# 基于 KVM 的虚拟化技术研究

□彭晓平　张雪坚　云南电网有限责任公司信息中心　黄波　昆明能讯科技有限责任公司

【摘要】　近年来，虚拟化技术发展迅速，尤其是 KVM 虚拟化技术已经被整合到 linux 内核中，其特点是能快速进行资源的整合，进而最大限度的对已整合资源进行分配，KVM 虚拟化是基于 Linux 内核的开源虚拟机平台，是硬件虚拟化的扩展及 QEMU 的升级版，在很大程度上已取代 Xen 成为 Linux 系统上创建和支持虚拟机的默认开源方案，文中解释了基于硬件虚拟化技术解决方案 Kernel-based Virtual Machine（KVM）的系统架构，深入地剖析了 KVM 虚拟机的核心技术和工作原理，分析了 KVM 的虚拟化拓扑结构灵活、硬件配置方案简单以及数据统一集中管理。

【关键词】　虚拟化　KVM　架构

## 引言

虚拟化技术最早出现在大型机时代。上世纪 60 年代，IBM 开始在其 CP-40 大型机系统中尝试虚拟化的实现，后来在 System/360-67 中采用，并衍生出 VM/CMS 到后来的 z/VM 等产品线，这项技术极大地提高了大型机资源利用率，由于当时软硬件水平的限制，只应用于少数型号的大型机上。20 世纪 90 年代后期，随着微处理器性能的不断提升以及多核处理器的发展，使得 KVM 虚拟化技术开始迅速升温，这项属于大型机及专利的技术开始在普通 X86 计算机上应用并成为当前计算机发展和研究的一个热点方向。目前，无论是在高性能服务器领域，还是在云计算等领域，KVM 虚拟化技术的应用都得到了蓬勃发展。

## 一、KVM 虚拟化技术概述

KVM (Kernel-based Virtual Machine，基于内核的虚拟机)，是一种用于 Linux 内核中的虚拟化基础设施，是硬件支持虚拟化技术（Intel VT 或 AMD-V）的 Linux 的全虚拟化解决方案。KVM 采用寄居式虚拟化架构，是 Linux 内核中的一个可装载模块，KVM 本身不执行任何硬件模拟，需要修改过的 QEMU 向它提供模拟的 I/O，其功能是将 Linux 内核转换成一个裸金属架构的 Hypervisor，实际上 KVM 只是虚拟化解决方案的一部分，其底层需要处理器支持，为多个操作系统提供虚拟化处理器，I/O 通过修改过的 QE-MU 进行的。KVM 最初是由以色列的公司 Qumranet 开发，并于 2006 年 12 月被合并在发布于 2007 年 2 月的 linux 2.6.20 内核中。RedHat 公司在 2008 年 9 月收购 Qumranet 公司后，在 RHEL 6 及以后发行版中使用 KVM 作为默认的虚拟化引擎。KVM 是免费的开源系统并在迅速发展当中，是目前唯一进入 Linux 核心的虚拟化解决方案。其架构如图 1 所示。

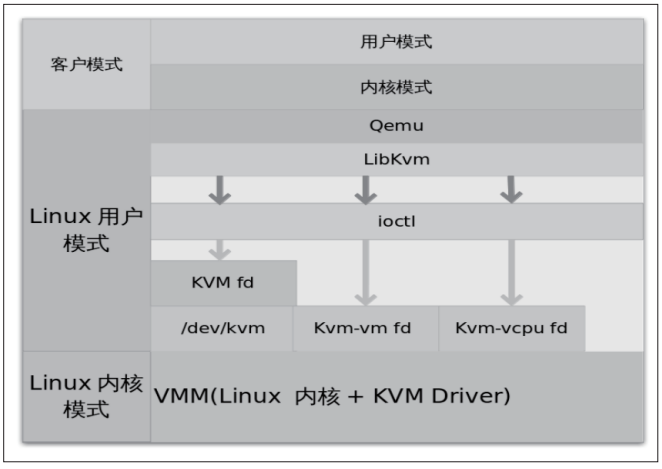


图 1　KVM 系统架构

上图 kvm 已经是内核模块，被看作是一个标准的 linux 字符集设备（/dev/kvm）。Qemu 通过 libkvm 应用程序接口，用 fd 通过 ioctl 向设备驱动来发送创建，运行虚拟机命令。KVM 包含内核模块和处理器模块两部分，内核模块 kvm.ko 提供核心的虚拟化支持，处理器模块 kvm-inte1. ko 和 kvm-

