УДК 004.056.55

Аналіз колізійних властивостей режиму вироблення імітовставок із вибірковим гамуванням

Д. В. Іваненко[1], О. О. Кузнецов[2], Є. П. Колованова[2]
1 Харківський національний університет радіоелектроніки, Україна
2 Харківський національний університет імені В.Н. Каразіна, Україна

Досліджується режим вироблення імітовставки із вибірковим гамуванням, який призначено для забезпечення цілісності та конфіденційності повідомлень. Розглядаються основні криптографічні перетворення, які застосовуються при реалізації цього режиму, даються теоретичні оцінки ймовірності збігів (колізій) формованих імітовставок. Обґрунтовуються пропозиції щодо вдосконалення дослідженого режиму при застосуванні в інформаційно-комунікаційних системах

Ключові слова: *колізія, хешування, імітовставка, криптографічний захист, блокове симетричне шифрування, цикл, підстановка.*

Исследуется режим выработки имитовставки с выборочным гаммированием, который предназначен для обеспечения целостности и конфиденциальности сообщений. Рассматриваются основные криптографические преобразования, которые используются при реализации этого режима, приводятся теоретические оценки вероятности коллизий сформированных имитовставок. Обосновываются предложения по усовершенствованию исследуемого режима при использовании в информационно-коммуникационных системах

Ключевые слова: *коллизия, хеширование, имитовставка, криптографическая защита, блоковое симметричное шифрование, цикл, подстановка.*

In the paper, we investigate formation of the Galois Message Authentication Code with selective Counter, which is designed to ensure the integrity and confidentiality of communications. The paper describes the basic cryptographic transformations used to implement this mode, and gives the theoretical estimates of collisions probability. We propose and substantiate improvements of this mode when applied to information and communication systems.

Key words:*. collision, hashing, authentication code, cryptographic protection, block symmetric encryption, cycle, permutation.*

1. Вступ

Однією з важливих складових забезпечення інформаційної безпеки є криптографічний захист інформації, тому дослідження сучасних криптоперетворень та обґрунтування перспективних напрямків зі створення надійних національних технологій захисту інформації є важливим та складним науковим завданням.

З метою побудови сучасних механізмів криптографічного захисту інформації широко застосовується блокове симетричне шифрування, яке полягає у перетворенні інформації з використанням ключових даних з метою приховування (відновлення) змісту інформаційного повідомлення, підтвердження його справжності, цілісності, авторства, тощо. При цьому рівень захищеності інформації залежить не лише від властивостей блокового симетричного шифру, але і від режиму шифрування, під яким зазвичай розуміється такий метод його використання, який дозволяє реалізувати

© Іваненко Д. В., Кузнецов О. О., Колованова Є. П., 2014

перетворення послідовності блоків відкритих даних в послідовність блоків зашифрованих даних із отриманням певних, наперед визначених криптографічних властивостей [10].

Надійним механізмом забезпечення цілісності та конфіденційності інформації в сучасних інформаційно-комунікаційних системах є режим формування імітовставки із вибірковим гамуванням (Galois/Counter Mode and GMAC), специфікацію якого наведено у міжнародному стандарті NIST SP 800-38D [4]. Цей режим призначено для реалізації швидкого криптоперетворення при забезпеченні послуг безпеки інформації із використанням різних криптографічних примітивів, зокрема поліноміального ґешування, гамування, тощо.

Метою даної роботи є аналіз колізійних властивостей формованих імітовставок нового режиму шифрування Galois/Counter Mode and GMAC (GCM & GMAC) та обґрунтування умов його застосування в сучасних інформаційно-комунікаційних системах.

2. Аналіз криптоперетворень GCM & GMAC

Новий режим шифрування Galois/Counter Mode and GMAC призначено для забезпечення послуг конфіденційності та цілісності даних, перш за все, при реалізації інформаційно-комунікаційних протоколів, зокрема в межах протоколів безпеки IPSec.

Структурна схема режиму вироблення імітовставки із вибірковим гамуванням GCM-AE_K (IV, P, A) наведено на рис. 1, де позначення 0^s визначає рядок довжини s, який складається з бітів ‘0’; len(X) - бітова довжина рядка X; $[x]_s$ повертає бінарне представлення x як рядка бітової довжини s; MSBs(X) повертає s найбільш значущих бітів X; CIPH - затверджений 128-бітний блоковий симетричний шифр. Для забезпечення конфіденційності відкритого тексту P застосовується функція $GCTR_K$ - деяка варіація режиму гамування [1-3], де перший блок лічильника для шифрування відкритого тексту генерується шляхом збільшення (inc_{32}) блоку J_0, сформованого з вектору ініціалізації IV. Для забезпечення цілісності застосовується інший механізм, який засновано на функції ґешування $GHASH_H$. Функцію ґешування використовують для стискання зашифрованих доданих автентифікованих даних A (Additional Authenticated Data – AAD) та шифротексту C в єдиний блок, який далі проходить шифрування для створення коду справжності T (імітовставки).

Зворотне перетворення полягає в перевірці справжності шифртексту C із доданими даними A та реалізується функцією GCM-AD_k (IV, C, A, T), структурну схему якої наведеною на рис. 2. При підтвердженні справжності (знов обчислений код T’ дорівнює отриманому T) виконується розшифрування шифртексту C та формується відкритий текст P.

Таким чином, як видно з рис. 1 та 2, основними перетвореннями нового режиму Galois/Counter Mode and GMAC є ґешування даних із використанням функції $GHASH_H$ та шифрування/розшифрування функцією $GCTR_K$. Розглянемо їх більш детально.

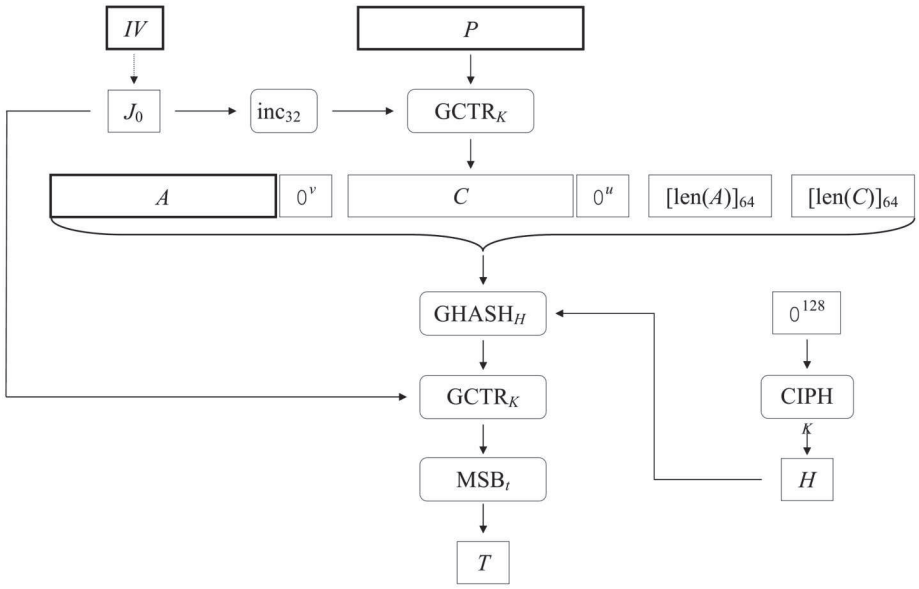

Рис. 1. $\mathrm{GCM-AE_K}\left(\mathrm{IV},\ \mathrm{P},\ \mathrm{A}\right)\ =\ \left(\mathrm{C},\ \mathrm{T}\right).$

