secure
sba-research.org

COMET

Competence Centers for
Excellent Technologies

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

secure
sba-research.org

FFG

TU WIEN
Technische Universität Wien
Vienna University of Technology

Main Contributions

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

Section 2

Background

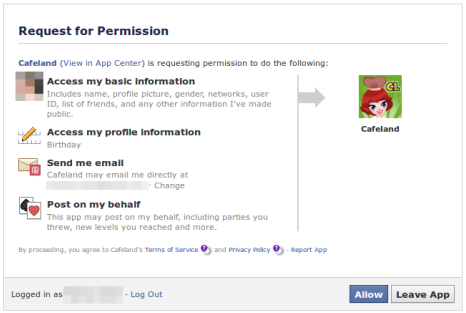
OSN apps

- 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

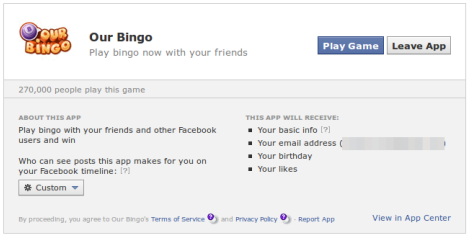
Modus operandi of OSN apps

- 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

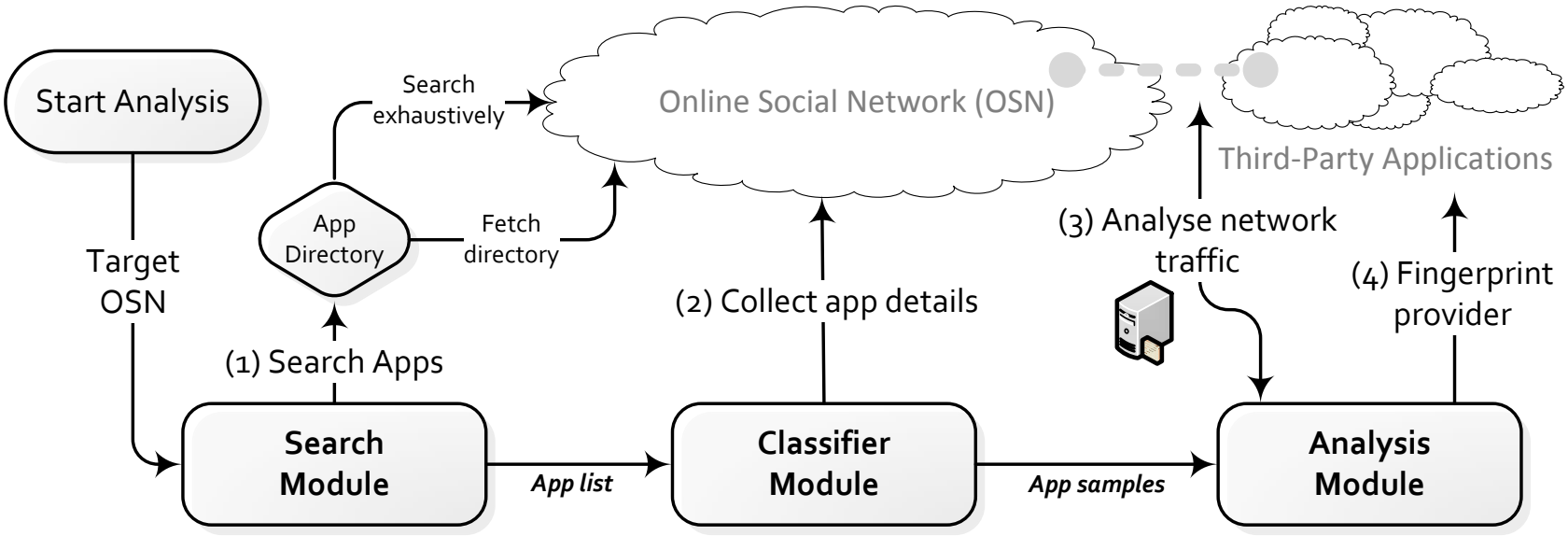


Figure: AppInspect, a framework for automated security and privacy analysis of social network ecosystems.

