Sybil In Online Social Networks (OSNs)

- Sybil ($\textit{sibl}$): fake identities controlled by attackers
  - Friendship is a pre-cursor to other malicious activities
  - Does not include benign fakes (secondary accounts)

- Research has identified malicious Sybils on OSNs
  - Twitter [CCS 2010]
  - Facebook [IMC 2010]
  - Renren [IMC 2011], Tuenti [NSDI 2012]
Real-world Impact of Sybil (Twitter)

- Russian political protests on Twitter (2011)
  - 25,000 Sybils sent 440,000 tweets
  - Drown out the genuine tweets from protesters
Security Threats of Sybil (Facebook)

- Large Sybil population on Facebook
  - August 2012: 83 million (8.7%)
- Sybils are used to:
  - Share or Send Spam
  - Theft of user's personal information
  - Fake like and click fraud

- Malicious URL: [Mark Zuckerberg - Official Announcement](http://apps.facebook.com/acnts_chkr/)

- 50 likes per dollar: 
  - 1,000 Facebook Likes: $34.90
  - 3,000 Facebook Likes: $79.90
  - 5,000 Facebook Likes: $129.90
  - 10,000 Facebook Likes: $199.90
Community-based Sybil Detectors

Prior work on Sybil detectors

- SybilGuard [SIGCOMM’06], SybilLimit [Oakland ’08], SybilInfer [NDSS’09]

Key assumption: **Sybils form tight-knit communities**
- Sybils have difficulty “friending” normal users?
Do Sybils Form Sybil Communities?

- Measurement study on Sybils in the wild [IMC’11]
  - Study Sybils in Renren (Chinese Facebook)
  - Ground-truth data on 560K Sybils collected over 3 years

- Sybil components: sub-graphs of connected Sybils

- Sybil components are internally sparse
- Not amenable to community detection
- New Sybil detection system is needed
Detect Sybils without Graphs

- Anecdotal evidence that people can spot Sybil profiles
  - 75% of friend requests from Sybils are rejected
  - Human intuition detects even slight inconsistencies in Sybil profiles

- Idea: build a crowdsourced Sybil detector
  - Focus on user profiles
  - Leverage human intelligence and intuition

- Open Questions
  - How accurate are users?
    - What factors affect detection accuracy?
  - How can we make crowdsourced Sybil detection cost effective?
Outline

- Introduction
- User Study
  - Feasibility Experiment
  - Accuracy Analysis
  - Factors Impacting User Accuracy
- Scalable Sybil Detection System
- Conclusion

Details in Paper
User Study Setup*

- User study with 2 groups of testers on 3 datasets
- 2 groups of users
  - Experts – Our friends (CS professors and graduate students)
  - Turkers – Crowdsworkers from online crowdsourcing systems
- 3 ground-truth datasets of full user profiles
  - Renren – given to us by Renren Inc.
  - Facebook US and India – crawled
    - Sybils profiles – banned profiles by Facebook
    - Legitimate profiles – 2-hops from our own profiles

*IRB Approved
Classifying Profiles

Real or fake?

Why?

Navigation Buttons

Classifying Profiles

Browsing Profiles

Screenshot of Profile
(Links Cannot be Clicked)
<table>
<thead>
<tr>
<th>Dataset</th>
<th># of Profiles</th>
<th>Test Group</th>
<th># of Testers</th>
<th>Profile per Tester</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sybil</td>
<td>Legit.</td>
<td>Chinese Expert</td>
<td>24</td>
</tr>
<tr>
<td>Renren</td>
<td>100</td>
<td>100</td>
<td>Chinese Turker</td>
<td>418</td>
</tr>
<tr>
<td>Facebook US</td>
<td>32</td>
<td>50</td>
<td>US Expert</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>49</td>
<td>US Turker</td>
<td>299</td>
</tr>
<tr>
<td>Facebook India</td>
<td>50</td>
<td>49</td>
<td>India Expert</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>India Turker</td>
<td>342</td>
</tr>
</tbody>
</table>

**More Profiles per Experts**
Individual Tester Accuracy

- Experts prove that humans can be accurate
- Turkers need extra help…

>80% accuracy!
Wisdom of the Crowd

- Is wisdom of the crowd enough?

