Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра дискретной математики и
информационных технологий

Генерация псевдослучайных чисел с помощью p-адических эргодических преобразований: равномерно распределённые последовательности и их линейная сложность

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 271 группы
направления 09.04.01 «Информатика и вычислительная техника»
факультета компьютерных наук и информационных технологий
Кардаша Василия Ивановича

Научный руководитель
к. ф.- м.н., доцент _____ _____ Л.Б. Тяпаев
подпись, дата

Зав. кафедрой
к. ф.- м.н., доцент _____ _____ Л.Б. Тяпаев
подпись, дата

Саратов 2018

ВВЕДЕНИЕ Актуальность работы. Теория динамических систем в полях p-адических чисел является важной частью алгебраической и арифметической динамики. Изучение p-адических динамических систем мотивировано их применениями в различных областях математики, физики, генетики, биологии, когнитивной науки, нейрофизиологии, информатики, криптологии и т.д. В частности, p-адические динамические системы нашли применение в криптографии, что

стимулировало интерес к изучению проекций непрерывных p-адических отображений в евклидовой плоскости их распределению и линейной сложности. С точки зрения приложений, потребность изучения эргодичности p-адических динамических систем возникает при решении задачи генерации псевдослучайных чисел: p-адические эргодические динамические системы

мы дают широкий класс превосходных генераторов псевдослучайных чисел, которые играют столь важную роль в криптографии, а также в других прикладных областях, таких как численный анализ и компьютерное моделирование.

Генератор гаммы в простейшем случае, на самом деле, есть автономный автомат. Любое состояние автомата x_i представляет собой некоторую бинарную последовательность длины n, значит размер пространства состояний, очевидно, равно 2^n, а само пространство $X = Z/2^nZ$ – кольцо вычетов по модулю 2^n; элемент этого кольца ассоциирован с бинарной строкой длины n; начальное же состояние x_0 (выбранная в начальный момент времени некоторая бинарная строка длины n) это ключ (и он секретен). Значение n, очевидно, определяет длину ключа. Выходная реакция автомата y_i есть бинарная последовательность длины m, а значит $Y = Z/2^mZ$. Последовательность x_0, $f(x_0)$, $f^2(x_0)$, . . . – есть траектория ключа, а, последовательность $y_0, y_1, y_2, \ldots$ – требуемая гамма.

Генератор гаммы должен удовлетворять следующим требованиям:

1. Псевдослучайность: функция переходов f должна обеспечить псевдослучайность, в частности, равномерное распределение и максимальную длину периода последовательности состояний.

2. Стойкость: функция выходов g должна обеспечивать псевдослучайность выходной последовательности – равномерное распределение и длинный период. Более того, для данного y_i нахождение x_i из уравнения $y_i = g(x_i)$ должно быть вычислительно трудоемкой задачей.

3. Быстродействие: для программной реализации генератора функции f и g должны быть простыми композициями элементарных арифметических и поразрядных логических операций.

Для выполнения данных требований необходимо использовать транзитивные на $Z/2^nZ$ (а, на самом деле, эргодические на кольце целых 2-адических чисел Z_2) и сбалансированные (на самом деле, сохраняющие меру на Z_2) отображения в качестве функций f и g соответственно (здесь потребуются методы алгебраической динамики.

Целью данной работы является определение линейной сложности последовательностей, порождаемых с помощью эргодических совместимых функций, используемых в псевдослучайных генераторах при синтезе поточных шифров.

Задачи:

1) Изучение линейной сложности последовательностей;
2) Рассмотреть конгруэнтные генераторы: инверсные, полиномиальные, экспоненциальные;
3) Разработать и описать инструмент для удобного представления графиков рассматриваемых отображений на евклидовой плоскости;
4) Экспериментальное наблюдение графиков функций и их композиций, используемых в псевдослучайных генераторах.

Работа выполнена на 62 страницах машинописного текста, состоит из введения,7 глав, заключения, содержит 13 рисунков, 1 приложения, список литературных источников содержит 29 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ. В первой главе «Элементы теории p-адических чисел » говорится об области p-адических чисел, для таких чисел не выполняется аксиома Архимеда. Иными словами, речь идет о неархимедовой области математики. Разработка новых подходов для анализа защищённости информационных систем – актуальная задача в области информационной безопасности. Особое место занимает проблема оценки криптографических средств защиты на предмет их соответствия потребностям организации. Задача определения эффективности системы защиты информации при использовании криптографических методов защиты зачастую более трудоёмкая, чем их разработка, требует наличия специальных знаний и более высокой квалификации. Анализ криптостойкости шифра – научная, а не инженерная задача. Даются основные определения и понятия из теории p-адических чисел.

