A Family of NFA's Which Need $2^n - \alpha$ Deterministic States

Kazuo Iwama[1], Akihiro Matsuura[1], and Mike Paterson[2]

[1] School of Informatics, Kyoto University
[2] Department of Computer Science, University of Warwick

E-mail: {iwama, matsu}@kuis.kyoto-u.ac.jp, msp@dcs.warwick.ac.uk

Abstract. We show that for all integers $n \geq 7$ and α, such that $5 \leq \alpha \leq 2n - 2$ and satisfying some coprimality conditions, there exists a minimum n-state nondeterministic finite automaton that is equivalent to a minimum deterministic finite automaton with exactly $2^n - \alpha$ states.

1 Introduction

Finite automata theory is obviously a popular first step to theoretical computer science, through which students learn several basic notions of computation models. Nondeterminism might be the most important one among those notions. The subset construction [1], which shows that any nondeterministic finite automaton (NFA) can be simulated by a deterministic finite automaton (DFA), is probably one of the oldest, non-trivial theorems in this field. This theorem is often stated as above, i.e., "NFA's are no stronger than DFA's", but we have to be careful since the simulation is only possible "by increasing the number of states". Since the number of states is the principal complexity measure for finite automata, the extent to which NFA's are more efficient than DFA's is an important feature and provides the basis for the same relationship in stronger models.

It is known [2], [3] that there is an NFA of n states which needs 2^n states to be simulated by a DFA. Thus some NFA's are exponentially more efficient than DFA's in terms of the number of states. Of course, however, this is not always true; for example, the DFA which counts the number of 1's modulo k needs k states and equivalent NFA's need the *same* number of states. So, nondeterminism works very well for some kind of languages and does not for others. Thus it is of interest to ask which kinds of language belong to the first category and which to the second.

It is hard to give a general answer to this problem. However, one simple and concrete question regarding this problem is the following: *For a positive integer n, is there an integer Z, $n < Z < 2^n$, such that no DFA of Z states can be simulated by any NFA of n states?* Such a number Z or the one that satisfies the above question for all n can be regarded as a "magic number" for which nondeterminism is especially weak. It turns out that to answer this question,

M. Nielsen and B. Rovan (Eds.): MFCS 2000, LNCS 1893, pp. 436–445, 2000.

we have only to consider $2^{n-1} \leq Z < 2^n$. Furthermore, 2^{n-1} cannot be such a magic number [4]. If there are no such magic numbers at all, which seems more likely to us, that means that for any integer $0 \leq \alpha \leq 2^{n-1} - 1$, there is an NFA of n states which needs $2^n - \alpha$ deterministic states.

This question was first considered by Iwama, Kambayashi and Takaki [5]. They show that if an integer α can be expressed as 2^k or $2^k + 1$ for some integer $0 \leq k \leq n/2-2$, then there is an NFA of n states which needs $2^n - \alpha$ deterministic states, i.e., such $2^n - \alpha$ cannot be a magic number in the above sense. In this paper, we give a somewhat (but not yet completely) general answer. Namely, for all integers $n \geq 7$ and α, such that $5 \leq \alpha \leq 2n - 2$ and with some coprimality condition, $2^n - \alpha$ cannot be a magic number. Furthermore, we show that $2^n - 6$ cannot be a magic number, unconditionally. Note that $2^n - 6$ is the largest number which cannot be expressed as 2^k or $2^k + 1$, and so was left open in [5].

2 Main Results

A finite automaton M is determined by the following five items: a finite set of states; a finite set of input symbols Σ, which is always $\{0, 1\}$ in this paper; an initial state; a set of accepting states; and a state transition function δ.

Our main task in this paper is (i) to give an NFA M, (ii) to find the equivalent DFA, (iii) to analyze the number of states in the DFA which can be reached from its initial state, and finally (iv) to show that all such states are inequivalent. For (ii), we use the so-called subset construction [1], i.e., each state of the DFA is given as a subset of M's states and the resulting DFA is written as $D(M)$. To avoid confusion, a state of $D(M)$ will be called an *f-state* (*f* stands for family). We always use δ for the state transition function of $D(M)$. Two f-states Q_1 and Q_2 are *equivalent* if for all $x \in \Sigma^*$, $\delta(Q_1,\ x) \in F$ iff $\delta(Q_2,\ x) \in F$, where F is the set of accepting states in $D(M)$. Suppose on the other hand that we wish to show that two f-states Q_1 and Q_2 are not equivalent. Then, what we should do is (i) to show that $Q_1 \in F$ and $Q_2 \notin F$ (or vice versa), or (ii) to find a string $x \in \Sigma^*$ such that $\delta(Q_1,\ x)$ and $\delta(Q_2,\ x)$ are already known to be inequivalent. For an NFA M of n states, $\Delta(M)$ denotes the number of states of a minimum DFA which is equivalent to M. It is well known [1] that a DFA is minimum if all of its states can be reached from the initial state and no two states are equivalent. Now we are ready to give our results.

