УДК 681.3.06:519.248.681

О. В. ПОТІЙ, канд. техн. наук, Ю. І. ГОРБЕНКО, Є. В. ПОПОВИЧ

МЕТОД ОЦІНКИ ІМОВІРНОСТЕЙ КОЛІЗІЙ У БЕЗУМОВНО СТІЙКИХ ТА ОБЧИСЛЮВАЛЬНО СТІЙКИХ КРИПТОСИСТЕМАХ

В останні роки знаходять застосування системи криптографічного захисту інформації, що забезпечують безумовну або обчислювальну стійкість. В безумовно стійких системах для шифрування (зашифрування/розшифрування) використовуються ключові послідовності K, що будуються на основі випадкових процесів [1,2]. При цьому довжина ключової послідовності $l_{кл}$ повинна бути не менше довжини l_M шифруємої інформації. Зашифрування здійснюється згідно правила

$$C_i = (M_i + K_i) \bmod m, \tag{1}$$

де C_i – i-й зашифрований символ; M_i – i-й – символ відкритої інформації; K_i – i-й – символ ключової послідовності; m – модуль криптографічного перетворення.

Розшифрування здійснюється за правилом

$$M_i = (C_i - K_i) \bmod m. \tag{2}$$

Одним із слабких місць такої криптосистеми є можливість перекриття ключових послідовностей деякої критичної довжини $l_{кр}$ в просторі чи часі. Тут та далі за текстом таке перекриття ключових послідовностей будемо називати колізією гами шифру. Дійсно, якщо $C_i' = (M_i' + K_i) \bmod m$, тобто відбулося перекриття K_i з довжиною $l_{кр}$, то

$$C_i + C_i' = (M_i + K_i + M_i' + K_i) \bmod m = (M_i + M_i') \bmod m.$$

Далі розклад суми $(M_i + M_i') \bmod m$ є практично поліноміальною задачею [3].

Тому, на наш погляд, дуже важливою є задача дослідження можливостей виникнення таких колізій та визначення обмежень на величину $l_{кр}$ в залежності від допустимої імовірності виникнення колізії P_k та числа k сформованих відрізків ключової послідовності з довжиною $l_i \le l_{kp}$. Слід зазначити, що, не враховуючи методи формування гами шифру, дані показники будуть справедливі для будь-якого шифру гамування.

Розглядаючи клас блокових шифрів, необхідно зазначити, що блокові симетричні криптоперетворення (шифри) можуть використовуватись принаймі в п'ятьох режимах [4]. Запропоновано ще декілька режимів криптографічних перетворень [4], що здійснюються на основі блокового симетричного шифрування. Основними з них є слідуючи режими:

- блокового шифрування,
- потокового шифрування зі зворотнім зв'язком по шифротексту,
- потокового шифрування зі зв'язком по гамі шифруючій,
- потокового шифрування з лічильником,
- потокового шифрування з лічильником та автентифікацією,
- блокового шифрування зі зв'язком блоків (виробки імітоприкладки).

Проведемо аналіз можливостей виникнення колізій в режимі блокового шифрування. В цьому режимі результат зашифрування є деякий блок C_i, причому

$$C_i = F(M_i, K_j, R_{i-v}) \tag{3}$$

де M_i – блок відкритого тексту; K_j – ключ зашифрування; R_{i-v} – сінхромаркер (інформація зворотного зв'язку або стан лічильника).

Метою даної статті є обґрунтування та розробка методу оцінки колізій та розробка рекомендацій на допустимі мінімальні значення $l_{кр}$ для безумовно стійких та обчислювально стійких симетричних шифрів.

