УДК 004.056.53

С. В. Ковригина[1]*

Модуль обнаружения контейнеров цифровой стеганографии, функционирующий на базе нейронных сетей

[1] Национальный исследовательский ядерный университет «МИФИ», г. Москва, Российская Федерация
* e-mail: sofkakovr@gmail.com

Аннотация. В данной работе рассмотрен модуль для обнаружения контейнеров цифровой стеганографии, функционирующий на базе нейронных сетей. Освещена проблематика скрытого извлечения данных с использованием стеганографических методов в современных информационных системах и предложен новый подход к решению этой задачи. Объектом исследования выступают методы и инструменты для выявления цифровой стеганографии. Цель работы заключается в разработке и внедрении модуля, который позволит эффективно обнаруживать скрытые данные, обеспечивая тем самым повышение уровня информационной безопасности. В статье подчеркивается значимость разработки таких решений на фоне растущих рисков утечек данных. В результате исследования были разработаны методы для детектирования цифровой стеганографии, сформирована архитектура модуля в виде диаграмм и определен комплект необходимых компонентов. В рамках проектирования и программной реализации модуля были использованы современные технологии глубокого обучения, что позволило достичь высокой точности в обнаружении стеганографических вмешательств. Предложенное решение способствует повышению уровня информационной безопасности в коммерческих и частных информационных средах за счет точного обнаружения скрытых каналов утечки данных и может быть интегрировано в существующие системы защиты для эффективной борьбы с цифровой стеганографией.

Ключевые слова: нейронные сети, стеганография, LSB, метод наименее значимого бита

S. V. Kovrigina[1]*

Digital Steganography Container Detection Module Based on Neural Networks

[1] National research nuclear university MEPhI, Moscow, Russian Federation
* e-mail: sofkakovr@gmail.com

Abstract. This paper presents a module for detecting digital steganography containers, operating on the basis of neural networks. It addresses the problem of hidden data extraction through steganographic methods in modern information systems and proposes a new approach to solving this issue. The research focuses on methods and tools for detecting steganographically contained data. The aim of the work is to develop and implement a module that allows for effective detection of hidden data, thereby improving the protection of information security. The article emphasizes the importance of developing such solutions against the backdrop of increasing data leakage risks. As a result of the research, methods for detecting digital steganography were developed, the architecture of the module was formed in diagrams, and a set of necessary components was determined. During the design and software implementation of the module, modern deep learning technologies were used, which allowed achieving high accuracy in detecting steganographic interventions. The proposed solution contributes to enhancing the level of information security in both commercial and private in-

formation environments by precisely detecting hidden data leakage channels and can be integrated into existing security systems for effective combat against digital steganography.

Keywords: neural network, steganography, LSB, least significant bit method

Введение

Утечки персональных данных и корпоративный шпионаж представляют собой значительную угрозу для организаций [1]. Современные методы извлечения ценной информации из внутренних сетей предприятий уже не включают использование чатов и файлообменников ввиду их высокой распознаваемости существующими системами предотвращения утечек данных (DLP-системы). Вместо этого злоумышленники применяют альтернативные методы. Один из распространенных подходов заключается во внедрении конфиденциальной информации в изображения с использованием различных техник стеганографии, в том числе метода наименьшего значащего бита (LSB).

Цифровая стеганография представляет собой метод скрытого встраивания информации в цифровые изображения, что является как полезным инструментом в области информационной безопасности, так и потенциальной угрозой, поскольку может быть использована для неправомерной передачи конфиденциальных данных [2]. Современные методы обнаружения стеганографии часто неэффективны против сложных техник встраивания, которые могут быть невидимы для невооруженного глаза или стандартного анализа изображений [3].

Модуль выявления контейнеров цифровой стеганографии разрабатывается с целью последующего его внедрения в уже существующие DLP-решения. В отличие от традиционных методов данный подход не полагается на обнаружение заранее известных сигнатур стеганографии, что делает его более гибким и эффективным в обнаружении новых и неизвестных методов встраивания информации.

