Linux下桥接模式详解一

邮箱：zhunxun@gmail.com

< 2020年5月 >

日	一	二	三	四	五	六
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

搜索

找找看

谷歌搜索

PostCategories

C语言(2)
IO Virtualization(3)
KVM虚拟化技术(26)
linux 内核源码分析(61)
Linux日常应用(3)
linux时间子系统(3)
qemu(10)
seLinux(1)
windows内核(5)
调试技巧(2)
内存管理(8)
日常技能(3)
容器技术(2)
生活杂谈(1)
网络(5)
文件系统(4)
硬件(4)

PostArchives

2018/4(1)
2018/2(1)
2018/1(3)
2017/12(2)
2017/11(4)
2017/9(3)
2017/8(1)
2017/7(8)
2017/6(6)
2017/5(9)
2017/4(15)
2017/3(5)
2017/2(1)
2016/12(1)
2016/11(11)
2016/10(8)
2016/9(13)

ArticleCategories

时态分析(1)

Recent Comments

1. Re:virtio前端驱动详解
我看了下，Linux-4.18.2中的vp_notify()函数。bool vp_notify(struct virtqueue *vq){ /* we write the queue's selec...
--Linux-inside
2. Re:virtIO之VHOST工作原理简析

注册博客园已经好长时间，一直以来也没有在上面写过文章，都是随意的记录在了未知笔记上，今天开始本着分享和学习的精神想把之前总结的笔记逐步分享到博客园，和大家一起学习，一起进步吧！

2016-09-20 17:11:05

其实之前已经有分析过网桥的原理，但是当时对其理解还是局限于表面，对于其本身的实现原理并没有结合linux源代码进行分析，那么本次实际上是要分析qemu对于网卡的模拟，那么 从源头来说，首先分析下桥接模式下数据是如何转发的。

既然说到了桥接就不得不提到一个数据链路层设备------网桥。在计算机网络中，网桥作为一个网络设备应用也许并没有那么广泛。但是作为它的扩展---交换机就显得火的多了。交换机其本质就是相当于一个多端口的网桥，大多数交换机是工作在数据链路层即L2层，但目前三层交换机也逐步浮出水面。在本片文章中我们说讨论的仅仅是普通的二层交换机。
交换机既然工作在链路层，那么其必定要比工作在物理层的集线器高明一点。即其转发是基于帧的转发，交换机可以提取到每个帧 的源MAC和目的MAC，并且其本身会维护一个转发表，表内包含有MAC地址和其相应端口的映射。交换机收到一个帧之后的处理流程如下：
1、每当交换机收到一个帧，就提取出帧的源MAC，然后查找转发表，表中若没有这一项，就把该地址和进入的端口写入转发表，若存在，就更新!
2、然后提取目的MAC，查找转发表，表中若有对应的项，就从对应的端口转发出去，若没有，就从除了接受该帧的所有端口转发出去。
上面是对网桥（交换机基本原理的介绍），下面就谈谈LInux内核中对这一特性的支持。
首先在桥接模式下，网卡需要设置成混杂模式。即接收所有到达的数据包，不管目的地址是否是网卡的MAC。
在软件方面，就要依赖与Linux内核中的Tun/Tap驱动了。

Tun/Tap驱动详解

Tun/Tap 用于虚拟网络设备，Tun虚拟网络层的设备，而Tap虚拟数据链路层的设备。所以Tun/Tap驱动提供一种功能,即可以通过这种驱动创建接口（Tun/Tap），通过这些接口可以在内核和用户空间传送网络数据包，其中Tun类型的接口可以传送IP数据包即从链路层交付上来的包，不包含以太网头。Tap接口可以传送以太网数据帧，就是链路层的包。
Tun/Tap驱动包含两个部分，一个是网卡驱动，一个是字符设备驱动。网卡驱动接收来自TCP/IP 协议栈的网络分包并发送或者反过来将接收到的网络分包传给协议栈处理；而字符驱动部分则将网络分包在内核与用户态之间传送，模拟物理链路的接收和发送。Tun/Tap 驱动的字符设备驱动部分向用户态暴漏了一个接口/dev/net/tun，其实是一个字符设备文件，用户空间的应用程序可以通过这个设备文件来和内核中的驱动程序进行交互，其操作方式和普通的文件操作无异。

再问一个问题，从设置ioeventfd那个流程来看的话是guest发起一个IO，首先会陷入到kvm中，然后由kvm向qemu发送一个IO到来的event，最后IO才被处理，是这样的吗？

--Linux-inside

3. Re:virtIO之VHOST工作原理简析
你好。设置ioeventfd这个部分和guest里面的virtio前端驱动有关系吗？设置ioeventfd和virtio前端驱动是如何发生联系起来的？谢谢。

--Linux-inside

4. Re:QEMU IO事件处理框架
良心博主，怎么停跟了，太可惜了。

--黄铁牛

5. Re:linux 逆向映射机制浅析
小哥哥520脱单了么

--黄铁牛

Top Posts

1. 详解操作系统中断(21152)
2. PCI 设备详解一(15806)
3. 进程的挂起、阻塞和睡眠(13713)
4. Linux下桥接模式详解一(13465)
5. virtio后端驱动详解(10538)

