Modifying the upper Bound on the Length of Minimal Synchronizing Word

A.N. Trahtman*

Bar-Ilan University, Dep. of Math., 52900, Ramat Gan, Israel

Lecture Notes in Computer Science, 6914, 173-180, 2011

Abstract. A word w is called synchronizing (recurrent, reset, magic, directable) word of deterministic finite automaton (DFA) if w sends all states of the automaton to a unique state. In 1964 Jan Černy found a sequence of n-state complete DFA possessing a minimal synchronizing word of length $(n-1)^2$. He conjectured that it is an upper bound on the length of such words for complete DFA. Nevertheless, the best upper bound $(n^3-n)/6$ was found almost 30 years ago.
We reduce the upper bound on the length of the minimal synchronizing word to $n(7n^2+6n-16)/48$.
An implemented algorithm for finding synchronizing word with restricted upper bound is described. The work presents the distribution of all synchronizing automata of small size according to the length of an almost minimal synchronizing word.

Keywords: deterministic finite automaton, synchronizing word, Černy conjecture.

Introduction

The problem of synchronization of DFA is natural and various aspects of this problem were touched upon the literature. Synchronization makes the behavior of an automaton resistant against input errors since, after detection of an error, a synchronizing word can reset the automaton back to its original state, as if no error had occurred. Therefore different problems of synchronization draw the attention.

A problem with a long story is the estimation of the minimal length of synchronizing word. In 1964 Jan Černy found [3] n-state complete DFA with shortest synchronizing word of length $(n-1)^2$ for alphabet size $q=2$. He conjectured that it is an upper bound on the length of the shortest synchronizing word for any n-state complete DFA. Best known now as a Černy's conjecture, it was raised independently not once.

The problem encourages a lot of investigations and generalizations [2] and together with Road Coloring problem [10], [14] was considered as a most fascinating old problem in finite automata theory.

* Email: trakht@macs.biu.ac.il

The conjecture holds true for a lot of automata, but in general the problem still remains open in spite the fact that over hundred papers consider this problem from different points of view. Moreover, two conferences ("Workshop on Synchronizing Automata" (Turku, 2004) and "Around the Černy conjecture" (Wroclaw,2008) were dedicated to this longstanding conjecture. The problem was discussed in "Wikipedia" - the popular Internet Encyclopedia and on some other sites. // The problem can be reduced to automata with strongly connected graph [3]. The best known upper bound is now equal to $\frac{n^3-n}{6}$ [5], [9], [10]. This estimation was not improved almost 30 years.

We reduce the upper bound on the length of a minimal reset word. This length of n-state strongly connected automaton (and also for not necessary strongly connected) is not greater than

$$\frac{n(7n^2+6n-16)}{48}$$

The crucial estimation makes here the value $7n^3/48$. So the obtained result improves known upper bound $\frac{n^3-n}{6}$. A modification of the old bound makes here the coefficient $\frac{7}{8}$.

The search is essentially based on lemmas from [5] and [9]. The same lemmas were used in a polynomial time algorithm described below for finding synchronizing word. The algorithm is implemented in the package TESTAS [15]. The time complexity of the algorithm is $O(n^3)$ and the space complexity is quadratic. An important feature of the algorithm is that the length of the obtained synchronizing word is restricted by some given upper bound. We propose a modification of the algorithm that reduces this bound to the above-mentioned value of $n(7n^2 + 6n - 16)/48$.

There are no examples of automata such that the length of the shortest synchronizing word is greater than $(n-1)^2$. Moreover, the examples of automata with synchronizing word of length $(n-1)^2$ are infrequent. After the sequence found by Černy and the example of Černy, Piricka and Rosenauerova [4] of 1971 for alphabet size $q = 2$, a next example was found by Kari [7] only in 2001 for $n = 6$ and $q = 2$. Roman [11] had found an analogous example for $n = 5$ and $q = 3$ in 2004.

