УДК 004.056.53; 519.876.5

Ю. В. Задоя, бізнес-аналітик IT-департаменту HRM-системи КБ “ПриватБанк”
О. Н. Молотков, кандидат технічних наук, доцент кафедри інформаційних систем та технологій Академії митної служби України
Л. В. Кабак, кандидат технічних наук, доцент кафедри інформаційних систем та технологій Академії митної служби України

ОПТИМІЗАЦІЯ ПАРАМЕТРІВ ГЕНЕРАЦІЇ ПСЕВДОВИПАДКОВИХ ЧИСЕЛ З МЕТОЮ ГЕНЕРАЦІЇ ПАРОЛІВ

Стаття присвячена оптимізації параметрів сучасних генераторів псевдовипадкових чисел з метою генерації паролів. Проаналізовано сучасні генератори псевдовипадкових послідовностей, що використовуються для завдань автентифікації користувачів, та виявлено наявність впливу значень параметрів генерації на її якість, запропоновано діапазони вхідних параметрів генерації, які дозволяють підвищити відсоток проходження тестів на випадковість.
Ключові слова: *генерація; псевдовипадкові числа; оптимізація; параметри.*

The article is devoted to the optimization of the parameters of modern random number generators to generate passwords. Modern generators of pseudo-random sequences, which used for user authentication tasks, have been analyzed and it has been revealed the presence of the influence of parameters of generation on the quality of the generation, ranges input parameters of generation have been for increasing the percentage of passing tests of randomness.
Key words: *generation; pseudo-random numbers; optimization; parameters.*

Постановка проблеми. Зі зростанням інформатизації всіх складових сучасного життя і швидкодії обчислювальної техніки питання захисту інформації з кожним роком набуває дедалі важливішого значення.

З погляду програмного захисту інформації, на сучасному етапі провідне місце посідає авторизація користувачів. Водночас виникає нова проблема: як захистити пароль від зламування, адже існує безліч програм, які методами простого перебирання за певний час можуть підібрати пароль будь-якого рівня складності.

Нині одним із провідних і перспективних методів є використання генераторів псевдовипадкових послідовностей для створення надійних паролів. З їх допомогою можна створити таку послідовність чисел, властивості якої будуть схожі на властивості послідовності випадкових чисел.

Аналіз останніх досліджень і публікацій. Проблема досягнення “чистої випадковості” цікавить учених уже не одне століття. Творець теорії ймовірностей, французький математик Паскаль, у XVII ст. уперше згадує в одній зі своїх праць термін “генератор випадкових чисел”. Під таким генератором розумів підкидання гральних костей або монети. Ці свої прості генератори він використовував для збирання статистичних даних з метою перевірки своєї теорії ймовірностей. Але знаючи вагу, початкове розташування монети, силу, прикладену для її підкидання, характер поверхні приземлення та деякі інші параметри, можна точно розрахувати, якою стороною вона впаде. Тому навіть тепер досить складно знайти процеси, значення яких неможливо спрогнозувати, маючи необхідні знання та обладнання.

У доком'ютерні часи випадкові числа отримували, витягаючи різнокольорові м'ячі з мішків, карти, кидаючи кості. Зрозуміло, що серйозні дослідження так проводити не можна, тому у 1927 р. Тіппетт опублікував першу таблицю випадкових чисел, яка була сформована з довільно взятих чисел з різних звітів [1].

У 1939 р. розроблено машину, яка механічно виробляла 100 000 випадкових чисел. До 1955 р. компанією RAND Corporation було отримано вже 1 000 000 випадкових чисел, згенерованих машиною [2].

Нині як генератор випадкових чисел використовуються кількісні виміри (наприклад, клацання лічильника Гейгера) радіоактивного розпаду якоїсь радіоактивної речовини. Для генерації випадкових чисел в електронно-обчислювальній техніці зазвичай використовують чітко детерміновані ітераційні процеси, спектри яких являють собою ідеальний білий шум.

Оскільки комп'ютери та калькулятори працюють з обмеженою кількістю цифр після коми, то через якийсь час комп'ютерний генератор випадкових чисел зациклюється і починає повторювати ті числа, які були на самому початку і в тому ж порядку. Таким чином, комп'ютерні генератори випадкових чисел не є справжніми генераторами випадкових чисел, а їх спектр має одну виділену частоту повторення всієї послідовності.

