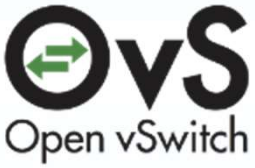
Introduction to Open vSwitch



Professor Chien-Chao Tseng

Department of Computer Science
National Yang Ming Chiao Tung University

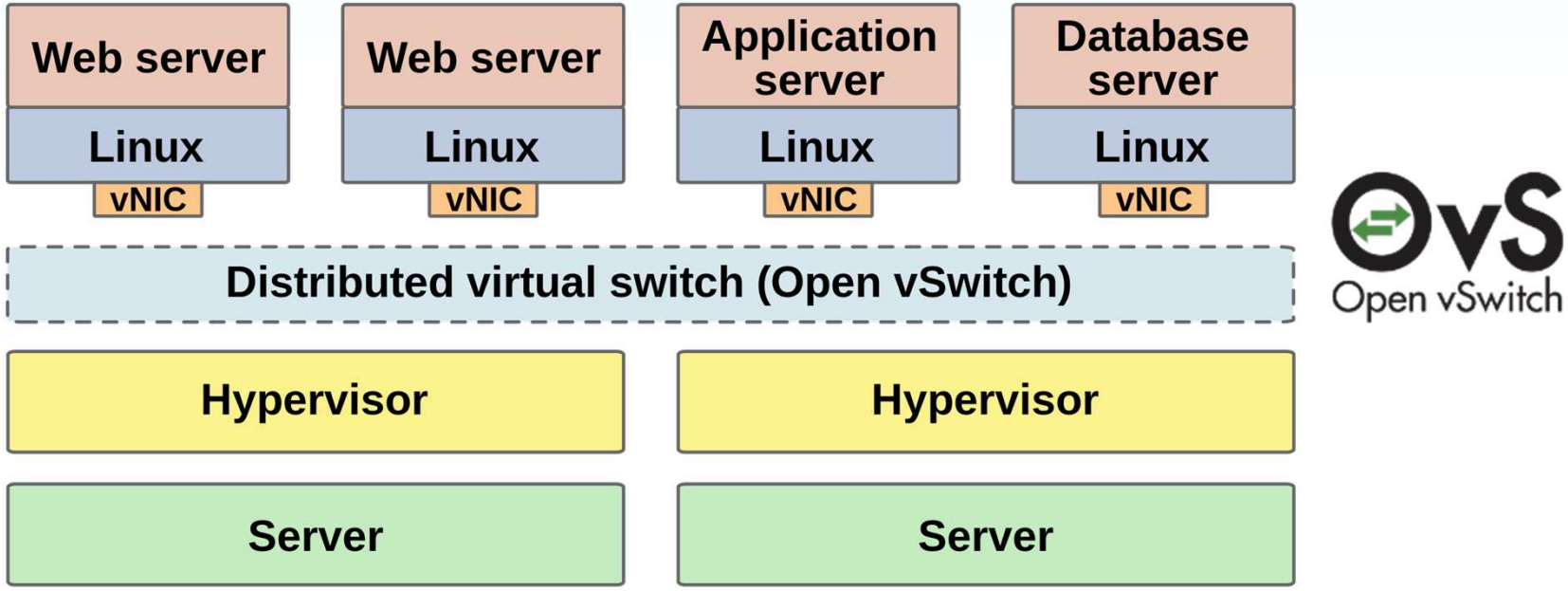
cctseng@cs.nctu.edu.tw

References:

1. *Dean Pemberton, Andy Linton and Sam Russell. Open vSwitch*, NSRC
2. Arthur Chiao. OVS Deep Dive.
3. Ben Pfaff et al. The Design and Implementation of Open vSwitch, USENIX.
4. Shie-Yuan Wang. Open vSwitch (OVS) Usages and Its Internal Design and Implementation
5. H. W. Chiu, https://www.hwchiu.com/openvswitch-overview.html

Open vSwitch (OVS)

- An open-source implementation of a distributed virtual **multilayer** (L2/L3) switch.
  - Deployed as a cross-server virtual network switch

