从 Bridge 到 OVS，探索虚拟交换机

> 本文首发于我的公众号 **CloudDeveloper(ID: cloud_dev)**，专注于云计算，但不止于云计算，努力打造干货平台，欢迎大
> 末可以扫。

CONTENTS

Linux Bridge

和物理网络一样，虚拟网络要通信，必须借助一些交换设备来转发数据。因此，对于网络虚拟化来说，交换设备的虚拟化是很关键的一环。

上文「网络虚拟化」已经大致介绍了 Linux 内核为了满足网络虚拟化的要求，实现了一套虚拟交换设备——Bridge。本文重点介绍下 Bridge 的加强版——Open vSwitch（OVS），并从 Bridge 过渡到 OVS 的缘由讲起，让大家有个全面的认识。

借助 Linux Bridge 功能，同主机或跨主机的虚拟机之间能够轻松实现通信，也能够让虚拟机访问到外网，这就是我们所熟知的桥接模式，一般在装 VMware 虚拟机或者 VirtualBox 虚拟机的时候，都会提示我们要选择哪种模式，常用的两种模式是桥接和 NAT。

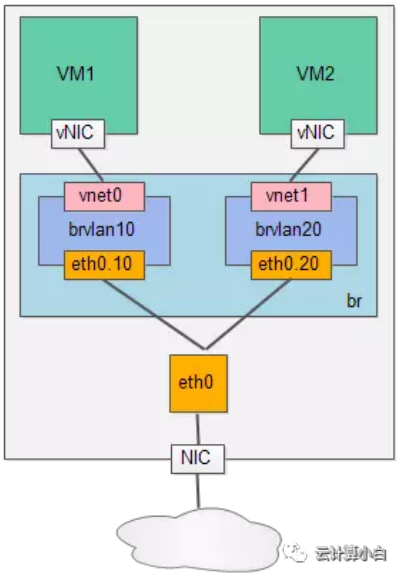
NAT 也很好理解，可以简单理解为当虚拟机启用了 NAT 模式之后，宿主机便通过 DHCP 为其生成可以访问外网的 IP，当 VM 访问外网的时候，就可以用该 IP 访问，其实就是宿主机为其做了地址转换。更详细的内容请自行搜索了解。

物理交换机有个重要的功能，就是虚拟局域网（VLAN），是对局域网（LAN）的软件化升级。一般，两台计算机通过一台交换机连接在一起就构成了一个 LAN。

一个 LAN 表示一个广播域，这意味着这个 LAN 中的任何节点发的数据包，其他节点都能收到，这有两个问题，一个是容易形成广播风暴，造成网络拥塞，另一个是广播包无法隔离，比如节点 B 不想接收节点 A 的包，但节点 A 强行要发，这就有点说不过去了。

解决这个问题的方案就是 VLAN，VLAN 能够对广播包进行有效隔离，它的做法是从软件上将交换机的端口虚拟出多个子端口，用 tag 来标记，相当于将交换机的端口划分多个 LAN，同一个 LAN 中的节点发出的数据包打上本 LAN 的 tag，这样，其他 LAN 中的节点就无法收到包，达到隔离的目的。

Bridge 本身是支持 VLAN 功能的，如下图所示，通过配置，Bridge 可以将一个物理网卡设备 eth0 划分成两个子设备 eth0.10，eth0.20，分别挂到 Bridge 虚拟出的两个 VLAN 上，VLAN id 分别为 VLAN 10 和 VLAN 20。同样，两个 VM 的虚拟网卡设备 vnet0 和 vnet 1 也分别挂到相应的 VLAN 上。这样配好的最终效果就是 VM1 不能和 VM2 通信了，达到了隔离。



Linux Bridge + VLAN 便可以构成一个和物理交换机具备相同功能的虚拟交换机了。对于网络虚拟化来说，Bridge 已经能够很好地充当交换设备的角色了。

OVS

但是为什么还有很多厂商都在做自己的虚拟交换机，比如比较流行的有 VMware virtual switch、Cisco Nexus 1000V，以及 Open vSwitch。究其原因，主要有以下几点（我们重点关注 OVS）：

1）方便网络管理与监控。OVS 的引入，可以方便管理员对整套云环境中的网络状态和数据流量进行监控，比如可以分析网络中流淌的
是来自哪个 VM、哪个 OS 及哪个用户，这些都可以借助 OVS 提供的工具来达到。