The package TESTAS has studied all automata with strongly connected transition graph of size $n \leq 10$ for $q = 2$, of size $n \leq 8$ for $q \leq 3$ and of size $n \leq 7$ for $q \leq 4$ [15]. Our work presents the distribution of all considered synchronizing automata of small size according to the length of an almost minimal synchronizing word.

Five new examples of DFA with shortest synchronizing word of length $(n-1)^2$ from this class of automata were found. The size of the alphabet of these and all presently known examples is two or three.

Preliminaries

We consider a complete n-state DFA with state transition graph Γ and transition semigroup S over alphabet Σ ($|\Sigma| = q$). Let us exclude the trivial cases $n \leq 2$ and $q = 1$.

The states of the automaton are considered also as vertices of the transition graph Γ and let $|\Gamma| = n$ be the number of states.

If there exists a path in an automaton from the state $\mathbf{p}$ to the state $\mathbf{q}$ and the edges of the path are consecutively labelled by $\sigma_1, ..., \sigma_k$, then for $s = \sigma_1 ... \sigma_k \in \Sigma^+$ let us write $\mathbf{q} = \mathbf{p}s$.

Let Ps be the set of states $\mathbf{q} = \mathbf{p}s$ for all $\mathbf{p}$ from the subset P of states and $s \in \Sigma^+$. For the transition graph Γ of an automaton let Γs denote the set Ps for the set P of all states of the automaton.

A word $v \in \Sigma^+$ is called *synchronizing word* of an automaton A with transition graph Γ if $|\Gamma v| = 1$. An automaton (and its transition graph) possessing a synchronizing word is called *synchronizing*.

A state $\mathbf{p}$ will be called *empty* state by mapping of the word s if $\mathbf{p} \in \Gamma \setminus \Gamma s$.

The direct product Γ^2 of two copies of the transition graph Γ over an alphabet Σ consists of pairs $(\mathbf{p}, \mathbf{q})$ and edges $(\mathbf{p}, \mathbf{q}) \to (\mathbf{p}\sigma, \mathbf{q}\sigma)$ labelled by σ. Here $\mathbf{p}, \mathbf{q} \in \Gamma$, $\sigma \in \Sigma$ [13].

1 A state outside the image

Lemma 1 *Suppose $\mathbf{p}_i \notin \Gamma s$. Then $\mathbf{p}_i \notin \Gamma us$ for any word u.*

Proof follows from $\Gamma u \subseteq \Gamma$.

Lemma 2 *Let Γ be transition graph of a DFA. If there are words s and t such that $\mathbf{p}s \notin \Gamma ts$ for some $\mathbf{p}$ from $\Gamma \setminus \Gamma t$ then Γts is a proper subset of Γs.*

Proof. One has $\Gamma s = ((\Gamma \setminus \Gamma t) \cup \Gamma t)s = (\Gamma \setminus \Gamma t)s \cup \Gamma ts$. The state $\mathbf{p}s$ from Γs is outside Γts and also outside $(\Gamma \setminus \Gamma t)s$. Now from $\Gamma ts \subseteq \Gamma s$ follows $\Gamma ts \subset \Gamma s$.

Lemma 3 *Let Γ be a transition graph of a synchronizing strongly connected n-state DFA. Then for any state $\mathbf{q}$ there exists a word t of length not greater than n such that $\mathbf{q} \notin \Gamma t$. For any $k < n$ there are at least k states $\mathbf{q}_k$ and a words u_k of length not greater than k such that $\mathbf{q}_k \notin \Gamma u_k$.*

Proof. The automaton is synchronizing, whence for some letter β, $\Gamma\beta \subset \Gamma$ and at least one state is empty by mapping β. The set $\Gamma \setminus \Gamma s$ is the set of empty states by mapping of the word s. Let R_k be a union of all $\Gamma \setminus \Gamma t$ for all words t such that $|t| \leq k$. Obviously that $R_k \subseteq R_m$ for $k \leq m$. From $\Gamma\beta \subset \Gamma$ follows that R_1 is non-empty.

For complement C_k of the set R_k we have $C_k = \cap \Gamma s$ for all words s of length not greater than k.

