nftables from ingress

Pablo Neira Ayuso
<pablo@netfilter.org>

Netdev 1.1
February 2016
Sevilla, Spain

nftables?

- What is?
  - Network specific virtual machine on kernelspace
    - 32-bit/128-bit registers.
    - Simple bytecode verification.
  - Netlink frontend
  - 2-phase commit protocol
  - Better dynamic/incremental update support
  - Userspace library: libnftnl
  - nft command line tool
    - Interactive shell
    - Scripting

nftables from ingress?

- Yes, since Linux 4.2.
- Placed just after the tc ingress hook.
- Transparent access to existing features.
- Potential reusage of the existing Netfilter building blocks: conntrack, NAT, logging and userspace queueing (although not yet implemented).

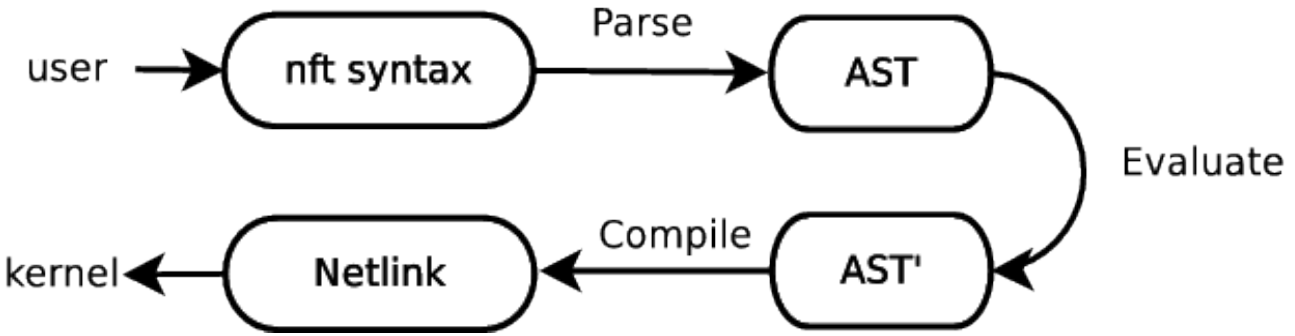
nftables

- **# nft --debug=netlink add rule netdev filter ingress **
  vlan id 1 ip saddr 10.0.0.0/23 counter

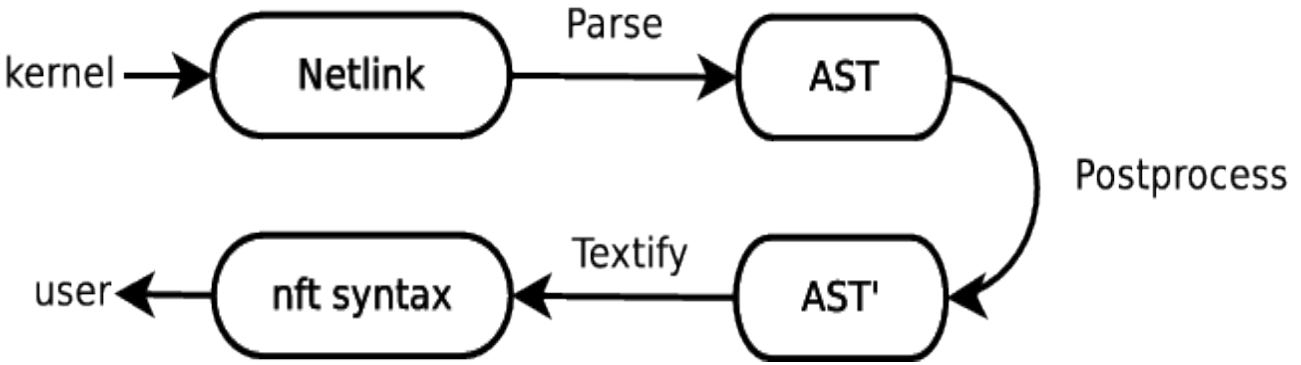
  netdev test-netdev ingress
  [meta load iiftype => reg 1]
  [cmp eq reg 1 0x00000001]
  [payload load 2b @ link header + 12 => reg 1]
  [cmp eq reg 1 0x00000081]
  [payload load 2b @ link header + 14 => reg 1]
  [bitwise reg 1 = (reg=1 & 0x0000ff0f) ^ 0x00000000]
  [cmp eq reg 1 0x00000100]
  [payload load 2b @ link header + 16 => reg 1]
  [cmp eq reg 1 0x00000008]
  [payload load 4b @ network header + 12 => reg 1]
  [bitwise reg 1 = (reg=1 & 0x00feffff) ^ 0x00000000]
  [cmp eq reg 1 0x0000000a]
  [counter pkts 0 bytes 0]

nftables

- From userspace to kernel:



Dump from kernel to userspace:



Tables, chains and rules

- nft add table netdev foo
- nft add chain netdev foo bar { \
  type filter hook ingress device eth0 priority 0\;
  }
- nft add rule netdev foo bar counter

Non-base chains

- nft add chain netdev foo blah
- nft add rule netdev foo bar counter jump blah

Rules: Expressions

- nft add rule netdev foo bar tcp dport != 80
- nft add rule netdev foo bar tcp dport 1-1024
- nft add rule netdev foo bar meta skuid 1000-1100
- nft add rule netdev foo bar meta length \> 1000
- nft add rule netdev foo bar ip daddr 192.168.10.0/24
- nft add rule netdev foo bar meta mark 0xffffffff/24
- nft add rule netdev foo bar meta mark and 0x0000ffff == 0x0000123
- nft add rule netdev foo bar meta mark set 0x0000321

Rules: Statements

- nft add rule netdev foo bar meta mark set 0x0000321
- nft add rule netdev foo bar ether daddr set be:ef:00:ca:fe:00

Sets and maps

- nft add rule netdev foo bar tcp dport { 22, 80, 443 } counter
- nft add set netdev foo whitelist { type ipv4_addr \; }

  nft add rule netdev foo bar ip daddr @whitelist \
  counter accept

  nft add element ip foo whitelist { \
  192.168.0.1, \
  192.168.0.10 \
  }
- nft add rule netdev foo bar dup to ip saddr map { \
  1.1.1.0/24 : “eth0” , \
  2.2.2.0/24 : “eth1” \
  }

Verdict maps

- nft add chain netdev foo tcp-chain
  nft add chain netdev foo udp-chain
  nft add chain netdev foo icmp-chain

- nft add rule netdev foo bar ip protocol vmap { \
  tcp : jump tcp-chain, \
  udp : jump udp-chain, \
  icmp : jump icmp-chain
  }

Sets timeouts

- nft add set netdev foo whitelist { \
  type ipv4_addr; \
  timeout 1h; \
  }
- nft add element netdev foo whitelist { \
  192.168.2.123, \
  192.168.2.124, \
  }
- nft add set netdev foo whitelist { \
  type ipv4_addr; flags timeout; \
  }
- nft add element netdev foo whitelist { 192.168.2.123 timeout 10s }

Comments

- nft add rule netdev foo bar \
  ip daddr 8.8.8.8 counter accept\
  comment \“google dns\”
- nft add set netdev foo dns-whitelist {\
  type ipv4_addr\;
  }
- nft add element netdev foo dns-whitelist { \
  8.8.8.8 comment “google dns”, \
  192.203.230.10 comment “nasa dns”,
  }

Concatenations

- nft add rule netdev foo bar \
  ether saddr . ip saddr . tcp dport { \
  c0:fe:00:c0:fe:00 . 192.168.1.123 . 80, \
  be:ef:00:be:ef:00 . 192.168.1.120 . 22} \
  counter accept
- nft add rule netdev foo bar ip saddr . tcp dport vmap { \
  192.168.1.123 . 22 : jump whitelist, \
  192.168.1.123 . 80 : jump whitelist, \
  }

Concatenations (2)

- nft add set netdev foo bar { \
  type ether_addr . ipv4_addr \; }
- nft add element netdev foo bar { \
  00:ca:fe:00:be:ef . 192.168.1.123, \
  00:ab:cd:ef:00:12 . 192.168.1.124 \
  }

Statements

- nft add rule netdev foo bar \
  rate 10 mbytes/second burst 9000 kbytes
  accept
- nft add rule netdev foo bar \
  limit rate 10/second counter accept
- nft add rule netdev foo bar icmp type echo-request \
  limit rate over 10 mbytes/second counter drop
- nft add rule netdev foo bar dup to eth1
- nft add rule netdev foo bar fwd to vethSEF72

Restoring ruleset

- echo “nft flush ruleset” > ruleset.nft
- nft list ruleset > ruleset.nft
- nft -f ruleset.nft

Monitoring updates

- nft monitor
- nft monitor new rules

Scripting

```
#!/usr/sbin/nft

include "another-ruleset.nft"


#
# Allowed NTP servers
#
define ntp_servers = { 84.77.40.132, 176.31.53.99, 81.19.96.148,
138.100.62.8 }

add rule netdev foo bar ip saddr $ntp_servers udp dport 123 counter
```

Learn more and help us

- Grab the code
  - Kernel: http://www.kernel.org
  - Library: git://git.netfilter.org/libnftnl
  - User-space: git://git.netfilter.org/nftables
- Documentation
  - http://wiki.nftables.org
  - man nft
- Report bugs:
  - https://bugzilla.netfilter.org

nftables from ingress

Pablo Neira Ayuso
<pablo@netfilter.org>

Netdev 1.1
February 2016
Sevilla, Spain