Calico问题

Subtitle

2022/10/05

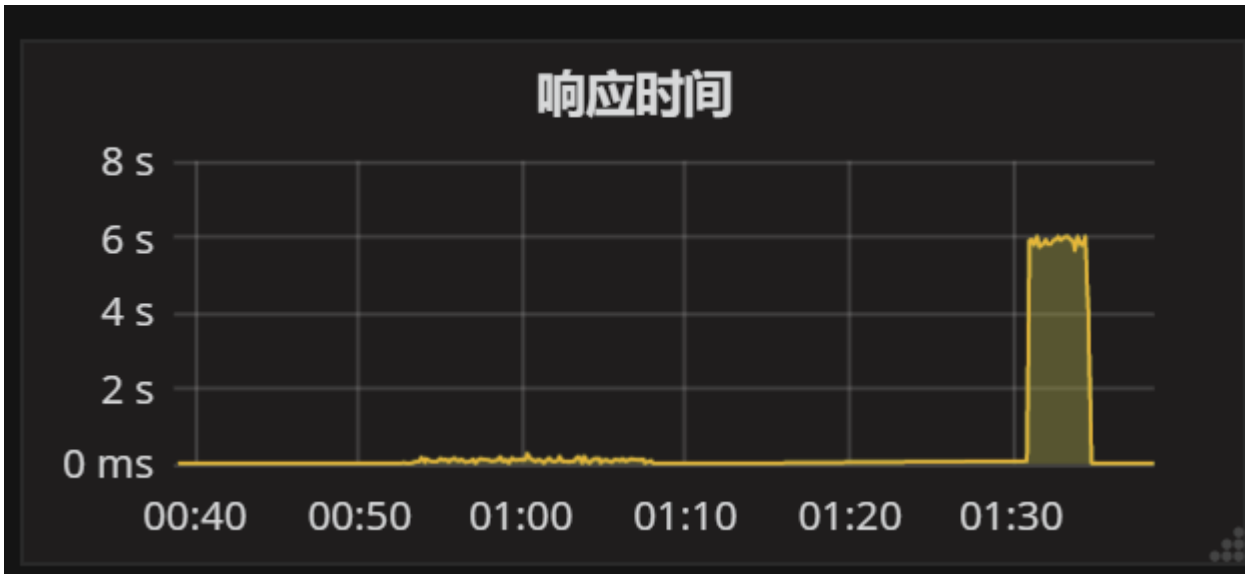
Table of Contents

Calico问题 1
机器重启后路由表建立很慢 1
可能的解决方案 1
bind a non-exists ip address 3
typha healthcheck bind localhost 3
误删calico-node daemonset 3

Calico问题

机器重启后路由表建立很慢

严重影响ingress，导致ingress不能访问容器网络



see:

- https://github.com/projectcalico/calico/issues/2211
- https://github.com/kubernetes/kops/issues/3224

```
2019-08-21 18:02:48.005 [INFO][67] route_table.go 133: Calculated interface name regexp
regex="^cali.*"
2019-08-21 18:02:48.007 [INFO][67] int_dataplane.go 614: Will refresh routes on timer
interval=1m30s
2019-08-21 18:02:48.054 [INFO][67] route_table.go 199: Trying to connect to netlink
```

可能的解决方案

1. 在 postStart 里检测
2. taint: https://cloud.google.com/kubernetes-engine/docs/how-to/node-taints?hl=zh-cn
3. 开机检测容器网络（dig @169.169.0.2），如不通则用 iptables reject 80, 443端口，网络就绪之后删除对应规则（此方法仅对ingress有用，如果刚启动的节点就有容器调度过来，会受到影响）

