ИНФОРМАТИКА

УДК 681.3.06

ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ КВАЗИОБРАТИМЫХ И НЕОБРАТИМЫХ ФУНКЦИЙ, ПРИМЕНЯЕМЫХ ПРИ ЗАЩИТЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ОТ НЕСАНКЦИОНИРОВАННОГО ИСПОЛЬЗОВАНИЯ

канд. техн. наук, доц. Д.О. ГЛУХОВ, М.В. МАТЮШ
(Полоцкий государственный университет)

Исследуется псевдослучайный характер квазиобратимых и необратимых функций. Проверяется гипотеза о законе распределения последовательности обращенных значений. Стохастические параметры рассматриваемой величины характеризуют ее с точки зрения применимости в системах защиты программного обеспечения от несанкционированного использования. Предложен подход к построению систем защиты программного обеспечения (ПО) от несанкционированного использования, основанный на применении методов криптографии, чувствительных к погрешности компьютерных вычислений. Предложены варианты реализации шифрующих, дешифрующих и хеш-функций избирательного действия на основе квазиобратимых и необратимых функций представимых вещественных чисел.

Введение. В настоящее время активно совершенствуется государственная политика Республики Беларусь и стран СНГ по защите авторских прав на программы для ЭВМ и базы данных, принимаются жесткие меры по борьбе с пиратством, преступлениями в информационных сетях. Такая политика позволяет всерьез говорить о том, что формируется рынок интеллектуальной собственности, в частности, программного обеспечения (ПО) и компонентов программного обеспечения [1].

Эти процессы требуют соответствующих шагов и от разработчиков программного обеспечения. Программа становится товаром, со всеми присущими для него характеристиками. Отрабатываются различные схемы продаж программного обеспечения. Разрабатываются новые способы защиты программного обеспечения от несанкционированного копирования и использования.

Когда речь идет о защите ПО как особого способа представления данных и знаний о механизмах манипулирования ими (алгоритмах), то необходимо отметить общие особенности ПО как объекта защиты:

1) необходимо учитывать, что защита выполняется в интерпретации того либо иного микропроцессорного устройства, а значит, оперирует системой команд и учитывает особенности аппаратной платформы;

2) необходимо учитывать отличия устройств вычислений с плавающей точкой различных производителей (Intel, AMD, MAC, SPARC) и возможные ошибки в различных моделях и архитектурах микропроцессоров [5].

Отмеченные особенности крайне важны для разработки средств защиты ПО. Учет особенностей вычислений – ключевой момент предлагаемой нами защиты. Причем само процессорное устройство уже определяет набор «естественных» ограничений, в рамках которых не только будет функционировать взламываемое ПО, но и будет вынужден протекать процесс взлома. **Это те правила игры, которые мы навязываем злоумышленнику и которые должны заведомо дать нам преимущество**.

Сформулируем цель разработки как метод защиты ПО:

1) метод, не нарушающий ход вычислительного алгоритма;

2) метод, имеющий множество инвариантных быстрых фрагментов верификации ключа, распределенных в программном коде;

3) метод, обеспечивающий заданную криптоаналитическую стойкость, заданную сложность обратного преобразования;

4) метод, для взлома которого потребовался бы более (на порядки) точный компьютер.

ФОРМАЛИЗАЦИЯ

Для формального определения рассмотрим функции, заданные на дискретном множестве представимых в машинном формате вещественных чисел $\boldsymbol{D}$. Будем говорить, что функция f двух переменных замкнута относительно $\boldsymbol{D}$, если $f(u, k) \in \boldsymbol{D}$ при $u, k \in \boldsymbol{D}$.

Функцию f будем называть обратимой, если существует обратная функция g такая, что для всех $u \in \boldsymbol{D}$:

$$g(f(u, k), k) = u. \quad (1)$$

Определение 1. Квазиобратимой функции представимых вещественных чисел

Функция f квазиобратима, если существует такая обратная функция g, что для всех $u \in \boldsymbol{D}$ выполняется неравенство:

$$u - \Delta u < g(f(u, k), k) < u + \Delta u, \quad (2)$$

где Δu – абсолютная погрешность обращения; $\Delta u = u\delta$ (δ – относительная погрешность обращения).

