Evaluating the Wisdom of Crowds in Assessing Phishing Websites

Tyler Moore and Richard Clayton

University of Cambridge
Computer Laboratory

12th International Financial Cryptography and Data Security Conference (FC08)
Jan. 28, 2008, Cozumel Mexico

UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

Outline

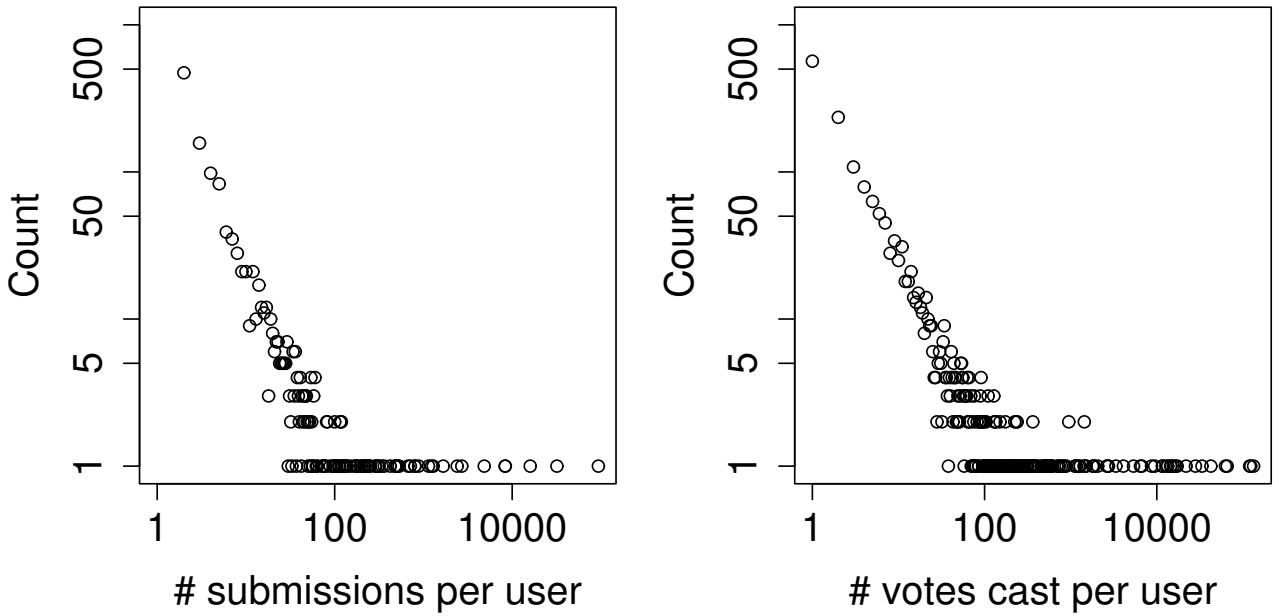
UNIVERSITY OF CAMBRIDGE

PhishTank

- Online community established in 2006 using the 'wisdom of crowds' to fight phishing
- Users contribute in two ways
  1. Submit reports of suspected phishing sites
  2. Vote on whether others' submissions are really phishing or not
- Data collection
  - We examined reports from 200 908 phishing URLs submitted between February and September 2007
  - For 24 254 reports, the site was removed before voting was completed, leaving 176 366 complete submissions
  - 3 798 users participated, casting 881 511 votes
  - $\Longrightarrow$ 53 submissions and 232 votes per user. But . . .

