

实时识别 P2P-TV 视频流的方法研究

胡超* 陈鸣 许博 李兵
(解放军理工大学指挥自动化学院 南京 210007)

摘要: 基于 P2P 的 IPTV(P2P-TV)是当前发展最为迅速的因特网应用之一, 实时识别 P2P-TV 视频流是管理网络 P2P-TV 流量和理解网络行为的关键一步。通过分析以 PPLive 为代表的 P2P-TV 体系结构、通信过程、报文结构以及系统特征, 该文提出了一种实时的基于爬虫的识别视频流 CIVF 算法和一种实时的基于协议特征的识别视频流 PIVF 算法, CIVF 算法通过爬虫程序获取 P2P-TV 节点信息来识别 P2P-TV 视频流, 而 PIVF 算法则基于视频流的通信时序和应用层负载特征实现实时识别。在因特网环境的试验分析结果表明, CIVF 算法具有实现便捷但识别率不够高且节点信息残存时间较长的特点, PIVF 算法则具有准确率较高、识别速度较快和扩展性强的特点。

关键词: 实时识别; P2P; P2P-TV; 爬虫; 协议特征

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2011)09-2219-06

DOI: 10.3724/SP.J.1146.2010.00975

Research on the Real-time Identification Methods for P2P-TV Flows

Hu Chao Chen Ming Xu Bo Li Bing

(Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: P2P-based IPTV (P2P-TV) is one of the Internet applications grown dramatically currently. It is the key to identify P2P-TV video streaming in real-time for managing the P2P-TV traffic and understanding the networking behaviors. Analyzing the architecture, communications processes, message frames and system characteristics of P2P-TV represented by PPLive, a Crawler-based Identifying Video Flows (CIVF) algorithm and a Protocol characteristic-based Identifying Video Flows (PIVF) algorithm identifying video flows in real-time are proposed. CIVF identifies P2P-TV flows based on the peer information, which is obtained from crawler program, while PIVF makes use of the communication sequence and payload characteristic in P2P-TV applications to achieve real-time identification. The results of experiments in Internet environment shows that CIVF algorithm has features such as ease to realize, not high accurateness and long outdated peer information, and PIVF algorithm has features such as high accurateness, fast working and high expansibility.

Key words: Real-time identification; P2P; P2P-TV; Crawler; Protocol characteristic

1 引言

经过 IP 协议向桌面用户分发电视流式内容是当前因特网的一种主流应用, 并且由于 P2P 技术具有的资源聚合的经济性、较好的扩展性和可靠性等优点, 使这些系统大都利用 P2P 覆盖网络传输流式(streaming)内容。目前我们见到的最为流行的应用包括 P2P 应用无线电、P2P 流式音乐和基于 P2P 的因特网电视(P2P-TV)。

由于电视媒体对人类文化和生活方式有重要影响力, 并且在因特网中产生了的巨大流量, 理解、管理和引导 P2P-TV 应用的行为是当前各研究机

构、因特网服务提供商研究和关注的主题, 而这一切的前提是能识别出 P2P-TV 流。所谓 P2P-TV 流是指以 P2P 方式传输的电视信号报文集合, 它们符合五元组{源 IP 地址, 源端口号, 目的 IP 地址, 目的端口号, 运输层, 协议类型}流规范和 64 秒超时定义^[1]。本文重点研究 P2P-TV 视频流的识别技术, 鉴于 PPLive^[2]是现有研究的主要对象^[3-6], 我们的研究以 PPLive 流为实例进行。由于其他流行的 P2P-TV 系统如 PPStream^[7], SopCast^[8]和 UUSee^[9]等与 PPLive 有不少相似之处, 因此本文工作具有普遍意义。

当前识别 P2P 应用的方法有 3 类: 基于应用协议特征字的识别方法^[10]通过分析 P2P 协议的应用层负载, 提取出能唯一标识出协议类型的特征字符串, 来识别 P2P 应用。这种方法准确率高, 但无法识别

