基于流量工程的端到端网络性能监测系统*)

Measurement System of Internet Dynamics Based on Traffic Engineering

何 飞 李 健 有 悦

(清华大学中国教育与科研计算机网络中心 北京100084)

Abstract Internet traffic engineering is defined as that aspect of Internet network engineering dealing with the issue of performance evaluation and performance optimization of operational IP networks. *Traffic Engineering encompasses the application of technology and scientific principles to the measurement, characterization, modeling, and control of Internet traffic.* There are several research projects are applying their energies to improve the performents of Internet. With the internationalization and advancement of CERNET's operation and management, it's necessary to create the measurement system in the backbone of CERNET. As a result of the application of the solution, we will be able to create Q-Bone in CERNET to provide guaranteed traffic service in high speed.

Keywords Traffic Engineering, Measurement

1. 引言

网络技术的迅猛发展,Internet 的广泛渗透,作为网络的最终用户希望从 ISP(Internet 提供商)得到高质量的服务,所以 Internet 运行性能的研究变得更加重要、更具有挑战性,大型 IP 网络(特别是 Internet 主干网)的性能监控变成了一个非常值得研究和重视的问题。目前在国际上有很多组织致力于网络性能监控的项目,例如 IETF(Internet Engineer Task Force) IPPM 工作组,提出了对于网络性能进行评价的标准,定义大范围网络性能测试的模型,为正确进行网络性能的测量提供高效的测量工具和测量技术[RFC2330];CAIDA(Cooperative Association for Internet Data Analysis)是专门对 Internet 流量进行研究和分析的国际组织。CAIDA 已经开发出了许多的工具,对于 Internet 网络流量进行采集、统计和分析,以期达到全球网络基础设施的合理利用,优化网络拓扑结构,提高网络的使用性能等目的;APAN(Asia-Pacific Advanced Network,亚太先进科研网)的网络测量工作组主要针对 APAN 成员国(主要包括美国、韩国、日本、中国)进行主干网络性能的监测。除了这些国际组织之外,在国际上还有一些针对性能监控的项目,例如:Surveyor 是基于 IETF IPPM 提出的测量标准而设计的测量系统,特别是测量网络端到端单程的性能如传输延迟,丢包率等参数。Surveyor 同时也开发了对于原始数据进行测量和分析的工具;NIMI(National *Internet Measurement Infrastructure*):是为进行大规模网络测试而设计的软件系统,它在 Internet 上建立大范围的监测点,采用"traceroute"进行网络端到端和 hop 到 hop 测量,根据测量结果进行网络路由分析:是否存在病态路由,路由是否稳定,路由是否对称[3];IPMA(Internet Performance Measurement and Analysis project):它的功能是进行网络性能数据采集,并将结果进行分析和实时显示。国家级大型计算机网 CERNET(中国教育与科研网)作为 APAN 的成员国,随着自身运行和管理工作不断走向正规化并开始与国际接轨,以及远程教育和网上招生等应用的实施,迫切要求建立 CERNET 主干网、CERNET 到国际互联节点以及 CETNET 到国内不同网络之间的网络性能监测与控制系统。系统的建立与实施,将起到充分发挥网络效益的重大作用,为国内其他大型互联网络提供示范。

流量工程是网络工程中重要部分,是面向性能评价和 IP 网络运行性能优化的,IETF Internet Draft 定义 Internet 流量工程为:从 Internet 网络工程角度处理运行 IP 网络中的性能评估和性能优化问题,它包括测量,描述,建模和控制 Internet 流量。文[4]RFC 2330 Framework for IP Performance Metric 定义了一个 IP 性能测量尝试应该考虑的框架。为了最大程度地使 Internet 用户和 Internet 供应商共同理解端到端的性能和可信度,框架定义了一些测量的准则。

