On the Consequences of Using Social Botnets for Spam Distribution and Digital-influence Manipulation

Jinxue Zhang, Rui Zhang, Yanchao Zhang, Guanhua Yan

Arizona State University
University of Hawaii
Los Alamos National Laboratory
Outline

- Introduction to social botnets
- Social botnets for spam distribution
- Social botnets for digital-influence manipulation
- Countermeasures and future work
Social bots

- **Social bots**: fake or compromised Online Social Network (OSN) accounts controlled by a botmaster
  - 40% accounts on social media sites are fake or fake-like
  - 8% messages on OSNs are spams

- **Impact of social bots**
  - Befriending and grabbing the personal information [ACSAC’11]
  - Spam campaign [CCS’10, IMC’11]
  - Political astroturf: manipulating/misleading the public [WWW’11]
Social botnets

- A group of social bots collaborating to mimic normal users, embed into the OSNs, evade detection and conduct malicious behaviors
  - E.g., maintain a balanced followings/followers ratio in Twitter
- Several measurement studies have confirmed the existence of the social botnets
  - Social bots are inclined to connect other bots [WWW’11]
  - Build the link farms and also form the spammer social networks [WWW’12]
- What else can social botnets do?
Our contribution

- We demonstrate the advantages of a social botnet for spam distribution
  - Balance the multiple objectives of spam distribution
- We show that social botnets can easily manipulate the digital influence
  - Online influence of OSN users
  - The digital influence is emerging with 20+ vendors
  - Can be easily manipulated by social botnets
- We point out possible countermeasures against social botnets
Build a social botnet on Twitter

- We choose Twitter as the OSN example
  - 500 million users, 340 million tweets per day
  - Unidirectional relationship: follower/friend
  - Information message: 140-character tweets
  - Interaction: retweet, reply and mention
- Bought 1000 Twitter accounts with only $57
- Control the bots by a Twitter app
Social botnet on spam distribution

- **Spam**: unwanted tweet containing malicious URLs

- Distributing spams is one of the main objectives for the botmaster

- We will demonstrate the advantages of using social botnets for spam distribution
- Independent method:
  - Each bot publishes a tweet with the malicious URL
  - All the followers of the bot will see it
Botnet method:

- The botmaster organizes the bots as a retweeting tree/forest
- Only the root(s) of the tree (forest) will publish the spam, while all the children retweet them
Our conjecture

- Twitter is continuously suspending the accounts who posted the spams

- We conjecture that the botnet method is safer than the independent one
Experimental discovery

- We conducted the experiments to compare these two methods.
- For the independent method, we used 100 bots to publish the spams hourly.
- For the botnet method, we built a 10-ary complete tree with 3 levels and total 111 bots. The root published the spams hourly which was then retweeted by its children in a random delay.
- Result: all bots in the independent method have been suspended; only the roots in botnet method have been suspended.
Moving forward...

- We further formulate the spam distribution problem.

- Given a set of social bots, what is the optimal strategy for spam distribution from the standpoint of the botmaster?
Goals

- **Maximize the Coverage** $C$: the number of non-bots who received the spams
- **Minimize the Delay** $\tau$: the average time for each user in $C$ to see a given spam since it is generated by the root bot
- **Minimizing the Cost** $|S|$: #suspended bots
- **Minimizing the Cost** $|\bar{S}|$: #non-bot followers which will be lost due to the suspension of bots
The botmaster/attacker needs to avoid the existing detection methods (rules) from both Twitter and the research community.

- **Strict rules**
  - The violation will cause suspension immediately
  - e.g., tweeting malicious URLs

- **Loose rules**
  - The violation will cause suspicion
  - The accumulated violation may cause suspension
  - e.g., repeatedly posting others’ tweets as its own
Constraints (cont’d)

1. The max. height of a retweeting tree is $K = 10$ [WWW’10]
2. Following before retweeting [RAID’11]
3. The ratio of followers to retweet the message is no more than $r$ [NDSS’12, ACNS’12]
4. $t_i$: the retweeting lag at any hop $i \in [1, K]$ [WWW’10]
5. The social bots within $M$ hops will be suspended [experiment]
Single-objective optimization problem

\[
\min \ f(\{\mathcal{V}_k\}_{k=1}^K) = \alpha \beta \frac{|\tilde{S}|}{|\mathcal{C}|} + (1 - \alpha) \tau
\]

s.t.

