Security & Privacy in Online Social Networks

Konstantin (Kosta) Beznosov
lerssse.ece.ubc.ca
outline

• why OSNs?
• rewards and challenges of research in OSN
• current research directions
  – de-anonymization
  – privacy (game)
  – Sybil & compromised account detection/resistance
WHY ONLINE SOCIAL NETWORKS?
why OSNs?: multitude

<table>
<thead>
<tr>
<th>Site</th>
<th>Traffic Rank</th>
<th>Users (M)</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Live</td>
<td>4</td>
<td>120</td>
<td>USA</td>
</tr>
<tr>
<td>Facebook</td>
<td>4</td>
<td>175</td>
<td>USA</td>
</tr>
<tr>
<td>MySpace</td>
<td>7</td>
<td>250</td>
<td>USA</td>
</tr>
<tr>
<td>Hi5</td>
<td>17</td>
<td>60</td>
<td>USA</td>
</tr>
<tr>
<td>SkyRock</td>
<td>43</td>
<td>13</td>
<td>France</td>
</tr>
<tr>
<td>Friendster</td>
<td>45</td>
<td>95</td>
<td>USA</td>
</tr>
<tr>
<td>NetLog</td>
<td>71</td>
<td>35</td>
<td>Belgium</td>
</tr>
<tr>
<td>Tagged</td>
<td>75</td>
<td>70</td>
<td>USA</td>
</tr>
<tr>
<td>Orkut</td>
<td>83</td>
<td>67</td>
<td>USA</td>
</tr>
<tr>
<td>LiveJournal</td>
<td>85</td>
<td>18</td>
<td>Russia</td>
</tr>
<tr>
<td>Bebo</td>
<td>119</td>
<td>40</td>
<td>USA</td>
</tr>
<tr>
<td>PerfSpot</td>
<td>124</td>
<td>20</td>
<td>USA</td>
</tr>
<tr>
<td>meinVZ</td>
<td>156</td>
<td>12</td>
<td>Germany</td>
</tr>
<tr>
<td>Multiply</td>
<td>161</td>
<td>12</td>
<td>USA</td>
</tr>
<tr>
<td>Badoo</td>
<td>168</td>
<td>19</td>
<td>UK</td>
</tr>
<tr>
<td>Sonico</td>
<td>183</td>
<td>33</td>
<td>Argentina</td>
</tr>
<tr>
<td>Ning</td>
<td>187</td>
<td>1</td>
<td>USA</td>
</tr>
<tr>
<td>CyWorld</td>
<td>315</td>
<td>20</td>
<td>South Korea</td>
</tr>
<tr>
<td>Xanga</td>
<td>346</td>
<td>40</td>
<td>USA</td>
</tr>
<tr>
<td>MyYearbook</td>
<td>406</td>
<td>15</td>
<td>USA</td>
</tr>
</tbody>
</table>

why OSNs?: sheer scale

Monthly Active Users (MAUs)

Millions of MAUs

- Rest of World
- Asia
- Europe
- US & Canada

Q4'10 | Q1'11 | Q2'11 | Q3'11 | Q4'11 | Q1'12 | Q2'12 | Q3'12 | Q4'12
---|---|---|---|---|---|---|---|---
608 | 133 | 183 | 207 | 225 | 245 | 268 | 1,007 | 1,056
680 | 156 | 174 | 196 | 212 | 234 | 255 | 288 | 304
739 | 138 | 201 | 212 | 229 | 239 | 246 | 253 | 298
800 | 183 | 212 | 221 | 179 | 183 | 186 | 189 | 261
901 | 154 | 169 | 176 | 179 | 183 | 193 | | 

Please see Facebook’s Form 10-K for the year ended December 31, 2012 for definitions of user activity used to determine the number of our MAUs, DAUs and mobile MAUs. The number of MAUs, DAUs, and mobile MAUs do not include Instagram users unless such users would otherwise qualify as MAUs, DAUs, and mobile MAUs based on activity that is shared back to Facebook.

In June 2012, we discovered an error in the algorithm we used to estimate the geographic location of our users that affected our attribution of certain user locations for the first quarter of 2012. The first quarter of 2012 user metrics reflect a recategorization to more correctly attribute users by geographic region.

why OSNs?: this is where users are!

• 20% of US page-views are on Facebook [1]

• Each Facebook user spends on average 15 hours and 33 minutes a month on the site [2]

• Twitter is handling 1.6B queries per day [2]

nearly 2/3 of OSN users use them daily

“The Social Habit” Edison Research and Arbitron, 2012-June
In the last 25 hours, approximately how many times did you check your Facebook account?

