

## 基于爬行器的大规模 P2P IPTV 测量\*

姜志宏<sup>+</sup>, 王 晖, 樊鹏翼

(国防科学技术大学 信息系统与管理学院, 湖南 长沙 410073)

### Crawler-Based Measurement of Large Scale P2P IPTV Systems

JIANG Zhi-Hong<sup>+</sup>, WANG Hui, FAN Peng-Yi

(College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: jonathan.jiang@263.net

Jiang ZH, Wang H, Fan PY. Crawler-Based measurement of large scale P2P IPTV systems. *Journal of Software*, 2011,22(6):1373-1388. <http://www.jos.org.cn/1000-9825/3849.htm>

**Abstract:** In order to gain insight on the topology characteristics of P2P IPTV systems and their online user behavior characteristics, This paper develop and deploys a multiprotocol P2P IPTV network crawler, called TVCrawler, which enables users to launch an in-depth measurement and comparative research for several well-known P2P IPTV systems, including PPLive, PPStream, and UUSee. This paper presents results from experiments and research efforts on these large-scale P2P IPTV overlay graphs. Major findings include 1) more than 50% of online users are unreachable because they lie behind NAT or firewall; 2) fluctuation range of churn increases with the increase of the population of channel, and there exists a power-law upper bound for fluctuation range of churn; 3) session length of peer follows a stretched exponential distribution; 4) while the in-degree of PPLive follows the power-law distribution with exponential cutoffs, the in-degree of PPStream appears to have multiple separate power-law regimes with different exponents, and the in-degree of UUSee approaches Weibull distribution. 5) All of measured P2P IPTV overlay graphs appear as a disassortative mixing; 6) small-world networks are ubiquitous in measured P2P IPTV systems. 7) the PPLive overlay graph is clustering, but the others are not; 8) all of measured P2P IPTV overlay graphs have robustness similar to that of a power-law graph, but PPLive is more vulnerable to malicious attacks and more robust to random failures than the other. These research and findings Will be good for designing and optimizing of P2P IPTV protocols, but also for monitoring, directing, and dominating the P2P IPTV system.

**Key words:** peer-to-peer; IPTV; crawler; network measurement; user behavior; topology characteristics

**摘 要:** 为了解大规模 P2P IPTV 系统中的用户行为特征和拓扑结构特征等内在信息,开发和部署了一个多协议 P2P IPTV 爬行器 TVCrawler,对 3 个主流的 P2P IPTV 系统——PPLive,PPStream 和 UUSee 进行了大量的主动测量,并对 P2P IPTV 系统中的用户行为和网络拓扑特征进行了分析和比较.主要发现包括:1) P2P IPTV 系统的频道在线人数中,有一半以上位于不可达的 NAT 或者防火墙后面;2) 节点动态性的波动范围随频道人数的增加而增加,但是其取值范围具有幂律上限;3) 节点会话长度符合广延指数分布;4) PPLive 的入度分布属于具有指数截断的幂次分

\* 基金项目: 国家高技术研究发展计划(863) (2008AA01Z407)

收稿时间: 2009-12-03; 修改时间: 2010-01-20; 定稿时间: 2010-03-11

布,PPStream 的入度表现为某种分段幂律函数,UUSee 的入度接近威布尔分布;5) P2P IPTV 系统都是异配网络;6) P2P IPTV 系统都表现为小世界网络;7) PPLive 网络具有聚类特征,而 PPStream 和 UUSee 则不存在明显聚类特征;8) 3 个系统都表现出类似于无标度网络的鲁棒性特征,而与其他两个系统比较时,PPLive 具有更高的故障容错性和更明显的攻击脆弱性.这些测量研究和发现不仅有助于设计出更符合真实网络应用环境的系统或协议,也是实现对 P2P IPTV 进行监测、引导、控制等方面的重要依据和基础.

关键词: 对等网络;网络电视;爬行者;网络测量;用户行为;拓扑特征

中图法分类号: TP393 文献标识码: A

随着基于 Peer-to-Peer(P2P)技术的网络电视的广泛流行,大量成功的商业 IPTV 系统迅速发展起来,IPTV 用户人数也快速膨胀.MRG 统计数据 displays,2008 年全球 IPTV 用户的人数为 2 040 万,到了 2009 年增长为 2 670 万,到 2013 年预计将增长为 8 100 万.通常,这些 IPTV 系统(如 PPLive<sup>[1]</sup>,PPStream<sup>[2]</sup>和 UUSee<sup>[3]</sup>等)提供数百个网络电视频道,吸引了数百万的用户在线观看.而根据节目内容的热门程度不同,其单个频道的同时在线人数则从几十人到数千,甚至数万不等,说明了基于 P2P 技术的网络电视已经具备了提供大规模网络视频服务的技术能力.这些系统的成功,得益于两个重要的 P2P 技术:一是采用基于数据驱动(data-driven)的视频流数据分发机制,二是采用类 Gossip 方式维护用于数据分发的覆盖网络.显然,P2P 覆盖网络拓扑在 P2P 网络电视的成功中起到了重要作用.而同时,P2P IPTV 网络协议的设计通常面向节点个体行为,对于大规模 P2P IPTV 网络的拓扑和整体行为仍然知之甚少.

目前,P2P IPTV 测量方面的相关研究一般采取以下 3 种方法:一是基于嗅探的被动测量方法,即利用网络嗅探软件来捕捉特定环境下的 IPTV 客户端的通信流量,并对通信流量进行分析和统计.基于嗅探的被动测量方法通常以流量相关的系统局部特征为侧重点,包括本地流量统计、上传/下载带宽使用情况和数据包大小分布等,相关研究一般是在特定实验环境下搭建测试平台.虽然得到的统计数据是基于真实流量,但其实验规模限制了结论的普适性;二是基于通信协议的主动测量方法,即通过对被测系统的通信协议进行分析和理解,设计一个 P2P 协议爬行者来主动探测系统和收集信息.由于目前的 P2P IPTV 系统大多采用私有协议,因此,基于通信协议的主动测量方法的难点在于需要结合 P2P IPTV 系统的基本设计原理,对 P2P IPTV 系统采用逆向工程的方法,分析和得到特定 P2P IPTV 的数据通信模式,包括通信协议的格式和语义信息,从而设计和实现协议爬行者;三是 IPTV 系统提供商参与的测量方法,即研究机构与 P2P IPTV 系统提供商展开合作,系统提供商在其客户端软件中增加数据采集接口,或在整个系统中部署数据采集方案,供研究者进行数据的收集和分析使用.这一类研究方法由于要求有比较好的合作条件,因此相对比较困难.

## 1 相关研究及本文工作

P2P 网络的测量起源于对 P2P 文件共享系统的测量研究.早在 2002 年,Saroiu 等人<sup>[4]</sup>针对当时流行的两种 P2P 文件共享系统 Napster 和 Gnutella 分别设计了爬行者进行测量,发现在 P2P 文件共享系统中,不仅在节点带宽、延迟和可用性方面存在巨大的异构性,而且对等节点的行为仍然表现出明显的类客户端(client-like)或类服务器(server-like)的特征;他们的研究表明,Gnutella 网络的抗攻击性与幂律网络一致,即对于随机攻击鲁棒性较强,而对于选择性攻击鲁棒性较弱.Ripeanu<sup>[5]</sup>也通过爬行者对 Gnutella 进行了测量研究,他认为,Gnutella 网络的节点度分布符合两段式分布,当节点度大于 10 时为幂律分布,当小于 10 时近似于常数分布;Stutzbach 等人<sup>[6]</sup>改进了 Gnutella 爬行器的爬行速度,得到了更为精确的测量数据.其研究表明,Gnutella 网络具有小世界特征,但是度分布则表现得更为复杂,尤其在度数为 30 左右时表现出一个峰值;此外,Gnutella 网络在面对随机攻击和选择性攻击时都表现出较强的鲁棒性.与协议公开的 Gnutella 的主动测量研究不同,对于其他采用私有/加密协议的 P2P 文件共享系统,进行主动测量存在较大困难,相对研究也较少.Liang 等人<sup>[7]</sup>对采用私有协议的 KaZaA 系统的网络行为进行了协议分析,设计出专门的 Crawler,对 KaZaA 网络的结构以及动态特性等进行了测量研究.

