DOI 10.34185/1562-9945-4-153-2024-05
УДК 004.4

К.Ю. Островська, І.В. Стовпченко, Є.В. Островський

РОЗРОБКА ПІДХОДУ ДО ВИЯВЛЕННЯ ШКІДЛИВОГО ПЗ ДЛЯ ANDROID З ВИКОРИСТАННЯМ МЕТОДІВ ГЛИБИННОГО НАВЧАННЯ

Анотація. Метою роботи є розробка підходу для виявлення шкідливого програмного за безпечення для операційної системи Android на основі статистичного аналізу с викори станням методів глибокого навчання. Для досягнення поставленої мети були вирішені наступні задачі: 1. Дослідження особливостей Android додатків і розробка способу по дання додатка для подальшого аналізу безпеки. 2. Дослідження методів глибокого на вчання і вибір найбільш відповідного з них. 3. Розробка підходу до виявлення шкідливого програмного забезпечення для Android з використанням методів глибокого навчання. Ос новна ідея підходу уявлення Android додатки у вигляді зображення для подальшого аналізу згортовою нейронною мережею, причому в цьому зображенні пікселі представля ють послідовність пар API виклику і відпорному йому рівня захисту, який виводиться з дозволу, яке необхідно для виклику API.
Ключові слова: операційна система, Android, Android додатки, штучна нейронна мережа, згорткова нейронна мережа, API, алгоритм, класифікація, машинне навчання.

Згідно з останніми звітами про поширеність на ринку різних операційних систем [1] Android займає більше 75% ринку значно випереджає своїх найближчих суперників. В даний час Android виповнюється на великий збір різних типів пристроїв і форм-факторів, будь то смартфон, планшет, електронна книга, наручний годинник, фітнес-браслет, ноутбук, телевізор, бортовий комп'ютер автомобіля, розумні окуляри і багато інших пристроїв. З огляду на таку поширеність, гнучкість і багатофункціонального нальность Android, не дивно, що дана операційна система є і найпопулярнішою серед розробників шкідливого програмного забезпеченняня (ШПЗ), причому Android є не тільки найпопулярнішою платформою для ШПЗ серед мобільних ОС, але і серед настільних, значно випереджаючи свого найближчого конкурента - ОС Windows [2].

© Островська К.Ю., Стовпченко І.В., Островський Є.В., 2024

Метою роботи є розробка підходу для виявлення шкідливого програмного забезпечення для операційної системи Android на основі статистичного аналізу с використанням методів глибокого навчання.

Основна ідея підходу - уявлення Android-додатки у вигляді зображення для подальшого аналізу згортовою нейронною мережею, причому в цьому зображенні пікселі представляють послідовність пар API виклику і відпорному йому рівня захисту, який виводиться з дозволу, яке необхідно для виклику API. На рисунку 1 представлено пропонуємий підхід.

