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.

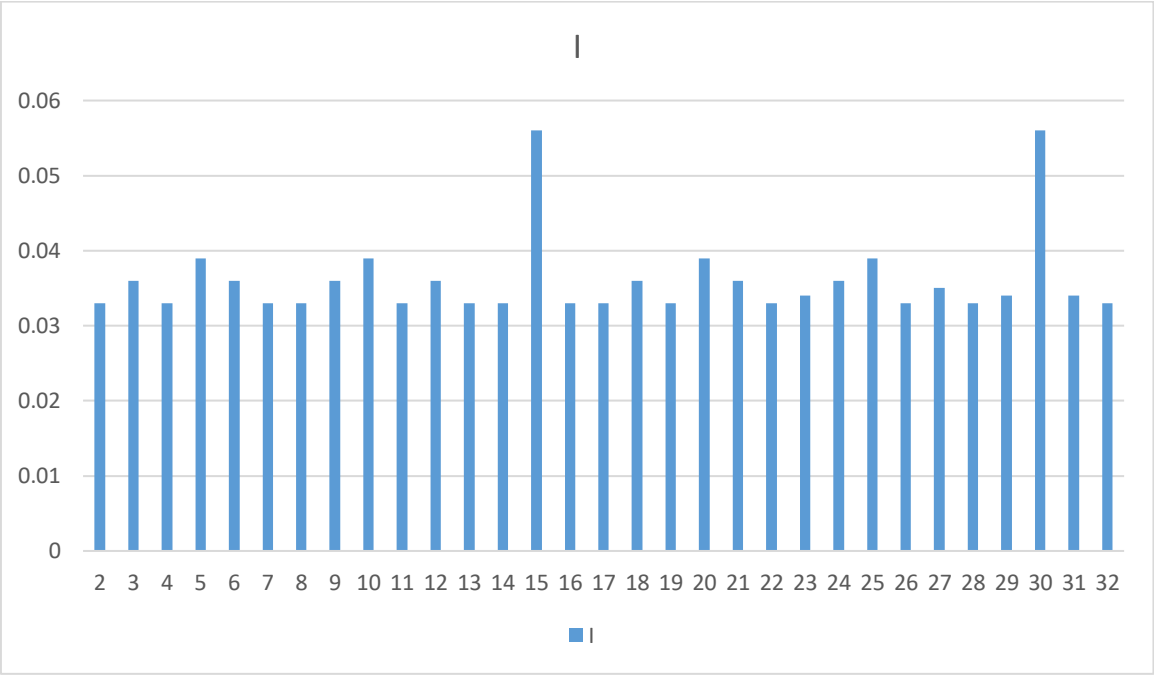
r	2	3	4	5	6	7	8	9	10	11
I_r	0.033	0.036	0.033	0.039	0.036	0.033	0.033	0.036	0.039	0.033

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.

