An Empirical Analysis of Phishing Attack and Defense

Tyler Moore and Richard Clayton

University of Cambridge
Computer Laboratory

Computer Lab Security Seminar
April 8, 2008

UNIVERSITY OF CAMBRIDGE

Outline

1. Who's winning the phishing arm's race?
   - The mechanics of phishing
   - Rock-phish attacks
   - Phishing-website lifetimes
2. Non-cooperation when countering phishing
   - Comparing lifetimes for different feeds
   - Estimating the cost of phishing attacks
3. Evaluating the 'wisdom' of PhishTank's crowd
   - PhishTank vs. proprietary feeds
   - User participation in PhishTank
   - Disrupting PhishTank's verification system

UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

Technical requirements for phishing attacks

- Attackers send out spam impersonating banks with link to fake website
- Hosting options for fake website
  - Free webspace (`http://www.bankname.freespacesitename.com/signin/`)
  - Compromised machine (`http://www.example.com/~user/images/www.bankname.com/`)
  - Registered domain (`bankname-variant.com`) which then points to free webspace or compromised machine
- Personal detail recovery
  - Completed forms forwarded to a webmail address
  - Stored in a text file on the spoof website

UNIVERSITY OF CAMBRIDGE

Defending against phishing attacks

- Proactive measures
  - Web browser mechanisms to detect fake sites, multi-factor authentication procedures, restricted top-level domains, etc.
  - Not the focus of our research
- Reactive measures
  - Banks tally phishing URLs
  - Reported phishing URLs are added to a blacklist, which is disseminated via anti-phishing toolbars
  - Banks send take-down requests to the free webspace operator or ISP of compromised machine
  - If a malicious domain has been registered, banks ask the domain name registrar to suspend the offending domain

UNIVERSITY OF CAMBRIDGE

Defending against phishing attacks

- Proactive measures
  - Web browser mechanisms to detect fake sites, multi-factor authentication procedures, restricted top-level domains, etc.
  - Not the focus of our research
- Reactive measures
  - Banks tally phishing URLs
  - Reported phishing URLs are added to a blacklist, which is disseminated via anti-phishing toolbars
  - Banks send take-down requests to the free webspace operator or ISP of compromised machine
  - If a malicious domain has been registered, banks ask the domain name registrar to suspend the offending domain

UNIVERSITY OF CAMBRIDGE

Data collection methodology

- Phishing website availability
  - Several organizations collate phishing reports; we selected reports from PhishTank
  - PhishTank DB records phishing URLs and relies on volunteers to confirm whether a site is wicked
  - 33 710 PhishTank reports overs 8 weeks early 2007
  - We constructed our own testing system to continuously query sites until they stop responding or change
- Caveats to our data collection
  - Sites removed before appearing in PhishTank are ignored
  - We do not follow web-page redirectors

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks are different!

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (e.g., `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 address of one of several compromised machines
  4. Compromised machines run a proxy to a back-end server
  5. Server loaded with many fake websites (around 20), all of which can be accessed from any domain or compromised machine

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks are different!

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (e.g., `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 address of one of several compromised machines
  4. Compromised machines run a proxy to a back-end server
  5. Server loaded with many fake websites (around 20), all of which can be accessed from any domain or compromised machine

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks are different!

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (e.g., `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 address of one of several compromised machines
  4. Compromised machines run a proxy to a back-end server
  5. Server loaded with many fake websites (around 20), all of which can be accessed from any domain or compromised machine

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks are different!

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (e.g., `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 address of one of several compromised machines
  4. Compromised machines run a proxy to a back-end server
  5. Server loaded with many fake websites (around 20), all of which can be accessed from any domain or compromised machine

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks are different!

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (e.g., `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 address of one of several compromised machines
  4. Compromised machines run a proxy to a back-end server
  5. Server loaded with many fake websites (around 20), all of which can be accessed from any domain or compromised machine

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks are different!

- 'Rock-phish' gang operate different to 'ordinary' phishing sites
  1. Purchase several innocuous-sounding domains (e.g., `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 address of one of several compromised machines
  4. Compromised machines run a proxy to a back-end server
  5. Server loaded with many fake websites (around 20), all of which can be accessed from any domain or compromised machine

UNIVERSITY OF CAMBRIDGE

Rock-phish attacks (cont'd.)

- Rock-phish strategy is more resilient to failure
  - Dynamic pool of domains maps to another pool of IP addresses
- Also increase confusion by splitting the attack components over disjoint authorities
  - Registrars see non-bank domains
  - Compromised machine owners don't see bank webpages

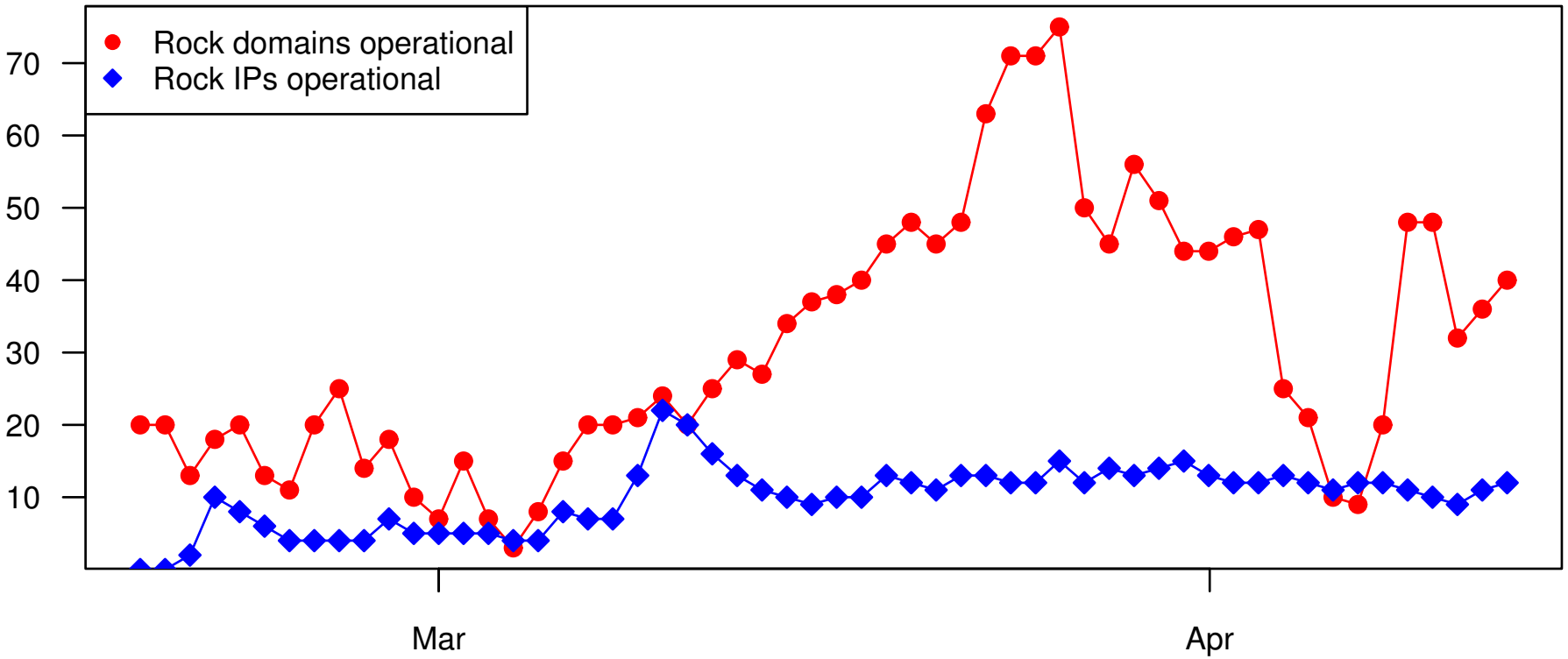
UNIVERSITY OF CAMBRIDGE

'Fast-flux' phishing domains

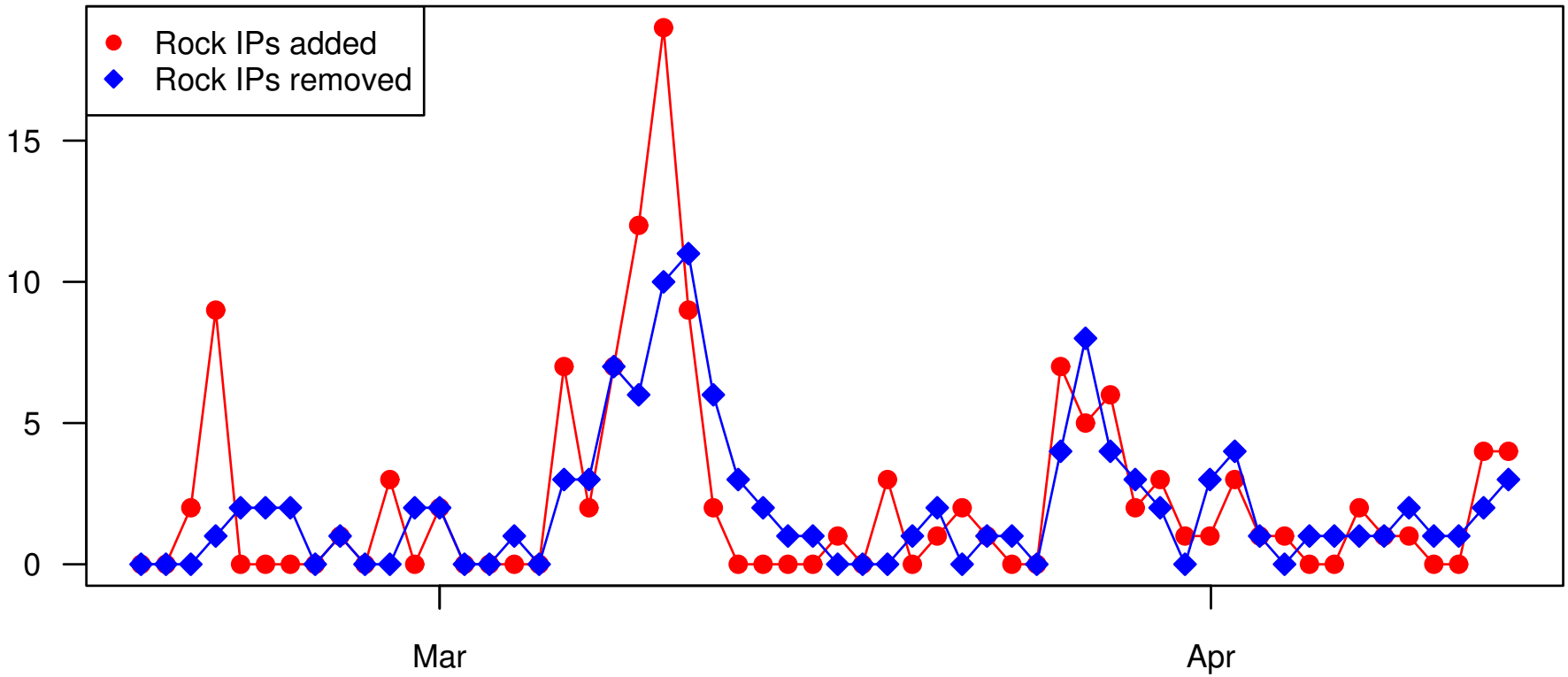
- Rock-phish gang's strategy is evolving fast
- In a fast-flux variant, domains resolve to a set of 5 IP addresses for a short time, then abandon them for another 5
- Burn through 400 IP addresses per week, but the upside (for the attacker) is that machine take-down becomes impractical
- Fast-flux strategy demonstrates just how cheap compromised machines are

