

面向多自治域网格的信息服务模型及其实现

张海辉¹⁾ 周兴社¹⁾ 杨志义¹⁾ 吴小钧²⁾ 杨 刚¹⁾

¹⁾(西北工业大学计算机学院 西安 710072)

²⁾(长安大学信息工程学院 西安 710064)

摘 要 网格是实现分布异构资源共享的有效模式,而信息服务实现系统服务与资源的有效管理,是网格系统的重要组成部分. ChinaGrid 是由多个自治域组成的大规模网格,现有的信息服务不能满足此类系统特性与应用需求. 文中提出网格信息服务体系 GISA2.0,强化了域自治管理和资源信息的安全性. GISA2.0 实现了可扩展的网格信息模型和面向服务、支持多种监控信息聚集的层次化信息管理框架. 提出了基于分布 XPath 引擎的多域资源信息检索机制,实现了安全、快速和用户相关的虚拟全局资源视图.

关键词 多自治域网格; ChinaGrid; 信息服务; 分布 XPath 引擎; 资源视图
中图法分类号 TP311

Research and Implementation of Information Service Model on Grid with Multiple Autonomous Domains

ZHANG Hai-Hui¹⁾ ZHOU Xing-She¹⁾ YANG Zhi-Yi¹⁾ WU Xiao-Jun²⁾ YANG Gang¹⁾

¹⁾(College of Computer Science, Northwestern Polytechnical University, Xi'an 710072)

²⁾(College of Information Engineering, Chang'an University, Xi'an 710064)

Abstract Grid is an effective pattern for implementing heterogeneous resource sharing. Information service is a vital component of grid system, it carries out efficient management of system services and resources. ChinaGrid is a large-scale grid containing multiple autonomous domains. Existing information services can not meet the requirements of systems and applications with such kind of characteristics. This paper presents GISA2.0 (Grid Information Service Architecture, GISA), focusing on autonomous domain management and resource information security. GISA2.0 implements a scalable grid information model and service-oriented hierarchical information management framework, which supports the aggregation of monitoring information from multiple monitoring systems. To achieve secure, fast, and user-related virtual global resource view, GISA2.0 uses a cross-domain information retrieval mechanism based on distributed XPath engine.

Keywords multiple autonomous domain grid; ChinaGrid; information service; distributed XPath engine; resource view

1 引 言

网格被描述成在大规模分布系统中,支持分散团

体构成虚拟组织,为实现共同目标进行协同计算^[1]. 目前,对超大规模网格的研究较少,这类网格系统具有由众多自治域组成,存在丰富网格类型和海量网格信息等特点. ChinaGrid^[2]是由教育部实施的大型

收稿日期:2006-04-27;最终修改稿收到日期:2007-09-14. 本课题得到国家自然科学基金(2001CG1101)、国家“八六三”高技术研究发展计划项目基金(2006AA01Z162)和教育部中国教育科研网络计划 ChinaGrid 资助. 张海辉,男,1977 年生,博士研究生,研究方向为分布计算与网格计算. E-mail: zhh409@tom.com. 周兴社,男,1955 年生,教授,博士生导师,研究领域为分布计算与嵌入式计算等. 杨志义,男,1952 年生,教授,博士生导师,研究领域为分布计算. 吴小钧,男,1972 年生,博士,讲师,研究方向为分布计算. 杨 刚,男,1974 年生,博士,讲师,研究方向为分布计算.

网格项目,目标是整合 CERNET 网络资源,为教育和科研提供高质量的计算与数据服务. ChinaGrid 不仅包括图像处理、大学数字博物馆、生物信息、流体力学等跨学校的专业化应用网格,而且包括相关高校的校园网格,从而形成了多个自治域的网格系统.

ChinaGrid 的各个自治域在网格资源属性、运行模式上具有不同特点,从而对信息服务提出了多种需求:(1)多类型网格信息描述.大量网格系统^[3]中的资源通常指以服务形式封装的计算资源和存储资源,而在描述网络资源、设备资源(如图像处理网格中的扫描设备、三维显示设备)时存在不足.(2)自治域个性化管理.大部分系统的虚拟组织^[1]突出了其动态性,而很少考虑其不同个性的自治域组织.(3)可扩展性.满足连接众多子网格和大量网格资源的需求,表现在自治域的可扩展和管理资源的可扩展.(4)灵活、安全的信息检索.网格中存在大量不同权限的用户,信息服务应实现灵活的跨域和用户相关的查询.

已经存在的网格信息服务并未完全地满足以上需求. Globus 的 MDS4^[4]以信息聚集的方式实现统一的状态收集,多级的信息聚集容易导致性能瓶颈,限制了可扩展性;资源监控服务 Ganglia^[5]、MonALISA^[6]等依赖组播机制进行查找,应用范围有限,当物理和虚拟组织结构不一致时不正确;GGF 提出网格监控体系 GMA^[7],基于事件的方式支持不同协议的信息交互,实现了基于 GMA 的信息服务,由于存在集中注册器,可扩展性差.同时,大部分网格信息服务^[4,8-10]还未考虑域自治性和信息安全带来的管理需求.

CGSP^[11]为 ChinaGrid 网格系统的开发支撑平台提供支持,2.0 版本于 2006 年 4 月发布,已成功应用于多个专业网格和 20 多所高校的校园网格. 作者在设计 CGSP 信息服务时,提出并实现了网格信息服务体系 GISA2.0,其建立了多自治域环境中统一的信息模型、资源监控框架和信息获取机制. GISA2.0 包含以下 3 部分:(1)可扩展的网格信息模型,包括资源类型的可扩展和资源信息存储结构的可扩展以及全局资源文档模型;(2)层次化信息管理:将网格资源信息和自治域属性信息分离,实现网格信息本地管理和域信息全局管理两层结构,并将信息安全贯穿其中;(3)虚拟全局资源视图:通过分布的 XPath 查询引擎,将物理上分布在多个域的网格信息组织成虚拟全局资源文档,实现安全、快速和用户相关的虚拟全局资源视图.

本文第 2 节阐述可扩展的网格资源信息模型;第 3 节提出层次化的网格信息管理;第 4 节为虚拟全局资源视图的相关实现技术;第 5 节介绍相关工作及评价;第 6 节给出结论并指出下一步研究内容.

2 可扩展的网格信息模型

网格信息模型涉及如何描述、存储和组织资源信息,GISA2.0 提出基于 XML 的信息模型,包含网格信息模板服务(GISS)和网格信息文档管理(GIDM)两部分. GISS 提供了可扩充的资源元数据管理;GIDM 定义了组织网格信息的核心架构,存储和管理各种网格信息,GIDM 管理的网格信息都符合 GISS 中的元数据定义.

2.1 网格资源描述

网格中任何事物都是信息,信息描述必须在可扩展性、自描述性和紧凑性上取得平衡^[3]. 前者是不断增加的网格资源类型的需求,后者是满足系统可扩展的前提. GISA2.0 扩展了 GLUE^[12]实现各种网格资源的模板定义,提出基于 XML/XML Schema 的可扩展元数据管理. XML 是一种元标记语言,可描述任意资源;XML 具有层次结构,适合表达资源间的包含关系;XML Schema 技术可对资源描述严格定义,并通过修改 XML Schema 可灵活地调整信息模型;XML 支持 XPath/XQuery 语言,可对信息进行灵活查询;XSLT 等技术可将 XML 文档以不同方式转化或显示.

