INFORMATION AND CONTROL **50**, 276–284 (1981)

On the Equivalence Problem for Binary DOL Systems

JUHANI KARHUMÄKI

Department of Mathematics, University of Turku, Turku, Finland

It is shown that to test whether two DOL sequences in the binary case coincide it is enough to test whether four first words of these sequences are the same. The result is optimal.

1. INTRODUCTION

For several years the DOL equivalence problem was one of the most interesting open problems within the theory of formal languages. The problem is as follows. Given two morphisms h and g of a finitely generated free monoid Σ^* and an element ω of Σ^*. Does there exist an algorithm to decide whether or not the equation $h^n(\omega) = g^n(\omega)$ holds true for all $n \geqslant 0$?

The problem was solved positively by Culik and Fris (1977). Later a shorter proof was given by Ehrenfeucht and Rozenberg (1978). Moreover, from the arguments of Ehrenfeucht and Rozenberg (1978) it was deduced in Ehrenfeucht and Rozenberg (1980) an explicit bound n_0 depending on the two systems such that if the two sequences coincide up to the level n_0, then they will coincide forever.

There are no known examples of two nonequivalent DOL systems such that their sequences would differ from each other for first time "far" from the beginning. This situation has led to the following $2n$-conjecture; see Salomaa (1978b): For two DOL systems over an n-letter alphabet, to test the equivalence of these systems it is enough to test whether $2n$ first words of the sequences are the same.

It is known that $2n$ would be close to optimal. Indeed, there are examples showing that $[(3/2)\,n]$ is not enough. The gap between the n_0 given in Ehrenfeucht and Rozenberg (1980) and $2n$ is huge.

Our purpose here is to fill this gap in the case of binary DOL systems. We shall prove the $2n$-conjecture in this case. It follows from the known example, see Nielsen (1974), that our bound is optimal. Our proof is based on a characterization of equality languages of binary morphisms given in Ehrenfeucht *et al.* (1981).

Our approach gives also solutions to some related problems concerning

0019-9958/81/090276–09$02.00/0

DOL and DTOL systems. For instance, we show that the $2n$-conjecture, interpreted in a natural way, holds true for DTOL systems over the binary alphabet, too.

2. Preliminaries

In this note we need only very basic terminology of the theory of formal languages and the theory of free monoids. For few unexplained notions we refer to Harrison (1978). More background material concerning DOL systems can be found in Rozenberg and Salomaa (1980).

A free monoid generated by a finite alphabet Σ is denoted by Σ^* and its identity, so-called empty word, by λ. Elements of Σ^* are words. For a word x the notation $|x|$ denotes its length and the notation $\mathrm{pref}_k(x)$, for $k \geqslant 1$, its prefix of the length k. If $|x| < k$, we set $\mathrm{pref}_k(x) = x$. For a word x in Σ^* and a letter c in Σ, $\#_c(x)$ denotes the number of c's in x. In the case of the binary alphabet $\{0, 1\}$ the *ratio* $r(x)$ of a nonempty word x is defined as $\#_0(x){:}\,\#_1(x)$. We call such a word *ratio-primitive* if none of its prefixes has the same ratio as the whole word. By a *primitive* word we mean, as usual, a nonempty word x which is not a proper power of any word, i.e., the relation $x = z^n$ implies that $x = z$ and $n = 1$. Finally, for two words x and y the notation $x^{-1}y$ (resp. yx^{-1}) is used to denote the left (resp. right) quotient of y by x.

A *DOL system* is a triple (Σ, h, ω) where Σ is a finite alphabet, h is a morphism from Σ^* into itself and ω is a nonempty word of Σ^*. A DOL system $G = (\Sigma, h, \omega)$ defines the sequence

$$\omega, h(\omega), h^2(\omega), \ldots$$

Such a sequence (resp. set of words) is called a *DOL sequence* (resp. *DOL language*) or a *DOL sequence* (resp. *DOL language*) *generated by G*. We call two DOL systems *equivalent* if they generate the same DOL sequence. The *DOL sequence equivalence problem* is the problem of whether there exists or not an algorithm to decide the equivalence of two given DOL systems.