2010-09-07 收到, 2011-07-07 改回

国家 863 计划项目(2007AA01Z418)和江苏省自然科学基金(BK2009058, BK2010133)资助课题

*通信作者: 胡超 huchaonj@hotmail.com

加密的数据。基于行为特征的识别方法^[4,11,12]综合利用了流的属性、统计和行为特性,根据启发式规则对流进行分析,达到识别 P2P 应用的目的。这种方法识别时间较长,而且不同 P2P 应用之间的行为特征差异往往不够明显,识别准确率不够高。基于机器学习的分类方法^[13,14]通过提取各类应用的流、分组等信息对分类器进行训练,然后利用训练好的分类器对流进行分类。这种方法训练分类器的时间长,而各类应用的流和分组信息不存在明显的区分,因此准确性有待提高。目前这些识别 P2P-TV 视频流的方法仍存在着识别率不高、识别滞后等缺陷,需要我们从 P2P-TV 的特点出发,研究出更好的识别算法。

2 P2P-TV 系统分析

2.1 P2P-TV 体系结构概述

在 P2P-TV 系统中,为了降低成本和便于系统构建,通常利用覆盖网络多播技术实现内容分发。P2P-TV 的覆盖网络可以分为结构化的覆盖网络和无结构的覆盖网络^[15],结构化的 P2P-TV 覆盖网络一般为树型(tree-based)结构,它通过在应用层构建一棵或多棵多播树以实现视频流的广播或多播。在树型系统中,数据总是从父节点向子节点传输,容易产生单点故障;特别是在波动性大的环境中,整个系统很可能会崩溃,因此,基于网状(mesh-based)结构的 P2P-TV 覆盖网络应运而生。这种覆盖网络通常是无结构,每个节点都同时会与一定数量的节点保持邻居关系,并采用数据驱动的策略来确定向哪些节点下载哪些数据,即使有节点因故退出系统,节点也能够及时从其他的节点处下载数据,所以具有很强的鲁棒性和扩展性,当前绝大部分主流 P2P-TV 系统都采用了这种结构。

2.2 P2P-TV 通信过程分析

P2P-TV 系统通常包括频道列表服务器、节点列表服务器和视频流服务器等部分。频道列表服务器主要负责向节点提供最新的频道列表;节点列表服务器记录了当前正在网络中运行 P2P-TV 客户端程序的节点(以下简称节点)的 IP 地址和端口号等信息,节点向节点列表服务器发送请求报文从而获取正在观看相同频道的其他节点的信息;视频流服务器对电视节目进行编码,并产生适合于网络传输的视频流。

P2P-TV 通信过程可以分为两个阶段:启动阶段和重放阶段。启动阶段是指从用户启动程序到用户选中频道之前这段时间,其间完成从频道列表服务器下载并更新频道列表以及检测软件升级等任

务;重放阶段指从用户选中某个频道进行播放直至用户退出系统这段时间,视频流分组的交互是在该阶段完成的。此外,P2P-TV 系统的通信过程具有以下特点:

(1)节点与服务器以及与其他节点的通信过程具有严格的时序关系:用户在启动 P2P-TV 客户端之后,节点首先主要采用 TCP 协议与服务器进行初始化过程。节点先向频道列表服务器请求并下载最新的频道列表;在用户选中频道后,节点开始向节点列表服务器发送请求报文以获取网络中当前其他节点的信息。此后,才开始下载和交换视频流分组。

(2)节点间在建立邻居关系时需要进行应用层的握手过程:邻居节点间在传输视频流分组之前需要有应用层的连接,之后的数据传输都经过该连接传输。因此,通过识别这种连接报文就可判断该流是否为 P2P-TV 视频流。

(3)节点均设有 UDP 监听端口:为了便于与其他节点交互,每个节点都设有一个 UDP 监听端口,该端口号对不同的节点可能是不同的,其他节点要通过该端口才能与该节点传送视频流分组。因此,一旦某节点要主动与另一个节点的某个 UDP 端口通信,被连接方的 UDP 端口必定是该节点的 UDP 监听端口。

2.3 节点列表报文分析

P2P-TV 节点在从节点列表服务器处获取节点的 IP 地址和端口信息,图 1 解析了 PPLive 节点列表请求报文的应用层信息。其中,应用层前 2 个字节(e9 03)以及第 5 至第 8 个字节(98 ab 01 02)是固定的;第 3 个字节(02)表示的是节点向服务器发送的报文序列;第 4 个字节表示的是该报文是请求报文还是响应报文,图中的“01”表示该报文为请求报文;第 9 个字节表示节点在本次节点信息请求阶段累计发送的报文数目;后面的字段表示频道的 ID。

