(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105657051 B
(45)授权公告日 2020.03.24

(21)申请号 201610121012.9

(22)申请日 2016.03.03

(65)同一申请的已公布的文献号
申请公布号 CN 105657051 A

(43)申请公布日 2016.06.08

(73)专利权人 广东顺德中山大学卡内基梅隆大学国际联合研究院
地址 528300 广东省佛山市顺德区大良南国东路9号
专利权人 中山大学

(72)发明人 丁泽伟 余顺争

(74)专利代理机构 广州华进联合专利商标代理有限公司 44224
代理人 王程

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

(56)对比文件
CN 103220329 A,2013.07.24,
CN 101505314 A,2009.08.12,
CN 101442519 A,2009.05.27,
US 2008141358 A1,2008.06.12,
杨锐.特征字符串匹配在P2P流量控制中的应用.《科技信息》.2006,(第11期),

审查员 张诗纬

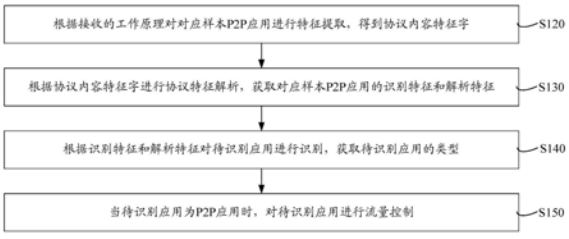
权利要求书2页 说明书8页 附图2页

(54)发明名称

P2P类应用的识别控制方法和系统

(57)摘要

本发明涉及一种P2P类应用的识别控制方法和系统，根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字。根据协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征。根据识别特征和解析特征对待识别应用进行识别，获取待识别应用的类型；当待识别应用为P2P应用时，对待识别应用进行流量控制。通过结合P2P应用的识别特征和解析特征，实现对某个或者某类P2P应用流量的准确定位，并进行识别与控制，能有效提高P2P类应用的识别率，降低误识别率。



CN 105657051 B

1.一种P2P类应用的识别控制方法,其特征在于,包括以下步骤:

根据接收的工作原理对对应样本P2P应用进行特征提取,得到协议内容特征字;具体包括:分析P2P应用的工作原理,针对在P2P应用中客户端与服务器连线的行为做分类架构P2P网络,使用Wireshark抓取一步连线时传输与接收的数据包,进而分析抓取下来的数据包,提取P2P应用的协议内容特征字提并用正则表达式表达形成协议特征库;

根据所述协议内容特征字进行协议特征解析,获取对应样本P2P应用的识别特征和解析特征,具体包括:对所述协议内容特征字进行解析,得到用于识别P2P应用类型的特征字,作为P2P应用的识别特征;提取所述协议内容特征字与协议、格式、报文结构相关的解析特征;将P2P应用的识别特征和解析特征用正则表达式表示;

根据所述识别特征和解析特征对待识别应用进行识别,具体将识别特征和解析特征应用于Linux系统下的L7-filter,对待识别应用进行识别,获取所述待识别应用的类型;

当所述待识别应用为P2P应用时,利用L7-filter对所述待识别应用进行流量控制,限制所述待识别应用的带宽。

2.根据权利要求1所述的P2P类应用的识别控制方法,其特征在于,根据接收的工作原理对对应样本P2P应用进行特征提取,得到协议内容特征字的步骤之前,还包括以下步骤:

分析在样本P2P应用中客户端与服务器的连接行为、对等节点的相关信息获取以及与对等节点进行信息交互过程,得到对应样本P2P应用的工作原理。

3.根据权利要求1所述的P2P类应用的识别控制方法,其特征在于,所述对所述待识别应用进行流量控制包括:使用iptables进行流量过滤控制,或使用TC设置流量控制策略进行流量控制。

4.一种P2P类应用的识别控制系统,其特征在于,包括:

特征提取模块,用于根据接收的工作原理对对应样本P2P应用进行特征提取,得到协议内容特征字;具体包括:分析P2P应用的工作原理,针对在P2P应用中客户端与服务器连线的行为做分类架构P2P网络,使用Wireshark抓取一步连线时传输与接收的数据包,进而分析抓取下来的数据包,提取P2P应用的协议内容特征字提并用正则表达式表达形成协议特征库;

