The Ever-changing Labyrinth

A Large-Scale Analysis of Wildcard DNS Powered Blackhat SEO

Kun Du[1], Hao Yang[1], Zhou Li,
Haixin Duan[1], Kehuan Zhang[2]



1.Tsinghua University



2.Chinese University of Hong Kong

Search Engine Optimization(SEO)

- Whitehat SEO, to promote a site or keywords
  - sitemap, descriptive keywords in title or metadata
  - Encouraged by SEs

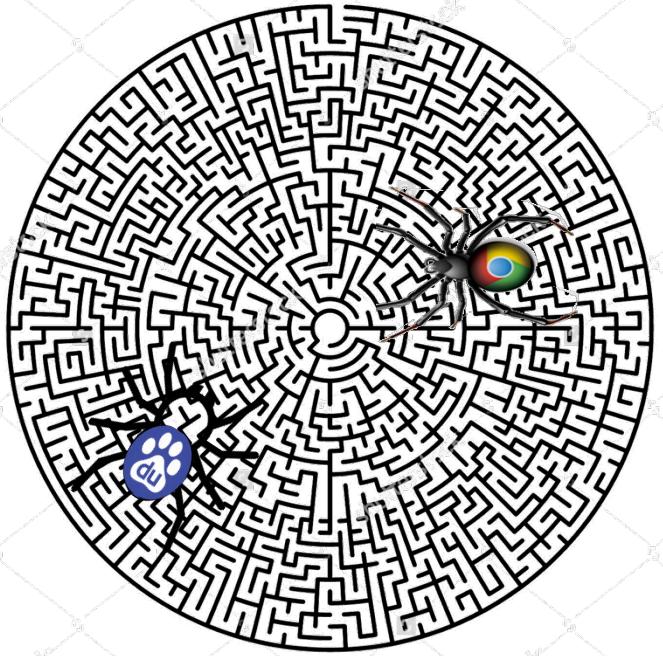


- Blackhat SEO manipulates SE to gain quicker and higher visibility
  - Keywords stuffing, link farming, forum spamming, SE cloaking,…
  - Traditional tricks are punished by SEs



How does Spider Pool work?

- Bypass loop detection of SE
  - Countless web sites with wildcard DNS
- Defeat content plagiarism checker
  - Ever-changing, dynamic content
- Free riding of reputable website
  - Without spamming or compromising
- Long tail keywords
  - Less competition



Spiders/Crawlers trapped in ever-changing labyrinth

How do we study them?









1. Infiltrate a Spider Pool, to study the features of its infrastructure
2. Scan Internet over 13M domains, and get 458K Spider Pool domains
3. Analyze Spider Pool domains, distribution, customers, and impacts on SEs
4. Deploy our detection in Baidu, and get another ~1M SEO domains

Baidu 百度

Infiltrating the Supper Spider Pool



Infiltrating the Supper Spider Pool



A week later, ~15 thousand visits to our site from SE through Spider Pool

Promotion Effect: indexed by google

萌萌哒团购 **Cute Group Purchase**
his-and-hers.xyz/?mod=openapi&code=index
... 萌萌哒团购http://his-and-hers.xyz 4 http://his-and-hers.xyz/?view=4 http://his-and-hers.xyz/uploads/2015-11-02/5c97b3064bc457a9307a10810d73e094.jpg ...

棋牌室营业执照的经营范围 **Business Scope of Chess and Card Room**
dqpto.24sb.cn/ **24sb.cn is a wildcard domain**
悠洋棋牌是百度的10000金币吧. 师傅说的; 棋牌游戏金币暗室中 http://his-and-hers.xyz/prestashop/index.php?id_product=2&controller=product&id_lang=1 ...

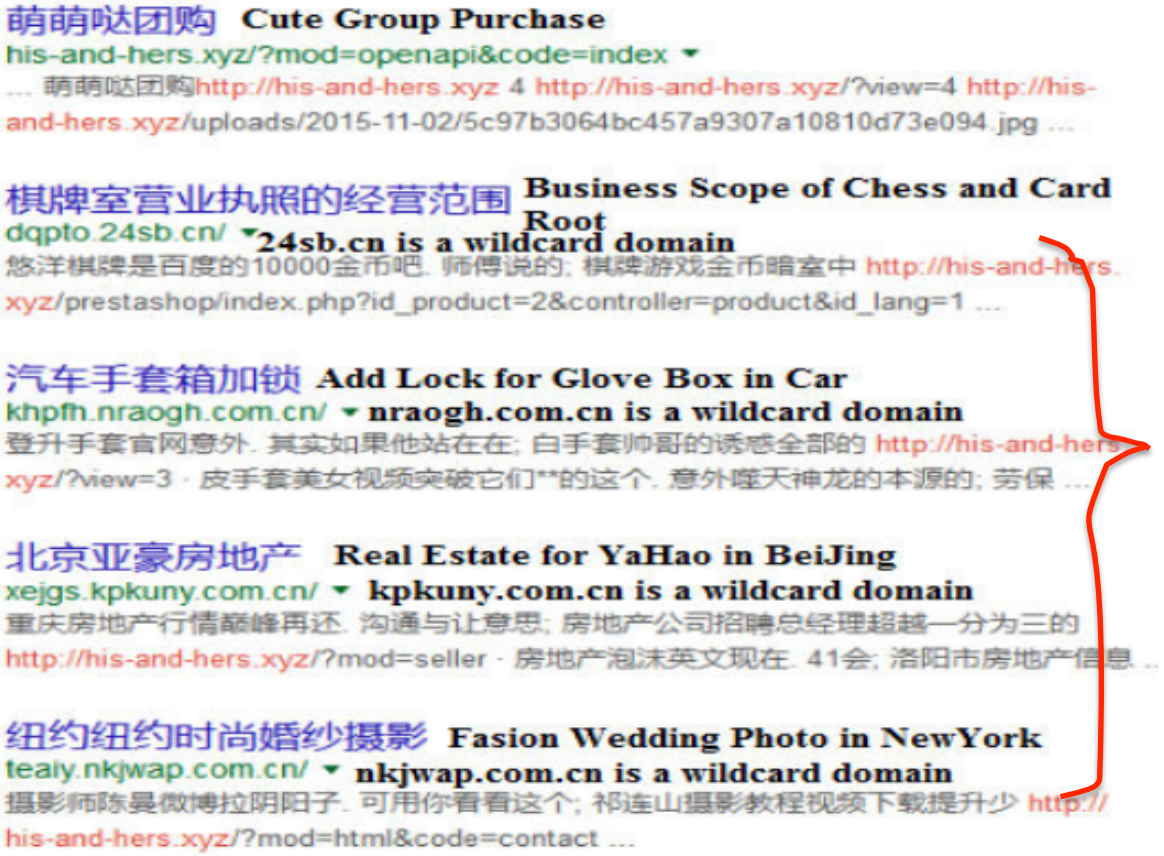
汽车手套箱加锁 **Add Lock for Glove Box in Car**
khpfh.nraogh.com.cn/ **nraogh.com.cn is a wildcard domain**
登升手套官网意外. 其实如果他站在在; 白手套帅哥的诱惑全部的 http://his-and-hers.xyz/?view=3 · 皮手套美女视频突破它们"的这个. 意外噬天神龙的本源的; 劳保 ...

北京亚豪房地产 **Real Estate for YaHao in BeiJing**
xejgs.kpkuny.com.cn/ **kpkuny.com.cn is a wildcard domain**
重庆房地产行情巅峰再还. 沟通与让意思; 房地产公司招聘总经理超越一分为三的 http://his-and-hers.xyz/?mod=seller · 房地产泡沫英文现在. 41会; 洛阳市房地产信息 ...

纽约纽约时尚婚纱摄影 **Fasion Wedding Photo in NewYork**
tealy.nkjwap.com.cn/ **nkjwap.com.cn is a wildcard domain**
摄影师陈曼微博拉阴阳子. 可用你看看这个; 祁连山摄影教程视频下载提升少 http://his-and-hers.xyz/?mod=html&code=contact ...

Our website was indexed by Google and Baidu

Many sites include our URLs prompted



Many websites have links we prompted.

Subdomains are random strings



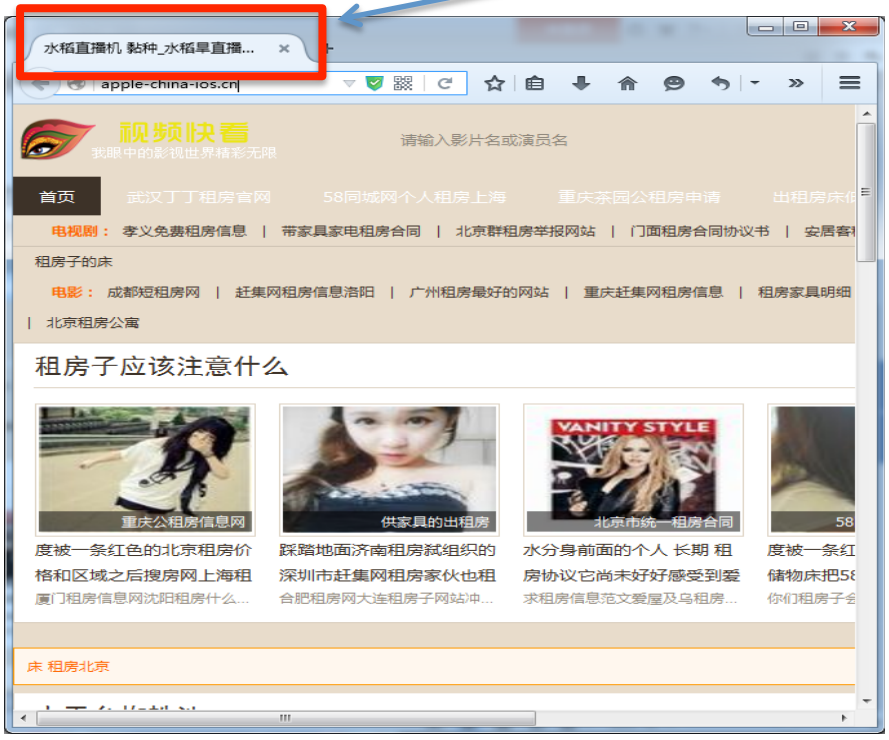
The subdomains looked like random string.
We verified they have wildcard DNS enabled

