Claim What You Need: Towards a Simplified Permission Requests Management for Android Mobile Apps

Mingkui Wei, Xi Gong and Wenye Wang
IEEE GlobeCom 2016
NC State University
April, 2013
Outline

1 Motivation
2 Related Work
3 System Model
4 Evaluation
5 Conclusion
Outline

1 Motivation
2 Related Work
3 System Model
4 Evaluation
5 Conclusion
Motivation

Malware

Malware, short for malicious (or malevolent) software, is software used or programmed by attackers to disrupt computer operation, gather sensitive information, or gain access to private computer systems.

Malware in Mobile phone market

(a) Platform Distribution

(b) Mobile Malware Distribution

Kaspersky Lab
Motivation

Background: Android permission mechanism
Motivation

Background: Android permission mechanism

However...
Motivation

Android permission mechanism: not effective

Don’t understand...
Got used to...

...
Motivation

Android permission mechanism: not effective

Therefore...
Motivation

Question

can we precisely summarize the functionality of an app and furthermore predict the corresponding necessary permissions?
Background and Related Work

Enck et al: Compare requested permissions against a preset-rule, which defines “dangerous” permissions-combination

Peng et al: App risk ranking based on score, which was derived from studying permission combination pattern

Code based approach.

Shortage

Computationally intensive: Download package, unzip package, find certain file, parse permission requested, ....

Code not always available: packages are encrypted

Error prone: “Dangerous combination” is not dangerous, as long as developers mean it, and users know it
Background and Related Work

**Related Work**

Enck et al: Compare requested permissions against a preset-rule, which defines “dangerous” permissions-combination

Peng et al: App risk ranking based on score, which was derived from studying permission combination pattern

Code based approach.
Background and Related Work

Related Work

Enck et al: Compare requested permissions against a preset-rule, which defines “dangerous” permissions-combination

Peng et al: App risk ranking based on score, which was derived from studying permission combination pattern

Code based approach.

Shortage

- Computationally intensive: Download package, unzip package, find certain file, parse permission requested, ....
- Code not always available: packages are encrypted
- Error prone: “Dangerous combination” is not dangerous, as long as developers mean it, and users know it
Outline

1 Motivation
2 Related Work
3 System Model
4 Evaluation
5 Conclusion
System Model: Introduction

Intuition

Goal

Detect inconsistency between Description and Permission requested.

Regulate developers’ behavior: write sensible description, request reasonable permission.
OfficeSuite Pro includes a File Browser and integration with Cloud service readers: Box, DropBox, Google Drive, SkyDrive and SugarSync to give users a multitude of options on how to manage files and attachments.

**STORAGE**
- MODIFY OR DELETE THE CONTENTS OF YOUR USB STORAGE
- MODIFY OR DELETE THE CONTENTS OF YOUR SD CARD
  
  Allows the app to write to the USB storage. Allows the app to modify or delete the contents of the USB storage.

**NETWORK COMMUNICATION**
- FULL NETWORK ACCESS
  
  Allows the app to create network sockets and use custom network protocols. The browser and other applications provide means to send data to the internet, so this permission is not required to send data to the internet.
System Model: Introduction

Intuition

OfficeSuite Pro includes a File Browser and integration with Cloud service providers: Box, DropBox, Google Drive, SkyDrive and SugarSync to give users a multitude of options on how to manage files and attachments.

Goal

- Detect inconsistency between *Description* and *Permission requested*.
- Regulate developers’ behavior: write sensible description, request reasonable permission.
Example 1: Reasonable
System Model: Example

Example 1: Reasonable

NBA 3D Live Wallpaper
Cellfish Studios
🌟 TOP DEVELOPER

🌟🌟🌟🌟 (4.738)
INSTALL
Example 1: Reasonable

NBA 3D Live Wallpaper
Cellfish Studios
TOP DEVELOPER

⭐⭐⭐⭐⭐ (4.738)
INSTALL

YOUR MESSAGES
READ YOUR TEXT MESSAGES (SMS OR MMS)
Allows the app to read SMS messages stored on your tablet or SIM card. This allows the app to read all SMS messages, regardless of content or confidentiality. Allows the app to read SMS messages stored on your phone or SIM card. This allows the app to read all SMS messages, regardless of content or confidentiality.

