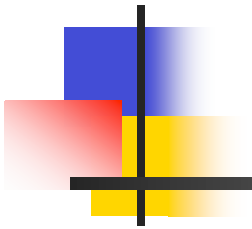


Protocol Offloading Using an IXP2400 Network Processor

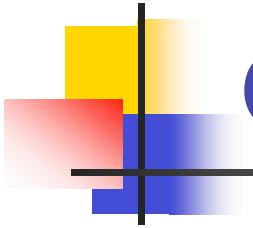
Chris Baron, Yan Luo*, Laxmi Bhuyan

Computer Science & Engineering

University of California Riverside

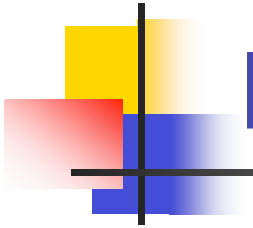
*now with Dept. of Electrical & Computer Engineering of UMass Lowell

http://faculty.uml.edu/yluo/



Outline

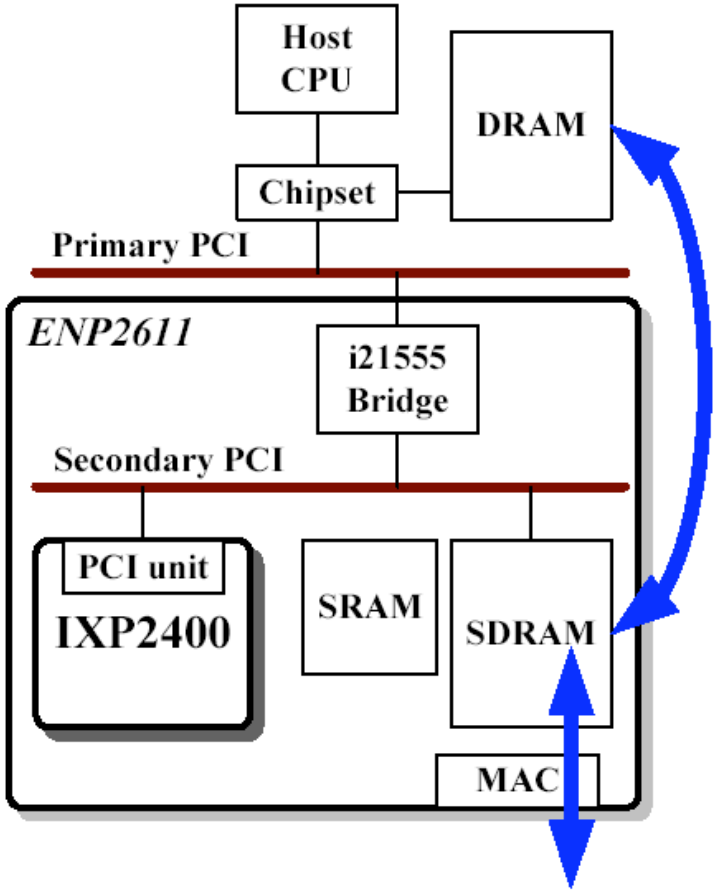
- Motivation
- Design Goals and Challenges
- Implementation of Protocol Offloading
- Performance Evaluation
- Ongoing Research and Teaching



