

组织：中国互动出版网 (<http://www.china-pub.com/>)

RFC 文档中文翻译计划 (<http://www.china-pub.com/compters/emook/aboutemook.htm>)

E-mail: ouyang@china-pub.com

译者：田金勇 ([tany tjy308@263.net](mailto:tany_tjy308@263.net))

译文发布时间：2001-6-15

版权：本中文翻译文档版权归中国互动出版网所有。可以用于非商业用途自由转载，但必须保留本文档的翻译及版权信息。

Network Working Group
Request for Comments: 1112
Obsoletes: RFCs 988 , 1054

S. Deering
Stanford University
August 1989

主机扩展用于 IP 多点传送

(RFC1112 Host Extensions for IP Multicasting)

1. 备忘录状态

该备忘录说明了为了支持多播而对互联网协议 (IP) 的主机实现的扩展。它是互联网上 IP 多播的推荐标准。该备忘录可以任意发布。

2. 简介

IP 多播是指一个 IP 数据报向一个主机组的传送，该主机组是由一个单独的 IP 目的地址标记的多个或零个主机。一个多播数据报被尽可能地传递给它的目的主机组的所有成员，就像常规的单播 IP 数据报一样。也就是说，不能保证数据报能完好无损的到达目的组所有成员，也不能保证它以相对于其它数据报同样的顺序到达。

主机组的成员是动态变化的；也就是说，主机可以随意加入和离开组。对一个主机组的位置和组成员的数量并没有限制。一个主机可以同时是一个或多个组的成员。不是某一组的成员的主机也能向该组发送数据报。

主机组可以是永久的或是暂时的。永久组是一个众所周知的，由 Internet 管理机构分配的 IP 地址，它的地址是永久的，而该组的成员是可以改变的。在任意时刻，一个永久组可以有任意数量的成员，甚至没有成员。另外一些不是为永久组保留的 IP 多播地址是为暂时组动态分配的地址，这些暂时组只有当它们有成员时才存在。

IP 多播数据报的网间传递是通过“多播路由器”实现的，多播路由器可以驻留在互联网网关上，也可以与互联网网关分离。主机以本地网络多播的方式传送 IP 多播数据报，这些数据报只到达目的主机组中所有与该主机直接邻近的成员。如果数据报的 IP 首部中的生存时间(TTL)字段大于 1，则在本地网络上的一个或多个多播路由器负责将它传递到有目的主机组的成员的其他网络上。在那些可以在 IP 生存时间为零以前到达的其它成员网络里，

当地一个多播路由器以本地多播的方式传递该数据报来完成传送。

该备忘录说明了为了支持 IP 多播而对主机的 IP 实现的扩展，这里的“主机”是除了用作多播路由器以外的任何互联网主机和网关。用在多播路由器内和之间的算法和协议对主机是透明的，将在独立的文档中说明。该备忘录也不说明本地网络多播是如何在所有不同类型的网络中实现的，尽管它说明了任意的本地网络所需的服务界面，并给出了一个以太网技术要求说明作为例子。其他类型的网络技术说明将会在将来的备忘录中说明。

3. 一致性等级

这个规范有三个一致性等级：

0 级：不支持 IP 多播。

在这一等级中不需要所有的 IP 实现支持 IP 多播。0 级主机一般不受多播活动的影响。在一些类型的本地网中，由于 1 级和 2 级主机的存在可能导致多播 IP 数据报错误递交到 0 级主机上，这就出现了唯一的异常。因为这些数据报目的地址字段中出现 D 类 IP 地址，所以它们能被很容易地被检测出来。它们会被不支持 IP 多播的主机毫无声息的丢弃。D 类地址在本备忘录的第四部分说明。

1 级：能发送但不能接收多播 IP 数据报。

1 级主机可以参加一些基于多播的服务，诸如资源定位或状态报告，但它不允许主机加入任何主机组。IP 实现可以非常容易地从 0 级升到 1 级，仅需增加少量的代码。在本备忘录中仅仅第 4、5、6 部分适用于 1 级的实现。

