ON A BOUND FOR THE DOL SEQUENCE EQUIVALENCE PROBLEM

by

Andrzej Ehrenfeucht*

and

Grzegorz Rozenberg**

#CU-CS-163-79 August, 1979

*A. Ehrenfeucht, Department of Computer Science, University of Colorado at Boulder, 80309 USA

**G. Rozenberg, Institute of Applied Mathematics and Computer Science, University of Leiden, 2300 RA Leiden, Holland

All correspondence to G. Rozenberg.

ABSTRACT

An explicit bound is given for a solution of the DOL sequence equivalence problem; that is, given two DOL systems G_1 and G_2, we compute explicitly a positive integer constant $C(G_1,G_2)$ such that the sequences generated by G_1 and G_2 are equal if and only if the corresponding first $C(G_1,G_2)$ elements of the sequences are equal.

INTRODUCTION

The DOL sequence equivalence problem was first solved in [CF]. In [ER] a considerably simplier solution of this problem is presented. In this note we show that the solution method from [ER] allows one to compute the explicit bound on how long one has to compare two DOL sequences before deciding whether or not they are equal. It seems that the method from [CF] does not allow one to compute explicitly such a bound.

This note is a companion of [ER] in the sense that it assumes the thorough familiarity with [ER] and it uses the terminology and the notation from [ER]. The only additional notation we need is that of <u>maxr</u> h which for a homomorphism h on Σ^* stands for $\max\{|\alpha| : h(b) = \alpha \text{ for a letter } b \text{ in } \Sigma\}$.

THE RESULT

First of all we need the following lemma.

Lemma. Let $G_1 = (\Sigma, h_1, \omega)$ and $G_2 = (\Sigma, h_2, \omega)$ be two elementary DOL systems with $E(G_1) = \omega_0^{(1)}, \omega_1^{(1)}, \ldots$ and $E(G_2) = \omega_0^{(2)}, \omega_1^{(2)}, \ldots$. Let $n = \#\Sigma$, $m = \max\{\underline{\max r}\, h_1, \underline{\max r}\, h_2\}$, $s = |\omega|$, r be the constant separating short states from long states in A_{h_1,h_2} (it is called r_0 in point (iii) of the proof of Lemma 3 in [ER]) and let $q = 2 \cdot n^r$. Let ψ be a function from N^+ into N^+ defined by $\Psi(x) = s \cdot m^{x^{n \cdot x}} + x$. Then $E(G_1) = E(G_2)$ if and only if $\omega_i^{(1)} = \omega_i^{(2)}$ for $0 \leq i \leq u$ where $u = (\Psi^q(q))^{n \cdot \Psi^q(q)}$.

Proof.

It suffices to show that if $\omega_u^{(1)} = \omega_u^{(2)}$ then for every $v \geq u$, $\omega_v^{(1)} = \omega_v^{(2)}$.

Let $B_1, B_2, \ldots$ be a sequence approximating A_{h_1,h_2} defined as follows (recall that $A_{h_1,h_2} = A_{h_1,h_2}^{(1)}, A_{h_1,h_2}^{(2)}, \ldots$).

(1). $B_1 = A_{h_1,h_2}^{(1)}$, i.e., it consists of all the short states of A_{h_1,h_2}.

(2). Let $\omega_{k_1}^{(1)}$ be the first element of $E(G_1)$ that does not belong to $T(B_1)$. Assume that $\omega_{k_1}^{(1)} \in \mathrm{MID}(h_1,h_2)$ and let B_2 results from B_1 by adding to it a path from A_{h_1,h_2} traversed by $\omega_{k_1}^{(1)}$.

(3). In general, for $i \geq 1$, let $\omega_{k_i}^{(1)}$ be the first element of $E(G_1)$ which do not belong to $T(B_i)$ and let (assuming that $\omega_{k_i}^{(1)} \in \mathrm{MID}(h_1,h_2)$) B_{i+1} results from B_i by adding to it a path from A_{h_1,h_2} traversed by $\omega_{k_i}^{(1)}$.

If $E(G_1) = E(G_2)$ then there exists an ℓ such that $L(G_1) = T(B_\ell)$; in the "worst case" B_ℓ will be such that $T(b_\ell) = MID(h_1, h_2)$. Let us estimate ℓ first.

(i). $\ell \leq 2 \cdot n^r$.

Proof of (i).

Note that in A_{h_1, h_2} each short state which has a transition to a long state and can lead back to a short state gives rise to a unique trace (path) in A_{h_1, h_2}. Hence the number of these paths is bounded by the number of short states which is clearly bounded by $2 \cdot n^r$. But each B_{i+1} results from B_i by adding one such path to B_i. Hence the bound.

Let $d(B_i)$ denote the number of states in B_i.

(ii). $d(B_1) \leq 2 \cdot n^r$.

Proof of (ii).

Obvious.

(iii). $k_i \leq d(B_i)^{n \cdot d(B_i)}$

Proof of (iii).

