RAIRO-Theor. Inf. Appl. 46 (2012) 451–457
DOI: 10.1051/ita/2012015

Available online at:
www.rairo-ita.org

UNDECIDABILITY OF INFINITE POST CORRESPONDENCE PROBLEM FOR INSTANCES OF SIZE 8

Jing Dong[1] and Qinghui Liu[1]

Abstract. The infinite Post Correspondence Problem (ωPCP) was shown to be undecidable by Ruohonen (1985) in general. Blondel and Canterini [*Theory Comput. Syst.* **36** (2003) 231–245] showed that ωPCP is undecidable for domain alphabets of size 105, Halava and Harju [*RAIRO–Theor. Inf. Appl.* **40** (2006) 551–557] showed that ωPCP is undecidable for domain alphabets of size 9. By designing a special coding, we delete a letter from Halava and Harju's construction. So we prove that ωPCP is undecidable for domain alphabets of size 8.

Mathematics Subject Classification. 03D35, 03D40, 68R15.

1. Introduction

An *instance* of the *Post Correspondence Problem* (PCP, for short) consists of two morphisms $h, g : A^* \to B^*$, where A and B are two finite alphabets. If the cardinality $|A| = n$, we say that the *size* of the instance (h, g) is n, and we denote it as $PCP(n)$. If there is a nonempty word $w \in A^*$ such that $h(w) = g(w)$, then we call w a *solution* of (h, g). In the PCP, it is asked whether or not an instance (h, g) has a solution.

The PCP is an undecidable problem in its general form; see Post [6]. It was proved that the $PCP(2)$ is decidable by Ehrenfeucht *et al.* [2]; see also [4]. On the other hand, Matiyasevich and Sénizergues proved in [5] that PCP(7) is undecidable.

Keywords and phrases. ωPCP, semi-Thue system, undecidable, theory of computation.

[1] Dept. Comput. Sci., Beijing Institute of Technology, Beijing 100081, P.R. China.
dongjing@bit.edu.cn; qhliu@bit.edu.cn

Article published by EDP Sciences

We denote the empty word by ε. For any finite alphabet Σ and two strings $u, v \in \Sigma^*$, we denote $u \lhd v$, if u is a *prefix* of v, *i.e.*, there exists $w \in \Sigma^*$ such that $v = uw$ (we also denote $w = u^{-1}v$); and denote $v \rhd u$, if u is a *suffix* of v, *i.e.*, there exists $s \in \Sigma^*$ such that $v = su$ (we also denote $s = vu^{-1}$). Given an instance (h, g) of PCP, an infinite word $w = a_1 a_2 \ldots$ over A with $a_i \in A$ for each $i = 1, 2, \ldots$, we call w an *infinite solution* of the instance (h, g), if for any finite prefix u of w, either $h(u) \lhd g(u)$ or $g(u) \lhd h(u)$.

The *infinite PCP* (ωPCP, for short) is to determine whether or not a given instance of the PCP has an infinite solution. Ruohonen proved in [7] that ωPCP is undecidable. Blondel and Canterini [1] used undecidability of the halting problem of the Turing machine and proved that the ωPCP is undecidable for instances of size 105, or in short, ωPCP(105) is undecidable. It was proved by Halava and Harju [3] using undecidability of the termination problem of 3-rule semi-Thue systems that the ωPCP(9) is undecidable.

In this paper we shall prove that the ωPCP(8) is undecidable. As in [3], our proof relies on the undecidability of semi-Thue system.

A *semi-Thue system* $T = (\Sigma, R)$ consists of an alphabet $\Sigma = \{a_1, \ldots, a_n\}$ and a relation $R \subseteq \Sigma^* \times \Sigma^*$, the elements of which are called the *rules* of T. For any two words $w_1, w_2 \in \Sigma^*$, we write $w_1 \to_T w_2$, if there are $x, y \in \Sigma^*$ and $(u, v) \in R$ such that

$$w_1 = xuy, \quad w_2 = xvy.$$

For any word $w_0 \in \Sigma^*$, if there does not exist any infinite sequences of words $w_1, w_2, \ldots$ such that $w_i \to_T w_{i+1}$ for all $i \geq 0$, then we say that T *terminates* on w_0. The *termination problem* asks if $w_0 \in \text{TERMINATE}_T$, where

$$\text{TERMINATE}_T = \{w_0 \in \Sigma^* \mid T \text{ does not terminate on } w_0\}.$$

