

Προηγμένες αρχιτεκτονικές υπολογιστών και προγραμματισμός παραλλήλων συστημάτων

LAB-05A. MPI (Μέρος Α') –
Εισαγωγή – Περιγραφή μοντέλου

Υπόβαθρο

- Παράλληλη Επεξεργασία.
- Επικοινωνία με άλλες διεργασίες.
- Συνεργατικές λειτουργίες.
- Μονόπλευρες λειτουργίες.
- Η διεργασία MPI.

Παράλληλη Επεξεργασία

- Πολλαπλοί εργάτες (workers) ή διεργασίες (processes).
- Με ή χωρίς συντονιστή (master).
- Αλληλεπίδραση με ανταλλαγή πληροφορίας υπό μορφή μηνυμάτων.

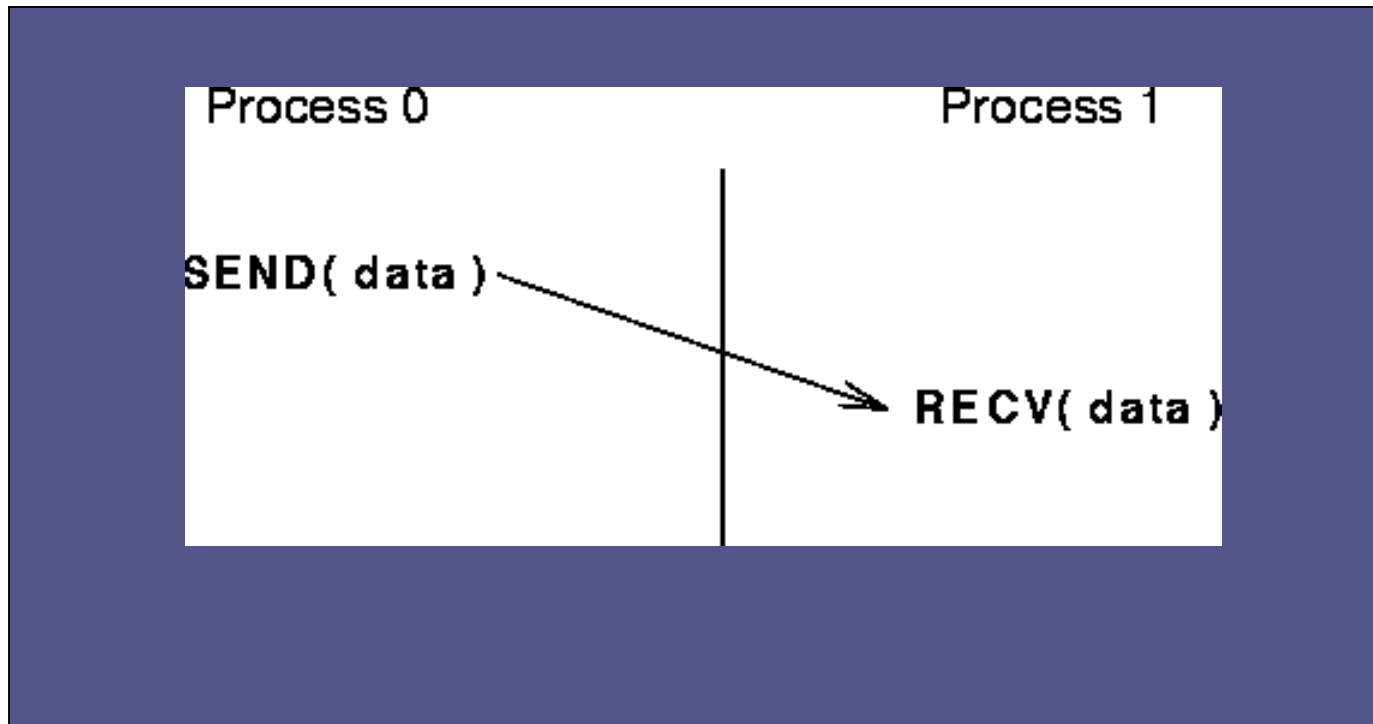
Είδη Παράλληλης Επεξεργασίας

- **Data-parallel** : Ίδιες λειτουργίες σε διαφορετικά δεδομένα (SIMD).
- **SPMD**: Ίδιο πρόγραμμα, διαφορετικά δεδομένα.
- **MIMD**: Διαφορετικό πρόγραμμα, διαφορετικά δεδομένα.
- Τα SPMD και MIMD είναι ουσιαστικά το ίδιο αφού κάθε MIMD μπορεί να γίνει SPMD.
- Το SIMD είναι ειδική περίπτωση, αλλά λιγότερο πρακτικό.
- Το MPI είναι κυρίως για SPMD/MIMD.

Επικοινωνία Διεργασιών

- Τα δεδομένα πρέπει να ανταλλάσσονται με άλλους εργάτες (επεξεργαστές/διεργασίες/κόμβους).
- Συνεργατική επικοινωνία: τα δύο (ή περισσότερα) μέρη συμφωνούν στην ανταλλαγή δεδομένων.
- Μονόπλευρη επικοινωνία: ένας εργάτης εκτελεί τη μετακίνηση των δεδομένων.

