FACULTY
OF INFORMATICS
Masaryk University

Analýza inštalačných APK súborov pre OS Android

Bakalárska práca

Martin Styk

Outline for Section 1

Ciele práce

- Vytvoriť rozsiahlu databázu inštalačných súborov pre OS Android (APK súborov)
- Popísať štruktúru a formát inštalačných súborov
- Vytvoriť nástroj na analýzu a získavanie metadát o inštalačných súboroch
- Určiť štatistické informácie o aplikáciách vo vytvorenej databáze
- Implementovať mechanizmus detekcie potenciálne modifikovaných APK súborov

OS Android

- 84,7% podiel na trhu s mobilnými operačnými systémami [1]
- Obľúbenosť vďaka veľkému počtu aplikácií
- Riziko jednoduchej možnosti modifikácie inštalačných súborov

Outline for Section 2

Distribúcia inštalačných APK súborov

- Oficiálne zdroje
  - Obchod s aplikáciami Google Play Store
- Neoficiálne zdroje
  - Neoficiálne obchody s aplikáciami
    - SlideMe
    - Amazon Appstore
  - Stránky na zdieľanie obsahu
    - ZippyShare
    - UlozTo
  - Torrenty

Získavanie aplikácií z Google Play

- Možné len do zaregistrovaného Android zariadenia
- Aplikácia Google Play Crawler [2]
- Projekt Playdrone
  - Databáza 1 100 000 APK súborov z Google Play
  - November 2014
  - Apk súbory dostupne vo verejnom archíve

Automatizácia hromadného preberania APK súborov

- Implementovaný nástroj ApkDownloader[3]
- Založený na princípe analýzy HTML kódu
  - Vyhľadanie priamych odkazov na APK súbory
  - Prevzatie APK súborov
- Implementovaný pre nasledujúce lokality
  - archív projektu Playdrone
  - www.appsapk.com
  - www.apkmaniafull.com
  - ww.androidapksfree.com
- Jendoducho rozšíritelný, open source

Zdroje prevzatých aplikácií

- Celkovo získaných 20 060 APK súborov, z toho viac ako 90 % pomocou ApkDownloader
- Celková veľkosť prebraných APK súborov 192 GB

Zdroj	**Počet aplikácií**	**%**
Playdrone	8 200	40,9
www.appsapk.com	6 470	32,3
www.apkmaniafull.com	2 870	14,3
www.androidapksfree.com	1 030	5,1
www.zippyshare.com	750	3,7
torrenty	550	2,7
www.uloz.to	190	0,9
Spolu	20 060	

Outline for Section 3

APK súbory

- Android application package file
- Slúžia na distribúciu aplikácií na platforme Android
- Štruktúra vychádza z formátu JAR
- Archívne súbory vo formáte ZIP
- Typicka štruktúra s povinnými súbormi v pevne stanovenom formáte
- Skompilované súbory
- Binárne XML

Štruktúra APK súborov

- Priečinok META-INF
  - CERT.RSA
  - MANIFEST.MF
  - CERT.SF
- Priečinok RES
- Priečinok LIB
- Priečinok ASSETS
- Súbor resources.arsv
- Súbor classes.dex
- Súbor AndroidManifest.xml

Analýza APK súborov

- Získanie detailných metadát o rôznych aspektoch APK súborov
- Automatizované pomocou vyvinutej aplikácie ApkAnalyzer[4]
- Využitie nástroja ApkTool na dekompiláciu [5]
- Použitá knižnica AXML na konverziu binárnych XML súborov do čitateľnej formy [6]
- Celkovo získaných 63 atribútov každého APK súboru

