УДК 681.3

АЛГЕБРАЇЧНІ АТАКИ НА ПОТОКОВІ ШИФРАТОРИ ЯК УЗАГАЛЬНЕННЯ КОРЕЛЯЦІЙНИХ АТАК

С.О. ПОМЕТУН

Запропоновано нові теоретичні поняття для булевих функцій: кореляція при відомому значенні функції та її розширення. Доведено, що алгебраїчна атака на потокові шифратори без пам'яті зводиться до апроксимації ускладнюючої функції шифратора низькостепеневими поліномами в термінах введеної кореляції. Ця кореляція може бути використана і для опису алгебраїчних атак на інші типи шифраторів.

ВСТУП

Останнім часом почали інтенсивно розвиватися методи криптоаналізу, які базуються на запропонованих декілька років тому так званих алгебраїчних атаках [1,2]. Криптоаналіз великого класу шифраторів можна звести до розв'язання системи нелінійних рівнянь від багатьох змінних над деяким скінченним полем (найчастіше $GF(2)$). Проте обчислювальна складність розв'язання таких систем «лобовими» методами настільки велика, що не дозволяє скільки-небудь суттєво знизити стійкість криптосистеми. Складання і дослідження таких систем та пошуки відносно швидких способів їх розв'язання для конкретних шифраторів відносять, в першу чергу, до аналітичних методів криптоаналізу. Окремий випадок аналітичних методів криптоаналізу — способи зниження степеня рівнянь таких систем шляхом домноження їх на спеціально підібрані поліноми — і називають алгебраїчними атаками (хоча в зарубіжній літературі під алгебраїчними атаками розуміють вже будь-який спосіб відносно швидкого розв'язання таких систем). Із застосуванням аналітичного криптоаналізу можна ознайомитись, наприклад, в роботі [3], де криптоаналіз (знаходження ключа) знаменитого алгоритму DES (стандарт шифрування США до 2001 р. і фактичний світовий стандарт шифрування комерційної (і не тільки) інформації протягом останньої чверті 20-го століття) зводиться до розв'язання системи рівнянь.

Алгебраїчні атаки на потокові шифратори вперше запропоновані у 2003 р. [2], де була показана їх ефективність для певного класу шифраторів — потокових шифраторів без пам'яті, побудованих на основі регістрів

зсуву з лінійним зворотним зв'язком (РЗЛЗЗ). Для деяких шифраторів, наприклад, LILI-128 (пропонувався на Європейський конкурс шифраторів Nessie), Toyocrypt (пропонувався на Японський конкурс шифраторів Cryptrec) алгебраїчна атака має відносно невелику обчислювальну складність (2^{57} та 2^{49} операцій відповідно). Для шифратора Toyocrypt атака потребує лише 20 Кбайт гами (або шифрованого тексту за умови атаки при відомому відкритому тексті) і може бути реалізована на практиці [2]. Зауважимо, що система рівнянь для Toyocrypt має 128 змінних і складність у 2^{49} операцій, що набагато менше, ніж складність повного перебору — 2^{128} операцій. Обчислювальна складність розв'язання системи рівнянь сильно залежить від її степеня. У випадку з Toyocrypt за допомогою алгебраїчної атаки степінь системи вдалося знизити з 63 до 3.

Незабаром з'явилося узагальнення алгебраїчних атак на потокові шифратори з пам'яттю [4]. Функціонування S-блоків (основних нелінійних елементів більшості блокових шифраторів) теж описується системою рівнянь, тому алгебраїчні атаки можуть застосовуватися і до них (див., наприклад, [5]). Таким чином, на сьогоднішній день алгебраїчні атаки потенційно можуть застосовуватися для всіх основних типів сучасних шифраторів. Для шифраторів, побудованих на основі РЗЛЗЗ, алгебраїчні атаки вдалося вдосконалити за рахунок додаткових передобчислень. Такі атаки названо швидкими алгебраїчними атаками [6].

Метою даної роботи є дослідження алгебраїчних атак та розробка відповідних формальних понять.

У першому розділі поставлено, власне, задачу криптоаналізу, в другому — коротко наведено вже відомі результати алгебраїчних атак, у третьому — дається їх формалізація. Нарешті, у четвертому розділі наведено приклади опису алгебраїчних атак за допомогою введених понять та завершено їх логічний зв'язок із кореляційними атаками.