UNIVERSITY OF CAMBRIDGE

Rock-phish site activity per day



New and removed rock-phish IPs per day

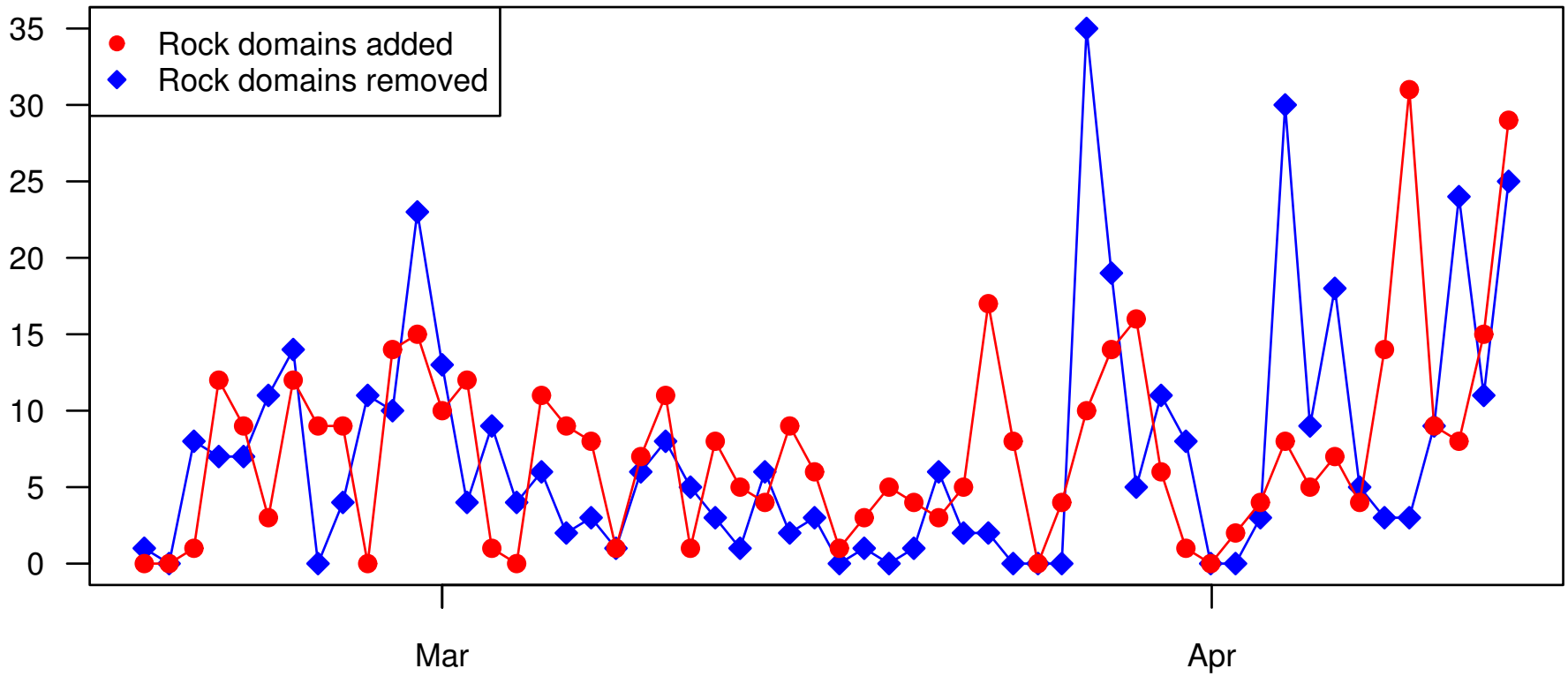


Correlation coefficient r: 0.740

Synchronized $\Longrightarrow$ automated replenishment

UNIVERSITY OF CAMBRIDGE

New and removed rock-phish domains per day

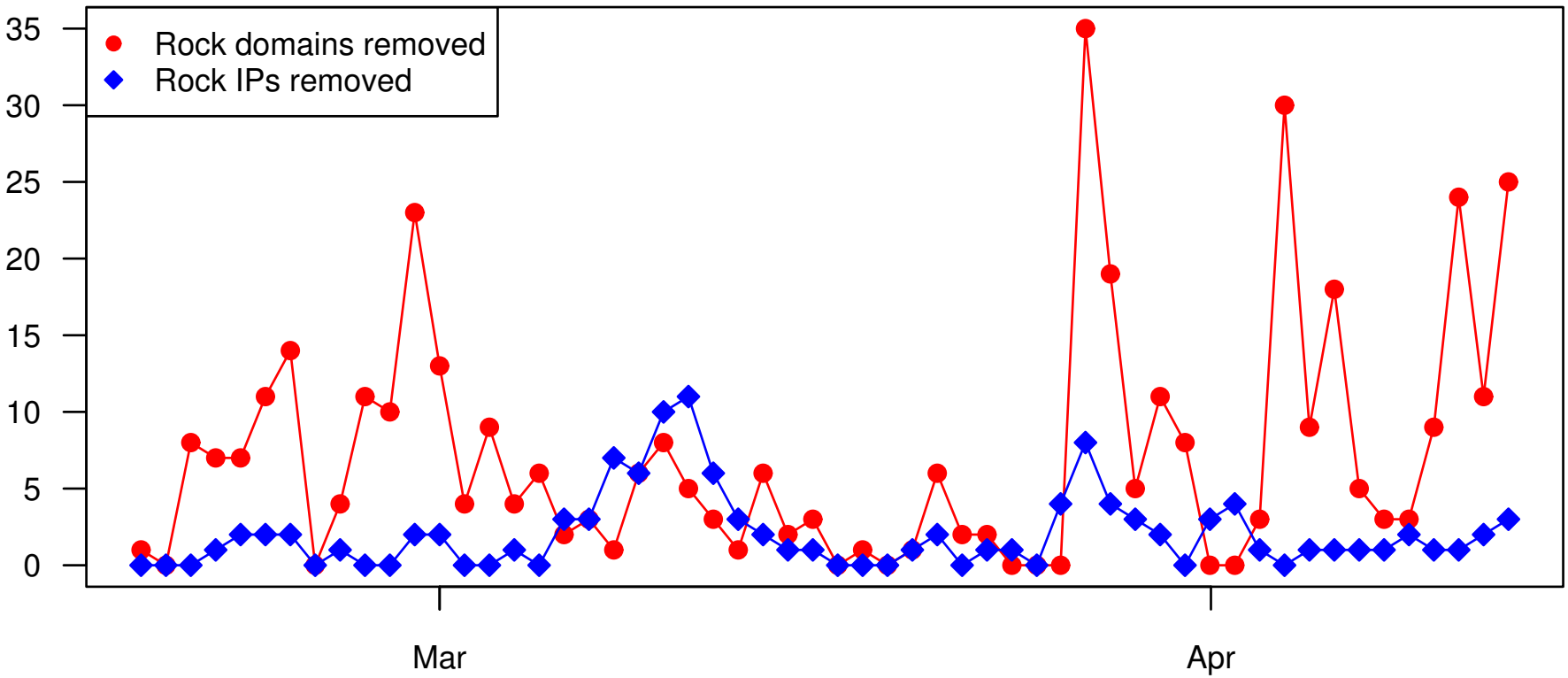


Correlation coefficient r: 0.340

Unsynchronized $\Longrightarrow$ manual replenishment

UNIVERSITY OF CAMBRIDGE

Rock-phish domain and IP removal per day



Correlation coefficient r: 0.142

Unsynchronized $\Longrightarrow$ uncoordinated response

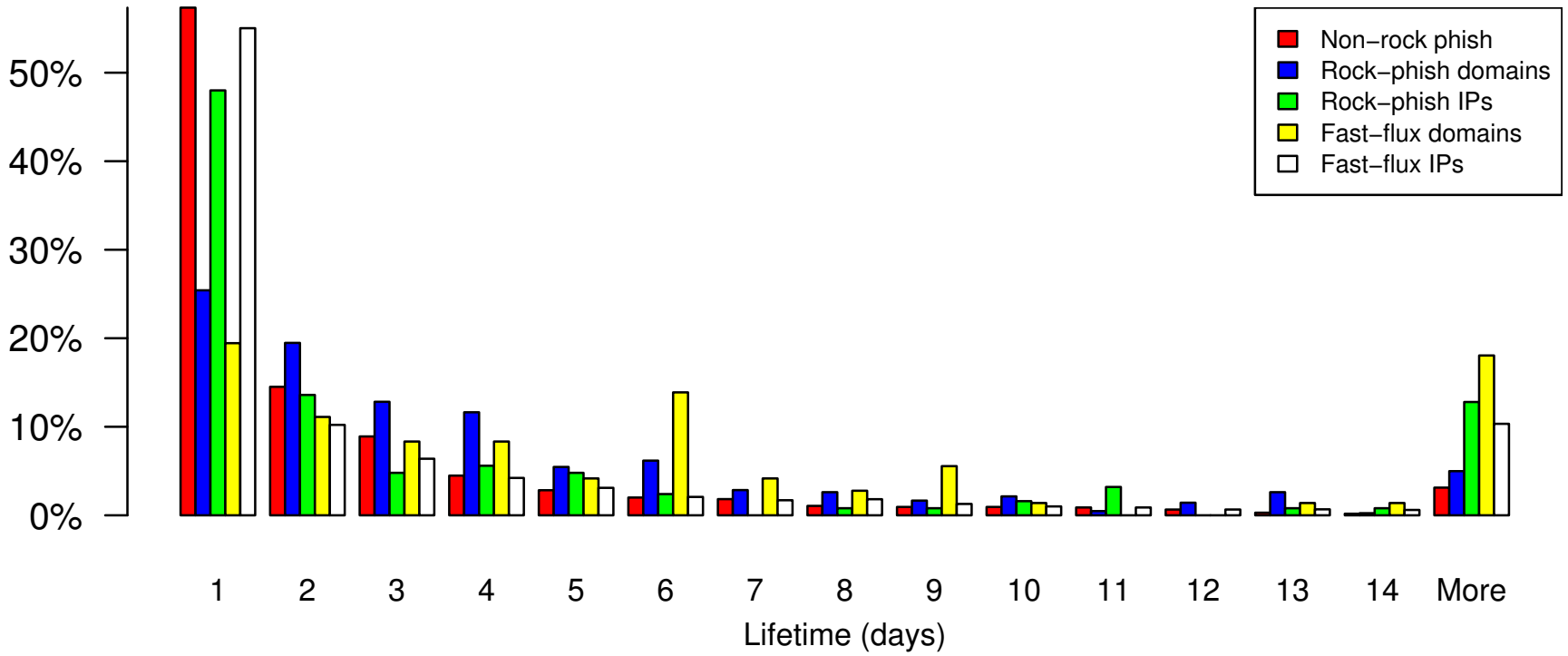
UNIVERSITY OF CAMBRIDGE

Phishing-website lifetimes

	Sites	Mean lifetime (hrs)	Median lifetime (hrs)
Non-rock	1 695	61.7	19.5
Rock domains	421	94.7	55.1
Rock IPs	125	171.8	25.5
Fast-flux domains	57	196.2	111.0
Fast-flux IPs	4 287	138.6	18.0

UNIVERSITY OF CAMBRIDGE

Histogram of phishing-site lifetimes



