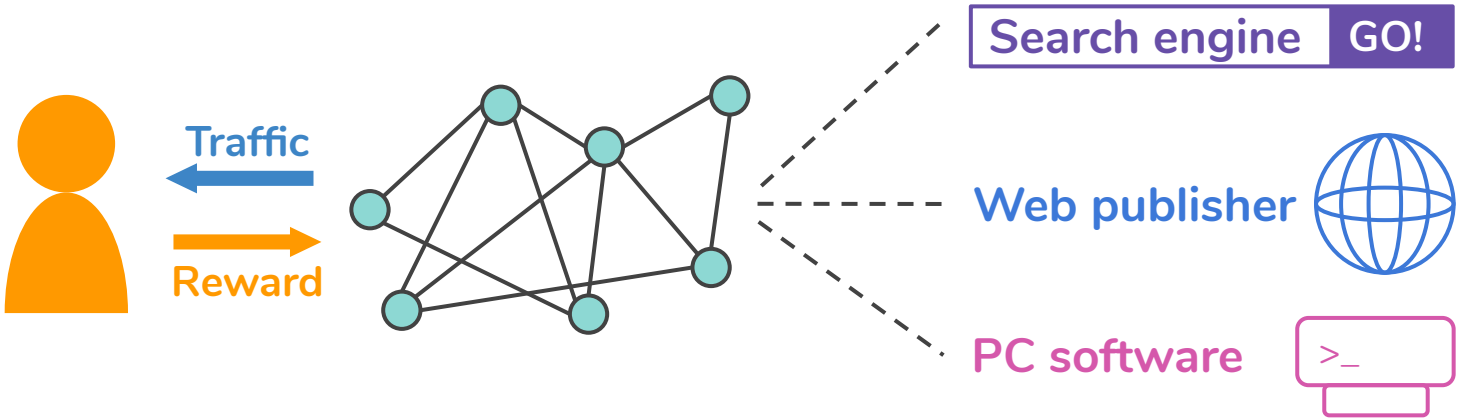
TraffickStop: Detecting and Measuring Illicit Traffic Monetization Through Large-scale DNS Analysis

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

Traffic Network

Connects site owners and affiliates.



Site owner
(Needs traffic)

Traffic Network
(Finds affiliates)

Affiliate
(Refers traffic)

