Social Turing Tests: Crowdsourcing Sybil Detection

Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang,

Miriam Metzger, Haitao Zheng, Ben Y. Zhao

Crowdsourcing vs. Sybil Identities

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

- Idea: build a crowdsourced Sybil detector
  - Leverage human intelligence and intuition
  - Resilient to changing attacker strategies

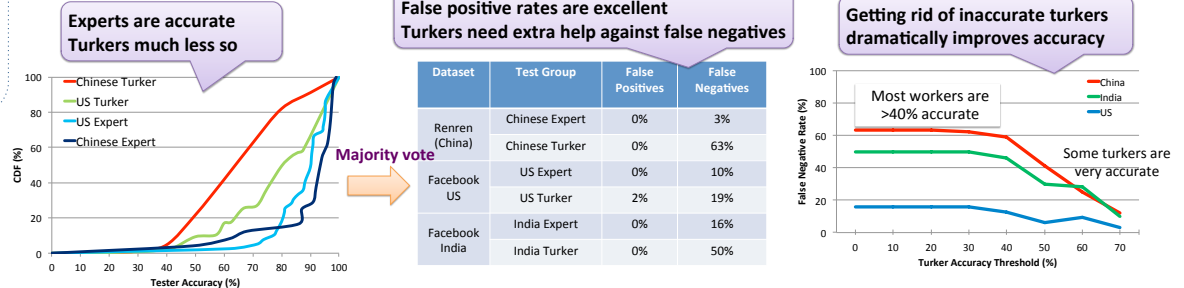*Crowdsourcing:* a process that enlists many people to do small jobs to solve problems that software cannot

Tasks/Problems → Solution

Effective when a big problem can be decomposed into small easy tasks

- Open Questions
  - How accurate is human based detection?
  - What factors affect detection accuracy?
  - Is this approach scalable, i.e. cost effective for large systems?
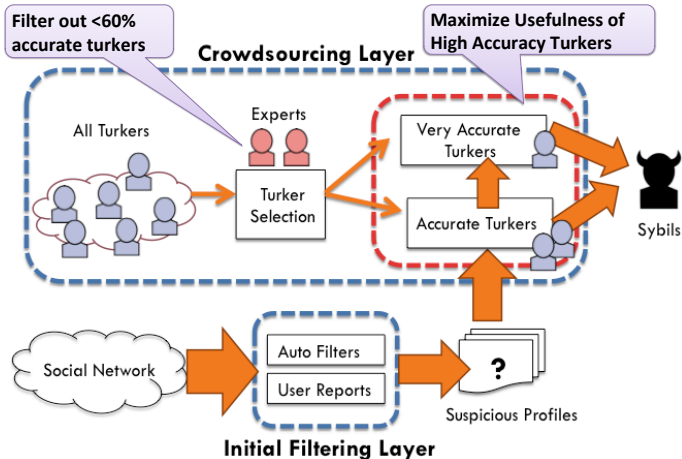
Large User Study

- Two groups of users
  - Experts – CS professors, masters, and PhD students
  - Turkers – crowdworkers from Mechanical Turk and 猪八戒

- Three ground-truth datasets of full user profiles
  - Both fake (Sybil) and legitimate user profiles
  - 人人 – given to us by Renren Inc.
  - 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

Dataset	# of Profiles			Test Group	# of Testers	Profile per Tester
	Sybil	Susp.	Legit.			
Renren (China)	100	0	100	Chinese Expert	24	100
				Chinese Turker	418	10
Facebook US	32	117	50	US Expert	40	50
				US Turker	299	12
Facebook India	50	101	49	India Expert	20	100
				India Turker	342	12

Experts are accurate Turkers much less so

Majority vote →

False positive rates are excellent Turkers need extra help against false negatives

Dataset	Test Group	False Positives	False Negatives
Renren (China)	Chinese Expert	0%	3%
	Chinese Turker	0%	63%
Facebook US	US Expert	0%	10%
	US Turker	2%	19%
Facebook India	India Expert	0%	16%
	India Turker	0%	50%

Getting rid of inaccurate turkers dramatically improves accuracy

Most workers are >40% accurate

Some turkers are very accurate

A Crowdsourced Sybil Detection System

Filter out <60% accurate turkers
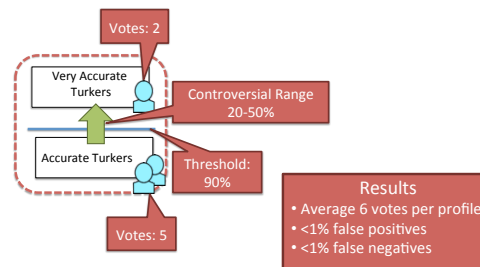
Maximize Usefulness of High Accuracy Turkers

Advantages

- Scale to many millions of users, low relative cost
- Extremely high accuracy
- Limit information exposure when giving data to turkers

Votes: 2

Very Accurate Turkers

Controversial Range 20-50%

Accurate Turkers

Threshold: 90%

Votes: 5

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

Cost Estimation

- Estimated cost in a real-world social networks: Tuenti
  - 12,000 profiles to verify daily
  - 14 full-time employees
  - Minimum wage ($8 per hour) → $890 per day

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

- Cost with malicious turkers
  - Estimate that 25% of turkers are malicious
  - 63 turkers
  - $1 per hour → $504 per day