第34卷 第12期 计 算 机 工 程 2008年6月
Vol.34 No.12 Computer Engineering June 2008
·开发研究与设计技术· 文章编号：1000—3428(2008)12—0250—04 文献标识码：A 中图分类号：TP309.3

基于CWDM的存储扩展及其性能仿真

蔡昭权[1,2]
(1. 惠州学院网络中心，惠州 516015；2. 清华大学计算机科学与技术系，北京 100084)

摘 要：为研究影响基于粗波分复用存储扩展系统性能的因素，采用OPNET仿真软件建立基于粗波分复用的存储扩展仿真模型，对扩展距离、信用值、FC帧大小以及链路带宽等因素对存储扩展性能的影响进行仿真分析。结果表明，随着扩展距离、链路带宽和FC帧长的增加，需要相应用于流量控制的信用值，才能维持存储扩展系统的高性能。

关键词：粗波分复用；容灾；性能仿真

Storage Extension Based on CWDM and Its Performance Simulation

CAI Zhao-quan[1,2]
(1. Network Center, Huizhou University, Huizhou 516015;
2. Department of Computer Science & Technology, Tsinghua University, Beijing 100084)

【Abstract】In order to analyze influential factors of performance of storage extension based on Coarse Wavelength Division Multiplexing (CWDM), this paper presents simulation environment by using OPNET software to study how distance, credit, FC frame size or bandwidth of extension link impact the performance of stroage extension based on CWDM. Simulation results show that with the increase of distance, credit, FC frame size or bandwidth of link, more credits need to be allocated to sustain high throughput.

【Key words】Coarse Wavelength Division Multiplexing(CWDM); disaster recovery; performance simulation

1 存储扩展的必要性及其模型

在数字化和互联网时代，信息资源的积累呈爆炸性增长。IDC 研究报告显示，2006年创建的数字信息总量为161 EB (1 EB=10亿GB)，预计还将以每年约57%的速度增长，到2010年将达到988 EB[1]。为了应对大容量、高性能存储系统的需求，越来越多的企业在其数据中心大量部署SAN[2-3]，为了面对来自自然灾害或人为因素可能导致的数据破坏，企业开始寻求保护数据安全的有效方法，大量数据灾难事件证明，容灾是提高数据可用性，保持企业业务连续运行的有效手段。通过存储扩展将关键数据从主数据中心复制到远程容灾备份中心是解决数据安全问题的有效方法，存储扩展已成为存储领域新的研究热点。Cisco, Nortel, EMC, Ciena, MCDATA等国际知名大公司都开始关注存储扩展产品和解决方案。存储网络工业协会 (Storage Networking Industry Association, SNIA)提出了存储扩展模型，如图1所示。

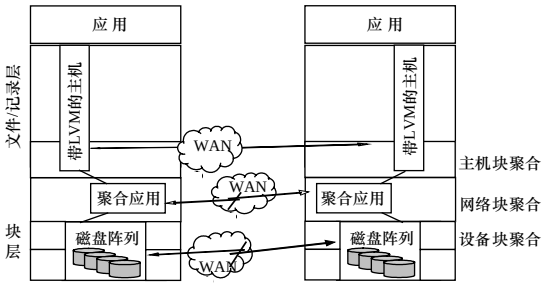

图1 多站点块存储模型

在图1中，2个地理位置上分散的数据站点之间，根据应用对性能、复制数据量的大小、数据备份的方式等要求的不同，通过高速广域网实现在主机逻辑卷、SAN交换构架之间或存储设备之间的扩展，本文所研究的基于 CWDM 的存储扩展是在SAN交换构架层次的存储扩展。

2 粗波分复用应用于存储优势

目前，3 种常见的存储扩展技术分别是基于SONET/SDH[4]、基于密集波分复用(DWDM)和基于IP的存储扩展，这3 种存储扩展技术的性能、吞吐率、可扩展性和经济性对比如表1所示。

表1 3种存储扩展的比较

项目	基于IP	基于SONET/SDH	基于DWDM
经济性	好	好	差
性能(抖动等)	低	高	高
吞吐率	低	中等	高
系统的可靠性	不可预见	高	高
安全性	一般	高	高

