【反欺诈专栏】互联网黑产剖析——代理和匿名

同小盾 / 2017-07-13 03:26:00 / 浏览数 5918 技术文章 技术文章 顶(0) 踩(0)

技术原理

定义 ：代理，英文:Proxy，在这里特指网络代理。
代理是一种网络服务，用于为客户端和服务端提供非直接的链接。

一般的，根据代理所使用的协议，可以对代理做如下分类:

代理类型	代理协议	用途
HTTP代理	HTTP	转发HTTP请求，一般是web访问
HTTPS代理	HTTPS	转发HTTPS请求，一般HTTPS代理，也能够进行HTTP代理
Socks代理	Socks4/5	可以转发任意类型的请求
VPN	PPTP/L2TP/OpenVPN/SSL VPN/IPSec VPN	可以转发任意类型的请求
Tor	Tor	可以转发任意类型的请求
ShadowSocks	私有协议	主要用于转发HTTP/HTTPS请求
RTSP	Real流媒体	一般用在视频缓存的场景中
POP3/SMTP	POP3/SMTP	一般用在邮件转发和缓存
FTP代理	FTP	FTP转发、跳板、缓存
TURN/STUN	TRUN/STUN	这个协议本来是在电话会议的一些场景中使用的 也可以被用来做代理
…	…	…

由于近年来"网络安全"和"隐私保护"的问题成为了热点，越来越多的人开始追求互联网上的匿名。因此，代理的匿名程度，也成为区分代理的一个重要标志。

根据代理的匿名程度，也有做如下区分：

代理类型	特征描述
高匿代理	高匿代理本质上只是单纯的转发TCP数据包，不会做对数据包做任何修改； 服务端记录的来源IP，是代理服务器的IP； 高匿代理的流量无法追查最终来源。
普匿代理	普匿代理一般会对数据包做一定修改，偶尔会留下一些可供追查的真实信息； 比如：X-Forwarded-For。
透明代理	透明代理不但会对数据包进行修改，还会传递真实的用户IP，一般用在硬件的防火墙上。
间谍代理	用于进行记录、监控和研究而部署的代理。

比如，目前很多人在用的SS，就是一种高匿代理。当我使用一个SS代理去访问互联网时，平台或网站获取到的IP地址，就不再是我的真实IP，并且整个访问过程中，都不会

部署一个代理的成本其实并不高。相信很多人都拥有自己专属的代理服务器，用来访问一些国外的“资源”。

Nginx中开启代理只需要做如下配置：

```
server {
    listen            8080;
    location / {
        proxy_pass $scheme://$host$request_uri;
        proxy_set_header Host $http_host;
        ......
    }
    ......
}
```

