A Double-edged Sword: Implications of Crowdsourcing on Web Security

Ben Y. Zhao
UC Santa Barbara
IPCCC 2012 Keynote
Austin, TX
What is Crowdsourcing?

- **Crowdsourcing**: a process that enlists many people to do small jobs to solve problems that software cannot.

Effective when a **big** problem can be decomposed into **small** easy tasks.
Search for Jim Gray (2007)

- Jim Gray, Turing Award winner
- Missing with his sailboat outside San Francisco Bay, Jan 2007
- No result from searches of coastguard and private planes

- Use satellite image to search for Jim Gray’s sailboat
- Problem: search cannot be automated by machine
- Solution
  - Split the satellite image into many small images
  - Volunteers look for his boat in each image

100,000 tasks completed in 2 days
The Original

- Popular crowdsourcing services
  - Amazon Mechanical Turk, FreeLancer
  - Tasks: translation, transcription, product survey, etc.

Crowdsourcing is much more pervasive today!
Community Q&A Sites

• Crowdsourcing Q&A

• Yahoo Answers, 1\textsuperscript{st} and largest
  o Started in 2005
  o 1 Billion answers by May 2010
  o 300 Million questions by July 2012
  o But activity and traffic both dropping

• The “hottest” site today: Quora
  o Well curated, deep questions
  o Leveraged power-law graph to focus attention
  o True domain experts write answers politicians, actors, startup founders etc…
Traffic Navigation: Waze

- Crowdsourced traffic system
  - Warns of congestion, police, accidents
  - Automatically measures traffic speed
- 20M users, July’12
- Recommended by Tim Cook as a replacement for Apple Maps
TaskRabbit

- Crowdsourcing physical tasks
  - Rigging up Christmas lights
  - Holiday shopping, Gift wrapping
  - Food delivery
  - Campaign/election worker

- Workers (rabbits) have background checks
  - Interviews, essay, background checks
  - Gain levels with work experience

- Helps with
  - Job posting
  - Figuring out task price
  - Locating open jobs
To date, most crowdsourcing research focused on efficiency and quality of answers.

Recent work examined real-time crowdsourcing, HCI applications.

**But what about security?**
Outline

• Crowdsourcing, a brief introduction

• Crowdturfing: the dark side of crowdsourcing

• Crowdsourcing Sybil detection
Malicious Content / Spam on the Web

• We tend to think of spam as "low quality"
• What about high quality spam?
  o Fake reviews that seem legit
  o Sybil accounts that look like normal users

Christo Wilson
Max Gentleman
is the bestest male enhancement system available.

HYp://cid-ce6ec5.space.live.com/

Stock Photographs
Reviews for Arigato Sushi

7/19/2007

Amandeep Kaur
LPN at NIKEID Lives in Amritsar, Punjab In a relationship Knows Punjabi, English, Hindi From Amritsar, Punjab

Work and Education
Employers I NIKEID LPN - Jan 2011 - present Amritsar, Punjab

Reviews for Arigato Sushi

Arigato is amazing, the service is perfect, the atmosphere is french seaside bistro with a hint of japanese fishing village and cozy and perfect for dates. they have an excellent sake selection as well.

Arts and Entertainment

Rama Choudary
Aparna Goyal

Music

Shakira
Sade
Sade Song
Defending Against Automated Spam

- Variety of CAPTCHA tests
  - Read fuzzy text, solve logic questions
  - Rotate images to natural orientation
  - Identify friends (Social CAPTCHA)

- Detecting bots using behavioral models
  - Detect high volume in per-IP API requests

But what if the enemy is a real human being?

Who is tagged in the photo?
Black Market Crowdsourcing

- Amazon’s Mechanical Turk

- Observation: **black market** crowdsourcing websites
  - Social spam generated by **real people**
  - Major force in China (water army), growing in USA and India

Astroturfing + Crowdsourcing = Crowdturfing

- Admins remove spammy jobs
A Recent Example (May 2012)

Connect to LinkedIn Profiles by hand

Bids: 23 Avg Bid: $318 USD

**Project ID:** 1572287  **Project Type:** Fixed

**Budget:** $250-$750 USD

**Project Description:**

It is very simple:

You will get the 1st LinkedIn Account.

You will get a LinkedIn Industry (http://www.linkedin.com/directory/companies/), from each of the companies within the Industry you will friend 5-20 people.

You will friend 300 people with that LinkedIn Account, sending 1 of 15 different versions of a text message.

You will save the Company Name, the Profile Names, Profile URLs in an excel Sheet that we provide.

