

# 数据中心虚拟化技术的 TCP 性能分析与研究

杨 健,陈春玲,欧阳志友

(南京邮电大学 计算机学院,江苏 南京 210046)

**摘 要:**文中的工作主要包括针对服务器虚拟化对 TCP 协议性能影响的测试,为云计算数据中心服务器虚拟化的部署提供参考与指导。虚拟化目的在于有效提高硬件资源的利用率,实施计算资源的隔离,避免不同用户或不同应用间计算资源的干扰,提升计算资源的可靠性与容错能力等。XEN 是一种非常流行的虚拟化软件技术,现已被大量应用于数据中心服务器应用整合等领域。文中测试了基于 XEN 虚拟化平台上虚拟机 TCP 通信的性能,分析了 XEN 虚拟化技术对 TCP 性能的影响。在云计算数据中心越来越多采用虚拟化技术对其服务器进行管理的背景下,文中的研究表明云计算数据中心的虚拟化应避免盲目性。

**关键词:**XEN;服务器虚拟化;超虚拟化;完全虚拟化;TCP;性能

**中图分类号:**TP393

**文献标识码:**A

**文章编号:**1673-629X(2012)12-0115-04

## Analysis and Research on TCP Performance of Virtualization Technology in Data Center

YANG Jian, CHEN Chun-ling, OUYANG Zhi-you

(Computer Dept of Nanjing University of Post and Telecommunications, Nanjing 210046, China)

**Abstract:** The work in this paper mainly includes the test of server virtualization for TCP performance effect, providing reference and guidance for the deployment of cloud computing data center server virtualization. Virtualization purpose is to improve the utilization rate of hardware resources, implement computing resources isolation, avoid computing resources interference for different users or between different application, improve the reliability of the computing resources and fault tolerant ability, etc. XEN is a popular software technology for virtualization which has been widely applied to fields like application integration of servers of data centers. It tests TCP communication capability of virtual machine based on XEN virtualized platform and analyzes the influence of XEN virtualization technology on TCP performance. In the background of cloud computing data center more and more using virtualization technology on the server management, the studies show that cloud computing data center virtualization should avoid blindness.

**Key words:** XEN; virtualization of server; hyper-virtualization; complete virtualization; TCP; performance

## 0 引 言

通过虚拟化技术既可以将多台计算机组合成一台计算机的形式呈现出来,也可以将一台物理服务器分割成多个相互独立、互不干扰的虚拟化环境。文中的服务器虚拟化指的是将一台物理服务器逻辑上划分成多个虚拟机,在其上可以并发地运行一个或多个操作系统实例的技术<sup>[1-5]</sup>。虚拟化技术在服务器整合、服务器安全、灾难恢复以及测试环境等领域具有很好的应用前景。目前,根据是否需要修改客户机操作系统的源代码,可将虚拟化技术分为超虚拟化(Para-virtualization)

和完全虚拟化(Full-virtualization)。超虚拟化技术较之完全虚拟化技术,具有更好的性能,而完全虚拟化不需要修改客户机操作系统,因此具有良好的兼容性<sup>[6]</sup>。

XEN 源于英国剑桥大学计算机实验室开发的一个虚拟化开源项目,目前成为了一种非常流行的虚拟化软件技术,并得到了业界许多供应商的支持,其中 Novell SUSE Linux Enterprise Server 最先采用了 XEN 虚拟技术。XEN 一开始只支持泛虚拟化,2005 年初,INTEL 和 AMD 推出了基于处理器硬件的虚拟化技术,使得 XEN 既可以支持完全虚拟化(又被称为硬件辅助的虚拟化,Hardware assisted Virtualization),也可以支持超虚拟化。XEN 提供了 Linux 上免费的虚拟化解决方案,现已被大量应用于数据中心服务器应用整合等领域<sup>[7,8]</sup>。

研究表明,目前数据中心中 TCP 占了 99% 以上的

收稿日期:2012-05-03;修回日期:2012-08-06

基金项目:国家自然科学基金资助项目(60973046,60905040);南京邮电大学青蓝项目(NY208021)