Theorem 1. *Let n and α be any integers such that $5 \leq \alpha \leq n-1$, $6 \leq \alpha \leq n$, or $9 \leq \alpha \leq 2n - 2$ and such that n is relatively prime with $\alpha - 1$, $\alpha - 2$, or $\lceil \alpha/2 \rceil - 1$, respectively. Then there exists a minimum n-state NFA whose equivalent minimum DFA has $2^n - \alpha$ states.*

Corollary 1. *For all integers $n \geq 7$ and α, such that $5 \leq \alpha \leq 2n - 2$ and satisfying the comprimality condition in Theorem 1, there exists an n-state NFA whose equivalent minimum DFA has $2^n - \alpha$ states.*

Note that for $\alpha \leq 5$, it was shown in [5] that there exists an n-state NFA M such that $\Delta(M) = 2^n - \alpha$ for $n \geq 8$.

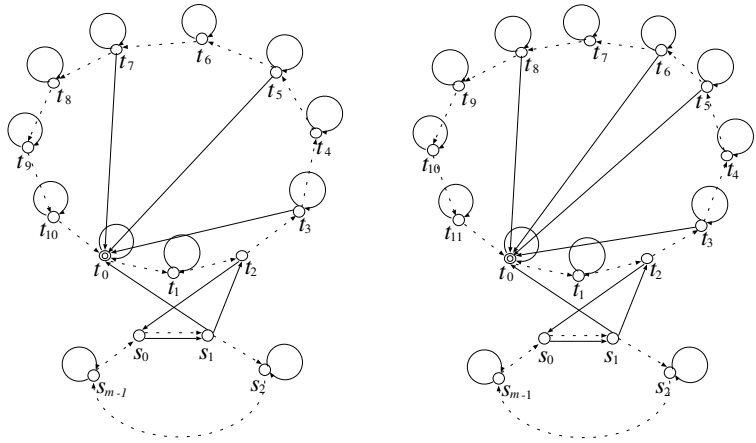


Fig. 1. (a) M_1 when k is odd ($k = 11$) and (b) M_1 when k is even ($k = 12$)

The next theorem is less general, but does not need the coprimality condition. Recall that $2^n - 6$ was the first unsettled number in [5].

Theorem 2. *For any $n \geq 5$, there exists an n-state NFA whose equivalent minimum DFA has $2^n - 6$ states.*

3 Proof of Theorem 1

For ease of explanation, we introduce the parameter k that represents $\alpha - 1$, $\alpha - 2$, or $\lceil \alpha/2 \rceil - 1$, corresponding to the three cases in the hypothesis, and we suppose that k and n have no common divisor. Let m denote $n - k$, i.e., $n = k + m$, then k and m also have no common divisor. In this section, we first give an NFA M_1 whose equivalent minimum DFA has $2^n - (k + 1)$ states. Then we give five lemmas which give the number of f-states in $D(M_1)$ and claim that no two f-states are equivalent. M_1 is illustrated in Fig. 1. Its state set is the union of $T = \{t_0, t_1, \cdots, t_{k-1}\}$ and $S = \{s_0, s_1, \cdots, s_{m-1}\}$. Its initial state is t_0. Note that $|T \bigcup S| = k + n = m$. A set in T (in S, resp.) is called a T-state (S-state, resp.). State transitions on reading 0 (denoted by dotted arrows in Fig. 1) are cyclic, i.e., $t_0 \xrightarrow{0} t_1$, $t_1 \xrightarrow{0} t_2$, $\cdots$, $t_{k-1} \xrightarrow{0} t_0$, and $s_0 \xrightarrow{0} s_1$, $\cdots$, $s_{m-1} \xrightarrow{0} s_0$. Transitions on reading 1 are as follows: For all i, $0 \leq i \leq k-1$ excepting $i = 2$, there are self-loops as $t_i \xrightarrow{1} t_i$. Similarly, $s_j \xrightarrow{1} s_j$ for $2 \leq j \leq m - 1$. In addition, there are transitions of the form $t_i \xrightarrow{1} t_0$ where $i = 3,\ 5,\ \cdots,\ k-4$ when k is odd. When k is even, these transitions are defined for $i = 3,\ 5,\ \cdots,\ 2r-3,\ 2r-1,\ 2r,\ 2r+2,\ \cdots,\ k-4$ where $r = \lceil k/4 \rceil$. The remaining four transitions are $s_0 \xrightarrow{1} s_1$, $s_1 \xrightarrow{1} t_0$, $s_1 \xrightarrow{1} t_2$, and $t_2 \xrightarrow{1} s_0$.

