Security Analysis of Malicious Socialbots on the Web

Living in the (malicious) social web: Beyond friendships

Yazan Boshmaf, Konstantin Beznosov, Matei Ripeanu, Dionysious Logothetis, Georgios Siganos, Jose Lorenzo
Social bots

Automated fake accounts in online social networks (OSNs)

Designed to deceive and appear human

The threat of malicious social bots

Automated fake accounts in online social networks (OSNs)

What is at stake?

Designed to deceive and appear human

Fake accounts are bad for business

“… If advertisers, developers, or investors do not perceive our user metrics to be accurate representations of our user base, or if we discover material inaccuracies in our user metrics, our reputation may be harmed and advertisers and developers may be less willing to allocate their budgets or resources to Facebook, which could negatively affect our business and financial results…”
Fake accounts are bad for users

OSNs are attractive medium for abusive users

Social Infiltration

Connecting with many benign users (friend request spam)
Fake accounts are bad for users

OSNs are attractive medium for abusive users

Social Infiltration → Data collection

Online surveillance, profiling, and data commoditization

Fake accounts are bad for users

OSNs are attractive medium for abusive users

Social Infiltration
Data collection
Misinformation

Influencing users, biasing public opinion, propaganda

Ratkiewicz et al. Detecting and tracking political abuse in social media. Proc. of ICWSM. 2011
Fake accounts are bad for users

OSNs are attractive medium for abusive users

Infecting computers and use it for DDoS, spamming, and fraud

Thomas et al. The Koobface botnet and the rise of social malware. Proc. of MALWARE, 2010
Fake accounts are bad for users

Our work

OSNs are attractive medium for abusive content


Infecting computers and use it for DDoS, spamming, and fraud

Questions

1. **Vulnerability analysis**
   - How vulnerable are OSNs to social infiltration?

2. **Characterization of user behavior**
   - What are the security and privacy implications of social infiltration?
   - Quantification of privacy breaches
   - Effectiveness of security defenses

3. **Scalability from economic context**
   - Scalability from economic context
   - Profit-maximizing infiltration strategy

4. **Victim prediction for robust detection**
   - How can OSNs detect fakes or social bots that infiltrate on a large scale?
   - Victim prediction for robust detection
   - Framework for evaluation

**Note:** The diagram shows a flow of questions and sub-questions related to the security and privacy implications of social infiltration in online social networks (OSNs). The questions are designed to explore various aspects including vulnerability analysis, characterization of user behavior, scalability from an economic perspective, and victim prediction for robust detection.
Questions

1. How vulnerable are OSNs to social infiltration?
   - Vulnerability analysis
   - Characterization of user behavior

2. What are the security and privacy implications of social infiltration?
   - Quantifying privacy breaches
   - Effectiveness of security defenses

3. How is the economic rationale behind infiltrating OSNs at scale?
   - Scalability from economic context
   - Profit-maximizing infiltration strategy

4. How can OSNs detect fakes or social bots that infiltrate on a large scale?
   - Victim prediction for robust detection
   - Framework for evaluation
Questions

1. Vulnerability analysis
   - Characterization of user behavior

   **How vulnerable are OSNs to social infiltration?**

2. What are the security and privacy implications of social infiltration?
   - Quantifying privacy breaches
   - Effectiveness of security defenses

3. Scalability in economic context
   - Profit-maximizing infiltration strategy

   **What is the economic rationale behind infiltrating OSNs at scale?**

4. How can OSNs detect fakes or social bots that infiltrate on a large scale?
   - Victim prediction for robust detection
   - Framework for evaluation
Questions

Threat Characterization

1. How vulnerable are OSNs to social infiltration?
   - Vulnerability analysis of OSN platforms
   - Characterization of user behavior

2. What are the security and privacy implications of social infiltration?
   - Quantification of privacy breaches
   - Effectiveness of security defenses