建立Socks代理更为简单：

```
$ssh -D 1080 user@163.44.154.189
```

用户只需要使用简单的代理工具，或者在浏览器中设置代理，就可以使用。

总的来说，代理，提供了一种低成本的隐藏用户真实信息的途径和方式。

保护隐私的同时，代理也为各种风险行为提供了良好的庇护。就像Tor(洋葱路由)一样，最早期Tor由美国海军研究实验室赞助开发，用来为特工提供一种高度隐蔽和高度匿名

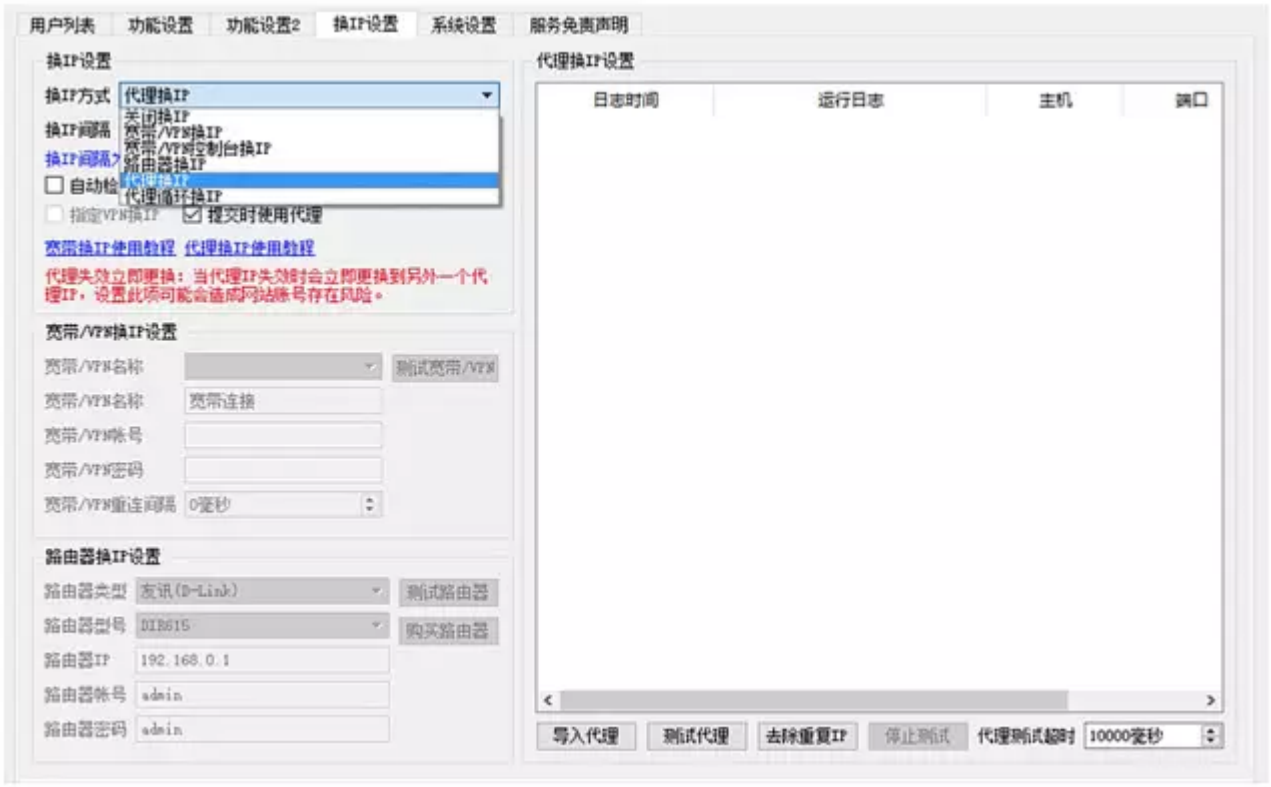
代理在欺诈行为中的使用场景

TCP/IP，是互联网的基础，没有TCP/IP，可能就没有今天的互联网了。

TCP/IP协议，赋予了每个上网的人都会拥有一个IP。因此，IP地址也成为了风控中非常重要的一环。

一般的，我们的风控策略中，会设置很多频度方面的策略。比如：同一个IP上一个小时内登陆次数超过50次。

上一期的分享中，提到了某工作室的一个自动注册机。其中包含了一个更换IP的功能，可以通过设定代理IP或VPN来实现。



那么，对于一般的风控规则：单个IP在一小时内注册/登陆/交易次数超过N次，就可以使用代理来进行规避。

根据目前了解到的情报，代理已经成为欺诈分子的必备工具，每完成一次欺诈活动，就会更换一个新的代理。

比如，从注册账号、登陆账号、领取优惠券到使用优惠券下单都使用同一个代理，下单完毕，就标志这次欺诈活动结束，然后使用一个新的代理继续进行下一次，绕开IP地址

除了规避简单的频次限制规则，代理的另外一个用途在于隐藏自己的真实位置。

IP地址的划分和使用是有迹可循的，根据用户访问的IP，可以判断该用户所处的大概位置，一般可以精确到城市级别。网站或平台会根据用户的IP，解析当前用户的位置，通

比如，当你从一个长期生活的城市，忽然去到另外一个城市的时候，很多APP都会有异地登陆的提醒。

代理，提供了一种规避位置判断的手段。

此前我们遇到的一个案例中，欺诈分子通过代理来实施盗卡。我们通过设备指纹的定位信息，确定欺诈分子身处南京，但是使用了一个上海的代理IP，盗用了一张上海的银行

随着移动技术的发展，产生出了很多其他的定位手段，正在逐步替换掉IP地址的位置解析结果。但是依然存在很多场景，我们不得不使用IP来对用户的位置进行判断。

此前，我们曾实验性地部署了一台代理服务器，详细地记录通过这台代理发生的各种请求。事后我们根据服务器上的日志来分析用户们使用这个代理做了些什么。

大致得出了以下一些统计数据(统计时间片为一周)：

行为	特征	量级
垃圾信息	频繁访问发帖/回复/弹幕/私信的接口； 发送的内容主体都是广告或违禁词。	660万次
撞库攻击	频繁访问网站的登录接口； 每次请求中的用户名和密码字段都不相同。	250万次
垃圾注册	频繁访问网站的注册接口； 每次请求中的用户名和密码都不同，而且都很随机。	45万次

