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

Zoznam tabuliek

Zoznam obrázkov

1 Úvod

TBD

2 Databáza inštalačných APK súborov

Základnou úlohou tejto práce je vytvoriť dostatočne veľkú databázu inštalačných APK balíčkov. Pre ďalšie potreby práce bolo požadované, aby veľká časť aplikácií pochádzala z neoficiálnych zdrojov, čím sa zvyšuje pravdepodobnosť, že aplikácia obsahuje malvér.

Naša databáza pozostáva približne z 20000 Android aplikácií. Tie boli zaobstarané v časovom rozmedzí medzi novembrom 2015 a februárom 2016. Žiadna z aplikácií nebola stiahnutá priamo z obchodu *Google Play*, ale veľká časť bola získaná s využitím projektu *Playdrone*. V rámci tohto projektu bolo v novembri 2014 z *Google Play* stiahnutých viac ako milión aplikácií dostupný pre zariadenie *Galaxy Nexus* s operátorom *T-Mobile* [**Viennot2014**]. Naša databáza obsahuje 8200 najsťahovanejších aplikácií z *Google Play* v období november 2014, ktoré boli stiahnuté z archívu projektu *Playdrone*.

Celková veľkosť všetkých stiahnutých APK súborov je 192 GB. Prehľad všetkých zdrojov APK súborov a ich počet zobrazuje tabuľka 2.1.

Zdroj	Počet stiahnutých aplikácií
Playdrone[1]	8200
www.appsapk.com	6470
www.apkmaniafull.com	2870
www.androidapksfree.com	1030
www.zippyshare.com	750
torrenty	550
www.uloz.to	190
Spolu	20060

Tabuľka 2.1: Zdroje prevzatých APK súborov

2.1 Implementácia

Viac ako 90 % aplikácií bolo stiahnutých automatizovane prostredníctvom aplikácie *ApkDownloader* implementovanej v rámci tejto práce. Aplikácia neposkytuje grafické užívateľské rozhranie, ale užívateľ môže zadávať parametre prostredníctvom príkazového riadku. Podporuje sťahovanie aplikácií získaných pomocou projektu *Playdrone* alebo z neoficiálnych lokalít zameraným na distribúciu Android aplikácií *www.appsapk.com*, *www.apkmaniafull.com* alebo *www.androidapksfree.com*. Aplikácia funguje na jednoduchom princípe, keď najskôr získa zoznam URL odkazov na APK súbory, ktoré následne stiahne. Užívateľ pomocou parametrov špecifikuje z ktorej podporovanej lokality chce APK súbory stiahnuť, ich želaný počet, umiestnenie prebraných súborov a maximálny počet súbežných preberaní. Pri vyhľadávaní URL odkazov je na prácu s HTML súbormi použitá open source knižnica *jsoup*. Pri sťahovaní sa využíva knižnica *HtmlUnit*, ktorá poskytuje funkcionalitu internetového prehliadača. Na preberanie súborov z URL odkazov je použitá knižnica *Apache Commons IO*. Keďže je *ApkDownloader* open source, môže byť jednoducho rozšírený o podporu sťahovania APK súborov z nových lokalít.
Torrent súbory boli získane automatizovane s využitím knižnice *flux*[2].

[2] https://github.com/ProjectMoon/flux

3 These are

3.1 the available

3.1.1 sectioning commands.

Paragraphs and

subparagraphs are available as well. Inside the text, you can also use unnumbered lists,

- such as
- this one
    - and they can be nested as well.
    - » You can even turn the bullets into something fancier,
    - § if you so desire.

Numbered lists are

1. very
    (a) similar

and so are description lists:

Description list A list of terms with a description of each term

The spacing of these lists is geared towards paragraphs of text. For lists of words and phrases, the paralist package offers commands

- that
    - are
        - ∗ better
            - · suited

1. to
    (a) this
        i. kind of
            A. content.

4 Floats and references

The logo of the Masaryk University is shown in Figure 4.1 and Figure 4.2 at pages 7 and 8. The weather forecast is shown in Table 4.1 at page 8. The following chapter is Chapter 5 and starts at page 9. Items 3, 3b, and 3(c)iv are starred in the following list:

1. some text
2. some other text
3. ⋆
   (a) some text
   (b) ⋆
   (c) some other text
      i. some text
      ii. some other text
      iii. yet another piece of text
      iv. ⋆
   (d) yet another piece of text
4. yet another piece of text

If your reference points to a place that has not yet been typeset, the `\ref` command will expand to **??** during the first run of `pdflatex thesis.tex` and a second run is going to be needed for the references to resolve. With online services – such as Overleaf – this is performed automatically.

Obr. 4.1: The logo of the Masaryk University at 40 mm

UNIVERSITAS · MASARYKIANA BRUNENSIS ·
UNIVERSITAS · MASARYKIANA BRUNENSIS ·
UNIVERSITAS · MASARYKIANA BRUNENSIS ·

Obr. 4.2: The logo of the Masaryk University at $\frac{2}{3}$ and $\frac{1}{3}$ of text width

Day	Min Temp	Max Temp	Summary
Monday	13°C	21°C	A clear day with low wind and no adverse current advisories.
Tuesday	11°C	17°C	A trough of low pressure will come from the northwest.
Wednesday	10°C	21°C	Rain will spread to all parts during the morning.

