[19] 中华人民共和国国家知识产权局

[12] 发明专利说明书

专利号 ZL 200480004531.2

[51] Int. Cl.
H04L 29/08 (2006.01)
H04L 12/26 (2006.01)
H04L 12/18 (2006.01)
H04N 7/24 (2006.01)
H04L 29/06 (2006.01)

[45] 授权公告日 2009 年 10 月 28 日

[11] 授权公告号 CN 100556045C

[22] 申请日 2004.2.19
[21] 申请号 200480004531.2
[30] 优先权
[32] 2003.2.19 [33] GB [31] 0303812.2
[86] 国际申请 PCT/GB2004/000681 2004.2.19
[87] 国际公布 WO2004/075511 英 2004.9.2
[85] 进入国家阶段日期 2005.8.18
[73] 专利权人 英国电讯有限公司
地址 英国伦敦
[72] 发明人 特雷弗·伯布里奇
梅齐亚尔·内科韦
安德列·亚索伯拉
[56] 参考文献
CN1337108A 2002.2.20

Multicast Session Membership Size Estimation. Timur Friedman, Don Towsley. INFOCOM '99. 1999

Optimal on – line estimation of the size of a dynamicmulticastgroup. Sara Alouf, Eitan Altman, Philippe Nain. IEEE INFOCOM 2002. 2002
审查员 汤广强
[74] 专利代理机构 北京三友知识产权代理有限公司
代理人 孙海龙

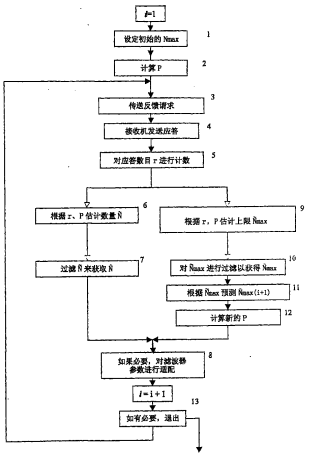
权利要求书 2 页 说明书 12 页 附图 2 页

[54] 发明名称

用于追踪组播观众数量的方法

[57] 摘要

本方法包括向接收组播的接收机传送多个要求反馈的请求(3)，每一个请求都包含概率参数(P)。每一个终端以相应的概率来对此进行应答(或是不应答)(4)。 然后，其中对针对各请求的应答数目进行计数(5)；根据计数和参数确定接收机数量估计(6)；并且对估计进行过滤(7)。 本方法还包括通过根据计数和参数预测接收机数目上限来重复计算后续反馈请求中包含的新的概率参数(9，10，11)，并且从中确定新的概率参数(12)，由此将应答数目超出预定阈值的风险保持在某个预定值以下。



1．一种估计组播观众数量的方法，包括以下步骤：

（a）向接收组播的多个接收机传送一个请求，所述请求包含概率参数（P），由此各接收机都能以相应概率做出应答或是不应答；

（b）对所述请求的应答数量（r）进行计数；

（c）根据所述计数和所述概率参数确定对接收机数量的估计；

（d）过滤所述估计；

其中该方法还包括：通过根据所述计数和所述概率参数预测接收机数目的上限来对后续步骤（a）中包含的新的概率参数进行计算。

2．根据权利要求1所述的方法，该方法包括：

确定新的概率参数，由此将应答数目超出预定阈值的风险保持在预定值以下。

3．根据权利要求2的方法，其中计算新的概率参数的步骤包括：

估计最大观众数量，该最大观众数量使得接收到的应答数目与使用所述概率参数的情况下观察到的应答数目相等的概率为预定概率；

使用所述估计得到的最大观众数量以及所述最大观众数量的至少一个在先值来执行所述预测；

结合预测得到的最大观众数量，确定新的概率参数（$P(t_{i+1})$），其中，所述新的概率参数使应答数目超出可用于对应答进行接收的容量的风险落在预定风险阈值以下。

