

AIRWeb 2009



The Potential for **Research and Development in**
Adversarial Information Retrieval



Brian D. Davison
Computer Science and Engr., Lehigh University



2

AIRWeb after 5 years

- Self-examination natural
- Redirection possibilities



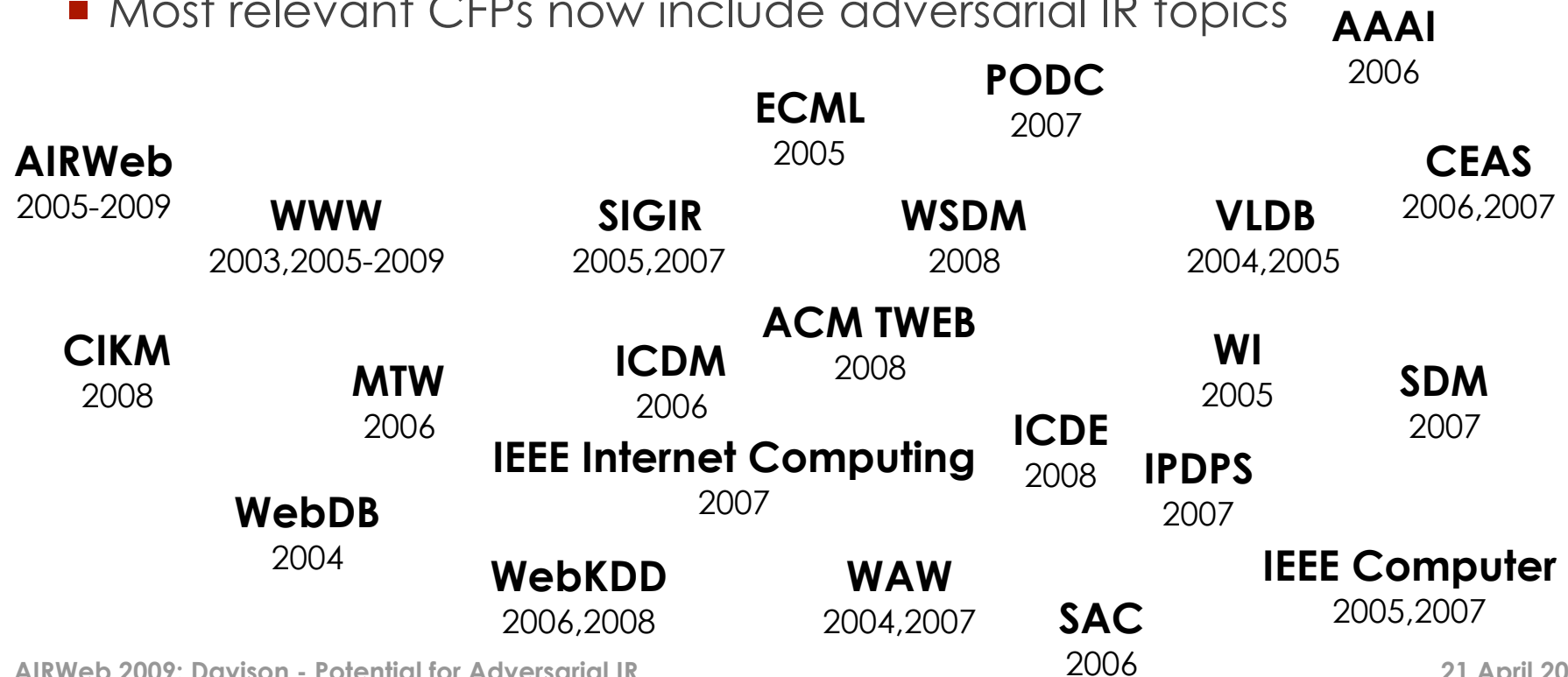
AIRWeb Topics Have a History

- Brin and Page, 1998
- Kleinberg, 1998/1999
- Bharat and Henzinger, 1998
- Lempel and Moran, 2000
- “Adversarial IR” coined by Broder in 2000

Work in AIRWeb topics has blossomed over the years



- Papers have been published in high-visibility venues
- Most relevant CFPs now include adversarial IR topics



Has the AIRWeb workshop
become superfluous?



Potential for Research and Development in Adversarial IR



- Not just AIRWeb
- Not strictly for the Web



Introduction

- Why am I here?
 - To remind you of things you might already know, but perhaps haven't thought about for a while
- Definitions
 - Adversarial: Assumes competing parties trying to affect the outcome of a system (system could be an algorithm, a market, etc)
 - Adversarial IR: Information retrieval, ranking, or classification system affected by multiple parties acting in their own interest



The Future

Search is Power

- The world now looks to the Web
 - through the eyes of search engines
 - to see what is happening
 - to answer questions
 - to learn
- “For the user, search is the power to find things, and for whoever controls the engine, search is the power to shape what you see.” —*Blown to Bits*
- Thus, adversarial web IR is tremendously important as it affects who controls search engine results





Perspectives

- It is common to find organizations (sometimes even extremist) that cater to a specific audience, both offline and online
 - Often telling them what they want to hear
- Every society has competing factions
 - liberal vs. conservative
 - orthodox vs. secular
- Many media organizations are aligned with, or at least cater to particular mindsets
 - News companies





Media/mind control

- Concentrated ownership of mass media long believed to be dangerous
 - Monopoly concerns
 - Desire for diversity of opinion and unfettered/unfiltered access to information
- The same kinds of divisions of perspective do not appear in today's search engines
 - Might expect them to develop as engines get better in answering non-factoid questions
 - Engines may still be manipulated by particular ideologies!

Surprising!



The truth

- What information can be considered true or objective?
 - Important to find out!
 - The Web is becoming the sum of human knowledge
- Imagine an adversary that does not want to sell anything, but instead wishes to influence public perception on some topic
 - Link bombing ("Google-bombing") is of this type
 - Future attacks might affect summarization, automated Q&A systems
 - Could be subtle! Extremist organizations, even (esp!) governments, may be willing to have a low-profile but effective impact on public perception of events and issues before us
- So this leads to a futuristic research challenge
 - Discover people/pages that are intentionally distorting the truth



The Present



Adversarial IR Today

- The field has typically focused on immediate responses to immediate problems
 - How to address specific kinds of search engine spam
 - Sometimes also considers the effect of publishing the method
- This is a war (of sorts)

“Know your enemy.”