Let us consider the sequence of automata $\tau_i = B_i, (B_i)_h, (B_i)_{h^2}, \ldots$ (see the notation from the proof of Theorem 6 in [ER]). Then k_i is the smallest integer j such that $\omega \notin T((B_i)_{h_1^j})$. However, the number of all different automata in the sequence τ_i cannot exceed $d(B_i)^{n \cdot d(B_i)}$ and hence the bound.

(iv). $d(B_{i+1}) \leq s \cdot m^{d(B_i)^{n \cdot d(B_i)}} + d(B_i)$.

Proof of (iv).

Going from B_i to B_{i+1} we get ω_{k_i} to be in $T(B_{i+1})$. In other words ω_{k_i} gives rise to a trace starting with q_{in} then coming to a short state from which a unique path leads through the sequence of long states (which were not in B_i !!) and come back to a short state (which was in B_i). Thus we added no more than $|\omega_{k_i}|$ new states to B_i to obtain B_{i+1}. Hence $d(B_{i+1}) \le |\omega_{k_i}| + d(B_i)$. But $|\omega_{k_i}| \le s \cdot m^{k_i}$ and so $d(B_{i+1}) \le s \cdot m^{k_i} + d(B_i)$.

Combining this with (iii) we get

$$d(B_{i+1}) \le s \cdot m^{d(B_i)^{n \cdot d(B_i)}} + d(B_i).$$

(v). Now we complete the proof of the lemma as follows.

From (iv) we have that $d(B_{i+1}) \le \Psi(d(B_i))$, hence from (i) it follows that $d(B_\ell) \le \Psi^{2 \cdot n^r}(2 \cdot n^r)$.

Then from (iii) it follows that $k_\ell \le (\Psi^{2 \cdot n^r}(2 \cdot n^r))^{n \cdot \Psi^{2 \cdot n^r}(2 \cdot n^r)}$.

But clearly $\omega^{(1)}_{k_{\ell-1}} \in T(B_\ell)$ if and only if $\omega^{(1)}_{k_{\ell-1}} \in T(A_{h_1,h_2})$ if and only if for every $v \ge k_{\ell-1}$, $\omega^{(1)}_v \in T(A_{h_1,h_2})$ if and only if for every $v \ge k_\ell$, $\omega^{(1)}_v = \omega^{(2)}_v$.

Hence the lemma holds. □

<u>Theorem</u>. Let $G_1 = (\Sigma, h_1, \omega)$ and $G_2 = (\Sigma, h_2, \omega)$ be two DOL systems with $E(G_1) = \omega^{(1)}_0, \omega^{(1)}_1 \ldots$ and $E(G_2) = \omega^{(2)}_0, \omega^{(2)}_1, \ldots$. Let

$n = \#\Sigma$, $\overline{m} = \max\{(\underline{\text{maxr}}\ h_1)^{2^n}, (\underline{\text{maxr}}\ h_2)^{2^n}\}$,

$\overline{s} = \max\{|h_1^{2^n}(\omega)|, |h_2^{2^n}(\omega)|\}$, $\overline{r}$ is defined analogously to the way that r is defined in the statement of Lemma 1 except that one replaces h_1 by $h_1^{2^n}$ and h_2 by $h_2^{2^n}$ and let $\overline{q} = 2 \cdot n^{\overline{r}}$. Let $\overline{\Psi}$ be a function from $\mathbf{N}^+$ into $\mathbf{N}^+$ defined by $\overline{\Psi}(x) = \overline{s} \cdot \overline{m}^{x^{n \cdot x}} + x$. Then $E(G_1) = E(G_2)$ if and only if $\omega^{(1)}_i = \omega^{(2)}_i$ for $0 \le i \le u$ where

$u = (\overline{\Psi}^{\overline{q}}(\overline{q}))^n \cdot \overline{\Psi}^{\overline{q}}(\overline{q}) + 2^{n+1}$.

Proof.

This follows directly from our Lemma 1, Theorem 6.2 and its proof in [ER] and from Lemma 7.1 in [ER]. □

Remark. Clearly, we realize that the explicit bound we provide is "awfully big." The aim of this note was to show that the method of solving the DOL equivalence problems presented in [ER] allows one to compute such a bound. One should remind here that the common conjecture is that such a bound equals 2n where n is the size of the alphabet involved (see, e.g., [S]).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of NSF grant number MCS79-03838.

REFERENCES

[CF]. K. Culik II and I. Fris, The decidability of the equivalence problem for DOL systems, Information and Control 35 (1977) 20-39.

[ER]. A. Ehrenfeucht and G. Rozenberg, Elementary homomorphisms and a solution of the DOL sequence equivalence problem, Theoretical Computer Science 7 (1978) 169-183.

[S]. A. Salomaa, DOL equivalence: the problem of iterated morphisms, Bulletin of the European Association for Theoretical Computer Science, No. 4, January 1978.