

Programska podrška mjernih i procesnih sustava

Branko Jeren i Predrag Pale

Fakultet elektrotehnike i računarstva
Zavod za elektroničke sustave i obradbu signala

Distribuirani sustavi

višeprosesorski, paralelni, mrežni i distribuirani

Čemu to ?

- računalni sustav u središtu ima CPU
- brzina sustava izravno ovisi o brzini CPU
- smanjenje cijene i složenosti primjene CPU
- LAN

Kakvi sve ?

LS&S

- višeprocorski
 - procesori na raznim hijerarhijama i zadaćama
- paralelni
 - ista hijerarhija, ista kutija
 - ravnopravan pristup svim resursima
- distribuirani
 - potpuni sustavi
 - različiti ili isti
 - povezani porukama
- mrežni
 - dijeljenje podataka
 - dijeljenje uređaja

B.Jeren i P.Pale: Programska podrška mjernih i procesnih sustava

PMPS © 1998-2003

Prednost distribuiranog sustava (pred centraliziranim)

LS&S

- ekonomska
 - više manjih CPU je jeftinije nego jedan jači
- brzina
 - više sporijih procesora može biti brži od praktičnih i teoretskih granica brzine jednog procesora
- inherentna distribuiranost
 - proizvodni odjeli, poslovnice
- pouzdanost
 - kvar dijela sustava ne paralizira cijeli sustav
- inkrementalni rast
 - kod jednog CPU moramo ga "baciti" i kupiti novi

B.Jeren i P.Pale: Programska podrška mjernih i procesnih sustava

PMPS © 1998-2003

Prednost distribuiranog sustava (pred nakupinom nezavisnih PS i WS)

LS&S

- dijeljenje podataka
 - na kopijama se ne vide promjene koje upravo nastaju
 - rezervacija karata, banke
- dijeljenje uređaja
 - printeri, archive
- komunikacija
 - pojednostavnjuje rad
- fleksibilnost
 - teret posla može se (privremeno) rpebaciti na drugi sustav

B.Jeren i P.Pale: Programska podrška mjernih i procesnih sustava

PMPS © 1998-2003

Nedostaci distribuiranih sustava

- složeni softver
- mreže
 - kapaciteti
 - nepouzdaost
 - nesigurnost
- sigurnost
 - veće mogućnosti da ju i veću mogućnost zloupotrebe

Organizacijski koncepti (Flynn 1972)

- SISD (Single Instruction Stream and Single Data Stream)
 - klasični jednoprocorski sustavi (uključujući PC)
- SIMD (Single Instruction Stream and Multiple Data Stream)
 - ista se naredba izvodi na različitim podacima
 - numerički proračuni (array processors)
- MISD (Multiple Instruction Stream and Single Data Stream)
 - istovremene različite naredbe na istim podacima
 - do sada ne postoji takvo računalo
- MIMD (Multiple Instruction Stream and Multiple Data Stream)
 - zapravo nakupina nezavisnih računala

MIMD u odnosu na dijeljenje memorije

- multiprocesori
 - imaju zajedničku memoriju
- multikomputeri
 - nemaju zajedničku memoriju

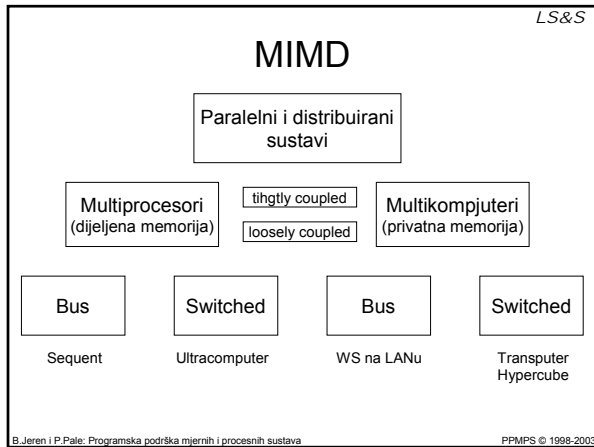
LS&S

MIMD

u odnosu na međusobnu povezanost

- bus
 - sabirnicom
- switched
 - veznom matricom
- tightly coupled
 - vrlo brzo prenošenje podataka
 - obično multiprocesori
- loosely coupled
 - relativno sporo prenošenje poruka
 - obično multikomputeri

B.Jeren i P.Pale: Programaska podrška mjernih i procesnih sustava PMPS © 1998-2003



