

Sybil In Online Social Networks (OSNs)

1

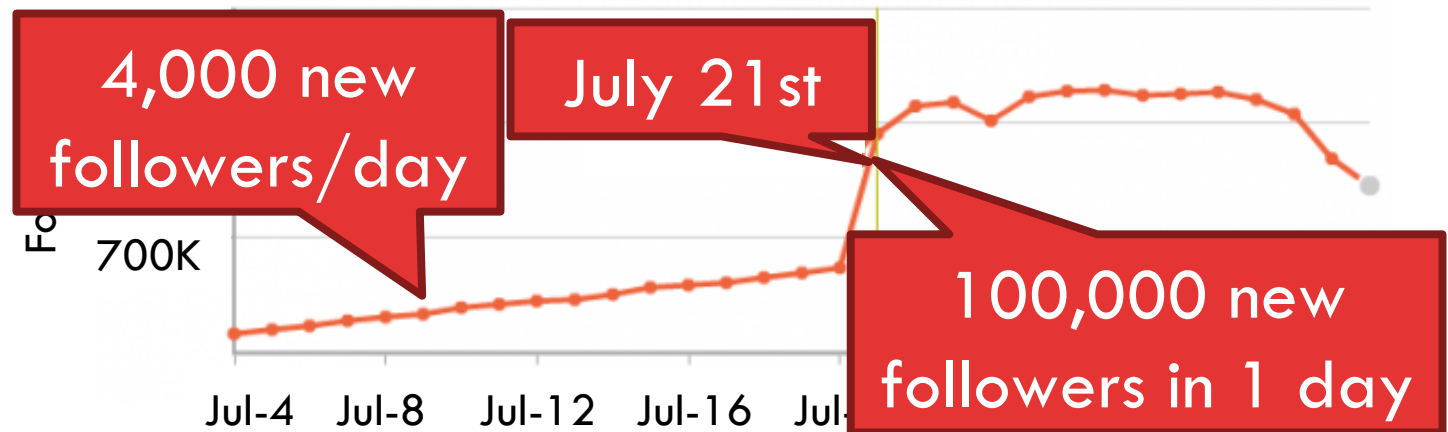


- Sybil (*stɪbəl*): fake identities controlled by attackers
 - ▣ Friendship is a pre-cursor to other malicious activities
 - ▣ Does not include benign fakes (secondary accounts)

- 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)

2



- 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)

3

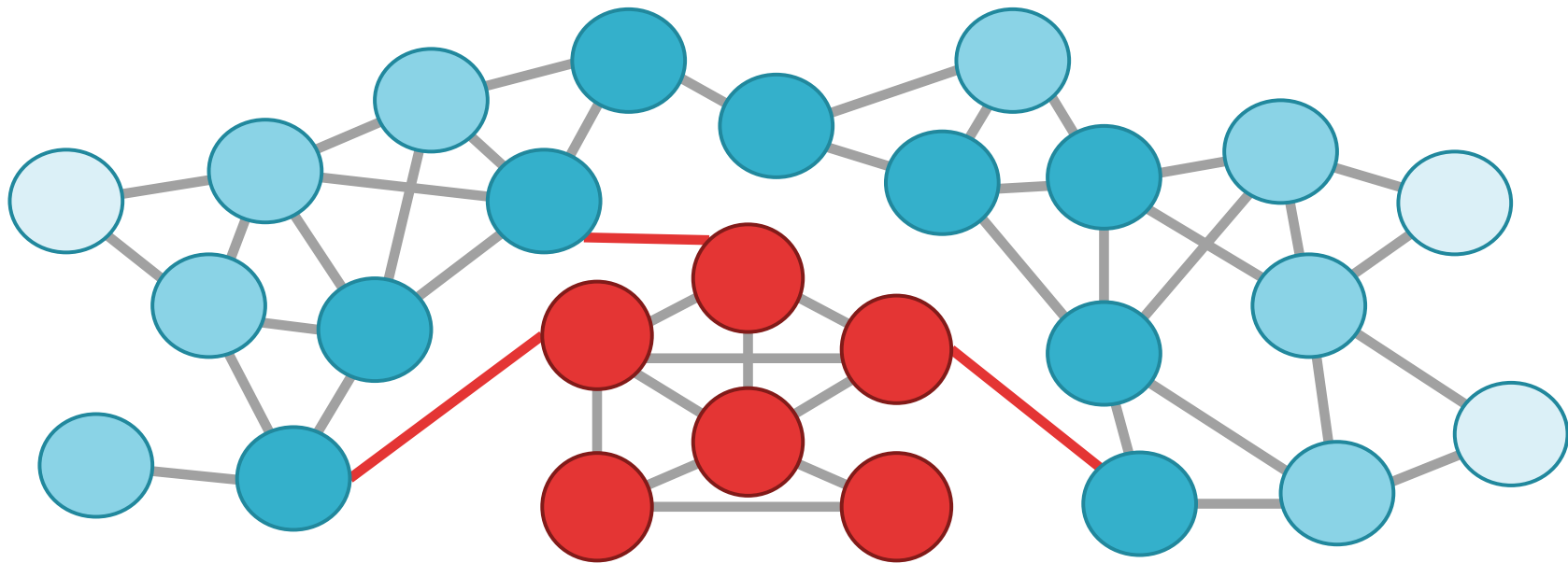
- ❑ Large Sybil population on Facebook
 - ▣ August 2012: 83 million (8.7%)
- ❑ Sybils are used to:
 - ▣ Share or Send Spam



