Mesos容器网络解决方案

数人云 肖德时

Mesos 1.0 全面支持 Container Network Interface (CNI)

Mesos的网络问题：
1、一容器一IP
2、DNS方式的服务发现
3、网络隔离

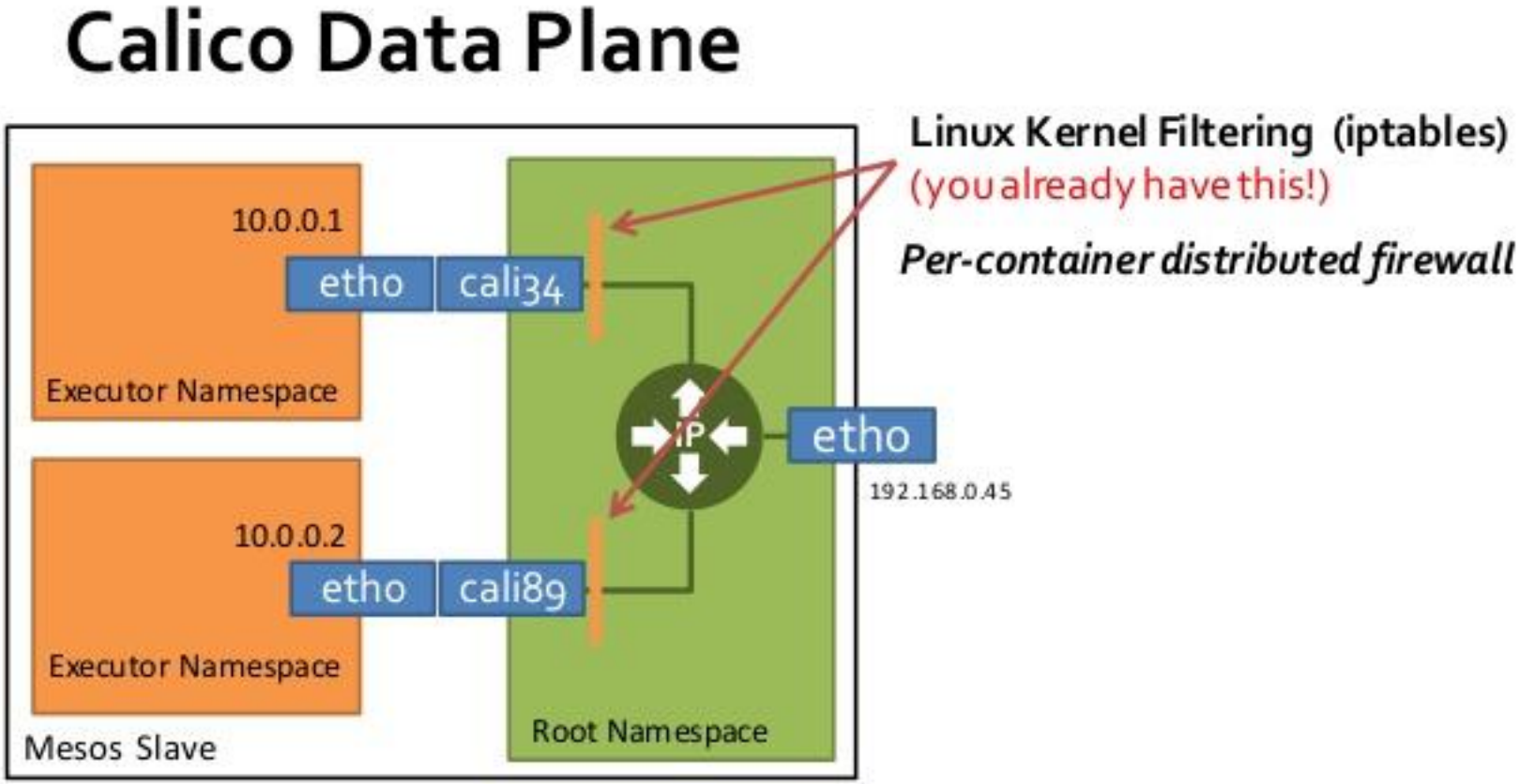
Calico 是当前唯一一套可行的网络解决方案

今天我们只讲 **一容器一IP** 的实现

一容器一IP，我们为啥只讲Calico

Calico的网络实现优点：

- L3网络虚拟化以及隔离
- 简单、可扩展，并且开源



CNI这个方案需要介绍下

- 由 CoreOS 引入的CNI是新一代的容器网络插件模型

- 使用单一JSON配置文件描述网络
- Core plugins，IPAM内置2插件host-local和dhcp
- 第三方实现
  - Project Calico - a layer 3 virtual network
  - Weave - a multi-host Docker network
  - Contiv Networking - policy networking for various use cases
  - SR-IOV

```
{
    "name": "mynet",
    "type": "bridge",
    "bridge": "mynet0",
    "isDefaultGateway": true,
    "forceAddress": false,
    "ipMasq": true,
    "hairpinMode": true,
    "ipam": {
        "type": "host-local",
        "subnet": "10.10.0.0/16"
    }
}
```

Mesos Cluster需要的是数据中心的网络结构

- **North-South Traffic** = Traffic in and out of the DC
  East-West Traffic = Traffic between servers (Containers) within the DC

- Calico，基于BGP协议的路由方案，支持很细致的ACL控制，对混合云亲和度比较高
- Macvlan，从逻辑和Kernel层来看隔离性和性能最优的方案，基于二层隔离，所以需要二层路由器支持，和主机网络强绑定，内部可以使用，上云迁移不可能。

