Spamming for Science: Active Measurement in Web 2.0 Abuse Research

Andrew G. West [1] -- Pedram Hayati [2]
Vidyasagar Potdar [2] -- Insup Lee [1]

WECSR -- March 2, 2012

2 Curtin University of Technology
1 Penn Engineering

Outline

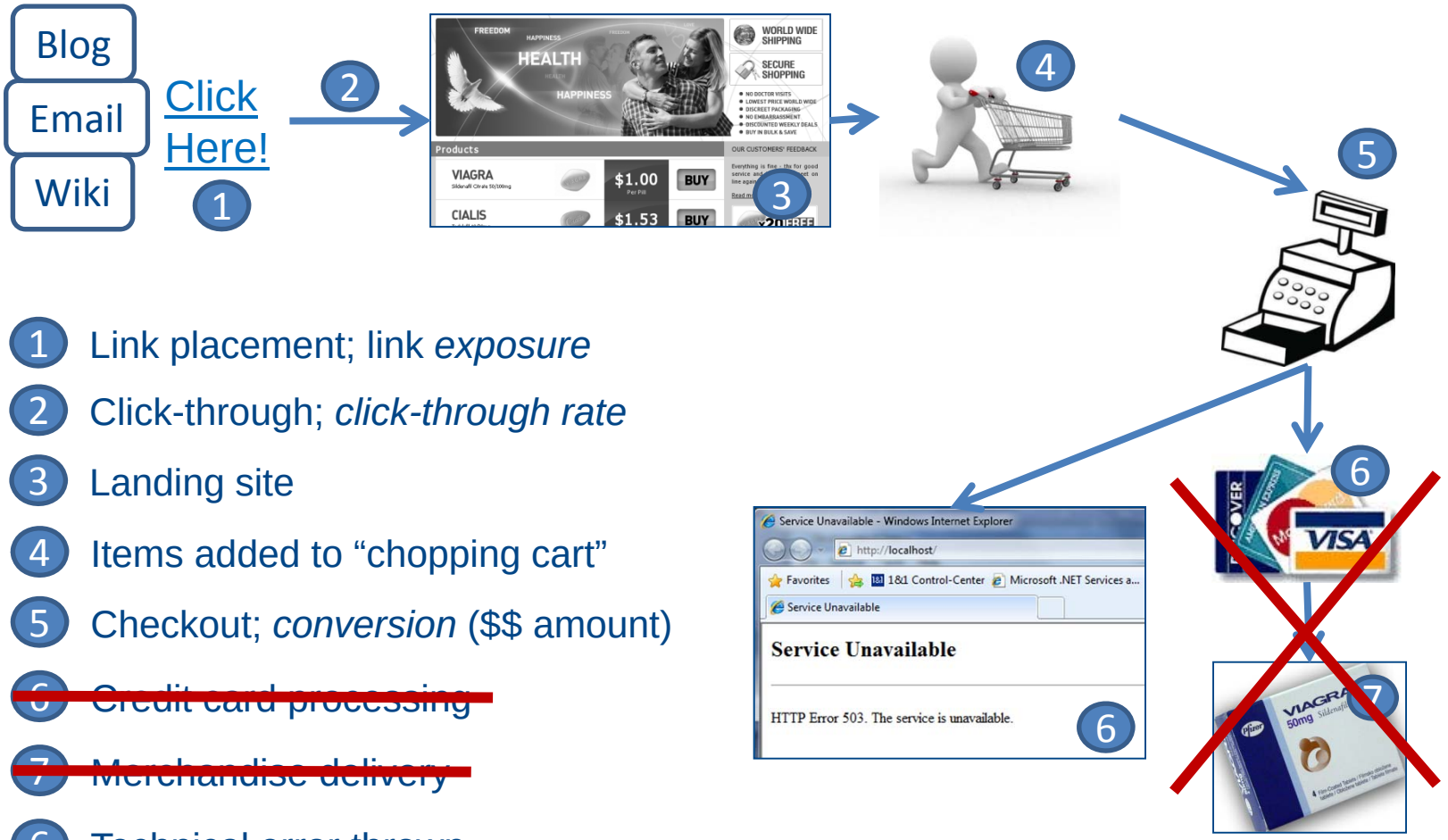
- Economic spam experiments
  - UPenn: Novel attack on Wikipedia
  - Curtin: Status quo of blog/forum spam
- Approvals process
  - IRB/HREC criteria/conditions
  - Legal and other considerations
- Community reaction
- Related work
- Discussion

