高速网络性能测量研究*)

王俊峰　周明天

(电子科技大学计算机科学与工程学院　成都610054)

摘　要　近年来,随着网络技术的迅猛发展和互联网(Internet)应用的日益广泛,网络性能测量技术成为网络研究领域的热点。网络测量的研究成果对网络的发展具有重要的指导作用。本文综述了网络测量的重要意义、测量指标体系与常用性能指标;回顾国际上在网络性能测量领域所做的工作及取得的进展;对性能测量方法进行分类,总结有效网络测量方法的评价准则;阐述了高速网络环境下进行性能测量的关键技术与难点。最后对进一步的研究工作进行初步探讨。

关键词　网络测量,性能评价,主动测量,被动测量,端到端指标

Research on High Speed Network Performance Measurements

WANG Jun-Feng　ZHOU Ming-Tian

(College of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054)

Abstract　In recent years, with the rapid development of network technology and the wide deployment of the Internet-based applications, network performance measurement has been becoming the hot spot in network research filed. Results of network measurements play an important role in driving the network-related researches and development. This paper surveys the state of the art in network performance measurements, including motivations, metrics architecture and common performance metrics, and the past progresses in this area. The classification of measurement methodologies is presented, and the criteria for valid measurement methods are summarized. Key issues for high speed network, system scale measurements are detailed. Finally, the paper illustrates potential works in high speed network performance measurements.

Keywords　Network measurements, Performance evaluation, Active measurement, Passive measurement, End-to-end metrics

1　引言

网络技术的迅猛发展特别是基于 TCP/IP 协议的互联网(Internet)的广泛应用深刻地改变了人们工作、学习与生活的各个方面。网络规模的不断扩大、异构性与复杂性的提高、新应用的不断涌现,使系统级网络的可管理性和可控性越来越小,网络性能的可知性越来越复杂。尽管网络基础设施的建设取得了长足的进展,但是用户对网络总体满意程度仍然没有随着网络带宽的增加而相应提高[1]。

网络测量技术特别是性能测量技术是当前计算机网络领域重要研究热点之一。对网络各项指标的测量或推断,评价网络的性能与行为,是寻求建立高效、稳定、安全、可靠、可控网络的重要前提与必要手段。本文主要讨论了进行高速网络下性能测量的目的、国内外相关工作的进展及 IETF 在网络性能测量标准化方面的努力、测量方法的分类与方法的评价、系统级网络测量的主要关键技术等,最后探讨了网络测量领域进一步的研究内容与可行方法。

2　网络测量的意义

由于历史的原因,互联网一直没有得到有效的测量。近年来,由于互联网的飞速发展与新应用的出现,网络的应用已经从单纯的数据传输发展到支持各种类型信息的传输。网络流量与行为发生了极大的改变,动摇了互联网的理论基础:即网络流量模型从基于泊松分布转变为具有自相似性的特性[2,3]。由于对网络协议分布、网络流量特征、用户与网络行为缺乏准确的理解与精确的描述,从而严重影响对网络资源有效利用与网络自身的发展。

网络测量技术是对网络进行认识与深入研究的重要手段,同时也是实施协议工程、流量工程、进行网络管理与规划设计的重要前提条件。其主要意义体现在:

(1)网络测量技术是及时了解网络运行状况,检测网络拥塞、发现网络性能瓶颈,进行网络资源动态配置与管理,保证 QoS 的手段。不同应用对 QoS 的特定需求,也要求必须通过实时网络性能测量以评估当前网络对应用的支持程度。如对于流媒体应用,网络服务提供商(ISP)必须要保证网络的延迟抖动指标,满足用户对 QoS 的要求。

(2)网络性能指标的长期测量与统计分析是进行高性能协议设计、网络设备开发、网络规划与建设、高效网络管理的基础。通过网络测量与网络行为分析可以为网络的科学管理和有效控制、网络的发展与利用提供科学的依据。

(3)网络测量是提供网络辅助控制与管理、增值服务的前提。在高速网络中,基于使用的计费与用户服务水平协议(SLA)的保证都涉及到对网络的有效测量。

(4)网络测量是建立精确网络行为模型的重要手段之一。网络的日益复杂性与新应用的不断涌现,使得不同的应用具有不同的流量特征与行为特征,仅仅利用数学仿真、经典的排队模型进行建模是远远不够的。互联网是复杂的巨系统,基于测量的网络建模与分析是对理论模型进行验证与修正的重要基准。

3　指标体系与测量指标

在进行网络测量过程中,必须要定义一系列定量的参数

*)本文研究得到国家863基金项目(2002AA121032)的资助。王俊峰　博士研究生,主要研究方向为高速互联网性能测量、下一代互联网协议建模与一致性测试。周明天　教授,博士生导师,主要研究领域为计算机网络、分布对象技术等。

用以描述网元、数据链路与端到端路径的性能。这些经过严格定义的定量参数称之为测量指标(Metrics)。IETF BM(Benchmarking Methodology)工作组最初定义了IP网络测量指标框架和相应的基准测量方法。网络性能测量指标的定义、测量方法的研究主要由IETF的IPPM(IP Performance Measurement)工作组制定相关的建议和草案。文[4]定义了进行网络性能指标的框架,并规范性能指标所必须遵循的标准:(1)测量指标的定义必须是严格、具体的;(2)指标测量方法的可重复性;(3)指标的无偏性;(4)指标必须是易于理解的。