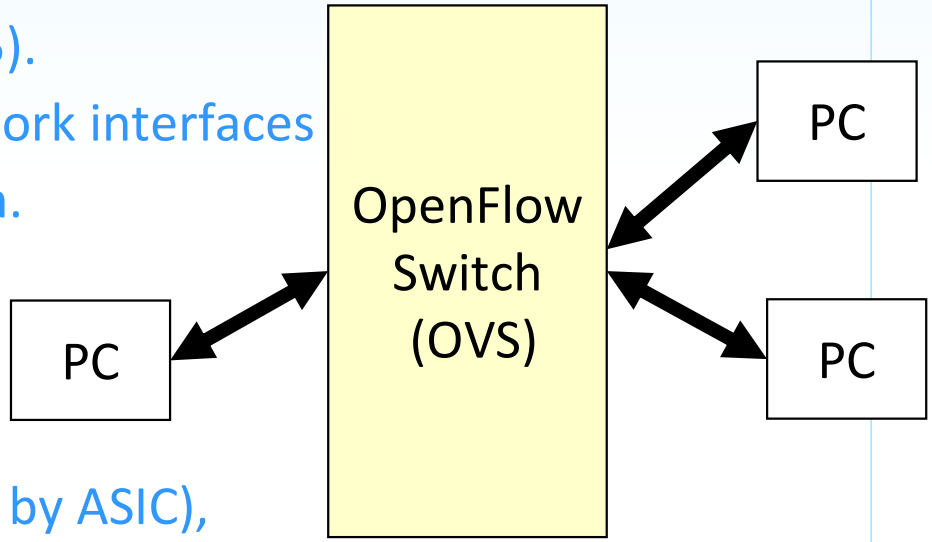


By Goran tek-en, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=32070011

- Originally used by Hypervisor to bridge traffic between VMs and with outside world
  - Why not use the built-in L2 switch (Linux bridge)?

An OVS Usage Example

- OVS is a widely-used software virtual switch
- supports the latest OpenFlow protocol (ver 1.5).
- Runs OVS on a Linux server with multiple network interfaces
  - Can turn the server into an OpenFlow switch.

