

网络性能监测技术综述

唐海娜, 李俊

(中国科学院 计算机网络信息中心 网络室, 北京 100080)

摘要: 互联网络的结构和上面的应用正变得越来越复杂, 出现了对性能要求敏感的应用, 如视频会议、VOIP 等。网络性能监测越来越重要, 它对于 ISP 以及网络研究人员来说是一个大的课题。分析了互联网络的动态特性, 然后概述了当前网络性能监测的理论、技术和工具, 以及国际上目前开展的网络性能监测课题的情况。

关键词: 性能监测; 延迟; 丢包率; 延迟抖动; 带宽; 后挡板; 流量工程

中图法分类号: TP393.07 文献标识码: A 文章编号: 1001-3695(2004)08-0010-04

A Survey: Network Performance Monitoring Technology

TANG Hai-na, LI Jun

(Network Room, Computer Network Information Center, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: The structure of Internet is becoming more and more complex together with its applications. New applications appear which are sensitive to network performance such as video conference, VOIP and so on. Network performance monitoring is becoming more and more important, it is a big subject to ISP and network researcher. The article analyzes the dynamic character of network, and summarizes the theory, technology, tools of network performance monitoring, including the current projects of the world in this field.

Key words: Network Performance Monitoring; Delay; Loss; Delay Jitter; Bandwidth; Tailgating; Traffic Engineering

1 网络性能监测的意义

在过去的几十年里整个互联网发生了翻天覆地的变化。互联网的规模越来越庞大, 结构越来越复杂。在网络上流通的应用也越来越趋于多样化, 出现了对网络性能要求高的应用如 VOIP、视频会议等。互联网络体系结构的复杂化使得网络的运行控制、管理维护、分析设计日趋困难。

网络性能监测提供了一种在实际环境中探索网络特性的手段。网络性能监测是一个从网络设备上采集数据、解码数据、分析数据的过程。它从网络中采集一些具体的指标性数据, 并反馈给监测者。就像人体的体温计一样, 这些数据可以用来作为分析网络性能、了解网络运行动态、诊断可能存在的问题, 甚至预测可能出现问题的“度量值”。这一技术目前被广泛应用于商业和科研领域。在商业领域, 网络性能监测作为网络管理的一部分, 对于网络性能分析、异常监测、链路状态监测、容量规划等方面发挥着重要作用。在科研领域, 网络性能监测技术是实现具体建模、分析的必要前提和手段。

2 网络传输模型

要分析网络性能, 先要分析现有网络的传输模型^[1], 从而了解性能问题出现的原因。现有网络是基于“存储—转发”的包转发模式。数据包在到达路由器先被存储在等待队列里, 然后按先进先出(FIFO)转发。数据包网络性能分析的模型基础是上述的排队理论。其中, 路由器的进出口线路的传输速率、

路由器的延迟、最大包队列长度等相关参数决定了数据包的传输性能。以一个数据包的传输过程为例, 首先, 它从发送端到达第一个路由器的这段信号传播时间 (Signal Propagation Time) 长短取决于物理传输介质, 也就是物理线路的长度 / 信号传播速度 (d_i/c), 这段时间对于局域网来说可能是几毫秒, 但对于无线网来说可能是几微秒, 然后它经历包传输时间 (Packet Transmission Time) 取决于包大小和链路带宽 (s/b_i), 包从进入路由器队列到被发送这段时间是转发时间 f_i (Forwarding Time), 如图 1 所示。

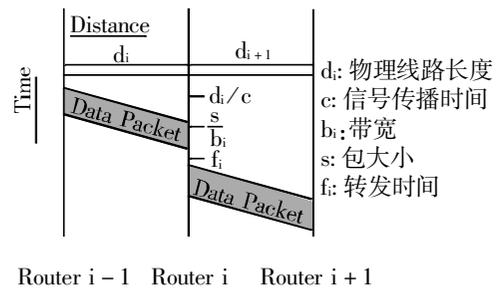


图 1 包传输过程

这样, 包从发送端到接收端的时间可以表示为:

$$OTT(n, s) = \sum_{i=1}^n [\frac{s}{b_i} + \frac{d_i}{c} + f_i] = \sum_{i=1}^n [\frac{s}{b_i} + \frac{d_i}{c}] + \sum_{i=1}^n f_i$$