1 Оцінка імовірностей колізій для безумовно стійких шифрів

Нехай генератор $Г$ формує випадкові ключові послідовності K_i, еквівалентна двійкова довжина яких є l_i. На виході такого генератора може з'явитись рівноімовірно та незалежно кожна із $n = 2^{l_i}$ ключових послідовностей. Нехай також необхідно сформувати k ключових послідовностей довжини l_i. Необхідно знайти математичне співвідношення, що зв'язує між собою такі величини, як $n(l_i)$, k та допустиму імовірність колізій P_k, а також визначити допустиму довжину l_k ключових послідовностей з урахуванням можливих застосувань безумовно стійких криптосистем та допустимих імовірностей колізії.

Розв'язок виконаємо на основі узагальненого «парадоксу дня народження» [5].

В даному генераторі усього може бути сформовано $\bar{n} = 2^{l_i}$ ключових послідовностей K_i, і усі вони є рівноімовірними та незалежними. Нехай усього сформовано k ключових послідовностей довільної довжини l_i. Визначимо імовірність події, що при цьому відбудеться колізія, тобто два ключі із k співпадуть.

Визначимо загальну множину подій N_Σ, які можливі при формуванні послідовності $K_1, K_2, ..., K_k$. Оскільки усі n значень є незалежними та рівноімовірними, то

$$N_\Sigma = \underbrace{n \cdot n \cdot n \cdot ... \cdot n}_{k} = n^k. \tag{4}$$

Далі визначимо множину подій, її величину, при якій не буде жодної колізії.

В перший раз без колізії може відбутися n подій, в другий n-1, третій n-2 і т.д., і при k-й події $n-(k-1)$. Оскільки усі ці події рівноімовірні та незалежні, то загальне число подій при k експериментах, при яких колізій не буде, можна визначити як

$$N_k = n \cdot (n-1)(n-2)...(n-(k-1)). \tag{5}$$

Знаючи N_k та N_Σ, імовірність того, що при експериментах формування ключових послідовностей довжини l_i колізії не буде, визначимо імовірність відсутності колізії $P_b(k,n)$ як

$$P_b(k,n) = \frac{N_k}{N_\Sigma} = \frac{n(n-1)(n-2)...(n-(k-1))}{n^k} = \frac{n!}{n^k (n-k)!}. \tag{6}$$

Оскільки імовірність колізії $P_k(k,n)$ та відсутність колізії складають повну групу подій, то

$$P_k(k,n) = 1 - P_b(k,n). \tag{7}$$

Враховуючи складність розрахунків під час роботи з великими числами, підставимо (6) в (7):

$$P_k(k,n) = 1 - \frac{n(n-1)(n-2)...(n-(k-1))}{n \quad n \quad n \qquad\qquad n} =$$
$$= 1 - 1\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)...\left(\frac{n-(k-1)}{n}\right) = \qquad (8)$$
$$= 1 - \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right).$$

Оскільки в реальних випадках $k < 0{,}1n$, то для спрощення (8) можна зробити заміну $(1-x) \le e^{-x}$, в результаті маємо

$$P_k(k,n) = 1 - e^{-\frac{1}{n}} \cdot e^{-\frac{2}{n}} ... e^{-\frac{k-1}{n}} = 1 - e^{-\frac{1}{n(1+2+3+...+k-1)}} = 1 - e^{-\frac{k(k-1)}{2n}} = 1 - e^{-\frac{k(k-1)}{2 \cdot 2^{l_i}}}. \qquad (9)$$

Таким чином при вказаних обмеженнях одержано аналітичне співвідношення, що зв'язує між собою імовірність колізії $P_k(k,n)=P_k$, число сформованих ключових послідовностей k довжиною $l_i \ge l_{kp}$ та загальне число імовірних послідовностей $n = 2^{l_i}$.

Наявність співвідношення (9) дозволяє:

1. Оцінити імовірності колізій, змінюючи значення k та l_i.
2. Визначити критичне значення l_{kp} в залежності від допустимого значення імовірності колізії P_k для різних величин k (вони визначаються практичними додатками).
3. Визначити обмеження на число сформованих в системі (просторі та часі) послідовностей k_o, при яких імовірність колізії не перевищує P_k, якщо довжина ключової послідовності є $l_i \ge l_{kp}$.

