Úvod do Virtual Memory Managementu

Martin Očkaják, mandos@hq.sk

1 Ciele a problémy

1.1 Pamať (memory)

- veľkosť (size)
- čakacia doba (latency)
- priepustnosť (throughput)

1.2 Pamäte je vždy príliš málo a je príliš pomalá !

1.3 Pracovná sada (working set)

- pamät ktorú bude program používať v blízkej budúcnosti

1.4 Pamäťová hierarchia (memory hierarchy, storage hierarchy)

- registre procesora (veľkost rádovo 100b)
- vyrovnávacie pamäte procesora (veľkost rádovo 100Kb)
- hlavná pamäť (veľkost rádovo 100 Mb)
- disk (veľkost rádovo 100Gb)

1.5 Hlavná pamäť (main memory)

- stredná časť pamäťovej hierarchie (obvykle RAM)

1.6 Cieľ správy pamäte

- zabezpečit aby bola pracovná sada v čo najrýchlejšej pamäti

1.7 Problémy

- vhodná migrácia dát medzi rôznymi úrovňami pamäťovej hierarchie
- čo najmenšie používanie najpomalších pamätí
- vhodné využitie prenosu vačších blokov dát pri pamätiach s veľkou čakacou dobou ale veľkou priepustnosťou
- predvídanie zmien pracovnej sady
- ochrana pamäte
- zaručenie behu systému aj v prípade preťaženia pamäťového subsystému

2 Základné pojmy

2.1 Vyrovnávacia pamäť (cache memory)

- úspešnosť zásahov (hit ratio)
- čas prístupu (access time)
- lokalita prístupov (locality of reference)
  - rozsah pamäte do ktorej program pristupuje počas krátkeho casu
- koherencia (cache coherency)

2.2 Stratégia vyrovnávacej pamäte (cache policy)

- načítavacia (fetch policy)
- nahradzovacia (eviction policy)
- zapisovacia (write policy)

2.3 Prideľovanie pamäte (allocation)

- fragmentácia (fragmentation)
  - vnutorná (internal)
  - vonkajšia (external)

2.4 Uvoľnovanie pamäťe (deallocation)

- spájanie voľných blokov (coalescing)

2.5 Virtuálna pamäť (virtual memory, VM)

- pracuje s virtuálnymi adresami, ktoré správa virtuálnej pamäte preklada na fyzické

- stránka (page)
  - použitá (used, accessed)
  - zmenená (modified, dirty)
- mapovanie (mapping)
- tabuľka stránok (page table)
  - viacurovňová (multi-level)

- výpadok stránky (page fault)
- vyrovnávacia pamäť stránok (page cache)
- odkladací priestor (swap space, backing store)
- jednotka správy pamäte (memory management unit, MMU)
- vyrovnávacia pamäť na preklad adries (translation lookaside buffer (TLB))

3 Prostriedky správy pamäte

3.1 Prideľovanie a uvoľňovanie pamäťe (allocation and deallocation)

3.2 Nahradzovacie stratégie (replacement policies)

3.3 Spätný zápis do nižších urovní pamäťovej hierarchie (write-back)

3.4 Odswapovanie (page out, swap out)

3.5 Namapovanie virtuálnej pamäťe, ktorá nie je fyzicky pridelená (overcommiting)

3.6 Načítavanie stránky na požiadanie (demand paging)

3.7 Replikácia stránky na požiadanie (copy on write)

3.8 Predbežné načítavanie dát (read ahead)

3.9 Pozastavovanie zápisov (write throttling)

3.10 Zhlukovanie prístupov (access clustering)

3.11 Rušenie procesov zaťažujúcich pamäťový subsystém v prípade nedostatku pamäťe (out of memory (OOM) condition process killing)

4 Algoritmy

4.1 Prideľovanie a uvoľnovanie pamäťe (allocation and deallocation)

- blok presnej veľkosti (perfect fit)
- prvý dostatočne veľký blok (first fit)
  - zoradenie podľa adries (adrees ordered)
  - zoradenie podľa kruhového zoznamu (first in first out (FIFO) ordered)
  - zoradenie podľa zásobníka (last in first out (LIFO) ordered)

- oddelené zoznamy blokov rôznych rozmedzí veľkosti (segregated fit)
  - striktne oddelené zoznamy blokov pevných veľkostí (strict segregated fit)
  - umožnuje rozdeliť blok na dva menšie
    - buddy system (buddy system)
      - spája iba susedné rozdelené bloky
      - dvojitý (double buddies)
      - vážený (weighted buddies)
      - Fibonacciho (Fibonacci buddies)

- najmenší možný dostatočne veľký blok (best fit)
  - použitie indexovacej štruktúry (indexed fit)
- vhodný dostatočne veľký blok (good fit)
  - aproximácia best fit

4.2 Nahradzovanie vo vyrovnávacej pamäti a swapovanie (replacement and paging)

- režimy vzhľadom na proces
  - lokálny (local mode)
  - globálny (global mode)
- hodinový algoritmus (clock algorithm)
  - v prvom prechode označí všetky stránky ako nepoužité a v druhom prechode stránky ktoré odvtedy neboli znovu použité označí ako kandidátov na odswapovanie

- dvojručičkový hodinový algoritmus (double-handed clock algorithm)
  - druhý prechod sa začne skôr ako sa prvý skončí
- kruhový zoznam (first in first out (FIFO))
  - nahradí stránku ktorá je v pamäti najdlhšie
- dokonalý (perfect)
  - nahradí stránku, ktorá zostane v budúcnosti najdlhšie nepoužitá (nerealizovateľné)

- najdlhšie nepoužitá (least recently used (LRU))
  - nahradí stránku, ktorá bola najdlhšie nepoužitá
- nepoužité (not recently used (NRU))
  - nahradi niektorú zo stránok, ktoré neboli použité za určitý časový interval
- najmenej používaná (least frequently used (LFU))
  - nahradi stránku, ktorá bola najmenej používaná za určitý časový interval

- starnutie stránok (page aging)
  - forma váženého LFU - použitie stránky v blízkej minulosti má vačšiu váhu ako použitie dávnejšie v minulosti
- rôzne vylepšenia
  - spiace procesy
  - streaming data

5 Ladenie

5.1 Vyhodnocovanie správania sa programov pri prístupe do pamäťe

- referenčný reťazec (reference string)
  - označuje prístupy k rôznym častiam pamäti podľa času behu procesu

5.2 Žiadny proces nemá v pamäti svoju pracovnú sadu a systém trávi viac času obsluhovaním výpadkov stránok ako vlastným behom procesov (thrashing)

- detekcia
  - pomocou odhadu súčtu pracovných sád procesov
- zabránenie
  - ak sa súčet blíži veľkosti hlavnej pamäťe, systém prestane prideľovat ďalšiu pamäť

6 Implementácie

6.1 Linux

6.2 FreeBSD

6.3 NetBSD

6.4 Windows NT

7 Zaujimavé odkazy

- http://www.memorymanagement.org
- http://cne.gmu.edu/modules/vm/
- http://citeseer.nj.nec.com/OperatingSystems/MemoryManagement/
- http://www.linux-mm.org
- http://www.freebsd.org/doc
- http://www.netbsd.org/Documentation

Ďakujem za pozornosť