Теоретические основы влияния шумоподобных сигналов на процессы шифрования и дешифрования активно изучаются в теории кодирования телевизионных сигналов. Источником погрешности при кодировании таких сигналов являются погрешности АЦП, ЦАП преобразователей, электромагнитные шумы с известной плотностью распределения [2]. В нашем случае источником погрешности выступает вычислительное устройство, причем характеристики «зашумленности» шифруемых данных при определенном выборе квазиобратимых функций будут также приближаться к случайным последовательностям. Псевдослучайная погрешность вычисления, однако, обладает одной отличительной чертой – она повторима в точности при одинаковых начальных условиях. Ею можно управлять, **ее можно использовать**.

Определение 2. Необратимой функции представимых вещественных чисел

Функция f необратима, если не существует такая обратная функция g, что для всех $u \in \boldsymbol{D}$ выполняется неравенство (2) при заданной константе δ.

Необратимость функции отражает тот факт, что функция крайне чувствительна к погрешности вычисления. Поиск решения обратной задачи (вычисления обратной функции) затруднен ее плохой обусловленностью и малейшие изменения аргументов приводят к значительным отличиям результатов вычислений [6]. Малые невязки отражают значительные ошибки при вычислении неизвестных.

Определение 3. Функций шифрования и дешифрования представимых вещественных чисел

Если мы имеем две функции $f(u_1, k_1)$ и $g(u_2, k_2)$, зависящие от u_1 и u_2 и ключей k_1 и k_2, такие, что $f(g(u, k_2), k_1) = u$ для некоторой пары ключей k_1 и k_2 и любого u, тогда $g(.)$ и $f(.)$ – соответственно функции шифрования и дешифрования. Результат $s = g(u, k_1)$ называется криптограммой (или шифрограммой).

К функциям $g(.)$ и $f(.)$ предъявляются следующие требования:

1) они должны быть легко вычислимы для любых u, если известны k_1 и k_2;

2) вычисление $f(u, k_1)$ – трудная задача при неизвестном ключе k_1 (т.е., не зная ключа дешифрования, мы не можем вычислить исходное сообщение по криптограмме);

3) вычисление ключей k_1 и k_2 – трудная задача при наличии некоторого набора пар $<u, s>$ (имея набор криптограмм и исходных сообщений, мы не можем вычислить ключи);

4) вычисление k_2 (в случае k_2, отличного от k_1) при известном k_1 также должно быть трудной задачей (это требование относится к асимметричным шифрам. Для цифровой подписи k_2 и k_1 меняются ролями).

Надежность шифра определяется именно по его соответствию вышеперечисленным требованиям; при этом считается, что алгоритмы вычисления $g(.)$ и $f(.)$ известны всем (есть еще вариант так называемой Security by obscurity, т.е. защиты из-за неизвестности алгоритма, но он при оценке стойкости шифров не рассматривается) [4].

Если $k_1 = k_2$, то шифр называется симметричным, в противном случае – асимметричным. Примеры известных симметричных шифров – DES, IDEA, Blowfish, ГОСТ, асимметричных – RSA, ECC [3].

Определение 4. Хеш-функций представимых вещественных чисел

Основное предназначение криптографических хешей – контроль подлинности данных путем вычисления от них некоторой функции $h(.)$, дающей результат фиксированной (и обычно небольшой длины). Функция $h(.)$ должна удовлетворять следующим требованиям:

1) для любых сообщений u, $s = h(u)$ легко вычисляема;

2) задача нахождения такого u_2 (отличного от u), чтобы $h(u_2) = s$, должна являться трудной при неизвестном u;

3) задача нахождения такого u_2, что $h(u) = h(u_2)$ является трудной при известном u.

Большинство популярных хеш-функций генерируют хеш длиной 128 бит и более. Примерами наиболее распространенных хеш-функций являются MD5 и SHA. Значения хеш-функций часто используются в системах электронной цифровой подписи для генерации дайджеста сообщения, который затем и подписывается тем или иным алгоритмом. Также хеш-функции применяются в системах аутентификации для проверки паролей – открытый пароль пользователя не должен храниться в системе, вместо него хранится его хеш, который затем и сравнивается с хешем от пароля, вводимого пользователем при входе в систему [3].

Векторные функции представимых вещественных чисел

Поскольку рассмотренные квазиобратимые и необратимые функции представимых вещественных чисел замкнуты относительно универсума $\boldsymbol{D}$, то и в векторной постановке они сохраняют свои свойства квазиобратимости и необратимости. Верхней оценкой отклонения для квазиобращения в относительных величинах будет рассматриваться δ_u, определяющая отклонение некоторой векторной нормы по вектору $\boldsymbol{\delta}$.

