Chat Program Censorship and Surveillance in China: Tracking TOM-Skype and Sina UC

Jeffrey Knockel, University of New Mexico
Greg Wiseman, Citizen Lab, Munk School, University of Toronto
We...

- Reverse engineered censorship and surveillance of TOM-Skype & Sina UC Chinese chat clients
- Analyzed changes to the triggering keyword lists made over an 18 month period
- Had an unbiased view of all triggering keywords whenever we wanted
TOM-Skype Censorship

- Different versions censor differently
- 6.1 censorship and surveillance: [a1.skype.tom.com/installer/agent/keyfile5.5/keyfile](a1.skype.tom.com/installer/agent/keyfile5.5/keyfile)
- 6.1 surveillance-only: [a1.skype.tom.com/installer/agent/keyfile5.5/keyfile_u](a1.skype.tom.com/installer/agent/keyfile5.5/keyfile_u)
- Each encrypted with DES+ECB with key: "\x7a\xdd\xe7\xdc\x23\x25\x53\x75"
TOM-Skype Surveillance

- Different versions surveil differently
- All send to same php script: a1.skype.tom.com/installer/tomad/ContentFilterMsg.php
- 6.1 example message:
  JohnDoe fuck you 12/31/2011 6:00:00 PM 1 JaneDoe
- Breakdown:
  sender message date-time is-incoming receiver
- Encrypted with DES+ECB with key: "X7sRUjL\0"
消息包含敏感字符！
Sina UC Censorship

- Downloads from URL:
- Five lists JSON-encoded
  - List 1 censors chat and usernames
  - List 2 censors usernames
  - List 4 censors chat
  - Others have unknown purpose
- Encrypted with Blowfish+ECB with key: "H177UC09VI67KASI"
- No *client-side* surveillance
Dataset

- 2,576 distinct keywords across 8 TOM-Skype sources
- 1,818 distinct keywords across 5 Sina UC sources
- Only 138 keywords common between clients
- Lists range in size from 1 to 1,421 unique keywords
- 87% of keywords contain Chinese characters
Analysis

- All keywords human translated
- Associated with political / social context
- Categorized and tagged
- Visualizations
- Comparisons over time and between clients
- Correlating list changes with political events
Keyword Content

● Adaptation to circumvention attempts
  ○ Unicode (e.g. six—four, ⑥④, I Ⅸ Ⅷ Ⅸ)
  ○ Neologisms, homophones (e.g. Bo Xilai written as 薄熙来, 博西莱, BO稀莱)

● Highly specific keywords
  ○ 西大直街康宁路路口世纪联华 ("Corning West and Da Zhi Street intersection, Century Lianhua gate")

● Extremely broad keywords
  ○ 华人 ("Chinese person")
  ○ 互联网 ("Internet")

● WTFs
  ○ Baby Mama Drama
List changes: TOM-Skype
List changes: Sina UC
Reaction to Sensitive Events

- Identified current events referenced in the dataset and correlated them with keyword list updates
- Inconsistent patterns across selected events
- Seemingly important events were not always represented
- Two examples
  - Ferrari crash (March 2012)
  - Wenzhou train crash (July 2011)
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Mar 2012</td>
<td>Three people were thrown out of the car (527)</td>
</tr>
<tr>
<td>18 Mar 2012</td>
<td>Three people were thrown out (1232)</td>
</tr>
<tr>
<td>25 Mar 2012</td>
<td>Three people were thrown out of the car (517)</td>
</tr>
<tr>
<td>01 Apr 2012</td>
<td>Beijing Ferrari car accident (1260)</td>
</tr>
<tr>
<td></td>
<td>Beijing Ferrari car accident (515)</td>
</tr>
<tr>
<td></td>
<td>Beijing Ferrari car accident (1259)</td>
</tr>
<tr>
<td></td>
<td>Beijing Ferrari car accident (1258)</td>
</tr>
<tr>
<td></td>
<td>Beijing Ferrari car accident (1257)</td>
</tr>
<tr>
<td></td>
<td>Latest news on Beijing Ferrari car crash (1256)</td>
</tr>
<tr>
<td></td>
<td>Beijing fourth ring (highway) car accident (533)</td>
</tr>
<tr>
<td></td>
<td>Ferrar -- one death two injured (531)</td>
</tr>
<tr>
<td></td>
<td>Ferrar -- one man two women (520)</td>
</tr>
<tr>
<td></td>
<td>Ferrar -- Baful [Bridge] (529)</td>
</tr>
<tr>
<td></td>
<td>Ferrar -- 4 am (518)</td>
</tr>
<tr>
<td></td>
<td>Ferrar -- Beijing Fourth ring (highway) (532)</td>
</tr>
<tr>
<td></td>
<td>Ferrar crashes into wall disintegrates (516)</td>
</tr>
<tr>
<td></td>
<td>Ferrar hits bridge (526)</td>
</tr>
<tr>
<td></td>
<td>Ferrar hits the bridge (1231)</td>
</tr>
<tr>
<td></td>
<td>Ferrar death (528)</td>
</tr>
<tr>
<td></td>
<td>Ferrar death (1233)</td>
</tr>
<tr>
<td></td>
<td>Ferrari car accident (530)</td>
</tr>
<tr>
<td></td>
<td>One man and two women in car accident (519)</td>
</tr>
<tr>
<td></td>
<td>Car accident -- 4am (522)</td>
</tr>
</tbody>
</table>
Wenzhou Train Crash
Conclusion

- Data set is unbiased and comprehensive, but many open questions remain
- Changes in censorship / surveillance focus
- Importance of interdisciplinary research
- Upcoming paper
- Website and (processed) data will be public soon