
基于 SMI-S 的网络存储系统实时监控与管理系统的实现

高岩 孟承运 刘晓光 王刚

南开大学计算机系, 天津, 300071

monciel@163.com, mcy19851217@yahoo.com.cn, liuxguang@gmail.com, wgzwp@163.com

Implementation of Real-time Monitoring and Management System for Network Storage System Based on SMI-S

Gao Yan, Meng Chengyun, Liu Xiaoguang, and Wang Gang

(Department of CS, Nankai University, Tianjin, 300071)

Abstract With the rapid development of storage technology, the capacity and scale of network storage system are growing. This requires that storage management systems must be of the high manageability, maintainability and interoperability among different management softwares are needed. SMI-S, an interface standard of storage management, is proposed by SNIA. It makes storage management can be vendor-independent, and also reduces management cost, raise the management efficiency. In this paper, we design and implement a storage management system which is based on SMI-S standard and WBEM/CIM framework. This system can achieve real-time monitoring and remote management of the remote hosts, snapshot systems and remote copy systems well although there are many differences between these managed systems. This storage management system has many functions such as real-time dynamic monitoring, remote command transmission, logging, error information statistics, alarming and so on. And users can get all information and results of this management system by a graphic interface through web directly. It improves the operability and maintainability of this system. Test results show that this system can run stably and be well-behaved, afford an available solution for the research on real-time monitoring and remote management of storage systems, also for storage management of enterprise customers.

Keywords SMI-S; WBEM/CIM; storage management

摘要 随着存储技术的高速发展, 网络存储系统容量和规模不断扩大。这就要求存储管理系统必须具备高可管理性、可维护性, 并且不同管理软件之间要具有互可操作性。SMI-S (Storage Management Initiative Specification) 是 SNIA (Storage Networking Industry Association) 提出的存储管理的接口标准。它使存储管理可以真正实现与厂商无关, 从而降低管理成本, 提高管理效率。本文以基于 SMI-S 标准的 WBEM/CIM 框架为基础设计并实现了一个存储管理系统, 该系统能够在屏蔽异构性的情况下, 实现对远端主机和快照系统和远程复制系统的实时监控和远程管理。该系统具备了实时动态监控、远程命令传输、日志记录、错误信息统计和预警报警处理等功能, 并且以基于 Web 的图形化界面将获取的各种信息和分析结果直观地展现给用户, 提高了系统的可操作性和可维护性。测试结果表明这个管理平台运行稳定, 功能良好, 为存储系统实时监控和远程管理技术的研究以及企业级用户的存储管理提供了一种可行的解决方案。

关键词 SMI-S; WBEM/CIM; 存储管理

随着制造工艺的进步和网络技术的发展, 存储系统容量从 GB 发展到现在的 TB, 正在向 PB 和 EB 发展。它为计算机和网络技术带来蓬勃生命力的同时也使得存储管理面临着巨大的挑战: 首先是兼容性问

题。早期的存储管理软件都是由硬件厂商提供。而不同厂商的管理软件往往不兼容, 这增加了系统的管理难度。如何降低管理的复杂性以及提高管理软件之间的互可操作性, 有效地管理不同厂商提供的设备已经

基金项目: 论文工作得到国家自然科学基金(60903028), 国家“863”计划(2008AA01Z401), 教育部博士点基金(20070055054)和天津市科技发展计划(08JCYBJC13000)资助。

成为设计和实现存储管理系统的重大课题。除此之外，存储系统的容量和规模的增大也会带来的可管理性和可维护性问题。随着系统中设备数量和型号的不断增多，系统维护人员的工作量也将逐渐增加。同时，系统的设备运行状况、设备输入输出数据流量、交换网络传输流量等直接反映了系统的工作效能，维护人员也必须及时了解系统的整体状况，使系统运行在最佳状态。但由于设备数量大，种类繁多，当系统出现故障时，很难在极短的时间内定位故障点。综上所述，为存储系统管理建立一个集中统一的远程管理和实时监控平台已成为一个迫切的需求。