P2P IPTV 的发展稍晚于 P2P 文件共享应用.与 P2P 文件共享系统相比较,P2P IPTV 系统的数据传输具有严

格的实时性要求,因而其网络具有更强的动态性特点.加上主流 P2P IPTV 系统的通信协议大部分都是私有、未公开的,这些都增加了测量研究的难度.CoolStreaming<sup>[8]</sup>是最早在互联网上大规模部署的 P2P IPTV 系统,Zhang 等人<sup>[9]</sup>在系统中增加了一个 P2P 报告机制用于采集测量数据,对 CoolStreaming 的在线人数、用户行为和用户的地理分布等进行了分析和研究.Li 等人<sup>[10]</sup>也利用上述报告机制对 CoolStreaming 运行日志数据进行了分析,研究了用户类型和分布、系统动态性和资源分布等.Wu 等人<sup>[11]</sup>与悠视网合作,通过在 UUSee 客户端中增加日志和报告功能,采集了大量实际运行数据,对 UUSee 系统的在线人数、用户行为、用户分布和 UUSee 网络拓扑特征以及流量特征进行了统计分析.这些研究都是通过与 P2P IPTV 系统提供商合作进行的,或者研究者本身就是系统开发者.而对于不公开协议的 P2P IPTV 系统的测量研究,主要还是通过被动嗅探和主动爬行两种方式.Hei 等人<sup>[12,13]</sup>通过主动爬行的方式研究了 PPLive 的用户行为,包括频道在线人数和系统在线人数变化、用户动态性等,认为 PPLive 用户与传统电视用户的收看时间一致;此外,他们还结合嗅探测量的方式研究了 PPLive 的流量特征和视频服务质量.Vu 等人<sup>[14,15]</sup>也设计了 PPLive 协议爬行器,对 PPLive 进行了测量研究,发现 PPLive 的频道在线人数变化具有时间敏感(time-sensitive)、自重复(self-repeated)和事件相关(event-dependent)的特点;也证实了 PPLive 网络动态性更强于 P2P 文件共享系统,而其节点平均度与频道在线人数无关;并且认为,PPLive 覆盖网络对随机性攻击和选择性攻击的鲁棒性几乎一致,尤其是小规模 PPLive 网络接近于随机网络.Ali 等人<sup>[16]</sup>则采用被动嗅探的方式研究了 PPLive 和 SOPCast 两个系统的网络资源利用率、节点交换数据的局部性以及数据分发的稳定性等流量特征.Silverston 等人<sup>[17]</sup>采用嗅探的方式研究和比较了 PPLive,PPStream, SOPCast 和 TVAnts 这 4 个 P2P IPTV 系统的平均包大小、节点吞吐量、节点生存周期等特征.

我们采用协议爬行器的方式对 PPLive,PPStream 和 UUSee 这 3 个 P2P IPTV 网络进行了主动测量,比较和研究了其各自的用户行为特征,并且对其覆盖网络的拓扑特征进行了分析和研究.从目前已知的研究文献看,PPLive 的测量研究相对较多,而本文所设计的爬行器能够对多个 P2P IPTV 网络实施爬行测量,同时本文也是第一个对多个 P2P IPTV 网络开展主动测量和分析比较的研究工作.本文第 2 节介绍了 P2P IPTV 的主要协议框架以及多协议爬行器的设计思想.第 3 节介绍了多协议爬行器的实验运行环境和数据采集方法以及测量数据集的基本情况.第 4 节基于主动测量数据研究了各个 P2P IPTV 的用户行为特征.第 5 节对 P2P IPTV 覆盖网络的拓扑特征进行了研究.最后做出总结,并对后续相关研究工作进行了展望.

## 2 P2P IPTV 协议爬行器

### 2.1 P2P IPTV 基本原理

目前,主流的商用 P2P IPTV 系统,如 PPLive,PPStream 和 UUSee 等,都属于 mesh-pull 结构,即通过类 Gossip 方式构造和维护一个 mesh 结构的覆盖网络,同时采用基于数据驱动的方式设计视频流数据分发机制.系统中播放的视频流被分成数据块(chunk),从视频源服务器通过覆盖网络向其他节点分发,其简单流程(如图 1 所示)是:

当一个新节点加入系统时:

- (1) 首先访问频道服务器(channel server),获得当前的播放的频道列表,用户从中选择目标频道;
- (2) 新节点随后向成员服务器(tracker server)进行注册,并获得多个在线节点的信息,用于形成自己的初始邻居集.新节点加入系统后,在其生存周期内的网络活动主要分为两个类型:
- (3) 通过 Gossip 的方式更新和维护自己的邻居集,即每个节点定期与本地邻居集中的节点交换邻居列表,并根据交换结果更新本地的邻居集,以实时保持与 P2P 网络其他部分的连通性;
- (4) 每个节点与本地邻居集中的节点交换各自的视频数据缓存映像(buffer map),获得有效的视频缓存信息,并采用高效调度算法选择合适的邻居节点下载数据块.

这两种类型的网络活动贯穿于节点的整个生存周期.

P2P IPTV 协议爬行器的基本思想就是,遵循 IPTV 系统的协议流程和语义,模拟客户端与 Tracker Server 及其他节点通信,从而获取节点邻接关系并加入本地节点队列,反复迭代直至获取到足够的 P2P 覆盖网络信息.由于目前主要的 P2P IPTV 系统都是私有协议,为了获得爬行器所必需的协议语义和数据模式,需要其网络协议进

行逆向工程.我们通过网络嗅探软件 WireShark,对 PPLive,PPStream 和 UUSee 这 3 个 P2P IPTV 系统分别捕获了大量的实际通信流量,根据图 1 所示工作模式,重点分析了其中与节点邻接关系相关的部分协议,获得了图 1 中第(2)部分、第(3)部分的协议语义.与 P2P 文件共享系统的区别在于,上述 3 种 P2P IPTV 系统大量基于 UDP,而不是 TCP 连接方式来实现协议交互.另外,PPLive 和 UUSee 两个系统都采用 Pull 的方式交换邻居信息,即节点主动向邻居节点查询,邻居节点则产生一个本地邻接关系列表作为应答返回;而 PPStream 则采用了 Push 的方式,即由节点定期产生一个本地邻接关系列表,发送给随机选择的邻居节点.

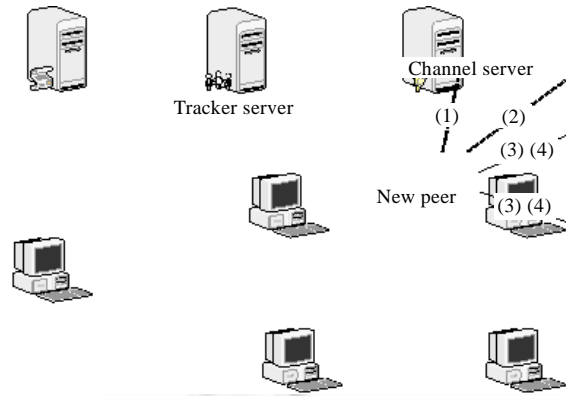


Fig.1 Primary activity in P2P IPTV communication protocol

图 1 P2P IPTV 主要通信协议过程

## 2.2 P2P IPTV多协议爬行器

针对主流 P2P IPTV 系统在软件结构和协议设计上的相似性,我们设计和实现了多协议的 P2P IPTV 爬行器——TVCrawler<sup>[18]</sup>,实现对 PPLive,PPStream 和 UUSee 这 3 种流行 P2P IPTV 系统的主动测量.TVCrawler 由爬行控制器(controller)和爬行器终端(crawler)两部分组成,采用主从结构分布式部署.其中,爬行器终端是整个系统的核心部分,通过覆盖网络爬行获取拓扑的终端视图,并将终端视图以增量方式汇总到控制器;爬行控制器根据爬行策略控制多个爬行器客户端的同步启动和终止爬行,并接收多个爬行器上传的拓扑数据,合并生成覆盖网络拓扑总表.多个爬行器终端同时运行在不同网络终端上,在控制器的控制管理下,对同一个 P2P IPTV 覆盖网络并行地爬行.

P2P IPTV 协议爬行器从开始爬行到爬行结束,得到覆盖网络拓扑总表,称为快照(snapshot),其中包含大量的在线节点以及节点之间的邻居关系,用于表征在一个短时间内 P2P IPTV 频道的覆盖网络.为了验证爬行数据的有效性,需要检验快照覆盖率,即快照中包含的节点集  $S_{ss}$  对实际在线节点集  $S_R$  的覆盖程度.快照覆盖率定义为  $P_c = P(v \in S_{ss} | \forall v \in S_R)$ .但是由于 P2P IPTV 的无中心化特点,快照覆盖率无法进行精确地量化计算,因此本文设计了一个实验进行粗略估计:

- (1) 运行爬行器采集指定频道的拓扑快照,得到快照中的节点集  $S_{ss}$ ,所花时间为  $t$ ;
- (2) 在运行爬行器的同时,启动 P2P IPTV 客户端  $C$  加入相同频道,并采用 WireShark 监测客户端  $C$  的网络通信流量,得到在  $t$  时间内所有与该客户端发生数据交换的远端节点集  $S_m$ ;
- (3) 计算概率  $P_r = P(v \in S_{ss} | \forall v \in S_m)$ .