Именно векторная постановка задачи построения шифрующих, дешифрующих и хеш-функций, определенных на дискретном множестве представимых в машинном формате вещественных чисел, и является ключевой в данной работе.

Новизна состоит в предложении избирательного криптографического преобразования, в рамках которого шифрующая и дешифрующая функции сохраняют свои функции обратимости только для некоторого конечного подмножества шифруемых сообщений, для остальных, строго говоря, обладают свойством квазиобратимости или даже необратимости.

Особенности множества представимых вещественных чисел *D*

Дискретное конечное множество представимых в машинном формате вещественных чисел ***D*** формируется из бесконечного множества вещественных чисел ***R*** заданием формата числа (разрядности мантиссы и порядка, а также правил интерпретации выбранного представления).

Вещественные числа из множества ***D*** расположены в метрическом пространстве ***R*** с заданной оценкой расстояния (метрикой) (–) неравномерно. Блоки чисел, образованные полным перебором всех вариантов мантиссы, следуют с двукратным расширением/сжатием при увеличении/уменьшении порядка.

Это множество обладает симметрией относительно числа 0. Оно при рассмотрении подмножества положительных чисел ограничено сверху и снизу максимальным и минимальным числом.

Особенности квазиобратимых функций представимых вещественных чисел

Класс рассматриваемых функций очень широк. Практически все известные нам трансцендентные функции обладают свойством квазиобратимости или необратимости, если они рассматриваются как функции, замкнутые относительно ***D***. Иными словами, компьютерные реализации трансцендентных функций вносят погрешности, которые превращают их в квазиобратимые или необратимые функции.

Приведем простейший пример функции, заданной произведением аргумента на число, и обратной ей функции деления (умножения на обратное число):

$$\begin{aligned} f(x,v,k) &= \frac{v+k}{x}; \\ g(x,f,k) &= f\cdot x - k, \end{aligned} \tag{3}$$

где $x \in \boldsymbol{D}$, что v – вектор эталонов; k – дополнительный ключ.

Отклонение значений восстанавливаемого аргумента в результате такого функционального преобразования иллюстрирует рисунок 1.

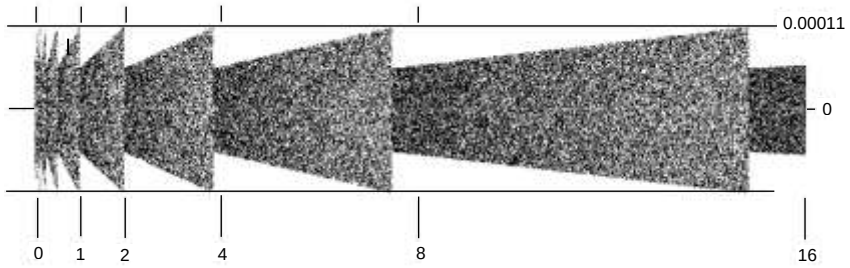

Рис. 1. Распределение погрешности квазиобращения для операций /,* для чисел в интервале от 0 до 16

Введение параметра k позволяет дополнительно увеличить погрешность на заданное количество разрядов, в частности, в рассматриваемом примере погрешность колебалась в пределах ±0,00011. Погрешность квазиобращения будет определяться по формуле:

$$\Delta_v = \left\| g(x,f,k) - v \right\|. \tag{4}$$

Распределение погрешности носит случайный характер или, по крайней мере, позволяет говорить о распределениях погрешности как о псевдослучайных последовательностях, соответствующих равномерному закону распределения вероятности в некотором диапазоне значений, однако четко прослеживается скачкообразное изменение пределов разброса восстанавливаемых значений, связанное с переходом степени двойки. Причем уникальные штампы разброса формируют полные переборы мантиссы, а это 4 503,599 627 370 496 триллиона комбинаций для мантиссы 52 бита (без учета знака) и 9 007,199 254 740 992 триллиона комбинаций для мантиссы 53 бита.

Изменение степени двойки, по сути, смещает уникальный штамп, пропорционально расширяя (в 2 раза) или сужая его.

Так, штамп, полученный прострелом 1000 точек интервала от 1 до 2, идентичен штампу по интервалу от 2 до 4.

