一种基于非对称结构的虚拟存储网络的实现*)

刘朝斌　谢长生　黄建忠
(华中科技大学计算机学院外存储系统国家重点实验室　武汉430074)

摘　要　全球范围内数据量的飞速增长,对存储系统提出了更高的要求,网络存储越来越受到人们的关注。目前流行的网络存储技术是NAS和SAN。NAS和SAN各有自己的体系架构、互联协议、文件组织和管理方式等。而将这两种技术融合到一个统一的结构框架下已经成为目前人们关注的热点问题。本文利用存储虚拟化技术,采用非对称结构模型,实现了一种融合NAS和SAN的虚拟存储网络。

关键词　附网存储,存储区域网,存储虚拟化

The Realization of Virtual Storage Networking Based on Asymmetric Architecture

LIU Zhao-Bin　XIE Chang-Sheng　HUANG Jian-Zhong
(National Storage System Laboratory, Huazhong University of Science and Technology, Wuhan 430074)

Abstract　The explosive growth of global data is dramatically driving the demand for storage system. Network storage has been giving more and more attentions. Now the two most popular ones are NAS (Network Attached Storage) and SAN (Storage Area Network). Both the NAS and SAN have themselves architecture, interconnected protocol, file organization and manage method etc.. Recently, it has become the trends to merge them into universal storage architecture. In this paper, based on the storage virtualization, we have realized one asymmetric virtual storage network prototype to merge the NAS and SAN.

Keywords　Network attached storage, Storage area network, Storage virtualization

1. 引言

随着全球经济与网络的飞速发展,企业数据量飞速增长,对计算机的存储需求越来越高。而在现在的计算机系统中,芯片处理能力和网络传输能力都得到了很大的提高,这就更加导致数据存储越来越成为计算机发展的瓶颈[1,3]。另一方面,单位容量硬件存储成本不断降低,因此,从上世纪90年代末开始,存储尤其是网络存储,逐渐走上了计算机系统设计的前台,成为业界研究的热点[2],而许多公司和企业也开始以数据为中心布置业务。

传统的本地直接存储(DAS)的方式是存储设备专属于某个服务器,数据之管理附属于某个主机的控制之下,服务器的网络传输和数据处理能力是数据利用的制约因素。而且如果服务器因为某种原因无法工作(这种情况在实际应用中经常发生),则整个存储数据将无法使用,因此,这种传统存储方式已远远不能满足企业高可用性、可扩展性和集中统一管理数据的需求,在这种情况下网络存储技术应运而生[10]。网络存储技术的两个典型代表是附网存储(NAS)和存储区域网(SAN),它们都在各自的领域得到了很大的应用和发展。NAS是面向文件级应用的,是可以直接联到网络上向用户提供文件级服务的存储设备[5,6]。而SAN则有所不同,它是一种专用的存储网络体系结构,目前一般是用光纤通道技术互连,SAN中存储设备和应用服务器之间的存储I/O一般是以数据块I/O的方式进行[7,9]。虽然NAS和SAN各有自己的体系架构、互联协议、文件系统和管理方式等,但是它们并不是互相排斥的关系。对于不同的数据应用和业务需求,它们也可以互为补充,整合在一个统一的系统框架下。但是对于同时拥有配置了这两种类型设备的企业来说,则很难把它们统一起来。本文研究的虚拟存储技术就是统一NAS和SAN技术的有效途径,通过一种非对称结构模型,设计了一种基于成熟TCP/IP技术的虚拟存储网络的原型系统。

2. 非对称结构模型设计

2.1　存储虚拟化技术原理

在典型的存储网络环境下,由于历史的原因,各种设备往往呈现出异构性:首先是经常会有不同类型的服务器,如Windows、UNIX和Linux等;其次是不同提供商的存储设备,即便是同一厂商的产品,往往也会有不同的性能等物理参数的差别;最后,FC,ESCON,iSCSI,SSA,Infiniband等存储网络的接入方法和互连协议,也有很大不同。

