Characterization of Infinite Solutions of Marked and Binary Post Correspondence Problems

Vesa Halava* Tero Harju
Juhani Karhumäki†

Department of Mathematics and
TUCS - Turku Centre for Computer Science
University of Turku
FIN-20014 Turku
Finland
E-mail: {vehalava,harju}@utu.fi
karhumak@cs.utu.fi

16th February 2004

Abstract

In the infinite Post Correspondence Problem an instance (h, g) consists of two morphisms h and g, and the problem is to determine whether or not there exist an infinite word ω such that $h(\omega) = g(\omega)$. This problem is undecidable in the general case, but it is known to be decidable for binary and marked instances. A morphism is binary, if the domain alphabet is of size 2, and marked, if each image of a letter begins with different letter. We prove that the solutions in these decidable cases form a set $P \cup K^\omega \cup K^* F$, where P is a finite set of ultimately periodic words, K is a finite set of solutions of the PCP, and F is a finite set of morphic images of fixed points of D0L systems.

1 Introduction

In the *Post Correspondence Problem* (PCP, for short), we are given two morphisms $h, g \colon A^* \to B^*$, where A and B are finite alphabets, and we are asked whether or not there exists a nonempty word $w \in A^*$ such that $h(w) = g(w)$. The pair (h, g) is called an *instance* of the PCP and a word $w \in A^+$ is a *solution* of the instance (h, g) if $h(w) = g(w)$. The set of all solutions,

$$E(h, g) = \{w \in A^+ \mid h(w) = g(w)\},$$

is called the *equality set* of the instance (h, g).

[*] Corresponding author.

[†] Supported by the Academy of Finland under grant 44087.

The PCP is undecidable in this general form (see Post [13]). The borderline between decidable and undecidable sets of instances has been investigated in several occasions by restricting the instances of the PCP. For example, it is an easy exercise to show that the *unary PCP*, where the domain alphabet has only one letter, is decidable. An instance (h, g) of the PCP, where $h, g \colon A^* \to B^*$, is *binary* if $|A| = 2$. It was proved in [2] that the PCP is decidable for binary instances; see also [5] or [6] for a somewhat simpler proof. On the other hand, the PCP is undecidable for instance with domain alphabets A satisfying $|A| \geq 7$ (see [12]).

Another known borderline between decidability and undecidability is provided by *marked* and *prefix* morphisms. A morphisms $h \colon A^* \to B^*$ is said to be marked if the images $h(a)$ and $h(b)$ of any two different letters $a, b \in A$ begin with a different letter. The problem where the instances are pairs of marked morphisms is called the *marked PCP* (consisting of *marked instances*). It was proved in [8], that the marked PCP is decidable in general. On the other hand, in [14] it was shown that the PCP is undecidable for instance of prefix morphisms. A morphism is called prefix if no image of a letter is a prefix of an image of another letter.

Note that the solutions of an instance $I = (h, g)$ of the PCP form the set $E(I) = E_{\min}(I)^+$, where

$$E_{\min}(I) = E(I) \setminus E(I)^2$$

consists of the *minimal solutions*, or the *solutions of minimal length*, of the instance. Now, if the instance I is marked, then $E_{\min}$ is a finite set, since different minimal solutions start with different letters.

In this paper we shall consider infinite solutions of the instances (h, g). Two (finite) words u and v are said to be *comparable*, denoted by $u \bowtie v$, if one is a prefix of the other. Let $\omega = a_1 a_2 \ldots$ be an infinite word over A where $a_i \in A$ for each index i. We write $h(\omega) = g(\omega)$ if the morphisms h and g *agree* on ω, that is, if $g(u) \bowtie h(u)$ for all finite prefixes u of ω. We also say that such an infinite word ω is an *infinite solution* of the instance $I = (h, g)$. We denote by $E^\omega(I)$ the set of all infinite solutions of the instance I.

The problem where it is asked whether a given instance of the PCP has or does not have an infinite solution is called naturally the *infinite* PCP. It was shown in [14] that there is no algorithm to determine whether a general instance of the PCP has an infinite solution. Indeed, by [14], the infinite PCP is undecidable for instances where the morphisms are prefix.

