

АНАЛІЗ КОЛІЗІЙНИХ ВЛАСТИВОСТЕЙ РЕЖИМУ ВИРОБЛЕННЯ ІМІТОВСТАВОК ІЗ ВИБІРКОВИМ ГАМУВАННЯМ

О.О. КУЗНЕЦОВ, Д.В. ІВАНЕНКО, Є.П. КОЛОВАНОВА

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

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

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 розуміють такий метод його використання, який дозволяє реалізувати перетворення послідовності блоків відкритих даних у 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 вибірковою 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 гешування, 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.

1. АНАЛІЗ КРИПТОПЕРЕТВОРЕНЬ GCM & GMAC

Новий режим шифрування Galois/Counter Mode and GMAC призначено для забезпечення послуг конфіденційності та цілісності даних,

перш за все, в ході реалізації інформаційно-комунікаційних протоколів, зокрема в межах протоколів безпеки IPsec.

Структурна схема режиму вироблення імітовставки із вибірковою 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 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 (inc₃₂) блоку J₀, сформованого з вектора ініціалізації IV. Для забезпечення цілісності застосовується інший механізм, який засновано на функції гешування GHASH_H. Функцію гешування 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 шифротексту C із доданими даними A та реалізується функцією GCM-AD_K (IV, C, A, T), структурну схему якої наведено на рис. 2. При підтвердженні справжності (знов обчислений код T' дорівнює отриманому T) виконується розшифрування шифротексту 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 гешування даних із використанням функції GHASH_H та шифрування/розшифрування функцією GCTR_K. Розглянемо їх більш детально.

На рис. 3 зображено структурну схему функції GCTR_K для реалізації шифрування/розшифрування, де ICV – початковий блок лічильника; CB_i – i-й блок лічильника; inc – функція лічильника. Шифрування відбувається за допомогою затвердженого блокового симетричного шифру AES зі 128-бітним розміром блоку з використанням ключа шифрування K.

