WHYPER: Towards Automating Risk Assessment of Mobile Applications

Rahul Pandita, Xusheng Xiao, Wei Yang, William Enck, and Tao Xie
Department of Computer Science
North Carolina State University
\textsuperscript{\textdagger} University of Illinois at Urbana-Champaign
Application markets have played an important role in the popularity of smartphones and mobile devices.
Predominant approaches towards Market Security/Privacy

- **Apple** (*Market’s Responsibility*)
  - Apple performs manual inspection

- **Google** (*User’s Responsibility*)
  - Users approve permissions for security/privacy
  - Bouncer (static/dynamic malware analysis)

- **Windows Phone** (*Hybrid*)
  - Permissions / manual inspection
Previous approaches look at permissions, code, and runtime behaviors

Caveat: what does the users expect?

- **GPS Tracker**: record and send location
- **Phone-Call Recorder**: record audio during phone call
- **One-Click Root**: exploit a privilege escalation vulnerability
- Others are more subtle
Vision

“Bridging the gap between user expectation and app behaviors”

- A first step in this direction
- Focus on permission and application descriptions
  - permissions protecting user understandable resources should be discussed
  - low-level system permissions are unlikely to be mentioned
WHYPER Overview

Application

Market

WHYPER

DEVELOPERS

USERS
Use Cases

- Enhance user experience
  - while installing Apps
- Functionality disclosure
  - enforce on part of developers
- Complementing program analysis
  - to ensure more appropriate justifications
Solution
Simple Solution?

Keyword-based search on application descriptions

Photo Credit: Ahora estoy en via Flickr
Problems with Ctrl + F

- **Confounding effects:**
  - Certain keywords such as “contact” have a confounding meaning.
  - For instance, “… displays user contacts, …” vs “… contact me at abc@xyz.com”.

- **Semantic Inference:**
  - Sentences often describe a sensitive operation such as reading contacts without actually referring to keyword “contact”.
  - For instance, “share yoga exercises with your friends via email, sms”.
Natural Language Processing (NLP)

• NLP techniques help computers understand NL artifacts

• NLP is still difficult

• NLP on domain specific sentences with specific styles is feasible
Parts Of Speech (POS) Tagging
- E.g., noun, verb, prepositions...

Phrase and Clause Parsing
- E.g., noun phrases (basketball players) and verb phrases (make sure)...

Stanford-Typed Dependencies
- E.g., subject, object, adverbial modifiers...

Named Entity Recognition
- E.g., ‘Pandora Internet Radio’ is a name, ‘$5’ refers to a currency amount...
Preprocessor

○ Period Handling
  ○ Decimals, ellipsis, shorthand notations (Mr., Dr.)

○ Sentence Boundaries
  ○ Tabs, bullet points, delimiters (:) 
  ○ Symbols (*,-) and enumeration sentence

○ Named Entity Handling
  ○ E.g., “Pandora internet radio”,

○ Abbreviation Handling
  ○ E.g., “Instant Message (IM)”
Also you can share the yoga exercise to your friends via Email and SMS.
WHYPER Framework

APP Description

APP Permission

API Docs

WHYPER

Preprocessor

Intermediate Representation Generator

FOL Representation

Semantic Graphs

Semantic Graph Generator

Semantic Engine

Annotated Description
“Also you can share the yoga exercise to your friends via Email and SMS.”
Semantic-Graph Generator

- **WHY**
  - to perform deep semantic analysis

- **HOW**
  - infer graphs from API documents
Systematic approach to infer graphs

- Find related API documents based on PScoat [CCS 2012]
- Identify resource associated with the permissions from the API class name
- Inspect the member variables and methods to identify actions and subordinate resources

ContactsContract.Contacts

Class Overview

Contacts is the Contacts table, which contains a record per entry of raw contacts representing the same person. Operations

Insert

A Contact record can be inserted explicitly. If a raw contact is inserted, the provider will first find the contact representing the same person. If one is found, the raw contact's CONTACT_ID column gets the ID of the aggregate Contact. If no match is found, the provider automatically inserts a new Contact and puts its ID into the CONTACT_ID column of the newly inserted raw contact.

Update

To update the contact, the provider automatically inserts raw contact of the updated contact. The update columns of the Contact also changes them on all constituent raw contacts.

Delete

Capacity providers themselves do not delete raw contacts and all constituent raw contacts will get deleted. However, the context is given to the user to delete the raw contact with appropriate confirmation prompts from the delete function.

Query

To retrieve a single raw contact, consider using CONTENT_LOOKUP_URI instead of CONTENT_URI.

