УДК 519.688
DOI https://doi.org/10.32838/2663-5941/2020.2-1/23

Політанський Р.Л.
Чернівецький національний університет імені Юрія Федьковича

ДОСЛІДЖЕННЯ ПЕРІОДИЧНОСТІ ПСЕВДОВИПАДКОВИХ ПОСЛІДОВНОСТЕЙ МЕТОДОМ БУЛЕВОГО ГІПЕРКУБУ

У роботі розглянута методика визначення періоду генерованих дискретними відображеннями псевдовипадкових послідовностей, які мають властивість марківських ланцюгів. Додатній показник Ляпунова зазначених послідовностей вказує на те, що вони володіють властивостями динамічного хаосу і не мають періоду повторення. Однак слід зауважити, що виконання обчислень із дійними числами в реальних комп'ютерних системах завжди виконується зі скінченною точністю, що визначається кількістю цифр у досліджуваних числах.

Максимальна кількість неповторюваних елементів послідовності, представлених у десятковому форматі, становить 10^n. Задача визначення періодів генерованих послідовностей вимагає значних витрат машинного часу. В роботі запропонований алгоритм, який базується на відображенні десяткового дробу на багатовимірний масив двійкових чисел (булевий гіперкуб), який значно скорочує мінімально необхідну кількість обчислень.

Дослідження проводилися з використанням програмного середовища Dev C++. Точність обчислень становила до 9 десяткових знаків. Ефективність запропонованого алгоритму оцінювалася на прикладі логістичного відображення, для виконання програми використовувався комп'ютер класу ПЕОМ.

Отримані якісні дані щодо характеру залежності значення періоду від точності проведених обчислень. Метод має переваги з теоретичної точки зору і скерований на його практичне застосування для генерування псевдовипадкових послідовностей. Запропонований метод є більш точним порівняно з методом визначення періоду шляхом виявлення співпадань із заздалегідь вибраним значенням, оскільки при цьому повторення невідомих значень не можливо виявити. З практичної точки зору метод забезпечує визначення періоду в псевдовипадкових послідовностях, генерованих на основі будь-якого алгоритму з використанням програмних чи апаратних засобів.

Ключові слова: *булевий гіперкуб, псевдовипадкові послідовності, стандарти IEEE, ланцюг Маркова, легка криптографія.*

Постановка проблеми. Розроблення методів генерування псевдовипадкових послідовностей, які використовуються в системах криптографічного захисту, кодового розділення каналів, розширення спектру, є однією з основних задач, що підлягають вирішенню при побудові високоефективних систем передавання інформації [1, с. 175]. Дослідження статистичних властивостей псевдовипадкових послідовностей можуть бути проведені випробуваннями на відповідність вимогам NIST STS [2, с. 1].

Іншими методами перевірки генерованих послідовностей на їх відповідність критеріям псевдовипадковості є підходи Блюма, Голдвассера, Міккалі та Яо [3, с. 51], які базуються на поліноміальному алгоритмі їх генерування. Більшість зазначених послідовностей володіють властивістю марківських ланцюгів, кожен наступний елемент яких залежить лише від попереднього. Якщо в такому ланцюзі матиме місце повторення деякого значення, то генерована послідовність буде періодичною.

Визначення періоду є складною задачею з точки зору обчислювальної складності. В роботі розроблений алгоритм виявлення повторюваних значень у послідовності, в основу якого покладено відображення десяткового дробу на багатовимірний масив розмірності 10^n булевих змінних, де n – число десяткових знаків, які визначають дробову частину.

Аналіз останніх досліджень і публікацій. Розвиток технологій, пов'язаних із використанням під'єднаних до мереж загального користування комп'ютеризованих пристроїв (в тому числі пристроїв, підключених до інтернету), вимагає розроблення нових методів шифрування, які одержали назву легкої криптографії. Важливість алгоритмів легкої криптографії зазначається в доповідях [4, с. 1], опублікованих на сайті NIST у 2017 році. Метою цих доповідей є стандартизація алгоритмів шифрування, які не потребують значних машинних ресурсів і можуть бути застосовані в автономних комп'ютерних системах. Згідно з цим документом в якості апаратної

реалізації алгоритмів легкої криптографії використовуються перепрограмовані мікроконтролери та пристрої RFID.

