TraffickStop: Detecting and Measuring Illicit Traffic Monetization Through Large-scale DNS Analysis

Baojun Liu, Zhou Li, Peiyuan Zong, Chaoyi Lu, Haixin Duan, Ying Liu, Sumayah Alrwais, Xiaofeng Wang, Shuang Hao, Yaoqi Jia, Yiming Zhang, Kai Chen and Zaifeng Zhang
Illicit Traffic Monetization

How Pay-Per-View Networks Cost Advertisers $180 Million A Year In Impression Fraud

A significant percentage of the top 100 online pay-per-view (PPV) networks that perpetrate impression fraud are using ad secure platform recently spun off from


'Bigggest Ad Fraud Ever': Hackers Make $5M A Day By Faking 300M Video Views

https://www.forbes.com/sites/thomasbrewster/2016/12/20/methbot-biggest-ad-fraud-busted/#64ae66fe4899

JURY ORDERS $2.3 MILLION PAYMENT IN SEARCH-AD CLICK-FRAUD SCHEME

Traffic Network

Connects site owners and affiliates.

Site owner
(Needs traffic)

Traffic Network
(Finds affiliates)

Affiliate
(Refers traffic)
Traffic Network

Connects site owners and affiliates.

- eCommerce Network
- Advertising Network
- Navigation Network
Cheating in Traffic Networks

Cheaters earn profit from site owners using invalid traffic.
Cheating in Traffic Networks

Cheaters earn profit from site owners using invalid traffic.

A fraudulent site (FS) redirects user traffic to a program site (PS) of a traffic network.

The process violates rules of traffic networks.
Cheating happens EVERYWHERE!

**Client-side:**
Browser Hijacking

- Install PUP / Malware on client machines
- Reroute user traffic to targeted sites

---

**Adware.Yontoo**

**Caused $8M loss in 2013**

Adware.Yontoo is Malwarebytes' generic detection name for a large family of adware targeting Windows systems.

https://blog.malwarebytes.com/detections/adware-yontoo/
Transport-layer: ISP Injection

Inject extra ads into web responses

Mitigation: HTTPS
Relies on adoption rate

https://techscience.org/a/2015103003/
Cheating happens EVERYWHERE!

Server-side: Search Ad Impersonation

Publish fake ads in search engines

Impersonate popular brands to trap more users
Cheating happens EVERYWHERE!

Client-side: Browser Hijacking
- Install PUP / Malware on client machines
- Reroute user traffic to targeted sites

Transport-layer: ISP Injection
- Inject extra ads into web responses
- Mitigation: HTTPS
  - Relies on adoption rate

Server-side: Search Ad Impersonation
- Publish fake ads in search engines
- Impersonate popular brands to trap more users
A fraudulent site (FS) redirects user traffic to a program site (PS) of a traffic network.

The process violates rules of traffic networks.
Previous Works

“Active” approaches.

Honey ads [Dave 2012]

Inspection JS [Reis 2008, Thomas 2015]

Network probe [Dagon 2008, Kuhrer 2015]

Require deep involvement of publisher websites

Work on only one type of traffic fraud
Our approach: Passive Analysis
Ground Truth Collection

Manually collect **151 FSes** for empirical study.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Ad Impersonation</td>
<td>Cases from four-month Baidu search results of popular brand products</td>
<td>57 FS</td>
</tr>
<tr>
<td>Browser Hijacking</td>
<td>Cases from online posts and tech forums</td>
<td>50 FS</td>
</tr>
<tr>
<td>ISP Injection</td>
<td>Collected by custom Flash advertisement</td>
<td>44 FS</td>
</tr>
</tbody>
</table>
Key Features of FS

Manually collect **151 FSes** for empirical study.

Webpage of bd.114la6.com, a typical FS

**Key Feature 1:** AUTOMATIC & IMMEDIATE redirection to program sites.

**Result:** Strong domain correlation
Key Features of FS

Manually collect 151 FSes for empirical study.