4．根据权利要求1的方法，其中计算新的概率参数的步骤包括：

估计最大观众数量，该最大观众数量使得接收到的应答数目与使用所述概率参数的情况下观察到的应答数目相等的概率为预定概率；

过滤所述最大观众数量；

使用经过滤的所述最大观众数量以及所述最大观众数量的至少一个在先值来执行所述预测；

结合预测得到的最大观众数量，确定新的概率参数（$P(t_{i+1})$），其中，所述新的概率参数使应答数目超出可用于对应答进行接收的容量的风险

落在预定风险阈值以下。

5．根据权利要求 4 的方法，其中对所估计出的最大观众数量所进行的过滤是由维纳滤波器执行的。

6．根据权利要求 4 的方法，所述方法包括依据所述估计的功率谱来自适应地调整对所估计出的最大观众数量进行所述过滤所用的参数。

7．根据权利要求 2 所述的方法，其中通过推断所估计出的最大观众数量的过去值而改善预测。

8．根据权利要求 2 所述的方法，其中对所述估计所进行的所述过滤是由维纳滤波器执行的。

9．根据权利要求 2 所述的方法，所述方法包括按照所述估计的过去的值的功率谱的函数来自适应地调整对所述估计进行所述过滤所用的参数。

10．根据权利要求 2 所述的方法，其中对所述估计所进行的所述过滤是在停止确定所述估计之后执行的。

11．根据权利要求 2 所述的方法，其中对所述估计所进行的所述过滤是在每一次确定新的估计的时候执行的。

12．根据权利要求 1 所述的方法，所述方法包括测量请求或应答的丢失概率，并且对最初估计数量进行校正。

用于追踪组播观众数量的方法

技术领域

本发明涉及观众数量的测量，尤其涉及对组播传输中动态变化的观众数量进行估计。

背景技术

组播的一种形式是IP技术，它允许有效地将数据流从一点发送到多个目的地。组播会在通向不同组播群组成员的路径出现分叉处的路由器跳（hop）处复制分组，而不是为每一个目的地设定独立的单播会话。这样允许信源发送数据流的单个拷贝，而同时到达数以千计乃至数以百万计的接收机。在常规的单播通信中，发送器为每一个观众成员设定一个独立的传输会话，并且通过对服务器日志文件中被请求的流的数目进行计数来追踪观众数量。在组播通信中，服务器只设定一个数据流，所述数据流将被发送到一个单独的组播地址，而不是分别定址到各个接收机。随后加入相应组播群组的任何一个对数据流感兴趣的主机从它的本地路由器中获取数据流。因此，对发送器来说，观众数量是隐藏的，它会随时间而快速变化。

这种情形非常类似于常规的无线电和电视广播：无线电或电视电台送出节目来进行无线电广播，接收方则可以将其接收机调谐到发送节目的频道，以便收听或观看喜欢的节目。

从技术上讲，对某些应用而言，追踪组播群组的规模是非常重要的：

●组播被用于将因特网上的音频和视频流实时传递给大量观众。从这些应用中产生的收益主要来自广告，它在很大程度上依赖于观众数量的估计及其随时间的变化。

●组播被用于大规模的发布订阅应用。在这些应用中，订户会登记其对某个主题或事件模式的兴趣，并且异步接收与其兴趣相匹配的事件。

通过订阅某个主题，订户将会成为与该主题（例如某个公司的股价）相对应的组播群组中的成员，在该主题更新时，用户将会收到通知。发布订阅服务器可以定期更新用户对主题的兴趣（各主题的接收者数目），以便优化主题划分及其在组播群组中的分发。

组播观众数量（以及其他统计数据）的测量可以在网络层进行。在这种方法中，从网络路由器并行获取存储表，并且将其传送到监视单元。然后，监视单元会从这些表中提取组播群组数量（以及其他统计数据）（参见 K. Sarac、K. Almeroth 在 2000 年发表于 Journal of High Speed Networks 第 9 卷第 3-4 号的“Supporting multicast deployment efforts: a survey of tools for multicast monitoring”，以及 P. Rajvaida、K Almeroth 和 K. Claffy 在 2000 年 12 月发表于 Proc. Of 11th IIFIP/IEEE International Workshop on Distributed Systems (DSOM2002), Austin, Texas, USA 的“A scalable architecture for monitoring and visualizing multicast statistics”）。

这种方法存在三个问题：

- 它要求监视单元能够访问网络组播路由器。这个要求在很多情况下都是不可行的。此外，即使监视单元可以访问组播路由器，当前标准也没有要求这些路由器保持本地接收机数量计数。
- 它不适合应用层组播，在所述应用层组播中，构建组播传输作为在没有组播功能的网络顶端的重叠。
- 这种方法无法升级到上述发布订阅的情况（由于监视单元上的信息爆炸），在该发布订阅情况中，数以亿计的订户订阅了数百万个组播群组。

在文献中提出了很多解决方案，以便解决对大规模组播传输中的观众数量进行实时和端到端的估计的问题。其中两种已知的技术是由 Lieu 和 Nonnenmacher 提出的用于估计静态观众数量的方法（C. Lieu 和 J. Nonnemacher 在 2000 年 3 月发表于 Proc. Of IEEE INFOCOM2000, Tel Aviv, Israel 第 2 卷第 952～960 页的“Broadcast audience estimation”）以及由 Alouf、Altman 和 Nain 提出的估计动态变化的观众的方法（S. Alouf、

E. Altman 以及 P. Nain 在 2002 年 6 月发表于 Proceedings of INFOCOM 2002, New York, USA 的“Optimal on-line estimation of the size of a dynamic multicast group”，http://www-sop.inria.fr/mistral/personnel/Sara.Alouf/Publications/multicast.pdf)。这两种方法都是以发送器对观众群组的反馈消息所进行的随机取样为基础的。然后，通过使用统计技术，可以根据该随机取样推断出观众数量。

