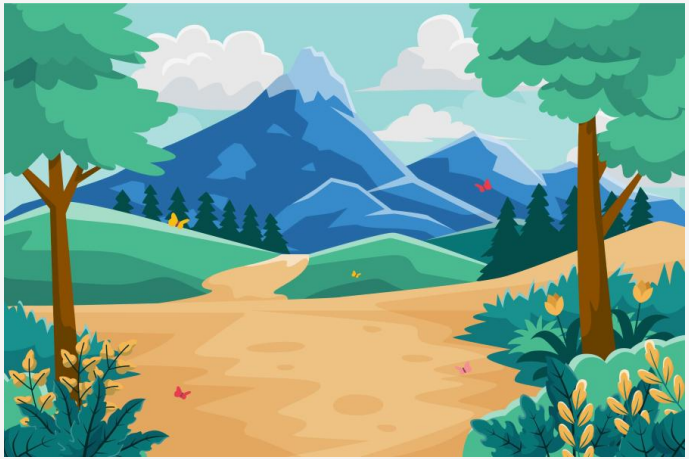
Літня школа з кібербезпеки

Основи стеганографії

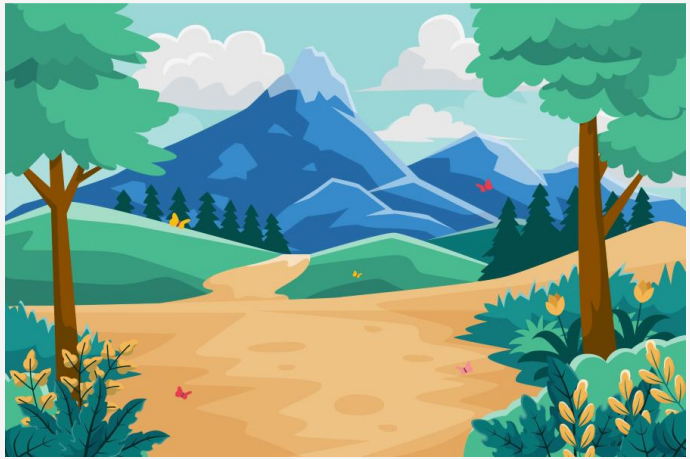


© Щур Н.О.

У якому зображенні приховано повідомлення?



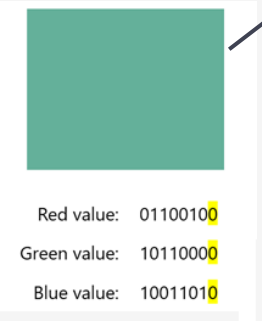
Зображення А

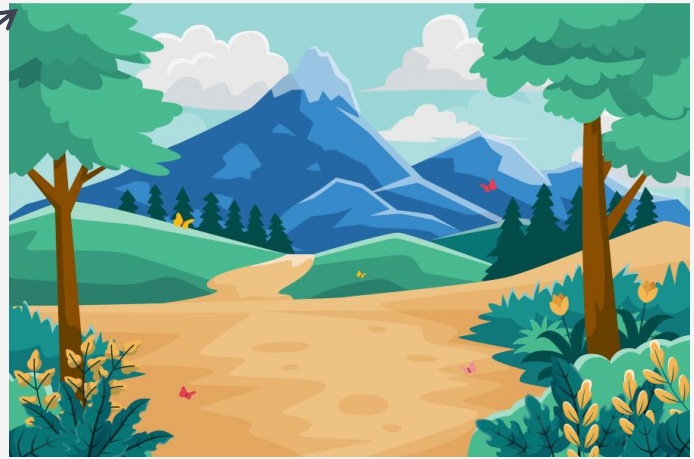


Зображення В

Повідомлення приховано в пікселях

Hi, this is a secret message!





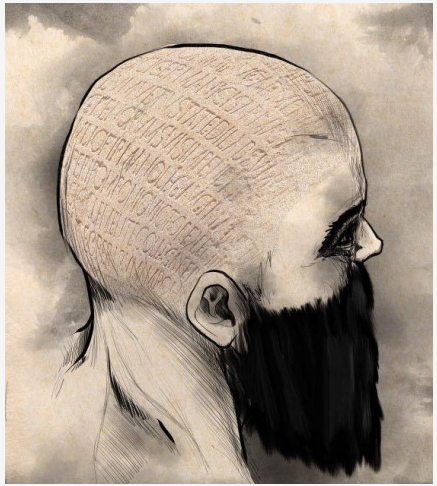
Зображення B

Що таке стеганографія?

