Honeybot
Your Man in the Middle for
Automated Social Engineering

Institute Eurecom

Tobias Lauinger          Davide Balzarotti
Veikko Pankakoski       Engin Kirda
Automated Social Engineering

- Spambot sending spam scales well
- Attack is “easy” to identify by users
- Phisher chatting with victims is “hard” to detect by users
- Attack does not scale

How could attackers improve this?

Click here if you want to see me naked: http://123.123.123.123/

Good morning sir
> Good morning
We need to verify your details
> Why?
We do this periodically
Could you give me your birth date?
> ...
Previous Work

- Introduced notion of ASE
- Chatterbot, identified by users after 3 messages (80%)

- A pathological chatterbot example (ELIZA):

  Emil

  Hello, I am Eliza.

  Hello Eliza, how are you?

  Would you prefer if I were not?

  Eliza
Honeybot in the Middle

- Bot initiates conversations with users on chat
- Bot uses human user to answer messages

Emil

Hi there!
Hello Barbara!
Hi, how are you?
ROFL: http://ww...

Barbara (Honeybot)

Fritz

Hello Fritz!
Hi, how are you?
Does This Work in Practice?

We want to test Honeybot in the wild... …in an ethical way.

- Risks for test subjects
  - Waste of time
  - Revealing personal information
  - Emotional consequences
- Careful setup to minimise these risks
- Evaluation on IRC during 74 days

For clarity of presentation, only results of channel *Dating 1*. 
Bootstrapping a Conversation

• Say *Hi, wanna chat?* to 1\textsuperscript{st} user & forward reply

- Total success probability 59.5%
- Total median bootstrapping delay 44s
Maintaining a Conversation

• Forwarding messages, median duration 112s

• Replacing male ↔ female words: duration 317s

I'm a *gentleman*, you know.

I'm a *lady*, you know.
Attack, Part 1: Links

- Different **contents** & **occasion** of links

<table>
<thead>
<tr>
<th>Link Type</th>
<th>Keyword</th>
<th>Random</th>
<th>Replacement</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP Address</td>
<td>50.5%</td>
<td>59.7%</td>
<td>58.3%</td>
<td>54.5%</td>
</tr>
<tr>
<td>TinyURL</td>
<td>61.3%</td>
<td>64.5%</td>
<td>87.5%</td>
<td>63.5%</td>
</tr>
<tr>
<td>MySpace</td>
<td>56.4%</td>
<td>71.3%</td>
<td>77.8%</td>
<td>62.8%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>55.9%</td>
<td>64.8%</td>
<td>76.1%</td>
<td>60.1%</td>
</tr>
</tbody>
</table>
Attack, Part 2: Questions

• *btw, what was US president Obama's first name again? I completely forgot*
  – 56.1% correct answers (keyword matching)

• *do u know where is the eiffel tower? I know it's in France but where???
  – 47.2% correct answers*
Countermeasures

• Technical
  – Prevent message forwarding, warning next to links, block links...
  – Can be circumvented

• Systematic
  – Talk to verified friends only, but: Profile cloning
  – Trust-based mechanisms
  – User education, but: Attack difficult to detect
Conclusion

• Towards automating social engineering
  – Using human to answer messages
  – Influence conversation
  – Automated & human (scalable and difficult to detect)

• Tested spamming & questioning
  – high click rates
  – good stealth: “you’ve got a virus, seek help!”

• Could be used to spy on conversations in underground economy channels
Questions?

xkcd.com