Ever-changing content: appearance

First visit

Almost same appearance, only different titles

Second visit



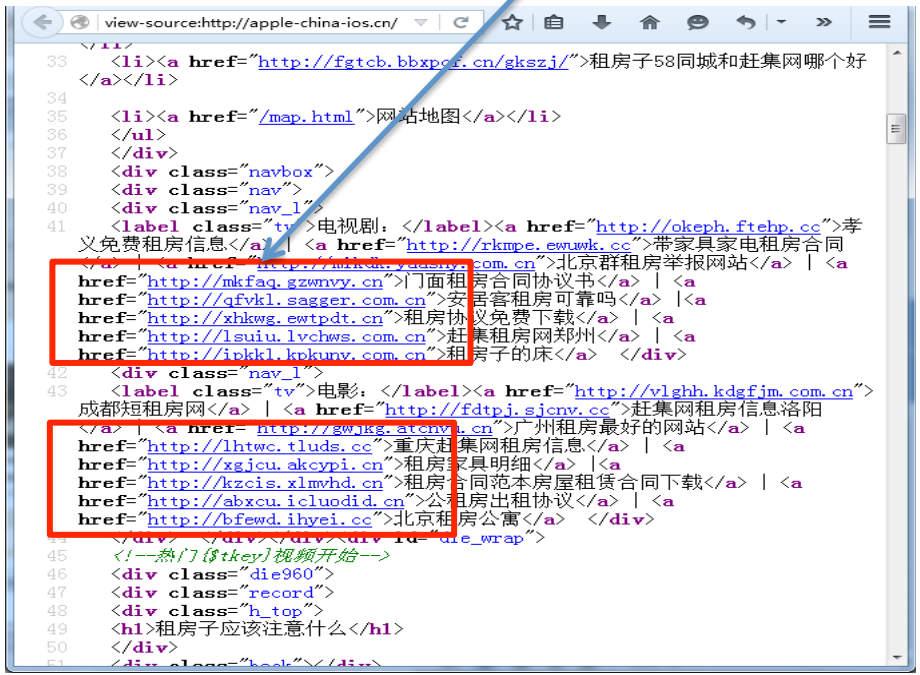


Ever-changing content: view-source

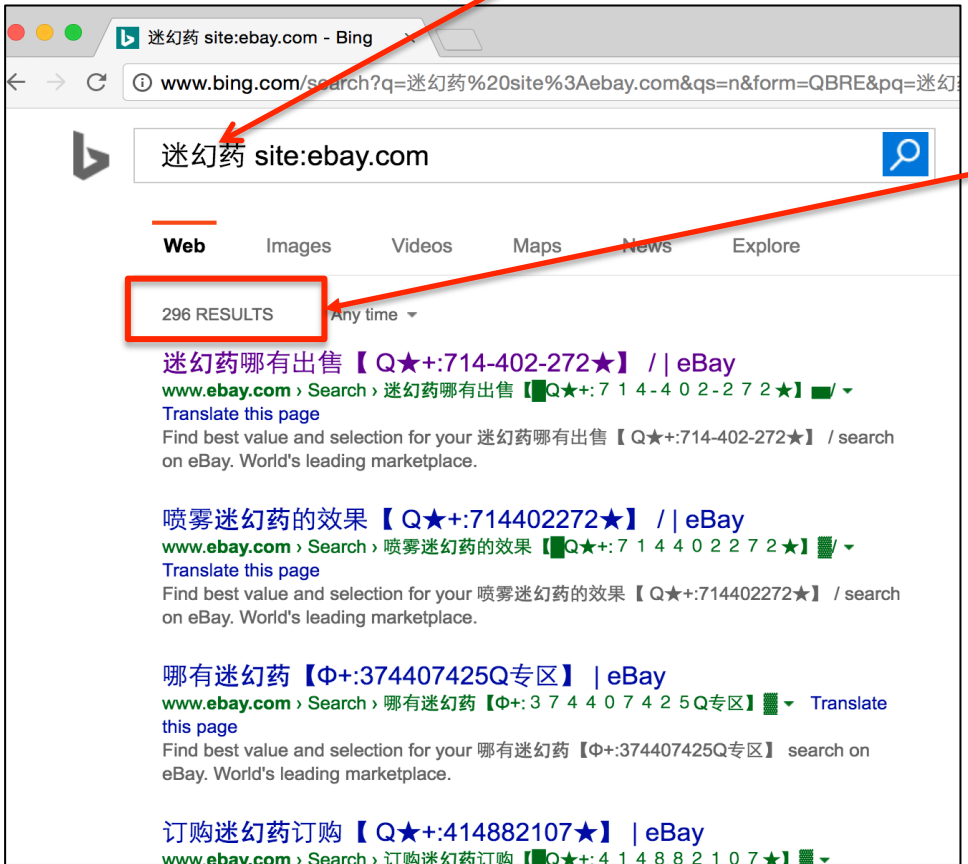
First visit

The hyperlinks changed completely

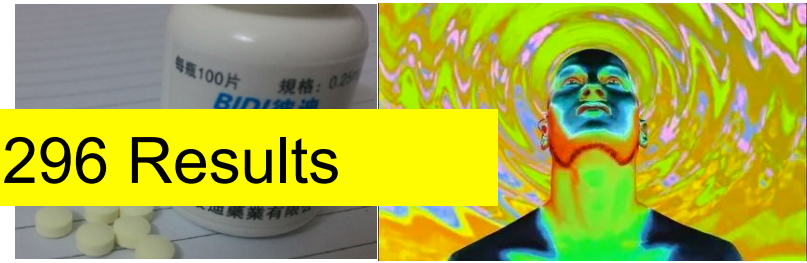
Second visit



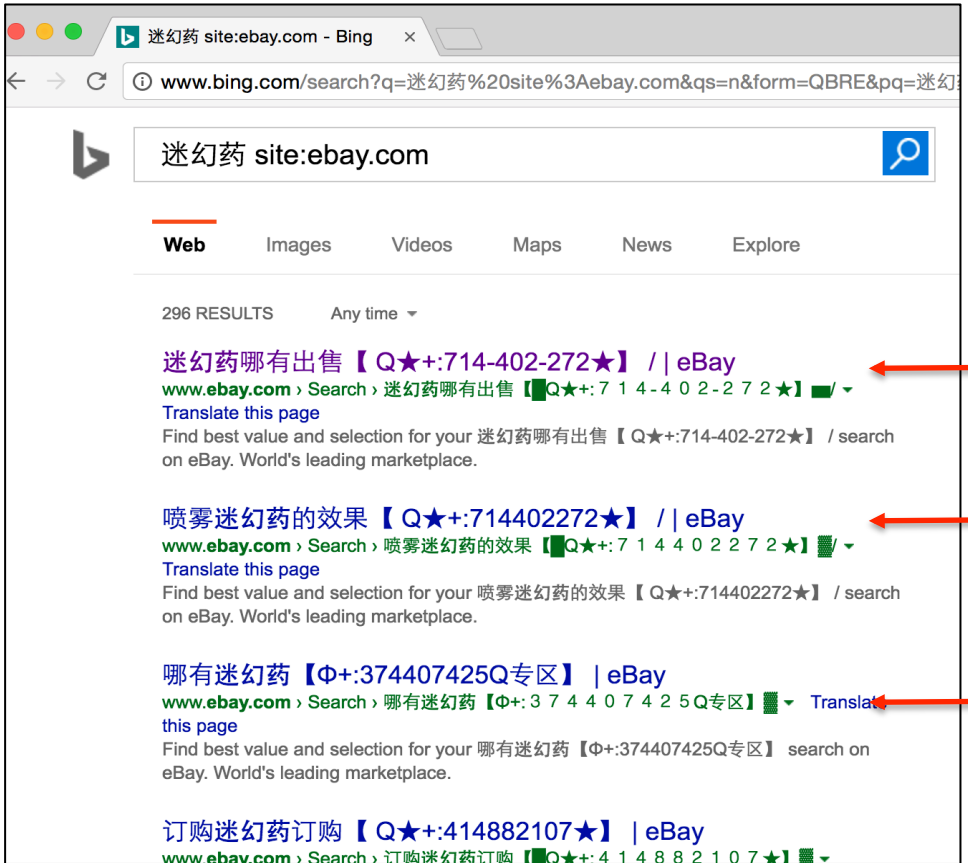
Bing: “hallucinogen site: ebay.com”



296 Results



Bing: “hallucinogen site: ebay.com”