The graph Γ is strongly connected. Therefore for non-empty complement C_k of R_k there exists a letter γ such that $C_k\gamma \not\subset C_k$, whence $C_k \setminus C_k\gamma$ is not empty. Suppose $\mathbf{r} \in C_k \setminus C_k\gamma$.

$C_{k+1} = \cap \Gamma s$ for all words s of length not greater than $k+1$ and $\mathbf{r} \notin \mathbf{C_k}\gamma$. Therefore $\mathbf{r} \notin C_{k+1}$. Thus $\mathbf{r} \in R_{k+1}$, whence $R_k \subset R_{k+1}$ and $|R_k| < |R_{k+1}|$.

Consequently, for any $k \leq n$ there exists a state $\mathbf{q}$ and a word u of length not greater than k such that $\mathbf{q} \notin \Gamma u$. One has $|R_k| \geq k$, whence for given k there are at least k such states $\mathbf{q}$.

Lemma 4 *Let Γ be a transition graph of synchronizing strongly connected n-state DFA. Then for every $k \leq \frac{n+1}{2}$ there exists a word s such that $|s| \leq k^2$ and $|\Gamma s| \leq n - k$.*

Proof. If $|\Gamma s| > \frac{n+1}{2}$ for a word s then there is at least one state $\mathbf{p} \in \Gamma s$ having only one preimage $\mathbf{q}$ by mapping s. In opposite case every state $\mathbf{p} \in \Gamma s$ has at least two preimages by mapping s, whence $|\Gamma s| \leq \frac{n}{2}$.

Let us consider for a word s_i the states from Γs_i having only single preimage by mapping s_i and let Q_i be the set of such single preimages.

Our aim is now to find a short word s such that $|\Gamma s| \leq \frac{n+1}{2}$. We construct a sequence of mappings s_i that reduce the size of the set Q_i and the size of Γs_i on every step i.

By Lemma 3 for every state $\mathbf{q}$ there exists a word t_q such that $\mathbf{q} \notin \Gamma t_q$ and for $k \leq n$ there are at least k states $\mathbf{q}$ with t_q of length not greater than k.

There exists a letter α such that $|\Gamma\alpha| < |\Gamma|$. Let α be the word $s_1 = t_1$. Let Q_1 be the set of single preimages of Γt_1. Then $|\Gamma t_1| < n$ and $|Q_1| \leq n - 2$. If $n - |\Gamma\alpha| = m$ then $m < n - |Q_1| \leq 2m$.

On every next step, let us take the state $\mathbf{q}$ from Q_{i-1} with t_q of minimal length. Suppose $t_{i-1} = t_q$ and $s_i = t_{i-1}s_{i-1}$. By Lemma 2, $\Gamma s_i \subset \Gamma s_{i-1}$ and so $|\Gamma s_i| < |\Gamma s_{i-1}|$.

For mapping of defect j the size of corresponding Q_j is at least $n - 2j$. In virtue of Lemma 3, the length of minimal t_q of $\mathbf{q}$ from Q_j of size r is not greater than $n - r + 1$. So for the defect j the length of minimal t_q for Q_j has upper bound $n - (n - 2j) + 1 = 2j + 1$. This upper bound depends only on the defect j of the mapping. The sequence of possible upper bounds grows together with the defect and consists of consecutive odd integers (sometimes with gaps).

The process continues until the set Q_i is not empty (in particular, if $|\Gamma s_i| > n/2$). Thus the length of s_k is restricted by the sum of k (or less) odd integers for every $k \leq \frac{n+1}{2}$.

Consequently, for some word s and $k \leq \frac{n+1}{2}$ the length of s in view of Lemma 3 is restricted by the sum $\sum_{i=1}^{k}(2i - 1) = k^2$ and $|\Gamma s| \leq n - k$.

2 Pairs of states

The next our step is based on the following result of Frankl and Klyachko et al.

Theorem 1 *[5], [9] Let N be set of size n with subset D of size $i > 1$. Then there exists a word s of length at most*

$$C^2_{n-i+2} = (n - i + 2) * (n - i + 1)/2$$

such that $|Ds| < |D|$.