Во второй главе «Линейный ранг » деется определение понятию линейная сложность. Это важная мера криптографической сложности бинарной последовательности.

Линейная сложность характеризует распределение элементов последовательности, порождаемой итерированием отображения $f : Z_p \to Z_p$:

$$x_0, x_1 = f(x_0), x_2 = f^2(x_0), \ldots .$$

Здесь Z_p обозначает пространство целых p-адических чисел; элементы этого пространства (т.е. целые p-адические числа) – это просто бесконечные (влево) строки символов над алфавитом $F_p = \{0, 1, 2, \ldots p-1\}$, p простое. Пространство Z_p есть коммутативное кольцо со структурой проективного предела: Z_p есть проективный предел колец вычетов

$$\ldots, Z/p^nZ, \ldots, Z/p^2Z, Z/pZ.$$

По определению, линейная сложность $\lambda_R(\chi)$ последовательности $\chi = x_0, x_1, x_2, \ldots$ над коммутативным кольцом R есть наименьшее неотрицательное целое $r \in N_0$ такое, что существуют $c, c_0, c_1, \ldots, c_{r-1} \in R$ (не все равные нулю) такие, что для любых $i = 0, 1, 2, \ldots$

$$C + \sum_{j=0}^{r-1} c_j \cdot x_{i+j} = 0.$$

Если же такого r не существует, то $\lambda_R(\chi) = \infty$. Линейную сложность еще можно определить как размерность регистра сдвига с линейной обратной связью (РСЛОС). Понятие линейной сложности имеет геометрическую интерпретацию. Например, если $R = Z/p^kZ$ есть кольцо вычетов по модулю p^k, то геометрически $\lambda_R(\chi) = r$ означает, что все точки с координатами $\left(\frac{x_i}{p^k}, \frac{x_{i+1}}{p^k}, \ldots \frac{x_{i+r-1}}{p^k}\right)$, $i = 0, 1, 2, \ldots$ единичного гиперкуба (размерности r) лежат на параллельных гиперплоскостях.

В криптографическом ракурсе нас интересует линейная сложность последовательности, генерируемой отображением $f : Z_p \to Z_p$ при $p = 2$.

Третья глава «Поточные шифры и генераторы псевдослучайных чисел» содержит сведения о регистрах сдвига с линейной обратной связью, линейных конгруэнтных генераторах, псевдослучайных генераторах, а так же в ней описан метод срединных квадратов.

По данному преобразованию F: $Z/NZ \to Z/NZ$ кольца вычетов Z / NZ по модулю $N > 1$ генератор, по определению, чей закон рекурсии есть f, а

Элемент $x_0 \in Z / NN$ есть начальное состояние генератора (источник),

порождает выходную последовательность

$$x_0, x_1 = f(x_0), x_2 = f(x_1), \ldots$$

элементов множества $\{0,1, \ldots N-1\}$ по следующему рекурсивному правилу

$$x_{i+1} \equiv f(x_i) \pmod N$$

Генератор называется конгруэнтным, тогда и только тогда, когда f сохраняет все конгруэнции (т.е. все сравнения $\equiv$ по модулю M) кольца вычетов Z / NZ: т.е. $f(x) \equiv f(y) \pmod M$ как только $x \equiv y \pmod M$ и M делит N ($M \neq 1$).

Из китайской теоремы об остатках следует, что выходная последовательность конгруэнтного генератора имеет наибольший период (длины) тогда и только тогда, когда любая функция $f \bmod p^n$ транзитивна по

модулю p^n, где n = $\mathrm{ord}_p N$ (здесь $\mathrm{ord}_p N$ обозначает показатель степени простого числа p в разложении натурального N по степеням простых сомножителей).

Транзитивная по модулю p^n функция, по определению, есть одноцикловая перестановка элементов множества $\{0, 1 \ \ldots \ p^n - 1\}$.