Комп'ютерні генератори випадкових чисел описано ще в кінці XIX і на початку XX ст. у працях багатьох математиків, зокрема Пуанкаре. Після появи комп'ютерів у 40-х рр. XX ст. ці генератори були реалізовані у вигляді комп'ютерних програм.

Природні генератори, що ґрунтуються на радіоактивному розпаді, теж повторюють попередні числа, якщо дуже довго чекати, адже всі вимірювальні прилади працюють з обмеженою точністю.

Тому говорячи про "випадкові числа", які були створені за допомогою генераторів випадкових чисел навіть природного характеру, завжди йдеться саме про послідовність псевдовипадкових чисел (ППВЧ).

Послідовність псевдовипадкових чисел – це така послідовність, статистичні властивості якої схожі на властивості послідовності випадкових чисел.

Уперше запропонував їх використовувати Джон фон Нейман у 1946 р. Його метод полягав у такому: n-розрядне число підносилось до квадрата і з нього вибиралися середні n цифр. Метод був дуже недосконалий, послідовності майже завжди перетворювалися на нуль або зациклювалися з коротким періодом.

В основі програмних генераторів, як правило, лежать рекурентні формули. Зазвичай, вони генерують цілі числа, рівномірно розподілені на відрізку від 0 до деякого максимального m. Щоб отримати числа з плаваючою комою, рівномірно розподілені на [0,1), кожен отриманий результат ділять на m.

Д. Е. Кнут у праці [3] запропонував генератор, що ґрунтується на ідеї отримання ППВЧ з гарними статистичними характеристиками під час складних математичних операцій.

Незважаючи на подібну складність, цей алгоритм швидко зійшовся до числа 6 065 038 420, яке через невелику кількість кроків перетворилося на себе ж [3].

Цікавим фактом генерації випадкових чисел є те, що у 2012 р. австралійські вчені навчилися використовувати флуктуації вакууму для отримання істинно випадкових чисел. Робота опублікована в журналі Applied Physics Letters, її короткий зміст можна прочитати на сайті ABC Science. Для отримання випадкових чисел дослідники використовували лазер, випромінювання якого поділяли на два промені. Інтенсивність променів порівнювали між собою. Оскільки світло має квантову природу, різниця в інтенсивності була не постійною, а коливалася навколо середнього значення. Використовуючи описаний метод, учені отримували випадкові числа зі швидкістю 2 Гб на секунду. Дослідники виклали приклади випадкових чисел на своєму сайті, крім того, там же здійснюється онлайн-трансляція одержуваних випадкових чисел безпосередньо з лабораторії [4].

Попри успіх наведеного дослідження, воно досить затратне з погляду необхідних ресурсів і прийнятне не для всіх можливих задач. Тому оптимізація сучасних методів генерації псевдовипадкових послідовностей – задача, що буде актуальною ще довгий час.

Дослідження даної предметної області актуальне, оскільки лише 40–60 % [5] паролів для різних генераторів успішно проходять усі тести на випадковість.

У наукових працях описано відсоткове співвідношення проходження тестів на випадковість та їх успішного результату для найпопулярніших генераторів псевдовипадкових чисел. Але не виділено ті діапазони вхідних значень коефіцієнтів, для яких генератори дають найкращий вихідний результат. Крім того, нині немає конкретних вимог до початкового значення послідовності, за винятком умови цілочисельності, описаної в праці Кнута [3].

Тому можна сказати, що вдосконалення генераторів, зокрема визначення впливу взаємозалежності коефіцієнтів на якість генерації випадкових послідовностей, – задача, що потребує свого розв'язання.

Мета статті – пошук наборів параметрів і стартових значень генерації псевдовипадкових чисел для найкращого проходження тестів на випадковість.

Виклад основного матеріалу. На основі аналізу найпоширеніших методів генерації псевдовипадкових чисел зроблено висновок, що в контексті поставленої задачі створення парольного захисту вибір змішаного лінійного конгруентного методу [6] найоптимальніший.

Для змішаного лінійного конгруентного методу кожен наступний член послідовності розраховується за формулою:

$$X_{n+1} = (aX_n + c) \bmod m,\ n \in [0;\infty),$$

де a, c, m – параметри генерації.

