Xen 网络解析

2014/5/13

renyl

1 Guest 网桥通信方式

1.1 能够与外部网络通讯

Guest 要和网桥通信，必须借助 Tap 设备与网桥建立连接，如图 1-1 所示。

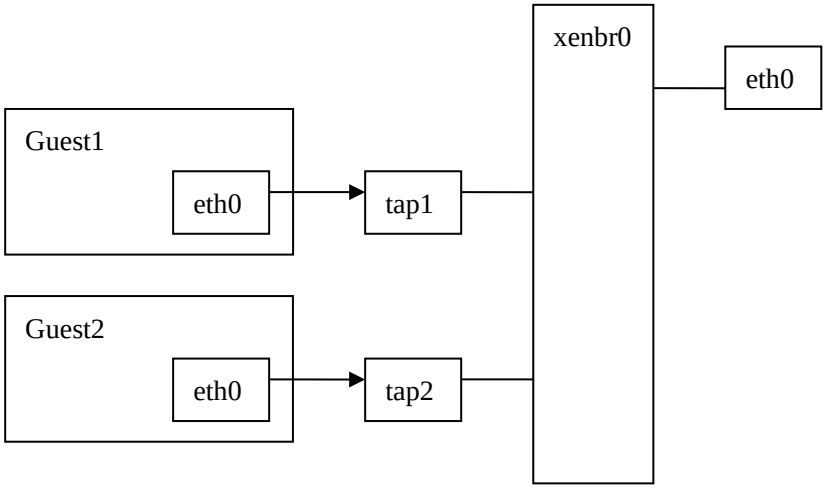


图 1-1 Guest 与网桥通信方式

分析：

1） Tap 设备是 Linux kernel 为用户空间提供的一种虚拟网络设备。

2） Tap 设备可以理解为网桥上的一个接口，Guest 的虚拟网卡 eth0 与这些接口相连，进而连接到网桥。

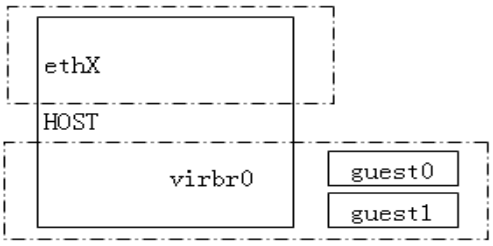
3） 同时，网桥与 Host 的物理网卡 eth0 相连，保证了 Guest 与外部网络的正常通信。

注：使用 Tap 设备的原因：

Guest 的虚拟网卡对于 Host 来说是不可见的。也就是说 Host 上看不见 Guest 里面的 eth0。所以 Host 的网桥必须借助一种机制，来区分连接在网桥上的各个 Guest，这就是 Tap 设备。

1.2 不能够与外部网络通讯

使用 virbr0 作为 guest 与 host 通信的网桥，网络拓扑如下图所示：



2 Xen 网络连接原理

Xen 将 Host 也看作一个 Guest，称为 Dom0。真正的 Guest 称为 DomU，如 Dom1，Dom2，Dom3 等。因此 Host（Dom0）的网络连接从逻辑结构上说，和其他的 Guest 是一样的，如图 2-1 所示。

