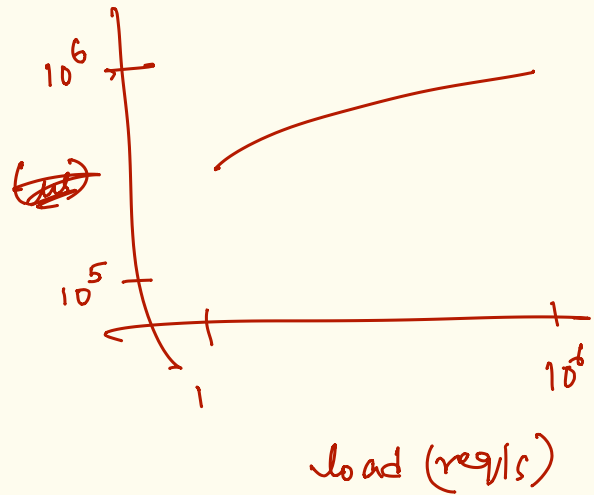
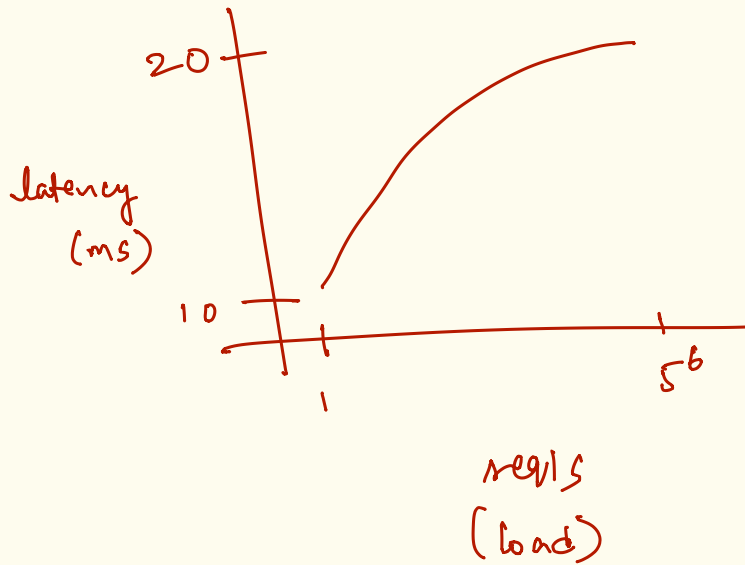
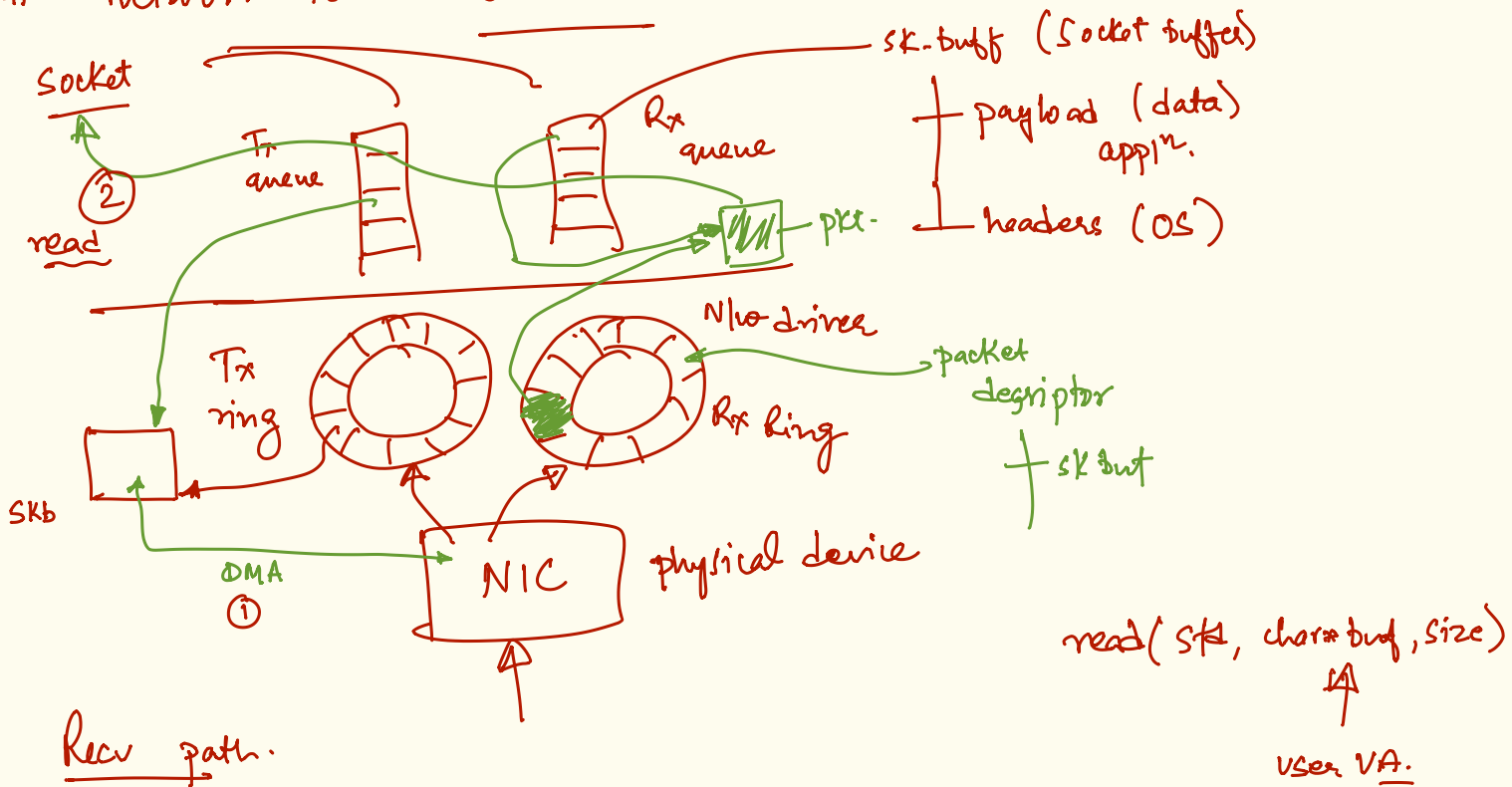