作者简介:杨 健(1978-),男,江苏徐州人,讲师,博士研究生,研究方向为信息网络、物联网。

流量,其中既包括时延敏感的查询类应用,也有对吞吐率敏感的长流。因此文中从这两个方面入手,测量了 XEN 虚拟化技术对其虚拟机上 TCP 通信性能的影响<sup>[9-12]</sup>。

## 1 XEN 虚拟化体系结构

XEN 虚拟化环境的构成如图 1 所示:

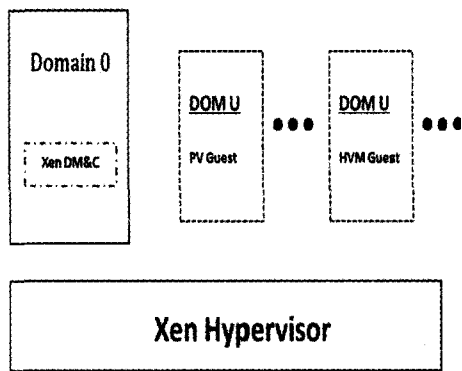


图 1 XEN 体系结构图

· XEN Hypervisor。

XEN Hypervisor 负责各虚拟机的 CPU 调度和内存的分配,控制着各虚拟机所共用的进程运行环境。

· Domain 0。

本质上也是一个虚拟机,但它是唯一的,需要运行在其它虚拟机之前。拥有操作 I/O 资源的特权,与其它的虚拟机(Domain U)进行交互。

· XEN DM&C。

在 XEN DM&C 中提供了一系列用于 Domain 管理和控制的 daemon 和工具,管理命令都是传递给 Domain 0,最终作用于 XEN Hypervisor 上。

· Domain U。

Domain U 分两种类型:PV Guest 和 HVM Guest,其中 PV Guest 是准虚拟化的虚拟机实例,要求操作系统来配合虚拟环境执行硬件操作,因此操作系统需要考虑如何与 Hypervisor 交互,所以一般在 PV Guest 实例上运行的客户操作系统是修改过的 Linux。HVM Guest 是完全虚拟化的虚拟机实例,在此实例上运行的操作系统会认为是工作在真实硬件环境中,因此不需要修改客户操作系统,就可以在上面运行。

## 2 TCP 性能测量

### 2.1 测试环境

测试使用的物理机器配置:

CPU: Intel Core I5, 4Core@3.2GHz

Memory: 4G

网卡: Intel 82578DM 千兆网卡

其中

Dom0 的配置: vCPU: 4Core, Memory: 1G

DomU 的配置: vCPU: 4Core, Memory: 1G

组网使用的交换机: 华为 S2300

软件环境: SUSE 11 Linux Enterprise Server

VMM: XEN4.0.2

VM Guest OS: SUSE 11 Linux Enterprise Server

测试网络拓扑如图 2 所示:

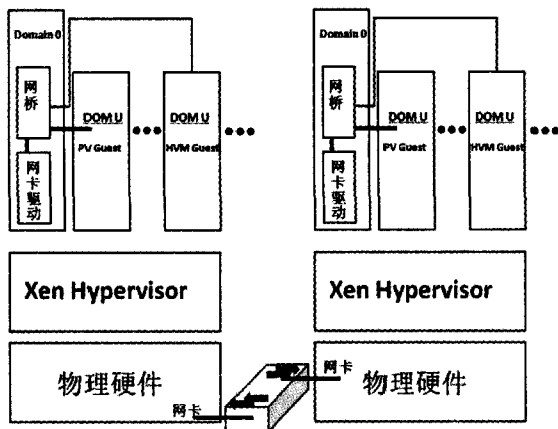


图 2 测试网络拓扑

### 2.2 测试设计

对以下六种场景进行了测试:

1) Dom0-Dom0。

Dom0 与 Dom0 间通信测试作为基准,用于研究虚拟化对 TCP 性能的影响。

2) Dom0-PVM。

Dom0 与超虚拟化虚拟机间的通信,测试单一的超虚拟化对 TCP 性能的影响。

3) Dom0-FVM。

Dom0 与完全虚拟化虚拟机间的通信,测试单一的完全虚拟化对 TCP 性能的影响。

4) PVM-PVM。

超虚拟化虚拟机与超虚拟化虚拟机间的通信,测试超虚拟化虚拟机间 TCP 的通信性能。

5) PVM-FVM。

超虚拟化虚拟机与完全虚拟化虚拟机间的通信,测试超虚拟化虚拟机与完全虚拟化虚拟机间 TCP 的通信性能。

6) FVM-FVM。

完全虚拟化虚拟机与完全虚拟化虚拟机间的通信,测试完全虚拟化虚拟机间 TCP 的通信性能。

### 2.3 测试结果

以下是使用 Netperf 测试工具,针对以上测试场景测得的数据。

#### 2.3.1 查询类应用测试一

每次交易不需要重新建立连接,服务器接收缓冲区的大小设为 873800 字节,客户端发送缓冲区大小为

16384 字节,发送的查询请求为 64 字节,应答消息的大小分别设为 256,512,1024,2048,4096,8192,16384 进行测试,图 3 中的横坐标表示应答消息的大小,其中  $K=1024$ ,纵坐标是每秒平均的交易次数,每个数据点的测试时间为 90 秒,结果如下:

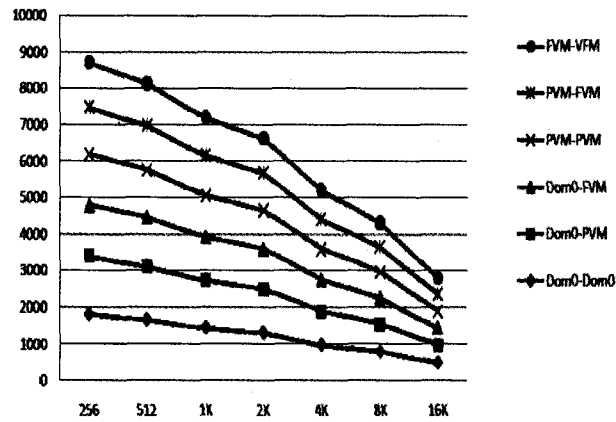


图 3 交互式应用测试结果一

从上面的图 3 数据可以看出,虚拟化,尤其是完全虚拟化,对较短的应答消息的查询类应用的性能影响比较大。

### 2.3.2 查询类应用测试二

每次交易需要重新建立连接,服务器接收缓冲区的大小设为 873800 字节,客户端发送缓冲区大小为 16384 字节,发送的查询请求为 64 字节,应答消息的大小分别设为 256,512,1024,2048,4096,8192,16384 字节进行测试,图 4 中的横坐标表示应答消息的大小,其中  $K=1024$ ,纵坐标是每秒平均的交易次数,每个数据点的测试时间为 90 秒,结果如下:

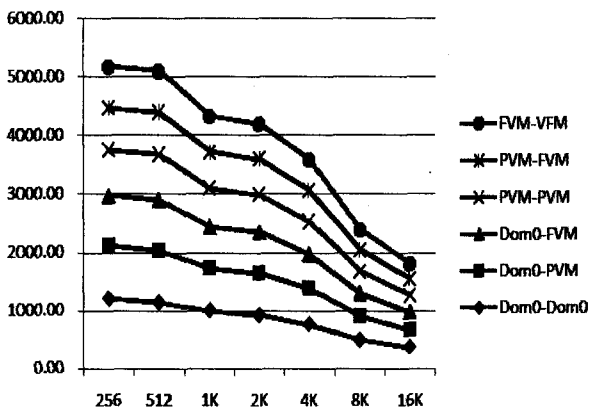


图 4 查询类应用测试结果二

从上面的图 4 数据可以看出,如果每次查询都需要新建连接时,虚拟化,尤其是完全虚拟化,对较短的应答消息的查询类应用的性能影响会更大。

### 2.3.3 吞吐率测试一

服务器接收缓冲区的大小设为 873800 字节,客户端发送缓冲区大小分别设为 1024,2048,4096,8192,16384,32768,65536 字节进行测试,图 5 中的横坐标表

示发送缓冲区的大小,其中  $K=1024$ ,纵坐标为吞吐率,单位为 Mbps,每个数据点的测试时间为 90 秒,结果如下:

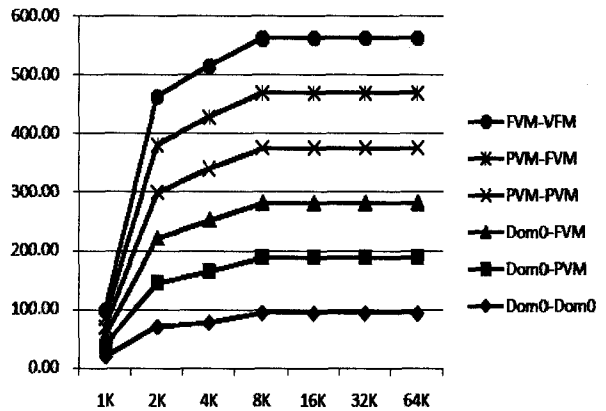


图 5 吞吐率测试结果一

上面的吞吐率测试表明,对于较小的发生缓冲区(比如 1024 字节),虚拟化技术对 TCP 产生了较大的影响,但当发送缓冲区较大时,无论哪种虚拟化技术对 TCP 性能几乎没有影响。

### 2.3.4 吞吐率测试二

客户端发送缓冲区大小设为 65536 字节,服务器接收缓冲区的大小分别设为 1024,2048,4096,8192,16384,32768,65536 字节进行测试,图 6 中的横坐标表示接收缓冲区的大小,其中  $K=1024$ ,纵坐标为吞吐率,单位为 Mbps,每个数据点的测试时间为 90 秒,结果如下:

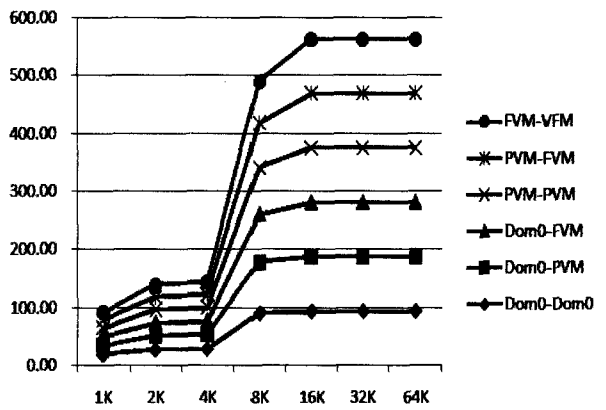


图 6 吞吐率测试结果二

上面的吞吐率测试表明,对于较小的接收缓冲区(比如不超过 8192 字节),虚拟化技术对 TCP 产生了较大的影响,但当接收缓冲区较大(超过 16384 字节)时,无论哪种虚拟化技术对 TCP 性能几乎没有影响。

## 3 结束语

以上测试仅仅是针对虚拟化对 TCP 性能的影响,测试过程中除了运行的测试工具以外并未运行其他的应用,而且未参与测试的 VM 只处于活动状态,也未运

行任何应用,这样设计测试过程的目的是为了排除其他因素对 TCP 性能测试的干扰,而仅仅分析虚拟化对 TCP 性能的影响。

通过测试可以看出,虚拟化技术,尤其是完全虚拟化,其开销明显影响了 TCP 交互式、查询类应用的性能。这主要是因为虚拟化技术中的网络 I/O 机制,会增加 TCP 的 RTT(Round Trip Time),因此会对数据中心中时延敏感的应用产生显著影响。但通过适当地设置收发缓冲区的大小,可以使得虚拟机上进行批量数据传输的 TCP 通信不受虚拟化技术产生的不利影响。

以上测试了虚拟化技术对 TCP 性能的影响,未在文中列出的测试还表明 Dom0 中的 Memory、CPU 分配会对 TCP 性能产生很大的影响。这是因为 VM 与外面的网络通信都需要经过 Dom0 进行中转,若 Dom0 未分配足够的资源,反而会成为性能瓶颈。因此,通过改进 XEN Hypervisor 的 I/O 机制,避免经过 Dom0 的中转应该是一个值得研究的方向,因为这不仅可以降低 I/O 时延,还可以有效提升资源的分配效率。

文中测试与分析了 XEN 桥接组网方式下的 TCP 性能,另外探讨 XEN 中的其他组网方式(比如 NAT、Routing 模式)下 TCP 的性能,以及其它类型的 Guest OS(比如 RedHatLinux、CentOS、Windows 等)下的 TCP 性能也是值得关注的一项工作。

#### 参考文献:

- [1] Alizadehzy M, Greenberg A, Maltzy D A, et al. DCTCP[C]//Proc of SIGCOMM. [s. l.]:[s. n.], 2010.

