Баєсівський підхід в криптоаналізі: побудова і дослідження детерміністичної та стохастичної вирішуючих функцій

Варіант 4

Мета роботи

Ознайомлення з принципами баєсівського підходу в криптоаналізі: побудова вирішуючих функцій для моделей схем шифрування та криптоаналіз моделей шифрів за допомогою програмної реалізації; порівняння детерміністичної та стохастичної вирішуючих функцій.

Задача

Маючи відомі розподіли відкритих текстів $\Pr\{M\}$ та ключів $\Pr\{K\}$ задані в файлі "prob_04.csv", а також правило шифрування $\mathrm{Enc}: K \times M \longrightarrow C$ задане таблицею в файлі "table_04.csv", побудувати оптимальні:

1. Детерміністичну вирішуючу функцію;
2. Стохастичну вирішуючу функцію.

Обчислити та порівняти середнє значення втрати для побудованих вирішуючих функцій.

Хід роботи

Основні кроки для вирішення поставленої задачі:

1. Із заданих розподілів та функції шифрування обчислити сумісний розподіл відкритих текстів та шифротекстів $\Pr\{M, C\}$.
2. Маючи розподіл $\Pr\{M, C\}$, обчислити оптимальні вирішуючі функції $\delta_{\text{stochastic}}$ та $\delta_{\text{deterministic}}$.
3. Для побудованих функцій обчислити середнє значення втрат $l(\delta)$.

Код та результати

Початкові дані. Зчитуємо таблиці із файлів та виведемо отримані розподіли і таблицю.

Розподіл на множині ключів:

```
Pr{K=1} =    0.050
Pr{K=2} =    0.050
Pr{K=3} =    0.050
Pr{K=4} =    0.050
Pr{K=5} =    0.050
Pr{K=6} =    0.050
Pr{K=7} =    0.050
Pr{K=8} =    0.050
Pr{K=9} =    0.050
Pr{K=10} =    0.050
Pr{K=11} =    0.050
Pr{K=12} =    0.050
```

```
Pr{K=13} =    0.050
Pr{K=14} =    0.050
Pr{K=15} =    0.050
Pr{K=16} =    0.050
Pr{K=17} =    0.050
Pr{K=18} =    0.050
Pr{K=19} =    0.050
Pr{K=20} =    0.050
```

На множині відкритих текстів:

```
Pr{M=1} =    0.090
Pr{M=2} =    0.090
Pr{M=3} =    0.090
Pr{M=4} =    0.090
Pr{M=5} =    0.040
Pr{M=6} =    0.040
Pr{M=7} =    0.040
Pr{M=8} =    0.040
Pr{M=9} =    0.040
Pr{M=10} =    0.040
Pr{M=11} =    0.040
Pr{M=12} =    0.040
Pr{M=13} =    0.040
Pr{M=14} =    0.040
Pr{M=15} =    0.040
Pr{M=16} =    0.040
Pr{M=17} =    0.040
Pr{M=18} =    0.040
Pr{M=19} =    0.040
Pr{M=20} =    0.040
```

А також таблицю шифрування:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	12	18	20	13	7	3	16	14	15	2	19	6	1	17	9	11	8	5	10
2	18	6	7	16	13	11	9	17	5	1	8	3	4	19	12	2	14	20	10	15
3	14	16	9	7	8	12	19	3	11	15	18	5	2	20	13	17	4	1	6	10
4	16	8	11	20	6	1	15	4	12	7	13	10	18	19	3	14	5	17	2	9
5	10	5	2	8	7	3	15	1	14	9	6	4	20	19	12	13	11	17	16	18
6	18	3	17	8	16	7	2	19	11	4	10	5	20	6	12	13	15	9	14	1
7	14	11	20	5	19	3	8	10	16	17	4	2	18	12	7	13	9	1	6	15
8	12	5	2	17	10	14	20	9	19	6	16	11	4	8	1	18	15	3	7	13
9	17	20	7	11	1	16	8	14	5	3	15	4	18	6	19	12	2	10	9	13
10	2	13	15	16	3	20	1	4	8	7	9	19	5	12	11	6	10	14	17	18
11	9	3	7	6	19	17	10	2	20	18	16	8	1	11	5	4	15	14	13	12
12	12	3	14	8	19	9	10	2	15	13	5	18	11	16	17	7	1	4	20	6
13	17	4	7	19	12	9	18	16	14	11	6	3	2	8	13	20	5	10	1	15
14	13	5	1	9	14	11	10	7	18	4	2	15	17	16	6	20	19	12	8	3
15	13	11	18	3	7	17	4	16	10	14	5	2	20	15	12	1	9	19	8	6
16	20	17	11	8	9	15	16	2	7	14	6	13	5	12	18	4	19	1	3	10
17	13	5	14	18	11	10	1	16	3	20	17	2	8	6	12	9	7	15	4	19
18	16	9	4	13	2	15	11	12	18	17	10	5	6	7	1	8	20	14	3	19
19	12	4	13	8	2	3	10	15	5	18	7	20	9	11	17	19	1	6	14	16
20	3	10	19	16	1	7	17	14	13	5	20	11	12	2	8	18	4	9	6	15

