КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1
З КУРСУ
МЕТОДИ КРИПТОАНАЛІЗУ 1

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

1 Мета роботи

Ознайомлення з принципами баєсівського підходу в криптоаналізі, побудова детерміністичної та стохастичної вирішуючих функцій для моделей схем шифрування та криптоаналіз моделей шифрів за допомогою програмної реалізації, зокрема здійснення порівняльного аналізу вирішуючих функцій.

2 Необхідні теоретичні відомості

2.1 Модель криптосистеми та означення шифру

Означення 1. *Модель криптосистеми* визначається як кортеж

$$\Sigma = \langle \mathcal{M},\ \mathcal{K},\ \mathcal{C},\ \mathcal{E},\ \mathcal{D} \rangle, \text{ де:}$$

- $\mathcal{M}$ — простір усіх можливих відкритих текстів (повідомлень, ВТ);
- $\mathcal{K}$ — простір особистих ключів;
- $\mathcal{C}$ — простір шифротекстів (криптограм, ШТ);
- $\mathcal{E}$ — простір алгоритмів шифрування:

$$\mathcal{E} = \{E_k,\ k \in \mathcal{K}\},$$

де E_k — конкретний алгоритм шифрування з ключем k;

- $\mathcal{D}$ — простір алгоритмів розшифрування:

$$\mathcal{D} = \{D_k,\ k \in \mathcal{K}\},$$

де D_k — конкретний алгоритм розшифрування з ключем k.

Нехай $M = (m_1, \ldots, m_n)$ — відкритий текст з простору $\mathcal{M}$, де символи m_i для $i = \overline{1,n}$ належать деякому алфавіту Z_m, що містить m букв, та $C = (c_1, \ldots, c_n)$ — шифротекст з простору $\mathcal{C}$, де c_j для $j = \overline{1,n}$ належать тому ж алфавіту Z_m.

Означення 2. *Шифром* називається відображення $\mathcal{M} \times \mathcal{K} \to \mathcal{C}$ таке, що виконується така умова:

$$\forall\, k \in \mathcal{K}, M \in \mathcal{M} :\ D_k(E_k(M)) = M,$$

де $E_k : M \to C$ — ін'єктивне відображення.

Нехай на множині $\mathcal{M} \times \mathcal{K}$ задано ймовірнісний розподіл, тобто для будь-якої пари (M, k) задана ймовірність $P(M, k)$ така, що виконується:

$$\sum_{M \in \mathcal{M}} \sum_{k \in \mathcal{K}} P(M,k) = 1;$$

$$\sum_{M \in \mathcal{M}} P(M,k) = P(k);$$

$$\sum_{k \in \mathcal{K}} P(M,k) = P(M).$$

Будемо вважати, що відкритий текст та ключ, що генерується, незалежні (як правило, на практиці так і є), тобто виконується:

$$P(M,k) = P(M) \cdot P(k).$$

Розподіл $P(M, k)$ індукує розподіли ймовірностей на інших просторах криптосистеми:

$$\forall C :\ P(C) = \sum_{(M,k):E_k(M)=C} P(M,k);$$

$$\forall (M,C) :\ P(M,C) = \sum_{k:E_k(M)=C} P(M,k).$$

Всі ці ймовірності відомі криптоаналітику (задані або можуть бути обчислені).

В даному комп'ютерному практикумі будуть розглядатись модельні відкриті тексти та шифротексти, що не є послідовностями літер з алфавіту Z_m. Тому простори $\mathcal{M}$ та $\mathcal{C}$ розглядаються як абстрактні множини, що містять відкриті тексти та шифротексти відповідно.

2.2 Детерміністична вирішуюча функція

Означення 3. *Детерміністична вирішуюча функція* — це правило (алгоритм, процедура), за допомогою якого криптоаналітик по шифротексту однозначно обирає відповідний йому відкритий текст (при цьому він може помилитись). *Задача детерміністичної вирішуючої функції* — по заданому шифротексту C без знання ключа k дати відповідь *«шифротекст C отримано внаслідок шифрування відкритого тексту M»*.

На рисунку 1 схематично зображено використання детерміністичної функції (істинний відкритий текст позначається M_0).