Стеганографія – це наука про методи та способи, які дозволяють **приховувати** сам факт **існування** секретного повідомлення у тому чи іншому середовищі або об’єкті

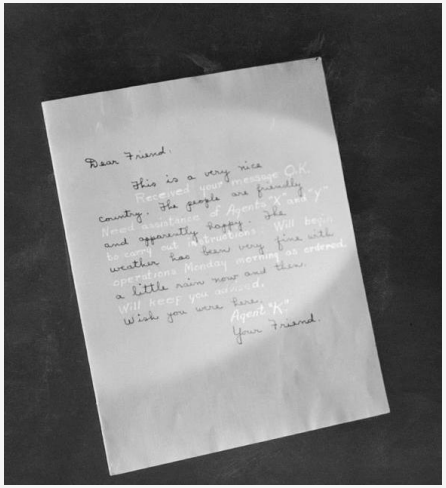


Стеганографія існувала задовго до комп'ютерів



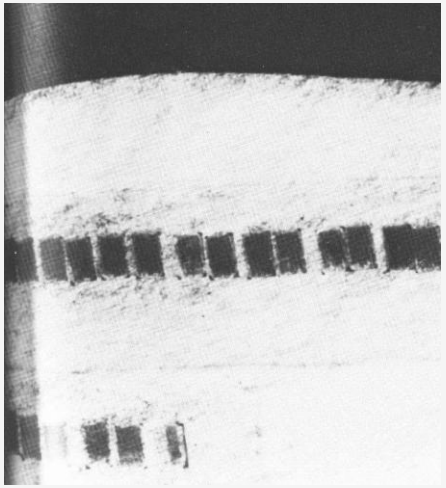
Послання на голові раба

Античність



Невидимі чорнила

Новий час



Мікроточки (зменшені фото)

XX століття

Стеганографія у цифрову епоху

Сучасна цифрова стеганографія вивчає методи **приховування** (вбудовування) додаткової інформації в цифрові об'єкти, такі як **зображення**, **аудіо**, **відео, текстові** файли та інші медіа



Напрямки використання стеганографії



Цифрова криміналістика
пошук прихованих даних, аналіз файлів



Конфіденційна комунікація
передавання повідомлень без привернення уваги



Ігри та медіа
приховані підказки, Easter eggs, ARG-квести



Захист авторських прав
цифрові водяні знаки, маркування контенту

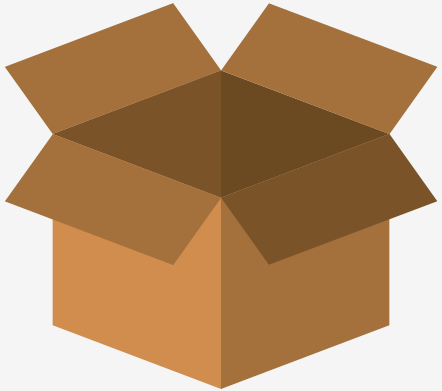
Стеганографія ≠ криптографія

	Стеганографія	Криптографія
Мета	Приховати сам факт існування повідомлення	Приховати зміст повідомлення
Що бачить сторонній?	Звичайний файл без очевидних ознак секрету	Зашифровані або незрозумілі дані
Де секрет?	Усередині іншого об'єкта-носія	У зашифрованому повідомленні

Найкращий підхід: спочатку зашифрувати повідомлення (криптографія), а потім приховати отриманий шифротекст (стеганографія)

Де ховати?

Контейнер – це цифровий файл або об'єкт, який використовується для приховування інформації



Порожнім контейнером називають контейнер без будь-якого секретного повідомлення

Стегооб’єкт (стегоконтейнер) – контейнер після вбудовування прихованої інформації

У які об’єкти можна приховати дані?

Типи контейнерів

Зображення

Дані приховуються в пікселях, колірних каналах (чи інших компонентах) зображення

Аудіо

Дані приховуються у звукових семплах або спектрі сигналу

Відео

Дані інтегруються у кадри або звукову доріжку відео

Текст

Дані приховуються через пробіли, невидимі символи або структуру тексту

Протокол

Дані інтегруються у заголовки або інші службові поля мережевих пакетів

Що ховати?

