МЕТОДИ ТА АЛГОРИТМИ ПРОЕКТУВАННЯ

УДК 004.056.55; 004.051

В.С. Яковина, Д.В. Федасюк, С.І. Салій, М.М. Сенів
Національний університет "Львівська політехніка",
кафедра програмного забезпечення автоматизованих систем

ДОСЛІДЖЕННЯ ХАРАКТЕРИСТИК КРИПТОСТІЙКОСТІ АЛГОРИТМУ СИМЕТРИЧНОГО ШИФРУВАННЯ DES

Здійснено дослідження характеристик дифузії та конфузії алгоритму DES та статистичне дослідження особливостей шифрування цього алгоритму. Встановлено, що середнє значення дифузії і конфузії алгоритму DES становить 49,7 % та 49,8 % відповідно, а дисперсія значень залежно від номера біта блока (чи ключа) становить 2,3 % і 3,5 %. Показано, що середня імовірність появи *i*-го байта шифрованого тексту, що збігається з *i*-м байтом відкритого тексту, становить 0,39 % для кожного байта. Обґрунтовано використання алгоритму DES як генератора псевдовипадкових чисел для розподіленої системи теплового проектування та показано можливі способи крипто-аналізу алгоритму за схемою "з вибраним відкритим текстом".

The studies of diffusion and confusion characteristics as well as encrypting peculiarities of DES algorithm have been performed. The diffusion and confusion mean value is 49.7 % and 49.8 % correspondingly with dispersion depending on bit number of 2.3 % and 3.5 %. It is shown that average probability of *i*-th plain text byte to be equal to *i*-th cipher text byte is about 0.39 % for each byte. The exploitation of the DES algorithm as a random number generator for distributed thermal design system has been substantiated as well as the possible ways for the algorithm cryptanalysis using chosen plaintext have been demonstrated.

Вступ

Криптографія має справу з нетривіальними моделями порушника, що припускають наявність у нього обчислювальних, математичних і криптоаналітичних можливостей [1]. Практична реалізація криптографічних засобів захисту інформації повинна забезпечувати захист і від порушника, що володіє відповідними лабораторними можливостями. Методи криптографії не залежать від виду інформаційної системи: не важливо, чи реалізована ця система на основі універсальних комп'ютерів або апаратно, чи є вона системою зв'язку, управління або базою даних.

Криптографічний захист даних забезпечує низку захисних характеристик, таких, як конфіденційність, цілісність, достовірність, неповторюваність і коректність даних, ідентифікація суб'єкта інформаційної системи тощо. Криптографічний захист даних забезпечується їхнім перетворенням, яке може бути подане функцією множини входів у множину виходів. Ця функція може бути як оборотною, так і необоротною. Крім того, вона може залежати від змінного параметра – ключа, який зазвичай буває секретним. Така функція називається криптографічним примітивом.

Алгоритм блокового шифрування DES, незважаючи на значний вік, є одним з найбільш досліджених та стійких алгоритмів. Основним його недоліком залишається мала довжина ключа та

відносна слабкість до диференціального криптоаналізу. Нині алгоритм DES все ще широко використовується у реальних криптосистемах, наприклад, в PGP, переважно у вигляді "потрійної" модифікації, яка збільшує довжину ключа, та як криптографічно стійкий генератор псевдовипадкових чисел у стандарті ANSI X9.17 [2].

Під час побудови системи захисту інформації для розподіленої системи теплового проектування однією з важливих складових є використання криптостійкого генератора псевдовипадкових чисел. Нами було здійснено дослідження різних алгоритмів генерування псевдовипадкових чисел на предмет статистичної однорідності розподілу [3] та дослідження швидкодії програмної реалізації алгоритму DES на сучасних апаратних платформах [4], однак при цьому не вивчались питання передбачуваності послідовності та криптостійкості алгоритму. На передбачуваність генерованих чисел послідовності у разі використання як генератора алгоритму ANSI X9.17 значно впливатимуть такі характеристики базового для стандарту алгоритму DES, як дифузія і конфузія, які визначатимуть, на скільки зміниться наступне число при зміні входу алгоритму на один біт.