与上述请求报文对应的响应报文的典型格式如图 2 所示。其中,除第 4 个字节为 00 表示该报文为响应报文外,其他前 28 字节格式与请求报文的完全相同;第 37 个字节表示在响应报文中所包含的信息

```

                                e9 03 01 01 98 ab
01 02 01 00 00 00 00 00 00 00 34 12 78 56 00 00
00 00 00 00 00 00 c9 5f 91 4e 0d 9a 32 4d a6 83
f0 9b 40 24 31 6d 8a 1c 8a 1c 00 00 00 00 00 00
00 00 00 00 00 00 01 3a c1 98 86

```

图 1 PPLive 节点列表请求报文的应用层信息

e9 03 02 00 98 ab
01 02 02 00 00 00 00 00 00 00 34 12 78 56 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 2c da
02 02 5d 89 1a 89 1a 00 00 00 00 00 00 00 00 da
5e 92 4d f3 1b f3 1b 00 00 00 00 00 00 00 00

图 2 PPLive 节点列表响应报文的应用层信息

节点数目, 图 2 中的值为 0x2c(对应的十进制数为 44), 表明后面有 44 个节点的信息; 从第 38 个字节开始则是以 16 字节为长度单位的节点信息, 其中前 8 个字节包含了节点的地址和端口信息, 第 1 至 4 个字节表示节点的 IP 地址, 第 7 和第 8 个字节表示端口号(高位字节与低位字节内容要换位)。

2.4 P2P-TV 系统的服务器域名以及应用层特征字

如前所述, P2P-TV 系统的节点在传输视频流前, 必须执行向服务器进行注册等操作, 因此, 凡是向 P2P-TV 服务器注册的节点, 随后将可能进行视频流通信, 我们可以通过监视与具有系统域名的服务器 IP 地址集合进行了通信的主机来过滤出这些注册的节点。具有系统域名服务器的 IP 地址集合被称为服务器地址集, 这可通过解析服务器域名得到。表 1 中列出了当前流行的 P2P-TV 系统主要的服务器域名。

PPLive, PPStream, SopCast, UUSEE 主要采用 UDP 协议传输视频流, 尽管它们与服务器通信时一般采用 TCP 协议, 不过这些通信报文数量较少, 可以忽略。分析结果还表明, 这些 P2P-TV 系统握手报文(即前两个分组)应用层的前几个字节可以作为特征字(被称为应用层特征字)。表 2 列出了几种特定版本下 P2P-TV 系统 UDP 报文应用层特征字的内容。

3 实时识别 P2P-TV 方法

基于上述分析结果, 本文提出了两种能够实时

表 1 各种 P2P-TV 系统主要的服务器域名

P2P-TV 系统	P2P-TV 服务器域名
PPLive	passport.pplive.com, iptable.pplive.com, pp.pplive.com, update.pplive.com, list.pplive.com
PPStream	fds.ppstream.com, tvguide.pps.tv, vodguide.pps.tv, msg.ppstream.com, download.ppstream.com, stat.ppstream.com, notice.ppstream.com
SopCast	as1.sopserv.com, home.sopserv.com, broker.sopcast.com
UUSEE	log.uusee.com, update.uusee.com, player.uusee.com, traffic.uusee.com, home.uusee.com

表 2 几种 P2P-TV 系统应用层特征字及其位置信息(0x 表示 16 进制)

P2P-TV 系统	版本号	UDP 报文特征字	起始位置
PPLive	2.2.26.0002	0x2100	第 1 个字节
PPStream	2.3.550.1950	0x004300	第 2 个字节
SopCast	3.0.3	0xffff	第 1 个字节
UUSEE	5.9.710.2	0x0909	第 1 个字节

识别 P2P-TV 流的方法, 一种是以主动方式基于爬虫程序的视频流识别(Crawler-based Identifying Video Flows, CIVF)算法, 另一种是以被动方式基于协议特征的视频识别(Protocol characteristic-based Identifying Video Flows, PIVF)算法。

