

面向存储资源管理的多协议存储系统

颜秉珩¹, 钱德沛^{1,2}

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 北京航空航天大学计算机学院, 100083, 北京)

摘要: 针对网格工作流中存在的跨域数据共享问题, 提出了一种基于存储资源管理的多协议支持的存储系统——MPSS. 利用管理 URL 来帮助传输服务区分数据类型; 通过将不同的用户映射到不同的私有空间, 实现访问空间的隔离; 把握控制读写令牌的授权来协调多个传输服务之间的合作关系. 这样, 透明地实现了跨域的数据共享, 极大地简化了网格工作流的执行过程, 提高了网格数据共享的效率. MPSS 支持常见的传输协议, 在实际的航空、药物网格以及气象等项目中表现出了良好的稳定性和效率. 压力测试表明, MPSS 至少可以处理每秒 300 次的访问, 平均响应时间在 1.7 s 左右.

关键词: 存储资源管理; 跨域; 多协议; 数据共享

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2009)06-0010-05

SRM-Based Multi-Protocol Supported Storage

YAN Bingheng¹, QIAN Depai^{1,2}

(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. School of Computer Science, Beihang University, Beijing 100083, China)

Abstract: An SRM-based multi-protocol supported storage (MPSS) is proposed to solve the data share problem in a grid workflow environment of cross-domain. The MPSS utilizes managed URL to distinguish different data types, insulate users' private space through mapping different users' space to different physical position, and coordinate different transfer services by controlling authorization of read/write tokens. With these strategies the MPSS implements the data sharing of cross-domain transparently, simplifies the grid workflow's running and improves the efficiency of grid data sharing. The MPSS supports popular transfer protocols, provides a powerful supporting for grid workflow of cross-domain, and shows a good stability and efficiency in practical projects such as aviation, DD grid and meteorology. The results of pressure test show that the MPSS can process at least 300 requests per second with an average response time 1.7 s.

Keywords: storage resource manager; cross-domain; multi-protocol; data sharing

网格工作流通过若干网格服务、协同工作可完成一个特定的目标, 在现有的网格基础设施环境下可将工作流模型实例化. 在这个过程中, 网格服务通过数据共享连接起来, 完成对共享数据的一系列操作, 可见跨域的数据共享是实现网格工作流的基础之一. 图 1 展示了一个飞机设计的分布式优化流程, 它要用到 GRIA^[1] 和 GOS 2 种网格设施上的服务:

网格工作流执行前先被解析为一系列的网格服务调用, 然后通过元调度将任务依次调度到合适的工作结点再执行, 数据共享在这里起到了串联多个网格服务的作用.

跨域数据共享的主要困难在于网格设施之间数据传输协议有所不同, 例如欧洲网格设施 gLite 主要使用 GridFTP 协议, GRIA 使用 HTTPS 协议,

而 CNGrid 的网格中间件 GOS 主要使用 FTP 协议,所以说不同的数据访问协议是当前网格设施之间完成数据共享任务的障碍。

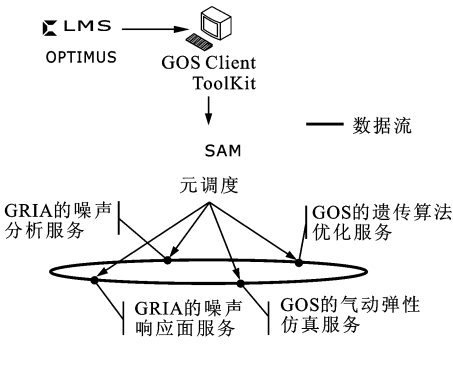


图 1 飞机设计的分布式优化流程

为了支持跨域的数据共享,提高网格工作流的执行效率,本文提出了一种基于存储资源管理 (SRM) 的多协议存储系统 (MPSS),在不改动上层网格基础设施的情况下,它可以透明、高效地完成跨域的数据共享任务。

1 跨域数据共享方案及存在的问题

在网格服务之间进行数据共享,最主要的是完成跨协议的数据访问,解决方案一般有 2 种形式:①直接访问,这需要访问站点和存储站点至少支持一种相同的协议;②间接访问,其先将数据迁移到可以直接访问的存储站点,然后再使用数据。显然,直接访问比间接访问在效率上更具优势,但是除了效率之外,还必须考虑方案的通用性和扩展性。

