Engineering Tripos Part IIA THIRD YEAR

Module 3F5: Computer and Network Systems

Solutions to 2011 Tripos Paper

Authors: Andrew Gee and Tim Wilkinson

1. Caches

(a) Cache access is faster than main memory access because of its physical proximity to the CPU and its construction out of static RAM, as opposed to slower dynamic RAM. Data still needs to be fetched from main memory to the cache, but this overhead is easily amortized through temporal locality of reference (so a fetched item is likely to be accessed again soon, and this time it will be in the cache) and spatial locality of reference (so a fetched item's neighbours are likely to be accessed soon, so fetch *blocks* of data at a time, paying the main memory latency price just once). [20%]

(b) (i) The cache capacity is 4 × 4 × 65536 = 1048576 = 1MB. The index is given by

$$\left\lfloor \frac{\text{byte address}}{\text{bytes per block}} \right\rfloor \text{ modulo (blocks in cache)} = \left\lfloor \frac{\text{0xA0973C8F}}{\text{0x10}} \right\rfloor \text{ modulo 0x10000}$$
$$= \text{0xA0973C8 modulo 0x10000} = \text{0x73C8}$$

This is the last byte in the block, so the word and byte offsets are both 3. The tag comprises the remaining upper bits of the address, in this case 0xA09.

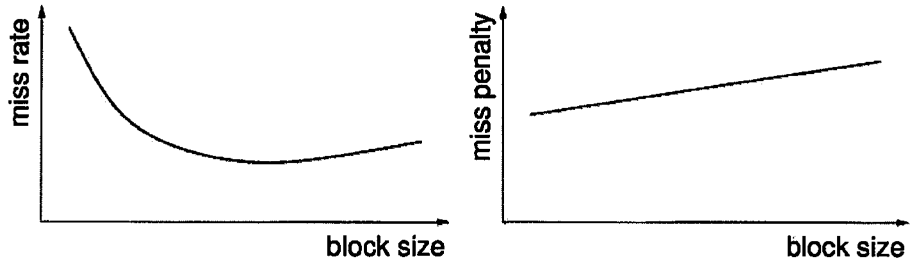
(ii) The cache capacity is 4 × 4 × 4 × 4096 = 262144 = 256KB. The index is given by

$$\left\lfloor \frac{\text{byte address}}{\text{bytes per block}} \right\rfloor \text{ modulo (sets in cache)} = \left\lfloor \frac{\text{0xA0973C8F}}{\text{0x10}} \right\rfloor \text{ modulo 0x1000}$$
$$= \text{0xA0973C8 modulo 0x1000} = \text{0x3C8}$$

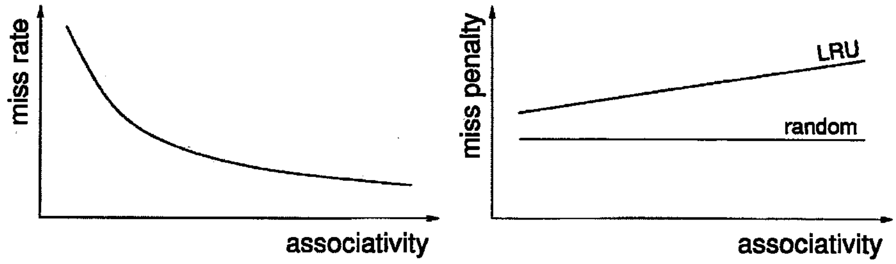
This is the last byte in the block, so the word and byte offsets are both 3. The tag comprises the remaining upper bits of the address, in this case 0xA097. The byte may reside in any of the four blocks in set 0x3C8.