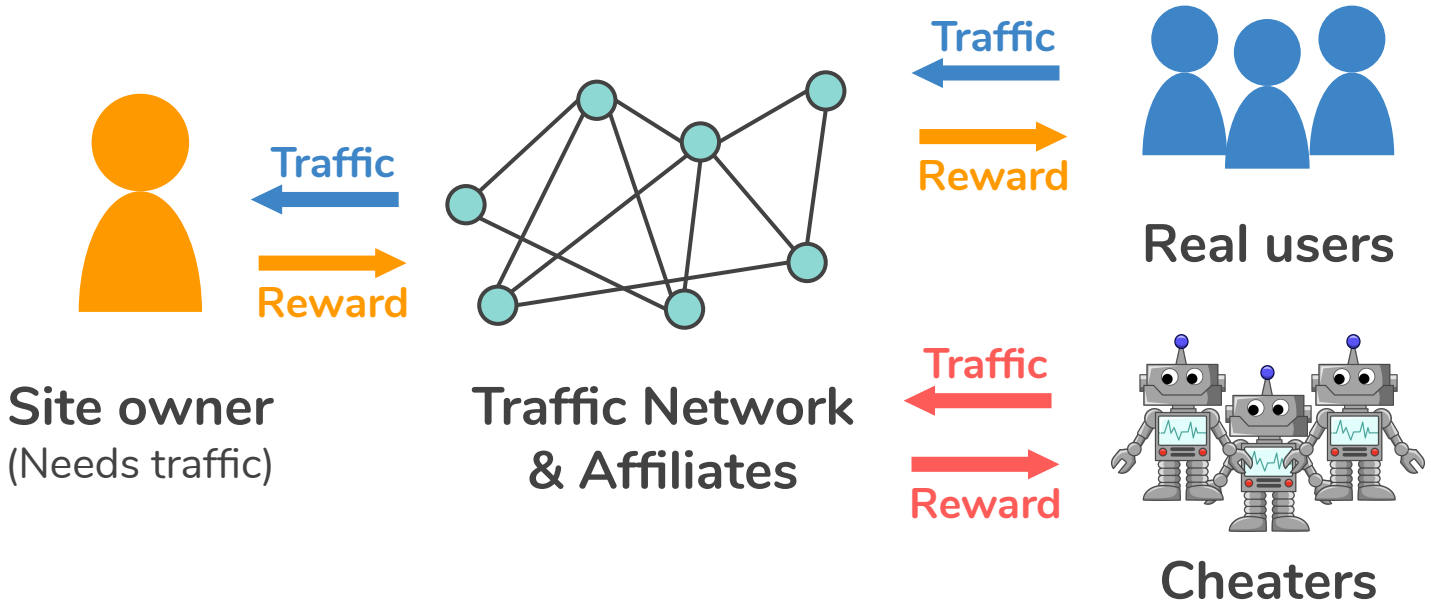
Traffic Network

Connects site owners and affiliates.

eCommerce Network	Advertising Network	Navigation Network
amazon associates	Google Ads	hǎo123
ebay partner network	Microsoft | Advertising	360导航
Rakuten Marketing	media.net	

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.

Cheating happens EVERYWHERE!

Client-side:
Browser Hijacking

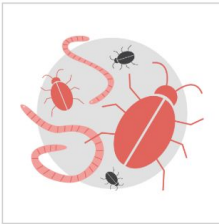
Install PUP / Malware on client machines

Reroute user traffic to targeted sites

Adware.Yontoo

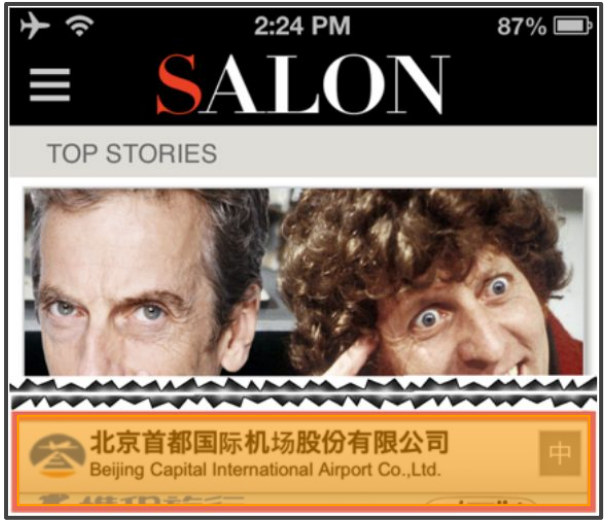
Short bio **Caused $8M loss in 2013**

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



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

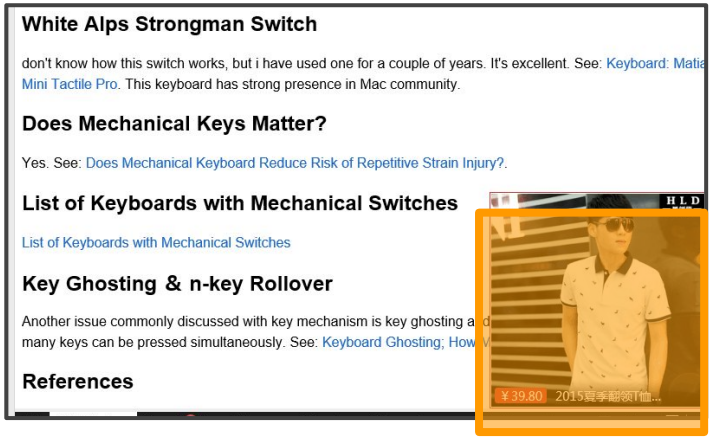
Cheating happens EVERYWHERE!



