INFORMATIQUE THÉORIQUE ET APPLICATIONS

VESA HALAVA
TERO HARJU

Undecidability of the equivalence of finite substitutions on regular language

Informatique théorique et applications, tome 33, n° 2 (1999), p. 117-124

<http://www.numdam.org/item?id=ITA_1999__33_2_117_0>

Theoretical Informatics and Applications
Theoret. Informatics Appl. **33** (1999) 117–124

UNDECIDABILITY OF THE EQUIVALENCE OF FINITE SUBSTITUTIONS ON REGULAR LANGUAGE

VESA HALAVA[1] AND TERO HARJU[2]

Abstract. A simplified proof is given for the following result due to Lisovik: it is undecidable for two given ε–free finite substitutions, whether they are equivalent on the regular language $b\{0,1\}^*c$.

1. INTRODUCTION

We give a simplified proof for the result due to Lisovik [4]. We also present an undecidability result considering whether a periodic morphism is included in finite substitution on regular language.

Let Σ and Δ be two finite alphabets and denote by ε the empty word. A mapping $\varphi : \Sigma^* \to 2^{\Delta^*}$, where 2^{Δ^*} denotes the power set of Δ^*, is called *substitution* if

1. $\varphi(\varepsilon) = \{\varepsilon\}$ and
2. $\varphi(xy) = \varphi(x)\varphi(y)$ for all x and y in Σ^*.

These conditions mean that a substitution is a monoid morphism from Σ^* into 2^{Δ^*}. A substitution φ is called *ε–free* if $\varepsilon \notin \varphi(a)$ for all $a \in \Sigma$, and *finite* if, for all $a \in \Sigma$, the set $\varphi(a)$ is a finite subset of Δ^*.

Let $L \subseteq \Sigma^*$ be a language. Two substitutions $\varphi, \xi : \Sigma^* \to 2^{\Delta^*}$ are *equivalent* on L if for all $w \in L$, $\varphi(w) = \xi(w)$.

We shall concentrate on ε–free finite substitutions and prove that the equivalence of two such substitutions on the regular language $b\{0,1\}^*c$ is undecidable. The undecidability of the equivalence problem is not trivial, since it is known that there are decidable special cases. For example, for finite prefix substitutions the equivalence problem on regular language is known to be

[1] Turku Centre for Computer Science TUCS, Lemminkäisenkatu 14 A, FIN-20520 Turku, Finland; e-mail: vehalava@cs.utu.fi

[2] Department of Mathematics, University of Turku, FIN-20014 Turku, Finland; e-mail: harju@utu.fi

decidable, see Karhumäki [2]. This follows from the fact that the monoid of all prefix codes is free and from the pumping property of regular languages.

There also exists some other undecidability results on ε–free finite substitutions, see Turakainen [5].

In the undecidability proofs we use the undecidability of the universe problem for the *integer weighted finite automata.*

Consider a (nondeterministic) finite automaton $\mathcal{A} = (Q, \Sigma, \delta, q_0)$ *without final states* with the states Q, the alphabet Σ, the transition function $\delta\colon Q \times \Sigma \to 2^Q$, and the initial state q_0. A transition $p \in \delta(q, a)$, where $q, p \in Q$ and $a \in \Sigma$, will also be written as (q, a, p) (in which case $\delta \subseteq Q \times \Sigma \times Q$ is regarded as a relation and sometimes also as an alphabet). Note that we allow only transitions that read a single letter.

Let $(\mathbb{Z}, +, 0)$ be the additive group of integers. An integer weighted finite automaton $\mathcal{A}^\gamma$ consists of a finite automaton $\mathcal{A}$ as above, and a *weight function* $\gamma\colon \delta \to \mathbb{Z}$ of the transitions. To simplify the notation, we will write

$$(p, \gamma(t)) \in \delta(q, a)$$

for each transition $t = (q, a, p) \in \delta$.