1. ПОСТАНОВКА ЗАДАЧІ

Алгебраїчні атаки будемо розглядати на прикладі алгебраїчних атак на потокові шифратори без пам'яті, побудовані на основі РЗЛЗЗ. Їх функціонування може бути представлено системою рівнянь

$$\begin{cases} f(k_0,...,k_{n-1}) = b_0, \\ f(L(k_0,...,k_{n-1})) = b_1, \\ ..... \\ f(L^{N-1}(k_0,...,k_{n-1})) = b_{N-1}, \end{cases} \qquad (1)$$

де $\overline{k} = (k_0,...,k_{n-1})$ — невідомий ключ (початковий стан регістрів шифратора) довжини n біт; $\overline{b} = (b_0,...,b_{N-1})$ — гама, яка при шифруванні побітово додається до відкритого тексту (сучасні шифратори будуються стійкими до

атак на основі відкритого тексту, тому гаму вважають відомою); N — кількість наявних біт гами (вважається, що доступна достатня кількість гами); L — деякий лінійний оператор (описує зміну станів РЗЛЗЗ); $f : GF(2)^n \to \to GF(2)$ — ускладнююча функція. Оператор L та ускладнююча функція f також відомі, оскільки відомою вважається вся конструкція шифратора.

Цією схемою описуються фільтруючі шифратори, побудовані на одному РЗЛЗЗ, шифратори на декількох РЗЛЗЗ із нелінійно скомбінованими виходами, а також будь-які шифратори без додаткової пам'яті (чи їх частини) на лінійних регістрах зсуву з регулярним (відомим) законом руху. Всі змінні розглядаються над полем $GF(2)$, хоча ідея атаки, в принципі, може бути розширена і на довільні скінченні поля. Задача криптоаналізу — знайти невідомий ключ $\bar{k}$.

2. АЛГЕБРАЇЧНА АТАКА НА ПОТОКОВІ ШИФРАТОРИ

Наведемо основні ідеї та результати, отримані Куртуа та Маєром у роботі [2], де вперше запропоновано та розробляються алгебраїчні атаки на шифратори, які описуються системою (1).

Нехай $\bar{s}^t = L^t(\bar{k}) = L^t(k_0,...,k_{n-1}) = (s_0^t,...,s_{n-1}^t)$ — стан шифратора (заповнення РЗЛЗЗ) в момент часу t. Тоді рівняння (1) будуть мати вигляд

$$\begin{cases} f(\bar{s}^0) = b_0, \\ f(\bar{s}^1) = b_1, \\ ..... \\ f(\bar{s}^{N-1}) = b_{N-1}, \end{cases} \qquad (2)$$

де для будь-якого індексу t, $\bar{s}^t$ — лінійна функція від ключа $\bar{k}$ (запис $L^t(\bar{k})$ позначає t-й степінь лінійного оператора L, тоді як $\bar{s}^t$ — заповнення РЗЛЗЗ у момент часу t, тут t є індексом). На функцію f накладаються вимоги по забезпеченню стійкості шифратора проти різних методів криптоаналізу. Зокрема, функцію f вибирають достатньо високого степеня (під степенем функції розуміється степінь поліному Жегалкіна, яким подається функція). Це робиться для того, щоб систему (2) важко було розв'язати методом лінеаризації, при якому кожний терм поліному замінюється новою змінною, і система (2) перетворюється на систему лінійних рівнянь з великою кількістю змінних. Як правило, функція f використовує лише $k << n$ з її вхідних аргументів. Таким чином, набір рівнянь (2) — це сильно перевизначена система нелінійних рівнянь від n змінних деякого степеня $r \le k$, де k — кількість суттєвих аргументів функції f. На практиці n часто лежить у діапазоні $80…256$, k — у діапазоні $10…17$, r близьке до k.

Метою алгебраїчної атаки є отримання іншої системи (3) степеня $k' < r$, яка *випливає* з системи (2), але *не є еквівалентною* їй. Для цього рівняння (2) домножаються на деяку функцію $g(\bar{s}^t)$, і отримуємо

$$\begin{cases} f(\bar{s}^0)g(\bar{s}^0) = b_0 g(\bar{s}^0), \\ f(\bar{s}^1)g(\bar{s}^1) = b_1 g(\bar{s}^1), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f(\bar{s}^{N-1})g(\bar{s}^{N-1}) = b_{N-1} g(\bar{s}^{N-1}). \end{cases} \qquad (3)$$

