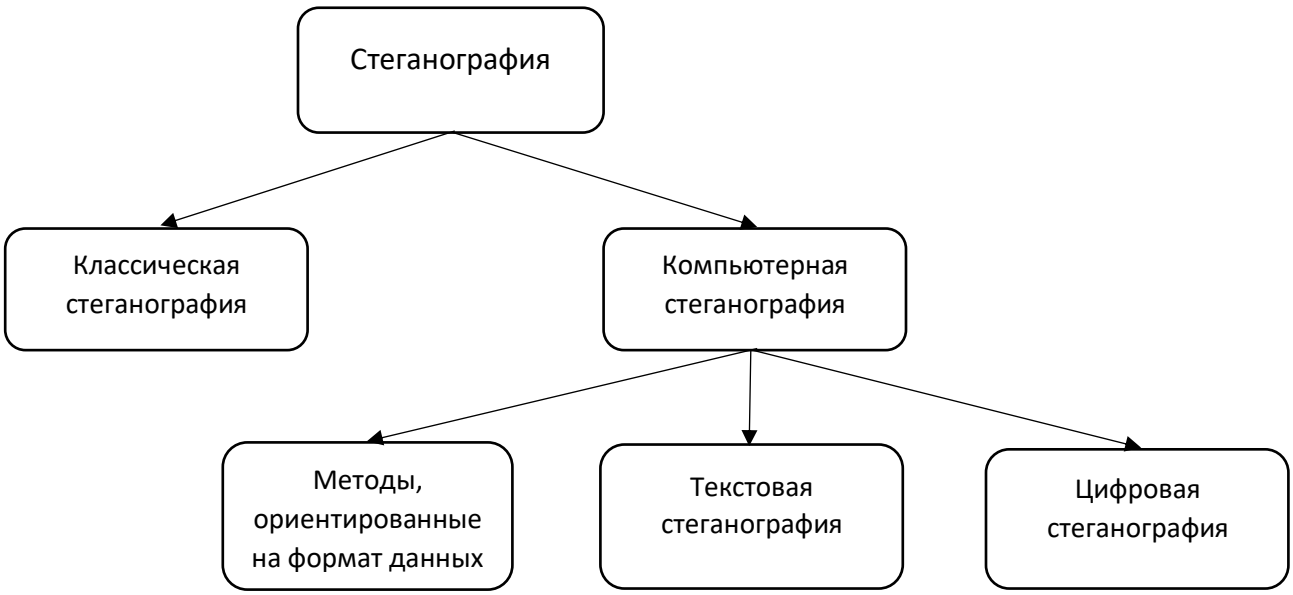
Теоретические основы встраивания информации

1. Что такое стеганография

Стеганография – это способ скрытия информации внутри другой информации или физического объекта так, чтобы она была не обнаружена. С использованием техники стеганографии можно скрыть практически любой цифровой контент, включая текстовые файлы, изображения, аудио и видео. И когда скрытая информация доходит до получателя, она может быть извлечена.

Стеганографию иногда сравнивают с криптографией, так как обе являются методами секретной коммуникации. Однако между ними есть разница, так как в случае со стеганографией данные не шифруются при отправке и для их получения не требуется ключ дешифрования.

Основные направления стеганографии:



2. Что такое цифровой водяной знак

Встраиванием ЦВЗ (Digital Watermarking) называется процесс внедрения в цифровой сигнал (как заметного, так и незаметного) информации, имеющей некоторое отношение к этому цифровому сигналу. Цифровым водяным знаком называется собственно внедряемая информация. Примером такой информации может быть идентификатор автора, предназначенный для защиты авторских прав на аудиовизуальное произведение, или электронная цифровая подпись, подтверждающая аутентичность цифровой мультимедийной информации.

Системой встраивания ЦВЗ (Watermarking system, система ЦВЗ, ЦВЗ-система) будем называть совокупность методов и средств, предназначенных для внедрения в цифровой сигнал информации, имеющей некоторое отношение к этому цифровому сигналу.

3. Какие существуют методы встраивания скрытой информации в контейнер

Метод наименее значимого бита (LSB): суть этого метода заключается в замене последних значащих битов в контейнере (изображения, аудио или видеозаписи) на биты скрываемого сообщения. Разница между пустым и заполненным контейнерами должна быть не ощутима для органов восприятия человека.