For any f-state P, $P \cap T$ is called the *T-portion* of P and denoted by P_T. Similarly, $P \cap S$ is called the *S-portion* of P and denoted by P_S. The *size* of P,

$|P|$, is the number of M_1's states included in P. The transition on T (or S) that occurs on reading 0 is called a *0-shift*. The index of a state is considered to be modulo k; namely, the 0-shift of t_i is always written as t_{i+1}.

The first lemma deals with exceptional f-states P such that $|P_S| = 0$ and $|P_T| = 0,\ 1$, and 2. We say that an f-state Q_1 is *reachable from* an f-state Q_2 if there is a string $x \in \Sigma^*$ such that $Q_1 = \delta(Q_2,\ x)$. If $Q_1 = \delta(\{t_0\},\ x)$, then we simply say that Q_1 is *reachable*. Q_1 is said to be *unreachable* if it is not reachable.

Lemma 1. *For an f-state P such that $|P_S| = 0$ and $0 \leq |P_T| \leq 2$, the following statements hold. (1) When $|P_T| = 0$, there is only one f-state, ϕ (the empty set), and this is unreachable. (2) When $|P_T| = 1$, P is reachable. (3) When $|P_T| = 2$, P is reachable unless P consists of two neighboring states of T, that is, $P = \{t_i,\ t_{i+1}\}$ $(i = 0,\ 1,\ \cdots,\ k-1)$.*

Note that there are $k+1$ unreachable f-states given in this lemma. The remaining $2^n - (k+1)$ f-states are all reachable, which is shown by the following four lemmas depending on (i) whether $|P_S| = 0$ or $|P_S| > 0$ and (ii) whether or not P_T contains two states of distance two, i.e., t_i and t_{i+2}. Distance-two states are important since the transition $s_1 \xrightarrow{1} \{t_0,\ t_2\}$ plays a special role in M_1.

Lemma 2. *For an f-state P such that $|P_S| = 0$ and $|P_T| \geq 3$, if P contains a pair of states, t_i and t_{i+2} for some $i = 0,\ 1,\ \cdots,\ k-1$ (P may include t_{i+1} as well), P is reachable from some f-state Q such that $(|Q_S|,\ |Q_T|) = (1,\ |P_T|-2)$.*

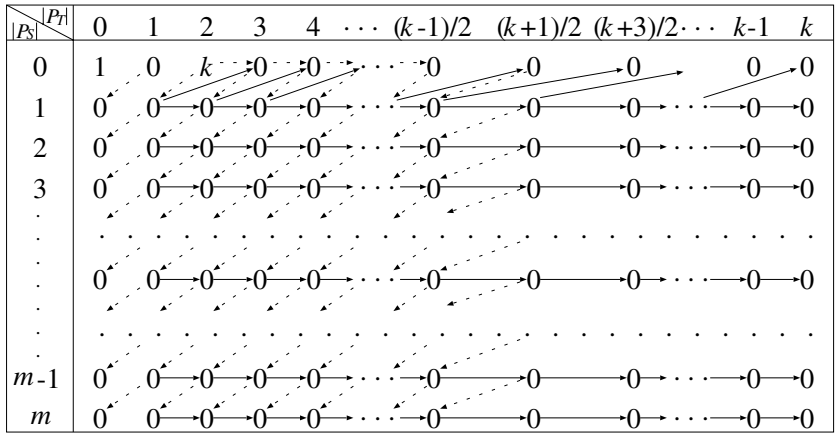
Lemma 3. *For an f-state P such that $|P_S| = 0$ and $|P_T| \geq 3$, if P does not contain a pair of states, t_i and t_{i+2} for any $i = 0,\ 1,\ \cdots,\ k-1$, P is reachable from some f-state Q such that $(|Q_S|,\ |Q_T|) = (0,\ |P_T| - 1)$. Furthermore, if $|P_T| = 3$, $Q_T \neq \{t_i,\ t_{i+1}\}$, i.e., the two states of Q_T are not neighboring.*