Let $\pi = t_0 t_1 \dots t_n$ be a path of $\mathcal{A}$, where $t_i = (q_i, a_i, q_{i+1})$ for $0 \le i \le n-1$. Define a morphism $\|\cdot\|\colon \delta^* \to \Sigma^*$ by setting $\|(q, a, p)\| = a$. The weight of the path π is the element

$$\gamma(\pi) = \gamma(t_0) + \gamma(t_1) + \cdots + \gamma(t_n)\,.$$

Further, we let $L(\mathcal{A}^\gamma) = \|\gamma^{-1}(0)\|$, that is

$$L(\mathcal{A}^\gamma) = \{w \in \Sigma^* \mid w = \|\pi\|,\ \gamma(\pi) = 0\}\,,$$

be the language *accepted by* $\mathcal{A}^\gamma$. In other words, a word $w \in \Sigma^*$ is accepted by $\mathcal{A}^\gamma$ if there is a path t reading w and having the weight 0. An integer weighted finite automata is denoted by $\mathrm{FA}(\mathbb{Z})$.

Next lemma is from [1].

Lemma 1.1. *The universe problem, asking whether $L(\mathcal{A}^\gamma)$ is the universal language Σ^*, is undecidable.*

Actually, in [1] it was proved that the universe problem is undecidable for a rather restricted type of 4-state integer weighted finite automata, if we allow transitions to have finitely many different weights *i.e.*, γ is a function from δ into $2^{\mathbb{Z}}$.

We note that the model of finite automata defined above is closely related to the counter automata. In our model the counter is replaced by a weight function of the transitions, and while doing so, the finite automaton becomes independent of the counter.

The next corollary follows from Lemma 1.1. We shall use it in the proof of the main result.

Corollary 1.2. *The universe problem is undecidable for an* FA($\mathbb{Z}$) *over binary alphabet.*

Proof. We shall construct an FA($\mathbb{Z}$) $\mathcal{B}^\gamma$ over a binary alphabet $\{a,b\}$ from $\mathcal{A}^\gamma$ over an alphabet Σ such that $L(\mathcal{B}^\gamma) = \{a,b\}^*$ if and only if $L(\mathcal{A}^\gamma) = \Sigma^*$.

First we encode the alphabet $\Sigma = \{a_1, \ldots, a_n\}$ to a binary alphabet. One such encoding is provided by $\psi(a_i) = a^i b$ for $a_i \in A$.

Each transition in $\mathcal{A}^\gamma$ is divided in a way that, for all $q \in Q$, we have new states $q^1, q^2, \ldots, q^n$ and a transition function δ' of $\mathcal{B}^\gamma$ is defined so that $\delta'(q,a) = \{(q^1,0)\}$, $\delta'(q^k,a) = \{(q^{k+1},0)\}$ for $1 \leq k \leq n-1$, and, for all $1 \leq i \leq n$, $\delta'(q^i,b) = \{(p,z) \mid (p,z) \in \delta(q,a_i)\}$. Clearly $\mathcal{B}^\gamma$ accepts a word $w \in \psi(A^*)$ if and only if w is accepted in $\mathcal{A}^\gamma$.

Finally, we add a part that accepts the words in the regular language $\{a,b\}^* \setminus \psi(\Sigma^*)$. Clearly each regular language is accepted by an FA($\mathbb{Z}$) and FA($\mathbb{Z}$) languages are closed under union. □

2. A SPECIAL INCLUSION PROBLEM

In this section we prove the undecidability of a special inclusion problem.

Fix alphabets $A = \{b,0,1,c\}$ and $B = \{a,b\}$. Let $h : A^+ \to B^+$ be a *periodic non-erasing* morphism, *i.e.* there exists $w \in B^+$ such that $h(u) \in w^+$ for all $u \in A^+$.

Theorem 2.1. *Let $h : A^+ \to B^+$ be a periodic morphism as above and $\chi : A^* \to 2^{B^*}$ be an ε-free finite substitution. It is undecidable, whether $h(u) \in \chi(u)$ for all $u \in b\{0,1\}^* c$.*

Proof. Let $\mathcal{A}^\gamma = (Q, \{0,1\}, \delta, q_1)$ be an FA($\mathbb{Z}$). We shall define h and χ such that $h(u) \in \chi(u)$ for all $u \in b\{0,1\}^*c$ if and only if $L(\mathcal{A}^\gamma) = \{0,1\}^*$.