DEFAULT
ACCESS MAIL INFORMATION
Allows the app to access information about your mail.
Example 1: Reasonable

NBA 3D Live Wallpaper
Cellfish Studios
TOP DEVELOPER

* Four stars (4.738)

INSTALL

**YOUR MESSAGES**
READ YOUR TEXT MESSAGES (SMS OR MMS)
Allows the app to read SMS messages stored on your tablet or SIM card. This allows the app to read all SMS messages, regardless of content or confidentiality. Allows the app to read SMS messages stored on your phone or SIM card. This allows the app to read all SMS messages, regardless of content or confidentiality.

**DEFAULT**
ACCESS MAIL INFORMATION
Allows the app to access information about your mail.

* Choose from 5 different basketball styles
* The ability to "shoot baskets" right on your home screen (with sounds)
* Animated fans in team colors
* SMS and Email notification in ticker
* Day and Night detection with in arena lighting.
* Classic Mode for a top-down court display
* Made with official release software to ensure quality and performance
* Officially licensed release from the NBA
* All 30 teams have a live wallpaper available for in-app purchase
System Model: Example

Example 2: Unreasonable
System Model: Example

Example 2: Unreasonable
**Example 2: Unreasonable**

![American Football Wallpapers](image)

**YOUR LOCATION**

**APPROXIMATE LOCATION (NETWORK-BASED)**
Allows the app to get your approximate location. This location is derived by location services using network location sources such as cell towers and Wi-Fi. These location services must be turned on and available to your device for the app to use them. Apps may use this to determine approximately where you are.

**PRECISE LOCATION (GPS AND NETWORK-BASED)**
Allows the app to get your precise location using the Global Positioning System (GPS) or network location sources such as cell towers and Wi-Fi. These location services must be turned on and available to your device for the app to use them. Apps may use this to determine where you are, and may consume additional battery power.

**YOUR PERSONAL INFORMATION**

**READ YOUR WEB BOOKMARKS AND HISTORY**
Allows the app to read the history of all URLs that the Browser has visited, and all of the Browser’s bookmarks. Note: this permission may not be enforced by third-party browsers or other applications with web browsing capabilities.

**WRITE WEB BOOKMARKS AND HISTORY**
Allows the app to modify the Browser’s history or bookmarks stored on your tablet. This may allow the app to erase or modify Browser data. Note: this permission may be enforced by third-party browsers or other applications with web browsing capabilities. Allows the app to modify the Browser’s history or bookmarks stored on your phone. This may allow the app to erase or modify Browser data. Note: this permission may be enforced by third-party browsers or other applications with web browsing capabilities.

**PHONE CALLS**

**READ PHONE STATUS AND IDENTITY**
Allows the app to access the phone features of the device. This permission allows the app to determine the phone number and device IDs, whether a call is active, and the remote number connected by a call.
System Model: Example

Example 2: Unreasonable

If you are an American football fan you will undoubtedly be an ardent fan of the American football wallpapers. A great number of wallpapers are offered, just take your time and select your favorite one. Throw a pass on your screen but don’t forget the helmet!

- Easy-to-browse interface enables you to find the desired background in no time!
- Customize the wallpaper to fit your taste and screen!
- Zoom, un-zoom, rotate, move, scale, adjust brightness, add various effects and save your background design!