本文的工作是基于 SMI-S 标准的 WBEM/CIM (Web-Based Enterprise Management/Common Information Model) 平台设计并实现了一个存储管理系统，该系统能够在屏蔽异构性的情况下，实现网络存储系统以及快照和远程复制等模块的实时监控和远程管理。该系统具备了实时动态监控、远程命令传输、日志记录、错误信息统计和预警报警处理等功能，并且以基于 Web 的图形化界面将获取的各种信息和分析结果直观地展现给用户，提高了系统的可操作性和可维护性。

1 相关工作

目前主流的系统管理体系结构有 SNMP/MIB[1], DMI/MIFI[2], CMIP/GDMO[3]等。虽然它们在各自的领域都有不错的表现，但在管理包括 IP 网和电信网的混合网络方面都存在一定缺陷。SNMP 是目前 IP 网络管理中的事实标准，但它的信息建模是面向过程的，并且具有事件上报机制有限等缺点；DMI 则更多地关注桌面管理，对大型复杂网络的支持有限；CMIP 虽然比 SNMP 的管理功能更为强大，但却由此带来了系统实现复杂性、用户操作不便捷等缺点而未被广泛接受。

存储业界也一直在向标准化存储管理接口的方向努力，存储管理接口规范 (Storage Management Initiative Specification, SMI-S) 是由美国存储网络工业协会[4] (Storage Networking Industry Association, SNIA) 提出的。SMI-S 标准在存储网络中的存储设备和管理软件之间提供了标准化的通信方式，其核心技术就是 WBEM/CIM 框架，遵守这个规范的产品都可以通过统一的应用程序接口 (API) 来进行管理，使存储管理真正实现了厂商无关性。WBEM/CIM 是一些领先的软硬件厂商如 Microsoft, Intel 等，基于 SMI-S 标准联合提出的了基于网络的企业管理系统架构[5]，对不同软硬件环境下的管理方式进行了统一。WBEM 采用面向对象的方法，通过多种网络管

理协议 (SNMP、DMI、CMIP 等) 获取资源信息并进行标准化处理，使这些信息能够以单一的协议形式出现。因此 WBEM 可以用来集成不同管理平台提供的各项服务，有力地加强了系统综合管理的能力。

2 网络存储系统实时监控和远程管理系统设计

2.1 系统架构

存储管理系统应该具备以下监控和管理功能：忠实地记录系统的运转状态，进行实时的动态监控，帮助管理员对网络和系统环境进行充分的了解，防患于未然；通过远程控制等手段，避免重复性的工作，快捷地解决用户的问题；及时报告系统出错信息，并提供相应的分析资料和建议，便于故障的分析和解决；将复杂的监控结果和管理手段用简洁直观的形式表示。根据以上的目标，我们基于 CIM[6] XML 接口的 WBEM 综合体系结构原型设计并实现了一个综合存储管理系统，系统框架如图 1 所示。

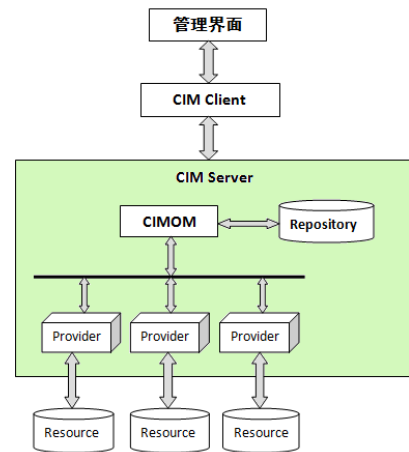


图 1 存储管理系统框架图

该系统由管理界面，CIM Client 和 CIM Server 构成。其中 CIM Server 和 CIM Client 均存在于被管对象资源上，管理界面则是一个基于 B/S 结构的图形化界面。

CIM Server 是整个存储管理系统的核心，主要包括 CIMOM 和 Provider 两个部分。CIMOM 有一个专属的数据库 Repository，负责存放静态数据。Provider 是真正与被管理的对象资源进行交互的部件。本文的存储管理系统中涉及的受管系统主要包括受管设备所在主机的操作系统、非写前拷贝快照系统[7]和远程复制系统[8]，并且可以方便地对受管系统进行扩展。

