УДК 003.26:004.056.55
С 17

МЕТОДИКА ВИЗНАЧЕННЯ ДОВЖИНИ ПАРОЛЯ ШИФРІВ ПОЛІАБЕТКОВОЇ ЗАМІНИ

Д. М. Самойленко, канд. фіз.-мат. наук, доц.;
В. В. Дронов, студ.

Національний університет кораблебудування, м. Миколаїв

Анотація. Розроблено методику визначення прихованої періодичності у послідовностях на основі модифікованої функції автокореляції. Методика випробувана на задачах визначення довжини пароля у шифрах поліабеткової заміни.
Ключові слова: криптоаналіз, автокореляція, періодичність.

Аннотация. Разработана методика определения скрытой периодичности в последовательностях на основе модифицированной функции автокорреляции. Методика испытана на задачах определения длины пароля в шифрах полиалфавитной замены.
Ключевые слова: криптоанализ, автокорреляция, периодичность.

Abstract. The hidden periodicity method of sequences determination based on the modified autocorrelation function has been developed. The method was tested in the problem of password length determination in polyalphabetic ciphers.
Keywords: cryptanalysis, autocorrelation, periodicity.

ПОСТАНОВКА ПРОБЛЕМИ

Останнім часом сфера застосування криптографії розширилася і включає не тільки передачу повідомлень, але і методи перевірки їхньої цілісності, ідентифікацію чи автентифікацію відправника та одержувача, цифрові підписи, інтерактивні підтвердження та технології безпечного спілкування тощо.

Для захисту інформації, яка має високу цінність, використовуються сучасні криптографічні алгоритми. У той же час для менш цінної інформації та у задачах, для яких критичним є час перетворень, можуть використовуватися простіші, проте значно швидші у реалізації алгоритми, наприклад алгоритми поліабеткової заміни. У статті розглянемо криптоаналіз зазначених шифрів поліабеткової заміни.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

З огляду на принципи побудови сучасних криптографічних систем (відомих також як принципи Керкгоффса) головна складність має визначатися паролем, за допомогою якого здійснюється перетворення [5]. Решту технічних, математичних та інших параметрів системи слід вважати відомими. Відповідно, головною задачею криптоаналізу є встановлення секретного пароля.

Методи криптоаналізу шифрів простої (моноабеткової) заміни відпрацьовані досить добре. Наявна достатня кількість публікацій та оглядів щодо зазначених методів [3]. У той же час криптоаналіз поліабеткових шифрів розглянуто значно менше. Переважна більшість методів аналізу зазначених шифрів ґрунтується на методах Казинські (1863 р.) та Фрідмана (поч. XX ст.) [1]. Слід зазначити, що мова йде про універсальні автоматизовані системи криптоаналізу, оскільки існує ряд методів, які ґрунтуються на відомостях про певні особливості відкритого тексту і, відповідно, ці методики є застосовними лише до шифротекстів, що мають такі особливості. Проте навіть і у цьому разі системи не є повністю автоматизованими і вимагають експертного втручання криптоаналітика.

Усі методи криптоаналізу моноабеткових шифрів можуть бути застосовані для поліабеткових шифрів за умови відомої довжини пароля [1, 2]. Так, при побудові криптосистем для поліабеткових шифрів слід передбачати, що довжина пароля може бути довільною і бажано секретною. А однією із задач криптоаналізу поліабеткових шифрів постає задача визначення довжини пароля, яка може інтерпретуватися як задача приведення поліабеткових шифрів до моноабеткових.

МЕТОЮ СТАТТІ є розвинення та випробовування методики автоматизованого визначення довжини пароля шифрів поліабеткової заміни з використанням кореляційних методів.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Довжина пароля визначає періодичність, з якою у шифротексті змішані результати моноабеткової заміни. Саме на дослідженні періодичності шифротексту базуються методи Казинські та Фрідмана [1].

Метод Казинські полягає у пошуку однакових фрагментів у шифротексті та встановленні чисел, кратних відстані між цими фрагментами. Метод має певну внутрішню суперечливість:

– якщо довжина фрагментів невелика (2–3 символи), то цілком імовірні випадкові збіги, безпосередньо не пов'язані з довжиною пароля;

– якщо довжина фрагментів велика, то їх кількість значно зменшується і, у кращому разі, вдається встановити числа, кратні довжині пароля. Виникає необхідність додаткового перебору всіх кратних чисел.