同时,IPPM还建议了相应指标测量方法需要满足的特征:测量方法的可重复性(Repeatable)、测量方法与指标的连续性(Continuity)特性。

当前所进行的网络测量主要集中在对网络性能方面的测量。与之相关的指标根据是否被IPPM标准化来进行区分,可分为两大类:

(1)标准化测量指标。IPPM工作组在定义与规范网络性能测量框架的同时,还标准化了一系列重要的网络性能指标。

·连通性指标(Connectivity)[5]。连通性是指在某时刻t,源、目的地址间对某种类型数据报的可达性,是描述网络可用性与可靠性最基本的指标,也是网络提供各种上层服务最基本的条件。

·单向延迟(One-way Delay)[6]。由于网络延迟的不对称性与应用对不同方向上延迟的不同需求,如在FTP或VOD服务中,应用的数据主要来自于下行的数据报,单向延迟指标可以比较准确地反映网络实际向应用提供的服务水平。

·单向丢包(One-way Packet Loss)[7]。网络发生拥塞使路由器缓存溢出或数据传输延迟过大而导致数据包的丢失。丢包进一步造成数据包重传,网络负载增大,性能恶化。因此单向丢包率是描述网络当前负载状况与进行网络性能预测的重标指标。

·时延变化(IPDV)[9]。时延变化定义为对于一给定网络流中,数据包之间的时延变化程度。IPDV对VoIP、网络视频会议等实时交互和流媒体应用的性能有很大的影响,这类应用对服务的基本要求是数据包的定期到达,较大的IPDV意味着应用可能出现停顿或中止。此外,IPDV也是网络负载特征的重要表征[11]。

·往返延迟指标(Round-trip Delay)[10]。单向延迟指标无疑可以反映网络的运行状况,进行单向延时测量必须在端到端系统间分别记录数据包发送和到达时刻、端到端系统间且必须进行时钟同步。进行往返延迟的测量因测量方法简单、易部署与容易理解的优点,仍然得到了广泛的应用。

IPPM还标准化了其它几个指标,包括单向包丢失模式(One-way Loss Pattern)[8],块传输能力(Bulk Transfer Capacity)[12]等。

(2)非标准化测量指标。标准化的网络性能指标主要侧重于对网络中端系统间性能的研究,这类指标直接反映了终端用户使用网络时可获得的性能。非标准化网络性能指标主要是面向网络的指标,有链路利用率(Link Utilization)、网络带宽(Bandwidth)、网络流量(Traffic)及突发性(Burstiness)等。其中,带宽指标是面向网络性能测量中最重要的指标。

网络带宽指标可根据其面向链路或路径分为链路容量(Link Capacity)、路径容量(Path Capacity)、链路可用带宽(Link Available Bandwidth)和路径可用带宽(Path Available Bandwidth)。

链路容量为任意时间间隔Δt内,向链路发送的最大数据包传输速率。

路径容量C定义为一条路径中最小链路容量,即:

$$C=\min_{i=1,\cdots,n} C_i \quad (1)$$

C_i为第i条链路的容量。路径容量又称为路径瓶颈带宽(Bottleneck bandwidth)。

链路/路径可用带宽是指当存在背景流量时,上层应用在该链路/路径上实际可获得的最大传输速率。

路径容量与路径可用带宽又分别称为狭窄链路(Narrow Link)带宽和紧张链路(Tight Link)带宽。

4 网络测量基础设施和项目

用户的持续增长与网络流量指数级的增加,如何详细测量网络性能使用户和ISP能够从各个层面上了解网络性能、定位性能瓶颈变得越来越重要。建立互联网性能测量基础设施(Infrastructure)可以在一定程度上满足越来越多的网络性能测量需求。由于网络规模及其复杂性,测量基础设施的建立与部署极为复杂困难,但是在协作网络间建立测量基础设施是可行的[14]。在IPPM致力于网络性能指标标准化同时,政府机构、不同的学术组织与企业分别根据不同的需求,建立了相应的网络性能测量基础设施。本节简要介绍几个著名的测量基础设施。

(1)网络分析基础设施(Network Analysis Infrastructure,NAI)[15]是由网络应用研究国家实验室(NLANR)所建立的网络性能测量基础设施,它由两个项目所组成即被动测量与分析(Passive Measurement and Analysis PMA)和主动测量项目(Active Measurement Project AMP)。PMA利用在不同测量点的流量监测工具OC3MON记录网络中流量的包头信息[17];在AMP体系结构中,多个测量点构成全连接网,测量点之间周期主动地发送探测包,实现对往返延迟、包丢失、拓扑和吞吐量的测量。利用DAG6被动网络测量卡,NAI已支持对OC192c/10GE链路上数据包的捕获。

NAI不仅采集反映网络性能的各类数据,同时,NLANR对数据的分析进行了深入研究,开发了一系列数据分析与可视化工具。NAI也是到目前为止最大的网络性能测量基础设施。

