netfilter hw offloads

NFWS 2017 - Faro, Portugal

Pablo Neira Ayuso <pablo@netfilter.org>

Conntrack offload

- Software conntrack is doing much more than probably hardware can do...
  - TCAM hardware just have a table with tuple
    - [ip src, ip dst, ip protocol, sport, dport]
    - Aging through
  - Smart NICs can probably do more than this, still probably just a subset of the features.
- Offload need to be configure from user context.
- A 1:1 mapping may give the wrong impression to the user.

Flow offload

- Idea: Populate nft flow table based in matching criteria.
  - We can limit the size of the flows that fit in.
  - Configurability: We can select what flows are offloaded.
  - … Flow table = Just a map with pointer to conntrack objects.
- Conntrack entries in that flow table are owned by HW
  - It has to release this resource
- Kernel thread to walk over the flow table, configure the hardware from user context.

Flow offload (2)

- We still have to keep conntrack entries in table:
  - NAT needs this for port allocation logic
  - Set on a flag to indicate that the entry is owned by HW
- Races:
  - For little time packets may go through software until hardware is configured.

Flow offload (3)

- Flushing flow table: Flow goes back to conntrack.
  - Set pickup flag in flow to infer state from the middle of connection.
  - Unset HW stolen flag.

nftables offload

- Resurrect patches I submitted to mailing list one year ago:
  - Software Intermediate representation (IR) for hardware
  - this time only for conntrack.
- Transform nftables internal representation to tree.
- Walk over that tree to populate hw IR structure.
- Chain flag to ask for nft to offload ruleset
- nf_tables VM description to userspace:
  - Express HW limitations **through** nft abstraction.