(iii) The cache capacity is 4 × 4 × 32768 = 524288 = 512KB. The byte may reside in any of the cache blocks. The tag would be the entire block address, which is 0xA0973C8 as before. This is the last byte in the block, so the word and byte offsets are both 3. [30%]

(c)

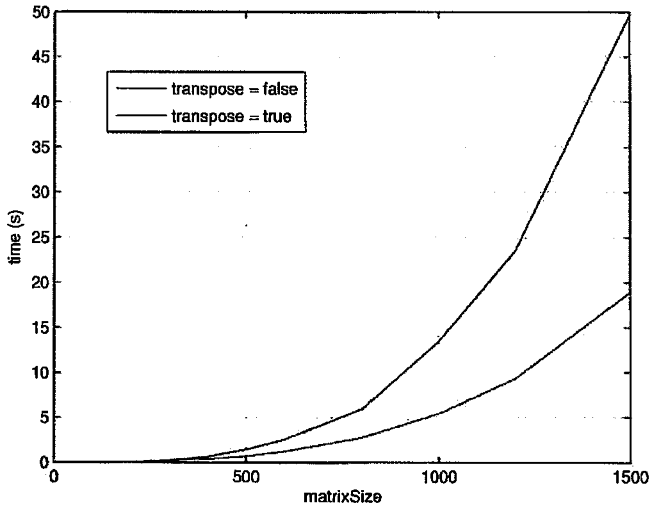


Spatial locality of reference means that the miss rate generally decreases with block size, though with very large blocks the miss rate may eventually increase (too many cache conflicts). The miss penalty increases with block size, since more words need to be transferred from main memory.



The miss rate decreases with increased associativity, since block replacement strategies like LRU can be used to replace blocks which are unlikely to be needed again soon. With LRU, the miss penalty may increase with more associativity, since the LRU algorithm needs to check more access times to decide which block to replace. For random replacement, the miss penalty is likely to be independent of associativity. [25%]

(d) Candidates are expected to deduce the curves' shapes themselves, but for interest here are some actual results from a 3.2 GHz Pentium 4.



There is the expected cubic dependency on `matrixSize`, but with a far lower cost per operation when `transpose = true`. This is because matrices are stored in memory row by row, so the elements of `b` and `ct` are accessed sequentially inside the inner loop, with excellent spatial locality of reference. The same cannot be said of `c`: cache misses account for the relatively poor performance when `transpose = false`. There is a small price to pay in taking the transpose: while it is hard to see on the graph, the curves do in fact cross at around `matrixSize = 225`. [25%]

2. **RISC datapaths**

(a) The distinguishing features of RISC architectures are:

- Fixed length instructions
- Just a few instruction formats
- General purpose register, load-store instruction set architecture
- Limited operations
- Simple addressing modes
- Complete one instruction per datapath cycle (with pipelining)
- No microcode to control the datapath [25%]

(b) (i) No, this datapath could not be used to execute an instruction set that is not load-store. The operands for arithmetic and logic instructions clearly come from the register file only. A non-load-store architecture would require an extra ALU after the data memory unit, so arithmetic and logic instructions could retrieve operands from memory. [10%]

(ii) The instruction rate is limited by the longest path through the datapath, which is for the `lw` instruction. This requires $2+1+2+2+1=8$ns. Switching to the compact ALU/adder design would slow this to $2+1+3+2+1=9$ns, so programs would take 12.5% longer to run. However, we could switch the adders only, with no impact on performance. This is because the output of the first adder would settle after 3 ns, both inputs to the second adder would be stable after 3 ns, so the output of the second adder would settle after a total of 6 ns. The adders would therefore not be on the critical path through the datapath, which would still be the 8 ns `lw`. [25%]

(c) The following changes would need to be made to the datapath.

- The register file needs to be able to write two registers simultaneously, so add a 5-bit Write reg 2 input and a 32-bit Write data 2 input.
- The register to be incremented is specified in the `rs` field, so connect Instruction [25-21] to the new Write reg 2 input.
- The contents of the unincremented register are output on Read data 1. Feed this signal into one input of a new adder, with the constant number 4 on the other input. Connect the output of this new adder to the Write data 2 input of the register file.
- Add a new control input RegWrite2 to the register file. When low, this has no effect. When high, the register identified on the Write reg 2 input is replaced with the value on the Write data 2 input.

- Modify the main control unit to set the new RegWrite2 control signal low for all instructions apart from `lw_inc`, when it should be set high.

