一个融合组播流媒体系统①

A Mix Multicast System for Video Streaming

彭墨青 (广东外语外贸大学 教育技术中心 广东 广州 510006)

谢建国 (广东外语外贸大学 信息学院 广东 广州 510006)

摘 要: 文章对目前网络中 IP 组播流媒体存在组播孤岛和不同接入带宽用户群组播数据不能共享等问题,提出了一种新的流媒体融合组播覆盖网络,研究通过 P2P 互联不同 IP 组播区,实现组播数据共享。每一个 IP 组播区内有一个 IP 组播源(IP-SM)和一到多个 P2P 组播源(P2P-SM),所有的 IP-SM 和 P2P-SM 组成一个融合组播树,将不同的 IP 组播区互联。IP 组播区内采用网络层组播技术,组播区之间采用应用层组播技术,实现不同组播模式的融合,使流媒体系统具有扩展性和网络适应能力。

关键词: 流视频 伸缩编码视频 IP 组播 P2P 融合组播树

1 引言

IP 组播视频是最有效的流媒体模式[1],但由于 IP 组播在拥塞控制,流量控制,扩展性,协议部署,以及网络异构性和终端多样性适应能力等方面存在一些问题,所以基于 IP 组播的流媒体应用一直未在互联网上得到实际的开展。P2P 流媒体模式,如应用层组播 ALM(Application Layer Multicast),是基于终端复制与转发的,因其适应能力强成为了 IP 组播模式的替换方案,然而这种模式,存在延迟,抖动,节点状态和传输效率等问题。

因协议部署和管理等问题,互联网络中出现的一些 IP 组播孤岛现象,可以通过 MBone 隧道技术进行连接,但需要主机操作系统的配置和支持 MBbone 中的 IP-in-IP 传输模式,具体实现存在难度。使用 UDP 隧道也可以实现 IP 组播孤岛互联[2,3],UDP 隧道形成一个应用层组播树,连接分布在因特网中 IP 组播孤岛,实现 IP 组播数据跨越因特网分发。融合组播技术将 P2P 技术和 IP 组播技术两种传输模式整合,发挥它们各自的优点,力图解决各自模式存在的固有缺陷,实现网络流媒体的大规模应用问题。本文提出了一种新的流视频融合组播树系统,和上述的 UDP 组播树有两个明显的不同:一是融合组播树中的支路由 IP 层传输链和 P2P 应用层传输链组成;二是文章的 IP 组播区不同于物理上的 IP 组播岛,是逻辑上的概念,且支持分层视频传输,如 PC 用户群和手机用户群可共享不同层次的组播视频。

2 融合组播流媒体系统

IP 组播工作在网络层,P2P 工作在应用层,是将组播作为一种叠加的业务来实现的,属于应用层组播。IP 组播的数据是沿着物理链路复制和转发,而应用层组播的数据则在主机端进行复制和转发,数据报沿着逻辑链路传输。IP 组播能更多的节约网络带宽,数据转发速率更高。而应用层组播与现今网络结构相符合,使用时不需变动现有的网络协议与硬件,部署相对容易,具有较好的扩展性,能适应网络条件的动态变化。融合组播技术用 P2P 连接 IP 组播区实现两种模式的融合,将 IP 组播的有效性和 P2P 组播的适应性结合起来,实现多 IP 组播区的互联互通。以下先给出融合系统中的一些定义。

2.1 主要术语

定义 1. IP 组播区 是一个尺寸任意的网络,由一个或多个主机组成,域中的主机之间支持网络层组播技术,并共享相同的数据流。把参与组播的主机称之为叶成员(Leaf Member,LM)。

如某以太网接入的计算机用户群,ADSL 接入用

① 基金项目:湖南省自然科学基金项目(05JJ30112);广州市科技计划项目(2007J1-C0321)
收稿时间:2008-10-07

© 中国科学院软件研究所 http://www.c-s-a.org.cn

户群，手机用户群等，这些用户群因物理网络的环境不同，组成不同的 IP 组播区。同一物理网络内因带宽或用户质量需求不同也可以单独建立 IP 组播区。在文献[3]中上述的组播物理阻隔被称为组播孤岛，本文的定义允许一个组播岛中可以有多个 IP 组播区，它们可以有不同服务质量需求。