Также наблюдается зеркальная симметрия относительно 0.

При постановке задачи в виде (3) будем говорить, что требуется определить такой вектор ключей x, что эталонный вектор v восстанавливается без погрешности. Для вектора v, элементы которого не превосходят 3-х порядков и ключа k 12-го порядка, происходит потеря 9 порядков, это для чисел двойной точности означает, что требование точного восстановления эталонного вектора обеспечивает одно число на 10^8 чисел. Оценка приблизительная и требует уточнения для каждого конкретного k.

Если предположить, что восстановленное значение аргумента представляет собой случайную величину с нормальным законом распределения, то она имеет следующие свойства:

1) может принимать непрерывный ряд значений от –∞ до +∞;

2) центр распределения случайной величины одновременно является центром симметрии, т.е. одинаковые отклонения результатов измерения в меньшую и в большую сторону от центра встречаются одинаково часто;

3) малые отклонения встречаются чаще больших, другими словами, реализуются с большей вероятностью [6].

Произведем проверку выполнения нормального закона распределения по критерию Пирсена (χ^2-критерию). В качестве исходных данных возьмем выборку из 100 случайных значений. Построим гистограмму распределения случайных величин и проверим, подчиняется ли выборка закону нормального распределения (распределения Гаусса):

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}. \qquad (5)$$

Причем в качестве $\bar{x}$ и σ берутся соответственно значения среднего и стандартного отклонения, вычисленные для тестируемой выборки.

Для выборки из 100 случайных значений погрешности квазиобращения для операций /,* (среднее = = 2,73437499999874e-06; стандартное отклонение = 7,06716286262362e-05).

Значения были упорядочены по возрастанию и разбиты на 10 интервалов с помощью интеграла функции Гаусса (табл. 1).

Таблица 1

Теоретическое количество точек, которые должны попасть в данный интервал, для распределения погрешности квазиобращения для операций /,*

№ интервала	Левая граница интервала	Правая граница интервала	Экспериментальное число точек	Теоретическое число точек
1	– ∞	–9,55078124999995e–05	10	8,38
2	–9,55078124999995e–05	–6,87500000000029e–05	10	7,49
3	–6,87500000000029e–05	–4,37500000000003e–05	10	9,91
4	–4,37500000000003e–05	–2,08984374999998e–05	10	11,67
5	–2,08984374999998e–05	1,75781249999938e–06	10	12,55
6	1,75781249999938e–06	2,65624999999971e–05	10	13,31
7	2,65624999999971e–05	5,15624999999999e–05	10	12,18
8	5,15624999999999e–05	7,61718749999994e–05	10	9,59
9	7,61718749999994e–05	9,90234374999991e–05	10	6,23
10	9,90234374999991e–05	+ ∞	10	8,69

На основании экспериментальных данных построена следующая гистограмма (рис. 2).

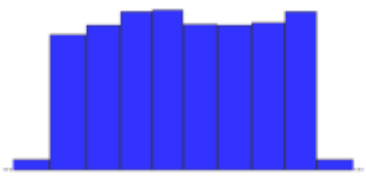

Рис. 2. Гистограмма для распределения погрешности квазиобращения для операций /,*.

Тестовая статистика: 5,62193574497136; она меньше табличного значения χ^2 распределения(hi(p = 0,95, f = 7) = 14.1), следовательно, данные **подчиняются** нормальному закону распределения (рис. 3).

В данном случае функциональное преобразование является простым и допускает анализ, позволяющий вычислить ключи, минимизируя перебор чисел. Но оно наглядно иллюстрирует разрабатываемую идею.

Более сложные функциональные преобразования, как будет показано далее, устраняют повторяемость штампов, симметрию и даже

обеспечивают невозможность обратного преобразования в силу того, что погрешность вычисления превышает результат.