特征解析模块,用于根据所述协议内容特征字进行协议特征解析,获取对应样本P2P应用的识别特征和解析特征;

应用识别模块,用于根据所述识别特征和解析特征对待识别应用进行识别,具体将识别特征和解析特征应用于Linux系统下的L7-filter,对待识别应用进行识别,获取所述待识别应用的类型;

流量控制模块,用于在所述待识别应用为P2P应用时,利用L7-filter对所述待识别应用进行流量控制,限制所述待识别应用的带宽;

所述特征解析模块包括:

第一解析单元,用于对所述协议内容特征字进行解析,得到用于识别P2P应用类型的特征字,作为P2P应用的识别特征;

第二解析单元,用于提取所述协议内容特征字与协议、格式、报文结构相关的解析特征;

第三解析单元,用于将P2P应用的识别特征和解析特征用正则表达式表示。

5.根据权利要求4所述的P2P类应用的识别控制系统，其特征在于，还包括原理分析模块，所述原理分析模块用于在特征提取模块根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字之前，分析在样本P2P应用中客户端与服务器的连接行为、对等节点的相关信息获取以及与对等节点进行信息交互过程，得到对应样本P2P应用的工作原理。

6.根据权利要求4所述的P2P类应用的识别控制系统，其特征在于，所述流量控制模块对所述待识别应用进行流量控制包括：使用iptables进行流量过滤控制，或使用TC设置流量控制策略进行流量控制。

P2P类应用的识别控制方法和系统

技术领域

[0001] 本发明涉及软件控制技术领域，特别是涉及一种P2P类应用的识别控制方法和系统。

背景技术

[0002] P2P(PEER-TO-PEER，对等网络)中的计算机之间可以互相通信和共享资源(文件、外设等)，在人们的工作生活中骑着重要作用。

[0003] 传统的P2P类应用流量识别方法主要是端口识别法，使用相对固定的端口进行连接控制和数据通信。例如，eDonkey使用4661或4662端口，BT使用的端口在6881-6890之间等等。端口识别法是基于协议端口固定来识别P2P流量，可用于通过检测端口来确定是否为目标流量。但随着动态端口和伪端口的出现，端口识别法已逐渐失去其原有的识别效果。传统的P2P类应用流量识别方法存在识别率低的缺点。

发明内容

[0004] 基于此，有必要针对上述问题，提供一种识别率高的P2P类应用的识别控制方法和系统。

[0005] 一种P2P类应用的识别控制方法，包括以下步骤：

[0006] 根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字；

[0007] 根据所述协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征；

[0008] 根据所述识别特征和解析特征对待识别应用进行识别，获取所述待识别应用的类型；

[0009] 当所述待识别应用为P2P应用时，对所述待识别应用进行流量控制。

[0010] 一种P2P类应用的识别控制系统，包括：

[0011] 特征提取模块，用于根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字；

[0012] 特征解析模块，用于根据所述协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征；

[0013] 应用识别模块，用于根据所述识别特征和解析特征对待识别应用进行识别，获取所述待识别应用的类型；

[0014] 流量控制模块，用于在所述待识别应用为P2P应用时，对所述待识别应用进行流量控制。

[0015] 上述P2P类应用的识别控制方法和系统，根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字。根据协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征。根据识别特征和解析特征对待识别应用进行识别，获取待识别应用的类型；当待识别应用为P2P应用时，对待识别应用进行流量控制。通过结合P2P

应用的识别特征和解析特征，实现对某个或者某类P2P应用流量的准确定位，并进行识别与控制，能有效提高P2P类应用的识别率，降低误识别率。

附图说明

[0016] 图1为一实施例中P2P类应用的识别控制方法的流程图；

[0017] 图2为一实施例中P2P类应用的识别与控制系统的基本框架示意图；

[0018] 图3为一实施例中P2P类应用的识别控制系统的结构图。

具体实施方式

[0019] 一种P2P类应用的识别控制方法，能够通过P2P类应用协议内容特征字实现P2P类应用的识别与控制，如图1所示，包括以下步骤：

[0020] 步骤S120：根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字。根据工作原理对样本P2P应用进行特征提取，得到协议内容特征字，在其中一个实施例中，步骤S120包括步骤122和步骤124。