As usual we shall identify a singleton set $\{u\}$ with its sole member u. Let $s = |Q|+1$, for $Q = \{q_1, \ldots, q_{s-1}\}$, where q_1 is the initial state of $\mathcal{A}^\gamma$. Further, let

$$r = |\min\{z \mid \exists (p,z) \in \delta(q,x) \text{ for some } p,q \in Q, x \in \{0,1\}\}|,$$

i.e. r is the absolute value of the minimal weight of the transitions in $\mathcal{A}^\gamma$. We write $\langle k,x,z,j\rangle$ for $(q_j,z) \in \delta(q_k,x)$. Define

$$w = a^s b, \qquad N = \bigcup_{k=1}^{s-1} a^k b, \quad \text{and} \quad T_x = \bigcup_{\langle k,x,z,j\rangle} a^{s-k} b w^{z+r} a^j. \tag{2.1}$$

Here T_x encodes the rules of $\mathcal{A}^\gamma$. Define

$$\chi(b) = a, \quad \chi(c) = Nw, \quad \chi(x) = T_x,$$

for $x \in \{0,1\}$, and

$$h(b) = w, \quad h(c) = w, \quad h(x) = w^{r+1}.$$

Let $L = b\{0,1\}^*c$. Clearly, for all $u \in \{0,1\}^*$, $h(buc) = w \cdot w^{|u|(r+1)} \cdot w$, and we shall prove that $w^{|u|(r+1)+2} \in \chi(buc)$ for $u \in \{0,1\}^*$ if and only if $u \in L(\mathcal{A}^\gamma)$.

Let $u = x_1 \dots x_m$, $x_i \in \{0,1\}$ for $1 \le i \le m$, and $v \in \chi(buc)$. We know that

$$v = a \cdot a^{s-k_1} b w^{z_1+r} a^{j_1} \cdot a^{s-k_2} b w^{z_2+r} a^{j_2} \cdots a^{s-k_m} b w^{z_m+r} a^{j_m} \cdot a^{\ell} b w,$$

for the sequence $\langle k_i, x_i, z_i, j_i \rangle$, $1 \le i \le m$.

Assume that $v \in w^+$. Then necessarily $k_1 = 1$, $k_{i+1} = j_i$, for $1 \le i \le m-1$, and $\ell = s - j_m$. This means that the sequence is a computation of u in $\mathcal{A}^\gamma$ starting from the initial state q_1 and ending in the state q_{j_m}. It follows that

$$v = w^{1+\sum_{i=1}^m (z_i+r+1)+1} = w^{\sum_{i=1}^m z_i + m(r+1)+2}.$$

Now $v = w^{|u|(r+1)+2}$ if and only if $\sum_{i=1}^m z_i = 0$. The latter condition is equivalent to saying that the sequence $\langle k_i, x_i, z_i, j_i \rangle$, $1 \le i \le m$, is an accepting computation of u in $\mathcal{A}^\gamma$.

The claim follows from the undecidability of the universe problem of the FA($\mathbb{Z}$) $\mathcal{A}^\gamma$. □

3. The equivalence problem

In this section we shall prove that the equivalence problem of finite substitutions is undecidable for the regular language $b\{0,1\}^*c$.

The proof we present modifies the proof of Lisovik [4]. In the original proof Lisovik used the undecidability of the reliability of *defense systems*, *cf.* [3]. The undecidability of the reliability of defense systems has a long proof using the undecidability of a special inclusion problem for *nondeterministic finite transducers* over alphabet $\{0,1\}^* \times c^*$, which also has a somewhat involved proof.

We shall use the undecidability of the universe problem for finite integer weighted automata over a binary alphabet. This model of an automaton is simple and the undecidability proofs for it are elementary.

Theorem 3.1. *The equivalence problem for ε-free finite substitutions on the regular language $b\{0,1\}^*c$ is undecidable.*

Proof. Let $\mathcal{A}^\gamma = (Q, \{0,1\}, \delta, q_1)$ be an FA($\mathbb{Z}$). We shall define two finite substitutions $\varphi, \xi : \{b,0,1,c\}^* \to \{a,b\}^*$ such that φ and ξ are equivalent on $b\{0,1\}^*c$ if and only if $L(\mathcal{A}^\gamma) = \{0,1\}^*$.

