Theoretical Computer Science 237 (2000) 485–494

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

Tight bounds on the number of states of DFAs that are equivalent to n-state NFAs ☆

Kazuo Iwama[a,*,1], Yahiko Kambayashi[a,2], Kazuya Takaki[b,3]

[a] *School of Informatics, Kyoto University, Kyoto 606-8501, Japan*

[b] *NTT Long-Distance Communications Sector, Nippon Telegraph And Telephone Corporation, Tokyo 150, Japan*

Received December 1997; revised October 1999

Communicated by A. Salomaa

Abstract

It is shown that if α is an integer which can be expressed as 2^k or $2^k + 1$ for some integer $0 \leqslant k \leqslant n/2 - 2$, then there exist nondeterministic finite automata with n states whose equivalent deterministic finite automata need exactly $2^n - \alpha$ states. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

After students start studying automata theory, they soon understand that nondeterministic automata are more efficient than deterministic ones. In the standard textbooks (e.g., [3–5, 9]), this important fact is first demonstrated using finite automata: Namely, given a nondeterministic finite automaton (NFA) M of n states, one needs up to 2^n states to construct a deterministic finite automaton (DFA) which is equivalent to M. Thus it appears that we need much more states to simulate NFAs by DFAs. Note that, however, this shows only an upper bound. To be more precise, let $\Delta(M, n)$ be the number of states that is necessary and sufficient to simulate the NFA M of n states by

☆ A preliminary version of this paper was presented at the 3rd Conference on Developments in Language Theory, Thessaloniki, July 1997.

* Corresponding author.

E-mail addresses: iwama@kuis.kyoto-u.ac.jp (K. Iwama), yahiko@kuis.kyoto-u.ac.jp (Y. Kambayashi), k.takaki@esc.longdist.ntt.co.jp (K. Takaki)

[1] Supported in part by Scientific Research Grant, Ministry of Education, Japan and Mitsubishi Foundation.

[2] Supported in part by Scientific Research Grant, Ministry of Education.

[3] This work was mainly done while this author was with Kyushu University, Fukuoka, Japan.

PII: S0304-3975(00)00029-3

some DFA. Then the above fact says that $\Delta(M,n) \leqslant 2^n$ for any NFA M, which is one of the oldest theorems in automata theory [8].

It was not so old that this bound was shown to be tight by Moore [6], i.e., there exists an NFA M such that $\Delta(M,n) = 2^n$. It is a little surprising that this result does not seem to be common; as far as the authors know this result is not included in any standard textbooks. (As a rare exception, [2] suggests, as one of chapter-end exercises, that an NFA M exists such that $\Delta(M,n) = 2^{n-1}$ without citing [6].) Even more surprising is that the research on $\Delta(M,n)$ completely stopped there; the literature does not answer any basic questions like whether there is an NFA M such that $\Delta(M,n) = 2^n - k$. Clearly, the most general and interesting question is whether there always exists an NFA M of n_1 states such that $\Delta(M,n_1) = n_2$ for any integers n_1 and n_2 satisfying that $n_1 \leqslant n_2 \leqslant 2^{n_1}$.

In this paper, we cannot give answers to this final question, but we show that if the integer n_2 can be expressed as $2^{n_1} - 2^k$ or $2^{n_1} - 2^k - 1$ for some integer $k \leqslant n_1/2 - 2$, then there is an NFA M of n_1 states such that $\Delta(M,n_1) = n_2$. An immediate corollary is that there are NFA's M of n states such that $\Delta(M,n) = 2^n - 1,\ 2^n - 2,\ 2^n - 3,\ 2^n - 4,\ 2^n - 5,\ 2^n - 8,\ 2^n - 9, \ldots$. Thus the first unsettled number is $2^n - 6$, or it is not known at this moment if there is an NFA M such that $\Delta(M,n) = 2^n - 6$ (although our strong conjecture is that there does exist one).

Note that finite automata in this paper are always one-way and use the binary input symbols 0 and 1. If we allow three or more input symbols, then the above question becomes easier, i.e., it is easier to find NFAs whose deterministic counterparts need a specific number of states. If we extend our attention to two-way and/or probabilistic finite automata, several other results on the number of states exist. Recently, for example, Ambainis shows in [1] that there exist probabilistic finite automata with an isolated cutpoint that need $\Omega(2^{n(\log\log n)/(\log n)})$ deterministic states. Berman and Lingas [2] show that there is a two-way NFA of $O(n)$ states that needs $\Omega(2^{n^2})$ deterministic (one-way) states. In [7] Nozaki investigates the minimum length of input strings to decide whether two NFAs are equivalent or not, which implies the result of Moore [6] as a corollary.

