NFA reductions

Motivation

Intrusion Detection Systems have to match large sets of regular expressions to detect malicious traffic on multi-gigabit networks. Regular expression matching is a time critical operation so hardware acceleration is needed. Many algorithms and architectures have been proposed to accelerate pattern matching, but formal methods for reduction of NFA have not been used yet. NFA minimization can not be used as it is a PSPACE-complete problem. Reduction algorithms usually have polynomial time and space complexity. Reductions of state graphs are commonly studied and used in formal verification for state reduction.

Reduction Algorithms

State reduction algorithms with polynomial time and space complexity can be divided into three groups acording to their mathematical foundations:

1. Modified subset construction (MSC)
   - Subset construction is modified to avert state explosion
2. Equivalence (EQ)
   - Left and right language equivalence is computed on NFA states
   - Left and right language equivalence can be used repeatedly for NFA reduction
3. Preorders
   - Left and right language preorder is computed on NFA states
   - Various algorithms for construction of state equivalence from preoreders and NFA reduction exists:
     1. Repeated use of left and right preorders (PRE1)
     2. Repeated simultaneous use of left and right preorders (PRE2)
     3. Mediated preorders (MPRE)

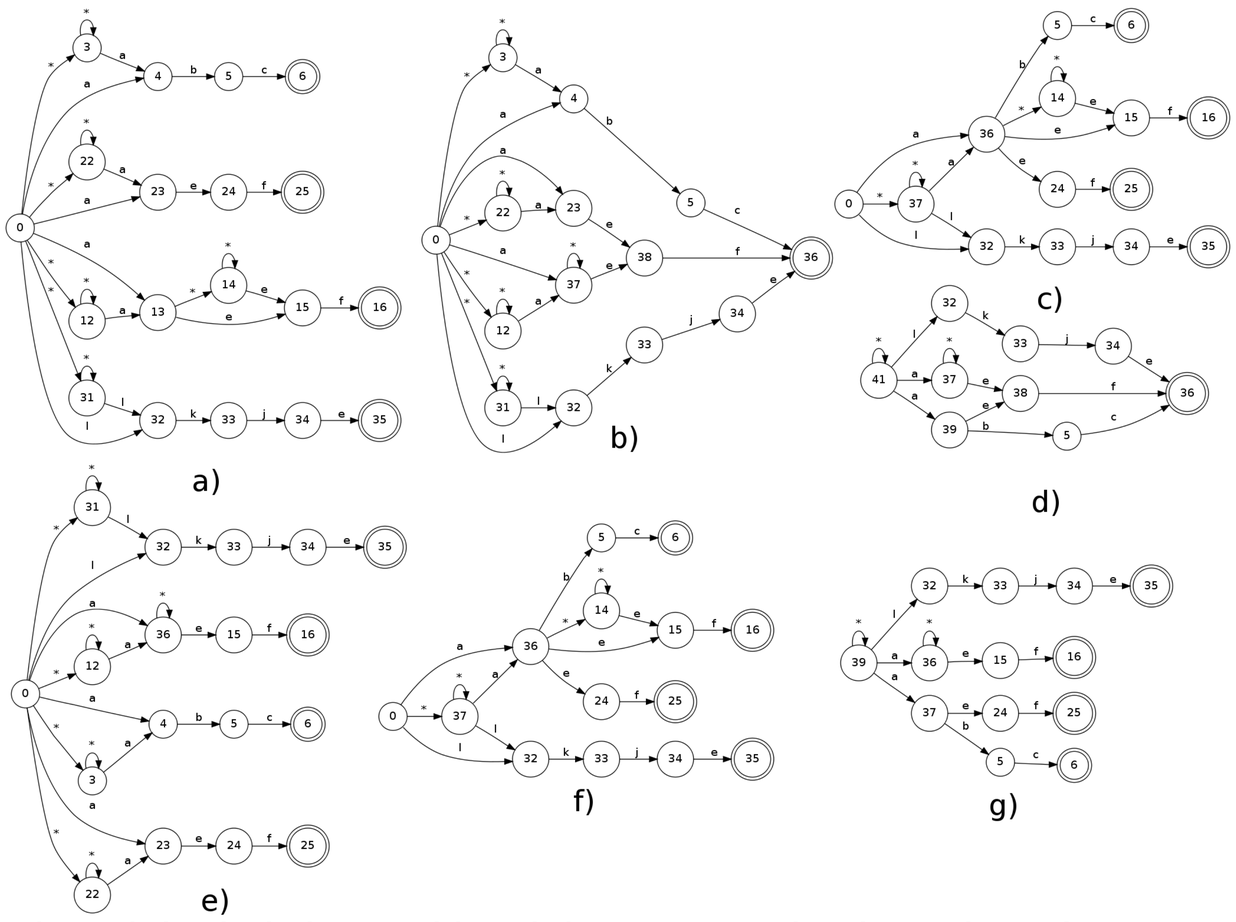
Extensions for Regular Expressions Searching

Because NFA reductions join final states of original NFA into one new final state, relation between final states and regular expression (RE) numbers is lost. We have therefore introduced extension of NFA – Multilingual NFA (MNFA). MNFA adds set of language labels and unary function for mapping final states to labels. We have also modified reduction methods mentioned above to work on MNFA.

