System design-related Optimization problems

Michela Milano
Joint work
DEIS Università di Bologna
Dip. Ingegneria Università di Ferrara
STI Università di Urbino

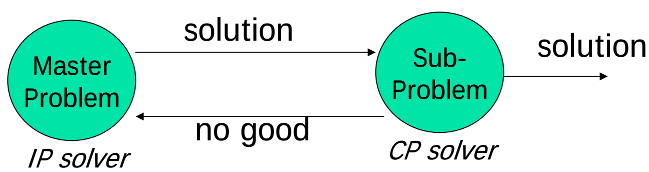
Allocation and scheduling for MPSoCs DATE'06

- Objective function: communication cost
- Developed a complete approach based on Logic Based Benders Decomposition
- Exploit the best solver for each sub-problem
  - IP for allocation
  - CP for scheduling
  - Nogood for communication between solvers
- Three order of magnitude speedup w.r.t. the single solvers
- Focussed on pipelined applications
- Results validated on the MPARM platform
  - Simplifying assumptions on the bus do not impact the expected throughput if the bus utilization is maintained under the 60% of the total available bandwidth
  - Average error 4.8% standard deviation 0.08.
  - We used random task graphs, a GSM and a MIMO processing

Problem decomposition

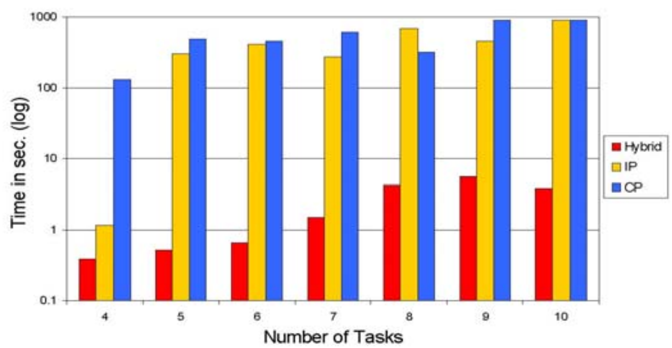
Objective function: Min(Communication Cost)

- Assignment of tasks and memory slots *(master problem)*
  - ✓ Obj. Func. Relates alternative resources to couples of tasks
  - ✓ Not a good scenario for Constraint Programming
- Task scheduling with static resource assignment *(subproblem)*
  - ✓ Integer Programming does not handle time efficiently
  - ✓ Constraint Programming is instead effective



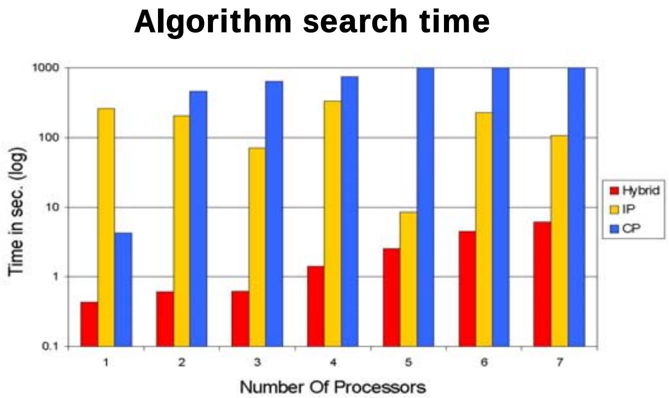
Results

Algorithm search time



The combined approach dominates, and its higher complexity comes out only for simple system configurations

Results



The combined approach dominates, and its higher complexity comes out only for simple system configurations

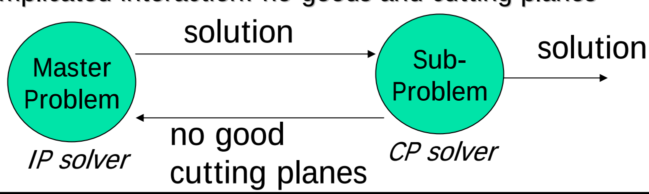
Dynamic Voltage and Freq. Scaling

- Objective function: computation and switching energy
- We optimize allocation, scheduling, and voltage scaling but not separately
- Again a complete approach based on Logic Based Benders Decomposition
  - IP for allocation and frequency selection
  - CP for scheduling
  - Nogood and cutting planes for communication between solvers
- Exploit the best solver for each sub-problem
- Pipeline and non-pipeline applications
- No way to solve the problem with a single solver
- Results validated on the MPARM platform
  - Average error throughput 4.51% standard deviation 1.94.
  - Average error energy 4.80% standard deviation 1.71.
  - We used random task graphs, and GSM

