Memory Capacity Planning - 1

The Performance of a memory Hierarchy is determined by the

Effective access time (T_{eff}) to any level in hierarchy. It depends on hit ratio and access frequency.

Hit ratio :- When an information item is found in M_i, we call it a hit, otherwise, a miss. For a memory hierarchy ($i = 1$ to n) the hit ratio h_i at M_i is the probability that an information item will be found in M_i.

~~It is a function~~ The miss ratio at M_i is defined as $1 - h_i$

CPU always accesses M_1 first and the access to the outermost memory M_n is always a hit.

The access frequency of M_i is defined as $f_i = (1-h_1)\cdot(1-h_2)\cdots\cdot(1-h_{i-1})\cdot h_i$. This

This is indeed the probability of successfully accessing M_i, when there are $i-1$ misses at the lower levels and a hit at M_i.

$$\sum_{i=1}^{n} f_i = 1, \quad f_1 = h_1$$

Due to the locality property, the access frequencies decrease very rapidly from low to high

$$\underset{\text{More often}}{f_1} \gg f_2 \gg f_3 \gg \cdots \gg \underset{\text{Less often}}{f_n}.$$

$$T_{eff} = \sum_{i=1}^{n} \underbrace{f_i}_{\text{Access frequency at } M_i} \cdot \underbrace{t_i}_{\text{Access time at } M_i}$$

$$= h_1 t_1 + (1-h_1) h_2 t_2 + \cdots + (1-h_{n-1}) t_n$$

$\boxed{\text{Not using } h_n \text{ because } h_n = 1}$

~~The~~ Every time a miss occurs, a penalty must be paid to access higher level memory.

Misses called block miss in cache and pagefault in M.M

M.C-P-3

C.M M.M HDD

$t_1 < t_2 < t_3$.

Stone (1990) has pointed out the a cache miss 2 to 4 time as costly as a cache ~~hit~~ hit.

Hierachy Optimization!-

The total cost of memory hierachy is estimated as cost/Byte.

$$C_{total} = \sum_{i=1}^{n} C_i \cdot S_i$$ size

Means that cost is distributed over n levels $C_1 > C_2 > C_3 > \cdots > C_n$, we have to choose $S_1 < S_2 < S_3 \cdots < S_n$.

The optimal design of a memory hierachy should result in a Teff close to the t_1 of M_1 and total cost to the C_n of M_n

M.C.P -4

So its a optimization Problem given ceiling C_0 on the total cost.

Minimize $$\boxed{T_{eff} = \sum_{i=1}^{n} f_i \cdot t_i}$$

Subject to constraints

$$s_i > 0, \ f_i > 0 \quad \forall\ i = 1, 2, \ldots, n$$

$$\boxed{C_{total} = \sum_{i=1}^{n} c_i \cdot s_i < C_0}$$

Example

M_i	Access time	Capacity	Cost/Kbyte
Cache	$t_1 = 25$ NS	$S_1 = 512$ KB	$C_1 = 1.25$
M.M	$t_2 = ?$	$S_2 = 32$ MB	$C_2 = 0.2$
Disk	$t_3 = 4$ MS	$S_3 = ?$	$C_3 = 0.0002$

M.C.P - 5

Goal to achieve ~~t~~ = $T_{eff} = 10.04\ \mu s$

with hit ratio $h_1 = 0.98$, $h_2 = 0.9$;

$C_0 = 15000$,

$C = C_1 S_1 + C_2 S_2 + C_3 S_3 \leq 15000$

$S_3 = 39.8$ G.B

$t_2 = ?$

$T_{eff} = h_1 t_1 + (1-h_1) h_2 t_2 + (1-h_1)(1-h_2) h_3 t_3 \leq 10.04$

$10.04 \times 10^{-6} = 0.98 \times 25 \times 10^{-9} + 0.02 \times 0.9 \times t_2$

$+ 0.02 \times 0.1 \times 1 \times 4 \times 10^{-3}$

$= 903$ ns.

" Now Let suppose someone double the size of M.M that what would happen"?

M.C.P - 5

Goal to achieve $t = T_{eff} = 10.04\ \mu s$

with hit ratio $h_1 = 0.98$, $h_2 = 0.9$;

$C_o = 15000$,

$C = C_1 S_1 + C_2 S_2 + C_3 S_3 \leq 15000$

$S_3 = 39.8$ G.B

$t_2 = ?$

$T_{eff} = h_1 t_1 + (1-h_1) h_2 t_2 + (1-h_1)(1-h_2) h_3 t_3 \leq 10.04$

$10.04 \times 10^{-6} = 0.98 \times 25 \times 10^{-9} + 0.02 \times 0.9 \times t_2$

$+ 0.02 \times 0.1 \times 1 \times 4 \times 10^{-3}$

$= 903$ ns.

" Now Let suppose someone double the size of M.M that what would happen"?