Community-based Sybil Detectors

4

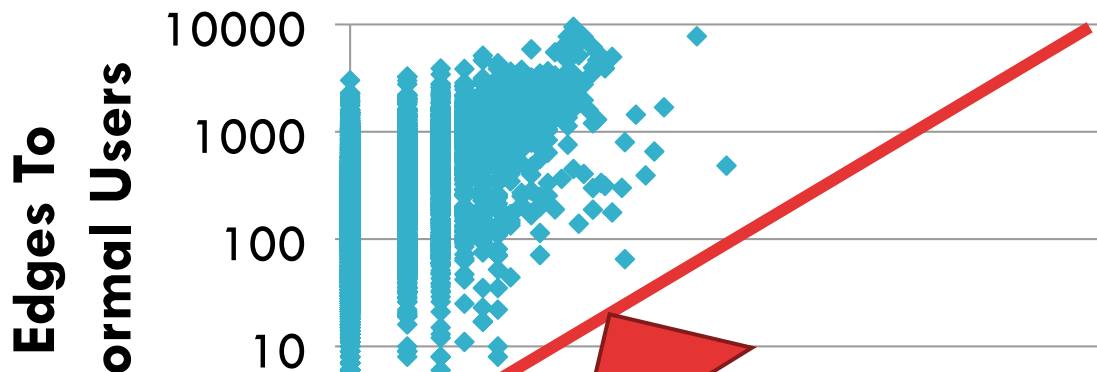
- Prior work on Sybil detectors
 - ▣ SybilGuard [SIGCOMM'06], SybilLimit [Oakland '08], SybilInfer [NDSS'09]
 - ▣ Key assumption: **Sybils form tight-knit communities**
 - Sybils have difficulty “friending” normal users?



Do Sybils Form Sybil Communities?

5

- 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

6

- 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**?

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

Details in
Paper

User Study Setup*

8

- User study with 2 groups of testers on 3 datasets
- 2 groups of users
 - ▣ Experts – Our friends (CS professors and graduate students)
 - ▣ Turkers – Crowdworkers 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** profile – crawled profiles by Facebook
 - **Legitimate** profile – crawled profiles by Facebook

Data collection details

Social Turing Tests: Crowdsourcing Sybil Detection
Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang[†],
Miriam Metzger[‡], Haitao Zheng and Ben Y. Zhao

Abstract

As popular tools for spreading spam and malware, Sybils (or fake accounts) pose a serious threat to online communities such as Online Social Networks (OSNs). Today, sophisticated attackers are creating realistic Sybils that effectively evade most automated Sybil detection systems. In this paper, we explore the effectiveness of crowdsourcing Sybil detection systems for OSNs. We conduct a large-scale study on the ability of human workers to identify Sybils using the Facebook and Renren datasets. Our results show that crowdsourcing is highly effective in identifying Sybils, and that the accuracy of human workers is significantly higher than that of automated systems. The research community has produced a substantial number of techniques for automated detection of Sybils [4, 32, 33]. However, with the exception of SybilRank [3], few techniques rely on the assumption that Sybil accounts have difficulty friending legitimate users, and thus tend to form their own communities, making them visible to community detection techniques applied to the social graph [29]. Unfortunately, the success of these detection schemes is likely to decrease over time as Sybils adopt more sophisticated strategies to ensnare legitimate users. Few user studies on OSNs such as Facebook show that Sybils rarely try to infiltrate connected components. For example, from [2], Second, despite the discovery of a million Sybils on the Renren network, these Sybils rarely created links to other users. Finally, these Sybils rarely try to infiltrate communities. For

*IRB Approved

Sybil.Detector

0 out of

Real or fake?

Why?

The below profile is: If fake, mark suspicious content (multiple choice)

Real

Fake

Navigation Buttons

← Classifying Profiles

Please browse the below profile



Rachel Thompson

Worked at Victoria Secret Studied at Harvard University Lives in New York, New York From Paris, France

Work and Education

Employers

Victoria Secret



College



Harvard

Class of 2

High School



Columbus High School

← Screenshot of Profile
(Links Cannot be Clicked)

