

# 支持多域协同的端到端组播监测机制

王栋平

曹争

东南大学计算机科学与工程系 江苏省计算机网络技术重点实验室, 南京, 210096

**摘要:** 随着组播技术的飞速发展, Internet 对组播技术进行大规模商业化应用的要求越来越迫切。如何在大范围内实现组播 SLA (性能) 监测是一个关键问题。为此, 本文针对包含多个管理域的组播网络, 提出了一种主动和被动相结合的端到端组播监测机制——MEMM (Multi-domain-supported End-to-end Multicast Monitor)。

**关键词:** 组播; 多域; 主动监测; 被动监测

## Multi-domain-supported End-to-End Multicast Monitor Mechanism

Wang Dongping, Cao Zheng

Southeast University, Computer Science and Engineering Dept.,  
JIANGSU Province Key Laboratory of network technology,  
Nanjing, 210096

**Abstract:** The commercial application of multicast becomes more important as the developing of multicast. The key to it is large-scale multicast SLA (performance) monitor. Multi-domain-supported end-to-end multicast monitor mechanism--MEMM is provided in this paper. The characteristics and methodology of MEMM are presented in detail in this paper.

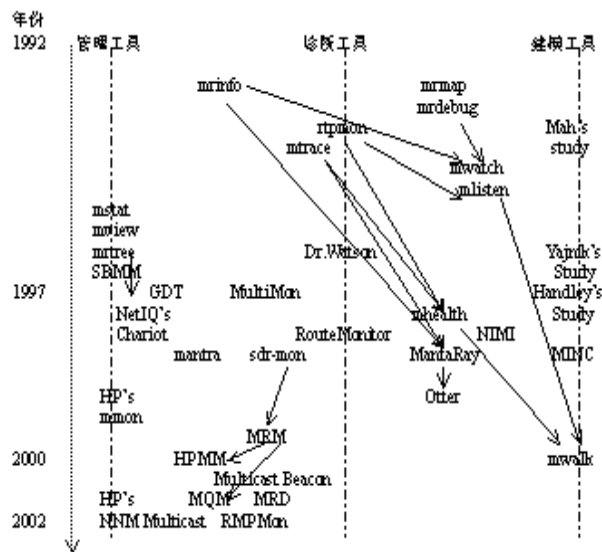
**Key words:** Multicast, Multi-domain, Active-Monitor, Passive-Monitor

### 1. 引言

随着 Internet 的飞速发展, 能够更有效地利用网络带宽资源的 IP 组播技术研究与应用成为当前科研领域的一个热点。越来越多的企业也参与到这一领域的研究工作之中。例如, Cisco、VBrick、RealNetworks、Microsoft 等公司均推出了支持 IP 组播的多媒体业务系统<sup>[1]</sup>。在日益强调服务质量的今天, 网络传输中断、网络拥塞以及组播机制失效等问题如果不能及时发现并解决, 将严重影响组播服务的质量。这会极大地阻碍组播应用的部署和推广进程。组播监测的任务就是及时发现组播网络所存在的故障, 并且提供相关信息指导故障的排除工作, 从而保证组播的服务质量。

### 2. 组播监测的现状

伴随着组播技术的发展, 研究者开发了许多组播监测工具。这些工具按照用途来分, 大致可以分为三类<sup>[2]</sup>: 管理, 诊断, 建模。图 1 表示这些工具所属类别和时间顺序, 部分工具兼具多种功能。



注: 2000 前数据来自<sup>[2]</sup>。

图 1 组播监测工具分类及时间表

组播诞生之初, 研究重点是组播内在的运行机制, 因此早期的组播监测工具都围绕着组播诊断这一目的。例如 rtpmon<sup>[3]</sup>和 mtrace<sup>[4]</sup>。这些工具适合于组播研究的专家用来获取组播网络的细节信息。它们要求使用者具备丰富的组播知识, 这一特征使此类工具的使用范围受到很大局限。自 2000 年以来, 基本上没有开发新的此类工具。

