360 INTERNET SECURITY CENTER

Long Tail domain Clustering Analysis

Zhang Zaifeng
Network Security Research Lab
QIHOO 360

2015.10

About passiveDNS.cn

360 INTERNET SECURITY CENTER

- PassiveDNS.cn Database
  - Based on ~ 15% DNS traffic in China
  - Last 13 months (2014.08~2015.09)
    - Unique domain: 4.7 billion
    - Rrsets:5.7 billion
    - Rdata:17.4 billion
  - First and biggest public database in China
  - Open to communities (nsp-sec, ops-trust…)

Agenda

360 INTERNET SECURITY CENTER

- Why clustering domains
- Data selecting
- Cluster methods
- Result analysis
- Future works

Why clustering domains

360 INTERNET SECURITY CENTER

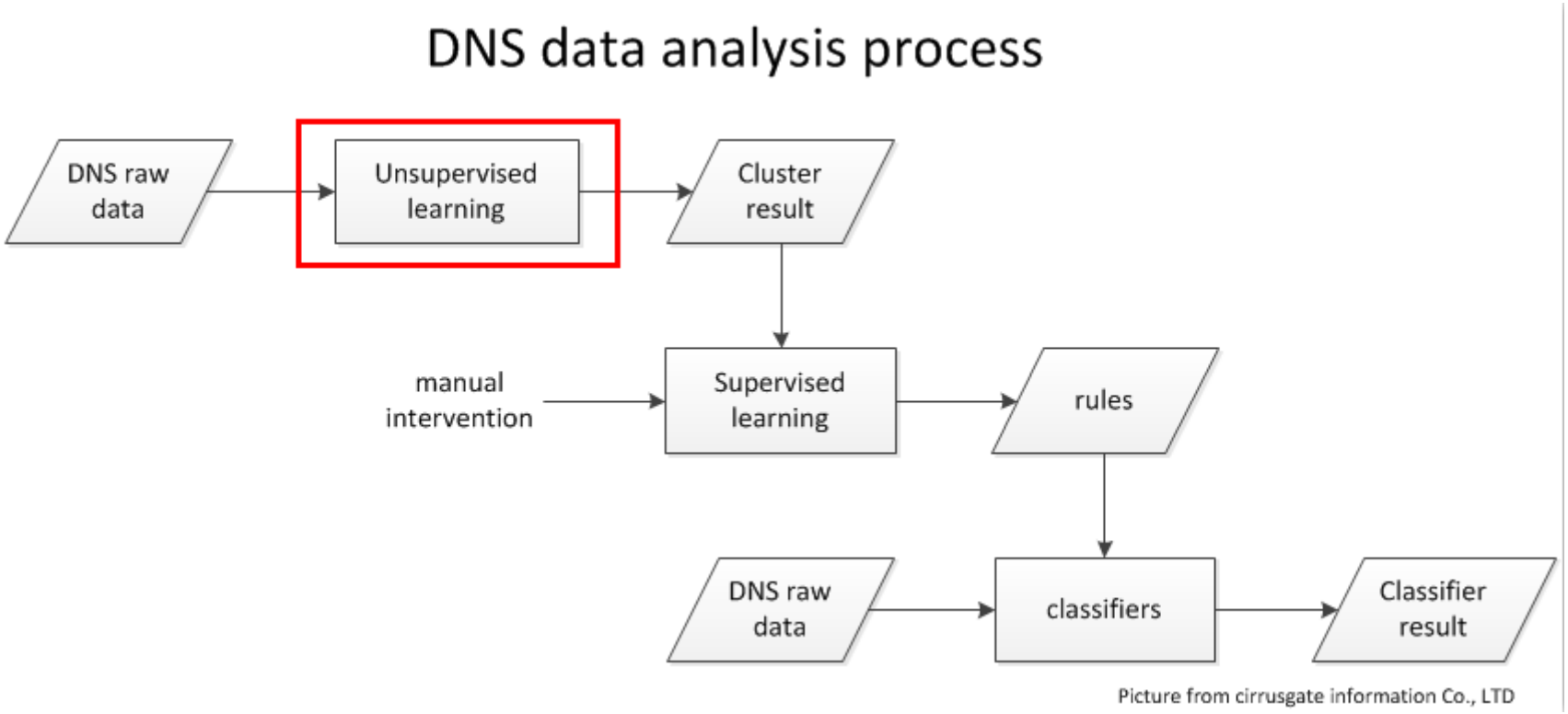
- Know lots applications based on DNS
  - DNS tunnel, CDN domains, Anti-virus/DNSBL, DGA domains(botnet), browser prefetching …
- What's the following domains used for?