Для простоти позначимо суттєві аргументи функцій f та g через $\bar{x} = (x_0, ..., x_{k-1})$ і будемо вважати, що f та g — функції від k аргументів. Виявляється, що в багатьох випадках можна підібрати таку g, що для степенів ($\deg$) буде виконуватись $\deg(f(\bar{x})g(\bar{x})) < \deg(f(\bar{x}))$. Наприклад, якщо $f(\bar{x}) = x_1x_2x_4x_5 + x_2x_3x_6 + x_1$, то при $g(\bar{x}) = x_2 + 1$

$$\deg(f(\bar{x})g(\bar{x})) = \deg[(x_1x_2x_4x_5 + x_2x_3x_6 + x_1)(x_2 + 1)] =$$

$$= \deg[(x_1x_3x_5 + x_3x_6)x_2(x_2 + 1) + x_1(x_2 + 1)] =$$

$$= \deg x_1(x_2 + 1) = 2 < \deg(f(\bar{x})) = 4,$$

бо $x_2(x_2 + 1) = 0$, $x_2 \in GF(2)$.

Степінь правої частини (3) дорівнює $\deg(g(\bar{x}))$ або нулю в залежності від того, b_t дорівнює одиниці чи нулю.

Випадок, коли до системи (3) включають всі рівняння з (2) ($\deg g(\bar{x}) \le \le \deg f(\bar{x})g(\bar{x})$) назвемо сценарієм атаки A1.

Якщо степінь $g(\bar{x})$ надто великий, тобто $\deg f(\bar{x})g(\bar{x}) < \deg f(\bar{x})$ та $\deg f(\bar{x})g(\bar{x}) < \deg g(\bar{x})$, то до системи (3) записують лише ті рівняння, для яких $b_t = 0$ (сценарій атаки A2).

Також можливий випадок, коли $f(\bar{x})g(\bar{x}) = 0$ тотожньо, і $\deg(g(\bar{x})) < < \deg(f(\bar{x}))$, тоді до (3) записують тільки ті рівняння, для яких $b_t = 1$, і система матиме вигляд $g(\bar{s}^t) = 0$, $t \in \{t : b_t = 1\}$ (сценарій атаки A3).

Отриману систему меншого степеня (3) часто розв'язувати простіше (тобто для розв'язання буде потрібно менше обчислювальних операцій), бо її степінь менший. Зазвичай це робиться методом лінеаризації або його модифікаціями.

Доведено твердження, що визначає умови, за яких, незалежно від функції f, степінь системи (2) може бути знижений [2].

Твердження 1. Для будь-якої функції $f : GF(2)^k \to GF(2)$ існує функція $g(\bar{x}) \neq 0$ степеня не більше, ніж $\lceil k/2 \rceil$, така, що степінь $f(\bar{x})g(\bar{x})$ не

більше, ніж $\lfloor k/2 \rfloor$ (і може дорівнювати нулю), де $\lceil k/2 \rceil, \lfloor k/2 \rfloor$ — відповідно верхнє і нижнє округлення числа $k/2$ до найближчого цілого.

Наслідок. Можна вибрати таку функцію $g(\bar{x})$, коли незалежно від вибраного сценарію атаки степінь системи (3) не перевищує $\lceil k/2 \rceil$.

Доведення твердження 1 також дає конструктивний шлях знаходження функції $g(\bar{x})$ зі складністю порядку 2^{3k} операцій.

В принципі можна розглядати випадки, коли рівняння отриманої системи (3) не будуть низького степеня, але вони будуть наближатися з високою точністю рівняннями низького степеня. Це утворює сценарії атак Б1–Б3.

3. ФОРМАЛІЗАЦІЯ АЛГЕБРАЇЧНОЇ АТАКИ

Наведемо запропоновану в даній роботі формалізацію описаної вище алгебраїчної атаки (з узагальненням на сценарії Б1–Б3). Вводемо поняття кореляції за відомого значення функції та розширення булевої функції, що дасть змогу описати алгебраїчну атаку на потокові шифратори без пам'яті як кореляційну атаку в термінах введеної кореляції.

Фактично вище мова йшла про спрощення розв'язання системи нелінійних рівнянь вигляду (2) шляхом заміни їх на систему (3).

$$\begin{cases} f(\bar{s}^0) = b_0, \\ f(\bar{s}^1) = b_1, \\ \ldots\ldots\ldots\ldots \\ f(\bar{s}^{N-1}) = b_{N-1}, \end{cases} \Rightarrow \begin{cases} f(\bar{s}^0)g(\bar{s}^0) = b_0 g(\bar{s}^0), \\ f(\bar{s}^1)g(\bar{s}^1) = b_1 g(\bar{s}^1), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f(\bar{s}^{N-1})g(\bar{s}^{N-1}) = b_{N-1} g(\bar{s}^{N-1}). \end{cases}$$

Тут і далі $\Rightarrow$ позначає логічне слідування (імплікацію), а $\Leftrightarrow$ — еквівалентність.