Purpose/Takeaway

- Case study -- story time!
  - A “how to” in getting research like this approved
  - What justifications/exemptions are feasible
- Exemplifies IRB and CS relationship issues
- Research has been approved and conducted, yet remains unpublished on ethical grounds
  - This is unsatisfactory on multiple levels
  - Proposals for fixing; discuss

THE EXPERIMENTS

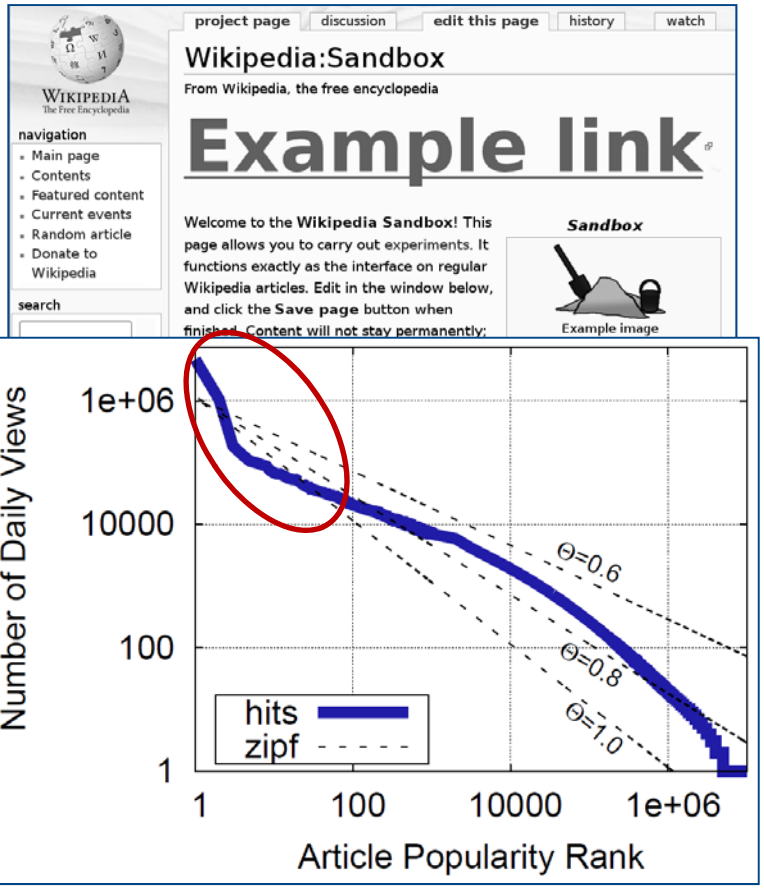
Spam Pipeline [3]