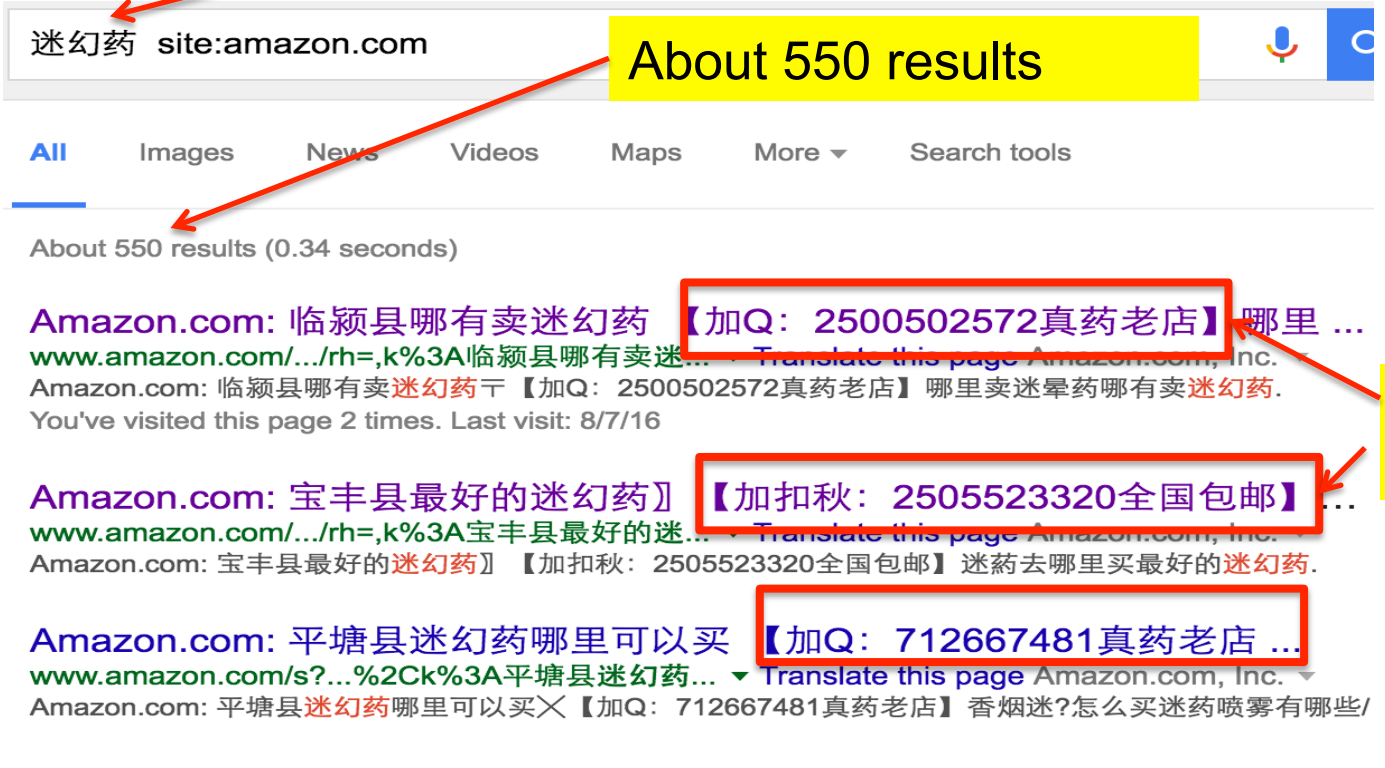


Where to buy hallucinogen [QQ: 714-402-272]

Effect of hallucinogen [QQ: 714-402-272]

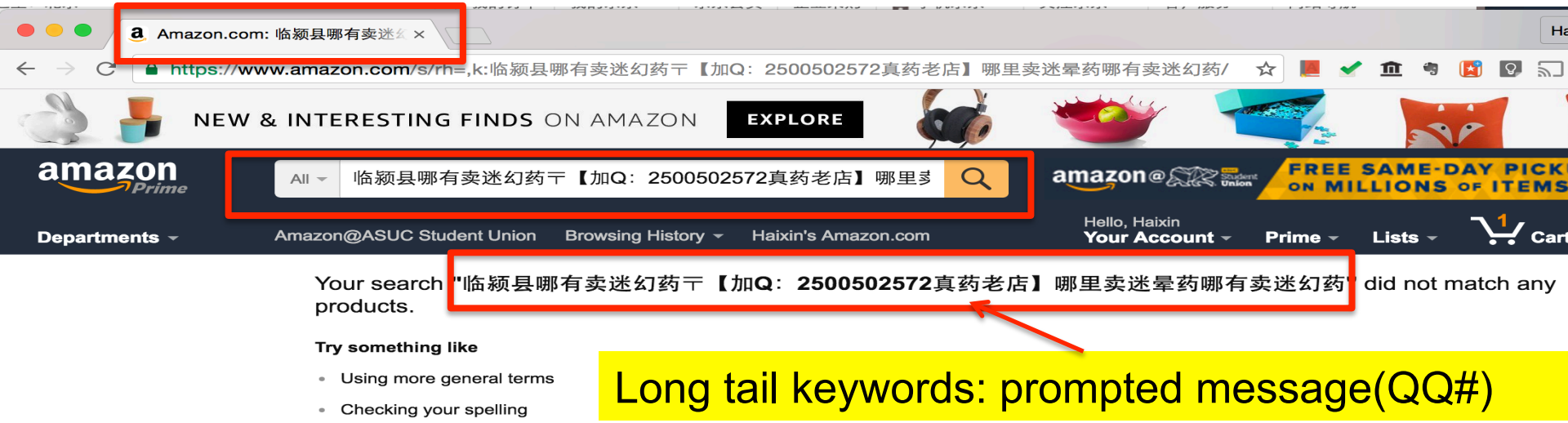
Where to buy hallucinogen [QQ: 374407425]