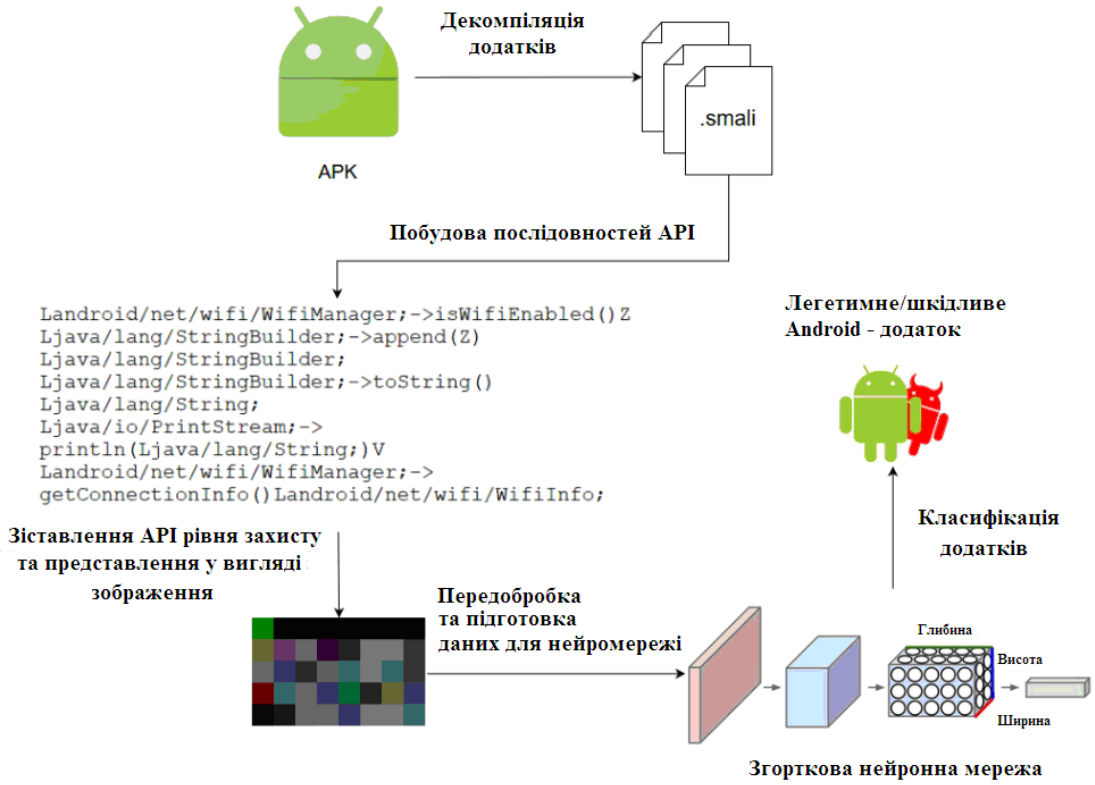

Рисунок 1 – Структурно-логічна схема розробленого підходу

Спочатку Android-додаток декомпілюються, байт-код classes.dex дізассемблюється в код smali. Далі проводиться аналіз отриманого коду і побудова графа потоку управління, за яким і виконується побудова послідовності API викликів. Варто враховувати, що великі програми можуть містити кілька файлів classes.dex, тому це необхідно враховувати при аналізі. Отриманим API викликам зіставляються відповідні рівні захисту, і таким чином формуються уявленні додатки. Один з ключових моментів даного підходу - це уявлення Android-додатку для згорткової нейронної мережі. Перетворення даної отриманої на попередньому кроці послідовності в зображення описується в наступному підрозділі.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

На останньому кроці заздалегідь навчена згортова нейронна мережа (на виборці, яка формується таким же чином, як зазначено вище), вирішує завдання класифікації заданого додатки на легітимні і шкідливі.

Подання пари API і рівня захисту. для даного дослідження була обрана згортова нейромережа, яка очікує на вхід дані в форматі зображення.

З кожною новою версією Android кількість доступних API росте, крім того в даній роботі враховуються API Java і Javax, тому точну кількість API назвати не можна, це сотні тисяч API, що безумовно лежить в межах між 2^{16} і 2^{24} . З огляду на, що все 3 канали пікселя вже зайняті кодом API, то для рівня захисту можна було б відвести альфа-канал, а коди рівнів впорядкувати таким чином, щоб більш небезпечного рівня захисту відповідало більше числове значення.

Таким чином, більш прозорі пікселі будуть відповідати менш небезпечними API, а менш прозорі - більш небезпечним.

Для створення більш щільного розподілу значень API по 3 байтам використовується швидка некріптографічна 32-бітна хеш-функція MurmurHash3 [3]. Після обчислення хеш-значення виділяються його молодші 24 біта, в результаті чого можливі колізії, однак, як виявилося на практиці їх кількість вкрай мала в порівнянні з кількістю API викликів, тому даною проблемою можна знехтувати. Рівень захисту виступає в якості альфа-каналу.