2. Preliminaries

A finite automaton M is determined by giving the following five items: (i) A finite set K of states, $S_0, S_1, \ldots, S_{n-1}$, (ii) A finite set Σ of input symbols, which is always $\{0,1\}$ in this paper. (iii) An initial state ($\in K$), which is always S_0 in this paper. (iv) A set F of accepting states ($\subseteq K$). (v) A state transition function δ. If δ is a mapping from $K \times \Sigma$ into K, then M is said to be *deterministic*. If δ is a mapping from $K \times \Sigma$ into 2^k, then M is said to be *nondeterministic*. The domain of δ is naturally extended from $K \times \Sigma$ into $K \times \Sigma^*$. The definition of the language accepted by M is as usual and may be omitted. If two finite automata M_1 and M_2 accept the same language, then M_1 and M_2 are said to be *equivalent*.

When we discuss the number of states of a DFA M, M must be a minimal DFA, i.e., it must be guaranteed that there is no other DFA M' that is equivalent to M and has fewer states than M. It is a fundamental fact [8] that a DFA M is minimal if (i) all states can be reachable from the initial state and (ii) there are no two equivalent states. Here, two states Q_1 and Q_2 are said to be *equivalent* if for all $x \in \Sigma^*$, $\delta(Q_1, x) \in F$ iff $\delta(Q_2, x) \in F$. For an NFA M of n states, $\Delta(M, n)$ denotes the number of states of a minimal DFA M' that is equivalent to M. NFAs should also be minimal. However, within this paper, we only consider NFAs whose $\Delta(M, n)$ value is large. So, it is not necessary to give explicit proofs for the minimality of NFAs because of the following fact:

Proposition 1. *If* $\Delta(M, n) > 2^{n-1}$, *then the NFA* M *is minimal.*

Proof. Obvious since $\Delta(M, n-1) \leqslant 2^{n-1}$ for any NFA M of $n-1$ states. □

Let M_1 be an NFA of n states $K_1 = \{S_0, S_1, \ldots, S_{n-1}\}$. Then one can construct an equivalent DFA M_2 as follows: We first introduce all the 2^n subsets of K_1, each of which can be a state of M_2. Thus a state of the DFA M_2 corresponds to a family of states of the NFA M_1. To avoid confusion, a state of M_2 is often called an *F-state*. If an F-state X consists of k (M_1's) states, then it is said that the *size* of X is k and also denoted by $|X| = k$. The initial F-state of M_2 is $\{S_0\}$ if the initial state of M_1 is S_0. An F-state $X \subseteq K_1$ of M_2 is an accepting state if X includes at least one accepting state of M_1. The transition function δ_2 of M_2 is defined using the transition function δ_1 of M_1 as follows: For F-states Q_1 and $Q_2 \subseteq K_1$, $\delta_2(Q_1, a) \equiv Q_2$ ($a \in \{0, 1\}$) if $\bigcup_{s \in Q_1} \delta_1(s, a) = Q_2$. After determining this δ_2, we remove all F-states which cannot be reached from the initial F-state $\{S_0\}$. Note that this DFA may still not be minimal since some two states might be equivalent. The whole procedure is usually called the "subset construction" [8].

3. Main results

Our main results are the following two theorems. Proofs are very similar for both theorems, so only the difference will be briefly given for the second one.

Theorem 1. *There is an NFA* M *of* n *states such that* $\Delta(M, n) = 2^n - 2^k - 1$ *for any integers* n *and* k *satisfying that* $0 \leqslant k \leqslant n/2 - 2$.

Theorem 2. *There is an NFA* M *of* n *states such that* $\Delta(M, n) = 2^n - 2^k$ *for any integers* n *and* k *satisfying that* $0 \leqslant k \leqslant n/2 - 2$.

