

Web 计算: 针对 Web 资源的网络计算

张昌利 吴 健 胡正国

(西北工业大学计算机学院 西安 710072)

摘要 随着 Web 技术的不断发展和普及,它已经从最初的人类社会的辅助设施逐渐形成一个与现实世界相对应的 Web 世界(Web-wide World),作为现实世界的映像,与现实世界紧密的融合在一起。因此,Web 计算已经成为推动并主导人类社会发展的越来越重要的因素。文章通过研究和分析 Web 的发展以及 Web 和 Grid 相互融合的趋势两大背景,将 Web 计算定位为一种 Web 资源的组织、管理和使用的分布式模型和技术。然后,分析了 Web 计算所具有的特点,如普遍性、连通性、位置透明性等;进一步,总结了 Web 计算研究所必须面临的一些任务,包括资源标识与定位、资源表示、资源传输以及信任管理。其中,Web 世界的信任管理研究,将是 Web 计算研究的一个重要的方向和趋势,其目的就是为了提供一个诚信、合作的信任 Web(Web of Trust)。

关键词 Web 计算, Web 世界, Web 资源, 信任 Web

Web Computing: The Network Computing of Web Resources

ZHANG Chang-Li WU Jian HU Zheng-Guo

(Department of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072)

Abstract After a long-term development and population, the World Wide Web (Web for simplification) has been becoming a "Web-Wide World" (Web World for simplification) as a mirror and extension of our real world, and has been interweaving and linking with the real world deeply. Thus, Web computing is becoming a more and more important, even leading, factor in the development of our human society. In this paper, the evolution of Web is analyzed and that Web will merge with Grid computing as a conclusion is drawn. Thus, Web Computing can be treated as models and relating technologies for organization, management and utilization of Web Resources. Then, the characteristics of Web Computing are listed, such as universality, connectivity and simplicity etc. At last, the emergent tasks of Web Computing researching, including resource identification and location, resource representation, resource transportation, and trust management, are talked. Among these tasks, trust management is the most emergent issue, and its aim is to provide a sincere, trustful and cooperative Web of Trust.

Keywords WSeb computing, Web world, Web resource, Web of trust

1 前言

以 URI、HTML、HTTP 为代表的 Web 技术,经过约 20 年的发展,已经从最初的实验室产品扩展成全世界无所不在的信息网,已经从最初的以面向浏览器用户的数据共享为目的的网络模式逐渐发展成面向智能体(Agent)的提供各种计算服务的网络计算环境^[1]。并且,Web 已经从最初的人类社会的辅助设施逐渐形成一个与现实世界(Real World)相对应的 Web 世界(Web-wide World),作为现实世界的映像(Mirror of the World),与现实世界紧密的联系(Link)并交织(Interweave)在一起^[2,3]。不可置疑,Web 计算——现实世界各个领域的形态和行为的 Web 世界映像(如 E-Business、E-Commerce、E-Manufacturing 等)——已经成为推动并主导人类社会发展的越来越重要的因素。

本文第 2 节通过分析 Web 发展过程,得出 Web 与网络计算相互融合的发展趋势的结论;基于这样的结论,在第 3 节重点讨论了 Web 计算的定位和所要解决的问题;进一步,文章第 4 节和第 5 节对 Web 计算的特点和任务进行了详细的归纳和讨论;最后,对全文进行了简要总结。

2 Web 的发展与展望

最初的 Web 是一个面向浏览器用户的一个文档网(Web

of Document)^[4]。随着 Web 的不断发展壮大,它已经逐渐成为一个信息资源的载体,承载着无数的(Zillions of)数据资源和软件资源。但是,由于这些资源的异构性以及应用程序边界和企业边界的限制,形成一个个“信息孤岛”。这些信息尽管有着不可估量的价值,但是却无法进行有效利用。于是,数据资源和软件资源的集成(Integration)成为一个重要的、亟需解决的问题。因此,许多 Web 技术应运而生,其中最主要的两个是:Web 服务和语义 Web。

Web 服务是一个崭新的分布式计算模型,是用来对 Web 上的应用程序进行跨应用程序边界和企业边界集成的有效机制,其目的就是为了创建一个服务网(Web of Services);语义 Web 是现有 Web 的扩展,它提出的目的,是为了解决跨越应用程序边界和企业边界的数据集成的问题。与 Web 不同,语义 Web 是数据网(Web of Data),语义 Web 中的信息被赋予定义良好的含义,更便于计算机和人的协同^[4]。

Web 服务和语义 Web 都是基于 Web 技术建构的。根据 Tim BL 的观点,Web 服务的创建是为了满足当前的立即需要(应用集成需要),而语义 Web 的建构则是为了一个更加长期的目的,它将是现有 Web 的进一步演进;尽管二者的发展相互独立,但是它们最终将自然而然的相互融合,合二为一^[5]!