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

Inject extra ads into web responses

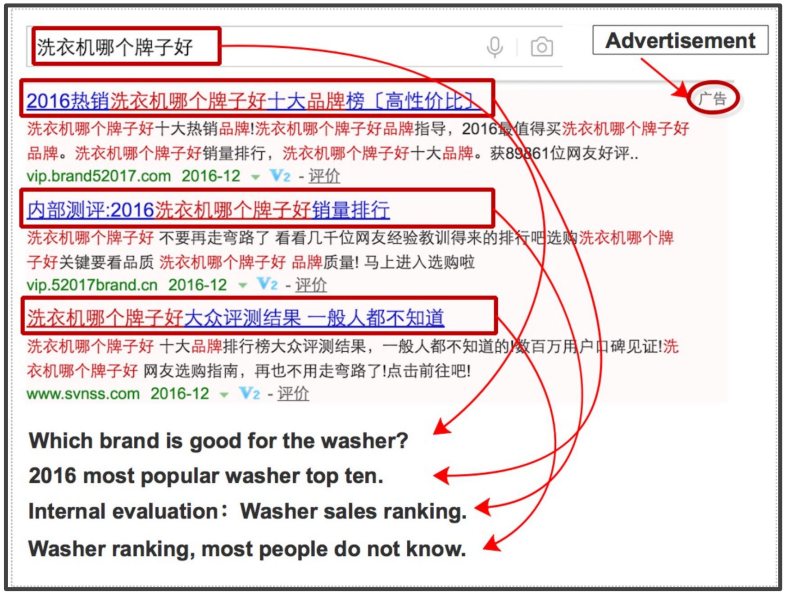
Mitigation: HTTPS Relies on adoption rate



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

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

Transport-layer:
ISP Injection

Server-side:
Search Ad Impersonation

Install PUP / Malware on client machines

Reroute user traffic to targeted sites

Inject extra ads into web responses

Mitigation: HTTPS Relies on adoption rate

Publish fake ads in search engines

Impersonate popular brands to trap more users

Cheating happens EVERYWHERE!

Client-side:
Browser Hijacking

Transport-layer:
ISP Injection

Server-side:
Search Ad Impersonation

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.