Mean = 4 times

“The Social Habit” Edison Research and Arbitron, 2012-June
what do users do there?

- social connection
- shared identities
- Photographs
- content
- social investments
- social networks
- status updates

What's your Status Update?

Turn Off Facebook and check your own Status
why OSNs?: reach out real world

Obama raised $690m online in 2012.

50m ‘likes’
why OSNs?: mobilize real world

Arab Spring in 2011

Photo credit: Peter Macdiarmid, Getty Images

Photo credit: Steve Crisp, Reuters
why OSNs?: reflect real world

predicting the future: UK General Election 2010

why OSNs?: predict real world

Twitter mood (*Calm*) predicts Dow Jones Industrial Average (*DJIA*)

why OSN?

- multitude
- sheer scale
- this is where users are!
- reach out real world
- mobilize real world
- reflect real world
- predict real world
WHAT MAKES OSNs ATTRACTIVE FOR RESEARCH?
what’s attractive?:
diverse stakeholders and actors

- users
- OSN operator
- advertisers
- OSN application developers
- user’s employers, ensures, etc.
- law enforcement, intelligence agencies, and other government organizations
- stalkers, investigators, users’ nosy colleagues and neighbors
what’s attractive?

also ...

• new phenomena
• volatile
• socio-technical systems
• new threats, vulnerabilities, defenses
• most of us are OSN users
  — easier to recruit study participants
  — relevance
Facebook easily infiltrated, mined for personal info
Socialbot network could mine 175 chunks of personal data per bot per day


Robots can easily pass as real users on Facebook, allowing them to betray real humans and mine personal information such as birthdays, addresses and phone numbers, Canadian researchers have found.

Socialbots and Facebook data

Researchers have demonstrated a technique capable of stealing information from Facebook.

Using 'socialbots', computer programs that mimic real Facebook profiles, were able to harvest vast quantities of data.

Socialbots are increasingly on the internet as criminals and spammers use them to glean information about users. Some socialbots use as little as $29 (£18) on the internet.

The Register

InfoWorld

Socialbots' stealth gigabytes of Facebook user data works prone to large-scale infiltration

A study by researchers at the University of British Columbia has found that a group of online impersonators could harvest as many as 175 chunks of personal data for each bot operating on the social network.

The Vancouver Sun

The Register

Socialbots steal gigabytes of Facebook user data

Researchers have demonstrated a technique capable of stealing information from Facebook.

Using 'socialbots', computer programs that mimic real Facebook profiles, were able to harvest vast quantities of data.

Socialbots are increasingly on the internet as criminals and spammers use them to glean information about users.

The Register

InfoWorld

Socialbots and Facebook data

Researchers have demonstrated a technique capable of stealing information from Facebook.

Using 'socialbots', computer programs that mimic real Facebook profiles, were able to harvest vast quantities of data.

Socialbots are increasingly on the internet as criminals and spammers use them to glean information about users. Some socialbots use as little as $29 (£18) on the internet.
• relevant

• easy to recruit study participants

• new threats, vulnerabilities, defenses

• socio-technical systems

• volatile

• new phenomena

• diverse stakeholders and actors

what's attractive?

summary
WHAT MAKES OSN RESEARCH CHALLENGING?
what’s challenging?

- overcrowded by researchers
- access to data becoming difficult
- hard to evaluate vulnerabilities/defenses

just in January-May 2013
what's challenging?: moving target

Personal Information Disclosure Trends 2005-2011

Disclosure

We are reluctant to take legal action against UBC, its researchers or students engaged in legitimate academic projects. However, this is not the first time UBC has ignored Facebook’s terms and the law, and UBC ethical obligations, we request that your offices: (1) ensure that UBC researchers cease and desist any and all unauthorized access to Facebook’s site and systems; (2) return to Facebook all illegally harvested user data obtained by UBC researchers and certify destruction of all copies that remain in UBC’s possession; (3) provide an accounting of all research activities involving Facebook and its users; (4) suspend any ongoing Facebook-related research unless and until Facebook provides consent; (5) explain the process by which UBC approved this particular study; and (6) preserve all materials that refer or relate to the UBC’s approval, or lack thereof, for studies involving Facebook and/or its users.
summary of challenges

- overcrowded by researchers
- access to data becoming difficult
- hard to evaluate vulnerabilities/defenses
- moving target
- OSN operators protective
RESEARCH IN OSN SECURITY & PRIVACY
research directions

- de-anonymization
- privacy (game)
- sybil detection/resistance
DE-ANONYMIZATION OF/WITH OSNs
social network data anonymization?

