Scheduling Parallel Applications on Heterogeneous Distributed Systems

Guoqi Xie • Gang Zeng • Renfa Li • Keqin Li

Scheduling Parallel Applications on Heterogeneous Distributed Systems

Springer

Guoqi Xie
College of Computer Science
and Electronic Engineering
Hunan University
Changsha, Hunan, China

Gang Zeng
Graduate School of Engineering
Nagoya University
Nagoya, Aichi, Japan

Renfa Li
Key Laboratory for Embedded and
Cyber-Physical Systems of Hunan Province
Hunan University
Changsha, Hunan, China

Keqin Li
Department of Computer Science
State University of New York
New Paltz, NY, USA

ISBN 978-981-13-6556-0 ISBN 978-981-13-6557-7 (eBook)
https://doi.org/10.1007/978-981-13-6557-7

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Today, heterogeneous distributed systems exhibit diversification and variation. These systems can be large enough, such as cloud computing systems, or small enough, such as embedded computing systems. Cyber-physical systems (CPS) are emerged heterogeneous distributed systems, which have created one of the hottest research fields in recent years. Automotive CPS (ACPS) are further development of automotive embedded systems. Similarly, cyber-physical cloud systems (CPCS) are further development of cloud computing systems. At present, the relevant results of CPS have gradually been reflected in practical applications. The scheduling problems of parallel applications have been studied for a long time, but for the abovementioned heterogeneous distributed systems, there is an urgent need for customized scheduling algorithms to deal with their individual characteristics and challenges. The publication of this book satisfies this need in a timely manner.

This book first introduces different forms of today's heterogeneous distributed systems. Then, it presents energy-efficient real-time scheduling algorithms for real-time parallel applications on heterogeneous distributed embedded systems, reliability-aware fault-tolerant scheduling algorithms for parallel applications on heterogeneous distributed cloud systems, and high-performance real-time scheduling algorithms for multiple parallel applications on heterogeneous distributed embedded systems. The book further discusses scheduling algorithms and applications on ACPS and CPCS, respectively, which are the two types of CPS.

This book introduces different types of heterogeneous distributed systems and proposes a variety of advanced scheduling algorithms for parallel applications on these systems. This book is rich in content and detailed in graphics. For each proposed algorithm, the book uses corresponding motivational examples to explain clearly and achieve the easy-to-understand purpose; such an approach is a major and unique feature of this book. The book contains the basic knowledge and the latest research development of scheduling theory and methods on heterogeneous distributed systems and is indeed a significant contribution to the field of heterogeneous distributed systems.

This book is a joint effort and creation of four scholars from three countries with dedication and distinction. The authors have published very extensively in the last few years in the fields of embedded computing, cloud computing, and cyber-physical systems and are undoubtedly the leading scholars in scheduling parallel applications on heterogeneous distributed systems. Finally, I would like to congratulate the authors on a job well done, and I look forward to see the book in print.

Albert Y. Zomaya

Chair Professor of High Performance Computing
and Networking
School of Computer Science
University of Sydney
Sydney, NSW, Australia

Preface

Motivation of the Book

Scheduling parallel applications on heterogeneous distributed systems is a classic research area in computer science and engineering. In recent years, with the emergence and development of embedded computing, cloud computing, and cyber-physical systems (CPS), this research area has shown renewed vitality, challenges, and breakthroughs. Today, heterogeneous distributed embedded systems (e.g., automotive embedded systems) and heterogeneous distributed cloud systems (e.g., cloud-based services, such as Amazon EC2) are typical scenarios of heterogeneous distributed systems. As advanced heterogeneous distributed systems, CPS further enhance the existing embedded and cloud systems. Specifically, automotive CPS (ACPS) and cyber-physical cloud systems (CPCS) are two types of CPS applied to the areas of embedded computing and cloud computing, respectively.

