Researches on the Network Support Environment of the Collaboration-based Virtual Laboratory

Li, Fangmin[1,2] Li, Renfa[1] Ye, Chengqing[2]

1: Dept. of Information Science Xiangtan Polytechnic University Xiangtan 411201

2: Dept. of Computer Science and Technology Zhejiang University Hangzhou 310027

Abstract: With the development of the broadband IP platform, the Internet-based collaboration applications not only improve efficiency, but save money. The paper investigates the network support environment of the virtual laboratory, including the gateway queue mechanism and end-to-end congestion mechanism. Then we analyze the shortcomings of current mechanisms, and present a adaptive end-to-end congestion mechanism based on the UDP which guarantees minimum rate. Finally, we simply evaluate our method and discuss how to improve it.

Keywords: collaboration, virtual laboratory, network support environment, QoS

基于协作的虚拟实验室的网络支撑环境研究

李方敏[1,2] 李仁发[1]

([1]湘潭工学院信息系 湘潭 411201)

叶澄清[2]

([2]浙江大学计算机系 杭州 310027)

摘要：随着宽带 IP 平台的发展，基于 Internet 的协作不但能大大提高研究人员的效率，而且能节省开支。本文以基于 Internet 的虚拟实验室为背景，研究底层的网络支撑环境，包括支持 QoS 的网关机制和端-端拥挤控制机制，以及在此基础上的软件协作环境。然后介绍了我们提出的基于 UDP 支持最少速率保证的端-端控制机制，最后简单评价了该算法以及我们下一步的改进。

关键词：协作，虚拟实验室，网络支撑环境，服务质量

虚拟实验室定义为：它是一个无墙的中心，通过计算机网络系统，研究人员或学生将不受时空的限制，能随时随地与同行协作，共享仪器设备，共享数据和计算资源，得到教师的远程指导以及同行间的相互研讨。

随着计算机网络技术、科学计算可视化技术、智能仪器仪表技术的不断发展，为研究并实现虚拟实验室奠定了基础。如何使研究人员不受时空的限制，使他们能相互协作完成大项目，共享稀缺的仪器设备，提供研究人员的科研效率，大大降低研究费用，在当前竞争激烈的信息社会里，这无疑是提高竞争力的一个有效途径。

网上大学已被认为是现代教育的一种新模式，越来越受到政府和企业及社会的关注。但如何解决网上实验，仍然是一个全新同时也是亟待解决的问题，而虚拟实验室是解决这一问题的一个有效途径。我国教育经费预算紧缩时代还会持续很长一段时间，虚拟实验室有助于提高资源的利用率，节省管理费用的支出。

协作机构的基本技术涉及数据与软件共享、仪器的遥控、具有 Qos 的通信机制、可视化等技术。

虚拟实验室研究，国际上始于 90 年代，有代表性的成果有：

- 环境与分子科学协作机构[1]
  提供一个协同工作环境，并且可远程使用核磁共振(NMR)频谱仪。
- 远程实验环境[2]
  实时参与在 General Atomics 的 D-IIID 托卡马克进行的实验。
- Beamline 7 协作结构[3]
  远程使用位于劳伦斯.伯克利实验室的先进光源以获取空间解析化学信息。
- 医学协作机构[4]
  通过 X 射线照片及超声影象进行的同步或异步远程咨询。

我们研究的目标是建造一个虚拟实验中心原形系统，通过计算机网络实现某个方面的远程实验。研究虚拟实验技术原理，解决建构虚拟实验室的几个关键

技术问题，为建造可用于网上大学的虚拟实验中心提供样板和可行的解决方案。

虚拟实验室研究的内容包括：

- 虚拟实验系统的模型建立
- 仪器设备的远程控制方法研究
- 实验数据的远程采集、处理与使用
- 接口技术研究

本文主要研究网络支撑环境存在的问题以及相应的解决方案。

一、系统模型

我们的虚拟实验室是基于协作构建在 Internet 之上的一个协同工作环境，框架结构如图一所示。

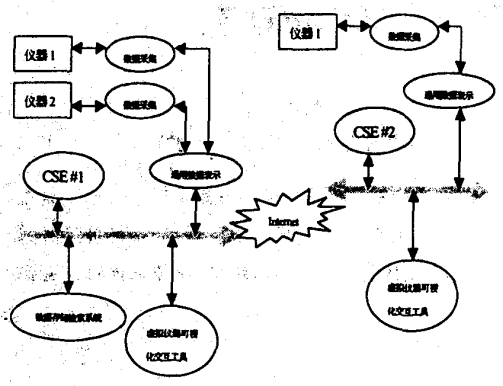

图 1 虚拟实验室系统模型

我们在研究支持 QoS 的底层网络协议的基础上，架构一个协作软件环境（collaboratory software enviroment:CSE），包括：

(1)文本消息交流工具
(2)音频视频会议工具
(3)白板
(4)多媒体数据存储、检索工具
(5)基于 Web 的访问工具
(6)虚拟仪器的可视化交互工具

二、网络支撑环境存在的问题

本文以网络虚拟实验室的原形设计为背景，主要研究网络支撑环境，解决在现有的 Internet 环境下如何有效地传输块数据流和实时数据流。我们将从网关和端主机两个方面着手。

