

Математичні методи захисту інформації: курс
лекцій. Частина 1

Завадська Л.О., Савчук М.М.

21 травня 2014 р.

Розділ 1

Задачі, напрямки та методи захисту інформації. Поняття про криптографічний захист інформації

1.1 Області застосування, мета, методи захисту інформації

Протягом свого існування людство пережило декілька інформаційних революцій: створення і розвиток мов, винахід писемності, винахід та широке застосування друкарства, створення комп'ютера та новітніх електронних технологій, що радикально змінили суспільство в усіх галузях, на всіх рівнях суспільного розвитку. Кількість інформації, що була доступна та використовувалась, постійно зростала, а за періоди інформаційних революцій — на декілька порядків. Володіння інформацією і в минулому і нині давало можливість досягти швидкого розвитку і успіху у різних галузях як у глобальному масштабі, так і в конкретних справах. Сьогодні світ переживає період, коли накопичено колосальний об'єм знань, що дозволяє перейти до здійснення справді революційних технологічних рішень. Основою розвитку нині може бути, перш за все, процес пізнання, і він посилюється тільки високоосвіченому суспільству, в якому праця приймає все більш інтелектуальні форми. Технологіям майбутнього потрібні широко освічені люди, які здатні орієнтуватися в нових умовах дійсності, що стрімко змінюється.

Однією з галузей, що найбільш динамічно змінюються протягом останніх десятиліть, є технології електронної обробки інформації, телекомунікацій, комп'ютерних мереж, технології захисту інформації.

В Україні широкий попит на методи і засоби захисту інформації почав виявлятися у другій половині 80-х років XX ст. З часом виникла нагальна потреба використання криптографічних та технічних методів захисту також у приватному секторі. Сьогодні велика кількість конфіденційної інформації передається в електронному вигляді, на електронних носіях, між

ЕОМ звичайними лініями зв'язку. Інформація може продаватися та купуватися, мати ціну, що незрівнянно перевищує ціну матеріального носія. Часто володіння інформацією дає переваги, ціну яких неможливо підрахувати, наприклад, у військовій справі. Термін збереження секретності інформації може коливатися від декількох годин до багатьох десятиліть. Тому вкрай потрібні спеціалісти, які володіють криптографічними, технічними, комплексними методами захисту, знають відповідні стандарти, здатні використовувати (або розробляти) програмне й апаратне забезпечення для гарантування таємності та цілісності конфіденційної інформації. Криптографічні методи захисту вважаються одними з найбільш надійних та ефективних.

- Області застосування захисту інформації (відповідно — види таємниці):

- | | |
|------------------|------------------------|
| 1. Військова; | 6. Промислова; |
| 2. Дипломатична; | 7. Наукова; |
| 3. Фінансова; | 8. Юридична; |
| 4. Банківська; | 9. Медична; |
| 5. Комерційна; | 10. Особиста таємниця. |

- Мета і головні задачі захисту інформації:

1. Конфіденційність (секретність) інформації;
2. Цілісність інформації;
3. Автентичність інформації;
4. Доступність інформації.

- Напрямки, аспекти, методи і засоби захисту інформації:

1. Юридичні, правові;
2. Методично-нормативні;
3. Організаційні;
4. Безпосередні (фізичні);
5. Технічні — захист від витоку по технічним каналам:
 - (a) електромагнітному
 - (b) оптичному
 - (c) акустичному
 - (d) віброакустичному;
6. Стеганографічні;
7. Криптографічні. Методи математичного захисту інформації;
8. Методи квантової криптографії;
9. Морально-етичні норми.

1.2 Перші поняття криптографічного захисту інформації

Визначення 1.1 (Криптографічний захист інформації). Криптографічний захист інформації — це різновид захисту інформації, який реалізується за допомогою криптографічних перетворень, спеціальних ключових даних з метою приховування та відновлення змісту інформації, підтвердження достовірності, авторства, запобігання несанкціонованому використанню тощо.

Визначення 1.2 (Криптографічне перетворення). Криптографічне перетворення — це перетворення інформації відповідно до певних правил (логічних, математичних) з метою забезпечення функціонування криптографічних протоколів.