3.1 基于爬虫的识别视频流算法 CIVF

基于爬虫的视频流识别算法的基本思想是: 爬虫程序根据应用层报文特征构造请求报文并向已知节点列表服务器发送, 然后接收该服务器返回的响应报文, 并根据应用层报文特征解析出当前网络中节点的 IP 地址和端口信息, 基于此构造源/目的 IP 地址和端口号列表, 以对从链路中的分组进行比对过滤, 进而识别出相关的 P2P-TV 视频流。

在根据 CIVF 算法识别视频流的过程中, 本文设置了两个线程协调工作, 更新线程根据从节点列表服务器获取的信息更新 P2P-TV 节点集, 并删除过期的节点信息, 识别线程则提取出流经分组的源/目的 IP 地址和源/目的端口号, 比对查询 P2P-TV 节点集来判断该流是否为 P2P-TV 流, 两个线程用 PV 算法实现同步和互斥。CIVF 识别算法描述如表 3 所示。

CIVF 算法对每个分组均只需对源地址、源端口号或者目的地址、目的端口号进行匹配, 因此便于在线实现, 对主机和网络环境要求不高且方便部署。爬虫程序获得的节点信息能准确标识出 P2P-TV 流, 避免了误判的出现, 但由于无法保证节点信息的完整性, 因而无法消除识别的漏判率。CIVF 同时也存在着以下几点不足: 首先, P2P-TV 系统的请求报文格式或交互规则比较复杂, 并且不同频道的请求报文内容不相同, 构造请求报文和解析响应报文的困难较大。其次, 爬虫程序必须运行一段时间后才能得知刚加入系统的节点信息, 故该识别算法有一定滞后。最后, 节点退出系统时, 节点信息服务器需要延迟一段时间才能删除该节点在服务器缓存中的记录, 这些过期的节点信息会对查找效率带来一定的影响。

3.2 基于协议特征视频流识别算法 PIVF

为了弥补 CIVF 算法的不足, 本文提出了一种

表 3 CIVF 识别算法描述

```

mutex: semaphore
procedure update;           //更新线程
begin
    向节点列表服务器发送请求报文;
    接收并解析响应报文相关信息;
    P(mutex);
    更新 P2P-TV 节点集;
    V(mutex);
end;
procedure identify;        //识别线程
begin
    if (分组为 TCP 报文)丢弃并返回;
    提取被捕获分组的源和目的 IP 地址、端口号;
    P(mutex);
    if P2P-TV 节点集包含源 IP 地址和源端口号 then
        begin
            将分组标识为 P2P-TV 分组;
            V(mutex);
        end
    else
        begin
            if P2P-TV 节点集包含源 IP 地址和源端口号 then 将分
            组标识为 P2P-TV 分组;
            else 将分组标识为非 P2P-TV 分组;
            V(mutex);
        end;
    end;
end.

```

基于协议特征的两阶段识别视频流算法 PIVF。PIVF 的基本思想如下：第 1 阶段针对 P2P-TV 节点必须访问服务器地址集的行为，记录下那些与服务器地址集内服务器通信的 IP 地址，并剔除非 P2P-TV 应用流分组；第 2 阶段对余下的流的握手报文（即第 1 个分组）比对是否具有应用层特征字的特征，如果相匹配则将该流识别为 P2P-TV 视频流。为了便于描述 PIVF 算法，进行如下定义：

定义 1 IP^s, IP^d 为分组的源和目的 IP 地址， $port^s, port^d$ 为分组的源和目的端口号， $proto$ 为分组的运输层协议类型， pl 为分组的应用负载。

定义 2 $F = \{(sIP, dIP, sport, dport, protocol, type)\}$ 为流表，它存储了每条经过探针的流的五元组及其所属应用类型的信息，初始时 $F = \emptyset$ 。

定义 3 $P_x = (IP_x, type_x)$ 为活跃节点，它表明地址为 IP_x 的主机正在运行类型为 $type_x$ 的 P2P-TV。

定义 4 $N = \{P_i | P_i \text{ 为识别出的活跃主机地址}\}$ 为活跃主机地址集，它记录了所有正在运行 P2P-TV 的主机的 IP 地址，初始时 $N = \emptyset$ 。