В данной статье рассмотрены предлагаемые научным сообществом решения, позволяющие определять, является ли изображение контейнером цифровой стеганографии. Рассмотрены технологические аспекты, примеры использования и потенциальные преимущества такого подхода.

Методы и материалы

В данном разделе будут рассмотрены существующие работы научного сообщества и конфигурация разработанной нейронной сети.

Научным сообществом разрабатываются новые методы для обнаружения скрытых посредством стеганографии данных в изображениях. Для решения данной задачи используются нейронные сети.

Все решения, приводимые в данных исследованиях, ограничены палитрой изображений – исключительно черно-белой, а также ограничены поддерживаемым форматом изображений – BMP, который не имеет широкого распространения в современном мире.

Для анализа были выбраны такие характеристики, как архитектура нейронной сети, BPP (от англ. bits per pixel), разрешения анализируемых изображений и точность модели. В научных работах в качестве архитектуры использовалась ANN (от англ. Artificial Neural Network) или CNN (от англ. Convolutional Neural Network). Результаты анализа существующих решений представлены в табл. 1.

Таблица 1

Сравнительная таблица существующих научных исследований

Метод	Источник	Год	BPP	Разрешение	Количество наблюдений	Точность (%)
ANN	[4]	2016	нет данных	от 128x128 до 512x512	200	97-99
ANN	[5]	2017	0.1, 0.25	512x512	4800	86-90
CNN	[6]	2016	[0.1, 0.5]	512x512	10000	84-86
CNN	[7]	2017	0.4	512x512	40000	96-98
CNN	[8]	2017	0.1 и 0.5	512x512	11600	91-92

Все эти работы не предоставляют исходных кодов, поэтому собственноручно протестировать решения не удалось.

На основе изучения уже разработанных решений были выявлены главные проблемы, которые необходимо решить разрабатываемым решением:

- поддержка только формата BMP;
- поддержка одного алгоритма сокрытия;
- невозможность интеграции с DLP системой.

В разрабатываемом решении должна присутствовать поддержка современных форматов изображений, например, такого как PNG.

По итогу были определены статистические характеристики изображений, которые наилучшим образом отражают наличие в изображении скрытой информации. Ими стали [9, 10, 11]:

- стандартное отклонение [12];
- асимметрия;
- эксцесс гистограммы;
- мобильность Хьорта;
- медиана гистограммы;
- геометрическая медиана;
- сложность по Хьорту;
- диапазон значений гистограммы.

Далее была разработана блок схема алгоритма процесса обучения модуля обнаружения (рис. 1).

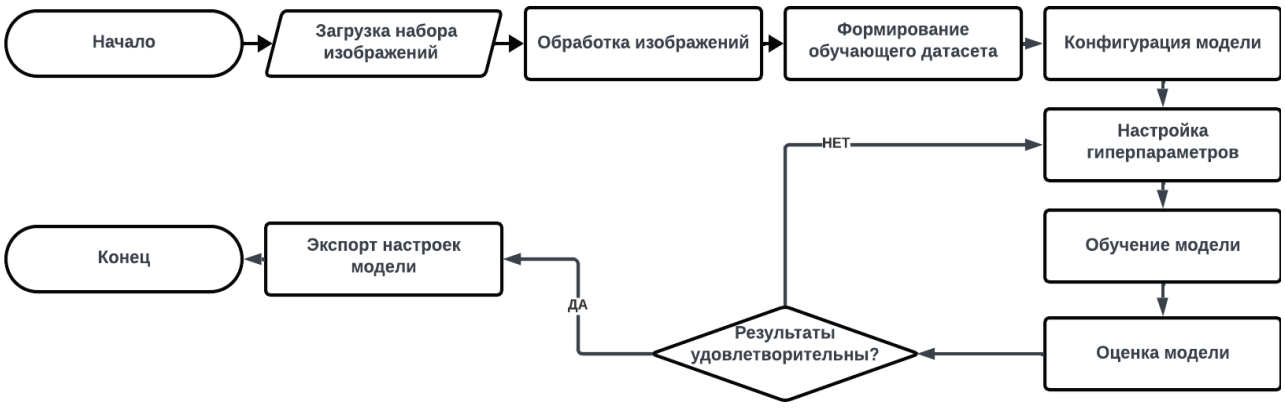

Рис. 1. Блок-схема алгоритма обучения модуля обнаружения

Так как при помощи контейнеров цифровой стеганографии часто происходит изъятие секретных корпоративных данных и персональных данных пользователей, то одним из вариантов использования этой модели нейронной сети может быть фильтрация входящей и исходящей корпоративной почты, а также мессенджеров.