← Browsing Profiles

Experiment Overview

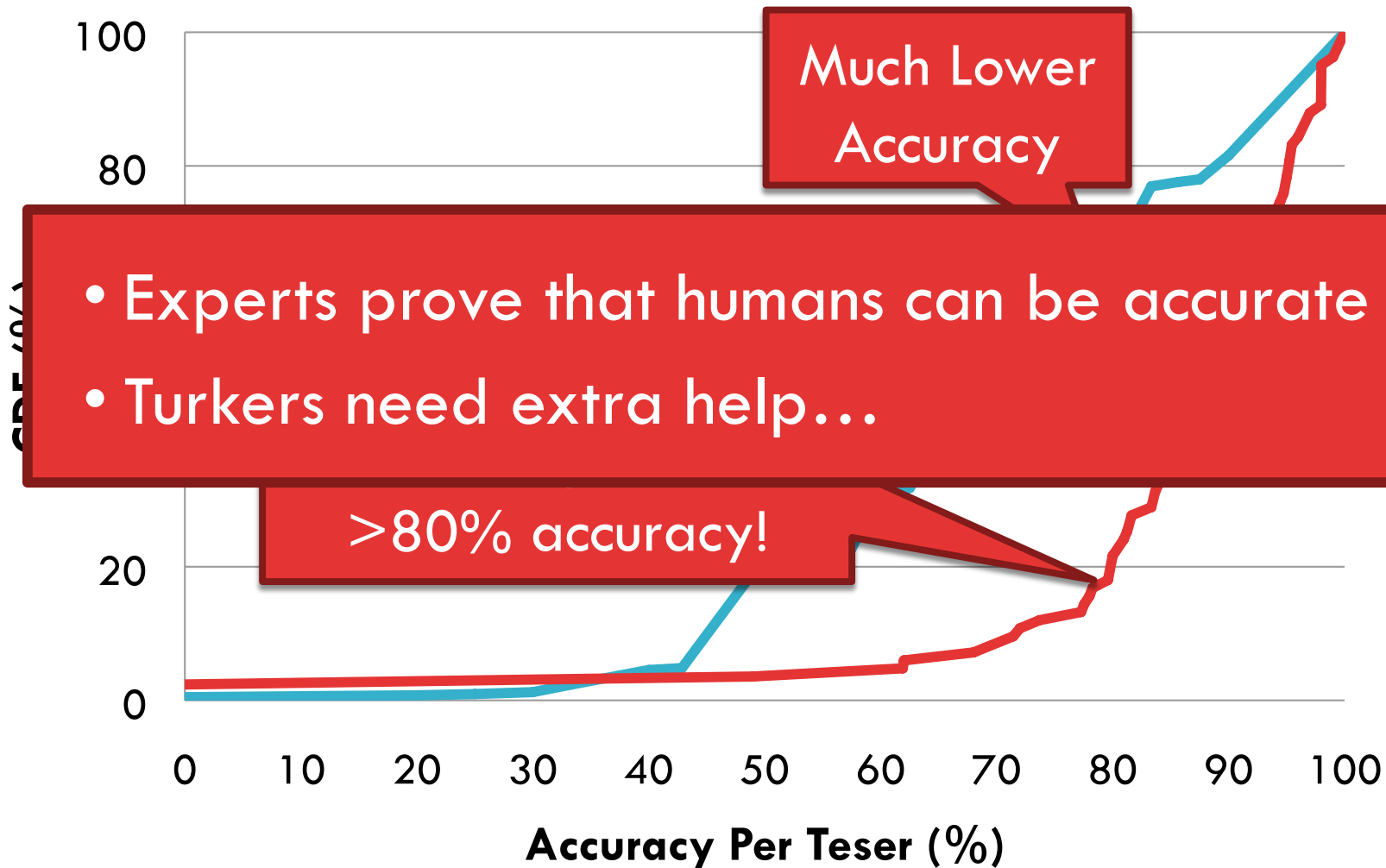
10

Dataset	# of Profiles		Test Group	# of Testers	Profile per Tester
	Sybil	Legit.			
Renren	100	100	Chinese Expert	24	100
			Chinese Turker	418	10
Facebook US	32	50	US Expert	40	50
			US Turker	299	12
Facebook India	50	49	India Expert	20	100
			India Turker	342	2

More Profiles per Experts

Individual Tester Accuracy

11



Wisdom of the Crowd

12

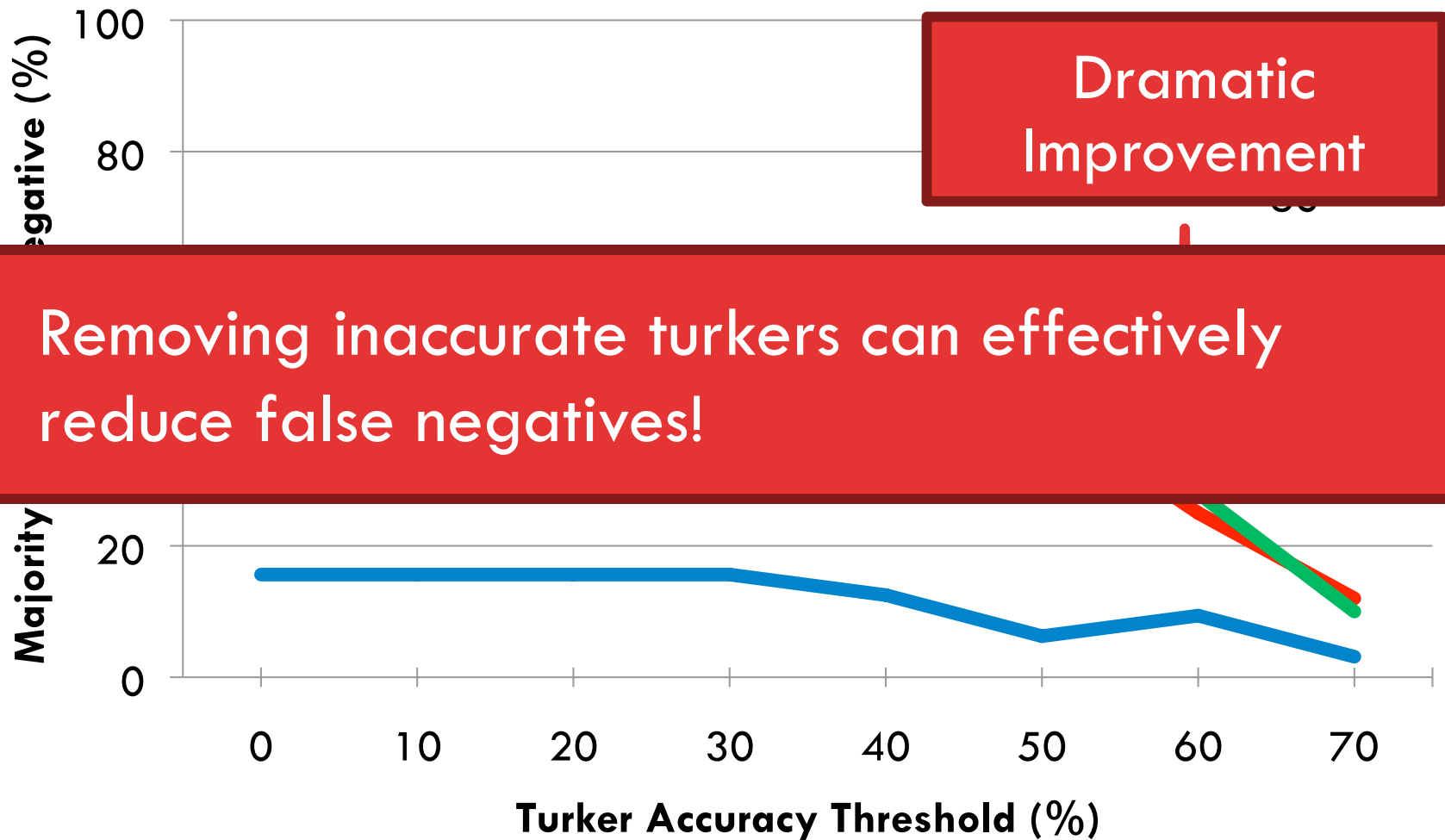
- Is wisdom of the crowd enough?

- 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

13



Removing inaccurate turkers can effectively reduce false negatives!

