Національний технічний університет України «Київський Політехнічний Інститут імені Ігоря Сікорського»

Комп’ютерний практикум №2
Криптоаналіз шифру Віженера

Виконали:
студенти групи ФІ-94
Маринін Іван Павло Ігорович
Немкович Ольга Михайлівна

Перевірив:
Чорний Олег Миколайович

Київ – 2022

Зміст

1. Мета комп’ютерного практикуму. 3
2. Постановка задачі та варіант завдання. 4
3. Хід роботи, опис труднощів, що виникали під час виконання завдання, та шляхи їх подолання. 5
3.1 Хід роботи: 5
3.2 Опис труднощів та шляхи їх подолання: 5
4. Обчислені значення індексів відповідності I_r для вказаних значень r для шифротексту варіанту №7. 6
5. Обчислена послідовність D_r. 7
6. Значення ключа. 8
6.1 Значення ключа, отримане шляхом співставлення найчастіших літер блоків та найчастіших літер мови. 8
6.2 Значення ключа, отримане використанням функції $M_i(g)$. 8
7. Фрагмент шифрованого тексту та його розифрування згідно варіанту №7. 9
8. Шифрування власного тексту шифром Віженера. 10
9. Висновки. 11

1. Мета комп’ютерного практикуму.

Засвоїти методи частотного криптоаналізу. Здобути уміння роботи та аналізу потокових шифрів гамування адитивного типу на прикладі шифру Віженера.

2. Постановка задачі та варіант завдання.

Нехай A = {a_0, a_1 , ..., a_{m-1}} – алфавіт відкритого (ВТ) та шифрованого (ШТ) текстів, що складається з m букв. Природнім чином можна замінити символи алфавіту їх номерами і перевести множину A у кільце Z = {0,1, ..., m−1} m із відповідними операціями додавання та множення. Шифр Віженера є прикладом поліалфавітноі підстановки. Ключем цього шифру є послідовність r букв алфавіту (k_0, k_1, ..., k_{r-1}), яку підписують під ВТ, повторюючи стільки разів, скільки потрібно. Часто в якості ключа використовують якусь фразу або уривок тексту. Число r називається періодом шифру Віженера.

Позначимо ВТ через X = x_0, x_1, …, x_n , а ШТ через n Y = y_0, y_1, …, y_n. Шифрування відбувається шляхом додавання букв ВТ до підписаних під ними букв ключа за модулем m, тобто $y_i = (x_i + k_i \bmod r) \bmod m$, i = 0,n . Криптоаналіз шифру Віженера починають з визначення періоду r. Зробити це можна тому, що шифр Віженера зберігає деякі статистичні властивості мови. Після встановлення значення періоду шифру подальше його розшифрування зводиться до серії розшифрувань шифрів Цезаря. У результаті, буде отримано відкритий текст із можливими викривленнями, які піддаються редагуванню.

Ми розшифровували текст під варіантом №7 (І.П.Маринін – ФІ-94).

Програмна реалізація виконувалася за допомогою мови програмування Python у середовищі розробки PyCharm. Задля кращої наглядності отримані результати роботи алгоритму конвертувалися у текстові файли або ж документи Exel.

3. Хід роботи, опис труднощів, що виникали під час виконання завдання, та шляхи їх подолання.

3.1 Хід роботи:

1. Самостійно підібрано текст для шифрування (2-3 кб) та ключі довжини r = 2, 3, 4, 5, а також довжини 10-20 знаків. Зашифровано обраний відкритий текст шифром Віженера з цими ключами.

2. Підрахувано індекси відповідності r_I для відкритого тексту та всіх одержаних шифротекстів і порівняно їх значення.

3. Використовуючи наведені теоретичні відомості, розшифровано наданий шифртекст (варіант №7). Зокрема, необхідно:

– визначити довжину ключа, використовуючи або метод індексів відповідності, або статистику співпадінь Dr (на вибір);

– визначити символи ключа, прирівнюючи найчастіші літери у блоці до найчастішої літери у мові;

– визначити символи ключа за допомогою функції M (g) i ;

– розшифрувати текст, використовуючи знайдений ключ; в разі необхідності скорегувати ключ.

3.2 Опис труднощів та шляхи їх подолання:

Алгоритм не викликав значних труднощів у програмній реалізації. Найбільш трудомісткою виявилася задача пошуку довжини ключа методом індексів відповідності. Ми декілька годин шукали проблему у написаному скрипті, тому що ні ключ, ні розшифрований текст не мали найменшого логічного сенсу.