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>
```

Traffic Network Affiliate Code

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_ic
<script language="JavaScript">
if(location.hostname=="bd.114la6.com")
          location="https://www.baidu.com/?tn=90578459_hao_pg";
</script>
                                  Traffic Network    Affiliate Code

</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

http://

URL

WHOIS

Association Finder

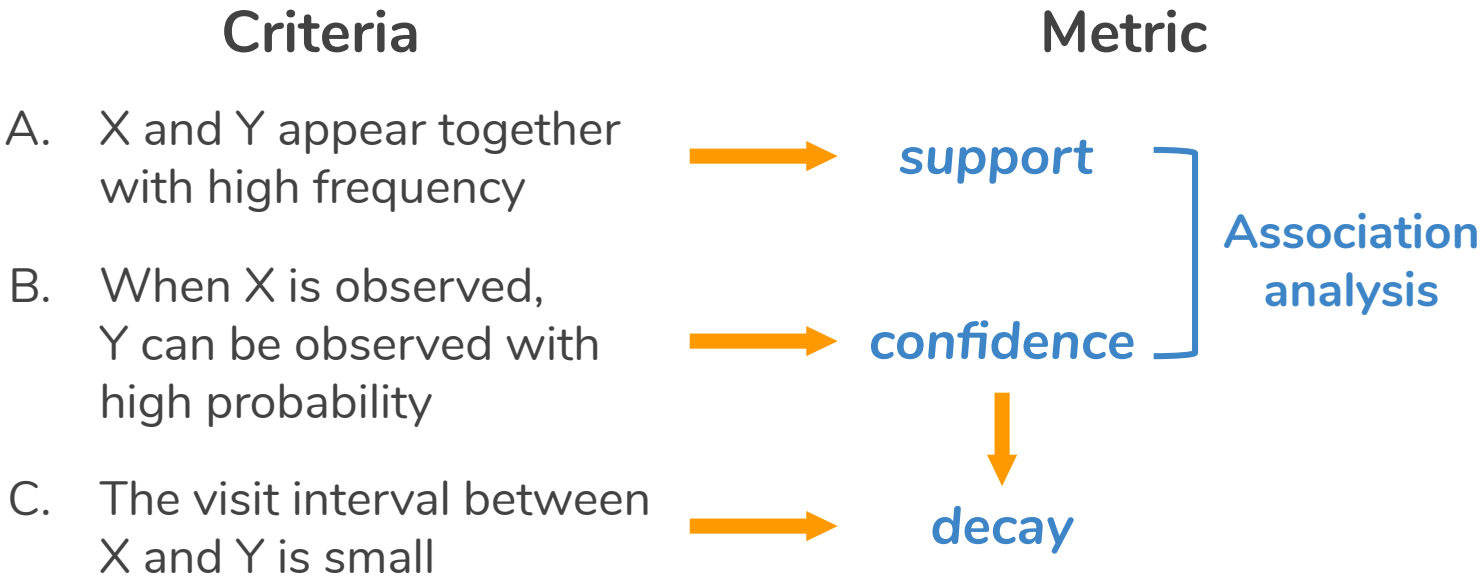
Finds domains with strong correlation

Content Analyzer

Examines suspicious behaviors between domains

Association Finder

