VXLAN

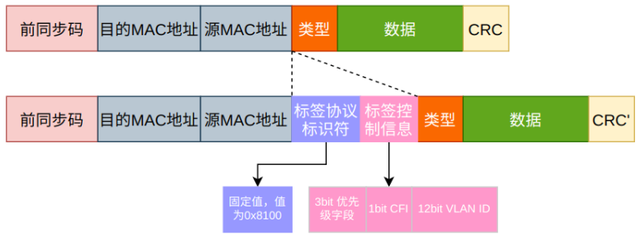
VLAN

VLAN建立在支持VLAN的交换机之上，换而言之，建立在链路层之上，通过将交换机的接口(interface)与特定VLAN ID绑定，进而实现虚拟局域网

交换机和交换机之间特殊的连接称之为trunk，或者是干线连接

VLAN标签直接添加在MAC数据帧之中，其中最为关键的为12bit的VLAN ID，表示当前数据包应该属于哪个虚拟局域网

12bit除去系统专用，只剩下4096个，这对当前的虚拟云计算来说根本不够用



VXLAN

VLAN需要特殊的交换机支持，并且所能够承载的虚拟局域网数量有限，不能够满足云计算的需求，所以需要另一种不依赖底层物理设备，并且能够承载足够数量的虚拟局域网的虚拟网络实现

VXLAN全称为Virtual eXtensible Local Area Network，虚拟扩展局域网，建立在三层网络(网络层，或者说IP层)之上，不需要物理设备的支持。只要主机间三层可达，就可以在现有网络上建立虚拟网络

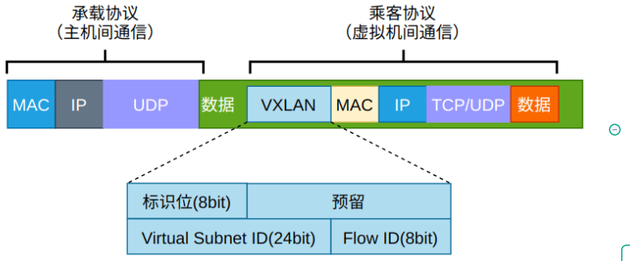
基本原理

主机和主机之间可以使用公有或者私有IP进行通信，那么运行在主机之上的虚拟机该如何与另外一台运行在其它主机上的虚拟机通信呢?

假设每台主机上都运行了一个进程，称之为T。进程T将虚拟机A的报文进行一个封装，封装成主机A发给主机B的报文，然后发送给主机B。报文到达主机B之后同样的进程进行解包，然后发给虚拟机

虚拟机的网络包乘着主机间通信的网络包到达目的地

就像俄罗斯套娃或者是特洛伊木马一样



VXLAN实际报文结构

在主机间通信数据包中，原来承载用户数据的部分现在用来承载另一个数据包

可以看到，这里的VXLAN ID为24位，相较于12位来说，足够使用了

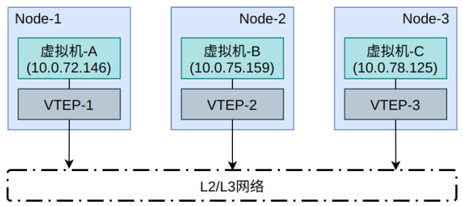
VTEP

既然要"套娃"，那就得有个人来做这个事情，做这个事情的进程称之为VTEP(VXLAN Tunnel Endpoint)，负责对VXLAN协议包的封包和解包，以及管理虚拟机网络

在每一个节点上都可以有一个VETP运行，每当节点中有新的虚拟机启动时，都得向VTEP进行注册，VTEP知道节点中有多少个虚拟机运行，以及其网络运行情况

当一个 VTEP 启动的时候，它们都需要通过 IGMP 协议添加进一个广播组中，可以简单的将广播组理解成用户组

通信过程



虚拟机A想要给虚拟机B(10.0.75.149)发送信息，只知道IP地址，不知道MAC地址，无法通信

VTEP-1封装ARP包，在加入的广播组中进行广播: 哪位兄弟手下有10.0.75.149这台虚拟机啊? 把MAC地址给我可好?

VTEP-2和VTEP-3收到广播请求之后，再在本地发起ARP广播，虚拟机B回应，此时VTEP-2将记录虚拟机B的IP地址与MAC地址的映射，并返回结果给VTEP-1

得到MAC地址之后，虚拟机A发给虚拟机B的MAC将填充完整，并封装在节点通信的IP包中，从网络中发出去

Node-2收到数据包后，根据UDP包中的端口发送至VTEP进程，进程进一步解包，获得虚拟机的MAC地址与IP地址，将其发送至指定的虚拟机，通信结束

为什么VXLAN使用UDP协议作为承载传输层协议

TCP为点对点协议，而UDP支持广播，想要在VXLAN中发送诸如ARP这种广播，必须使用UDP