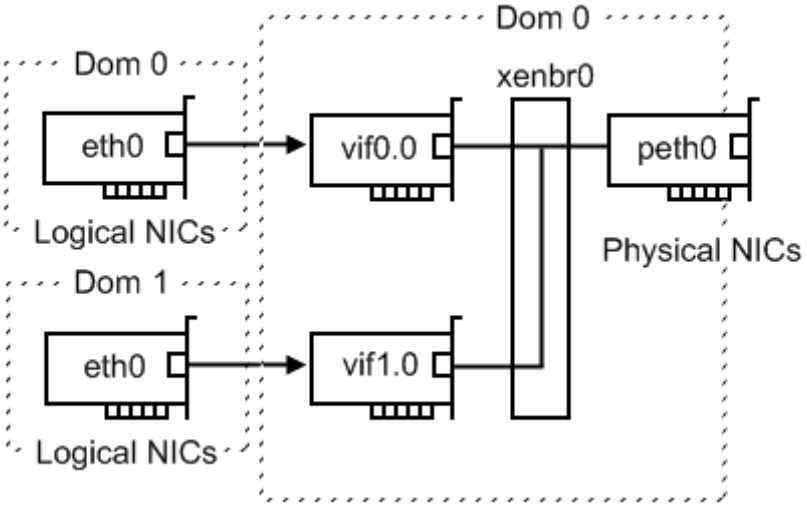


图 2-1 Xen 网络连接原理

注：

vif0.0 和 vif1.0 为 tap 设备的命名。vifx.x 与 guest 的网卡对应规则：vif 后的第一个数字为 guest 的 Dom 编号，第二个数字为 guest 中的网卡编号，如 vif1.0 对应 Dom1 上的 eth0。

分析：

由上图可知，Host 作为 Dom0 和其他的 Guest 一样，需要使用虚拟网卡 eth0，经过 tap 设备 vif0.0 与网桥通信。Xen 系统启动时，会创建以下两个虚拟设备，Dom0 虚拟网卡的实现如下：

1）创建 veth0，用于模拟 Dom0 中的 eth0

2）创建 Tap 设备 vif0.0，它是 Dom0 的 eth0 接入网桥所必须的 Tap 设备

Xen 系统启动时通过/etc/xen/scripts/network-bridge 脚本自动进行网络配置的流程如下：

1） 创建网桥 xenbr0；

2） 关闭 host 的物理网卡 eth0，将 eth0 的 IP 和 MAC 地址复制到 veth0 中（执行命令：ip addr add IPADDR dev veth0; ip link set veth0 addr MACADDR）；

3）将 host 的物理网卡 eth0 重命名为 peth0（执行命令：ip link set eth0 name peth0），并且修改 peth0 的 mac 地址为 FE:FF:FF:FF:FF:FF

4)将创建的虚拟网卡 veth0 重命名为 eth0(Dom0 的虚拟网卡)(执行命令：ip link set veth0 name eth0)；

5）将 peth0 和 vif0.0 加入 xenbr0；

6）启动所有网络接口：xenbr0，eth0，vif0.0，peth0。

因此，Xen 默认的网络配置如图 2-2 所示。

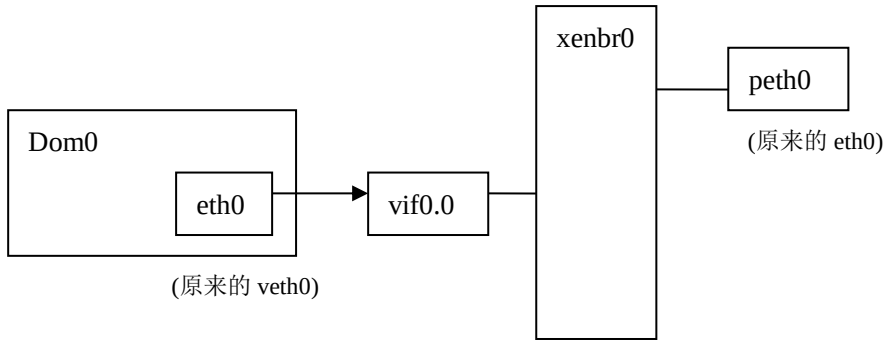


图 2-2 Xen 默认网络配置

这里，eth0 通过 vif0.0 间接地接入 xenbr0，Xen kernel 会将发送到 xenbr0 的数据包自动转发给 eth0，同样，从 eth0 发出的数据包，将会通过 vif0.0 转发出去。

可以使用以下命令检查网络配置：

```
# brctl show
bridge name     bridge id          STP enabled     interfaces
xenbr0          8000.feffffffffff  no              peth0
                                                   vif0.0     #表示eth0通过vif0.0接入到xenbr0
```

