# CSE 291: Operating Systems in Datacenters

Amy Ousterhout

Sept. 29, 2022

# Agenda for Today

- Reminders
- Introduction to network stacks
- IX discussion
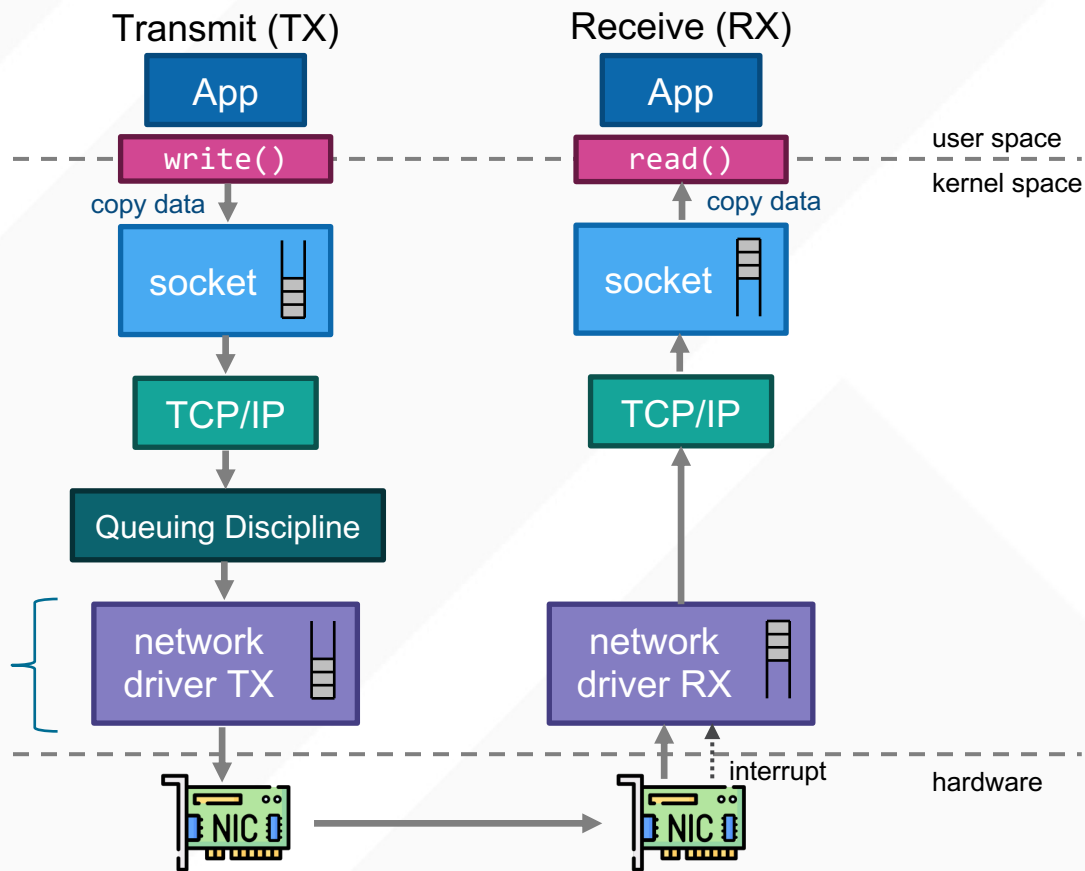- XDP discussion

# Reminders

- Sign up to lead a discussion
  - Due tonight at 11:59 pm
- First Day Survey #FinAid
  - Due Friday 10/7
- Warm-up assignment
  - Posted on Canvas
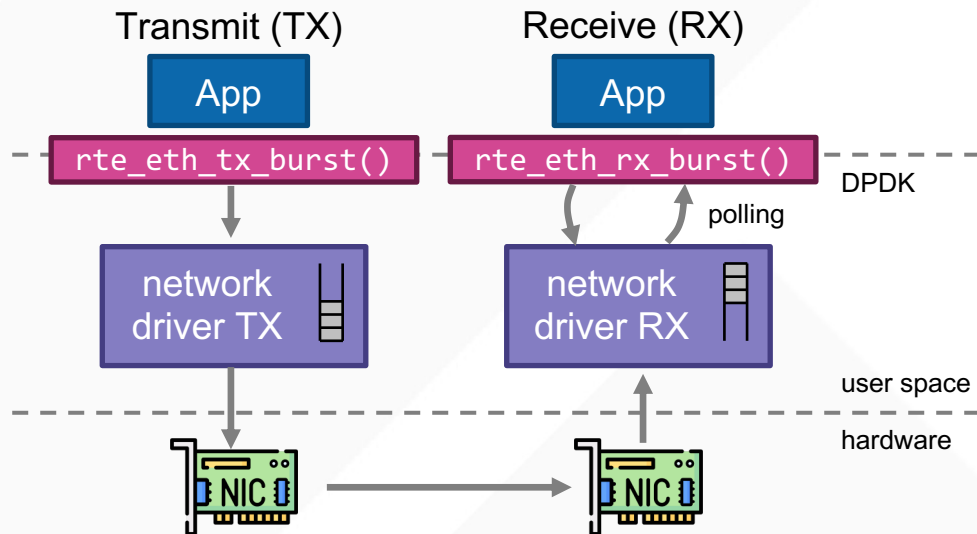  - Due Tuesday 10/11 at 11:59 pm

