



针对 IPv6 视频流的一种随流测量方案的设计

仇进^{1,2}, 曹争^{1,2}

(1 东南大学计算机学院, 南京, 211189; 2 计算机网络与信息集成教育部重点实验室, 南京, 211189)

摘要: IPv6 环境中使用组播传输视频流情况下, 需要实时监控视频流传输过程中的性能状况, 本文在考虑现有测量系统的不足以及新的方法后提出一种随流测量方案, 该方案通过构造随流测量报文减少对网络流量影响, 同时通过分布式测量点的部署监控传输路径中各段性能状况。

关键词: 组播; 随流测量; IPv6 视频流; 段测量

An Inline Packet Measurement Scheme For IPv6 Video Stream

Qiu Jin^{1,2}, Cao Zheng^{1,2}

(1. School of computer science & engineer, Southeast University, Nanjing, 211189;

2. Key Laboratory of Computer Network and Information Integration, Ministry of Education, Nanjing, 211189)

Abstract: There were many measurement systems to monitor performance in using multicast to transport IPv6 video stream. This paper puts forward an inline packet measurement scheme after considering the shortage of existing systems and new measurement methods. This scheme have two merits, using inline packet to reduce the influence to network traffic, providing segment measurement by distributed measurement point.

Key words: Multicast; Inline packet measurement; IPv6 Video; Segment Measurement

组播是一种有效的单点发送、多点接收的通信方式。目前已经被应用在一些领域如 IPTV, 视频会议等。通常组播应用对传输的实时性, 可靠性都有较高的要求, 所以服务提供商要尽量地保证传输的质量。为了达到这个目的, 需要一个有效的性能测量系统来监测视频流的实时情况。

目前已经有一些系统来测量组播流量, 比如路由层测量组播流量的 Mantra、能显示组播树和传输质量的 MHealth、测量应用 RTP 协议的控制信息的 RTPmon 等。

但目前的测量系统通常存在以下一些问题:

- 只提供端到端的测量, 这样仅仅反应出客户端收到数据时的状况, 不能定位出现问题的设备;
- 通过构造测量报文测量的主动测量方式可能由于并不与实际数据在同一组中, 从而转发路径

授, E-mail: zcao@njnet.edu.cn;

和实际路径不一致, 最终导致测量误差;

- 应用于特定的协议, 比如 RTP 协议。

本文针对以上问题, 提出一种组播数据的随流测量方案, 利用对 IPv6 视频报文头的扩展, 以及测量点的分布测量以达到在对实际网络环境影响较小的情况下测量出数据传输路径中各段的性能。

1 设计目标

构建一个基于 IPv6 的测量组播视频流的测量系统, 可以对受控的任意的 SSM 组播频道提供高效的性能测量和诊断信息, 以帮助服务提供商进行 SLA 确认, 错误定位和服务优化。

- 系统能够进行组播转发树拓扑发现, 并对测量点进行维护和调整;
- 系统能够利用随流测量获得段性能状况并对性能瓶颈提供定位;
- 用图表的形式将性能状态直观的显示给用

作者简介: 仇进, (1987-), 男, 硕士研究生, E-mail: jqiu@njnet.edu.cn; 曹争, (1958-), 男, 副教



户；

2 设计思想及涉及技术

2.1 组播转发树拓扑监控

第一步，根据已注册的设备及组播源信息形成备选设备集。

第二步，从组播源的第一跳路由器开始获得它的组播路由信息。并按照下面的规则 2 将下一跳路由放入队列。

第三步，以广度优先的方式从队列里面取出下一个网络设备，获取它的组播路由信息，并按照规则 1 将此下一跳路由放入队列。

第四步，重复第三步，直到所有备选设备已遍历完或者遍历到主机。

获取组播路由信息的方法，主要基于 SNMP 协议，因为本系统运行在受控网络中，所以对于此网络中的设备具有控制权，可以通过 SNMP 协议来获得 IP multicast route MIB 信息，对于查询路由器得到的 IP multicast route MIB

规则 1：如果 ipMRRouteGroup 为组播地址而且 ipMRRouteSource 为组播源主机地址，那么此设备在转发树上，否则从队列中取下一个设备查询。