UNIVERSITY OF CAMBRIDGE

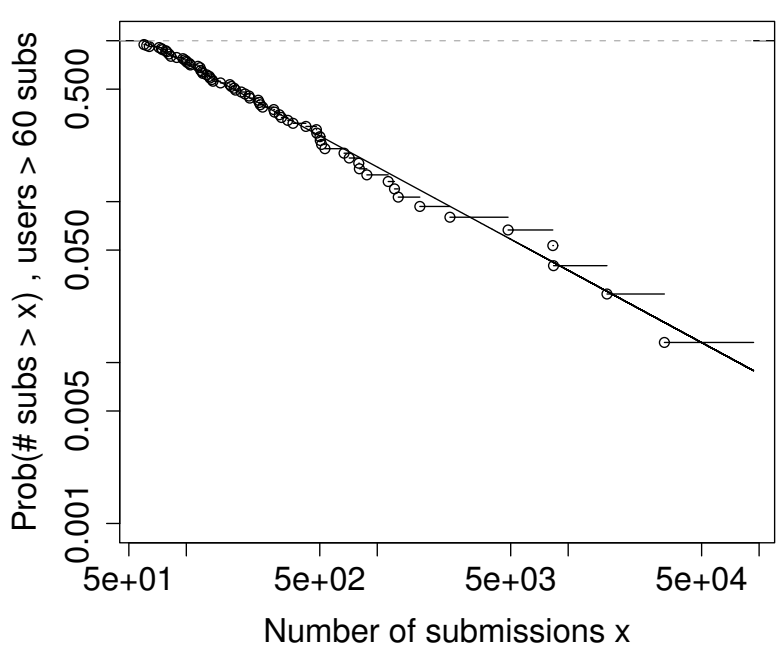
Density of user submissions and votes

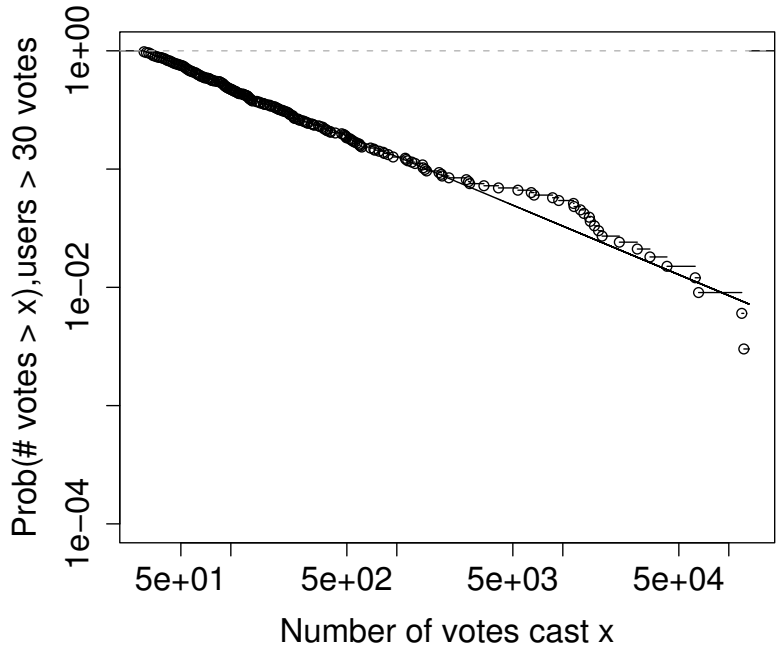


- Top two submitters (93 588 and 31 910) are anti-phishing organizations
- Some leading voters are PhishTank moderators – the 25 moderators cast 74% of votes

UNIVERSITY OF CAMBRIDGE

User participation in PhishTank follows power law





	Power-law dist.		Kolmogorov-Smirnov	
	α	x_{min}	D	p-value
Submissions	1.642	60	0.0533	0.9833
Votes	1.646	30	0.0368	0.7608

UNIVERSITY OF CAMBRIDGE

User participation in PhishTank follows power law

- What does a power-law distribution mean in this context?
  - A few highly-active users carry the load
  - Most users participate very little, but their aggregated contribution is substantial
- Why do we care?
  - Power-law distributions appear often in real-world contexts, including many types of social interaction
  - This suggests skewed participation naturally occurs for crowd-sourced applications
  - Power laws invalidate Byzantine fault tolerance – subverting one highly active participant can undermine system

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks and duplicate submissions to PhishTank

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (`lof80.info`)
  2. Send out phishing email with URL `http://www.volksbank.de.netw.oid3614061.lof80.info/vr`
  3. Gang-hosted DNS server resolves domain to IP addresses of compromised machines that proxy to a back-end server
- Wildcard DNS confuses phishing-report collators
  - 120 662 PhishTank reports (60% of all submissions)
  - Reduces to just 3 260 unique domains
  - 893 users voted 550 851 times on these domains, wasting users' resources that could be focused elsewhere

UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

Miscategorization in PhishTank

- Nearly all submitted URLs are verified as phishing – only 3% are voted down as invalid
- Many 'invalid' URLs are still dubious – 419 scams, malware hosts, mule-recruitment sites
- Even moderators sometimes get it wrong – 1.2% of their submissions are voted down
- PhishTank rewrites history when it is wrong, so we could identify 39 false positives and 3 false negatives
  - False positives include real institutions: `ebay.com`, `ebay.de`, `53.com`, `nationalcity.com`
  - False negatives include a rock-phish domain already voted down previously

UNIVERSITY OF CAMBRIDGE

Does experience improve user accuracy?

