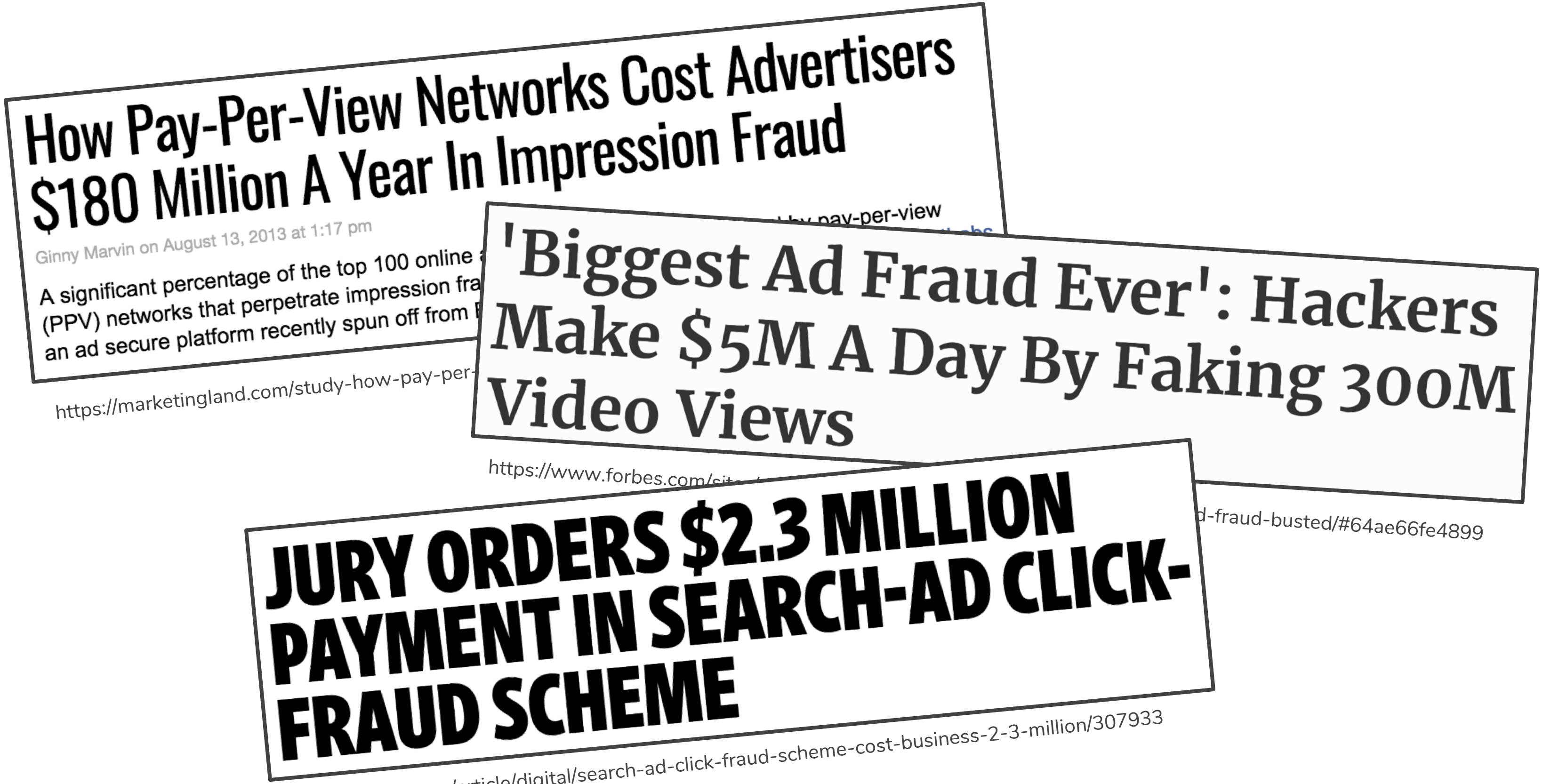
TrafficStop:
Detecting and Measuring Illicit Traffic Monetization Through Large-scale DNS Analysis

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

Ginny Marvin on August 13, 2013 at 1:17 pm

A significant percentage of the top 100 online a
(PPV) networks that perpetrate impression fra
an ad secure platform recently spun off from F
by pay-per-view

https://marketingland.com/study-how-pay-per-

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

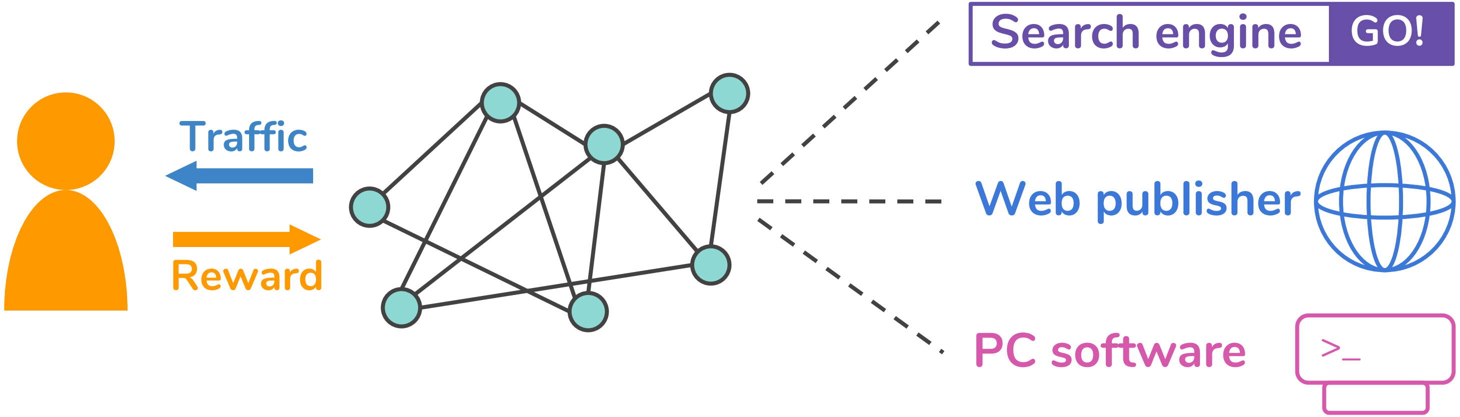
https://www.forbes.com/sit
d-fraud-busted/#64ae66fe4899