American Football NFL is a tough contact sport that dominates the American sporting scene and is gaining popularity in Europe. It is a game requiring all-round physical fitness. Players have to be fast, strong, and proficient in the six basic skills. Some of them are passing and catching. But running and blocking are not less important. American football has always been associated with a large number of injuries, not surprising when the game commonly invokes players in excess of 240 pounds colliding with each other at full speed. In the United States, it is estimated that more than 300,000 high school players, 36,000 college players, and half of the National Football League players are injured to some extent each season. Head and neck injuries are common: there are approximately 260,000 incidents of concussion each year. Some injuries have been fatal or have resulted in players becoming permanent quadriplegics. However, the number of very severe injuries has declined in recent years. This is probably due to rule changes concerning tackling and blocking with the head, and improvements in protective equipment which were introduced in 1976. There has also been a greater emphasis in training on strengthening the muscles of the neck and learning good tackling techniques. This is particularly important in the younger age group.
System Model: Approach

Description-based text-mining approach
Description-based text-mining approach

System Model: Approach

Web Extractor

Permission

Score Rule Finder

Keyword Vs. Permission

Score Calculator & Permission Judger

RESULT: Permission & Description Pair Reasonable? (N/Y)

Kea
Keywords Extractor

Description

Key Words

App: Permission & Description

URL

Training Set
System Model: Approach

Web Extractor (Python)
Web Extractor (Python)

GW Mail is an email client ... Please include information about your device and the version of GroupWise that you are using.

full network access
Agenda Meeting Minutes now on Android!!!!Record useful Meeting Minutes in your smartphone or tablet with your easy to use Agenda / Planning Software. Benefits: Record useful Meeting Minutes with our software:

- Use the Planning / Agenda / Calendar to plan your meeting.
- Prepare carefully: purpose of the meeting, attendees, topics, discussions.
- During the meeting, don’t miss any information and use your smartphone or tablet to record topics, conclusions, actions to take with responsible, deadline...
- After the meeting, review your notes and export a professional PDF version of your Meeting Minutes to print it or send it to all people involved in the meeting. - Don’t lose information about your meetings, the Planning / Agenda / Calendar will help you to see directly upcoming and recent meetings from the main view of the applications. - Search in the Planning / Agenda to find a meeting by date, purpose of topic. - Customize your Agenda with date format and time format. - Simplify the user interface with only needed information.

Main Functionalities:

- Home Screen: Agenda with direct access to upcoming and recent meetings. Search screen to find by date, purpose or topic. Agenda ordered by date, time, and showing Meeting purpose directly. Contact List.
- General information of the meeting: purpose, date, time, location and optional information: chair/facilitator, meeting call, note taker, timekeeper. Selection of the attendees from the contact database. Topics linked to meeting with title, description of the discussion, and optional information: conclusions, presenter. Actions linked to meeting with title, description, and optional information: deadline and person responsible. Export of each meeting in a professional Meeting Minute PDF document that can be used to print or send by mail to attendees.
Keywords and keyphrases (multi-word units) are widely used in large document collections. They describe the content of single documents and provide a kind of semantic metadata that is useful for a wide variety of purposes. The task of assigning keyphrases to a document is called keyphrase indexing. For example, academic papers are often accompanied by a set of keyphrases freely chosen by the author. In libraries, professional indexers select keyphrases from a controlled vocabulary (also called Subject Headings) according to defined cataloguing rules. On the Internet, digital libraries, or any depositories of data (flickr, del.icio.us, blog articles etc.) also use keyphrases (or here called content tags or content labels) to organize and provide a thematic access to their data.

KEA is an algorithm for extracting keyphrases from text documents. It can be either used for free indexing or for indexing with a controlled vocabulary.

KEA is implemented in Java and is platform independent. It is an open-source software distributed under the GNU General Public License.

In real life, the Kea is one of New Zealand’s native parrots, famed for theft, destroying cars and cameras, forming street gangs, pecking sheep to death for their delicious kidney fat, and other cutey antics.

Thanks to Gordon Boyter, who has created the original poster with the Kea.
Kea Keyword Extractor (Java)
System Model: Approach

Kea Keyword Extractor (Java)
System Model: Approach

Score-Rule Finder: The Idea

Statistics

- 1 110 contact
- 2 38 email
- 3 19 business
- 4 43 text
- 5 10 notes
- 6 10 call
- 7 10 calendar
- 8 10 location
- 9 7 download
- 10 3 tasks
- 11 6 office
- 12 6 exchange
- 13 6 caller
- 14 5 photos
- 15 5 outlook
- 16 5 sync
- 17 5 secure
- 18 4 card
- 19 4 events
- 20 4 usb
- 21 4 scan
- 22 4 invoice
- 23 4 gps
- 24 4 battery
- 25 4 box
- 26 4 sending
- 27 3 text messages