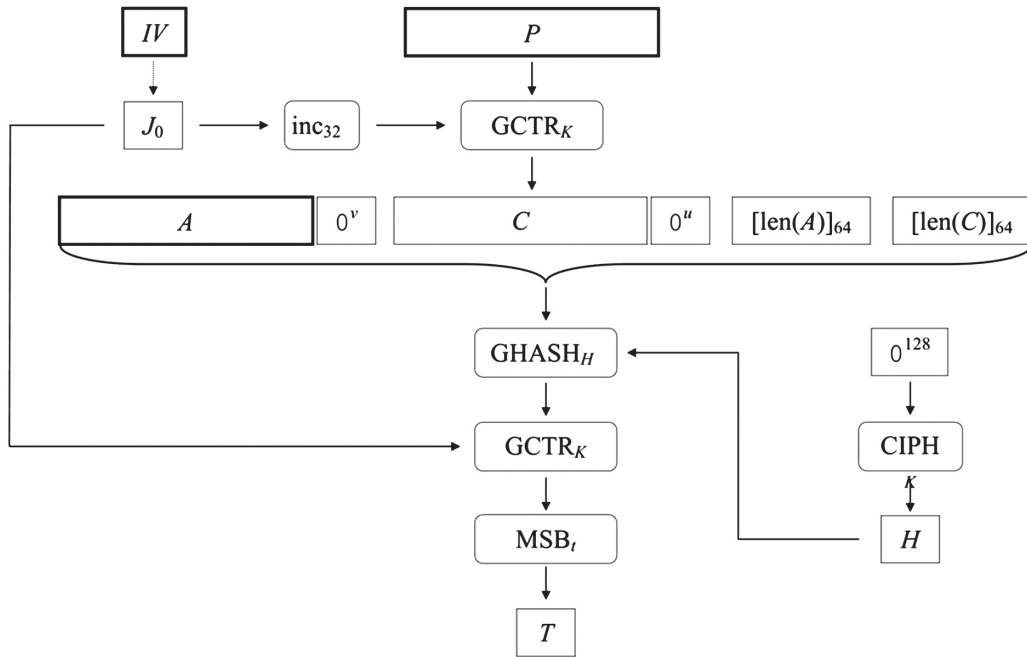


Рис. 1. $GCM - AE_K(IV, P, A) = (C, T)$

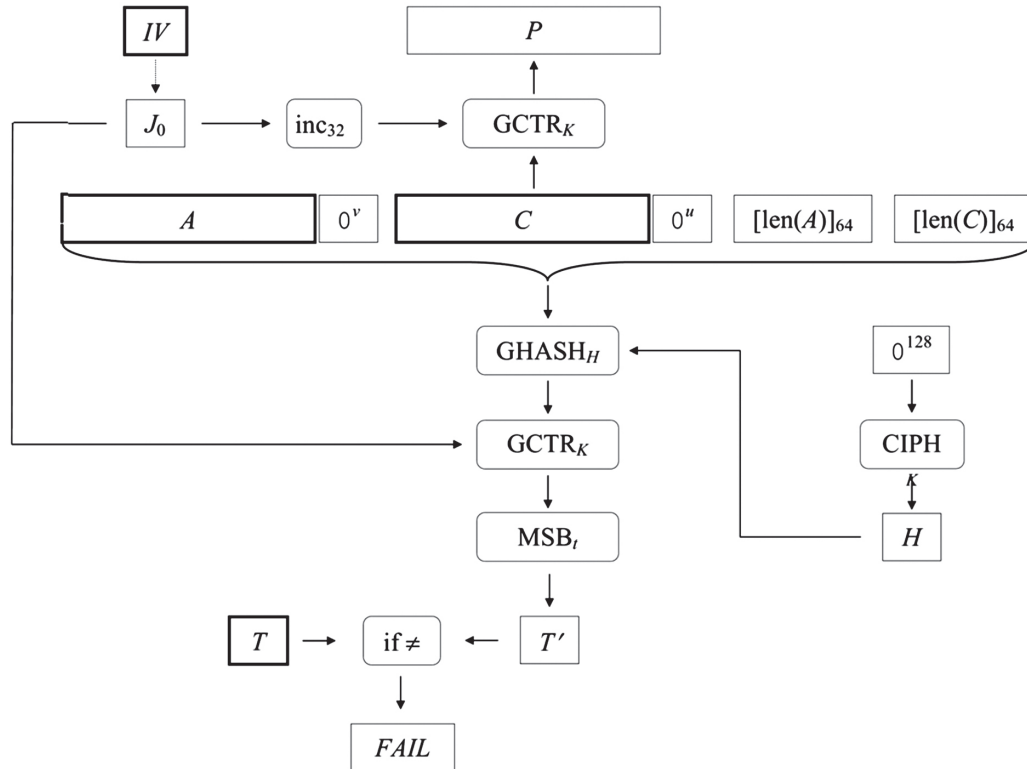


Рис. 2. $GCM - AD_k(IV, C, A, T) = P$ або $FAIL$

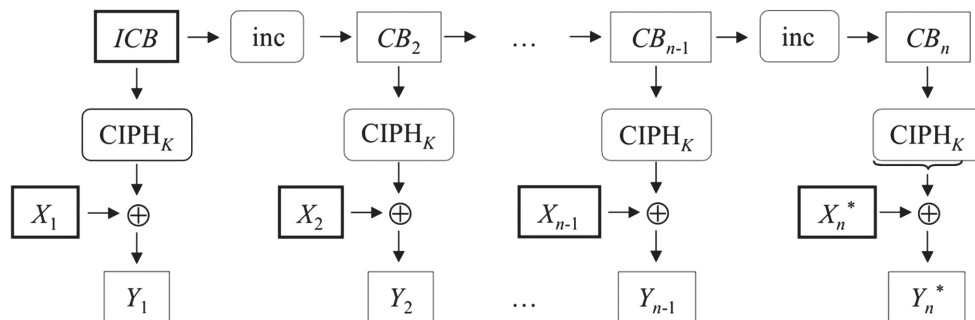


Рис. 3. $GCTR_K(ICB, X_1 || X_2 || \dots || X_n^*) = Y_1 || Y_2 || \dots || Y_n^*$

Для забезпечення цілісності інформації використовується функція гешування 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 гешування GHASH розраховує геш-значення з використанням схеми Горнера, структуру якої зображено на рис. 4.