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



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* ∈ *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* ∈ *session* **do**
      value ← DECAY(source.location, destination.location)
      Out: *source*, *destination*, *value*
    $index++$

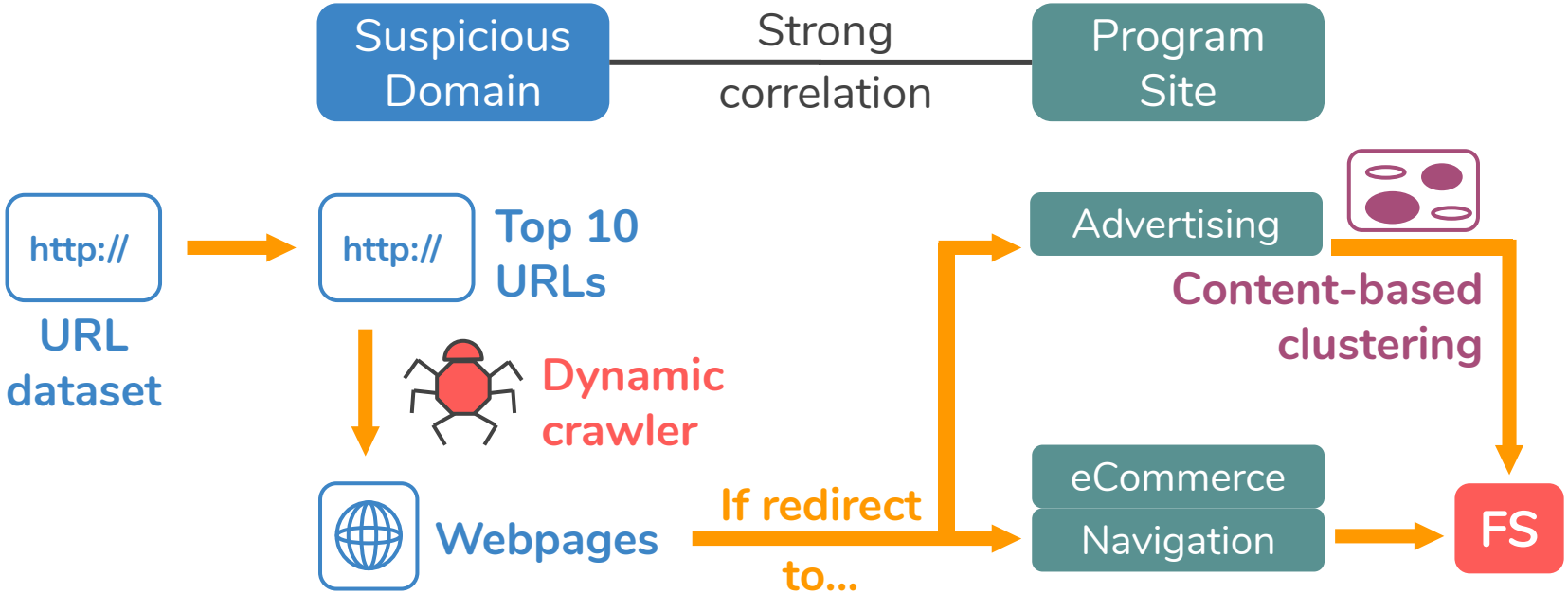
Procedure: ***Reduce***
$Group\_source \leftarrow$ GROUPBY(*source*)
$Rule \leftarrow$ MERGE($Group\_source$)
$Rule\_group \leftarrow$ FILTER_RULE($Rule, minsup, minconf$)
for *rule* ∈ *Rule_group* **do**
  Out: *source_domain*, *destination_domain*, *confidence*, *support*

Map procedure:
Calculate the interval between two domain visits

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

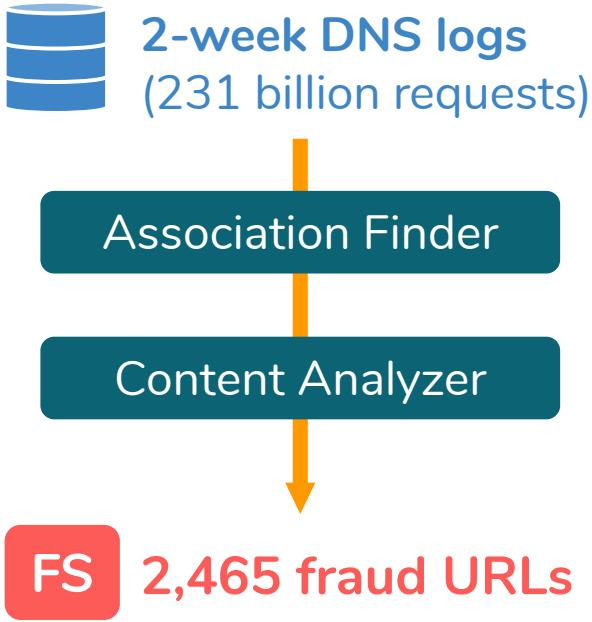
Content Analyzer

Examine **Redirection + Meaningless content**.



System Evaluation

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



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)

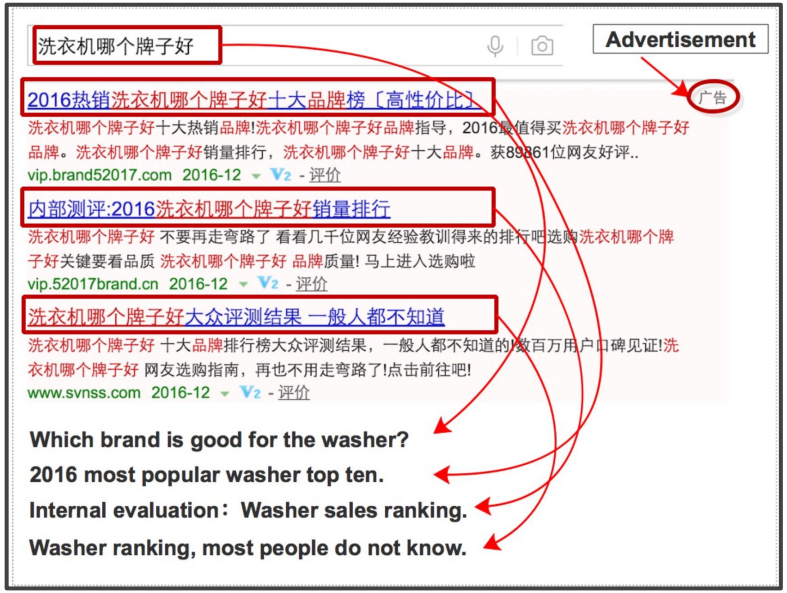
53 Billion
Total DNS queries to these FSes

100K+ Queries
96%+ FSes receive each

300+ Days
85%+ FSes are active for

Search Ad Impersonation

Buying ads on search engines to attract visits.



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

