SLOVENSKÁ AKADÉMIA VIED
MATEMATICKÝ ÚSTAV

Mgr. Kristína Čevorová

Autoreferát dizertačnej práce

State Complexity and Magic Number Problem for Unary Operations

Na získanie akdemického titulu philosophiae doctor
v odbore doktorandského štúdia: 9.1.9 aplikovaná matematika

Bratislava 2016

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Matematickom ústave Slovenskej akadémie vied v Bratislave.

Predkladateľ: Mgr. Kristína Čevorová
Matematický ústav
Slovenskej akadémie vied
Štefánikova 49
814 73 Bratislava

Školiteľ: RNDr. Galina Jirásková, CSc.
Matematický ústav
Slovenskej akadémie vied
Grešákova 6
040 01 Košice

9.1.9 aplikovaná matematika

Predseda odborovej komisie:
prof. RNDr. Daniel Ševčovič, CSc.
Fakulta matematiky, fyziky a informatiky
Univerzity Komenského
Mlynská dolina 842 48 Bratislava

1 Introduction

The state complexity studies the succinctness of regular language representation deterministic finite automata. For a regular language it is the number of states of the minimal deterministic finite automaton recognizing this languages. For a regular operation it represents the worst case state complexity of languages resulting from this operation, considered as a function of the state complexities of the operands.
The first investigated state complexity-related problem was the cost of NFA to DFA conversion. Rabin and Scott showed that any n-state NFA could be simulated by a 2^n-state DFA [RS59]. Moreover 2^n states are not only sufficient, but also necessary – this was at first shown by Soviet scientists [Yer62, Lup63], but this went unnoticed by the western world and equivalent results were independently proven in [Moo71, MF71].
The operational state complexity studies the relationship between the state complexity of inputs and the output of a regular operation. The pioneering works from 1970s [Mas70, Mir66] were isolated for a long time before the gradual renewal of interest in late 1980s. This period is characteristic by unsystematic results like [RI89, Bir91]; Birget introduced the term state complexity [Bir91]. A boom followed; it was initiated by the very systematic paper of Yu, Salomaa and Zhuang; they studied all basic regular operation on regular and unary languages [YZS94]. This interest still persists and the current state of art is well-documented in a technical report by Gao, Moreira, Reis and Yu [GMRY15].
One of the current avenues of research on state complexity is a systematic study of basic regular operations on various subregular families: finite languages [CCSY01], cofinite languages [BGN10], free languages [BJLS11], star-free languages [BL12], ideal languages [BJL13], and closed languages [BJZ14]. In the light of this extensive research on other regular operations, the square operation has been incomprehensibly neglected, the only results are on regular and unary languages [Ram06].

The state complexity reflects only the worst case. It does not say anything about any other case. Two different approaches to fix this have emerged so far. On the one hand, Nicaud studied an average case [Nic99]. Since the mere enumeration of all automata with a given state complexity is already too difficult, he limited himself to basic operations on unary automata.
The magic number problem approach seems to be more fruitful. The original magic number problem was introduced by Iwama, Kambayashi and Takaki at the Third Conference on Developments in Language Theory. Their question was whether, given any integers n and α with $n \leq \alpha \leq 2^n$, we are able to find a binary language with nondeterministic state complexity n and deterministic state complexity α [IKT00]. If this is not possible, the number α is called magic for n [IMP00].
Seemingly simple question turned out to be hard and in an attempt to tackle this problem, the condition on alphabet was relaxed. First it was shown, that if an exponentially growth of an alphabet is allowed, there are no magic numbers [Jir01]. Then the growth of an alphabet was limited to linear [Gef07b] but a breakthrough was a proof that a constant four letter alphabet suffices [JJS08]. Further improvement was a ternary alphabet [Jir11b]. Despite many efforts, it remains an open problem, whether binary alphabet suffices; several papers addressed this problem, identifying various families of non-magic numbers including [IKT00, IMP00, Jir01, Gef05, MS08, Jir08]. Moreover, if we restrict ourselves to certain subregular language classes, the hierarchy remains contiguous [HJK12]. The situation is different for unary languages. Since $e^{O(\sqrt{n \log n})}$ states are sufficient for a DFA simulating any n-state NFA [Chr86], the relevant interval is $[n, e^{O(\sqrt{n \log n})}]$. And indeed, not every α from this interval is attainable. Actually, in some sense the non-magic numbers are rare [Gef07a].
The generalization of the magic number problem from a determinization to regular operations has the same principle, as the generalization from the cost of NFA to DFA transformation to operational state complexity. The operational magic number problem is so far studied much less systematically than

