Masarykova univerzita
Fakulta informatiky

Analýza inštalačných APK súborov pre OS Android

Bakalárska práca

Martin Styk

Brno, jar 2016

Masarykova univerzita
Fakulta informatiky

Analýza inštalačných APK súborov pre OS Android

Bakalárska práca

Martin Styk

Brno, jar 2016

Namiesto tejto stránky vložte kópiu oficiálneho podpísaného zadania práce a prehlásenie autora školského diela.

Prehlásenie

Prehlasujem, že táto bakalárska práca je mojím pôvodným autorským dielom, ktoré som vypracoval samostatne. Všetky zdroje, pramene a literatúru, ktoré som pri vypracovaní používal alebo z nich čerpal, v práci riadne citujem s uvedením úplného odkazu na príslušný zdroj.

Martin Styk

Vedúci práce: Ing. Mgr. et Mgr. Zdeněk Říha, Ph.D.

Poďakovanie

Rád by som sa poďakoval vedúcemu práce Ing. Mgr. et Mgr. Zdeňkovi Říhovi, Ph.D. za venovaný čas, ochotu a cenné pripomienky, ktoré mi pomohli pri tvorbe tejto práce.

Zhrnutie

Práca sa zaoberá získavaním metadát o inštalačných APK súboroch pre mobilný operačný systém Android. V rámci práce je vytvorená rozsiahla databáza APK balíčkov. Na základe analýzy týchto súborov sú určené štatistické vlastnosti APK súborov a príslušných aplikácií. Ako súčasť tejto práce je implementovaný nástroj na hromadné sťahovanie APK súborov, ich analýzu a výpočet štatistických dát nad množinou APK súborov. Práca sa zaoberá aj bezpečnosťou aplikácií a detekciou modifikovaných APK súborov. V práci je navrhnutá metóda detekcie upravených a prebalených APK balíčkov, ktorá je aj prakticky implementovaná. V teoretickej časti je popísaná štruktúra APK balíčkov a súborov v nich obsiahnutych.

Kľúčové slová

APK súbor, Android, Apktool, malvér, analýza aplikácií, AndroidManifest.xml

Obsah

1 **Úvod** 1
2 **Štatistiky** 3
2.1 *Získané dáta* 3
2.1.1 Podpis APK balíčka 6
Literatúra 7
Register 9
A **An appendix** 11

Zoznam tabuliek

2.1 Najpoužívanejšie prístupové oprávnenia *5*

A.1 Lokalizácia aplikácií *11*
A.5 Zbierané metadata o APK súbore *13*
A.2 Najpoužívanejšie vlastnosti *14*
A.3 Hodnoty najnižsej vyžadovanej verzie Android SDK *14*
A.4 Hodnoty cieľovej verzie Android SDK *15*

Zoznam obrázkov

2.1 Hodnoty atribútu *android:installLocation* *4*
2.2 Algoritmus podpisu APK balíčku *6*

1 Úvod

TBD

2 Štatistiky

Analýzou jednotlivých APK súborov získame detailné informácie o jednotlivých aplikáciách. Pre ucelenejší pohľad na všeobecné vlastnosti a atribúty Android aplikácií je vhodné rozšíriť analýzu jednotlivých aplikácií na skúmanie väčšej množiny APK balíčkov. Keďže databáza APK súborov použitá v tejto práci obsahuje dostatočne veľkú vzorku približne 20000 APK súborov, ktoré pochádzajú z rôznych oficiálnych aj alternatívnych zdrojov, poskytuje dobrú vzorku na určenie štatistických údajov o Android aplikáciách. Štatistické informácie prezentované v tejto kapitole sa viažu k aplikáciám dostupným v rokoch 2014–2016 a teda sú aktuálne pre spomenuté obdobie. Vyvinutá aplikácia *ApkAnalyzer* poskytuje možnosť výpočtu štatistík nad množinou APK súborov. Funkcionalita výpočtu štatistických informácií sa aktivuje pomocou prepínača *–statistics* pri spustení programu z príkazového riadku. Ako vstup aplikácie slúžia JSON súbory vytvorené analýzou popísanou v kapitole **??**. Výstupom je súbor vo formáte JSON obsahujúci vypočítané štatistické dáta. Pri vlastnostiach, ktorých hodnota je vyjadrená číselne sú vypočítané základné matematické štatistiky ako je aritmetický priemer, modus, medián, rozptyl, smerodajná odchýlka, minimum a maximum. Pri najnižšej a najvyššej hodnote obsahuje výstup aj názov aplikácií, ktoré tieto hodnoty dosahujú. V prípade vlastností, ktoré nadobúdajú obmedzený počet predom definovaných hodnôt je určené percentuálne zastúpenie jednotlivých hodnôt.

