

基于 IP 的全局网络存储解决方法

刘福明¹ 郑纬民²

(重庆三峡学院计算机科学系 重庆万州 404000)¹ (清华大学计算机科学与技术系 北京 100084)²

Global Networking Storage Solution Based On IP

LIU Fu-Ming¹ FU Chang-Dong² ZHENG Wei-Min²

(Computer Science Dept of Chongqing Three Gorges College, Wanzhou 404000)¹

(Computer Science and Technology Dept of Tsinghua Univ., Beijing 100084)²

Abstract IP Storage achieves a method to transfer SCSI command and block-level data over IP network. It decreases cost of network storage realization and extending SANs over long distance, and can be a global storage solution. This paper discusses advantage of IP storage bring and the key technology and strategies of IP storage.

Keywords IP Storage, iSCSI, FCIP, iFCP, Global

1. 引言

随着 Internet 的应用和普及,信息爆炸性地增长。因此,改善存储性能和提供 7×24 的可用性的压力也越来越大。传统的存储技术(BAS)采用并行 SCSI 或 DAS 技术,基于总线结构,无法提供足够灵活和可扩展的功能以满足如今对应用和存储快速部署的要求。

网络存储采用了新的存储体系结构。它将网络与存储 I/O 路径相结合,使数据处理和数据存储分离,消除了不同存储设备和服务器之间的连接障碍,提高了数据的可用性、共享性、可扩展性和可管理性。

网络存储包括了网络和 I/O 的精华,提供资源共享,具有灵活的寻址能力,远距离的传输能力和 I/O 高效的性能等。SAN(Storage Area Network)和 NAS(Network Attached Storage)是目前网络存储技术发展的两个主要方向。NAS 具有提供以太网数据访问的传统,通常以与现有网络结构相容的应用软件包形式提供集成服务。它可以很经济地解决存储容量不足的问题。对于关键事务而言,它必须使用专用的宽带网络。安装调试容易,易于管理。但难以获得满意的性能。SAN 是一种新的存储连接拓扑结构,它被设计用于代替现有的系统和存储系统之间的 SCSI I/O 连接。它的实现一般采用光纤通道(FC)技术,速度快,性能好,已逐渐变成了主流的存储解决方案。但其设备价格高,总的实现费用昂贵,难以配置

和维护;并且由于缺乏统一标准,产品互操作性差。另外, SANs 是一个专用局域网络,不能对远程服务器和其它非本地计算资源进行访问。

为了解决这些问题,人们提出了 IP 存储。IP 存储结合了 NAS 和 SAN 存储系统的优点,实现了在 IP 网络上传输 SCSI 协议的技术。一方面通过 IP 网络通信,降低了实现费用;另一方面使用块协议,提供高的系统传输速率。

2. IP 存储

IP 存储是基于 IP 网络的存储解决方案。它通过一组存储协议(iSCSI、FCIP、iFCP 等)提供了在 IP 网络上封装块级请求的方法,使得我们可以使用标准的 TCP(Transmission Control Protocol)在 IP 网络上传输 SCSI 命令和块数据,提供高的系统传输速率,并能容易地进行长距离(超过 10km)的全局数据存取。这为很多相关应用,如:远程镜像和数据复制、统一资源、远程磁盘(或磁带)备份和恢复、容灾以及远程启动等带来好处。

其存储 I/O 路径(以 iSCSI 卡为例)见图 1。

IP 存储系统的 I/O 路径和 SAN 的相比,只是光纤通道卡和光纤通道协议变为 iSCSI 卡和 iSCSI 协议。其它方面完全一样。在 IP 存储系统中,SCSI 是通过 iSCSI 协议映射到 TCP/IP 上的,并将 SCSI 从并行总线结构中释放出来。