Key Feature 2:
The page only performs redirection, without anything else.

Result:
Meaningless content
TraffickStop: Passive Analysis

Data Collection
- Passive DNS & DNS logs
- http://
- WHOIS

Association Finder
Finds domains with strong correlation

Content Analyzer
Examines suspicious behaviors between domains
Association Finder

Find domain pairs \{X, Y\} with strong correlation.

Criteria

A. X and Y appear together with high frequency

B. When X is observed, Y can be observed with high probability

C. The visit interval between X and Y is small

Metric

- support
- confidence
- decay

Association analysis
Association Finder

Implementation: **FP-Growth algorithm** with MapReduce.

Algorithm 1 Pair discovery based on FP-Growth.

**Input:** Sorted DNS data

**Output:** Rule, confidence, support

1. function MERGE(Group_source)
2. for uniq_dest ∈ destination_set do
3.     confidence ← SUM_VALUE uniq_dest / source.support
4.     Rule[uniq_dest] ← uniq_dest.support, confidence
5. return Rule
6.
7. Procedure: Map
8. for DNS_Sequence ∈ DNS_database do
9.     while index < DNS_Sequence.length do
10.    source ← DNS_Sequence[index]
11.    session ← DNS_Sequence[index-window, index+window]
12.    for destination ∈ session do
13.        value ← DECAY source.location, destination.location
14.        Out: source, destination, value
15. index ++
16.
17. Procedure: Reduce
18. Group_source ← GROUPBY(source)
19. Rule ← MERGE Group source
20. Rule_group ← FILTER RULE Rule, minsup, minconf
21. for rule ∈ Rule_group do
22.     Out: source_domain, destination_domain, confidence, support

Map procedure:
Calculate the interval between two domain visits

Reduce procedure:
Calculate the frequency of domain pairs, to find those highly correlated.
Content Analyzer

Examine **Redirection + Meaningless content**.

- **Suspicious Domain**
- **Program Site**
- **Strong correlation**
- **Top 10 URLs**
- **Dynamic crawler**
- **Webpages**
- **If redirect to...**
- **Advertising**
- **Content-based clustering**
- **eCommerce Navigation**
- **FS**

**URL dataset**
System Evaluation

Detect three types of fraud at a time.

2-week DNS logs (231 billion requests)

Validation Rules:
A. Serving illegal or unreadable content
B. Forcing redirection
C. URL contains affiliate ID

FS 2,465 fraud URLs

89.4% eCommerce
67.5% Navigation
74.8% Advertising

72.7% accuracy (1,792/2,465)
Measurement & Analysis
1,457 FS SLDs are confirmed by TraffickStop.

1-year passive DNS data
(May 2017 - Apr 2018, ~15% of DNS traffic in China)

53 Billion
Total DNS queries to these FSes

100K+
Queries
96%+ FSes receive each

300+
Days
85%+ FSes are active for
Search Ad Impersonation

Buying ads on search engines to attract visits.

1,457 fraud SLDs

23 Ad fraud SLDs (All redirecting to taobao.com)
Search Ad Impersonation

23 Ad fraud SLDs redirecting to taobao.com.

TABLE V: Query volume of FS in Search Ad Impersonation

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Domain Name</th>
<th>Query Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hao1.dambolofashion.org</td>
<td>314,202</td>
</tr>
<tr>
<td>2</td>
<td><a href="http://www.svnss.com">www.svnss.com</a></td>
<td>232,153</td>
</tr>
<tr>
<td>3</td>
<td><a href="http://www.hxfus.com">www.hxfus.com</a></td>
<td>181,085</td>
</tr>
<tr>
<td>4</td>
<td>hao2.3506ygfs.com</td>
<td>180,063</td>
</tr>
<tr>
<td>5</td>
<td>hao2.csyycsyy.com</td>
<td>131,011</td>
</tr>
</tbody>
</table>

