РОЗДІЛИ СУЧАСНОЇ КРИПТОЛОГІЇ
Комп'ютерний практикум №2

Лінійний криптоаналіз блокових шифрів

1. Мета роботи

Опанування сучасних методів криптоаналізу блокових шифрів, набуття навичок у дослідженні стійкості блокових шифрів до лінійного криптоаналізу.

2. Основні теоретичні відомості

2.1. Загальні положення лінійного криптоаналізу

Лінійний криптоаналіз був розроблений та опублікований у 1992 році японським криптологом Міцуру Мацуі, який виявив певні слабкості у S-блоках алгоритму DES та запропонував використовувати їх для атаки на симетричні шифри.

Лінійний криптоаналіз розглядає залежності між бітами вхідного тексту, шифртексту та ключа. Нехай E – алгоритм шифрування і $Y = E_K(X)$. Трійка векторів α, β, γ називається *лінійною апроксимацією* шифру *E*, якщо виконується рівність

$$\forall X, Y, K: \quad \alpha \cdot X \oplus \beta \cdot Y \oplus \gamma \cdot K = const,$$

де $\cdot$ позначає скалярний добуток бітових векторів.

Кожна лінійна апроксимація, що виконується безумовно, дозволяє відновити один біт ключа. Дійсно, вираз $\gamma \cdot K$ – це просто сума деяких бітів ключа. Якщо відома одна пара *X* та *Y*, що відповідають одне одному, то з неї обчислюється значення цієї суми; отже, один біт ключа можна обчислити через значення інших бітів.

Звісно, найчастіше лінійна апроксимація буде виконуватись не безумовно, а з деякою імовірністю. В цьому випадку можна побудувати імовірнісний алгоритм відновлення біту ключа. Нехай $\Pr_{X,Y}\{\alpha \cdot X \oplus \beta \cdot Y \oplus \gamma \cdot K = 0\} = p > \frac{1}{2}$; тоді, накопичивши достатню кількість пар (X,Y), ми знайдемо значення суми $\gamma \cdot K = \alpha \cdot X \oplus \beta \cdot Y$. Якщо ж $p < \frac{1}{2}$, то тоді маємо $\Pr_{X,Y}\{\alpha \cdot X \oplus \beta \cdot Y \oplus \gamma \cdot K = 1\} = 1 - p > \frac{1}{2}$ і правильне значення суми будемо шукати як $\gamma \cdot K = \alpha \cdot X \oplus \beta \cdot Y \oplus 1$. Лише випадок, коли $p = \frac{1}{2}$, не дозволить нам визначити правильне значення $\gamma \cdot K$.

Зауважимо, що в описаній схемі атаки використовується одна лінійна апроксимація, яка дозволяє знаходити один біт ключа. Використання великої кількості (незалежних) лінійних апроксимацій дозволяє знаходити велику кількість бітів ключа. Однак даний підхід має той практичний недолік, що знаходити апроксимації, які поєднують біти відкритого та шифрованого текстів із бітами ключа, зазвичай дуже складно. Для того, щоб оминути цю проблему, Мацуі запропонував інший підхід, який використовує ідею лінійного криптоаналізу у загальній схемі атаки на ключ останнього раунду шифрування.

Для побудови та обґрунтування атаки нам знадобиться декілька визначень та понять з теорії імовірності.

Нехай $b \in \{0,1\}$ – випадкова величина, розподілена по Бернуллі із параметром *p*, тобто $\Pr\{b=1\} = p$. Тоді можна записати $p = \frac{1}{2} + s = \frac{1}{2}(1-\varepsilon)$; *s* називається *відхиленням* (*bias*) випадкової величини *b*, а ε – *кореляцією* випадкової величини *b*. Згідно визначення маємо:

$$\varepsilon = 1 - 2p = \Pr\{b=0\} - \Pr\{b=1\},$$

отже, якщо $\varepsilon > 0$, то *b* більше схильна до значення 0, а якщо $\varepsilon < 0$ – то до значення 1. Рівноімовірні величини мають кореляцію $\varepsilon = 0$.

Фундаментом для лінійного криптоаналізу є така лема.

Лема (Мацуї) (*про набігання знаків*[1]).

Нехай $X_1, X_2, ..., X_n$ – незалежні випадкові бітові величини, розподілені по Бернуллі, а ε_i – їх відповідні кореляції. Тоді бітова випадкова величина $X = X_1 \oplus X_2 \oplus ... \oplus X_n$ розподілена по Бернуллі із кореляцією $\varepsilon = \prod_{i=1}^{n} \varepsilon_i$.

