Автоматическая классификация вредоносного программного обеспечения для платформы Android

Анастасия Сковорода, Денис Гамаюнов

Московский Государственный Университет имени М. В. Ломоносова

Москва, 2016

Введение

- Широкое распространение смартфонов (4.5 миллиардов в мире)
- Устройства под управлением ОС Android (1.4 миллиарда)
  - Возможность установки приложений из неофициальных ресурсов

Методы борьбы с ВПО под Android

Подход к обнаружению

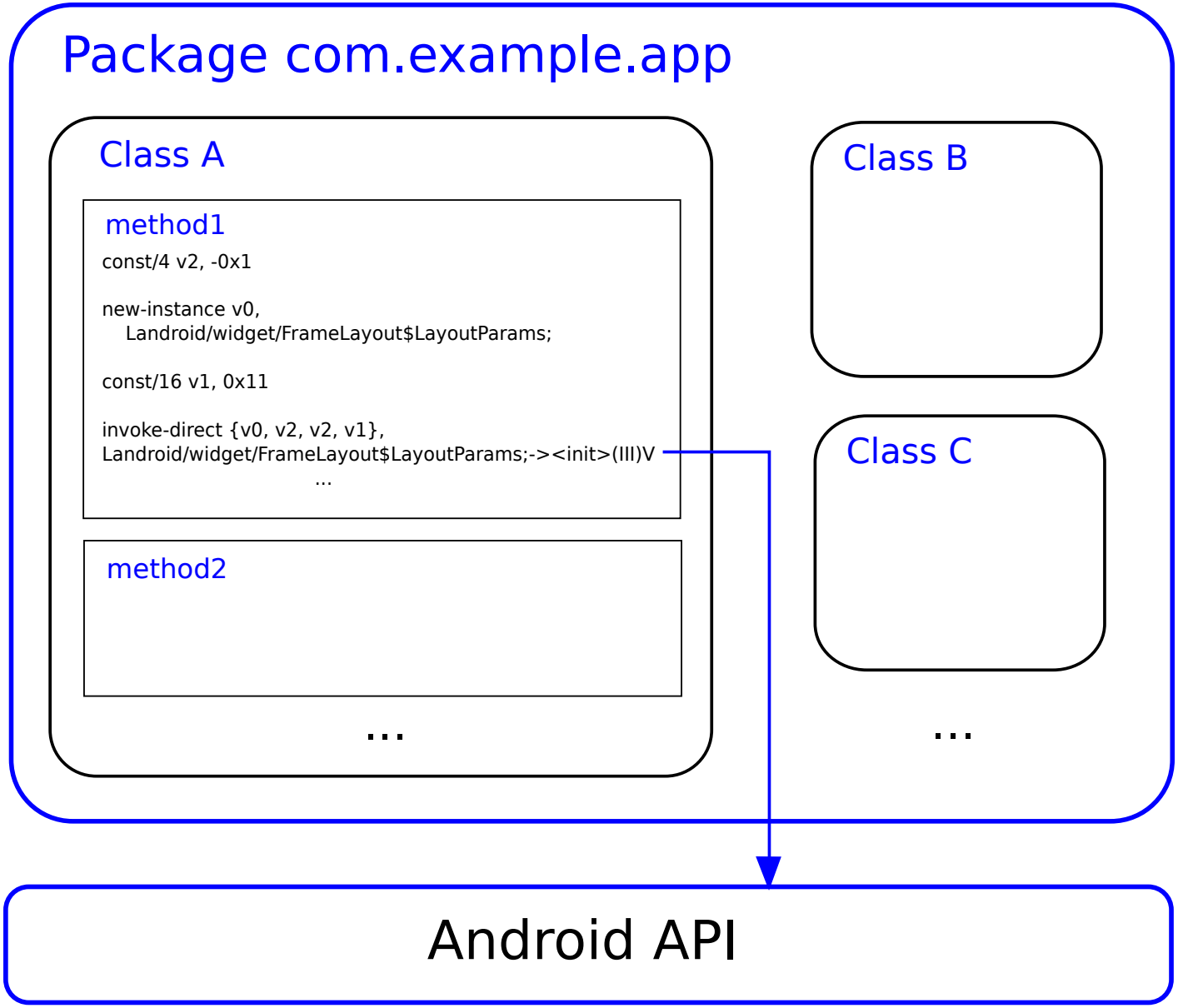
- Обнаружение аномалий
- Обнаружение злоупотреблений
  - Основаны на моделях/сигнатурах вредоносных приложений
    Проблема: составление моделей/сигнатур
  - Основаны на спецификации вредоносного поведения (taint-анализ)