Ganesh 等人<sup>[19]</sup>研究表明,采用基于 Gossip 的覆盖拓扑管理方式中,每个节点的邻居关系接近于随机连接,即可近似认为  $S_m$  为  $S_R$  的一个随机抽样子集,因此有  $P_r \rightarrow P_c$ .通过在不同的网络环境中运行 P2P IPTV 客户端,重复上述实验,最终得到  $P_r > 95\%$ ,说明 TVCrawler 获取的快照数据具有较高的覆盖率,可认为使用 TVCrawler 对 P2P IPTV 的覆盖网络的测量是有效的.

此外,爬行器的爬行速度也是快照质量的重要影响因素,如果爬行器爬行的时间过长,期间网络变化过大,

将导致测量视图与实际视图不一致,从而影响测量数据的准确性.在对 P2P 文件共享系统的测量中,早期的一些 Gnutella 爬行器的爬行速度约为 15K 节点/h<sup>[5]</sup>和 2.5K 节点/m<sup>[4]</sup>,而单次爬行时间长达数小时. Stutzbach<sup>[6,20-22]</sup>设计的 Gnutella 爬行器 Cruiser 能在 4 分钟以内爬行得到 300K~400K 节点,其研究还表明,当 Gnutella 爬行器速度过慢时,会给 P2P 系统测量结果引入明显的误差.P2P IPTV 爬行器一般针对单个电视频道,其在线人数规模小于 Gnutella 等 P2P 文件共享系统,但是网络动态性特征更加明显.Vu 等人<sup>[23]</sup>设计的 PPLive 爬行器针对大规模网络进行了调整,采用多个终端并行爬行,单次爬行时间根据频道规模的不同而不同,约为 3m~8m.TVCrawler 通过爬行统计数据自适应地调整爬行时长.我们的实验数据表明,在保证快照完整性的前提下,TVCrawler 对 3 种 P2P IPTV 的平均爬行速度均在 300 节点/s 左右,而获取单张快照的时长取决于网络规模,约为 30s~4m.

### 3 P2P IPTV 测量实验

作为国内的主流 P2P IPTV 系统,PPLive,PPStream 和 UUSEE 各自提供了包括地方卫视、财经、体育等内容的几十到上百个视频直播频道.与 Silverston 等人<sup>[17,24]</sup>的测量研究不同的是,我们选择了财经类电视频道  $F$  作为研究对象,而不是特定事件时期内突发出现的高流行度频道(如世界杯足球赛期间的体育频道).频道  $F$  是国内影响较大的财经类频道,频道的节目安排较为固定,长期稳定地拥有较为庞大的收看用户群.在 3 个 P2P IPTV 系统中,不仅同时提供频道  $F$  的 24 小时直播服务,而且各系统的统计数据都表明,频道  $F$  流行度经常排名在直播类频道的前三,属于稳定的高流行度频道.因此,选择频道  $F$  有利于分析研究日常的 P2P IPTV 用户行为,而不是特定事件影响下的短期用户行为.

在本文的测量实验中,TVCrawler 采用 1 台服务器运行爬行控制器,2 台网络终端运行爬行终端,测量节点的互联网出口带宽为 100M(TVCrawler 采用尽量发送方式,对网络带宽等资源要求较高,在高带宽环境中性能要优于低带宽环境).分别对 3 个 P2P IPTV 系统中的  $F$  频道进行主动测量,测量开始时间为 2009 年 11 月 9 日,实验持续 2 周.得到一系列连续的快照数据(见表 1),快照获取的时间间隔为 2m~3m,数据总容量达到 6GB.由于 3 个 P2P IPTV 系统的主要用户都来自中国,因此本文采用北京时间(GMT+8)对测量数据进行分析.

Table 1 Dataset from P2P IPTV active measurement

表 1 P2P IPTV 主动测量的结果数据集

	Number of Snapshots	Avg. interval (s)	Max interval (s)	Min interval (s)	Max size of Snapshots (number of nodes)	Min size of Snapshots (number of nodes)
PPLive	10 615	111	184	109	2 959	24
PPStream	9 202	128	286	111	28 931	16
UUSEE	10 460	112	224	111	4 576	15

同时,由于 P2P IPTV 系统中存在一些位于防火墙或 NAT(network address translation,网络地址转换)后面的不可达节点.在测量数据集中,这些节点被称为哑节点(dummy peer),它们包含在其他节点返回的邻居列表中,但是对测量请求报文不返回任何应答.与此相对应的其他节点称为响应节点(response peer).在 P2P 应用研究中,哑节点的存在对于系统性能的影响仍然是一个开放的问题.由于覆盖网络中其他节点无法主动与位于防火墙或 NAT 后面的哑节点建立连接,因而哑节点对于数据上传和转发没有贡献或很少贡献,表现在快照中则是哑节点的出度为 0.因此,本文在对快照进行网络拓扑特征的部分研究中,采用了与 Vu 等人<sup>[14,15]</sup>相似的做法,即仅对快照中的响应节点进行分析.

## 4 大规模 P2P IPTV 用户行为特征

### 4.1 频道在线人数

频道在线人数(population)是指同时停留在频道内收看节目的用户数.图 2 表示在 2009 年 11 月 9 日 18:00:00~2009 年 11 月 17 日 11:00:00 时间段内,3 个 P2P IPTV 系统中  $F$  频道的在线人数变化.可以看出:P2P IPTV 的频道在线人数随时间变化较大,被测量的  $F$  频道人数波动在 20~20000 之间;同时,频道在线人数的变化在时间上具

有一定的自重复性.与 Vu 等人认为用户倾向于在业余时间使用 PPLive<sup>[14,15]</sup>不同,图 2 中 3 个系统中  $F$  频道的主要收视时间位于工作日的白天,例如上午的 9 点~下午 16 点之间,同时在工作日中午时段在线人数会略有回落;而工作日晚上和周末时段的在线人数都保持较低水平.一个可能的原因是, $F$  作为财经类电视频道,其用户的收视兴趣与国内股票市场的营业时间有关.此外,频道在线人数的演化明显分为工作日和周末两种类型.在工作日,频道在线人数变化呈现出明显的以天为单位的周期性特征;而周末的频道在线人数虽然同样表现出以天为单位的周期性,但是最大数量和波动幅度都远低于工作日.图 2 中虚线表示快照中哑节点比例随时间的变化关系.PPLive,PPStream 和 UUSee 这 3 个系统的哑节点平均比例分别为 70.9%,56.2%和 49.8%,说明 P2P IPTV 的在线用户有一半以上均位于不可达的 NAT 或者防火墙后面.同时,哑节点所占比例的变化在时间上也具有自重复性,在工作日的白天,与在线人数高峰相对应的是哑节点所占比例最低,3 个系统都在 0.4 左右;而在凌晨时段,与在线人数低谷对应的是哑节点所占比例达到最高.

