6. УДК 004.421.5

Самофалов К.Г., чл.-кор. НАН України,
Марковский А.П., канд.техн.наук,
Абабне О.А.

ОЦІНКА ЯКОСТІ ГЕНЕРАТОРІВ ДВІЙКОВИХ ПОСЛІДОВНОСТЕЙ З ВИКОРИСТАННЯМ НЕЛІНІЙНОЇ ВІДТВОРЮЮЧОЇ МОДЕЛІ

Національний технічний університет України "КПІ"

В статті пропонується підхід до підвищення ефективності тестування генераторів двійкових послідовностей для систем захисту інформації. Розроблено технологію визначення складності відтворення двійкової послідовності з k-кратною помилкою з використанням нелінійної відтворюючої моделі. Показано, що реалізація запропонованого способу має суттєво меншу обчислювальну складність в порівнянні з відомими методами, які використовують лінійну відтворюючу модель. Використання запропонованої технології оцінки складності відтворення двійкової послідовності з k-кратною помилкою дозволяє тестувати більш довгі послідовності, забезпечуючи, за рахунок цього більшу достовірність оцінки якості генераторів двійкових послідовностей.

Вступ

Динамічний розвиток інформаційної інтеграції на основі мережових технологій не можливий без адекватного вдосконалення засобів захисту інформації. Чільне місце в системі засобів захисту інформації посідають генератори випадкових та псевдовипадкових двійкових послідовностей (ГДП). Вони є основою одного з трьох класів алгоритмів захисту інформації, а саме потокових алгоритмів. Забезпечуючи найвищу швидкість шифрування даних, потокові алгоритми широко використовуються в мобільному зв'язку та швидкісних системах передачі даних бездротових мереж.

Розширення використання згаданих технологій передачі даних зумовлює необхідність вдосконалення ГДП. З огляду на тенденцію збільшення продуктивності комп'ютерних систем і можливості їх інтеграції для порушення захисту інформації, в сучасних умовах зростає важливість оцінки ГДП з точки зору забезпечення необхідного рівня захищеності даних. Така оцінка проводяться в формі тестування сформованої за допомогою ГДП послідовності, яка потенційно може бути використана для порушення захисту. В ряді стандартизованих тестів [1,6], з точки зору задач захисту інформації, найважливішим є тест на складність відтворюючої моделі заданої послідовності, який дозволяє формально оцінити теоретичну можливість її передбачення [1]. Вдосконалення такої оцінки, підвищення її достовірності, розвиток методики її реалізації являє собою, в огляду на вказані вище чинники, актуальну і практично важливу задачу.

Аналіз попередніх розробок

Формальною оцінкою рівня непередбачуваності n-бітової двійкової послідовності $S=\{s_1,s_2,\ldots,s_n\}$, $\forall j\in\{1,\ldots,n\}$: $x_j\in\{0,1\}$ виступає її складність [1], яка, в свою чергу, вимірюється через складність моделі, яка здатна відтворити послідовність S. За стандартом [6] в якості відтворюючої моделі виступає зсувний регістр з лінійною функцією зворотного зв'язку (лінійна модель). Користуючись алгоритмом [4], для будь-якої двійкової послідовності S можна побудувати лінійну модель, при цьому, лінійною складністю $L(S)$ послідовності S називається довжина зсувного регістру лінійної моделі. Доведено [4] , що за умови $L(S)=n/2$ задана послідовність S не може бути продовжена (екстрапольована). З теоретичної точки зору це означає, що задана послідовність S містить недостатню кількість інформації для успішної екстраполяції.

Практика оцінки якості ГДП, що застосовуються для захисту інформації показала недостатність тесту лінійної складності для адекватної оцінки неможливості передбачення двійкових послідовностей [1]. Лінійна складність дозволяє оцінити можливість безпомилкової екстраполяції послідовності. На практиці важливо оцінювати можливість екстраполяції послідовності з деякими похибками. проблема полягає в тому, що можлива ситуація, за якою n-бітова послідовність S має лінійну складність $L(S)=n/2$, тобто відповідний ГДП проходить тест на непередбачуваність. Проте, якщо змінити в послідовності S k бітів (позначимо отриману послідовність через S''), то може бути, що $L(S'')<<n/2$, тобто послідовність S'' можна екстраполювати. Це означає, що можна екстраполювати й послідовність S з ймовірністю помилки k/n. Таким чином, практика тестування ГДП показала необхідність побудови профілю складності при введенні k-кратної помилки. Виходячи з цього, в останні року в світі інтенсивно розвивається концепція побудови лінійної моделі, яка відтворює задану послідовність з k-кратною похибкою.

