Revisiting the bicriteria (length,reliability) multiprocessor static scheduling problem

Alain Girault and Hamoudi Kalla

Workshop on the Foundations of Component-Based Design

September 30th, 2007 — ARTIST II

Problem and motivations

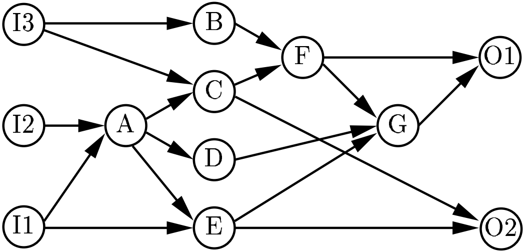
Problem

Schedule an application task graph onto a heterogeneous distributed memory architecture, with a guaranteed reliability and WCET

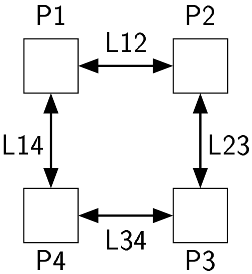
- Two criteria : maximize the reliability and minimize the WCET
- Belongs to the class of bicriteria optimization problems
- Reliability is crucial to assess the system's dependability
- Length is crucial to assess the system's real-time property
- Industrial applications : automotive (AUTOSAR), consumer electronics, ...

Algorithm and architecture model

Algorithm task graph



Distributed architecture graph



Reliability model

Definition of reliability

It measures the service continuity ➪ Probability that the system functions correctly during a given time interval.

Reliability model of [Lloyd & Lipow, 1962] [Shatz & Wang, IEEE TR'89]

$$R(X/P) = e^{-\lambda_P \, d(X/P)}$$

- λ_P is the failure rate of component P per time unit
- $d(X/P)$ is the WCET of operation X onto P
- All the HW components are fail-silent
- All the failures are transient (implies the "hot" failure model)
- All the failure occurrences are statistically independent events

State of the art in bicriteria scheduling

- [Qin, Jiang & Swanson, ICPP'02] : reliable point-to-point comm. links, re-execution of failed operations with overlap, each primary task is scheduled onto the processor minimizing the reliability cost
- [Dogan & Özgüner, IEEE TPDS'02] : no task replication, smart choice of assignments of the tasks to the processors, aggregation of the two criteria
- [Dogan & Özgüner, TCJ'05] : same as above with a tuning of the aggregation coefficients to tradeoff execution time for reliability
- [Assayad, Girault & Kalla, DSN'04] : active replication of operations, aggregation of the two criteria
- [Pop, Poulsen & Izosimov, CODES-ISSS'07] : reliable comm. bus, re-execution of failed operations
- Plus plenty of articles that assume the network is acyclic to make the terminal-pair problem tractable

Intuitions

Intuition 1 : antagonistic criteria

More replication is good for the reliability but bad for the schedule length (and vice-versa)

Intuition 2 : tasks' replication level vs. reliability

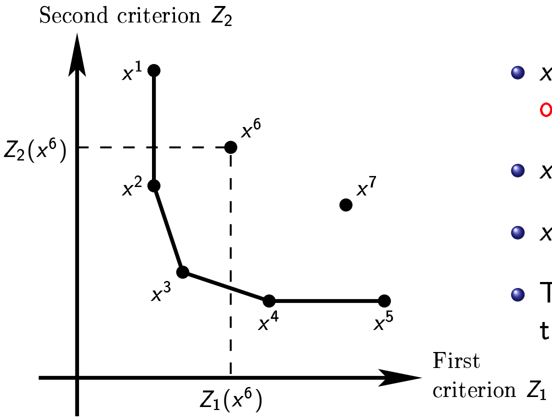
The level of replication is related to the reliability criteria

Intuition 3 : replication factor vs. processor reliability

Operations scheduled onto more reliable processors are replicated less (and vice-versa)

Shortcomings I : issues related to Pareto optima

The two criteria are antagonistic !
➪ Pareto optima and non-dominanted solutions [T'kindt & Billaut, 2006]



- x^1, x^2, x^3, x^4, and x^5 are Pareto optima
- x^1 and x^5 are weak optima
- x^2, x^3, and x^4 are strong optima
- The set of all Pareto optima is the Pareto curve

Usual approaches to bicriteria optimization

[T'kindt & Billaut, 2006]

1. Aggregation of the two criteria into a single one ➪ transform the problem into a classical single criterion optimization problem.
2. Transformation of one criterion into a constraint ➪ find the optimum among all the solutions that satisfy the constraint.
3. Hierarchization of the criteria ➪ optimize one criteria at a time.
4. Interaction with the user ➪ the user guides the search for a Pareto optimum.