2.1 Získané dáta

Veľkosť APK súborov

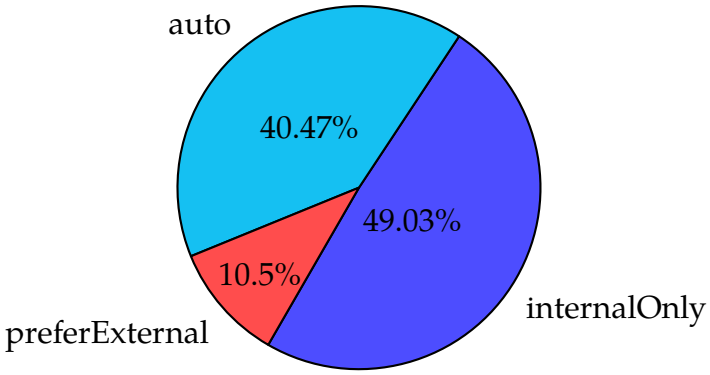
Analýzou databázy APK súborov sa zistilo, že stredná hodnota veľkosti APK súborov je 5,26 MB, priemer dosahuje hodnotu 10.19 MB.

Počet súborov v APK balíku

Strednou hodnotou celkového počtu súborov v APK balíčku je 397, aritmetický priemer má hodnotu približne 730 súborov.

Inštalačná politika

Android poskytuje aplikáciám možnosť špecifikácie preferovaného pamäťového priestoru (interná alebo externá pamäť) a prípadnú možnosť presunutia nainštalovanej aplikácie (viď **??**). Až 49 % aplikácií neumožňuje inštaláciu alebo presun na externé pamäťové médiá. 40 % aplikácií preferuje inštaláciu na interné úložisko s možnosťou presunu do externej pamäte. 10,5 % aplikácií uprednostňuje inštaláciu na externé pamäťové médium. Rozdelenie hodnôt je zobrazené v grafe 2.1.



Obr. 2.1: Hodnoty atribútu *android:installLocation*

Komponenty aplikácií

Základnou funkčnou jednotkou Android aplikácií sú aktivity (viď **??**). Stredná hodnota počtu aktivít medzi analyzovanými aplikáciami je 10, priemer dosahuje hodnotu 20,23. Aplikácie obsahujú najčastejšie 2 aktivity.
Priemerný počet služieb definovaných v aplikácií je 3,99, stredná hodnota je 1, no najčastejším prípadom je, že aplikácia nedefinuje žiadnu službu.

Verzie Android SDK

Najčastejšou najnižšou vyžadovanou verziou Android SDK je verzia 9 s 21,3% zastúpením. Naniższie vyžadované verzie Android SDK v našej databáze APK súborov sú zobrazené v grafe A.3. Až 25,64 % aplikácií je primárne určených na SDK verziu 19.

Prístupové oprávnenia

Android aplikácie najčastejšie deklarujú, že využívajú 4 prístupové oprávnenia (viď **??**). Stredná hodnota počtu vyžadovaných oprávnení je 8. 10 najčastejšie využívaných oprávnení spolu s ich percentuálnym zastúpením v analyzovanej vzorke aplikácií je uvádených v tabuľke 2.1.

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
android.permission.vending.billing	27,1

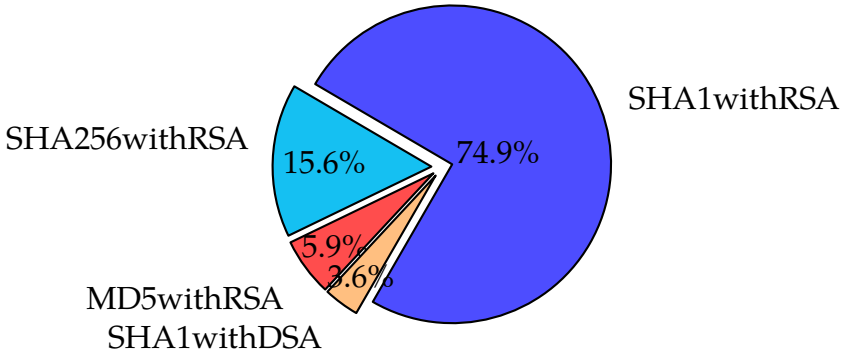
Tabuľka 2.1: Najpoužívanejšie prístupové oprávnenia

