Information Processing Letters 20 (1985) 95–98
North-Holland

15 February 1985

CODETERMINISTIC AUTOMATA ON INFINITE WORDS

D. BEAUQUIER and D. PERRIN

L.I.T.P., Université Paris 7, 2 Place Jussieu, 75251 Paris Cedex 05, France

Communicated by L. Boasson
Received June 1984

We prove that any recognizable set of infinite words is the infinite behaviour of some finite codeterministic automaton.

Keywords: Infinite word, finite automaton

1. Introduction

There are several possibilities for the definition of acceptance of an infinite word by a finite automaton. One of them, often referred to as *Buchi's condition*, requires that, for some computation, the automaton goes infinitely often through a final state. Another one, known as *Muller's condition* requires that, for some computation, the set of states reached infinitely often belongs to some prescribed family of sets of states. These two conditions can be proved to be equivalent in the sense that, given a finite automaton accepting a set of infinite words with Muller's condition, one may construct another accepting the same set with Buchi's condition and vice versa. The important difference between both conditions lies in their relationship with determinism. It is not difficult to see that, with Buchi's condition, it is not possible in general to use deterministic automata instead of nondeterministic ones. On the contrary, a deep theorem due to R. McNaughton shows that this is always possible with Muller's condition.

In this paper we prove that any finite automaton is equivalent, with Buchi's condition, to a reverse deterministic automaton, which is what we call a *codeterministic* automaton. It is possible to imagine these codeterministic automata as deterministic automata reading an infinite word from infinity to 0. Accordingly, the number of possible computations on a given infinite word is finite (and bounded by the number of states).

The above result was previously announced by Mostowski [3] but we were not able to understand his proof. Our own motivations for proving this result are related with the study of two-sided infinite words initiated in [2]. The presentation given here is mainly self-contained. The notation and the proof of some elementary results needed can be found in [1].

2. Preliminaries

Let A denote a (finite or infinite) alphabet. We denote by A^* the set of finite words over the alphabet A. The empty word is denoted by 1 and $A^+ = A^* - 1$. We denote by $A^{\mathbb{N}}$ the set of infinite words

$$\alpha = \alpha_0\alpha_1\alpha_2\ldots \quad (\alpha_n \in A).$$

A finite automaton is denoted by $\mathscr{A} = (Q, I, T)$ with Q the set of states of $\mathscr{A}$ and I (respectively T) the set of initial (respectively terminal) states. The set $\mathscr{F} \subset Q \times A \times Q$ of edges of $\mathscr{A}$ is considered as implicit. The *finite behaviour* of the automaton $\mathscr{A}$ is denoted by $\mathscr{A}^*$. It is the subset of $\mathscr{A}^*$ consisting of the labels of paths starting in I and ending in T. The *infinite behaviour* of $\mathscr{A}$ is denoted by $\mathscr{A}^{\mathbb{N}}$. It is the subset of $A^{\mathbb{N}}$ consisting of the labels of infinite paths starting in I and passing infinitely often through T.

An automaton $\mathscr{A}$ is said to be *deterministic* if for any pair $(p, a) \in Q \times A$ there is at most one $q \in Q$ such that (p, a, q) is an edge. Note that we drop the usual condition of uniqueness of the initial state for deterministic automata and the symmetrical condition of uniqueness of the terminal state for codeterministic ones.

A subset X of A^* is said to be *recognizable* if it is the finite behaviour of some finite automaton $\mathscr{A}$. Likewise, a subset U of $A^{\mathbb{N}}$ is said to be recognizable if it is the infinite behaviour of some finite automaton.

It is well known that, contrary to the case of finite words, not all recognizable subsets of $A^{\mathbb{N}}$ are the behaviour of a deterministic automaton. We shall prove below that they are always the behaviour of a codeterministic automaton. Before going to the proof of this fact, we recall some of the results needed for the proof. For the sake of completeness we shall give a short proof of these preliminary results.

For a set $Y \subset A^*$, we denote by Y^ω the subset of $A^{\mathbb{N}}$ composed by $\alpha = y_0 y_1 y_2 \ldots$ with $y_n \in Y - 1$ for all $n \geqslant 0$.

The first result can be considered as a consequence of Ramsey's theorem. It is of constant use in the study of automata on infinite words.

