Ph.D. in Information Technology
Thesis Defense

February 28th, 2025
at 13:00
Sala Donatori - Rettorato

Filippo CARLONI – XXXVII Cycle

Exploiting Hardware and Software Specialization for Efficient Regular Expressions Processing

Supervisor: Prof. Marco D. Santambrogio

Abstract:

The last decade has seen remarkable growth in the amount of data to analyze, challenging the current processing systems to keep pace with the increasingly high-computational and low-latency demand. These data must be correctly manipulated and filtered depending on the applicative fields to extract useful information. Within this scenario, pattern matching based on Regular Expressions (REs) is one of the most pervasive and helpful approaches for data analysis. Typical RE applications are packet filtering and malware analysis in computer security, genome analysis in bioinformatics, Natural Language Processing in speech recognition and synthesis, and database queries. However, REs are challenging to handle due to the intrinsical data-dependent computational pattern of their equivalent finite-state machine recognizers, severely limiting their optimization and use in computationally demanding or constrained scenarios. Additionally, dynamic scenarios such as computer security or database queries require updating REs at runtime, adding further complexity. Literature RE processing engines showcase a gap between flexibility and runtime adaptability on one side and efficient RE execution on the other. On the one hand, software libraries based on mainstream general-purpose architectures show extreme flexibility and runtime adaptability. However, they rely on high-end processors to provide attractive performance, and these solutions are generally unavailable in scenarios where processing has to be near the data, such as intrusion detection and prevention systems or in constraint devices, pushing the research in alternative directions. On the other hand, hardware-centric approaches exploit direct-logic mapping or in-memory automata representation to provide high-processing performance at the cost of limited flexibility, reduced RE features support, and high reconfiguration time of the patterns to analysis. For these reasons, other strategies consider using the REs as a language to represent automata via a sequence of instructions, thus increasing flexibility thanks to dedicated Instruction Set Architectures (ISAs) while providing efficient execution through optimized Domain-Specific Architectures (DSAs). However, these approaches are currently limited in advanced RE features support and simple REs-to-instructions translation, leaving plenty of room for improvement. In this context, this dissertation explores how to exploit hardware and software specialization through a tightly integrated approach, balancing software flexibility and optimization with hardware performance for efficient execution of REs. This dissertation starts by introducing a latency-oriented benchmarking methodology for REs, allowing a standard and replicable evaluation of current CPU and GPU RE-focused execution engines, thus

providing the basic structure for the proposed solution assessment. After that, this dissertation proposes an RE-tailored Instruction Set Architecture (ISA) based on the most widely employed RE features. On top of this ISA, this dissertation proposes a speculative DSA for efficient RE processing and a full-custom compiler based on the "RE as a Domain-Specific Language" concept to translate REs into optimized binaries, eventually executed on the target microarchitecture. These three components constitute an RE-tailored, completely optimized execution platform focused on low-latency and energy-efficient RE processing. Then, this dissertation explores the impact of multiple abstractions during the REs compilation and proposes a multi-dialect compiler based on the MLIR compiler infrastructure. Afterward, since the proposed approach focuses on the single-RE latency execution, this dissertation proposes a novel methodology to merge multiple REs into a single automaton, providing high throughput and reducing the memory utilization impact while processing multiple REs in parallel. Finally, this dissertation summarizes the proposed contributions, discusses its limitations, and explores potential future directions for the domain of REs.

Francesco PEVERELLI – XXXVII Cycle

On Novel Design Tools for Reconfigurable Spatial Architectures

Supervisor: Prof. Marco D. Santambrogio

Abstract:

FPGA-based reconfigurable architectures are set to play a pivotal role in the so-called new golden age of computer architecture, as the focus is shifting towards domain-specific hardware and accelerators. FPGAs not only represent a viable technology for hardware acceleration both in the embedded and data-center environment, but they are also an important prototyping tool to validate early prototypes of ASIC designs. Reconfigurable Spatial Architectures (RSA) want to abstract from the fine-grained programmability of FPGAs to a higher level of functional abstraction but without restricting themselves to a particular interconnection topology and programming model, as is more typical for CGRAs. However, maintaining this flexibility also means dealing with an extremely large design space. For these reasons, streamlining the process of implementing and optimizing RSAs is especially valuable, saving designers hours of development time and effort. This thesis proposes several tools designed to aid in several stages of FPGA-based RSA development, from identifying suitable acceleration targets to the early prototyping of several architectural variations to optimizing existing designs via Design Space Exploration. The works presented in this thesis push the state-of-the-art RSA design tools involving compilers, programming languages, and hardware architecture optimization problems to facilitate the design of better and more performing RSAs with reduced complexity and effort.

PhD Committee

Prof. Donatella Sciuto, Politecnico di Milano

Prof. Juergen Becker, Karlsruhe Institute of Technology

Prof. Dionisios Pnevmatikatos, University of Athens