Лема про набігання знаків дозволяє будувати лінійні апроксимації із високою очікуванною кореляцією для композицій булевих функцій. Дійсно, нехай $z = h_{k_1,k_2}(x) = g_{k_2}(f_{k_1}(x))$, де ключі k_1 та k_2 обираються незалежно один від одного. Якщо відомі лінійна апроксимація $\xi_f = \alpha \cdot x \oplus \beta \cdot y \oplus \gamma_1 \cdot k_1$ із кореляцією ε_1 (тут $y = f_{k_1}(x)$) та лінійна апроксимація $\xi_g = \beta \cdot y \oplus \lambda \cdot z \oplus \gamma_2 \cdot k_2$ із кореляцією ε_2, а інших апроксимацій із високою кореляцією ці функції не матимуть, то апроксимація $\xi_h = \xi_f \oplus \xi_g = \alpha \cdot x \oplus \lambda \cdot z \oplus \gamma_1 \cdot k_1 \oplus \gamma_2 \cdot k_2$ матиме очікувану кореляцію $\varepsilon_1 \varepsilon_2$. Точне значення кореляції буде визначатись іншими можливими апроксимаціями при довільних значеннях маски β.

Опишемо тепер схему атаки на окремий раундовий ключ блокового шифру. Через певну «антисиметрію» лінійного криптоаналізу у порівнянні з диференціальним, атака буде проводитись на раундовий ключ *першого* раунду.

Алгоритм M2.

0. Нехай для *r*-раундового шифру *E* відома лінійна апроксимація його останніх $r-1$ раундів $\alpha \cdot X_1 \oplus \beta \cdot X_r = 0$ із суттєвою кореляцією ε. Алгоритм дозволяє знайти істинне значення ключа першого раунду шифрування k_1.

1. Одержати *N* пар відкритих та шифрованих текстів (X, Y), де $Y = E_K(X)$.

2. Для кожного кандидата *k* у ключі k_1 виконати такі дії:

2.1. Зашифрувати відкриті тексти *X* на один раунд: $X_1 = F_1(X, k)$.

2.2. Обчислити значення

$$\hat{u}(k) = \left|\{(X_1, Y) : \alpha \cdot X_1 \oplus \beta \cdot Y = 0\}\right| - \left|\{(X_1, Y) : \alpha \cdot X_1 \oplus \beta \cdot Y = 1\}\right|.$$

3. Ключ k_1 визначається як $k_1 = \arg\max_k \left|\hat{u}(k)\right|$.

Пояснимо принцип роботи алгоритму M2.

Випадкова величина $\zeta = \alpha \cdot X_1 \oplus \beta \cdot Y$ для правильного значення ключа k_1 матиме математичне очікування $M\zeta = \frac{1}{2}(1 \pm \varepsilon)$ з міркувань, наведених вище. В той же час для

[1] В закордонній літературі ця лема відома під назвою "*pilling-up lemma*"

неправильного значення ключа k_1 замість «розшифрування» до правильного значення X_1 ми матимемо додаткове «шифрування» в інший бік на один раунд, тобто одержана величина ζ буде значно ближчою до рівноімовірної. Таким чином, алгоритм M2 розв'язує задачу розрізнення гіпотез щодо розподілу випадкової величини ζ, яке в одному випадку визначається через $M\zeta = \frac{1}{2}(1 \pm \varepsilon)$, а в іншому – через $M\zeta = \frac{1}{2}$.

Алгоритм M2 відноситься до класу атак за відомими відкритими текстами. Однак оцінка складності алгоритму виявилась непростою задачею (хоча на практиці він працював доволі ефективно), оскільки не завжди істинне значення ключа відповідає саме максимальному значенню лічильника $\hat{u}(k)$ або таких ключів виявлялось декілька, тому при реалізації атаки зазвичай розглядаються декілька можливих кандидатів у ключі k_{r+1}, які мають великі значення $\hat{u}(k)$. Імовірність успіху алгоритму M2 (за деяких припущень) визначається наступною теоремою, яку ми наводимо без доведення.

Якщо оригінальна довжина ключа шифрування складала m бітів, а після застосування алгоритму M2 залишилось 2^t кандидатів, то будемо казати, що алгоритм M2 має $(m-t)$-*бітову перевагу*.

Теорема (Сельчук).

Нехай лінійні апроксимації шифру для кожного значення ключа є незалежними. Якщо алгоритм M2 використовує лінійну апроксимацію із кореляцією ε та має при цьому перевагу a, то імовірність його успіху (для достатньо великих N та довжини ключа) приблизно дорівнює $\Phi\left(\varepsilon\sqrt{N} - \Phi^{-1}\left(1 - \frac{1}{2^{a+1}}\right)\right)$.

З теореми Сельчука випливає, що складність проведення атаки за алгоритмом M2 також визначається величиною $O(\varepsilon^{-2})$. Таким чином, можна вважати квадрат кореляції лінійної апроксимації основним параметром, що визначає складність проведення лінійного криптоаналізу.