The next lemma follows the ideas from [9].

Lemma 5 *Let Γ be a transition graph of a strongly connected n-state automaton and let D_k of size k be a subset of states of the automaton.*

Then the word of length at most $C^3_{n+1} - C^3_{n-k+2}$ synchronizes the set D_k.

Proof. By Theorem 1, every D_i has a pair of states with a minimal synchronizing word of length not greater than $C^2_{n-i+2} = (n-i+2)(n-i+1)/2$. So D_k has synchronizing word of length at most $S = \sum_{i=2}^{k} C^2_{n-i+2}$.

Suppose $j = n-i+2$. Then $n \geq j \geq n-k+2$. Now $S = \sum_{j=n-k+2}^{n} C^2_j = \sum_{j=2}^{n} C^2_j - \sum_{j=2}^{n-k+1} C^2_j$.

For every $m > 2$, $\sum_{j=2}^{m} C^2_j = C^3_{m+1}$. So $\sum_{j=2}^{n} C^2_j = C^3_{n+1}$ and $\sum_{j=2}^{n-k+1} C^2_j = C^3_{n-k+2}$. Therefore $S = C^3_{n+1} - C^3_{n-k+2}$.

Theorem 2 *Let Γ be a transition graph of a strongly connected n-state automaton. Then a word of length not greater than*

$$\frac{n(7n^2+6n-16)}{48}$$

synchronizes the automaton.

Proof. Let us combine the quadratic estimation from Lemma 4 and the cubic estimation of Lemma 5. The length of some synchronizing word is not greater than the sum of $S_1 = C^3_{n+1} - C^3_{n-k+2}$ (Lemma 5) and $S_2 = k^2$ (Lemma 4) for $k \leq \frac{n+1}{2}$.

We must consider $k \leq \frac{n+1}{2}$. Hence the maximum of $S = S_1 + S_2$ exists for even n and $k = \frac{n}{2}$ (the case of odd n and $k = \frac{n+1}{2}$ also will be calculated for clarity). In the case of even n

$S_1 = C^3_{n+1} - C^3_{n-k+2} = C^3_{n+1} - C^3_{n/2+2} = \frac{n^3-n}{6} - \frac{(n/2+1)^3-n/2-1}{6} = \frac{8n^3-8n-n^3-6n^2-12n-8+4n+8}{48} = \frac{n(7n^2-6n-16)}{48}$.

$S_2 = k^2 = \frac{n^2}{4}$.

So the length of a minimal synchronizing word has in the case of even n the following upper bound $S_1 + S_2 = \frac{n(7n^2-6n-16)}{48} + \frac{n^2}{4} = \frac{n(7n^2+6n-16)}{48}$.

In the case of odd n

$S_1 = C^3_{n+1} - C^3_{n-k+2} = C^3_{n+1} - C^3_{(n+1)/2+2} = \frac{n^3-n}{6} - \frac{((n+1)/2+1)^3-(n+1)/2-1}{6} = \frac{8n^3-8n-n^3-9n^2-27n-27+4n+12}{48} = \frac{7n^3-9n^2-31n-15}{48}$.

$S_2 = k^2 = \frac{n^2+2n+1}{4}$.

So the length of a minimal synchronizing word has in the case of odd n the following upper bound $S_1 + S_2 = \frac{7n^3-9n^2-31n-15}{48} + \frac{n^2+2n+1}{4} = \frac{7n^3+3n^2-7n-3}{48}$. This value is less than $\frac{n(7n^2+6n-16)}{48}$ for $n > 2$.

Thus the value $\frac{n(7n^2+6n-16)}{48}$ is an upper bound on the length of the minimal synchronizing word. The obtained result improves the old upper bound $\frac{n(n^2-1)}{6}$ by factor $\frac{7}{8}$.

Remark 1 *For odd n a word of length not greater than $\frac{7n^3+3n^2-7n-3}{48}$ synchronizes the automaton.*

2.1 An algorithm for finding synchronizing word of restricted length

The algorithm presents another useful application of the combinatorial ideas from [9]. The Theorem 1 gives us an estimation of the length of the reset word.