In [4] it was proved that the infinite PCP is decidable for marked instances of the PCP. Later, using the previous result, is was proved in [7] that The infinite PCP is decidable for the binary instances.

We shall consider the form of the solutions in these two decidable cases. By studying the proof of decidability of the marked infinite PCP, we establish a characterization of the solutions.

In order to state our main results, we need some definitions and notations.

Let A be an alphabet. For a set $K \subseteq A^*$ of finite words, let

$$K^\omega = \{w_1 w_2 \cdots \mid w_i \in K \text{ for all } i \geq 1\}$$

be the set of all infinite concatenation of words from K. For a singleton set $K = \{w\}$, we let w^ω denote $\{w\}^\omega$. In particular, A^ω consists of all (one-way) infinite words $a_1 a_2 \ldots$ over A. An infinite word $\omega \in A^\omega$ is called *ultimately periodic*, if is $\omega = uv^\omega$ for some finite words u and v.

A *D0L system* $H = (h, w)$ consists of a morphism $h\colon A^* \to A^*$ and an *axiom* $w \in A^*$. A D0L system $H = (h, w)$ defines the language $L_H = \left\{h^i(w) \mid i \geq 0\right\}$. An infinite word ω is a *fixed point* of the D0L system $H = (h, w)$, if

$$\omega = \lim_{n \to \infty} h^n(w),$$

that is, if $h^n(w)$ is a prefix of $h^{n+1}(w)$ for all $n \geq 0$. For further theory of D0L systems, we refer to [9]. We shall prove

Theorem 1 (Main). *Let $I = (h, g)$ be a marked instance of the Post Correspondence Problem. Then the infinite solutions of I form a set*

$$E^\omega(I) = E^\omega_{\min} \cup E^*_{\min}\left(P \cup F\right),$$

where P is a finite set of ultimately periodic words, and F is a finite set of morphic images of fixed points of D0L systems.

This characterization holds also for the binary infinite PCP, as we shall see.

We shall now fix some notation. The *empty word* is denoted by ε. The length of a word u is denoted by $|u|$. A word $u \in A^*$ is said to be a *prefix* of a word $v \in A^*$, denoted by $u \leq v$, if $v = uw$ for some $w \in A^*$. Also, if $u \neq \varepsilon$ and $w \neq \varepsilon$ in $v = uw$, then u is a *proper* prefix of v, and, this is denoted by $u < v$. Recall that u and v are comparable, $u \bowtie v$, if $u \leq v$ or $v \leq u$. The longest common prefix of the words u and v is denoted by $u \wedge v$. If $v = uw$ then we also write $u = vw^{-1}$ and $w = u^{-1}v$.

A word v is *primitive*, if $v = u^n$ implies that $n = 1$, and thus that $u = v$. It is well known (see [10]) that every word w has a unique *primitive root*, a primitive word v such that $w = v^n$ for some $n \geq 1$.

A morphism $h\colon A^* \to B^*$ is said to be *periodic*, if there exists a word $v \in B^*$ such that $h(a) \in v^*$ for all $a \in A$. The word v can be assume to be primitive.

2 The algorithm for marked PCP

The basic result on which we build on the results of this paper is the decidability of the marked instances of the PCP. In order to study the solutions of these instances, we need to consider in detail the decision algorithm for the infinite marked PCP. On the other hand, in order to study those details we need to study the algorithm for deciding the marked PCP. The detailed proofs of the decidability of marked PCP can be found from [8] or [3].

We begin with the following simpler problem with a simple solution:

> Given an instance $I = (h, g)$ of the PCP for $h, g\colon A^* \to B^*$, and a letter $a \in B$. Does there exist words $x, y \in A^+$ such that $h(x) = g(y)$ with $a \leq h(x)$?

