

非法在线博彩分析与研究

分享人：杨皓

清华大学NISL实验室

目录

1. 背景介绍
2. 检测系统
3. 测量与分析
4. 校园网部署实践

目录

1. 背景介绍
2. 检测系统
3. 测量与分析
4. 校园网部署实践

非法在线博彩危害大



葡京、杏彩
BET365.....



形式多样，种类繁多

在线博彩

十几万中国人在菲律宾被奴役:专坑同胞的陷阱,竟遍布在国内正规...



2018年8月24日 - 虽然在线博彩设立在菲律宾,但是几乎所有在上面投注的,都是中国人。这也是为什么在菲律宾,那么多中国人从事这个行业,这些形形色色的博彩公司幕后的老板,也...

[finance.ifeng.c...](https://finance.ifeng.com/) - 百度快照

成都创业青年身陷网络博彩平台 一年输掉300多万_大成网_腾讯网



2017年11月6日 - "天游娱乐"平台网站的简介显示,该在线博彩公司于2009年成立于菲律宾,在王鹏赌博期间,一位QQ名为"天游事务·力哥"的人与他保持着紧密联系,"是朋友...

<https://cd.qq.com/a/20171106/0...> - 百度快照

在线博彩公司GVC将收购英国博彩公司Ladbrokes Coral,交易价值约3

2017年12月22日 - 在线博彩公司GVC将收购英国博彩公司Ladbrokes Coral,交易价值约32亿英镑。进入【新浪财经股吧】讨论责任编辑:王永生 SF153...

finance.sina.com.cn/7x... - 百度快照

你下注时,博彩公司在做什么? 腾讯网 腾讯体育



2014年7月7日 - 时下,各种投注方式犹如瀑布般在实体店、在线商店以及移动客户端倾泻而下,但是无...世界杯开赛前,立博国际博彩公司预期在世界杯期间接受投注额1亿英镑。虽然英格兰早...

<https://sports.qq.com/original...> - 百度快照

虚假网站安排美女在线发牌 千余人赌博被骗(图)_新闻_腾讯网



2016年3月24日 - 误以为这就是澳门赌场的官方网站,每天多达上万人参与在线"赌博",案件...而假冒博彩网站的服务器则设在柬埔寨。该团伙从中国内地招募人员,以旅游签证...

<https://news.qq.com/a/20160324...> - 百度快照

经济损失、社会危害大

为什么要研究非法在线博彩

研究非法在线博彩的重要性

- 非法在线博彩是网络犯罪中的重要组成部分
- 当前缺乏对非法在线博彩产业链、运维模式的了解
- 理解非法在线博彩，对净化网络空间有很大帮助

主要挑战

- 为了逃避监管，非法在线博彩具有天然的隐蔽性，导致缺乏有效的检测入口
- 缺乏实时、有效、大规模的数据集用于分析

主要贡献

1. 实现了基于文本语义的非法博彩检测系统
 - 检测出967,954个博彩页面
2. 首次对非法博彩网站的游戏类型及网络基础设施进行了分析
3. 首次对第三方支付、第三方客服应用及第三方存储的滥用进行了分析
4. 首次对非法博彩网站的推广方法进行了分析