—Sun Tzu, *The Art of War*



- How many kinds of spammers?
 - Are they in identifiable camps?
 - Do they work together or against each other?
- How many spammers are there?
 - Is there a subset that is particularly effective?
 - Is the set of (effective) spammers growing?
- What are the methods that spammers use?
 - Do we need to distinguish between white hat and black hat SEO?

Fighting Search Engine Spam: The big(ger) picture

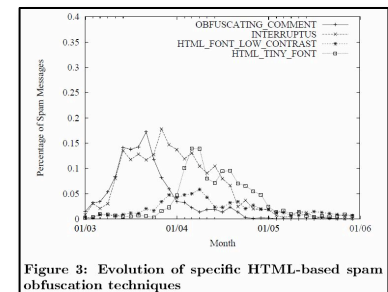
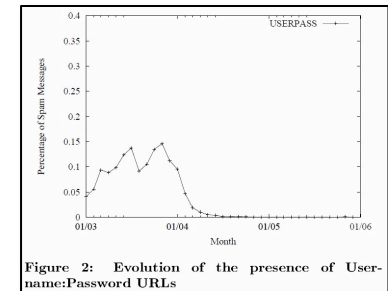


- Need to look beyond immediate actions and outcomes
- Need to examine and postulate the outcome of the larger adversarial system
 - Not easy!
 - Perhaps like a chess game with perpetual opportunities to change the rules
 - More complex than those typically studied in game theory
 - No one has all information (in the present or of the past)
- Goal: to model (and predict) actions and reactions of the adversaries



Guide: email spam research

- Observed Trends in Spam Construction Techniques: A Case Study of Spam Evolution
Pu and Webb, CEAS 2006
- Examined an email spam archive (three years)
- Celebrates "success stories" of spam methods that no longer are used
 - `http://user:password@host.domain`
 - `Vi<xxx>ag<yyy>ra`



GU

- Ob
- Tec
- Pu
- E
- C
- r
-
-

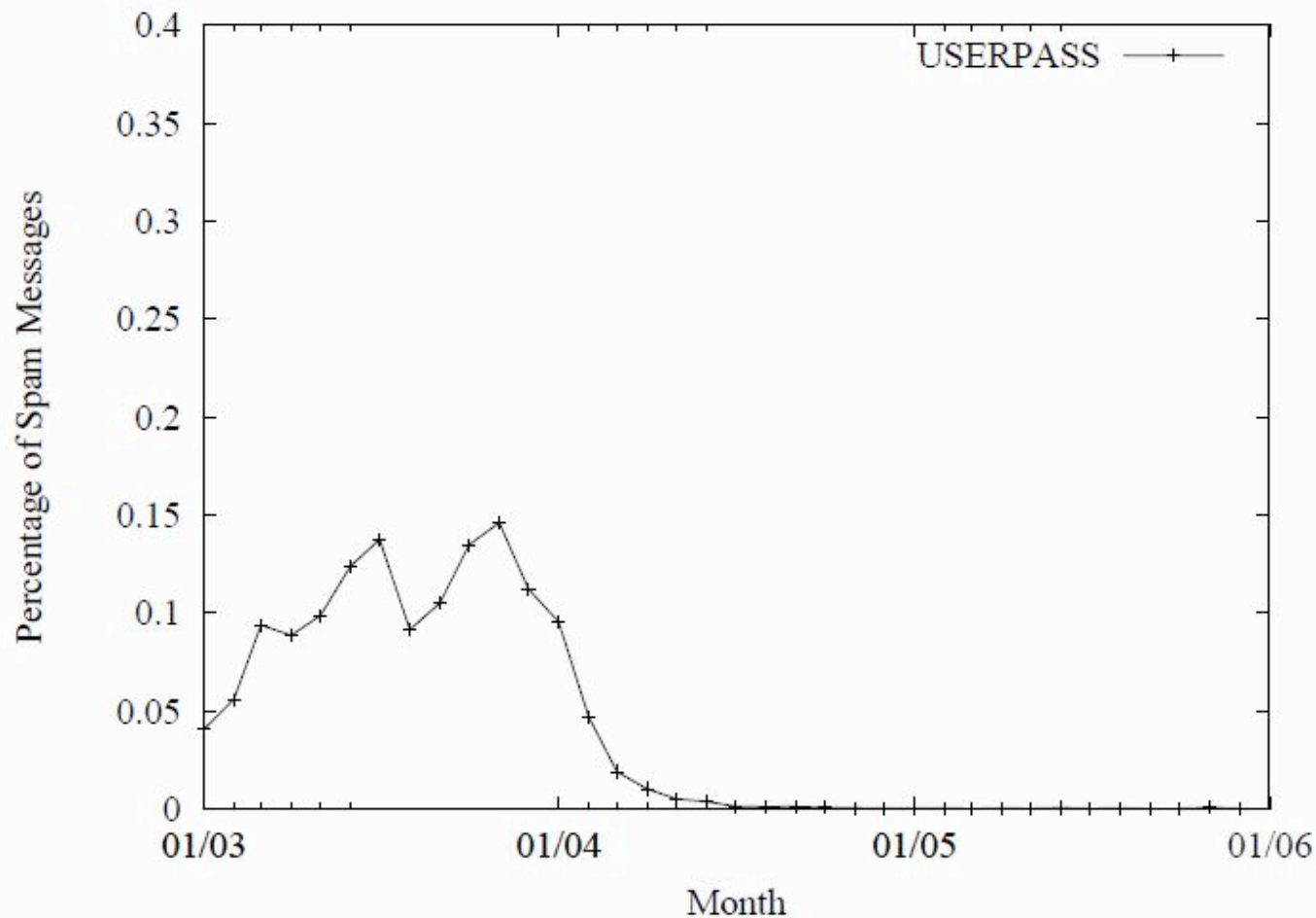
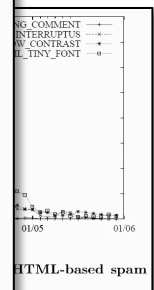