Актуальність досліджень алгоритмів потокового шифрування та розроблення пристроїв, які базуються на алгоритмах легкої криптографії і потребують обмежених обчислювальних ресурсів, підтверджується також низкою сучасних досліджень, результати яких представляються на конференціях IEEE по проблемах розвитку Інтернету речей (IoT) [5, с. 1]. До алгоритмів легкої криптографії належать також алгоритм блокового шифрування DESL, що на відміну від восьми перетворень використовує одне [6, с. 165] перетворення S-box, застосоване у класичному алгоритмі DES.

Крім того, заслуговують на увагу алгоритми потокового шифрування, більшість із яких базуються на псевдовипадкових послідовностях. Зокрема це послідовності Голда [1, с. 178], Уолша [7, с. 198], m-послідовності [1, с. 173], послідовності Баркера [7, с. 40], коди Френка [7, с. 110] та Чу [7, с. 143], які генеруються з використанням регістрів із оберненими зв'язками.

Постановка завдання. Основним недоліком зазначених методів є їх лінійність. Тому актуальним є розроблення нелінійних методів генерування, зокрема методів, заснованих на дискретних відображеннях систем із властивостями динамічного хаосу. Такі системи мають додатні показники Ляпунова і генерують числові послідовності, що теоретично є неперіодичними.

Метою роботи є розроблення та апробація нового алгоритму визначення періодів псевдовипадкових послідовностей. Вирішена задача генерування псевдовипадкових послідовностей зареєстрованої довжини, при яких відсутні повторення комбінацій заданого числа перших цифр десяткового дробу. Дослідження проводилися для різних значень кількості цифр, що позначають дріб (від трьох до дев'яти) на множині початкових значень $x_0 \in [0;1]$ із різними кроками дискретизації залежно від точності подання десяткового дробу від 0.01 до 0.000001.

Виклад основного матеріалу дослідження. Математичною моделлю довільного одновимірного відображення є такий вираз:

$$x_{n+1}=f(x_n), \qquad (1)$$

де f – деяка нелінійна функція, n – номер ітерації, x_n та x_{n+1} – поточний і наступний елементи псевдовипадкової послідовності.

Найвідомішим є логістичне відображення, яке описується функцією другого порядку (2):

$$x_{n+1}=\lambda x_n (1-x_n), \qquad (2)$$

де λ – параметр логістичного відображення, за яким генеруються хаотичні коливання.

Зумовлену кінцевою точністю обчислень періодичність хаотичних коливань, генерованих за логістичним відображенням із параметром λ = 4, автор пояснює на послідовності, утвореній в результаті обчислень із кінцевою точністю двох десяткових знаків:

0.1→0.36→**0.92**→0.29→0.94→
0.23→0.80→0.64→**0.92**.
0.2→**0.64**→0.92→0.29→0.94→0.23→0.80→**0.64**.
0.3→0.84→0.54→0.99→0.04→0.15→
0.51→1.00→**0.00**→**0.00**.
0.4→0.96→0.15→0.51→1.00→**0.00**→**0.00**.

У наведеному прикладі цифри, які повторюються, виділені більш жирним шрифтом. Очевидно, що після повторення однієї цифри відбуватиметься повторення всієї послідовності.

Визначення періоду повторення заданої кількості десяткових значень у генерованій послідовності здійснювалося методом проектування деякої кількості перших десяткових знаків, які відображають члени послідовності, на багатомірний масив булевих змінних (булевий гіперкуб), розмірність якого становила 10×10×…×10 (кількість множників залежить від кількості десяткових знаків).

Для прикладу автор розглядає відображення на масив булевих змінних bool_array, генерованих за логістичним відображенням псевдовипадкової послідовності із початковим значенням $x_0 = 0.10$, λ = 4 при умові, що точність обчислень становить два десяткові знаки (перший рядок із наведеного прикладу). При цьому розмірність масиву, що містить контрольовані значення, становить 10×10. Автор знову наводить генеровану послідовність, вказавши також процедуру заповнення масиву булевих змінних:

x_0 = 0.10- > bool_array [1] [0] = true
x_1 = 0.36- > bool_array [3] [6] = true.
x_2 = 0.92- > bool_array [9] [2] = true
x_3 = 0.29-> bool_array [2] [9] = true.
x_4 = 0.94- > bool_array [9] [4] = true
x_5 = 0.23- > bool_array [2] [3] = true.
x_6 = 0.80- > bool_array [8] [0] = true
x_7 = 0.64- > bool_array [6] [4] = true.
x_8 = 0.92- > bool_array [9] [2] = true.