Секретне повідомлення — це дані, які потрібно приховати всередині контейнера

Секретним повідомленням може бути:
текст, QR-код, зображення, аудіо, відео, архів, документ або інший файл

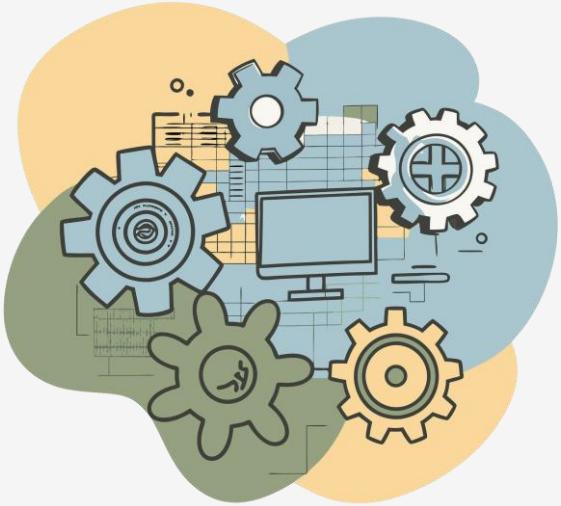
Стегодані – це приховані дані, які вбудовано в контейнер



Як ховати та вилучати?

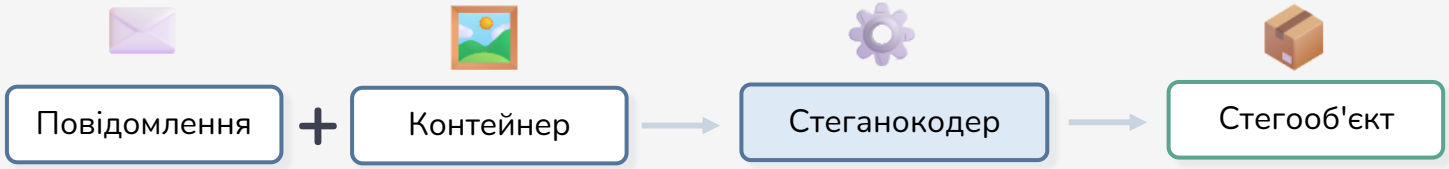
Кодер (стеганокодер) – це алгоритм або пристрій, що вбудовує стегодані в контейнер

Декодер (стеганодекодер) – це алгоритм або пристрій, який вилучає стегодані з контейнера

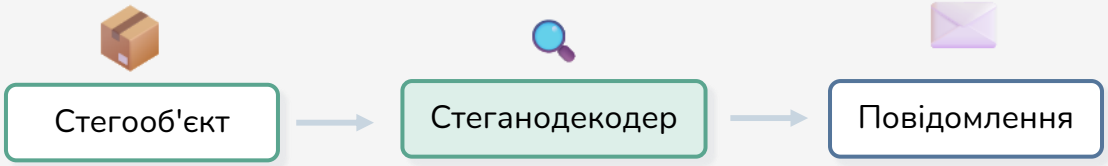


Як працює стегосистема?

Приховування

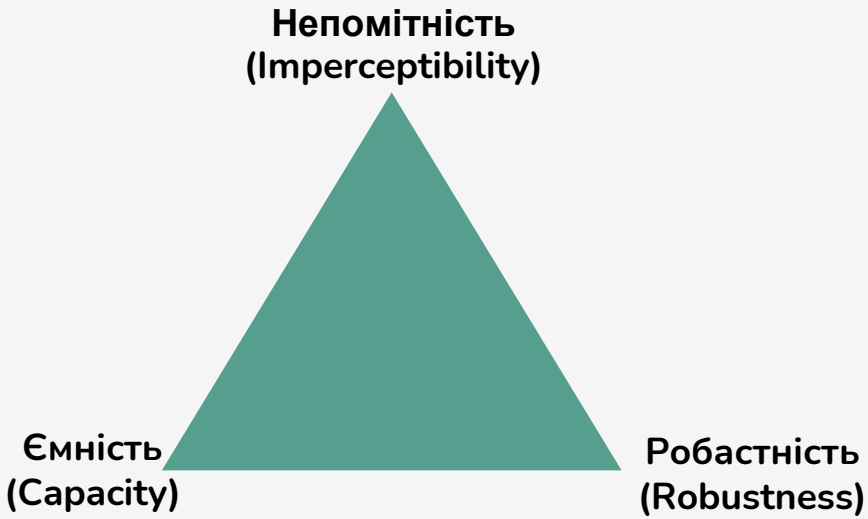


Вилучення



Три ключові характеристики стегосистеми

- **Ємність** – максимально можливий обсяг даних, які можна приховати у контейнері
- **Непомітність** – зміни в контейнері настільки малі, що не помітні для людського сприйняття
- **Робастність** – здатність прихованих даних залишатися неушкодженими після різних перетворень носія



Як виявляти?