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

1M+

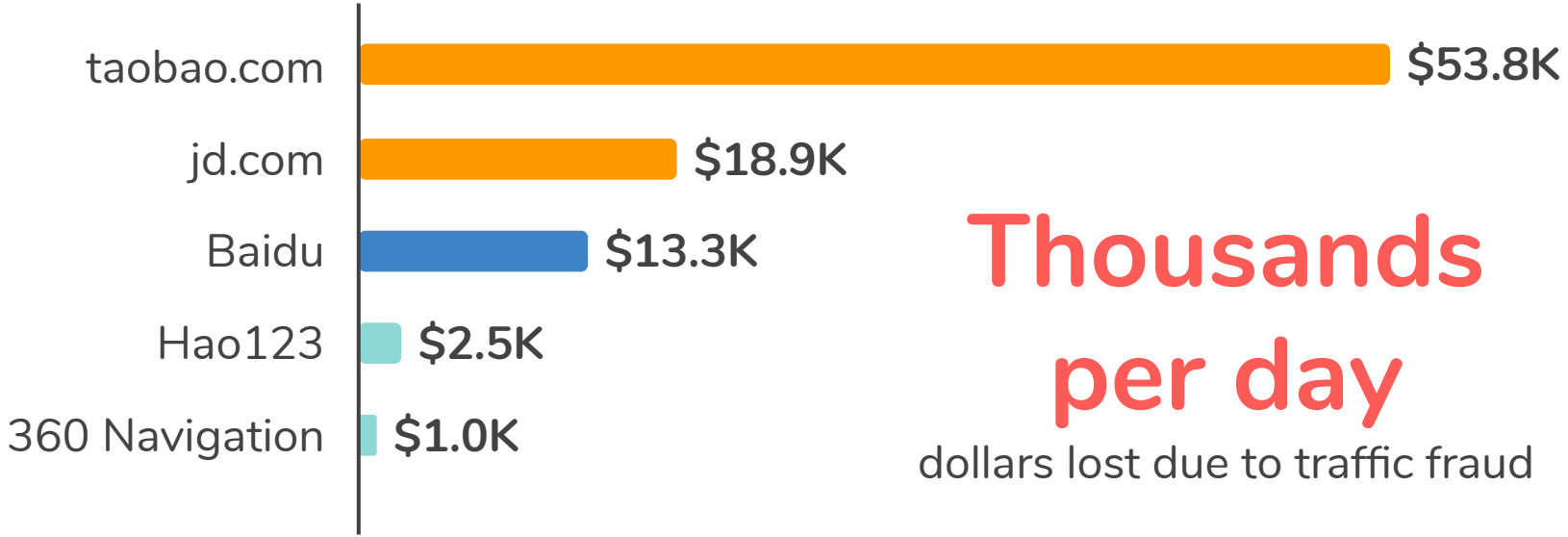
Total visits

Hundreds of

keywords bought under each domain

Economic Loss

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

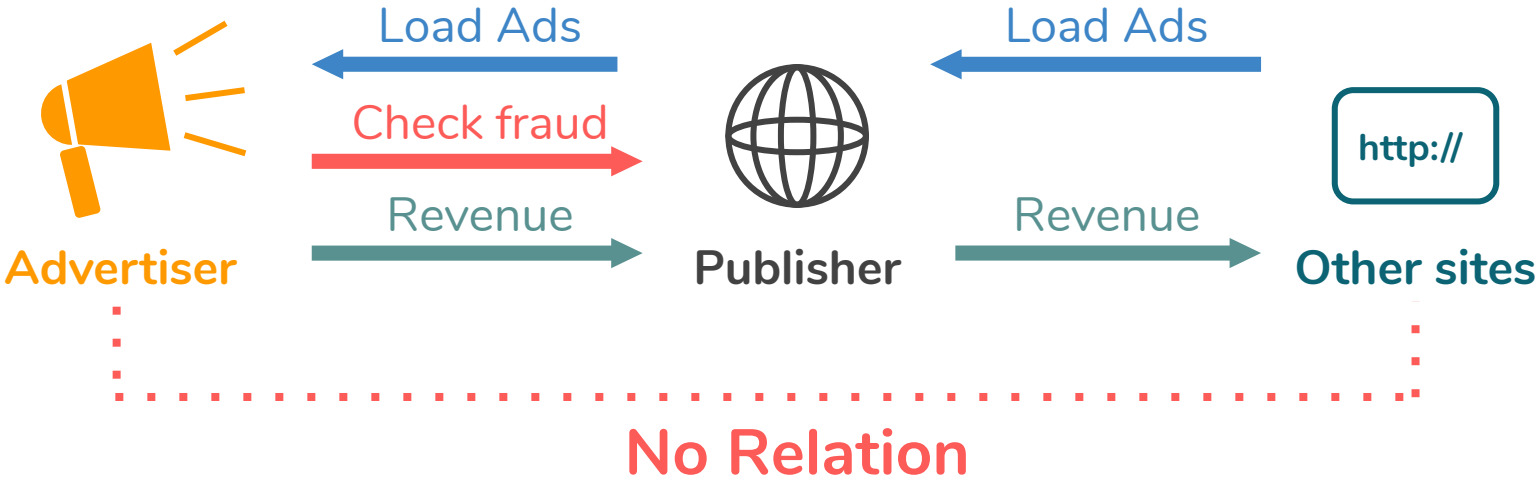


Thousands per day

dollars lost due to traffic fraud

New Strategy: Ad Reselling

Evading fraud detection of advertising platforms.

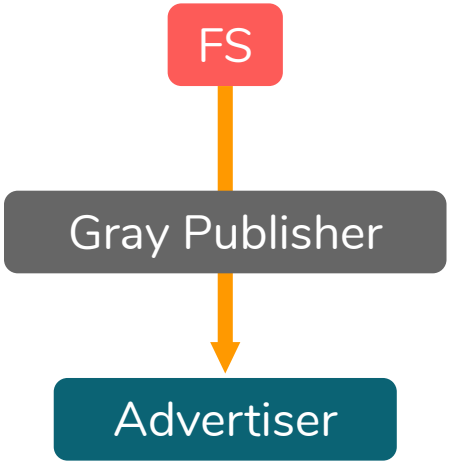


New Strategy: Ad Reselling

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

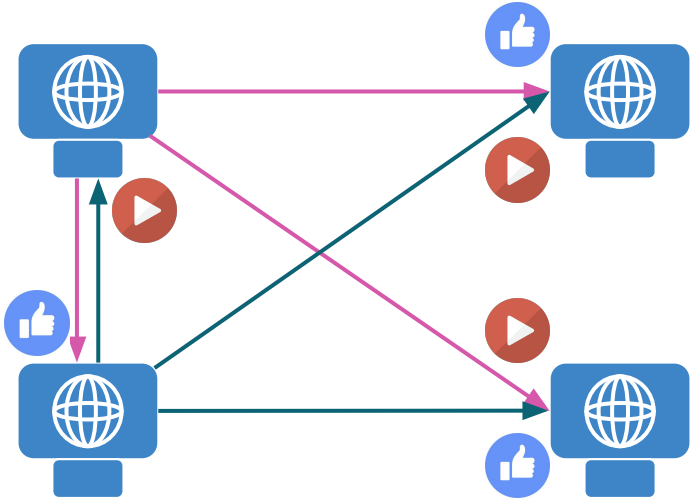


Case Study: P2P Traffic Pal

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