LS&S

Bus multiprocesori

sabirnica

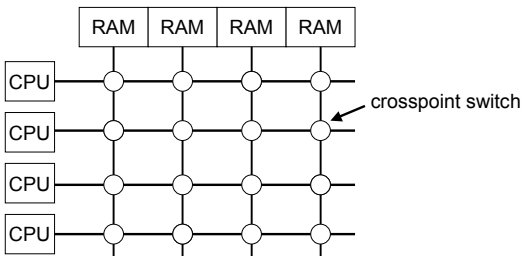
- svi CPU preko sabirnice pristupaju *istom* RAMu
- koherentnost
 - ono što jedan CPU zapíše, svi odmah vide
- usko grlo je sabirnica => cache (spremište)
- “hit rate”
 - vjerojatnost da se potrebni podaci nalaze u cache-u
- problem inkoherenosti => write-through cache
- snooping cache
 - osvježava svoje podatke čim ih netko mijenja

B.Jeren i P.Pale: Programaska podrška mjernih i procesnih sustava PMPS © 1998-2003

Bus multiprocesori - primjer

- SGI power challenge
- Sabirnički sutav se sastoji od 4x POWERpath-2 busa
 - svaka sabirnica ima propusnost 320MB/s
- 18 procesora tipa MIPS R8000, 64-bit RISC
- 16GB memorije
- OS=IRIX
- <http://www.crpc.rice.edu/NHSEreview/ORS/node23.html>

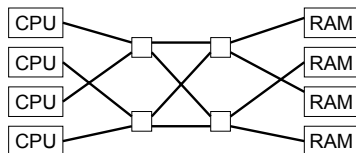
Switched multiprocesori



- crossbar switch

Switched multiprocesori

- prednost
 - dobrom organizacijom SW, svi CPU istovremeno mogu pristupati RAMu
 - => teoretski neograničen broj paralelnih CPU
- nedostatak
 - za N CPU potreban je N^2 preklopnika
- rješenje: omega switch



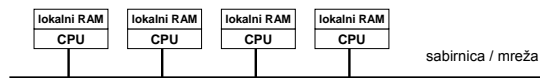
Switched multiprocesori

- omega switch
 - $\log_2 N$ preklopnih stupnjeva s po $N/2$ preklopnika
 - problem kašnjenja (zbog propagacije kroz stupnjeve preklapanja)
- rješenje: hijerarhijski sustav
 - svaki CPU ima svoj RAM s vrlo brzim pristupom
 - i pristup zajedničkom RAMu, nešto sporije
 - NUMA (NonUniform Memory Access)

Switched multiprocesori - primjer

- SGI Altix
- Proizveden 2003.
- SGI NUMAflex architektura povezivanja procesora
- 512 procesora Intel Itanium2
- 1TB memorije
- OS=Linux
- <http://www.sgi.com/servers/altix/>

Bus multikomputeri



- svaki CPU ima svoj privatni RAM
- komunikacija samo između CPUova
 - bitno manji zahtjevi nego kod CPU-RAM komunikacije
- vrlo jednostavno za realizirati
- najčešće računala na LANu

Primjeri - Clusteri

- *Fail-over Clusters*
 - sastoje od 2 ili više računala koji su povezani zasebnom "heartbeat" mrežom.
 - Čim jedna usluga na nekom od računala prestane raditi da netko drugi preuzima.
- *Load-balancing Clusters*
 - Kad dođe zahtjev odabire se ono računalo koje je najmanje zauzeto.
 - Često je i *fail-over* tipa.

Clusteri

- *High Performance Computing (HPC) Cluster*
 - Nudi vrhunske performanse izvođenja pojedinih aplikacija.
 - Rade i load-balancing. Procese razbacaju po različitim računalima.
 - Pokušavaju proces paralelizirati kako bi se rutine unutar njega mogle paralelno izvoditi.

Hrvatski clusteri

- Cluster se obično prevodi sa grozd
- 2 projekta SRCE i IRB
 - <http://isabella.srce.hr> i <http://grozd.irb.hr>
- Nakupina PC-a spojena na ethernet-LAN
- Jedno računalo služi kao pristupno (front-end).
 - Korisnik se ulogirava i izvodi aplikacije.
 - Raspodjeljuje posao po ostalim računalima
 - Korisnički direktorij je share-an ostalim rač.

Hrvatski clusteri - Hardware

- SRCE
 - 8 dvoprocesorskih PIII na 1.13GHz
 - 512MB RAM
 - pristupno računalo sa PIII na 1Ghz i 1GB RAM.
 - RedHat Linux 7.2
- IRB
 - 18 dvoprocesorskih PC-a (taktovi: 1GHz, 1.8Ghz, 2.0Ghz),
 - 1GB ili 2GB memorije svaki.
 - RedHat Linux 7.1

Clusteri – korištenje (tipično)

- Cluster ima batch sustav
 - Batch sustav se brine o opterećenosti pojedinih računala, dodijeli resursa, pokretanju/zaustavljanju pojedinih jobova
 - Primjeri: PBS i MOSIX
 - Sve aplikacije koje korisnik hoće izvršavati mora prvo kod takvog sustava registrirati.

