第 34 卷 第 16 期 计 算 机 工 程 2008 年 8 月
Vol.34 No.16 Computer Engineering August 2008
·网络与通信· 文章编号：1000—3428(2008)16—0119—03 文献标识码：A 中图分类号：TP393

分层的应用层组播协议研究

张秋余，随冬梅
(兰州理工大学计算机与通信学院，兰州 730050)

摘 要：在P2P视频直播网络中，用户频繁地加入或离开组播组会造成数据传输中断。该文提出一种新的基于分层分簇思想的应用层组播协议，基于视频直播中的用户行为分析，通过分层获得一个有效层来减少组播树中的节点失效次数。仿真实验表明，该协议能够有效提高稳定性，减少平均组播时延并具有可扩展性。

关键词：点对点；应用层组播；用户行为分析；视频直播

Research on Hierarchical Application Layer Multicast Protocol

ZHANG Qiu-yu, SUI Dong-mei
(College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050)

【Abstract】 Data delivery can be easily interrupted by departure of end hosts in P2P live streaming network, which may lead to degradation of QoS in time-sensitive applications. This paper proposes a new layered and clustering protocol. Based on user behavior analysis in live streaming, it can keep a level of efficiency while reduce the interruption times of Application Layer Multicast(ALM) tree. Simulation results show that the protocol can effectively boost up the robustness, reduce the average ALM latency, and satisfy the scalability.

【Key words】 P2P; Application Layer Multicast(ALM); user behavior analysis; live streaming

1 概述

P2P 视频直播对直播的实时性有很大的要求，而在 P2P 系统中，组成组播树的中间节点并不是专门的网络路由器，而是自主的终端主机，每个终端主机的行为都是不可控制的。当节点的数量巨大时，每时每刻都有很多节点的加入与离开，致使数据传输暂时中断，影响该节点的子孙用户的视频观看。

近年来，虽然有一些针对流媒体负载特性进行的研究，但涉及到视频直播特性分析的研究却很少。文献[1]通过对央视国际网站在线频道历时 103 天的视频直播日志共 1 000 多万条记录的统计发现，CCTV2、CCTV4 和 CCTV9 这 3 个直播频道中分别有 40.17%, 49.18%和 46.19%的用户在 60 s 内退出，而到 300 s 时该值分别达到 69.18%, 76.19%和 73.16%，这样的结果说明大部分用户持续观看某一频道的时间很短，非常不利于构建稳定的组播树。因此，需要对参与组播的节点退出组播的可能性进行预测。该统计数据验证了文献[2]中得到的在 P2P 视频直播中用户在线时间符合对数正态分布的结论，并指出用户平均剩余在线时间会随着用户已经在线时间的增大而增大。

显然，在P2P视频直播系统中，稳定性是一个非常关键的问题。现有应用层组播协议却很少考虑组播树的稳定性。HBM[3]采用冗余链路策略在中间节点离开时恢复其下游节点的数据接收。冗余链路策略通过在节点之间增加冗余链路来提高系统的稳定性，这增大了节点的处理负担。针对现有系统中存在的缺陷，本文通过预测节点的行为提出一种低中断频率的应用层组播协议(下文简称为NHP)。

2 NHP 系统模型

2.1 NHP 的网络拓扑

NHP 网络拓扑分为 2 层：Lower 层和 Upper 层。节点的组织如图 1 所示。

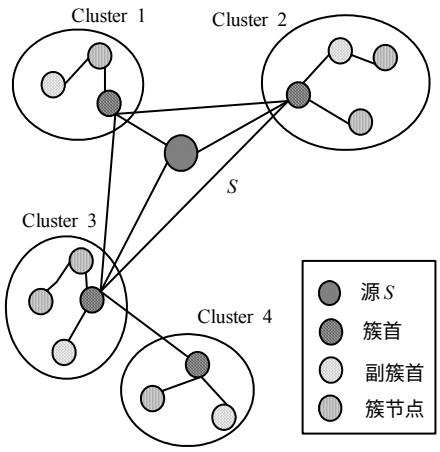

图 1 NHP 的网络组织结构

Lower 层包括参与组播的所有节点(源 *S* 除外)。节点按照距离远近组成簇。簇的划分主要根据节点的物理位置(IP 地址)，将物理地址临近的节点划为一个簇。Lower 层的每个簇都根据算法 1(见 2.3 节)选出一个服务能力和稳定性都较强的节点作为簇首，一个稳定性最强的节点为副簇首。所有簇首与源 *S* 一起组成 Upper 层。