[0021] 步骤122：根据工作原理捕捉样本P2P应用的流量数据包。在熟悉样本P2P应用工作原理后，具体可通过Wireshark等网络嗅探器捕捉P2P类各种具体应用的流量数据包，操作简便且数据捕捉可靠性高。

[0022] 步骤124：提取流量数据包的协议内容特征字，并用正则表达式表达形成协议特征库。正则表达式描述了一种字符串匹配的模式，可以用来检查一个串是否含有某种子串、将匹配的字串做替换或者从某个串中取出符合某个条件的字串等。通常是由普通字符以及特殊字符组成的文字模式。正则表达式作为一个模板，将某个字符模式与所搜索的字符串进行匹配。提取流量数据包的协议内容特征字并以正则表达式的方式形成协议特征库，以便于存储和调用。

[0023] 具体地，本实施例中进行协议内容特征提取可为，在熟悉要解析的P2P应用后，分析其工作原理，针对在P2P应用中客户端与服务器连线的行为做详细的分类，架构完整且无杂讯的P2P网络，使用Wireshark抓取一步连线时传输与接收的数据包，进而分析此步骤抓取下来的数据包，最后提取P2P应用的协议内容特征字形成协议特征库。其中，熟悉要解析的P2P应用时，需考虑选择哪一个版本的P2P应用当作测试目标，了解最少需要多少台机器参与其中。针对连线的行为做详细的分类时，需结合P2P应用工作原理的三个步骤。架构完整且无杂讯的P2P网络时，需考虑是否有服务器的角色存在，以及是否需要对外连线。使用Wireshark抓取一步连线时传输与接收的数据包时，需考虑传输是否使用固定接口，采用什么方式传输资料以及连线在什么条件下被终止。分析抓取下来的数据包时，需考虑协定如何分配数据栏位，资料是否有经过加密，资料是否有被切割与其切割的大小，以及连线动作可能用到的命令是什么。提取P2P协议内容特征字形成协议特征库时，需考虑2P协议内容特征字能否用正则表达式表达。

[0024] 本实施例中利用协议内容特征识别的准确性，能有效覆盖识别各类P2P类应用，提高P2P应用的识别率。此外，在分析抓取下来的数据包之后，还可判断是否所有数据包都抓取完毕，若是，则提取P2P应用的协议内容特征字形成协议特征库；若否，则可返回使用Wireshark抓取一步连线时传输与接收的数据包，再次进行数据包抓取。

[0025] 在提取P2P应用的协议内容特征字时有如下技巧:P2P应用的协议内容特征字一般位于应用层的前几个字节;P2P应用大多使用TCP(Transmission Control Protocol传输控制协议)协议传输控制信息,通过UDP(User Datagram Protocol,用户数据包协议)协议传输数据;通过Wireshark等网络嗅探器自带的协议通信量统计工具,可以为协议内容特征字提取提供线索;熟悉P2P应用的工作原理,特别是了解P2P应用特有通信协议的工作原理有利于快速提取协议内容特征字;同一P2P应用的不同版本,其对应的协议内容特征字会有不同;提取非P2P应用的协议内容特征字,不但可以缩小P2P协议内容特征字找寻的范围,而且可以为流量控制系统的设计提供帮助。

[0026] 步骤S130:根据协议内容特征字进行协议特征解析,获取对应样本P2P应用的识别特征和解析特征。在其中一个实施例中,步骤S130包括步骤132和步骤134。

[0027] 步骤132:对协议内容特征字进行解析,得到用于识别P2P应用类型的识别特征。将提取到的协议内容特征字进行深度解析,分析出能具体识别某种P2P应用的特征字,作为P2P应用的识别特征。

[0028] 步骤134:提取协议内容特征字与协议、格式、报文结构相关的解析特征。提取协议内容特征字与协议、格式、报文结构等相关的解析特征,通过该解析特征可以知道该P2P应用采用的是哪种协议或者使用的是哪种报文结构。