Lecture 20

CS744



network IO OS-view



- use free pkt. descriptors to DMA pkt. into memory. (NIC)

- raise interrupt.

- OS perform protocol processing.

- uses sk-buff in socket object & points to allocates packet in descriptor.

- allocates new packet descriptors for Rx ring.

interrupt processing.

Network IO optimizations.

① split n/w processing into two halves. [Kflushd]
[Kmigrationd]
[Ksoftirqd]

- | <u>top-half</u> | & | <u>bottom-half</u> |
|---|---|--|
| <ul style="list-style-type: none">- minimal processing- ^{very} quick interrupt handling- acks the interrupt- move the Rx ring ptr.- <u>schedule</u> bottom half work. } as processing Rx Queue- return from interrupt. | | <ul style="list-style-type: none">- not in interrupt context- OS pkt processing- skb handling in socket queues |

② RSS — receive side scaling. (hw feature)

- process interrupts simultaneously ~~on~~ on many cores.
- each core has its own Rx Ring / Tx Ring
- split pkts based on 4-tuple across multiple rings.
- each ϕ ring processed by a separate interrupt handler & Ksoftirqd per core.

③ NAPI — new API

- combination of interrupts + polling
- at high rate s/w to polling
- at low rates work w/ interrupts.

④ Kernel by-pass (skip ~~all~~ all OS processing for received pkts).

- + DPDK — hw feature
- AF-XDP — OS feature (s/w)

- DMA pkts directly to user VAs.
- no interrupts.

- user process polls & consumes & manages Rx/Tx rings.

high-speed IO
+ custom protocols
+ raw pkt. processing