Shortcomings II : issues related to reliability

Reliability model : $$R(X/P) = e^{-\lambda_P \, d(X/P)}$$

The reliability is a function of the length

➪ Three problems :

1. The length criteria overpowers the reliability criteria
2. It is impossible to control the replication factor of the operations onto the processors (potential funnel effect)
3. The reliability is not a monotonous function of the scheduling

Proposal

First contribution

Define a new criteria independent of the length : the GSFR

GSFR = Global System Failure Rate

Second contribution

Design a new bicriteria (length,GSFR) scheduling algorithm

Find $$\min_{S \in \mathcal{S}} \left(\mathtt{C}_{\mathtt{max}}(S), GSRF(S) \right)$$

Definition of the Global System Failure Rate (GSFR)

Reliability model : $$R(X/P) = e^{-\lambda_P \, d(X/P)}$$

The GSFR is the failure rate per time unit of the global system S, seen as if it were a single HW component :

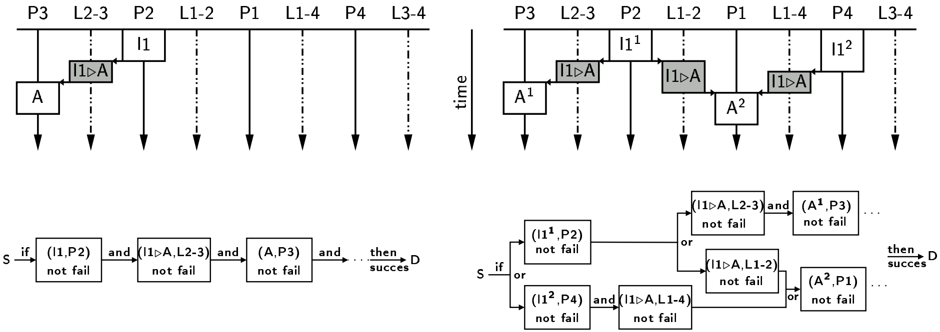
$$GSFR(S) = \Lambda(S) = \frac{-\log R(S)}{U(S)}$$

With : $$U(S) = \sum_{o_i \in S} \mathcal{E}xe(o_i)$$ (consistent with the “hot” model)

And of course the usual reliability formula holds :

$$R(S) = e^{-\Lambda(S) U(S)}$$

Computing the reliability : Reliability Block-Diagrams (RBD)



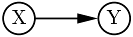
In general, the reliability computation exponential in the RBD size

(aka terminal-pair problem, NP-complete [Ball, IEEE TR'86])

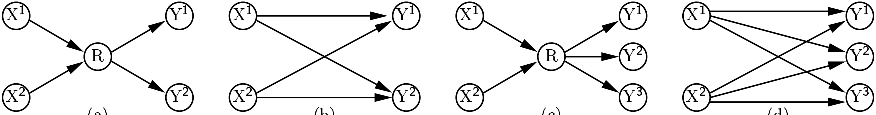
➪ Compute the reliability with the minimal cut sets method

Making the RBD serial-parallel

Simple algorithm graph :



We insert routing operations in the algorithm task graph :

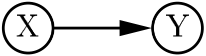


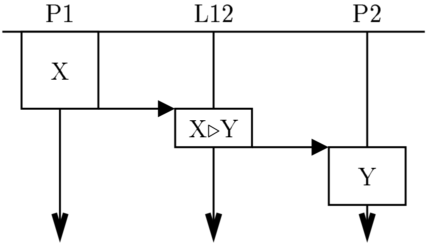
(a) (b) (c) (d)

They incur an additional overhead on the schedule length, because there is less concurrency between the communications.

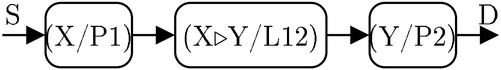
However, since there are also less communications, this additional overhead is reasonable.

