TC Classification with Open vSwitch

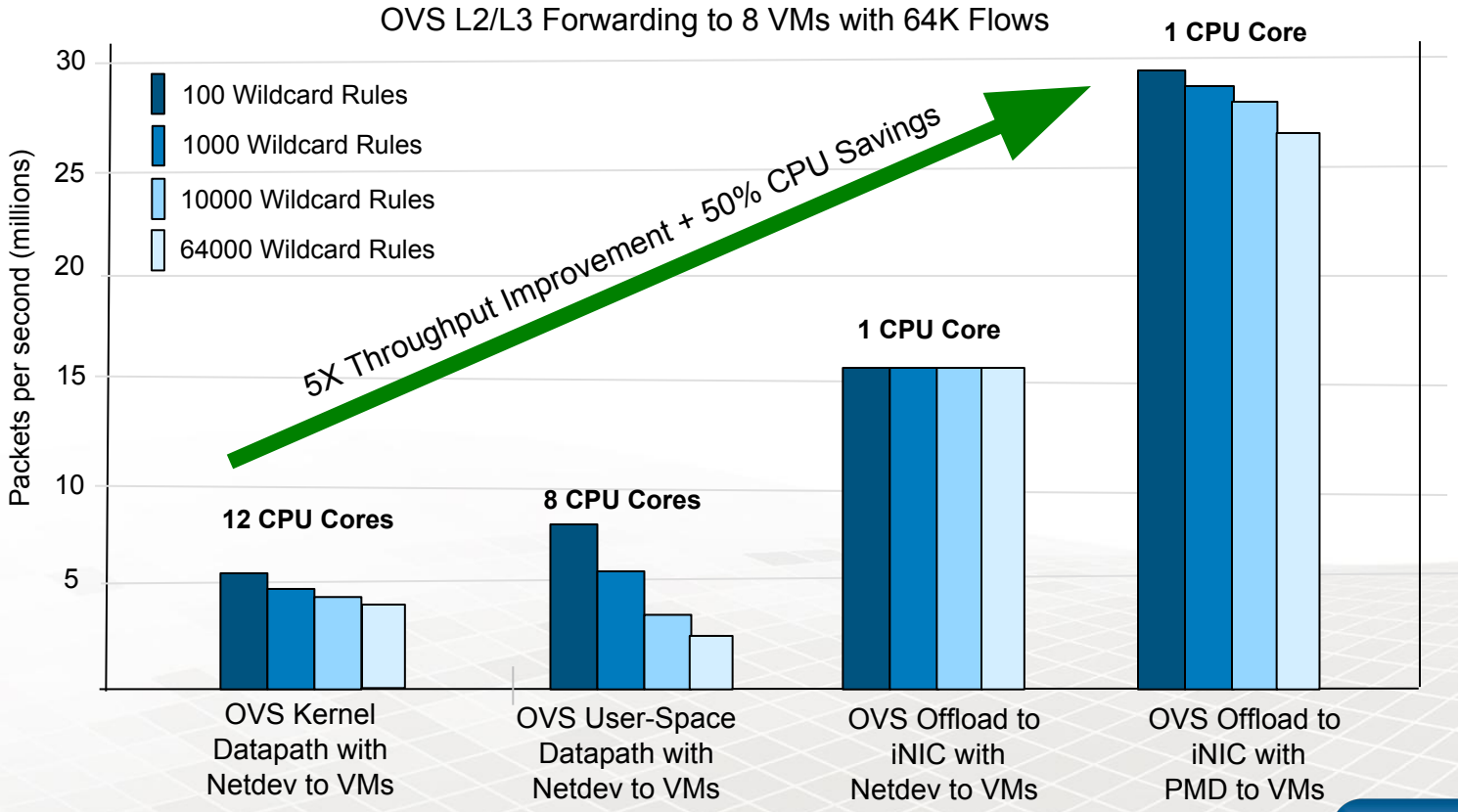
Simon Horman

Motivation

Would like to:

- Partially or fully offload Open vSwitch
  - Software: e.g. TC
  - Hardware: e.g. iNIC
- Do so using mechanisms present in upstream

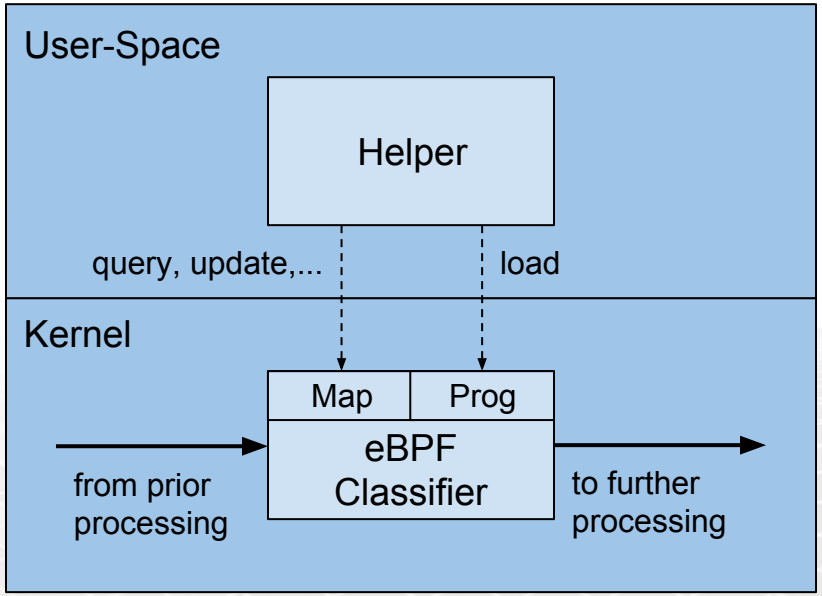
Importance of Offloading



NETRONOME

eBPF TC Classifier and Actions

- eBPF programs supplied to kernel by user-space
- Data may be shared with user-space using eBPF maps



