MUM2012

PCC 负载均衡

EDCwifi 林利钢

深圳捷联讯通科技有限公司

负载均衡(Load Balancing)

负载均衡：

负载均衡是建立在现有的网络结构之上，它提供了一种廉价、有效、透明的方法来扩展网络的带宽、增加吞吐量、加强网络数据处理能力、提高网络的灵活性和可用性。

负载均衡(Load Balancing)

当使用同一个类型的网络（即同一个运营商的网络）多线路接入时，可以采用负载均衡。

ECMP(Equal-Cost Multi-path):

ECMP是通过在route列表添加多网关的静态路由，然后由路由协议建立动态的多线路由。这种负载均衡有个缺点，缺点就是每十分钟内核会重新均衡线路，使一些连接会被指定到其他路线，出现频繁掉线的情况。

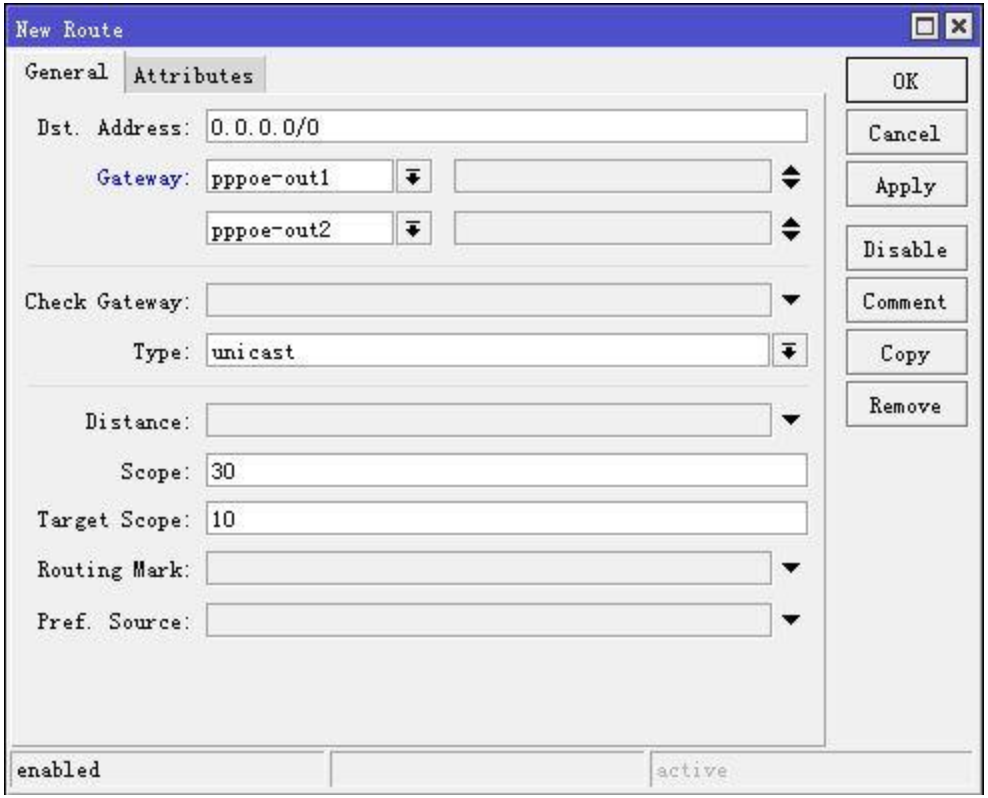
NTH:

NTH是采用第N次链接的负载均衡，它不仅可以实现基于IP的负载均衡，同时还能实现对端口负载均衡和对nat指定有序的访问。这样基本实现了不掉线的真正负载均衡。但是NTH存在着一个弊端，就是在某些对IP要求严格的网站会反复要求验证。比如，网银！这样我们需要通过策略将一些IP或者端口指定走固定的线路出去，从而避开网站繁琐的验证。

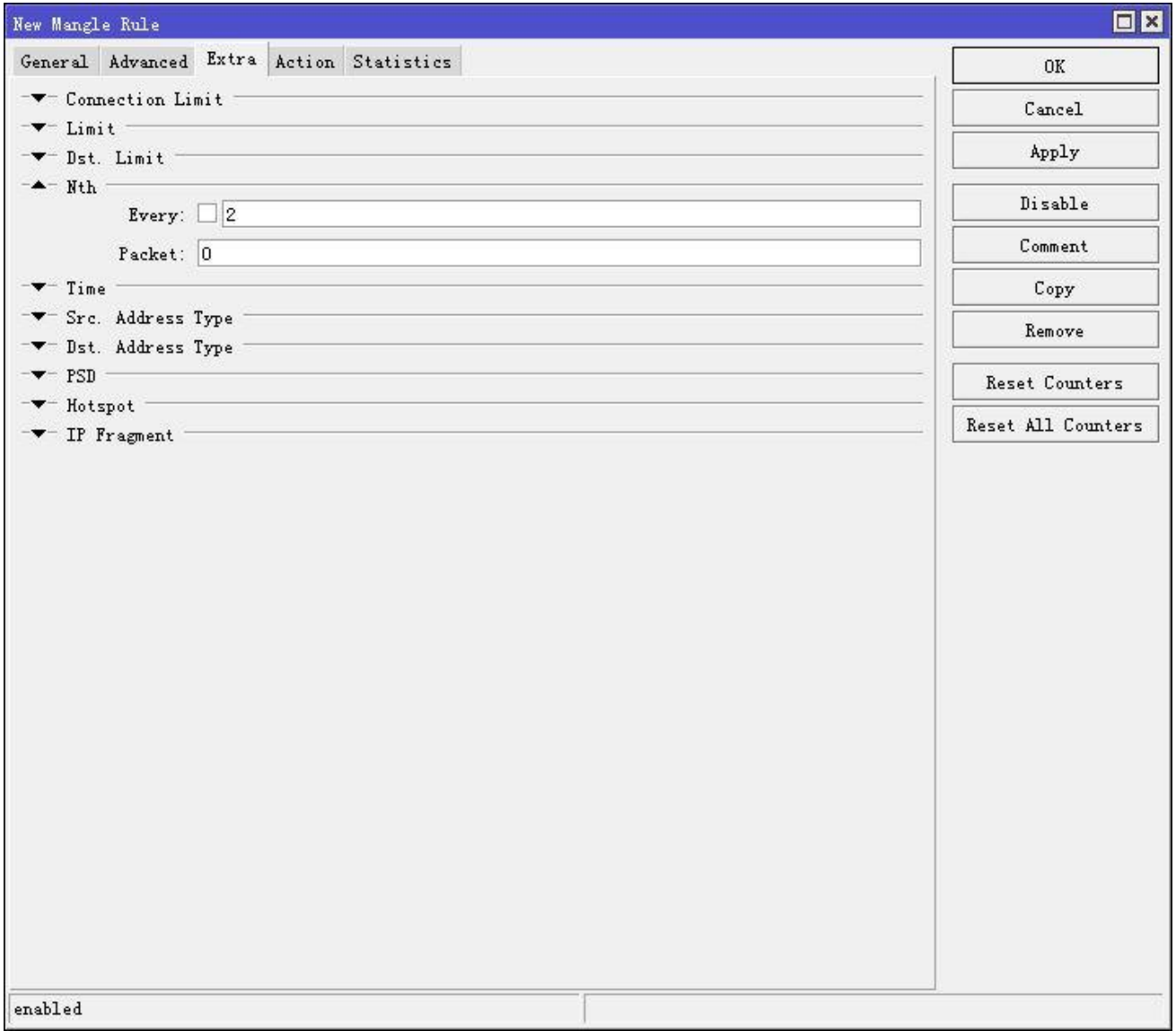
PCC（Per connection classified）：

PCC是通过判断源地址或者目的地址、源端口或者目的端口对数据进行分类来实现负载均衡，对每个连接进行分类大多保持了连续性，这样大大弥补了NTH的不足。

ECMP(Equal-Cost Multi-path)



NTH


New Mangle Rule
General
Advanced
Extra
Action
Statistics
Connection Limit
Limit
Dst. Limit
Nth
Every:
2
Packet:
0
Time
Src. Address Type
Dst. Address Type
PSD
Hotspot
IP Fragment
OK
Cancel
Apply
Disable
Comment
Copy
Remove
Reset Counters
Reset All Counters
enabled

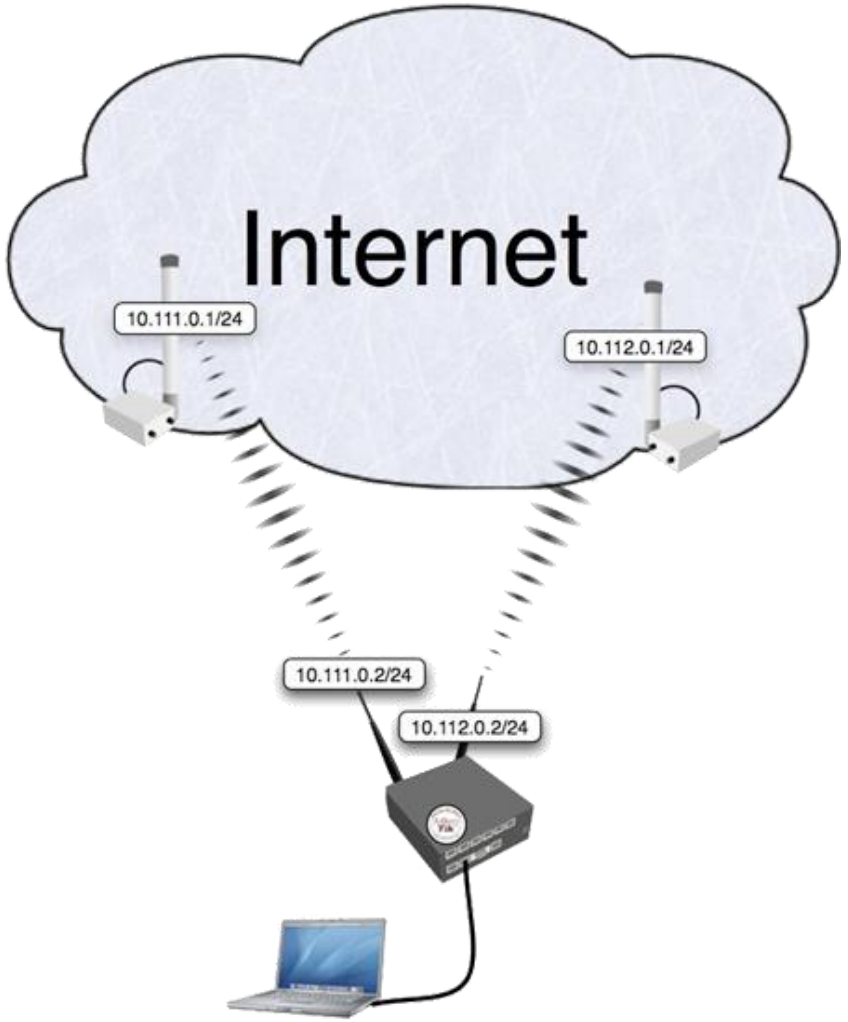
PCC 负载均衡

- 概述：PCC 允许分离传输流做到平衡流量的功能。（可以根据src-address、src-port、dst-address、dst-port选择）
- 原理：PCC从一定范围内分析选择IP数据包头，通过hash算法将选定的区域转换为32bit值。这个值除了指定分母（Denominator），余数将比较一个指定的余数（Remainder），如果相等这时数据包将会被提取。