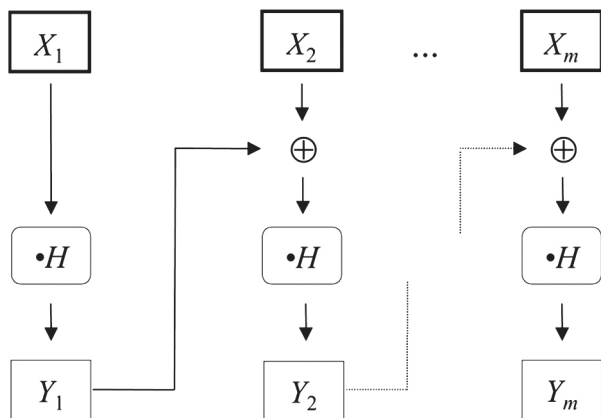


Рис. 4. $GHASH_H(X_1 \parallel X_2 \parallel \dots \parallel X_m) = Y_m$

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

- шифрування/розшифрування відкритого тексту Р функцією $GCTR_K$, яка по суті є деякою варіацією режиму гамування CTR [3, 4];
- обчислення геш-значення із використанням поліноміальної функції $GHASH_H$, яку використовують для стиснення шифротексту та автентифікованих даних А в єдиний блок;
- шифрування/розшифрування отриманого геш-значення, при цьому знову застосовується функція $GCTR_K$.

Отримане зашифроване геш-значення і є тим кодом справжності Т, який призначено для забезпечення цілісності та автентичності інформації. За вітчизняною термінологією код Т 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, геш-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. Під час проведення досліджень зосередимо увагу на вивченні властивостей геш-кодів, що сформовано функцією поліноміального гешування $GHASH_H$, та вихідних кодів справжності (імітовставок), що сформовано за результатом роботи режиму GCM & GMAC загалом.

2. ДОСЛІДЖЕННЯ КОЛІЗІЙНИХ ВЛАСТИВОСТЕЙ ПОЛІНОМІАЛЬНОГО ГЕШУВАННЯ $GHASH_H$

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

Ідеєю універсального гешування є визначення набору геш-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 геш-значення розраховані з використанням функції із визначеної множини. Ймовірність виникнення такої колізії можна підрахувати так [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 геш-значень, $|H|$ – потужність множини геш-функцій (або потужність множини ключів гешування, бо кожний ключ визначає окрему функцію із визначеної множини).

Для деякої геш-функції у 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 $\delta_y(X_a, X_b) = 1$ лише тоді, коли геш-значення від вхідних повідомлень X_a та X_b дадуть колізію. Для кінцевої множини H визначимо

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

Звідси $\delta_H(X_a, X_b)$ підраховує кількість геш-функцій (ключів гешування), які дають колізію для визначених X_a та X_b .

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

$$P_k = (n - 1) / |H|, \quad (1)$$

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

Дійсно, якщо правило гешування повідомлення $X = (X_1 \parallel X_2 \parallel \dots \parallel X_n)$ задається через обчислення у кінцевому полі $GF(2^k)$ значення поліному

$$Y_n = X_1 \cdot H_j^{n-1} \oplus X_2 \cdot H_j^{n-2} \oplus \dots \oplus X_n \cdot H_j^0, \quad (2)$$

де $H_j \in H$ – значення ключа гешування, $X_i, H_j \in GF(2^k)$, тоді колізія (співпадіння) геш-

значення Y_n із геш-кодом Y_n^* , який відповідає іншому повідомленню

$$X^* = (X_1^* \| X_2^* \| \dots \| X_n^*) \neq X = (X_1 \| X_2 \| \dots \| X_n)$$

відповідатиме випадку тотожності:

$$\begin{aligned} X_1 \cdot H_j^{n-1} \oplus X_2 \cdot H_j^{n-2} \oplus \dots \oplus X_n \cdot H_j^0 &= \\ = X_1^* \cdot H_j^{n-1} \oplus X_2^* \cdot H_j^{n-2} \oplus \dots \oplus X_n^* \cdot H_j^0 & \end{aligned}$$

для будь-якого введеного ключа гешування H_j .