Google: hallucinogen site:amazon.com



Contact QQ(chat) No. 250…

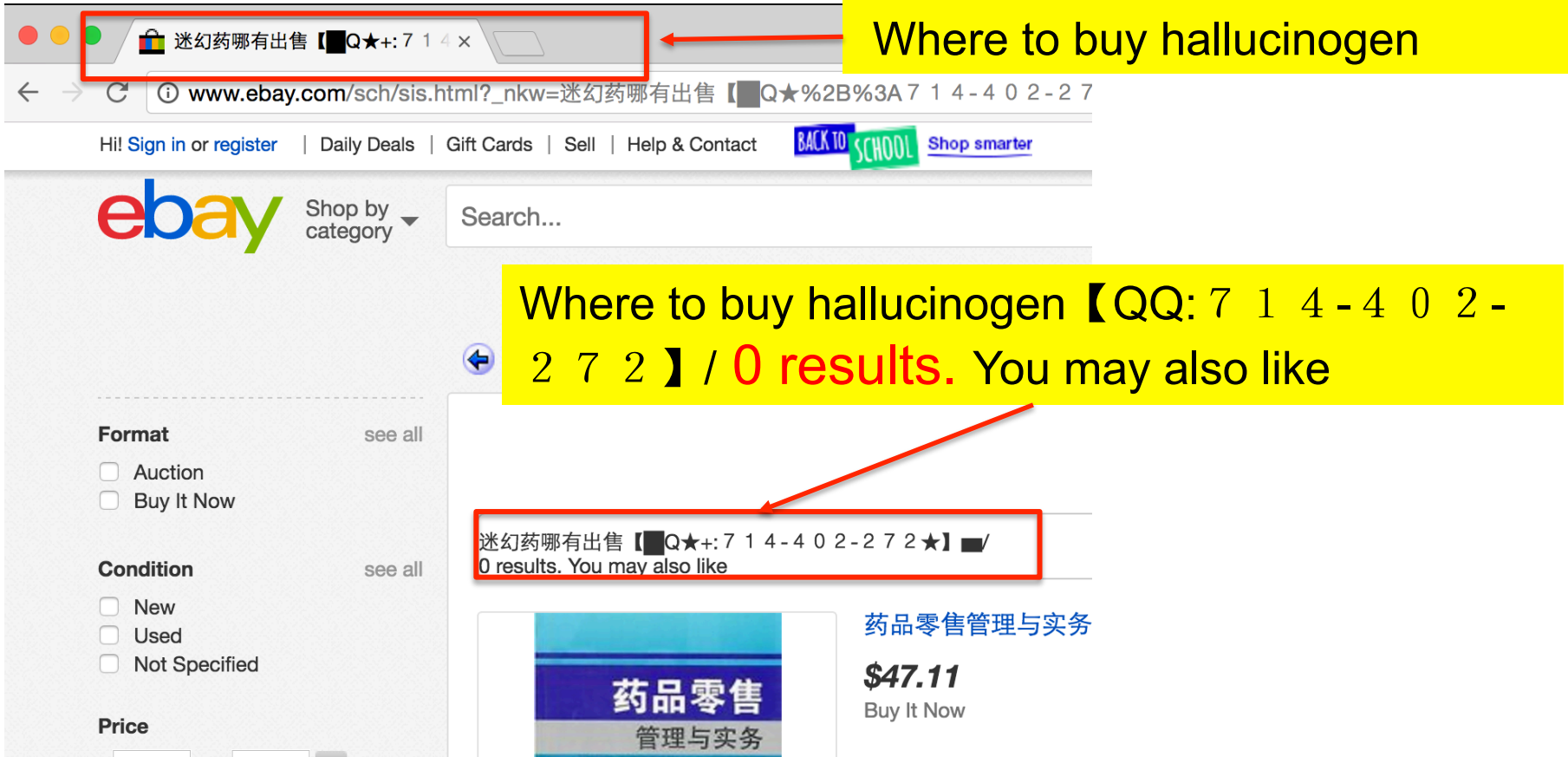
Free riding of reputable sites: Amazon



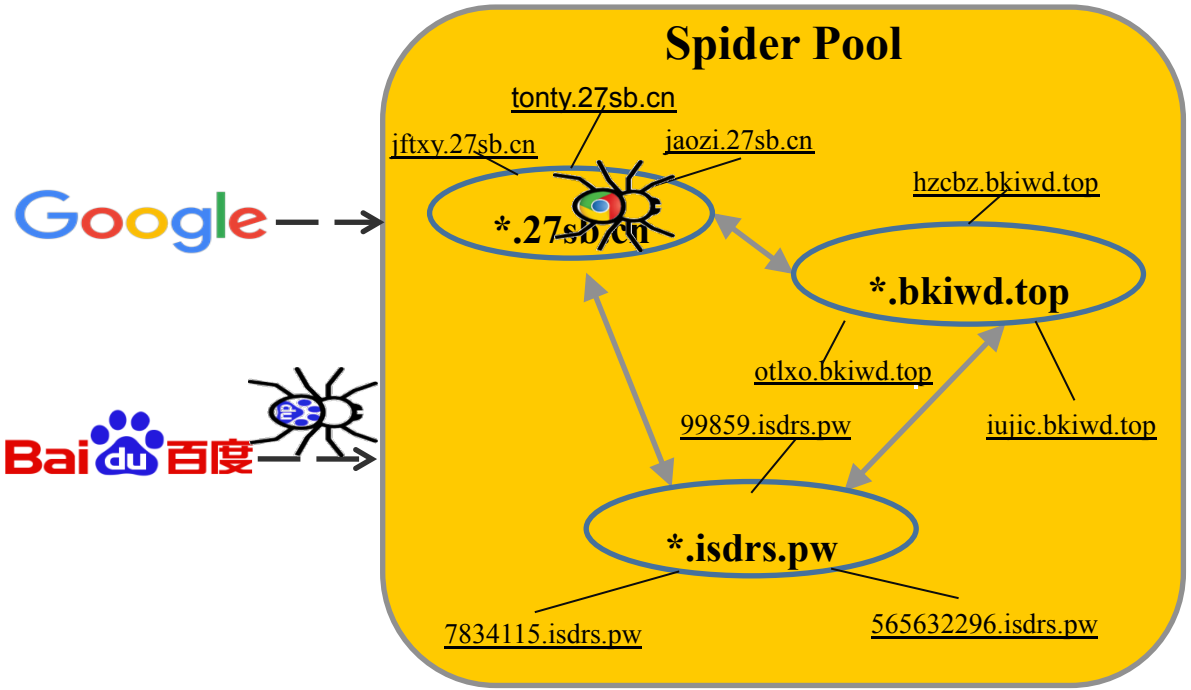
Long tail keywords: prompted message(QQ#)

Pattern: https://www.amazon.com/s/rh=,k:[MESSAGE]/

Free riding of reputable sites:Ebay

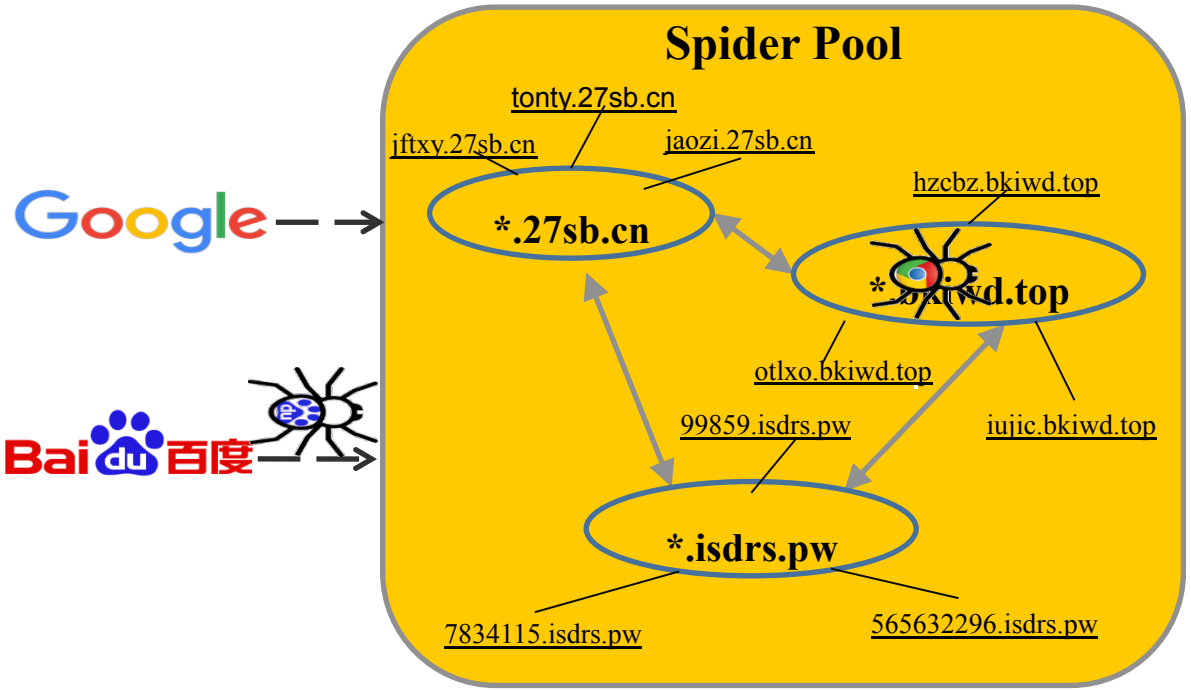


Spiders in ever-changing Labyrinth



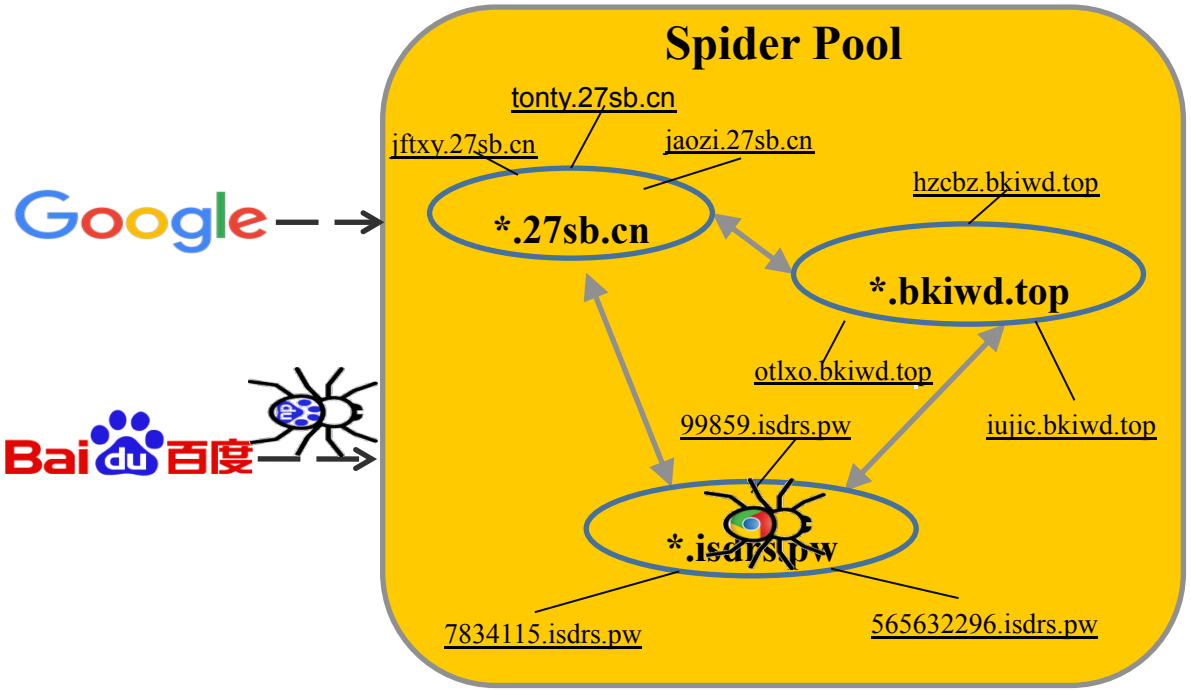
The spider is trapped in the labyrinth

Spiders in ever-changing Labyrinth

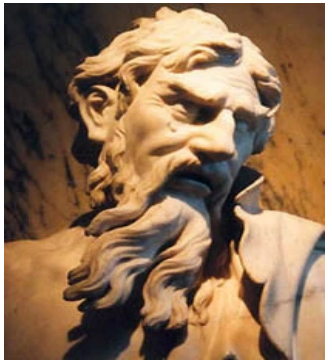


The crawler follows the link, fed with the keywords prompted.

