Preface

INTRODUCTION

Various attempts are conducted nowadays in order to bridge the gap between two worlds: real-time and distributed computing, where processing, storage, and networking need to be combined and delivered with guaranteed levels of service. The publication aims at becoming a reference point for the research community by providing a detailed analysis on the outcomes of the work performed on real-time technologies and concepts for Service Oriented Infrastructures (SOIs) - including more than 400 references to innovative works - since real-time is considered to be a multidisciplinary approach. Furthermore, each section that describes a specific area (e.g. workflow management) also presents comparisons of the research outcomes and Future Trends in order to highlight the potentials for future work in the field.

REAL-TIME & DISTRIBUTED COMPUTING

Traditionally, “real-time” refers to hard real-time systems, where even a single violation of the desired timing behavior is not acceptable, for example because it leads to total failure, possibly causing loss of human lives. However, there is also a wide range of applications that also have stringent timing and performance needs, but for which some deviations in Quality of Service (QoS) are acceptable, provided these are well understood and carefully managed. These are soft real-time applications and include a broad class of interactive and collaborative tools and environments, including concurrent design and visualization in the engineering sector, media production in the creative industries, and multi-user virtual environments in education and gaming. Soft real-time applications are traditionally developed without any real-time methodology or run-time support from the infrastructure on which they run. The result is that either expensive and dedicated hardware has to be purchased to ensure good interactivity levels and performance, or that general-purpose resources are used as a compromise (e.g. commodity operating systems and Internet networking) with no way to guarantee or control the behavior of the application as a result.

Real-time attributes in SOIs require time-constrained operation of services, both to satisfy deadlines of executing tasks that have been set by users and to sustain appropriate throughput levels. In particular, this time critical operation involves synchronous communication between services, something which has a direct impact in the design, implementation, and integration of services oriented infrastructures.
The most challenging aspect is the integration of real-time attributes into all levels of service oriented systems. Major focus is given to the need for methodologies, tools, and architectures for complex distributed systems that address the practical issues of performance guarantees, timed execution, real-time management of resources, synchronized communication under various load conditions, satisfaction of QoS constraints, as well as dealing with the trade-offs between these aspects.

The main objective of this publication is to provide an integrated view of the outcomes in the multidisciplinary area of real-time in distributed computing. In the core of the chapters, the readers can find answers to questions such as: what is the state of the art in the field with regard to real-time, how are the presented approaches compared, and what are the potentials in each area. Therefore, the chapters also include reflections on what are the merits and limits of the relevant approaches. The publication describes and compares more than 400 research efforts.

ORGANIZATION OF THE BOOK

Service Oriented Architectures refer to a specific architectural paradigm that emphasizes implementation of components as modular services that can be discovered and used by clients. Through the agility, scalability, elasticity, rapid self-service provisioning, and virtualization of hardware, Service Oriented Architecture principles are reflected into Grids and Clouds, which provide the ability to efficiently adapt resource provisioning to the dynamic demands of Internet users. Many architectural paradigms from distributed computing such as service-oriented infrastructures, Grids, and virtualization are incorporated into Clouds. According to these, there are three main classes in the cloud services stack which are generally agreed upon:

- Software as a Service (SaaS), which refers to the provision of an application as a service over the Internet or distributed environment.
- Platform as a Service (PaaS), which refers to the provision of a development platform and environment providing services and storage, hosted in the distributed system.
- Infrastructure as a Service (IaaS), which refers to the provision of “raw” machines (computing, storage, networking and other devices) on which the service consumers deploy their own software.

Based on the above, in this book the research efforts in different areas of real-time distributed systems are presented, classified under the aforementioned cloud services stack (i.e. SaaS, PaaS, IaaS). The book includes fifteen (15) chapters contributed by forty-three (43) scholars. In the first section (named “Software as a Service”), Chapter 1 introduces specific techniques used to model and analyze the required QoS level for real-time applications as well as mechanisms to find business models where value can be provided at all stages in the value-chain. For example, how can an application user identify a set of service providers that in combination provide the best solution to a workflow, how much risk is involved, et cetera. Authors of Chapter 2 provide an overview of model-driven engineering approaches, standards, resources, and tools that support different aspects of model-driven software engineering: language development, modeling services, and real-time applications. Chapter 3 discusses on the research efforts that aim to APIs for real-time computing, including real-time application interfaces, as well as available cloud APIs, along with the architecture and the technologies that they support. The last chapter of the
first section, Chapter 4 presents the current state-of-the-art technology and methodologies regarding the evaluation of the provided QoS in service-oriented environments.