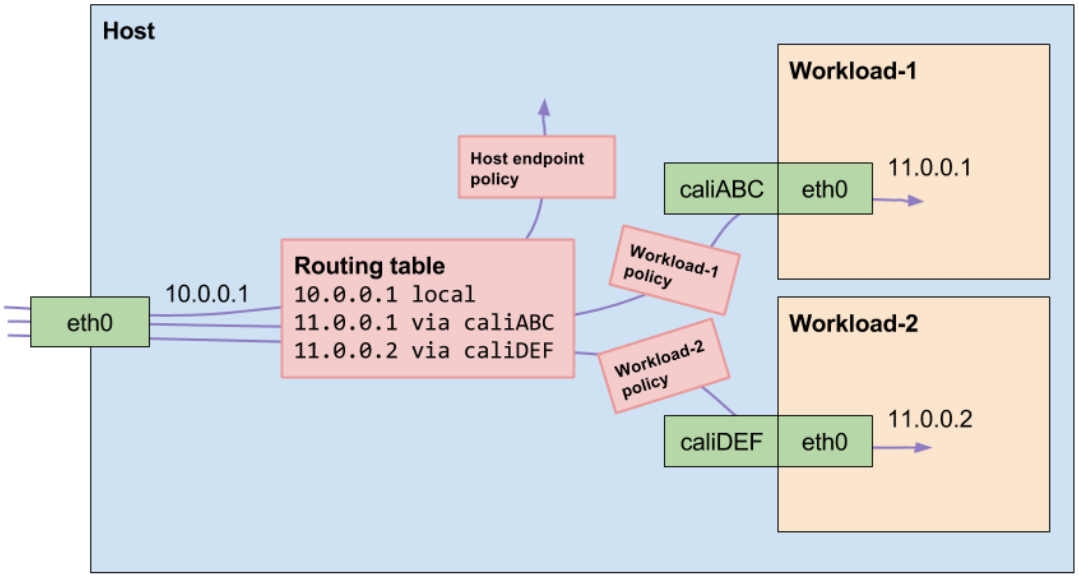
- 体会：Calico灵活性有保障

网络很复杂，基本功：什么是3层网络

为了节省时间，给一个简单的定义：

简单的说三层交换技术 = 二层交换技术 + 三层路由功能，我们也可以理解成三层交换机 = 二层交换机 + 传统的路由器

网络很复杂，基本功：网络数据包



传统Overlay网络

Outer MAC	Outer IP	Outer UDP	VXLAN	VM			
				MAC	IP	TCP/UDP	Data

Host	VM		
MAC	IP	TCP/UDP	Data

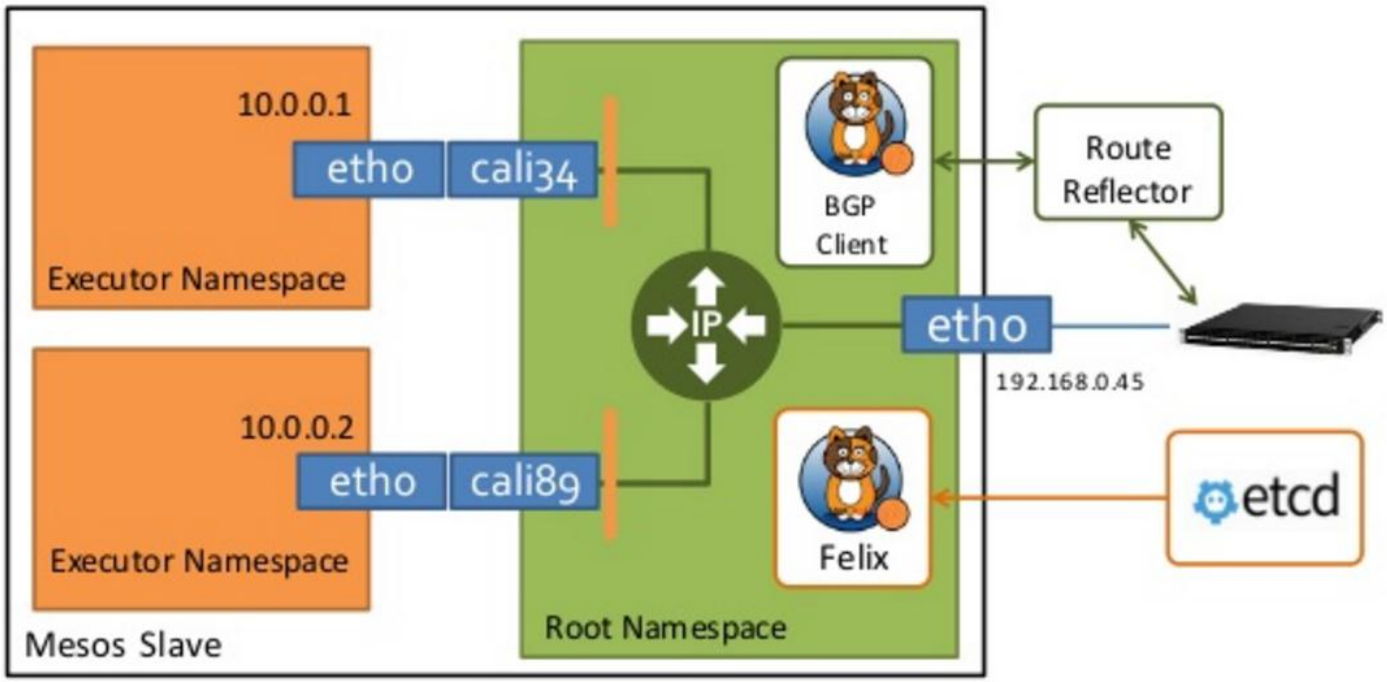
网络很复杂，基本功：性能对比

Client	Server	Throughput, tps	Ratio to "direct-direct"
Direct	Direct	282780	1.0
Direct	Host	280622	0.99
Direct	Bridge	250104	0.88
Bridge	Bridge	235052	0.83
overlay	overlay	120503	0.43
Calico overlay	**Calico overlay**	**246202**	**0.87**
Weave overlay	Weave overlay	11554	0.044