Крім того, нами розпочато серію досліджень стосовно використання нейронних мереж для задач криптоаналізу алгоритмів симетричного шифрування [5]. Одним з важливих етапів цього дослідження є формування навчальних вибірок та пошук особливостей алгоритму шифрування, таких, як недостатнє значення дифузії і конфузії чи збіг одного чи більше байтів відкритого тексту з такими у шифрованому тексті, що згідно з нашими попередніми результатами [5] є найефективнішим для використання в нейромережевих засобах криптоаналізу.

Отже, метою цієї роботи було експериментальне дослідження характеристик дифузії та конфузії алгоритму DES з метою як визначення множини вхідних даних для криптоаналізу засобами нейронних мереж, так і для дослідження криптостійкості генератора псевдовипадкових чисел на основі цього алгоритму. Другим завданням роботи було статистичне дослідження особливостей шифрування алгоритму DES, яке полягає у еквівалентності одного чи більше блоків відкритого тексту відповідним блокам шифрованого тексту при певних ключах шифрування, з метою формування навчальних та тестових вибірок для криптоаналізу алгоритму з використанням нейронних мереж.

Методика досліджень

Терміни дифузія (diffusion) і конфузія (confusion) були введені в шифрування Клодом Шенноном [6] для того, щоб охарактеризувати два основні будівельні блоки криптографічних систем. Основним завданням, яке ставив перед собою Шеннон, було перешкодити спробам криптоаналізу, основаним на статистичному аналізі повідомлень. Шеннон ввів поняття ідеального шифру – шифру, який повністю приховує в шифрованому тексті всі статистичні закономірності відкритого тексту. Крім характеристики ідеальних систем, Шеннон запропонував два методи, завданням яких є утруднення криптоаналізу: дифузію і конфузію. Суть дифузії полягає у розсіянні статистичних особливостей відкритого тексту по широкому діапазону статистичних характеристик шифрованого тексту. Це досягається тим, що значення кожного елемента відкритого тексту впливає на значення багатьох елементів шифрованого тексту або, що еквівалентно сказаному, будь-який з елементів шифрованого тексту залежить від багатьох елементів відкритого тексту. У блокових шифрах, що оперують двійковими даними, дифузії можна досягти за допомогою декількох послідовних перестановок даних з подальшим застосуванням до результату перестановки деякої функції – в результаті у формуванні кожного біта шифрованого тексту буде брати участь багато бітів відкритого тексту.

У будь-якому блоковому шифрі використовується залежне від ключа перетворення блока відкритого тексту на блок шифрованого тексту. Механізм дифузії покликаний зробити статистичний взаємозв'язок між відкритим і шифрованим текстами якомога складнішим, щоб максимально ускладнити задачу визначення ключа з такого взаємозв'язку. Що стосується

конфузії, то перед нею ставиться задача максимально ускладнити статистичний взаємозв'язок між шифрованим текстом і ключем з тією самою метою протистояти спробам визначити ключ. Отже, навіть якщо противник зуміє визначити деякі статистичні особливості шифрованого тексту, складність використання ключа для отримання шифрованого тексту повинна виявитись достатньою для того, щоб спроби визначити ключ на основі цих статистичних особливостей виявились безрезультатними. Це досягається використанням складних підстановочних алгоритмів.

Як зазначено в [7], поняття дифузії і конфузії виявились настільки вдалими з погляду опису суті бажаних характеристик блокових шифрів, що ці терміни стали базовими для усіх розробників сучасних шифрів цього типу.

