AppInspect: Large-scale Evaluation of Social Networking Apps

ACM COSN, Boston, 10/08/2013

Markus Huber, Martin Mulazzani, Sebastian Schrittwieser, Edgar Weippl

mhuber[AT]sba-research[DOT]org
Main Contributions

- **AppInspect**: privacy and security analysis of OSN apps
- Prototype for Facebook’s application ecosystem
- Detected information leaks, shortcomings in popular apps
- Cooperated with Facebook to fix apps and protect users
- AppInspect datasets available to the research community
Section 2

Background
OSN apps

- Apps used by hundreds of millions of social networking users
- Games, horoscopes, quizzes, etc.
- Access sensitive personal information (date of birth, email address, personal messages etc.)
- Access to information of application user’s friends
Modus operandi of OSN apps

• OSNs act as proxies between user and app developer
• Personal information is transferred to developers
• App developers themselves rely on third-parties (analytics, advertising products)
• Custom hosting infrastructures
• Approval of apps with authentication dialog
Facebook’s application authorization dialog

(a) Unified Auth Dialog, April 2010

(b) Enhanced Auth Dialog, January 2012

(c) App Center Auth Dialog, May 2012
Section 3

AppInspect Framework
AppInspect Framework

Figure: AppInspect, a framework for automated security and privacy analysis of social network ecosystems.
(1) Search Module

• Enumerate applications for target social network
• Simple scrapers
  ▶ Google+, single HTML page with few applications
  ▶ LinkedIn, easy to enumerate via applicationId
• Facebook
  ▶ Majority of apps not in directories
  ▶ Numeric identifier brute force not feasible \((10^{14})\)
  ▶ Exhaustive search: character n-grams, keywords, etc.

LinkedIn Example

GET /opensocialInstallation/preview?_applicationId=1000
Host: https://www.linkedin.com
(2) Classifier Module

- Application properties: rating, popularity, permissions, type
  - Web scraping
  - Redirection behavior
- Language
  - Detect and translate non-english applications

Redirect example

```
GET /apps/application.php?id=194699337231859
Host: www.facebook.com
Redirects to http://yahoo.com
```
(3) Analysis Module

• Traffic collection
  ▶ Applications are installed on test accounts
  ▶ HTTP(S) proxy collects network traffic

• Web tracker identification
  ▶ Detection of analytics and advertising products

• Information leaks
  ▶ Leakage of personal data, auth tokens to third parties

• Hosting infrastructure fingerprint
  ▶ Fingerprint the underlying hosting infrastructure
  ▶ Search vulnerability databases for detected services
Section 4

Evaluation
Prototype

- Analysis of Facebook’s application ecosystem
- Non-intrusive security audits

- AppInspect Prototype
  - Python with mechanize, Mozilla Firefox + Adobe Flash
  - Fast crawling, and realistic network samples

- Traffic Analysis
  - HTTP(S) interception proxy
  - XML parser for network samples

- Web tracker identification
  - Based on Ghostery DB

- Hosting infrastructure fingerprint
  - Standard unix tools (dig, nmap)
  - Exploit-DB, metasploit-DB
Enumerated Apps

- Exhaustive search with character trigrams
- 434,687 unique applications in two weeks
- Validation against Socialbakers’ Facebook applications
Application Sample

- 10,624 most popular apps ≈ 94.07% of cumulative usage
- In-depth analysis on 4,747 apps which transfer user data

<table>
<thead>
<tr>
<th>Application Type</th>
<th>Applications</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentication Dialog</td>
<td>4,747</td>
<td>44.68%</td>
</tr>
<tr>
<td>Canvas</td>
<td>2,365</td>
<td>22.26%</td>
</tr>
<tr>
<td>Connect</td>
<td>2,260</td>
<td>21.27%</td>
</tr>
<tr>
<td>Defect</td>
<td>865</td>
<td>8.14%</td>
</tr>
<tr>
<td>Page Add-ons</td>
<td>280</td>
<td>2.64%</td>
</tr>
<tr>
<td>Mobile</td>
<td>107</td>
<td>1.01%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10,624</strong></td>
<td><strong>100.00%</strong></td>
</tr>
</tbody>
</table>