Попередній аналіз допустимих джерел показав, що можливість здійснення на безумовно стійкі криптосистеми атаки методом колізій у них не враховано.

Розглянемо порядок розв'язку перелічених вище задач та проведемо їх дослідження.

При оцінці величин імовірності колізії можна використовувати вираз (9). При цьому для випадку, коли $k^2 >> k$, його можна спростити до виду

$$P_k(k,n) \approx 1 - e^{-\frac{k^2}{2^{(l_i+1)}}}. \qquad (10)$$

В табл. 1 наведені значення імовірностей колізії P_k в залежності від k та l_i.

Таблиця 1

l \ k	2	16	32	64	128	256	1024	65536	10^9	10^{12}
8	0,004	0,38	0,867	0,999	~1	~1	---	---	---	---
16	$3.05 \cdot 10^{-5}$	$1.9 \cdot 10^{-3}$	$7.7 \cdot 10^{-3}$	0,031	0,118	0,393	0,999	~1	---	---
32	$4{,}6 \cdot 10^{-10}$	$2{,}9 \cdot 10^{-8}$	$1{,}1 \cdot 10^{-7}$	$4{,}7 \cdot 10^{-7}$	$1{,}9 \cdot 10^{-6}$	$7{,}6 \cdot 10^{-6}$	$1{,}2 \cdot 10^{-4}$	0.393	~1	---
64	$9{,}7 \cdot 10^{-20}$	$6{,}2 \cdot 10^{-18}$	$2{,}5 \cdot 10^{-17}$	$9{,}9 \cdot 10^{-17}$	$3{,}9 \cdot 10^{-16}$	$1{,}7 \cdot 10^{-15}$	$2.8 \cdot 10^{-14}$	$1.1 \cdot 10^{-10}$	0,02	~1
128	~0	~0	~0	~0	~0	~0	~0	$1{,}7 \cdot 10^{-29}$	$1{,}2 \cdot 10^{-21}$	$1{,}2 \cdot 10^{-15}$
256	~0	~0	~0	~0	~0	~0	~0	~0	~0	~0
512	~0	~0	~0	~0	~0	~0	~0	~0	~0	~0

Аналіз даних табл. 1 дозволяє зробити такі висновки. Незалежно від довжини одноразової гами завжди існує ймовірність виникнення колізії і, як наслідок, здійснення атаки на безумовно стійку систему. Тому, якщо довжина шифруємого повідомлення не перевищує 64 бітів, то існують реальні імовірності колізій і, як наслідок, розкриття безумовно стійкої системи. Так при довжині гами l=64 бітів n=10^{13}, при k=10^{12} імовірність колізії дорівнює 1. Тому довжини повідомлень, що зашифровуються одноразовою гамою, повинні складати не менше 128 бітів.

Критичне значення l_{kp} можна знайти із співвідношень (9) або (10), подавши його у виді

$$1-P_k = e^{-\frac{k(k-1)}{2^{l_i+1}}} = e^{-\frac{k(k-1)}{2n}}. \qquad (11)$$

Прологарифмувавши даний вираз, маємо

$$\ln(1-P_k) = -\frac{(k^2+k)}{2^{l_i+1}} = -\frac{(k^2+k)}{2n}. \qquad (12)$$

Далі із (12) спочатку знаходимо n_{kp} як

$$n_{kp} = -\frac{(k^2+k)}{(2\ln(1-P_k))}. \qquad (13)$$

Критичне значення l_{kp} знаходимо із виразу $n_{kp} = 2^{l_{kp}}$, отже

$$l_{kp} = \log_2 n_{kp}. \qquad (14)$$

В табл. 2 наведено значення l_{kp} мінімально допустимих довжин ключових послідовностей (бітів) для безумовно стійких криптосистем в залежності від величин k та P_k.