(2)NIMI(National Internet Measurement Infrastructure)[17]是在Berkeley大学Vern Paxson设计的NPD(Network Probe Daemon)基础上建立的第一个进行全球大规模端到端Internet行为测量的基础设施。NIMI采用层次结构与主动测量技术,在网络中不同测量点部署大量的测量探针(Probe),实现对网络性能的测量。同时,NIMI还是一个开放的基础设施,任何性能指标的测量,都可作为第三方的测量工具集成到基础设施中。目前已经集成的测量工具和实现的测量指标主要有:

Ping:实现往返延迟、包丢失、包失序(Reordering)的测量;

Traceroute:实现对路由动态性的测量,分析路由的稳定性、抖动、路由循环等;

此外,还利用PathChar进行带宽的测量,TReno实现对块传输能力的测量等。

(3)CAIDA(the Cooperative Association for Internet Data Analysis)[18]是受NSF和DARPA所资助,由商业、政府与研究机构所组成的一个进行网络数据分析的合作组织。由CAIDA所开发的CoralReef是一个综合的软件包,可以实现对网络流量的被动测量与分析,还向高层应用提供了用于数据包捕获、数据分析、报表生成的编程接口(API)[19]。CAIDA开发了其它一系列测量工具,如Skitter用于网络拓扑的测量、反映路由的变化与路由的可视化以及流分析,cflowd对网络流量数据的分析[20]等。

其它网络性能测量基础设施还包括对 IPPM 定义指标进行测量的 Surveyor[21]、由 Stanford 大学线性加速器中心(SLAC)所主持的对高能核物理网络性能进行监控的 IEPM(Internet End-to-end Performance Monitoring)[22]、网络服务提供商 Sprint 建立的对其骨干网络性能进行监控与评估的 IPMON(IP Monitoring Project)项目[23]等。

5 网络指标测量方法与评价

5.1 测量方法分类

性能测量方法根据其实现手段主要可分为主动测量与被动测量两种。

主动网络性能测量按照一定的策略主动向网络中发送探测数据包(Probe Packets),通过对探测数据包所受网络影响而发生特性变化的分析,计算出所要测量的性能指标。如 Ping 发送 ICMP 包,计算网络的往返延迟与连通性;Traceroute 实现对路由的测量和网络拓扑的推断,Pathchar 进行带宽的测量等。

与主动测量相对应,被动性能测量不需主动向网络中发送数据包,而是在测量点捕获网络流量中数据包报头信息实现对网络行为的分析。

由于主动测量需向网络发送数据包,数据包本身会对网络性能造成影响,例如在网络发生拥塞情况下,探测数据包可能会使网络性能进一步恶化,造成测量结果的失真,影响网络性能指标的可用性与可靠性。被动测量不需向网络中发送探测包,不会对网络的性能造成不良影响,捕获的网络流量数据可以实现对多种网络性能指标分析与网络行为建模。如利用 Tcpdump 捕获的数据包记录(Trace),可计算网络协议的分布、链路的利用率等多项指标。被动测量只能获得局部网络性能信息,因此主要用于网元级性能指标的测量与评估上,而主动测量由于其实现上的灵活性,可以实现面向应用的端到端网络性能测量。由于测量任务的复杂性,性能测量趋向于采用主动与被动测量相结合的方法,如利用主动测量确定网络的整体性能,而当网络发现异常时,则采用网元级的被动测量方法确定问题所在位置,进行故障诊断。

在运营的网络中,网络中的网元设备如路由器或交换机有部分用于网络管理的数据,对这些数据的采集可以提供部分网络性能信息。如周期读取路由器管理信息库(MIB)中的信息,可以实现对链路利用率、流量等网络性能指标的计算。这成为被动测量之外的另一种网元级网络性能测量方法。对网元设备存储信息的存取,涉及到访问权限控制、系统的安全性、对网络系统性能的负面影响和可扩展性等问题通常只能由网络运营商所实施,测量的指标也有限。

IPPM 根据指标实现的复杂程度,将测量方法粗略分为以下4类:

直接测量法 即对部分网络性能指标,通过直接测量或轮询网元中存储信息就可实现的指标测量方法。如利用 ping 工具可以直接实现对某条路径在给定大小 ICMP 数据包情况下的往返延迟的测量;

分解测量法 通过对某指标的测量,可以推导出组成指标的不同分量指标的测量方法。如对单跳链路在不同大小探测包情况下延迟的测量,可以估算出链路的传播延迟;

组合测量法 从对一系列简单易实现指标的测量,实现对指标进行测量的方法;

推理法 根据指标在过去时刻的计算值,在一定的时变模型下计算指标当前值的方法。

5.2 测量方法标准

前一小节讨论了测量网络性能指标的方法及其分类,在实际的测量中,同一指标可由不同的方法来实现,要实现对网络性能指标精确、有效的测量,测量方法必须具有以下特征:

可重复性 在相同的条件下,对指标的多次测量结果应是一致的;

连续性 当测量条件发生微小变化时,测量结果的变化也应是很小的;

稳定性 网络的复杂性与测量条件的不可预知性,指标测量过程中会出现指标值的不稳定性与误差。有效的测量方法必须能够确定出现不稳性与误差的原因,最小化对测量指标的影响,且具有量化这种不稳定性和误差的能力;