[0029] 利用P2P应用的解析特征,能知道P2P应用采用的协议、帧格式或报文结构,细化P2P协议解析的颗粒度,实现对P2P应用网络流量的精准控制。通过结合P2P应用的识别特征和解析特征,可有效覆盖识别各类P2P类应用,提高P2P应用的识别率;能知道P2P应用采用的协议、帧格式或报文结构,细化P2P协议解析的颗粒度,实现对P2P应用网络流量的精准控制。

[0030] 以eDonkey/eMule协议解析为例,通过对支持eDonkey/eMule协议的P2P文件下载软件(如迅雷、eMule、MLDonkey、aMule、Shareaza、Morpheus、XoloX等)进行协议特征解析,可以发现,当TCP建立连接后的第一个报文的数据的第一字节为0xe3时,可以判定该P2P应用采用eDonkey协议,若为0xe5,可以判定该P2P应用采用eMule协议;接下来4个字节是package length,其值是TCP数据段的长度减去IP+TCP包头长度再减5。

[0031] 通过协议内容特征提取和协议特征解析,可以得到协议内容特征字,按照上述流程,将已分析的一些特征列出如下:

[0032] Emule&eDonkey协议解析:

[0033]

第一字节	第二字节	第三字节	其他字节	UDP长度	类型
0xe3	0x9a	any	any	26	edonkey
0xe3	0x96	any	any	14	edonkey
0xc5	0x91	!0	any	12	emule
0xc5	0x90	!0	any	26	emule
0xc5	0x92	any	any	10	emule
0xc5	0x93	any	any	10	emule
0xe4	0x50	any	any	12	kad
0xe4	0x58	!0	any	14	kad
0xe4	0x59	any	any	10	kad

0xe4	0x30	any	0x01(19)	>26	kad
0xe4	0x28	any	0x00(69)	>76	kad
0xe4	0x20	!0	!0(35)	43	kad
0xe4	0x00	any	0x00(27)	35	kad
0xe4	0x10	any	0x00(27)	35	kad
0xe4	0x18	any	0x00(27)	35	kad
0xe4	0x40	any	1(19)0(20)	>40	kad

[0034] Vagaa协议解析

[0035] 发送请求数据包特征:0x 78 01 7B DC C9 C0 C0 3F 90 B8 6E 97 E6 35 3E A6 92 73 F3 A5 64 1B 14 F2 77

[0036] 确认(握手)数据包特征:0x DE AD BE EF

[0037] BitTorrent协议解析

[0038] UDP包特征:UDP长度24字节(含UDP头),起始8个字节为:00 00 04 17 27 10 19 80

[0039] TCP包特征:负载第一字节为0x13,而且后续数据为:“BitTorrent Protocol”

Gnutella特征

[0040] UDP包起始数据为“GNUTELLA”或“GND”

[0041] gnutella命令特征:负载最后以“\r\n”结尾,而且起始数据为:“GET/get”,或者是:“GET/uri-res/”

[0042] KaZaA特征

[0043] UDP数据部分的结尾6个字节是:“KaZaA\0”

[0044] KaZaA命令的特征为:负载最后以“\r\n”结尾,而且起始数据为:“GET/.hash=”

[0045] SoulSeek特征

[0046] 情况1:前8个字节格式为:xx xx 00 00 yy zz 00 00,其中xx xx为16位负载长度减4,yy!=0,zz任意

[0047] 情况2:数据长度8字节全0

[0048] 情况3:数据格式为:01 xx 00 00 00 yy..zz 00 00 00..,其中负载长度大于xx+6,负载第xx+4+1字节(zz)不为0,而负载第xx+5+1字节,第xx+6+1字节为0。

[0049] WinMX特征

[0050] 负载长度为4字节时负载内容为“SEND”

[0051] 负载长度为3字节时负载内容为“GET”

[0052] 其他情况负载长度必须大于10,负载必须以“SEND”或“GET”开头,而且负载内容中出现0x20 0x22,之后出现0x22 0x20。

[0053] appleJuice特征

[0054] 负载起始数据为“ajport\r\n”