定义 2. 根数据源(Root Source，RS)是流视频服务的根起点，它可以是专门的流视频服务器，也可以是一台普通的 PC 主机如个人视频会议发布服务。

定义 3. P2P 数据源成员(Source Member，P2P-SM)是 IP 组播区中的一个 LM，在本 IP 组播区内它既是组播数据的接收者，同时它又是另一个 IP 区中 IP-SM 的数据发送者。P2P-SM 与 IP-SM 之间采用 P2P 模式通信，实现不同 IP 组播区之间传输数据。

定义 4. IP 数据源成员(IP-SM)是所在 IP 组播区中的组播源，其数据来自 RS 或其他 IP 组播区的 P2P-SM，为本区提供流视频组播服务。

定义 5. 视频信息表(Video Information List，VIL)用来记录视频属性的，包括：视频名、视频质量等级数(根据伸缩视频编码层数来定义，若等级数为 5，则包括一个基本层，4 个增强层)。

定义 6. IP 组播区列表(IP Multicast Area List，IP-MAL)，这些组播区信息是 P2P-SM 需要的，IP-ML 的信息包括：IP-SM 的 IP 地址、组播组号、质量等级、启播时间等。IP-MAL 表保存在 RS 上，并由 RS 实时维护。

定义 7. IP 组播区成员表(IP Multicast Member List，IP-MML)用来记录本 IP 组播区内的所有在线成员，基本信息包括：LM 的 IP 地址、LM 类型(IP-SM=0，P2P-SM=1，LM=2)等。

定义 8. 每一个 LM 都自己维持一个成员信息表(Member Information List, MIL)：质量等级(基本层=0，基本层+第 1 个增强层=1，依此类推)、接入带宽(Kb)、CPU 空闲状态(%)等。

定义 9. HELLO 报文是融合组播数运行过程中控制报文，用于节点之间的信息交互和参数传递，主要字段包括：报文类型(请求加入=0、请求退出=1、播放状态=目前收到的视频分组序列号)、报文负载(如当报文类型为 0 时，携带 MIL 表信息)等。

2.2 融合组播算法

一个融合组播流视频系统中，第一个请求 RS 流视频服务(如实况转播)的节点被默认为是一个空 IP 组播区的 IP-SM，并被记录到 RS 的 IP-MAL 列表中，规定从 RS 到 IP-SM 的数据传输是 P2P 模式的。其后有其他节点请求相同视频服务时，由 RS 根据新节点的网络属性和视频质量需求描述来决定它加入哪一个区或单独成立新区。当一个组播区的 LM 成员增加时，区中的 IP-SM 成员可以根据每一个成员的综合性能重新选定，以增加组播区的稳定性。任何 IP-SM 到同区的 LM 传输采用 IP 层组播，任何 P2P-SM 到 IP-SM 传输采用应用层组播。根据图 1，有 3 个流视频组播区：局域网接入的组播区、ADSL 接入的组播区和无线移动组播区。IP-SM1 首先请求 RS 的全视频，并形成第一个组播区；其后 IP-SM2 请求 RS 的全视频，被指定从 P2P-SM1 中获得全视频数据，从而形成第二组播区；最后 IP-SM3 请求 RS 的基本层视频，被指定从 P2P-SM2 中分流得到，形成第三个组播区。上述方式构造的融合组播网络可以互联网络中的组播孤岛，可以支持分区分层流视频服务，具有规模扩展性和网络适应性等特点。