If you are done with the Account you will take the next Account, that we prodive. (If you have already good fake LinkedIn Accounts, we would pay extra for that service)
Analysis of Crowdturfing Sites

• Focus on the two largest sites

![ZBJ](www.zhubajie.com)  ![SDH](sandaha.com)

• Crawling ZBJ and SDH
  o Details are completely open
  o Complete campaign history since going online
    • ZBJ 5-year history
    • SDH 2-year history
Crowdturfing Workflow

Customers
- Initiate campaigns
- May be legitimate businesses

Agents
- Manage campaign and workers
- Verify completed tasks

Workers
- Complete tasks for money
- Control Sybils on other websites

Campaign

Tasks

Reports
Campaign Information

Promote our product using your blog

Campaign ID: [40854]
Input: ¥100元
Category: Blog Promotion
Rewards: 100 tasks, each ¥0.8
77 submissions accepted
Still need 23 more
Status: Ongoing (177 reports submitted)

Report generated by workers

Report ID: 2814244号
WorkerID: WYQ951456
Experience: 10 中级
Reputation: 

URL

Accepted!

Screenshot
## Crowdturfing Growth in China

### Table: Site Growth Over Time

<table>
<thead>
<tr>
<th>Site</th>
<th>Active Since</th>
<th>Total Campaigns</th>
<th>Workers</th>
<th>Reports</th>
<th>$ for Workers</th>
<th>$ for Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZBJ</td>
<td>Nov. 2006</td>
<td>76K</td>
<td>169K</td>
<td>6.3M</td>
<td>$2.4M</td>
<td>$595K</td>
</tr>
</tbody>
</table>

### Diagram: Site Growth Over Time

- **ZBJ**
- **SDH**

The chart shows the growth in campaigns and dollars per month from Jan. 2008 to Jan. 2011.
• Campaign 1: promote a Travel Agent

Weibo (microblog)

Great deal! Trip to Maldives!

Workers

Create Spam

Task Info

ZBJ (Crowdturfing Site)

New Job Here!

Check Details

Measurement Server

Redirection

Travel Agent

Trip Info

Weibo Users

Weibo (microblog)
Campaign Results

<table>
<thead>
<tr>
<th>Campaign</th>
<th>About</th>
<th>Target</th>
<th>Cost</th>
<th>Task/Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip</td>
<td>Advertise for a trip organized by travel agent</td>
<td>Weibo</td>
<td>$15</td>
<td>100/108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QQ</td>
<td>$15</td>
<td>100/118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forum</td>
<td>$15</td>
<td>100/123</td>
</tr>
</tbody>
</table>

- Our Settings:
  - One-week campaign
  - $45 per Campaign ($15 per target)

- Cost per click:
  - Sina Weibo ($0.21), QQ ($0.09), Forum ($0.9)
  - Price > Web display Ads ($0.01)

- Shenzhen Travel Agency
  - Averaged 2 sales/month before campaign
  - 11 sales in 24 hours after campaign
  - Each trip sells for $1500

- 80% of reports are generated in the first few hours
Crowdturfing in the US

<table>
<thead>
<tr>
<th>US Sites</th>
<th>% Crowdturfing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Legit</strong></td>
<td></td>
</tr>
<tr>
<td>Mechanical Turk</td>
<td>12%</td>
</tr>
<tr>
<td>Freelancer</td>
<td>28%</td>
</tr>
<tr>
<td><strong>Black Market</strong></td>
<td></td>
</tr>
<tr>
<td>MinuteWorkers</td>
<td>70%</td>
</tr>
<tr>
<td>MyEasyTasks</td>
<td>83%</td>
</tr>
<tr>
<td>Microworkers</td>
<td>89%</td>
</tr>
<tr>
<td>ShortTasks</td>
<td>95%</td>
</tr>
</tbody>
</table>

- Growing problem in the United States
  - More black market sites popping up
  - Exponential growth trends match Chinese sites
  - International workers who speak English
Outline

• Crowdsourcing, a brief introduction

• Crowdturfing: malicious crowdsourcing for cheap

• Crowdsourcing Sybil detection
  o Can we use it for the forces of good?
Crowdsourcing as Force for Good

• Social networks losing the battle against fake accounts
  o Measurements show Sybils do not form clusters, target insertion into specific communities instead (IMC 2011)

• Idea: build a crowdsourced Sybil detector
  o Leverage human intelligence and intuition
  o Resilient to changing attacker strategies
  o Scalable?

• Open Questions
  o How accurate is human based detection?
  o What factors affect detection accuracy?
  o Is this approach scalable, i.e. cost effective to social networks (millions of users)?
Large User Study

• Two groups of users
  o Experts – CS professors, masters, and PhD students
  o Turkers – workers from Mechanical Turk and ZBJ

• Three ground-truth datasets of full user profiles
  o Both fake (Sybil) and legitimate user profiles
  o Renren – given to us by Renren Inc.
  o Facebook US and India
    • Crawled (only publicly accessible data)
    • Legitimate profiles – 2-hops from our own profiles
    • Suspicious profiles – generic profile images
    • Sybil profiles – Banned suspicious profiles
Testers may skip around and revisit profiles.

Real or fake? Why?

