

基于 Internet 的信息网格的软件框架研究

桂小林

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 为了实现数据信息在 Internet 上的广域共享和快捷获取, 提出了一种信息网格的软件框架. 参照计算网格的开放式网格服务构架, 通过拓展信息搜集、监测、登记、发现、预订和搜索等服务, 建立了面向 Internet 的信息网格的软件组件模型, 研究了基于计算网格构建信息网格的方法. 基于容错目的, 提出了主、从层次的网格目录结构, 研究了网格目录的复制和引用机制. 为了实现网格信息的搜集与发现功能, 设计了结点资源监测代理. 通过提出网格信息登记协议和网格文件传输协议, 实现了信息预订和信息搜索服务. 使用 JAVA 技术开发了专门的 Shell, 以实现信息网格的服务接口. 通过该接口, 远程用户可以方便、灵活地使用信息网格提供的计算与信息服务, 实现广域资源的单一访问入口.

关键词: 信息网格; 软件框架; 信息共享; 元数据; 网格计算

中图分类号: TP303; TP31 **文献标识码:** A **文章编号:** 0253 - 987X(2004)06 - 0551 - 04

Research of the Software Framework Based on Internet-Wide Information Grid

Gui Xiaolin

(School of Electronics and Information Engineering, Xi an Jiaotong University, Xi an 710049, China)

Abstract: To share the wide-area distributed data and information resources interconnected with Internet, a software framework for the information grid is introduced. Referencing to the open grid service architecture of the computational grid and extending the services such as information collection, monitoring, register, discovering, booking and searching, a component model of the information grid software is presented. Based on the campus computational grid named WADE, a software method for information grid is also addressed. For the purpose of the fault-tolerance, a mast-slave meta-data catalog structure is proposed and the replication and reference mechanism of the catalog are explored too. In addition, a resource monitor agent is designed to discover and collect dynamical resources, and a grid information register protocol and a grid file transfer protocol are introduced for booking and searching the grid information. A service interface to the information grid is then implemented using a shell which is a version called GridShell developed by using JAVA techniques. With this shell, remote users can utilize the computing and the information services provided by the grid conveniently and flexibly and the wide-area resource is realized through a single assess portal.

Key words: information grid; software infrastructure; information share; meta data; grid computing

网格从其功能强度上可分为知识网格、信息网格和计算网格^[1~2]. 计算网格聚合网络上分布的计算机、工作站、机群、群集、数据库、高级仪器和存储设备等, 形成对用户相对透明、虚拟的高性能计算环

境, 目标是解决大规模的科学计算问题. 基于计算网格的基础构件, 信息网格主要研制一体化的智能信息处理平台, 消除信息孤岛, 方便用户发布、处理和获取信息. 信息网格包含计算网格的全部功能, 重点

收稿日期: 2003 - 08 - 22. 作者简介: 桂小林(1966~), 男, 博士, 副教授. 基金项目: 国家自然科学基金资助项目(60273085); 教育部 China Grid 计划资助项目; 国家高技术研究发展计划资助项目(2001AA111081).

拓展了广域信息共享服务.一方面,网络中接入的计算机数量日益增多,但使用效率并不高;另一方面,互联网内容每天飞速增长,不可能有哪个单一的服务器或搜索引擎能够掌握所有资源.互联网上每年增加的内容大约只有百万分之一能为公众所用,即便功能强大的搜索引擎也只能查找其中的极少内容,而且找到的逻辑内容很大部分不能物理使用.信息网格利用数据挖掘、信息融合和资源引擎等技术完成网络资源的搜集与共享,并利用单一系统映像技术完成信息的单一映像和有效融合,保证逻辑资源的物理可用性,便于用网格引擎实现资源的获取.信息网格作为网格的高端研究领域,主要是参照计算网格的开放式网格服务架构(OGSA)^[3~4]进行软件开发,还未形成信息网格的体系结构标准.通过对广域分布式环境(WADE)^[5]进行技术拓展,本文提出了一种支持信息网格的软件框架的广域分布式环境 WIIG (WADE Infrastructure for Information Grid),并详细讨论了实现这种软件框架的关键技术.