传输过程中, 接口速率 b 、物理线路的长度 d 、信号传播速度 c 一般是不变的, 变化的是包大小 s 、等待队列长度和转发时间 f 。所以传输时间随包大小、等待队列长度、转发时间而变化。

3 性能监测内容

评价网络性能有几个关键的网络性能参数^[2], 介绍如下:

(1) 连通性 (Connectivity)。它是网络服务中最重要的一

项。除去底层线路问题,另外,路由转发表不一致也有可能导导致网络节点间的断连。路由方面的原因主要是路由的错误配置,或路由收敛不稳定。网络的路由错误可能会从一个路由器传到另一个路由器,从而导致大范围的影响。

(2) 延迟(Delay)。根据 RFC 1242, 存储转发模式下延迟定义为: 输入帧的最后一位到达输入端口和输出帧的第一位出现在输出端口的时间间隔, 即 LIFO(Last In First Out) 延迟。测量 Delay 出于很多原因: 一些应用在端到端延迟大时运行性能会发生明显的下降; 网络上延迟的不稳定(抖动) 导致不能支持某些实时性应用如 VOIP、视频会议。这一参数的最小值意味着在不考虑等待队列导致的延迟下, 仅由于信号传播时间和包传输时间导致的延迟, 借助这一值可以了解无负载条件下的网络特征。延迟值的大小反映了网络的拥塞等级。

延迟有两种定义方式: One-way Delay, 也就是单向延迟。仅仅测量数据包从发送端到接收端的过程。但这种测量有一个很大的问题, 就是发送端和接收端的时钟同步问题。解决这个问题一般采用 GPS。 Two-way Delay, 与 相比, 双向延迟则要简单得多。

由于时钟同步问题, 节点间的网络延迟很难精确测量, 一般都是测 Round-trip Time(RTT)。RTT 指的是一个包被发送到网络上, 它从发送端到接收端后又回到接收端所经历的时间。RTT 参数的影响因素有: 点到点的距离、标准队列的长度和拥塞等级。长期的 RTT 数据可用来作趋势分析和数据关联。RTT 的变化通常意味着配置的变化和拥塞等级的变化。

(3) 延迟抖动(Delay Jitter)。它也被称为 Delay Variation, Delay Jitter 指的是延迟的变化程度。数据包传输过程中在源节点中由于排队和访问而引起的延迟对于发出的所有包有所不同, 一般需要在目标节点中来缓存延迟。很多应用对于瞬间的 Delay Jitter 非常敏感, 而对于平滑的 Delay Jitter 则没有什么影响。

(4) 丢包率(Loss)。它指的是包丢失率。当它很大时, 意味着我们不能依赖网络的服务。如果端到端的 Loss 大于某门限值, 某些应用将不能正常运行(如 4% ~6% 的 Loss 将对严重影响视频会议的运行)。对于 TCP 连接来说, 丢失一个包会等待 4s, 这会造成数据重传, 从而增加了数据的延迟。

(5) 基本带宽和可用带宽(Bulk Transfer Capacity and Available Bandwidth)。基本带宽是指网络的 TCP 一次传输大数据量的能力, 也就是当没有竞争流量时一条路径能给一个流的最大带宽。它由节点间各个链路段中带宽最小的路段(瓶颈链路段)的带宽决定。拥塞控制对于基本带宽非常关键。利用 Packet Loss 和 Delay 来估计可得到的基本带宽是一个热门的研究领域。由于链路一般不是独占使用, 在链路上的其他通信会使应用程序获得的实际使用带宽小于基本带宽, 这个实际使用带宽称为可用带宽。

4 性能测量方法及工具

网络性能监测按采集流量数据的方法可以分为两种方式:

主动(Active)方式。它是指监测者主动发包去探测网络设备的运行情况, 从网络的反馈中分析发出包的具体性能来得到

需要的信息。被动(Passive)方式。它是指监测者被动地采集网络中现有的标志性数据以了解网络设备的运行情况。主动方式具备实时性, 不受管理权限、范围的限制, 但会对网络性能造成影响, 且不准确。被动方式实时性差, 一般需要管理权限, 但准确, 不会对网络性能造成影响。

