Public Abstract

First Name:Xiaodong

Middle Name:

Last Name:Yu

Adviser's First Name:Michela

Adviser's Last Name:Becchi

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:SS 2013

Department:Electrical Engineering

Degree:MS

Title:DEEP PACKET INSPECTION ON LARGE DATASETS: ALGORITHMIC AND PARALLELIZATION TECHNIQUES FOR ACCELERATING REGULAR EXPRESSION MATCHING ON MANY-CORE PROCESSORS

Regular expression matching is a central task in several networking (and search) applications and has been accelerated on a variety of parallel architectures, including general purpose multi-core processors, network processors, field programmable gate arrays, ASIC- and TCAM-based systems. All of these solutions are based on finite automata (either in deterministic or non-deterministic form), and mostly focus on effective memory representations for such automata. More recently, a handful of proposals have exploited the parallelism intrinsic in regular expression matching (i.e., coarse-grained packet-level parallelism and fine-grained data structure parallelism) to propose efficient regex matching designs for GPUs. However, most GPU solutions aim at achieving good performance on small datasets, which are far less complex and problematic than those used in real-world applications.

In this work, we provide a more comprehensive study of regular expression matching on GPUs. To this end, we consider datasets of practical size and complexity and explore advantages and limitations of different automata representations and of various GPU implementation techniques. Our goal is not to show optimal speedup on specific datasets, but to highlight advantages and disadvantages of the GPU hardware in supporting state-of-the-art automata representations and encoding schemes, schemes which have been broadly adopted on other parallel memory-based platforms.