UNIVERSITY OF CAMBRIDGE

And now for some curve fitting

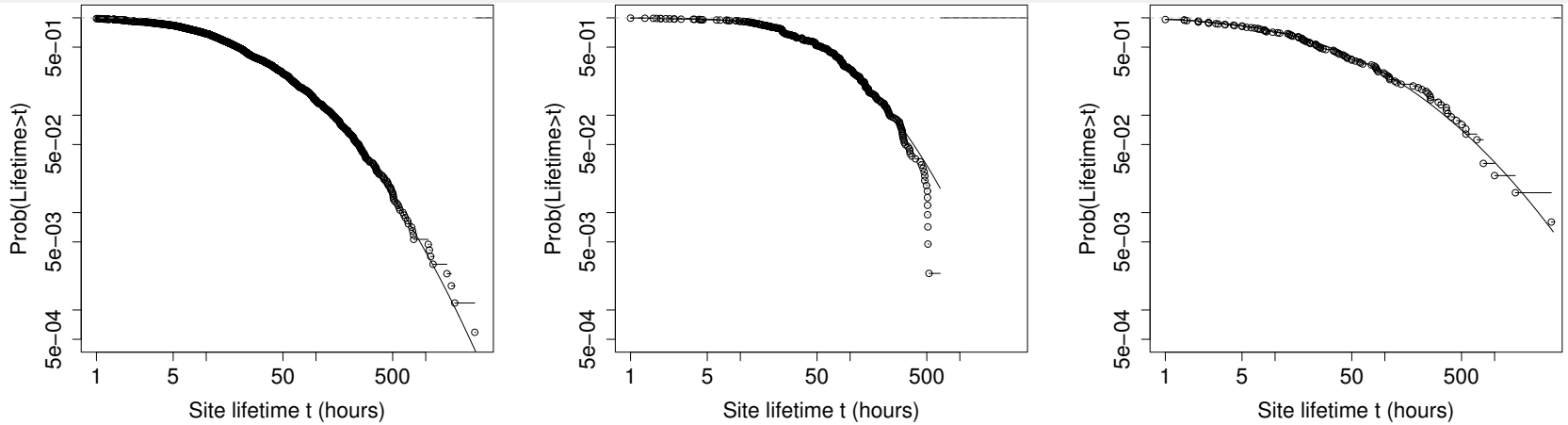


Figure: CDF of website lifetimes for non-rock (left), rock domains (center) and rock-phish IPs (right).

	Lognormal				Kolmogorov-Smirnov	
	μ	Std err.	σ	Std err.	D	p-value
Non-rock	3.011	0.03562	1.467	0.02518	0.03348	0.3781
Rock domains	3.922	0.05966	1.224	0.04219	0.06289	0.4374
Rock IPs	3.434	0.1689	1.888	0.1194	0.09078	0.6750

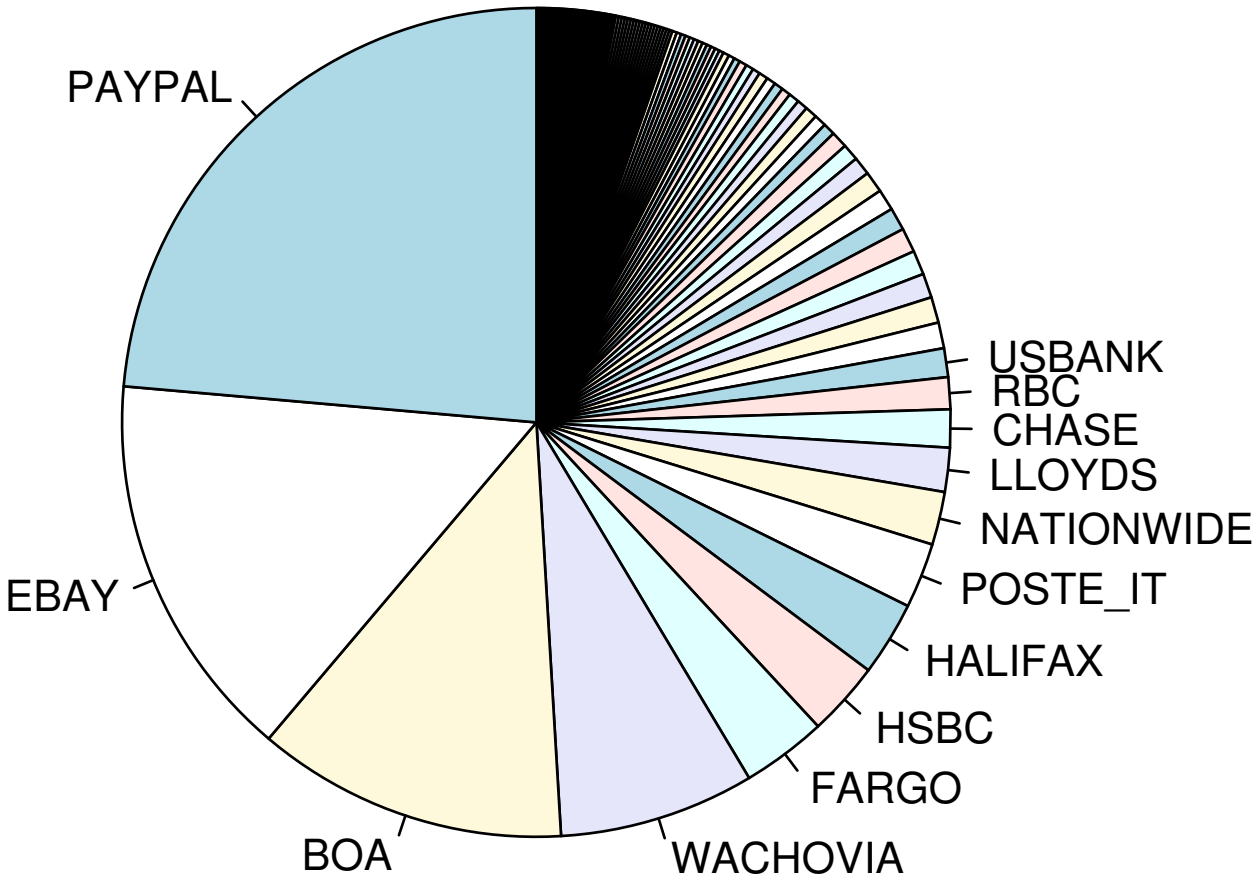
UNIVERSITY OF CAMBRIDGE

Breaking down site lifetimes

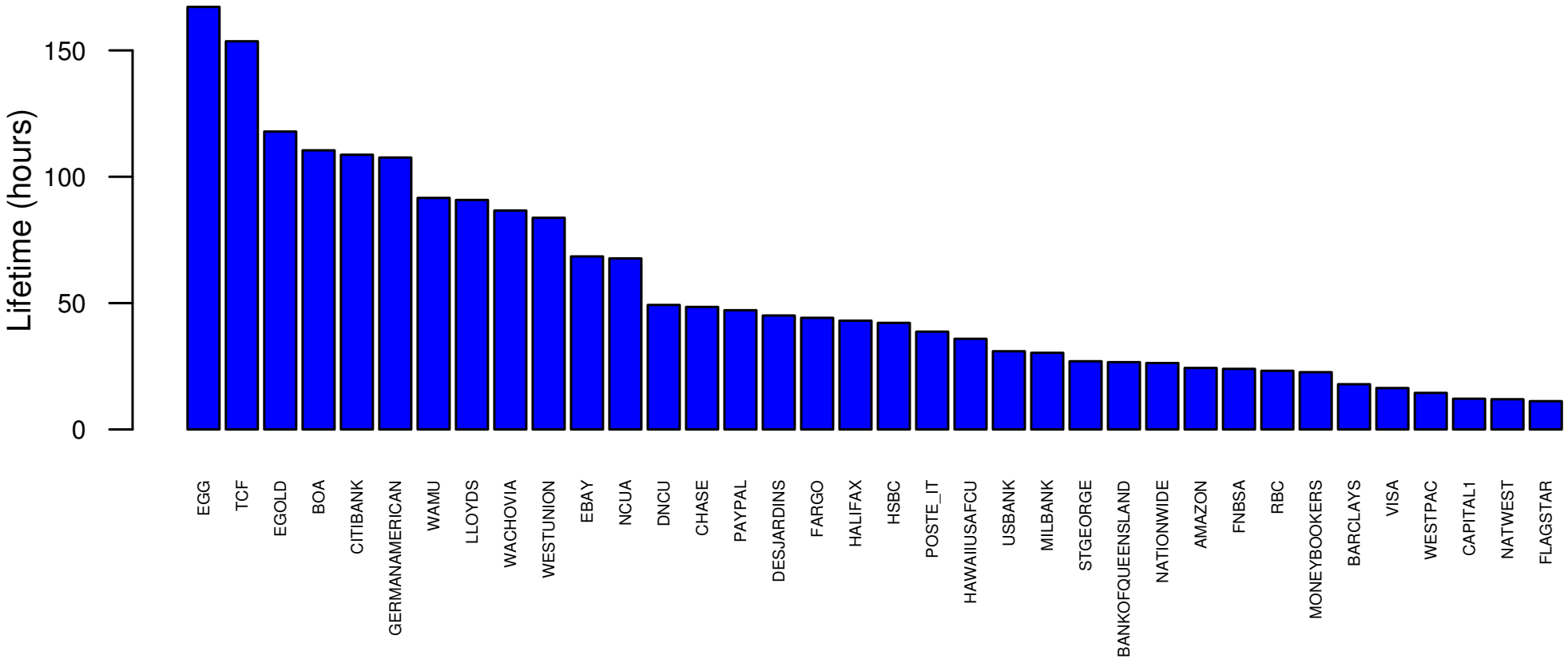
- Phishing site lifetimes vary greatly, but can we make sense of the differences?
  - We have already established that the rock-phish gang are more effective than other attackers
  - Do some banks perform better than others?
  - Do some ISPs respond better than others?
- Identifying exceptional performers (both good and bad) could encourage improved response times