Figure 2: Evolution of the presence of User-name:Password URLs



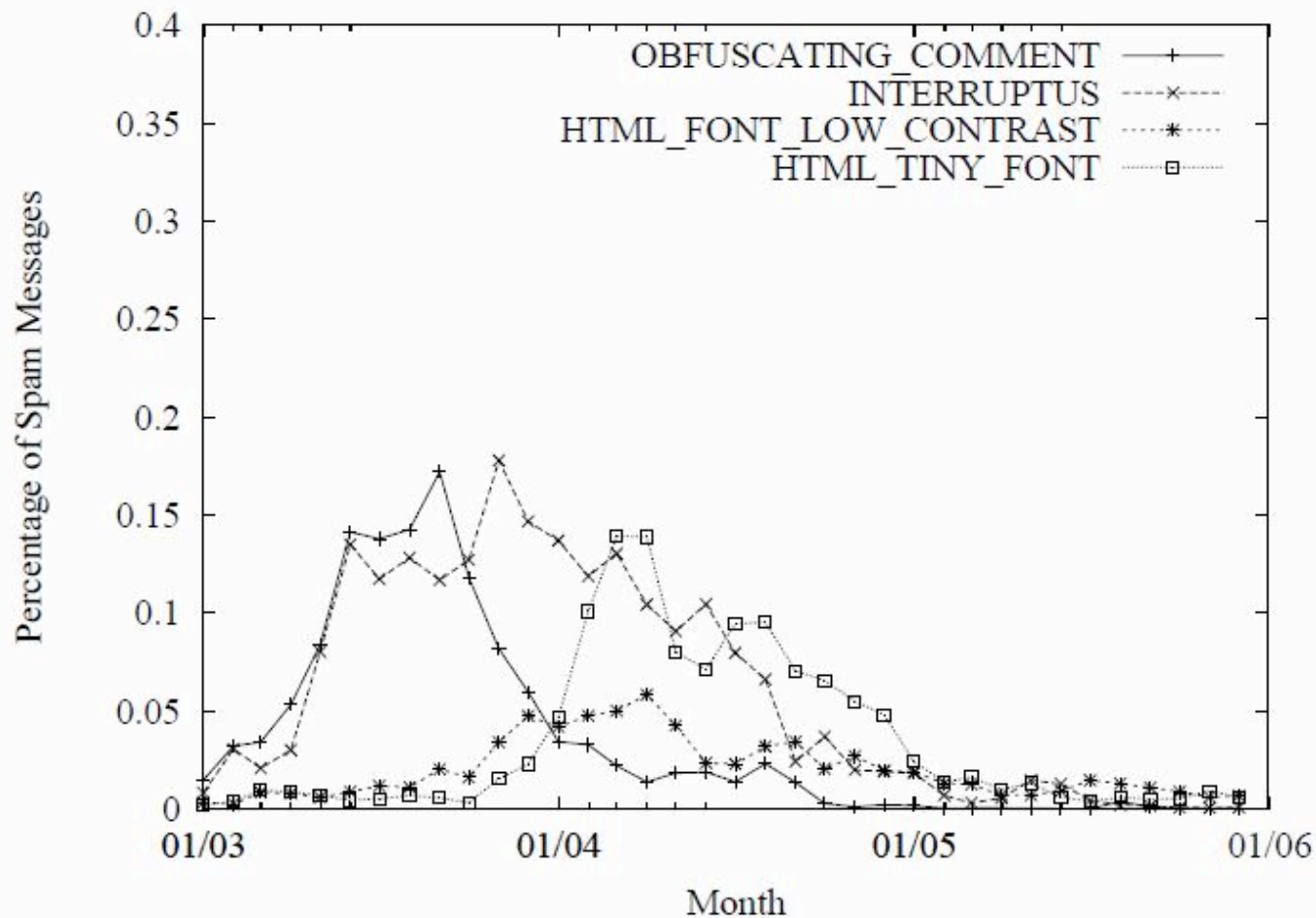


Figure 3: Evolution of specific HTML-based spam obfuscation techniques

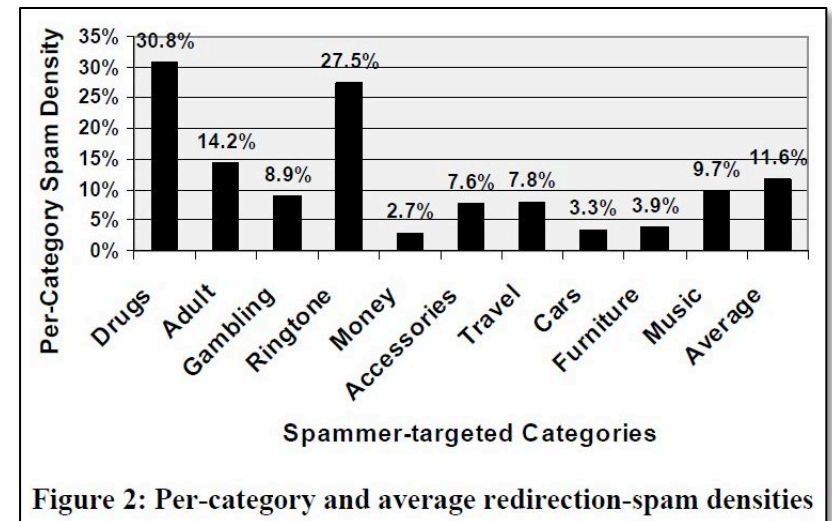
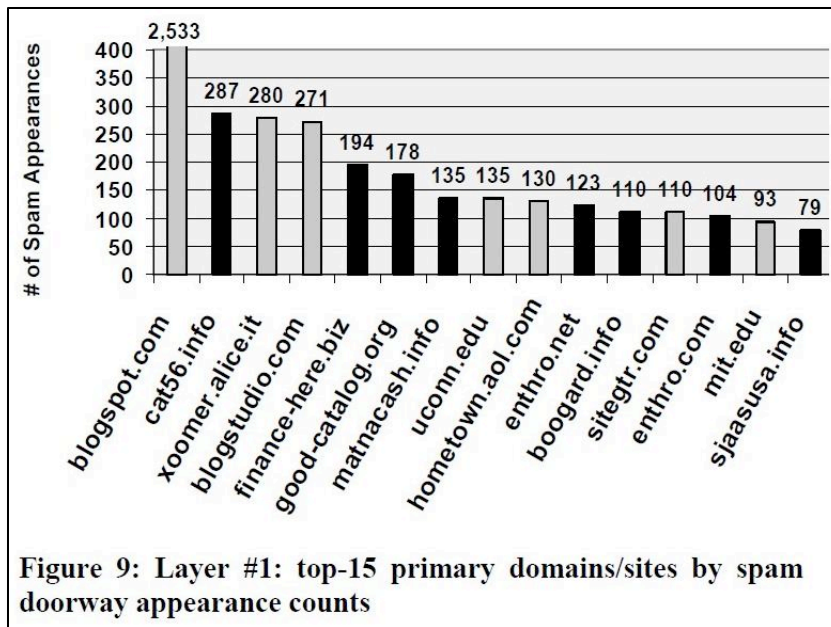