Останнє рівняння перепишемо у канонічному вигляді:

$$\begin{aligned} (X_1 \oplus X_1^*) \cdot H_j^{n-1} \oplus (X_2 \oplus X_2^*) \cdot H_j^{n-2} \oplus \dots \\ \oplus (X_n \oplus X_n^*) \cdot H_j^0 = 0, \end{aligned} \quad (3)$$

отже колізія геш-значень, тобто подія $Y_n = Y_n^* |_{X \neq X^*}$ виникатиме 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 урахуванням їхньої кратності, тобто на безлічі значень $H_j \in H$ не більше $n-1$ різних ключів гешування обертаються в нуль ліву частину рівності (3). Таким чином, при рівноймовірному обранні ключів гешування $H_j \in H$ ймовірність колізії $P_k = P(Y_n = Y_n^* |_{X \neq X^*})$ визначатиметься за виразом (1).

Беручи до уваги зазначене вище, можна зробити висновок, що, обираючи підходяще значення потужності $|H|$ множини геш-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 наведено на рис. 5.

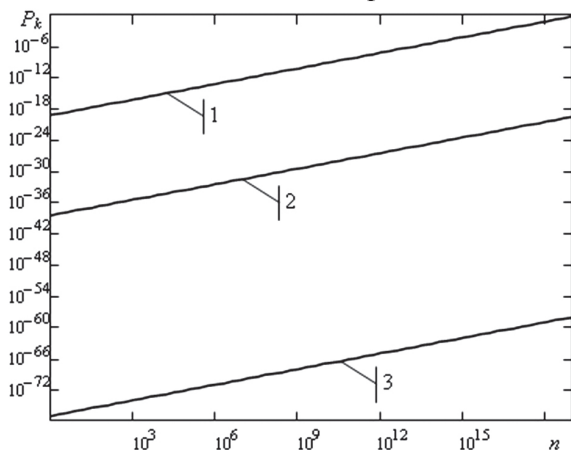


Рис. 5. Залежність ймовірності виникнення колізії від кількості блоків вхідного повідомлення:

- 1) $|H| = 2^{64}$; 2) $|H| = 2^{128}$; 3) $|H| = 2^{256}$

Як видно з наведених залежностей ймовірності колізій в ході застосування поліноміальної схеми гешування значно підвищується із збільшенням довжини вхідного повідомлення. При фіксованій довжині ключів гешування (та, відповідно, потужності $|H|$) ця залежність накладає додаткові обмеження на довжину пові-

домлень, які хешуються. Наприклад, для заданої ймовірності виникнення колізій $P_k = 2^{-32}$ та при довжині ключа гешування 128 бітів загальна кількість блоків повідомлення має задовольняти вимогу $n \leq 2^{96}$.

Обчислення геш-значення за правилом (2) має певні недоліки. Наприклад, якщо вхідне повідомлення складається лише з одного блоку, тоді формула (2) набуває вигляду $Y_1 = X_1 \cdot H_j^0 = X_1$ і правило гешування є тотожністю, тобто геш-значення дорівнюватиме блоку повідомлення і ключ гешування H_j при обчисленні Y_1 зовсім не використовується. Можливо саме тому в специфікації GCM & GMAC [4] застосовується дещо змінена поліноміальна форма обчислення геш-коду, а саме (див. рис. 4):

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

Випадок колізії спостерігатиметься так само під час виконання рівності:

$$\begin{aligned} X_1 \cdot H_j^m \oplus X_2 \cdot H_j^{m-1} \oplus \dots \oplus X_m \cdot H_j^1 &= \\ = X_1^* \cdot H_j^m \oplus X_2^* \cdot H_j^{m-1} \oplus \dots \oplus X_m^* \cdot H_j^1, \end{aligned}$$

яку запишемо у вигляді

$$\begin{aligned} (X_1 \oplus X_1^*) \cdot H_j^m \oplus (X_2 \oplus X_2^*) \cdot H_j^{m-1} \oplus \dots \\ \oplus (X_m \oplus X_m^*) \cdot H_j^1 = 0, \end{aligned}$$

що після скорочення на H_j^1 повністю відповідає (3) із оцінкою ймовірності колізій (1).

