

组播性能测量及其相关工具^{*}

Methods and Tools to Examine Multicast Traffic Performance

陈 瑾 吴文峻

(北京航空航天大学计算机系 国家软件开发环境重点实验室 北京100083)

Abstract In order to better design various kinds of multicast application, improve the performance of multicast traffic, learn more about the effect which is made by network flow control mechanisms, we have to monitor and analyze the process of multicast sessions. Based on the insight, we introduce some important parameters about multicast performance, and give examples and ways for examining the performance, and introduce some useful tools.

Keywords Performance of multicast, Loss rate, RTCP, Visualization

1. 引言

无论是从科学研究,还是从商业开发的角度来看,网络性能评测的研究已逐渐成为一个热点领域。但是直到现在,研究重点一直是单播通讯,组播通讯性能评测的研究相对比较少。虽然二者有许多共性,然而由于组播通讯语义与单播通讯的根本不同,组播通讯性能的评测需要不同的工具和策略。

组播技术是个新兴领域,在发展过程中,大部分注意力被放在如何发展组播协议,如何保证组播骨干网的连接质量,如何演化 MBGP (Multicast Border Gateway Protocol) 组播结构,以及如何保证端节点用户的连接质量等,然而,随着组播技术的不断成熟,组播性能测量技术也在不断发展。现在,有许多自由软件已经支持组播性能测量的部分功能,然而此类的商业软件还不是很多。本文将具体解释组播性能测量方法,介绍并分析现有的一些工具的优缺点。

2. 影响组播通讯性能的因素

2.1 组播通讯的特色

与单播通讯相比,组播通讯有多个动态接收者,这个简单的事实是二者最根本的区别,组播通讯需要把

数据包转发给多个接收者,相应的工具不应该只监视一组用户,而是要处理可能规模非常大的一组用户;不能只监视单一路径上的线路,而是要管理组播树上的所有线路;所进行的性能分析不能只考虑到一组用户,而应该兼顾多组用户。

2.2 有关组播网络配置的重要信息

组播路由协议:有两种组播路由协议得到了广泛的应用,分别是 DVMRP (Distance Vector Multicast Routing Protocol) 和 PIM (Protocol Independent Multicast), DVMRP 起源于 Steve Deering,它在 Internet 上的经典实现是 mroute。PIM 由 RFC2117 [PIM] 定义,在 Cisco 的路由器上得以实现。

链路的类型:是隧道类型还是支持组播的路由类型,这提供了一些有关网络配置的必要信息。

隧道的状态:如果链路的类型是隧道,那么隧道的状态是正常运行还是崩溃了,这有助于掌握组播网络的工作状态。

链路的权值(metric):每条链路的路由的权值。

报文的生存周期(live threshold):是报文的 TTL (Time to Live) 的初值。报文每经过一次路由转发, TTL 值就减小,减为0后报文就不再转发。这可以表示组播的范围。

^{*} 本文得到中国高速互连研究实验网项目 NSFCNet 和973海量信息系统项目的资助。陈 瑾 硕士生,主要研究方向为组播性能分析和测量,高性能计算环境。吴文峻 博士生,主要研究方向为可靠组播,拥塞控制算法。

参 考 文 献

- 1 Black U. ATM: foundation for broadband networks. 1995
- 2 Black U. ATM: signaling in broadband networks. 1998
- 3 张宏科,裴正定. ATM 网络互连原理与工程. 1996
- 4 陈永光,卢锡城. ATM 交换机缓冲策略研究. 计算机科学, 1998
- 5 陈永光. ATM 交换机输入缓冲系统 HOL 阻塞及其改善. 电子工程学院学报, 1998
- 6 Papoulis A. Probability, Random Variables, and Stochastic Processes. 1984
- 7 冯富强,陈永光. VBR 业务下的 ATM 交换机缓冲策略研究. 内部资料, 2000