张昌利 博士研究生,主要研究领域为计算机软件工程与网络软件;吴 健 教授,主要研究方向为嵌式系统、软件理论、网络计算;胡正国 教授,博导,主要研究领域为高性能计算、软件工程与理论。

在 Web 迅速发展的同时,许多不同领域的科学工作者对共享大型计算机(Supercomputer)、有效使用大量闲散计算机资源进行大规模的实验数据计算的需求,导致了网格计算(Grid Computing)的产生。从宏观上来看,Web 和网格计算的一个最根本的差别在于:Web 的计算是离散的、无状态的;正好相反,网格计算是有状态的,甚至其计算资源也是有状态的。但是,当前许多人认识到,Web 和网格计算最终将殊途同归^[5]!比如,最近提出的开放的网格服务架构 OGSA (Open Grid Service Architecture)和 Web 服务资源框架 WS-RF(WS-Resource Framework)就是相关组织在这方面的努力。

3 Web 计算:针对 Web 资源的网络计算

从 Web 资源的角度来看,随着 Web 技术和网格计算的发展与融合,网络资源已经不再是以前的以 HTML 静态页面为代表的面向浏览器用户的文档资源,而是通过 Web 服务技术产生的一系列的具有可操作性的软件和服务资源(有状态资源),以及通过语义 Web 技术而产生的一系列具有可交换和共享的数据资源(无状态资源),而且,这些资源都是直接面向网络实体的,是机器可识别、可操作的网络资源。

在这里,我们将 Web 计算定位为一种借助于 Web 技术(如 Web 服务和语义 Web 技术)的 Web 资源——数据资源和服务资源——的组织、管理和使用的分布式计算模型及其相关技术。其中,所有的数据资源(如以关系数据库为代表的结构化数据、以 Web 文档为代表的半结构化数据,乃至以多媒体数据为代表的各种无结构数据),均可组织成 XML(或者 RDF、OWL)形式进行管理,使用 SOAP 协议进行访问、交换和共享;同时,所有的软件和服务资源,都可以封装成 Web 服务的方式进行管理、共享。基于这样的观念,Web 计算和 Web 是统一的,是符合 Web 发展方向的。

4 Web 计算的特点

Web 计算是建构于 Web 之上的计算模式,因此它必将兼有 Web 的特性。也就是说,Web 计算应具备如下的特点:

(1)继承自 Web 的特点,诸如普遍性(University: Anything can link to anything)、连通性(Connectivity)、简单性(Simplicity)、位置和操作透明性(Location Transparent & Operation Transparent)等。

(2)完全基于 Web 架构。最初的 Web 架构由 URI、HTML、HTTP 组成;随着 Web 的发展,其架构从深度和广度上都有了很大的扩展,如 Web 服务和语义 Web 架构。但总体来看,其核心包括资源标识机制(URI)、资源描述标准(XML、RDF 或 OWL、WSDL、OWL-S 等)以及资源交换通信协议(SOAP)几个部分。这是 Web 计算的一个核心特点,也是判别一个计算模式是否为 Web 计算的重要标准。

(3)面向用户(Human)、组织(Organization)、或者各种智能终端——如机器人甚至智能灰尘(Smart Dust)——等普遍的现实世界主体(Principal),当前常用的客户端如浏览器(Browser)和 Agent 等。

(4)Web 计算是满足 Web 标准的数据资源和服务资源的集成机制。其中,数据资源必须满足 XML、RDF 和 OWL 等 Web 数据描述标准,服务资源必须满足 WSDL、OWL-S 等 Web 服务描述标准。

5 Web 计算的任务

Web 的创建之初,为了满足 Web 的简单性和普及性,不得不牺牲诸如数据一致性(Consistency)、数据真实性(Authenticity)、数据完整性(Integrity)、数据表述能力(HTML 具有局限性)、数据可证明性(Provability)等等特性作为代价^[1]。随着 Web 世界与现实世界的联系越来越紧密,这些 Web 的遗留问题逐渐凸现出来,成为 Web 计算研究必须面对的任务,主要有:

(1)资源的标识与定位。当前的 Web 使用 URI(URL)作为 Web 资源标识和定位机制,但是这一机制存在着巨大的缺陷,比如“死链接”、标识能力有限(无法有效标识 Web 用户,当前许多实践使用 Email 进行标识^[6],如 FOAF 项目,以及 Wallop.com、Hi5.com 等网站就是采用这种机制)等问题。OASIS 组织提出的 XRI(Extended Resource Identifier)^[7]机制是 URI 和 URN 的扩充,其目的就是为了解决这些问题,当前正处于完善阶段中。

