

МОДЕЛЮВАННЯ ТА КРИПТОАНАЛІЗ ФУНКЦІОНУВАННЯ ПОТОКОВИХ ШИФРІВ З ПЕРЕКРИТТЯМ ВІДРІЗКІВ ГАММИ

Н. А. Борис^{1, а}, М. М. Савчук^{1, б}

¹ Національний технічний університет України «Київський політехнічний інститут»

Анотація

В роботі проводиться дослідження математичної моделі функціонування поточкових шифрів гаммування, оцінюються ймовірності виникнення загроз накладання однакових відрізків гамми на відкриті тексти. Розглядаються варіанти методу безключового читання як методу відновлення за різницею шифрованих повідомлень відкритих текстів у випадку повторного використання відрізків ключової послідовності для шифрування. Наведено результати статистичного моделювання та експериментальних досліджень.

Ключові слова: поточковий шифр гаммування, гамма шифру, комбінаторно-ймовірнісна модель функціонування поточкових шифрів, безключове читання, статистичне моделювання.

Вступ

При сучасному швидкому зростанні темпу обігу та об'єму інформації в усіх сферах людської діяльності немає необхідності зазначати важливість використання ефективних методів захисту інформації, застосування стійких криптосистем, надійних режимів та протоколів шифрування. У даній роботі наведено результати дослідження роботи симетричних поточкових систем шифрування при інтенсивному використанні і можливого, так званого, перекриття відрізків гамми. Як відомо, поточкові криптосистеми, у разі коректної реалізації, мають ряд переваг, зокрема: низькі вимоги до затрат пам'яті, можливість ефективного використання для даних, об'єм яких не є відомим заздалегідь, більша швидкість у порівнянні з блоковими криптосистемами.

Зважаючи на всі переваги, а отже і широкий спектр використання поточкових систем шифрування, зрозуміла необхідність ретельного дослідження стійкості таких систем, режимів їх роботи до різних методів криптоаналізу та визначення умов їх правильного використання, невиконання яких може призвести як до спотворення змісту зашифрованого повідомлення, так і до збільшення ймовірності отримання інформації злоумисником.

У роботі проводиться дослідження вразливостей поведінки поточкових шифрів при можливому повторному застосуванні однакових відрізків гамми шифру. Наявність такого роду вразливості дає змогу отримання інформації про відкриті повідомлення без використання безпосередньої атаки на сам поточковий шифр, що значно полегшує криптоаналіз як з боку технічної бази, через можливість досягнення успіху при використанні менш потужного обладнання, так і ресурсів, витрачених на розробку та реалізацію атаки, завдяки можливості отримання додаткової

інформації про відкритий текст значно простішим способом.

У ході дослідження ставляться наступні задачі:

- дослідження математичної моделі функціонування симетричних поточкових шифрів гаммування при інтенсивному використанні,
- програмна реалізація запропонованої комбінаторно-ймовірнісної моделі,
- верифікація моделі та знаходження її основних статистичних характеристик,
- аналіз існуючих варіантів криптоаналізу для виявлення наявності перекриття гамм та відновлення відповідних відкритих текстів,
- створення та програмна реалізація алгоритму відновлення відкритих текстів при наявності перекриття ключової гамми.

1. Схема розміщень комплектів як модель функціонування поточкового шифру гаммування

1.1. Принцип роботи поточкових шифрів гаммування

Поточковим шифром є симетричний шифр, в якому кожен символ тексту перетворюється в символ шифротексту в залежності від ключа ключа, символу відкритого тексту, його позиції в потоці відкритого тексту та, можливо, попередніх символів. В симетричних поточкових шифрах гаммування символ шифротексту залежить тільки від ключа, символу відкритого тексту та його позиції, зокрема, шифрування здійснюється шляхом накладання (у двійковому випадку – XOR) гамми шифру на послідовність відкритого тексту, в результаті отримуємо шифрований текст. Операція шифрування, реалізована таким чином, називається гаммуванням. Для виконання операції гаммування довжина відрізка ключової послідовності повинна дорівнювати довжині відкритого тексту,