Comparison of various NFA reduction methods [states]

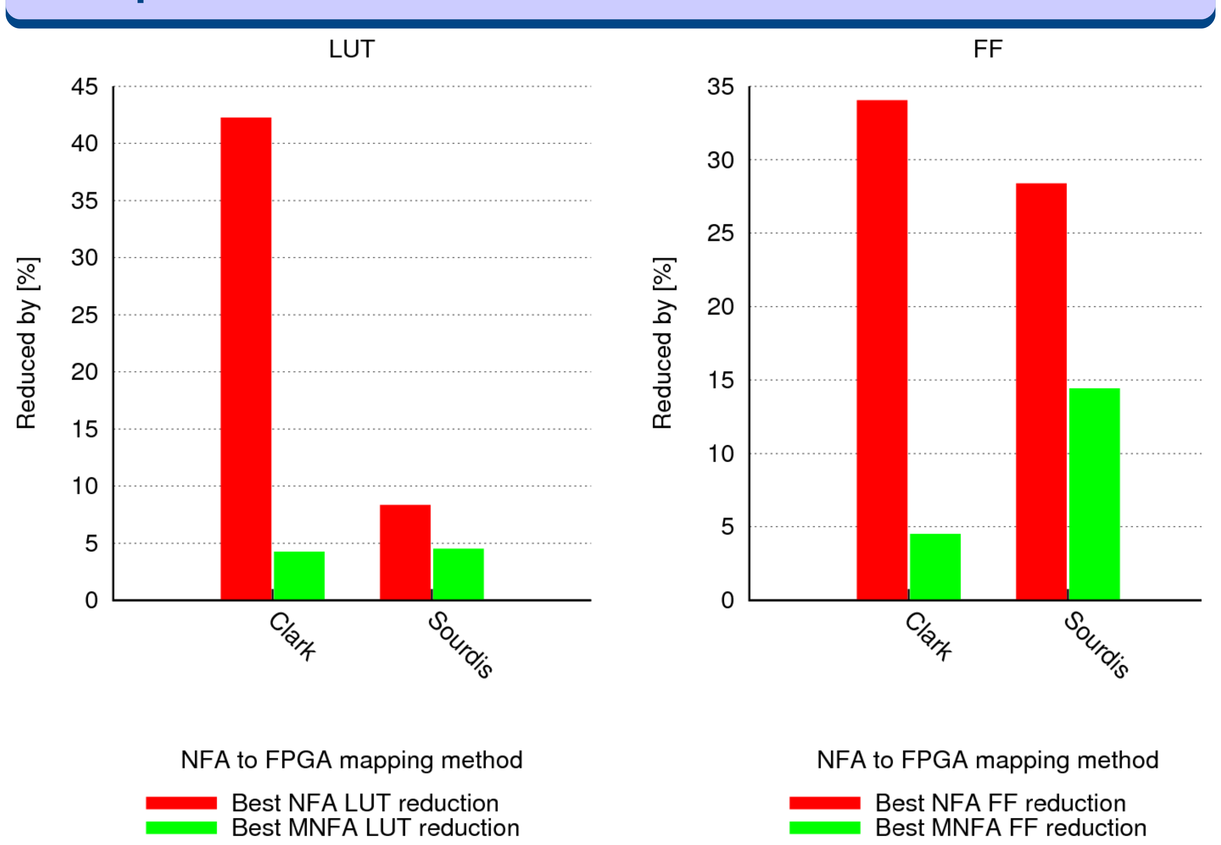
RE set	MSC [%]	EQ [%]	PRE1 [%]	PRE2 [%]	MPRE [%]
L7	0.26	18.84	18.97	18.97	18.84
backdoor	1.25	18.30	18.53	18.65	18.33
web-php	6.21	46.27	46.58	46.63	45.96
ftp	0.64	63.34	63.35	63.35	63.34
nntp	-0.2	57.49	57.64	57.64	57.49
voip	0.69	42.16	42.29	42.29	42.22
Snort 25	-0.24	34.86	35.03	35.17	35.02

Example of (M)NFA reduction by equivalence



a) Original (M)NFA b) Right NFA reduction c) Left NFA reduction d) Repeated left and right NFA reduction e) Right MNFA reduction f) Left MNFA EQ reduction g) Repeated left and right MNFA reduction

Comparison of NFA and MNFA reduction on Snort RE



Publications

- Košař, V., Kořenek, J.: Reduction of FPGA Resources for Regular Expression Matching by Relation Similarity, In: IEEE Design and Diagnostics of Electronic Circuits and Systems DDECS'2011, Cottbus, DE, IEEE CS, 2011, p. 401-402, ISBN 978-1-4244-9753-9
- Košař, V.: Redukce Zabraných Zdrojů FPGA pro Vyhledávání Vzorů Popsaných Regulárními Výrazy, In: Počítačové architektury a diagnostika 2011, Stará Lesná, SK, FIIT STU, 2011, p. 6, ISBN 978-80-227-3552-0

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