В частности, интерес представляют следующие конгруэнтные генераторы:

1) Инверсный: $f(x) = 1 + x + \frac{p^2}{1+p^x}$;

2) Экспоненциально-инверсный: $f(x) = 1 + x + p^2 \cdot (1 + px)^{\frac{1}{1+px}}$

3) Общий инверсный генератор $\mathrm{inv}(x) = p^{ord_{p^x}} \cdot \left(\frac{1}{p^{ord_{p^x}}} \right)^{-1}$; $x \in Z_p \setminus \{0\}$;

$f(0) = 0$

Генератор $\mathrm{inv}(x)$ есть 1-Липшицева функция (следовательно, совместимая). Более того, она дифференцируемая (хотя не равномерно) везде на Z_p кроме 0.

Если f 1-Липшицева эргодическая функция на Z_2, то и $f(\mathrm{inv}(x))$ и $\mathrm{inv}(f(x))$ эргодичны. Обратно, если функция $f(\mathrm{inv}(x))$ или $\mathrm{inv}(f(x))$ эргодична, то и $\mathrm{inv}(x)$ эргодическое отображение.

Тогда, например,

1. $f(x) = \mathrm{inv}(2x^2) + \mathrm{inv}(7x) + 1$ и $f(x) = \mathrm{inv}(2x^2 + 7x + 1)$ эргодичны на Z_2 (поскольку полином $2x^2 + 7x + 1$ эргодичен);
2. $f(x) = 3 \cdot \mathrm{inv}(x) + 3^{inv(x)}$ эргодическая функция на Z_2 (в силу эргодичности отображения $3 \cdot x + 3^x$).

Экспериментальное наблюдение линейной сложности последовательностей, построенных различными генераторами, в том числе линейным, квадратичным, обратным, с помощью «графиков» – основная цель работы. Экспериментальным инструментом является компьютерная программа, которая по заданной функции f: $Z_2 \to Z_2$ строит «график» редуцированной функции $f \bmod 2^k$: $Z/2^k Z \to Z/2^k Z$ в единичном квадрате плоскости. С ее

помощью можно наблюдать расслоения точек графика по гиперплоскостям (параллельным прямым).

В четвертой главе «Полиномиальные и экспоненциальные алгоритмы» даются определения большинству практических алгоритмов, с которыми работают программисты. Полиномиальным (или алгоритмом полиномиальной временной сложности) называется алгоритм, у которого временная сложность есть O(p(n)), где p(n) – полином от n. Задачи, для решения которых известен алгоритм, сложность которого составляет полином заданной, постоянной и не зависящей от размерности входной величины n степени, называют “хорошими” и относят их к классу P.

Экспоненциальной по природе считается задача, сложность которой не менее порядка xn , где x – константа или полином от n. Например, это задачи, в которых возможное число ответов уже экспоненциально. В частности, к ним относятся задачи, в которых требуется построить все подмножества заданного множества или все поддеревья заданного графа. Экспоненциальные задачи относят к классу E.

В пятой главе «Равномерное распределение» дается подробное определение строгому равномерному распределению, и приводятся примеры. Пусть $\chi(S, x)$ есть характеристическая функция подмножества S некоторого пространства M, т.е. $\chi(S, x) = 1$ в случае, если $x \in S$, и $\chi(S, x) = 0$, если $x \notin S$. Более того, пусть пространство M снабжено мерой μ. По определению, последовательность $(x_i)_{i=0}^{\infty}$ называется равномерно распределенной, если выполняется следующее равенство

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \chi(S, x_i) = \mu(S)$$

для всякого измеримого (относительно меры μ) подмножества $S \subset M$.

Отображение F: M→M называется сохраняющим меру, если для любого измеримого подмножества $S \subset M$ справедливо $\mu(F^{-1}(S)) = \mu(S)$. Сохраняющее меру отображение F: M →M называется эргодическим, если у

него нет инвариантных подмножеств (т.е. таких что $F^{-1}(S) = S$) кроме меры 0 или 1: если выполняется $F^{-1}(S) = S$, то значит $\mu(S) = 0$, либо $\mu(S) = 1$.