Як видно із наведеного прикладу довжина послідовності, утвореної неповторюваними значеннями, становить 9, тобто: $\{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$, тоді як період становить 6 {x2, x3, x4, x5, x6, x7}.

Блок-схема розробленого алгоритму, що визначає довжину послідовності, утвореної неповторюваними значеннями, наведена на рисунку 1.

Автор зауважує, що для визначення періоду потрібно виконати алгоритм ще раз із початковим значенням, яке повторюється і було визначене на попередньому кроці. В цьому випадку генерована послідовність до завершення алгоритму матиме такі значення:

x_0 = 0.92- > bool_array [9] [2] = true.
x_1 = 0.29-> bool_array [2] [9] = true.
x_2 = 0.94- > bool_array [9] [4] = true.
x_3 = 0.23- > bool_array [2] [3] = true.
x_4 = 0.80- > bool_array [8] [0] = true.
x_5 = 0.64- > bool_array [6] [4] = true.
x_6 = 0.92- > bool_array [9] [2] = true.

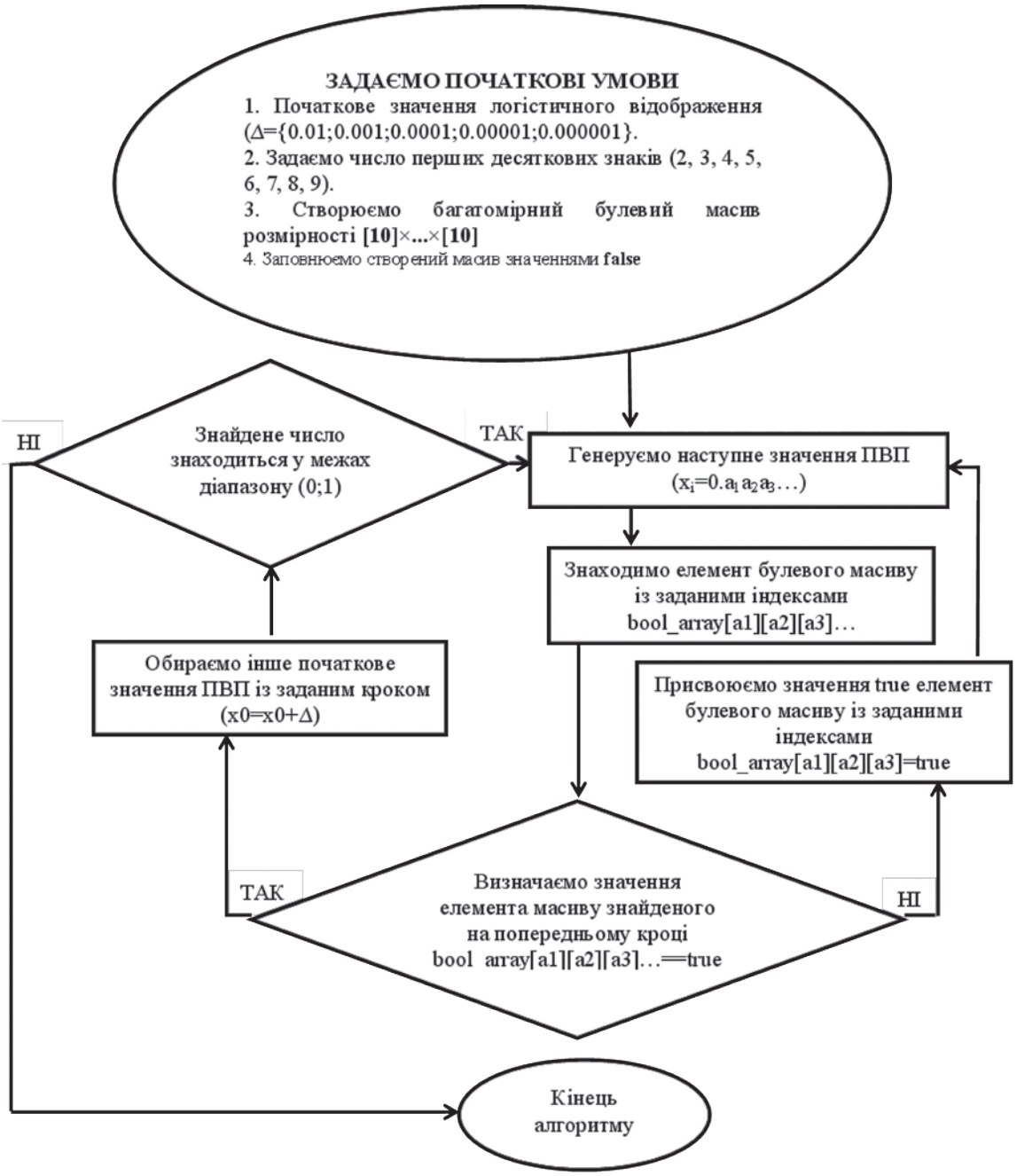

Рис. 1. Блок-схема алгоритму визначення довжини послідовності, утвореної неповторюваними значеннями