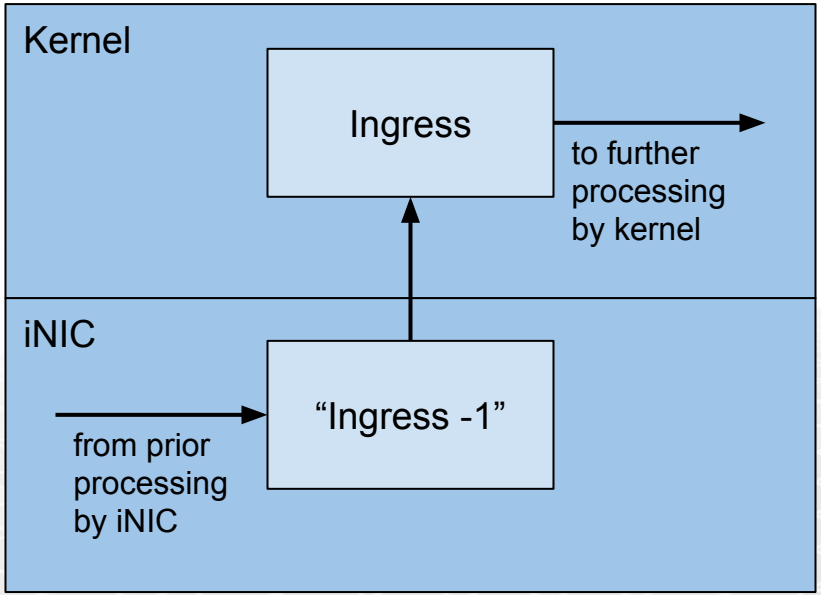
eBPF Offload

- eBPF seems well suited to offloading to programmable hardware
  - May be interpreted or;
  - JITed and run natively
  - Mechanism for very fast and flexible packet handling

NETRONOME

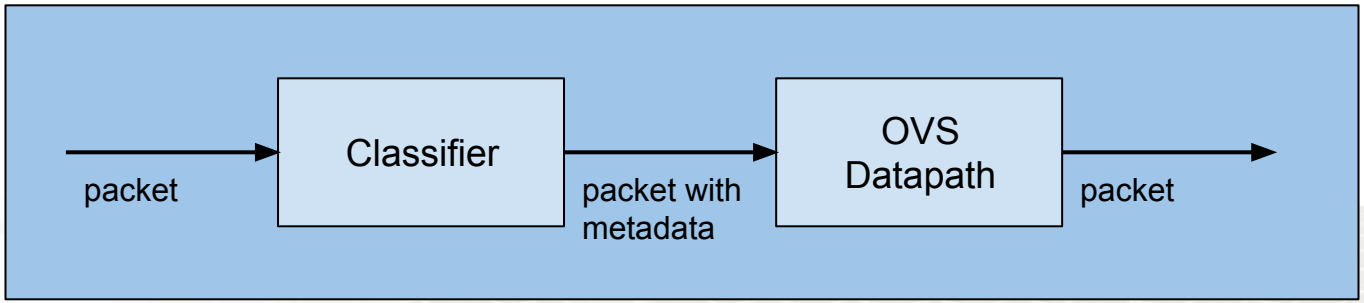
TC Offload Possibility

- One possibility is to add “ingress -1” support to TC
  - TC qdisc may be added before Ingress
  - Managed by Kernel; Executed in Hardware



Hinting

- Mechanism for offloading Open vSwitch classifier
- Offload classifies packet and tags it with metadata
- Open vSwitch performs flow lookup using metadata