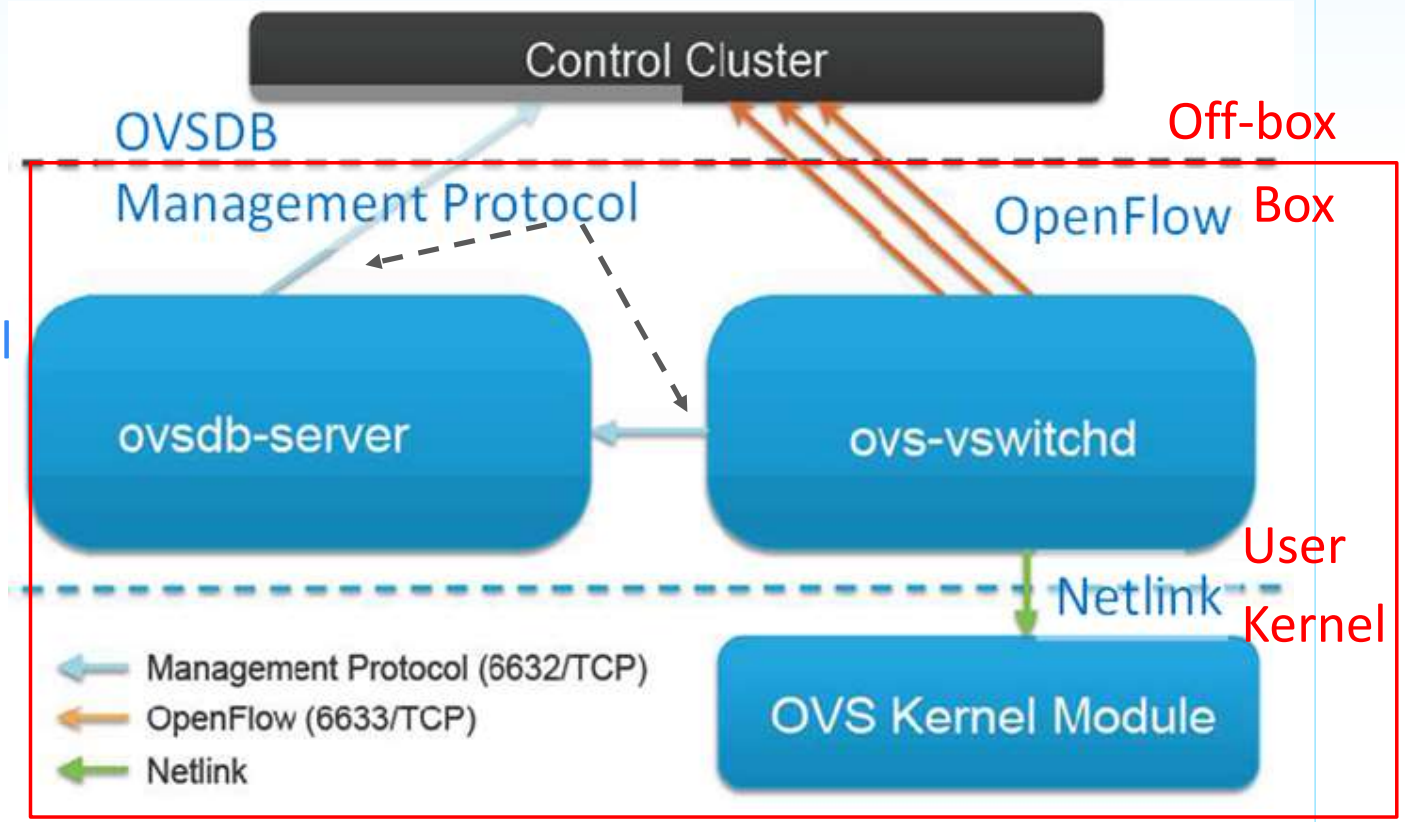


- Packet forwarding is done by CPU (rather than by ASIC),
  - Cannot support many ports and high data rate.
    - packets may be delayed or eventually dropped in OVS

- OVS is also widely used as an OpenFlow agent running inside commercial OpenFlow switches

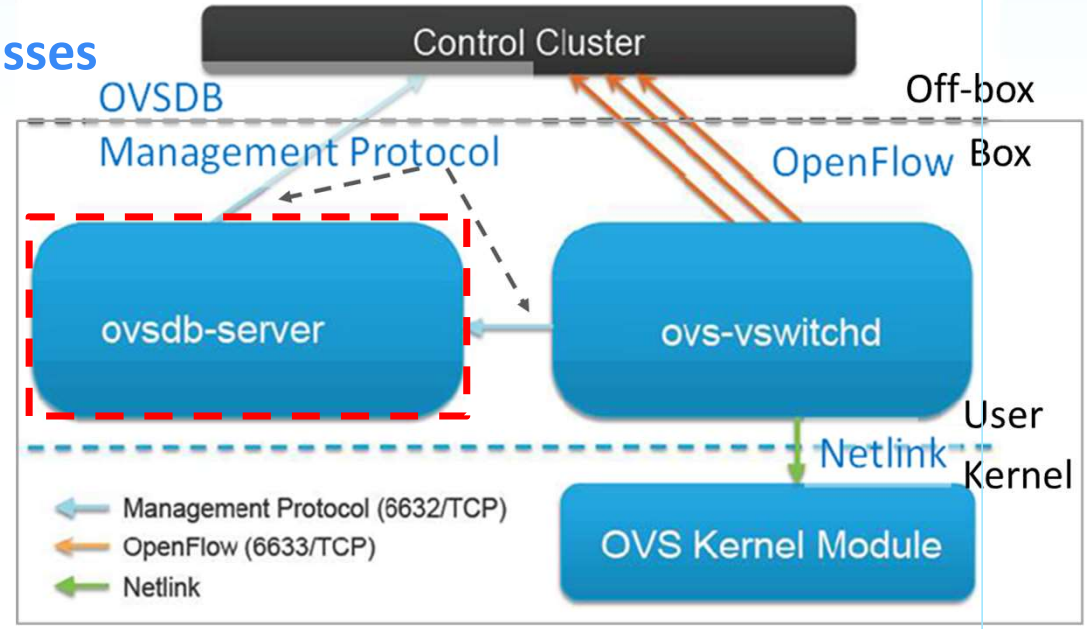
Open vSwitch Architecture

- **Main Components:**
- **ovs-vswitchd**
  - User space program
  - A daemon that controls all OVSs in the system
  - Tools: ovs-appctl, vs-ofctl
- **ovsdb-server**
  - User space program
  - Database server of ovs
- **Kernel Module** (datapath)
  - kernel space module,
  - OVS packet forwarder
  - Tools: ovs-dpctl