Docker Swarm的原生网络推荐在原型和测试环境中使用。生产级别网络推荐Calico或Host模式

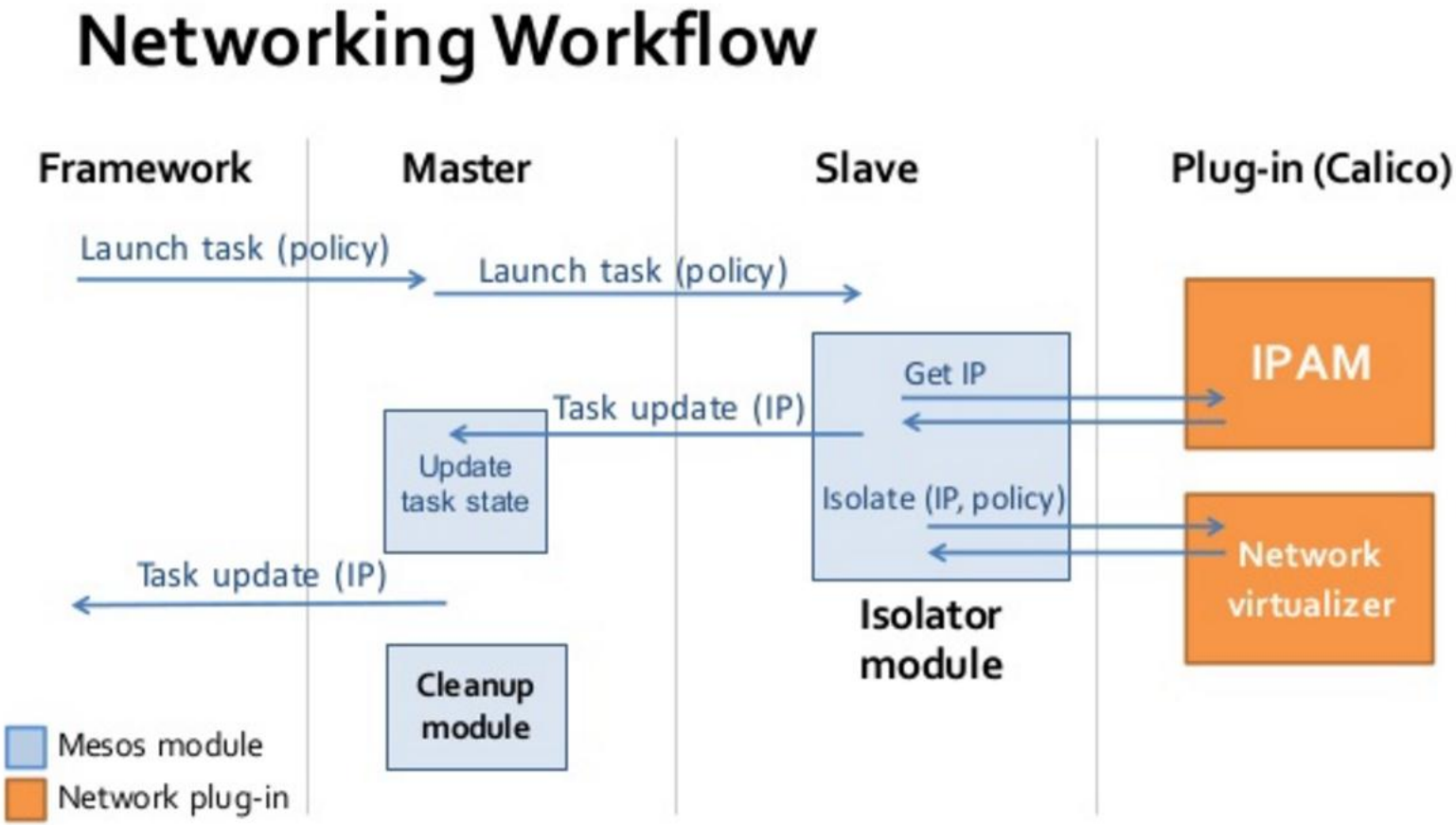
Source: https://www.percona.com/blog/2016/08/03/testing-docker-multi-host-network-performance/