链路的度(degree):连接这个节点的链路数目。

组播软件的版本号(multicast version):不同的组播软件可能支持的协议不同。

成员的资料数据:包括组成员的姓名、EMAIL地址、使用的工具等等,有助于对组播组的管理。

这些信息的获得对明晰组播的拓扑结构和在组播性能好坏原因的分析中起着很重要的作用。

2.3 评价组播通讯性能的重要参数

报文丢失率:接收到的报文的数目除以期望接收到的所有报文的数目,是评价组播通讯性能的关键因素,其有关时间、组播接收者数目、组播报文数目的分布函数是性能分析的重点。

组播接收者的数目:是指参加特定地址的组播通讯的数目。接收者的数目越多,可能会引起通讯的拥塞,从而增加报文的丢失率,降低组播的性能。

发送者的发送速率:组播发送者的发送速率,有的性能测试报告中表明,这个参数对报文的丢失率没有很大的影响^[1]。

延迟:报文从发出到接收所用的时间,体现了组播发送的质量。

抖动:接收到的前后两个报文延迟的差值,体现了报文发送质量的变化。

整个网络中的组播流量:用于评价组播通讯对整个网络的影响。

特定组的组播流量:用于分析单个组播流的对网络通讯的影响。

3. 组播性能测试范例

本节以 UCL 的 Mark Handley 对几次 NASA 航天飞机录像的组播会话的通讯性能的测试^[1]为例,说明测试以及分析的方法。

3.1 测试方法

测试者写了一个工具来收集组播会话中的 RTP 报文和 RTCP 报文的记录(logger),以及 mtrace(在第4节中有详细的介绍)得到的数据。

RTP 报文记录了同步源,报文长度,发送者的时间戳以及到达时间。

RTCP 报文记录了报告主机的同步源,有关 RTP 成员的一些诸如学名(CNAME)、姓名(NAME)、电子邮件(EMAIL)、电话(PHONE)之类的信息,报文丢失率等信息。

3.2 分析方法

利用这些数据,测试者生成了一棵组播树,并且在链路上标明了报文丢失率,明晰了组播的拓扑关系,以及比较了链路的性能。然后,将分析以下几组数据的关系。

·加入组播的节点数与报文的平均丢失率 每个

RTCP 报文的丢失率域给出在 RTCP 报告间隔内接收者的平均报文丢失率。RTCP 报告间隔依赖于接收者的数目,这个会话中间隔一般在10和60秒之间。曲线表明约有50%的接收者的报文平均丢失率小于10%,约有80%的接收者在一天中的某些间隔内没有报文丢失。然而,80%的接收者在一天中的一些间隔内报文丢失率大于20%,这可以看成是一个阈值,在这之上,音频信号如果不加入冗余信号就会变得不清楚。约有30%的接收者一天中至少有一个间隔中丢失率大于95%。

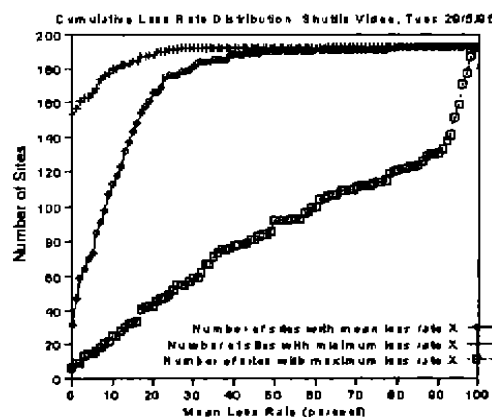


图1 关于接收者数目的平均、最小、最大报文丢失率的分布函数图

·报文数目与平均丢失率 把一天中的 RTCP 报文(包括所有接收者的报文)记录下来,做出平均的报文丢失率的分布函数图。考虑到数据可能受到 RTCP 报文丢失的影响,将原始数据规则化,即给接收到的报文数目乘以权值:

$$\bar{\omega}_i = n_i / \bar{n}$$

其中 n_i 表示接收者 i 接收到报告的数目, $\bar{n}$ 表示所有接收者收到报文的平均数目。规则化(Normalised)可以使得损失了报文的接收者得到相应补偿。