Результаты

В ходе программной разработки была определена топология нейронной сети (рис. 2), созданы программные модули для обучения и использования нейронной сети, а также веб-интерфейс с API для взаимодействия с моделью.

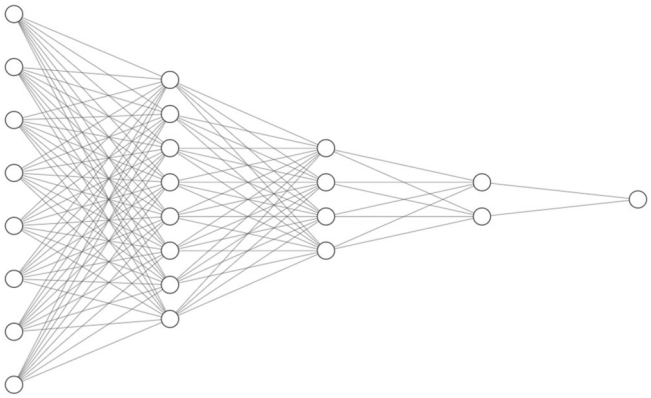

Рис. 2. Топология нейронной сети

Так как нейронная сеть должна выводить только один параметр, указывающий на то, является ли изображение контейнером цифровой стеганографии, то в качестве скрытых слоев были использованы сверточные, которые позволили перейти от 8 изначальных параметров (по количеству анализируемых статистических характеристик) к одному нейрону в выходном слое, который принимает значение в диапазоне от 0 до 1 включительно. Это значение показывает уверенность модели в том, что изображение содержит скрытые данные, где 0 – изобра-

жение не имеет скрытых данных и 1 – изображение является контейнером цифровой стеганографии.

Разработка велась на языке Python [13] и фреймворке TensorFlow [14], в процессе разработки были созданы несколько модулей (рис. 3):

- load_image.py: загрузка изображений на сервер;
- encode.py: скрытие информации для формирования обучающего набора;
- make_dataset.py: вычисление статистических характеристик;
- train.py: программа для обучения нейронной сети при помощи ранее созданного набора данных.