(2)资源描述。当前,Web 上的服务资源通常使用 WSDL 或者 OWL-S 描述,数据资源通常使用 XML、RDF、OWL 等标准格式描述。在 Web 计算体系下,需要一种统一的机制对这两种资源进行一致的描述。相应地,按照 Tim BL 的观点,所有的 Web 计算资源都将使用语义 Web 技术——即 RDF 和 OWL 进行描述。

(3)数据传输。在 Web 服务中,使用 SOAP 实现了基于 HTTP 协议和 XML 序列化技术的 RPC 机制;在语义 Web 中,Tim BL 则建议可以使用 SOAP 传递 RDF 数据,如 RDF 查询数据和 RDF 更新数据。OASIS 则提出了一套新的数据传输机制,意图实现基于 XRI 技术、包含契约和协商机制的数据共享功能^[8]。相应地,按照 Tim BL 的观点,所有的 Web 计算资源的交换与传输,将使用 Web 服务技术——即 SOAP 进行描述^[4]。

(4)信任(Trust)管理问题。Web 计算的信任管理问题包括两方面:首先,Web 上充斥着大量的数据资源和服务资源,这些资源或新或旧、或正确或错误,互相冲突甚至互相矛盾,需要进行有效的甄别、选择,从而找出最正确、最可信的资源;另一方面,Web 世界主体(如 Agent、浏览器等)具有匿名性,也需要有效的机制管理本地资源,防止非法使用和非法传播。总之,所谓信任就是 Web 计算主体(如 Agent)基于历史交互经验对其它主体未来行为的主观期望,是一种带有主观性、风险性和脆弱性的理性行为^[9]。当前对 Web 计算信任管理的研究尚处于初始阶段,仍集中在不同 Web 实体间互操作信任的层面。但是,正如许多人所指出,信任问题的研究将是 Web 计算研究的一个重点方向和趋势,其目的就是要建立一个诚信、合作的基于 Web 的分布式计算环境,也就是一个信任 Web(Web of Trust)^[2, 6, 10, 11]。

结束语 Web 的发展,已经远远超越了普遍的数据共享系统的初衷,已经逐步扩充并发展成为一个与现实世界相对应的虚拟世界,延伸并不断影响现实世界。正如文^[12]所说,基于 Web,我们所处的这个复杂、纷乱的世界才真正地掌握到了我们的手中。

Web 计算是 Web 资源的组织、管理和使用的分布式模型和技术,随着 Web 世界与现实世界的交织,它已经成为一个推动并主导人类社会发展的越来越重要的因素。但是,与 Web 一样,Web 计算仍旧还是一个发展中的技术,仍旧需要

从许多方面进行深入分析和研究。尤其 Web 世界的信任问题,更是 Web 计算亟需面对和解决的重要问题。

参考文献

- 1 TIM BERNERS-LEE. The world wide web; past, present and future[DB/OL]. http://www.w3.org/2002/04/Japan/Lecture.html, 1996. 8
- 2 Khare R, Rifkin A. Weaving a web of trust[DB/OL]. http://www.4k-associates.com/trust.html, 1997
- 3 Kunii T L. Cyberworld modeling integrating cyberworlds, the real world and conceptual worlds[A]. In: 2005 International Conference on Cyberworlds (CW'05), 2005. 3~11
- 4 Berners-Lee T. WS-SW: integrating applications[DB/OL]. http://www.w3.org/2003/Talks/05-gartner-tbl/slide1-0.html, 2003. 5
- 5 Wilson M, Matthews B. The future of the world wide web? [J]. BNCOD 2004, LNCS 3112, 2004. 4~15
- 6 Golbeck J, Parsia B, Hendler J. Trust networks on the semantic

- 7 web[A]. In: Proceedings of Cooperative Intelligent Agents. 2003 OASIS. XRI 2.0 FAQ[DB/OL]. http://www.oasis-open.org/committees/download.php/15695, 2005. 12
- 8 Reed D, Strongin G. The dataweb; an introduction to XDI[DB/OL]. http://www.oasis-open.org/committees/download.php/5115/wd-xdi-intro-white-paper-2004-01-20.pdf, 2004. 1
- 9 Mui L, Mohtashemi M, Halberstadt A. A Computational Model of Trust and Reputation[A]. In: Proc. of the 35th Hawaii International Conference on System Sciences, Big Island, Hawaii, IEEE (2002)
- 10 Heymans S, Nieuwenborgh D V, Vermeir D. Preferential reasoning on a web of trust[A]. ISWC 2005, LNCS 3729, 2005. 368~382
- 11 Golbeck J, Grau B C, Halaschek-Wiener C. Semantic web research trends and directions[A]. PReMI 2005, LNCS 3776, 2005. 160~169
- 12 Gelernter D. Mirror worlds; the day software puts the universe in a shoebox... how it will happen and what it will mean[M], Oxford University Press, 1991

(上接第 113 页)