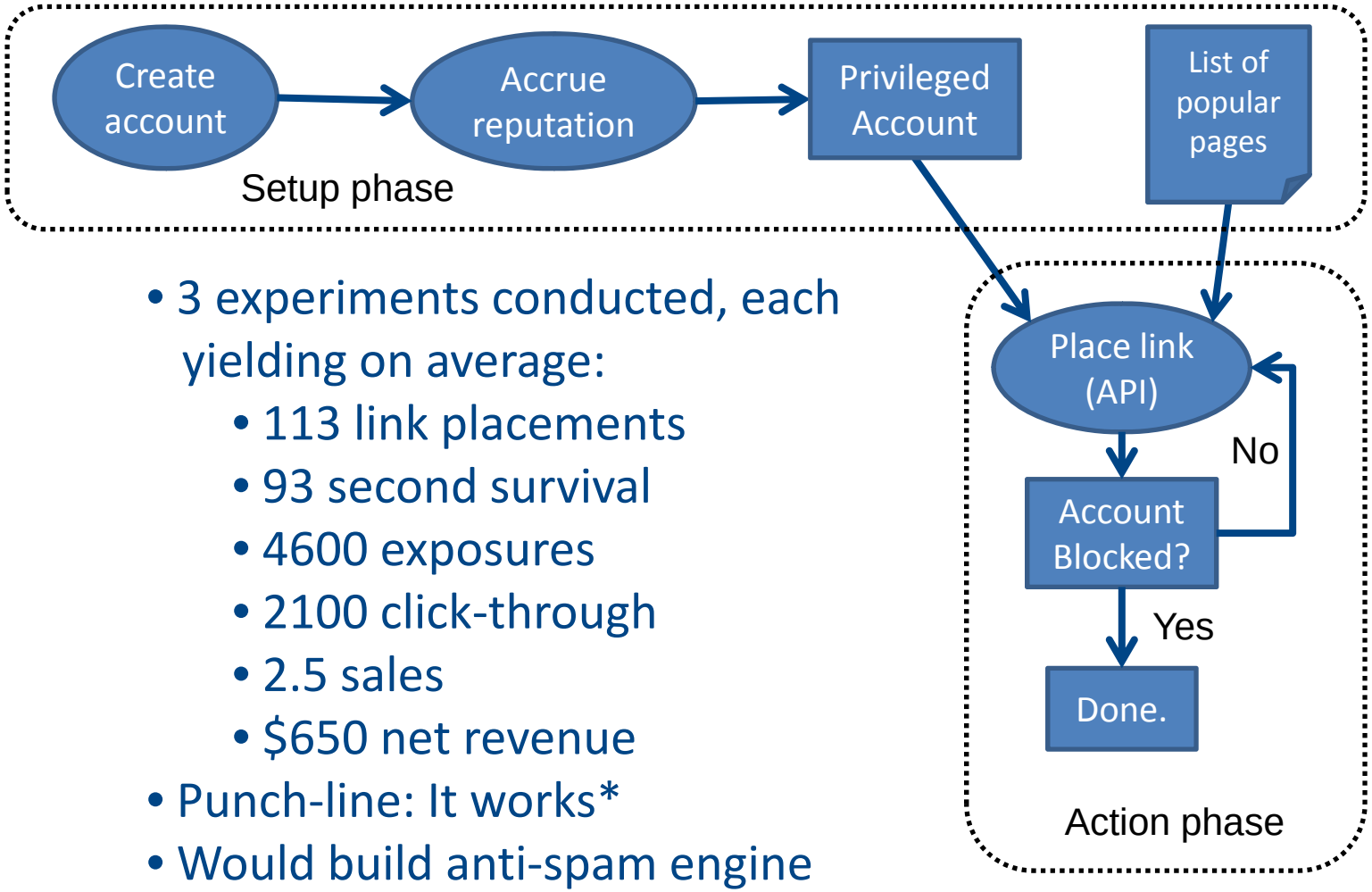
1. Link placement; link *exposure*
2. Click-through; *click-through rate*
3. Landing site
4. Items added to “chopping cart”
5. Checkout; *conversion* ($$ amount)
6. ~~Credit card processing~~
7. ~~Merchandise delivery~~
6. Technical error thrown

Wikipedia Model

- Status quo of spam: naïve and ineffective [8]
- Novel attack [8], characterized by:
  - Reputation*
  - Prominence
  - High-value targets
- Human mitigation introduces inherent detection latency