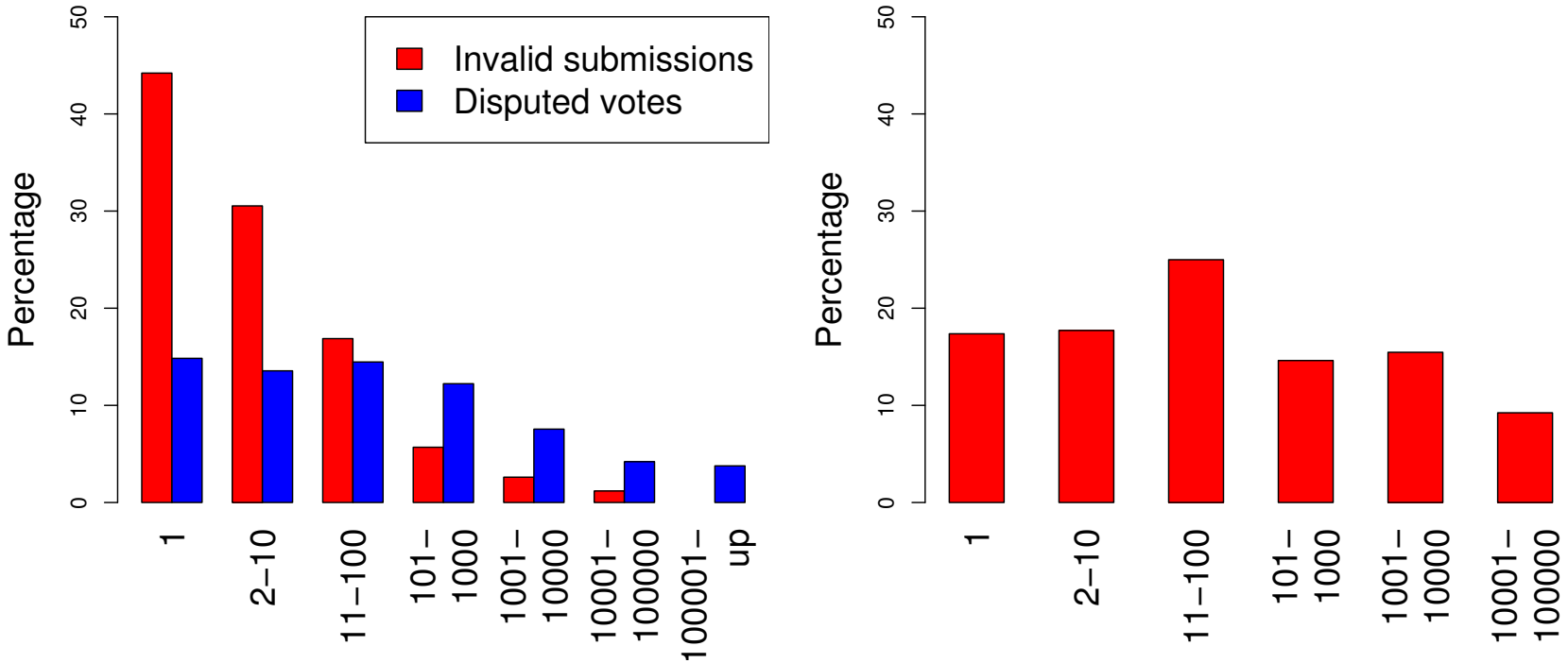


Figure: Inaccuracy of user submissions and votes according to the total number of submissions and votes per user, respectively (left). Proportion of all invalid user submissions grouped by number of submissions (right).

UNIVERSITY OF CAMBRIDGE

Do users with bad voting records vote together?

- High-conflict users: HC=93 users where most votes are bad
- Empirical measure of voting overlap

$$\text{overlap}(HC) = \sum_{A \in HC} \sum_{B \in HC, B \neq A} |V_A \cap V_B| = 254$$

- Expected overlap if relationship between users is random (thanks to Jaeyeon Jung)

$$E(\text{overlap}) = \sum_{A \in HC} \sum_{B \in HC, B \neq A} \sum_{i=1}^{min(|V_A|,|V_B|)} i \times \frac{\binom{|V_A|}{i} \times \binom{|T|-|V_A|}{|V_B|-i}}{\binom{|T|}{|V_B|}} = 0.225$$

UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

Disrupting PhishTank's verification system

- Can PhishTank's open submission and voting policies be exploited by attackers?
- Other anti-phishing groups have been targeted by DDoS attacks
- Attacks on PhishTank
  1. Submitting invalid reports accusing legitimate websites.
  2. Voting legitimate websites as phish.
  3. Voting illegitimate websites as not-phish.

  - Selfish attacker protects her own phishing websites by voting down any accusatory report as invalid
  - Undermining attacker goes after PhishTank's credibility by launching attacks 1&2 repeatedly

UNIVERSITY OF CAMBRIDGE

Simple countermeasures don't work

1. Place upper limit on the votes/submissions from a single user
   - Power-law distribution of participation means that restrictions would undermine the hardest-working users
   - Sybil attacks
2. Require users to participate correctly n times before counting contribution
   - PhishTank developers tell us they implement this countermeasure
   - Since 97% of submissions are valid, attacker can quickly build up reputation by voting 'is-phish' repeatedly – there is no honor among thieves
   - Savvy attacker can minimize positive contribution by only voting for rock-phish URLs

UNIVERSITY OF CAMBRIDGE

Simple countermeasures don't work (cont'd.)