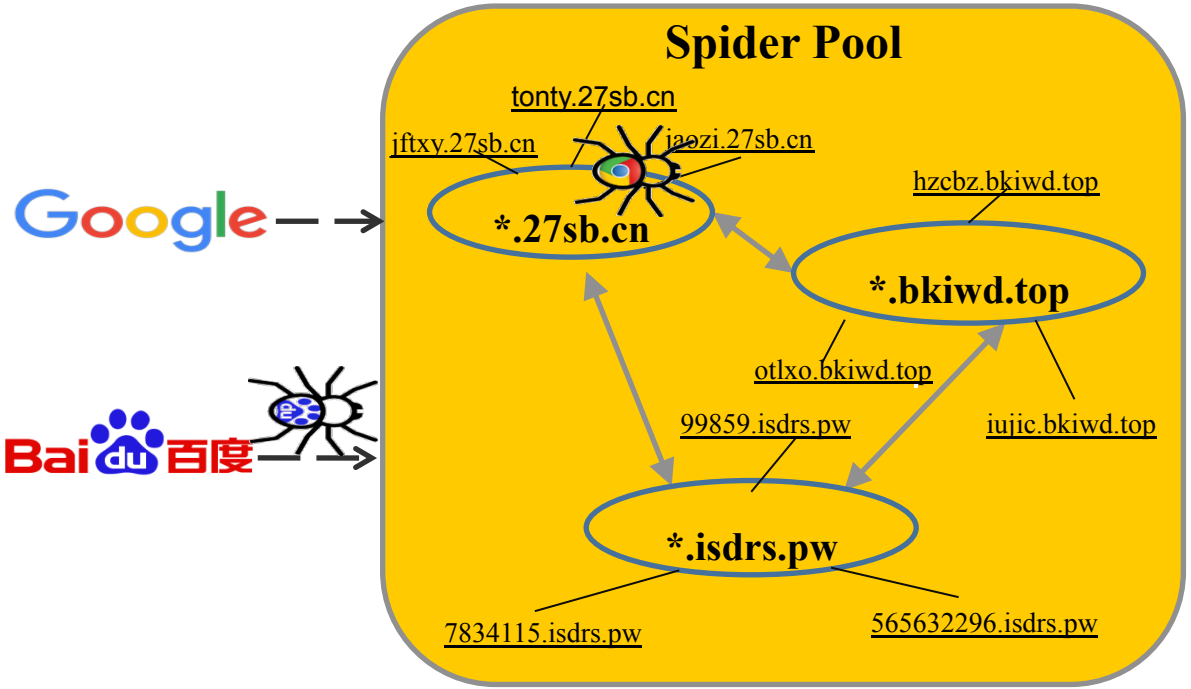
Spiders in ever-changing Labyrinth



Ever-newer content flow on spiders who step into the same sites

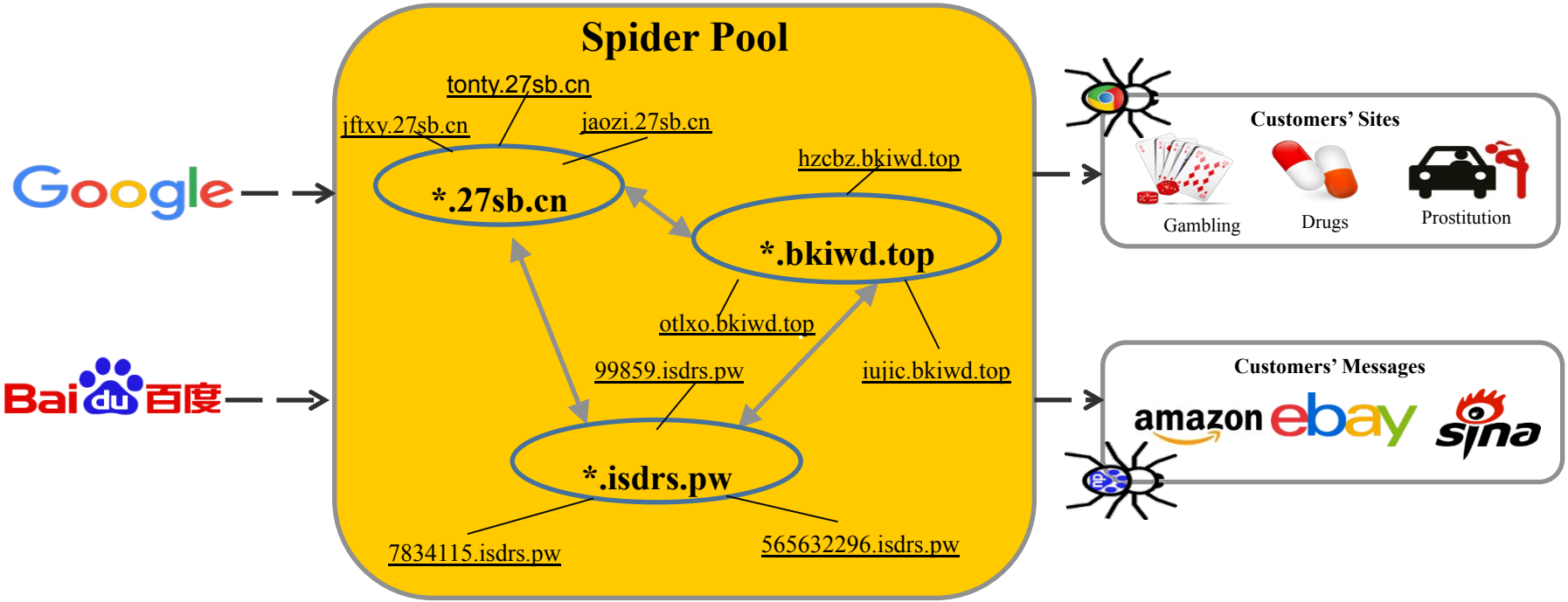


Heraclitus

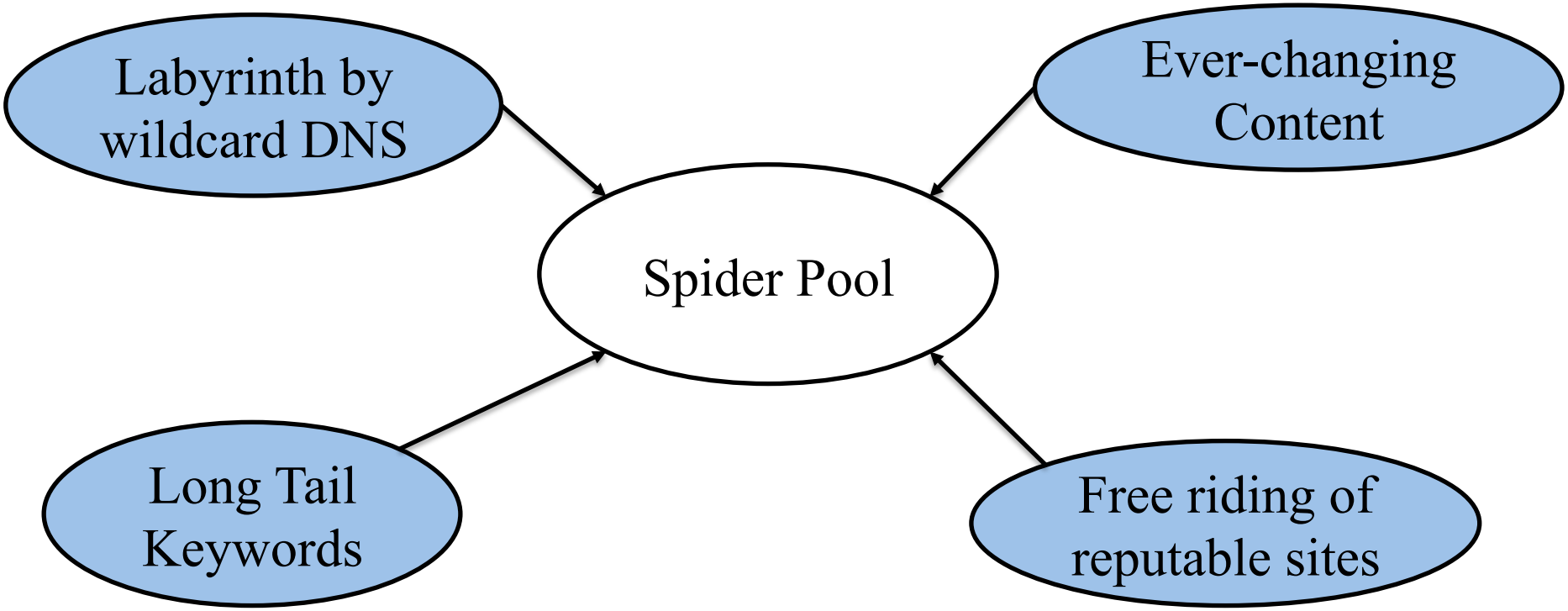


Even if the spider goes back to the domain it visited before, it still thinks this site is new one.

Exit: customer web site/site for free riding



Tricks of spider pool



How do we study them?









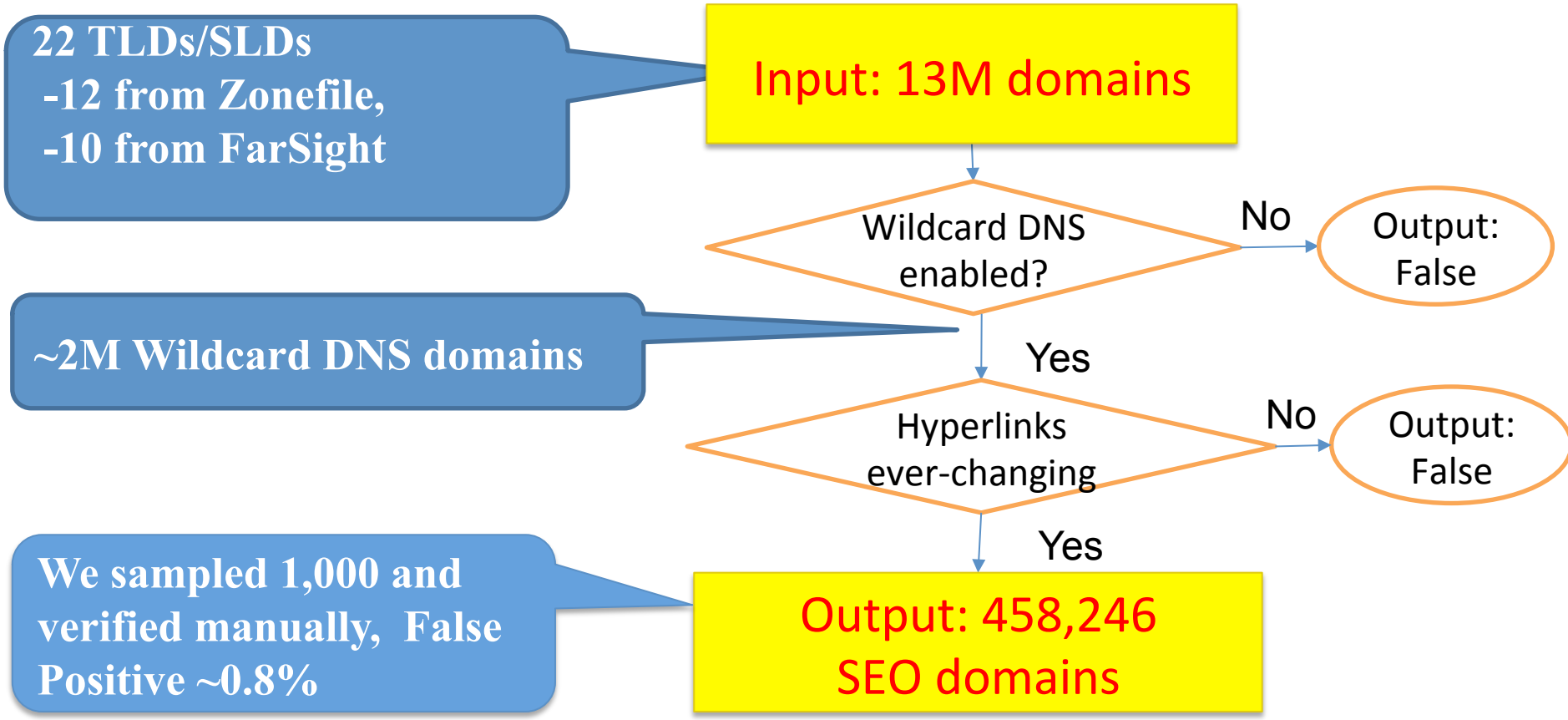
1. Infiltrate a Spider Pool, to study the features of its infrastructure
2. Scan Internet over 13M domains, and get 458K Spider Pool domains
3. Analyze Spider Pool domains, their distribution, customers, impacts on SEs
4. Deploy our detection in Baidu, and get another ~1M SEO domains