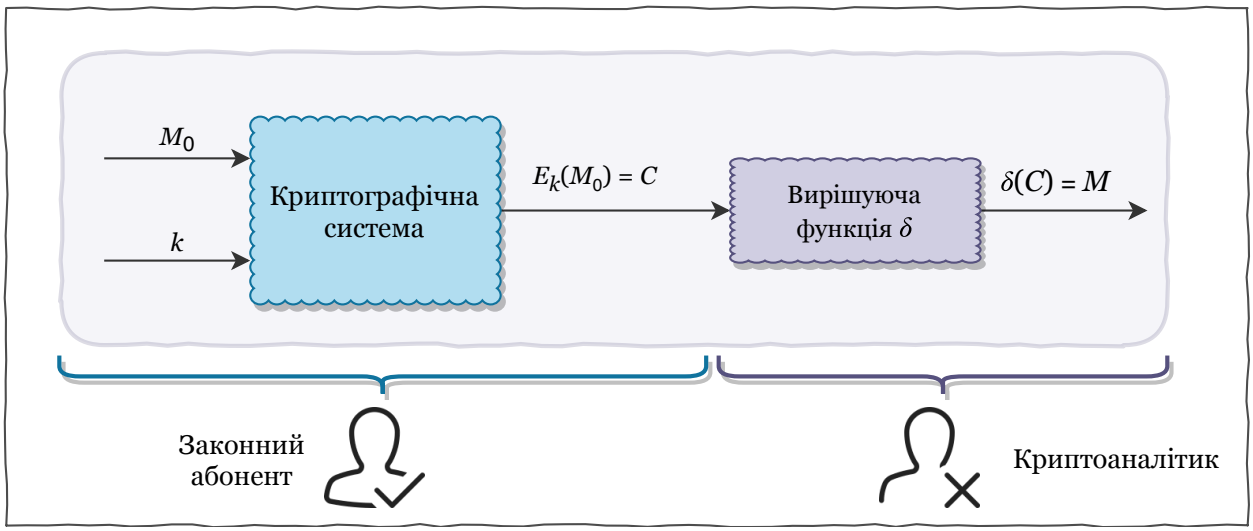

Рис. 1: Схематичне представлення баєсівського підходу

Розглянемо формальне визначення детерміністичної вирішуючої функції.

Означення 4. *Детерміністичною вирішуючою функцією* називається послідовність відображень:

$$\delta_D = \{\delta_D^{(n)} : \ Z_m^n \to Z_m^n, \ n \in \mathbb{N}\},$$

де Z_m — алфавіт шифротекстів та відкритих текстів, n — довжина шифротексту, що був перехоплений криптоаналітиком.

Зауваження. Оскільки в даному комп'ютерному практикумі розглядаються модельні простори $\mathcal{M}$ та $\mathcal{C}$, то детерміністичною вирішуючою функцією буде називатись таке відображення:

$$\delta_D : \mathcal{C} \to \mathcal{M}.$$

Байєсівське прийняття рішення за допомогою детерміністичної вирішуючої функції

1. На вхід вирішуючої функції δ_D приходить шифротекст C.
2. Обчислюється $\delta_D(C) = M$ та приймається рішення, що шифротекст C було отримано внаслідок шифрування відкритого тексту M.

Очевидно, що якщо існують декілька відкритих текстів $M_1, \ldots, M_r$ таких, що:

$$\exists \, k_1, \ldots, k_r : \ C = E_{k_i}(M_i), \ i = \overline{1, r},$$

то детерміністична вирішуюча функція на запит з шифротекстом C видасть один з відкритих текстів $M_1, \ldots, M_r$. В такому випадку лише при одному з ключів $k_1, \ldots, k_r$ відповідь детерміністичної вирішуючої функції буде правильна.

Означення 5. *Функцією втрат* для δ_D називається таке відображення:

$$L_{\delta_D} : \mathcal{M} \times \mathcal{C} \to \{0, 1\},$$

$$L_{\delta_D}(M, C) = \begin{cases} 1, & \text{якщо } \delta_D(C) \neq M_0; \\ 0, & \text{якщо } \delta_D(C) = M_0, \end{cases}$$

де M_0 — істинний відкритий текст.