1 信息网格的软件框架

广域一体化的信息服务是信息网格的核心.参照 OGSA,并以西安交通大学校园计算网格 WADE 为基础,提出了一种构建在计算网格之上的信息网格的软件框架 WIIS,如图 1 所示. WADE 支持面向计算网格的计算服务, WIIS 则重点拓展了信息服务组件.这些组件包括信息的搜集、监测、登记、发现、预订和搜索等. WIIS 信息服务通过元数据目录完成广域信息的单一映像和有效融合,使用网格引擎可方便地实现资源的获取.下面简要说明主要服务组件和网格目录的功能.

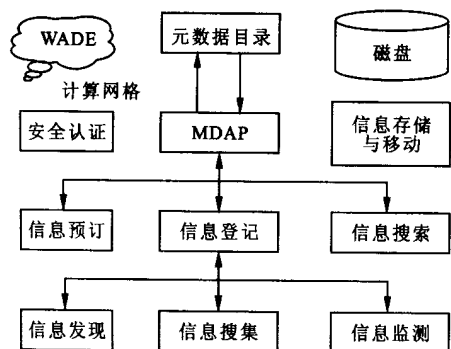


图 1 信息网格的软件框架

(1) 网格目录. 网格目录又称元数据目录 (MDC), 它构成一个分层的树形结构, 该结构有利

于实现网格管理功能的分布和网格系统的扩展. WIIG 的网格目录分为计算网格目录 (“/ GCS”) 和信息网格目录 (“/ GIS”), 前者为网格计算提供信息共享, 后者管理动态和随机的共享信息. 元数据访问协议 (MDAP) 为信息的搜集、登记、发现、预订和搜索提供了访问网格目录的协议接口.

(2) 信息登记. 它为信息的发现、搜集提供服务. 信息提供者使用驻留进程将搜集到的各实体信息周期性地登记到 MDC 中, 而 MDC 则为每一实体信息建立一个索引项. 信息登记是一个动态过程, 以适应网格资源的变化. 这种动态登记机制能够使网格内的信息始终保持最新.

(3) 信息搜集. 每个信息提供者通过驻留进程搜集自身的两大类信息, 一类是实体静态信息, 如操作系统版本、CPU 类型、处理机数、平台体系结构、内存容量、总磁盘空间、网卡型号等; 另一类是实体动态信息, 如文件信息、负载信息、CPU 利用率、可用内存或磁盘空间、网络带宽和延迟等.

(4) 信息搜索. 信息请求者通过 Broker, 利用 MDAP 查询距离它最近的元数据目录, 可能会发现很多与目标相关的实体信息, 这与 WWW 上的搜索引擎功能类似. 从发现的信息中选出感兴趣的若干项, 通过进程 “gftpd” 直接去查找这些实体信息的提供者, 从而获取更详细的信息.

2 元数据目录

MDC 按照一定的数据模型来存储和管理网格对象数据库. 一个信息网格至少要建立一个主网格目录, 一个管理域内要建立一个子网格目录. 主网格目录负责子网格目录的交互, 存放子网格目录的索引. 子网格目录可以采用多层结构, 整个信息网格目录构成一个树型结构. 为了简化, 校园网络内最好只建一级子网格目录. 本文基于标准轻量级目录访问协议^[6]实现网格目录. 在被开发的信息网格中支持并行计算服务和信息共享服务的元数据目录结构如图 2 所示.

网格目录的管理包括网格对象的存储、移动、删除, 以及网格目录的复制和引用. 网格目录的复制是为了增强网格信息服务的可靠性. 一个信息网格内只有一个主网格目录负责网格对象的管理是不够的, 一旦主网格目录崩溃, 整个信息服务将被迫终止, 无法提供连续可靠的服务, 因此需要引入从网格目录. 一个主网格目录可配备多个从网格目录, 备份它自己的全部内容. 主、从网格目录都在上级网格目

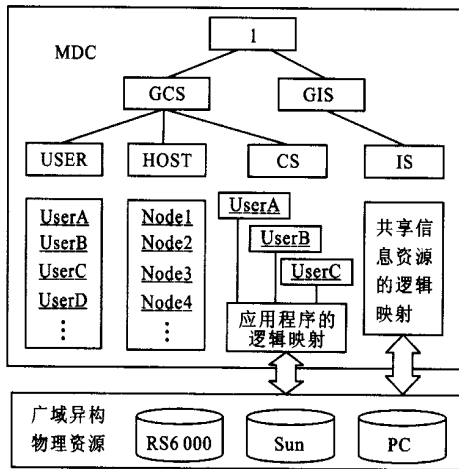


图 2 信息网络的元数据目录

录中存有索引,该索引指明下级网格目录所在服务器的 IP、端口和访问方法等.当主网格目录崩溃时,上级网格目录会自动地指向工作正常的从网格目录.通过网格目录之间的自适应复制功能可以保持网格目录之间的数据一致性.