Вибір такого підходу обумовлено низкою переваг зазначеного методу.

1. Відсутність чіткої прив'язки до конкретних значень параметрів генерації, що дозволяє проводити експерименти відповідно до мети дослідження.

2. Простий і зрозумілий алгоритм реалізації.

3. Невеликий період генерованої послідовності (порівняно з іншими методами) компенсується кращими властивостями випадковості послідовності.

4. У зв'язку з використанням методу тільки для задачі парольного захисту періодичність генерованої послідовності можна вважати достатньою, оскільки для створення надійного і зручного пароля використовують принцип розумної достатності: послідовність обмежується 8–10 символами.

У рамках даного дослідження обрано методи тестування псевдовипадкових послідовностей, запропоновані Д. Кнутом. Виконано перевірку згенерованої послідовності на критерій розподілу χ^2, а також критерію серій та серіальної кореляції.

Дослідження проводилося в 3 етапи.

1. Визначення відсотка проходження тестів на випадковість для випадкових значень параметрів генерації змішаним лінійним конгруентним методом.

2. Визначення області початкового наближення обраних параметрів генерації, якій відповідає найвищий відсоток проходження тестів на випадковість.

3. Оптимізація параметрів генерації методом крутого сходження.

У рамках першого етапу дослідження параметри генерації a, c та m обиралися випадково. Було згенеровано 2000 паролів довжиною 10 символів. За результатом тестування паролів 3 методами, лише 841 пароль успішно пройшов усі тести, а це становить лише 42,05 % від загальної кількості згенерованих паролів.

На другому етапі запропоновано початкове наближення до області оптимуму у вигляді лінійної залежності між параметрами *a, c* і *m*.

У ході описаного етапу дослідження проведено двадцять сім експериментів для різних лінійних залежностей між зазначеними коефіцієнтами. Для кожного експерименту згенеровано по 100 різних паролів, кожен з яких тестувався на випадковість трьома вищеописаними методами.

Результати експериментів можна подати у вигляді графіка.

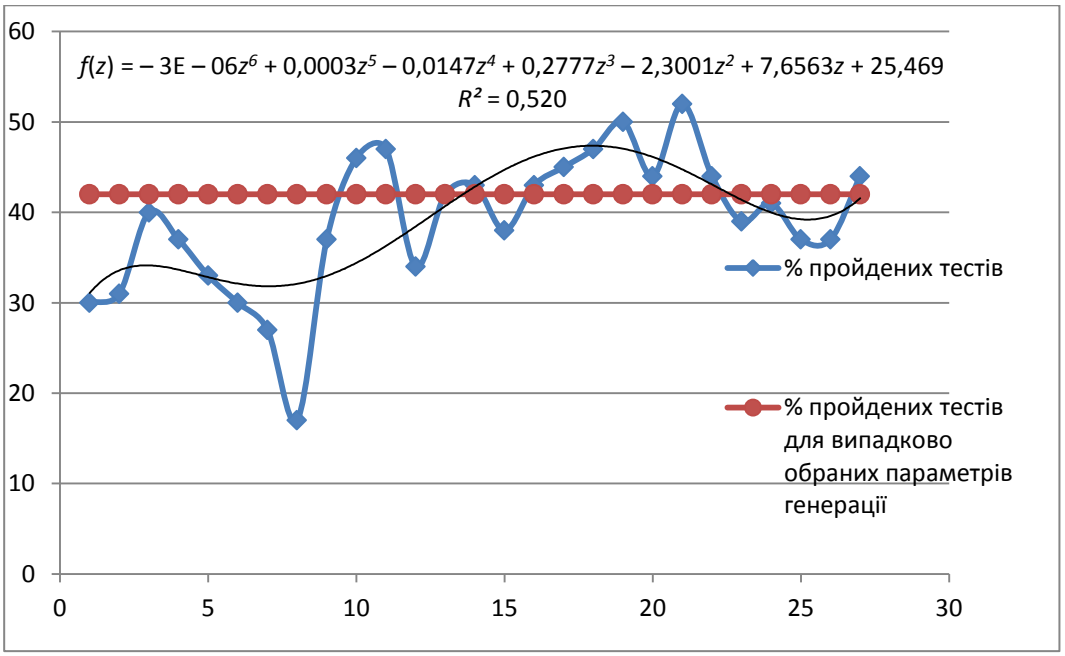

Рис. 1. Результати проходження тестів на випадковість

На графіку зображено три криві, перша з яких ілюструє відсоток успішного проходження тестів на випадковість для різних варіантів лінійної залежності між вхідними параметрами генерації; друга – середній відсоток проходження тестів на випадковість для випадково обраних параметрів генерації. Також побудована лінія тренду *f*(*z*) для першої кривої, яка показує, що існує тенденція до зростання відсотка успішного проходження зі збільшенням лінійної залежності між параметрами.