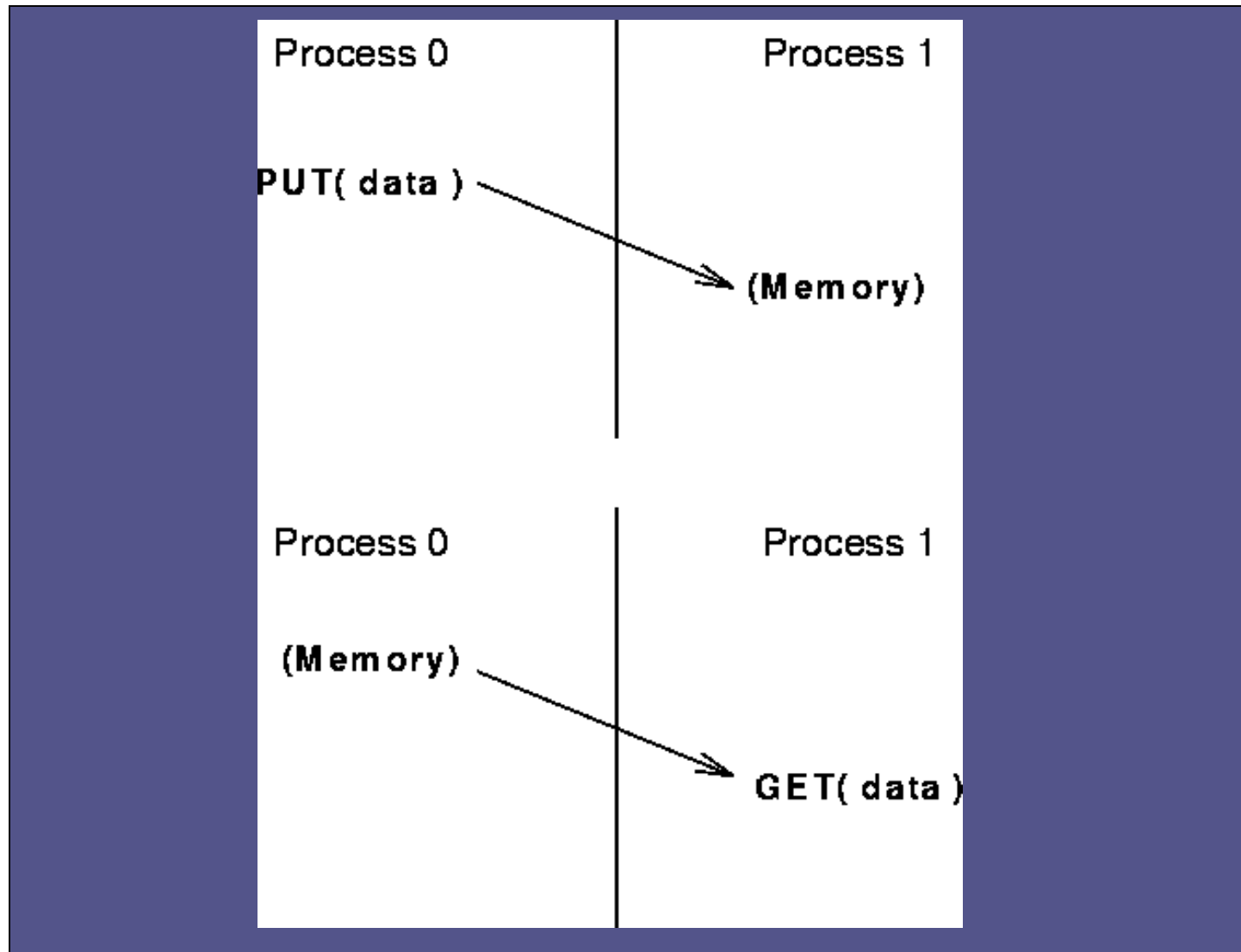
Συνεργατική Επικοινωνία



Συνεργατική Επικοινωνία (2)

- Η ανταλλαγή μηνυμάτων (message-passing) κάνει την ανταλλαγή δεδομένων συνεργατική.
- Τα δεδομένα πρέπει και να σταλούν και να ληφθούν ρητώς.
- Πλεονέκτημα: κάθε αλλαγή στη μνήμη του δέκτη (receiver) γίνεται με την συμμετοχή του δέκτη.

Μονόπλευρη Επικοινωνία



Μονόπλευρη Επικοινωνία (2)

- Η μονόπλευρη επικοινωνία περιλαμβάνει τις λειτουργίες remote memory read και remote memory write.
- Πλεονέκτημα: τα δεδομένα μπορούν να διακινηθούν χωρίς αναμονή για την απάντηση της άλλης διεργασίας.

Μοντέλα Hardware

- Μοντέλα hardware για την ανταλλαγή δεδομένων μεταξύ εργατών.
- Κατανεμημένη μνήμη (Distributed memory) π.χ., Paragon, IBM SPx, workstation network.
- Κοινή μνήμη (Shared memory) π.χ., SGI Power Challenge, Cray T3D.
- Και τα δύο είδη μπορούν να χρησιμοποιηθούν με τα μοντέλα SIMD ή MIMD.

MPI: Message Passing Interface