Спектральное встраивание: Изменение спектральных компонентов аудио или видео сигналов. Например, модификация коэффициентов дискретного косинусного преобразования (DCT) в изображениях JPEG.

Метод маскировки и фильтрации: Используется в аудиостеганографии, где скрытая информация добавляется к аудиофайлу на частотах, которые меньше всего воспринимаются человеческим ухом.

Разрезание и смешивание: Разделение скрываемой информации на части и встраивание их в различные участки контейнера.

Стеганография на основе разности: Изменение разности между парами пикселей или аудиосэмплов таким образом, чтобы внедрить скрытую информацию.

Форматно-специфическое встраивание: Использование особенностей конкретных форматов данных, например, заголовков файлов или метаданных для скрытия информации.

4. Как работает метод встраивания в наименее значимую битовую плоскость (Least Significant Bit, НЗБ, LSB)

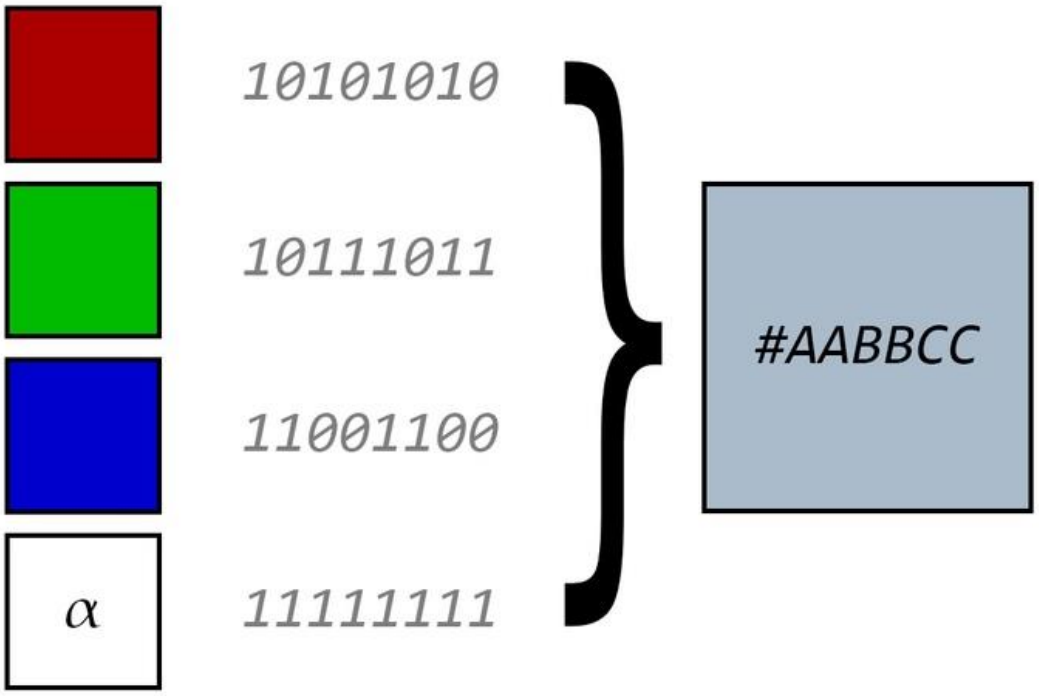
Метод LSB (англ. Least Significant Bit — Наименее значимый бит) относится к пространственным методам цифровой стеганографии в изображениях. То есть, в рамках данного метода мы будем «прятать» информацию в каждом пикселе или группе пикселей какого-либо изображения, изменяя при этом младший бит каждого цветового канала.

Данный метод заключается в выделении наименее значимых бит изображения-контейнера с последующей их заменой на биты сообщения. Поскольку замене подвергаются лишь наименее значимые биты, разница между исходным изображением-контейнером и контейнером, содержащим скрытые данные невелика и обычно незаметна для человеческого глаза.

Метод LSB применим лишь к изображениям в форматах без сжатия или со сжатием без потерь, так как для хранения скрытого сообщения используются наименее значимые биты значений пикселей, при сжатии с потерями эта информация может быть утеряна.