- 1 110 contact
- 2 38 email
- 3 19 business
- 4 43 text
- 5 10 notes
- 6 10 call
- 7 10 calendar
- 8 10 location
- 9 7 download
- 10 3 tasks
- 11 6 office
- 12 6 exchange
- 13 6 caller
- 14 5 photos
- 15 5 outlook
- 16 5 sync
- 17 5 secure
- 18 4 card
- 19 4 events
- 20 4 usb
- 21 4 scan
- 22 4 invoice
- 23 4 gps
- 24 4 battery
- 25 4 box
- 26 4 sending
- 27 3 text messages

- 1 110 contact
- 2 38 email
- 3 19 business
- 4 43 text
- 5 10 notes
- 6 10 call
- 7 10 calendar
- 8 10 location
- 9 7 download
- 10 3 tasks
- 11 6 office
- 12 6 exchange
- 13 6 caller
- 14 5 photos
- 15 5 outlook
- 16 5 sync
- 17 5 secure
- 18 4 card
- 19 4 events
- 20 4 usb
- 21 4 scan
- 22 4 invoice
- 23 4 gps
- 24 4 battery
- 25 4 box
- 26 4 sending
- 27 3 text messages

- 1 110 contact
- 2 38 email
- 3 19 business
- 4 43 text
- 5 10 notes
- 6 10 call
- 7 10 calendar
- 8 10 location
- 9 7 download
- 10 3 tasks
- 11 6 office
- 12 6 exchange
- 13 6 caller
- 14 5 photos
- 15 5 outlook
- 16 5 sync
- 17 5 secure
- 18 4 card
- 19 4 events
- 20 4 usb
- 21 4 scan
- 22 4 invoice
- 23 4 gps
- 24 4 battery
- 25 4 box
- 26 4 sending
- 27 3 text messages
### System Model: Approach

#### Score-Rule Finder: Keyword Score

<table>
<thead>
<tr>
<th>Rank</th>
<th>Keyword</th>
<th>Score (contact)</th>
<th>Score (location)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>contact</td>
<td>$\frac{38}{38 + 27 + 19 + \ldots + 1}$</td>
<td>$\frac{8}{38 + 27 + 19 + \ldots + 1}$</td>
</tr>
<tr>
<td>2</td>
<td>email</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>business</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>sms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>file</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>send</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>call</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>calendar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>location</td>
<td>$\frac{8}{38 + 27 + 19 + \ldots + 1}$</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>download</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>office</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>exchange</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>outlook</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>sync</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>secure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>card</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>usb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>scan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>invoice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>gps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>battery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>box</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>sending</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>text messages</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total words: 427
System Model: Approach

Final Score Calculator

**Final Score**

\[ \text{Final Score} = \text{Score}(\text{send}) + \text{Score}(\text{email}) + \cdots \]

\[ = \cdots \% \]
System Model: Approach

Score-Rule Finder: Distribution

- directly call phone numbers
  - w/ directly call phone numbers
  - w/o directly call phone numbers

- precise location (GPS and network-based)
  - w/ precise location (GPS and network-based)
  - w/o precise location (GPS and network-based)

- read your contacts
  - w/ read your contacts
  - w/o read your contacts

- send SMS messages
  - w/ send SMS messages
  - w/o send SMS messages
System Model: Approach

Score-Rule based Classification Model

Permissions

THIS APPLICATION HAS ACCESS TO THE FOLLOWING:

HARDWARE CONTROLS

TAKE PICTURES AND VIDEOS

Allows the app to take pictures and videos with the camera. This permission allows the app to use the camera at any time without your confirmation.

NETWORK COMMUNICATION

FULL NETWORK ACCESS

Allows the app to create network sockets and use custom network protocols. The browser and other applications provide means to send data to the internet, so this permission is not required to send data to the internet.

Description

Mobile Doc Scanner (MDScan) is a new business tool exclusively for your Android device!