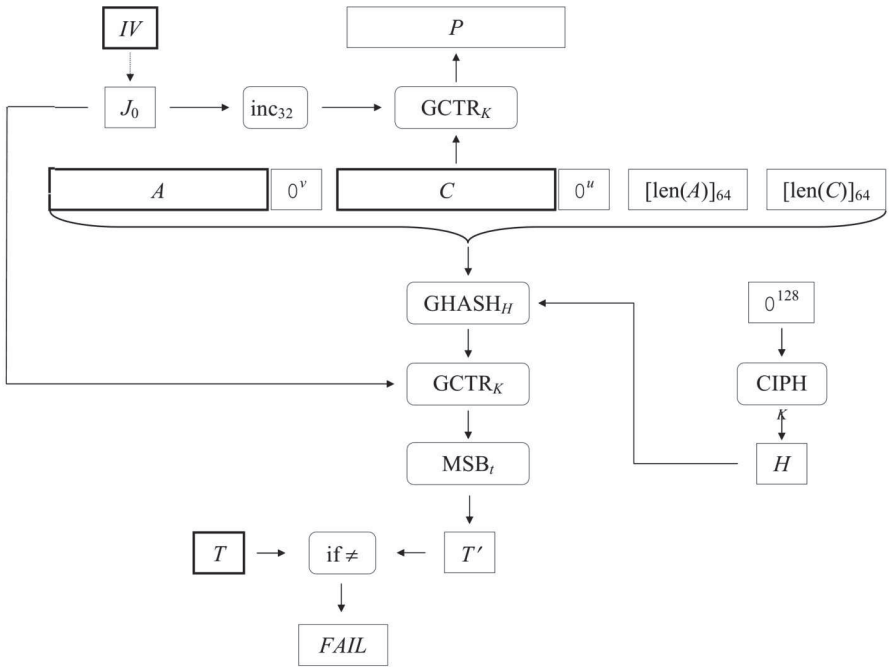

Рис. 2. $\mathrm{GCM-AD_k}\left(\mathrm{IV},\ \mathrm{C},\ \mathrm{A},\ \mathrm{T}\right)\ =\ \mathrm{P}$ *або* $\mathrm{FAIL}.$

На рис. 3 зображено структурну схему функції $\mathrm{GCTR_K}$ для реалізації шифрування/розшифрування, де ICB - початковий блок лічильника; $\mathrm{CB_i}$ – і-ий блок лічильника; inc – функція лічильника. Шифрування відбувається за

допомогою затвердженого блокового симетричного шифру AES зі 128-бітним розміром блоку з використанням ключа шифрування K.

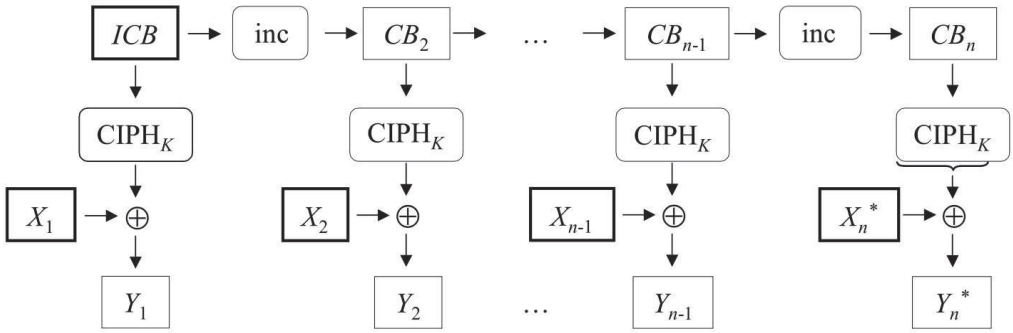

Рис. 3. $\mathrm{GCTR_K}(ICB,\ X_1 \| \mathrm{X}_2 \| ... \| \mathrm{X}_n^*) = Y_1 \| \mathrm{Y}_2 \| ... \| \mathrm{Y}_n^*$.

Для забезпечення цілісності інформації використовується функція ґешування GHASH (див. рис. 1, 2). Цю функцію побудовано на основі поліноміальної схеми та реалізовано за допомогою множення на фіксований параметр – субключ з операціями в двійковому полі Галуа. На вхід функції подається деяка унікальна послідовність блоків довжиною m. Функція ґешування GHASH розраховує ґеш-значенння з використанням схеми Горнера, структуру якої зображено на рис. 4.

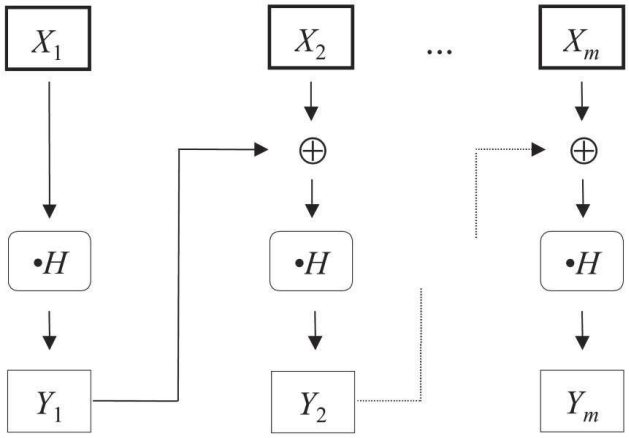

Рис. 4. $\mathrm{GHASH_H}(\mathrm{X}_1 \| \mathrm{X}_2 \| \ ... \ \| \mathrm{X}_\mathrm{m}) = \mathrm{Y}_\mathrm{m}$

Таким чином, проведений аналіз показав, що режим GCM & GMAC, який визначений у NIST Special Publication 800-38D: Galois/Counter Mode (GCM) and GMAC, містить наступні криптоперетворення:

- шифрування/розшифрування відкритого тексту P функцією $\mathrm{GCTR_K}$, яка по суті є деякою варіацією режиму гамування CTR [3, 4];

- обчислення ґеш-значення із використанням поліноміальної функції $GHASH_H$, яку використовують для стискання шифротексту та автентифікованих даних A в єдиний блок;

- шифрування/розшифрування отриманого ґеш-значення, при цьому знов застосовується функція $GCTR_K$.