(上接第 114 页)

资源,通过试验对比对其模型格式进行了深入剖析,解决了解码、绘制中的关键技术,进行了良好的组件化封装,以该组件作为重要组成部分的空间场景可视化控件可以方便地构建战场可视化系统。

#### 参考文献:

- [1] 张飞宇,闫晓勇.基于 Creator/Vega 的寻的导弹飞控系统[J].计算机工程,2010,36(9):43-45.
- [2] 黄权,学军.基于 OpenGL 的卫星跟踪仿真[J].计算机技术与发展,2007,17(2):131-134.
- [3] 张龙慧,宋杨,唐俊,等.基于 Open Inventor 的卫星姿态控制可视化仿真[J].计算机技术与发展,2009,19(11):214-217.
- [4] 胡平平,刘建明,王晶杰. OpenGL 显示 3DS 模型若干问题的研究[J].工程图学学报,2010(4):189-193.
- [5] Velsen M V. 3D-Studio File Format(.3ds)[EB/OL]. (1998-10-05)[1998-10-05]. <http://www.the-labs.com/>

- [2] Whiteaker J, Schneider F, Teixeira R. Explaining Packet Delays under Virtualization[J]. Computer Communication Review, 2011, 41(1):38-44.
- [3] Kangarlou A, Gamage S, Kompella R R, et al. vSnoop: Improving TCP Throughput in Virtualized Environments via Acknowledgement Offload[C]//Proc of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis. Washington, D C: IEEE Computer Society, 2010.
- [4] 胡翔羽,兰巨龙,邹江兴.基于固件虚拟的承载网络资源重构模型研究[J].通信学报,2011,32(12):36-42.
- [5] 张伟哲,张宏莉,张迪,等.云计算平台中多虚拟机内存协同优化策略研究[J].计算机学报,2011,34(12):2265-2277.
- [6] Wang Guohui, Eugene Ng T S E. The Impact of Virtualization on Network Performance of Amazon EC2 Data Center[C]//INFOCOM2010. [s. l.]:[s. n.], 2010.
- [7] 罗军舟,金嘉晖,宋爱波,等.云计算:体系架构与关键技术[J].通信学报,2011,32(7):3-21.
- [8] 顾晓峰,王健.基于 Intel VT-x 的 XEN 全虚拟化实现[J].计算机技术与发展,2009,19(9):242-245.
- [9] 朱团结,艾丽蓉.基于共享内存的 Xen 虚拟机间通信的研究[J].计算机技术与发展,2011,21(7):5-8.
- [10] 王位杰,王锋.基于网络的虚拟调试环境的研究[J].计算机技术与发展,2011,21(7):46-49.
- [11] 马飞,刘峰,刘真.云计算中高效节能的虚拟机分布式管理方法[J].计算机工程,2012,38(11):5-11.
- [12] 温少君,陈俊杰,郭涛.一种云平台中优化的虚拟机部署机制[J].计算机工程,2012,38(11):18-21.

Blender/3dsspec. html.

- [6] 肖羽,李光耀,王文举.基于 3DS 的 OpenFlight 模型构建方法[J].计算机应用,2009,29(6):302-304.
- [7] 刘立嘉,盛业华,周炯,等.基于 DXF 的 OpenFlight 数据创建方法研究[J].测绘科学,2008,33(4):215-217.
- [8] Autodesk Inc. AutoCAD 2007 DXF Reference v. u. 21. 1. 01 [EB/OL]. (2006-03-07)[2007-01-08]. <http://images.autodesk.com/adsk/files/acad-dxf.pdf>.
- [9] 宋殿宇,韩潮.关于在 VRML 技术中应用 STK 模型的研究[J].计算机仿真,2004,21(10):122-125.
- [10] 韩潮,曲艺. Open Inventor 在 STK 模型转换中的应用[J].计算机仿真,2005,22(10):63-66.
- [11] 张占月,徐艳丽,曾国强.基于 STK 的航天任务仿真方案分析[J].装备指挥技术学院学报,2006,17(1):48-51.
- [12] Wright R S, Lipchak B, Haemel N. OpenGL superbible[M]. 4th ed. American: Addison-Wesley Publication, 2007:409-414.