TECHNIQUES FOR EFFICIENT REGULAR EXPRESSION MATCHING ACROSS HARDWARE ARCHITECTURES

Xiang Wang

Dr. Michela Becchi, Thesis Supervisor

ABSTRACT

Regular expression matching is a central task for many networking and bioinformatics applications. Graphical Processing Units (GPUs) offer a highly parallel platform for memory-based implementations, while Field Programmable Gate Arrays (FPGAs) support reconfigurable, logic-based solutions for regular expression matching. In addition, Micron Technology's Automata Processor is a memory-based, reprogrammable hardware device. All these regular expression matching engines are based on finite automata, either in their non-deterministic or in their deterministic form (NFA and DFA, respectively).

In this work, we aim to implement highly parallel memory-based and logic-based regular expression matching solutions. First, we explored compression techniques and regular expression clustering algorithms to alleviate the memory pressure of DFA-based GPU implementations. Second, we developed a parser using our internal representation for Automata Networks (an extension of NFA) defined through Micron's Automata Network Markup Language (ANML), a XML-based language designed for the Automata Processor. Finally, we implemented a tool to convert our internal representation to Verilog, thus allowing automatic deployment on FPGA and the comparison of performances between FPGA and Micron's Automata Processor.