Tabuľka 4.1: A weather forecast

5 Mathematical equations

TeX comes pre-packed with the ability to typeset inline equations, such as $\mathrm{e}^{ix} = \cos x + i \sin x$, and display equations, such as

$$\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

LaTeX defines the automatically numbered `equation` environment:

$$\gamma Px = PAx = PAP^{-1}Px. \tag{5.1}$$

The package amsmath provides several additional environments that can be used to typeset complex equations:

1. An equation can be spread over multiple lines using the `multline` environment:

$$\begin{multline} a+b+c+d+e+f+b+c+d+e+f+b+c+d+e+f \\ +f+g+h+i+j+k+l+m+n+o+p+q \tag{5.2} \end{multline}$$

2. Several aligned equations can be typeset using the `align` environment:

$$\begin{aligned} a+b &= c+d & (5.3) \\ u &= v+w+x & (5.4) \\[1ex] i+j+k+l &= m & (5.5) \end{aligned}$$

3. The `alignat` environment is similar to `align`, but it doesn't insert horizontal spaces between the individual columns:

$$\begin{alignedat}{2} a+b+c+d & &&= 0 \qquad (5.6) \\ & e+f+g &&= 5 \qquad (5.7) \end{alignedat}$$

4. Much like chapter, sections, tables, figures, or list items, equations – such as (5.8) and (My equation) – can also be labeled and referenced:

$$\begin{alignedat}{5} b_{11}x_1 &+ b_{12}x_2 &&+ b_{13}x_3 && &&= y_1, &\qquad (5.8) \\ b_{21}x_1 &+ b_{22}x_2 && &&+ b_{24}x_4 &&= y_2. &\qquad \text{(My equation)} \end{alignedat}$$

5. The `gather` environment makes it possible to typeset several equations without any alignment:

$$\begin{gather}\psi = \psi\psi, \tag{5.9}\\ \eta = \eta\eta\eta\eta\eta\eta, \tag{5.10}\\ \theta = \theta. \tag{5.11}\end{gather}$$

6. Several cases can be typeset using the `cases` environment:

$$|y| = \begin{cases} y & \text{if } z \geq 0, \\ -y & \text{otherwise.} \end{cases} \tag{5.12}$$

For the complete list of environments and commands, consult the amsmath package manual[1].

[1] See `http://mirrors.ctan.org/macros/latex/required/amslatex/math/amsldoc.pdf`. The `\url` command is provided by the package url.

6 We have `several` FONTS *at* **disposal**

The serified roman font is used for the main body of the text. *Italics are typically used to denote emphasis or quotations.* `The teletype font is typically used for source code listings.` The **bold**, SMALL-CAPS and sans-serif variants of the base roman font can be used to denote specific types of information.

We can also change the font size, although it is usually not necessary.

A wide variety of mathematical fonts is also available, such as:

$$\mathrm{ABC}, \mathcal{ABC}, \mathbf{ABC}, \mathsf{ABC}, \mathit{ABC}, \mathtt{ABC}$$

By loading the amsfonts packages, several additional fonts will become available:

$$\mathfrak{ABC}, \mathbb{ABC}$$

Many other mathematical fonts are available[1].

[1] See `http://tex.stackexchange.com/a/58124/70941`.

7 Inserting the bibliography

After loading the `biblatex` package and linking a bibliography database file to the document using the `\addbibresource` command, you can start citing the entries. This is just dummy text [**inbook-full**] lightly sprinkled with citations [**incollection-full**]. Several sources can be cited at once [**whole-collection**, **manual-minimal**, **manual-full**]. **inbook-full** was written by **inbook-full** in **inbook-full** We can also produce **inbook-full** or (**inbook-full**, **inbook-full**). The full bibliographic citation is: **inbook-full**. We can easily insert a bibliographic citation into the footnote[1].

The `\nocite` command will not generate any output, but it will insert its argument into the bibliography. The `\nocite{*}` command will insert all the records in the bibliography database file into the bibliography. Try uncommenting the command and watch the bibliography section come apart at the seams.

When typesetting the document for the first time, citing a `work` will expand to [**work**] and the `\printbibliography` command will produce no output. It is now necessary to generate the bibliography by running `biber thesis.bcf` from the command line and then by typesetting the document again twice. During the first run, the bibliography section and the citations will be typeset, and in the second run, the bibliography section will appear in the table of contents.

The `biber` command needs to be executed from within the directory, where the LaTeX source file is located. In Windows, the command line can be opened in a directory by holding down the Shift key and by clicking the right mouse button while hovering the cursor over a directory. Select the Open Command Window Here option in the context menu that opens shortly afterwards.

With online services – such as Overleaf – all commands are executed automatically.

[1] **inbook-full**.

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.

8 Inserting the index

After using the `\makeindex` macro and loading the `makeidx` package that provides additional indexing commands, index entries can be created by issuing the `\index` command. It is possible to create ranged index entries, which will encompass a span of text. To insert complex typographic material – such as α or TEX – into the index, you need to specify a text string, which will determine how the entry will be sorted. It is also possible to create hierarchal entries.

After typesetting the document, it is necessary to generate the index by running

```
texindy -I latex -C utf8 -L ⟨locale⟩ thesis.idx
```

from the command line, where ⟨*locale*⟩ corresponds to the main locale of your thesis – such as `english`, and then typesetting the document again.

The `texindy` command needs to be executed from within the directory, where the LATEX source file is located. In Windows, the command line can be opened in a directory by holding down the Shift key and by clicking the right mouse button while hovering the cursor over a directory. Select the Open Command Window Here option in the context menu that opens shortly afterwards.

With online services – such as Overleaf – the commands are executed automatically, although the locale may be erroneously detected, or the `makeindex` tool (which is only able to sort entries that contain digits and letters of the English alphabet) may be used instead of `texindy`. In either case, the index will be ill-sorted.

Register

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	%
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

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 A.3: Najpoužívanejšie prístupové oprávnenia

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.4: 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.5: Hodnoty cieľovej verzie Android SDK