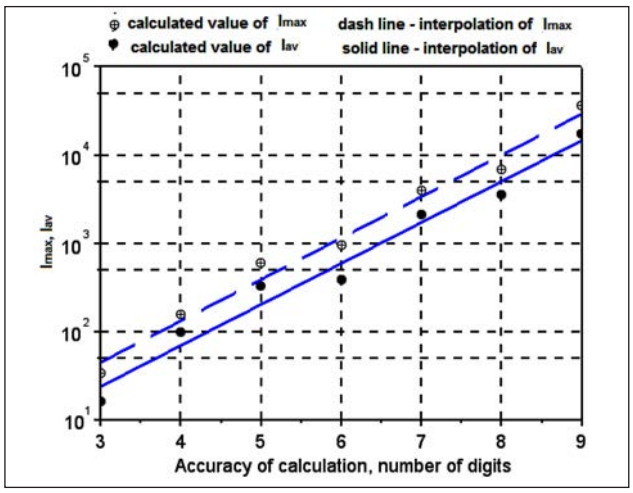

Рис. 2. Обчислені середні та найбільші значення довжин ПВП, генерованих за логістичним відображенням при λ = 4 і відповідні їм інтерполяційні прямі, розраховані у припущенні експоненційної залежності від точності обчислень

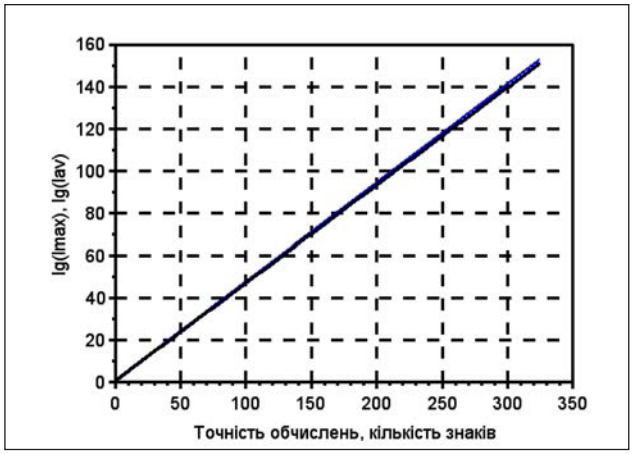

Рис. 3. Екстраполяція середніх і найбільших значень довжин ПВП, генерованих за логістичним відображенням при λ = 4

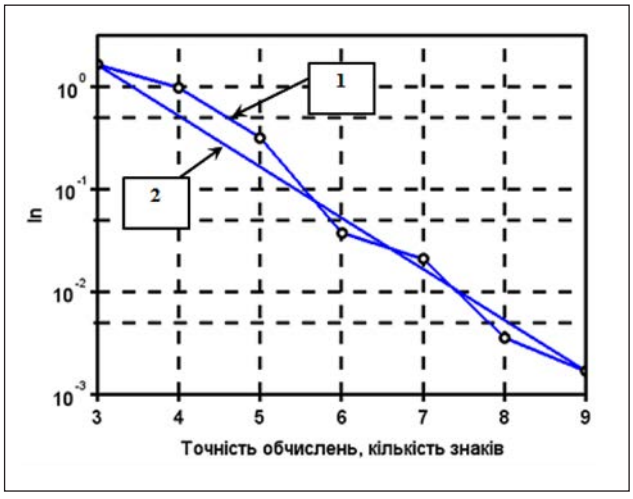

Рис. 4. Залежність нормованого значення довжини ПВП, генерованих за логістичним відображенням при λ = 4 від точності обчислень (1) та її лінійна інтерполяція (2)

Результати досліджень періодичності ПВП, генерованих за логістичним відображенням з використанням розробленого алгоритму, наведені в Табл. 1.