UNIVERSITY OF CAMBRIDGE

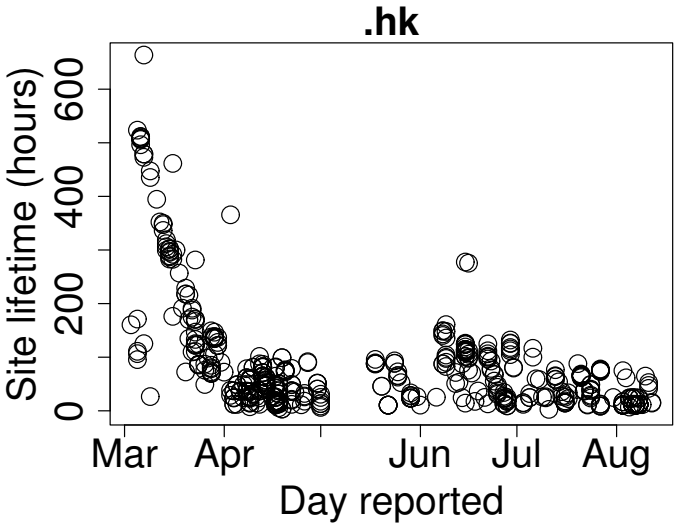
Number of phishing sites per bank

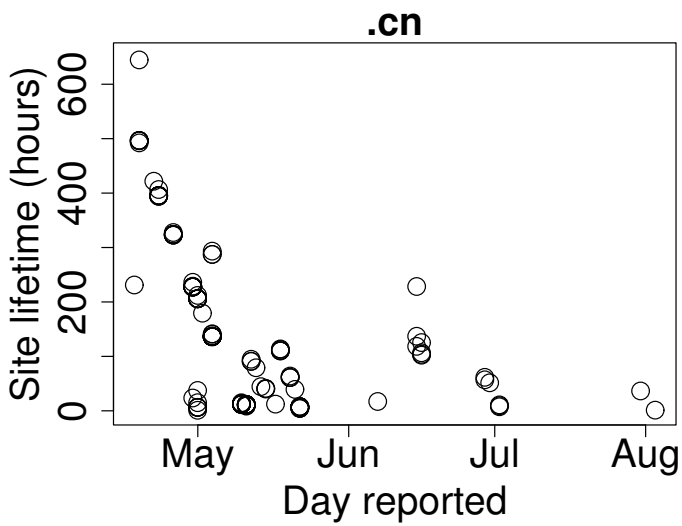


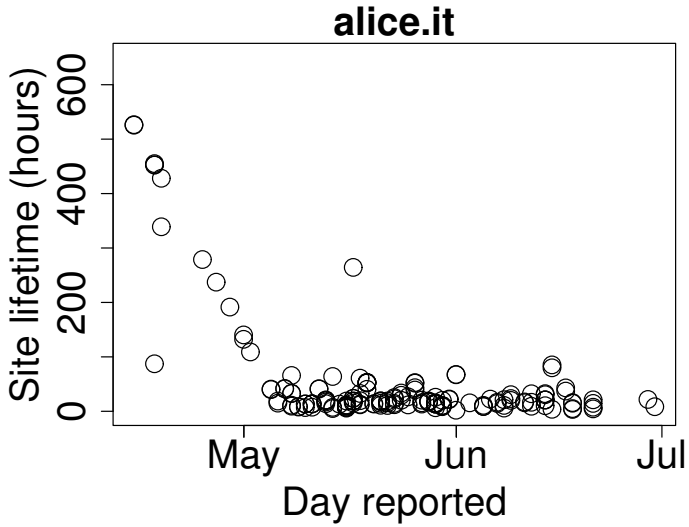
Phishing-site lifetimes per bank (only banks >= 5 sites)



'Clued-up' effect on free host & registrar take-down times







UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

Non-cooperation when countering phishing

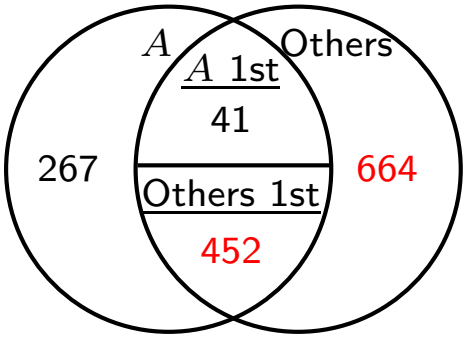
- The phishing-website lifetimes just presented are longer than those reported by banks and take-down companies
- We collected feeds of phishing URLs from two take-down companies, a brand owner, the Anti-Phishing Working Group and PhishTank
- Using this wider perspective, we can explain the disparity: websites unknown to the banks take much longer to be removed
- So we have examined the feeds from two take-down companies, called A and B, in greater detail during October–December 2007

UNIVERSITY OF CAMBRIDGE

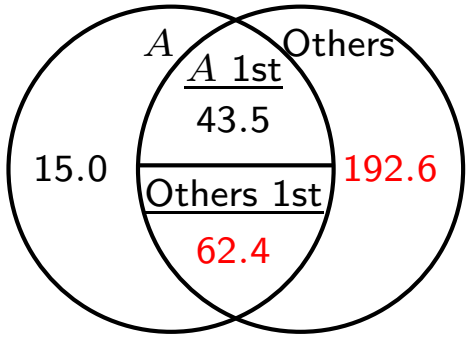
How one bank suffers when take-down companies don't share phishing URLs

A's client $A1$

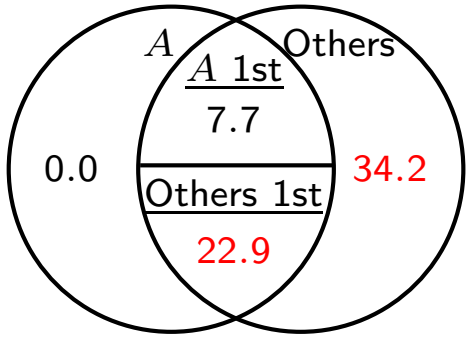
Ordinary phishing sites



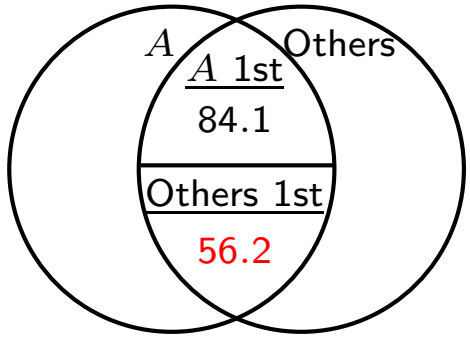
Mean lifetime (hours)



Median lifetime (hours)



Mean difference (hours)

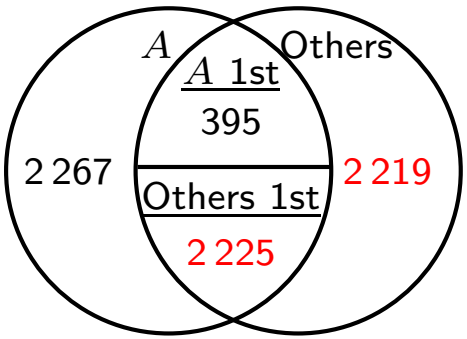


UNIVERSITY OF CAMBRIDGE

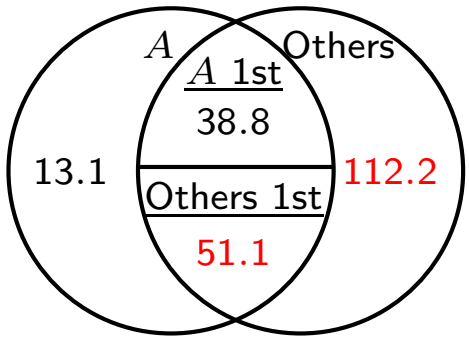
Most banks suffer when phishing URLs are not shared

A's 53 clients attacked during Q4 2007

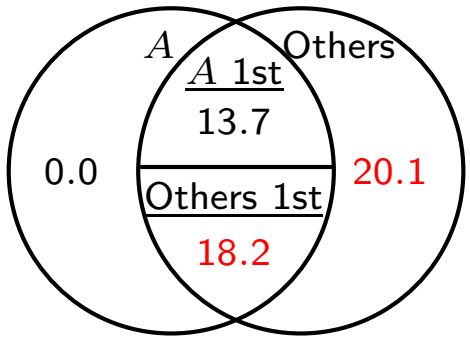
Ordinary phishing sites



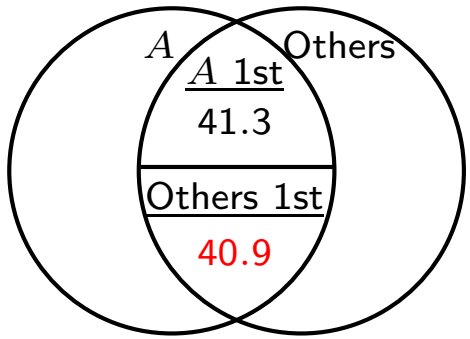
Mean lifetime (hours)