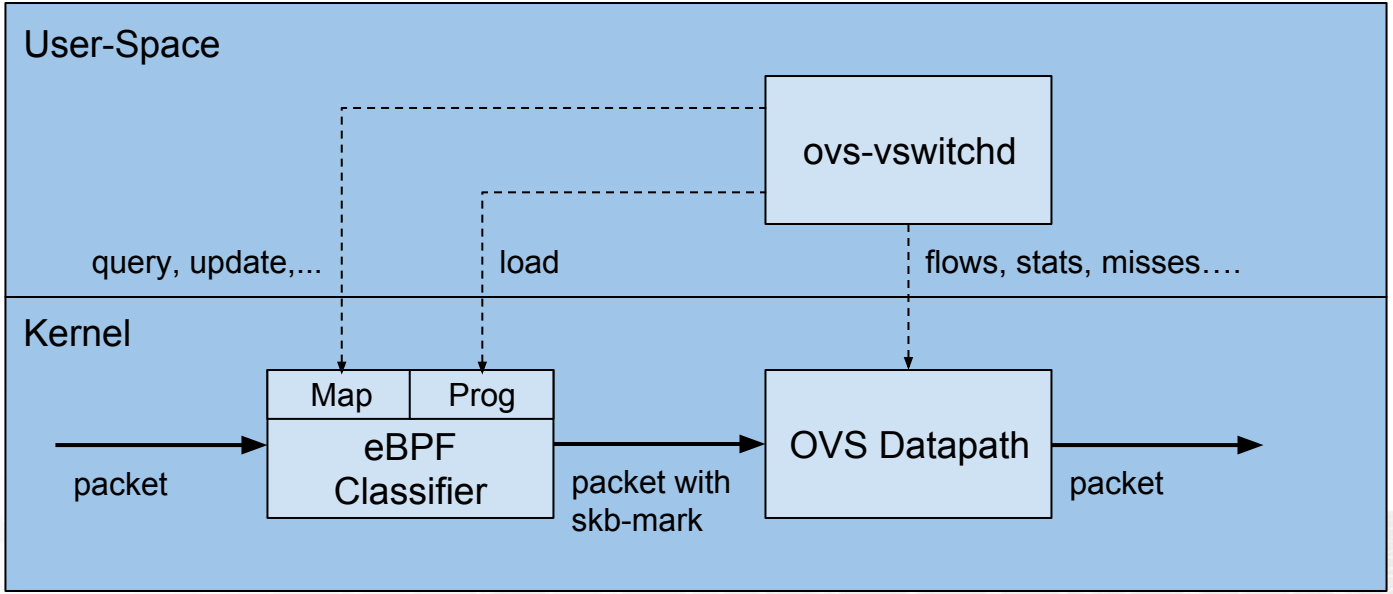


NETRONOME

Modest Proposal for Hinting

- TC Classification → metadata added as skb mark
- Open vSwitch looks up flows using skb mark
- eBPF TC Classifier allows eBPF maps to synchronise flow/mark mapping between TC classifier and Open vSwitch
- And the eBPF may be offloaded to hardware

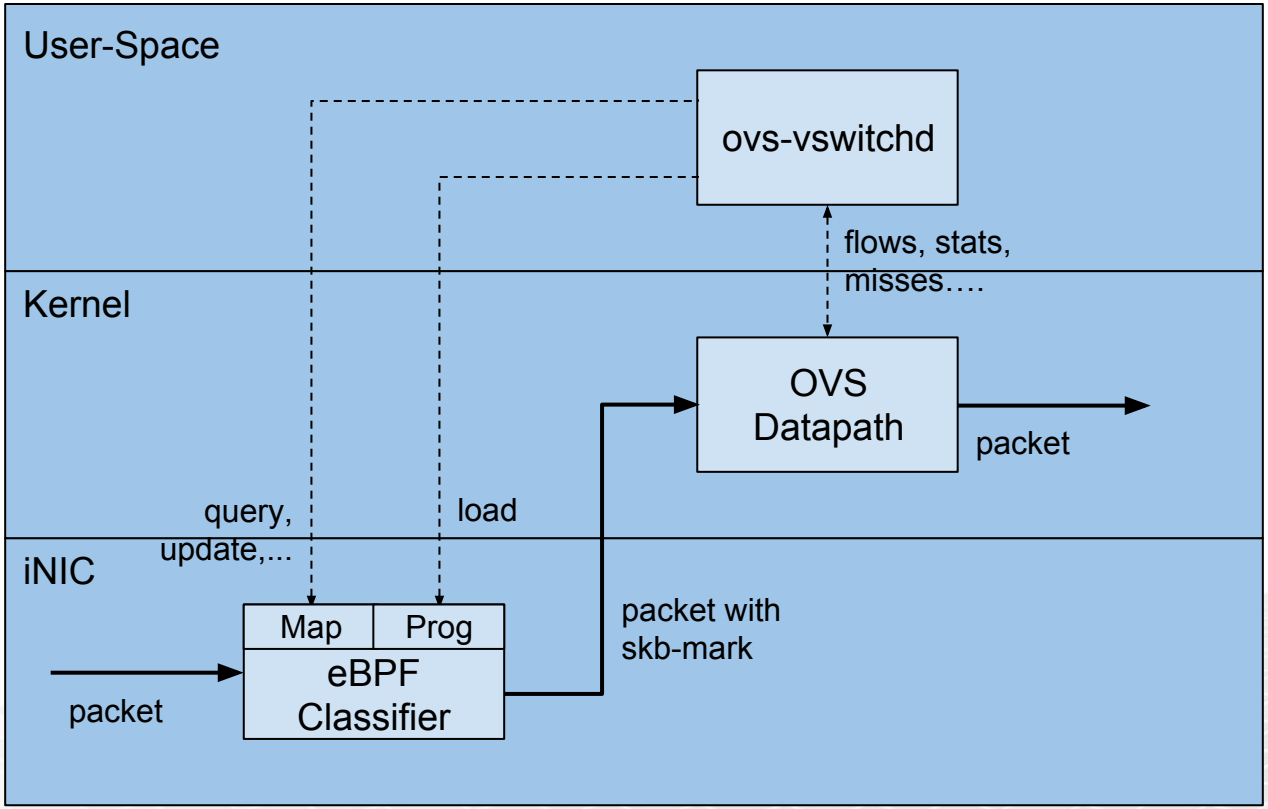
Hinting with eBPF Classifier



Hinting Packet Processing Path

NETRONOME

Offloaded Hinting



NETRONOME

Research at Netronome

Investigating:

- Integration of TC classifier with eBPF
- Hardware offload of Kernel OVS datapath using switchdev
- User-space driven hardware offload of OVS using match-interface
- ...