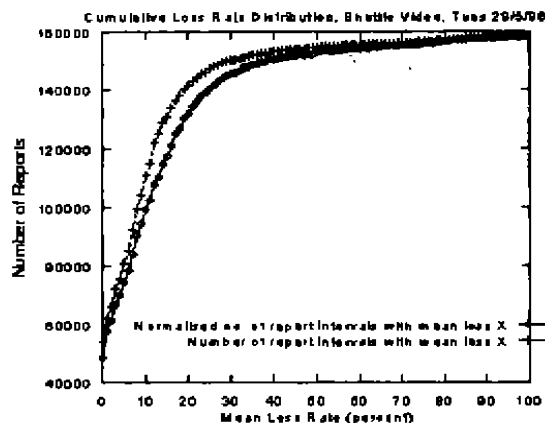


图2 关于报文数目的报文丢失率的分布函数

图2中画出了原始数据曲线和规则化后的曲线,真实情况的曲线最有可能位于两条曲线之间,从图中看出两条曲线很相近,说明报文丢失的影响不大。

曲线表明报文丢失率的分布呈长尾形,大量的报具有较低的丢失率,尽管高丢失率的报文存在,但是很少,80%的报文的丢失率小于20%。

• 报文丢失率与时间的关系 曲线表明了 UCL 处(一个接收者)接收的 RTP 报文丢失率随时间的变化趋势,图3(a)是在1996年5月28日早晨(UK 时间)测得,此时网络处于轻负载状态,而图3(b)是在一天中报文丢失率很高的时候测得。两幅图都体现出31秒的损失周期模式,由于图3(b)中还夹杂其他的损失,体现的不如图3(a)明显,测试者对于损失自相关性的分析中也表明当网络丢失率很高的情况下,这种周期效应也很明显,这种效应可能是由网络中的一些周期效应造成,例如组播路由中的相互作用,少数工作站上运行的某些周期性的 mrouted 进程的作用。

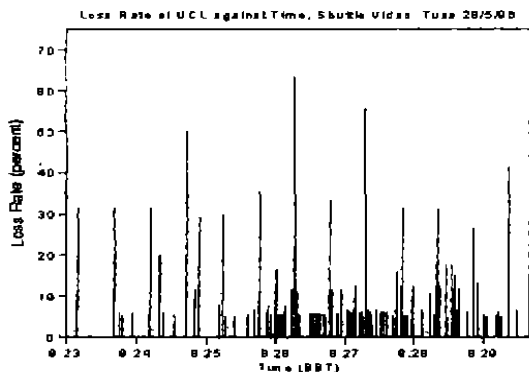


图3(a) 报文丢失率和时间关系

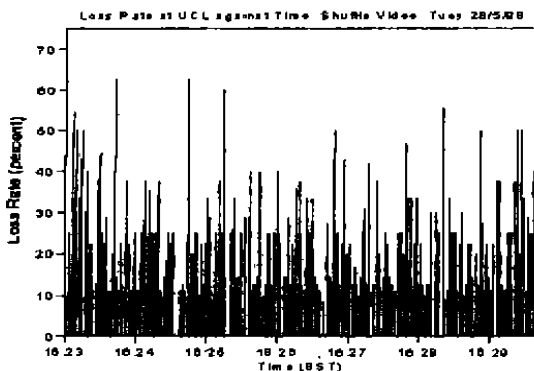


图3(b) 报文丢失率和时间关系

测试者还分析了所有节点的报文丢失率与时间的关系,所有发送者的数据发送率与报文丢失率的关系,以及报文的连续丢失的概率等,限于篇幅,就不再详

述。

4. 现有的性能评测工具及其优缺点

• mrinfo 目前较通用的用于收集 Mbone 节点和隧道当前状态信息的工具是 mrinfo, mrinfo 可以返回链路的类型和状态,表明用的是隧道技术,还是支持组播的路由器;隧道的状态是崩溃(down)还是正常,并返回权值(metric),生存周期(live threshold)等参数。一个典型的 mrinfo 的例子如下:

```
198.17.46.39 (mbone.sdsc.edu) [version 3.255, prune,
  genid, mtrace].
198.17.47.39 -> 0.0.0.0 (local) [1/1/disabled]
198.17.46.39 -> 198.17.46.56 (pmot.sdsc.edu) [1/1]
198.17.46.39 -> 198.17.46.43 (cs-f.vbns.sdsc.edu) [1/1]
198.17.46.39 -> 198.17.46.10 (medusa.sdsc.edu) [1/1]
198.17.46.39 -> 198.17.46.205 (bigmama.sdsc.edu) [1/1]
198.17.46.39 -> 138.18.5.224 (ncosc-mbone.dren.net) [1/
  32/tunnel]
198.17.46.39 -> 192.215.245.10 (mbone.cerf.net) [5/32/
  tunnel]
```