Сумісний розподіл. Із теоретичних викладок, наведених в документі лабораторної, можна сказати, що побудова оптимальної вирішуючої функції (як детерміністичної, так і стохастичної) зводиться до пошуку максимально ймовірних відкритих текстів для кожного шифротексту, тобто:

$$\forall\, C: \quad \text{знайти такі } M_{max} \text{ що } \; M_{max} = \operatorname{argmax}_{m \in M} \Pr\{m \mid C\}$$

Зауваження. Замість умовного розподілу $\Pr\{M \mid C\}$ можна розглядати сумісний – $\Pr\{M,\ C\}$, оскільки

$$\Pr\{M \mid C\} = \frac{\Pr\{M,\ C\}}{\Pr\{C\}},$$

і домножання кожного рядку на $\Pr\{C\}$ не змінює максимумів.

Сумісний розподіл обчислимо як:

$$\sum_{K:\ \mathrm{Enc}(K,M)=C} \Pr\{M\}\Pr\{K\}$$

Відповідно матриця сумісного розподілу M та C має вигляд:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	0	0.0045	0	0.0040	0.0020	0.0040	0.0020	0	0.0020	0	0	0.0020	0.0020	0.0040	0.0020	0.0040	0.0060	0.0020	0.0020
2	0.0045	0	0.0090	0	0.0040	0	0.0020	0.0060	0	0	0.0040	0.0060	0.0040	0.0020	0	0.0020	0.0020	0	0.0020	0
3	0.0045	0.0135	0	0.0045	0.0020	0.0060	0.0020	0.0020	0.0020	0.0020	0	0.0040	0	0	0.0020	0	0	0.0020	0.0040	0.0020
4	0.0045	0.0090	0.0045	0	0	0	0.0020	0.0040	0	0.0040	0.0020	0.0040	0.0040	0	0	0.0040	0.0040	0.0020	0.0020	0
5	0	0.0180	0	0.0045	0	0	0	0	0.0060	0.0020	0.0040	0.0060	0.0040	0	0.0020	0	0.0040	0	0.0020	0
6	0	0.0045	0	0.0045	0.0020	0	0	0	0	0.0020	0.0060	0	0.0040	0.0060	0.0020	0.0020	0	0.0020	0.0060	0.0040
7	0	0	0.0180	0.0045	0.0040	0.0060	0	0.0020	0.0020	0.0040	0.0020	0	0	0.0020	0.0020	0.0020	0.0020	0	0.0020	0
8	0	0.0045	0	0.0225	0.0020	0	0.0040	0	0.0020	0	0.0020	0.0020	0.0020	0.0040	0.0020	0.0020	0	0.0020	0.0040	0
9	0.0045	0.0045	0.0045	0.0045	0.0020	0.0040	0.0020	0.0020	0	0.0020	0.0020	0	0.0020	0	0	0.0040	0.0040	0.0040	0.0020	0.0020
10	0.0045	0.0045	0	0	0.0020	0.0020	0.0080	0.0020	0.0020	0	0.0040	0.0020	0	0	0	0	0.0020	0.0040	0.0020	0.0060
11	0	0.0090	0.0090	0.0045	0.0020	0.0040	0.0020	0	0.0040	0.0020	0	0.0040	0.0020	0.0040	0.0020	0	0.0040	0	0	0
12	0.0135	0.0045	0	0	0.0020	0.0020	0	0.0020	0.0020	0	0	0	0.0020	0.0060	0.0100	0.0020	0	0.0020	0	0.0020
13	0.0135	0.0045	0.0045	0.0045	0.0040	0	0	0	0.0020	0.0020	0.0020	0.0020	0	0	0.0040	0.0060	0	0	0.0020	0.0040
14	0.0090	0	0.0090	0	0.0020	0.0020	0	0.0040	0.0060	0.0040	0	0	0	0	0	0.0020	0.0020	0.0060	0.0040	0
15	0	0	0.0045	0	0	0.0040	0.0040	0.0020	0.0020	0.0040	0.0020	0.0020	0	0.0020	0	0	0.0060	0.0020	0	0.0080
16	0.0090	0.0045	0	0.0135	0.0020	0.0020	0.0020	0.0080	0.0020	0	0.0040	0	0	0.0040	0	0	0	0	0.0020	0.0020
17	0.0090	0.0045	0.0045	0.0045	0	0.0040	0.0020	0.0020	0	0.0040	0.0020	0	0.0020	0	0.0060	0.0020	0	0.0040	0.0020	0
18	0.0090	0	0.0090	0.0045	0	0	0.0020	0	0.0040	0.0040	0.0020	0.0020	0.0060	0	0.0020	0.0040	0	0	0	0.0040
19	0	0	0.0045	0.0045	0.0060	0	0.0020	0.0020	0.0020	0	0	0.0040	0	0.0060	0.0020	0.0020	0.0040	0.0020	0	0.0040
20	0.0045	0.0045	0.0045	0.0090	0	0.0020	0.0020	0	0.0020	0.0020	0.0020	0.0020	0.0060	0.0020	0	0.0040	0.0020	0.0020	0.0020	0