*)国家自然科学基金资助项目(69904006)。何 飞 硕士研究生,主要研究方向为网络服务质量监控与性能分析。

本文以 CERNET 先进的网络环境为研究背景，根据流量工程所定义的体系结构，设计并实现了网络性能监测系统。

2. 网络性能测量方法

网络性能的监测可分为主动式测量与被动式监听，主动式测量是利用测试工具在指定的时间向指定的主机发送数据包（ICMP、TCP、UDP）进行数据的采集。重要的测量参数包括：端到端的网络丢包率，延迟，吞吐量。经典的主动测试工具有 Ping，TraceRoute，Netperf，Pathchar 等。主动式测量有助于发现问题，跟踪问题，可以进行有针对性的方案设计，有助于分析网络行为特性。被动式监测可以在指定的站点或路由器上进行数据收集和分析，可以持续地监测进、出流量情况。

被动式性能监测方法主要是针对网络流量的采集与分析。采用被动测量方法的优点是不会给网络增加额外的负担，适合于数据量较大的网络流量监测，具体的方法可分为两类：

1. 基于 SNMP 协议的流量采集：SNMP 是 IETF 为 Internet 管理设计的信息交换协议，SNMP 协议简单而易于实现，以 UDP 数据包在管理者与被管对象间传递数据，是目前使用最广的网管协议。基于 SNMP 协议开发的客户端软件 MRTG 是目前应用最广泛的软件，它可以将采集到的网络流量以曲线图的形式实时显示。

2. 基于 RMON 的数据采集：RMON 是一种特殊的 SNMP MIB（Management Information Base）。MIB 描述被管设备所能提供的管理信息。标准的 SNMP MIB 通常只支持对网络设备的一般性管理，RMON 则定义了从网络体系结构的各个层次来管理一个网络所需要的信息。

主动式性能监测方法主要是通过向网络中发送“探测包”来实现测量，测试网络线路的敏感参数为：端到端的网络响应时间、丢包率、TCP/UDP 的吞吐率等。主动的测量方法应用相当广泛，不用特别的设备，可以方便地进行大范围网络的性能监测。它的缺点是“探测包”会增加网络负载。在这里主要介绍几种常用而高效的测量方法：

1. Ping：Ping 命令是使用 ICMP 协议的 ECHO-REQUEST 数据报强制从特定的主机上返送响应。ICMP 是相当低级的协议，它不需要在被检查的主机上运行服务器进程。用户可以根据自己的需求配置 Ping 命令的参数，来获得从源端到目的端的丢包率与延迟时间。但 Ping 的缺点是某些服务器会把 ICMP 的数据包过滤掉。

2. Traceroute：是用来查看一个分组到达其目的主机所经历的一系列路由器。在环路测试中，它是指探测包到目的地的单程路径，而不是返回路径。Traceroute 是通过设置待发送分组的存活时间 TTL（存活跳数）字段来工作的。TTL 值设得比较小时分组不到目的地就会超时，TTL 过期时分组所在的网关应向源主机返送一个 ICMP 出错信息。每次给 TTL 加1，分组就会再往前进一个网关。通过 Traceroute 可以发现错误路由、网关不工作以及目标主机不工作等问题。

3. Pathchar：是用来评价从源端到目的端网络路径中每一个节点的性能。它通过向网络线路上每一跳发送一系列大小不同的 UDP 数据包，来测量跳之间的可用带宽、延时、丢包率和排队等特点。它的主要功能是诊断网络瓶颈，找出网络瓶颈产生的原因。

4 Netperf：是 client-server 结构，有两个主要的元素：netperf 和 netperfserver。它可以用来测试不同类型网络的性能。它可以测试单向网络的吞吐率（包括 TCP 和 UDP）和端到端的延迟。它的缺点是：不合理的测试会给网络带来额外的负载，从而影响网络正常运行。

除了以上介绍的主动性能测试方法，还有为进行大规模精确测试而开发的专用软硬件系统，例如：Internet2的 Surveyor，有55个监控点，可以在1883条路径上进行单向延迟的测量；DOE 的 Pinger，有18个监控点可以监控1261条路径等。