In this paper we consider only the case when Σ is binary, say $\{0, 1\}$. We call a morphism h *periodic* if there exists a word p such that $h(\Sigma) \subseteq p^*$. The set $\{\alpha, \beta\}$ of two words is called *marked* if $\mathrm{pref}_1(\alpha) \neq \mathrm{pref}_1(\beta)$. Let h be a nonperiodic morphism on $\{0, 1\}^*$. It is well known that $h(01) \neq h(10)$. Let z_h be the maximal common prefix of $h(01)$ and $h(10)$. Consequently, $|z_h| < |h(01)|$. Now the following result is easy to see.

Lemma 1. *For any word $x \in \Sigma^*0\Sigma^* \cap \Sigma^*1\Sigma^*$, $h(x)$ has the prefix z_h.*

If x *and* y *are two words such that* $x, y \in \Sigma^*0\Sigma^* \cap \Sigma^*1\Sigma^*$ *and* $\mathrm{pref}_1(x) \neq \mathrm{pref}_1(y)$, *then* $\mathrm{pref}_{|z_h|+1}(h(x)) \neq \mathrm{pref}_{|z_h|+1}(h(y))$.

Let h and g be two morphisms on Σ^*. Following Salomaa (1978a) we define the *equality set* of the pair (h, g), in symbols $E(h, g)$, by

$$E(h, g) = \{x \in \Sigma^* \mid h(x) = g(x)\}.$$

A basic property of binary equality sets is as follows.

LEMMA 2. *For a given equality set over a binary alphabet, all of its nonempty words have the same ratio.*

From the arguments in Ehrenfeucht *et al.* (1981) the following characterization for the equality sets over a binary alphabet can be derived.

THEOREM 1. *For a pair* (h, g) *of binary morphisms such that at least one of them is injective the equality set* $E(h, g)$ *is one of the following forms*:

(i) $\{u, v\}^*$ *for some (possibly empty) words* u *and* v,

(ii) $\{uw^*v\}^*$ *for some nonempty words* u, w *and* v *satisfying* w, uw^iv, *for* $i \geqslant 0$, *and* vu *are ratio-primitive*, $\mathrm{pref}_1(w) \neq \mathrm{pref}_1(v)$ *and* $w \in \Sigma^*0\Sigma^* \cap \Sigma^*1\Sigma^*$.

Finally, we say that two morphisms h and g *agree on a word* x if $h(x) = g(x)$ and that they *agree on a language* L if they agree on each word of L, i.e., $L \subseteq E(h, g)$.

3. MAIN RESULT

Here we prove our main result.

THEOREM 2. *Let* $H = (\{0, 1\}, h, \omega)$ *and* $G = (\{0, 1\}, g, \omega)$ *be two DOL systems. The following conditions are equivalent*:

(i) *H and G are equivalent*,

(ii) $h^i(\omega) = g^i(\omega)$ *for* $i = 0, 1, 2, 3$.

Proof. Clearly, (i) implies (ii). So we assume that (ii) holds true, and we shall prove (i).

If both h and g are periodic, then the result is easily seen to hold. So let, e.g., h be nonperiodic, in other words injective. By Theorem 1, we have two cases.

(I) $E(h, g) = \{u, v\}^*$ for some (possibly empty) words u and v. Now, our assumption implies

$$\omega, h(\omega), h^2(\omega) \in \{u, v\}^*. \tag{1}$$

From this and Lemma 1 it follows that $r(\omega) = r(h(\omega))$. Consequently, for any word x such that $r(x) = r(\omega)$, we have $r(x) = r(h(x))$, i.e., h preserves the "correct ratio," Since $\omega \in \{u, v\}^*$, we have, e.g., $\omega = uz$ for some word z. Hence, $h(\omega) = h(u)\, h(z)$ where $r(h(u)) = r(\omega)$, and so, by the fact $h(\omega) \in E(h, g)$, we obtain that $h(u) \in \{u, v\}^*$.

If both ω and $h(\omega)$ are in u^* we are done. Indeed, in this case $h^n(\omega) \in u^* \subseteq E(h, g)$ for all $n \geqslant 0$. So assume that $\omega \in \{u, v\}^*\, v\{u, v\}^*$ or $h(\omega) \in \{u, v\}^*\, v\{u, v\}^*$. In the first case we obtain, as above, that $h(v) \in \{u, v\}^*$. The same conclusion can be drawn also in the second case when only the fact $h^2(\omega) \in \{u, v\}^*$ is used. Consequently, we have also now that $h^n(\omega) \in \{u, v\}^*$ for all $n \geqslant 0$. This completes the proof of case I.