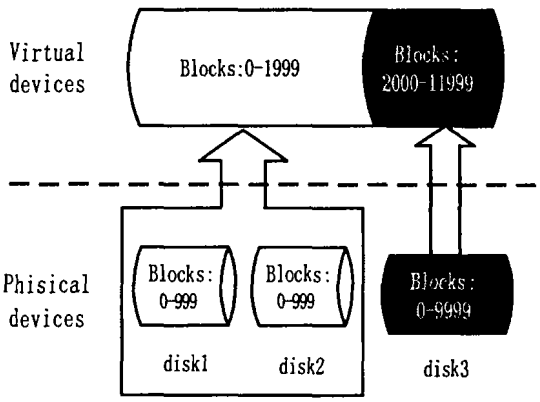

图1　存储虚拟化原理

存储虚拟化的核心工作是物理存储设备到单一逻辑存储

*)基金项目:本项目得到国家自然科学基金(No. 60173043)资助。刘朝斌　博士生,主要研究方向是新型存储系统与结构、高性能存储网络和计算机网络等。谢长生　教授,博士生导师,从事新型计算机外存储体系结构和网络海量存储技术等研究工作。

资源池的映射，通过虚拟化技术，把各种异构的存储资源统一成对用户来说是单一视图的存储资源(storage pool)，为用户或应用程序提供虚拟磁盘或虚拟卷，并为用户隐藏或屏蔽具体存储设备的各种物理特性[4,8]。利用存储虚拟化技术，同时配合以分条、分区和逻辑单元掩码等技术，用户可以根据自己的需求，方便地将这个大的存储池分割、分配给特定的主机或应用程序。

采用存储虚拟化技术，几个不同的存储设备可以抽象成一个连续的虚拟设备呈现给用户。当然，如果想扩展整个系统的存储能力，可以继续无缝地扩展系统容量。如图1所示，采用虚拟化技术后，从用户的角度来看，由于disk1和disk2的块号是连续的，所以如果存储需求超过了disk1的容量限制，那么可以不作任何更改地继续在disk2上提供存储服务。同时，如果向系统提出更高的存储要求，则可以再向系统中添加其他存储设备。从图1中可以看出，向系统添加了disk3后，扩展后的虚拟设备的块号仍然是连续的，这样就实现了系统在不影响系统服务的前提下的无缝扩展能力。

显然，如果不采用虚拟化技术，如果用户想更新一个更大容量的磁盘，那么必须实际地更换所有要更换的已存在的物理磁盘。这样，原来的磁盘就一般不会再用到而闲置起来，从而造成了用户投资的浪费。

2.2 非对称结构模型

根据存储网络中数据I/O与控制信息是否使用同一通道，存储虚拟化的实现方式有对称结构和非对称结构两种。我们这里主要研究非对称结构的虚拟存储化网络模型。即数据和控制信息分别使用不同的通道传输。

非对称结构模型的实现结构如图2所示。系统中有一个元数据服务器，它负责专门管理整个存储系统的各种存储资源，并保存元数据，在增加了存储虚拟化功能的同时，对整个系统的存储I/O实施定向、分配和控制等功能。当异构的客户端要访问存储设备时，首先向元数据服务器提交请求，经过服务器授权批准后，服务器对文件I/O和块I/O进行不同的处理，再通知相应的存储模块，这样客户应用程序就和相应的存储池建立了连接，并直接进行数据的I/O请求读写。此时服务器监控数据的传输，在传输完毕后负责关闭连接并进行一些善后的工作。显然，在非对称结构模型情况下，I/O的传输不经过元数据服务器，减少了由于数据服务器转发引起的I/O的传输层次，因此这种存储I/O访问方式提高了整个存储系统的响应性能。