PCC 负载均衡

双向地址负载均衡

通过分组源地址和目的地址实现负载均衡



PCC 负载均衡

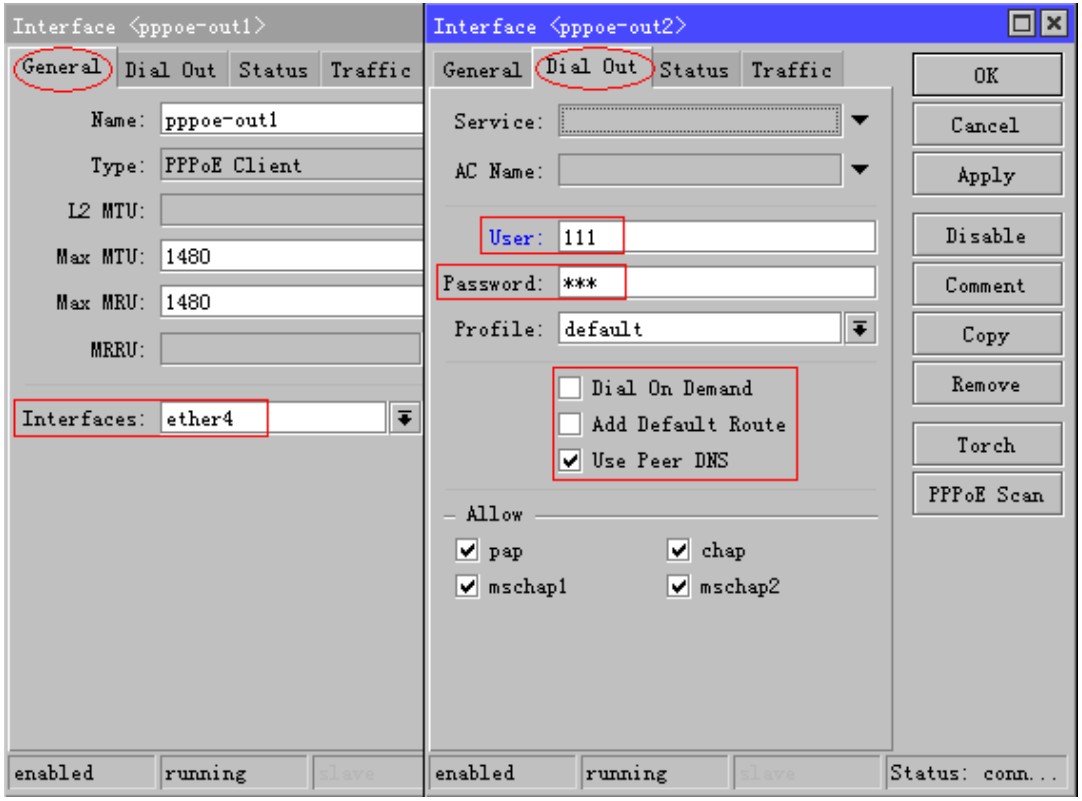
PCC的常规设置与命令

pppoe

添加PPPoE拨号。

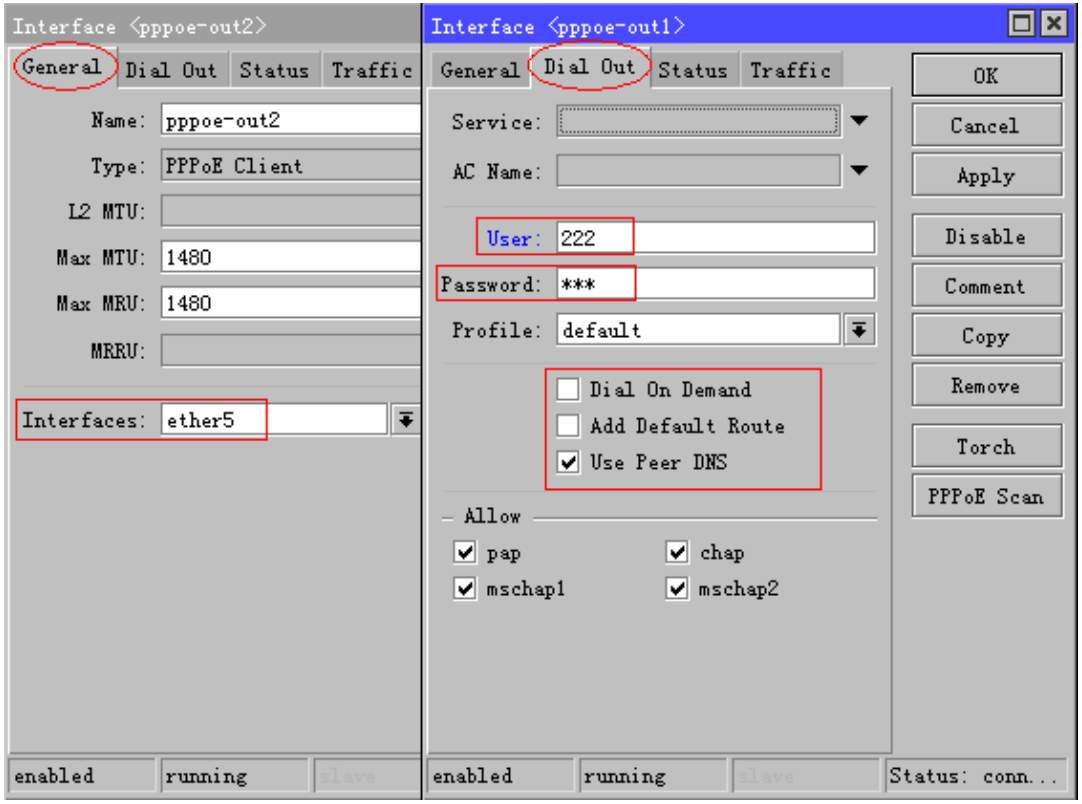
添加pppoe-client
在General下interface选择对应的网口
在Dial Out下填写账号、密码

Dial on demand：按需拨号
Add default route：添加默认路由
Use peer DNS：使用对等DNS

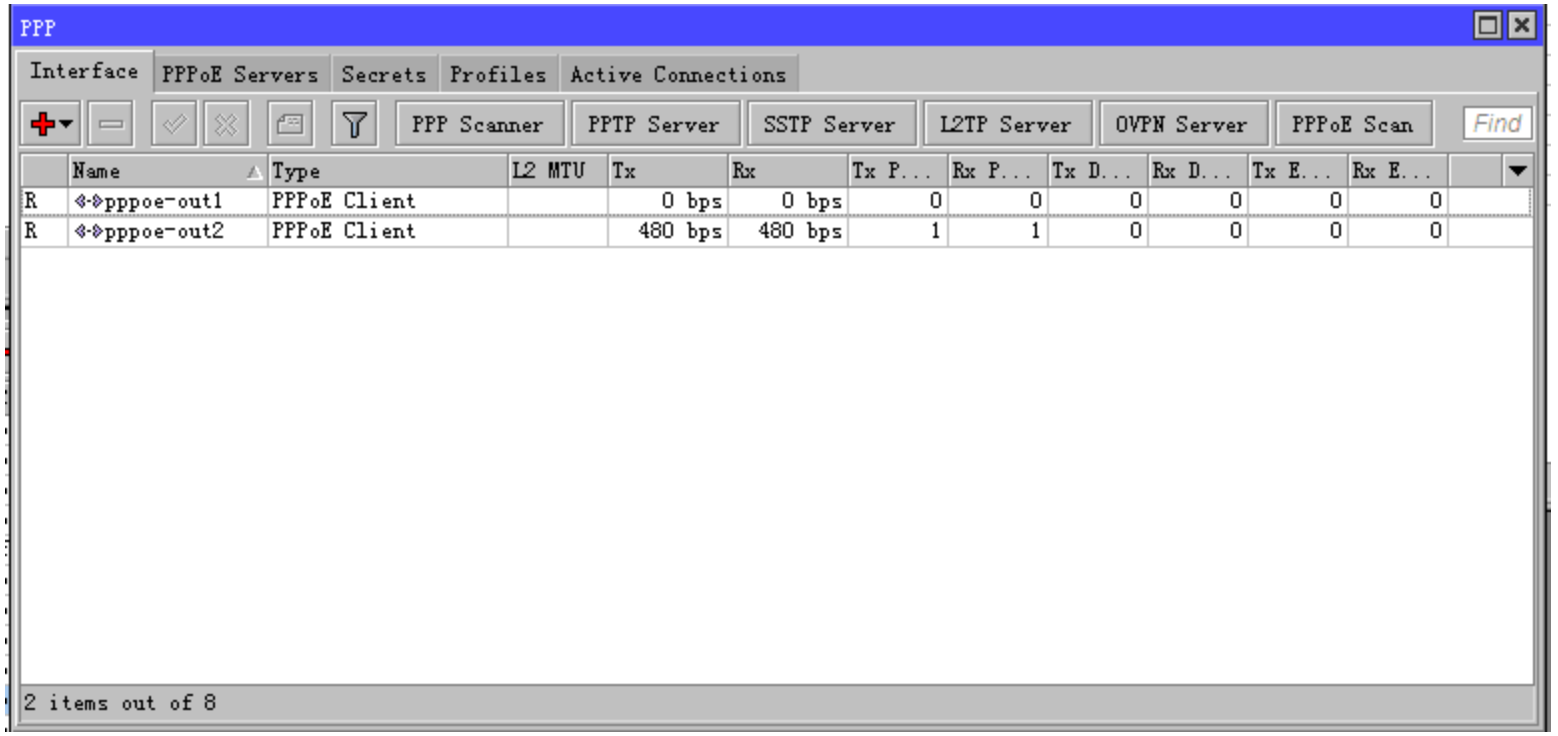


pppoe

添加另一个pppoe拨号



pppoe


PPP
Interface
PPPoE Servers
Secrets
Profiles
Active Connections
PPP Scanner
PPTP Server
SSTP Server
L2TP Server
OVPN Server
PPPoE Scan
Find
Name
Type
L2 MTU
Tx
Rx
Tx P...
Rx P...
Tx D...
Rx D...
Tx E...
Rx E...
R
pppoe-out1
PPPoE Client
0 bps
0 bps
0
0
0
0
0
0
R
pppoe-out2
PPPoE Client
480 bps
480 bps
1
1
0
0
0
0
2 items out of 8

pppoe

```
/interface pppoe-client
add ac-name="" add-default-route=no allow=pap,chap,mschap1,mschap2 \
   dial-on-demand=no disabled=no interface=ether4 max-mru=1480 max-
     mtu=1480 \
   mrru=disabled name=pppoe-out1 password="" profile=default service-
     name="" \
   use-peer-dns=yes user=""
add ac-name="" add-default-route=no allow=pap,chap,mschap1,mschap2 \
   dial-on-demand=no disabled=no interface=ether5 max-mru=1480 max-
     mtu=1480 \
   mrru=disabled name=pppoe-out2 password="" profile=default service-
     name="" \
   use-peer-dns=yes user=""
```