网格资源具有多样性和异构性,根据其特征进行分类,为每类资源制定统一的元数据,使网格环境中同类资源信息在语法和语义上得到一致. 对资源空间的分类将形成一个层次分类树,如图 1 所示. 资源之间存在包含、继承等关系,为便于问题研究和系统实现,假设:分类 T ,且对于每一个网格资源 $r \in R$,都有唯一的类型 $T(r)$, $T(r) \in T$,即资源具有唯一类型.

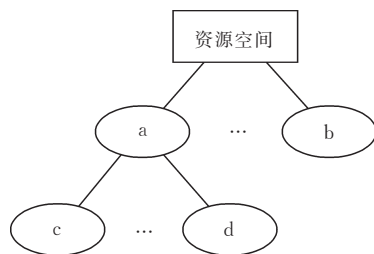


图 1 层次资源分类树

考虑分类在概念上的层次关系,为每个分类指

定全局唯一的层次分类码 (Category)，其对应了资源之间的层次关系。如 CGSP.Computer.Pc 对应个人计算机模板，而 CGSP.Computer.Pc.Hardware 对应个人计算机硬件信息模板。建立层次的信息分类系统，分级管理 Category，有利于提高系统可扩展性和可维护性。统一管理〈Category, Schema〉值对，保证网格资源定义的唯一性，是网格资源有效共享的前提。

注册已定义模板的信息时，用户指定其分类码和资源信息，GIDM 通过 GISS 得到分类码所对应的信息模板，以模板对资源信息进行有效性验证，验证通过方能注册。当注册未定义类型的信息时，必须

先注册其模板，并获取分类码。目前 GISS 提供了域信息框架模板、节点信息模板、通用服务信息模板等。支持运行时添加、修改模板信息，使网格开发人员能方便地扩展模板库，满足不同网格应用的需求。

2.2 资源文档组织

物理上，自治域独立管理其网格资源信息，为了实现逻辑上的全局资源视图，需要对网格信息的组织结构进行统一约定。根据网格信息的包含关系和层次结构，GIDM 将各种信息组织成一个树形的虚拟全局资源文档 GridInformation.xml，其核心架构如图 2 所示。

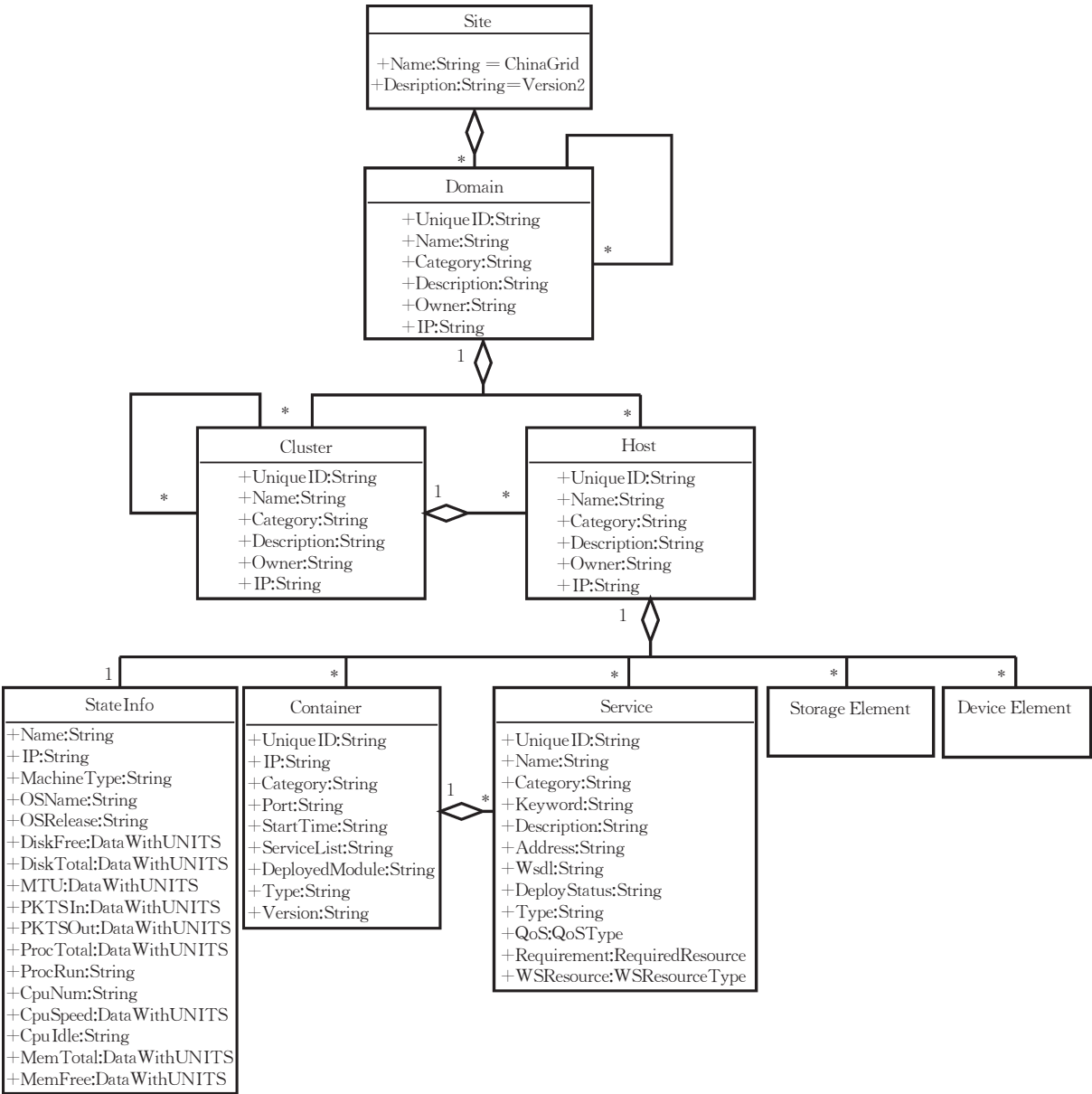


图 2 资源文档组织结构

虚拟全局资源文档的根元素为 Site,其定义与 GLUE 中一致;第二级为 Domain,表示 ChinaGrid 由若干个域组成.任何自治域的网格信息文档实际仅包含单个 Domain 域,即自身域内信息,通过分布查询引擎形成包含多个域信息的虚拟全局资源文档,详见第 4 节.

Domain 包含若干个 Host 元素和 Cluster 元素,Cluster 元素又由 Host 元素组成;Host 元素包含节点状态信息和若干表示资源的元素,如 Service, Container 等.其中 Container 元素包含容器的 IP、端口、启动时间等基本信息以及运行其上的 Service 元素.