snippet.bash

```
#!/bin/bash

# 机器重启时calico 路由表需要几分钟才能正常刷新，导致容器网络不通
# 影响ingress，此脚本用于检测容器网络，不通时禁用 80，443端口
# 让 lvs 健康检测将此 ingress 节点标记为不可用
# 此脚本应放置于 /etc/rc.local 在 开机时执行一次
# centos 7中需给 /etc/rc.d/rc.local加x权限，见
https://blog.csdn.net/chenghuikai/article/details/45173909
```

```
# 请勿使用cron执行
# 如果有未预料的情况，可能导致所有节点都被加上iptables规则

COREDNS="169.169.0.2"
DOMAIN="kubernetes.default.svc.cluster.local."
MAX=200

function comment() {
  [ "$1"x == ""x ] && echo "param port needed. exit..." && exit 1
  echo "calico route not ready.reject lvs port $1"
}

function checkroute() {
  count=`route -n |grep tunl0 |wc -l`
  dig @$COREDNS +time=1 +short $DOMAIN &>/dev/null
  [ $? -eq 0 -a $count -ge 2 ] && return 0 || return 1
}

function entryExists() {
  COMMENT=`comment "$1"`
   iptables -L -n |grep "$COMMENT" &>/dev/null && return 0 || return 1
}

function addEntry() {
   entryExists $1
  if [ $? -eq 1 ];then
      iptables -A INPUT -p tcp --dport $1 -j REJECT -m comment --comment "`comment $1`"
  else
    echo "entry exists. continue..."
  fi
}

function deleteEntry() {
  while true;do
    COMMENT=`comment $1`
    n=`iptables -L -n --line-number |grep "$COMMENT" |head -n1 |awk '{print $1}'`
    if [ "$n"x == ""x ];then
      echo "no entry found. exit..."
      break
    else
      echo "delete entry: $COMMENT"
        iptables -D INPUT $n
    fi
  done
}

n=0
while true;do
  echo "try $n times..."
   checkroute
  if [ $? -eq 0 ];then
    echo "checkroute pass. will delete iptables entry"
```

```
      deleteEntry 80
      deleteEntry 443
    break
  else
    echo "checkroute failed. will reject 80 and 443"
      addEntry 80
      addEntry 443
  fi
  ((n+=1))

  if [ $n -gt $MAX ];then
      echo "too many retrys. exit.."
      deleteEntry 80
      deleteEntry 443
    break
  fi
  sleep 3
done
```

bind a non-exists ip address

calico node try to bind a non-exists ip address: https://github.com/projectcalico/calico/issues/2146

typha healthcheck bind localhost

缺失/etc/nsswitch.conf，不能正确解析localhost

误删calico-node daemonset

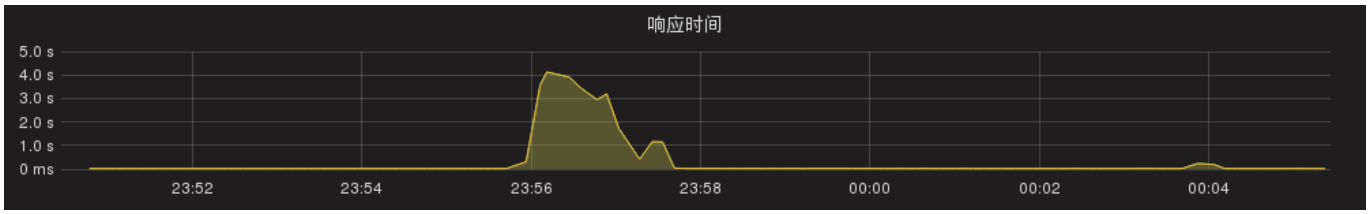
从dashboard上误删了整个calico-node 的daemonset，容器网络暂时正常。但是新增机器可能会有问题

误删之后 新增容器，网络不通

新增容器网络不通的原因

当我们在k8s中创建一个Pod时，kubelet组件会调用CNI插件来为Pod添加虚拟网卡，veth设备就是在这个时候被创建的。参见：https://sycki.com/articles/kubernetes/k8s-calico

重建calico-node之后，网络基本正常，仅有部分大流量业务受到短时影响



更新，以上响应时间抖动原因是网络正常之后，hpa恢复，触发此服务依赖服务的自动缩容

事件

消息	来源	子对象	总数	最早出现于	最近出现于
unable to get metrics for resource memory: unable to fetch metrics from resource metrics API: the server is currently unable to handle the request (get pods.metrics.k8s.io)	horizontal-pod-autoscaler	-	712	2018-11-21T09:50 UTC	2018-12-15T15:42 UTC
failed to get memory utilization: unable to get metrics for resource memory: unable to fetch metrics from resource metrics API: the server is currently unable to handle the request (get pods.metrics.k8s.io)	horizontal-pod-autoscaler	-	688	2018-11-21T09:50 UTC	2018-12-15T15:37 UTC
New size: 8; reason: All metrics below target	horizontal-pod-autoscaler	-	31	2018-11-21T15:47 UTC	2018-12-15T15:55 UTC
Scaled down replica set device-manager-api-internal-697977d888 to 8	deployment-controller	-	28	2018-11-23T06:17 UTC	2018-12-15T15:55 UTC

Printed on: 2022/10/05 18:00