定义 5 $I_d = \{(IP', type) | IP' \text{ 的域名为 } d \wedge d \text{ 所属 P2P-TV 类型为 } type\}$ 表示域名为 d 的服务器所对应的所有 IP 地址集合。

定义 6 $S = \bigcup_{d \in D} I_d$ 为所有服务器地址和类型的集合，其中 D 为 P2P-TV 应用服务器的域名集合。在开始进行检测之前 S 需要进行初始化，初始化过程是对所有已知的 P2P-TV 服务器域名进行 DNS 解析，并将解析出的 IP 地址添加到 S 中。

定义 7 C_{type} 表示类型为 $type$ 的 P2P-TV 应用的应用层特征字符串。

定义 8 \triangleleft 运算：判断前一个元组出现在后一个集合中，如果存在则返回最后一个元素的值，否则返回 0， \triangleleft 运算的优先级为最高。

PIVF 算法的描述如表 4 所示。

表 4 PIVF 算法的描述

```

输入：分组
输出：P2P-TV 类型，如果为 0 表示非 P2P-TV
if ( tmp = (IPs, IPd, ports, portd, proto) < F )
    return tmp;
if ( proto == TCP ) {
    if ( tmp = IPd < S ) N = N ∪ (IPs, tmp);
    F = F ∪ (IPs, IPd, ports, portd, UDP, 0) ∪ (IPd, IPs, portd,
        ports, UDP, 0);
    return 0;
}
else if ( proto == UDP ) {
    if ( tmp = (IPs < N | IPd < N) ) {
        if ( pl == Ctmp ) {
            F = F ∪ (IPs, IPd, ports, portd, UDP, tmp) ∪ (IPd, IPs,
                portd, ports, UDP, tmp);
            return tmp;
        }
    }
    F = F ∪ (IPs, IPd, ports, portd, UDP, 0) ∪ (IPd, IPs, portd,
        ports, UDP, 0);
    return 0;
}

```

分析 PIVF 算法可知，该算法对每条流的第 1 个分组进行处理后即可判断出该流是否为 P2P-TV 流，实时性强，因为它根据主机的通信过程过滤了绝大部分不可能为视频流的 P2P-TV 流，并避免了其他几类 P2P 应用识别方法中可能出现的漏判的情况，同时仅利用 2 至 3 个字节就可快速识别出 P2P-TV 流，此外该算法具有很高的效率和很强的扩展性。首先，在进行应用层特征字匹配前，通过查询活跃主机地址集能够过滤掉不可能进行 P2P-TV 视

频流通信的那些节点产生的流量，使识别效率大幅提高。其次，在进行行为特征检测时，已经限定了可能运行的P2P-TV类型，匹配仅需对这几种应用协议的特征字进行。即使P2P-TV类型有所增加，利用行为特征过滤的步骤也将保证该算法的耗费增加幅度较小。

4 试验及结果分析

为了测试CIVF和PIVF算法的性能，在Linux平台上编制了相关程序，试验主机的CPU主频为2.92 GHz，内存为2 GB，同时为了对比分析算法的性能，我们还在测试这些算法的同时运行了TcpDump工具俘获所有网络流量，并对俘获的分组进行人工离线分析结果，与试验结果进行对比。

4.1 CIVF 算法试验及结果

我们在CERNET上对CIVF算法进行了在线试验，试验时间为2010年5月28日20点至20点30分。表5中的“实际数据”即为人工分析TcpDump俘获数据的结果，而“识别数据”即为CIVF算法分析的结果。

表5 CIVF 算法试验结果

PPLive 数据	IP 地址 数量	流数量	分组数量	字节数量
实际数据	6204	12175	1121520	684223453
CIVF 识别 数据	5115	11217	904836	597596191

从表5可见，采用CIVF算法能够识别出大部分PPLive视频流量，而未识别出的流主要因为有部分节点的信息未能获取到，主机与服务器之间的信令报文也无法识别，以及爬虫程序滞后发现地址期间漏掉的视频流报文。当节点的信息过期时，其他节点向其发送连接请求后将得不到响应报文的回复。此外，从表5中发现两者的分组数量差距较大，而两者间的字节数量差距相对较小，这是因为视频流分组通常要比信令分组大的缘故。