Countermeasure Design

3. What is the economic rationale behind infiltrating OSNs at scale?
   - Scalability from economic context
   - Profit-maximizing infiltration strategy

4. How to detect social bots that infiltrate on a large scale?
   - Is victim prediction feasible
   - Can victim prediction enable robust detection
Attack side: Social infiltration in OSNs

Threat Characterization

1. How vulnerable are OSNs to social infiltration?
   - Vulnerability analysis of OSN platforms
   - Characterization of user behavior

2. What are the security and privacy implications of social infiltration?
   - Quantification of privacy breaches
   - Effectiveness of security defenses

3. What is the economic rationale behind infiltrating OSNs at scale?
   - Scalability from economic context
   - Profit-maximizing infiltration strategy
   - Victim prediction for robust detection
   - Framework for evaluation

4. How can OSNs detect fakes or social bots that infiltrate on a large scale?

---

2. *Key challenges in defending against malicious socialbots*, Boshmaf, Beznosov, Ripeanu, USENIX LEET, April 2012
Social botnet: Experiment

Operated 100 socialbots on Facebook, single botmaster

Bots sent 9.6K friend requests send in 8 weeks, 35.7% requests from bots accepted (victims)
Main findings

(Platform-level vulnerability)

It is feasible to automate social infiltration by exploiting platform and user vulnerabilities.
Main findings

(Data breaches)

Social infiltration results in serious privacy breaches, where personally identifiable information is compromised
Victims are highly affected

<table>
<thead>
<tr>
<th>Profile Info</th>
<th>Direct (%)</th>
<th></th>
<th>Extended (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Before</td>
</tr>
<tr>
<td>Birth Date</td>
<td>3.5</td>
<td>72.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Email Address</td>
<td>2.4</td>
<td>71.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Gender</td>
<td>69.1</td>
<td>69.2</td>
<td>84.2</td>
</tr>
<tr>
<td>Home City</td>
<td>26.5</td>
<td>46.2</td>
<td>29.2</td>
</tr>
<tr>
<td>Current City</td>
<td>25.4</td>
<td>42.9</td>
<td>27.8</td>
</tr>
<tr>
<td>Phone Number</td>
<td>0.9</td>
<td>21.1</td>
<td>1.0</td>
</tr>
<tr>
<td>School Name</td>
<td>10.8</td>
<td>19.7</td>
<td>12.0</td>
</tr>
<tr>
<td>Postal Address</td>
<td>0.9</td>
<td>19.0</td>
<td>0.7</td>
</tr>
<tr>
<td>IM Account ID</td>
<td>0.6</td>
<td>10.9</td>
<td>0.5</td>
</tr>
<tr>
<td>Married To</td>
<td>2.9</td>
<td>6.4</td>
<td>3.9</td>
</tr>
<tr>
<td>Worked At</td>
<td>2.8</td>
<td>4.0</td>
<td>2.8</td>
</tr>
<tr>
<td>Average</td>
<td>13.3</td>
<td>34.9</td>
<td>15.4</td>
</tr>
</tbody>
</table>

2.62 times more private data collected after infiltration
Friends of victims are affected too

<table>
<thead>
<tr>
<th>Profile Info</th>
<th>Direct (%)</th>
<th>Extended (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>Birth Date</td>
<td>3.5</td>
<td>72.4</td>
</tr>
<tr>
<td>Email Address</td>
<td>2.4</td>
<td>71.8</td>
</tr>
<tr>
<td>Gender</td>
<td>69.1</td>
<td>69.2</td>
</tr>
<tr>
<td>Home City</td>
<td>26.5</td>
<td>46.2</td>
</tr>
<tr>
<td>Current City</td>
<td>25.4</td>
<td>42.9</td>
</tr>
<tr>
<td>Phone Number</td>
<td>0.9</td>
<td>21.1</td>
</tr>
<tr>
<td>School Name</td>
<td>10.8</td>
<td>19.7</td>
</tr>
<tr>
<td>Postal Address</td>
<td>0.9</td>
<td>19.0</td>
</tr>
<tr>
<td>IM Account ID</td>
<td>0.6</td>
<td>10.9</td>
</tr>
<tr>
<td>Married To</td>
<td>2.9</td>
<td>6.4</td>
</tr>
<tr>
<td>Worked At</td>
<td>2.8</td>
<td>4.0</td>
</tr>
<tr>
<td>Average</td>
<td>13.3</td>
<td>34.9</td>
</tr>
</tbody>
</table>

