НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського»
Навчально-науковий Фізико-технічний інститут

Криптографія

Комп’ютерний практикум №2

Криптоаналіз шифру Віженера

Варіант №6

Виконали:
Студенти 3 курсу НН ФТІ
групи ФБ-31
Гаврилюк Володимир
Гек Роман

Мета роботи

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

Порядок виконання роботи

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
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.
2. Підрахувати індекси відповідності для відкритого тексту та всіх одержаних шифртекстів і порівняти їх значення.
3. Використовуючи наведені теоретичні відомості, розшифрувати наданий шифртекст (згідно свого номеру варіанта).

Виконання роботи

Для виконання першого пункту роботи було обрано текст обсягом приблизно 2.7 КБ (уривок з твору А.П. Чехова "Остров Сахалин"). Його ми назвали TEXT.txt .

Далі ми зробили попередню обробку тексту за доп. Програми filter.py . Тут все доволі просто:

```
lowered_text = original_text.lower().replace('ё', 'е')

filtered_chars = []
for char in lowered_text:
        if char in ALPHABET:
            filtered_chars.append(char)
```

1) Переведення всіх символів у нижній регістр
2) Заміна букви «ё» на «е»
3) Видалення всіх символів, окрім букв російського алфавіту (32 літери)
4) Видалення пробілів та знаків пунктуації

Результатом роботи програми став файл filtered_text.txt розміром 2743 символи.

```
поворачиваемнаправонавсемнашемпутипоставленызнакипоказывающиефарватеркомандирнесходитсмостикаимеханикневыходитизмашиныбайкалначинаетидтивсетишеитишеиидет
встановилсявсемельчеймельчеглубинауменьшалась черезкаждуюмилюнаоднусаженьплылонксеверудотехпорпокаемупозволялиразмерыегокорабляидойдядоглубинысаженостанов
```

Далі було саме шифрування. Для шифрування було обрано ключі різної довжини:
r = 2: ключ "ок"
r = 3: ключ "нет"
r = 4: ключ "ключ"
r = 5: ключ "осень"

r = 10: ключ "терминатор"
r = 16: ключ "командирнесходит"

```
keys = {
        'key2': "ок",
        'key3': "нет",
        'key4': "ключ",
        'key5': "осень",
        'key10': "терминатор",
        'key16': "командирнесходит"
}
```

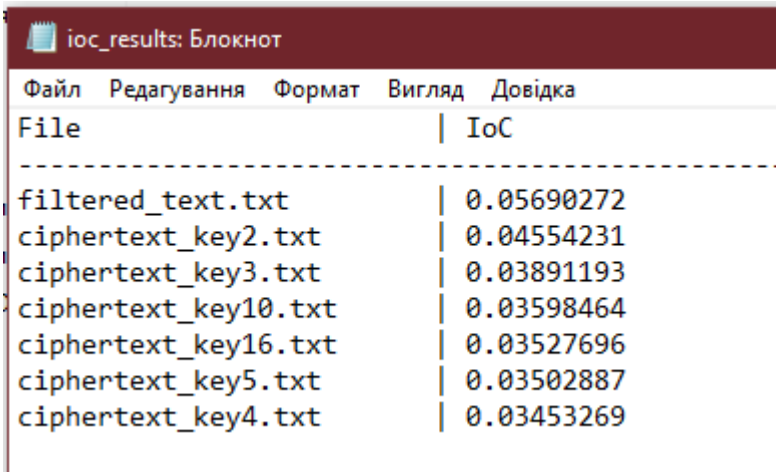
```
    # y = (x + k) mod m
    for i in range(len(plaintext)):
        char_num = letter_to_number[plaintext[i]]

        key_char = key[i % key_length]
        k = letter_to_number[key_char]

        c = (char_num + k) % M

        ciphertext += number_to_letter[c]
```