Отже, бажаною властивістю будь-якого алгоритму шифрування повинна бути висока чутливість результату до зміни початкових даних – будь-які незначні зміни відкритого тексту чи ключа повинні приводити до значних змін в шифрованому тексті (лавинний ефект). Зокрема, зміна значення одного біта відкритого тексту чи ключа повинна відображатись у зміні значень багатьох бітів шифрованого тексту. Аналогічний сильніший критерій [8], що називається строгим критерієм лавинного ефекту (SAC – strict avalanche criterion), вимагає, щоб для будь-яких i та j при інвертуванні вхідного біта i будь-який вихідний біт j змінювався з імовірністю ½. Іншим критерієм, запропонованим у [8], є критерій незалежності бітів (BIC – bit independence criterion), згідно з яким для будь-яких значень i, j та k при інвертуванні вхідного біта i вихідні біти j та k повинні змінюватись незалежно. Строгий критерій лавинного ефекту і критерій незалежності бітів покликані гарантувати алгоритму необхідний рівень дифузії.

Для експериментальних досліджень було створено програмну реалізацію алгоритму шифрування DES у режимі ECB (електронної записної книжки), яка б дала можливість накопичувати результати шифрування та змінювати параметри шифрування. Зокрема, в програмі передбачено автоматичну генерацію пар вхідних блоків для дослідження дифузії, пар ключів шифрування для дослідження конфузії та статистичну обробку результатів. Крім того, для вивчення особливостей шифрування алгоритму DES створений програмний продукт уможливлював автоматизований пошук пар "відкритий текст – шифрований текст", в яких один чи більше байтів відкритого тексту збігаються з такими у шифрованому тексті; для можливості аналізу пар такий пошук можна вести у напівавтоматичному режимі, коли оператор може фіксувати один чи більше байтів вхідного блока чи ключа та здійснювати повну вибірку байтів за допомогою циклічних зсувів.

Для дослідження дифузії було сформовано 640 пар вхідних блоків, які відрізнялися на 1 біт, а для дослідження конфузії – 560 пар ключів, які відрізнялися на 1 біт. Дифузія визначалась як кількість (у %) різних бітів для пари шифрованих текстів, якщо відповідна пара відкритих текстів відрізнялась на 1 біт, а ключ шифрування був однаковим. Конфузія визначалась як кількість (у %) різних бітів для пари шифрованих текстів, якщо відповідна пара ключів відрізнялась на 1 біт, а відкритий текст був однаковим. Експеримент здійснювали 10 разів для кожного біта блока відкритого тексту та ключа шифрування, а результат усереднювався.

Для статистичного дослідження особливостей шифрування алгоритму DES були виконані експерименти, які полягали у такому:

- Фіксувався ключ шифрування (випадковий), для 10000 випадкових вхідних блоків вели пошук шифрованих блоків, в яких 1 чи більше байтів збігаються з відповідними байтами вхідного блока. Кожен експеримент повторювався по 3 рази для 10 різних ключів шифрування.
- Фіксувався вхідний блок (випадковий), для 10000 випадкових ключів шифрування вели пошук шифрованих блоків, в яких 1 чи більше байтів збігаються з відповідними байтами вхідного блоку. Кожен експеримент повторювався по 3 рази для 10 різних вхідних блоків.

• З метою пошуку пар "відритий текст – шифрований текст", в яких збіг спостерігається для 2 і більше байтів, було здійснено експеримент для 1000000 випадкових ключів шифрування та випадкових вхідних блоків.

• Для кожного вхідного блока та ключа шифрування, для яких збіг становив 3 байти, було виконано по 100000 експериментів з фіксуванням як ключа шифрування, так і вхідного блока з метою пошуку інших пар "відритий текст – шифрований текст", в яких збіг спостерігається для 3 байтів.

Результати тестування

Усереднене для 10 експериментів для кожного біта блока значення дифузії алгоритму DES наведено на рис. 1. Максимальне значення дифузії становить 53 %, мінімальне – 45 %. Крім того, були розраховані основні статистичні показники для ряду значень дифузії, які наведено в табл. 1. Основними характеристиками в нашому випадку є структурні характеристики вибірки: мода і медіана. Вони практично визначають структуру вибіркових даних і визначаються через ці дані. Ці показники мають статус основних або головних при асиметричному розподілі даних, причому у разі асиметрії розподілу мода або медіана беруть на себе роль середнього значення, відстань між ними може характеризувати ступінь асиметрії, крім того, медіана вважається найстійкішою характеристикою вибірки, а тому може бути основою для критерію оптимального розподілу даних в інтервалах. Як бачимо, в нашому випадку значення медіани і моди практично збігається з середнім значенням показника дифузії, що підтверджує хорошу статистичну однорідність даних.

