

网络资源及服务的监测、分析与评估集成技术*

□ 乔晓东 李鹏 梁健 张寅生 / 中国科学技术信息研究所 北京 100038

摘要: 本文分析了网络资源监测分析评估技术当前的局限性,即只能完成TCP/IP协议栈部分协议事件、Web服务的不完整过程和网络资源部分要素的监测,为此建立了包括网络实体、用户和信息资源等多要素的网络资源(一体化)模型,并明确了各要素间的服务关系及其与TCP/IP协议的关系。在此模型基础上,构建了一个针对网络资源通用模型对象(典型网站)的监测、分析、评估的指标体系。为实现这些指标的监测、分析和评估,设计了网络(一体化)资源监测、分析、评估系统的体系结构和技术集成方法,从而实现了网络资源完整服务过程和多种资源要素的监测、分析和评估。

关键词: 网络资源模型, 网络资源监测, 网络资源监测指标, 网络资源监测技术集成, 网站监测方法, Web服务监测, 网络行为分析, 信息资源评估, 数字图书馆

DOI: 10.3772/j.issn.1673-2286.2008.11.010

1 网络资源(一体化)模型

作为术语的“网络资源”,其意义显而易见是明确的,它应该包括Internet中的网络实体,即网络协议对象的集合(物理实体),如交换机、路由器、服务器,还应包括以下资源:网络实体各种设备存储和处理的信息资源,以及它们的用户。

在学科、研究领域划分上,网络资源的物理实体模型在网络管理研究领域有大量研究,例如基于OSI协议栈对应的网络实体抽象模型。网络环境下的信息资源在信息管理研究领域有大量研究。^[1]在这些不同的研究领域中,网络物理实体的模型和网络环境下的信息资源模型都被大量地建立,但是,跨领域的——即既表示网络物理实体又表示网络环境下的信息资源的——网络资源一体化模型尚没有建立。这里所说的模型应包括资源的类别以及对资源进行操作的逻辑关系和相关的协议特征。

显然,网络资源模型的建立是资源理论研究及其相关应用开发的基础。但是,尚没有检索到包括上述意义的网络资源模型。本文提出网络资源一体化模型如下。

在这个模型中,资源分为4种,访问者(用户)、表示层数据资源(XML)、应用层数据资源、应用层以下网络

| |
|--------------|
| 访问者(用户) |
| 表示层数据资源(XML) |
| 应用层数据资源 |
| 应用层以下网络协议对象 |

图1 网络资源模型

协议对象。层次自下而上表示相关的协议间的逻辑关系及其资源之间的服务关系。

2 网络资源和网络服务监测技术发展概要及其与用户需求的差距

对网络资源及其服务进行监测、分析和评估是对网络资源及其相关服务的一项基本需求。本文所理解的监测基本上是对原始数据和基本信息特别是客观性信息的获取,分析主要指对于获取的监测数据进一步处理使获得者对被监测对象进行的相关判断(因果性、本质意义等),评价指对于被监测对象的价值(合目的性)以及相同相似对象的对比性进行的判断。

与网络资源模型相对应,实际上对网络资源及其相应的网络服务的监测往往被分割在网络(实

* 本文是国家科技基础条件平台应用服务支撑系统“网络科技资源监测、分析评估体系建设”项目2005DKA53700的研究成果之一。

体) 监测和网络信息资源监测两个研究领域。对于网络监测技术, 往往不考虑网络物理实体所存储的信息的总体特性, 只考虑网络物理实体的状态信息; 网络信息资源监测技术, 往往只关注静态信息的特性而不考虑网络实体的状态信息。这样, 网络资源和网络服务监测技术实际上需要考虑两个技术领域的发展概况。

对网络物理实体状态信息的监测, 实际上就

是网络管理技术中的重要内容, 即通常所说的“网络监测”。换言之, 当前的“网络监测”基本上指“网络物理实体的状态信息的监测”, 而非对各种网络资源(包括自然语言信息、用户)的监测。