在 Lieu 等人的方法中，随机取样是使用基于定时器的方法完成的（参见 J. Nonnemacher 和 E. Biersack 在 1999 年 6 月发表于 IEEE/ACM Trans. On Networking 第 7 卷第 3 号第 375～386 页的“Scalable feedback for large groups”，以及我们在 2002 年 6 月 21 日提交的名为“Timer-based feedback in multicast communication”的欧洲专利申请 02254355.7，其中该申请的发明人是 M. Nekovee 和 S. Olafsson)。在该方法中，发送器将定时器分布 f(t)连同反馈请求一起发送到所有接收机。一旦收到这个反馈，则每一个接收机都根据 f(t) 进行一段补偿时间（backoff time）的采样。

在该时间终止之后，如果接收机检测到来自任一其他接收机的反馈信息，那么该接收机将会保持沉默，否则它会向发送器（以及所有其他接收机）发送包含有接收机定时器终止时间的反馈。发送器则收集所有这些反馈。通过将反馈计数与接收机定时器终止时间相组合，接收机对观众数量进行统计估计。为了在观众估计过程中取得很高的精确度，上述过程必须重复进行多个循环，这是因为估计误差仅仅会降低为$1/\sqrt{M}$，其中 M 是循环数。而这种方法存在的主要技术问题在于：

- 这个估计过程适用于静态观众数量，但是并不适合动态变化的观众数量。这是因为该方法需要多次从相同观众那里收集反馈，以便准确估计数量，当观众数量在这些收集循环过程中发生变化时，所述估计将是不正确的。

- 基于定时器而对观众所进行的随机采样需要将各个反馈消息经由组播发送到所有观众。当采样数量（也就是在各个循环中发送消息的接

收机数目）很少并且我们只考虑一个组播群组时，处理这些消息的开销很小。但是当接收机同时处于若干个组播群组并且由此接收到来自所有这些群组的反馈消息时，开销有可能会变得很大。

● 平均起来，与发送器具有最小往返时间（RTT）的接收机更有可能发送反馈消息，因此，基于“定时器”的采样机制具有向这些接收机的偏差。这种偏差将会降低观众数量估计的精确度。此外，由于发送器“听到”那些 RTT 大于平均值的接收机的可能性较小，因此，该偏差将会导致轮询机制不公平。

Alouf、Altman 和 Nain 的方法使用了一种更简单的机制来实施基于概率组播的随机采样，由此解决了其中某些问题（参见 M. H. Ammar 在 1994年6月发表于Proc. Of IEEE INFOCOM94, Toronto, Canada, 第848～855 页的“Probabilistic multicast: Generalising the Multicast Paradigm to Improve Scalability”）。在该方法中，发送器向所有接收机发送一个要求反馈的请求。一旦收到该请求，每一个接收机将会依照指定概率来发送反馈或是不发送反馈。由于每一个接收机所做出的是否发送消息的决定都是独立于剩余各方的，因此采样机制不会受到发送器-接收机以及接收机-接收机的 RTT 时间的影响，并且不会向 RTT 较小的接收机偏斜。

所发送消息的平均数量是具有二项式概率分布函数（pdf）的随机变量。与基于定时器的方法相似，发送器使用公知的统计估计技术从反馈计数中估计观众数量。

Alouf、Altman 和 Nain 的方法利用了处于连续反馈循环中的观众数量之间的统计相关性，由此提供了一种针对动态变化的观众数量问题的解决方案，从而提高了即时估计的准确性。这种方法是通过下列假设实现的：（i）观众数量变化是通过通信业务繁忙时的所谓的 M/M/∞队列的总数变化来描述的（参见 Leonard Kleinrock，Queueing Systems, Vol 1, John Wiley & Sons，New York，1975），（ii）反馈请求和反馈消息不会丢失，以及（iii）观众数量很少，由此不存在反馈爆炸的风险。

这种方法存在的技术问题是：

• 该方法不能扩展到大型组播群组，但是这种群组在我们较早论述的情况中（因特网和内部网的电视和无线电广播、发布订阅服务）是非常典型的情形。这是因为响应概率是在独立于观众数量的情况下选择的，因此，反馈消息的平均数量会随着观众数量增长而增大，由此最终导致反馈爆炸（feedback implosion），虽然据记载如果应答数目过大，这种可能性将会略微减小。