Таблиця 2

P_k \ k	2	8	16	32	64	128	256	1024	32768	65536	10^6	10^9	10^{12}
10^{-3}	11	15	17	19	20	22	24	28	38	40	48	68	88
10^{-6}	21	25	27	28	30	32	34	38	48	50	58	78	98
10^{-9}	31	35	36	38	40	42	44	48	58	60	68	88	108
10^{-12}	41	45	46	48	50	52	54	58	68	70	78	98	118
10^{-16}	54	58	60	62	64	66	68	72	82	84	91	111	131

Таким чином ми отримали значення мінімальних довжин блоків, на які повинна буті поділена інформація з огляду на її загальний обсяг та імовірність виникнення колізій. Добуток значень k – кількості блоків та l – довжини блоку, отримані з таблиці, дадуть загальну довжину відкритого тексту, який може бути зашифрований, з відповідною імовірністю виникнення колізій. Слід зазначити, що під блоком інформації розуміється

таким чином скомпонована інформація, що її подальша обробка або аналіз можливі лише цілим блоком. Такі властивості інформація може отримати завдяки рандомізації.

Розв'язок третьої задачі, тобто визначення величини k_o, можна також здійснити, використавши (12). В результаті маємо

$$-\frac{(k^2+k)}{2n} = \ln(1-P_k)$$

або

$$k^2 + k + 2n\ln(1-P_k) = 0. \tag{15}$$

При значенні $k^2>>k$ можна використовувати співвідношення

$$k^2 + 2n\ln(1-P_k) = 0.$$

Тоді оцінкою величини k є значення

$$k = \sqrt{-2n\ln(1-P_k)} = \sqrt{-2^{l+1}\ln(1-P_k)}. \tag{16}$$

При відомих P_k та n або l можна обчислити величину k та загальну довжину відкритого тексту як

$$L = l \cdot k = l\sqrt{-2^{l+1}\ln(1-P_k)}. \tag{17}$$

Таким чином ми отримали показник загальної довжини тексту, який можливо зашифрувати в залежності від довжини блоків відкритого тексту та ймовірності колізій.

2 Оцінка імовірностей колізій для обчислювально стійких шифрів

Аналіз показує, що під час використання блокових шифрів може виникнути ситуація, коли два блоки криптограми дорівнюють один одному, тобто $C_i=C_j$. Виходячи з властивостей блокових шифрів, такий збіг може статися лише за наступних умов:

1. $M_i \neq M_j$, $K_i \neq K_j$, $R_i \neq R_j$.
2. $M_i = M_j$, $K_i \neq K_j$, $R_i \neq R_j$.
3. $M_i \neq M_j$, $K_i = K_j$, $R_i \neq R_j$.
4. $M_i \neq M_j$, $K_i \neq K_j$, $R_i = R_j$.
5. $M_i = M_j$, $K_i = K_j$, $R_i = R_j$.

Розглядаючи дані умови, слід зазначити, що по-перше інтерес викликає імовірність колізії ключів, тобто $C_i=C_j$ за умов $M_i \neq M_j$, $K_i = K_j$, $R_i \neq R_j$, та імовірність колізії текстів, тобто $C_i=C_j$ за умов $M_i = M_j$, $K_i \neq K_j$, $R_i \neq R_j$. Якщо M_i рандомізовано, то це дозволяє розглядати M_i як рівноімовірні та незалежні реалізації. Крім того, оскільки K_j є випадковим, то і C_i можна вважати випадковими, рівноімовірними, незалежними (однорідними) на повній множині подій. Тому задачу можливо звести до розрахунку імовірності появи на виході шифратора двох однакових криптограм в залежності від випадкового параметра K, який може бути блоком відкритого тексту або ключем. Використовуючи співвідношення (11), імовірність колізії можна визначити як

$$P_k(k,n) = 1 - e^{\frac{-k(k-1)}{2n}} = P_k, \tag{18}$$

де k – кількість оброблених або використаних блоків (кількість експериментів); n – розмір повної множини виходу перетворювача F, причому $n = 2^{l_б}$, де $l_б$ – довжина блоку.