ovsdb-server

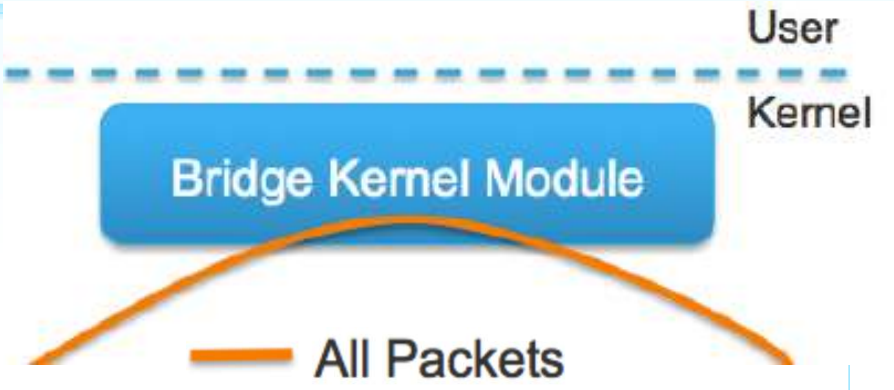
- OVSDB: Database that holds switch-level configuration, such as
  - Bridge, interface, tunnel definitions
  - OVSDB and OpenFlow **Controller addresses**
- Configuration is stored on disk and survives a reboot
- A custom database with nice properties
  - Value constraints
  - Weak references
  - Garbage collection
- Log-based (fantastic for debugging!)



- **ovsdb-server** speaks OVSDB protocol (JSON-RPC) to **Controller** and **ovs-vswitchd**
  - OVSDB: RFC 7047 (The Open vSwitch Database Management Protocol)
    - OpenFlow configuration protocol designed to manage OVS implementations.

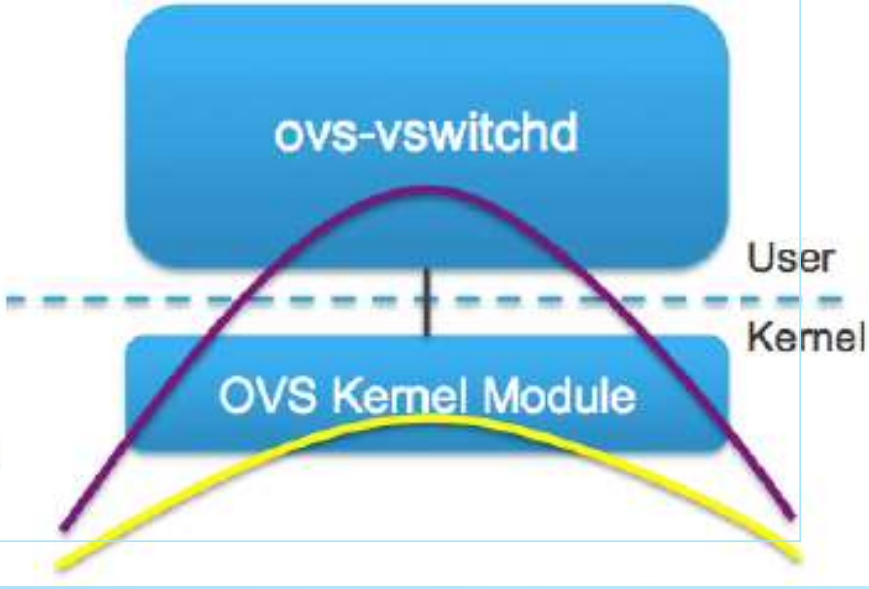
Packet Handling

- **Linux Bridge**
  - Simple forwarding
  - Matches destination MAC address and forwards
  - Packet never leaves kernel



