DOI 10.37882/2223-2966.2024.7-2.14

РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА ЭВРИСТИЧЕСКОГО АНАЛИЗА ПРИЛОЖЕНИЙ ДЛЯ ОПЕРАЦИОННОЙ СИСТЕМЫ ANDROID

HEURISTIC APPLICATIONS ANALYSIS SOFTWARE TOOL DEVELOPMENT FOR THE ANDROID OPERATING SYSTEM

S. Kozhemyakin
A. Selin
S. Gorelik
A. Bugaev

Summary. This article presents a study of the security of mobile applications by identifying encrypted content within them. As part of the study, a software tool was developed for statistical analysis of the contents of application installation files for the Android operating system. The open-source package of statistical tests — NIST (The National Institute of Standards and Technology) — is used as a mathematical algorithm. For implementation, the Android Studio development environment and Java programming language were used. The proposed analysis method can be used as a complement to existing heuristic tools and algorithms used by antivirus software.

Keywords: analysis of mobile applications, statistical analysis of files, information security.

Кожемякин Сергей Юрьевич
ассистент,
МИРЭА Российский технологический университет
sergei.unk@yandex.ru
Селин Андрей Александрович
доцент,
МИРЭА Российский технологический университет
chuknor@yandex.ru
Горелик Сергей Сергеевич
старший преподаватель,
МИРЭА Российский технологический университет
gorelikss@rambler.ru
Бугаев Александр Александрович
старший преподаватель,
МИРЭА Российский технологический университет
sansanych.bugaev@yandex.ru

Аннотация. В данной статье представлено исследование безопасности мобильных приложений путем выявления шифрованного контента в их составе. В рамках исследования был разработан программный инструмент статистического анализа содержимого установочных файлов приложений для операционной системы Android. В качестве математического аппарата используется пакет статистических тестов с открытым исходным кодом — NIST (англ. The National Institute of Standards and Technology). Для реализации использовалась среда разработки Android Studio, язык программирования — Java. Предложенный способ анализа может использоваться как дополнение к существующим эвристическим инструментам и алгоритмам, применяемым антивирусным программным обеспечением.

Ключевые слова: анализ мобильных приложений, статистический анализ файлов, информационная безопасность.

На сегодняшний день мобильные телефоны стали неотъемлемой частью нашей жизни. Помимо основного средства коммуникации, смартфоны используются для покупок, хранения личных данных, управления финансами, получения государственных и иных услуг. Вместе с этим появились новые угрозы конфиденциальности и безопасности. Мошенники и киберпреступники активно используют вредоносные мобильные приложения в своих целях.

Защитить от такого рода атак призваны антивирусные приложения, а также системные компоненты защиты, предусмотренные производителями. Для операционной системы Android таким компонентом является Google Play Protect [1]. Также проверки приложений на безопасность осуществляются на этапе публикации в официальном магазине приложений Play Market. Однако в связи с многочисленными блокировками российских приложений (в том числе приложений финансового сектора) пользователи всё чаще загружают установочные файлы программ на свой смартфон из различных сторонних источников.

Установочный файл приложения для операционной системы Android — это, как правило, файл с расширением *.apk, являющийся в свою очередь ZIP-архивом. Помимо скомпилированного кода в архиве могут содержаться различные файлы и компоненты. Согласно отчетам [2–3] исследовательских компаний в области кибербезопасности, зачастую такие файлы представляют собой зашифрованный исполняемый код, который динамически расшифровывается во время работы приложения.

В антивирусном программном обеспечении наиболее распространены два метода обнаружения вредоносных приложений — это эвристический и сигнатурный анализ [4].

Сигнатурный анализ оценивает форму файла, занимаясь поиском строк и шаблонов, совпадающих с уже известными образцами вредоносного ПО. Однако при использовании зашифрованного кода такой метод анализа не демонстрирует достаточной точности.

Эвристический анализ оценивает функционал файла, используя специальные алгоритмы и шаблоны для отслеживания возможной подозрительной активности или поведения. В основе работы эвристического анализатора лежит набор предположений или эвристик о характерных признаках вредоносного или безопасного кода. Каждый признак кода имеет определенный вес (т. е. число, показывающее важность и достоверность этого признака). Таким признаком может являться зашифрованный контент в составе установочного пакета приложения.