Median lifetime (hours)

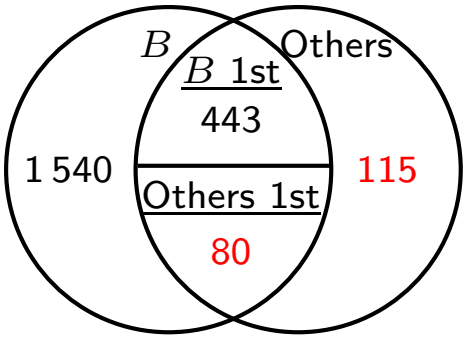


Mean difference (hours)

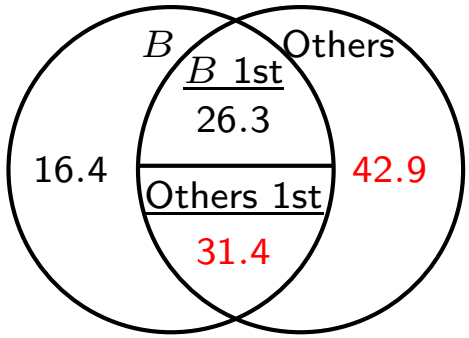


B's 66 clients attacked during Q4 2007

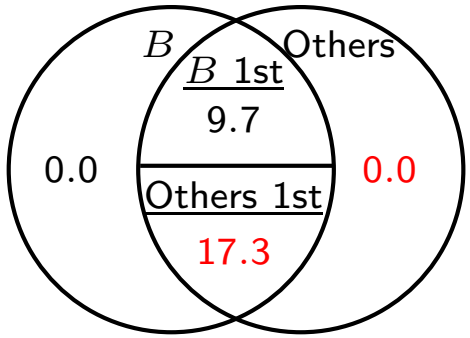
Ordinary phishing sites



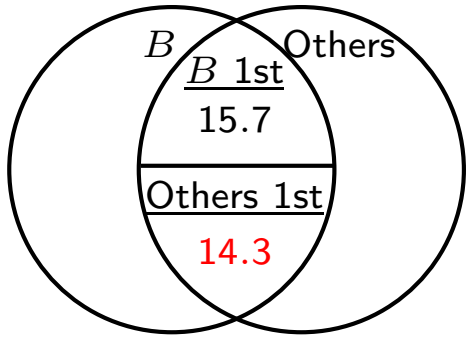
Mean lifetime (hours)



Median lifetime (hours)

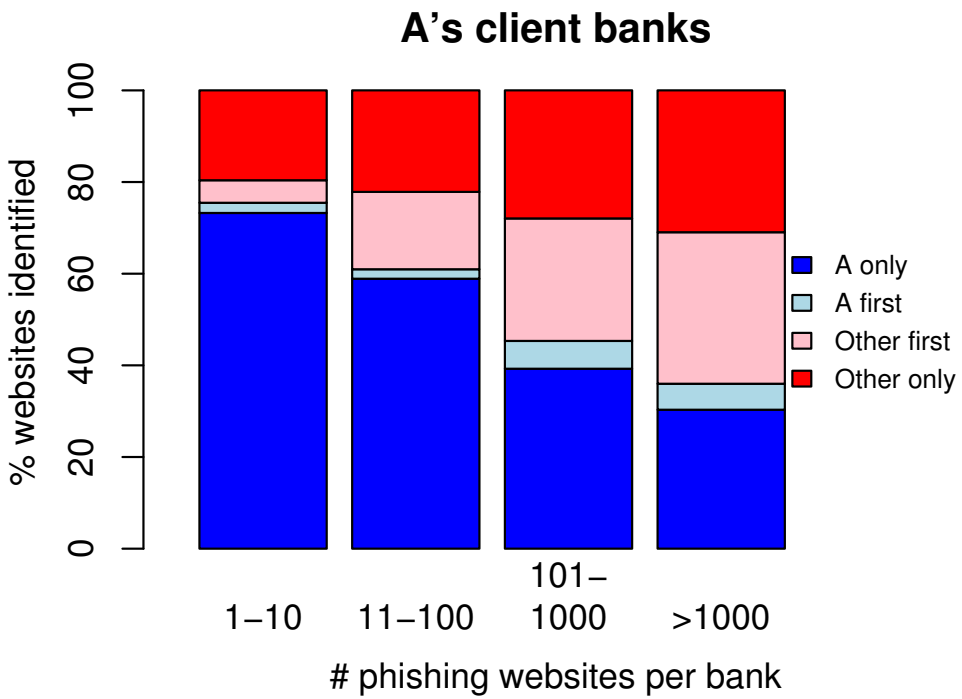


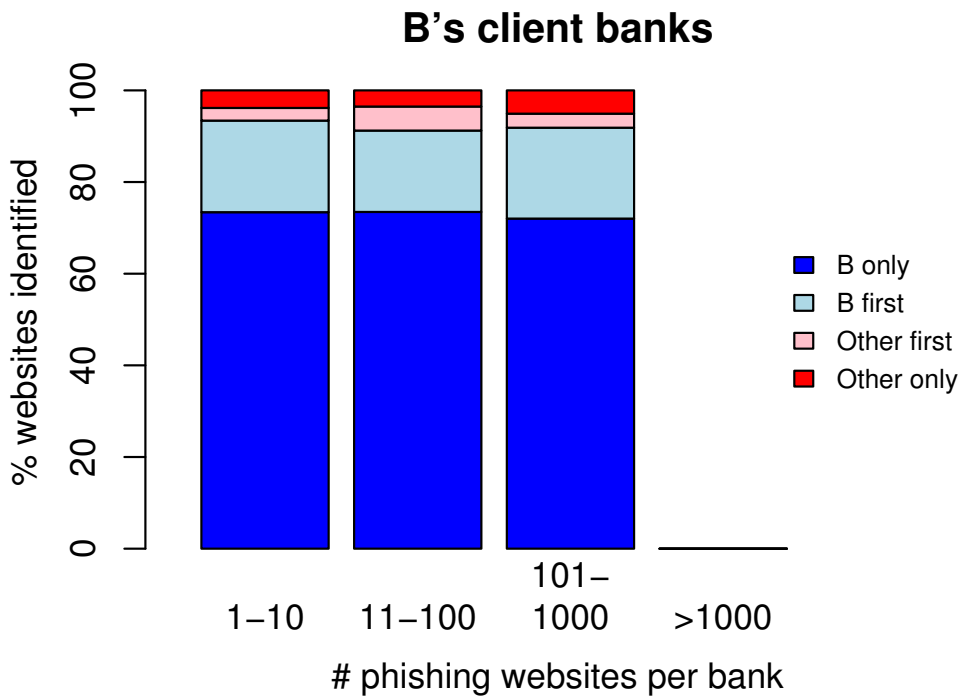
Mean difference (hours)



