Минобрнауки России

Федеральное государственное автономное образовательное учреждение высшего образования   
«Национальный исследовательский университет   
«Московский институт электронной техники»

Факультет Микроприборов и технической кибернетики

Кафедра Высшей математики 1

Сабурова Виктория Игоревна

Бакалаврская работа   
по направлению 01.03.04 «Прикладная математика»

Критерий на открытый текст на базе решающего леса

Студент Сабурова В.И.

Научный руководитель,

к.ф.-м.н. Козлитин И.А.

Москва 2017

# Оглавление

[Оглавление 2](#_Toc478913425)

[Введение 3](#_Toc478913426)

[Глава 1 4](#_Toc478913427)

[Часть 1. Базовые понятия криптографии 4](#_Toc478913428)

[Часть 2. Изучение линейно – рекуррентных последовательностей 7](#_Toc478913429)

[Часть 3. Изучение деревьев решений и рандомизированных лесов 10](#_Toc478913430)

[Глава 2 11](#_Toc478913431)

[Часть 1. Реализация процедуры потокового шифрования с помощью линейно – реккурентных последовательностей 11](#_Toc478913432)

[1.1 Генератор потока ключей 11](#_Toc478913433)

[1.2 Принцип работы ЛРСОС. 13](#_Toc478913434)

[1.3 Критерий Вальда 13](#_Toc478913435)

[Часть 2. Построение простого критерия на открытый текст на основе принципа запрещенных биграмм 15](#_Toc478913436)

[Часть 3. Построение простого критерия на основе рандомизированного леса 16](#_Toc478913437)

[Глава 3 17](#_Toc478913438)

[Часть 1. Исследование эффективности критерия 17](#_Toc478913439)

# Введение

В современном мире необходимость в защите информации возникла задолго до появления информационных технологий. Потребность в сокрытии передаваемой информации и ознакомления с ней лиц, которым данное сообщение не предназначалось, была реализована несколькими подходами, одному из которых и будет посвящена данная работа.   
 Подход, который заключался в преобразовании символов в набор неупорядоченных знаков, где получатель мог интерпретировать полученную информацию в осмысленное сообщение, если имел ключ к построению, позже был назван криптографией.   
 Говоря о криптографии важно упомянуть о том, что она является всего лишь небольшой частью более общей дисциплины – криптологии, науке о шифрах, которую принято разделять на две части – криптографию и криптоанализ, коим будет уделено равное внимание в работе.   
 В данной работе будут рассмотрены две гипотезы: основополагающей из них является гипотеза о том, что рассматриваемый текст, полученный в результате дешифрования, является открытым, то есть биграммы данного текста распределены в соответствии с априори известной гистограммой.

Другая гипотеза, являющаяся альтернативной, подразумевает, что получаемый на вход текст является шифрованным, то есть для него все биграммы равновероятны.   
 Для данных гипотез, соответственно, будут реализованы различные методы шифрования и дешифрования: создание ключа на основе примитивного многочлена, перебор ключей на основе последовательного анализа Вальда, критерий на основе запрещенных биграмм и т.д.

# **Глава 1**

# Часть 1. Базовые понятия криптографии

Прежде чем приступить к описанию данной работы, к конкретно поставленной задаче и ее решения введем некоторые основные понятия и определения, которые здесь используются.

Учитывая, что криптография является довольно широко используемой наукой, мы можем легко дать ей определение: криптография - наука, изучающая построение и использование систем шифрования, в том числе их стойкость, слабости и степень уязвимости относительно различных методов вскрытия, называется криптографией.

Все методы преобразования информации с целью защиты от несанкционированного доступа делятся на две группы: шифрование с закрытым ключом и шифрование с открытым ключом.

*Шифрование с открытым ключом* (ассиметричное шифрование) используется со второй половины XX века. В данную группу относятся методы шифрования, в которых для шифровки и расшифровки данных используются два разных ключа, при котором один из них (открытый ключ) может передаваться по открытому (незащищенному) каналу связи.