Theorem 1.1 (see [5]). *There exists a* 3*-rule semi-Thue system with an undecidable termination problem.*

Halava and Harju proved

Theorem 1.2 (see [3]). *If the termination problem is undecidable for a semi-Thue system T with n rules, then the ωPCP is undecidable for instances of size $n + 6$.*

To introduce our idea, we first sketch the proof of Theorem 1.2 in [3].

As a start, any semi-Thue system has an equivalent 2-letter alphabet semi-Thue system with some special coding.

Let $T = (\Sigma, R)$ be a semi-Thue system. We can assume that Σ is binary. Indeed, for $\Sigma = \{a_1, a_2, \ldots, a_k\}$, define a coding $\varphi : \Sigma^* \to \{a, b\}^*$ by

$$\varphi(a_i) = ab^i a, \quad i = 1, \ldots, k. \tag{1.1}$$

Then let $R' = \{(\varphi(u), \varphi(v)) | (u, v) \in R\}$ be a new set of rules, and define $T' = (\{a, b\}, R')$. We see that $\omega \to_T \omega'$ in T if and only if $\varphi(\omega) \to_{T'} \varphi(\omega')$ in T'.

Note that $\varphi(\text{TERMINATE}_T)$ is a strict subset of $\text{TERMINATE}_{T'}$, and the latter may have more complex structure than the former. So, for simplicity, we define the termination problem of T' with coding φ as

$$\text{TERMINATE}_{T',\varphi} = \{\varphi(w_0) \mid w_0 \in \Sigma^*,\ T' \text{ does not terminate on } \varphi(w_0)\}.$$

Hence $\varphi(\text{TERMINATE}_T) = \text{TERMINATE}_{T',\varphi}$. It follows that the termination problem of T is undecidable, if and only if the termination problem of T' with coding φ is undecidable.

Halava and Harju [3] showed that, for any word $u \in \{a,b\}^*$ with coding φ defined in (1.1), they can construct an instance of PCP (h,g), so that T' does not terminate on u if and only if (h,g) has an infinite solution. Their construction is as follows.

Given any two alphabets Y, Z and a nonempty word $s \in Z^*$, define morphisms $l_s, r_s : Y^* \to (Y \cup \{s\})^*$ by $l_s(a) = sa$ and $r_s(a) = as$ for all letters $a \in Y$. Here, we require Y and Z be disjoint.

Let $T = (\{a,b\}, R)$ be a semi-Thue system with $R = \{t_1, t_2, \ldots, t_n\}$ such that $t_i = (u_i, v_i)$ and u_i, v_i are encoded by φ. For any word $u \in \Sigma^*$ encoded by φ, they constructed an instance of PCP of size $n+6$, *i.e.*, $\Phi(u) = (h,g)$, where the morphisms $h, g : (\{a_1, a_2, b_1, b_2, d, \#\} \cup R)^* \to \{a,b,d,\#\}^*$ are defined by

$$\begin{array}{ll|ll}
h(a_1) = dad, & & g(a_1) = add, & \\
h(a_2) = dda, & & g(a_2) = add, & \\
h(b_1) = dbd, & & g(b_1) = bdd, & \\
h(b_2) = ddb, & & g(b_2) = bdd, & \\
h(d) = l_{dd}(u)dd\#d, & & g(d) = dd, & \\
h(\#) = dd\#d, & & g(\#) = \#dd, & \\
h(t_i) = d^{-1}l_{dd}(v_i), & & g(t_i) = r_{dd}(u_i), & \text{for } i = 1, \ldots, n.
\end{array} \tag{1.2}$$

In the special case of $v_i = \varepsilon$, define $h(t_i) = d$.

