

ECE 697J

Final Exam

Name: Solutions

	Maximum	Achieved
Question 1	16	
Question 2	12	
Question 3	6	
Question 4	20	
Question 5	10	
Question 6	12	
Question 7	14	
Question 8	10	
Total	100	

Please write legibly! Be concise. Unreadable answers will not be graded.

Time: 90 minutes.

Question 1 (16 points (4+4+8)):

Answer the following general questions regarding Network Processors:

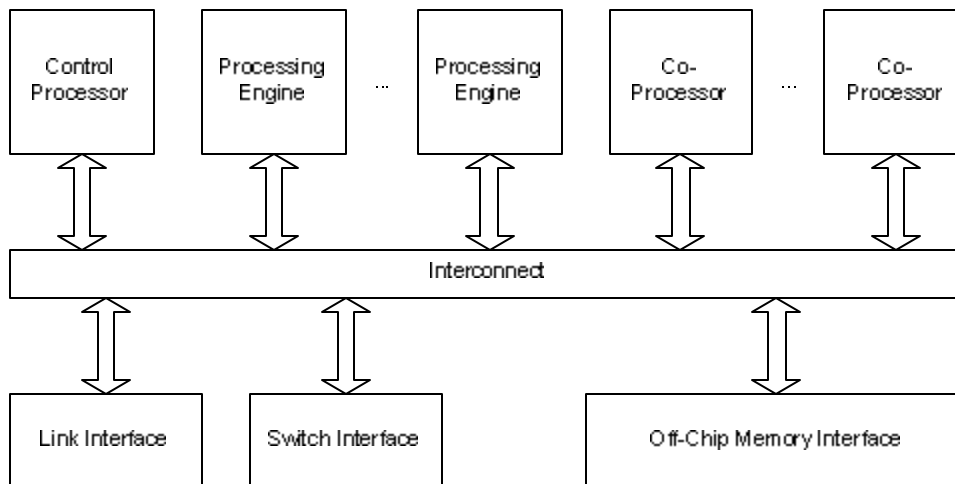
- a) What is the motivation behind developing network processors instead of simply using workstation processors?

The workload for network processors is very different from workstations. A single processor does not meet the performance and scalability requirements of networks.

- b) What are the main characteristics of a network processor?

System-on-a-chip (processors, memory, I/O on a single chip), CMP (chip-multiprocessor), usually numerous simple processors and one control processor, a few co-processors (checksum, lookups).

- c) Show a system outline of a network processor (does not have to be a realistic commercial processor).



Question 2 (12 points (4+4+4)):

Answer the following questions regarding Benchmarking and Network Processor Design:

- a) What are the differences in the workload characteristics of network processors as compared to workstation processors?

Network Processor: simple tasks, many tasks, repetitive tasks, small amounts of data processed (packet), overall workload highly parallelizable, much I/O, real-time constraints

Workstation Processor: opposite

- b) Suggest three benchmark applications that cover a broad range of network processing and argue why these are good choices.

IP forwarding: most commonly used “application”, requires simple lookup, requires simple checksum computation.

Firewall: requires more complex lookup (policies), requires maintenance of per-flow state.

Encryption of packets (e.g. IPsec): requires complex processing (could be done by co-processor), payload modifications.

- c) In the context of CommBench, “computational complexity” is defined as a benchmark metric. What is the exact definition and why is it relevant?

Complexity = (number of instructions executed) / (amount of data processed)

It gives a system-independent metric for the amount of processing that is required for a data stream of given bandwidth. Note that complexity is measured in instructions/byte, not cycles/byte (wouldn't be system-independent).

Question 3 (6 points (3+1+2)):

Answer the following questions regarding Parallelism in Networks

- a) In class we have discussed that there are several levels of parallelism in networking. Describe where this parallelism occurs.

System-level parallelism: router ports can operate independently

Flow-level parallelism: packet of different flows can be processed independently on a network processor

Packet-level parallelism: in some cases packets of one flow can be processed independently

Instruction-level parallelism: instructions of one flow can execute in parallel

- b) How can these levels of parallelism be exploited by a network processor architecture?

One NP on each port, multiple processing engines on each NP.

- c) What are the limitations to exploiting parallelism on a router and on a network processor?

Shared data structures (e.g., routing tables)

Shared hardware (e.g., off-chip memory channel, co-processors)