1.54 times more, with more than 1 million affected users
Friends of victims are affected too

<table>
<thead>
<tr>
<th>Profile Info</th>
<th>Direct (%) Before</th>
<th>Direct (%) After</th>
<th>Extended (%) Before</th>
<th>Extended (%) After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Date</td>
<td>3.5</td>
<td>72.4</td>
<td>4.5</td>
<td>53.8</td>
</tr>
<tr>
<td>Email Address</td>
<td>2.4</td>
<td>71.8</td>
<td>2.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Gender</td>
<td>69.1</td>
<td>69.2</td>
<td>84.2</td>
<td>84.2</td>
</tr>
<tr>
<td>Home City</td>
<td>26.5</td>
<td>46.2</td>
<td>29.2</td>
<td>45.2</td>
</tr>
<tr>
<td>Current City</td>
<td>25.4</td>
<td>42.9</td>
<td>27.8</td>
<td>41.6</td>
</tr>
<tr>
<td>Phone Number</td>
<td>0.9</td>
<td>21.1</td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td>School Name</td>
<td>10.8</td>
<td>19.7</td>
<td>12.0</td>
<td>20.4</td>
</tr>
<tr>
<td>Postal Address</td>
<td>0.9</td>
<td>19.0</td>
<td>0.7</td>
<td>1.3</td>
</tr>
<tr>
<td>IM Account ID</td>
<td>0.6</td>
<td>10.9</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Married To</td>
<td>2.9</td>
<td>6.4</td>
<td>3.9</td>
<td>4.9</td>
</tr>
<tr>
<td>Worked At</td>
<td>2.8</td>
<td>4.0</td>
<td>2.8</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>13.3</strong></td>
<td><strong>34.9</strong></td>
<td><strong>15.4</strong></td>
<td><strong>23.7</strong></td>
</tr>
</tbody>
</table>

From 49K birthdates to 584K

1.54 times more, with more than 1 million affected users

Acquisti et al. Predicting social security numbers from public data. Proc. Of Nat. Acad. of Sc. 106(27), 2009
Vulnerabilities exploited to automate infiltration

(User behavior characterization)

Some users are more susceptible to social infiltration, which partly depends on factors related to their social structure.
User susceptibility to become a victim correlates with social structure

More friends, more susceptible to infiltration

More mutual friends, more susceptible to infiltration

Without mutual friends

Pearson’s $r = 0.85$

Acceptance rate (%) vs. Number of friends

Acceptance rate (%) vs. Number of mutual friends

- Without mutual friends: $60\%$
- With mutual friends: $80\%$
- More mutual friends lead to higher susceptibility.
Fake accounts mimic real accounts

Only 20% of fakes were “detected”

All manually flagged by concerned users
Defense side: Infiltration-resilient fake account detection

Countermeasure Design

1. Vulnerability analysis of OSN platforms
2. Characterization of user behavior
3. Quantification of privacy breaches
4. Effectiveness of security defenses

1. How vulnerable are OSNs to social infiltration?
2. What are the security and privacy implications of social infiltration?
3. What is the economic rationale behind infiltrating OSNs at scale?
4. How can OSNs detect fakes or social bots that infiltrate on a large scale?