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

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

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

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

3. ОЦІНКА ЙМОВІРНІСТІ ВИНИКНЕННЯ НУЛЬОВОГО СУБКЛЮЧА ГЕШУВАННЯ GHASH_n.

Розглянемо множину всіх бієктивних перетворень множини $Y = \{y_1, y_2, \dots, y_n\}$ 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

$$s^{-1} \cdot s(y) = s \cdot s^{-1}(y) = e(y), \quad y \in Y,$$

де $e(y) \in S_n$ – одинична підстановка, тобто $e(y) = y$ для всіх $y \in Y$.

Введемо такі позначення: $s \cdot s \cdot \dots \cdot s = s^k$, $s^{-1} \cdot s^{-1} \cdot \dots \cdot s^{-1} = s^{-k}$, 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 вважатимемо $y \sim y'$ 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 $y \sim y'$ випливає $y' \sim y$, оскільки із рівності $y' = s^j(y)$ випливає, що $y = s^{-j}(y')$;
- із $y \sim y'$ та $y' \sim y''$ випливає, 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 так:

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

де $s_i^l(y) = y$.

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

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

Наприклад, підстановка s степеня 4-го 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:

$$s_1 = (y_1) = (1), \quad l_1 = 1;$$

$$s_2 = (y_2, s_2(y_2)) = (2, 4), \quad l_2 = 2;$$

$$s_3 = (y_3) = (3), \quad l_3 = 1,$$

тобто маємо такий розклад:

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

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

$$s = (y_1)(y_2) \dots (y_{\alpha_1})(y'_1, y''_1)(y'_2, y''_2) \dots (y'_{\alpha_2}, y''_{\alpha_2}) \dots,$$

$$1\alpha_1 + 2\alpha_2 + \dots + n\alpha_n = n.$$

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

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

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

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

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

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 відповідатиме випадковому і рівномірно вибраному шифру тексту для кожного відкритого тексту при будь-якому ключі, тобто на всіх можливих варіантах відображень відкритого тексту у шифрограму.

За визначенням блоковий симетричний шифр є функцією відображення множини текстів і множини ключів у множину шифртекстів: $E: K \times M \rightarrow E$, 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 $E: K \times M \rightarrow E$ має 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!. Утім розмір ключа більшості блокових шифрів не перевищує невеликого числа, кратного розміру блоку, відповідно такі шифри можуть забезпечити лише невелику частку від повної кількості можливих перестано-

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

вок. Наприклад, для 128-бітного шифру AES маємо $2^{128} \approx 2^{128 \cdot 2^{128}}$ можливих перестановок 128-бітових блоків, з яких, залежно 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 і 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 у випадково вибраній підстановці $s \in S_n$. Оцінимо ймовірність випадкової події $\xi_n = k$, тобто такого випадку, коли у випадково вибраній підстановці спостерігатиметься точно k циклів (див. вираз (5)). З формули (6) безпосередньо впливає вираз для точного розподілу ймовірностей через числа Стірлінга першого роду:

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

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

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

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

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

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

У 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. Водночас для оцінки ймовірності виникнення нульового субключа гешування 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 рівномірній підстановці степеня n . Тоді розподіл ймовірностей випадкової події $\chi_{n,l} = k$ визначається як:

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

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

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

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 мають переводити фіксований елемент y з множини Y сам у себе. Тобто необхідно оцінити кількість таких підстановок $s \in S_n$, які для визначеного $y \in Y$ містять цикли $s(y) = y$ довжини $l=1$. У криптографії при розгляді блокових симетричних криптоперетворень такі випадки прийнято називати фіксованими точками підстановки [11].