NETRONOME

Questions

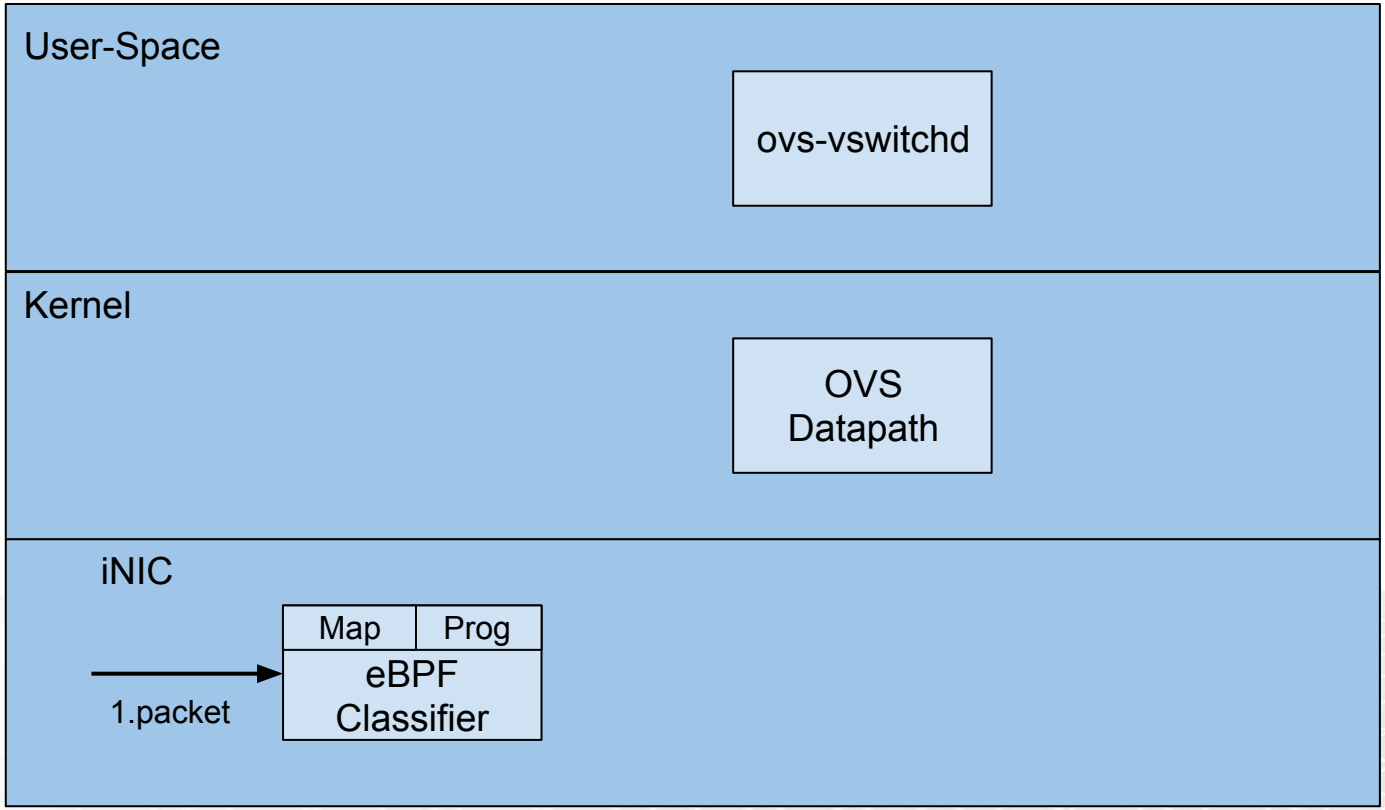
Thanks to the community for many of the ideas presented here and thanks for your time