Фактично, задача полягає в віднаходженні для фіксованого значення k такої локалізації одиничних компонент вектору $C=\{c_1,c_2,\ldots c_n\}$, $\forall j\in\{1,\ldots,n\}$: $c_j\in\{0,1\}$, $c_1+c_2+\ldots+c_n=k$ для якої значення $L(S\oplus C)$ приймає мінімальне значення. Відповідно, мінімальне значення $L(S\oplus C)$ називають лінійною складністю послідовності S з k-кратною похибкою (k-error linear complexity of sequences).

Основною проблемою практичної реалізації врахування лінійної складності послідовності з k-кратною похибкою є значна обчислювальна складність. Дійсно, для заданої n-бітової послідовності S обчислювальна складність побудови лінійної відтворюючої моделі за алгоритмом Berlekamp-Massey [4] становить $O(n^2)$, при об'ємі пам'яті - $2\cdot n^2$. Кількість q_k можливих варіантів локалізації k модифікованих бітів послідовності становить $q_k=\binom{n}{k}$. За умови k<<n, можна вважати, що $q_k\approx n^k$. Відповідно, обчислювальна складність віднаходження мінімального значення $L(S\oplus C)$ становить $O(n^{k+2})$. Враховуючи, що довжина n послідовності, яка тестується становить в сучасних умовах 10^5-10^6 біт [1], то очевидною є проблема практичної реалізації оцінки якості ГДП с урахуванням можливості наближеної екстраполяції послідовностей.

В останні роки запропоновано ряд підходів [3,5], використання яких дозволяє частково знизити обчислювальну складність вказаної задачі тестування ГДП. Зниження обчислювальної складності, в основному, досягається за рахунок часткового використання при формуванні кожної нової відтворюючої лінійної моделі результатів побудови передуючої їй.

Ціллю досліджень є розробка підходу, який дозволяє зменшити затрати обчислювальних ресурсів для визначення лінійної складності послідовності S з k-кратною похибкою, тим самим, підвищити достовірність оцінки якості ГДП, орієнтованих на використання в системах захисту інформації.

Визначення складності двійкової послідовності з використанням нелінійної моделі

Як зазначалося вище, в умовах розширення використання бездротових засобів передачі даних динамічно зростає довжина послідовності, яка потенційно може бути використана для порушення системи захисту даних. Це звужує можливості тестування ГДП з використанням лінійної відтворюючої моделі, обчислювальна складність побудови я якої становить $O(n^2)$.

Для розширення можливостей тестування ГДП було запропоновано використовували нелінійну відтворюючу модель [2]. Така модель здатна відтворити задану послідовність S і являє собою зсувний m-бітовий регістр, поточні значення розрядів якого утворюють вектор

$X=\{x_1,x_2,\ldots,x_m\}$, з нелінійною, частковозначеною нелінійною функцією $f(X)$ зворотного зв'язку. Нелінійною складністю (максимальним порядком складності) $N(S)$ двійкової n-бітової послідовності S є мінімальна довжина - m зсувного регістру з нелінійною функцією зворотного зв'язку, який відтворює послідовність S. В роботі [2] теоретично доведено, що послідовність не може бути екстрапольована, якщо $N(S)=2\cdot\log_2 n$ і запропоновано алгоритм формування нелінійної відтворюючої моделі. Основна перевага в використанні нелінійної відтворюючої моделі в порівнянні з лінійною полягає в меншій обчислювальній складності її побудови. Так, обчислювальна складність запропонованого в роботі [2] алгоритму формування нелінійної моделі складає $O(n\cdot\log_2 n)$, що суттєво менше складності $O(n^2)$ побудови лінійної відтворюючої моделі. Сутність алгоритму полягає в побудові бінарного дерева G, число ярусів якого відповідає значенню нелінійної складності.

Фактично, в результаті роботи алгоритму [2] будується частково визначена нелінійна функція $f(X)$ зворотного зв'язку. Недоліком нелінійної моделі, на відміну від лінійної, є неможливість її використання для екстраполяції послідовності.

В роботі пропонується розвиток вказаної концепції використання нелінійної відтворюючої моделі для тестування двійкових послідовностей. Зокрема, пропонується розширення можливостей нелінійної моделі за рахунок її використання для визначення складності послідовності S з k-кратною похибкою.