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

https://adage.com/article/digital/search-ad-click-fraud-scheme-cost-business-2-3-million/307933

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

- amazonassociates
- ebay partner network
- Rakuten Marketing

Advertising Network

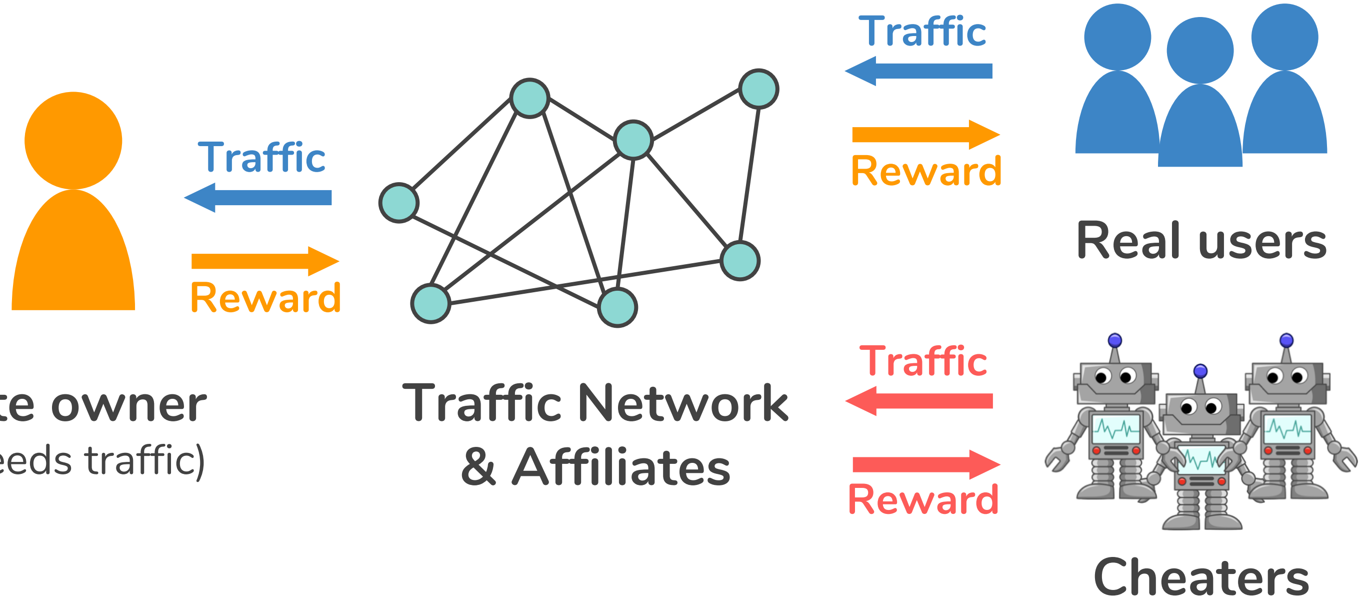
- Google Ads
- Microsoft | Advertising
- media.net

Navigation Network

- hao123
- 360导航

Cheating in Traffic Networks

Cheaters earn profit from site owners using invalid traffic.



Traffic

Reward

Site owner

(Needs traffic)

Traffic Network & Affiliates

Traffic

Reward

Real users

Traffic

Reward

Cheaters

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.

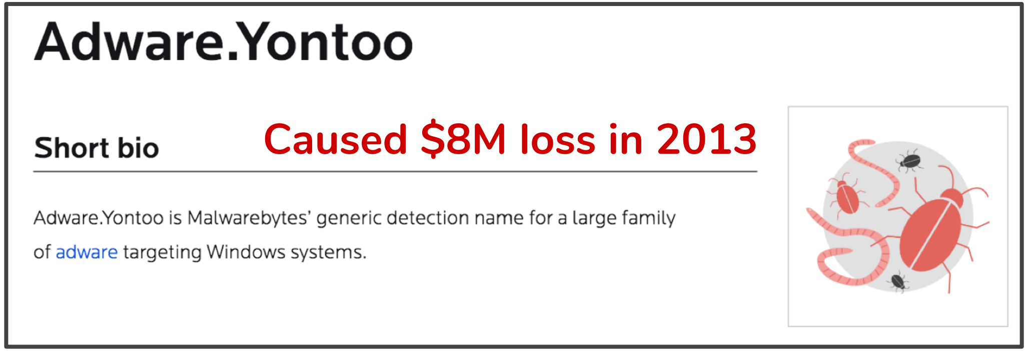
The process violates rules of traffic networks.

Cheating happens EVERYWHERE!

Client-side: Browser Hijacking

Install PUP / Malware on client machines

Reroute user traffic to targeted sites



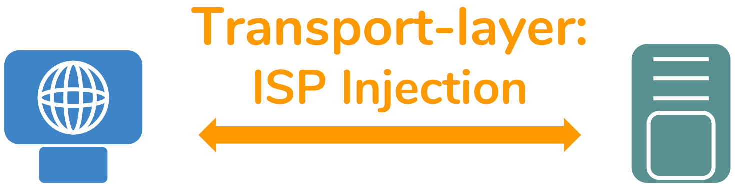
Adware.Yontoo

Short bio Caused $8M loss in 2013

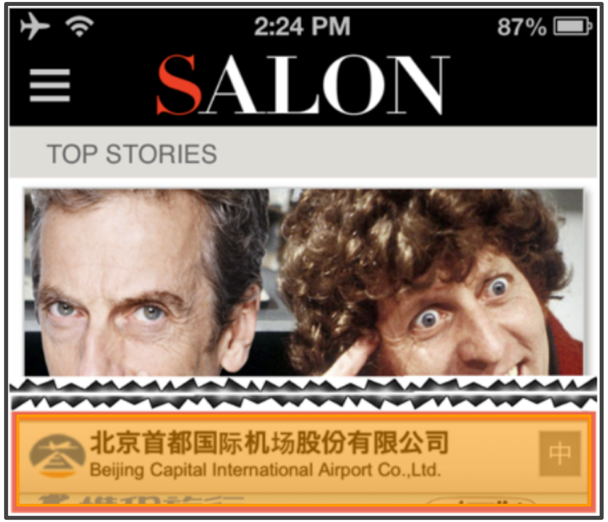
Adware.Yontoo is Malwarebytes' generic detection name for a large family of adware targeting Windows systems.

https://blog.malwarebytes.com/detections/adware-yontoo/

Cheating happens EVERYWHERE!