目录

1. 背景介绍
2. 检测系统
3. 测量与分析
4. 校园网部署实践

用户访问博彩网站流程



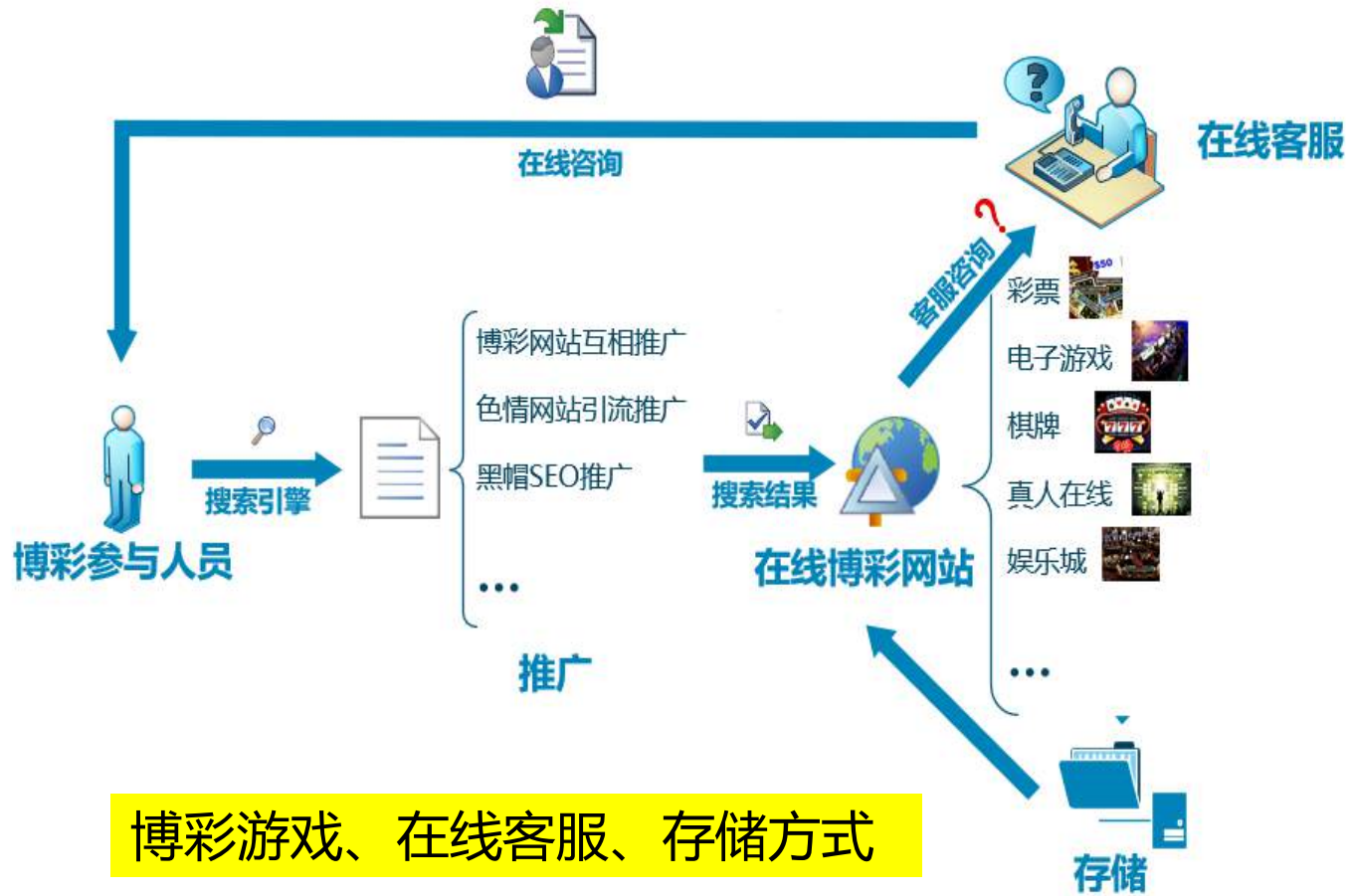
博彩参与人员

导流入博彩页面

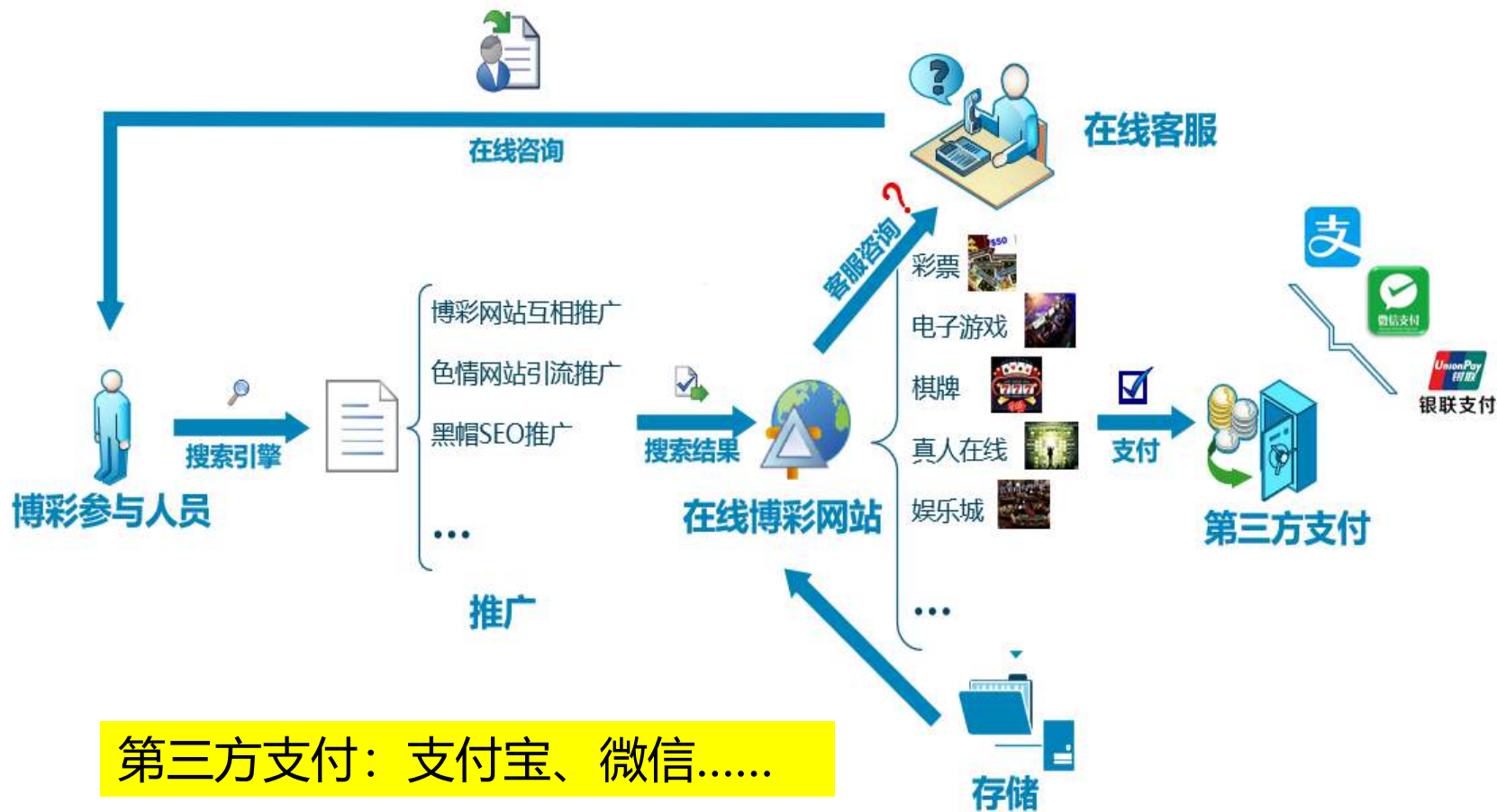


博彩、色情网站引流、SEO推广

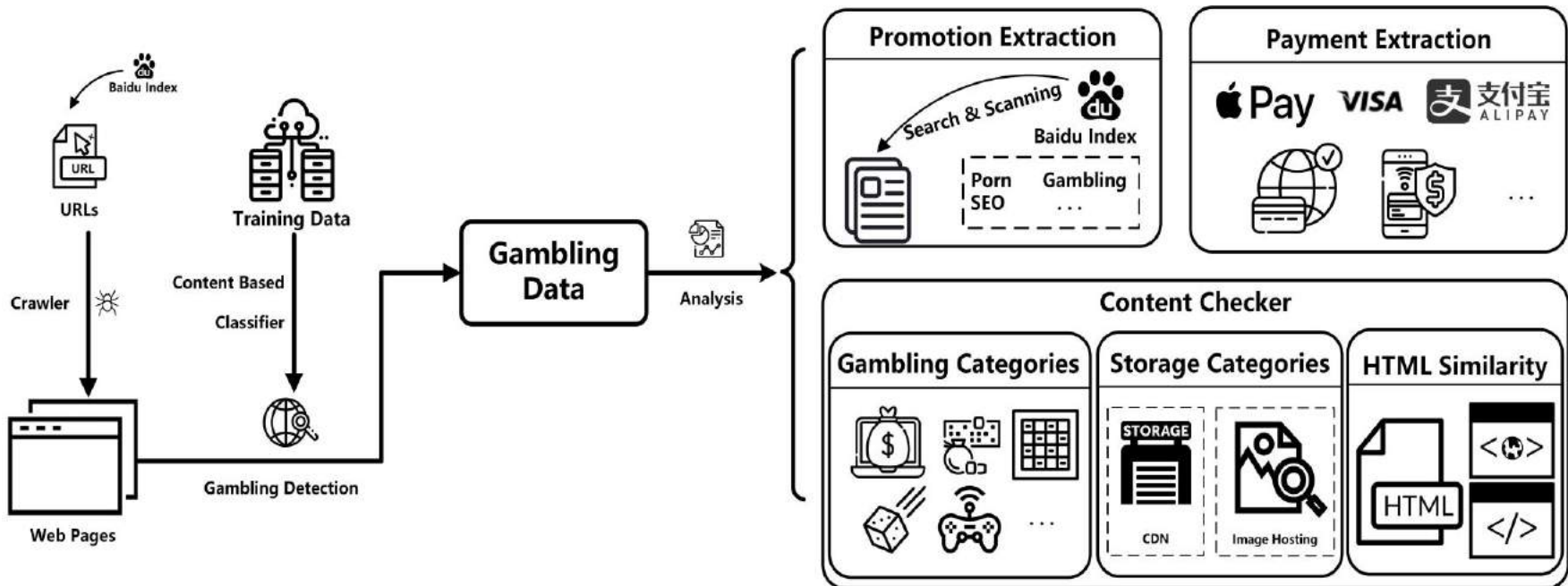
博彩页面内容



完成支付



系统架构



博彩页面 vs 正常页面

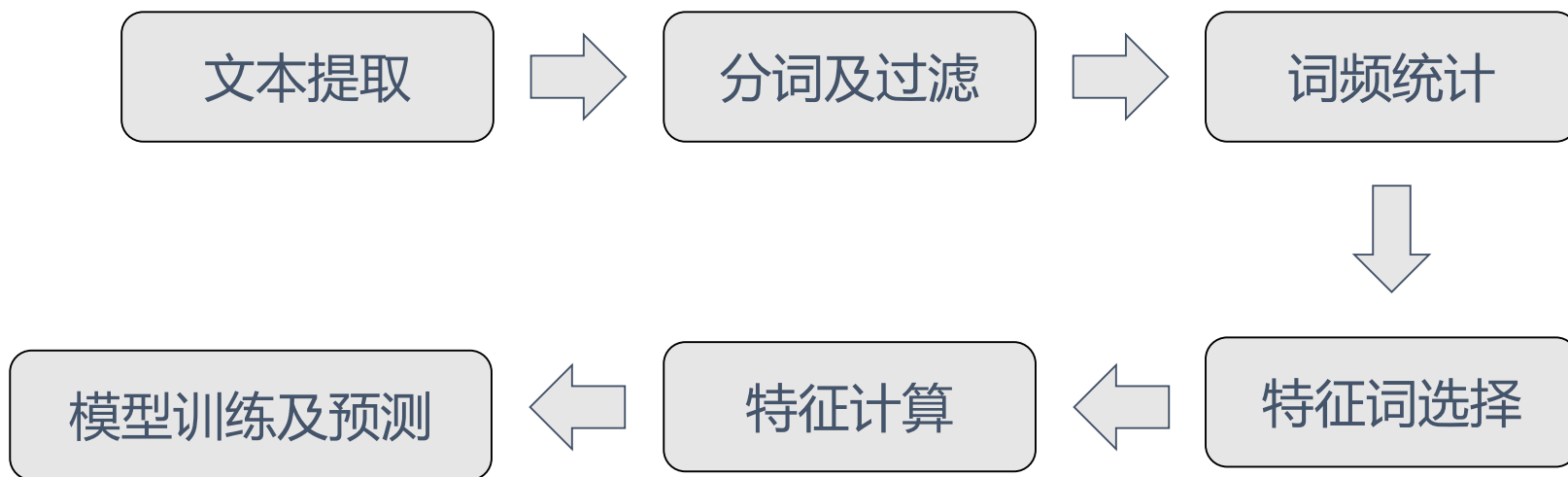


关键词：捕鱼大厅、太阳城、百家乐、老虎机.....



关键词：新闻、军事、体育、科技、电影....

检测模块



基于文本内容的博彩、色情页面检测系统

检测模块

• 数据预处理

文本提取



分词及过滤



词频统计

- 网页文本提取：提取title、meta、body等标签，去除style、script等不含语义信息的内容
- 过滤非常用字符：仅保留中文、英文及标点
- 使用Jieba对文本进行分词（使用已有黑词作为分词词库），并保留词性
- 停用词过滤：去除中英文停用词
- 词性过滤：仅保留动词、名词
- 统计每个文档中词j出现在title、keywords、description、body text中的个数

检测模块

• 特征选择

特征词选择



特征计算