Визначення 1.3 (Криптографічний ключ). Криптографічний ключ — це параметр, який використовується в криптографічному алгоритмі для вибору конкретного криптографічного перетворення; ключі можуть бути таємними або відкритими.

Визначення 1.4 (Криптографічний протокол). Криптографічний протокол — це послідовність узгоджених дій згідно з деякими правилами, у відповідності з якими відбувається обмін інформацією між сторонами або учасниками протоколу та її перетворення з використанням криптографічних методів і засобів. Простий приклад криптографічного протоколу — це зашифрування та розшифрування повідомлення.

Визначення 1.5 (Криптографія). Криптографія — науково-технічна дисципліна, яка вивчає принципи, методи і засоби криптографічного захисту інформації і інформаційних технологій, предметом якої є розробка криптографічних систем.

Визначення 1.6 (Криптоаналіз). Криптоаналіз — це науково-технічна дисципліна, яка вивчає методи, способи і засоби аналізу криптографічного захисту інформації: криптографічних систем, криптографічних алгоритмів, протоколів з метою знайти способи їх розкриття без знання секретних ключів і, можливо, будови криптосистем, знайти способи несанкціонованого доступу, підробки даних тощо. Криптоаналіз оцінює складність таких способів розкриття (злому) і стійкість криптографічного захисту інформації. Фахівець, який займається криптоаналізом будемо називати криптоаналітиком.

Визначення 1.7 (Криптологія). Криптологія за найбільш поширеною сучасною термінологією, об'єднує в собі дисципліни криптографію і криптоаналіз.

Зауваження 1. *Не всі країни дотримуються останньої термінології щодо дисциплін. Так, наприклад, в Росії назва „криптографія” об'єднує в собі власне криптографію (криптосинтез) у вище наведеному розумінні і криптоаналіз, а криптологія розглядається як галузь криптографії, що вивчає математичні моделі криптографічних систем, і також поділяється на криптосинтез та криптоаналіз.*

1.3 Етапи розвитку технологічних засобів криптографії

1. “Ручна” криптографія (із давнини до середини-кінця XIX століття)
Основні види шифрів — заміни і перестановки.

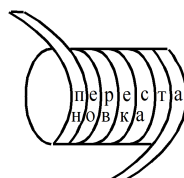


Рис. 1.1: Шифр Скитала

Перший шифр перестановки, застосування якого зафіксоване у військовій справі, (Спарта, V ст. до н.е.) — шифр Скитала (рис. 1.1). Таємний ключ — діаметр барабана.

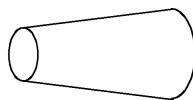


Рис. 1.2: Засіб криптоаналізу шифру Скитала

Для шифрування на стрічці, що намотувалась на барабан (скитал), писалось вздовж барабана повідомлення. Після знімання з барабану на стрічці була зовні випадкова послідовність літер — шифроване повідомлення. Криптоаналіз шифру Скитала запропонував Арістотель за допомогою барабана змінного діаметру (рис. 1.2): якщо на намотаній на нього стрічці з шифрованим повідомленням у деякому місці вгадувались якісь частини слів, то цьому місцю відповідав діаметр справжнього барабану.

Прикладом шифру заміни є шифр Цезаря — заміна кожної букви повідомлення на букву циклічно віддалену в алфавіті на фіксоване число позицій.

2. Застосування телеграфу для шифрування і кодування (з середини XIX ст.)
3. Використання механічних машин (кінець XIX ст. — 20і роки XX ст.)
4. Електромеханічні машини (з 20-х років XX ст. — середина XX ст.)
Приклад — ENIGMA — основна шифрувальна машина Вермахту у Другій світовій війні.
5. Електронні машини (з кінця 40-х років XX ст.)
6. Напівпровідникові криптосистеми
7. Криптосистеми, засновані на мікросхемах

8. Використання комп'ютерної техніки для криптографічного захисту

9. Квантова криптографія

1.4 Про розвиток теоретичної криптографії

До епохи Відродження криптографією займалися, можна сказати, не професіонали. Найчастіше шифри уявляли собою деякі головоломки. Але слід зазначити, що іноді винаходилися шифри, які не були розкриті на протязі 100 і навіть більше років! Хоча за сучасними поняттями, враховуючи залучення до криптоаналізу ЕОМ, загалом це були слабкі, нестійкі шифри.