当前网络监测技术实现了对图1(网络资源模型)中不同层次的网络资源的监测。在这些不同的技术中, 有许多技术是比较成熟的。这些不同的技术及其功能列于表1。

表1 当前网络监测主要技术及其功能

| 监测对象 | Windows 网络监视器 | HP OVO ^[2] | IBMITCAM ^[3] | SNMP ^[4,7] | Netflow ^[9] | LibPcap ^[5] | Big Brother ^[6] | MRTG ^[6,7,8] |
|--------------------|---------------|-----------------------|-------------------------|-----------------------|------------------------|------------------------|----------------------------|-------------------------|
| Internet Browser | | | | | | | | |
| XML文件 | | | | | | | | |
| Data Base | | | | | | | | |
| JDBC/ODBC | | √ | √ | | | | | |
| TELNET, HTTP, FTP等 | | | | | | | ▽ | |
| TCP/UDP 对象和事件 | | | | √ | | ▽ | | ▽ |
| IP对象和事件 | | | | ▽ | √ | | | ▽ |
| 链路层协议对象和事件 | ▽ | ▽ | ▽ | | | | | |
| 主机: 系统 | √ | √ | √ | | | | √ | |
| 主机: CPU系统 | √ | √ | √ | | | | √ | |
| 主机: 磁盘 | √ | √ | √ | | | | √ | |
| 主机: 内存 | √ | √ | √ | | | | √ | |

√表示能够实现主要功能; ▽表示能够实现部分功能

对照表1和图1可以看出, 当前的网络管理和网络监测技术主要局限于图1所示的应用层协议以下网络实体运行状态的监测, 对于用户(浏览者)行为以及静态(以自然语言为主的)信息资源的监测功能几乎是没有的。

对于网络(以自然语言为主的)信息资源监测目前尚没有明确专门术语和专门实现技术。但是对于网络信息资源监测的需求是大量存在的, 特别是动态地掌握某一网络环境下特定信息资源的不同情况对于某些用户仍然是非常必要的。满足这种需求可能和当前的网络信息资源的发现、整合技术密切相关。信息资源的发现、整合技术具有专门的论述, 在此不再综述。

显然, 对这些综合情况的监测、分析和评估需求的确是非常常见和合理的, 即因特网用户不只需要掌握某个网络服务系统的环境(网络实体及其运行), 也要掌握网络服务系统环境下的信息资源及

其客户的主要特征和重要动态变化情况。对于网络资源(一体化资源)的综合监测、分析和评估, 上述当前的网络监测技术和信息资源的发现、整合技术单独方面都难以实现。特别是用户最需要了解的某些指标例如资源受欢迎的程度、资源的易用性, 往往是难以在逻辑或语义上区分是网络环境、信息和用户的单方面特性, 也不只是TCP/IP协议栈局部层次所规定的对象的行为, 而是TCP/IP协议栈所有协议层次共同执行的结果。

因此, 对于网络资源的监测应该综合考虑网络实体监测及其所负载的信息资源及其用户的监测、分析和评估, 需要对资源不同方面的相关技术进行集成。

3 面向web服务的网络资源的通用评价指标

由于网络资源在实际上包含网络实体和负载的

信息资源,因此其网络资源评价指标体系应该考虑两个方面的特征。

对网络实体的技术描述指标当前已经是比较成熟的。

国际电信联盟(International Telecommunication Union, ITU)和中华人民共和国信息产业部分别制订了一系列国际和中国关于TCP层和IP层服务对象性能和技术特征描述的标准。这些标准包括了一些网络基本技术、性能的评价指标。其中,《IP数据通信业务—IP数据包传输和可用性性能参数》(ITU-T Y.1540)^[10]和《互联网协议通信业务——IP数据包传输及可用性目标和分配》(ITU-T Y.1541)^[11]和《IP网络技术要求—网络性能参数与指标》(YD/T 1171-2001)^[12]是比较重要的网络层及其相邻层技术性能描述和评价性指标体系。对于Web服务高层协议以及应用层之上的业务目前形成了SOAP (Version 1.2)一系列技术规范。^[13]由于web服务功能过于复杂,其功能性的评价指标尚未形成通用的体系,对于非功能性的评价指标已经进行了一定研究。^[14]概要而言,这些指标主要关注网络及其服务的速度、可靠性、安全性。

关于网络信息资源的评价指标体系目前国内外形成了各种观点和方案。^[15]国外比较有代表性的如Betsy Rechmond提出了网络信息资源的评价指标的原则,包括:内容、可信度、范围、连续性、批判性思考等10项。Wilkinson等提出了网络信息资源的评价指标体系,包括11个门类(125个指标)。11个门类是:可检索性和可用性,信息资源的识别和验证,作者身份的鉴别,作者的权威性,信息结构与内容,信息内容相关性和范围,内容的正确性,内容的准确性与公正性,导航系统,链接质量,美观与效果。此外,Robert Harris, David Stocker 和Alison Cooke, Jim Kapoun, Alistair G Smith, John R Henderson, LII(Librarians' Index to Internet), IPL(Internet Public Library), OPLIN(Ohio Public Library Information Network)也提出了不同的网络信息资源的评价指标相关的标准或方法。国内的蒋颖较早提出了网络信息资源的评价标准,即:信息质量、范围、易用性、稳定性,此外,左艺、魏良、赵玉虹,黄奇、郭晓苗,罗春荣、曹树金,栗慧,田菁等分别提出了网络信息资源的评价指标标准。