Як правило, приклади використання методу прив'язані до конкретного тексту. При цьому не гарантується, що метод може бути використаний для подібних фрагментів такої ж довжини в інших шифротекстах. Загальних рекомендацій до встановлення довжини фрагментів, а також до виду цих фрагментів не наводиться, що значно ускладнює процес автоматизації методу Казинські.

Група методів Фрідмана [1] базується на обчисленні індексу збігів, величина якого відповідає ймовірності збігу двох випадково взятих символів шифротексту. Сутність методу полягає в аналізі статистичних особливостей, пов'язаних з надлишковістю інформації у мовах. Так, імовірність випадкового збігу двох символів з повністю випадкового тексту становитиме 1/30 ≈ 0,033 (якщо в абетці 30 літер), проте для англійської мови цей індекс становить 0,0662, для російської – 0,0529.

Практичне використання методів Фрідмана полягає у підрахунку індексу збігів для вибірок із шифротексту з різними періодами. Величини періодів, для яких спостерігається збільшення індексу збігів, будуть кратними довжині пароля.

Метод може бути легко автоматизований, проте його недоліком є обмеження на довжину абетки. Вважається, що шифротекст складається з абетки тієї ж довжини, що і початковий текст. Тобто передбачається афінне шифрування [4] з модулем, рівним довжині абетки.

При використанні сучасної обчислювальної техніки набагато швидше та ефективніше виконувати операції логічної арифметики, які відповідають розрядності процесора. Навіть для відносно застарілих процесорів з розрядністю 8 біт розмірність вихідної абетки становитиме 256 символів. На фоні абетки такої довжини статистичні особливості тридцяти з її символів значно зменшуються.

Запропонуємо комбіновану методику для визначення довжини пароля, яка грунтується на використанні функції, подібної до функції автокореляції. Будемо називати функцією автокореляції шифротексту $K(s)$ результат, розрахований за формулою

$$K(s) = \sum_{i=1}^{N} \delta\big(T(i) \oplus T(i+s), 0\big),$$

де i – номер символу шифротексту; N – довжина шифротексту (кількість символів); $T(i)$ – символ з номером i (поточний символ); $T(i + s)$ – символ, що знаходиться на відстані s від поточного (s – величина зсуву); δ – дельта-функція Дірака; $\oplus$ – операція «виключне або» (XOR).

Принцип визначення періоду появи однакових символів базується на властивості операції XOR, яка полягає у тому, що

$$\forall x : x \oplus x = 0,$$

тобто за умови, що на відстані s один від одного знаходяться однакові символи, результат операції $T(i) \oplus T(i + s)$ дорівнює нулю.

Дельта-функція передбачає порівняння обчисленого результату з нулем. За умови рівності значення функції беруть рівним одиниці. У протилежному випадку значення функції є нульовим.

Ураховуючи вказані особливості, значенням функції автокореляції шифротексту $K(s)$ буде кількість однакових символів у шифротексті, які знаходяться на відстані s один від одного.

Також слід зазначити, що у разі перевищення результату ($i + s$) довжини шифротексту обирається символ з початку шифротексту, тобто для практичного використання зсув розраховується як $i + s \pmod N$, що еквівалентно циклічному зсуву.

За умови наявності у шифротексті однакових символів з періодом s очікується зростання значення функції автокореляції шифротексту для даного s. При цьому у функції враховуються всі символи шифротексту, тобто не робиться жодних припущень щодо виду символів, повторення яких очікується.

У разі відсутності періодично повторюваних символів величина функції автокореляції шифротексту має відповідати статистичним збігам, обмежуваним розмірністю шифроабетки.

Випробування методики полягало у дослідженні поведінки функції автокореляції шифротексту для файлів різного типу. Як об'єкт дослідження обрано матеріал статті [6] та його переклад, які подано у різних форматах: текстового процесора MS Word (DOC), форматі переносних документів (PDF) та текстовому форматі (TXT). Додатково досліджено архів зазначеного об'єкта (RAR). Результати випробування наведені на рис. 1–3.

На рис. 1 наведено функції автокореляції шифротексту $K(s)$ для файлів текстового процесора MS Word. Вміст файла крім тексту включав формули, рисунки, засоби форматування. Кількість текстових символів в англомовному варіанті – 22 136, усього символів (розмір файла) – 274 944. В українському перекладі – 20 492 та 293 888 відповідно. Слід зазначити, що при відносній кількості текстової складової у 8 % загального об'єму файла екстремуми функції $K(s)$ виражені достатньо чітко.

