Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный технологический университет»

Факультет информационных технологий

Кафедра информационных систем и технологий

**Отчет к лабораторной работе**:

«Исследование криптографических шифров на основе подстановки (замены) символов»

Выполнил:

студент 3 курса 4 группы

специальности ПОИТ

Карленок Ю.А.

Минск 2020

1. **Теоретические сведения**

Сущность подстановочного шифрования состоит в том, что, исходный текст (из множества М) и зашифрованный текст (из множества С) основаны на использовании одного и того же или разных алфавитов, а тайной или ключевой информацией является алгоритм подстановки.

Если исходить из того, что используемые алфавиты являются конечными множествами, то в общем случае каждой букве ax алфавита AM (ax∈ AM) для создания сообщения Мi (Мi ∈ M) соответствует буква ay или множество букв{АхС} для создания шифртекста Сi (Сi ∈ С). Важно, чтобы во втором случае любые два множества (например, {АхС}b и {АхС}n, b≠n, 1 ≤ b, n, x, y ≤ N, N – мощность алфавита), используемые для замены разных букв открытого текста, не пересекались:

{АхС}b ∩ {АхС}n=0.

Если в сообщении Мi содержится несколько букв ax, то каждая из них заменяется на символ ay либо на любой из символов {АхС}. За счет этого с помощью одного ключа можно сгенерировать различные Сi для одного и того же Мi. Так как множества {АхС}b и {АхС}n попарно не пересекаются, то по каждому символу Сi можно однозначно определить, какому множеству он принадлежит, и, следовательно, какую букву открытого сообщения Мi он заменяет. В силу этого открытое сообщение восстанавливается из зашифрованного однозначно.

Приведенные утверждения справедливы для следующих типов подстановочных шифров:

* моноалфавитных (шифры однозначной замены или простые подстановочные),
* полиграммных,
* омофонических (однозвучные шифры или шифры многозначной замены),
* полиалфавитных.

**Моноалфавитные шифры подстановки**

В данных шифрах операция замена производится только над каждым одиночным символом сообщения Мi. Для наглядной демонстрации шифра простой замены достаточно выписать под заданным алфавитом тот же алфавит, но в другом порядке или, например, со смещением. Записанный таким образом алфавит называют алфавитом замены.

Максимальное количество ключей для любого шифра этого вида не превышает N!, где N – количество символов в алфавите.

Для математического описания криптографического преобразования предполагаем, что зашифрованная буква ay (ay ∈ Сi), соответствующая символу aх (aх∈ Мi), находится на позиции

y ≡ x + k (mod N),

где x, y – индекс (порядковый номер, начиная с 0) символа в используемом алфавите, k – ключ.

Для расшифрования сообщения Сi необходимо произвести расчеты обратные, т. е.:

х ≡ у – k (mod N).

Одним из существенных недостатков моноалфавитных шифров является их низкая криптостойкость. Зачастую метод криптоанализа базируется на частоте встречаемости букв исходного текста.

Система шифрования Цезаря с ключевым словом (лозунгом) также является одноалфавитной системой подстановки. Особенностью этой системы является использование ключевого слова (лозунга) для смещения и изменения порядка символов в алфавите подстановки (желательно, чтобы все буквы ключевого слова были различными). Ключевое слово пишется в начале алфавита подстановки.

Применяя одновременно операции сложения и умножения по модулю n над элементами множества (индексами букв алфавита), можно получить систему подстановок, которую называют аффинной системой подстановок Цезаря. Определим процедуру зашифрования в такой системе:

y ≡ ax + b (mod N),

где a и b – целые числа.

При этом взаимно однозначные соответствия между открытым текстом и шифртекстом будут иметь место только при выполнении следующих условий: 0 ≤ a, b< N, наибольший общий делитель (НОД) чисел a, b равен 1, т.е. эти числа являются взаимно простыми.

**Полиграммные шифры**

В таких шифрах одна подстановка соответствует сразу нескольким символам исходного текста.

Первым известным шифром этого типа является шифр Порты [9]. Шифр представляется в виде таблицы. Наверху горизонтально и слева вертикально записывался стандартный алфавит. В ячейках таблицы записываются числа в определенном порядке.

