

Cisco数据中心容灾 及业务永续解决方案

第二部分

Cisco容灾及业务永续 解决方案 —**SAN**扩展技术详述

2.1 城域DWDM的SAN扩展实现业务永续.....	5
2.1.1 业务永续应用.....	5
2.1.2 用于业务永续和灾难备份的DWDM技术.....	6
2.1.3 案例分析.....	7
案例 1 – 银行数据中心应用.....	7
案例 2 – 公共基金多地网络整合.....	8
案例 3 – 保险公司.....	10
2.1.4 总结.....	12
2.2 在SONET/SDH上的光纤通道支持.....	13
2.2.1 Cisco存储及光产品.....	13
Cisco ONS 15000 光产品系列.....	13
Cisco MDS 9000 存储产品系列.....	14
2.2.3 灾难恢复和业务永续应用.....	15
同步复制.....	15
异步复制.....	16
灾难恢复和业务永续应用的需求.....	16
远程磁带备份应用.....	17
远程磁盘镜像.....	18
光纤通道SAN扩展设计—基于存储的复制.....	19
架构和网络设计.....	20
存储网络管理.....	21
2.2.4 总结.....	22
2.3 以FCIP实现SAN扩展.....	23
2.3.1 概述.....	23
为什么要使用FCIP?.....	23
解决方案拓扑.....	23
必要组件.....	24
2.3.2 使用压缩功能.....	25
压缩和可用带宽.....	25
压缩模式和速率.....	26
2.3.3 使用写加速功能.....	28
写加速概述.....	28
往返次数超过两次的写交换.....	30
在使用写加速时的预期性能增益.....	31
何时应当使用写加速.....	33
使用写加速的考虑因素和指导原则.....	35
2.4 利用Cisco MDS 9000 系列多层交换机构建SAN互联解决方案.....	36
2.4.1 SAN互联.....	36
2.4.2 SAN岛的发展.....	37
2.4.3 虚拟SAN.....	37
2.4.4 VSAN间路由简介.....	38
VSAN间分区 (IVZ).....	39
IVR边界交换机.....	39
IVR通道(Transit)VSAN.....	40
网络的稳定性和永续性.....	40
采用了VSAN间路由的永续性设计.....	41

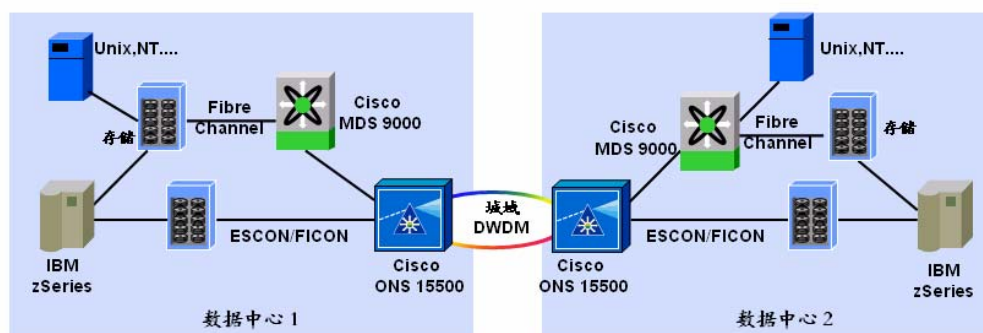
2.4.5 VSAN间路由的用途	41
共享对跨VSAN边界的存储设备的访问	41
为数据复制提供扩展SAN解决方案	41
VSAN间路由的最佳实践经验	43

2.1 城域 DWDM 的 SAN 扩展实现业务永续

2.1.1 业务永续应用

为保障业务永续的应用包括了很多种，其中最重要的两个应用是同步数据镜像和跨地域分散集群系统。它们包含了如下面图 2-1 和图 2-2 所示的实时数据镜像。

图 2-1 同步数据镜像

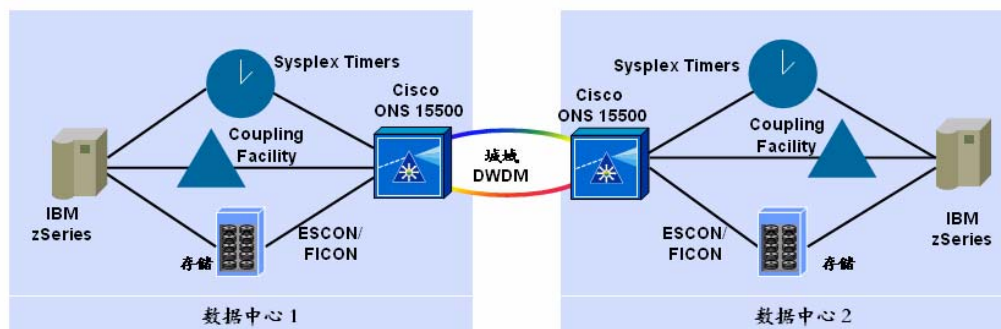


同步数据镜像和跨地域分散集群系统被大量使用，包括金融（银行、基金、经纪所、保险等）、卫生保健（医院、保健服务商）以及各级政府安全等。

同步镜像提供了：

- 在两个地点的存储设备实时镜像，实现快速恢复（一般为1-2个小时）
- 一般情况下距离为20-300公里
- 支持的存储环境包含IBM PPRC, EMC SRDF, HDS TrueCopy, HP Continuous Access
- 需要高带宽、低延迟的网络来支持光纤通道、IBM FICON和ESCON协议
- Cisco ONS 15500系列，包含ONS 15530和ONS15540城域密集波分(DWDM)平台，在高速、关键业务网络上 提供了对于这些存储接口的高密度支持。

图 2-2 IBM 跨地域分散集群系统



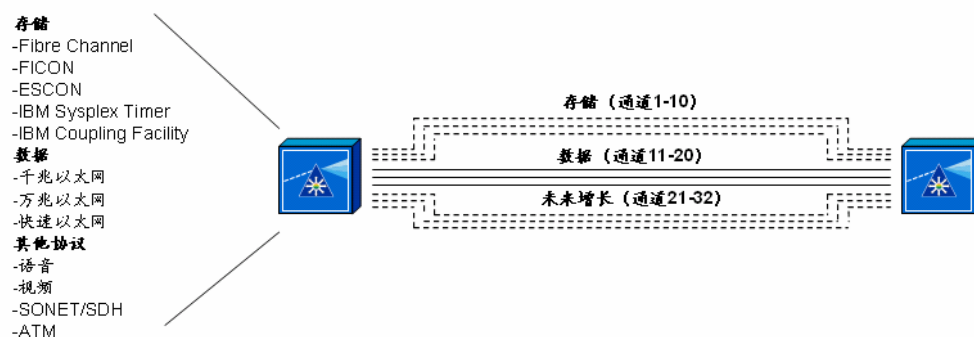
跨地域分散集群系统提供了：

- 在两个地点的存储设备和服务器应用系统的实时镜像
- 通过计算资源的负载均衡和即时向备用节点的切换，使得恢复时间从由小时计下降到以分钟计
- IBM Geographically Dispersed Parallel Sysplex (GDPS)支持最远100公里的同步主机计算
- 需要高带宽、低延迟的网络来支持存储（ESCON/FICON）和IBM Sysplex时钟和耦合设备
- Cisco ONS 15500城域密集波分(DWDM)解决方案在提供了高密度FICON和ESCON存储接口的同时，还提供了对IBM特定协议（如Sysplex始终和耦合链路等）的支持。

2.1.2 用于业务永续和灾难备份的 DWDM 技术

密集波分（DWDM）技术在—对光纤上提供了多个不同波长的光通道。在一个 32 通道的 DWDM 系统上，—对光纤上能够承载多达 32 种业务，从而提供了业务永续所需要的高密度的业务支持能力。例如，光通道 1-10 可分配给诸如光纤通道、FICON 或 ESCON 的存储接口，而光通道 11-20 则可以分配给诸如千兆以太网或万兆以太网等数据接口使用，其余的 21-32 通道可以用来支持未来业务和网络增长所需要的更多接口（图 2-3）。

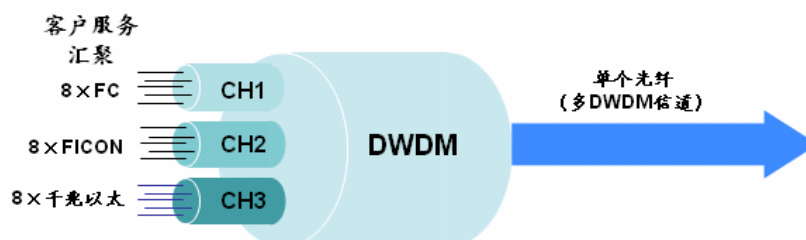
图 2-3 密集波分（DWDM）



再进一步，在使用 DWDM 技术将多个波长（光通道）复用在一对光纤之前，还可以在使用业务汇聚技术将多种业务或数据类型整合在一个单个的 DWDM 光通道内。图 2-4 就展示了将八个光纤通道连接汇聚到 DWDM 光通道 1 内，八个 FICON 连接汇聚到光通道 2 内，还有八个千兆以太网连接汇聚到光通道 3 内。继而，使用 DWDM 将这些光通道合并到—对光纤上。

这样的业务汇聚和 DWDM 的合并提供了能够支持数百业务种类的可扩展解决方案，而这些对于在大型数据中心之间的同步数据镜像或跨地域分散集群应用系统非常有用。

图 2-4 业务汇聚



基于Cisco ONS 15500系列的城域DWDM解决方案，能够支持实时业务永续应用，这包括了IBM GDPS、EMC Symmetrix Remote Data Facility (SRDF), HP Continuous Access和HDS TrueCopy。

使用Cisco ONS 15500城域DWDM解决方案来实现业务永续具有以下优点：

- 在—对光纤上提供多协议支持（如存储、数据、语音、视频、SDH/SONET和ATM等）
- 容量和吞吐率的极大提高
- 提供业务汇聚和业务透明

- 支持大量的业务永续应用，其中包括同步数据镜像、跨地域分散集群系统、异步镜像和本地/远端磁带备份
- 通常在20-300公里的范围内
- 地域透明的存储整合和共享
- 业务永续的延迟降低到最小
- 降低了技术复杂度，提供了快速的恢复
- 公共管理支持

2.1.3 案例分析

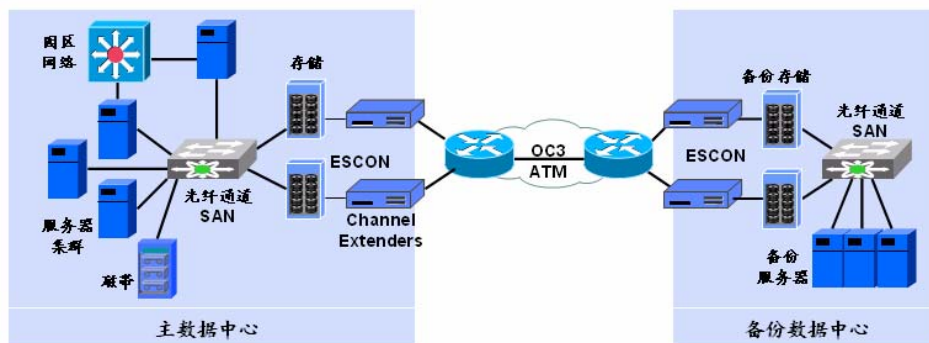
本节展示了城域DWDM解决方案如何实现高带宽、可扩展解决方案，在城域扩展业务永续的策略。

案例 1 – 银行数据中心应用

案例1展示一个中型的银行，它准备将其目前基于ATM的网络升级为基于城域DWDM进行SAN扩展，从而在两个主要数据中心之间部署数据镜像、联机交易处理（OLTP）和数据网络应用。

网络描述	◆ 节点数：2（数据中心点对点） ◆ 距离：30公里（东向），20公里（西向） ◆ 应用： — 存储：部署新的同步磁盘镜像应用 — 数据网络：园区局域网扩展 — 多媒体业务：语音、视频 ◆ 业务：光纤通道、FICON、ESCON（存储）；千兆以太网、万兆以太网（数据网络）
问题	◆ 需要增强灾难备份的性能以满足美国联邦和证交会及沙式法案所规定的新的“无数据丢失”政策 ◆ 必须能够应对业务增长对网络带来的新的流量负载 ◆ 当机的风险及信息的丢失 ◆ 升级到2.5Gbps速率的ATM的昂贵费用：\$250,000的第三方ATM转换器及每月\$45,000的线路租用费，还有日常的维护 ◆ 服务器内未用资源超过\$1,000,000 ◆ 单协议；不能支持其他流量（诸如光纤通道、NAS、LAN和PSTN） ◆ 由于共享一个连接，导致延迟影响到最终用户的体验
ONS 15530业务永续解决方案优点	◆ 提供高可用的解决方案，保护了关键业务数据并符合行业规范 ◆ 降低方案费用并提高了灵活性 ◆ 使用专门的连接改善了业务应用的性能 ◆ 可扩展的网络，支持存储、数据、语音和影响应用

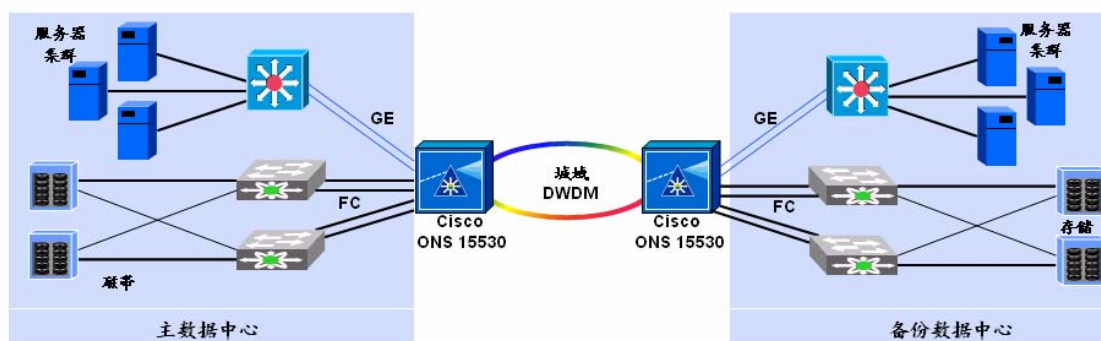
图 2-5 之前的灾难备份解决方案



图中所示的该银行以前的灾难恢复解决方案具有以下缺点：

- 存储设备间的同步镜像—该网络不能满足“零数据丢失”的需求
- 关键业务OLTP(联机交易处理)应用收到最终用户由于系统延迟的抱怨
- 目前的基于ATM的OC-3（155Mbps）有局限的解决方案，以及昂贵的升级为OC-48的带宽扩展
- 备份服务器需要人工切换

图 2-6 使用城域 DWDM 的灾难恢复解决方案



通过向 Cisco ONS 15500 城域 DWDM 网络的过渡，实现 SAN 扩展和网络整合，该银行能够在以下方面获得收益：

- 无需支付每月及每年的维护费用
- 通过跨不同地点的存储共享，使设备固定投资得到更好的利用
- 改善了Oracle OLTP应用的性能
- 更灵活地支持未来的商务及应用需求
- 支持众多的存储和数据协议，包括了ESCON、1Gbps和2Gbps光纤通道及FICON接口。从而为SAN应用的大容量、高可靠性的需求提供支持。

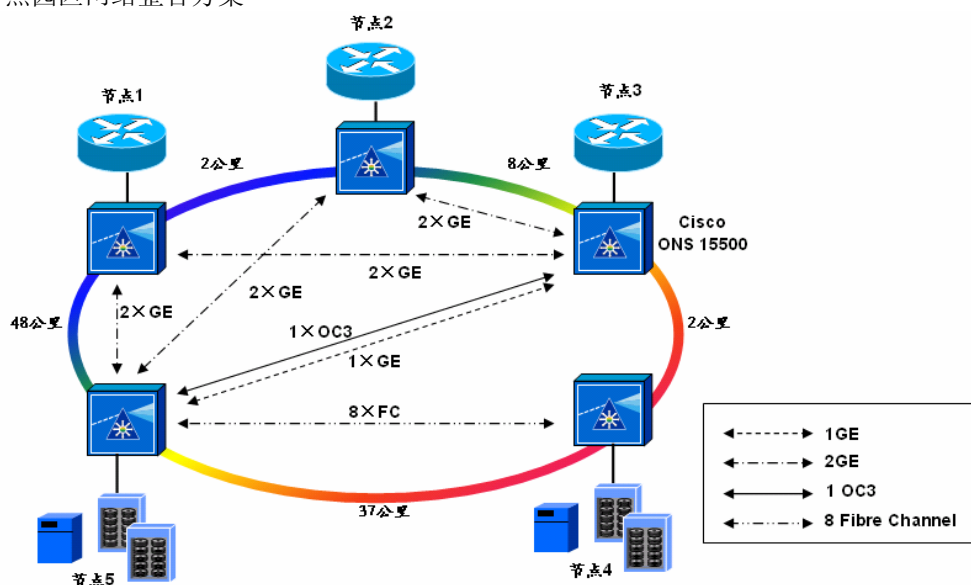
案例 2 – 公共基金多地网络整合

案例2展示一个公共基金机构，它准备建立一个城域DWDM网络来实现多地网络整合。

网络描述	◆ 节点数：5（网状连接环） ◆ 距离： 124公里（环结构） ◆ 应用： —存储：部署新的同步磁盘镜像应用 —数据网络：园区局域网扩展 —多媒体业务：语音、视频 ◆ 业务：初始为光纤通道、千兆以太网和OC-3，在随后的两
------	---

	年扩展到更多业务
问题	◆ 需要高可用的环境以防止关键业务数据的丢失（沙式法案） ◆ 需要将平均当机事件从24小时降下来，因为每小时的当机损失费\$1,495,134 ◆ 需要自动从主SAN复制20TB的数据到备用SAN ◆ 需要降低方案的费用并提高灵活性以适应未来流量的增长
ONS 15530业务永续解决方案优点	◆ 高可用性业务永续解决方案满足了对关键业务数据的保护 ◆ 平均当机时间降低为1小时（相当于节约了三千四百万美元） ◆ 单个网络支持SAN、数据、语音和视频流量节约网络费用 ◆ 富裕的容量能够更好的支持未来增长

