

---

# Honeybot

Your Man in the Middle for  
Automated Social Engineering

*Institute Eurecom*

Tobias Lauinger

Veikko Pankakoski

Davide Balzarotti

Engin Kirda

# Automated Social Engineering

*iSecLab*  
Institute Eurecom

- 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

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

How could attackers improve this?

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

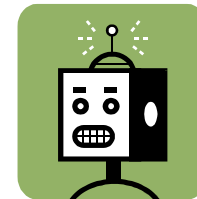
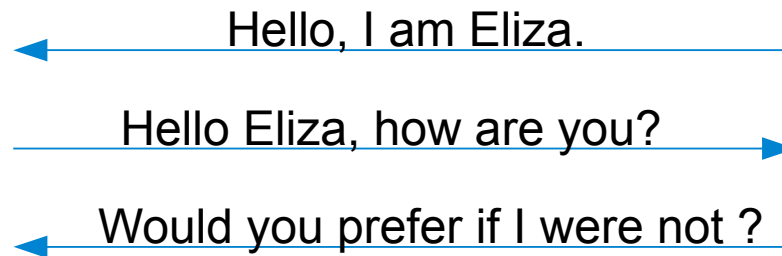
# Previous Work

Huber, Kowalski, Nohlberg, Tjoa. *Towards automating social engineering using social networking sites*. In CSE, 2009.

- Introduced notion of ASE
- Chatterbot, identified by users after 3 messages (80%)
- A pathological chatterbot example (ELIZA):