2 级：完全支持 IP 多播。

2 级主机可以加入或离开主机组，以及发送数据报给主机组。它要求 Internet 组管理协议 (IGMP) 的实现和在主机内对 IP 和本地网服务接口的扩展。本备忘录中以下所有部分都适用于 2 级的实现。

4. 主机组地址

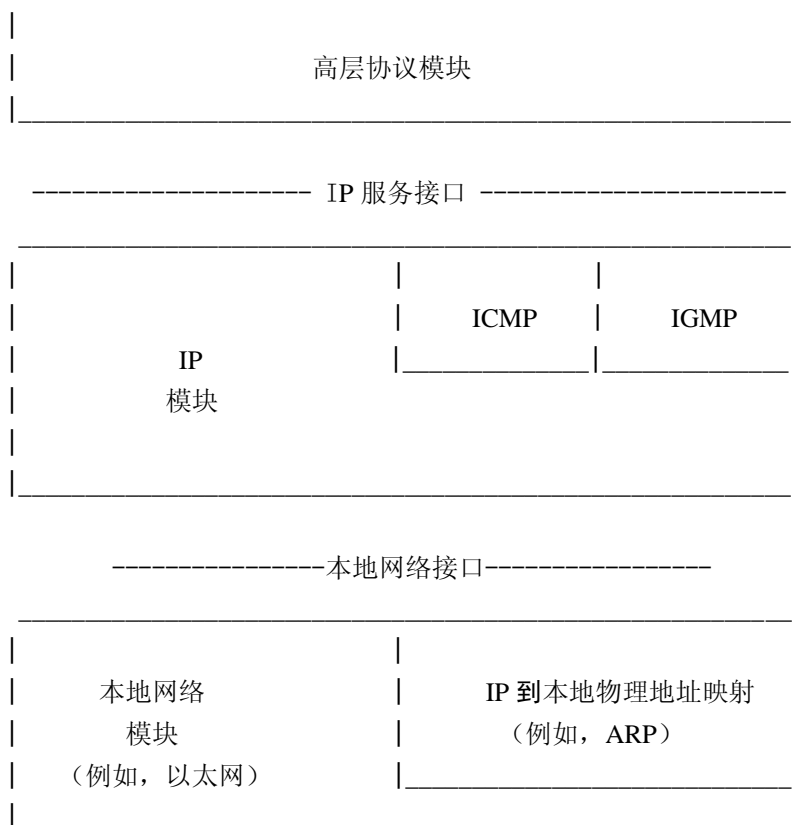
主机组由 D 类 IP 地址标记，即高四位为“1110”的那些 IP 地址。E 类 IP 地址，即那些高四位为“1111”的 IP 地址，是为了将来的编址方式而保留的。

在 Internet 标准的点分十进制表示中，主机组地址的范围是从 244.0.0.0 到 239.255.255.255。地址 224.0.0.0 被保证不分配给任何组（包括网关）。它被用于标记在直接相连的网络中所有多播主机。没有多播地址（或其它 IP 地址）用来标记 Internet 上的所有主机。其它众所周知的地址、永久组将在“已分配编号”（Assigned Numbers）文档中公布。

附录 2 包括与主机组地址相关的一些问题的背景的讨论。

5. 主机 IP 实现模型

对主机 IP 实现的多播扩展将在下面的层次模型图中说明。在这种模型中, ICMP 和 IGMP (对于 2 级主机) 应在 IP 模块中实现, IP 地址到本地网络地址的映射应由本地网络模块负责。这种模型仅用于说明目的, 而不应该认为它限制了实际的实现。