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

A Practical Sybil Detection System

15

1. Scalability

- ▣ Must scale to millions of users
- ▣ High accuracy with low costs

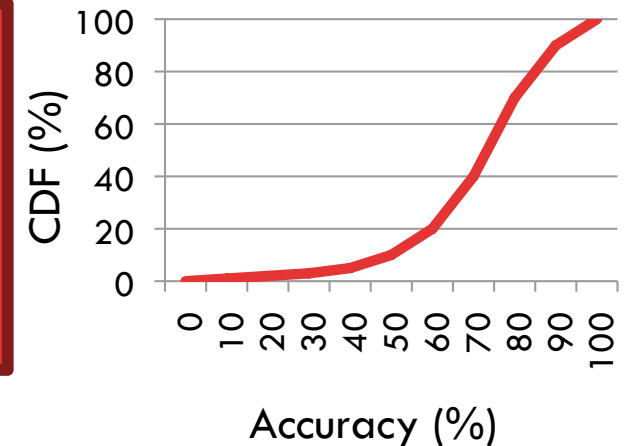


Details in
Paper

2. Preserve user privacy when giving data to turkers

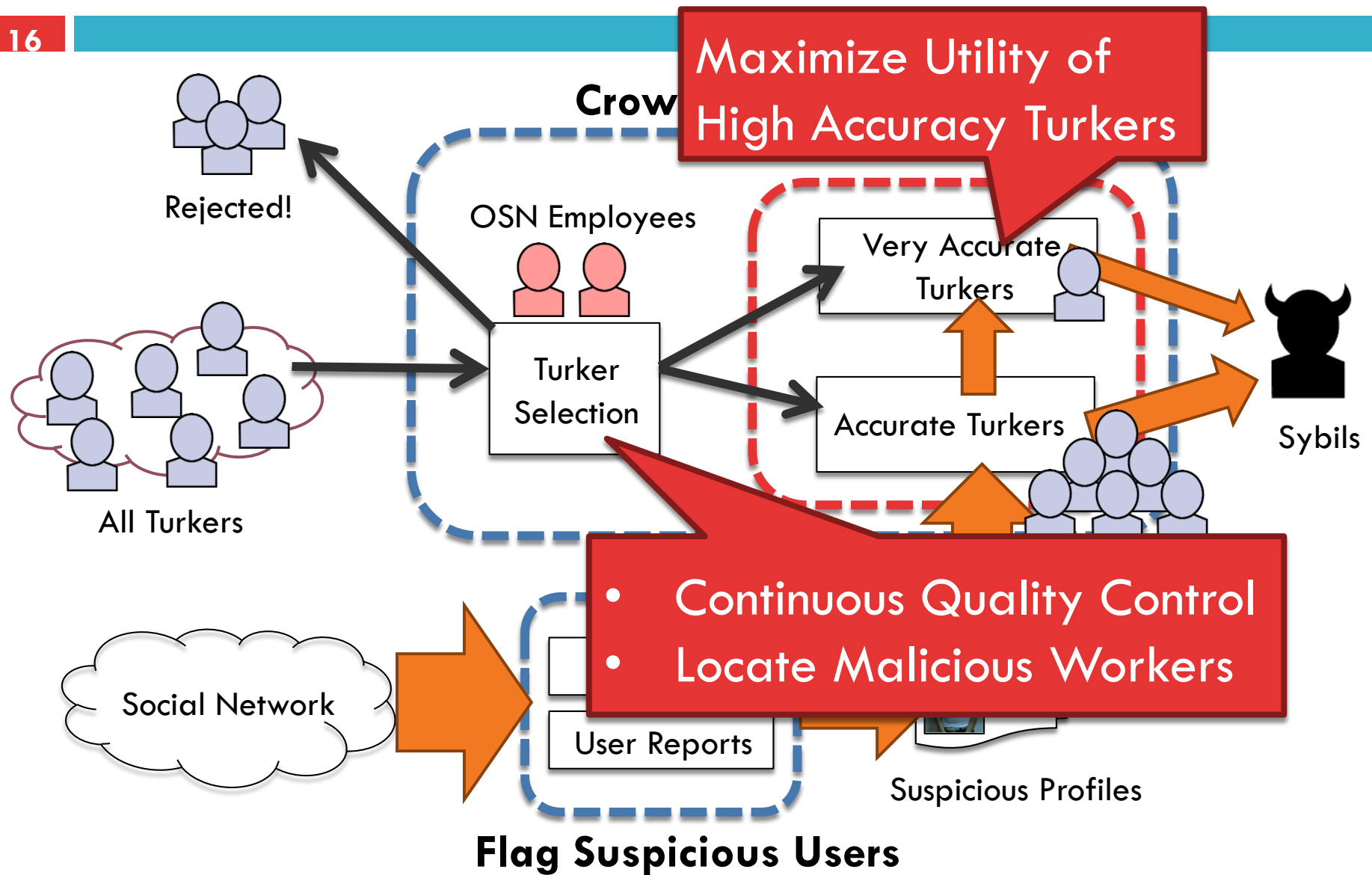
Key insight to designing our system

- Accuracy in turker population highly skewed
- Only 10% turkers $>$ 90% accurate



System Architecture

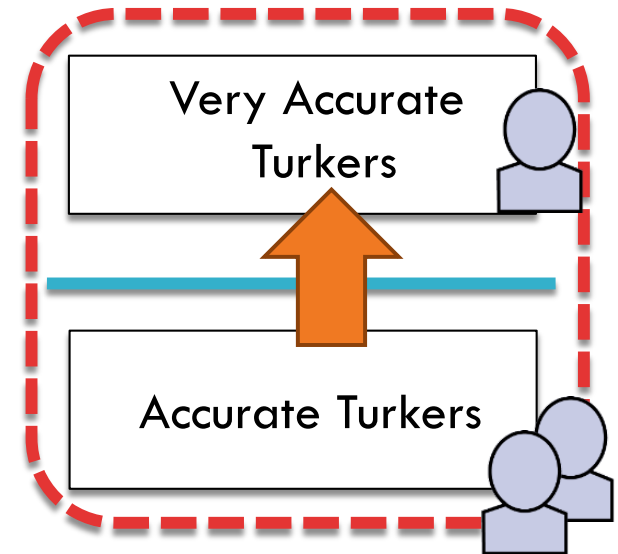
16



Trace Driven Simulations

17

- ▣ 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 8 votes per profile
- <0.1% false positives
- <0.1% false negatives