图2-7 多点园区网络整合方案



客户对于第一年的最初需求包含了以下的接口：

- ◆ 2个GE接口（节点1和3）
- ◆ 2个GE接口（节点1和5）
- ◆ 2个GE接口（节点2和3）
- ◆ 2个GE接口（节点2和5）
- ◆ 1个GE接口（节点3和5）
- ◆ 1个OC-3接口（节点3和5）
- ◆ 8个光纤通道(FC)接口（节点4和5）

通过提供最多32个DWDM光通道的支持，Cisco ONS 15500网络使公共基金能够扩展其网络，以支持将来更多的光纤通道和千兆以太网端口。另外，Cisco ONS 15500的设计还提供了将网络扩展到其余节点的能力。部署Cisco ONS 15500网络还有其他方面的优点：

- ◆ 通过IBM、EMC、HDS和HP的认证
- ◆ 网络整合 — 客户能够通过建立一个单一、可扩展、高速端到段的光网络，提供强大的业务永续策略，满足其所有在存储、数据（千兆以太网、万兆以太网）、语音和视频等的需求
- ◆ 方便地将光加入到Cisco环境 — 客户能够方便及快速地通过使用相同的Cisco管理工具，通过已城域光网络有的Cisco LAN/WAN网络

◆ Cisco公司的服务和支持 — 非常简便地与已有的Cisco服务和支持计划融合

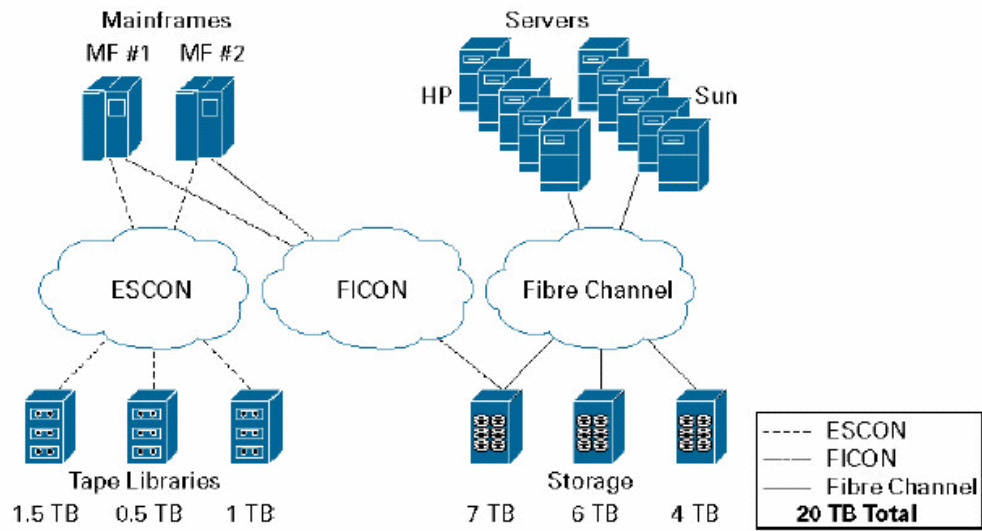
案例 3 – 保险公司

案例3展示一个保险公司，作为其业务永续计划的一部份，准备建设第二数据中心用于灾难恢复。

网络描述	◆ 节点数：2（点对点保护环） ◆ 距离： 40公里 ◆ 应用： — 存储： 部署新的IBM GDPS和同步磁盘镜像应用 — 数据网络： 千兆以太网和未来对万兆以太网的支持 — 多媒体业务： 语音、视频 ◆ 业务： 光纤通道、FICON、ESCON、千兆以太网、IBM Sysplex 时钟河耦合设备
问题	◆ 需要第二数据中心以满足沙式法案对灾难恢复的要求 ◆ 目前的网络有较长当机时间及信息丢失的风险
ONS 15530业务永续解决方案优点	◆ 很少的当机事件，从而处理更多交易并提高用户满意度 ◆ 竞争优势及对有价值客户数据的保护 ◆ 网络整合 — 一个网络满足存储、数据、语音和视频需求

建立第二灾难恢复节点的最先步骤是考察当前存储网络和应用在接口方面的需求。图2-8展示了初始的存储网络，包含了服务器、存储设备，以及基于ESCON、光纤通道和FICON连接的存储网络来提供设备间通信。整体存储（包括直连存储设备和存储区域网络）为20TB的数据。

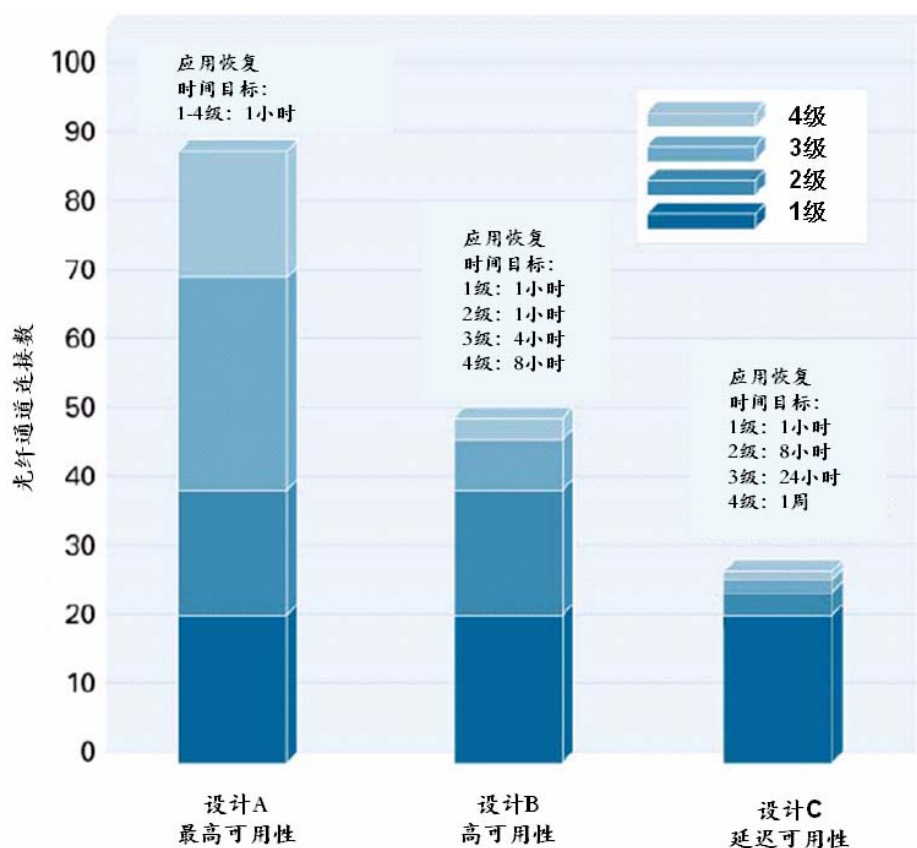
图2-8 初始存储网络



在作初始网络评估时，应用被划分为一级（最高优先级）、二级（次高优先级）、第三级和第四级。基于不同的恢复时间的目标（图2-9），考虑了三种设计（设计A,B和C）。设计A具有最高的可用性，能够满足为从一级到四级的所有应用提供最严格的一小时的恢复时间目标，它需要89个光纤通道连接。而B设计中则使用了较为宽松的恢复时间要求，其中为一级和二级应用提供一小时的恢复时间，而为三级应用提供四小时的恢复时间，四级应用提供八小时恢复时间。通过降低对三级和四级应用

的恢复时间要求，整个对于连接的需求降低到50根光纤通道连接。C设计为一级应用保持了一小时的恢复时间，但放宽了对于二级应用（8小时）、三级应用（24小时）和四级应用（1周）的恢复时间，这样就将所需的连接数目降到了28根光纤通道连接。因此，为了支持更为严格的恢复时间，就需要更多的网络连接。最后，该保险公司决定使用A设计“最高可用性”来保证所有的应用在一小时内恢复。

图 2-9 恢复时间和网络连接数目的关系



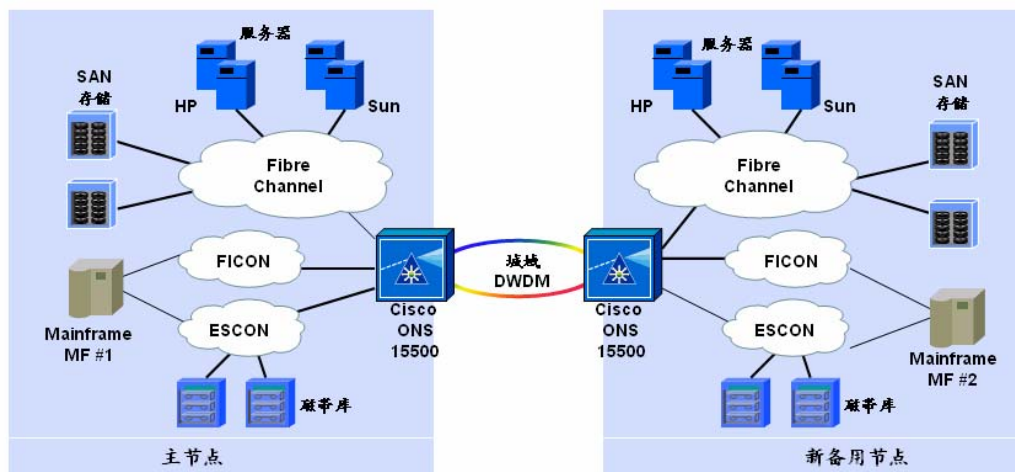
在设计考虑中包含了对于潜在的未来增长的分析以及对后面的三年所需的网络应用的估计。从而确定在第一年内，存储数据将由第一年开始的20TB增长到第三年超过30TB。最终选择了ONS 15500 DWDM解决方案，它不仅能够支持初始的89根光纤通道连接，还能够很方便的扩展到支持最密集连接需求（每一对光纤上最多可以支持256个光纤通道）

设计考虑

- 光容量 — 光容量能够通过DWDM得到非常大的提高；每个DWDM光通道能够支持最多8个光纤通道（FC）或千兆以太网，或是最多支持40个ESCON通道。
- 灵活性 — 支持存储（光纤通道、FICON、ESCON、IBM Sysplex Timer和耦合链路）和数据（千兆以太网、万兆以太网）
- 扩展性 — Cisco ONS 15500解决方案提供了满足从当前到今后三年的网络需求的能力，而无需更换设备
- 经济性 — 使用DWDM下的每单位带宽的费用急剧降低，使得第二备份节点成为可能
- 资源利用率 — 使得所有的存储空间能够被公司的所有用户共享

为满足A设计“最高可用性”（所有应用一小时恢复）的需求，部署了城域DWDM网络（图2-10）

图2-10 使用城域DWDM网络的新的灾难备份设计



部署Cisco ONS 15500城域DWDM网络的优点包括

- 通过IBM、EMC、HDS和HP的认证
- 与AT&T、BellSouth、Qwest、SBC和Verizon是合作伙伴
- 业界领先的技术（密度、保护和管理）以及对未来更高能力的承诺
- 可扩展的解决方案，能够以小空间和低电源功率的需求，满足无论低密度（两个光纤通道/千兆以太网或10个ESCON通道）还是高密度（最多256个光纤通道/千兆以太网或最多1280个ESCON通道）应用。

2.1.4 总结

业务永续策略包含了对于关键业务应用的保护，将当机时间缩到最短。像美国联储和SEC对于金融和银行机构的规定以及HIPAA对于卫生保健及议员的规定，都要求其成员机构必须部署业务永续策略，以保护关键业务信息。而诸如同步数据镜像和跨地域分散集群系统提供了在不同地点保持相同数据拷贝的能力，提供了当其中一个地方失效时的数据保护。这些应用都需要一个强壮、高速、高容量的城域DWDM网络来实现在不同地点之间的大容量的数据传送。

使用Cisco ONS 15500系列建立网络架构提供对于业务永续应用的支持，具有以下优点：

认证：通过IBM、EMC、HDS和HP的认证

合作伙伴：IBM全球服务（IGS）、HP服务、AT&T、英国电信、大韩电信、BellSouth、Qwest、SBC、Verizon及其他

业界领先：业界领先的技术和预见，出类拔萃之处：

- 业务密度= 扩展能力
- 保护机制
- 性能监控
- Cisco 端到段的管理

扩展能力：可扩展的解决方案，能够以小空间和低电源功率的需求，满足无论低密度（两个光纤通道/千兆以太网或10个ESCON通道）还是高密度（最多256个光纤通道/千兆以太网或最多1280个ESCON通道）应用

易于管理：客户能够方便及快速地通过使用相同的Cisco管理工具，通过已城域光网络有的Cisco LAN/WAN网络

整合网络：一个网络满足存储、数据、语音和视频需求

协议支持：支持众多的存储和数据协议，包括了ESCON、1Gbps和2Gbps光纤通道及FICON接口。从而为SAN应用的大容量、高可靠性的需求提供支

2.2 在 SONET/SDH 上的光纤通道支持

一个强壮的业务永续计划的核心策略之一，是如何更好地以位置分散的方式实现数据保护和数据安全。自 2001 年以来，基于这种考虑的数据部署，使得通过网络传送的存储流量大量增长。IDC 在 2003 年 7 月的报告中显示，网络化的存储首次超过了整个外接磁盘存储系统市场规模的一半，超过了服务器直连存储。这意味着向网络存储的过渡比预期的更快，同时清楚地显示了跨地域数据分散数据的增长。

随着对于数据的位置分散需求的增加的同时，需要备份和存储的数据量也在增长。在这个报告中，IDC 还指出在 2003 年的第一季度，快速增长的企业存储需求带来存储设备行业容量的 49% 的增长，对于数据保护及数据安全的需求以及数据量的增长，给企业带来了现实的挑战，他们需要寻找能够负担得起的解决方案，来实现在 SAN 孤岛之间的连接并部署地域分散数据恢复解决方案。

这些浮现出来的需求为运营商带来了新的机会——在已有的 SONET/SDH 网络提供大有受益的 SAN 扩展服务。SONET/SDH 以其所具有的可用性、可靠性、可达性和低延迟的特性，成为支持存储协议在短距和长距的理想传输机制。

本文介绍了 Cisco 提供的存储和光产品，并重点对 Cisco ONS 15454 中的 SL 系列基于 SONET/SDH 的光纤通道/FICON 接口卡以及 Cisco MDS 9000 系列产品进行介绍。进而概要介绍他们在通用灾难备份和业务永续应用解决方案上的使用。

2.2.1 Cisco 存储及光产品

Cisco ONS 15000 光产品系列

针对包括运营商和企业用户各方面的需求，Cisco 光产品平台提供了最大的业务密度和种类，从而满足了多种数据、存储和传统网络协议等无可比拟的业务能力。Cisco COMET 体系构架的解决方案设计是以运营商能够以低的网络花费、运营费用和架构复杂性，提高利润为出发点。这些解决方案包括了 Cisco ONS 15454 多业务部署平台 (MSPP)、ONS 15600 多业务交换平台 (MSSP) 和 ONS 15454 多业务传输平台。尽管这些是为满足运营商的可靠性和可用性需求而设计，企业用户也可以使用这些产品来实现光专线和以太网服务。使用 SL 系列板卡能够帮助企业实现 SAN 网络在 SONET/SDH 上的无缝扩展。

Cisco 提供了集成的网元管理系统，Cisco 传输管理器 (CTM) 可以部署、管理和监控边缘、城域及核心网络的所有多业务光平台。另外，还可以通过 Cisco IOS 软件实施对 Cisco ONS 15500 产品及 ONS 15454 ML 系列卡进行管理。通过这些，能够非常方便地融合到使用 CiscoWorks 的管理平台，这降低了培训和管理费用。

Cisco ONS 15454 — 领先的 SONET/SDH 多业务部署平台

自其发布而引发的多业务部署平台市场以来，Cisco ONS 15454 始终保持了市场的领导地位，提供了无可相比的多业务接口，包括以太网、存储和 TDM 等。Cisco ONS 15454 有效地将数据、存储、语音和视频业务汇聚在高密度、高可用的城域网络。

Cisco ONS 15454 通过将 DWDM 集成而推出 Cisco ONS 15454 多业务传输平台，这个产品为 TDM、数据、存储和视频业务提供了多速率的收发器和 Muxponder、灵活的 OADM、为传送数十至数百公里的高级放大器，以及通过增强软件实现的点击配置的网络智能。另外，还有自动光功率管理和增加波长及 A-Z 的波长部署。

Cisco 还以 SL 系列板卡进一步扩大了 ONS 15454 多业务部署平台的市场领导地位，提供对于在 SONET/SDH 网络上的存储网络支持。

Cisco ONS 15454 SL 系列板卡

SL 系列板卡是一个四端口的单槽位板卡，每个端口都支持 1Gbps 或 2Gbps 光纤通道/FICON。它使用可插拔的 GBIC 光接口，为用户提供了更大的灵活性和投资保护。用户接口的流量载荷通过透明的 GFP-T 封装被直接映射为 SDH 的载荷中，这个载荷然后通过交叉连接传送到系统的光中继接口（最高 OC-192），与其他业务一起传送到其他的网络设备。这个板卡进一步完善了光纤通道的传输解决方案。这不仅使 Cisco 能够提供各种距离上的光纤通道传输，还提供在城域、区域和广域网络上对于数据中心和企业存储网络解决方案的端到端的覆盖。

SL 系列板卡可以插入已有的 Cisco ONS 15454 机箱，并能够通过目前的管理系统进行管理。因此，采用这个板卡不需要在固定和运行费用上作很大的投资，就能够提供业务上的很大拓展。对于企业来说，意味着可以将存储网络通过 SONET/SDH 进行扩展，从而满足其业务永续的目标。

SL 系列板卡的重要特性

在 Cisco ONS 15454 多业务部署平台提供对于 TDM、以太网等业务支持的基础上，SL 系列板卡使其增加了对于存储网络的支持，并具有光纤通道速率和密度上的领先地位。