У таблиці наведені максимальні та відповідні їм початкові значення ПВП, середні та нормовані середні значення довжин ПВП, для яких спостерігається максимальне значення довжини. Розрахунки проводилися для точності обчислень 3-, 4-, 5-, 6-, 7-, 8-, 9-десяткових значень. Для кожного значення точності обчислень найменшим ступенем дискретизації, для якого було доцільно проводити розрахунки, є 0.01.

Найбільший ступінь дискретизації обмежувався двома чинниками: точністю обчислень і максимальним часом розрахунків, який не перевищував 2 години. Так, для точності обчислень 9-десяткових знаків і кроку дискретизації 4-десяткових знаків час обчислень становив приблизно 15 годин. Наведені в таблиці часові витрати на обчислення довжини ПВП отримані з використанням процесора Intel з частотою 2.00 GHz і програмного середовища Dev C++.

Нормоване значення довжини ПВП розраховано як відношення її середнього значення до потужності множини, утвореної десятковими числами розрядності n (3):

$$l_{norm} = l_{av} / 10^{n}, \qquad (3)$$

де l_{av} та l_{norm} – середнє та нормоване значення довжини ПВП, n – точність обчислень.

Розраховані значення середніх і максимальних довжин ПВП наведені на Рис. 2 у напівлогарифмічному масштабі. Їх можна інтерполювати експоненційними залежностями, описаними такими виразами:

$$l_{av} = 0.957 \cdot e^{1.07 \cdot n}, \qquad (4)$$

$$l_{max} = 1.746 \cdot e^{1.08 \cdot n}, \qquad (5)$$

На Рис. 2 зображені також інтерполяційні прямі, побудовані на основі виразів (4) і (5) у напівлогарифмічному масштабі.

Екстрапольовані значення десяткових логарифмів середніх і максимальних значень довжин ПВП, які побудовані за формулами (4) і (5), практично збігаються (Рис. 3) і можуть бути використані для оцінювальних значень довжин послідовностей для високих значень точностей обчислень,

Таблиця 1

Результати статистичних досліджень періодичності ПВП при різних значеннях кроку дискретизації і початкових значень хаотичних коливань в діапазоні $x_0 \in (0;1)$

Точність обчислень, кількість десяткових знаків	Крок дискретизації	Максимальні значення довжин ПВП / Відповідні їм початкові значення	Середні значення довжин ПВП	Нормовані середні значення	Час обчислень, с
3	0.01	30 / 0.28	15.38	1.54	<1
	0.001	34 / 0.322	16.37	1.64	<1
4	0.01	150 / 0.22	89.94	0.90	<1
	0.001	154 / 0.072; 0.347	94	0.94	<1
	0.0001	156 / 0.1976	97.7	0.98	5
5	0.01	499 / 0.34	273	0.273	<1
	0.001	581 / 0.069	319.8	0.32	1
	0.0001	595 / 0.0779	324.6	0.32	9.72
	0.00001	603 / 0.27618	323	0.32	115.7
6	0.01	801 / 0.39	384.68	0.038	<1
	0.001	911 / 0.416	375.46	0.038	4.36
	0.0001	931 / 0.0778	382.76	0.038	42.15
	0.00001	947 / 0.39731	383.71	0.038	348.6
	0.000001	948 / 0.357161	382.93	0.038	4390
7	0.01	3371 / 0.24	2150.9	0.022	4
	0.001	3799 / 0.394	2133.1	0.021	42.57
	0.0001	3892 / 0.2536	2109.33	0.021	394
	0.00001	3942 / 0.41866	2116.27	0.021	3865
8	0.01	6194 / 0.34	3258	0.0033	45.24
	0.001	6654 / 0.209	3475.21	0.0035	425
	0.0001	6772 / 0.2478	3548.84	0.0036	4286
9	0.01	34272 / 0.1;0.2	17200.36	0.0017	494.8
	0.001	36525 / 0.03	17040.99	0.0017	5396

що потребують використання спеціалізованих потужних обчислювальних кластерів.

Наприклад, при точності обчислень 250 знаків середня довжина послідовності становитиме 10^{120} чисел, що приблизно дорівнює 2^{360}.

На Рис. 4 у напівлогарифмічному масштабі наведена залежність нормованих на потужність множини десяткових чисел заданої розрядності значень довжини ПВП, генерованих за логістичним відображенням при $\lambda = 4$.

