РЕФЕРАТ

Пояснительная записка к работе содержит 56 страниц, 17 источников, 3 рисунка, 2 приложения, 11 формул.

Главная цель работы состоит в определении криптостойкости систем, основанных на групповой математике, хаотических преобразованиях и ДНК.

Провести анализ криптосистем, построенных на разных базах, определить недостатки и преимущества каждой.

В работе рассмотрены алгоритмы построения криптосистемы на хаосе и ДНК, также проанализированы базовые методы защиты информации, которые используют групповую математику. Реализован алгоритм RSA, и дополнения, которые позволяют применять его для ДНК-криптографии, с использованием нуклеотидов.

НЕАБЕЛЕВЫ ГРУППЫ, ДНК, ХРАНЕНИЕ ДАННЫХ В ДНК, ДНК-КРИПТОГРАФИЯ, КРИПТОГРАФИЧЕСКИЕ СИСТЕМЫ, ХАОТИЧЕСКИЕ КРИПТОСИСТЕМЫ, ВНЕДРЕНИЕ ХАОТИЧЕСКИХ КРИПТОСИСТЕМ, РЕКОМЕНДАЦИИ ПО РАЗРОБОТКЕ ХАОТИЧЕСКИХ КРИПТОСИСТЕМ.

РЕФЕРАТ

Пояснювальна записка до роботи містить 56 сторінок, 17 посилань, 3 рисунка, 2 додатки, 11 формул.

Головна мета роботи полягає у визначенні криптостійкості систем, заснованих на груповій математиці, хаотичних перетвореннях і ДНК. Провести аналіз криптосистем, побудованих на різних базах, визначити недоліки та переваги кожної.

В роботі розглянуті алгоритми побудови криптосистеми на хаосі і ДНК, також проаналізовані базові методи захисту інформації, які використовують групову математику. Реалізовано алгоритм RSA, та доповнення, які дозволяють застосовувати його для ДНК-криптографії, з використанням нуклеотидів.

НЕАБЕЛЕВІ ГРУППЫ, ДНК, ЗБЕРІГАННЯ ДАННИХ У ДНК, ДНК-КРИПТОГРАФІЯ, КРИПТОГРАФІЧНІ СИСТЕМИ, ХАОТИЧНІ КРИПТОСИСТЕМИ, ВПРОВАДЖЕННЯ ХАОТИЧНИХ КРИПТОСИСТЕМ, РЕКОМЕНДАЦІЇЇ ДО РОЗРОБКИ ХАОТИЧНИХ КРИПТОСИСТЕМ.

ABSTRACT

Explanatory note for the work contains 56 pages, 17 sources, 3 figures, 2 applications, 11 formulas.

The main goal of the paper is to determine the cryptographic strength of systems based on group mathematics, chaotic transformations and DNA. Conduct an analysis of cryptosystems built on different bases, determine the disadvantages and advantages of each.

In work were considered algorithms of construction of a cryptosystem on chaos and DNA, basic methods of information protection that use group mathematics were analyzed. Implemented the RSA algorithm, and additions that allow it to be used for DNA cryptography, using nucleotides.

NON-ABELIANS GROUPS, DNA, STORAGE OF DATA IN DNA, DNA CRYPTOGRAPHY, CRYPTOGRAPHIC SYSTEMS, CHAOTIC CRYPTOSYSTEMS, INTRODUCTION OF CHAOTIC CRYPTOSYSTEMS, RECOMMENDATIONS FOR DEVELOPMENT OF CHAOTIC CRYPTOSYSTEMS.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

RSA – аббревиатура от фамилий Rivest, Shamir и Adleman

CSP – задач поиска сопряжений (Conjugacy Search Problem)

ДНК – дезоксирибонуклеи́новая кислота́

DH – Diffie-Hellman

A – аденин

T – тимин

C – цитозин

G – гуанин

Base4 – четверичная система счисления

Base3 – троичная система счисления

ПЦР – полимеразная цепная реакция

PCR – polymerase Chain Reaction

XOR – операция, исключающая “ИЛИ”