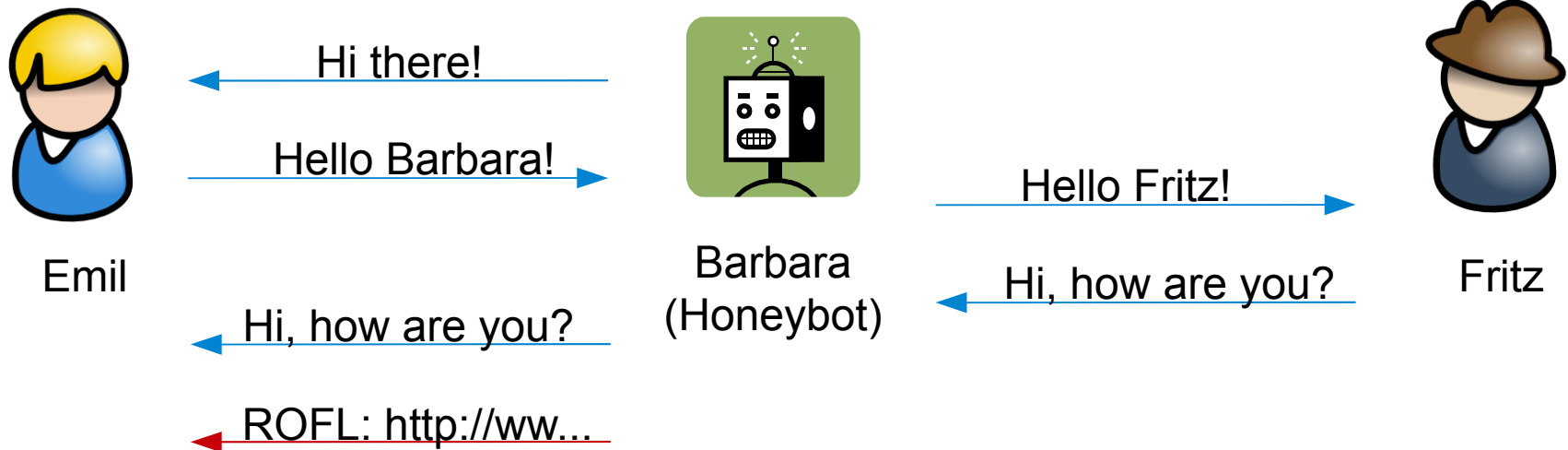
Emil



Eliza

# Honeybot in the Middle

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



# Does This Work in Practice?

iSecLab  
Institute Eurecom

---

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

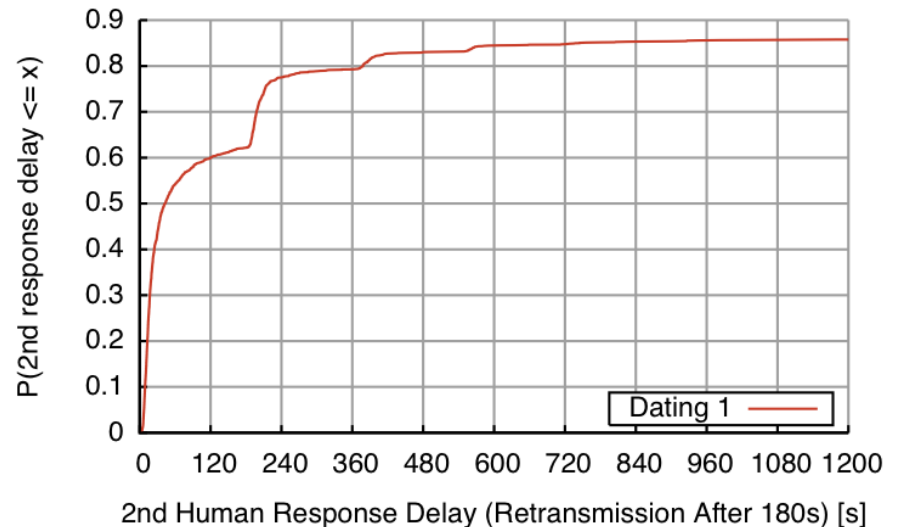
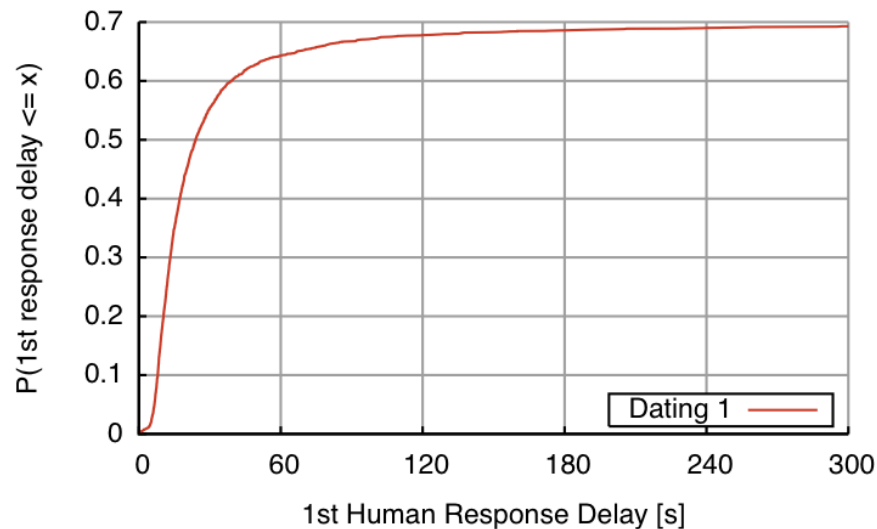
- Risks for test subjects
  - Waste of time
  - Revealing personal information
  - Emotional consequences
- Careful setup to minimise these risks
- Evaluation on IRC during 74 days

---

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

# Bootstrapping a Conversation

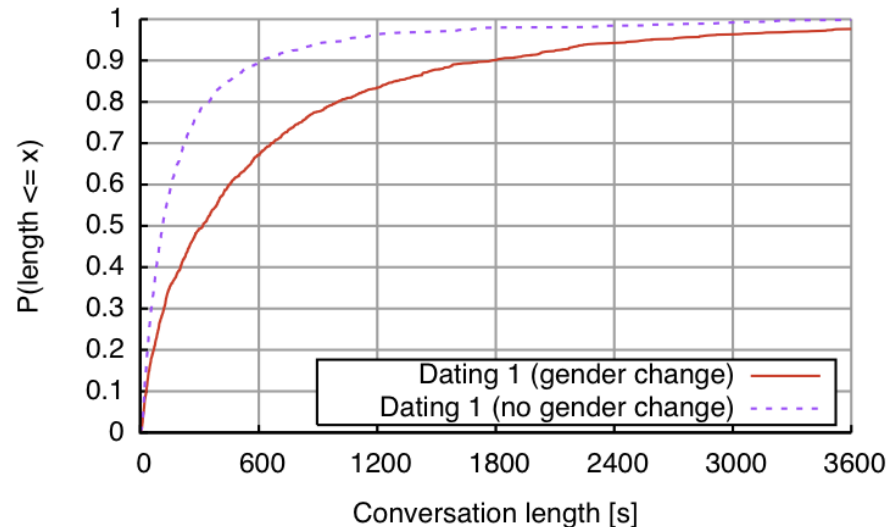
- Say *Hi, wanna chat?* to 1<sup>st</sup> user & forward reply