В случае, когда M = Z_p (либо M = $Z_p{}^n$) эргодичность означает, что последовательность $(F^i(x_0))_{i=0}^{\infty} = \{x_0, F(x_0), F^2(x_0), \ldots\}$ является равномерно распределенной. Однако в криптографическом ракурсе эргодичность не гарантирует высокую линейную сложность (в идеале, бесконечную). Например, известная в криптографии функция Климова и Шамира f (x) = x + (x^2OR5), заданная на Z_2, эргодична, однако линейная сложность последовательности λ_{Z_2} $((f^i(x_0))_{i=0}^{\infty})$ равна двум (!), как и у последовательности, порожденной эргодической же функцией f (x) = x-1. Отображение f : Zp→Zp можно изучать с помощью, так называемого, «графика». Пусть для k = 1, 2, . . . множество Ek (f) есть множество точек $e_k^f(x)$ единичного квадрата [0, 1] × [0, 1] ⊂ R2 вида:

$$e_k^f(x) = \left(\frac{x \bmod p^k}{p^k}, \frac{f(x) \bmod p^k}{p^k} \right),$$

где $x \in Z_p$ – прообразу отображения f – целому p-адическому числу x, $x = \ldots x_k x_{k-1} \ldots x_1 x_0 \in Z_p$, ставиться в соответствие элемент кольца вычетов по модулю p^k (так называемая редукция по модулю p^k): $x \bmod p^k = x_{k-1} \ldots x_1 x_0 \in Z/p^kZ$; образ отображения f (x) Z_p заменяется своей редукцией (по модулю p^k): $f(x) \bmod p^k \in Z/p^kZ$; затем редукции $x \bmod p^k$ и $f(x) \bmod p^k$ (по сути, неотрицательные целые, не превосходящие p^k) делятся на p^k (тем самым, становятся точками отрезка [0, 1]).

Пусть E (f) есть замыкание множества E (f) = $\cup_{k=1}^{\infty} Ek$ (f) в топологии плоскости R^2 – «график» отображения f: $Z_p \to Z_p$. При фиксированном значении $k \in N$, мы имеем дело с «приближенным графиком» отображения f: $Z_p \to Z_p$, и «точным графиком» редуцированного отображения f mod p^k: Z/p^kZ Z/p^kZ.

Здесь через mod p^k обозначена редукция по модулю p^k, которая есть, на самом деле, эпиморфизм кольца целых p-адических чисел на кольцо вычетов по модулю p^k,

$$\mathrm{mod}\ p^k\colon\ Zp \to Z/p^kZ.$$

Целому p-адическому числу $x = \sum_{i=0}^{\infty} x_i \cdot p^{\,i}$ этот эпиморфизм ставит в соответствие неотрицательное целое число $\sum_{i=0}^{k-1} x_i \cdot p^{\,i}$ (записанное в системе счисления с основанием p). Ядром эпиморфизма mod p^k является шар $p^kZ_p = B_{p\text{-}k}\,(0)$ радиуса p^{-k} с центром в 0.

Любой полином с целыми (в том числе, и с целыми p-адическими коэффициентами) есть совместимое преобразование пространства Zp, или, например, композиция операций сложения, умножения и поразрядных логических операций (XOR, OR, AND, NOT) есть совместимая функция на Z_2. Более того, совместимое отображение f: $Z_p \to Z_p$ удовлетворяет p-адическому условию Липшица с константой 1:

$$|f(x) - f(y)|_p \le |x - y|_p$$

для любых x, y ∈ Zp; здесь символом $|\bullet|_p$ обозначена p-адическая норма (абсолютная величина). Из этого условия следует, что совместимая функция непрерывна относительно p-адической метрики (метрика (расстояние) d_p(x, y) для любой пары x, y ∈ Z_p, по определению, есть норма разности $|x - y|_p$). Метрика d_p(x, y) удовлетворяет усиленному неравенству треугольника

$$d_p(x, y) \le \max\ \{d_p(x, z),\ d_p(z, y)\},$$

для любой тройки x, y, z ∈ Zp a, следовательно, является неархимедовой (ультраметрикой). Таким образом, заключаем, что кольцо Z_p, наделенное метрикой d_p(x, y) является неархимедовым (ультраметрическим) пространством.

Пространство Zp наделяется не только метрикой, но и вероятностной мерой μp – нормализованной мерой Хаара (мера всего пространства Z_p полагается равной 1; элементарными измеримыми подмножествами являются шары $B_{p^{-l}}$(a) радиуса p^{-l} с центром в точке a ∈ Z_p, $\ell = 0, 1, 2, \ldots$ и, по определению, $\mu_p\,(B_{p^{-l}}(a)) = p^{-l}$).

Итерирование отображения f: $Z_p \to Z_p$ порождает динамическую систему на Z_p (динамическая система, по определению, есть тройка (S, μ, f),

где S пространство, наделенное мерой μ, f – измеримое и непрерывное преобразование пространства S).