Далі було обчислення та аналіз індексів відповідності.
За доп. програми index.py порахували:

$$I(Y) = \frac{1}{n(n-1)} \sum_{i \in Z_m} N_i(Y)(N_i(Y) - 1),$$

```
def calculate_ioc(text):
    n = len(text)
    if n <= 1:
        return 0.0

    letter_counts = Counter(text)

    summation_term = 0.0
    for char in ALPHABET:
        nt = letter_counts.get(char, 0)
        summation_term += nt * (nt - 1)

    denominator = n * (n - 1)

    ioc = summation_term / denominator
    return ioc
```

В результаті програма обчислила індекси відповідності для відкритого тексту та всіх шифртекстів.
Результати представили у таблиці ioc_results.txt :



З отриманих результатів видно:

I) ВТ має найвищий індекс відповідності (0.0569), що близьке до теоретичного значення для російської мови (яке плаває десь поряд з 0.0529).

II) Одразу видно залежність від довжини ключа, а саме те що для коротких ключів (r=2, r=3) індекс відповідності залишається відносно високим (0.0455 та 0.0389 відповідно), що пояснюється збереженням частотних характеристик мови в кожному блоці.

III) Для довших ключів (r=4, 5, 10, 16) спостерігається неоднозначна картина, бо значення IoC не завжди монотонно спадають. Припускаємо, що це пов'язано з обмеженим обсягом тексту (2.7 КБ), через що статистична похибка стає суттєвою при розбитті на велику к-ть блоків.

IV) Згідно з теорією, при збільшенні довжини ключа індекс відповідності має прямувати до значення 1/m = 1/32 ≈ 0.0313 (індекс для рівноймовірного алфавіту) і в нас тут чудово отримані значення для довгих ключів (0.0346-0.0353) підтверджують цю тенденцію.

Далі було розшифрування. Був наданий зашифр. текст нашого варіанту 6.
За доп. програми vigenere_decrypt_v2.py ми використали обидва представлених методи визначення періоду:
Метод 1: Через індекс відповідності блоків
Для кожного кандидата періоду r (від 2 до 30) шифртекст розбивається на r блоків, для кожного блоку обчислюється індекс відповідності, після чого знаходиться середнє значення. Період вважається знайденим, коли сер. індекс близький до теоретичного значення для російської мови (0.0529).

```
55      for period_candidate in range(2, max_period + 1):
56          blocks = split_into_blocks(ciphertext, period_candidate)
57
58          block_indices = [compute_coincidence_index(block) for block in blocks]
59          average_index = sum(block_indices) / len(block_indices)
60          deviation = abs(average_index - theoretical_value) #до теор.
61          print(f"Період {period_candidate:2d}: середній IC = {average_index:.5f}, "
62                f"відхилення = {deviation:.5f}")
63
64          if deviation < best_score:
65              best_score = deviation
66              best_period = period_candidate
```

Метод 2: Статистика співпадінь
Для кожної відстані r обчислюється к-ть співпадінь символів D_r:

$$D_r = \sum_{i=1}^{n-r} \delta(y_i, y_{i+r}),$$

де δ(a,b) - символ Кронекера (дорівнює 1, якщо a=b, інакше 0).
Локальні максимуми D_r вказують на можливі значення періоду або кратні йому величини.

```
86      coincidence_stats = {}
87      for distance in range(1, max_period + 1):
88          coincidences = count_coincidences(ciphertext, distance)
89          coincidence_stats[distance] = coincidences
90          print(f"D_{distance:2d} = {coincidences:4d}")
91
92      sorted_by_value = sorted(coincidence_stats.items(),
93                               key=lambda x: x[1], reverse=True) #лок.макс.
```