目前用于组播管理的组播监测工具主要有 MRM, HPMM, Multicast Beacon……等等。使用的测量手段分为两类: 第一类是主动方式, 如 MRM<sup>[5]</sup>, Multicast Beacon<sup>[6]</sup>; 第二类是被动方式, 如 HPMM<sup>[7]</sup>。主动方式虽然可以实现灵活多样的测量, 但是会给目标网络注入流量, 对其产生额外的影响, 因而影响监测精度。被动方式虽然可以将对目标网络的干扰降至最小, 但是它的应用却受到许多方面条件的制约。例如, 对于不在组播转发树拓扑上的子网就无法采用被动方式进行监测。显然, 单独采

用主动或被动方式都不能取得最佳的监测效果。在组播监测代理的分布方面，MRM 和 HPMM 都要求在监测网络内部的路由器中配置监测代理。其最大的缺陷就是部署的代价非常高，需要将路由器（特别是骨干路由器）全部更新，这几乎是不可能的。

本文针对端到端组播监测的需求，提出一个支持多管理域，采用主动和被动相结合的端到端组播监测机制——MEMM(Multi-domain-supported End-to-end Multicast Monitor)。MEMM 针对管理域内部和外部监测的不同特点，利用不同的监测策略，较好的解决了域间监测和组播监测扩展性的问题，同时通过主动和被动相结合的测量手段可以兼顾监测的灵活性和准确性。

### 3. MEMM 机制

#### 3.1 MEMM 的体系结构

MEMM 通过收发探测报文（或者利用网络已有的组播流量）来监测组播网络的端到端性能。MEMM 中定义的端到端性能包括：单向延时，单向抖动，丢报率。MEMM 在被监测网络边缘部署 MEMM 代理（agent），它们被称为 MA (Monitor Agent)。在主动方式中，MA 将加入测试组播组，定期发送测试报文到测试组播组，同时接收来自其他 MA 的测试报文，并以此为依据来计算组播网络的性能信息。在被动方式中，MA 不仅进行测试组播组的测量，MA 还通过网络监听的方式采集已有用户组播流量的信息，不同 MA 采集的用户流量信息汇集到组播监测管理者 MM (Monitor Manager) 计算组播网络的性能。MA 定期向 MM 发送监测报告，汇报监测结果。MM 发现性能劣化超过阈值后，将产生告警信息通知组播网络管理系统，与其他系统协同实现故障定位和故障恢复。

MM 是 MEMM 系统中另外一种实体，它的任务是依据监测目标配置 MA，收集和分析由 MA 发送的网络性能信息，产生性能报告，提出故障指示，与组播网络管理系统的其它部分协同工作来隔离故障区域和指导故障恢复。每个管理域中只有一个 MM，在多管理域协同监测的情况下，MM 还负责选择参与域间监测的 MA，收集和分析域间组播传输性能。

MEMM 系统之中存在两类通信：MA 和 MA 之间存在测试数据流作为测量的对象，此类通信采用普通组播方式进行传输；MM 和 MA 以及 MM 和 MM 之间存在控制流，用来交换测量报告和传输控

制命令，一个 MM 需要和多个 MA 通信，所以采用 UDP 协议传输，该类通信要求的可靠性高，因此由应用层保证传输的可靠性。MEMM 的控制报文传输不依赖于组播，使系统在网络的组播机能发生故障时仍然可以工作。