Here we do not look for a solution of the instance (h, g), but for a pair of words such that $h(x) = g(y)$, where we additionally require that $h(x)$ starts with a specific letter a. If $h(u) = g(v)$ and $h(u') \neq g(v')$ for all $u' \leq u$ and $v' \leq v$ with $(u, v) \neq (u', v')$, then the pair (u, v) is called a *block solution* to the equation $h(x) = g(y)$.

Let (h, g) be a marked instance of the PCP. Notice that, for each letter a, a block solution is unique, if it exists. We shall now give a procedure for constructing the block solution for a letter a. Define a sequence $(x_i, y_i) \in A^* \times A^*$, for $i = 1, 2, \ldots$, inductively by the following *block procedure* **BP**:

(1) If there are (necessarily unique) letters $b, c \in A$ such that $a \leq h(b)$ and $a \leq g(c)$, and $h(b) \bowtie g(c)$, then set $x_1 = b$ and $y_1 = c$, and let $s_1 \in B^*$ be such that either $h(x_1)s_1 = g(y_1)$ or $g(y_1)s_1 = h(x_1)$. We call such a word s_1 an *overflow of* h or g, respectively.

(2) If $s_{i-1} = \varepsilon$, then (x_{i-1}, y_{i-1}) is the block solution. Otherwise, if $h(x_{i-1})s_i = g(y_{i-1})$ and $s_{i-1} \bowtie h(d)$ for some $d \in A$, then set $x_i = x_{i-1}d$ and $y_i = y_{i-1}$. Similarly, if $g(y_{i-1})s = h(x_{i-1})$ and $s_{i-1} \bowtie g(d)$ for some $d \in A$, then set $x_i = x_{i-1}$ and $y_i = y_{i-1}d$. Let then the new overflow be

$$s_i = \begin{cases} g(y_i)^{-1}h(x_i), & \text{if } g(y_i) \leq h(x_i), \\ h(x_i)^{-1}g(y_i), & \text{if } h(x_i) \leq g(y_i). \end{cases}$$

Since the morphisms are marked and the sequence (x_i, y_i) is constructed in **BP** letter by letter, there are only finitely many different possible overflows. Therefore eventually we reach one of the following choices:

(C1) The words $h(x_i)$ and $g(y_i)$ are not comparable.

(C2) The same overflow s of h or of g reappears.

(C3) For one overflow $s_i = \varepsilon$. Then (x_i, y_i) is the block solution.

Note that in the case (C2) the overflows form an ultimately periodic sequence, and $h(x_i) \bowtie g(y_i)$ for all $i \geq 1$. Although the case (C2) is not important in the marked PCP, this case needs to be studied when we consider infinite solutions of the marked instances in Section 3.

The following lemma is immediate from the conditions (C1), (C2) and (C3) (see [5]).

Lemma 1. *Let $h, g\colon A^* \to B^*$ be marked morphisms. There exists at most one block solution (u, v) such that $a \leq h(u)$ and $a \leq g(v)$. Moreover, such a block solution for a given letter a can be effectively found.*

For marked instance (h, g), if (u, v) is the block solution with $a \leq h(x)$ and $a \leq g(y)$, then (u, v) is called a *block for the letter* a, and it is denote by $\beta(a) = (u, v)$. If no such block solution exists, then $\beta(a)$ is undefined. Furthermore, if $\beta(a)$ is defined, then a is called a *block letter*.

Assume that $w \in A^+$ is a solution of the marked instance (h, g). Then there exists a unique *block decomposition* of w,

$$w = u_1u_2\cdots u_k = v_1v_2\cdots v_k, \tag{1}$$

where $(u_i, v_i) = \beta(a_i)$ for $a_i \in A$ for $i = 1, \ldots, k$. This means that each solution is a concatenation of blocks.

Let (h, g) be a marked instance with $h, g\colon A^* \to B^*$. We can make two assumptions: first we can assume that $A \subset B$ and secondly that

$$a \leq h(a) \quad \text{for all } a \in A. \tag{2}$$

Indeed, both assumptions can be fulfilled by renaming and permuting the letters of the domain alphabet A.