Packet-dependencies within a flow

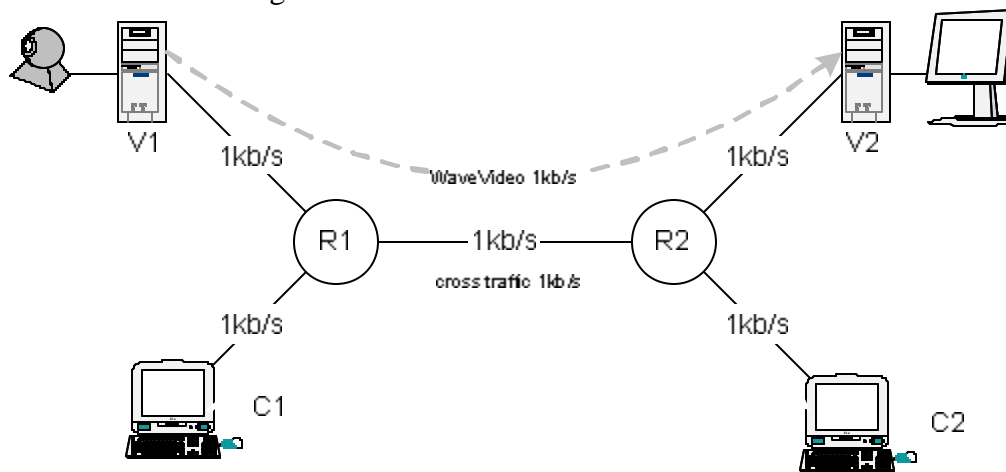
Question 4 (20 points (2+6+1+5+6)):

Answer the following questions regarding the WaveVideo application. This question is similar to question 4 in the midterm.

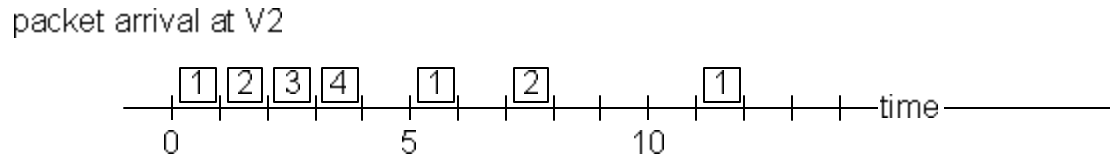
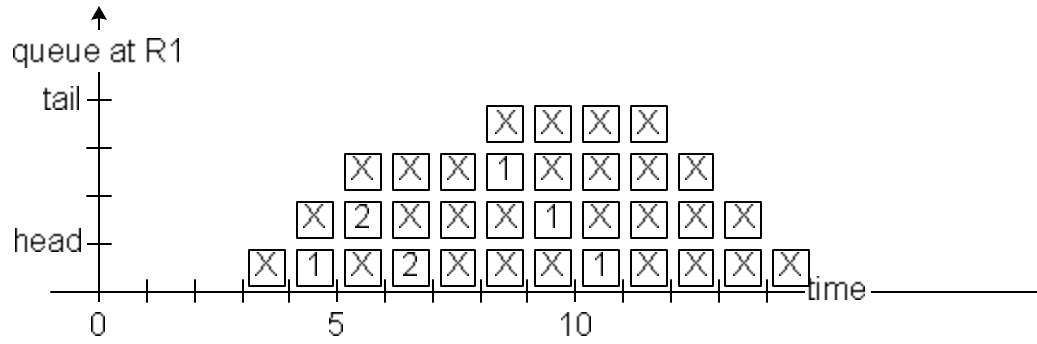
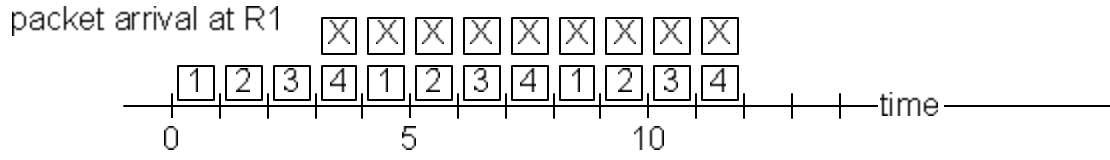
- a) What is the idea behind WaveVideo scaling on active routers? For what type of applications is this mechanism useful?

WaveVideo is a scalable video encoding that can be adapted to the available bandwidth. An active network node can detect congestion and perform the scaling by dropping less important high-frequency packets. It is particular useful for real-time media application that can accept lossy transmissions and multicast applications where receivers have different amounts of available bandwidth.

- b) Assume the following scenario:

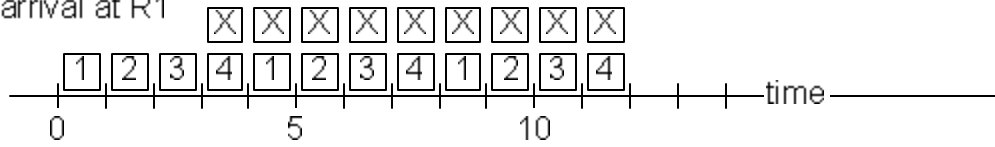


Assume the video transmission starts at time 0 and cross traffic starts at time 3. All sources stop sending at time 11. Then the arrival of packets on router R1 is shown below. Show the state of the packet queue at router R1 (show the packets as boxes). Also, show the WaveVideo packets that arrive at a given time at V2:

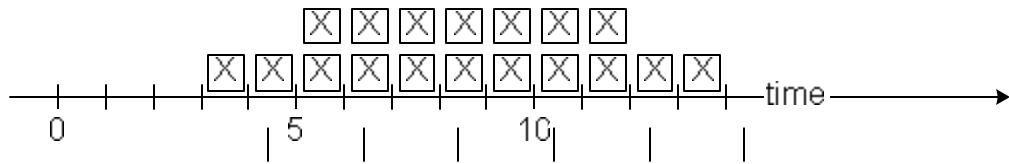


e) Show how your solution works and how it solves the problems you have observed in c) using the above example. You can use the following time lines. Please label clearly what is shown on each time line that you use.