3 环路问题

由于 Xen kernel 将 Host 看作特殊的 Guest：Dom0，因此导致 Dom0 中的网卡 eth0 在 Host 上可见。如下所示：

```
# ifconfig
eth0        Link encap:Ethernet   HWaddr 00:19:99:0E:DE:39
            inet addr:133.162.10.15   Bcast:133.162.10.255   Mask:255.255.255.0
……
peth0       Link encap:Ethernet   HWaddr FE:FF:FF:FF:FF:FF
            inet6 addr: fe80::fcff:ffff:feff:ffff/64 Scope:Link
……
vif0.0      Link encap:Ethernet   HWaddr FE:FF:FF:FF:FF:FF
            inet6 addr: fe80::fcff:ffff:feff:ffff/64 Scope:Link
……
```

但必须注意，此时的 eth0 是 Dom0 的虚拟网卡，peth0 才是 Host 的物理网卡。这是两个不同的设备。

- **误操作：将 eth0 接入网桥 xenbr0**

  导致本次网络故障的原因是，使用 brctl 命令将 eth0 接入了 xenbr0。

```
# brctl addif xenbr0 eth0
# brctl show
bridge name    bridge id          STP enabled    interfaces
xenbr0         8000.feffffffffff  no             peth0
                                                 vif0.0     #表示 eth0 间接接入 xenbr0
                                                 eth0       #表示 eth0 直接接入 xenbr0
```

1） 由于 brctl 命令以及 Linux 平台上的虚拟网桥是独立于 Xen kernel 实现的
2） 而且 eth0 并不是直接接入 xenbr0 的
3） 所以执行这一步操作时 brctl 命令无法检测出 eth0 已经与 xenbr0 连通
4） 进而可以第二次将 eth0 接入 xenbr0，形成了环路。

将 eth0 接入 xenbr0 以后，网络连接如图 3-1 中的红色线路所示：

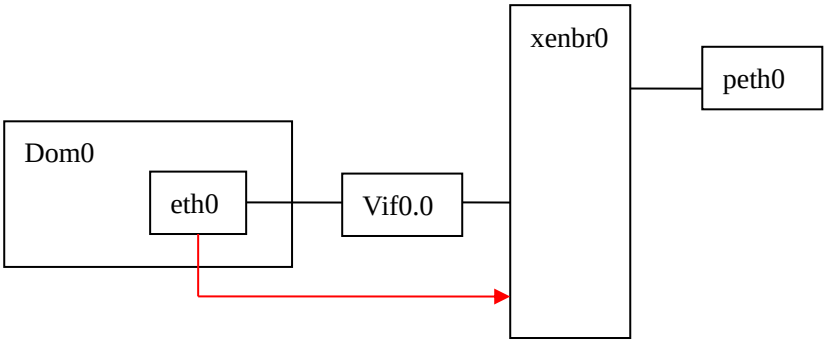


图 3-1 误操作后的网络配置

当 Dom0 向外部网络发送数据包时，会形成图 3-2 所示的环路。

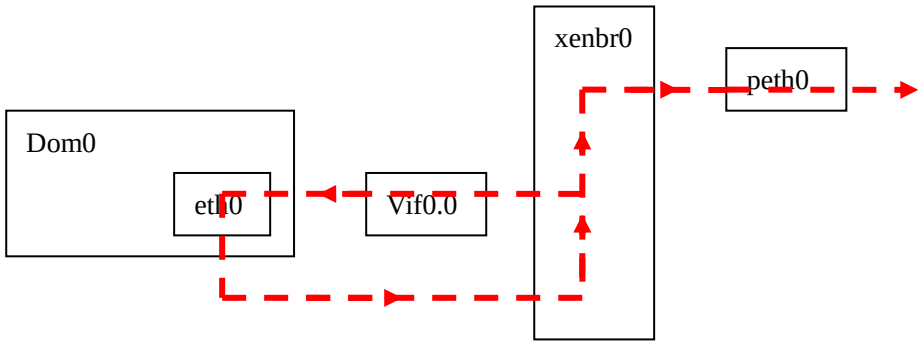


图 3-2 网络环路

- **网络阻塞的根本原因**

网络内部形成环路，造成数据流量过大，网络阻塞。