2）加速数据包的寻路与转发。相比 Bridge 单纯的基于 MAC 地址学习的转发规则，OVS 引入流缓存的机制，可以加快数据包的转发效

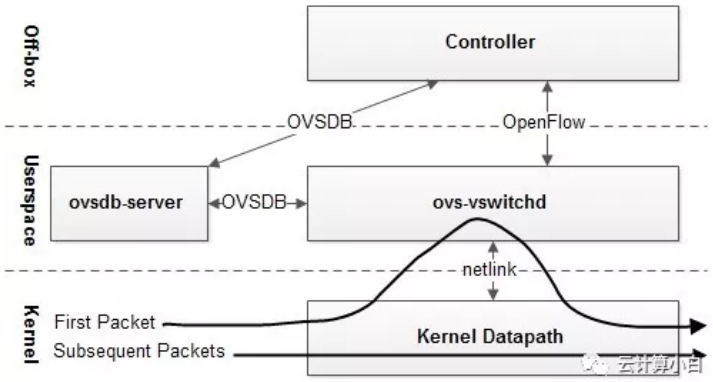
4）隧道协议支持。Bridge 只支持 VxLAN，OVS 支持 gre/vxlan/IPsec 等。

5）适用于 Xen、KVM、VirtualBox、VMware 等多种 Hypervisors。

.....

除此之外，OVS 还有很多高级特性，详情可以查阅官网自行了解。

下面简单看下 OVS 的整体架构，如下图所示，OVS 在 Linux 用户态和内核态都实现了相应的模块，用户态主要组件有数据库 server 和守护进程 ovs-vswitchd。内核态中实现了 datapath 模块。



其中， ovs-vswitchd 和 datapath 共同构成了 OVS 的数据面，控制面由 controller 模块来完成，controller 一般表示的是 OpenFlow 控制器，在 OVS 中，它可以借由第三方来完成，只要支持 OpenFlow 协议即可。

这里额外提一点，很多的一些产品级的虚拟交换机都是自身集成了控制器，比如 Cisco 1000V 的 Virtual Supervisor Manager(VSM)，VMware 的分布式交换机中的 vCenter，而 OVS 是把这个事交由第三方去做，这么做的意义还是比较大的，可以让自己的产品很好地融入到各种解决方案中。

OpenFlow

OpenFlow 是控制面和数据面通信的一套协议，我们常常把支持 OpenFlow 协议的交换机称为 OpenFlow 交换机，控制器称为 OpenFlow 控制器，业界比较知名的 OpenFlow 控制器有 OpenDaylight、ONOS 等。

OpenFlow 是一个独立的完整的流表协议，不依赖于 OVS，OVS 只是支持 OpenFlow 协议，有了支持，就可以使用 OpenFlow 控制器来管理 OVS 中的流表。OpenFlow 不仅仅支持虚拟交换机，某些硬件交换机也支持 OpenFlow 协议。

ovs-vswitchd

ovs-vswitchd 是 OVS 的核心组件，它和内核模块 datapath 共同构成了 OVS 的数据面。它使用 OpenFlow 协议与 OpenFlow 控制器通信，使用 OVSDB 协议与 ovsdb-server 通信，使用 netlink 和 datapath 内核模块通信。

ovsdb-server

ovsdb-server 是 OVS 轻量级的数据库服务，用于整个 OVS 的配置信息，包括接口、交换内容、VLAN 等，ovs-vswitchd 根据这些配置信息工作。

OpenFlow 控制器

OpenFlow 控制器可以通过 OpenFlow 协议连接到任何支持 OpenFlow 的交换机，比如 OVS 。控制器通过向交换机下发流表规则来控制数据流向。

Kernel Datapath

datapath 内核模块和 ovs-vswitchd 是相互协作工作的，datapath 负责具体的收发包，而 ovs-vswitchd 通过 controller 下发的流表规则指导 datapath 如何转发包。

举个例子，datapath 从主机物理网卡 NIC 或者 VM 的 虚拟网卡 vNIC 收到包，如果是第一次收到包，datapath 不知道怎么处理这个包，于是将其丢给 ovs-vswitchd ， ovs-vswitchd 决定该如何处理这个包之后又丢给 datapath，datapath 根据 ovs-vswitchd 的指示执行相应的动作，是丢弃还是从哪个口传出去。同时，ovs-vswitchd 会让 datapath 缓存好这个包的动作，下次再来就可以直接执行动作。

如果不是第一次收到包，就是按照之前缓存好的动作执行，这样极大地提高了数据处理的速度。

本文先对 OVS 有个初步印象，下文再详细介绍 OVS 的其他组件。

CONTENTS

分类：云计算，虚拟化
标签：虚拟化，云计算，OVS

« 上一篇： 网络虚拟化
» 下一篇： OVS 总体架构、源码结构及数据流程全面解析

posted @ 2017-12-24 10:20 CloudDeveloper 阅读(16256) 评论(0) 编辑 收藏