CIS 700/003: Distributed Systems meet Social Networks

Measurement of behavior

January 26, 2010
Where are we?

- Measurement and analysis of MAD systems
  - Measuring the social component
  - Measuring the technical component
  - Measuring behavior

- Building MAD systems
- Faults and misbehavior
- Internet crime
- Privacy and confidentiality
- Novel opportunities
- Experience
My discussion points

- Is their approach legitimate?
  - Comparison to medical research
  - "Strictly reduce harm"?
  - Are there any circumstances or results that could justify an active spam campaign?

- How to avoid measuring other researchers?
- Economic approach to spam prevention?
- Arms race: How to stay ahead?
- What can we learn from this paper?
  - "You did WHAT?"
- Other potential applications of this approach?
Some of your discussion points

- **Ethics / "strictly reduce harm"**
  - Why not impersonate an actual worker bot - can estimate MTA delivery rate
  - Should they have attempted a 'cleanup' after the experiment?

- **Is a click on checkout really a conversion?**
  - Regional differences

- **Profits are estimated**
  - What about other sources of revenue, e.g., using CC number for fraud, DDoS, distributed calculations for nefarious purposes?
  - Do the numbers justify the conclusions?

- **Not representative**
  - Low absolute numbers / high error margin
  - Just one botnet and two campaigns
  - No explanation for differences in response rates. Impact of product quality?
  - Types of spam other than self-propagation and pharmaceuticals?
  - Time variance, e.g., due to impact of social engineering?

- **Effectiveness of CBL**
Countermeasures against spam

- URL blacklists
- IP blacklists
- Reputation filtering
- Machine learning
  - Bayesian inference
  - SVM
- Social relationships (RE:, Ostra)
- Copyright (Habeas, Inc.)
- Micropayments, pledges
Mystery paper

This paper is not published yet, so these slides will NOT appear on the course web page.
SLIDES DELETED - THIS PAPER IS NOT YET PUBLISHED
Where are we?

- Measurement and analysis of MAD systems
  - Measuring the social component
  - Measuring the technical component
  - Measuring behavior

- Building MAD systems

- Faults and misbehavior

- Internet crime

- Privacy and confidentiality

- Novel opportunities

- Experience

© 2010 Andreas Haeberlen
Recap: Measurement of MAD systems

- **OSN study**
- **DisCarte**
- **Spamalytics**

- **News cycle study**
- **Broadband study**
- **Mystery paper**

**Social component**

**Technical component**

**Behavior**
Stay tuned!

Next week you will learn:

How you can distribute content more effectively using many trees