the original magic number problem or the operational state complexity. These are all known results concerning DFAs:
There are no magic numbers for *union* on a binary alphabet [HJS05]. From the De Morgan's law and the fact that complementation preserves state complexity follows that neither there are magic numbers for *intersection* on a binary alphabet. If we allow a growing alphabet, there are no magic numbers for *concatenation* [Jir11a], for *reversal* on a linear $2n-2$ letter alphabet [Šeb13], and for a *Kleene star* on a linear $2n$ alphabet [JPŠ14].
On the other hand, there are three cases, when magic number do exist: *cyclic shift* on fixed alphabet [JO08], *conversion of a unary DFA to NFA* [Gef07a], and *conversion of unary symmetric difference NFA to classic DFA* [vZ05].

Questions concerning state complexity and the magic number problem are natural and would be understood by researchers since the very beginning of computer science. Why nobody asked them? One of the reasons for a long hibernation of this field during 70s and 80s may be the fact that until the 1990s, the computational power was not sufficient enough to aid in constructing hypotheses. This may be especially true for the magic number problem that needs even more computational resources and indeed it did not emerge until the early 2000s.

2 Aims

Th first of our aims is to determine the state complexity of the square operation on subregular classes in free, closed and ideal languages families and fill the gap left in the literature. Our second aim is to study the magic number problem on unary languages for Kleene star and square.

3 Main Results – State Complexity

Results on concatenation in [BJLS11, BJZ14, BJL13] provide an instant upper bound on the state complexity of square operation. Yet poorer upper bound can be obtained from the results on square on regular languages in [Ram06].
Table 1 provides a summary of our results and a comparison with these results.

		square	$\lvert\sum\rvert$	concatenation	$\lvert\sum\rvert$
ideal	unary	$\mathbf{2n-1}$		$m+n-1$	
	right	$\mathbf{n+2^{n-2}}$	2	$m+2^{n-2}$	2
	left, 2-sided, all-sided	$\mathbf{2n-1}$	1	$m+n-1$	1
closed	unary	$\mathbf{2n-2}$		$m+n-2$	
	suffix	$\mathbf{\frac{1}{2}(n^2+n)-1}$	3	$(m-1)n+1$	3
	prefix	$\mathbf{(n+4)2^{n-3}-1}$	2	$(m+1)2^{n-1}$	3
	factor, subword	$\mathbf{2n-1}$	2	$m+n-1$	2
free	unary	$\mathbf{2n-2}$		$m+n-2$	
	prefix, bifix, factor, subword	$\mathbf{2n-2}$	1	$m+n-2$	1
	suffix	$\mathbf{n2^{n-3}+1}$	3	$(m-1)2^{n-1}+1$	3
regular	unary	$2n-1$		mn	if $(m,n)=1$
	general	$n2^n-2^{n-1}$	2	$m2^n-2^{n-1}$	2

Table 1: Comparison of our results on the state complexity of square with concatenation.

Let us discuss these results in more detail. Witnesses for the bounds for concatenation could be reused as witnesses for square for all ideal languages except for right ideal languages; all free languages except for suffix-free languages and for unary-closed languages. We showed that the bound obtained from concatenation is also tight for right-ideal and factor- or subword-closed languages, but we had to find new witnesses to prove it.
The rest of results shows that upper bounds from concatenation are only asymptotically tight. For suffix-closed languages the quadratic bound differs by a factor $\frac{1}{2}$. We provided a new upper bound and proved its optimality on a ternary alphabet. Tightness on a binary alphabet remains an open question, but computations suggest that this upper bound cannot be attained by a binary language. For suffix-free languages the state complexity is exponential and differs by a factor $\frac{1}{4}$. We obtained an upper bound and proved its tightness again on a ternary alphabet. Tightness on a binary alphabet is an open problem. Also for prefix-closed languages, the new bound differs from the exponential bound for concatenation by a factor of $\frac{1}{4}$. In this case, the tightness is shown by a binary alphabet, as opposed to the bound on concatenation, where a ternary alphabet is used.
Note that all results are shown using a constant alphabet of size at most 3 and except for suffix-free and suffix-closed languages, the alphabet size is optimal.

