Linux Network Namespaces in Open vSwitch

Jiri Benc
Red Hat
November 2015

Network Namespaces

- Partitioning of Linux network stack
- Resources isolation
- Used heavily by containers, Open Stack, ...

Current State of Open vSwitch Support

- Interfaces in an OVS bridge may be moved to a different netns

```
ovs-vsctl add-port br0 eth0
ip link set eth0 netns otherns
```

  - But cannot be added from a different netns

- Weird behavior of some OVS tools

```
ovs-vsctl show
```

```
ovs-ofctl show br0
```

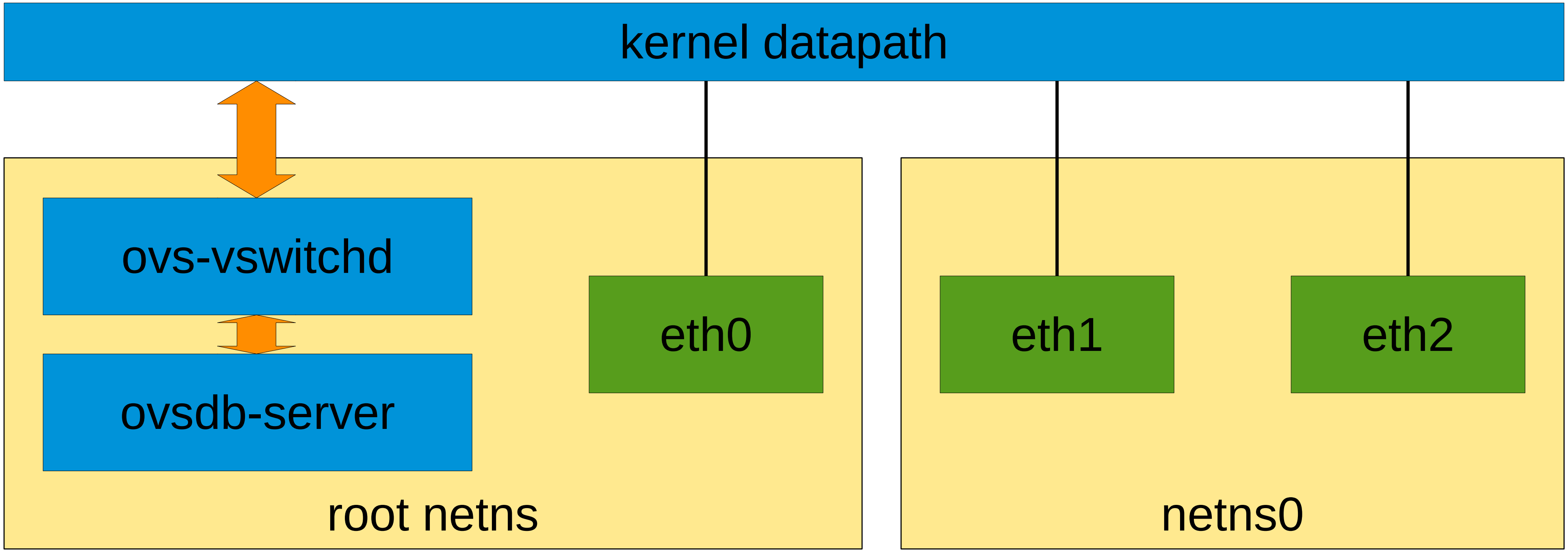
Kernel Datapath

- Isolation: skb_scrub_packet
- Recently added to ovs_vport_receive:

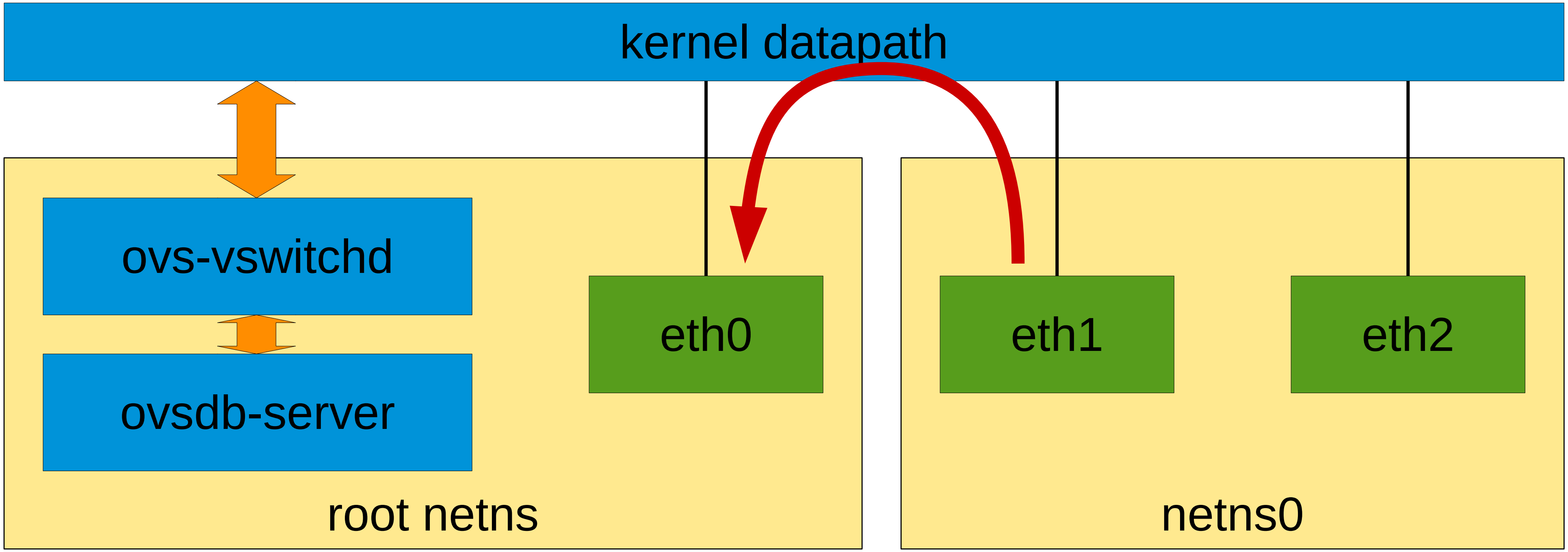
```
if (unlikely(dev_net(skb->dev) != ovs_dp_get_net(vport->dp)))
        skb_scrub_packet(skb, true);
```

- What is the netns of the datapath?