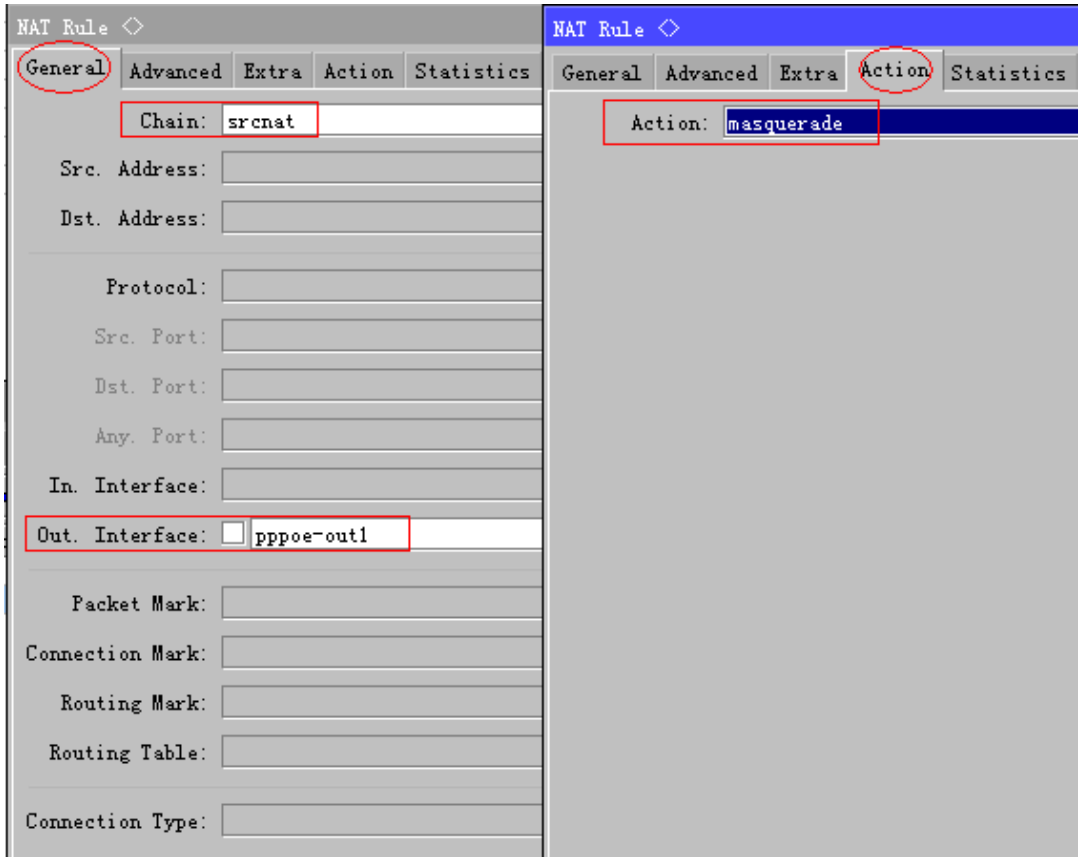
nat

masquerade 地址伪装：

可以根据src address、out interface或者src address list做地址伪装。

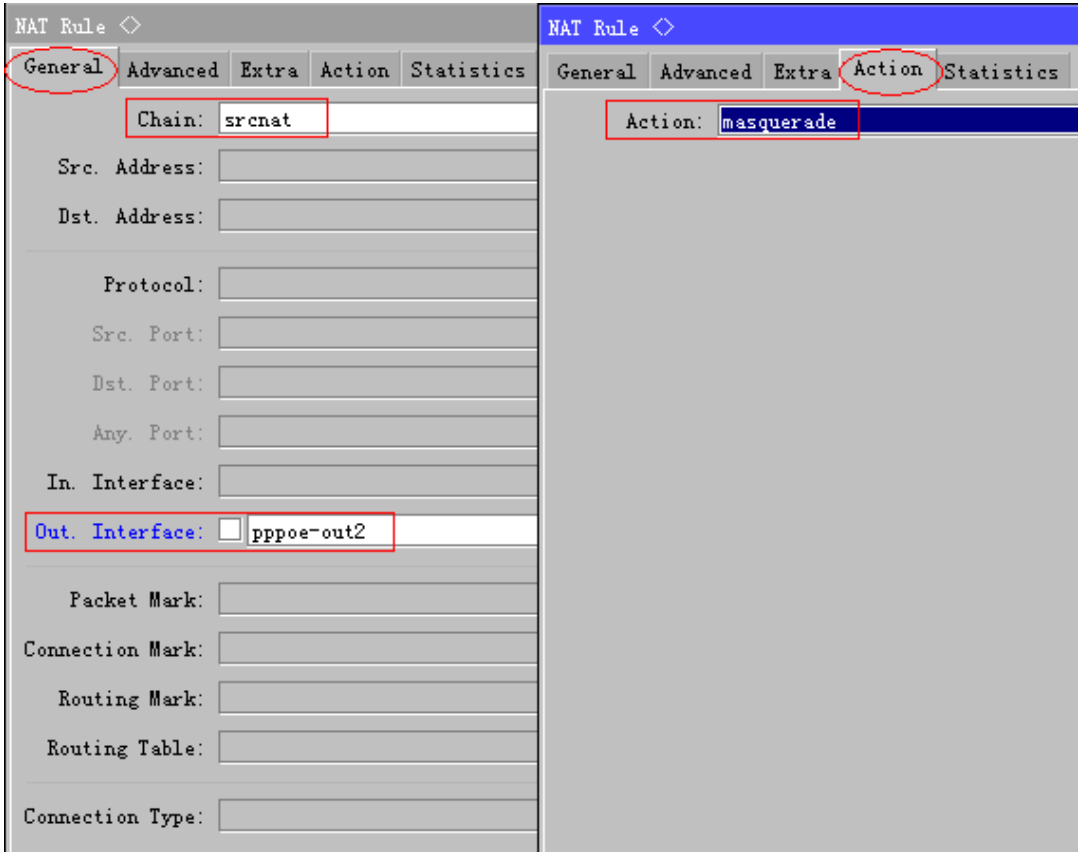
如果是PPPo拨号上网，这里需要注意的是out. Interface

应该选择对应的pppoe-out接口，而不是ether口！

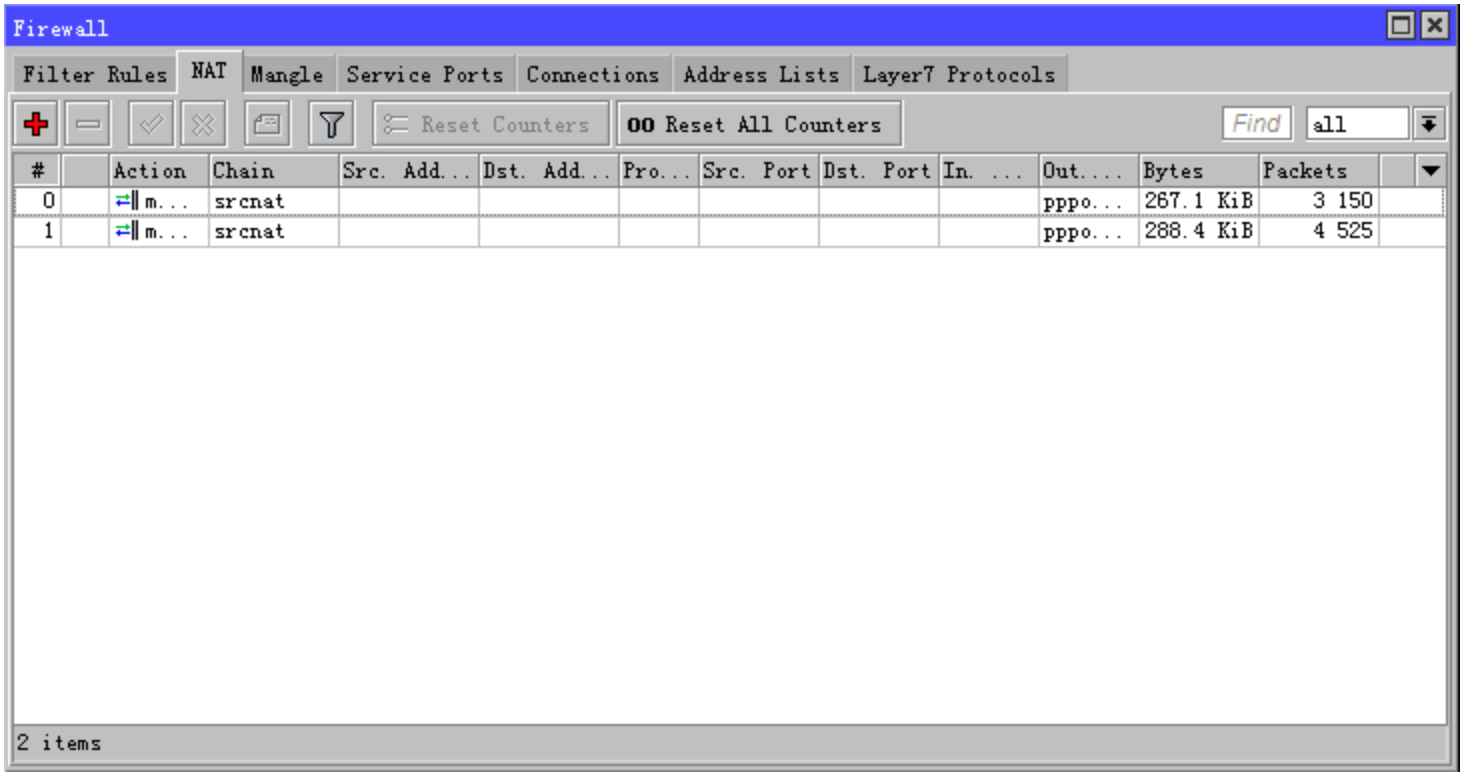


nat

另一个masquerade



nat


Firewall
Filter Rules
NAT
Mangle
Service Ports
Connections
Address Lists
Layer7 Protocols
Reset Counters
00 Reset All Counters
Find
all
Action Chain Src. Add... Dst. Add... Pro... Src. Port Dst. Port In. ... Out.... Bytes Packets
0 m... srcnat pppo... 267.1 KiB 3 150
1 m... srcnat pppo... 288.4 KiB 4 525
2 items