容灾国际标准 Share 78 通过引入恢复点目标(Recovery Point Objective, RPO)和恢复时间目标(Recovery Time Objective, RTO)来区分不同容灾等级对数据可用性的要求。其中，RPO衡量系统能够容忍的数据丢失量；RTO衡量数据恢复需要的时间。根据Share 78标准，实时应用容灾系统的RPO 和 RTO 分别为分钟级和小时级，在数据量一定的情况下，网络吞吐率将成为影响RPO 和 RTO 的主要因素。计算表明，在不考虑传播时延的情况下，采用千兆以太网传输200 GB 的数据约需27 min，而且随着数据容量的不断增加，传输时间还将不断增加。因此，要满足实时应用容灾系统RPO和RTO的要求，必须采用具有高吞吐率的网络。

基金项目：广东省自然科学基金资助项目(7008368)
作者简介：蔡昭权(1970－)，男，副教授、硕士，主研方向：计算机网络，软件技术，数据库，信息安全
收稿日期：2007-10-24 **E-mail：**cai@hzu.edu.cn

由表 1 的比较结果可知，对于大容量的存储系统而言，只有采用 DWDM 才能为存储扩展提供高吞吐率的网络。目前，商用 DWDM 能提供 8 路、32 路和 40 路的密集波分复用，按每个波长 2.5 Gb/s 的带宽来计算，复用后的网络总带宽将分别达到 10 Gb/s, 80 Gb/s 和 100 Gb/s。虽然 DWDM 能为存储扩展提供高带宽，但由于密集波分复用要求采用高波长稳定度的激光器和密集波分复用与解复用器，并且在整个光路上还需进行光功率均衡，因此在远程容灾系统中使用 DWDM 费用很高，这已成为制约 DWDM 应用于远程容灾的最大障碍之一。

粗波分复用(Coarse Wavelength Division Multiplexing, CWDM)技术的出现使运营商找到一种低价、高性能的传输解决方案。相对于DWDM而言，CWDM具有下列 4 个优点[5]：

(1)器件成本低。由于 CWDM 系统波长间隔为 20 nm，远远大于 DWDM 系统的波长间隔(小于 1 nm)，因此允许在使用无致冷光源条件下，各个波长同时传输，而且不需要激光器制冷、波长锁定和精确镀膜等复杂技术，大大降低了设备成本。

(2)功耗低。CWDM 的无致冷激光器及其控制电路每波长只需要 0.5 W 左右，大大降低了功耗。

(3)体积小、集成度高。由于 CWDM 激光器的结构和控制电路简单，因此激光器的物理尺寸远小于 DFB 激光器。另外，CWDM 系统不使用光放大器，更能设计成结构紧凑的 SFF(Small Form Factor)和 SFP(Small Form Factor Pluggable)2 种形式，方便安装和维护。

(4)支持更广泛的光纤类型。城域网内铺设的多数是 G.652 光纤，该类光纤在 1 385 nm 窗口的损耗为 1 dB 左右，虽然不能开通 DWDM 系统，但可开通 8 波 CWDM 系统；对于 G.655B 光纤，也可开通 8 波 CWDM 系统；对于全频带全波的 G.652C 光纤，则可开通 16 波 CWDM 系统。

目前，商用的粗波分复用系统有 8 波段和 16 波段 2 种，按照每波长承载 2.5 Gb/s 计算，8 波和 16 波 CWDM 的网络带宽将分别达到 20 Gb/s 和 40 Gb/s，远远大于 IP 网和 SONET 的网络带宽，是一种能满足目前存储扩展应用要求的高性价比传输技术。

3 基于 CWDM 的存储扩展模型

CWDM 借助于频分复用(Frequency Division Multiplexing, FDM)[6]的思想，将一根光纤能支持的信道按照波长分成多个子信道，不同的子信道在发送端通过复用技术组合到一根光纤中传输，在接收端通过解复用器将相应的波长从光纤中分离出来。因此，利用CWDM技术能大大提高光网络的传输容量，很好地满足存储扩展应用对高带宽网络的需求。基于CWDM的SAN扩展原理如图 2 所示。

