

Lab — Policy-Based Routing, hands-on

Lesson 03 · ~15 min · Single host, uses ip netns + iptables mangle

Caution

Modifies kernel routing rules and iptables mangle table. Run on a throwaway VM if you can. The cleanup section at the end reverses every change.

Step 1 — Read your current rule list

```
ip rule show
# Default kernel rules:
# 0:    from all lookup local
# 32766: from all lookup main
# 32767: from all lookup default
```

Step 2 — Add a source-based rule + populate a new table

```
# Pretend traffic from 10.10.0.0/24 should use a different gateway.
sudo ip rule add pref 100 from 10.10.0.0/24 lookup 200
sudo ip route add default via 10.99.0.1 dev eth0 table 200

ip rule show          # confirm pref 100 now appears
ip route show table 200
```

Step 3 — Ask the kernel which table it would consult

```
ip route get 8.8.8.8 from 10.10.0.5
# Expected: shows table 200 selected, gateway 10.99.0.1
```

Step 4 — fwmark-based rule

```
# Add a fwmark rule (higher priority than the source-based one)
sudo ip rule add pref 90 fwmark 0x42 lookup 300
sudo ip route add default via 10.99.1.1 dev eth0 table 300

# Mark outbound SSH (port 22) with iptables
sudo iptables -t mangle -A OUTPUT -p tcp --dport 22 -j MARK --set-mark 0x42

# See the rule fire (no actual SSH needed — just check the kernel agrees)
ip route get 8.8.8.8 mark 0x42
# Expected: table 300 selected, via 10.99.1.1
```

Step 5 — Verify with conntrack (no save/restore yet)

```
# Without --save-mark, only the first packet of a NEW connection gets marked by
# the MARK rule. Subsequent packets in the same flow won't reliably hit it.
conntrack -L 2>/dev/null | head
```

Step 6 — Sticky marks with CONNMARK

```
# The canonical "mark once, persist" pattern.
sudo iptables -t mangle -A PREROUTING -m conntrack --ctstate NEW -j MARK --set-mark 0x42
sudo iptables -t mangle -A PREROUTING -j CONNMARK --save-mark
sudo iptables -t mangle -A PREROUTING -j CONNMARK --restore-mark

# Now: every packet in the flow inherits the mark from conntrack.
# Open a long-lived TCP connection between two namespaces, then:
conntrack -L | grep mark
```

Step 7 — Cleanup

```
sudo ip rule del pref 90 2>/dev/null
sudo ip rule del pref 100 2>/dev/null
sudo ip route flush table 200 2>/dev/null
sudo ip route flush table 300 2>/dev/null
sudo iptables -t mangle -F
```

Self-check

- ✓ You can list and add `ip rule` entries with custom priorities.
- ✓ You have populated a non-main routing table.
- ✓ You have used `ip route get ... from ...` to verify a policy match.
- ✓ You have wired iptables MARK to an `ip rule fwmark` entry.
- ✓ You have used the three-rule CONNMARK pattern (`--set-mark` + `--save-mark` + `--restore-mark`).