#### 3.2 MEMM 系统初始化

MEMM 系统进行初始化的过程如图 2 所示：(1) 系统默认的监测任务是以主动方式对管理域内进行监测，MM 发送 SR (Start Request) 报文以默认配置启动全部 MA。(2) 成功启动的 MA 将返回 SS (Start Success) 报文通知 MM。(3) MM 通过分析 SS 报文建立将协同工作的主动 MA 列表，此时被动 MA 列表为空。(4) 然后 MM 通过 RR (Reconfigure Request) 报文将 MA 列表配置到 MA。(5) MA 收到新的 MA 列表后将更新其配置文件，然后返回 RS (Reconfigure Success) 报文通知 MM。(6) MM 收到全部 MA 的 RS 报文后，初始化过程结束。(7) 如果 MM 发现新的可以作为测量对象的已有组播流，就向需要变更测量方式的 MA 发送 MCR (Mode Change Request) 报文。(8) MA 成功完成测量模式变更后返回 MCS (Mode Change Success) 报文。(9) MM 根据 MCS 报文建立新的主动 MA 列表和被动 MA 列表。系统重复步骤 (4) -(6) 完成新的初始化过程。



图 2 MEMM 的初始化过程

MEMM 系统的初始化过程是由 MM 发起的，MM 可以在任何时间对指定 MA 集合进行初始化以更新 MA 的配置。MM 发起初始化的触发条件包括：新的 MA 注册，MA 注销，监测配置改变，MA 测量模式改变，以及其它域要求协同监测等等。其中 MA 测量模式改变主要存在两种触发原因：(1) 系统探测到适合作为测量对象的已有组播流，MM 通过初始化使该组播流所达范围的 MA 改成被动模式，该组播流结束后 MM 通过新的初始化使这些 MA 恢复成主动模式，这是一种自动控制方式，无需人工

干预。(2) 网络管理者需要对未来的特定时期内的特定组播服务进行监测, 则由网络管理者通过 MM 的管理界面指定上述信息, 到指定时间后 MM 通过初始化使所涉及到的 MA 改成被动模式, 服务结束后 MM 通过新的初始化使这些 MA 恢复成主动模式, 这种方式需要人工干预, 具有比较明确的目的性。

### 3.3 MEMM 的域内监测

在同一管理域内部, 组播性监测的目标是实时地获得任意两点间的端到端性能, 因此, 需要在任意两个 MA 之间进行双向测试。这部分监测行为对于网络的干扰是全部干扰的主体, 不同管理域应该将其限制在自己的管理域中, 通过设定 TTL 值并且在管理域边界路由器上施加转发控制来实现。

在图 3 所示的域, MM 为组播监测管理者, A、B、C、D 为 MA。经过初始化过程, 所用 MA 均得到了协同监测 MA 的列表, 即任何 MA 均与其他三个 MA 协同进行端到端性能监测。例如, A 将进行 A-B、A-C 和 A-D 之间的端到端性能监测。协同监测开始后, MA 发送测试报文到测试组播组, 同时加入测试组播组, 接收来自其他 MA 的测试报文。测试报文的发送间隔和报文长度是由 MM 预先配置的, 报文中还包含发送 MA 标识、测试报文序号和发送时标。通过对发送间隔和报文长度的控制, MEMM 可以实现指定强度的测试。报文中的发送 MA 标识, 便于接收者过滤测试报文。测试报文序号和发送时标可以使接收者方便地计算单向延时、单向抖动和丢报率这些性能指标。MA 的报告间隔由 MM 预先配置 (例如为 15 分钟)。在报告时间到达后, MA 首先对报告间隔内的性能指标进行预处理, 计算 15 分钟内的指标均值并且选出极值, 然后 MA 将上述特征值报告给 MM。由于经过预处理, 报告信息得到压缩, 可以有效解决 MM 处可能会出现的数据爆炸现象, 同时也减轻了 MM 的计算压力。MM 主要负责将来自 MA 的数据汇总整理, 保存在数据库中, 以便于将来查询。一旦 MM 发现任何端到端性能指标劣化超过阈值, 就会产生告警信息通知组播网络管理系统, 并且与其他系统协同实现故障定位和故障恢复。

上述主动测量方式虽然可以随时监测目标网络任意端到端的性能, 但是为了避免网络带宽的浪费和减小监测系统本身对网络性能的影响, 测试流的强度通常难以达到实际用户组播流的水平, 这会造成监测结果存在一定的误差。因此, 在条件允许的