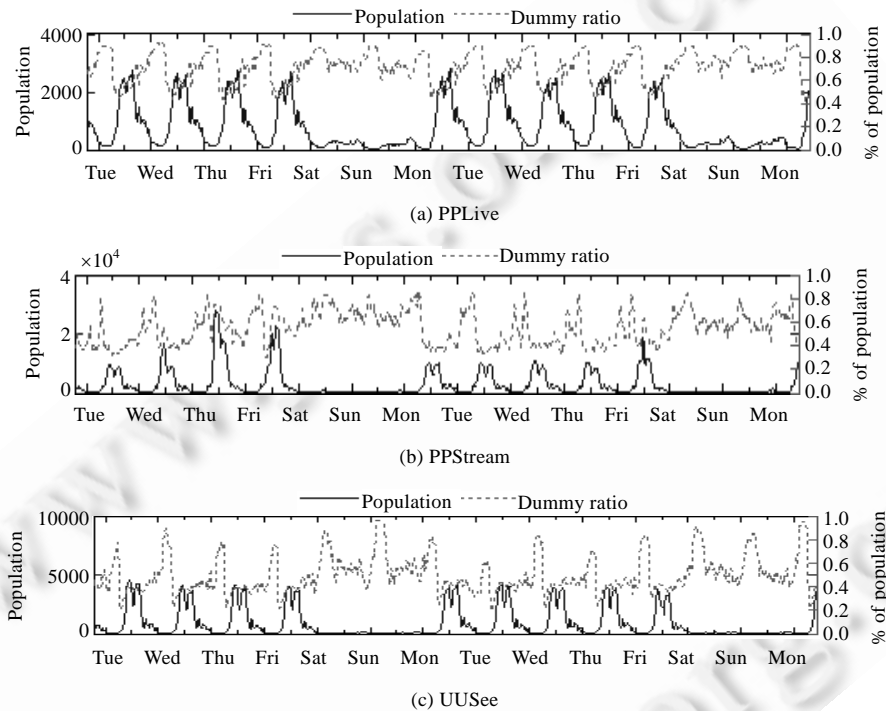


Fig.2 Statistics of population in channel  
图 2 频道在线人数统计

#### 4.2 节点地域分布

3 个 P2P IPTV 系统都是面向整个互联网提供服务,但是被测量的财经类电视频道  $F$  仅提供中文节目,而且节目内容侧重于中国国内的财经相关信息,因而频道在线用户主要来源于中国国内.图 3 表示 3 个系统中在线节点的地域分布.图 3(a)表示 3 个 P2P IPTV 系统的  $F$  频道中,国内用户所占比例随时间的演化情况.可以看出,由于节目语种、内容和时差等关系, $F$  频道的主要用户群位于中国大陆地区,尤其在白天时间段内,来自中国大陆地区用户数常保持在 95% 以上.图 3(b)~图 3(d)分别表示在 3 个系统的  $F$  频道中,来自中国大陆地区以外的用户组成,可以看出,此类用户主要来源于亚洲的其他国家和地区、北美和欧洲.PPStream 和 UUSee 的非中国大陆用户中,来自亚洲其他国家和地区的比例接近 50%,PPLive 的非中国大陆用户中则有一半来自北美.此外,与其他两个系统不同的是,PPStream 还拥有 2% 左右的非中国大陆用户来自大洋洲.

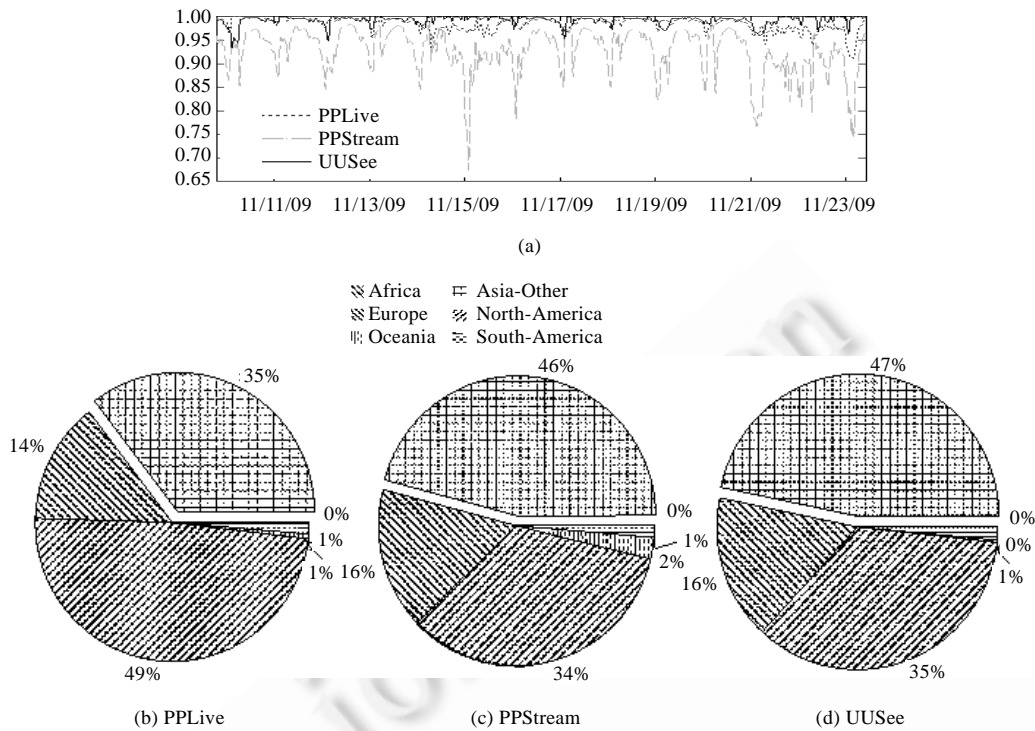


Fig.3 Statistics of population geographic distribution in channel

图 3 频道在线人数地域分布统计

### 4.3 节点动态性

节点动态性是指在 P2P IPTV 网络中,节点频繁地加入或离开 P2P 网络的行为.节点动态性对于 P2P 网络的传播动力学特征有着重要的影响,本文侧重于研究 P2P IPTV 网络动态特征的两个度量值:churn 变化率(churn ratio)和节点的会话时长(session length).

churn 变化率反映出单位时间内节点加入和离开导致网络大小变化的速度,Kumar 等人<sup>[25]</sup>定义 Churn 变化率见公式(1):

$$CR = \frac{N_{arr} - D_{dep}}{N_T} \quad (1)$$

其中, $N_{arr}$ 是单位观测时段  $T$  内加入节点的总数, $D_{dep}$ 是  $T$  内离开节点的总数, $N_T$ 是开始时刻节点的总数.在本文研究中,以分钟作为单位观测时长,即  $CR$  为每分钟内的 churn 变化率.图 4(a)表示 3 个系统中的  $CR$  时间演化规律,可以看出, $CR$  在时间上的自重复特点与频道在线人数相似,通常在夜间和周末  $CR$  取值范围较小;而在工作日的白天, $CR$  取值范围较大.图 4(b)和图 4(c)采用双对数尺度,分别给出了  $CR \geq 0$ ,  $CR < 0$  两种情况下与频道在线人数的关系,显示出  $CR$  与频道在线人数呈现出正相关性.随着频道在线人数的变化, $CR$  的取值范围具有较为明显的幂律上限;而且从图中还可以看出,在 3 个 P2P IPTV 系统中,对应的幂律上限具有相同或者相近的幂指数,通过非线性最小二乘拟合得到幂指数  $\gamma=0.067$ .  $CR$  取值范围的幂率上限说明,随着频道人数增长,节点动态性也随之增长,但是由于增长的非线性特点,使得整个网络的相对动态性却趋于下降.一个可能的原因是,频道在线人数是与频道节目相关的,当频道内容更加让用户感兴趣的时候,不仅用户数量保持在较高的水平,而且有更高比例的用户在较长时间内一直停留在频道内,从而使得网络的相对动态性保持在一个更小的范围内波动.

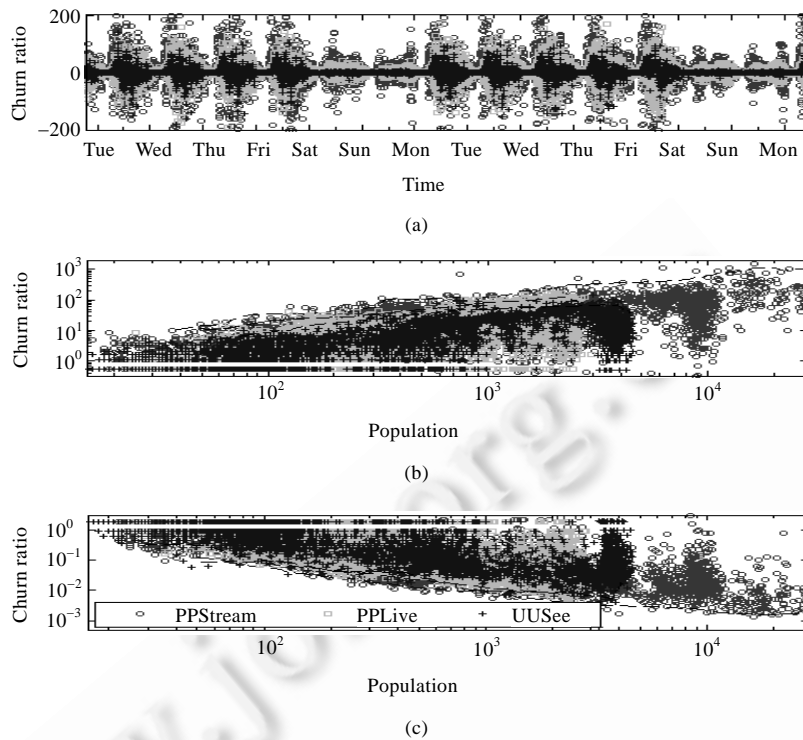


Fig.4 Evolution of churn ratio in channel

图4 频道 churn 变化率的演化

P2P IPTV 节点会话长度是指节点一次加入频道至离开频道的的时间,反映了节点一次停留在频道中的时间长度.我们的测量数据是一系列在时间上连续的快照集  $S=\{s_1,s_2,\dots,s_N\}$ ,  $s_i$  表示第  $i$  张快照.对测量得到的每一个节点  $p$ ,查找对应的快照子序列  $S_p=\{s_n,s_{n+1},\dots,s_{n+i}\}$ ,表示在第  $n$  张快照中首次发现节点  $p$ ,并在随后的连续  $i$  张快照中都包含节点  $p$ ,直至快照  $s_{n+i+1}$  该节点消失.将快照序列  $S_p$  的总时长  $T_s$  作为节点的会话长度.对于退出频道后再次加入的节点,同样当作新节点加入处理.实际上,由于存在快照测量时长和测量间隔时长,  $T_s$  只是节点实际会话长度的近似值,偏差最大时,约为连续两张快照形成的间隔时间.