Аналізуючи графік, можна зробити висновок, що з використанням досить малих значень параметрів генерації *a* і *c*, знайдені паролі не проходять тести на випадковість навіть на рівні, отриманому для випадково обраних значень. Для більшості результатів відсоток проходження тестів перебуває в діапазоні від 37 до 47 %, тобто в радіусі отриманих результатів для випадкових значень параметрів генерації з 5 % допустимою похибкою. Але також виявлено область, для якої отримані значення досягають 52 %, тобто перевищують результати з випадковими параметрами навіть з урахуванням очікуваної помилки. Саме ця область і була обрана як початкове наближення до області оптимуму.

Отримані значення використано для побудови рівняння регресії на першому кроці крутого сходження [7].

На останньому етапі дослідження проведено два кроки оптимізації параметрів генерації методом крутого сходження, з побудовою рівняння регресії на кожному кроці за допомогою повного факторного експерименту. Як функцію відгуку експерименту (*y*) було обрано відсоток успішного проходження тестів на випадковість.

Оскільки обрано три фактори, то відповідно до методики повного факторного експерименту слід провести розрахунки відгуку *y* зі значеннями факторів на двох рівнях – нижньому та верхньому. Середнє значення факторів, нижній та верхній рівень відповідно становлять:

1) a_c = 74 623; $a_н$ = 16 077; $a_в$ = 133 169.

2) c_c = 43 035; $c_н$ = 9263; $c_в$ = 76 788.

3) X_{0c} = 41; $X_{0н}$ = 2; $X_{0в}$ = 81.

Матриця плану повного факторного експерименту в істинних значеннях параметрів та отримані значення відсотків успішного проходження тестів на випадковість *y* наведено в табл. 1.

Таблиця 1

Матриця плану та результатів повного факторного експерименту

№ п/п	x_0	x_1	x_2	x_3	y
1	1	16 077	9263	2	56
2	1	133 169	9263	2	39
3	1	16 077	76 788	2	51
4	1	133 169	76 788	2	40
5	1	16 077	9263	81	76
6	1	133 169	9263	81	50
7	1	16 077	76 788	81	39
8	1	133 169	76 788	81	53

де x_0 – додатковий фактор, уведений для оцінки вільного члена моделі [7], x_1 – параметр генерації *a*, x_2 – параметр генерації *c*, x_3 – параметр генерації X_0 *(початкове значення генерації)*, *y* – відсоток успішного проходження тестів на випадковість. Відповідно до матриці експерименту можна побудувати систему лінійних рівнянь, котра відображатиме лінійну модель процесу:

$$\begin{cases} a_0 + 16077a_1 + 9263a_2 + 2a_3 = 56 \\ a_0 + 133169a_1 + 9263a_2 + 2a_3 = 39 \\ a_0 + 16077a_1 + 76788a_2 + 2a_3 = 51 \\ a_0 + 133169a_1 + 76788a_2 + 2a_3 = 40 \\ a_0 + 16077a_1 + 9263a_2 + 81a_3 = 76 \\ a_0 + 133169a_1 + 9263a_2 + 81a_3 = 50 \\ a_0 + 16077a_1 + 76788a_2 + 81a_3 = 39 \\ a_0 + 133169a_1 + 76788a_2 + 81a_3 = 53 \end{cases}.$$

Для знаходження значень коефіцієнтів a_0, a_1, a_2 та a_3 перейдемо до нормованих факторів x за формулою:

$$x_i = \frac{2(x_i - x_{iC})}{x_{iB} - x_{iH}}, \quad i \in [1;4]$$

та розрахуємо значення коефіцієнтів за формулою [7]:

$$a_k = \frac{1}{2^n} \sum_{i=1}^{2^n} y_i x_{ki}.$$

Нормований план повного факторного експерименту і відповідні відгуки наведено в табл. 2.