• 滤波器参数是预先固定的。它们并不适应于总数数量和动态特性的可能变化。

• 该算法假设反馈请求和反馈消息都不会在网络中丢失，但在有可能出现高分组丢失率的实际网络状况中，这种算法是很容易受到破坏的。

发明内容

依照本发明，提供了一种追踪组播观众数量的方法，包括：

（a）向接收组播的接收机传送多个请求，其中每一个请求都包含概率参数，由此每一个终端都能以相应概率进行应答或不应答；

（b）计数对每一个请求的应答的数量；

（c）从计数和参数中确定接收机数量估计；

（d）过滤这些估计；

其中该方法还包括：通过根据计数和参数来预测接收机数目的上限对后续步骤（a）中包含的新的概率参数进行反复计算，并且从中确定新的概率参数，由此将应答数目超出预定阈值的风险保持在预定值以下。

本发明的其他方面是在权利要求中限定的。

附图说明

现在将参照附图来描述本发明的某些实施例，其中：

图 1 是网络示意图；以及

图 2 是描述本发明一个实施例的操作的流程图。

具体实施方式

图 1 示出了一种组播方案，其中服务器 100 经由 IP 网络 102（例如因特网）而与多个接收终端 103（只显示了其中的三个）相连。组播操作是以常规方式运作的，因此，现在仅仅描述那些专为测量观众数量而提出的附加功能。

我们设想了将消息发送到组播群组的发送器（即服务器 100）。群组成员（103）可以在任何时候加入和离开该群组。其中一个实例是使用组播而在因特网上广播电视。观众可以通过加入一个与广播节目的频道相对应的组播群组来收看某个节目，并且可以在节目中的任何时间离开这个组播群组。

现在将要描述的测量过程是如下执行的：

处于发送器上的附加软件进行：（a）构造并且借助组播机制来周期性传送请求消息，（b）对来自终端的应答进行计数，以及（c）对结果进行处理。

处于每一个终端上的附加软件对这种请求进行处理并且酌情产生应答。观众数量是依赖于时间的随机变量，并且我们用 N(t)来表示这个变量。我们假设发送器可以在每一个反馈循环中处理最多为 r_{max} 的反馈消息。目的是借助来自接收机的反馈计数 r(t)而对 N(t)（或是其他可行的接收机属性）进行实时正确的估计，同时使反馈爆炸（这种情况是在 r(t)超出 r_{max} 的时候发生的）的风险一直最小。

接下来将参照图 2 所示的流程图来对该处理进行描述。

1．这个估计始于初始猜测的最大观众数量 N_{max}。

2．然后，在给出了 N_{max} 的情况下，通过将反馈爆炸风险低于用户定义的风险阈值δ的最大响应概率选作 P，可以获取响应概率 P 的初始值。由于反馈消息数目具有二项式概率分布函数，因此，在给出了对象总体数量 N_{max} 的情况下，反馈消息数量超出 r_{max} 的概率是如下给出的：

$$\Pr(r \geq r_{\max}) = 1 - \Pr(r < r_{\max}) = 1 - \sum_{r=1}^{r_{\max}} B(N_{\max}, P) = I_p(r_{\max} + 1, N_{\max} - r_{\max})$$

由此，通过求解下式，可以获取 P 的最大可能值

$$I_p(r_{\max} + 1, N_{\max} - r_{\max}) = \delta \qquad (1)$$

在上述等式中，$B(N_{max}, P)$是二项式分布，并且 I_p 表示不完全的β函

数，即

$$I_P(a,b)=\frac{\int_0^P t^{a-1}(1-t)^{b-1}dt}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}$$

在当前实施方式中使用了某种形式的牛顿-拉弗森（Newton-Raphson）迭代法而以从等式（1）中找到数值P。

3．一旦找到了 $P(t_1)$，则发送器组播要求反馈的请求（其中包含 P 值）。

4．每一个接收机选择一个范围介于 $0\leqslant X\leqslant 1$ 的随机数 X，并且只在 $X\leqslant P(t_1)$ 的情况下才传送反馈消息。

5．发送器从该循环中收集总数为 r 的反馈消息。

6．然后如下估计观众数量：

$$\tilde{N}(t_1)=\frac{r(t_1)}{P(t_1)}\pm\gamma \qquad (2)$$

其中 γ 是随机估计误差。在这里不必对其进行计算，但是如果需要的话，则它与 $\frac{1-P(t_1)}{P(t_1)}$ 成比例。

（等式 2 以及相应的误差表达式是通过使用统计估计理论中公知的最大似然法并且将其应用于具有未知参数 N(t) 的二项式分布而被推导出来的）。

应该指出的是，流程图中显示的两个分支是为了清楚显示有两个估计处理正在工作而被绘制的；它们并不是替换的路径。

7．观众数量估计（等式 2）包含了统计估计误差，其幅度与所采样的接收机数量成反比。我们提供了一种通过利用那些可用的过去数据来实时减小估计误差的方法。如所示，这种方法是在每一次测量之后的测量处理中进行的（参见下文中的替换实施方式）。这种处理是通过将估计误差 n(t) 视为依赖于时间的噪声来实现的，其中该噪声叠加在观众数量的准确数值 N(t) 之上。然后，通过使用维纳滤波器（Wiener filter）过滤测量得到的观众数量 $\tilde{N}(t)$（$\tilde{N}(t)=N(t)+n(t)$），而使其中的信噪比最大化，所述滤波器提供了关于观众总数的最佳均方估计。一开始，滤波器参数是基于观众数量变化的历史信息以及所考虑的应用（例如因特网电视、