nat

/ip firewall nat

add action=masquerade chain=srcnat
disabled=no out-interface=pppoe-out1

add action=masquerade chain=srcnat
disabled=no out-interface=pppoe-out2

mangle

标记连接

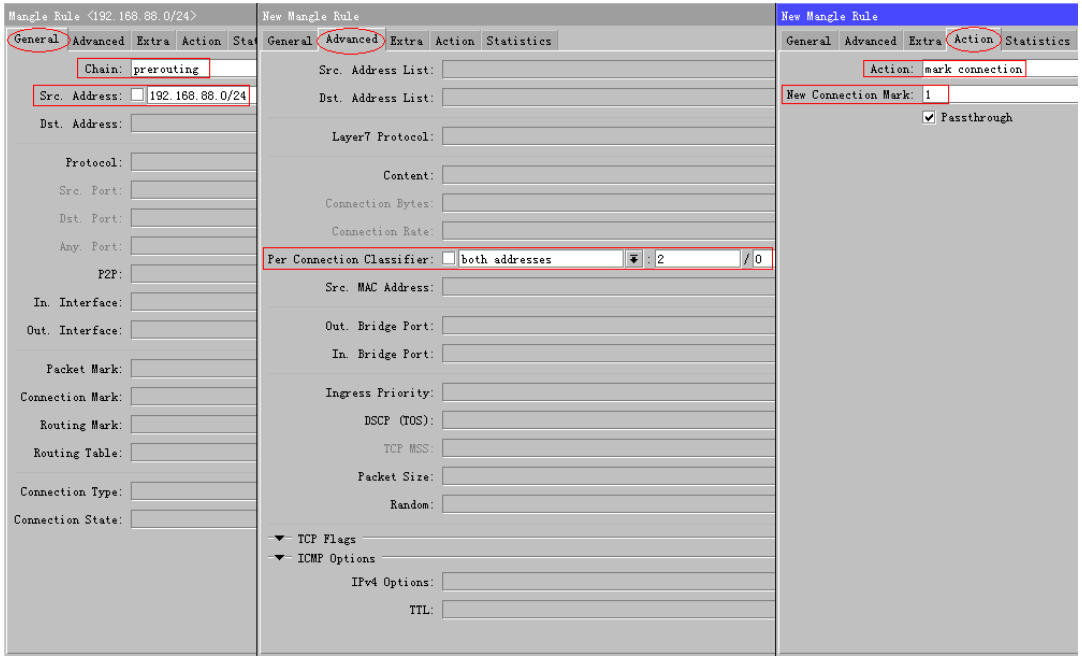
将进入路由的连接进行标记

这里我们用src address来标记

当然，也可以通过in interface选择内网接口来标记

这里我们使用both addresses来将源地址进行分类

2/0：2条路线，0定义为第一条



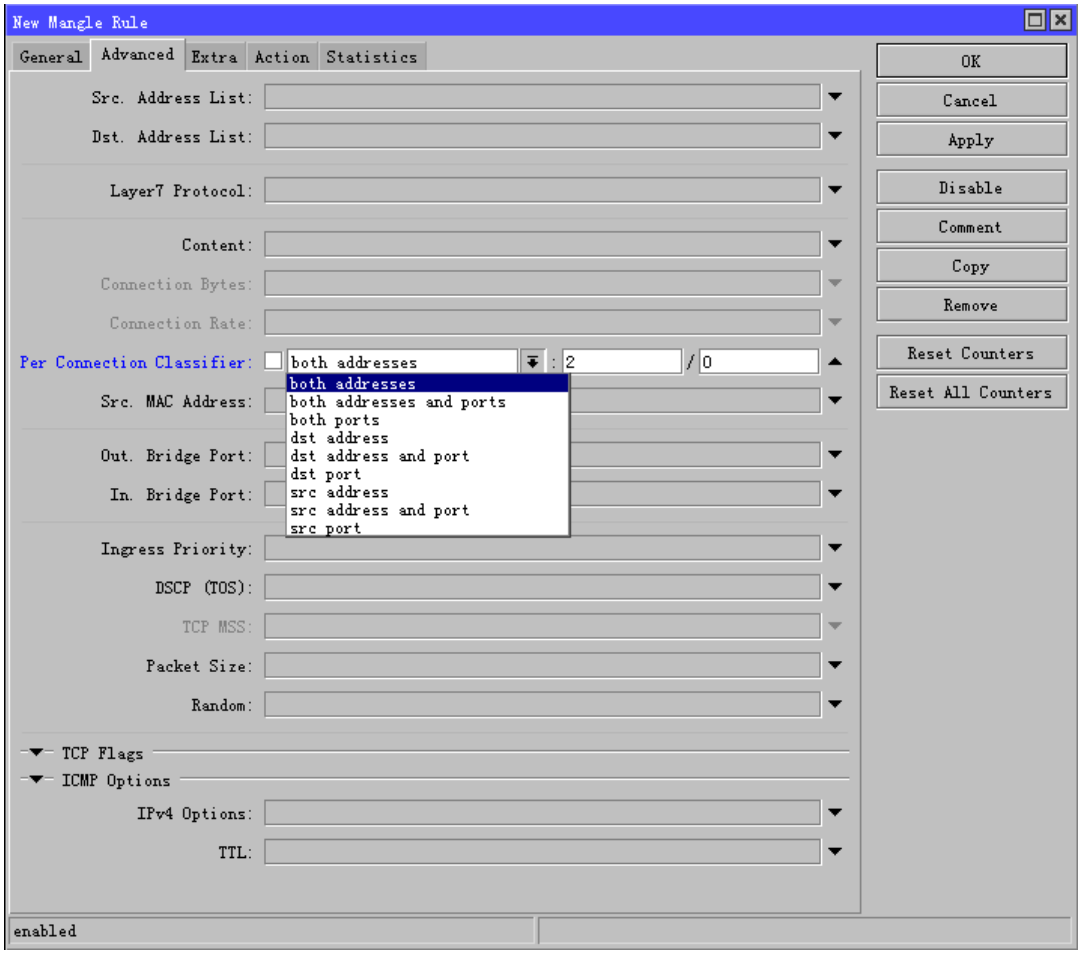
mangle

both-address(双向地址)

是以源地址和目的地址作为输入值。

如果数据包的源地址和目的地址相同，则连接被分为一组，将得到相同的哈希值。然后把所有组进行平分标记(当然也可以不平分标记)。这些数据包将被分配到同一条外网链接上。

both-address是比较稳定的一种设置方法，但不是很平均。



src address

这种是负载均衡里面最稳健的。在某些环境中，甚至用both adress都会出现问题。但是src-address是所有模式里面均衡效果最差的，因为兼容性和均衡效果不可得兼。

src-address and port

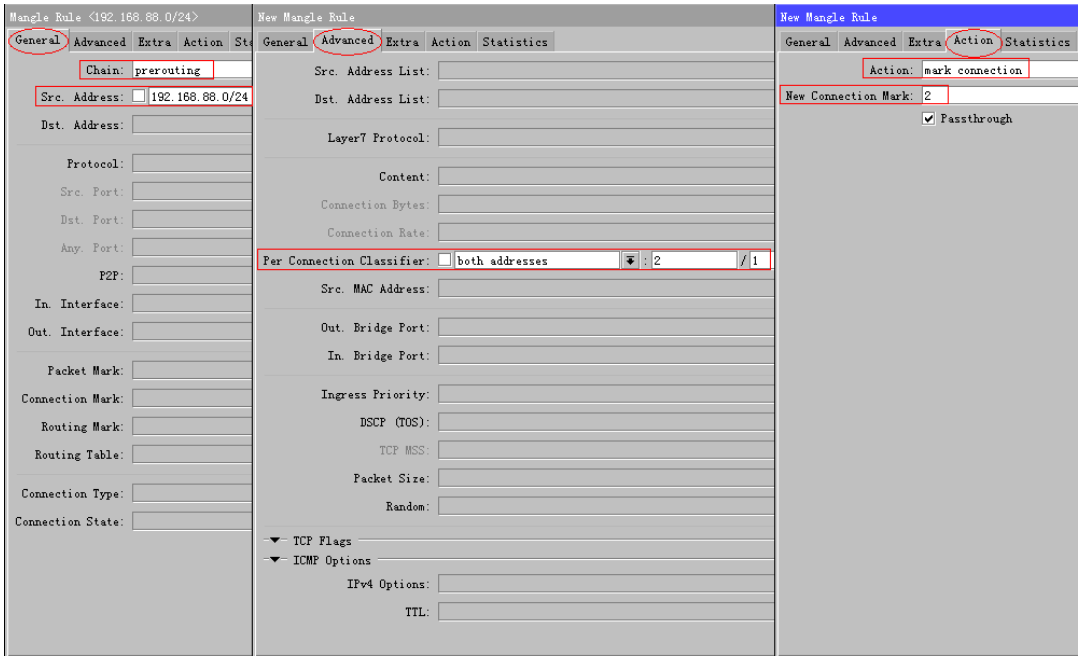
对相同源地址和相同源端口的数据包来说，将会被分配到同一条线路上。这样的话，对于目的地址相同的数据包，也有可能分别走不同的线路，这在对安全性要求比较高的环境中，是不能被接受的。

both addrss and ports

是均衡效果最好的。因为带有port的输入参数，引入了port，而port数值从1-65535，因而hash的输入样本大大增加，使数据包平均分流到各条线路的概率也就大大增加了！

mangle

标记连接


Mangle Rule <192.168.88.0/24>
General
Advanced
Extra
Action
Chain: prerouting
Src. Address: 192.168.88.0/24
Dst. Address:
Protocol:
Src. Port:
Dst. Port:
Any. Port:
P2P:
In. Interface:
Out. Interface:
Packet Mark:
Connection Mark:
Routing Mark:
Routing Table:
Connection Type:
Connection State:
New Mangle Rule
General
Advanced
Extra
Action
Statistics
Src. Address List:
Dst. Address List:
Layer7 Protocol:
Content:
Connection Bytes:
Connection Rate:
Per Connection Classifier: both addresses : 2 / 1
Src. MAC Address:
Out. Bridge Port:
In. Bridge Port:
Ingress Priority:
DSCP (TOS):
TCP MSS:
Packet Size:
Random:
TCP Flags
ICMP Options
IPv4 Options:
TTL:
New Mangle Rule
General
Advanced
Extra
Action
Statistics
Action: mark connection
New Connection Mark: 2
Passthrough

