Исследование и разработка методов маркирования текстовых документов

Якушев Алексей

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Кафедра системного программирования

Научный руководитель к. т. н. Ю. В. Маркин

Москва, 2020 г.

Расследование утечек конфиденциальной информации

Data Leakage Prevention

Data Leakage Prevention (DLP) – совокупность технологий предотвращения и/или расследования утечек данных.

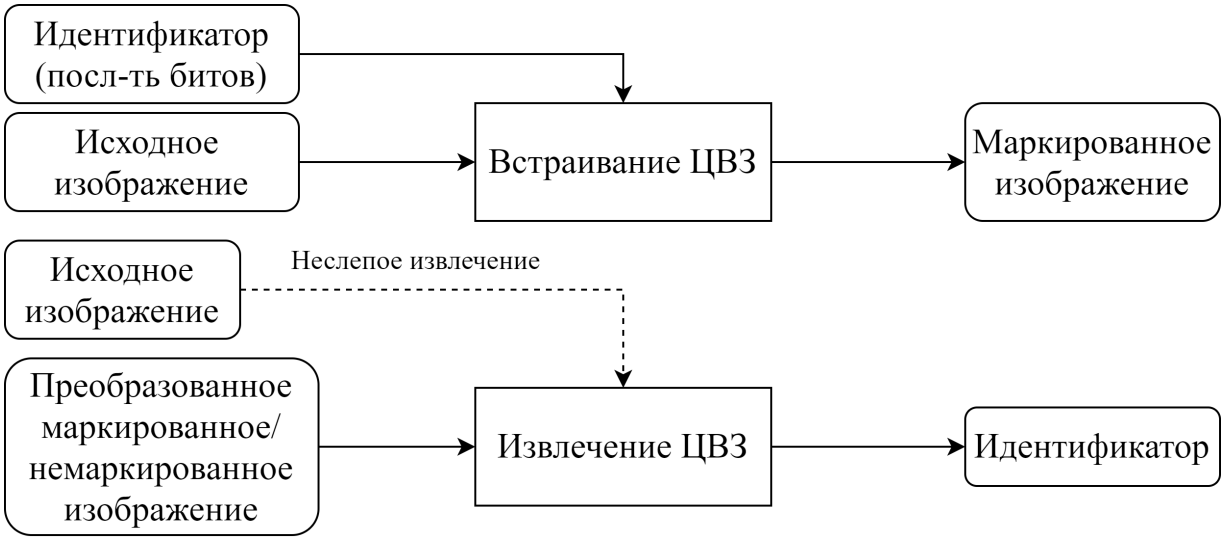
Проблема

Необходимо расследовать утечки конфиденциальных текстовых документов, вызванные фотографированием экрана и последующей публикацией фотографий в публичных источниках.

Решение

Встраивать идентификатор пользователя устройства в изображение документа, выводимого на экран. Встраивание идентификатора пользователя в изображение относится к технологии внедрения цифровых водяных знаков (ЦВЗ).

Общие сведения о ЦВЗ



Сценарии использования ЦВЗ

- screen-cam
  Фотографирование изображения на экране
- print-scan
  Сканирование распечатанного изображения
- print-cam
  Фотографирование распечатанного изображения

Атаки на ЦВЗ в сценарии screen-cam

- Расположение камеры относительно экрана
- Фокусировка камеры
- Эффект муара
- Неравномерная яркость на фотографии
- Сжатие фотографии

Существующие решения

Методы маркирования текстовых документов

- Изменение длины пробелов между словами. Биты кодируются соотношением суммарной длины половин пробелов в строке.
  Zou D., Shi Y. «Formatted Text Document Data Hiding Robust to Printing, Copying and Scanning», 2005
- Слова в тексте заменяются синонимами в зависимости от кодируемого бита. Корректность замены проверяется с помощью Google n-gram corpus.
  Chang C.-Y., Clark S. «Practical Linguistic Steganography using Contextual Synonym Substitution and a Novel Vertex Coding Method», 2014
- Добавление двух симметричных групп пикселей к буквам. Расстояние между группами кодируют бит.
  Varna A.L., Rane S. «Data hiding in hard-copy text documents robust to print, scan and photocopy operations», 2009