适配器模式 (例如 GridFTP DSI) 是一种流行的支持直接访问的跨域数据共享方案,它将存储站点原有的传输服务封装为访问站点可以直接访问的服务,以此屏蔽协议之间的差异,从而完成数据的跨域透明访问。这种方案一般具有较好的数据共享效率,但是扩展性较差,即在一个涉及 n 个不同传输协议的环境中,往往需要 $n(n-1)$ 个适配器才能实现跨协议的数据访问。

存储-转发方案是典型的间接访问模式,具有很多优点,例如可以针对数据共享任务的特点提供缓冲、调度以及错误恢复等功能,便于形成独立、通用的模块,提高网格工作流的并行性^[2]。

网格存储管理工作组 (GSMWG) 制定的 SRM^[3] 协议为异构的网格存储设施之间的互联互通奠定了必要的基础,它提供了一个通用的 SRM

接口,包括空间的申请和文件的管理等。SRM 独立于存储系统的传输协议,为跨协议数据共享提供了动态的协议协商机制:在访问站点的传输协议与存储站点的存储协议之间可进行协商,并选取一个双方均支持的传输协议完成数据传输任务。SRM 服务与传输服务有所不同,它们之间的关系如图 2 所示。

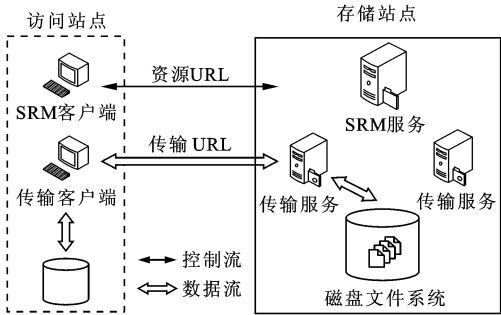


图 2 SRM 客户端访问站点与 SRM 服务端存储站点的交互关系

由于 SRM 本身不涉及到具体的数据传输工作,因此它并不能直接完成跨域的数据共享。如果存储站点同时支持多种传输协议,SRM 就能发挥它的动态协商功能,访问站点就能通过 SRM 获取所需的传输 URL (例如 FTP 形式的 URL),完成数据访问^[4]。遗憾的是,当前能全面支持网格设施的 SRM 协议 (既支持 SRM 服务端存储站点,又支持 SRM 客户端访问站点) 很少,递进地关系到一个存储站点支持多种传输协议的情形更是极为少见。

综上,跨域数据共享的困难主要在于网格所用的数据传输协议有所不同。为了解决这一问题,我们设计了一种简单、高效的数据共享方案——支持多协议的存储系统 (MPSS),它既具有直接访问模式的高效性和透明性,还具有很好的扩展性。采用 MPSS 的数据共享方式如图 3 所示,从中看出这种以存储为中心的数据共享方式,大大提高了网格服务中的数据访问速度,并简化了网格工作流的处理过程。

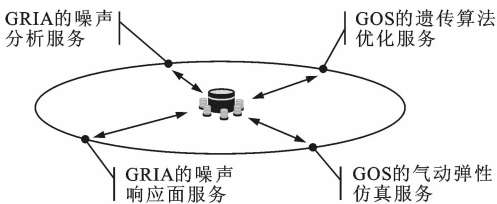


图 3 采用 MPSS 的数据共享方式

2 MPSS 设计

虽然 SRM 无法独立完成跨域的数据共享,但是它给数据访问带来了以下便利:①无论站点支持的传输协议有多少,存储站点上的文件都可以用惟一的源 URL(SURL)来标识,具体的传输 URL(TURL)可以在运行时通过动态转换来获得;②SRM 服务通过 SRM 接口和内部的数据访问机制可协调多个传输服务,以完成对同一个文件系统的访问。

MPSS 在继承了这些优点的同时,还进行了必要的扩展,克服了通过 SRM 进行跨域数据访问存在的问题。这些扩展主要包括:对 TURL 的直接支持、对用户私有空间的隔离以及传输服务之间采用协作机制。

2.1 对 TURL 的直接支持

访问站点通过 SRM 访问存储站点,可以看作是一个通过 SURL 获取 TURL 的过程。MPSS 为了减少对 SRM 的依赖,实现了透明的数据共享,扩展了对传输协议的直接支持,也就是访问站点可以直接通过 TURL 来访问数据。