Результати визначення періоду

Метод 1 дав період: r = 17

```
=== M1: Індекс відповідності блоків ===
Теор. індекс для російської: 0.05590
Індекс для випадкового тексту: 0.03125

Період  2: середній IC = 0.03406, відхилення = 0.02185
Період  3: середній IC = 0.03408, відхилення = 0.02182
Період  4: середній IC = 0.03409, відхилення = 0.02181
Період  5: середній IC = 0.03412, відхилення = 0.02178
Період  6: середній IC = 0.03398, відхилення = 0.02193
Період  7: середній IC = 0.03416, відхилення = 0.02175
Період  8: середній IC = 0.03403, відхилення = 0.02188
Період  9: середній IC = 0.03392, відхилення = 0.02199
Період 10: середній IC = 0.03393, відхилення = 0.02198
Період 11: середній IC = 0.03401, відхилення = 0.02189
Період 12: середній IC = 0.03406, відхилення = 0.02185
Період 13: середній IC = 0.03404, відхилення = 0.02187
Період 14: середній IC = 0.03408, відхилення = 0.02182
Період 15: середній IC = 0.03423, відхилення = 0.02168
Період 16: середній IC = 0.03403, відхилення = 0.02188
Період 17: середній IC = 0.05551, відхилення = 0.00040
Період 18: середній IC = 0.03384, відхилення = 0.02207
Період 19: середній IC = 0.03393, відхилення = 0.02198
Період 20: середній IC = 0.03380, відхилення = 0.02210
Період 21: середній IC = 0.03424, відхилення = 0.02166
Період 22: середній IC = 0.03417, відхилення = 0.02173
Період 23: середній IC = 0.03394, відхилення = 0.02196
Період 24: середній IC = 0.03408, відхилення = 0.02182
Період 25: середній IC = 0.03374, відхилення = 0.02216
Період 26: середній IC = 0.03386, відхилення = 0.02205
Період 27: середній IC = 0.03409, відхилення = 0.02181
Період 28: середній IC = 0.03388, відхилення = 0.02203
Період 29: середній IC = 0.03408, відхилення = 0.02183
Період 30: середній IC = 0.03394, відхилення = 0.02197

>>> Найкр. період за M1: 17
```

Метод 2 дав період: r = 17

```
=== M2: Стат. співпадінь Dr ===

D_ 1 =  199
D_ 2 =  207
D_ 3 =  220
D_ 4 =  257
D_ 5 =  212
D_ 6 =  234
D_ 7 =  220
D_ 8 =  226
D_ 9 =  220
D_10 =  244
D_11 =  233
D_12 =  227
D_13 =  242
D_14 =  225
D_15 =  218
D_16 =  214
D_17 =  394
D_18 =  212
D_19 =  202
D_20 =  205
D_21 =  228
D_22 =  203
D_23 =  254
D_24 =  227
D_25 =  218
D_26 =  204
D_27 =  248
D_28 =  258
D_29 =  210
D_30 =  223

Топ5 відстаней за к-тю співпадінь:
  Відстань 17: 394 співпадінь
  Відстань 28: 258 співпадінь
  Відстань 4: 257 співпадінь
  Відстань 23: 254 співпадінь
  Відстань 27: 248 співпадінь

>>> Найкращий період за м2: 17
```

Обидва методи однозначно вказали на період **r = 17**.

Далі залишалося знайти логічний підходящий ключ:
Після встановлення періоду шифртекст було розбито на 17 блоків. Для кожного блоку:
1) Визначаємо найчаст. букву
2) Припускаємо, що вона відповідає найімовірнішій букві рос. мови
3)Обчислюємо ключ за формулою: k=(y- x)modm
4)І нарешті обираємо варіант з найкращою оцінкою якості тексту (мінімальна квадратична різниця частот)

```
=== Розшифр. з періодом 17 ===

Блок 0: найчаста буква 'р', ключ = 2 ('в')
Блок 1: найчаста буква 'ь', ключ = 14 ('о')
Блок 2: найчаста буква 'х', ключ = 7 ('з')
Блок 3: найчаста буква 'р', ключ = 2 ('в')
Блок 4: найчаста буква 'ю', ключ = 16 ('р')
Блок 5: найчаста буква 'о', ключ = 0 ('а')
Блок 6: найчаста буква 'з', ключ = 25 ('щ')
Блок 7: найчаста буква 'у', ключ = 5 ('е')
Блок 8: найчаста буква 'ы', ключ = 13 ('н')
Блок 9: найчаста буква 'ц', ключ = 8 ('и')
Блок 10: найчаста буква 'у', ключ = 5 ('е')
Блок 11: найчаста буква 'т', ключ = 4 ('д')
Блок 12: найчаста буква 'ф', ключ = 6 ('ж')
Блок 13: найчаста буква 'щ', ключ = 11 ('л')
Блок 14: найчаста буква 'ы', ключ = 13 ('н')
Блок 15: найчаста буква 'т', ключ = 13 ('н')
Блок 16: найчаста буква 'о', ключ = 0 ('а')
```

