基于千兆以太网的SAN的探讨

白俊峰　曾文方　伍良富　范雪莉
（成都四川大学（西区）计算机系　610065）

摘要：通过比较分析NSA和SAN、基于光纤通道SAN和基于千兆以太网SAN，得出基于千兆以太网SAN的跨平台和硬件兼容的优越性，进而提出千兆以太网SAN的实现要求。

关键词：存储区域网络　SAN　NAS　千兆以太网

1　引言

随着信息量的急剧增加，数据备份工作日益重要，要求存储技术进一步有所突破，新一代网络存储技术——存储区域网络（SAN）获得了飞速的发展，但各个厂商在残酷的竞争和过于太快的发展速度下，基于光纤通道技术和串行SCSI-3技术的SAN就面临硬件设备的不兼容和管理软件的不可跨平台操作等问题，这给SAN的发展带来不利的影响。但如果在SAN中引入千兆以太网技术，并精心设计其管理软件，那么就可较为容易地解决SAN发展中存在的这两大问题。

2　网络存储(NAS)和存储区网络(SAN)

2.1　网络存储

NAS(Network Attached Storage)是一种存储容量很大的设备。通过服务器I/O路径连接在网络上，它是与服务器共存的存储系统。具有使磁盘空间扩展简单、管理方便，同时具有保证数据完整性的RAID功能。其特点是：使用Ethernet网络技术和SCSI即插即用的存储技术。因此，NAS易于实现硬件和软件的集成，同时支持多平台，其存储设备（如磁带库、磁盘等）易于为不同类型服务器、客户机共享。但它是基于网络的存储设备，当存储介质增多时面临对它的读写竞争，会因网络带宽的限制，造成网络性能下降。NAS存储介质的管理是分散等原因，管理成本将随网络上的NAS设备增加而增大。因此，采用NAS难以获得满意的效果，不适宜关键事务的应用。

2.2　存储区网络

SAN(Storage Arean Networks)是为了提高I/O的性能要求，把存储设备从数据网中分离出来，使用高速网络技术为存储设备组建一个专用存储网络，将存储设备作为网络的节点而直接连接到这个网络上，如图1所示。SAN与传统的存储方式相比，具有很大优越性：隔离存储流量，不占用数据网的通信资源，既可提高数据网的效率，也可提高存储速度；SAN作为一种专用网络，可以为它开发专用网络技术，从近几年的发展中可以看出，SAN技术比数据网技术发展更快；SAN作为一个专用网络，易于实现整个存储设备的集群、易于扩展存储介质、易于实现集中管理和控制；使用光纤通道技术和千兆以太网技术，易于实现异地备份，更好实现灾难恢复。SAN的拓扑结构有三种：点到点式SAN，环行SAN和交换式SAN。目前广泛使用而又具有良好的扩展性的是交换式SAN。

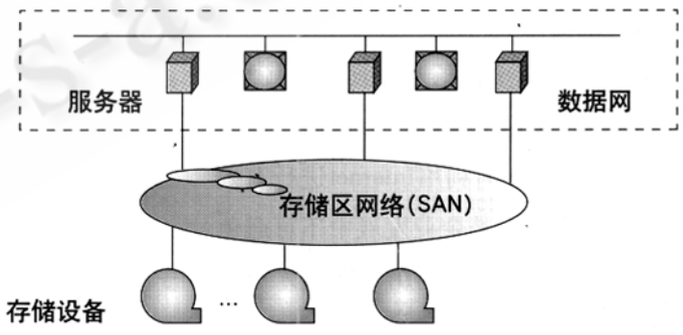

图1　SAN与数据网的分离

3　基于千兆以太网的SAN及其实现

实现SAN的方法有两种，一是基于光纤通道的SAN（简称FC SAN），二是基于千兆以太网的SAN（简称IP SAN）。.

3.1　基于光纤通道的SAN

FC SAN是实现SAN较早而又

比较成熟的技术,它是基于光纤通道物理网络的基础上实现的，其协议（总称为FCP）共分为五层：FC-0(物理层协议)、FC-1(传输层协议)、FC-2(帧协议层)、FC-3(公共服务层)、FC-4(ULP映射层)。并且在这五层之上映射相应的上层协议。FCP提供了四种服务类型：第一种是由交换机纠错排序，第二种是由终端节点底层纠错排序，第三种是由终端节点交由上层软件纠错排序，第四种是混合型。FCP完全不同于以太网的协议分层。其特点是：把存储设备作为节点连接到SAN网络上的一种高速通道，从而无需占用数据网的带宽，在上层可映射多重上层协议,如SISC-3协议。所以FC SAN采用协议是:下层为光纤通道协议(FCP),上层为SISC-3串行协议。