[0055] Fasttrack特征

[0056] Get/.hash 0x270000002980

[0057] DirectConnect特征

[0058] TCP建立连接后第一个报文的数据的第一个字节值匹配于“$”,最后一个字节值匹

配与“|”。

[0059] 在“$”标识后出现的command_type字符串匹配于下面的command_type列表中的一个。command_type列表包括:MyNick,Lock,Key,Direction,GetListLen,ListLen,MaxedOut,Error,Send,Get,FileLength,Canceled,HubName,ValidateNick,ValidateDenide,GetPass,Mypass,BadPass,Version,Hello,Logedin,MyINFO,GetINFO,GetNickList,NickList,OpList,To,Connect-ToMe,MultiConnectToMe,RevConnectToMe,Search,MultiSearch,SR,Kick,OpForceMove,ForceMove,Quit。

[0060] PPlive特征

[0061] TCP数据传输的签名特征:TCP建立连接后第一个报文为4个字节,内容为**、0x00、0x00和0x00,其中**表示非0。

[0062] UDP数据传输的签名特征:报文前8字节为0xe9、0x03、**、**、0x98、0xab、0x01和0x02,其中**表示任意字节。

[0063] PPstream特征

[0064] TCP建立连接后第一个报文的数据前两个字节为0x50和0x53,对应字符串“PS”。

[0065] 接下来8个字节的值匹配于字符串“Protocol”。

[0066] QQllive特征

[0067] UDP报文的前5个字节为0xfe、0x29、0x04、0x04和0x29或者为0xfe、**、0x00、0x00和**,其中**表示非0。

[0068] 且第2个字节与第5个字节值相同。

[0069] UUsee特征

[0070] TCP数据传输的签名特征:TCP建立连接后第一个报文的数据前5个字节为0x39、0x00、0x00、0x00和0x8d或者为0x39、0x00、0x00、0x00和0x28。

[0071] UDP数据传输的签名特征:报文的前8个字节为0x13、**、**、**、**、0x00、0x00和0x00或者为0x14、**、**、**、**、0x00、0x00和0x00,其中**表示任意字节。

[0072] SopCast特征

[0073] UDP报文的前5个字节为0x00、**、0x01、**和0x00或者0x00、**、0x01、**和0x22,其中**表示任意字节。

[0074] 步骤S140:根据识别特征和解析特征对待识别应用进行识别,获取待识别应用的类型。具体可将识别特征和解析协议特征应用于Linux系统下的L7-filter,监控主机流量,L7-filter会将待识别应用于协议解析特征进行匹配,确定待识别应用是否为P2P应用,具体是哪种P2P应用。对待识别应用进行识别,具体同样可按照步骤S120和步骤S130得到待识别应用的识别特征和解析特征后,与样本P2P应用的识别特征和解析特征进行比较,若存在样本P2P应用的识别特征和解析特征与待识别应用的识别特征和解析特征相同,则可确认该待识别应用为P2P应用,且可知道具体为哪种P2P应用。

[0075] 此外,在步骤134之后,步骤S130还可包括将P2P应用的识别特征和解析特征用正则表达式表示的步骤。L7-filter使用的是V8正则表达式语法。将P2P应用的识别特征和解析特征用正则表达式表示,可以有效应用于L7-filter的字符串匹配。

[0076] 步骤S150:当待识别应用为P2P应用时,对待识别应用进行流量控制。利用L7-filter实现P2P流量识别控制,若识别出为某种P2P应用,可以通过L7-filter的相关命令实

现流量控制。本实施例中对待识别应用进行流量控制包括:使用iptables进行流量过滤控制,或使用TC(Traffic Control,流量控制)设置流量控制策略进行流量控制。

[0077] 在其中一个实施例中,步骤S120之前,P2P类应用的识别控制方法还可包括步骤110。

[0078] 步骤110:分析在样本P2P应用中客户端与服务器的连接行为、对等节点的相关信息获取以及与对等节点进行信息交互过程,得到对应样本P2P应用的工作原理。对预设的样本P2P应用进行分析,熟悉P2P应用工作原理,以便于后续进行特征提取。