NHP 通过分层分簇来获得一个有效稳定层，“分层”、“分簇”思想也使协议具有良好的扩展性。其中，每个簇的簇首负责创建和维护本簇内的组播树。副簇首用于备份簇首所维护的信息数据，以防簇首失效后造成的信息丢失。Upper 层节点在源 *S* 的协助下互相连接。这样所有参与组播的节点

作者简介：张秋余(1966－)，男，副研究员，主研方向：图像处理，模式识别，多媒体通信；随冬梅，硕士研究生

收稿日期：2007-09-28 **E-mail**：suilemon@163.com

构成覆盖网络，任何 2 个节点之间在逻辑上都存在一条路径。

NHP的数据传输有 2 个策略。在簇内，簇首以算法 2(见 2.4 节)构建树的方法向本簇内的所有节点进行应用层组播(Application Layer Multicast, ALM)，而在Upper层中，源*S*运行modified Dijkstra's算法[4]进行ALM。其数据传输路径如图 2 所示。

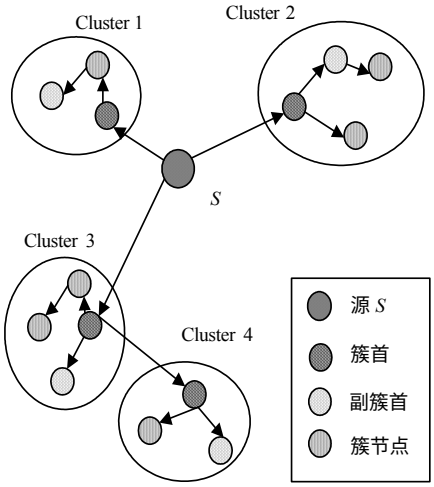

图 2　NHP 的数据传输结构

2.2　影响失效次数的因素及系统稳定性评价函数

在 NHP 中，要构建有效稳定的 Upper 层，应先分析影响应用层组播失效次数的因素。根据文献[1]的分析，节点在线的时间越长，其继续停留的平均时间越大，退出概率越小。本文定义 $prob(i, t)$为节点 i 已在线时间为 t 时离开组播树的概率，其中，$prob(i, t)$的大小仅仅与节点 i 的已在线时间有关。但是，文献[1]没有考虑到在线时间越长的节点，也有可能是最快离开组播组的节点。如时长为 60 min 的节目，当一个节点已在线 56 min 时，它退出的概率可能是最小的，但是 4 min 后，它是一定离开的。因此，综合考虑节点的已在线时间和节点的可能剩余在线时间，本文定义一个新的指标-稳定度 d 来表示节点在组播组中的稳定性，设一个直播节目总时长为 T，把 T 分成 n 等份，令$\triangle t=T/n$, $n\in 2N$。定义 d 为

$$d=\begin{cases}\left\lceil \dfrac{t}{\triangle t}\right\rceil & t\le \dfrac{T}{2}\\ \left\lceil \dfrac{T-t}{\triangle t}\right\rceil & t>\dfrac{T}{2}\end{cases} \tag{1}$$

设P_d是稳定度为d的节点失效的概率，令

$$P_d=\frac{\frac{n}{2}+1-d}{1+2+\cdots+n} \tag{2}$$

则稳定度为 d 的节点失效受影响的节点数的期望为

$$E[N_d]=\sum_{d=1}^{n/2}P_d\times N_d \tag{3}$$

其中，N_d是稳定度为d的节点失效受影响的节点个数。组播树的期望失效恢复代价越小，系统越稳定。

由式(3)可知，影响失效次数的因素有 2 个，一个是P_d，另一个是N_d。因为P_d的大小仅与d有关，不能改变其大小，所以降低失效恢复代价可行的方法应该遵循以下 2 个原则：

(1)尽量减少稳定度 d 小的节点处于组播树的上层。

(2)使节点能力较大的节点位置靠近根节点，使构建的树尽量短而宽。

本文使用应用层组播稳定性评价函数(Rb)来评价系统的稳定性。

$$Rb=(1-\frac{E[N_d]}{N})\times 100\% \tag{4}$$

其中，N 是除源 S 之外的组播成员数；Rb 是在单点失效情况下组播树的稳定函数，Rb 越大，组播树越稳定。

2.3　簇首和副簇首选择

根据上文中为了降低失效恢复代价应该遵循的 2 个原则，在选择簇首时，应把服务能力不足和稳定度小的节点排除，在服务能力强和稳定度大的节点中选择簇首。这样，就可以构建一个有效稳定的 Upper 层。本文的节点服务能力仅考虑带宽服务，用节点的度表示。