Estimating Cost

18

Estimated cost in a real-world social networks: Tuenti

- 12,000 profiles to verify daily
- 14 full-time employees



Cost with malicious turkers

- 25% of turkers are malicious
 - \$504 per day
- 20sec/profile, 8 hour day → 10 turkers (per hour) → \$2240 per day
 - Facebook wage (\$1 per hour) → \$400 per day

Augment existing automated systems

Conclusion

19

- 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



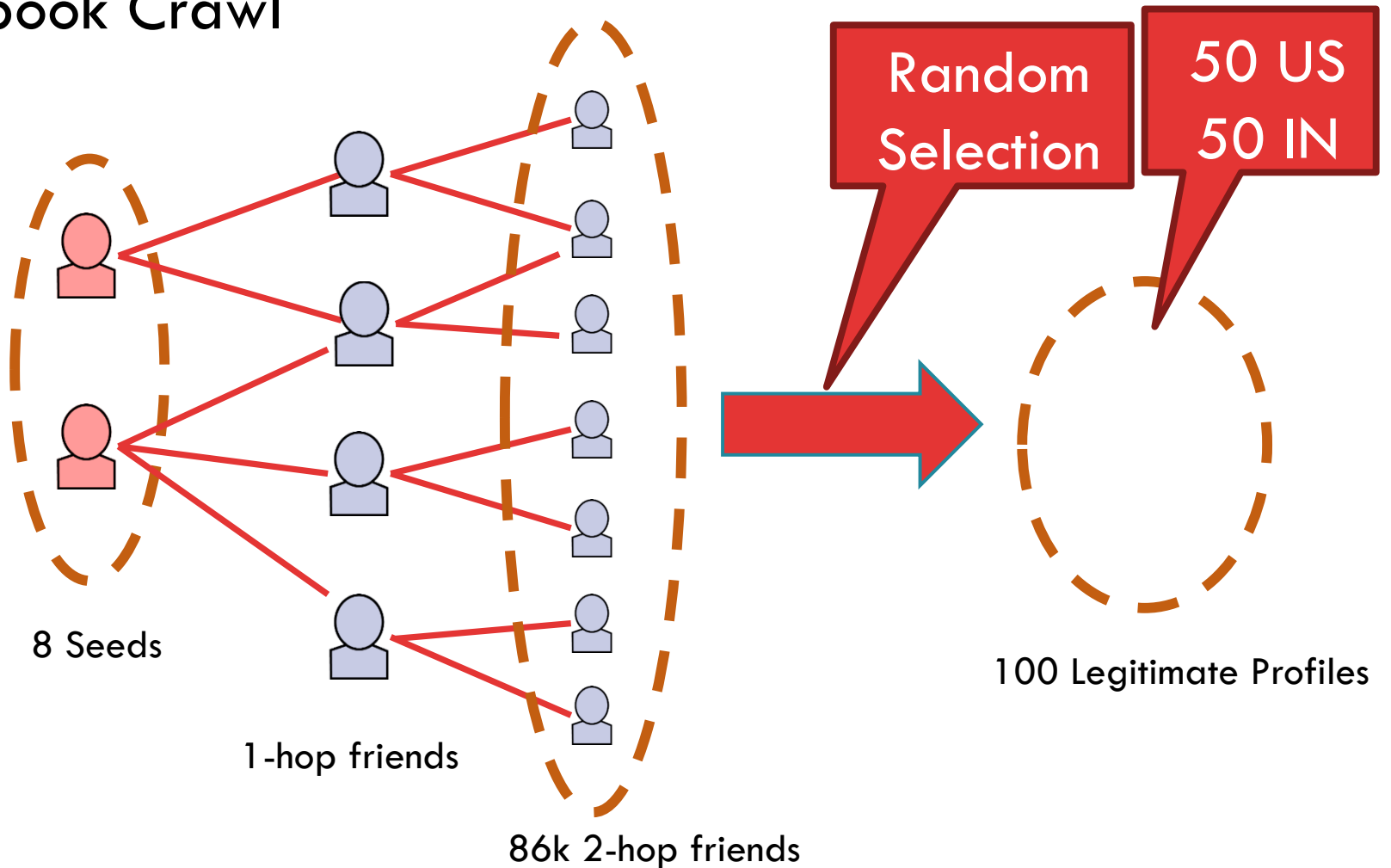


Thank you!

Ground-truth Data Collection (Legit.)