Так, залежність нормованої довжини від точності обчислень є обернено експоненційною, а відповідна апроксимуюча залежність описується так:

$$l_{norm} = 97.047 \cdot e^{-1.23 \cdot d}, \quad (6)$$

З таблиці 1 випливає, що подальше зменшення кроку дискретизації на порядок призводить до зростання часових затрат приблизно до 10 год. Збільшення точності обчислень вимагає використання спеціалізованих обчислювальних комплексів зі значним обсягом оперативної пам'яті. При точності обчислень 10 десяткових знаків необхідний обсяг пам'яті становить 10^{10} байт, що приблизно дорівнює 9 Гбайт, що не забезпечується технічними можливостями користувацького програмного та апаратного забезпечення та вимагає використання програмного-апаратних обчислювальних комплексів із спеціалізованою архітектурою.

Підвищення швидкодії обчислень на множині початкових значень із найменшим можливим кроком дискретизації, що дорівнює 10^{-n}, можливе шляхом створення паралельних алгоритмів, що є відносно нескладною задачею, в той час як виконання паралельних обчислень при збільшенні точності обчислень є більш складним.

Висновки. Запропонований алгоритм обчислення довжини ПВП шляхом відображення десяткового дробу на багатовимірний масив булевих змінних уможливлює встановлення значень періоду генерованих ПВП та його залежності від початкових значень хаотичних коливань і параметру логістичного відображення з використанням ПЕОМ із відносно невисокими обчислювальними характеристиками.

Одержані результати можуть слугувати основою аналізу поведінки ПВП шляхом екстраполяції даних для більш високих значень точності обчислень, що уможливлює вибір оптимальної конфігурації автономних обчислювальних пристроїв для генерування псевдовипадкових послідовностей із заданими значеннями періоду.

Список літератури:

1. Климаш М.М., Пелішок В.О. Проектування ефективних систем безпровідного зв'язку. Львів, 2010. 232 с.

2. NIST. URL: https://www.nist.gov/publications/statistical-test-suite-random-and-pseudorandom-number-generators-cryptographic.

3. Гулак Г.М., Мухачов В.А., Хорошко В.О., Яремчук Ю.Є. Основи криптографічного захисту інформації : підручник. Вінниця : ВНТУ, 2011. 199 с.

4. NIST. URL: https://www.nist.gov/publications/report-lightweight-cryptography.

5. Liu B., Wu R., Xie M., Li Q.P. Loong: A Family of Involutional Lightweight Block Cipher Based on SPN Structure. IEEE Access. Vol. 7. 2019. P. 136023–136035.

6. Stallings W. Cryptography and Network Security Principles and Practices, Fourth Edition. Prentice Hall. 2005. 592 p.

7. Сумик М.М., Прудиус І.Н., Сумик Р.М. Теорія сигналів : підручник. Львів : Бескид біт, 2008. 232 с.

Politanskyi R.L. INVESTIGATION OF THE PERIODICITY OF PSEUDORANDOM SEQUENCES BY THE BOOLEAN HYPERCUBE METHOD

The method of determining the period of discrete mapping of pseudorandom sequences having the property of Markov chains is considered in the paper. The Lyapunov exponent of these sequences indicates that they have dynamic chaos properties and, as a consequence, do not have a repetition period. However, it should be noted that performing calculations with real numbers on real computer systems is always done with finite accuracy, which is determined by the number of digits in the numbers being investigated. The maximum number of non-repeating elements presented in decimal format in any sequence is 10^n.

The task of defining periods of generated sequences requires considerable machine time. An algorithm based on the mapping of a decimal fraction to a multidimensional array of binary numbers (Boolean hypercube) is proposed, and it significantly reduces the minimum required number of calculations. The studies were conducted using the Dev C++ software environment. The accuracy of the calculations was up to 9 decimal places.

The effectiveness of the proposed algorithm was evaluated by the example of logistic mapping; to use the program used a computer class PC. The qualitative data about the character of the dependence of the value of the period on the accuracy of the calculations are obtained. The method has advantages from the theoretical point of view and is useful for practical application for the generation of pseudorandom sequences.

The proposed method is more accurate than the period determination method based on detecting matches with a pre-selected value, since the repetition of unknown values cannot be detected. From a practical point of view, the method provides for determining the period in pseudorandom sequences generated by any algorithm using software or hardware.

***Key words**: Boolean hypercube, pseudorandom sequences, IEEE standard, Markov sequences, light cryptography.*