1059.xiaoyun.com	coquc24864.bpjer.cn	10510185.weimob.com.cn-aute.com.cn	rvix.qtye175.pw
1027.xiaoyun.com	rylcx79812.holcj.cn	13154585.weimob.com.yiqunlongye.cn	iiii.qtye173.pw
1011.xiaoyun.com	ybjbc50404.argvi.cn	40973102.weimob.com.datainterest.cn	rdpq.qtye174.pw
10669.xiaoyun.com	dcvrg37213.nwbey.cn	10012962.weimob.com.meitaoxuan.com.cn	yfq8rpttnwwtr.qtye173.pw
10693.xiaoyun.com	hghol74680.sqhcy.cn	14553372.weimob.com.6666666.xj.cn	dd5qnkdzher.qtye174.pw
10722.xiaoyun.com	tfbkt12118.jmjko.cn	1023935.weimob.com.yuanfangyaoye.cn	saswzzspthy2tkqy.qtye174.pw
10658.xiaoyun.com	gvffk16862.zcpls.cn	12996323.weimob.com.loveshenghuo.cn	3e3eteshkf2.qtye173.pw
10302.xiaoyun.com	xxekn73496.bpjer.cn	26845518.weimob.com.yuerentang.com.cn	pppjrr73ctyg4.qtye174.pw
10075.xiaoyun.com	czexv26357.holcj.cn	55004702.weimob.com.ccdiamond.cn	iiiq.qtye172.pw
101.xiaoyun.com	pprka90575.ytxjc.cn	10445762.weimob.com.cdjjzy2009.com.cn	sjlz.tcye170.pw
10480.xiaoyun.com	twjga84396.nwbey.cn	20584492.weimob.com.duanxin.sd.cn	zpc.zhy163.pw
10464.xiaoyun.com	107343.jhbv.science	1065011.weimob.com.6666666.xj.cn	b1.tcye169.pw
10384.xiaoyun.com	nafqjq.tabyi.science	53569154.weimob.com.kangjingshun.cn	bscmds.zhy172.pw

- Emerge thousands of new applications each day

Why clustering domains cont.

360 INTERNET SECURITY CENTER



Data selecting

360 INTERNET SECURITY CENTER

- Focus on long tail domains
  - Malicious domains generally not popular
  - New business domains generally not popular

- How we pick the domains:
  - active domains(Last 7 days)
  - SLD not on Alexa top 500,000
  - First seen in last 3 months
  - other filters
    - DNSBL
    - Bit torrent tracker server
    - PTR records domains
    - Wrong configure servers
    - …
- Data set scale
  - FQDN: 2 million/day

Cluster methods

360 INTERNET SECURITY CENTER

- Domain structure
  - 13 features
    - Domain entropy(max, min, mean, median)
    - Alphabets number
    - Max length of alphabets
    - Digits number
    - Max length of digits
    - Hyphen number
    - Dots number
    - Sub-domain number
    - Domain length
    - TLD length
  - Clustering methods
    - Kmeans

Cluster methods cont.

360 INTERNET SECURITY CENTER

- Domain structure
  - some clusters are accurate

5	zxu3.ermv.kv1a-zzc.waic13.pw	12	shop100001068.tj88.com	25	yeyecao.daiyun888.com
5	zxxh.egtw.kv1a-zzc.waic15.pw	12	shop100001079.tj88.com	25	yeyecao.daiyun999.com
5	zxxj.utey.kv1a-zzc.waic15.pw	12	shop100001084.tj88.com	25	yeyegan.daiyun888.com
5	zxy1.blui.kv1a-zzc.waic21.pw	12	shop100001087.tj88.com	25	yeyegan.daiyun999.com
5	zxzv.ximd.kv1a-zzc.waic13.pw	12	shop100001092.tj88.com	25	yeyelu.nvzhuang88.com
5	zy2t.bvtv.kv1a-zzc.waic15.pw	12	shop100001102.tj88.com	25	yeyelu.yxblgfj888.cn
5	zy6c.joua.kv1a-zzc.waic14.pw	12	shop100001103.tj88.com	25	yeyeqing.daiyun888.com
5	zycb.ccra.kv1a-zzc.waic22.pw	12	shop100001105.tj88.com	25	yeyexxx.daiyun888.com