Отримане зашифроване ґеш-значення і є тим кодом справжності T, який призначено для забезпечення цілісності та автентичності інформації. За вітчизняною термінологією код T є імітовставкою, формування якої призначено для забезпечення захищеності від підміни та/або перекручування даних, здатності протистояти нав'язуванню помилкових повідомлень чи підміні повідомлення з метою зміни його сенсу.

Основним ймовірнісним показником ефективності режиму вироблення імітовставки (коду справжності, ґеш-значення) є ймовірність збігу (колізії) формованих кодів, тобто ймовірність такої події, коли для різних вхідних даних формовані імітовставки співпадають. Тобто ймовірність колізій визначає властивості формованих імітовставок щодо їх співпадіння на повній множині ключів та вхідних даних і є вихідним параметром щодо оцінки імітостійкості [10].

Проведемо дослідження колізійних властивостей розглянутої схеми вироблення імітовставок. При проведенні досліджень зосередимо увагу на вивченні властивостей ґеш-кодів, що сформовано функцією поліноміального ґешування $GHASH_H$, та вихідних кодів справжності (імітовставок), що сформовано за результатом роботи режиму GCM & GMAC загалом.

3. Дослідження колізійних властивостей поліноміального ґешування $GHASH_H$

Проведений аналіз показав, що при формуванні імітовставок у режимі GCM & GMAC використана функція ґешування за поліноміальною схемою Горнера (функція $GHASH_H$). За визначенням вона належить до класу універсальних ґеш-функцій [5].

Ідеєю універсального ґешування є визначення набору ґеш-функцій таким чином, що випадковий вибір функції забезпечить низьку ймовірність того, що будь-які два різних вхідних повідомлення X_a та X_b дадуть колізію, коли їх ґеш-значення розраховані з використанням функції із визначеної множини. Ймовірність виникнення такої колізії можна підрахувати наступним чином [5, 6]:

$$P_k = \delta_y(X_a, X_b) / |H|,$$

де P_k - ймовірність виникнення колізії, $\delta_y(X_a, X_b)$ - кількість співпадінь ґеш-значень, $|H|$ - потужність множини ґеш-функцій (або потужність множини ключів ґешування, бо кожний ключ визначає окрему функцію із визначеної множини).

Для деякої ґеш-функції y та вхідних повідомлень X_a, X_b буде $\delta_y(X_a, X_b) = 1$ якщо $y(X_a) = y(X_b)$, та $\delta_y(X_a, X_b) = 0$ в іншому випадку. Тобто

$\delta_y(\mathrm{X_a},\mathrm{X_b})=1$ лише тоді, коли ґеш-значення від вхідних повідомлень $\mathrm{X_a}$ та $\mathrm{X_b}$ дадуть колізію. Для кінцевої множини H визначимо

$$\delta_H(\mathrm{X_a},\mathrm{X_b})=\sum\nolimits_{y\in Y}\delta_y(\mathrm{X_a},\mathrm{X_b}).$$

Звідси $\delta_H(\mathrm{X_a},\mathrm{X_b})$ підраховує кількість ґеш-функцій (ключів ґешування), які дають колізію для визначених $\mathrm{X_a}$ та $\mathrm{X_b}$.

Згідно з [6] ймовірність виникнення колізії для функції ґешування, яку побудовано на основі поліноміальної схеми, визначається як:

$$P_k=(n-1)/\left|H\right|, \tag{1}$$

де n - довжина вхідного повідомлення (кількість блоків вхідного повідомлення), $\left|H\right|$ - потужність множини ключів (потужність двійкового поля Галуа).

Дійсно, якщо правило ґешування повідомлення $X=(X_1\left\|X_2\right\|...\left\|X_n\right)$ задається через обчислення у кінцевому полі $GF(2^k)$ значення поліному

$$Y_n=X_1\cdot H_j^{n-1}\oplus X_2\cdot H_j^{n-2}\oplus...\oplus X_n\cdot H_j^0, \tag{2}$$

де $H_j\in H$ - значення ключа ґешування, $X_i,H_j\in GF(2^k)$, тоді колізія (співпадіння) ґеш-значення Y_n із ґеш-кодом Y_n^*, який відповідає іншому повідомленню $X^*=(X_1^*\left\|X_2^*\right\|...\left\|X_n^*\right)\neq X=(X_1\left\|X_2\right\|...\left\|X_n\right)$ буде відповідати випадку тотожності:

$$X_1\cdot H_j^{n-1}\oplus X_2\cdot H_j^{n-2}\oplus...\oplus X_n\cdot H_j^0=X_1^*\cdot H_j^{n-1}\oplus X_2^*\cdot H_j^{n-2}\oplus...\oplus X_n^*\cdot H_j^0$$

для будь-якого введеного ключа ґешування H_j.

Останнє рівняння перепишемо у канонічному вигляді:

$$(X_1\oplus X_1^*)\cdot H_j^{n-1}\oplus(X_2\oplus X_2^*)\cdot H_j^{n-2}\oplus...\oplus(X_n\oplus X_n^*)\cdot H_j^0=0, \tag{3}$$

отже колізія ґеш-значень, тобто подія $Y_n=Y_n^*\Big|_{X\neq X^*}$ буде виникати лише тоді, коли H_j є коренем рівняння (3).

Однак за основною теоремою алгебри будь який многочлен степеня $n-1$ має точно $n-1$ коренів з врахуванням їхньої кратності, тобто на безлічі значень $H_j\in H$ не більше $n-1$ різних ключів ґешування будуть обертати в нуль ліву

частину рівності (3). Таким чином, при рівноймовірному обранні ключів ґешування $H_j \in H$ ймовірність колізії $P_к = P\left(Y_n = Y_n^* \middle|_{X \neq X^*}\right)$ буде визначатися за виразом (1).

Беручи до уваги зазначене вище, можна зробити висновок, що, обираючи підходяще значення потужності $|H|$ множини ґеш-функцій (ключів) та довжину вхідного повідомлення, можна досягти необхідного рівня ймовірності виникнення колізії. Графіки залежності ймовірності виникнення колізії від довжини повідомлення та потужності множини ключів для поліноміальної схеми представлено на рис. 5.