3. Ignore any user with more than n invalid submissions/votes
   - Power-law distribution of participation means that good users make many mistakes
   - One top valid submitter, *antiphishing*, also has the most invalid submissions (578)
4. Ignore any user with more than x% invalid submissions/votes
   - Power law still causes problems – attackers can pad their 'good' statistics to also do bad
   - Significant collateral damage – ignoring users with $> 5\%$ bad submissions wipes out 44% of users and 5% of phishing URLs
5. Use moderators exclusively if suspect an attack
   - Moderators already cast 74% of votes, so it might work OK
   - Silencing the whole crowd to root out attackers is intellectually unsatisfying, though

UNIVERSITY OF CAMBRIDGE

Lessons for secure crowd-sourcing

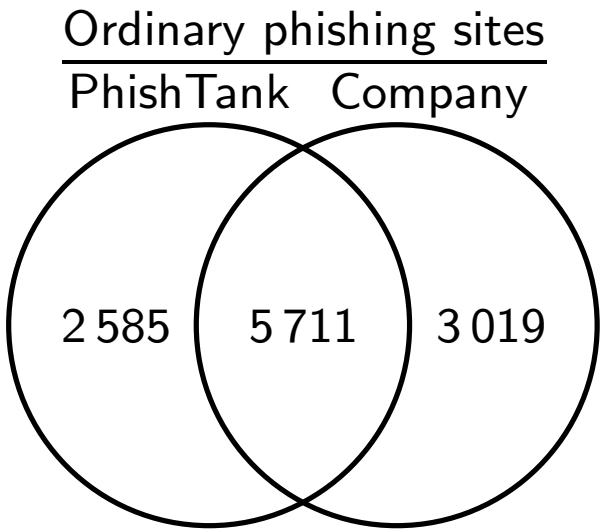
1. *The distribution of user participation matters*
   - Skewed distributions such as power laws are a natural consequence of user participation
   - Corrupting a few key users can undermine system security
   - Since good users can participate extensively, bad users can too
2. *Crowd-sourced decisions should be difficult to guess*
   - Any decision that can be reliably guessed can be automated and exploited by an attacker
   - Underlying accuracy of PhishTank (97% phish) makes boosting reputation by guessing easy
3. *Do not make users work harder than necessary*
   - Requiring users to vote multiple times for rock-phish is a bad use of the crowd's intelligence

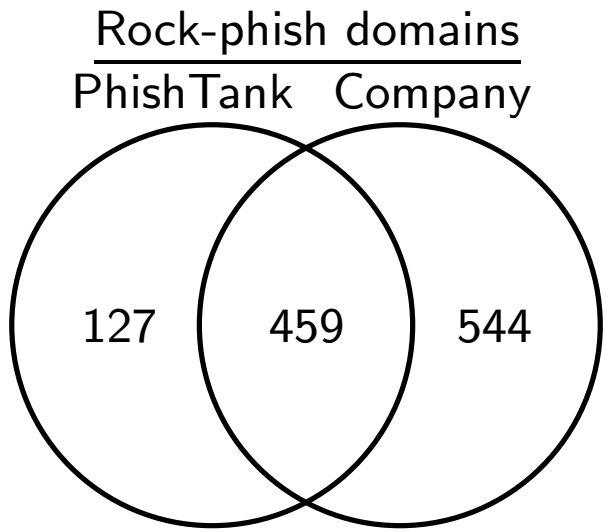
UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

PhishTank's open feed vs. company's closed feed





- Verdict
  - PhishTank and the company's feeds are similar for ordinary sites, but the company is much more comprehensive on rock-phish
  - Both have significant gaps in coverage, which motivates sharing feeds

UNIVERSITY OF CAMBRIDGE

Verification speed: PhishTank vs. company

- Voting introduces significant delays to verification
  - 46 hr average delay (15 hr median)
  - Company, by contrast, uses employees to verify immediately
  - Impact can be seen by examining sites reported to both feeds

ΔPhishTank − Company	Ordinary phishing URLs		Rock-phish domains	
	Submission	Verification	Submission	Verification
Mean (hrs)	−0.188	15.9	12.4	24.7
Median (hrs)	−0.0481	10.9	9.37	20.8

UNIVERSITY OF CAMBRIDGE

Conclusions

- While leveraging the wisdom of crowds sounds appealing, it may not always be appropriate for information security tasks
- After examining one such effort, we found its decisions to be mostly accurate but vulnerable to manipulation
- Compared to a similar proprietary effort, PhishTank is less complete and less timely
- Moving forward, user participation as input to security mechanisms should be treated with caution
- For more, see `http://www.cl.cam.ac.uk/~twm29/` and `http://www.lightbluetouchpaper.org/`

UNIVERSITY OF CAMBRIDGE