*Шифрование с закрытым ключом* (шифрование с секретным ключом или симметричное шифрование) используется довольно давно; в данном случае для шифрования и расшифровывания текста необходимо использовать один и тот же ключ, который обе стороны хранят в секрете от взломщиков. В данной работе представлен метод шифрования с закрытым ключом, так как для шифрования и расшифровывания текста используется один и тот же примитивный многочлен и применяется однотипная операция XOR.

Для того, чтобы лучше понять принципы данного исследования в первую очередь необходимо знать, что такое шифр. *Шифр* - это совокупность заранее оговоренных способов преобразования исходного секретного сообщения с целью его защиты. Как говорилось выше, в рассмотрение берется две гипотезы, в первой из которой используется открытый текст - исходное сообщение, где символом является любой знак, в том числе буква, цифра или знак препинания. В данной работе будет использоваться английский алфавит, состоящий из 26 букв, а также знаки препинания: точка, запятая, восклицательный знак, вопросительный знак, кавычки. Данное дополнение обусловлено тем, что необходима минимальная избыточность работы программы, в связи, с чем была выбрана пяти битная кодировка, позволяющая верно зашифровать все 26 букв алфавита без учета регистра. Чтобы расшифрованный текст состоял из тех же символов, что и зашифрованный, необходимо дополнить исходный алфавит любыми символами, встречаемыми в исходном открытом тексте. Вообще, *алфавитом* в криптографии называется конечное множество используемых для кодирования информации символов, преобразование открытого текста в криптограмму - зашифрованием, а обратное действие - расшифрованием.

Еще одним важным определением в данной области является понятие *ключа* - информации, необходимой для шифрования и расшифрования сообщений. В данной работе система шифрования - любая система, использующаяся для обратимого изменения текста сообщения с целью сделать его непонятным для всех, кроме тех, кому оно предназначено - использует *регистр сдвига с линейно обратной связью (linear feedback shift register, LSFR*) на основе примитивного многочлена тридцать первой степени. Данный выбор обусловлен множественными преимуществами РСЛОС:

* высокое быстродействие создаваемых алгоритмов;
* пользование простейших битовых операций, которые аппаратно были реализованы практически во всех вычислительных устройствах;
* РСЛОС легко анализируются с использованием алгебраических методов;
* широкие исследования в области рекуррентных последовательностей с относительно удовлетворительными криптографическими свойствами;
* стойкие криптографические свойства.

Но наряду с этим, РСЛОС так же имеют недостатки (неэффективная программная реализация, линейность последовательности на выходе регистра позволяет однозначно определить многочлен обратной связи последовательным битам с помощью алгоритма Евклида, относительная легкость анализа алгебраическими методами увеличивает шансы на взлом), которыми в данной работе можно пренебречь.

# Часть 2. Изучение линейно – рекуррентных последовательностей

Прежде чем приступить к изучение рекуррентных последовательностей, как и в первой части данной главы, нам необходимо ввести некоторые определения.

Последовательностью над полем называют любую функцию : , заданную на множестве целых неотрицательных чисел и принимающую значения в поле.

*Последовательность* называют линейной рекуррентной последовательностью порядка надо полем , если существуют константы , такие, что

.

Важным фактом является то, что линейная рекуррентная последовательность реализуется описанной в данной работе схемой линейного регистра сдвига.

Принцип работы РСЛОС заключается в том, что значения, которые содержатся в ячейках накопления, умножаются на заданные коэффициенты и затем суммируются (в поле P), что приводит к следующему результату: происходит сдвиг информации в регистре, в пустую ячейку (крайнюю) записывается значение суммы, которые было вычислено.

Запишем теперь закон рекурсии, выражающий зависимость между знаками линейной рекуррентной последовательности:

.

В данном равенстве многочлен u будет называться *характеристическим многочленом ЛРП* (линейная рекуррентная последовательность), вектор – *начальный вектор ЛРП*. Если данный вектор имеет наименьшую степень, то он называется *минимальным многочленом ЛРП*, а сама степень в данном случае *линейной сложностью ЛРП* *u*, которая и определяет минимальную длину линейного регистра сдвига, реализующего данную последовательность.