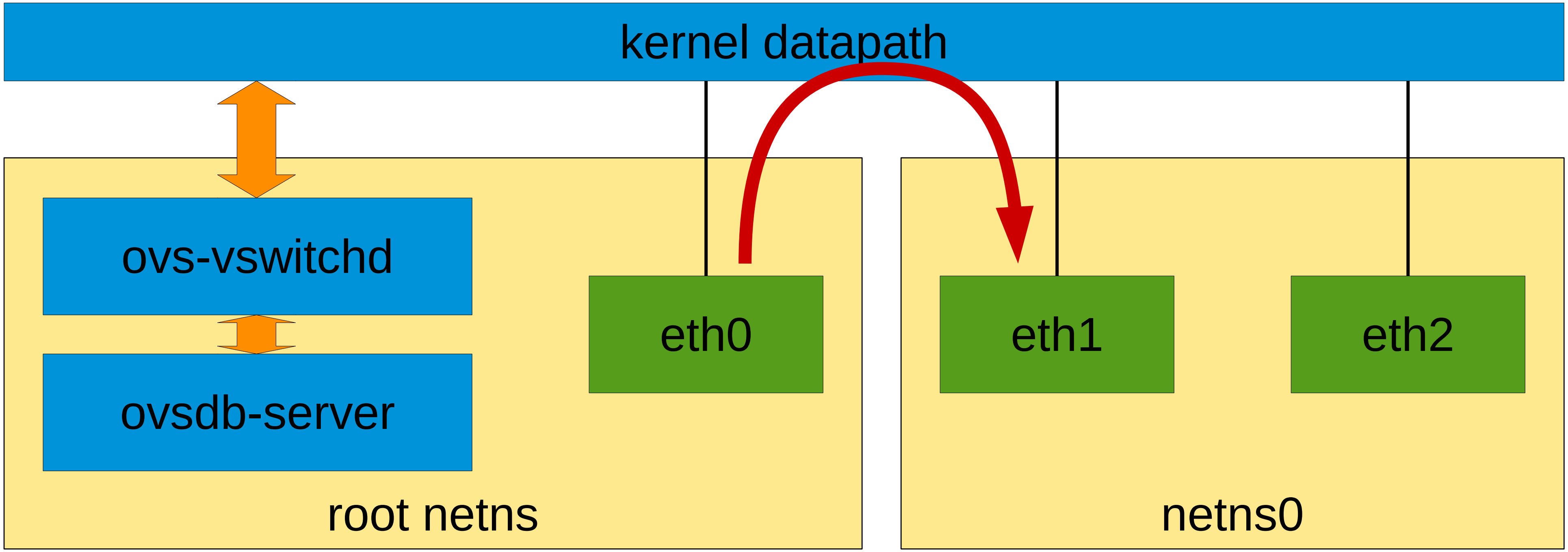
Kernel Datapath



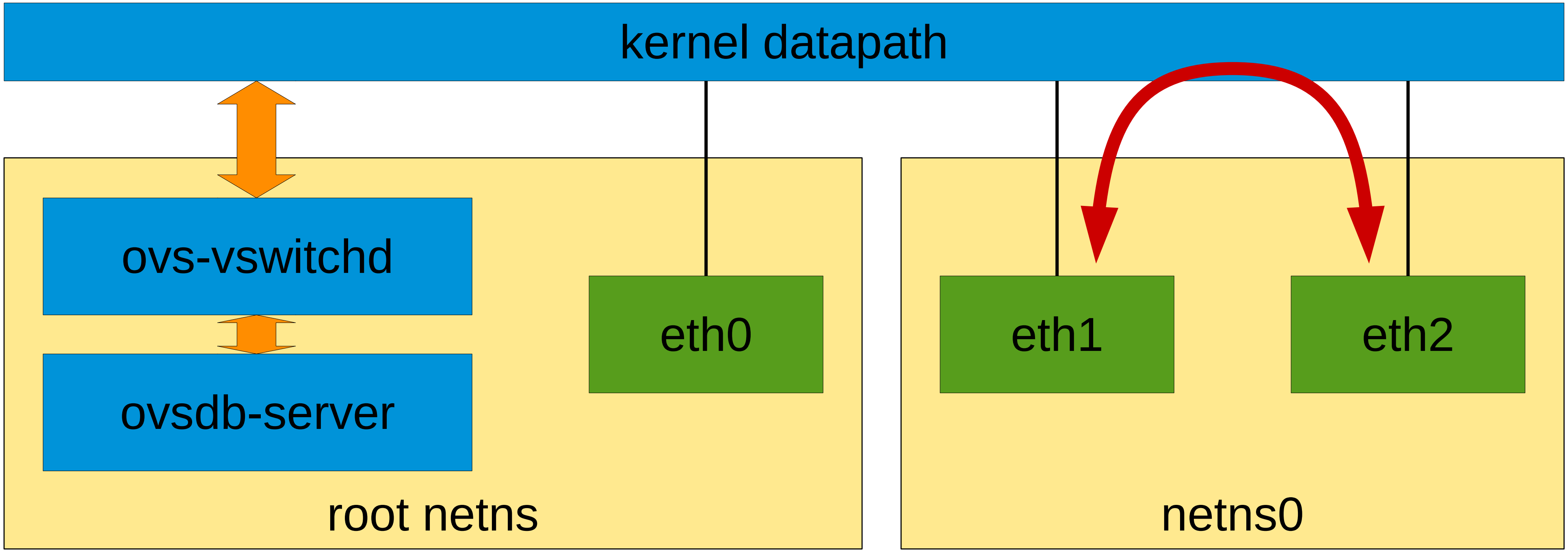
Kernel Datapath – the Easy Case



Kernel Datapath – the Easy Case Reversed



Kernel Datapath – Switching Inside Netns



Kernel Datapath – skb scrubbing

- Call skb_scrub_packet on **send** (ovs_vport_send)
  - compare netns of the ingress and egress interface
  - ignore netns of the datapath

Kernel Datapath – skb scrubbing