概括而言,网络信息资源的评价指标可能因为资源的类别和环境不同而存在差异。需要进一步提出更为通用型的描述或评价指标。

为此,本文提出基于web服务的网络资源的通用评价指标。这些指标旨在描述和评价以下对象:以Web服务为主的网络资源,资源包括网络实体(作为环境),也包括它所负载的信息资源和客户。

指标分为4类,即:I.用户行为或特征, II.Web服务特征, III.以自然语言为主的信息的特征, IV.应用层协议以下的网络对象性能特征。这些类别与图1:网络资源模型的层次是基本对应的。除所描述的基本方面大致是交互性、类别、质量、数量、速度。每类的指标可以按用户要求扩展、收缩或更改。见表2。

4 基于Web服务的网络资源监测、分析、评估系统的体系结构

为获得表2指标的数据,本文作者及其合作单位已经开发了实现表2指标的应用系统——基于Web服务的网络资源监测、分析、评估系统。系统的基本特征如下。

- (1)能够实现图1(网络资源模型)所有类别资源和表2所有指标的监测分析和评估;
- (2)对于指标,系统具有可扩展性,允许在本类别的指标中进行再开发以扩展新的指标;
- (3)能够对同类监测对象进行表2指标的对比分析和评估;
- (4)监测、分析和评估的数据能够进行统一展示和登陆认证;
- (5)监测、分析和评估的结果能够可视化展示;
- (6)可以远程浏览可视化展示。
- (7)系统各个服务的用户可以统一身份认证。

为此,所设计的系统体系结构见图2。

总体上,系统由被监测端(A端)和监测端(监测中心,B端)构成。

在物理结构上,由4个在不同物理地址运行的部分构成,即:①被监测端服务器上运行的程序(含模块A1~A3);②被监测端服务器网络出口的网络设备上运行的程序(模块A4);③浏览者计算机上

表2 基于Web服务的网络资源监测、分析、评估的主要指标

| | I. 用户行为或特征 | II. Web服务特征 | III. 以自然语言为主的 信息的特征 | IV. 应用层协议以下的 网络对象性能特征 | |
|----|----------------|-------------------|------------------------|---------------------------------|----------------------|
| 1 | 用户数 | 网站管理信息 | 数据库管理信息 | 主机 特征 及 运 行 | 服务器管理信息 (管理员、用户等) |
| 2 | 按业务(或资源)统计的用户数 | 网站无故障运行时间 | 信息资源类别 | | 服务器基本参数 |
| 3 | 按IP地址统计的用户数 | 访问响应时间 | 信息总量(数据库、网页数) | | 服务器无故障运行时间 |
| 4 | 按国家和地区统计的用户数 | 网站访问量 | 不同类别的信息总量 | | 操作系统基本信息 |
| 5 | | 网站访问量变化率 | 不同类别的信息资源利用程度 | | 操作系统无故障运行时间 |
| 6 | | (各种)网络服务提供量 | 数据库系统无故障运行时间 | 网 层 资 源 和 事 件 | 网络设备信息 |
| 7 | | 最受欢迎的服务 | | | 网络设备管理信息 |
| 8 | | 重要网页访问量 | | | 包传输时延(包速率) |
| 9 | | 按用户IP地址分布的重要网页访问量 | | | IP包吞吐量 |
| 10 | | 按国家和地区统计的重要网页访问量 | | | 网络端到端丢包率 |
| 11 | | 最受欢迎的网页 | | | |