Extremely useful mobile tool gives users the power to transform any file into PDF format and share it right from your device.

• Ranked in Top 5 Business application in 35 countries (March, 2013)
• Ranked in Top 10 Business application in 43 countries (March, 2013)

With MDScan users are free to scan any type of document, ranging from receipts to test pages and almost everything in between, on the go anywhere and anytime.

A perfect companion for busy users, MDScan allows anyone to transform everything from invoices to whiteboards and classes notes, magazine articles and more into a PDF format in order to easily export them, share and e-mail it to others, or for any other reason they can come up with. MDScan automatically detects borders, corrects distortion, and equalizes brightness to create clear, legible documents at a very high speed for your convenience.

Going above and beyond other mobile scanning solutions MDScan offers Batch shooting mode for a super fast scanning so users can easily scan within seconds multiple pages and save scanned files for batch processing later.
Outline

1 Motivation

2 Related Work

3 System Model

4 Evaluation

5 Conclusion
### Table: Prediction Summary

<table>
<thead>
<tr>
<th>Activity</th>
<th>Precision</th>
<th>False Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>directly call phone numbers</td>
<td>0.695652</td>
<td>0.304348</td>
</tr>
<tr>
<td>precise location</td>
<td>0.652174</td>
<td>0.347826</td>
</tr>
<tr>
<td>read your contacts</td>
<td>0.782609</td>
<td>0.173913</td>
</tr>
<tr>
<td>send SMS messages</td>
<td>0.739130</td>
<td>0.173913</td>
</tr>
</tbody>
</table>
Conclusion

We propose a machine learning based framework discovering permission-description patterns behind Android apps, and utilize this framework to detect the potential inconsistency between the description and the permission of a untrusted app.

Contribution
1. First to propose permission-abuse detection from text-mining approach
2. First to propose finding relationship link between description and permission
3. Rudimentary study shows a promising result (Our future work!)
We propose a machine learning based framework discovering permission-description patterns behind Android apps, and utilize this framework to detect the potential inconsistency between the description and the permission of a untrusted app.
Conclusion

We propose a machine learning based framework discovering permission-description patterns behind Android apps, and utilize this framework to detect the potential inconsistency between the description and the permission of a untrusted app.

Contribution

1. First to propose permission-abuse detection from text-mining approach
2. First to propose finding relationship link between description and permission
3. Rudimentary study shows a promising result (Our future work!)
Conclusion

We propose a machine learning based framework discovering permission-description patterns behind Android apps, and utilize this framework to detect the potential inconsistency between the *description* and the *permission* of a untrusted app.

Contribution

1. First to propose permission-abuse detection from *text-mining* approach
We propose a machine learning based framework discovering permission-description patterns behind Android apps, and utilize this framework to detect the potential inconsistency between the description and the permission of a untrusted app.

Contribution

1. First to propose permission-abuse detection from text-mining approach
2. First to propose finding relationship link between description and permission
We proposep a machine learning based framework discovering permission-description patterns behind Android apps, and utilize this framework to detect the potential inconsistency between the description and the permission of a untrusted app.

Contribution

1. First to propose permission-abuse detection from text-mining approach
2. First to propose finding relationship link between description and permission
3. Rudimentary study shows a promising result (Our future work!)
Conclusion: Future Work

1. Diversity
   - Larger data set within one category
   - More categories
   - Category-less

2. More precise key word extraction:
   - More, more comprehensive training set for Kea
   - More precise keyword extractor algorithm?
Conclusion: Future Work

Future Work

1. Diversity
   - Larger data set within one category
   - More categories
   - Category-less

2. More precise key word extraction:
   - More, more comprehensive training set for Kea
   - More precise keyword extractor algorithm?
## Future Work

### 1. Diversity
- Larger data set within one category
- More categories
- Category-less
Conclusion: Future Work

Future Work

1. Diversity
   - Larger data set within one category
   - More categories
   - Category-less

2. More precise key word extraction:
   - More, more comprehensive training set for Kea
   - More precise keyword extractor algorithm?
Thank You!
Questions please