Після простих перетворень (18) може бути подано у вигляді аналогічно (15), тобто

$$k^2 + k + 2n\ln(1 - P_k) = 0.$$

Оскільки для блокового криптоперетворення значення $l_б$ фіксоване, то $n = 2^{l_б}$ і також є величиною сталою. Тому для блокових криптоперетворень можна розв'язувати два типи задач. Визначення імовірностей колізії для різних значень K, а також визначення величини k, при якій імовірність колізії не перевищує допустимого значення P_g.

Для розв'язку першої задачі використовуємо співвідношення (10). В табл. 3 наведено значення імовірностей колізії в залежності від величини k для $l_б$=64, 128, 256 бітів.

Таблиця 3

$l_б$ \ k	2	2^8	2^{16}	2^{32}	2^{64}	2^{96}	2^{128}	2^{192}	2^{256}
64	$9{,}5\cdot10^{-20}$	$1{,}55\cdot10^{-15}$	$1{,}02\cdot10^{-10}$	0,393	~1	---	---	---	---
128	0	0	$2{,}2\cdot10^{-29}$	$9{,}49\cdot10^{-20}$	0,393	~1	~1	---	---
256	0	0	0	0	0	$2{,}37\cdot10^{-20}$	0,393	~1	~1

Аналіз даних табл. 3 показує, що імовірність колізії суттєво залежить від довжини блоку $l_б$, і довжини блоків повинні складати не менше 128 бітів. Цей висновок дозволяє зрозуміти, чому в європейському проекті Nessie для стійких шифрів довжина блоку відкритого тексту повинна складати не менше 128 бітів.

Для розв'язку другої задачі краще використовувати (15). Причому, якщо $k^2 >> k$, то його можна спростити до виду (16).

В табл. 4 наведені значення величин k в залежності від довжин l блоку та допустимих імовірностей колізій P_k.

Дані, наведені в табл. 4, дозволяють зрозуміти причини та необхідність збільшення в перспективних блокових симетричних шифрах довжин блоку $l_б$.

Таблиця 4

$l_б$ \ P_k	10^{-16}	10^{-12}	10^{-9}	10^{-6}	10^{-3}	10^{-1}	0,5
64	64	$6{,}07\cdot10^{3}$	$1{,}92\cdot10^{5}$	$6{,}07\cdot10^{6}$	$1{,}92\cdot10^{8}$	$1{,}97\cdot10^{9}$	$5{,}05\cdot10^{9}$
128	$8{,}2\cdot10^{11}$	$8{,}2\cdot10^{13}$	$2{,}6\cdot10^{15}$	$8{,}2\cdot10^{16}$	$8{,}2\cdot10^{17}$	$8{,}4\cdot10^{18}$	$2{,}1\cdot10^{19}$
256	$5{,}07\cdot10^{30}$	$4{,}8\cdot10^{32}$	$1{,}5\cdot10^{34}$	$4{,}8\cdot10^{35}$	$1{,}5\cdot10^{37}$	$1{,}5\cdot10^{38}$	$4\cdot10^{38}$

Використовуючи допустимі значення величини k, можна визначити також допустимі строки дії ключів, у тому числі в залежності від обсягу трафіка, швидкості шифрування, а також параметрів блокового симетричного шифру. Так, при умові, що допустима імовірність колізії $P_k \le 10^{-16}$, при $n = 2^{64}(l_б = 64)$ допустиме значення $k \le 64$, при $n = 2^{128}(l_б = 128)$ допустиме значення $k \le 8{,}2\cdot10^{11}$, при $n = 2^{256}(l_б = 256)$ допустиме значення $k \le 5\cdot10^{30}$. А знаючи розміри ключів, можна розрахувати максимально допустимий обсяг мережі.