Велике просування в криптографії відбулося після залучення до вирішення її проблем відомих математиків періоду відродження, серед яких були Ф. Вієт, Д. Кардано та ін. Пізніше почали створюватись спеціалізовані державні служби шифрування і дешифрування (розкриття). Такі служби створили, наприклад, Кромвель у Англії, кардинал Рішельє у Франції, Петро I в Росії.

У 1883 році була опублікована книга Керкгоффа «Військова криптографія», в якій були вперше сформульовані деякі вимоги до криптосистем, правила щодо утворення, експлуатації, стійкості криптографічних пристроїв. Частина цих правил і зараз вважаються обов'язковими.

Наукою у повному розумінні цього слова теоретичну криптографію стали визнавати з 1949 року після публікації у відкритому друці статті К. Шеннона «Теорія зв'язку в секретних системах». А з 1976 року завдяки ідеям статті Діффі і Хеллмана «Нові напрямки в криптографії» почався новий етап у розвитку криптографії – застосування криптосистем з відкритим ключем, яке дало можливість успішно вирішити низку назрілих проблем криптографічного захисту інформації.

1.5 Класифікація сучасних криптосистем

1. *Симетричні (одноключові, з секретним ключем)*. Підрозділяються на блокові й потокові. У відправника та одержувача повідомлення один і той самий секретний ключ, вони знаходяться у рівних (симетричних) умовах, можуть як зашифрувати повідомлення так і розшифрувати за допомогою таємного ключа.
2. *Асиметричні (двохключові, з відкритим ключем, із загальнодоступними ключами)*. Відомі з 1976 року, активно використовуються на практиці з 1978 року. У найпростішому випадку мають два ключі: один (відкритий) — у відправника для шифрування, інший — у одержувача (секретний) для розшифрування.
3. *Квантові* (знаходяться у стадії експерименту та розвитку)

1.6 Контрольні питання

1. Які мета та задачі захисту інформації?

2. Назвіть напрямки та методи захисту інформації.
3. Що таке криптографічний ключ?
4. З яких частин складається криптологія?
5. Чи були відомі способи захисту інформації шифруванням до нашої ери?
6. Коли використовувались роторні шифрувальні машини?
7. Яка дата вважається початком розвитку криптографії з відкритим ключем?
8. Чим відрізняються симетричні криптосистеми від асиметричних?

Розділ 2

Моделі джерел відкритого тексту. Ентропія на символ джерела

При шифруванні текст перетворюється таким чином, щоб зробити його зміст незрозумілим для того, хто не знає секретного ключа. Для побудови математичної теорії криптографічних систем шифрування потрібно насамперед дати математичний опис (або математичну модель) тексту та перетворень, які відбуваються з ним під час шифрування.

Визначення 2.1 (Алфавіт). Надалі вважатимемо, що алфавіт є скінченним. Позначимо алфавіт як $Z_m = \{z_1, \dots, z_m\}$, де z_1, \dots, z_m — букви (символи) алфавіту. Елементами алфавіту можуть бути власне букви; букви та цифри; букви, цифри та знаки пунктуації, взагалі скінченний набір будь-яких символів, наприклад, танцюючі чоловічки. Як правило, ми будемо розглядати український, російський чи латинський алфавіт (малі букви) зі знаком пропуску, що вважається буквою, або без нього, або ж двійковий алфавіт, що складається з двох символів: 0 та 1.

Визначення 2.2 (Текст). Під текстом будемо розуміти послідовність букв деякого алфавіту.

Визначення 2.3 (Відкритий текст). Відкритий текст (ВТ) — це текст, що підлягає шифруванню.

Визначення 2.4 (Шифрований текст). Шифрований текст (ШТ) — це текст, що утворюється в результаті шифрування.