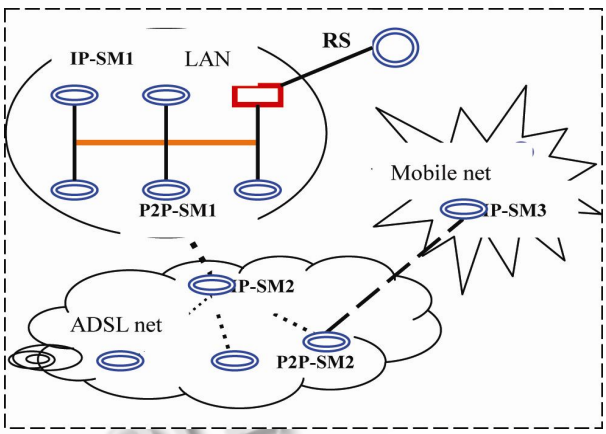

图 1　融合组播流视频系统

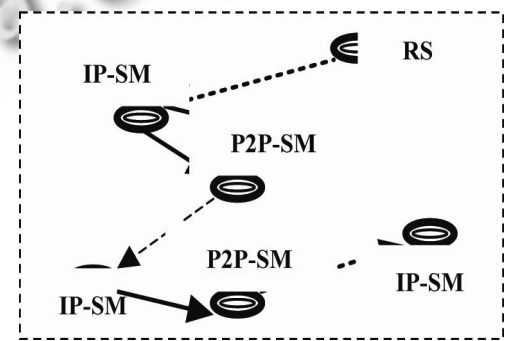

图 2　融合覆盖组播树

将图 1 中的 LM 成员屏蔽，得到图 2 所示的融合组播树，树的分支由成员 IP-SM 和 P2P-SM 交替构成，IP-SM 到 P2P-SM 分支是 IP 组播链，P2P-SM(含 RS)到 IP-SM 是应用层组播链，一个 P2P-SM 可以有多个分支，分支数量由它的综合性能决定。

© 中国科学院软件研究所 http://www.c-s-a.org.cn

节点加入算法：

(1)节点 A 首先 HELLO 到 RS，RS 根据 A 的网络号和质量等级需求等信息遍历 IP-MAL 表；

(2)若 IP-MAL 表中有 IP-SM 的网络号和节点 A 的网络号相同的项，且质量等级与需求一致，则令 A 向自己所在的网络区发出相应的组播请求，执行组播协议(如 IGMP 协议)，注册自己到 IP-MML 表中，启动客户端程序，算法结束；

(3)否则，RS 传回节点 A 所需的 IP-SM 节点集合(也可以包括 RS 自己)，根据 RS 提供信息，A 逐个发起 HELLO 请求，进行搜索，每一个得到这个 HELLO 信息的 IP-SM 必须回答，节点 A 根据回答，选择网络路由跳数最小的 IP-SM。选中的 IP-SM 遍历 IP-MM 表，在本区的 LM 中为节点 A 选择 P2P-SM，从而建立了一个新的 P2P 链，启动客户端程序，算法结束。

节点 A 退出算法：

(1)若节点 A 是 LM 成员，则向 IP-SM 发出含退出的 HELLO 报文，IP-SM 从 IP-MML 表中注销节点 A；

(2)若节点 A 是 P2P-SM 成员，则向 IP-SM 发出含退出的 HELLO 报文并携带其子节点 IP-SM 的信息，IP-SM 从 IP-MML 表中注销节点 A，再由 IP-SM 指定新的 P2P-SM；

(3)若节点 A 是 IP-SM 成员，则向 RS 发出含退出的 HELLO 报文并携带 IP-MML 表信息，RS 从 IP-MAL 表中注销节点 A，再由 RS 为该 IP 组播区寻找新的 IP-SM。

融合组播树的维护算法：

(1)每一个 IP 组播区需要运行必要的组播协议，除 IP-SM 外，每一个节点还要定期向 IP-SM 发送包含当前视频序列号的 HELLO 报文，若连续 2 次未收到节点 A 的 HELLO 报文，则启动节点 A(LM)的退出算法；

(2)每一个 IP-SM 要定期向父节点(P2P-SM 或 RS)发送含当前视频序列号的 HELLO 报文，若连续 2 次未收到这样的报文，则 P2P-SM 或 RS 停止其视频报文发送服务；

(3)每一个节点根据其收包情况，若得不到正常的数据服务，则根据其成员性质，重新请求组播服务；

IP-SM 选取算法：

在一个 IP 组播区内，IP-SM 除了要正常处理自己的视频接收和回放等工作外，还要为本组区组播视频数据和其他通信处理，所以要求节点的性能是越高越好。在开始一个新 IP 组播区时，第一个请求的点自然被 RS 登记为 IP-SM，随后本区若有更多的点加入，则由 IP-SM 和 LM 交流重新确定 IP-SM。