Graph-based Sybil detection in social and information systems. In Proc. of ASONAM, Aug 2013
2. Integro: Leveraging victim prediction for robust fake account detection in OSNs. NDSS, Feb 2015
3. Thwarting fake accounts by predicting their victims. Submitted to TISSEC, Feb 2015
Feature-based detection is ineffective

Only 20% of fakes were “detected”

(Graph-based detection)

Social infiltration invalidates the assumption behind graph-based fake account detection

All manually flagged by concerned users
Graph-based detection

Assumes social infiltration on a large scale is infeasible

Finds a (provably) sparse cut between the regions by ranking
Graph-based detection

Ranks computed from landing probability of a short random walk

Most real accounts rank higher than fakes

Graph-based detection is not resilient to social infiltration

50% of bots had more than 35 attack edges
Premise: Regions can be tightly connected

Cut size = 10 (densest)

Real region

Fake region

- Real
- Trusted
- Victim
- Fake
**Key idea:** Identify potential victims with some probability

Potential victim with probability 0.9

- Real
- Trusted
- Victim
- Fake

Real region
Fake region
**Key idea:** Leverage victim prediction to reduce cut size

Assign lower weight to edges incident to potential victims

Cut size $= 1.9 << 10$
Delimit the real region by ranking accounts

Ranks computed from landing probability of a short random walk

Most real accounts are ranked higher than fake accounts
Result 1: Bound on ranking quality

Number of fake accounts that rank equal to or higher than real accounts is $O(\text{vol}(E_A) \log n)$ where $\text{vol}(E_A) \leq |E_A|$. Assuming a fast mixing real region and an attacker who establishes attack edges at random.
Result 2: Victim classification is feasible (even using low-cost features)

Random Forests (RF) achieves up to 52% better than random

No need to train on more than 40K feature vectors on Tuenti

Integro: Leveraging victim prediction for robust fake account detection in OSNs. NDSS, Feb 2015
Thwarting fake accounts by predicting their victims. Submitted to TISSEC, Feb 2015.
Result 3: Ranking is resilient to infiltration

Integro delivers up to 30% higher AUC, and AUC is always > 0.92
Deployment at Tuenti confirms results

Integro delivers up to an order or magnitude better precision

Highly-infiltrating fakes

Precision at lower intervals

Precision at higher intervals
Research Questions and Contributions

Threat Characterization

1. How vulnerable are OSNs to social infiltration?
2. What are the security and privacy implications of social infiltration?
3. What is the economic rationale behind infiltrating OSNs at scale?

Countermeasure Design

4. How can OSNs detect fakes or social bots that infiltrate on a large scale?

- Vulnerability analysis of OSN platforms
- Characterization of user behavior
- Quantification of privacy breaches
- Effectiveness of security defenses
- Scalability from economic context
- Profit-maximizing infiltration strategy
- Victim prediction for robust detection
- Framework for evaluation
Impact

Public education & further studies

Production-class deployment

Open-source, public release
Research Questions and Contributions

- Vulnerability analysis of OSN platforms
- Characterization of user behavior
  - How vulnerable are OSNs to social infiltration?
- Quantification of privacy breaches
-Effectiveness of security defenses
  - What are the security and privacy implications of social infiltration?
- Scalability from economic context
  - Profit-maximizing infiltration strategy
  - What is the economic rationale behind infiltrating OSNs at scale?
- Victim prediction for robust detection
  - Framework for evaluation
  - How can OSNs detect fakes or social bots that infiltrate on a large scale?

Threat Characterization
Countermeasure Design

Publications

**Primary:**


1. Boshmaf et al. *Key challenges in defending against malicious socialbots.* In Proc. of USENIX LEET, April 2012 (18% acceptance rate)

1. Boshmaf et al. *Design and analysis of a social botnet.* J. Comp. Net., 57(2), Feb 2013 (1.9 impact factor)

1. Boshmaf et al. *Graph-based Sybil detection in social and information systems.* In Proc. of ASONAM, Aug 2013 (13% acceptance rate, **best paper award**)

**Related:**

