

ECOLE NATIONALE SUPERIEURE  
DES MINES DE SAINT ETIENNE

N° d'ordre: 158 ID

**THESE**  
présentée

pour obtenir le grade de **DOCTEUR EN SCIENCE** de

**L'UNIVERSITE JEAN MONNET et de L'ECOLE NATIONALE  
SUPERIEURE DES MINES DE SAINT ETIENNE**

Spécialité: **Informatique**

PAR

**Mustapha Boukhalfa HADIM**

**Sur une Méthode de Routage des Messages dans les  
Architectures Parallèles à Mémoire Distribuée :  
Application à la Grille Torique**

Soutenue le 30 Juin 1997 devant la Commission d'examen composée de

Mr. Bernard PEROCHE	Président
Mr. Joffroy BEAUQUIER	Rapporteur
Mr. Yousef SAAD	Rapporteur (excusé)
Mme. Marie-Claude HEYDEMANN	Examineur
Mr. Jean-Claude KÖNIG	Examineur
Mr. Jean AZEMA	Examineur
Mr. Jean-Jacques GIRARDOT	Examineur
Mr. Ibrahima SAKHO	Directeur

## Résumé

Dans les architectures parallèles à mémoire distribuée, la communication entre processus est un des facteurs de performance les plus importants pour les applications. Le système qui en a la charge, i.e, *le noyau de communication*, doit intégrer une fonctionnalité essentielle pour de telles architectures : *le routage des messages*. Cette fonctionnalité est assurée par une composante spécifique du noyau de communication : *le noyau de routage*, dont le rôle est l'acheminement d'un message d'un nœud émetteur vers un nœud récepteur.

L'acheminement des messages nécessite une *stratégie de routage* qui spécifie les chemins de communication pour toute paire de processeurs (source, destination) du réseau d'interconnexion. Une telle stratégie de routage doit satisfaire d'une part, des *critères de correction* et d'autres part, des *critères d'efficacité*.

Le but de cette thèse est la conception de stratégies de routage pour les réseaux de processeurs qui satisfont à la fois, les critères de correction et les critères d'efficacité. Nous proposons une méthode de conception de stratégies de routage, permettant par une démarche incrémentale, de satisfaire les deux types de critère : *la communication multi-niveaux* et le *schéma de communication primaire* associé.

Pour mesurer l'efficacité de la méthode, nous l'appliquons à un réseau particulier : *la grille torique*. Les différents algorithmes de routage obtenus sont corrects et très efficaces.

Nous proposons également une technique d'implantation de notre méthode de routage, permettant le calcul des tables de routage directement sur le réseau de processeurs. Cette technique permet ainsi l'obtention d'un système *auto-constructif*.

**Mots clés :** Architectures parallèles, Communication, Routage des messages, Communication multi-niveaux, Schéma de communication primaire, Grille torique.

## Abstract

In distributed memory parallel architectures, interprocess communication is one of the main efficiency factor for the applications. The system which is in charge of this task, i.e, *the communication kernel*, must integrate an essential functionality for such architectures : *message routing*. This functionality is ensured by a specific component of the communication kernel : the *routing kernel*, whose purpose is to convey a message from a sender node to a receiver node.

Messages conveying requires a *routing strategy* which specify the communication paths for each pair of (source,destination) processors of the interconnection network. Such routing strategy must satisfy in one hand, correction criteria and in other hand, *efficiency criteria*.

The purpose of this thesis is the design of routing strategies for networks of processors, which satisfy at the same time, correction criteria and efficiency criteria. We propose a method to design routing strategies, which allows by an incremental approach, to satisfy both of the criteria : the *multi-level communication* and the associated *primary communication scheme*.

To prove the efficiency of the method, we apply it to a particular topology : The *torus* network. The several obtained routing algorithms are correct and very efficient.

We propose also a technique to implement our routing method, which allows to compute the routing tables directly on the target network processors. Thus, this technique allows to obtain a *self constructing* system.

**Keywords :** Parallel architectures, Communication, Message routing, Multi-level communication, Primary communication scheme, Torus network.