测量和研究节点会话长度的分布,有助于了解用户加入和离开频道的动态行为,并可用于用户动态行为的精确建模.Vu 等人<sup>[23]</sup>在研究 PPLive 的测量数据时认为,其中的节点会话长度符合指数分布.图 5(a)是我们测量的 3 个系统节点会话长度  $T_s$  的累积概率分布  $P(T_s)$ ,其中的虚线表示分别进行指数拟合的结果.可以看出,3 个系统的会话时长分布并不符合指数分布,其衰减速度虽然慢于指数分布但是快于幂律分布,从形状上看具有广延指数分布(stretched exponential distribution,简称 SED)的特征.

广延指数分布的累积分布函数如公式(2):

$$P(x) = e^{-\left[\frac{x}{x_0}\right]^c} \quad (2)$$

其中,  $c$  和  $x_0$  是常数.其中:后者是归一化参数,前者为广延参数(或形状参数),是决定分布形状的关键参数.即当  $c=1$  时表示指数分布,当  $c \rightarrow 0$  时表示幂律分布.  $c$  在 1 和 0 之间时,分布介于指数分布和幂律分布之间.对公式(2)两边取自然对数,则有

$$x^c = -x_0^c \ln P(x) \quad (3)$$



因此,广延指数分布分布在 SE 坐标系( $x^c \sim \ln P(x)$ )下是一条直线.图 5(b)是 3 个系统节点会话长度累积概率分布的 SED 拟合结果,从子图中可以看出,3 个系统的会话长度分布在 SE 坐标系下都具有明显的线性相关性,说明节点会话长度的累积概率分布对 SED 的符合程度较高.我们通过最小二乘拟合得到广延参数  $c$  在 3 个系统中的值分别为  $c_{PPLive} \approx 0.36$ ,  $c_{PPStream} \approx 0.46$  和  $c_{UUSEe} \approx 0.60$ ,说明会话长度分布均介于幂律分布和指数分布之间.图 5(b)主图中的虚线表示节点会话长度累积分布的 SED 拟合曲线,其拟合程度远高于指数分布,进一步说明广延指数分布能够更好地拟合 3 个系统节点会话长度的实际分布.

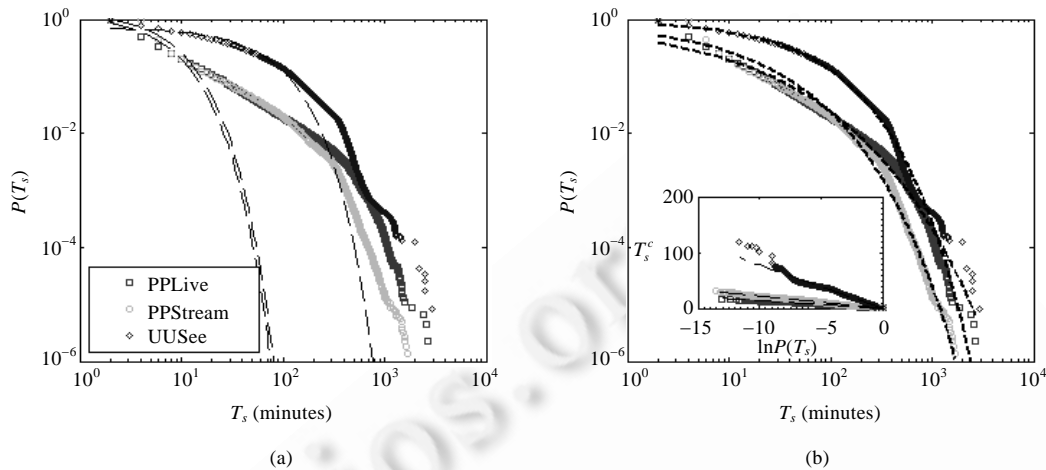


Fig.5 Session length distribution of peers in channel

图 5 节点会话长度分布

Saroiu 等人<sup>[26]</sup>对 Napster 和 Gnutella 的研究表明,P2P 文件共享系统的节点会话长度平均 60m,80%的节点小于 180m.从图 5 可以看出:PPLive 和 PPStream 有 80%的节点会话长度在 10m 以下,平均会话长度为 12m 左右;UUSEe 平均会话时间略长,为 46m,其中 80%的节点小于 70m.说明 P2P IPTV 系统中的节点会话长度显著低于 P2P 文件共享系统,这与 Vu 等人<sup>[23]</sup>的“PPLive 用户具有 impatient 特点”的研究结论是一致的,反映了 IPTV 用户和文件下载用户在使用习惯上的不同.

## 5 大规模 P2P IPTV 网络拓扑特征

### 5.1 平均度与度分布

在测量 P2P IPTV 平均度时,考虑到哑节点由于位于 NAT 或防火墙后面,在快照数据中表现为出度未知或者为零,我们仅对快照中响应节点计算平均度,在图 6 中表示 3 个系统的平均度对比.在 3 个系统中,响应节点的平均度并不相同,PPLive 的平均度均值最小,为 11.9;UUSEe 平均度均值最大,为 36.4;PPStream 平均度均值则为 25.6.不同的平均度可能体现了各系统在邻居节点管理的实现(如参数设置)上的差异.同时还清楚表明,当在线人数较大(>200)时,3 个系统平均度的取值虽然具有波动性,但是波动变化范围较小,而且平均度大小与频道在线人数取值无关.这与 Vu 等人<sup>[23]</sup>单独对 PPLive 的测量分析结果是一致的.

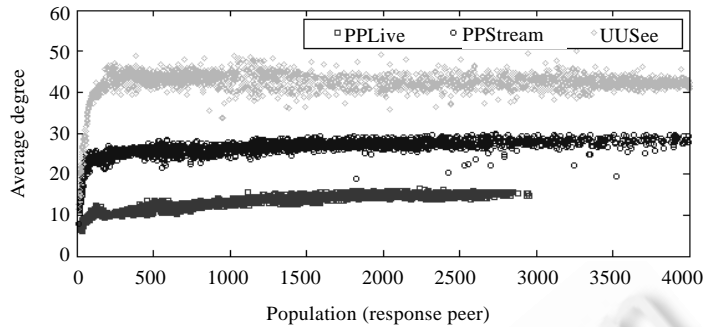


Fig.6 Correlation between average degree for response peers and online population in channel

图 6 响应节点平均度与频道在线人数的关系

早期研究表明,P2P 网络的度分布具有幂次规律;而近期的一些研究认为,度分布具有更为复杂的特性. Stutzbach 等人<sup>[6]</sup>测量 Gnutella 网络发现,其不具备明显的幂次规律.Wu 等人<sup>[11]</sup>认为,UUSee 网络的入度分布较为复杂,而出度近似于两段幂次分布.Vu 等人<sup>[15]</sup>则没有区分入度和出度,认为 PPLive 网络度分布符合一个分段分布.其中,在低度数和一定范围内的高度数下表现为均匀分布,而在中等度数范围内表现为指数分布.但是目前对于 P2P 网络的度分布的研究仍然没有定论.