网络性能监测按采集数据的来源可以分为链路级(Per-link)监测和端到端(End-to-End)监测。链路级的监测一般都是基于设备端口的, 采集该端口链路上的流量数据。端到端到点的监测, 监测两个设备间的性能。所有主动(Active)方式的监测都是端到端(End-to-End)的。

下面从主动方式和被动方式两个方面来说明性能监测的方法, 并列举了一些流行的监测软件。

4.1 主动方法

主动测量的方法是指主动发送数据包去探测被测量的对象的情况。被测对象的响应作为性能评价的结果来分析。测量者一般采用模拟现实的流量(如 Web Server 的请求、FTP 下载、DNS 反应时间等)来测量一个应用的性能或者网络的性能。由于测量点一般都靠近终端, 所以这种方法能够代表从监测者的角度反映的性能。然而由于性能实际上受多种因素的影响(如流量模式、包长分布、服务类型等), 所以这种测量并不准确, 不一定能反映实际网络数据的性能, 而且会对网络的实时性能造成影响。采用主动方法监测时可以从传输层和网络层进行。传输层的协议一般是 TCP 或 UDP。因为 TCP 是面向连接的, 所以测试 TCP 的性能能够反映发送端与接收端的端与端之间的性能参数, 如重传个数、建立和关闭 TCP 连接的时间、平均段大小、吞吐率等。采用这种方式的工具有 Treno, Netperf, Iperf, Ttcp 等。而网络层测量的对象一般是节点、链路或经过这个传输设备的包。监测的属性一般是: Delay, Throughput, Loss, Connectivity, Resource Utilization 等。另外, 路由对性能的影响, 也是网络层要监测的关键对象。网络层性能测量的方法有单播和多播两种。就单播而言, 它通过发送探测包探测发送端与接收端之间路径上的性能参数。采用这种方式的工具有 Ping, Traceroute, Pathchar, Nettimer 等。研究多播的有美国 Umass 大学的 MINC 等。

下面介绍一些常用的主动方式监测工具:

(1) Treno^[3]。它被设计用来测量两主机间的基本带宽。Treno 从应用层来模拟 TCP, 采用与 Traceroute 相同的技术来探测网络。通过发送低 TTL 值的 UDP 包, 沿路上的主机和路由器会回送 ICMP TTL 越界消息, 它具有和 TCP Ack 包相同的属性。它可以测量可得到的最大 TCP 吞吐量。但是, 由于一些路由器可能对 Treno 探测包的反应速度不像它们转发数据包的速度那样快, 因此, 这就意味着 Treno 测试的结果并不能准确反映节点的带宽。

(2) Netperf^[4]。它是一个复杂的基准测试(Benchmarking)工具, 可以用来测量多方面的网络性能。它主要集中于测量基本带宽和传输性能等方面。它可以使用在 TCP/UDP 的 Socket, DLPI, UNIX Domain Socket, ATM API 等环境中。Netperf 包括两部分: Server 端和 Client 端。启动 Netperf 后, 一个 TCP 控制隧道就建立了, 用来协商测试参数的特性。然后, Netperf 通

过传输大数据量来计算得到基本带宽;通过计算指每秒处理的请求 - 响应数得到传输性能。

(3) Ping Traceroute。Ping 的原理基于:该命令引发 IP 层发送一个类型是 Echo-request 的 ICMP 包,而目的方收到这个包之后,将源地址与目的地址进行交换,回应一个 Echo-reply 类型的 ICMP 封包给查询端以确定连线的可行性;Traceroute 与 Ping 原理同,通过设置发送包序列的生存时间 (Time-To-Live, TTL) 来获取包经过的所有机器的时间信息。

(4) Pathchar^[5]。它使用与 Traceroute 相似的技术。它通过发送一系列不同大小的 UDP 包到沿途每一个路由器(递增 TTL),来评价从源节点到目的节点的性能。利用前面各跳和本跳的 RTT 分布,Pathchar 能够分析出本跳的带宽、延迟、丢包率等属性。但是,由于它要发送大量的包到网络,所以会严重影响网络的性能。Pathchar 使用 RTT 作为传播延迟、队列延迟、转发延迟、发送包时间等的和,即 $RTT = 2 \times Latency + Packet\ Size / Bandwidth$ 。通过发送不同长度的包来分析结果。它采用的是最小 RTT,因此能够尽可能地减少队列的影响。RTT/包大小就给出了一个近似的线性图像。那么,可用带宽就可以通过计算与数据集最吻合的直线(使用线性回归算法计算)斜率的倒数得到。为了推导出每一跳的特性,设置了递增的 TTL。这样对于每一跳都会有一个单独的计算。每一跳的延迟和带宽也就可以计算出来。