Email/web spam analysis

- Characterizing Web Spam Using Content and HTTP Session Analysis
Webb et al., CEAS 2007
 - ~350K URLs in full Webb corpus (from email spam)
 - 263K unique landing page URLs
 - 202K unique content pages
 - 109K clusters of duplicate and near-duplicate pages (after shingling)
 - 84% of pages hosted on 63.*-69.* and 204.* - 216.* IP addresses
 - Finds dominant sets of spammers



Web spam advertising analysis



Spam Double-Funnel: Connecting Web Spammers with Advertisers
Wang et al., WWW2007




Adversarial Situations are **Everywhere!**

- Email spam
- Search engine spam
- Many more...

[Back to Inbox](#)
[Archive](#)
[Report spam](#)
[Delete](#)
[More Actions](#)

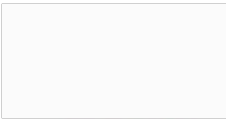
Photobucket - Lynda Smith wanted you to check out this image
[Inbox](#) | X


☆ [josephnathan@mail.md](#) to sam
 [show details](#) 17:52
 [Reply](#) | ▼



[lynda.jpg](#)

Good Day, I am using this opportunity to thank you for your effort to our unfinished transfer of fund into your account,I want to inform you that I have successfully transferred the Cheque out of the company to someone else who was capable of assisting me in this great venture .Due to your effort, sincerity, courage you showed at the course of the transaction I want to compensate you with the sum of \$1,200,000.00 (One Million, Two Hundred Thousand Dollars only).You are to contact the COMPENSATION FINANCE & INVESTMENT LTD where I deposited the fund to issue you an international certified bank draft cashable at your bank or make a direct transfer to you. GLOBAL SECURITY AND INVESTMENT LTD. Doctor Dr. Suleiman dako EMAIL: dr_suleimandako2@live.com Tel: +234 806 166 9491 At the moment, I am very busy here because of the investment projects I am having at hand. Finally,I have forwarded instruction to the finance house on your behalf to send the bank draft to you as soon as you con



 photobucket

If you are having problems viewing this email, copy and paste the following into your browser:
<http://s464.photobucket.com/albums/rr2/lyndasmith111/?action=view¤t=lynda.jpg>