As in the proof of Theorem 2.1, let $s = |Q| + 1$, for $Q = \{q_1, \dots, q_{s-1}\}$, where q_1 is the initial state of $\mathcal{A}^\gamma$, and

$$r = |\min\{z \mid \exists (p,z) \in \delta(q,x) \text{ for some } p,q \in Q, x \in \{0,1\}\}|.$$

We write $\langle k, x, z, j\rangle$ for $(q_j, z) \in \delta(q_k, x)$, and define w, N and T_x, for $x \in \{0,1\}$, as in (2.1). Further, for $x \in \{0,1\}$, define

$$C_x = \bigcup_{\langle k,x,z,j\rangle} a^{s-k} b w^{z+r+1}.$$

Finally we define the substitutions ξ and φ,

$$\begin{aligned}
\xi(b) &= a \cup wN, \\
\xi(c) &= (\varepsilon \cup N)w, \\
\xi(x) &= w^{r+1} \cup (\varepsilon \cup N)(T_x \cup C_x N),
\end{aligned}$$

for $x \in \{0,1\}$, and

$$\varphi(b) = \xi(b) \cup w, \quad \varphi(c) = \xi(c), \quad \varphi(x) = \xi(x).$$

Let $L = b\{0,1\}^*c$. Clearly $\xi(v) \subseteq \varphi(v)$ for all $v \in L$, since $\xi(x) \subseteq \varphi(x)$ for all letters $x \in \{b, 0, 1, c\}$. Therefore to prove that the equivalence of ξ and φ is undecidable, we need to show that $\varphi(v) \subseteq \xi(v)$ for all $v \in L$ if and only if $L(\mathcal{A}^\gamma) = \{0,1\}^*$.

Assume first that $L(\mathcal{A}^\gamma) \neq \{0,1\}^*$. Let $u \in \{0,1\}^*$, $u \notin L(\mathcal{A}^\gamma)$, and assume that h and χ are as in the proof of Theorem 2.1. Clearly $h(buc) \in \varphi(buc)$ and $\chi(buc) \subseteq \xi(buc)$, and by the proof of Theorem 2.1, and the form of N, $h(buc) \in \xi(buc)$ if and only if $u \in L(\mathcal{A}^\gamma)$. By our assumption, $h(buc) = w^{|u|(r+1)+2} \neq \xi(buc)$. It follows that if $L(\mathcal{A}^\gamma) \neq \{0,1\}^*$, then $\varphi(x) \neq \xi(x)$ for some $x \in b\{0,1\}^*c$.

Assume next that $L(\mathcal{A}^\gamma) = \{0,1\}^*$ and let $x = x_0 \dots x_{n+1} \in L$, $u = u_0 \dots u_{n+1}$, where $x_i \in \{b, 0, 1, c\}$ and $u_i \in \varphi(x_i)$ for all $0 \leq i \leq n+1$. Note that $x_0 = b$ and $x_{n+1} = c$. We have to show that there exists $v_i \in \xi(x_i)$ for all $0 \leq i \leq n+1$ such that $v_0 \dots v_{n+1} = u$.

First, we note that the only difference in the images of ξ and φ is in the image of b, and $\xi(b) \setminus \varphi(b) = w$. If $u_0 \neq w$, then we have a trivial solution $u_i = v_i$ for all $0 \leq i \leq n+1$. Therefore we assume that $u_0 = w$.

We shall use parenthesis to illustrate the factorizations by φ and ξ to u_i's and v_i's, respectively. We divide our consideration into three cases:

(i) if $n = 0$, then $x = bc$ and we have two subcases:

(1) if $u_1 = w \in \varphi(c)$, then $u_0u_1 = (w)(w) = (a)(a^{s-1}bw) \in \xi(x)$.

(2) If $u_1 \in Nw \subseteq \varphi(c)$, *i.e.* for some $1 \leq k \leq s-1$, $u_0u_1 = (w)(a^{s-k}bw) = (wa^{s-k}b)(w) \in \xi(x)$.