Využité vlastnosti

Aplikácie deklarujú nízky počet využívaných vlastností (viď **??**). Aritmetický priemer je 1,44, stredná hodnota a modus sú nulové. Najčastejšie deklarované je využívanie vlastností uvedených v tabľke A.2.

2.1.1 Podpis APK balíčka

Na podpisovanie APK balíčkov je v najčastejšie využitý algoritmus *SHA1withRSA*, ktorý využíva až 74,87 % aplikácií. 15,56 % aplikácií je podpísaných pomocou algoritmu *SHA256withRSA*, podpis pomocou *MD5withRSA* je využitý v 5,88 % prípadoch.



Obr. 2.2: Algoritmus podpisu APK balíčku

Lokalizácia

Aplikácie sú okrem základného jazyka lokalizované najčastejšie v 17 iných jazykoch. Aritmetický priemer počtu lokalizácií je 31,42. Najčastejšie lokalizácie aplikácií sú uvedené v tabuľke A.1. Lokalizácie sú ovplyvnené tým, že časť aplikácií bola stiahnutá z portálov určených pre strednú európu. V českom jazyku je lokalizovaných 49 % aplikácií, v slovenčine 46 %.

Obrázkové súbory

Analýza ukázala, že aplikácie využívajú množstvo obrázkových súborov. Stredná hodnota celkového počtu obrázkových súborov je 210, aritmetický priemer dosahuje hodnotu 462 a najčastejším počtom obrázkových súborov je 5 . Stredná hodnota počtu rozdielnych obrázkov (veľkosť a rozlíšenie sa neberie do úvahy) je 134. Najčastejším formátom je PNG, aplikácia obsahuje priemerne 342,8 takýchto súborov.

Literatúra

1. Freed, Ned; Kucherawy, Murray; Baker, Mark; Hoehrmann, Bjoern. *Media Types* [online]. 2016 [visited on 2016-03-23]. Available from WWW: ⟨http://www.iana.org/assignments/media-types/media-types.xhtml⟩.
2. *Building and Running Overview* [online]. 2016 [visited on 2016-03-23]. Available from WWW: ⟨http://developer.android.com/tools/building/index.html⟩.
3. Yang, Herong. *META-INF Files - Digests, Signature and Certificate* [online]. 2015 [visited on 2016-03-24]. Available from WWW: ⟨http://www.herongyang.com/Android/Project-META-INF-Files-Digest-Signature-and-Certificate.html⟩.
4. *Accessing Resources* [online]. 2015 [visited on 2016-03-24]. Available from WWW: ⟨http://developer.android.com/guide/topics/resources/accessing-resources.html⟩.
5. *Providing Resources* [online]. 2015 [visited on 2016-03-24]. Available from WWW: ⟨http://developer.android.com/guide/topics/resources/providing-resources.html⟩.
6. *Providing Alternative Resources* [online]. 2015 [visited on 2016-03-24]. Available from WWW: ⟨http://developer.android.com/guide/topics/resources/providing-resources.html%5C#AlternativeResources⟩.
7. Reddy, Satheesh. *Android Application Build Process or Compilation Process* [online]. 2014 [visited on 2016-03-24]. Available from WWW: ⟨http://www.c-sharpcorner.com/UploadFile/34ef56/android-application-build-process-or-compilation-process/⟩.
8. *ART and Dalvik* [online]. 2015 [visited on 2016-03-23]. Available from WWW: ⟨https://source.android.com/devices/tech/dalvik/⟩.
9. *App Manifest* [online]. 2015 [visited on 2016-03-24]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/manifest-intro.html⟩.
10. *Manifest element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/manifest-element.html⟩.
11. *Uses-permission element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/uses-permission-element.html⟩.