测量时间 从实施指标测量开始,到获得指标所需花费的时间。测量时间的实时性反映了对网络运行状况的描述能力;

失真度 除了被动测量利用在测量点捕获的流量数据进行指标计算与网络行为分析外,主动测量与基于轮询网元设备中信息的测量都会引起被测网络行为的变化,造成测量指标的失真。如采用主动方法测量网络路径带宽时,需向网络中注入大量的测量流量,额外的测量流造成被测指标的失真;轮询网元设备致使设备性能的下降,也引起被测指标的失真。测量方法对网络行为的影响程度是设计性能测量方法所必须考虑的另一重要因素。

可重复性与连续性是进行网络性能指标测量的前提条件。测量方法的稳定性、测量过程的实时性是进行有效网络性能测量的重要保证。此外,测量方法的可操作性、灵活性、可扩展性也是进行指标测量所需考虑的因素。

6 系统级性能测量关键技术

针对不同的网络性能指标,如带宽、单向延迟、包丢失等,已提出了多种测量方法并对其进行了深入的研究。集成各种指标测量方法,建立高效、可扩展的分布式网络测量系统是实现大规模网络控制、管理与性能评估的前提。本节主要讨论系统级性能测量所面临的主要问题与研究进展。

6.1 数据采集方法

广义上讲,不论是主动测量、被动测量或基于轮询网元中信息的测量,均涉及数据采集问题。文[11]设计了一种利用包对(Packet Pair)进行有效端到端时延变化的主动测量方法。包对之间的时间间隔服从泊松分布,即通过控制样本点的时间间隔来计算时延变化,对网络负载特征进行估计。在被动测量中,由于网络传输速率越来越高,利用基于软件的数据包捕获方法如 libpcap 软件包已经不能满足需求[24];利用专用硬件如 DAG 采集卡,因其代价昂贵,在大规模网络性能测量中的使用受到一定的限制[25]。抽样方法是高速网络环境下数据采集的有效手段。Zseby 采用分层抽样方法进行单向延迟的被动测量[26],Duffield 研究了根据流大小来确定采样频率的准确网络计费算法[27]。

常用的抽样方法可以分为以下几类:系统抽样、简单随机抽样和分层抽样[28]。在系统抽样方法中,样本点之间的间隔是相同的,由于被测指标本身可能存在的周期性特性,抽样过程可能会出现与被测指标的变化同步的现象,形成对指标的有偏(Biased)测量[4]。几何抽样、泊松抽样是常用的简单随机抽样法,由于其无偏性(Unbiased)和良好的数学特性,成为 IPPM 所推荐的数据采集方法。分层抽样根据样本总体(Parent)的分布特性,将总体分为多个组,然后应用系统抽样或随即抽样法对每组中的对象进行抽样。在利用了总体分布信息情况下,分层指标具有良好的性能[29]。由于需要总体分布的先验知识,分层抽样方法难以用于对指标的在线测量(不能事先获得进行指标计算的总体分布特性)。

由于每个样本的采集均需一定的处理、存储与通信代价,寻求具有较小代价、测量指标精度可控、支持在线指标测量的

抽样方法是数据采集中重点研究的内容。

文[30]提出了一种基于拟合的自适应抽样算法(Fitting-based Adaptive Sampling Methodology,FASM),FASM 本质上是一种改进的分层抽样方法,它利用已获得的样本来对总体的变化趋势进行估计,通过自动调节当前样本与下一个样本之间的间隔,实现在线的分层抽样。对网络流量指标与单向延迟指标的测量实验表明,在保持相同的误差情况下,FASM 与泊松抽样和几何抽样相比,大大减少了所需的样本数量,降低了数据采集过程的代价。如何针对不同被测指标估计 FASM 相关的参数是应用该算法时所需解决的问题。

6.2 时钟同步方法

进行大规模网络性能测量所面临的另一课题是时钟同步问题。如带宽、单向延迟及其衍生指标(单向带宽、时延变化等)均需要测量点之间的时钟同步。由于部署与实现的限制,当前的性能测量主要是将各测量点分布在被测网络的边缘,进行端到端的性能测量。因此时钟同步主要在端到端的测量点之间实现。

端到端的时钟同步,从同步时钟的来源可分为利用外部时钟源的同步方式与测量点之间的自同步。常用的外部时钟同步方式有基于全球定位系统(Global Positioning System,GPS)和采用网络时间协议(Network Time Protocol,NTP)[31]的时钟同步。GPS 方案可获得准确、高可靠性的时钟同步,其精度达纳秒级,但其代价昂贵,部署条件受到空间地理位置的限制,难以使用在大规模网络测量中。NTP 时钟同步机制则是将测量点的时钟与层次化结构(Hierarchically-structured)的 NTP 服务器的时钟进行同步,最高层的 NTP 服务器利用 GPS 接收器与外部高精度时钟源同步。时钟信息是通过网络从服务器传送到测量点的,NTP 同步方式的精度较低,通常在几毫秒到几十毫秒。对于同步要求较高的指标测量,如延迟、单向带宽等则不可用。