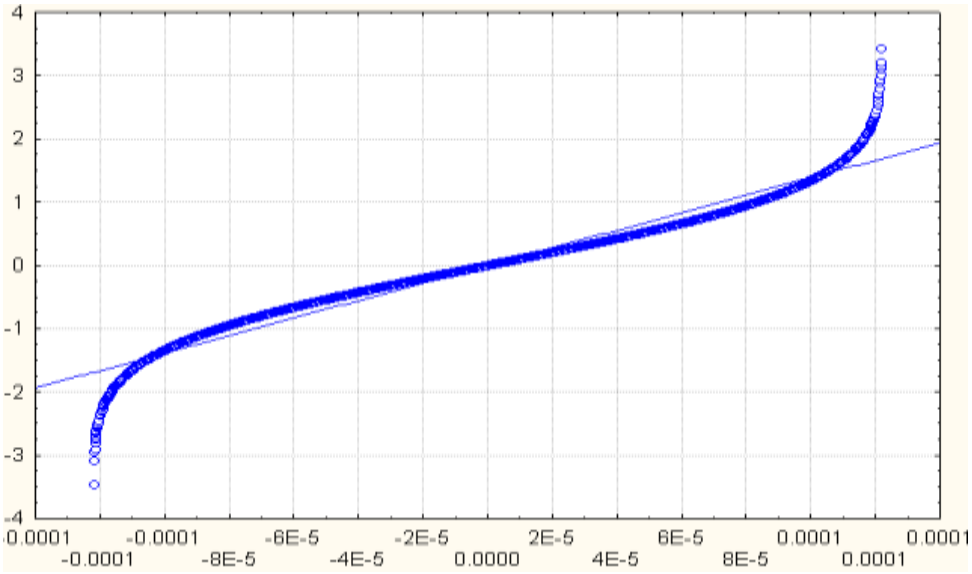

Рис. 3. График подтверждения нормального распределения для распределения погрешности квазиобращения для операций /,* в интервале от 0 до 10000

Рассмотрим экспоненциально-логарифмическое и логарифмически-синусоидальное преобразование.
Зададим ключи $P1 := 1{,}018 \cdot 10^{8}$ $P2 := 1 \cdot 10^{-6}$.
Зададим экспоненциальную функцию

$$y(x) := \exp(x\ P2) \tag{6}$$

и обратную ей функцию логарифма

$$x2(x) := \frac{\ln(y(x))}{P2} - x\,. \tag{7}$$

Более сложную топологию погрешности квазиобращения иллюстрирует рисунок 4.

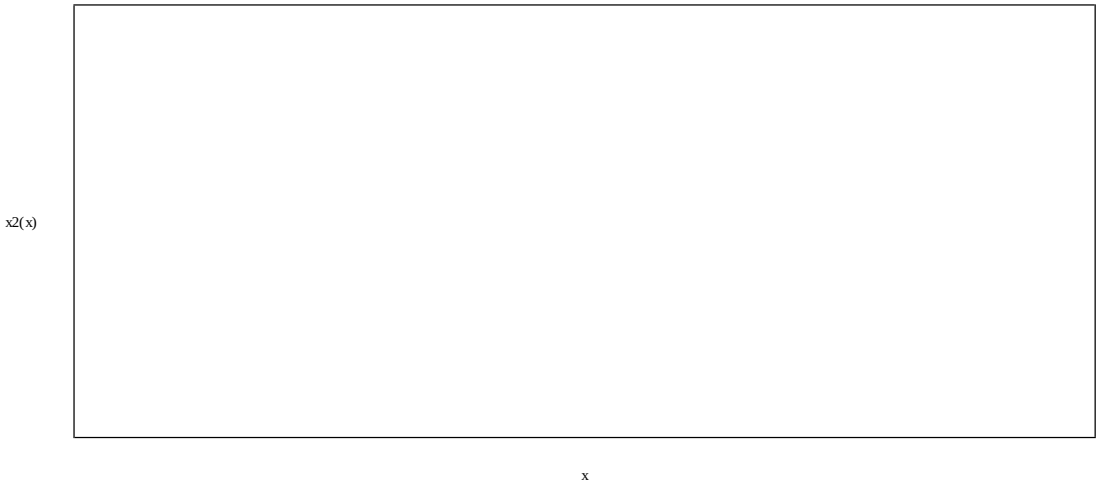

Рис. 4. Распределение погрешности квазиобращения экспоненциально-логарифмической функции на интервале до 30 000

Для проверки выполнения нормального закона распределения по критерию Пирсена (χ^2-критерию). Так же как и в первом случае, в качестве исходных данных возьмем выборку из 100 случайных значений и определим теоретическое количество точек, которые должны попасть в данный интервал (табл. 2). Построим гистограмму распределения случайных значений (рис. 5).