Proposition 2.1. *Let* $\phi : A^* \to M$ *be a morphism from* A^* *into a finite monoid* M. *For each infinite word* $\alpha \in A^{\mathbb{N}}$, *there exists an* $m \in M$ *and an idempotent* $e = e^2 \in M$ *with* $me = m$ *such that*

$$\alpha \in XY^\omega$$

with $X = \phi^{-1}(m)$, $Y = \phi^{-1}(e)$.

Proof. The proof is by induction on the cardinality of M. If $\mathrm{Card}(M) = 1$, there is nothing to prove. Otherwise, let

$$R = \{ r \in M \mid rM = r \}.$$

Consider first the case where α has a factorization

$$\alpha = u_0 v_0 u_1 v_1 \ldots$$

with $\phi(v_i) \in R$ for all $i \geqslant 0$. It is possible to suppose that one has identically $\phi(v_i) = r$ for some $r \in R$. Then $\phi(v_i u_{i+1}) = r$ for all $i \geqslant 0$ and the property is proved with $e = r$, $m = \phi(u_0 v_0 u_1)$.

Otherwise, one has $\alpha = u\beta$ where β has no left factor in R. Let $p \in M$ be such that β has an infinity of left factors in $\phi^{-1}(p)$ and let $P = \{ t \in M \mid pt = p \}$. Then $\beta = v_0 v_1 v_2 \ldots$ with $\phi(v_0 v_1 \ldots v_n) = p$ for all $n > 0$, whence $\phi(v_n) \in P$. Since $p \notin R$, one has $\mathrm{Card}(P) < \mathrm{Card}(M)$. Let $B = \{ b_1, b_2, \ldots \}$ and let $\psi : B^* \to P$ be the morphism defined by $\psi(b_n) = \phi(v_n)$. By the induction hypothesis there is an $m' \in P$ and an idempotent $e \in P$ with $m'e = m'$ such that

$$b_1 b_2 \ldots \in \psi^{-1}(m') [\psi^{-1}(e)]^\omega.$$

Therefore $v_1 v_2 \ldots \in \phi^{-1}(m')[\phi^{-1}(e)]^\omega$ and the property is proved with $m = \phi(uv_0)m'$. □

We say that a subset U of $A^{\mathbb{N}}$ is a *simple recognizable set* if it is of the form $U = XY^\omega$ with $X = \phi^{-1}(m)$, $Y = \phi^{-1}(e)$ for $\phi : A^* \to M$ a morphism onto a finite monoid and $m \in M$, $e = e^2 \in M$, $m = me$ as in Proposition 2.1. These sets are indeed recognizable since it is well known that

(1) $U \subset A^{\mathbb{N}}$ is recognizable iff it is a finite union of sets XY^ω with X, Y recognizable subsets of $A^{\mathbb{N}}$.

(2) $X \subset A^*$ is recognizable iff it is of the form $\phi^{-1}(P)$ for $\phi : A^* \to M$ a morphism onto a finite monoid and $P \subset M$.

Proposition 2.2. *Any recognizable subset* U *of* $A^{\mathbb{N}}$ *is a finite union of simple recognizable sets.*

Proof. In view of property (1) above it is enough to prove the statement for a set $U = XY^\omega$ with X, $Y \subset A^*$ recognizable. In view of property (2) above, there exists a morphism ϕ from A^* into a finite monoid M such that $XY^* = \phi^{-1}(P)$, $Y^* = \phi^{-1}(Q)$ with P, $Q \subset M$. Let then I be the set of pairs (m, e) such that $m \in P$, $e \in Q$, $me = m$ and $e = e^2$. For $i = (m, e) \in I$ let $X_i = \phi^{-1}(m)$, $Y_i = \phi^{-1}(e)$. Then U is the union over I of the sets $X_i Y_i^\omega$. Indeed, one clearly has $X_i Y_i^\omega \subset U$ for all $i \in I$. Conversely, let $\alpha = x y_0 y_1 y_2 \ldots$ with $x \in X$, $y_n \in Y$ for all $n \geqslant 0$. Applying Proposition 2.1 for an alphabet $B = \{ b_0, b_1, \ldots \}$ in one-to-one correspondence with $(x, y_0, y_1, \ldots)$ one has an idempotent $e \in M$ such that $\alpha = uv_0 v_1 v_2 \ldots$ with $u \in XY^*$, $v_n \in Y^*$, $\phi(v_n)$