因自然灾害造成的断纤有多种形式，本文以其中一种常见的断纤情况进行分析：

接入层的芦山隆兴—芦山青龙乡—芦山玉溪—芦山太平任意断纤，芦山中心站 2—宝兴中心站—芦山太平任意断纤。其中芦山隆兴—芦山青龙乡—芦山玉溪—芦山太平任意断纤，芦山隆兴节点的 LSP1:1 保护会将业务倒换到备用，即图上黑色箭头的路径。芦山中心站 2—宝兴中心站—芦山太平处断纤，则会触发汇聚层的 Wrapping 环网保护，业务不会中断。若断纤的顺序不同，只是触发不同倒换机制的顺序不同，不会造成业务中断。

另外，由于 OTN 下沉到汇聚环，PTN 接入层通过汇聚 PTN 节点后，业务直接在 OTN 上承载，OTN 网状网多路由等保护方式也能有效的保护业务。

论文通过分析雅安传送网存在的问题，对传送网的抗灾应用策略进行较深入的研究，并以此提出传送网系统的抗灾建设思路，以增强传送网系统在地震等自然灾害中的可靠性与可用性。

总之，随着传送网正朝着智能化、高速率大容量和多业务能力的方向发展，传送网的建设应该秉承总体规划，计划实施的思想，结合三年滚动规划，逐渐完成传送网自身的强壮优化，并适时引入新技术新思路，优化网络结构，保证传送网的再生和抗灾能力。

amd. ko 分别提供了对 Intel 和 AMD 处理器虚拟化技术的支持，实现客户模式的切换，处理因为 I/O 或者其他指令引起的从客户模式退出（VM_EXIT）,kvm 模块工作在这个模式下。KVM 通过加载 kvm.ko 内核模块将 Linux 内核转换为一个 VirtualMachineMonitor(VMM，Hypervisor)，因此 KVM 可以随着 Linux 标准内核的升级而获得性能提升 ( 如调度程序、内存支持等 )。虚拟机对应成为标准的 Linux 进程，因而可以用标准的 Linux 进程管理机制进行管理。普通的 Linux 进程有 Linux 内核模式和 Linux 用户模式两种运行模式，内核模式表示代码执行的特权模式，用户模式表示代码执行的非特权模式。在 KVM 系统中为 Linux 引入了一种新的进程模式，新的模式称为客户模式，客户模式用来执行虚拟机操作系统的非 I/O 代码。在客户模式中包含内核模式和用户模式两种标准模式，VM 操作系统可在内核模式下运行标准的内核，在用户模式下支持自己的内核和用户空间应用程序，在 kvm 的模型中，每一个 Gust OS 都是作为一个标准的 linux 进程，都可以使用 linux 进程管理命令管理。

虚拟机操作系统的 I/O 操作是由修改过的 QEMU 支持的，QEMU 是一种用动态翻译技术实现的快速指令集层虚拟机，它支持整个计算机系统的模拟，包括多种处理器 (X86、ARM、PowerPC 等 )、磁盘、图形适配器和网络设备等。KVM 是用硬件虚拟化技术代替了 QEMU 的动态翻译技术，实现虚拟机操作系统代码直接由硬件处理从而提高系统性能。VM 操作系统生成的 I/O 请求会被截获并转发到用户空间，由 QEMU 的设备模型来模拟 I/O 操作，在需要的情况下触发真实的 I/O 操作。