3.1. Proof of Theorem 1

For simpler exposition, we first prove the theorem for $k = 2$ and $n \geqslant 10$. Let M be the NFA of n states whose transition function is given in Fig. 1. Its initial state is S_0

current state	next states	
	0	1
S_0	S_1	S_1
S_1	S_2	$S_1,\ S_2$
S_2	S_3	$S_1,\ S_3$
.	.	.
.	.	.
S_i	S_{i+1}	$S_1,\ S_{i+1}$
.	.	.
.	.	.
S_{n-7}	S_{n-6}	$S_1,\ S_{n-6}$
S_{n-6}	S_{n-5}	$S_1,\ S_{n-5}$
S_{n-5}	S_{n-4}	S_{n-2}
S_{n-4}	S_{n-3}	S_{n-1}
S_{n-3}	S_0	S_1
S_{n-2}	S_{n-1}	$S_1,\ S_{n-4}$
S_{n-1}	S_{n-2}	$S_1,\ S_{n-3}$

Fig. 1. Transition function of the NFA *M*.

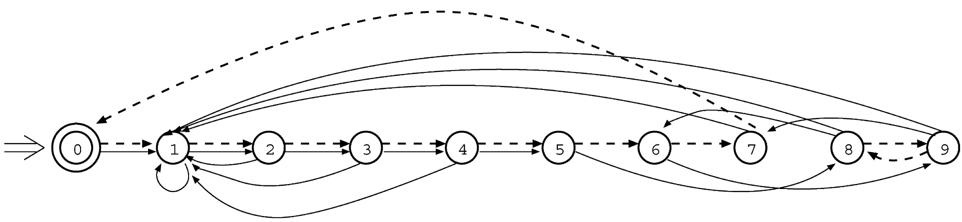

Fig. 2. State diagram of *M* for $k=2$ and $n=10$.

and its accepting states are also only S_0. Fig. 2 illustrates the state diagram of M for $k=2$ and $n=10$ where plain lines denote state transitions by symbol 1, and dotted lines by reading symbol 0. We first construct the DFA, denoted by T, by the subset construction and show the number of states in T is 2^n-5 and all of them can be reached from the initial state. After that we shall show that no two states among those 2^n-5 ones are equivalent. Before describing details, we first take a look at the basic structure of this NFA M and its deterministic counterpart T.

The state set of M is divided into two groups $A=\{S_0,\dots,S_{n-3}\}$ and $B=\{S_{n-2},S_{n-1}\}$. If M reads 0's, its state is preserved within group A or B. In group A, M's state is shifted on the cycle of $S_0\to S_1\to\cdots\to S_{n-3}\to S_0$ by reading 0's. This is the same for the DFA T obtained by the subset construction in the following sense: Let X be its F-state consisting of M's states. If T reads symbol 0, X changes to X' where each state in X is shifted one position on the above cycle. It is said that X' is obtained from X by a 0-*shift* and conversely X is obtained from X' by a 0-*inv-shift*. In group B, M's state is shifted on the cycle of $S_{n-2} \to S_{n-1} \to S_{n-2}$ by reading symbol 0.

State transitions by reading symbol 1 are also divided into two groups, *Back-transitions* (*B-transitions*) and *Forward-transitions* (*F-transitions*). B-transitions include every transition to S_1 i.e., those from $S_0, S_1, \ldots, S_{n-6}, S_{n-3}, S_{n-2}$ and S_{n-1}. F-transitions are all the other transitions. If we consider only F-transitions, then M's state is again shifted on the path $S_1 \to S_2 \to \cdots \to S_{n-5} \to S_{n-2} \to S_{n-4} \to S_{n-1} \to S_{n-3}$. Similarly as 0-shifts and 0-inv-shifts, we can consider a 1-*shift* and a 1-*inv-shift* on this path. However, it is not a cyclic shift this time; If an F-state X contains S_1, then by a 1-inv-shift, this S_1 disappears, i.e., $|X|$ decreases by one. Similarly for a 1-shift when X includes S_{n-3}. (Note that we in fact have a transition by reading 1 from S_{n-3} to S_1, but this transition was defined as a B-transition.)

Now we introduce an important definition: An F-state X is called an S_1-*pattern* if it satisfies the following three conditions: (i) $2 \leqslant |X| \leqslant n-3$ and all the (M's) states included by X are in group A. (ii) $S_0 \notin X$ and $S_1 \in X$. (iii) X includes at least one S_i such that $2 \leqslant i \leqslant n-5$.