注册的资源文档信息必须符合资源模板约束,结合建立全局资源文档和基于索引的快速信息检索(见 5.1 节)的需求,进行以下约定:

(1) 每个注册资源具有全局唯一的 ID 属性,并以此作为资源标志,含有 ID 属性的节点称为资源节点,没有 ID 属性的节点称为属性节点.

(2) 资源节点可以嵌套资源节点和属性节点,属性节点可以嵌套属性节点和文本值.当两个资源存在包含关系,如一个资源在另一个资源中运行,可以方便地利用资源信息的嵌套关系进行选择.

(3) 一个资源文档具有若干设置 primaryKey 属性的节点,表示通过这些节点可以唯一标志该资源,如节点的 IP,服务的 Url 等.

3 层次化信息管理

ChinaGrid 二期中,域的安全、自治性质得到加强,GISA2.0 不直接提供跨域信息聚集,所有的网格信息仅存储在本地域.层次化信息管理分为本地资源管理和自治域信息管理两层,如图 3 所示.本地资源管理完成对网格资源的发现和监控,是整个网格信息服务的基础;自治域信息管理通过维护域的拓扑结构,动态收集所有域信息;信息安全框架提供资源粒度的信息安全.

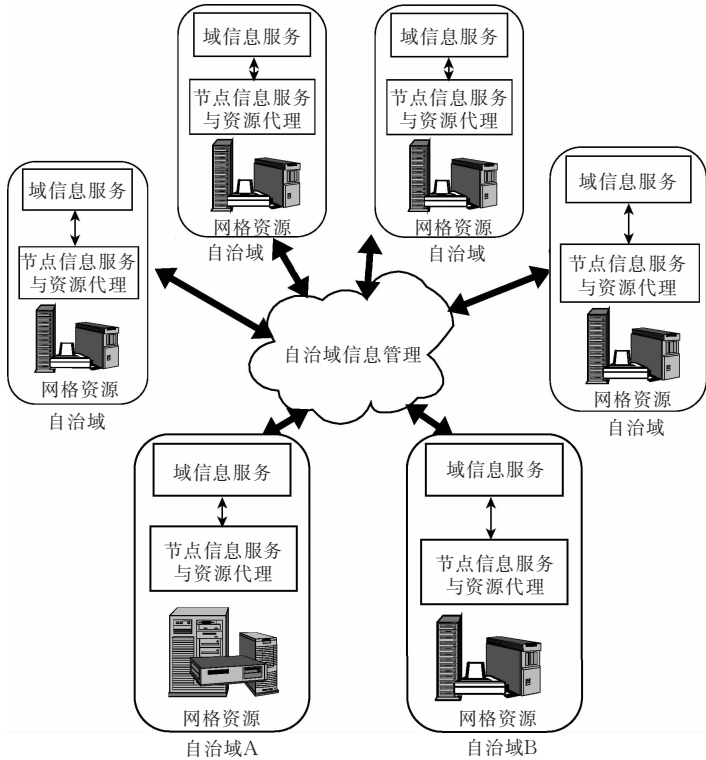


图 3 层次化网格信息服务

3.1 分层模块组成

GISA2.0 遵循 WSRF 和 WS-Notification 规范,采用多层模块化设计,如图 4.

网格资源层为实际的资源提供者,包括各种资源监控服务和异构的网格资源,其具有异构、多样和动态等特点.

资源代理层对资源进行统一表示和提供通用访问接口、各种资源代理封装资源的多样性,完成对信息源的监控,对上层展现为具有共同行为的抽象资源.

核心服务层由一组可独立运行的 WSRF 服务组成,完成信息收集、存储、访问、安全实现、域自身

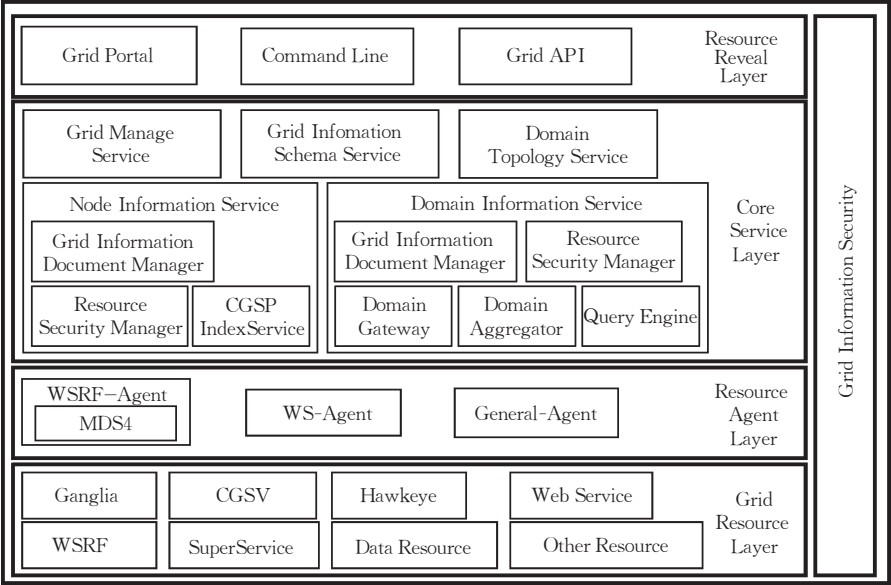


图 4 GISA2.0 分层模块组成

信息维护和跨域访问等功能。

信息展现层描述了信息的展现方式和访问接口,网格应用和开发者可以通过 Portal、命令行方式和 API 进行系统管理和信息检索。

网格信息安全贯穿信息服务的各个层次,安全的信息共享体现在资源访问授权、用户身份认证、权限匹配、跨域身份映射和访问控制等方面。

3.2 本地资源管理

本地资源管理实现通用的资源发现和监控框架,支持多种监控系统接入,如 Hawkeye、MDS4^[4]、Ganglia^[5]、CGSV^[13]等,通过定义其发布信息的 Schema,可充分利用已有监控系统的功能,如图 5 所示。CGSV 独立于 CGSP,为 ChinaGrid 提供监控功能,收集实体(如资源、服务、用户、作业、网络)的状态信息。

网格资源层支持大量分布的信息源,如 WSRF

服务、普通 Web 服务、超级服务^[2]、监控信息、数据资源和特殊设备资源等。

资源代理(Agent)封装不同资源的异构性,完成资源的信息收集,形成符合资源模板约束的 XML 资源描述,并为上层提供统一的访问接口。XML 文档的一部分是资源注册时提交的静态属性,另一部分是动态收集的状态信息。动态信息的监控通过以下代理实现:(1)WSRF 代理采用 MDS4 资源聚集框架,实现 WSRF 服务和资源的聚集;(2)WS 代理收集普通 Web 服务属性,周期检查其状态信息;(3)通用代理原则上可收集任何资源信息,通过配置资源动态信息的相关属性,可周期获取其信息。

节点信息服务(NIS)管理资源代理汇集的信息,生成节点信息表。NIS 包含三个子模块:CGSP 索引服务,提供资源聚集的功能;文档管理负责组织和管理以 XML 表示的资源信息;资源安全管理负