Let us consider the inverse of the graph Γ^2. So the incoming edges of every pair $(\mathbf{p}, \mathbf{q})$ from Γ^2 together with its ancestors are known. Then let us enumerate the pairs of vertices. There are vertices $\mathbf{p}, \mathbf{q}$ from Γ such that for some letter α $\mathbf{p}\alpha = \mathbf{q}\alpha$. For every such pair $(\mathbf{p}, \mathbf{q})$ from Γ^2 suppose $n(\mathbf{p}, \mathbf{q}) = 1$. Let us connect with the pair $(\mathbf{p}, \mathbf{q})$ the letter α.

Then for every enumerated pair $(\mathbf{p}, \mathbf{q})$ from Γ^2 with $n(\mathbf{p}, \mathbf{q}) = k$ let us consider all its ancestors without enumeration. These pairs obtain the number $k + 1$ and are connected with the letter on the edge of the graph Γ^2 from this pair to the pair $(\mathbf{p}, \mathbf{q})$.

We find a sequence of mappings of the graph of the automaton induced by the letters on the labels. Let us consider the graph Γs for some word s and find a pair $(\mathbf{p}, \mathbf{q})$ from Γs with a minimal number. The letter α of the pair is the first letter of the word w we build. The next letter of the word w is the letter of the pair $\mathbf{p}\alpha, \mathbf{q}\alpha$. The number of this pair is less than the number of $(\mathbf{p}, \mathbf{q})$. We proceed on this way until the number of the pair exists. The last pair is synchronizing by a letter. The obtained word w synchronizes the vertices $\mathbf{p}$ and $\mathbf{q}$. The length of the word w is at most $\frac{(n-|\Gamma s|+2)(n-|\Gamma s|+1)}{2}$ (theorem 1).

The search of the first letter of the word w needs $O(|\Gamma|(|\Gamma| - 1)/2)$ steps. Then the building of the word w needs $|\Gamma|$ steps. The number of the words w is less than n. Therefore the time complexity of considered procedure can be estimated by $O(|\Gamma|^3)$ in the worst case. The space complexity of the algorithm is $O(|\Gamma|^2)$ because of the size of Γ^2. The algorithm is correct in view of the Lemma 5 and is implemented in the package TESTAS.

2.2 A modification of the algorithm

The modification is based on Lemmas 3 and 4. There exists a letter α and a state $\mathbf{p}$ such that $\mathbf{p} \notin \Gamma\alpha$. Let R_1 be the set of such states $\mathbf{p}$ (as in Lemma 3). We associate the word $u_1 = \alpha$ of length one with every state.

Suppose the set R_k of states and its non-empty complement C_k with corresponding words exist. From the proof of Lemma 3 follows that for some letter β there exists a state $\mathbf{q}$ in $C_k \setminus C_k\beta$. For every state $\mathbf{r}$ from C_k with corresponding word u we associate the word $u_{k+1} = u\beta$ and add $\mathbf{q}$ to R_{k+1}.

Let us keep with every state $\mathbf{p}s \in \Gamma s$ its preimage by mapping s and fix the case of more than one preimage. If $\mathbf{p}s_i$ has only one preimage by mapping s_i then suppose $\mathbf{p} \in Q_i$.

All states of the graph belong to Q_0, after the mapping v_1 we have $|Q_1| \leq n - 2$. The set Q_i lost states on every step. If $|\Gamma s| > |\Gamma|/2)$ then there exists a state in Γs having only one preimage and so Q_i is not empty by mapping s. We proceed until Q_i is not empty.

Let us choose a state $\mathbf{p} \in Q_i$ with word u_k of minimal length and suppose $v_{i+1} = u_k v_i$. The length of the word v_i is restricted according to Lemma 3. We continue until Γv_i has states with only one preimage (Q_i is not empty). So we obtain the set of states Γv_i of size less than $(n + 1)/2$ (Lemma 4) and then proceed by the main algorithm.

