(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108566380 A
(43)申请公布日 2018.09.21

(21)申请号 201810212973.X

(22)申请日 2018.03.15

(71)申请人 国家计算机网络与信息安全管理中心四川分中心
地址 610071 四川省成都市青羊区文庙后街80号

(72)发明人 昝家玮 朱魏魏 张旭 李文佑

(74)专利代理机构 成都信博专利代理有限责任公司 51200
代理人 张辉

(51)Int.Cl.
*H04L 29/06*(2006.01)
*H04L 29/08*(2006.01)

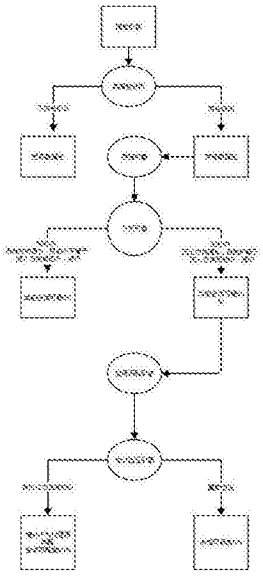
权利要求书1页 说明书5页 附图6页

(54)发明名称

一种代理上网行为识别与检测方法

(57)摘要

本发明公开了一种代理上网行为识别与检测方法，包括：初步处理网络中数据，在数据流出时，记录目的地址不在境内的IP，在数据流入时，记录源地址不在境内的IP；将数据流出时和数据流入时记录的IP保存在存储引擎中；储存引擎中的IP数据再提供给分析引擎，供分析引擎进行分析；将标识为可疑访问会话的数据流入方向的源地址加入到应用识别引擎中去进行应用识别；调用协议验证引擎进行协议验证，若验证出有协议为伪装协议，则判定境外IP为代理服务器访问会话或者判定境外IP为代理服务器访问会话的可疑度高。本发明方法能够精准地识别使用全局代理访问出口以外网站的用户及代理服务器。



CN 108566380 A

1.一种代理上网行为识别与检测方法,其特征在于,

采用分流设备将网络中数据流量全流量镜像到捕包引擎中进行初步处理,即根据数据流入和数据流出分别应对不同的规则,在数据流出时,记录目的地址不在境内的IP,在数据流入时,记录源地址不在境内的IP;

将数据流出时和数据流入时记录的IP保存在存储引擎中;

储存引擎中的IP数据再提供给分析引擎,供分析引擎进行分析;所述分析引擎为一个时时运行的守护进程,其不停地扫描存储的IP数据并进行匹配,如果发现流出方向的数据源地址与目的地址在设定时间以内同流入方向的数据目的地址和源地址匹配,则将这个访问会话标识为可疑访问会话;

将标识为可疑访问会话的数据流入方向的源地址加入到应用识别引擎中去进行应用识别;

当应用识别引擎识别出目标IP开放的所有协议后,再调用协议验证引擎进行协议验证,若验证出有协议为伪装协议,则判定境外IP为代理服务器访问会话或者判定境外IP为代理服务器访问会话的可疑度高,若验证出有协议为真实协议,则判定境外IP为无代理服务器访问会话。

2.如权利要求1所述的一种代理上网行为识别与检测方法,其特征在于,所述代理上网行为识别与检测方法采用C/S模式结构。

3.如权利要求1所述的一种代理上网行为识别与检测方法,其特征在于,还包括后台结果展示,即采用B/S模式结构,基于轻量级WEB开发语言PHP开发,结合laravel或jquery框架形成展示系统。

4.如权利要求1所述的一种代理上网行为识别与检测方法,其特征在于,在进行协议验证时,采用传统的网络安全端口扫描的方式识别开放的应用,并使用协议“真”识别的方式去验证开放的这些应用是否真实。

5.如权利要求1所述的一种代理上网行为识别与检测方法,其特征在于,分析引擎进行分析时,设定时间为60秒。

6.如权利要求4所述的一种代理上网行为识别与检测方法,其特征在于,所述传统的网络安全端口扫描采用的扫描工具为NMAP或MASSCAN。

# 一种代理上网行为识别与检测方法

## 技术领域

[0001] 本发明涉及代理上网行为识别技术领域，特别是一种代理上网行为识别与检测方法。

## 背景技术

[0002] 目前国际、国内反恐维稳任务艰巨，新技术新业务日益更新，传统的手段和方法已不能满足新形势下的工作。特别的，对于一些重要的地区，特殊敏感时期保障任务艰巨，责任重大，急需能够识别使用加密代理上网行为的技术或手段。