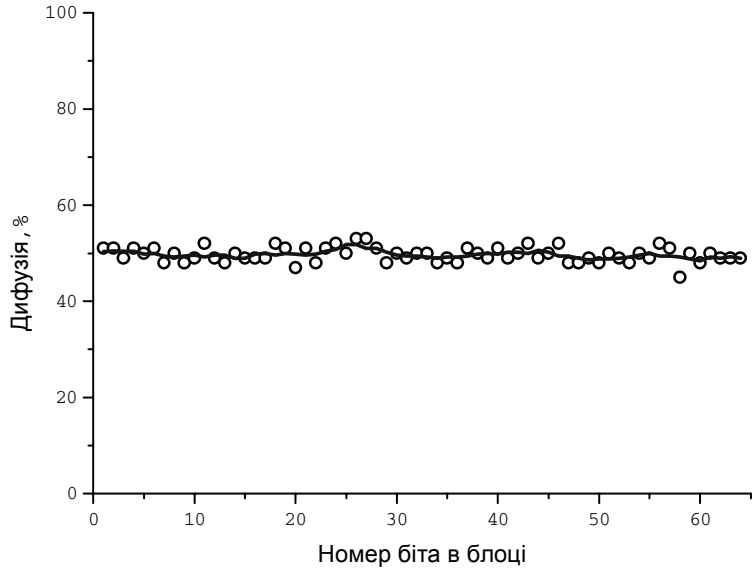


Рис. 1. Усереднене значення дифузії для алгоритму DES (точки – експериментальні значення, лінія – згладжування за 5 точками)

Таблиця 1

Статистичні характеристики розподілу значення дифузії та конфузії алгоритму DES від номеру біта блока

Параметр	Дифузія	Конфузія
1	2	3
Найбільше значення	53	54
Найменше значення	45	46
Розмах	8	8

Продовження табл.1

1	2	3
Медіана	50	50
Середнє арифметичне значення	49,7	49,8
Значення моди	49	48
Дисперсія*	2,3	3,5
Стандартне відхилення*	1,5	1,9

* – Дисперсія D і стандартне відхилення σ^2 визначались як: $D = \overline{x^2} - \overline{x}^2$, $\sigma^2 = \sqrt{D}$.

На рис. 2 наведено типові значення дифузії алгоритму DES, отримані під час виконання серії з 10 експериментів для окремих бітів блока тексту. На цьому рисунку подано значення дифузії для біта з максимальним (рис. 2, *б*), мінімальним (рис. 2, *г*) та типовими (рис. 2 *а*, *в*) значеннями дифузії. Як видно з рис. 2, розподіл значень дифузії у межах виконаних експериментів є достатньо однорідним, що підтверджує високу надійність алгоритму шифрування DES.

На рис. 3 наведено усереднене по 10 експериментах для кожного біта ключа шифрування значення конфузії алгоритму DES. Максимальне значення дифузії становить 54 %, мінімальне – 46 %. Основні статистичні показники залежності конфузії від номера біта ключа наведено в табл. 1. Як видно з цієї таблиці, однорідність розподілу значень конфузії є доволі високою з дисперсією 3,5, що підтверджується також значеннями моди та медіани, які становлять 48 і 50 відповідно, при середньому значенні показника конфузії 49,8 %.