情况下应该尽量使用被动测量方式获取性能指标。例如, 假设图 3 中 A-D 之间存在组播视频点播应用, 存在由 A 到 D 的组播流, 流量为 50kb/s, 使用组播地址 239.10.10.5。MM 将配置 A 和 D 的 MA 进行被动模式的测量。MA 通过网络监听采集目标地址为 239.10.10.5 的组播报文, 然后将 IP 报头取出并且增加采集它的 MA 标识和时标, 形成该报文的摘录。MA 定期将采集到的报文摘录发送给 MM, 由 MM 根据 A、D 两个 MA 的报文摘录计算从 A 到 D 的性能指标。被动测量结果和主动测量的结果按照一定的权重合成最终结果, 保存在数据库中, 作为将来查询的数据和判别网络性能的依据。假设用户组播报文平均长度为 1000 字节, 形成的报文摘录不足 100 字节 (原 IP 报头+MA 标识+时标), 因此, 测量所增加的网络流量不足用户流量的 1/10, 即被动测量比用主动测量方式达到同样强度极大的减少了对网络的干扰和影响。另一方面, 由于监测对象为用户组播流, 所以监测结果和用户所感受到的结果相一致, 监测的准确度也比较好。

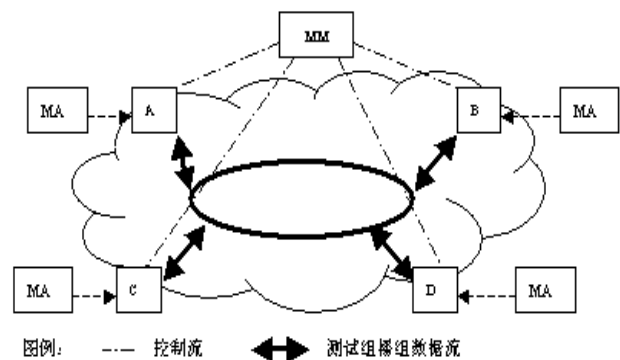


图 3 MEMM 的域内监测过程

### 3.4 MEMM 的域间协同监测

不同管理域之间的组播监测目标是保证域间组播的性能。由于域内组播的性能已经由域内组播监测机制所保证, 所以域间协同监测的重点是监测域间互联链路和外部组播路由协议是否正常工作。因此不必在域间进行任意两点间的双向测试, MEMM 在每个域中各选出若干个 MA, 仅有这些 MA 负责域间监测工作。

域间组播协同监测如图 4 所示。如果管理域 A 和管理域 B 之间存在  $n$  (本例中  $n=2$ ) 条不同外部路由, 那么在每个管理域中可以仅选择  $k_A$  ( $k_A \geq n$ ) 和  $k_B$  ( $k_B \geq n$ ) 个 MA (这些 MA 必须覆盖了所有的外部路由, 例如, A-B 和 A-D 使用域间互连

链路 1, C-B 和 C-D 使用域间互连链路 2) 进行组播监测, 即可实现对于域间组播的监测。

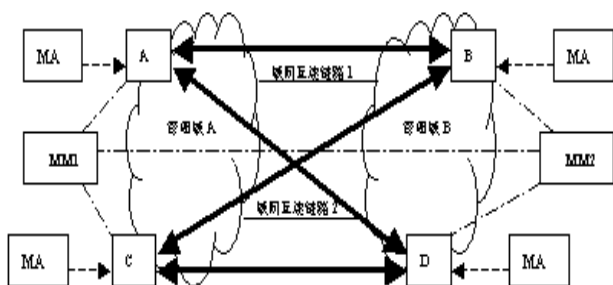


图 4 MEMM 的域间协同监测过程

参与域间协同监测 MA 的选择算法如下(以图 4 为例,  $k_A=k_B=n=2$ ,  $i$  的初始值为 0):