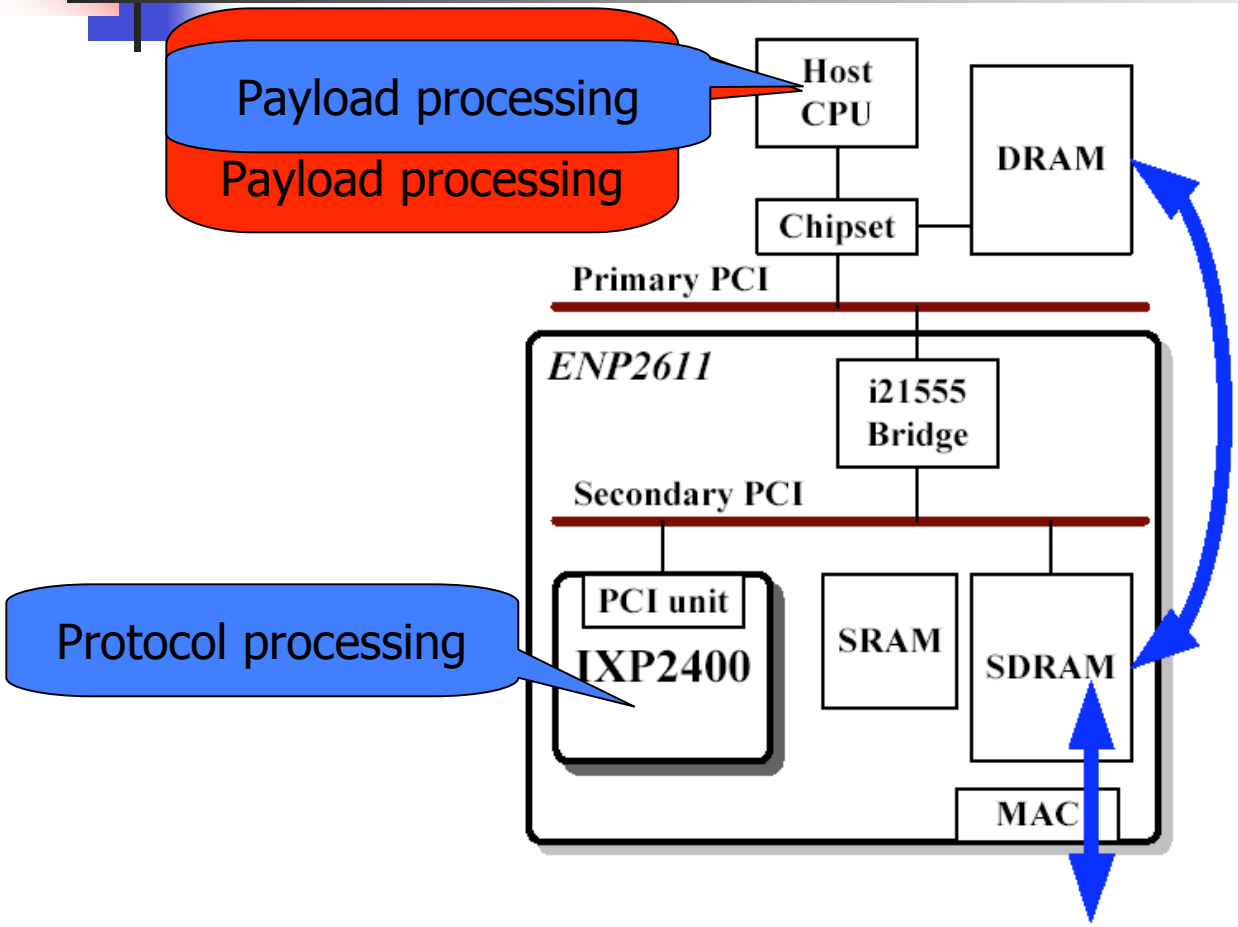
Motivation

- High performance interconnects required in clusters
  - High bandwidth
  - Low latency
- Protocol processing overhead increases
  - “1GHz for 1Gbps”
- Related work
  - Ethernet Message Passing [Shivam ’01]
  - TCP/IP Offload Engine [Adaptec’03]
  - IXP1200-based NIC [Mackenzie’03]

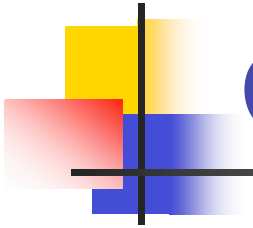
System Architecture



Design Goals

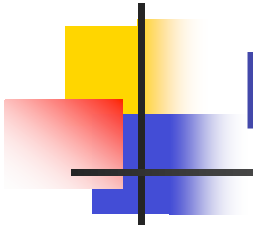


- Transparency
- UDP-like protocol
- Mem access across PCI domain
- Interrupt and polling