Options ▼

<http://www.costpernews.com/archives/social-media-spam-sucks/>

Adversario
everywhere

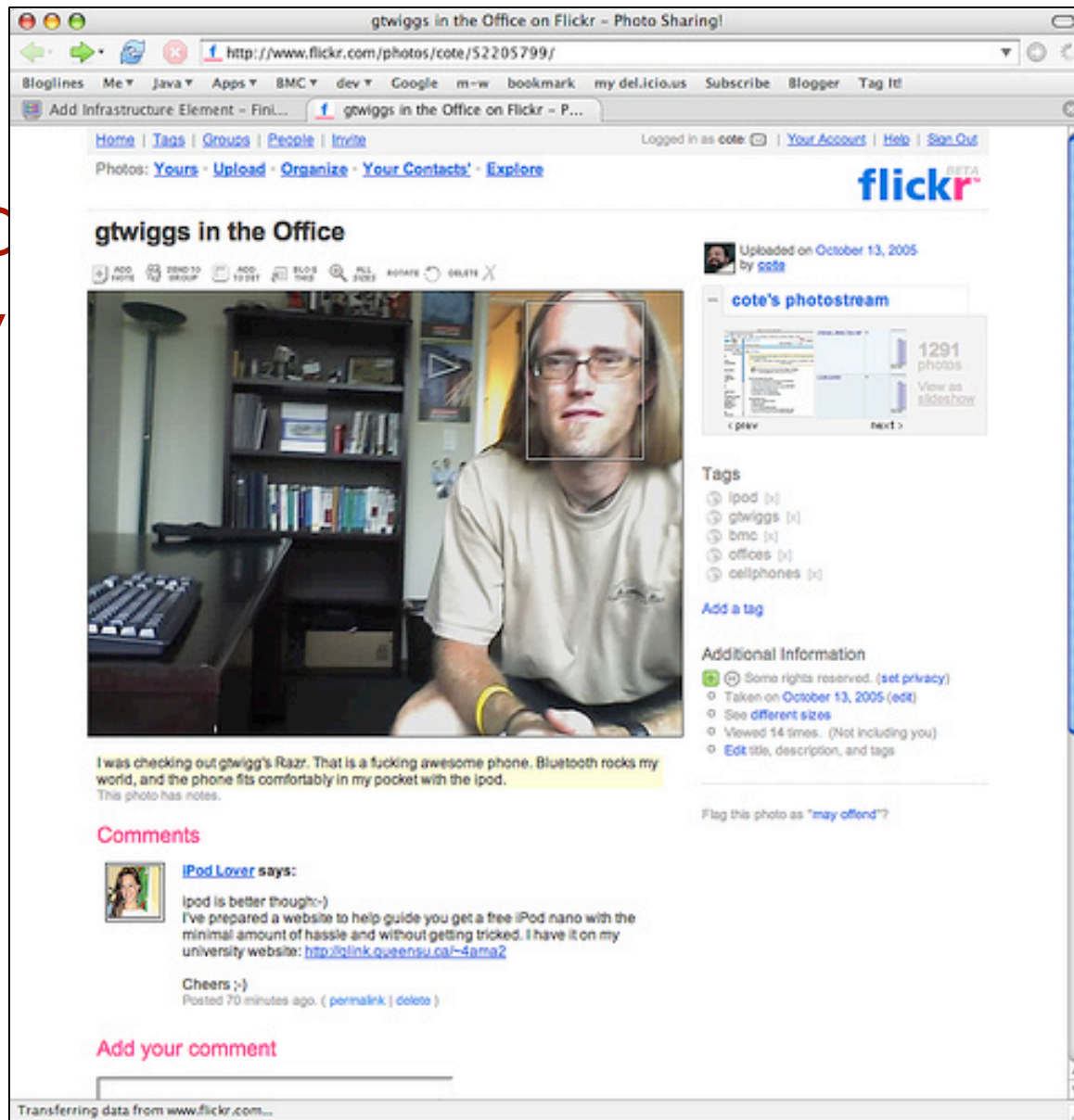


Who goes there?

Sorry, the account you were headed to has been suspended due to strange activity. [Mosey along now](#), nothing to see here.