3.2 IP SAN的实现

基于千兆以太网的SAN的基本思想是：千兆以太网在传输速度上大于光纤通道网络，用千兆以太网代替光纤通道物理网络，可实现硬件上的性价比的最优化；在上层协议中使用TCP/IP协议，实现存储及管理功能的底层通信，还可实现管理及存储的跨平台功能。这样，在FC SAN中难以实现的设备兼容性、多平台管理功能，而在IP SAN中实现就更加容易了。

IP SAN的实现方发有多种，其中之一是通过软件将SCSI数据及命令直接映射成TCP/IP上的报文，在存储设备和服务器间通过千兆以太网SAN传输它们。本文着重讨论这种方式。

3.2.1 IP SAN的物理网络

千兆以太网的物理层提供半双工和全双工两种工作方式。在IP SAN中与大部分操作都是备份和恢复数据有关，通常占用的时间长，如果采用传统的半双工方式（不能同时在网络收发数据），因传输过程中会产生错误，这样上层的协议要求重传，但接收重传等命令和发送数据不能同时进行，再加上发送数据间有96bit的时隙，显然会降低效率。所以必须使用双工方式来提高效率。

千兆以太网的数据链路层共有3个子层：MAC子层、MAC控制子层、LLC子层。其中MAC控制子层是可选的，在千兆以太网中用作流量控制，而在IP SAN中是必选的，其原因为：千兆以太网的带宽宽，发送数据的速度也非常快，如果接收方的数据缓冲区填满后，而又无MAC控制子层功能，就会丢弃已到达的数据，另外备份和恢复所涉及的数据流量大，在无流量控制时，数据经常会溢出，而在上层传输层使用的是面向连接的TCP协议，被抛弃的数据会经常要求重传，从而造成传输效率低下。如果在IP SAN中采用了MAC控制子层，接收方和发送方可使用PAUSE帧对链路进行对流量控制，这将大大减少数据重发，提高存储效率。所以要在IP SAN中必须选用MAC控制子层。

3.2.2 构造IP SAN的备份管理软件

备份是SAN最重要的功能之一，SAN备份管理软件主要功能是在处理数据流的基础上用选定备份方法完成备份及其相关工作。因此，数据流的处理和备份方法的实现是备份管理软件的主体。

IP SAN备份管理软件的工作原理如下：

(1) 数据流的处理原理。IP SAN的备份管理软件是建立在TCP/IP协议的基础上，工作在TCP/IP协议的应用层上。其中传输层采用TCP协议进行可靠传输，即在IP SAN上给SCSI命令和数据传输架设一个TCP通道。IP SAN备份管理软件是负责管理IP SAN有关备份操作的模块，它负责对SCSI数据和命令在TCP/IP协议栈里的封装和解封、收发，负责把SCSI命令及数据交给备份驱动程序和对它们的执行结果进行处理。使用面向连接的TCP协议传输SCSI命令和数据，是因为存储命令执行的时延不能太长，对执行结果需要快速处理；传输的数据量很大；要求在命令和数据在设备上操作后，需要及时传送存储设备的状态信息。

A：在千兆以太网上传输的被封装成IP数据包的SCSI命令和数据
B：在SCSI-3总线上传输的SCSI命令和数据
C：交换机
E：返回的设备状态信息

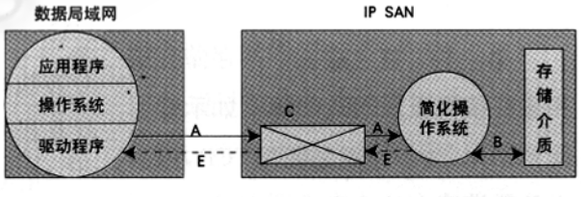

图 2 IP SAN备份管理软件数据流处理

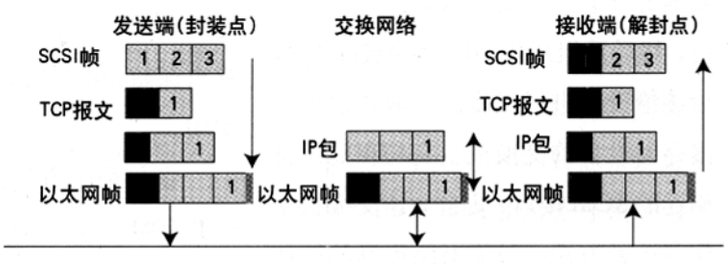

图 3 SCSI命令和数据在网络中的变化

IP SAN备份管理软件工作原理是：数据局域网服务器上的应用程序

发出的SISC数据和命令就被封装成IP数据包，再发送到IP SAN上指定的设备上解包、重组，还原成SISC数据和命令，并交给相应的存储介质执行命令和存储数据；然后简化操作系统把设备系统状态信息打包后，回送给LAN上的服务器，让它根据这些信息调用其他处理程序处理，比如错误处理等，见图2。SCSI命令和数据封装、传输、解封的变化如图3所示。