21

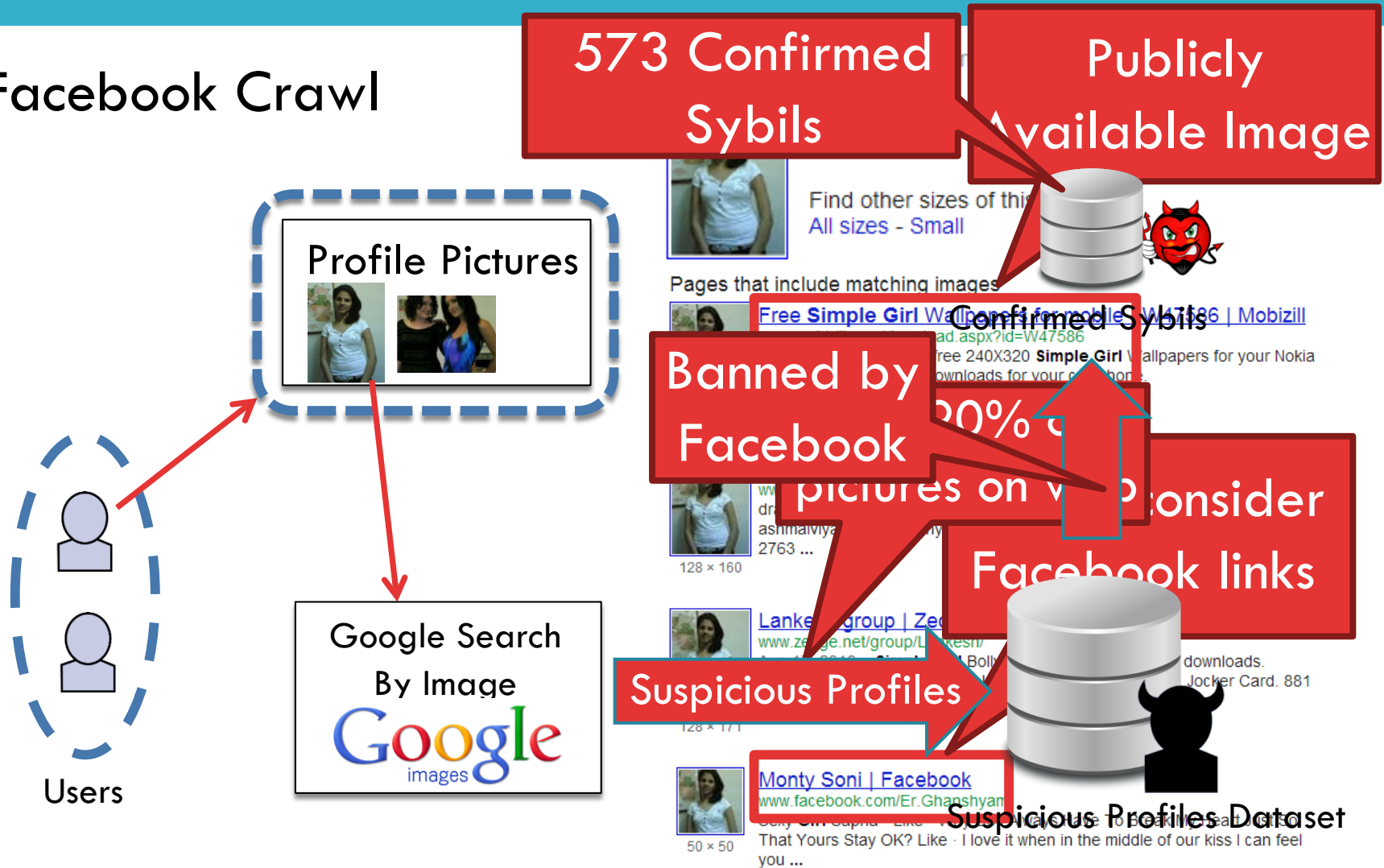
Facebook Crawl



Ground-truth Data Collection (Sybil)