设$d_{\max}[i]$表示节点i的总带宽，$d_n[i]$表示节点i已使用带宽，节点的稳定度为d，SPH_j是簇j的节点集合，该簇节点个数为C_n。

算法 1

(1)对SPH_j中所有节点按照节点的可用带宽$d_{\max}[i]$-$d_n[i]$进行降序排序，得到集合SPH_j1。这样具有较大带宽服务能力的节点占据SPH_j1 中的靠前位置。

(2)取SPH_j1 中前αC_n个节点按照节点的稳定度d进行降序排序，得到SPH_j2。参数α可根据应用需求选择($\alpha\le 1$)。用αC_n调节在带宽服务能力强的节点中选择稳定度大的节点的个数。

(3)选取SPH_j2 中排名最前的节点为该簇的簇首。由簇首选取稳定度d最大的节点作副簇首。若副簇首与簇首相同，则在其他稳定度最大的节点中任选一个作副簇首。

2.4　节点加入

一个新节点要加入某个组播组，应获得源S的地址。一般情况下，系统中服务器S的地址是已知的。待加入者j生成Inquiryforcluster报文发送给源S，源节点收到Inquiryfor cluster报文后，将它维护的簇首集合set of clusterheader生成Ackoncluster报文返回给j。j根据收到的簇首地址，向每个簇首发出Probecluster报文，探测到各簇首之间的延迟，根据返回的探测结果，j向距离最近的簇首发出Joinforcluster请求。若簇已满，则簇仿照NICE[5]协议进行簇分裂；若簇未满，则该簇簇首调用算法 2 选出一个父节点并以报文Ackonpp形式返回给j一个父节点；j向父节点发送Subscribe 报文，父节点收到报文后，将j加入自己维护的孩子列表中。父节点向j发送AckOnJoin报文，表示加入成功。待加入者向簇首节点发送UpdateTreeJoin 报文，簇首更新维护的组播树结构。

算法 2

```
SPP1=SPP2=NULL;//候选父节点集合
 for p= 1 to Cn {
    if dn[p]<dmax[p]
       p 添加到 SPP1; };
  for each p of SPP1{
    if l[p]<=⌈L⌉
// l[p]是节点 P 的深度，L 是组播树的平均深度
       p 添加到 SPP2;
    else
       深度最小的节点 p 组成 SPP2; };
 //待加入节点 j 计算到本簇簇首的延时
 delay(r,j)min=∞;
//r 为本簇簇首节点
 parent=NULL;
 for each PPi {
//选延时最小的路径
```

$delay_{PPi}(r,j)=delay(r,PPi)+delay(PPi,j)$;
if $delay(r,j)<delay(r,j)_{min}$ {
$delay(r,j)_{min}=delay_{PPi}(r,j)$;
parent=PPi; };
retrun parent;

2.5 节点的离开

簇中组播树是动态平衡的，每时每刻都有节点退出。节点退出分为正常退出和非正常退出。对于节点非正常退出，本文采用心跳机制(heart beats)。即组播树上的每个节点周期性地向其父节点和孩子节点发送一个探测报文，表示自己仍处于活动状态，当一个节点在一段时间内没有收到来自某个节点的信息时，就认为该节点非正常退出。簇中不同节点的退出采用的修复策略不同：

(1)若待离开节点为非簇首节点，则向簇首和与其直接相连的邻居节点发送 Leave 报文。若有孩子节点，则每个孩子节点向簇首发送 Joinforcluster 报文，重新加入组播树。

(2)若待离开节点是簇首节点，则向本簇副簇首和与其直接相连的邻居节点及服务器 *S* 发送 Leave 报文，启动簇首选择机制，重新选出一个新的簇首。

(3)若待离开节点是副簇首节点，则向簇首和与其直接相连的邻居节点发送 Leave 报文，由簇首重新选出一个新节点作为副簇首。

3 性能评价

3.1 模拟环境

为验证NHP协议的性能，本文将NICE与NHP协议进行对比实验。实验使用GT-ITM生成不同规模(50 个、100 个、200 个、500 个、1 000 个节点)的网络拓扑。所有节点在 200 s内随机加入。NHP的簇大小限制参数C_{max}随着节点规模的增大从 5 增加到 25。整个模拟实验的时间为 1 000 s。