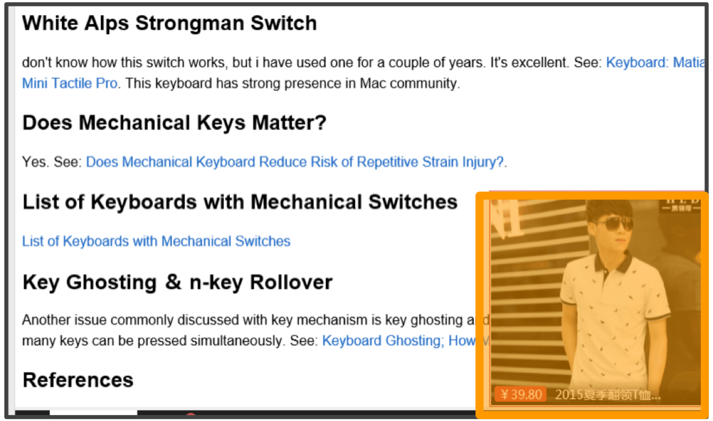
Transport-layer:
ISP Injection


2:24 PM
87%
SALON
TOP STORIES
北京首都国际机场股份有限公司
Beijing Capital International Airport Co.,Ltd.
中

https://techscience.org/a/2015103003/

Inject extra ads into web responses

Mitigation: HTTPS
Relies on adoption rate



White Alps Strongman Switch

don't know how this switch works, but i have used one for a couple of years. It's excellent. See: Keyboard: Matia Mini Tactile Pro. This keyboard has strong presence in Mac community.

Does Mechanical Keys Matter?

Yes. See: Does Mechanical Keyboard Reduce Risk of Repetitive Strain Injury?.

List of Keyboards with Mechanical Switches

List of Keyboards with Mechanical Switches

Key Ghosting & n-key Rollover

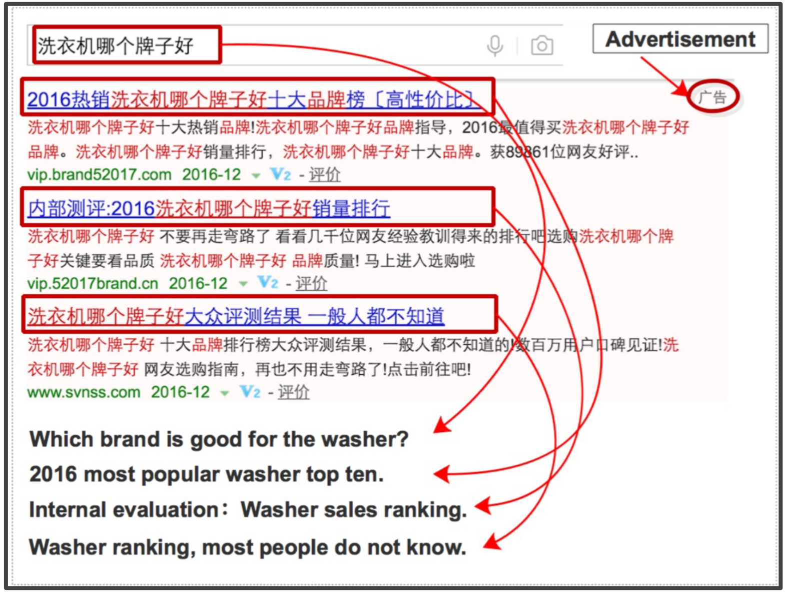
Another issue commonly discussed with key mechanism is key ghosting a d many keys can be pressed simultaneously. See: Keyboard Ghosting; How V

References

http://xahlee.info/w/china_ISP_ad_injection.html

Cheating happens EVERYWHERE!

Server-side:
Search Ad Impersonation



洗衣机哪个牌子好

Advertisement

广告

2016热销洗衣机哪个牌子好十大品牌榜〔高性价比〕

洗衣机哪个牌子好十大热销品牌!洗衣机哪个牌子好品牌指导，2016最值得买洗衣机哪个牌子好品牌。洗衣机哪个牌子好销量排行，洗衣机哪个牌子好十大品牌。获89861位网友好评..

vip.brand52017.com 2016-12 - 评价

内部测评:2016洗衣机哪个牌子好销量排行

洗衣机哪个牌子好 不要再走弯路了 看看几千位网友经验教训得来的排行吧选购洗衣机哪个牌子好关键要看品质 洗衣机哪个牌子好 品牌质量! 马上进入选购啦

vip.52017brand.cn 2016-12 - 评价

洗衣机哪个牌子好大众评测结果 一般人都不知道

洗衣机哪个牌子好 十大品牌排行榜大众评测结果，一般人都不知道的!数百万用户口碑见证!洗衣机哪个牌子好 网友选购指南，再也不用走弯路了!点击前往吧!

www.svnss.com 2016-12 - 评价

Which brand is good for the washer?

2016 most popular washer top ten.

Internal evaluation：Washer sales ranking.

Washer ranking, most people do not know.

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

Previous Works

“Active” approaches.

Honey ads
[Dave 2012]

JavaScript

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

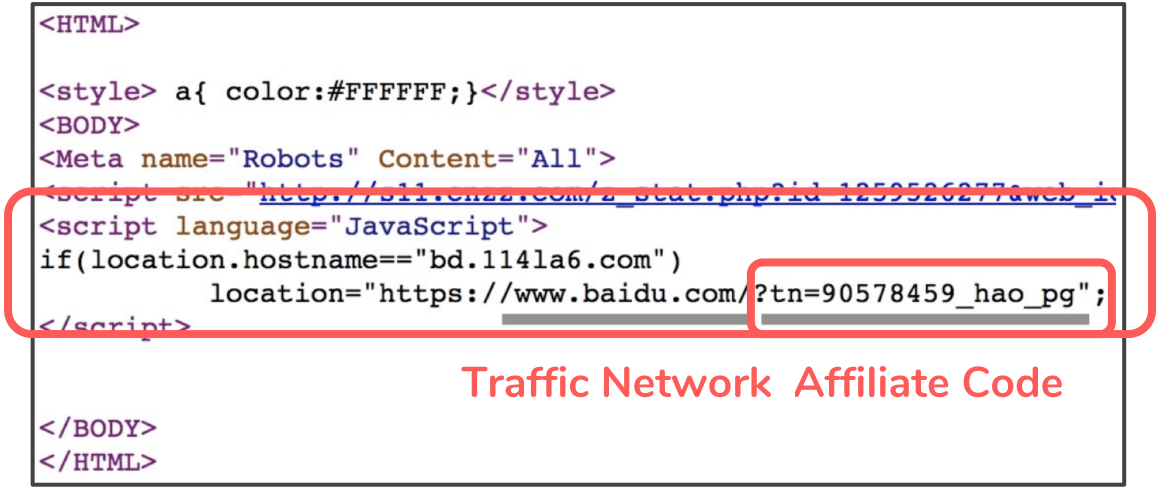
Ground Truth Collection

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

Search Ad Impersonation	Cases from four-month Baidu search results of popular brand products	57 FS
Browser Hijacking	Cases from online posts and tech forums	50 FS
ISP Injection	Collected by custom Flash advertisement	44 FS

