Attack of the Spam Bots- Banish Spammers, Keep Customers

Who: Jake Munson

Company: Idaho Power

Website: http://techfeed.net/blog/

Email: yacoubean@gmail.com

Location: Kuna, ID

What is a spam bot?

- Any kind of spam that comes in through web forms.
  - Comment spam in blogs
  - Feedback forms
  - Registrations forms

How do spam bots work?

- Automated software
  - Directly attack form processor
  - Cached forms
  - http://www.botmaster.net/
    - "This autosubmitter uses a huge database of forums, guestbooks, wikis and blogs to post messages...its ability to work around most types of 'captchas'."
- Manual spammers
  - Armies of cheap labor

Spam bot in action

How do you stop them?

- Remove feedback options
- Moderation queues
- CAPTCHA
  - The user has to prove they are human
- Emerging methods
  - Make the spammer prove *they* aren't a spammer

CAPTCHA

Completely Automated Public Turing test to tell Computers and Humans Apart

The Good

- Can be effective
  - ICR software has difficulty reading the image
- Automated-no moderation is necessary

CAPTCHA In ColdFusion

- CF8's built-in CAPTCHA (<cfimage action="captcha"...>)
- Alagad Captcha-http://www.alagad.com/index.cfm/name-captcha
- Lyla Captcha-http://lyla.maestropublishing.com/

CAPTCHA

The Bad

- Accessibility problems
  - Captcha is designed to defeat automated screen readers
  - Blind people use screen readers
- #1 web design rule: “Don't make me think”-Steve Krug
  - Captcha is designed to make the user think, which is bad for usability
  - Some Captchas are so difficult the user needs to make multiple attempts
- More and more bots break CAPTCHA
  - Youtube problem: “spammers have used intelligent character recognition (ICR) software to circumvent the verification system commonly known as Captcha”

Programmatically Identify Spammers

Users are innocent until proven guilty.

Body of Evidence to Prove Innocence

- Mouse movement
- Keyboard usage
- Empty hidden field is empty
- Normal time to fill out form
- 1 or less URLs in form contents
- Suspect IP address
- Form contents are not “spammy”

Mouse Movement

Users move mice, spam bots don't

Keyboard Usage

Users bang on keyboards, spam bots don't

4 More Key Clues

The evidence is starting to pile up

- Empty hidden field is empty
  - Spammers fill out all fields
- Normal time to fill out form
  - Software is a lot faster than users
- 1 or less URLs in form contents
  - Spammers like to...well, spam
  - Dave Shuck's idea
- IP address is from a known spammer
  - http://www.projecthoneypot.org/

Project Honeypot Demo

The Final Straw

If all else fails, call in the Dream Team

- If you want to use any of these ideas, use Akismet
  - http://www.akismet.com/
- Similar to virus definitions
  - You send form contents to a web service, it returns true or false
  - Compares form contents to vast database of known form spam
  - Community of web developers contributes to database
- Extremely accurate

If it walks like a duck...

Users don't do spammy things

- Each test is unreliable by itself
- Many tests together can identify spammers
- CFFormProtect
  - http://cfformprotect.riaforge.org/
- Others are doing it
  - Ben Nadel-http://bennadel.com/index.cfm?dax=blog:405.view
- Be creative!

Questions?

Who: Jake Munson

Website:
http://techfeed.net/blog/

Email: yacoubean@gmail.com