网关机制[5, 6, 7, 8] Internet 路由器的队列管理对不同的数据流量行为有很大的影响，传统的队列管理主要采用 Drop-Tail 队列算法，一旦缓冲区满则丢弃所有到达的包，这种算法显然对具有突发性的 Internet 流量是不合适的。在此基础上提出的随机早期检测（RED）[5]算法当平均队列长度超过一个最小值 min_{th}，则根据概率 max_p 随机丢弃新到来的包，如超过一个最大值 max_{th}，则丢弃所有进来的包，这样在一定程度上缓解了突发性的影响。但 RED 有几个重要的问题还是没有解决：

- RED 没有分离拥挤控制和错误控制，它是通过丢包作为拥挤通知，这样一方面加大了丢包率，另一方面使小窗口的 TCP 源只能等待超时重传。
- 当拥挤发生时，RED 根据平均队列长度丢包，而平均队列基本上不能反应各连接的有关信息，因此即使对使用带宽很少的连接也会丢包，TCP 的内在的拥挤控制策略限制了小窗口的连接获得其公平共享。
- RED 不能限制非自适应的连接(如 UDP)过度地从自适应的连接（如 TCP）不公平地掠夺带宽，从而有可能导致拥挤崩溃。
- RED 不能支持优先级服务。

基于以上的讨论，我们准备设计一种新的或改进的网关队列管理机制取得以下目标：

- 分离拥挤控制和错误控制，如利用 ECN[6]。
- 识别并限制非自适应连接。
- 使各连接能公平地共享瓶颈带宽。
- 减少丢包率，提高整个网络资源的利用率。
- 支持同时大数目的连接。
- 支持优先级服务。

端主机的拥挤控制机制 在我们的虚拟实验室的软件协作环境中，不但要用到 TCP 作为一种可靠的数据传输机制，而且要使用 UDP、RTP 传输实时的音频、视频数据，故端主机的拥挤控制主要解决三个问题，一是改进现有 TCP 拥挤控制的不足，二是研究一种 TCP 友好的基于 UDP 的自适应拥挤控制机制，三是研究一种 TCP 友好、伸缩性好的组播(multicast)拥挤控制机制。

TCP 拥挤控制机制[9, 10, 11] 目前主流的四种 TCP 为：Tahoe 只实现了快速重传；Reno 实现了快速重传和快速恢复；New-Reno 一旦收到一个 ACK 则每 RTT 重传一个丢失的包，当整个窗口的包都被确认之后则终止恢复阶段；SACK 使用附加的 SACK 块能重传多个丢失的包。以上的 TCP 拥挤机制都存在以下问题：

- 多丢失包的恢复问题：虽然 New-Reno 和 SACK 不经历超时实现多个丢失包的恢复，但研究表明当平均拥挤窗口小于 12 时，85%的超时是由于没有触发快速重传引发的。
- 没有解决对 RTT 的敏感性，小 RTT 的连接能获得更多的带宽。
- 如何集成对优先级服务的支持。

我们将对现有的 TCP 拥挤控制机制进行改进，通过实验证明能解决上述存在的问题。

TCP 友好的基于 UDP 的自适应拥挤控制机制[12, 13, 14] UDP 本身没有任何拥挤控制机制，如果建立在其上的应用没有相应的拥挤控制机制，则应用不会对网关的拥挤信令作出任何反应，而是仍然以原有速率发送数据，这样使得 UDP 流能不公平地掠夺大量的网络带宽，从而饥饿 TCP 连接，最后有可能导致网络崩溃。

基于以上分析，我们必须在 UDP 上的应用中嵌入一种类似 TCP 的拥挤控制机制，能对网络拥挤作出合理的反应，从而可与 TCP 连接公平地共享网络带宽。本文将在第四节介绍一种基于 UDP 的自适应控制机制。

组播拥挤控制机制 组播(multicast)和传统的点-点通信有很大的不同，它支持点到多点通信，因此视频会议之类的应用可节省大量的网络带宽，目前广为使用的 Mbone 就是建立在 IGMP 之上的组播骨干网。

在组播环境下，每个发送者可能对应成百上千的接收者，而不同接收者在可用带宽和主机本身处理性能方面的异构性使得其拥挤控制机制更加复杂化，就发送者速率的调节来说，如采用较高的速率则会导致低带宽的接收者大量的丢包，如以最低速率发送则会使具有较高带宽的接收者的可用带宽得不到有效利用。