На рис. 2 наведено значення функції $K(s)$ для файлів формату TXT без формул та рисунків. Текст статті англійською мовою склав 26 658 символів (з пробілами), а переклад українською мовою – 24 586 символів. Як видно з рисунка, інтенсивність екстремумів порівнянна з рис. 1.

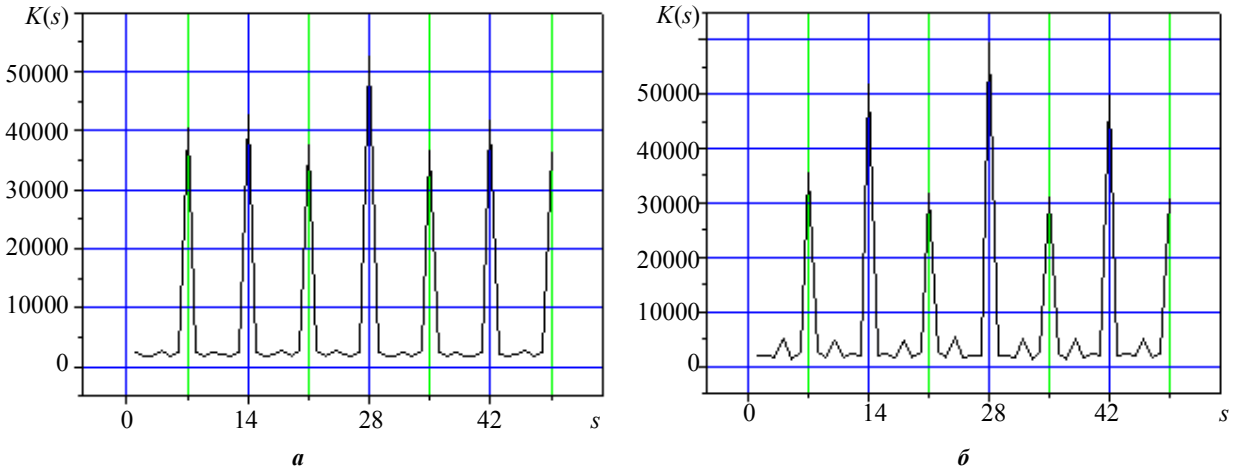

Рис. 1. Функція автокореляції шифротексту для форматів DOC тексту статті [6] (***а***) та його перекладу українською мовою (***б***)

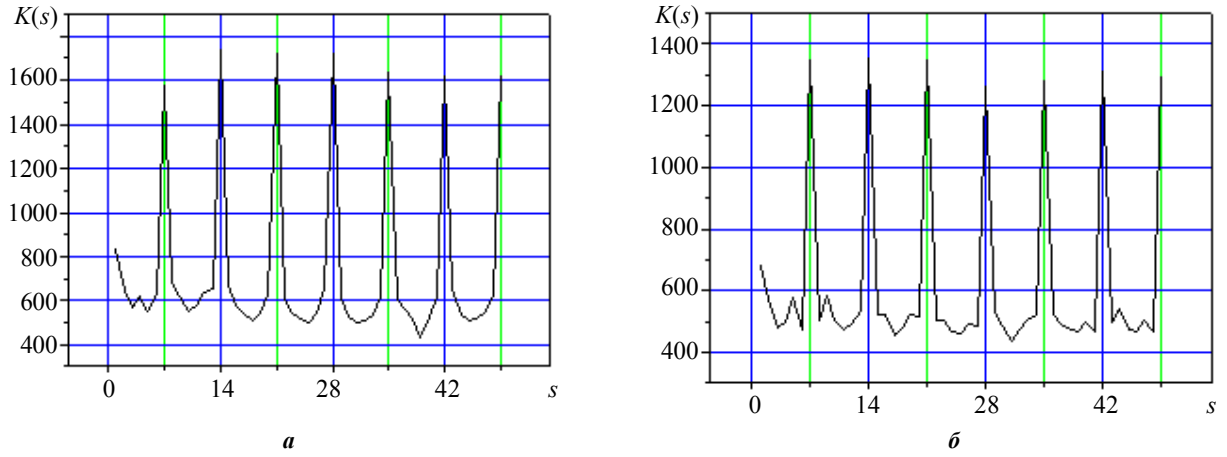

Рис. 2. Функція автокореляції шифротексту для форматів TXT тексту статті [6] (***а***) та його перекладу українською мовою (***б***)