12. *Permission element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/permission-element.html⟩.
13. *Uses-sdk element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/uses-sdk-element.html⟩.
14. *Uses-feature element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/uses-feature-element.html⟩.
15. *Supports-screens element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/supports-screens-element.html⟩.
16. *Activity* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/reference/android/app/Activity.html⟩.
17. *Service* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/reference/android/app/Service.html⟩.
18. *ContentProvider* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/reference/android/content/ContentProvider.html⟩.
19. *Receiver element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/receiver-element.html⟩.
20. *Uses-library element* [online]. 2015 [visited on 2016-03-26]. Available from WWW: ⟨http://developer.android.com/guide/topics/manifest/uses-library-element.html⟩.
21. Westenberg, Jimmy. *Gartner: Android and iOS dominate smartphone market with 98 percent marketshare* [online]. 2015 [visited on 2016-03-23]. Available from WWW: ⟨http://www.androidauthority.com/android-ios-hold-98-percent-marketshare-656624/⟩.
22. Thomas, Daniel R.; Beresford, Alastair R.; Rice, Andrew. Security Metrics for the Android Ecosystem. *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices - SPSM '15*. 2015, pp. 87–98. Available also from WWW: ⟨http://dl.acm.org/citation.cfm?doid=2808117.2808118⟩.

23. *Industry Leaders Announce Open Platform for Mobile Devices* [online]. 2007 [visited on 2016-03-23]. Available from WWW: ⟨http://www.openhandsetalliance.com/press_110507.html⟩.
24. Rosoff, Matt. *Google's Biggest Acquisitions So Far, And What They Became* [online]. 2011 [visited on 2016-03-23]. Available from WWW: ⟨http://www.gizmodo.com.au/2011/08/googles-16-biggest-acquisitions-so-far-and-what-happened-to-them/⟩.
25. Beavis, Gareth. *A complete history of Android* [online]. 2008 [visited on 2016-03-23]. Available from WWW: ⟨http://www.techradar.com/news/phone-and-communications/mobile-phones/a-complete-history-of-android-470327⟩.
26. *An Overview of the Android Architecture* [online]. 2013 [visited on 2016-03-23]. Available from WWW: ⟨http://www.techotopia.com/index.php/An_Overview_of_the_Android_Architecture⟩.
27. Parmar, Ketan. *In Depth: Android Package Manager and Package Installer*. 2013. Available also from WWW: ⟨https://dzone.com/articles/depth-android-package-manager⟩.
28. Elenkov, Nikolay. *Android security internals: an in-depth guide to android's security architecture*. San Francisco: No Starch Press, 2015. ISBN 978-1-59327-641-6.

Register

α, 31

dummy text, 31

TEX, 31

vehicles
    speed cars, 31
    trucks, 31

A An appendix

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
zh-rtw	čínsky (tradičný)	54,0
pt	portugalský	52,6

Tabuľka A.1: Lokalizácia aplikácií

Názov atribútu	Dátový typ	popis
fileName	String	Názov analyzovaného APK súbor
sourceOfFile	String	Zdroj súboru
fileSize	Long	Veľkosť APK súboru v bajtoch
dexSize	Long	Veľkosť súboru *classes.dex* v bajtoc
arscSize	Long	Veľkosť súboru *arscSize.dex* v bajtc
packageName	String	Hodnota atribútu *package* v eleme
versionCode	String	Hodnota atribútu *android:version*
installLocation	String	Hodnota atribútu *android:installLo*
numberOfActivities	Integer	Počet aktivít definovaných aplikác
numberOfServices	Integer	Počet služieb definovaných aplikác
numberOfContentProviders	Integer	Počet poskytovateľov obsahu defi
numberOfBroadcastReceivers	Integer	Počet komponent typu *BroadcastR*
namesOfActivities	List<String>	Názvy aktivít definovaných aplik