^аninaborys5@gmail.com

^бmikhail.savchuk@gmail.com

який необхідно зашифрувати. Зазвичай, формується псевдовипадкова ключова послідовність, що створюється на основі секретного ключа невеликого розміру шляхом його перетворення алгоритмом потокового шифру на гамму. Ключова послідовність повинна відповідати певним вимогам, зокрема, необхідно використання стійкого генератора псевдовипадкової послідовності (шифру) при її створенні. Якщо два шифрованих повідомлення отримані накладанням однакового відрізка гамми, то по двом таким криптограммам можливо відновити обидва відкритих тексти.

У даній роботі предметом дослідження є математичні моделі, що дають змогу оцінити ймовірності перекриття відрізків гамм в різних ситуаціях, та методи криптоаналізу поточкових шифрів при перекритті гамм.

1.2. Схема розміщення комплектів елементів на колі

Для визначення характеристик відрізків перекриттів гамм, що виникли при використанні однієї і тієї ж шифруючої послідовності для різних повідомлень, доцільно використовувати альтернативну постановку задачі, а саме розглядати процес шифрування як схему розміщення комплектів елементів на колі. [1].

Нехай N пронумерованих комірок розташовані на колі послідовно так, що за i -ю коміркою лежить комірка з номером $j \equiv i + 1(\text{mod } N)$, а перед нею розташована комірка $j \equiv i - 1(\text{mod } N)$, $i = \overline{0, N-1}$.

В комірки розміщуємо n комплектів по m елементів в кожному так, щоб в кожному комплекті елементи з однаковою ймовірністю займають будь-які m позицій підряд, а події, що полягають в розміщенні елементів різних комплектів є незалежними. Вважатимемо, що два комплекти перетинаються, якщо існує принаймні одна комірка, що містить елементи обох комплектів.

Для такого представлення коло, з розташованому на ньому комірками, співвідноситься з періодичною гаммою, де її період відповідає повному проходженню по колу. Тоді розміщення комплектів відповідає накладанню n відкритих текстів по m символів на гамму, а перетин комплектів - використанню однакової гамми для шифрування двох різних відкритих текстів.

Нехай $\eta_r(M, m, n)$ — випадкова величина, що дорівнює числу комплектів, що мають перетин рівно із r іншими.

1.3. Верифікація моделі розміщень

Для перевірки отриманої програмної реалізації схеми розміщень та визначення областей зміни параметрів моделі, в яких асимптотичні оцінки достатньо якісні, використовуються наступні результати:

- 1) Значення математичного сподівання випадкової величини

Нехай $n, m \rightarrow \infty, \alpha \rightarrow 0, r \geq 0$

$$\alpha = mn/N$$

$$\gamma_r = n(2\alpha)^r / r!$$

Тоді

$$M\eta_r = \gamma_r (1 + O(\frac{\delta_{r0}}{\min(m, n)} + \alpha))$$

2) Теорема 1. Нехай

$$\zeta_r = \frac{\eta_r - \gamma_r}{\sqrt{\gamma_r(3 - 2^{1-r})}},$$

Якщо $N, n, m \rightarrow \infty$ та при деякому $r \geq 1$, $\gamma_r \rightarrow \infty$, а $\gamma_r(\alpha + m^{-1}) \rightarrow 0$, то випадкова величина ζ має в якості граничного розподілу нормальний розподіл з параметрами $(0, 1)$, тобто $\zeta_r \sim N(0, 1)$.