Стегоаналіз – це набір методів та алгоритмів, спрямованих на **виявлення** прихованих даних у цифрових носіях або мережевих потоках шляхом **аналізу** їхніх структурних чи статистичних характеристик



Зображення як популярний контейнер

Ємність

Пікселі зображення містять достатньо інформації для приховування даних

Непомітність

Незначні зміни в пікселях непомітні для людського ока

Гнучкість форматів

Великий вибір форматів зображень дозволяє обрати оптимальний спосіб приховування

Широке використання

Зображення легко передавати в мережі без привернення зайвої уваги

Підходи до вбудовування даних у зображення

Способи вбудовування

Просторова область

Зміни на рівні пікселів чи інших базових елементів

- **LSB** – заміна молодших бітів RGB-пікселів
- **PVD** – приховування через різницю між сусідніми пікселями
- **Histogram Shifting** – зміщення гістограми яскравості
- **Palette-Based** – зміна індексів палітри у PNG/GIF

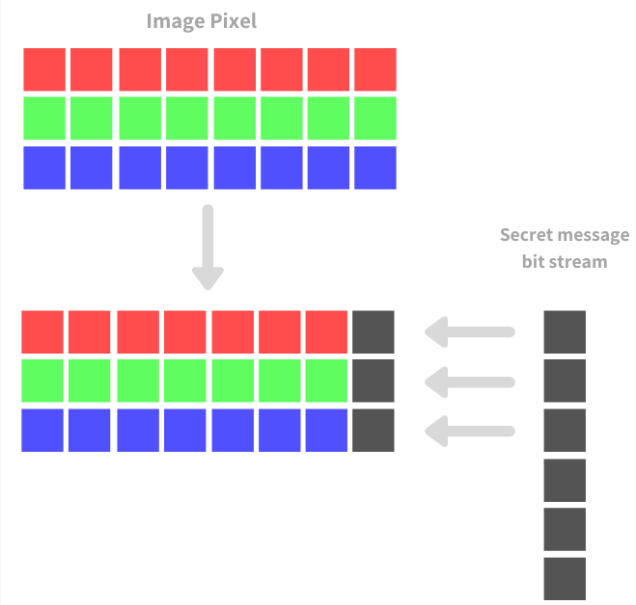
Частотна область

Вбудовування даних у частотні компоненти носія

- **DCT** – перетворення блоку зображення на набір частотних коефіцієнтів
- **DWT** – вбудовування даних у різні рівні деталізації зображення
- **DFT** – приховування даних у частотній структурі зображення
- **Spread Spectrum** – розподіл секретного повідомлення по багатьох частинах зображення

Метод LSB в зображеннях

Метод заміни найменшого значущого біта (LSB, Least Significant Bit) полягає у приховуванні даних шляхом **заміни** найменших значущих бітів пікселів зображення бітами секретного повідомлення



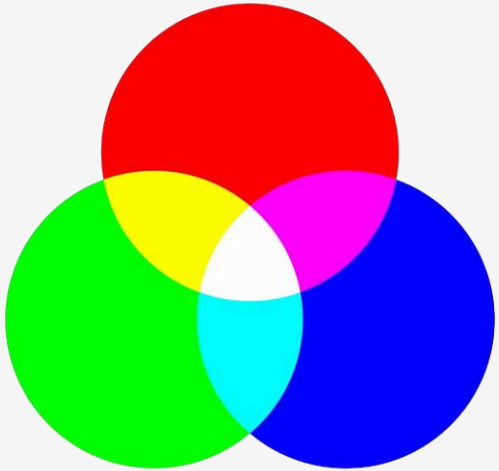
Анатомія цифрового пікселя

Цифрове (растрове) зображення складається з **пікселів**.

У колірній моделі RGB кожен піксель:

- складається із **трьох** компонентів:

  червоний (R), зелений (G) і синій (B)
- представляється за допомогою **24 бітів** (по 8 бітів на кожен колірний компонент)
- кодується числом у діапазоні від 0 до 255, що відповідає **інтенсивності** кольору

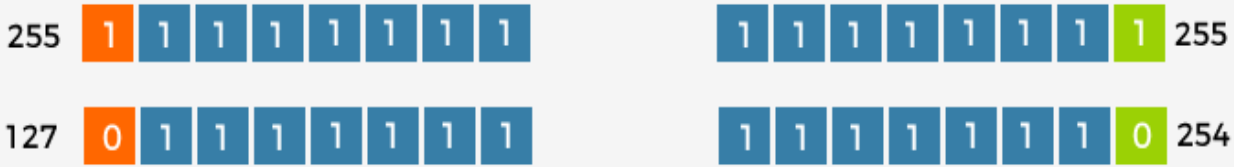


MSB vs LSB

Найбільший значущий біт (MSB, Most Significant Bit) розташований ліворуч і відповідає за основну вагу числа



