PART IIA 2007
3F5 Computer and network systems
Dr A Gee

Engineering Tripos Part IIA THIRD YEAR

Module 3F5: Computer and Network Systems

Solutions to 2007 Tripos Paper

1. **Caches and TLBs**

(a) The two properties are:

Temporal locality of reference: if an item is referenced, it tends to be referenced again soon (loops, local variables).

Spatial locality of reference: if an item is referenced, items with nearby addresses will tend to be referenced soon (instructions, data in arrays) [10%]

(b) If the page size is 4 KBytes ($= 2^{12}$), the page offset must be 12 bits, leaving 20 bits for the virtual and physical page numbers. The TLB has 64 entries and therefore 32 ($= 2^5$) sets. The index into the TLB is therefore 5 bits, leaving the remaining 15 bits of the virtual page number to form the TLB tag. The primary data cache is 16 KBytes large, which is 4K words or 1K ($= 2^{10}$) blocks. The index into the cache is therefore 10 bits. There is a block offset of 2 bits and a byte offset of 2 bits, leaving 18 bits for the cache tag. A hardware schematic of the TLB and cache can be found on the next page. [50%]

(c) The clock cycle time is $1/(1\text{GHz}) = 1$ ns. So the cost of accessing main memory is 100 clock cycles. The effective CPI with a single level cache is thus:

$$\begin{aligned}\text{effective CPI} &= \text{baseline CPI} + \text{memory stall cycles per instruction}\\ &= 1.0 + 100 \times 5\% = 6.0\end{aligned}$$

The cost of accessing the secondary cache is 10 clock cycles. The effective CPI with a two-level cache is thus:

$$\begin{aligned}\text{effective CPI} &= \text{baseline CPI} + \text{secondary cache stall cycles per instruction}\\ &\quad + \text{memory stall cycles per instruction}\\ &= 1.0 + 10 \times 5\% + 100 \times 40\% \times 5\% = 3.5\end{aligned}$$

The speedup is therefore $6.0/3.5 = 1.71$ times. We have assumed that the main memory access time, in particular the component associated with cache miss handling, remains the same irrespective of which cache it is servicing[1]. [30%]

(d) The primary cache filters accesses to the secondary cache, especially those with good spatial and temporal locality of reference. Hence, accesses to the secondary cache tend to have poor locality of reference, resulting in a high miss rate at this level. [10%]

[1] If the secondary cache has a larger block size than the primary cache, one might suppose that its miss penalty would be larger. However, there are schemes like *early restart* and *critical word first* that allow the CPU to restart before the entire block has been transferred from main memory to the cache.

2. **Multiprocessor systems**

(a)

SIMD, Single Instruction Stream Multiple Data Streams. A computer classification in Flynn's taxonomy of parallel processing machines. Multiple execution units respond to the same instruction at the same time, but operate on different data. Useful for vector processing, nowadays commonly implemented in graphics hardware.

MIMD, Multiple Instruction Streams Multiple Data Streams. A computer classification in Flynn's taxonomy of parallel processing machines. Multiple uniprocessors connected on a single bus or via a network. The most general form of parallelism.

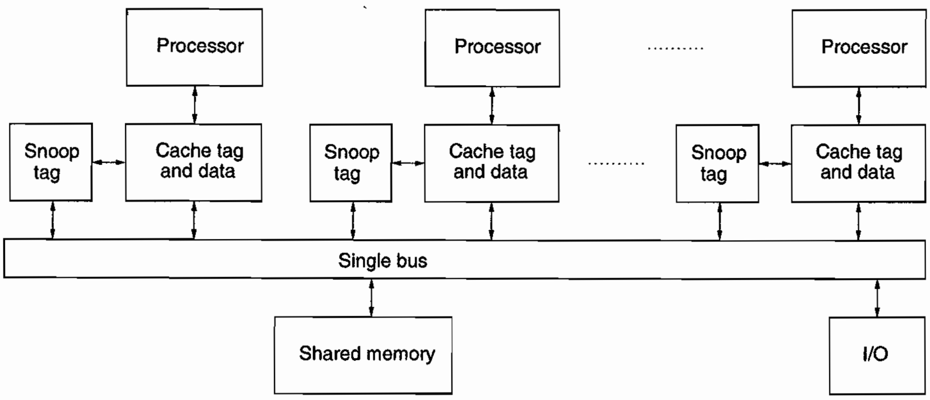
SMP, Symmetric Multiprocessor. A type of single address space multiprocessor in which accesses to main memory take the same amount of time no matter which processor requests the word and no matter which word is requested.

UMA, Uniform Memory Access. Means the same thing as symmetric multiprocessor (SMP).

NUMA, Nonuniform Memory Access. A type of single address space multiprocessor in which some memory accesses are faster than others depending on which processor asks for which word.

SMT, Simultaneous Multithreading. An operating mode for superscalar hardware that allows several processes to run concurrently, with instructions from distinct processes fetched in the same clock cycle. Compared with conventional process switching, the superscalar hardware spends far less time idling since there are fewer dependencies between the running instructions. [35%]

(b) The most common method is bus snooping.