Bonus Slides

Possible Flow Handling

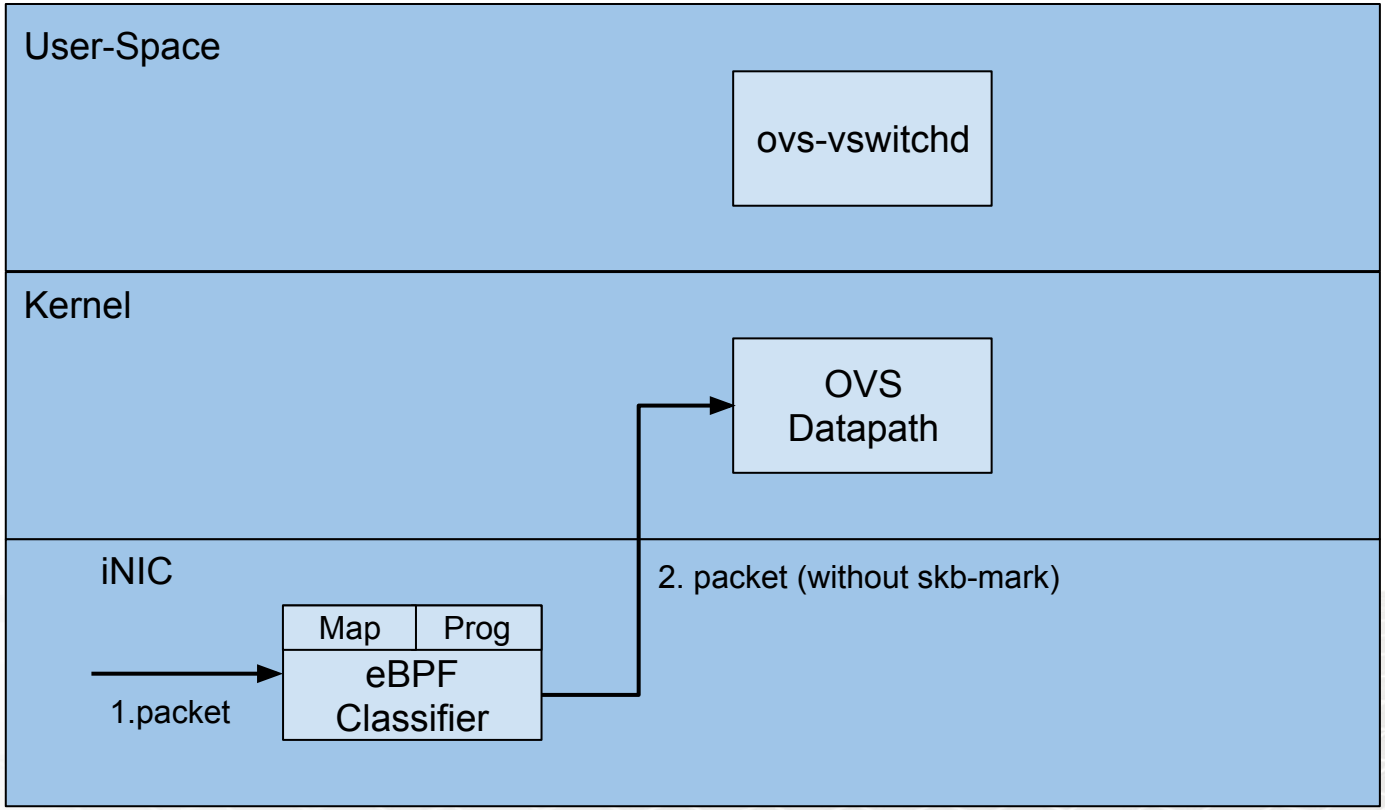
- Makes use of skb-mark for hinting
- Flow handling should work with and without hw offload