Шифрование выполняется парами букв исходного сообщения. Первая буква пары указывает на строку, вторая – на столбец. В случае нечетного количества букв в сообщении Мi к нему добавляется вспомогательный символ, например, «А».

С точки зрения криптостойкости рассматриваемый тип шифров имеет преимущества перед моноалфавитными шифрами. Это связано с тем, что распределение частот групп букв значительно более равномерное, чем отдельных символов. Во-вторых, для эффективного частотного анализа требуется больший размер зашифрованного текста, так как число различных групп букв значительно больше, чем мощность алфавита.

**Омофонические шифры**

Омофонические шифры (омофоническая замена) или однозвучные шифры подстановки создавались с целью увеличить сложность частотного анализа шифртекстов путем маскировки реальных частот появления символов текста с помощью омофонии.

При шифровании символ исходного сообщения заменяется на любую подстановку из «своего» столбца. Если символ встречается повторно, то, как правило, используют разные подстановки. Например, исходное сообщение «АБРАМОВ» после зашифрования может выглядеть так: «357 990 374 678 037 828 175».

Книжный шифр. Заметным вкладом греческого ученого Энея Тактика в криптографию является предложенный им так называемый книжный шифр. После Первой мировой войны книжный шифр приобрел иной вид. Шифрозамена для каждой буквы определялась набором цифр, которые указывали на номер страницы, строки и позиции в строке (вспомните пример использования такого шифра известными героями фильма «17 мгновений весны»). Даже с формальной стороны отсутствие полной электронной базы изданных к настоящему времени книг делает процедуру взлома шифра практически не выполнимой.

**Полиалфавитные шифры**

Полиалфавитные (или многоалфавитные) шифры состоят из нескольких шифров однозначной замены. Выбор варианта алфавита для зашифрования одного символа зависит от особенностей метода шифрования.

Диск Альберти. В «Трактате о шифрах» [10] Альберти приводит первое точное описание многоалфавитного шифра на основе шифровального диска.

Он состоял из двух дисков – внешнего неподвижного и внутреннего подвижного дисков, на которые были нанесены буквы алфавита. Процесс шифрования заключался в нахождении буквы открытого текста на внешнем диске и замене ее на букву с внутреннего диска, стоящую под ней. После этого внутренний диск сдвигался на одну позицию и шифрование второй буквы производилось уже по-новому шифралфавиту. Ключом данного шифра являлся порядок расположения букв на дисках и начальное положение внутреннего диска относительно внешнего.

Таблица Трисемуса. В 1518 году в развитии криптографии был сделан важный шаг благодаря появлению в Германии первой печатной книги по криптографии. Аббат Иоганнес Трисемус, настоятель монастыря в Вюрцбурге, написал книгу «Полиграфия», в которой он описал ряда шифров, один из которых развивает идею многоалфавитной подстановки. Зашифрование осуществляется так: заготавливается таблица подстановки (так называемая «таблица Трисемуса» – таблица со стороной равной N, где N – мощность алфавита), где первая строка – это алфавит, вторая – алфавит, сдвинутый на один символ, и т. д. При зашифровании первая буква открытого текста заменяется на букву, стоящую в первой строке, вторая – на букву, стоящую во второй строке, и т.д. После использования последней строки вновь возвращаются к первой.

Таким образом, ключом в таблицах Трисемуса является ключевое слово и размер таблицы. При шифровании буква открытого текста заменяется буквой, расположенной ниже нее в том же столбце. Если буква текста оказывается в нижней строке таблицы, тогда для шифртекста берут самую верхнюю букву из того же столбца.

Шифр Виженера. В 1586 г. французский дипломат Блез Виженер представил перед комиссией Генриха III описание простого, но довольно стойкого шифра, в основе которого лежит таблица Трисемуса.

В этом шифре мы имеем дело с последовательностью сдвигов, циклически повторяющейся. Основная идея заключается в следующем. Создается таблица (таблица Виженера) размером N·N (N – число знаков в используемом алфавите). Эти знаки могут включать не только буквы, но и, например, пробел или иные знаки. В первой строке таблицы записывается весь используемый алфавит. Каждая последующая строка получается из предыдущего циклического сдвига последней на 1 символ влево. Таким образом, при мощности алфавита (английского языка) равной 26, необходимо выполнить последовательно 25 сдвигов для формирования всей таблицы.