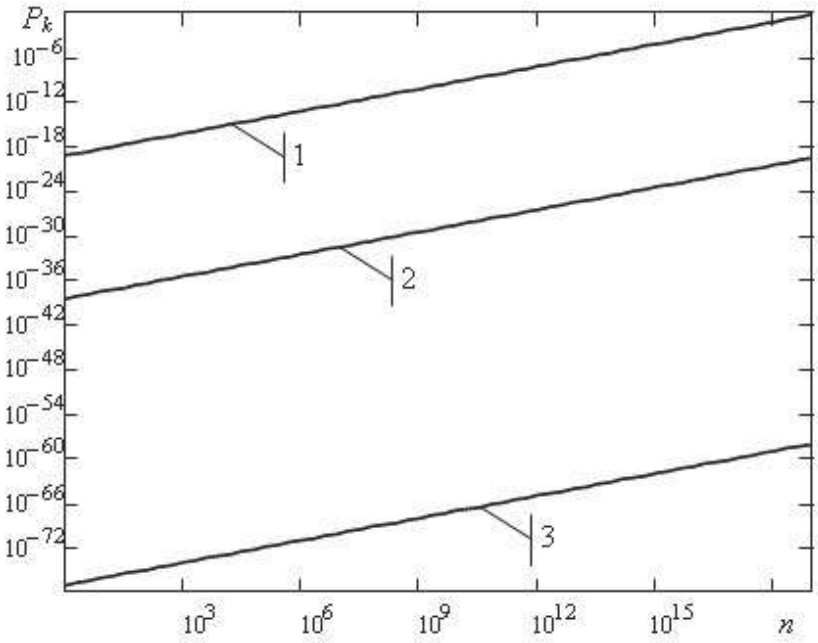

Рис. 5. Залежність ймовірності виникнення колізії від кількості блоків вхідного повідомлення: 1) $|H| = 2^{64}$ *; 2)* $|H| = 2^{128}$ *; 3)* $|H| = 2^{256}$.

Як видно з наведених залежностей ймовірність колізій при застосуванні поліноміальної схеми ґешування значно підвищується із збільшенням довжини вхідного повідомлення. При фіксованій довжині ключів ґешування (та, відповідно, потужності $|H|$) ця залежність накладає додаткові обмеження на довжину повідомлень, які ґешуються. Наприклад, для заданої ймовірності виникнення колізій $P_k = 2^{-32}$ та при довжині ключа ґешування 128 бітів загальна кількість блоків повідомлення повинна задовольняти вимозі $n \le 2^{96}$.

Обчислення ґеш-значення за правилом (2) має певні недоліки. Наприклад, якщо вхідне повідомлення складається лише з одного блоку, тоді формула (2) прийме вигляд $Y_1 = X_1 \cdot H_j^0 = X_1$ і правило ґешування є тотожністю, тобто ґеш-

значення буде дорівнювати блоку повідомлення і ключ ґешування H_j при обчисленні Y_1 зовсім не використовується. Можливо саме тому в специфікації GCM & GMAC [4] застосовується дещо змінена поліноміальна форма обчислення ґеш-коду, а саме (див. рис. 4):

$$Y_m = X_1 \cdot H_j^m \oplus X_2 \cdot H_j^{m-1} \oplus ... \oplus X_m \cdot H_j^1. \quad (4)$$

Випадок колізії буде спостерігатися так само при виконанні рівності:

$$X_1 \cdot H_j^m \oplus X_2 \cdot H_j^{m-1} \oplus ... \oplus X_m \cdot H_j^1 = X_1^* \cdot H_j^m \oplus X_2^* \cdot H_j^{m-1} \oplus ... \oplus X_m^* \cdot H_j^1,$$

яку запишемо у вигляді

$$(X_1 \oplus X_1^*) \cdot H_j^m \oplus (X_2 \oplus X_2^*) \cdot H_j^{m-1} \oplus ... \oplus (X_n \oplus X_n^*) \cdot H_j^1 = 0,$$

що після скорочення на H_j^1 при $n = m-1$ повністю відповідає (3) із оцінкою ймовірності колізій (1).

Таким чином, правило обчислення ґеш-значень (4) є за колізійними властивостями тотожним правилу (2), але навіть у випадку ґешування одного блоку обчислений ґеш-код $Y_1 = X_1 \cdot H_j$ залежить як від X_1, так і від значення ключа ґешування H_j.

Втім, слід відмітити недоліки застосованого правила (4). Якщо ключ ґешування H_j дорівнює нулю, тоді ґеш-код буде також дорівнювати нулю для будь якого вхідного повідомлення. Це накладає додаткові обмеження на схему формування ключових даних схеми ґешування, втім специфікацією режиму GCM & GMAC ніяких обмежень та вказівок з цього приводу не наводиться [4]. Вказано лише, що субключ ґешування H_j формується як зашифрована двійкова послідовність (див. рис. 1, 2).

Розглянемо цей випадок більш докладніше, бо формування нульового ключа ґешування $H_j = 0$, як з'ясувалося, призводить до виродженої роботи поліноміальної схеми $GHASH_H$, і формована імітовставка T у цьому випадку зовсім не буде залежати від ґеш-коду повідомлення, а визначатиметься лише значенням вектору ініціалізації (синхропосилки) IV (див. рис. 1, 2).

Оцінимо ймовірність виникнення нульового субключа ґешування, тобто ймовірність такої події, коли при шифруванні нульового вектору 0^{128} буде отримано значення $H_j = 0$. Для цього скористаємося деякими визначеннями та поняттями теорії підстановок [8].

4. Оцінка ймовірності виникнення нульового субключа гешування $GHASH_H$.

Розглянемо множину всіх бієктивних перетворень множини $Y = \{y_1, y_2, ..., y_n\}$ саму в себе. Ці перетворення, які мають назву підстановок степеня n, утворюють групу відносно операції послідовного виконання перетворень. Така група має назву симетричної групи підстановок степеня n та позначається як S_n [8]. Її потужність визначається потужністю множини всіх підстановок степеня n, тобто дорівнює $n!$.

Кожній підстановці $s \in S_n$ відповідає єдина підстановка $s^{-1} \in S_n$, така, що $s^{-1} \cdot s(y) = s \cdot s^{-1}(y) = e(y)$, $y \in Y$, де $e(y) \in S_n$ - одинична підстановка, тобто $e(y) = y$ для всіх $y \in Y$.

Введемо наступні позначення: $s \cdot s \cdot ... \cdot s = s^k$, $s^{-1} \cdot s^{-1} \cdot ... \cdot s^{-1} = s^{-k}$, де добутки містять k множників. Відповідно маємо $s^k \cdot s^{-k} = s^{-k} \cdot s^k = s^0 = e$.