规则 2：如果 ipMRRouteNextHopGroup 为组播地址，ipMRRouteNextHopSource 为组播源主机地址且 ipMRRouteNextHopState 为“forward”，那么将下一跳放入队列。

表 1 已注册设备表

设备	关联路由设备
随流报文重构服务器	路由器 1（第一跳）
MMP1	路由器 1
MMP2	路由器 2
MMP3	路由器 4
MMP4	路由器 7
MMP5	路由器 11

例如已注册的设备为表 1 所示的设备，在图 1 所示的传输路径上。获取组播转发树的方法如下：

MCS 向路由器 1 发送 SNMP QUERY，得到 multicast route MIB 信息，按照规则 2 找到下一跳的路由，并放入队列，同时记录 1 作为它们的父节点。然后从队列取出第一个按照规则 1，判断是否在转发树上，如果在，再判断是否在已注册设备表中，在设备表中则标记，否则按规则 2

找到下一跳路由，放入队列时将本路由的父节点设置为下一跳路由的父节点。如果不在转发树上直接取队列中下一个设备，直至队列为空，或者已注册设备表全部标记为经过。最后形成的转发树拓扑如图 3-2 所示。

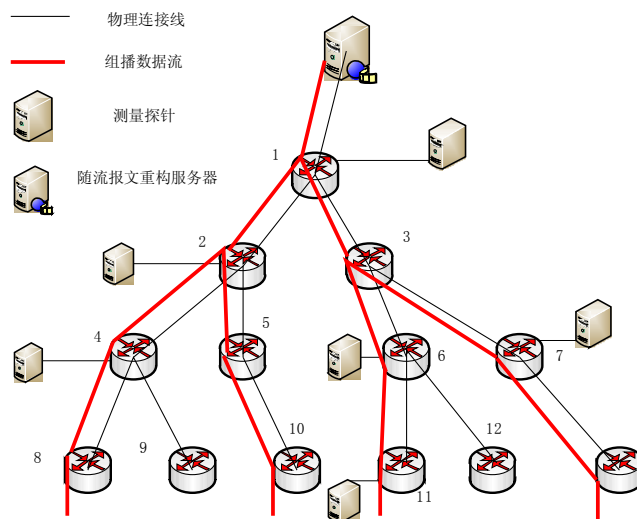


图 1 组播数据传输图

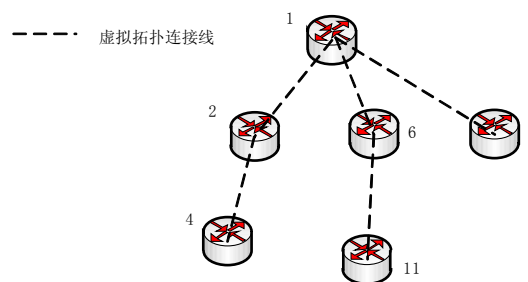


图 2 转发树拓扑图

2.2 随流报文重构

2.2.1 随流测量

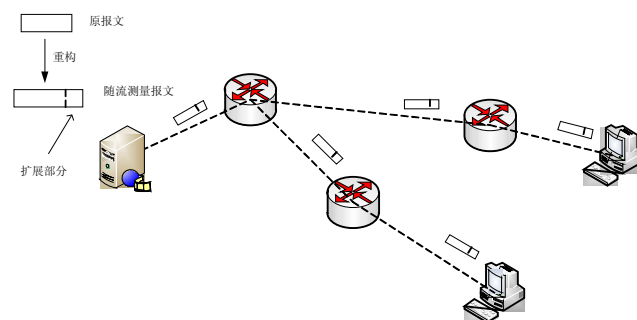


图 3 随流报文传输



所谓随流测量是指在真实的报文中嵌入测量信息，这样在测量设备捕获到实际报文时能得知一些测度信息，用来计算网络性能。如图 3，按照用户要求捕捉一定量报文，并将其重构，这些报文会与原报文一样抵达终端。

随流测量保证用户收到的报文和测量的报文时一致的，因此相比较于主动测量它能更准确的反应通信数据的性能，当然由于延伸了报文长度肯定会带来一定的误差。相较于被动测量，随流测量不需要捕获大量的报文，同时在达到同样目的的前提下，系统复杂性也大大降低。

