Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Фізико-технічний інститут

КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №2

Криптоаналіз шифру Віженера

Виконали:
ФБ-31 Острун Катерина
ФБ-31 Острун Михайло

Мета роботи:

Засвоєння методів частотного криптоаналізу. Здобуття навичок роботи та аналізу потокових шифрів гамування адитивного типу на прикладі шифру Віженера.

Хід роботи:

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

Спочатку ми підготували дані. Взявши текст з фрагменту Л. Толстого «Анна Кареніна», очищений від неалфавітних символів, довжиною близько 1000 символів, ми зашифрували його шифром Віженера для ключів довжиною r від 2 до 30.

Правила шифрування та дешифрування наступні. Припустимо, ВТ позначено як $X = (x_0, x_1, x_2, x_3, .. x_n)$, а ШТ як $Y = (y_0, y_1, y_2, y_3, .. y_n)$, тоді шифрування відбувається шляхом додавання букв ВТ до підписаних під ними букв ключа за модулем m, тобто

$$y_i = (x_i + k_{i \bmod r}) \bmod m, \qquad i = 0 \ldots . n$$

Ключі були обрані як випадкові літери алфавіту (m=32), наприклад

```
Ключ довжини 2: ео
Ключ довжини 3: яйа
Ключ довжини 4: ьндш
Ключ довжини 5: аыщць
Ключ довжини 6: пийсыд
Ключ довжини 7: ынщсуэд
Ключ довжини 8: ьефчяяот
Ключ довжини 9: ззэюухнмг
Ключ довжини 10: уняфбйбгиф
Ключ довжини 11: щнкбпссюзпщ
Ключ довжини 12: цькюэуыяцйлэ
Ключ довжини 13: ртйыйвюлвжщчэ
Ключ довжини 14: юисзшсюкзцувжо
Ключ довжини 15: чэозюсшещыблждй
Ключ довжини 16: эрлээыцюгшдяцаяа
Ключ довжини 17: бщецезлсочэлызузн
Ключ довжини 18: чухяджофрхпбгшмыхц
Ключ довжини 19: пюасяхтэнюбьоймччов
Ключ довжини 20: гхщтьацнсгахмоорнеба
```

Результат записано у файли cipher_r*.txt для кожного r. Відповідно текст до

```
Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива
по-своему. Все смешалось в доме Облонских. Жена узнала, что муж был в связи с
бывшею в их доме француженкою-гувернанткой, и объявила мужу, что не может жить с
ним в одном доме. Положение это продолжалось уже третий день и мучительно
чувствовалось и самими супругами, и всеми членами…
```

та після шифрування (наприклад для ключа довжини 3):

```
бъераарылзлыдъелеиочхнпигщувцагщувйкяпдяиндъчяътксвяисдхьюцераарылзлаочсб
чельврослошяфоревгчмдчбкчнруифпемйужцакйчсчмтпбъфврляжссадвчоюбсхгчмдэряц
цтпемуоэмубормйнсуоисоаеябсляхуеьчсчндхоеотестыънзхвнннхднхеочлнпемсеьыо
ощогчл…
```

2. Підрахувати індекси відповідності для відкритого тексту та всіх одержаних шифртекстів і порівняти їх значення.

Індекс відповідності I(Y) для тексту Y довжиною n обчислювався як

$$I(Y) = \frac{1}{n(n-1)} \sum_{t \in Z_m} N_t(Y)(N_t(Y) - 1)$$

Теоретичне математичне очікування IC для мови дорівнює

$$MI(Y) = \sum_{t \in Z_m} {p^2}_i$$

Після обчислень отримано результати:

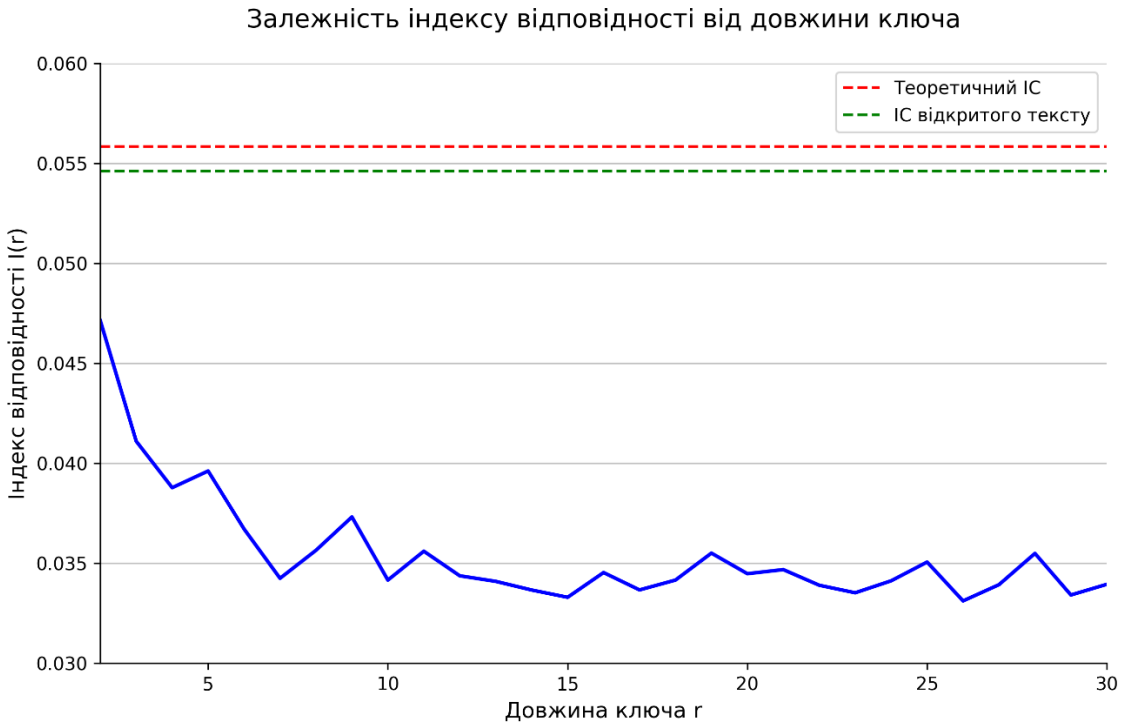
```
IC теоретичне: 0.055839
IC відкритого тексту: 0.054606
```

Далі, обчислили IC для кожного шифртексту cipher_r*.txt:

Довжина ключа (r)	IC для шифртексту
2	0.047138
3	0.041104
4	0.038781
5	0.039616
6	0.036713
7	0.034248
8	0.035649
9	0.037317
10	0.034159
11	0.035603
12	0.034373
13	0.034095
14	0.033662
15	0.033298
16	0.034536
17	0.033667
18	0.034159
19	0.035508
20	0.034479
21	0.034683
22	0.033901
23	0.033524
24	0.034123
25	0.035061
26	0.033118
27	0.033930
28	0.035493
29	0.033414
30	0.033948

Можна помітити, що IC стрімко падає з ростом r, наближаючись до 1/m ≈ 0.03125 для рівноймовірного алфавіту.

Далі побудовано графік залежності IC шифртексту від r. З графіка видно, що при зменшенні довжини ключа ймовірність появи літер стає більш відповідною природній мові, що знижує якість шифрування (легше виявити статистику). Навпаки, з ростом r шифр краще маскує текст.



3. Використовуючи наведені теоретичні відомості, розшифрувати наданий шифртекст (згідно свого номеру варіанта).