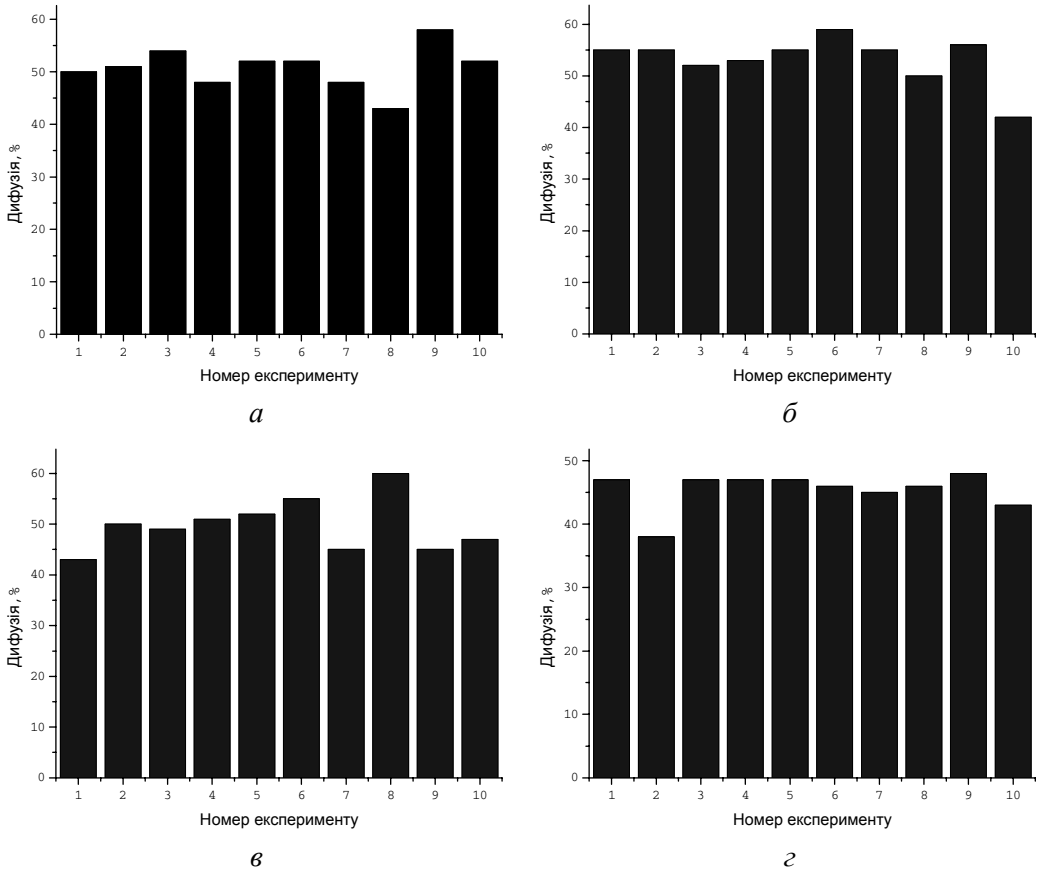


Рис. 2. Розкид значень дифузії для алгоритму DES в межах серії експериментів для одного біта блока (а – біт № 1, б – біт № 28, в – біт № 42, г – біт № 58)