Знайдений ключ: [2, 14, 7, 2, 16, 0, 25, 5, 13, 8, 5, 4, 6, 11, 13, 13, 0]

У символьному вигляді: "возвращениеджлнна"

Ключ вийшов майже змістовною фразою російською мовою. Звісно я спочатку подумав, що це "возвращение джедая", але вийшов у результат нелогічний тектс. Потім просто пошукав в Інтернеті відповідний ключ і виявляється існує така книга: "возвращение джинна".

Використовуючи знайдений ключ, виконали розшифрування:

```
=== РЕЗУЛЬТАТ ===

Знайд. період: 17
Знайд. ключ (числа): [2, 14, 7, 2, 16, 0, 25, 5, 13, 8, 5, 4, 6, 8, 13, 13, 0]
Знайд. ключ (текст): возвращениеджинна

=== РОЗШИФР. ТЕКСТ ===

дорофейльвовичпивторыкобылыниразувжизнинепокидалземлихотяпрожилужебольшешестидес
ятилетработалпрорабомстроительнойкомпаниидомостройвхарьковестолицевкраинылюбилпо
рыбачитьсдрузьяминаозерахроганьскогокраязачертойгородавыращивалнадачномучасткеов
ощифруктывоспитывалвнуковавотуезжатьзапределыроднойвкраинынелюбилнесмотрянавозм
ожностивсвязиссозданиемглобальнойсетиметропобыватьналюбойпланетесолнечнойсистемы
идажезаеепределамичтоподвиглоегосогласитьсянаэкскурсиюполунеонисамневсостоянииобы
лответитьвероятносыгралисвоюрольрассказыдрузейхваставшихсясвоимипутешествиямииун
еговзыгралолюбопытствопосмотретьвблизичтожеэтотакоеспутницаземлиокоторойтакмного
говорятдетивнукиидрузьякакбытонибылоаутромдвадцатьтретьегодекабряаккуратвначалос
вятокдорофейльвовичвтайнеотродныхиблизкихпозвонилвбюроэкскурсийсолнечнойсистемыз
апинаясьобъяснилчегохочетивтотжеденьспомощьюметродобралсядоаполлонтаунагороданал
унеоткудадолжнабыланачатьсяэкскурсияпосамымкрасивымизагадочнымместамспутницыземл
иаполлонтаунрасполагалсянаравнинеморяспокойствиянедалекоотзнаменитойбороздымаске
лайнпохожейнаиизвилистоеруслорекиименноздесьогдатовконцедвадцатоговекасовершилп
осадкуамериканскийпилотируемыйкорабльаполлонодиннадцатьаточнееегопосадочныймодул
ьестественноэкскурсантамзанимавшимкабинудвадцатиместногоэкскурсионногофлайтаснач
алапоказалипамятникаполлонуодиннадцатьпирамидуизлунногобазальтаспосадочнойплатфо
рмойиамериканскимфлагомазатемфлайтотправилсявпутешествиепоморюспокойствиязалитом
уяркимсолнечнымсветомэкскурсантамиоказалисьмолодыелюдиввозрастеотвосемнадцатидод
вадцатилетпоэтомупоначалудорофейльвовиччувствовалсебяневсвоейтарелкесмущаясьподл
юбопытнымивзглядамиспутниковнопотомегозахватиласуроваякрасоталунныхпейзажейионпе
ресталобращатьвниманиенавеселящуюсякомпаниюжадноразглядываяпроплывающиеподднищем
флайтациркиэскарпыкратерыживописныегруппыскалмореспокойствияполучилосвоеназвани
енеслучайноегоровнаясглаженнаяповерхностьтипичнадляобширныхморейнадневнойстороне
луныредкорадуетнаблюдателейпроявлениемвулканическойдеятельностиоднакоиздесьимел
осьнемалоинтересныхместиобъектовкоторыедесяткилетволновалиастрономовизучающихспу
тницуземлизагадочнаяцепочкакратеровподназваниемтеннисная ракеткаоколодвухдесятков
ямокдиаметромотпятидесятидостаметровпротянулисьудивительноровнойлинией заканчивая
ськратеромпобольшедиаметромоколошестисотметроввпечатлениескладываетсятакоебудтоп
олунной поверхностидействительнопрокатилсяподпрыгиваятеннисныймячоставиввпылицепо
чкуследовсовиныймосткаменная аркачерезбороздумаскелайндлинойколотрехкилометровиз
умительноровнаястенаобрывадлинойколотридцатикилометромбудтоктотоотхватилножомку
соклунной поверхностиивыбросилвкосмосоставивсрезиложбинуглубинойвкилометрбороздаз
олотойручейсамоенастоящееруслорекиширинойвполторакилометраидлинойвполторастасвер
кающееподлучамисолнцакристалликамипиритацветочная клумбавозвышениерыхлойпородыора
нжевогоцветадиаметромоколодвухкилометровивысотойвдвестиметровдействительноклумба
еслипосмотретьсверхустоунхенджгруппаскалсплоскимивершинамисоединенныхповерхудост
аточноровнымиплитамипрактическинеотличаетсяотземногомегалитическогокомплексаванг
лииинаконецбороздамаскелайндлинойколочетырехсоткилометровтакжездоровопохожаянар
услорекиширинойокилометрадотрехкакобъяснилгидбороздынасамомделепредставляетсобо
йсдвиговыйразломлуннойкорыслучившийсядесяткимиллионовлетназадврезультатеподвижки
щитаотудараметеоританосверхубороздавсеравнонапоминаетрекуидорофейльвовичдажепред
ставилкакпоруслутечетводаостанавливалисьивыходилиизфлайтаодетыевпузыривакуумплот
ныхспецкостюмовнесколькоразвкабинеаппаратаподдерживаласьнормальнаясилатяжестипоч
тиземнаяавнееецарилолунноетяготениевшестьразслабееземногопоэтомунеобошлосьбезкур
```