DES – data encryption standard

IDEA – international Data Encryption Algorithm

RC5 – Rivest's Cipher 5

RC4 – Rivest's Cipher 4

SEAL – software-optimized Encryption Algorithm

ЦП – центральный процессор

ОС – операционная система

ПБайт – петабайт

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 9](#_Toc484633688)

[1 КРИПТОГРАФИЯ В НЕАБЕЛЕВЫХ ГРУППАХ 11](#_Toc484633689)

[1.1 Основы криптографии с открытым ключом 12](#_Toc484633690)

[1.2 Основы криптографии на свободных группах 18](#_Toc484633691)

[1.3 Обмен публичными ключами с помощью неабелевых групп 25](#_Toc484633692)

[1.4 Трехэтапный протокол Шамира и протоколы обмена ключами 29](#_Toc484633693)

[1.5 Криптография в полициклических группах 31](#_Toc484633694)

[2 ИСПОЛЬЗОВАНИЕ ДНК В КРИПТОГРАФИИ 34](#_Toc484633695)

[2.1 Хранение данных с использование ДНК 34](#_Toc484633696)

[2.2 Процесс кодирования информации 36](#_Toc484633697)

[2.3 ДНК-криптография 38](#_Toc484633698)

[2.3.1 Криптосистема ДНК с использованием замены 39](#_Toc484633699)

[2.4 Преимущество и недостатки в использовании ДНК 41](#_Toc484633700)

[3 КРИПТОГРАФИЯ ОСНОВАННАЯ НА ХАОСЕ 42](#_Toc484633701)

[3.1 Внедрение хаотических систем 42](#_Toc484633702)

[3.2 Внедрение криптосистем 43](#_Toc484633703)

[3.3 Анализ безопасности 45](#_Toc484633704)

[3.4 Базовые правила формирования устойчивой криптосистемы 47](#_Toc484633705)

[ВЫВОДЫ 48](#_Toc484633706)

[ПЕРЕЧЕНЬ ССЫЛОК 49](#_Toc484633707)

[ПРИЛОЖЕНИЕ А ВХОДНЫЕ И ПОЛУЧЕННЫЕ ДАННЫЕ 51](#_Toc484633708)

[ПРИЛОЖЕНИЕ Б ИСХОДНЫЕ КОДЫ ПРОГРАММЫ 52](#_Toc484633709)

ВВЕДЕНИЕ

Благодаря междисциплинарному развитию науки об информации, физической науки и биологической науки в области криптографии появилось много новых технологий. Новые области криптографии в основном состоят из квантовой криптографии, хаотической криптографии, ДНК-криптографии и т. д. Безопасность квантовой криптографии основана на принципе неопределенности Гейзенберга. Квантовая криптография - единственная, которая может реализовать безоговорочную безопасность в настоящее время. Мэттьюс впервые применил теорию хаоса в криптографии и предложил хаотическую схему шифрования потока, основанную на пересмотренной логистической карте. С тех пор хаотическая криптография привлекла широкое внимание, большинство исследований в области хаотической криптографии сосредоточено на криптографии с секретным ключом. Криптография ДНК, которая использует вычисления ДНК, является новой ветвью криптографии в последние годы. Используя высокую плотность хранения и высокий параллелизм молекул ДНК, такая криптография может реализовать шифрование, аутентификацию, подпись и т. д.

Между тем, криптографы с нетерпением ждут применения новых трудноразрешимых математических задач в классической криптографии. В настоящее время большинство публичных криптографических примитивов основаны на высокой стойкости некоторых математических задач в очень больших конечных абелевых группах. Выдающиеся трудные задачи состоят из проблемы факторизации больших целых чисел, задачи дискретного логарифмирования над конечным полем или эллиптической кривой и т. д. Однако благодаря квантовым алгоритмам для факторизации целых чисел и решению задачи дискретного логарифмирования большинство известных криптосистем с открытым ключом будут небезопасными, когда квантовые компьютеры станут практичными. Таким образом, это неизбежная работа по разработке эффективных криптографических схем, которые могут противостоять квантовым атакам. Собственно, с 1980-х годов несколько экспертов пытались разработать новые схемы криптографии, основанные на сложных проблемах теории групп. В 1985 году Вагнер и Мадьярик предложили подход к разработке криптосистем с открытым ключом на основе групп и полугрупп с неразрешимой проблемой слов. В 1986 году, Магливерас предложил симметричную криптосистему, основанную на специальном типе факторизации бесконечных групп, названных логарифмическими сигнатурами для конечных групп перестановок. В 2000 году Ko разработал теорию криптографии на основе кос, основанную на жесткости задачи поиска сопряжения (CSP) в группах кос. В 2004 году Эйк и Кахробаи предложили новую криптосистему на основе полициклических групп. В 2005 году Шпильрайн и Ушаков предположили, что группа Томпсона может быть хорошей платформой для построения криптосистем с открытым ключом. Между тем, активная ветвь некоммутативной криптографии, основанная на жесткости проблемы групповой факторизации, добилась больших успехов в течение последних двух десятилетий.