为了达到 1 级多播, 主机实现必须支持多播 IP 数据报的传送。为了达到 2 级多播, 主机必须还需支持多播数据报的接收。这两个新服务将在以下各自独立的章节中分别讲述。对于每种服务, 将分别在 IP 服务接口、IP 模块、本地网络服务接口、以太网本地网络模块中说明需要的扩展。对于非以太网的本地网络模块的扩展只是简要地提及, 而不详细说明。

6. 发送多播 IP 数据报

6.1 IP 服务接口的扩展

多播 IP 数据报的发送和单播 IP 数据报的发送一样的, 都使用“发送 IP”操作; 高层协议模块仅仅指定一个 IP 主机组地址, 而不是单个 IP 地址, 作为目的地址。然而, 许多扩展可能是必须或是可取的。

首先, 服务接口必须提供一种方式, 让高层协议能够指定要送出的多播数据报的 IP 生存时间 (TTL) 字段的值, 如果这种能力以前并不存在的话。如果高层协议没有指定生存时

间，那么对所有多播 IP 数据报，生存时间（TTL）缺省为 1。因此，要让多播跨过单个网络，就必须明确指定生存时间。

第二，对于那些属于多个网络的主机，服务接口必须提供一种方式，能让高层协议标记哪个网络接口将被用于多播传送。仅有一个接口用于初始传送；如果需要的话，多播路由器负责向所有其它网络传递。如果高层协议没有标记一个外出接口，缺省接口将被使用，这可由系统管理来控制。

第三（这仅用于 2 级实现），如果发送数据报的主机本身是它发向的主机组的成员，那么服务接口应该提供一种方式，让高层协议禁止这个数据报的本地传送；缺省情况下，该数据报的一个拷贝被回送。这对于限制了组成员关系只能是每个主机一个进程（例如路由选择协议）的高层协议，或对于在更高协议上处理组通信回送（例如多播传输协议）的高层协议，这是一种性能优化。

6.2 IP 模块的扩展

为了支持多播 IP 数据报的发送，IP 模块必须被扩展以在为外出数据报选择路由时识别 IP 主机组地址。大多数 IP 实现包括以下逻辑：

```
if IP 目的地在同样的本地网络上
    在本地发送数据报给 IP 目的地
else
    在本地发送数据报给网关（IP 目的地）
为了达到多播传送，路由选择逻辑必须改为：
if IP 目的地在同样的本地网络上
or IP 目的地是一个主机组
    在本地发送数据报给 IP 目的地
else
    在本地发数据报给网关（IP 目的地）
```

如果发送主机本身是外出接口上目的的主机组的成员，则外出数据报一个拷贝必须回送来达到本地传送，除非由发送者禁止。（仅用于 2 级实现）

外出数据报的源地址必须是与外出接口相对应的单个地址中的一个。

主机组地址一定不会出现在外出 IP 数据报的源地址字段，和源路由或记录路由选项中的任何地方。

6.3 本地网络服务接口的扩展

为了支持多播 IP 数据报的发送，对本地网络服务接口没有任何改变。当 IP 模块调用现存的“send local”操作时，只不过指定一个 IP 主机组目的地址，而不是单个 IP 目的地址。

6.4 以太网本地网络模块的扩展

通过允许多播地址出现在以太网分组中的目的地址字段，以太网就能支持本地多播分组的发送。为了支持多播 IP 数据报的发送，所需的是一个将 IP 主机组地址映射到以太网多播地址的过程。

IP 主机组地址映射到以太网多播地址,是通过将 IP 地址的低 23 位放在以太网多播地址 01-00-5E-00-00-00(十六进制)的低 23 位上。因为在 IP 主机组地址上有 28 个有意义的位,所以多个主机组地址可以映射到相同的以太网多播地址上。