Втім необхідно зауважити, що два припущення, які зазвичай використовуються у лінійному криптоаналізі, а саме припущення про незалежність раундів шифрування (для застосування леми про набігання знаків) та припущення про незалежність лінійних апроксимацій при різних значеннях ключа шифрування (для оцінки складності алгоритму M2) виконуються не завжди. Наприклад, було доведено, що ці припущення невірні для алгоритму шифрування RC5. В той же час, для багатьох класичних алгоритмів шифрування, в першу чергу DES-подібних та AES-подібних, такі припущення дають адекватні результати на практиці. Наразі можемо стверджувати, що перевірка наведених гіпотез перед застосуванням лінійного криптоаналізу обов'язково потрібна у випадку, коли алгоритм шифрування для замішування із ключем та перетворення даних окрім побітового додавання використовує різні алгебраїчні операції (модульне додавання, множення, керовані циклічні зсуви тощо)

2.2. Лінійні апроксимації та лінійні потенціали

Розглянемо булеву функцію $f : V_n \to V_n$.

Лінійна апроксимація булевої функції f – це довільна пара двійкових векторів (α, β) з V_n, з якою пов'язується випадкова величина $\alpha \cdot x \oplus \beta \cdot f(x)$.

Коефіцієнтом кореляції лінійної апроксимації булевої функції є величина

$$C_f(\alpha, \beta) = \frac{1}{2^n} \sum_{x \in V_n} (-1)^{\alpha \cdot x \oplus \beta \cdot f(x)},$$

а *лінійним потенціалом* апроксимації булевої функції – величина

$$LP^f(\alpha,\beta) = (C_f(\alpha,\beta))^2 = \left(\frac{1}{2^n}\sum_{x\in V_n}(-1)^{\alpha\cdot x\oplus\beta\cdot f(x)}\right)^2.$$

Для лінійних потенціалів виконується багато властивостей, аналогічних властивостям диференціальних імовірностей. Зокрема:

1) $LP^f(\alpha, 0) = [\alpha = 0]$.

2) Якщо f – бієктивна функція, то $LP^f(\alpha,\beta) = LP^{f^{-1}}(\beta,\alpha)$.

3) Якщо f – бієктивна функція, то $LP^f(0, \beta) = [\beta = 0]$.

4) $\sum_{\alpha} LP^f(\alpha,\beta) = 1$.

5) Якщо f – бієктивна функція, то $\sum_{\beta} LP^f(\alpha,\beta) = 1$.

Наявність такої подібності дозволяє будувати формальну теорію лінійного криптоаналізу на тих самих засадах, що й формальну теорію диференціального криптоаналізу.

Для лінійних апроксимацій шифруючого перетворення $f_k : V_n \times K \to V_n$ розглядається *усереднений лінійний потенціал* $ELP^{f_k}(\alpha,\beta) = \sum_k LP^{f_k}(a,b)$. Ця величина фактично виступає параметром стійкості до лінійного криптоаналізу. К. Ніберг показала, що усереднені потенціали для апроксимацій виду $\alpha\cdot X \oplus \beta\cdot f_k(X) \oplus \gamma\cdot K$ не залежать від значення γ та також дорівнюють $ELP^{f_k}(\alpha,\beta)$.

Нехай E – ітеративний блочний шифр, що складається з послідовних раундових перетворень $f_{k_1}^{(1)}$, $f_{k_2}^{(2)}$, …, $f_{k_r}^{(r)}$. Раундові ключі k_i вважаються незалежними та рівномірно розподіленими.

Лінійна характеристика шифру E – послідовність бітових векторів $\Omega = (\omega_0, \omega_1, ..., \omega_r)$, де всі $\omega_i \in V_n \setminus \{0\}$. Лінійна характеристика розглядається як послідовність лінійних апроксимацій раундів шифрування: $\omega_0\cdot X_0 \oplus \omega_1\cdot X_1$ на першому раунді, $\omega_1\cdot X_1 \oplus \omega_2\cdot X_2$ на другому тощо. *Усереднений лінійний потенціал* лінійної характеристики формально визначається як

$$ELCP^E(\Omega) = \prod_{i=1}^{r} ELP^{f_{k_i}^{(i)}}(\omega_{i-1},\omega_i).$$

Коректність такого означення випливає з леми про набігання знаків.

Позначимо через $\Omega(a,b)$ множину таких лінійних характеристик Ω, в яких $\omega_0 = a$, $\omega_r = b$. Апроксимацію (a,b) будемо називати *обвідною апроксимацією* лінійної характеристики $\Omega \in \Omega(a,b)$, а таку характеристику будемо називати *вкладеною* для апроксимації (a,b). На відміну від диференціальних імовірностей, для лінійних потенціалів їх зв'язок із потенціалами вкладених лінійних характеристик зовсім не очевидний. Але для окремих класів ітеративних шифрів це питання досліджено в повній мірі. Зокрема, має місце така теорема.