iSCSI 卡将 TCP/IP 协议作为引擎在卡内实现。

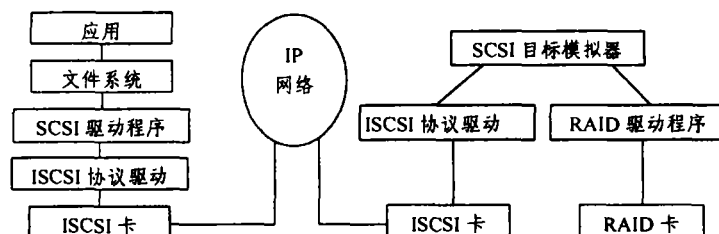


图 1 IP Storage 存储系统的 I/O 路径

IP 是当前流行的通用网络协议。它之上的技术及管理工具为大家所熟悉和接受。并且有统一的标准,产品具有通用

性。将其与存储相结合,是如今研发的热点之一。IP 存储让人们使用公共硬件及熟悉的管理工具,这样 IT 部门不必在 IP

刘福明 讲师,主要研究方向:网络存储、软件工程。郑纬民 教授,博士生导师,主要研究方向:并行处理/分布计算、计算机体系结构。

和 FC 两者中维护设备、技术工人以及专门技术,降低了实现费用。

很多 IT 部门,尤其是那些想利用网络存储先进特性的中小公司,均倾向于 IP 存储。据统计,到 2005 年,用户将看到采用光纤技术的 SAN 的市场份额逐渐下降,而 iSCSI、IP over Storage 等技术的市场占有率不断上升。其中 iSCSI 更因为解决了开放性、容量、传输速度、兼容性、安全性等问题,以及价格低、性能好的优点而将占领 30% 以上的市场份额。

3. 关键技术

目前,IETF(Internet Engineering Task Force)在这个领域开发的标准有三个:iSCSI、FCIP、iFCP。FCIP 描绘了不朽的光纤通道,iFCP 描绘了 FC 到 IP 存储网络的转换,而 iSCSI 描绘了最终的 IP 存储网络的目标。虽然它们有不同的实现和设计目标,但每一个都将允许用户通过标准的 IP 网络,甚至 Internet 延伸它们的存储环境。

3.1 iSCSI(Internet SCSI)

这是一个基于 IP 网络(尤其是以太网)的执行块数据调用的主机到存储、端到端的协议标准。与 FC SAN 结构相同,iSCSI 技术将包括 iSCSI 主机和 iSCSI 存储阵列,它们之间通过 IP 交换机进行通讯。iSCSI 建立在两个基本的技术之上:存储 SCSI 协议及网络 IP 协议。它将提供必要的映射,使 TCP/IP 成为 SCSI 命令的传送器。其协议栈如图 2。

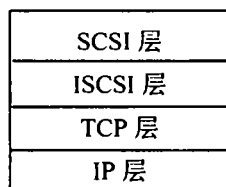


图 2

iSCSI 使用 TCP 作为它的下层传输层,提供一个可靠的保证按序传送的传输器。SCSI 命令封装后进入 iSCSI PDU (Protocol Data Unit),依次加上 iSCSI 和 TCP 包头。此时,封装的 SCSI 命令就被像其它的 IP 包一样对待,通过标准的 IP 结构传送到目标(基于 IP 地址)。目标设备接收到该包后,就剥落每一层直到 SCSI,从而得到 SCSI 命令。

iSCSI 采用全局的 IP 寻址方案。其设备有两种类型的标识:iSCSI 名和 iSCSI 地址。

与 FC 全局名(WWN)一样,iSCSI 发起者和目标都将被给一个永久的 iSCSI 名,这个名字将标识这个设备,而不管它的位置和 IP 地址。iSCSI 地址指定了 iSCSI 发起者或目标的位置,它由 IP 地址、端口号和设备的 iSCSI 名组成。如:

iSCSI 名:fqn.com.disk.-vendor.diskarray.45678

iSCSI 地址:iSCSI:/diskfarm1.acme.corn:80/fqn.com.disk-vendor.diskarray.45678

iSCSI 是一个使用 IP 结构连接服务器和存储系统资源的主要方法。理想的 iSCSI 结构将是一个 All-IP 方法,所有的服务器和目标设备都将有 iSCSI 接口。

3.2 FCIP(FC over IP)

FCIP 提供了一个允许 FC SANs 孤岛间通过 IP 网络互连从而形成一个大的、统一的 FC SAN 的机制。这种扩展的 FC SAN 结构可以继续使用标准的 FC 寻址。

FCIP 能以多种方式实现,或单独的形式(gateway、switch、router 等),或嵌入 IP 路由器里,或通过 FCIP 端口与 FC 交换机结合。典型的 FCIP 设备一边具有 FC 端口,与 FC SAN 相连;另一边具有 Gb 以太网端口,与 TCP/IP 网络连接。