不利用外部时钟源,实现端系统间的相对同步是近来研究的热点。其基本思想是在一定的时钟模型下,在端系统间周期发送探测数据包并构造单向延迟序列,并假定同步情况下单向延迟序列中最小的单向延迟具有相同的值,分析实测单向延迟序列中的隐含的时钟动态性,在消除由于时钟动态变化而造成的时钟偏差(Clock skew)后,实现端系统间的相对同步。因此,自同步所需解决的关键问题是首先建立合适的端系统间时钟模型,利用单向延迟序列检测出时钟动态变化的位置,最后消除时钟偏差。

最初的研究假定端系统间时钟频率的偏差在整个测量周期内都是固定的,同时任何端系统的时钟不存在时钟值的跳变和频率漂移(Drift)。基于这种严格的假设,并在假定端系统间的前向(Forward)与反向(Reverse)路径的单向延迟相等的情况下,Paxson 引入了一种强线性拟合算法(Robust line-fitting algorithm)来消除时钟动态性[32]。Moon 摒弃了前向与反向单向延迟相等的假设,利用单向延迟序列,将时钟偏差消除问题转化为线性规划问题,提出了利用线性规划算法(Linear Programming Algorithm,LPA)实现端系统间的时钟同步算法[33]。

观察发现,端系统时钟的实际行为比理论上的假设更为复杂。如端系统的时钟频率会随环境温度、压力等因素的变化而发生变化,时钟值本身也会因外界干预而发生突变[32],Li Zhang 研究了当端系统中出现时钟复位时,基于凸包(Convex Hull based CH)进行端系统间的相对同步算法,其算法复杂度为 $O(Nw)+O((N/R)^{R}NR)$[34]。在 CH 算法中,引入了额外的假设条件即时钟复位发生的频率较低,发生时钟复位前后端到端的时钟频率偏差保持不变。文[35]在对端系统间时钟相对变化建模的基础上,将时钟动态性分成时钟值突变与相对频率突变(假设不存在频率漂移)两种情况,利用对单向延迟时间序列的分段(聚类分析的一种特例),检测出测量过程中的时钟动态性,最后在每个时钟稳定区间,使用 LPA 算法消除时钟偏差,其复杂度为 $O(KN^2)$。由于聚类算法所固有的鲁棒性,基于时间序列分段方法不需对测量环境做其它假设,可以有效实现端系统间相对时钟同步。

6.3 基于流的数据压缩方法

采用被动测量方法,捕获经过测量点的包头数据是进行网络行为与演化分析的重要手段。网络带宽的增加及高带宽需求的实时业务的出现,被动测量面临高带宽、大规模、实时分析、提供各种粒度指标的挑战。传统的基于数据包捕获、存储、分析的方式已经不能满足高速链路下的要求,主要表现在:

·PCI 总线的限制。数据包的捕获、存储需经过 PCI 总线传输2次,对于 OC48/2.5G,OC196/10G 以上链路,需新的总线技术才能满足对总线带宽的要求;

·存储设备的容量限制和海量数据管理也是高速网络性能测量所面对的问题;

·内存带宽与访问速度的限制;

·I/O 和 CPU 速度的增长远远慢于链路速率的增长。

因此,在当前硬件体系结构下,进行高速网络的被动测量所需解决的问题主要有:

·如何在总线与内存带宽、速度受限情况下进行高速流量的捕获;

·如何在保证获得完整流量详细信息的同时传输与存储尽可能少的信息。

基于流的流量捕获是解决高速链路数据采集的可行方案之一。Iannaccone 将流定义为一五元组,即:传输层协议类型、源 IP 地址、目标 IP 地址、源端口号、目标端口号[36]。基于流的捕获,将传统基于数据包(Packet-based)捕获的每个包头中含完整且冗余的信息分成共享的流标志信息与每个数据包所含特定信息。由于消除了数据包头中的冗余信息,大大减少了所需传输与存储的数据量,实现了网络流量数据信息的无损压缩。

对通过测量点的流量进行基于流的捕获时,系统所维护流的个数、每个流所持续的时间受到系统内存空间的限制。而减少流的个数与持续时间,不仅可能会将本属同一流的数据包分成多个流,而且还带来冗余信息的存储。在基于流的方法中,流结束的判定是研究的重点。

网络中的60%~90%的负载是 TCP 的流[37],TCP 流有明确的结束标识(TCP 流中 FIN 或 RST 标志),可用于确定流的结束信息。这种基于协议的流结束判定方法,实现简单,但是网络环境的不确定性即使具有流结束标志,重传的数据包可能会在流结束标志后到达,从而造成流分裂与不完整流的出现。

Claffy 对 NSFNET 上采集的数据包利用事后分析法(Post facto),研究了不同协议下流的大小、长度分布后,首先提出了采用固定超时时间(60s~120s)来作为流的截止时间,即在一定的时间间隔内收不到属于该流的数据包,则认为流已经结束[38]。引入时间间隔确定判断流的结束,解决了非 TCP 流的结束位的判定问题,但固定的时间间隔会存在以下两个问题:

·当超时的时间过长,保存的流数量过大,需消耗更多的系统资源。

·过短的时间间隔使系统不断的销毁旧的流和创建新的流,造成系统的抖动。其次,过短的时间间隔将大的流分割成多个较小的流,一方面造成数据存储与传输的压力,且不利于对网络的持续流量进行有效分析。

Ryu 在对大量网络流量记录(Trace)分析后,设计了一种具有自适应超时时间隔的流结束确定算法——二进指数超时间隔(Measurement-based Binary Exponential Timeout,MBET)算法[39]。MBET 根据一个流中数据包的到达情况,对整个流的当前行为进行预测,从一系列离散化的固定时间间隔中确定新的流结束的超时时间。由于考虑到流的行为特性,自适应超时算法性能与固定超时时间算法相比在对资源的消耗上有明显改善。

在 MBET 算法中,一个流超时时间间隔的变化总体上是呈递减趋势的,其起始时间间隔设为最大值,一旦超时间隔减小后,不能再调整为较大值。这导致对数据量较大、持续时间长的流被分成多个流进行处理。由于一个流通常代表用户一次会话过程(Session),不同的应用在流量、持续时间、行为等上均有不同的特征,结合应用层协议特征来确定自适应超时间隔应是研究方向之一。

6.4 系统部署策略(Deployment policies)

测量系统通常采用分布测量、集中控制的体系结构,系统由位于测量点的测量任务执行部分(Probes)与存储控制部分组成。在主动或被动测量中,各个测量点所获得的性能相关数据均需要传送到控制系统进行数据存储与性能分析。性能数据特别是被动测量下捕获的大量数据包头信息在传输过程中,需消耗网络的带宽、处理器与存储等资源。系统部署策略需要解决二个方面的问题:对于被测网络系统,如何分布测量点,实现对网络有效测量;如何确定数据存储的位置,实现最少性能数据的传输。

测量点的数量,分布位置是与测量任务相关的。测量点的部署策略是在尽可能少的测量点下,实现对指标的测量。在被动测量中,Breitbart 假设网元设备满足流守恒约束条件下,将对网络流量的测量、测量点数量及部署位置形式化为求解图 $G=(V,E)$ 上的最小弱顶点覆盖问题[40]。而在主动端到端测量中,由于利用端到端的性能测量进行网络断层(Network Tomography)分析已成为网络性能测量中的又一研究热点[41]。因此,主动测量点的部署不仅需完成端到端性能指标的测量,还应具有对关键链路相关性能指标推断的能力,如 Horton 研究了在具有 n 个节点拓扑未知的网络中,进行正确拓扑推测所需的测量点数目上下界为 $(n+1)/3$ 和 $(n-1)/3$,以及测量点的部署位置策略[42]。

存储系统的部署策略可形式化为由多个测量点所构成网络拓扑下的中心布局问题(Center Placement Problem,CPP)[43],因而,可借鉴网络中镜像服务器[44]或 Web 复制服务器的部署算法来确定存储系统位置[45]。

7 进一步的研究

网络性能测量因需求的驱动从方法简单、指标单一的测量向大规模、多指标协同测量方向发展。其趋势主要表现在:

·多种测量方法相结合;

·对指标不仅进行短期的测量,还对指标的中、长期特性进行统计测量;

·由于骨干链路的速率日益增加,大规模测量系统主要使用端到端的测量方法,测量点通常部署在被测网络的边缘;

·利用端到端的性能测量进行网络内链路级性能推断的需求日益凸现;

·为基于网络的各种应用提供参数化的网络性能测量服务。网络性能测量服务成为网络基础设施的必要组成部分。

因此,在系统级的网络性能测量中,进一步的研究工作主要有:

·高速网络中性能指标数据的压缩存储研究。基于流的数据捕获是一种无损、高效的数据压缩存储方法,适用于被动网络性能指标的测量。对于主动测量,根据测量数据内在的相关性,研究精度可控的有损数据压缩是可行的[30];

·数据分析方法上的研究。对于海量的网络性能数据,研究从中挖掘反映网络性能及动态行为的信息是大规模网络性能评价的一项重要课题;

·网络性能断层(Tomography)分析技术。端到端测量提供了对路径级、面向用户的网络性能指标的测量,但核心网或骨干网性能不能得到有效的测量。研究表明,通过端到端的测量来推测网络内部的性能是可能的[46]。当前所采用的方法主要是在进行组播(Multicast)或结合包对技术(Packet pair)进行端到端的测量后,然后基于极大似然估计(Maximum Likelihood Estimation MLE)或期望最大化(Expectation Maximization EM)算法对具有加型(如链路延迟[47~49])和乘积型(如链路包丢失[50])指标进行测量。对凹型指标(如带宽)的推断研究较少。如何对凹型指标建模,实现有效的推断是网络断层分析中的一个难点。网络性能的断层分析技术本质上是信号盲分离(Blind Separation BS)的一个特例。结合 BS 中的研究成果将有助于对网络性能推断的研究。