[0003] 当下使用新技术来对抗监管的手段越来越多，如Shadowsocks等强加密流量且流量无明显特征的代理软件来传递消息或者获取非法内容。但是每个人使用的电脑、手机上都安装了多个应用软件，如手机上基本人人都是用QQ、微信等APP。这些应用软件随时都会向服务器发送数据或者检查升级，如果在这些设备上使用了VPN全局代理或者Shadowsocks全局代理来躲避监管，那么上述软件的行为也将会通过这些通道返回到国内服务器中，因此，通过上述特征结合一定的手段可以精准地识别出使用代理的用户以及代理服务器。

## 发明内容

[0004] 本发明所要解决的技术问题是提供一种代理上网行为识别与检测方法，能够精准地识别使用全局代理访问出口以外网站的用户及代理服务器。

[0005] 为解决上述技术问题，本发明采用的技术方案是：

[0006] 一种代理上网行为识别与检测方法，采用分流设备将网络中数据流量全流量镜像到捕包引擎中进行初步处理，即根据数据流入和数据流出分别应对不同的规则，在数据流出时，记录目的地址不在境内的IP，在数据流入时，记录源地址不在境内的IP；

[0007] 将数据流出时和数据流入时记录的IP保存在存储引擎中；

[0008] 储存引擎中的IP数据再提供给分析引擎，供分析引擎进行分析；所述分析引擎为一个时时运行的守护进程，其不停地扫描存储的IP数据并进行匹配，如果发现流出方向的数据源地址与目的地址在设定时间以内同流入方向的数据目的地址和源地址匹配，则将这个访问会话标识为可疑访问会话；

[0009] 将标识为可疑访问会话的数据流入方向的源地址加入到应用识别引擎中去进行应用识别；

[0010] 当应用识别引擎识别出目标IP开放的所有协议后，再调用协议验证引擎进行协议验证，若验证出有协议为伪装协议，则判定境外IP为代理服务器访问会话或者判定境外IP为代理服务器访问会话的可疑度高，若验证出有协议为真实协议，则判定境外IP为无代理服务器访问会话。

[0011] 进一步的，所述代理上网行为识别与检测方法采用C/S模式结构。

[0012] 进一步的，还包括后台结果展示，即采用B/S模式结构，基于轻量级WEB 开发语言PHP开发，结合laravel或jquery框架形成展示系统。

[0013] 进一步的，在进行协议验证时，采用传统的网络安全端口扫描的方式识别开放的应用，并使用协议“真”识别的方式去验证开放的这些应用是否真实。

[0014] 进一步的，分析引擎进行分析时，设定时间为60秒。

[0015] 进一步的，所述传统的网络安全端口扫描采用的扫描工具为NMAP或 MASSCAN。

[0016] 与现有技术相比，本发明的有益效果是：可以无需关注流量封装的具体内容，也可不关注流量是否加密即可精准地识别使用全局代理访问出口以外网站的用户及代理服务器。为了适应更多的协议验证，还可以采用开放式的开发设计，即任何人都可以根据规范编写验证插件，不需要改动系统程序结构。

## 附图说明

[0017] 图1是本发明方法整体架构示意图。

[0018] 图2是本发明方法具体流程示意图。

[0019] 图3是采用本发明方法后的界面展示图。

[0020] 图4是PC的正常连网状态图。

[0021] 图5是正常连网时QQ显示图。

[0022] 图6是网络连接断开时状态图。

[0023] 图7是网络连接断开时QQ显示图。

[0024] 图8是采用本发明方法的实验系统图。

[0025] 图9是本发明中实验时捕获的最新数据的前五位。

[0026] 图10是本发明实验时得到的疑似代理服务器IP。

[0027] 图11是本发明实验时得到的开放的端口信息。

[0028] 图12是与图10中IP有过数据交互的信息的前五位。

[0029] 图13是本发明实验分析和验证结果。

## 具体实施方式

[0030] 下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0031] 若在需要管控地区出口处区分境内IP和境外IP，并将所有非常规的访问源 IP和目的IP做一个异常列表并重点监控，在之后发现有异常列表中目的IP向国内服务器请求已知应用数据且目的IP并没有开放常见的应用服务的情况下，那么基本可以判断此种行为是代理行为。

