Monetizing ZeroAccess

Characterizing Large-Scale Click Fraud

Paul Pearce
University of California, Berkeley

With: Vacha Dave (Microsoft), Chris Grier (Berkeley/ICSI), Kirill Levchenko (UCSD), Saikat Guha (Microsoft Research India), Damon McCoy (George Mason), Vern Paxson (Berkeley/ICSI), Stefan Savage (UCSD), and Geoffrey M. Voelker (UCSD)
Overview

• Goal: Illuminate the nature and behavior of large scale click fraud
  – How does modern click fraud look at scale?
  – \( \rightarrow \) $$$

• Our lens: ZeroAccess
  – Peer-to-peer command & control, fraud payloads
  – Measurement opportunities

• Takedown and Resurrection
  – Leveraging impulse events can enable fraud detection and measurement

• Aggregate click fraud behavior
  – Driving towards estimating lost advertiser $$$
ZeroAccess: A Malware Delivery Platform

• Core ZA: Simply a mechanism to distribute other pieces of malware
  – Payload decoupled from infection
• Estimated size: 1.9 million (Mid 2013, Symantec)
• ZA’s payload monetization strategy has evolved with changes in the underground economy
  – 4 known monetization strategies across 5 years
• Click Fraud is the dominant form of monetization
Online Advertising: Click Anatomy
Online Advertising: Click Anatomy
Online Advertising: Click Anatomy

[$$$]
Online Advertising: Click Anatomy
Online Advertising: Click Anatomy
Online Advertising: Click Anatomy
Online Advertising: Click Anatomy

MyBlog.com  bing Ads  amazon.com

JS To Show Ads  Ad To Serve

$  $$$  $$$$
Online Advertising: Primer

Time

$ MyBlog.com

$$ bing Ads

$$$$ amazon.com


Online Advertising: Primer

Time

Page Visit

MyBlog.com

bing Ads

amazon.com

$ $$ $$$ $$$$
Online Advertising: Primer

- Page Visit
- Page Request
- Page w/ JS
- JavaScript requests Ad
- Log Impression

Time

MyBlog.com

$ $ $$$

bing Ads

amazon.com
Online Advertising: Primer

Time

Page Visit

Page Request

Page w/ JS

JavaScript requests Ad

Returns Ad

Log Impression

MyBlog.com

$ $

bing Ads

$$$

amazon.com

$$$$
Online Advertising: Primer

- Page Visit
- Page Request
- Page w/ JS
- JavaScript requests Ad
- Returns Ad
- Log Impression
- User Ad Click

Time

MyBlog.com

$$_$

$$$

$$_$$$

bing Ads

amazon.com
Online Advertising: Primer

Time

Page Visit

Page Request

Page w/ JS

JavaScript requests Ad

Returns Ad

User Ad Click

Ad Click Request

Log Impression

Log Ad Click

MyBlog.com

bing Ads

amazon.com

$ $ $$ $$
Online Advertising: Primer

- **Page Visit**
- **Page Request**
- **Page w/ JS**
- **JavaScript requests Ad**
- **Returns Ad**
- **Log Impression**
- **User Ad Click**
- **Ad Click Request**
- **Redirect**
- **Log Ad Click**
- **Page Visit**
- **Advertiser Page**

MyBlog.com

$ $ $$ $$
Online Advertising: Primer

Page Visit → Page Request → Page w/ JS → JavaScript requests Ad → Returns Ad → Log Impression → User Ad Click → Ad Click Request → Redirect → Log Ad Click → Page Visit → MyBlog.com → Clicks Buy → Advertiser Page → $ → $$$ → $$$$
Online Advertising: Click Anatomy

- Page Visit
- Page Request
- Page w/ JS
- JavaScript requests Ad
- Returns Ad
- Log Impression
- User Ad Click
- Ad Click Request
- Redirect
- Log Ad Click
- Page Visit
- Advertiser Page
- Conversion Request
- Clicks Buy
- Log Conversion

Complicated Affiliate Ecosystem
Online Advertising: Click Anatomy

MyBlog.com $ bing Ads $$$

amazon.com $$$$
Online Advertising: Click Anatomy

MyBlog.com

bing Ads

amazon.com

$ $ $$ $$$
Online Advertising: Click Anatomy

$ MyBlog.com $ Search.com $$$$ bing Ads $$$$ amazon.com
Online Advertising: Click Anatomy
Online Advertising: Click Anatomy