Table: Classification of subsample with popular applications
Section 5

Results
### Requested Permissions (n=4,747)

<table>
<thead>
<tr>
<th>Permission</th>
<th>App Category</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>game</td>
<td>app</td>
</tr>
<tr>
<td>Publish posts to stream</td>
<td>1,617</td>
<td>819</td>
</tr>
<tr>
<td>Personal email address</td>
<td>1,055</td>
<td>1,132</td>
</tr>
<tr>
<td>Publish action</td>
<td>435</td>
<td>857</td>
</tr>
<tr>
<td>Access user’s birthday</td>
<td>582</td>
<td>428</td>
</tr>
<tr>
<td>Access user’s photos</td>
<td>721</td>
<td>99</td>
</tr>
<tr>
<td>Access data offline</td>
<td>517</td>
<td>120</td>
</tr>
<tr>
<td>Access user likes</td>
<td>438</td>
<td>153</td>
</tr>
<tr>
<td>Access user location</td>
<td>350</td>
<td>143</td>
</tr>
<tr>
<td>Read stream</td>
<td>409</td>
<td>80</td>
</tr>
<tr>
<td>Access friends’ photos</td>
<td>319</td>
<td>17</td>
</tr>
</tbody>
</table>

Table: Most common requested permissions by third-party applications
Permissions per Provider

- 4,747 applications belonged to 1,646 distinct providers
- 60.24% of all providers requested personal email address
Developers with $\geq 10$ Permission Requests

- 40 providers requested more than 10 permissions
- Manually verified requested permissions vs. app functionality
- Legitimate uses
  - Dating and job hunting applications
  - XBOX application (not available anymore)
- Excessive permission requests
  - Horóscopo Diário, 2.5 million monthly users
  - Would require data of birth, 25 different permissions
  - Request permission but do not use them
  - Users do not seem to verify requested permissions
Internet Hosting Services

- 55% of applications hosted in the US
- 64 different countries in total

<table>
<thead>
<tr>
<th>Provider</th>
<th>Location</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon EC2</td>
<td>US (755), IE (82), SG (52)</td>
<td>18.72%</td>
</tr>
<tr>
<td>SoftLayer</td>
<td>US (505)</td>
<td>10.65%</td>
</tr>
<tr>
<td>Peak Hosting</td>
<td>US (244)</td>
<td>5.14%</td>
</tr>
<tr>
<td>Rackspace</td>
<td>US (147), GB (11), HK (4)</td>
<td>3.41%</td>
</tr>
<tr>
<td>GoDaddy</td>
<td>SG (51), US (29), NL (6)</td>
<td>1.82%</td>
</tr>
<tr>
<td>Linode</td>
<td>US (72), GB (6), JP (2)</td>
<td>1.69%</td>
</tr>
<tr>
<td>OVH</td>
<td>FR (42), PL (7), ES (2)</td>
<td>1.04%</td>
</tr>
<tr>
<td>Hetzner</td>
<td>DE (47)</td>
<td>0.99%</td>
</tr>
<tr>
<td>Internap</td>
<td>US (35)</td>
<td>0.73%</td>
</tr>
</tbody>
</table>
Discovered Web Services

- 55% Apache httpd, nginx (15.63%), Microsoft IIS (9.4%)
- 2 hosts source code disclosure vulnerability (CVE-2010-2263)
- 8 hosts ProFTPD buffer overflow (CVE-2006-5815, CVE-2010-4221)
- Host with 1.2 million monthly users and sensitive information