Розшифрований текст легко гуглиться. І далі розраховуємо оцінку якості розшифрування (квадратична різниця частот):

```
Оцінка якості розшифрування: 0.000371
(менше => краще)
```

Підсумок та висновки

Індекс відповідності дійсно є ефективним інструментом для аналізу поліалфавітних шифрів:
для відкритого тексту IoC = 0.0569 (близько до теоретичного 0.0529)
для коротких ключів (r=2,3) індекс залишається високим (0.0455, 0.0389)
для довгих ключів (r=10,16) індекс наближається до 1/m ≈ 0.0313
При обмеженому обсязі тексту (2.7 КБ) спостерігали статистичні похибки для великих значень r.

Виконали криптоаналіз ШТ варіанту 6:
визначили період шифру r = 17 (обидва методи дали однаковий результат)
знайдено ключ "возвращениеджинна" (змістовна фраза, яка є назвою книги)
розшифрували текст, який є відповідно текстом книги.

Період визначали через:
Метод1 - індекс відповідності блоків ефективний для періодів від 2 до 20-25
Метод2 - статистики співпадінь дає чіткі локальні максимуми на значеннях, кратних періоду
Комбінація обох методів підвищує надійність визначення періоду.

Щодо частотного аналізу, то ми підтвердили, що шифр Віженера зберігає статистичні властивості мови всередині кожного блоку, що робить його вразливим до криптоаналізу при відомій довжині ключа. Загалом можна сказати, що для надійного шифрування довжина ключа має бути співрозмірною з довжиною повідомлення. Використання змістовних ключів полегшує дешифрування після визначення початкових символів (бо в нас різні версії розшифрувальників давали помилкові 1-3 літери ключа, які можна було і не вгадати, якби ключ не був логічним висловом).
Обсяг тексту критично важливий для статистичного аналізу (мінімум 3-5 КБ для r > 10)