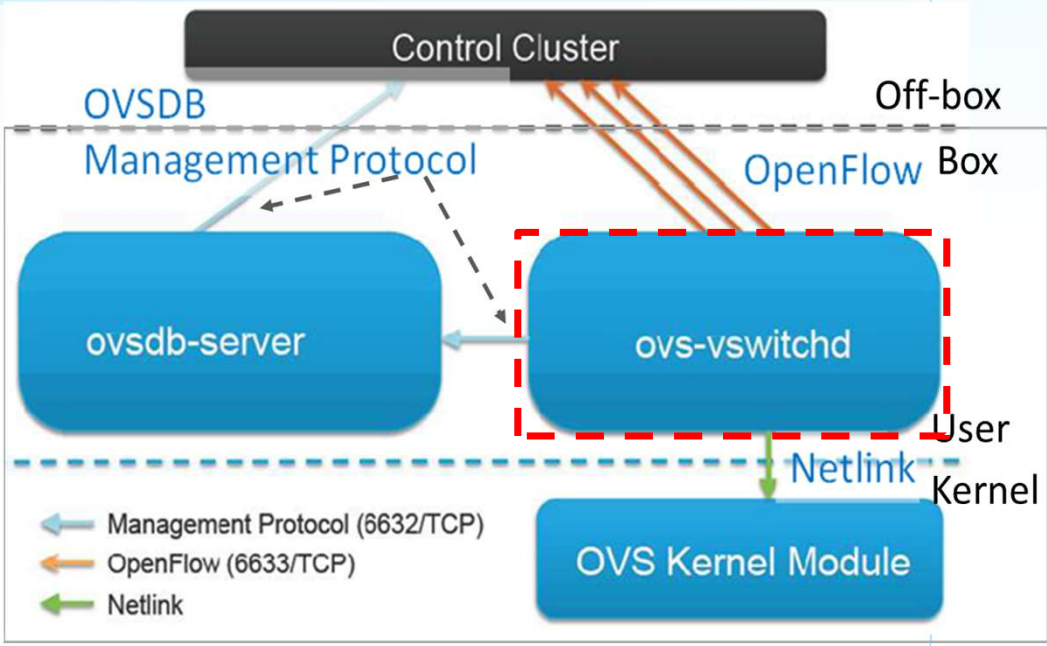
- **OVS**

  Decision about how to process packet made in **user space**
  - **First packet** of new flow goes to **ovs-vswitchd** (slowpath)
  - **Subsequent packets** hit **cached entry** in **kernel** (fastpath)

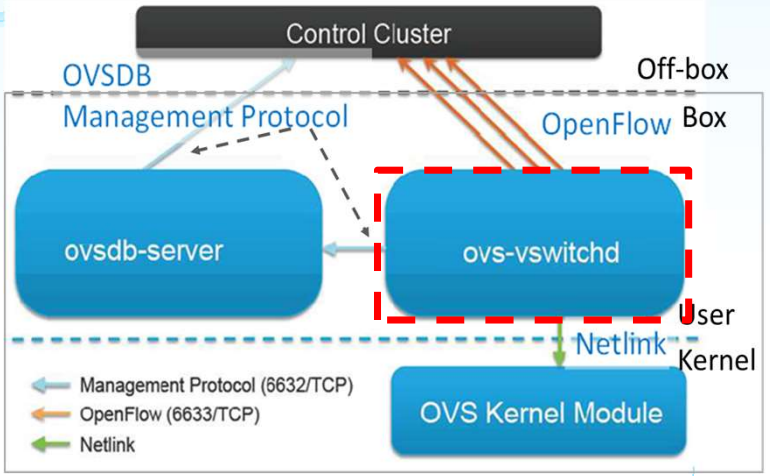


ovs-vswitchd

- Communicates with
  - **Outside world** using **OpenFlow**
  - **ovsdb-server** using **ovsdb management protocol**
  - **Kernel module** over **netlink**
  - **System** through **netdev** abstract interface
- Implements **mirroring, bonding, and VLANs**
  - through modifications of **flow table**
- **Handle OpenFlow messages**
- Also handle **flow expiration** and **statistics requests**
  - By checking **datapath** flow counters



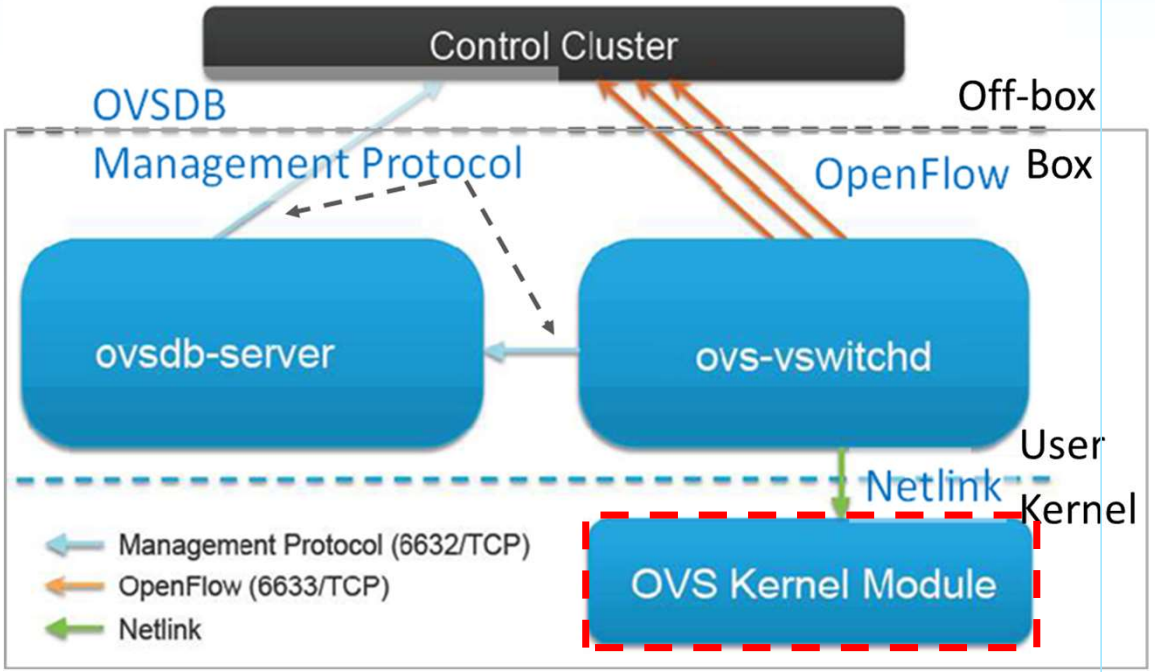
ovs-vswitchd – Flow Tables and Packet Processing