[0079] 具体地,P2P应用工作一般包括以下三个步骤:P2P应用客户端通过种子文件或DNS(Domain Name System,域名系统)查询获取服务器的IP地址、客户端向服务器请求对等节点的相关信息(IP地址、资源拥有情况等)、与对等节点进行信息交互。根据P2P应用的工作步骤熟悉P2P应用的工作原理,可准确获取P2P应用的相关信息,提高后续特征提取的准确性和全面性。

[0080] 上述P2P类应用的识别控制方法,根据工作原理对对应样本P2P应用进行特征提取,得到协议内容特征字。根据协议内容特征字进行协议特征解析,获取对应样本P2P应用的识别特征和解析特征。根据识别特征和解析特征对待识别应用进行识别,获取待识别应用的类型;当待识别应用为P2P应用时,对待识别应用进行流量控制。通过结合P2P应用的识别特征和解析特征,实现对某个或者某类P2P应用流量的准确定位,并进行识别与控制,能有效提高P2P类应用的识别率,降低误识别率。

[0081] 此外,还可将上述P2P类应用的识别控制方法应用于P2P类应用的识别与控制系统。具体地,在充分研究Linux系统下L7-filter源代码的基础上,对源代码进行适度改写,要求能选择或输入P2P应用名称、P2P协议、报文结构、帧格式等来确定要进行操作的P2P应用,并进行过滤、流量控制、设置优先级等操作。另外,还需要设计一个界面用于P2P类应用识别控制输入和结果呈现。从而形成一种专门针对P2P类应用的识别与控制系统。

[0082] P2P类应用的识别与控制系统的基本框架如图2所示,包括流量识别层、流量控制层和系统管理层。主要基于Linux 2.6.17内核,根据P2P流量识别的需求,在防火墙netfilter/iptables框架下有捕获分析数据流的功能;对识别出的P2P数据流配置不同策略,对不同类型的数据流进行流量控制。再上层是TC带宽控制的管理层,是管理配置TC的应用程序。最上层是WEB配置界面,提供用户配置接口。系统的主要功能包括识别和控制两个部分,最终由流量控制模块按照数据属性和配置策略裁决处理方法。

[0083] P2P流量识别模块是P2P流量识别控制系统的重要组成部分,主要完成数据包的接收、规则匹配、P2P流量识别、数据包分发等工作。本模块设计主要采用DPI(Deep Packet Inspection,深度数据包检测技术)的识别方法。

[0084] 本系统中的P2P流量控制模块,可以不用彻底封杀P2P应用,把P2P流量限制在一定的带宽中,既满足一部分人使用P2P下载文件的愿望,又不妨碍正常的网络使用,缓解大量的P2P应用导致的网络压力。

[0085] 本系统最上层是系统管理层,设计采用基于Web的远程管理方式,只要有标准浏览器,用户就可以很方便的对系统进行远程配置。因此系统需要Web服务器对用户的管理的请求进行响应。

[0086] 本发明还提供了一种P2P类应用的识别控制系统,能够通过P2P类应用协议内容特

征字实现P2P类应用的识别与控制，如图3所示，包括特征提取模块120、特征解析模块130、应用识别模块140和流量控制模块150。

[0087] 特征提取模块120用于根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字。根据工作原理对样本P2P应用进行特征提取，得到协议内容特征字，在其中一个实施例中，特征提取模块120包括第一提取单元和第二提取单元。

[0088] 第一提取单元用于根据工作原理捕捉样本P2P应用的流量数据包。在熟悉样本P2P应用工作原理后，具体可通过Wireshark等网络嗅探器捕捉P2P类各种具体应用的流量数据包，操作简便且数据捕捉可靠性高。

[0089] 第二提取单元用于提取流量数据包的协议内容特征字，并用正则表达式表达形成协议特征库。提取流量数据包的协议内容特征字并以正则表达式的方式形成协议特征库，以便于存储和调用。

[0090] 具体地，本实施例中进行协议内容特征提取可为，在熟悉要解析的P2P应用后，分析其工作原理，针对在P2P应用中客户端与服务器连线的行为做详细的分类，架构完整且无杂讯的P2P网络，使用Wireshark抓取一步连线时传输与接收的数据包，进而分析此步骤抓取下来的数据包，最后提取P2P应用的协议内容特征字形成协议特征库。