MyBlog.com

- Relationships with traffic sources
- Relationships with advertisers and ad networks

$ $ $ $ 

MyBlog.com

- MyBlog.com
- Search.com
- Bing Ads
- Amazon.com

Relationships with advertisers and ad networks
Online Advertising: Click Anatomy

- Relationships with traffic sources
- Relationships with advertisers and ad networks
Online Advertising: Click Anatomy

- Click fraud is:
  - Delivering bogus traffic to advertiser pages
    - Impressions, Clicks, and/or conversions
Online Advertising: Click Anatomy

- Click fraud is:
  - Delivering bogus traffic to advertiser pages
    - Impressions, Clicks, and/or conversions

Click fraud points:
- MyBlog.com
- Search.com
- Bing Ads
- Amazon.com

Relationships with traffic sources
Relationships with advertisers and ad networks
Fraud Pain Points
Online Advertising: Click Anatomy

- Click fraud is:
  - Delivering bogus traffic to advertiser pages
    - Impressions, Clicks, and/or conversions
- Middle men can obscure badness from ad network visibility
How ZA Works: Peer-to-peer C&C
How ZA Works: Peer-to-peer C&C
How ZA Works: Peer-to-peer C&C
How ZA Works: Peer-to-peer C&C
Click Fraud Payloads

- Two distinct click fraud payloads
  - Each with distinct C&C servers separate from P2P
- z00clicker
  - Produces high velocity, low quality clicks
  - Ads not visible to users
  - No chance of conversion
- Search Engine Result Page (SERP) hijacker: Serpent
  - More sophisticated fraud model
  - Intercepts user search queries
  - Higher chance of conversion ➔ $$$
# Serpent: Detailed Behavior

<table>
<thead>
<tr>
<th>Browser</th>
<th>Serpent</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Google Search" /></td>
<td><img src="image" alt="Search Engine" /></td>
</tr>
</tbody>
</table>

- **Browser**:
  - ![Google Search](image)
  - ![I'm Feeling Lucky](image)

- **Serpent**: Search Engine
Serpent: Detailed Behavior

Browser

Search Engine

Serpent

Bikes

Google Search  I'm Feeling Lucky
Serpent: Detailed Behavior

Browser

Serpent

Page Fetch

Search Engine
Serpent: Detailed Behavior

Browser

![Google Search]

Bikes ①

Serpent

Page Fetch ②

Search Engine

Serpent C&C (Bikes) ③

Serpent-C&C
Serpent: Detailed Behavior

1. Bikes

2. Page Fetch (Search Results)

3. Serpent C&C (Bikes)
Serpent: Detailed Behavior

1. Bikes
2. Page Fetch (Search Results)
3. Serpent C&C (Bikes) (Ad URLs)

Browser

Serpent

Search Engine

Serpent-C&C
Serpent: Detailed Behavior

Browser

Google Search

Serpent

Page Fetch

(Search Results)

Search Engine

Serpent C&C (Bikes)

(Ad URLs)

Serpent-C&C
Serpent: Detailed Behavior

Browser

Google

Bikes

Search Engine

Page Fetch

(Search Results)

Serpent C&C (Bikes)

(Ad URLs)

Serpent-C&C

Intended Server
Serpent: Detailed Behavior

1. Browser
   - Google search for "Bikes"

2. Page Fetch (Search Results)
   - Serpent fetches search results from the search engine.

3. Serpent C&C (Bikes)
   - Sends traffic to Serpent C&C (Bikes).

4. Intended Server
   - Intended server receives the traffic.

Additionally, the diagram includes:
- Ad URLs:
  - Trek Bicycle advertisement
  - BikeSD advertisement
- San Diego County Bicycle Coalition Home advertisement
Serpent: Detailed Behavior

Browser

1. Bikes

Serpent

2. Page Fetch (Search Results)
3. Serpent C&C (Bikes)
4. (Ad URLs)

Search Engine

Serpent-C&C

Intended Server
Serpent: Detailed Behavior

1. Browser
   - Google search for "Bikes"

2. Serpent
   - Page Fetch
   - (Search Results)

3. Serpent C&C (Bikes)
   - (Ad URLs)

4. Browser
   - Click on search result

5. Intended Server
   - Page Fetch

6. Ad Server
Serpent: Detailed Behavior

1. **Browser**
   - Google search for "Bikes"