  - But some not

133	xiaobaohanxiaoqian1.lighting86.com.cn	170	www69caocom.akkpp.com	155	zxm42fkhzdcq6.qtye173.pw
133	xnnigdt.devlabrps.dom2.redprairie.com	170	www-6.cloudcenter.fi	155	zy95c3c587.dedeadmin.com
133	x8.xf.packetix.servers-v6.ddns.sesvc.cn	170	www-6pdy-com.diyys.com	155	zz190c3c150.dedeadmin.com
133	x9eieh6okaa4tfa.dns.turbobytes.com	170	www74eeecom.wzrga.com	155	zz294c3c485.dedeadmin.com
133	xagt8c.mbj1ibioioegy.ovhdnllxxg.info	170	www81seseoyg.dmmzz.net	155	zz2j6zn397.nenggaoxq.com
133	xbkenoy.devlabrps.dom2.redprairie.com	170	www82aaacom.nkkaa.com	155	zz3r5kdj8tqskp.tly177.pw
133	xcdqkmwkjhs4y3qpxfgknekdrrz.tly177.pw	170	www84ytcom.dabolu.cc	155	zz599c3c843.dedeadmin.com
133	xcpytmskhxmjzjpp7yierjisyh.tly177.pw	170	www88ququco.dxxpp.com	155	zz660c3c700.dedeadmin.com
133	xcywscppwhkrcdj5yjfdscnng.tly179.pw	170	www95abcdco.hrraa.com	155	zz69c3c250.oceanweek.net
133	xfl4mpcbz6.beihaikaiyuanmingdu.com	170	www97paocom.taahh.net	155	zz6xmwd75mqmcrs.tly177.pw
133	xfvcycnl8vnv.1mqustm.egkqcuhg.info	170	www9zyl.yagegezhw.org	155	zzjx58.w221-e0.ezcname.com
133	xianggangcaipiaot35cczilіao.6hc139.top	170	www-daraz-pk.ax4z.com	155	zzp3tjhdhpc63f.tly179.pw

Cluster methods cont.

360 INTERNET SECURITY CENTER

- Problem with domain structure cluster
  - Feature adjusting inefficiency
    - A big chunk of domains can be nicely classified
    - But the rest needs lots of tweaking|feature adjusting
  - Domain structure is only one perspective
    - Need consider other aspects
      - Server IPs, Client IPs, Name Servers, Whois data…
  - Missing co-occurrence relations

Cluster methods cont.

360 INTERNET SECURITY CENTER

- Adding new dimensions
  - client IP addresses
  - Domain IP addresses
- Steps:
  - Generate domain-IP matrix(A) and IP-domain matrix(A^T)
  - Calculate similarity matrix $B = A*A^T$
  - Calculate the pairwise domains' similarity
  - Cluster the similar domains