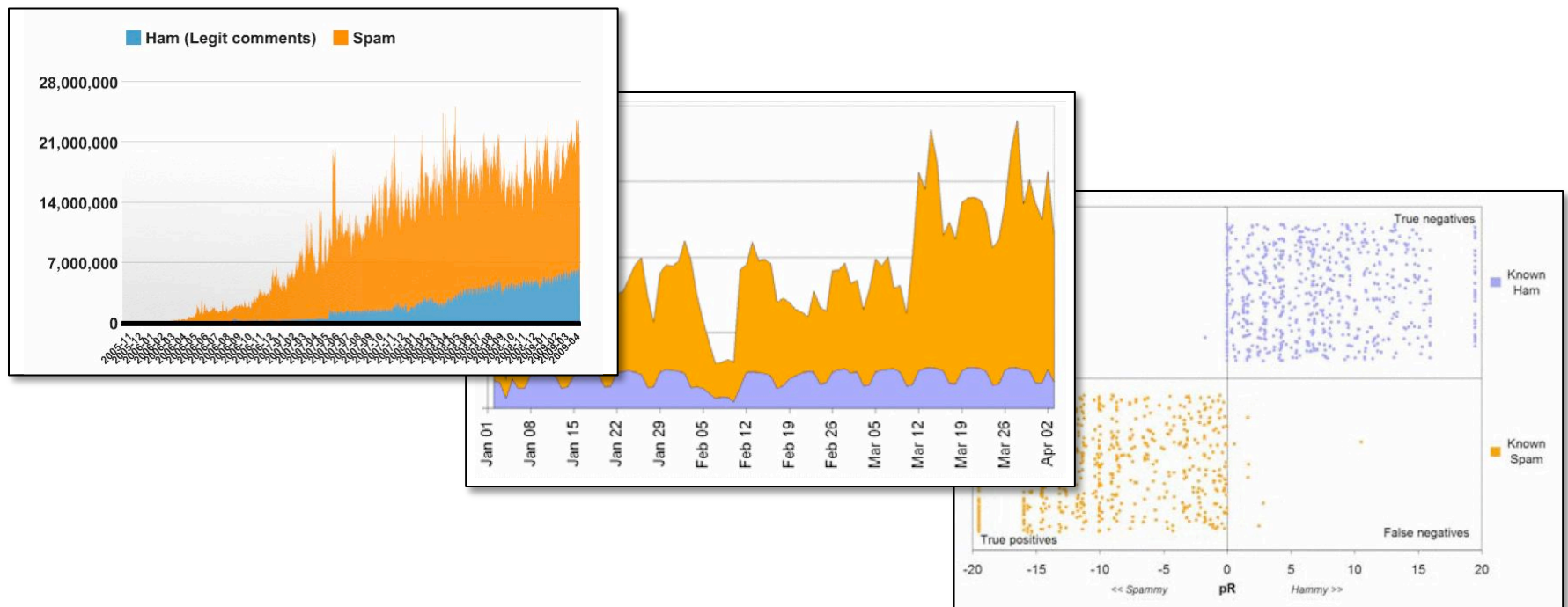
Ac
ev



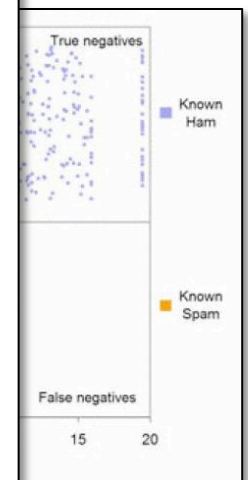
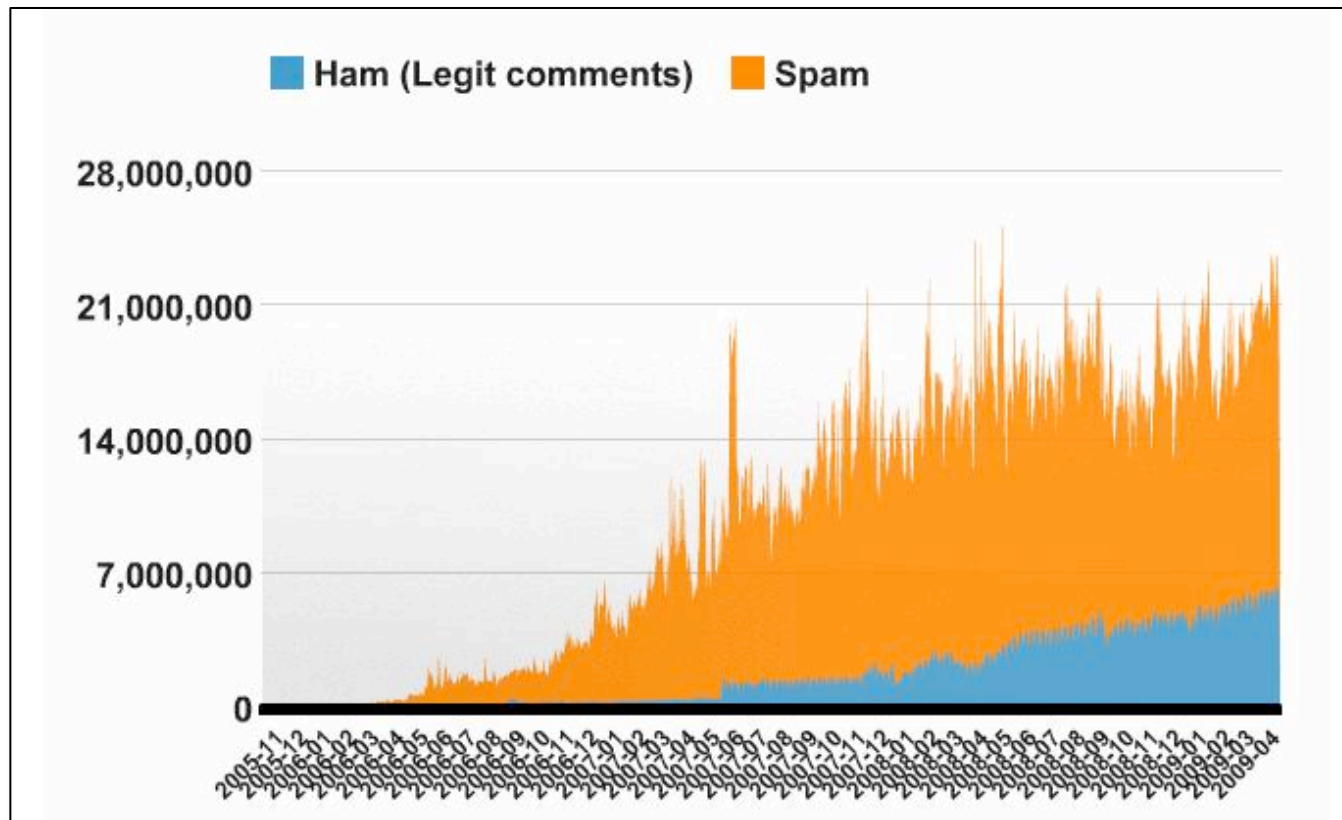
26