mrinfo 第一行输出显示调查的主机是 mbone.sdsc.edu, 它运行的 mrouted 版本号是 3.255, 并且支持 mtrace, 下面的各行表示到另外一个组播路由器的连接, 括弧中列举了它的相关属性。第一个属性表示权值(metric), 第二个属性表示报文的生存周期(live threshold), 属性中的“tunnel”表示在这两个主机之间的组播路径是通过隧道技术的, “down”表示隧道的状态是崩溃了。

利用 mrinfo 提供的信息, 可以构造组播网络的拓扑结构。

mrinfo 也有着缺陷, 如果 mrinfo 是在 Cisco 的路由器上运行, 执行的是 PIM 协议, 而与之相连的路由器执行 DVMRP 协议, 那么 mrinfo 得不到该路由器的状态, 这样得到的网络状态就不完全了。另一个缺点是, 如果防火墙过滤 IGMP (Internet Group Multicast Protocol) 包, 那么 mrinfo 无法访问防火墙后的节点。另外, 一些站点出于安全考虑, 限制路由器不响应 mrinfo 的调查; 有的路由器配置不对; 有的软件不兼容等原因都会造成 mrinfo 得到的某些信息不准确。

• mtrace 是跟踪两个主机之间的组播路径的工具。它侦听由接收者发送给组播组的 RTCP 报文, 并建立发送者和接收者之间的列表, 它跟踪从选定的发送者到所有的接收者之间的组播路由, mtrace 允许用户看到每秒流过每个网络链接的组播包的数目, 还可以给出针对特定组播地址的网络链路上组播包丢失的数目。mtrace 可以说是监测组播流的状态及性能的最好的工具之一。

mtrace 的运行受很多限制, 版本号小于 3.3 的组播路由器不能正确地支持 mtrace, 另外由于链路上严重的网络拥塞, 造成 mtrace 包的丢失, 遗漏路由状态,

使得 mtrace 无法得到需要的信息, mtrace 的另一个缺点是引入额外的通讯, 增加网络的负载。因而有的路由器给予 mtrace 请求比较低的优先权以减少拥塞, 这样又会造成 mtrace 的请求被路由器忽略, mtrace 得到的信息不全。

• **RTPmon** 结合了 mtrace 的 RTPmon, 它可以加入到一个特定的组播地址, 接收所有接收者的反馈报告(RTCP 报文), 每一个发送者和每一个接收者的报文丢失率显示在一个实时的表格中(参看图4)。点击表中特定单元, 就会显示有关丢包率和抖动的信息, 在 RTPmon 界面中的按钮允许一个用户对特定的组、来源和目的地执行 mtrace, 使得用户可以监测组播流的质量。

• **mhealth** 在 RTPmon 的基础上更进一步, 将一个特定组的组播树型结构(包括丢包信息等), 用实时、图形化的方式显示出来, 使得网络管理者更好地“看到”在网络的哪个部分发生了丢包。mhealth 是用 JAVA 写成, 所以它具有跨平台的特性。但它还是要求首先安装 mtrace。