2. **Serpent**
   - Page fetch
   - (Search Results)

3. **Serpent C&C (Bikes)**
   - (Ad URLs)

4. **Advertising Victim**
   - Ad website

5. **Intended Server**

6. **Ad Website**
   - Page fetch

- Ad Server
Serpent: Detailed Behavior

Browser

1. Bikes

2. Google Search

3. Serpent C&C (Bikes)

4. Click on search result

5. Page Fetch

6. Advertising Victim

Search Engine

Serpent

C&C Milking Opportunity

(Search Results)

Serpent C&C

(Ad URLs)

Intended Server

Ad Website

Ad Server

SALE!
Serpent: C&C Milking

• Reverse engineered the C&C
• Once we understood the C&C, we could interact with it without running malware
• Safely clicked on a small number of the ads
• Goal: Understand the ecosystem and composition of the ad fraud
  — Critical in formulating a global fraud estimation
Serpent: Detailed Behavior
The Takedown

• December 5th, ZeroAccess Serpent and z00clicker C&C servers seized simultaneously
• Several hours later, updated payloads distributed via P2P
  – Partially resuming fraud
• These impulse events enable fraud attribution within advertising networks
Serpent: Detailed Behavior

Browser

Serpent

Search Engine

Serpent C&C (Bikes)

(Ad URLs)

Intended Server

Page Fetch

(Email Results)

Ad Website

Ad Server

Advertising Victim
Serpent: Detailed Behavior

1. Browser (Search Engine)
2. Serpent (Serpent C&C)
3. (Search Results)
4. Serpent C&C (Bikes)
5. (Ad URLs)
6. Intended Server
7. Page Fetch
8. Ad Website
9. Ad Server
10. Advertising Victim
Aggregate Ad Behavior

• What can we say about the significance of ZA fraud?

• How can we do this?
  – Collaboration with a large real-world ad network that we observed in our milking

• Send our IPs and chains off to our ad network partner
  – What do we find?
Ad Network Background

• Key Concept:
  – Ad Units

• Ad units loosely correspond to distinct traffic resellers that have a syndication relationship with our ad network partner
  – Subsyndication: These resellers may buy traffic from other parties
Ad Unit Basics
Ad Unit Basics
Ad Unit Basics
Ad Unit Basics
## Putting It All Together

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Quantity</th>
<th>Dates (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supernode Peers (Contacts)</td>
<td>260,811,204</td>
<td>Dec 1-6</td>
</tr>
<tr>
<td>Serpent DNS Queries (/24 Buckets)</td>
<td>16,208,758</td>
<td>Dec 1-4</td>
</tr>
<tr>
<td>Milker Click Chains</td>
<td>1,766</td>
<td>Sept 10-Dec 5</td>
</tr>
<tr>
<td>Ad Click Data (/24 Buckets)</td>
<td>Over 10TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
<tr>
<td>Ad Unit Data (/24 Buckets)</td>
<td>Around 2TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
</tbody>
</table>
# Putting It All Together

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Quantity</th>
<th>Dates (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supernode Peers (Contacts)</td>
<td>260,811,204</td>
<td>Dec 1-6</td>
</tr>
<tr>
<td>Serpent DNS Queries (/24 Buckets)</td>
<td>16,208,758</td>
<td>Dec 1-4</td>
</tr>
<tr>
<td>Milker Click Chains</td>
<td>1,766</td>
<td>Sept 10-Dec 5</td>
</tr>
<tr>
<td>Ad Click Data (/24 Buckets)</td>
<td>Over 10TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
<tr>
<td>Ad Unit Data (/24 Buckets)</td>
<td>Around 2TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
</tbody>
</table>
# Putting It All Together

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Quantity</th>
<th>Dates (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supernode Peers (Contacts)</td>
<td>260,811,204</td>
<td>Dec 1-6</td>
</tr>
<tr>
<td>Serpent DNS Queries (/24 Buckets)</td>
<td>16,208,758</td>
<td>Dec 1-4</td>
</tr>
<tr>
<td>Milker Click Chains</td>
<td>1,766</td>
<td>Sept 10-Dec 5</td>
</tr>
<tr>
<td>Ad Click Data (/24 Buckets)</td>
<td>Over 10TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
<tr>
<td>Ad Unit Data (/24 Buckets)</td>
<td>Around 2TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
</tbody>
</table>
## Putting It All Together