Possible Flow-Miss Handling

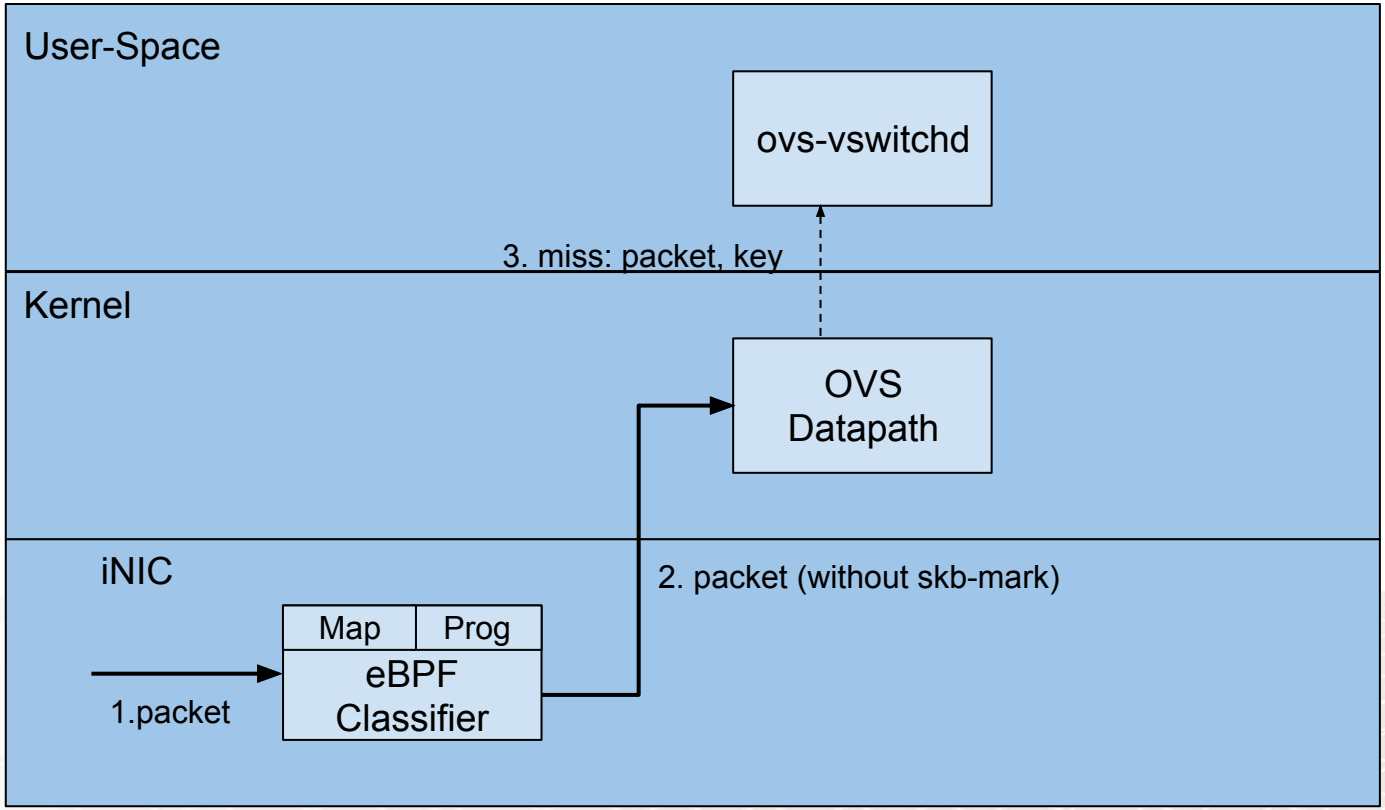


NETRONOME

Possible Flow-Miss Handling

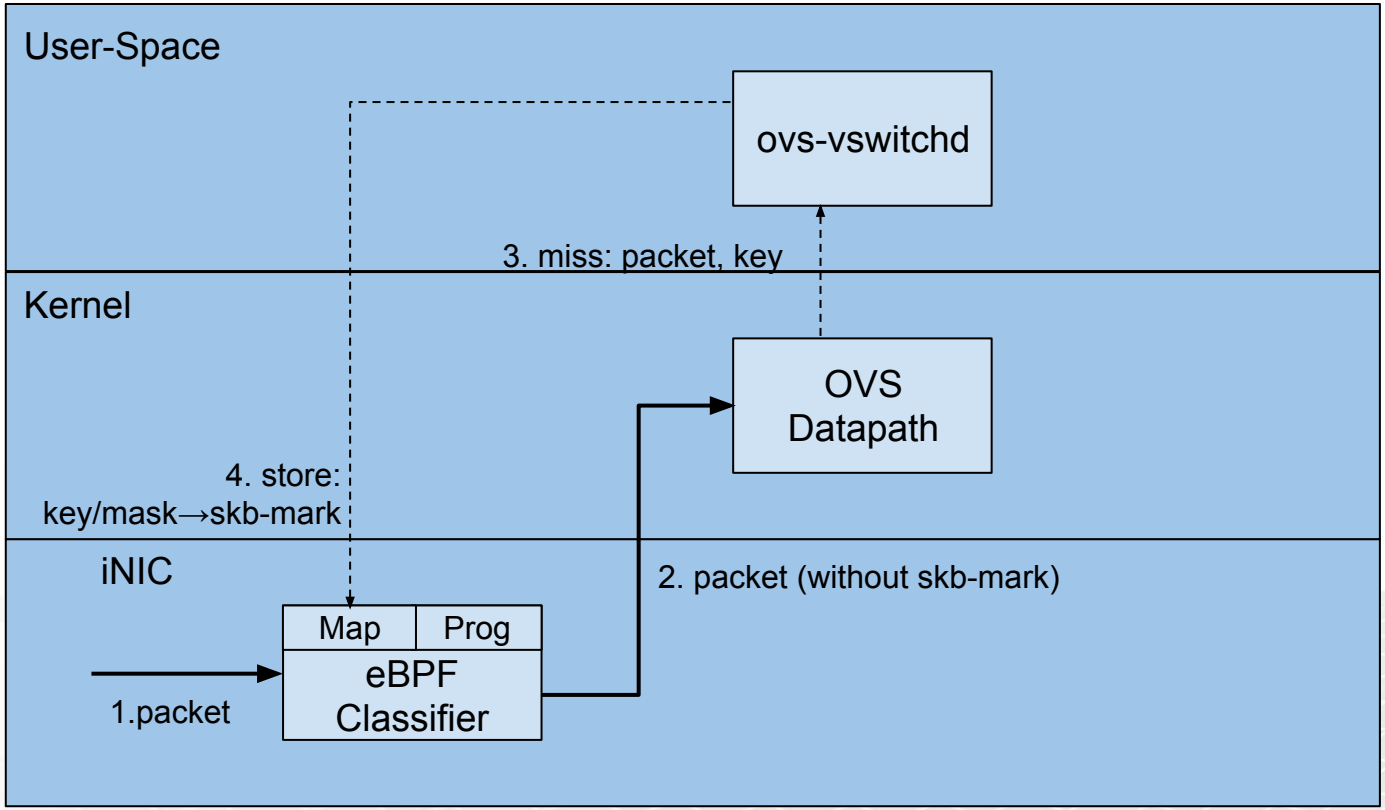