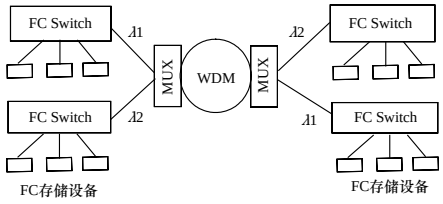

图 2　基于 CWDM 的 SAN 扩展原理

如图 2 所示，利用 CWDM 传输存储业务之前，必须将 FC 交换机输出的存储业务信号(光信号或电信号)变换成符合 CWDM 波长要求的光信号，图 2 中的波长转换部件可以通过在 FC 交换机中插入 CWDM 光模块的方式实现，也可以利用相应的光模块自己设计转换电路，但所选用的波长必须与 CWDM 环网中分配的波长相适应。

图 2 所示的是基于 CWDM 的点到点存储扩展网络拓扑结构，除这种结构外，还可根据实际应用的需要，非常方便地构建基于 CWDM 的点到多点、环行等存储扩展拓扑结构。

4 性能仿真

基于 CWDM 的存储扩展应用中，采用的是 FC 协议基于信用(credit)的流量控制协议。研究表明，在基于 CWDM 的存储扩展应用中，与性能有关的因素主要包括扩展距离、信用值、FC 帧的大小、链路的带宽等。

4.1 仿真模型建立

利用 OPNET 建立如图 3 所示的模型。

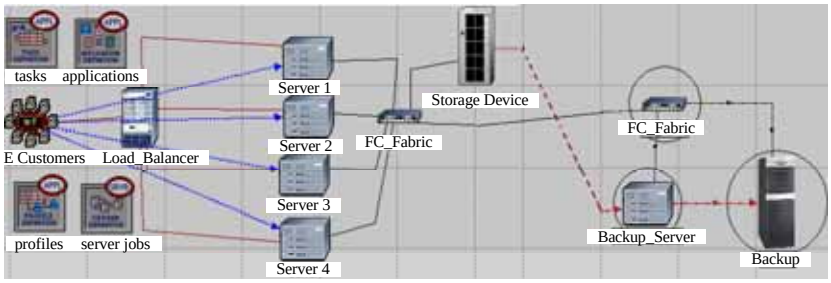

图 3　基于 OPENT 的存储扩展模型

在该模型中，原数据节点有 4 个应用服务器 Server1~Server4、1 个 FC 存储设备、1 个 FC 交换机；远程节点包含一个 FC 交换机、一个备份服务器和一个备份存储设备，2 个 FC 交换机之间通过 1 Gb/s 或 2 Gb/s 的 FC 链路连接，因为基于 CWDM 的存储扩展不涉及协议之间的转换，所以在该模型中利用 FC 链路模仿 CWDM 中一个波长承载的 FC 业务的设置是合理的。图 3 中虚线代表 FC 业务流，存储设备(storage device)和备份存储设备(backup)的虚线表示本地和远程之间的 FC 数据备份流。

4.2 性能与扩展距离的关系

为了研究数据远程备份性能与扩展距离的关系，建立了 4 个场景，其中，FC 帧的大小为 2 112 B，链路带宽为 1 Gb/s，FC 交换机的缓存大小为 4 MB(影响信用值的大小)，4 个场景的扩展距离不同，分别用不同的链路传播时延表示，4 个场景 6~9 的链路传播时延分别为 5 s, 3 s, 2.95 s 和 2.85 s。4 个不同场景下存储扩展的性能和对应的链路利用率分别如图 4 和图 5 所示。