Challenges

- Transparency to Applications
  - Device driver for ENP2611 card
  - API of connection setup, send and receive
- Memory Access across PCI domains
  - Memory mapping between two domains
  - Configuring PCI bridge
  - DMA/PIO on host, DMA on IXP
- Efficient Utilization of IXP2400 resources
  - Memory allocation (scratchpad, SRAM, SDRAM)
  - Workload partition on MEs



Resource Allocation

- Memory
  - Host side: Bigphys memory for DMA, send buf and receive buf, connection table
  - IXP side: downbuf and upbuf in SDRAM, queue in scratchpad
- ME allocation
  - Downbuf manager (1)
  - Downbuf worker (1 to 4)
  - Upbuf manager (1)
  - XMIT and RCV (2)

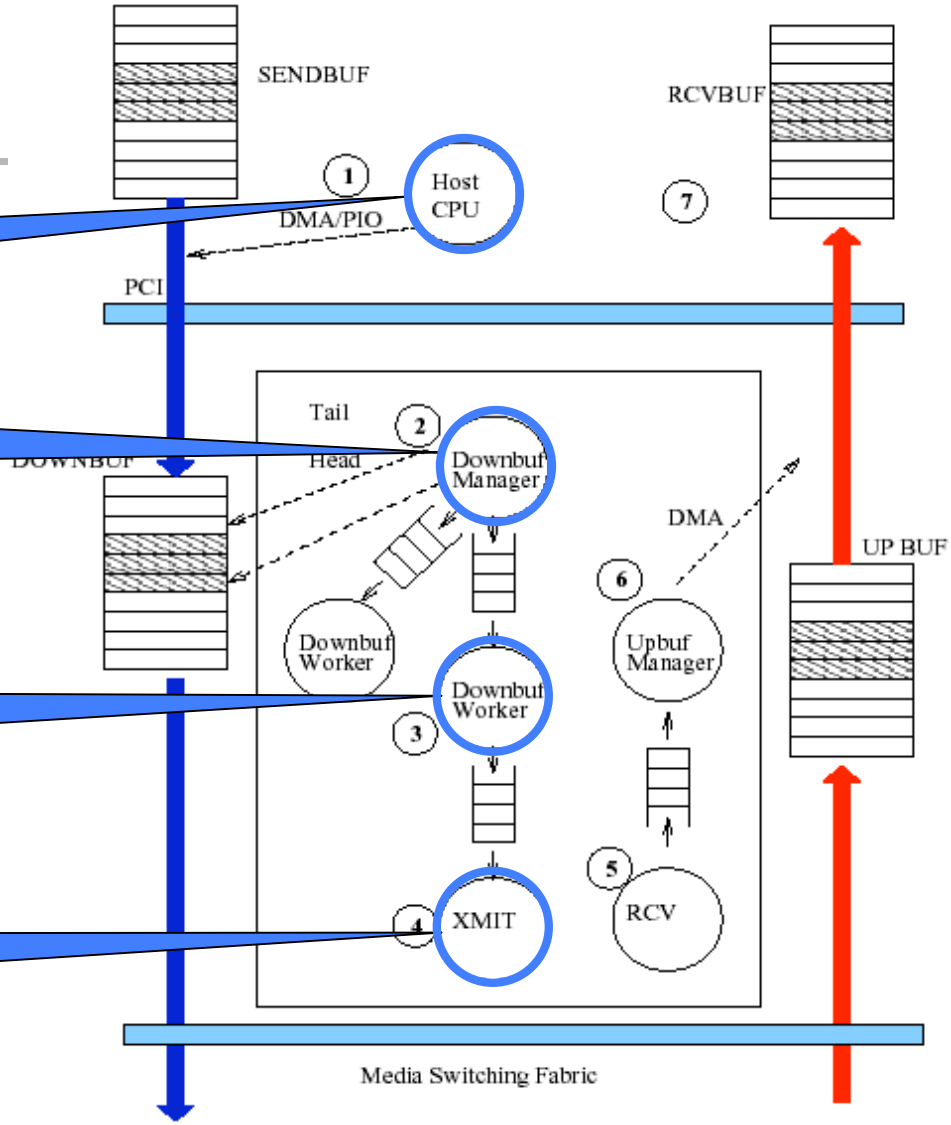
Send

1. Host CPU "send"

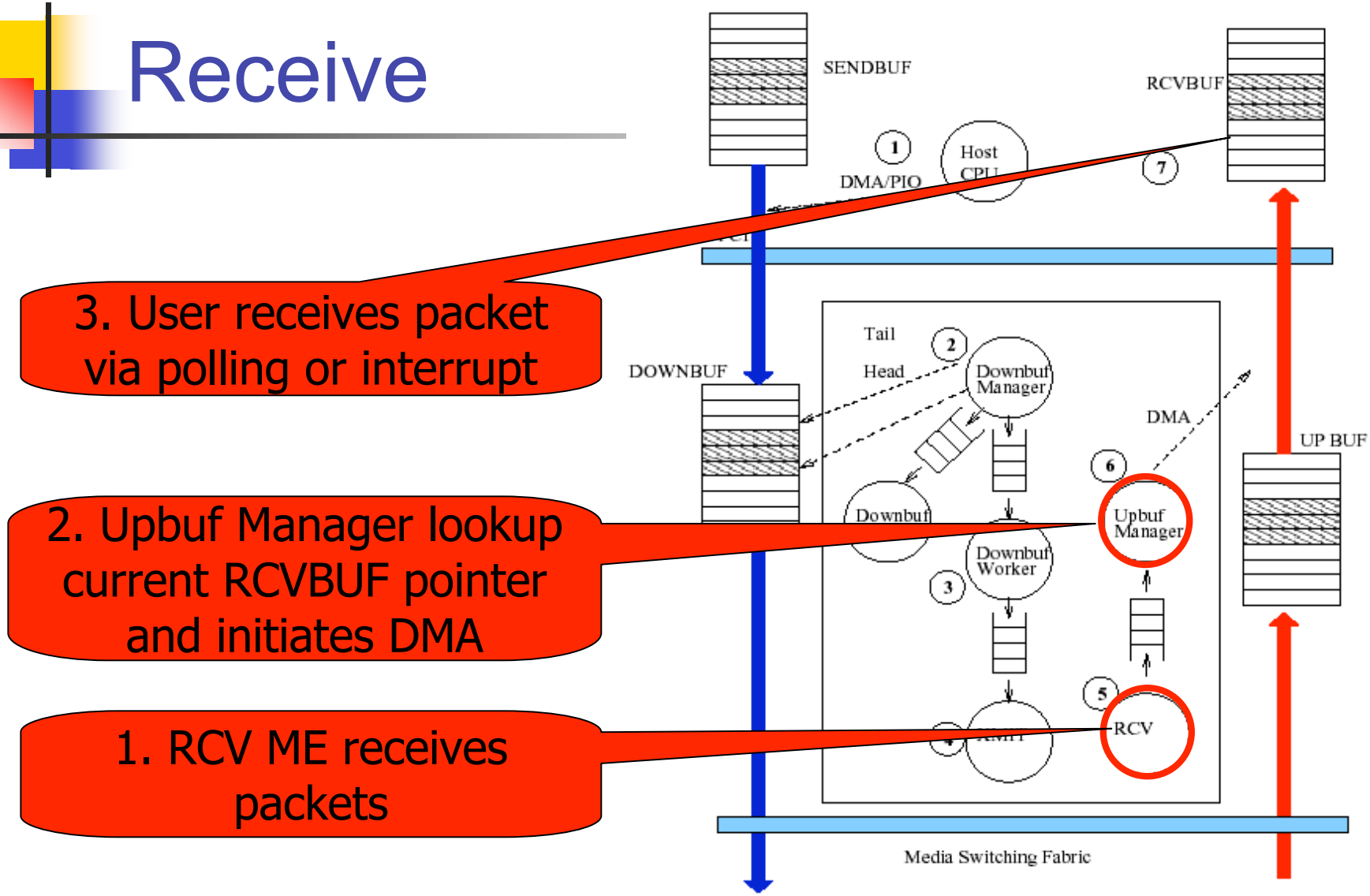
2. Downbuf manager dispatch new packet

3. Downbuf worker prepare new packet

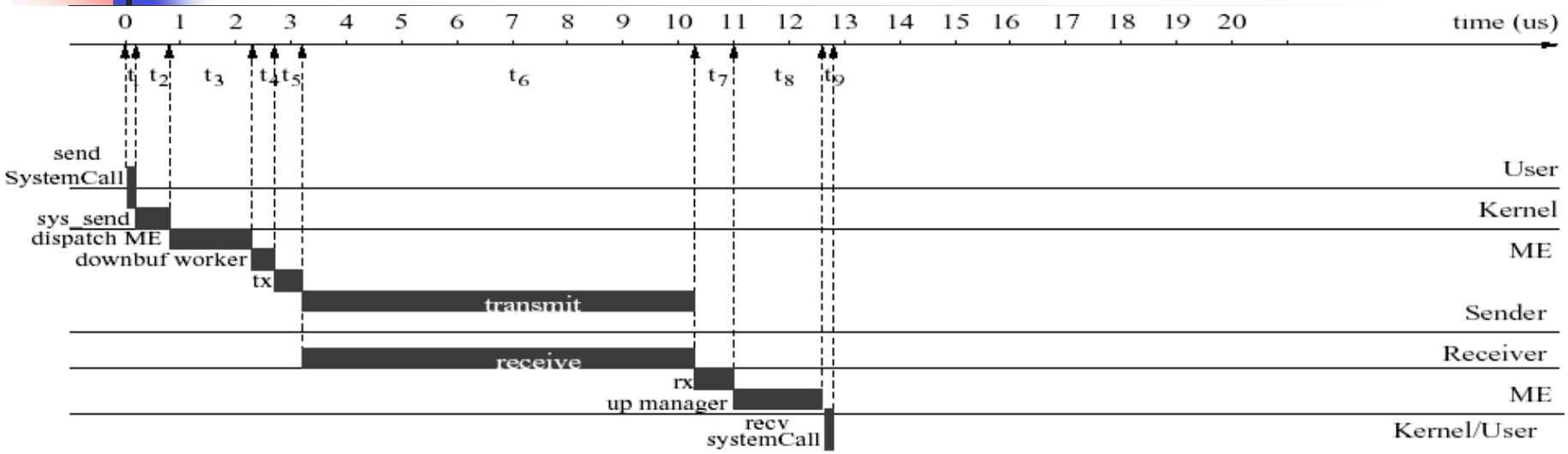
4. XMIT ME send packet onto wire



Receive


SENDBUF
RCVBUF
1
Host CPU
DMA/PIO
7
3. User receives packet via polling or interrupt
DOWNBUF
Tail
Head
2
Downbuf Manager
DMA
UP BUF
6
Downbuf
Upbuf Manager
2. Upbuf Manager lookup current RCVBUF pointer and initiates DMA
Downbuf Worker
3
5
RCV
1. RCV ME receives packets
Media Switching Fabric

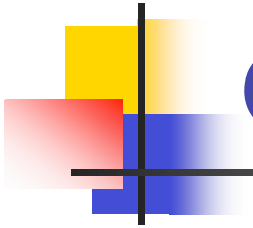
Latency and Throughput



Time line analysis of transfering a 1-byte payload

Processing Latency: 4.56 us (IXP2400) vs 7.07 (Intel 82544GC NIC)

Throughput: 950Mbps (with one worker ME), headroom for 3Gbps



Ongoing Research/Teaching

- NePSim 2
  - Models IXP2400/2800 (NePSim 1.0 models 1200)
  - Under internal testing
  - Expected to release early 2006
- NP-based switches
  - Content-aware switch
  - VoIP gateway
- A graduate level course on network processors to be offered in Spring 2006

Thank You !

Questions ?

Yan Luo

Yan_Luo@uml.edu

http://faculty.uml.edu/yluo/