Антивирусные компании не раскрывают особенностей алгоритмов работы своих анализаторов, чтобы разработчики вредоносного программного обеспечения не смогли адаптировать свои приложения к средствам защиты. В рамках данной работы предлагается использовать математический аппарат с открытым исходным кодом, который позволит расширить возможности по детектированию подозрительного контента и ограничивать установку приложения до момента его запуска.

Постановка задачи

Целью работы является повышение безопасности использования мобильных приложения для операционной системы Android за счет разработки специального программного решения для статистического анализа компонентов установочного пакета формата APK (англ. Android Package Kit).

Основные задачи, решаемые в работе:

- исследование предметной области;
- обзор существующих методов оценки статистических свойств файлов;
- адаптация алгоритмов оценки статистических свойств битовых данных;
- создание программного продукта для операционной системы Android на основе проведенных исследований.

Объект исследования — безопасность мобильных приложений для операционной системы Android.

Предмет исследования — выявление небезопасных мобильных приложений на основе статистических свойств файлов из состава установочного пакета приложения для операционной системы Android.

Для проведения исследования были подготовлены 2 тестовые выборки. Каждая выборка содержит 100 мобильных приложений. Первая выборка [5] состоит из легитимных приложений для личного использования, вторая [6] — из образцов известных вредоносных приложений. Приложения из обеих выборок содержат дополнительные файлы в составе установочных пакетов.

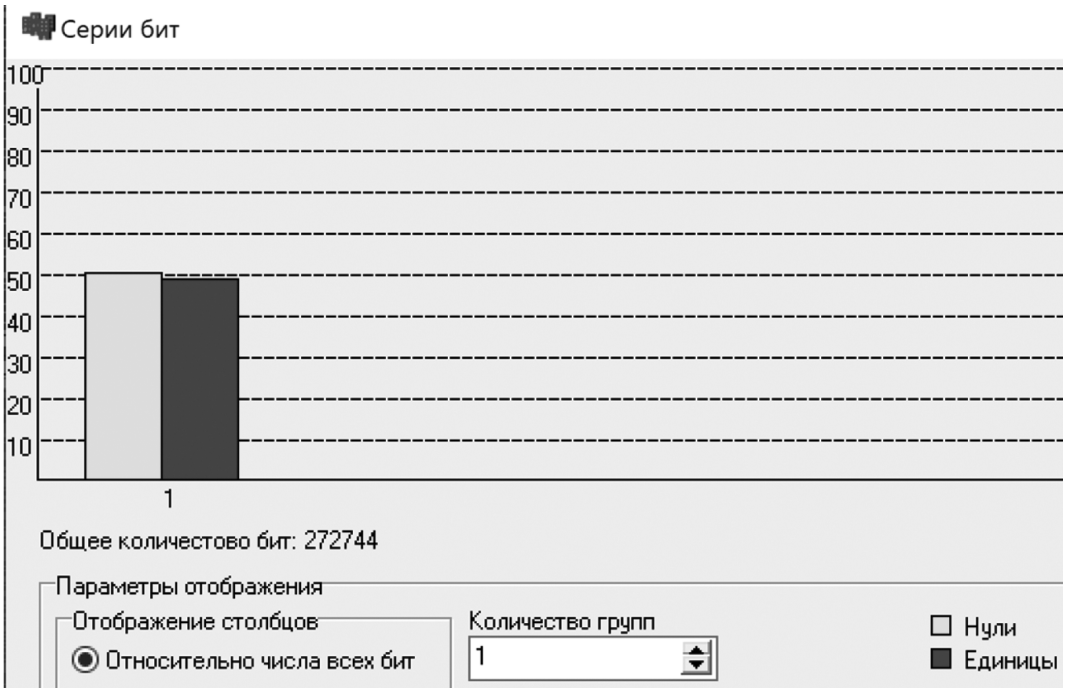

Рис. 1. Распределение битов в файле omsdk-v1.js из состава установочного пакета легитимного приложения

Особенности существующих методов оценки статистических свойств цифровых данных

В результате исследования известных подходов к оценке статистических характеристик файла были рассмотрены графические и оценочные тесты, применяемые для анализа цифровых последовательностей на стохастичность.

Графические тесты обладают недостатком с позиции точности в интерпретации результатов. Решение о стохастичности входной последовательности в конечном счете принимается оператором визуально, что в ряде случаев приводит к ошибкам.