# 

ИСПОЛЬЗОВАНИЕ ДНК В КРИПТОГРАФИИ

Размеры «цифровой вселенной» должны были превысить 16 зеттабайт к 2017 году. Значительная доля этих данных хранится в виде архивов. К примеру, компания Facebook недавно построила отдельный дата-центр для «холодного» хранения 1 эксабайта данных. Такое же количество информации способно уместиться в 1 мм3 ДНК. Сохранение данных в ДНК проходит в три этапа: преобразование цифровых данных в последовательность нуклеотидов ДНК, синтез молекул ДНК и, непосредственно, хранение данных. Чтобы данные считать, необходимо выделить требуемую последовательность из молекулы ДНК и преобразовать её в первоначальный вид. Время на проведение синтеза и секвенирования уменьшается экспоненциально как показано на рисунке 2.1 [8].

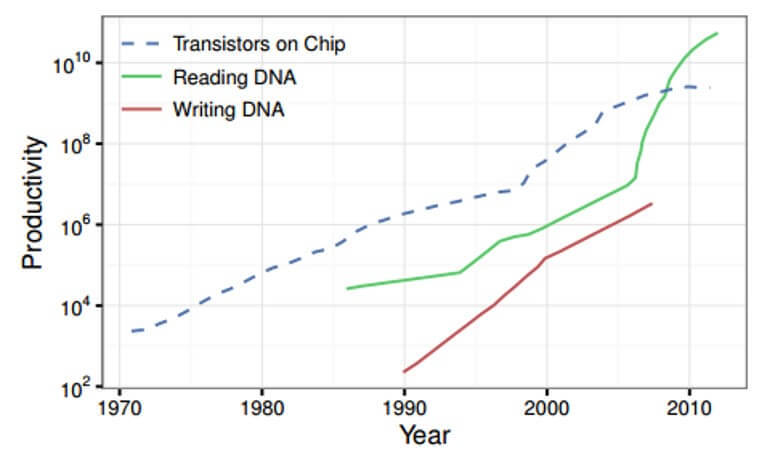


Рисунок 2.1 – Рост продуктивности использования ДНК

## 2.1 Хранение данных с использование ДНК

Молекула ДНК хранит информацию в четверичной системе счисления, по количеству нуклеотидов. Это компактный контейнер с плотностью записи в тысячи раз больше, чем у существующих носителей. Однако, чтобы технология перешла от научных испытаний к коммерческому использованию, требуется решить ряд проблем. Одна из них — специфика цифровой информации, в которой одни и те же биты могут многократно повторяться. Если многократно повторять один и тот же нуклеотид в молекуле ДНК, то это негативно влияет на стабильность кластера и информация может быть потеряна, даже при использовании избыточного дублирования и коррекции ошибок.  
 Исследователи из Европейского института биоинформатики  опубликовали работу с описанием способа, как можно существенно повысить стабильность ДНК. Попросту, они предлагают отказаться от четверичной системы (Base-4) в пользу троичной (Base-3), а четвёртый нуклеотид использовать в служебных целях для разбиения длинных цепочек согласно рисунка 2.2 [9].