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Quantity</th>
<th>Dates (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supernode Peers (Contacts)</td>
<td>260,811,204</td>
<td>Dec 1-6</td>
</tr>
<tr>
<td>Serpent DNS Queries (/24 Buckets)</td>
<td>16,208,758</td>
<td>Dec 1-4</td>
</tr>
<tr>
<td>Milker Click Chains</td>
<td>1,766</td>
<td>Sept 10-Dec 5</td>
</tr>
<tr>
<td>Ad Click Data (/24 Buckets)</td>
<td>Over 10TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
<tr>
<td>Ad Unit Data (/24 Buckets)</td>
<td>Around 2TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
</tbody>
</table>
## Putting It All Together

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Quantity</th>
<th>Dates (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supernode Peers (Contacts)</td>
<td>260,811,204</td>
<td>Dec 1-6</td>
</tr>
<tr>
<td>Serpent DNS Queries (/24 Buckets)</td>
<td>16,208,758</td>
<td>Dec 1-4</td>
</tr>
<tr>
<td>Milker Click Chains</td>
<td>1,766</td>
<td>Sept 10-Dec 5</td>
</tr>
<tr>
<td>Ad Click Data (/24 Buckets)</td>
<td>Over 10TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
<tr>
<td>Ad Unit Data (/24 Buckets)</td>
<td>Around 2TB of raw logs</td>
<td>Nov 28-Dec 6</td>
</tr>
</tbody>
</table>
Applying Taint and Takedown Dynamics

Pre Takedown Click Volume vs ZA Taint
Applying Taint and Takedown Dynamics

Pre Takedown Click Volume vs ZA Taint
Applying Taint and Takedown Dynamics

Pre-Takedown Click Volume vs ZA Taint
Applying Taint and Takedown Dynamics
Applying Taint and Takedown Dynamics

Pre Takedown Click Volume vs ZA Taint

Takedown Click Volume vs ZA Taint
Applying Taint and Takedown Dynamics

Pre Takedown Click Volume vs ZA Taint

Takedown Click Volume vs ZA Taint

Ratio of ZA tainted subnets, pre H hour

H/[H-1]
Applying Taint and Takedown Dynamics

Pre Takedown Click Volume vs ZA Taint

Takedown Click Volume vs ZA Taint

Taint + Dropoff Promising, but noisy
Second Impulse Volume

H–Hour Click Volume Ratio VS R–Hour Ratio
Second Impulse Volume

H–Hour Click Volume Ratio VS R–Hour Ratio

Takedown day R / \[R

Takedown day H / \[H

Hour Click Volume Ratio VS R–Hour Ratio
Combining Impulses Holds Promise
Second Impulse Taint

ZA Taint Prior To Takedown

ZA Taint Hours H→R−1 vs ZA Taint Hours R→R+4
Second Impulse Taint
Second Impulse Taint

ZA Taint Prior To Takedown

ZA Taint After Takedown Vs After New Module
Second Impulse Taint

ZA Taint Prior To Takedown

ZA Taint After Takedown Vs After New Module
Second Impulse Taint

Taint Over Time Instrumental
Challenges and Estimations

• Manually verified ad units identified from taint analysis
• 54 ad units identified with high confidence
  – Fraud clearly taints other ad units in small amounts not included in this total
• Ad partner estimates ZA fraud they suffered likely 2x higher
Challenges and Estimations Con’t

• Hard to extrapolate across ad networks
  – Milking can help, but has limitations
• Ad network partner estimates ZA clicks were worth 10-30 cents
• With these limitations, global fraud impact of ZeroAccess estimate:
  – $100,000 per day
• See paper for more in-depth discussion of assumptions and limitations
Discussion and Conclusion

• Modern web advertising ideally suits cybercrime
• Subsyndication is a problem
  – A single trusted syndicator can open the gates to fraudulent subsyndicators
  – Evidence of this in mixed but dirty ad units
• Attribution is hard
  – We encountered redirection chains up to 14 entities deep
• Low signal to noise ratio
  – Analysis was not possible without the introduction of the takedown impulses
  – The ability to leverage impulses from ecosystem disruptions can sharpen the signal
Questions?

pearce@berkeley.edu