Record and Replay for Android: Are We There Yet in Industrial Cases?

Wing Lam*, Zhengkai Wu*, Dengfeng Li*, Wenyu Wang*, Haibing Zheng*, Hui Luo*, Peng Yan*, Yuetang Deng* and Tao Xie*

*University of Illinois at Urbana-Champaign, USA. Email: {winglam2, zw3, dli46, wenyu2, taoxie}@illinois.edu
*Tencent, Inc., China. Email: {mattheng, huiluo, peteryan, yuetangdeng}@tencent.com

Motivation

- 9.2m Android app developers
- WeChat: 900m users worldwide
- Record and replay tools help developers conduct
  - Regression testing
  - Compatibility testing
  - Performance testing
- Existing tools: good enough in practice?

Qualitative Study

Highly Desirable Characteristics

- Coordinate sensitive
- Widget sensitive
- State sensitive
- Timing sensitive
- Whether the tool requires developers to install a custom operation system

Secondary Desirable Characteristics

- Events are recorded based on the coordinates of the events on the GUI screen
- Example Trace: TOUCH(\{'x':513,'y':1692,'type':'downAndUp',\})

Timing Sensitive

- Tool automatically sets the timing between recorded events
- Example Trace: WAIT(\{'seconds':2.0,\})

State Sensitive

- Tool can be replayed successfully only if the device is currently in the state that it was in when the trace was recorded
- Tool can be replayed even if the phone is in any state

Selected Tools and Their Characteristics

<table>
<thead>
<tr>
<th>Name</th>
<th>Open source</th>
<th>Coordinate</th>
<th>Widget</th>
<th>State</th>
<th>Timing</th>
<th>Instruments app</th>
<th>Custom OS</th>
<th>Root access</th>
<th>Needs source</th>
</tr>
</thead>
<tbody>
<tr>
<td>appetizer</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>monkeyrunner</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>RERAN</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Coordinate Sensitive

- Events are recorded based on the coordinates of the events on the GUI screen
- Example Trace: TOUCH(\{'x':513,'y':1692,'type':'downAndUp',\})

Widget Sensitive

- Events are recorded based on the widgets (e.g., buttons, text fields) where the events happen
- Example Trace: TOUCH(\{'id':'R.id.loginButton','type':'downAndUp',\})

Quantitative Study

Selected Apps and Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Category</th>
<th>Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>112.0.17.71</td>
<td>Social</td>
<td>72.2</td>
</tr>
<tr>
<td>File Explorer</td>
<td>5.1.3.0</td>
<td>Business</td>
<td>9.2</td>
</tr>
<tr>
<td>WeChat</td>
<td>6.5.7</td>
<td>Communication</td>
<td>40.5</td>
</tr>
</tbody>
</table>

Scenarios

Facebook
- Create post ✔
- Watch video ✔
- Share photo ✔
- Play games ✔

File Explorer
- Tutorial ✔
- Explore menu ✔
- Browse directories ✔
- New windows ✔

WeChat
- Group chat/Link ✔
- Group chat/Photo ✔
- Gallery/Photo ✔
- Moments/Photo&Link ✔
- Moments/Post ✔

Space Overhead

Traces produced by RERAN occupy about 67 times the space that traces from monkeyrunner would occupy

<table>
<thead>
<tr>
<th>Name</th>
<th>appetizer</th>
<th>monkeyrunner</th>
<th>RERAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>10</td>
<td>3</td>
<td>215</td>
</tr>
<tr>
<td>File Explorer</td>
<td>4</td>
<td>3</td>
<td>94</td>
</tr>
<tr>
<td>WeChat</td>
<td>24</td>
<td>8</td>
<td>625</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>14</td>
<td>954</td>
</tr>
</tbody>
</table>

Replay Robustness

Recorded traces from common usage scenarios replayed on different-resolution devices
- For WeChat and Facebook scenarios, none of the 3 tools were successful in replaying them
- None of these tools are able to replay complicated scenarios on different screen sizes successfully

Conclusion

Qualitative Summary

- There exist many record-and-replay tools, but few of them are available for use in practice
- Our interactions with developers derive desirable characteristics of a record-and-replay tool
- Unfortunately, none of the tools that we find exhibit all of the desirable characteristics

Quantitative Summary

- We record and replay common usage scenarios on 3 popular apps
- RERAN can successfully replay all scenarios at least some of the time
- RERAN is inferior in its use of space to store recorded traces
- All 3 tools cannot replay the scenarios on different screen sizes

The work of UIUC authors is supported in part by NSF under grants no. CCF-1409423, CNS-1434582, CNS-1513939, CNS-1564274.