Multipath Routing
NFWS 2015

uw@ocedo.com

Where

- Core routers

  - Dynamic routing protocols

  - L3 network based distribution

- Edge routers

  - Manual setup

  - Source IP bound to uplink

  - L3/L4 based distribution

  - Link monitoring

Current state

- IPv4 with multiple default routes:

```
ip route add default via 2.2.2.2 dev eth0
ip route append default via 3.3.3.3 dev eth1
```

- ~~No balancing~~, but Layer 2 failover detection via fib_detect_death()

Current state

- IPv4 multipath route:

```
ip route add default \
    nexthop via 2.2.2.2 weight 2 \
    nexthop via 3.3.3.3 weight 1
```

- Kernel config: IP_ROUTE_MULTIPATH

- No failover detection

- Allows weights

- „Broken" since routing cache removal

Current state

- IPv6 multipath routes:

  ```
  ip -6 route add default via fe80::1 dev eth0
  ip -6 route append default via fe80::2 dev eth1
                        OR
  ip -6 route add default \
     nexthop via fe80::1 dev eth0 \
     nexthop via fe80::2 dev eth1
  ```

- Nexthop route resolves to multiple routes

- Only static routes allowed, no autoconf

- L3+L4+flowlabel hash for persistence

IPv4 persistence

- Use a L3/L4 hash function like rt6_info_hash_nhsfn

- Patch from Peter Christensen to add RTA_MP_ALGO again

SNAT persistance

- IP policy rule (IP_MULTIPLE_TABLES):

  ```
  ip rule add from 3.3.3.3 lookup 101
  ```

- Need to run multipath_select() only once per flow

- Broken since reroute after SNAT removed 2006

  ```
  commit ee68cea2c26b7a8222f9020f54d22c6067011e8b
  ```

- Lookup conntrack within ip_route_input_slow() and replace fl4.saddr with SNAT IP

L4 policy rules

- Balance traffic based on L3 and L4

- Skip rules when link is unavailable

- Lookup „before" socket is bound

L4 policy rules

- Examples:

  - from 2.0.0.0/24 to 3.0.0.0/24 dev eth0

  - from 0.0.0.0/0 udp port 5060 dev eth3 sticky

  - from 0.0.0.0/0 tcp port 80 dev eth0

  - from 0.0.0.0/0 tcp port 80 dev eth1

  - default hash distribution

L4 policy rules

- How:

  - Extend multipath_select()

  - Extend ip policy rules

    - protocol/port already available in flowi struct

  - Netfilter routing table to set an outgoing interface hint