Таким образом, тройка (Zp, μp, f) есть неархимедова динамическая система, а, последовательность x_0, $x_1 = f(x_0)$, $x_2 = f(x_1)$, . . . – орбита точки x_0.

В шестой главе «Описание инструмента для представления отображений в единичной евклидовой плоскости» основываясь на изученном теоретическом материале, приводится подробная характеристика программы, главной функцией которой является визуальное отображение графиков функций на экран монитора для дальнейшего исследования. Описываются основные функции, интерфейс и особенности программы.

В седьмой главе «Эксперименты по построению проекций совместимых отображений на евклидову плоскость» проводятся эксперименты в программе с описанием характера изображения распределения точек на графике, а также с визуальным вычислением линейного ранга по характеру заполнения плоскости.

Пример 1. Зададим функцию f (x) = 3 + 5 * x . В Результате работы программы наблюдаем отображение функции в евклидовой плоскости, на котором с увеличеним k четко видно, что точки графика превращаются в параллельные симметричные друг-другу линии. Мы можем сказать, что линейная сложность данной функции равна 2, она не подходит для применения в криптографии.

Пример 2. Рассмотрим функцию f (x) = x*x+3*x+2 .В Результате ее отображения видим явно фрактальную структуру, картинки являются симметричными, относительно своего центра, самоподобными или приближённо самоподобными. С увеличением длины слов k лакун становится все меньше, при k = 20 наблюдаем равномерное распределение – точки закрашивают весь единичный квадрат. Линейный ранг функции стремиться к бесконечности.

Пример 3. Рассмотрим функцию f (x)= 3^x + 3x + 3 .В Результате ее отображения видим, что видим, что при увеличении дины слов k точки

графика все больше заполняют единичный квадрат, линейный ранг функции стремиться к бесконечности.

Пример 4. Рассмотрим функцию $f(x)= 2x^2 + 3x + 1$. В результате ее отображения видим, что с увеличением k возрастает количество точек отображения в евклидовой плоскости, и она становится все более заполненной, в какой-то момент на «графике» мы видим даже отсутствие лакун. Это значит, что линейный ранг заданной нами функции стремиться к бесконечности.

Пример 5. Рассмотрим функцию $f(x) = x + ((x^2)\ OR\ (-131065))$. В Результате ее отображения видим, что с увеличением k, как и в предыдущем примере, наша плоскость все более заполняется точками, но, после длинны слов k = 18 видно расслоение графика на четкие горизонтальные полосы, а при k = 22 наблюдается 1 полоса на графике ровно по диагонали. Это функция «Климова и Шамира».

Пример 6. Рассмотрим функцию $f(x) = x + (x^2\ OR\ 5)$.В Результате ее отображения видим, что для каждого отображения характерно присутствие симметрии на «графике», с увеличением k лакун становится все меньше.

Пример 7. Рассмотрим функцию $f(x) = x + (x^2\ OR\ 87305)$. В Результате ее отображения видим, что при длине слов k = 15 и особенно при k = 18, точки графика распределяются по гиперплоскостям, на нашей единичной плоскости наблюдаем, что все точки графика разбились на параллельные друг - другу прямые линии. Можно сказать, что у этой функции высокий линейный ранг.

ЗАКЛЮЧЕНИЕ

В ходе проделанной научно-исследовательской работы был разработан инструментарий, реализующий построение «графиков» совместимых (заданных функцией) 1-Липшицевых отображений на единичную евклидову плоскость, в том числе используя композиции различных непрерывных функций p-адических чисел. С помощью этого инструментария было проведено множество экспериментов и наблюдений за поведением большого

числа отображений «графиков» функций и композиций. Подробно и изучена линейная сложность, так же изучены генераторы псевдослучайной последовательности. Хорошие статистические свойства последовательности не гарантируют ее хороших криптографических свойств. Для достижения криптостойкости необходимо использовать комбинирующие генераторы и комбинирующую функцию высокой степени (либо увеличивать длину РСЛОС). Рассмотренные условия криптостойкости необходимы, но не достаточны.

Если динамическая система эргодична (т.е. f есть эргодическое преобразование), то орбита любой точки – равномерно распределенная последовательность, но «график» отображения f это свойство орбит не отражает. Однако, если орбиты, порожденные эргодическим отображением имеют высокую линейную сложность, то на «графике» (в единичном квадрате) будет наблюдаться распределение точек орбит по параллельным прямым (в случае бесконечной линейной сложности – «график» содержит все точки единичного квадрата).