Виграш в тому, що система (3) може розв'язуватися простіше. У даному випадку вона має менший степінь. Програш — можливість появи зайвих коренів.

В середньому (за умови збалансованості функцій f та g) кожному рівнянню системи (3), які після перейменування аргументів мають вигляд $f(\bar{x})g(\bar{x}) = a\,g(\bar{x})$, задовольняє $3/4$ всіх аргументів (окрім тих, що задовольняють одночасно рівнянням $g(\bar{x}) = 1$ та $f(\bar{x}) = a + 1$). Вважаємо, що рівняння системи (3) — незалежні, тоді для відсіювання всіх зайвих коренів необхідна така кількість рівнянь N, що $(3/4)^N \sim 1/2^n$ (2^n — кількість різних можливих розв'язків). Звідки $N = n\log_{3/4}(1/2) \approx 2{,}4n$. Для незбалансованих функцій зміниться лише константа перед n. Для розв'язання системи (3) методом лінеаризації необхідно близько $\sum_{i=0}^{k'=\deg(fg)} C_n^i >> 2{,}4n$ рівнянь, тому зайвих коренів практично ніколи немає.

Заміна системи (2) на (3) для кожного окремого рівняння записується так: $f(\bar{x}) = a \Rightarrow f(\bar{x})g(\bar{x}) = ag(\bar{x})$, але $f(\bar{x})g(\bar{x}) = ag(\bar{x}) \Leftrightarrow (f(\bar{x}) + a)\, g(\bar{x}) + a = a$. Позначимо $h_a(\bar{x}) \equiv (f(\bar{x}) + a)\, g(\bar{x}) + a$, тоді маємо $h_a(\bar{x}) = a$, і заміна (2) на (3) запишеться так:

$$\begin{cases} f(\bar{s}^0) = b_0, \\ f(\bar{s}^1) = b_1, \\ \ldots\ldots\ldots\ldots \\ f(\bar{s}^{N-1}) = b_{N-1}, \end{cases} \Rightarrow \begin{cases} h_{b_0}(\bar{s}^0) = b_0, \\ h_{b_1}(\bar{s}^1) = b_1, \\ \ldots\ldots\ldots\ldots \\ h_{b_{N-1}}(\bar{s}^{N-1}) = b_{N-1}. \end{cases} \tag{4}$$

Для більш ґрунтовного опису зручно ввести означення.

Означення 1. Множиною a-значень функції f назвемо множину $X_a(f) = \{\bar{x} \in GF(2)^k : f(\bar{x}) = a\}$ всіх аргументів функції f, на яких вона приймає конкретне значення $a \in GF(2)$.

Так, $X_0(f)$ — множина нулів f. Будь-яка булева функція f взаємнооднозначно визначається множиною своїх a-значень $X_a \subset GF(2)^k$.

Мають місце рівності $X_a(f) = X_{a+1}(f+1)$, $X_a(f+a) = X_0(f)$, $X_0(fh) = X_0(f) \cup X_0(h)$, $X_1(fh) = X_1(h) \cap X_1(h)$, $X_0(f) \oplus X_1(f) = GF(2)^k$.

Означення 2. Функцію $h(\bar{x})$ назвемо a-розширенням $f(\bar{x})$, якщо $X_a(f) \subset X_a(h)$.

Тепер імплікацію $f(\bar{x}) = a \Rightarrow h(\bar{x}) = a$ можна записати так: $h(\bar{x})$ є a-розширенням $f(\bar{x})$ або $X_a(f) \subset X_a(h)$.

Має місце еквівалентність $X_a(f) \subset X_a(h) \Leftrightarrow X_{a+1}(h) \subset X_{a+1}(f)$, тобто, наприклад, якщо h є 0-розширенням f, то f є 1-розширенням h.

Будь-яка функція завжди має два тривіальних розширення: $X_0(f) \subset \subset X_0(0)$, $X_1(f) \subset X_1(1)$.

Означення 3. Функцію $h(\bar{x})$ назвемо розширенням $f(\bar{x})$ (позначимо $X(f) \subset X(h)$), якщо $X_0(f) \subset X_0(h)$ або $X_1(f) \subset X_1(h)$).

За таких означень, якщо $h(\bar{x})$ є розширенням $f(\bar{x})$, то й $f(\bar{x})$ є розширенням $h(\bar{x})$.

Доведемо наступне основне твердження про еквівалентність.