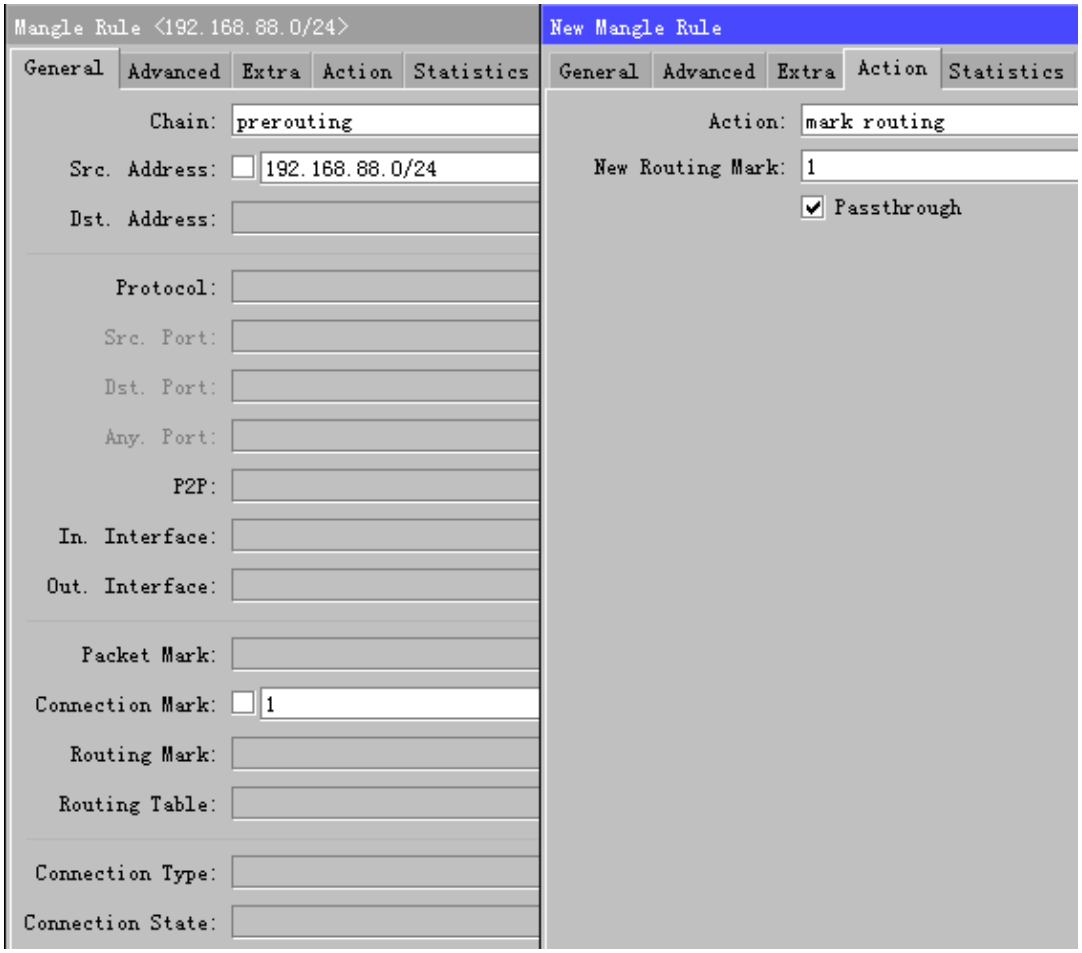
mangle

标记路由

同样标记路由也可以通过src address或者in interface进行标记

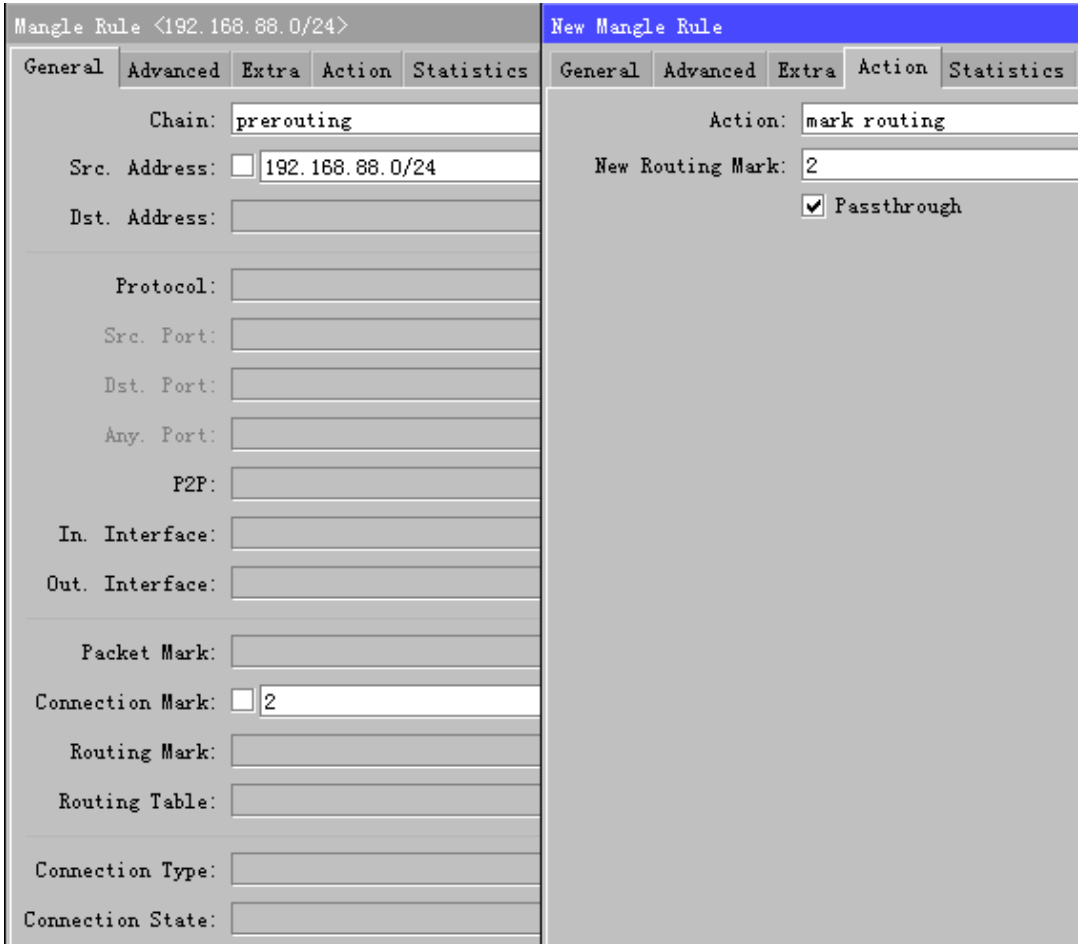
选择对应的连接标记

然后标记路由，新建一个新的路由标记

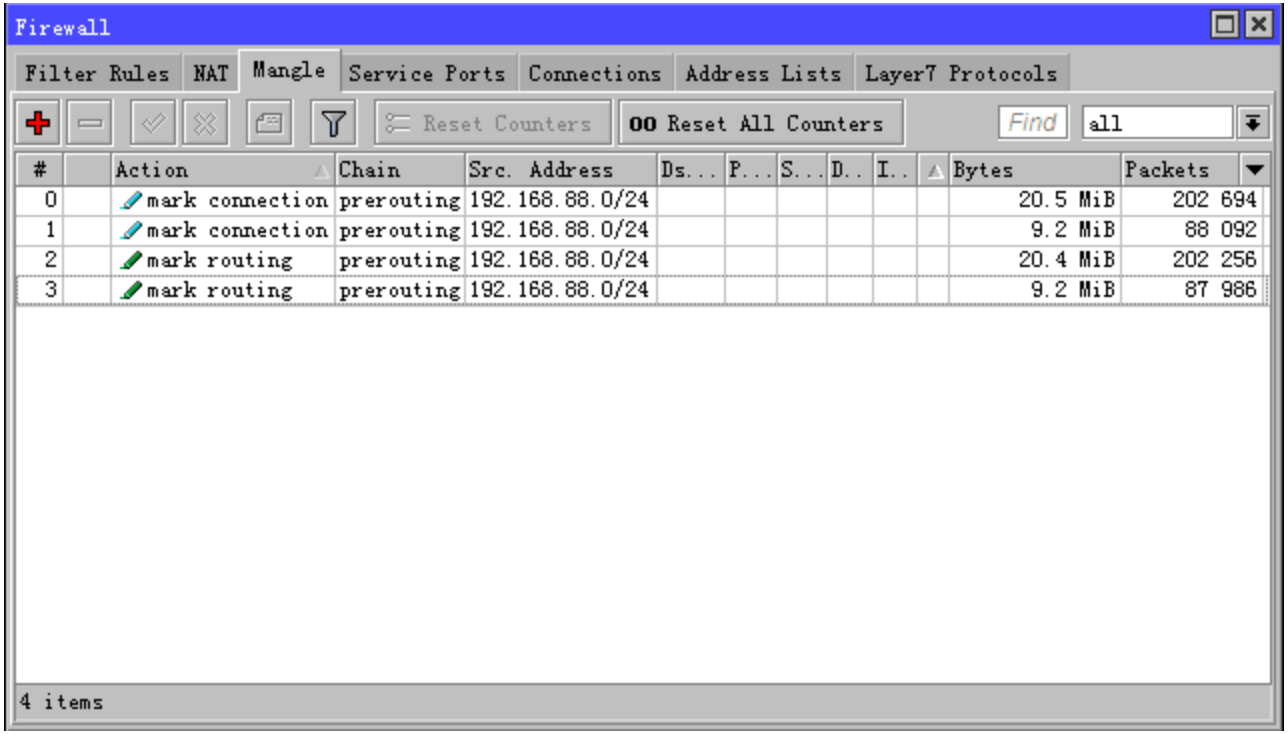


mangle

标记路由


Mangle Rule <192.168.88.0/24>
General
Advanced
Extra
Action
Statistics
Chain: prerouting
Src. Address: 192.168.88.0/24
Dst. Address:
Protocol:
Src. Port:
Dst. Port:
Any. Port:
P2P:
In. Interface:
Out. Interface:
Packet Mark:
Connection Mark: 2
Routing Mark:
Routing Table:
Connection Type:
Connection State:
New Mangle Rule
General
Advanced
Extra
Action
Statistics
Action: mark routing
New Routing Mark: 2
Passthrough

mangle


Firewall
Filter Rules
NAT
Mangle
Service Ports
Connections
Address Lists
Layer7 Protocols
Reset Counters
00 Reset All Counters
Find
all
Action Chain Src. Address Ds... P... S... D.. I.. Bytes Packets
0 mark connection prerouting 192.168.88.0/24 20.5 MiB 202 694
1 mark connection prerouting 192.168.88.0/24 9.2 MiB 88 092
2 mark routing prerouting 192.168.88.0/24 20.4 MiB 202 256
3 mark routing prerouting 192.168.88.0/24 9.2 MiB 87 986
4 items

mangle

```
/ip firewall mangle
add action=mark-connection chain=prerouting disabled=no new-connection-
    mark=1 \
   passthrough=yes per-connection-classifier=both-addresses:2/0 src-
    address=\
   192.168.88.0/24
add action=mark-connection chain=prerouting disabled=no new-connection-
    mark=2 \
   passthrough=yes per-connection-classifier=both-addresses:2/1 src-
    address=\
   192.168.88.0/24
add action=mark-routing chain=prerouting connection-mark=1 disabled=no \
   new-routing-mark=1 passthrough=yes src-address=192.168.88.0/24
add action=mark-routing chain=prerouting connection-mark=2 disabled=no \
   new-routing-mark=2 passthrough=yes src-address=192.168.88.0/24
```