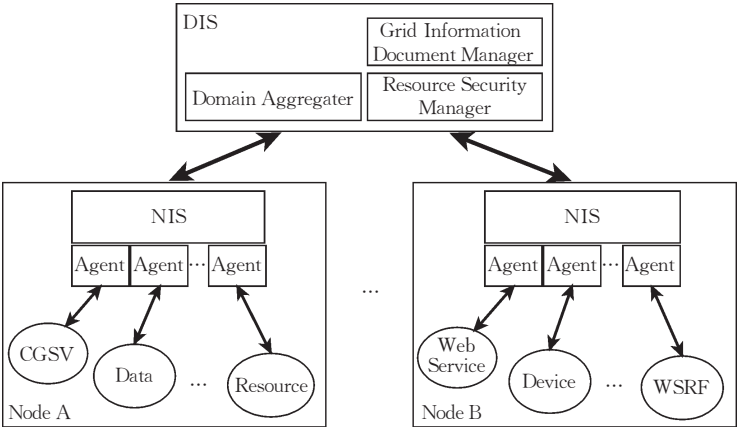


图 5 域内资源管理

责管理资源相应的授权信息. NIS 启动时,生成全局唯一的 ID,并以此作为身份标识,为新注册的资源分配 ID. NIS 信息变化时,采用同步更新或延迟更新的方式向域信息服务(DIS)发送局部更新消息,包括 NIS 的 ID、更新的内容、更新位置以及更新序号. 更新位置为基于节点信息表的 XPath 表达式,更新序号保证有序更新.

DIS 作为信息服务的核心,实现集中的域内网格信息的聚集、管理、安全和检索. DIS 中聚集模块动态收集 NIS 的信息,并通过 NIS 周期发送的心跳信号,维护其可用状态;资源安全管理模块统一管理网格资源授权信息;文档管理模块将域属性信息文档和多个以节点 ID 命名的节点信息文档组织为虚拟的域信息文档. 虚拟域信息文档避免网格信息过多造成单个文档管理效率低下的问题,通过索引机制和多个节点文档的并发访问,可大大提高访问效率.

3.3 自治域信息管理

在 ChinaGrid 中,自治域可动态地加入或退出,自治域信息管理通过域拓扑服务(DTS)维护了树形的域拓扑结构,如图 7(左). DTS 实现域信息的一致视图,其目的是获取全局域信息表,包括信息服务、作业管理器和域管理等地址. 全局域信息表如图 6 所示.

```
<ChinaGrid>
  <Domain ID="NWPU" relation="self" Timestamp="">
    <BasicInfo>
      <Name>name</Name>
      <Address>http://ip:port</Address>
      <DIS>http://ip:port/wsrf/services/DInfoService</DIS>
      <DMS>http://ip:port/wsrf/services/DMSservice</DRS>
      .....
    </BasicInfo>
    <Parent id="">http://ip:port</Parent>
  </Domain>
  <Domain ID="PKU" Timestamp="">... </Domain>
  .....
</ChinaGrid>
```

图 6 全局域信息表

DTS 运行了一个包含域注册、定时更新、注销、超时检测以及相应的应答报文的协议,其语义如下:

- (1) 一个域加入 ChinaGrid 时,向指定的 DTS 发送注册报文,批准后建立逻辑的双向连接,形成父子关系,并交换域信息表.
- (2) DTS 正常退出时,向父 DTS 发送注销报文,父 DTS 更新其域信息表.
- (3) 除根 DTS 外,其它 DTS 定时向父 DTS 发

送 Hello 报文,报告其状态信息.

- (4) 若某 DTS 未在规定时间内发送 Hello 报文,父 DTS 发送超时探测报文,如不能得到响应,父 DTS 则认为该 DTS 发生故障.
- (5) 当 DTS 的域信息表发生变化时,向根 DTS 发送更新报文.
- (6) 根 DTS 将更新报文发给子 DTS,任何 DTS 收到更新报文后,更新自身域信息表,并向其子 DTS 转发.

该协议第(6)条采用逐步同步子 DTS,可减少根 DTS 开销. 在域比较少时,也可由根 DTS 直接同步所有域,加快域拓扑收敛的速度(收敛状态指每个域所保存的拓扑信息完全一致的状态). 新增域的过程如图 7(右)所示.

树形结构的缺点在于根域容易形成单点故障,采用备份 DTS 的方式提高可靠性. 备份服务周期检测主服务状态,一旦主服务失效,备份服务向根 DTS 发送更新消息,接替主服务工作.

树形结构的缺点在于根域容易形成单点故障,采用备份 DTS 的方式提高可靠性. 备份服务周期检测主服务状态,一旦主服务失效,备份服务向根 DTS 发送更新消息,接替主服务工作.

树形结构的缺点在于根域容易形成单点故障,采用备份 DTS 的方式提高可靠性. 备份服务周期检测主服务状态,一旦主服务失效,备份服务向根 DTS 发送更新消息,接替主服务工作.

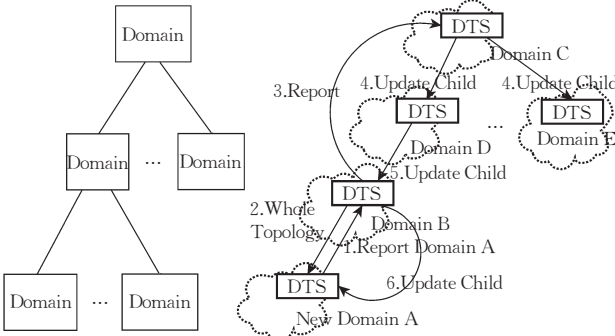


图 7 GISA2.0 域拓扑结构图

3.4 信息安全框架

CGSP 实现了统一的授权管理机制,包括用户管理、授权管理和文件级访问控制. 用户管理提供了身份证书、代理证书管理和身份映射;授权管理基于 GT4.0 实现多级访问控制和动态授权机制;文件级访问控制基于文件元数据信息实现存储资源中的安全.

信息服务的安全包括服务自身安全和网格资源信息安全,服务自身安全由域管理器统一完成. 本文提出的信息安全框架目标是实现网格资源信息安全,包含资源授权描述、授权信息管理、权限匹配等三个方面.

