Crowdsourced Testing

Research and Practice gap

Zhenyu Chen
Software Institute, Nanjing University
http://software.nju.edu.cn/zychen

Dec. 20, 2015, Peking University
Internet+Testing

Testing as a Service

PTS of Alibaba
pts.aliyun.com

MTC of Baidu
mtc.baidu.com
Internet+Testing

Power of Devices

Power of People
Crowdsourcing in SE

Open call | Undefined workers | Often large scale | Often Internet
Utest : $43 M Investment (2014.1)

Crowdsourced Testing of Mobile Apps
Testin: $54.90 M Investment (2015.7)

Crowdsourced Testing of Mobile Apps
Why Mobile App Testing?

- Android fragmentation
- Diversified devices
- Diversified scenarios
# Crowdsourced Testing Platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>Website</th>
<th>Service Types</th>
<th>On-demand Matching, Online Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>uTest</td>
<td><a href="http://www.ustest.com">www.ustest.com</a></td>
<td>Software Testing</td>
<td>On-demand Matching</td>
</tr>
<tr>
<td>Passbrains</td>
<td><a href="http://www.passbrains.com">www.passbrains.com</a></td>
<td>Software Testing</td>
<td>On-demand Matching</td>
</tr>
<tr>
<td>TestBirds</td>
<td><a href="http://www.testbirds.com">www.testbirds.com</a></td>
<td>Software Testing</td>
<td>On-demand Matching</td>
</tr>
<tr>
<td>Testbats</td>
<td><a href="http://www.testbats.com">www.testbats.com</a></td>
<td>Software Testing</td>
<td>On-demand Matching</td>
</tr>
<tr>
<td>Pay4Bugs</td>
<td><a href="http://www.pay4bugs.com">www.pay4bugs.com</a></td>
<td>Software Testing</td>
<td>On-demand Matching</td>
</tr>
<tr>
<td>TestFlight</td>
<td><a href="http://www.testflightapp.com">www.testflightapp.com</a></td>
<td>Mobile App Testing</td>
<td>On-demand Matching</td>
</tr>
</tbody>
</table>

Crowdsourced Testing (2013.6)

- Quasi-Crowdsourcing Testing for Educational Projects, ICSE 2014
- Test Report Prioritization to Assist Crowdsourced Testing. ESEC/FSE 2015

Driven by Industry

- Baidu-Input-Android
- Baidu-Browser
- Baidu-Media-Player
Results of Case Study

Table 1: Summary of Industry Feedback

<table>
<thead>
<tr>
<th>Project</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td># Tester</td>
<td>20</td>
<td>55</td>
<td>54</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td># F-test</td>
<td>274</td>
<td>231</td>
<td>252</td>
<td>316</td>
<td>574</td>
</tr>
<tr>
<td># V-test</td>
<td>186</td>
<td>47</td>
<td>62</td>
<td>113</td>
<td>114</td>
</tr>
<tr>
<td>% V-test</td>
<td>67.88%</td>
<td>20.35%</td>
<td>24.60%</td>
<td>35.76%</td>
<td>19.86%</td>
</tr>
</tbody>
</table>

- F-test: Failed test reports submitted by crowdsourcing testers
- V-test: Test reports validated by industrial actors

- Quasi-Crowdsourcing Testing for Educational Projects, ICSE 2014
### An Example of Test Report (Chinese)

<table>
<thead>
<tr>
<th>E: Environment</th>
<th>I: Input</th>
<th>O: Output</th>
<th>D: Description</th>
</tr>
</thead>
</table>
| 前置条件: 手机开机
允许横屏模式
测试机型: 小米2A、魅族MX2
测试环境: 手机QQ2012 | 1. 进入设置页面
2. 关闭浮动显示主题特效
3. 开启横屏时启用全屏模式
4. 进入超级皮肤页面
5. 选择有凸出图片的主题
6. 返回后进入横屏模式 | 使用默认主题列表中第一列第一个主题，关闭浮动显示主题特效和打开横屏全屏模式，横屏时布局出错，主题图片会被遮挡。 |
| Prerequisite: Turn on the landscape mode
Mobile Type: MIUI 2A, SUMSUNG MX2
Testing Environment: Mobile QQ2012 | 1. Go to settings page
2. Close the floating effect and turn on the topic effects
3. Turn on the landscape mode and enable the full-screen mode
4. Go to the super skin setting page
5. Select any theme with a floating picture
6. Exit setting page and enter the landscape mode | Use the first column the first topic of default topic list, close floating display effect and turn on landscape screen and full-screen mode. A landscape screen layout error occur, the theme pictures will be blocked. |

- Test Report Prioritization to Assist Crowdsourced Testing. ESEC/FSE 2015
Crowdsourcing to mooctest

Industry

kikbug.net

Driven by Industry

Agent

Teacher

Student

mooctest.net

An online exam system
mooctest.net

- Automatic Scoring
- Programming
- Testing
- Debugging
• Do/How spectrum-based fault localization techniques work?
• 2 buggy programs for everyone (with/without tool)
• Capture all keyboard and cursor behaviors

• Revisit of Automatic Debugging via Human Focus-tracking Analysis, ICSE 2016
Organization and Excitation

