МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

ІНФОРМАТИКА, МАТЕМАТИКА, АВТОМАТИКА

IMA :: 2017

МАТЕРІАЛИ
та програма

НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

(Суми, 17–21 квітня 2017 року)

Суми
Сумський державний університет
2017

Классификация вредоносного программного обеспечения с помощью методов машинного обучения

Чумаченко К.И., *студент*
Харьковский национальный университет радиоэлектроники,
г. Харьков

Современные методы обнаружения и классификации вредоносного программного обеспечения, основанные на сигнатурах, не способны эффективно обнаруживать полиморфные вирусы и вирусы нулевого дня. Таким образом, существует необходимость в новых методах анализа, таких как методы, основанные на машинном обучении.

В данной работе было проведено исследование, направленное на определение наиболее точного метода обнаружения вирусов, представленного в виде задач бинарной классификации (вирус / легитимная программа) и многоклассовой классификации. В качестве признаков были использованы успешные и безуспешные вызовы API, а также соответствующие коды возврата, которые были получены во время выполнения вирусов в изолированной виртуальной среде.

Для работы было использовано 1150 вредоносных файлов, принадлежащих к семьям Dridex, Locky, TeslaCrypt, Vawtrak, Zeus, DarkComet, CyberGate, Xtreme, CTB-Locker, а также 980 легитимных файлов.

Для задач бинарной и многоклассовой классификации были протестированы методы Decision Tree, Naive Bayes, Support Vector Machines, Random Forest, k-Nearest Neighbors. Данные методы дали следующие результаты для бинарной классификации: kNN – 94.6%, SVM – 94.6%, Naive Bayes – 55%, Decision Tree – 94.6%, Random Forest – 96.8%. Для многоклассовой классификации были получены следующие результаты: kNN – 87%, SVM – 87.6%, Decision Tree – 93.3%, Random Forest – 95.69%.

Таким образом, наилучший результат в обоих случаях был достигнут с помощью Random Forest. Была достигнута высокая точность, что подтверждает актуальность и эффективность использования методов машинного обучения для обнаружения вредоносного прогарммного обеспечения.