[0032] 本发明方法采用前后端分离模式设计，前端采用B/S架构，后端采用C/S 模式设计，整体架构如图1所示。

[0033] 后端C/S模式系统结构如下：分流设备将网络中数据流量全流量镜像到捕包引擎中进行初步处理，在这里将区分出数据流入和流出分别应对不同的规则，即：流出方向只记录目的地址不在境内的IP，流入方向的数据只记录源地址不再境内的IP，将这些数据记录并保存在存储引擎中供分析引擎进行分析，分析引擎为一个时时运行的守护进程，将不停地扫描存储的数据并进行匹配，如果发现流出方向的数据源地址与目的地址在60秒以内同流入方向的目的地址和源地址匹配，则将这个访问会话标识为高度可疑，并将流入方向的源地址(境外 IP)加入到应用识别系统中去进行应用识别；当应用识别引擎识别出目标IP

开放的所有协议后，为了保证数据的可靠性，将会在调用协议验证引擎进行协议验证，并最终判断结果。具体流程如图2所示。

[0034] 前端B/S模式系统结构如下：基于轻量级WEB开发语言PHP开发，结合热门的laravel、jquery等框架形成一个展示系统，主要将后台结果展示出来，主界面如图3所示。

[0035] 本发明方法的原理之一为：1)、用户设备上安装有多个国内的热门应用，如微信、QQ、微博、360杀毒等；2)、用户使用VPN上网后上述热门应用会自动重新连接服务器，而这些热门应用发起的连接请求均都回通过VPN服务器反连回来。本发明方法要收集常用热门APP的服务器IP。

[0036] 从国际IP分配机构获取属于中国IP，并做成一个KEY-VALUE的高效查询库用于查询，但是这种方式也有一定的缺陷如对于广播IP的方式则可能会发生误判。

[0037] 采用64位LINUX系统加libpacp库实现流量的高效提取和分析，分析传输中的源地址、目的地址、源端口和目的端口。经过实验这种方法在服务器配置：CPU XEN E5-2609*2，内存2GB，千兆网卡的接入环境中可以实现实际传输 50MB/S的流量捕获和分析。

[0038] 针对部分代理服务器可能开启了混淆和伪装的协议，本发明方法在研究中采用了传统的网络安全端口扫描的方式识别开放的应用，并使用协议“真”识别的方式去验证开放的这些应用是否真实，比如我SHADOWSOCKS开放的端口是 443，则一般传统的应用识别会将此协议标注为HTTPS，而为了验证这个端口开放的协议是不是真的HTTPS，则验证引擎会模拟一个真实的HTTPS去请求服务器，如果是真实的HTTPS协议，则会对模拟的请求给出正确的反馈，如果是伪装的HTTPS协议，则不会反馈或给出错误的反馈。

[0039] 本发明方法为了适应更多的协议验证，则采用了开放式的开发设计，即任何人都可以根据规范编写验证插件，而不需要改动系统程序结构。

[0040] 经常使用手机或者PC都可以发现一个规律，当你网络环境改变后系统内安装的时时网络程序都会及时的重新连接服务器，这一点尤其是在腾讯QQ、微信等即时通信程序中最为常见。例如：

[0041] a)正常连网的PC，安装腾讯QQ并登陆，如图4、图5所示，可以看到腾讯QQ正常连接状态。

[0042] b)手动将网络连接断开，并开始计时，大约5秒后，腾讯QQ会自动下线 (断开)。

[0043] c)手动将网络连接上，并开始计时，大约5秒后，腾讯QQ会自动上线(连接上服务器)。

[0044] 当连接VPN或者全局代理时腾讯QQ也会实现如上的下线-上线流程，以检测和改变自身网络环境，这是其特征之一。而一般正常的境内访问境外的业务，如跨国公司，跨地区公司在通过VPN访问公司资源一般都只是境内发起请求到境外，而极少境外发起请求到境内，更不可能当一个境内向境外IP发起请求后再极短的时间内境外IP又向境内热门的应用发起请求，这是其特征之二。正是利用这几点可以在网络出口地方进行流量捕获分析并结合一定的规则算法就可以识别出异常会话，判断公式如表1所示。