Основным преимуществом, которым обладают оценочные тесты по сравнению с графическими, является то, что оценочные тесты характеризуются численной характеристикой, которая позволяет однозначно определить, пройден тест или нет.

Простейшим методом оценки статистических свойств является подсчет доли битов 1 и 0 в исследуемом файле. При использовании этого метода на легитимной выборке приложений эксперимент показал большую долю ошибки второго рода (β-ошибка) — более 20 %. Некоторые файлы из состава установочного пакета приложения имеют близкое распределение битов:

То есть, безопасные файлы легитимных приложений могут быть отмечены как шифрованные.

Наиболее известные оценочные тесты для анализа последовательностей на случайность: тесты Кнута, DIEHARD и тесты NIST. Для тестов Кнута отмечаются следующие недостатки [7]:

- отсутствие рекомендуемых параметров тестирования;
- методика оценки результатов является недостаточно четкой.

В результате анализа тестов DIEHARD определены следующие недостатки:

- отсутствие четко сформулированной методики трактовки результатов;
- большинство тестов основано на результатах испытаний;
- жесткое задание параметров тестирования.

В ходе анализа тестов, описанных в руководстве [8] NIST, было установлено, что недостатки, присущие исследованным ранее подборкам, в наибольшей степени были устранены. Подборка состоит из 15 различных тестов:

В тестах введена достаточно простая интерпретация алгоритмов, появилась возможность сведения оценок

Таблица 1.

Список тестов NIST

Частотный побитовый тест	Частотный блочный тест	Тест на последовательность одинаковых битов
Тест на самую длинную последовательность единиц в блоке	Тест рангов бинарных матриц	Спектральный тест
Тест на совпадение неперекрывающихся шаблонов	Тест на совпадение перекрывающихся шаблонов	Универсальный статистический тест Маурера
Тест на линейную сложность	Тест на периодичность	Тест приблизительной энтропии
Тест кумулятивных сумм	Тест на произвольные отклонения	Альтернативный тест на произвольные отклонения

полученных результатов для каждого теста в одну общую оценку и возможность тестирования последовательностей практически любого размера. В алгоритмах тестирования использованы функции математической статистики, что позволяет более точно получить наблюдаемое значение для критериального сравнения. Данные тесты основаны на различных статистических свойствах, присущих только стохастическим последовательностям. Такой подход способен послужить мерой оценки случайности цифрового потока для его дальнейшего анализа.

Адаптация алгоритмов оценки статистических свойств битовых данных и разработка программного решения

Для разработки программного средства (приложения) для операционной системы Android адаптирован код тестов NIST на языках программирования Java [9] и C++ [10]. Для принятия решения о результате выполненного теста происходит оценка полученного итогового значения с плавающей точкой p_value. Корректным значением является любое число в диапазоне (0;1]. В результате запуска тестов на двух выборках приложений выявлены 11 тестов, показавших наиболее точные и корректные оценочные величины. Частотный побитовый тест, частотный блочный тест, тест на последовательность одинаковых битов и альтернативный тест на произвольные отклонения были исключены вследствие нестабильных результатов для приложений из разных выборок.

Исходя из полученных данных, разработано мобильное приложение, получающее у пользователя разрешение на установку других пакетов. При попытке установить любое приложение (файл APK) разработанное программное средство распаковывает указанный пакет