6.5 非以太网的本地网络模块的扩展

其它的那些本身就支持多播的网络,例如符合 IEEE 802.2 标准的环和总线网络。它们可以像以太网那样达到发送多播 IP 数据报的目的。对于那些支持广播而不支持多播的网络,例如实验性以太网,所有的 IP 主机组地址可以被映射为单个本地广播地址(以增加所有本地主机费用为代价)。对于连接两个主机(或者一个主机,一个多播路由器)的一个的点对点连接,多播应该完全像单播那样传递。对于像 ARPANET 这样的存储转发网络,或公共的 X.25 网络,所有 IP 主机组地址必须映射为 IP 多播路由器的众所周知的本地地址;在这些网络上路由器负责完成在一个网络内或所有网络中的多播传递。

7. 接收多播 IP 数据报

7.1 对 IP 服务接口的扩展

高层协议使用与接收标准的单播数据报一样的“Receive IP”操作,接收到来的多播 IP 数据报。目的高层协议的选择由 IP 首部的协议字段决定,而与目的 IP 地址无关。然而,在属于一个特定组的任意数据报被接收之前,高层协议必须请求 IP 模块加入该组。因此,必须扩展 IP 服务接口,提供以下两个操作:

```
JoinHostGroup(group-address, interface)
```

```
LeaveHostGroup(group-address, interface)
```

JoinHostGroup 操作要求本主机在给定的本地网络接口上成为以“group-address”标记的主机组中的成员。LeaveGroup 操作要求本主机在给定的本地网络接口上离开以“group-address”标记的主机组。如果主机只有一个接口,则 interface 参数可以省略。对于属于多个网络的主机,高层协议可以不指明 interface,这要求使用缺省的 interface 发送多播数据报(参看 6.1 节)。

允许在多个 interface 上加入一个相同的组,这种情况下可能会有重复的多播数据报被接收;也允许多个高层协议要求加入同一个组。

这两种操作都立即返回(即它们是非阻塞操作),表示成功或是失败。每种操作都可能因为无效的组地址或接口标识符而出错。JoinHostGroup 操作可能由于缺少本地资源而失败。LeaveHostGroup 可能因为本主机不属于在给定的接口上的某一组而失败。LeaveHostGroup 如果成功,组成员关系可能还存在,因为可能有多个高层协议是同一个组的成员。

7.2 IP 模块的扩展

为了能够接收多播 IP 数据报,必须扩展 IP 模块,维持一个与每个网络接口相关的主机组成员列表。在到来的数据报中,目的地属于这些组地址中某一个的数据报的处理方式与目的地属于主机的单播地址中某一个数据报的处理方式相同。

如果主机接收到目的地址是组地址的数据报，而该主机不属于该组，则这个数据报将被丢弃而不生成任何错误或日志记录。在有多个网络接口的主机上，如果目的地属于某一组的数据报经由一个接口到达，而该主机只在另一个网络接口上属于这个组，则这个数据报将被毫无声息的丢弃。（出现这种情况的唯一原因是本地网络模块的多播地址过滤功能不完善。）

如果到来的数据报的 IP 生存时间 (TTL) 为 1，则该数据报不会被拒收（即，如果到来的数据报不会再被转发，则它的生存时间字段不会自动减少）。如果到来的数据报的源地址字段出现了 IP 主机组地址，则该数据报将被毫无声息的丢弃。ICMP 出错信息（目的主机不可达，超时，参数问题，源站被抑制，重定向）不会因为目的地为 IP 主机组的数据报而产生。

主机组成员列表根据高层协议的 JoinHostGroup 和 LeaveHostGroup 操作要求而更新。每个成员应该有一个相应的参照数或其它同样的机制，来处理加入和离开同一组的多个要求。当第一个要求加入和最后一个要求离开在给定接口上的某个组时，那个接口的网络模块就被告知，去更新它的多播接收过滤器（参看 7.3 部分）。

IP 模块也必须扩展，实现在附录 1 中说明的 IGMP 协议。IGMP 是用来让邻近的多播路由器获知在一个特定的本地网络中出现的主机组成员。为了支持 IGMP，每个 2 级主机必须在初始化时加入“所有主机”组（地址为 224.0.0.1），而且只要主机是开着的，它都要是这个组的成员。