用式(1)来量化节点的综合性能，一些量化参数如接入带宽、CPU 处理能力及和父节点的路由长度等，把它们作为一个综合因素加以考虑并量化：

$$P(n)=q_1\times W(n)+q_2\times C(n)+q_3T(n)+q_4\times L(n) \quad (1)$$

其中，W(n)表示节点 n 的接入可用带宽。C(n)表示节点 n 的 CPU 可用处理能力。T(n)表示节点 n 的加入组播组的在线时间，刚加入的节点其值为 1，其后每一段时间会递增 1。T(n)值越高，说明该节点离开的意愿越低，稳定性好，拥有更高的优先级。因此，可以通过它来间接反映出该节点的稳定性。L(n)表示节点距离其候选父节点(RS 或 P2P-SM)的路由跳数。(q_1,q_2,q_3,q_4) 为权值，满足 $0<q_i<1$ 和 $q_1+q_2+q_3+q_4=1$。

P2P-SM 选取算法：

每一个 IP-SM 节点维持一个本区的 IP 组播区成员表(IP-MML)，从 IP-MML 中选择综合性能为优的 LM 作为 P2P-SM。综合性能的量化也使用公式(1)，但其中 $L(n)$ 表示节点距离其子节点(IP-SM)的路由跳数。

3 拥塞控制

融合组播树的拥塞控制或速率控制可能发生在 IP-SM 和 P2P-SM 上，目前系统规定除 RS 可以有多个子节点(IP-SM)外，其余每一个 P2P-SM 只允许有一个子节点 IP-SM。这样一来，所有的 SM 类节点都只要转发一份视频拷贝，极大地降低了拥塞的可能性。

3.1 IP-SM 的反馈风暴

反馈风暴是 IP 组播中的固有缺陷，建立融合组播体系除了扩大组播规模外，还可以预防反馈风暴的产生。但和 P2P-SM 相比，IP-SM 还要处理大量本区内的 HELLO 报文，尤其是当本 IP 组播区的成员数目达到一定规模后(如支持 IP 组播技术的校园网络)，可能存在反馈风暴问题。当上述情况发生时，需要一分为二拆分原 IP 组播区，其原则是按照物理网络分布属性(如 VLAN、子网等)来进行，这样形成 2 个 IP 组播分区后，原 IP-SM 的 HELLO 报文的处理会减少一倍，反馈风暴问题可得到解决。

反馈风暴的预防可以进行量化，可考虑的因素应包括：接入带宽的占用率或丢包率、CPU 处理能力的占用率或延迟性。当丢包率达到给定的阀值或视频包处理存在延迟时，就启动 IP 组播分区算法。

3.2 P2P-SM 的率控制

基于发送端的速率控制通过调整发送端的转发速率来匹配网络的可用带宽。P2P-SM 向它的子节点 IP-SM 转发视频数据包，IP-SM 周期性反馈含视频包

© 中国科学院软件研究所 http://www.c-s-a.org.cn

序列号的 HELLO 报文给父亲节点 P2P-SM。P2P-SM 可以从序列号中监测两个 SM 节点之间的网络状态。根据序列号来判断其丢包率，当达到给定的阀值时，启动率控制算法，根据接收端的带宽来调整一下发送视频的层数，如一次丢弃一个视频编码中的一个增强层，直到和目前网络带宽相一致。其后，若网络状况向好，则可以逐渐增加发送视频编码的层数。

4　系统实现

本节实现一个真实的流媒体融合组播网络，环境构建如图 3 所示，由 4 个网段构成，主机 B 作为 RS 发送源视频数据，主机 C、C1 和 F 具有双网卡和路由功能，同时启动主机 C 和 C1 的 IGMP 组管协议。实验模拟了 3 个 IP 组播区：{D，A，E}、{A1，D1}和{E1，B1}。组播系统建立过程如下：

- 主机 B 启动视频发送服务，主机 D 先请求主机 B 获得视频服务,主机 A 和 E 从主机 D 处获得组播服务，它们构成第一个 IP 组播区。
- 主机 A1 从主机 A 处获得 P2P 流媒体数据后和主机 D1 构成第二个 IP 组播区。
- 主机 E1 和 B1 分别启动手机模拟平台，从主机 D1 处只获得视频基本层编码数据后，和主机 B1 形成窄带 IP 组播区。