[0045] 表1判断公式及结论

[0046]

| 输入数据 | 计算公式 | 结论 |
| --- | --- | --- |
| 境内 IP(Y1) ->请求->境外 IP(M1)<br>境外 IP(Y2)->请求->境内 IP(M2)<br>应用识别结果 (R1) | M1 等于 Y2 且 R1 为空 | 上网行为：使用代理 50%<br>代理服务器：50%代理服务器 |
| | M1 等于 Y2 且 R1 等于常规应用 | 正常行为 |
| | M1 等于 Y2 且 R1 等于代理服务器 或 R1 等于 VPN 服务器 | 上网行为：使用代理 90%<br>代理服务器：90%代理服务器 |

[0047] 代理服务器识别技术一直是研究的难点，现在为了对抗识别，更是使用混淆、伪装技术。并且作为一个网络出口处工作的系统，必须要有极快的速度来识别。本发明方法利用了网络安全端口扫描技术来初步实现目标服务器的端口开放应用协议识别，目前开源的NMAP、MASSCAN都是比较优秀的端口扫描工具，并且速度很快。采用NMAP，例如使用常规方法对一个IP进行端口扫描，扫描 1000个端口，不考虑网络因素，实际测试约5秒左右。

[0048] 为了更准确的识别协议，本发明方法采用了“插件”的方式验证识别出的协议，主要实现方式为：如果NMAP识别出是imap协议，则将IP和端口提交要验证组件中，验证组件会模拟imap的协议去访问目标IP，如果能得到正确的连接，则表明目标协议为真IMAP协议，如果不能建立连接，则目标协议为假的协议，也表明目标IP有存在代理服务器的可能。

[0049] 下面通过具体实例对本发明方法及其有益效果进行进一步的验证。

[0050] 实验环境说明：

[0051] 1、部署如图8所示的网络拓扑的实验环境；

[0052] 2、境外VPN服务器为在LINODE上购买的VPS服务器；

[0053] 3、捕包服务器为本地服务器上的虚拟机服务器，上面安装了开发的后端C/S 模式的程序组件；

[0054] 4、WEB服务器为模拟热门应用服务器(因网络限制模拟为热门应用服务器)；

[0055] 5、普通PC为发起VPN连接请求并模拟热门应用自动访问服务器(安装了 CISCO VPN客户端)。

[0056] 实验步骤：

[0057] 1、在捕包服务器上启动各个组件，让系统进入分析状态，在系统调试日志中可以看到各项输出信息。

[0058] 2、在普通PC 172.16.0.172上面使用VPN客户端与服务器建立连接，在建立连接成功后的60秒内访问http://221.237.189.127(因为实验环境不能捕获到真正的热门应用IP

数据，因此采用手动访问221.237.189.127这一IP来模拟热门应用自动连接服务器这一动作)。

[0059] 3、当第2步完成后查看系统界面可以发现捕获有172.16.0.172与45.79.69.72 的数据记录，并将这个会话标注为可疑会话。查看该IP信息可以发现此IP开放了多个端口，以及和该IP进行过数据交互的最新TOP5IP信息。

[0060] 4、当45.79.69.72这个IP的检测任务完成后，在查看可疑列表发现该IP为代理服务器的可能性有90％，端口443为虚假协议，所有与该IP的443端口连接的境内IP都有可能是使用代理上网用户。