Key Features of FS

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



```
<HTML>

<style> a{ color:#FFFFFF;}</style>
<BODY>
<Meta name="Robots" Content="All">
<script src="http://s11.cnzz.com/z_stat.php?id=1259526277&web_i
<script language="JavaScript">
if(location.hostname=="bd.114la6.com")
          location="https://www.baidu.com/?tn=90578459_hao_pg";
</script>


</BODY>
</HTML>
```

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

Key Feature 1:

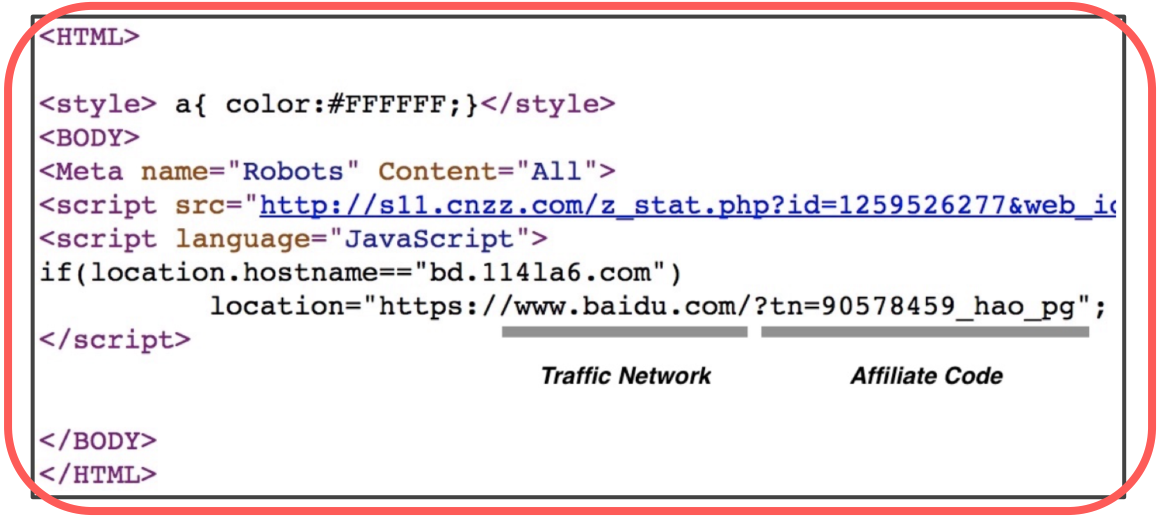
AUTOMATIC & IMMEDIATE redirection to program sites.

Result:

Strong domain correlation

Key Features of FS

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



```
<HTML>

<style> a{ color:#FFFFFF;}</style>
<BODY>
<Meta name="Robots" Content="All">
<script src="http://s11.cnzz.com/z_stat.php?id=1259526277&web_i
<script language="JavaScript">
if(location.hostname=="bd.114la6.com")
          location="https://www.baidu.com/?tn=90578459_hao_pg";
</script>


</BODY>
</HTML>
```

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

Key Feature 2:

The page only performs redirection, without anything else.

Result:

Meaningless content

TraffickStop: Passive Analysis

Data Collection

Passive DNS & DNS logs

URL

WHOIS

Association Finder

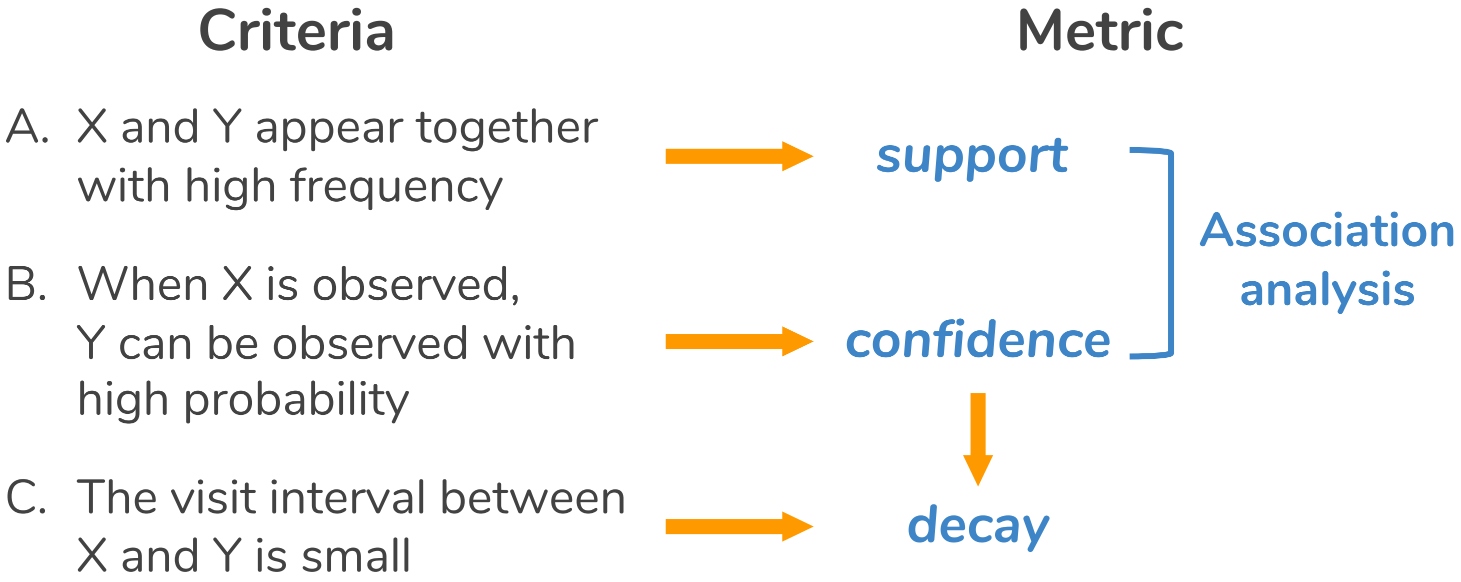
Finds domains with strong correlation

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

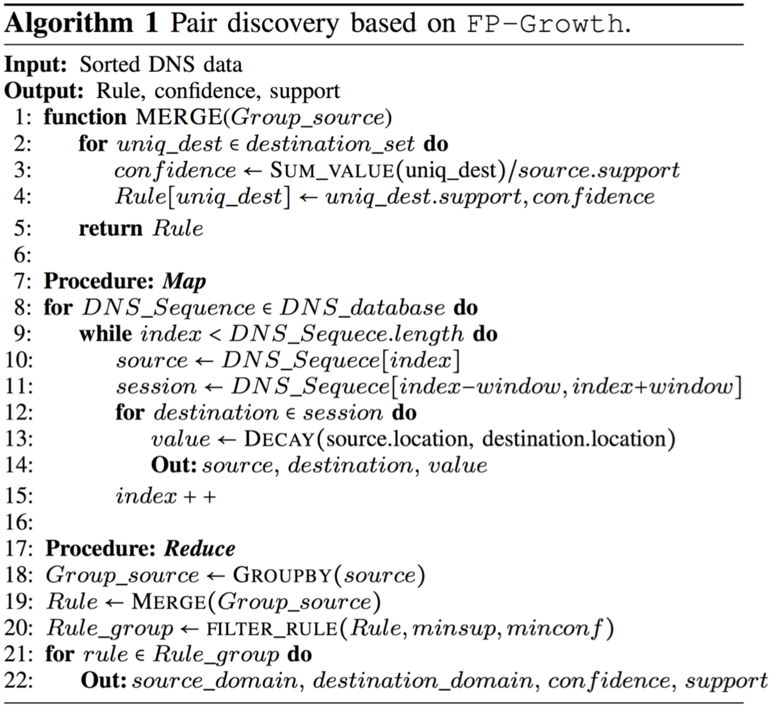
confidence

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

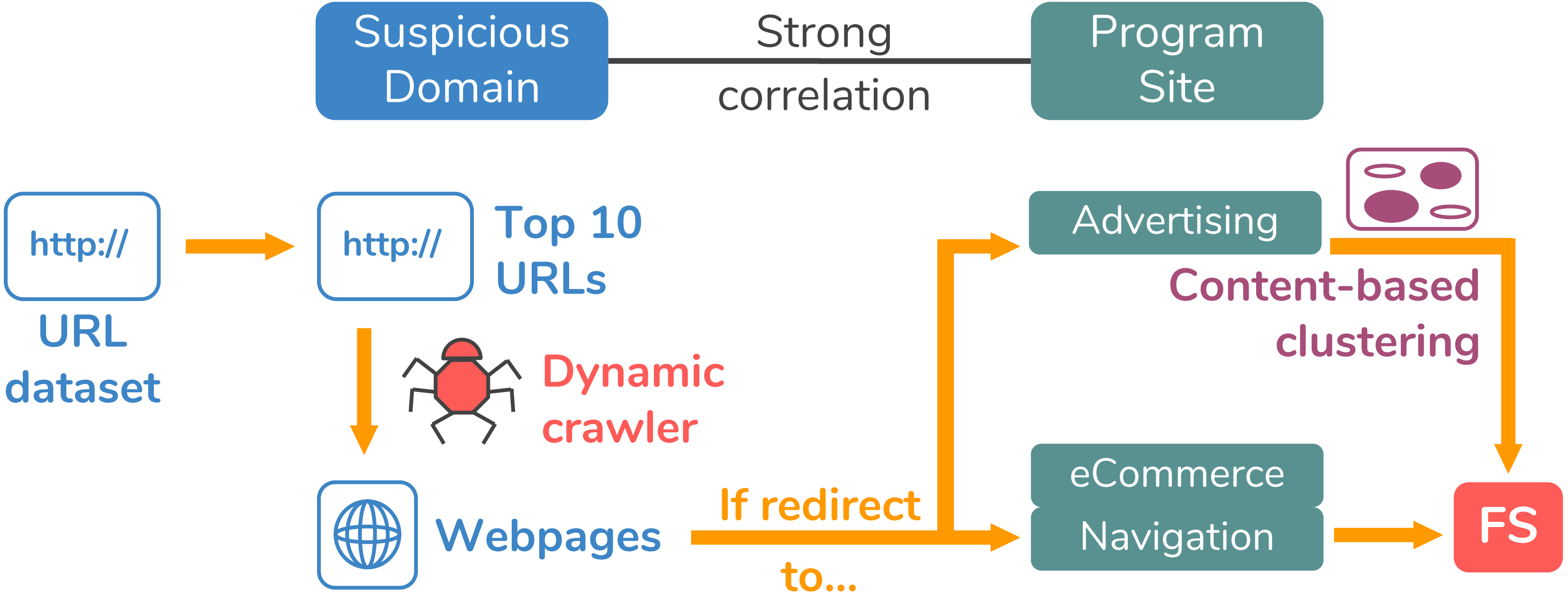
```
function MERGE($Group\_source$)
    for $uniq\_dest \in destination\_set$ do
        $confidence \leftarrow$ SUM_VALUE(uniq_dest)$/source.support$
        $Rule[uniq\_dest] \leftarrow uniq\_dest.support, confidence$
    return $Rule$

Procedure: Map
for $DNS\_Sequence \in DNS\_database$ do
    while $index < DNS\_Sequece.length$ do
        $source \leftarrow DNS\_Sequece[index]$
        $session \leftarrow DNS\_Sequece[index{-}window, index{+}window]$
        for $destination \in session$ do
            $value \leftarrow$ DECAY(source.location, destination.location)
            Out: $source$, $destination$, $value$