如前所述,传输服务与 SRM 服务是松耦合的,传输服务并不知道 SRM 服务是否存在,事实上访问站点通过 SRM 访问数据时也无需知道具体的传输服务。但是,MPSS 为了直接支持传输协议,就需要区分普通数据和共享数据,只有那些需要跨域共享的数据才有接受 MPSS 协调管理的必要。MPSS 用 MURL(Managed URL,以/srm/home为根路径的一种特殊的 TURL)来帮助传输服务区分数据的类型,即用 MURL 访问共享数据,否则访问普通数据。下面例子的前 3 个 TURL 都是 MURL,最后一个 SURL,均指向存储在 MPSS 上的同一个共享文件 data.share:

```
ftp://user:pass@host:2121/srm/home/data.share;  
https://host:8043/srm/home/data.share;  
gsiftp://host:2811/srm/home/data.share;  
srm://host:8443/srm/v2/server?SFN=~/  
data.share.
```

此处 URL 中的路径都是逻辑路径,MPSS 是通过将 MURL 或者 SURL 映射到特定的物理路径上,来控制传输服务时的数据可见性。MPSS 对 TURL 的直接支持,使得网格服务可以通过 MURL 直接访问共享数据,因而大大提高了数据共享的透

明性和效率。

2.2 对用户私有空间的隔离

MPSS 在完成跨域数据共享的同时,还保证了共享数据的安全性,即只有合法的用户才能访问相应的共享数据。MPSS 为每个用户维护一个私有的存储空间,该用户的所有文件都存储在这个私有的空间内。MPSS 通过将不同的用户映射到不同的私有空间,实现了访问空间的隔离。

MPSS 采用了 2 种方式来确定访问者的用户标识:①采用用户名/密码的方式来支持 FTP 形式的 MURL;②采用证书的方式来支持 GridFTP 和 HTTPS 形式的 MURL。MPSS 通过分析 MURL 携带的信息就可以确定访问用户及其访问的私有空间:①FTP 形式的 MURL 通过登陆名确定访问的用户空间;②HTTPS 和 GridFTP 形式的 MURL 通过携带的证书信息确定访问的用户空间。

事实上,我们建立了证书、登录名与 MPSS 用户之间的映射关系:FTP 登录名与 MPSS 用户是一对一的,而证书 DN(Distinguished Name)与 MPSS 用户是多对一的。值得注意的是,数据传输的安全性由访问站点偏好的协议来保证,因为 MPSS 无法改变传输协议的固有属性,只是负责自身存储数据的安全。

2.3 传输服务之间的协作

若允许用户通过多个传输服务来并发地访问一个文件系统,就必须解决服务之间的协作问题。例如,同一个数据文件上的并发读/写操作是不允许的。MPSS 通过控制读写令牌的授权来协调多个传输服务之间的合作关系,并规定传输服务在读写共享文件之前必须申请相应的访问令牌,即读令牌或者写令牌,操作完成后释放令牌,如果申请令牌失败,那么操作将不能进行。因为存储在 MPSS 上的共享文件都可以用一个惟一的 SURL 标识,而且读写令牌都是针对 SURL 的,所以 MPSS 为每个 SURL 维护一个授权的令牌队列。对于一个 SURL 的读写令牌的申请,MPSS 使用如下机制进行授权:①写令牌只有在 SURL 上的令牌队列为空时才能授权;②读令牌只有在 SURL 上的令牌队列全为读令牌时才能授权。