4 Main Results – Magic Numbers

The magic number problem for the square (the Kleene star) is stated as follows: is there for every n and and α such that $1 \leq \alpha \leq 2n-1$ (for the Kleene star such that $1 \leq \alpha \leq (n-1)^2+1$) a language with the state complexity n and the state complexity of its square (its Kleene star) α? Or are there any gaps – magic numbers?

Square We showed that unless $n < 5$, all values between 1 and $2n-1$ are attainable and there are no magic numbers. Value 1 is magic for $n = 2$ and value 2 is magic for $n = 3$ or 4.

Kleene Star First we decided the magicness of values higher than $n^2 - 4n + 6$. We found that only three or four are non-magic, depending on the parity of n, and integers in two linear intervals between these values are magic. We conjure that there are many more magic numbers below this bound but their existence is probably highly dependent on the number-theoretical properties of n.
Magic numbers cannot be too small – we gave a lower bound $n + 1$ on the smallest magic number. However, computations suggest that this bound is not even asymptotically optimal. A better bound is an open problem. Our results on which numbers are magic are summarized in Figure 1.

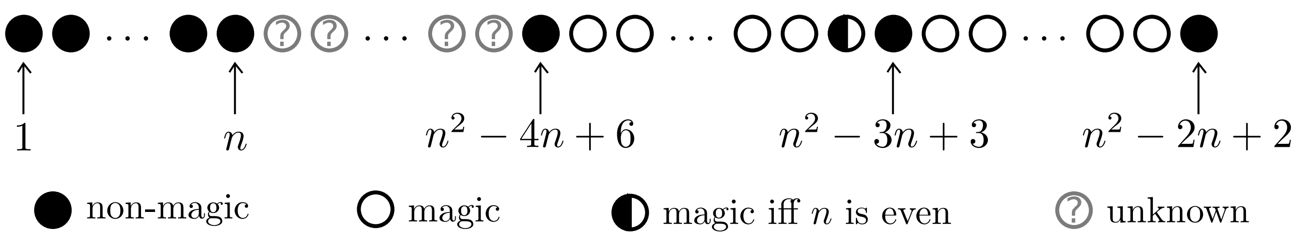

Figure 1: Summary of known magic numbers for Kleene star on unary languages.

As a corollary we also proved in this section that the conjecture from [EHJ13, Section 6] that the state complexity of Kleene star on unary non-returning DFAs is $n^2 - 4n + 6$, is true.
In the end we outlined a connection with the Frobenius problem and derived a formula for a special case of its generalization.

Related Published Papers (AEC)

[Čev13] K. Čevorová. Kleene star on unary regular languages. In H. Jürgensen and R. Reis, editors, *DCFS*, volume 8031 of *LNCS*, pages 277–288. Springer, 2013.

[Čev15] K. Čevorová. Square on ideal, closed and free languages. In J. Shallit and A. Okhotin, editors, *DCFS*, volume 9118 of *LNCS*, pages 70–80. Springer, 2015.

[ČJK14] K. Čevorová, G. Jirásková, and I. Krajňáková. On the square of regular languages. In M. Holzer and M. Kutrib, editors, *CIAA*, volume 8587 of *LNCS*, pages 136–147. Springer, 2014.

[ČJM+14] K. Čevorová, G. Jirásková, P. Mlynárčik, M. Palmovský, and J. Šebej. Operations on automata with all states final. In Z. Ésik and Z. Fülöp, editors, *AFL*, volume 151 of *EPTCS*, pages 201–215, 2014.

Cited in

[ČJK14] K. Čevorová, G. Jirásková, and I. Krajňáková. On the square of regular languages. In M. Holzer and M. Kutrib, editors, *CIAA*, volume 8587 of *LNCS*, pages 136–147. Springer, 2014.