上面的这些风险行为，累计涉及了300多家网站和平台。

你可能会注意到，前面介绍的注册工具中，可以批量导入代理。而互联网上，代理的资源非常丰富，欺诈分子轻而易举就能获取到上千甚至上万个代理IP。欺诈份子刚好选中

想必，你也能感觉到这其中涌动的暗流了吧？

代理总数: 17424188个 1分钟前更新　　10分钟内可用: 6071个 1分钟前更新

免费高速HTTP代理IP列表（2017-02-27）

IP	PORT	匿名度	类型	get/post支持	位置	响应速度	最后验证时间
122.96.59.99	83	匿名	HTTP	GET, POST	江苏省南京市 联通	3秒	2分钟前
121.232.145.20	9000	高匿名	HTTP	GET, POST	中国 江苏省 镇江市 电信	3秒	46秒前
121.232.146.65	9000	高匿名	HTTP	GET, POST	中国 江苏省 镇江市 电信	0.5秒	3分钟前
122.5.129.218	808	高匿名	HTTP, HTTPS	GET, POST	中国 山东省 德州市 电信	1秒	7分钟前
116.9.74.226	8998	高匿名	HTTP, HTTPS	GET, POST	中国 广西壮族自治区 来宾市 电信	2秒	9分钟前
223.72.128.22	8000	高匿名	HTTP, HTTPS	GET, POST	中国 北京市 北京市 移动	1秒	12分钟前
27.38.97.71	9797	透明	HTTP	GET, POST	中国 广东省 深圳市 联通	3秒	15分钟前
116.52.17.141	8998	高匿名	HTTP, HTTPS	GET, POST	中国 云南省 昆明市 电信	3秒	18分钟前
121.232.147.176	9000	高匿名	HTTP	GET, POST	中国 江苏省 镇江市 电信	2秒	21分钟前
101.200.38.16	80	高匿名	HTTP	GET, POST	中国 北京市 北京市 阿里云	3秒	24分钟前

代理检测和规避检测

代理，是一种非常廉价的资源。因为互联网是开放的，一个开放的代理，只要知道服务器的IP，
代理协议和代理端口，任何人都可以使用它。从而产生了很多提供代理检测服务的平台，他们会对整个互联网进行代理扫描，把可用的代理IP
记录下来，提供给爬虫或其他工具使用。

开放性的代理虽然很容易获取，但是使用的人非常多，鱼龙混杂，爬虫、广告机、注册机，甚至某些网络攻击也会使用这些代理。于是，安全产商和一些软件联盟，开始了对

代理运行的时间越长，使用的人越多，经过这个代理产生的风险行为也越多，然后被各大平台拉入黑名单。比如，Wordpress根据全球范围内的WP博客接收到的垃圾评论及
IP黑名单，其中绝大部分都是代理。 著名的开源入侵检测系统Snort，付费版规则集中屏蔽了上百万个IP，大部分也都是代理IP。
这些由安全产商、软件联盟整理的黑名单数据，目前已经收录到同盾IP画像第三方风险证据库中进行维护。

虽然无法透过代理，去追查幕后的欺诈分子，但是如果能够有效的识别代理，就可以有效识别出大批的风险行为。为此，同盾建立了一套高性能的代理检测系统，用于对全网

文章开头，我们提到了代理可以根据协议进行分类，对代理进行检测的时候，一般也按照代理的协议进行。在反欺诈领域，绝大多数欺诈行为都通过HTTP协议进行，所以，
Socks和VPN。

一种监测HTTP代理的简单方式如下:

```
$curl -i -k -s -H 'Host:www.target.com' -XGET 'http://27.38.97.71:8000/'
```