通过控制令牌的授权机制,MPSS 可以协调多个传输服务之间的合作关系,保护共享数据的一致性。

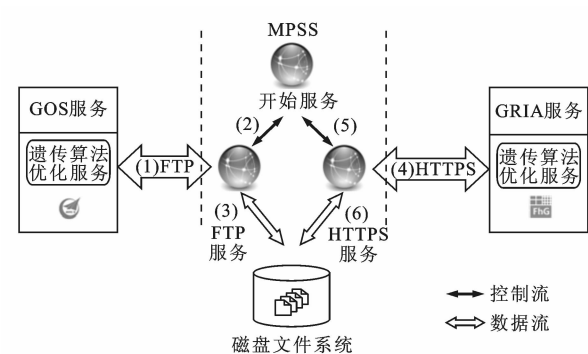
2.4 传输服务的修改

规范的传输服务不能识别 MURL,运行在

MPSS 上的存储服务需要进行修改以支持上述的几种扩展.在此,本文给出了修改传输服务的关键代码,如图 4 所示,主要涉及登陆和读写操作.登陆部分主要完成 MPSS 用户空间的映射,而读写部分主要完成 3 个功能:①MURL 的识别;②相关令牌的申请;③MURL 到物理路径的映射.

```
LOGIN handler // called when a new session constructed
begin
    call old LOGIN handler ;
    if (LOGIN is successful) then
        // mapping to SRM user for locating SRM space
        MPSS- user = mapping (credential or name)
        change home directory to MPSS -user's directory
    endif
end
READ handler//called when read data from MPSS
begin
    if (URL is NOT a MURL) then
        call old READ handler
    else
        file=get physical file WRT URL
        read-token=acquire read-token of file
        if (read-token is valid) then
            READ data from file
            release read-token
        endif
    endif
end
WRITE handler//called when write data to MPSS
begin
    if (URL is NOT a MURL) then
        call old WRITE handler
    else
        file = getphysical file WRT URL
        write-token=acquire write-token of file
        if (write-token is valid) then
            WRITE data to file
            release write-token
        endif
    endif
end
end
```

图 4 修改传输协议的关键代码



(1):遗传算法优化服务使用 FTP 形式的 MURL 写入数据;(2):FTP 服务向 MPSS 申请写令牌;(3):GOS 节点使用 FTP 协议写入数据并在完成后释放写令牌;(4):噪声响应面服务使用 HTTPS 形式的 MURL 读取数据;(5):HTTPS 服务向 MPSS 申请读令牌;(6):GRIA 节点使用 HTTPS 协议读取数据

图 5 使用 MPSS 在 GOS 和 GRIA 之间实现数据共享

修改后的传输服务可以实现通过 MURL 对共

享数据的直接访问,实现图 3 所示的数据共享模式.图 5 展示了 GOS 和 GRIA 之间使用 MPSS 完成数据共享的详细过程,此时 GOS 和 GRIA 上的节点可以直接使用自己的协议访问数据,而无需任何改变.

3 实验与分析

MPSS 大大提高了跨域数据共享的效率,作为一个独立功能的模块,它可以方便地进行部署.从图 3 中可以看出,MPSS 位于数据共享的中心,面临着巨大的访问压力,因此有必要对 MPSS 进行压力测试,并对测试结果进行分析.

3.1 测试指标

目前,MPSS 支持 3 种传输协议:FTP、Grid-FTP 和 HTTPS.在飞机、药物网格设计中,观察到 MPSS 的以下几个特点(见表 1).

表 1 飞机、药物网格设计中 MPSS 的特点

特点	类型
最频繁的操作	文件列表、数据上传/下载
高峰期并发连接数	≈150
最耗费时间/资源的操作	数据上传/下载

不同协议的高峰期并发连接数略有不同,实际运行中观察到的最大并发连接数是 150 (不同传输协议连接的总和),访问类型主要是获取文件列表、写数据和读数据,数据规模在几十个 kB 到几十个 GB 不等,这是 MPSS 面临的主要压力.依据这些特点,我们设计了如下的测试指标^[5].

(1)并发成功率.针对不同数目的并发连接,计算 MPSS 完成操作的成功率.

(2)吞吐量,即

吞吐量 = 并发量 / 全部操作完成时间

对于上传/下载来说,操作完成时间是指用户发出请求到开始上传/下载数据的时间.

(3)响应时间.响应时间指用户从发出请求到等待回应结果的平均时间,对于上传/下载来说,是指用户发出请求到开始上传/下载数据的时间.

3.2 实验环境及结果分析

按照以上内容,在实验室内搭建了一个压力测试环境,如图 6 所示,它以百兆级局域网连接.图 6 还列出了相应的测试系统平台和测试内容,背景显示了压力测试过程中并发量的变化情况.