        $index++$

Procedure: Reduce
$Group\_source \leftarrow$ GROUPBY($source$)
$Rule \leftarrow$ MERGE($Group\_source$)
$Rule\_group \leftarrow$ FILTER_RULE($Rule, minsup, minconf$)
for $rule \in Rule\_group$ do
    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.

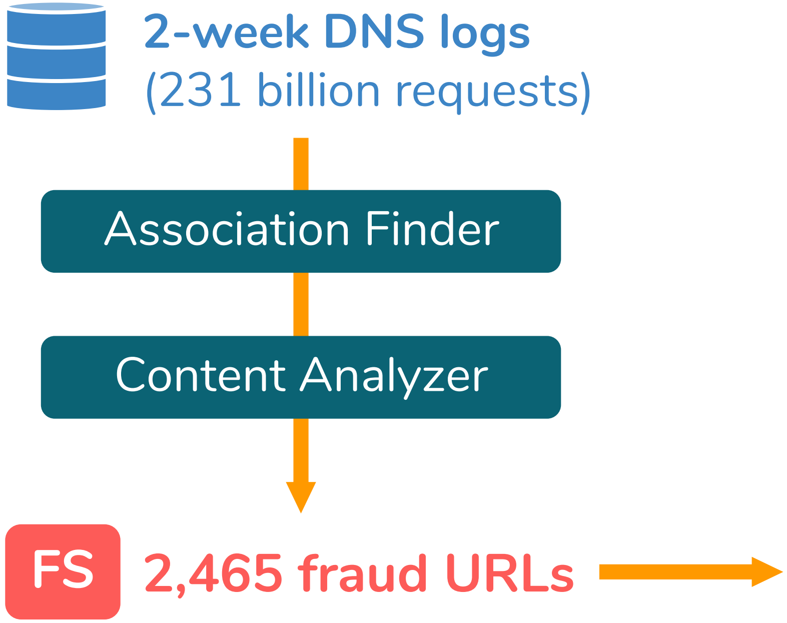
Content Analyzer

Examine **Redirection + Meaningless content.**



System Evaluation

Detect three types of fraud **at a time**.



2-week DNS logs

(231 billion requests)

Association Finder

Content Analyzer

FS 2,465 fraud URLs

Validation Rules:

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

72.7% accuracy

(1,792/2,465)

89.4%	eCommerce
67.5%	Navigation
74.8%	Advertising

Measurement & Analysis

Fraud Scale

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

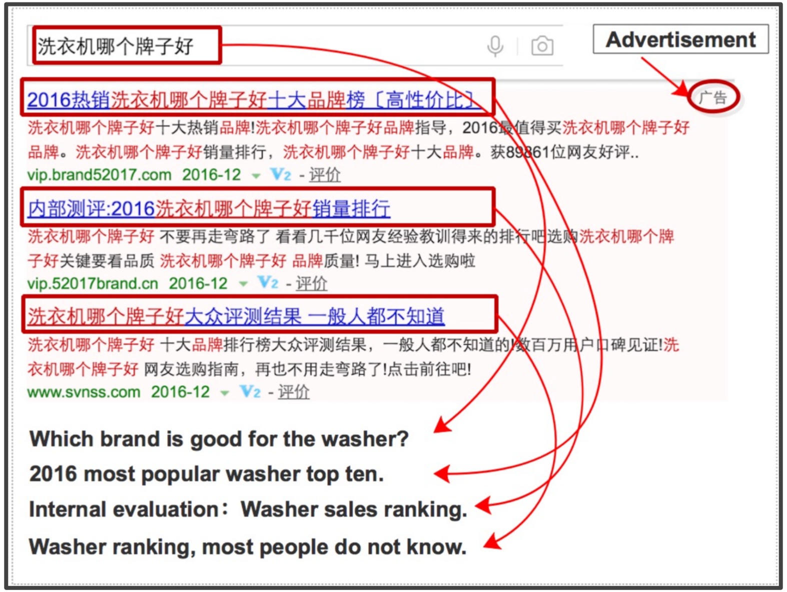
96%+ FSes receive each

300+ Days

85%+ FSes are active for

Search Ad Impersonation

Buying ads on search engines to attract visits.



洗衣机哪个牌子好