CIM Client 位于 CIM 架构中的上层，主要负责将用户请求及相关操作 XML 封装，通过 HTTP 传送给 CIMOM，同时将 CIMOM 返回的结果传递给客户端。在本系统中，一个 Client 对应一个 Provider 并对其调用。

2.2 系统功能设计

存储系统实时监控和远程管理的设计主要体现在以下功能中：实时动态监控功能、远程命令控制功能、日志记录功能、错误信息统计功能、预警报警功能和数据分析功能，这些功能都将通过管理界面展现给用户，存储管理系统的纵向管理模块如图 2 所示。

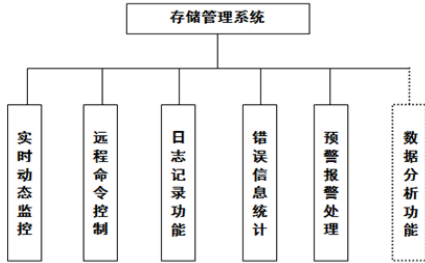


图 2 存储管理系统的纵向管理模块图

目前设计的存储系统实时监控和远程管理可以应用在任意主机、非写前拷贝快照系统和远程复制系统这几个存储系统上，以后可以根据用户的需要对横向管理模块进行扩充，可以很方便地加入其它的受管设备或存储系统。

2.3 CIM 框架设计

(1) CIM Server 的设计

CIM Server 开发的核心工作是针对不同受管系统的功能设计实现相应的 Provider 接口。CIM Provider 是完成资源操作请求的实际部分，根据所设计的各项具体功能实现相应的 Provider 程序，以动态链接库的形式被 CIM Server 调用，并提供给远端的 CIM Client。

由于远程复制系统和快照系统都是在 LVM2[9] 上实现的，是内核模块的扩展，而 Provider 多为用户态程序，如何高效地进行两者之间的交互是一个需要解决的问题。

/proc 文件系统[10]是一个虚拟文件系统，它只存在于内存中，而不占用外存空间，专门用于内核和用户态程序的信息交互。/proc 文件系统动态从系统内核读出所需信息并提交给用户。操作系统内核中的 CPU、进程等各种信息都可以通过 /proc 文件系统直接访问。对于内核的不同版本，/proc 中的信息位置是一致的，因此这种方法具有良好的可移植性。本系统的 Provider 就是通过 /proc 文件系统获得受管对象的信息。

(2) CIM Client 的设计

CIM Client 是具有 CIM 交互能力的客户端程序，向 CIM Server 提交 CIM Operation 请求并接收 CIM Operation 响应，通过 CIM provider 程序访问和操作 CIM 模型对象。存储管理系统中，一个 Provider

对应一个 Client。CIM Client 和 CIM Server 的通信如图 3 所示。

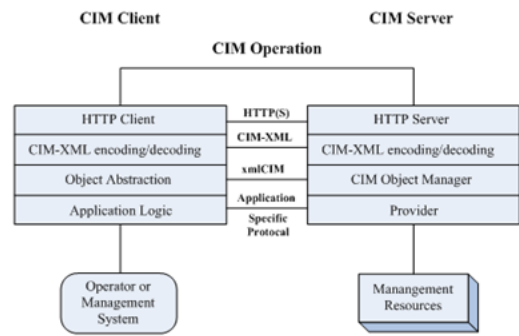


图 3 CIM Client 和 CIM Server 的通信图

3 网络存储系统实时监控和远程管理系统实现

3.1 远端主机实时监控功能的实现

针对远端主机的实时监控和远程管理主要体现在实时查看主机操作系统信息、CPU 信息和使用率、内存信息和使用率以及硬盘信息等。

(1) 远端主机 CIM Server 实现

根据 2.3 所述，CIM Server 的实现需要做的具体工作就是实现相应的 Provider 接口。对于 Provider 的开发我们采用 CIMPLE2.0.0，开发环境为 SuSE2.6.25.5，CIM Server 平台为 OpenPegasus2.8.1[11]。