- Call skb_scrub_packet on **send** (ovs_vport_send)
  - compare netns of the ingress and egress interface
  - ignore netns of the datapath
- What about tunnels?

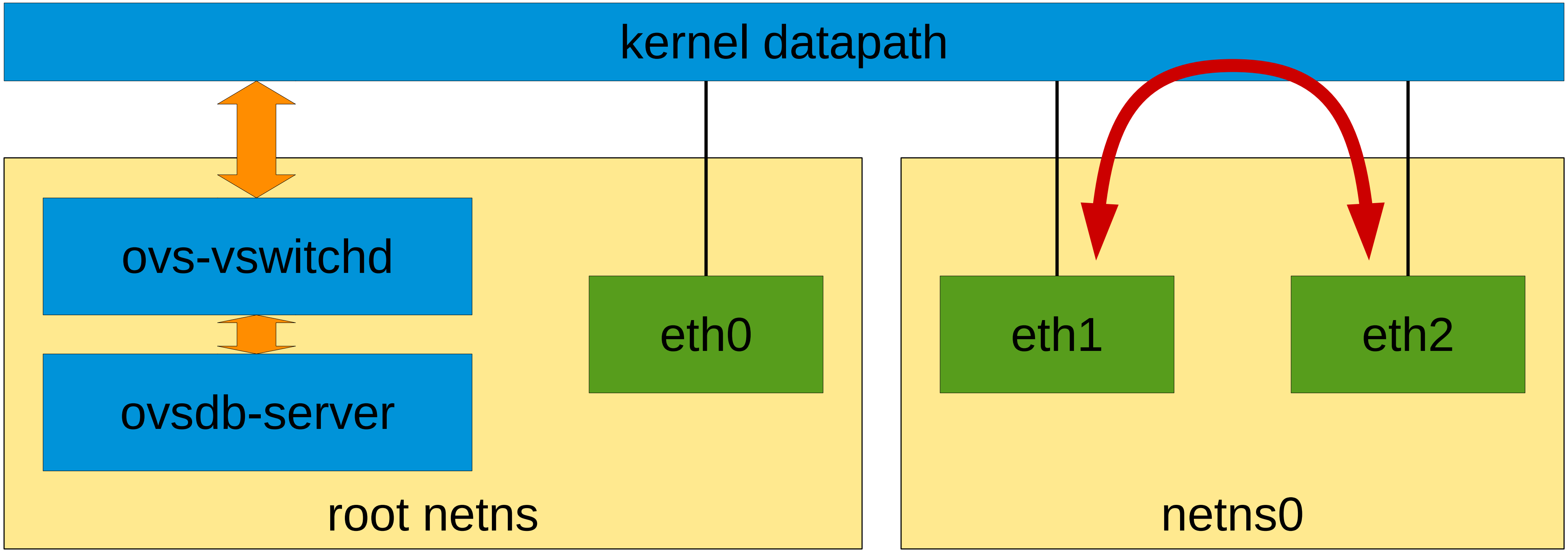
Kernel Datapath – skb scrubbing

- Call skb_scrub_packet on **send** (ovs_vport_send)
  - compare netns of the ingress and egress interface
  - ignore netns of the datapath
- What about tunnels?
  - nothing special since lwtunnels
- What about conntrack?

