nftables tcp option mangling

Florian Westphal
4096R/AD5FF600 fw@strlen.de
80A9 20C5 B203 E069 F586
AE9F 7091 A8D9 AD5F F600

Red Hat

July 2017

background

manuel added tcp option matching to nftables

```
tcp option maxseg size 1460
tcp option sack0 left > 1000
```

phil added exist/missing style matching:

```
tcp option window exists
tcp option maxseg missing
```

want to add set (write) support

use cases

1. iptables TCPMSS target
2. iptables TCPOPTSTRIP target

first one has PoC implementation

problems (1)

- iptables TCPMSS doesn't increase mss, ever
  - nft poc has check for kind `TCPOPT_MSS` to validate sreg $<$ size
- iptables TCPMSS has `clamp-to-pmtu` option
  - does `dst_mtu` on skb dst and route lookup w. reverse tuple
  - q: do we want to support nft equivalent? how? note: only works with symmetric routes
  - not enough to add `rt pmtu`

problems (2)

- user syntax: `tcp option maxseg size set 1340` looks ugly
- add short version? `tcp option mss set 1340`?
- should we disallow set on kind and length? always generates corrupt packets otherwise

optstrip support

- iptables OPSTRSTRIP target to remove options (nop'd out)
- could overload `tcp option window kind set nop` to clear out entire length
- or add new netlink option? explicit syntax?
- `tcp option window delete`?