World of Tech WOT

51CTO传媒

WOT 2015 互联网运维与开发者大会

■ 2015年04月10日-11日 ■ 北京珠三角JW万豪酒店

Docker的精细化运维

李雨来

运维关注的内容

- 性能
- 资源使用监控
- 网络
- 自动化


不爽
改造
舒服

Docker网络的改造

Docker的网络改造

- Docker的默认网络模型饱受诟病
- 一大波Docker SDN项目袭来 (pipework, weave, flannel…)

Docker的网络改造

- OpenVSwitch (或者Linux Bridge)
- iproute2
- Linux Net Namespace
- Docker

Docker的网络改造

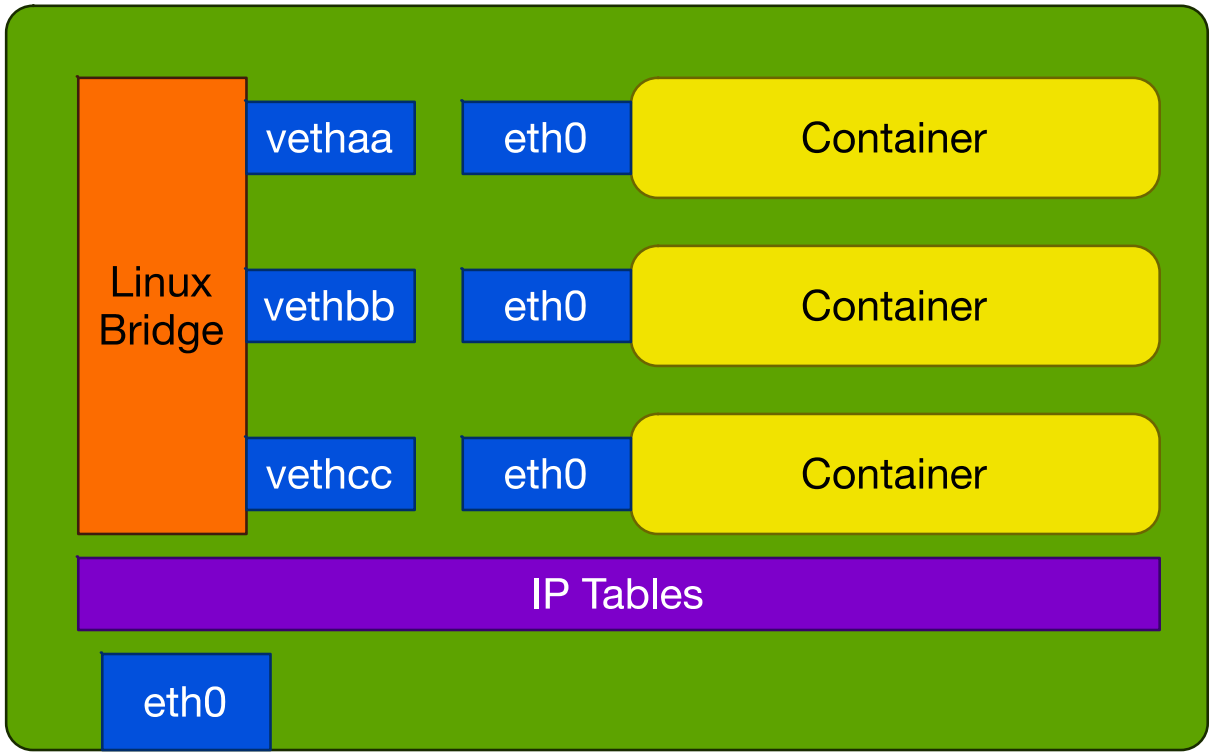
```
#!/bin/bash
PID=`docker inspect -f '{{.State.Pid}}'` $1
ID=`docker inspect -f '{{.Id}}'` $1
ETHNAME=$2
ln -s /proc/$PID/ns/net /var/run/netns/$ID
ip link add dev $ETHNAME.0 type veth peer name $ETHNAME.1
ip link set dev $ETHNAME.1 netns $ID
ip link set dev $ETHNAME.0 up
ip netns exec $ID ifconfig $ETHNAME.1 $3 up
rm -rf /var/run/netns/$ID
```