然后判断返回的页面，是否和http://www.target.com/页面内容一致，就可以确定这个IP是否是代理。

其他的代理协议，检测起来要更复杂一些，这里就不再逐一列举了。代理检测虽然原理简单，一套建立够满足业务需要的代理检测系统，还是需要投入很大的成本，主要是硬

具备足够实力的互联网公司，都会尝试建立自己的代理检测系统。长此以往，欺诈分子也意识到，代理一旦被发现，就不能再被使用。于是产生了很多用于规避代理检测的方

一场旷日持久的对抗，由此展开。我们来细数一下，常见的规避手段都有哪些。

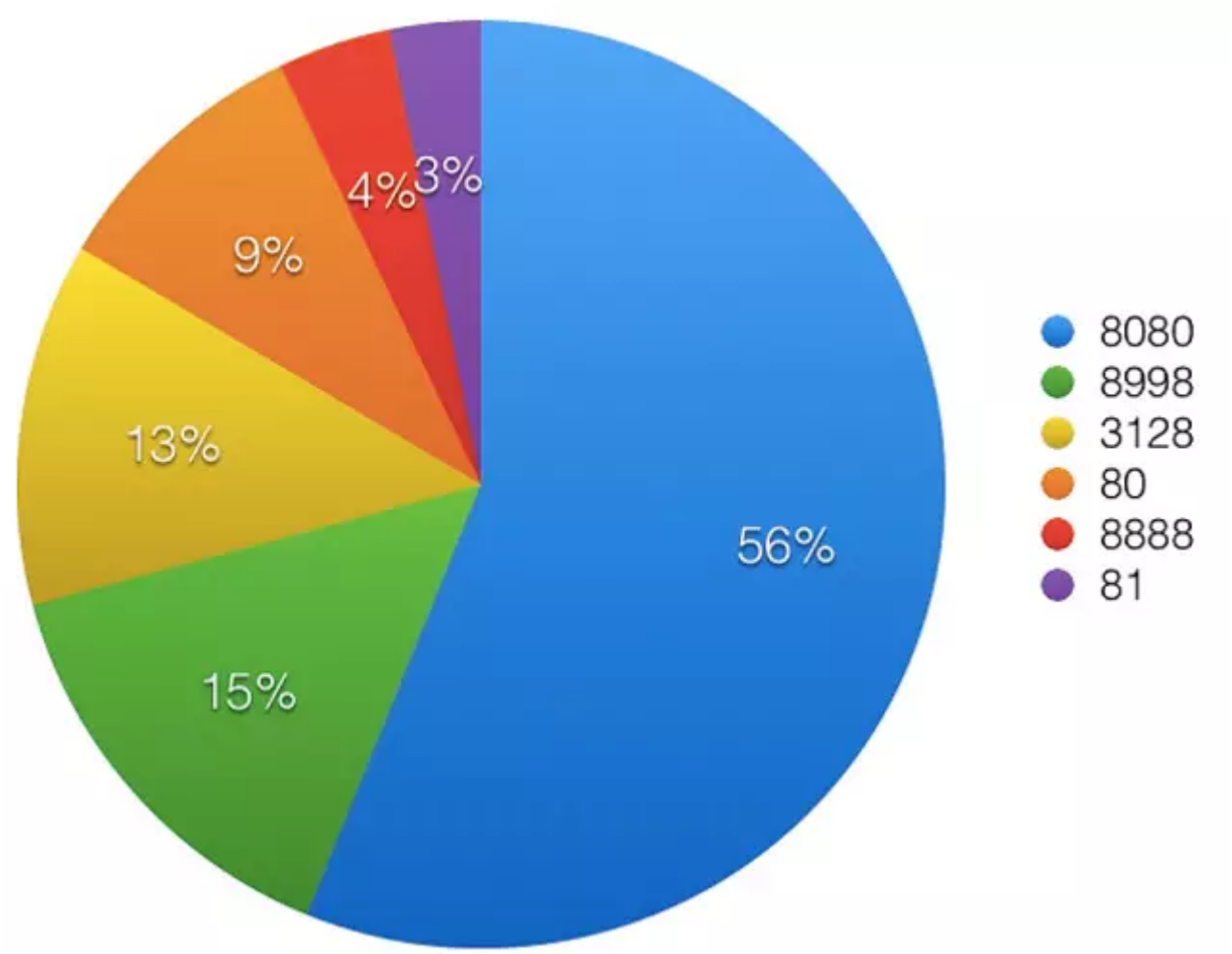
使用非常规端口

每个IP上可以使用的端口数量是有限的，从1~65535，不可能对所有端口都进行检测，这个代价是无法承受的。

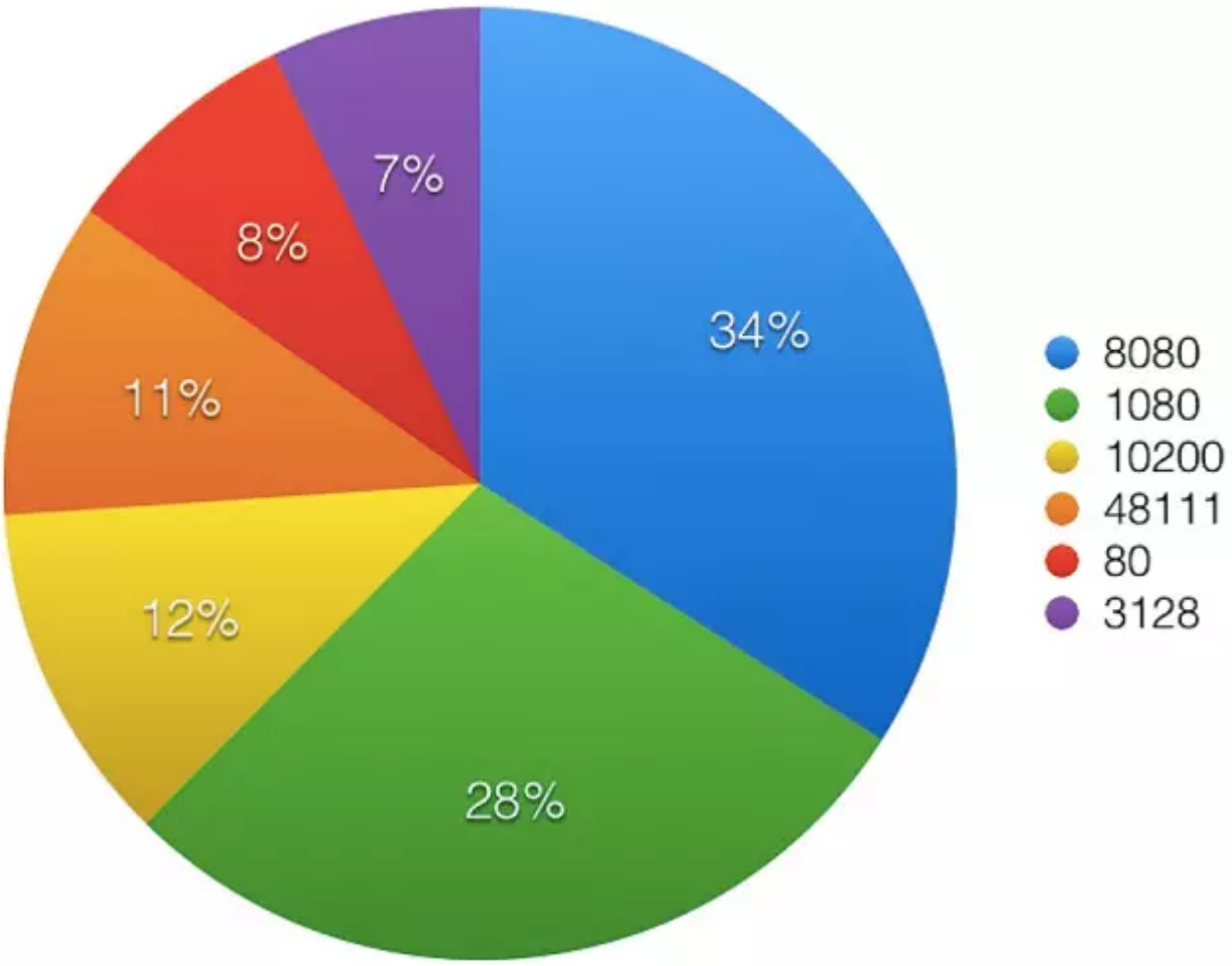
一般的代理检测，会针对特定的端口进行。

我们此前对全网的代理服务器做过分析，统计各种代理协议和代理端口的对应关系，部分如下：