Множина підстановок степеня n, яка є замкнутою відносно операції множення та обчислення оберненого для $s \in S_n$ елементу $s^{-1} \in S_n$, має назву групи підстановок. Кожна така група є підгрупою симетричної групи S_n [8].

Розглянемо деяку підстановку $s \in S_n$, яка діє на множині Y. Визначимо на множині Y бінарне відношення, при цьому будемо вважати $y \sim y'$ для $y, y' \in Y$ якщо існує таке j, що $y' = s^j(y)$. Це бінарне відношення є рефлексивним, симетричним та транзитивним, тобто є відношенням еквівалентності. Дійсно, відповідно до [8] маємо:

- $y \sim y$, оскільки $y = s^0(y) = e(y)$;

- із умови $y \sim y'$ витікає $y' \sim y$, оскільки із рівності $y' = s^j(y)$ випливає, що $y = s^{-j}(y')$;

- із $y \sim y'$ та $y' \sim y''$ витікає, що $y \sim y''$, бо з рівностей $y' = s^j(y)$ та $y'' = s^i(y')$ випливає, що $y'' = s^i(s^j(y)) = s^{i+j}(y)$.

Цикл s_i підстановки $s \in S_n$ довжини l_i визначається наступним чином:

$$s_i = (y, s_i(y), s_i^2(y), ..., s_i^{l_i - 1}(y)),$$

де $s_i^{l_i}(y) = y$.

Таким чином, довільну підстановку $s \in S_n$ можна розкласти на відповідні цикли [8]:

$$s = (y_1, s_1(y_1), s_1^2(y_1), ..., s_1^{l_1 - 1}(y_1))...(y_k, s_k(y_k), s_k^2(y_k), ..., s_k^{l_k - 1}(y_k)). \quad (5)$$

Наприклад, підстановка s степеня 4 виду

$$s=\begin{pmatrix} y_1 & y_2 & y_3 & y_4 \\ s(y_1) & s(y_2) & s(y_3) & s(y_4) \end{pmatrix}=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$$

може бути подана у вигляді розкладу на 3 цикли:

$$s_1=(y_1)=(1),\ l_1=1;$$
$$s_2=(y_2,s_2(y_2))=(2,4),\ l_2=2;$$
$$s_3=(y_3)=(3),\ l_3=1,$$

тобто маємо наступний розклад:

$$s=(y_1)(y_2,s_2(y_y))(y_3)=(1)(2,4)(3).$$

Загалом, підстановка $s\in S_n$ належить до циклового класу $\{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\}$, якщо вона містить α_1 циклів довжини 1, α_2 циклів довжини 2, і так далі, тобто:

$$s=(y_1)(y_2)...(y_{\alpha_1})(y'_1,y''_1)(y'_2,y''_2)...(y'_{\alpha_2},y''_{\alpha_2})...,$$
$$1\alpha_1+2\alpha_2+...+n\alpha_n=n.$$

Позначимо через $C(\alpha_1,\alpha_2,...,\alpha_n)$ число підстановок в цикловому класі $\{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\}$, а через $C(n,k)$ - число підстановок степеня n, які мають k циклів. Тоді маємо [8]:

$$C(\alpha_1,\alpha_2,...,\alpha_n)=\frac{n!}{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\alpha_1!\alpha_2!...\alpha_n!},$$

$$C(n,k)=\sum_{\substack{1\alpha_1+2\alpha_2+...+n\alpha_n=n\\ \alpha_1+\alpha_2+...+\alpha_n=k,\ \alpha_i\geq 0}}\frac{n!}{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\alpha_1!\alpha_2!...\alpha_n!}=\left|s(n,k)\right|, \qquad (6)$$

де $s(n,k)$ - числа Стірлінга першого роду, які визначаються через співвідношення:

$$x(x-1)...(x-n+1)=\sum_{k=0}^{n}s(n,k)x^k.$$

На множині всіх підстановок степеня n, які утворюють симетричну групу S_n, задамо рівномірний ймовірнісний розподіл, тобто кожній вибраній підстановці $s\in S_n$ поставимо у відповідність ймовірність її обрання, що дорівнює $1/n!$. За сучасними поглядами симетричної криптографії така множина рівноймовірних відображень відповідає уявленню про «ідеальний» шифр, бо якщо обрання окремої підстановки $s\in S_n$ пов'язати із значенням введеного ключа шифрування, тоді отримане перетворення буде відповідати випадковому і рівномірно вибраному шифр тексту для кожного відкритого

тексту при будь-якому ключі, тобто на всіх можливих варіантах відображень відкритого тексту у шифрограму.

За визначенням блочний симетричний шифр є функцією відображення множини текстів і множини ключів в множину шифртекстів: E: K × M → E, де K, M та E – множини ключів, відкритих та шифртекстів, відповідно. Для шифру AES потужність множини ключів $|K| \in \{2^{128}, 2^{192}, 2^{256}\}$, а $|M| = |E| = 2^{128}$. Оскільки при зашифруванні необхідно мати можливість відновити текст за допомогою ключа, потрібно, щоб для всіх ключів $k \in K$ функція зашифрування була перестановкою (підстановкою), тобто відображення E: K × M → E повинно бути бієктивним.

На практиці для довільного n-бітового блокового шифру існує $2^n!$ можливих перестановок відкритого тексту. Практично це означає, що кількість бітів ключа, яку необхідно для отримання всіх можливих перестановок, становить близько $\ln 2^n! \approx n \cdot 2^n$ бітів[1]. Втім розмір ключа більшості блочних шифрів не перевищує невеликого числа, кратного розміру блоку, відповідно такі шифри можуть забезпечити лише невелику частку від повної кількості можливих перестановок. Наприклад, для 128-бітного шифру AES маємо $2^{128}! \approx 2^{128 \cdot 2^{128}}$ можливих перестановок 128-бітових блоків, з яких, в залежності від довжини ключа, використовується тільки 2^{128}, 2^{192} або 2^{256} перетворень. Таким чином, кожен шифр є деяка підмножина повної множини всіх можливих підстановок, що діють на множині блоків оброблюваних даних. Основне припущення, яке приймається при обґрунтуванні стійкості симетричного криптоперетворення полягає саме у збереженні ймовірнісних властивостей випадкової підстановки, тобто припускається, що хоча при шифруванні і застосовується обмежений набір підстановок із S_n, та певні розподіли ймовірностей елементів цієї підмножини відповідають властивостям випадково и рівномірно обраної підстановки із всієї множини S_n.