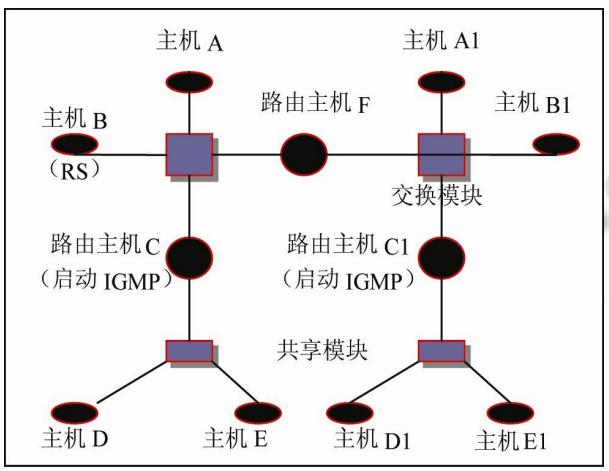

图 3　融合系统实验图

HELLO 报文使用 UDP 传输，所有参与组播的主机开放相同的 UDP 端口来收发 HELLO 报文。发送端用 RTP/RTCP 协议模拟 FGS 流视频发送。根据 FGS 编码视频 Thefirn[4]原始帧数据，增强层编码由 8 个位平面编码层构成(MSB0~MSB7)，编码类型有 I、P 和 B 三种帧，实验中根据这些数据的帧尺寸来模拟流视频发送，并使用了文献[5]提出的优化算法来发送视频，在接收端用一个缓冲区为 3MB 的模拟解码器来统计实验需要的性能数据(成员属性为 P2P-SM 和 IP-SM 的节点另增 5MB 缓冲空间)。

L(n)的值使用 TraceRT 命令来测定，节点 E1 和 B1 的 $W(n)=C(n)=5$，其余节点的 $W(n)=C(n)=10$，$T(n)$ 取值按照成员加入顺序来计数，初值为 1，当同区每当有新成员加入时原有成员的 $T(n)$ 增 1，考虑到都是同质主机，且实验网络规模小，所以，$T(n)$ 的权值取 $q=0.1$，其余各项的权值取 $q=0.3$。实验系统稳定后，各 IP 组播区的构成及各成员的属性是：{D(P2P-SM)，A(IP-SM)，E}、{A1(IP-SM)，D1(P2P-SM)}和{E1(IP-SM)，B1}。

实验中对主机 C 加上 NAT 功能，主机 D 和 E 成为私有网络，将实验中的第二个 IP 区改为：主机 A1 从主机 D 处获得 P2P 流媒体数据后和主机 D1 构成第二个 IP 组播区，这种环境下的融合网络构造过程中需要增加一个穿越 NAT 的功能模块。

5　总结

文章提出了一个融合组播的流媒体系统及相关的一些构造协议，并实现了这个系统，验证了融合组播流媒体是可行的。在理论上融合系统解决了 IP 组播中的孤岛及异构网络中组播适应性问题，但系统大规模化后的稳定性、容错性、拥塞控制等还需要花费大量的工作去做，这是我们今后要研究和完善的主要内容。

参考文献

1 Byers JW, Kwon GI, Luby M, Mitzenmacher M. Fine-grained layered multicast with STAIR. IEEE/ACM Transactions on Net- working，2006, 2(14): 81－93.

2 Finlayson R, Perlman R, Rajwan D. Accelerating the deployment of multicast using automatic tunneling. Internet Draft, IETF, 2001(2).

3 Zhang BC, Jamin S, Zhang LX. Host Multicast: A Framework for Delivering Multicast to End Users. Proceedings of INFOCOM, 2002.

4 Cuetos P, Ross KM. Adaptive rate control for streaming stored fine-grained scalable video. Proce- edings of NOSSDAV, Miami, FL, USA, 2002.

5 谢建国,姜灵敏,陈松乔. VBR流式视频的最短路径率平滑传输算法.计算机学报, 2004, 27(3):357－364.

© 中国科学院软件研究所 http://www.c-s-a.org.cn