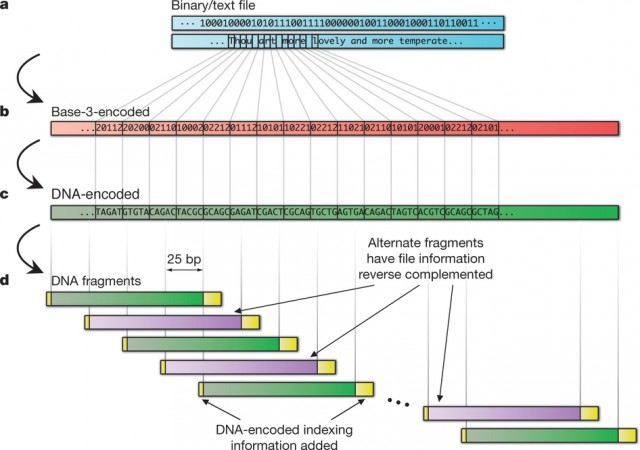
. 

Рисунок 2.2 – Процесс записи данных в ДНК

Во время эксперимента исследователи записали в ДНК почти мегабайт информации, в том числе все 154 сонета Шекспира в формате .txt, видеоролик с записью выступления Мартина Лютера Кинга продолжительностью 26 секунд, обложку журнала Bioinformatics Institute в формате .jpeg, научную работу с описанием структуры ДНК в формате .pdf, а также ещё один файл с описанием процесса кодирования. В общей сложности всё уместилось в 739 килобайт.  
При переходе с Base-4 на Base-3 мы теряем 25% информационной ёмкости, но даже в таком варианте учёные сообщают об информационной плотности записи 2,2 петабайта на 1 грамм биологического материала. Эксперимент показал надёжность считывания информации 100%. Теоретически, эта схема способна масштабироваться в пределах, превышающих объёмы всей существующей цифровой информации, пишут авторы исследования.

Исходя из нынешнего технологического прогресса в области синтеза и секвенирования, носители ДНК для записи информации должны появиться в открытой продаже в течение десяти лет. Хотя ДНК позволяет хранить информацию тысячелетиями, первые коммерческие носители будут продаваться с гарантией до 50-ти лет, считают исследователи.

## 2.2 Процесс кодирования информации

ДНК содержит четыре типа нуклеотидов: аденин (A), цитозин (С), гуанин (G) и тимин (T). Нить ДНК представляет собой линейную последовательность этих нуклеотидов. Таким образом, у нас есть четыре кодовых   
символа (A, C, G и T), поэтому очевидным подходом к хранению двоичных данных будет их кодирование в четверичной системе счисления, например, 0=A, 1=C, 2=G, и 3=T, на основе этих данных, в приложении А, представлен реализованный алгоритм шифрования основанный на RSA с использованием нуклеотидов. Однако стоит учитывать, что синтез и секвенирование подвержены ошибкам. Вероятность ошибок можно снизить, если закодировать двоичную информацию не в четверичной, а в троичной   
системе счисления, как показано на рисунке ниже. Чтобы избежать неэффективного преобразования исходных двоичных данных в троичную систему счисления, используется код Хаффмана. Пример такого преобразования показан на рисунке 2.3.

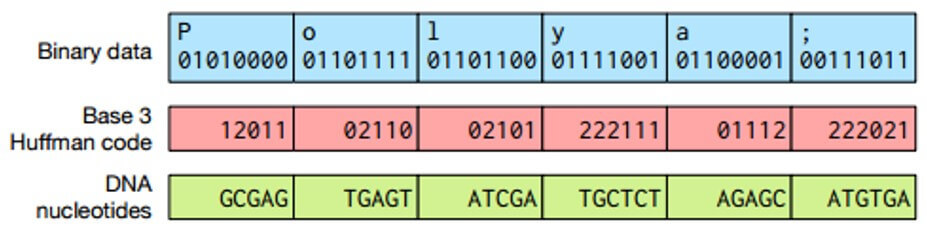


Рисунок 2.3 – Процесс преобразование слова в нуклеотиды ДНК

Каждая из трех цифр соотносится с нуклеотидом ДНК в соответствии с рисунком 2.4, причем нуклеотиды в цепочке не повторяются, что приводит к снижению ошибки секвенирования.