The upper bound on the length of the synchronizing word in virtue of Theorem 2 is $\frac{n(7n^2+6n-16)}{48}$.

3 Distribution of the length of synchronizing word of small automata

A program based on the synchronization algorithms of the package TESTAS was used for a search of automata with a minimal reset word of relatively great length. The program has investigated all complete DFA for $n \leq 10$ over an alphabet of size 2, $n \leq 8$ over an alphabet of size 3 and for $n \leq 7$ over an alphabet of size 4 [14].

Maximal value of the length of a synchronizing word for $n = 10$ found by the algorithm on the set of considered automata of size n is 93. The length found by the minimal length algorithm is 81 ($Err < 0.13$). So the shift of the size of the synchronizing word is relatively small.

The program consistently sifts non-synchronizing automata, the automata with a very short reset word and a part of isomorphic automata. The following table presents the distribution of all remaining automata of size 10 over an alphabet of two letters (see also [1]).

interval of size of the automata	n - 2n	2n - 3n	3n - 4n	4n - 5n	5n - 6n	6n -7n
percent of automata in interval	81.01	16.2	1.82	0.8	0.05	0.006

The distribution for three and four letters does not differ noticeable and is omitted.

The synchronizing words of minimal length are found only for automata having great minimal reset words. The maximal number of considered n-state automata has its length of the reset word near n+1.

Thus one can conclude that the polynomial synchronizing algorithms of the package find synchronizing words of a length not far of the minimal, especially for automata with very great reset words. The presented distribution does not differ essentially from the distribution of the lengths of the minimal synchronizing words.

4 Acknowledgments

I would like to express my gratitude to the referees for many helpful and useful remarks and for improving the calculation of the result.

References

1. Ananichev D., Gusev V., Volkov M.: Slowly Synchronizing Automata and Digraphs. Springer, Lect. Notes in Comp. Sci., 6281, MFCS, 55-65, (2010).
2. Béal M.P., Czeizler E., Kari J., Perrin D.: Unambiguous automata. Math. Comput. Sci., 1, 625638, (2008).

3. Černy J.: Poznamka k homogenym eksperimentom s konechnymi automatami, Math.-Fyz. Čas., 14, 208-215, (1964).
4. Černy J., Piricka A., Rosenauerova B.: On directable automata. Kybernetika, 7, 289-298, (1971).
5. Frankl P.: An extremal problem for two families of sets. Eur. J. Comb., 3, 125-127, (1982).
6. Friedman J.: On the road coloring problem. Proc. of the Amer. Math. Soc., 110, 1133-1135, (1990).
7. Kari J.: A counter example to a conjecture concerning synchronizing word in finite automata. EATCS Bulletin, 73, 146-147, (2001).
8. Kari, J.: Synchronizing finite automata on Eulerian digraphs. Springer, Lect. Notes in Comp. Sci., 2136, 432-438, (2001).
9. Kljachko A.A., Rystsov I.K., Spivak M.A.: An extremely combinatorial problem connected with the bound on the length of a recurrent word in an automata. Kybernetika, 216-225, (1987).
10. Pin J.-E.: On two combinatorial problems arising from automata theory. Annals of Discrete Math., 17, 535-548, (1983).
11. Roman A. Experiments on Synchronizing Automata. Schedae Informaticae, Versita, Warsaw, V. 19, 35-51, (2010).
12. Steinberg B.: The Averaging Trick and the Cerny Conjecture. Developments in Language Theory, Springer, NY, Springer, Lect. Notes in Comp. Sci., 6224, 423-431, (2010).
13. Trahtman A.N.: The Cerny Conjecture for Aperiodic Automata. Discr. Math. and Theoret. Comput. Sci., 9, 2(2007), 3-10.
14. Trahtman A.N: The Road Coloring and Černy Conjecture. Proc. of Prague Stringology Conference, 1-12 (2008).
15. Trahtman A.N.: Notable trends concerning the synchronization of graphs and automata. CTW06, El. Notes in Discrete Math., 25, 173-175, (2006).