RBD of a schedule without replication





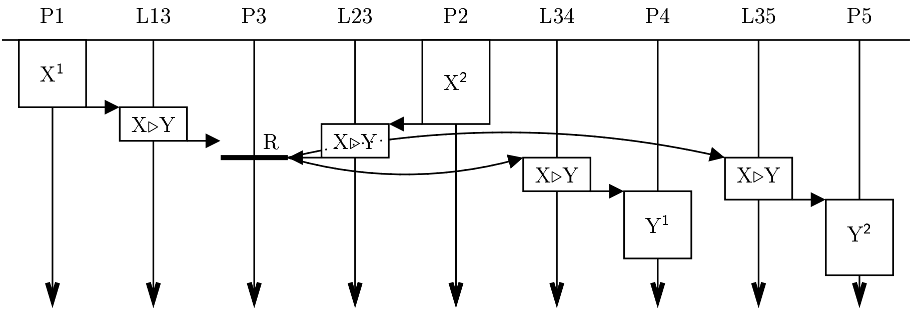
The RBD is :

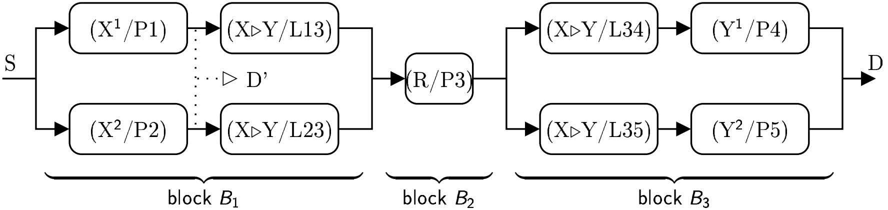


$$\begin{aligned} R &= R(X, P1) R(X \triangleright Y, L12) R(Y, P2) = e^{-\lambda_1 t_X^1} e^{-\lambda_{12} t_{XY}^{12}} e^{-\lambda_2 t_Y^2} \\ &= e^{-(\lambda_1 t_X^1 + \lambda_{12} t_{XY}^{12} + \lambda_2 t_Y^2)} \end{aligned}$$

$$\Lambda = \frac{-\log R}{U} = \frac{\lambda_1 t_X^1 + \lambda_{12} t_{XY}^{12} + \lambda_2 t_Y^2}{t_X^1 + t_{XY}^{12} + t_Y^2}$$

RBD of a schedule with replication (I)





RBD of a schedule with replication (II)

$$
\begin{aligned}
R(S) &= R(B_1) \cdot R(B_2) \cdot R(B_3) \\
R(B_1) &= 1 - \left(1 - e^{-(\lambda_1 t_X^1 + \lambda_{13} t_{XY}^{13})}\right) \left(1 - e^{-(\lambda_2 t_X^2 + \lambda_{23} t_{XY}^{23})}\right) \\
R(B_2) &= 1 \text{ because the WCET of R is always 0} \\
R(B_3) &= 1 - \left(1 - e^{-(\lambda_{34} t_{XY}^{34} + \lambda_4 t_X^4)}\right) \left(1 - e^{-(\lambda_{35} t_{XY}^{35} + \lambda_5 t_X^5)}\right)
\end{aligned}
$$

For each processor Pi, we take $\lambda_i = 10^{-5}$ and $t_X^i = t_Y^i = 5$.
For each link Lij, we take $\lambda_{ij} = 10^{-4}$ and $t_{XY}^{ij} = 3$.

$$
\left.\begin{array}{r} R(B_1) = 0.99999988 \\ R(B_2) = 1 \\ R(B_3) = 0.99999988 \end{array}\right\} \Longrightarrow R(S) = 0.99999976
$$