The control signals for `lw_inc` should be set as follows:

Signal	Setting
MemRead	High (memory read)
MemWrite	Low (no memory write)
ALUSrc	Sign-extended instruction[15-0]
RegDst	Instruction[20-16]
RegWrite	High (write register)
RegWrite2	High (write register 2)
Branch	Low (replace PC with PC+4)
MemtoReg	High (write memory output to register file)
ALUOp	00 (add independent of instruction[5-0])

We would expect this new instruction to have a noticeably positive impact on performance. The extra adder operates in parallel with the existing ALU and does not extend the longest (`lw`) path through the datapath: we could still clock the datapath at 8 ns. The compiler could take advantage of the new instruction when stepping through data arrays word by word, a fairly common programming procedure (Amdahl's Law). Previously, each word fetch would require two instructions, a `lw` to fetch the word followed by a `addi` to increment the base register. With the modifications, these two instructions could be replaced by a single `lw_inc`, saving 8 ns per iteration. [40%]

Question 3

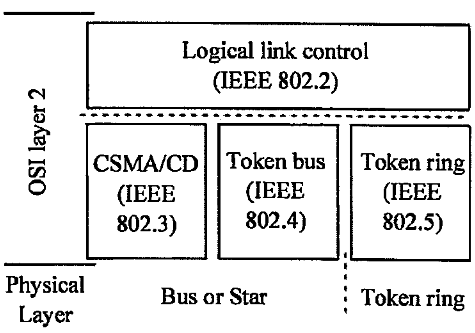
a)The network environment. The protocols and standards relating to the physical communications networks. The lowest three layers (1-3) are network dependent and are concerned with the protocols associated with the data communication network being used to link the two computers. Communication appears to take place **across** each layer. Each data exchange passes down to the bottom layer (the physical layer) at the sending terminal, crosses the network to the receiving terminal and then passes up again.

Layer 3: **The network layer.** This is concerned with the operation of the network between the terminals. It is responsible for establishing the correct connections between the appropriate network nodes, including network routing (addressing) and in some instances, flow control across the computer to network interface. Example – routing protocols, IP address

Layer 2: **The data link layer.** This provides error detection and correction. Provides error free data to the network layer. Adds control information to the data blocks and a frame check sequence (FCS) for error-checking or bits for synchronisation. Retransmits data if errors have occurred. The data-link layer is also responsible for determining the length of data segments to be sent to the physical layer. If the data provided by the network layer is too long, then it is broken into suitable sized packets. Transparency is preserved for the data bits in these blocks. In the case of LANs, the data-link layer provides local area address management on top of that provided by the network layer through medium access protocols (MACs). Example – error correction codes, fragmentation

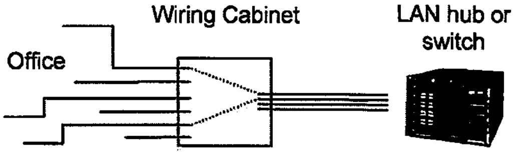
Layer 1: **The physical layer.** This defines an interface in terms of the connections, voltage levels and data transmission rate, in order for signals to be transmitted bit by bit. Example – optical signal, Manchester encoding.

b) The local area network (LAN) emerged in the late 1980s as the most important means of conveying data between different computers and computer peripherals (printer, file server, email server, fax gateway, host gateway, scanner, etc.) within a single office, office building or small campus. LANs are constrained by their mode of operation to a geographically limited area, but are ideally suited to short distance data communications. A high bit speed LAN can carry high volumes of data with rapid response times. The different types of LAN were characterised by their distinctive topologies. They all comprise a single transmission path interconnecting all the data terminal devices, with a bit speed typically between 1 and 30Mbits/s, together with the appropriate protocols (called the logical link control (LLC) and the medium access control (MAC)) to enable data transfer. The LLC and MAC protocols split the data-link layer (Layer 2)

