

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ
компьютерной безопасности и
криптографии

Безопасность при использовании графических файлов

АВТОРЕФЕРАТ

дипломной работы

студента 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Жильцова Вячеслава Алексеевича

Научный руководитель

доцент

И.Ю. Юрин

22.01.2022 г.

Заведующий кафедрой

д. ф.-м. н., доцент

М. Б. Абросимов

22.01.2022 г.

Саратов 2022

ВВЕДЕНИЕ

Огромное количество изображений, распространяемых через Интернет, передаются в форматах PNG, JPG и TIFF. Эти изображения доступны многим людям без какой-либо защиты личной информации или контроля распространения. В последние годы социальные сети и облачные вычисления ускоряют обмен изображениями. Кроме того, многие портативные устройства имеют функции связи, которые позволяют мгновенно распространять фотографии после их захвата. В сочетании, например, с потенциально включенной информацией GPS в формате файла, фотография может предоставлять личную информацию и информацию о географическом местоположении миру.

Поиск метаданных в файлах фотографий — одно из звеньев доксинга (англ. «doxing», от «docs»), ставшей уже весьма популярной практики сбора сведений об интересующем человеке в Интернет-источниках в тех или иных целях.

Один из главных «коллекционеров» метаданных — блок EXIF, добавляемый к графическим файлам. Стандарт Exchangeable Image File Format разработан Японской ассоциацией электронной и ИТ-индустрии JEITA, его первая версия была опубликована в 1995 году. EXIF предназначен для графических форматов JPEG и TIFF. Файлы других популярных форматов, таких как PNG и GIF, тоже могут содержать подобные метаданные — в частности, по стандарту XMP, разработанному Adobe. Кроме того, производители камер используют собственные форматы метаданных, отчасти дублирующие EXIF.

Присутствующие в файлах с фотографиями метаданные не раз приносили вред авторам или действующим лицам снимков. Один из самых ярких примеров — арест Джона Макафи в Гватемале в 2012 году. Скрываясь от следствия по делу об убийстве его соседа, Макафи дал интервью изданию Vice,

проиллюстрированное фотографией. По метаданным этого снимка и был найден Джон Макафи: геометка, обнаружившаяся в файле снимка, помогла правоохранителям найти и схватить беглеца.

Во избежание таких нежелательных ситуаций в структуре графических файлов должны быть предусмотрены механизмы защиты конфиденциальности и безопасности. Например, шифрование является хорошо разработанным подходом для полной и / или частичной защиты данных изображения и метаданных в файлах JPEG.

Другой подход заключается в удалении метаданных из структуры изображений.

Цель данной курсовой работы заключается в изучении метаданных формата EXIF, исследовании с целью выяснить, какие веб-сервисы и социальные сети удаляют метаданные из графических файлов, а какие не удаляют. Также будет написана программа для удаления потенциально вредоносных для владельца метаданных из изображений.

Дипломная работа состоит из введения, 7 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 54 страницы, из них 38 страниц – основное содержание, включая 53 рисунков и 3 таблицы, список использованных источников из 14 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел дипломной работы «Безопасность при использовании графических файлов» содержит описание графического формата TIFF.

TIFF - это гибкий, адаптируемый формат файла для обработки изображений и данных в одном файле за счет включения тегов заголовков (размер, определение, расположение данных изображения, применяемое сжатие изображения), определяющих геометрию изображения. Например, файл TIFF может быть контейнером, содержащим сжатые изображения JPEG (с потерями) и PackBits (без потерь). Файл TIFF также может включать векторную траекторию отсечения (контуры, обрезки, рамки изображений). Возможность хранить данные изображения в формате без потерь делает файл TIFF полезным архивом изображений, поскольку, в отличие от стандартных файлов JPEG, файл TIFF, использующий сжатие без потерь (или его отсутствие), можно редактировать и повторно сохранять без потери качества изображения. Это не тот случай, когда TIFF используется в качестве контейнера, содержащего сжатый JPEG. Другие варианты TIFF - это слои и страницы.

Второй раздел работы содержит описание графического формата JPEG и алгоритма сжатия JPEG.

JPEG – один из достаточно мощных алгоритмов сжатия. Он оперирует областями 8x8, на которых цвет и яркость меняются относительно плавно. Таким образом, сжатие в JPEG осуществляется за счёт плавности изменения цветов в изображении.

JPEG (произносится «джейпег», англ. Joint Photographic Experts Group, по названию организации-разработчика) — один из популярных растровых графических форматов, применяемый для хранения фотоизображений и подобных им изображений. Файлы, содержащие данные JPEG, обычно имеют расширения (суффиксы) .jpg, .jfif, .jpe или .jpeg. Однако из них .jpg является самым популярным на всех платформах. MIME-типом является image/jpeg.