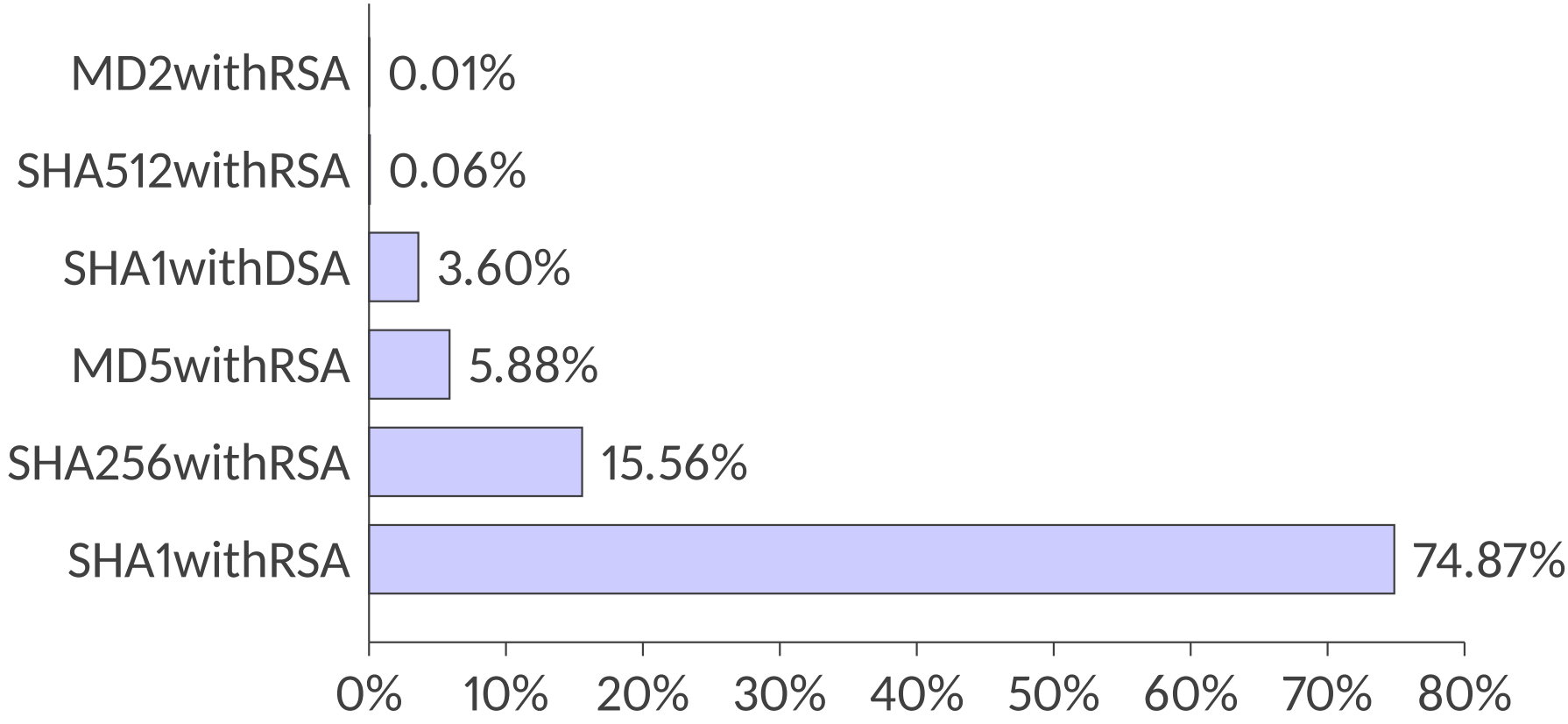
Príklady zbieraných informácií

- Základné informácie
  - veľkosť APK súboru, veľkosti dôležitých súborov
- Dáta zo súboru AndroidManifest.xml
  - komponenty aplikácie
  - prístupové oprávnenia
  - verzia Android SDK
- Informácie o certifikáte
  - algoritmus podpisu
- Informácie o zdrojových súboroch
  - formáty obrázkov
  - lokalizácia
- Hashe jednotlivých súborov

Najčastejšie prístupové oprávnenia

Názov	%
android.permission.internet	92,9
android.permission.access_network_state	87,9
android.permission.write_external_storage	75,2
android.permission.wake_lock	49,5
android.permission.read_phone_state	49,4
android.permission.access_wifi_state	44,7
android.permission.vibrate	43,6
android.permission.get_accounts	31,3
android.permission.receive_boot_completed	30,5

Algoritmy podpisu APK balíčkov



Obr.: Algoritmus podpisu APK balíčku

Lokalizácie APK balíčkov

Kód	Jazyk	%
es	španielsky	61,7
de	nemecký	59,6
fr	francúzsky	59,4
ru	ruský	58,1
ja	japonský	57,6
it	taliansky	57,4
ko	korejský	56,9
zh-rcn	čínsky (zjednodušený)	55,6

- Anglický jazyk považovaný za základ, nevyskytuje sa v tabuľke.
- V českom jazyku je lokalizovaných 49 % aplikácií, v slovenčine 46 %

