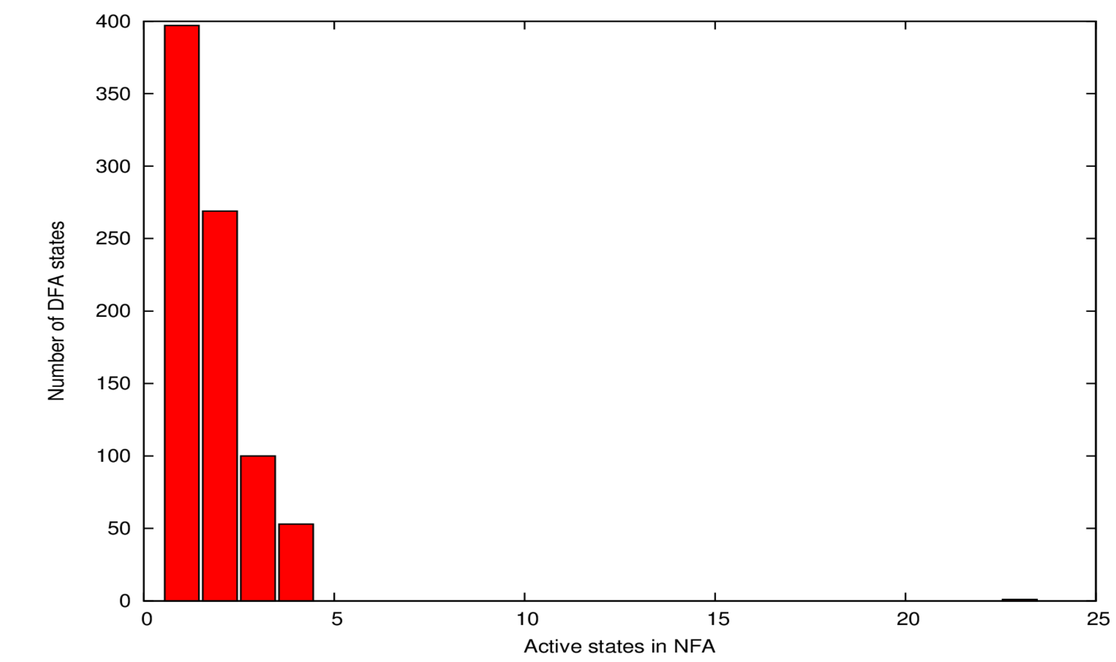
NFA split architecture

Motivation

Regular expression matching is a time-critical operation in network security devices. Thousands of regular expressions have to be matched at multi-gigabit speed to detect attack, worms or viruses in current IDS/IPS systems. Software implementations have limited throughput to hundreds Mbps. Many hardware architectures have been proposed but they can achieve high throughput only at the cost of many FPGA resources or large and fast memories. Reduction of FPGA logic or memory requirements have direct impact to required size of FPGA and price of security devices.

Efficiency of Current NFA Mapping Techniques

We have evaluated efficiency of current NFA mapping techniques. Regular expressions from L7 decoder ware transformed to NFA and than to DFA. Every DFA state is represented by set of active NFA states. We have created histogram of DFA states according to the number of representing NFA states.



Maximal amount of active NFA states is usually less than 5%.

Design

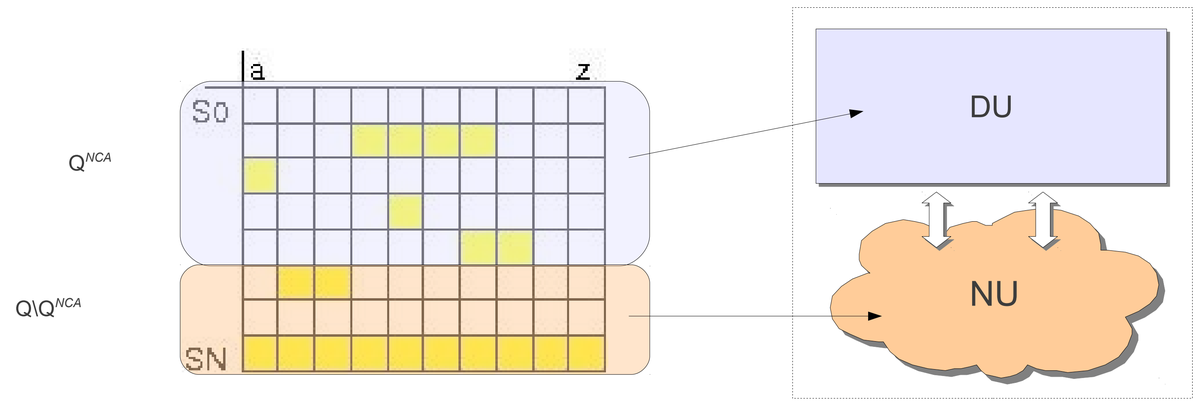
Because only very small subset of states can be active, it is possible to split NFA into several parts. States which can not be active at once are mapped into DFA and remaining states are mapped into NFA. More than one DFA can be created. Those parts communicate when transitions exist between them.