Розглядаючи умови виникнення колізій та враховуючи незалежність потоку відкритих повідомлень та ключів, можна стверджувати, що імовірність співбігання криптограм за рахунок виникнення умови дорівнює добутку імовірності колізії ключів та імовірності колізії відкритих текстів.

Проведемо також дослідження можливостей появи колізії при застосуванні блокових симетричних шифрів в потоковому режимі, а також для потокових шифрів. Будемо вважати, що початковий стан шифруючого пристрою задається початковим ключем з довжиною l_n. При двійковій основі всього можна задати (ввести) $N_{кл} = 2^{l_n}$ ключів. Оскільки для початкових шифрів входом шифроутворюючого алгоритму є ключ, то число різних значень входів співпадає з повним простором ключів $N_{кл}$.

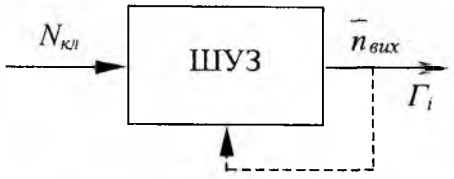

Рис. 1

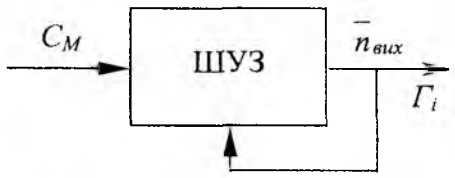

Рис. 2

На рис. 1 розглядається випадок, коли заміна початкового стану здійснюється з деяким інтервалом.

На рис. 2 розглядається випадок, коли здійснюється режим зі зворотним зв'язком по шифртексту чи лічильнику.

Виходом ШУЗ є гама зашифрування/розшифрування. В обох випадках настання колізії по суті буде призводити до перекриття шифру. Дійсно, якщо відрізки M_i^1 та M_i^2 зашифровано Γ_i, то

$$M_i^1 \oplus \Gamma_i \oplus M_i^2 \oplus \Gamma_i = M_i^1 \oplus M_i^2.$$

Далі розклад $M_i^1 \oplus M_i^2$ на складові може бути здійснений з поліноміальною складністю. Тому перекриття Γ_i у просторі або часі при одному і тому ж ключі є дуже загрозливим фактором.

Висновок

При розробці безумовно стійких шифрів необхідно враховувати можливість здійснення криптоаналізу на основі створення колізій. При цьому для забезпечення допустимої імовірності колізії мінімальна довжина блока інформації, а відповідно і одноразової гами, повинна бути обмежена. Для обчитслювально стійких шифрів імовірність перекриття шифру може бути визначена з використанням парадоксу «дні народження» і розрахунку імовірностей колізій, при цьому довжина шифруючої гами також обмежується з низу і залежить від допустимої імовірності колізії.

Список літератури: 1. *Шеннон К.* Теория связи в секретных системах // Работы по теории информации и кибернетике. М.: Изд-во. иностр. лит., 1963. С. 333 – 402. 2. *Замула А.А., Попович Е.В., Горбенко Ю.И.* Условия и возможности создания безусловно стойких криптосистем // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып. 119. С. 95 – 100. 3. *Шнайер Б.* Прикладная криптография. М.: Изд-во. «Триумф», 2002. 797 с. 4. *Стандарт* симетричного шифрування XXI століття: властивості, режими роботи, реалізація / І.Д. Горбенко, Л.В. Скрипник, С.О. Головашич, Т.О. Гріненко // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып. 119. С. 22 – 35. 5. *В. Столлингс.* Криптография и защита сетей: Принципы и практика. 2-е изд. Киев: Изд. дом «Вильямс», 2001. 669 с.

Харківський національний
університет радіоелектроніки
Департамент спеціальних телекомунікаційних
систем та захисту інформації СБ України

Надійшла до редколегії 15.05.2003