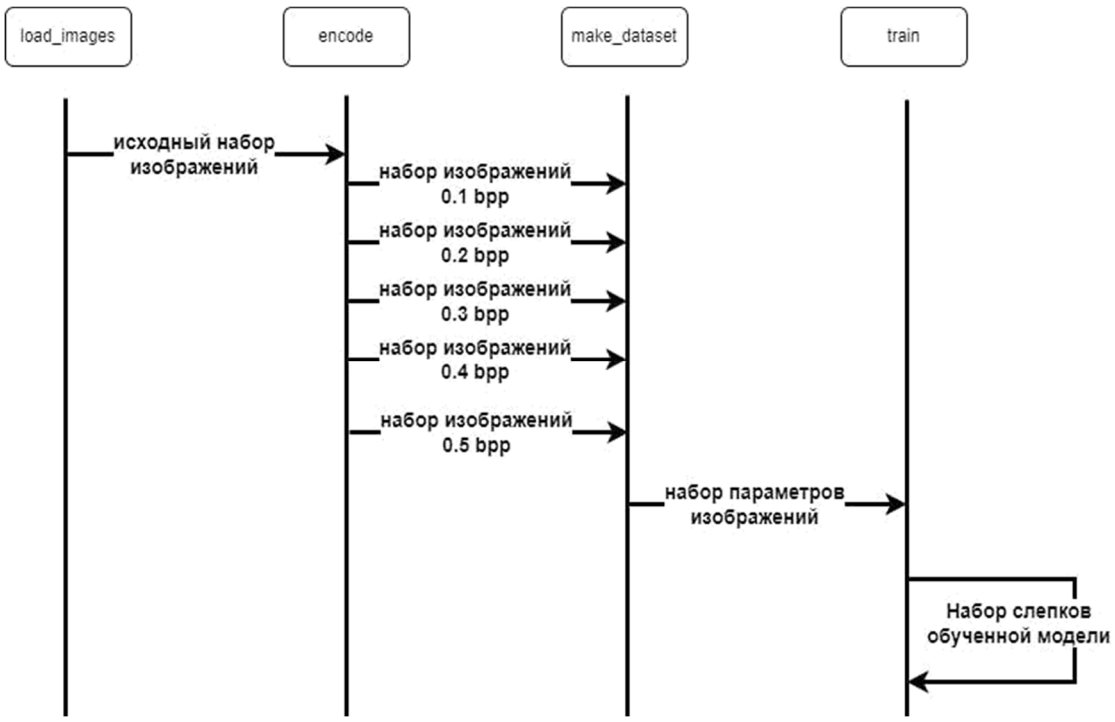

Рис. 3. UML-диаграмма взаимодействия разработанных модулей

В процессе тестирования главной целью было определить точность и полноту нейронной сети, а также скорость обработки изображений.

Результаты тестирования представлены в табл. 2.

Таблица 2

Результаты тестирования модели

Параметр	Значение
Время обучения	320 с
Скорость обработки изображения	50 мс
Скорость инициализации	1,5 с
Точность определения	94,5 %
Полнота	90 %

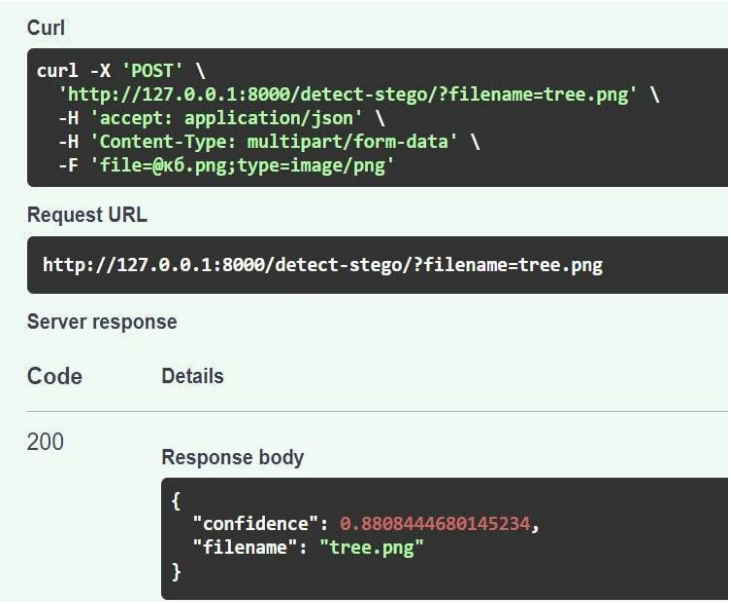

Рис. 4. Пример взаимодействия с веб-сервером

Также для удаленного взаимодействия с моделью был разработан веб-интерфейс на основе Python FastAPI [15], который позволяет при помощи HTTP запроса отправить изображение на анализ и получить ответ модели. Пример взаимодействия с веб-сервером продемонстрирован на рис. 4.

Обсуждение

В данной статье была представлен модуль обнаружения контейнеров цифровой стеганографии, функционирующий на базе нейронных сетей.

Существующие программные решения не позволяют достаточно эффективно определять, является ли изображение контейнером цифровой стеганографии. Они не позволяют обнаруживать способ сокрытия, отличный от того, что уже заложен в программу, и зачастую не имеют программного интерфейса взаимодействия, что очень сильно ограничивает сферу их применения.

Основной проблемой решений научного сообщества является поддержка исключительно черно-белых изображений, поддержка исключительно формата BMP и отсутствие интерфейса для взаимодействия с ними.