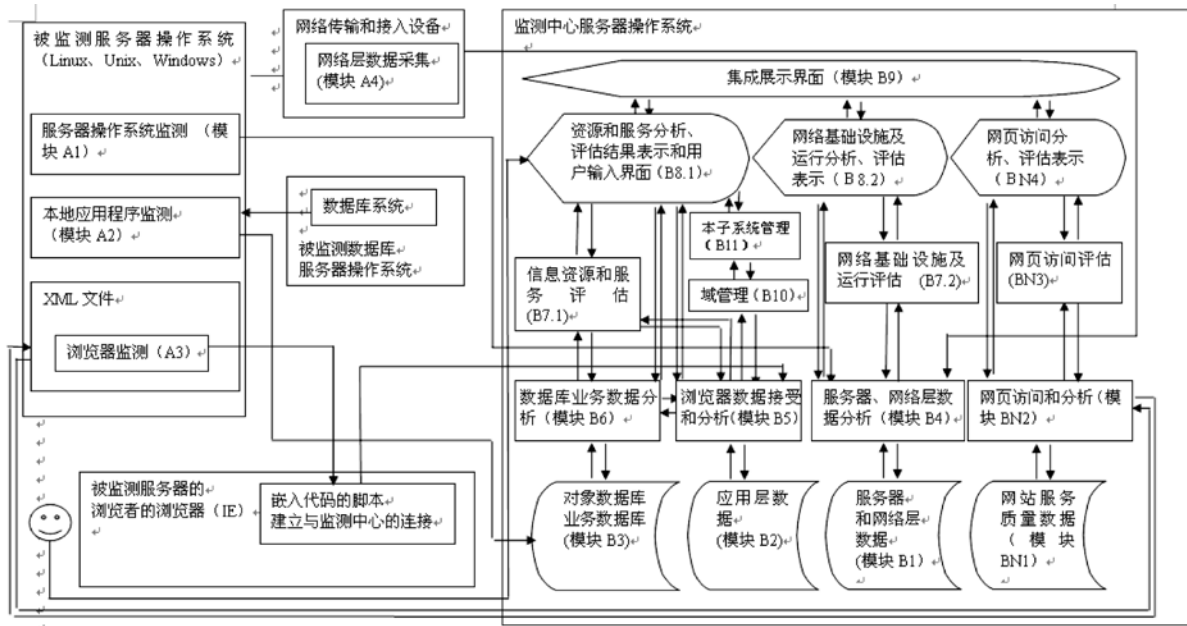


图2 基于Web服务的网络资源的监测分析评估系统体系结构

运行的程序（由A3自动加载到该被监测网站的用户浏览器的浏览器上）；④监测中心端运行的程序（含模块B1~B9, BN1-BN4）。

在服务器端运行的程序分为4个层次,即: ①数据存储层 (B1~ B3,BN1): 存储采集的数据; ②分析层 (B4~ B6,BN2): 输入存储采集的数据, 输出监测指标; ③评估层 (B7.1~ B7.2,BN3): 输入监测指标, 输出监测结果 (知识); ④指标和知识 (分析和评估结果) 表示层 (B8.1~ B8.2,BN4,B9): 输入监测指标和评估知识, 输出监测指标和评估知识的图形表示。

在功能划分上, 本系统由4个流程上连贯的子系统。

子系统1——网络基础设施及运行监测分析评估子系统数。

该子系统在被监测网络的接入设备 (交换机和路由器) 入口嵌入设备 (模块A4), 以模拟链接、截获数据等方法采集网络层数据。采集的数据传入监测中心的服务器、网络层数据分析 (模块B4), 对于多个网站的同类指标的对比分析构成网络基础设施及运行评估 (模块B7.2), 分析和评估结果均输入可视化工具展示 (模块B8.2), 也能够集成环

境下展示(模块B9)。子系统1完成IV.应用层协议以下的网络对象性能特征的所有指标。

子系统2——数据库业务监测分析评估子系统。该子系统在被监测网站的业务数据库运行环境下,在获得数据库访问授权的情况下以SQL语言程序(模块A2)获得数据库指标相关的数据。采集到的数据发送到监测中心对象数据库业务数据库(模块B3),经过数据库业务数据分析(模块B6)分析和多网站同类指标的对比分析(评估,模块B7.1)输入给资源和服务分析、评估结果表示和用户输入界面(B8.1),也能够在集成环境下展示(模块B9)。

子系统2主要完成III.以自然语言为主的信息的特征指标,其中,信息总量指标只完成数据库部分。

子系统3——用户行为分析子系统。子系统3采集数据的方法是脚本语言代码嵌入被监测网站中所选择的网页上(浏览器监测模块A3),当该网页的浏览者访问网页时运行此程序将浏览行为信息发到监测中心的浏览器数据接受和分析模块(B5),当多个网站同类指标进行分析时由信息资源和服务评估(模块B7.1)完成。分析(模块B5)和评估(模块B7.1)结果都由资源和服务分析、评估结果表示