（地址是所有主机组的数据报会被多播路由器当成一个特殊情况，不管它的生存时间是多少，它不会被传递出单一的网络。因此，所有主机地址不用于互联网范围上的广播地址。为了达到 IGMP 的目的，属于所有主机组的这项要求仅仅在主机属于至少一个其它组时才真正是必须的。然而，要求主机在任何时候是一个所有主机组的成员的原因是(1)简单，(2)接收不必要的 IGMP 请求的频率应足够的低，以使开销可以忽略不计。(3)所有主机地址可以用于其它面向路由选路的目的，例如通知网关的存在或解析本地地址)

7.3 对本地网络服务接口的扩展

把到来的本地网络多播分组传递给 IP 模块使用的方式与处理本地网络单播分组时一样，都是使用“Receive Local”操作。为了允许 IP 模块通知本地网络模块它要接收哪种多播分组，必须扩展本地网络服务接口，以提供以下两个新操作：

JoinLocalGroup(group-address)

LeaveLocalGroup (group-address)

“group-address”是 IP 主机组地址，JoinLocalGroup 操作要求本地网络模块接收和向上传递目的地为给定 IP 主机组地址的分组。LeaveLocalGroup 操作要求本地网络模块停止向上传递目的地为给定 IP 主机组地址的分组。本地网络模块应该能够映射 IP 主机组地址到本地网络地址，

这是更新本地网络模块的多播接收过滤器所要求的。任何本地网络模块都可以忽略 LeaveLocalGroup 操作的要求。如果它不能充分过滤到来的分组，也可以向上传递超出 JoinLocalGroup 操作所要求的更多的多播地址的分组。

本地网络模块不会向上传递任何从该模块出发的多播分组；多播回送是在 IP 层或更高层处理。

7.4 对以太网本地网络模块的扩展

为了支持多播 IP 数据报的接收，以太网模块必须能接收地址是与主机的 IP 主机组地址相关的以太网多播地址的分组。利用以太网硬件接口可能具有的地址过滤能力是非常可取的，这样主机只接收目的地指向它的那些分组。

不幸的是，许多当前的以太网接口对硬件能识别的地址数目有很少的限制。而且，一个具体的实现必须能侦听任意数目的以太网多播地址，这意味着当地址数目超过过滤器的限制时，“打开”地址过滤器以接收所有的多播分组。

对于那些硬件地址过滤能力不强的接口，为了性能的原因，在以太网模块的软件中执行以太网地址过滤可能要好一些。然而，这并不是强制性的，因为 IP 模块基于 IP 目的地址执行它自己的过滤。

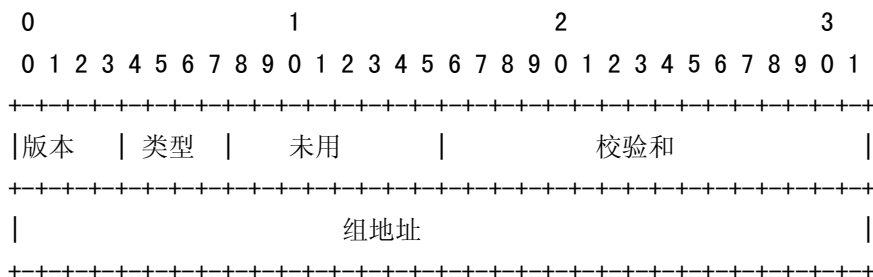
7.5 对以太网以外的本地网络模块的扩展

像 IEEE 802.2 网络这样的其它多播网络，为了达到接收多播 IP 数据报的目的而进行的处理方式与以太网相同。对于纯广播网络，如实验性以太网，所有到达的广播分组能够被接收并传递给 IP 模块以实现 IP 层过滤。在点对点或存储转发网络，多播 IP 数据报以本地网络单播的形式到达，所以没有必要修改本地网络模块。