Lemma 1. *Let X be any F-state such that $2 \leqslant |X| \leqslant n-3$ and all states in X are in group A. Then there is an S_1-pattern Y such that X can be obtained from Y by some number* (*may be zero*) *of* 0-*shifts.*

Proof. If X itself is an S_1-pattern, then we need zero 0-shift. So suppose that X is not an S_1-pattern. Since $|X| \leqslant n-3$, at least one state in group A is missing. Hence, one can change X into X_1 by some number of 0-inv-shifts such that X_1 does not include S_0 but does include S_1. Now check if X_1 is an S_1-pattern. If so, then we are done since X can be obtained from X_1 by 0-shifts. Otherwise, let $X_1 = \{S_1, S_{i_1}, S_{i_2}, \ldots\}$ where $1 \leqslant i_1 \leqslant i_2 \leqslant \cdots$. Then since X_1 is not an S_1-pattern, $i_1 \geqslant n-4$. Now apply 0-inv-shifts until this S_{i_1} changes to S_1 and let the resulting F-state be X_2. Then this X_2 does not include S_0 since S_{i_1-1} is not in X_1. Also this X_2 must include some S_i such that $2 \leqslant i \leqslant n-5$, that may be the former S_{i_2} in X_1 or the former S_1 in X_1 (recall that X_1 contains at least two states). Thus it turns out that this X_2 must be an S_1-pattern and that is what we wanted. □

Lemma 2. *Let X be any F-state such that its intersection with A, i.e., $X \cap A$, is an S_1-pattern. Then there is another F-state Y such that $|Y| = |X| - 1$ and the DFA T changes from Y to X by reading a single* 1.

Proof. Since $X \cap A$ is an S_1-pattern, X can be written as $X = \{S_1, S_{i_1}, \ldots\}$ where $2 \leqslant i_1 \leqslant n-5$. Now let Y be the F-state obtained from X by a 1-inv-shift. Y can be written as $\{S_{i_1-1}, \ldots\}$ and $|Y| = |X| - 1$. Now let Z be the F-state into which T changes from Y by reading 1. (We wish to show that $Z = X$.) Then, since the 1-inv-shift of X is Y, the 1-shift of Y is $X - \{S_1\}$, which means Z must include this $X - \{S_1\}$. Also, Z must include S_1 since there is a B-transition to S_1 from S_{i_1-1} in Y (this is the reason why we introduced the third condition for the S_1-pattern). Since the states reached by reading 1 are at most those by 1-shift and S_1 by means of B-transitions, no extra states are included in Z, i.e., $X = Z$. □

Now we are ready to show that $\Delta(M,n) = 2^n - 5$. To do so, we will first show that the DFA T has $2^n - 5$ states and then that T is minimal. It will turn out that among 2^n all subsets of $\Sigma = \{S_0, S_1, \ldots, S_{n-1}\}$, the following five subsets (five F-states) are missing in T; (i) ϕ (the empty set), (ii) $A = \{S_0, S_1, \ldots, S_{n-3}\}$, (iii) $A \cup \{S_{n-2}\}$, (iv) $A \cup \{S_{n-1}\}$ and (v) $A \cup \{S_{n-2}, S_{n-1}\}$. Let Γ be the set of those five F-states. In the following we shall use mathematical induction to show that all the F-states but those in Γ appear in the DFA T. The base of the induction is $m = 2$. So, we first consider the case that $m = 1$, then the case that $m = 2$ and then the general case, i.e., for $m \geqslant 2$.

Case 1: ($m = 1$). $\{S_0\}$ is the initial state of T. Each of $\{S_1\}$ through $\{S_{n-3}\}$ can be reached by 0-shifts from $\{S_0\}$. $\{S_{n-2}\}$ and $\{S_{n-1}\}$ are reached from $\{S_{n-5}\}$ and $\{S_{n-4}\}$ by reading 1, respectively.

Case 2: ($m = 2$). All F-states X of size two are divided into the following three groups: (2.1) Both states in X are in group A (see Case 2.1 and similarly below). (2.2) One of the two states is in group A. (2.3) None is in group A (i.e., both are in group B).

Case 2.1: X satisfies the conditions of Lemma 1. So there exists another F-state, say, Y, such that Y is an S_1-pattern and T can change from Y to X by reading 0's. Y satisfies the condition of Lemma 2. So there exists another F-state, Z, such that $|Z| = 1$ and T can change from Z to Y by reading 1. Existence of such Z is guaranteed by the argument in Case 1, and hence such an F-state X must exist in T.