Твердження 2. Для будь-якої $h(\bar{x})$, яка є a-розширенням $f(\bar{x})$, існує $g(\bar{x})$ така, що $h(\bar{x}) = a \Leftrightarrow f(\bar{x})g(\bar{x}) = ag(\bar{x})$, і навпаки, для будь-якої $g(\bar{x})$ існує єдина $h(\bar{x})$ така, що $h(\bar{x})$ є a-розширенням $f(\bar{x})$ та $f(\bar{x})g(\bar{x}) = = ag(\bar{x}) \Leftrightarrow h(\bar{x}) = a$.

Доведення.

1. Існування $g(\bar{x})$.

За умовою $X_a(f) \subset X_a(h)$. Покладемо $X_0(g) = X_a(h) \setminus X_a(f) \oplus A$, де $A \subset X_a(f)$ — довільна підмножина a-значень функції f. Тоді $fg = ag \Leftrightarrow$ $\Leftrightarrow (f + a)g + a = a$ та

$$X_a((f + a)g + a) = X_0((f + a)g) = X_0(f + a) \cup X_0(g) = X_a(f) \cup X_0(g) =$$

$$= X_a(f) \cup [X_a(h) \setminus X_a(f) \oplus A] = X_a(f) \cup (X_a(h) \setminus X_a(f)) \cup X_a(f) \cup A =$$

$$= X_a(f) \cup (X_a(h) \setminus X_a(f)) = X_a(h).$$

Отже,

$$(f + a)g + a = h, \text{ тобто } fg = ag \Leftrightarrow h = a.$$

2. Існування $h(\bar{x})$.

$fg = ag \Leftrightarrow (f + a)g + a = a$, отже шукана $h = (f + a)g + a$.

Зауваження. Фактично це твердження означає, що домноження рівняння $f(\bar{x}) = a$ на деяку функцію $g(\bar{x})$ (метод алгебраїчної атаки) еквівалентне його заміні рівнянням $h(\bar{x}) = a$, де $h(\bar{x})$ — a-розширення $f(\bar{x})$ (див. рівняння (4)). Більше того, кожному розширенню $h(\bar{x})$ відповідає $2^{|X_a(f)|}$ різних $g(\bar{x})$ (відповідно до кількості множин A), і такий опис алгебраїчної атаки, на відміну від домноження, не є надлишковим.

Метод алгебраїчної атаки також припускає, що $h(\bar{x})$ може не бути низького степеня, але з високою ймовірністю наближається функцією низького степеня $h'(\bar{x})$. Тобто $f(\bar{x}) = a \Rightarrow h(\bar{x}) = a$ (або, що теж саме, $P(h = f \mid f = a) = 1$) та $P(h' = h) = 1 - \varepsilon$, де ε — маленьке. Тоді $P(h' = f \mid f = a) = 1 - \delta$, де δ — теж маленьке. Для опису цього випадку узагальнимо поняття a-розширення функції і введемо кореляцію за відомого значення функції.

Означення 4. Кореляцією функцій $h(\bar{x})$ та $f(\bar{x})$ за відомого значення $f(\bar{x}) = a$ назвемо величину

$$C_a(h, f) = P(h = f \mid f = a) - P(h \neq f \mid f = a).$$

Якщо $P(f = a) = 0$, то $C_a(h, f) = 1$ для будь-якої $h(\bar{x})$.

Нагадаємо означення звичайної кореляції.

Означення 5. Кореляцією булевих функцій $h(\bar{x})$ та $f(\bar{x})$ називають величину

$$C(h, f) = P(h = f) - P(h \neq f).$$

Мають місце рівності

$$C_a(h, f) = P(h = f \mid f = a) - P(h \neq f \mid f = a) =$$

$$= P(h = f \mid f = a) - (1 - P(h = f \mid f = a)) =$$

$$= 2P(h = f \mid f = a) - 1 = 2P(h = a \mid f = a) - 1,$$

$$C_a(h, f) + C_a(h+1, f) = 0.$$

Таким чином, запис $C_a(h, f) = 1 - \varepsilon$ означає, що функція $h(\bar{x})$ є апроксимацією з точністю $1 - \varepsilon$ функції $f(\bar{x})$ в термінах кореляції за відомого значення $f(\bar{x}) = a$. Фактично це звичайна апроксимація, але не на всій множині аргументів $GF(2)^k$, а на підмножині $X_a(f)$ a-значень $f(\bar{x})$.

Твердження 3. Функція $h(\bar{x})$ є a-розширенням $f(\bar{x})$ тоді і лише тоді, коли кореляція $h(\bar{x})$ та $f(\bar{x})$ за відомого значення $f(\bar{x}) = a$ дорівнює 1.