Calico的核心组件

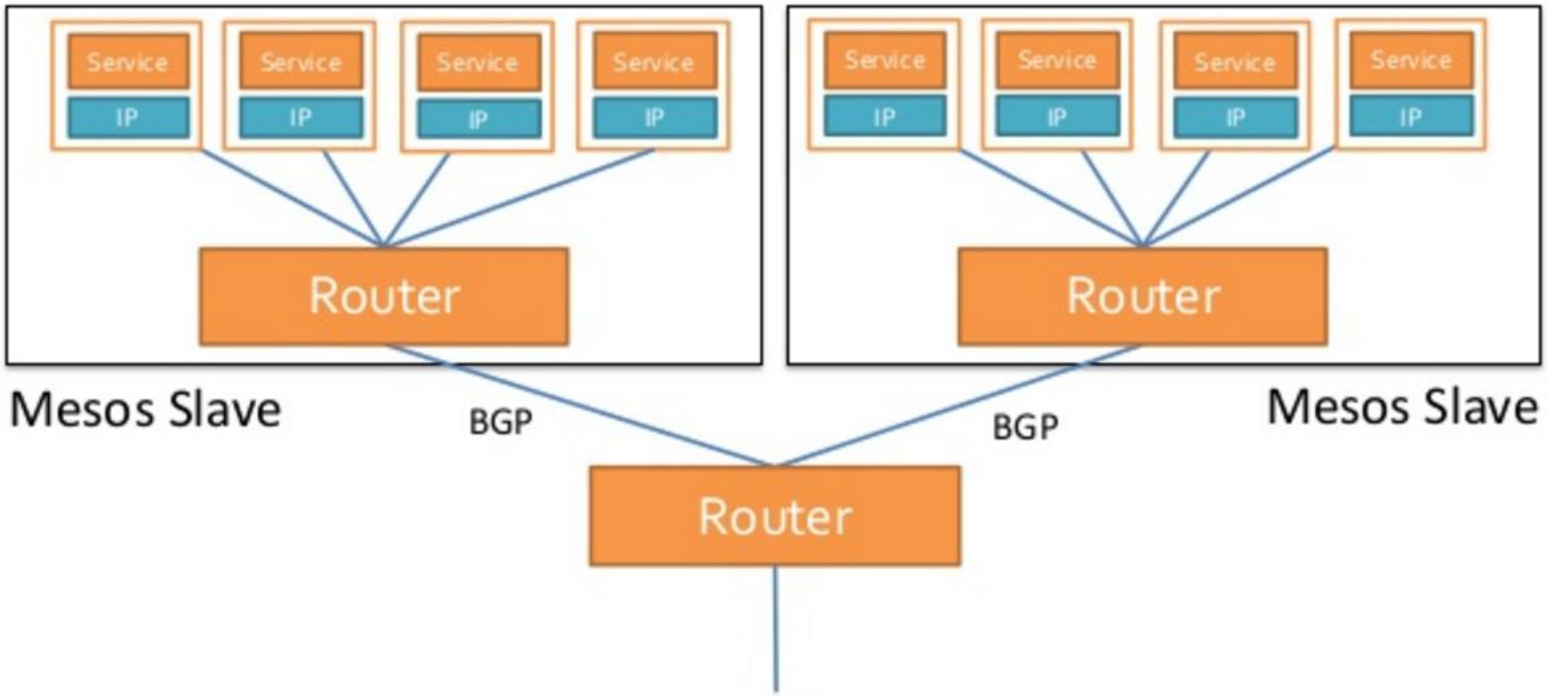


- Felix，Calico Agent，跑在每台需要运行Workload的节点上，主要负责配置路由及ACLs等信息来确保Endpoint的连通状态；
- etcd，分布式键值存储，主要负责网络元数据一致性，确保Calico网络状态的准确性；
- BGP Client（BIRD），主要负责把Felix写入Kernel的路由信息分发到当前Calico网络，确保Workload间的通信的有效性；
- BGP Route Reflector（BIRD），大规模部署时使用，摒弃所有节点互联的 mesh 模式，通过一个或者多个BGP Route Reflector来完成集中式的路由分发。

Calico网络访问流程



Mesos Cluster 最终网络图



THANKS!