Проведемо дослідження цих розподілів з метою оцінки ймовірності виникнення нульового шифр тексту при шифруванні нульового відкритого тексту. Для цього розглянемо випадкову величину ξ_n, яка дорівнює числу циклів в випадково вибраній підстановці $s \in S_n$. Оцінимо ймовірність випадкової події $\xi_n = k$, тобто такого випадку, коли у випадково вибраній підстановці буде спостерігатися точно k циклів (див. вираз (5)). З формули (6) безпосередньо випливає вираз для точного розподілу ймовірностей через числа Стірлінга першого роду:

$$P(\xi_n = k) = \frac{C(n,k)}{n!} = \frac{|s(n,k)|}{n!},\ k = 0,1,...,n.$$

В роботах [8, 15] отримано математичне очікування $M\xi_n$ та дисперсію $D\xi_n$ випадкової величини ξ_n:

[1] За формулою Стірлінга $\ln(x!) = x\ln(x) - x - O(\ln(x))$

$$M\xi_n = \sum_{j=1}^{n} \frac{1}{j} = \ln n + C + o(1),\ C = 0,5772...,$$

$$D\xi_n = \sum_{j=1}^{n} \frac{1}{j} - \sum_{j=1}^{n} \frac{1}{j^2} = \ln n + C + o(1),$$

крім того показано, що при $n \to \infty$ випадкова величина $\xi'_n = (\xi_n - \ln n)/(\ln n)$ розподілена асимптотично нормально з параметрами $(0,\ 1)$, тобто

$$\lim_{n\to\infty} P(\xi'_n < u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-y^2/2} dy.$$

В роботах [11-14] досліджено емпіричний розподіл ймовірності виникнення циклу певної довжини у зменшених моделях шифру, встановлено, що цей розподіл дуже близький до розглянутого теоретичного розподілу випадкової підстановки, тобто за цим критерієм можна стверджувати, що шифр за розподілом кількості циклів подібний властивостям випадкової підстановки. В той же час для оцінки ймовірності виникнення нульового субключа ґешування в схемі GCM & GMAC потрібна інша характеристика шифру, а саме розподіл числа циклів заданої довжини. Відповідно до [8], ця характеристика у випадковій підстановці визначається наступним чином.

Позначимо як $\chi_{n,l}$ число циклів довжини l у випадковій рівно ймовірній підстановці степеня n. Тоді розподіл ймовірностей випадкової події $\chi_{n,l} = k$ визначається як:

$$P(\chi_{n,l} = k) = \frac{1}{l^k k!} \sum_{j=0}^{[n/l]-k} \frac{(-1)^j}{l^j j!},\ k = 0,1,...,\lfloor n/l \rfloor. \qquad (7)$$

При $n \to \infty$ випадкова величина $\chi_{n,l}$ має в межі розподіл Пуасона з параметрами $\lambda = 1/l$, тобто

$$\lim_{n\to\infty} P(\chi_{n,l} = k) = \frac{1}{l^k k!} e^{-1/l},\ k = 0,1,....$$

Скористаємося формулою точного розподілу ймовірностей випадкової події $\chi_{n,l} = k$ у вигляді (7). Значення $n!P(\chi_{n,l} = k)$ відповідає кількості підстановок, які містять k циклів довжини l. Нас цікавить кількість підстановок, які обов'язково мають цикли довжини $l = 1$, причому ці цикли повинні переводити фіксований елемент y з множині Y сам у себе. Тобто треба оцінити кількість таких підстановок $s \in S_n$, які для визначеного $y \in Y$ містять цикли $s(y) = y$ довжини $l = 1$. В криптографії при розгляді блокових симетричних

криптоперетвореннь такі випадки прийнято називати фіксованими точками підстановки [11].

Для $l=1$ формула (7) прийме вигляд

$$P(\chi_{n,l=1}=k)=\frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!},\ k=0,1,...,n,$$

причому для $k=1,...,n$ кожен з $n!P(\chi_{n,l=1}=k)$ випадків для фіксованого $y\in Y$ буде спостерігатися

$$\frac{C_{n-1}^{k-1}}{C_n^k}=\frac{(n-1)!}{(k-1)!(n-k)!}\frac{k!(n-k)!}{n!}=\frac{k}{n}$$

разів, тобто кількість фіксованих точок $s(y)=y$ для визначеного $y\in Y$ буде визначатися за формулою:

$$\sum_{k=1}^{n}n!P(\chi_{n,l=1}=k)\frac{C_{n-1}^{k-1}}{C_n^k}=\sum_{k=1}^{n}\left(\frac{(n-1)!}{(k-1)!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!}\right)=(n-1)!,$$

а відповідна ймовірність фіксованої точки у випадковій рівноймовірній підстановці степеня n матиме вигляд:

$$\sum_{k=1}^{n}P(\chi_{n,l=1}=k)\frac{C_{n-1}^{k-1}}{C_n^k}=\frac{(n-1)!}{n!}=\frac{1}{n}. \qquad (8)$$

Останній вираз може бути отриманий значно простіше з тривіальних комбінаторних міркувань. Дійсно, якщо на множині $Y=\{y_1,y_2,...,y_n\}$ зафіксувати m елементів, тоді можливі $(n-m)!$ варіантів перестановок решти елементів. Тобто на всій множині підстановок з S_n при їх випадковому рівноймовірному розподілі ймовірність обрати підстановку з m фіксованими точками буде дорівнювати

$$\frac{(n-m)!}{n!}=\frac{1}{(n-m+1)(n-m+2)...n}=\frac{1}{(n)_m}, \qquad (9)$$

що при $m=1$ співпадає з (8) [2].

Припустимо, що n-бітний блоковий симетричний шифр, який складається з деякої підмножини симетричної групи підстановок степеня 2^n, відповідає ймовірнісним властивостям випадкової підстановки, зокрема приймемо

[2] $(n)_m=n(n-1)...(n-m+1)$ - загальноприйняте позначення спадаючого факторіалу

припущення, що на всій множині ключів шифру, які задають обрання конкретної підстановки s із S_{2^n}, ймовірність обрати підстановку з m фіксованими точками буде визначатися за (9), тобто буде дорівнювати

$$\frac{(2^n - m)!}{2^n!} = \frac{1}{(2^n)_m}.$$

Тоді кількість ключів шифру, які призведуть до появи m фіксованих точок $s(y) = y$ для визначених блоків відкритого тексту $y \in Y$ буде визначатися за формулою $N(m) = \frac{2^k}{(2^n)_m}$, де k - бітова довжина ключа шифру.