Доведення.

1. Необхідність.

Нехай $f = a \Rightarrow h = a$, тоді

$$C_a(h, f) = 2P(h = a \mid f = a) - 1 = 2 - 1 = 1.$$

2. Достатність.

$$2P(h = a \mid f = a) - 1 = 1 \Rightarrow P(h = a \mid f = a) = 1 \Rightarrow (f = a \Rightarrow h = a).$$

Тепер, в термінах введених означень ідея алгебраїчної атаки має такий вигляд:

$$\begin{cases} f(\bar{s}^0) = b_0, \\ f(\bar{s}^1) = b_1, \\ \ldots\ldots\ldots\ldots \\ f(\bar{s}^{N-1}) = b_{N-1}, \end{cases} \Rightarrow \begin{cases} h_{b_0}(\bar{s}^0) = b_0, \\ h_{b_1}(\bar{s}^1) = b_1, \\ \ldots\ldots\ldots\ldots \\ h_{b_{N-1}}(\bar{s}^{N-1}) = b_{N-1}. \end{cases}$$

Відповідні кореляції для функцій h_0 та h_1 дорівнюють: $C_0(h_0, f) = 1 - \varepsilon_0$ та $C_1(h_1, f) = 1 - \varepsilon_1$. Ймовірності істинності рівнянь системи (4)

$$P(h_0 = 0 \mid f = 0) = \frac{C_0(h, f) + 1}{2} = 1 - \varepsilon_0 / 2$$

та

$$P(h_1 = 1 \mid f = 1) = \frac{C_1(h, f) + 1}{2} = 1 - \varepsilon_1 / 2.$$

Якщо для якогось $a \in GF(2)$ $\varepsilon_a = 0$, то маємо окремий випадок $C_a(h, f) = 1$, $f(\bar{x}) = a \Rightarrow h(\bar{x}) = a$, що дає детерміновану систему. Цей випадок описує сценарії А1–А3. Випадок $\varepsilon \neq 0$ дає систему зі спотвореними правими частинами (ймовірності спотворення для h_0 та h_1 — $\varepsilon_0 / 2$ та $\varepsilon_1 / 2$ відповідно) і описує сценарії Б1–Б3 (див. розд. 2).

Таким чином, алгебраїчна атака зведена до пошуку хороших апроксимацій ускладнюючої функції в термінах введеної кореляції, тобто представлена як певний вид вже досить давно відомих кореляційниих атак [7, 8].

Твердження 4. За умови збалансованості $f(\bar{x})$ має місце співвідношення

$$C(h, f) = \tfrac{1}{2}(C_0(h, f) + C_1(h, f)).$$

Доведення.

$$\tfrac{1}{2}(C_0(h,f)+C_1(h,f))=\tfrac{1}{2}(2P(h=f\mid f=0)-1+2P(h=f\mid f=1)-1)=$$

$$=P(h=f\mid f=0)+P(h=f\mid f=1)-1=$$

$$=P(h=f,f=0)/P(f=0)+P(h=f,f=1)/P(f=1)-1=$$

$$=(P(h=f,f=0)+P(h=f,f=1))/(1/2)-1=2P(h=f)-1=C(h,f).$$

Наслідок. $\left|C(h,f)\right|\le\max(\left|C_0(h,f)\right|,\left|C_1(h,f)\right|)$.

Твердження 5. При фіксованій $f(\bar{x})$ та випадковій $h(\bar{x})$ випадкові величини $C_0(h,f)$ та $C_1(h,f)$ є незалежними.

Доведення. $C_0(h,f)$ є функцією від звуження $h(\bar{x})$ на множину $X_0(f)$, а $C_1(h,f)$ — функцією від звуження $h(\bar{x})$ на множину $X_1(f)$. Оскільки $X_0(f)\cap X_1(f)=0$, то ці звуження розподілені незалежно, значить, і $C_0(h,f)$ та $C_1(h,f)$ теж незалежні.