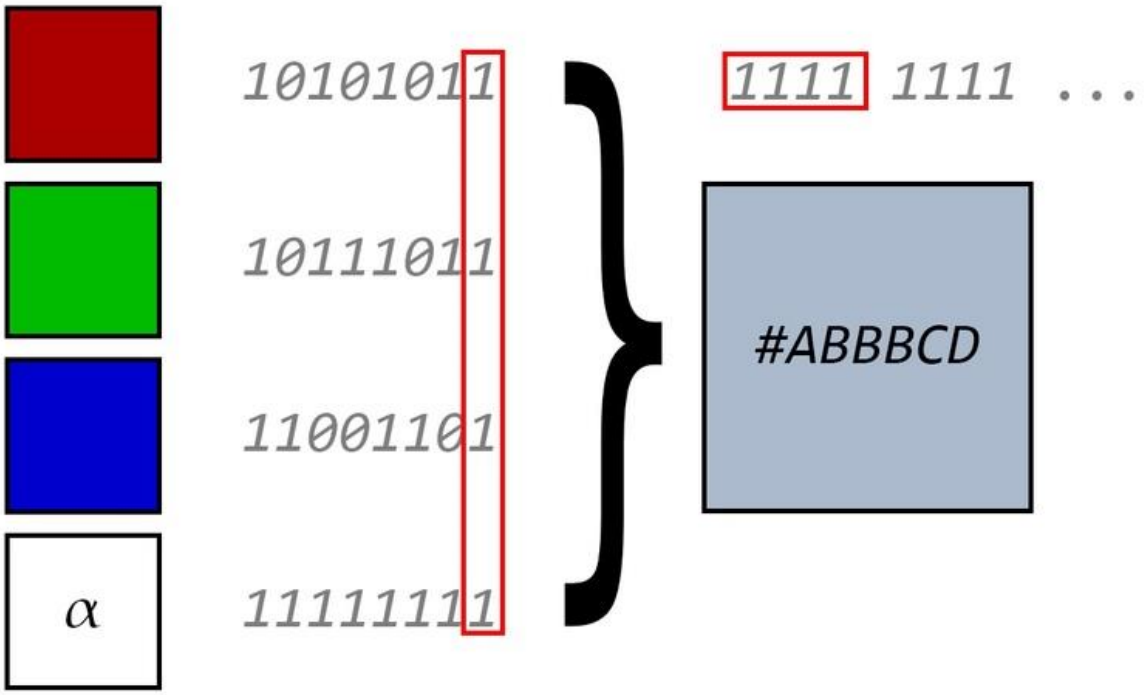
Встраивание

Как известно, в цифровом изображении каждый цвет кодируется при помощи нескольких основополагающих цветов, иногда с использованием дополнительных каналов. В зависимости от конечного применения это могут быть модели RGB (красный, зеленый, синий), RGBA (красный, зеленый, синий, альфа-канал) или, например, CMYK (циан, маджента, желтый, черный). Каждый цвет можно представить в виде последовательности чисел, где для каждого канала-цвета определено свое числовое значение.



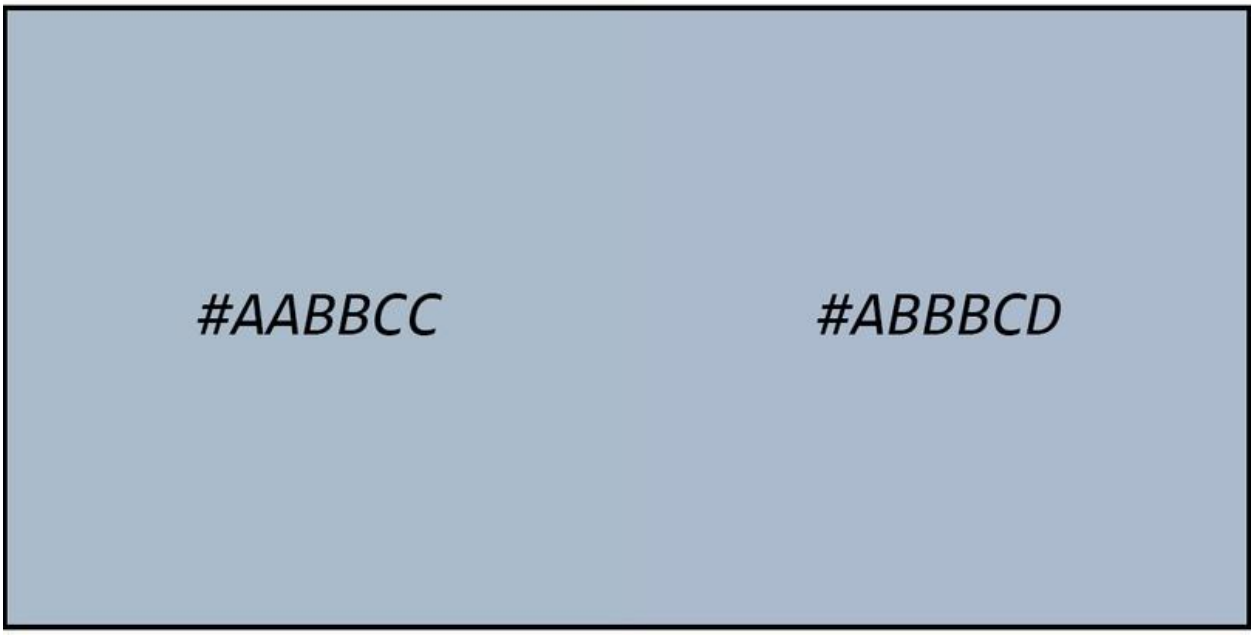
Кодирование одного цвета в формате RGBA (8 бит).