Методы маркирования, устойчивые к атакам сценария screen-cam

- ЦВЗ внедряется в домен ДКП областей, задаваемых ключевыми точками, полученными с помощью алгоритма I-SIFT.
  Fang H., Zhang W.«Screen-Shooting Resilient Watermarking», 2019
- Изменение яркости областей на экране. Биты кодируются уменьшением/увеличением яркости круговых областей.
  Gugelmann D., Sommer D.«Screen watermarking for data theft investigation and attribution», 2018

Постановка задачи

Разработать и реализовать метод маркирования текстовых документов, отображаемых на экране монитора, а также метод извлечения ЦВЗ с фотографии экрана. Метод должен удовлетворять требованиям:

- Незаметность для пользователя устройства
- Встраивание в режиме реального времени
- Независимость от формата файла докумета
- Устойчивость к атакам сценария screen-cam
- Работа в режиме слепого извлечения (без оригинала документа)

Разработанный подход к маркированию документов

- ЦВЗ представляет собой шаблон, состоящий из прямоугольных областей разной яркости, называемых маркерами
  - Светлые маркеры кодируют бит «0», темные — «1»
  - Кодирующие маркеры чередуются с промежуточными для упрощения извлечения ЦВЗ
  - В начало, середину и конец шаблона встраиваются маркеры, состоящие из двух частей. Эти маркеры позволяют определить положение шаблона при извлечении ЦВЗ
  - На шаблон накладывается фильтр Гаусса для сглаживания переходов между маркерами, что снижает заметность ЦВЗ
- Шаблон отображается при помощи частично прозрачного окна, лежащего поверх остальных окон
- Шаблон встраивается в области межстрочных интервалов нижележащих окон

Разработанный подход к маркированию документов

Алгоритм встраивания ЦВЗ

1. Создание снимка экрана
2. Очистка снимка экрана от предыдущей цифровой метки
3. Определение частей окон, видимых пользователем
4. Определение областей с текстом
5. Определение областей межстрочных интервалов
6. Генерация шаблона яркости, соответствующего внедряемой информации
7. Встраивание шаблона на оверлей в области межстрочных интервалов маркируемого текста.

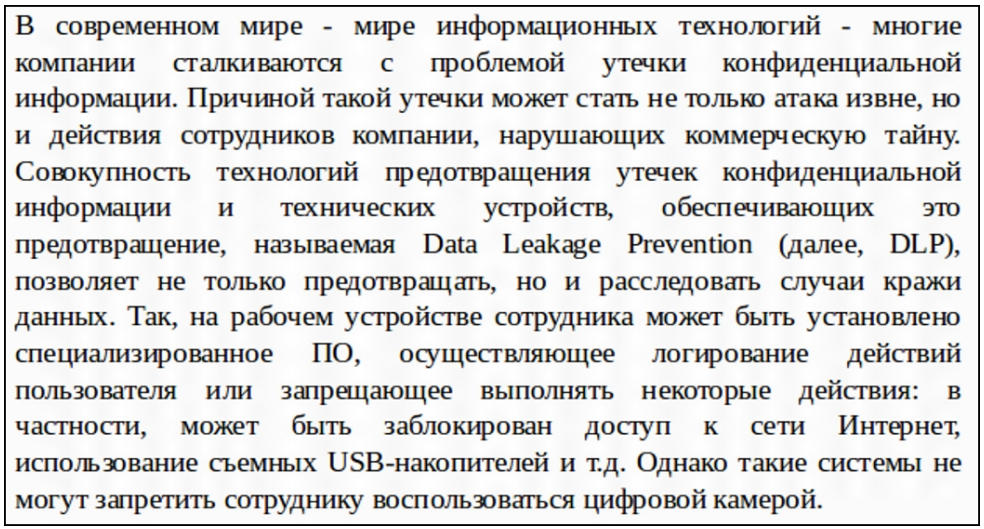
Алгоритм извлечения ЦВЗ

1. Коррекция перспективы и обрезка фотографии
2. Определение областей с текстом
3. Определение областей межстрочных интервалов
4. Удаление из межстрочных интервалов фрагментов букв
5. Определение наличия ЦВЗ в межстрочных интервалах
6. Извлечение ЦВЗ из межстрочных интервалов
7. Объединение меток, извлеченных из разных межстрочных интервалов и текстовых областей.