利用 CIMPLE 实现 Provider 主要有以下几个步骤：

- 1) 设计相关的 CIM Class。CIM Provider 的操作对象必须先在 CIMOM 中定义好相应的 CIM Class，用 MOF 文件对 CIM Class 进行描述。
- 2) 使用 CIMPLE 工具生成代码框架。
- 3) 填写 CIM Provider 的框架代码，完成具体的功能。
- 4) 编译工程。
- 5) 注册模块。

3.2 快照系统自动扩容监控的实现

本文介绍的存储管理系统针对快照系统的监控和远程管理具体体现在可以查看快照系统的 CPU 使用率、内存使用率、物理空间使用率，并且在出现异常情况时进行预警，系统发生错误时报警，还实现了快照系统的自动扩容功能。其中对快照物理卷的自动扩容功能既需要实时监控物理空间使用率又必须在使用率达到阈值时远程发送自动扩容命令，是存储管理系统实时监控和远程管理功能相结合的最佳体现。

(1) 快照系统自动扩容方法

为提高磁盘空间利用率，快照系统采用了按需分配的机制。数据卷初始分配的实际的物理空间可以小

于用户需求的逻辑空间大小。随着数据的增多,可用的物理空间将渐渐减少,系统应适时地自动为其增加一定的物理空间以满足用户的实际需求。为了实现数据卷的物理空间与逻辑空间的分离,我们设计了两层卷的结构。在此基础上,设计并实现了对物理空间及逻辑空间自动扩容的解决方案。对每个卷来说,都有一个守护进程对磁盘空间使用率进行监控,当该使用率达到阈值时将自动启动扩容进程。较之人工扩容,这种自动物理空间扩容技术有效地避免了溢出。对于逻辑空间的扩容,我们采用了索引数组的方式,可以方便地管理分散的元数据区,并且将原有卷与新增部分有效地组织在一起。

(2) CIM 框架实现方法

存储管理系统对快照系统物理卷自动扩容的实现主要涉及两个具体功能:实时获取物理卷的空间使用率和远程发送自动扩容命令,因此管理界面需要实时动态地显示空间使用率的曲线变化图表,管理系统针对上述两项功能分别实现了相应的 CIM Client 和 CIM Provider。

Snap_full_provider 是用来获取快照系统物理空间使用率的 provider 程序,它可以通过/proc 接口读取空间使用率。Snap_full_client 是用来发送获取空间使用率的 CIM Operation 请求的 CIM Client。Snap_extent_provider 和 Snap_extent_client 分别是管理系统为远程管理快照系统进行物理卷自动扩容所实现的 CIM provider 和 CIM client。

3.3 远程复制系统监控及管理功能的实现

对于远程复制系统,存储管理的主要目的是监控其状态,并通过管理界面来调整远程复制系统地参数。远程复制的主要参数是初始的组间隔和一个组中的请求数目。远程复制系统会在创建卷的时候根据设置的值进行自适应调整,达到由用户远程控制受管复制系统传输规模的目的。这一过程是通过把管理界面的相关参数传递给远程复制系统的 CIM Client 和 CIM Provider 完成的。

RC_parameter_provider 是用来向远程复制系统传递参数的 provider 程序,目前只需要传递初始的组间隔时间和请求数目,随着功能的扩展,需要传递更多参数时,可以直接在这个 provider 程序里添加。

RC_parameter_client 是用来完成向远程复制系统传递参数的 CIM Operation 请求的 CIM Client。先连接 CIM Server,然后向 Server 发送请求并接收响应。

4 测试及结果分析

本文对存储系统实时监控和远程管理功能进行了较为全面的测试。测试对象分别为存在于

SuSE2.6.25.5 上的主机以及运行在 RedHat AS server 5 (内核版本是 2.6.18-128.el5)上的快照系统和远程复制系统,并使用 LVM2 2.02.39, device mapper 1.02.28 和 NBD 2.8.8。测试服务器为戴尔 Power Edge Server, 3.20GHz Intel Xeon CPU, 3GB 内存。实验用 IOMeter[12]来产生负载。