2.2.2 IPv6 扩展报文头

IPv6数据报报头对 IPV4 做的一项重大改进就是将所有可选字段移出IPv6报头，置于扩展头中。在 Internet Protocol, Version 6 (IPv6) 文档中定义了少数几个扩展报头及其选项，比如 Hop-By-Hop Options，Destination Options，Routing，Fragment, Authentication 等，这些定义的扩展报头中除 Hop-by-Hop 选项扩展头外，其他扩展头不受中转路由器检查或处理，这样就能提高路由器处理包含选项的IPv6分组的性能。

由于目的选项扩展报头不仅容易设置测量信息，而且与其他报头不冲突，因此本系统选择目的选项扩展头作为随流测量信息的载体。将报文扩展为图 4 所示结构：

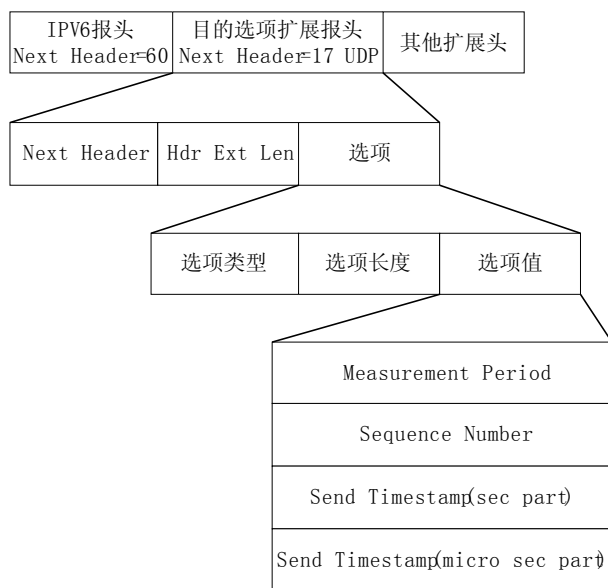


图 4 扩展报文结构

2.3 性能测量方法

根据 ITU-T Y.1541 对 IP 服务的网络性能目标的定义我们可以通过时延，时延变化，丢包三个测度反应网络性能。再跟据段测量需求延伸出组播段时延，段时延变化，段丢包这几个测度。

2.3.1 单向时延及抖动的测量方法

时延由传播时延、传输时延和路由器时延组成，路由器时延又由处理时延和排队时延组成。其中，传输时延、传播时延和处理时延由路径特征和报文长度决定，对于同一条路径，同样长度的报文，它们的时延是相同的，因此被称为固定时延或者确定性时延；每个数据报文要经历的排队时延则是随着网络状况的变化而变化的，这部分时延称为变化时延或者随机时延。本系统根据 IPv6 扩展报头携带时戳信息与测量点接受到信息时的时戳差异来计算。

对于给定的源和端如图 3-5 所示，Ts表示数据包发送时的时戳，Tr表示端在接收到数据包时的时戳，那么单向时延计算公式见公式 (1)

$$D = Tr - Ts \quad (1)$$

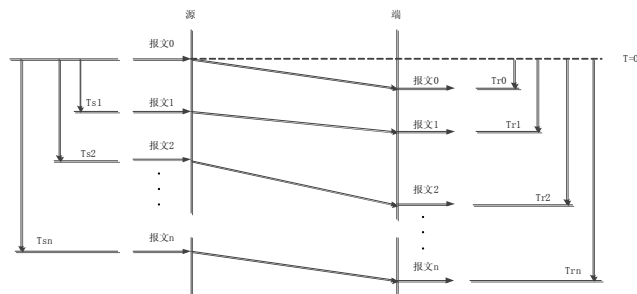


图 5 报文传输中的时延及时延变化

流应用中通常会使用 IP 时延变化的总体范围来避免缓冲区的上溢或下溢。极端的变化将导致 TCP IP 重传计时器阈值增大和也可能知道包重传延时或者不必要的重传。

在这里使用是第 i 个报文的时延与参考时延之间的差来表示 IP 时延变化。其中根据 ITU-T Y.1541 参考时延在 IPDV 对象中一般取最小值，因为这样可以保证所有的变化都为正值。

