Chapter 4: Notes on Address Translation Mechanism and TLB

Fig on Slide 82 shows the mechanism for virtual memory address translation. Each virtual address is generated by the processor, whether it is an instruction fetch or an operand fetch/store operation, is interpreted as a virtual page number (higher order bits) (VPN) followed by an offset (low-order bits) that specifies the location of a particular byte (or word) within a page. Information about the main memory (MM) location of each page is kept in a page table (PT). This information includes the MM address where the page is stored and the current status of the page. An area in the MM that can hold one page is called a page frame. The starting address of the page table is kept in a page-table register (PTR). By adding the VPN to the contents of this register, the address of the corresponding entry in the page table is obtained. The contents of this location give the starting address of the page, if that page currently resides in the MM.

Each entry in the PT also includes some control bits that describe the status of each page while it is in MM. One of the bits indicate the validity of the page, i.e., whether the page is actually loaded in the MM. This bit allows the OS to invalidate the page without removing it actually.

The PT information is used by the Memory Management Unit (MMU) for every read and write access, so ideally, the PT should be placed within the MMU. Unfortunately, the PT may be rather large, and since MMU is normally implemented as a part of CPU chip (along with the primary cache), it is impossible to include a complete PT on this chip. Therefore, the PT is kept in the MM ! However, a copy of small portion of this PT can be accommodated within MMU. This portion of PT consists of the PT entries that are most recently accessed pages. A small cache, usually referred to as Translation Lookaside Buffer (TLB), is incorporated into the MMU for this purpose. The operation of TLB w.r.t the PT in the MM is essentially same as the operation we have discussed for the cache memory mechanisms. In addition to the information that constitutes a PT entry, the TLB must also include the virtual address of the entry. Figure on Slide 86 shows a possible organization of a TLB, where associative mapping technique is used.

How does the address translation scheme work? Given a virtual address, the MMU looks in the TLB for the referenced page. If the PT entry for this page is found in the TLB, the physical address is obtained immediately. If there is a miss in the TLB, then the required entry is obtained from the PT in the MM, and the TLB is updated. When a program generates an access request to a page that is not in the MM, a page fault is said to have occurred. The whole page must be brought in from the disk in to the memory before access can proceed. When it detects a page fault, the MMU asks the OS to intervene by raising an exception (interrupt). Processing of the active task is interrupted, and the control is transferred to the OS. The OS then copies the requested page from the disk to the MM and returns control to the interrupted task.

Concepts similar to LRU are used in this case, when a new page is brought in and we need to

provide sufficient space for it.