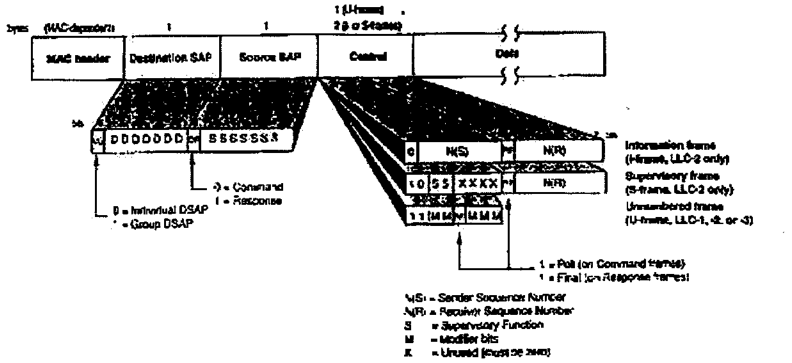


It made sense to define a standard interface between the LAN and the application intended to communicate across it. This standard is the logical link control (LLC) defined by IEEE 802.2 or ISO 8802.2. The LLC provides a standard communication interface equivalent to that provided by OSI layer 2 to OSI layer 3 and uses HDLC calls like those used in the X.25 data-link layer protocols.

c) **Ethernet** was designed as a layer 2 contention based protocol. On a CSMA/CD LAN the terminals do not request permission from a central controller before transmitting data onto the transmission channel; they contend for its use. Before transmitting a packet of data, a sending terminal 'listens' to check whether a path is in use and if so, it waits before transmitting its data. Even when it starts to send data, it needs to continue checking the path to make sure that no other stations have started sending data at the same time. Theory suggests that random collision of a large number of devices can lead to transmission degradation under heavy traffic, however traffic is rarely random as most LAN transmissions involve a central server system. Traffic problems can be alleviated by subdivision into smaller LANs. LAN hubs (or switches) have become necessary as the data rate is too high for shared media. This caused the development of LANs with structured cabling, using hubs and twisted pair (10 and 100baseT), in a star configuration. An Ethernet MAC sub layer frame can take two possible forms. The preamble/SFD, address and frame check sequence (FCS) are common to both types. They are referred to as *type* and *length encapsulation* frame formats. The key to MAC layer protocols is a uniform addressing structure so that LAN hardware can easily identify stations on the network and transmit packets between them. In the evolution of LAN protocols, Ethernet has dominated and has become popular in MANs and WANs. This is due to the advancement of data rates beyond what is capable of buses and the shrinking of the bus into a LAN switch system. This eliminates a large degree of the MAC protocols and



also allows much higher layer protocols such as IP to be used in place of systems ate would once have been hardwired via the LLC sublayer. These have been largely disabled and removed making Ethernet purely a frame format.



d) A voice service cannot tolerate significant delays but is tolerant to date bit errors. The internet protocol is fundamentally a connectionless protocol and cannot offer QoS such as delay free service. Protocols such as VoIP try to offer this but fundamentally cannot over an IP network unless all of the servers and switches in the system are designed to run the particular VoIP protocol. TCP/IP has several features that could be invoked but there is a risk that not all parties in the network connection will obey these in order to gain QoS.

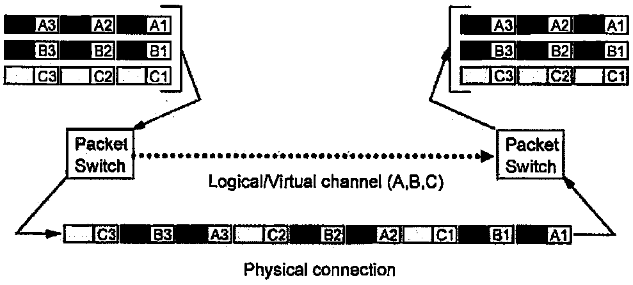
TCP can use features such as port allocation, acks, flags and windowing to try and implement a connection oriented services, but all are flawed if TCP is not used in the intermediate nodes. This is referred to as layer 4 switching and is not always part of IP network backbones. IP itself also has internal QoS mechanisms such as ICMP and or 'Type of Service' fields as used by protocols such as diffserv. These will only work if all of the intermediate nodes are enabled to the same protocol system, which cannot be guaranteed over a native IP network. In fact, these features can be a risk and delay a packet if the router is using the fast Path algorithm to parse its packet stream as the added value packets are removed from the fast path.

It is also possible to implement a connection oriented pseudoservice via layer 2 and the LLC part of Ethernet. This is also a problem, however as most intermediate nodes in an IP network do not interpret the bit fields in the layer 2 frame structure.