在图4中，简化的操作系统可以位于存储设备里，也可位于交换机上。它的作用是把来自千兆以太网上的IP数据包中SCSI命令和数据解封重组，并将重组好的命令和数据交给设备。因此根据简化的操作系统放置的位置（即IP数据包解封的位置）不同，可把IP SAN的方式又划分为两种：

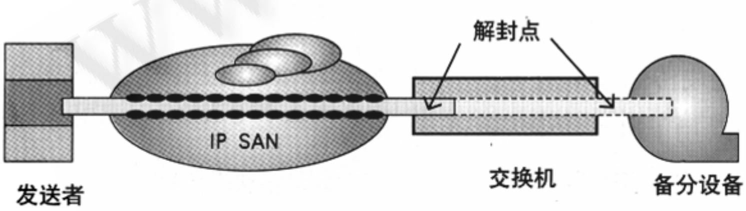

图 4 IP SAN上通道

第一种，解包点位于与存储介质相连交换机上，通道的建立如示意图图4所示。这条通道是使用TCP方式，它是数据网上的发送者和交换机（连接SAN和存储设备）间建立的；发送者通过通道发送封装有SCSI命令及数据的IP报文，交换机收到后把它解包，然后重组成SCSI命令和数据，最后传给设备进行操作。这显然要求连接存储设备的交换机上有与存储设备相连的SCSI接口，因此，让交换机工作在这种方式，就需对路由器的软件和硬件都要作修改，所以此方案的实施代价太大。

第二种，解包点在存储设备里，即存储设备通过千兆以太网卡连到IP SAN。见图4，这条通道也是使用TCP方式，在数据网上的发送者和存储设备间建立的，存储设备收到来自发送者通过该通道发送的封装有SCSI命令及数据的IP报文，把它解包重组成SCSI命令和数据，然后传给I/O总线进行存储。现在的存储设备多为智能设备，即有微处理器、内存等，在这种方式下只需要装入简化的操作系统，如广泛使用的、可裁剪内核式LINUX，装入IP-SCSI管理软件就可以完成工作了。本文的IP SAN管理软件就是基于这种模式构建的。

(2) 备份方法的选择。备份是SAN的主要功能，SAN的备份方法有三种，分别是LAN-free、集成介质和设备备份以及无服务器备分，其中LAN-free是最重要的，它是其他备份分方法的基础，本文讨论的备份分方法是指LAN-free。

LAN-fee这种备份方式是将SAN的存储流量与LAN上的数据流量分离。实现LAN-free方法应考虑两种情况：一是混合平台下，采用备分分区方法实现；二是在同一平台下，采用具有保留/释放功能的路径控制方法实现。

备份分区方法是：当SAN上多平台共存时，为不同平台生成虚拟的专用备份设备网络。具体的做法是：将所有SAN设备集成到一个简单的连线结构中，将它们通过逻辑上隔离到不同I/0路径上。目的是避免不同平台的服务器对备份设备的交叉访问。

路径控制方法是：在SAN上，当某一种类型的平台的服务器多于一个时，让备份应用程序可以保留一个特定的备份设备，并留给自己使用，直到它不再需要该设备，此时它可以释放设备给其他备份程序。其目的是避免两个备份发起者不会因为同时向相同的设备发送备份数据而纠缠在一起。

4 结束语

本文通过比较分析NSA和SAN、基于光纤通道SAN和基于千兆以太网SAN，得出基于千兆以太网SAN的跨平台和硬件兼容的优越性，进而提出千兆以太网SAN硬件的实现要求。基于千兆以太网的SAN具有了硬件兼容性和管理软件跨平台管理的特点，其管理软件的实现，还要在今后做更细致的工作。■

参考文献

1 Farley•M. Building Storaage Networks. McGraw Hill, 2000。
2 W.Richard Stevens. UNIX 网络编程，北京清华大学出版社，1999。
3 W.Richard Stevens. Advaced Programming in the UNIX Environment. Addion-Wesly, 2000。
4 SCI与光纤通道之争，网络世界，2001，4。
5 SAN解决方案，www.jdxfcg.com/project2/thm-doc/ad1.htm。