Outline for Section 4

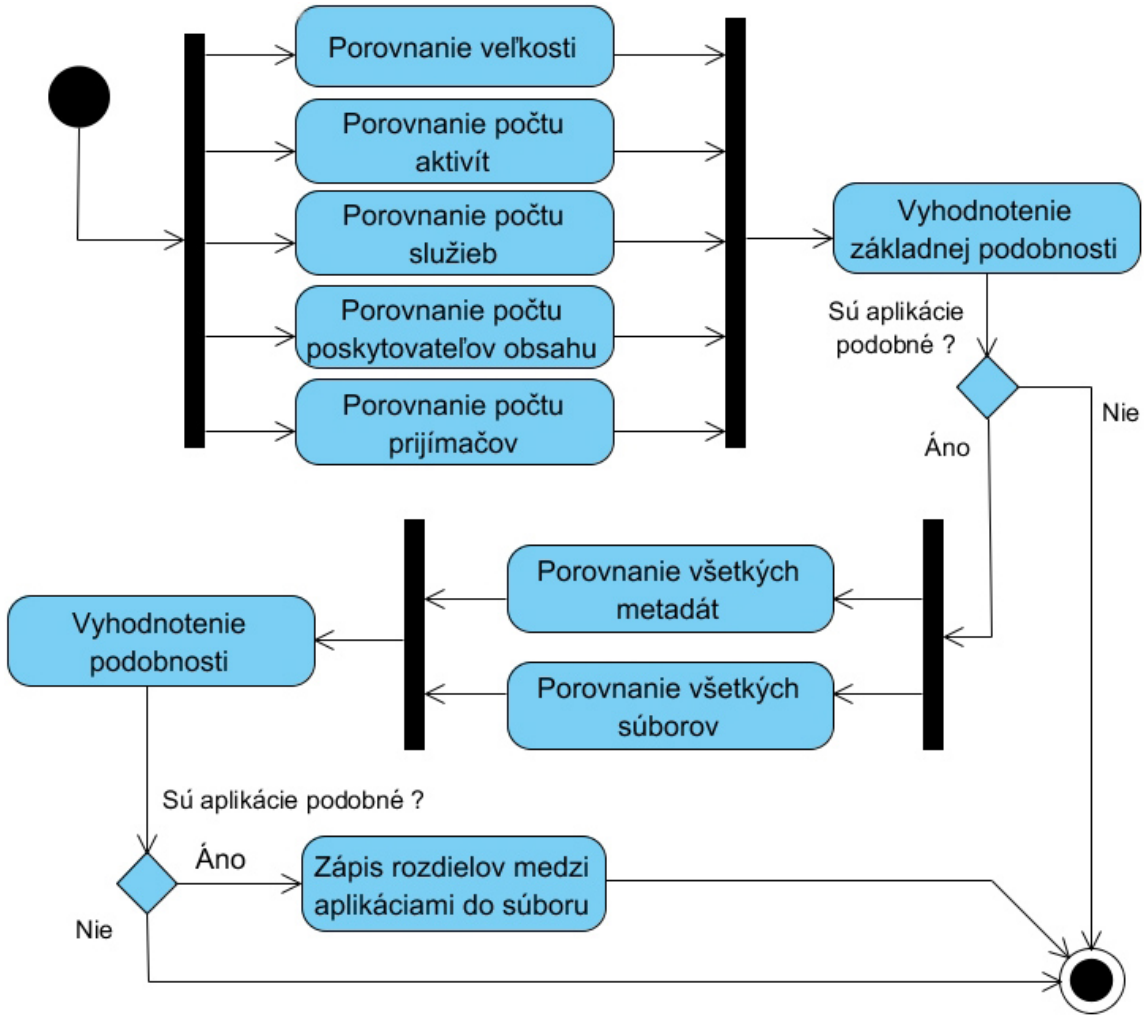
Prebalené APK súbory

- Apk súbory je možné rozbaliť, modifikovať a zabaliť do pôvodnej podoby
- Android obsahuje ochranný mechanizmus
  - súbor MANIFEST.MF obsahuje hash každého súboru - ochrana integrity
- Jednoduché vyhnutie ochrannému mechanizmu
  - po každej zmene nutné balíček podpísať
- Spôsob šírenia malvéru

Detekcia prebalených APK súborov

- Existujúce riešenia
  - analýza zdrojového kódu
  - analýza súborov v APK balíčku
- Implementovaná metóda detekcie nadmieru podobných APK súborov pomocou podobnosti súborov a metadát o aplikáciách
- Malvérové aplikácie zachovávajú look & feel pôvodných
- Funkcionalita obsiahnutá v aplikácií ApkAnalyzer
- Vyuzitie metadát získaných analýzou APK súborov
- Párové porovnanie APK súborov

Navrhnutá metóda detekcie prebalených APK súborov



Výstup detekcie prebalených APK súborov

- Výstup párového porovnania obsahuje detailné rozdiely medzi jednotlivými aplikáciami
- Na základe zhody certifikátu a verzie aplikácie sú vyhodnotené potenciálne škodlivé aplikácie
- Identifikovaných 161 nadmieru podobných APK súborov s rovnakou verziou podpísaných rôznymi certifikátmi

Bibliografia

Analýza inštalačných APK súborov pre OS Android

[1] www.gartner.com/newsroom/id/3169417.

[2] https://github.com/Akdeniz/google-play-crawler

[3] github.com/MartinStyk/ApkDownloader

[4] github.com/MartinStyk/ApkAnalyzer

[5] ibotpeaches.github.io/Apktool/

[6] axml