Тепер відповідно криптоаналіз наданого шифртексту з варіанту 8 (довжина 6138 символів) перші 50:

```
рэаюцугкьелаяюиутбхигцичопщпюиермтгсфюлхутвныкрчюрэънфожэчыцфуттщююуфрйэм
идтэяршххаяоняихнтбктяусунаыфетшкткампэгынсфеууаллхекцчакцуяфйз…
```

Використаємо перший алгоритм, тому для кожного кандидата r = 2...30 розбиваємо шифртекст Y на r блоків. Після цього обчислюємо IC для кожного блоку за формулою I(Y). Усереднюємо IC блоків (оскільки розміри блоків великі) і якщо середнє близьке до теоретичного IC мови, то r знайдено, якщо до $1/m \approx 0.03125$, то ні.

Результати середніх IC блоків для шифротексту:

Довжина ключа (r)	Середній IC блоків
2	0.036278
3	0.034787
4	0.037706
5	0.041734
6	0.036314
7	0.034701
8	0.037635
9	0.034897
10	0.047975
11	0.034772
12	0.037702
13	0.034561
14	0.036137
15	0.041620
16	0.037748
17	0.034901
18	0.036328
19	0.034638
20	0.057507
21	0.034615
22	0.036159
23	0.034658
24	0.037613
25	0.041888
26	0.036433
27	0.034811
28	0.037284
29	0.034652
30	0.047860

Можна побачити, шо найкраще r = 20 (максимальний середній IC: 0.057507, близьке до теоретичного 0.055839).

Далі для кожного блоку обчислюємо кореляцію частот зі зсувом k (0-31) та вибираємо k з максимальною кореляцією (відповідає шифру Цезаря).

Отримуємо деталі по блоках (наприклад для перших чотирьох блоків):

```
Блок 0:
  - Зсув 'у' (k=19): кореляція 0.054230
  - Зсув 'ц' (k=22): кореляція 0.041612
  - Зсув 'р' (k=16): кореляція 0.040124
Блок 1:
  - Зсув 'л' (k=11): кореляція 0.055930
  - Зсув 'о' (k=14): кореляція 0.046854
  - Зсув 'к' (k=10): кореляція 0.039332
Блок 2:
  - Зсув 'а' (k=0): кореляція 0.055781
  - Зсув 'э' (k=29): кореляція 0.040626
  - Зсув 'г' (k=3): кореляція 0.038325
Блок 3:
  - Зсув 'н' (k=13): кореляція 0.056255
  - Зсув 'о' (k=14): кореляція 0.042855
  - Зсув 'к' (k=10): кореляція 0.041629
Блок 4:
  - Зсув 'о' (k=14): кореляція 0.059202
  - Зсув 'с' (k=17): кореляція 0.042757
  - Зсув 'л' (k=11): кореляція 0.041549
```

Відповідно збираємо найімовірніші зсуви для кожного блоку та отримаємо початковий ключ:

```
улановсеребряныепули
```

Після застосування отриманого ключа для розшифрування ІС тексту становило 0.055593 і отримано наступний текст :

```
этасистемакрасногокарликаникогданеимеланазваниятольмозубодробительнодлинн
ыйномервкаталогеисследовавшийеекиберзондотметилналичиетрехгазовыхгигантов
двучастероидныхсолейкометногооблакаизанесвсеэтиданныев...
```

Висновки:

У ході аналізу ми дослідили методи частотного криптоаналізу шифру Віженера та підтвердили теоретичні відомості про залежність індексу відповідності (ІС) від довжини ключа r.

Для наданого шифртексту (варіант 8) перший алгоритм виявив довжину ключа 20 з максимальним середнім ІС блоків 0.057507. Частотний аналіз блоків дав ключ "улановсеребряныепули", який утворює змістовну фразу. Розшифрований текст має ІС 0.055593.

Загалом, робота показала, як шифр Віженера зберігає статистичні властивості мови для коротких ключів, роблячи його вразливим до частотного аналізу. Перший алгоритм, який ми використовували, ефективний для довгих періодів за умови достатньої статистики в блоках. Це підкреслює важливість вибору довгого ключа для підвищення стійкості шифру.