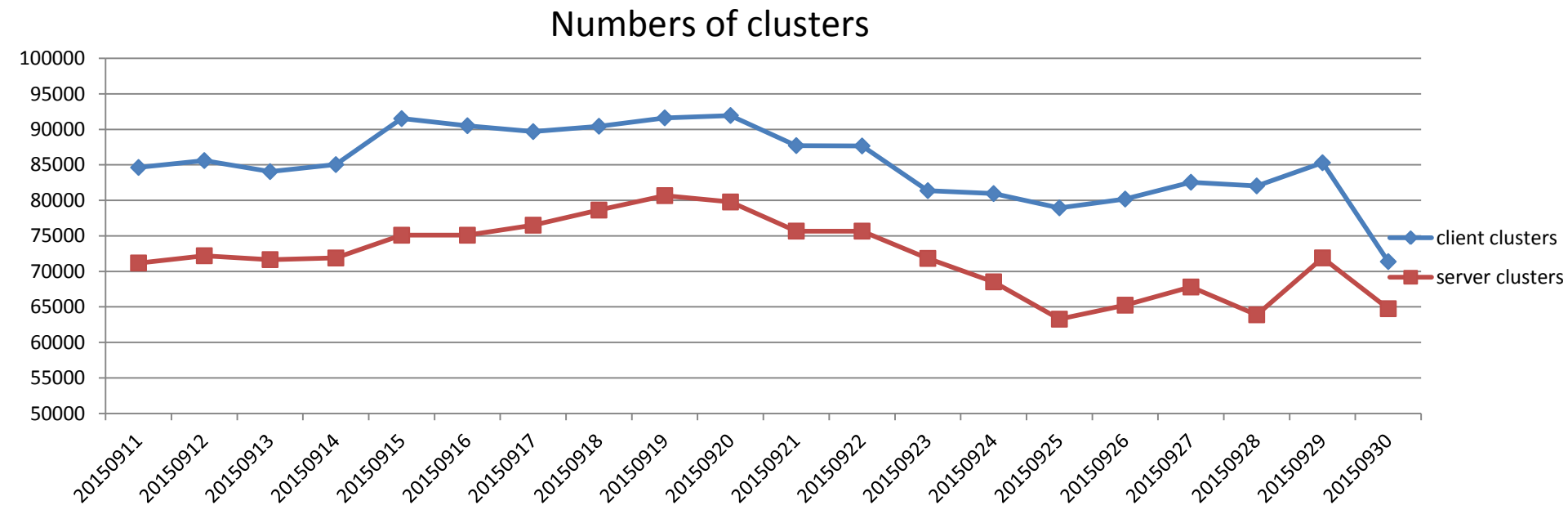
Cluster methods cont.

360 INTERNET SECURITY CENTER

- Challenges & solutions
  - Expensive (similarity) calculation
    - Complexity N^2 (N is the number of domains)
    - Compress the data size
      - Discard the single client domains
    - Spark + sparse matrix multiplication
      - 300,000 vertex, 15,000,000 edges , about 1 hour
  - Algorithm (cluster):
    - Louvain method

Result analysis

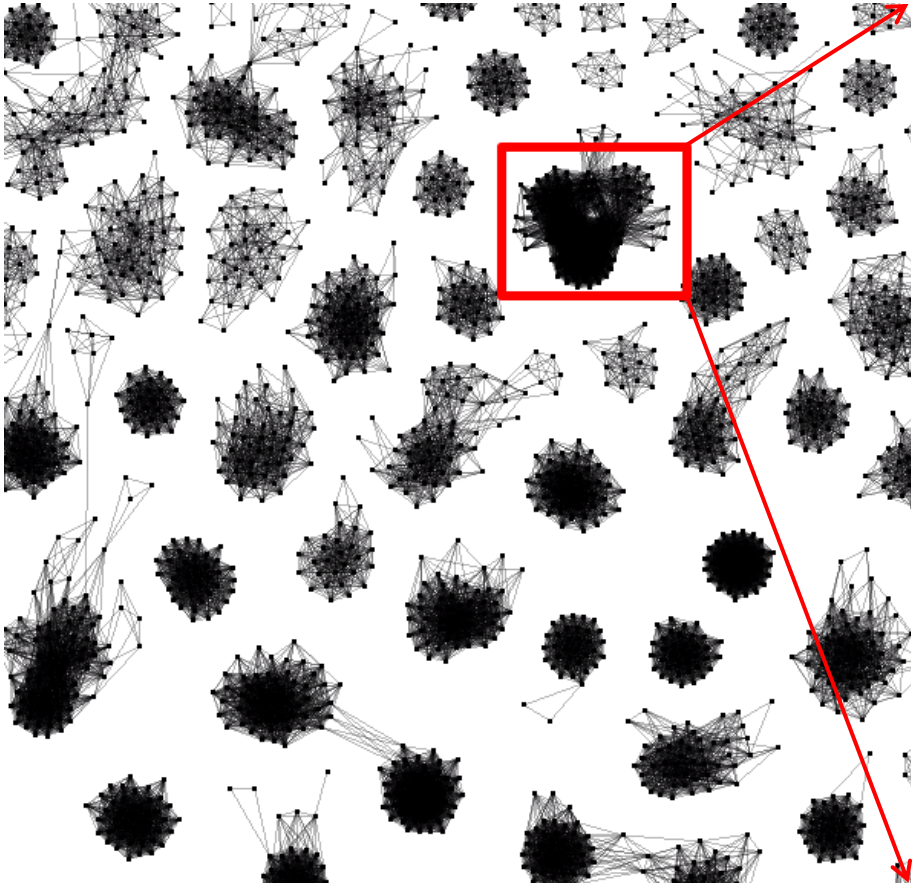
- Numbers of clusters
  - Client clusters: 85,163/day
  - Server clusters: 72,064/day