# Network Stacks

# Linux's Network Stack

- A simplified version
- Kernel data structure for packets: sk_buff
- Overheads:
  - Copying data
  - Context switches
  - Interrupts
  - Lots of queueing
- For more details, see "Understanding Host Network Stack Overheads" [SIGCOMM '21]



Transmit (TX)

App

write()

copy data

socket

TCP/IP

Queuing Discipline

network driver TX

Device driver (NIC-specific)

Receive (RX)

App

read()

copy data

socket

TCP/IP

network driver RX
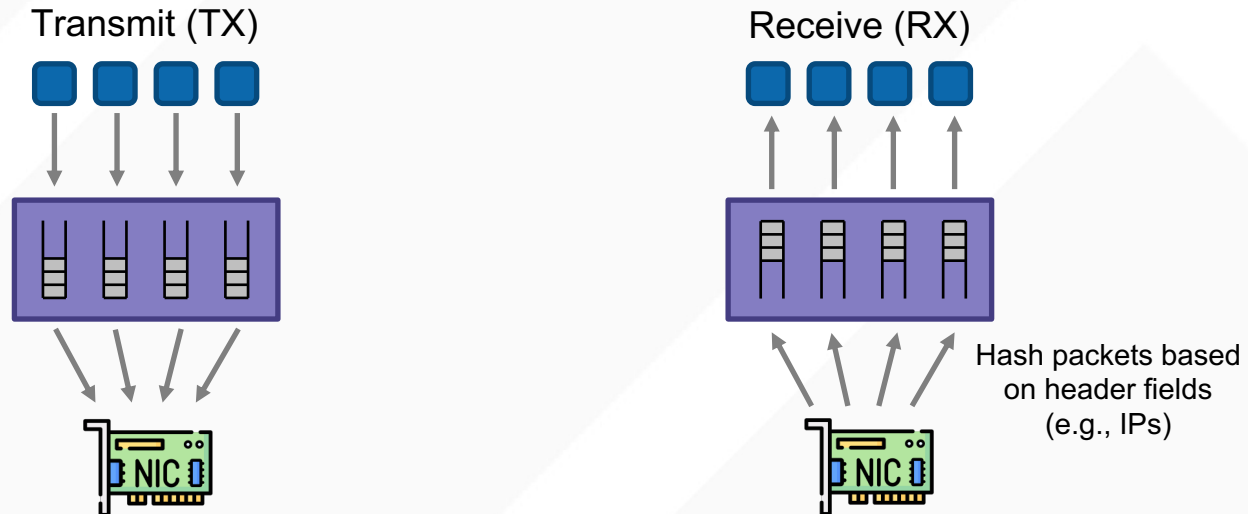
user space
kernel space

interrupt

hardware

NIC

NIC

# Data Plane Development Kit (DPDK)

- One example of kernel bypass
- DPDK data structure for packets: rte_mbuf
- Differences from Linux:
  - Much less functionality
  - No copying
  - Polling instead of interrupts

Transmit (TX)

Receive (RX)

App

App

`rte_eth_tx_burst()`

`rte_eth_rx_burst()`

DPDK

polling

network driver TX

network driver RX

user space

hardware

NIC

NIC

# Handling Multicore

- One NIC queue per core
- Transmit path: each core sends to its own queue
- Receive path:
  - Flow-consistent hashing using receive-side scaling (RSS)



Transmit (TX)

Receive (RX)

Hash packets based on header fields (e.g., IPs)

# Prep for Tuesday

- FaRM discussion
  - Read FaRM and submit review as usual
- Where do research ideas come from? To prepare, consider:
  - What is a systems paper you have really liked?
  - Why did you like it?
  - How do you think the authors came up with the idea?
  - (no need to submit anything)