- Ένα πρότυπο ανταλλαγής μηνυμάτων για παράλληλους υπολογιστές, clusters, και ετερογενή δίκτυα.
- Σχεδιάστηκε για:
 - να βοηθήσει στην ανάπτυξη βιβλιοθηκών παράλληλου λογισμικού.
 - να προσφέρει πρόσβαση σε προηγμένο παράλληλο hardware σε τελικούς χρήστες, συγγραφείς βιβλιοθηκών και σχεδιαστές εργαλείων.
- Διάφορες υλοποιήσεις: MPICH/2/3, Open MPI, Microsoft MPI, κλπ.
- Εμπορικές υλοποιήσεις από HP, Intel, NEC, κλπ.

Χαρακτηριστικά MPI

- Δεν είναι compiler. Απαιτεί την ύπαρξη κάποιου άλλου compiler (π.χ. GCC για να λειτουργήσει).
- Αρχική υποστήριξη: C, C++, Fortran.
- Βιβλιοθήκες που επεκτείνουν άλλες γλώσσες ώστε να υποστηρίζουν το MPI: Java, MATLAB, Python, R, Julia, κλπ.

Ιστορία MPI

- Πρωτοπαρουσιάστηκε στο Williamsburg Workshop, Απρίλιος 1992.
- Οργανώθηκε στο Supercomputing '92 (Νοέμβριος 1992).
- Pre-final draft στο Supercomputing '93.
- Final draft Μάιος 1994.
- Υπάρχουν διαθέσιμες δημόσιες υλοποιήσεις στο Web:
- MPICH: <https://www.mpich.org/>
- Δεκέμβριος 2020: MPICH 3.3.2.