模型训练

- 利用**文档词频**、**信息增益**、**卡方统计**等三类特征，确定筛选阈值
- 文档词频：每个词出现的文档数量
- 信息增益：某个词为分类能够提供的信息量
- 卡方统计：体现词与类别之间的独立性缺乏程度

- 综合三类特征计算方法，明确特征词
- 计算全部特征词的tfidf

- 设置正常、博彩、色情样本的比例为20:5:5，对比不同比例对训练效果的影响
- 对比不同算法检测效果，最终采用SVM算法，检测出准确率**99.93%**

目录

1. 背景介绍
2. 检测系统
3. 测量与分析
4. 校园网部署实践

博彩域名及推广渠道

	TLD&SLD	域名类别	数量	百分比
1	.com	Traditional TLD	600,959	62.08%
2	.cn	Traditional TLD	188,613	19.48%
3	.club	New gTLD	33,211	3.43%
4	.com.cn	SLD	30,950	3.20%
5	.top	New gTLD	28,414	2.94%
6	.net	Traditional TLD	22,726	2.34%
7	.cc	Traditional TLD	12,374	1.28%
8	.tw	Traditional TLD	5,955	0.61%
9	.vip	New gTLD	4,448	0.46%
10	.org	Traditional TLD	4,401	0.45%
总	-	-	932,051	96.29%