- 用法： network.sh docker-test veth0 192.168.1.10/24

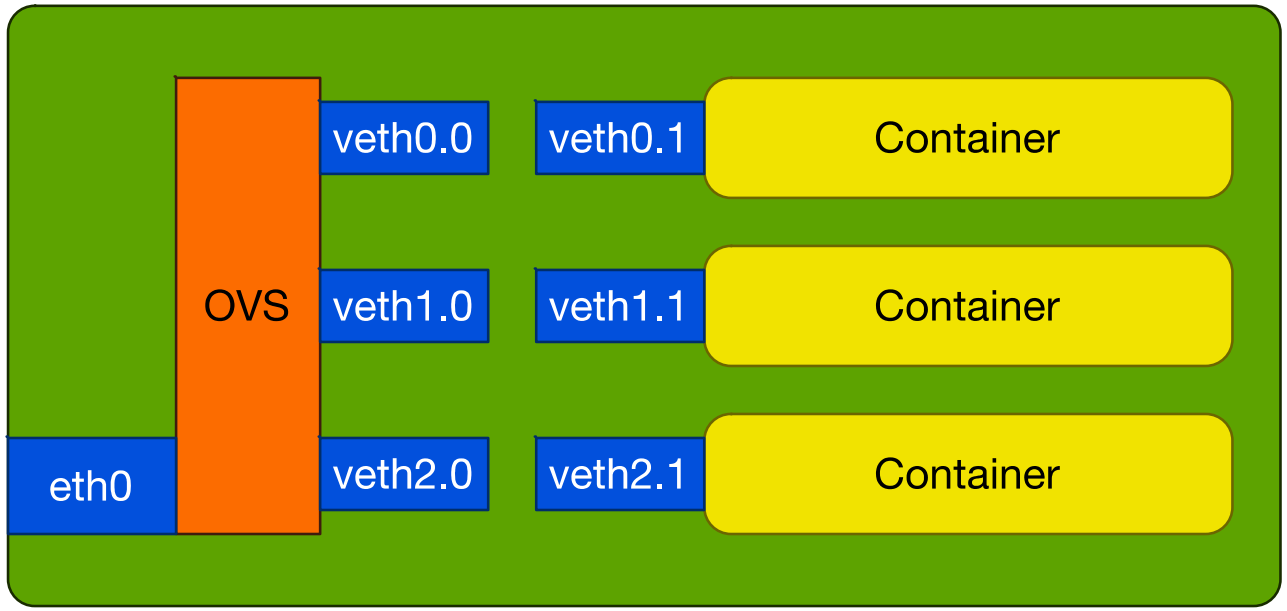
Docker的网络改造

- Linux Net Namespace: /proc/PID/ns/net
- ip link add 命令创建 veth 网卡
- ip netns 命令通过检测 /var/run/netns 路径调整netns的配置
- 使用ovs-vsctl（或者brctl）来设置新网卡的网络

Docker的网络改造



Docker的网络改造



想象力

- OpenFlow
- OpenStack Neutron
- SR-IOV
- VLAN,NVGRE,VXLAN

Docker资源隔离的改进

Docker的资源隔离改进

- 如果某个Container的CPU使用过多
- 如果某个Container的磁盘IO使用过多
- 如果某个Container的网络资源使用过多
- 如果某个Container的磁盘容量使用过多

Docker的资源隔离改进

- Docker的cgroups模型
  - Libcontainer： /sys/fs/cgroup/<subsystem>/docker/<containerID>
  - Linux Container: /sys/fs/cgroup/<subsystem>/lxc/<containerID>

- Kernel启动参数
  - cgroup_enable=memory swapaccount=1

CPU和内存的限制

- cpuset.cpus
- cpu.shares
- memory.memsw.limit_in_bytes
- memory.limit_in_bytes

磁盘IO的限制

- blkio.throttle.read_bps_device
- blkio.throttle.write_bps_device
- blkio.throttle.read_iops_device
- blkio.throttle.write_iops_device

- echo "<major>:<minor> <limit>" > throttle.write_iops_device

- cat /proc/partitions

网络带宽的限制

- OpenVSwitch
  - ovs-vsctl set interface veth1 ingress_policing_rate=1000
- ebtables + tc
  - ebtables -A FORWARD -i veth1 -j mark --mark-set 0x1 --mark-target ACCEPT
  - qdics, class, filter
  - tc filter add dev eth0 parent 1:0 protocol ip handle 1 fw flowid 1:1

磁盘容量的限制

- LVM创建卷，挂载之后通过--volume参数让Container访问

- 使用btrfs，并设置quota
  - btrfs qgroup limit –e 100G /var/lib/docker/btrfs/subvolumes/CONTAINER_ID

Container的监控

Container的监控

- CPU： cpuacct.usage
- 内存：
  - memory.usage_in_bytes
  - memory.memsw.usage_in_bytes
- 磁盘IO：
  - blkio.throttle.io_serviced
  - blkio.throttle.io_service_bytes
- 带宽：
  - /sys/class/net/<ethname>/statistics/
- docker stats <containerID>

升级Docker

- 慎重重启Docker Daemon进程

谢谢大家