Lemma 4. *For an f-state P such that $|P_S| \geq 1$, if P contains a pair of states t_i and t_{i+2} for some $i = 0,\ 1,\ \cdots,\ k-1$, P is reachable from some f-state Q such that $(|Q_S|,\ |Q_T|) = (|P_S|,\ |P_T| - 1)$.*

Lemma 5. *For an f-state P such that $|P_S| \geq 1$, if P does not contain a pair of states t_i and t_{i+2} for any $i = 0,\ 1,\ \cdots,\ k-1$, P is reachable from some f-state Q such that $(|Q_S|,\ |Q_T|) = (|P_S| - 1,\ |P_T| + 1)$. Furthermore, when $(|P_S|,\ |P_T|) = (1,\ 1)$, $Q_T \neq \{t_i,\ t_{i+1}\}$, i.e., the two states of Q_T are not neighboring.*

See Table 1, which summarizes Lemmas 1 to 5 and also summarizes our induction scheme to claim how each f-state is reachable for an odd k. The leftmost three entries $(1, 0,$ and $k)$ in its first row show the numbers of unreachable f-states described in Lemma 1. Dotted arrows show the reachability described in Lemmas 3 and 5. Solid arrows show the reachability given in Lemmas 2 and 4. For example, the entry for $(|P_S|,\ |P_T|) = (0,\ 4)$ receives a dotted arrow from $(|P_S|,\ |P_T|) = (0,\ 3)$ and a solid arrow from $(|P_S|,\ |P_T|) = (1,\ 2)$. Two dotted arrows from $(|P_S|,\ |P_T|) = (0,\ 2)$ need special care since this entry includes unreachable f-states, or we have to show that those reachabilities do not start from such unreachable states. Also, one should notice that there are no dotted arrows to any P such that $|P_T| \geq (k+1)/2$. The reason is that if

Table 1. Number of unreachable f-states when k is odd



$|P_T| \geq (k+1)/2$, then P_T must include a pair of distance-two states. Altogether, each f-state P such that $(|P_S|,\ |P_T|) = (1,\ 0)$ is reachable from Q such that $(|Q_S|,\ |Q_T|) = (0,\ 1)$, and such a Q is reachable from $\{t_0\}$ by Lemma 1. All the other f-states are reachable by traversing solid and dotted arrows starting from $(|P_S|,\ |P_T|) = (0,\ 2)$, and the latter f-states are reachable by Lemma 1.

3.1 Proof of Lemma 1

ϕ is obviously unreachable since every state in M_1 has non-empty transitions on reading 0 and 1. When $|P_T| = 1$, P can be written as $\{t_i\}$, which is reachable from $\{t_0\}$, the initial f-state, by 0-shifts.

We now consider the case $(|P_S|,\ |P_T|) = (0,\ 2)$, which is divided into two cases according to whether or not P contains a neighboring pair of states in T. The argument is a little different for odd and even k's. In the following, we only consider the odd case. First, suppose that $P = \{t_0, t_i\}$, where $i \neq 1,\ k-1$, namely the two states of P are not neighboring. When $i = 3,\ 5,\ 7,\ \cdots,\ k-4$, we can use the following transitions:

$$\{t_0\} \xrightarrow{0^i} \{t_i\} \xrightarrow{1} \{t_0,\ t_i\}.$$

When $i = 2$ and $k-2$, we can follow

$$\{t_0\} \xrightarrow{0^2} \{t_2\} \xrightarrow{1} \{s_0\} \xrightarrow{1} \{s_1\} \xrightarrow{1} \{t_0,\ t_2\} \xrightarrow{0^{k-2}} \{t_0,\ t_{k-2}\}.$$

When $i = 4,\ 6,\ 8,\ \cdots,\ k-3$, we can follow

$$\{t_0\} \xrightarrow{0^{k-i}} \{t_{k-i}\} \xrightarrow{1} \{t_0,\ t_{k-i}\} \xrightarrow{0^i} \{t_0,\ t_i\}.$$

Thus $P = \{t_0,\ t_i\}$ is reachable unless $i = 1$ or $i = k-1$. All other non-neighboring f-states are reachable from $\{t_0,\ t_i\}$ by 0-shifts.