Таблица 2

Теоретическое количество точек, которые должны попасть в данный интервал,
для распределения погрешности квазиобращения экспоненциально-логарифмической функции

№ интервала	Левая граница интервала	Правая граница интервала	Экспериментальное число точек	Теоретическое число точек
1	$-\infty$	–1,01195496425e–10	10	7,78
2	–1,01195496425e–10	–7,6219919265e–11	10	6,91
3	–7,6219919265e–11	–4,4767745065e–11	10	13,41
4	–4,4767745065e–11	–2,9363178555e–11	10	8,22
5	–2,9363178555e–11	–8,541611865e–12	10	12,48
6	–8,541611865e–12	1,8067325415e–11	10	14,88
7	1,8067325415e–11	4,06217282e–11	10	11,81
8	4,06217282e–11	6,225064908e–11	10	9,119999999
9	6,225064908e–11	8,3312023945e–11	10	6,21
10	8,3312023945e–11	$+\infty$	10	9,18

Вычислена тестовая статистика: 8,1096913164748; она меньше табличного значения χ^2 распределения (hi(p = 0,95, f = 7) = 14,1). Вывод – данные для распределения погрешности квазиобращения экспоненциально-логарифмической функции **подчиняются** нормальному закону распределения (рис. 6).

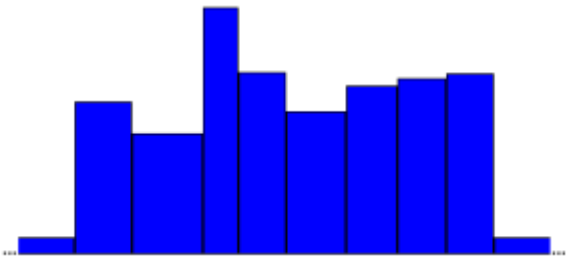

Рис. 5. Гистограмма для распределения погрешности квазиобращения
экспоненциально-логарифмической функции

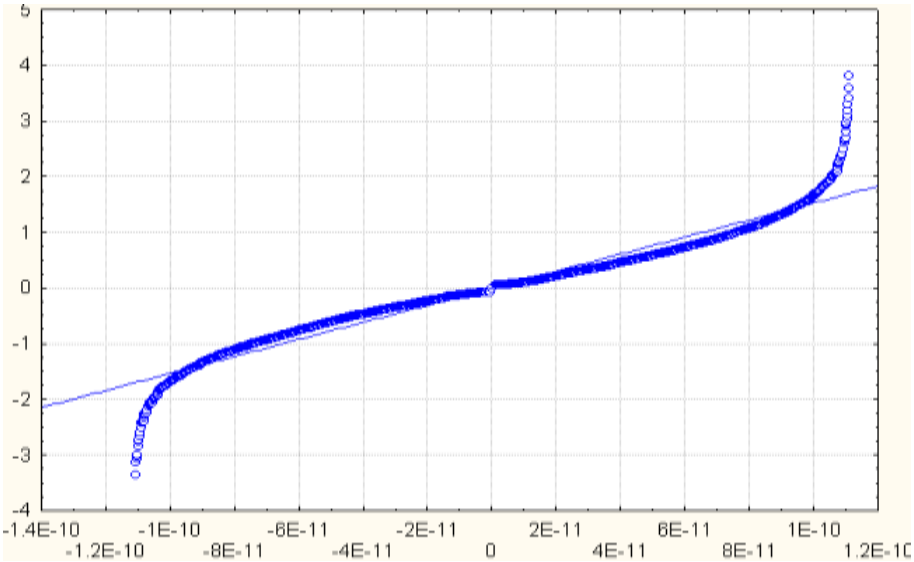

Рис. 6. График подтверждения нормального распределения для распределения погрешности квазиобращения
экспоненциально-логарифмической функции на интервале до 10000