- Majority voting
  - Treat each classification by each tester as a vote
  - Majority vote determines final decision of the crowd

- False positive rates are excellent
- What can be done to improve turker accuracy?
Eliminating Inaccurate Turkers

Removing inaccurate turkers can effectively reduce false negatives!
Outline

- Introduction
- User Study
- Scalable Sybil Detection System
  - System Design
  - Trace-driven Simulation
- Conclusion
A Practical Sybil Detection System

1. **Scalability**
   - Must scale to millions of users
   - High accuracy with low costs

2. **Preserve user privacy when giving data to turkers**

Key insight to designing our system
- Accuracy in turker population highly skewed
- Only 10% turkers > 90% accurate

Details in Paper
System Architecture

Maximize Utility of High Accuracy Turkers

- Continuous Quality Control
- Locate Malicious Workers

Flag Suspicious Users

Social Network

User Reports

Suspicious Profiles

Turker Selection

OSN Employees

Very Accurate Turkers

Accurate Turkers

Turker

Rejected!

All Turkers

Sybils
Trace Driven Simulations

- Simulation on 2000 profiles
- Error rates drawn from survey data
- Calibrate 4 parameters to:
  - Minimize false positives & false negatives
  - Minimize votes per profile (minimize cost)

**Results**

- Average 6 votes per profile
- <1% false positives
- <1% false negatives

**Results++**

- Average 8 votes per profile
- <0.1% false positives
- <0.1% false negatives
Estimating Cost

- Estimated cost in a real-world social networks: Tuenti
  - 12,000 profiles to verify daily
  - 14 full-time employees
    - Annual salary 30,000 EUR (~$20 per hour) → $2240 per day

- Crowdsourced Sybil Detection
  - 20 sec/profile, 8 hour day, 50 turkers
    - Facebook wage ($1 per hour) → $400 per day

- Cost with malicious turkers
  - 25% of turkers are malicious
  - $504 per day

Augment existing automated systems

*http://www.glassdoor.com/Salary/Tuenti-Salaries-E245751.htm*
Conclusion

- Designed a crowdsourced Sybil detection system
  - False positives and negatives <1%
  - Resistant to infiltration by malicious workers
  - Low cost

- Currently exploring prototypes in real-world OSNs
Questions?

Thank you!
Ground-truth Data Collection (Legit.)

- **Facebook Crawl**
  - 8 Seeds
  - 1-hop friends
  - 86k 2-hop friends
  - Random Selection
  - 50 US
  - 50 IN
  - 100 Legitimate Profiles
Ground-truth Data Collection (Sybil)

Facebook Crawl

Profile Pictures

Google Search By Image

Suspicious Profiles

Confirmed Sybils

Publicly Available Image

573 Confirmed Sybils

Banned by Facebook

Google Search

Suspicious Profiles Dataset

Users

50% of pictures on web suspicious profiles

If >90% of pictures on web suspicious, consider Facebook links

Confirmed Sybils

Users

Facebook Crawl

Profile Pictures

Google Search By Image

Suspicious Profiles

Confirmed Sybils

Publicly Available Image

573 Confirmed Sybils

Banned by Facebook

Google Search

Suspicious Profiles Dataset

Users

50% of pictures on web suspicious profiles

If >90% of pictures on web suspicious, consider Facebook links
Preserving User Privacy

- Showing profiles to crowdworkers raises privacy issues
- Solution: reveal profile information in context

[Diagram showing crowdsourced evaluation and friend-only profile information]
Survey Fatigue

US Experts

US Turkers

No fatigue: All testers speed up over time matters.
Wisdom of the Crowd

- Treat each classification by each tester as a vote
- Majority vote determines final decision

<table>
<thead>
<tr>
<th>Dataset</th>
<th>False Positives</th>
<th>False Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renren</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>Chinese Expert</td>
<td>0%</td>
<td>63%</td>
</tr>
<tr>
<td>Chinese Turker</td>
<td>2%</td>
<td>19%</td>
</tr>
<tr>
<td>Facebook</td>
<td>0%</td>
<td>10%</td>
</tr>
<tr>
<td>US Expert</td>
<td>0%</td>
<td>19%</td>
</tr>
<tr>
<td>US Turker</td>
<td>2%</td>
<td>19%</td>
</tr>
<tr>
<td>Facebook</td>
<td>0%</td>
<td>16%</td>
</tr>
<tr>
<td>India Expert</td>
<td>0%</td>
<td>16%</td>
</tr>
<tr>
<td>India Turker</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

- False positive rates are excellent
- Turkers need extra help against false negatives
- What can be done to improve accuracy?
Sybil Profile Difficulty

Experts perform well on most difficult Sybils

- Some Sybils are more stealthy
- Experts catch more tough Sybils than turkers

Sybil Profiles Ordered By Turker Accuracy

Average Sybil (%)

Sybil Profiles Ordered By Turker Accuracy
How Many Votes Do You Need?

- Only need a few votes
- False positives reduce quickly
- Fewer votes = less cost
Individual Tester Accuracy

- **Experts prove that humans can be accurate.**
- **Turkers need extra help…**

**Chinese Turker**

**Much Lower Accuracy**

**Excellent!**

80% of experts have >90% accuracy!