CATCHING CLICK-SPAM IN SEARCH AD NETWORKS

Vacha Dave †, Saikat Guha★ and Yin Zhang *

† University of California, San Diego  
★ Microsoft Research India  
* The University of Texas at Austin
Internet Advertising Today

- Online advertising is a 40 billion dollar industry *
- Advertisers can reach a massive audience
- Publishers can monetize traffic
  - Blogs, News sites, Syndicated search engines
  - Revenue for content development
- Pay-per-click advertising

*Based on Interactive Advertising Bureau Report, a consortium of Online Ad Networks
Pay-per-click advertising

Publisher

User visits and publisher site

Ad network

Ad Request

Ad Content + URL

Advertiser billed

Advertiser

$7

Publishers make 70% cut

$10
Click-spam in Ad Networks

- Click-spam
  - Fraudulent or invalid clicks
  - Users delivered to the advertiser site are uninterested
  - Advertisers lose money

- Ant-smasher
  - Squish the ant to win the game
  - Ads close to where user is expected to click
Evolution of click-spam

Google gets gruff over click fraud

By Stefanie Olsen
Staff Writer, CNET News

Google filed a lawsuit against an Internet operation that it claims systematically clicked on text ads to defraud its advertising network.

The case, filed Nov. 15 in the Superior Court of Santa Clara County in California, is among the first civil lawsuits to relate to click fraud. The lawsuit charges that Texas-based Auctions Expert International signed up to display Google's targeted text advertising on its Web site, and then fraudulently clicked on the ads to profit from its pay-per-click system.
ViceROI: Click-spam mitigation algorithm
- Can be used by ad network
- Looks at the financial motives
- Catches diverse click-spam attacks

Let us begin by looking at an example
- sophisticated botnet driven click-fraud
Malware driven click fraud

Jane searches for books

Botmaster generates list of publishers

Publisher List

Publisher URL

User wouldn’t know the malware was doing click-fraud

Auto-Redirect (Fraud)

AD URL
Malware driven click fraud

- Malware: TDL4

- Peculiar behavior:
  - Can intercept and redirect all browser requests
  - Only 1 click per IP address per day
  - Gates clicks on user actions

- Why?
  - Tries to evade possible rules that an ad network have
    - Javascript – CSS based signatures
    - IP thresholds
    - Timing analysis – (e.g. when is a user most active)

- Defending against click-spam is getting hard
Conversions as a signal

- Conversions are desirable actions
  - On advertiser page: Email sign-up, purchase etc.
  - Conversion tracking is an optional service
    - Pixel on the checkout page

- Using conversion to gauge traffic quality
  - Cost-per-action (CPA) payment model
  - Conversion discounted Cost-per-click (Smart pricing)
    - Discount clicks from publishers that don’t convert
Conversions being gamed too…

- Experiment
  - Bluff ad
    - Concentrate bad traffic[1]
  - Bluff form
    - Garbage form

- Over 200 form fills
  - In a week

- Means
  - Automated
  - Human assisted
  - Crowd-sourced

Gaming Conversions: Conversion Fraud

- Click-spammers now *generate conversions*
  - On non-financial advertisers
    - Email signups, form filling, CAPTCHAs

- Financial conversions don’t work either
  - Stolen credit card can be used

- Conversions don’t solve the problem
  - Need to go back to basics
Follow the money

- Click-spammers exist to make money
- Clicks, conversions are only side effects
  - Can be gamed
- Key idea: Follow the money trail

- Click-spammers need to pay to acquire users
  - Rent-a-bot, install browser plugin
- Use acquired user aggressively
Milking the users: Ad injectors

User searches for ACM membership in search engine

Ad: whatisthesymbol.com
Query: ACM membership

After install,
Acts as a publisher
 Inject ads in all websites
Milking the users: Search Hijacking

User has a Search toolbar bundled with browser

Entire area clickable

Show ads for all queries - Informational
Different Attacks: one goal

- Click-spam **turns profit for spammer**
  - Cost: Rent-a-bot, pay-per-install cost
  - Revenue: click payout

- Click-spam **carries inherent risk**
  - Arrest - E.g. Operation Ghost click [1]
  - Take down

- **Strategy**: use acquired user aggressively

- **Signature**: Extremely high revenue/user for a publisher
  - Regardless of means of click-fraud
  - As seen by the ad network

[1] Seven charged in malware-driven click fraud case, Ars Technica, Nov 2011
ViceROI: Key Challenges

- Publisher diversity
  - Diverse business models
    - Search engines, blogs, online retailers
  - Different volume scale
    - Blog sites to large companies
  - No single revenue/user number

- Click-spammers mix good and bad traffic
  - For covering bad traffic
ViceROI: Intuition

Revenue (log scale)

User Percentile

Several orders of magnitude

Ethical Publishers
Click-spammers
Mixed traffic
Contributions

- **ViceROI**
  - Single algorithm to catch diverse click-spam attacks
    - All four attacks described and others
  - No tuning knobs
  - Runs at the ad network

- **Works at Internet scale**
  - Piloted it at a large ad network
  - Across diverse publishers and users

- **Bluff form** for catching conversion fraud
Evaluation

- Ad data from a large ad network
  - Three weeks, millions of clicks
  - Thousands of publishers

- Ground truth
  - Ad network’s own heuristic

- Evaluation Criteria
  - Classifier performance (TP, FP, TN, FN)
  - Compare against existing filtration rules
  - Types of attack caught
Evaluation — TPR vs. FPR

TPR = TP/P, FPR = FP/P
Diverse attacks caught

- Bot driven click-fraud
  - Two different botnets, ZeroAccess and TDL4

- Conversion fraud enhanced click-fraud

- Search Hijacking
  - Toolbar based
  - Browser based
  - DNS based

- Ad injectors

- Parked domains, Arbitrage and others..
Summary

- ViceROI: algorithm to catch click-fraud
  - No tuning knobs
  - Based on click-spammers’ high profit motive
  - To beat ViceROI, spammer must reduce profit

- Good classifier performance

- Catches a wide variety of attacks
  - Malware-driven, conversion fraud, ad injectors and others..

- Piloted at a major ad network
Thanks!
Comparison against existing rules
Precision-Recall Curve
Effect of low intensity bot traffic

- **Search engine**
- **Click-spammer**

Steady Bot traffic
Ad revenue spend [IAB quarterly report]

Ad formats – Q2 2012
Total - $8.7 billion*

Ad formats – Q2 2013
Total - $10.3 billion*

*Approximate figures.