As for a neighboring pair of states such as $\{t_i,\ t_{i+1}\}$, this is shown to be unreachable as follows. First of all, one can see that if we do not use the transition from t_2 to s_0, we can never reach $\{t_i,\ t_{i+1}\}$, for the following reasons. We start from $\{t_0\}$. Then, if we use only transitions between T-states, which we call T-*transitions*, then the size $|P|$ of the current f-state P monotonically increases. Hence, consider the moment when $|P|$ changes from one to two. The transition used at this moment must be $t_i \xrightarrow{1} t_i$ and $t_i \xrightarrow{1} t_0$. It then follows that P cannot be neighboring since we have no such transitions from t_1 or t_{k-1}. It is easy to see that such a P cannot later change to a pair of neighboring states while $|P| = 2$. Thus there must be an f-state which includes some S-state on the way from $\{t_0\}$ to $\{t_i, t_{i+1}\}$ (if any). Let K be the last f-state including S-states. Then, symbol 1 must be read on state K, since otherwise $\delta(K,\ 0)$ still contains both S- and T-states. Furthermore, K never contains s_1, since otherwise $\delta(K,\ 1)$ includes $\{t_0,\ t_2\}$, which cannot change to a pair of neighboring states by using T-transitions. Hence, K must only contain some S-state other than s_1, but this contradicts our assumption for K. □

3.2 Proof of Lemma 2

Suppose that $\{t_i,\ t_{i+2}\} \subset P$. Obviously, P is reachable from some P' such that $\{t_0,\ t_2\} \subset P'$. Now we can see that $Q = (P' \backslash \{t_0,\ t_2\}) \bigcup \{s_1\} \xrightarrow{1} P'$, where $P' \backslash \{t_0,\ t_2\}$ means that $\{t_0,\ t_2\}$ is removed from P'. Thus Q satisfies the condition of the lemma, i.e., $(|Q_S|,\ |Q_T|) = (1,\ |P_T| - 2)$. □

3.3 Proof of Lemma 3

Now P does not include any $\{t_i,\ t_{i+2}\}$. Without loss of generality, we can assume that P contains t_0 (otherwise, P is reachable from such an f-state by 0-shifts). Hence, let $P = \{t_0,\ t_{p_1},\ t_{p_2},\ \cdots,\ t_{p_{r-1}}\}$, where $|P| \geq 3$ and $p_1 = 1$ or $p_1 \geq 3$ since there is no pair of distance-two states. The proof differs slightly according to whether k is odd or even (recall that our machine M_1 is different for odd and even k's). We first prove the lemma for an odd k and the difference in the even case will be briefly given. There are several cases to be considered.

(*Case 1*) $p_1 = 1$, namely, $P = \{t_0,\ t_1,\ t_{p_2},\ \cdots,\ t_{p_{r-1}}\}$. This case is further divided into two subcases according to whether p_2 is odd or even.

(*Case 1-1*) p_2 is odd. By the assumption of Lemma 3, $4 \leq p_2 \leq k-3$, and since p_2 was assumed to be odd, $5 \leq p_2 \leq k-4$. Therefore, we can use the transition $t_{p_2} \xrightarrow{1} \{t_0,\ t_{p_2}\}$, namely:

$$Q = P \backslash \{t_0\} = \{t_1,\ t_{p_2},\ \cdots,\ t_{p_{r-1}}\} \xrightarrow{1} \{t_0,\ t_1,\ t_{p_2},\ \cdots,\ t_{p_{r-1}}\} = P.$$

Note that Q satisfies the condition of the lemma, i.e., $(|Q_S|,\ |Q_T|) = (0,\ |P_T|-1)$. It should be noted that for $r = 3$, $Q = \{t_1,\ t_{p_2}\}$ is known to be reachable by Lemma 1 because t_1 and t_{p_2} are not neighboring.

(*Case 1-2*) p_2 is even. Since $4 \leq p_2 \leq k-3$ and p_2-1 is odd, $3 \leq p_2-1 \leq k-4$. Therefore, there is a transition $t_{p_2-1} \stackrel{1}{\longrightarrow} \{t_0,\ t_{p_2-1}\}$. Let $P' = \delta(P,\ 0^{k-1}) = \{t_{k-1},\ t_0,\ t_{p_2-1},\ t_{p_3-1}, \cdots,\ t_{p_{r-1}-1}\}$. Then, we can use the following sequence of transitions:

$$Q = P' \backslash \{t_0\} \stackrel{1}{\longrightarrow} P' \stackrel{0}{\longrightarrow} P.$$

For $r = 3$, $Q = \{t_{k-1},\ t_{p_2-1}\}$ is not neighboring again.