1M+ Total visits

TABLE VI: Number of URLs under each FS

<table>
<thead>
<tr>
<th>FS</th>
<th># URL</th>
<th>FS</th>
<th># URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>hao360.dawanbiao.cn</td>
<td>2,457</td>
<td>hao2.3506ygfs.com</td>
<td>660</td>
</tr>
<tr>
<td><a href="http://www.hxfus.com">www.hxfus.com</a></td>
<td>594</td>
<td><a href="http://www.wlzyx.com">www.wlzyx.com</a></td>
<td>279</td>
</tr>
<tr>
<td>t.iavip.cn</td>
<td>250</td>
<td>vip.1314dian.cn</td>
<td>98</td>
</tr>
</tbody>
</table>

Hundreds of keywords bought under each domain
Economic Loss

Loss = (Total Visits x Traffic Ratio) x Reward x Probability

- taobao.com: $53.8K
- jd.com: $18.9K
- Baidu: $13.3K
- Hao123: $2.5K
- 360 Navigation: $1.0K

Thousands per day dollars lost due to traffic fraud
New Strategy: Ad Reselling

Evading fraud detection of advertising platforms.

Diagram:
- Advertiser
  - Load Ads
  - Check fraud
  - Revenue

- Publisher
  - Load Ads
  - Revenue

- Other sites
  - http://

No Relation
New Strategy: Ad Reselling

Evading fraud detection of advertising platforms.

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Alexa Ranking</th>
<th>Evidence (redirection chain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher-1</td>
<td>~ 200</td>
<td><a href="http://hao.67it.com:86/dfadtz023.js">http://hao.67it.com:86/dfadtz023.js</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://mini.e*s*d*y.com/?qid=sytest23">http://mini.e*s*d*y.com/?qid=sytest23</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://dup.b*i*u*t*t*c.com/js/ds.js">http://dup.b*i*u*t*t*c.com/js/ds.js</a></td>
</tr>
<tr>
<td>Publisher-2</td>
<td>~ 1000</td>
<td><a href="http://t.5txs.cn/rb/i9.js">http://t.5txs.cn/rb/i9.js</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://11.m*d*i*e*s.com/****/baiduAfxId.html">http://11.m*d*i*e*s.com/****/baiduAfxId.html</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.d***.com/union2.html?u207">http://www.d***.com/union2.html?u207</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://cpro.b*i*u*t*t*c.com/cpro/ui/c.js">http://cpro.b*i*u*t*t*c.com/cpro/ui/c.js</a></td>
</tr>
<tr>
<td>Publisher-3</td>
<td>~ 4000</td>
<td><a href="http://m.cnepin.cn/cl/html/jd34.html">http://m.cnepin.cn/cl/html/jd34.html</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://bj.g****.com/content/contentbranch.php">http://bj.g****.com/content/contentbranch.php</a>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://cpro.b*i*u*t*t*c.com/cpro/ui/c.js">http://cpro.b*i*u*t*t*c.com/cpro/ui/c.js</a></td>
</tr>
</tbody>
</table>
Case Study: P2P Traffic Pal

Distributed platform that generate traffic from real users.

“Help me like this post at http://xxx!”

“Help me play this video: http://yyy!”

Clients with this software
Summary

A new passive approach to detect three kinds of illicit traffic monetization

1,457 fraudulent sites detected
72.7% overall accuracy

Measurement on scale, evasion and impact on legitimate parties
TraffickStop: Detecting and Measuring Illicit Traffic Monetization Through Large-scale DNS Analysis

Baojun Liu, Zhou Li, Peiyuan Zong, Chaoyi Lu, Haixin Duan, Ying Liu, Sumayah Alrwais, Xiaofeng Wang, Shuang Hao, Yaoqi Jia, Yiming Zhang, Kai Chen and Zaifeng Zhang