22

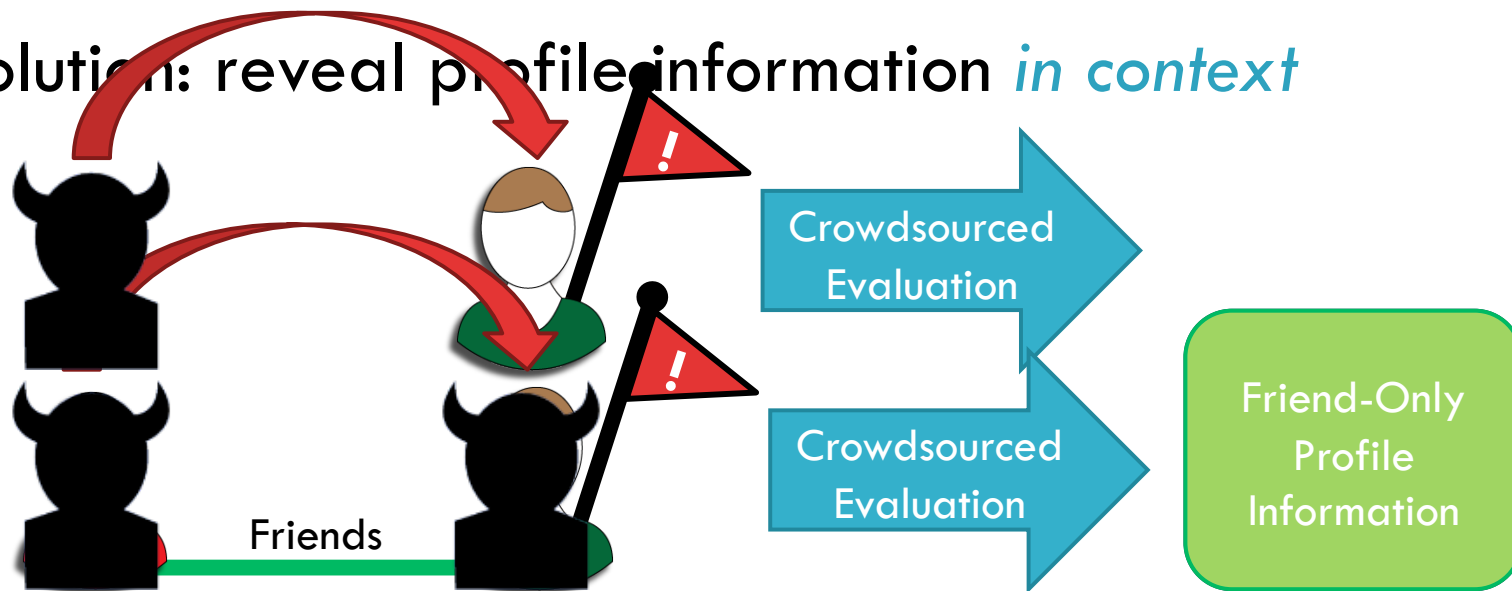
Facebook Crawl



Preserving User Privacy

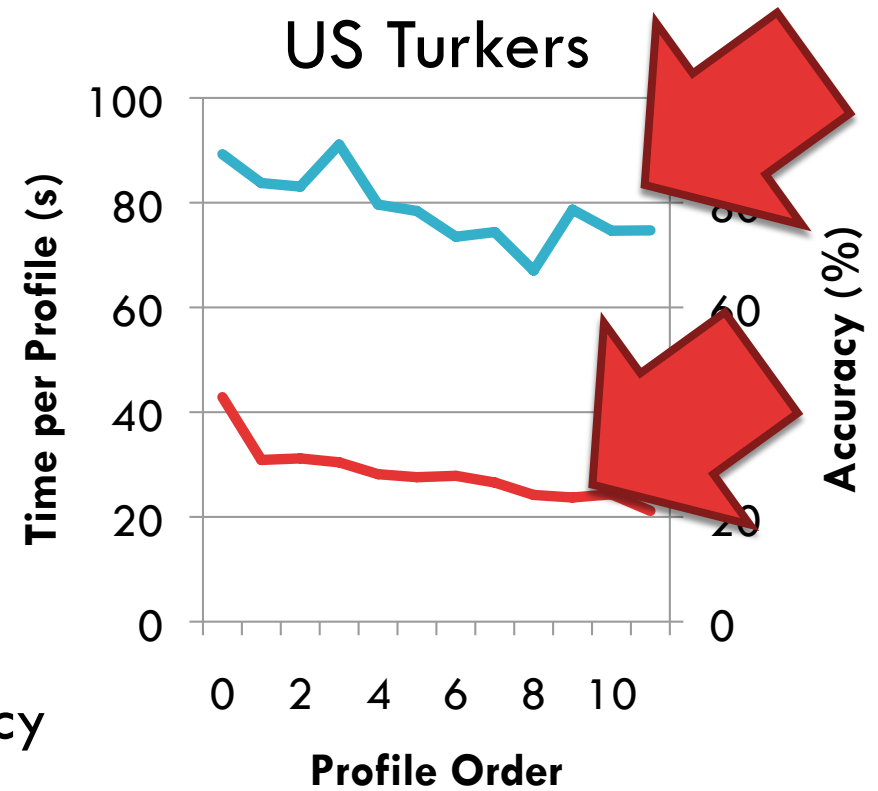
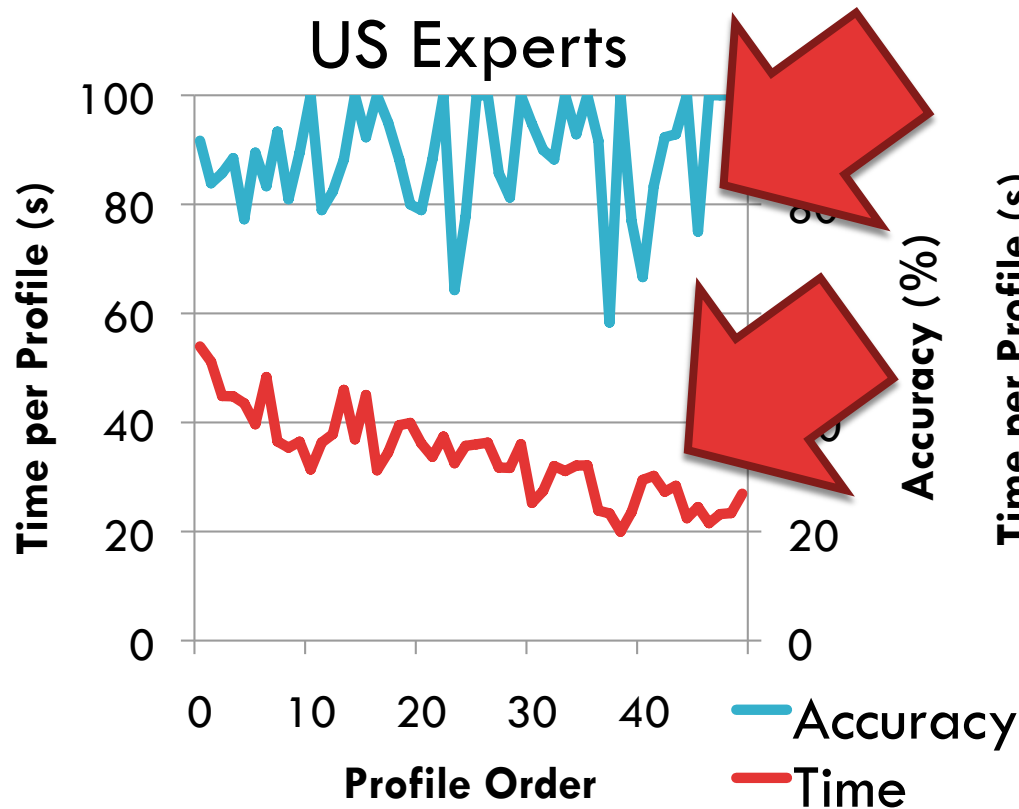
23

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



Survey Fatigue

24



No fa All testers speed up over time matters

Wisdom of the Crowd

25

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

Almost Zero

False Pos

Experts

False
Negatives

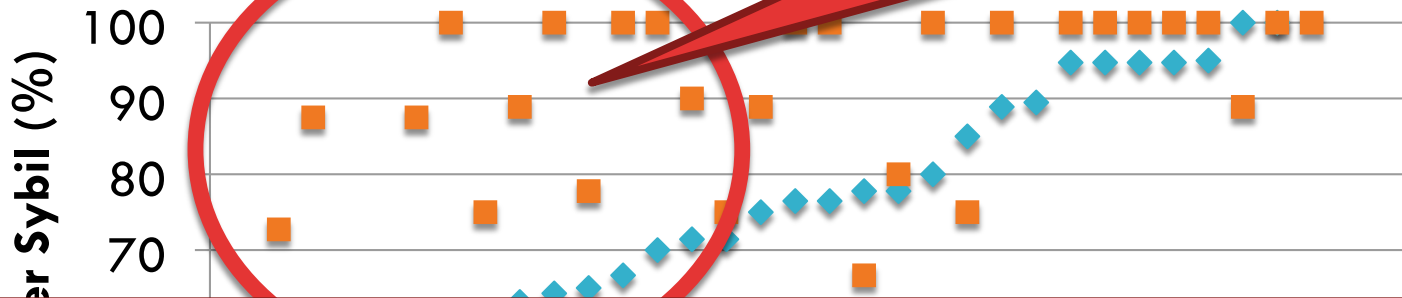
- False positive rates are excellent
- Turkers need extra help against false negatives
- What can be done to improve accuracy?