$= e$ for all $n \leqslant 0$ and $\phi(u)e = \phi(u)$. Then $(\phi(u), e) \in I$ and the property is proved. □

Example 2.3. Let $A = \{a, b\}$ and let

$$U = a^*bA^\omega.$$

A decomposition of U into a union of simple recognizable sets is

$$U = A^*ba^\omega \cup (a^*b)^\omega$$

separating the elements of U according to their finite or infinite number of occurrences of letter b. The corresponding morphism ϕ maps A^* onto the multiplicative submonoid $M = \{0, 1\}$ of the integers with $\phi(a) = 1$, $\phi(b) = 0$. For the first set, one has

$$X_1 = A^*ba^* = \phi^{-1}(0), \qquad Y_1 = a^* = \phi^{-1}(1).$$

For the second set, one has

$$X_2 = Y_2 = A^*ba^* = \phi^{-1}(0).$$

3. Codeterminism

We now come to the proof of our main result.

Theorem 3.1. *Any recognizable subset of $A^{\mathbb{N}}$ is the infinite behaviour of a codeterministic finite automaton.*

Proof. If $\mathscr{A} = (Q, I, T)$ and $\mathscr{A}' = (Q', I', T')$ are two codeterministic automata with $Q \cap Q' = \emptyset$, their union $\mathscr{B} = (Q \cup Q', I \cup I', T \cup T')$ is again codeterministic and $\mathscr{B}^{\mathbb{N}} = \mathscr{A}^{\mathbb{N}} \cup \mathscr{A}'^{\mathbb{N}}$. In view of Proposition 2.2 it is therefore enough to prove the result for a simple recognizable set $U = XY^\omega$ with $X = \phi^{-1}(m)$, $Y = \phi^{-1}(e)$, $e = e^2$, $me = m$ and $M = \phi(A^*)$.

Let $R = Y - A^+Y$ be the set of elements of Y having no proper right factor in Y. Let $\mathscr{A} = (Q, T, t)$ be a codeterministic automaton such that $\mathscr{A}^* = RY$ and such that, with the same set of edges, one has, for some $I \subset Q$,

$$(Q, I, t)^* = X.$$

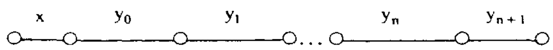
We also suppose that each state $q \in Q$ is coaccessible from t, that is, there exists some $u \in A^*$ such that $q \xrightarrow{u} t$.

We shall prove the theorem by showing that, still with the same set of edges, one has

$$(Q, I, T)^{\mathbb{N}} = XY^\omega.$$

First, consider $\alpha = xy_0y_1y_2\ldots$ with $x \in X$, $y_n \in Y$ for all $n \geqslant 0$.

For each $i \geqslant 0$, denote by r_i the unique right factor of y_i which belongs to R and let $y_i = s_ir_i$. Then, for each $n \geqslant 0$, one has a path c_n of the form



$$i \xrightarrow{x} j_0 \xrightarrow{s_0} t_0 \xrightarrow{r_0} j_1 \xrightarrow{s_1} t_1 \xrightarrow{r_1} j_2 \cdots j_n \xrightarrow{s_n} t_n \xrightarrow{r_n} j_{n+1} \xrightarrow{y_{n+1}} t$$

with $i \in I$ and $t_k \in T$ for $0 \leqslant k \leqslant n$. Since Q is finite, there is a subsequence of the sequence $(c_n)_{n \geqslant 0}$ with is convergent and therefore $\alpha \in (Q, I, T)^{\mathbb{N}}$. This proves that $XY^\omega \subset (Q, I, T)^{\mathbb{N}}$.

To prove the converse inclusion, we need a lemma which is stated below with the same notation as above.

Lemma 3.2. *Let* $f \in M$ *be an idempotent and let* $u, v \in \phi^{-1}(f)$ *be nonempty words such that* $t' \xrightarrow{u} t' \xrightarrow{v} t'$ *for some* $t' \in T$.