\[
\bigcup_{k=1}^{K} \mathcal{V}_i \subseteq \{1, \ldots, n\}
\]

\[
\mathcal{V}_i \cap \mathcal{V}_j = \phi, \ \forall i \neq j \in [1, K]
\]

\[
|\bigcup_{k=1}^{M} \mathcal{V}_i| \leq c
\]

\[
\sum_{i \in \mathcal{V}_k} \left[ \frac{r |\mathcal{F}_i|}{1 - r} \right] \geq |\mathcal{V}_{k+1}|, \ \forall k \in [M - 1, K - 1]
\]
A heuristic solution

- The problem is a variation of classical Set Partition Problem (SPP), which is NP-hard
- A heuristic solution
  - Build the retweeting forest level by level
  - Each level aims to achieve some goals
Trace-driven evaluation

- We carefully set the parameters from the existing measurement work
- Demonstrate the trade-off of multiple goals in the spam distribution problem
- Demonstrate the advantages of distributing the spam by botnet method over the independent method
Outline

- Introduction to social botnet
- Social botnet for spam distribution
- Social botnet for digital-influence manipulation
- Countermeasures and future work
The rise of digital influence

- **Digital influence**: online influence of OSN users

- **Usages:**
  - Viral marketing campaign
  - Rumor diffusion
  - Personal use
Popular vendors

- **Klout**
  - Measure the digital influence based on the OSN activities during the past 90 days
  - Score from 1 to 100 for OSN users
  - 100 million registered users

- **Kred**
  - Measure the digital influence in real-time based on the OSN activities during the past 1000 days
  - Score from 1 to 1000

- **Retweet Rank**
  - Score from 1 to 100
  - Rank the Twitter users on how many times they have been retweeted
Motivation

- Can these digital influence scores be manipulated by a social botnet?

- Intuition: all vendors compute the digital influence scores mainly based on the number of interactions
  - Being retweeted/replied
  - Being mentioned
  - Being followed

- These interactions can be forged by the social botnet
Experiment setup

- We want to show the manipulation of digital influence:
  - Different actions: following, retweeting, mentioning
  - Different vendors: Klout, Kred, Retweet Rank
  - Different speeds: 10, 100, 1000 times per day

- For each action, we choose 10 bots as the target group to manipulate. These target bots have initial scores

- Manipulation speed: 10, 100, 1000 times for the first, second, and third 10 days, respectively

- Record the Klout, Kred and Retweet Rank scores everyday
All three scores can be manipulated by the retweeting and mentions from the social botnet
# Speed of the manipulation

## #days used to manipulate the scores to the top 20%

<table>
<thead>
<tr>
<th></th>
<th>Klout</th>
<th>Kred</th>
<th>Retweet Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score Scale</td>
<td>1-100</td>
<td>1-1000</td>
<td>1-100</td>
</tr>
<tr>
<td>80&lt;sup&gt;th&lt;/sup&gt; percentile score</td>
<td>40</td>
<td>600</td>
<td>80</td>
</tr>
<tr>
<td>#days needs to reach the 80&lt;sup&gt;th&lt;/sup&gt; percentile scores by retweeting</td>
<td>10 per day</td>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>100 per day</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1000 per day</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- By 1000 retweetings per day, a user’s digital influence scores can be manipulated from nothing to top 20% in just one day.
Outline

- Introduction to social botnet
- Social botnet for spam distribution
- Social botnet for digital-influence manipulation
- Countermeasures and future work
Possible countermeasures

- Use an observation: although bots frequently interact with other bots, non-bots are less likely to interact with social bots

- Identify the malicious apps which are exclusively registered by the social bots
Conclusion

- We show the effectiveness and advantages of social botnet on spam distribution and digital influence manipulation

- Future work
  - Defend against the social botnet
  - Design a manipulation-resilient digital-influence system
Thanks!

Q&A
Digital influence: “the ability to cause effect, change behavior, and drive measurable outcomes online”—online version of influence

**Business model**

1. Track the influential users
2. Select the influential users
3. Free perks