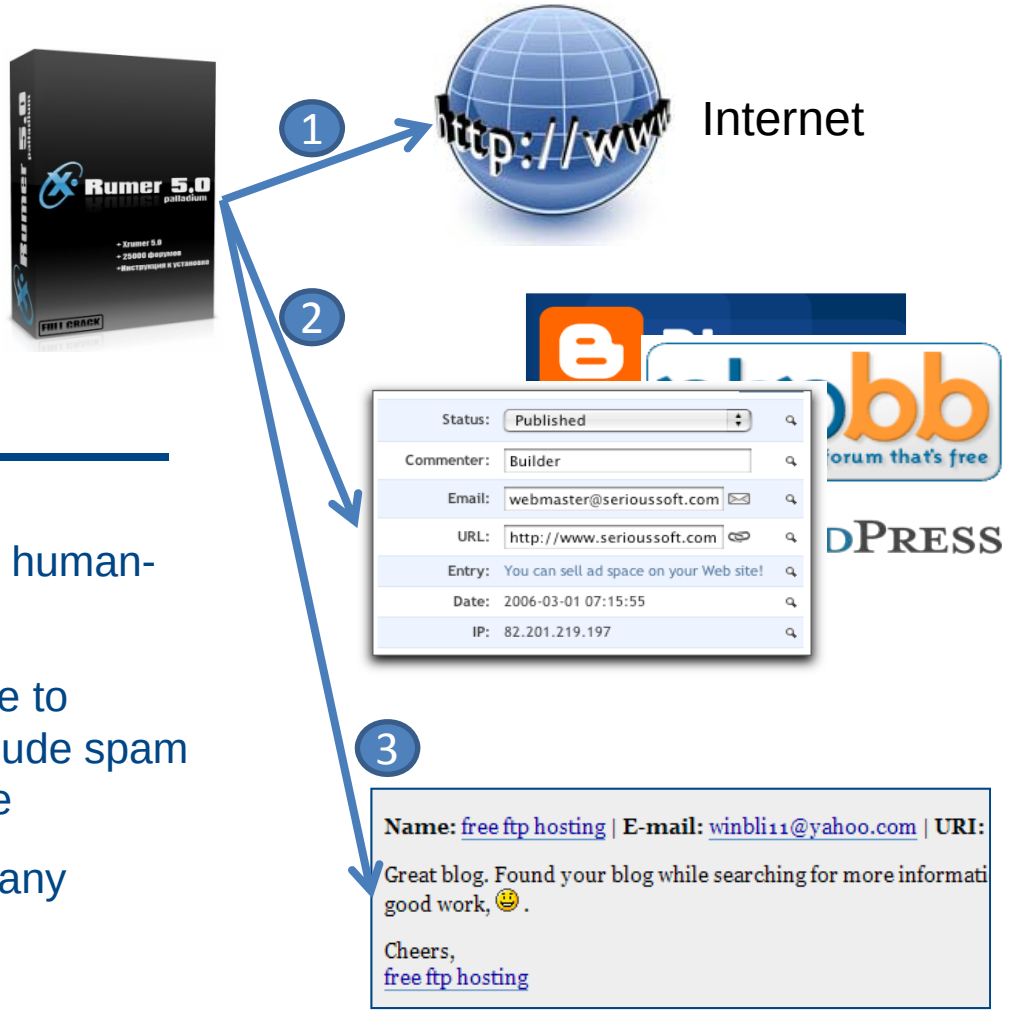
Wikipedia Outcome



- 3 experiments conducted, each yielding on average:
  - 113 link placements
  - 93 second survival
  - 4600 exposures
  - 2100 click-through
  - 2.5 sales
  - $650 net revenue
- Punch-line: It works*
- Would build anti-spam engine

Blog Model

Just do it as attackers already do [1]; keep statistics, point to payment-disabled pharmacy



1. Crawl the Internet looking for human-complete-able web forms
2. Exploit template-based nature to fill out in intelligent ways. Include spam URL somewhere in message
3. Submit form. Hope to evade any spam filters

Blog Outcome [2,7]

98,000 - Web forms harvested

66,226 - Actually targeted

7,772 - Successful link placements

2,059 - Visits to the pharmacy

3 - Attempts-to-purchase

- $400-$2000 estimated profit, over 1 month
- Note “link exposures” were not quantified
- Motivated research into behavioral recognition of how bots interact w/forms

GETTING APPROVAL*

*(remember, authors are non-experts)

Big Picture

Location:	Univ. of Pennsylvania, United States	Curtin University, Australia
Review org.:	IRB	HREC
Approval time:	≈ 14 weeks	≈ 1 meeting
Paraphrased justifications and concerns:	Full review; Real concern is publishing attack details and wiki-damage by actual malicious actors	Expedited review; "does not pose a greater risk than participants would face in their normal daily routine."

Informed Consent

Why would informed consent be biasing?	Experiments take place on 3rd party sites where: (1) consent form might be spam, (2) consent awkwardly adjacent to measurement, (3) no consent allows ignore (as is typical)	
Any notion of debriefing?	(1) Community notified in central "newspaper" (2) Notification of exploit	None
Why not more?	(1) If we notify them when they leave "pipeline"; could affect how future users interact w/experiment (2) Preserve anonymity; no contact details for later	

PII & Etc.

Data retained:	Hash of IP address Purchase items/amount Basic metadata	First name, last name, email address of "customers" (!)
Computing setup:	Web-host: 3rd party Launch: Cloud	Host: Self (non-Univ. ISP) Launch: VPN/Proxies
Legal approval	"No objection to publication"	Not required to seek

COMMUNITY REACTION
(*i.e.*, submission and review)

Reviews (1)

"… measurement study is a bit offensive, but the IRB approval seems to cover this … While the IRB problem is discussed, I am still not convinced that such experiments with Wikipedia are good from an ethical point of view."

* Underlining added by presentation authors

Reviews (2)

"I personally am concerned about the ethics of the active link-spamming research … research should not cause harm or damage to subjects without their informed consent .. it appears that harm or damage may have been done without prior consent of the Wikipedia Foundation … not persuaded ..."

* Underlining added by presentation authors

Reviews (3)

"Although they did get their institution's IRB to approve it, IRB approval is a necessary, but not sufficient, step for justifying such an experiment … such a cost, which is involuntary to the participants, needs to be justified by a significant gain in scientific understanding."