Зокрема, для випадку $m = 1$ маємо

$$N(m=1) = \frac{2^k}{(2^n)_1} = 2^{k-n},$$

тобто:

- якщо $k < n$ з високою ймовірністю можна стверджувати, що шифр не містить фіксованих точок, тобто на всій множині ключів шифрування не знайдеться жодного, який призведе до появи хоча б одного нешифрованого відкритого тексту;

- якщо $k = n$ з високою ймовірністю знайдеться один ключ із $|K| = 2^k$ можливих, який призведе до появи фіксованої точки, тобто отримаємо випадок нешифрованого відкритого тексту;

- якщо $k > n$ кількість ключів, які призводять до появи фіксованої точки, стрімко зростає, їх кількість експоненційно залежить від бітової довжини ключа.

У випадку застосування шифру AES маємо таке:

- при довжині ключа $k = 128$ бітів один ключ буде створювати фіксовану точку, тобто, наприклад, буде спостерігатися випадок, коли при шифруванні нульового вектору 0^{128} буде отримано значення $H_j = 0$;

- при довжині ключа $k = 192$ бітів загалом $2^{k-n} = 2^{64}$ ключів будуть створювати фіксовану точку, тобто випадок із шифруванням нульового вектору 0^{128} в нульове значення $H_j = 0$ буде спостерігатися 2^{64} разів;

- при довжині ключа $k = 256$ бітів вже $2^{k-n} = 2^{128}$ ключів відповідати випадку шифрування нульового вектору 0^{128} в те ж саме значення, тобто в $H_j = 0$.

Отримані оцінки дозволяють стверджувати, що при виконанні зроблених припущень, ймовірність виникнення фіксованої точки шифру не залежить від довжини ключа, вона визначається лише довжиною блоків даних, які обробляє шифр. Цей висновок збігається із загальною інтерпретацією шифру як випадкової підстановки, фіксовані точки в якій – звичайне явище. Однак для розглянутого вище ключового поліноміального гешування випадки фіксованих

точок для нульового блоку є неприпустимими, бо це повністю спотворює схему формування ґеш-значеннь, які вже не будуть залежати від вихідних даних, колізійні властивості порушуються і не відповідають теоретичним оцінкам. Очевидно, що при збільшенні потужності множини ключів кількість фіксованих точок також буде зростати і число так би мовити «слабких» ключів, які призводять до виродженої роботи функції ґешування збільшується пропорційно потужності ключового простору.

Розглянемо тепер вплив останнього шару перетворень режиму GCM & GMAC, зокрема шифрування/розшифрування отриманого за допомогою функції $GHASH_H$ ґеш-значення, оцінимо ймовірності показники формованих імітовставок.

5. Дослідження колізійних властивостей формованих режимом GCM & GMAC імітовставок.

Останній шар перетворень режиму GCM & GMAC полягає у застосуванні функції $GCTR_K$, яка по суті є деякою варіацією режиму гамування CTR [3, 4]. Тобто до отриманого на попередньому шарі перетворень ґеш-значення додається результат зашифрування деякого вектору J_0, який отримано із вектору ініціалізації IV (див. рис. 1, 2). Тобто фактично, до сформованого ґеш-значення додається зашифрована константа, яка формується встановленим порядком.

Відповідно до специфікації режиму GCM & GMAC на формування вектору ініціалізації (за вітчизняною термінологією - синхропосилка) накладаються такі обмеження:

- ймовірність того, що при обчисленні імітовставок для двох різних вихідних даних будуть застосовані однакові вектори ініціалізації при однакових ключах не повинна бути більшою ніж 2^{-32};

- для кожного заданого ключа повне число формованих імітовставок для будь яких вихідних повідомлень не повинно перевищувати 2^{32}.

Перша умова визначає максимальну ймовірність застосування однакових параметрів криптографічного перетворення для різних вихідних даних. Тобто навіть якщо при введеному ключі та векторі ініціалізації виникне колізія (співпадіння імітовставок для різних вихідних даних), ця подія буде повторена не частіше ніж один раз на 2^{32} [3].

Друга умова визначає максимальну кількість можливих застосувань режиму GCM & GMAC для кожного з введених ключів, тобто ця умова задає обмеження на термін дії ключа. Разом з першою умовою вона визначає, що кожен ключ буде поєднаний із деяким вектором ініціалізації тільки один раз, тобто при обробленні різних вихідних даних одні й ті ж пари «ключ - вектор ініціалізації» застосовуватися двічі не будуть ніколи.

Таким чином можна зробити наступні висновки:

[3] Це твердження не враховує випадку, коли колізія імітовставок відбувається при різних параметрах криптоперетворення

- для кожного введеного ключа при обробленні різних вихідних даних кожен раз будуть застосовуватися різні вектори ініціалізації;

- для кожного введеного вектору ініціалізації при обробленні різних вихідних даних кожен раз будуть застосовуватися різні ключі;

- одні й ті ж самі вихідні дані можуть бути оброблені із застосуванням одного і того ж самого вектору ініціалізації, але ключі кожен раз повинні бути різними;

- одні й ті ж самі вихідні дані можуть бути оброблені із застосуванням одного і того ж самого ключа, але вектори ініціалізації кожен раз повинні бути різними.

Тобто якщо повторюється ключ – тоді не повинні повторюватися вектори ініціалізації, а якщо повторюється вектор ініціалізації – повинні бути різними ключі.

Позначимо через $Y_m = GHASH_H(X_1 \| X_2 \| \ ... \| X_m)$ результат поліноміального ґешування як на рис. 4, а через $T = MSB_t\left(GCTR_K(J_0, Y_m)\right) = MSB_t(Z \oplus Y_m)$ формовану імітовставку як на рис. 1, 2, де $Z = CHIP_K(J_0)$ результат зашифрування J_0 на ключі K.

Очевидно, що колізійні властивості формованих імітовставок залежать як від властивостей ґеш-значень Y_m, так і від результатів зашифрування Z. Колізійні властивості Y_m було оцінено у розділі 3. Оцінимо ймовірність співпадіння Z та кінцевих результатів T.

За визначенням блоковий симетричний шифр є бієктивним відображенням, тому різні вихідні значення J_0 після зашифрування будуть співставленні із різними значеннями Z, тобто колізій виникати не буде. Із умов формування векторів ініціалізації, які розглянуто вище, слідує, що навіть для однакових вихідних повідомлень $\left(X_1 \| X_2 \| ... \| X_m\right)$ і однакових ключів відповідні J_0 будуть різними, тобто ймовірність колізій проміжних значень Z буде завжди дорівнювати нулю.