## 二、KVM 与 QEMU、libvirt 组件关系

KVM 是 linux 内核的模块，它需要 CPU 的支持，采用硬件辅助虚拟化技术 Intel-VT，AMD-V，内存的相关如 Intel 的 EPT 和 AMD 的 RVI 技术，通过 /dev/kvm 暴露接口，用户态程序可以通过 ioctl 函数来访问这个接口，而 Qemu 将 KVM 整合进来，通过 ioctl 调用 /dev/kvm 接口，将有关 CPU 指令的部分交由内核模块来做。kvm 负责 cpu 虚拟化 + 内存虚拟化，实现了 cpu 和内存的虚拟化，但 kvm 不能模拟其他设备，所以用 qemu 模拟 IO 设备（网卡，磁盘等），kvm 加上 qemu 之后就能实现真正意义上服务器虚拟化。因为用到了上面两个东西，所以称之为 qemu-kvm。libvirt 是目前使用最为广泛的对 KVM 虚拟机进行管理的工具和 API，Libvirtd 是一个 daemon 进程，可以被本地的 virsh 调用，也可以被远程的 virsh 调用，Libvirtd 调用 qemu-kvm 操作虚拟机。

KVM 、QEMU 、libvirt 三者之间的关系如下图 2。

### 2.1 KVM 与 QEMU 的关系

QEMU 是一个开源的模拟器，作者是法布里斯·贝拉 (Fabrice Bellard)。准确来说，KVM 是 Linux kernel 的一个模块。可以用命令 modprobe 去加载 KVM 模块。加载了模块后，才能进一步通过其他工具创建虚拟机，但仅有 KVM 模块是 远远不够的，因为用户无法直接控制内核模块去作事情，你还必须有一个运行在用户空间的工具才行。这个用户空间的工具，kvm 开发者选择了已经成型的开源虚拟化软件 QEMU。说起来 QEMU 也是一个虚拟化软件。它的特点是可虚拟不同的 CPU。比如说在 x86 的 CPU 上可虚拟一个 Power 的 CPU，并可利用它编译出可运行在 Power 上的程序，KVM 使用了 QEMU 的一部分，并稍加改造，就成了可控制 KVM 的用户空间工具了。所以你会看到，官方提供的 KVM 下载有两 大部分 (qemu 和 kvm) 三个文件 (KVM 模块、QEMU 工具以及二者的合集 )。也就是说，你可以只升级 KVM 模块，也可以只升级 QEMU 工具。