Означення 6. *Середні втрати детерміністичної вирішуючої функції* δ_D позначаються l_{δ_D} та обчислюються за таким співвідношенням:

$$l_{\delta_D} = \sum_{M \in \mathcal{M}} \sum_{C \in \mathcal{C}} P(M, C) \cdot L_{\delta_D}(M, C).$$

Означення 7. Детерміністична вирішуюча функція δ_D^* називається *оптимальною*, якщо виконується така умова:

$$\forall \delta_D \in \Delta_D : \; l_{\delta_D^*} \leq l_{\delta_D},$$

де Δ_D — клас усіх детерміністичних вирішуючих функцій.

Означення 8. *Баєсівською* вирішуючою функцією називається така детерміністична вирішуюча функція δ_B, для якої виконується умова:

$$P(\delta_B|C) = \max_{M \in \mathcal{M}} P(M|C).$$

Твердження 1. Детерміністична вирішуюча функція є оптимальною тоді і лише тоді, коли вона є баєсівською.

Доведення. Розглянемо середні втрати деякої детерміністичної вирішуючої функції $\delta_0 \in \Delta_D$:

$$l_{\delta_0} = \sum_{M \in \mathcal{M}} \sum_{C \in \mathcal{C}} P(M, C) \cdot L_{\delta_0}(M, C) = \sum_{C \in \mathcal{C}} P(C) \sum_{M \in \mathcal{M}} P(M|C) \cdot L_{\delta_0}(M, C) = \sum_{C \in \mathcal{C}} P(C)(1 - P(\delta_0(C)|C)).$$

Таким чином, значення l_{δ_0} досягає мінімуму тоді та лише тоді, коли ймовірність $P(\delta_0(C)|C))$ максимальна, тобто для таких M, для яких ймовірність $P(M|C)$ досягає максимуму, причому $M = \delta_0(C)$. А це і є баєсівською вирішуючою функцією за визначенням.

□

Таким чином, твердження 1 визначає практичний спосіб побудови оптимальних вирішуючих функцій: для побудови оптимальної вирішуючої функції необхідно та достатньо побудувати баєсівську детерміністичну вирішуючу функцію.

2.3 Стохастична вирішуюча функція

Означення 9. *Стохастичною вирішуючою функцією* називається послідовність $m^n \times m^n$-матриць:

$$\delta_S = \left\{ \delta_S^{(n)} : \left|\left| \delta_S^{(n)}(C, M) \right|\right|_1^{m^n}, \; n \in \mathbb{N} \right\},$$

де $\delta_S^{(n)}(C, M) \geq 0$ та для усіх $C \in \mathcal{C}$ виконується:

$$\sum_{M \in \mathcal{M}} \delta_S^{(n)}(C, M) = 1.$$

Таким чином, стохастична вирішуюча функція при фіксованому значенні n — це $m^n \times m^n$-матриця такого вигляду:

$$\delta_S^n = \left|\left| \delta_S^{(n)}(C, M) \right|\right|_1^{m^n}.$$

Для модельних $\mathcal{C}$ та $\mathcal{M}$ стохастична вирішуюча функція є матрицею розміру $\mathcal{C} \times \mathcal{M}$. Дана матриця є стохастичною по рядках. Детерміністична вирішуюча функція є частковим випадком стохастичної вирішуючої функції, у якої в кожному рядку стоїть рівно одна одиниця, а на всіх інших місцях стоїть нуль.

	M_1	M_2	...	M_l	...	$M_{\lvert\mathcal{M}\rvert}$
C_1	$\delta_S(C_1, M_1)$	$\delta_S(C_1, M_2)$	...	$\delta_S(C_1, M_l)$	...	$\delta_S(C_1, M_{\lvert\mathcal{M}\rvert})$
C_2	$\delta_S(C_2, M_1)$	$\delta_S(C_2, M_2)$	...	$\delta_S(C_2, M_l)$	...	$\delta_S(C_2, M_{\lvert\mathcal{M}\rvert})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_k	$\delta_S(C_k, M_1)$	$\delta_S(C_k, M_2)$	...	$\delta_S(C_k, M_l)$	...	$\delta_S(C_k, M_{\lvert\mathcal{M}\rvert})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_{\lvert\mathcal{C}\rvert}$	$\delta_S(C_{\lvert\mathcal{C}\rvert}, M_1)$	$\delta_S(C_{\lvert\mathcal{C}\rvert}, M_2)$	...	$\delta_S(C_{\lvert\mathcal{C}\rvert}, M_l)$	...	$\delta_S(C_{\lvert\mathcal{C}\rvert}, M_{\lvert\mathcal{M}\rvert})$

Табл. 1: Матричне представлення стохастичної функції

Баєсівське прийняття рішення за допомогою статистичної вирішуючої функції

1. За шифротекстом C_k криптоаналітик обирає рядок k у побудованій стохастичній вирішуючій функції δ_S.
2. Криптоаналітик реалізовує випадкову величину з розподілом $p_i = \delta_S(C_k, M_i)$, $i = \overline{1, |\mathcal{M}|}$.
3. Якщо випадкова величина приймає значення M_j, то приймається рішення, що був зашифрований відкритий текст M_j.