Set of states without collisions (Q^{NCA}) is computed from DFA states. NFA states forming DFA states are in collision. The set of states without collision is then generated by heuristic from sets of collision states.

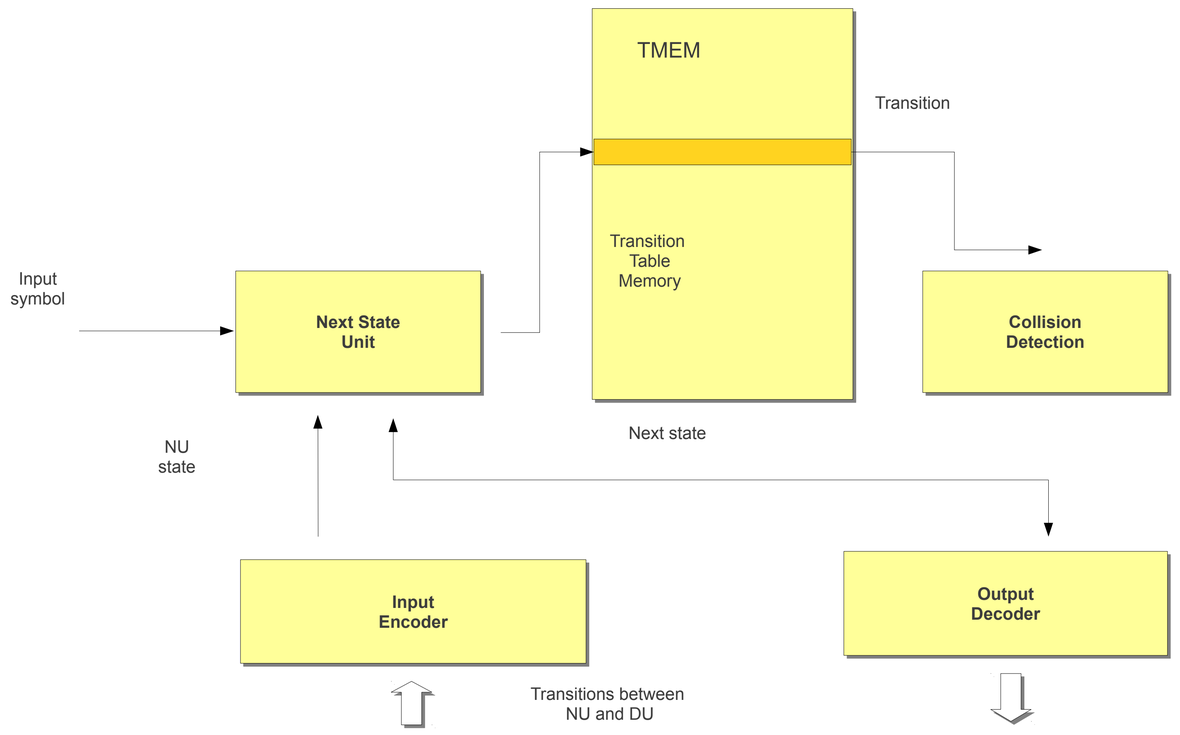
The NFA split architecture reduces 66.8% LUTs and 43.3% flip-flops in average for all selected sets of regular expressions (from L7 Decoder and Snort IDS).

NFA split architecture

States without collisions is represented by deterministic unit (DU), which uses binary encoding of states and appropriate part of transition table can be stored in memory. Others states and transitions are mapped to the FPGA logic.



Transition table of DU is stored in a memory (TMEM) and owerlapping of rows is used. Collisions are detected by input symbol and they are also stored in TMEM for every transition.



Publications

- Kořenek, J., Košař, V.: Efficient Mapping of Nondeterministic Automata to FPGA for Fast Regular Expression Matching, In: Proceedings of the 13th IEEE International Symposium on Design and Diagnostics of Electronic Circuits and Systems DDECS 2010, Vienna, AT, IEEE CS, 2010, p. 6, ISBN 978-1-4244-6610-8

- Kořenek, J., Košař, V.: NFA Split Architecture for Fast Regular Expression Matching, In: Proceedings of the 6th ACM/IEEE Symposium on Architectures for Networking and Communications Systems, La Jolla, US, ACM, 2010, p. 2, ISBN 978-1-4503-0379-8

- Kořenek, J.: Fast Regular Expression Matching Using FPGA, In: Information Sciences and Technologies Bulletin of the ACM Slovakia, Vol. 2, No. 2, 2010, Bratislava, SK, p. 103-111, ISSN 1338-1237

- Kořenek, J.: Rychlé vyhledávání regulárních výrazů s využitím technologie FPGA, Brno, CZ, UPSY FIT VUT, 2010, p. 105

esf european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

Leták byl vytvořen v rámci projektu
TeamIT - Budování konkurenceschopných výzkumných týmů pro IT,
MŠMT, CZ.1.07/2.3.00/09.0067, 2009-2012