Для $l=1$ формула (7) набуде вигляду

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

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

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

разів, тобто кількість фіксованих точок $s(y) = y$ для визначеного $y \in Y$ визначатиметься 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 рівномірній підстановці степеня n матиме вигляд:

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

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

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

що при $m=1$ співпадає з (8) ².

² $(n)_m = n(n-1)\dots(n-m+1)$ – загальноприйняте позначення спадаючого факторіалу

Припустимо, що n -бітний блоковий симетричний шифр, який складається з деякої підмножини симетричної групи підстановок степеня 2^n , відповідає ймовірнісним властивостям випадкової підстановки, зокрема прийемо припущення, що на всій множині ключів шифру, які задають обрання конкретної підстановки s із S_{2^n} , ймовірність обрати підстановку з m фіксованими точками визначатиметься за (9), тобто дорівнюватиме

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

Тоді кількість ключів шифру, які призведуть до появи m фіксованих точок $s(y) = y$ для визначених блоків відкритого тексту $y \in Y$ визначатиметься 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 створюватиме фіксовану точку, тобто, наприклад, спостерігатиметься 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 створюватимуть фіксовану точку, тобто випадок із шифруванням нульового вектору 0^{128} в нульове значення $H_j = 0$ спостерігатиметься 2^{64} разів;

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

Отримані оцінки дозволяють стверджувати, що в ході виконання 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 геш-значень, які вже не залежатимуть від вихідних даних, колізійні властивості порушуються і не відповідають теоретичним оцінкам. Очевидно, що зі збільшенням потужності множини ключів кількість фіксованих точок також зростатиме і число так би мовити «слабких» ключів, які призводять до виродженої роботи функції гешування, збільшується пропорційно потужності ключового простору.

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

4. 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 геш-значення додається результат зашифрування деякого вектору J_0 , який отримано із вектора ініціалізації IV (див. рис. 1, 2). Тобто фактично, до сформованого геш-значення додається зашифрована константа, яка формується встановленим порядком.

Відповідно до специфікації режиму GCM & GMAC на формування вектора ініціалізації (за вітчизняною термінологією – синхросилка) накладаються такі обмеження:

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

– для кожного заданого ключа повне число формованих імітовставок для будь-яких вихідних повідомлень не має 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} .

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

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

з першою умовою вона визначає, що кожен ключ буде поєднаний із деяким вектором ініціалізації тільки один раз, тобто в ході оброблення різних вихідних даних одні й ті ж пари «ключ – вектор ініціалізації» застосовуватися двічі не будуть ніколи.

Таким чином можна зробити такі висновки:

– для кожного введеного ключа в ході оброблення різних вихідних даних щоразу застосовуватимуться різні вектори ініціалізації;

– для кожного введеного вектора ініціалізації в ході оброблення різних вихідних даних щоразу застосовуватимуться різні ключі;

– одні й ті ж самі вихідні дані можуть бути оброблені із застосуванням одного і того ж самого вектора ініціалізації, але ключі щоразу мають бути різними;

– одні й ті ж самі вихідні дані можуть бути оброблені із застосуванням одного і того ж самого ключа, але вектори ініціалізації щоразу мають бути різними.

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

Позначимо через

$$Y_m = \text{GHASH}_H(X_1 \| X_2 \| \dots \| X_m)$$

результат поліноміального гешування як на рис. 4, а через $T = \text{MSB}_t(\text{GCTR}_K(J_0, Y_m)) = \text{MSB}_t(Z \oplus Y_m)$ формовану імітовставку як на рис. 1, 2, де $Z = \text{CHIP}_K(J_0)$ результат зашифрування J_0 на ключі K .

Очевидно, що колізійні властивості формованих імітовставок залежать як від властивостей геш-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 не виникатиме. Із умов формування векторів ініціалізації, які розглянуто вище, випливає, що навіть для однакових вихідних повідомлень $(X_1 \| X_2 \| \dots \| X_m)$ і однакових ключів відповідні J_0 будуть різними, тобто ймовірність колізій проміжних значень Z завжди дорівнюватиме нулю.