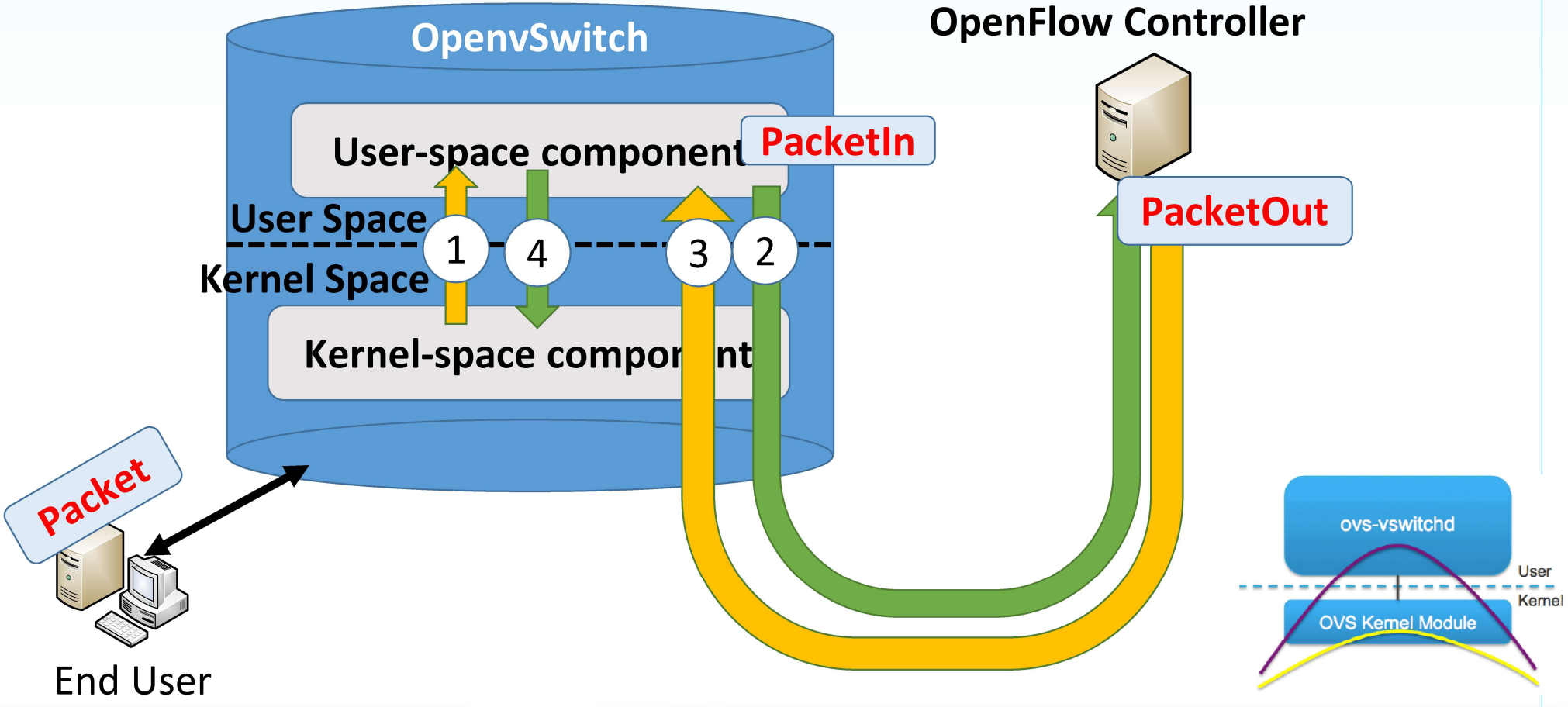
- **ovs-vswitchd** maintains **two** flow tables
  - **Exact Match** flow table
  - **Wildcard Match** flow table
  - OpenFlow Controller control Flow tables
- **Exact Match** has high priority than **Wildcard Match**,
- Lookup **Exact Match** flow table first.
  - if found, send **two** netlink message to **OVS Kernel Module**
    - One for packet processing action
    - The other for **Exact Match** flow entry **installation** (in OVS Kernel Module)
  - Otherwise, look up **Wildcard Match** flow table,
    - if found, generate a corresponding **Exact Match** flow entry in **ovs-vswitchd** and send **two** netlink messages to **OVS Kernel Module**.
    - Otherwise, issue a **Packet_In** to the controller.