4.1 远端主机

存储管理系统获得的受管主机硬盘信息的界面。远端主机其它信息可以通过界面左侧所示的相关链接获得,比如:远端主机的 CPU 信息和使用率,内存信息和使用率,操作系统信息等。如图 4 是远端主机的硬盘信息显示界面。



图 4 远端主机的硬盘信息显示界面

4.2 快照系统自动扩容

实验的源卷大小为 1G,快照容量初始为 100M,阈值设置为 80%,每次比例到达阈值后自动扩容源卷的 10%即 100M 的空间,所以当快照数据到达 80M 后,快照容量自动扩充到 200M,比例由 80%下降到 40%,当对源卷进行写操作使得比例又达到 80%,即快照数据为 160M 时,快照容量再一次自动扩充到 300M,比例由 80%下降到 53.3%,如此不断地进行自动扩容后。图 5 所示,图中界面上方显示启动自动扩容这项命令被成功执行,界面中间是物理空间使用率的动态实时显示图表,下方是检测时间和相应的检测数据记录。



图 5 快照系统自动扩容功能的显示界面

4.3 远程复制系统批次间隔传输

存储管理系统对远程复制系统的实现如图 6 所

示, 界面左侧为实时监控和远程管理的具体实现, 主要是查看 CPU 使用率, 查看内存使用率, 查看相关参数 (延迟数据), 命令操作以及错误信息记录等。以后对远程复制系统进行功能扩展只需要在左侧继续添加具体的功能即可。图 6 展示的是远程复制系统各项命令操作的界面, 如: 加载 dm-rs 模块, 加载 lvm 模块, 加载 nbd 模块, 建立 nbd 连接, 创建源卷等, 命令执行成功或者失败都会有相应的提示, 图中表示的即为成功加载 dm-rs 模块。



图 6 远程复制系统各项命令操作的界面

4.4 性能测试

(1) 远端主机

测试的方法是选取任意一台主机, 首先分别针对以下六种情况测试主机的 CPU 使用率及内存使用率: 未启动 CIM 框架时、启动 CIM 框架并且每隔 2 秒提出一次请求、启动 CIM 框架并且每隔 1 秒提出一次请求、启动 CIM 框架且每隔 0.5 秒提出一次请求、启动 CIM 框架且每隔 0.2 秒提出一次请求以及启动 CIM 框架且每隔 0.1 秒提出一次请求。

存储管理系统对受管主机 CPU 使用率的影响如图 7 所示。未启动 CIM 框架时 CPU 使用率为 2.1%, 启动 CIM 框架并且每隔 2 秒提出一次请求后, CPU 使用率上升到 2.5%, 几乎未受影响。当分别每隔 1 秒, 0.5 秒, 0.2 秒, 0.1 秒提出一次请求, CPU 使用率也一直都是缓慢上升, 上升率在可以接收的范围内, 说明存储管理系统对受管主机 CPU 使用率的影响比较小。

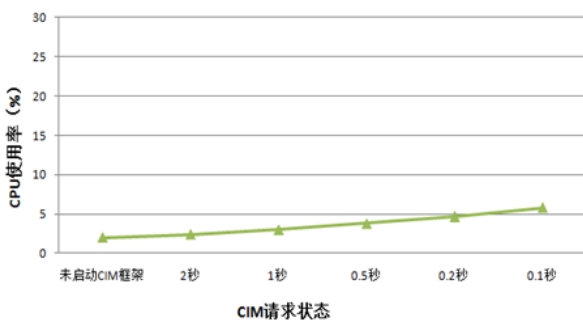


图 7 存储管理系统对受管主机 CPU 使用率的影响

同样, 从图 8 中可以看出, 存储管理系统对受管主机内存使用率的影响几乎可以忽略, 当启动 CIM 框架并且每隔 0.1 秒提出一次 CIM 请求时, 内存使用率也只是由未启动 CIM 框架时的 95.1% 上升到了 95.8%。