Роторные машины. Идеи Альберти и Виженера использовались при создании электромеханических роторных машин первой половины ХХ века. Некоторые из них использовались в разных странах вплоть до 1980-х годов. В большинстве из них использовались роторы (механические колеса), взаимное расположение которых определяло текущий алфавит шифрозамен, используемый для выполнения подстановки. Наиболее известной из роторных машин является немецкая машина времен Второй мировой войны «Энигма». Более детальному изучению и практическому анализу «Энигмы» далее будет посвящена отдельная лабораторная работа.

К полиалфавитным относится также шифр на основе «одноразового блокнота».

1. **Практическая часть**

В данной лабораторной работе необходимо разработать пользовательское приложение, которое должно реализовывать следующие операции:

* выполнять зашифрование/расшифрование текстовых документов (объемом не менее 5 тысяч знаков) созданных на основе алфавита языка в соответствии с нижеследующей таблицей вариантов задания; при этом следует использовать шифры подстановки:

1. На основе аффинной системы подстановок Цезаря; a=3, b=7
2. Шифр Порты

* формировать гистограммы частот появления символов для исходного и зашифрованного сообщений;
* оценивать время выполнения операций зашифрования/расшифрования.
* алфавит – полький.

Для выполнения первого задания была написана функция, которая позволяет зашифровать сообщение шифром аффинной системы подстановок Цезаря с использованием коэффициентов a и b. Разберем работу шифра на примере 1 буквы сообщения - «H».

В польском алфавите [aąbcćdeęfghijklłmnńoópqrsśtuvwxyzźż] символ «H» занимает 10 место (нумерация с нуля). Обозначим это как x = 10.

Введенные коэффициенты – a = 3, b = 7.

Тогда y = (ax + b) mod n,

где n – количество символов в алфавите (в нашем случае - 35).

y = (3 \* 10 + 7) mod 35 = 2.

Т.е. зашифрованный символ – это символ, стоящий в позиции 2 – «B».

Все остальные символы шифруем тем же способом:

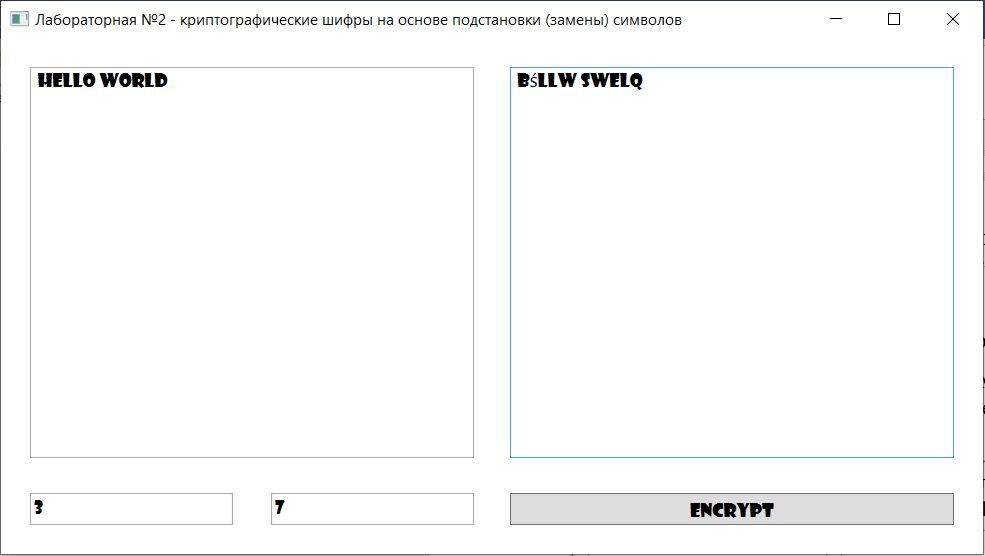


Рисунок 2.1 – Аффинная система подстановок Цезаря - шифрование

Гистограмма встречаемости символов приведена на рисунке 2.2:

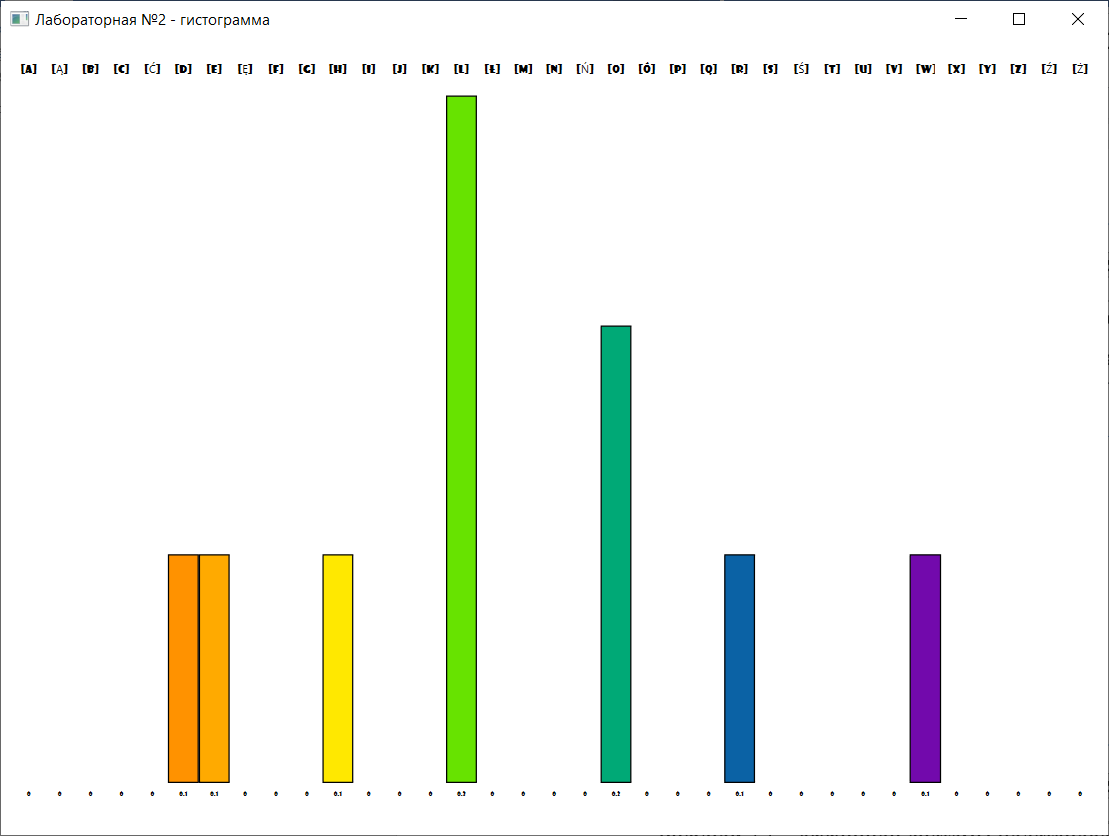


Рисунок 2.2 – Гистограмма частоты появления символа

Также нужно было разработать функцию, которая бы дешифровала текст. Для наглядности будем использовать шифротекст из предыдущего примера, а также соответствующие коэффициенты.

В польском алфавите [aąbcćdeęfghijklłmnńoópqrsśtuvwxyzźż] символ «B» занимает 2 место (нумерация с нуля). Обозначим это как y = 2.

Введенные коэффициенты – a = 3, b = 7.

Тогда x = (a-1 (y + n – b)) mod n,

где a-1 – обратное к a число по модулю n, т. е. оно удовлетворяет уравнению а \* a-1 = 1 mod n, а n – количество символов в алфавите (в нашем случае - 35).

a-1 = 12, т. к. 12 \* 3 = 1 mod n.

x = (12 \* (2 + 35 - 7)) mod 35 = 10.

Т.е. расшифрованный символ – это символ, стоящий в позиции 10 – «H».