route

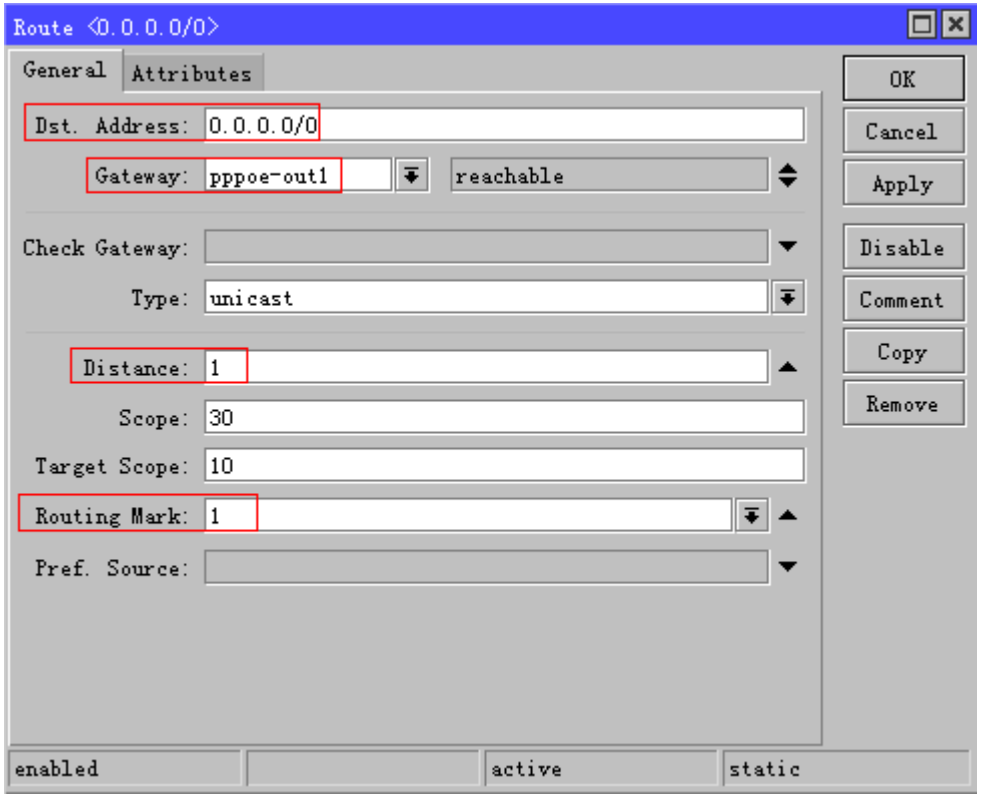
路由标记

所有目的地址走pppoe-out1网关

Distance（距离）在这里是路由的优先级

分为1~255个级别，1是最优先（数值越小越优先）

路由标记选择刚才mangle标记的路由1

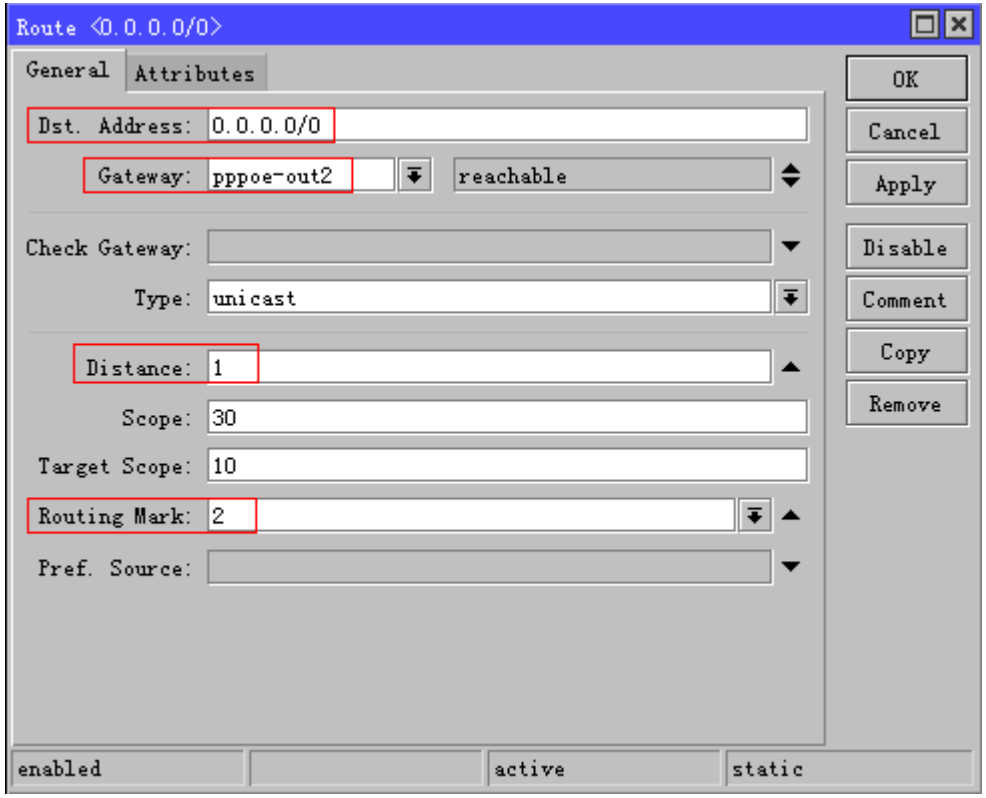


route

路由标记

所有目的地址走pppoe-out2网关

路由标记选择刚才mangle标记的路由2



route

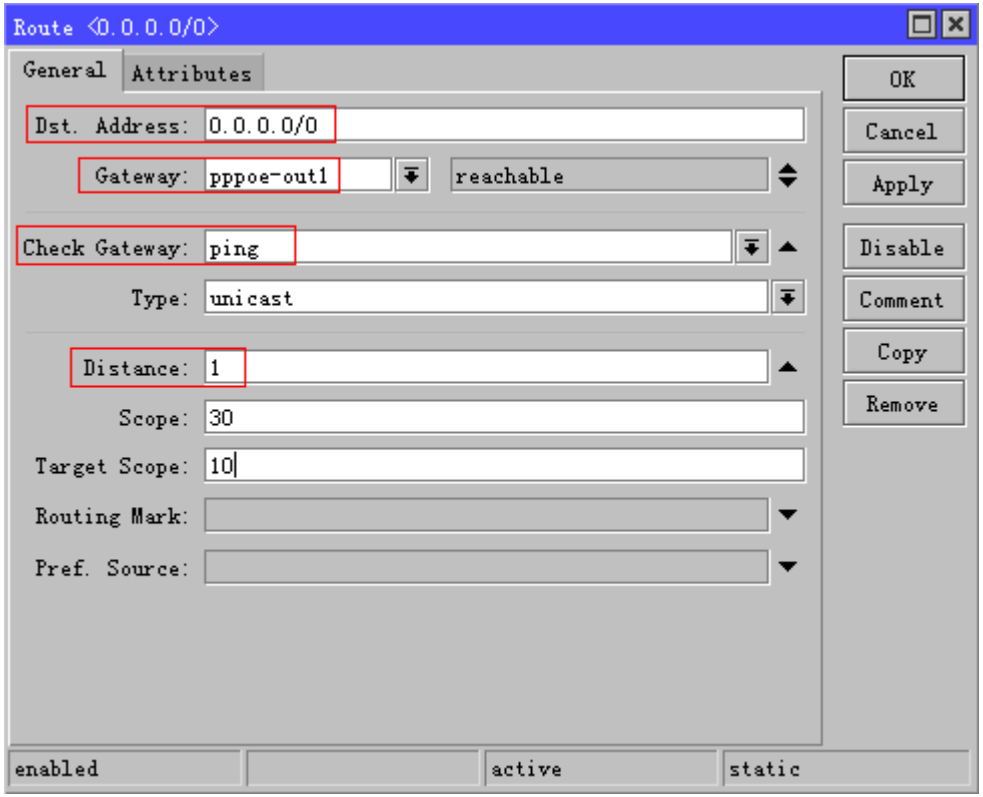
默认网关

Check gateway 检查网关

这里我们用ping网关的方式来判断网络是否通，如果不通路由表则自动切换到另一条路线做默认网关

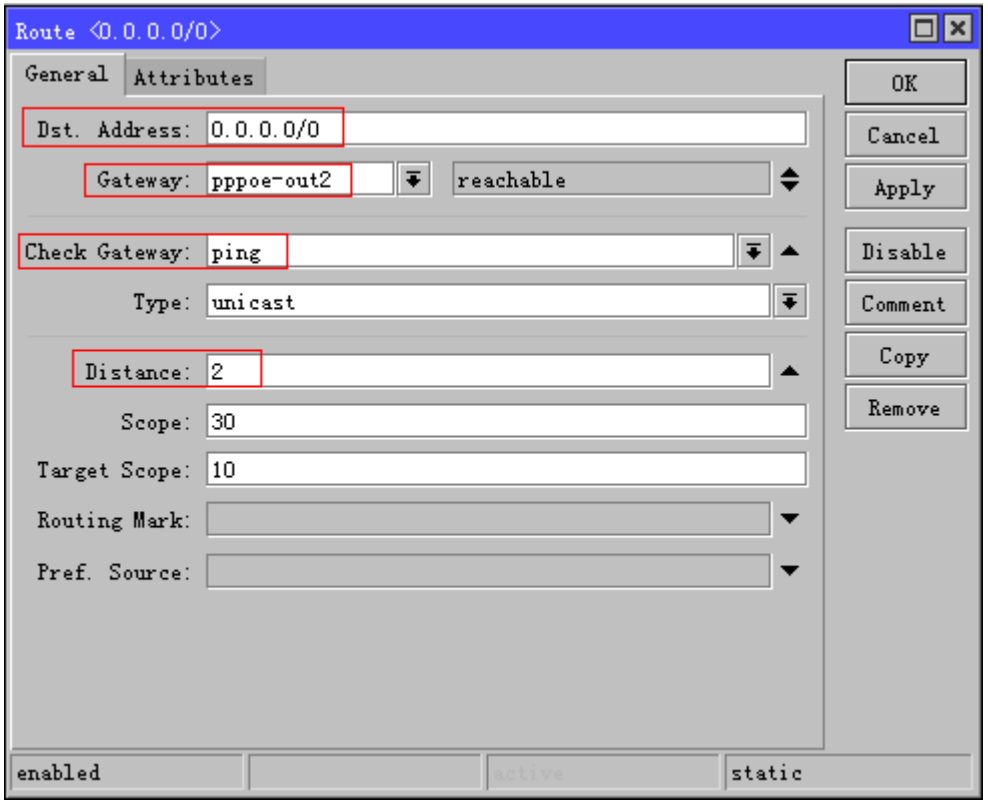
Distance（距离） 在这里是默认网关的优先级

分为1~255个级别，1是最优先（数值越小越优先）

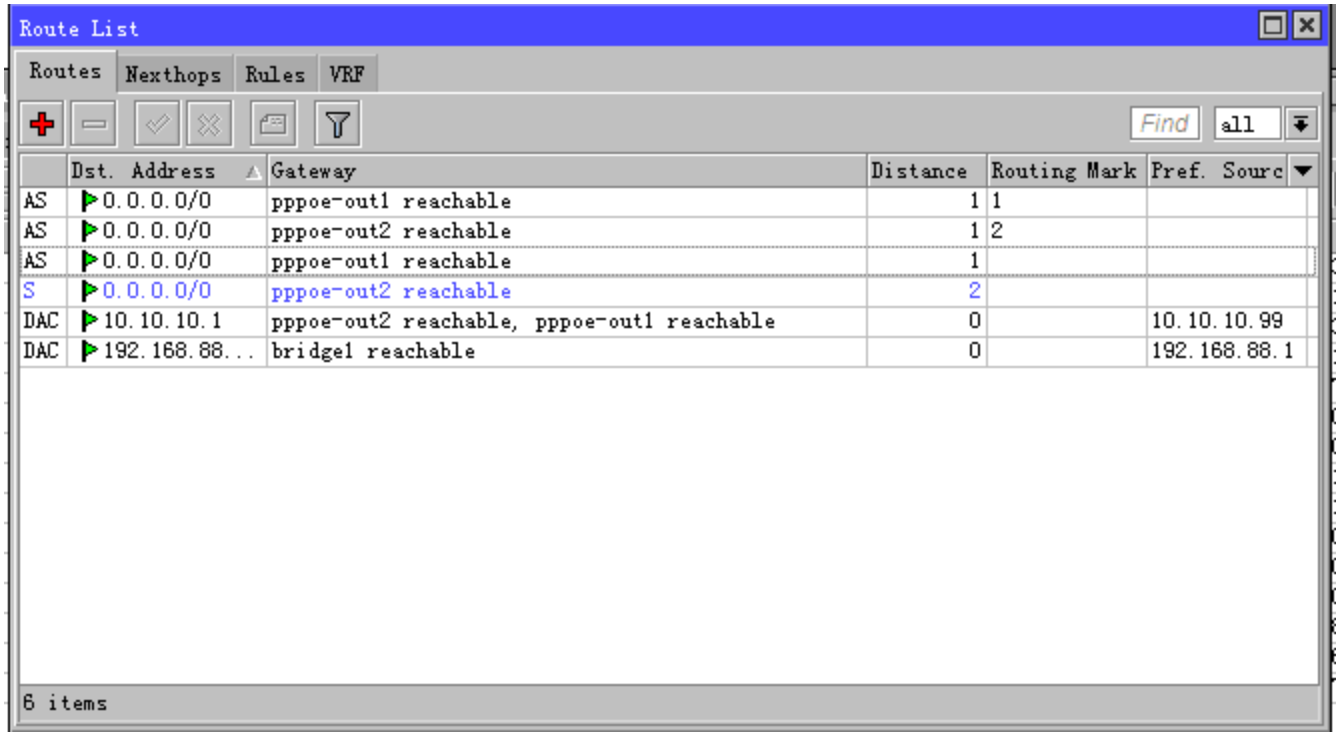


route

备用网关


Route <0.0.0.0/0>
General
Attributes
OK
Cancel
Apply
Disable
Comment
Copy
Remove
Dst. Address: 0.0.0.0/0
Gateway: pppoe-out2
reachable
Check Gateway: ping
Type: unicast
Distance: 2
Scope: 30
Target Scope: 10
Routing Mark:
Pref. Source:
enabled
active
static

route


Route List
Routes
Nexthops
Rules
VRF
Find
all
Dst. Address
Gateway
Distance
Routing Mark
Pref. Sourc
AS 0.0.0.0/0 pppoe-out1 reachable 1 1
AS 0.0.0.0/0 pppoe-out2 reachable 1 2
AS 0.0.0.0/0 pppoe-out1 reachable 1
S 0.0.0.0/0 pppoe-out2 reachable 2
DAC 10.10.10.1 pppoe-out2 reachable, pppoe-out1 reachable 0 10.10.10.99
DAC 192.168.88... bridge1 reachable 0 192.168.88.1
6 items

route

```
/ip route
add disabled=no distance=1 dst-address=0.0.0.0/0 gateway=pppoe-
    out1 \
   routing-mark=1 scope=30 target-scope=10
add disabled=no distance=1 dst-address=0.0.0.0/0 gateway=pppoe-
    out2 \
   routing-mark=2 scope=30 target-scope=10
add check-gateway=ping disabled=no distance=1 dst-
    address=0.0.0.0/0 gateway=\
   pppoe-out1 scope=30 target-scope=10
add check-gateway=ping disabled=no distance=2 dst-
    address=0.0.0.0/0 gateway=\
   pppoe-out2 scope=30 target-scope=10
```