And then, they proved that each infinite solution of (h,g) can only take the form

$$dw_1\#w_2\#w_3\#\ldots, \tag{1.3}$$

where for all j, $w_j = x_j t_{i_j} y_j$ for some $t_{i_j} \in R$, $x_j \in \{a_1, b_1\}^*$ and $y_j \in \{a_2, b_2\}^*$.

After proving that (h,g) has an infinite solution (1.3) if and only if T does not terminate on u, they obtained that, for any n-rule semi-Thue system, the termination problem is reduced to an ωPCP of alphabet size $n+6$.

Our observation starts from w_j. It is interesting to analysis the structure of $w_j = x_j t_{i_j} y_j$. It is composed of three parts. We call t_{i_j} the *rule part*; x_j, the *left part*; y_j, the *right part*. In the left part, they used the alphabet $\{a_1, b_1\}$, while in the right part, they used the alphabet $\{a_2, b_2\}$.

Our idea is to combine b_1 and b_2 to one letter. It needs some adjustment to make this combination work.

First of all, we change the coding of alphabet. Suppose $T_1 = (\Sigma_1, R_1)$ is a semi-Thue system with $\Sigma_1 = \{a_1, \ldots, a_k\}$. We define

$$\Gamma = \{abbaa(bb)^{i+1}a \mid 1 \le i \le k\}, \tag{1.4}$$

and a coding $\psi : \Sigma_1 \to \Gamma$ such that, for any $i = 1, \ldots, k$,

$$\psi(a_i) = abbaa(bb)^{i+1}a,$$

where we call *abba* the *guide gadget*, and $a(bb)^{i+1}a$ the *distinguish gadget* of $\psi(a_i)$.

Let $R = \{(\psi(u), \psi(v))|(u, v) \in R_1\} = \{t_i|i = 1, \ldots, n\}$ be a new set of rules, where $t_i = (u_i, v_i)$. Define $T = (\{a, b\}, R)$, then T is also a semi-Thue system. It is straightforward that for any word $u \in \Sigma_1^*$, T_1 terminates on u if and only if T terminates on $\psi(u)$.

Now we can define our reduction. For any u coded by ψ, *i.e.*, $u \in \Gamma^*$, we define $\Psi(u) = (h_1, g_1)$, an instance of PCP, by

$$h_1, g_1 : (\{d, a_1, a_2, b_1, \#\} \cup R)^* \to \{d, a, b, \#\}^*$$

with

$$\begin{array}{llll} h_1(a_1) = a, & & g_1(a_1) = a, & \\ h_1(b_1) = bb, & & g_1(b_1) = bb, & \\ h_1(a_2) = baab, & & g_1(a_2) = aabb, & \\ h_1(d) = d\#uabba, & & g_1(d) = d, & \\ h_1(\#) = \#ab, & & g_1(\#) = \#abb, & \\ h_1(t_i) = (ab)^{-1}v_i, & & g_1(t_i) = (abb)^{-1}u_i, & \text{for } i = 1, \ldots, n. \end{array} \tag{1.5}$$

In the special case of $v_i = \varepsilon$, we define $h_1(t_i) = ba$, $g_1(t_i) = (abb)^{-1}u_i abba$. Without loss of generality, we suppose that $u_i \neq \varepsilon$ (otherwise, $\text{TERMINATE}_{T,\psi} = \Gamma^*$ and hence decidable). We call $\#$ the *separating letter*, d the *initial letter*, and t_i the *rule letters*.

We will prove that T does not terminate on u if and only if $\Psi(u)$ has an infinite solution with prefix d, and then we can prove

Theorem 1.3. *If there is a semi-Thue system with n rules having an undecidable termination problem, then ωPCP is undecidable for instances of size $n + 5$.*

By Theorems 1.1 and 1.3, we have

Corollary 1.4. *ωPCP is undecidable for instances of size 8.*

Whether ωPCP is undecidable for instances of size $3 \leq n \leq 7$ is still open.

The same argument as in [3] yields that, by Theorem 4.1 of [1] and Corollary 1.4, the isolation threshold problem for the probabilistic finite automata with two letters and 32 states and the isolated threshold existence problem for probabilistic finite automata with two letters and 220 states, are undecidable.