3. CERNET 测量系统的设计与实现

3.1 端-端网络性能监测系统的结构

黑箱式的端到端测量与逐跳测量相结合，测量基于 IP 和 TCP（UDP）的数据传输。由于 TCP（UDP）/IP 的广泛应用，TCP（UDP）/IP 传输是一个真正的网络传输，我们可以从测量得到它的路径属性、时间属性，从而分析得到被测网络的性能。端到端的测量可以跨越不同的 ISP 和 AS，能够反映出用户的实际应用需要，但是它不能回答问题产生的原因。而逐跳测量则有助于发现网络的瓶颈和路由的稳定性问题，经过分析生成的 Policy 可以帮助网络管理，提高 CERNET 服务质量。

测量时间周期采用规律性测量和针对性测量相结合，规律性测量是指按固定的时间间隔进行网络的测量，例如10分钟。固定时间间隔的测量可以完整地反映网络线路的拥塞情况；针对性是指为了发现系统的带宽瓶颈、病态路由、诊断网络问题等而专门设计的测量方案。根据规律性测量的结果针对拥塞发生的高峰时间对于线路进行测试，这样有助于问题的诊断。

3.2 系统的组成

系统主要由三个主要的子系统组成：数据采集、数据分析和数据表示，如图1所示。

数据采集：本系统是针对端到端网络性能进行监控，所以数据采集是采用主动式测量方法。数据采集包括：MMS(Monitor and Measurement System)，包括测量网，测量主服务器，分布在主干网上的测量机，在其它自治域的志愿测量机；Measurement Tools：对于端到端网络状态进行监控主要采用主动式测试工具，例如：ping，traceroute，netperf 等。

数据分析：对采集到的网络性能数据进行分析。它包括：Data 为存储测量结果的数据库；ACS：Analysis and Control System，对于测量结果进行分析、采取有效控制的系统，并将结果通过 Web 的形式显示在网上。

数据表示：将采集与分析得到的网络性能结果通过图表显示在网上，主要包括 Monitor Server：根据测量结果提供基于 Web 的 Cernet Weather Report，Visualized Cernet Topology 服务，或者是以 Email 的方式报告给网络管理员。

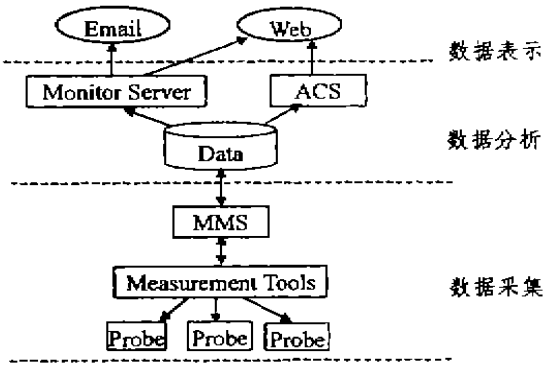

图1 监测系统结构

3.3 监测点的选取

对于网络性能的监测不可能针对所有的网络实体，需要选取网络中具有代表性的一些节点作为监测点。通过记录基点与监控点之间的数据通信过程，来反映端到端网络的性能，所以监控点的选取直接影响网络性能监测的质量，对于监测点的选取有以下几个原则：

(1)可靠性：监测点应该是24小时不断电，例如一个网络中心的服务器。

(2)有效性：有效性表示从监测点获得的数据绝对真实，所以监测点不应该在防火墙内部，它的域名和 IP 应该唯一标示这个点。

(3)代表性：监测点性能的变化应该反映所研究网络的性能变化，所以监测点应该靠近网络的出口，从而避免局域网流量的干扰；另外监测点的负载不能过重，否则由于监测点自身的原因而导致了对整个监测网络错误的指导。鉴于这一点，不宜选择 WWW 服务器。