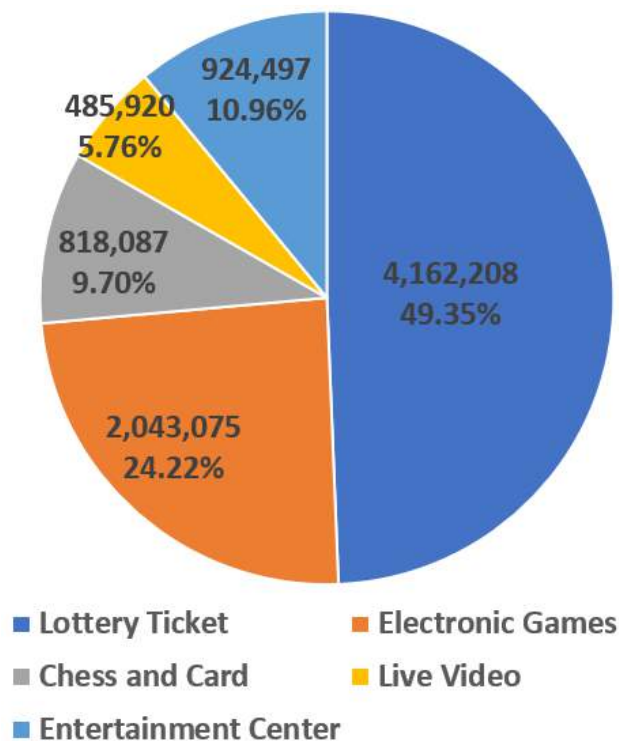
博彩站点IP地址主要位于美国

推广站点类型	出现数量	百分比
博彩	60,258,410	56.6%
色情	25,520,594	23.9%
Blackhat SEO	20,651,751	19.5%
总计	106,430,755	100%