3) Теорема 2. Якщо $N, n, m \rightarrow \infty$ і при деякому цілому $r \geq 1$, $\gamma_r \rightarrow \gamma (0 < \gamma < \infty)$, то розподіл випадкової величини η_r збігається до розподілу з генератрисою

$$P(x, \gamma) = \exp\left\{\frac{\gamma}{2} \sum_{k=1}^{r-1} 2^{-k}(x^k - 1) + \gamma 2^{-r}(x^{r+1} - 1)\right\}$$

Граничний розподіл величини η_r при цьому є розподілом величини

$$\xi(\gamma) = \sum_{k=1}^{r-1} k \xi_k\left(\frac{\gamma}{2^{k+1}}\right) + (r+1) \xi_{r+1}\left(\frac{\gamma}{2^r}\right),$$

де випадкові величини $\xi_k(a_k)$ незалежні та мають розподіл Пуассона з параметром a_k .

Для перевірки коректності програмної реалізації схеми розміщень та визначення областей зміни параметрів моделі, в яких асимптотичні оцінки достатньо коректні, за допомогою статистичного моделювання проводиться серія експериментів на ЕОМ: генерується вибірка достатнього об'єму та здійснюється перевірка виконання вказаних теорем за допомогою методів статистичного аналізу, а саме за допомогою критерію згоди Колмогорова-Смірнова та критерію χ^2 -Пірсона.

2. Відновлення відкритого тексту за наявності перекриття ключової послідовності

Іншим напрямком дослідження є створення алгоритму відновлення відкритих текстів за умови наявності перекриття відрізків шифруючої гамми.

Вважатимемо, що символи гамми утворюють послідовність незалежних, рівноймовірно розподілених випадкових величин, а на множині відкритих текстів задано певний ймовірнісний розподіл. Неважко показати, що за умови, якщо шифруюча гамма була використана однакою, то, при виконанні умови рівноймовірності гамми, отримуємо різницю двох шифртекстів з нерівноймовірним розподілом літер,

а при різних гаммах різницю двох шифртекстів з рівноймовірним розподілом літер [2].

Отже, статистичний критерій, що перевіряє правильність гіпотези про рівноймовірний розподіл різниці наявних криптограм, буде однознано виявляти наявність чи відсутність перекриття шифруючих гамм у досліджуваних відрізках.

Нехай алфавіт $Z_n = \{0, 1, \dots, n-1\}$ - єдиний для відкритих текстів, гамми та шифртекстів, а для криптоаналізу наявні дві криптограми довжини t , про які відомо, що вони були отриманні при накладанні однакових відрізків гамми на два різних відкриті тексти.

В такому випадку отримуємо послідовність, що дорівнює посимвольній різниці відомих шифртекстів, а отже, є різницею відповідних відкритих текстів.

$$S = \{s_i\} = \{x_i - x'_i\} = \{y_i - y'_i\}, i = \overline{0, t-1} \quad (1)$$

Тепер задачу можна переформулювати наступним чином: необхідно підібрати пару відкритих текстів, посимвольна різниця яких відповідатиме відомій послідовності S . При цьому припустимо, що відкриті тексти є текстами певної природної мови, а отже вважаємо, що розклад послідовності в різницю двох осмислених текстів є однозначним.

1) Метод протягування ймовірного слова

В основі методу лежить припущення, що одне з ймовірних слів мови довжини l зустрічається на початку першої криптограми. В такому випадку можна однозначно відновити l відповідних літер другої криптограми. При достатній довжині l легко встановити читабельність отриманого відрізка другої криптограми, а отже і правильність гіпотези про наявність даного ймовірного слова на вказаному місці першого шифртексту. У випадку вірності даної гіпотези, продовжуємо цю послідовність дій, доки це буде можливо, намагаючись відновити якомога більший уривок відкритих текстів. Коли подальше розшифрування тексту стає неможливим, просовуємо початок ймовірного слова на наступну позицію криптограми і повторюємо вказані дії. По закінченню аналізу з цим словом, таку ж процедуру можна виконувати і для наступних ймовірних слів [3].