Найменш значущий біт (LSB, Least Significant Bit), знаходиться у крайньому правому положенні і має найменшу вагу



Принцип LSB-стеганографії

Якщо виконати заміну LSB компонентів пікселя, то ці зміни практично **не впливають** на колір зображення і залишаються **непомітними** для людського ока



150 = 1001011**0** (R)
200 = 1100100**0** (G)
80 = 0101000**0** (B)



151=1001011**1** (R)
201 = 1100100**1** (G)
81 = 0101000**1** (B)

Якщо LSB уже дорівнює потрібному біту – значення не змінюється

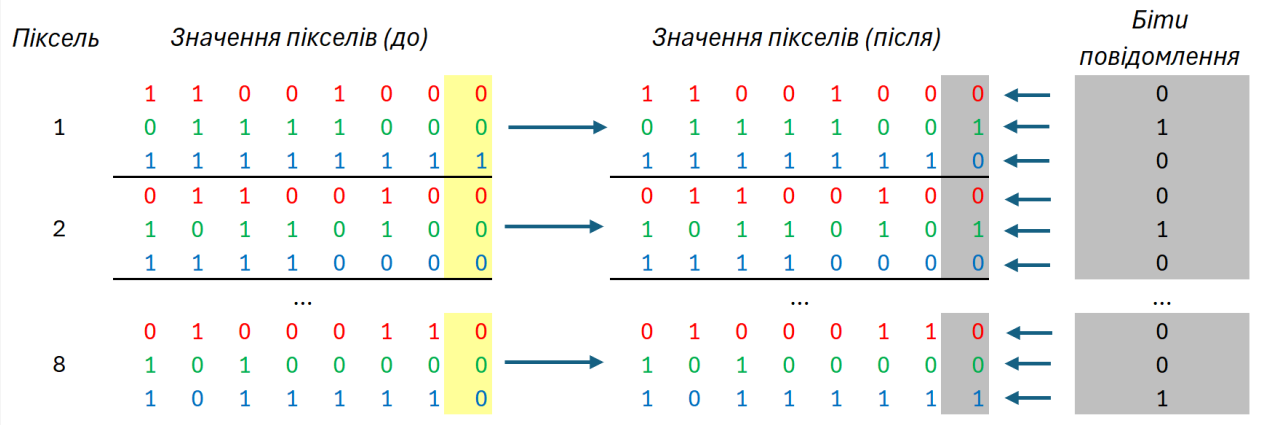
Алгоритм вбудовування методом LSB

1. **Підготовка даних.** Секретне повідомлення перетворюється у бінарний вигляд
2. **Вибір контейнера.** Обирається зображення для приховування даних (PNG/BMP)
3. **Заміна LSB у пікселях.** Кожен біт секретного повідомлення послідовно замінює LSB компоненти пікселів
4. **Формування нового зображення.** Після заміни LSB у всіх необхідних пікселях отримуємо змінену копію зображення

Приховаємо текст «Hi!» у LSB пікселів

Кожен символ перетворимо у ASCII-код, а потім у бінарний вигляд: 01001000 01101001 00100001 (24 біти)

Нам буде потрібно 8 пікселів, кожен з яких зберігає 3 біти повідомлення (по 1 біту на R, G, B)



Аудіо як контейнер

Цифрове аудіо – це послідовність дискретних числових аудіовідліків, отриманих під час оцифрування звукового сигналу



Аудіостеганографія вбудовує дані шляхом **контрольованої модифікації** аудіовідліків або їх спектрального представлення так, щоб зміни були **непомітні** для слухача

Основи цифрового аудіо

Поняття	Пояснення	Приклад
Амплітуда	числове значення відліку, що описує миттєвий рівень сигналу	-32768...32767 для 16-bit PCM
Семпл	дискретне числове значення амплітуди аудіосигналу в певний момент часу	16-bit PCM sample
Частота дискретизації	кількість семплів, що записуються або відтворюються за одну секунду	44 100 Гц = 44 100 семплів/с
Бітова глибина	кількість бітів, для кодування одного семплу; визначає кількість рівнів квантування	16-bit = 65 536 рівнів
Канали	незалежні послідовності аудіосемплів, що зберігаються синхронно	mono = 1 канал, stereo = 2 канали
Формати	спосіб кодування, стиснення та збереження аудіоданих у файлі	WAV, MP3, FLAC