(5) Nettimer。对于基本带宽的测量,如果能保证两个包在链路上是一直连续传输的(中间没有加入其他包),那么最后收到这两个包的间隔时间就反映了瓶颈链路上处理第二个包的时间。如图 2 所示,因此由第二个包的大小就能算出基本带宽。

Nettimer^[6]正是基于这样的原理。测量分两个阶段:测量整个路径的特性。这个阶段采用和 Pathchar 差不多的技术。

采用后挡板 (Tailgating) 的技术来得到带宽。它发送两个紧接的包:第一个包尽量大;第二个包尽量小。由于每一跳的位置路由器在输出链路上都会引入延迟,而第二个包的延迟总小于第一个包,所以第二个包会紧紧跟着第一个包。把第一个包的 TTL 设为 L,假设第一个包在被丢弃前不会遇到队列延迟;假设第二个包在被测量链路后不会遇到队列延迟。带宽的计算依靠前面带宽计算的积累。Tailgating 阶段最初的 TTL 设为 1,然后一直增长 TTL 直到全都测量完。但是由于测量路径上积聚的误差,这种方法并不准确。

4.2 被动方法

被动方法的监测是在监测点采集真实的网络数据包并统计的。这种方法的监测不会对网络运行造成影响。一种方法是采集和分析由应用程序自身产生的数据。如 Web, FTP, Dns 都维护一个行为的日志以用来分析该应用的性能。一个标准的 API 是 Application Response Measurements。利用 ARM 的信息,很多工具都可以用来监测应用程序的性能。但是,很多应用并没有采用 ARM API 来编码。另外,还有以下方法:

(1) TCPDump。Sniffer 是局域网上的抓包技术。在共享式的网络中,信息包会广播到网络中所有主机的网络接口。Sniffer 通过把网卡设成混杂模式,使主机接收所有到达的信息包。

Sniffer 技术既适合于黑客的使用,也适合于满足网络管理员分析网络性能的需要。Tcpdump 是一个强大的 Sniffer 工具。用尽量简单的话来定义 Tcpdump,就是 dump the traffic on a network。顾名思义,Tcpdump 可以将网络中传送的数据包的“头”完全截获下来提供分析。它支持针对网络层、协议、主机、网络或端口的过滤,并提供 and, or, not 等逻辑语句来帮助你去掉无用的信息。

(2) SNMP MIB。它定义了一系列对象组。一些对象是网络设备必须提供的,作为强制型对象组,另一些对象只在某些设备才用到。对象本身是按层次结构定义的,因此很容易扩充。采集 SNMP MIB 中的数据,可以得到网络设备的各种统计信息。采用这种方法的工具有 MRTG 等。

(3) NetFlow。它是 Cisco 的专用协议,用来根据路由器中 IP 层的信息,了解该路由器所传输的包表头内容。CFlowd 是一个流分析工具,用来分析 Cisco 的 NetFlow 数据。分析的数据可以用来作容量规划、趋势分析、载荷分类等。

(4) Mmdump。众所周知,互联网上的多媒体应用越来越多。但在实际应用中,很难监测多媒体流量对网络的影响。这是因为多媒体应用经常采用控制协议来动态分配端口号。这些协议包括:RTSP, H. 323, SIP 等。虽然控制协议采用一个众所周知的端口号,但传输数据的端口号是动态的。Mmdump^[7]是 Tcpdump 的一个扩展,它从低层采集流量数据,对于每一个多媒体控制协议,Mmdump 都有一个分析模块。在控制协议端口上采集到的流量都交给分析模块。分析模块识别各个单独的控制流来解析出动态分配的端口号。解析模块于是动态改变包过滤表达式来允许关联这些端口的包被捕获。通过这种途径,Mmdump 可以分析多媒体流对网络的负载影响。