附录 1 INTERNET 组管理协议 (IGMP)

Internet 组管理协议 (IGMP) 用于 IP 主机向所有紧邻的多播路由器报告它们的主机组成员关系。IGMP 是不对称的协议，将从主机的视角而不是从多播路由器的视角描述它。(IGMP 也可以在多播路由器之间对称或非对称的使用。这样的用法这里没有指定。)

像 ICMP 一样，IGMP 是 IP 的一个组成部分。要在所有符合 IP 多播规范的 2 级主机上实现。IGMP 报文封装在 IP 数据报中，数据报的 IP 协议字段为 2。所有与主机相关的 IGMP 报文具有如下格式：



版本

本备忘录说明 IGMP 版本 1。版本 0 在 RFC-988 中说明，它现在已经废弃。

类型

有两种与主机相关的 IGMP 报文：

1=主机成员请求

2=主机成员报告

未用

未用字段，在发送时为零，接收时被忽略。

组地址

主机成员请求报文在被发送时组地址字段为零，被接收时忽略该字段。
主机成员报告报文中，组地址字段为被报告组的 IP 主机组地址。

非正式的协议描述

多播路由器发送主机成员请求报文（在下文叫做请求），以发现哪些主机组在它们所属的本地网络上有成员。请求的目的地址为所有主机组（224.0.0.1），携带的 IP 生存时间为 1。

主机为了响应一个请求，生成主机成员报告报文，报告它们在接到请求的网络接口上所属的每个主机组。为了避免并发报告“爆炸”，减少被传送的报告的总数，使用了两种技术：

1. 当主机接收到一个请求，不是立即发送报告，而是为到达请求的网络接口上每个组成员启动一个报告延时定时器。每个定时器被设置为在 0 到 D 秒之间的一个不同的、随机选择的值。当一个定时器超时，会为相应的主机组产生一个报告。因此，报告的产生时间分散在 D 秒的间隔内，而不是都立即产生。

2. 被发送的报告的 IP 目的地址等于它要报告的主机组的地址，它的 IP 生存时间为 1，所以在同一网络上这一组的其它成员能听到该报告。如果主机听到在那个网络上它所属的组的报告，则主机停止那个组的定时器，不再产生那个组的报告。因此，在通常情况下，仅仅为出现在网络上的每个组生成一个报告，这是由延迟定时器最先超时的成员主机生成的。要注意的是，多播路由器接收所有 IP 多播数据报，因此不需要明确的指明它的地址。还要注意的，路由器不需知道哪些主机属于一个组，而仅需知道是否至少有一个主机属于在特定网络上的一个组。

对以上描述的特性有两个例外。第一，如果在请求收到以前，已经有一个报告延时定时器在为一个组成员运行，则该定时器不会重置为一个新的随机值，而允许以它的当前值运行。第二，报告定时器不会因为主机是所有主机组（224.0.0.1）的成员而生成，这个成员关系不会被报告。

如果主机使用伪随机数生成器计算报告延迟时间，主机的单个 IP 地址之一应用作生成器种子的一部分，以减少多宿主机生成相同的延迟序列的可能性。

主机应该确保接收到的报告的 IP 目的地址字段和 IGMP 组地址字段有相同的 IP 组地址。以保证主机自己的报告不会因为错误的接收报告而取消。主机应该丢弃除了主机成员请求或主机成员报告类型以外的任何 IGMP 报文。

多播路由器周期性地发送请求，以更新它所知的，在特定网络上目前的成员关系情况。如果在几次请求之后，没有接收到特定组的报告，则路由器假设这个组已经没有本地成员，它们就不需转发从远端传来的那个组的多播数据报到本地网络上。正常情况下，请求被发送的频率很低（每分钟不超过一次），因此 IGMP 对主机和网络的开销非常小。然而，当多播路由器启动时，它可以发出几个时间间隔很小的请求，以便尽快获知本地成员的情况。