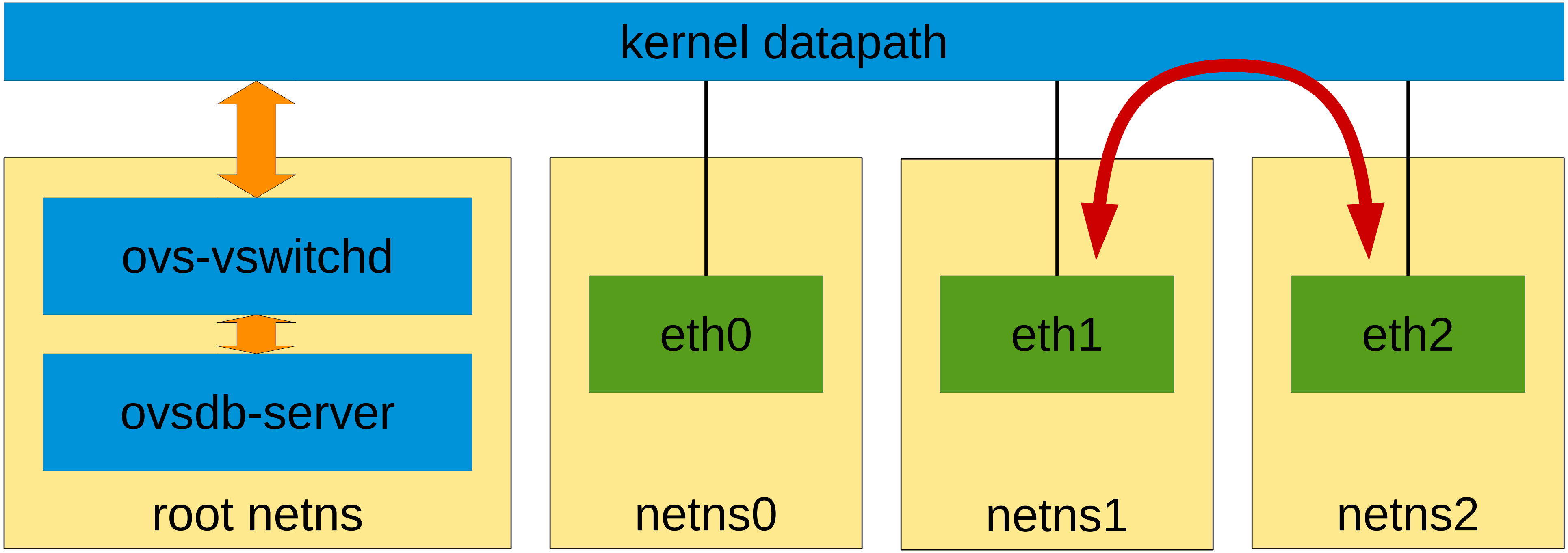
Kernel Datapath – skb scrubbing

- Call skb_scrub_packet on **send** (ovs_vport_send)
  - compare netns of the ingress and egress interface
  - ignore netns of the datapath
- What about tunnels?
  - nothing special since lwtunnels
- What about conntrack?
  - conntrack is done in datapath netns
  - egress scrubbing is too late

Kernel Datapath – Conntrack



Kernel Datapath – Conntrack



Matching in User Space

- ovsdb contains only the interface name
- Kernel datapath may have a different view
  - interface renames
  - moving interfaces between net namespaces
- Example:

```
ovs-vsctl add-port br0 eth0
ip link set eth0 name shadow0
ip link set eth1 name eth0
ovs-ofctl show br0
ovs-dpctl show
```

Detecting Interface Changes

- Listening to netlink events, updating the db
- What to do on interface deletion?