Особенностями разработанного решения являются довольно высокая точность обнаружения (94,5 %), возможность удаленного взаимодействия с моделью благодаря наличию веб-интерфейса и модуль конвертации изображений, который позволяет модели обрабатывать в том числе и изображения в формате PNG. Несмотря на значительные преимущества представленного решения, остаются некоторые нерешенные проблемы. Например, отсутствие возможности работы с изображениями формата JPEG, который использует сжатие с потерями, и его конвертация в формат BMP нетривиальна. Также требует дополнительной разработки интеграция с существующими DLP системами.

Заключение

В ходе исследования предметной области был проведен обзор научных статей в рассматриваемой предметной области. В рамках моделирования предметной области была определена структура нейронной сети, набор анализируемых статистических характеристик и набор компонентов, из которых будет состоять разрабатываемый модуль, разработаны алгоритмы обучения и использования нейронной сети. На основе полученных результатов был спроектирован модуль, который позволяет обнаруживать контейнеры цифровой стеганографии.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Juma'h A. H., Alnsour Y. The effect of data breaches on company performance //International Journal of Accounting & Information Management. – 2020. – Т. 28. – №. 2. – С. 275-301.

2. Williams O. C. What are the cybersecurity risks of artificial intelligence generated steganography? : дис. – Utica College, 2019.

3. Федосеев В. А. Цифровые водяные знаки и стеганография [Текст] / В.А. Федосеев // Издательство Самарского университета, 2019.

4. Ingale A.K., Dharwadkar N.V., Kodulkar P. Universal steganalysis using DWT and entropy // 2016 International Conference on Signal and Information Processing (IConSIP). – Vishnupuri, India, 2016.

5. Aljarf A., Amin S., Shuttelworth J., Filippas J. Detection system for gray and color images based on extracting features of difference image and renormalized histogram // Journal of Information Hiding and Multimedia Signal Processing. – 2017. – Vol. 8. – No. 2. – Pp. 16.

6. Qian Y., Dong J., Wang W., Tan T. Learning and representations for image steganalysis transferring using convolutional neural network // IEEE 2016 International Conference on Image Processing (ICIP). – Phoenix, AZ, USA, 2016.

7. Kim D.-H., Lee H.-Y. Convolutional neural network-based steganalysis on spatial domain // International Journal of Mathematics, Computer Science and Simulation. – 2017. – Vol. 11. – P. 5.

8. SSun Y., Zhang H., Zhang T., Wang R. Deep neural networks for efficient steganographic payload location // Journal of Real-Time Image Processing. – 2019. – Vol. 16. – No. 3. – Pp. 635–647.

9. Басыня Е. А., Сафронов А. В. Разработка и исследование системы управления метаданными изображений //Актуальные проблемы электронного приборостроения АПЭП-2016. – 2016. – С. 155-157.

10. Французова Г. А., Гунько А. В., Басыня Е. А. Применение искусственного интеллекта в сфере сетевой информационной безопасности // Под редакцией д. филос. н. ЕА Никитиной Рецензенты: д. ф. м. н., проф. ВГ Редько д. филос. н., проф. Т. Н Семенова. – 2013. – С. 110.

11. Грибунин В., Оков И., Туринцев И. Цифровая стеганография. – Litres, 2022

12. Федосеев В.А. Цифровые водяные знаки и стеганография [Текст] / В.А. Федосеев // Издательство Самарского университета, 2019.

13. Ирматова Д., Хаджаев С. Нейронные сети и искусственный интеллект на языке Python: обзор библиотек и фреймворков //Research and implementation. – 2023.

14. TensorFlow vs PyTorch: сравнение фреймворков глубокого обучения [Электронный ресурс] // Статья на сайте habr.com – Доступ к ресурсу URL: https://habr.com/ru/companies/ru_mts/articles/565456/

15. Басыня Е. А., Лукина М. С. Безопасность и анонимизация автоматизированной настройки серверных решений //Материалы конференций ГНИИ «Нацразвитие». – 2016. – С. 69-76.