每个资源对应一个资源安全描述文件,其定义了资源授权信息,包括具有访问或更改权限的用户和组信息. 资源安全描述文件基于统一的模板定义,采用 XML 描述各种权限,如图 8 所示.

```
<?xml version="1.0" encoding="UTF-8"?>
<Atomic id="NWPU-Aviation01">
  <WriteRight>
    <group>admin</group>
    <user>owner</user>
  </WriteRight>
  <ReadRight>
    <group>admin</group>
    <group>group1</group>
    <user>IamATester</user>
    <user>user1</user>
  </ReadRight>
</Atomic>
```

图 8 网络资源安全描述文件

资源注册时提交安全描述文件,资源信息和安全描述文件通过资源 ID 关联。NIS 和 DIS 中资源安全管理模块(RSM)统一管理授权信息,并提供管理接口。

信息查询时,将匹配用户信息和安全描述文件,过滤查询结果,实现安全的信息检索。资源访问时,CGSP 容器同样将验证用户信息,保证只有合法用户才能访问资源。

若需要域间交互,必须先设定域互信关系,即通过 CGSP 域管理器进行域间用户映射,以此确定信任邻居域和信任用户。DIS 中的域网关模块维护信任邻居表、域间用户映射表、查询代理以及信息缓存等。进行跨域访问时,域网关模块向远程的信任域 DIS 发送访问请求,远程 DIS 通过身份映射后,进行请求处理,并将结果返回。

4 虚拟全局资源视图

GISA2.0 按图 2 所示的虚拟全局资源文档组

织所有信息,而网络资源信息实际上分散在各个域的 DIS 中。因此,需要一种统一的信息检索机制,网络用户可以像操作一个实际存在的文档一样进行查询,从而得到全局的资源视图。

XML 的查询语言得到广泛研究,如 XML-QL, Xpath, Xquery 等,它们采用正则路径表达式获取 XML 中的结构关系和内容。XPath 是实现 XML 数据周游的基本语言,可构建复杂的查询语句,结合效率和功能因素,GISA2.0 采用其作为统一的查询语言。一般 XPath 引擎执行时,顺序读取 XML 文档中的节点,如果该节点的路径满足 XPath 中定义的条件(包括结构关系、测试条件和谓词关系),则作为查询的一个输出。这种线性处理方式导致必须遍历整个 XML 文档来获取所有满足查询路径的结果,在网络信息量大时,执行效率低下。

本文提出分布式 XPath 查询引擎,通过扩展被广泛应用的 XPath 查询引擎 Jaxen,实现透明的全局信息检索。基于索引的 XML 查询得到广泛研究^[14-17],但对基于索引查询的整体处理框架研究较少,本文提出了一个完整的索引框架,包括基于索引的文档管理和支持索引的 Xpath 查询算法,实现用户相关的虚拟全局资源视图。

4.1 分布式查询引擎

分布式查询引擎是实现虚拟全局资源视图的核心技术,其原理是根据域拓扑管理所获得的域信息服务地址,将其展开为虚拟的包含所有域网络信息的 XML 文档。分布式查询引擎内部模块和查询处理过程如图 9 所示。

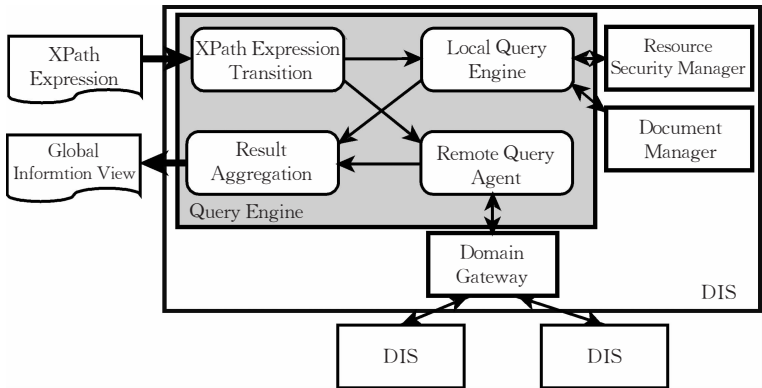


图 9 全局资源视图获取过程

用户提交 XPath 请求后,由 XPath 语句转换模块进行分析。XPath 表达式逻辑上分成两部分,域查询请求和域内查询请求。XPath 语句转换模块将域查询请求转发给远程查询代理,由其决定将向哪些

域的 DIS 发送查询请求。转发的 XPath 表达式经过了 XPath 语句转换器重新构建,查询将只在目标 DIS 执行。

域网关模块为每次跨域查询生成一个查询代

理,其封装了远程域参数和访问请求.查询代理提供同步和异步查询模式,后者以回调接口形式实现.当查询多个远程域时,通过实例多个查询代理,查询时间近似于查询单个外域的时间.

本地查询引擎结合安全管理和文档管理模块,得到符合用户权限的本域结果集(查询算法见 4.3 节).结果聚合模块将各个域查询的结果信息与域信息进行组合,并将结果文档返回给用户.对用户而言,查询的是一个完整的虚拟全局资源文档,得到的是具有访问权限的全局资源视图.

例 1. XPath 表达式: “/Site/Domain/Host[StateInfo[CpuNum>1 and MemFree>500 and DiskFree>2]]//Service[Type='GRS']”.

查询后返回所有域中的 GRS 服务,其条件是部署 GRS 服务的主机必须有多于 1 个 CPU, 500MB 空闲内存和 2GB 的空闲磁盘空间. XPath 语句转换模块根据 “/Site/Domain” 向所有的域发送 “//Host[StateInfo[CpuNum>1 and MemFree>500 and DiskFree>2]]//Service[Type='GRS']” 请求,获取结果后统一返回给用户.

4.2 基于索引的文档管理

网格信息检索包含值查询和结构查询,值查询通过限定元素内容或属性值进行选择;结构查询通过路径表达式对资源间的结构关系进行查询.为了有效支持查询,本文提出一种复合索引方式,包含值索引,如 ID 值、具有 PrimaryKey 属性的节点值以及用户自定义标签的值;关键路径索引,对每个资源节点生成路径索引,如路径 “ChinaGrid/Domain/Host” 将形成以 Host 为单位的一组元素子树,资源节点指含有 ID 属性的节点,见 2.2 节中定义.

另外通过资源 ID 提供资源级别的编码索引,资源 ID 采用前缀编码方式,形成域、节点、资源三层编码结构,其长度分别为 4,6,4,由上层服务保证下层编码分配的唯一性.要判断资源节点 v 是否属于另一资源 u 的后裔,只需要判断字符串 $C(u)$ 是否是字符串 $C(v)$ 的前缀.前缀编码不仅能有效支持包含关系计算,而且能够有效支持位置的关系计算.

以资源节点为单位的索引,类似于目标节点导向的思想^[14],一方面可加快常用信息的查询,另一方面有效控制索引所占用空间. DIS 中的 GIDM 生成和维护各类索引表,并建立和管理索引间的逻辑关系.如图 10 所示.

GIDM 进行 XML 文档解析时,根据模板中的索引信息和用户自定义的索引标签生成索引表,具

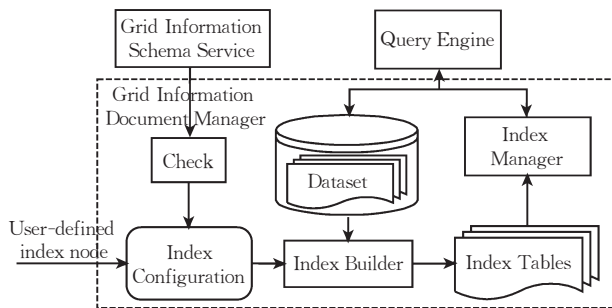


图 10 DIS 的 GIDM 模块框架图

体支持的索引类型包括:

ID 索引表,存储 ID 值与所属元素的对应关系,可加速已知资源 ID,查找资源信息的情况;

PrimaryKey 索引表,存储所属元素的文本值与所属元素的对应关系,如 IP 元素具有 PrimaryKey 属性,索引表中记录了 IP 元素的文本值,即 IP 地址与 IP 元素的一一映射;

关键路径索引表,形成关键路径对应的元素子树 Sub-tree 集.