PCC 负载均衡

实际应用中可能遇到的情况：

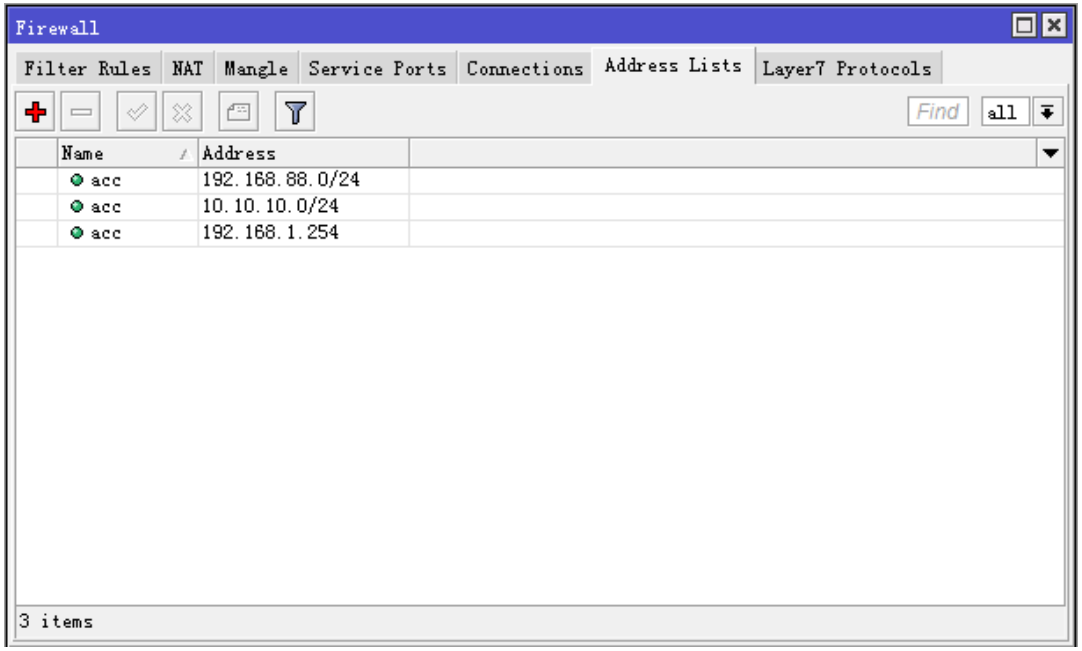
内网多网段或者有服务器需要相互访问（排除内网通讯）

外网带宽大小不同，进行PCC负载均衡（例如1条4M的带宽跟1条8M的带宽）

排除内网通讯

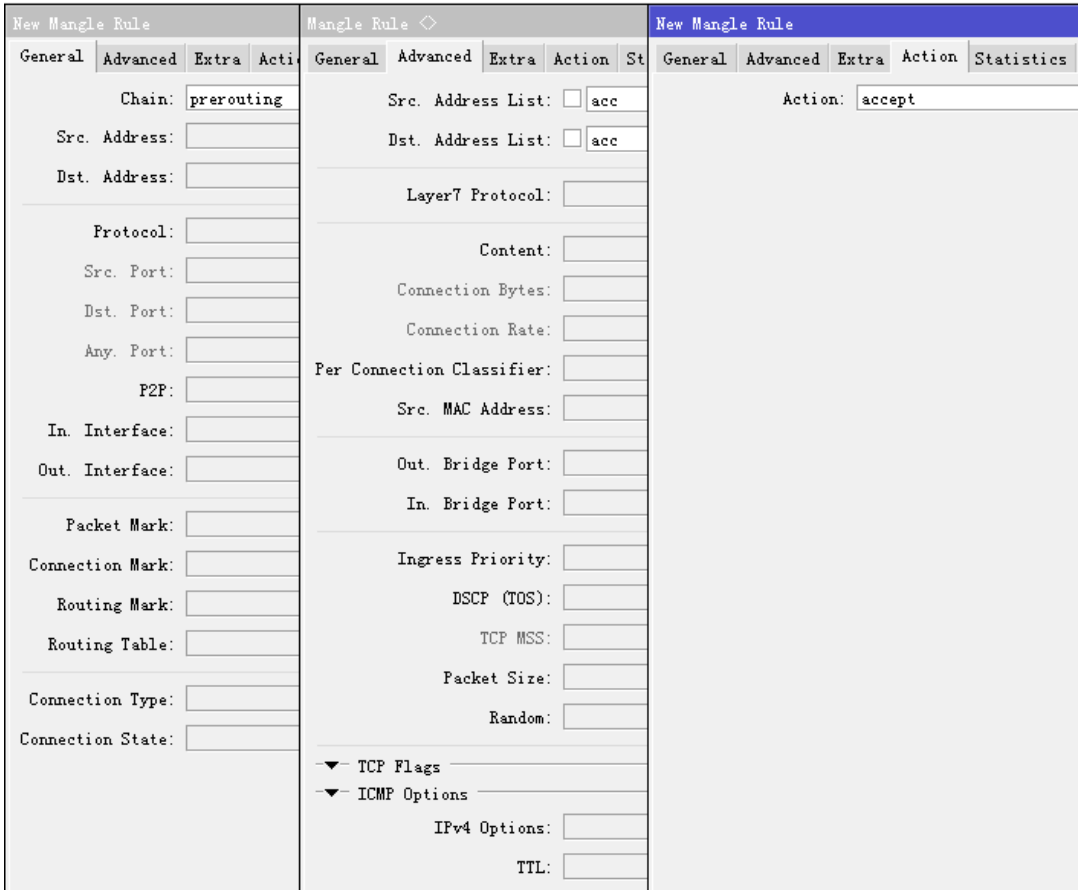
地址列表（address lists）

添加需要互访的IP或者网段



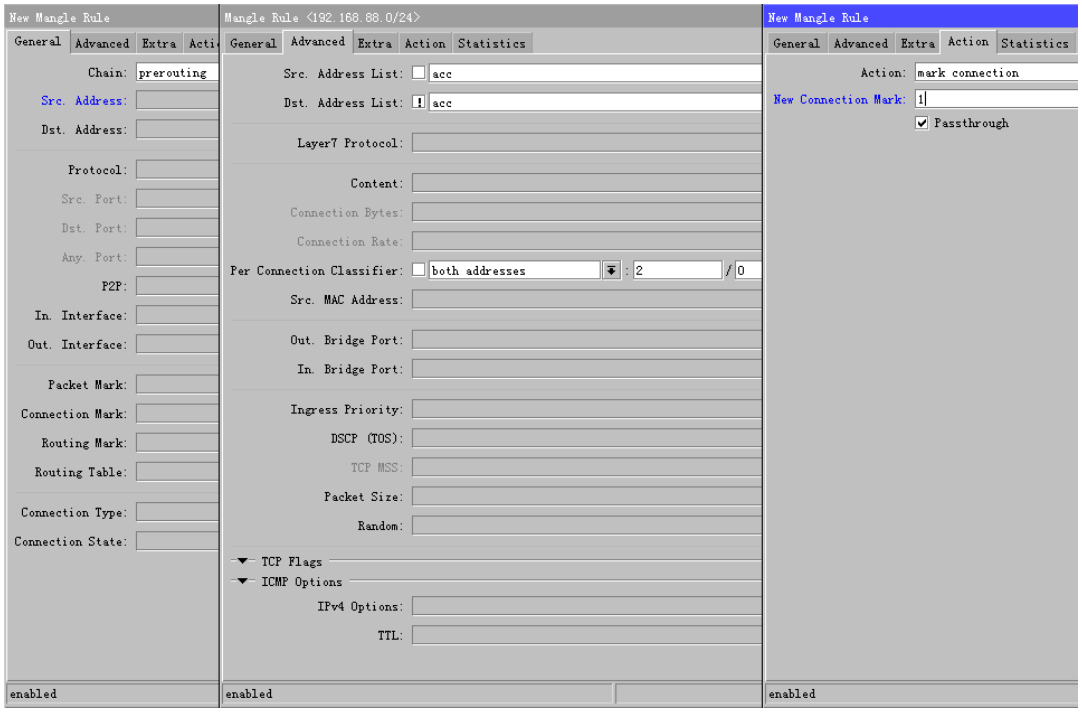
排除内网通讯

在PCC mangle之前先做个accept规则，让内网可以互访