Логарифмически-синусоидальная функция (рис. 7) обладает свойством необратимости:

$$y(x) := \sin(\ln(x)\, P1). \tag{8}$$

Особенностью данного преобразования является то, что функция логарифма нелинейно сжимает пространство с 308 порядка до верхней границы в 709 так, что исходное пространство оказывается разбито на классы эквивалентности, в пределах которых значение логарифма не изменяется. Умножение на значительный параметр $P1$ приводит к тому, что малейшее изменение мантиссы логарифма изменяет синус на величину, равную 10^{-6}. Причем характер распределения результатов вычисления функции не изменяется в зависимости от масштаба анализа:

$$x2(x) := 1 + \frac{P2}{\left(\frac{1}{y(x)}\right)}. \tag{9}$$

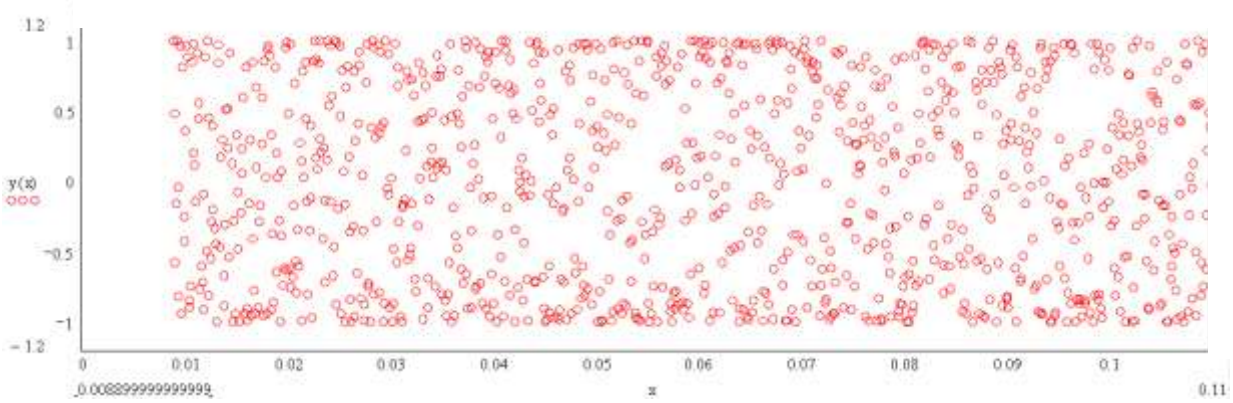

Рис. 7. График логарифмически-синусоидального преобразования для чисел в интервале 0...0,11

Интересной задачей в случае рассмотрения необратимой функции является задача поиска таких x, которые не изменяют число 1, как показано (9), или не изменяют заданный эталонный вектор. Иными словами, такие x, значения необратимой функции от которых лежат в узком коридоре $\pm 10^{-6}$, например (рис. 8). Варьируя ширину коридора, можно обеспечить управление степенью криптозащиты.

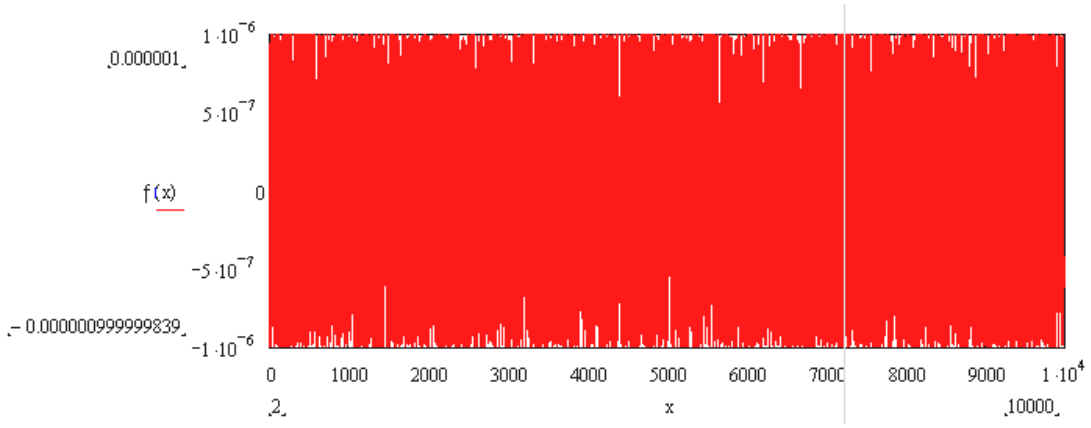

Рис. 8. График значения необратимой функции логарифмически-синусоидального преобразования в интервале от 2...10000

В результате проверки выполнения нормального закона распределения по критерию Пирсена (χ^2-критерию) из 100 случайных значений определили теоретическое количество точек, которые должны попасть в данный интервал (табл. 3), и построили гистограмму распределения случайных (рис. 9).