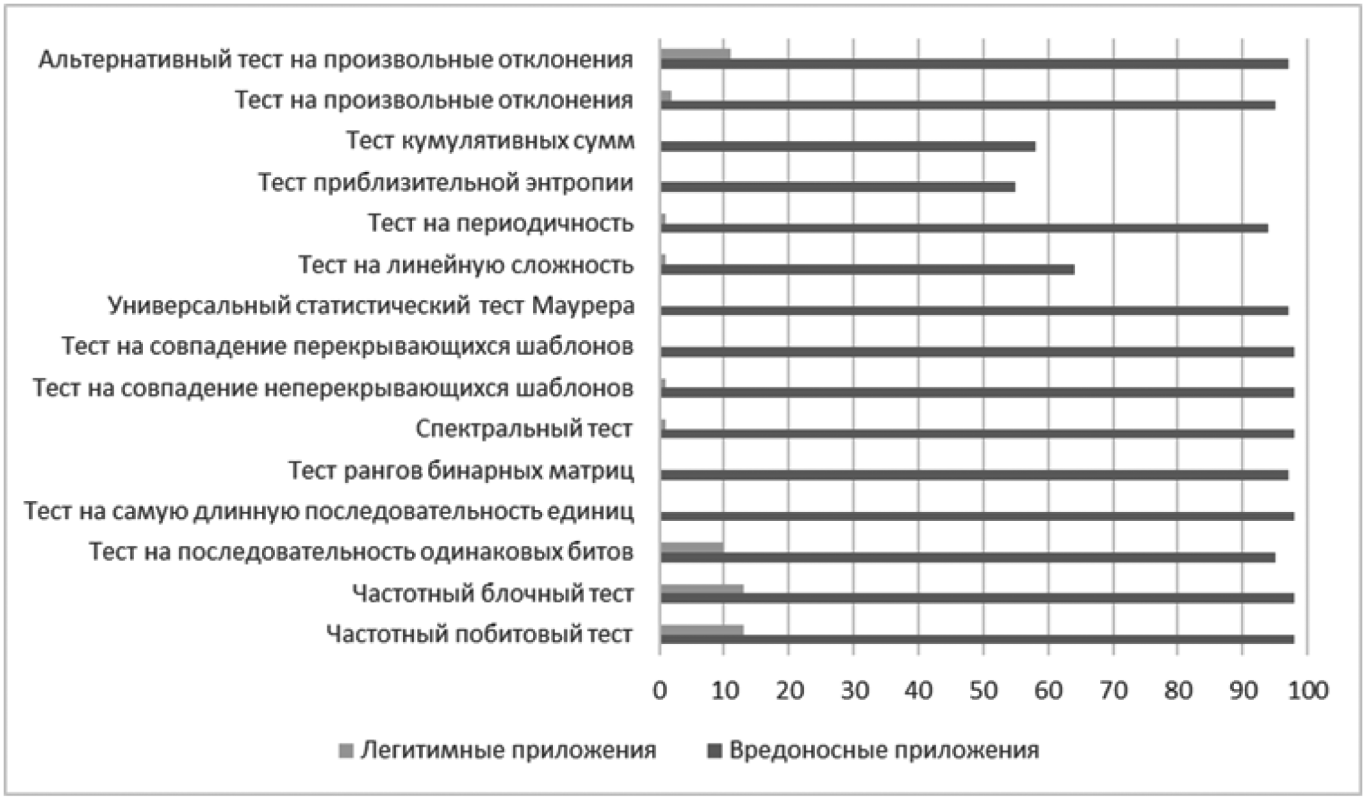

Рис. 2. Статистика успешно выполненных тестов для двух типов выборок

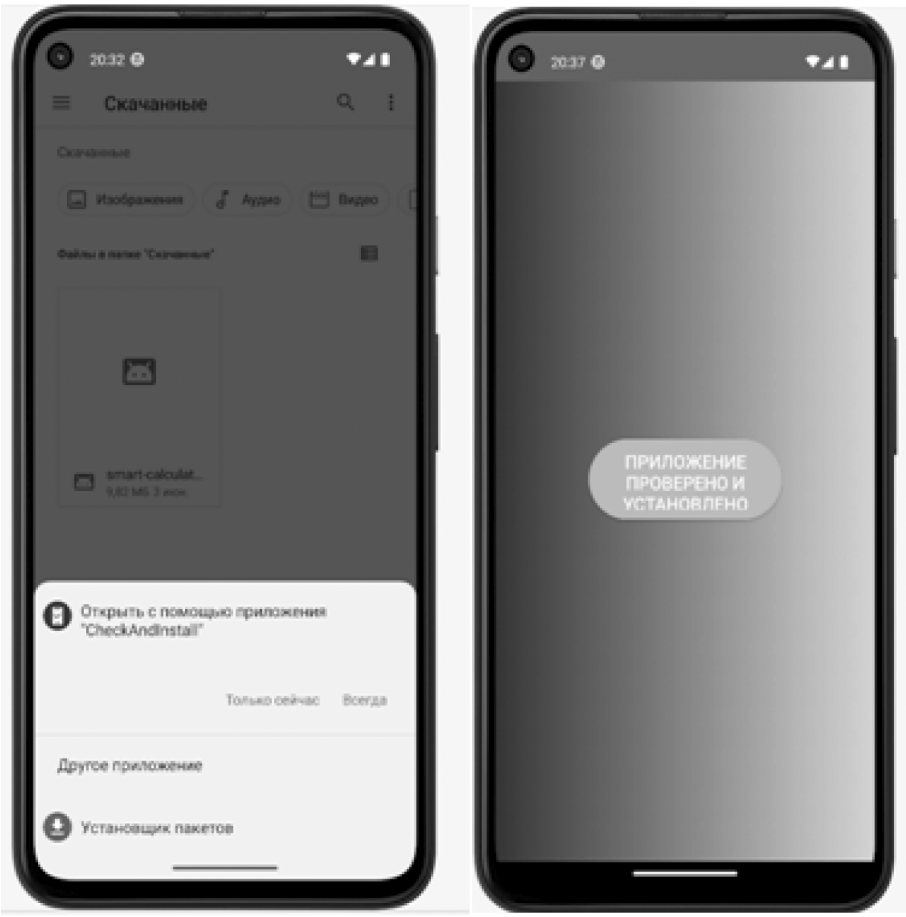

Рис. 3. Установка и проверка приложения с помощью разработанного программного средства

во внутреннюю память, разбивает полученные файлы из состава пакета на блоки (в рамках эксперимента — 100 Кб) и проводит их анализ по 11 выбранным тестам. Область принятия гипотезы о наличии псевдослучайных последовательностей — 6 и более успешно выполненных тестов. В таком случае установка приложения останавливается и пользователю отображается информирующее сообщение. При отсутствии таких блоков — установка продолжается и приложение запускается.

Заключение

В работе рассмотрены проблемы детектирования зашифрованного контента в составе установочных пакетов для операционной системы Android, являющегося одним из отличительных признаков вредоносных приложений. Для решения данной проблемы предложен подход к оценке статистических свойств таких файлов с помощью математического аппарата тестов NIST. На основе открытого кода тестов NIST разработано мобильное приложение для демонстрации предложенной гипотезы. Данный способ может использоваться как дополнительный инструмент к антивирусным продуктам, расширяя возможности эвристического анализа.

ЛИТЕРАТУРА

1. Google Play Protect. — Текст: электронный — URL: https://developers.google.com/android/play-protect (дата обращения 10.05.2024)
2. Отчет компании «AtRedPiranha» — Текст: электронный — URL: https://redpiranha.net/news/android-malware-dvmap-infects-devices-google-play-store-first-android-malware-has-code (дата обращения 10.05.2024)