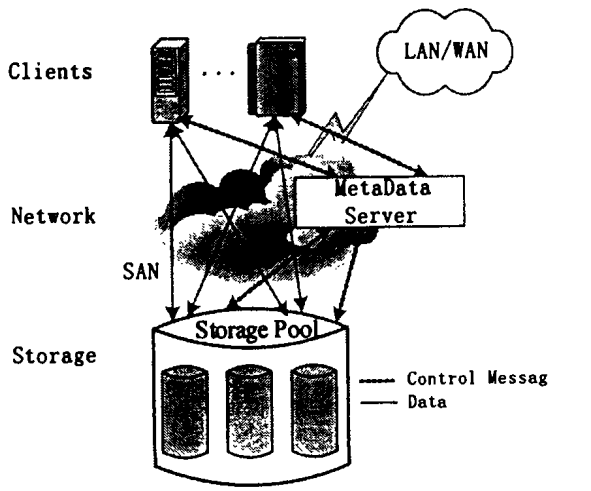

图2 非对称结构模型

3. 元数据服务器设计

3.1 元数据服务器软件结构

元数据是描述数据的数据。研究非对称结构的虚拟化存储网络系统，我们设计的元数据服务器是核心。图3所示为元数据服务器的软件结构模型。当NAS或SAN客户提出I/O请求时，经过本地的文件系统、磁盘设备驱动和网络层发往元数据服务器，由元数据服务器建立客户与存储设备的数据传输通道。由于NAS是面向文件级应用的存储设备，一般通过NFS或CIFS直接联到网络上，因此对NAS配置来说无需改变。而SAN是直接面向数据块请求的，所以我们增加了一层存储I/O虚拟化代理层，与SAN客户端的I/O虚拟化代理协同工作来截取并重定向I/O。

3.2 存储I/O数据流程分配

采用存储虚拟化技术，我们把面向文件I/O的NAS和面向数据块I/O的SAN统一在一个通用的系统结构下，透明地为两种不同的应用提供服务。同时由于我们实现的是基于非对称结构的原型系统，所以两种存储I/O的映射和分配是关键。本系统中我们通过服务器和SAN客户端的I/O虚拟化代理的协同工作来实现存储虚拟化的功能。因为针对NAS的文件I/O无需改动太多，相对简单，所以我们针对SAN的块I/O来介绍采用虚拟化技术后的存储I/O的总体流程，具体步骤如下(参见图3)：

(1)SAN客户机向存储系统提出读写请求，通过本地文件系统、设备驱动等进行必要的处理转换后到达I/O虚拟化代理层，通过控制信息通道，由其重定向到元数据服务器。

(2)元数据服务器首先进行身份认证，如果通过认证则将描述请求数据的元数据返回给客户端，建立客户端和存储池的直接数据请求通信机制，与此同时，采用某种锁机制策略将被请求的数据块加锁。

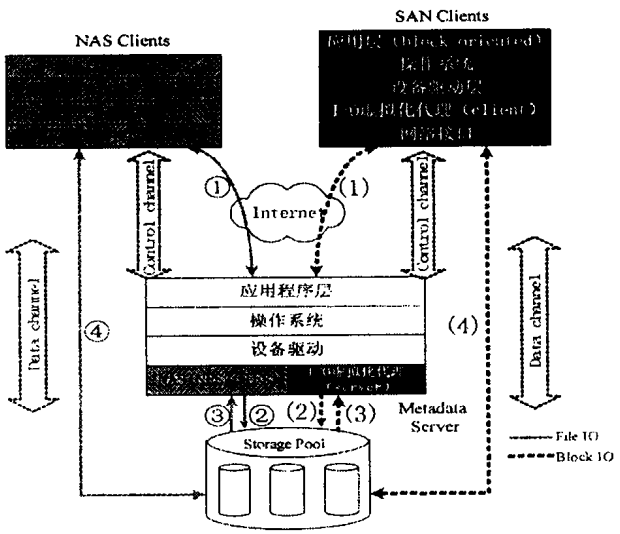

图3 存储I/O流程

(3)通过数据通道，客户端直接和被授权访问的存储池进行数据读写操作，再经过客户端软件的配合返回给上层应用程序。

(4)数据请求完毕后，数据通信关闭，存储设备向服务器报告作业完毕，由服务器进行数据解锁。

总结与展望 通过虚拟化技术，用户可以透明地使用存储资源，屏蔽具体的存储资源的物理细节，这样用户可以专注

(下转第63页)