一般现在较为流行的解决异构性的方法是基于接收者的层次组播技术（RLM）[15]，本文准备在 RLM 的基础上进行改进，通过合适地在瓶颈点设置代理[16]，并且考虑与 TCP 连接的公平性，从而更有效地解决组播拥挤控制。

软件协作环境的建立 在以上对网络协议研究的基础上，我们将构建软件协作环境。为了加快原形系统的研制，我们准备将伯克利实验室的 MASH 项目研制的多媒体平台进行改进和移植。

三、基于 UDP 的自适应控制机制

由于在基于协作的虚拟实验室中，软件协作环境有大量的视频和音频等须在 UDP 上传输，为了在异构的 Internet 环境中有效地支持实时数据传输，很多应用都要求网络能提供最低的带宽保证，本接基于 IETF 定义的框架在端主机实现基于令牌桶的自适应的支持标记的速率调节机制，在网关采用加强的 RED（随机早期检测）队列调度算法对不同的流量进行相应的处理，从而支持 controlled-load 服务。

包标记处理 IETF 的 INTSERV 工作组负责定义关于集成服务的相关协议和标准，为了支持 controlled-load 服务，必须在源主机或网络接入点有相应的侦察（policing）和标记（marking）处理，使进入网络的流量符合通过 RSVP 预流的带宽。

本文描述的机制不需要 RSVP 信令，只是在网络入口点（源主机或 Intranet 和 Internet 的接口处）根据预先协商的流量说明 Tspec 对连接进行侦察和标记处理，然后在路由器根据包是否标记进行区分处理。Tspec 包括以下参数：r_a 说明该连接的平均速率，r_p 说明其峰值速率，b 是应用生成的最大突发块大小，此外还包括包的最大和最小长度。

我们使用令牌桶在网络入口点侦察流量，令牌生成进程根据源的 Tspec 生成令牌，平均生成速率为 t_a，短期峰值速率为 t_p，令牌桶的深度为 b，每当有包被注入网络时如果有相应数目的令牌则消耗这些令牌，这些包称之为顺从流（conformant traffic），并对这些包进行标记，否则归入非顺从流（nonconformant traffic），不标记这些包。

路由器除了标记区分处理外，一般还需有接入控制机制，以便预留的聚合流量带宽不会超过路由器的容量，因为本文的主要目的不是讨论接入控制，故假设聚合流量预留级别不超过路由器的容量。

为了能在路由器处理不同优先级的包，可以采用基于类（CBQ）的分离队列，然后根据队列优先级进行处理，为简化队列处理，本文采用 FIFO 队列，队列调度采用加强的 RED 算法（ERED）[6]。一般的 RED 算法当平均队列长度超过一个最小值 min_{th}，则根据概率 max_p 随机丢弃新到来的包，如超过一个最大值 max_{th}，则丢弃所有进来的包。而 ERED 区分标记和未标记的包，当未标记包的平均队列长度超过 min_{th}（unmarked）时，则根据概率 max_p 随机丢弃新到来的未标记的包，如超过一个最大值 max_{th}（unmarked），则丢弃所有进来的未标记包，对标记包只有队列溢出时才丢弃。其调度算法如下：

```
if (包未标记) {
    if  (Qavg > max_th (unmarked) )
        丢弃该包;
    else if  (min_th (unmarked) < Qavg )
        随机丢弃该包;
    else
        将该包入队列;
} else  if (Qavg > =Qsize)
    丢弃该包;
else
    将该包入队列;
```

端-端控制机制　我们的端-端拥挤控制的目的是要能对网络的拥挤作出反应，并且能与 TCP 流公平共享瓶颈带宽，减少丢包率，提高带宽使用效率。目前一些 ISP 为了让用户合理地使用网络带宽，一般通过价格机制来平衡其带宽的使用，例如采用在单位时间传输的数据量，这样，用户会合理地使用优先级而不会在短期内产生大量的突发流量。基于这样的背景，我们的发送者的速率由两个部分组成，r_m 表示需标记的顺从流量速率，r_u 是不须标记的 best-effort 流量，我们假使接入控制机制保证所有发送者的 r_m 聚合速率不超过瓶颈带宽。发送者的标记引擎根据令牌的生成塑料标记发送的包。发送者根据接收者的反馈信息调整发送速率。

接收者反馈信息分析　接收者定期反馈其接收状态信息，发送者根据接收者的报告计算丢包率和 RTT，在计算丢包率时我们分别计算标记流的丢包率 λm 和非标记流的丢包率 λu，为避免 QoS 的波动，我们采用低通滤波器(1-α) λ+α b 代替 λ，b 表示当前的丢包率，α 大于 0 小于 1，根据 α 的大小可调节 b 对 λ 的影响。