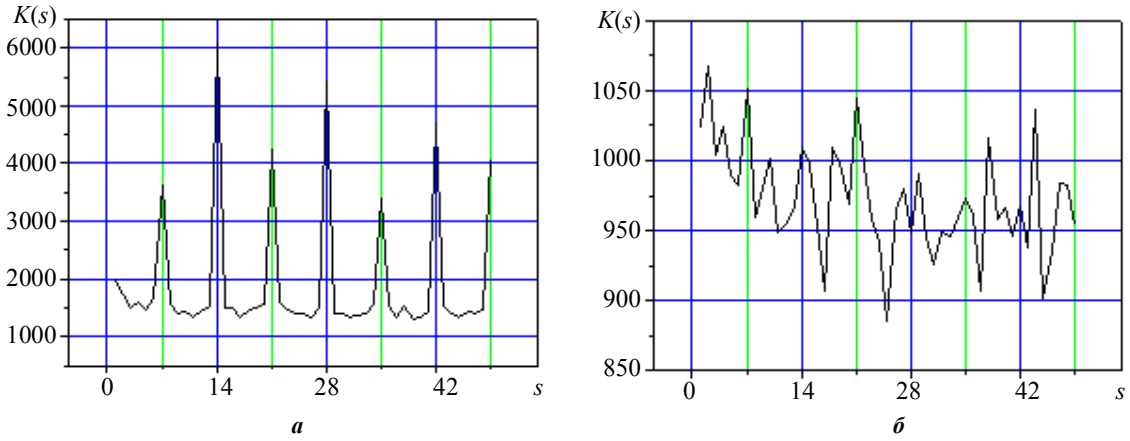

Рис. 3. Функція автокореляції шифротексту статті [6] у форматі PDF (***а***) та її архіву у форматі RAR (***б***)

На рис. 3 наведено значення функції $K(s)$ для файлів формату PDF (розмір 274 386 байт) та RAR (244 068 байт). Можна відзначити, що формат PDF зберігає характер $K(s)$, у той час як формат RAR не відхиляється від загальної картини.

Причиною відсутності екстремумів для файла у форматі RAR при незначній відмінності у розмірах файлів можна вважати небайтне подання інформації в архіві. Для аналізу таких файлів слід використовувати інший спосіб формування символів різної бітової довжини.

ВИСНОВКИ

Запропоновано методику визначення довжини пароля шифрів поліабеткової заміни за допомогою функції автокореляції шифротексту. Створено програмне забезпечення, що реалізує розроблену методику. Методика позитивно апробована на комп'ютерних текстово-графічних файлах різного формату англійською та українською мовами. Перспективи подальших досліджень вбачаються у розвиненні кількісних характеристик функції автокореляції шифротексту, вивченні їх поведінки для файлів різного об'єму, різними мовами тощо.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] **Аграновский, А. В.** Практическая криптография: алгоритмы и их программирование [Текст] / А. В. Аграновский, Р. А. Хадим. – М. : Соломон-пресс, 2009. – 256 с.

[2] **Волков, С. В.** Криптология [Текст] / С. В. Волков. – Рубежное : ИХТ ВНУ, 2010. – 90 с.

[3] **Кан, Д.** Взломщики кодов [Текст] / Д. Кан. – М. : Центрполиграф, 2000. – 480 с.

[4] Розвинення криптології та її місце в сучасному суспільстві [Текст] / М. В. Захарченко, Л. Г. Йона, Ю. В. Щербина, О. В. Онацький. – О. : ОНАЗ ім. О.С. Попова, 2003. – 80 с.

[5] **Шнайер, Б.** Прикладная криптография [Текст] / Б. Шнайер. – М. : Триумф, 2002. – 816 с.

[6] **Samoilenko, D. N.** Knowledge diagnostics by search methods in the semantic space [Text] / D. N. Samoilenko // Electrotechnic and computer systems. – 2012. – № 7 (83).

Надійшла до редколегії 11.03.13
Статтю рекомендує до друку член редколегії Вісника НУК
д-р техн. наук, проф. *В. С. Блінцов*
Статтю розміщено у Віснику НУК № 2, 2013