取 D_i 代表第 i 个数据包的单向时延值， D_{min} 表示一组测量报文中单向时延的最小值，则可以通过公式 (2) 计算时延变化。

$$v_i = D_i - D_{min} \quad (2)$$



2.3.2 单向丢包测量方法

根据扩展报文头中的 Sequence Number 字段测量开始后, 序号逐个递增。在测量点中计算收到数据包的单向时延, 并记录在序号所标识的位置, 一个周期完成后将 Sequence 归零, 若下一个收到的测量报文比之前的 Sequence 小很多, 则之前属于一个周期, 缺少的 Sequence 均计为丢包, 在计算一个间隔时延数据的同时, 我们会统计处这个间隔丢掉的报文的个数 Lost。因此在在一个周期 T 中, 我们设定测量报文总数为 Total = Measurement Packet Count, 因此端到端单向丢包率可由公式 (3) 计算:

$$L = \frac{\text{Lost}}{\text{Total}} \quad (3)$$

2.3.3 组播测量段时延测量方法

段测量中的基本定义如下:

- 1) Src 表示组播源;
- 2) Di 表示组播数据传输过程中, 组播转发树上的一个设备;
- 3) (a,b)(b>a) 表示转发树中一个段;
- 4) <Src, D1, D2, Da, ..., Db, ..., Dn> 表示设备序列;
- 5) <T0, T1, T2, Ta, ..., Tb, ..., Tn> 表示接收到测量报文时的时间序列。

段时延测量就是在 <Src, D1, D2, Da, ..., Db, ..., Dn> 设备序列中任意指定的段 <Src/Da, Db> 的时延。

在组播数据传输过程中, 从组播源 Src 到各个设备在不同时刻发出的测量报文, 在经过设备时记录下时戳, 从而形成:

<T1, dT1.1, dT1.2, ..., dT1.a, ..., dT1.b, ..., dT1.n, dT1>
 <T2, dT2.1, dT2.2, ..., dT2.a, ..., dT2.b, ..., dT2.n, dT2>
 ...
 <Tm, dTm.1, dTm.2, ..., dTm.a, ..., dTm.b, ..., dTm.n, dTm>

因此对于本系统来说, 在时刻 Tk, 任意段

(a,b)的时延为

$$DTk = dTk.b - dTk.a \quad (4)$$

2.3.4 组播测量段时延变化测量方法

段抖动的计算机要基于2.3.3得出的时延矩阵之上, 对于给定的段 (a, b) 上的时延变化 IPDV 矩阵为,

<T1, dT1.ab - dTmin.ab>

< T2, dT2.ab - dTmin.ab >

...

< Tm, dTm.ab - dTmin.ab >

其中dTmin.ab为一段时间内, a, b 之间的最小时延值。

因此每一时刻 Tk, 任意段 (a, b) 的时延变化为

$$vTk = dTk.ab - dTmin.ab \quad (5)$$

2.3.5 组播测量段丢包测量方法

在2.3.3小节段测量基本定义基础上: <L0, L1, L2, ..., Ln> 表示从组播源到传输过程涉及各设备上传输中, 设备有没有接收到测量报文。如果接收到Li值为1, 否则为0。

测量时, 根据单向丢包测量中的方法, 获得从组播源到设备之间的单向丢包情况, 结合转发树形成丢包矩阵:

<T1, L1.1, L1.2, L1.a, ..., L1.b, ..., L1.n, L>

<T2, L2.1, L2.2, L2.a, ..., L2.b, ..., L2.n, L>

...

<Tm, Lm.1, Lm.2, Lm.a, ..., Lm.b, ..., Lm.n, L>

在计算段 (a, b) 丢包率时需要注意的是, 与丢包率相关的是, 基数是设备 a 接收到的报文总数和设备 a 处接收到, 而设备 b 没有收到的报文数。

丢包率为

$$\text{Sum}(Li.b)/\text{Sum}(Li.a) \text{ where } Li.a == 1 \quad (6)$$

2.3 时钟同步

测量涉及多台分布的计算机, 因此需要对它们进行时钟同步, 通常时钟同步有 NTP(Network Time protocol)和GPS两种方法, 然而NTP时钟协议的主要目的是保持一种较粗粒度范围内的时钟同步, 无法很好的消除时滞的影响, 而GPS系统的价格又很昂贵。因此本系统选择论文中同步时