Используемые техники

- Статические
- Динамические

Структура Android-приложений

Структура Android-приложений



Модель привилегий

На основе анализа привилегий, запрашиваемых вредоносными приложениями, составлен список из 101 привилегии.

android.permission.ACCESS_FINE_LOCATION	1
android.permission.BLUETOOTH_ADMIN	0
android.permission.CAMERA	1
android.permission.INTERNET	1
android.permission.NFC	0

Модель API-вызовов Android Framework

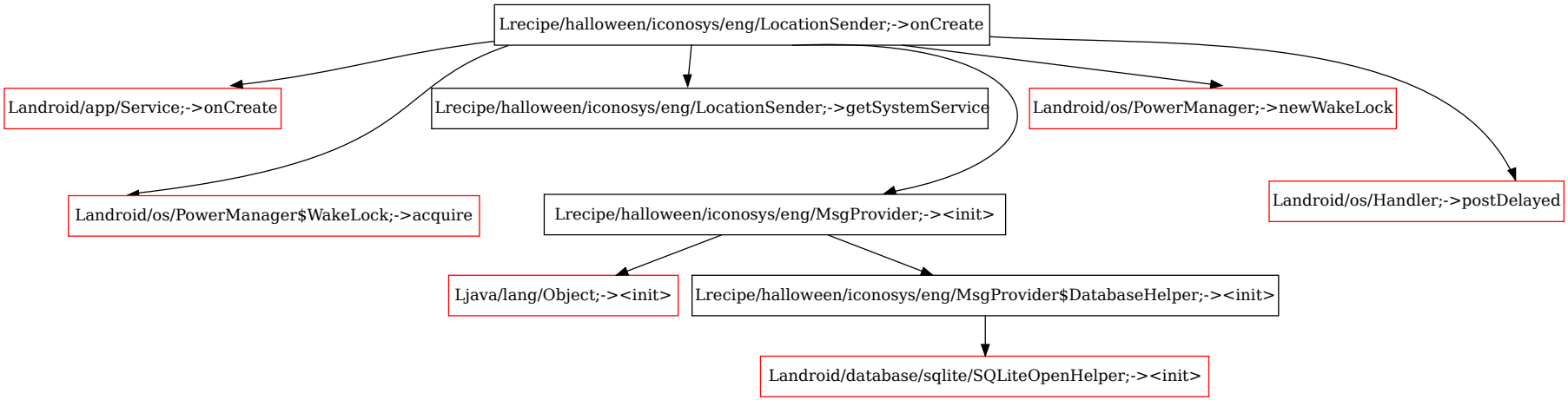
Составлен список из 384 API-вызовов, используемых во вредоносных приложениях (большинство из которых является *защищёнными*, т.е. требует наличия определённых привилегий).