Таблиця 2

Нормований план повного факторного експерименту

№ п/п	x_0	x_1	x_2	x_3	y
1	1	–1	–1	–1	56
2	1	1	–1	–1	39
3	1	–1	1	–1	51
4	1	1	1	–1	40
5	1	–1	–1	1	76
6	1	1	–1	1	50
7	1	–1	1	1	39
8	1	1	1	1	53

Розрахувавши значення коефіцієнтів регресійної моделі, отримуємо такий результат.
Лінійна модель

$a_0 = 50{,}5$
$a_1 = -5$
$a_2 = -4{,}75$
$a_3 = 4$

Лінійна модель процесу матиме такий вигляд:

$$y = 50{,}5 - 5x_1 - 4{,}75x_2 + 4x_3.$$

Для аналізу результатів необхідно перейти до моделі у звичайних змінних.
Коефіцієнти моделі у звичайних координатах

$a_0 = 58{,}723\ 690$
$a_1 = -0{,}000\ 085$
$a_2 = -0{,}000\ 141$
$a_3 = 0{,}101\ 266$

З урахуванням розрахованих коефіцієнтів модель у звичайних змінних набуває такого вигляду:

$$y = 58{,}7369 - 0{,}00008_{5x_1} - 0{,}00014_{1x_2} + 0{,}10126_{6x_3},$$

де x_1 – параметр генерації *a*, x_2 – параметр генерації *c*, x_3 – параметр генерації X_0 *(початкове значення генерації)*, y – відсоток успішного проходження тестів на випадковість.

Аналізуючи отриману модель, можна зробити висновок, що найбільший вплив на функцію відгуку має фактор початкового значення. При детальному розгляді значень функції відгуку мож-

на сказати, що три експерименти показали значення, яке ненабагато менше результату, отриманого з випадковою генерацією параметрів; але навіть при цьому вони перебувають у радіусі допустимої похибки. Один з результатів перевищує очікуваний на 34 %. Це досить гарний показник, проте лише за одним значенням не можна стверджувати про покращання якості генерації. Для її оцінки пропонується розглядати 2 показники: середнє значення за всіма експериментами для моделі та різницю між максимальним і мінімальним значеннями для функції відгуку. Для досліджуваної системи середнє значення становить 50,5 %, різниці між границями функції відгуку – 37 %. Отримані результати свідчать про необхідність зменшення різниці граничних значень і збільшення середнього показника проходження тестів.

Також важливо проаналізувати безпосередньо поведінку моделі. Якщо обчислити очікувані результати функції відгуку для розрахованої моделі з обраними значеннями параметрів генерації, то можна отримати результати, які вказують на середній показник абсолютного відхилення, що становить 6,88 %. Таке значення досить велике, особливо якщо взяти до уваги максимальне значення відхилення в понад 15 %. Тому даний аспект також потребує доопрацювання в наступній ітерації експерименту.

Таблиця 3

Відхилення розрахованої та експериментально отриманої функції відгуку моделі

№ п/п	*a*	*c*	X_0	*y*	*y* (розрах.)	відхилення
1	16 077	9263	2	56	56,266 804	–0,266 804
2	133 169	9263	2	39	46,313 984	–7,313 984
3	16 077	76 788	2	51	46,745 779	4,254 221
4	133 169	76 788	2	40	36,792 959	3,207 041
5	16 077	9263	81	76	64,266 818	11,733 182
6	133 169	9263	81	50	54,313 998	–4,313 998
7	16 077	76 788	81	39	54,745 793	–15,745 793
8	133 169	76 788	81	53	44,792 973	8,207 027

Хоча вплив усіх факторів досить незначний, можна виконати ще один крок крутого сходження, збільшивши ті параметри генерації, коефіцієнти за яких у результуючому рівнянні мають додатне значення, і зменшивши ті, коефіцієнти за яких мають від'ємні значення. Збільшенню підлягає початкове значення генерації, але в рамках поставленої задачі як верхню межу вже було використано максимальне значення вказаного параметра. Якщо абстрагуватися від питання генерації паролів, тобто тісної прив'язки до числово-символьних значень, можна згадати, що під час генерації довільної псевдовипадкової послідовності з використанням лінійного конгруентного методу немає верхньої межі для початкового значення генерації. Тому в рамках задачі дослідження пропонується збільшити початкове значення генерації, а для відображення його в паролі – присвоїти деякий довільний символьний знак.