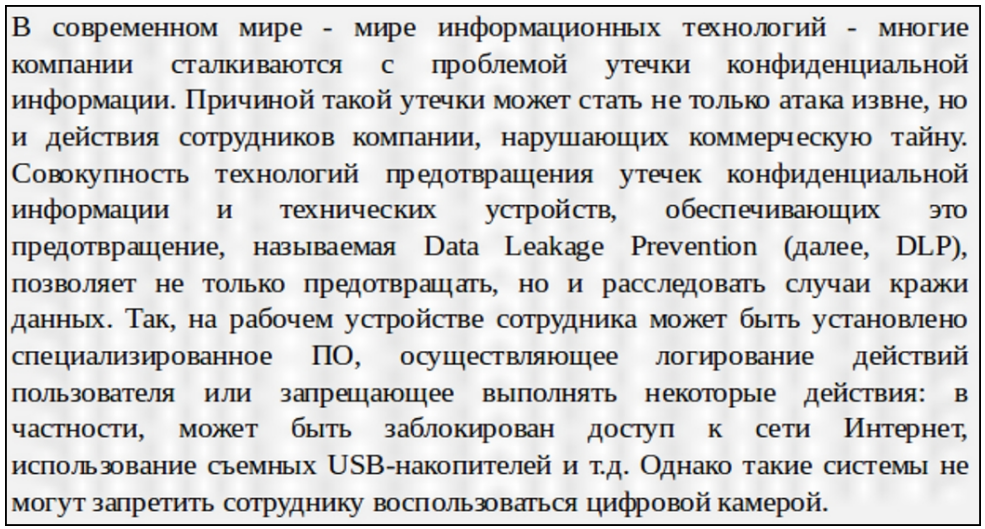
Результаты работы алгоритма встраивания ЦВЗ

Маркеры цифровой метки могут быть встроены с разной непрозрачностью (opacity). Повышение непрозрачности маркеров делает ЦВЗ более устойчивым к атакам, но в то же время ведет к большей заметности метки.

- Непрозрачность 0% — цвет маркеров совпадает с цветом фона
- Непрозрачность 100% — пользователь видит на экране маркеры черного и белого цвета



Непрозрачность маркеров 3%



Непрозрачность маркеров 10%

Результаты работы алгоритма извлечения ЦВЗ

Извлекаемое значение k-го бита сообщения зависит от яркости k-го кодирующего маркера.

- GS_k — оценка отклонения яркости маркера от среднего значения яркости в межстрочном интервале
- CS_k — оценка выпуклости функции яркости около положения кодирующего маркера в межстрочном интервале

Тогда общая оценка яркости кодирующего маркера:

$$S_k = GS_k + \alpha \cdot CS_k,\ \alpha > 0$$

Значение бита $m_k,\ k \in [1, N]$, соответствующего k-му кодирующему маркеру, вычисляется как:

$$m_k = \begin{cases} 0, & \text{если } S_k > 0; \\ 1, & \text{иначе.} \end{cases}$$

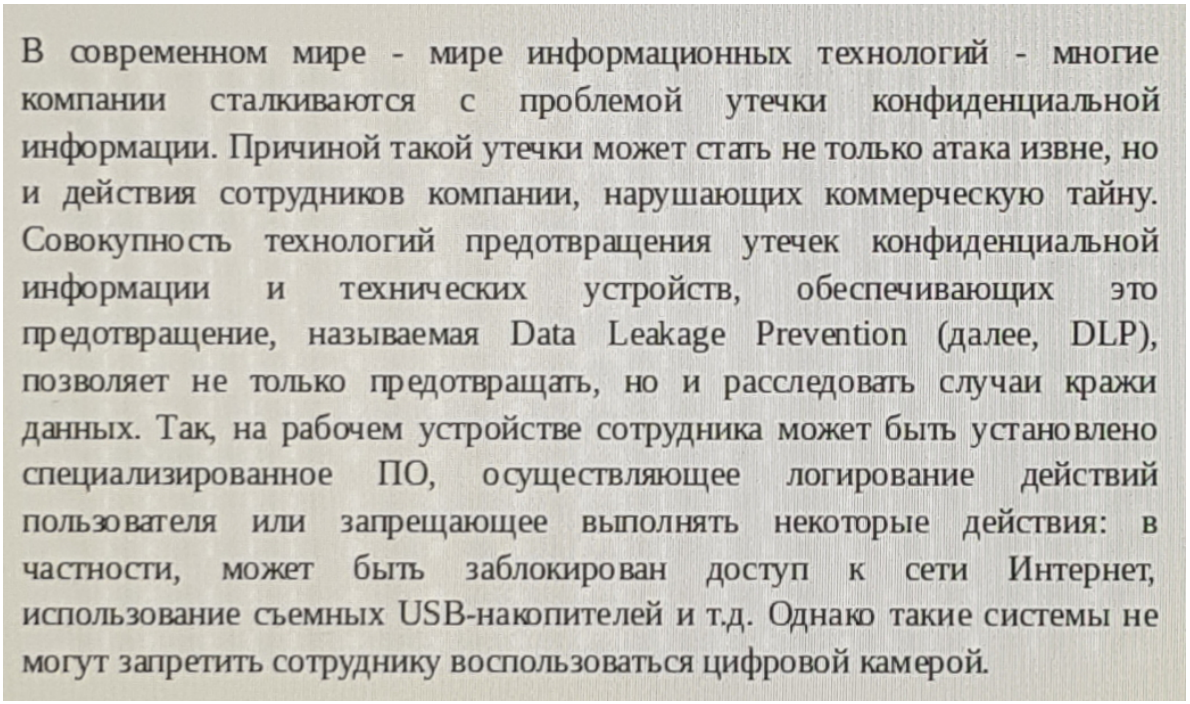
$|S_k|$ характеризует степень уверенности в извлеченном значении.