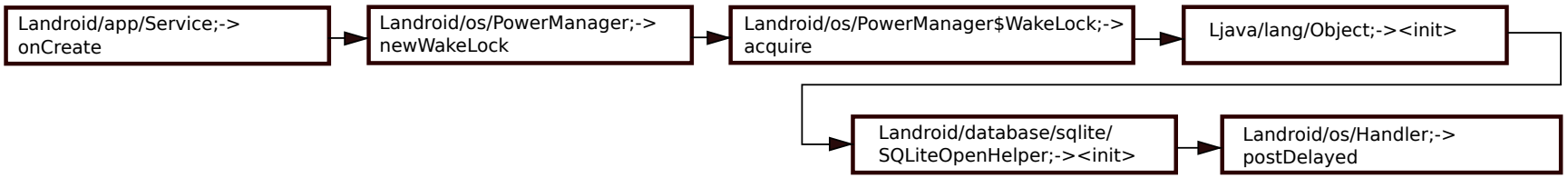
Ljava/net/URL;openConnection	1
Landroid/location/LocationManager;getBestProvider	0
Landroid/net/wifi/WifiManager;setWifiEnabled	0
Landroid/telephony/TelephonyManager;getDeviceId	1
Ljava/util/regex/Pattern;matcher	0

Модель цепочек API-вызовов

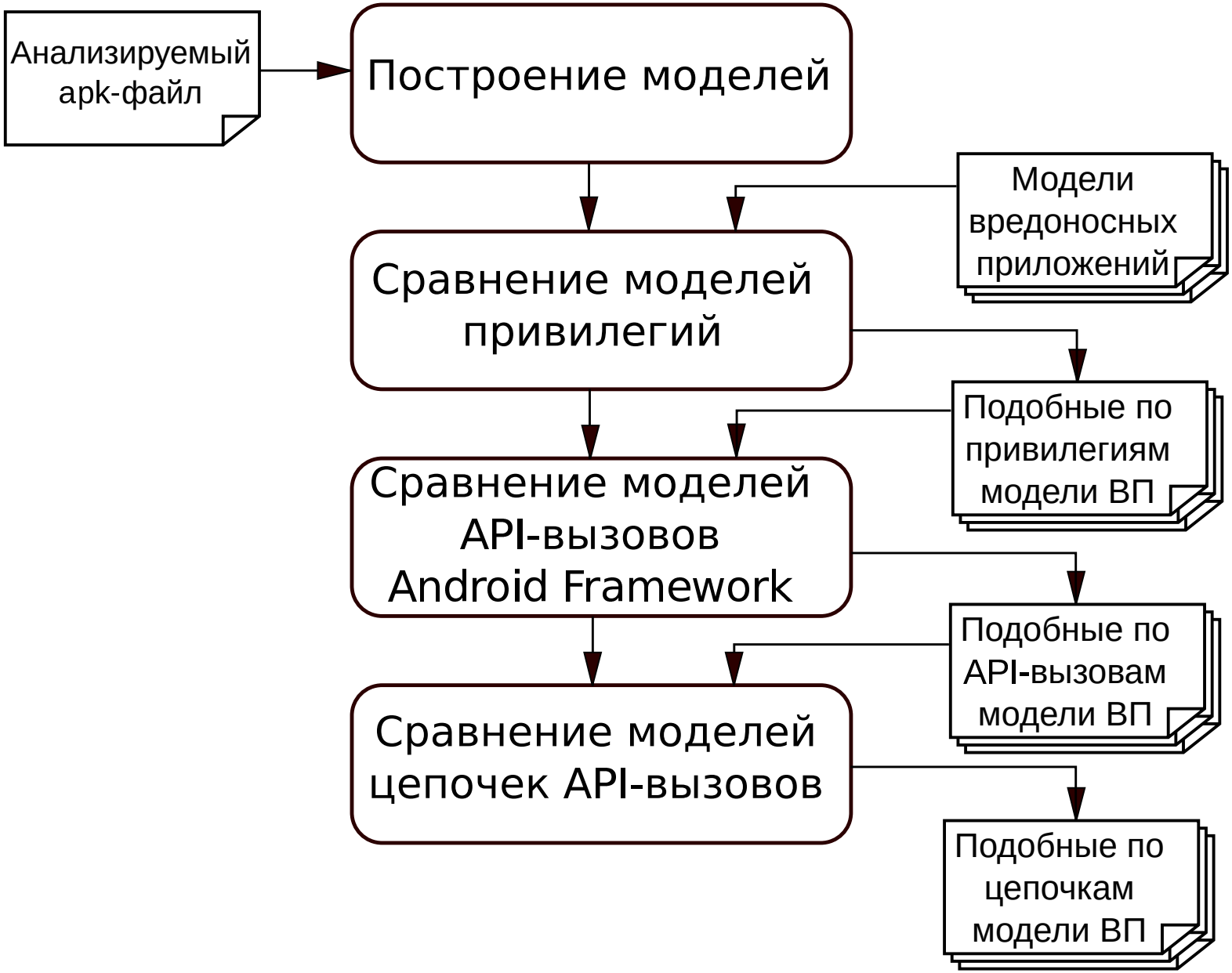
1. Поиск возможных точек входа в приложение;
2. Построение графа функциональных вызовов для каждой из точек входа;
3. Формирование цепочек API-вызовов, исходя из листьев и вершин, соответствующих вызовам вкомпилированных в приложение библиотечных функций, в графах функциональных вызовов.