(ii) If $n \geq 1$ and $u_1 \neq w^{r+1}$, then we show that there is a factorization such that $u_i = v_i$ for $2 \leq i \leq n+1$ and $u_0u_1 = v_0v_1$. For this, there are four subcases:

(1) if $u_1 \in T_{x_1}$ then, for $\langle k, x_1, z, j\rangle$,

$$u_0u_1 = (w)(a^{s-k}bw^{z+r}a^j) = (a)(a^{s-1}ba^{s-k}bw^{z+r}a^j) = v_0v_1 \in a \cdot NT_{x_1}.$$

(2) If $u_1 \in NT_{x_1}$ then, for $\langle k, x_1, z, j\rangle$ and $1 \le \ell \le s-1$,

$$u_0u_1 = (w)(a^{s-\ell}ba^{s-k}bw^{z+r}a^j) = (wa^{s-\ell}b)(a^{s-k}bw^{z+r}a^j) = v_0v_1 \in wN \cdot T_{x_1}.$$

(3) If $u_1 \in C_{x_1}N$ then, for $\langle k, x_1, z, j\rangle$ and $1 \le \ell \le s-1$,

$$\begin{aligned} u_0u_1 &= (w)(a^{s-k}bw^{z+r+1}a^{s-\ell}b) \\ &= (a)(a^{s-1}ba^{s-k}bw^{z+r+1}a^{s-\ell}b) = v_0v_1 \in a \cdot NC_{x_1}N. \end{aligned}$$

(4) If $u_1 \in NC_{x_1}N$ then, for $\langle k, x_1, z, j\rangle$ and $1 \le \ell, t \le s-1$,

$$\begin{aligned} u_0u_1 &= (w)(a^{s-\ell}ba^{s-k}bw^{z+r+1}a^{s-t}b) \\ &= (wa^{s-\ell}b)(a^{s-k}bw^{z+r+1}a^{s-t}b) = v_0v_1 \in wN \cdot C_{x_1}N. \end{aligned}$$

(iii) If $n \ge 1$ and $u_1 = w^{r+1}$ then we need the fact that $L(\mathcal{A}^\gamma) = \{0,1\}^*$. Let $t = \min\{i \mid i \ge 1, u_i \ne w^{r+1}\}$. Now $u_0u_1 \dots u_{t-1} = w(w^{r+1}) \cdots (w^{r+1}) = w^{(t-1)(r+1)+1}$.

By our assumption, for the sequence $x' = x_1 \dots x_{t-1} \in \{0,1\}^*$, we have transitions

$$(q_{j_i}, z_i) \in \delta(q_{j_{i-1}}, x_i), \quad i = \{1, \dots, t-1\},$$

in $\mathcal{A}^\gamma$, where $1 \le j_i \le s-1$ and $j_0 = 1$, and

$$\sum_{i=1}^{t-1} z_i = 0, \tag{3.1}$$

since $x' \in L(\mathcal{A}^\gamma)$. Therefore there exists

$$v_i' = a^{s-j_{i-1}}bw^{z_i+r}a^{j_i} \in T_{x_i},$$

for $1 \le i \le t-1$, and by setting $v_0 = a$, we get

$$\begin{aligned} v_0v_1' \dots v_{t-1}' &= aa^{s-1}bw^{z_1+r}a^{j_1}a^{s-j_1}bw^{z_2+r}a^{j_2} \dots a^{s-j_{t-2}}bw^{z_{t-1}+r}a^{j_t-1} \\ &= ww^{z_1+r}ww^{z_2+r}w \dots ww^{z_{t-1}+r}a^{j_{t-1}}. \end{aligned}$$

By (3.1)

$$v_0v_1' \dots v_{t-1}' = w^{(t-1)r+(t-1)}a^{j_{t-1}} = w^{(t-1)(r+1)}a^{j_{t-1}},$$

and therefore $u_0u_1 \dots u_{t-1} = v_0'v_1' \dots v_{t-1}'a^{s-j_{t-1}}b$. We may already assume that $v_i = v_i'$ for $1 \le i \le t-2$.

Now there are two cases depending on t. First if $t = n+1$ then we have two subcases:

(1) if $u_{n+1} = w$ then $v_{t-1} = v_{t-1}'$ and $v_{n+1} = a^{s-j_{t-1}}bw \in Nw \in \xi(c)$ and we get that $u = v$.