Все остальные символы дешифруем тем же способом:

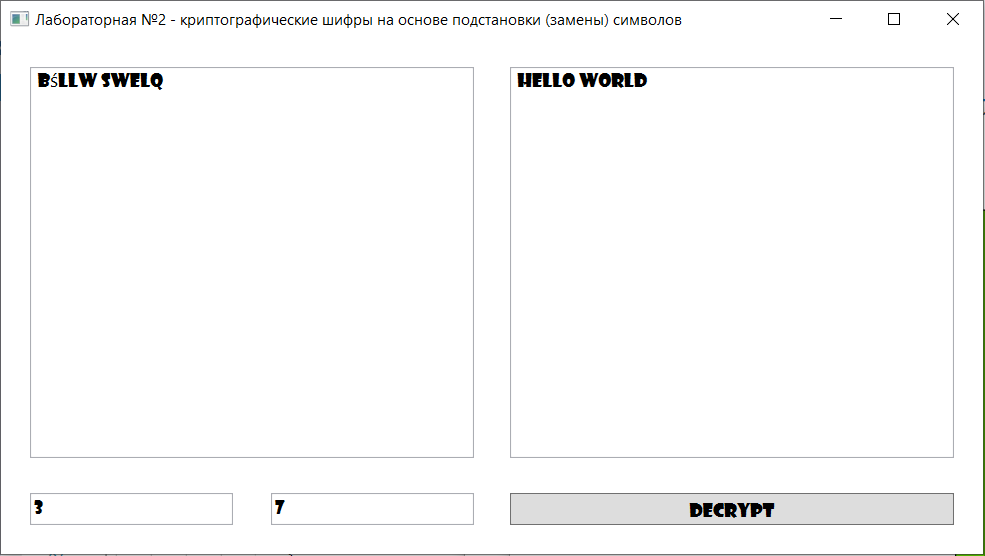


Рисунок 2.3 – Аффинная система подстановок Цезаря - дешифрование

Гистограмма встречаемости символов приведена на рисунке 2.2:

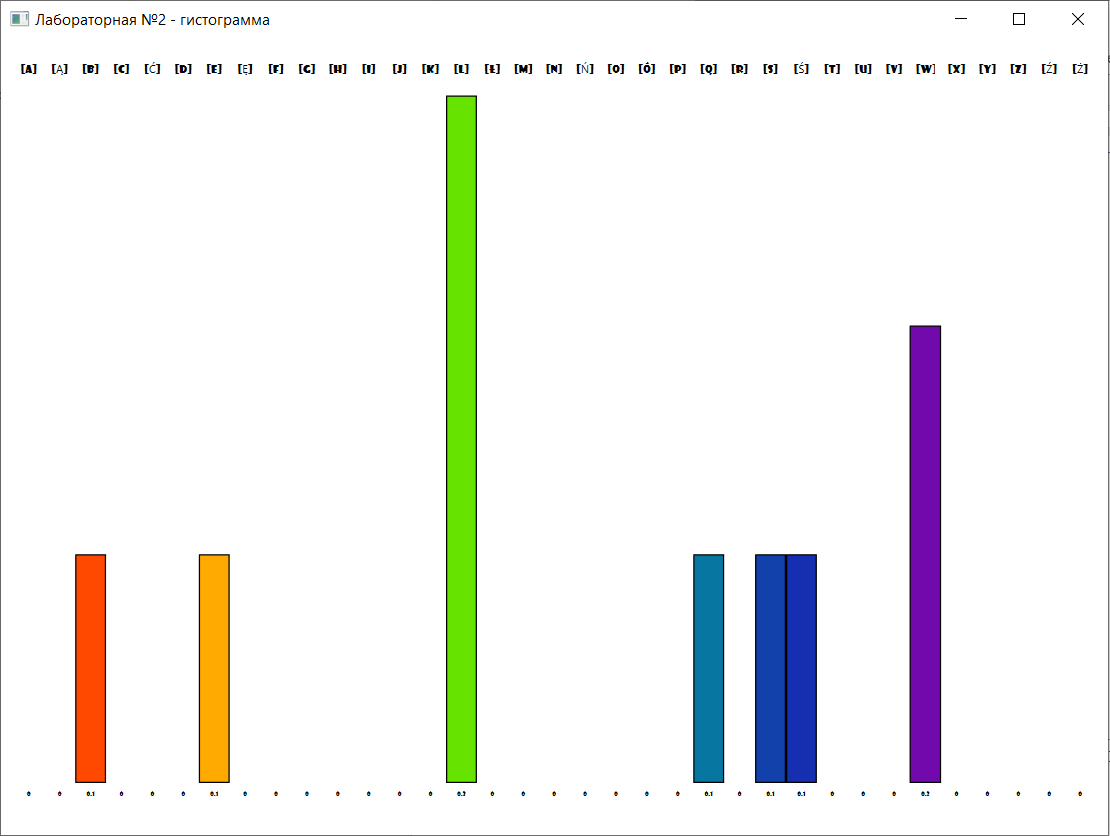


Рисунок 2.4 – Гистограмма частоты появления символа

Вторым шифром, который нужно было использовать в лабораторной работе, являет шифр Порты. Таблица, которая генерируется кодом программы, приведена на рисунке 2.5.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | Ą | B | … | Z | Ź | Ż |
| A | 0000 | 0001 | 0002 | … | 0032 | 0033 | 0034 |
| Ą | 0035 | 0036 | 0037 | … | 0067 | 0068 | 0069 |
| B | 0070 | 0071 | 0072 | … | 0102 | 0103 | 0104 |
| … | … | … | … | … | … | … | … |
| Z | 1120 | 1121 | 1122 | … | 1152 | 1153 | 1154 |
| Ź | 1155 | 1156 | 1157 | … | 1187 | 1188 | 1189 |
| Ż | 1190 | 1191 | 1192 | … | 1222 | 1223 | 1224 |

Рисунок 2.5 – Таблица шифра Порты для польского алфавита

Введем текст, дополнительный символ (если не найдется пары другому символу – берет его). Рассмотрим принцип действия алгоритма на первых двух символах текста - «HE».

«H» - 10 символ польского алфавита, «E» - 6 (нумерация с нуля).

Цепочка символов, получаемая после шифрации, считается по формуле:

y = x[i] \* n + x[i+1],

где x[i], x[i+1] – соответствующие номера символов, n – число символов в алфавите.

Тогда для нашего примера y = 10 \* 35 + 6 = 0356.

Все остальные символы дешифруем тем же способом:

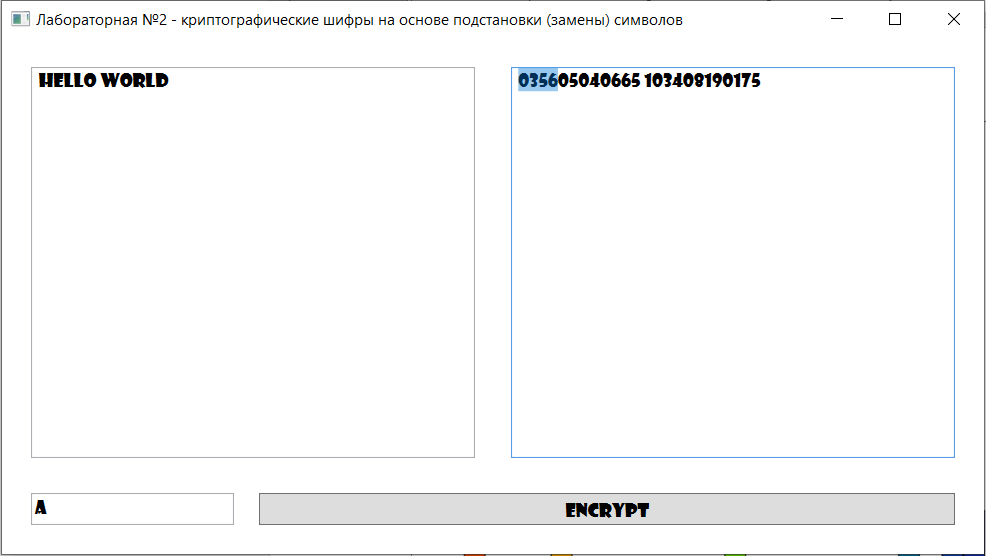


Рисунок 2.6 – Шифр Порты - шифрование

Гистограмма встречаемости символов приведена на рисунке 2.7:

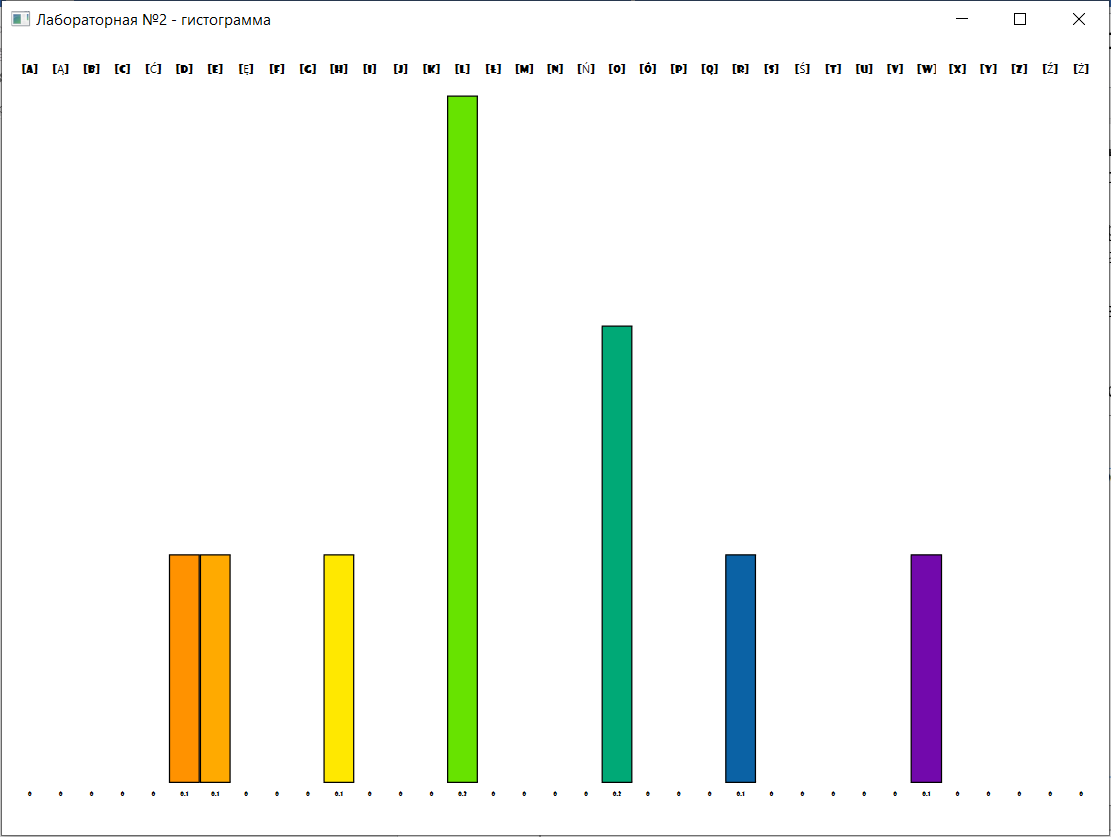


Рисунок 2.7 – Гистограмма частоты появления символа

Теперь дешифруем полученный шифротекст. Берем первые 4 символа – 0356. Согласно предыдущей формуле:

x[i] = y / n;

x[i+1] = y % n;

Тогда для нашего примера x[i] = 10, x[i+1] = 6.

10 символ польского алфавита – «H», 6 – «E».

Все остальные символы дешифруем тем же способом:

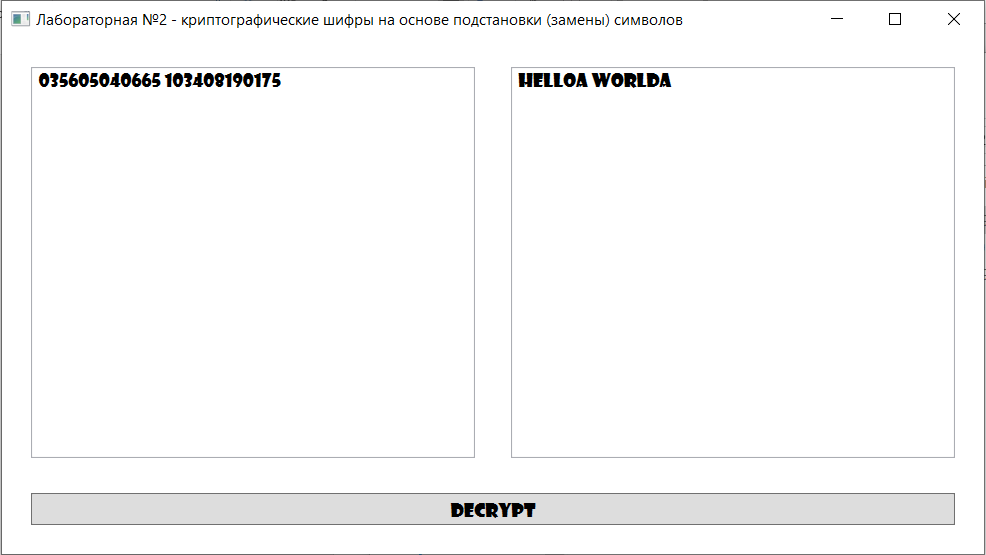


Рисунок 2.8 – Шифр Порты - дешифрование

Гистограмма встречаемости символов приведена на рисунке 2.8:

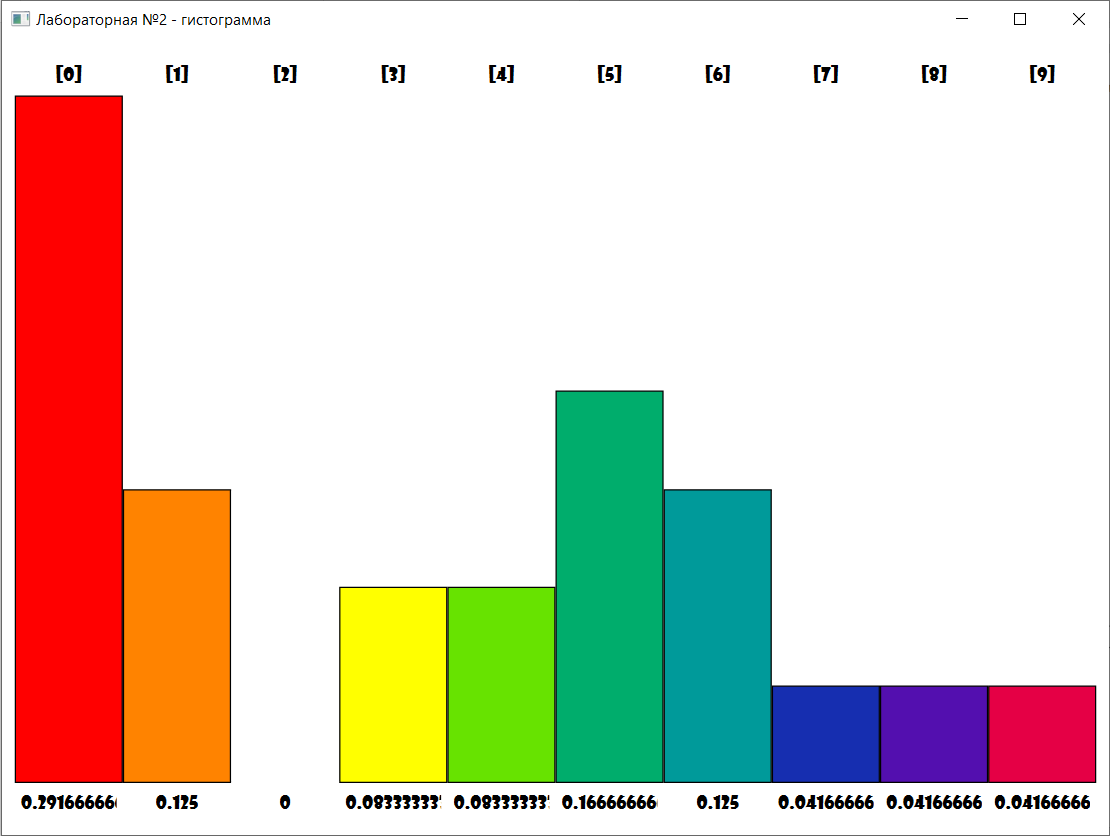


Рисунок 2.9 – Гистограмма частоты появления символа

**Вывод**

В данной лабораторной работе я закрепил теоретические знания по алгебраическому описанию, алгоритмам реализации операций зашифрования/расшифрования и оценке криптостойкости шифров. Ознакомился с особенностями реализации и свойствами различных подстановочных шифров. Разработал приложения для реализации указанных преподавателем методов подстановочного зашифрования/расшифрования. А также выполнял исследование криптостойкости шифров на основе статистических данных о частотах появления символов в исходном и зашифрованном сообщениях.