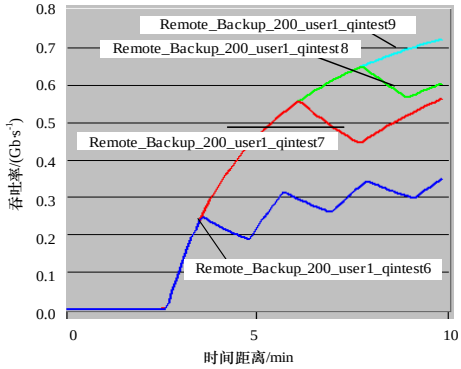

图 4　吞吐率与时间距离的关系

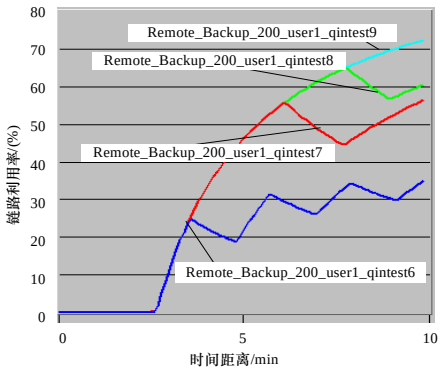

图 5　链路利用率与扩展距离的关系

从图 4 中可以发现，在其他参数相同的情况下，随着扩展距离(链路传播时延)的增加，存储扩展的性能将逐步降低，这是由于在信用值一定的情况下，扩展距离越长，单位时间内传输的数据帧的数量越少，导致存储扩展的性能降低，链路的利用率也降低，从图 5 中可以清晰地发现，链路利用率随扩展距离增加而降低。图 4 和图 5 中的 4 条曲线从上到下分别对应链路的传播时延从高到低的 4 个不同仿真场景的数据。

4.3　性能与链路带宽的关系

在距离、FC 帧的大小等参数相同时，随着扩展链路带宽的增加，维持链路全部被利用所需要的信用值也会增加，如果信用值不能随链路带宽的增加而增加，则链路的带宽不能被利用。在 4.1 节所建立的 4 个场景中，只将存储扩展链路的带宽修改为 2 Gb/s，重新仿真，得到 4 个不同场景下存储扩展的性能和对应的链路利用率，并与 4.1 节得到的结果进行比较，结果分别如图 6～图 9 所示。图 6 和图 7 是在其他参数相同、链路带宽分别为 1 Gb/s 和 2 Gb/s 的条件下，存储扩展的性能。对比图 6 和图 7 不难发现，存储扩展的性能并没有随网络带宽的增加而增加。

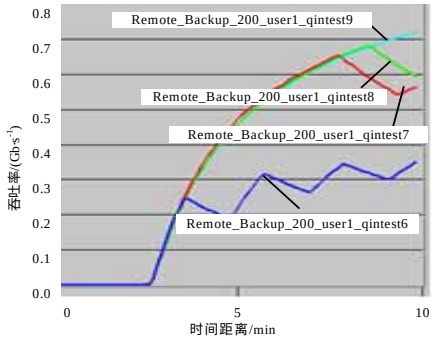

图 6　吞吐率与带宽的关系(1 Gb/s)

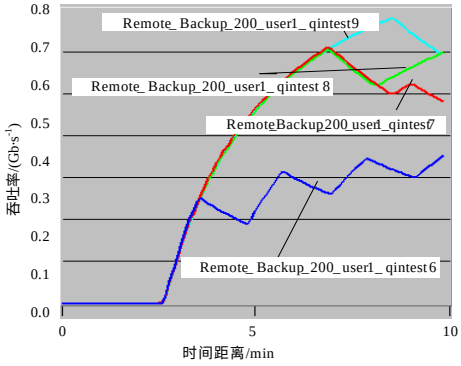

图 7　吞吐率与带宽的关系(2 Gb/s)

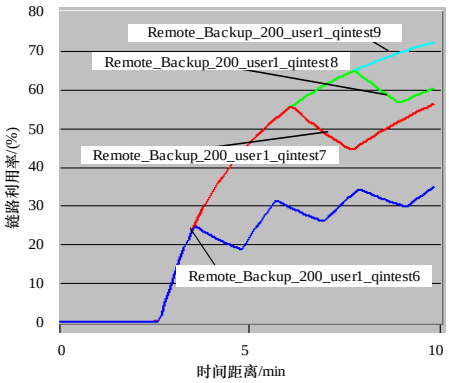

图 8　链路利用率与带宽的关系(1 Gb/s)

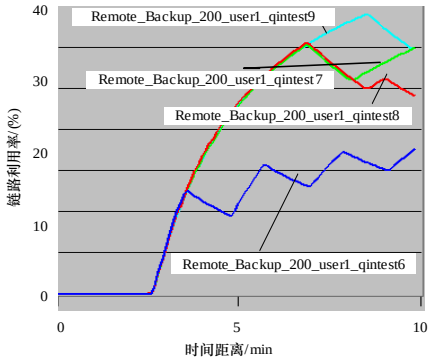

图 9　链路利用率与带宽的关系(2 Gb/s)

