分层自治的Multicast地址管理和连接控制*

王 箭 张福炎

(南京大学计算机科学与技术系 南京 210093)

E-mail: wangjian@graphics.nju.edu.cn/fyzhang@nju.edu.cn

摘要 Multicast动态地址管理和连接控制是多点会话中的重要问题.基于分层自治结构,文章提出了Multicast地址分配管理机制、主从结构连接管理体系和一种简单的多点连接建立算法,简述了它们的工作过程,并通过模拟实验比较了3种地址分配方式,即集中管理方式、分布方式(由Eleftheriadis提出)与分层自治方式.分层自治结构与Internet自组织拓扑结构一致,分层自治地址分配机制结合了集中方式和分布方式的优点,具有较高的整体效率,主从结构连接管理体系也具有较高的控制效率,它们都具有较高的鲁棒性、柔韧性和伸缩性.

关键词 自治系统,Multicast地址管理,连接控制,多点应用,Multicast通信,图,Steiner树,递归算法,阻塞概率,Time-out周期.

中图法分类号 TP393

多点会话需定义Multicast地址[1].与Unicast网络地址不同,它可动态分配.一个Multicast地址不能分配给两个并发会话,否则将因互相干扰而引起混乱.目前,IP环境还不具有鲁棒性和伸缩性的动态分配策略,为建立会话,用户必须预先交换所有相关参数,这些阻碍了多点会话的应用和开发,也增加了用户网络管理负担.多点会话的增长要求有效的动态地址分配策略.

Multicast传输协议,如XTP,ST-II,MTP和RTP[2]等,都假定存在一个分配与管理Multicast地址的独立实体.Braudes[3]提出了Multicast地址管理框架,地址由分层的MGA(multicast group authority)进行管理,中心控制器作为管理树的根结点,靠近根的结点将承担大量的控制负载,且如果中间结点或链接失效,上下两边结点就不能交换Multicast地址.Eleftheriadis[4]提出基于网络的分布Multicast地址分配(network-divided mode,简称NDM)和连接管理机制,没有利用网间结构关系有效地解决阻塞概率与会话建立延迟之间的矛盾,且Multicast地址空间划分不完备.Pejhan[5]基于主机的地址分配机制,阻塞概率和会话建立延迟大大降低,但难于实现有效的连接控制,并需调整路由和传输协议,消除干扰需耗费大量的计算资源,也将大大增加网络负载.

广播型应用不需连接控制,用户之间松散耦合,称为开式多点通信.闭式多点通信是指高交互应用,用户之间紧密耦合.地址分配应满足这种需要,为实时管理多点连接,需要连接控制,它负责与会方之间连接建立和管理.

本文第1节介绍分层自治域结构.第2节是Multicast地址分配.第3节是连接控制和Time-out机制.第4节是简单的多点连接建立算法.第5节是失效处理与性能分析.第6节是模拟实验.最后是总结.

1 分层自治结构

INTERNET是由若干网络组成的全球性网络,可看成是由分层自治域构成的,如CERNet由8个区域网络组成,区域网络包含本地区若干高校,又通过国际线路与INTERNET连接,CERNet可看成一个自治域,区域网

---

* 作者王箭,1968年生,博士,助教,主要研究领域为多媒体通信,CSCW,视频会议.张福炎,1939年生,教授,博士生导师,主要研究领域为多媒体技术,计算机图形学.

本文通讯联系人:王箭,南京 210093,南京大学计算机科学与技术系

本文1997-12-23收到原稿,1998-09-07收到修改稿

© 中国科学院软件研究所 http://www.jos.org.cn

络是子自治域,校园网是更低一层的自治域.自治域是个整体,知其接口就可与域内主机通信.它应包括一个IPAM(IP address manager)、一个MAM(multicast address manager)、若干CCs(connection controllers)、一个ICM (inter-domain connection manager)以及若干路由结点.IPAM记录IP地址范围并检测IP地址是否在本域内.MAM负责Multicast地址管理,每个MAM都有一个MAS(multicast address set).子域MAS是父域的子集,同父的任意两个子域MAS不相交.只有底层域(bottom-layer domain,简称BLD)MAM向用户分配Multicast地址,并负责回收.BLD可以是网段、LAN或ISP管理域等.ICM动态或静态管理本域与其他域的连接,它构成了INTERNET的连接框架.分层自治结构如图1所示.