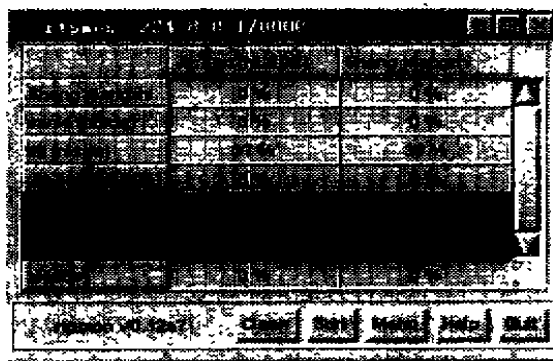


图4 RTPmon 的一个界面

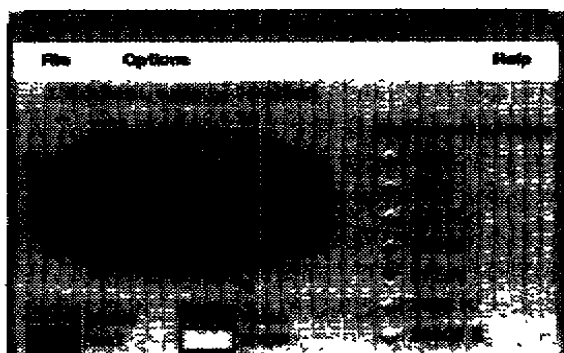


图5 multimon 的界面

• **multimon** 采取了和 mhealth 相反的方法, 不提供一个特定组播组的端对端性能的信息, 而是代以给

出在一个特定局域网中所有组播流量的信息。它借助 TCPdump 来监视网络中报文, 收集一个局域网中流量的数目和类型。

然而, Multimon 是为一个 UNIX 系统开发出来的, 需要 Tcl8.0, Tk4.2, 一个修改过的 TCPdump(内含), xplot(内含), tcl-dp(内含)和 stooop(内含), 所以安装它会花掉不少时间的。

• **Mlisten** 其基本用途是收集一个特定组的成员资料数据。这个组的信息是通过 SDR 报文发布而得到的。Mlisten 一般收集所有已经发布的组, 它也可以调整特征值(例如小于16的 TTL 值)来收听某些特定的组。

使用这个工具的风险在于它加入了所有监测的组, 它有可能因此成为流量的汇集点。目前, 它只有 sunUNIX(SunOS, Solaris)版本可用。有一个版本具有图形化界面, 允许管理者实时查看组的大小, 用户也可以点击一个组查看组中成员的 IP 地址列表。

• **Dr. Watson** 是一个多用途的诊断工具。它能够执行可达性测试, 产生 SNMP 询问, 观察网络流量, 检查路由表信息, 它实现了许多单播和组播协议, 对于组播通讯, 它能监视流量、组播源的流量、传输 IGMP 信息到路由器。它特别有用之处在于用户可以测试路由器和其他局域网主机是否正确地响应加入和撤离组播组的消息。

Dr. Watson 是一个商业软件, 具有详细的文档和支持。但它是一个很底层的工具, 涉及到许多细节, 要求用户具有相当的专业技能, 对于大规模网络管理和组播路由问题的处理不一定有用。

• 基于 SNMP 的工具和组播相关的管理信息库(MIBs) 组播管理信息库(MIBs)的访问需要使用简单网络管理协议(SNMP)。先前的许多 Mbone 中的管理和调试函数没有使用到 SNMP 和 MIB, 这是由于在 Mbone 发展过程中, 人们曾经关心建立基本的连接胜于关心发展可靠的管理工具, 然而, 这个趋势正在发生改变。由于 SNMP 是一个很好的网络管理协议, 一些研究者已经认识到定义组播管理信息库(MIBs)的重要性。现在 IETF 已经有许多的有关组播的 MIBs 被提出, 但还处于修改的过程中, 还没有发展成为 RFC。

使用组播 MIB 的两个重要的免费工具是 mstart 和 mview, 都是由 Merit SNMP-Based Management Projects 完成的。Mstart 询问一个路由器或支持 SNMP 的 IP 隧道来得到各种表信息, 包括: 路由表、网络配置、缓存内容等等, Mview 是一个可视化的管理 Mbone 的程序。Mview 允许用户交互地以各种方法显示 Mbone 的拓扑结构, 收集路由和链接的性能信息, 能够辅助解决网络问题。

5. 现有的组播拓扑可视化工具及其优缺点

为了更加清楚地了解组播网络的拓扑信息,了解它的生长趋势,了解从基于隧道的结构向组播路由的结构发展的过程,组播网络的拓扑信息可视化也是一个组播研究的热点,下面介绍两个软件。