引用机制实现信息网络的分布性和扩展性.随着网格规模的扩大,引用机制可保证跨管理域的查询依然有效.其方法是:一个子网格目录保存着上级主网格目录的指针,当不能满足信息查询要求时,就把它保存的上级网格目录的指针返回给信息请求者,信息请求者再按照该指针去查找上级网格目录;若找到所要的结果,直接返回给信息请求者,不需再经过下级的网格目录.引用机制既提供了信息查询的上溯功能,又避免了层层搜索网格带来的时间消耗,提高了查询效率.

3 信息网格的关键技术

3.1 信息监测、搜集与发现

信息搜集与信息发现模块可以通过信息动态监测软件来实现.网格是一种地理上分布的、跨管理域的广域共享资源,各种资源性能会随时间而变化,甚至不可用.资源监测器周期地检测网格目录中的各种资源的可用性.每个注册结点安装一个守护进程,监视本机的资源变化情况,定期刷新网格目录,监视结点的开、关机情况.与手机开、关机类似,守护进程在开机后启动,并通过 MDAP 刷新结点状态.当结点正常关机时,通知网格目录更新相关结点信息;如果结点非正常关机,则无法刷新结点状态信息,这时启用服务器守护进程,定时刷新注册节点信息.

信息多样性和分散性增加了资源管理的难度.

将信息资源限定以文件形式表示的软件、程序和数据库,可以简化信息共享方案.如果信息共享范围包括每个网格结点的全部文件的话,会带来安全隐患,所以,将共享对象限制到每个网格结点的一个目录(称为共享目录)上,该目录的全部文件作为信息共享的物理目标.当共享目录的物理资源发生变化时,逻辑资源也要随之变化.文件资源动态监测系统分为两个层次:资源状态监视层和通信层.资源状态监视层负责资源扫描,监视状态变化,发现变化立即调用通信层,并向网格目录报告资源的变化状态.通信层接收资源状态监视层的调用,根据资源变化状态刷新元数据目录.

在资源状态监视层使用了 2 种算法:一个是目录遍历算法,另一个是扫描监视算法.前者采用了递归的深度优先遍历算法,后者是在目录遍历算法的基础上扫描遍历到的所有文件,记录遍历到的每一级目录下的每一个文件的路径和文件名,并保存扫描结果.在每次目录遍历之后,将本次与前次扫描结果进行比较,确定文件的增删及改名等变化.文件改名是增加一个新文件和删除一个旧文件的过程,需要两次通信,虽增加了开销,但简化了设计.

3.2 信息登记

信息监测、搜集与发现模块获得的结果通过信息登记模块映射至元数据目录.当信息提供者搜集到网格实体的原始信息后,先进行格式化,并使用 MDAP 将信息登记到 MDC 之中.网格信息登记过程使用了基于事件驱动的登记机制和网格对象安全传输机制,下面分别叙述.

(1) 事件驱动的登记机制.网格实体的动态性使得网格信息的登记既不能是连续的,也不能是静态的,需要根据网格内产生的某种条件来动态登记,这种条件的产生称为事件.只有当事件触发的时候,也就是条件满足的时候,才临时使用已提供的对象元素来生成网格对象,并将其登记到网格目录中.

(2) 网格对象安全传输机制.网格信息登记过程是在信息提供者和网格目录之间传递网格对象的过程,为了防止信息泄漏,尤其是防止用户认证信息被非法网格实体窃取,必须采取某种安全传输机制进行数据传输.

3.3 信息搜索

信息网格的信息搜索过程如图 3 所示.图中,信息请求者根据需要请求其他网格实体的信息;信息提供者动态搜集自身或其邻近实体的信息,并将搜集到的信息登记至元数据目录;利用元数据访问协

议,信息请求者可以发现或查找其请求的实体信息.