(II) $E(h, g) = \{uw^*v\}^*$ for some nonempty words u, w and v. Moreover, u, w and v satisfy the conditions: w, uw^iv, for $i \geqslant 0$, and vu are ratio-primitive, $\mathrm{pref}_1(w) \neq \mathrm{pref}_1(v)$ and w contains both 0 and 1 as a subword. Since h is nonperiodic we set, as earlier, z_h to be the maximal common prefix of $h(01)$ and $h(10)$. We define

$$\begin{aligned} \alpha &= z_h && \text{if} \quad |z_h| < |h(v)| \\ &= h(v) && \text{if} \quad |z_h| \geqslant |h(v)|. \end{aligned} \tag{2}$$

Now, by Lemma 1, α is a prefix of both $h(w)$ and $h(v)$, and moreover $\mathrm{pref}_{|\alpha|+1}(h(w)) \neq \mathrm{pref}_{|\alpha|+1}(h(v))$.

Let us recall our assumption

$$\omega, h(\omega), h^2(\omega) \in \{uw^*v\}^*. \tag{3}$$

If ω and $h(\omega)$ are both in $\{uw^iv\}^*$, for some i, we are done, the reasoning being as in case I. In the other case there exist integers i and j, with $i > j$, such that $\{\omega, h(\omega)\}$ contains a word both from $\{uw^*v\}^*\, uw^iv\{uw^*v\}^*$ and from $\{uw^*v\}^*\, uw^jv\{uw^*v\}^*$.

By our assumption, $h(uw^iv)$ and $h(uw^jv)$ are in $\{uw^*v\}^*$. Let $C = \{w, vu\}$. Then C is marked and, therefore, there exists a unique word γ in C^* such that

$$u\gamma = h(uw^j)\, \alpha y$$

for some word y not containing either w or vu as a suffix. We claim that $y = \lambda$ or $y = u$. This follows since $h(uw^iv)$ and $h(uw^jv)$ are both in uC^*v, C

is marked and $\mathrm{pref}_{|\alpha|+1}(h(w)) \neq \mathrm{pref}_{|\alpha|+1}(h(v))$. Indeed, if in (2), $|h(v)| > |\alpha|$, then $y = \lambda$, and if $|h(v)| = |\alpha|$, then $y = u$. Consequently, we have either

$$\begin{aligned} h(uw^j)\,\alpha &\in uC^*, \\ \alpha^{-1}h(v) &\in C^*v, \\ \alpha^{-1}h(w^{i-j})\,\alpha &\in C^* \end{aligned} \tag{4}$$

or

$$\begin{aligned} h(uw^j)\,\alpha &\in uC^*v, \\ \alpha^{-1}h(v) &= \lambda, \\ \alpha^{-1}h(w^{i-j})\,\alpha &\in uC^*v. \end{aligned} \tag{4'}$$

Now, we look at the third relation of (4) in detail. By Lemma 2 and the form of $E(h, g)$, we have $r(w) = r(vu) = r(\omega)$. Further, as shown in case I, h preserves the "correct ratio." Therefore, $r(\alpha^{-1}h(w)\,\alpha) = r(\omega)$. So it follows from the ratio-primitivinesses of w and vu and from the third relation of (4) that

$$\alpha^{-1}h(w)\,\alpha \in C^*. \tag{5}$$

This, in turn, applied to the first relation of (4) yields

$$h(u)\,\alpha \in uC^*. \tag{6}$$

Here the fact that C is a code is needed.

In the case (4′) similar arguments can be used. Then the facts that the words uw^iv, for $i \geqslant 0$, are ratio-primitive and that also the set $\{uw^iv \mid i \geqslant 0\}$ is a code yield

$$\alpha^{-1}h(w)\,\alpha \in uC^*v \tag{5'}$$

and

$$h(u)\,\alpha \in uC^*v. \tag{6'}$$

Now, we are ready to finish this proof. Indeed, by (4), (5) and (6) or alternatively (4′), (5′) and (6′), we obtain

$$h(uw^*v) \subseteq uC^*v = \{uw^*v\}^* = E(h, g),$$

which together with $\omega \in E(h, g)$ implies (i).