HTTP代理端口分布：



Socks代理端口分布：



如果代理IP使用的端口，不在常规检测的列表之内，或者使用一些非常特别的端口，就可以避免被检测。

我们曾经遇到过一个代理，运行端口为1，测试的时候还以为我们程序出现了问题，反复确认了好几遍。

根据我们长期的全网监测，统计全网各类代理最为集中的端口。结果表明，只要对其中20个端口进行代理检测，就能够覆盖全网超过70%的代理，如果扫描端口增加到40个

同盾会定期统计代理端口的分布情况，并更新扫描端口列表，保证代理的检出率。

用后即毁

代理有两个非常重要的指标："响应延迟"和"生存期"。

很多提供代理检测服务的平台，都会把代理节点的延迟作为一个重要参数，代理的延迟越低，使用的人就越多。

另一方面，绝大部分代理并不会长期运行，根据我们的统计，80%的代理存活时间只能以分钟计算。如果一个代理的存活时间非常短，就可以完全避免被代理检测发现。

短效优质代理API包天	
API链接提取	✔
每日IP流水量	5万左右
有效率	95-100%
私密性	万人骑
平均响应时间	0.03秒
一次可提取IP量	60-80
每个代理并发限制	10个
每个代理存活期限	2分钟
全部高匿级别	✔

短效优质代理API包月	
API链接提取	✔
每日IP流水量	5万左右
有效率	95-100%
私密性	万人骑
平均响应时间	0.03秒
一次可提取IP量	60-80
每个代理并发限制	10个
每个代理存活期限	2分钟
全部高匿级别	✔

短效优质代理API包半年	
API链接提取	✔
每日IP流水量	5万左右
有效率	95-100%
私密性	万人骑
平均响应时间	0.03秒
一次可提取IP量	60-80
每个代理并发限制	10个
每个代理存活期限	2分钟
全部高匿级别	✔

短效优质代理API包年	
API链接提取	✔
每日IP流水量	5万左右
有效率	95-100%
私密性	万人骑
平均响应时间	0.03秒
一次可提取IP量	60-80
每个代理并发限制	10个
每个代理存活期限	2分钟
全部高匿级别	✔