Section 2 (named “Platform as a Service”), starts with Chapter 5, in which authors extract a taxonomy of characteristics found in modern service discovery systems and produce a categorization of existing implementations in a grouped and comparative way, based on these features. Furthermore, the authors introduce a mapping of these characteristics to the cloud business model, in order to assist in selecting the suitable solutions for each provider based on his location in the value chain or identify gaps in the existing implementations. Chapter 6 introduces the reasons and difficulties for monitoring and metering on Cloud infrastructures. The approaches for monitoring of the execution environment and the network on virtualized infrastructures are described along with the existing monitoring tools present on different commercial and research platforms. Still in the PaaS section of the book, Chapter 7 gives an overview of various research efforts regarding workflow semantics and languages as well as their enactment within the scope of distributed systems. Scheduling algorithms and advance reservation techniques are also discussed as these are one of the hottest research topics in workflow management systems. Authors of Chapter 8 analyze different fields where Service Level Agreements are used, examine the proposed solutions, and investigate how these can be improved in order to better support the creation of real-time service-oriented architectures. Addressing security concerns is the topic of Chapter 9, in which authors examine cloud security, privacy, and trust issues from three levels: business (to identify issues arising from the motivations and concerns of business stakeholders), jurisdiction (to identify risks that arise from legislation, gaps in legislation, or conflicts between legislation in different jurisdictions related to a cloud deployment), and technical (to identify issues that arise from technical causes such as ICT vulnerabilities, and/or require technical solutions, such as data confidentiality and integrity protection). The last chapter of the second section of the book, Chapter 10, introduces different Web service specifications, analyses their different communication mechanisms, and compares the existing implementations. Moreover, various approaches to implement Web services are described by the authors followed by the middleware that makes use of specific web service specifications, including the description of the commercial interfaces and development tools to create services for distributed systems. Interoperability problems are also mentioned in Chapter 10.

The last section of the book, named “Infrastructure as a Service,” discusses how different research efforts address issues related to virtualization of computing, storage, and network resources on the infrastructure level. In this context, Chapter 11 provides recent results in the research literature on virtualized large-scale systems and soft real-time systems. These concepts constitute the fundamental background over which the execution environment of any large-scale service-oriented real-time architecture for highly interactive, distributed, and virtualized applications will be built in the future. Authors of Chapter 12 introduce related state of the art technologies regarding topics such as QoS provisioning, virtualization, and network resource management. This background is enriched with latest research results on future trends and advances in network management. Storage approaches are tackled in Chapter 13. Authors discuss how QoS provision within storage in distributed systems is possible given the behavior of the storage system under a variety of conditions dictated by the application and the network infrastructure. Besides, the authors present a QoS mechanism for data storage keeping in view the important parameters that come into play for the storage subsystem in a soft real-time cloud environment. Chapter 14 presents the concepts of fault detection and recovery, including terminology, classification of faults, and analysis of the key processes taking place in a system in order to diagnose and recover from failures. The state of the art mechanisms and techniques for fault detection and recovery are also analyzed, while recom-
mendations for applying them in Service Oriented Infrastructure are presented. The last book chapter, Chapter 15 provides an overview of different research efforts in the direction of enriching general purpose operating systems with real-time capabilities, with a particular focus on the Linux OS, since due to its open-source nature and wide diffusion and availability, Linux is one of the most widely used operating systems for such experimentations.

TARGET AUDIENCE

The target audience for the specific book ranges from PhD or Master’s students to researchers and teachers. The publication is a point of reference for real-time distributed systems, classifying more than 400 research efforts, to which the aforementioned audience can refer. From the domain or community point of view, researchers in real-time and distributed systems are naturally meant since they could discover through this book the particularities, strengths, weaknesses, and future trends of different research areas and potentially draft a “research agenda” for future projects according to these. Given the number of laboratories in universities and research institutes worldwide that deal with real-time systems, we expect this book to become a point of reference for the students and researchers in the multidisciplinary area of real-time in distributed computing.

CONCLUSION

Today, the Internet is the backbone of modern society and the global economy supporting almost every aspect of social and business interaction. The success of the global connectivity principle within the original Internet architecture has created and promoted new continuously evolving interaction models offering faster, more efficient, and richer collaborative and community experiences. In this context and given that sensors networks are emerging, focus is being put upon infrastructures that will be able to process and deliver in real-time interactive applications. To this direction, the publication provides a point of reference for the research community in the field of distributed environments that allows for the adoption of interactive real-time applications.

Dimosthenis P. Kyriazis
National Technical University of Athens, Greece

Theodora A. Varvarigou
National Technical University of Athens, Greece

Kleopatra G. Konstanteli
National Technical University of Athens, Greece