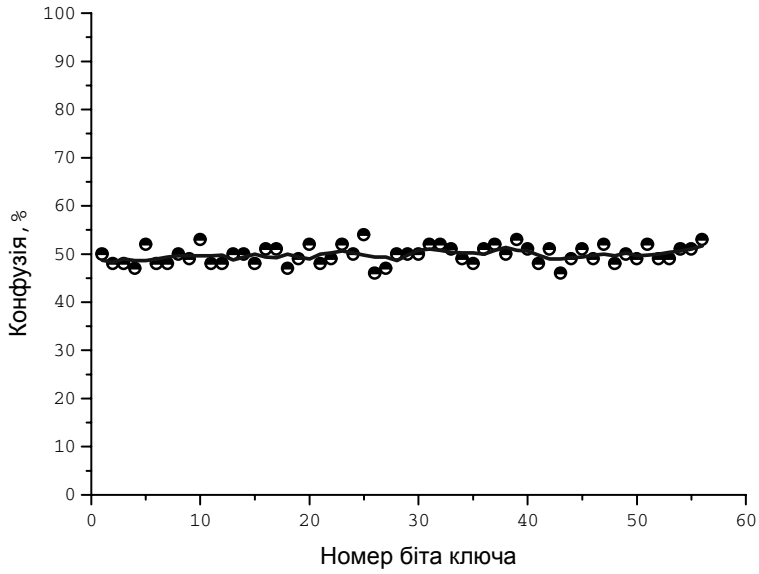
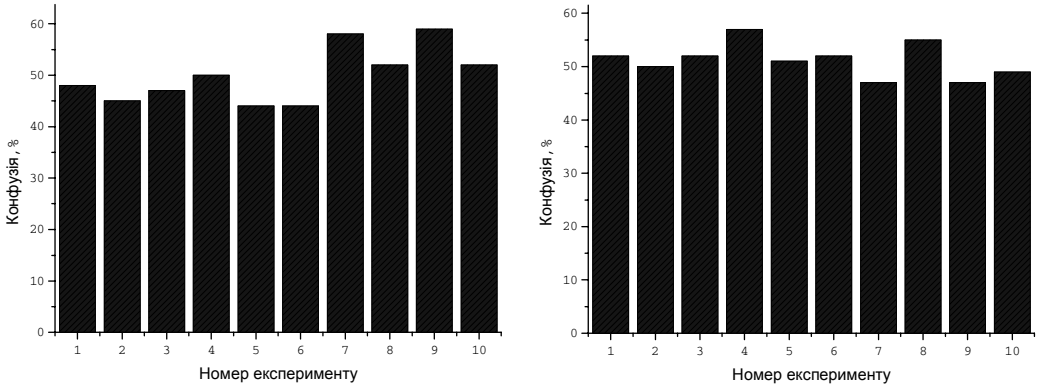


Рис. 3. Усереднене значення конфузії для алгоритму DES (точки – експериментальні значення, лінія – згладжування за 5 точками)

Типові значення розкиду конфузії алгоритму DES в серії з 10 експериментів для окремих бітів ключа шифрування показані на рис. 4, *а*, *б*. Як видно з рис. 4, розподіл значень конфузії у межах виконаних експериментів показує хорошу статистичну однорідність, що також підтверджує високу надійність алгоритму шифрування DES.



а *б*

Рис. 4. Розкид значень конфузії для алгоритму DES у межах серії експериментів для одного біта ключа (а – біт № 8, б – біт № 41)

Наступним етапом експериментів було статистичного дослідження особливостей шифрування алгоритму DES. Під час пошуку шифрованих блоків, в яких 1 чи більше байтів збігалися з відповідними байтами вхідного блока, здійснювався як підрахунок загальної кількості пар "відкритий текст – шифрований текст", так і підрахунок пар залежно від номера байта, що збігся. Результати експерименту з 1000000 блоків та ключів шифрування наведено на рис. 5. На ньому зображено кількість пар "відкритий текст – шифрований текст", в яких були однакові байти, залежно від номера цього байта. Розрахована з цих результатів середня імовірність появи *i*-го байта шифрованого тексту, що не відрізняється від *i*-го байта відкритого тексту, становить 0,39 % для кожного байта. При збільшенні кількості вхідних блоків/ключів шифрування в одному експерименті спостерігається покращання рівномірності розподілу однакових байтів залежно від номера байта в блоці.

Під час експерименту з 1000000 випадкових вхідних блоків та ключів шифрування було виявлено 3 пари "відкритий текст – шифрований текст", для яких збіг становив 3 байти, і ще 2 пари було виявлено під час попередніх 20 експериментів зі 100000 пар блоків/ключів. Виявлені пари в шістнадцятковому вигляді наведені в табл. 2.