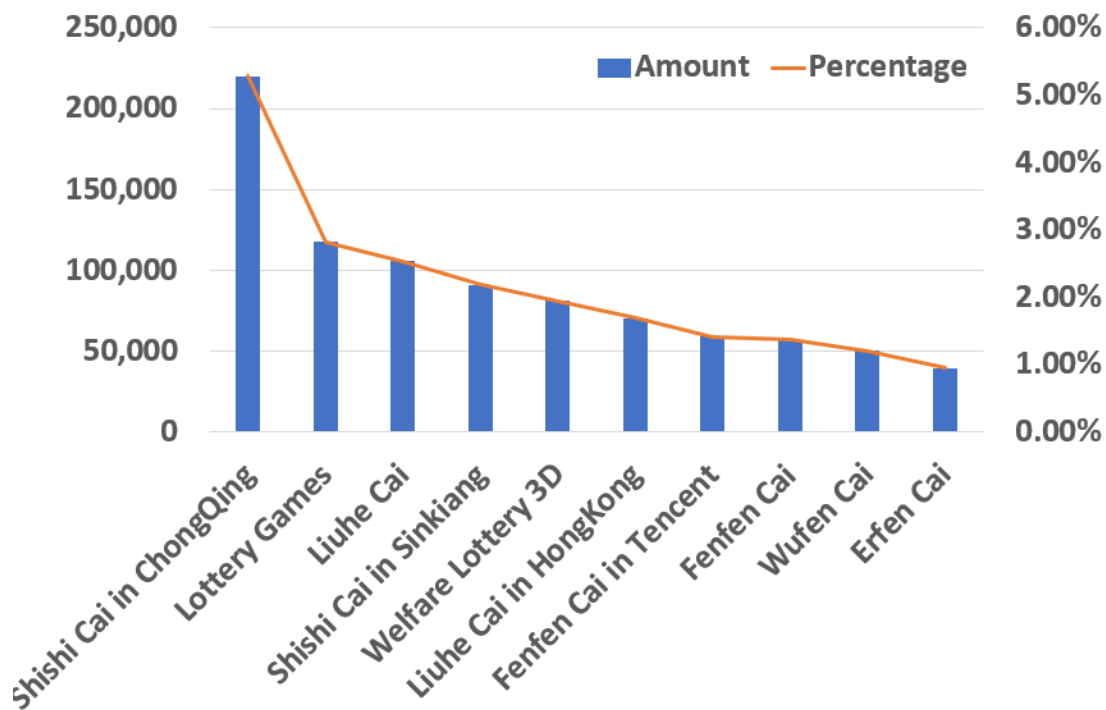


典型博彩导航

博彩游戏类型



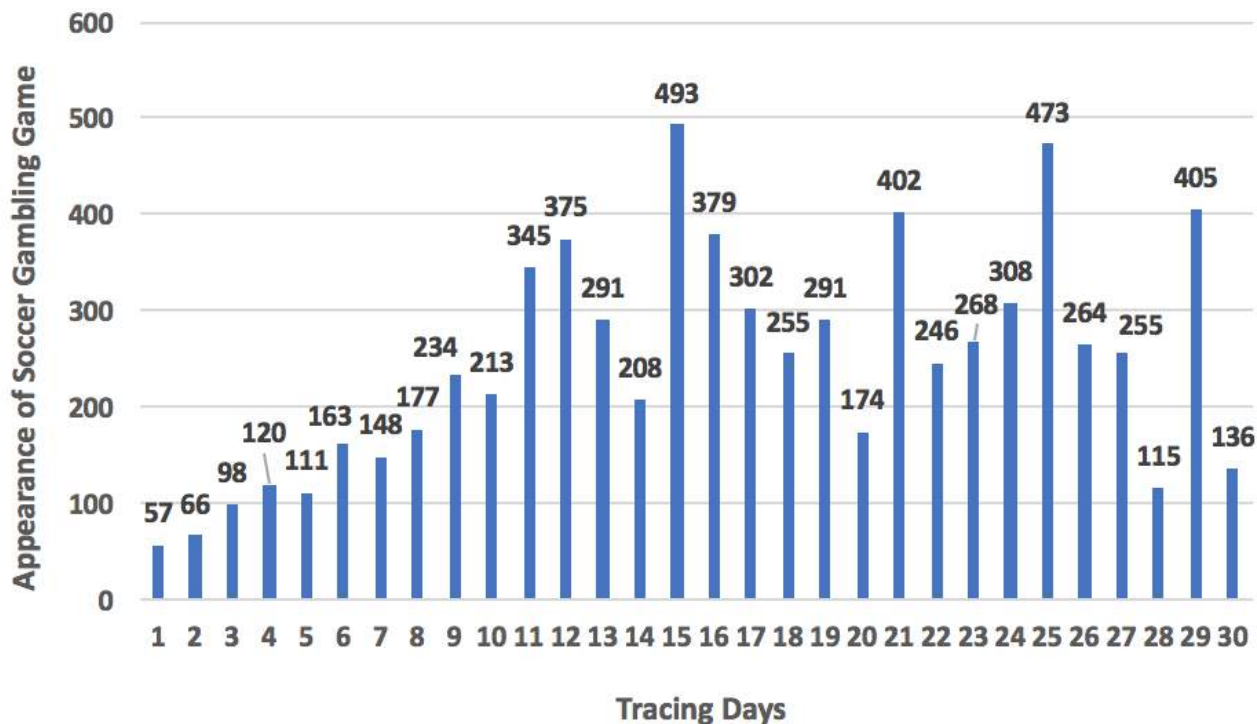
博彩项目类型



彩票类游戏类型

彩票类游戏在各类博彩网站中出现频次最高

博彩游戏变化跟踪



在世界杯期间，对博彩站点进行持续跟踪

在跟踪的**10000**个博彩站点中，**7372**个出现了世界杯相关的博彩内容

网络存储设施的滥用

存储类型	网站数量	URL数量	百分比
本地存储	440,253	8,702,216	45.48%
第三方图床	237,197	4,831,317	24.50%
公共服务器	215,278	4,517,975	22.24%
内容分发网络	75,226	1,375,212	7.78%
总计	967,954	19,426,720	100%

	域名	出现次数	URL数量
1	Weibo-hk.com	59,563	406,315
2	51yes.com	43,201	46,413
3	Alicdn.com	26,823	203,226
4	Clsj365.com	25,666	2,871,335
5	Sinaimg.cn	24,846	229,972

常见存储站点



<http://wx2.sinaimg.cn/large/0065w7B7gy1fi2xauuq47g30qo01omyp.gif>



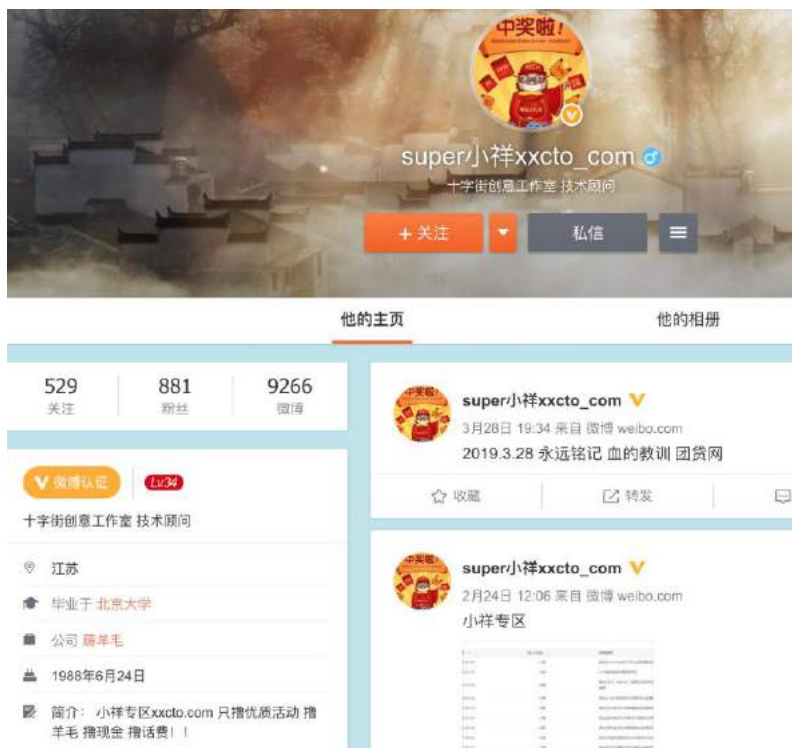
https://cbu01.alicdn.com/img/ibank/2018/783/382/9460283387_1746120392.jpg

CDN、公共服务器、第三方图床等存储工具被大规模滥用

基于新浪图床的博彩团伙识别



<http://wx2.sinaimg.cn/large/0065w7B7gy1fi2xauuq47g30qo01omyp.gif>



基于图床URL反查微博用户，定位犯罪团伙

从博彩行业出发，会牵出一系列地下黑色产业

博彩客服系统

- 客户留言页面

- 例: <https://www.78977c.com/pages/contact-us.html>

- 第三方客服应用

- 例: providesupport.com、meiqia.com

- 社交论坛滥用

- http://tieba.baidu.com/home/main?un=ManBetX_%E5%A9%B7%E5%A9%B7&fr=home&id=cd474d616e426574585fe5a9b7e5a9b74f93

博彩客服系统

- 客户留言页面

- 例：<https://www.78977c.com/pages/contact-us.html>

- 第三方客服应用

购买第三方客服应用

- 例：providesupport.com

- 社交论坛滥用

- <http://tieba.baidu.com/p/7727406616>
<http://tieba.baidu.com/p/7727406616?fr=home>
<http://tieba.baidu.com/p/7727406616?fr=home>



博彩客服系统

- 客户留言页面

- 例: <https://www.78977c.com/pages/contact-us.html>

- 第三方客服应用

- 例: providesupport.com

- **社交论坛滥用**

- <http://tieba.baidu.com/home/B7%E5%A9%B7&fr=home&id:7e5a9b74f93>



论坛用户主页

博彩客服系统

- 社交工具滥用：QQ、微博、微信
 - 752826666、 2835771727、 1654367017
- 邮箱客服
 - CEO@cais88.com、 ayxxzx@163.com
- 电话客服
 - tel:13594444243、 tel:400-889-8869