- 支持 1Gbps 及 2Gbps 光纤通道接口，采用低延迟的 GFP-T 映射
- 在保护的 SONET/SDH 传输网络上，支持业界最高的单台设备光纤通道接口密度：在全保护的传输网络上（诸如 4F BLSR OC-192、双 2F-BLSR/USPR OC-192），一个机柜 16 个线速光纤通道

SL 系列板卡可以插入已有的 ONS 15454 多业务部署平台上，保护了用户的投资，提高了回报：

- 使用已有的架构和管理工具降低了固定投资和运营费用
- 提高了业务的提供能力
- 无需对平台的昂贵部件进行升级，诸如交换矩阵等。

SL 系列板卡还将提供对于更高级和更灵活的技术的支持。

Cisco MDS 9000 存储产品系列

Cisco MDS 9000 系列存储产品，将业界最强大和灵活的硬件结构与多层存储智能相结合，降低了存储网络的总拥有成本。使得客户能够构建高可用、可扩展的存储网络，并享受完善的安全保护和统一管理的好处。

完整的多层存储网络产品

Cisco MDS 9000 系列，包含 Cisco MDS9500 系列多层导向器和 Cisco MDS9216/9216i 多层矩阵交换机，可以帮助您建设高可用的、可扩展的存储网络，并为其提供先进的安全性和统一的管理。Cisco MDS 9000 系列可以提供各种智能化网络功能，例如多协议/多传输集成（在一个系统中无缝地集成 FC、iSCSI 和 FCIP），虚拟 SAN（VSAN），全面的安全性，先进的流量管理，完善的诊断功能，以及统一的 SAN 管理。此外，Cisco MDS 9000 系列还可以为嵌入各种智能化存储服务（例如基于网络的虚拟化和复制）提供一个开放的平台。Cisco MDS 9000 系列用一种层次化的方式来实现网络和存储智能，为存储网络的发展开辟了一个新的纪元。

Cisco MDS 9500 系列多层导向器

Cisco MDS 9500 系列多层导向器(Director)提升了导向器(Director)级交换机的标准。Cisco MDS 9500 系列可以提供业界领先的高可用性、可扩展性、安全性和易于管理，让您能够以最低的整体运

营成本部署高性能的存储网络。通过将多种智能化功能分层部署到一个高性能的、协议无关的光通道交换机结构, Cisco MDS 9500系列多层导向器(Director)可以满足大型数据中心存储环境的严格要求。Cisco MDS 9500系列分别提供了六槽、九槽和十三槽三种配置, 每个设备最多可以支持256个1/2Gbps自检测光纤通道端口, 每个机架最多可以支持768个光纤通道端口。利用业界领先的1.44Tbps背板带宽, Cisco MDS 9500系列可以集成未来的10Gbps模块。

Cisco MDS 9216/9216i 多层矩阵交换机

Cisco MDS 9216 的系统架构与 Cisco MDS 9500 系列相同, 它将多层的智能和一个模块化的机型紧密地结合在一起, 从而成为了业界最智能、最灵活的光通道交换机。Cisco MDS 9216 具有 16 个1/2Gbps 自检测光纤通道端口, 而它的扩展槽让用户可以添加 Cisco MDS 9000 系列的任何模块, 因而总共最多可以支持 48 个光纤通道端口。随着存储网络的进一步发展, 用户可以将 Cisco MDS 9000 系列模块从 Cisco MDS 9216 光通道交换机移植到 Cisco MDS 9500 系列多层导向器(Director)中, 从而获得平稳的移植、通用的部件和出色的投资保护。

而 Cisco MDS 9216i 的丰富特性和强大能力更使其成为企业容灾存储网络及 IP 存储连接的理想选择。Cisco MDS 9216i 存储网络容灾交换机在一个设备中同时提供了对光纤通道和 IP 的支持, 它针对运营关键任务的企业存储网络设计, 同时充分发挥了 IP 的高性价比和普遍性, 从而更加强了对于商务业务连续性的支持程度。其提供的关键特性有: 硬件优化 FCIP 特性提高业务持续和灾难恢复能力、虚拟存储网络 (VSAN) 和 VSAN 间路由 (IVR) 增强 SAN 安全性和稳定性、以硬件压缩和 FCIP 加速实现高性能 SAN 拓展、高级流量管理、iSCSI 为以太网连接的服务器提供经济的 SAN 存储拓展等

2.2.3 灾难恢复和业务永续应用

灾难恢复和业务永续计划要求保护关键业务信息并保障对重要数据不间断访问。灾难恢复应用倾向于将数据复制到远程备份地点。备份节点可能是位于同城, 也可能是跨洲的距离。业务永续的最紧要的需求是实时恢复—当灾难发生时, 切换应即时启动, 提供非常快速的恢复。业务永续在中断发生时实现业务应用的保护, 通常用于复制和保护关键信息的应用包括同步和异步复制以及磁带备份。

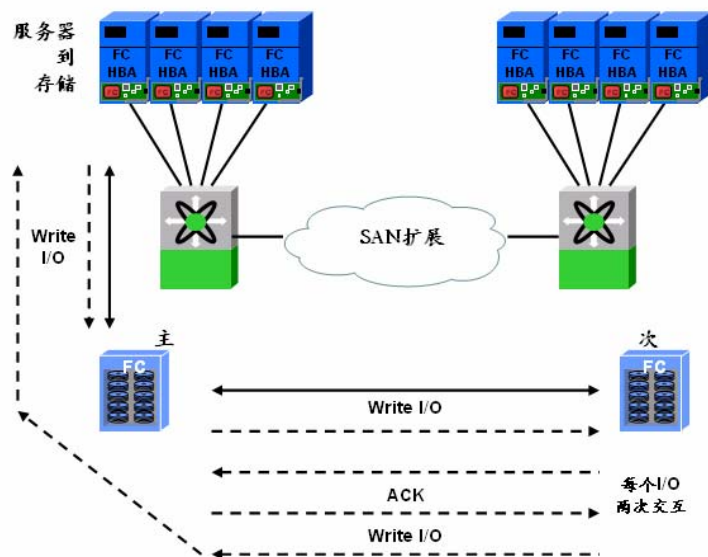
同步复制

同步复制用于保护对可用性有着严格要求的数据和应用。某些应用 (诸如在线交易) 的设计和部署, 必须保证灾难发生时无数据丢失。为实现这个目标, 一个交易必须同步写入主节点和备份节点, 以确保数据库的一致。当一个应用向磁盘写数据时, 响应返回之前, 数据需被复制到远程节点, 只有当一批数据同时被写入两个节点后, 该写输入/输出操所彩绘被响应。因此, 网络带来的延迟会直接影响应用的性能。而带来的延迟通过光速和距离间的函数计算:

延迟=5 微秒 X 往返距离 *2

例如, 两个相距 50 公里的节点的延迟为 1 毫秒。而为完成输入/输出操所, 需要两次往返。为避免复制的影响, 存储阵列厂商制定了对于同步复制应用的距离限制, 通常情况下这个距离为 100 公里左右。

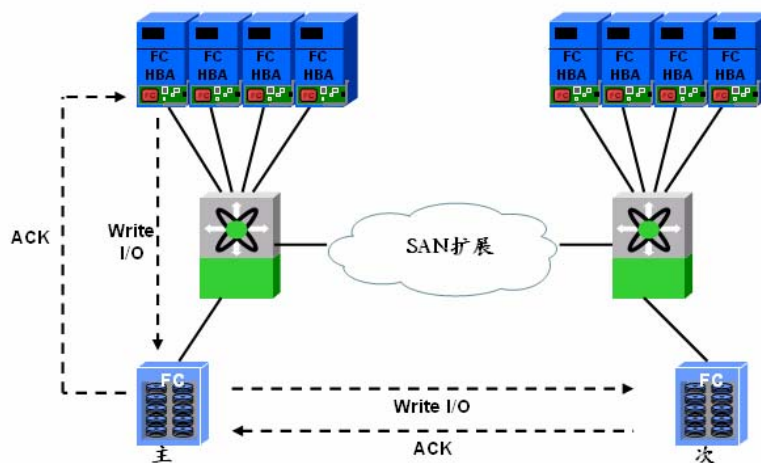
图 2-11 同步复制



异步复制

如果在灾难发生时，可以容忍一定数量的业务信息的丢失，就可以考虑异步复制技术。异步复制提供了非常好的保护，但是某些交易在灾难发生时丢失。在异步复制下，一旦向主存储阵列的写操作完成，输入/输出操作就完成了。服务期并不等待向其他存储阵列的写操作的完成。异步复制没有距离限制，能够跨越数千公里或是更远。

图 2-12 异步复制



灾难恢复和业务永续应用的需求

在设计广域网络前，对于每个应用的需求及其在网络上如何运行的了解非常重要。现实有大量的应用，表 2-2-1 归纳了具有严格需求的主要应用。城域/广域网应能够满足每个应用的需要以达到最大效果。

表 2-2-1 应用需求

应用	带宽	延迟	备注	异步/同步
磁带备份	典型情况每个磁	<1-5 毫秒	延迟敏感，利用	同步

	带驱动器 10-15 MB, 最高 40MB		SCSI 协议提供超时和差错恢复。 注：一旦会话中断, 会话的所有备份会丢失。	
磁盘镜像	各存储阵列不同, 典型情况每个存储阵列最大 50MB	同步<1-5 毫秒 异步复制可容忍较高延迟(100ms)	同步应用对延迟非常敏感。异步应用敏感性比较低	同步及异步
文件存取	与操作系统相关		取决于操作系统及其上的应用	

Cisco 端到端解决方案以 Cisco 的平台作为构件（包括 Cisco ONS 15454, MDS 9000 和 ONS 15540/15530），支持几乎所有市场主导的灾难恢复和业务永续应用。下面的部分将讨论一些应用以及支持这些应用的 Cisco SAN 扩展解决方案的典型架构。

远程磁带备份应用

在一个企业网络中, 某些节点（远端办事处或小型办公室）会有一个小型的 SAN 用以将服务器连接到存储阵列。而通过广域网对这些服务器的备份和恢复是灾难备份操作的基础。在广域网上扩展磁带备份, 在保障有效性上有严格的需求, 这些需求包含无数据丢失、低延迟和抖动、链路监控及高安全性。

慢速的广域网链路会增加备份的时间, 而且在分配的时间片（窗口）内完成备份变得不可能。对于磁带备份应用来说, 由于延迟需求可以预计, 因而距离不是主要的限制。为保证磁带备份尽可能有效, Cisco 建议保持一个特定的速度, 以保证连续的数据流送达磁带, 当磁带能够接收到稳定的数据流的情况下, 备份的性能可以达到最佳。通过广域网络的磁带备份是非对称的。图 2-13 展示了 Cisco 的将两个提供备份的远程节点与主节点互连的解决方案。磁带备份数据传送的非对称特性对于 SAN 网络扩展设计, 提出了特别的挑战。而磁带“管道”技术提供了帮助, 使磁带机可以位于数千公里远以外, 从而使磁带备份成为业务永续和灾难备份应用的重要部分。这些是通过利用缓存和纠错机制达成的, 其概念上与伪装类似—尽管服务器和磁带控制器相隔很远, 但在运行表现上就好像是他们在一起。磁带管道技术放宽了 SAN 扩展设计的技术压力。

图 2-13 磁带备份

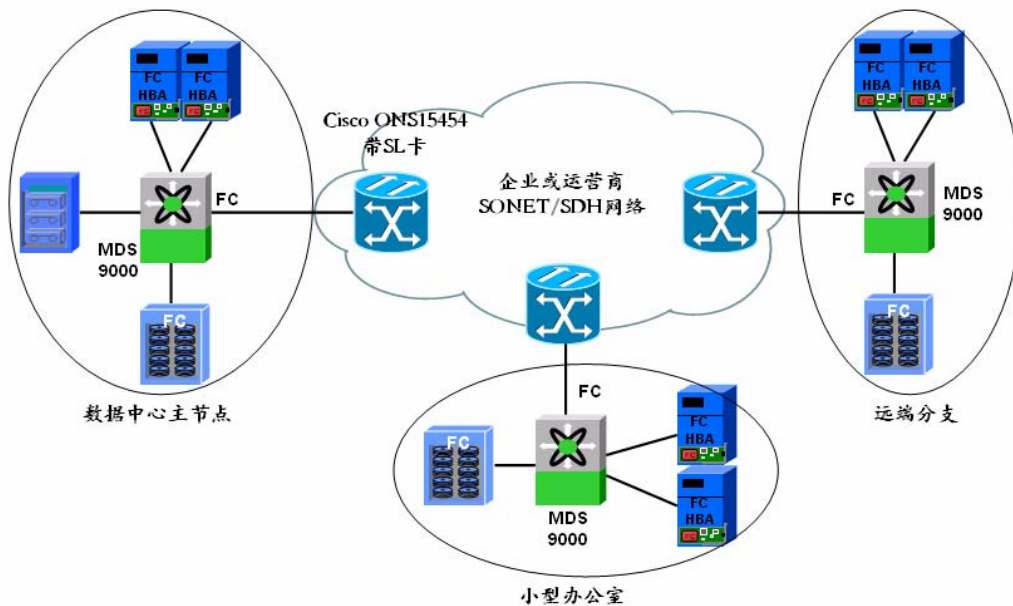
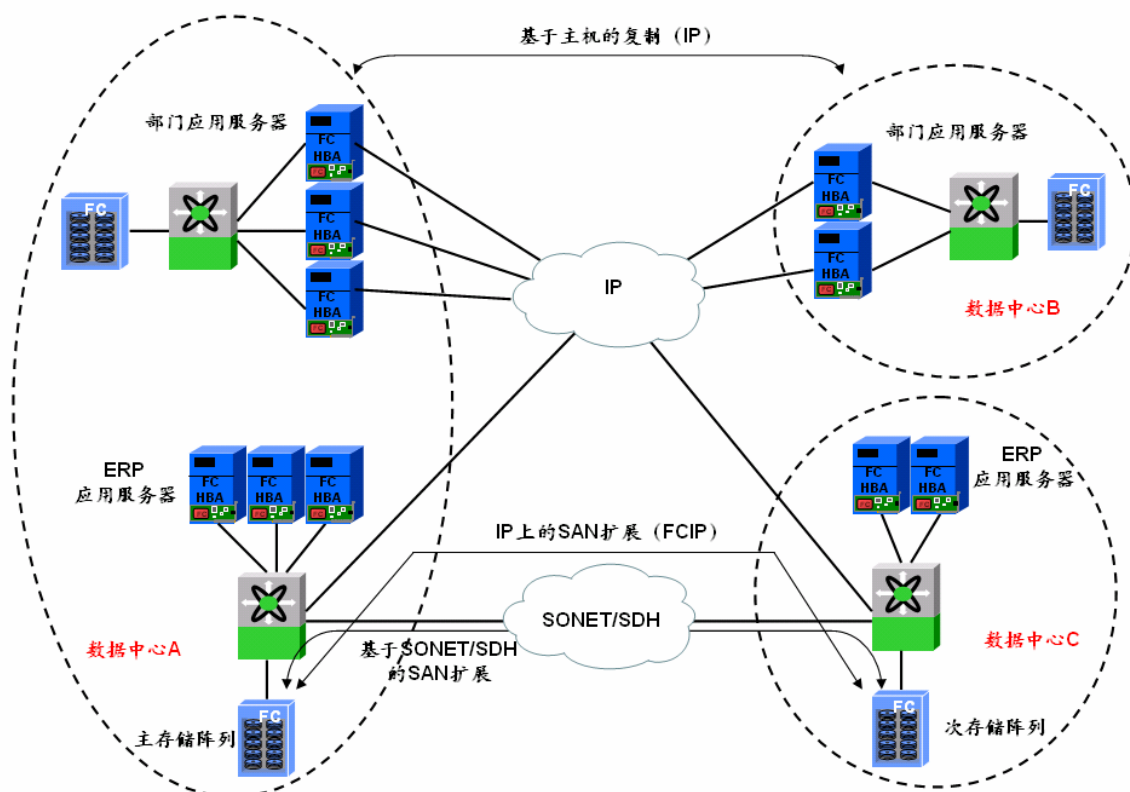


图 2-13 所示的 Cisco 解决方案提供了所需的部署、管理、带宽优化以及对磁带备份应用非常关键的性能。Cisco 的解决方案提供了对带宽需求增长的扩展能力，从而始终保持对这些业务的服务质量保障。

远程磁盘镜像

磁盘镜像是一种同时将数据写入两个不同磁盘的技术，这样就确保了在主磁盘失败时能够即时切换到第二个磁盘上。防止或是将数据丢失的可能降到最小。磁盘镜像在数据库应用中经常会用到，出于灾难恢复和业务永续的要求，企业需要位置分散镜像技术，即远程磁盘镜像。混合了光纤通道和千兆以太网的典型远程磁盘镜像网络架构（层级网络架构）如图 2-14 所示。

图 2-14 广域网上的层级数据中心网络（磁盘镜像应用）



光纤通道 SAN 扩展设计—基于存储的复制

Cisco SAN 扩展解决方案支持大多数的存储阵列厂商的基于存储阵列的复制。这里的 SAN 扩展网络的设计原则和方式完全一致，提供了冗余和在网络故障时的快速收敛。这里描述两种解决方案设计。第一种设计凭借冗余的光纤通道交换机构成稳定的矩阵。第二种设计则使用 MDS 9000 的高可靠性和冗余特性。

在这部分提出的设计适用于各个不同的存储厂家的复制应用，包括 EMC SRDF、HDS Truecopy、HP DRM 及 IBM PPRC。

EMC SRDF(对称远程数据设备)

EMC SRDF 是一种独立于主机的联机镜像数据解决方案，它将生产节点的数据复制到物理分立的目的地 Symmetrix 系统。这个 SRDF SAN 扩展很大程度上降低了影响业务永续的主要作业的风险，这些作业包括：灾难恢复、备份和定期维护、数据中心迁移及整合，以及测试。