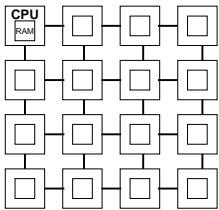
Clusteri – korištenje (tipično)

- Biblioteke za razvoj paralelnih programa
 - Nema zajedničke memorije pa se podatci razmjenjuju pomoću sustava poruka.
- MPI (**message passing interface**)
 - Omogućava 'point-to-point' i kolektivnu međuprocesnu komunikaciju grupe jednodretvenih procesa koji se izvršavaju paralelno.
 - MPI_Send i MPI_Recv() služe za slanje zadataka drugim procesima i primanje rezultata.
 - Primjer koda je na
 - <http://www.iam-mpi.org/tutorials/one-step/ezstart.php>.

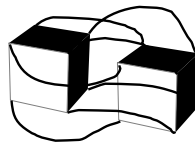
Clusteri – korištenje (tipično)

- Postoje već paralelizirane biblioteke koje sadrže standardne ili napredne matematičke funkcije.
- Primjeri su:
 - mpiblacs - MPI Basic Linear Algebra Communication Subprograms
 - scalapack -Scalable Linear Algebra PACKage

Switched multikomputeri



Grid



Hypercube

- Grid
 - najprikladnija za dvodimenzionalne probleme
- Hypercube
 - n-dimenzionalna kocka

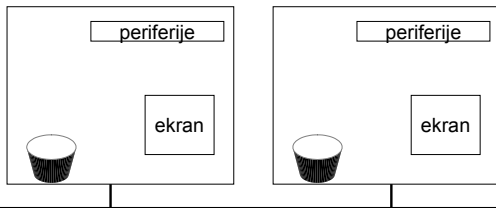
Switched multikomputeri - primjer

- Intel Paragon
- Iz 1992.
- 1024 nodeova povezanih kao hiperkocka
- Svaki node je imao 3x80860 Intel 32bitni RISC procesor sa 128MB RAM
 - Dva procesora služe za računanje a jedan za komunikaciju
- Operacijski sutav je bio OSF1
- <http://ed-thelen.org/comp-hist/intel-paragon.html>

Softver za distribuirane sustave

- loosely coupled
 - WS na LANu
- tightly coupled
- network OS
- “pravi” distribuirani sustavi
- multiprocesorski time-sharing sustav

Network operating system

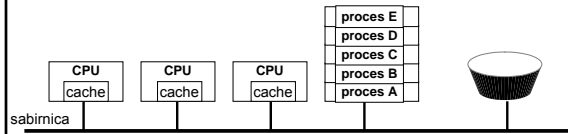


- omogućava
 - slanje poruka
 - pristup podacima
 - izvršenje naredbi

“Pravi” distribuirani sustav

- izvodi se na nakupini nezavisnih računala bez zajedničke memorije
- korisnicima se čini da se radi o jednom računalu
- za sada ni jedan sustav u potpunosti ne zadovoljava definiciju

Multiprocesor time-sharing sustav



- svaki CPU izvodi jedan proces

Elementi dizajna SW

- transparentnost
- fleksibilnost
- pouzdanost
- performansa
- skalabilnost

Transparentnost

- korisnik ne može razlikovati jedno i višeprocesorski rad
- location transparency
 - korisnik ne zna gdje se proces izvodi
- migration transparency
 - resursi se mogu samostalno seliti
- replication transparency
 - OS samostalno kopira podatke među lokacijama
- concureny transparency
 - korisnik ne prepoznaje postojanje drugih korisnika
- parallelism transparency
 - "sveti cilj"

Fleksibilnost

- učiniti jezgru OS što jednostavnijom
- podijeliti zadatke korisničkim serverima, dislociranima

Pouzdanost

- po definicije je veća
- ali treba paziti da sustav ne ovisi o nekom servisu koji se izvodi samo an jednom CPU
- osim pouzdanosti i dostupnost (availability)
- neosjetljivost na greške (fault tolerance)

Performanse

- fine-grained parallelism
 - mnogo sitnih poslova
 - neprikladno za distribuiranost
- coarse-grained parallelism
 - malo velikih poslova
 - prikladno za distribuiranost

Skalabilnost

- potrebno je izbjegavati
 - centralizirane komponente
 - centralizirane podatke
 - centralizirane algoritme

Programska podrška mjernih i procesnih sustava

www.zesoi.fer.hr
(<http://www.ZESOI.FER.hr/hrzesoi/dodip/ppmips.htm>)

ppmps@zesoi.fer.hr