根据以上原则和 CERNET 国家网络运行的需求，端到端网络监控系统的监测点设计分布如图2所示，其中：

CERNET 主干节点：八大地区网络节点包括：北京、上海、南京、成都、武汉、沈阳、西安、广州。所选取的监测点是 CERNET 到地区网络中心关键路由器。

CERNET 国际测量节点：美国、加拿大、日本、韩国等，监视点为关键路由器。

CERNET 到国内其他网络的监测点也是线路的关键路由器。

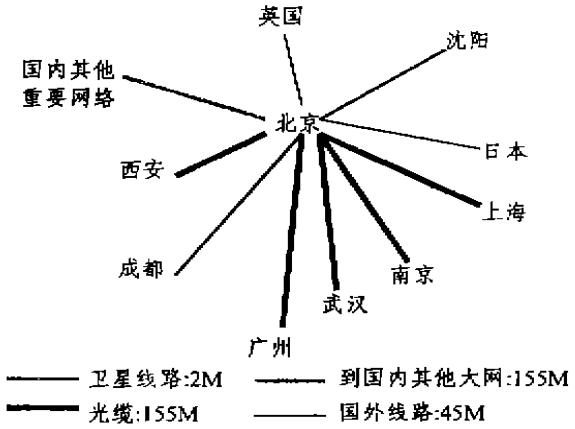

图2 端到端网络监控系统的监测点拓扑图

4. 系统测量结果

(1)测量参数：丢包率，响应时间 以 CERNET(中国教育与科研计算机网络中心)到 CERNET 东北地区网的网络情况为例。此网络状况图为以10分钟为数据采集间隔，连续发送10个56Bytes 的 ICMP 数据包，对于端到端网络的丢包率、响应时间进行实时显示，并统计出丢包率与响应时间的最大值和平均值。通过曲线统计图网络管理员可以直观看到当前线路的情况，以及24小时之内的情况，能够及早地发现问题；系统可根据采集到的原始数据进行分析得到网络的繁忙时间段(网络负载最大的时间段)，为分析子系统提供依据。

(2)TCP/UDP 的吞吐量 以 CERNET(中国教育与科研计算机网络中心)到 CERNET 华北地区网的网络情况为例。测量的条件是：对于 TCP 数据包的吞吐量的测量，接受窗口和发送窗口的大小均为16k Bytes，TCP 信息包为1k Bytes，对于 UDP 数据包吞吐量的测量参数为数据包为1k Byte。对于端到端网络吞

吐量的测试时间间隔为20分钟，因为时间间隔太短会　　增加网络负载。

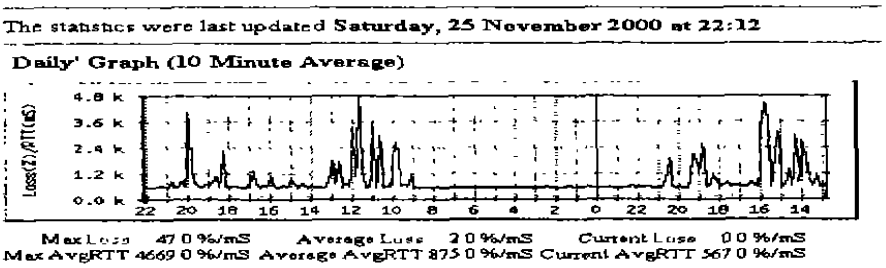

图3　丢包率与响应时间测试结果

图4　TCP/UDP 吞吐量测试结果

结论　本系统是根据 IETF 定义的网络流量工程实现对 IP 大型主干网络性能进行主动式实时监测，并采用直观曲线图的形式将统计结果显示在 Web 页面上。在前面的部分中，我们对本系统的结构，以及各部分的组成进行了介绍，并给出了性能监测的结果。本系统所采用的测试原理、技术和实施方案与国际流行的技术水平保持一致。此系统通过实时的网络性能监测，不但可以及时地了解网络运行状况，而且为网络设计者提供有力的依据。

参 考 文 献

1　Paxson V. End-to-End Routing Behavior in the Internet IEEE/ACM Transactions on Networking, 1997.5(5)

2　Awduche D, et al. Requirements for Traffic Engineering over MPLS. RFC 2702, September 1999

3　Awduche D. MPLS and Traffic Engineering in IP Networks. IEEE Communications Magazine, December 1999

4　Awduche D, et al. A Framework for Internet Traffic Engineering. Internet Engineering Task Force Internet-Draft Working Group, May 2000

5　Paxson V, et al Framework for IP Performance Metrics Network Working Group Request for Comments: 2330 May 1998

6　王继龙，吴建平. 大规模计算机互联网络性能监控模型研究. 计算机研究与发展，2000，37(4)：443～453