为了测试识别方法的时延大小，我们用一台试验机运行PPLive客户机程序，同时用另一台主机运行CIVF程序，并记录试验机运行PPLive至被CIVF程序发现的时间差，以此作为识别的时延，该试验共进行了50次，图3中显示了识别时延的平均结果。

从图3可见，采用CIVF算法的时延约为400 ms，这说明该算法能够满足P2P-TV流识别的实时性要求。其中出现的一次大于5000 ms的时延是因为测试中节点的杀毒软件升级而导致，可忽略此次

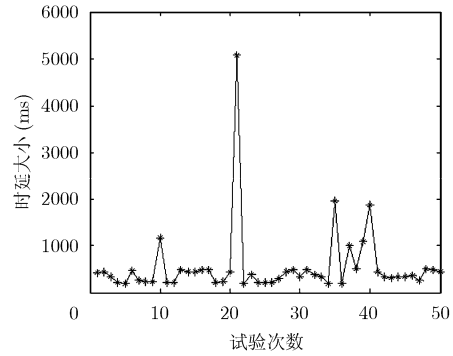


图3 CIVF 算法的时延

测试结果。

为了测试节点退出PPLive系统后其信息残存时间，我们分别在PPLive退出后的每15 s运行一次CIVF程序，来判断服务器清除退出节点信息所需要的时间。试验表明，在节点退出PPLive后至少270 s内，该节点的信息还没有被所有服务器完全清除掉。

4.2 PIVF 试验及结果

(1)准确性测试 在CERNET上采集了40 GB的样本数据，采用PIVF算法对样本数据进行了处理，表6显示了PIVF算法的准确性试验结果，同时与文献[4]中所提出的Abacus算法进行了对比。表中对每种算法都列出了正确识别该种P2P-TV应用和该种P2P-TV应用被错误地识别为其他应用的比率。其中，PPLive被PIVF算法和Abacus正确识别的比率分别是94.8%和81.66%；PPLive没有被PIVF算法误判为PPStream, SopCast和UUSee，但却以9.55%的比率被Abacus算法误判为SopCast。PIVF和Abacus算法对SopCast的试验有与对PPLive相似的结果。PIVF对于PPStream和UUSee都有非常好的试验结果，而Abacus没有进行相关试验。

由此可见，由于PIVF算法对应用层字符串进行了匹配，使它对这4种P2P-TV系统的正确识别率都达到了很高的水平，识别率最低的PPLive也接近95%，误判率则保持为0。而Abacus算法仅采用P2P-TV的行为特征进行检测，算法识别率不高，误判率也较大。

(2)效率测试 由于只需对每条流的第1个分组进行匹配即可判断出整条流是否为P2P-TV流，而且由于在进行应用协议特征字匹配之前的主机行为特征检测已经确定了协议的类型，不需要对所有特征字都进行匹配，因此PIVF算法的识别效率较高，而且扩展性也较强。图4显示了随P2P-TV类型的逐渐增加时，PIVF算法对样本数据进行处理所需的

表 6 PIVF 算法的准确性试验结果

P2P-TV 系统	PPLive		PPStream		SopCast		UUSee	
	PIVF	Abacus	PIVF	Abacus	PIVF	Abacus	PIVF	Abacus
PPLive	94.8%	81.66%	0	/	0	9.55%	0	/
PPStream	0	/	99.9%	/	0	/	0	/
SopCast	0	3.76%	0	/	99.9%	89.62%	0	/
UUSee	0	/	0	/	0	/	99.7%	/