HDS Truecopy

HDS Truecopy 提供了一个基于存储的灾难备份硬件解决方案，它实现了快速精确的系统恢复。TrueCopy 操作一经建立，会自动维护复制的数据拷贝，以为备份和灾难恢复之用。在正常的 TrueCopy 操作中，主卷处于激活状态，为所有主机提供读写输入/输出服务。一旦发生灾难或系统故障，第二份数据拷贝会被快速地调用，从而达到具有很高程度数据完整性的应用恢复。TrueCopy 还可用于数据复制和迁移作业。

HP DRM

DRM 提供的容灾存储解决方案通过硬件冗余和一定距离上的两个节点的数据复制实现。多个主机可以连接到一个或多个共享的存储子系统，基本的 DRM 配置包含了两个节点，一个发起端(Initiator)和一个目的端(Target)。发起端节点作为数据处理的第一承载，而目的端节点则用于数据复制。在发起端有数据处理时，数据会被复制或镜像到目的端。如果任何一端的部件出现问题，DRM 将会切换

到同节点的冗余部件继续运行，例如，如果两个节点间的双冗余光纤通道链接失效，DRM 会切换到其他的链接。而当发起端发生严重故障（灾难）时，数据处理在目的端的得到恢复，那里的数据是完整的。这种流程叫做端点切换（Site Failover）。而在发起端节点故障排除之后，通过叫做端点回切（Site Failback）的流程，数据处理将移回发起端。

IBM 对等远程复制（PPRC）

对等远程复制（PPRC）是企业存储服务器（ESS）的一项功能，将一个节点（通常称作应用节点）的应用系统数据影射到第二节点（称作恢复节点）。应用节点中的ESS内承载数据的逻辑卷叫做主逻辑卷，而恢复节点中的相应逻辑卷则称为次逻辑卷。在主服务器和次服务器间通过ESCON链路连接。

架构和网络设计

灾难恢复应用需要强壮的网络架构，其带来的延迟需要非常小（至少对于同步复制如此）。SONET/SDH网络可以在城域或更长的距离用作存储传送，SONET/SDH的收敛时间小于50毫秒。其环状结构使得在一条光纤切断时还有另外一条路径。图2-15展示了两个光纤通道交换机连接到Cisco ONS 15454 SONET/SDH平台的架构。

图2-15 EMC SRDF应用网络架构

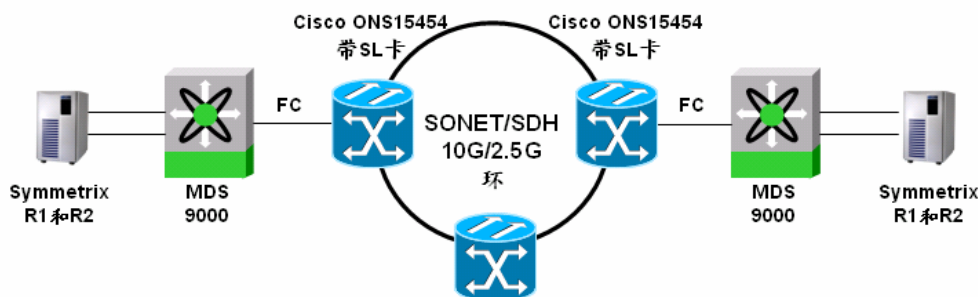
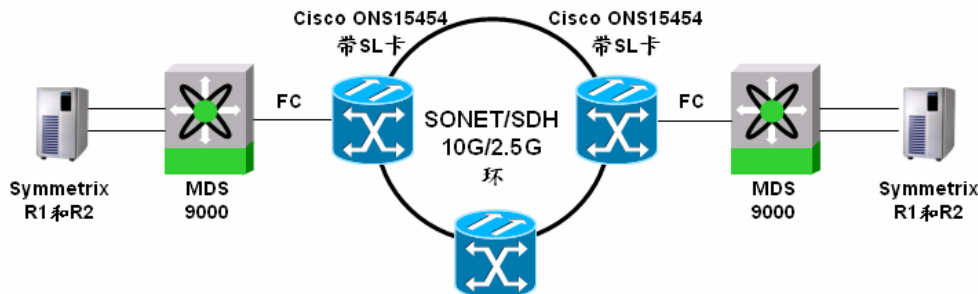


图 2-16 则展示了两套 EMC Symmetrix 存储阵列通过 SAN 扩展系统进行连接。每台 Symmetrix 设备连接到两个光纤通道端口上，从而在两个 Symmetrix 设备之间构成双光纤通道路径。在 Cisco ONS 15454 上，SL 系列板卡提供到 Cisco MDS 9000 光纤通道交换机的连接。在两个 SONET/SDH 端节点之间部署 SONET/SDH 电路，这是全保护电路。

图 2-16 SRDF 应用网络架构（非冗余）



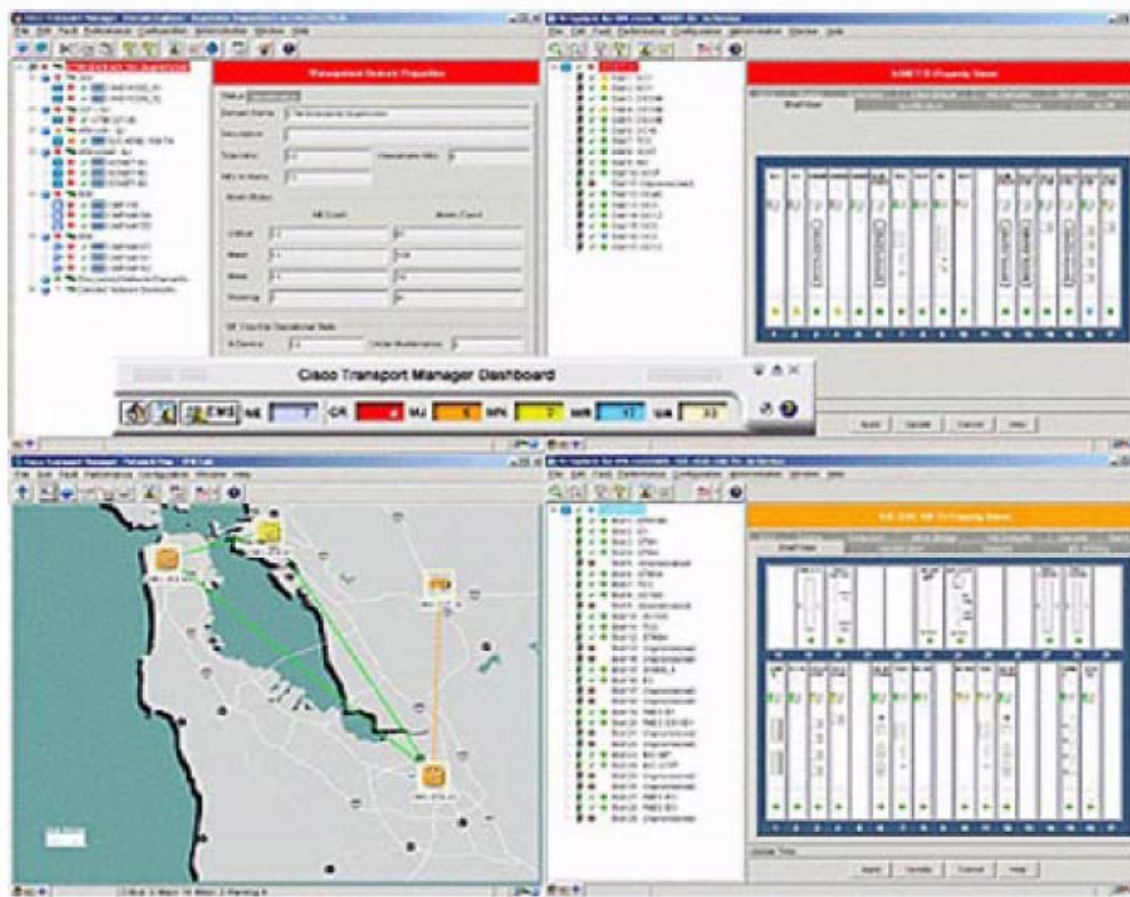
图中，两边都创建了 R1 卷和 R2 卷，这就是说，流量会在双方向上产生。在这里，两边的 R1 生成信息并复制到另外一个地点（R2）。这对于在两地负载共享的企业业务应用来说比较常见。

存储网络管理

随着存储网络环境得不断增长，对于完整的存储网络管理服务的需求会显现出来。Cisco ONS 15454 光平台和 Cisco MDS 9000 系列产品提供了许多的特性和服务，能够很好地支持这些应用。

Cisco 传输管理器（CTM）是为 Cisco ONS 15000 光系统设计的集成网元管理系统，简化了管理，提高了 Cisco ONS 15454 的部署速度。CTM 提供了对于 Cisco 光网网元、子网和网络的配置、故障、性能及安全管理的高级管理功能。支持 SONET、SDH、DWDM、光纤通道/FICON 及以太网的管理，同时为 OSS 提供了开放的接口，在可靠性和可扩展性上得到验证。图 2-17 中展示了 CTM 的典型的外观。

图 2-17 CTM 图形用户界面



Cisco 矩阵管理器，包含的对于网远及矩阵的管理，提供了对于 Cisco MDS 9000 系列多样的管理能力。网元管理使网管人员能够配置交换机并跟踪统计和事件。Cisco 矩阵管理器提供了对于多个个被构成的矩阵的管理能力，它给予矩阵拓扑建立管理应用。当被调用时，拓扑搜寻过程就开始了，使用从 Cisco MDS 9000 系列设备中获取的信息（包括 NameServer 注册和 Fibre Channel-GS-3 矩阵配置服务器程序等信息），矩阵管理器能够重新生成矩阵拓扑并以客户化的方式展示出来。Cisco MDS 9000 系列交换机本身还带有一系列面向矩阵的管理工具和手段，可以用于搜寻并监控整个互连的矩阵。这些工具为光纤通道 Ping 和 Traceroute、SCSI 目标发现等，支持光纤通道-GS-3 和光纤通道中继的所有服务。通过使用 Cisco 矩阵分析器，用户能够从交换机种抓取控制信息并进行解码，而无需中断任何连接，也无需一定要在需要分析的设备在现场。

2.2.4 总结

本文展示了Cisco端到端的存储产品，并重点介绍了用于在SONET/SDH上承载光纤通道/FICON的ONS 15454的SL系列板卡，以及Cisco MDS 9000系列。

对于SAN扩展以及Cisco存储和光产品，请参考以下资料：

- IBM Redbooks “IBM SAN Survival Guide”
- http://www.emc.com/products/systems/symmetrix.jsp?openfolder=storage_systems.com
- 产品介绍和手册 — Cisco ONS 15454 SL系列
- 产品介绍和手册 — Cisco MDS 9000

2.3 以 FCIP 实现 SAN 扩展

本章提供了关于 FCIP 协议的背景信息，介绍了在设计某个特定的部署方案时需要考虑的性能因素。本章包括下列主题：

- 概述
- 使用压缩
- 使用写加速

2.3.1 概述

本节介绍了 FCIP 的好处和部署 FCIP 所需要的组件。本节包括下列主题：

- 为什么要使用 FCIP？
- 解决方案拓扑
- 必要的组件

为什么要使用 FCIP？

利用 FCIP，您可以将光纤通道 SAN 扩展到任何存在 IP 网络和具有一定带宽的地方。您可以在只能用 IP 传输数据的城域、园区或者大陆间距离上使用 FCIP。

对于短距应用，例如在一个数据中心内部，SAN 可以借助多模光纤，在光网连接的基础上进行扩展。随着距离的增加，例如在某个大型数据中心或者园区中，您可以利用 Cisco MDS 9000 上的粗波分复用（CWDM）SFP 光网终端，使用单模光纤。在需要进行整合或者光纤短缺时，可以使用 CWDM。在城域或者地区范围内，密集波分复用（DWDM）是首选的传输方式。在需要较高的整合密度，或者需要汇聚 FICON、ESCON 和 10GE 数据中心连接时，也可以使用 DWDM。

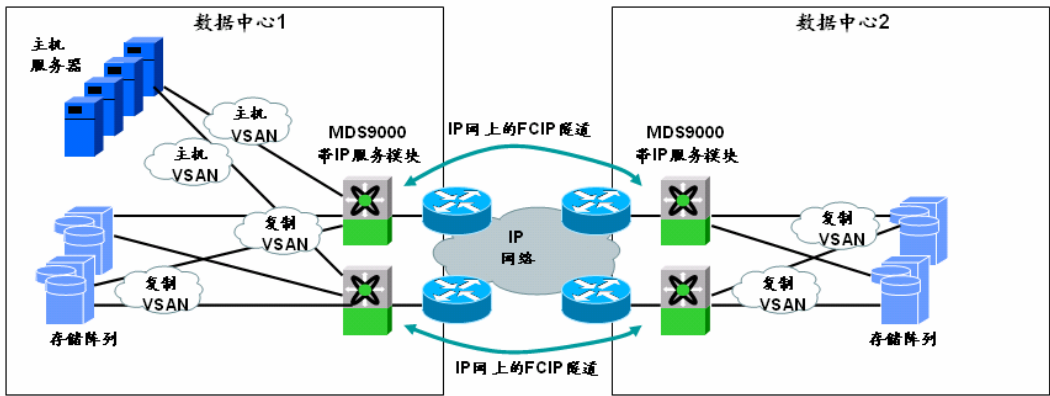
FCIP 在 SAN 扩展方面的典型应用包括：

- 同步数据复制——利用专门的复制软件，在智能存储阵列之间实现零恢复点目标（RPO）。在进行同步复制时，网络延时会影响磁盘 I/O 服务时间和应用性能，因此必须考虑各种影响延时的因素。这些因素包括距离；路由器和交换机的存储和转发延时、流量拥塞以及 QoS 设置。光网 SAN 扩展通常更适合同步复制的需要，因为它对延时的容忍程度较低。
- 异步数据复制——利用专门的复制软件，在智能存储阵列之间运行低恢复点目标应用。网络延时对应用性能的影响远远小于同步复制。您可能需要调整复制软件或者上层协议，以确保对 FCIP 连接的充分利用。
- 远程磁带库——利用磁带或者磁盘为灾难恢复进行远程备份。磁带应用通常只允许保留一个尚未完成的输入/输出（I/O）操作，从而限制了远程连接的吞吐率。在这种情况下，您可以利用写加速和压缩功能优化吞吐率。
- 发起主机到远程存储池——实现对位于另外一个地点或者数据中心的、与 FC 连接的存储阵列池的访问。

解决方案拓扑

[图 2-18](#) 显示了一个数据复制拓扑示例。在该图中，两个包含主机服务器和存储阵列的数据中心通过虚拟 SAN（VSAN）连接到 MDS 9000 系列交换机或者导向器。存储阵列通过 MDS 9000 IP 存储服务模块，连接到通过 FCIP 连接或者隧道扩展的复制 VSAN。

图 2-18 使用 FCIP 的数据复制拓扑



必要组件

您可以在表 2-3-1 中列出的三个思科平台上部署 FCIP。

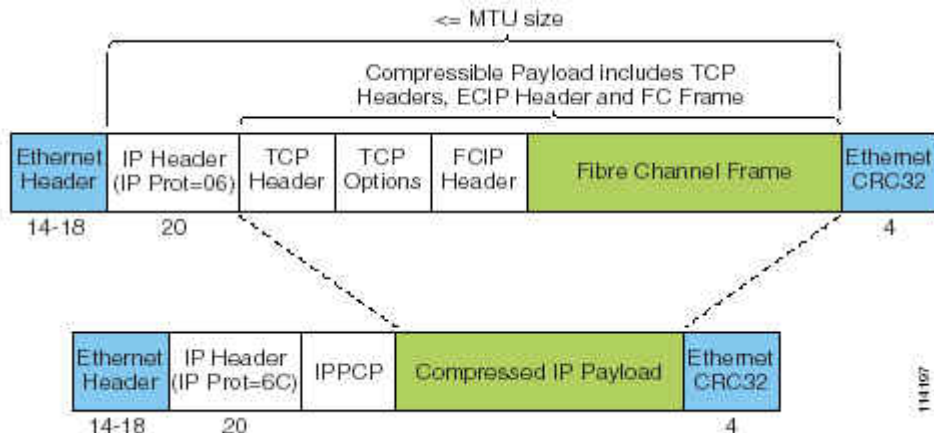
表 2-3-1 思科 FCIP 平台和光纤通道端口类型

思科 FCIP 平台	所支持的端口类型	IP 端口类型和密度	FC 端口类型和密度
MDS 9000 IP 存储服务模块	B_Port , E_Port , TE_Port	每个模块配有 8 个千兆位以太网端口 每个机箱 (MDS 9505) 最多 6 个模块 (总共 48 个端口) 每个接口最多 3 个 FCIP 隧道 (每台 MDS 9509 可以支持 144 个 FCIP 连接) 基于 LC 连接器的 SFP 连接 (MM 或者 SM 光纤)	在速率为 1 或者 2Gbps 时, 最多 224 个端口 (MDS 9509)
针对 Cisco 7200 和 7400 的 PA-FC-1G 端口适配器	B_Port	依赖于 WAN 端口适配器 (从 T1/E1 到千兆位以太网) 每个 PA-FC-1G PA 支持 1 个 FCIP 隧道 每个配有 NPE-400 和 C7200-I-O 控制器的 7200 机箱最多支持 2 个 PA-FC-1G (建议为每个机箱使用 1 个 PA-FC-1G)	在速率为 1Gbps 时, 每个 PA-FC-1G 支持 1 个端口
SN 5428-2 存储路由器	B_Port	两个千兆位以太网端口 基于 LC 连接器的 SFP 连接 (MM 或者 SM 光纤)	在速率为 1 或者 2Gbps 时, 支持 8 个 FC 端口