$$
\Longrightarrow \Lambda(S) = \frac{-\log R(S)}{U(S)} = 7.500\,10^{-9}
$$

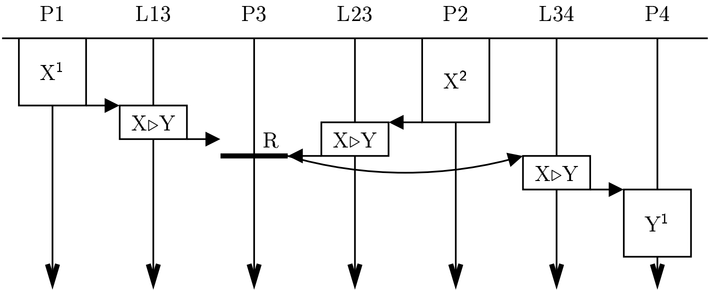
Computing compositionally the GSFR

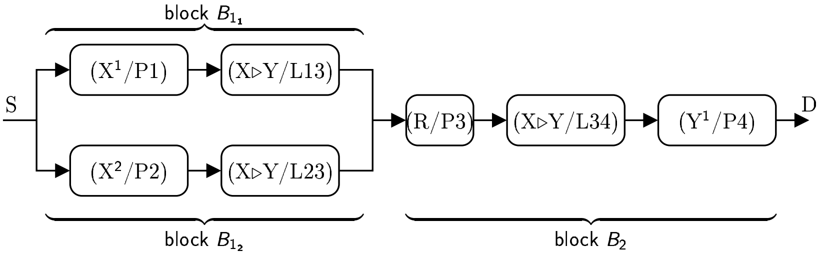
Suppose we have two blocks B_1 and B_2, with respective failure rates λ_1 and λ_2, and respective WCET t_1 and t_2

Serial schedule : $$\Lambda(B_1 \cdot B_2) = \frac{\lambda_1 t_1 + \lambda_2 t_2}{t_1 + t_2}$$

Parallel schedule : $$\Lambda(B_1 \| B_2) \simeq \frac{\lambda_1 t_1 \lambda_2 t_2}{t_1 + t_2}$$

How redundancy improves the GSFR (I)





How redundancy improves the GSFR (II)

$$\Lambda_{1_1} = \frac{\lambda_1 t_X^1 + \lambda_{13} t_{XY}^{13}}{t_X^1 + t_{XY}^{13}} = 4.375\,10^{-5} \qquad T_{1_1} = t_X^1 + t_{XY}^{13} = 8$$

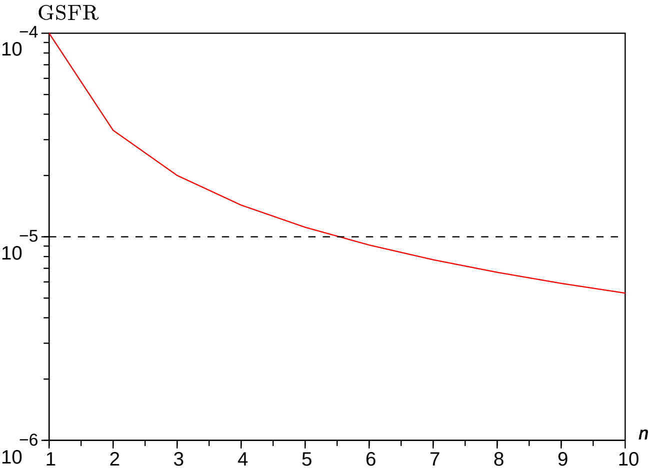
$$\Lambda_{1_2} = \frac{\lambda_2 t_X^2 + \lambda_{23} t_{XY}^{23}}{t_X^2 + t_{XY}^{23}} = 4.375\,10^{-5} \qquad T_{1_2} = t_X^2 + t_{XY}^{23} = 8$$

$$\Lambda_1 \simeq \frac{\Lambda_{1_1} T_{1_1} \Lambda_{1_2} T_{1_2}}{T_{1_1} + T_{1_2}} \simeq 7.656\,10^{-9} \qquad T_1 = T_{1_1} + T_{1_2} = 16$$

$$\Lambda_2 = \frac{0 + \lambda_{34} t_{XY}^{34} + \lambda_4 t_Y^4}{0 + t_{XY}^{34} + t_Y^4} = 4.375\,10^{-5} \qquad T_2 = 0 + t_{XY}^{34} + t_Y^4 = 8$$

$$\Lambda = \frac{\Lambda_1 T_1 + \Lambda_2 T_2}{T_1 + T_2} = 2.917\,10^{-5} \qquad T = T_1 + T_2 = 24$$

How redundancy improves the GSFR (III)



➪ If one operation is not replicated, then we replicate twice six other operations to regain one order of magnitude of the GSFR !