Open vSwitch Kernel Module

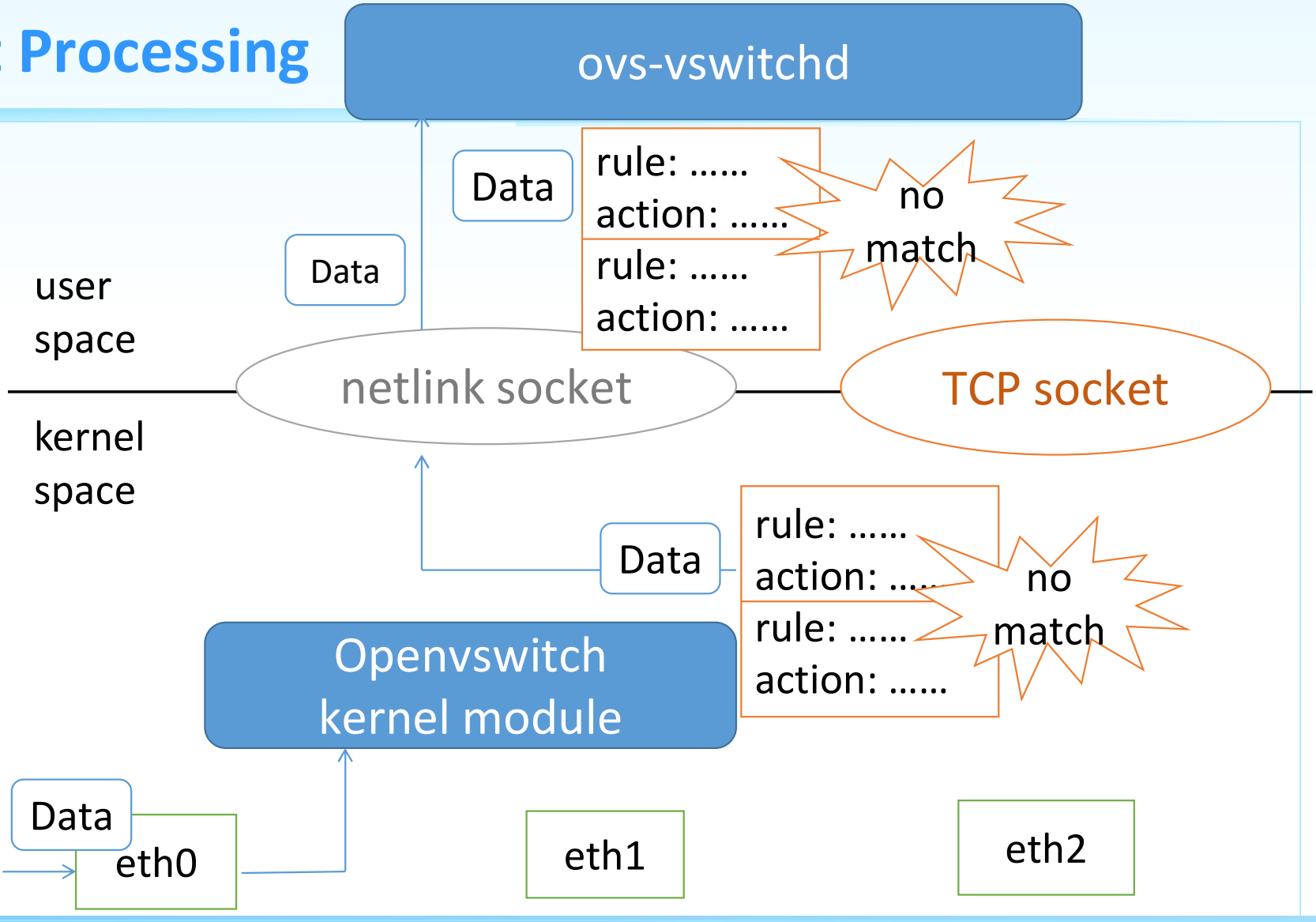
- Designed to be fast and simple
- Maintains **Exact-Match** flow table (cache)
  - **Exact-Match** flows:
    flows without wildcards
  - – Compares an incoming packet,
    - if a match found,
      executes associated actions
      and updates counters
    - Otherwise,
      sent to **ovs-vswitchd**
- Implements tunnels
- ✓ Remarks:
  - – Knows nothing of OpenFlow protocol
  - – Does not perform flow expiration