* Underlining added by presentation authors

Reviews (4)

"… their active experiment is ethically deficient … I view each [of multiple issues, the `ethical deficiency' included] as a deal-breaker … The ethical standing is dubious enough that it does *not* suffice to simply tell us that you had IRB approval. We need to know the wording of what the IRB approved. In addition, while the text briefly mentions (un)informed consent, there is no mention of *post facto debriefing* … [this] makes the reviewer wonder to what degree the authors really did obtain IRB approval that was itself informed."

* Underlining added by presentation authors

Reviews (5)

"…the paper is rather offensive … the discussion in the appendix is also not very convincing… I suggest to revise the appendix and maybe even publish all IRC documents … Apart from this aspect, the study is interesting and the authors demonstrate convincingly that Wikipedia is an attractive target for link spam."

* Underlining added by presentation authors

Related Work

- Studies doing spam-like things [5]
  - Malicious posts to celebrity social profiles [9]
    - No IRB; some complaints; still published
  - Personalized phishing mails to students [6]
    - Approved; considerable justification in-document
  - Race-based response-rate among professors [4]
    - Cross-domain; IRB as gold-standard in other fields
- Studies paying bad guys
  - Blackhat software, spam products, CAPTCHAs

DISCUSSION

SQ Unsatisfactory

Curtin University of Technology

Penn Engineering

Review stage is a poor place for enforcement!

- Harm has already done; but is not allowed to contribute to scientific understanding
- Personal sour grapes; less interest in pursuing similar research (globally true)
- Lost time and grant resources
- Need to allocate space for ethical justification
- Reviewers (and possibly PCs) become ethical regulators; likely outside their technical expertise

Possible Solutions

1. New and/or corollary organization to "approve" CS research a priori

2. Wider respect for IRB decisions; greater community education to this end.

3. Incremental improvements during conference CFP, submission, and review

CIS-IRB 2.0

- A theoretical “win” for all involved…
- … but technical challenges aplenty:
  - Replace or operate along-side the IRB?
  - Who pays for it? Who does reviewing?
  - Handling multiple legal/cultural jurisdictions?
  - Is this novel or just IRB 2.0?
    - What policies are basis for review?
    - What forces PCs/reviewers to respect?
    - Why haven’t other fields required this?

Respect the IRB

1. New and/or corollary organization to "approve" CS research a priori

2. Wider respect for IRB decisions; greater community education to this end.

3. Incremental improvements during conference CFP, submission, and review

Practical Progress

- Part of a PC's job is to set an agenda; ethics can be part of that (BlackHat/WOOT/LEET)
- Try to indicate leanings in CFP?
- Easier submission of IRB documentation; make it available to reviewers (anonymity)
- Ethics as a new dimension of paper scoring
- Quick-review by PC; possibility to DQ paper
  - Are ethics orthogonal to technical merit?

References

[1] **XRumer** (Blackhat SEO software). http://www.xrumerseo.com/

[2] Hayati, P., Firoozeh, N., Potdar, V., Chai, K.: **How much money do spammers make from your website?** (Working paper, in submission)

[3] Kanich, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G.M., Paxson, V., Savage, S.: **Spamalytics: An empirical market analysis of spam marketing conversion**. In: *CCS'08: Proc. of the Conf. on Computer and Comm. Security* (2008)

[4] Milkman, K.L., Akinola, M., Chugh, D.: **The temporal discrimination effect: An audit study of university professors**, (Working paper)

[5] Moore, T., Anderson, R.: **Economics and Internet security: A survey of recent analytical, empirical and behavioral research**. *Tech. Rep. TR-03-11, Harvard University, Department of Computer Science* (2011)

[6] Nathaniel, T.J., Johnson, N., Jakobsson, M.: **Social phishing**. *Communications of the ACM* 50(10) (October 2007).

[7] Thomson, C. **Who you gonna call?** *Curtin University News* (June 2011), http://news.curtin.edu.au/news/who-you-gonna-call/

[8] West, A.G., Chang, J., Venkatasubramanian, K., Sokolsky, O., Lee, I.: **Link spamming Wikipedia for profit.** In: *CEAS `11: Proc. of the Eighth Annual Collaboration, Electronic Messaging, Anti-Abuse, and Spam Conference* (September 2011)

[9] Ur, B.E., Ganapathy, V.: **Evaluating attack amplification in online social networks**. In: *W2SP'09: The Workshop on Web 2.0 Security and Privacy* (2009)