(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

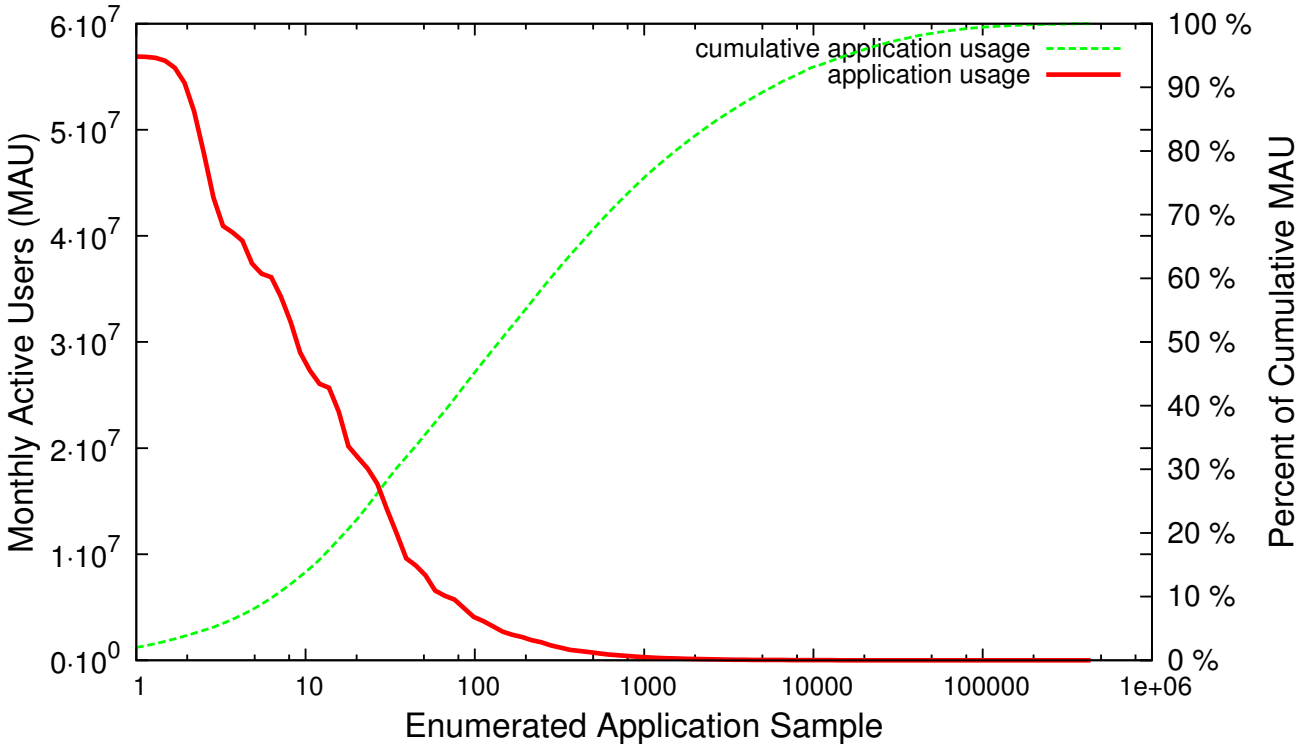
Prototype

- 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 $\simeq$ 94.07% of cumulative usage
- In-depth analysis on 4,747 apps which transfer user data

Application Type	**Applications**	**Total %**
Authentication Dialog	*4,747*	*44.68%*
Canvas	2,365	22.26%
Connect	2,260	21.27%
Defect	865	8.14%
Page Add-ons	280	2.64%
Mobile	107	1.01%
Total	*10,624*	*100.00%*

