Yanlong Zhai · Muhammad Mudassar ·
Liehuang Zhu

Edge Computing Resilience

Overcoming Resource
Constraints in Unstable
Computing Environments

Springer

SpringerBriefs in Computer Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

**Indexing: This series is indexed in Scopus, Ei-Compendex, and zbMATH **

Yanlong Zhai • Muhammad Mudassar •
Liehuang Zhu

Edge Computing Resilience

Overcoming Resource Constraints
in Unstable Computing Environments

🐴 Springer

Yanlong Zhai 🆔
Cyberspace Science and Technology
Beijing Institute of Technology
Beijing, China

Muhammad Mudassar 🆔
Computer Science Department
COMSATS University Islamabad
Vehari, Punjab, Pakistan

Liehuang Zhu 🆔
Cyberspace Science and Technology
Beijing Institute of Technology
Beijing, China

If disposing of this product, please recycle the paper.

To the visionaries and innovators in the domain of edge computing, whose tireless quest for excellence and a determined focus on forging ahead in technology have motivated us to venture into new horizons of computing resilience.

This book pays tribute to those who demonstrate anticipatory wisdom to identify potential amidst difficulties, the engineers who persistently refine resource utilization, and the researchers who boldly address the ever-shifting landscape, all in gratitude for your unswerving commitment and passion.

We also extend our heartfelt appreciation to our families, whose steady solidarity and deep concern have sustained us on this voyage. With your love and support, we've found the strength to fulfill ambitious objectives.

May this book light the way for those enthusiastic about enhancing the resilience, efficiency, and adaptability of computing environments, serving as a source of knowledge and inspiration.

Foreword

In an era of growing global interconnectivity, the essence of technological innovation is changing at its periphery. Edge computing, a forward-thinking notion that spurring readily available infrastructure, is altering our perspective of data processing, scalability, and fault tolerance, and it's increasing the potential for innovation.

"Edge Computing Resilience: Overcoming Resource Constraints in Unstable Computing Environments" demonstrates the unwavering commitment to pursue knowledge and distinguish the relevant research and industry sectors in this dynamically changing sector. Dr. Yanlong Zhai and a dedicated team of experts have scripted this book, thoroughly examining edge computing that spans from its foundational principles to progressive approaches. It presents valuable insights for those tracing the complex roadmap of edge systems. In this book, readers will discover a rich source of information that probes into the complicated world of edge computing from multiple angles.

Beginning with an introduction to edge computing systems, the monograph describes their real-world usability and the challenges they confront. The authors then focus on creating resilient edge systems, highlighting scalability and fault tolerance as essential aspects. By exploring resource-constrained offloading, energy-aware offloading, and optimization methodologies in subsequent chapters, readers will acquire essential knowledge to deal with these specific edge computing challenges. Looking ahead, the book also scrutinizes forthcoming ventures, staging a glance at the emerging cooperative edge computing models, wireless environments, and atypical scenarios.

In this masterwork, Dr. Yanlong Zhai's leadership and the authors' accumulated expertise glow brightly. This book reflects their commitment to establishing new paths in the domain of edge computing. It functions as a navigational tool for researchers, practitioners, and students embarking on their journey into this field. Whether you're an experienced professional looking to enhance your knowledge or a newcomer determined to explore this dynamic field, you will find invaluable enlightenment throughout the content.

Congratulations to Dr. Yanlong Zhai and the team for this distinguished contribution to the realm of edge computing. May "Edge Computing Resilience: Overcoming Resource Constraints in Unstable Computing Environments" encourage and brighten the way ahead for all who read it.

University of Wollongong, Wollongong, NSW,   Professor Jun Shen
Australia
Beijing, China
June 2024

Preface

The recent development of Edge Computing resulted in a new age of opportunities and problems in the ever-evolving field of computing. This book explores the complex domain of edge computing systems and offers a thorough examination of its core ideas, uses, and approaches for boosting effectiveness. The journey starts with the foundational work by outlining the fundamental ideas behind edge computing. While highlighting the significance of resilience in edge computing settings, we investigate its broad range of applications and the distinctive problems it brings.

The essential domain of scalability and fault tolerance considered critical to real-time edge applications is explored. We outline the function of fault tolerance and analyze the different sources and fault kinds. We focus on scaling in distributed edge systems and bottleneck identification, which prompts us to investigate scaling solutions and state backup techniques. The relevance of offloading in edge computing is explained along with the relationship between local and offloading devices. With an emphasis on inductive learning and privacy constraints, we investigate privacy-preserving offloading techniques since privacy is an important concern.