2) Метод читання в колонках

Для реалізації цього методу будемо таблицю, в якій для кожної літери, отриманої в послідовності S , ставимо у відповідність n пар літер, різниця яких утворює необхідну літеру. Далі необхідно обирати з кожної колонки по одній парі літер так, щоб отримати два читабельні тексти. Таким чином, виходячи з припущення про належність відкритих текстів до певної природної мови, отримуємо можливість відновити відкриті тексти, використовуючи статистичний аналіз мови.

Розглянемо модифікації методу читання в колонках, реалізовані в ході дослідження.

Зважаючи на припущення, що відрізок, на якому різні відкриті тексти накладуються на однакову гамму уже виявлено, процес отримання криптограм з відкритих текстів у ході дослідження не розглядає-

ться. Тому в якості вхідних даних використовуємо послідовність S (1).

Першим етапом є частотний аналіз досліджуваної мови, а саме отримання статистики біграм та ймовірностей їх появи у тексті. Далі формується таблиця пар всіх можливих варіантів розкладу кожної літери з алфавіту в різницю двох інших: $(i, j) : k = i - j(\text{mod } N)$, $i, j = \overline{0, N-1}$.

Далі відбувається читання в колонках відкритих текстів. Однак, враховуючи, що для тексту довжиною t кількість можливих варіантів становить N^t для кожного з текстів, розташування літер у кожній з колонок необхідно оптимізувати, використовуючи отримані статистичні властивості мови.

У роботі розглядаються варіанти розташування пар в колонках в порядку спадання сумарної ймовірності букв з пари та в порядку спадання добутку ймовірностей літер. Кожній літері з послідовності S ставиться у відповідність стовпчик із впорядкованої таблиці. Для подальшого аналізу відбирається лише певна кількість перших - найбільш ймовірних пар, оптимальну для отримання максимально вдалого результату розшифрування відкритих текстів.

Для переходу між колонками використовується один з двох підходів: по поодиначному максимуму біграм з відкритих текстів або по максимуму добутку ймовірностей біграми переходу з i -ї в j -ту позицію першого відкритого тексту на відповідну біграму переходу з i -ї в j -ту позицію другого.

Повним перебором стовпчиків визначається перехід між першим та другим символом, а далі, виходячи з обраної позиції, розглядаються лише варіанти з кожного наступного стовпця.

Можливе покращення вказаного алгоритму за рахунок використання інших статистичних відомостей про досліджувану мову, наприклад статистики триграм.

3. Результати дослідження

3.1. Верифікація реалізації схеми розміщення елементів на колі

Результати тесту на значення математичного сподівання випадкової величини η_r (табл. 1) (позначення N, m, n та η_r з пункту 1.2), а також відповідні значення середньоквадратичного відхилення:

Табл. 1. Результати для $M\eta_r$

(N,m,n)	Epected		Actual		StDev	
	r=1	r=2	r=1	r=2	r=1	r=2
(10 ⁶ ,500,200)	100	10	98	8	1.41	1.41
(10 ⁹ ,10 ⁴ ,10 ³)	200	2	196	3	2.83	0.71
(10 ¹² ,10 ⁵ ,10 ⁴)	200	0.2	203	0.7	2.12	0.35

Нижче наведено результати проходження статистичного критерію згоди Колмогорова-Смірнова та критерію χ^2 для теореми про граничний стандартний нормальний розподіл 1 (табл. 2) та для теореми про розподіл суміші пуассонівських величин 2 (табл. 3). В таблицях наведені значення відповідних статистик та критичні порогові значення критеріїв.

Табл. 2. Результати для Теореми 1

(N,m,n)	K-S test		χ^2 test	
	r=1	r=2	r=1	r=2
(10 ⁶ ,500,200)	5.702	2.067	340.02	95.986
(10 ⁹ ,10 ⁴ ,10 ³)	1.657	2.344	8.365	50.7
(10 ¹² ,10 ⁵ ,10 ⁴)	0.415	0.867	1.951	25.326
Critical	0.895		82.358	

Табл. 3. Результати для Теореми 2

(N,m,n)	K-S test		χ^2 test	
	r=1	r=2	r=1	r=2
(10 ⁶ ,500,200)	5.289	1.137	171.98	91.242
(10 ⁹ ,10 ⁴ ,10 ³)	0.891	0.679	7.176	79.2
(10 ¹² ,10 ⁵ ,10 ⁴)	0.694	0.099	2.779	81.25
Critical	0.895		82.358	

Виконання першого етапу перевірок моделі показало відсутність випадку перетину трьох або більше комплектів елементів для діапазонів стартових значень, що розглядаються в роботі.