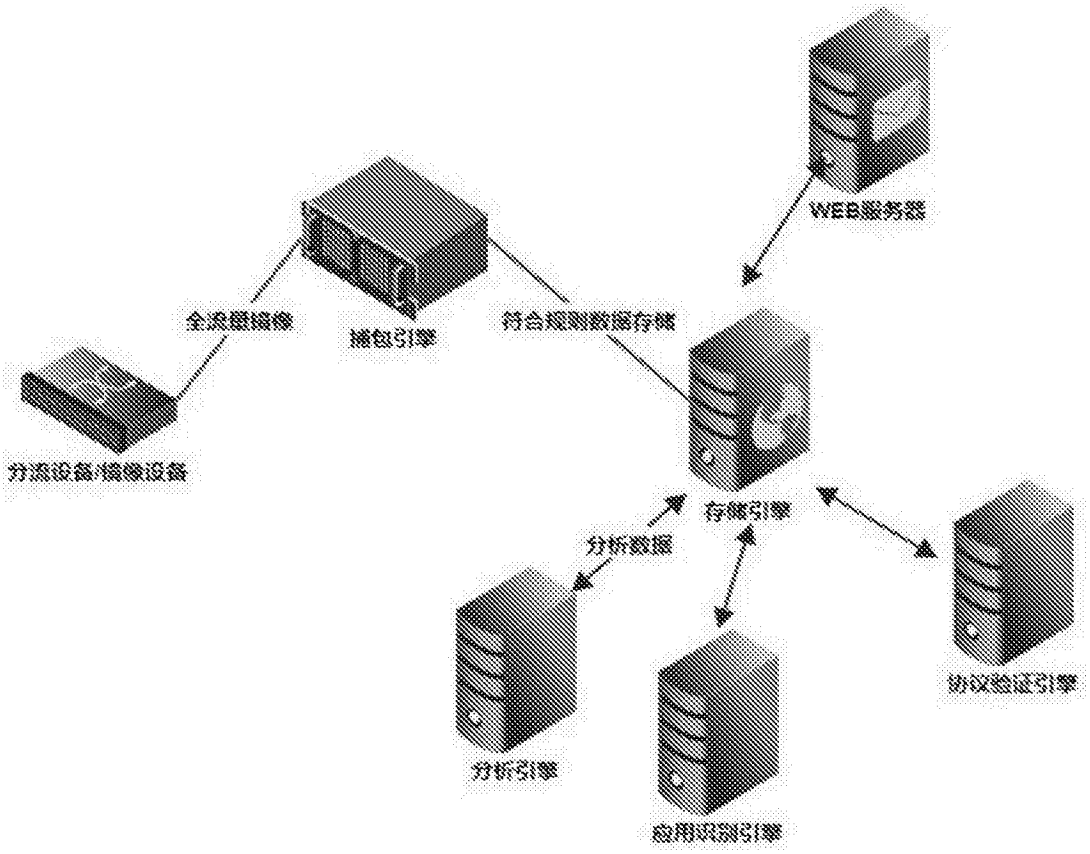

WEB服务器
全流量镜像
捕包引擎
符合规则数据存储
分流设备/镜像设备
存储引擎
分析数据
协议验证引擎
分析引擎
应用识别引擎

图1

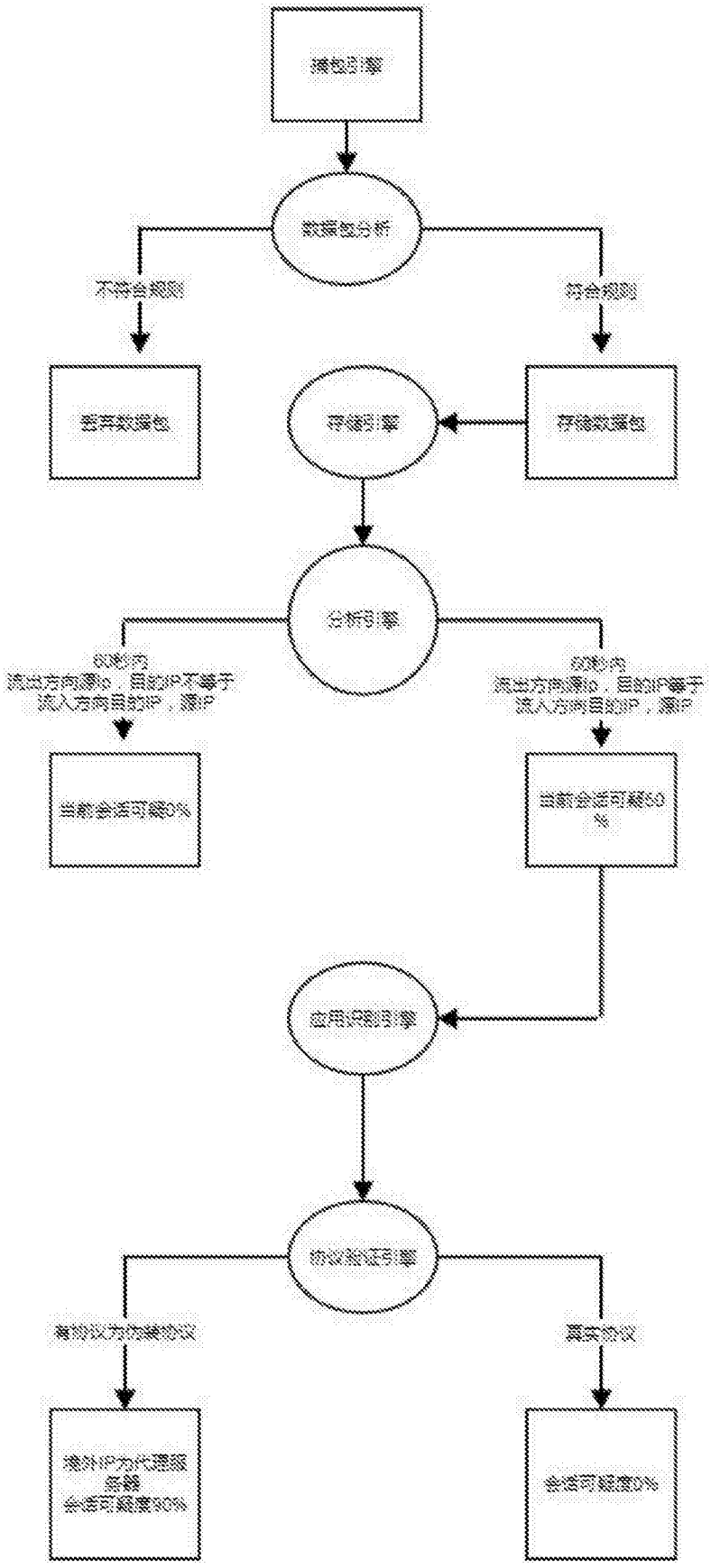

抓包引擎
数据包分析
不符合规则
符合规则
丢弃数据包
存储引擎
存储数据包
分析引擎
双向为
流出方向源ip，目的IP等于
流入方向目的IP，源IP
双向为
流出方向源ip，目的IP等于
流入方向目的IP，源IP
当前会话可疑0%
当前会话可疑50%
应用识别引擎
协议验证引擎
有协议方向的协议
真实协议
源IP为代理服务器
会话可疑度30%
会话可疑度0%

