chliny

Free as Freedom

检查IP是否可用的python脚本

明显的，本文所说的检查IP可用是指墙内可用的google，twitter，facebook等网站的IP。本文讲的是我折腾脚本查找可用IP的经验，脚本已经扔在github上，欢迎fork && pr。脚本只能在python3，Linux，root权限下可用，这也是本文描述中默认的环境。

在开始检查IP是否可用之前，我们得先有找到IP。google有公开其持有的IP段。详见google帮助。但这个IP列表就很长了，我的做法是偷懒，github上跟踪数个hosts源，把历史hosts文件都取出来，再做检查。

历史hosts文件还能使用是基于这么一个假设：GFW出于性能或者其他什么原因，会随机释放一些之前被block的IP，另外GFW存在地域和运营商的差别，对别人不可用的IP在我这可能还是可以用的。在我的实践中，这个假设确实是成立的。

进入正题，首先是跟目标IP建立一个socket连接，类似于telnet。目标端口首先是80，然后443。对目标IPconnect之后，还需要再send一个随机字符串，没抛出异常才是正常可用的。因为存在一些IP，是能建立socket的，只是建立完会马上断掉。然后是443，如果目标IP是支持HTTPS的，只是被GFW了，那么这里一般是抛出socket.timeout的异常，少数会是socket.error的，而如果目标IP不支持HTTPS，会是普通的异常，通过这个可以避免误伤只有80的IP。

然后是对支持HTTPS的IP进行证书的检查。现在网上有一些hosts是用一个google的IP来通吃google的所有域名，但因为各个IP上证书存在差异，而且大部分google的客户端是不支持SNI的，所以像googlepais.com得用和google.com不一样的IP。但是我不知道怎么直接用python去取到证书的DNS里配置的域名，但我有另外一个简单的办法，用requests，带验证ssl地去打开https://IP，这时候会抛出requests.exception.SSLError的异常，而在异常的信息里就有IP证书支持的域名。

也有一些域名，虽然支持https，但证书不对也还能用，这种我也无法判断，只能在配置文件里加上一个白名单，好在这种情况不多，目前只发现包括ingress在用的几个appspot.com域名。

再然后是打开页面测试。这里因为有一些IP虽然端口证书都正常，但真正用到hosts上，打开域名是404，或者502，还有可能是204没有内容。最常见的就是google北京的203.208开关的那一批IP了。
打开测试的方法是用urllib.request.urlopen，然后重写HTTPHandler和HTTPSHandler，主要是重写DNS，将域名解析到想要测试的IP上。

这里打开测试又有一个另外的问题是一些CDN类型的域名，本来就是无法直接打开的，这个我也没有很好的办法，只有在配置文件里加一个配置节，将不需要打开测试的域名，比如*client*, *static*这类的写入打开测试的黑名单。好在使用多次之后，这个黑名单也比较稳定不经常要调整了。

我还有遇到一个奇葩的是mtalk.google.com这个域名，用于GMS推送用的，即使IP证书使用的正常也不能正常使用，目前发现只有188结尾的googleIP能用于这个域名。。

一个IP通过上面的测试就基本能用了，最后就是用ping结果来排序，取延时的丢包最小的用。这里为了优雅一点，我没有调用shell使用系统ping，而是用了python-ping这个模块，原生python操作raw ICMP包，这也是脚本需要root权限的原因。

以上～ 后面有新发现再补充。

This entry was posted in 学习 and tagged hosts, python on 三月 15, 2015 [http://chliny.me/?p=104204] .