Navigation buttons:

Progress
Classifying Profiles
Browsing Profiles

Screenshot of Profile (Links Cannot be Clicked)
## Experiment Overview

### Dataset Overview

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of Profiles</th>
<th># of Testers</th>
<th># of Profile per Tester</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renren China</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Facebook US</td>
<td>32</td>
<td>117</td>
<td>50</td>
</tr>
<tr>
<td>Facebook India</td>
<td>50</td>
<td>101</td>
<td>49</td>
</tr>
</tbody>
</table>

### Data from Renren

- Small # of Experts
- More Profiles for Experts

### Test Group

<table>
<thead>
<tr>
<th>Test Group</th>
<th># of Testers</th>
<th>Profile per Tester</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese Expert</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td>Chinese Turker</td>
<td>418</td>
<td>10</td>
</tr>
<tr>
<td>US Expert</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>US Turker</td>
<td>299</td>
<td>12</td>
</tr>
<tr>
<td>India Expert</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>India Turker</td>
<td>342</td>
<td>12</td>
</tr>
</tbody>
</table>

Crawled I More Profiles for Experts
**Individual Tester Accuracy**

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

Not so good :(

false positives
Are certain profiles more difficult?

Experts perform well on most difficult Sybils

- Some Sybils are more stealthy
- Experts catch more tough Sybils than turkers
Other Factors: Survey Fatigue?

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

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

## False Positives

<table>
<thead>
<tr>
<th>Dataset</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>India Expert</td>
<td>0%</td>
</tr>
<tr>
<td>India Turkers</td>
<td>0%</td>
</tr>
<tr>
<td>Renren</td>
<td>3%</td>
</tr>
<tr>
<td>China Experts</td>
<td>0%</td>
</tr>
<tr>
<td>China Turkers</td>
<td>63%</td>
</tr>
<tr>
<td>Facebook US Expert</td>
<td>0%</td>
</tr>
<tr>
<td>Facebook US Turkers</td>
<td>2%</td>
</tr>
<tr>
<td>Facebook India Expert</td>
<td>0%</td>
</tr>
<tr>
<td>Facebook India Turkers</td>
<td>0%</td>
</tr>
<tr>
<td>India Expert</td>
<td>16%</td>
</tr>
<tr>
<td>India Turkers</td>
<td>50%</td>
</tr>
</tbody>
</table>

- False positive rates are excellent
- Turkers need extra help against false negatives
- How can we improve system accuracy?
Most workers are >40% accurate

Dramatic
From 60% to 10%

- Only a subset of workers are removed (<50%)
- Getting rid of inaccurate turkers is no-brainer
How Many Votes to Convergence?

- Only need a few votes
- False positives reduce quickly
- Less survey fatigue
Crowdsourced Sybil Detection Advantages

- **Scalability**
  - Must scale to many millions of users

- **Performance**
  - Improves turker accuracy
  - Reduces costs

- **Privacy**
  - Preserves user privacy when giving data to turkers
System Architecture

Initial Filtering Layer

Filter Out <60% Accurate Turkers

Leverage Existing Techniques
• Helps the System Scale

Auto Filters
User Reports

Social Network

Experts

Turker Selection

Very Accurate Turkers

Maximize Usefulness of High Accuracy Turkers

Suspicious Profiles
Trace Driven Simulations

- Simulation run w/ 2000 profiles
- Error rates drawn from survey data
- Vary 4 parameters

**Results:**
- Average 6 votes per profile
- <1% false positives
- <1% false negatives
Estimating Cost

• Estimated cost in a real-world social networks: Tuenti
  o 12,000 profiles to verify daily, 14 full-time employees
  o Salary (30K Euros/yr salary) $2520 per day
  o Minimum wage ($8 per hour) $890 per day

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

• Cost with malicious turkers
  o Estimate that 25% of turkers are malicious
  o 63 turkers
  o $1 per hour $504 per day
The Next Security Battleground

• Human labor: the next battle ground for security
  o Internet: the great global equalizer
  o Outsourcing at its best in tough economic times

• A double-edged sword
  o Attackers got here first…
    • Crowdturfing growing exponentially around the world
    • We need to design defenses today
  o Turning the tide…
    • Even cheap labor can be effective at Sybil detection
    • Understanding of user attention/behavior a critical step
What’s Next

- Crowdsourcing applications expand in scope
  - Physical tasks are next, likely for a higher price
  - Deliver packages, engage in group behavior
  - Given sufficient funds, there are few limits
  - Worst case: innocuous actions combine to enable illegal activity

- “Responsibility” or “opportunity” for research?
  - How do you stop a crowdsourcing C&C channel?
  - How do you separate the good from the dangerous?
  - How do you navigate tradeoffs between security and privacy?
  - Highly interdisciplinary challenges involving issues of psychology, communications, economics, law
Thank You.

Questions?

Papers, data, etc… at http://current.cs.ucsb.edu

Most of content here available in our WWW 2012 and NDSS 2013 papers