Case 2.2: Let $X = \{S_i, S_j\}$ when $0 \leqslant i \leqslant n-3$ and $S_j = S_{n-1}$ or S_{n-2}. Obviously, there exists $Y = \{S_1, S_{j'}\}$ ($S_{j'} = S_{n-1}$ or S_{n-2}) such that T moves from Y to X by reading 0's. Now consider $Z = \{S_{n-3}, S_{j''}\}$ where $j'' = n - 4$ if $j' = n - 1$ and $j'' = n - 5$ if $j' = n - 2$. One can see that T moves from Z to Y by reading 1. Since $Z \subseteq A$, its existence is guaranteed by Case 2.1.

Case 2.3: $X = \{S_{n-2}, S_{n-1}\}$. Let $Z = \{S_{n-5}, S_{n-4}\}$. T moves from Z to X by reading 1. Z must exist as shown in Case 2.1.

Case 3: (For general $m \geqslant 2$). Now our induction hypothesis is that every F-state of size m ($\geqslant 2$) exists in T if it is not in Γ (recall that Γ is the set of the five F-states given before). Under this assumption we shall show any F-state, X, of size $m+1$ exists unless X is in Γ. As before, the F-states of size $m + 1$ are divided into three groups: (3.1) All states in X are in group A. (3.2) One of them is in group B. (3.3) Two of them are in B.

Case 3.1: Recall that X (of size $m + 1$) is not in Γ. Then X is different from the whole A and hence it satisfies the condition of Lemma 1. The proof is very similar to Case 2.1, i.e., we can find an F-state Z of size m from which T can change to X and whose existence is guaranteed by the induction hypothesis.

Case 3.2: X can be written as $X = X_1 \cup X_2$, where $|X_1| = m$ ($\geqslant 2$) and $X_1 \subseteq A$ and $X_2 = \{S_{n-2}\}$ or $\{S_{n-1}\}$. One can easily verify that X_1 satisfies the condition of

Lemma 1. So, we can obtain an S_1-pattern Y_1 by applying some number of 0-inv-shifts. Also Y_2 (again $\{S_{n-2}\}$ or $\{S_{n-1}\}$) is obtained from X_2 by the same number of 0-inv-shifts. Let $Y = Y_1 \cup Y_2$. Then this Y satisfies the condition of Lemma 2 and we can get an F-state Z of size m by a 1-inv-shift. Thus X can be reached from Z whose existence is guaranteed by the induction hypothesis.

Case 3.3: $X = X_1 \cup X_2$ where $|X_1| = m-1$ and $X_2 = \{S_{n-2}, S_{n-1}\}$. We need to consider further two cases.

Case 3.3.1: $m=2$. In this case $|X_1|=1$. T can change from $\{S_{n-5}, S_{n-4}, S_{n-3}\}$ to $\{S_1, S_{n-2}, S_{n-1}\}$ by reading symbol 1 and then to X by reading some number of 0's. The existence of $\{S_{n-5}, S_{n-4}, S_{n-3}\}$ is guaranteed by Case 3.1.

Case 3.3.2: $m \geqslant 3$. In this case $|X_1| \geqslant 2$. Hence we can make very similar argument as Case 3.2, which may be omitted.

Thus we have shown that any F-state $\notin \Gamma$ appears in T.

Lemma 3. *Any F-state in Γ does not appear in T.*

Proof. First of all, ϕ cannot be reached from $\{S_0\}$ since we have no next-state entry in Fig. 1 that contains ϕ. The other four F-states in Γ are $\{S_0, S_1, \ldots, S_{n-3}\}$, $\{S_0, S_1, \ldots, S_{n-3}, S_{n-2}\}$, $\{S_0, S_1, \ldots, S_{n-3}, S_{n-1}\}$ and $\{S_0, S_1, \ldots, S_{n-3}, S_{n-2}, S_{n-1}\}$. Now one can see that if T could reach one of those state from $\{S_0\}$, then there must be an F-state X such that X is different from any of those four states and T can move from X to one of the four states, say, Y, by reading symbol 0 or 1.