Результати проходження статистичних тестів для моделі, що використовувала значення періоду $N = 10^6$, показали відсутність успішного проходження жодного тесту. Починаючи зі значення періоду в $N = 10^9$ модель проходить тести для теореми 2. Успішне проходження усіх тестів отримано для періоду $N = 10^{12}$, отже для області зміни параметрів моделі, що відповідає таким або більшим значенням періоду, кількості та довжини комплектів даної комбінаторно-ймовірнісної моделі, асимптотичні оцінки можна вважати достатньо якісними.

3.2. Відновлення різниці відкритих текстів на основі методу безключового читання

По завершенню реалізації процедури відновлення відкритих текстів шляхом читання у колонках, отримано ряд результатів, зокрема статистика індексів

входження правильних варіантів розкладу літери з результуючої різниці шифртекстів до стовпчика усіх можливих варіантів розкладу. Ця частина дослідження показала, що більш ефективним є використання впорядкування пар у колонці у порядку спадання ймовірності добутку літер, а також дозволила визначити оптимальну кількість пар, необхідну для розгляду при аналізі, для отримання максимально вдалого варіанту розшифрування. Для розглянутих текстів довжиною 37 символів вона становила 7 пар - близько 18% довжини тексту. У такому випадку в якості найбільш вдалого розшифрування тексту можемо отримати 25 його літер, що становить 67.6% тексту.

Наступним етапом дослідження була процедура відновлення символів відкритого тексту з послідовності S (1). Варіант переходу по максимуму з біграм максимальної ймовірності для припущень виявився неефективним, вірно підбравши 3 літери, що становить 12% від максимально вдалого варіанту. Інший варіант переходу - по максимуму добутку ймовірностей біграм-припущень. Для нього вдалося відновити 9 літер, що становить 36% від максимально вдалого варіанту розшифрування. Варто зазначити, що відносне значення результатів залишається постійним для вхідних текстів різної довжини.

Висновки

В даній роботі досліджено модель функціонування поточкових шифрів гаммування, отримано оцінки ймовірностей виникнення загроз накладання однакових відрізків гамми на відкриті тексти. Проведено реалізацію та верифікацію комбінаторно-ймовірнісної моделі функціонування поточкових шифрів - схеми розміщення комплектів елементів на колі. Знайдено області коректного використання наведених у роботі асимптотичних оцінок ймовірностей і розподілів різних перетинів відрізків гамм.

Іншим напрямком дослідження є спроба відновлення відкритих текстів за їх різницею при припущенні, що відрізок перекриття послідовності гамми уже було виявлено. На основі отриманих результатів можна сказати, що використання статистики біграм природної мови виявилось недостатньо для відновлення відкритих текстів до читабельного варіанту. Проводиться робота по вдосконаленню запропонованих алгоритмів.

Перелік використаних джерел

1. Коваленко И. Н., Левитская А. А., Савчук М. Н. Избранные задачи вероятностной комбинаторики. — Киев: Наукова думка, 1986. — С. 84 – 90, 106 – 111 с.
2. Грушо А. А., Применко Э. А., Тимонина Е. Е. Анализ и синтез криптоалгоритмов. — Йошкар-Ола, изд. МФ МОСУ, 2000. — С. 26 – 35 с.
3. Основы криптографии. / А. П. Альферов, А. Ю. Зубов, А. С. Кузьмин, А. В. Черемушкин. — Москва: Гелиос АРВ, 2002. — С. 126 – 155 с.