Bicriteria Scheduling Heuristics (BSH)

Theorem

In a serial-parallel RBD, if each macro-block in the sequence is such that its GSFR is less than Λ_{obj}, then the GSFR of the whole RBD is also less than Λ_{obj}.

Outline of BSH, our Bicriteria Scheduling Heuristic :

- It is a list scheduling heuristic
- Candidate operations are sorted by a smart cost function
- The dependable schedule pressure selects the most urgent candidate operation
- This most urgent operation is scheduled on a subset of processors such that the GSFR of the block is less than Λ_{obj} and such that the increase in schedule length is minimal

Simulations : architecture's reliability

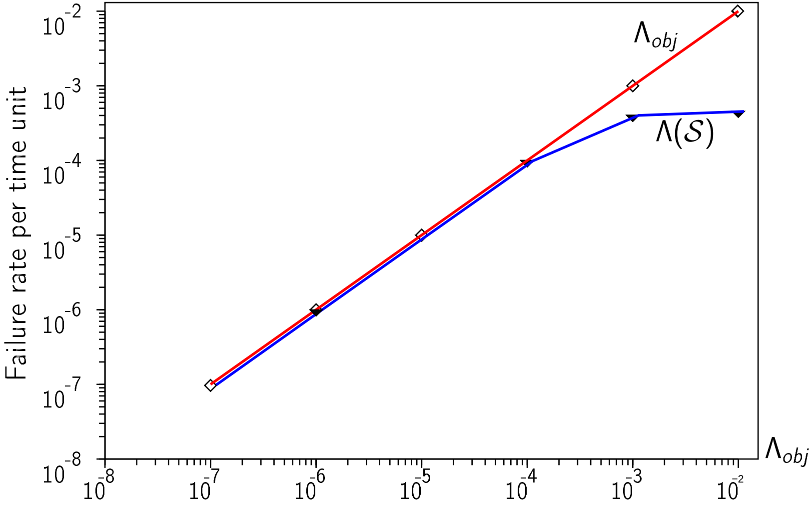
4 processors fully connected architecture :

P1,P2	P5,P6	L12,L15,L16,L25,L26,L56
$\lambda_{1,2} = 10^{-4}$	$\lambda_{5,6} = 10^{-5}$	$\lambda_m = 10^{-3}$

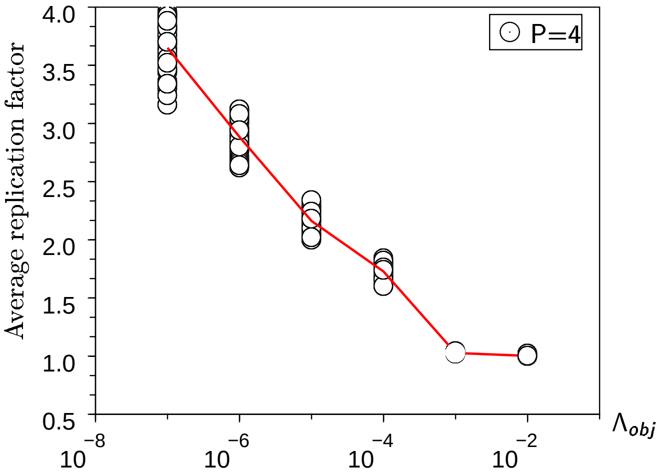
6 processors fully connected architecture :

P3,P4	L13,L14,L23,L24,L34,L35,L36,L45,L46
$\lambda_{3,4} = 5.10^{-5}$	$\lambda_m = 10^{-3}$

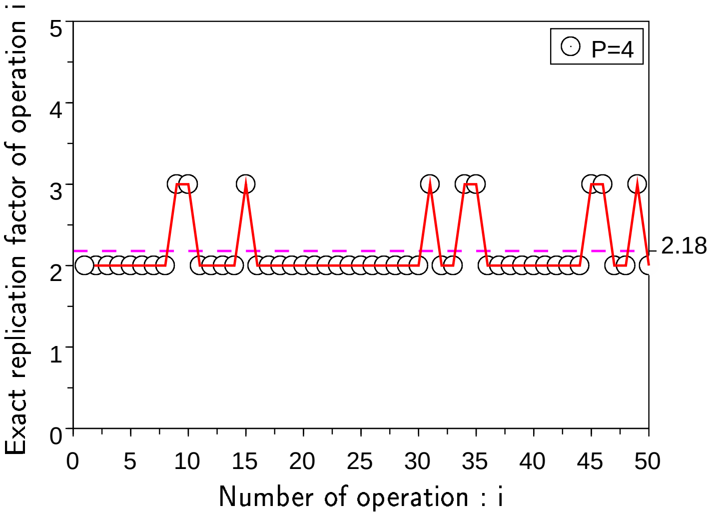
Variation of the obtained GSFR Λ in function of Λ_{obj}