第三方客服统计

序号	e2LD	覆盖域名
1	Livechatvalue.com	11,633
2	live800.com	11,573
3	Providesupport.com	6,396
4	qq.com	6,395
5	Tencent://	6,271
6	Providesupport.net	3,870
7	Meiqia.com	2,459
8	Learnsaas.com	6,677
9	53kf.com	3,574
10	Duokebo.com	2,111
总计	-	131,928

常见第三方客服URL

	URL	覆盖 域名
1	https://messenger.providesupport.net/messenger/0n7w61u9pi8zo0uvfo1jtmraq.html	6,202
2	tencent://message/?uin=67393111&Menu=yes	6,202
3	https://chat.manbetx800.net/chat/chatClient/chatbox.jsp?companyID=666&configID=6	5,055
4	http://tieba.baidu.com/home/main?un=ManBetX_%E5%A9%B7%E5%A9%B7&fr=frs&d=cd474d616e426574585fe6c3e6c34f93&red_tag=o2127582729	5,054
5	http://tieba.baidu.com/home/main?id=d7c3e4b887e58d9a5fe4bd93e882b24e93?t=1532162862&fr=userbar&red_tag=a1194738600	5,011
6	https://vp8.livechatvalue.com/chat/chatClient/chatbox.jsp?companyID=80002422&configID=2826	3,108
7	http://wpa.qq.com/msggrd?v=3&uin=937382222&amp;site=qq&menu=yes	2,160

博彩客服联系方式提取

	URL	覆盖域名
1	https://messenger.providesupport.net/messenger/0n7w61u9pi8zo0uvfo1jttmraq.html	6,202
2	tencent://message/?uin=67393111&Menu=yes	6,202
3	https://chat.manbetx800.net/chat/chatClient/chatbox.jsp?companyID=666&conf	5,055
4	http://tieba.baidu.com/post/12127582729?fr=frs&d=cd47	5,054
5	http://tieba.baidu.com/post/3e882b24e93?fr=frs&t=153	5,011
6	https://vp8.livechat.com/800024	3,108
7	http://wpa.qq.com/msgrd?v=3&uin=67393111	2,160

vnsr 客服: 静静(67393111)

对方需要您回答一下验证问题:

问题1: 您的会员账号是多少

基于客服信息能够识别博彩团伙, 共提取
1,721个从业者QQ号

支付工具提取 (快速支付)



支付工具提取 (快速支付)



 扫一扫支付 手机也能支付, 输入支付网址: 29818pay.com—键入款, 立即到账!	 支付流程 输入并确认正确的会员账号> 输入存款额度>点击确认支付> 付款成功后1-10秒自动到账。	 存款范围 支付宝/微信存款金额范围为 1-5000元, 需交大额入款可 分多次存入或使用其它方式存款。
会员账号:	请填写太阳城会员账号	*必填
确认账号:	请确认会员账号是否正确, 否则无法充值	*必填
在线支付:	银联扫码  银联银联扫码  支付宝扫码  银联支付宝扫码	*必填
提示信息:	不同的支付方式, 对存款额度有不同的限制, 请依据规则存款	
确认额度:	请输入存款的数额	*必填

确认支付




支付工具提取 (快速支付)



 <p>扫一扫支付</p> <p>手机也能支付, 输入支付网址: 29818pay.com—输入款立即到账!</p>	 <p>支付流程</p> <p>输入并确认正确的会员账号> 输入存款额度>点击确认支付> 付款成功后1-10秒自动到账。</p>	 <p>存款范围</p> <p>支付宝/微信存款金额范围为 1-5000元, 需交大额入款可分多次存入或使用其它方式存款。</p>
会员账号:	请填写太阳城会员账户	*必填
确认账号:	请确认会员账户是否正确, 否则无法充值	*必填
在线支付:	 银联扫码  银联银联扫码  支付宝扫码  银联支付宝扫码	*必填
提示信息:	不同的支付方式, 对存款额度有不同的限制, 请依据规则存款	
确认额度:	请输入存款的存款金额	*必填

确认支付



 <p>扫一扫支付</p> <p>手机也能支付, 输入支付网址: 29818pay.com—输入款立即到账!</p>	 <p>支付流程</p> <p>输入并确认正确的会员账号> 输入存款额度>点击确认支付> 付款成功后1-10秒自动到账。</p>	 <p>存款范围</p> <p>支付宝/微信存款金额范围为 1-5000元, 需交大额入款可分多次存入或使用其它方式存款。</p>
会员账号:	ghow123	*必填
确认账号:	ghow123	*必填
在线支付:	 银联扫码  银联银联扫码  支付宝扫码  银联支付宝扫码	*必填
提示信息:	本支付方式最低金额 100	
确认额度:	123	*必填

确认支付

支付工具提取 (快速支付)



	扫一扫支付 手机也能支付, 输入支付网址: 29818pay.com—输入款立即到账!		支付流程 输入并确认正确的会员账号>输入存款额度>点击确认支付>付款成功后1-10秒自动到账。		存款范围 支付宝/微信存款金额范围为1-5000元, 需交大额入款可分多次存入或使用其它方式存款。
会员账号:	请填写太阳城会员账号				*必填
确认账号:	请确认会员账号是否正确, 否则无法充值				*必填
在线支付:	<input checked="" type="radio"/> 银联扫码 <input type="radio"/> 银联银联扫码 <input type="radio"/> 支付宝扫码 <input type="radio"/> 银联支付宝扫码				*必填
提示信息:	不同的支付方式, 对存款额度有不同的限制, 请依据规则存款				
确认额度:	请填写实际的存款金额				*必填

确认支付



	扫一扫支付 手机也能支付, 输入支付网址: 29818pay.com—输入款立即到账!		支付流程 输入并确认正确的会员账号>输入存款额度>点击确认支付>付款成功后1-10秒自动到账。		存款范围 支付宝/微信存款金额范围为1-5000元, 需交大额入款可分多次存入或使用其它方式存款。
会员账号:	ghcw123				*必填
确认账号:	ghcw123				*必填
在线支付:	<input checked="" type="radio"/> 银联扫码 <input type="radio"/> 银联银联扫码 <input type="radio"/> 支付宝扫码 <input type="radio"/> 银联支付宝扫码				*必填
提示信息:	本支付方式最低金额 100				
确认额度:	123				*必填