• **MantaRay** 通过域名或者 IP 地址从相应的地理信息数据库(例如 InterNIC, ARIN 等)中获得有关节点的地理位置信息(经度/纬度),然后结合 mwatch 或 mrinfo 得到的数据,诸如组播路由协议、链接类型、隧道状况、链路权值、报文生存周期、每个节点的链接数、组播软件的版本号等,进行相关参数组播拓扑结构的可视化。

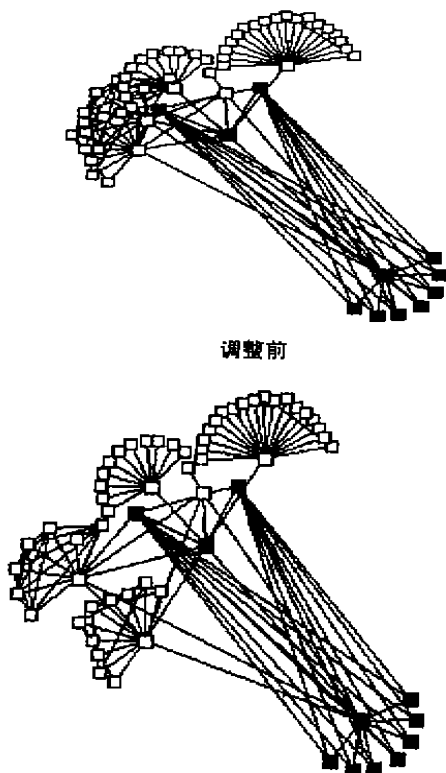


图6 调整后

但它过于依赖节点的地理信息,如果查询不到该节点的地理信息,那么就不会显示该节点。然而,该工具的研究者们在10,000个节点,25,000个链接的组播网中用 mwatch 做的调查显示,表明只有40%的节点

能够在他们的地理信息数据库访问到,由于缺乏地理信息数据库的支持,MantaRay 不能反映完全的组播拓扑信息。

• **Otter** 为了克服 MantaRay 的局限性,研究者们发展了 Otter 来支持基于拓扑的半地理信息位置算法,允许用户从一个更抽象的级别来观察组播网络,算法得分分为两步,首先放置根节点的位置,根节点可以按环形放置,也可以按经度/纬度信息放置。其他不知道地理信息的节点就可以按照一定的算法摆放在根节点的周围。算法的目的主要是避免子节点重叠在一起。用户也可以手工调整节点的拓扑位置。但由于生成的拓扑图表示的不是精确的物理距离,有的时候可能会产生误导。

图6是一个组播网的拓扑结构在手工调整前和调整后的示例,实心的节点是根节点,空心的节点是子节点。

这两个工具都是基于 mrinfo/mwatch 的数据,所以在 mrinfo/mwatch 不被响应的地方,这两个工具也无法使用。

总结与展望 随着组播通讯的发展,组播性能测量的方法以及工具都在不断地改良,虽然有多种多样的监测工具,但是缺少一个完善的组播通讯测量的集成环境,缺少规范化,标准化的测试方法体系。从长远的角度来看,有必要制定组播性能测试方法的标准,生成比较全面的测试辅助工具,融合现有的多种性能测试工具的功能,支持多种应用,具有组播监测、管理等多种用途,将来甚至可以集成到现有的网络管理的体系结构中。

感谢李未老师给我们指定了论文的研究方向,并在研究方法方面给予了指导。

参考文献

- 1 Handley M. An Examination of Mbone Performance. UCL and ISI, January 1997
- 2 Managing IP Multicast Traffic: A First Look at the Issues, Tools, and Challenges, A Write Paper from the IP Multicast Initiative (IPMI) and Standust Forums for the Benefit of Attendees of the 3rd Annual IP Multicast Summit, DoubleTree Hotel, CA USA, 1999. 7~9
- 3 Huffaker B, Kc Claffy, Nemeth E. Tools to Visualize the Internet Multicast Backbone URL: <http://www.caida.org/Tools/Manta/Paper/>