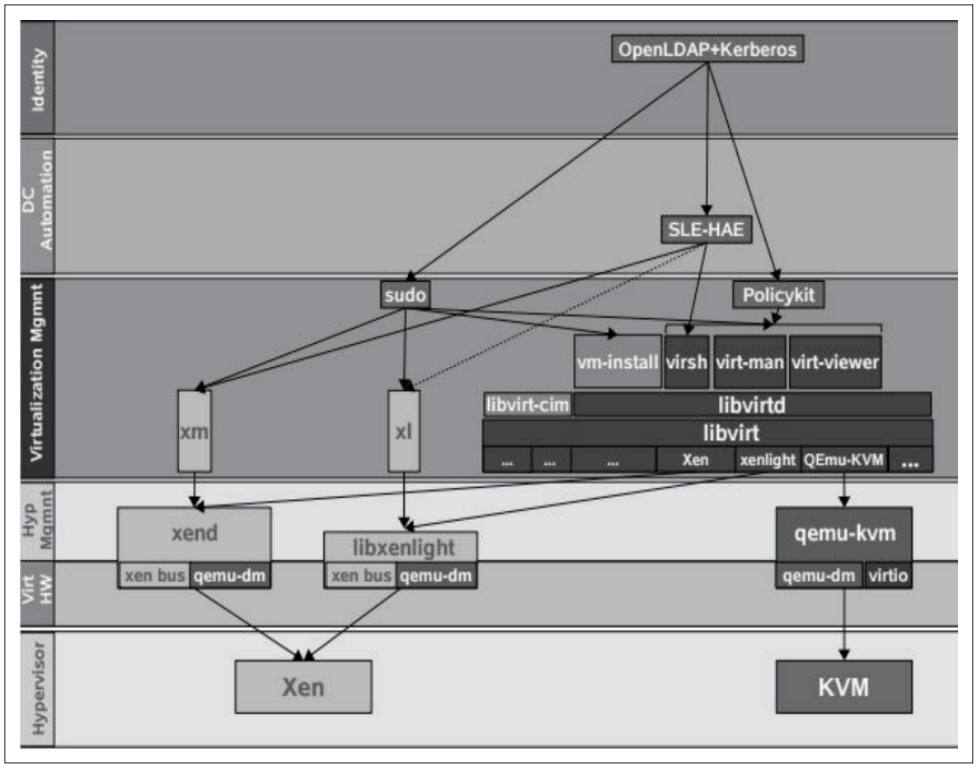


图 2

### 2.2 KVM 与 libvirt 的关系

libvirt 是一套实现 Linux 虚拟化功能的开源 API，旨在提供一种单一的方式管理多种不同的虚拟化方案。libvirt 由一套 API 库，一个 libvirtd 服务，以及一个 virsh 命令行管理工具组成。虽然 libvirt 是 C 开发的，但是可以很好的支持主流的编程语言，包括 C, Python, Perl, Java 等等。最新的发行版还包含了一系列基于 libvirt 的工具，用于简化虚拟机的维护管理：

1.virt-install: 一个创建虚拟机的工具，支持从本地镜像或者网络镜像 ( NFS、FTP 等等 )启动。

2.virsh: 一个交互式 / 批处理 shell 工具，可以用于完成虚拟机的日常管理工作。

3.virt-manager: 一个通用的图形化管理工具，可以用来管理本地或远程的 Hypervisor 及其虚拟机。

4.virt-viewer: 一个轻量级的、能够安全连接到远程虚拟机的图形控制台工具。

## 三、KVM 虚拟机部署

### 3.1 环境检查及软件配置

KVM 的虚拟化需要硬件支持（如 Intel VT 技术或者 AMD V 技术），是基于硬件的完全虚拟化技术。使用 grep –E ––color ‘vmx|svm’ /proc/cpuinfo 命令进行检测，有输出结果说明 CPU 支持虚拟化。部分服务器默认是关闭虚拟化技术的，需要进入 BIOS 打开 CPU 的虚拟化支持。 Linux 内核是从 2. 6. 20 版本开始集成 KVM，因此 Linux 内核版本必须在此之上。使用 uname—a 命令即可查看 Linux 内核版本。使用 lsmod 命令可查看 KVM 模块是否加载成功，如果未加载成功可以使用命令 modprobekvm 重新载入。

### 3.2 VM 操作系统实例化

3.2.1 建立 KVM 虚拟磁盘镜像

虚拟磁盘镜像在逻辑上是提供给虚拟机使用的硬盘，在物理上可以是 Linux 系 统内一普通镜像文件，也可以是真实的物理磁盘或分区。本方案设计中将虚拟机集中存储在 SAN 存储阵列中，采用文件方式，用 dd 命令创建如下 dd if= /dev/ zero of= hdisk.img bs= 2G count= 15,dd 命令创建一个名为 hdisk.img 的容量为 15G 的虚拟磁盘。虚拟磁盘并不会立即分配全部空间，而是根据使用情况在不超过 15G 范围内动态分配。

3.2.2 配置 KVM 虚拟网络

KVM 虚拟化有两种网络模式：Bridge 网桥模式和 NAT 网络地址转换模式，其中 Bridge 方式适用于服务器主机的虚拟化，NAT 方式适用于桌面主机的虚拟化。客户机安装完成后，需要为其设置网络接口，以便和主机网络，客户机之间的网络通信。事实上，如果要在安装时使用网络通信，需要提前设置客户机的网络连接。KVM 客户机网络连接有两种方式：NAT 模式即用户网络（User Networking）：让虚拟机访问主机、互联网或本地网络上的资源的简单方法，但是不能从网络或其他的客户机访问客户机，性能上也需要大的调整。Bridge 模式即虚拟网桥（Virtual Bridge）：这种方式要比用户网络复杂一些，但是设置好后客户机与互联网，客户机与主机之间的通信都很容易。