Визначення складності послідовності з k-кратною помилкою за допомогою нелінійноївідтворюючої моделі

Аналіз можливостей застосування нелінійної відтворюючої моделі для визначення складності n-бітовой двійкової послідовності S з k-кратною похибкою дозволяє виділити два способи вирішення цієї задачі.

Перший спосіб дозволяє оцінити можливість спрощення відтворюючої моделі при заміні k бітів послідовності на якісному рівні, другий - дозволяє виконати кількісну оцінку.

Сутність першого способу полягає в оцінці отриманого значення нелінійної складності M(S): перевищення значення M(S) рівня $2\cdot\log_2 n$ непрямо пов'язано з незбалансованістю бінарного дерева G. В свою чергу, незбалансованість дерева G свідчить при можливість зменшення числа його ярусів за рахунок балансування шляхом зміни k кінцевих вершин, що є тотожним інвертуванню відповідних k бітів заданої послідовності S. Доведено [2], що нелінійна складність розподілена на нормальним законом с математичним очікуванням $2\cdot\log_2 n$ та середньоквадратичним відхиленням $0.5\cdot\log_2 n$. Відповідно, ймовірність P_1 того, що для випадкової послідовності значення нелінійної складності M(S) перевищить $2\cdot\log_2 n$ визначається як:

$$P_1=\frac{\sqrt{2}}{\sqrt{\pi}\cdot\log_2 n}\cdot\int_{2\cdot\log_2 n}^{M(S)} e^{\frac{(x-2\cdot\log_2)^2}{0.5\cdot\log_2^2 n}}dx$$

Обчислене за допомогою функції Лапласа значення P_1 порівнюється з пороговим значенням $P_П$ і при умові $P_1>P_П$ приймається рішення щодо невідповідності послідовності, що проходить тестування, критерію неможливості екстраполяції з невеликою помилкою.

Запропонований спосіб простий в реалізації і може бути, в принципі, модифікований для лінійної відтворюючої моделі. В роботі [3] зазначається, що "послідовність з великим (відносно рівня $n/2$-*авт.*) значенням лінійної складності може бути апроксимована послідовністю з низьким рівнем складності".

Сутність другого способу полягає в обчисленні ваги Хемінга диференціалу частково визначеної булевої нелінійної функції $f(X)$ зворотного зв'язку по змінній x_1. Визначений диференціал порівнюються зі значенням k - кількості помилок апроксимації.

Хемінгова вага $HW(\partial f(X)/\partial x_1)$ диференціалу частково визначеної функції нелінійної функції $f(X)$ по змінній x_1 визначається згідно з наступного виразу:

$$HW(\partial f(X)/\partial x_1) = \sum_{x_2 \ldots x_m \in Z_1} \delta(f(x_1, x_2 \ldots x_m), f(\overline{x_1}, x_2, \ldots, x_m)),$$

де функція $\delta(y_1,y_2) = y_1 \oplus y_2$, якщо обидві булеві змінні y_1, y_2 визначені, тобто $y_1, y_2 \in \{0,1\}$ і $\delta(y_1,y_2)=0$, якщо значення хоча б однієї з змінних y_1, y_2 не визначено, Z_1 - множина з 2^{m-1} всіх можливих наборів значень булевих змінних $x_2, x_3, \ldots, x_m$. Враховуючи, що нелінійна модель відтворює послідовність S, то фактично $q = HW(\partial f(X)/\partial x_1)$ показує кількість бітів послідовності змінюються при зміні функції $f(x_1, x_2, \ldots, x_m)$ зворотного зв'язку на функцію $f(x_1 \oplus 1, x_2, \ldots, x_m)$. Іншими словами, значення $q = HW(\partial f(X)/\partial x_1)$ дорівнює кількості бітів послідовності S, які будуть невірно відтворені нелінійною моделлю без врахування старшого розряду x_1 зсувного регістру, тобто за допомогою (m-1)-розрядного зсувного регістру. Зі сказано випливає, що якщо $HW(\partial f(X)/\partial x_1) \le k$, або $HW(\partial f(X)/\partial x_1) > n-k$, то задана послідовність S може бути апроксимована послідовністю S' з помилкою $d = \min\{q, n-q\}$ не більше ніж в k бітів, причому нелінійна складність S' на одиницю менша за складність заданої послідовності S: $M(S')=M(S)-1$.