Вычислена тестовая статистика: 57,3829755140338; она больше табличного значения χ^2 распределения ($\mathrm{hi}(p = 0{,}95;\ f = 7) = 14{,}1$). Вывод – данные для распределения погрешности логарифмически синусоидального преобразования **не подчиняются** нормальному закону распределения.

Но данная случайная величина подчиняется закону равномерного распределения, так как все ее значения лежат внутри определенного малого интервала вещественных чисел, близких к **1** (рис. 8), и внутри этого интервала значения этой случайной величины равновероятны.

Недостатком необратимых функций является ресурсоемкая процедура поиска валидных ключей даже для разработчика защиты.

Таблица 3

Теоретическое количество точек, которые должны попасть в данный интервал, для распределения логарифмически-синусоидального преобразования

№ интервала	Левая граница интервала	Правая граница интервала	Экспериментальное число точек	Теоретическое число точек
1	$-\infty$	0,999999036642937	10	10,93
2	0,999999036642937	0,999999104591577	10	1,78
3	0,999999104591577	0,999999339410361	10	8,48
4	0,999999339410361	0,999999486548965	10	6,57
5	0,999999486548965	0,999999919752455	10	23,04
6	0,999999919752455	1,0000001193169	10	10,99
7	1,0000001193169	1,00000050366871	10	18,44
8	1,00000050366871	1,00000068221659	10	6,42
9	1,00000068221659	1,00000090765923	10	5,709999999
10	1,00000090765923	$+\infty$	10	7,64

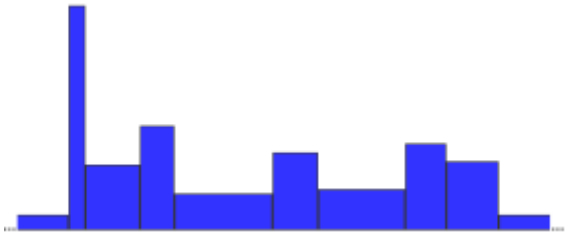

Рис. 9. Гистограмма для распределения погрешности логарифмически-синусоидального преобразования

Рассмотренные примеры позволяют говорить о большом разнообразии вариантов защиты ПО, основанного на квазиобратимых и необратимых функциях представимых вещественных чисел. Нами разработан программный комплекс защиты ПО электронными ключами, предоставляющий разработчику научного ПО полный комплекс услуг: от включения средств защиты в код приложения до создания дистрибутива и сайта технической поддержки лицензионных пользователей защищаемого ПО.

Выводы

В данной работе предложен ряд **новых** положений:

1. Предложен подход к построению систем защиты ПО от несанкционированного использования, основанный на применении методов криптографии, чувствительных к погрешности компьютерных вычислений.
2. Рассмотрены особенности дискретного пространства представимых в машинном формате вещественных чисел двойной точности, функций замкнутых относительно данного пространства, а также свойства обратимости, квазиобратимости и необратимости таких функций.
3. Рассмотрена векторная постановка задачи обратимости для квазиобратимых функций представимых вещественных чисел, обеспечивающая избирательный характер действия криптографического преобразования.
4. Предложены варианты реализации шифрующих, дешифрующих и хеш-функций избирательного действия на основе квазиобратимых и необратимых функций представимых вещественных чисел.
5. Выполнена проверка гипотезы о равномерном и нормальном законе распределения квазиобратимых и необратимых функций.

ЛИТЕРАТУРА

1. Щеглов, А.Ю. Защита компьютерной информации от несанкционированного доступа / А.Ю. Щеглов. – СПб.: Наука и техника, 2004.
2. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования: ГОСТ 28147-89.
3. Ковалевский, В. Криптографические методы / В. Ковалевский, В. Максимов // КомпьютерПресс. – 1993. – № 5. – С. 31 – 34.
4. Водолазский, В. Коммерческие системы шифрования: основные алгоритмы и их реализация. Ч. 1 / В. Водолазский // Монитор. – 1992. – № 6 – 7. – С. 14 – 19.
5. [Электронный ресурс]. – Режим доступа: http://www.x86.org/secrets/Dan0411.html.

6. Каханер, Д. Численные методы и математическое обеспечение: Пер с англ. / Д. Каханер, К. Моулер, С. Нэш. – М.: Мир, 1998. – 575 с.
7. [Электронный ресурс]. – Режим доступа: http://chemstat.com.ru/online/pirsen.html.

Поступила 26.02.2007