User-defined K-tag 索引表,即用户自定义关键字段索引表.设定 K-tag 标签以及所属关键路径,建立关键字段值与元素子树的映射关系,如 “Service” 子树中,采用 “Name” 和 “Category” 等作为 K-tag,可建立 Category 值与 Service 子树的对应.全局资源文档对应的索引表实例如图 11 所示.

索引表中 Element 项为指向实际存储位置的引用. ID、PrimaryKey 和 K-tag 索引表,采用 Hash 表存储,而 Sub-tree 索引表以数组形式存储.

4.3 基于索引的查询算法

通常 XPath 查询算法将查询表达式分解为若干个定位步(location step),然后从左到右依次计算.定位步对上下文节点集中的每个节点进行求值,第一个定位步基于初始的上下文节点,计算得到的结果节点集作为下一个定位步的上下文节点集,最后的定位步产生的结果集是查询表达式结果.

一个定位步包含三个部分:一个轴(axis),指定了定位步选择节点与上下文的树状关系;一个节点测试,指定了定位步选择节点的节点类型或节点名称;零个或多个谓词,使用专有的谓词表达式来进一步刷选定位步所选择的节点集合.定位步的计算先从轴和节点测试开始,产生中间节点集;然后利用各个谓词对中间节点集进行过滤,得到结果节点集合.

本文提出基于索引的查询算法,对原有算法的优化体现在三个方面:根据索引表快速计算节点测试,得到中间节点集;根据谓词中出现的已建立索引

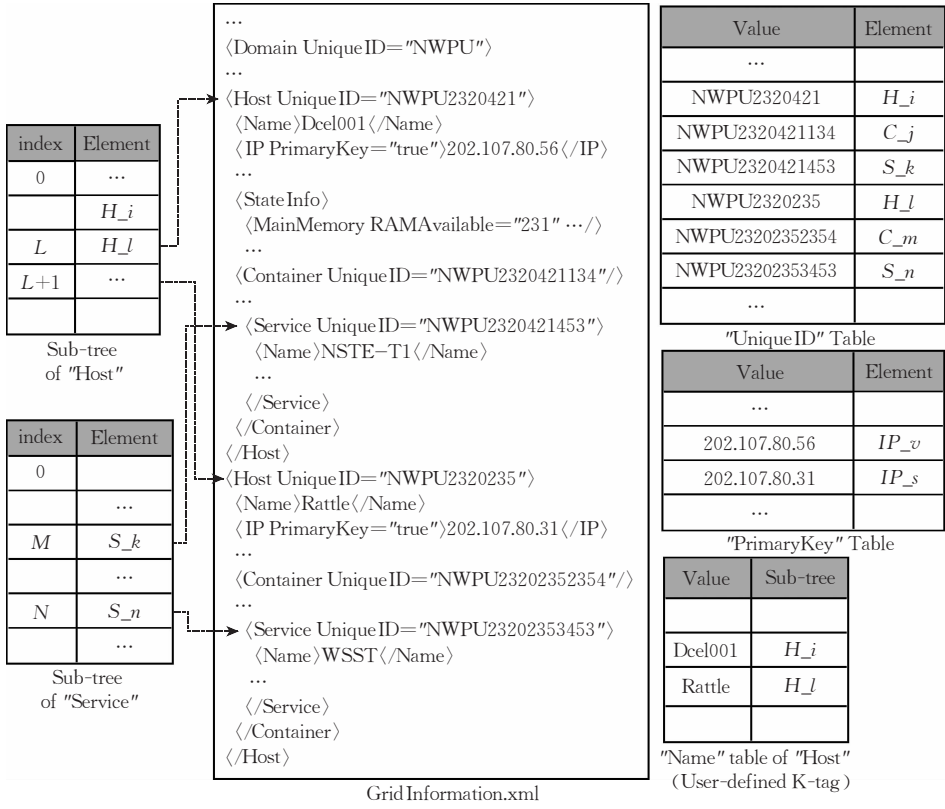


图 11 网络信息索引表

的标签,快速过滤中间结果集;根据用户信息和信息 id 所对应的安全描述文件,快速过滤用户不具有访问权限的结果集。

算法 1. 用户相关的索引查询算法,斜体部分描述了对原有算法的改进。

//输入: 查询表达式 Q , 初始上下文 $initContext$, 用户 id
//输出: 用户具有访问权限的结果集 S
//内部变量: $nextStepContext$ 存放定位步产生的结果,
 $interimSet$ 为定位步中完成轴和节点测试
 后所产生的中间节点集

```
Query-Function( $Q$ ) {
  List  $S = \emptyset$ ;
  XPathReader 解析  $Q$ , 分解为  $N$  个定位步  $step$ ;
  设置  $stepContext$  为  $initContext$ ;
  For each  $step_i (0 \leq i < N)$  of  $Q$  do {
     $nextStepContext = \emptyset$ ;
    For each  $context_j (0 \leq j < M)$  of  $stepContext$  do {
      IF 存在和  $step_i$  节点测试匹配的索引表 THEN
        将节点集赋值给  $interimSet$ 
      ELSE
         $interimSet = X_i.matchAxisNode(context_j)$ ;
        IF  $PredicateSet$  中存在值索引标签 THEN
          快速过滤  $interimSet$ , 并加入  $nextStepContext$ 
        ELSE
          For each  $predicate_k (0 \leq k < K)$  of  $PredicateSet$ 
```

```
do {
   $nextStepContext.add(Predicate_k, valueate(interimSet))$ ;
}
}
Set  $stepContext = nextStepContext$ ;
}
 $S = nextStepContext.getNodeList()$ ;
For each node  $b$  of  $S$  do {
  Check 用户  $id$  with  $b$  的安全描述文件
  IF 非授权用户 THEN
    将  $b$  排除出  $S$ 
}
return  $S$ 
}
```

4.4 两种 XPath 引擎性能比较

DOM4J 是一个非常优秀的开源代码的 Java XML API,已广泛应用于 Java 软件开发. 其默认以 Xerces 遍历文档,以 Jaxen 执行 XPath 查询. 通过对其进行扩展,包括生成索引表和支持带索引的查询,我们将新的 API 称为 IBP4J.

对 XML 查询的性能通常取决于文档特征,如标签和数据比率,属性使用程度^[18],元素子树数目以及平均元素子树大小等. 本文采用现实的网格资

源文档进行测试,文档结构如图 2. 文档大小分别为 121KB, 460KB, 1.1MB, 4.66MB, 10.2MB, 20.1MB, 它们包含了不同数量的资源信息. 然后量化比较采用 DOM4J 和 IBP4J 初始化文档和执行 XPath 查询的时间开销.

所显示的计时结果是来自使用带有 512MB RAM 的 P4 2.4GHz 的 PC 机,操作系统为 Redhat Linux9,运行 JDK1.4.2,初始 JVM 最大内存大小为 128MB.

在图 12 中,由于 IBP4J 在 DOM4J 的基础上,增加了根据索引配置生成索引表的过程,增加了大约 10%的时间开销.