2.3.2 使用压缩功能

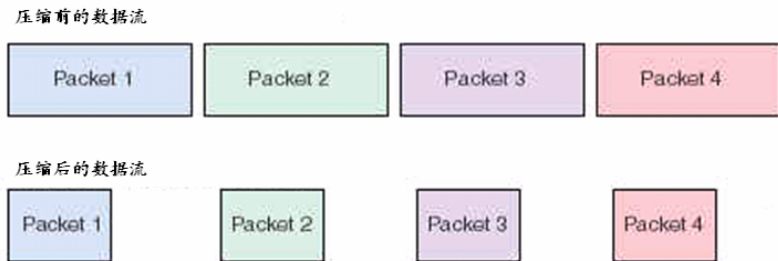
MDS 9000 为 FCIP 连接集成了一种可选的软件压缩功能。它建立在 IP 载荷压缩协议 Lempel-Ziv-Stac (IPPCP LZAS)¹ 的基础上。压缩是针对 IP 载荷, 对每个分组进行的 (如图 2-19 所示)。因为压缩发生在分段之后, 所以以太网帧数在压缩前后不会发生变化, 只是每个帧的尺寸会变小 (如图 2-20 所示)。

图 2-19 LZS IP 载荷压缩



大帧的压缩率要高于小帧。这是因为 IPPCP LZS 压缩算法为每个分组记录了一个 2KB 的令牌历史信息。压缩历史会随着每个帧而重新开始记录。与任何压缩机制一样, 实际的压缩率完全取决于数据流的内容。实验室测试已经证明, 一个包含无效数据 (零) 的、MTU 为 2300 的全尺寸帧的压缩率比 MTU 为 1500 的全尺寸帧高 30%。另外一种极端情况——即数据流包含的全部是随机数据——通常无法压缩, 因此压缩前后的分组长度将不会发生变化。实际数据流的压缩率将介于这两种极端情况之间。

图 2-20 压缩前后的分组



经过加密的数据流通常不能进行压缩, 因为它会导致数据流成为随机排列的比特流。如果需要进行加密和压缩操作, 必须先压缩数据, 再进行加密。接收端则应当先解密数据, 再解压数据。例如, 某个经过压缩的 FCIP 数据流可能会被发送到一台配有 VPN 适配器模块 (SA-VAM2) 的 Cisco 7200, 以 OC-3 数据速率进行加密; 或者被发送到一台配有 VPN 服务模块 (VPNSM) 的 Catalyst 6500, 以千兆位数据速率进行加密。

压缩和可用带宽

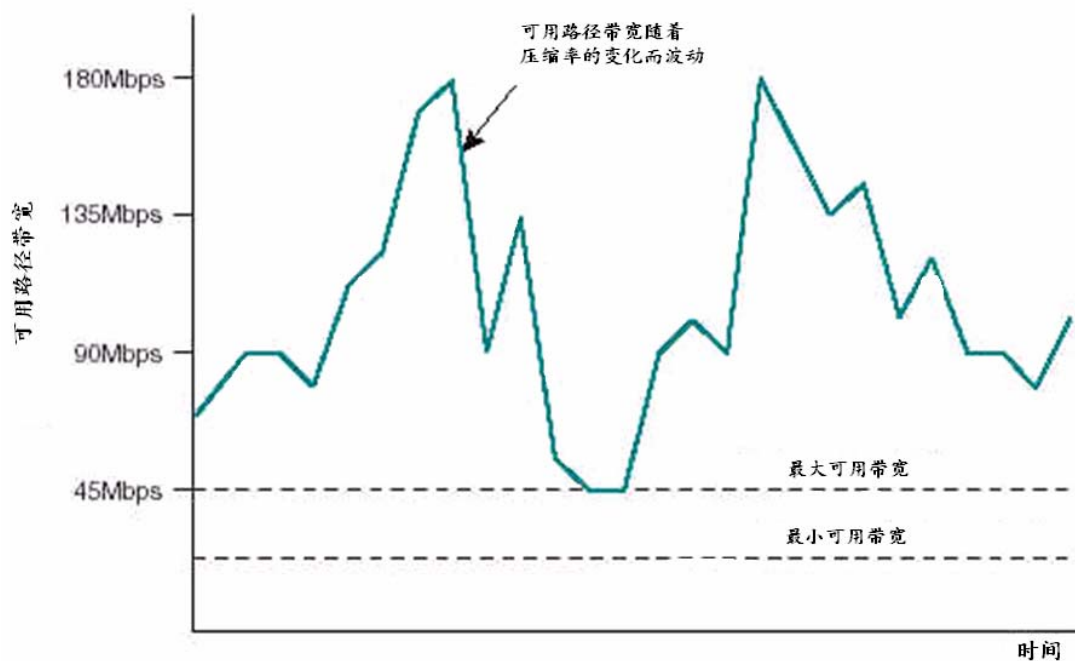
如果对某个 FCIP 连接进行加密, 该连接将会对 TCP 显示为一个可变带宽路径 (如图 2-21 所示)。路径带宽将介于以下两种极限之间:

- 低带宽极限等于未压缩的路径带宽, 即等于实际路径带宽。这与压缩率为 1:1 时的高极限相同, 这时所传输的数据通常被视为无法压缩的随机字符串。
- 高带宽极限等于路径带宽的 n 倍。 n 表示分组的压缩率 (在计算未被压缩的以太网和 IP 报头之

后)。

压缩是针对每个分组进行的，压缩率会根据原始数据流的不同，随着不同的分组而发生变化。无效数据和重复字符串的压缩率较高；随机字符串——例如经过加密的分组——则根本不能压缩。

图 2-21 压缩对于可用 FCIP 路径带宽的影响



压缩模式和速率

IP 存储服务模块允许为每个 FCIP 接口选择三种压缩配置模式：

- 无压缩——不对该接口上的 FCIP 分组进行压缩。
- 高吞吐率——对该接口上的 FCIP 分组进行压缩，但是通过较低的压缩率获得较高的吞吐率。
- 高压缩率——对该接口上的 FCIP 分组进行压缩，同时通过较低的吞吐率获得较高的压缩率。

图 2-22 利用 Canterbury Corpus 测试套件在 MDS 9000 上获得的 FCIP 压缩率

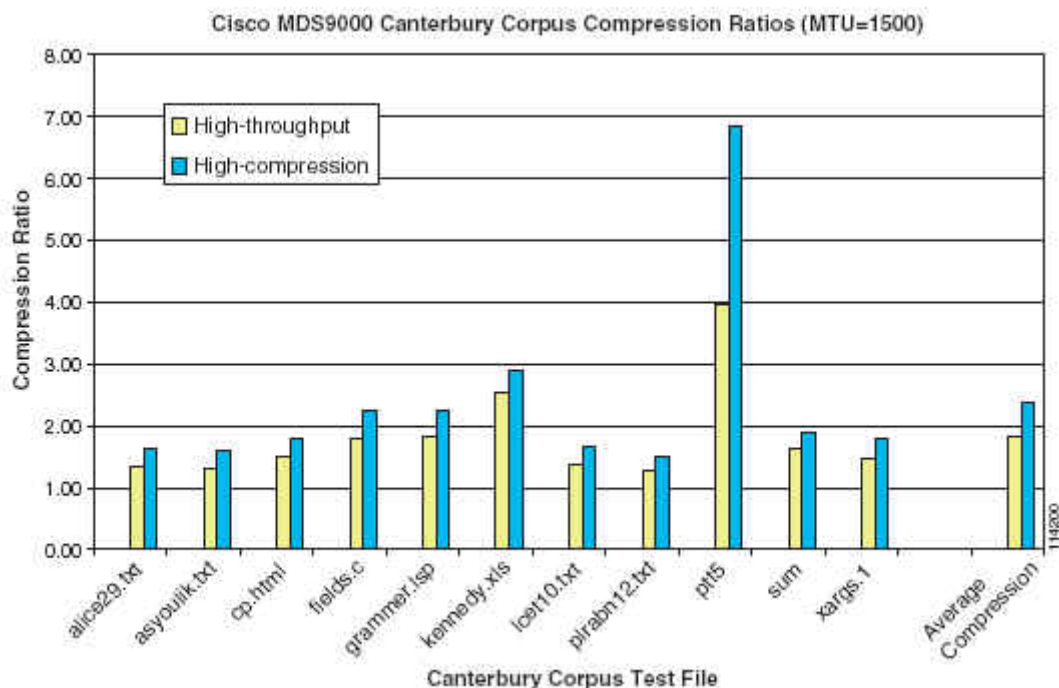


图 2-22 显示了在使用 Canterbury Corpus 测试套件（用于对比压缩率的标准测试套件）的情况下，在两种压缩设置中使用缺省的 MTU 大小（1500 字节）时获得的压缩率。

图 2-23 每个 FCIP 千兆位以太网接口在使用不同的 MTU 和压缩率时的吞吐量

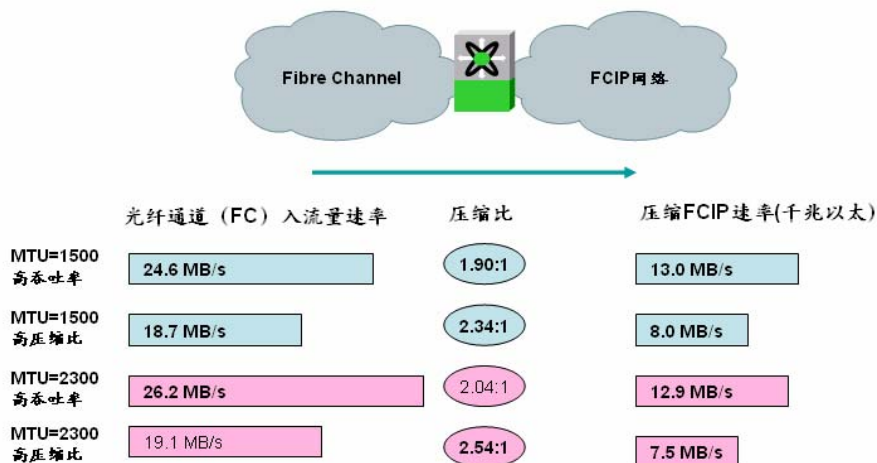


图 2-23 显示了 MDS 9000 IP 存储服务模块针对典型流量的相对性能。最佳的压缩率和吞吐量可以通过巨型帧（MTU=2300）获得。注意，越容易压缩的数据的吞吐量越高；随机性越高或者越难压缩的数据的总体吞吐量越低。表 2-3-2 显示了近似的性能对比。

表 2-3-2 压缩性能对比

	下界				上界		
MTU	压缩机制	光纤通道 输入 (MB/s)	GigE 压缩 (MB/s)	压缩率	光纤通道 输入 (MB/s)	GigE 压缩 (MB/s)	压缩率

1500	高吞吐率	~18	~17	~1.05 :1	~40	~6	~6.7 :1
	高压缩率	~13	~9	~1.4 :1	~38	~5	~7.6 :1
2300	高吞吐率	~19	~17	~1.1 :1	~52	~5	~10 :1
	高压缩率	~12	~8	~1.5 :1	~46	~4	~11 :1

2.3.3 使用写加速功能

本节提供了关于使用 SAN-OS 1.3 提供的写加速功能的背景信息和设计建议。本节包括下列主题：

- 概述
- 超过两次往返的写入数据交换
- 写加速对于 SCSI 写入和复制流量的影响
- 在使用写加速时的预期性能增益
- 何时应当使用写加速
- 估计尚未处理的 I/O 个数
- 使用写加速的考虑因素和指导原则

写加速概述

写加速是 SAN-OS 1.3 中推出的一种可配置的功能。您可以利用 IP 存储服务模块，将其用于 FCIP SAN 扩展。它是一种 SCSI 协议伪装机制，旨在通过缩短 SCSI 写入输入/输出操作（I/O）和远程复制写入 I/O 的总体服务时间，提高应用性能。

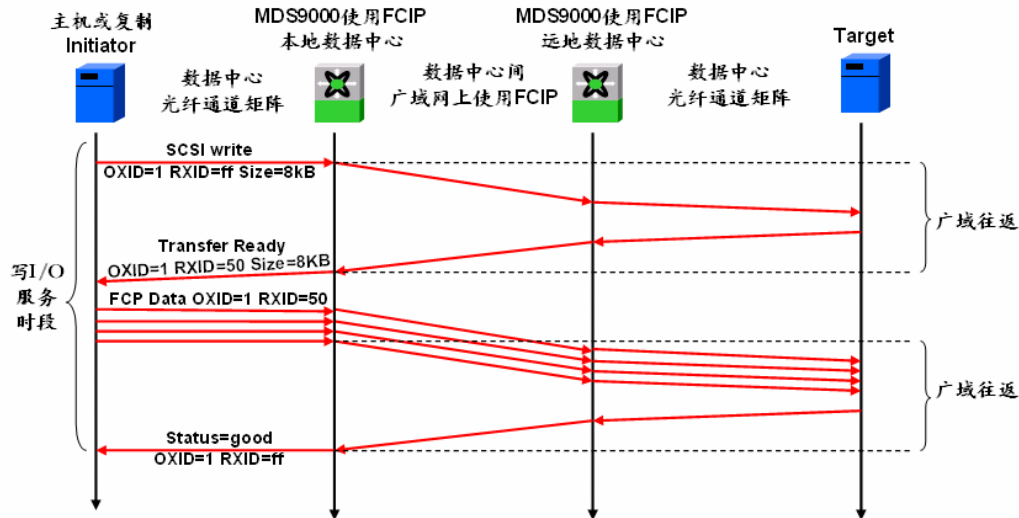
写入操作适用于下列 FCIP SAN 扩展应用：

- 数据中心之间的距离和延时会降低同步复制的性能，影响总体应用性能
- 复制吞吐率受到高层协议的限制，而底层 FCIP 和 IP 传输没有得到充分的利用。
- 在使用远程磁带备份时，距离和延时会严重影响磁带写入性能，因为磁带通常只允许一个尚未处理的 I/O。在这种情况下，写加速可能会有效地将所支持的传输距离增加一倍，或者将传输速率增加一倍。
- 在数据中心之间使用发起主机到磁带目标，尤其是在共享数据群集分布在数据中心之间，而且一个主机必须将数据写入到一个远程存储阵列时。

写加速所带来的性能提升幅度通常接近 2 :1，但是具体幅度则随情况而定。在某些使用较大的写入数据区块的情况下，性能提升幅度可能会超过两倍，因为减少了为每次写入操作传输就绪命令的次数。

写加速可以把每个 SCSI FCP 写入 I/O 的 FCIP WAN 往返次数减少到一次。大部分 SCSI FCP 写入 I/O 交换都需要在发起主机和目标阵列或者磁带之间往返两次以上。图 2-24 显示了一个 8KB 的写入数据交换过程。在线交易处理数据库应用经常需要使用这种过程。

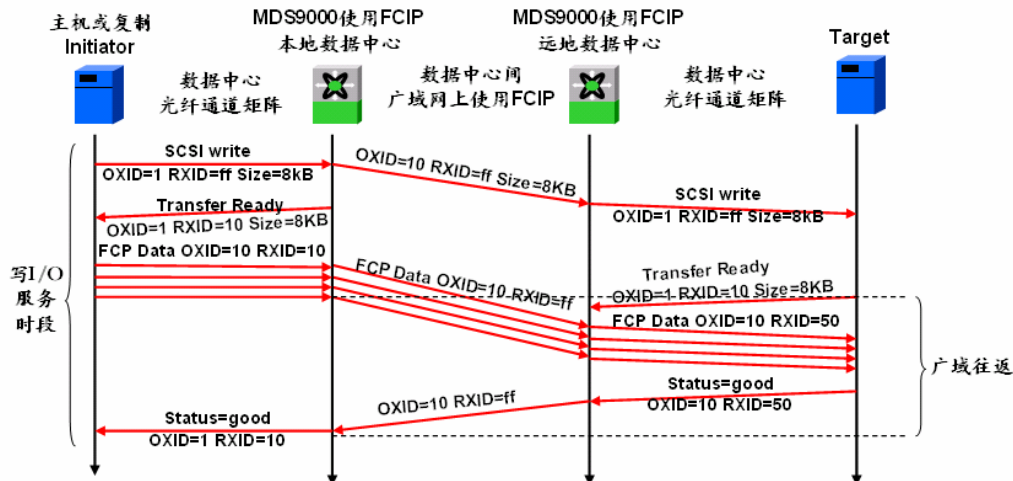
图 2-24 不采用写加速的短程 SCSI 写入交换（往返 WAN 两次）



在不使用写加速功能的情况下，一个典型的 SCSI FCP 写入的流程包括：

1. 发起主机发出一个 SCSI 写入命令，其中包括写入数据的总尺寸（在这个例子中为 8KB）；一个源交换标识符（OXID），以及一个接收端交换标识符（RXID）。
2. 目标返回一个 FCP 传输就绪命令。这将让发起方知道目标愿意在下一个写入序列中接收多少数据，以及目标为 RXID 分配的值（在这个例子中为 50）。
3. 发起方按照之前的传输就绪命令中指定的数量发送 FCP 数据帧。
4. 如果 I/O 成功完成，目标返回一个 SCSI “状态=良好” 帧。