ACM Press, Jan. 1999
2 Zhang T, Kuo C-C J. Video Content Parsing Based on Combined Audio and Visual Information. SPIE 1999, 1999, IV
3 Wold E, Blum T, Wheaton J. Content-base classification, search, and retrieval of audio. IEEE Multimedia, 1996, 3(3): 27～36
4 Tzanetakis G, Cook P. Multifeature audio segmentation for browsing and annotation. In: Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA99, New Paltz, NY, 1999
5 Tzanetakis G, Cook P. Experiments in computer-assisted annotation of audio. In: Proc. Int. Conf on Auditory Display, ICAD, 2000
6 Birmingham W, Dannenbert R, et al. Musart: Music Retrieval Via Aural Queries. In: Proc of the Annual Symposium on Music Information Retrieval (ISMIR 2001), Bloomington, ID, Oct. 2001
7 Lu Lie, Jiang Hao, Zhang Hongjiang. A robust audio classification and segmentation method. in ACM Multimedia, 2001
8 Foote J. Content-based retrieval of music and audio. In: C. C. J. Kuo et al., eds. Multimedia Storage and Archiving Systems Ⅰ. In: Proc. of SPIE, volume 3229, 1997. 138～147
9 Srinivasan S, Petkovic D, Ponceleon D. Towards robust features for classifying audio in the CueVideo System. In: Proc. of the seventh ACM intl. conf. on Multimeida, 1999. 393～400
10 Liu Z, Wang Y, Chen T. Audio feature extraction and analysis for scene segmentation and classification. Journal of VLSI Signal Processing Systems, June 1998
11 Boreczky J S, Wilcox L D. A Hidden Markov Model framework for video segmentation using audio and image features. In: Proc. of ICASSP'98, Seattle, May 1998. 3741～3744
12 Pfeiffer S, Fischer S, Effelsberg W. Automatic audio content analysis. In: Proc. of the fourth ACM intl. conf. on Multimedia, 1996. 21～30
13 Saunders J. Real-time discrimination of broadcast speech/music. In: Proc. ICASSP96, vol. II, Atlanta, May, 1996. 993～996
14 Scheirer E, Slaney M. Construction and evaluation of a robust multifeature music/speech discriminator. In: Proc. ICASSP 97, vol. II, IEEE, April 1997. 1331～1334
15 http://www-3.ibm.com/software/speech/windows/index.shtml
16 James G. Droppo Ⅱ. Time-Frequency features for speech recognition: [PhD thesis]. University of Washington
17 Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 1989, 77(2): 257～286
18 Jelinek F, Mercer R, Roukos S. Principles of lexical language modeling for speech recognition. in Readings in Speech Recognition, ed. A Waibel and K. F. Lee, Morgan Kaufmann Publishers, 1990
19 Riccardi G, Bocchieri E, Pieraccini P. Non-deterministic stochastic language models for speech recognition. In: Proc. ICASSP'95, Detroit, May 1995. 237～240
20 Ron D, Singer S, Tishby N. The power of amnesia. In: J. Cowan et al., eds. Advances in Nueral Information Processing Systems, Morgan Kauffmann, 1996, 6
21 Teuhola J, Raita T. Application of a finite-state model to text compression. The Computer Journal, 1993, 36: 607～614
22 Ghias A, Logan J, Camberlin D, Smith B C. Query by humming: Musical information retrieval in audio database. In: Proc. ACM Int. Conf. on Multimedia, San Fancisco, CA, ACM, 1995. 231～236
23 Kosugi N, et al. A practical query-by-humming system for a large music database. In: Proc. ACM Int. Conf. on Multimedia, Los Angeles, CA, 2000. 333～342
24 Pampalk E, Rauber A, Merkl D. Using smoothed data histograms for cluster visualization in self-organizing maps. In: Proc of the Int. Conf on Neural Networks, 2002
25 Bainbridge D. The role of music IR in the New Zealand Digital Music Library project. In: Proc. of the Intl. Symposium on Music Information Retrieval, 2000
26 Tseng Y H. Content-based retrieval for music collections. in SIGIR, 1999, ACM
27 Rolland P Y, Raskinis G, Ganasicia J G. Musical content based retrieval: an overview of the Melodiscov approach and system. in Multimedia, Orlando, FL: ACM, 1999
28 Blackburn S, DeRoure D. A Tool for Content Based Navigation of Music. In: Proc. ACM Multimedia 98, 1998
29 McNab R J, Smith L A, Bainbridge D, Witten I H. The New Zealand Digital Library MELody inDEX. http://www.dlib.org/dlib/may97/meldex/05written.html, May 1997
30 Uitdenbogerd A, Zobel J. Melodic Matching Techniques for Large Music Database. In: Proc. ACM Multimedia 99, Nov. 1999. 57～66
31 Tzanetakis G, Essl G, Cook P. Automatic musical genre classification of audio signals. In: Proc. Int. Smposium on Music Information Retrieval (ISMIR), 2001
32 Pampalk E. Islands of music: analysis, organization, and visualization of music archives. Master's thesis, Vienna University of Technology, 2001
33 Pampalk E, Rauber A, Merkl D. Content-based organization and visualization of music archives. In: Proc of ACM Multimedia 2002, Juan-les-Pins, France, ACM, 2002
34 Rauber A, Pampalk E, Merk D. Using psycho-acoustic models and self-organizing maps o create a hierarchical structuring of music by sound similarities. In: Porc. Int. Symposium on Music Information Retrieval (ISMIR), Paris, France, 2002