These new distributed systems not only have their own characteristics but also raise new and special requirements for scheduling parallel applications. Parallel application scheduling in the context of energy-efficient scheduling, reliability-aware scheduling and high-performance real-time scheduling, is worthy of investigation on heterogeneous distributed embedded systems, heterogeneous distributed cloud systems, ACPS, and CPCS. The motivation of this book is to provide a comprehensive coverage of our research results in the aforementioned topics in recent years.

Summary of Contents

Chapter 1 introduces heterogeneous distributed systems, including heterogeneous distributed embedded systems and heterogeneous distributed cloud systems. At the same time, ACPS and CPCS are taken as examples to introduce the advanced heterogeneous distributed systems of CPS. Then, this chapter explains the characteristics of parallel applications and introduces the automotive and workflow applications as examples. For scheduling parallel applications on heterogeneous distributed systems, this chapter analyzes the necessity and challenges of energy-efficient scheduling, reliability-aware scheduling, high performance real-time scheduling, scheduling on ACPS, and scheduling on CPCS. Finally, this chapter lists the outline of this book.

Chapter 2 discusses energy-efficient real-time scheduling for parallel applications. First, this chapter presents an effective energy consumption minimization method for real-time parallel applications on heterogeneous distributed embedded systems implemented by combining the presented non-DVFS and global DVFS-enabled energy-efficient scheduling algorithms. Second, this chapter presents two energy-aware processor merging algorithms for real-time parallel applications in heterogeneous distribution cloud systems by turning off partial processors from an energy-saving perspective.

Chapter 3 studies reliability-aware fault-tolerant scheduling for parallel applications. First, this chapter presents enough and heuristic replication algorithms, respectively, to minimize the redundancy for a parallel application on heterogeneous distributed cloud systems. Second, this chapter further presents quantitative fault-tolerant scheduling algorithms with minimum execution costs and shortest schedule lengths, respectively, for a parallel application on heterogeneous distributed cloud systems. Third, this chapter proposes reliability-aware fault-tolerant scheduling toward resource cost optimization for heterogeneous embedded systems, such as ACPS, during the design phase.

Chapter 4 focuses on high-performance real-time scheduling for parallel applications. First, this chapter presents mixed real-time scheduling of multiple parallel applications on heterogeneous distributed embedded systems. Then, this chapter further presents high-performance mixed-criticality real-time scheduling of multiple parallel applications on heterogeneous distributed embedded systems.

Chapter 5 considers the application on ACPS. This chapter presents fairness-based and adaptive dynamic scheduling algorithms, respectively, on mixed-criticality ACPS, which aims to minimize individual schedule lengths of applications with short overall schedule length of ACPS from a high-performance perspective. The adaptive dynamic scheduling algorithm can respond autonomously to the joint challenges of heterogeneity, dynamics, and parallelism of ACPS.

Chapter 6 covers the application on CPCS. This chapter presents an effective energy management algorithm for multiple real-time parallel workflow applications on CPCS. First, the algorithm can achieve lower deadline missed ratios (DMRs) than the state-of-the-art algorithm. Second, this algorithm can save as much energy as possible with lower DMRs.

Chapter 7 summarizes the book and mentions future research.

Audience and Readership

This book should be a useful reference for the researchers, engineers, and practitioners interested in scheduling theory for embedded computing systems, cloud computing systems, and CPS. The book can be used as a supplement for an advanced undergraduate or a graduate course in distributed computing, embedded computing, cloud computing, and cyber-physical systems in computer science, computing engineering, and electrical engineering. By reading this book, the graduate students and doctoral students will be familiar with the new types of heterogeneous distributed systems and their features, learn a variety of scheduling algorithms, and find a source of inspiration for their own research.

Acknowledgments

The authors are grateful to Professor Albert Zomaya of the University of Sydney in Australia for writing the foreword of this book. We also thank Ms. Jasmine Dou and Mr. Ganesan Senthil Kumar of Springer for their efforts and support in helping to publish this book.

Changsha, Hunan, China Guoqi Xie
Nagoya, Aichi, Japan Gang Zeng
Changsha, Hunan, China Renfa Li
New Paltz, NY, USA Keqin Li

Contents