钟的设计。

该设计使用层次同步结构，让节点使用组播的方式来寻找可为其提供同步服务的上层服务器。同时系统通过检测时滞是否稳定的方法来动态地调节节点的同步间隔，使节点能够根据自身的情况做出调整，确保让节点的时钟保持在一定范围内，

3 设计与实现

测量系统的测量对象是被管网络中的组播视频流，由主干网和接入网构成。接入网用户加入组播组后将收到组播源发来的数据，而在转发路径的关键路由器或交换机上都会连接一台测量计算机。同时测量控制服务器用来调度测量任务和计算测量信息。

3.1 总体设计

在测量系统中主要包含三种设备，分别是测量控制服务器（Measurement Manage Server），测量点（Measurement Point），随流报文重构服务器（Inline Packet Reconstruct Server），如图1所示：

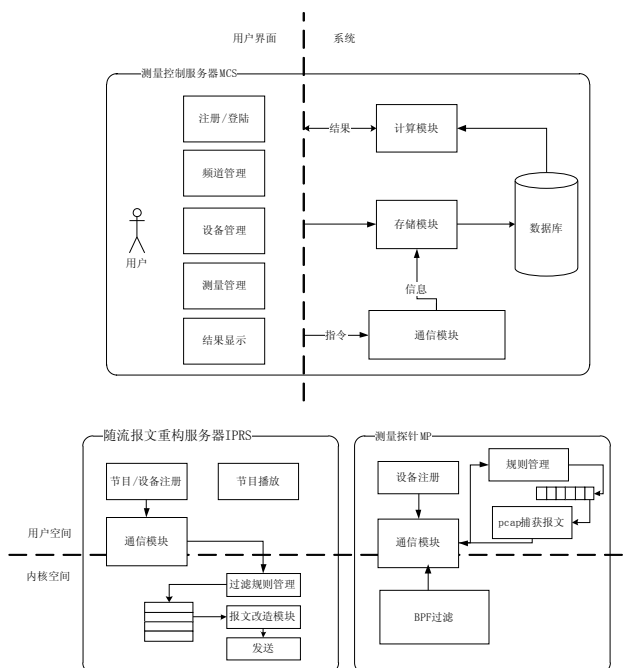


图 6 系统总体结构图

(1) 测量控制服务器功能

- 开通通信服务，等待各设备的注册；
- 根据已注册设备，计算组播转发树拓扑；
- 检测测量点工作状态；
- d)
- 向 IPRS 发送测量命令；
- e)
- 向 MP 发送测量指令；
- f) 接受 MP 发送的消息，并将结果汇总计算。

(2) 测量点的功能

- 向 MCS 注册设备信息；
- 接受 MCS 指令并构成规则集；
- 依据规则集加入或退出组播组；
- d) 依据规则集捕获测量报文，简单计算后发给 MCS。

(3) 随流报文重构服务器的功能

- a)
- 向 MCS 注册设备信息以及播放节目组信息；
- b)
- 接收 MCS 指令，并依此构成规则集；
- c) 内核根据规则集重构报文，形成测量报文。

3.2 测量控制服务器设计

MCS 主要分为两大块，用户控制界面和服务通信模块，用户控制界面负责将信息展示给用户，同时获取用户的需求以触发测量指令，而服务通信模块则主要负责与 MP 和 IPRS 通信。

MCS 工作步骤如下：

- (1) 系统启动后，启动 Web 服务，同时启动服务通信进程；
- (2) 等待各设备注册，注册信息存入数据库；
- (3) 按组播组信息以及 IPRS 设备获取组播转发树，并每隔 T_1 时间，重新获取一次；
- (4) 将组播转发树及频道，设备信息都在 UI 上呈现给用户，等待用户指令；
- (5) 用户下达指令后，根据指令内容，向对应设备 MP，IPRS 发出指令；
- (6) 服务通信进程每隔 T_2 时间，询问设备工作状态；
- (7) 服务通信进程收集 MP 测量信息，并呈现给用户。