Викладене ілюструється наступним прикладом. Нехай задана послідовність: $S=\{1,0,1,1,0,0,1,0\}$, така ж сама, як наведена в [2]. Її нелінійна складність дорівнює трьом: $M(S)=3$. Значення частково-визначеної нелінійної функції $f(x_1,x_2,x_3)$ зворотного зв'язку наведені в таблиці 1.

Таблиця 1

Таблиця істинності $f(x_1,x_2,x_3)$

x_1 x_2 x_3	$f(x_1,x_2,x_3)$
0 0 0	-
0 0 1	0
0 1 0	-
0 1 1	0
1 0 0	1
1 0 1	1
1 1 0	0
1 1 1	-

Значення $HW(\partial f(x_1,x_2,x_3)/\partial x_1) = 1$. Це означає, що задана послідовність S може бути апроксимована з одним помилковим бітом послідовністю $S'=\{1,0,1,1,0,1,1,0\}$, яка має нелінійну складність $M(S')=2$.

Це означає, що послідовність S' може бути відтворена зсувним регістром з розрядністю 2 і функцією $\varphi(x_2,x_3)$ зворотного зв'язку, значення якої представлені в таблиці 2.

Таблиця 2

Таблиця істинності $\varphi(x_2,x_3)$

x_2 x_3	$\varphi(x_2,x_3)$
0 0	-
0 1	1
1 0	1
1 1	0

Важливим є оцінка затрат обчислювальних ресурсів, потрібних на реалізацію запропонованого способу тестування можливого спрощення відтворюючої моделі послідовності. Обчислювальна складність формування Хемінгової ваги диференціалу частково визначеної нелінійної функції $f(X)$ по змінній x_1 визначається перебором половини таблиці істинності $f(X)$ і дорівнює $O(2^{m-1})=O(2 \cdot n)$. Враховуючи обчислювальну складність побудови нелінійної моделі - $O(n \cdot \log_2 n)$ [2], в результаті чого формується таблиця функції $f(X)$, отримуємо, що обчислювальна складність визначення можливості зменшення на одиницю розрядності нелінійної відтворюючої моделі за рахунок d помилок апроксимації

заданої двійкової послідовності становить $O(n\cdot(\log_2 n + 2))$. Це означає, що витрати часу на вирішення вказаної задачі тобто незначно перевищує витрати на побудову самої нелінійної відтворюючої моделі.

Вказане значення обчислювальної складності є суттєво меншим від вирішення аналогічної задачі з використанням лінійної відтворюючої моделі [5], обчислювальна складність якої експоненційно залежить від значення k - $O(n^k)$. Слід зазначити, що значне зменшення ресурсів часу, потрібного для визначення можливості спрощення відтворюючої моделі за рахунок допустимої похибки апроксимації заданої послідовності досягається за рахунок трьох факторів:

1. Використання нелінійної відтворюючої моделі, розмірність якої $(2\cdot\log_2 n)$ значно менша за розмірність лінійної моделі $(n/2)$;

2. Суттєвого збільшення, а порівнянні з відомими методами [3] необхідного об'єму пам'яті, потрібної для формування таблиці істинності нелінійної булевої функції $f(X)$ зворотного зв'язку. Згаданий об'єм становить $2\cdot 2^m = 8\cdot n$ бітів. Наприклад, для типової довжини послідовності 10^6 бітів, потрібний об'єм пам'яті становить близько одного мегабайту, що цілком реалізуємо сучасними технічними засобами;

3. Звуження класу задачі тестування: фактично в попередніх розробках [3,5] визначалась мінімальна складність відтворюючої лінійної моделі при всіх можливих k-бітових помилках апроксимації. Запропоноване рішення дозволяє виявити можливість зменшення складності відтворюючої моделі при всіх можливих k-бітових помилках апроксимації. На практиці тестування ГДП визначення такої можливості цілком може бути прийнятним результатом [1].

Для визначення мінімальної складності нелінійної відтворюючої моделі при всіх можливих k-бітових помилках апроксимації заданої послідовності, запропонована процедура може бути застосована рекурсивно.

Нехай задана послідовність S довжиною n біт і допустима кількість k бітових помилок апроксимації S. Потрібно визначити мінімальну довжину зсувного регістру з нелінійною функцією зворотного зв'язку, який здатен відтворити послідовність S з похибкою не більше k біт.