namesOfServices	List<String>	Názvy služieb definovaných aplikáciou
namesOfContentProviders	List<String>	Názvy poskytovateľov obsahu definov
namesOfBroadcastReceivers	List<String>	Názvy komponent typu *BroadcastRecei*
usesPermissions	List<String>	Názvy povolení využívaných aplikáci
usesLibrary	List<String>	Názvy knižníc využívaných aplikáciou
permissions	List<String>	Názvy povolení definovaných aplikáci
permissionsProtectionLevel	List<String>	Level ochrany povolení definovaných a
usesFeature	List<String>	Názvy vlastností využívaných aplikáci
usesMinSdkVersion	String	Hodnota atibútu *android:minSdkVersion*
usesTargetSdkVersion	String	Hodnota atribútu *android:targetSdkVers*
usesMaxSdkVersion	String	Hodnota atribútu *android:maxSdkVersio*
supportsScreensResizeable	Boolean	Hodnota atribútu *android:resizeable* ele
supportsScreensSmall	Boolean	Hodnota atribútu *android:smallScreens*
supportsScreensNormal	Boolean	Hodnota atribútu *android:normalScreen*
supportsScreensLarge	Boolean	Hodnota atribútu *android:largeScreens* e
supportsScreensXlarge	Boolean	Hodnota atribútu *android:xlargeScreens*
supportsScreensAnyDensity	Boolean	Hodnota atribútu *android:anyDensity* e
fileName	String	Názov súboru s certifikátom
signAlgorithm	String	Algoritmus použitý na podpis
signAlgorithmOID	String	OID algoritmu použitého na podpis
startDate	Date	začiatok platnosti certifikátu
endDate	Date	koniec platnosti certifikátu
publicKeyMd5	String	MD5 hash verejného klúča
certBase64Md5	String	Base64 MD5 hash certifikátu
certMd5	String	MD5 hash certifikátu
version	Integer	Verzia cetifikátu
issuerName	String	Názov vydávateľa vo formáte definova
subjectName	String	Názov subjektu vo formáte definovanc
locale	List<String>	Lokalizácie súboru *string.xml*
numberOfStringResource	Integer	Počet záznamov v súbore *string.xml*

pngDrawables	Integer	Počet PNG obrázkov
ninePatchDrawables	Integer	Počet 9.PNG obrázkov
jpgDrawables	Integer	Počet JPG obrázkov
gifDrawables	Integer	Počet GIF obrázkov
xmlDrawables	Integer	Počet XML obrázkov
ldpiDrawables	Integer	Počet obrázkov v ldpi priečinku
mdpiDrawables	Integer	Počet obrázkov v mdpi priečinku
hdpiDrawables	Integer	Počet obrázkov v hdpi priečinku
xhdpiDrawables	Integer	Počet obrázkov v xhdpi priečinku
xxhdpiDrawables	Integer	Počet obrázkov v xxhdpi priečink
xxxhdpiDrawables	Integer	Počet obrázkov v xxxhdpi priečin
tvdpiDrawables	Integer	Počet obrázkov v tvdpi priečinku
nodpiDrawables	Integer	Počet obrázkov v nodpi priečinku
unspecifiedDpiDrawables	Integer	Počet obrázkov nezaradených v *
rawResources	Integer	Počet súborov v *raw/* priečinku
layouts	Integer	Počet súborov v *res/layout** prieči
differentLayouts	Integer	Počet rôznych súborov v *res/layou*
menu	Integer	Počet súborov v *res/menu* priečink
dexHash	String	Hash súboru *classes.dex* prevzatý
arscHash	String	Hash súboru *arscHash.dex* prevza
drawableHash	Map<String,String>	Hashe a cesty k súborov z prečink
layoutHash	Map<String,String>	Hashe a cesty k súborov z prečink
otherHash	Map<String,String>	Hashe a cesty k všetkým ostatný

Tabuľka A.5: Zbierané metadata o AP

Názov	%
android.hardware.camera	18,1
android.hardware.touchscreen	16,1
android.hardware.telephony	14,8
android.hardware.camera.autofocus	10,6
android.hardware.location.gps	10,2
android.hardware.location	8,8
android.hardware.wifi	8,4
android.hardware.location.network	7,0
android.hardware.bluetooth	6,6
android.hardware.touchscreen.multitouch	6,0

Tabuľka A.2: Najpoužívanejšie vlastnosti

Verzia Android SDK	%
9	21,3
8	18,4
7	14,2
14	10,5
10	8,1
4	7,0
3	5,6
15	3,7
5	3,7
11	2,1

Tabuľka A.3: Hodnoty najnižsej vyžadovanej verzie Android SDK

Verzia Android SDK	%
19	25,6
17	11,8
21	11,7
15	6,8
14	6,3
22	6,0
16	5,7
18	5,6
20	3,8
8	2,9

Tabuľka A.4: Hodnoty cieľovej verzie Android SDK