Відкритий та шифрований тексти можуть бути записані як у одному й тому ж, так і у різних алфавітах (більш докладно поняття ВТ та ШТ розглядаються у лекції 3).

Визначення 2.5 (n -грама). n -грамою називається послідовність n символів тексту, що стоять підряд. При $n=2$ це біграма, при $n=3$ — триграма.

Будь-який текст має певну статистичну структуру. Для опису цієї структури використовуються різноманітні ймовірнісні моделі мови.

Визначення 2.6 (Джерело відкритого тексту). Джерело відкритого тексту генерує послідовність символів алфавіту $x_1, x_2, \dots, x_n, \dots$ випадковим чином. Джерело визначається алфавітом та ймовірностями появи n -грам: $\mathbb{P}(x_{i+1} = z_1, x_{i+2} = z_2, \dots, x_{i+n} = z_n)$ для будь-яких цілих $n \geq 1, i \geq 0$ (тут x_1, x_2, \dots, x_n — випадкові величини, а z_1, \dots, z_n — букви алфавіту), які мають задовольняти умовам:

1. Вихід джерела ВТ є випадковим процесом з дискретним часом та множиною станів Z_m

$$\sum_{z_1, z_2, \dots, z_n \in Z_m} \mathbb{P}(x_{i+1} = z_1, \dots, x_{i+n} = z_n) = 1$$

2. Умова узгодженості скінченновимірних розподілів виходу джерела ВТ: Для будь-якого цілого $s \geq 1$

$$\begin{aligned} \sum_{z_1, z_2, \dots, z_n \in Z_m} \mathbb{P}(x_{i+1} = z_1, \dots, x_{i+n} = z_n, \dots, x_{i+n+s} = z_{n+s}) = \\ = \mathbb{P}(x_{i+1} = z_1, \dots, x_{i+n} = z_n) \end{aligned}$$

Визначення 2.7 (Стаціонарне джерело відкритого тексту). Джерело називають стаціонарним, якщо для будь-яких цілих $n \geq 1, 1 \leq i_1 < \dots < i_n, j \geq 0$ і будь-якого набору букв алфавіту z_1, \dots, z_n виконується рівність:

$$\mathbb{P}(x_{i_1+j} = z_1, x_{i_2+j} = z_2, \dots, x_{i_n+j} = z_n) = \mathbb{P}(x_{i_1} = z_1, x_{i_2} = z_2, \dots, x_{i_n} = z_n)$$

У подальшому будемо розглядати лише стаціонарні джерела, тобто такі, у яких немає залежності від зсуву j . Для стаціонарних джерел достатньо задати ймовірності $\mathbb{P}(x_1 = z_1, \dots, x_n = z_n)$ для $n \geq 1$.

В залежності від властивостей сумісних розподілів $\mathbb{P}(x_1 = z_1, \dots, x_n = z_n)$, $n \geq 1$ можна побудувати різні моделі джерела ВТ.

Найбільш уживаними є описані нижче чотири моделі, з яких кожна наступна все більш адекватно відображає структуру мови. Перша з них є простою й менш за всі враховує реальні статистичні властивості мови. Назвемо її моделлю М0.

Визначення 2.8 (Модель М0). У цій моделі джерело у кожен момент часу генерує символи із Z_m незалежно та рівноімовірно:

$$\mathbb{P}(x_i = z) = \frac{1}{m}, z \in Z_m, i = 1, 2, \dots$$

Всі n -грами в моделі М0 є рівноймовірними для будь-яких цілих $n \geq 1$ та $z_1, \dots, z_n \in Z_m$

$$\mathbb{P}(x_1 = z_1, \dots, x_n = z_n) = \frac{1}{m^n}$$

Модель М0 має допоміжний характер, бо вона не враховує навіть найпростіших властивостей мови.

Наступна модель ВТ враховує частоти, з якими окремі букви зустрічаються у мові.