<http://www.flickr.com/photos/cote/52231621/>

Adversarial situations are everywhere: blog comments

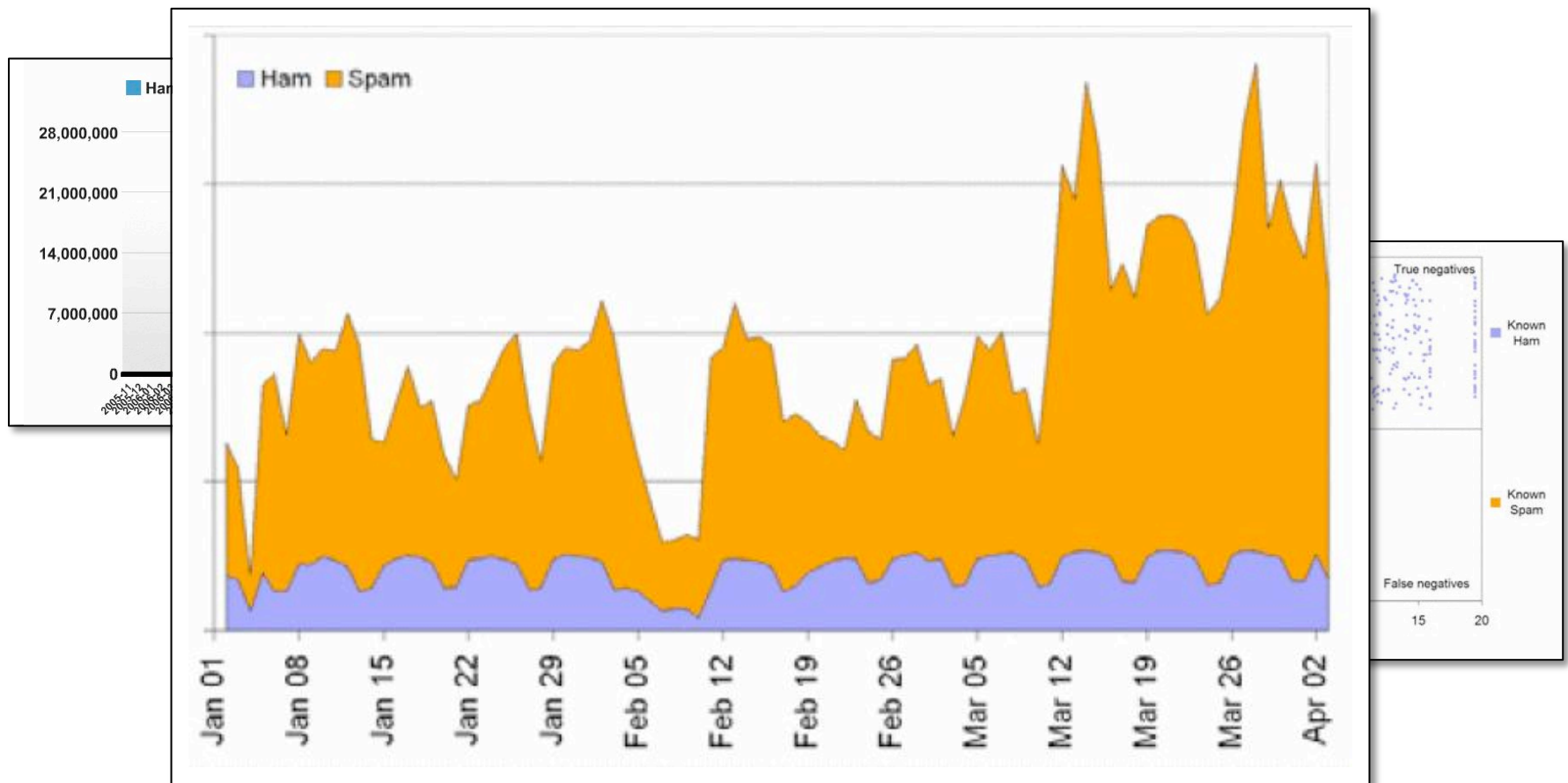


Adversarial situations are everywhere: blog comments

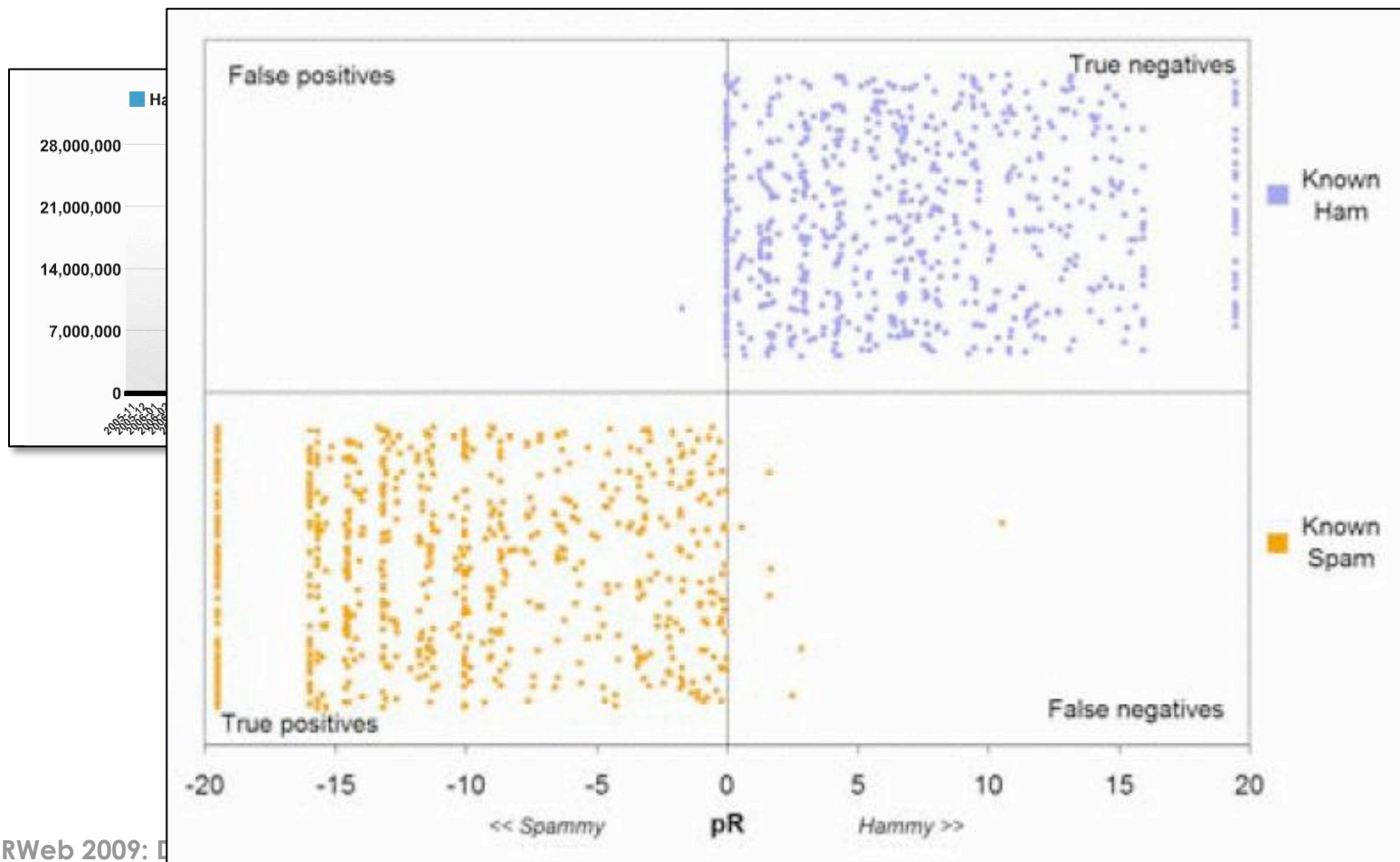


Akismet

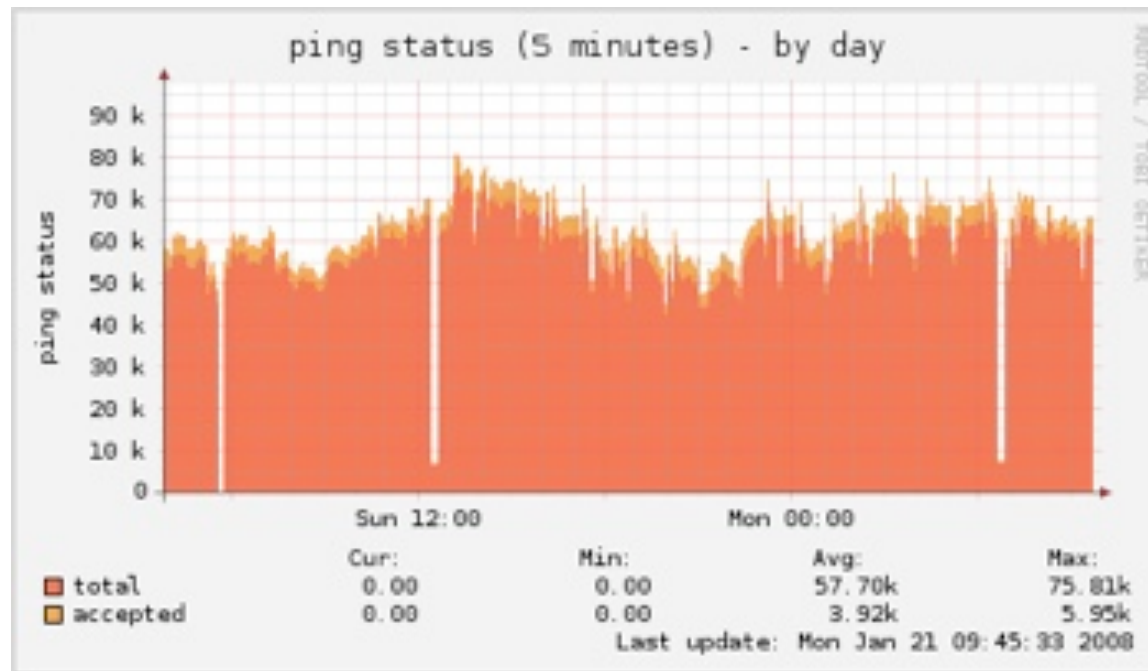
Adversarial situations are everywhere: blog comments