Detecting Interface Changes

- Listening to netlink events, updating the db
- What to do on interface deletion?
  - netns move is reported as delete + create
  - create is reported in the target netns

Detecting Interface Changes

- Listening to netlink events, updating the db
- What to do on interface deletion?
  - netns move is reported as delete + create
  - create is reported in the target netns
  - missing kernel API

Detecting Interface Changes

- Listening to netlink events, updating the db
- What to do on interface deletion?
  - netns move is reported as delete + create
  - create is reported in the target netns
  - missing kernel API
- Listening in other namespaces
  - NETLINK_LISTEN_ALL_NSID

Detecting Interface Changes

- Listening to netlink events, updating the db
- What to do on interface deletion?
  - netns move is reported as delete + create
  - create is reported in the target netns
  - missing kernel API
- Listening in other namespaces
  - NETLINK_LISTEN_ALL_NSID
  - no way to detect newly created namespaces
  - missing kernel API

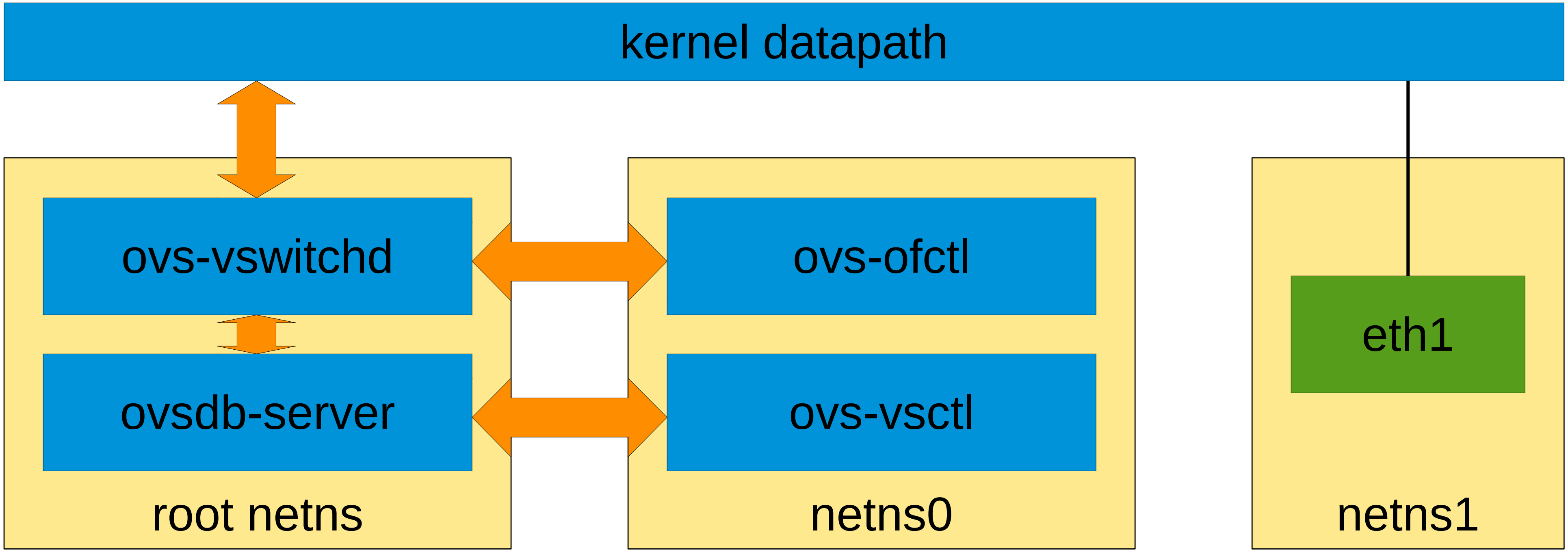
Namespaces in ovsdb

- Conflicting interface names
- Need to store netns in ovsdb
  - netnsid (from the ovsdb-server namespace)

