Fraudulent Activity in Online Advertising

Kevin Springborn, Paul Barford

Summer 2013
Motivation

1990’s – flat fee popups

C. 2000 – CPC adwords

Today – CPM display ads
The online ad ecosystem

**Brands**
- Pay for ad placements on websites and apps
- Specify target audience for ad campaigns

**Advertisers**

**Intermediaries**
- Receive payment for each ad placed
- Facilitate ad placements
- Examples: ad servers, ad exchanges, etc.
- Incentives for click and impression fraud

**Publishers**
- Sell ad placements on websites and apps
- Generate content that draws target audiences
- Incentives for click and impression fraud

**Users**
Threats

- Simple threats: script-based page retrieval
  - Ubiquitous - $12/10K impressions
- Mechanical turk
  - Humans who are paid (or requested) to access sites
- More complex threats: botnets
  - Geotargeting, clicks, and other characteristics
  - As much as $100/10K impressions
- New threats: pay-per-view networks
  - Websites that load 3rd party pages in an obfuscated fashion when accessed by users
Web traffic generation

• Search is the original form of traffic generation
  – Search engine optimization (SEO) is a popular service

• Legitimate forms of traffic generation
  – Adwords
  – Content widgets (e.g., OutBrain)

• Type “purchase web traffic” in Google search
  – MANY traffic generation offerings
  – Wide variety of features (geo, uniques, etc.)

• Investigation of 34 target sites
  – Assurance that “no black hat methods” are used
  – Great deal of repackaging and ambiguous DNS registrations
Honeypot websites

- Series of websites developed to be targets for traffic generation
  - Look and feel of a “real” site
- Instrumentation
  - Gather as much data per access as possible
Case study traffic purchases

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Amount</th>
<th>Runtime</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxVisits</td>
<td>10,000</td>
<td>5 days</td>
<td>$11.99</td>
</tr>
<tr>
<td>BuildTraffic</td>
<td>20,000</td>
<td>60 days</td>
<td>$55.00</td>
</tr>
<tr>
<td>AeTraffic</td>
<td>10,000</td>
<td>7 days</td>
<td>$39.95</td>
</tr>
<tr>
<td>BuyBulkVisitor</td>
<td>20,000</td>
<td>5 days</td>
<td>$53.00</td>
</tr>
<tr>
<td>TrafficMasters</td>
<td>50,000</td>
<td>2 days</td>
<td>$70.00</td>
</tr>
</tbody>
</table>

US-based traffic, unique users over 24 hour period
Purchased traffic profiles

BuildTraffic arrival process

TrafficMasters arrival process
Purchased traffic observations

- 69K out of the 110K purchased impressions were delivered
  - BuildTraffic ended 28 days into its 60 day cycle
  - AeTraffic is still delivering at a low level
  - BuyBulkVisitor delivered 1 impression
- Tiny (1%) overlap with public IP blacklists
- Virtually no interactions on our site
- Only modest reuse of IP addresses
- Inconsistent ad loads
- Most traffic was from pop-unders
  - Only 53% of view had non-zero frame size
PPV networks

• Traffic generation objective: look like a real user (and offer features at low price)
• Some TG services offer a simple tag that when included on a site pays attractive CPM
  – “…will not block any of your site content and does not lead to actions where users might be led to leave your site”
  – Tag will “display” 3rd party websites
• PPV network: groups of sites that run tags from a single TG service
Deep dive into PPV nets

• When a user accesses a site running a PPV tag a pop-under window is generated
  – Typically requires a user action
• Pop-under calls PPV network server
  – Delivers details on user and site
• PPV network will deliver URL’s of sites buying traffic
  – Often to 0 height frames
  – Frequent reloads
Scope of PPV nets

• Many PPV sites publish their volume
  – Average of 17.16M unique visitors and 6.29B page views per provider per day are claimed

• We identified 10 candidate PPV tags from review of purchased traffic and online forums

• We searched Jan-June ‘12 Common Crawl DB for publisher sites that run PPV tags from 10 providers
  – Javascript checked for pop-under generation

• Over 4M PPV tags found on over 11K domains
  – Largest: ero-advertising.com on 5.8K domains
Impact of PPV nets

• MuStats used to estimate daily page views on identified pages (domains and subdomains)
  – Over 168M estimated daily page views

• Assumptions
  – 25% of pop-unders are blocked
  – 4 destination sites per pop-under
  – Destination sites run 4 ads @ $0.25 CPM

• Over $15M/month in wasted ad spend from 10 PPV networks alone!
Mitigating PPV nets

- **Viewport size filters**
  - Advertisers or intermediaries that run ad servers can augment their Javascript to check viewport
  - Would filter 46% of impressions in our data

- **Referrer blacklist**
  - Block traffic originating from PPV networks
  - Would block 99% of our purchased traffic

- **Publisher blacklist**
  - Identify publishers that participate in PPV nets
  - Block ad requests from those publishers
  - Would discourage publishers from PPV nets
Summary and status

• Fraud is a huge problem in online advertising
  – From simple scripts to sophisticated bots to PPV networks
• Ongoing honeypot-based data gathering
• m.Labs has developed TQ platform to identify fraud
• m.Labs open experimental platform
  – Data repository and API for ad fraud detection
• New directions for data gathering and fraud analysis
Thank you!

- Paul Barford
- Broadcast Interactive Media

- Kevin Springborn (kevin@mdotlabs.com)