Possible Flow-Miss Handling



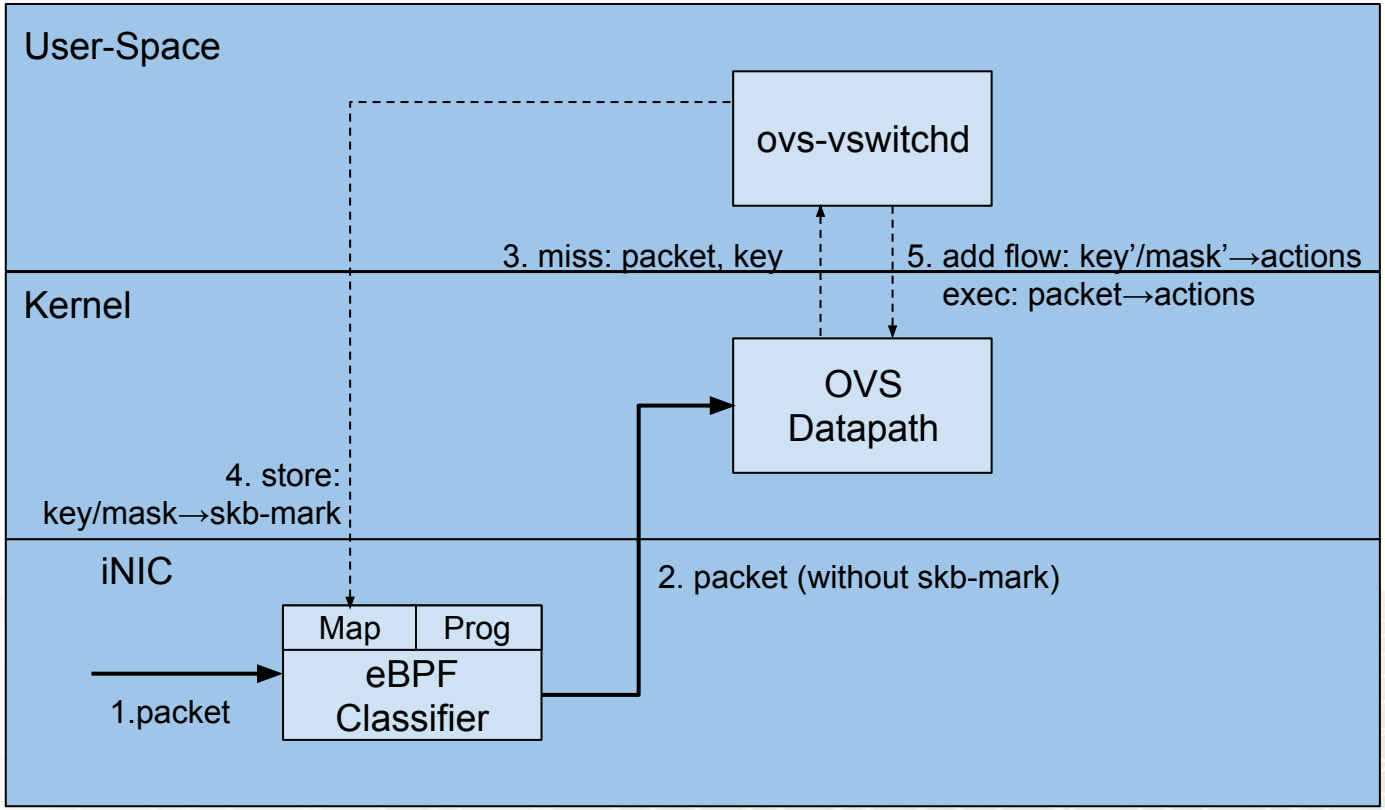
NETRONOME

Possible Flow-Miss Handling



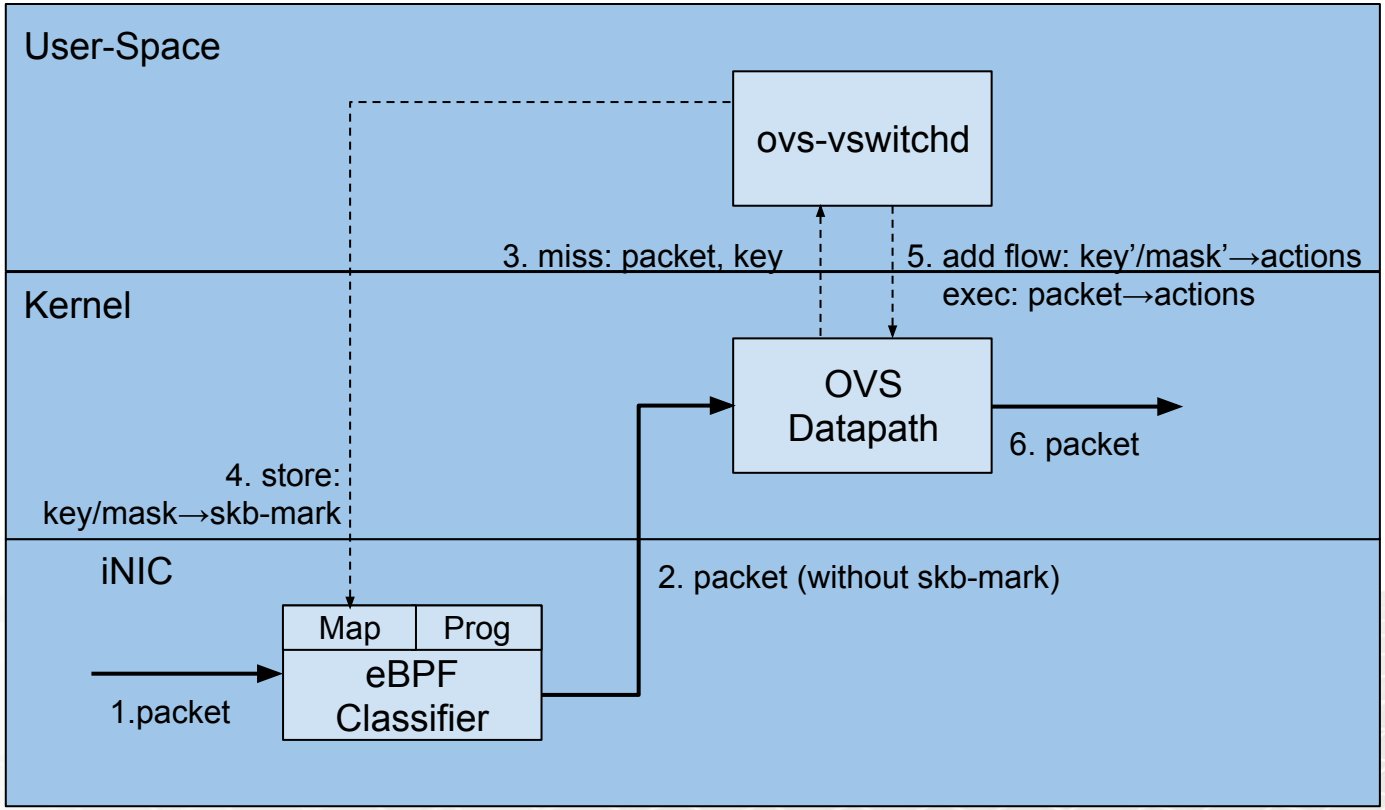
NETRONOME