图 2-25 采用写加速的短程 SCSI 写入交换（往返 WAN 一次）



如图 2-25 所示，在使用写加速功能时，通信流程为：

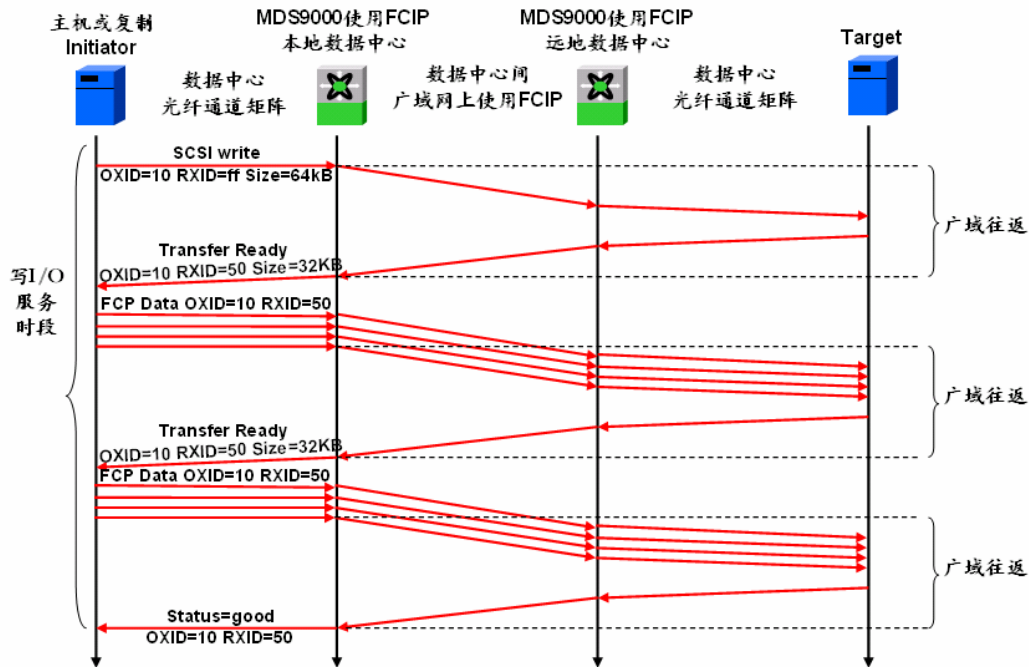
1. 在发起方发出一个 SCSI FCP 写入命令之后，MDS 9000 立即向发起方返回一个传输就绪命令。这个传输就绪命令包含一个本地分配的 RXID。
2. 在远端，并不知道写加速功能的目标返回一个传输就绪命令。其中的 RXID 被保存在一个本地表中。
3. 当发起方发出的 FCP 数据到达远程 MDS 9000 时，每个帧中的 RXID 值都将被替换为本地表的相应 RXID 值。

SCSI “状态=良好” 帧的 RXID 在本地 MDS 9000 中被第一步中设置的值替换。

往返次数超过两次的写交换

每个 FCP 数据队列中发送的数据量取决于目标之前发送的传输就绪响应。

图 2-26 包含多个传输就绪响应的 FCP SCSI 写入流程



例如，某些存储阵列只接受尺寸不超过 32KB 的 FCP 数据队列。在这种情况下，一个 64KB 的写入操作可能需要往返三次，从而导致更多的写入 I/O 服务时间延时，如图 2-26 所示。

图 2-27 将写加速功能用于需要多个传输就绪响应的写入交换流程

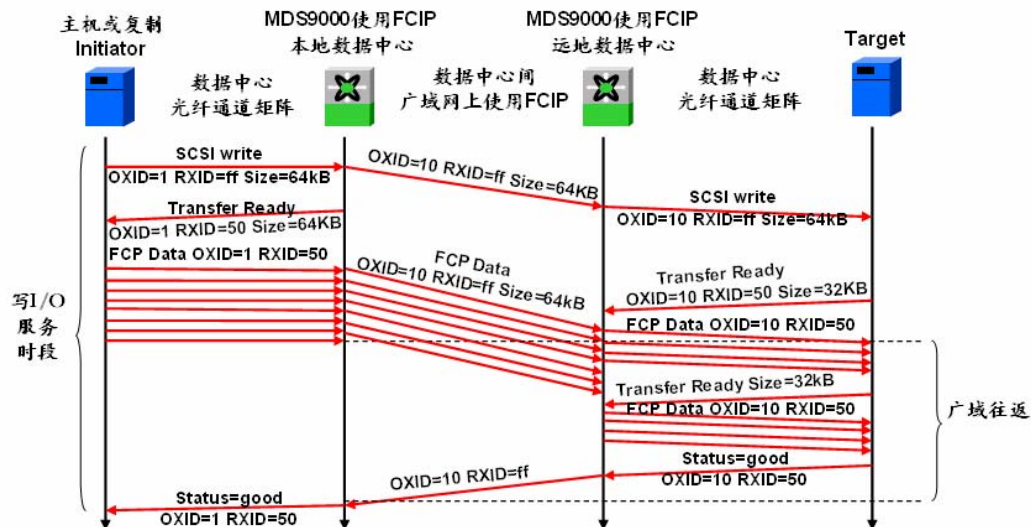


图 2-27 显示了使用写加速功能时的协议流程。在这种情况下，尽管目标一次只能接收 32KB，但是本地 MDS 9000 会针对总共 64KB 发出一个伪装的传输就绪命令。这可以将 WAN 往返次数从三次减少到一次。在这种情况下，写加速功能可以将性能或者传输距离增加三倍。

表 2-3-3 列出了写加速功能对 SCSI 写入和复制流量的影响。我们在两个通过一条距离/延时可变的 FCIP 路径隔离的存储阵列中的 LUN 之间进行了一系列同步复制测试。磁盘 I/O 的测试采用的是 8Kb 写入操作。

表 2-3-3 写加速的影响

FCIP 连接特性		不采用写加速功能			采用写加速功能			写加速性能比
RTT 延时 (ms)	FCIP 路径带宽	每秒 I/O 次数	吞吐量 (MB/s)	平均 I/O 响应时间 (ms)	每秒 I/O 次数	吞吐量 (MB/s)	平均 I/O 响应时间 (ms)	
0	155Mbps	235	1.8	4.3	366	2.7	2.88	1.5
2	155Mbps	125	0.98	8.0	213	1.67	4.7	1.7
4	155Mbps	84	0.63	12.3	150	1.17	6.7	1.9
6	155Mbps	62	0.49	15.9	116	0.91	8.6	1.9
8	155Mbps	49	0.38	20.4	89	0.70	11.0	1.8
10	155Mbps	35	0.29	26.7	78	0.62	12.7	2.1
15	155Mbps	29	0.23	34.7	55	0.43	18.3	1.9
20	155Mbps	22	0.18	43.8	44	0.34	22.6	1.9
30	155Mbps	16	0.12	66	29.4	0.23	33.5	1.9

在使用写加速时的预期性能增益

图 2-28 到图 2-31 显示了对基于单个尚未处理的 I/O 的同步复制使用写入复制时的预期性能增益。每对图形显示的分别是：

- 在启用和禁用写加速功能时的 I/O 响应时间（单位：毫秒）。每秒 I/O 次数（IOPS）与 I/O 响应时间成反比（ $IOPS=1000/I/O$ 响应时间）。吞吐量是通过用写入数据尺寸乘以 IOPS 而得到的。这是针对每个共享 FCIP 连接的数据流计算的。
- 写加速率。这是使用写加速功能时的 I/O 响应时间（倒数）和 IOPS 与禁用加速功能时的 I/O 响应时间（倒数）和 IOPS 的比值。这个图形是针对 RTT 绘制的，写入数据的大小分别为 4Kb、8Kb、16Kb 和 32Kb。

对于下列图形，需要注意以下几点：

- 数值来自于一个启用和禁用写加速功能的模型（经过了实验室测试的验证）。
- 图形只能作为预期性能增益和差异的参考。实际的 I/O 响应时间将会随着存储阵列供应商、模型、配置和并发负载的变化而变化。
- 网络延时假定在 MDS 9000 的 FCIP 千兆位以太网接口之间有一对路由器。附加的路由器或者交换机将会添加更多的顺序发送延时。这种现象在速度较慢的连接——例如 45Mbps——上表现得更为明显，但是在速度较高的连接——例如 622Mbps 或者 1Gbps——上则不太显著。
- 所有图形都假定存储阵列的写入传输数据的尺寸至少为 32Kb，这是大部分阵列的典型尺寸。这是每个阵列在每个写入队列中可以接收的写入数据量。较低的写入数据尺寸将会导致较高的写加速性能率，因为在禁用写加速功能时执行写入操作需要往返三次以上才能完成。

图 2-28 速率为 45Mbps 时的近似写加速 I/O 响应时间和写加速率

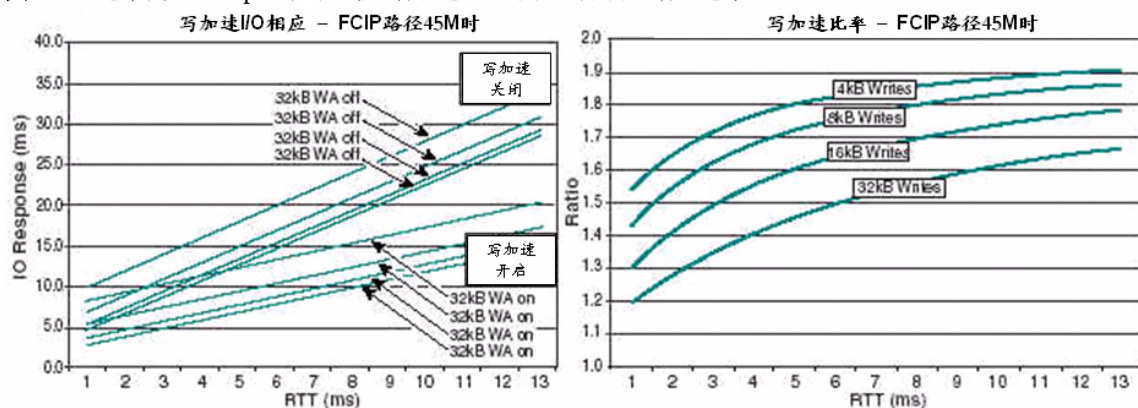


图 2-29 速率为 155Mbps 时的近似写加速 I/O 响应时间和写加速率（单个未被处理的 I/O）

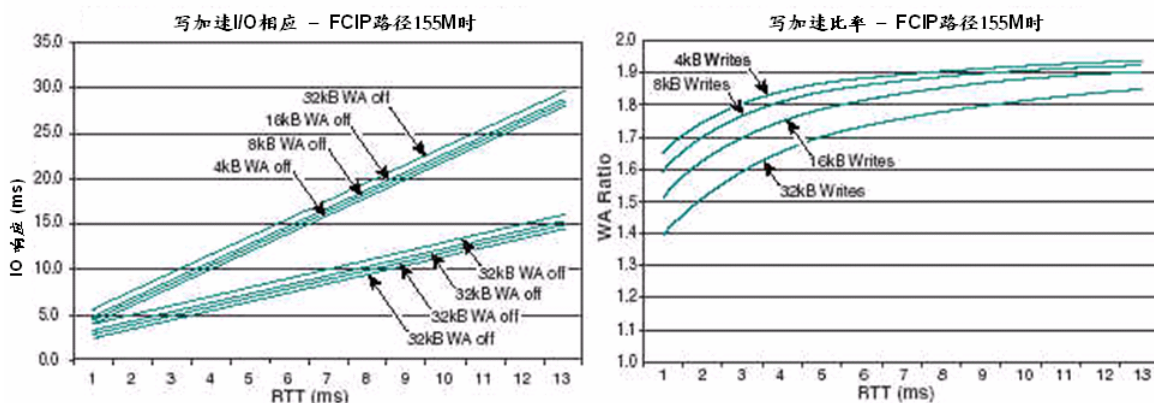


图 2-30 速率为 622Mbps 时的近似写加速 I/O 响应时间和写加速率（单个未被处理的 I/O）

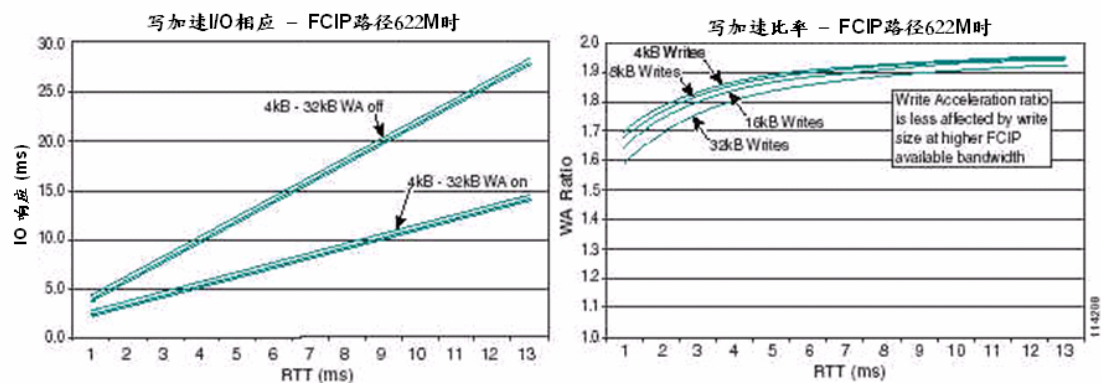
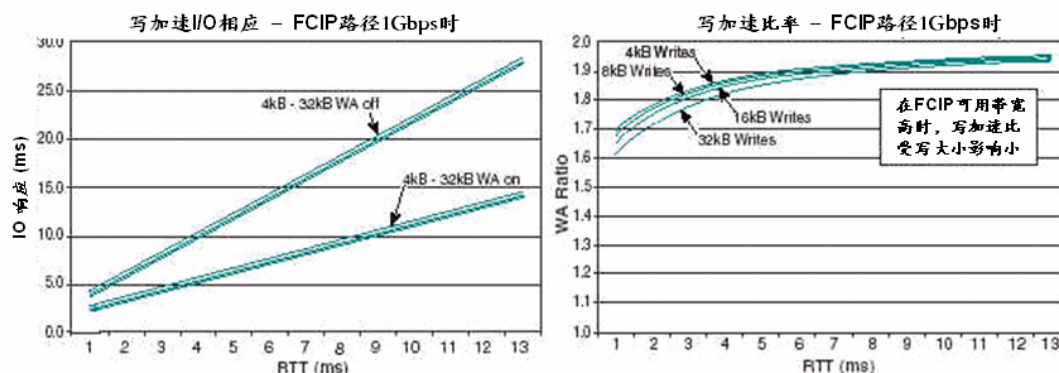


图 2-31 速率为 1Gbps 时的近似写加速 I/O 响应时间和写加速率（单个未被处理的 I/O）



何时应当使用写加速

在大多数情况下，写加速可以提高复制或者写入 I/O 的吞吐率，缩短 I/O 响应时间，尤其是在 FCIP RTT 提高时。每个 FCIP 连接都可以传输大量并发的或者尚未处理的 I/O。这些 I/O 可能来自于某一个复制源，或者大量的复制源。在未被处理的 I/O 达到某个特定的阈值时，FCIP 连接将会饱和。

这个阈值在很大程度上取决于 RTT、写入数据尺寸和可用的 FCIP 带宽。如果所有复制进程（单向）的尚未处理的 I/O 总数低于这个阈值，FCIP 连接就没有得到充分利用，因而可以得益于写加速。

表 2-3-4 显示了在不启用写加速功能的情况下，FCIP 连接对于不同的 RTT、FCIP 最大可用带宽和写入数据尺寸的未处理 I/O 阈值。

表 2-3-5 显示了在启用写加速功能的情况下，导致某个 FCIP 连接饱和所需要的未处理 I/O 总数。注意：这些数值低于启用写加速功能时的阈值，这意味着某个应用只需要少量未处理 I/O，就可以充分地使用一个 FCIP 连接。

注：下表中的值都是近似值，仅供参考。实际数值将会随着存储供应商、模型、配置和中间网络的变化而变化。

1. 确定 FCIP 往返时间（如果不清楚，请使用 `MDS 9000 ips measure-rtt` 命令）。
2. 确定最大可用带宽。这应当与 FCIP 设置中指定的 FCIP 连接最大可用带宽相对应，不得超过 FCIP 路径中的最低连接速度。
3. 确定应用所使用的写入数据尺寸和最大未处理 I/O 数量。注意：同步应用只允许一个未处理 I/O；半同步应用允许两个；异步应用允许的数量会随着供应商产品和配置的变化而变化。
4. 计算所有使用 FCIP 连接的复制进程的未处理 I/O 的总数。
5. 在表 2-3-4 和 2-3-5 中查找未处理 I/O 阈值。如果总数低于表 2-3-4 中显示的阈值，则启用写入加速。如果总数高于阈值，则 FCIP 很可能已经饱和，这时写加速功能不会带来任何好处。

表 2-3-4 在不启用写加速时，完全利用连接容量的未处理 I/O 近似总数

FCIP 往返 时间	45Mbps			155Mbps			622Mbps			1000Mbps		
	8kB	16kB	32kB	8kB	16kB	32kB	8kB	16kB	32kB	8kB	16kB	32kB
1	4	3	2	11	6	4	38	20	11	60	31	16
2	6	4	3	16	9	5	57	30	16	91	46	24
3	7	4	3	20	11	6	77	39	20	123	62	32
4	8	5	3	25	13	7	96	49	25	154	78	40
5	10	6	4	30	16	9	116	59	30	185	93	47
6	11	6	4	35	18	10	135	68	35	216	109	55
7	13	7	4	40	21	11	155	78	40	248	125	63
8	14	8	5	45	23	12	174	88	45	279	140	71
9	16	9	5	49	26	14	193	98	50	310	156	79
10	17	9	5	54	28	15	213	107	54	341	171	87
15	24	13	7	79	40	21	310	156	79	498	250	126
20	31	16	9	103	52	27	407	204	103	654	328	165
30	45	23	12	151	76	39	602	302	152	966	484	243
40	59	30	16	200	101	51	796	399	200	1279	640	321
50	73	37	19	248	125	63	990	496	249	1591	796	399
60	87	44	23	296	149	75	1185	593	297	1904	953	477
70	101	51	27	345	173	87	1379	690	346	2216	1109	555
80	115	58	30	393	197	100	1574	788	395	2529	1265	633
90	129	65	34	442	222	112	1768	885	443	2841	1421	712
100	143	73	37	490	246	124	1962	982	492	3154	1578	790