Question 4

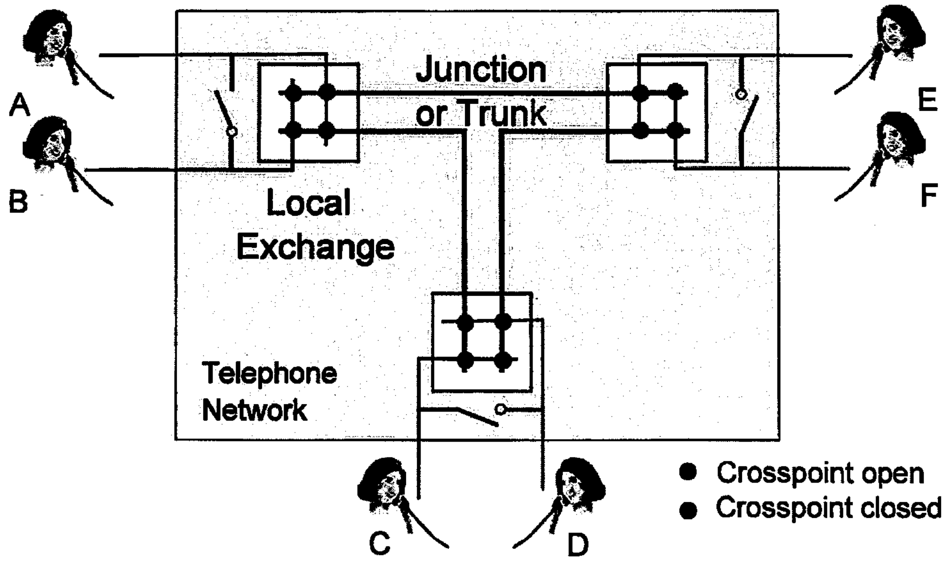
a) Circuit switched networks are examples of Connection Oriented network systems. In a connection oriented technique a circuit, virtual circuit, connection or virtual connection (VC) is established between sender and receiver before information is transported. A lot of time division based circuit switched networks are connection oriented as the users are connected by a virtual connection that is actually sent as a sequential line of timed framed multiplexed through a physical media common with other users such as used in SDH. A voice service is simple as the connection oriented nature means that delay free calls can be guaranteed. The problem is that due to the bursty nature of data traffic, the efficiency of data traffic will be low as the setting of a connection will be for a fixed time which should be longer than the bursts of traffic.

Packet switching is so called because the user's overall message is broken up into a number of smaller packets, each of which is sent separately. The receiving end re-assembles the packets in their proper order, with the aid of sequence numbers and the other PCI fields. Each packet is carried across the network in a store and forward fashion, taking the most efficient route available at the time. Problems arise when more than one or all transmitters try to send packets at once. This is accommodated by buffers at each end of the connection. These delay some of the simultaneous packets in a first in first out (FIFO) system until the line becomes free.



Being connectionless is ideally suited to needs of data traffic on packet switched networks. Connections are only needed for the duration of data and leads to very efficient data services. Voice data is more difficult as the penalty of packet switching is that if a connection is not available at that instant, then the packet will be buffered and delayed.

b) In this example each of the crossbar switches (labelled exchanges) is shared by a number of stations, each of which is switched and connected to the exchange by a local line or local loop. This example now resembles a public switched telephone network (PSTN). The lines are referred to at junctions or trunks. In a real exchange, the numbers of exchanges and their locations are governed by the overall number and geographical density of the telephone users. The number of junctions or trunks will be made sufficient to cater for normal telephone demand. The network below is classed as having full availability as any of the stations can be connected to any of the available junctions. For six users in three groups, the layout is as below using 3 junctions.



c) There are several different configurations for a users. Below is one, with local exchanges having 2x2 switch matrices and the central exchange being either 4x2x2 or 8x2 switches. This system uses 8 junctions for full availability. If only 4 were used, then the availability would depend on local users not calling simultaneously. The addition of the dotted crossbar lines below would further increase availability and reduce the complexity of the engagement signalling.