Предлагаемый метод / Построение моделей





Предлагаемый метод / Схема



Сравнение моделей привилегий

$$S_P(v,u) = \frac{\sum\limits_{p \in P} w_p(2 * 1_{v_p=u_p \& u_p=1} + 0.5 * 1_{v_p=1 \& u_p=0})}{\sum\limits_{p \in P} 1_{v_p=1} + 1_{u_p=1}} \quad (1)$$

Сравнение моделей API-вызовов Android Framework

$$S_{API}(v,u) = \frac{\sum\limits_{f \in A} 1_{v_f=u_f \& u_f=1}}{\sum\limits_{f \in A} 1_{u_f=1}} \quad (2)$$

- Учитывается соотношение совпавших компонентов и общего количества ненулевых компонентов в модели ВП.
- Если значение функции сходства превосходит некоторый порог, модели считаются подобными.

Сравнение моделей цепочек API-вызовов

1. Попарное сравнение цепочек в модели анализируемого приложения и модели вредоносного приложения.
   - Поиск наибольшей общей подпоследовательности между цепочками — *общей подцепочки*.
   - Учитывается количество *компонентов*, на которые разбивается цепочка в модели анализируемого приложения *общей подцепочкой*.
2. Для каждой из цепочек в модели анализируемого приложения находятся все подобные цепочки во вредоносной модели, и если их несколько, выбирается наиболее подходящая (с учётом длины, наличия защищённых API-вызовов и количества *компонентов* в цепочке).

Сравнение моделей цепочек API-вызовов

- Результат сравнения моделей:
  1. Количество общих цепочек;
  2. Суммарная длина общих цепочек;
  3. Количество длинных общих цепочек;
  4. Количество общих цепочек, содержащих защищенные API-вызовы.
- Для каждого из этих параметров эмпирически были установлены пороговые значения и условия с учётом этих пороговых значений, при выполнении которых модели считаются подобными.

Эксперименты

Данные для экспериментов

- Легитимные: 43932 приложения из Google Play;
- Вредоносные: 5451 приложение – коллекция drebin (собрана в Гёттингенском университете имени Георга-Августа), 1929 приложений – коллекция icsx (собрана центром ISCX университета Нью-Браунсуик)

Результаты

- Модели ВП: 2026 приложений из коллекции drebin; среди вредоносных приложений для тестирования определены как подобные одной из заданных моделей 98.5%, доля ложных срабатываний 4%.
- Модели ВП: 80 приложений из коллекции iscx; среди вредоносных приложений для тестирования определены как подобные одной из заданных моделей 75%, доля ложных срабатываний 0.06%.

Контактная информация

Анастасия Сковорода
e-mail: nastya_jane@seclab.cs.msu.su

Денис Гамаюнов
e-mail: gamajun@seclab.cs.msu.su

Репозиторий с исходным кодом анализатора:
https://github.com/nastya/apk-analysis