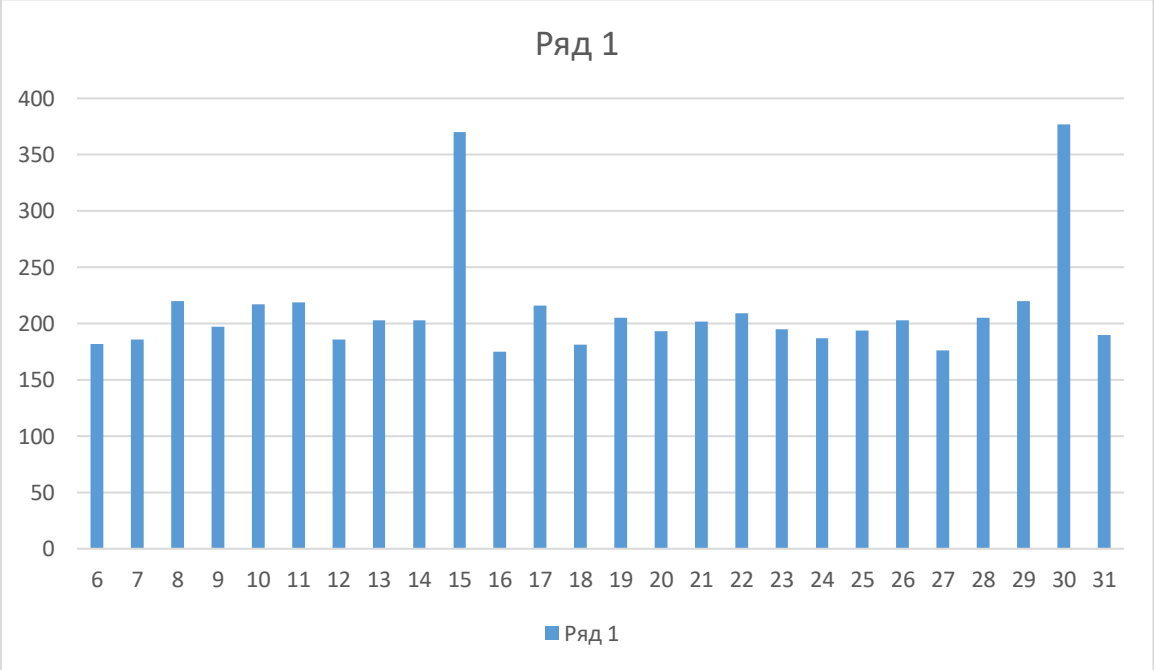
Згідно отриманих даних можна зробити висновок, що 15 – довжина ключа із відповідним значенням I = 0.05605177331202787. Також схоже значення має місце при r = 30:
I = 0.05600313156972977, але воно менше, і 30 – кратне для 15, чим і пояснюється схожість.