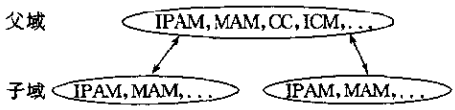

图1 分层自治结构

多点会话在最小包容域(minimal magnanimous domain,简称MMD)中进行,MMD是指包含所有与会点的最小自治域,不一定是BLD,地址由BLD分配给用户,需建立多点连接(multipoint connection,简称MC),MC跨越相关域并将包含与会点的所有BLD连接起来.

2 Multicast地址分配

Eleftheriadis[4]按照网络地址来划分Multicast地址空间,若网络或子网地址是A1.A2.A3,则地址范围是[224～239].A1.A2.A3,若为A1.A2,则地址范围为[224～239].A1.A2.X,X=[0,255].这种方式实现简单,但存在一些不足.

- 划分不完备.由于网络地址第1字节的特殊规定及它们没有都被使用,使得大量Multicast地址被遗漏.
- 地址不能被充分使用.静态划分既不能充分有效地使用地址资源,也不能满足用户的需要.

本文按域划分地址,建立分层域之间的划分关系,即子域MAS是对父域MAS的划分,大大提高了伸缩性和柔韧性.但划分策略还有待进一步研究.BLD的MAM向多点会话分配Multicast地址,上层域MAM负责MAS的静态与动态管理.预先设定各子域MAS,子域地址用尽且有请求时,父域MAM可将直接管理的或其他子域的空闲地址借与该子域,会话结束时借用地址需归还.MAM记录本域Multicast地址的使用情况,若借用频率大于预定门限,则向父域MAM申请增加Multicast地址,父域根据一定的策略选择直接管理的或将其他子域富余的地址授权给该子域.

Multicast地址分配的大致过程(如图2所示)如下.

(1) 若BLD的MAM有空闲地址,分配之,并转(3);否则,设BLD为当前域;

(2) 若向父域MAM借用地址成功,分配之,并转(3);否则,设父域为当前域,并转(2);

(3) 在BLD中选择一个CC,将Multicast地址和与会方地址等传送给CC;

(4) CC建立多点连接与连接控制的分层结构;

(5) MAM收到CC确认信息后,将Multicast地址等会话信息通知用户.

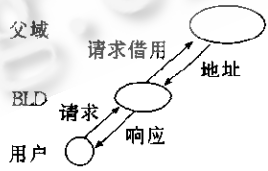

图2 地址分配

3 多点连接控制

3.1 连接控制体系

CC建立、管理和维护会话的多点连接,交换状态信息,管理与会结点的进入和退出等.将地址管理与连接控制分开,主机分布操作,进程和主机失效处理的鲁棒性获得显著提高.在BLD中由SCC(secondary connection controller)承担局部连接控制功能以实现分布管理,有效地处理连接状态和拓扑变化.

MAM在该BLD中选择一个CC.该CC根据与会点地址建立多点连接,并在MMD中选择一个CC作为MCC(master connection controller),而后,MCC在含有与会点的BLD中选择SCC(MAM选择的CC成为所在域的SCC),就建立了连接控制的主从结构.CCs都包含所有会话相关信息,如连接、会话状态以及MAM地址等,并始终保持一致.MAM需包含MCC和所在BLD的SCC地址.

© 中国科学院软件研究所 http://www.jos.org.cn

主从结构连接控制的优点是：(1) 与分层自治结构一致；(2) 与会点可较快获得会话信息；(3) 较高的鲁棒性，少数 CC 失效不会影响会话正常进行，并可较快恢复；(4) MCC 到 SCCs 的 hop 数大致相等，简化 Time-out 估算；(5) 可利用建立的多点连接来交换信息.

当用户加入时，若该 BLD 中有属于该会话的 SCC，则该 SCC 连接操作后通知 MCC，否则，由 MCC 建立该 BLD 的 SCC，然后由 MCC 通知其他 SCC. 当用户退出时由相连的 SCC 进行处理，若 SCC 管理的用户数为 0，则该 SCC 从会话中退出，卸下会话数据，并通知 MCC. MMD 改变时，有两种处理方法：① MCC 不变，使 Time-out 估算变得复杂，但 CCs 和 MAM 会话数据只作较小调整；② 在 MMD 中重新选择 MCC，Time-out 估算简单，但使 CCs 和 MAM 的会话数据作较大调整，不适合 MMD 频繁变动的情况. 折衷的办法是建立门限，使相连前后两次 MCC 变动的时间间隔大于该门限.