Next we recall an example, due to Nielsen (1974), which shows that our

Theorem 2 is optimal. Let H and G be DOL systems with the starting word ab and the morphisms h and g defined by

$$h(a) = abb, \qquad g(a) = abbaabb,$$
$$h(b) = aabba, \qquad g(b) = a.$$

Then,

$$h^0(ab) = ab = g^0(ab)$$
$$h(ab) = abbaabba = g(ab)$$
$$h^2(ab) = (abbaabbaaabbaabb)^2 = g^2(ab)$$

and

$$\mathrm{suff}_2(h^3(ab)) = ba \neq aa = \mathrm{suff}_2(g^3(ab)),$$

where the notation suff_2 denotes the suffix of length 2.

It is instructive to consider the above example in the light of equality sets. Clearly, $E(h, g) = \{ab, ba\}^*$. Since the starting word ab belongs to $E(h, g)$ we must have $h(ab) = g(ab)$. Moreover, we have $h(ab) \in E(h, g)$. So it follows that $h(h(ab)) = g(h(ab)) = g(g(ab))$. But now this word is no longer in $E(h, g)$ since $\mathrm{suff}_2(h^2(ab)) = bb$ and so we can at once conclude that $h^3(ab) \neq g^3(ab)$.

4. Generalizations

In this section we discuss the generalizations of Theorem 2. The equivalence of two DOL systems, with morphisms h and g, can be interpreted as "morphisms h and g agree on the DOL language generated by one of the systems." So an obvious generalization is to allow that the morphism of the DOL system is different from h and g. We have the result.

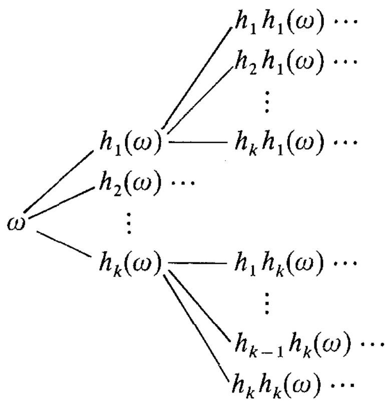
Theorem 3. *Let h and g be morphisms $\{0, 1\}^* \to \{0, 1\}^*$ and $G = (\{0, 1\}, f, \omega)$ a DOL system. Then the following conditions are equivalent*:

(i) *h and g agree on the language generated by G,*

(ii) *$h(f^i(\omega)) = g(f^i(\omega))$ for $i = 0, 1, 2, 3$.*

Proof. The proof of Theorem 2 is valid also now after one observation: In that proof it is unnecessary to require that the morphism h in (1) and (3) is one of those used in the considered equality sets.

Another way of generalizing Theorem 2 is to consider so-called DTOL systems, cf. Rozenberg and Salomaa (1980). A *DTOL system* is a $(k+2)$-tuple $(\Sigma, h_1, \ldots, h_k, \omega)$ where $k \geqslant 1$ and each of the triples (Σ, h_j, ω) is a DOL system. A DTOL system $(\Sigma, h_1, \ldots, h_k, \omega)$ generates a *tree* of words as follows:



The set of all words in this tree is called the *DTOL language generated by G*. As in the case of DOL systems we call two DTOL systems $(\Sigma, h_1, \ldots, h_k, \omega)$ and $(\Sigma, g_1, \ldots, g_k, \omega)$ *equivalent* if they generate the same trees of words, i.e., if $h_{i_1} \cdots h_{i_s}(\omega) = g_{i_1} \cdots g_{i_s}(\omega)$ holds true for all $s \geqslant 0$ and $i_j \in \{1, \ldots, k\}$.

Theorem 2 can be generalized to

THEOREM 4. *Let $H = (\{0, 1\}, h_1, \ldots, h_k, \omega)$ and $G = (\{0, 1\}, g_1, \ldots, g_k, \omega)$ be two DTOL systems. Then the following conditions are equivalent*:

(i) *H and G are equivalent,*

(ii) $h_{i_1} \cdots h_{i_s}(\omega) = g_{i_1} \cdots g_{i_s}(\omega)$ *for* $s \leqslant 3$ *and* $i_j \in \{1, \ldots, k\}$.

Proof. Again the proof is basically that of Theorem 2. Indeed, as there, we can derive from the assumption (ii) that either all words of the language generated by H are in u^* for some word u or for each i and j in $\{1, \ldots, k\}$

$$h_j(E(h_i, g_i)) \subseteq E(h_i, g_i)$$

which together with the assumption $\omega \in \bigcap_{i=1}^{k} E(h_i, g_i)$ implies the result.