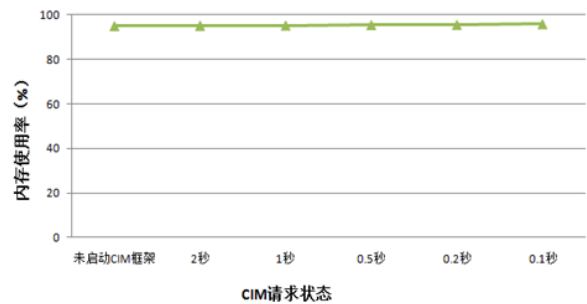


图 8 存储管理系统对受管主机内存使用率的影响

最后本文针对未启动 CIM 框架、启动 CIM 框架并且每隔 2 秒、1 秒提出一次请求三种情况测试了主机文件系统磁盘写速度。测试结果如图 9 所示, 三组数据的值相差不大, 磁盘写速率没有因为 CIM 框架的应用而受到明显影响。

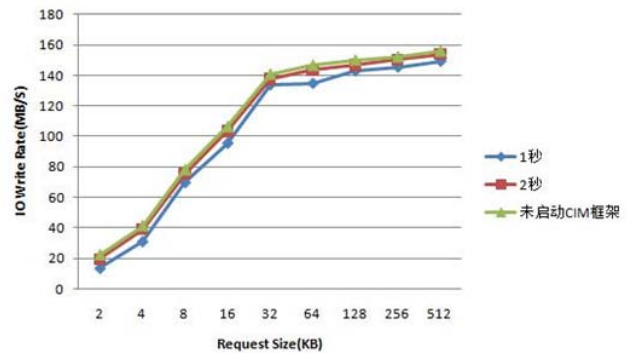


图 9 存储管理系统对受管主机写速度影响

(2) 快照系统自动扩容

测试分三组, 每组创建的卷的大小分别是 1G、10G 和 100G。在每组实验中, 进行了十次扩容, 测试在数据卷大小相同的情况下, 扩容不同增量时所用的扩容时间, 增量取值分别是 10%、20%、30%、...、100%。对于第一组来说, 第一次扩容后, 数据卷将增加 1G*10% 的空间, 测试结束后, 将卷销毁并重建一个 1G 的数据卷, 然后按照 20% 的增量对其扩容, 以此类推。其他两组也是同样的测试方法。

实验结果如图 10 所示, 三组数据的值相差不大, 不会因为卷容量或扩容增量的不同而有明显的变化, 几乎是常量, 而且这个值很小。由此可见, 利用本文介绍的存储管理系统进行数据卷的自动扩容速度很快, 用户不会因扩容而暂停正常业务。

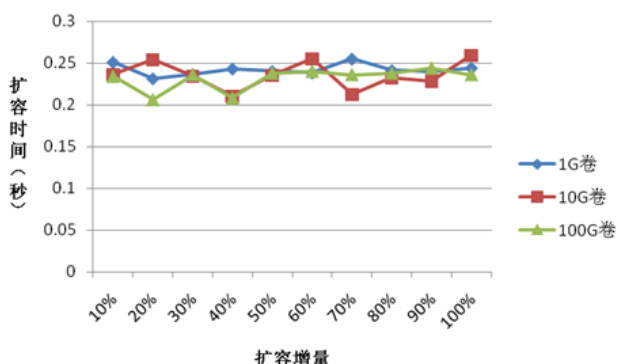


图 10 快照系统自动扩容性能测试

(3) 远程复制系统

实验是在 2.66GHz 的英特尔双核 Xeon 节点下进行的。一个作为主站点，另一个作为备份站点。每台机器有 2GB 的内存和由 6 个 37GB 的 SAS 硬盘组成的 RAID-5 硬件，两个节点由千兆以太网相连。