Possible Flow-Miss Handling



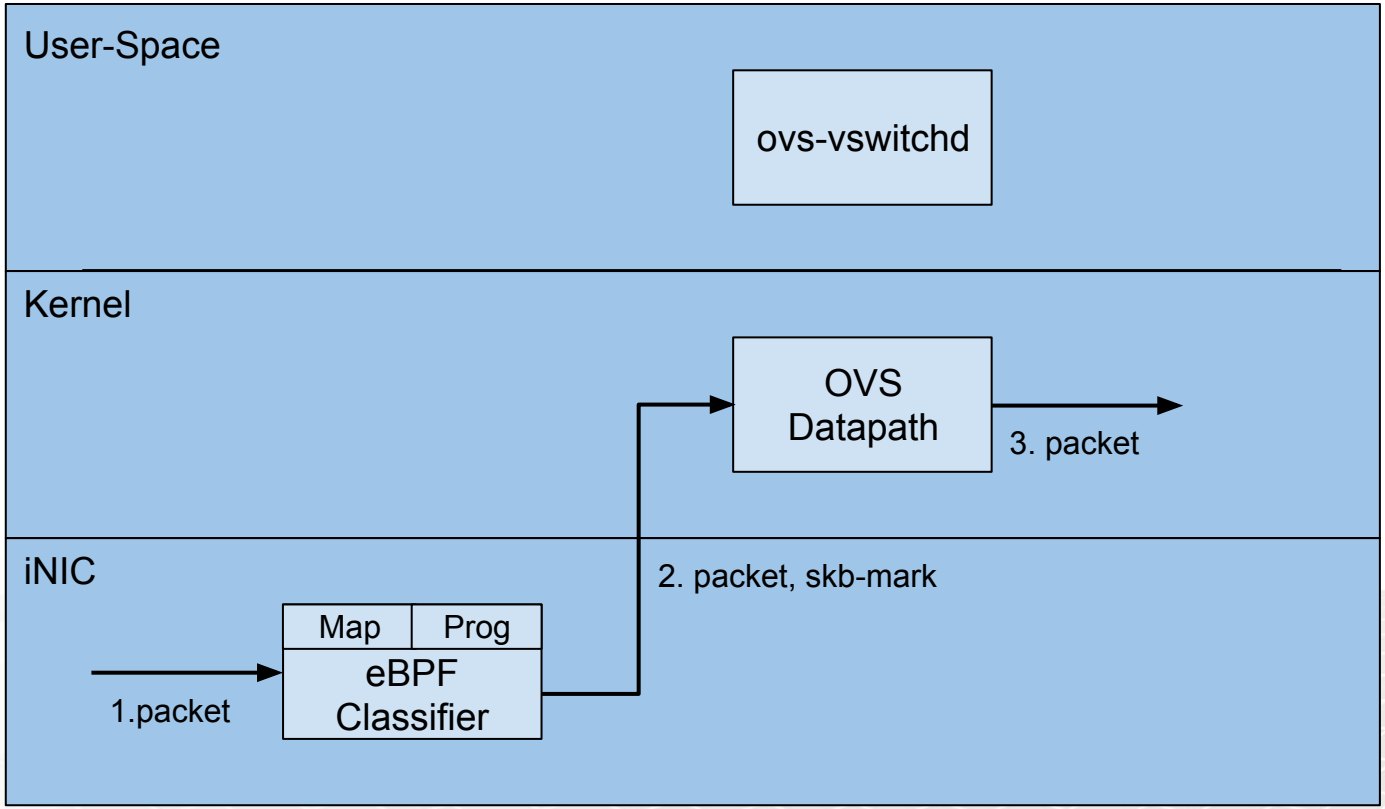
NETRONOME

Possible Flow-Miss Handling



NETRONOME

Possible Flow-Hit Handling



NETRONOME