表 2-3-5 在启用写加速时，完全利用连接容量的未处理 I/O 近似总数

FCIP 往返 时间	45Mbps			155Mbps			622Mbps			1000Mbps		
	8kB	16kB	32kB	8kB	16kB	32kB	8kB	16kB	32kB	8kB	16kB	32kB
1	3	2	2	7	4	3	23	12	7	36	19	10
2	4	3	2	9	5	4	33	17	9	51	26	14
3	5	3	2	12	7	4	42	22	12	67	34	18
4	5	3	2	14	8	5	52	27	14	83	42	22
5	6	4	3	17	9	5	62	32	17	98	50	26
6	7	4	3	19	10	6	71	37	19	114	58	30
7	7	4	3	21	11	7	81	41	21	130	66	34
8	8	5	3	24	13	7	91	46	24	145	73	37
9	9	5	3	26	14	8	101	51	26	161	81	41
10	9	6	4	29	15	8	110	56	29	176	89	45
15	13	7	4	41	21	11	159	80	41	255	128	65
20	16	9	5	53	27	14	208	105	53	333	167	84
30	24	13	7	77	39	20	305	153	77	489	245	123
40	31	16	9	101	51	27	402	202	102	645	323	162
50	38	20	11	126	64	33	499	250	126	801	401	202
60	45	23	12	150	76	39	596	299	150	958	480	241
70	52	27	14	174	88	45	693	348	175	1114	558	280
80	59	30	16	198	100	51	791	396	199	1270	636	319
90	66	34	18	222	112	57	888	445	223	1426	714	358
100	73	37	19	247	124	63	985	493	247	1583	792	397

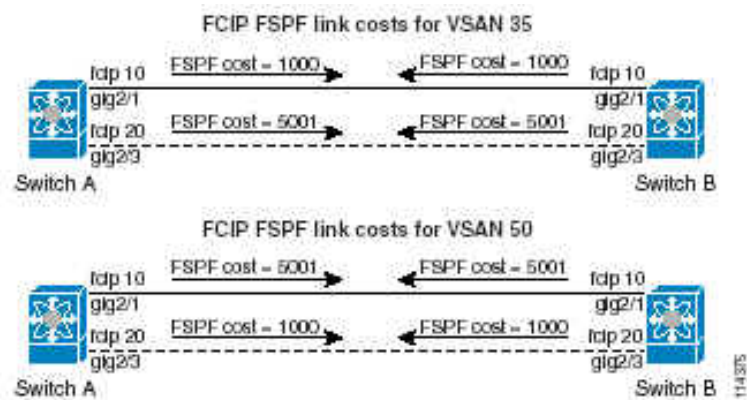
使用写加速的考虑因素和指导原则

在利用写加速部署 FCIP SAN 扩展时，需要考虑下列影响：

- 对于高可用性的影响——一次只能使用一个 FCIP 连接，不能使用 FCIP 端口通道。可以部署多个连接，并为每个方向上的 FCIP 接口分配不同的 FSPF 连接成本。
- 对于负载平衡的影响——因为不能使用 FCIP 端口通道，考虑为基于 FCIP 连接的中继创建多个 VSAN，并为 FCIP 连接分配 FSPF 权重，从而让一半 VSAN 选择一个 FCIP 连接，另外一半选择另外一个 FCIP 连接（假设有两个 FCIP 连接可用）。

图 2-32 显示了通过使用多个 FCIP 连接，在多个 VSAN 上传输流量的方法。这个例子为每端使用了相同的编号，但实际上也可以在每端独立地对 FCIP 编号。

图 2-32 为每个 VSAN 的 FCIP 接口分配不等的 FSPF 连接成本



在这个例子中，两个 VSAN（35 和 50）在两台交换机（交换机 A 和交换机 B）之间通过 FCIP 进行扩展。VSAN 35 将基于 FCIP 10 的 FCIP 连接作为首选路径。只有在 FCIP 10 发生故障时，才用 FCIP 20 传输 VSAN 35 的流量。VSAN 50 则将 FCIP20 作为传输 VSAN 50 流量的首选路径。只有在 FCIP 20 发生故障时，才用 FCIP 10 传输流量。

FCIP 接口的缺省 FSPF 成本为 1000。建议为替代 FCIP 路径单独分配一个成本值（例如 5001），以便在更加复杂的光纤通道矩阵发生故障时选择 FCIP 路径。奇数值（5001）确保了没有其他 FSPF 路径可以拥有相等的成本。

2.4 利用 Cisco MDS 9000 系列多层交换机构建 SAN 互联解决方案

减少和控制 IT 数据中心存储基础设施开支的持续压力迫使企业向存储局域网移植。随着越来越多的应用移植到 SAN，SAN 的永续性变得越来越重要。当越来越多的应用移植到通用 SAN 之后，通用 SAN 的错误和间断对应用的影响力也不断升高。另外，作为业务连续性计划的一部分，将 SAN 解决方案扩展到远程数据中心后，虽然会扩展 SAN 连接，但也会扩展潜在错误分区，因为远程设施中的间断也会影响到中央数据中心。

在 IP 世界里，随着 IP 网络的持续扩展，第 2 层 VLAN 技术与第 3 层 IP 路由协议的使用，不但能将错误分区限制在 IP 子网边界内，还能实现这些分区之间的必要连接。但是，作为第 2 层基础协议的光纤通道并不具备这种错误分区隔离功能。如果想不断推进能够降低成本的融合 SAN 基础设施的发展，又不降低这些环境的稳定性或安全性，必须在 SAN 内实现这一功能。为将传统数据网络中的可扩展性和稳定性延伸到 SAN 领域中，思科在 MDS 9000 系列中采用了很多创新技术。与其它传统数据网络一样，虚拟 SAN（VSAN）和 VSAN 间路由等新型数据中心技术，以及 FCIP 等集成式 SAN 扩展传输方式能够提高光纤通道 SAN 的可扩展性和永续性。

本文将探讨与光纤通道和单物理 fabric 相关的潜在问题——单错误域模式以及无法有效隔离本地网络和远程网络的事件和错误。思科的解决方案不但实现了 SAN Fabric 网络的可扩展性，还消除了单错误分区带来的影响，因而可以利用 VSAN 的逻辑 fabric 网络模式建立光纤通道 SAN，在此过程中，利用 VSAN 间路由和 FCIP 等永续性增强技术提供连接。另外，本文还将讨论实现 SAN 互联的其它方法，包括 iFCP，及其运行机制和固有缺陷。

2.4.1 SAN 互联

SAN 已经从为小型隔离 SAN 岛提供连接的基于仲裁环集线器的小网络发展为当今连接数百台设备的大型 SAN 网络。像以太网发展过程中需要考虑可扩展性和永续性一样，依靠一个 SAN 网络连接众多设备也会降低 SAN 的永续性和稳定性。另外，以前只局限在一个数据中心内，专为一个应用或一个部门使用的 SAN 网络，现在可能已经跨越国界或洲界。大型 SAN 网络的优势是能够集中管理，能够更加有效地使用网络资源（端口、ISL），但是，为保证扩展过程中能够保持适当的永续性，原有 SAN 交换机的体系结构以及基础光纤通道协议本身很快就会达到使用极限。

以太网协议一样，光纤通道属于交换式第 2 层协议，由于没有等级网络域概念，因而不能将控制协议和消息传送隔离和限制在网络的某个分区。与以太网类似，光纤通道在一定网络范围内（例如分区）使用一套网络 fabric 控制协议，以及交换机产生的一套状态变化通知。因此，即使是本地的光纤通道控制协议消息也只能在小 Fabric 网络范围内传送，因此，它们的影响只局限在 Fabric 之内。这种体系结构上的缺陷会在网络中形成很大的高可用性漏洞，尤其是地理分布范围较大的 SAN。造成这种高可用性风险的原因是，潜在干扰性局域控制协议事件可能会干扰到扩展链路两边的整个 Fabric 网络。另外两个大问题是网络扩展过程中遇到的质量问题和上述事件影响问题。许多客户都计划建立一个网络中包含 3000 多个端口的超大型 Fabric 网络。随着这些 Fabric 网络的增长，网络控制信息量将不断增加，交换机 CPU 和内存资源的负担也将加重，从而破坏网络的稳定性。在 MDS 9000 系列交换机中，思科满足了增加 CPU 处理资源的需要，使 CPU 处理能力达到了市场上其它导向器或 Fabric 交换机的 10 倍以上。但是，因 Fabric 网络控制协议流量的增加而提高的 Fabric 网络级干扰风险依然存在。

为在以太网域中解决这个问题，建议使用第 2 层和第 3 层网络协议层为大型以太网提供等级和必要的隔离。VLAN 能够在第 2 层提供隔离功能，OSPF 和 BGP 等 IP 路由协议则能够提供有目的的高层 IP 连接。生成树等以太网控制协议与对应的 VLAN 隔离开，这样，在发生生成树事件时就不会影响到其它 VLAN。遗憾的是，SAN 中的通用上层协议——光纤通道协议（FCP）不提供 IP 世界里的隔离功能。随着 SAN 的增长，以及传输方式、线速和延迟的进一步改善，这个问题将变得越来越重要。

2.4.2 SAN 岛的发展

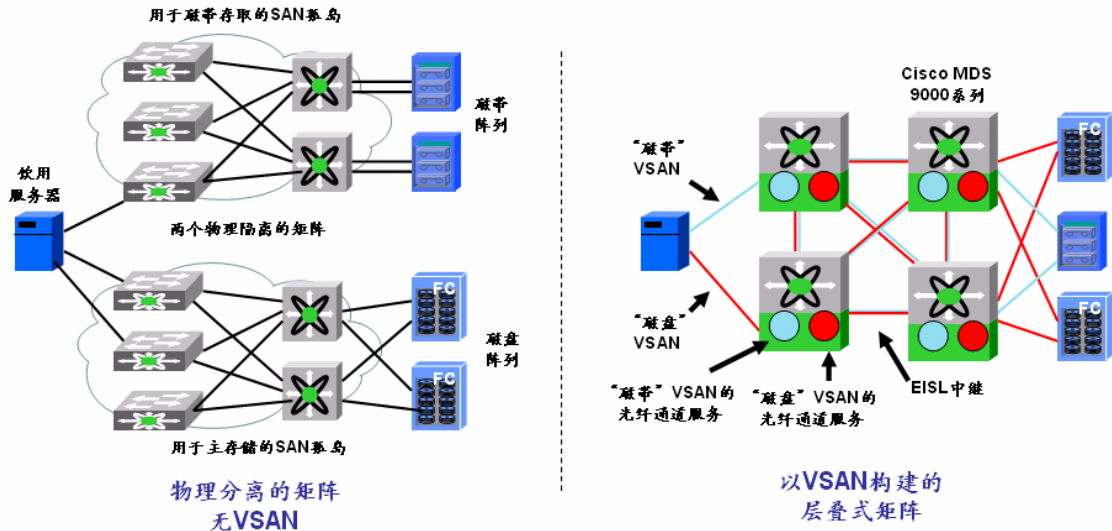
以前的 SAN 设计使用了许多较小的 SAN 连接岛。这些 SAN 岛一般都以应用或部门为单位部署，以满足应用的安全性和稳定性要求。许多稳定性问题的根源都是因为光纤通道协议内缺乏在上述大 Fabric 网络中建立分区的相应功能。不仅如此，由于以前没有高端口数的大型 SAN 交换机，不能在 SAN 中支持大量交换机，再加上前面提到的光纤通道协议低效的现象，使客户不得不在较小的“SAN 岛”配置中部署 SAN。SAN 岛指利用少量交换机建立的 SAN，这种 SAN 为每个企业应用或部门提供的总端口数量很少。多个小 SAN 岛的管理非常困难。现在，老交换机已经被更大的导向器级交换机所取代，因而能够支持端口密度较高的核心-边缘配置。但是，出于许多技术和非技术原因，建立 SAN 岛仍是首选方案，SAN 岛还在不断涌现。部署多个 SAN 岛是一种浪费，因为每个 SAN 岛的所有端口都为某一个应用服务，而且许多端口都处于闲置状态，从而进一步提高了这种设计方法的有效成本。

2.4.3 虚拟 SAN

为解决 SAN 岛的低效问题，思科开发了一种称为虚拟 SAN（VSAN）的技术。虚拟 SAN 利用了以太网中使用虚拟 LAN 的概念，成功地克服了 SAN 岛的低效问题。VSAN 将相互隔离的 SAN 岛连接到同一物理基础设施上，但并不合并独立 Fabric 的控制和管理。Cisco MDS 9000 系列交换机的每个端口被分配到建好的 VSAN，物理 Fabric 网络内的每台交换机最多支持 256 个 VSAN。通过使用独立 SAN 岛实现的所有 Fabric 网络隔离都采用了 VSAN 技术。在 VSAN 型网络中，称为 EISL 的通用 ISL 同时为多个 VSAN 服务。来自所有 VSAN 的流量都将打上 VSAN 成员信息标记，通过通用 ISL 传输，都仍然保留逻辑上的隔离。在多个虚拟网络和应用中使用通用 ISL 能够进一步降低基础设施成本。借助 VSAN，不但能够在更大的物理 Fabric 网络中实现第 2 层隔离，还可以集中执行 VSAN 管理。另外，还可以将单个 VSAN 的管理全部或部分分配给应用管理员，即允许指定管理员管理与分配与相应应用有关的虚拟网络的所有服务和策略。利用思科推出的基于角色的访问控制（RBAC）技术以及 RBAC VSAN 增强，可以为管理员指定单项角色，允许管理员有选择地访问交换机配置功能，例如以 VSAN 为基础。例如，如果指定了称为 role_1 的角色，管理员将可以查看整个物理网络（所有 VSAN），但只能对 VSAN_10 定义和激活分区和静态 FCID 分配。

与通常会浪费端口的 SAN 岛不同，物理 Fabric 网络中的端口可以顺利地从一个 VSAN 转移到另一个 VSAN，大大降低供应成本。与 VLAN 和以太网相似，称为 EISL 的 VSAN 型 ISL 能够承载与交换机间多个 VSAN 相关的光纤通道流量。利用 EISL，网络设计者可以为多个虚拟网络使用相同的交换机间链路，并对每个 VSAN 实施最终的流量控制。另外，利用 Cisco MDS 9000 系列内称为端口通道的 ISL 或 EISL 汇聚，还可以将多个（E）ISL 捆绑汇聚在一起，形成永续性和带宽更高的大型虚拟（E）ISL。

图 2-32 VSAN 型 SAN 举例



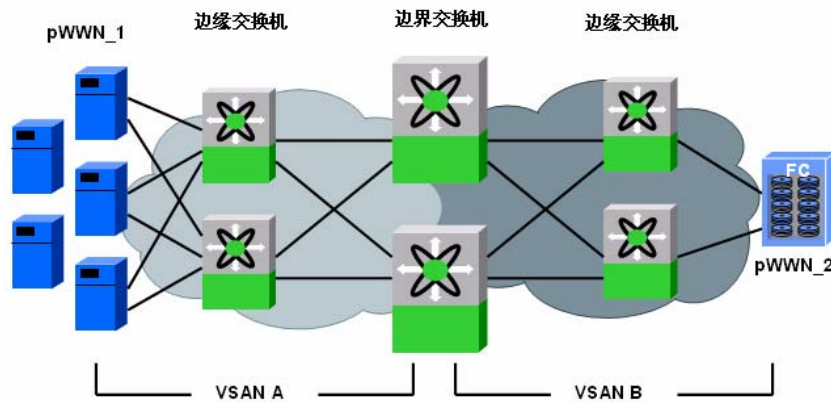
VSAN 内的所有控制和数据流量都与 VSAN 隔开，只能随着特定 VSAN 网络范围的扩展而扩展。VSAN 的这种设计特性是提高大物理 Fabric 网络扩展能力的关键。利用 VSAN，不但可以在 Fabric 网络内建立任意到任意物理连接，还可以在网络范围内动态建立更小、可管理性更高的虚拟网络。由于这种 VSAN 型设计能够提供离散性和必要的控制隔离，因而能提高网络的永续性和可扩展性。小虚拟网络中发生的网络事件只对虚拟网络构成影响，而不会影响到更大的物理 Fabric 网络。

VSAN 不但有助于开发规模更大、永续性更高、可管理性更好的 SAN 网络，还有助于建立端口数量高于 1000 的 Fabric 网络。但是，VSAN 的设计有一个局限。由于所有流量都被隔离在相应 VSAN 之内，因此，与 Fabric 网络相连的设备只能包含在一个 VSAN 中。虽然对于多数主机 HBA 连接，这都不成问题，但是，如果想在包含多个 VSAN 的多个应用和/或多个部门之间共享更多存储阵列端口，问题就暴露出来了。在长距离的扩展网络解决方案中，用户可能并不希望利用高延迟传输服务扩展通用网络或 VSAN，并以此提高网络永续性。为使 VSAN 解决方案满足这些要求，MDS 9000 交换机系列中采用了称为 VSAN 间路由或 IVR 的 VSAN 技术扩展。

2.4.4 VSAN 间路由简介

如前所述，VSAN 能够在普通物理 Fabric 网络之上构建虚拟矩阵网络，从而实现高度隔离。但是，SAN 构建者可能既要求某些设备在多个 VSAN 内通信，又不希望将虚拟 Fabric 网络合并在一起，因为这样会降低永续性。磁带驱动器或大型子系统等设备通常会出现在多个 VSAN 中。利用思科 VSAN 间路由（IVR）技术，可以建立从一个 VSAN 到不同 VSAN 内多台设备的路径，而且不需要合并单个 VSAN Fabric 网络，即不会造成错误分区融合。如果没有 IVR 功能，大型磁盘阵列等设备必须与它想参与的所有 VSAN 的特定交换机端口相连。利用 IVR 技术，只需要定义多个 VSAN 中需要连接的设备 IVR 分区，共享设备就可以在多个 VSAN 之间通信，而且不需要合并参与的虚拟矩阵网络（VSAN）。

图 2-33 VSAN 间路由



VSAN 间分区 (IVZ)