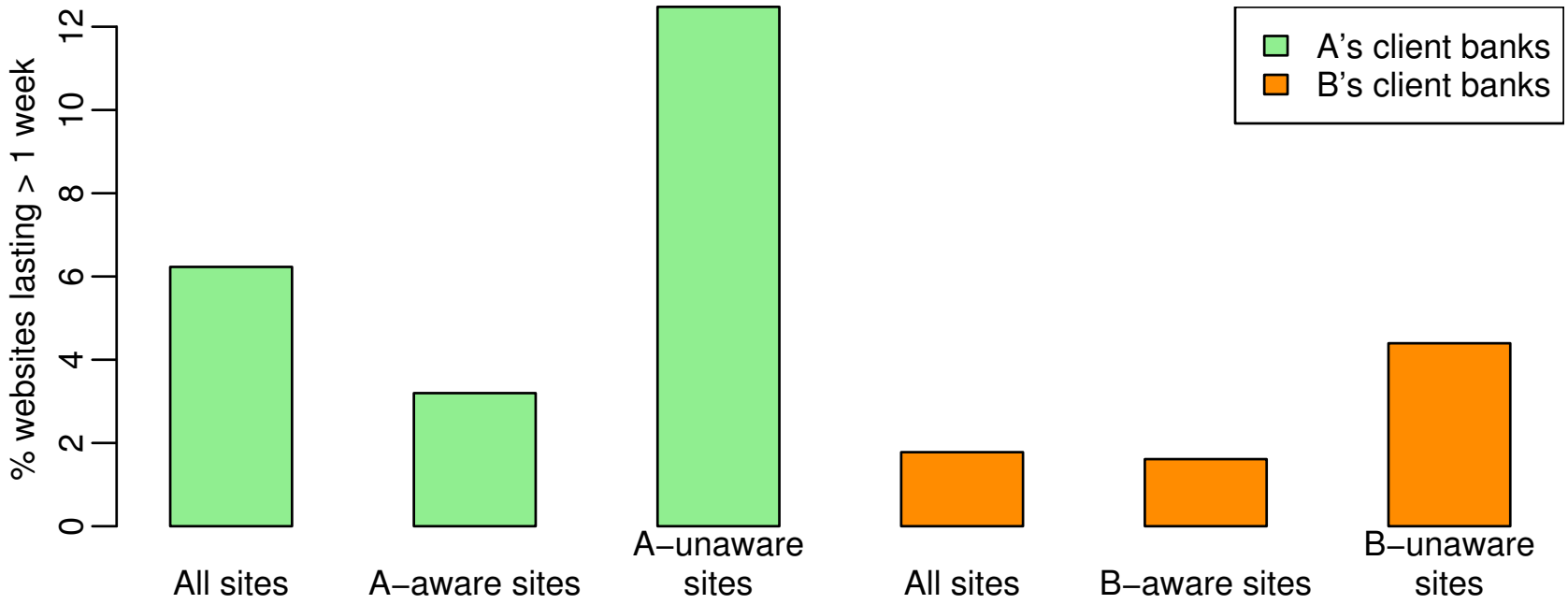
UNIVERSITY OF CAMBRIDGE

Popularity of phishing target affects gain from sharing

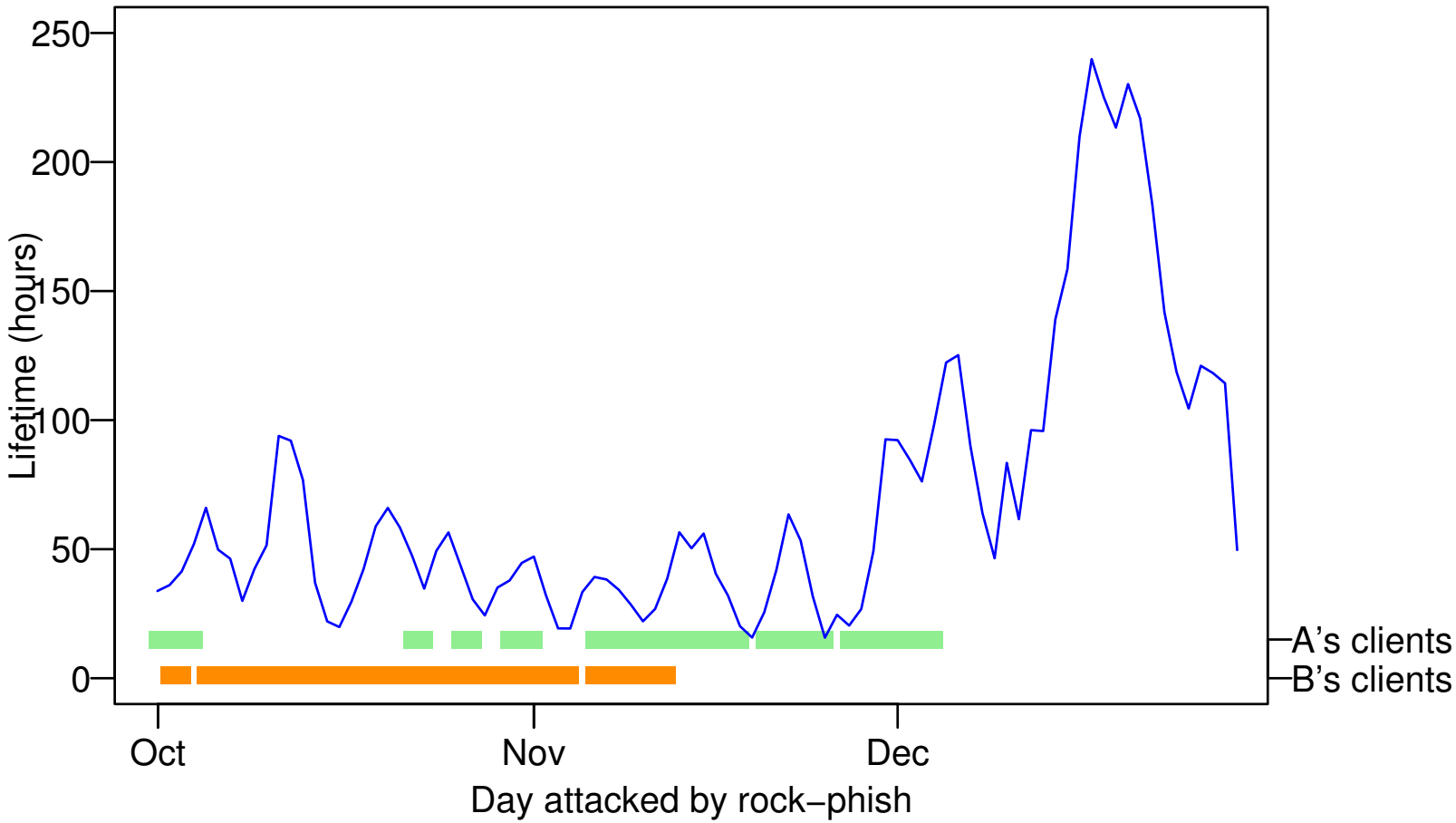




Long-lived phishing websites caused by not sharing URLs



Rock-phish website lifetimes depend on A and B's effort



User response to phishing

- Webalizer data
  - Web page usage statistics are sometimes set up by default in a world-readable state
  - Gives daily updates of which URLs are visited
  - We can view how many times a 'thank you' page is visited
  - We automatically checked all reported websites for the Webalizer package, revealing over 700 sites
- On-site text files
  - We retrieved around two dozen text files with completed user details from phishing sites
  - 200 of the 414 responses appeared legitimate

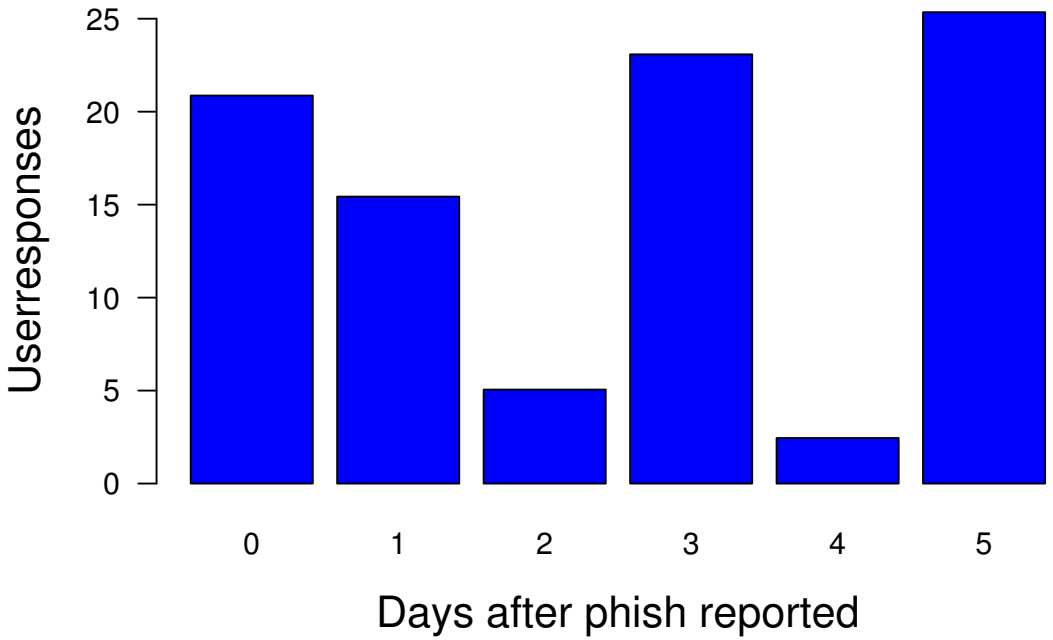
UNIVERSITY OF CAMBRIDGE

User response to phishing

- Webalizer data
  - Web page usage statistics are sometimes set up by default in a world-readable state
  - Gives daily updates of which URLs are visited
  - We can view how many times a 'thank you' page is visited
  - We automatically checked all reported websites for the Webalizer package, revealing over 700 sites
- On-site text files
  - We retrieved around two dozen text files with completed user details from phishing sites
  - 200 of the 414 responses appeared legitimate

UNIVERSITY OF CAMBRIDGE

User responses to phishing sites over time



$$\frac{\#\text{victims}}{\text{site}} = \text{mean lifetime} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims before detection.}$$

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks

- Having measured how many phishing sites exist, how long they stick around, and how many people give away their details, we can estimate the losses due to phishing
- DISCLAIMER: Cost is the product of several fuzzy estimates
  1. $61 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims on 1st day} = 30 \frac{\text{victims}}{\text{site}}$
  2. PhishTank identified 1 438 banking phishing sites, which implies 9 347 p.a.
  3. Upon examining other feeds, we conclude PhishTank identifies just 34.9% of phishing sites
  4. We therefore estimate $\frac{9\,347}{0.349} = 26\,800$ phishing websites p.a.
  5. Gartner estimate cost of identity theft to be \$572 per victim
  6. Estimated loss $= 30 \frac{\text{victims}}{\text{site}} \times 26\,800 \text{ sites} \times \$572 = \$460\text{m}$

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks

- Having measured how many phishing sites exist, how long they stick around, and how many people give away their details, we can estimate the losses due to phishing
- DISCLAIMER: Cost is the product of several fuzzy estimates
  1. $61 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims on 1st day} = 30 \frac{\text{victims}}{\text{site}}$
  2. PhishTank identified 1 438 banking phishing sites, which implies 9 347 p.a.
  3. Upon examining other feeds, we conclude PhishTank identifies just 34.9% of phishing sites
  4. We therefore estimate $\frac{9\,347}{0.349} = 26\,800$ phishing websites p.a.
  5. Gartner estimate cost of identity theft to be \$572 per victim
  6. Estimated loss $= 30 \frac{\text{victims}}{\text{site}} \times 26\,800 \text{ sites} \times \$572 = \$460\text{m}$

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks

- Having measured how many phishing sites exist, how long they stick around, and how many people give away their details, we can estimate the losses due to phishing
- DISCLAIMER: Cost is the product of several fuzzy estimates
  1. $61 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims on 1st day} = 30 \frac{\text{victims}}{\text{site}}$
  2. PhishTank identified 1 438 banking phishing sites, which implies 9 347 p.a.
  3. Upon examining other feeds, we conclude PhishTank identifies just 34.9% of phishing sites
  4. We therefore estimate $\frac{9\,347}{0.349} = 26\,800$ phishing websites p.a.
  5. Gartner estimate cost of identity theft to be \$572 per victim
  6. Estimated loss $= 30 \frac{\text{victims}}{\text{site}} \times 26\,800 \text{ sites} \times \$572 = \$460\text{m}$

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks

- Having measured how many phishing sites exist, how long they stick around, and how many people give away their details, we can estimate the losses due to phishing
- DISCLAIMER: Cost is the product of several fuzzy estimates
  1. $61 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims on 1st day} = 30 \frac{\text{victims}}{\text{site}}$
  2. PhishTank identified 1 438 banking phishing sites, which implies 9 347 p.a.
  3. Upon examining other feeds, we conclude PhishTank identifies just 34.9% of phishing sites
  4. We therefore estimate $\frac{9\,347}{0.349} = 26\,800$ phishing websites p.a.
  5. Gartner estimate cost of identity theft to be \$572 per victim
  6. $\text{Estimated loss} = 30 \frac{\text{victims}}{\text{site}} \times 26\,800 \text{ sites} \times \$572 = \$460\text{m}$

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks

- Having measured how many phishing sites exist, how long they stick around, and how many people give away their details, we can estimate the losses due to phishing
- DISCLAIMER: Cost is the product of several fuzzy estimates
  1. $61 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims on 1st day} = 30 \frac{\text{victims}}{\text{site}}$
  2. PhishTank identified 1 438 banking phishing sites, which implies 9 347 p.a.
  3. Upon examining other feeds, we conclude PhishTank identifies just 34.9% of phishing sites
  4. We therefore estimate $\frac{9\,347}{0.349} = 26\,800$ phishing websites p.a.
  5. Gartner estimate cost of identity theft to be \$572 per victim
  6. $\text{Estimated loss} = 30 \frac{\text{victims}}{\text{site}} \times 26\,800 \text{ sites} \times \$572 = \$460\text{m}$

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks (cont'd.)