Альфа-канал представлений значенням від 0 до 1, відповідно від абсолютно прозорі до абсолютно непрозорого. У зв'язку з цим варто зазначити, яким чином перетворюються рівні захисту. У таблиці 1 представлені присутні в роботі рівні захисту.

Таблиця 1

Опис рівнів захисту

Рівень захисту	Числове значення	Короткий опис
NONE	1	Присвоюється Android API, яким не було поставлено у відповідність ніякий дозвіл
UNKNOWN	2	Присвоюється Java і Javax API викликів
NORMAL	3	Відповідає групі дозволів Android SDK з рівнем захисту normal
SIGNATURE	4	Відповідає групі дозволів AndroidSDK з рівнем захисту signature
DANGEROUS	5	Відповідає групі дозволів AndroidSDK з рівнем захисту dangerous

Перетворення значення рівня захисту в діапазон від 0 до 1 виконується елементарно згідно з формулою (1).

$$PL^{\alpha} = \frac{PL_i}{PL_{dangerous}} \quad (1)$$

де $PL_i = \{1, ..., 5\}$, $PL_{dangerous} = 5$, а PL^{α} представляє значення PL_i в діапазоні від 0 до 1.

Приклад вищеописаного RGBA-RGB перетворення представлений на рисунку 2.

Тут в якості RGBAtoRGB представлена процедура перетворення RGBA пікселя в RGB з урахуванням чорного фону. Як видно на рисунку 2, RGBA пікселі з однаковими значеннями RGB каналів, але різним значенням альфа-каналу перетворюються в відповідно темніший і світліший RGB пікселі.

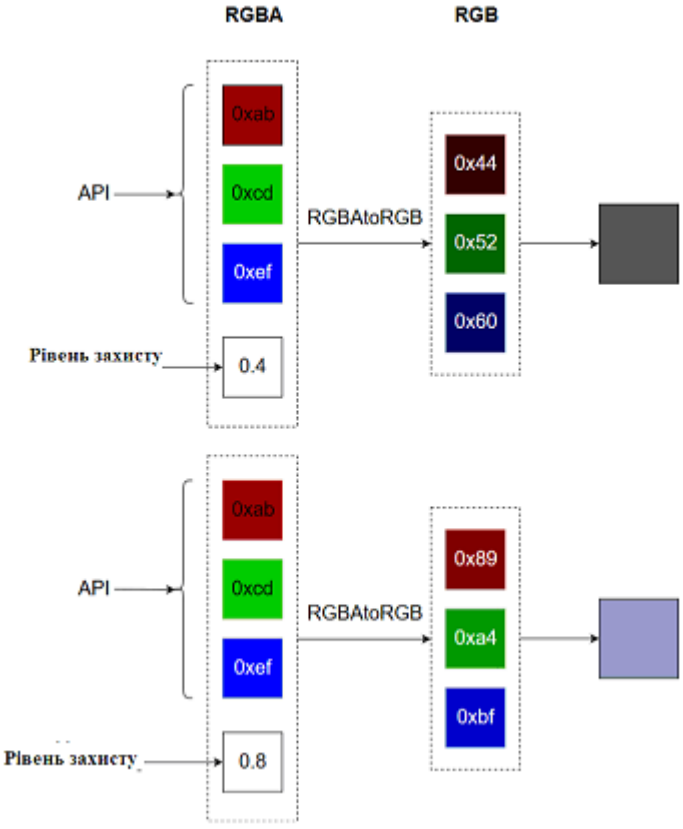

Рисунок 2 - Приклад RGBA-RGB перетворення

Нейронна мережа в якості класифікатора. Останній компонент підходу – це згорткова нейронна мережа.