(*Case 2*) $p_1 \geq 3$. We can assume that t_{p_1} and t_{p_2} are not neighboring, since otherwise we can apply the argument of Case 1. Also note that $p_1 \leq k-3$ (otherwise P, $|P| \geq 3$, clearly includes a pair of distance-two states).

(*Case 2-1*) p_1 is odd. We have the following direct transition to P using the transitions $t_{p_1} \stackrel{1}{\longrightarrow} \{t_0,\ t_{p_1}\}$ in M_1:

$$P \backslash \{t_0\} = \{t_{p_1},\ t_{p_2},\ \cdots,\ t_{p_{r-1}}\} \stackrel{0}{\longrightarrow} P.$$

(*Case 2-2*) p_1 is even. Since $4 \leq p_1 \leq k-3$ and $k-p_1$ is odd, $3 \leq k-p_1 \leq k-4$. This time we use $t_{k-p_1} \stackrel{1}{\longrightarrow} \{t_0,\ t_{k-p_1}\}$. Let $P' = \delta(P,\ 0^{k-p_1}) = \{t_{k-p_1},\ t_0,\ t_{p_2-p_1},\ \cdots,\ t_{p_{r-1}-p_1}\}$. Then, we obtain the following sequence of transitions:

$$P' \backslash \{t_0\} \stackrel{1}{\longrightarrow} P' \stackrel{0^{p_1}}{\longrightarrow} P.$$

Thus, P is reachable from $Q = P' \backslash \{t_0\}$. Again for $r = 3$, $Q = \{t_{k-p_1},\ t_{p_2-p_1}\}$ is not neighboring and is known to be reachable by Lemma 1. Consequently, it has been shown that in all cases, there is a transition of the form $Q \longrightarrow P$, where Q satisfies $(|Q_S|,\ |Q_T|) = (0,\ |P_T|-1)$. □

For an even k, it is divided into three cases: $p_1 = 1$, $3 \leq p_1 \leq k/2-1$, and $p_1 \geq k/2$. In each case, the reachability of the f-states are shown in the similar way to the odd case, where the transitions of type $t_i \stackrel{1}{\longrightarrow} \{t_0,\ t_i\}$ play the essential role again.

3.4 Proof of Lemma 4

Recall that $\{t_i,\ t_{i+2}\} \subset P$ and $|P_S| \geq 1$. We consider two cases, one for $|P_S| \leq m-1$ and the other for $|P_S| = m$.

(*Case 1*) $|P_S| \leq m-1$. Since k and n have no common divisor and since $P_S \neq S$, there is an f-state P' such that (i) P is reachable from P', (ii) $\{t_0,\ t_2\} \subset P'$, and (iii) $s_0 \in P'$ and $s_1 \notin P'$. Let $P'_1 = P'_T \backslash \{t_0,\ t_2\}$ and $P'_2 = P'_S \backslash \{s_0\}$. Then, one can use the following transition:

$$Q = (P'_1 \bigcup \{t_2\}) \bigcup (P'_2 \bigcup \{s_1\}) \stackrel{1}{\longrightarrow} (P'_1 \bigcup \{s_0\}) \bigcup (P'_2 \bigcup \{t_0,\ t_2\}) = P',$$

since P'_1 and P'_2 do not change on reading 1, $\{t_2\} \stackrel{1}{\longrightarrow} \{s_0\}$, and $\{s_1\} \stackrel{1}{\longrightarrow} \{t_0,\ t_2\}$. Note that $(|Q_S|,\ |Q_T|) = (|P_S|,\ |P_T|-1)$ and the lemma follows.

(*Case 2*) $|P_S| = m$. Namely, $P_S = S$. Similarly to Case 1, there is an f-state P' such that (i) P is reachable from P', (ii) $\{t_0,\ t_2\} \subset P'$, and (iii) $P'_S = P_S = S$.

Let $P'_1 = P'_T \backslash \{t_0,\ t_2\}$ and $P'_2 = P_S \backslash \{s_0,\ s_1\}$. Then one can use the following transition:

$$Q = (P'_1 \bigcup \{t_2\}) \bigcup (P'_2 \bigcup \{s_0,\ s_1\}) \xrightarrow{1} (P'_1 \bigcup \{s_0\}) \bigcup (P'_2 \bigcup \{s_1,\ t_0,\ t_2\}) = P',$$

where $(|Q_S|,\ |Q_T|) = (|P_S|,\ |P_T| - 1)$. □

3.5 Proof of Lemma 5

Suppose that P does not include any $\{t_i,\ t_{i+2}\}$. We consider two cases similarly to Section 3.4.