Аудіо можна змінювати на різних рівнях

Способи вбудовування у аудіо

Часова область

Зміни самих семплів у часі

- **LSB** – змінюється молодший біт кожного семплу
- **Parity Coding** – парність групи семплів = секретний біт
- **Echo Hiding** – додається слабке відлуння з різними параметрами

Частотна область

Приховування у спектрі частот

- **DFT/FFT** – вбудовування у вибрані частотні компоненти
- **DCT** – змінюються коефіцієнти після косинусного перетворення
- **Phase Coding** – змінюється фаза окремих частотних компонентів
- **Spread Spectrum** – розподіл даних по широкому діапазону частот

Частотно-часова область

Зміни частотних компонентів у часі

- **Spectrogram Hiding** – додавання слабких частотних компонентів, видимих на спектрограмі
- **DWT**– дані ховаються на різних рівнях деталізації сигналу

Приховування у спектрограмі аудіо

Спектрограма – це візуальне відображення звуку

- вісь X – час,
- вісь Y – частота,
- яскравість – гучність

Якщо зображення перетворити у звук, воно може звучати як шум, скрегіт або «космічний» сигнал



Текст як контейнер

Текстова стеганографія – це приховування даних всередині текстових файлів шляхом маніпуляції форматуванням, невидимими символами або лінгвістичними особливостями тексту



Головний виклик: текст має невеликий розмір і меншу природну надлишковість, ніж зображення чи аудіо

Методи приховування у тексті

Способи вбудовування

Символьні та форматні методи

Зміна зовнішнього вигляду тексту або службових символів без зміни змісту

- **Zero-Width Characters** – невидимі Unicode-символи у тексті
- **Whitespace** – приховування за допомогою маніпуляції пробілами чи табуляцією
- **Font / Style** – зміна розміру, кольору шрифту
- **Homoglyphs** – схожі символи з різних алфавітів
- **HTML / Markdown** – приховані коментарі або службові елементи

Лінгвістичні методи

Маніпуляція самими словами, синонімами чи структурою тексту

- **Акровірш** – повідомлення з перших літер рядків
- **Синоніми** – вибір слів кодує біти або символи
- **Структура тексту** – порядок слів, речень або абзаців

Акровірш – історичний метод приховування у тексті

Акровірш (acrostic) являє собою вірш чи інший словесний твір, у якому перша літера кожного нового рядка або абзацу, або іншого повторюваного елемента в тексті утворює повідомлення

History unfolds in unexpected ways.
Ideas shape the world around us.
Discovery drives progress.
Every innovation matters.

Історія розгортається несподіваними шляхами.
Думки формують світ навколо нас.
Експерименти рухають прогрес.
Якщо є ідея – з’явиться можливість.

Магія Unicode

Стандарт **Unicode** містить спеціальні символи, що не мають візуальної ширини і не відображаються на екрані:

\u200B

Zero Width Space

UTF-8: E2 80 8B

часто як двійковий **0**

\u200C

Zero Width Non-Joiner

UTF-8: E2 80 8C

часто як двійкова **1**

Секретний текст переводять у біти, після чого 0 і 1 кодують невидимими Unicode-символами

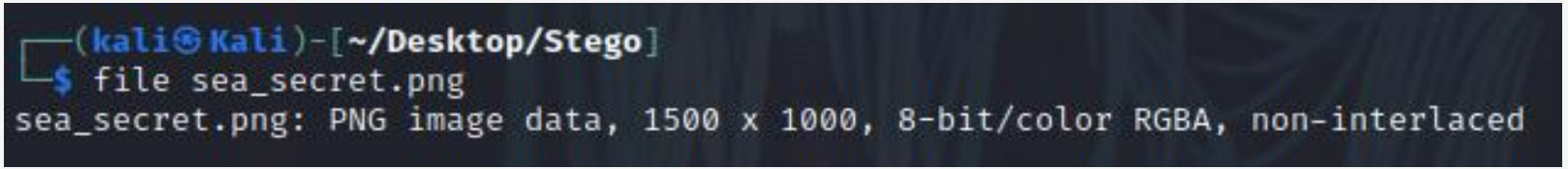
Базовий алгоритм аналізу CTF-файлу

Золоте правило стегоаналізу: починайте з найпростішого

Крок	Що робимо	Інструменти
1	Перевіряємо реальний тип файлу	`file`, `hexdump`
2	Знаходимо текстові рядки	`strings`
3	Переглядаємо метадані	`exiftool`
4	Шукаємо вкладені файли	`binwalk`
5	Аналізуємо канали, LSB або спектрограму	StegSolve, StegOnline, `zsteg`, Sonic Visualiser
6	Перевіряємо кодування та паролі	CyberChef, `steghide`

З чого насправді складається файл?

