БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра математического моделирования и анализа данных

Аннотация к магистерской диссертации

«Обнаружение вредоносных файлов с использованием методов машинного обучения»

Круглик Карина Сергеевна

Научный руководитель – кандидат физико-математических наук, доцент, заведующий НИЛ статистического анализа и моделирования НИИ ППМИ БГУ Абрамович М. С.

Минск, 2022

Реферат

Магистерская диссертация, 46 страниц, 13 рисунков, 12 таблиц, 24 источника.

Ключевые слова: ВРЕДОНОСНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ, СТАТИЧЕСКИЙ КОД, КЛАССИФИКАЦИЯ, КЛАСТЕРИЗАЦИЯ, МАШИННОЕ ОБУЧЕНИЕ, НЕЙРОННАЯ СЕТЬ, TF-IDF, ДОЛГАЯ КРАТКОСРОЧНАЯ ПАМЯТЬ.

Объектом исследования являются исполняемые файлы вредоносного и легитимного программного обеспечения.

Целью работы является разработка и реализация алгоритмов для обнаружения и классификации вредоносного программного обеспечения.

В ходе работы рассматриваются следующие подходы для классификации и кластеризации вредоносного программного обеспечения:

1) классификация и кластеризация на вредоносное и легитимное программное обеспечение, основанные на статических текстовых характеристиках байтового кода файла;

2) классификация на вредоносное и легитимное программное обеспечение, используя сверточную нейронную сеть и представление файла в виде изображения;

3) классификация вредоносных файлов на типы вредоносного программного обеспечения, используя сверточные и рекуррентные сети.

Полученный результат – разработанное программное обеспечение для классификации и кластеризации файлов на два класса: вредоносное и легитимное программного обеспечения, а также классификации вредоносного программного обеспечения на типы вредоносного ПО.

Область применения – обнаружение вредоносного программного обеспечения.

Abstract

Master thesis, 46 pages, 13 figures, 12 table, 24 resources.

Keywords: MALWARE, STATIC CODE, CLASSIFICATION, CLUSTERING, MACHINE LEARNING, NEURAL NETWORK, TF-IDF, LONG SHORT-TERM MEMORY.

The object of research is executable files of malicious and legitimate software.

The aim of this work is development and implementation of algorithms for detection and classification of malicious software.

During the work, the following approaches are considered for the classification and clustering of malicious software:

1) classification and clustering into malicious and legitimate software, based on the static text characteristics of the file's byte code;

2) classification into malicious and legitimate software using a convolutional neural network and file representation as an image;

3) classification of malicious files into types of malware using convolutional and recurrent networks.

Results of work – developed software for classifying and clustering files into two classes: malicious and legitimate software, also classifying malicious software into types of malware.

Area of application – malware detection.