发布订阅等等）而被确定的，但在对观众数量进行测量的过程中将会周期性地重新计算这些参数，由此它们可以适应于观众数量变化的实际模式。

在时间 t_i，经过改进的观众数量估计是从下式中获取的：

$$\hat{N}(t_i)=\sum_{j=1}^{i} h(t_i-t_j)\tilde{N}(t_j)=(\beta-\alpha)\sum_{j=1}^{i}\tilde{N}(t_j)\exp(-\beta(t_i-t_j)) \quad (3)$$

在这里

$$h(t_i-t_j)=(\beta-\alpha)\exp(-\beta(t_i-t_j)) \quad (4)$$

是最佳的维纳滤波器内核。

该内核是通过以下假设获取的，其中信号（N(t)）和噪声（n(t)）在统计上是无关的，并且这些分量的功率谱 $N_P(\omega)$ 和 $n_P(\omega)$ 可以如下地近似估计：

$$N_P(\omega)=\frac{\alpha K}{\omega^2+\alpha^2} \quad (5)$$

$$n_P(\omega)=A \quad (6)$$

其中 K、α 和 A 是可调整的参数，并且

$$\beta^2=\alpha^2+\frac{K}{A} \quad (7)$$

由此可见，这表示的是观众数量 $\tilde{N}(t)$ 的功率谱 $\tilde{N}_P(\omega)$ 的模型，其中

$$\tilde{N}_P(\omega)=N_P(\omega)+n_P(\omega)=\frac{\alpha K}{\omega^2+\alpha^2}+A \quad (8)$$

上述关于 $N_P(\omega)$ 的选择源于以下事实：等式（5）的限制 $\alpha \to 0$ 准确模拟了缓慢变化的观众数量，同时，通过增大 α，可以对常规的时变观众数量进行模拟，并且这种模拟的精度足以确定滤波器参数。此外，为 $n_P(\omega)$ 假设的形式与白噪声的假设是对应的。我们的模拟研究显示，在对观众数量进行测量的过程中，这是关于统计噪声的合理估计。

8．在对观众数量进行测量的过程中，滤波器参数 β 的自适应是如下执行的。发送器在大小为 M 的滑动窗口上累积 $\tilde{N}(t)$ 的过去值，然后通过对 $\tilde{N}(t)$ 的过去值进行傅立叶变换来进行信号和噪声的功率谱分析（如果数据点并非均匀分布，则使用 Lomb 算法来估计功率谱）。然后，通过使用最小均方法使这些功率谱适于等式（8）所给出的参数化形式，来获取参

数 K、α 和 A 的新值，由此可以使用等式（7）来获取滤波器参数 β 的新值。

应该指出的是，由于滤波器内核 h(t)在时间上是按指数衰减的，因此在实践中只要为大小约为 1/β 的间隔累积过去的统计数据即可。假设对象总体是以平均速率 $f=\frac{1}{T}$ 采样的，那么需要累积的过去统计数应该是 $M\approx\frac{f}{\beta}$。

应该指出的是，自适应步骤不必在每一次迭代中都被执行。

9．除了估计观众数量之外，我们还提供了一种用于动态估计观众数量上限的方法。这种方法是依照以下方式实现的：在给出了接收到的反馈消息 $r(t_1)$ 的数目以及风险参数 ε 的情况下，找出可能出现的最大观众数量，其中该数量能以概率 1-ε 来得到 $r(t_1)$ 个反馈消息。还表明，在给出了观察值 $r(t_i)$ 的情况下，通过使用贝斯定理并执行二进制的概率分布函数的泊松近似，观众数量超出了某个值 N_{max} 的条件概率是由下式给出的：

$$\Pr(N(t_i)\geq N_{\max}\mid r(t_i))\propto Pg(r(t_i)+1,PN_{\max})$$

其中 Pg 是不完全的 Γ 函数。

由此，最大可能数量 $\tilde{N}_{\max}(t_i)$ 是通过求解下式（使用某种形式的 Newton-Raphson 迭代）得到的

$$Pg\left(r(t_i),P(t_i)\tilde{N}_{\max}(t_i)\right)=1-\varepsilon \qquad (9)$$

其中 Pg 是不完全的 Γ 函数。

10．然后，通过使用与步骤（7）的描述相同的方式来过滤这个估计 $\tilde{N}_{\max}(t_i)$，由此使用等式（10）来提供已过滤的观众数量上限估计 $\hat{N}_{\max}(t_i)$：

$$\hat{N}_{\max}(t_i)=\sum_{j=1}^{i}h(t_i-t_j)\tilde{N}_{\max}(t_j)=(\beta-\alpha)\sum_{j=1}^{i}\tilde{N}_{\max}(t_j)\exp(-\beta(t_i-t_j)) \qquad (10)$$