- Total success probability 59.5%
- Total median bootstrapping delay 44s

# Maintaining a Conversation

- Forwarding messages, median duration 112s

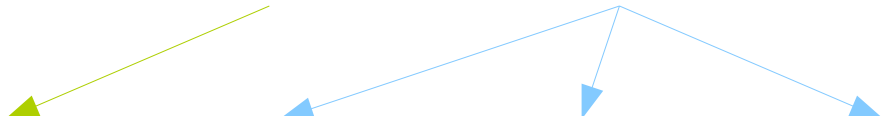


- Replacing male ↔ female words: duration 317s



# Attack, Part 1: Links

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



Link Type	Keyword	Random	Replacement	TOTAL
IP Address	50.5%	59.7%	58.3%	<b>54.5%</b>
TinyURL	61.3%	64.5%	87.5%	<b>63.5%</b>
MySpace	56.4%	71.3%	77.8%	<b>62.8%</b>
<b>TOTAL</b>	<b>55.9%</b>	<b>64.8%</b>	<b>76.1%</b>	<b>60.1%</b>



# 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

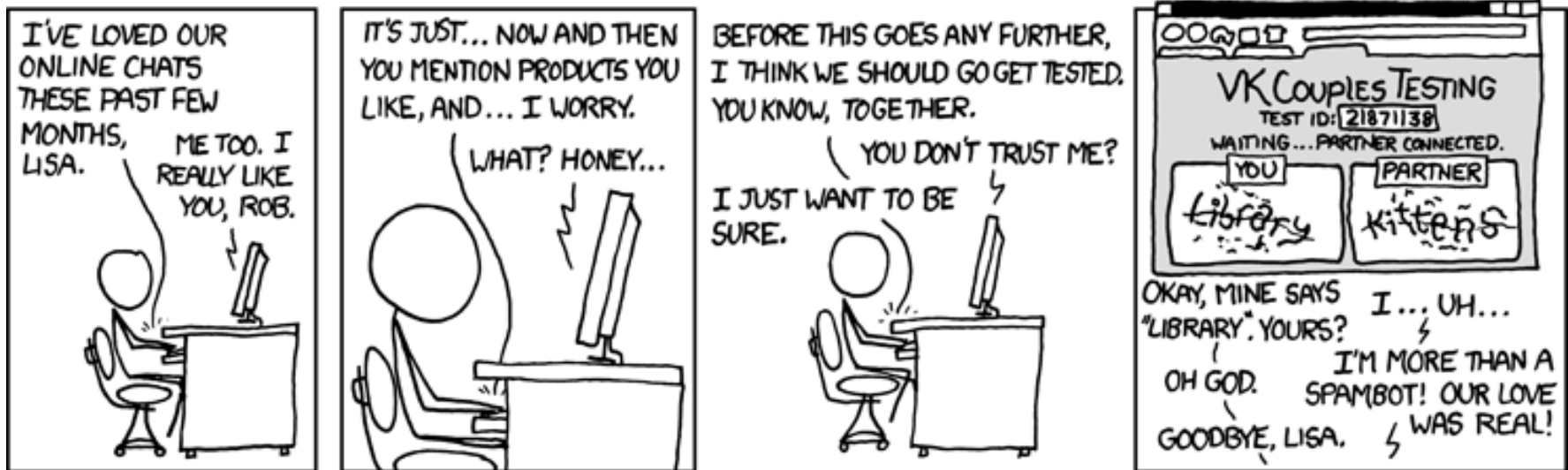
# Countermeasures

- 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