Ιστορία MPI (2)

- Δημιουργήθηκε με ευρεία συμμετοχή:
 - Εταιριών: IBM, Intel, TMC, Meiko, Cray, Convex, Ncube.
 - Συγγραφών βιβλιοθηκών: PVM, p4, Zipcode, TCGMSG, Chameleon, Express, Linda
 - Ειδικών εφαρμογών και συμβούλων.

Ιδέες και νοήματα (1)

- **Communicators:** Συνδέουν ομάδες διεργασιών σε μία συνεδρία (session) MPI.
 - Προσδίδουν ένα ανεξάρτητο αναγνωριστικό σε κάθε διεργασία της ομάδας.
 - Οργανώνουν τις διεργασίες της ομάδας σε μία διατεταγμένη τοπολογία.
- **Επικοινωνία Point-to-Point:** Επικοινωνία διεργασίας/CPU με άλλη διεργασία/CPU.
 - Τρόποι επικοινωνίας: normal (blocking και non-blocking), synchronous, ready, buffered.
 - Buffers και derived datatypes.

Ιδέες και νοήματα (2)

- **Συλλεκτική επικοινωνία:** Επικοινωνία ομάδων διεργασιών/CPU με άλλες (συγκεκριμένες ή μη) ομάδες διεργασιών/CPU.
 - Built-in και user-defined συλλεκτικές λειτουργίες.
 - Πλήθος συναρτήσεων μετακίνησης δεδομένων.
 - Ορισμός υποομάδων.
- Παράδειγμα: Message broadcasting: Μετάδοση ενός μηνύματος από μία διεργασία προς όλες τις υπόλοιπες διεργασίες.
- **Εφαρμογο-στρεφείς τοπολογίες:**
Ενσωματωμένη υποστήριξη για γράφους και grids.

Ιδέες και νοήματα (3)

- **Profiling:** Επιτρέπεται στο χρήστη να συλλάβει τα MPI calls και να τα αντικαταστήσει με δικά του.
- **Περιβάλλον προγραμματισμού:**
 - Inquiry.
 - Έλεγχος σφαλμάτων (debugging).

Τι δεν είναι το MPI

- Υλοποιούνται ιδέες και έννοιες που είναι σχετικές μόνο με message-passing.
- Συνεπώς, ιδέες που δεν σχετίζονται με ανταλλαγή μηνυμάτων δεν υλοποιούνται.
- Ιδέες που δεν υλοποιούνται από το MPI:
 - Process management.
 - Remote memory transfers.
 - Active messages.
 - Threads.
 - Virtual shared memory.

Μέγεθος

- Το MPI είναι μεγάλο (125 συναρτήσεις).
 - Η εκτεταμένη λειτουργικότητα του MPI απαιτεί παροχή πολλών συναρτήσεων.
 - Το πλήθος αυτό δεν συνεπάγεται πολυπλοκότητα ή βραδύτητα.
- Το MPI είναι μικρό (6 συναρτήσεις).
 - Πολλά παράλληλα προγράμματα μπορούν να γραφτούν κάνοντας χρήση 6 μόνο βασικών συναρτήσεων.
- Παρέχεται δυνατότητα επιλογών και ευελιξία αν αυτή είναι απαραίτητη.
- Δεν απαραίτητη η γνώση όλου του πακέτου του MPI προκειμένου να χρησιμοποιηθεί.

Use cases

- Φορητά (portable) παράλληλα προγράμματα.
- Παράλληλες βιβλιοθήκες.
- Αλγόριθμοι που περιλαμβάνουν ακανόνιστες ή δυναμικές σχέσεις μεταξύ των δεδομένων που δεν ταιριάζουν σε κάποιο μοντέλο παράλληλης επεξεργασίας.
- Μεγάλος όγκος δεδομένων.
- Μεγάλη ανάγκη υπολογιστικής ισχύος.
- Το MPI συναντάται σήμερα στην πλειοψηφία των Top-500 υπερ-υπολογιστών.