在这个形式中，所使用的参数是在步骤 7 中使用（以及在 8 中适配）的那些参数。

11．然后，通过使用估计 $\hat{N}_{\max}(t_i)$ 并以如下方式来预测下一个循环中的最大观众数量：

如果 $\hat{N}_{max}(t_i) > \hat{N}_{max}(t_{i-1})$ 则

$$\hat{N}_{max}(t_{i+1}) = \hat{N}_{max}(t_i) + (t_{i+1} - t_i)\frac{\hat{N}_{max}(t_i) - \hat{N}_{max}(t_{i-1})}{t_i - t_{i-1}}$$

否则 $$\hat{N}_{max}(t_{i+1}) = \frac{\hat{N}_{max}(t_i) + \hat{N}_{max}(t_{i-1})}{2} \qquad (11)$$

12. 通过使用这个预测，可以从等式（1）中计算采样概率 $P(t_{i+1})$，此外，包含了 $P(t_{i+1})$ 的新的反馈请求将被发送到接收机。也就是说，上述步骤是从步骤（3）开始重复的。一旦收集了预期数量的测量结果，则在步骤 13 停止迭代。

现在将对某些可选的修改和附加步骤进行描述：

A. 在上文中显示，步骤 7 中的过滤是在继续测量的同时进行的，这在需要以准实时方式进行测量的情况下是非常有利的，这样一来，在继续测量处理的同时，可以使用已过滤的结果 $\hat{N}(t)$。然而，以离线方式执行如上所述的这种过滤有可能会非常有利，这是因为滤波器由此可以访问所有数据。在这种情况下，过滤 7（以及自适应 8）是在退出迭代之后执行的。然而，在循环内部必须保持步骤 10 中对于上限 $\tilde{N}_{max}$ 所进行的过滤，由此自适应也必须在循环内部进行。在这种情况下有可能使用非因果滤波器；如果使用的话，那么应该对滤波器核心进行修改，使用下式而不是等式（4）来适应 $(t_i - t_j)$ 的负值：

$$h(t) = \frac{\alpha K}{2A\beta} e^{-\beta|t|} \qquad (12)$$

B. 增大估计算法应对网络损失的健壮性

如果网络有损，那么某些要求反馈的请求可能会在到达接收机之前丢失。此外，某些来自接收机的反馈消息有可能并未到达发送器。对观众数量所进行的最大似然估计（等式 2）是在假设网络无损的情况下获取的，由此所有接收机会以相等概率 P 来响应要求反馈的请求。

这种方法提供了一种以平均方式顾及估计观众数量过程中的损失效果的方法。

如果接收机上的损失概率是 q_j（对接收机 j 而言），那么，来自这个接收机的反馈消息到达发送器的概率不再是 P，而是 $P_j = P(1-q_j)^2$。我们

在自己估计过程中以如下方式顾及了这种影响。我们假设每一个接收机测量自己的损失概率（例如，接收机可以通过查看分组序列来对那些没有从其所属的各个组播群组到来的分组数据进行计数，由此完成这个操作）。在收集反馈的过程中，那些响应受到采样的接收机会将这个测量结果回送到发送器。这样一来，发送器将会估计平均分组损失$\bar{q}$，并且会在观众数量的最大似然估计中使用$(1-\bar{q})^2 P$来替换 P，由此对其加以考虑。然后，从下式中获取经过校正的观众数量估计：

$$\tilde{N}_{loss}(t)=\frac{r(t)}{(1-\bar{q})^2 P(t)}=\tilde{N}(t)(1+2\bar{q}+3\bar{q}^2+...) \qquad (13)$$

其中，括号中的项是对估计观众数量所进行的校正，它是从网络损耗中导出的。这种变型是用等式（13）而不是等式（2）来实现的。

C．放宽反馈爆炸限制

反馈爆炸是在同时到达发送器的反馈消息数目超出 r_{max} 的时候发生的。除了选择响应概率值 P 而使爆炸机会低于某个阈值之外，发送器还可以通过展宽接收响应的间隔来进一步减少爆炸的可能性。这种处理可以通过发送参数 P 以及第二参数 S 来实现。然后，反馈过程是如下修改的：