依据设计需求,将并发连接数设定在 100~300

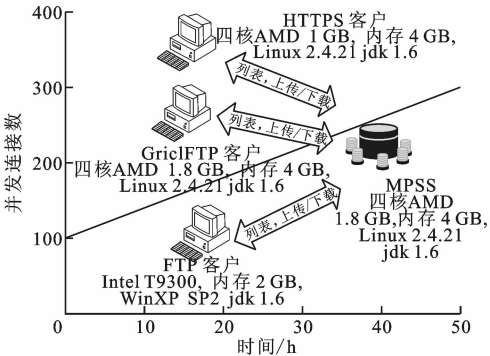


图 6 实验环境和测试中并发量的变化情况

之间,这是 3 种客户端总的并发数目.按照每小时增加 4 个并发连接数的变化率,完成一次测量需要 50 h,如图 6 背景所示,经过多次测量可获取实验结果的平均数据.结果表明,所有并发成功率均为 100%,MPSS 表现出了良好的稳定性.

图 7 是 MPSS 的吞吐量随并发量变化的情况.曲线是一次实验的结果,经多次实验后发现,吞吐量走势大致沿着“+”线发展,说明在 300 个并发访问的时候尚未到达系统处理能力的上限.图 8 是 MPSS 的响应时间,经多次测试发现,MPSS 的平均响应时间维持在 1.7 s 左右,达到了实际所需的处理速度.总的来看,MPSS 完全有能力为现有网格服务提供稳定、高效的跨域数据共享服务.

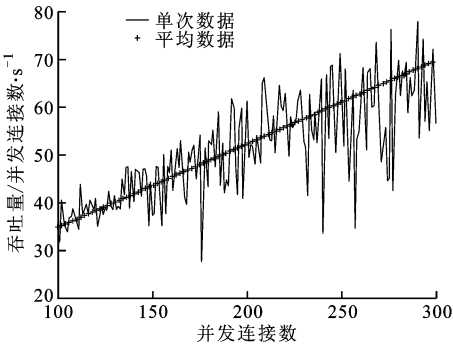


图 7 MPSS 吞吐量的变化情况

4 结束语

本文设计、实现了一种面向 SRM 的多协议存储系统——MPSS.它以SRM服务为基础,对外支

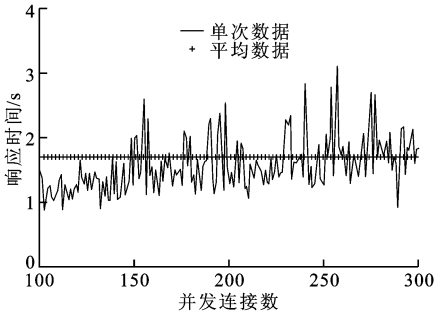


图 8 MPSS 的响应时间

持多种协议,可以有效地完成跨域的数据共享,提供了最佳的数据访问透明性,提高了网格工作流的执行效率,简化了网格服务之间的协作. MPSS 已经成功地应用于气象、药物网格以及航空项目之中,实际运行以及实验结果表明,MPSS 具有较高的稳定性.

参考文献:

[1] SURRIDGE M, TAYLOR S, ROURE D, et al. Experiences with GRIA - industrial applications on a web services grid[C]// 1st International Conference on e-Science and Grid Computing. Los Alamitos, CA, USA: IEEE Computer Society, 2005:98-105.

[2] 颜秉珩,钱德沛.一种基于网关的网格数据互操作框架[J].西安交通大学学报,2009,43(2):20-24.
YAN Bingheng, QIAN Depei. A gateway-based framework for data interoperation between grids[J]. Journal of Xi'an Jiaotong University, 2009, 43(2):20-24.

[3] Scientific Data Management Research Group. Storage resource manager[EB/OL]. [2008-05-21]. [http://sdm. lbl. gov/srm-wg](http://sdm.lbl.gov/srm-wg).

[4] YAN Bingheng, WU Zhongxin, YANG Dongbo, et al. Experiences with the EUChinaGrid project-implementing interoperation between gLite and GOS [C]// The 2nd IEEE Asia-Pacific Service Computing Conference. Piscataway, NJ, USA: IEEE, 2007: 224-231.

[5] 宾建伟. 压力测试 [J]. 电子产品可靠性与环境试验, 2008, 26(1):48-51.
BIN Jianwei. Pressure test [J]. Electronic Product Reliability and Environmental Testing, 2008, 26(1): 48-51.

(编辑 苗凌)