[0091] 本实施例中利用协议内容特征识别的准确性，能有效覆盖识别各类P2P类应用，提高P2P应用的识别率。此外，在分析抓取下来的数据包之后，还可判断步骤是否所有数据包都抓取完毕，若是，则提取P2P应用的协议内容特征字形成协议特征库；若否，则可返回使用Wireshark抓取一步连线时传输与接收的数据包，再次进行数据包抓取。

[0092] 特征解析模块130用于根据协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征。在其中一个实施例中，特征解析模块130包括第一解析单元和第二解析单元。

[0093] 第一解析单元用于对协议内容特征字进行解析，得到用于识别P2P应用类型的识别特征。将提取到的协议内容特征字进行深度解析，分析出能具体识别某种P2P应用的特征字，作为P2P应用的识别特征。

[0094] 第二解析单元用于提取协议内容特征字与协议、格式、报文结构相关的解析特征。提取协议内容特征字与协议、格式、报文结构等相关的解析特征，通过该解析特征可以知道该P2P应用采用的是哪种协议或者使用的是哪种报文结构。

[0095] 利用P2P应用的解析特征，能知道P2P应用采用的协议、帧格式或报文结构，细化P2P协议解析的颗粒度，实现对P2P应用网络流量的精准控制。通过结合P2P应用的识别特征和解析特征，可有效覆盖识别各类P2P类应用，提高P2P应用的识别率；能知道P2P应用采用的协议、帧格式或报文结构，细化P2P协议解析的颗粒度，实现对P2P应用网络流量的精准控制。

[0096] 应用识别模块140用于根据识别特征和解析特征对待识别应用进行识别，获取待识别应用的类型。具体可将识别特征和解析协议特征应用于Linux系统下的L7-filter，监控主机流量，L7-filter会将待识别应用于协议解析特征进行匹配，确定待识别应用是否为P2P应用，具体是哪种P2P应用。对待识别应用进行识别，具体同样可得到待识别应用的识别特征和解析特征后，与样本P2P应用的识别特征和解析特征进行比较，若存在样本P2P应用的识别特征和解析特征与待识别应用的识别特征和解析特征相同，则可确认该待识别应用

为P2P应用，且可知道具体为哪种P2P应用。

[0097] 此外，特征解析模块130还可包括第三解析单元，用于将P2P应用的识别特征和解析特征用正则表达式表示。L7-filter使用的是V8正则表达式语法。将P2P应用的识别特征和解析特征用正则表达式表示，可以有效应用于L7-filter的字符串匹配。

[0098] 流量控制模块150用于在待识别应用为P2P应用时，对待识别应用进行流量控制。利用L7-filter实现P2P流量识别控制，若识别出为某种P2P应用，可以通过L7-filter的相关命令实现流量控制。本实施例中对待识别应用进行流量控制包括：使用iptables进行流量过滤控制，或使用TC设置流量控制策略进行流量控制。

[0099] 在其中一个实施例中，P2P类应用的识别控制系统还可包括原理分析模块，原理分析模块用于在特征提取模块120根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字之前，分析在样本P2P应用中客户端与服务器的连接行为、对等节点的相关信息获取以及与对等节点进行信息交互过程，得到对应样本P2P应用的工作原理。对预设的样本P2P应用进行分析，熟悉P2P应用工作原理，以便于后续进行特征提取。

[0100] 具体地，P2P应用工作一般包括以下三个步骤：P2P应用客户端通过种子文件或DNS查询获取服务器的IP地址、客户端向服务器请求对等节点的相关信息(IP地址、资源拥有情况等)、与对等节点进行信息交互。根据P2P应用的工作步骤熟悉P2P应用的工作原理，可准确获取P2P应用的相关信息，提高后续特征提取的准确性和全面性。

[0101] 上述P2P类应用的识别控制系统，根据工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字。根据协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征。根据识别特征和解析特征对待识别应用进行识别，获取待识别应用的类型；当待识别应用为P2P应用时，对待识别应用进行流量控制。通过结合P2P应用的识别特征和解析特征，实现对某个或者某类P2P应用流量的准确定位，并进行识别与控制，能有效提高P2P类应用的识别率，降低误识别率。