传输服务提供者便由这 30 个函数组成。

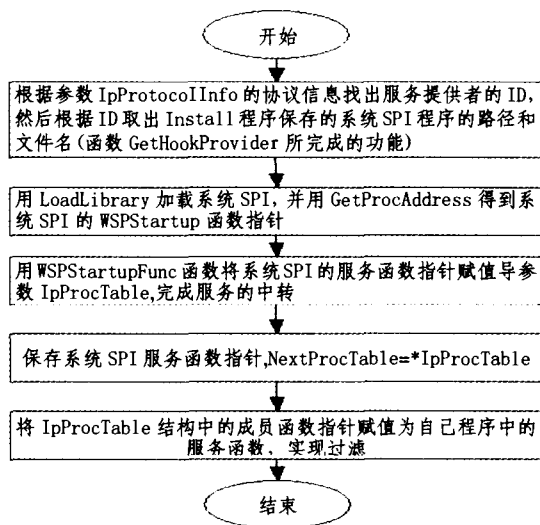


图 3 WSPStartUp 程序流程

WSPStartUp 完成了对 SPI 服务函数的过滤, 所有的 Windows Sockets 调用都会首先经自己的服务函数, 然后自己的服务函数大多调用已经保存在 NextProcTable 中的系统服务提供者函数。

(3) 得到服务提供者路径的相关函数

一个是用于得到此防火墙自己的服务提供者路径的函数, 另外一个是用来得到已经被防火墙自己的服务提供者函数所替换掉的系统自己的服务提供者函数。这两个函数的实现都是通过读取注册表的相关信息来实现的。而注册表的相关内容则是在安装防火墙系统时, 由防火墙安装程序在安装过程中写入的。

(4) 用于提供服务的相关函数

Windows 用于提供服务的函数共有 30 个。可以视需要来重载这些函数。

4.2 管制服务提供者函数部分

这些函数通过调用与控管规则直接打交道的控管规则匹配管制函数来得到结果。

4.3 控管规则管制函数

这一部分代码主要用于匹配内存中的 Session 与控管规则, 通过匹配得到的内容来决定是否放行封包。所有关于控管规则与 Session 函数的决策都在此一部分完成。

4.4 封包处理函数部分

封包处理函数部分的主要任务是在收到或发送封包的时候会在内存中创建 Session 来保存封包的相关信息, 以供底层动态链接库的其他函数调用。这一部分的封包处理函数正是用来处理这些 Session 所用。在这一部分的函数中, 主要包括了 Session 的初始化、Session 的创建、Session 的删除、查找、修改等。另外还有几个函数负责与界面应用程序的通信等。

4.5 界面程序设计与编码

界面程序设计部分, 首先需要考虑的便是程序的人机界面。由于要获得用户输入的控管规则, 因此控管规则添加部分是必不可少的, 同时也需要让用户在添加完控管规则以后有个实时的反馈, 所以需要有一个窗口单独显示控管规则的内容, 选择使用 MFC 提供的 List 控件来完成这项工作。除了考虑控管规则的添加以外, 还需要考虑控管规则的编辑与删除等。因此, 在添加按钮后, 又新加了编辑与删除按钮。同时, 在此页面上也包含了设置系统工作模式的几个控件。用户只需通过点击这几个控件就可以选择系统的工作模式: 全部放行、全部拒绝、询问。而底层动态链接库也会根据这里的工作模式的设置来决定其工作方式。在用户点击添加规则的按钮以后, 还需要有一个页面来显示添加规则的界面, 用户可以自行定制规则。另外, 为了使该防火墙在用户最小化界面的情况下仍然能够使用, 因此, 构建了一个隐藏的主窗体作为在界面最小化时与底层动态链接库进行消息通信之用。

结论 Winsock SPI 编程, 所编写的文件以 DLL 的方式存在。编程与调试都比较方便, 同时其可移植性也较佳。只有告之开发者 DLL 所提供的接口便可以借用此 DLL 对表现层进行开发, 而不需要另写底层代码。此防火墙工作对于系统资源的占有很少, 同时, 由于其工作在应用层, 因此能够很灵活的对使用网络的应用程序进行管理。通过在过滤规则中选择相关的应用程序名称, 用户就可以决定对这些应用程序的网络封包的操作动作, 十分方便。除却必须到驱动程序层进行控制的协议类型以外, 该防火墙对于其他的网络服务都可以有效的进行监控。

参考文献

- 1 张琳, 李璇华. 网络组建、管理与安全[M]. 人民邮电出版社, 2000
- 2 Ziegler R. Linux 防火墙. 余青霞, 周钢译. 人民邮电出版社, 2000. 10
- 3 Vacca J. Intranet 的安全性[M]. 电子工业出版社, 2000
- 4 Aggarwal C, Wolf J L, Yu P S. Caching on the World Wide Web[J]. IEEE Transactions on Knowledge and Data Engineering, 1999, 11(1)