对比图 8 和图 9 可以看出，存储扩展链路带宽加倍后，链路的利用率只有原来的一半，得到的结论与图 6、图 7 相符。即在一定条件下，如果不增加信用值而仅仅增加链路的带宽，不仅不能提高存储扩展的性能，而且还会导致链路资源的浪费。

4.4　性能与 FC 帧大小的关系

在缓存大小一定的情况下，信用值的大小将随 FC 帧容量的增加而减小，由此影响存储扩展的性能和链路的利用率。为了研究存储扩展的性能与 FC 帧大小之间的关系，建立 3 个仿真环境，每个环境中又设立 4 个场景。3 个仿真环境中 FC 帧的有效负载分别为 128 bit, 2 100 bit 和 2 400 bit，每个场景中链路的传播时延分别为 5 s, 3 s, 2 s 和 1 s。3 种仿真环境中存储扩展性能与 FC 帧大小之间的关系见图 10~图 12。

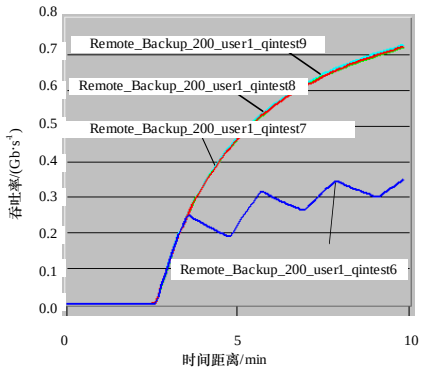

图 10　FC 帧有效载荷 128 bit

从图 10 可以看出，当 FC 帧有效负载为 128 bit 时，传播时延为 3 s, 2 s 和 1 s 的 3 条性能曲线基本重合，存储扩展性的性能在 720 Mb 左右，只有传播时延为 5 s 的场景性能较低，存储扩展的性能约为 320 Mb；当 FC 帧有效负载为 2 100 bit 时(对应图 11))，传播时延为 2 s 和 1 s 的 2 条性能曲线基本重

合并维持图 10 的性能，传播时延为 3 s 的性能开始下降，传播时延为 5 s 的性能保持在低位不变。

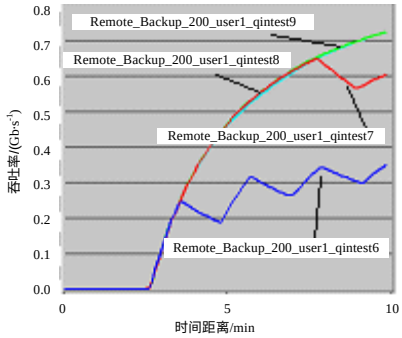

图 11　FC 帧有效载荷 2 100 bit

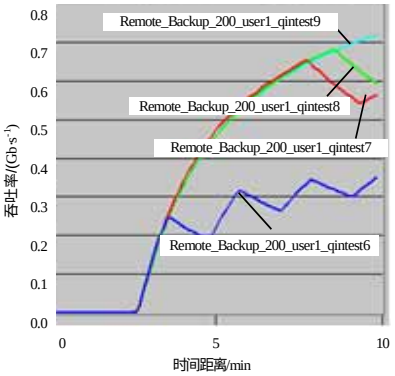

图 12　FC 帧有效载荷 2 400 bit

当 FC 帧有效负载为 2 400 bit 时(对应图 12)。只有传播时延为 1 s 的性能曲线维持在高位不变，传播时延为 3 s 和 2 s 的 2 条性能曲线继续下降，传播时延为 5 s 的性能保持在低位不变。对比图 10~图 12 的 3 种情况不难发现，随着 FC 帧容量的增加，信用值也需要增加，否则，存储扩展的性能将随 FC 帧的增加而降低。

5　结束语

CWDM 能为高性能的存储扩展提供高带宽的网络，满足高级别容灾应用对时间的要求。在存储扩展应用中，为了保持存储扩展的高性能，应用中要根据实际情况调整 FC 流量控制协议的 credit。随着扩展距离和 FC 帧负载长度的增加，需要更多的 credit。

参考文献

[1] Chen Lin. DoSTOR 存储分析 IDC: 2010 年数据量达 988 EB[EB/OL]. (2007-03-07). http://www.dostor.com/a/r/2007-03-07/0002231883.shtml.