每一个接收机选择一个均匀随机数 X，并且只在 X<P 的情况下才决定发送反馈。在做出决定之后，它会选择一个均匀分布在间隔[0,2S]中的随机数 s，并且在发送其反馈之前将会等待一段时间 s。这种算法的有效效果（net effect）是在某个间隔 S 上扩展来自接收机的反馈，从而在每一个单位时间发送的反馈消息的平均数量是由$\frac{r}{1+S}$给出的，并且如果 $r<(1+S)r_{max}$，则可以避免反馈爆炸。因此，爆炸限制实际是$(1+S)r_{max}$而不是 r_{max}。在步骤（2），等式（1）是通过插入$(1+S)r_{max}$而不是 r_{max}来修改的，由此可以使用新的限制来计算 P。在这里计算的 P 值将会大于使用 r_{max}所计算的值（我们是从 γ 函数形状中了解的），因此，在每一个循环中会产生更多反馈，并且提高了对象总体估计的精确度。

上述方法涉及的是端到端的方法，其中观众估计可以在没有来自网络路由器的任何帮助的情况下在应用级完成，并且顾及了大规模的组播

方案。

通过使用自适应方法来对源于接收机的反馈进行采样，以及使用一种用于将统计误差过滤以与观众数量的动态变化相适配的方法，可以解决较早论述的某些问题。

这种自适应采样方法将反馈爆炸的风险减至最小，同时有助于确保发送器接收到最大可能数目的反馈消息。自适应过滤还能及时动态调整成观众数量变化的模式，以便将估计得到的信号噪声比增至最大。另外，我们还描述了用于提高估计方法相对于网络中的大量分组损失的健壮性的过程，以及一种使用随机定时器来放宽爆炸限制的方法。