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

$$Z \oplus Y_m = Z' \oplus Y'_m, \quad (10)$$

де Z і Y_m є проміжними результатами формування імітовставки для вихідного повідомлення $(X_1 \| X_2 \| \dots \| X_m)$, а Z' і Y'_m є відповідними

значеннями для іншого повідомлення $(X'_1 \| X'_2 \| \dots \| X'_m) \neq (X_1 \| X_2 \| \dots \| X_m)$.

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

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

ВИСНОВКИ

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

В ході досліджень було з'ясовано, що випадок формування нульового субключа поліноміального гешування GHASH_H виникає за 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. Зі збільшенням потужності множини ключів кількість фіксованих точок також зростає і число «слабких» ключів, які призводять до виродженої роботи функції гешування збільшується пропорційно потужності ключового простору. У випадку застосування шифру AES при довжині ключа $k = 128$ бітів один ключ створюватиме фіксовану точку; при довжині ключа $k = 192$ бітів загалом $2^{k-n} = 2^{64}$ ключів створюватимуть фіксовану точку; при довжині ключа $k = 256$ бітів $2^{k-n} = 2^{128}$ ключів відповідають випадку шифрування нульового вектора 0^{128} в те ж саме значення, тобто відповідають формуванню нульового субключа гешування.

Ймовірність колізій формованих імітовставок визначатиметься комбінаторними властивостями шифру, тобто залежатиме від кількості

випадків, коли бітова різниця зашифрованих на одному ключі різних констант дорівнюватиме бітовій різниці сформованих на цих же ключах геш-значень. Випадок колізій імітовставок можливий лише для різних геш-значень. Емпірична оцінка числа колізій та відповідні оцінки ймовірностей із застосуванням зменшених моделей шифрів є перспективним напрямком подальших досліджень. Перспективним також є 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. [Электронный ресурс]. Режим доступа: <http://docload.spb.ru>
- [3] ISO/IEC 10116. Information technology — Security techniques — Modes of operation for an n -bit block cipher. [Электронный ресурс]. Режим доступа: <http://www.iso.org>
- [4] NIST Special Publication 800-38D. Block Cipher Modes. [Электронный ресурс]. Режим доступа: <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. P. 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. — 100 с.
- [10] Горбенко І.Д., Горбенко Ю.І. Прикладна криптологія. Теорія. Практика. Застосування: Підручник для вищих навчальних закладів. — Харків: Вид-во «Форт», 2013. — 880 с.
- [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. // Системы обработки информации. — X: 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. — 145 с.

Надійшла до редколегії 13.08.2014

Кузнецов Олександр Олександрович, фото та відомості про автора див. на стор. 207.

Іваненко Дмитро Вікторович, фото та відомості про автора див. на стор. 207.

Колованова Євгенія Павлівна, фото та відомості про автора див. на стор. 207.

УДК 004.056.55

Анализ коллизионных свойств режима выработки имитовставок с выборочным гаммированием / А. А. Кузнецов, Д. В. Иваненко, Е. П. Колованова // Прикладная радиоэлектроника: научн.-техн. журнал. — 2014. — Том 13. — № 3. — С. 226–234.

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, блочное симметричное шифрование, цикл, подстановка.

Ил.: 5. Библиогр.: 15 назв.

UDK 004.056.55

Analysis of collision properties of the mode of generating message authentication codes with selective gamming / A.A. Kuznetsov, D.V. Ivanenko, E.P. Kolovanova // Applied Radio Electronics: Sci. Journ. — 2014. — Vol. 13. — № 3. — P. 226–234.

The paper investigates the mode of generating a message authentication code with selective gamming which is designed to ensure the integrity and confidentiality of communications. Basic cryptographic transformations are considered that are used to implement this mode, theoretical estimates of collisions probability are provided. Offers to improve the mode under study are justified when used in information and communication systems.

Keywords: collision, hashing, authentication code, cryptographic protection, block symmetric encryption, cycle, substitution.

Fig.: 5. Ref.: 15 items.