`file` – команда, що аналізує вміст файлу, а не його розширення



Якщо розширення і реальний тип не збігаються – це перша підказка

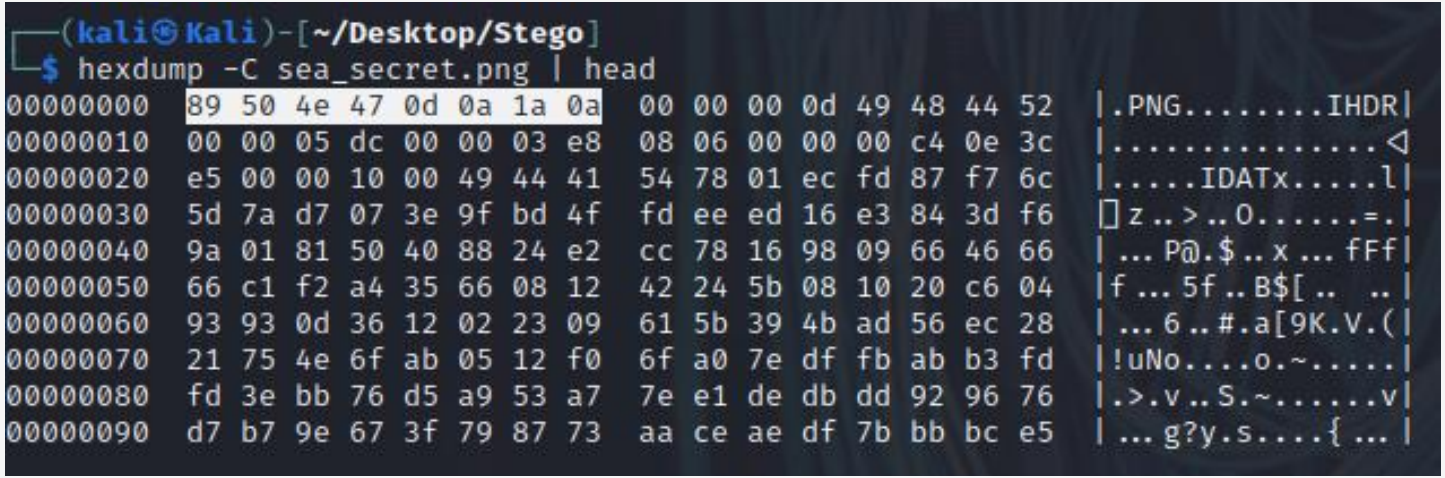
Що таке Magic Numbers?

Файлова сигнатура – це перші байти файлу, за якими можна визначити його справжній формат

Розширення	Тип файлу	Сигнатура (hex)
.png	PNG-зображення	`89 50 4E 47 0D 0A 1A 0A`
.jpg / .jpeg	JPEG-зображення	`FF D8 FF`
.pdf	PDF-документ	`25 50 44 46`
.zip	ZIP-архів	`50 4B 03 04`
.rar	RAR-архів	`52 61 72 21 1A 07 00`
.gif	GIF-зображення	`47 49 46 38`
.bmp	BMP-зображення	`42 4D`
.txt	Текстовий файл	Залежить від кодування
.html	HTML-документ	`3C 21 44 4F 43 54 59 50 45`
.json	JSON-текст	`7B`

Погляд на файл як на байти

`hexdump` – утиліта для перегляду вмісту файлу у шістнадцятковому вигляді



Пошук видимого тексту

`strings` – утиліта для пошуку читабельних рядків усередині файлів незалежно від їх типу

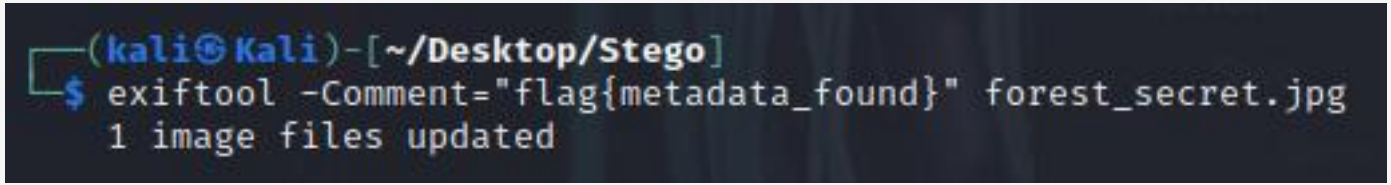
```
┌──(kali㉿Kali)-[~/Desktop/Stego]
└─$ echo "flag{strings_found_me}" >> bird_secret.jpg
```

```
┌──(kali㉿Kali)-[~/Desktop/Stego]
└─$ strings bird_secret.jpg | grep flag
flag{strings_found_me}
```