[GMRY15] Y. Gao, N. Moreira, R. Reis, and S. Yu. A survey on operational state complexity. *CoRR*, abs/1509.03254, 2015.

[JJS15] J. Š. Jirásek, G. Jirásková, and A. Szabari. Operations on self-verifying finite automata. In L. D. Beklemishev and D. V. Musatov, editors, *CSR*, volume 9139 of *LNCS*, pages 231–261. Springer, 2015.

[JP13] G. Jirásková and M. Palmovský. Kleene closure and state complexity. In T. Vinař, editor, *ITAT*, volume 1003 of *CEUR Workshop Proceedings*, pages 94–100. CEUR-WS.org, 2013.

[JPŠ14] G. Jirásková, M. Palmovský, and J. Šebej. Kleene closure on regular and prefix-free languages. In M. Holzer and M. Kutrib, editors, *CIAA*, volume 8587 of *LNCS*, pages 226–237. Springer, 2014.

Bibliography

[BGN10] F. Bassino, L. Giambruno, and C. Nicaud. Complexity of operations on cofinite languages. In A. López-Ortiz, editor, *LATIN*, volume 6034 of *LNCS*, pages 222–233. Springer, 2010.

[BHK09] H. Bordihn, M. Holzer, and M. Kutrib. Determination of finite automata accepting subregular languages. *Theoretical Computer Science*, 410(35):3209–3222, 2009.

[Bir91] J.-C. Birget. Intersection of regular languages and state complexity. *SIGACT News*, 22(2):49, may 1991.

[BJL13] J. A. Brzozowski, G. Jirásková, and B. Li. Quotient complexity of ideal languages. *Theoretical Computer Science*, 470:36–52, 2013.

[BJLS11] J. A. Brzozowski, G. Jirásková, B. Li, and J. Smith. Quotient complexity of bifix-, factor-, and subword-free regular languages. In P. Dömösi and S. Iván, editors, *AFL*, pages 123–137, 2011.

[BJZ14] J. A. Brzozowski, G. Jirásková, and C. Zou. Quotient complexity of closed languages. *Theory of Computing Systems*, 54(2):277–292, feb 2014.

[BL12] J. A. Brzozowski and B. Liu. Quotient complexity of star-free languages. *International Journal of Foundations of Computer Science*, 23(06):1261–1276, 2012.

[Brz10] J. A. Brzozowski. Quotient complexity of regular languages. *Journal of Automata, Languages and Combinatorics*, 15(1/2):71–89, 2010.

[CCSY01] C. Câmpeanu, K. Culik, K. Salomaa, and S. Yu. State complexity of basic operations on finite languages. In O. Boldt and H. Jürgensen, editors, *WIA*, volume 2214 of *LNCS*, pages 60–70. Springer, 2001.

[Čev13] K. Čevorová. Kleene star on unary regular languages. In H. Jürgensen and R. Reis, editors, *DCFS*, volume 8031 of *LNCS*, pages 277–288. Springer, 2013.

[Čev15] K. Čevorová. Square on ideal, closed and free languages. In J. Shallit and A. Okhotin, editors, *DCFS*, volume 9118 of *LNCS*, pages 70–80. Springer, 2015.

[Chr86] M. Chrobak. Finite automata and unary languages. *Theoretical Computer Science*, 47:149 – 158, 1986.

[CJ11] R. Cmorik and G. Jirásková. Basic operations on binary suffix-free languages. In Z. Kotásek, J. Bouda, I. Černá, L. Sekanina, T. Vojnar, and D. Antos, editors, *MEMICS*, volume 7119 of *LNCS*, pages 94–102. Springer, 2011.

[ČJK14] K. Čevorová, G. Jirásková, and I. Krajňáková. On the square of regular languages. In M. Holzer and M. Kutrib, editors, *CIAA*, volume 8587 of *LNCS*, pages 136–147. Springer, 2014.

[EHJ13] H.-S. Eom, Y.-S. Han, and G. Jirásková. State complexity of basic operations on non-returning regular languages. In H. Jurgensen and R. Reis, editors, *Descriptional Complexity of Formal Systems*, volume 8031 of *LNCS*, pages 54–65. Springer, 2013.