FCIP 的目标是在 IP 结构上传送 FC 帧。FCIP 网关封装 FC 帧,然后使用 TCP 作为下层传送器(也可使用 UDP 实现,但 IETF 提倡使用 TCP)。一旦 FC 帧被映射到 IP 上,它便能象 IP 包一样通过 IP 结构传到目标设备。

使用 FCIP 设备连接 FC SANs 必须在 FCIP 端点间建立 IP 隧道。FCIP 隧道建立后,链接对 FC 设备是透明的。FC 交换机看到的就像标准的 ISLs(Interswitch Link)。此时,FC 设备就把这些扩展的连接看作是标准的 FC 连接,并运用 FC 寻址。其典型的实现是使用 FCIP 端点在 ISL 方式下,通过一个标准 IP 结构连接两个(或更多)FC 交换机,从而组合两个(或多个)分散的 SAN。

3.3 iFCP(Internet FC Protocol)

iFCP 介于 iSCSI 和 FCIP 之间。像 FCIP 一样,iFCP 封装 FC 帧在 IP 结构上传输。因为这一点,IETF 制定了公共的 FC 封装格式。它们的不同主要在于它们的寻址方案。FCIP 建立的是 point-to-point 隧道,而 iFCP 是一个 gateway-to-gateway 的协议,其组合 FC 和 IP 寻址,允许传送 FC 帧到适当的目标地址,并允许每一互连的 SAN 保持自己独特的名字空间。

当远程 SANs 用 IP 结构通过 iFCP 连接时,每个 SAN 提供一个 iFCP 端口(或入口)。每个 iFCP 入口将其所在的局部 FC 结构里的所有设备通过 IP 网络呈现给其它的 iFCP 入口。每个入口维护一张远程设备表,并把它们作为它的局部 FC 结构的局部设备。这个过程对实际的本地 SAN 设备是透明的,远程设备看起来就像是本地的。当一个本地设备需要存取远程设备时,它就发出一个 FC 帧到远程设备的本地 FC 地址。这个帧转到本地 iFCP 入口,封装后进入 IP 包,然后传送到适当的远程 iFCP 入口。这个远程入口解包这个 FC 帧,并传输到指定设备。

这些 iFCP 入口既能进行 FC 寻址,也能进行 IP 寻址,所以它也可以作为标准的 FC 或 IP 交换机,从而代替 FC 结构。同时,FC 的存储阵列、HBAs、路由器、交换机和 HUBs 等也能直接加入 iFCP 存储交换机。

iFCP 的下层协议一般是 TCP。当我们用不同的协议,如 UDP 代替 TCP 时,也是可行的。我们称它为 mFCP (Metropolitan FCP)。

4. 实现策略

IP 存储带来的好处吸引了人们研发的热情,但它存在的问题,如:网络传输延迟、安全级别、性能价格比以及 TOE (TCP Offload Engine)机制的实现等,也限制了它的具体实施。一般来说,IP 存储的实现有这样三个策略(或阶段)。

4.1 存储 Tunneling

由于 FC SANs 的存在和增长,远距离地通过 IP 网络连接这些 SAN 的需求也就应运而生。一般我们使用 FCIP 设备来完成这种连接。如图 3。当然,iFCP 和 iSCSI 也可使用。

这种连接对一些应用,如 SANs 间的数据共享,远程容灾等带来了好处。

虽然存储 Tunneling 使用了 IP 网络,但它对 IP 的优势利用很少。Tunneling 只不过提供了一种连接 FC SANs 的方

法,并没有解决互操作性及分散管理等问题。

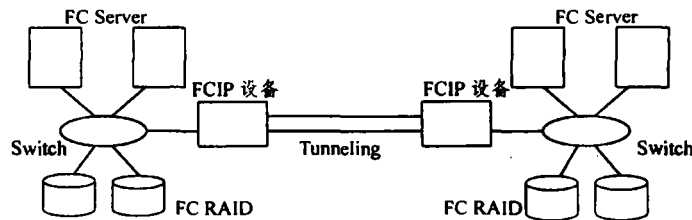


图 3 存储 Tunneling

4.2 Native IP Storage

这是有限范围的 IP 存储。它使用 iFCP(或 iSCSI)设备代替 FC SAN 中的 FC 交换机、HUB,形成一个本地的小范围的 IP SAN,而 FC 服务器和 FC 存储设备可以不做任何修改。从