(上接第53页)

于自己的业务，不必关心具体物理设备的大小、类型、位置及特性等物理参数，增加了用户的投资回报(ROI)。同时，采用存储虚拟化技术，用户已经购买的NAS产品和SAN设备都可以有效地利用起来，这样就保护了用户的已有投资，减少了总体拥有成本(TCO)。在一个集中存储池中能更充分地利用存储空间，使得存储可以被共享在一个异构服务器的网络环境集中管理。采用非对称的结构来实现虚拟网络存储系统，由于数据传输与控制信息采用不同的通道，因此系统I/O性能得到了显著的提高。

同时，在本文设计的非对称结构模型中，由于元数据服务器负责整个系统的全局存储管理，NAS和SAN的任何一个客户端的访问请求都要经过元数据存储服务器的统一管理和分配，这样就容易产生单点失效的问题，而且系统性能也会受到不同程度的影响，所以下一步工作，有待于开发冗余结构配置，研究冗余服务器配置环境下负载平衡和高效的数据锁机制算法等问题。

参 考 文 献

1 Arunkundram R S, et al. Special Edition Using Storage Area Networks. Que, 2001. 11
2 Liu Zhaobin, Xie Changsheng, Fu Xianglin, Cao Qiang. A high scalable and performance storage architecture for multimedia applications. In: Proc. of SPIE, v 4861, 2002. 116～120
3 Anderson D C, et al. Interposed Request Routing for Scalable Network Storage. ACM Transactions on Computer Systems, 2002, 20 (1): 25～48
4 Blunden M, Berx-Debeys M, Sim D. Storage Networking Virtualization. Redbooks Publications (IBM), Dec. 2000
5 Phillips B. Have Storage Area Networks Come of Age? IEEE Computers, 1998. 7
6 Clark C T. The Virtualization of Storage. TidalWare white paper
7 Heath J R, Yakutis P J. High speed storage area networks using a fibre channel arbitrated loop. interconnect IEEE Network, 2000, 14(2): 51～56
8 Milligan C, Selkirk S. Online storage virtualization: the key to managing the data explosion. System Sciences, 2002. HICSS. In: Proc. of the 35th Annual Hawaii Intl. Conf. 2002. 2905～2913
9 Milanovic S, Petrovic Z. Building the enterprise-wide storage area network. EUROCON'2001, Trends in Communications, International Conference, 2001, 1(1)136～139
10 Griswold R. Storage topologies. Computer, 2002, 35(12): 56～63