[Gef05] V. Geffert. (Non)determinism and the size of one-way finite automata. In C. Mereghetti, B. Palano, G. Pighizzini, and D. Wotschke, editors, *DCFS*, pages 23–37. Università degli Studi di Milano, Milan, Italy, 2005.

[Gef07a] V. Geffert. Magic numbers in the state hierarchy of finite automata. *Information and Computation*, 205(11):1652–1670, 2007.

[Gef07b] V. Geffert. State hierarchy for one-way finite automata. *Journal of Automata, Languages and Combinatorics*, 12(1-2):139–145, 2007.

[GMRY15] Y. Gao, N. Moreira, R. Reis, and S. Yu. A survey on operational state complexity. *CoRR*, abs/1509.03254, 2015.

[HJK12] M. Holzer, S. Jakobi, and M. Kutrib. The magic number problem for subregular language families. *International Journal of Foundations of Computer Science*, 23(1):115–131, 2012.

[HJS05] M. Hricko, G. Jirásková, and A. Szabari. Union and intersection of regular languages and descriptional complexity. In C. Mereghetti, B. Palano, G. Pighizzini, and D. Wotschke, editors, *DCFS*, pages 170–181. Università degli Studi di Milano, Milan, Italy, 2005.

[Hop71] J. E. Hopcroft. An n log n algorithm for minimizing states in a finite automaton. Technical report, Stanford University, Stanford, CA, USA, 1971.

[HS07] Y.-S. Han and K. Salomaa. State complexity of basic operations on suffix-free regular languages. In L. Kučera and A. Kučera, editors, *Mathematical Foundations of Computer Science 2007*, volume 4708 of *LNCS*, pages 501–512. Springer, 2007.

[HS09] Y.-S. Han and K. Salomaa. State complexity of basic operations on suffix-free regular languages. *Theoretical Computer Science*, 410(27–29):2537 – 2548, 2009.

[HSW06] Y. Han, K. Salomaa, and D. Wood. State complexity of prefix-free regular languages. In H. Leung and G. Pighizzini, editors, *DCFS*, pages 165–176. New Mexico State University, Las Cruces, New Mexico, USA, 2006.

[HSW09] Y.-S. Han, K. Salomaa, and D. Wood. Operational state complexity of prefix-free regular languages. In Z. Ésik and Z. Fülöp, editors, *AFL*, pages 99–115. Institute of Informatics, University of Szeged, Hungary, 2009.

[IKT00] K. Iwama, Y. Kambayashi, and K. Takaki. Tight bounds on the number of states of DFAs that are equivalent to n-state NFAs. *Theoretical Computer Science*, 237(1-2):485–494, 2000.

[IMP00] K. Iwama, A. Matsuura, and M. Paterson. A family of NFAs which need $2n - \alpha$ deterministic states. In M. Nielsen and B. Rovan, editors, *MFCS*, volume 1893 of *LNCS*, pages 436–445, London, UK, UK, 2000. Springer-Verlag.

[Jir01] G. Jirásková. Note on minimal finite automata. In *MFCS*, volume 2136 of *LNCS*, pages 421–431. Springer, 2001.

[Jir08] G. Jirásková. On the state complexity of complements, stars, and reversals of regular languages. In M. Ito and M. Toyama, editors, *DLT*, volume 5257 of *LNCS*, pages 431–442. Springer, 2008.

[Jir11a] G. Jirásková. Concatenation of regular languages and descriptional complexity. *Theory of Computing Systems*, 49(2):306–318, 2011.

[Jir11b] G. Jirásková. Magic numbers and ternary alphabet. *International Journal of Foundations of Computer Science*, 22(2):331–344, 2011.

[JJS08] J. Jirásek, G. Jirásková, and A. Szabari. Deterministic blow-ups of minimal nondeterministic finite automata over a fixed alphabet. *International Journal of Foundations of Computer Science*, 19(3):617–631, 2008.

[JK10] G. Jirásková and M. Krausová. Complexity in prefix-free regular languages. In I. McQuillan and G. Pighizzini, editors, *DCFS*, volume 31 of *EPTCS*, pages 197–204, 2010.