Еще одним важным определением является определение периода последовательности: наименьшее натуральное число t, для которого такое, что для всех верно следующее равенство: называется *периодом последовательности*.

Если вернуться к выше описанному закону рекурсии, то можно увидеть, что если *P* является конечным полем из элементов, то тогда максимальное значение периода данной последовательности порядка *n* будет равно . Это очевидно, ведь нулевой начальный вектор создает последовательность из одних нулей, тогда как число различных заполнений регистра длины n равно .

Линейно рекуррентные последовательности, которые имеют максимально возможный период получили название *максимальных ЛРП*.

Чтобы определить значение периода ЛРП необходимо обратить внимание на свойства минимальных многочленов. Таким образом, для того, чтобы ЛРП порядка n над полем из pэлементов имела максимальный период, необходимо и достаточно, чтобы минимальный многочлен являлся *примитивным* – неприводимый многочлен, корень которого в мультипликативной группе поля разложения имеет порядок . Существует достаточное количество критериев проверки неприводимости многочлена над конечными полями и методов их построения, но доказательство примитивности неприводимого многочлена удается в редких случаях.

Задача построения примитивных многочленов не является тривиальной, решение которой даже в исключительных случаях связано со многими трудностями, отчего на практике используют таблицы неприводимых многочленов над кончеными полями.

*Утверждение 1.* Если – неприводимый многочлен над полем *GF(2)* ( обозначение конечного поля из определенного количества элементов) степени *n,* и - простое число, то – примитивный многочлен.

*Утверждение 2.* Неприводимый многочлен примитивен в том и только в том случае, когда для любого просто числа p, делящего , многочлен не сравним с 1 по модулю многочлена .

Для удобного вычисления очередного знака через предыдущие используется представление ЛРП в виде формулы общего члена последовательности, которая являет собой аналитическое выражение члена последовательности в виде функции от его номера.

Рассмотрим следующее предположение. Пусть и *Q –* поле из элементов, являющееся расширением поля *P.* Тогда функцией «след» из поля *Q* в поле *Р* называется отображение вида . В силу свойств конечных полей справедливы равенства:

которые означают, что искомая функция «след» является линейным отображением над полем *P*.

Утверждение 3. Пусть *F(x)* – примитивный многочлен степени n над полем *P* и *Q* – корень *F(x)* в поле *Q*. Тогда любая ненулевая мультиграмма встречается на периоде ЛРП и ровно раз, а число вхождений нулевой мультиграммы на единицу меньше.

Приведенные выше утверждения показывают, что ЛРП над полем позволяют обеспечить первые два из трех требований к псевдослучайным последовательностям, используемым при построении управляющих блоков поточных шифросистем. При выборе закона рекурсии можно гарантировать достаточную величину периода получаемой псевдослучайной последовательности, а также хорошие статистические качества. Действительно, ведь период последовательности совпадает с числом всех ненулевых векторов длиной равной степени минимального многочлена ЛРП, где каждый из них встречается из этих векторов на периоде последовательности ровно один раз.

При это стоит отметить, что аналитическое строение ЛРП довольно простое. Для определения начального вектора по некоторому отрезку последовательности достаточно решить несложную систему линейных уравнений. Поэтому при использовании РСЛОС в криптографических задачах следует учитывать процедуры, заметно повышающие сложность аналитического строения разрабатываемых ими ЛРП.

Часть 3. Изучение деревьев решений и рандомизированных лесов

Глава 2

# Часть 1. Реализация процедуры потокового шифрования с помощью линейно – реккурентных последовательностей

Прежде чем приступить к детальному описанию реализации процедуры шифрования и дешифрования опишем общий алгоритм процесса:

* использование ключевого неприводимого многочлена для формирования линейной рекуррентной последовательности с помощью линейного регистра сдвига с обратной связью;
* генерация шифротекста на основное открытого текст с помощью выведенной последовательности;
* применение операции шифрования к шифротексту и как следствие вывод открытого текста (введенного изначально).