Advertisement

广告

2016热销洗衣机哪个牌子好十大品牌榜〔高性价比〕

洗衣机哪个牌子好十大热销品牌!洗衣机哪个牌子好品牌指导，2016最值得买洗衣机哪个牌子好品牌。洗衣机哪个牌子好销量排行，洗衣机哪个牌子好十大品牌。获89861位网友好评..

vip.brand52017.com 2016-12 V2 - 评价

内部测评:2016洗衣机哪个牌子好销量排行

洗衣机哪个牌子好 不要再走弯路了 看看几千位网友经验教训得来的排行吧选购洗衣机哪个牌子好关键要看品质 洗衣机哪个牌子好 品牌质量! 马上进入选购啦

vip.52017brand.cn 2016-12 V2 - 评价

洗衣机哪个牌子好大众评测结果 一般人都不知道

洗衣机哪个牌子好 十大品牌排行榜大众评测结果，一般人都不知道的!数百万用户口碑见证!洗衣机哪个牌子好 网友选购指南，再也不用走弯路了!点击前往吧!

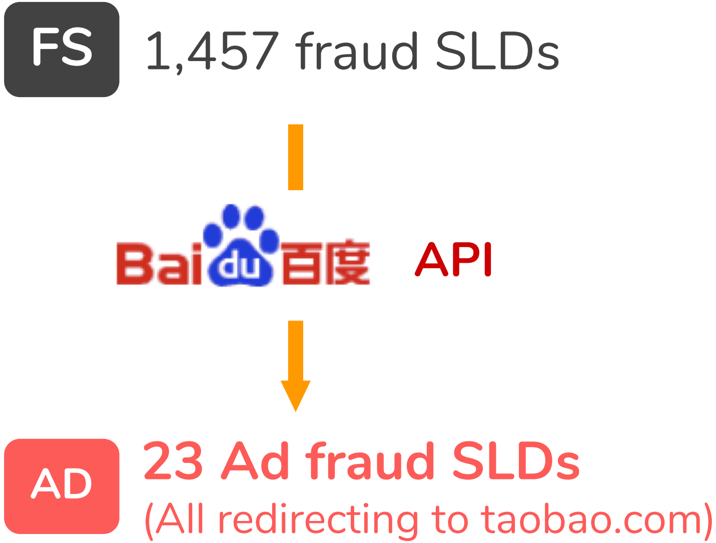
www.svnss.com 2016-12 V2 - 评价

Which brand is good for the washer?

2016 most popular washer top ten.

Internal evaluation: Washer sales ranking.

Washer ranking, most people do not know.



FS 1,457 fraud SLDs

Baidu 百度 API

AD **23 Ad fraud SLDs**

(All redirecting to taobao.com)

Search Ad Impersonation

23 Ad fraud SLDs redirecting to *taobao.com*.

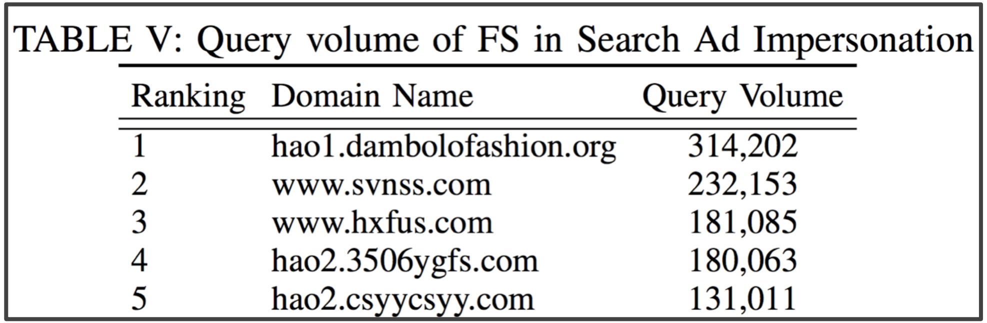


TABLE V: Query volume of FS in Search Ad Impersonation

Ranking	Domain Name	Query Volume
1	hao1.dambolofashion.org	314,202
2	www.svnss.com	232,153
3	www.hxfus.com	181,085