上图是国内一个比较大的代理服务平台，他们提供的短效代理，平均存活时间只有2分钟。

是否可以提升代理扫描器的性能，来满足分钟级别的监控呢？

如果有足够的软硬件资源和带宽资源，理论上是可行的。但是大规模的端口扫描，本身并不是一件好事。大量的数据包发送，有可能会导致运营商层面的设施故障、网络拥堵

那么，在现有的条件下，如何对这类代理进行防控呢？

部署代理虽然很简单，但并不是任何地方都可以部署。首先需要具备独立的公网IP，其次是有稳定的线路，保证虚构的上下行带宽，这些条件限制了代理只能在数据中心中部

具体如何识别一个IP是否来自于某个数据中心，我们将在下一篇文章中进行介绍。

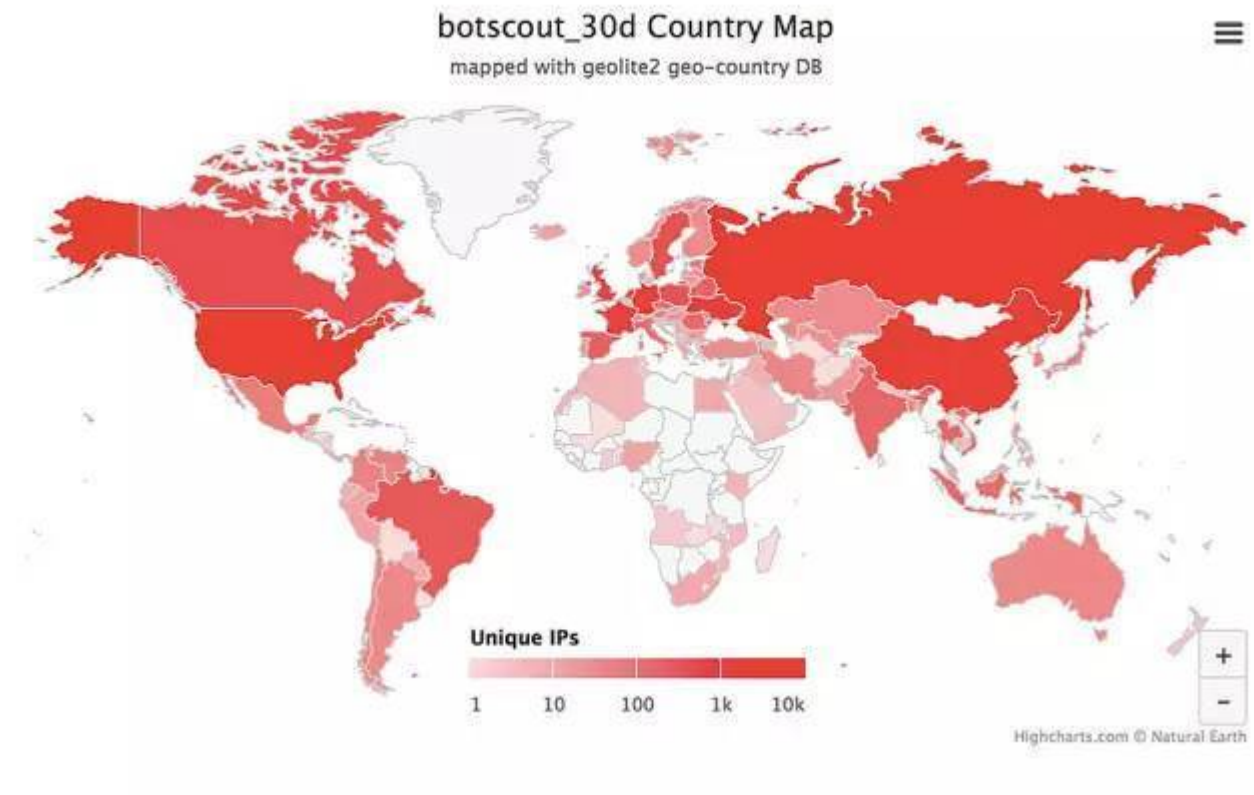
僵尸网络

这个词业内人士应该并不陌生。僵尸网络是通过各种远程控制程序组建起来的庞大网络，通过C&C服务器，向各个僵尸节点下发指令，进行各种网络攻击。

某些远程控制程序中提供简单的代理功能，欺诈分析可以通过这些僵尸节点来发起欺诈活动。
僵尸节点又称为“肉鸡”，虽然并不像“四要素”那样炙手可热，但是销路一直很不错。黑产会收集大量的肉鸡节点，用于发起DDOS攻击。

僵尸网络的分析和研究，是全社会面临的一个严峻的问题。国内外众多安全公司投入了巨大的人力和物力来对僵尸网络进行监控，有的安全公司也会公开自己长期监控的僵尸

国内有不少杰出的安全公司在这个领域有着深入的研究，在与之深度合作下，同盾也能够在反欺诈场景中，对僵尸节点进行有效的识别。



（上图是2017年3月15日，僵尸节点在全球范围内的分布情况）

结语

判断IP是否有风险，有很多种途径，代理检测只是其中一种，和虚假号码一样，仅仅是同盾众多风控手段中的一个环节。并不能单纯的因为一个IP被判断为代理，就直接封杀

为此，我们建立了同盾IP画像，尽可能多地提供关于IP的所有信息，在风险决策中进行综合评定。

关于同盾IP画像的详细内容，将在下一篇文章中进行介绍，尽请期待。

上一篇：【反欺诈专栏】互联网黑产剖析——虚假号码 下一篇：【反欺诈专栏】关于IP，这里有你想...
1. 3 条回复


泉泉圈圈 2017-08-17 03:01:38

死等

0 回复Ta


hades 2017-08-17 03:35:48

更新了~

0 回复Ta


ih0cker 2017-09-04 02:25:25

厉害了，学习了

0 回复Ta

登录 后跟帖

先知社区

现在登录

热门节点

技术文章

社区小黑板

目录