当主机加入一个新的组时，它应该立即发送一个那个组的报告，而不是等待一个请求，以防万一它是这个网络上该组的第一个成员。考虑到初始的报告可能丢失或损坏，建议在一个短的延迟后重复报告一到两次。（达到这一目的的一种简单的方式是好像仅仅那个组的请求被接收，设置组随机报告延迟定时器。以下的状态变换图解释了这种方法。）

要注意的是，在一个没有多播路由器的网络上，当有主机加入一个新组时，唯一的 IGMP 通信量是一个或多个报告的传送。

状态变换图

通过下面的状态转换图，将更加正式地说明 IGMP 特性。每个主机可能处于三种可能

的状态中的一种，这与在任意的单个网络接口上的一个单独的 IP 主机组有关。

—非成员状态，当主机不属于这个接口上的组时。这是所有网络接口上所有成员的初始状态；它不需要主机的存储空间。

—延时成员状态，当主机属于那个接口上的组时，需要为这个成员关系运行一个报告延时定时器。

—空闲成员状态，当主机属于那个接口上的组时，不需要为这个成员关系运行一个报告延时定时器。

有五种重大的事件能使 IGMP 状态变换：

—当主机决定加入这个接口的组时，出现“加入组”事件。这只能出现在非成员状态。

—当主机决定离开这个接口的组时，出现“离开组”事件。这只能出现在延迟成员和空闲成员状态。

—当主机收到一个有效的 IGMP 主机成员请求报文时，出现“请求收到”事件。要达到有效状态，请求报文必须至少有 8 字节长，有一个正确的 IGMP 校验和和地址为 224.0.0.1 的 IP 目的地址。单一的请求适合于收到请求的接口的所有成员关系。这里不考虑非成员状态或延迟成员状态的情况。

—当主机接收到一个有效的 IGMP 主机成员报告报文时，出现“报告收到”状态。要达到有效状态，报告报文必须至少有 8 字节长，有一个正确的 IGMP 校验和，而且在它的 IP 目的地址字段和 IGMP 组地址字段有相同的 IP 组地址。报告仅用于在接收到报告的接口上，由报告标识的组组成员。这里不考虑非成员状态或空闲成员状态的情况。