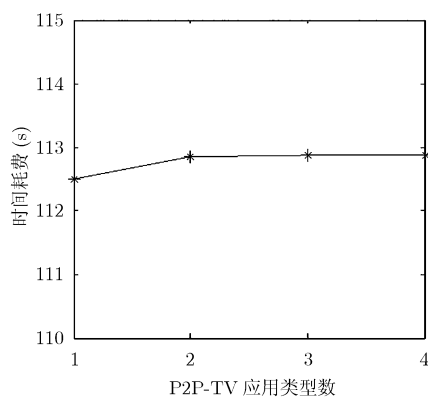


图 4 PIVF 识别方法的时间耗费

时间耗费。

从图 4 可以看出, PIVF 算法对 40 GB 大小的样本数据进行处理所需的时间耗费不超过 113 s, 具有较快的识别速度, 在 1 Gbps 以上的网络带宽环境下该算法也能够实时识别出网络中的 P2P-TV 流。而随着 P2P-TV 类型不断增加, PIVF 算法的时间耗费增长也非常缓慢, 因此它的扩展性较强。

5 结束语

本文以 PPLive 为代表分析现有 P2P-TV 体系结构、通信过程、报文结构以及系统特征, 进而提出了两种实时识别网络中的 P2P-TV 视频流的方法: 实时的基于爬虫的识别视频流 CIVF 算法和实时的基于协议特征的识别视频流 PIVF 算法。在因特网环境的试验分析的结果表明, CIVF 算法具有便于实现但识别率不够高、识别滞后且节点信息残存时间较长的特点; PIVF 算法则具有准确率较高、识别速度较快和扩展性强的特点。

参考文献

- [1] Claffy K. Internet Traffic Characterization [M]. San Diego: University of California, 1994: 55-57.
- [2] PPLive. <http://www.pplive.com>. 2008-3-20.
- [3] Spoto S, Gaeta R, Grangetto M, *et al.* Analysis of PPLive through active and passive measurements [C]. IEEE International Symposium on Parallel and Distributed

Processing, Rome, Italy, 2009: 1-7.

- [4] Valenti S, Rossi D, Meo M, *et al.* Accurate, fine-grained classification of P2P-TV applications by simply counting packets [C]. First International Workshop on Traffic Monitoring and Analysis, Aachen, Germany, 2009: 84-92.
- [5] Ciullo D, Garcia M A, Horvath A, *et al.* Network awareness of P2P live streaming applications: a measurement study [J]. *IEEE Transactions on Multimedia*, 2010, 12(1): 54-63.
- [6] Liu Yao, Guo Lei, Li Fei, *et al.* A case study of traffic locality in Internet P2P live streaming systems [C]. 29th IEEE International Conference on Distributed Computing Systems, Montreal, Quebec, Canada, 2009: 423-432.
- [7] PPStream. <http://www.ppstream.com>. 2009-5-10.
- [8] SopCast. <http://www.sopcast.cn>. 2009-5-10.
- [9] UUSee. <http://www.uusee.com>. 2009-5-10.
- [10] Sen S, Spatscheck O, and Wang D. Accurate scalable in-network identification of P2P traffic using application signatures [C]. Proceedings of the 13th International Conference on World Wide Web, New York, 2004: 512-521.
- [11] Karagiannis T, Papagiannaki K, and Faloutsos M. BLINC: multilevel traffic classification in the dark [C]. ACM SIGCOMM Conference, Philadelphia, USA, 2005: 229-240.
- [12] Constantinou F and Mavrommats P. Identifying known and unknown P2P traffic [C]. Proceedings of Fifth IEEE International Symposium on Network Computing and Applications, Cambridge, MA, USA, 2006: 92-102.
- [13] McGregor A, Hall M, Lorier P, *et al.* Flow clustering using machine learning techniques [C]. Passive and Active Network Measurement 2004, LNCS 3015: 205-214.
- [14] Zander S, Nguyen T, and Armitage G. Automated traffic classification and application identification using machine learning [C]. Proceedings of the IEEE 30th Conference on Local Computer Networks, Sydney, 2005: 250-257.
- [15] Yong Liu, Yang Guo, and Chao Liang. A survey on peer-to-peer video streaming systems [J]. *Peer-to-Peer Networking and Applications*, 2008, 1(1): 18-28.

胡超: 男, 1984年生, 博士生, 研究方向为 P2P 系统、网络测量等。

陈鸣: 男, 1956年生, 教授, 博士生导师, 研究方向为计算机网络体系结构、分布式计算、网络拓扑发现、网络管理等。

许博: 男, 1980年生, 博士, 研究方向为分布式计算、网络测量等。