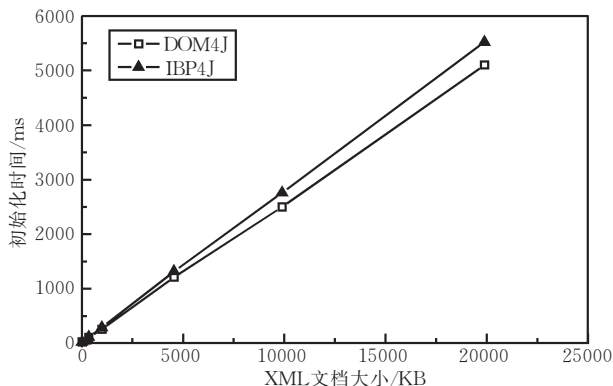


图 12 初始化文档时间比较

抽取实际应用中的 XPath 查询语句,随机查找各种网格资源信息,在不进行语法缓存的情况下查找 1000 次,平均解析时间反映了 XPath 执行的整体性能,如图 13 所示.

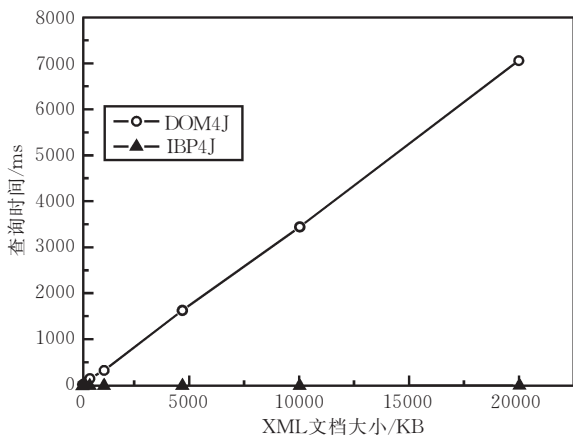


图 13 XPath 查询时间比较

曲线说明 DOM4J 中 XPath 引擎的查询时间随文档大小的增加递增,而 IBP4J 由于建立了所有资源节点为单位的路径索引,可快速定位到网格资源. 其查询开销为索引定位到资源节点的开销和资源内部查询开销,而单个资源信息量有限,因此能获得稳

定的查找性能.

5 相关研究及评价

为了实现 DataTag, iVDGL, Globus 等多个网格的互连,DataTag 项目中的 GLUE^[12]定义了一个抽象的信息模型来描述计算机系统对象、属性、行为和关系. GLUE 重在实现网格节点之间的交互性,定义了计算资源、存储资源等核心信息,但在具体的网格系统中,需要考虑更广泛的资源,因此, GISA2.0 扩展 GLUE 实现各种网格资源的模板定义. 另外, DMTF 的 CIM^[19]定义了一个面向对象与平台无关的标准,用于系统、网络应用、服务和厂商开发的管理信息的公共定义. CGS WG 扩展了网格作业提交服务模型, CRM WG 扩展了描述类似 OGSA 服务的可管理资源模型.

网格中资源发现和监控体系得到了广泛研究. MDS4^[4]广泛应用于采用 GT4.0 的网格系统中,它采用 XML 表示资源,并遵循 GLUE 规范,以逐层聚集的方式为服务状态的收集提供一种清晰和易于实现的机制. MDS 中的网格资源指以 WSRF 封装的服务资源,资源属性向上聚集时,包含所有未变化信息,导致无关的数据量很大,资源聚集的方式没有考虑域自治性. 由于 MDS4 实现了 WS-Notification 机制,并提供了灵活的信息收集,因此, GISA2.0 以 WSRF 代理封装 MDS4 的功能完成 WSRF 服务的信息收集.

Ganglia^[5]是一个开源的层次化监控系统,信息的发布和收集采用组播的方式,通过联邦集群, Ganglia 可用于广范围的信息收集. 但其主要针对集群设计,只能通过静态配置实现集群间交互. MonALISA^[6]采用 DDSA 体系,实现了基于 Jini 的可扩展的大规模分布系统的主机和网络监控框架,由于 Jini 采用组播方式,可扩展性受到一定限制. R-GMA^[8]使用了 GMA 中的基本概念,其采用关系数据库保存网格信息,生产者、消费者使用 SQL 语句来进行注册和数据获取,传输信息采用 XML 格式. GridRM^[9]包含全局层和本地层,通过网关封装资源的异构性,以 GMA 实现网关间的信息通信. 另外, PyGMA, AutoPilot, DMF, TopoMon 等基于 GMA 实现了不同程度的资源监控. 以上监控系统在虚拟组织的自治性、网格信息的安全性等方面考虑不足,而本文提出的层次资源监控框架,可集成多种已存在的监控系统,从而具有更好的灵活性和扩展性.

近期,将 P2P 技术应用于大规模网络成为一个热点^[10,21-22],其避免了由于信息聚集造成的性能瓶颈. 超级节点模型^[10]中,超级节点对应了网络中一个管理域,多个超级节点通过高层的 P2P 网络互连,主要研究了 P2P 网络中的成员管理和资源发现服务. OSGA^[21]实现不同服务发现机制和协议(如 SLP, UPnP, Jini 和 UDDI 等)之间的信息共享. 提出了统一服务描述 USD,在域上构建覆盖层(Overlay),采用 P2P 的方式实现不同域之间的服务共享,其中域的概念指单一服务发现协议所管理的服务和用户. 本文的自治域信息管理在域拓扑结构相对稳定,域数目较少时,与 P2P 相比具有简单高效等特点,跨域信息访问并不受限于域拓扑结构,可根据需要采用灵活的访问机制.

纯 XML 数据库得到了广泛研究,研究者已提出了 XML 数据的各种索引技术^[14-17]和编码方案^[22-23]. 与通用的数据库相比,而本文的 XML 信息的存储和管理与网格信息管理框架相结合,在节点信息服务和域信息服务可进行资源信息编码、并采用一致的存储架构等,因此可加快信息检索效率.

6 结论和展望

本文针对多自治域的大规模网格环境,提出了网格信息服务体系 GISA2.0. 充分考虑了自治域的自主管理和信息访问的安全性,实现了层次化信息管理和用户相关的虚拟全局资源视图. GISA2.0 采用 XML 描述、存储资源信息,并采用信息模板服务,支持各种网格资源注册,具有良好的可扩充性;建立了可扩展的域资源管理框架,灵活收集域内各种资源信息;强化了域自治特性和网格信息安全,提供域间共享机制,实现了安全的全局信息共享;优化了 XPath 查询引擎,实现了安全、分布和快速的信息查询.

后续的工作是基于 P2P 实现域拓扑维护,并与层次化管理方式进行性能对比;对邻居域建立信任度函数,能自动地调整域间信任关系;目前分布 XPath 查询的实现依赖于 XPath 语句转换模块,后续的研究将以资源链接节点的形式,在文档中包含远程文档的连接信息,直接实现远程查询操作.