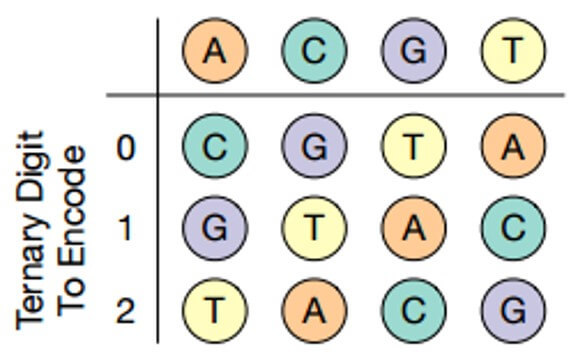


Рисунок 2.4 – Кодирования нуклеотидов

Чтобы обеспечить возможность произвольного доступа к данным, ученые организовали перевод ключей в уникальные последовательности праймеров. Праймеры – это короткие синтетические нити, определяющие начало и конец области, которую необходимо амплифицировать. Праймеры обеспечивают произвольный доступ с помощью полимеразной цепной реакции, которая генерирует множество копий ДНК. Цепи конкретного объекта имеют общий праймер, а разные цепи с одним и тем же праймером различаются по адресам.  «Контролируя последовательности, которые используются как праймеры для полимеразной цепной реакции (ПЦР), мы можем указать, какие нити в решении будут проходить амплификацию. Для того чтобы считать значение ключа в решении, мы просто проводим ПЦР, используя соответствующий этому ключу праймер», – говорят ученые.

## 2.3 ДНК-криптография

ДНК криптография – это новое поле, основанное на исследованиях в ДНК вычислений и новых технологиях, таких как PCR (Polymerase Chain Reaction),. ДНК-вычисления обладают высоким уровнем вычислительной способности и способны хранить огромное количество данные. Грамм ДНК содержит 1021 основание ДНК, что эквивалентно 108 терабайтам данных. В криптографии ДНК мы используем существующую биологическую информацию из общедоступных баз данных ДНК для кодирования открытого текста. Криптографический процесс может использовать разные методы. В методах одноразовых блокнотов описаны наиболее эффективные алгоритмы безопасности. В случае одноразового блокнота, открытый текст комбинируется с секретным случайным ключом или нашим блокнотом, который используется только один раз. Блокнот сочетается с открытым текстом, используя типичное модульное сложение или операцию XOR. Скорость алгоритма должна быть довольно высокой. Биологический фон: ДНК является аббревиатурой для дезоксирибонуклеиновой кислоты, которая является зародышевой плазмой всех стилей жизни. ДНК является своего рода биологической макромолекулой и состоит из нуклеотидов. Как уже выше упоминалось, каждый нуклеотид содержит одно основание и существует четыре типа оснований: аденин (А) и тимин (Т) или цитозин (С) и гуанин (G), соответствующий четырем типам нуклеотидов. Одноцепочная ДНК построена с ориентацией: один конец называется 5 ', а другой конец называется 3'. Обычно ДНК существует как двухцепочные молекулы в природе. Две взаимодополняющие нити ДНК удерживаются вместе, образуя структуру двойной спирали. Структура двойной спирали была обнаружена Уотсоном и Криком; Таким образом, комплементарная структура называется комплементарностью Уотсона Крика. Их открытие является одним из величайших научных открытий 20-го века. Развитие ДНК-криптографии, выигрывает от развития ДНК-вычислений (также называемых молекулярными вычислениями или биологическими вычислениями). С одной стороны, криптография всегда имеет некоторые отношения с соответствующей вычислительной моделью более или менее. С другой стороны, некоторые биологические технологии используются в ДНК-криптографии.

### 2.3.1 Криптосистема ДНК с использованием замены

В одноразовой системе замещения используется группа двоичного сообщения открытого текста и группа таблиц, определяющая случайное отображение на зашифрованный текст. Входная цепочка имеет длину и разбивается на незашифрованные слова фиксированной длины. Таблица отображает все возможные строки незашифрованного текста с фиксированной длиной в соответствующие строки зашифрованного текста, так что существует уникальное обратное отображение.