Dataset	Experts	False Pos	False Negatives
Facebook	India Expert	0%	10%
India	India Turker	0%	50%

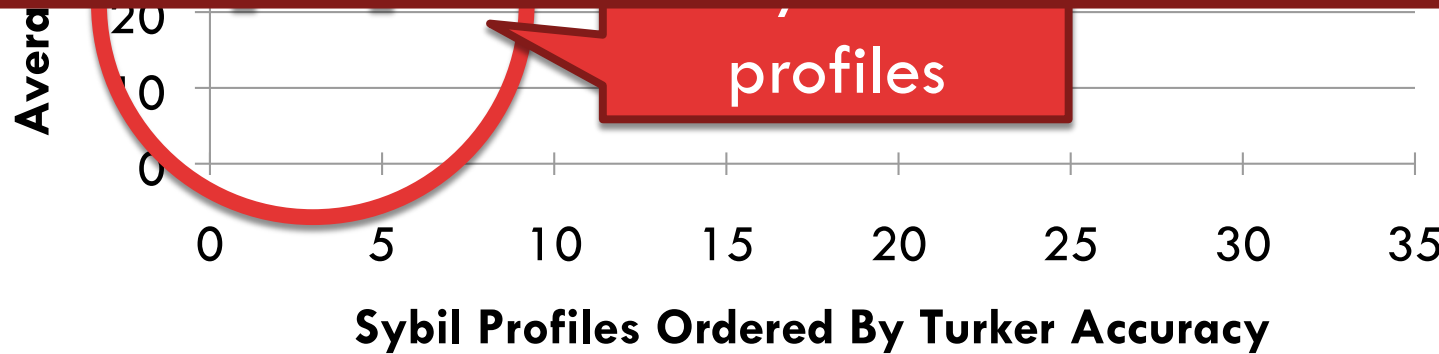
Sybil Profile Difficulty

26

Experts perform well on most difficult Sybils



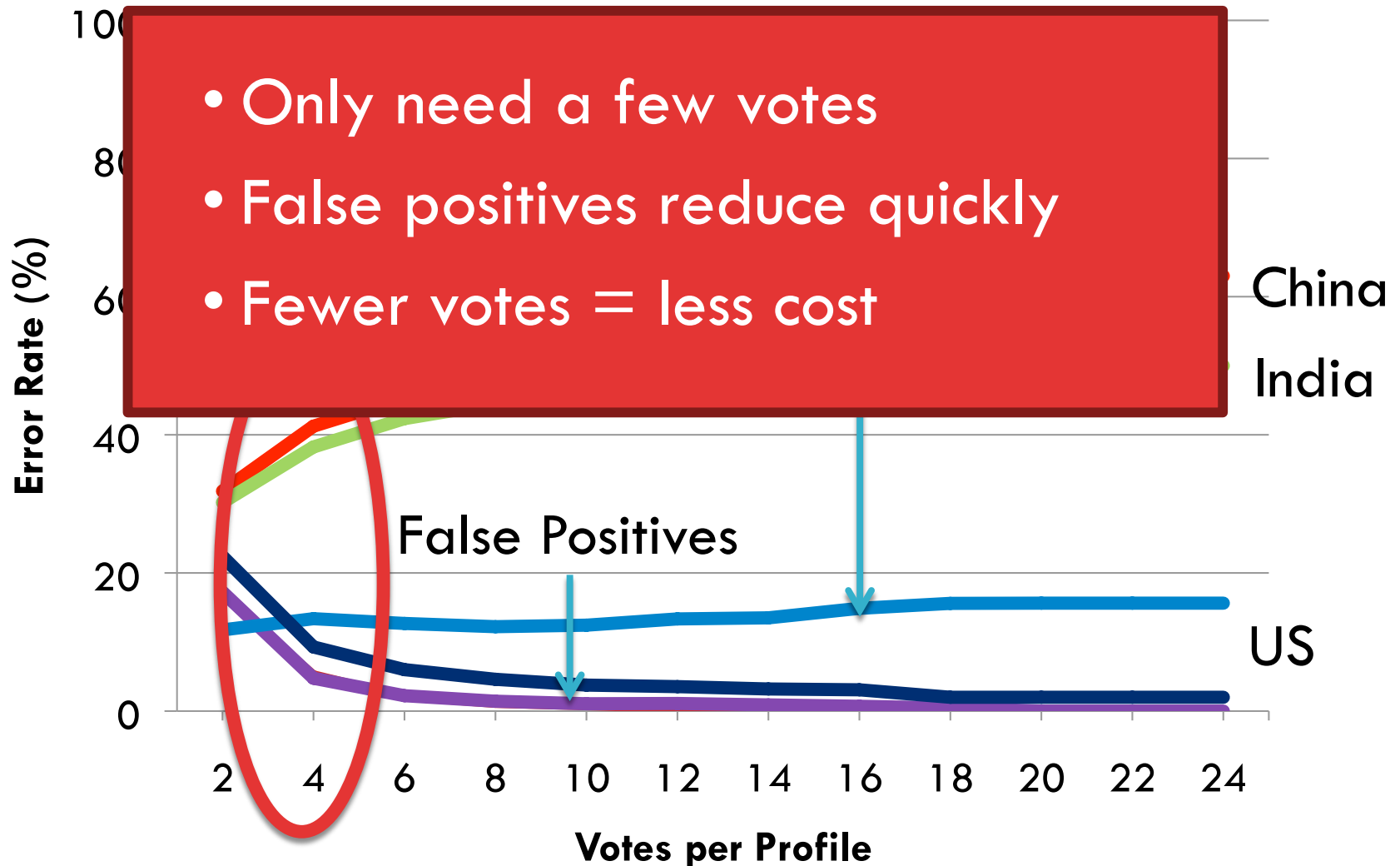
- Some Sybils are more stealthy
- Experts catch more tough Sybils than turkers



profiles

How Many Votes Do You Need?

27



Individual Tester Accuracy

28