Означення 10. *Функцією втрат* для δ_S називається відображення $L_{\delta_S} : \mathcal{M} \times \mathcal{C} \to [0, 1]$:

$$L_{\delta_S}(M, C) = \sum_{M' \in \mathcal{M},\ M' \neq M} \delta_S(C, M').$$

Означення 11. *Середні втрати стохастичної вирішуючої функції* δ_S позначаються l_{δ_S} та обчислюються за таким співвідношенням:

$$l_{\delta_S} = \sum_{M \in \mathcal{M}} \sum_{C \in \mathcal{C}} P(M, C) \cdot L_{\delta_S}(M, C).$$

Означення 12. Стохастична вирішуюча функція δ_S^* називається *оптимальною*, якщо виконується:

$$\forall \delta_S \in \Delta_S :\ l_{\delta_S^*} \leq l_{\delta_S},$$

де Δ_S — клас усіх детерміністичних вирішуючих функцій.

Твердження 2. Стохастична вирішуюча функція є оптимальною тоді та лише тоді, коли $\delta_S(C, M)$ при $P(C) > 0$ максимізує функцію:

$$f(C) = \sum_{M \in \mathcal{M}} P(M|C) \cdot \delta_S(C, M).$$

Доведення. Розглянемо середні втрати деякої стохастичної вирішуючої функції $\delta_0 \in \Delta_S$:

$$l_{\delta_0} = \sum_{M \in \mathcal{M}} \sum_{C \in \mathcal{C}} P(M, C) \cdot L_{\delta_0}(M, C) = \sum_{C \in \mathcal{C}} P(C) \sum_{M \in \mathcal{M}} P(M|C) \cdot L_{\delta_0}(M, C) =$$

$$= \sum_{C \in \mathcal{C}} P(C) \sum_{M \in \mathcal{M}} P(M|C) \sum_{M' \in \mathcal{M},\ M' \neq M} \delta_0(C, M') = \sum_{C \in \mathcal{C}} P(C) \left(1 - \sum_{M \in \mathcal{M}} P(M|C) \cdot \delta_0(C, M)\right).$$

Таким чином, l_{δ_0} досягає мінімуму тоді та лише тоді, коли значення $\sum_{M \in \mathcal{M}} P(M|C) \cdot \delta_0(C, M)$ досягає максимуму, якщо $P(C) > 0$, що і треба було довести.

□

Твердження 3. Стохастична вирішуюча функція є оптимальною тоді та лише тоді, коли виконується така умова: якщо $\delta_S(C, M) > 0$, то $P(M|C) = \max_{M' \in \mathcal{M}} P(M'|C)$.

Доведення. Доведення цього твердження випливає з твердження 2: для максимізації $f(C)$ необхідно щоб при кожному ненульовому значенні $\delta_S(C, M)$ ймовірність $P(M|C)$ була максимальною.

□

3 Дані для аналізу

В даному комп'ютерному практикумі пропонується модель криптосистеми, для якої

$$|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}| = 20.$$

Вихідні дані до комп'ютерного практикуму подані в форматі `.csv` у вигляді двох таблиць: `prob_XX.csv` та `table_XX.csv`, де `XX` — номер варіанта. Таблиця у файлі `prob_XX.csv` складається з двох рядків та 20 стовпців і містить ймовірнісний розподіл відкритих текстів в першому рядку та ймовірнісний розподіл ключів в другому. Файл `table_XX.csv` містить таблицю шифрування 20×20. Вона представлена таким чином: по стовпцях індексується відкритий текст, по рядках ключ; на перетині i-го рядка і j-го стовпчика міститься індекс h шифротексту, який отриманий в результаті шифрування j-го відкритого тексту на i-му ключі, тобто:

$$C_h = E_{k_i}(M_j).$$

4 Порядок виконання роботи і методичні вказівки

-1. Ознайомитись з порядком виконання комп'ютерного практикуму та відповідними вимогами до виконання роботи.