• why?
  – academic and government research
  – advertising
  – third-party applications
  – aggregation
• how?
  – remove node or edge attributes
  – inject random noise

threat agents in de-anonymization attacks

- Large-scale collection of detailed information on individuals
- Abusive marketing aimed at specific individuals
- Craft a highly individualized, believable message
- Recognize the victim’s node in the anonymized network and to learn sensitive information

- Stalkers
- Investigators
- Nosy colleagues
- Employers
- Neighbors
de-anonymization

- active [1]
  - “mark” regions of the graph with injected nodes (Sybils) and/or edges
  - costly on large scale
- passive [2]
  - use “auxiliary” network to re-identify nodes
  - self-reinforcing: seed population increases

PRIVACY IN OSNs
operators collect lots of personal data ...

- Email
- Date of Birth
- Name
- Sex
- Country
- Photo
- Email password
- Username
- Postcode
- City
- State
- Employer
- University

Proportion of sites asking for this data item

... and then make it widely visible

<table>
<thead>
<tr>
<th>Visibility Level</th>
<th>Default</th>
<th>Optional</th>
<th>Unavailable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Internet</td>
<td>41%</td>
<td>-</td>
<td>59%</td>
</tr>
<tr>
<td>All site users</td>
<td>48%</td>
<td>28%</td>
<td>24%</td>
</tr>
<tr>
<td>Sub-networks only</td>
<td>7%</td>
<td>17%</td>
<td>76%</td>
</tr>
<tr>
<td>Friends of friends</td>
<td>-</td>
<td>24%</td>
<td>76%</td>
</tr>
<tr>
<td>Friends only</td>
<td>3%</td>
<td>79%</td>
<td>17%</td>
</tr>
</tbody>
</table>

how do users (re)act?

what makes young users to “lock” data?

1. **no marginal gain for maintenance of an open profile**, when user’s network becomes “large”

2. **expectancy violations by weak-ties** generate privacy concerns among females

3. **engaging in conversational management of privacy** leads to customizing controls

Fred Stutzman, Jacob Kramer-Duffield, “Friends Only: Examining a Privacy-Enhancing Behavior in Facebook” CHI, April 10–15, 2010, Atlanta, GA.
improving models and UIs

privacy communication game

- optimizes OSN interaction with each user group
  - pragmatic majority
    - claims to be interested in privacy
    - forgets about privacy when given an attractive service or monetary rewards
    - more assurance of privacy can make them less comfortable than simply ignoring privacy
  - privacy fundamentalists
    - care deeply about privacy, and
    - may actively investigate a site and complain to non-fundamentalists
- minimize the concerns of the fundamentalists while simultaneously minimizing the awareness of privacy for the pragmatic majority
- poor privacy may be a rational choice for operators

research directions

• de-anonymization
• privacy (game)
• sybil detection/resistance
• detection of compromised accounts
SYBIL DETECTION/RESISTANCE IN OSNs
Hyungshick Kim

About

To see what he shares with friends, send him a friend request.

Contact Information
Website  http://seclab.skku.edu/
Facebook  http://facebook.com/hyoungshick.kim

Friends

Yazan Bashmaf
2 mutual friends

Ildar Musukhov
1 mutual friend

Tai Chung
Works at 알라바 magna

Yongdae Kim
Professor @ 한국과학기술원 (KAIST)

Kevin Yoon
Works at Hila USA Inc

Huy Kang Kim
Seoul, Korea

Seungjoo Kim
(Professor @ 고려대학교 (Korea University))

Ji Won Yoon

Add Friend  Add Friend  Add Friend  Add Friend

Add Friend  Add Friend
Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai, “Uncovering social network sybils in the wild,” In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC '11).
Sybils can be helpful …

... or dangerous

distribute malware

spread misinformation

collect data
can collect personal data

most importantly: can erode trust in ecosystem

Facebook Applications

Facebook Connect
socialbots

Software + Social media account =

Realboy
believable twitter bots

ECE, Olin College

The Web Ecology Project

approaches to reducing sybils in OSNs

• admit into OSN carefully
  – increase trust slowly, by observing actions
  – hurts growth of OSNs, turns users away

• detect (and disable) sybils
  – give full trust right away
  – analyze graph or individual accounts
    • graph-based detection
    • classification based on account “behavior”
  – challenge suspects