3.2 Time-out 机制

为实现系统的鲁棒性，MCC 与 SCCs 之间和 SCC 与用户之间定期交换 Keep-alive 信息. 首先由 MCC 发出消息，SCCs 收到后附上本地信息传给用户，SCCs 开始收集用户响应，收齐后就向 MCC 传送响应(如图 3 所示). 为避免冲突，将消息编号. MCC 与 MAM 之间也定期地交换信息，只是频率会较低.

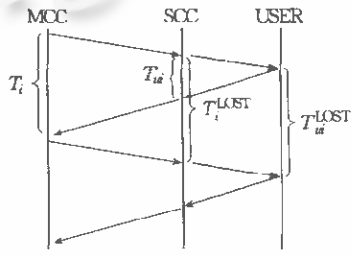

图3　Time-out机制

MCC 与 SCCs 之间的 hop 数相等，MCC 与与会点之间的 hop 数也大致相等，使 RTT_c 值大致相等，MCC 对每一会话维持一个 μ_{RTT_c}(RTT_c 平滑估计值)和 σ_{RTT_c}(RTT_c 方差)值. Time-out 值初始设为 $T_i=\mu_{RTT_c}+2_{\alpha}\sigma_{RTT_c}$，超过这个值，响应仍未收到，Time-out 值就增加一倍. 达到预定补偿次数 L 或最大值 T_{max} 时，没有响应的 SCC 可看做失效. 为便于 CCs 和 MAM 从失效中恢复，设置 Time-out 的上限是必要的，Time-out 下限也需设置，以使小 RTT 会话保持较低控制通信量.

SCC 端也设置一个 Time-out$_c$ 值，在此时间内没有再次收到 MCC 的消息，可认为 MCC 失效. 该值设为 $T_i^{LOST}=\min\{T_{max},\sum_{i=0}^{L}2^i(\mu_{RTT_c}+2_{\alpha}\sigma_{RTT_c})\}$. 同时，SCCs 还应设置一个 Time-out$_{\alpha}$ 值，以判断用户是否失效. 当收到 MCC 信息时，时间计数器启动. 该值设为 $T_{ui}=\mu'_{RTT_c}+2\sigma'_{RTT_c}$，当超过这个值，用户响应仍未收到时，该值就增加一倍. 达到预定补偿次数 L' 或最大值 T'_{max} 时，没有响应的用户可看做失效.

为判断 SCC 是否失败，用户端也应建立一个 Time-out$_u$ 值，在此时间内没有再次收到 SCC 的信息可认为 SCC 失效. 这个值设为 $T_{ui}^{LOST}=\min\{T_{max}+T'_{max},\sum_{i=0}^{L}2^i(\mu_{RTT_c}+2\sigma_{RTT_c})+\sum_{i=0}^{L'}2^i(\mu'_{RTT_c}+2\sigma'_{RTT_c})\}$.

4　多点连接(MC)建立算法

视频会议是一种闭式多点会话，由于与会方之间实时交互的需求，需要有效地建立 MC，连接本身应是高效柔韧的，连接建立算法应简单、快速. 基于分层自治结构，本文提出一个简单、高效的 MC 构成算法，如图 4 所示.

1. 寻找 MMD
查询本地 BLD 的 IPAM，
看所有与会点的 IP 地址是否在本域内；
若是，则本域即为 MMD；
否则，则将上级域指定为当前域；
while (当前域不包含所有与会结点)
　{当前域的上级域指定为当前域；
　查询当前域的 IPAM；}
当前域即为 MMD.
记 MMD 为 W_0，联接各子域的路由结点为 r_0.
IF (MMD 是 BLD), THEN RETURN.

2. 建立 MC
(1) 将与会结点集分成若干子集，分别记为 A_1，$A_2,\ldots,A_{k(0)}$，$A_i\neq\varnothing$，$A_i\cap A_j=\varnothing$，使 $A_i\in W_i$ (W_0 子域). 记 W_i 中与 r_0 直接相连的路由结点 r_i，并建立 r_0 与 r_i 之间的连接.
(2) 再分别将 A_i 分成子集为 $A_{i1},A_{i2},\ldots,A_{ik(i)}$，$A_{ij}\neq\varnothing$，$A_{i1}\cap A_{ij}=\varnothing$，使 $A_{ij}\in W_{ij}$ (W_i 子域). 记 W_{ij} 中与 r_i 直接相连的路由结点 r_{ij}，并建立 r_i 与 r_{ij} 之间的连接.
(3) 如此往复直至 BLD，这样就建立了具有中心路由结点的多点连接.

图 4　基于分层结构的 MC 建立算法

查询相应域 IPAM，确定是否有与会点，并查询 ICM 以建立包含与会点的域间连接以及与会点之间的 MC.