在本文的研究中,为消除测量随机性和在线人数规模的影响,我们先统计测量数据集中每一张快照响应节点的度分布,并计算其累加均值作为整个系统度分布,最终得到了 P2P IPTV 系统的入度和出度概率分布.从图 7(a)中看出,PPLive 响应节点的入度分布具有明显的幂次分布特征,通过最小二乘法进行曲线拟合得到幂指数为-1.9,而且在分布曲线的尾部还表现出指数截断(exponential cutoffs)性质<sup>[27]</sup>;PPStream 和 UUSee 响应节点的入度分布比 PPLive 相对均匀,但是都不具备幂律特征.其中,UUSee 的入度分布接近于威布尔分布,而 PPStream 的入度分布可能是某种类型的分段幂律或广延指数分布.一个有意思的发现是,PPStream 和 UUSee 在较低的节点度范围内(PPStream 中节点入度小于 20,UUSee 中节点入度小于 30)出现类似小变量饱和(saturation for small variables)现象,并且在小范围内形成了“双峰”,这种分布特征同样见于 Ripeanu 等人<sup>[5]</sup>测量的 Gnutella 覆盖拓扑中.考虑到我们的测量和分析方法中丢失了大量低入度的哑节点信息,这可能也是造成这种现象的原因之一.而在图 7(b)中,3 个系统均表现出不同程度的泊松分布特征,其中在出度为 15,30 和 50 左右,PPLive,PPStream 和 UUSee 分别表现出分布的高峰,反映了 3 个系统在邻居节点集初始大小等参数上设计的不同.

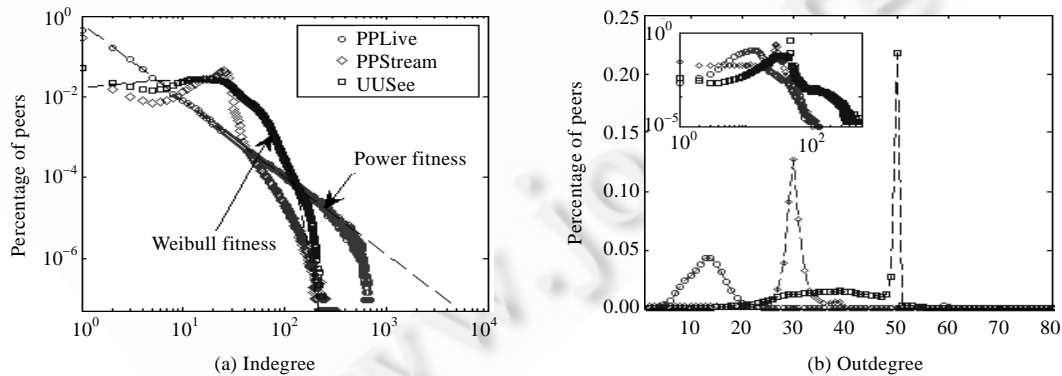


Fig.7 Degree distribution for response peers

图 7 响应节点的度分布

## 5.2 节点度相关性

在多数网络中,两节点之间存在连接边的概率,通常依赖于这两个节点的类型,这种现象通常称为网络的混合模式.依据节点度这一标量特征对节点进行分类从而表现出的网络混合模式,即通常所说的顶点度相关性.节点之间的连接通常会表现出某种倾向性,例如高度节点倾向于与其他高度节点连接,称为同配网络;反之,高度节点倾向于和低度节点连接,称为异配网络.这种节点之间连接的相关性,对网络可靠性、传播动力学等都拥有着不同于度分布特征的重要影响,越来越受到研究者的重视.Newman 提出了一种用 Pearson 相关系数来定量描述混合模式的量化方法<sup>[28]</sup>,即计算有向网络中节点度-度的相关系数  $r$ (公式(4)):

$$r = \frac{\sum_i j_i k_i - M^{-1} \sum_i j_i \sum_i k_i}{\sqrt{[\sum_i j_i^2 - M^{-1} (\sum_i j_i)^2][\sum_i k_i^2 - M^{-1} (\sum_i k_i)^2]}} \quad (4)$$

其中,  $M$  是总边数,  $j_i$  和  $k_i$  是第  $i$  条边关联的两个节点的剩余度,  $i=1,2,\dots,M$ .

图 8 表示在 3 个 P2P IPTV 系统中  $r$  与频道在线人数的关系,图中曲线是进行跨度为 0.1 的 loess 平滑处理后的结果.可以看出:3 个系统中,PPLive 和 UUSee 整体上具有负的度-度的相关系数;而 PPStream 的度-度相关系数虽然波动范围较大,但是其平滑曲线表明它仍然取得负的度-度相关系数.同时还可以看出,当在线节点数较大(500~4000)时,  $r$  趋向于平稳,其取值变化与频道在线人数无明显相关性,即  $r$  取值独立于频道在线人数.在 3 个 P2P IPTV 系统中,PPLive 网络的异配特征最为明显,平均  $r$  为 -0.26,标准差为 0.04;PPStream 平均 DCC 为 -0.12,但是随机波动范围较大,标准差为 0.18;而 UUSee 的异配特征较弱,平均  $r$  为 0.08,标准差为 0.04.显然,3 个 P2P IPTV 系统同时表现出异配特性,即,其中高度数节点是不同程度地偏向于与低度数节点相关联.根据 Brede 等人<sup>[29]</sup>的研究,虽然同配网络具有更低的传播临界值,但是异配网络却具有更大的传播范围,同时对网络动态变化具有更大的弹性.因此,我们分析得到的网络异配性结果也从一定程度上解释了 P2P IPTV 网络在频道在线人数规模上具备良好可扩充性的原因.

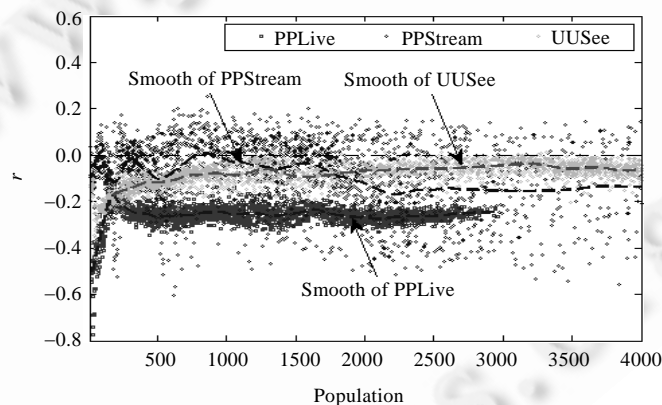


Fig.8 Degree-Degree correlations for response peers

图 8 响应节点的度-度相关性

## 5.3 平均最短路径

平均路径长度是指网络中所有节点对之间最短路径的平均值,通常以节点间的跳数作为度量来计算.平均路径长度与网络的连通性、可达性以及传输延迟等特征密切相关,也是判断复杂网络具有小世界效应的主要指标.平均最短路径的计算采用公式(5)<sup>[30]</sup>:

$$l = \frac{1}{N(N-1)} \sum_{ij} d_{ij} \quad (5)$$

其中,  $d_{ij}$  为网络中节点  $i$  和节点  $j$  之间的最短路径长度.通常认为,在网络平均顶点度固定的前提下,平均最短路径的值随网络大小以对数的速度(或慢于对数的速度)增长,即  $l \sim \log N$ ,称此网络具有小世界效应<sup>[30]</sup>.考虑到哑节

点的存在,我们同样仅计算响应节点之间的平均最短路径.从图 9 中可看出,当频道在线人数较大时,在半对数坐标下,3 个系统中的平均最短路径  $l$  与频道在线人数的对数  $\log(pop)$  具有较明显的线性相关性,即  $l$  与  $pop$  之间存在关系如公式(6):

$$l = a \log(pop) + b \quad (6)$$

图 9 中的虚线是对 3 条  $l$  曲线进行最小二乘拟合的结果,拟合参数见表 2.对应于斜率参数  $a$  还可看出,PPStream 的平均最短路径随频道在线人数的增长速度最快,UUSee 次之,而 PPLive 的增长速度最慢.

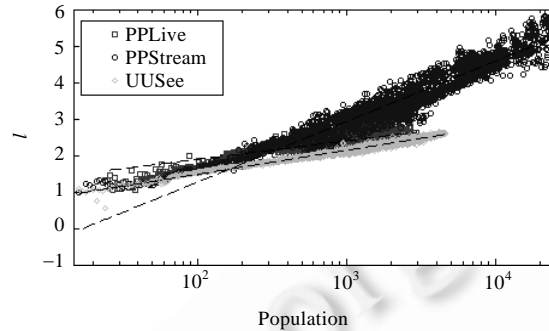


Fig.9 Correlation between length of Average shortest path and online population in channel

图 9 响应节点平均最短路径与频道在线人数的关系

Table 2 Fitness parameter for  $l$ -pop log curve

表 2  $l$ -pop 对数曲线的参数拟合结果

	$a$	$b$
PPLive	0.2204	0.8744
PPStream	0.6843	-1.756
UUSee	0.2853	0.1662

由于平均最短路径相对于频道在线人数明显倾向于对数增长,可认为 PPLive,PPStream 和 UUSee 这 3 个网络都具有典型的小世界特征.这也解释了这些 P2P IPTV 系统对在线用户数量上具有较强可扩展性的原因,即使大量用户在线,所形成的大规模 P2P IPTV 覆盖网络仍然具有较短的平均最短路径,从而保证了实时媒体数据的平均传播时间,能在全系统范围内提供均衡的、满意的服务质量.