[0102] 以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

[0103] 以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

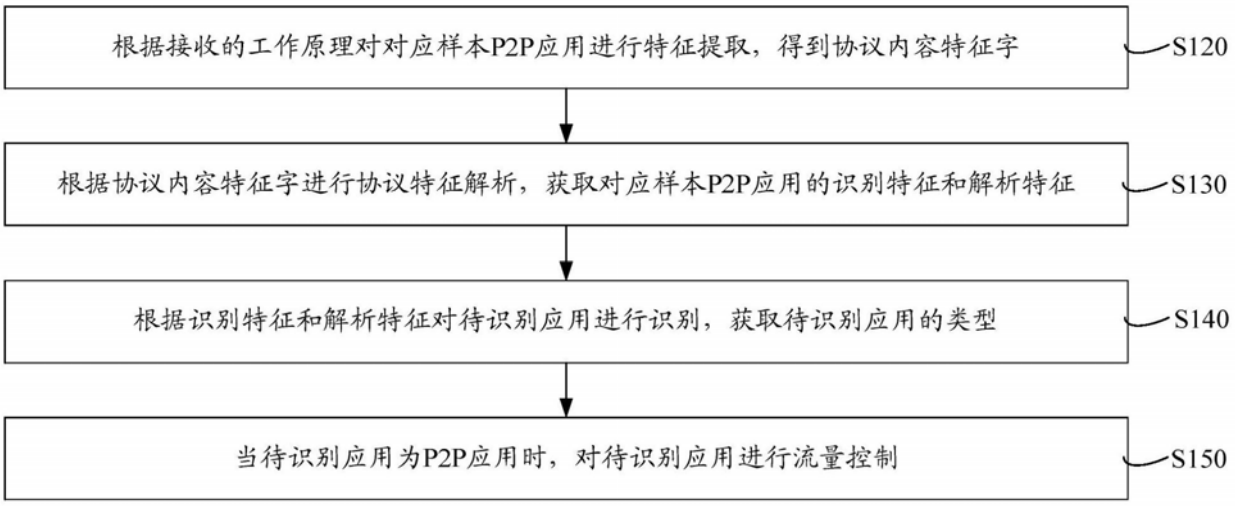

根据接收的工作原理对对应样本P2P应用进行特征提取，得到协议内容特征字
S120
根据协议内容特征字进行协议特征解析，获取对应样本P2P应用的识别特征和解析特征
S130
根据识别特征和解析特征对待识别应用进行识别，获取待识别应用的类型
S140
当待识别应用为P2P应用时，对待识别应用进行流量控制
S150

图1


远程WEB服务
Shell
系统管理层
P2P流量控制策略
流量控制层
P2P流量控制
P2P流量识别模块
流量识别层
Linux netfilter/iptables 框架

图2

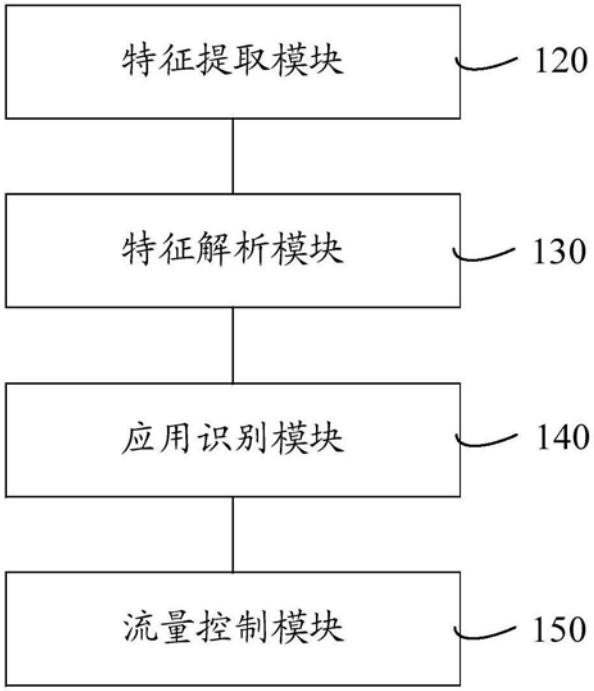


图3