2. Proof of Theorem 1.3

For any $u \in \{a, b\}^*$ with coding ψ, *i.e.*, $u \in \Gamma^*$, where Γ is defined in (1.4), we prove first that T does not terminate on u if and only if (h_1, g_1) has an infinite solution starting from letter d, where h_1, g_1 are defined in (1.5). And then we

construct an instance (h_2, g_2) so that T does not terminate on u if and only if (h_2, g_2) has an infinite solution (without limitation on the starting letter).

(i) Assume that T does not terminate on u, *i.e.*, there exists a sequence $(w_i)_{i\geq 1}$ such that

$$u = w_1 \to_T w_2 \to_T w_3 \to_T \dots, \tag{2.1}$$

where $u = w_1 = x_1 u_{i_1} y_1$ and $w_j = x_{j-1} v_{i_{j-1}} y_{j-1} = x_j u_{i_j} y_j$ for all $j \geq 2$.

Since $g_1(a_1) = a$, $g_1(b_1) = bb$, for any string $x \in \Gamma^*$, there exists a unique $\tilde{x} \in \{a_1, b_1\}^*$ such that $g_1(\tilde{x}) = x$. Letting $\alpha(x) = \tilde{x}$, the mapping $\alpha : \Gamma^* \to \{a_1, b_1\}^*$ is well defined. For example, we have $\alpha(abbaabbbba) = a_1b_1a_1a_1b_1b_1a_1$. Note that,

$$g_1(\alpha(x)) = h_1(\alpha(x)) = x.$$

Since $g_1(a_2) = aabb$, $g_1(b_1) = bb$, for any non-empty string $x \in \Gamma^*$, there exists a unique string $\tilde{x} \in \{a_2, b_1\}^*$ such that $g_1(\tilde{x}) = (abb)^{-1}xabb$. Letting $\beta(x) = \tilde{x}$, the mapping $\beta : \Gamma^* \to \{a_2, b_1\}^*$ is well defined. For example, let $x = abbaabbbbaabbaabbbbbba$, we have $\beta(x) = a_2b_1a_2a_2b_1b_1a_2$, where guide gadget disappears. Note that

$$g_1(\beta(x)) = (abb)^{-1}xabb, \quad h_1(\beta(x)) = (ab)^{-1}xab.$$

Let us start from d. We have $g_1(d) = d$ and

$$h_1(d) = d\#uabba = d\#x_1u_{i_1}y_1abba.$$

To match $\#x_1u_{i_1}y_1abba$, we see

$$\begin{array}{lll} \text{if } v_{i_1} \neq \varepsilon, & g_1(\#\beta(x_1)t_{i_1}\alpha(y_1)a_1b_1a_1) & = \#x_1u_{i_1}y_1abba, \\ & h_1(\#\beta(x_1)t_{i_1}\alpha(y_1)a_1b_1a_1) & = \#x_1v_{i_1}y_1abba; \\ \text{if } v_{i_1} = \varepsilon, & g_1(\#\beta(x_1)t_{i_1}[(a_1b_1a_1)^{-1}\alpha(y_1)a_1b_1a_1]) & = \#x_1u_{i_1}y_1abba, \\ & h_1(\#\beta(x_1)t_{i_1}[(a_1b_1a_1)^{-1}\alpha(y_1)a_1b_1a_1]) & = \#x_1v_{i_1}y_1abba. \end{array} \tag{2.2}$$

Define $\delta : \Gamma^* \to \{a_1b_1a_1, \varepsilon\}$ as, for any $x \in \Gamma^*$, if $x = \varepsilon$ then $\delta(x) = a_1b_1a_1$, otherwise $\delta(x) = \varepsilon$. So we can summarize the above two cases as

$$\begin{array}{l} g_1(\#\beta(x_1)t_{i_1}[\delta(v_{i_1})^{-1}y_1a_1b_1a_1]) = \#x_1u_{i_1}y_1abba, \\ h_1(\#\beta(x_1)t_{i_1}[\delta(v_{i_1})^{-1}y_1a_1b_1a_1]) = \#x_1v_{i_1}y_1abba = \#x_2u_{i_2}y_2abba. \end{array} \tag{2.3}$$