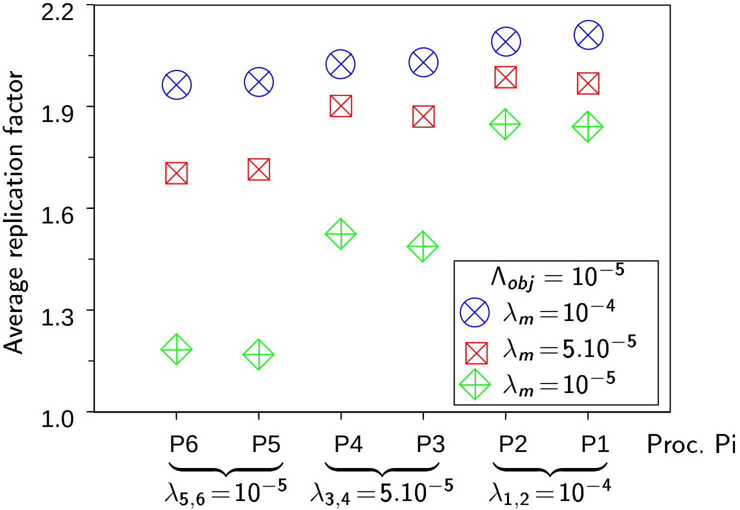
Variation of the average replication factor in function of Λ_{obj} on a 4 processors architecture



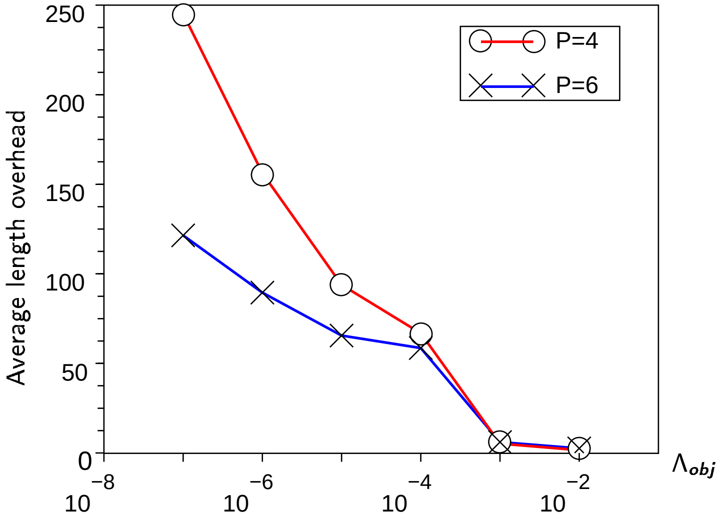
Variation of the exact replication factor in function of Λ_{obj}



Variation of the average replication factor in function of the processors' failure rate



Variation of the schedule length overhead in function of Λ_{obj}



Average schedule length overhead due to the routing operations in function of λ_m

Average schedule length overhead due to the routing operations :

λ_m	10^{-3}	10^{-4}	10^{-5}
P= 4	$-4.12\,\%$	$+2.43\,\%$	$+4.09\,\%$
P= 6	$+2.44\,\%$	$+8.47\,\%$	$+9.96\,\%$

Average replication factor for the schedules with routing operations :

λ_m	10^{-3}	10^{-4}	10^{-5}
P= 4	2.07	1.50	1.33
P= 6	2.10	1.52	1.35

Conclusions

The new bicriteria (length,GSFR) scheduling algorithm works remarkably well.

The simulation results match the three intuitions.

Adding the routing operations to compute the reliability incurs less than 4% overhead on average.

An important lesson learnt

Any bicriteria optimization problem in which the two criteria are not "independent" one from the other will always suffer for the three problems identified.