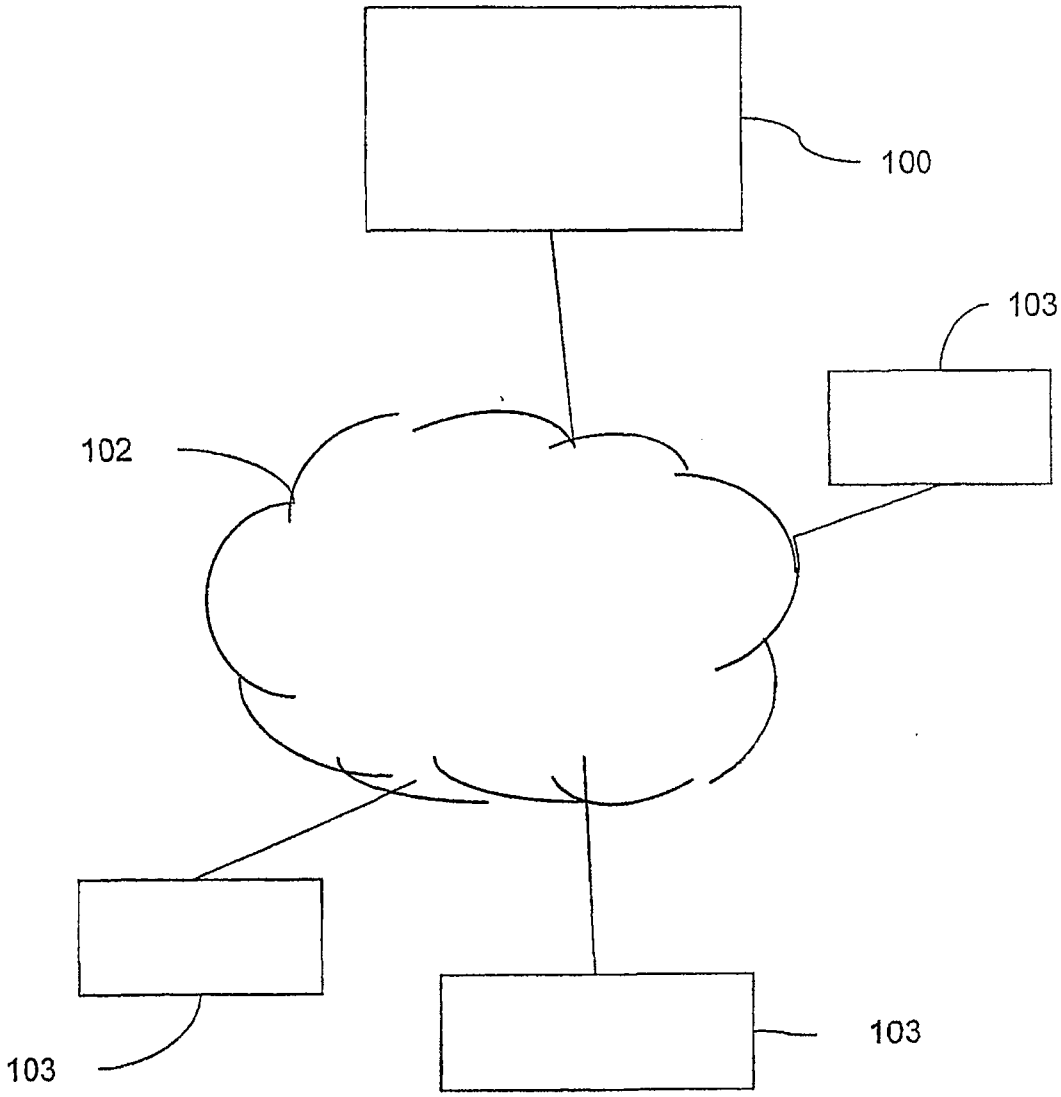


图 1

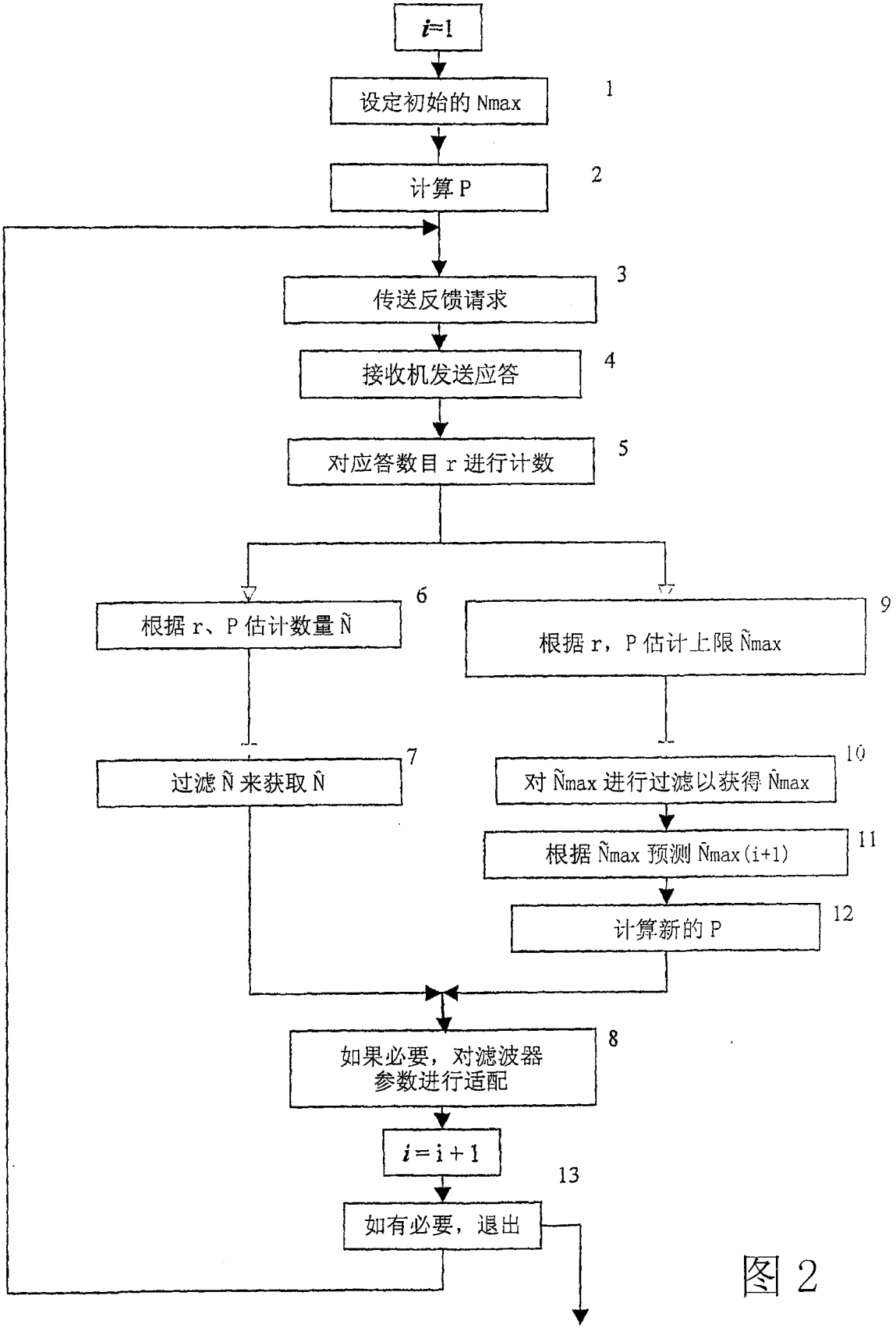

i=1
设定初始的 Nmax
1
计算 P
2
传送反馈请求
3
接收机发送应答
4
对应答数目 r 进行计数
5
根据 r、P 估计数量 Ñ
6
根据 r，P 估计上限 Ñmax
9
过滤 Ñ 来获取 N̂
7
对 Ñmax 进行过滤以获得 N̂max
10
根据 N̂max 预测 Ñmax(i+1)
11
计算新的 P
12
如果必要，对滤波器参数进行适配
8
i=i+1
13
如有必要，退出

图 2