Then there exist $r, s \in A^*$ *such that*

$$uv = rs, \qquad r \in Y, \qquad sY \subset Y.$$

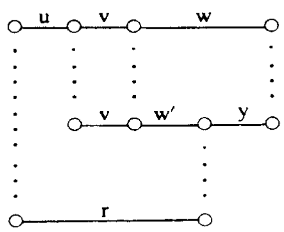
Proof. Since t' is coaccessible, there exists some $w \in A^*$ such that $t' \xrightarrow{w} t$. Since $t' \in T$, one has $uvw = ry$ for some $r \in R$, $y \in Y$. We show that w cannot be a right factor of y. Indeed, one would have in this case

$$r = uvw', \qquad w = w'y.$$

Then, since f is an idempotent and $\phi(u) = \phi(v) = f$, one has

$$\phi(vw') = \phi(v)\phi(w') = f\phi(w') = \phi(uv)\phi(w') = \phi(r).$$

But this implies that $\phi(vw') \in Y$ contradicting the fact that an element of R has no proper right factor in Y (see diagram below):



Therefore, one has, for some $s \in A^*$,

$$uv = rs, \qquad sw = y,$$

and there remains to show that $sY \subset Y$. But since $t' \rightarrow^{w} t$, one has $w \in RY$ whence $\phi(w) = e$. From $sw = y$ one then deduces that $\phi(s)e = e$ and therefore $sY \subset Y$. This proves the lemma. □

Proof of Theorem 3.1 (*continued*). We now consider an infinite word $\alpha \in (Q, I, T)^{\mathbb{N}}$. One then has

$$\alpha = uv_0v_1v_2 \ldots$$

with $i \rightarrow^{u} t'$, $t' \rightarrow^{v_n} t'$ for all $n \geqslant 0$ and $i \in I$, $t' \in T$. Applying Proposition 2.1, one may suppose that $\phi(v_n) = f$ for all $n \geqslant 0$ with $f = f^2$. Then, by the above lemma, for all $n \geqslant 0$,

$$v_{2n}v_{2n+1} = r_ns_n$$

with $r_n \in Y$ and $s_nY \subset Y$. We thus obtain

$$\alpha = ur_0s_0r_1s_1 \ldots$$

with $s_nr_{n+1} \in Y$ for all $n \geqslant 0$. Let us finally show that $ur_0 \in X$. For this, let $w \in A^*$ be such that $t' \rightarrow^{w} t$. Then $w \in RY$ and therefore $\phi(w) = e$. Now

$$\phi(ur_0) = \phi(u)\phi(r_0) = \phi(u)e = \phi(u)\phi(w) = \phi(uw).$$

But since $i \rightarrow^{uw} t$, one has $uw \in XY$. Hence $\phi(uw) = m$ and finally $\phi(ur_0) = m$. This completes the proof of the theorem. □

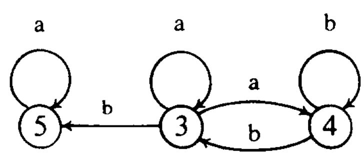
Example 3.3. The set

$$U = a^*bA^{\omega}$$

of Example 2.3 is the behaviour of the codeterministic automaton given below:

(1) (2)

with $I = \{1, 2, 3\}$, $T = \{2, 4\}$. The two components of this automaton correspond to the decomposition into simple sets of Example 2.3. Let us follow on this example the construction of the proof of Theorem 3.1. For the first one, one has $X_1 = A^*ba^*$, $Y_1 = a^*$, $R_1 = 1$ giving automaton (1). For the second one, one has $X_2 = Y_2 = A^*ba^*$, $R_2 = ba^*$, giving the automaton below:



with $I_2 = \{3, 4\}$, $T_2 = \{4\}$, $t_2 = 5$. This gives automaton (2) after removing state 5, which is not coaccessible from state 4.

References

[1] S. Eilenberg, Automata, Languages and Machines, Vol. A (Academic Press, New York, 1974).

[2] M. Nivat and D. Perrin, Ensembles reconnaissables de mots biinfinis, Proc. 14th ACM Symp. on Theory of Computing (1982) 47–59.

[3] A.W. Mostowski, Determinacy of sinking automata on infinite trees and inequalities between various pairs of Rabin's indices, Inform. Process. Lett. 15 (1982) 159–163.