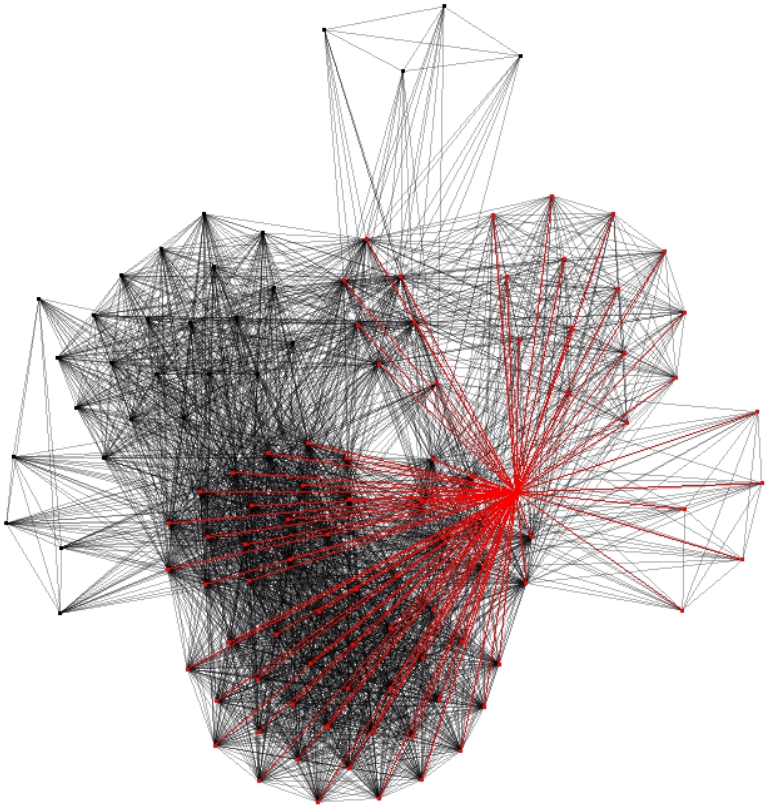


Result analysis

360 INTERNET SECURITY CENTER

- Cluster of client similarity
  - Conficker DGA domains





Result analysis

360 INTERNET SECURITY CENTER

- Some interesting/strange findings
  - Lots of NX government domain
    - maybe some type of scanner

5033	tlfsyc.gov.cn	5033	synjrd.gov.cn	5033	dachang.gov.cn
5033	gyetdz.gov.cn	5033	abghjs.gov.cn	5033	nlkxwsj.gov.cn
5033	nyzfcg.gov.cn	5033	jncqmz.gov.cn	5033	tulufan.gov.cn
5033	ldxgsj.gov.cn	5033	wyxrlj.gov.cn	5033	wlcb12365.gov.cn
5033	qhdagele.gov.cn	5033	lygafj.gov.cn	5033	fzja12315.gov.cn
5033	ynztglzj.gov.cn	5033	gzltax.gov.cn	5033	ktcj12319.gov.cn
5033	lpszsqcg.gov.cn	5033	ptxydz.gov.cn	5033	tzjj12345.gov.cn
5033	zqboftec.gov.cn	5033	xnghjs.gov.cn	5033	sxxz12319.gov.cn
5033	snqizhen.gov.cn	5033	jnflzx.gov.cn	5033	hhht12319.gov.cn
5033	pyhengli.gov.cn	5033	lntazj.gov.cn	5033	jiangning2013.gov.cn
5033	gcdongba.gov.cn	5033	myqjsw.gov.cn	5033	shuozhou81.gov.cn
5033	hdwaishi.gov.cn	5033	hblsgt.gov.cn	5033	huazhoucaizhengju.gov.cn

  - online marketing (Chinese most popular APP WeChat)