0. Уважно прочитати необхідні теоретичні відомості до комп'ютерного практикуму.

1. Для заданого варіанта моделі шифру описати алгоритм побудови детерміністичної та стохастичної вирішуючих функцій.

2. Створити репозиторій в системі контролю версій `Git` (бажано використовувати вебсервіс `GitHub` *). Важливо:

   (а) репозиторій створюється перед початком роботи над програмним кодом (якщо репозиторій приватний, то перед початком роботи має бути надано доступ викладачу до даного репозиторію);

   (б) весь процес створення програмного коду має бути відображений у відповідних комітах проекту (для кожної атомарної зміни коду має бути власний коміт);

   (в) програмна реалізація не допускається до захисту при недотриманні вищевизначених вимог.

3. Реалізувати алгоритми програмно і подати результати побудови детерміністичної та стохастичної вирішуючих функцій у вигляді таблиць. Для цього необхідно:

   (а) порахувати розподіли $P(C)$ та $P(M, C)$;

   (б) ґрунтуючись на цих розподілах обчислити $P(M|C)$;

   (в) побудова оптимальних детерміністичної та стохастичної вирішуючих функцій зводиться до максимізації $P(M|C)$.

4. Обчислити середні втрати, провести порівняльний аналіз вирішуючих функцій.

5. Оформити звіт до комп'ютерного практикуму.

Комп'ютерний практикум виконується або індивідуально, або бригадою з двох студентів. Протягом 2 тижнів з початку навчального семестру студентам необхідно надати викладачеві інформацію про спосіб виконання роботи (індивідуальний/бригадний) та склад бригади. Зміна складу бригади та способу виконання роботи протягом семестру можлива лише при узгодженні цього з викладачем комп'ютерних практикумів.

*Використання інших сервісів необхідно попередньо узгодити з викладачем

5 Оформлення звіту

Звіт про виконання комп'ютерного практикуму оформлюється згідно зі стандартними правилами оформлення наукових робіт за допомогою системи набору і верстки LaTeX, причому дозволяється використовувати розмір шрифту 12pt та одинарний міжрядковий інтервал. Звіт обов'язково має містити:

- мету комп'ютерного практикуму;
- постановку задачі та варіант завдання;
- хід роботи;
- опис алгоритму побудови детерміністичної та стохастичної вирішуючих функцій;
- таблицю ймовірностей $P(M|C)$;
- знайдені детерміністичну та стохастичну функції у вигляді таблиць;
- середні втрати для вирішуючих функцій;
- опис труднощів, що виникали при виконанні комп'ютерного практикуму, та шляхи їх розв'язання;
- висновки.

Лістинги програми дозволяється не включати у звіт.

6 Порядок захисту комп'ютерного практикуму

Для зарахування комп'ютерного практикуму студенту необхідно виконати захист теоретичної та практичної частин роботи (за умови своєчасного надання доступу викладачеві до `Git`-репозиторію, що містить код програми). Студент має можливість здавати теоретичну та практичну частини комп'ютерного практикуму в різні дні в довільному порядку.

7 Контрольні питання

1. Класифікація атак по типу відомої інформації.
2. Математична модель системи шифрування. Цілком таємна криптосистема в моделі Шеннона.
3. Детерміністична вирішуюча функція.
4. Функція втрат детерміністичної вирішуючої функції, середні втрати детерміністичної вирішуючої функції.
5. Оптимальна детерміністична вирішуюча функція, байєсівська детерміністична вирішуюча функція.
6. Стохастична вирішуюча функція.
7. Втрати та середні втрати стохастичної вирішуючої функції, оптимальна стохастична вирішуюча функція.
8. Процедура прийняття рішення по заданій стохастичній вирішуючій функції.
9. Яка вирішуюча функція краща: детерміністична чи стохастична?

Оцінювання комп'ютерного практикуму

Можлива кількість рейтингових балів 12

- Програмна реалізація 5
- Теоретичний захист роботи 6
- Своєчасне виконання роботи 1

Несвоєчасне виконання роботи ..... -1 бал за кожен тиждень пропуску

Академічний плагіат ..... -10 балів до рейтингу з вимогою виконати комп'ютерний практикум повторно та без можливості складання іспиту на основній сесії