Стеганография — искусство скрытой передачи данных, но её оборотная сторона — стегоанализ — наука, которая занимается обнаружением таких скрытых сообщений. В отличие от криптографии, где защита обеспечивается шифрованием, стеганография маскирует сам факт передачи, а стегоанализ вскрывает эту маскировку.

Актуальность темы:

* Рост киберпреступности.
* Развитие инструментов стеганографии (например, скрытие данных в DeepFake).
* Правовые аспекты: во многих странах использование стеганографии требует регулирования.

Цель стегоанализа:

* Обнаружение факта скрытой передачи.
* Определение метода стеганографии (LSB, DCT и др.).
* Возможность извлечения или блокировки сообщения.

Основные понятия

* Объект-носитель — исходный файл (изображение, аудио, текст).
* Стегоконтейнер — файл с внедрённым сообщением.
* Стегоключ — параметр для извлечения данных.

Классификация стегоанализа:

1. Пассивный: только детекция (например, анализ статистик).
2. Активный: попытка удалить или повредить скрытые данные.
3. Слепой: без знания метода стеганографии.
4. Адаптивный: с использованием машинного обучения.

Методы стегоанализа

Стегоанализ использует различные подходы в зависимости от типа носителя (изображение, аудио, текст) и метода стеганографии.

Статистические методы

Основаны на выявлении аномалий в статистических характеристиках данных после внедрения сообщения.

Примеры:

* Анализ гистограмм (для изображений):
  + В LSB-стеганографии гистограмма яркостей становится «ступенчатой» из-за парных значений.
  + Метод: проверка частоты соседних значений в гистограмме (например, тест χ²).
* Анализ DCT-коэффициентов (для JPEG):
  + Встраивание данных искажает распределение коэффициентов дискретного косинусного преобразования.
  + Метод: сравнение статистик (например, среднего, дисперсии) с эталонными.
* Энтропийный анализ:
  + Скрытые данные могут увеличивать энтропию файла.
  + Метод: сравнение энтропии исходного и модифицированного файла (например, для аудио).

Структурные методы

Выявляют нарушения внутренней структуры данных, вызванные стеганографией.

Примеры:

* Анализ заголовков файлов:
  + Несоответствие размера файла и его содержимого (например, в PNG).
* Проверка артефактов сжатия:
  + В JPEG стегоалгоритмы (например, F5) оставляют характерные искажения в блоках 8×8.

Методы машинного обучения

Современный тренд — использование ИИ для автоматического обнаружения стегоконтейнеров.

Примеры:

* Классические алгоритмы:
  + SVM (опорные векторы), Random Forest для классификации "стего/не стего" на основе признаков.
* Нейронные сети:
  + CNN (свёрточные сети) для анализа изображений (например, XuNet, YeNet).
  + RNN (рекуррентные сети) для аудио и текста.

Данные для обучения:

* Базы изображений (BOSSBase, ALASKA) с чистыми и модифицированными файлами.

Универсальные (blind) методы

Работают без знания конкретного стегоалгоритма, выявляя общие аномалии.

Примеры:

* SPAM-фильтры (SPatial stegAnalysis):
  + Анализ пространственных корреляций между пикселями.
* Rich Models:
  + Комбинация сотен признаков (например, SRM — Spatial Rich Model).

Инструменты стегоанализа

Существуют специализированные программы и фреймворки для обнаружения стегоконтейнеров.

1. StegExpose
   * Назначение: Обнаружение LSB-стеганографии в изображениях (PNG, BMP).
   * Метод: Анализ гистограмм и статистические тесты (χ²).
   * Плюсы: Высокая точность для простых методов.
2. StegDetect
   * Назначение: Поиск стегосообщений в JPEG (F5, JSteg, OutGuess).
   * Метод: Анализ DCT-коэффициентов.
   * Минусы: Устарел для современных алгоритмов.
3. Aletheia
   * Назначение: Автоматический стегоанализ с использованием ML.
   * Метод: Поддерживает SVM, CNN (например, для обнаружения стеганографии в JPEG).
   * Плюсы: Гибкость и поддержка нейросетей.
4. OpenStego
   * Назначение: Универсальный анализ (включая watermarking).
   * Минусы: Низкая эффективность против сложных методов.

Современные фреймворки

* StegBench (2023): Тестирование стойкости стегоалгоритмов на датасетах (ALASKA, BOSSBase).
* Steganography-Toolkit (GitHub): Интеграция с ИИ-моделями (например, для обнаружения стеганографии в VoIP).

Современные сложности и тренды

С развитием технологий стегоанализ сталкивается с новыми угрозами и задачами:

1. Генеративные модели (GAN, Diffusion)

* Проблема: Злоумышленники используют GAN (например, StyleGAN) для создания стегоконтейнеров, которые неотличимы от реальных изображений.
* Пример: Скрытие данных в сгенерированных артефактах (шумах).
* Методы борьбы:
  + Обучение детекторов на синтетических данных (например, с использованием Contrastive Learning).

2. Стеганография в нейросетях

* Тренд: Встраивание данных в веса нейросетей (например, в модели Stable Diffusion).
* Сложность: Традиционные методы стегоанализа не работают — требуется анализ градиентов и активаций сети.

3. Атаки на стегоанализ

* Adversarial-атаки: Искажение стегоконтейнера, чтобы «обмануть» детектор (аналогично атакам на ИИ-классификаторы).
* Пример: Добавление шума, который незаметен для человека, но ломает статистические тесты.

Особенности стеганографии для соцсетей

* Потеря данных при загрузке:
  + Соцсети пережимают изображения (JPEG, WebP), удаляя скрытые сообщения в LSB или DCT-коэффициентах.
  + Пример: Instagram конвертирует PNG в JPEG с качеством ~85%.
* Обрезка и масштабирование:
  + Автоматическое кадрирование (например, для превью) может обрезать стегоканал.
* Детекция водяных знаков:
  + Нейросети соцсетей (например, Facebook’s PhotoDNA) ищут аномалии в пикселях.

|  |  |  |
| --- | --- | --- |
| Метод | Устойчивость к пережатию | Риск детекции |
| LSB в PNG | Низкая (пережатие → уничтожение) | Высокий (легко обнаружить) |
| DCT в JPEG | Средняя (если избегать нулевых коэффициентов) | Средний (анализ статистик) |
| GAN-стеганография | Высокая (данные в шумах нейросети) | Низкий (идёт развитие в данном направлении) |
| Метаданные (Exif) | Нулевая (соцсети удаляют) | Высокий |

Устойчивые методы

1. Адаптивная стеганография в JPEG:
   * Внедрять данные только в среднечастотные DCT-коэффициенты (менее заметно при пережатии).
   * Использовать синдромное кодирование (устойчивость к потерям).
2. GAN-методы:
   * HiDDeN (нейросеть, которая учится скрывать данные, устойчивые к пережатию).
   * StegaStamp (встраивание в текстуры, которые сохраняются даже после редактирования).
3. Стеганография в цветовых профилях:
   * Данные в ICC-профиле (редко удаляется соцсетями).

Сокрытие факта стеганографии:

* Избегание LSB. Использовать QIM (Quantization Index Modulation) для JPEG.
* Имитация артефактов сжатия. Данные маскируются под естественный шум JPEG.
* Дозированное внедрение. Не более 1-3% от общего размера файла (иначе риск детекции).

Предложения для диплома: использования автоэнкодеров, визуальных трансформеров, diffusion-моделей, возможно сокрытие в весах нейронной сети. Возможно комбинация автоэнкодера и lsb/dct