5 相关组织

(1) IPPM (IP Performance Metrics)。为了从不同角度研究网络行为,需要定义不同的测度。IETF 的 IPPM^[8]工作组现已经定义了一整套测度用来度量 IP 数据传送的质量、性能和可靠性等。IPPM 的目标就是为服务商和用户提供一个准确的、普遍的网络性能的理解。IPPM 还提供准确测量这些测度的技术,并且鼓励开发测量的工具。RFC2330 定义了一个完整的 IPPM 框架。

(2) RTFM (Real-Time Flow Measurement)。RTFM^[9]于 1996 年成立,用来发展一个测量框架及一个 SNMP MIB。该组织最重要的结果就是在 RFC2722 中提出了一个测量框架(图 3)。

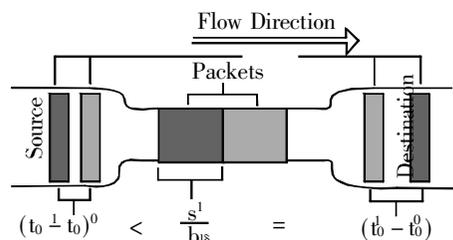


图 2 瓶颈链路上的包传输

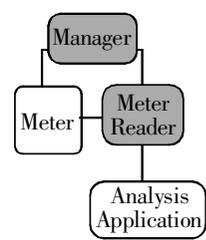


图 3 RTFM 测量框架

该框架包括 Meters, Meter Readers, Managers。Meter 用来采集数据和处理数据。Meter Reader 把多个 Meters 的测量数据传给应用程序,以用来作进一步的分析。Manager 用来配置和控制 Meters 和 Meter Readers。RTFM-Meters 的采集对象包

括性能参数(如 Loss, Delay, Throughput, Jitter, Congestion), 详细分析 IPPM 定义的流, 确定网络上的拥塞路径等。在 RFC 2724 “New Attributes for Traffic Flow Measurement”上, 对 RTFM 作了扩展, 如包括了 IPv6 的流标记、RSVP 流的监测、IntServ QoS 的相关参数等。它有三个主要任务: 总结当前流测量技术工作、提出改进的流模型、提出标准的 IETF 的 Meter MIB。

(3) Internet Traffic Engineering Working Group (ITEWG)。网络流量工程用来优化网络性能。流量工程的一个主要目标是在两个节点间的流量分配到两个不同的路由路径上去, 这能够为网络的容量扩充、QoS 等提供保证。另外, 它制定相关的技术来做路由控制和网络资源分配。目前的技术有基于约束的路由 (Constraint-based Routing), ATM, Frame Relay, MPLS。这个组织还试图解决是否可以利用流量工程的技术在一个 ISP 网络中实现区分服务。

ITEWG^[10] 定义、建立和推荐流量工程的原则和技术。它主要的一点是测量和控制内域的流量工程。它包括预留、测量和控制内域的路由、测量和控制内域的网络资源分配。已经开展的工作有 ATM 和帧中继模型、MPLS、受限路由、区分服务网上的流量工程技术等。它还考虑流量工程在跨 AS 域时遇到的问题。

(4) NLANR (the National Laboratory for Applied Network Research)。它是一个分布式的组织, 其旨在支持美国的高性能连接团体 (HPC Community)。HPC 由两个国家科学基金 (NSF) 支持, 提供高性能研究网络。目前有 vBNS 和 Abilene 网络。Moat 团队 (the Measurement and Operation Analysis Team) 的目标是建立一个网络分析框架 (Network Analysis Infrastructure, NAI), 通过采集原始数据、图形化来分析结果。它的测量包括基于包头的分析、基于 SNMP 的采集、基于 BGP 路由数据的采集。目前的课题有被动测量课题 (OCXmon) 和主动测量课题 (AMP)。