[2] 林　闯. 计算机网络和计算机系统的性能评价[M]. 北京: 清华大学出版社, 2001.

[3] 余寅辉, 余镇危, 杨传栋, 等. SAN 存储系统的性能分析模型[J]. 计算机工程, 2007, 33(10): 271-273.

[4] 王亚兵. 新一代 SDH/SONET 技术[J]. 光通信技术, 2005, (2): 4-6.

[5] 白　杉, 周　洁. CWDM 技术浅析[J]. 现代通信, 2003, (8): 14-16.

[6] 吴德明, 徐安士, 朱立新, 等. 波分复用和频分复用光纤通信技术研究[J]. 北京大学学报: 自然科学版, 1998, 34(2/3): 214-219.

(上接第 249 页)

3.4　配置模块

配置模块主要由 Config 包实现，其功能主要是支持配置管理的类和接口。配置模块提供了一种将边缘服务器程序从属性文件、XML 文件和数据库等配置文件的使用中松耦合出来的机制。配置模块使得在不修改代码的情况下加入新的配置文件格式成为可能。其中配置项可以是布尔型、整型、字符串或者文件，也可以扩展其他新的类型，配置项也可以组织成子配置模块，实现层级管理。

如图 5 所示，ConfigManager 执行读卡器和事件数据的集中管理，ConfigManagerImpl 是具体实现 ConfigManager 接口的一个 singleton 类。ReaderConfig 表示针对读卡器的配置数据，ReaderConfig.ReaderConfigData 是包含特定读卡器的配置数据，其他的几类配置文件也是类似的含义。其中，ConfigException 表示配置处理中出现的错误。

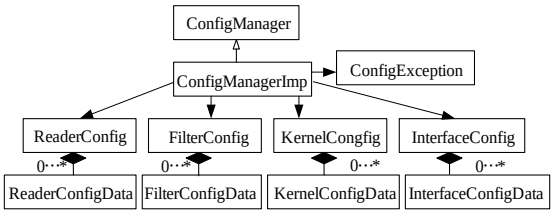

图 5　ALE 引擎设配置模块主要类图

4　结束语

虽然目前国外已有不少的大学和研究团体在做 RFID 中间件方面的设计和研究，但是，RFID中间件的研究还只能说是刚刚起步，研究还大都处于试验阶段，还都存在着或多或少的不足和限制[5]。而SUN, Microsoft, IBM, Oracle, BEA等厂商所提的RFID中间件的设计大多是基于自己目前所研发的核心产品或技术的应用，有太大的依赖性和较小的扩展性。

本文提出的基于 EPC ALE 引擎设计和实现，在一些技术和方法方向也还需要经过更多的分析，以进一步增强其功能，下一步的研究工作主要是通过完善目前的引擎，并且将其进行扩展来实现支持完整的 RFID 协议堆栈。

参考文献

[1] Clark S, Traub K, Anarkat D, et al. Auto-ID Savant Specification 1.0[R]. Auto-ID Center, Tech. Rep: MIT-AUTOID-TM-003, 2003.

[2] Ranasinghe D C, Leong K S, Ng M L, et al. A Distributed Architecture for a Ubiquitous RFID Sensing Network[C]//Proc. of International Conference on Intelligent Sensors, Sensor Networks and Information Processing. [S. l.]: IEEE Press, 2005: 7- 12.

[3] Rooney S, Bauer D, Scotton P. Edge Server Software Architecture for Sensor Applications[C]//Proc. of the 2005 Symposium on Applications and the Internet. Washington D. C., USA: IEEE Computer Society, 2005: 64-71.

[4] Luckham D C, Frasca B. Complex Event Processing in Distributed Systems[R]. Stanford University, Technical Report: CSL-TR-98-754, 1998.

[5] 丁振华, 李锦涛, 冯　波, 等. RFID 中间件研究进展[J]. 计算机工程, 2006, 32(21): 9-11.