(2) If $u_{n+1} = a^{s-k}bw \in Nw$, for some $1 \leq k \leq s-1$, then set

$$v_{t-1} = v_n = a^{s-j_{t-2}}bw^{z_{t-1}+r+1}a^{s-k}b \in C_{x_n}N \quad \text{and} \quad v_{n+1} = w.$$

This yields that $u = v$.

The second case is that $t \leq n$. Then set $v_i = u_i$ for $t+1 \leq i \leq n+1$ and there are four subcases for v_{t-1} and v_t depending on u_t:

(1) if $u_t = a^{s-k}bw^{z+r}a^j \in T_{x_t}$, for some $\langle k, x, z, j\rangle$, then set

$$v_{t-1} = v'_{t-1} \text{ and } v_t = a^{s-j_{t-1}}ba^{s-k}bw^{z+r}a^j \in NT_{x_t},$$

to obtain $u = v$.

(2) If $u_t = a^{s-\ell}ba^{s-k}bw^{z+r}a^j \in NT_{x_t}$, for some $\langle k, x, z, j\rangle$ and $1 \leq \ell \leq s-1$, then set

$$v_{t-1} = a^{s-j_{t-2}}bw^{z_{t-1}+r}a^{j_{t-1}}a^{s-j_{t-1}}ba^{s-\ell}b = a^{s-j_{t-2}}bw^{z_{t-1}+r+1}a^{s-\ell}b \in C_{x_t}N$$

and

$$v_t = a^{s-k}bw^{z+r}a^j \in T_{x_t},$$

to obtain $u = v$.

(3) If $u_t = a^{s-k}bw^{z+r+1}a^{s-\ell}b \in C_{x_t}N$, for some $\langle k, x, z, j\rangle$ and $1 \leq \ell \leq s-1$, then set

$$v_{t-1} = v'_{t-1} \text{ and } v_t = a^{s-j_{t-1}}ba^{s-k}bw^{z+r+1}a^{s-\ell}b \in NC_{x_t}N,$$

to obtain $u = v$.

(4) If $u_t = a^{s-\ell}ba^{s-k}bw^{z+r+1}a^{s-h}b \in NC_{x_t}N$, for some $\langle k, x, z, j\rangle$ and $1 \leq \ell, h \leq s-1$, then set

$$v_{t-1} = a^{s-j_{t-2}}bw^{z_{t-1}+r}a^{j_{t-1}}a^{s-j_{t-1}}ba^{s-\ell} = a^{s-j_{t-2}}bw^{z_{t-1}+r+1}a^{s-\ell}b \in C_{x_t}N$$

and

$$v_t = a^{s-k}bw^{z+r+1}a^{s-h}b \in C_{x_t}N,$$

to obtain $u = v$.

We have proved that if $L(\mathcal{A}^\gamma) = \{0,1\}^*$ then $\varphi(L) \subseteq \xi(L)$. The undecidability follows from the undecidability of the universe problem for FA($\mathbb{Z}$) over a binary alphabet, Corollary 1.2. □

In the previous proof we actually proved that it is undecidable, whether the inclusion

$$\varphi(x) \subseteq \xi(x),$$

holds for all word $x \in \Sigma^*$. This also follows from Theorem 3.1.

REFERENCES

[1] V. Halava and T. Harju, Undecidability in integer weighted finite automata. *Fund. Inform.*, to appear.
[2] J. Karhumäki, Equations over finite sets of words and equivalence problems in automata theory. *Theoret. Comput. Sci.* **108** (1993) 103–118.
[3] L.P. Lisovik, An undecidable problem for countable Markov chains. *Cybernetics* **27** (1991) 163–169.
[4] L.P. Lisovik, Nondeterministic systems and finite substitutions on regular language. *Bull. European Assoc. Theoret. Comput. Sci.* (1997) 156–160.
[5] P. Turakainen, The undecidability of some equivalence problems concerning ngsm's and finite substitutions. *Theoret. Comput. Sci.* **174** (1997) 269–274.

Communicated by W. Rytter.
Received February, 1998. Accepted December, 1998.