Шифрование происходит путем замены каждого слова открытого текста ДНК соответствующим шифрованным словом ДНК. Отображение реализуется с использованием длинного ДНК-блокнота, состоящего из множества сегментов, каждый из которых указывает на одно слово с открытого текста для шифрования словосочетаний. Слово с открытым текстом действует связывания праймера, который затем удлиняется. Это приводит к формированию текстовой пары открытого текста и зашифрованного текста.

Идеальная одноразовая библиотека будет содержать огромное количество блокнотов, и каждый из них обеспечит совершенно уникальное, случайное отображение из незащищенных слов в зашифрованные слова.

Повторяющийся блок состоит из: одного слова последовательности , из набора слов, соответствующих шифру или кодовой книге и префикса . Приметим, что включает в себя уникальную подпоследовательность, которая предотвращает атаки частотного анализа путем сопоставления нескольких экземпляров одного и того же открытого текста сообщения с различными зашифрованными словами. Кроме того, этот префикс может быть необязательно использован для кодирования положение слова в сообщении.

Каждая пара последовательностей однозначно связывает текстовое слово с шифровым словом.

Последовательность стопоров запрещает расширение растущей цепи ДНК за границу парного шифрованного слова. Используя эту тему, создается библиотека уникальных цепочек кодовых книг. Каждая отдельная цепочка из этой библиотеки кодовой книги задает конкретный уникальный набор парных слов[10].

Одноразовый блокнот состоит из цепочки ДНК длиной , содержащей копииповторяющегося шаблона: шифрованное слово длиной , слово открытого текста длиной и стопорной последовательности длины . Заметим, что длина слова растет логарифмически по всей длине блокнота. Определенно ; и = , где фиксированные целочисленные константы и . Каждый повторитель задает единую пару сопоставления, и ни одно слово из кодовой книги или текстовое слово не будет использоваться более одного раза в любом блокноте. Поэтому, учитывая шифрованное слово , мы уверены, что оно отображает только одно слово с открытым текстом и наоборот. Последовательность стоппера действует как «пунктуация» между повторяющимися звеньями, так что ДНК-полимераза не сможет продолжать копирование нити матрицы (блокнота). Стоппер – последовательности состоят из последовательности идентичных нуклеотидов, которые действуют для прекращения копирования цепи ДНК-полимеразой, учитывая отсутствие комплементарного нуклеотидтрифосфата в пробирке. Например, последовательность TTTT будет действовать в качестве точки остановки, если полимеризационной смеси не хватает своего комплементарного основания.

Экспериментальная осуществимость зависит от следующих факторов: размер лексикона, который представляет собой число пар слов открытого текста-зашифрованного текста, размер каждого слова, количество одноразовых ДНК, блокноты, которые могут быть созданы в цикле синтеза, и длина каждого сообщения, которое должно быть зашифровано. Если бы используемый лексикон состоял из слов английского языка, его размер находился бы в диапазоне от 10 000 до 25 000 словных пар. Если по экспериментальным причинам требуется меньший лексикон, тогда используемые слова могут представлять собой более малый набор, такой как ASCII-символы, что приводит к размеру словаря 128.

## 2.4 Преимущество и недостатки в использовании ДНК

В будущем такие системы потенциально позволят сохранить огромное количество данных на микроскопических носителях. Представьте себе «флешку» объемом 100 мм3, способную хранить порядка 100 000 ПБайт данных. Однако пока что самым крупным препятствием к внедрению подобных технологий остается время. Расшифровка и чтение молекулы ДНК занимает многие часы. Поэтому такой тип хранилищ вряд ли подойдет для содержания часто используемых данных, однако способен перевернуть наше представление о долговременных хранилищах в дата-центрах. Также на сегодняшний день стоимость кодирования информации в ДНК оценивается примерно в [$12400 за мегабайт](http://mashable.com/2013/01/23/dna-replace-hard-drive/), стоимость считывания — $220 за 1 МБ, что является явным недостатком. Еще одним минусом является то, что необходимы специальные емкости для хранения ДНК, во избежание ее распада и потери данных.