参 考 文 献

- [1] Ian Foster, Carl Kesselman, Steven Tuecke. The anatomy of the grid enabling scalable virtual organizations. *International Journal of High Performance Computing Applications*, 2001, 15(3): 200-222
- [2] Hai Jin. ChinaGrid: Making grid computing a reality//*Proceedings of the 7th International Conference of Asian Digital Libraries*. Shanghai, China, 2004: 13-24
- [3] Serafeim Zanikolas, Rizos Sakellariou. A taxonomy of grid monitoring systems. *Future Generation Computer Systems*, 2005, 21(1): 163-188
- [4] Schopf J, Raicu I, Pearlman L et al. Monitoring and discovery in a Web services framework: Functionality and performance of Globus Toolkit MDS4. Argonne National Laboratory, Chicago, USA: Technical Report MCS-P1315-0106, 2004
- [5] Massie M L, Chun B N, Culler D E. The ganglia distributed monitoring system: Design, implementation, and experience. *Parallel Computing*, 2004, 30(7): 817-840
- [6] Newman H, Legrand I, Galvez P, Voicu R, Cirstoiu C. MonALISA: A distributed monitoring service architecture//*Proceedings of the 2003 Conference for Computing in High Energy and Nuclear Physics(CHEP03)*. La Jolla, California, USA, 2003: 1-8
- [7] Tierney B, Aydt R, Gunter D et al. A grid monitoring architecture. GGF Performance Working Group: Report GWD-Perf-16-3, 2002
- [8] Andrew Cooke, Gray Alasdair, Ma Lisha et al. R-GMA: An information integration system for grid monitoring//*Proceedings of the International Conference on Cooperative Information Systems (CoopIS 2003)*. Catania, Sicily Italy, 2003: 462-481
- [9] Baker M, Smith G. GridRM: An extensible resource monitoring system//*Proceedings of the IEEE International Conference on Cluster Computing*, Portsmouth University, UK, 2003: 207-214
- [10] Carlo Mastroianni, Domenico Talia, Oreste Verta. A super-peer model for resource discovery services in large-scale grids. *Future Generation Computer Systems*, 2005, 21(8): 1235-1248
- [11] Wu Yong-Wei, Wu Song, Yu Hua-Shan et al. CGSP: An extensible and reconfigurable grid framework//*Proceedings of the 6th International Workshop on Advanced Parallel Processing Technologies (APPT2005)*. Hong Kong, China, 2005: 292-300
- [12] Andreozzi S, Burke S, Field L et al. GLUE schema implementation for the LDAP data model. Istituto Nazionale di Fisica Nucleare, Padova, Italy: INFN Technical Report INFN/TC-04/16, 2004
- [13] Zheng Wei-Min, Liu Lin, Hu Mei-Zhi et al. CGSV: An adaptable stream-integrated grid monitoring system//*Proceedings of the International Conference on Network and Parallel Computing(NPC2005)*. Beijing, China, 2005: 22-31
- [14] Wang Jing, Meng Xiao-Feng, Wang Yu, Wang Shan. Target node aimed path expression processing for XML data. *Journal of Software*, 2005, 16(5): 827-837(in Chinese)
- [15] Ian Foster, Carl Kesselman, Steven Tuecke. The anatomy of the grid enabling scalable virtual organizations. *International*

(王静, 孟小峰, 王宇, 王珊. 以目标节点为导向的 XML 路径查询处理. 软件学报, 2005, 16(5): 827-837)

- [15] Kong Ling-Bo, Tang Shi-Wei, Yang Dong-Qing, Wang Teng-Jiao, Gao Jun. XML indices. *Journal of Software*, 2005, 16(12): 2063-2079(in Chinese)
(孔令波, 唐世渭, 杨冬青, 王腾蛟, 高军. XML 数据索引技术. 软件学报, 2005, 16(12): 2063-2079)
- [16] Florescu D, Kossmann D, Manolescu I. Integrating keyword search into XML query processing. *Computer Networks*, 2000, 33(1): 119-135
- [17] Chan C Y, Felber P, Garofalakis M, Rastogi R. Efficient filtering of XML documents with XPath expressions. *The International Journal on Very Large Data Bases*, 2002, 11(4): 354-379
- [18] Matthias Nicola, Jasmi John. XML parsing: A threat to database performance//*Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM'03)*. New Orleans, LA, USA, 2003: 175-178
- [19] Bumpus Winston, Sweitzer John W, Thompson Patrick,

Westerinen Andrea R, Williams Raymond C. *Common Information Model: Implementing the Object Model for Enterprise Management*. Indianapolis, US: John Wiley & Sons, 1999

- [20] Noura Limam, Joanna Ziembicki, Reaz Ahmed et al. OSDA: Open service discovery architecture for efficient cross-domain service provisioning. *Computer Communications*, 2007, 30(3): 546-563
- [21] Wei Jie, Hung Terence, Turner S J, Cai Wen-Tong. Architecture model for information service in large scale grid environments//*Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06)*. Singapore, 2006: 107-114
- [22] Wang Wei, Jiang Hai-Feng, Lu Hong-Jun et al. PbiTree coding and efficient processing of containment joins//*Proceedings of the 19th ICDE*. Bangalore, India, 2003: 391-404
- [23] Torsten Grust. Accelerating XPath location steps//*Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*. Madison, Wisconsin, 2002: 109-120



ZHANG Hai-Hui, born in 1977, Ph.D. candidate. His research interests include distributed computing and grid computing.

ZHOU Xing-She, born in 1955, professor, Ph.D. supervisor. His research interests include distributed computing and embedded computing.

ting and embedded computing.

YANG Zhi-Yi, born in 1952, professor, Ph.D. supervisor. His research interests include distributed computing and embedded computing.

WU Xiao-Jun, born in 1972, Ph.D., lecturer. His research interests focus on distributed computing.

YANG Gang, born in 1974, Ph.D., lecturer. His research interests focus on distributed computing.

Background

This paper conducts research on the information service model for large-scale grid. Many researches have been done on grid information service, but they generally have not taken into consideration of the autonomy issue of virtual organization management. Most of the past research has taken the approach of direct information aggregation. Some of the researches use general information to represent interconnecting among multiple grids, but few of them study information management with autonomy requirement within grid.

The GISA2.0 presented in this paper focuses on the characteristics of large-scale environments, such as heterogeneity, diversity, dynamic, and autonomy, supports autonomous management of information service, secure information access, consistent global view, and fast information retrieval.

The research in this paper is part of the CGSP(China-Grid Support Platform) project, which is a core sub-project of ChinaGrid programme.

ChinaGrid is the largest grid projects in China. The goal

of ChinaGrid project is to integrate various resources distributed in CERNET(China Education and Research Network), providing high quality services to scientific research and study. CGSP provides a set of development tools for those China-Grid application developers; it has been successfully applied in the development of campus grid applications in more than twenty universities.

In the first phase of CGSP, the research team presented GISA1.0, which implemented efficient layered information aggregation, but paid little attention in security and information retrieval.

The research content presented in this paper is applied in the second phase of CGSP. Each campus grid is an autonomous ChinaGrid domain, with its own resource share scheme, and the mechanism to process interaction with other domains. CGSA2.0 can provide information management support for multiple-domain grids.