1）Bridge 方式

Bridge 方式原理：Bridge 方式即虚拟网桥的网络连接方式，是客户机和子网里面的机器能够互相通信。可以使虚拟机成为网络中具有独立 IP 的主机。

桥接网络（也叫物理设备共享）被用作把一个物理设备复制到一台虚拟机。网桥多用作高级设置，特别是主机多个网络接口的情况。如图 3 是其 bridge 的原理图。

如图 3，网桥的基本原理就是创建一个桥接接口 br0，在物理网卡和虚拟网络接口之间传递数据。

2）NAT 方式

NAT 方式原理：NAT 方式是 kvm 安装后的默认方式。它支持主机与虚拟机的互访，同时也支持虚拟机访问互联网，但不支持外界访问虚拟机。如图 4 是其 NAT 的原理图。

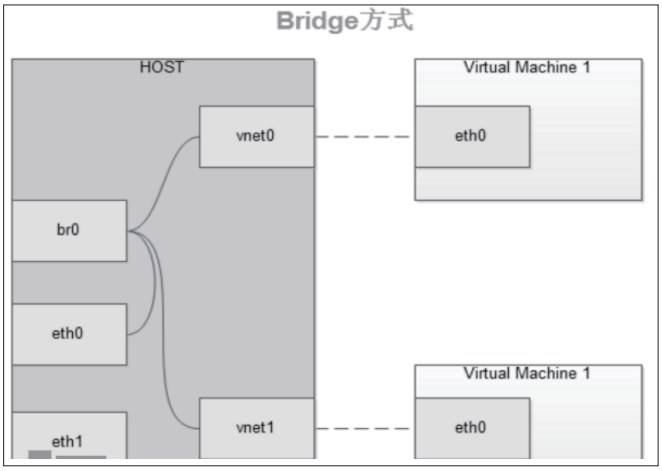


图 3

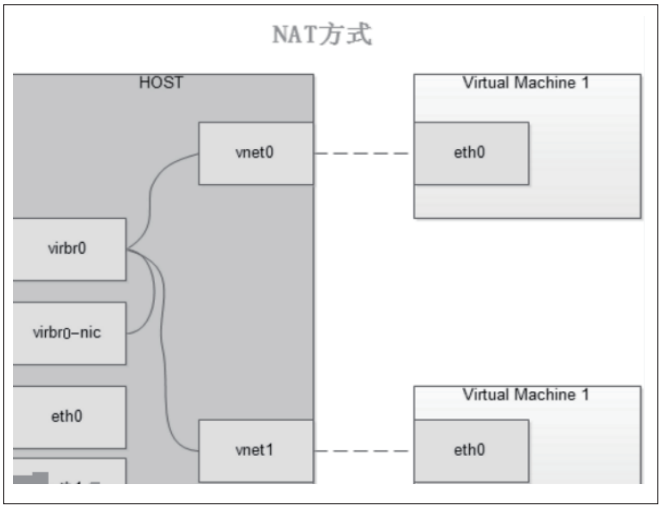


图 4

从图 4 上可以看出，虚拟接口和物理接口之间没有连接关系，所以虚拟机只能在通过虚拟的网络访问外部世界，无法从网络上定位和访问虚拟主机。

3.2.3 部署 KVM 操作系统

KVM 虚拟机硬件配置的设定或更改非常灵活，KVM 通过虚拟机启动命令参数指 定虚拟机所对应的 CPU、内存、硬盘、网卡、声卡、系统时钟等硬件配置。使用启动命令将虚拟磁盘文件和虚拟机关联起来，启动后开始安装操作系统。

kvm – m 1024 – hda /hom e/kvm/hdisk.img– cdrom /dev / cdrom – boot d – localtime

此命令是设置虚拟机使用磁盘镜像文件 /home/kvm / hdisk.img 作为硬盘，设置内存容量为 1024 兆，从光驱启动虚拟机安装操作系统，安装界面出现后和在物理机器上正常安装操作系统一致。安装完毕后将启动命令中的 – boot d 参数修改为 –boot c 即可实现从磁盘镜像正常启动虚拟机。