<table>
<thead>
<tr>
<th>TCP Port</th>
<th>Service</th>
<th>Hosts</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>ssh</td>
<td>662</td>
<td>40.22%</td>
</tr>
<tr>
<td>21</td>
<td>ftp</td>
<td>640</td>
<td>38.88%</td>
</tr>
<tr>
<td>25</td>
<td>smtp</td>
<td>572</td>
<td>34.75%</td>
</tr>
<tr>
<td>110</td>
<td>pop3</td>
<td>439</td>
<td>26.67%</td>
</tr>
<tr>
<td>143</td>
<td>imap</td>
<td>417</td>
<td>25.33%</td>
</tr>
</tbody>
</table>

Table: Most common additional services on application hosts
## Tracking and Advertisement Products

<table>
<thead>
<tr>
<th>Web bug</th>
<th>Type</th>
<th>Apps</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Analytics</td>
<td>analytics</td>
<td>3,378</td>
<td>71.16%</td>
</tr>
<tr>
<td>DoubleClick</td>
<td>advertising</td>
<td>529</td>
<td>11.14%</td>
</tr>
<tr>
<td>Google Adsense</td>
<td>advertising</td>
<td>361</td>
<td>7.61%</td>
</tr>
<tr>
<td>AdMeld</td>
<td>advertising</td>
<td>276</td>
<td>5.81%</td>
</tr>
<tr>
<td>Cubics</td>
<td>advertising</td>
<td>153</td>
<td>3.22%</td>
</tr>
<tr>
<td>LifeStreet Media</td>
<td>advertising</td>
<td>94</td>
<td>1.98%</td>
</tr>
<tr>
<td>Google AdWords</td>
<td>advertising</td>
<td>91</td>
<td>1.92%</td>
</tr>
<tr>
<td>OpenX</td>
<td>advertising</td>
<td>82</td>
<td>1.73%</td>
</tr>
<tr>
<td>Quantcast</td>
<td>analytics</td>
<td>49</td>
<td>1.03%</td>
</tr>
<tr>
<td>ScoreCard Beacon</td>
<td>analytics</td>
<td>48</td>
<td>1.01%</td>
</tr>
</tbody>
</table>

Table: Common web trackers included in third-party applications
Information Leaks

• 315 apps directly transferred personally identifiable information (via HTTP parameter)

uuid, birthdate, gender

GET socialanalytic-web-rest/rest/action/16/100000000000000000/wpc/landingbirthday=5%2F2%2F2013&gender=male
Host: removed from online version

uuid, tracking!

GET /delivery/brandConnect.php?callback=siteUserId=1000000000000&siteId=1111&popup=0
Host: removed from online version
Information leaks II

- 51 applications leaked unique user identifiers (HTTP Referer)
- 14 out of 51 applications also leaked OAuth tokens

Example leak, app with 4.7 million MAU

```plaintext
GET /fnf/flash.php?hhref=&u=&page=-1&frli=&oauth_token=AAAAAAAAAAAAAAAAAAAA&fbid=100000000000000&issec=0&locale=en_US:
Host: removed from online version
```
Section 6

Discussion and Conclusion
Discussion

- Reported our findings to Facebook in November 2012
  - Facebook responded quickly
  - Facebook acknowledged problems and contacted developers
  - Application issues fixed in May 2013
- Security and privacy implications
  - Since January 2010 unproxied access to email address
  - 60% of application providers request email address
  - Social phishing, context-aware spam
  - Users trackable with real name
- Hosting
  - Number of hosts possible vulnerable
  - FTP/SSH bruteforce
Limitations

• Limitation to Facebook canvas applications
  ▶ AppInspect adaption to other OSNs
  ▶ Mobile applications and websites

• Detection of excessive permission requests
  ▶ App functionality vs. requested permissions
  ▶ Requires manual reviews

• Detection of information leaks
  ▶ Obfuscated personal information
  ▶ Hidden back-ends for data transfer
  ▶ Offline passing on of data
Conclusion

- Automated social app analysis is feasible
- Helped to fix shortcomings in popular applications
- Framework and dataset
  - Plan: Release opensource version of code
  - Datasets for social app research

http://ai.sba-research.org