum_{i=1}^{\Omega^2} (\Delta(j_1, ..., j_{t_2}) \oplus 1) = \frac{H^{(2)}(u) + \Omega^{(2)}}{2},$$

где $H^{(2)}$ - это преобразование Уолша-Адамара для функции $h^{(2)}$ и вектор u равен $\left(a'_{k^{(1)}}, ..., a'_{k^{(1)}+k^{(2)}-1}\right)$. Как и прежде, введем пороговое значение $T^{(2)}$ для принятия решения о том, правильно ли угадано начальное состояние. Подбор этого параметра выполняется так же, как и на первом шаге (с заменой всех верхних индексов 1 на 2).

Теперь у нас есть $k^{(1)} + k^{(2)}$ битов начального состояния. На следующих шагах действуем аналогично, пока у нас не закончатся неизвестные биты начального состояния, или пока их не останется настолько мало, что их будет быстрее найти полным перебором. Тогда можно перебрать оставшиеся биты по обычной атаке Зигенитайлера.

3. Схема атаки

Алгоритм атаки заключается в следующем:

1) делим регистр на части по k_i бит, для каждой части проделываем следующие шаги:

2) Подбираем параметр t для всех частей регистра (в реализации он зафиксирован)

3) Набираем уравнения проверки четности (если набралось недостаточно, то мы либо берем гамму большей длины, если есть возможность (больше гамма -> больше линейных соотношений для выбора), либо увеличиваем t (но при этом у нас упадет q и соотношение будет выполняться с очень плохой вероятностью, почти случайно))

4) строим на основе наших уравнений находим значения для функции h (если мы набрали достаточно уравнений проверки четности, то они 'закроют' значения функции на всех точках, к этому нужно стремиться, если позволяет длина регистра).

5) находим значение T

6) Вычисляем преобразование Уолша-Адамара, находим разницу $\Omega_0 - \Omega_1$ для каждого возможного начального заполнения

7) Считаем все состояния, для которых Ω_0 больше порога T, верными

8) для каждого верного состояния восстанавливаем следующий кусок регистра (составляем уравнения проверки четности для куска длины $k_{i_1}+k_{i_2}$, где первые k_{i_1} бит фиксированы).

3.1. Особенности реализации атаки

К реализации атаки на языке Python3, которая будет приведена ниже, хочется сделать некоторые комментарии: при наборе уравнений проверок четности у нас для разных комбинаций бит гаммы возникнуть одинаковая линейная комбинация начальных бит - значение правой части можно определить принципом большинства голосов (значения '0' мы представляем как единицу, а '1' - как минус единицу с помощью выражения '-1^z'). По принципу большинства голосов в какую сторону произошел перевес, то значение и оставляем. Данный подход значительно повышает вероятность правильности линейного уравнения и 'учитывает' информацию из всей гаммы.

Еще один важный момент: при восстановлении второй и последующих частей при формировании функции h уже восстановленные части регистра будут фиксированы - эти значения мы переносим в правую часть равенства:

$$\underbrace{\sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i^{'}}_{\text{эту часть мы можем вычислить}} \oplus \sum_{i=k^{(1)}}^{k^{(2)}-1} x_i^{(1)} a_i = z_{i_1} \oplus ... \oplus z_{i_{t_1}}$$

$$\sum_{i=k^{(1)}}^{k^{(2)}-1} x_i^{(1)} a_i = z_{i_1} \oplus ... \oplus z_{i_{t_1}} \oplus \sum_{i=0}^{k^{(1)}-1} x_i^{(1)} a_i^{'} = const_i$$

В результате мы опять получили несколько одинаковых линейных соотношений с разными правыми частями - опять находим значение принципом большинства голосов.

4. Модификация входных данных

Атака Чжаня и Фэня может быть использована для вскрытия только тех регистров сдвига, у которых корреляционная вероятность выше 0.5. Но мы можем модифицировать входные данные так, чтобы использовать атаку и для вскрытия РСЛОС с корреляционной вероятностью меньше 0.5.

Пусть нам дана функция f, и корреляционная вероятность ее i-ой переменной $p_i = r < 0.5$. В таблице истинности f заменим все 0 на 1 и все 1 на 0. Обозначим полученную функцию $\overline{f}$. Сравним столбцы значений i-ой переменной и функции $\overline{f}$. В тех строках, где раньше значения совпадали, теперь они не совпадают, и наоборот. Тогда корреляционная вероятность p_i стала равна $1 - r > 0.5$.

Мы знаем фрагмент гаммы z, которая получилась при шифровании со старой комбинирующей функцией f. Если бы при шифровании использовалась функция

$\overline{f}$, то получилась бы гамма $\overline{z}$. Каждый ее бит является инверсией соответствующего бита z (то есть все 1 заменены на 0, а 0 на 1).

Таким образом, если нам надо вскрыть регистр сдвига с корреляционной вероятностью $p_i = r < 0.5$, мы должны инверсией из гаммы z получить гамму $\overline{z}$ и считать, что корреляционная вероятность регистра сдвига равна $1 - r$.

Использованная литература

1) "Исслелование методов криптоанализа поточных шифров", Александр Потий, Юрий Избенко

2) "Методы криптоанализа комбинирующих генераторов", выпускная квалификационная работа, Соловьева Д.Г., 2017

3) Zhang, B. Multi-pass fast correlation attack on stream ciphers / B. Zhang, D. Feng // Biham E., Youssef A. M. (eds.) SAC 2006. Lecture Notes in Computer Science. - 2006. - Vol. 4356. - P. 234-248.

4) "Поточные шифры. Результаты зарубежной открытой криптологии", 1997

5) "Основные понятия корреляционных методов криптоанализа поточных шифров."

6) https://en.wikipedia.org/wiki/Hadamard_transform

7) "Constructing of Analysis Mathematical Model for Stream Cipher Cryptosystems", Ayad Ghazi Naser, Fatin A. H. Majeed

8) J. O. Brüer, "On nonlinear combinations of linear shift register sequences,"in Proc.IEEE ISIT, les Arcs, France, June 21-25 1982

9) "Frequency Postulate's Theoretical Calculation for the Sequences Produced by Modified Geffe Generator", Hussein Ali Mohammed Al-Sharifi

10) P. R. Geffe, "How to protect data with ciphers that are really hard to break Electronics, Jan. 4, 1973, pp 99-101

11) David Wagner 'A Generalized Birthday Problem'