- (1) 在管理域 A 中任选一个新的 MA 作为第  $i$  个 MA, 称为  $MA_{Ai}$ ; 如果不存在新的 MA 候选, 算法失败, 否则, 令  $j=0$ , 执行下一步。
- (2) 在管理域 B 中选择一 MA 作为第  $j$  个 MA, 称为  $MAB_j$ , 使它满足如下条件: 如果  $j < n$ ,  $MA_{Ai}-MAB_j$  使用  $\{ MA_{Ai}-MAB_0, \dots, MA_{Ai}-MAB_{j-1}, \}$  所没有覆盖的域间互连链路; 如果  $n \leq j < k_B$ , 则任选一个 MA。如果能够选出满足条件的 MA 就执行下一步, 否则本轮选择失败, 重新执行第 (1) 步。
- (3)  $j=j+1$ ; 如果  $j < k_B$ , 重新执行第 (2) 步, 否则执行下一步。
- (4)  $i=i+1$ ; 如果  $i < k_A$ , 重新执行第 (1) 步, 否则算法结束。

如果算法失败, 说明存在域间组播不使用的域间互连链路, 请将其从监测链路集合中删除, 再重新利用上述算法求解。由于管理域间的互连链路通常比较少, 即  $n$  比较小, 而且通常比较固定, 所以选择参与域间组播监测 MA 的计算代价是可以承受的。为了便于管理和实现, 域间组播协同监测仅采用主动方式。监测过程中 A 和 C 将监测结果报告给 MM1, B 和 D 将监测结果报告给 MM2。MM1 和 MM2 交换监测结果, 形成最终结果, 保存在数据库中, 作为将来查询的数据和判别网络性能的依据。通过上述监测策略, 参与域间组播协同监测的 MA 数量将非常小, 因此仅会产生很少的跨域组播测试流量。

#### 4. 结论

MEMM 通过预先分布在网络边缘的监测代理, 根据实际情况选择测量主动或者被动方式, 来实现目标网络端到端组播性能的监测。充分发挥了两种测量方式各自的优势, 同时也避免了其不足。MEMM 将监测代理部署在网络边缘, 不需要对路由器(特别是骨干路由器)进行更新, 因此 MEMM 便于部署可以很方便的应用到现有网络之中。MEMM 针对管理域内和多管理域间监测目标的不同, 采用了不同的组播监测策略, 在仅增加少量测量开销的前提下很好地解决了域间组播监测的问题。今后我们将对 MEMM 作进一步改进, 包括增强 MA 功能, 例如, 将已有组播流探测的功能加入 MA, 从而进一步减轻 MM 的负担, 缓解 MM 作为系统性能瓶颈对整体性能的影响。

#### 5. 参考文献

- [1] <http://multicast.internet2.edu/wg-multicast-applications.shtml>
- [2] K. Sarac and K. Almeroth, "Supporting Multicast Deployment Efforts: A Survey of Tools for Multicast Monitoring", March 2001
- [3] A. Swan, D. Bacher, and L. Rowe, rtpmon 1.0a7. University of California at Berkeley, January 1997. Available from <ftp://mm-ftp.cs.berkeley.edu/pub/rtpmon/>.
- [4] W. Fenner and S. Casner, "A 'traceroute' facility for IP multicast", Internet Engineering Task Force (IETF), draft-ietf-idmr-traceroute-ipm-\*.txt, August 1998.
- [5] K. Almeroth, L. Wei, and D. Farinacci, "Multicast reachability monitor (MRM)", Internet Engineering Task Force (IETF), draft-ietf-mboned-mrm-\*.txt, July 2000.
- [6] Multicast Beacon: <http://dast.nlanr.net/projects/beacon/>
- [7] J. Walz and B. Levine, "A hierarchical multicast monitoring scheme", in International Workshop on Networked Group Communication (NGC), (Palo Alto, California, USA), November 2000.