Branch Coverage : 25%
Debugging : 15%
Performance : 15%
Mobile App : 25%
Final Exam : 10%
Class : 10%

Programming
Industry
Wechat
Weixin

6000+
Kikbug Web+Android App

- Texts
- Snapshots
- Events
- Logcats

2015/12/21
Two datasets from kikbug.net
Test Reports (Chinese) + Test Scripts (2015.6)

- 知识（知乎）V2.0.2
  - 219 TRs
    - 15 F-TRs

- 网易云音乐 V2.5.1
  - 227 TRs
    - 26 F-TRs

- 有道词典 V5.3.3
  - 296 TRs
    - 53 F-TRs

- QQ邮箱 V4.0.1
  - 163 TRs
    - 17 F-TRs

- UC浏览器 V8.8.3
  - 220 TRs
    - 22 F-TRs

- 美团外卖 V3.0.3
  - 258 TRs
    - 40 F-TRs

- 金山词霸 V7.2.4
  - 216 TRs
    - 19 F-TRs

- 飞信 V5.4.1
  - 288 TRs
    - 47 F-TRs

- 哔哩哔哩 V3.0.0
  - 173 TRs
    - 19 F-TRs
23 Brands

- **Xiaomi**, 69, 25%
- **Huawei**, 68, 25%
- **Samsung**, 44, 16%
- **Meizu**, 22, 8%
- **Coolpad**, 12, 4%
- **OPPO**, 11, 4%
- **HTC**, 8, 3%
- **Google**, 5, 2%
- **Lenovo**, 5, 2%
- **ZTE**, 7, 2%
- **Others**, 26, 9%
12 Android Versions

- 4.4.4: 4.4.4, 87, 32%
- 4.4.2: 4.4.2, 68, 25%
- 4.1.2: 4.1.2, 26, 10%
- 4.3.0: 4.3.0, 24, 9%
- 4.1.1: 4.1.1, 21, 8%
- 4.2.2: 4.2.2, 17, 6%
- 4.0.4: 4.0.4, 6, 2%
- 4.3.1: 4.3.1, 4, 1%
- 4.0.0: 4.0.0, 1, 0%
- 2.3.6: 2.3.6, 1, 0%
Test Reports + Test Logs (2015.12)

- App
  - 爱逛街 – Alibaba
  - 网易云音乐 – NetEase
  - 优播客 - Alibaba+新东方
  - 图个乐-Alibaba
  - SE-1800- Panneng

- Test Report+Test Logs
  - 爱逛街: 431TRs (208 Fail-TRs)
  - 网易云音乐: 234TRs (80Fail-TRs)
  - 优播客: 434TRs (189Fail-TRs)
  - 图个乐: 293TRs (215Fail-TRs)
  - SE-1800(电力:) 348TRs (175Fail-TRs)
181 Mobile Phones

- MI 4LTE: 7.0%
- M1 Note: 5.5%
- MI 2S: 3.6%
- M2 Mote: 2.5%
- PLK-TL01H: 2.4%
- MI 3W: 2.3%
- Vivo X5M: 2.0%
- Others: 73.8%

Gionee Elife E7: 1.8%
21 Android Versions

- 5.1: 12.6%
- 4.1: 8.8%
- 4.2: 9.1%
- 4.3: 8.1%
- 5.0: 19.7%
- Others: 1.1%
- 4.4%
- 41.3%
27 Resolutions

- Others: 17.4%
- 4.6: 60.9%
- 4.3: 7.9%
- 4.4: 5.9%
- 5.0: 5.2%
- 3.9: 2.7%
206 Combinations (282 Testers)

- 4.4.4 MI 4LTE: 7.0%
- 5.1.0 M1 note: 3.3%
- 5.0.2 MI 2S: 3.0%
- 5.1.0 M2 note: 2.5%
- 5.0.2 PLK-TL01H: 2.4%
- 4.4.4 MI 3W: 2.3%
- 4.4.4 M1 note: 2.1%
- 5.0.2 vivo X5M: 2.0%
- Others: 75.3%
Number of test reports from one tester
阿里云测
找BUG大赛

送！最新款YunOS智能手机

一等奖3名  YUNOS最佳应用一部
二等奖30名  支持移动设备一个
(优秀同学可获得阿里云测试团队定制T恤，实习生优先录取机会)

比赛内容：
1. 从Kikbug前端或后端提交“聚众”应用进行手工测试
2. 使用MQC（mqc.aliyun.com）进行自动化功能测试
3. 登录kikbug.net进行提交和反馈测试报告（用户体验、功能、性能、安全等Bug）

比赛时间：2015.12.27 14.00-18.00 比赛地点：南京大学信息学院701

参赛方式：
扫描右下角二维码或访问 http://www.kikbug.net/

参与条件：有一定前端后端能力，对APP开发与测试充满兴趣，任何高校师生均可参与。

主办方：阿里巴巴YUNOS, kikbug
协办方：阿里巴巴云测中心

www.kikbug.net
• Texts
• Snapshots
• Events
• Logcats
• Scripts
mqc.aliyun.com
Hybrid Crowdsourced Testing

- Automated Testing
  - Installation
  - Random
  - Compatibility
- 100+ mobile devices

- Crowdsourced Testing
  - Functional
  - Usability
  - Compatibility
- 6000+ testers
Beyond Research