·网络性能测量服务基础设施的应用研究。基于网络应用的发展对网络性能服务提供的方式提出了严峻的挑战,网络性能测量服务逐渐成为网络基础设施的必要组成部分。如何存储、组织和访问性能服务基础设施是提供服务的前提条件。如在网格计算环境下,由于地理上的分散性,大规模的资源共享与协同特性,为确保网格应用的性能,全球网格论坛(Global Grid Forum,GGF)在 IPPM 定义的网络性能指标基础上提出了网格监控体系结构(Grid Monitoring Architecture,GMA)[51],并对网络性能监控的实施、性能结果的发布、和存取控制原则进行了规范[52,53]。文[54]结合主动与被动测量方法,对基于网格应用的网络性能测量与结果存储进行了初步研究。

结论 网络测量技术经过数十年的研究,其一系列的成果对网络自身的发展、网络设备的研发等方面具有重要的指导意义。在国际上,政府、企业、大学和研究机构分别对网络性能测量的各个方面做了深入的研究。我国在该领域的研究起步较晚,当前主要还处于跟踪研究状态。我国在网络基础设施的建设方面取得了重要成就,研究网络测量技术,准确理解网络与用户行为特性,优化网络资源配置、提高服务水平是我们当前网络建设所面临的重要、迫切的课题。

致谢 解放军理工大学计算机系谢希仁教授为本文提供了部分有价值的参考资料,在此深表感谢。

参考文献

1 中国互联网络信息中心. 中国互联网络发展状况统计报告(2003年7月). http://www.cnnic.cn/develst/2003-7

2 Paxson V, Floyd S. Wide-area Traffic: The Failure of Poisson Modeling. IEEE/ACM Transactions on Networking, June 1995. 226~244

3 Leland W, Taqqu M, Willinger W, et al. On the Self-similar Nature of Ethernet Traffic (extended version). IEEE/ACM Transactions on Networking, 1994. 1~15