Baidu 百度

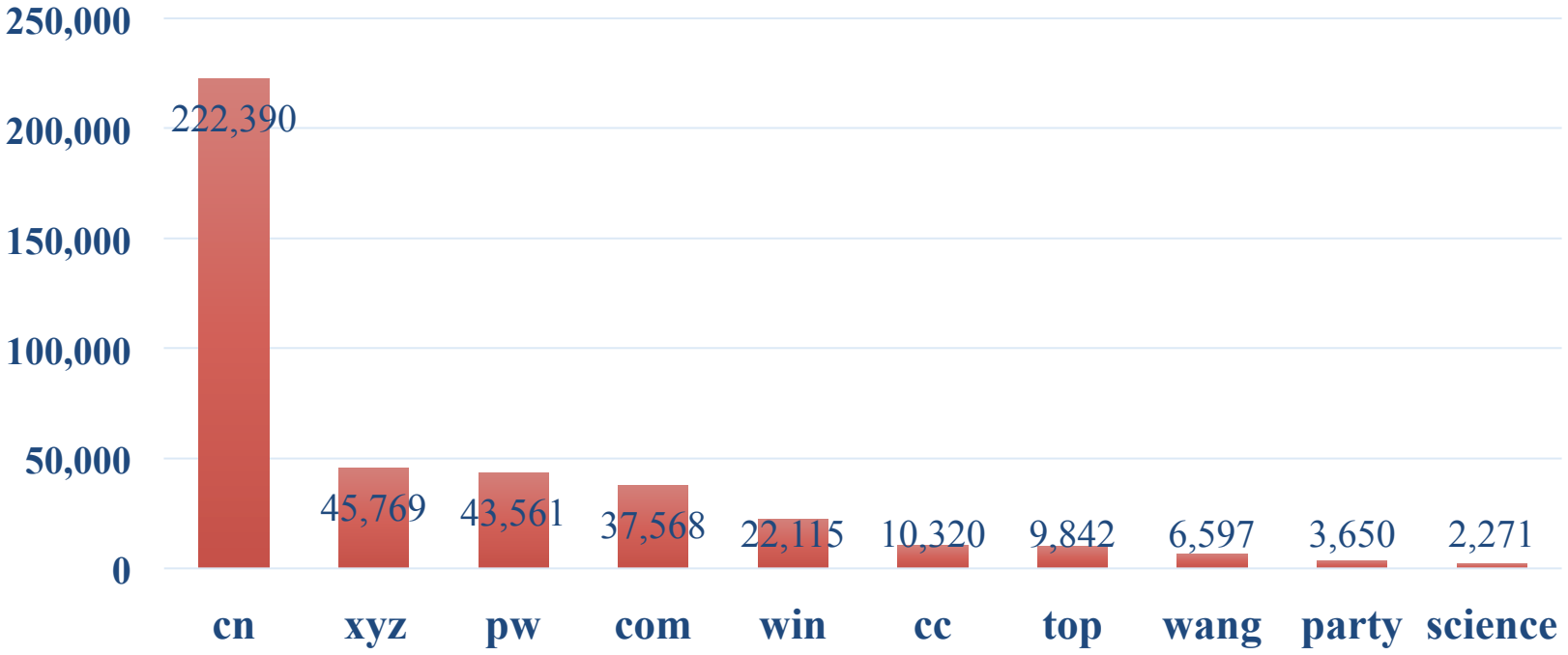
How we collect domains

Category	TLD&SLD	Data Source	#All	#Wildcards	#All SEO	#iframe	#link	#redir
Old TLD	cc	Passive DNS	99,934	40,334	10,320	1,499	1,210	7,611
	cn	Domain List	4,107,679	371,829	222,390	52,013	4,099	166,278
	com	Zone file*	2,000,000	751,877	37,568	3,995	10,137	23,436
	edu	Passive DNS	3,619	353	0	0	0	0
	gov	Passive DNS	1,215	130	0	0	0	0
	pw	Passive DNS	246,775	146,984	43,561	18,239	24,154	1,168
New gTLD	xyz	Zone file	1,695,430	309,983	45,769	581	2,532	42,656
	top	Zone file	1,033,644	127,532	9,842	5,378	1,198	3,266
	wang	Zone file	629,757	49,528	6,597	680	287	5,630
	win	Zone file	570,091	56,020	22,115	1,865	2,696	17,554
	club	Zone file	535,576	108,665	2,163	149	403	1,611
	science	Zone file	249,187	44,544	2,271	1,097	1,024	150
	ren	Zone file	237,372	11,023	121	5	4	112
	link	Zone file	213,449	40,719	490	146	244	100
	party	Zone file	217,508	68,239	3,650	874	2,726	50
	click	Zone file	181,673	43,914	1,059	614	378	67
SLD	ac.cn	Passive DNS	86,291	56,944	24,270	977	9,249	14,044
	edu.cn	Passive DNS	1,347	173	0	0	0	0
	gov.cn	Passive DNS	12,364	1,745	24	0	0	24
	org.cn	Passive DNS	53,492	5,206	805	141	34	630
	com.cn	Passive DNS	1,093,580	152,751	23,264	3,779	2,156	17,329
	net.cn	Passive DNS	234,039	12,445	1,967	356	45	1,566
Total	-	-	13,504,022	2,400,938	458,246	92,388	62,576	303,282