(*Case 1*) $|P_S| \leq m - 1$. As before, there is an f-state P' such that (i) P is reachable from P', (ii) $t_0 \in P'$ and $t_2 \notin P'$, and (iii) $s_0 \in P'$ and $s_1 \notin P'$. Let $P'_1 = P'_T \backslash \{t_0\}$ and $P'_2 = P'_S \backslash \{s_0\}$. Then, one can use the following transition:

$$Q = (P'_1 \bigcup \{t_0,\ t_2\}) \bigcup P'_2 \xrightarrow{1} (P'_1 \bigcup \{t_0,\ s_0\}) \bigcup P'_2 = P',$$

where $(|Q_S|,\ |Q_T|) = (|P_S| - 1,\ |P_T| + 1)$.

(*Case 2*) $|P_S| = m$. In this case, there is an f-state P' such that (i) P is reachable from P', (ii) $t_0 \in P'$ and $t_2 \notin P'$, and (iii) $P'_S = P_S = S$. Let $P'_1 = P_T \backslash \{t_0\}$ and $P'_2 = P_S \backslash \{s_0,\ s_1\}$. Then, one can use the following transition:

$$(P'_1 \bigcup \{t_0,\ t_2\}) \bigcup (P'_2 \bigcup \{s_0\}) \xrightarrow{1} (P'_1 \bigcup \{t_0,\ s_0\}) \bigcup (P'_2 \bigcup \{s_1\}) = P',$$

where $(|Q_S|,\ |Q_T|) = (|P_S| - 1,\ |P_T| + 1)$. □

We note that the proofs of Lemmas 4 and 5 do not depend on whether k is odd or even.

3.6 Inequivalence of Reachable f-States

We have so far shown that the number of reachable f-states in $D(M_1)$ is $2^{k+m} - (k+1) = 2^n - (k+1)$. Now we prove that those f-states are pair-wise inequivalent.

Lemma 6. *Any two reachable f-states of $D(M_1)$ are not equivalent.*

Proof. Let X and Y be two f-states such that $X \neq Y$. If $X_T \neq Y_T$, there must be an integer j such that $t_0 \in \delta(X_T,\ 0^j)$ and $t_0 \notin \delta(Y_T,\ 0^j)$. Thus X and Y are not equivalent. Next, suppose that $X_T = Y_T$ and $X_S \neq Y_S$. Then, there is an integer j such that $s_1 \in \delta(X_S,\ 0^j)$ $(= X')$ and $s_1 \notin \delta(Y_S,\ 0^j)$ $(= Y')$. We then read a 1, and $t_0 \in \delta(X',\ 1)$ while $t_0 \notin \delta(Y',\ 1)$. Therefore, $\delta(X',\ 1)$ and $\delta(Y',\ 1)$ have different T-portions and so are not equivalent, as shown previously. Hence, X and Y are not equivalent. □

Table 2. Numbers of unreachable f-states

NFA \ $\lvert P_T\rvert$	0	1	2	3	4	$\cdots$	k -1	k	Total
M_1	1	0	k	0	0	$\cdots$	0	0	k +1
M_2	2	0	k	0	0	$\cdots$	0	0	k +2
M_3	1	0	k	k	0	$\cdots$	0	0	2k +1
M_4	2	0	k	k	0	$\cdots$	0	0	2k +2

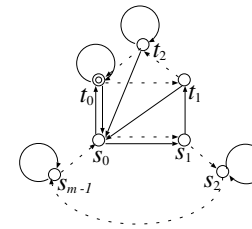


Fig. 2. M_5

3.7 Theorem 1 for $k = \alpha - 2$ and $\lceil \alpha/2 \rceil - 1$

We consider several modifications of M_1 to construct NFA's that realize other numbers of unreachable f-states. The modifications we consider are (i) to eliminate or add some transitions at M_1, and (ii) to modify slightly some transitions of the type $t_i \xrightarrow{1} t_0$ to increase the number of unreachable states. Using the first type of modification, we obtain the following lemma.