Theorem 4 shows that the equivalence problem for binary DTOL trees (sequences) is decidable—a result which, as far as I know, is not explicitly mentioned anywere, but which can be easily derived from the main theorem of Culik and Richier (1979). Although our algorithm for deciding the equivalence of two DTOL trees is very simple, the problem is, in general, and even in a three-letter case, still open. On the other hand, the problem of whether two DTOL systems generate the same language is shown to be undecidable in Rozenberg (1972).

We also have the following generalization of Theorem 3.

THEOREM 5. *Let h and g be morphisms of $\{0, 1\}^*$ and $G = (\{0, 1\}, f_1, \ldots, f_k, \omega)$ a DTOL system. Then the following conditions are equivalent*:

(i) *h and g agree on the language generated by G,*

(ii) *$h(f_{i_1} \cdots f_{i_k}(\omega)) = g(f_{i_1} \cdots f_{i_k}(\omega))$ for all $s \leqslant 3$ and $i_j \in \{1, \ldots, k\}$.*

Proof. Now the observation of the proof of Theorem 3 is valid. Indeed, we obtain along the lines of the proof of Theorem 2 that (ii) implies that either all words of the language generated by G are in u^* for some word u or

$$f_j(E(h, g)) \subseteq E(h, g) \qquad \text{for} \quad j = 1, \ldots, k$$

which together with the fact $\omega \in E(h, g)$ yields (i).

Theorems 3 and 5 give simple solutions to the problems which are referred to as morphism equivalence problems for binary DOL and DTOL languages. To be precise, such problems are as follows, cf. Culik and Salomaa (1979): The *morphism equivalence problem* for the family $\mathscr{L}$ of languages is to decide whether for a given language L in $\mathscr{L}$ and for two morphisms h and g, h and g agree on L. Culik and Richier showed that this problem is decidable for ETOL languages, cf. Rozenberg and Salomaa (1980), over a binary alphabet. Our Theorems 3 and 5 give considerably simpler algorithms for some subfamilies of this family, namely, for DOL and DTOL languages over a binary alphabet.

ACKNOWLEDGMENTS

The author is grateful to Dr. M. Linna for useful comments and to the Academy of Finland for the excellent working conditions under which this research was carried out.

RECEIVED: January 8, 1982.

REFERENCES

CULIK, K., II AND FRIS, I. (1977), The decidability of the equivalence problem for DOL-systems, *Inform. Contr.* **35**, 20–39.

CULIK, K., II AND RICHIER, J. L. (1979), Homomorphism equivalence on ETOL languages, *Internat. J. Comput Math.* **7**, 43–51.

CULIK, K., II AND SALOMAA, A. (1979), On the decidability of homomorhism equivalence for languages, *J. Comput. Systems Sci.* **17**, 163–175.

EHRENFEUCHT, A., KARHUMÄKI, J., AND ROZENBERG, G. (1981), On binary equality sets and a solution to the Test Set Conjecture in the binary case, *J. Algebra* (to appear).

EHRENFEUCHT, A., AND ROZENBERG, G. (1978), Elementary homomorphisms and a solution to the DOL sequence equivalence problem, *Theoret. Comput. Sci.* **7**, 169–183.

EHRENFEUCHT, A., AND ROZENBERG, G. (1980), On a bound for the DOL sequence equivalence problem, *Theoret. Comput. Sci.* **12**, 339–342.

HARRISON, M. (1978), "Introduction to Formal Language Theory," Addison-Wesley, Reading, Mass.

NIELSEN, M. (1974), On the decidability of some equivalence problems for DOL systems, *Inform. Contr.* **25**, 166–193.

ROZENBERG, G. (1972), The equivalence problem for deterministic TOL systems is undecidable, *Inform. Process. Lett.* **1**, 201–204.

ROZENBERG, G., AND SALOMAA, A. (1980), "The Mathematical Theory of *L* Systems," Academic Press, New York.

SALOMAA, A. (1978a), Equality sets for homomorphisms of free monoids, *Acta Cybernet.* **4**, 127–139.

SALOMAA, A. (1978b), DOL equivalence: The problem of iterated morphisms, *E.A.T.C.S. Bull.* **4**, 5–12.