На перший погляд отримані значення індексів відповідності мали деякий логічний сенс: були отримані 2 значення при довжині ключа у 15 та 30 символів, що відрізнялися від решти (0.040611776250711704 при r = 15 та I = 0.04060552171114859 при r = 30 проти всіх інших, які були приблизно рівні 0,031), але їх значення не збігалися до теоретичного I.

Зрештою, було вирішено спробувати алгоритм на шифротекстах інших варіантів. Для кращого розуміння роботи алгоритму потрібно зауважити, що шифротексти було попередньо завантажені у вигляді текстових файлів і мали пройти фільтрацію у вигляді усунення переносів рядків, а далі вже завантажувалися у середовище розробки. Власне, саме тут і була помилка, яка спричинила затримку у виконанні практикуму: початкова фільтрація для нашого варіанту не була проведена коректно. Такий висновок ми зробили проводячи дослідити із іншими шифротекстами, для яких усунення переносів відбувалося належним чином. Результатом вирішення даної проблеми є адекватний ключ та майже коректно розшифрований текст.

4. Обчислені значення індексів відповідності I_r для вказаних значень r для шифротексту варіанту №7.

Повна таблиця з усіма значеннями I_{Yi} для кожного r = 2, 3, …, 32 надсилається окремо у вигляді таблиці Exel.

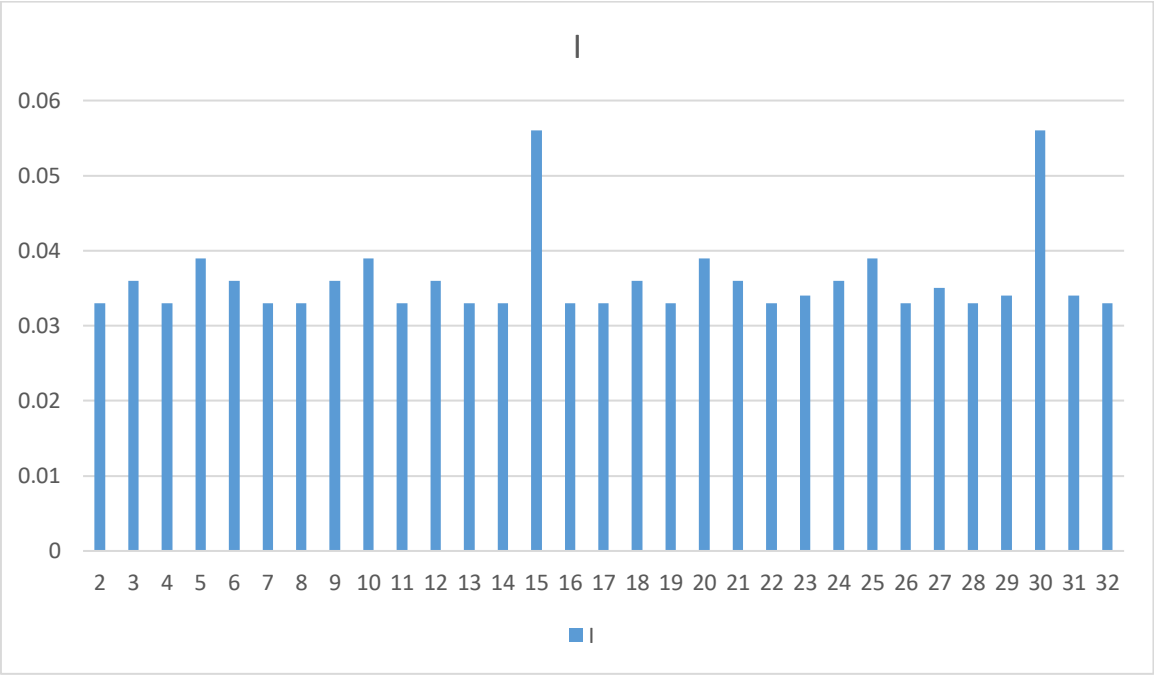
r	2	3	4	5	6	7	8	9	10	11
I_r	0.033	0.036	0.033	0.039	0.036	0.033	0.033	0.036	0.039	0.033

r	12	13	14	15	16	17	18	19	20	21
I_r	0.036	0.033	0.033	0.056	0.033	0.033	0.036	0.033	0.039	0.036

r	22	23	24	25	26	27	28	29	30	31
I_r	0.033	0.034	0.036	0.039	0.033	0.035	0.033	0.034	0.056	0.034

r	32
I_r	0.033

Таблиця 1.



Діаграма 1.