- Notes regarding the $460m annual loss estimate
  - Ignores rock-phish attacks, which account for around half of phishing spam
  - Less than Gartner's estimate that 3.5m people fall victim to identity theft at annual cost of $2 Bn
  - Much of the gap can be attributed to rock-phish, keyloggers, and other causes of identity theft not related to phishing
  - Microsoft Research estimated 2m victims (vs. our 800k estimate) using a completely different technique
- We can similarly estimate losses caused by not sharing feeds
  - Compare the lifetimes of phishing websites known to A and B to the lifetimes of websites unknown to them
  - This time difference is a direct consequence of not sharing feeds

UNIVERSITY OF CAMBRIDGE

Estimating the cost of phishing attacks (cont'd.)

- Notes regarding the $460m annual loss estimate
  - Ignores rock-phish attacks, which account for around half of phishing spam
  - Less than Gartner's estimate that 3.5m people fall victim to identity theft at annual cost of $2 Bn
  - Much of the gap can be attributed to rock-phish, keyloggers, and other causes of identity theft not related to phishing
  - Microsoft Research estimated 2m victims (vs. our 800k estimate) using a completely different technique
- We can similarly estimate losses caused by not sharing feeds
  - Compare the lifetimes of phishing websites known to A and B to the lifetimes of websites unknown to them
  - This time difference is a direct consequence of not sharing feeds

UNIVERSITY OF CAMBRIDGE

What is the cost of non-cooperation?

- Total exposure of A's 53 targeted clients during Q4 2007:

$$(57.4 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims}) \times 7\,106 \text{ sites} \times \$572 = \$117\text{m}$$

- 2 219 websites impersonating A's clients missed by A:

$$(112.2 - 13.9) \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,219 \text{ sites} \times \$572 = \$44\text{m}$$

- 2 205 websites found by A 40.9 hours after other sources:

$$40.9 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,225 \text{ sites} \times \$572 = \$18\text{m}$$

- \$62m of A's clients' \$117m put at risk during Q4 2007 is due to not sharing feeds

UNIVERSITY OF CAMBRIDGE

What is the cost of non-cooperation?

- Total exposure of A's 53 targeted clients during Q4 2007:

$$(57.4 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims}) \times 7\,106 \text{ sites} \times \$572 = \$117\text{m}$$

- 2 219 websites impersonating A's clients missed by A:

$$(112.2 - 13.9) \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,219 \text{ sites} \times \$572 = \$44\text{m}$$

- 2 205 websites found by A 40.9 hours after other sources:

$$40.9 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,225 \text{ sites} \times \$572 = \$18\text{m}$$

- \$62m of A's clients' \$117m put at risk during Q4 2007 is due to not sharing feeds

UNIVERSITY OF CAMBRIDGE

What is the cost of non-cooperation?

- Total exposure of A's 53 targeted clients during Q4 2007:

$$(57.4 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims}) \times 7\,106 \text{ sites} \times \$572 = \$117\text{m}$$

- 2 219 websites impersonating A's clients missed by A:

$$(112.2 - 13.9) \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,219 \text{ sites} \times \$572 = \$44\text{m}$$

- 2 205 websites found by A 40.9 hours after other sources:

$$40.9 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,225 \text{ sites} \times \$572 = \$18\text{m}$$

- \$62m of A's clients' \$117m put at risk during Q4 2007 is due to not sharing feeds

UNIVERSITY OF CAMBRIDGE

What is the cost of non-cooperation?

- Total exposure of A's 53 targeted clients during Q4 2007:

$$(57.4 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} + 8.5 \text{ victims}) \times 7\,106 \text{ sites} \times \$572 = \$117\text{m}$$

- 2 219 websites impersonating A's clients missed by A:

$$(112.2 - 13.9) \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,219 \text{ sites} \times \$572 = \$44\text{m}$$

- 2 205 websites found by A 40.9 hours after other sources:

$$40.9 \text{ hrs} \times \frac{8.5 \text{ victims}}{24 \text{ hrs}} \times 2\,225 \text{ sites} \times \$572 = \$18\text{m}$$

- \$62m of A's clients' \$117m put at risk during Q4 2007 is due to not sharing feeds

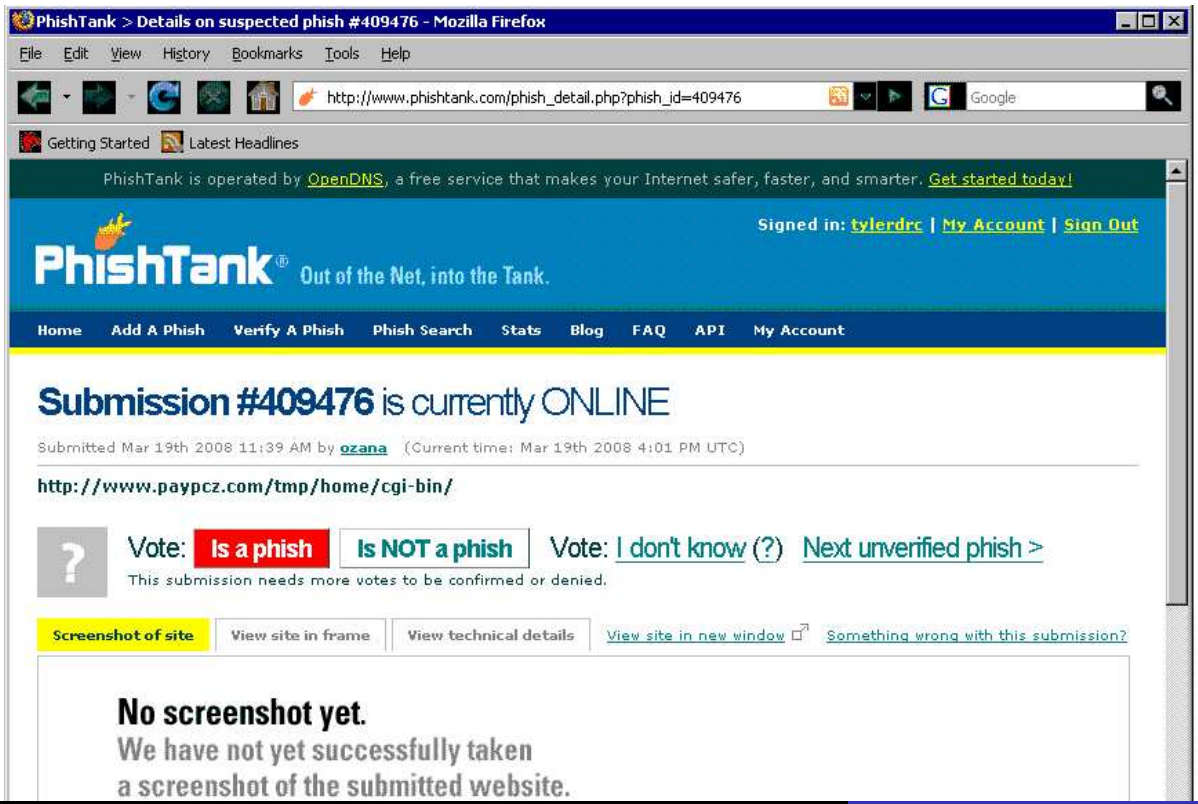
UNIVERSITY OF CAMBRIDGE

Outline

UNIVERSITY OF CAMBRIDGE

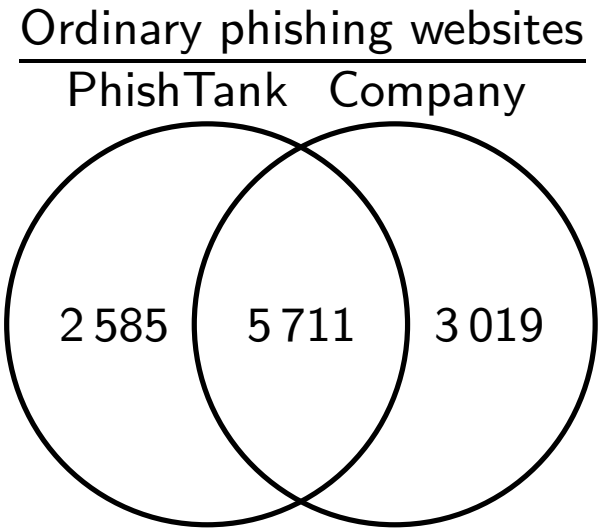
PhishTank

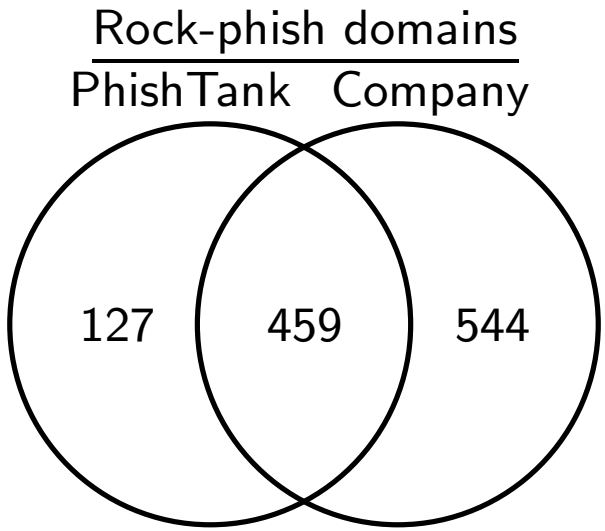
- Online community established in 2006 using the 'wisdom of crowds' to fight phishing
- Users contribute in two ways
  1. Submit reports of suspected phishing sites
  2. Vote on whether others' submissions are really phishing or not