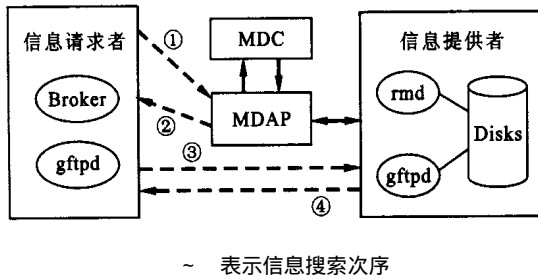


图3 信息网格的信息搜索过程

信息搜索的目的通常是下载所要资源,这时需要进行物理资源访问.信息请求者搜索网格信息(虚线),实际上就是对信息的物理资源进行访问.在访问物理资源之前,要进行逻辑资源访问,这是一般意义上的搜索引擎的功能.信息网格将搜索与下载合为一体,体现了共享资源的本地化.图3中的、实现逻辑访问,从MDC中获得物理资源的结点位置(通常意义上的搜索);、根据资源位置,使用进程“gftpd”实现物理资源的远程读写操作.“gftpd”进程是一个以Bash 2.0.5为基准、简化了的网格Shell程序,它是一个以C/S模式构建的程序,客户端是命令行解释器,服务器端是命令操作解释器,支持文件的读写与执行功能.为了实现磁盘操作的远程执行,“gftpd”劫持Bash的`execution-disk-command()`和`execv()`函数,完成文件的远程读写与执行.服务器端接收劫持函数发送来的命令,调用本地命令执行读写操作,操作结果定向传至客户端.

4 用户接口

信息网格的中间件软件需要友好的人机交互接口.WADE计算网格将被移植到信息网格中,它也支持像Web、Telnet那样的一般分布式系统的用户接口.Shell是用户与类Unix操作系统内核的常用接口.校园信息网格提供Grid Shell(简称GSh),支持命令行操作的用户.用户使用GSh就像使用本机Shell那样自然、方便,用它可以访问到整个校园网格的信息.GSh命令规则与Bash 2.0.5类似,在计算网格实现的GSh命令可以方便地移植到信息网格之中.GSh程序是由GNU的bash程序改造而来,程序执行后会启动一个新的客户端GSh,命令由它解释执行.GSh在WADE信息网格中实现了7个基

本命令,含义如下.

(1) `gsh`:用户登陆或退出Grid Shell的命令,输入“`gsh - l`”,用户名与口令,成功后进入GSh界面;输入“`gsh - q`”退出Grid Shell,返回Bash界面.

(2) `ls`:列出元数据目录中指定目录下的内容.

(3) `cd`:进入元数据目录中指定的目录.

(4) `pwd`:显示用户当前所在的元数据目录的位置.

(5) `reg`:登记命令,其功能是将用户名、口令以及本地磁盘文件登记到网格目录中.

(6) `mv`:在网格目录下进行文件改名.

5 结论

基于计算网格,提出了信息网格的软件框架,设计了信息网格的功能模型,讨论了相关功能模块的实现方法及关键技术的解决手段.作为网格的高端研究领域,信息网格参照计算网格OGSA架构,通过扩展信息搜集、监测、登记、发现、预订等服务,建立了面向Internet的信息网格的软件体系结构.在该结构基础上,开发了实现计算与信息服务一体化的接口软件GridShell,为信息网格的中间件开发提供了一条可行之路.进一步的研究工作将包括可信服务模型,可靠计算模型和信息安全模型.

参考文献:

- [1] Foster I, Kesselman C, Tueche S. The Anatomy of the grid: enabling scalable virtual organizations [J]. International Journal of High Performance Computing Applications, 2001, 15(3): 200~222.
- [2] Czajkowski K, Fitzgerald S, Foster I, et al. Grid information services for distributed resource sharing [A]. Proceedings of the Tenth IEEE International Symposium on High Performance Distributed Computing [C]. Los Alamitos, USA: IEEE Comput Soc, 2001.
- [3] OGSA-WG. GS-Spec-draft03 [EB/OL]. <http://www.gridforum.org/ogsa-wg/drafts/>, 2003-05-11.
- [4] OGSA-WG. Ogsa-draft2.9 [EB/OL]. <http://www.gridforum.org/ogsa-wg/drafts/>, 2003-05-11.
- [5] 桂小林,钱德沛.基于Internet的网格计算模型研究[J].西安交通大学学报,2001,35(10):1008~1011.
- [6] RFC2849-2000, The LDAP data interchange format (LDIF):technical specification [S].

(编辑 苗凌)