76072	61951	086ae2e.makax.net	76072	61951	cb70f0d.changfenxiang.cn
76072	61951	71d36cc.makax.net	76072	61951	e56739b.makax.net
76072	61951	e8b04c9.changfenxiang.com	76072	61951	70ec3f5.changfenxiang.net
76072	61951	4b20768.changfenxiang.net	76072	61951	64e6ac0.makax.net
76072	61951	0e49a09.changfenxiang.com	76072	61951	273a5f9.changfenxiang.net
76072	61951	c360931.changfenxiang.cn	76072	61951	17836a1.changfenxiang.net
76072	61951	e46b6f9.makax.net	76072	61951	f8bc986.changfenxiang.net
76072	61951	c484e57.changfenxiang.com	76072	61951	6574ad3.changfenxiang.net
76072	61951	0207f3b.changfenxiang.net	76072	61951	83e6910.makax.net
76072	61951	715336f.makax.net	76072	61951	4c76f4b.makax.net
76072	61951	2644317.changfenxiang.cn	76072	61951	086ae2e.makax.net
76072	61951	1490650.makax.net	76072	61951	71d36cc.makax.net

Result analysis

360 INTERNET SECURITY CENTER

- DGA domains
  - Data has been checked by DGArchive

GOZ		UnKnow DGA		necurs	
597	1709wh4mbm9yd5ltkfu1iao07c.com	51969	feaoru.com	18169	nflwdoithuydftxyk.to
597	19zglhjnac1w6115wl1i1iq00tj.com	51969	xaeypl.com	18169	jxuxpndjtdqcpmxum.nf
597	4tr3i313tu11k1dfkf0s1folijr.net	51969	qfsijb.com	18169	xetsjrhodlvsnptri.ga
597	eydwi81di3c7613unybh1v119f5.org	51969	iknjat.com	18169	shcuinurdcvdlxjec.xxx
597	14wxrv21021dr3qlrh3x13vtq4r.com	51969	icqxap.com	18169	miwkouuabultnfwlg.ir
597	1ad55gs4klp7h8ll8ut1ntz61w.net	51969	asyzwc.com	18169	lndnfpfmxx.co
597	170xbtba47w6k1lxito218bh5fo.com	51969	sjxtui.com	18169	nuhuhwrfhp.to
597	16l7jbeag3sn11p4aqeu35zm40.net	51969	qkpeay.com	18169	ykxqurpuyv.net
597	1s57bt17dj61i1ms3ibf1jzvdjc.biz	51969	dvmpay.com	18169	hanjvgikps.nf
597	13kxm6m1pj4zhz1sdn6x71mf7cws.org	51969	uryafd.com	18169	nffjmgtwla.pro
597	yfv1wrchmg5h146e2y92i8nxd.com	51969	peutxa.com	18169	laccheqbgj.kz
Conficker		**UnKnowDGA**			
27698	gdtlab.ws	3839	alzovuk-ahuz.ru		
27698	pfnago.ws	3839	albobib-efal.ru		
27698	bdgtqi.biz	3839	albijyw-ybew.ru		
27698	xqdpmi.biz	3839	aldozah-ifir.ru		
27698	krvbop.ws	3839	almezy-kivib.com		
27698	gxifzk.biz	3839	alqisy-bivar.com		
27698	ndqvia.biz	3839	albevar-ifeb.ru		
27698	kwannfcrxfy.biz	3839	alzafag-ywuz.ru		
27698	pqsymoizulz.ws	3839	algibi-vyfad.com		
27698	nprileqoihp.biz	3839	alqifo-fyfaq.com		
27698	bmhlwsedwtf.biz	3839	alkyni-fajav.com		

Future works

360 INTERNET SECURITY CENTER

- Domain, more accurate
  - quantitative evaluation result
  - Adjust the algorithm parameter
  - Adjust input data
  - More domain dimensions
    - Whois, name server…
  - Combine/intersect different cluster type

- IP cluster
  - Find the similar IP addresses
    - Port, domain, ASn

https://passivedns.cn

Thanks

zhangzaifeng@360.cn

360 INTERNET SECURITY CENTER | SAFETY FIRST TRUST 360

Reference

- https://dgarchive.caad.fkie.fraunhofer.de/site/
- http://faculty.cs.tamu.edu/guofei/paper/SMASH-ICDCS15.pdf