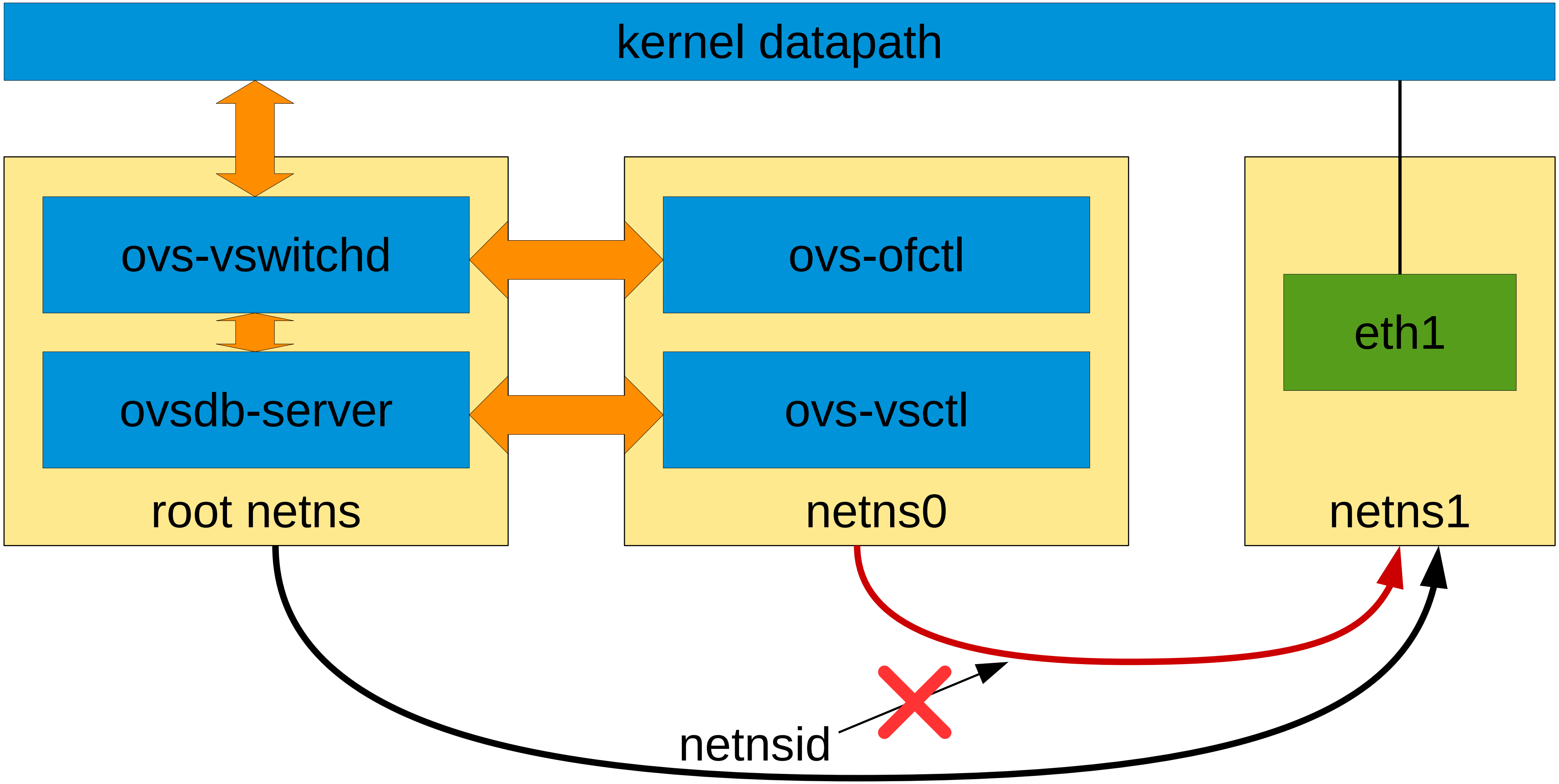
Namespaces in ovsdb

- Conflicting interface names
- Need to store netns in ovsdb
  - netnsid (from the ovsdb-server namespace)
- Cannot switch to netns using netnsid
  - missing kernel API

Netnsid Problem



Netnsid Problem



Questions? **Ideas?**