Розглянемо тепер умови колізій імітовставок, зокрема такий випадок, коли для різних вихідних повідомлень $\left(X_1 \| X_2 \| ... \| X_m\right)$ на однаковому ключі K відповідні результати $Z \oplus Y_m$ будуть тотожними. Така умова формально може бути подана у вигляді:

$$Z \oplus Y_m = Z' \oplus Y'_m, \tag{10}$$

де Z і Y_m є проміжними результатами формування імітовставки для вихідного повідомлення $\left(X_1 \| X_2 \| ... \| X_m\right)$, а Z' і Y'_m є відповідними значеннями для іншого повідомлення $\left(X'_1 \| X'_2 \| ... \| X'_m\right) \neq \left(X_1 \| X_2 \| ... \| X_m\right)$.

Вище показано, що для таких повідомлень завжди виконується нерівність $Z \neq Z'$ і умова (10) буде виконуватися лише при $Y_m \oplus Y'_m = Z \oplus Z' \neq 0^{128}$, тобто лише тоді, коли не буде виникати колізій ґеш-значень Y_m та Y'_m.

Останній висновок найбільш вражаючий, оскільки метою поліноміального ґешування було саме зменшення ймовірності колізій формованих ґеш-значень. Точне значення ймовірності колізій формованих імітовставок буде визначатися комбінаторними властивостями шифру, тобто буде залежати від кількості випадків $Y_m \oplus Y'_m = Z \oplus Z'$. Однак можна з впевненістю стверджувати, що колізійні властивості імітовставок не будуть повторювати відповідні властивості ґеш-значень, що досліджувалися у розділі 3.

6. Висновки.

Проведений аналіз показав, що для використання в режимі шифрування Galois/Counter Mode and GMAC визначеної функції поліноміального ґешування $GHASH_H$ повинні бути застосовані певні обмеження, зокрема не можна використовувати нульовий субключ ґешування, оскільки при цьому значення функція завжди обертається в нуль для будь якого вхідного повідомлення, що створює передумови для зниження рівня імітостійкості. Це є неприпустимим, бо схема формування ґеш-значень спотворюється, формовані імітовставки не будуть залежати від вихідних даних колізійні властивості порушуються і не відповідають теоретичним оцінкам.

В ході досліджень було з'ясовано, що випадок формування нульового субключа поліноміального ґешування $GHASH_H$ виникає при наявності фіксованих точок шифру, тобто таких ключів, які шифрують нульову послідовність саму у себе. Отримані оцінки показують, що ймовірність виникнення фіксованої точки шифру не залежить від довжини ключа, вона визначається лише довжиною блоків даних, які обробляє шифр. При збільшенні потужності множини ключів кількість фіксованих точок також зростає і число «слабких» ключів, які призводять до виродженої роботи функції ґешування збільшується пропорційно потужності ключового простору. У випадку застосування шифру AES при довжині ключа $k = 128$ бітів один ключ буде створювати фіксовану точку; при довжині ключа $k = 192$ бітів загалом $2^{k-n} = 2^{64}$ ключів будуть створювати фіксовану точку; при довжині ключа $k = 256$ бітів $2^{k-n} = 2^{128}$ ключів відповідають випадку шифрування нульового вектору 0^{128} в те ж саме значення, тобто відповідають формуванню нульового субключа ґешування.

Ймовірність колізій формованих імітовставок буде визначатися комбінаторними властивостями шифру, тобто буде залежати від кількості випадків, коли бітова різниця зашифрованих на одному ключі різних констант буде дорівнювати бітовій різниці зформованих на цих же ключах ґеш-значень. Випадок колізій імітовставок можливий лише для різних ґеш-значень. Емпірична оцінка числа колізій та відповідні оцінки ймовірностей із застосуванням зменшених моделей шифрів є перспективним напрямком подальших досліджень. Перспективним також бачиться перевірка зроблених припущень щодо тотожності деяких властивостей випадкової підстановки певним характеристикам застосовуваного шифру, зокрема це безпосередньо відноситься до вивчення ймовірнісних властивостей виникнення фіксованої точки та, що еквівалентно, кількості циклів довжини $l = 1$.

ЛІТЕРАТУРА

1. ГОСТ 28147-89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. – М.: Изд-во стандартов, 1989. – 20 с.
2. ГОСТ Р ИСО/МЭК 10116-93. Информационная технология. Режимы работы для алгоритма n-разрядного блочного шифрования. [Електронний ресурс]. Режим доступу: http://docload.spb.ru
3. ISO/IEC 10116. Information technology – Security techniques – Modes of operation for an n-bit block cipher. [Електронний ресурс]. Режим доступу: http://www.iso.org
4. NIST Special Publication 800-38D. Block Cipher Modes. [Електронний ресурс]. Режим доступу: http://csrc.nist.gov
5. Stinson D.R. Universal hashing and authentication codes // Designs, Codes and Cryptography – 1994. - Volume 4, Issue 3. - pp 369-380.
6. Polynomial hashing: 4,588,985 United States Patent: H 03 M 7/00, field of search 340/347 DD / Carter J. L., Wegman M. N.; International Business Machines Corporation, Armonk, N.Y. – filed Dec. 30, 1983 – May 13, 1986
7. Raphael Chung-Wei Phan. Mini Advanced Encryption Standard (Mini-AES): A testbed for Cryptanalysis Students // Cryptologia, XXVI(4), October 2002, pp. 283-306.
8. Сачков В.Н. Введение в комбинаторные методы дискретной математики. – М.: Наука. Гл. ред. физ.-мат. лит., 1982. – 384 с.
9. Тронин С.Н. Введение в теорию групп. – Казань: Казанский государственный университет, 2006. – 100с.
10. Горбенко І.Д., Горбенко Ю.І. Прикладна криптологія. Теорія. Практика. Застосування: Підручник для вищих навчальних закладів. – Харків: Вид-во «Форт», 2013. – 880с.
11. Долгов В.И., Лисицкая И.В., Руженцев В.И. Анализ циклических свойств блочных шифров // Прикладная радиоэлектроника – 2007. – Т.6, №2 – С. 257-263.
12. Кузнецов А.А., Лисицкая И.В., Исаев С.А. Линейные свойства блочных симметричных шифров, представленных на украинский конкурс // Прикладная радиоэлектроника. – 2011. – Т.10, №2 – С. 135-140.
13. Сорока Л.С., Кузнецов А.А., Московченко И.В., Исаев С.А. Исследование дифференциальных свойств блочно-симметричных шифров. // Системи обробки інформації. – Х: ХУПС. –2010 – Вип. 6(87). – С. 286 – 294.
14. Долгов В.И., Родинко М.Ю. Блочные симметричные шифры – случайные подстановки. Комбинаторные показатели // Прикладная радиоэлектроника – 2013. – Т.12, №2 – С. 236-239.
15. Александров П.С. Введение в теорию групп. – М.: Наука, – 1980. – 145с.

Надійшла - 14.05.2014.