Рекурсивне використання запропонованої процедури виконується в наступному порядку:

1. Для послідовності S будується нелінійна відтворююча модель і, відповідно, функція $f(X)$ зворотного зв'язку, кількість аргументів вказаної функції визначає нелінійну складність -$M(S)$.

2. Для функції $f(X)$ визначається значення її диференціалу по змінній x_1, яка відповідає старшому розрядові зсувного регістру: $q = HW(\partial f(x_1, x_2, x_3)/\partial x_1)$. Обчислюється значення $d = \min\{q, n-q\}$

3. Якщо $d \le k$, то в послідовності S інвертуються біти, які зумовлюють залежність функції $f(X)$ від змінної x_1. Значення k зменшується на величину d: $k:=k-d$. Якщо k>0, то повернення на п.1, інакше $M(S):=M(S)-1$.

4. Якщо $M(S) < \log_2 n$, то тестування послідовності свідчить про невідповідність ГДП використанню в системах захисту інформації.

Обчислювальна складність реалізації наведеної рекурсивної процедури визначається кількістю h кроків рекурсії і становить $O(h\cdot(n+2)\cdot\log_2 n)$. Це значно менше в порівнянні з обчислювальною складністю $O(n^{k+2})$. вирішення цієї задачі з використанням відомих способів.

Таким чином, запропонований спосіб тестування ГДП через визначення складності апроксимації послідовностей, що генеруються ГДП дозволяє, за рахунок меншої обчислювальної складності, суттєво збільшити довжину послідовностей. Експериментальними дослідженнями доведено, що тестування можливості апроксимації послідовностей довжиною 10^8 відомими методами [3] пов'язане зі значними технічними труднощами, в той час,

як запропонований спосіб дозволяє реалізувати тестування таких послідовностей достатньо просто. Можливість тестування довших послідовностей дозволяє значно підвищити достовірність оцінки якості ГДП.

Висновки

В результаті проведених досліджень запропоновано новий спосіб тестування ГДП, орієнтованих на використання для захисту інформації. Запропонований спосіб полягає в дослідженні складності отриманої з допомогою ГДП послідовності, на основі побудови нелінійної відтворюючої моделі у вигляді зсувного регістру з нелінійною функцією зворотного зв'язку. Розроблена методика дозволяє оцінити можливість апроксимувати послідовність при допустимості k-кратної помилки більш простою відтворюючою моделлю.

Основною перевагою запропонованого способу є значно менша обчислювальна складність процедури оцінки можливості апроксимації n-бітової послідовності з кратністю помилок k, що не перевищує задане значення. Доведено, що обчислювальна складність тестування за запропонованим способом становить $O(n\cdot(\log_2 n + 2))$, тоді як для відомих способів обчислювальна складність експоненційно залежить від k . Суттєве зменшення обчислювальної складності, яке досягається використанням запропонованого способу дозволяє реалізувати дослідження довших тестових послідовностей і, тим самим, підвищити достовірність та оперативність контролю придатності ГДП для використання в системах захисту інформації.

Отримані в роботі результати можуть бути використані для підвищення ефективності тестування як алгоритмів одержання псевдовипадкових двійкових послідовностей, так і засобів формування фізично випадкових двійкових послідовностей.

Список літератури

1. Иванов М.А., Чугунков И.В. Теория, применение и оценка качества генераторов псевдослучайных последовательностей. М.: "Кудиц-Образ",-2003.- 238 с.

2. Марковский А.П., Мустафа Акрам Ареф Найеф, Бойко А.В. Об одном подходе к определению сложности случайных и псевдослучайных двоичных последовательностей // Вісник національного технічного університету України "КПІ". Інформатика, управління та обчислювальна техніка. – 2002.- №37.- С.120-129.

3. Kurosava K., Sato F., Sakata T., Kishimoto W. A relationship between linear complexity and k-error linear complexity. // IEEE Trans. Information Theory, 2000.- V.46,-№3.-P.694-698.

4. Massey J.L. Shift register sequences and BCH decoding // IEEE Transaction of Information Theory.- 1969.- Vol. 15.- № 1. - P.122-127.

5. Meidl W., Niederreiter H. On the expected value of the linear complexity and the k-error linear complexity of periodic sequences // IEEE Transaction of Information Theory.- 2002.- Vol. 48.- № 11. - P.2817-2825.

6. NIST Special Publicaion 800-22: A Statistical Test Suite for Random and Pseudorandom Number. 2000. -348 p.