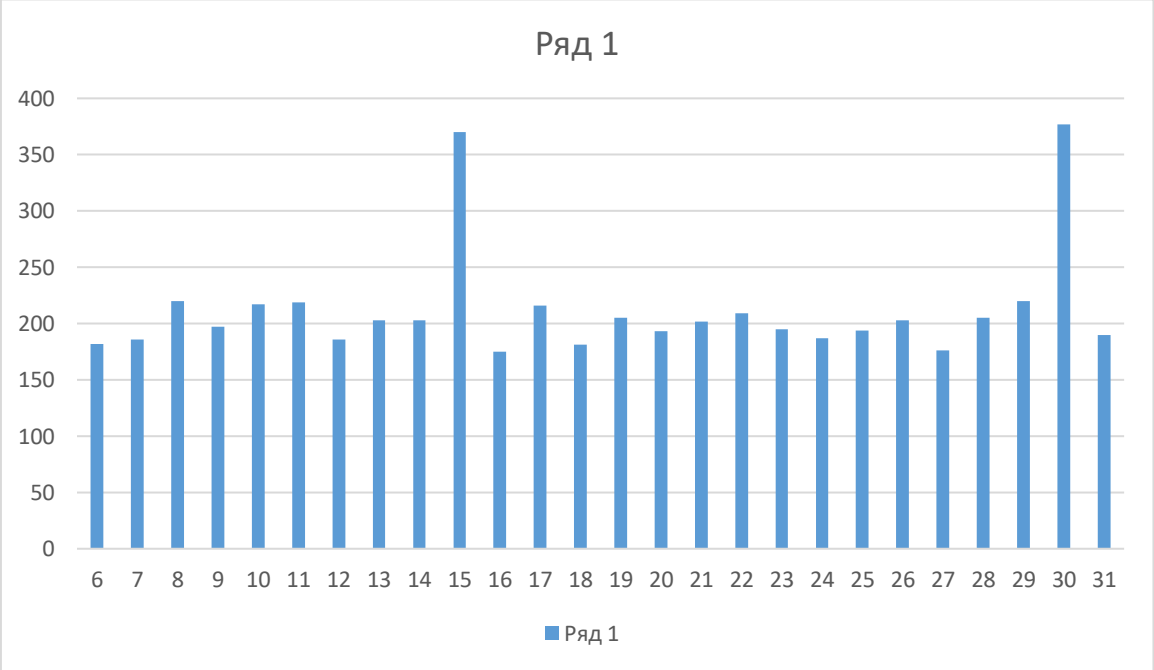
Згідно отриманих даних можна зробити висновок, що 15 – довжина ключа із відповідним значенням I = 0.05605177331202787. Також схоже значення має місце при r = 30:
I = 0.05600313156972977, але воно менше, і 30 – кратне для 15, чим і пояснюється схожість.

5. Обчислена послідовність D_r.

D_r	182	186	220	197	217	219	186	203	203	370	175	216	181
r	6	7	8	9	10	11	12	13	14	15	16	17	18

205	193	202	209	195	187	194	203	176	205	220	377	190
19	20	21	22	23	24	25	26	27	28	29	30	31

Таблиця 2.



Діаграма 2.

Як і в попередньому випадку бачимо 2 значення, які суттєво більші відносно інших, що приблизно рівні між собою. Тоді можемо зробити висновок, що довжина ключа рівна 15.

6. Значення ключа.

6.1 Значення ключа, отримане шляхом співставлення найчастіших літер блоків та найчастіших літер мови.

['а', 'р', 'у', 'д', 'а', 'з', 'е', 'в', 'а', 'р', 'х', 'и', 'м', 'а', 'г']

Саме такий ключ було отримано даним методом. У подальшому під час розшифрування було замічено спотворення літер на позиції №6 у ключі. Отже, дана літера була вгадана неправильно. Було зроблено припущення, що можна знайти у відритих джерелах правильність написання імені цього архимага. Ми знайшли твір про РудазОва архимага. Далі замінили літеру ‘е’ на літеру ‘о’ у ключі, розшифрували та отримали цілком правильний текст.

6.2 Значення ключа, отримане використанням функції $M_i(g)$.

[0, 16, 19, 4, 0, 7, 14, 2, 0, 16, 21, 8, 12, 0, 3] =
= ['а', 'р', 'у', 'д', 'а', 'з', 'о', 'в', 'а', 'р', 'х', 'и', 'м', 'а', 'г']

Даним методом нам вдалося одразу знайти правильний ключ, що й підтвердило на прикладі твердження з методички до КП№2, що даний метод є більш надійним, оскільки використовує увесь розподіл частот літер у блоці. Повна таблиця залежності g від $M_i(g)$ надсилається окремо і підписана відповідно.