# ВЫВОДЫ

В данной работе были проанализированы криптосистемы, основанные на неабелевых группах, были рассмотрены алгоритмы обмена ключами реализованные на модулярной арифметике, так же эти алгоритмы были представлены на абелевых группах, что дает большую стойкость, за счет сложности нахождения сопряжений между группами и подгруппами.

Анализ ДНК-криптографии показал, что использование ее на данном этапе развития не целесообразно за счет потребление больших вычислительных средств и дороговизны всей процедуры, также большим минусом является то, что процесс записи и извлечения занимает большое время, примерно в 5 раз больше, чем с использованием алгоритмах в групповой математике. Положительной стороной есть то, что технологии развиваются чрезвычайно быстро и через несколько лет данный вид защиты информации может быть реализован без больших финансовых и вычислительных трудностей.

Криптосистемы, основанные на хаосе, являются чрезвычайно чувствительными к выбору параметров, и для их создания был сформирован ряд правил, которые не являются необходимыми, но есть желательными. Для построения успешной такой криптосистемы были определены два свойства: путаницу и различие. Путаница используется для усложнения понимания алгоритма связи ключа с шифротекстом, различие – должно минимизировать количество шаблонов в самой криптосистеме.

# ПЕРЕЧЕНЬ ССЫЛОК

1. A.G. Myasnikov, V.Shpilrain and A. Ushakov, Group-Based Cryptography Advanced Courses in Mathematics, CRM Barcelona, 415, 2007.
2. P.Dehornoy, Braid-Based Cryptography Cont. Math., 360, 2004, 5–34.
3. W. Magnus, A. Karass and D. Solitar Combinatorial Group Theory, Wiley Interscience,New York, 527, 1968.
4. G. Baumslag, B.Fine, and X.Xu, Cryptosystems Using Linear Groups Appl. Alg. in Engineering, Communication and Computing 17, 290, 2006, 205-217.
5. K.Ko, J.Lee, J.H. Cheon, J.W. Han, J.Kang, C.Park, New Public-Key Cryptosystem Using Braid Groups , Advances in Cryptology - CRYPTO 2000 Santa Barbara CA , Lecture Notes in Computer Science, Springer 1880, 2000, 166-183.
6. D.Kahrobaei and B.Khan, A Non-Commutative Generalization of the ElGamal Key Exchange using Polycyclic Groups Proceeding of IEEE, GLOBECOM, 2006.
7. V.Shpilrain and A. Ushakov, The Conjugacy Search Problem in Public Key Cryptography; Unnecessary and Insufficient Applicable Algebra in Engineering, Communication and computing, 17, 2006 285-289.
8. «Заложено природой»: Система хранения данных на основе ДНК [Электронный ресурс] // URL: <https://m.habrahabr.ru/company/1cloud/blog/306656/> (дата обращения: 11.05.2017).
9. Надёжное хранение информации в ДНК (2,2 петабайта на грамм) [Электронный ресурс] // URL: <https://geektimes.ru/post/166889/> (дата обращения: 21.05.2017).
10. DNA Data cryptography [Электронный ресурс] // URL: <https://www.slideshare.net/mayukhmaitra/dna-cryptography> (дата обращения: 25.05.2017).
11. Alvarez, G. & Li, S. “Breaking cryptography with chaos at the physical level,” arXiv:nlin.CD/0403029, 215, 2004.
12. Cerm´ak, J. “Digital generators of chaos,” Phys. Lett. 525, 1996, 151–160.
13. Stinson, D. R. [1995] Cryptography: Theory and Practice (CRC Press).
14. Schneier, B. [1996] Applied Cryptography – Protocols, algorithms, and souce code in C second edn. (John Wiley & Sons, Inc., New York, USA).
15. Menezes, A. J., van Oorschot, P. C. & Vanstone, S. A. Handbook of Applied Cryptography (CRC Press), 1997.
16. Fog, A. [2000] “How to optimize for the Pentium family of microprocessors,” Online document, <http://www.codingnow.com/2000/download/pentopt.htm>.
17. Fridrich, J. “Symmetric ciphers based on two-dimensional chaotic maps,” Int. J. Bifurc. Chaos 8, 1465, 1998, 1259–1284.

ДОДАТОК А

ДОДАТОК Б