注： T_1 ， T_2 均为可配置时间。

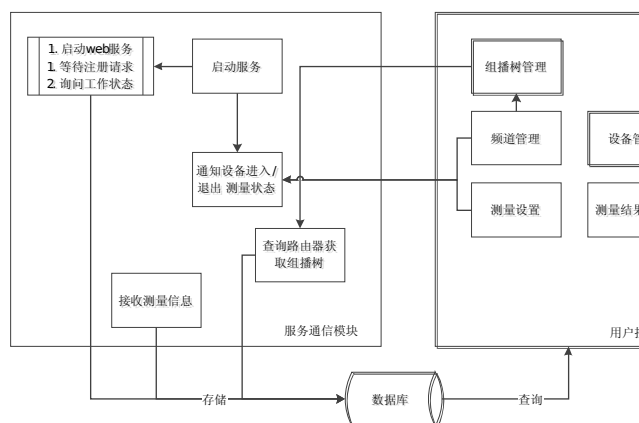


图 7 测量控制服务器的模块设计

3.3 随流报文重构服务器设计

IPRS 同样分两个大模块，第一个是通信模块，负责与 MCS 进行信息交互包括向其注册设备信息以及组播组信息，还包括接收 MCS 指令；第二大模块是内核重构报文模块，重构报文模块基于 MCS 指令得到的规则集对视频流报文进行抽样并扩展报文头。

内核重构模块的设计基于 Netfilter 框架，因为 Netfilter 框架是自 2.4.x 之后整合进 Linux 内核的用来提供防火墙功能的工具，它不仅开源，而且易于配置使用。所有从 IPRS 出去的报文都会经过 IP_LOCAL_OUT 挂载点，在此点的钩子函数中我们可以根据规则集来扩展报文头。

规则集的管理是基于通信模块和内核重构报文模块的通信，通信模块工作与用户空间，而内核重构报文模块工作与内核空间，它们之间的通信使用 Netlink 套接字。

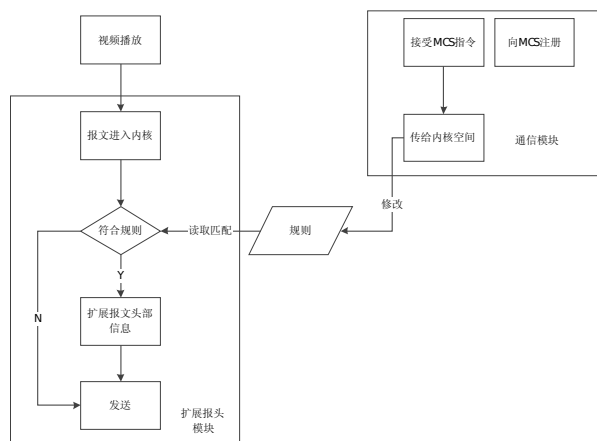


图 8 随流报文重构服务器结构图

3.4 测量探针设计

测量探针主要有三个模块，通信模块，规则集管理模块和报文捕捉模块。通信模块主要负责向 MCS 注册，以及接收 MCS 指令，而规则集管理模块则按 MCS 发送过来的指令加入或退出组播组，并且反应到规则集中。报文捕捉模块则捕获所有收到的报文，并将符合规则集的报文捕获，加上时戳后向 MCS 汇报。

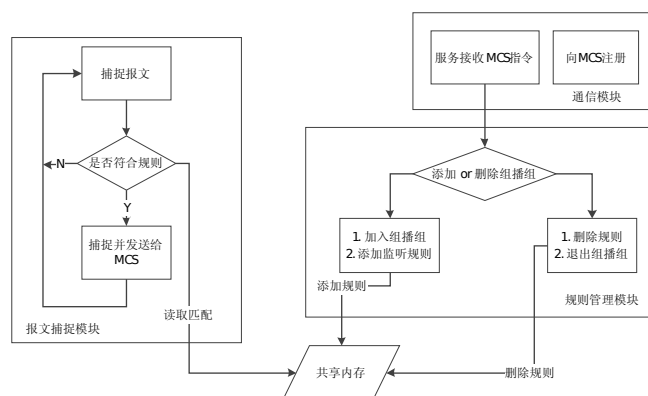


图 9 测量探针结构图

4 总结

本文依据可控网络中视频流传输性能检测的需求设计了一个针对 IPv6 组播视频流的随流测量方案；该方案主要有两个优点：

- 1) 利用 IPv6 扩展报头，尽量减少测量信息给网络带来的负担。
- 2) 通过多个测量点的反馈可以获得分段的数据传输性能情况。

该方案的不足之处在于需要通过 SNMP 协议获得路由器信息，而这需要路由器支持。

参考文献