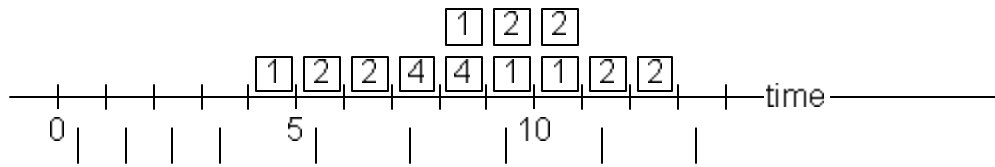
packet arrival at R1



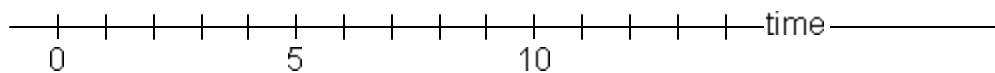
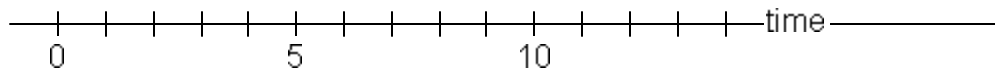
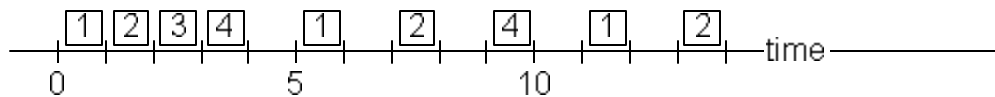
cross traffic queue



WaveVideo queue



packet arrival at V2



Question 5 (10 points):

Active networks and network processors are both approaches to providing flexibility in how network traffic is handled on a router. Discuss and contrast the goals of both research areas. Can the systems developed in these areas be combined easily? Where do you see differences between active networking and network processors?

Active network research aims at developing software environments that support dynamic deployment of network services. Active routers usually perform complex packet processing of novel network protocols and applications. Network processor research aims at the hardware infrastructure for packet processing. Network processors are useful for non-active protocols (e.g., plain IPv4 forwarding) as well as more complex services.

It is difficult to combine both approaches, as active network environments usually are complex and are designed to run on a workstation processor. Network processors have only limited on-chip memory and very simple operating systems.

Active networks are a more extreme approach to introducing new network services. Network processors are more conservative “programmable switches” that also implement traditional protocols.

Question 6 (12 points (3+4+4+1)):

In class, we discussed a paper on “Design Space Exploration of Network Processor Architectures.” Answer the following questions relating to this research area.

a) What is the “design space” in this context?

The choice of configuration of the network processor in terms of number/type of processors, size of on-chip memory, number/type of coprocessors.

b) What process is used to explore the design space?

An analytic performance model is developed.

System constraints are defined.

The optimization criterion is defined.

The system is optimized by varying the configurations.

c) What performance criteria can be used to determine the quality of a point in the design space?

Most processing power,

highest throughput,

least area given processing/throughput,

least power consumption given processing/throughput,

lowest cost given processing/throughput.

d) What alternatives are available to b)?

Simulation

Prototyping

Question 7 (14 points (5+4+5)):

Assume a network processor implements the typical functions of a port on a programmable router.

- a) What scheduling problems appear in the system (i.e., what resource needs to be scheduled for and who/what is competing for the resource and what is the scheduler trying to achieve)? Can they all be solved by the same algorithms? Explain your answer.

Processing: Packets of different flows compete for processing engines, goals of the scheduler: fair sharing, high throughput, low delay

Link: Packets of different flows compete for link bandwidth, goals of scheduler: fair sharing, keeping link filled (avoid underflow)

Switching Interface: Packets assigned to different ports compete for access to switch fabric, goals of scheduler: avoid buffer overflow, low delay

- b) Why are OS scheduling algorithms only of limited use for network processor scheduling problems?

OS schedulers assume that context switches are possible. This is too costly for network processing environments. Also, OS processes do not usually have different weights assigned – different flows might have different weights.

- c) In class, we briefly discussed a scheduling algorithm that is “locality-aware”? What are the ideas behind this scheduling algorithm and why is it useful for network processors?

The idea is to group packets that use the same instruction code to assign them to the same set of processing engines. As a result, instruction code in memory caches can be reused and the overall throughput increases. This is particularly useful for NPs where the overall processing time is short to the time it takes to “warm up” the cache.

Question 8 (10 points):

Consider all research ideas and systems discussed in this course. Which do you think will have the biggest impact on commercial networks (i.e., the Internet)? Support your argument. Which do you think will have least impact or is most unrealistic and why?

No single correct answer.