Adversarial situations are everywhere: blog comments



Adversarial situations are everywhere: blog pings



<http://blog.spinn3r.com/2008/01/blog-ping-and-s.html>

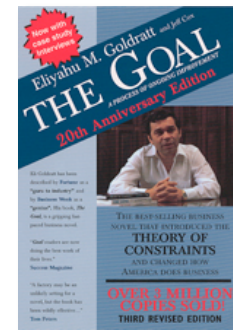


Spam in Social Systems

- Adversarial activities can be found in many social systems
 - Where they can impact the web (spam)
 - Either by creating links, or as secondary signals for search
 - E.g., Tag spam, comment spam
 - Potential for short-term (at least) research
 - Where they can garner social reputation
 - Masquerade as connectors, mavens, etc.
 - People with thousands of 'friends'

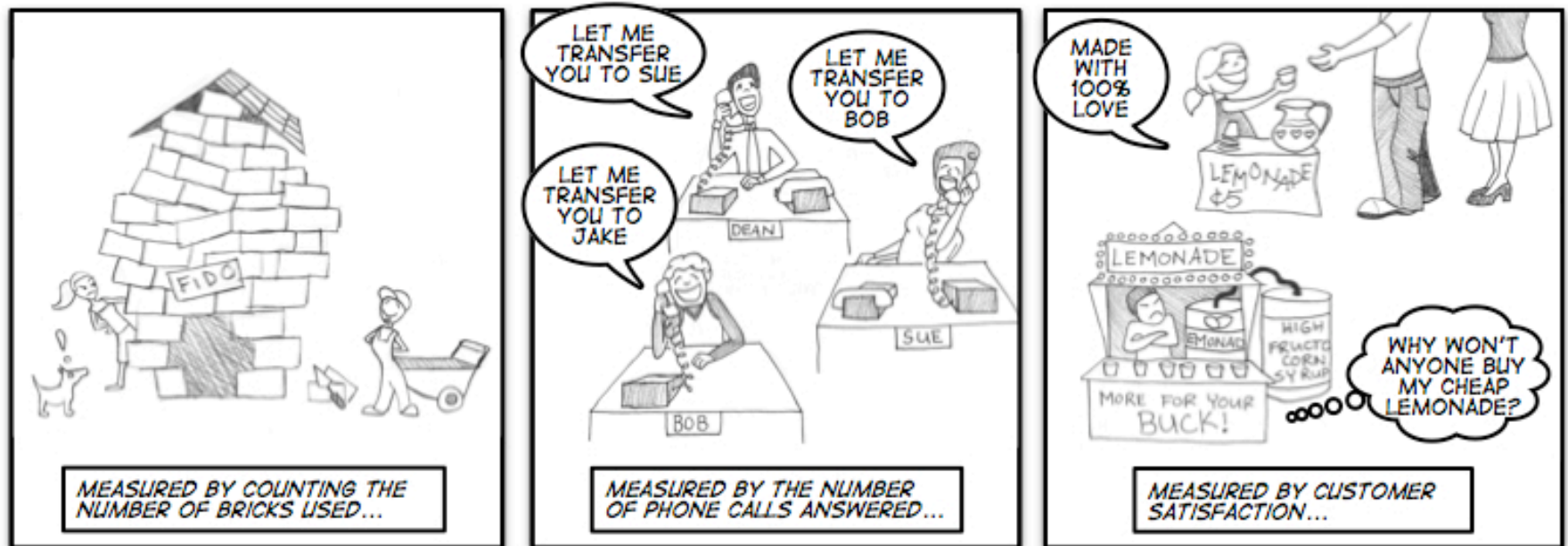
Acting in self-interest

- It is what (many!) people do
- “Tell me how [and when] you’ll measure me, and I’ll tell you how I’ll behave” –Eliyaho M. Goldratt, *The Goal*
- People are trained to satisfy metrics!





Acting in self-interest



<http://fridayreflections.typepad.com/weblog/2007/09/tell-me-how-you.html>



All warfare is based on deception

—Sun Tzu, *The Art of War*

- What if we had a transparent ranking system?
 - Publicize desired/utilized information
 - Expect self-promotion (and collusion, etc.)
 - But expose it
 - Penalize undesirable behavior
 - Reward desired behavior
- Might require strong identity management
 - (e.g., make activities traceable and thus have a social cost)



What do users want?

- To find information that satisfies their information need
 - To find relevant information...
 - To find reputable information...
 - To find truthful information...
- To maximize their opportunities in business and life
 - To increase visibility
 - To increase (perceived) stature/reputation
 - To increase (perceived) value



Research Topics Summary

- Find inaccurate information
 - Fact-checking, truth estimation, more subtle distortions
- Model adversarial scenario
 - Discover, understand and model the characteristics, knowledge and activities of adversaries
 - Examine history in order to consider the future of the larger adversarial system
- Consider new ranking systems such as transparent ones
 - Expecting and leveraging adversarial behavior
 - Explicitly (transparently) penalize poor behavior that should be discouraged
 - Reward desired behavior (explicitly)
 - Perhaps needing strong identification and tracking



References Cited

- Blown to Bits: Your Life, Liberty, and Happiness After the Digital Explosion
Hal Abelson, Ken Ledeen, Harry Lewis, Addison-Wesley, 2008
- The Goal: A Process of Ongoing Improvement, Rev. 3rd Ed.
Eliyahu M. Goldratt, Jeff Cox, North River Press, 2004
- Observed Trends in Spam Construction Techniques: A Case Study of Spam Evolution
Pu and Webb, CEAS 2006
- Spam Double-Funnel: Connecting Web Spammers with Advertisers
Wang et al., WWW2007
- Characterizing Web Spam Using Content and HTTP Session Analysis
Webb et al., CEAS 2007
- Blog Spam: A Review
Adam Thomason, Six Apart, CEAS 2007
- Email Spamming Campaign Analyses: A Campaign-based Characterization of Spamming Strategies
Calais et al., CEAS 2008



Thank You!

- I welcome your comments, questions, & discussion
- Brian D. Davison
davison(at)cse.lehigh.edu