© 中国科学院软件研究所 http://www.jos.org.cn

由于BLD通常是局域网,则无需考虑MC问题.

5 失效处理与性能分析

5.1 失效处理

(1) MCC失效.当SCCs超过预定时间(Time-out$_c$)还没收到MCC消息时,就向分配Multicast地址的MAM发送"MCC不响应"的消息,MAM通知同BLD的SCC卸下MCC地址,在MMD中寻找一个CC作为MCC,将会话数据拷贝到新的MCC之中,将新MCC地址等会话数据传给SCCs,建立MCC与SCCs之间的连接.若MAM检测到MCC失效,采用同样方法恢复连接控制.选择新MCC并恢复状态,对应用是透明的,并不引起任何通信的混乱.当MCC失效时,用户可以离开,有时也可以加入(当新加入者BLD有该会话的SCC).

(2) SCC失效.当用户或MCC超过预定时间(Time-out)没收到SCC的信息时,就向MCC发送"SCC不响应"消息,MCC在相应的BLD中寻找一个CC作为该会话的SCC.

(3) MAM失效.BLD的MAM失效,会话继续进行,因为只有CCs负责连接控制.但对Multicast地址请求将不会响应.该MAM重新启动后,向父域MAM注册并查询本域所有CC,得到回答后,MAM将相应CC标为在用,并将Multicast地址和MCC地址相联系.非BLD的MAM失效不影响Multicast地址分配,但对子域地址借用请求不响应,重新启动后,向父域MAM注册并查询子域MAM.

5.2 性能分析

基于分层自治结构地址分配机制,既有集中处理的低阻塞概率,也有分布处理的较小会话建立延迟,结合了集中和分布机制两者之长.地址请求由整个Multicast地址空间来服务,其阻塞概率相当于集中方式.地址请求首先由BLD服务,只有当本域地址用尽时,才向上层域请求,会话建立平均延迟接近于分布机制.

分层自治结构不但较好地解决了地址动态分配问题,而且使多点连接较易建立.通过查询相应域,IPAM和ICM可以很方便地将相应的BLD连接起来.域有外界接口,通过分层结构组织起来,可以通过少量数据表达出全局信息.

主从结构连接控制体系具有较高的控制效率和鲁棒性.BLD往往是采用同一网络协议的局域网,用户与SCC之间的消息交换是快速和高效的,连接控制的分布管理提高了控制效率.主从结构使得多点会话在广域环境中可以实现高效连接控制.少数CC失效不会影响整个会话的进行,而只影响局部用户,只有当大部分CCs失效时才会严重影响会话.

6 模拟实验

本文通过模拟实验,比较了3种Multicast的地址分配方式,即集中管理方式(central mode,简称CM)、Eleftheriadis[4]的NDM(network-divided mode)和分层自治方式(hierarchical autonomous mode,简称HAM).

令$s(0)=r_0(0)$且$s(t)=s(t-1)+r_0(t)$.以随机函数$s(t)$来模拟Multicast地址请求,若$s(t)>a(a\geqslant 1)$,则申请Multicast地址且令$s(t)=0$.同时,用$g_i=100*r_1(t)$来模拟多点会话持续时间.设定一个网络,包含3个子网.CM只有一个MAM,NDM有3个MAM,分别在3个子网中,HAM建立两层域,即顶层域和3个子域,每个域都有一个MAM.为简化对会话建立时间的模拟,对网络和子网范围分别设定会话建立时间t_1和t_2($t_1>t_2$).令$t_1=0.2$,$t_2=0.1$.令$h=3*r_2(t)$,若$i-1<h<i(i=1,2,3)$,则在第i个子网中.结点总数为600,每个子网有200结点,可用地址数为42,$a=1$.r_0,r_1与r_2是随机函数.

每次实验持续1小时,每秒钟做一次计算,判断是否有地址请求和会话终止,会话持续时间以秒为单位.记录每个MAM的地址请求数和成功分配数并相加得到总请求数n_0和成功会话数n_1,则阻塞概率$P_B=1-n_1/n_0$,平均会话建立时间T_s,CM为t_1,NDM为t_2,HAM需区分n_1中借用与没有借用地址的数目,设分别为n_2与n_3,则HAM为$(n_2*t_1+n_3*t_2)/n_1$.每种方式分别进行6次实验,实验结果如表1和表2所示.P_B单位为%,T_s单位为秒(s).