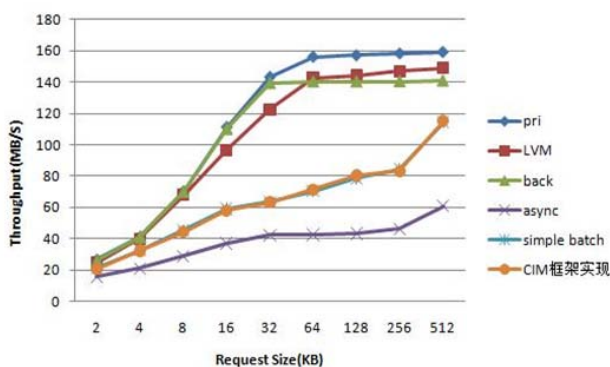


图 11 远程复制系统批次间隔传输基准测试

图 11 显示了实验结果。其中“pri”代表主站点中的 RAID-5 设备，“back”代表备份站点中的 RAID-5 设备，“LVM”代表主站点的源卷，“async”代表没有自适应成组算法的异步远程复制系统，“simple batch”代表有自适应成组算法的异步远程复制系统，“CIM 框架实现”代表利用存储管理系统实现远程复制系统批次间隔传输的实验结果。从图中六组数据的情况可以看出，利用存储管理系统实现的批次间隔传输和远程复制系统本身运用自适应成组算法实现批次间隔两者的性能几乎一致，而且效果很好。

5 结束语

本文基于 SMI-S 标准和 WBEM/CIM 框架设计和实现了一个存储管理系统。该系统能够屏蔽底层操作系统，实现对远端主机、快照系统和远程复制系统的实时监控和远程管理。测试结果表明本文描述的存储管理系统功能良好，运行稳定并且对受管设备或系统的性能影响很小，可以进行实际的应用，也为存储系统实时监控和远程管理的研究提供了一个可行的解

决方案。

在今后的工作中，我们将在如下几个方面进一步努力：扩展实时监控和远程管理的受管存储系统规模；扩充存储管理系统的监控和管理功能，增加数据分析，环境监控等措施；提高存储管理系统的安全性能等。

参考文献

- [1] Stallings W. SNMP and SNMPv2: The Infrastructure for Network Management. IEEE Communications Magazine, 1998-03
- [2] Desktop Management Interface Standards. <http://www.dmtf.org/standards/dmi>
- [3] CMIP ITU-T X.710-ISO/IEC 9595. Common Management Information Service Element. 1995
- [4] Storage Networking Industry Association (SNIA). ISO/IEC 24775:2007. SNIA Storage Management Initiative Specification. International Organization for Standardization (ISO). 2007
- [5] Andrea Westerinen, Jim Davis. WBEM Standards. DMTF 2002 Developers' Conference, 2002
- [6] W. Bumpus, J. W. Sweitzer, et al. Common Information Model: Implementing the Object Model for Enterprise Management. Wiley, 2000
- [7] 谢广军. 虚拟化存储系统关键技术研究: [博士学位论文]. 天津: 南开大学, 2009
- [8] 葛广贺. 基于 LVM2 和 NBD 的远程复制系统设计与性能优化: [硕士学位论文]. 天津: 南开大学, 2009
- [9] LVM2 user manual. <http://sources.redhat.com/lvm2>
- [10] 卢宁, 易雅鑫, 何锐等. 基于 WBEM 的系统管理技术研究. 电脑应用技术, 2007, 70: 13~15
- [11] 祁曼丽. 一个开放标准的网络存储系统管理框架设计与实现: [硕士学位论文]. 天津: 南开大学, 2009
- [12] <http://www.iometer.org>
- [13] OpenPegasus. <http://www.openpegasus.org>
- [14] CIMPLE. <http://cimple.org>

高岩, 女, 1986 年生, 硕士研究生, 研究方向: 快照、非写前拷贝快照系统的自动扩容、网络存储系统。

孟承运, 女, 1985 年, 硕士研究生, 研究方向: 网络存储系统。

刘晓光, 男, 1974 年生, 副教授, 研究领域: 并行计算、网络存储。

王刚, 男, 1974 年生, 教授, 研究方向: 并行计算、网络存储。