#### 5.4 聚类系数

聚类系数用于描述一个节点邻居之间相互连接的紧密程度,即网络的集团化程度,是网络拓扑的另一个重要参数.对于网络聚类系数的量化定义包括平均聚类系数(mean clustering coefficient)和全局聚类系数(global clustering coefficient)两种.由于平均聚类系数是对每个节点的局部聚类系数求均值,过多地倾向于加强低度节点在计算结果中的权重<sup>[30]</sup>,因此我们采用全局聚类系数对实验数据集中的聚类特性进行刻画,计算方法如下:

$$C = \frac{3 \times (\text{Number of triangles in the network})}{\text{Number of connected triples of vertices}} \quad (7)$$

当网络规模充分大时,SW 模型和 BA 模型的聚类系数趋向于 0.而许多大规模的实际网络通常都具有明显的聚类特性,他们的聚类系数尽管远小于 1,但是却比  $O(N^{-1})$  要大得多<sup>[30]</sup>.图 10 表示 3 个系统中响应节点之间的聚类系数与频道在线人数的关系,可以看出,PPLive 的响应节点间虽然聚类系数较小,但是其聚类系数在频道在线人数增加时无显著的变化趋势,统计聚类系数均值为 0.03,可以认为 PPLive 存在轻度的聚类特性.而从双对数坐标下可以看出,PPStream 和 UUSee 的聚类系数随频道在线人数的变化趋向于  $O(N^{-1})$ ,说明这两个系统不具备明显的聚类特性.3 个系统在聚类特性上的不同以及由此可能带来的对系统性能的影响,还需进一步研究.

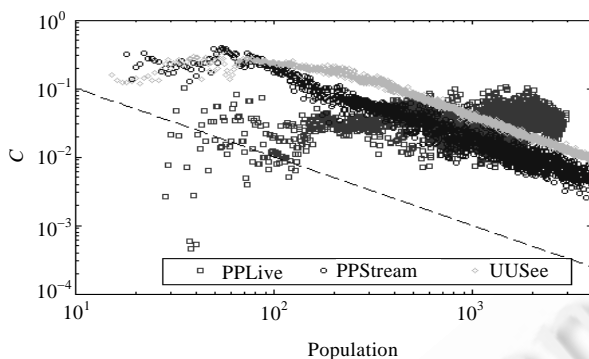


Fig.10 Correlation between clustering coefficient and online population in channel  
图 10 响应节点聚类系数与频道在线人数的关系

5.5 拓扑鲁棒性

鲁棒性是 P2P 覆盖网络的重要研究内容之一,包括 Internet,WWW 等在内的无标度网络的一个基本特征,就是具有对随机故障的鲁棒性和对蓄意攻击的脆弱性.研究 P2P IPTV 覆盖拓扑的鲁棒性特征,是实现 P2P IPTV 系统可靠性和容错性的基础.本文在测量数据集中随机选择了若干节点规模较大的快照作为研究对象,考虑两类节点去除策略:一是随机故障策略,即随机删除网络中的部分节点;二是蓄意攻击策略,即删除网络中度最高的部分节点.假设删除节点数占网络总节点数的比例为  $f$ ,考察删除节点后的最大连通子图  $G'$ ,采用  $G'$  的相对大小  $S$  和  $G'$  的平均路径长度  $l$  与  $f$  的关系来度量网络的鲁棒性.图 11 是 3 个 P2P IPTV 系统中各取 20 张快照进行鲁棒性实验的平均结果,可以看出,这 3 个系统都具有类似无标度网络的鲁棒性特征,即对随机故障容错性较高,而对蓄意攻击则表现脆弱性.其中,对于随机故障,随着删除节点比例增加直到 0.7,PPLive 覆盖网络最大连通团的平均最短路径  $l$  基本保持不变;而对于蓄意攻击,当删除节点比例达到 0.2 时,PPLive 覆盖网络最大连通团的相对大小  $S$  已经降低至 0.25,而平均最短路径  $l$  则增加了 4 倍.说明与其他两个系统比较,PPLive 具有更高的故障容错性和更明显的攻击脆弱性.一个可能的原因是,尽管三者都不是完全无尺度网络,但 PPLive 的度分布具有更加明显的幂次特征,其覆盖网络中少量高度数节点在连通性中扮演了重要的角色;而在 PPStream 和 UUSee 的覆盖拓扑中,出度分布相对均匀,具有相对更强的随机性特征,从而降低了它们面对蓄意攻击的脆弱性.

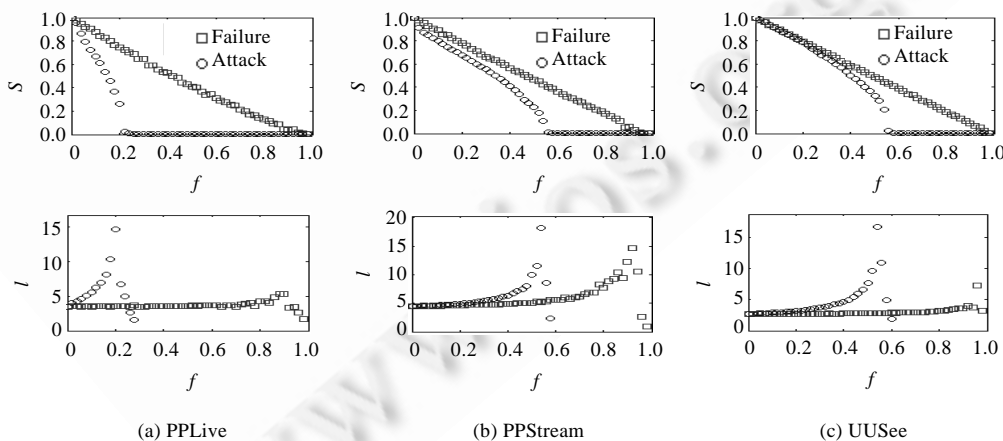


Fig.11 Comparison of robustness in overlay topology  
图 11 覆盖拓扑的鲁棒性比较

## 6 结 论

随着对等网络技术的发展,P2P IPTV 网络流量已经成为互联网通信流量的主要构成部分.为了实现对 P2P IPTV 系统进行有效监测、引导、控制以及进行协议改进,需要对已有的 P2P IPTV 系统进行测量研究,了解其网络拓扑特性、流量特征以及其用户行为特征.由于 P2P IPTV 网络具有大规模、自组织、高度动态和节点异构等特性,要对其实施完整、准确的测量存在很大的困难.目前,基于爬行器的主动测量是 P2P IPTV 测量研究中的最主要技术手段.它通过模拟 P2P IPTV 网络客户端的部分协议交互行为,对 P2P IPTV 覆盖网络进行迭代地爬行,获取拓扑数据.

本文采用自行设计的 P2P IPTV 多协议爬行器 TVCrawler,对 PPLive,PPStream 和 UUSEE 这 3 个系统的财经类直播频道  $F$  进行了连续两周的测量,并根据采集的数据集对 P2P IPTV 用户行为和网络拓扑特征进行了分析,主要发现包括:

- (1) P2P IPTV 系统的频道在线人数依赖于频道节目播出的内容,而且在线用户中有一半以上位于不可达的 NAT 或者防火墙后面.就  $F$  频道而言,其主要用户来自于中国大陆地区(>95%),其他用户主要分布在亚洲的其他国家、北美和欧洲;
- (2) 节点 churn 变化率的波动范围随频道人数增加而增加,但是其取值范围相对频道人数的变化具有幂律上限;
- (3) 与已有研究不同,我们发现节点会话长度并不符合指数分布,却能够较好地符合广延指数分布;
- (4) 平均度大小与频道在线人数取值无关,PPLive 的入度分布服从具有指数截断的幂次分布,PPStream 和 UUSEE 的入度分布比 PPLive 相对均匀,其中,PPStream 的入度表现为某种分段幂律函数,而 UUSEE 的出度接近威布尔分布.而 3 个系统的出度分布都具有明显的泊松分布特征;
- (5) 3 个系统同时表现出异配特性,即其中高度数节点不同程度地偏向于与低度数节点相关联;
- (6) 3 个系统的平均最短路径与频道在线人数成对数增长关系,即都表现为小世界网络;
- (7) PPLive 网络具有聚类特征,而 PPStream 和 UUSEE 则不存在明显聚类特征;
- (8) 3 个系统都表现出类似于无标度网络的鲁棒性特征,其中,PPLive 与其他两个系统比较,具有更高的故障容错性和更明显的攻击脆弱性.