All cache controllers monitor the bus and react to reads and writes in other caches. There are two ways of dealing with writes:

Q3 Crib (longer than expected from candidates)

a) Voice services: Voice services are intolerant to any form of delay (fixed or variable). If there is any significant delay (more than 20-30msec) then reflections can lead to echoes being generated. Voice services are tolerant to transmission errors. We can put up with considerable distortion to a voice signal, hence the tolerable bandwidth limits of voice are set at 3.4kHz. Assuming that at some point in its transmission, voice signals are digitised and sent as binary data, then a bit error rate of 1 error in 10^5 bits is tolerable.

An example of a voice service network is SDH. The delay minimum is guaranteed by offering a connection oriented service with guaranteed bandwidth in every 125usec STM1 frame for the entire length of the call. Network control is external to the data.

Data services. Data services are highly sensitive to bit errors. For this reason, data services like electronic banking require bit error rates as high as 1 bit error in 10^{12} (at a data rate of 155Mbits/sec this is one bit error every 1.8 hours!). Data networks are much more tolerant to delays in transmission. Often there is no penalty for delay as long as the message gets there.

An example of a data service is frame relay (or IP) where errors are controlled through CRC and FCS sections on each packet. Retransmission is possible if data is corrupted.

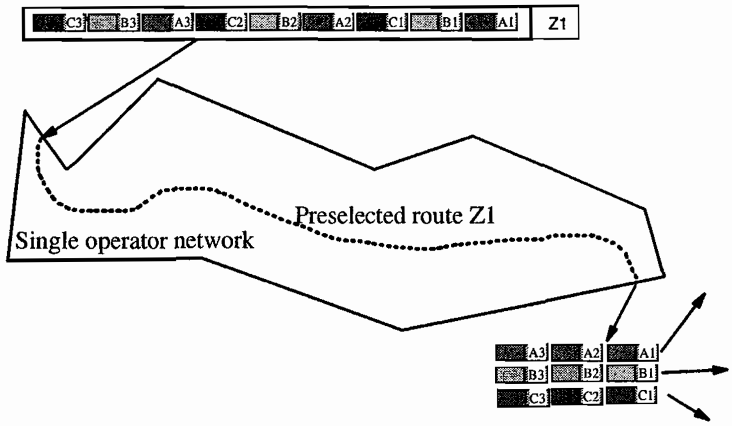
b) How to transmit broadcast quality video without delay or error. This is difficult when interactive services such as voice and video are offered as any delay at all is difficult to accept. The requirements for bit errors is also high as there is no point in downloading corrupted data. Hence there is a high demand on networks by these services in terms of both delay and transmission quality.

Broadcasting video over SDH is a good way of guaranteeing the delay requirements, but it is not a very efficient or cost effective means of delivering data at high rates to many different users as the control mechanisms for SDH are based on a connection oriented service model which is normally one to one. If SDH is used to broadcast then it makes the control very complex. If IP or FR is used to transmit video data, then the system is very reliable as long as the network is reliable and not under heavy usage. However neither service has a good mechanism for supplying a delay free service, which means that the service relies heavily on other network operations at the same time. There are priority based IP services available but they require all of the nodes in the IP network to use the same type of service which is often very rare when networks are interconnections between different manufacturers and service providers. For such a service to be delivered a new end to end packet system is ultimately required.

c) A key concept on routing IP packets quickly is the Fast Path. Packets are vetted to see if they will delay those that follow them. Packets that use fragmentation or that involve the routing mechanism itself are taken off the fast path. ICMP packets use the options field in the IP packet to implement more sophisticated services. These are not included in the fast path. The diagram below shows the fast path mechanism in its fully structure.

The protocol indication determines whether TCP or UDP protocol is used in the next higher layer, and therefore determines to which protocol agent, the datagram should be delivered. . The header checksum is akin to a cyclic redundancy check (CRC) or frame check sequence (FCS). It is only applied to the IP header and is recomputed each time the header is amended during passage through the network. The options field indicates the use of ICMP which is not on the fast path. The source and destination addresses are both 32 bits.

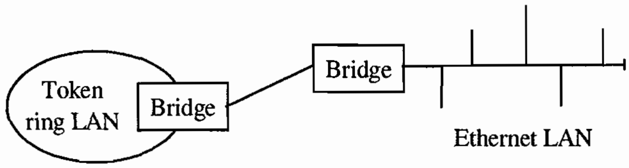
d) Tagging or label switching tries to group packets through a common network together with a common global header (or label or tag) which gets the group of packets to the other side of the network with minimum delay.



This puts a lot of pressure on the routers at the edge of the network to find suitable packet groups and also to link together common service types. This would be an efficient means of broadcasting video as the packets have a single common sources and could be distributed locally by the edge routers with low latency using a broadcast IP address structure. The main problem with this sort of network structure is that it requires all of the nodes in the label system to use the same address and priority mechanism and hence probably come from the same manufacturers. Label based networks are almost always proprietary systems from a single company.