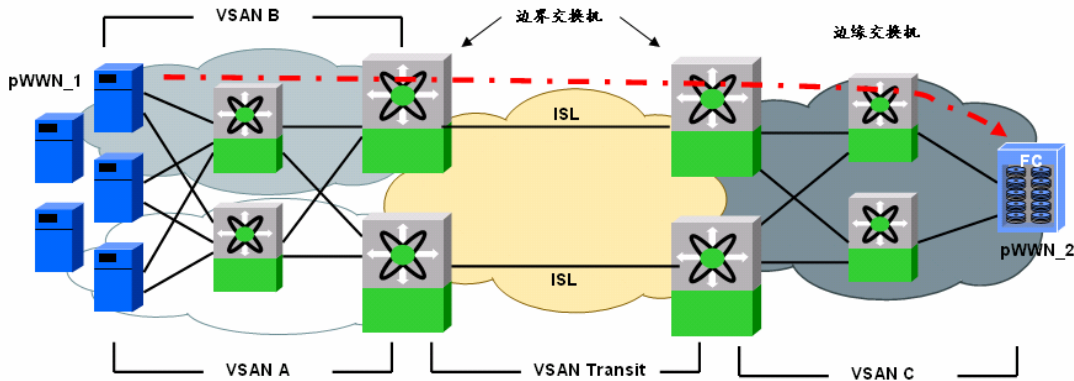
VSAN 间路由技术的最大优点之一是易于供应和管理。IVR 的主要功能之一，即 IVR 与 iFCP 等其它方法之间的主要区别是，IVR 的供应是利用已知步骤提供给 SAN 管理员的。利用 SAN 管理员再熟悉不过的分区管理步骤和管理任务，只需定义一种称为 VSAN 间分区的新分区，就可以建立 VSAN 间路由实例。VSAN 间分区的定义方法与其它分区相同，但它包含参与了边缘 VSAN，需要跨 VSAN 连接的设备。IVR 技术能按照 VSAN 间分区的定义自动确定哪些路径和名称服务器需要交叉传播。利用类似的分区机制，SAN 管理员不需要学习 iFCP 和 iSNS 或新的管理工具就可以管理 IVR 服务。

IVR 边界交换机

为了实现 IVR 连接，网络中至少有一台交换机必须作为传输或 IVR 边界交换机，以提供 VSAN 之间的传输路径。在图 2-33 所举的例子中，VSAN_1 中的设备 pWWN_1 需要与 VSAN_2 中的设备 pWWN_2 通信。边界交换机作为独立 VSAN 中设备间的传输交换机，如图 2 所示，可以利用多台 IVR 边界交换机提供冗余路径。边界交换机只将某些 FSPF 路径加入相应的本地 VSAN，这些本地 VSAN 与需要 IVR 连接的预定 IVR 型 VSAN 中的设备有关。换言之，只有通往包含这些设备的预定 IVR VSAN 中的域（交换机）的 FSPF 路径，才会交叉传播。另外，包含预定 VSAN 间分区内通信设备的身份识别的名称服务器数据库项，也将在规定的 IVR VSAN 内交叉传播。如果所有设备都包含在一个 VSAN 中，每台交换机里的所有数据库信息都将同步。利用 IVR 特性，按照 VSAN 间分区的规定，只有需要 IVR 通信的设备的一小部分路由和名称服务器信息才需要交叉传播，而且不需要合并参与的虚拟网络（VSAN）。

IVR 方法对 SAN 互联的另一个优点是，IVR 依靠标准 FSPF 路由实现路径选择和路径永续性。利用 iFCP 等其它互联解决方案，由于可以转换为 iFCP，因而没有端到端使用 FSPF 协议，也就没有端到端路由智能和永续性。iFCP 等解决方案严格依赖基于逐级的多路径解决方案提供永续性。但是，由于 IVR 通过 IVR 边界交换机传播某些 FSPF 路径，因而可以使用简单的标准 FSPF 从端到端网络角度选择路径和确定冗余路径。对于 iFCP 的解决方案不能担保的拓扑中的任何路径故障，这个功能都能提供快速、可靠的恢复机制。

图 2-34 使用了通道 VSAN 的 VSAN 间路由



IVR 通道(Transit)VSAN

对扩展 SAN DR 解决方案特别有用的另一个 IVR 配置选项是使用 IVR 通道 VSAN。通道 VSAN 是用于连接两个边缘 IVR VSAN 的 VSAN，有了它，无需从一个 IVR VSAN 扩展到另一个 VSAN 就能建立 IVR 服务。事实上，通道 VSAN 能够扩展边界交换机的范围，到达参与各 IVR VSAN。如图 2-33 所示，设备需要通信的两个边缘 IVR VSAN 已经与普通 VSAN 边界交换机相连。在这个简单的配置中，不需要通道 VSAN。事实上，虽然通道 VSAN 通常是任选设备，但在某些配置中，最好保持隔离状态，例如广域网 SAN 扩展解决方案。图 2-34 所示的通道 VSAN 可用于连接广域网上三个不同的边缘 VSAN。值得注意的是，通道 VSAN 中也可以包含各种设备，而且边缘 VSAN 之间可以存在多个通道 VSAN。如图 2-34 所示，IVR 建立 IVR 服务时，不需要将任何边缘 IVR VSAN 扩展到普通边界交换机。在这个例子中，这种做法不需要在高延迟长途网上扩展本地 IVR VSAN。虽然 IVR 功能支持选择性连接，但利用 MDS 9000 系列交换机内基于 VSAN、基于角色的访问控制 (RBAC)，无论是从控制协议角度，还是从管理角度，三个边缘 VSAN 都完全隔绝。另外，原来的非思科光纤通道交换机也可以加入任一边缘 IVR 型 VSAN，跨 VSAN 间边界通信，并成为通道 VSAN 的一部分，但不能作为边界交换机参与。

在图 2-34 所示的例子中，IVR 型 VSAN_A 和 VSAN_B 的网络控制流量在边界交换机处本地终止，IVR 型 VSAN_C 的控制流量在远程边界交换机处本地终止。通道 VSAN 是横跨边界交换机间广域连接的唯一 VSAN。因此，发生在每个边缘 IVR VSAN 中的网络事件，无论是普通控制流量，还是错误事件，都将局限在各自的边缘 VSAN 内，不会穿越和影响其它 IVR 型 VSAN 或通道 VSAN。在稳定环境中，参与 IVR 的所有交换机都将交换巨大的控制流量。在某两台交换机之间的正常 ISL 链路初始化过程中，交换机将交换少量控制协议信息。除这些信息外，位于通道 VSAN 边缘的两台交换机还将交换严格属于将通过通道 VSAN 建立 IVR 连接的那些设备的 FSPF 和名称服务器数据库信息。这种交换为需要 IVR 连接的每台设备增加了大约相当于两个光纤通道帧的流量。

与 iFCP 等其它互联方法相比，IVR 解决方案的另一个优点是，它对通道 VSAN 可以使用的传输技术没有限制。IVR 服务严格依赖基本光纤通道服务，因而可以利用光纤通道能够使用的所有传输选项，例如光纤通道本身、DWDM/CWDM 传输、SONET 传输或 FCIP。在基于 iFCP 的解决方案中，iFCP 必须是以以太网/IP 协议，不需要增加网关设备。

IVR 解决方案的其它优点还包括，它不生成也不需要任何其它新控制协议。为提供相应功能，IVR 服务严格使用 VSAN 和分区等通用机制，以及 Cisco Fabric Manager 中的某些功能。

网络的稳定性和永续性

需要强调的是，VSAN 能够将 Fabric 网络服务限制在每个虚拟 Fabric 网络或 VSAN 范围内，因而能提供 Fabric 网络稳定性。为新的 VSAN 配置 MDS 9000 系列交换机时，新虚拟 Fabric 网络中的所有交换机都交换设备、转发和分区数据库，以保证新 VSAN 的一致性。另外，VSAN 中的交换机还按

照交换的 VSAN 专用型 FSPF 链路状态信息建立路由拓扑。在虚拟 Fabric 网络初始化过程中，一台交换机被选为“首(principal)”交换机。首交换机将为其它交换机指定一组 VSAN 专用域 ID，用于在新虚拟 Fabric 网络中定址和路由。如果交换机失去了与首交换机的联络，其余交换机将重新执行初始化过程，为 VSAN 选择另一台首交换机。交换机将重新交换数据库，根据虚拟 Fabric 网络内交换机的数量以及 VSAN 的设备数量，开始时间不等的信息交换过程。

考虑扩展 SAN 解决方案时，长途网络传输服务的可靠性相当于服务供应商提供的服务等级协议（SLA）的规定。虽然长途链路存在连接中断问题，但出问题的概率非常低。在缺少长途传输手段的场合，VSAN 间路由能够减少数据库同步流量的中断和泛滥。在图 2-34 中，扩展 SAN 解决方案采用了 VSAN 间路由，并使用了通道 VSAN。通道 VSAN 选举驻留在长途链路一边的交换机为首交换机。另外，单边 VSAN 还选举了各自 VSAN 中的首交换机。如果长途传输服务出现故障，VSAN 间分区内拥有设备的边缘 IVR VSAN 中的交换机将收到远程 IVR VSAN 上某交换机（域）不再可用的通知。但是，与长途传输服务两边的所有交换机都位于同一个 SAN 或 VSAN 的情况不同，与通道 VSAN 连接的边缘 IVR VSAN 将不间断地继续操作，因为其 VSAN 内的首交换机仍然可用。IVR 配置和通道 VSAN 中的控制板隔离功能消除了边缘 VSAN 以外链路故障引发的边缘 VSAN 网络中断。

与 iFCP 相比，IVR 解决方案的最大优点是，IVR 解决方案使用了端到端标准光纤通道网络路由和通知服务，例如 RSCN，而且不需要将这些控制信息转换成另一协议中的其它控制信息，例如 iFCP。由于采用了连续的通用网络警报机制 RSCN，IVR 解决方案的可用性高于基于 iFCP 的解决方案。

采用了 VSAN 间路由的永续性设计

VSAN 和 VSAN 间路由基于在光纤通道中建立的路由准则和机制，称为基于 FSPF 的拓扑图。IVR 边界交换机之间横跨通道 VSAN 的链路就好像是光纤通道交换机间链路（ISL）。它们基于光纤通道，可以使用多种方式传输，包括 CWDM/DWDM、SONET/SDH 或 FCIP。多条 ISL 可以汇聚成几个端口通道，以提高链路永续性，并将最高带宽增加到 32Gbps。如果利用 FCIP 实施长途传输，还可以利用基于网络的端口通道技术，将来自 FCIP 通道的虚拟 ISL 汇聚成最高 16Gbps 的虚拟端口通道。提供 VSAN 间路由服务时，管理员不需要配置任何其它组件或传输/网关协议。

2.4.5 VSAN 间路由的用途

VSAN 间路由是思科 VSAN 技术的逻辑功能扩展。IVR 可用于如下多种解决方案。

共享对跨 VSAN 边界的存储设备的访问

VSAN 提供的永续性来自于每个 VSAN 的 Fabric 网络服务复制，以及属于每个 VSAN 的帧的显性标记。这种永续性或隔离度相当于用更经济、有效的方式建立了多个物理隔离的 SAN 岛。在 IVR 出现之前，流量无法穿越 VSAN 边界，因为硬件中的所有隔离机制都采用了显性帧标记。

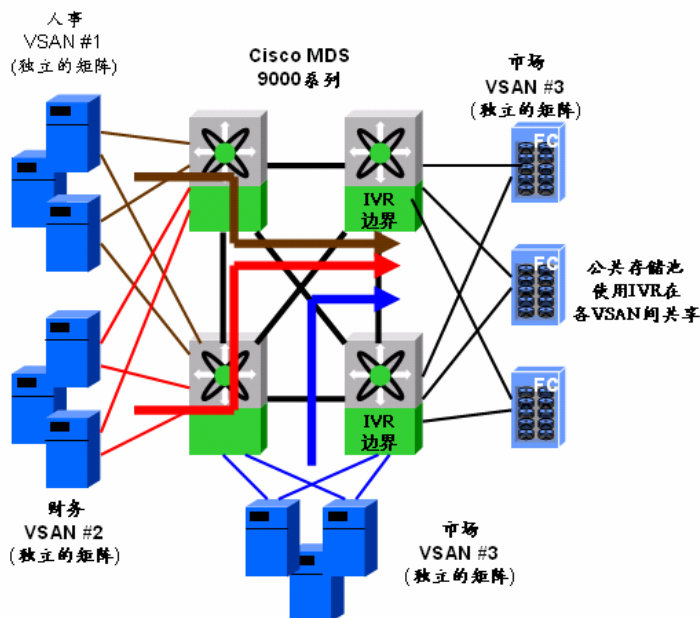
VSAN 间路由是允许跨 VSAN 边界选择发送流量的唯一机制。IVR 无需合并 VSAN 网络就能实现交叉 VSAN 通信。这个特性很重要，因为每个 VSAN 内仍然保留着独立的错误域。只有必须跨 VSAN 边界与其它设备通信的设备才使用 VSAN 间路由技术。从控制协议和管理角度看，每个边缘 VSAN 仍然保持着其 Fabric 网络永续性。由于 VSAN 允许大大扩充网络端口数量，因此，借助 IVR 提供的可以扩展的永续性方法，无需完全合并整个 Fabric 网络就能共享与这些大 Fabric 网络内通用存储设备的连接。例如，用户可以考虑建立应用专用或 OS 专用 VSAN，将与某应用或 OS 相关的所有服务器都放置在一个 VSAN 中，以提高永续性。利用 IVR 技术，无需合并虚拟 Fabric 网络和错误域就能在多个 VSAN 之间共享通用大型企业阵列。利用 IVR 在三个独立的部门 VSAN 之间共享磁盘阵列的例子如图 4 所示——为提高可管理性和稳定性，所有 VSAN 都保留在独立 Fabric 网络中。

为数据复制提供扩展 SAN 解决方案

VSAN 间路由能够改进 VSAN 为建立设备连接而提供的解决方案，这种解决方案只通过 SAN 扩展服

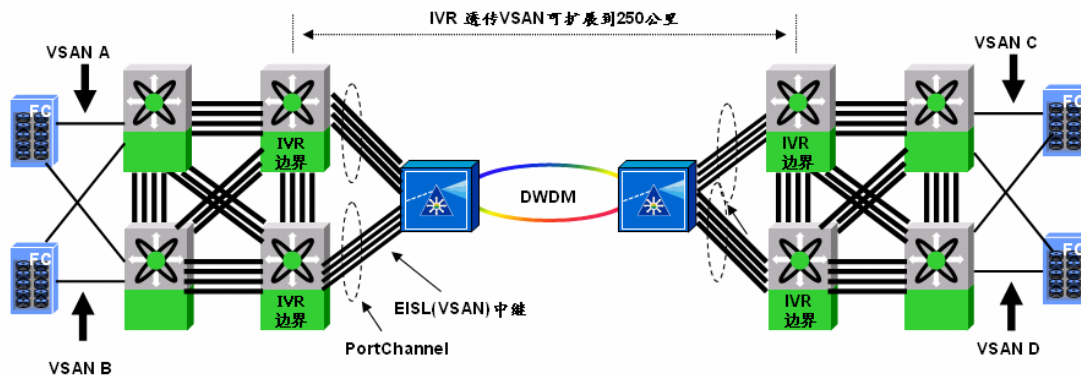
务通信，通常不在任何本地 SAN 中使用。例如，用大型企业存储阵列将数据复制到远程数据中心。这种远程复制解决方案包括基于阵列的服务，例如 EMC SRDF、HDS Truecopy、IBM PPRC 和 HP Continuous Access 复制服务。多数阵列要求控制器端口专门用于复制服务，不允许主机使用这些端口。最简单的解决方案是建立另一个 VSAN，只包含解决方案两边用于复制服务的存储阵列端口，而且只允许这个 VSAN 穿越长途链路。在这种配置中，指定 VSAN 中的控制流量最少，因为 VSAN 中的配置相对稳定，只包含少量设备。但是，由于 VSAN 间路由在解决方案两边之间提供了控制协议边界，只允许在两边之间交换相关 FSPF 路径和名称服务器项，因而提高了永续性。无论是计划内还是计划外 Fabric 网络事件，都将限制在各自的边缘 VSAN 内，不会通过长途链路传播。

图 2-35 IVR 例 1——共享目标端口



利用 IVR，可以减少 SAN 之间的长途交换机间流量。对于 SAN 扩展解决方案，如果 SAN 扩展的两个端点之间的延迟相对高于数据中心内部，使用 VSAN 间路由也很有利。只有定义为 VSAN 间分区一部分的设备的相关控制协议信息才会通过长途链路传输。另外，还可以为通道 VSAN 调整 ED_TOV 等独立网络定时器，以适应较高的传输延迟。SAN 扩展两边的 Fabric 网络保持隔离状态，并作为独立分区控制，因而能够为整体数据复制解决方案提供非常高的永续性。如图 2-36 所示，可以将 IVR 与 DWDM 和端口通道结合起来使用，建立高永续性的扩展 SAN 解决方案。

图 2-36 IVR 例 2——采用了 DWDM 和端口通道的 SAN 扩展



VSAN 间路由的最佳实践经验

部署 IVR 解决方案时，建议使用以下最佳实践经验：

为长途解决方案使用通道 VSAN

虽然并不完全必要，但 IVR 通道 VSAN 能够为 SAN 解决方案提供最佳永续性和隔离度。这些扩展 SAN 解决方案可以是点到点，或者点到多点，而且可以包含一个或多个通道 VSAN。作者建议为这些解决方案使用通道 VSAN。

必要时使用分区、VSAN 和 IVR

Cisco MDS 9000 系列多层光纤通道交换机使用硬件机制增强分区配置。事实上，来自某设备的流量只能传输到规定分区的其它设备。分区不但能提供硬件安全性和数据流量隔离，还是通过 VSAN 提供的隔离增强功能。如果控制协议和数据水平上需要 Fabric 网络隔离，最好使用 VSAN，因为这样能够模拟独立物理网络提供的隔离。如果既希望提供网络控制和数据流量隔离，又希望实现不同 VSAN 内的设备之间的选择性通信，最好使用 VSAN 间分区。如果想详细了解分区和 VSAN，请参考思科白皮书《在 Cisco MDS 9000 多层光纤通道交换机中使用 VSAN 和分区》。

建立冗余的 VSAN 间路径

VSAN 间路由的一个最大优点是使用了基于标准的光纤通道路由准则和机制。因此，最好使用同样的方法，利用建立高度可用的光纤通道 SAN 时所用的方式建立高度可用的 VSAN 间路由 SAN。在大型 SAN 网络中，最好使用冗余的 VSAN 间边界交换机，最好在交换机之间使用端口通道，最好使用冗余设备链路。