Ph.D. in Information Technology: Theses Defenses

July 17th, 2018

Seminar Room Alessandra Alario (building 21) – 2.30 pm

First Ph.D. presentation and discussion:

Eugenio GIANNITI – XXX Cycle

“Performance Models, Design and Run Time Management of Big Data Applications”

Advisor: Prof. **Danilo Ardagna**

Abstract:

Nowadays the big data paradigm is consolidating its central position in the industry, as well as in society at large. As data intensive applications (DIAs) gain more and more importance over time, it is fundamental for developers and maintainers to have the support of tools that enhance their efforts since early design stages and until run time.

Technically interesting scenarios, such as cloud deployments supporting a mix of heterogeneous applications, pose a series of challenges when it comes to predicting performance and exploiting this information for optimal design and management. Performance models, with their potential for what if analyses and informed design choices about DIAs, can be a major tool for both users and providers, yet they bring about a trade-off between accuracy and efficiency that may be tough to generally address. The picture is further complicated by the adoption of the cloud technology, which means that assessing operating costs in advance becomes harder, but also that the contention observed in data centers strongly affects big data applications' behavior. For all these reasons, ensuring quality of service (QoS) for novel DIAs is a difficult task that needs to be addressed in order to favor further development of the field.

Over this background, the present dissertation takes two main routes towards facing such challenges. At first we describe and discuss a number of performance models based on queueing networks (QNs), stochastic well formed nets (SWNs), and machine learning (ML). Among these, there are both basic models aimed at predicting specific metrics, like response time or throughput, and more specialized extensions that target the impact on big data systems of some design decisions, e.g., privacy preserving mechanisms or cloud pricing models. On top of this, the proposed models are variously positioned across the spectrum between efficiency and accuracy, thus enabling different trade-offs depending on the main requirements at hand. This is relevant in the second main part of this dissertation, where performance prediction is at the core of some formulations for capacity allocation and cluster management. In order to obtain optimal solutions to these problems, in one case at design time and in the other at run time, we adopt both mathematical programming an several performance models, according to the different constraints on solving times and accuracy.

Second Ph.D. presentation and discussion:

Francesco MARCONI – XXIX Cycle

"Formal Verification of TImed Properties for Data-Intensive Applications"

Advisor: Prof. **Matteo Rossi**

Abstract:

In the past few years, cloud-based enterprise applications, leveraging the so-called data-intensive technologies, have emerged as pervasive solutions for modern computing systems. Their adoption has been motivated by the growing need for systems that are able to collect, process, analyze and store huge quantities of data coming from various sources (social media, sensors, bank transactions, etc.) in a reasonable time. Data-intensive applications (DIAs), taking advantage of those technologies, natively support horizontal scalability, and constitute a significant asset for the production of large-scale software systems. However, the adoption of data-intensive technologies in the small and medium enterprises (SMEs), constituting the vast majority of the European industry, is still slow for a number of reasons, such as the steep learning curve of the technologies, the lack of experience and resources to keep up with such innovation. The definition of methodologies and principles for good software design is, therefore, fundamental to support the development of DIAs.

Non-functional (quality) requirements are an aspect of software design that is typically overlooked at design time and turns out to be crucial in later stages of development. Usually expressed in terms of Service Level Agreements (SLAs), if they are not met, further refinements of the applications are needed, resulting in additional costs. Design time quality analysis aims at detecting the presence of potential design flaws that could lead to later quality incidents, fostering the early detection of problems.

This thesis presents a unified model-driven approach for the formal analysis and verification of time-related quality properties for data-intensive applications. It addresses the computation of the two main classes of DIAs, i. e., streaming and batch processing, by proposing two formal models based on metric temporal logic and analyzable through state-of-the-art solvers.

Specifically, the formal models allow for the analysis of queue-boundedness in Storm-based streaming applications and deadline feasibility for Spark batch applications.

Both formalizations have been devised by extending the CLTLoc metric temporal logic with positive discrete counters and---albeit with relevant differences---capture the computational models of DIAs as DAGs, enriched with the most relevant quality aspects of the applications.

D-VerT is a model-driven tool that allows designers to perform formal analysis on streaming and batch applications by means of an automated toolchain, starting from suitably annotated UML diagrams. D-VerT embeds the models and analysis devised for Storm and Spark applications, and it automatically translates the UML design diagrams to the corresponding instances of the formal models, which are then fed to a state-of-the-art satisfiability solver. In this way, users with limited expertise of formal methods can benefit from the analysis without directly dealing with the specific formalisms.

The approach has been evaluated through experiments over a variety of use cases for the two kinds of analysis performed by D-VerT, showing promising results for both streaming and batch applications.

PhD Committee:

Prof. **Matteo Rossi**, DEIB
Prof. **David Carrera**, Universitat Politècnica de Catalunya (UPC)
Prof. **Oscar Corcho**, Universidad Politécnica de Madrid