Note that in case of $v_{i_1} = y_1 = \varepsilon$, we have $\delta(v_{i_1})^{-1}y_1a_1b_1a_1 = \varepsilon$. Analogous to (2.2) and (2.3), we define for any $j \geq 1$,

$$s_j = \beta(x_j)t_{i_j}[\delta(v_{i_j})^{-1}\alpha(y_j)a_1b_1a_1]. \tag{2.4}$$

Then we have

$$\begin{array}{l} g_1(\#s_j) = \#x_ju_{i_j}y_jabba, \\ h_1(\#s_j) = \#x_jv_{i_j}y_jabba = \#x_{j+1}u_{i_{j+1}}y_{j+1}abba. \end{array} \tag{2.5}$$

Now, we see

$$\begin{array}{l} g_1(d\#s_1) = d\#x_1u_{i_1}y_1abba, \\ h_1(d\#s_1) = d\#x_1u_{i_1}y_1abba\#x_2u_{i_2}y_2abba. \end{array}$$

Continuing this process, we see $d\#s_1\#s_2\#\cdots$ is an infinite solution of (h_1, g_1) starting from letter d.

(ii) Assume that w is an infinite solution of the instance (h_1, g_1) starting from letter d.

Step 1. We prove that the separating letter # occurs in w infinitely many times. In fact, there is one occurrence of # in $h_1(d)$ and no occurrences of # in $g_1(d)$. Note that # only appears in $h_1(\#) = \#ab$ and $g_1(\#) = \#abb$ simultaneously. Therefore there are infinitely many occurrences of letter # in any solution of (h_1, g_1) that starts from the letter d.

Step 2. We show that one can write w as

$$w = d\#\tilde{w}_1\#\tilde{w}_2\#\ldots,$$

where for $j \geq 1$, $\tilde{w}_j$ is of the form

$$\tilde{w}_j = \tilde{x}_j t_{i_j} \tilde{y}_j \tag{2.6}$$

for some $t_{i_j} \in R$, $\tilde{x}_j \in \{a_2, b_1\}^*$, $\tilde{y}_j \in \{a_1, b_1\}^*$.

Indeed, for any $s \in \{d, a_1, a_2, b_1, \#, t_1, \ldots, t_n\}^*$, in the string $g_1(s)$, the letter b always appear in pair. Therefore, by definition of h_1 and especially $h_1(\#) = \#ab$, there must exist exactly one rule letter between two successive separating letters #; the sub-word between a separating letter # and the followed rule letter are in $\{a_2, b_1\}^*$, and the sub-word between a rule letter and the followed separating letter # are in $\{a_1, b_1\}^*$.

Step 3. We prove that there exist sequences $(x_j)_{j=1}^{\infty}$, $(y_j)_{j=1}^{\infty}$ and $(i_j)_{j=1}^{\infty}$ in Γ^* such that for any $j \geq 1$,

$$\tilde{x}_j = \beta(x_j), \quad \tilde{y}_j = \delta(v_{i_j})^{-1}\alpha(y_j)a_1b_1a_1, \tag{2.7}$$

and moreover, by setting $w_j = x_j u_{i_j} y_j$ for any $j \geq 1$, we have

$$u = w_1 \to_T w_2 \to_T w_3 \to_T \ldots \tag{2.8}$$

Since w is a solution of (h_1, g_1), we have $g_1(\#\tilde{x}_1 t_{i_1} \tilde{y}_1) = \#uabba$. Now, the most important thing is whether u_{i_1} is a substring of u. Notice that $u \in \Gamma^*$, that is the guide gadget and distinguish gadget appear alternatively in u.

To prove it, we show first that the last letter of $\tilde{x}_1$ is a_2. By the fact that the guide gadget $abba \lhd u_{i_1}$ and $g_1(t_{i_1}) = (abb)^{-1}u_{i_1}$ or $(abb)^{-1}u_{i_1}abba$, we see $aabbbb \lhd g_1(t_{i_1})$. If the last letter of $\tilde{x}_1$ is b_1, then $g_1(\#\tilde{x}_1 t_{i_1}\tilde{y}_1)$ contains a substring $a(bb)^{m+2}aabbbb$ for some $m \geq 0$, there are two continuous distinguish gadgets without a guide gadget between them, which contradicts the fact that $u \in \Gamma^*$.