3. Отчет компании «INFOSEC» — Текст: электронный — URL: https://www.infosecinstitute.com/resources/malware-analysis/mallocker-android-ransomware-what-it-is-how-it-works-and-how-to-prevent-it-malware-spotlight/ (дата обращения 10.05.2024)
4. Как киберпреступники пытаются обойти антивирусную защиту — Текст: электронный — URL: https://www.kaspersky.ru/resource-center/threats/combating-antivirus (дата обращения 10.05.2024)
5. Хранилище мобильных приложений. URL: https://ru.uptodown.com/android/personal (дата обращения 10.05.2024)
6. Хранилище образцов вредоносного программного обеспечения. URL: https://bazaar.abuse.ch/browse.php?search=tag%3Aandroid (дата обращения 10.05.2024)
7. Иванов М.А. Теория, применение и оценка качества генераторов псевдослучайных последовательностей / М.А. Иванов, И.В. Чугунков. — М.: КУДИЦ-ОБРАЗ, 2003. — 240 с.
8. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST SP 800-22 Rev. 1 — Текст: электронный — URL: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf (дата обращения 10.05.2024)
9. steamfest /random test Электронный ресурс https://github.com. URL: https://github.com/stamfest/randomtests (дата обращения 10.05.2024)
10. jkerdels /NIST-Statistical-Test-Suite-C---Wrapper Электронный ресурс https://github.com. URL: https://github.com/jkerdels/NIST-Statistical-Test-Suite-C---Wrapper (дата обращения 10.05.2024)

© Кожемякин Сергей Юрьевич (sergei.unk@yandex.ru); Селин Андрей Александрович (chuknor@yandex.ru);
Горелик Сергей Сергеевич (gorelikss@rambler.ru); Бугаев Александр Александрович (sansanych.bugaev@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»