根据 λ 的值确定当前网络的状态，这里要用到两个阈值 λ1 和 λ2，当 λu 小于 λ1 时，网络处于欠载状态；当 λu 小于 λ2 而大于等于 λ1 时，网络处于良好运行状态；当 λu 大于等于 λ2 时，网络处于过载状态。我们根据网络当前所处的状态来调整发送者的非标记部分发送速率 r_u。

我们在前面提到过，接入控制保证所有发送者的 r_m 聚合速率不超过瓶颈带宽，也就是说一般情况下 λm 应该是 0，如果 λm 持续一段时间不为 0，并且超过了阈值 λ3，我们认为网络发生了严重的拥挤或网关可能不支持优先级标记。

发送者速率调整　每当发送者接收到接收者反馈的报告信息后，发送者据此调整其发送速率，其算法如下：

```
if (λu <λ1)
    r_u = r_u + v;
else if (λu >=λ2)
    r_u = r_u * μ;
r = r_m + r_u;
```

如果有足够的令牌可用，则将包标记后发送，否则不标记包。

性能评价和分析　我们在 NS[17] 下选择 λ1=1%，λ2=3，μ=0.7，v=50，α=0.3，其网络拓扑如图二所示，然后进行了初步的仿真实验，证明我们的机制在公平性、带宽使用效率、丢包率、优先级支持等方面基本上能达到我们的要求。下一步我们将进行更全面的实验对所有的参数进行合适的选取，并且让发送者根据接收者的反馈信息自适应地选取合适的参数。

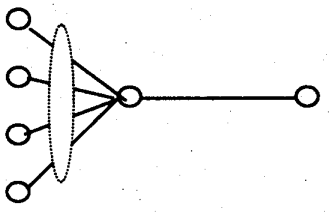

图 2　网络拓扑

四、结束语

网络虚拟实验室是一个涉及面非常广的项目，本文首先简单介绍了虚拟实验室的系统组成，然后分析了现有网络机制存在的缺陷，介绍了我们将进行的有关工作，并提出了基于 UDP 的自适应控制机制。我们的最终目标是研制一个网络支撑环境，在对现有的网络基础设施不作重大改变的前提下提供 QoS 支持，对所有的协议改进或新提出的机制我们都将在 NS[17] 仿真实验环境和自己搭建实验平台进行验证。

致谢　本文受国家自然科学基金资助（编号：69974031）。

参考文献

1. http://www.emsl.pnl.gov:2080/docs/collab/CollabHome.html
2. http://www.nersc.gov/Projects/REE
3. http://www-itg.lbl.gov/~deba/ALS.DCEE/project.html
4. http://www.sils.umich.edu/~weymouth/Medical-Collab
5. Floyd S. and Henderson T., ***Random Early Detection Gateways***

for Congestion Avoidance. ACM/IEEE Transaction on Networking, August 1993, 1(4):297-413.

6. Lin D. And Morris R., ***Dynamics of Random Early Detection.*** In Proc. Of ACM SIGCOMM, September 1997.
7. Demers A.,Keshav S.,Shenker S. , ***Analysis and Simulation of Fair Queuing Algorithm.*** In Proc. of SIGCOMM,1989.
8. Stooica I.,Shenkerr S.,and Zhang H., ***Core-Stateless Fair Queuing:A Scalable Architecture to Approximate Fair bandwidth Allocation in High Speed Networks.*** In Proc. of SIGCOMM,1998.
9. Hoe J., ***Improving the Start-up Behavior of a Congestion Control Scheme for TCP.*** In proc. of ACM SIGCOMM,1996.
10. Mathis M., Semke J. . ***The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm.*** Computer Communication Review, 27(1),July 1997.
11. Feng W., Kandlur D., ***Understanding and improving TCP performance over networks with minimum rate guarantees.*** IEEE/ACM Transactions on Networks,1999,7(2):173-186.
12. Busse I, Deffner B, and Schulzrinne H., ***Dynamic QoS control of multimedia application based on RTP.*** Computer Communication, 1996,19(1):49-58.
13. Sisalem D, Schulzrinne H. ,***The loss-delay based adjustment algorithm: A TCP-friendly adaption scheme.*** Workshop on Network and Operating System Support for Digital Audio and Video, July 1998.
14. Rejaie R, Handley M,and Estrin D., ***RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the Internet.*** In Proc IEEE Infocom, March 1999.
15. McCanne S., Jacobson V., and Vetteeerli M., ***Receiver-driven Layered Multicasting,*** In proc. of ACM SIGCOMM,1996.
16. Amir E., ***An Agent-based Approach to Real-time Multimedia Transmission over Heterogeneous Environment.*** PhD thesis, University of California at Berkeley 1998. http://www.bell-labs.com/user
17 McCanne S. and Floyd S. http://www-mash.cs.berkeley.edu/ns. NS-LBNL Network Simulator,1998.