和用户输入界面(模块B8.1)展示,并可由集成展示界面(模块B9)展示。对于代码的嵌入管理由两级管理员管理(管理本被监测网站和多个被监测网站),管理模块分别是域管理(模块B10)和本子系统管理(模块B11)。子系统3完成I所有指标和II.Web服务特征指标中的II.4-11。

子系统4——网页访问分析子系统从监测中心访问被监测网站以获得数据。访问、分析、评估、展示、集成展示分别模块BN2~BN4, B9。子系统4完成II.Web服务特征指标中的1~3,以及III3中的网页数指标。

这样,基本完成了用户对于典型的web服务一体化资源的主要特征、性能的监测、分析和评估。显示界面用ChartDirector图形插件实现,见图3。

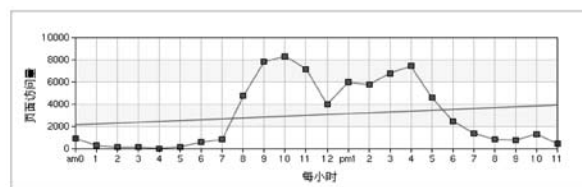


图3 监测指标显示结果界面

参考文献

- [1] 谷琦主编. 网络信息资源组织管理与利用[M]. 北京, 科学出版社, 2008:1-5.
- [2] [2008-8-14]. <http://openview.hp.com/products/ovowin/index.html>.
- [3] [2008-8-14]. <http://www-306.ibm.com/software/tivoli/products/composite-application-mgr-soa>.
- [4] [2008-8-14]. <http://www.cisco.com/en/US/docs/internetworking/technology/handbook/SNMP.html>.
- [5] [2008-8-14]. <http://sourceforge.net/projects/libpcap/>.
- [6] 余卫华. 基于开源软件的网络监测系统[J]. 微计算机信息, 2007(4): 93-94, 141.
- [7] 吴鸣虎. 网络监测和管理技术平台的设计与初步构建[J]. 医学信息, 2006(1): 6-9.
- [8] [2008-8-14]. <http://www.mrtg.com>.
- [9] [2008-8-14]. <http://www.cisco.com/web/go/netflow>.
- [10] 《IP数据通信业务—IP数据包传输和可用性性能参数》, ITU-T Y.1540.
- [11] 《互联网协议通信业务—IP数据包传输及可用性目标和分配》, ITU-T Y.1541.
- [12] 《IP网络技术要求—网络性能参数与指标》, YD/T 1171-2001.
- [13] [2008-8-14]. SOAP (Version 1.2) <http://www.w3.org/2001/09/soap-envelope>.
- [14] 熊鹏程, 范玉顺. 一种自动选择网络服务的UDDI扩展模型[EB/OL]. [2008-8-14] <http://www.simflow.net/Publications/Papers/Year2006/xpczydh-0607>.
- [15] 谷琦主编. 网络信息资源组织管理与利用[M]. 北京, 科学出版社, 2008: 97-100.

作者简介

乔晓东 (1965-), 英国谢菲尔德大学硕士, 研究员, 中国科学

技术信息研究所信息技术支持中心主任, 研究方向为信息资源管理, 发表论文数十篇。通讯地址: 中国科学技术信息研究所信息技术支持中心 100038

李鹏 (1979-), 梁健 (1981-), 张寅生 (1962-) 均为中国科学技术信息研究所信息技术支持中心研究和开发人员。研究方向: 智能信息处理 (李), 网络应用技术开发和自然语言理解 (梁)、人工智能 (张)。通讯地址: 同上

The Integrated Technologies Monitoring, Analyzing and Assessing Network Resources and Services

Qiao Xiaodong, Li Peng, Liang Jian, Zhang Yinsheng / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: This paper analyses the current technical limitations in monitoring, analyzing and assessing network resources. In the author's opinion, the limitations consist in that only partial protocol events in TCP/IP stack, incomplete course of web services and partial elements of network resources can be monitored. So a new model of network resources is established which includes multi-elements of network resources such as network entities, users and information, and the relations of the resources elements in TCP/IP layers. Based on the model, the network resources monitoring indicators are given. To fulfill these indicators for monitoring, analysis and assessment, an application system are designed, which integrates the relevant technologies realizing monitoring, analysis and assessment in the complete service processes and a variety of resources elements.

Keywords: Network resources, Network resources model, Network resources monitoring, Network resources monitoring indicators, Network resources integration, Website monitoring technology, Web service monitoring, Network behavior analysis, Information resources assessment, Digital library

(收稿时间: 2008-8-29; 责任编辑: 虞敏)