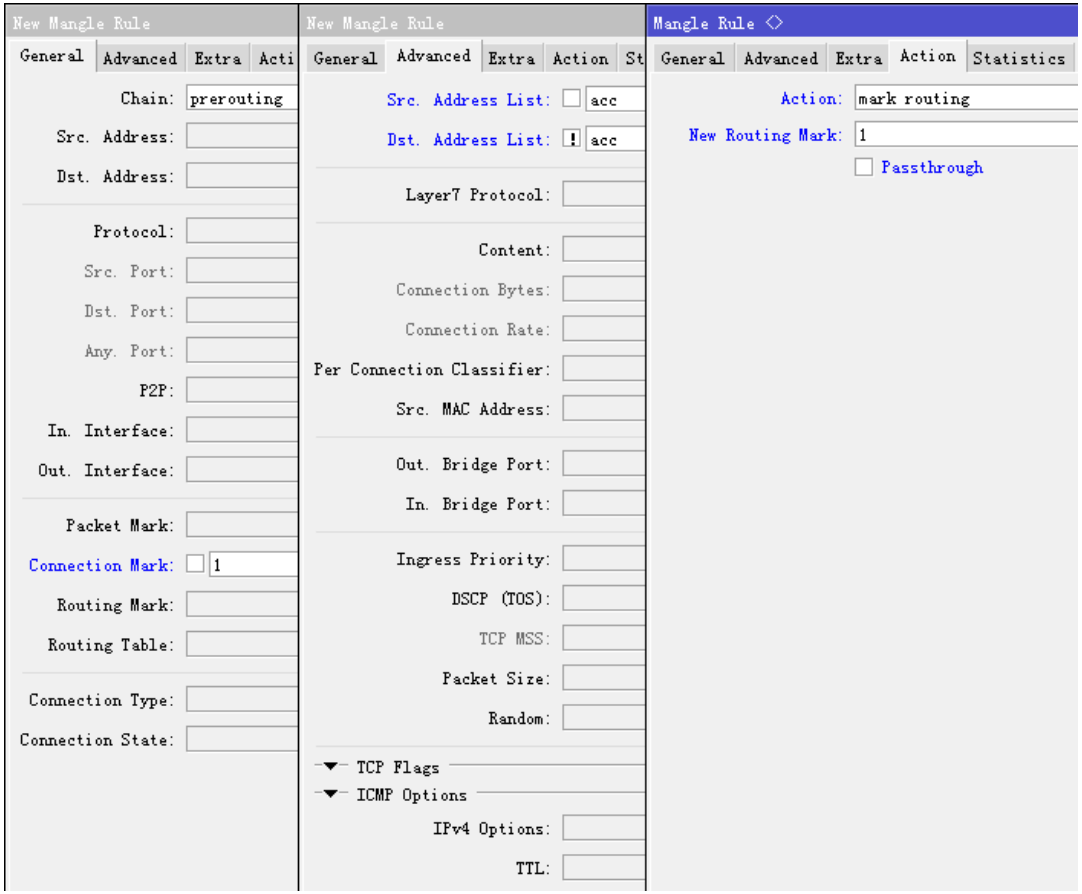
排除内网通讯

根据地址列表（address list）做PCC标记连接



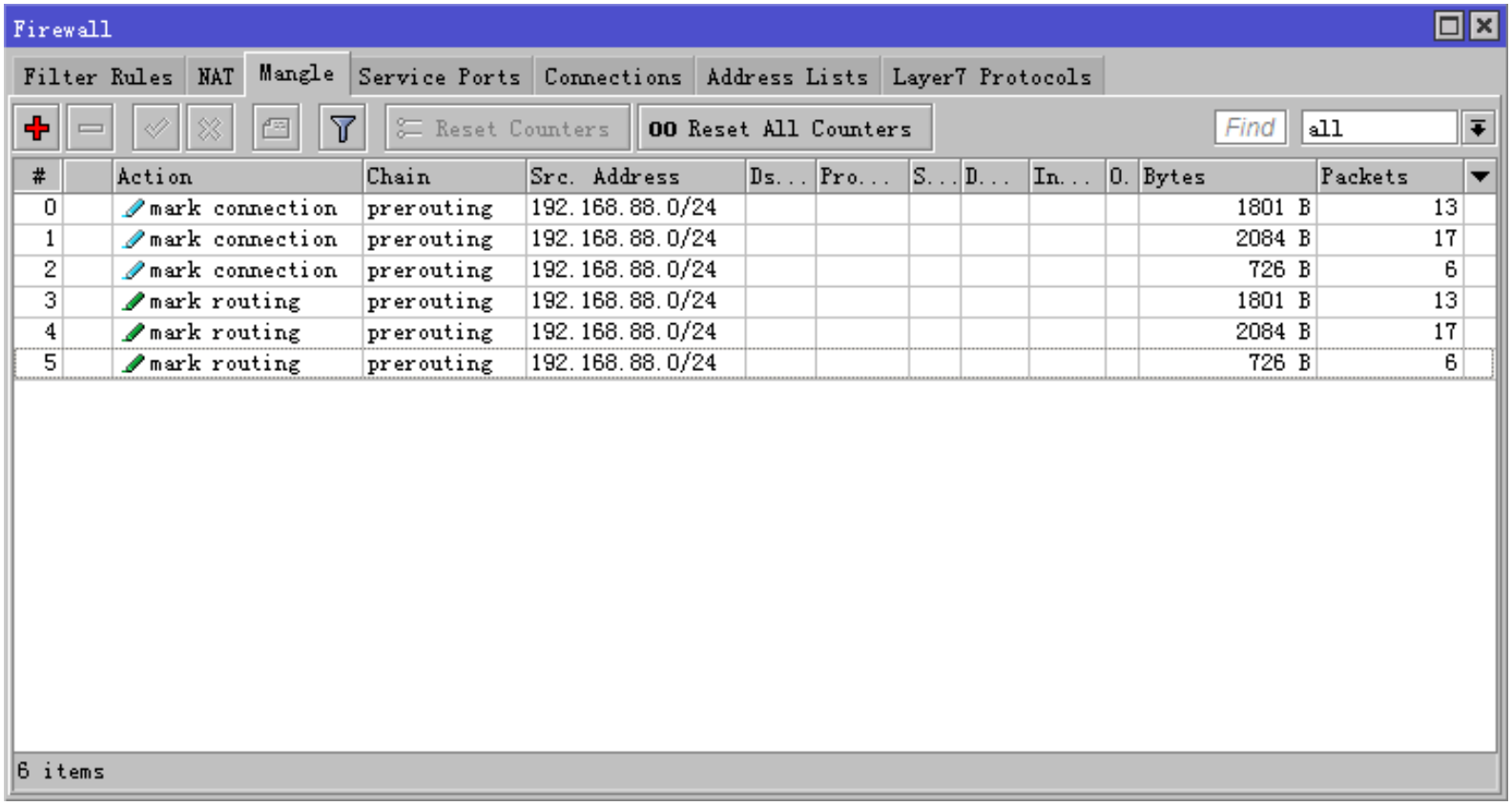
排除内网通讯

根据地址列表（address list）做PCC标记路由



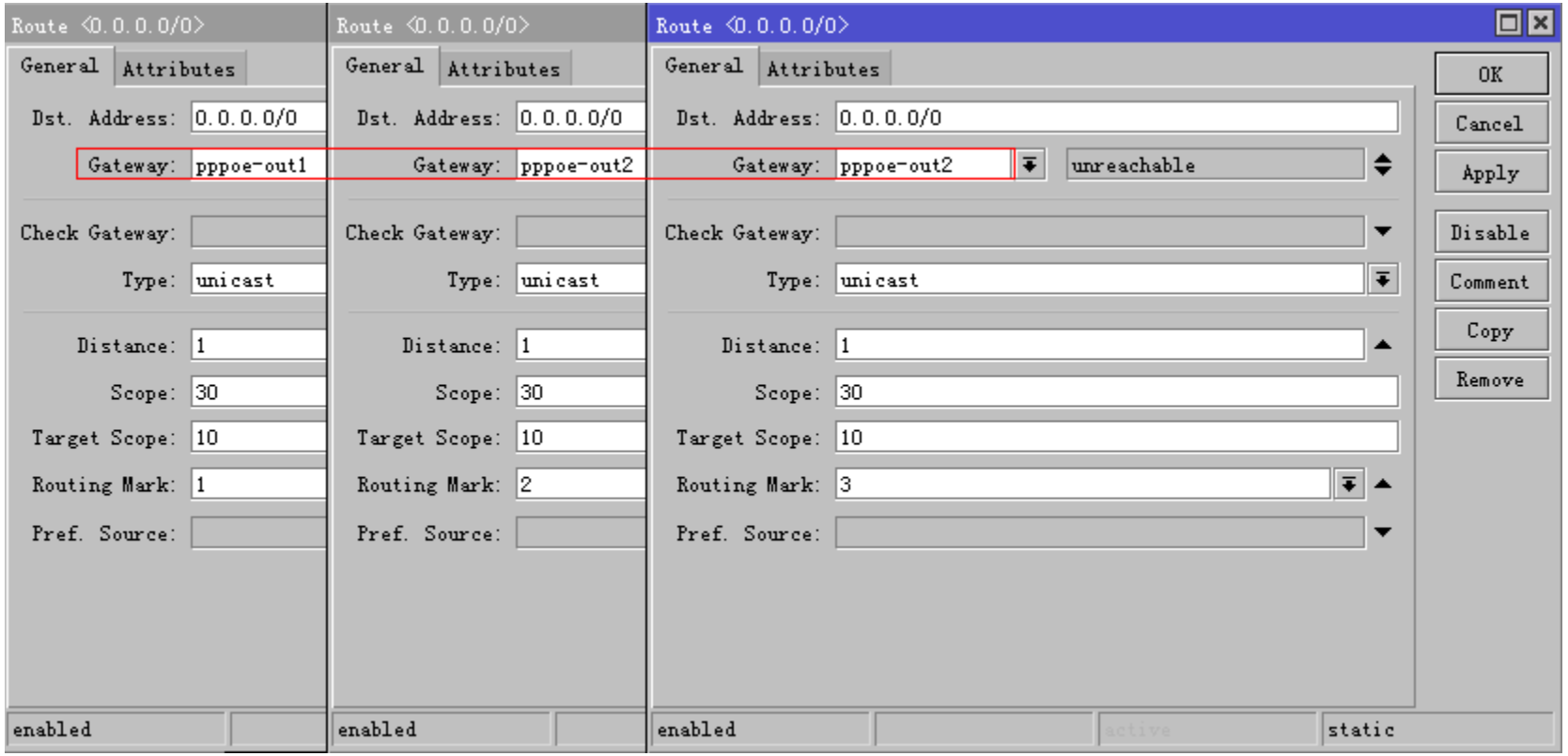
外网带宽：
4M+8M

做3个PCC标记



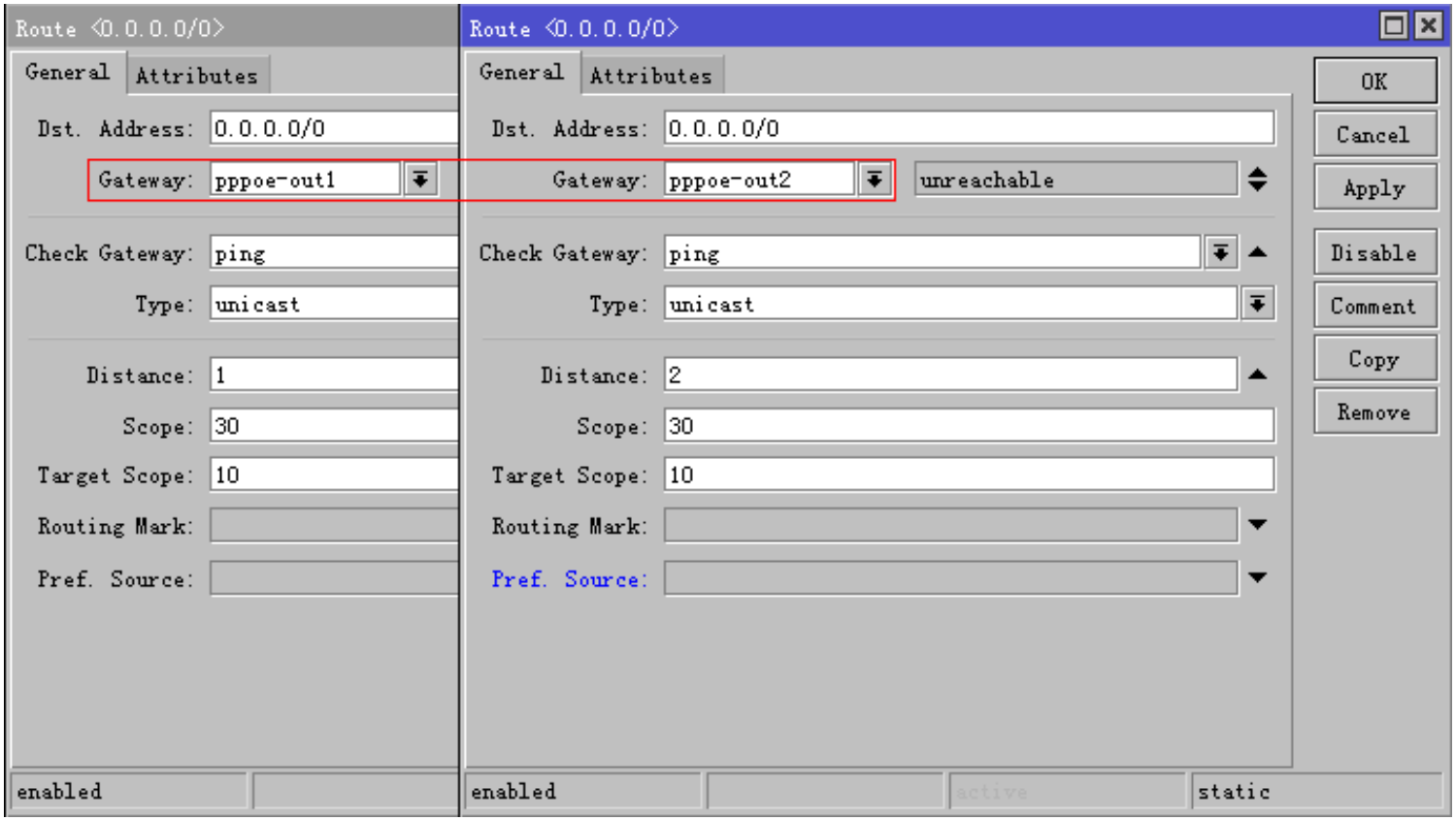
外网带宽：
4M+8M

路由标记



外网带宽：
4M+8M

默认路由



结束

EDCwifi 林利钢

深圳捷联讯通科技有限公司