[JO08] G. Jirásková and A. Okhotin. State complexity of cyclic shift. *Theoretical Informatics and Applications*, 42(2):335–360, 2008.

[JPŠ14] G. Jirásková, M. Palmovský, and J. Šebej. Kleene closure on regular and prefix-free languages. In M. Holzer and M. Kutrib, editors, *CIAA*, volume 8587 of *LNCS*, pages 226–237. Springer, 2014.

[JŠ11] G. Jirásková and J. Šebej. Note on reversal of binary regular languages. In M. Holzer, M. Kutrib, and G. Pighizzini, editors, *DCFS*, volume 6808 of *LNCS*, pages 212–221. Springer, 2011.

[KRS09] J.-Y. Kao, N. Rampersad, and J. Shallit. On NFAs where all states are final, initial, or both. *Theoretical Computer Science*, 410(47–49):5010 – 5021, 2009.

[Lei81] E. Leiss. Succinct representation of regular languages by boolean automata. *Theoretical Computer Science*, 13(3):323 – 330, 1981.

[Lup63] O. B. Lupanov. A comparison of two types of finite automata. *Problemy Kibernetiki*, 9:321–326, 1963. (in Russian).

[Mas70] A. N. Maslov. Estimates of the number of states of finite automata. *Doklady Akademii Nauk SSSR*, 194:1266–1268, 1970. English translation in Soviet Mathematics Doklady 11 (1970) 1373–1375.

[MF71] A. R. Meyer and M. J. Fischer. Economy of description by automata, grammars, and formal systems. In *SWAT*, pages 188–191, Washington, DC, USA, 1971. IEEE Computer Society.

[Mir66] B. G. Mirkin. On dual automata. *Kibernetica*, 2(1):7–10, 1966. English translation in Cybernetics 2, 6–9 (1970).

[Moo71] F. R. Moore. On the bounds for state-set size in the proofs of equivalence between deterministic, nondeterministic, and two-way finite automata. *IEEE Transactions on Computers*, 20(10):1211–1214, 1971.

[MS08] A. Matsuura and Y. Saito. Equivalent transformation of minimal finite automata over a two-letter alphabet. In C. Câmpeanu and G. Pighizzini, editors, *DCFS*, pages 224–232. University of Prince Edward Island, 2008.

[Nic99] C. Nicaud. Average state complexity of operations on unary automata. In M. Kutylowski, L. Pacholski, and T. Wierzbicki, editors, *MFCS*, volume 1672 of *LNCS*, pages 231–240. Springer, 1999.

[PS02] G. Pighizzini and J. Shallit. Unary language operation, state complexity and Jacobsthal's function. *International Journal of Foundations of Computer Science*, 13(01):145–159, 2002.

[Ram06] N. Rampersad. The state complexity of L^2 and L^k. *Information Processing Letters*, 98(6):231–234, 2006.

[RI89] B. Ravikumar and O. H. Ibarra. Relating the type of ambiguity of finite automata to the succinctness of their representation. *SIAM Journal on Computing*, 18(6):1263–1282, dec 1989.

[Rob56] J. B. Roberts. Note on linear forms. *Proceedings of the American Mathematical Society*, 7(3):465–469, 1956.

[RS59] M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM Journal of Research and Development*, 3(2):114–125, apr 1959.

[Šeb13] J. Šebej. Reversal on regular languages and descriptional complexity. In H. Jürgensen and R. Reis, editors, *DCFS*, volume 8031 of *LNCS*, pages 265–276. Springer, 2013.

[Sip06] M. Sipser. *Introduction to the Theory of Computation*. Course Technology, second edition, 2006.

[vZ05] L. van Zijl. Magic numbers for symmetric difference NFAs. *International Journal of Foundations of Computer Science*, 16(5):1027–1038, 2005.

[Yer62] Y. L. Yershov. On a conjecture of V. A. Uspenskii. *Algebra i logika (seminar)*, 1:45–48, 1962. (in Russian).

[YZS94] S. Yu, Q. Zhuang, and K. Salomaa. The state complexities of some basic operations on regular languages. *Theoretical Computer Science*, 125(2):315 – 328, 1994.