Алгоритм JPEG позволяет сжимать изображение как с потерями, так и без потерь (режим сжатия lossless JPEG). Поддерживаются изображения с линейным размером не более 65535×65535 пикселей.

JPEG – один из популярных растровых графических форматов, применяемый для хранения фотоизображений и подобных им изображений. Обычно файлы формата JPEG имеют расширения **.jpg**, **.jif**, **.jpe** или **.jpeg**. Из них **.jpg** является самым популярным.

Файл JPEG состоит из последовательности маркеров, каждый из которых начинается с байта 0xFF, свидетельствующего о начале маркера, и байта-идентификатора.

Маркеры также могут содержать дополнительные данные, которые состоят из двухбайтового поля. В этом поле содержатся длина информационной части маркера и сами данные. За некоторыми маркерами следуют энтропийно-кодированные данные.

Стоит заметить, что последовательные байты 0xFF используются для заполнения, однако это заполнение должно происходить для маркеров, сразу следующих за данными с энтропийным сканированием.

Внутри энтропийно закодированных данных после любого байта 0xFF кодировщиком вставляется байт 0x00 перед следующим байтом. Это делается для того, чтобы в неподобающем месте не появлялся маркер. Это помогает избежать ошибок при кадрировании. Декодеры пропускают байт 0x00.

Такая техника действует только для энтропийно закодированных данных.

На рисунке 1 отображены основные маркеры, встречающиеся в структуре JPEG-файла.

Тип маркера	Идентификатор	Обозначение стандарта	Определение
SOF ₀	C ₀₁₆	Baseline DCT	Начало кадра, базовый метод
SOF ₁	C ₁₁₆	Extended sequential DCT	Начало кадра, расширенный, последовательный метод
SOF ₂	C ₂₁₆	Progressive DCT	Начало кадра, прогрессивный метод
SOF ₃	C ₃₁₆	Lossless (sequential)	Начало кадра, метод сжатия без потерь
SOF ₅	C ₅₁₆	Differential sequential DCT	Начало кадра, дифференциальный последовательный метод
SOF ₆	C ₆₁₆	Differential progressive DCT	Начало кадра, дифференциальный прогрессивный метод
SOF ₇	C ₇₁₆	Differential lossless (sequential)	Начало кадра, дифференциальный метод без потерь
JPG	C ₈₁₆	Reserved for JPEG extensions	Резерв для последующих расширений формата JPEG
SOF ₉	C ₉₁₆	Extended sequential DCT	Начало кадра, расширенный последовательный метод, арифметическое кодирование
SOF ₁₀	C _{A16}	Progressive DCT	Начало кадра, прогрессивный метод, арифметическое кодирование
SOF ₁₁	C _{B16}	Lossless (sequential)	Начало кадра, метод без потерь, арифметическое кодирование
SOF ₁₃	C _{D16}	Differential sequential DCT	Начало кадра, дифференциальный последовательный метод, арифметическое кодирование
SOF ₁₄	C _{E16}	Differential progressive DCT	Начало кадра, дифференциальный прогрессивный метод, арифметическое кодирование
SOF ₁₅	C _{F16}	Differential lossless (sequential)	Начало кадра, дифференциальный метод без потерь, арифметическое кодирование
DAC	C _{C16}	Define arithmetic coding conditioning(s)	Определение условий арифметического кодирования
DHT	C ₄₁₆	Define Huffman table(s)	Определение таблиц Хаффмана
RST ₀ ...RST ₇	D ₀₁₆ ...D ₇₁₆	Restart marker number 0...7	Определение интервала перезапуска от 0 до 7
SOI	D ₈₁₆	Start of image	Начало изображения
EOI	D ₉₁₆	End of image	Конец изображения
SOS	D _{A16}	Start of scan	Начало скана
DQT	D _{B16}	Define quantization table(s)	Определение таблиц квантования
DNL	D _{C16}	Define number of lines	Определение числа линий
DRI	D _{D16}	Define restart interval	Определение интервала перезапуска

Рисунок 1 – основные маркеры

Третий раздел работы содержит описание графического формата PNG и алгоритма сжатия Deflate.

PNG (англ. portable network graphics) – растровый формат хранения графической информации, использующий сжатие без потерь по алгоритму Deflate.

PNG был разработан как улучшенная, не запатентованная замена формату GIF.

PNG поддерживает изображения на основе палитры (с палитрами 24-битных цветов RGB или 32-битных цветов RGBA), изображения в градациях серого (с альфа-каналом или без него для прозрачности) и полноцветные изображения RGB или RGBA без палитры. Разработчики PNG создали формат для передачи изображений в Интернете, а не для печатной графики профессионального качества; поэтому цветовые пространства, отличные от RGB, такие как CMYK, не поддерживаются. Файл PNG содержит одно изображение в расширяемой структуре блоков, кодирование основных пикселей и другую информацию, такую как текстовые комментарии и проверки целостности, задокументированные в RFC 2083.