Входом для нейромережі є певний обсяг, тобто тривимірний тензор розмірності W× H×D, де W- це ширина зображення, H- висота зображення, а D-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

глибина зображення або кількість каналів. У даній роботі D = 3 для вхідного обсягу, оскільки подаються на вхід зображення мають 3 канали - R, G і B. Навчання нейронної мережі проводиться з учителем, тобто разом з вхідним зображенням X, якє представлє деякий Android-додаток, подається мітка y = {0,1}, де y = 1, якщо додаток шкідливе, і y = 0 - якщо легітимне. Навчання будь якої глибокої нейронної мережі, в тому числі і згорткової, проводиться протягом декількох епох, тобто декількох прогонів тренуючих виборок. По завершенню кожної епохи нейромережі подається тестова вибірка, на якій нейронна мережа не навчалася, для оцінки точності.

В результаті навчання згортокої нейронної мережі фактично виходить готовий класифікатор Android-додатків, який для кожного раніше невідомого Android-додатку видасть своє припущення про те, чи є він шкідливим або легітимним.

Висновки. Розроблено підхід до виявлення шкідливого програмного забезпечення для ОС Android на основі уявлення Android-додатків, а також згорткової нейронної мережі, яка була спеціально розроблена для розпізнавання зображень. Послідовність пар API викликів і рівнів захисту Android-додатків перетворюють в RGB зображення, яке потім подається на вхід згорткової нейронної мережі. Навчившись на вибірці з подібних зображень, нейронна мережа виступає в якості класифікатора входять Android-додатків на легітимні і шкідливі.

ЛІТЕРАТУРА / REFERENCES

1. Mobile Operating System Market Share Worldwide [Електронний ресурс] // StatCounter.com - URL: http://gs.statcounter.com/os-market-share/mobile/worldwide

2. Android vs iOS vs Windows: Which suffers most infections? Nokia reveals all [Електронний ресурс] // ZDNet.com. - URL: https://www.zdnet.com/article / android-vs-ios-vs-windows-which-suffers-most-infections-nokia-reveals-all /

3. McAfee Mobile Threat Report [Електронний ресурс] // McAfee.com. - URL: https://www.mcafee.com/cn/resources/reports/rp-mobile-threat-report2024.pdf.

Received 29.04.2024.
Accepted 30.04.2024.

Development of an approach to the detection of Android software using deep learning methods

The purpose of the work is to develop an approach to detect malicious software for the Android operating system based on statistical analysis using deep learning methods. To achieve the goal, the following tasks were solved: 1. Study of the features of Android

applications and development of a method of submitting the application for further security analysis. 2. Research of deep learning methods and selection of the most appropriate of them. 3. Development of an Android malware detection approach using deep learning techniques. The main idea of the approach is to represent the Android application in the form of an image for further analysis by a convolutional neural network, and in this image the pixels represent a sequence of API call pairs and the level of protection against it, which is derived from the permission required for the API call.

An Android malware detection approach is developed based on the representation of Android applications, as well as a convolutional neural network that has been specially developed for image recognition. A sequence of pairs of API calls and security levels of Android applications is converted into an RGB image, which is then fed to the input of a convolutional neural network. Having trained on a sample of similar images, the neural network acts as a classifier of included Android applications into legitimate and malicious ones.

Островська Катерина Юріївна - к.т.н., доцент, доцент кафедри Інформаційних технологій і систем, Український державний університет науки і технологій.

Стовпченко Іван Володимирович – старший викладач кафедри Інформаційних технологій і систем, Український державний університет науки і технологій.

Островський Євген Вікторович – магістр кафедри Інформаційних технологій і систем зі спеціальності «Компютерні науки», Український державний університет науки і технологій.

Ostrovska Kateryna - candidate of technical sciences, associate professor, associate professor of the Department of Information Technologies and Systems, Ukrainian State University of Science and Technology.

Stovpchenko Ivan - senior teacher кафедри Інформаційних технологій і систем, Український державний університет науки і технологій.

Ostrovsky Yevhen - master of the Department of Information Technologies and Systems, from the specialty "Computer Science", Ukrainian State University of Science and Technology.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)