

# High-speed Regular Expression Matching with Pipelined Memory-based Automata

Denis Matoušek, Jiří Matoušek and Jan Kořenek

Brno University of Technology, Faculty of Information Technology, Brno, Czech Republic

Email: {imatousekd|imatousek|korenek}@fit.vutbr.cz

**Abstract**—The paper proposes an architecture of a high-speed regular expression (RE) matching system with fast updates of an RE set. The architecture uses highly memory-efficient Delayed Input DFAs (D<sup>2</sup>FAs), which are organized to a processing pipeline. The architecture is designed so that it communicates only locally among its components in order to achieve high frequency even for a large number of parallel matching engines (MEs), which allows scaling throughput to hundreds of gigabits per second (Gbps). The architecture is able to achieve processing throughput of up to 400 Gbps on current FPGA chips.

**Keywords**—Regular expression matching; 100 Gbps; 400 Gbps; Delayed Input DFA; Pipelined automata

The proposed architecture addresses the drawbacks of previously published architectures that include limited processing throughput not scaling above 100 Gbps and inability to change the RE set without FPGA reconfiguration. The architecture is based on *pipelined automata* [1]. They are used to avoid a complex interconnection network between parallel MEs and a shared packet buffer. Instead, the parallel MEs are connected directly to the shared packet buffer and only a state is passed between the MEs. In order to achieve fast RE set change without interruption of data processing, a memory-based approach, D<sup>2</sup>FA [2], is used. D<sup>2</sup>FA also reduces the size of a transition table by replacing multiple transitions of the original automaton with *default* transitions.

The architecture is depicted in Figure 1. It uses an array of  $k$  MEs with pipelined automata, which is shown in the upper part of the schematics. Input data is stored into the shared packet buffer in a parallel way. The detail of a ME is shown in the lower part of the schematics. Each ME employs two D<sup>2</sup>FAs to ensure full throughput for D<sup>2</sup>FAs with radius of one. However, this arrangement also allows to execute more than one default transition to accept a single input symbol. Default transitions are stored in FIFO<sub>default</sub> while standard transitions in FIFO<sub>standard</sub>.

Synthesis results for a test set of 12 REs summed up in Table I show that the most critical resources of the architecture are block memories. The table shows resource utilization for the whole parallel pipeline with 64 MEs (to achieve 100 Gbps throughput) and 256 MEs (to achieve 400 Gbps throughput). Note that 64 parallel MEs fit into Xilinx Virtex-7 VH580T chip (61.3% of block memories are used) and 256 MEs fit into Xilinx Virtex UltraScale+ VU13P chip (85.7% of block memories are used). Since the circuit is deeply pipelined using synchronous block memories and

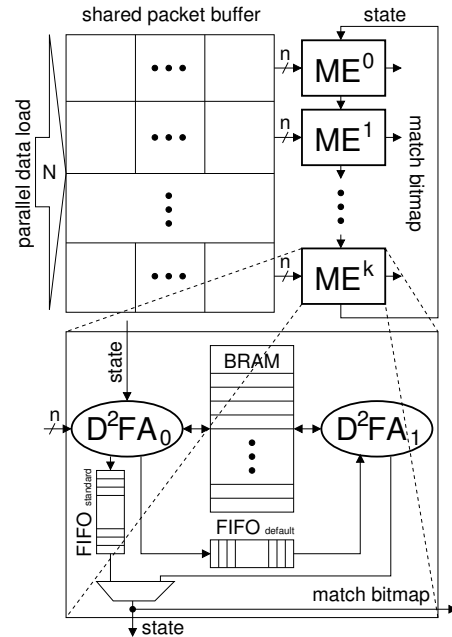


Figure 1. Pipelined Memory-based Automata

FIFOs in all stages, the required frequency of 200 MHz is met. It means that the architecture is scalable and allows to achieve 400 Gbps throughput with current FPGA chips.

Table I  
RESOURCE UTILIZATION (A TEST SET OF 12 RES, D<sup>2</sup>FA RADIUS TWO)

Component	Throughput	BRAMs	LUTs	FFs
Single ME	–	9	201	124
<b>64 MEs</b>	<b>100 Gbps</b>	<b>576</b>	<b>12 864</b>	<b>7936</b>
<b>256 MEs</b>	<b>400 Gbps</b>	<b>2304</b>	<b>51 456</b>	<b>31 744</b>

## REFERENCES

- [1] D. Matoušek, J. Kořenek, and V. Puš, “High-speed regular expression matching with pipelined automata,” in *Proc. of FPT’16*, Dec. 2016, pp. 93–100.
- [2] S. Kumar, S. Dharmapurikar, F. Yu, P. Crowley, and J. Turner, “Algorithms to accelerate multiple regular expressions matching for deep packet inspection,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 339–350, Aug. 2006, ISSN 0146-4833.