## 四、KVM 虚拟机三大存储模式

KVM 的存储选项有多种，包括虚拟磁盘文件、基于文件系统的存储和基于设备的存储。为实现 KVM(Kernel–based Virtual Machine) 存 储 管 理， 可 以 使 用 LVM(Logical Volume Manager) 和创建存储池。当系统创建 KVM 虚拟机的时候，

默认使用虚拟磁盘文件作为后端存储。安装后，虚拟机认为在使用真实的磁盘，但实际上看到的是用于模拟硬盘的虚拟磁盘文件。这一额外的文件系统层会降低系统速度。当然，基于磁盘镜像的虚拟磁盘并非全无益处，磁盘文件系统可以很轻松地用于其它的 KVM 虚拟化宿主机。但是如果您希望优化 KVM 虚拟化性能，最好考虑其它的存储方式。

**4.1 基于文件系统的 KVM 存储**

在安装 KVM 宿主机时，可选文件系统为 dir(directory) 或 fs(formatted block storage) 作为初始 KVM 存储格式。默认选项为 dir, 用户指定本地文件系统中的一个目录用于创建磁盘镜像文件。fs 选项可以允许用户指定某个格式化文件系统的名称，把它作为专用的磁盘镜像文件存储。两种 KVM 存储选项之间最主要的区别在于：fs 文件系统不需要挂载到某个特定的分区。两种选项所指定的文件系统，都可以是本地文件系统或位于 SAN 上某个物理宿主机上的网络文件系统。后者具备一定的优势，因为 SAN 可以很轻易地实现多个主机同时访问，而本地磁盘或文件系统则无法实现。还有一种基于文件的磁盘存储方式是 netfs, 用户可以指定一个网络文件系统的名称，如 Samba. 用这种方式作为 KVM 存储很方便，因为这样很容易访问到位于其它服务器上的文件系统，同时用户也可以通过多台宿主机访问磁盘文件。所有的这些基于文件的 KVM 存储方式都有一个缺点：文件系统固有缺陷。因为虚拟机的磁盘文件不能直接读取或写入 KVM 存储设备，而是写入宿主机 OS 之上的文件系统。这也就意味着在访问和写入文件时中间增加了额外一层，这通常会降低性能。所以，如果您希望寻找 KVM 虚拟化性能最优方案，最好考虑基于设备的存储。

**4.2 基于设备的 KVM 存储**

另外一种 KVM 存储的方式就是使用基于设备的方式。共支持四种不同的物理存储：磁盘、iSCSI、SCSI 和逻辑盘。磁盘方式指直接读写硬盘设备。iSCSI 和 SCSI 方式可选，取决于用户采取 SCSI 或 iSCSI 地址把磁盘设备连接。这种 KVM 存储方式的优势在于，磁盘的名称是固定的，而不需要取决于宿主机 OS 搜索到磁盘设备的顺序，这种连接磁盘的方式也有缺点：灵活性不足。虚拟磁盘的大小很难改变，而且基于设备的 KVM 存储不支持快照。如果要优化 KVM 存储的灵活性，可以使用 LVM(Logical Volume Manager)。LVM 的优势在于可以使用快照，而快照并不是 KVM 虚拟化自带的功能。LVM 可以把所有存储放到一个卷组里，从而轻松创建一个逻辑卷。该卷组是物理磁盘设备的一个抽象，所以如果超出可用磁盘空间最大值，还可以向卷组中添加新的设备，从而极大简化了增加存储空间的过程，增加的空间在逻辑卷中直接可以使用。使用 LVM 使得磁盘空间分配更加灵活，而且增加和删除存储也更为容易。最后，LVM 无论是在单宿主机或多宿主机环境中都可以很好工作。在多宿主机环境中，您可以在 SAN 上创建逻辑卷。如果使用 Cluster LVM, 可以很容易的配置成多个主机同时访问某个逻辑卷。