Table: Classification of subsample with popular applications

Section 5

Results

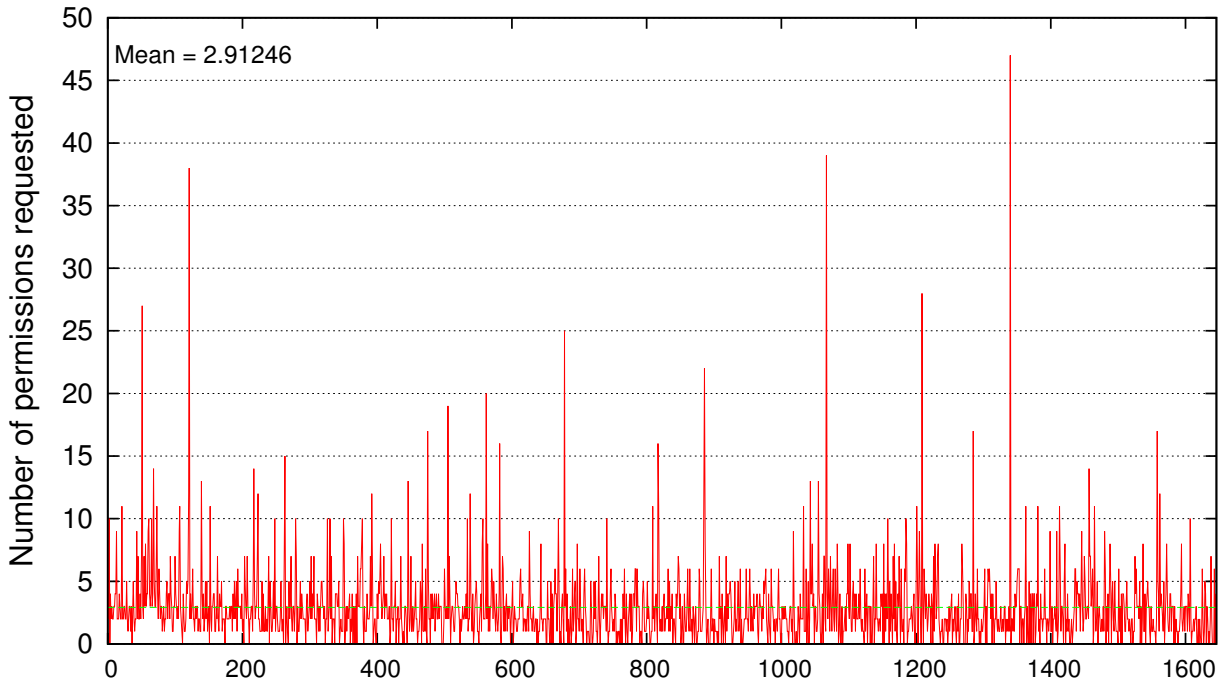
Requested Permissions (n=4,747)

	App Category		
Permission	game	app	**Total %**
Publish posts to stream	*1,617*	819	51.32%
Personal email address	1,055	*1,132*	46.07%
Publish action	435	857	27.22%
Access user's birthday	582	428	21.28%
Access user's photos	721	99	17.27%
Access data offline	517	120	13.42%
Access user likes	438	153	12.45%
Access user location	350	143	10.39%
Read stream	409	80	10.3%
Access friends' photos	319	17	7.08%

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

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

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

TCP Port	Service	Hosts	% Total
22	ssh	662	40.22%
21	ftp	640	38.88%
25	smtp	572	34.75%
110	pop3	439	26.67%
143	imap	417	25.33%

Table: Most common additional services on application hosts

Tracking and Advertisement Products

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

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
    /16/1000000000000/wpc/landingbirthday=5%2F2%2
    F2013&gender=male
Host: removed from online version
```

uuid, tracking!

```
GET /delivery/brandConnect.php?callback=siteUserId
    =100000000000&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

```
GET /fnf/flash.php?hbref=&u=&page=-1&frli=&
    oauth_token=AAAAAAAAAAAAAAAAA&fbid
    =1000000000000&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`