- If you need to look up a contact by the phone number, use PhoneLookup.CONTENT_FILTER_URI, which is optimized for this purpose.
- If you need to look up a contact by partial name, e.g. to produce filter-as-you-type suggestions, use the CONTENT_FILTER_URI.
Subjects

- Permissions:
  - READ_CONTACTS
  - READ_CALENDAR
  - RECORD_AUDIO
- 581/600* application descriptions (only English descriptions)
- 9,953 sentences

Research Questions (RQs)

- RQ1: What are the precision, recall and F-Score of WHYPER in identifying permission sentences?
- RQ2: How effective WHYPER is in identifying permission sentences, compared to keyword-based searching?
## Statistics of Subject Applications

<table>
<thead>
<tr>
<th>Permissions</th>
<th>#N</th>
<th>#S</th>
<th>$S_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>READ_CONTACTS</td>
<td>190</td>
<td>3,379</td>
<td>235</td>
</tr>
<tr>
<td>READCALENDAR</td>
<td>191</td>
<td>2,752</td>
<td>283</td>
</tr>
<tr>
<td>RECORD_AUDIO</td>
<td>200</td>
<td>3,822</td>
<td>245</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>581</td>
<td>9,953</td>
<td>763</td>
</tr>
</tbody>
</table>
Classification

- **True Positive (TP):**
  - $\text{WHYPER}(sentence) \&\& \text{Manual}(sentence)$

- **False Positive (FP):**
  - $\text{WHYPER}(sentence) \&\& \neg \text{Manual}(sentence)$

- **True Negative (TN):**
  - $\neg \text{WHYPER}(sentence) \&\& \neg \text{Manual}(sentence)$

- **False Negative (FN):**
  - $\neg \text{WHYPER}(sentence) \&\& \neg \text{Manual}(sentence)$
RQ1 Results: Effectiveness of WHYPER

- **Low FPs and FNs**
  - out of 9,061 sentences, only 129 are flagged as FPs
  - among 581 applications, 109 applications (18.8%) contain at least one FP
  - among 581 applications, 86 applications (14.8%) contain at least one FN

<table>
<thead>
<tr>
<th>Permission</th>
<th>S</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Prec.</th>
<th>Recall</th>
<th>F-Score</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>READ_CONTACTS</td>
<td>204</td>
<td>186</td>
<td>18</td>
<td>49</td>
<td>2,930</td>
<td>91.2</td>
<td>79.2</td>
<td>84.8</td>
<td>97.9</td>
</tr>
<tr>
<td>READCALENDAR</td>
<td>288</td>
<td>241</td>
<td>47</td>
<td>42</td>
<td>2,422</td>
<td>83.7</td>
<td>85.2</td>
<td>84.5</td>
<td>96.8</td>
</tr>
<tr>
<td>RECORD_AUDIO</td>
<td>259</td>
<td>195</td>
<td>64</td>
<td>50</td>
<td>3,470</td>
<td>75.3</td>
<td>79.6</td>
<td>77.4</td>
<td>97.0</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>751</td>
<td>622</td>
<td>129</td>
<td>141</td>
<td>9,061</td>
<td>82.8</td>
<td>81.5</td>
<td>82.2</td>
<td>97.3</td>
</tr>
</tbody>
</table>

24
RQ2 Results: Comparison to Keyword-based Search

<table>
<thead>
<tr>
<th>Permission</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>READ_CONTACTS</td>
<td>contact, data, number, name, email</td>
</tr>
<tr>
<td>READCALENDAR</td>
<td>calendar, event, date, month, day, year</td>
</tr>
<tr>
<td>RECORD_AUDIO</td>
<td>record, audio, voice, capture, microphone</td>
</tr>
</tbody>
</table>
## RQ2 Results: Comparison to Keyword-based Search

<table>
<thead>
<tr>
<th>Permission</th>
<th>Delta Precision</th>
<th>Delta Recall</th>
<th>Delta F-score</th>
<th>Delta Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>READ_CONTACTS</td>
<td>50.4</td>
<td>1.3</td>
<td>31.2</td>
<td>7.3</td>
</tr>
<tr>
<td>READ_CALENDAR</td>
<td>39.3</td>
<td>1.5</td>
<td>26.4</td>
<td>9.2</td>
</tr>
<tr>
<td>RECORD_AUDIO</td>
<td>36.9</td>
<td>-6.6</td>
<td>24.3</td>
<td>6.8</td>
</tr>
<tr>
<td>WHYPER Improvement</td>
<td>41.6</td>
<td>-1.2</td>
<td>27.2</td>
<td>7.7</td>
</tr>
</tbody>
</table>
• **Incorrect parsing**
  • “MyLink Advanced provides full synchronization of all Microsoft Outlook emails (inbox, sent, outbox and drafts), contacts, calendar, tasks and notes with all Android phones via USB”

• **Synonym analysis**
  • “You can now turn recordings into ringtones.”
Result Analysis (False Negatives)

- Incorrect parsing
  - Incorrect identification of sentence boundaries and limitations of underlying NLP infrastructure

- Limitations of Semantic Graphs
  - Manual Augmentation
    - microphone-*blow into* and call-*record*
    - *significant improvement of Delta Recalls:* -6.6% to 0.6%
  - Automatic mining from user comments and forums
Discussions

• Generalization to other permissions
  • user-understandable permissions: calls, SMS
  • location and phone identifiers
  • internet
Conclusion

We propose the use of NLP techniques to help bridge the semantic gap between what mobile applications do and what users expect them to do.

Our evaluation demonstrates an improvement over a simple keyword-based searching.
Thank You

Questions