Для другого кроку крутого сходження фактори набувають таких значень:

4) $a_c = 8859$; $a_н = 16\,077$; $a_в = 16\,077$;

5) $c_c = 5010$; $c_н = 938$; $c_в = 9263$;

6) $X_{0c} = 41$; $X_{0н} = 83$; $X_{0в} = 101$.

Провівши повний факторний експеримент з усіма можливими комбінаціями зазначених параметрів та отримавши функції відгуку, можна зробити аналогічні розрахунки і побудувати модель, яка характеризуватиме поведінку системи для вхідних значень факторів (параметрів генерації *a, c* та X_0) у вигляді:

$$y = 57{,}599594 + 0{,}00005_{2x_1} - 0{,}0003_{3x_2} + 0{,}12_{5x_3},$$

де x_1 – параметр генерації *a,* x_2 – параметр генерації *c,* x_3 – параметр генерації X_0 *(початкове значення генерації), y* – відсоток успішного проходження тестів на випадковість.

Як видно з рівняння залежності, найбільший вплив на успішність проходження тестів на випадковість має саме вдало обране початкове значення генерації X_0.

Розрахунки функції відгуку за отриманою моделлю дозволяють зробити висновок, що середнє відхилення розрахованої та експериментально визначеної поведінки системи становить за модулем менше 2 % з максимальним відхиленням 4 %. Такі показники є прийнятними та свідчать про те, що модель досить добре описує поведінку реальної системи. Окрім того, слід звернути увагу на безпосередні показники функції відгуку системи. Різниця між максимальним і мінімальним значенням становить 9 %, а середній показник проходження тестів на випадковість за всіма дослідами дорівнює 67,88 %, що на 25 % перевищує отриманий результат для генерації з випадково обраними параметрами. Отримані результати наглядно ілюструють дієвість обраного підходу до оптимізації параметрів генерації псевдовипадкових чисел.

На основі проведеного дослідження можна рекомендувати такі діапазони значень параметрів генерації для створення паролів з використанням лінійного конгруентного методу:

1) a = [1641; 16 077];

2) c = [938; 9263];

3) X_0 = [81; 101].

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямку. Основними практичними результатами даної роботи можна вважати виявлення залежності між параметрами генерації та відсотком успішного проходження тестів на випадковість, побудову математичної моделі, яка відображає залежність між параметрами генерації та відсотком успішного проходження тестів на випадковість, та визначення оптимальних діапазонів значень параметрів генерації, які сприяють підвищенню випадковості згенерованих послідовностей. За результатами дослідження вдалося:

– підвищити середній рівень успішного проходження тестів на випадковість лінійного конгруентного генератора на 25 %;

– підтвердити дієвість запропонованого підходу до оптимізації параметрів генерації псевдовипадкових чисел шляхом використання методу крутого сходження та повного факторного експерименту;

– поставити задачу поширення отриманих результатів на інші генератори псевдовипадкових чисел.

Список використаних джерел:

1. Генерація псевдовипадкових чисел [Електронний ресурс]. – Режим доступу : http://www.habrahabr.ru/post/132217

2. Кадан А. М. Генерация псевдослучайных чисел : курс лекций / Кадан А. М. – Гродно.

3. Кнут Д. Искусство программирования / Кнут Д. – 3-е изд. – М. : Вильямс, 2007. – Т. 2 : Получисленные алгоритмы. – С. 832.

4. Из квантового вакуума получили случайные числа [Электронный ресурс]. – Режим доступа : http://www.lenta.ru/news/2012/04/16/randomiser

5. Молотков О. Дослідження генераторів псевдовипадкових чисел як засобу генерації паролів / О. Молотков, О. Карнін // Вісник Академії митної служби України. – 2006. – № 4. – С. 103–108.

6. Иванов М. А. Теория, применение и оценка качества генераторов псевдослучайных последовательностей / М. А. Иванов, И. В. Чугунков. – М. : КУДИЦ-ОБРАЗ, 2003. – 240 с.

7. Адлер Ю. П. Планирование эксперимента при поиске оптимальных условий / Адлер Ю. П. – М. : Наука, 1976. – 279 с.