**4.3 基于使用 KVM 存储池**

为简化 KVM 存储管理的目的，可以创建存储池。在宿主机上创建存储池，可以简化 KVM 存储设备的管理。采用存储池的方式还可以实现对提前预留的存储空间的分配。这种策略对于大型应用环境很有效，存储管理员和创建虚拟机的管理经常不是同一个人。这样，在创建首台虚拟机之前先完成 KVM 存储池的创建是很好的方法。当您决定开始 KVM 虚拟化时，先在宿主机端创建一个 KVM 存储池，然后通过这个池提供 LVM 逻辑卷。对于使用 LVM 增加的快照功能，用户是不会感到后悔的，这种 KVM 存储方法提供了极大地灵活性。

**五、KVM 技术在虚拟化中的优势**

KVM 作为一个快速成长的 Linux 虚拟化技术，已经获得了许多厂商的支持，如 Canonical、Novell 等。Canonical 公司的 Ubuntu 服务器版操作系统是第一个提供全功能的 KVM 虚拟化栈的主要 Linux 发行版，之所以很多厂商采用 KVM，主要是 KVM 虚拟化技术具有较强的灵活性，能较好地将不同操作系统和特殊硬件设备加以利用，降低不同系统间维护的复杂度。而且 KVM 支持的 VM 操作系统种类很多，常见的基于 X86 架构的 Windows、Linux、Unix 操作系统绝大部分都可以稳定运行。KVM 本身运行在 Linux 系统内核当中，是 Linux 内核的一部分，属于瘦虚拟化方案，这个轻量级的虚拟化管理程序模块能直接与硬件交互，不需要修改虚拟化操作系统，因此性能更好。同时 KVM 具有优良的稳定性，系统更新便捷，并且补丁包能够和 Linux 内核兼容，轻松控制虚拟化进程，同时减轻管理负担。而且 KVM 本身体积很小，其支持硬件取决于 Linux 系统本身对硬件的支持，目前主流硬件设备均有对应的 Linux 驱动，这也就决定了 KVM 可以在最广泛的硬件系统之上运行，从以上几点足以证明 KVM 技术其优势是显而易见的。

**六、结语**

KVM 是一个发展时间比较短，但是性能和稳定性都表现优秀的虚拟化解决方案，由于操作系统直接和整合到 Linux 内核中的 KVM 虚拟化管理程序交互，所以在任何场景下都可以直接和硬件进行交互，而不需要修改虚拟化的操作系统，另一方面其灵活的网络拓扑结构、简便的硬件配置方案、集中统一管理可满足于大多数数据中心虚拟化实践的要求。当然 KVM 也有很多不足之处，比如对一些虚拟化扩展特性，如泛虚拟化支持、虚拟机动态迁移、图形化管理界面等新功能不是很好的支持，但是随着时间的推移，Ret Hat（目前掌握 KVM 技术），作为 Linux 企业市场中份额最大的企业，将会使 KVM 虚拟化技术的功能更加齐全，我相信未来虚拟化市场必定是 KVM 的。

**参 考 文 献**

[1] 广小明，胡杰，陈龙，等 . 虚拟化技术原理与实现 [M]. 北京：电子工业出版社，2012.
[2] 高清华 . 基于 Intel VT 技术的虚拟化系统性能测试研究 [D], 浙江大学 , 2008: 37–39.
[3] 崔泽永 , 赵会群 . 基于 KVM 的虚拟化研究及应用 [J], 计算机技术与发展 , 2011, 21(6) :109.
[4] 邓秀春 , 王超云 . 基于虚拟化技术的数据中心构建 [R], 科技创新导报 , 29, 2010.
[5] 李勇 ; 郭玉东 ; 王晓睿 ; 时光 ; 基于 EPT 的内存虚拟化研究与实现 [J]; 计算机工程与设计 ;2010 年 18 期