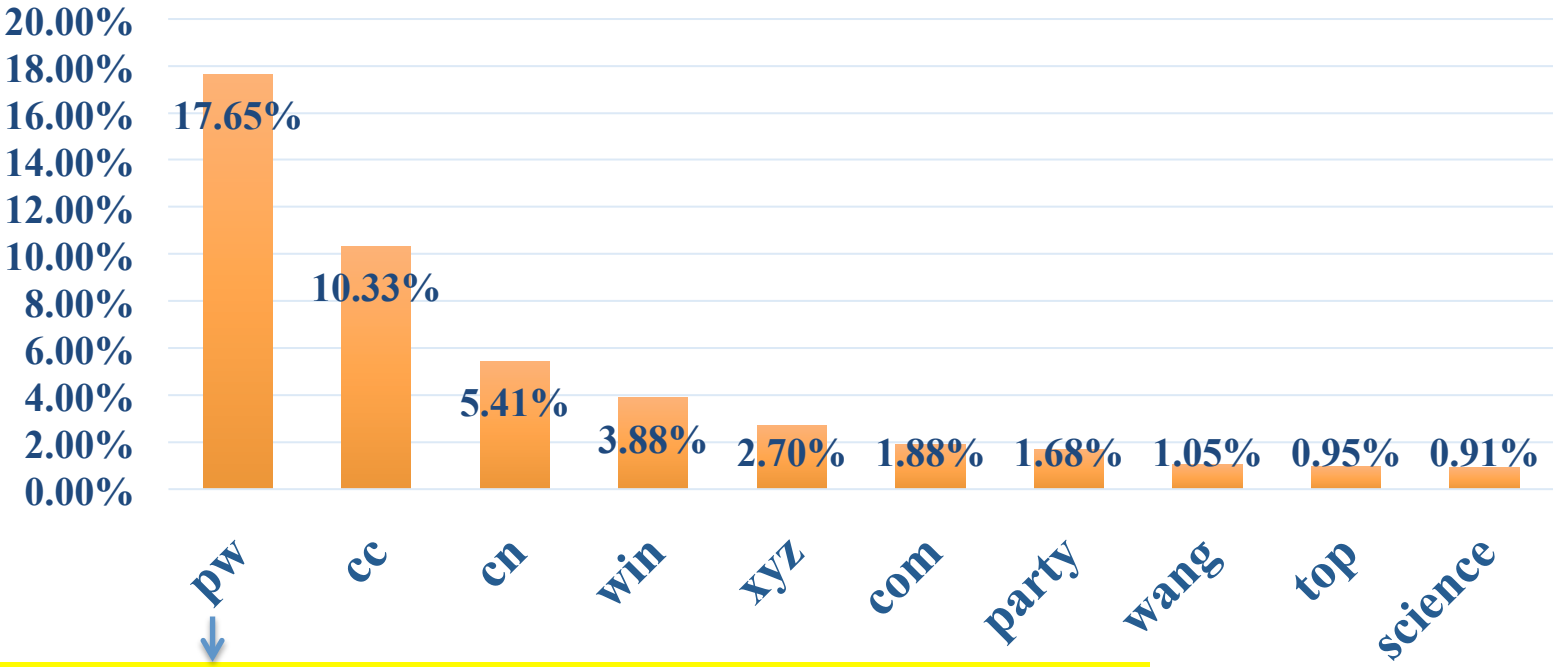
Spider Pool Detection



Spider Pool Domains: Top 10 TLDs



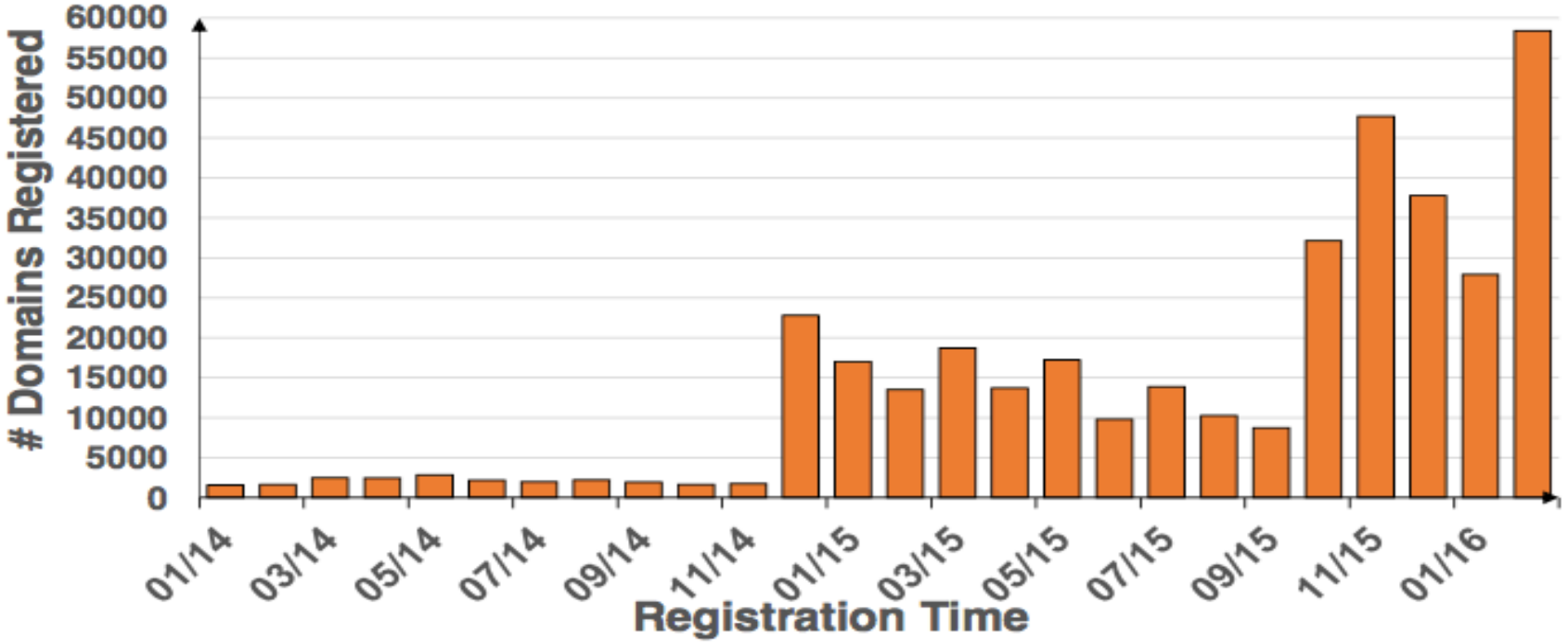
Top 10 Spider Pool Domain: Percentage



PW: ccTLD (Country Code TLD) for Republic of Palau

In 2013, Symantec claimed .PW was a significant source of spam e-mail

Registration time of spider pool domains



Top 10 registrars of SEO domains

- Top 10 registrar accounts for 81% SEO domains
- Most in China
- China has strict policies on domain registration, right?

NO.	Registrar	#Domain
1	Chengdu West Dimension Digital Technology CO., Ltd.	89,378
2	ERANET International CO., Ltd.	53,307
3	Xiamen Nawang Technology CO., Ltd.	39,837
4	FABULOUS.COM PTY LTD.	38,398
5	Alpnames Limited	38,247
6	Alibaba Cloud Computing Ltd. d/b/a HiChina (www.net.cn)	23,583
7	PDR Ltd. d/b/a PublicDomainRegistry.com	19,358
8	HANGZHOU Dnbiz Network CO., Ltd.	15,918
9	XIN NET TECHNOLOGY CORPORATION	15,379
10	HANGZHOU AIMING NETWORK CO.,LTD	14,587
Total	-	347,992

Top 10 registrants of SEO domains

- Top 10 registrants control 51.9% SEO domains
- The business is quite centralized, Chinese

NO.	Email	#Domain
1	11*44*42*2@qq.com	38,727
2	je*ny*ro*nb*lk*@gmail.com	32,667
3	13*77*17*5@qq.com	31,285
4	whois.private.service@gmail.com	26,606
5	15*59*72*73@163.com	19,837
6	yu*on*li*g5*1@163.com	19,237
7	du*on*56*02*8@126.com	15,839
8	go*go*@365.com	13,731
9	xi*os*ou*um*ng@163.com	12,712
10	13*19*96*@qq.com	10,213
Total	-	220,854

Where are the web servers hosted?

Top 5 countries hosting SEO domains.

Country	#IP	#Domain
United States	22,958	349,879
HongKong	2,319	19,483
China	1,958	13,295
Australia	135	6,387
Japan	285	3,459
Total	27,655	392,503

Register domains in China, but host servers in the U.S. Why?

Top 10 ASNs hosting SEO domains (All in US)

NO.	ASN	#IP	#Domain
1	AS6939	35	94,372
2	AS18779	534	57,286
3	AS38197	1,354	47,790
4	AS8100	546	45,808
5	AS15003	6,398	34,583
6	AS18978	3,719	27,325
7	AS40676	5,132	25,384
8	AS32097	431	18,152
9	AS209	3	17,668
10	AS32787	6	12,673
Total	-	18,158	381,061

Sample Spider Pool Campaigns

- 21 independent spider pools
- 15,816 SEO Domain
- 1,453 customer domains
- 7,236,315 promoted URLs

ID	Seed	SEO	Customer	
		#Dom	#Dom(S)	#URL(M)
SSP	mianmodaili14.cn	514	14	463
S_1	annunciincontri.top	494	0	0
S_2	jjytkvk.xyz	100	18	308
S_3	1559535.pw	945	22	13,247
S_4	604462.win	738	21	10,292
S_5	10086wxu.com	699	295	35,295
S_6	00u56m.pw	537	59	63
S_7	2janp3.science	583	710	1,828,669
S_8	mzysw.cn	740	213	208,486
S_9	zhcocu.cn	582	105	61,444
S_{10}	01q.ac.cn	984	65	19,006
S_{11}	432364.party	309	19	156,322
S_{12}	ckocn.club	768	319	2,651,720
S_{13}	srkros.com.cn	713	14	1,144
S_{14}	0acrn.pw	677	219	45,171
S_{15}	miead.cn	713	591	4,917
S_{16}	noykr.cn	879	45	6,703
S_{17}	4be9l.ac.cn	1,243	81	1,778,360
S_{18}	exzgyh.science	1,933	89	138,656
S_{19}	lingganpj099.science	1,033	75	102,847
S_{20}	usa4.win	632	94	172,613
Total	-	15,816	1,453	7,236,315

Customers of Spider Pools

Topic	#Domains	Ratio
Sales and Services	202	21.72%
Gambling	190	20.43%
Surrogacy	156	16.77%
News	156	16.77%
Sex	114	12.26%
Games	84	9.03%
Hospitals and Drugs	28	3.02%
Total	930	100%

More than half of their customers are illegal in China

QQ numbers to contact customers

NO.	QQ Number	Topic	#URL
1	53*95*1	Sex	156,306
2	34*06*58*	Drugs	261,135
3	37*40*42*	Sex	225,413
4	41*88*10*	Drugs	186,370
5	51*89*17*	Drugs, Sex	224,310
6	51*45*63*	Drugs, Sex	155,459
7	63*11*95*	Drugs, Sex	150,176
8	71*40*27*	Sex	225,655
9	76*26*01*	Sex	222,864
10	77*72*10*	Drugs, Sex	224,476
11	79*43*07*	Sex	225,000
12	10*36*03*3	Sex	5,986
13	10*33*91*1	Drugs, Sex	23,944
14	10*38*34*5	Drugs, Sex	17,267
15	11*07*07*2	Drugs, Sex	19,405
16	11*35*39*2	Sex	19,883
17	11*58*92*2	Sex	16,393
18	14*14*38*6	Gambling, Sex	6,906
19	15*53*65*8	SEO	5,339
20	19*82*12*3	Gambling, Sex	6,644
21	24*50*13*5	Sex	24,949
22	24*92*74*4	Gambling	24,736
23	25*17*02*5	Gambling, Sex	6,495
Total	-	-	2,435,111

