The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions

Tao Zhu Independent Researcher
David Phipps Bowdoin College
Adam Pridgen Rice University
Jedidiah R. Crandall University of New Mexico
Dan S. Wallach Rice University
Microblogging sites in China

March 2006

July 2009

August 2009

http://en.wikipedia.org/wiki/Microblogging_in_China
Sina Weibo

- 503 million registered users as of Dec 2012.
  - More than half are from mobile devices.

- About 100 million messages are posted each day on Sina Weibo.

- Promote visibility of social issues.

http://en.wikipedia.org/wiki/Sina_Weibo
Weibo’s influence: Wukan incident - 2011

烏坎 (The village name) vs 鸟坎 (Neologism)
Sina Weibo

- Strict controls over the posts.
Introduction of our research

- Detecting a censorship event within 1-2 minutes of its occurrence.
- Three strategies Weibo system uses to target sensitive content quickly.
Methodology

1. Identifying the sensitive user group

2. Crawling posts of sensitive user groups

3. Detecting deletions
Identifying the sensitive user group

- Use outdated sensitive keywords from China Digital Times

<table>
<thead>
<tr>
<th>Sensitive Word</th>
<th>CDT English Post</th>
<th>Date/Result of Test 1</th>
<th>Date/Result of Test 2</th>
<th>Date/Result of Test 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bo Liu</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bo death</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bo+medical parole</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bo+(Zhou) Yongkang</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bo Wang</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bo Hu</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five Chongqings</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>internal Party struggle</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>support Bo</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master Bo</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>think fondly of+Bo</td>
<td><a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a></td>
<td>2013.7.25 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South China (Morning Post)+Jack Ma</td>
<td><a href="http://goo.gl/mm0Ja">http://goo.gl/mm0Ja</a></td>
<td>2013.7.20 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Identifying the sensitive user group

- Use outdated sensitive keywords from China Digital Times.
- Start with 25 sensitive users.
Identifying the sensitive user group

- Use outdated sensitive keywords from China Digital Times.
- Start with 25 sensitive users.
- Sensitive group reaches 3,567 users after 15 days.
- More than 4,500 deletion daily
Crawling

● **User timeline:**
  ○ Weibo user timeline API returns the most recent 50 posts of the specified user.
  ○ Query 3,567 sensitive users once per minute
    ■ 100 accounts for API call
    ■ 300 concurrent Tor circuits.
  ○ Four-node cluster running Hadoop and Hbase
    ■ 2.38 million posts from July 20 to September 8, 2012.
Detecting deletions

Latest 50 posts

Our database

Deleted Post
Detecting deletions

The lifetime of deleted Post = tn - t0
Detecting deletions

- **Permission-denied or system deletion**
  - “Permission denied” error.
  - Caused by censorship events.
  - The post still exists but cannot be accessed by users.

- **General deletion**
  - “Post does not exist” error.
  - May caused by user self deletion or censorship events.
  - The post does not exist.
Detecting deletions

2.38 Million user timeline posts

- Permission-denied deletion: 4.5%
- General deletion: 8.3%
Detecting deletions

- Permission-denied deletion or system deletion
  - Around 1,500 permission denied deletions.
  - Comparing with WeiboScope, which is tracking around 300,000 users and have no more than 100 permission denied deletions daily.
Distribution of deleted posts

Whole lifetime

First two hours
Strategies to target sensitive contents

1. Weibo has filtering mechanisms as a proactive, automated defense.

2. Weibo targets specific users, such as those who frequently post sensitive content.

3. When a sensitive post is found, a moderator will use automated searching tools to find all of its related reposts, and delete them all at once.
1. Keywords list filtering

- Weibo has filtering mechanisms as a proactive, automated defense
  - Explicit filtering

Sorry, The content violates the relevant laws and regulations. If need help, please contact customer service.
1. Keywords list filtering

- Weibo has filtering mechanisms as a proactive, automated defense
  - Explicit filtering
  - Implicit filtering

Your post has been submitted successfully. Currently, there is a delay caused by server data synchronization. Please wait for 1 to 2 minutes. Thank you very much.
1. Keywords list filtering

- Weibo has filtering mechanisms as a proactive, automated defense
  - Explicit filtering
  - Implicit filtering
  - Camouflaged posts
1. Keywords list filtering

- Weibo has filtering mechanisms as a proactive, automated defense
  - Explicit filtering
  - Implicit filtering
  - Camouflaged posts
  - Surveillance keywords list?
    - If no such list the cost will be too expansive
      - Suppose to censor 70,000 new posts in one minute, it needs 1400 simultaneous workers (50 posts per minute per worker).
      - Assuming 8 hour shift, 4200 works is required.
      - It is not cost efficient
2. Targeting specific users

- Weibo targets specific users, such as those who frequently post sensitive content.
3. Finding all related reposts

- When a sensitive post is found, a moderator can find all of its related reposts, and delete them all at once.
Censors work in the night
Censors catch up on overnight posts by late morning
Conclusion

Whole lifetime

First two hours
Thank you!

Q & A