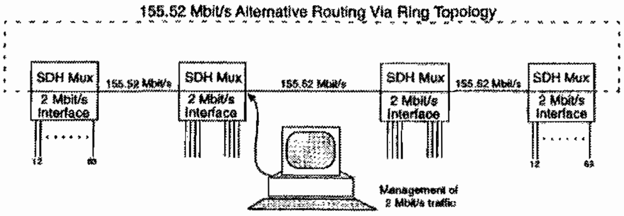
Q4 crib (longer than expected from candidates)

a) Aside from being small and fragile, optical fibre is the perfect transmission medium offering an attenuation of less than 1dB/km over a wavelength range of 700nm. This means that we could modulate an optical carrier at data rates beyond 100Tbits/sec (100, 000Gbits/sec) and still have attenuation less than 1dB/km. Optical fibres are virtually immune (unless you tie them in a knot!) to external interference and do not produce any radiation that could cause interference.

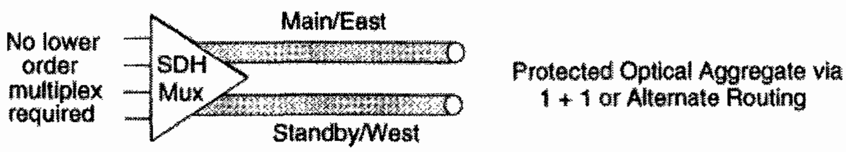


Modern LANs such as Gigabit Ethernet and the new 10G standards are struggling to utilise copper based physical wiring, hence fibre is also part of these standards. The bandwidth now justifies the expense of the lasers and connectors making them a commercially viable LAN technology.

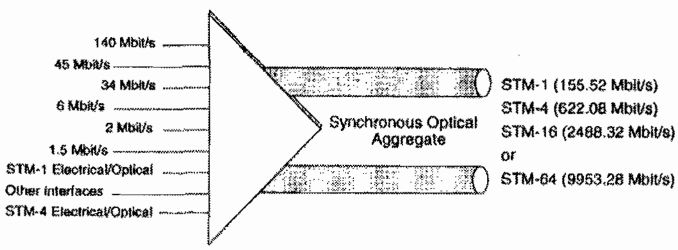
c) An SDH network is normally made from a ring of synchronous muxes, interconnected with SM fibre pairs in a East/West fashion to provide redundancy should the ring be broken.



No longer need to climb the mux mountain. We can now provide synchronisation and network management to the SDH network, which is entirely software controllable. Networks can be invisibly configure if there is a failure. Resources can be dynamically allocated to allow maintenance. Network capacity can be dynamically allocated to allow high bandwidth users like video conferencing or LAN access. Room for expansion and future-proofing.



The synchronous mux (basic function only), ITU-T defined. Allows optical integration at the STM-N level. Optical interfaces are often repeated for redundancy as main/standby or East/West pairs. Often in a ring topology. Must offer digital mux to STM-1 with flexible inputs.



The role of the mux in add/drop mode is central to SDH as it allow any mux to be added or dropped without having to demux the other\data streams in the mux.
d) One possible means of putting packet data over SDH section of the B-ISDN standard called asynchronous transfer mode or ATM. An ATM packet is a fixed cell of 48 bytes with a 5 byte header which can be mapped directly into an STM-1frame.

Assessors' remarks for question 1: A popular question testing candidates' understanding of caches and virtual memory systems. In (a), almost everyone knew that caches are motivated by spatial and temporal locality of reference. There were some good answers to (b), though many candidates struggled with the details (ie. precisely how wide each address field is). Most candidates got the gist of (c), though again the details proved problematic (eg. forgetting that the memory access time is *additional* to the baseline CPI, not instead of it). In (d), a pleasing number of candidates identified why secondary cache miss rates are much higher than primary cache miss rates.

Assessors' remarks for question 2: This essay-style question tested the candidates' understanding of multiprocessor systems. Most candidates knew the basics, especially when it came to cache coherency protocols, though there was some confusion about UMA/NUMA architectures. As usual with essay questions, the best answers were distinguished by sound editorial judgement: they highlighted the key points and, just as importantly, omitted irrelevancies and trivia. The average mark was distorted by a small number of very incomplete answers.

Assessors' remarks for question 3: Part (a) was basic bookwork: most candidates produced good answers with a few missing that a switched network is a class, not a type. Part (b) was done well as it required a bit of thought about the role of multimedia services. Part (c) was the worst answered, with many forgetting the role of the fast path and making something up. Part (d) was speculative but most did quite well.

Assessors' remarks for question 4: Part (a) was entirely bookwork: most candidates got it right with a few getting multimode and single mode properties the wrong way round. Part (b) was not so well answered, with most missing the cost issues of multimode fibre. Part (c) was well answered with a few nice diagrams produced. Part (d) was not well answered, with most missing the connection between IP and the connection-oriented nature of SDH.

Andrew Gee & Tim Wilkinson
May 2007

Part IIA 2007

Module 3F5: Computer and Network Systems

Numerical Answers

1. (d) 1.71 times faster.