Problem decomposition

Objective function:
Min(CompEnergy+FreqSwitchEnergy)

- Assignment of frequencies tasks and memory slots minimizing the Computation energy *(master problem)*
  - ✓ Not a good scenario for Constraint Programming
- Task scheduling with static resource/frequency assignment minimizing the *(subproblem)* switching overhead
  - ✓ Constraint Programming is instead effective
- Much more complicated interaction: no-goods and cutting planes



Allocation and scheduling of CTG

- On going research
- Up to now only the optimization part has been completed, the validation still missing
- Objective function: communication cost
- Technique: Logic based Benders Decomposition. We transform a stochastic problem in an approximation based on the CTG analisys. The approximation turns out to be exact.
- Pipelined and non-pipelined applications
- Performances comparable with the deterministic case

Some extremely hard instances: possibly solved with randomization in complete search

Objective function

$$\sum_{i=0}^{n-1} \boxed{f_i(X(\omega))} \left[\overline{m}_i (1 - \sum_{j=0}^{p-1} \boxed{M_{ij}}) + \overline{s}_i (1 - \sum_{j=0}^{p-1} \boxed{S_{ij}}) + \sum_{r \in \mathcal{E}_{i \to k}} \boxed{f_k(X(\omega))} \overline{e}_r (1 - \sum_{j=0}^{p-1} \boxed{E_{rj}}) \right]$$

- Depends on decision variables and on stochastic variables
- When the allocation is fixed only on stochastic vars.

$$\sum_{\omega \in \Omega} p(\omega) \left[\boxed{C_1} f_i(X(\omega)) + f_i(X(\omega)) \sum_{r \in \mathcal{E}_{i \to k}} f_k(X(\omega)) \boxed{C_{2_k}} \right]$$

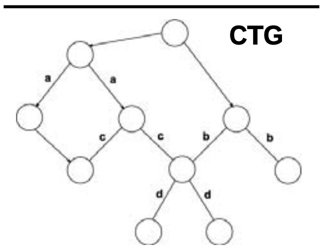
Substituting: $\sum_{\omega \in \Omega} p(\omega) f_i(X(\omega)) C_1 \blacktriangleright C_1 \sum_{\omega \in \Omega_i^1} p(\omega)$

$$\sum_{\omega \in \Omega} p(\omega) f_i(X(\omega)) \sum_{r \in \mathcal{E}_{i \to k}} f_k(X(\omega)) C_{2_k} \blacktriangleright \sum_{r \in \mathcal{E}_{i \to k}} C_{2_k} \left(\sum_{\omega \in \Omega_i^1 \cap \Omega_k^1} p(\omega) \right)$$

The expected value reduces to a deterministic function

CTG analysis

- **We need to know the probability of existence and co-exostence of nodes**
  - We have developed polynomial algorithms



▶ Data structures:
Activation set (AS)
Sequence matrix

Coexistence set — Probability of AS existence

Coexistence probability

Complexity $O(c^3)$

Allocation and scheduling of Multiple Task Graphs

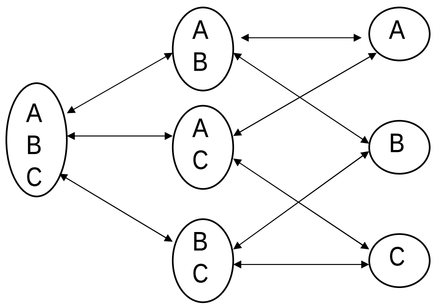
- On going research
- Up to now we are developing the optimization part, the validation still missing
- Objective function: communication cost + migration cost
- Technique: Logic based Benders Decomposition.
- We start from a situation where a TG1 is running and the second TG2 starts. We minimize the communication cost overall plus the migration cost of TG2.
- Many pareto optimal solutions, choose at runtime
- Pipelined applications
- Problem: transition graph with multiple nodes for each configuration

Allocation and scheduling of Multiple Task Graphs

- First solution
  - TG1 is running and TG2 starts its execution. We optimally allocate the second task by possibly migrating some tasks in TG1.
  - Various combination of communication cost and migration cost. Try to find pareto optimal points
  - Choose at run-time
- Same technique when a task graph stops its execution.

Allocation and scheduling of Multiple Task Graphs

- Second solution
  - Compute different minimum communication cost transition graphs with a bounded migration cost
  - Example: task graphs A, B and C



Each arc is labelled with the minimum delta communication cost. Each node is an allocation

Other on-going research

- Traffic Optimization + scheduling on NoC. For the moment we are facing an heuristic approach.
- Optimizing the communication
- Allocation and scheduling with stochastic duration