• make it hard for sybils to infiltrate the OSN
  – do users care?
  – how can they make better decisions?
innocent by association

approaches to reducing sybils in OSNs

• admit into OSN carefully
  – increase trust slowly, by observing actions
  – hurts growth of OSNs, turns users away
• detect (and disable) sybils
  – give full trust right away
  – analyze graph or individual accounts
    • graph-based detection
    • classification based on account “behavior”
  – challenge suspects
• make it hard for sybils to infiltrate the OSN
  – do users care?
  – how can they make better decisions?
Graph-theoretic Defense Techniques

Honest node

Attack edges

Honest region

Sybil region

Sybil detection via social networks [1]: SybilRank, SybilLimit, SybilGuard, SybillInfer, GateKeeper

in reality it could be like this [2]

a counter-example

do Sybils form connected components?

80% have degree = 0
No edges to other Sybils!

Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai, “Uncovering social network sybils in the wild,” In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC '11).
from 100% TPR to 0% in 2 weeks

Using state-of-the-art local community detection algorithm to detect Sybils during the first two weeks

Yazan Boshmaf, Konstantin Beznosov, Matei Ripeanu, “Graph-based Sybil detection in social and information systems,” in the Proceedings of IEEE/ACM ASONAM, Niagara Falls, ON, Canada (August 2013).
HOW FEASIBLE IS THE RISK OF SYBILS?
a more real example: a social botnet

Bot master

C&C Channel

Bot herder

Social bots

Online Social Network

- SocialBot
- Infiltrated user (randomly picked)
- Infiltrated user (with mutual friends)
more friends, more Sybils

mutual friends matter

possible Sybil indicators

- friend request frequency
- outgoing friend requests accepted

Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai, "Uncovering social network sybils in the wild," In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC '11).
experts detect Sybils much better

approaches to reducing sybils in OSNs

• admit into OSN carefully
  – increase trust slowly, by observing actions
  – hurts growth of OSNs, turns users away

• detect (and disable) sybils
  – give full trust right away
  – analyze graph or individual accounts
    • graph-based detection
    • classification based on account “behavior”
  – challenge suspects

• make it hard for sybils to infiltrate the OSN
  – do users care?
  – how can they make better decisions?
DETECTION OF COMPROMISED ACCOUNTS
cost of compromised accounts

- leverage existing trust relationships
- fake account detection not applicable
- cannot be removed easily
- involves costly password-reset process
a recent approach: COMPA

- statistical modeling
  - Extract behavioral profile for accounts

- anomaly detection
  - Compare new messages against observed behavior

- identify campaigns: similar messages & similar new behavior

COMPA: example

July 4th 2011, @foxnewspolitics

BREAKING NEWS: President @BarackObama assassinated, 2 gunshot wounds have proved too much. It's a sad 4th for #america. #obamadead RIP

Anomaly scores
- Time: 1.00 (1:24am EST, usually 8-10am EST)
- Source: 0.94 (Web, commonly using TweetDeck) – Hashtag: 0.88
- Domain: 0.26
- Mention: 0.67
- Lang: 0.00
COMPA evaluation

Twitter
• Text similarity:
  – 374,920 groups identiKied
  – 9,362 compromised (343,229 accounts)
  – FP: 377 groups (4%), 12,382 accounts (3.6%)
• Landing page similarity:
  – 14,548 groups identiKied
  – 1,236 compromised (54,907 accounts)
  – FP: 72 groups (5.8%), 2,141 accounts (3.8%)

Facebook:
– 48,586 groups identiKied
– 671 compromised (11,499 accounts)
– FP: 22 groups (3.3%), 412 accounts (3.6%)
summary

• why OSNs?
• rewards and challenges of research in OSN
• current research directions
  – de-anonymization
  – privacy (game)
  – Sybil detection/resistance
  – detecting compromised accounts
Security & Privacy in Online Social Networks

Konstantin (Kosta) Beznosov
lerss.ece.ubc.ca
Laboratory for Education and Research in Secure Systems Engineering
BACK UP SLIDES
OSN size, popularity, and age matter

larger, more popular, and more mature sites
• better privacy protection
• longer privacy policies

can Sybil components be detected?

- Sybil components are internally sparse
- Not amenable to community detection

Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai, “Uncovering social network sybils in the wild,” In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC '11).
establishing attack edges takes time

Real-world Sybil activity in Facebook (100 Sybils, fully connected)

SybilTrack
Incremental GSD Algorithm
from graph statistics to graph dynamics

T=1

T=2

T=3

T=4

T=5

T=6