Bit Error Rate (BER) позволяет оценить успешность извлечения ЦВЗ.

$$BER = \frac{\{\text{число неверно извлеченных бит цифровой метки}\}}{\{\text{общее число бит цифровой метки}\}} \cdot 100\%$$

Описание эксперимента

- Текст: 14 строк, 13 межстрочных интервалов, кегль шрифта 14 пт, множитель межстрочного интервала 1.15, масштаб текста 100%
- Монитор AOC-i2769Vm: диагональ 27 дюймов, разрешение 1920×1080 пикселей, тип матрицы IPS
- Камера смартфона Samsung Galaxy S8: 12 мегапикселей, апертура f/1.7, фокусное расстояние 26 мм
- Во все межстрочные интервалы текста встраивалась одинаковая цифровая метка, состоящая из 24 бит


В современном мире - мире информационных технологий - многие компании сталкиваются с проблемой утечки конфиденциальной информации. Причиной такой утечки может стать не только атака извне, но и действия сотрудников компании, нарушающих коммерческую тайну. Совокупность технологий предотвращения утечек конфиденциальной информации и технических устройств, обеспечивающих это предотвращение, называемая Data Leakage Prevention (далее, DLP), позволяет не только предотвращать, но и расследовать случаи кражи данных. Так, на рабочем устройстве сотрудника может быть установлено специализированное ПО, осуществляющее логирование действий пользователя или запрещающее выполнять некоторые действия: в частности, может быть заблокирован доступ к сети Интернет, использование съемных USB-накопителей и т.д. Однако такие системы не могут запретить сотруднику воспользоваться цифровой камерой.

Результаты тестирования

Извлекаемость оценивалась с помощью

- Оценка BER-24 рассчитывалась после комбинирования меток, извлеченных из разных межстрочных интервалов
- Оценка BER-312 рассчитывалась при предположении, что биты, встроенные в разные межстрочные интервалы, независимы
- BER-24 $= 0$ и BER-312 $< 15\%$ — метка извлечена успешно
- BER-24 > 0 или BER-312 $\geqslant 15\%$ — метка извлечена с ошибками

Эксперимент	Метка успешно извлечена	Метка извлечена с ошибками	Метка не обнаружена
Непрозрачность маркеров	$\geqslant 3\%$	–	$< 3\%$
Расстояние от камеры до экрана	30–40 см, 60–100 см	50 см	$\leqslant 25$ см
Угол между камерой и экраном	$\leqslant 45°$	$60°$	$\geqslant 75°$
Коэффициент качества JPEG	20–100	15	$\leqslant 10$

Заключение

Полученные результаты

- Разработан и реализован алгоритм маркирования текстовых документов, отображаемых на экране монитора.
- Разработан и реализован алгоритм извлечения цифровой метки с фотографии экрана.
- Метод показал высокие результаты устойчивости к атакам сценария screen-cam.

Дальнейшие исследования

- Развитие и оптимизация предложенного метода
- Тестирование метода при других конфигурациях экран/камера
- Разработка альтернативных схем маркирования текстовых документов

К публикации

Планируется публикация по теме диплома.