Strings добре працює лише тоді, коли текст зберігається відкрито

Метадані як джерело підказок

`exiftool` – утиліта для перегляду, аналізу та редагування метаданих файлів



У CTF використовується для пошуку прихованих підказок, прапорів, паролів або технічної інформації у службових полях файлу

Пошук вкладених файлів

`binwalk` – утиліта для пошуку вкладених файлів, архівів, стиснених даних або сигнатур відомих форматів усередині інших файлів

```
┌──(kali㉿Kali)-[~/Desktop/Stego]
└─$ binwalk sky_secret.png

DECIMAL       HEXADECIMAL     DESCRIPTION
--------------------------------------------------------------------------------
0             0×0             PNG image, 1500 x 1001, 8-bit/color RGBA, non-interlaced
91            0×5B            Zlib compressed data, compressed
1545881       0×179699        Zip archive data, at least v1.0 to extract, compressed size: 21, uncompressed size: 21, name: m
essage.txt
1546052       0×179744        End of Zip archive, footer length: 22
```

Приховування й вилучення з паролем

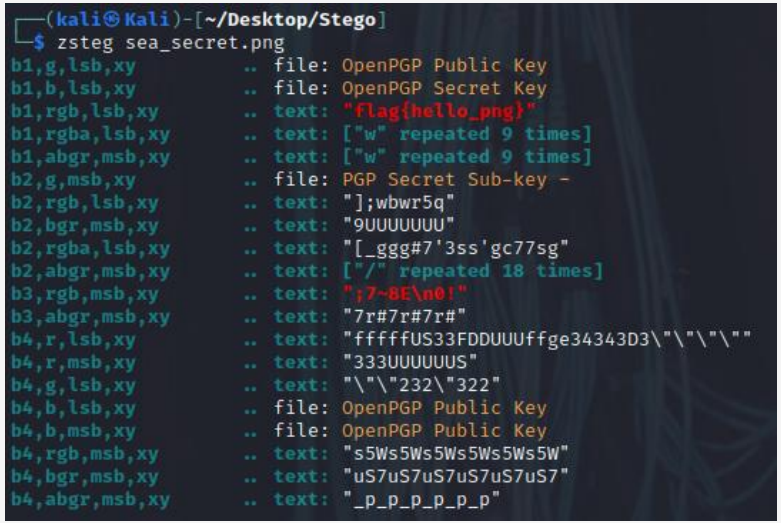
`steghide` – утиліта для приховування та вилучення файлів у JPG, BMP, WAV, AU

```
┌──(kali㉿Kali)-[~/Desktop/Stego]
└─$ steghide embed -cf cat.bmp -ef secret.txt -sf cat_secret.bmp
Enter passphrase:
Re-Enter passphrase:
embedding "secret.txt" in "cat.bmp"... done
writing stego file "cat_secret.bmp"... done
```

```
┌──(kali㉿Kali)-[~/Desktop/Stego/extracted]
└─$ steghide extract -sf ../cat_secret.bmp
Enter passphrase:
wrote extracted data to "secret.txt".
```

Пошук LSB у PNG та BMP

`zsteg` – утиліта для аналізу графічних файлів з метою пошуку прихованих даних (у LSB пікселів)



Корисні інструменти стегоаналізу

Інструмент	Опис
CyberChef	Універсальний інструмент для аналізу, перетворення та декодування даних
StegOnline	Онлайн-інструмент для роботи з LSB-стеганографією у зображеннях
StegSolve	Графічна Java-утиліта для візуального аналізу зображень за кольоровими каналами та бітовими площинами
Sonic Visualiser, Audacity	Програми для аналізу аудіофайлів і перегляду спектрограм

Підсумок

- Стеганографія приховує сам факт існування секретного повідомлення
- Контейнером може бути зображення, аудіо, відео, текст або мережевий протокол
- Контейнер + секретне повідомлення → стегооб’єкт
- Ємність, непомітність і робастність завжди балансують між собою
- Стегоаналіз починається з простих перевірок