# Table des matières

<b>Résumé</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Remerciements</b>	<b>iii</b>
<b>Table des matières</b>	<b>v</b>
<b>1 Introduction générale</b>	<b>1</b>
1.1 Du séquentiel au parallèle . . . . .	1
1.2 Les architectures parallèles . . . . .	4
1.3 Contrôle des architectures parallèles . . . . .	6
1.3.1 Le partitionnement . . . . .	6
1.3.2 L'allocation et l'ordonnancement . . . . .	7
1.3.3 Le contrôle des communications et synchronisations . . . . .	7
1.4 Problématique et apport de la thèse . . . . .	9
1.5 Organisation du mémoire . . . . .	11
<b>I Acheminement des Messages par Routage</b>	<b>12</b>
<b>2 Etat de l'art sur le routage des messages</b>	<b>13</b>
2.1 Spécification du problème de routage . . . . .	13
2.2 Une classification des solutions . . . . .	15
2.3 Les types de routage . . . . .	15
2.4 Les mécanismes de routage . . . . .	17
2.4.1 Les techniques de commutation de données . . . . .	18
2.4.2 Le contrôle de flot . . . . .	21
2.5 Les stratégies de routage . . . . .	22
2.5.1 Le routage avec des tables de routage compactes . . . . .	26
2.5.2 Le routage hiérarchique . . . . .	27

2.5.3	Le routage par préfixes . . . . .	28
2.5.4	Le routage par intervalles . . . . .	28
2.6	Les principaux problèmes de routage . . . . .	30
2.6.1	L'interblocage . . . . .	31
2.6.1.1	Interblocage et graphe de dépendance . . . . .	32
2.6.1.2	Les méthodes de détection-guérison . . . . .	35
2.6.1.3	Les méthodes préventives . . . . .	36
2.6.2	Les pannes physiques . . . . .	39
2.6.3	La congestion, la famine et le refus des messages . . . . .	40
2.7	Conclusion . . . . .	45
<b>3</b>	<b>De l'antagonisme des critères de routage</b>	<b>47</b>
3.1	Corrélation entre les critères de routage . . . . .	48
3.2	Conception de stratégies de routage . . . . .	51
3.3	Proposition d'une méthode . . . . .	53
3.4	Interblocage et élongation des chemins . . . . .	55
<b>II</b>	<b>Une Méthode de Routage</b>	<b>56</b>
<b>4</b>	<b>Modèle et définitions</b>	<b>57</b>
<b>5</b>	<b>La communication multi-niveaux</b>	<b>62</b>
5.1	Organisation du chapitre . . . . .	62
5.2	Le principe de la communication multi-niveaux . . . . .	62
5.3	Propriété caractéristique de la communication multi-niveaux . . . . .	64
5.4	Modélisation de la communication multi-niveaux . . . . .	70
5.5	Application 1 : la communication multi-niveaux selon les règles de routage par cycle eulérien . . . . .	72
5.5.1	La méthode de routage par cycle eulérien . . . . .	73
5.5.2	Application à la grille torique . . . . .	74
5.5.2.1	Un algorithme de routage dans les grilles toriques . . . . .	77
5.5.2.2	Un second algorithme de routage dans les grilles toriques . . . . .	86
5.6	Application 2: La communication multi-niveaux selon le routage <i>e-cube</i> . . . . .	87
5.7	Interprétation des résultats . . . . .	87
5.8	Conclusion . . . . .	89

<b>6</b>	<b>Implantation de la communication multi-niveaux</b>	<b>90</b>
6.1	Une implantation distribuée . . . . .	90
6.2	Un algorithme distribué de calcul d'un cycle eulérien dans un réseau	92
6.2.1	Le principe de l'algorithme . . . . .	92
6.2.2	Correction de l'algorithme . . . . .	96
6.2.3	Le code de l'algorithme . . . . .	99
6.2.4	Les performances de l'algorithme . . . . .	103
6.3	Le calcul des tables de coût et de routage . . . . .	103
<b>7</b>	<b>Un schéma de communication primaire</b>	<b>112</b>
7.1	Introduction . . . . .	112
7.2	Routage sans interblocage dans un anneau . . . . .	113
7.3	La méthode de dérivation de graphe de dépendance acircuitique .	115
7.3.1	Définitions et Notations . . . . .	115
7.3.2	Le principe de la méthode . . . . .	116
7.3.3	Une heuristique pour le calcul du graphe de dépendance .	120
7.3.4	Extension de l'heuristique . . . . .	124
7.3.5	Le calcul de la fonction de routage . . . . .	125
7.4	Correction de la méthode . . . . .	128
7.5	Application de la méthode aux tores de dimension 2 . . . . .	129
7.6	Conclusion . . . . .	136
	<b>Conclusion générale et perspectives</b>	<b>138</b>
<b>III</b>	<b>Annexe</b>	<b>142</b>
<b>A</b>	<b>The Multi-Level Communication: Minimal, Deadlock Free, Storage Optimal Routing for Torus Networks [50]</b>	<b>143</b>
<b>B</b>	<b>Minimal, Deadlock Free and <math>O(n)</math> Space Memory Routing for <math>k</math>-ary <math>n</math>-cubes with Wraparound Connections [52]</b>	<b>170</b>
	<b>Bibliographie</b>	<b>181</b>
	<b>Table des figures</b>	<b>189</b>