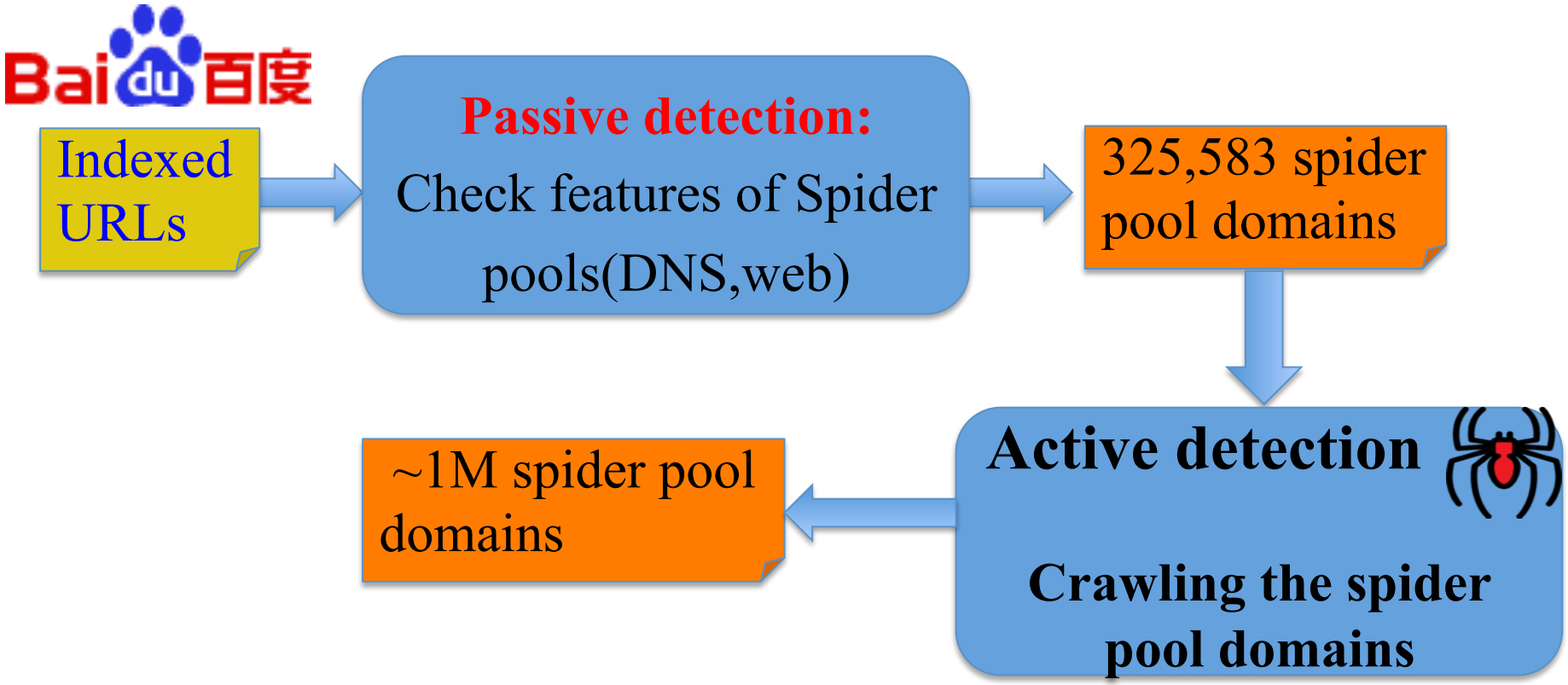
Search in Google and Baidu, in the first page…

Keyword	Category	Google	Baidu
深圳南山小姐一夜多少钱 (How much is prostitute in Shenzhen)	Porn	100%	100%
铜川哪里有迷情水卖的 (Where to buy hallucinogen in Tongchuan)	Drug	70%	22.5%
海南省三亚市哪里有汽枪买 How to buy air gun in Hainan Sanya city	Gun	37.5%	10 %
岳阳开发票公司 How to buy fake invoice in Yueyang city	Fake certificate	87.5%	32.5%

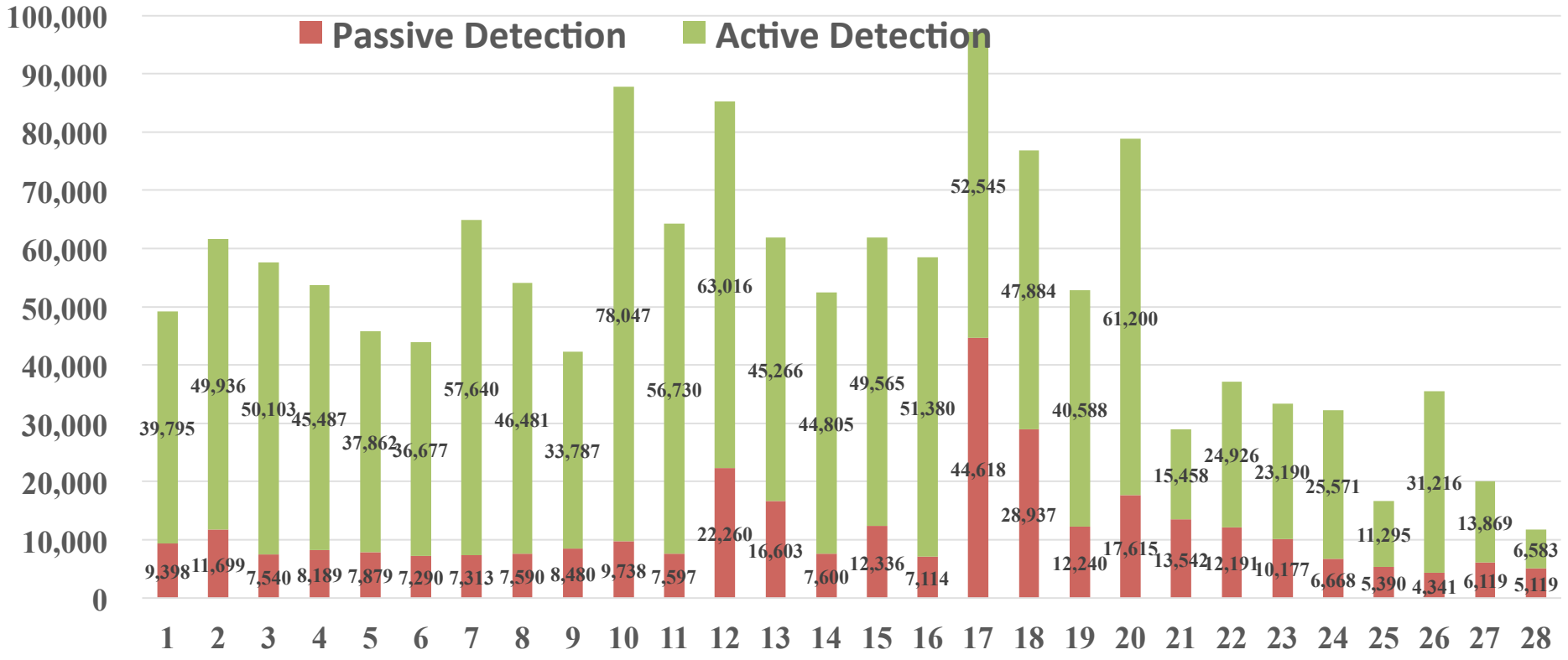
Top 20 reputable sites abused by spider pools and URL pattern

Alexa Rank	Abused Site	URL Pattern
323,377	baicai.com	http://www.baicai.com/salary-[MESSAGE]/
104	sogou.com	http://www.sogou.com/tx?word=[MESSAGE]
13	sina.com.cn	http://search.sina.com.cn/?q=[MESSAGE]
766,933	taofang.com	http://www.taofang.com/w_[MESSAGE]/
56,687	poco.cn	http://my.poco.cn/tags/tag_search.php?q=[MESSAGE]
3	amazon.com	http://www.amazon.com/s/ref=nb_sb_noss? url=search-alias%3daps&field-keywords=[MESSAGE]
22	ebay.com	http://www.ebay.com/sch/i.html?_nkw=[MESSAGE]
12,520	qzone.cc	http://www.qzone.cc/zipai/search/[MESSAGE]
1,274	xiami.com	http://www.xiami.com/search/song-lyric/h?key=[MESSAGE]
8	qq.com	http://v.qq.com/page/j/d/s/[MESSAGE]
88	jd.com	http://search.jd.com/search?keyword=[MESSAGE]
2,677	mafengwo.cn	http://www.mafengwo.cn/group/s.php?q=[MESSAGE]
35,810	chazidian.com	http://zuowen.chazidian.com/index.php?q=[MESSAGE]
38,942	mininova.org	http://www.mininova.org/search/?search=[MESSAGE]
1,489	bab.la	http://it.bab.la/dizionario/cinese-inglese/[MESSAGE]
3,57	enet.com.cn	http://www.enet.com.cn/enews/[MESSAGE]
65	tianya.cn	http://bbs.tianya.cn/index_self.jsp?key=[MESSAGE]
9,314	wasu.cn	http://www.wasu.cn/search/show/k/[MESSAGE]
93,010	yododo.com	http://www.yododo.com/search/searches.ydd?keyword=[MESSAGE]
277	douban.com	http://www.douban.com/group/search?q=[MESSAGE]

Deployment in Baidu.com, 1 month



Spider Pool Domains Detection in Baidu



July 11

Aug 7

Summary

- We infiltrated a new type of Blackhat SEO and and get the features of its infrastructure.
- Together with Baidu, we detected ~1.4M spider pool
- To SEs: be aware of this new threat and purify search result,
- To reputable sites: not be abused for free riding
- Useful information to attack cyber crimes, e.g. QQ# of who sells hallucinogen

Thank You!

Q&A