### 1.1 Генератор потока ключей

Потоковые шифры преобразуют открытый текст в шифротекст по одному биту за операцию. На рисунке ниже отображена простейшая реализация потокового шифра (*Рисунок 1*).



Рисунок 1. Процедура шифрования/ дешифрования

Данный поток ключей и поток битов открытого текста подвергаются операции XOR, называемой «исключающее или», результатом которой и будет служить поток битов широфтекста:

*.*

Операция дешифрования аналогична операции шифрования, но выполняется над битами шифротекста, используя тот же самый поток ключей, восстанавливая биты открытого текст:

*.*

Говоря о безопасности данного генератора, можно с уверенностью утверждать, что она полностью зависит от свойств генератора: выдавая поток ключей как бесконечную строку нулей, шифрованный текст полностью будет соответствовать открытому текста. В таком случае вся проведенная операция окажется бессмысленной. Если же генератор будет формировать поток ключей как бесконечный поток случайных битов, на выходе будет получена надежная система шифрования.

Рассмотри потоковый шифр, где генератор потока формирует псевдослучайный битовый поток, который частично или полностью похож на случайный, хотя в действительности детерминирован, и значит, может быть безошибочно воспроизведен при дешифровании. Получается, что чем ближе выход генератора потока ключей случайному, тем потребуется больше времени для взлома шифра, что, в принципе, очевидно. Если же наоборот, то используемая криптосистема будет легко подвержена взлому.

Опишем принцип работы генератора поток ключей. Сам генератор состоит из трех основных частей:

* внутреннее состояние описывает текущее состояние генератора потока ключей. Одинаковые потоки ключей (одинаковый ключ и одинаковое внутреннее состояние) формируют два генератора поток ключей;
* генерация потока ключей происходит путем использования функции выхода по внутреннему состоянию;
* генерация нового внутреннего состояния осуществляется с помощью функции следующего состояния (*Рисунок 2*).



Рисунок 2. Принцип работы генератора ключей

### 1.2 Принцип работы ЛРСОС.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*описываю принцип работы*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### 1.3 Критерий Вальда

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*описание критерия Вальда*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### 1.4 Описание работы алгоритма

Основной задачей в данной части главы является реализация процедуры шифрования и дешифрования с помощью ЛРСОС на основе рекуррентных последовательностей.

Шифрование текста производится с несколько этапов:

1. перевод текста в десятеричную кодировку по таблице ASCII,
2. формирование рекуррентной последовательности на основе выбранного примитивного многочлена для вывода ключа шифрования,
3. сложение векторов десятеричной кодировки символов теста и сформированного ключа; вывод суммы в качестве результата процедуры шифрования.

Для начала выберем примитивный многочлен из таблицы:

Приведенный выше многочлен используется в разрабатываемом алгоритме; в качестве многочлена может быть выбран любой многочлен, являющийся примитивным.

Напомним, что в качестве исходного алфавита используется английский алфавит, где количество букв является равным 26, в связи с чем была выбрана пятибитная кодировка. Поэтому следующим шагом дополним исходный алфавит символами знаков препинания (что наиболее логично в рамках поставленной задачи), чтобы при обратной процедуре символы однозначно декодировались в исходные символы.

...//

(symbol >= 'a' && symbol <= 'z');

symbol -= 96; //1-26;

//...

Далее составим гистограмму частоты употребления знаков и биграмму употребления сочетаний символов в открытом тексте для определения верхнего и нижнего порога по критерию Вальда. Исходя из экспериментальных данных, можем утверждать, что верхним порогом будет являться число \_\_\_\_, а нижним - \_\_\_\_\_. Данные пороги определяют открытость текста, то есть является ли поданный на вход текст открытым или же, наоборот, шифрованным.

# Часть 2. Построение простого критерия на открытый текст на основе принципа запрещенных биграмм

# Часть 3. Построение простого критерия на основе рандомизированного леса

# **Глава 3**

# Часть 1. Исследование эффективности критерия