Now we shall show that such X does not exist: (i) If T could move from X to Y, then the symbol read by T is not 1. (The reason: Y contains S_0 but S_0 is not included in the column for symbol 1 of Fig. 1.) (ii) So, the symbol read by T must be 0. Let $X = X_1 \cup X_2$ where $X_1 = X \cap A$. Then since $X \notin \Gamma$, $X_1 \neq A$. Recall that a state transition by symbol 0 is a "cyclic shift", so by reading 0, X_1 is shifted to some X_1' that must not coincide A again. Hence the next state of X by reading 0 must be different from Y since Y's group-A portion is the whole A. □

Now what remains to be shown is that the DFA T is minimal:

Lemma 4. *Any two states X and Y of T are not equivalent.*

Proof. We first consider the case that X and Y differ in their group-A portion. Let $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ where X_1 and Y_1 are their group-A portions. Once again recall that the transition by reading 0 is a "cyclic shift". Therefore, if $X_1 \neq Y_1$ then there exists some $i \geqslant 0$ such that $\delta(X_1, 0^i)$ contains S_0 but $\delta(Y_1, 0^i)$ does not or vice versa (δ is the transition function of T). In either case one of them is accepting and the other is not. (Actually, the states in X_2 and Y_2 are also involved but they have no effect on whether or not those F-states are accepting.) Thus if $X_1 \neq Y_1$ then X and Y are not equivalent.

Next suppose that $X_1 = Y_1$. Then X_2 and Y_2 must be different. Let $X' = \delta(X, 1)$ and $Y' = \delta(Y, 1)$. Then one can see that the group-A portions of X' and Y' are different.

current state	next states	
	0	1
S_0	S_1	S_1
S_1	S_2	S_1, S_2
S_2	S_3	S_1, S_3
.	.	.
.	.	.
S_i	S_{i+1}	S_1, S_{i+1}
.	.	.
.	.	.
S_{n-2k-3}	S_{n-2k-2}	S_1, S_{n-2k-2}
S_{n-2k-2}	S_{n-2k-1}	S_1, S_{n-2k-1}
S_{n-2k-1}	S_{n-2k}	S_{n-k}
S_{n-2k}	S_{n-2k+1}	S_{n-k+1}
.	.	.
.	.	.
S_{n-k-2}	S_{n-k-1}	S_{n-1}
S_{n-k-1}	S_0	S_1
S_{n-k}	S_{n-k+1}	S_1, S_{n-2k}
S_{n-k+1}	S_{n-k+2}	S_1, S_{n-2k+1}
.	.	.
.	.	.
S_{n-2}	S_{n-1}	S_1, S_{n-k-2}
S_{n-1}	S_{n-k}	S_1, S_{n-k-1}

Fig. 3. Transition function of the NFA M.

The reason is that when T reads 1, S_{n-1} moves to S_{n-3} (and also to S_1) and S_{n-2} moves to S_{n-4} (and also to S_1). Since there are no other transitions to S_{n-3} or to S_{n-4} by reading 1, if X_2 and Y_2 are different then the corresponding states in group-A reached from X_2 and Y_2 by reading 1 are also different. Thus, it turns out that X' and Y' are not equivalent for the same reason as above and hence X and Y are not either. □

3.2. *The case for a general k*

The transition function of T for general k ($0 \leqslant k \leqslant n/2 - 2$) is illustrated in Fig. 3. Its state diagram for $k = 3$ and $n = 10$ is given in Fig. 4. Again the whole state set is partitioned into $A = \{S_0, S_1, \ldots, S_{n-k-1}\}$ and $B = \{S_{n-k}, \ldots, S_{n-1}\}$. What we should be careful in the general case is the following: Recall that one of the key facts in the previous proof is that any F-state $X \subsetneq A$ of size at least two can be changed, by 0-inv-shifts, to an S_1-pattern Y such that T can reach Y from yet another F-state Z, whose size is one state less than Y, by reading 1. This is due to the fact that Z does

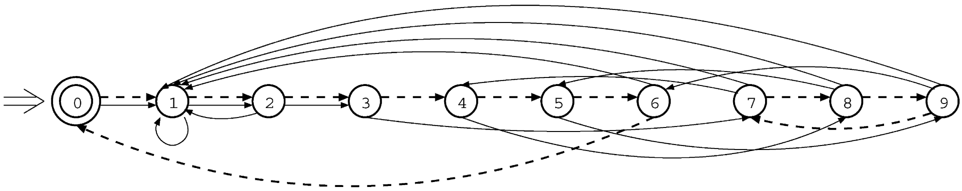

Fig. 4. State diagram of M for $k=3$ and $n=10$.