3.2 性能评价参数

(1)伸张度(*stretch*)。该参数以每个接收者成员作为统计单位，计算从源沿着覆盖网络中路径到达组成员的路径长度和直接从源到组成员的单播路径长度的比值。伸张度实质上反映了数据传输的相对延迟损失。

(2)应用层组播稳定性评价函数(*Rb*)。

3.3 仿真结果分析

NICE 与 NHP 在平均伸张度上的比较如图 3 所示。

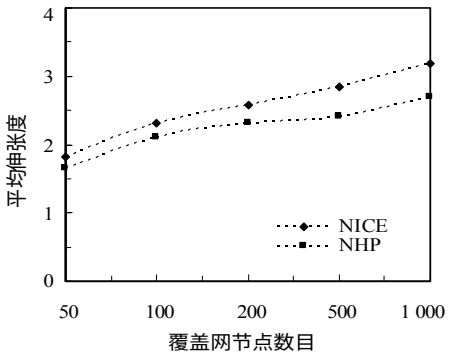

图 3 NICE 与 NHP 在平均伸张度上的比较

由图 3 可知，NHP 在 *stretch* 上优于 NICE。这是因为 NICE 仅考虑待加入节点和已加入节点之间的距离最短，但不能保证源到待加入节点之间的距离最短。而在 NHP 中，虽然待加入节点在选择父节点时同样没有考虑距离最优，但它通过分层和改进算法减少了覆盖路径的覆盖跳数，优化了整个组播树的平均时延。

图 4 显示了 NICE 和 NHP 的稳定函数都高于 85%。在单点失效的情况下，NHP 比 NICE 有更高的稳定性。这是因为 NHP 通过分层获得一个有效的稳定层，使能力强、稳定度大的节点处于组播树的上层，所以减少了节点的平均失效次数，达到较高的稳定函数。

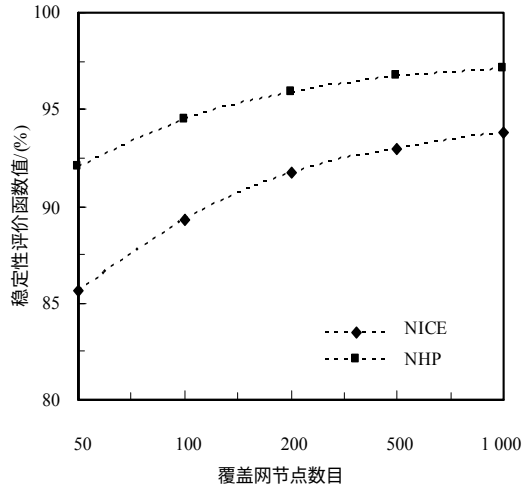

图 4 NICE 与 NHP 在单点失效时的稳定函数的比较

4 结束语

在 P2P 视频直播系统中，节点行为总是不可控制的，当节点规模增大时，每时每刻都有节点的加入与离开，严重影响应用层组播系统的稳定性。NHP 通过分层构建一个有效稳定层来提高组播树的稳定性，通过分簇由簇首来分担服务器的控制负载。簇内节点的信息只须由簇首节点负责维护，而无须服务器的维护，这样能使系统具有较好的可扩展性。与 NICE 协议的模拟对比结果显示，NHP 协议在平均延迟上略有优化，而在稳定性方面的优化更为明显。

参考文献

[1] 罗建光, 赵 黎, 杨士强. 基于用户行为分析的应用层组播树生成算法[J]. 计算机研究与发展, 2006, 43(9): 1557-1563.

[2] Veloso E, Almeida V, Meira W. A Hierarchical Characterization of a Live Streaming Media Workload[C]//Proc. of ACM SIGCOMM Workshop on Internet Measurement. New York, USA: [s. n.], 2002.

[3] Roca V. A Host-based Multicast(HBM) Solution for Group Communications[C]//Proc. of IEEE Int'1 Conf. on Networking. Colmar, France: [s. n.], 2001.

[4] Song H, Lee D S. Application Layer Multicast Tree for Real-time Media Delivery[J]. Computer Communications, 2006, 29(8): 1480-1491.

[5] Banerjee S, Bhattacharjee B, Kommareddy C. Scalable Application Layer Multicast[C]//Proc. of ACM SIGCOMM. Pittsburgh, PA, USA: [s. n.], 2002.