nf_tables sets overview

NFWS 2017 Faro, Portugal

Pablo Neira Ayuso <pablo@netfilter.org>

nf_tables set overview

- Selects backend based on description
  - Number of elements (if known)
  - Key length
  - Intervals
- Sets come with big O notation to indicate scalability
  - lookup
  - space
- User doesn't need to know need to learn about datastructures and play tuning games
- Two policies:
  - Performance, select the faster implementation (default behaviour)
  - Memory, selects the one that consumes less memory

nf_tables set overview (2)

- Existing set backend implementations
  - Hashtable
    - Two variants: fixed size and resizable
    - With timeout implementation.
  - Bitmap, up to 16 bit keys
    - 64 bytes for 8 bits.
    - 16 Kbytes for 16 bits.
  - Rbtree, for intervals
- Performance evaluation from nft ingress
  - one rule with anonymous, default policy drop

hashtable

- Resizable hashtable
  - With timeout support
  - 11076337pps, **5316Mb/sec**
- Fixed size hashtable
  - Selected if userspace indicates size:
    - Used for anonymous sets
    - User specifies 'size' statement in set definition
  - No timeout support, but could be done
  - 16-bit or 32-bit key: 13109944pps **6292Mb/sec**
  - Generic: 12670233pps **6081Mb/sec**

bitmap

- Keeps a list of existing dummy objects
  - Keeps element comments, only used for dumping
  - Increases memory consumption
- May add timeouts
- From lookup path, uses bitmap representation
  - Two bits to represent current and next/previous generation
- 16-bit key: 16755207pps **8042Mb/sec**

rb-tree

- For set intervals
  - Userspace expands interval in two elements, one with the end interval flag set on
  - Central rw-spinlock
  - No timeout support
  - Deprecate this: Replace it with rcu tree?
    - RCU Bonsai?

Discussion

- 16-bit key sets, with performance policy:
  - Number of elements known:
    - >= 380 elements, selects bitmap (faster)
    - <380 selects hashtable (slower)
  - Unknown number of elements:
    - Selects bitmap
- New set implementations?
- Select set backend from userspace?
  - Expose sets available via nft VM description
    - Still allows to deprecate set implementations
  - Userspace selects the set time
- Add more set decorations, eg. percpu.