4 Paxson V, Almes G, Mahdavi J, et al. Framework for IP Performance Metrics. IETF RFC 2330, May 1998

5 Mahdavi J, Paxsion V. IPPM Metrics for Measuring Connectivity. IETF RFC 2678, Sep. 1999

6 Almes G, Kalidindi S, Zekauskas M. A One-way Delay Metric for IPPM. IETF RFC 2679, Sep. 1999

7 Almes G, Kalidindi S, Zekauskas M. A One-way Packet Loss Metric for IPPM. IETF RFC 2680, Sep. 1999

8 Koodli R, Ravikanth R. One-way Loss Pattern Sample Metrics. I-

ETF RFC 3357, Aug. 2002
9 Demichelis C, Chimento P. IP Packet Delay Variation Metric for IP Performance Metrics (IPPM). IETF RFC 3393, Nov. 2002
10 Almes G, Kalidinidi S, Zekauskas M. A Round-trip Delay Metric for IPPM. IETF RFC 2681, Sep. 1999
11 黎文伟,王俊峰,谢高岗,等. 基于包对采样的IP网络时延变化测量方法. 计算机研究与发展,已录用
12 Mathis M, Allman M. A Framework for Defining Empirical Bulk Transfer Capacity Metrics. IETF RFC 3148, July 2001
13 Lai K, Baker M. Measuring Bandwidth. In: Proc. of IEEE INFOCOM'99, New York, March 1999
14 Hou Y T, Dong Yingfei, Zhang Zhili. Network Performance Measurement and Analysis Part 1: A Server-Based Measurement Infrastructure (Concept Paper): [Technical Report FLANCRTM98-01]. Fujitsu Laboratories of American, July 1998
15 Home page of NAI. http://moat.nlanr.net, Oct. 2003
16 Apisdorf J, Claffy K, Thompson K, et al. OC3MON: flexible, affordable, high performance statistics collection. In: Proc. of INET'97
17 Paxson V, Mahdavi J, Adams A, et al. An Architecture for Large-Scale Internet Measurement. IEEE Communications. 1998, 36 (8): 48～54
18 Homepage of CAIDA. http://www.caida.org, Oct. 2003
19 Keys K, Moore D, Koga R, et al. The Architecture of CoralReef: an Internet Traffic Monitoring Software Suite. In: Proc. of Passive and Active Measurements 2001 (PAM 2001), 2001
20 CAIDA Tools. http://www.caida.org/tools, Oct. 2003
21 Homepage of Surveyor. http://www.advanced.org/surveyor, Oct. 2003
22 Homepage of IEPM. http://www-iepm.slac.stanford.edu/, Oct. 2003
23 Homepage of IPMON. http://ipmon.sprintlabs.com/ipmon.php, Oct. 2003
24 Weigle E, Feng W-c. TICKETing High Speed Traffic with Commodity Hardware and Software. In: Proc. of Passive and Active Measurement Workshop 2002 (PAM 2002). Fort Collins, Colorado, USA. 2002
25 Pásztor A, Veitch D. PC Based Precision Timing Without GPS. In: Proc. of Intl. Conf. on Measurements and Modeling of Computer Systems (ACM SIGMETRICS 2002), Marina Del Rey, California, USA, June 2002. 1～10
26 Zseby T. Stratification Strategies for Sampling-based Non-intrusive Measurements of One-way Delay. In: Proc. of Passive and Active Measurement Workshop 2003 (PAM 2003), San Diego, CA, USA. April 2003
27 Duffield N, Lund C, Thorup M. Charging from Sampled Network Usage. In: Proc. of Internet Measurement Workshop, San Diego, CA, USA. Nov. 2001
28 Claffy K C, Polyzos G C, Braun H-W. Application of Sampling Methodologies to Network Traffic Characterization. In: Proc. of the ACM SIGCOMM'93, San Francisco, CA, USA. Sep. 1993. 194～203
29 Zseby T. Deployment of Sampling Methods for SLA Validation with Non-Intrusive Measurements. In: Proc. of the Passive and Active Measurements Workshop 2002 (PAM 2002). Fort Collins, Colorado, USA. 2002
30 Wang JF, Yang JH, Zhou HX, et al. Adaptive Sampling Methodology in Network Measurements. Journal of Software, Accepted for publication
31 Mills D. Network Time Protocol (version 3): Specification, implementation and analysis. IETF RFC 1305. March 1992
32 Paxson V. On calibrating measurements of packet transit times. In: Proc. of SIGMETRICS. June 1998
33 Moon S B. Measurement and Analysis of End-to-end Delay and Loss in the Internet: [PH. D dissertation]. University of Massachusetts, Amherst. Jan. 2000
34 Zhang L, Liu Z, Xia C H. Clock Synchronization Algorithm for Network Measurements. In: Proc. of IEEE INFOCOM 2002. June 2002
35 Wang JF, Yang JH, Zhou HX, et al. Detecting Clock Dynamics in One-way Delay Measurement. Journal of Software, Accepted for publication
36 Iannaccone G, et al. Monitoring very high speed links. In: Proc. of ACM SIGCOMM Internet Measurement Workshop 2001, San Francisco, Nov. 2001. 267～271
37 Fomenkov M, Keys K, Moore D, et al. Longitudinal study of Internet traffic in 1998-2003. Cooperative Association for Internet Data Analysis (CAIDA), 2003
38 Claffy K C, et al. A parameterizable methodology for Internet traffic flow profiling. IEEE JSAC, 1997
39 Ryu B, Cheney D, Braun H-W. Internet Flow Characterization: Adaptive Timeout Strategy and Statistical Modeling. In: Proc. of Passive and Active Measurement Workshop, Amsterdam. April 2001
40 Breitbart Y, et al. Efficiently Monitoring Bandwidth and Latency in IP Networks. In: Proc. of IEEE INFOCOM 2001, Anchorage, Alaska, USA, April 2001. 933～942
41 Chen Y, Bindel D, Katz R H. Tomograpy-based Overlay Network Monitoring. In: Proc. of ACM SIGCOMM Internet Measurement Conference 2003, Miami Beach, Florida, USA, Oct. 2003
42 Horton J D, Lopez-Ortiz A. On the Number of Distributed Measurement Points for Network Tomography. In: Proc. of ACM SIGCOMM Internet Measurement Conf. 2003, Miami Beach, Florida, USA, Oct. 2003
43 Charikar M, Guha S, Tardos E, et al. A Constant-Factor Approximation Algorithm for the k-Median Problem (Extended Abstract). In: Proc. of 31st Annual ACM Symposium on Theory of Computing, 1999. 1～10
44 Jamin S, Jin C, Jin Y. On the Placement of Internet Instrumentation. In: Proc. of IEEE INFOCOM 2000, March 2000
45 Qiu L, Padmanabhan V N, Voelker G M. One the Placement of Web Server Replicas. In: Proc. of IEEE INFOCOM 2001, Anchorage, Alaska, USA, April 2001. 1587～1596
46 Vardi Y. Network Tomography: Estimating Source-Destination Traffic Intnsities From Link Data. Journal of the American Statistical Association, March 1996. 365～377
47 Xia Y, Tse D. Inference of Link Delay through Measurement Redundancy in Communication Networks: [Technical Report, Memorandum UCB/ERL M00/57]. University of California, Berkeley. Nov. 2000
48 Duffield N G, Horowitz J, Presti F L, et al. Network Delay Tomography from End-to-end Unicast Measurements. In: Proc. of 2001 Intl. Workshop on Digital Communication 2001, Taormina, Italy. Sep. 2001
49 Coates M J, Nowak R. Network Delay Distribution Inference from End-to-end Unicast Measurement. In: Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing. May 2001
50 Coates M J, Nowak R. Network Loss Inference using Unicast End-to-End Measurement. In: Proc. of ITC Conf. on IP Traffic, Modeling and Management, Monterey, CA. Sep. 2000
51 Tierney B, Aydt R, Gunter D, et al. A Grid Monitoring Architecture: [Technical Report GWD-PERF-16-2]. GGF Performance Work Group. Jan. 2002
52 Lowekamp B, Tierney B, Cottrell L, et al. A Hierarchy of Network Performance Characteristics for Grid Applications and Services. Global Grid Forum Network Measurements Working Group. June 2003
53 Schema/Profile for Network Performance Measurements for Grids. Global Grid Forum Network Measurements Working Group working document. http://www-didc.lbl.gov/NMWG/
54 Wang Junfeng, Zhou Mingtian. Providing Network Monitoring Service for Grid Computing. In: 2nd Intl. Workshop on Grid and Cooperative Computing