4	hao2.3506ygfs.com	180,063
5	hao2.csyycsyy.com	131,011

1M+

Total visits

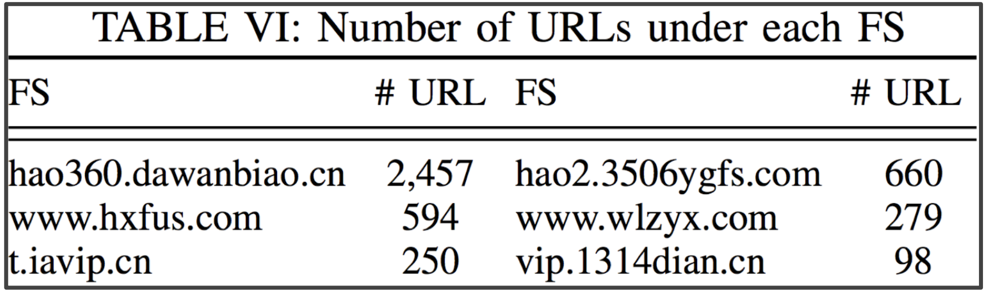


TABLE VI: Number of URLs under each FS

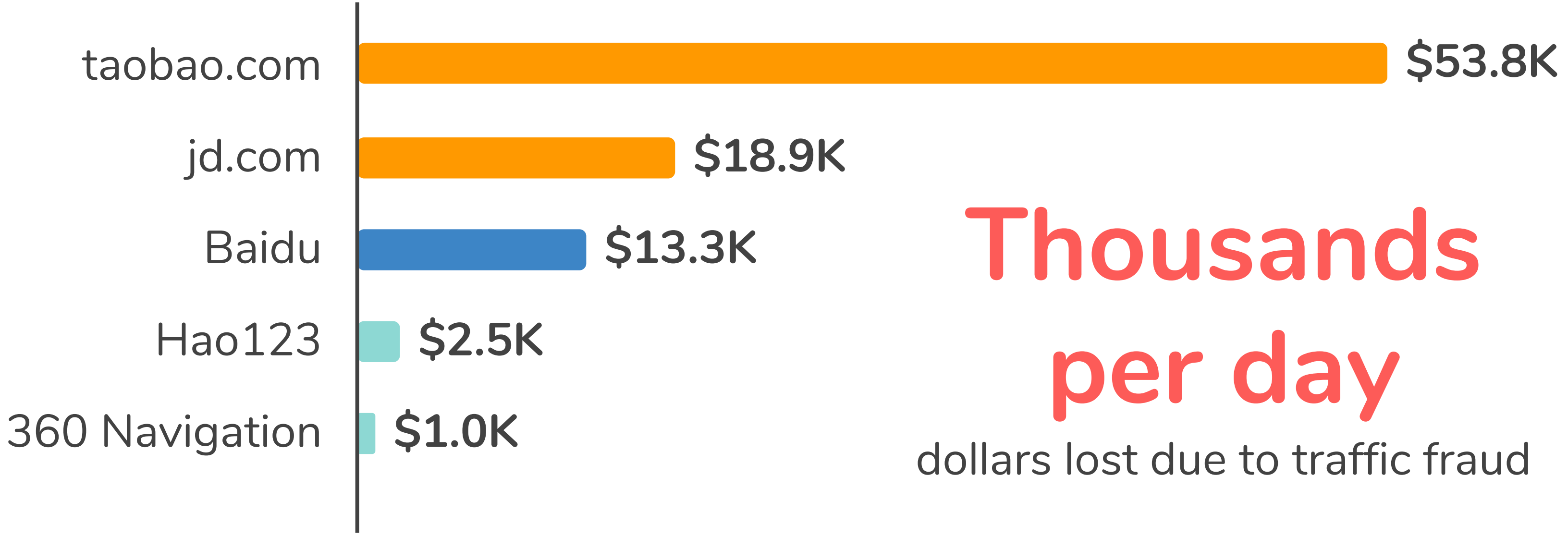
FS	# URL	FS	# URL
hao360.dawanbiao.cn	2,457	hao2.3506ygfs.com	660
www.hxfus.com	594	www.wlzyx.com	279
t.iavip.cn	250	vip.1314dian.cn	98

Hundreds of

keywords bought under each domain

Economic Loss

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

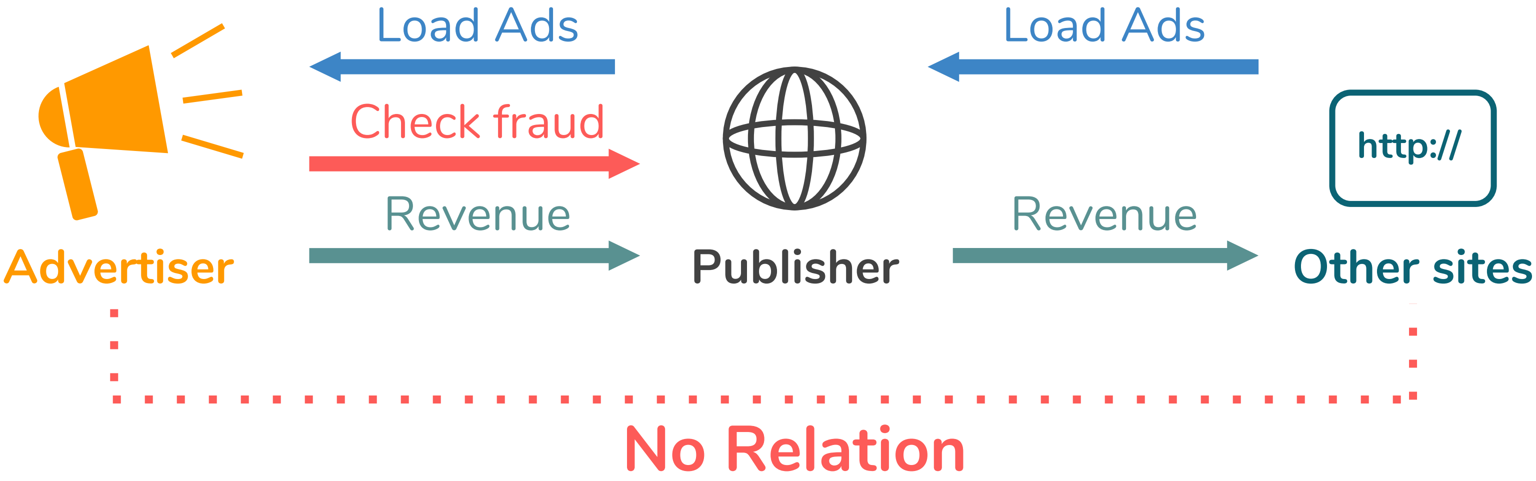


Thousands per day

dollars lost due to traffic fraud

New Strategy: Ad Reselling

Evading fraud detection of advertising platforms.



New Strategy: Ad Reselling

Evading fraud detection of advertising platforms.

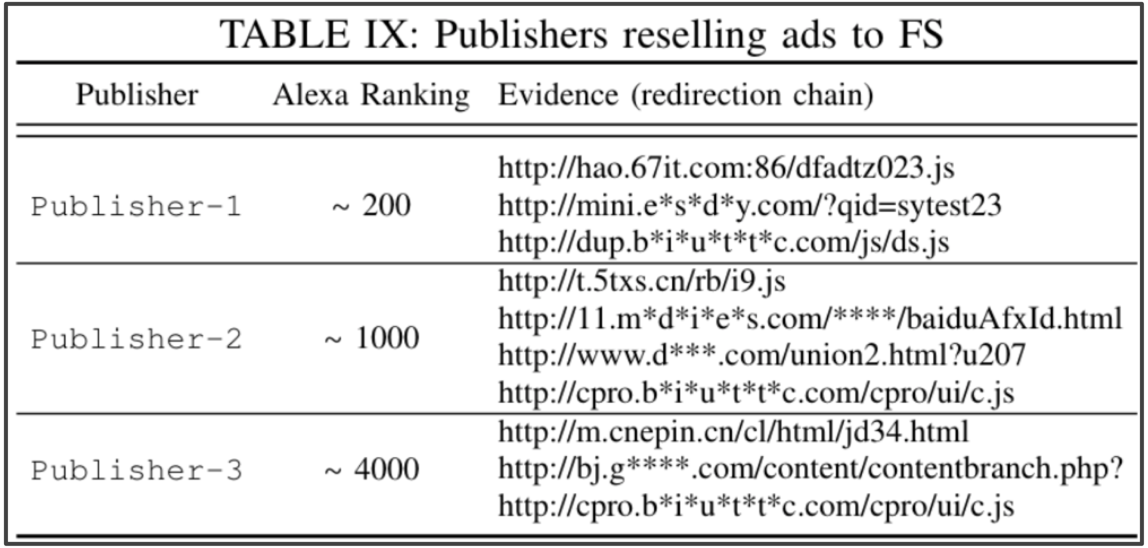
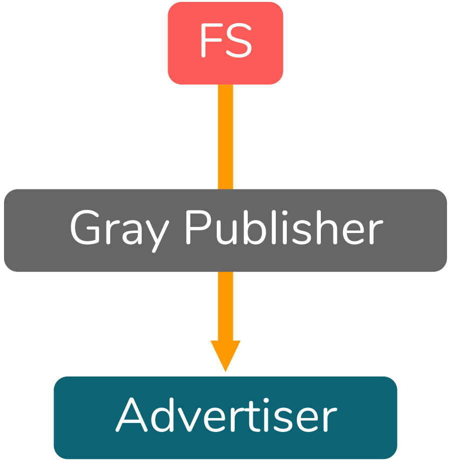


TABLE IX: Publishers reselling ads to FS

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

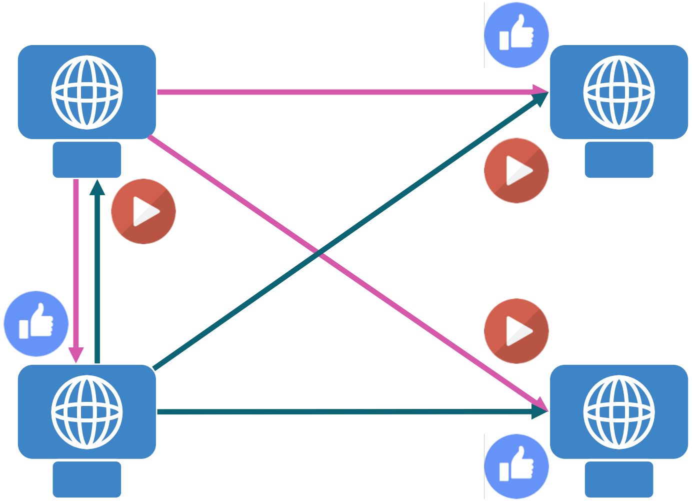


Case Study: P2P Traffic Pal

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