Теорема. Якщо в ітеративному шифрі E раундові перетворення мають вид $f_k^{(i)}(x) = g_i(x\oplus k)$, де $g_i : V_n \to V_n$ – деякі безключові булеві відображення, то

$$ELP^E(a,b) = \sum_{\Omega \in \Omega(a,b)} ELCP^E(\Omega, x) = \sum_{\Omega \in \Omega(a,b)} \prod_{i=1}^{r} ELP^{g_i}(\omega_{i-1}, \omega_i).$$

Це твердження дозволяє будувати атаки та оцінювати їх складність так само, як в диференціальному криптоаналізі. Зокрема, коректно працює метод гілок та границь для пошуку характеристик та апроксимацій з високими потенціалами, складність аналізу (кількість необхідних пар відкритих текстів та шифротекстів) дійсно визначається величиною, оберненою до $ELP^E(a,b)$ і таке інше.

3. Порядок і рекомендації щодо виконання роботи

1. Узяти реалізацію шифру Хейса із комп'ютерного практикуму №1 (із таким само варіантом).

2. Реалізувати методом «гілок та границь» пошук п'ятираундових лінійних апроксимацій шифру Хейса із великим потенціалом. Так само, як і в попередньому практикумі, для пошуку рекомендується використовувати початкові маски α із однією ненульовою тетрадою. Для виконання практикуму вам знадобиться 300-700 різних апроксимацій.

3. Реалізувати атаку на перший раундовий ключ шифру Хейса за такою схемою.

а) Одержати необхідну кількість пар «відкритий текст-шифротекст». Зауважимо, що кількість пар повинна бути обернено пропорційна до найменшого лінійного потенціалу серед усіх апроксимацій, які використовуються для атаки (краще за все із коефіцієнтом 8 або 16).

б) Для кожної апроксимації реалізувати алгоритм атаки M2. Відмітити кожний кандидат у ключі, для якого лічильник алгоритму M2 перевищив певний поріг (значення цього порогу залишається на ваш розсуд; наприклад, ви можете розглядати першу десятку, перші 50 або перші 100 ключів).

в) Серед усіх відмічених кандидатів для усіх знайдених апроксимацій обрати десять, які найчастіше обирались алгоритмом M2.

Атака вважається успішною, якщо правильний раундовий ключ потрапив у фінальну десятку кандидатів.

Необхідний статистичний матеріал (шифровані тексти) одержується із тестової програми **Heys.exe**, що додається.

Зауваження. Програма **Heys.exe** має консольний інтерфейс.

4. Оформити звіт з практикуму.

4. Оформлення звіту

Звіт до комп'ютерного практикуму оформлюється згідно зі стандартними правилами оформлення наукових робіт, за такими винятками:

- дозволяється використовувати шрифт Times New Roman 12pt та одинарний інтервал між рядками;
- для оформлення текстів програм дозволяється використовувати шрифт Courier New 10pt (8pt) та друкувати тексти в дві колонки;
- дозволяється не починати нові розділи з окремої сторінки.

До звіту можна не включати анотацію, перелік термінів та позначень та перелік використаних джерел. Також не обов'язково оформлювати зміст.

Звіт має містити:

– мету комп’ютерного практикуму;

– постановку задачі;

– хід роботи, опис труднощів, що виникали, та шляхів їх розв’язання;

– опис методу пошуку лінійних апроксимацій із великими потенціалами, обрані порогові значення потенціалів (із обґрунтуванням вибору);

– таблицю лінійних апроксимацій S-блоку вашого варіанту;

– знайдені за допомогою методу «гілок та границь» лінійні апроксимації для кожного раунду шифрування та їх потенціали (якщо перелік відповідних апроксимацій завеликий, дозволяється обмежитись певною вибіркою значень);

– знайдені в ході атаки кандидати у ключі першого раунду шифрування тестової програми (перша десятка), із зазначенням кількості шифртекстів, що були потрібні для знаходження;

– висновки до роботи.

5. Контрольні запитання

Дивіться лекції

6. Оцінювання комп’ютерного практикуму

За виконання комп’ютерного практикуму студент може одержати до 10 рейтингових балів; зокрема, оцінюються такі позиції:

- реалізація програм – до 5-х балів (в залежності від правильності та швидкодії);
- теоретичний захист роботи – до 5-ти балів;
- несвоєчасне виконання роботи – (-1) бал за кожні два тижні пропуску.

7. Рекомендовані джерела

1. Heys Howard M. A Tutorial on Linear and Differential Cryptanalysis [електронний ресурс] / Howard M. Heys. – Режим доступу : http://www.engr.mun.ca/~howard/PAPERS/ldc_tutorial.pdf