推荐排行榜

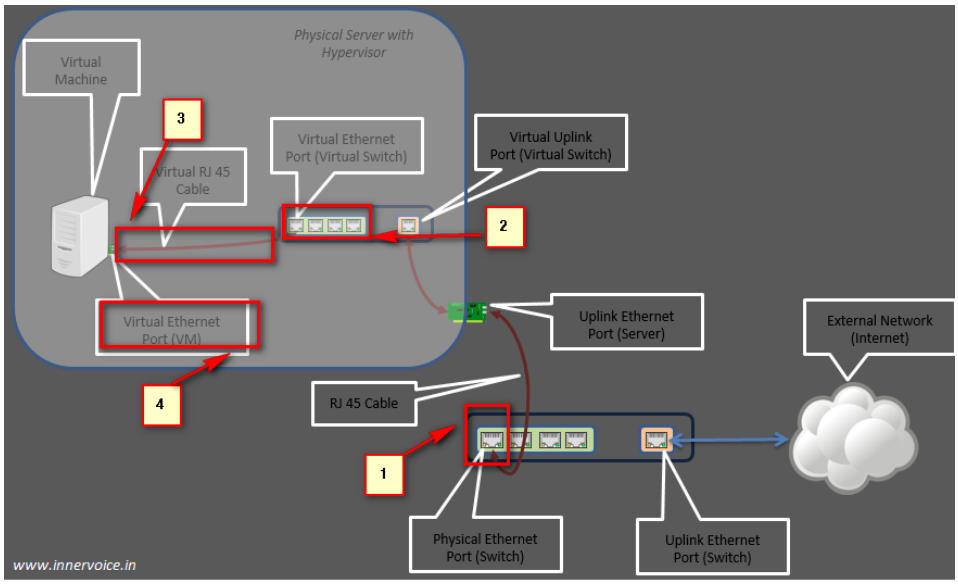
1. 进程的挂起、阻塞和睡眠(6)
2. 为何要写博客(2)
3. virtIO前后端notify机制详解(2)
4. 详解操作系统中断(2)
5. qemu-kvm内存虚拟化1(2)

```
[root@chendomain code]# ll /dev/net/
total 0
crw-rw-rw-. 1 root root 10, 200 Sep  8 13:19 tun
[root@chendomain code]#
```

回想下连接一台交换机的PC机和外网交互的流程。一下几个条件是必须的：
1、交换机上行端口--------用于连接外网
2、交换机下行端口--------用于连接交换机下面的各个PC机
3、网线
4、PC机（物理网卡）

外部数据通过交换机上行端口进入交换机，交换机工作在数据链路层，对数据帧提取MAC地址，首先进行地址学习后，就进行数据帧的转发，这里根据地址表中是否包含做出不同的选择，这不是重点，不在赘述。
然后数据帧通过网线到达PC机的物理网卡，然后系统就根据协议栈对数据包进行逐层处理最后交付了。

那么上述情况迁移到虚拟化环境下，即在一台LInux host上运行的虚拟机，如何通过LInux bridge上网?
其必要条件仍然有效，只不过需要采用另一种方式而已，见下图



通过查看上图，我们不难发现，虽然迁移到了虚拟化的环境下，前面提到的四个条件依然存在
1、上行网络接口--------这里表现为host的物理网卡，主要用以和外网交互
2、Tap接口---------这里就是虚拟出的网络设备，称之为Tap device,这里就是作为bridge的一个端口，连接虚拟机的虚拟网卡。
3、虚拟网线---------这里主要表现为一个设备文件/dev/net/tun ,用户程序和Tap接口交互就是通过这个设备文件，从功能上来讲就是类似于一个网线。
4、虚拟网卡----------这里就是从客户端的角度，一台虚拟机要发送和接收网络数据包必定需要虚拟网卡
LInux内部实现的bridge可以把一台机器上的多张网卡桥接起来，从而把自己作为一台交换机。同时，LInux bridge还支持虚拟端口，即桥接的不一定都是物理网卡接口，还可以是虚拟接口。目前主要表现为Tap接口，Tap接口在逻辑上和物理网卡实现相同的功能，都可以接收和发送数据包。所以这一应用也成就了虚拟化环境下的bridge实现。
到此基础理论知识就介绍完了，那么接下来就结合Linux源代码分析下桥接模式下数据包的转发流程。

Linux下桥接模式详解2

分类：KVM虚拟化技术, linux 内核源码分析, qemu

jack.chen
关注 - 12
粉丝 - 44
+加关注

1 0

» 下一篇： windows中根据进程PID查找进程对象过程深入分析

posted @ 2016-09-20 17:12 jack.chen Views(13465) Comments(2) Edit 收藏

Post Comment

#1楼 2017-11-30 11:18 | 最后一个亮亮

感谢大牛！
你的讲解非常清晰，最近一直卡在openstack网络这块，主要就是虚拟网络这块比较懵，看了之后收益匪浅。

支持(0) 反对(

#2楼 2018-08-18 22:43 | 被罚站的树

mark

支持(0) 反对(

刷新评论 刷新页面 返回顶部

注册用户登录后才能发表评论，请 登录 或 注册， 访问 网站首页。

相关博文：
· 什么是交换机初级网络工程师必看
· HUB和Switch
· 2017.8.11 交换机
· 中继器、集线器、交换机、网桥和路由器分别相应于哪一层?
· 2017.3.16下午
» 更多推荐...

最新 IT 新闻:
· Uber一季度营收35.4亿美元 亏损29亿美元
· 跟谁学回应香橼第三份做空报告：谴责用不实指控做空行为
· 视频会议需求激增，为什么苹果FaceTime却没火起来?
· 百度"度小镜"京东开卖：接入百度网盘 视频永不丢失

· 支付宝发布全国首份《网络互助白皮书》：8成用户年收入低于10万
» 更多新闻...

以马内利