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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_10	Y_11	Y_12	Y_13	Y_14
2	а	24,3042	10,3652	11,659	16,258	23,5	10,9169	14,1667	16,2574	24,9912	10,9076	11,1139	12,1073	10,8597	24,0182	17,5902
3	б	15,4736	10,3823	14,2283	17,2244	15,7827	13,335	13,4015	16,3355	15,7871	8,94939	10,5759	10,0957	10,6337	15,4466	15,3773
4	в	14,9026	14,5556	10,1551	14,3146	14,7844	14,4527	11,2976	24,1117	14,7563	13,1177	11,5891	13,8421	12,1601	16,2327	15,7607
5	г	16,8498	12,2569	10,3718	15,6887	16,011	16,5655	10,5842	15,8421	16,1733	12,44	14,9879	15,3911	13,2239	16,2657	24,9565
6	д	14,0924	10,4686	10,5264	23,3537	15,1903	17,0011	11,1705	14,5116	13,2426	10,9615	10,6282	15,3773	12,0473	14,011	15,901
7	е	14,4439	9,72057	13,9824	15,5809	14,8152	15,0798	13,8927	16,7294	15,8212	9,50935	11,8663	16,8955	10,341	14,4422	15,412
8	ж	14,3394	11,9323	12,6535	14,2992	14,1089	16,0209	12,5957	14,9103	13,8295	11,0545	10,5033	15,3977	13,1469	14,6271	16,5528
9	з	9,99395	12,857	9,94554	16,4752	10,2712	23,8647	11,5127	14,3372	9,02365	12,7255	14,2277	15,7041	14,8289	9,478	13,4048
10	и	10,7811	11,8922	9,56931	13,7123	11,5957	15,0611	13,5116	13,0072	11,7431	11,6491	12,7492	25,4433	15,511	12,2734	15,8465
11	й	14,1293	10,2112	11,7129	15,2431	14,297	14,7794	15,6029	9,36194	14,0292	11,3872	11,3674	15,5787	15,6436	13,6865	14,0682
12	к	10,7756	13,9862	12,516	13,9824	11,4059	16,2063	14,5864	10,9692	10,8487	13,4615	9,78053	15,495	15,5908	10,8493	9,44114
13	л	11,7184	15,852	11,0517	10,0435	11,6689	13,4488	17,4477	13,0578	10,8069	14,7591	10,9131	16,6975	17,3344	11,0561	11,1601
14	м	10,0517	16,9131	10,6403	11,2019	10,1128	14,5281	14,7305	10,4763	9,65732	16,538	13,0743	15,3311	23,6623	10,1045	14,0545
15	н	11,3905	16,5473	13,6139	14,2717	11,9054	13,8691	15,632	9,63696	11,0913	18,6931	11,2041	14,7651	15,0721	10,8108	10,9235
16	о	14,0831	14,9659	15,566	12,1254	14,8234	8,80638	22,9576	10,2976	14,6067	15,2849	10,6793	14,7222	16,0935	14,2712	10,4736
17	п	10,3982	16,3141	16,6095	11,2915	10,5215	11,5671	14,7415	10,8949	10,4153	15,6661	12,8344	9,76238	16,4527	9,33333	10,1419
18	р	11,3878	24,4103	17,1205	10,1496	10,6172	13,6095	14,0396	14,3295	10,4235	23,3966	14,9367	12,1458	15,2536	10,8075	9,93839
19	с	10,2151	15,3097	15,3438	11,9285	10,0528	10,4791	16,4769	10,4862	9,93784	15,8817	16,533	13,2448	15,8559	10,429	14,6381
20	т	14,5479	14,1678	17,5996	15,071	14,0028	10,4758	13,8663	10,5957	13,6562	14,6441	17,0275	11,0699	13,0396	13,313	9,98295
21	у	12,0886	16,0281	24,9466	11,1249	12,6826	9,4791	15,3658	10,3058	13,5099	15,5446	14,8559	10,6755	9,71727	12,5242	10,0534
22	ф	11,0347	13,6843	15,2849	11,0072	9,60891	11,0176	13,6469	14,5369	10,5385	13,9626	16,4961	10,1331	12,1727	10,8553	9,59736
23	х	10,0418	14,9824	14,3977	10,1843	9,76128	14,5237	9,72717	13,0627	10,1392	15,1342	23,5099	11,3377	12,7074	10,4549	13,1463
24	ц	10,2035	13,555	17,0457	13,3955	11,363	10,8608	11,6804	10,0358	11,1903	14,247	14,7767	14,2937	10,4037	11,6694	13,4791
25	ч	11,6738	8,94774	14,1997	13,2437	12,2327	10,5803	12,7783	10,4142	13,7222	10,445	14,9945	9,65127	10,6628	13,2651	10,4263
26	ш	10,956	12,1243	14,7948	10,8658	11,687	10,7426	11,2772	11,9873	11,8113	11,2288	15,9065	10,956	9,79153	11,8971	9,90979
27	щ	11,2145	13,6777	13,6617	9,5308	11,1513	14,335	11,0077	13,2629	9,9582	13,5886	13,0171	9,78878	11,5039	9,94939	10,5726
28	ъ	12,1782	10,6639	9,04785	11,0451	12,4048	13,967	9,59131	12,7096	13,4285	12,3729	15,2123	14,0275	13,6892	13,0462	14,1007
29	ы	15,0798	10,9224	11,5726	12,995	15,2596	11,0578	11,3619	11,1711	15,4175	11,6023	13,5699	11,5809	9,28383	14,4428	12,2822
30	ь	15,148	9,82618	14,3031	11,4417	16,077	10,2822	14,6128	13,484	16,1221	9,79373	9,70187	10,2266	10,8603	16,4516	10,538
31	э	18,7662	11,4345	10,1535	10,7398	16,1837	11,9197	10,192	15,4692	17,6337	11,4395	10,8757	9,70847	10,4631	16,7706	13,1051
32	ю	15,7415	15,2849	10,2833	13,1579	15,8064	13,5358	11,2233	16,3625	15,1755	14,2547	14,8031	11,0017	14,5138	16,5028	14,3515
33	я	16,995	10,7602	10,4433	14,0528	15,3146	12,6315	10,3207	16,0495	15,5121	10,3592	10,6887	12,5523	12,4796	15,7145	16,8135

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.