The last letter of $\tilde{x}_1$ is a_2 implies that $g_1(\#\tilde{x}_1)$ have abb as a suffix. Notice that $abbg_1(t_{i_1})$ contains u_{i_1}, so u_{i_1} is a substring of u.

Now, we can write $u = x_1 u_{i_1} y_1$ with $x_1, y_1 \in \Gamma^*$, and then

$$g_1(\tilde{x}_1) = (abb)^{-1} x_1 abb, \quad g_1(\tilde{y}_1) = \begin{cases} y_1 abba, & \text{if } v_{i_1} \neq \varepsilon, \\ (abba)^{-1} y_1 abba, & \text{if } v_{i_1} = \varepsilon. \end{cases}$$

Since the mappings α and β are well defined, we have

$$\tilde{x}_1 = \beta(x_1), \quad \tilde{y}_1 = \delta(v_{i_1})^{-1} \alpha(y_1) a_1 b_1 a_1,$$

and moreover, $h_1(\#\tilde{w}_1) = \#x_1 v_{i_1} y_1 abba$. Setting $w_1 = u = x_1 u_{i_1} y_1$, $w_2 = x_1 v_{i_1} y_1$, we have

$$w_1 \to_T w_2.$$

Continue this process, we prove (2.7) and (2.8).

(iii) So we have T does not terminate on u if and only if (h_1, g_1) has an infinite solution starting from the letter d. We modify the instance (h_1, g_1) corresponding u to morphisms

$$\begin{array}{l} h_2, g_2 : (\{d, a_1, a_2, b_1, \#\} \cup R)^* \to \{d, a, b, \#, \theta\}^* \\ h_2(\xi) = l_\theta(h_1(\xi)), \quad g_2(\xi) = r_\theta(g_1(\xi)), \quad \forall \xi \in \{a_1, a_2, b_1, \#\} \cup R, \\ h_2(d) = l_\theta(h_1(d)), \quad g_2(d) = \theta \mathrm{d} \theta. \end{array}$$

We get a new instance of PCP. Since every infinite solution of (h_1, g_1) starting from the letter d uses letter d only one time, we see that (h_1, g_1) has an infinite solution starting from the letter d if and only if (h_2, g_2) has an infinite solution. This proves Theorem 1.3.

Acknowledgements. We thank Morningside Center of Mathematics for the partial support, thank Prof. Jacques Peyriere, Wen Zhiying, Xi Lifeng for useful discussions. We thank the referees for helpful comments and suggestions.

References

[1] V.D. Blondel and V. Canterini, Undecidable problems for probabilistic automata of fixed dimension. *Theor. Comput. Syst.* **36** (2003) 231–245.

[2] A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, The (generalized) Post Correspondence Problem with lists consisting of two words is decidable. *Theoret. Comput. Sci.* **21** (1982) 119–144.

[3] V. Halava and T. Harju, Undecibability of infinite Post Correspondence Problem for instances of size 9. *RAIRO–Theor. Inf. Appl.* **40** (2006) 551–557.

[4] V. Halava, T. Harju and M. Hirvensalo, Binary (generalized) Post Correspondence Problem. *Theoret. Comput. Sci.* **276** (2002) 183–204.

[5] Y. Matiyasevich and G. Sénizergues, Decision problems for semi-Thue systems with a few rules. *Theoret. Comput. Sci.* **330** (2005) 145–169.

[6] E. Post, A variant of a recursively unsolvable problem. *Bull. Amer. Math. Soc.* **52** (1946) 264–268.

[7] K. Ruohonen, Reversible machines and Posts Correspondence Problem for biprefix morphisms. *J. Inform. Process. Cybernet. EIK* **21** (1985) 579–595.

Communicated by J. Kari.
Received November 11, 2011. Accepted May 9, 2012.