not include S_1 but does include at least one state S_i for $2\leqslant i\leqslant n-6$ in Fig. 1, from which S_1 is "generated" by reading 1. Let us call such S_i an S_1-*generating state*. In the case of Fig. 3, S_1-generating states are states S_i such that $2\leqslant i\leqslant n-2k-2$. Then one can see that the number of the S_1-generating states decreases as k increases. For example, there are four S_1-generating states, S_1, S_2, S_3 and S_4, in Fig. 2, but only two, S_1 and S_2, in Fig. 4. It is not hard to see that the above fact no longer holds if there are too few S_1-generating states. In other words, if there are an enough number of S_1-generating states, or if k is relatively small (up to some $n/3$), then the proof of the general case is virtually the same as before.

When k is large, we thus have few S_1-generating states. Instead, however, one should notice that we have more and more states in group B. Looking at the state transition, fortunately, it turns out that the group-B states can play the same role as S_1-generating states. See Fig. 2 again and recall that any F-state of size two in group A can be reached from some F-state of size one, which played an important role in the proof. For example, $\{S_1,S_2\}$ from $\{S_1\}$, $\{S_1,S_3\}$ from $\{S_2\}$, $\{S_1,S_6\}$ from $\{S_1,S_3\}$ (by 0-shifts), and so on. This is very similar in the case of Fig. 4: F-states $\{S_1,S_2\}$, $\{S_1,S_3\}$, $\{S_1,S_4\}$, $\{S_1,S_5\}$, and $\{S_1,S_6\}$ are reached from $\{S_1\}$, $\{S_2\}$, $\{S_7\}$, $\{S_8\}$, and $\{S_9\}$, respectively, by reading 1. Although details are omitted, this is the reason why we can enlarge k up to almost $n/2$.

4. Proof of Theorem 2

The transition function of the NFA M is exactly the same as Fig. 3 except only one entry. Namely, the next states from S_0 by reading 1 is changed from S_1 to ϕ. Thus, the F-state ϕ must appear in the equivalent DFA T and ϕ is not equivalent to any other F-state since it is completely impossible to reach any accepting F-state from ϕ. (One can see that there is a path to S_0 from every other state in Fig. 3, which means T can reach some accepting F-state from any F-state of size at least one.)

Thus what we have to prove is that (i) T has all the F-states but $\Gamma-\{\phi\}$ and (ii) any two of them are not equivalent. (ii) is exactly the same as before. To show (i), one should notice that we did not use the transition from S_0 by reading 1 anywhere in Section 3.1. Details may be omitted.

5. Concluding remarks

An apparent future goal is to find an NFA M such that $\Delta(M,n)=2^n-6$. Note that our basic approach in this paper is to divide the whole F-states into two groups and to prohibit the whole group-A states from appearing in the equivalent DFA. Thus the number of disappearing states has to be the size of the power set of group B, which is to be in the form of 2^k. The above number, 6, is exactly the middle between $4(=2^2)$ and $8(=2^3)$, which clearly makes it difficult to apply the above basic approach.

References

[1] A. Ambainis, The complexity of probabilistic versus deterministic finite automata, Proc. 7th ISAAC, Lecture Notes in Computer Science, Vol. 1178, Springer, Berlin, 1996, pp. 231–238.

[2] P. Berman, A. Lingas, On complexity of regular languages in terms of finite automata, ICS PAS Report 304, Warszawa, 1979.

[3] M. Harrison, Introduction to Formal Language Theory, Addison-Wesley, Reading, MA, 1978.

[4] J.E. Hopcroft, J.D. Ullman, Introduction to Automata Theory, Languages and Computation, Addison-Wesley, Reading, MA, 1979.

[5] H. Lewis, C. Papdimitoriou, Elements of the Theory of Computation, Prentice-Hall, Englewood Cliffs, NJ, 1981.

[6] F. Moore, On the bounds for state-set size in the proofs of equivalence between deterministic, nondeterministic, and two-way finite automata, IEEE Trans. Comput. C-20 (1971) 1211–1214.

[7] A. Nozaki, Equivalence problem of non-deterministic finite automata, J. Comput. System Sci. 18 (1979) 8–17.

[8] M. Rabin, D. Scott, Finite automata and their decision problems, IBM J. Res. Dev. 3 (1959) 114–125.

[9] A. Salomaa, Computation and Automata, Cambridge University Press, Cambridge, 1985.