7. Фрагмент шифрованого тексту та його розифрування згідно варіанту №7.

- Візьмемо фрагмент шифрованого тексту у вигляді його перших 385 символів:

  пабьлхэбтэхмвахьфаййпяфаарсроппюдцецупнювигаооцыжащкуоагтчехвэшрнпшфоэъофлтоэу
  хтхныеьипмэхотгймжьпсьъхфлсдшасалдвтмкцуяивэбсисаричврбнивлчйрнцдаыччъдсбэбрммя
  фесгуишиташщммябцхчтьеслшхднмяуабзичизвхаддэофыьэфмгтоыатсщкапюшшязллбтжрзпр
  тггхътуытупсжарлмяцуахеькцоийсохжъиастбадиопввыфуэякаъюгтпуобхжщънрижосолщбкаъ
  цчаатютжнхызпагэъдллюфйзфомачххщожлръдуфуеоягтьафнхюмайумиэхйьянлшыттйцулш

- Після розшифрування ключем ['а', 'р', 'у', 'д', 'а', 'з', 'е', 'в', 'а', 'р', 'х', 'и', 'м', 'а', 'г'] маємо:

  прошлошятнадцатьднейиътарыйдомпостепонноначаложиватесороклетвнемниутонежилпонас
  тоищемузаэтовремячнсменилодиннадяатьхозяевноникыоизнихневыдержсвалвподобноммеъте
  большетрехмеъяцевкреоливанеъсасталидвенадцйтымимагполностеюпогрузилсявракотуонотр
  ывалсяыолькозатемчтобдпоестьаотснаизкавлялсязаклятиомбессонницынодфякреолаэтоявноц
  епроходилобезнйказанноглазауногопокраснелиавокинабряклииотвссли

- Після розшифрування скорегованим ключем ['а', 'р', 'у', 'д', 'а', 'з', 'о', 'в', 'а', 'р', 'х', 'и', 'м', 'а', 'г']:

  прошлопятнадцатьднейистарыйдомпостепенноначаложиватьсороклетвнемниктонежилпонаст
  оящемузаэтовремяонсменилодиннадцатьхозяевноникгоизнихневыдерживалвподобномместеб
  ольшетрехмесяцевкреоливанессасталидвенадцатымимагполностьюпогрузилсявработуонотры
  валсятолькозатемчтобыпоестьаотснаизбавлялсязаклятиембессонницынодлякреолаэтоявнонеп
  роходилобезнаказанноглазаунегопокраснелиавекинабряклииотвисли

8. Шифрування власного тексту шифром Віженера.

У ході комп’ютерного практикуму ми самостійно обрали текст російською мовою ваги 2кБ. Його було зашифровано випадковими ключами із довжинами 2,3,4,5,10,11,…,20 символів.

Наведений відкритий текст та результати шифрування надсилаються окремо у вигляді текстових файлів, що підписані відповідно. Нижче наведена таблиця значень індекса відповідності при відповідних довжинах ключів для даного тексту.

r	I_r
0	0.059291509778385654
2	0.05075642401884657
3	0.03715187148874491
4	0.03702593704609867
5	0.03673289728532567
10	0.032928829481737386
11	0.03369896703484328
12	0.03223437368502948
13	0.033701994305099195
14	0.03250440619185749
15	0.031835984919350496
16	0.03161378328256601
17	0.032345171776396134
18	0.031604096017747066
19	0.031428514342903745
20	0.031201469073709794

9. Висновки.

У даному комп’ютернопму практикумі ми проводили криптоаналіз шифру Віженера. Спершу ми нагадали собі теоретичні відомості на тему шифру Віженера, далі ознайомилися із необхідними для дешифрування методами та алгоритмами. Ми дізналися алгоритм дій, який потрібен для того, щоб розшифрувати текст, шифрований шифром Віженера із невідомим ключем, частково, або навіть повністю.

У роботі було розглянуто 2 метода знаходження довжини ключа Віженера: 1) за індексами відповідності, 2) за статистикою ключів. Також ми опрацьовували 2 методи знаходження символ ключа при його відомій довжині: 1) прирівнюючи найчастіші літери у блоці до найчастіший літер у мові, 2) за допомогою функції $M_i(g)$. Ми навчилися корегувати ключ за такої необхідності. Також ми розшифрували шифротекст вагою 7кБ та зашифрували власний відкритий текст довжиною 2кБ із ключами довжини від 2 до 20 символів.