确认支付



支付工具分析

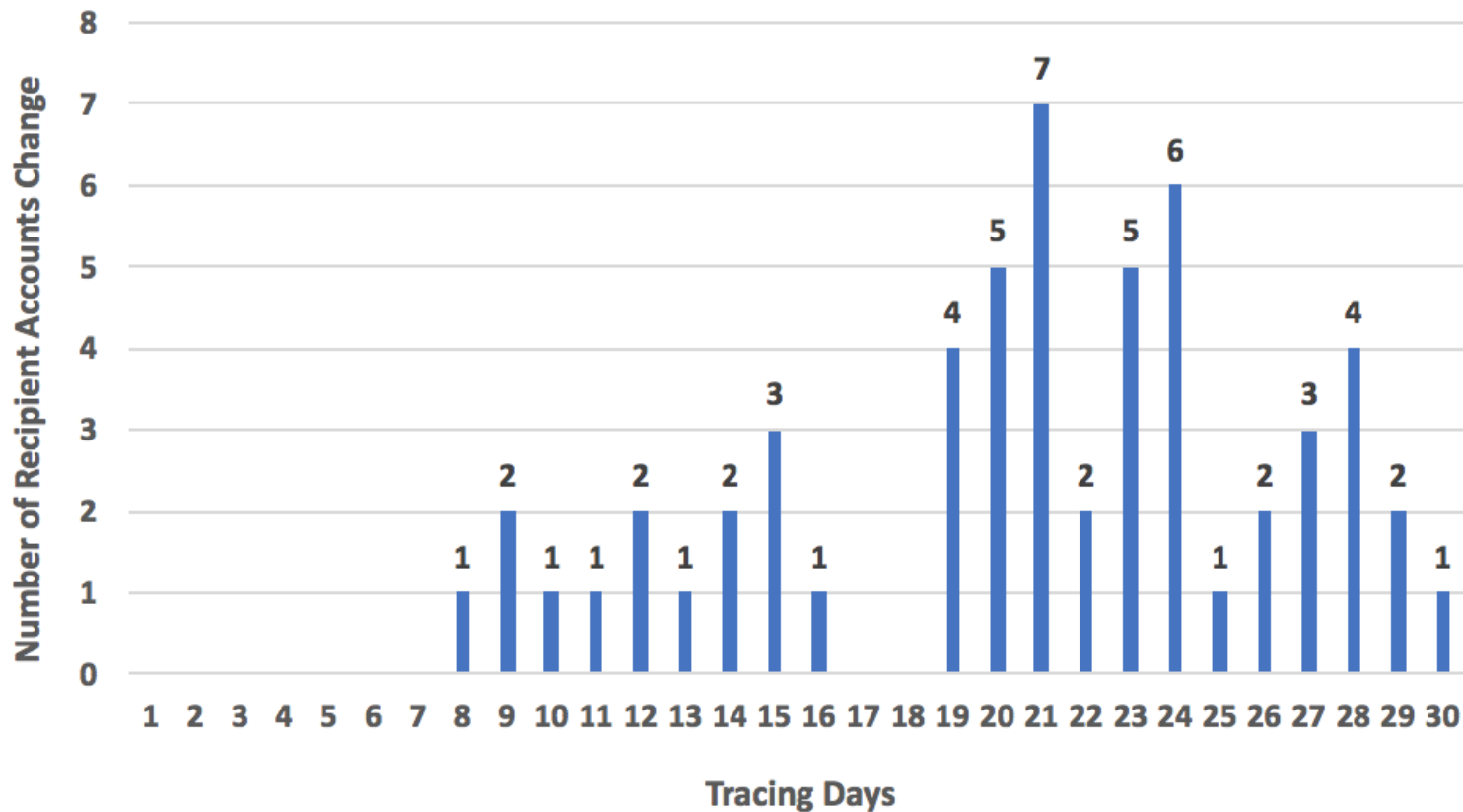
会员账号:	<input type="text" value="请填写会员账号"/>
确认账号:	<input type="text" value="请确认会员账号是否正确, 否则无法充值"/>
支付类型:	<input checked="" type="radio"/>  支付宝支付 <input type="radio"/>  京东支付 <input type="radio"/>  银联扫码

博彩网站会同时使用**多种**支付工具

通过对1,140个支付站点进行分析, 支付宝在接近**90%**的博彩站点中被使用

序号	支付类型	数量	百分比
1	支付宝	1,019	89.4%
2	微信支付	584	51.2%
3	银联支付	565	49.6%
4	QQ支付	414	36.3%
5	网银支付	389	34.1%
6	京东金融	247	21.7%

收款账户变化



2018.8.1-2018.8.31, 对**100**个博彩网站进行监控, 其中**56**个更新了收款账户, **43**个变换了支付工具

博彩团伙识别

- 基于公共的客服、支付链接，识别出**6,581**个不同的博彩站点群

序号	域名数量	模板数量	百分比
1	19,936	79	2.06%
2	11,325	53	1.17%
3	9,311	30	0.96%
4	7,689	127	0.79%
5	6,986	74	0.72%
6	6,474	35	0.67%
7	6,264	99	0.65%
8	5,853	33	0.60%
9	5,447	61	0.56%
10	5,312	42	0.55%
总计	84,597	-	8.74%



目录

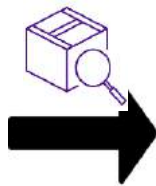
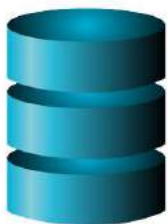
1. 背景介绍
2. 检测系统
3. 测量与分析
4. 校园网部署实践

校园网实践

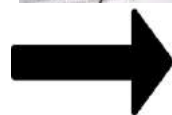
每日访问数据

four-threads.com
fpb.org
fingerlickingdutch.com
fyl.bergvalls.com
homefinance.businessbureau.co.za
goodbulldesigns.com
imap.photosintheloft.com
m.fuguodu520.cc
www.bhmy518.com
gardainsurance.com
www.betterbodies.com
fornewgames.com
giftprint.com.ec
florescatarroja.com
histamin.sk
jonsurfer.com
foundationbrands.org
www.hotel-cosmos-bg.com
flowerthinking.com
grubelasgallivanting.com
www.meingottwalther.de
www.yihan.it
www.zsdlw.com
formacaotecnica.com.br
jumpingfeet.nl
jong-online.info
gracielevelinglift.com
cdn1.69teensex.com
gardensportingcenter.it
imap.acappellahub.com
fitfoody.nl
froma2z.org
frobabies.com