—当接口上的组的报告延迟定时器超时，出现“定时器超时”事件。它只出现在延迟成员状态。

所有其它的事件，如收到无效的 IGMP 报文，或收到除请求或报告之外的 IGMP 报文，被忽略了。

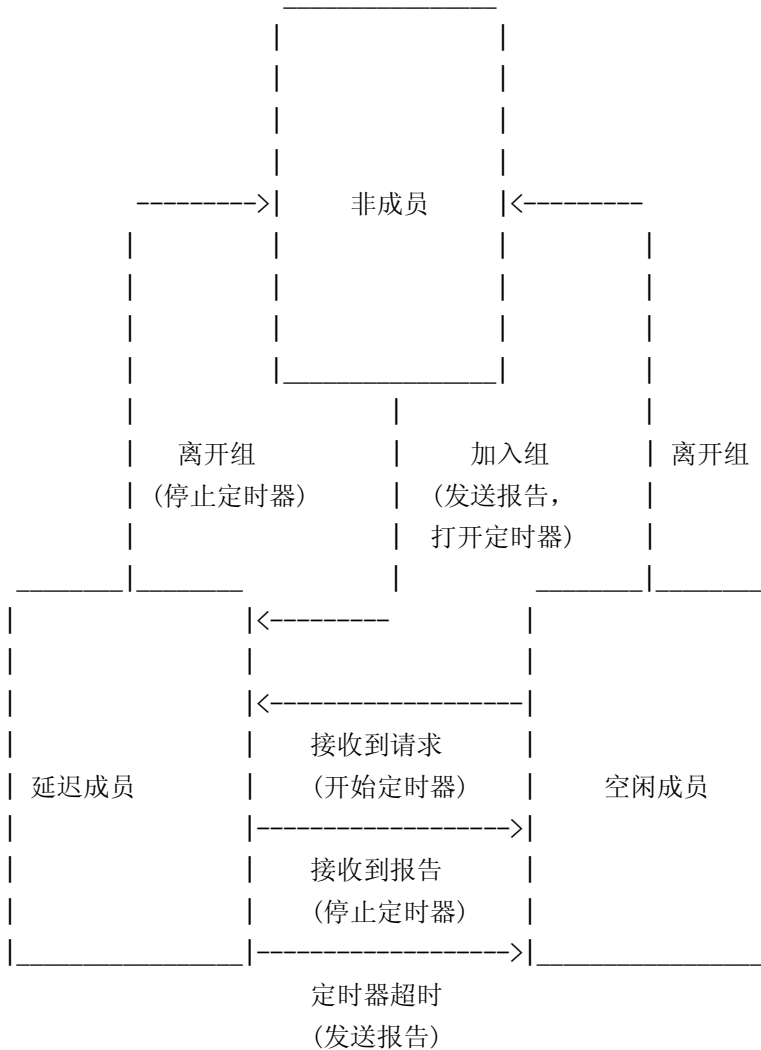
有三种可能的动作，可以用来响应以上事件：

—为那个接口上的组“发送报告”。

—为那个接口上的组“打开定时器”，使用在 0 到 D 秒之间的随机延时值。

—为那个接口上的组“停止定时器”。

在下面的图表中，每个状态转换弧由促使转换的事件标记，在圆括号中的是在转换过程中所执行的动作。



所有主机组（地址为 224.0.0.1）的处理属于特殊情况。主机开始时是每个接口上的那个组的空闲成员状态，不会变换到其它状态，也不会为那个组发送报告。

协议参数

最大报告延时 D 为 10 秒。

附录 2 主机组地址问题

本附录不是 IP 多播规范的一部分，但提供了有关 IP 主机组地址的一些问题的背景讨论。

组地址绑定

IP 主机组地址到物理主机地址的绑定可以被认为是 IP 单播地址绑定的推广。一个 IP 单播地址静态地绑定到一个单一的 IP 网络上的一个单一的本地网络地址接口上。一个 IP 主机组地址动态地绑定到一个 IP 网络集合上的本地网络接口集合上。

IP 主机组地址不会绑定到 IP 单播地址集合上，理解这一点很重要。多播路由器不会维持每个主机组的单个成员的列表。例如，属于一个以太网的多播路由器仅需有一个与有本地成员的每个主机组对应的单一的以太网多播地址，而不是那些成员的单个 IP 或以太网地址列表。

临时主机组地址的分配

本备忘录没有说明暂时的组地址是如何分配的。预期是 IP 暂时主机组地址空间的不同部分的分配使用不同的技术。例如，可能有许多服务器联系在一起获得一个新的暂时组地址。一些高层协议（比如 VMTP，在 RFC-1045 中说明）可以生成高层暂时“进程组”或“实体组”地址，这些地址会在算法中映射到 IP 暂时主机组地址的子集中，这与 IP 主机组地址映射到以太网多播地址上类似。IP 组地址空间的一部分可以留给一些应用程序随机分配，这些应用程序能容忍和其它多播用户偶尔冲突，在发现一个合适的“安静”地址之前，一直生成新的地址。

一般来说，主机不能假设被送到主机组地址的数据报将仅仅到达目的主机，或作为暂时主机组的一个成员接收数据报的目的就是为了接受。在 IP 层以上，使用高层标识符或认证标志来检测错误投递。如果发送者不想让不必要的用户收到，则传递给主机组地址的信息应该被加密或用路由管理控制方法进行控制。

作者地址

Steve Deering
Stanford University
Computer Science Department
Stanford, CA 94305-2140

Phone: (415) 723-9427

EMail: deering@PESCADERO.STANFORD.EDU