Метою останнього з серії експериментів, в якому для кожного вхідного блока та ключа шифрування, для яких збіг становив 3 байти, було здійснено по 100000 експериментів з фіксуванням як ключа шифрування, так і вхідного блока, було встановити кореляцію між імовірністю появи трьох однакових байтів у парі з ключем шифрування чи блоком відкритого тексту. В ході цього експерименту встановлено, що за імовірності появи трьох однакових байтів приблизно 2–3 на мільйон блоків, при використанні як вхідних даних фіксованого вхідного блока, що мав 3 однакові байти при певному ключі, і здійсненні експерименту на 100000 ключів для усіх 5 експериментів виявлявся принаймні ще один ключ шифрування, для якого у шифрованому тексті було 3 однакові байти з відкритим текстом. Так, наприклад, для вхідного блока № 1 (табл. 2) було знайдено ключ шифрування 0x9F DD 96 94 EF FE EC, при якому шифрований блок (0x89 A8 01 6F C3 5C C2 ED) має спільні байти № 2, 5 та 6 з вхідним блоком. При фіксуванні ключа шифрування такого результату отримати не вдалось. Отже, можна зробити висновок, що поява трьох однакових байтів у парі "відкритий текст – шифрований текст" залежить від блока відритого тексту, а не від ключа шифрування, що може бути пов'язано з деякою лінійною залежністю S-матриць алгоритму DES, і може бути використано при атаці на алгоритм DES за схемою "з вибраним відкритим текстом".

Таблиця 2

Пари "відкритий текст – шифрований текст" з 3 однаковими байтами

№	Вхідний блок	Шифрований блок	Ключ шифрування
1	A2 A8 8D A5 C3 5C FC 94	82 A8 6C 97 B5 68 FC 94	39 7C 56 DB FE 01 71
2	2C C7 2A 3D 5D A3 28 47	2C 8A AF 3D C3 6C 68 47	C5 97 2D A9 DB F7 8C
3	25 D8 55 37 76 59 B7 DE	25 5B 78 59 95 59 99 DE	79 71 91 B7 07 EB B0
4	10 EA C4 FF 0B C8 1D 28	10 EF 56 52 F5 60 1D 28	34 C0 A4 9A FB 41 D4
5	10 89 58 6B C3 D2 4A 0A	10 B3 58 B6 3B C1 EA 0A	AD 84 CE 9C 82 B4 21

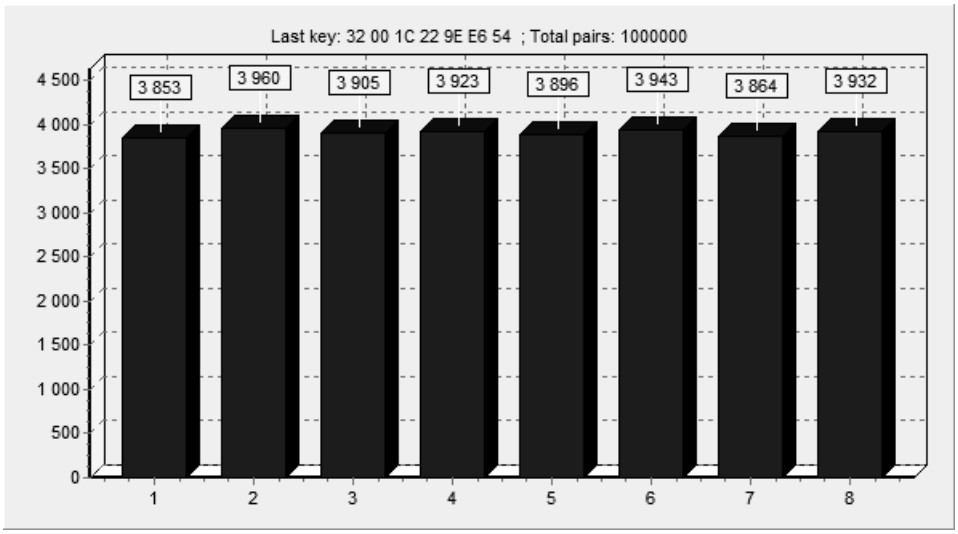


Рис. 5. Кількість пар "відкритий текст – шифрований текст" з однаковими байтами залежно від номера байта для експерименту з 1000000 блоків

Висновки

У роботі здійснено експериментальне дослідження характеристик дифузії та конфузії алгоритму DES для усіх бітів блока відкритого тексту і ключа відповідно; виконано статистичне дослідження особливостей шифрування цього алгоритму, зокрема пошук пар "відкритий текст – шифрований текст" в яких *i*-й байт блока шифрованого тексту збігається з *i*-м байтом блока відкритого тексту.