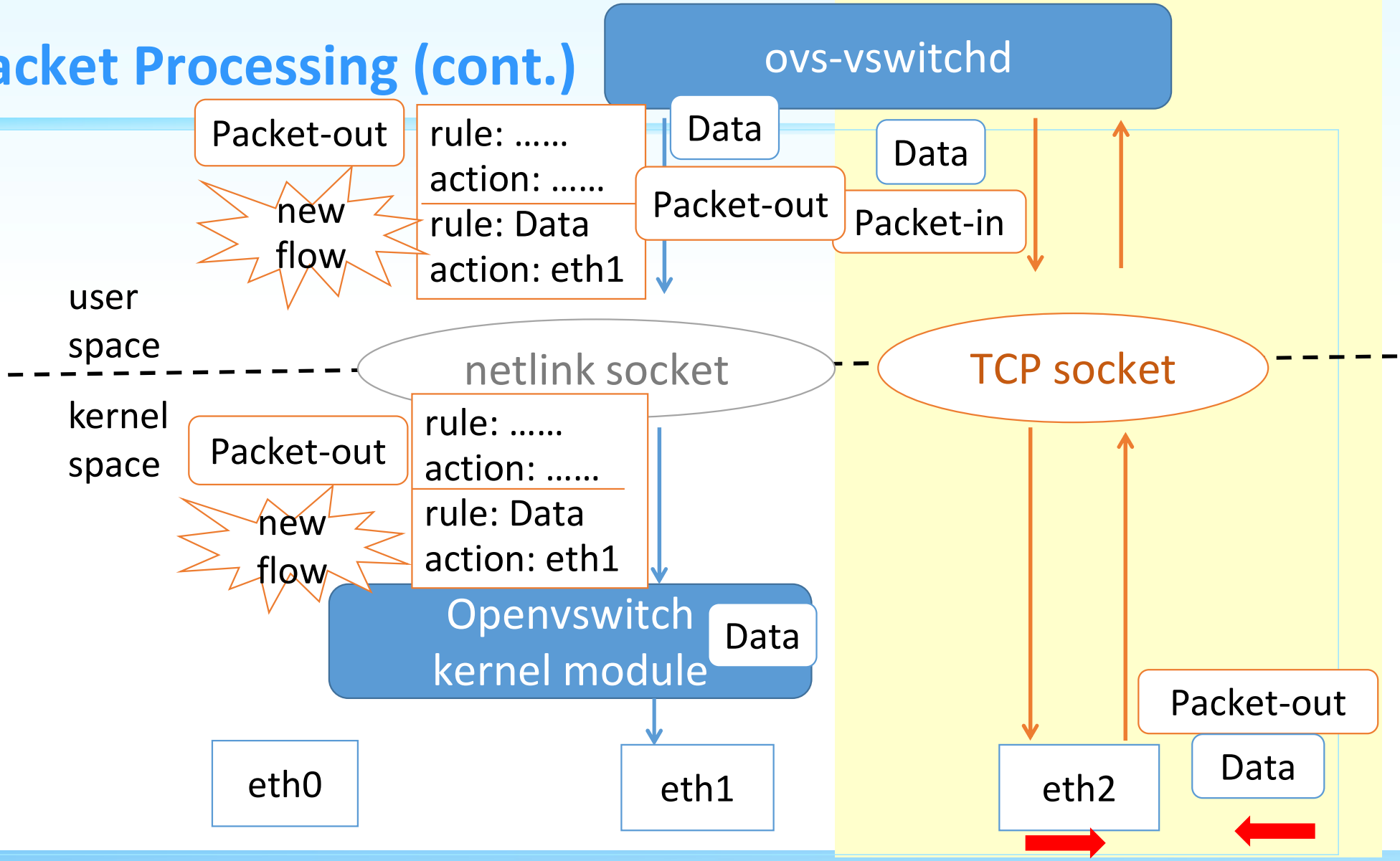
OVS Workflow



Packet Processing



Packet Processing (cont.)



OVS Flow Tables and Manipulations