而让 SCSI 和 FC 协议直接与 IP 协议结合,使存储与网络无缝合并,统一了存储和网络的管理。

Nishan System 提出了一个完整的框架 SoIP (Storage over IP)实现这种策略。如图 4。

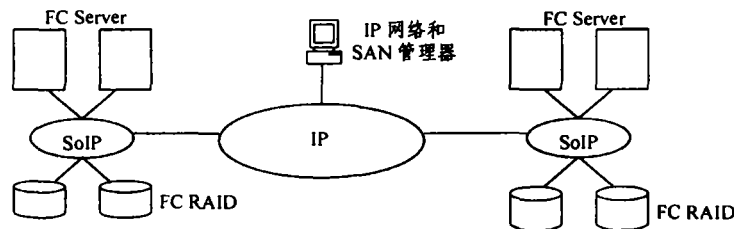


图 4 Native IP Storage

4.3 IP SAN

当人们克服了 IP 存储的诸多障碍后,就可以建立完全的端到端的基于 IP 的全局 SAN 了。如图 5。这最适合用 iSCSI 协议来实现。这种基于 iSCSI 的 IP SAN 将包含 iSCSI HBA 卡以卸载大量的 TCP 开销,并让本地的所有 iSCSI 设备通过标准 IP 结构通信。这样,很多先进的 IP 功能,如带宽集成、QoS 的保证等都将进入 SAN 空间。

和网络通信协议的转换协议,和 IP-SAN 的协议结构基本相似。

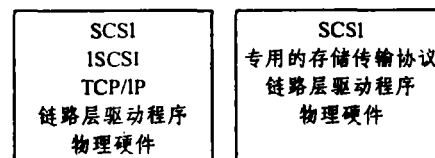


图 6 协议栈

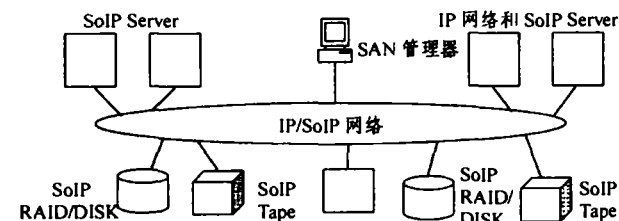


图 5 IP SAN

结束语 当前已实现了在 IP 上封装存储通信的技术及方法。IETF 标准草案(如 iSCSI)设计提供了一个标准化方法,以在现有的 IP 网络之上传输块级存储。

IP 网络结合进存储网络,不管什么形式,从它们基于今天的 FC 技术到一个广泛范围的基于 IP 的(如 Ethernet、ATM)全球配置技术,都将扩展 SANs。无论是互连远程 FC SANs,还是建立完全的端到端的 IP SAN,基于 iSCSI、FCIP、iFCP 的 IP 存储都将被期望成为将来存储行业的关键角色。

5. 基于高速数据通信网络的全局网络存储

基于 IP 网络的全局网络存储方案实际上是基于高速数据通信网络的全局网络存储方案,网络存储的网络通信物理硬件不仅仅可以基于以太网,还可以基于 MYRINET、ATM、SDH 等等。在高速数据通信网络上实现块级数据存储,将可以统一各式各样的数据通信网络,而不仅仅是 IP 网络。基于高速数据通信网络的全局网络存储在协议栈存在 2 个不同的结构(图 6):

在各种数据通信网络的链路层上实现专用的存储传输协议,它负责映射 SCSI 协议和网络通信协议。而另一种方法是在链路层上实现 TCP/IP 协议,然后以 iSCSI 作为 SCSI 协议

参考文献

- 1 IP Storage: A Technology Overview <http://www.storageareanetworking.net>
- 2 Building Global Storage Networks <http://www.storageareanetworking.net>
- 3 Farley M, 孙功星等译. SAN 存储区域网络. 机械工业出版社, 2001. 1
- 4 <http://www.emass.com>
- 5 <http://www.snia.org>
- 6 <http://www.sannas.ca/>
- 7 <http://www.adic.com>