Встановлено, що середнє значення дифузії і конфузії алгоритму DES становить 49,7 % та 49,8 % відповідно, а дисперсія значень залежно від номера біта блока (чи ключа) становить 2,3 % і 3,5 %. Однорідність розподілу значень цих характеристик підтверджується також значеннями моди та медіани, які майже не відрізняються від середнього арифметичного значення. Дифузійні характеристики алгоритму DES практично відповідають сильному лавинному критерію, і мають статистично однорідний розподіл.

Отже, завдяки високій швидкості програмної реалізації та високим характеристикам криптостійкості при побудові модуля захисту розподіленої системи теплового проектування доцільно використати як генератор псевдовипадкових чисел стандарт ANSI X9.17-1985, який ґрунтується на алгоритмі DES.

Під час дослідження особливостей шифрування алгоритму DES показано, що середня імовірність появи *i*-го байта шифрованого тексту, що збігається з *i*-м байтом відкритого тексту, становить 0,39 % для кожного байта. Імовірність появи 3 байтів, що збігаються, в одному блоці становить приблизно 0,0003 %. При збільшенні кількості вхідних блоків/ключів шифрування в одному експерименті спостерігається покращання рівномірності розподілу однакових байтів залежно від номера байта у блоці.

Поява трьох однакових байтів у парі "відкритий текст – шифрований текст" залежить від блока відритого тексту, а не від ключа шифрування, що може бути пов'язано з певною функціональною залежністю S-матриць алгоритму DES, і може бути використано при атаці на алгоритм DES за схемою "з вибраним відкритим текстом".

Подяки

Ця робота виконувалась за грантом Національного університету "Львівська політехніка" для молодих вчених "Використання нейронних мереж для задачі криптоаналізу алгоритму симетричного шифрування DES" № 4/ГП та держбюджетною темою "Розробка методів та засобів розподілення обчислень в задачах теплового проектування електронних пристроїв нового покоління" ДБ Діаграма.

1. Ростовцев А.Г., Маховенко Е.Б. Теоретическая криптография. – СПб.: НПО "Профессионал", 2004. – 478 с. 2. ANSI X9.17-1985, Financial Institution Key Management (Wholesale). 3. Yakovyna V., Odukha A., Smirnov V. Testing random number generators // Proceedings of the 2nd International Conference of Young Scientists Computer Science and Engineering CSE-2007, Lviv, Ukraine, 2007, P. 25–28. 4. Яковина В., Федасюк Д., Сенів М., Білас О. Порівняння швидкодії програмної реалізації алгоритмів симетричного (DES) та асиметричного (RSA) шифрування // Вісник Національного університету "Львівська політехніка" Комп'ютерні науки та інформаційні технології. – № 598 (2007). – С. 181–185. 5. Yakovyna V., Saliy S., Bilas O. The analysis of neural networks architecture for DES encryption task // Proceedings of the International Conference on Computer Science and Information Technologies CSIT'2007, Lviv, Ukraine, 2007. – P. 56–58. 6. Шеннон К. Теория связи в секретных системах // В сб.: Шеннон К. "Работы по теории информации и кибернетике". – М.: ИЛ, 1963. – С. 333–402. 7. Robsaw M. Block Ciphers // RSA Laboratories Technical Report TR-601, August 1995. 8. Webster A.F. and Tavares S.E. On the Design of S-Boxes // Advances in Cryptology CRYPTO '85 Proceedings, Springer-Verlag, 1986, pp. 523–534.