Приклад до тверджень 4, 5.

$$f(\bar{x})=f(x_1,x_2)=x_1+x_2,\ h(\bar{x})=h(x_1,x_2)=x_1x_2+1,$$

$$C(h,f)=2P(h=f)-1=2\cdot\tfrac{3}{4}-1=1/2,$$

$$C_0(h,f)=2P(h=0\mid f=0)-1=2\cdot\tfrac{1}{2}-1=0,$$

$$C_1(h,f)=2P(h=1\mid f=1)-1=2\cdot1-1=1,$$

$$C(h,f)=1/2=\tfrac{1}{2}(C_0(h,f)+C_1(h,f))=\tfrac{1}{2}(0+1)=1/2.$$

Твердження 4 та 5 теоретично обґрунтовують можливість існування хороших (з досить високим абсолютним значенням кореляції) низькостепеневих апроксимацій в термінах C_0 та C_1, навіть якщо не існує хороших низькостепеневих апроксимацій в термінах звичайної кореляції C. Наслідок з твердження 4 означає, що апроксимація хоча б за однією з кореляцій C_0 чи C_1 завжди є, принаймні, не гіршою, ніж апроксимація за звичайною кореляцією C. Тобто, якщо функція не має хороших апроксимацій по відношенню до кореляцій C_0 та C_1, то вона автоматично не має хороших апроксимацій по відношенню до кореляції C, але не навпаки. Це означає, що кореляція за відомого значення функції є в певному розумінні узагальненням звичайної кореляції. Відповідно алгебраїчні атаки можна розуміти не просто як певний вид кореляційних атак, а як їх узагальнення, завдяки узагальненню поняття кореляції булевих функцій.

Спираючись на результати, отримані в роботі [5], де уніфікуються алгебраїчні атаки на потокові шифратори без пам'яті, з пам'яттю та S-блоки (відображення $GF(2)^n\to GF(2)^m$, які часто є основою блокових шифраторів), можна стверджувати, що алгебраїчні атаки на блокові шифратори та

потокові шифратори з пам'яттю теж можуть бути зведені до кореляційних в термінах введеної кореляції.

4. ЗАСТОСУВАННЯ АТАКИ

Таким чином, вразливість того чи іншого шифратора до алгебраїчної атаки визначається наявністю низькостепеневих функцій $h(\bar{x})$, сильно скорельованих (часто з кореляцією, що дорівнює 1) з функцією $f(\bar{x})$, в термінах кореляції за відомого значення $f(\bar{x}) = a$, тобто наявністю наближень (апроксимацій) в термінах такої кореляції для $a = 0$ або $a = 1$. Більш детально набір рівнянь (2) розбивається на дві приблизно рівні за потужністю множини $A_0 = \{f(\bar{s}^t) = 0 | t : b_t = 0\}$ та $A_1 = \{f(\bar{s}^t) = 1 | t : b_t = 1\}$ (нагадаємо, що гама відома). І якщо, наприклад, знайдеться така низькостепенева $h_0(\bar{x})$, що $C_0(h_0, f) = 1 - \varepsilon$, то будується множина рівнянь $A_0' = \{h_0(\bar{s}^t) = 0 | t : b_t = 0\}$. Складність її розв'язання часто набагато менша, ніж складність розв'язання вихідної системи (при $\varepsilon = 0$ A_0' детермінована, а при $\varepsilon \neq 0$ має спотворені праві частини з ймовірністю спотворення $\varepsilon / 2$). Аналогічні дії виконуються для множини A_1, якщо $f(\bar{x})$ добре наближається деякою низькостепеневою функцією в термінах кореляції C_1.

Приклади з практики. Для фільтруючої функції $f(\bar{x})$ (63-го степеня від 128 змінних) шифратора Toyocrypt знайшлися такі функції $h_0(\bar{x})$, $h_1(\bar{x})$ 3-го степеня, що $C_0(h_0, f) = 1$ та $C_1(h_1, f) = 1$(!) [2]. Складність розв'язання нової системи порядку 2^{49} операцій. Така атака може бути реалізована на практиці. Степінь ускладнюючої функції знижений набагато нижче границі, яка дається твердженням 2, тому можна припустити, що шифратор не розроблявся стійким проти цієї атаки. Для фільтруючої функції $g(\bar{x})$ (6-го степеня від 10 змінних) шифратора LILI-128 знайшлися такі $h_0(\bar{x})$, $h_1(\bar{x})$ 4-го степеня, що $C_0(h_0, f) = 1$ та $C_1(h_1, f) = 1$ [2]. Це теж суттєвий виграш (адже при оцінці складності криптоаналітичних атак розробники вважали, що степінь ускладнюючої функції дорівнює все ж таки шести, бо алгебраїчні атаки ще були невідомі). Окрім того, твердження 2 дає верхню границю для знижуваного степеня. В цьому випадку $\lceil 10/2 \rceil = 5 > 4$, тобто ускладнююча функція вибрана не оптимально.

Варто зауважити, що придатні тільки нетривіальні апроксимації, оскільки завжди мають місце рівності $C_0(0, f) = 1$, $C_1(1, f) = 1$, але рівняння $0 = 0$ та $1 = 1$ не дають ніякої інформації про ключ.