NLANR 主动测量课题 (AMP) 在 vBNS 上实行端到端的监测。目前有遍布美国的 100 个 AMP 监测点。从这 100 个站点采集的数据被加工处理。有三种类型的监测: RTT, Loss 和 Topology。这些监测数据被持续地从监测点采集, 每一个 AMP 监测点发送一个 ICMP 包到其他站点, 并且记录响应时间。另外, 每 10 分钟用 Traceroute 记录到其他站点的路由, 同时任两点间的吞吐量也被计算。采集到的所有数据被送到圣地亚哥超级计算中心加工处理, 并显示在 Web 上。被动测量课题 (PMA) 旨在为高级网络 (如 vBNS, Abilene) 提供协作性的服务支持, 目前已经在 11 个 OC3/ATM、两个 OC12/ATM、一个 FDDI 上展开。采集到的数据使用 CAIDA 的 CoralReef 软件分析, 然后统一送到中央机器。

(5) CAIDA^[11] (the Cooperative Association for Internet Data Analysis)。它位于圣地亚哥超级计算中心, 其目标是提供健壮的、可扩展的全球网络设施。它主要针对商业网络。CoralReef 是基于 OCXmon 建立的, 它是一个综合性的软件, 用来采集和分析流量数据, 还提供可编程 API。这些包由 CAIDA 维护。CoralReef 软件被广泛使用。CAIDA 还维持一些测量工具。

(6) IEPM。斯坦福大学的 IEPM^[12] Group 发起于 1995 年。The Energy Sciences Network (ESNet) 是一个高速核物理网络用来为上千的科学家提供服务, 高能核物理对于广域网提出了挑战。IEPM 是用来监测网络连接和端到端性能的一个项目。PingER 是 IEPM 的一个监测软件, 它采用标准的 Ping 来采集统计数据。每 30 分钟一系列的主机用十一个 100 字节的数据相互 Ping。每个 Pinger 站点的操作是独立的, 没有中心管理机制, Pinger 的主机只提供数据, 需要使用者对得到的数据进行加工处理。

6 总结

随着计算机网络的普遍使用, 网络范围的扩大, 网络性能管理成为一个重要的问题。目前我国网络性能监测方面还没有规模较大的组织和课题。本文概括了目前在网络性能监测方面的发展、技术和理论, 这对于网络性能的研究、QoS 的研究, 以及网络运营等方面都具有现实意义。

参考文献:

- [1] am ́ Varga Supervisors: Ferenc Baumann, Technical University of Budapest [EB/OL]. <http://hsnlab.ttt.bme.hu/~varga/pub/tva-MScThesis-98.pdf> Performance monitoring of IP based networks, 1998.
- [2] T n nes Brekne, et al. State of the Art in Performance Monitoring and Measurements URL [EB/OL]. http://www.telenor.no/fou/publisering/rapporter/R_15.pdf, 2002.
- [3] Treno [EB/OL]. http://www.psc.edu/networks/treno_info.html, 2002-02-11.
- [4] Netperf [EB/OL]. <http://www.netperf.org/netperf/netperfPage.html>, 2002-02-11.
- [5] Van Jacobson. Pathchar: a Tool to Infer Characteristics of Internet Paths [EB/OL]. <ftp://ftp.ee.lbl.gov/pathchar/mrsi-talk.pdf>, 1997-04-21.
- [6] Lai K Baker. M Measuring Link Bandwidths Using a Deterministic Model of Packet Delay [C]. Proceedings of ACM SIGCOMM 2000 Stockholm, 2000.
- [7] Mmdump Project [EB/OL]. <http://www.research.att.com/info/Projects/mmdump>.
- [8] IEFT IP Performance Metrics Work Group Web Site [EB/OL]. <http://www.ietf.org/html.charters/ippm-charter.html>.
- [9] IETF RTFM (Real-time Flow Measurement) Work Group Web Site [EB/OL]. <http://www.ietf.cnri.reston.va.us/proceedings/96mar/charters/rtfm-charter.html>.
- [10] Retf Tewg Work Group Web Site [EB/OL]. <http://www.ietf.org/html.charters/tewg-charter.html>.
- [11] CAIDA Tools Site. [EB/OL]. <http://www.caida.org/tools>.
- [12] IEPM Project [EB/OL]. <http://www-iepm.slac.stanford.edu/>.

作者简介:

唐海娜 (1977-), 山东日照人, 硕士, 主要研究领域为计算机网络安全、网络性能分析; 李俊 (1967-), 安徽人, 研究生导师, 博士, 主要研究领域为计算机网络安全、网络管理。