Lemma 7. *Let M_2 be the NFA such that $s_0 \xrightarrow{1} t_0$ is added to M_1 and such that m is relatively prime with $k = \alpha - 2$. Then, the f-state S (i.e.,$|P_T| = 0$ and $|P_S| = m$) is unreachable, while the reachability of the other f-states is the same as for M_1.*

We omit the detailed proof but the intuition is as follows. Since $|P_T| = 0$ and $|P_S| = m$, we have to "remove" all T-states and "fill" all the S-states on reading the final 1. Previously, i.e., when there was no transition from s_0 to t_0, we could do this by using $\{s_0,\ t_2\} \xrightarrow{1} \{s_0,\ s_1\}$. This is now impossible, since we have the transition $s_0 \xrightarrow{1} t_0$. □

Using the second type of modification, we construct the NFA M_3. M_3 has the transitions of the type $t_i \xrightarrow{1} t_0$ as follows. When k is odd, transitions $t_i \xrightarrow{1} t_0$ are defined for $i = 3,\ 4,\ 6,\ 8,\ \cdots,\ k-5$. When $k \bmod 4 = 0$, they are defined for $i = 3,\ 4,\ 6,\ 7,\ \cdots,\ k-6,\ k-5$. When $k \bmod 4 = 2$, they are defined for $i = 3,\ 4,\ 6,\ 7,\ \cdots,\ k-8,\ k-7,\ k-5$. Suppose that m is relatively prime with $k = \lceil \alpha/2 \rceil - 1$. Then with regard to unreachable f-states of M_3, we obtain the following lemma.

Lemma 8. *In addition to the unreachable f-states for M_1, M_3 has new unreachable f-states of the type $\{t_i,\ t_{i+3},\ t_{i+4}\}$ $(0 \le i \le k-1)$.*

The numbers of unreachable states for the M_i's are summarized in Table 2.

Remark. Our assumption that k and m have no common divisor is necessary. For example, consider a simple case where $k = m$ or $|T| = |S|$. Then, $\{t_1,\ t_2,\ s_0\}$, which was formerly reachable, turns out to be unreachable.

4 NFA for Theorem 2

For the NFA M_5 given in Fig. 2, there are six unreachable f-states, i.e., ϕ, $S = \{s_0,\ s_1,\ \cdots,\ s_{m-1}\}$, $\{t_0,\ t_1\}$, $\{t_1,\ t_2\}$, $\{t_2,\ t_0\}$, and $\{t_0,\ t_1,\ t_2\}$. Furthermore, all the reachable f-states are inequivalent; thus, $\Delta(M_5) = 2^{m+3} - 6 = 2^n - 6$. The proof for the reachability of f-states is similar to Theorem 1 except for the divisible case, i.e., $n \bmod 3 = 0$. In this case, we explicitly construct transitions for each f-state instead of using the coprimality condition and the 0-shifts.

5 Concluding Remarks

In this paper, we presented families of NFAs with n states, whose equivalent minimum DFAs have $2^n - \alpha$ states, subject to coprimality conditions on n and α. These NFAs are minimum since the equivalent DFAs have more than 2^{n-1} states. Finally, we conjecture that for all n there exists an n-state NFA M such that $\Delta(M) = 2^n - \alpha$ for any $0 \leq \alpha < 2^{n-1}$. To reach this range of α (without "holes" as in [5]) will need some new ideas.

Acknowledgments. We thank the anonymous referees for their helpful comments on our earlier version. We also gratefully acknowledge the help provided by the package kbmag, developed by Derek Holt (Department of Mathematics, University of Warwick), which contains a program for computing the minimum DFA equivalent to a given NFA. We used this program extensively during the design phase of our constructions; it enabled us to check potential NFAs for reasonable values of k and m, e.g., $k = 15$, $m = 10$.

References

1. M. Rabin and D. Scott, "Finite automata and their decision problems," *IBM J. Res. Develop.* 3, pp. 114-125, 1959.
2. O. B. Lupanov, "Uber den Vergleich zweier Typen endlicher Quellen," *Probleme der Kybernetik*, Vol. 6, pp. 329-335, Akademie-Verlag, Berlin, 1966.
3. 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, pp. 1211-1214, 1971.
4. J. E. Hopcroft and J. D. Ullman, Introduction to automata theory, languages and computation, Addison-Wesley, 1979.
5. K. Iwama, Y. Kambayashi, and K. Takaki, "Tight bounds on the number of states of DFA's that are equivalent to n-state NFA's," *Theoretical Computer Science*, to appear. (http://www.lab2.kuis.kyoto-u.ac.jp/~iwama/NfaDfa.ps)