Для «упаковки» секретного сообщения в пиксели изображения необходима некоторая подготовка: сообщение преобразуется в битовую цепочку и делится «порциями» по n бит, где n — количество каналов, кодирующих итоговое изображение. Например, при «упаковке» сообщения в кодировке UTF-8 в изображение вида RGBA для записи одного символа текста понадобится 2 пикселя (т.к. 1 символ «весит» 1 байт, а цветовых канала мы имеем 4).



«Упаковка» сообщения в пиксель в формате RGBA (8 бит).

Вполне очевидно, что при таком незначительном изменении цветов отличить один пиксель от другого «на глаз» не представляется возможным.



Видите разницу? А она есть!

Поскольку крайне редко секретное сообщение имеет тот размер, который идеально вписывается в изображение, то нам необходимо записать некоторый «шум» в неиспользованные пиксели. В качестве «шума» подойдет любая цепочка бит, например, можно повторно записать секретное сообщение.

Извлечение

Для извлечения секретного сообщения следует знать следующие данные:

- Количество младших бит, используемых для «упаковки» сообщения.
- Исходная кодировка сообщения.
- Порядок кодирования бит сообщения по каналам (если каналов больше одного).
- Порядок кодирования пикселей (например, слева направо, сверху вниз).
- Длина сообщения либо любая последовательность, указывающая на конец сообщения (для предотвращения попытки извлечь данные из случайного шума).

Само извлечение происходит следующим образом:

1. Из каждого цветового канала каждого пикселя извлекается значение младших бит и склеивается в одну цепочку.
2. Цепочка бит декодируется в сообщение при помощи используемой отправителем сообщения кодировки.
3. В случае, если имеются данные о длине сообщения, то извлечение заканчивается в момент получения сообщения известной длины. Если же для указания окончания используется парольная фраза, то извлечение заканчивается в тот момент, когда последние n символов извлеченного сообщения соответствуют парольной фразе.

5. **Какие существуют методы стегоанализа.**

Статистический анализ:

- Гистограммы: Анализ распределения цветовых значений или значений аудиосигналов для выявления аномалий.

- Анализ частотного спектра: Выявление изменений в частотных компонентах аудиофайлов или изображений, которые могут указывать на скрытую информацию.

Анализ наименее значимых битов (LSB):

- Chi-квадрат тест (χ^2-тест): Проверка равномерности распределения значений LSB для выявления аномалий.

- Анализ последовательностей битов: Поиск неестественных последовательностей в наименее значимых битах.

Методы на основе машинного обучения:

- Классификация: Обучение моделей для распознавания изображений или аудио с встраиванием и без встраивания скрытой информации.

- Кластеризация: Группировка медиафайлов на основе схожих характеристик, чтобы выявить потенциальные аномалии.

Анализ визуальных артефактов:

- Сравнительный анализ: Сравнение оригинальных и подозреваемых файлов на наличие визуальных артефактов, вызванных встраиванием.

- Анализ остаточных сигналов: Выявление следов, оставленных стеганографическими алгоритмами.

Анализ метаданных:

- Изучение заголовков и метаданных файлов: Поиск нестандартных или необычных записей, которые могут указывать на наличие скрытой информации.

Методы на основе измененной корреляции:

- Проверка корреляции пикселей: Анализ корреляции между соседними пикселями для выявления нарушений, вызванных встраиванием данных.