UNIVERSITY OF CAMBRIDGE

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





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

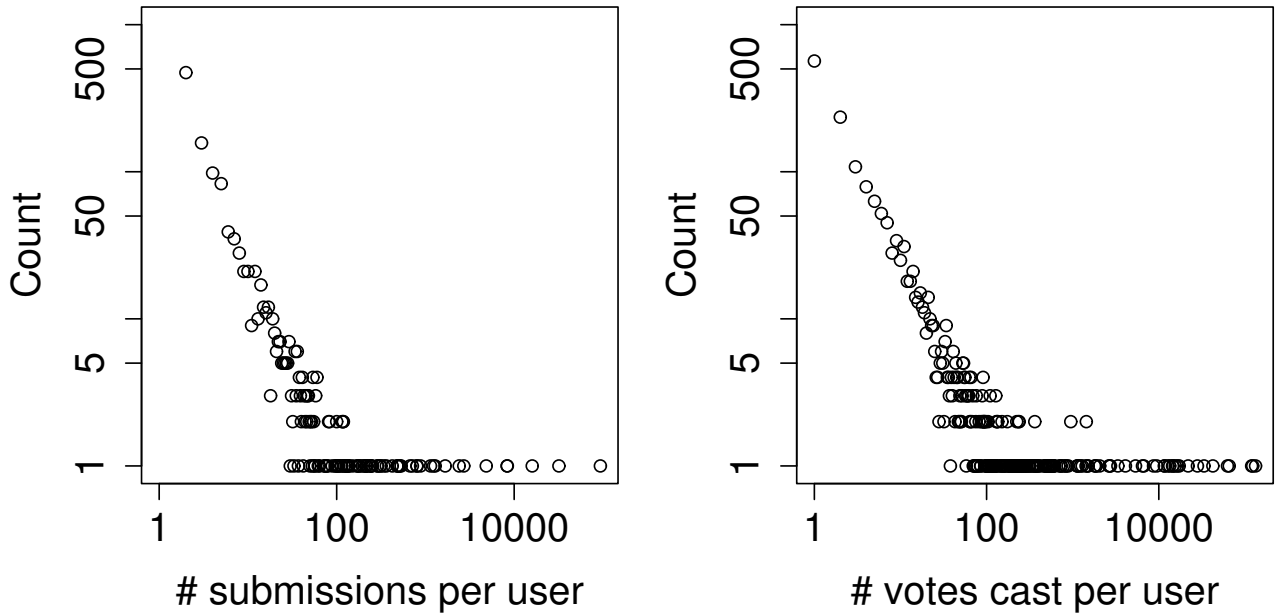
UNIVERSITY OF CAMBRIDGE

PhishTank data collection

- We examined reports from 176 366 phishing URLs submitted between February and September 2007
- 3 798 users participated, casting 881 511 votes
- $\Longrightarrow$ 53 submissions and 232 votes per user. But ...

UNIVERSITY OF CAMBRIDGE

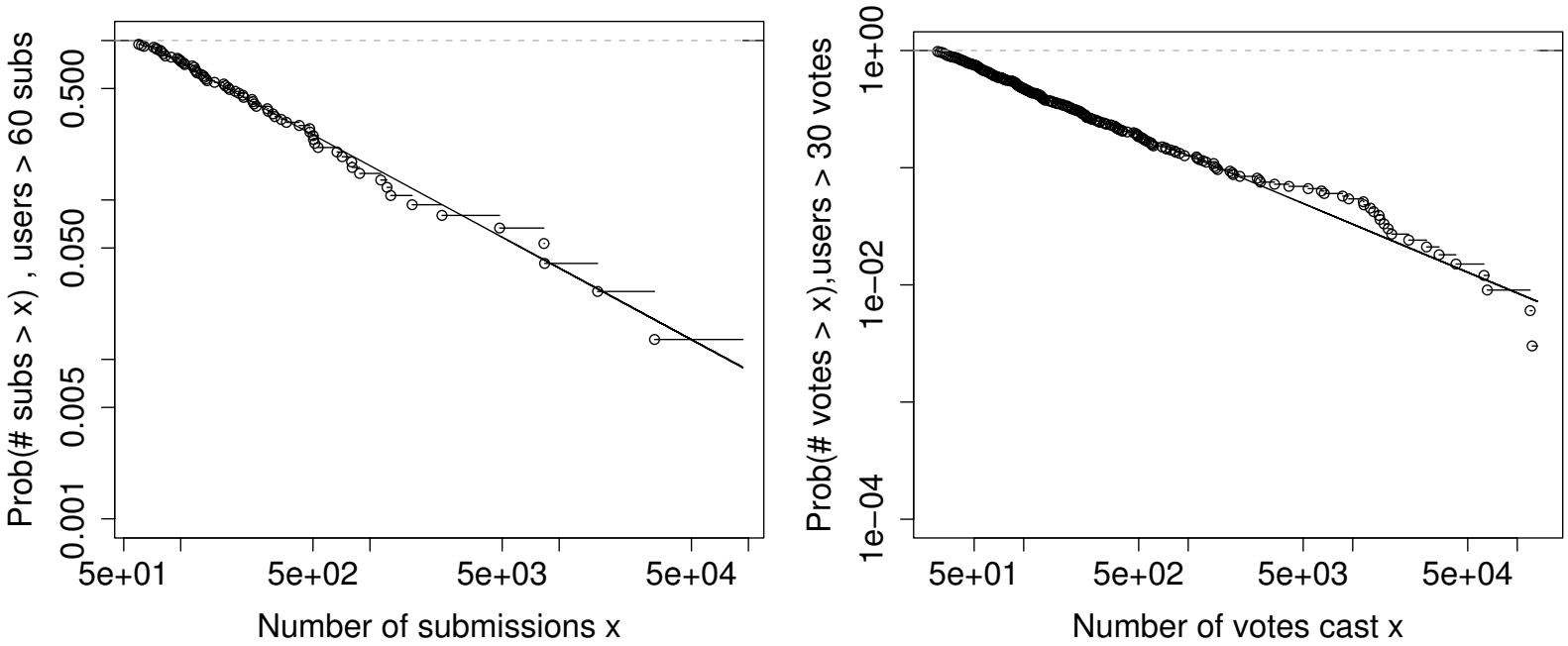
Density of user submissions and votes



- 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 sends out unique URLs

  `http://www.volksbank.de.netw.oid3614061.lof80.info/vr`
- 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

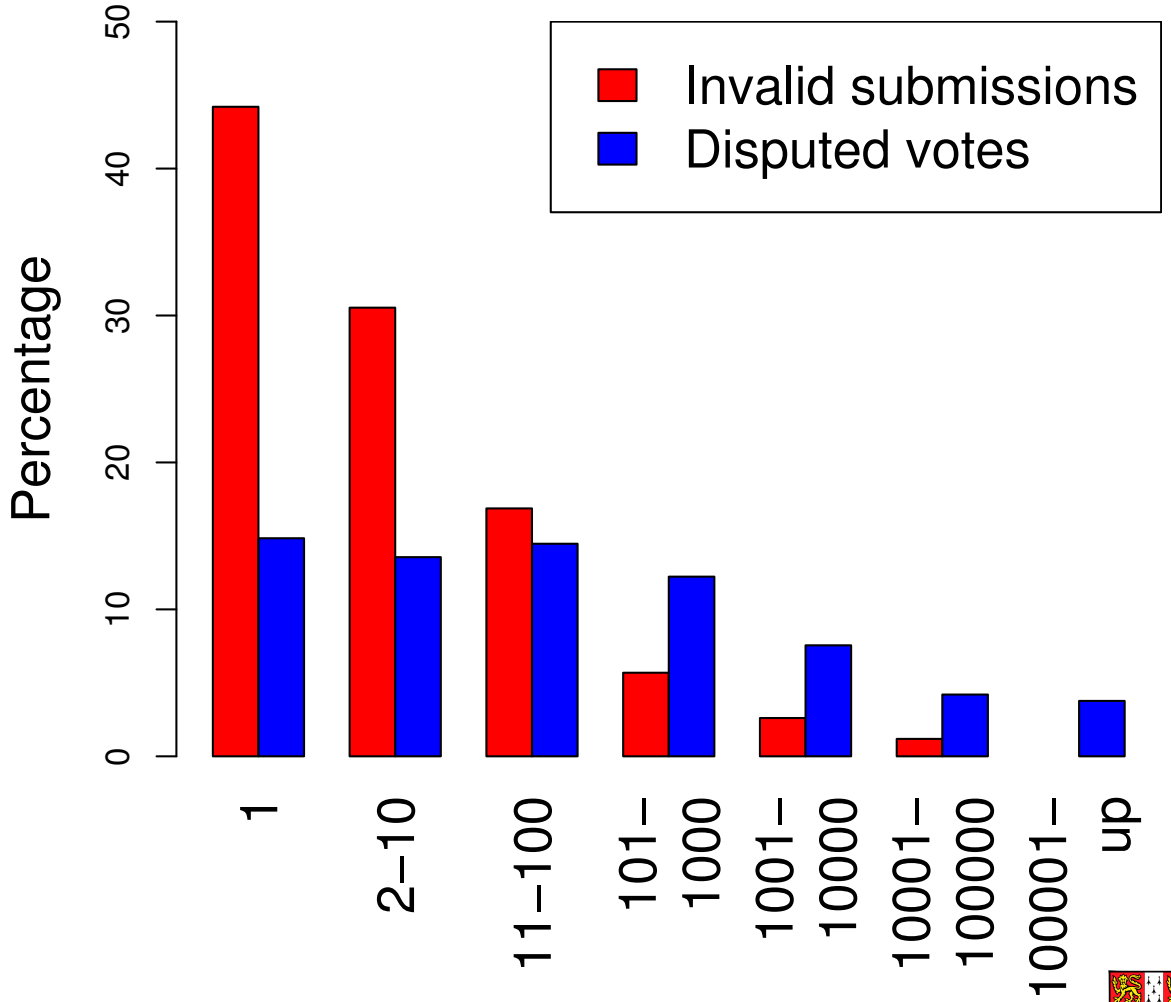
UNIVERSITY OF CAMBRIDGE

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?



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

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

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

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

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

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

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

Conclusions

- Empirically examining attacks leads to many insights!
- We have established that there is wide disparity in phishing website lifetimes
- Banks should demand take-down companies share URL feeds
- We have also seen attackers innovate: rock-phish sites outlive ordinary phishing sites through clever adaptations in strategy
- While leveraging the wisdom of crowds sounds appealing, it may not always be appropriate for information security tasks
- For more, see `http://www.cl.cam.ac.uk/~twm29/` and `http://www.lightbluetouchpaper.org/`

UNIVERSITY OF CAMBRIDGE