结果显示,CM的阻塞概率最小,NDM的平均会话建立时间最小,对于这两个指标,HAM都接近最小值,若

© 中国科学院软件研究所 http://www.jos.org.cn

选择适当地址划分算法则几乎等于最小值,因此,HAM 的总体性能最好.

表 1　实验数据

	HAM				CM		NDM	
	N_0	N_1	N_2	N_3	N_0	N_1	N_0	N_0
1	3 011	3 008	301	2 707	2 869	2 867	3 002	2 987
2	2 703	2 700	540	2 160	3 300	3 298	2 881	2 868
3	3 302	3 299	297	3 002	2 857	2 855	2 909	2 983
4	2 730	2 727	191	2 536	3 310	3 308	2 998	2 986
5	3 005	3 002	240	2 762	2 005	2 004	2 888	2 875
6	2 705	2 702	270	2 432	2 850	2 848	3 009	2 997

N_0:总请求数,N_1:成功会话数,N_2:借用地址数,N_3:没有借用地址数

表 2　实验结果

	1		2		3		4		5		6		A	
	P_B	T_s	P_B	T_s	P_B	T_s	P_B	T_s	P_B	T_s	P_B	T_s	P_B	T_s
HAM	0.1	0.11	0.11	0.12	0.09	0.109	0.11	0.107	0.1	0.108	0.11	0.11	0.103	0.117
CM	0.07	0.2	0.06	0.2	0.07	0.2	0.06	0.2	0.05	0.2	0.07	0.2	0.063	0.2
NDM	0.5	0.1	0.45	0.1	0.55	0.1	0.4	0.1	0.45	0.1	0.4	0.1	0.458	0.1

P_B:阻塞概率,T_s:平均会话建立时间,A:平均值

7　总　结

分层自治结构与 INTERNET 网络的体系结构较为接近.基于分层自治结构,本文提出了一种 Multicast 地址分布管理和分配策略,有效地降低了地址分配阻塞概率,并减少了平均会话建立延迟,充分结合了集中和分布两种方式之长,是高效可伸缩的,而且具有鲁棒性.极端情况最小包容域可以达到整个 INTERNET,并不影响地址分配阻塞概率,而只是使连接建立时间增加,与其他方式相比,依然具有较高的效率.主从结构连接控制解决广域环境多点会话的广域连接控制与局部地址分配之间的矛盾,并且充分利用分层自治结构特性,提出了一个简单高效的多点连接建立算法.

参考文献

1　Deering S. Host Extensions for IP Multicasting. RFC1112, Stanford University, 1989

2　Steinmetz R, Nahrstedt K. Multimedia Computing, Communications & Applications. Beijing: Tsinghua University Press, 1997

3　Braudes R, Zabele S. Requirements for Multicast Protocol. RFC1458, TASC, Reading, MA, May 1993

4　Eleftheriadis A, Pejhan S, Anastassiou D. Address management and connection control for multicast communication application. In: Proceedings of the 14th IEEE INFOCOM'95 Conference on Computer Communications. Boston: Computer Society Press, 1995. 386～393

5　Pejhan S, Eleftheriadis A, Anastassiou D. Distributed multicast address management in the global internet. IEEE Journal on Selected Areas in Communications (special issue on the global internet), 1995,13(8):1445～1456

Hierarchical Multicast Address Management and Connection Control

WANG Jian　ZHANG Fu-yan

(*Department of Computer Science and Technology　Nanjing University　Nanjing*　210093)

Abstract　Multicast address management and connection control are two essential components needing to be solved in the multipoint applications. Based on hierarchical autonomous structure in accordance with the

© 中国科学院软件研究所 http://www.jos.org.cn

self-organization topologies of INTERNET, a multicast address management scheme which sets up the hierarchical partition is put forward in this paper, which is dynamic, of Multicast address space in order to reduce the blocking probability. A connection control hierarchy (CCH) based on master/slave relationship and a simple efficient building algorithm of multi-point connection are also advanced. The normal operations of multicast address management and multi-point connections controller are also described. With simulation experiment, three of multicast address allocation modes, hierarchical autonomous mode (HAM), central mode (CM) and network-divided mode (NDM), are compared. The result shows that the hierarchical autonomous mode integrates the merits of central mode (CM) and network-divided mode (NDM), owns high efficiency in the whole. Connection control hierarchy (CCM) is also efficient in the connection control. They are shown to be highly robust, flexible and scalable.

Key words Autonomous system, multicast address management, connection control, multipoint application, multicast communication, graph, Steiner tree, recursive algorithm, blocking probability, time-out periods.

© 中国科学院软件研究所 http://www.jos.org.cn