Визначення 2.9 (Модель М1). Символи тексту $x_1, x_2, \dots, x_n, \dots$ є незалежними, але вони генеруються із різними ймовірностями

$$\mathbb{P}(x_i = z) = p(z), z \in Z_m, i = 1, 2, \dots$$

Розподіл ймовірностей $p(z)$ відповідає частотам появи букв $z \in Z_m$ у мові. У цій моделі ймовірність появи n -грами має вигляд:

$$\mathbb{P}(x_1 = z_1, x_2 = z_2, \dots, x_n = z_n) = \prod_{i=1}^n p(z_i)$$

Зазначимо, що ймовірності букв у природних мовах значно відрізняються. Наприклад, найбільшу частоту в українській та російській мові має буква “о”, в англійській — буква “е”. В російській мові “о” зустрічається майже в 50 раз частіше букви “ф”, що має найменшу частоту.*

Наступна модель враховує залежність в мові між двома буквами, що стоять поряд.

Визначення 2.10 (Модель М2). Джерело генерує біграми $x_1x_2, x_3x_4, x_5x_6, \dots$ і т.д. незалежно одну від одної. Тобто на множині всіх біграм заданий розподіл ймовірностей $p(z_i, z_j), i, j = 1, \dots, m$, і кожна нова біграма джерела генерується незалежно від інших.

Більш складні залежності мови враховуються за допомогою марковської моделі.

Визначення 2.11 (Однорідний ланцюг Маркова). Однорідний ланцюг Маркова — це послідовність випадкових величин $\{x_i\}, i = 1, 2, \dots$, що приймають значення у дискретній множині Z , така, що для будь-якого $n \geq 2$

$$\begin{aligned} \mathbb{P}(x_n = z_n \mid x_1 = z_1, x_2 = z_2, \dots, x_{n-1} = z_{n-1}) = \\ = \mathbb{P}(x_n = z_n \mid x_{n-1} = z_{n-1}) = p_{z_n z_{n-1}} \end{aligned}$$

Для неоднорідного марковського ланцюга ймовірності переходу $p_{z_n z_{n-1}}$ залежали б від n , а в нашому випадку вони залежать лише від станів z_{n-1}, z_n .

Визначення 2.12 (Модель М3). У цій моделі джерела ВТ послідовність x_1, x_2, \dots утворює однорідний ланцюг Маркова. Для завдання такого марковського ланцюга достатньо задати розподіл початкових станів $p_0(i), i \in Z_m$ та перехідні ймовірності $p_{ij} = \mathbb{P}(x_{n+1} = j \mid x_n = i), i, j \in Z_m$, які в силу однорідності не залежать від n .

При виконанні деяких умов на ланцюг Маркова (які не суперечать властивостям природних мов) існує граничний розподіл

$$\pi_i = \lim_{n \rightarrow \infty} \mathbb{P}(x_n = i \mid x_1 = j)$$

що не залежить від початкового стану j . Він називається стаціонарним розподілом ймовірностей марковського ланцюга. Ці ймовірності задовольняють наступним рівностям

*За умови, що твердий знак отожднюється з м'яким.

$$\begin{cases} \sum_{i \in Z} \pi_i = 1 \\ \sum_{i \in Z} \pi_i p_{ij} = \pi_j, j \in Z_m \end{cases}$$

Ймовірність n -грами у моделі МЗ у стаціонарному режимі можна записати у вигляді

$$\mathbb{P}(x_1 = z_1, \dots, x_n = z_n) = \pi_{z_1} p_{z_1 z_2} p_{z_2 z_3} \dots p_{z_{n-1} z_n}$$

Якщо існує стаціонарний розподіл і $p_0(i) = \pi_i, i \in Z_m$, то одноріний ланцюг Маркова — стаціонарний випадковий процес.

Зміст

1	Задачі, напрямки та методи захисту інформації. Поняття про криптографічний захист інформації	3
1.1	Області застосування, мета, методи захисту інформації	3
1.2	Перші поняття криптографічного захисту інформації	5
1.3	Етапи розвитку технологічних засобів криптографії	6
1.4	Про розвиток теоретичної криптографії	7
1.5	Класифікація сучасних криптосистем	7
1.6	Контрольні питання	7
2	Моделі джерел відкритого тексту. Ентропія на символ джерела	9