Однією з основних вимог до ускладнюючої функції шифратора (для запобігання кореляційним атакам) є неможливість її хорошого наближення за допомогою афінних функцій. Чому ж перевірялась можливість наближення в термінах звичайної кореляції C, а кореляції C_0 та C_1 випали з уваги? Одна з можливих відповідей дається таким твердженням.

Твердження 6. Якщо $h(\bar{x})$, $f(\bar{x})$ — збалансовані функції, то $C_0(h, f) = C_1(h, f) = C(h, f)$.

Доведення. Для збалансованих функцій $f(\bar{x})$ та $h(\bar{x})$ мають місце рівності

$$\left.\begin{aligned} P(f=0) = P(f=0, h=0) + P(f=0, h=1) = 1/2 \\ P(h=1) = P(f=1, h=1) + P(f=0, h=1) = 1/2 \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow P(f=0, h=0) = P(f=1, h=1).$$

Отже, для будь-якого $a \in GF(2)$

$$C(h, f) = 2P(h = f) - 1 = 2(P(h=0,\ f=0) + P(h=1, f=1)) - 1 =$$

$$= 4P(h=a,\ f=a) - 1 = \frac{2P(h=a,\ f=a)}{1/2} - 1 =$$

$$= \frac{2P(h=a,\ f=a)}{P(f=a)} - 1 = 2P(h=a \mid f=a) - 1 = C_a(h, f).$$

Взявши до уваги, що ускладнююча функція $f(\bar{x})$ шифратора завжди вибирається збалансованою і будь-яка невироджена афінна функція $l(\bar{x})$ теж збалансована, бачимо, що, поки $f(\bar{x})$ наближалася афінними функціями $l(\bar{x})$, різниці між цими кореляціями не було. Різниця проявляється лише тоді, коли $f(\bar{x})$ наближається незбалансованими (тобто, з необхідністю не афінними) функціями $h(\bar{x})$, що й має місце в алгебраїчних атаках.

ВИСНОВКИ

У роботі досліджувались алгебраїчні атаки на потокові шифратори без пам'яті. Було введено новий тип кореляції булевих функцій — кореляцію за відомого значення функції. Це дало змогу об'єднати шість різних сценаріїв алгебраїчної атаки в єдине логічне ціле. До цього досліджувалися лише три простіші сценарії, які приводили до детермінованих систем рівнянь (у запропонованому описі атаки це відповідає пошуку апроксимацій ускладнюючої функції з кореляцією, що дорівнює строго одиниці). Показано, що введена кореляція є в певному сенсі узагальненням звичайної кореляції, і тому алгебраїчні атаки (на всі типи шифраторів) можуть розглядатися як узагальнення кореляційних атак.

Запропонований опис дозволяє теоретично досліджувати шість сценаріїв атаки одночасно, а також може бути теоретичною основою для розробки алгоритмів перевірки шифратора на вразливість проти всіх шести сценаріїв атаки.

Захищеність інформації (фізична, організаційна і криптографічна) є принципово важливою властивістю інформації у сучасному світі. Дана робота може бути корисною для покращення криптографічного захисту інформації.

ЛІТЕРАТУРА

1. *Courtois N., Pieprzyk J.* Cryptanalysis of block ciphers with overdefined systems of equations, Advances in Cryptology // Lecture Notes in Computer Science. — 2002. — **2501**. — P. 267–287.
2. *Courtois N., Meier W.* Algebraic Attacks on Stream Ciphers with Linear Feedback // Ibd. — 2003. — **2656**. — P. 345–359.
3. *Schaumuller-Bichl I.* Cryptanalysis of the Data Encryption Standard by the Method of Formal Coding // Ibd. — 1983.— **149**. — P. 235–255.
4. *Armknecht F., Krause M.* Algebraic attacks on Combiners with Memory // Ibd. — 2003. — **2729**. — P. 162–176.
5. *Armknecht F.* On the Existence of low-degree Equations for Algebraic Attacks // http://eprint.iacr.org/2004/185/.
6. *Courtois N.* Fast Algebraic Attacks on Stream Ciphers with Linear Feedback // Lecture Notes in Computer Science. — 2003. —**2729**. — P. 177–194.
7. *Siegenthaler T.* Cryptanalyst's representation of nonlinearly filtered ml-sequences // Ibd. — 1986. — **219**. — P. 103–110.
8. *Meier W., Staffelbach O.* Fast correlation attacks on certain stream ciphers // Journal of Cryptology. — 1989. — **I**, № 3. — P. 159–176.

Надійшла 08.11.2006