де жовтим позначені максимальні значення в кожному рядку (тобто максимально ймовірні відкиті тексти для кожного шифротексту).

Побудова вирішуючих функцій. Оптимальна *детерміністична* вирішуюча функція є баєсівською, тобто:

$$\Pr\{\delta_B(C) \mid C\} \ = \ \max_{m \in M} \Pr\{m \mid C\} \ = \ \max_{m \in M} \Pr\{m,\ C\}.$$

Отже, маючи таблицю сумісного розподілу, отриману вище, побудуємо оптимальну детерміністичну вирішуючу функцію таким чином:

$$\delta_D^{\mathrm{opt}}(C) \ = M_{max}\ ,$$

де M_{max} – це довільний фіксований шифротекст, що відповідає максимуму у відповідному рядку таблиці.

Оптимальна *стохастична* вирішуюча функція будується за схожим принципом, але (only god knows why) визначається як матриця. Згідно з твердженням 3. побудуємо оптимальну стохастичну вирішуючу функцію як:

$$\delta_D^{\text{opt}}(C,\, m) = \begin{cases} 1/d & \text{якщо } \Pr\{m \mid C\} = \max_{m \in M} \Pr\{m \mid C\} \\ 0 & \text{інакше} \end{cases},$$

де $d = \#\{\, m : \Pr\{m \mid C\} = \max_{m \in M} \Pr\{m \mid C\} \,\}$ (тобто кількість максимально ймовірних відкритих текстів для шифротексту C.

Оптимальна детерміністична вирішуюча функція:

```
1       ->    18
2       ->    3
3       ->    2
4       ->    2
5       ->    2
6       ->    11
7       ->    3
8       ->    4
9       ->    1
10       ->    7
11       ->    2
12       ->    1
13       ->    1
14       ->    1
15       ->    20
16       ->    4
17       ->    1
18       ->    1
19       ->    5
20       ->    4
```

Середнє значення втрат для побудованої функції:

```
0.7860
```

Оптимальна стохастична вирішуюча функція:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0	0
2	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0.3333	0	0	0.3333	0	0	0	0	0.3333	0
7	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0.2500	0.2500	0.2500	0.2500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0.5000	0.5000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0.5000	0	0.5000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000
16	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0.5000	0	0.5000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0.5000	0	0	0	0	0	0	0	0	0.5000	0	0	0	0	0	0
20	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Та відповідні середні втрати:

```
0.7860
```

Труднощі життя

Основними труднощами, з якими ми стикнулись під час виконання лабораторної роботи, були:

- Придумати гарну назву для файлу
- Складність розуміння теоретичного підгрунття лабораторної роботи
- Неможливість зрозуміти, де в таблиці відкриті тексти, а де ключі
- Виведення гарних таблиць в матлабі

Висновки

Середні значення втрат для обох функцій вийшли однаковими, тому не дуже зрозуміло, для чого ми тут взагалі зібрались. Але стохастична функція дозволяє покрити більше потенційних відкритих текстів, що відповідають отриманому шифротексту, під час вибору, тому з практичної точки зору дає більше можливостей. В той же час, для певних випадків, стохастична вирішуюча функція може бути складніша для зберігання в пам'яті (якщо її матриця не sparse).