Deflate — это алгоритм сжатия без потерь, использующий комбинацию алгоритмов LZ77 и Хаффмана. Изначально был описан Филом Кацем для

второй версии его архиватора PKZIP, который впоследствии был определён в RFC 1951 (1996 год).

Deflate считается свободным от всех существующих патентов, и пока оставался в силе патент на LZW (он применяется в формате GIF), это привело к использованию Deflate не только в формате ZIP, для которого Кац изначально его спроектировал, но также в компрессоре/декомпрессоре gzip и в PNG-изображениях.

В четвёртом разделе описывается формат метаданных EXIF.

EXIF (англ. Exchangeable Image File Format) — стандарт, позволяющий добавлять к изображениям и прочим медиафайлам дополнительную информацию (метаданные), комментирующую этот файл, описывающий условия и способы его получения, авторство и т. п. Получил широкое распространение в связи с появлением цифровых фотокамер. Информация, записанная в этом формате, может использоваться как пользователем, так и различными устройствами, например, принтером. Стандарт EXIF является чрезвычайно гибким (например, позволяет сохранить полученные с приёмника GPS координаты места съёмки) и допускает широкое развитие — как правило, фотоаппараты добавляют к файлу информацию, специфичную только для данной конкретной камеры. Правильно интерпретировать такую информацию могут только программы от изготовителя фотоаппарата.

Разработчик формата — Japan Electronics and Information Technology Industries Association[en] (JEITA). История версий формата EXIF указана в таблице 1.

EXIF является частью более широкого стандарта DCF.

Таблица 1 – история версий формата EXIF

Дата	Версия	Описание
Октябрь 1995	1	Установлен как стандарт JEIDA. Определяется структура, состоящая из формата данных изображения и информации об атрибутах (теги) и основных тегов.
Ноябрь 1997	1.1	Сохранены основные положения версии 1.0 и добавлены положения для необязательной информации об атрибутах и операции форматирования
Июнь 1998	2	Добавлено цветовое пространство sRGB, сжатые миниатюры и аудиофайлы
Декабрь 1998	2.1	Обновлен и расширен формат хранения и атрибут информации. Добавлены рекомендуемые сведения о совместимости в качестве дополнения к версии 2.0
Февраль 2002	2.2	Добавлена информация в версию 2.1 для улучшения качества печати
Сентябрь 2003	2.21	Добавлено необязательное цветовое пространство (Adobe RGB)

Пятый раздел посвящён исследованию веб-сервисов и социальных сетей на предмет удаления метаданных из графических файлов при загрузке.

Таблица 2 – результаты исследования

Название сервиса / формат изображения	JPG	PNG	TIFF
Mail.ru	-	-	-
Google.com	-	-	-
Облако Mail.ru	-	-	-
Яндекс Диск	-	-	-
Google Drive	-	-	-
Facebook	+	+	+
Twitter	+	+	
Вконтакте	+	+	
Telegram	+	-	-
Viber	+	-	-
Instagram	+	+	
Flickr	-	-	-
Tumblr	+	+	
Google Photo	+	-	+
Авито	+	+	
Юла	+	+	
КупиПродай	+	+	
hh.ru	+	-	

В шестом разделе описывается исследование программ и веб-сервисов, предназначенных для удаления метаданных из графических файлов. Были рассмотрены следующие программы и веб-сервисы: ExifTool, Exif Remover,

Verexif.com, Exif Tag Remover . Был сделан вывод, что ни одна из них не может удобно распознавать, а затем удалять метаданные из изображений любого формата.

Седьмой раздел посвящён программной реализации удаления метаданных из графических файлов. В результате была написана программа, удаляющая метаданные из графических файлов формата JPG, PNG, TIFF. Программа удаляет не все метаданные, а только потенциально самые полезные для злоумышленников: дата и время съёмки или создания, изготовитель и модель устройства, GPS-координаты, данные о ПО и т.д. Пользователь может сам выбрать, какие метаданные удалить, а какие оставить на месте. Реализованы следующие возможности: открытие графического файла из каталога с последующим выводом сообщения об успешном или неудачном открытии файла, удаление метаданных из файла с помощью специальной кнопки, сохранение изображения в файле. В программе использован графический интерфейс Tkinter.

Полный код программы приведен в Приложении А.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены особенности формата графических изображений, во второй части были рассмотрены особенности формата метаданных EXIF, исследованы различные веб-сервисы и приложения на предмет удаления метаданных из графических файлов.

В ходе работы была написана программа на языке Python, реализующая удаление потенциально опасных для владельца метаданных из графических изображений. Обзор программы дан в разделе 7, а полный листинг программы размещён в приложении А.