下一步的研究工作包括两个主要部分:其一是对频道在线人数、用户地域分布、动态性等用户行为特征进行时间序列分析,挖掘 P2P IPTV 用户行为模式,建立用户行为预测模型,实现对异常用户行为的检测;其二,Newman 等人<sup>[30]</sup>的统计数据表明,社会网络都是同配网络,而其他类型的网络(信息网络、技术网络、生物网络)都是异配网络.虽然这一现象仍然无法得到足够解释,但是 Newman 同时也认为,其中的一个重要原因是由于社会网络具有社区结构(community structure),而非社会网络则不具有明显的社区结构,从而造成其混合模式上的不同<sup>[31]</sup>.我们的测量结果表明,P2P IPTV 网络都具有异配性,那么这是否表明自组织演化的 P2P IPTV 不具备社区结构呢?这类问题对于大规模无结构化的 P2P 流媒体应用的协议设计、优化以及引导和监管也具有重要的意义.

### References:

- [1] PPLive website. 2009. <http://www.PPLive.com>
- [2] PPStream website. 2009. <http://www.ppstream.com>
- [3] UUSEE website. 2009. <http://www.uusee.com>
- [4] Saroiu S, Gummadi PK, Gribble SD. A measurement study of peer-to-peer file sharing systems. In: Kienzie MG, Shenoy PJ, eds. Proc. of the Multimedia Computing and Networking 2002. San Jose: SPIE, 2002. 156-170.
- [5] Ripeanu M, Foster I, Iamnitchi A. Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. IEEE Internet Computing Journal, 2002,6(1):50-57. [doi: 10.1109/4236.978369]
- [6] Stutzbach D, Rejaie R, Sen S. Characterizing unstructured overlay topologies in modern P2P file-sharing systems. IEEE/ACM Trans. on Networking, 2008,16(2):267-280. [doi: 10.1109/TNET.2007.900406]

- [7] Liang J, Kumar R, Ross KW. The KaZaA overlay: A measurement study. In: Proc. of the 19th IEEE Annual Computer Communications Workshop. Bonita Springs, 2004.
- [8] CoolStreaming website. 2009. <http://www.coolstreaming.us/>
- [9] Zhang XY, Liu J, Li B. On large scale peer-to-peer live video distribution: CoolStreaming and its preliminary experimental results. In: Proc. of the IEEE Int'l Workshop on Multimedia Signal Processing (MMSP 2005). Piscataway: IEEE Press, 2005.
- [10] Li B, Xie SS, Qu Y, Keung GY, Lin CC, Liu J, Zhang XY. Inside the new CoolStreaming: Principles, measurements and performance implications. In: Proc. of the 27th IEEE Conf. on Computer Communications. Piscataway: IEEE Press, 2008. 1031–1039.
- [11] Wu C, Li BC, Zhao SQ. Exploring large-scale peer-to-peer live streaming topologies. ACM Trans. on Multimedia Computing, Communications, and Applications (TOMCCAP), 2008,4(3):19. [doi: 10.1145/1386109.1386112]
- [12] Hei XJ, Liang C, Liang J, Liu Y, Ross KW. Insights into PPLive: A measurement study of a large-scale P2P IPTV system. In: Proc. of the IPTV Workshop, Int'l World Wide Web Conf. New York: ACM Press, 2006. [doi: 10.1109/TMM.2007.907451]
- [13] Hei XJ, Liang C, Liang J, Liu Y, Rossy KW. A measurement study of a large-scale P2P IPTV system. IEEE Trans. on Multimedia, 2007,9(8):1672–1687. [doi: 10.1109/TMM.2007.907451]
- [14] Vu L, Gupta I, Liang J, Nahrstedt K. Measurement of a large-scale overlay for multimedia streaming. In: Proc. of the 16th Int'l Symp. on High Performance Distributed Computing. New York: ACM Press, 2007. 241–242. [doi: 10.1145/1272366.1272410]
- [15] Vu L, Gupta I, Nahrstedt K, Liang J. Understanding overlay characteristics of a large-scale peer-to-peer IPTV system. ACM Trans. on Multimedia Computing, Communications and Applications, 2010,6(4):1–24.
- [16] Ali S, Mathur A, Zhang H. Measurement of commercial peer-to-peer live video streaming. In: Proc. of the 1st Int'l Workshop on Recent Advances in Peer-to-Peer Streaming (WRAIPS 2006). Waterloo, 2006.
- [17] Silverston T, Fourmaux O. Measuring P2P IPTV systems. In: Proc. of the 17th Int'l Workshop on Network and Operating Systems Support for Digital Audio & Video (NOSSDAV 2007). New York: ACM Press, 2007. [doi: 10.1145/1368436.1368490]
- [18] Jiang ZH, Wang H, Fan PY, Yuan XM. TVCrawler: Multi-protocol P2P IPTV crawler. Journal of Computer Applications, 2010, 30(3):715–718 (in Chinese with English abstract). [doi: 10.3724/SP.J.1087.2010.00715]
- [19] Ganesh AJ, Kermarrec AM, Massoulié L. Peer-to-Peer membership management for gossip-based protocols. IEEE Trans. on Computer, 2003,52(2):139–149. [doi: 10.1109/TC.2003.1176982]
- [20] Stutzbach D, Rejaie R. Capturing accurate snapshots of the Gnutella network. In: Proc. of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005). Piscataway: IEEE Press, 2005. 2825–2830. [doi: 10.1109/INFCOM.2005.1498570]
- [21] Stutzbach D, Rejaie R. Evaluating the accuracy of captured snapshots by peer-to-peer crawlers. In: Dovrolis C, ed. Proc. of the 6th Int'l Workshop on Passive and Active Measurement (PAM 2005). New York: Springer-Verlag, 2005. 353–357.
- [22] Stutzbach D, Rejaie R. Characterizing today's Gnutella topology. Technical Report, CIS-TR-04-02, University of Oregon, 2004.
- [23] Vu L, Gupta I, Liang J, Nahrstedt K. Measurement and modeling of a large-scale overlay for multimedia streaming. In: Proc. of the 4th Int'l Conf. on Heterogeneous Networking for Quality, Reliability, Security and Robustness & Workshops. New York: ACM Press, 2007. 3. [doi: 10.1145/1577222.1577227]
- [24] Silverston T, Fourmaux O. P2P IPTV measurement: A comparison study. The Computing Research Repository (CoRR), abs/cs/0610133, 2006.
- [25] Kumar P, Sridhar G, Sridhar V. Bandwidth and latency model for DHT based peer-to-peer networks under variable churn. In: Proc. of the 2005 Systems Communications. Washington: IEEE Computer Society, 2005. 320–325. [doi: 10.1109/ICW.2005.31]
- [26] Saroiu S, Gummadi KP, Gribble SD. Measuring and analyzing the characteristics of Napster and Gnutella hosts. Multimedia Systems, 2003,9(2):170–184. [doi: 10.1007/s00530-003-0088-1]
- [27] Trevor F, Mark L, George L. A stochastic evolutionary model exhibiting power-law behaviour with an exponential cutoff. Physica A: Statistical Mechanics and its Applications, 2005,355(2-4):641–656. [doi: 10.1016/j.physa.2005.01.007]
- [28] Newman MEJ. Mixing patterns in networks. Physical Review E, 2003,67(2):026126. [doi: 10.1103/PhysRevE.67.026126]
- [29] Brede M, Sinha S. Assortative mixing by degree makes a network more unstable. Arxiv Preprint, cond-mat/0507710, 2007.

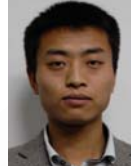
- [30] Newman MEJ. The structure and function of complex networks. *SIAM Review*, 2003,45(2):167–256. [doi: 10.1137/S003614450342480]
- [31] Newman MEJ, Park J. Why social networks are different from other types of networks. *Physical Review E*, 2003,68(3):036122. [doi: 10.1103/PhysRevE.68.036122]

附中文参考文献:

- [18] 姜志宏,王晖,樊鹏翼,袁雪美.一个 P2P IPTV 多协议爬行器——TV Crawler. *计算机应用*,2010,30(3):715–718.



姜志宏(1975—),男,湖南新化人,博士生,讲师,主要研究领域为复杂网络,P2P 流媒体.



樊鹏翼(1984—),男,博士生,主要研究领域为 P2P 流媒体建模与仿真.



王晖(1968—),男,博士,教授,博士生导师,CCF 会员,主要研究领域为信息系统工程,多媒体网络.