图2


图3

图4

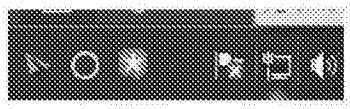


图5


图6

图7

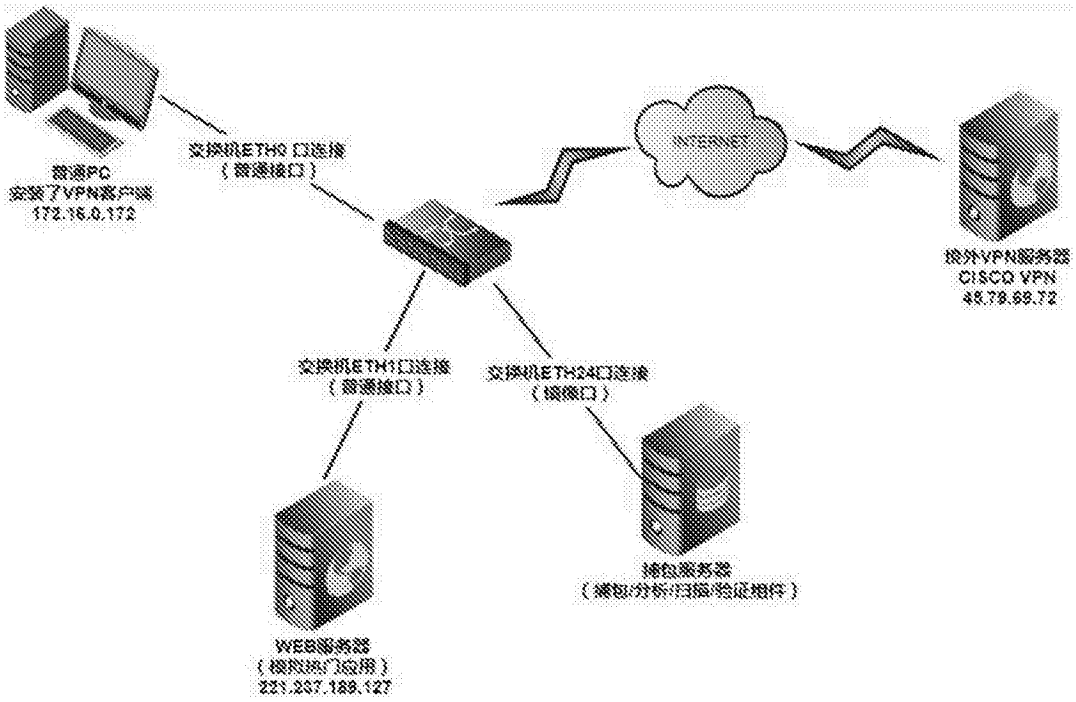

普通PC
安装了VPN客户端
172.16.0.172
交换机ETH0 口连接
（普通接口）
INTERNET
境外VPN服务器
CISCO VPN
45.79.69.72
交换机ETH1口连接
（普通接口）
交换机ETH24口连接
（镜像口）
WEB服务器
（模拟热门应用）
221.237.169.127
抓包服务器
（抓包/分析/扫描/验证相件）

图8

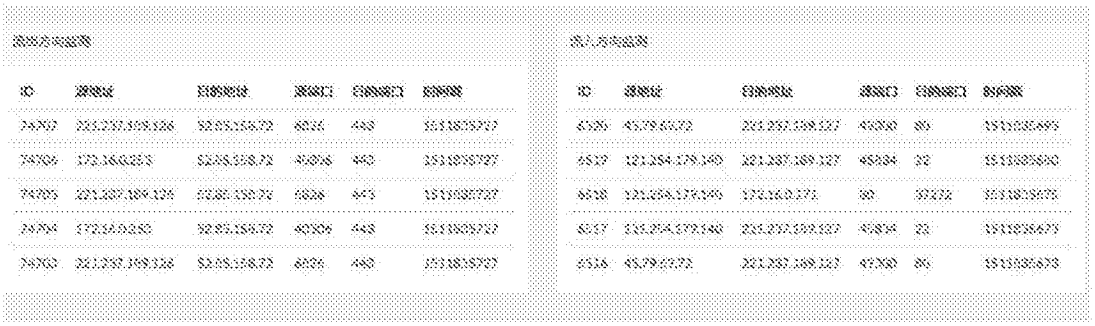


图9

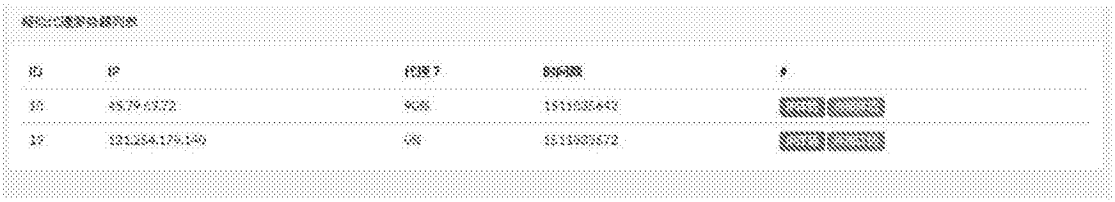


图10


图11

浏览信息

出口

| ID | 源地址 | 目的地址 | 源端口 | 目的端口 | 时间戳 |
|---|---|---|---|---|---|
| 74662 | 221.237.189.126 | 45.79.69.72 | 19121 | 993 | 1511835725 |
| 74661 | 172.16.0.172 | 45.79.69.72 | 49296 | 993 | 1511835725 |
| 74645 | 221.237.189.126 | 45.79.69.72 | 19121 | 993 | 1511835725 |
| 74644 | 172.16.0.172 | 45.79.69.72 | 49296 | 993 | 1511835725 |
| 73680 | 221.237.189.127 | 45.79.69.72 | 80 | 49300 | 1511835693 |

入口

| ID | 源地址 | 目的地址 | 源端口 | 目的端口 | 时间戳 |
|---|---|---|---|---|---|
| 6520 | 45.79.69.72 | 221.237.189.127 | 49300 | 80 | 1511835693 |
| 6516 | 45.79.69.72 | 221.237.189.127 | 49300 | 80 | 1511835673 |
| 6513 | 45.79.69.72 | 172.16.0.171 | 443 | 57842 | 1511835645 |
| 6512 | 45.79.69.72 | 221.237.189.127 | 49301 | 80 | 1511835641 |

图12

图13