Furthermore, we focus on energy optimization and efficiency. We examine the energy used by offloading devices and compare the execution of local and offloading tasks. New offloading optimization techniques have been presented to support both single and multiple battery-operated devices. A thorough investigation of optimization in edge computing systems is presented. We explore several optimization strategies, such as task offloading, edge caching, network optimization, resource allocation, and privacy-aware optimization. In addition, edge node clustering and heterogeneous device settings are examined. We firmly think that readers' research into the field of edge computing will provide them with new perspectives and sources of inspiration.

Overview of the Book

This book is aimed to focus on three major aspects of edge computing resilience. Specifically, this includes fault detection and tolerance in edge computing, offloading, and optimization techniques in edge computing. As depicted in the book title, the book provides a comprehensive analysis of resilience in edge computing applications based on the aforementioned key areas. The discussion includes a thorough analysis of edge computing challenges, problem formulation, design methodologies, and implementation details of some application scenarios. It also provides recommendations for overcoming resource constraints in unstable computing environments based on the existing system designs and the future scope. It's significant to point out that optimization techniques, offloading methods, and fault-handling principles provided in this book are discussed within the scope and context of the edge computing paradigm.

In total, this book contains seven chapters, with multiple sections and subsections in each. Chapter 1 introduces the background of edge computing systems, their applications, and the challenges which prompted the need for resilience in edge computing applications. Several other challenges including security, system and statistical heterogeneity, resource management, and disaster management are discussed briefly, while the detailed discussion on fault tolerance, offloading, and optimization continues in the later section. Particularly, fault tolerance is covered in Chap. 2, whereas Chaps. 3 and 4 discuss about resource-constrained offloading and privacy preserving offloading respectively. SDN-based and energy-aware offloading is presented in Chap. 5. The edge computing optimization is covered in Chap. 6 while Chap. 7 focuses on the future scope.

In Chap. 2, we introduce the concept of scalability and fault tolerance for real-time edge applications. Here, we discuss various aspects of fault tolerances including their significance, causes, and different fault types. We presented design methodology and provided implementation of scale-out operation for fault tolerance. In detail, we give an insight into scaling for distributed edge system and provide a detailed design and operation of scale-out methods and state backup methods for fault tolerance respectively. Under this, we presented the principle of dynamic scaling where the system scales out resources in real-time according to the fluctuations of incoming streams with minimal overhead and latency. Furthermore, we also provide a concept and framework for check-pointing-based fault tolerance with scale-out operations.

In Chap. 3, we explore resource-constrained offloading, and then discuss offloading strategies and possible scenarios. A computational model for local execution and task offloading is introduced. The chapter examines existing research on task dependency, offloading to fog devices, dynamic offloading frameworks, and centralized versus distributed task offloading. Finally, the challenges linked to resource-constrained offloading are highlighted. Besides introducing resource-constrained offloading in edge computing, this chapter serves as an important prerequisite for the next chapter.

In Chap. 4, we uncover the need for privacy-preserving offloading in resource-constrained environments. A privacy-preserving offloading scheme for mobile edge intelligent systems is also presented. In this context, we discuss the privacy issues associated with task offloading and then present local differential privacy and inductive learning as a solution to address privacy concerns in edge computing. Further discussion includes energy-aware offloading which is provided in the next chapter.

In Chap. 5, our discussion is mainly centered on energy-aware offloading based on SDN. Three key areas are covered in this chapter, namely, the energy consumed by offloading devices, offloading task execution including the local and offloading execution tasks, and finally the offloading optimization which presents a detailed implementation of offloading optimization both for single and multiple devices on battery.

In Chap. 6, we introduce the concept and goal optimization for edge computing applications and then provide an in-depth discussion on various optimization techniques. These include task offloading, network optimization techniques, resource allocation schemes, the concept of edge caching as well, and privacy-aware optimization. We also presented the idea of optimization for heterogeneous devices followed by edge node clustering mechanisms for distributed computing introduced earlier. This chapter ends with a discussion on performance analyses and evaluation metrics such as profiling tools, benchmarking, tracing, simulations, containerization, and edge analytics.

The last chapter of the book, Chap. 7, focuses on future work in edge computing systems. In particular, we introduce the various challenges in the new cooperative edge computing. In addition, we provide a detailed discussion on optimization in wireless edge computing environment and provide potential recommendation and potential research directions for future work. Other aspects of edge computing resilience including edge computing in denied, disconnected, intermittent, or limited (D-DIL) environments are also discussed.

Beijing, China   Yanlong Zhai
Vehari, Pakistan   Mudassar Muhammad
Beijing, China   Liehuang Zhu
June 2024

Acknowledgments

Contents