数据库查询模块



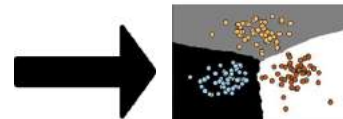
爬虫子系统 Python+ Selenium+ Firefox



HTML+访问截图



检测子系统



检测结果展示



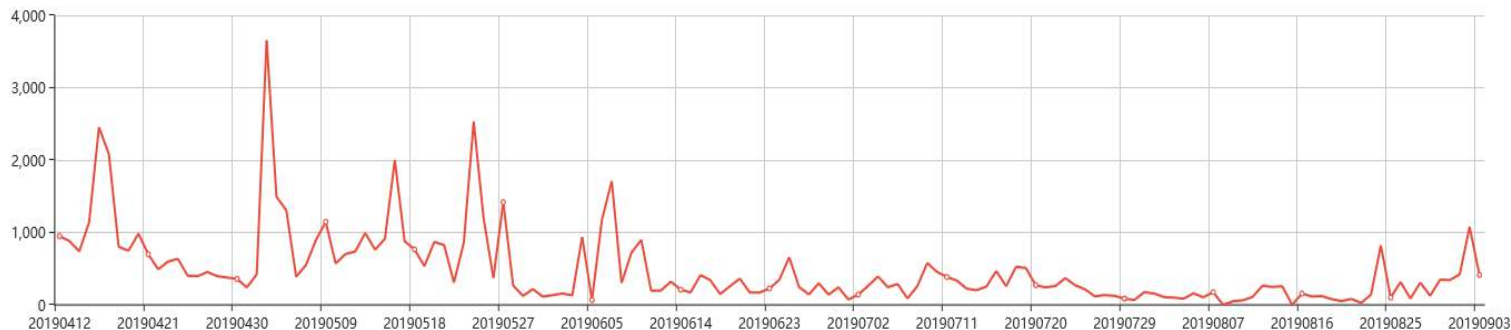
系统运行情况

博彩域名按天统计与每天详情

域名安全检测

详细

○ 博彩域名数量



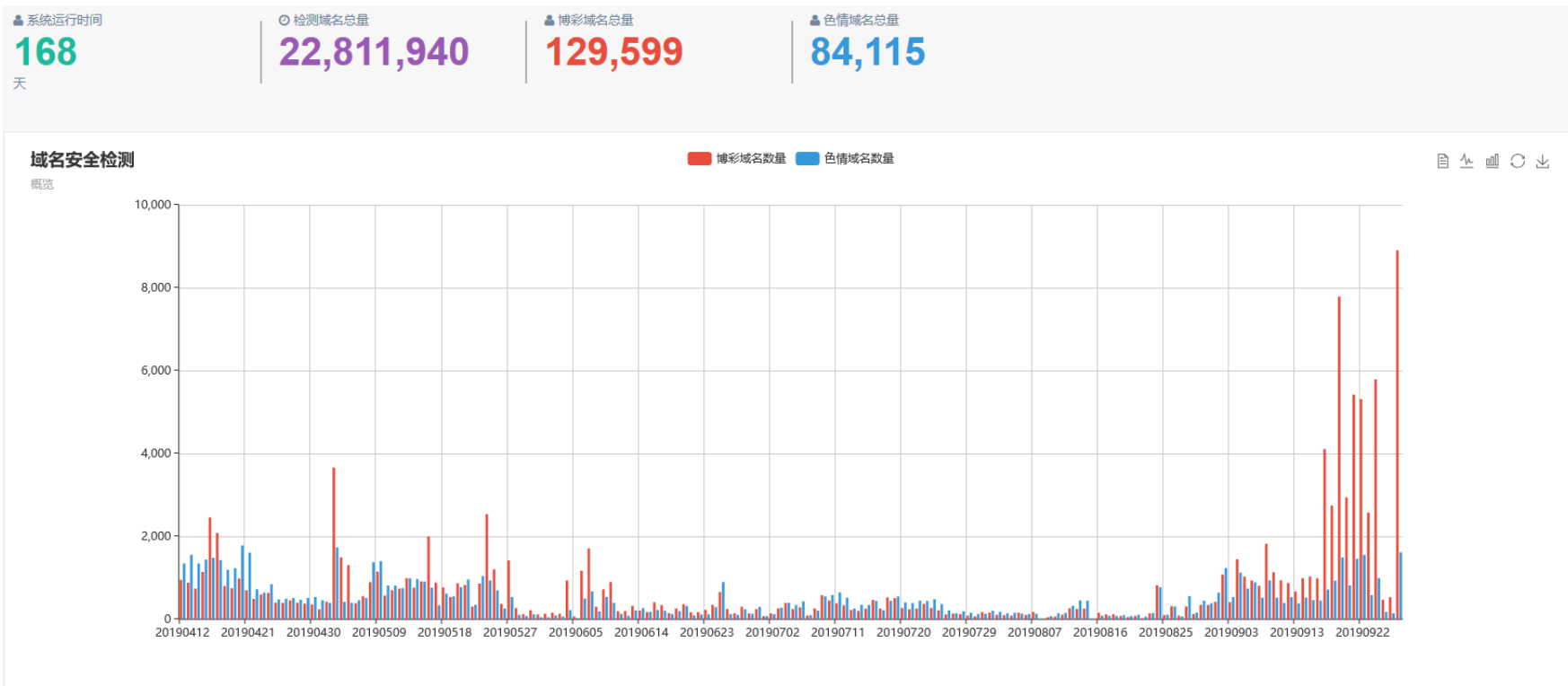
20190412

共有 949 个检测结果

1 0024aagg.com	2 00882410.com	3 0091331.com	4 0125345.com	5 0393918.com	6 046688.com	7 0679app.com	8 0806p.cc	9 08820088.com	10 1024.hegongchar	11 103111.com	12 111239.com

系统检测成果

- 部署**168**天来，共检测域名**22,811,940**个，检测出包含博彩内容的域名**129,599**个，包含色情内容的域名**84,115**。



感谢聆听

Q&A