By using blocks we define for a marked instance $I = (h, g)$ its *successor* $I' = (h', g')$ as follows. The new domain alphabet of I' is

$$A' = \{a \mid \beta(a) \text{ exists}\}$$

Note that $A' \subseteq A \subseteq B$. Define the morphisms $h', g' \colon (A')^* \to A^*$ by

$$h'(a) = u \quad \text{and} \quad g'(a) = v, \qquad \text{if } \beta(a) = (u, v).$$

Note that the successor instance I' is marked, since the morphisms h and g are marked and for each letter in B, there is at most one block. The following lemma states a connection between an instance and its successor (see [5] or [4]).

Lemma 2. *Let $I = (h, g)$ be a marked instance and $I' = (h', g')$ be its successor. Then*

(i) $hh'(x) = gg'(x)$ for all $x \in (A')^*$.

(ii) I has a solution if and only if I' has.

(iii) if w' is a solution of I', then $w = h'(w') = g'(w')$ is a solution of I.

(iv) $a \leq h'(a)$ for each $a \in A'$.

It can be proved that the successor I' is at least as simple as the original instance I. Using the construction of the successors inductively we end up in an instance, for which the existence of a solution can be easily decided. We use two measures for the hardness of an instance. The first measure is the size of the domain alphabet. It is immediate that $|A'| \leq |A|$, since, trivially, there can be at most $|A|$ block letters.

The second measure is the *suffix complexity.* For a morphism $h \colon A^* \to B^*$, the suffix complexity $\sigma(h)$ is defined to be the number of different suffixes of the image words,

$$\sigma(h) = |\{x \mid x \text{ proper suffix of } h(a) \text{ for some } a \in A\}|\,.$$

For an instance $I = (h, g)$ of the PCP, the suffix complexity is defined as

$$\sigma(I) = \sigma(h) + \sigma(g).$$

The next lemma was proved in [8].

Lemma 3. *Let $I = (h, g)$ be a marked instance and $I' = (h', g')$ be its successor. Then $\sigma(I') \leq \sigma(I)$.*

The decision procedure for the marked PCP uses the successors iteratively, i.e., it generates the successors as long as an instance is obtained where the decision is easy to make. Therefore we define the *successor sequence* as follows: Let $I_0 = (h, g)$ be a marked instance for $h, g \colon A_0^* \to B^*$. We define inductively $I_i = (h_i, g_i)$ by $I_{i+1} = I_i'$. Moreover, we have that $h_i, g_i \colon A_i^* \to A_{i-1}^*$ for all $i \geq 1$. Recall that $A \subseteq B$ and therefore $A_i \subseteq B$ for all $i \geq 0$.

Now since the size of the alphabet and the suffix complexity do not increase, one of the following three cases occurs in the successor sequence I_i:

(C1) $|A_j| = 1$ for some $j \geq 0$,

(C2) $\sigma(I_j) = 0$ for some $j \geq 0$,

(C3) the successor sequence enters a cycle, i.e., there exists n_0 and $d \geq 1$ such that, for all $j \geq n_0$, $I_j = I_{j+d}$.

The first two cases of the following lemma are trivial; for the third case (the cycling case), see [8], [3], or [4].

Lemma 4. *In case (C1), the instance is unary. In case (C2), all images of letters are of length one, and therefore the block solutions are of length one. The length of each block solution in the case (C3) is one.*

We have now the following decidability theorem from [5].

Theorem 2. *The marked PCP is decidable for all domain alphabets. Moreover, the set of minimal solutions of a marked instance can be effectively constructed.*

3 Infinite solutions in marked PCP

In this section we present a decision algorithm for infinite solutions of the marked PCP. The constructions of the previous section for the marked PCP are crucial in these considerations. For the complete proofs, see [4]. We begin with a lemma, which follows from Lemma 2(i).

Lemma 5. *Let $I = (h, g)$ be a marked instance and $I' = (h', g')$ be its successor. If $x \bowtie y$, then also $h(h'(x)) \bowtie g(g'(y))$.*

Assume that ω is an infinite solution of a marked instance $I = (h, g)$. We say that ω has a *block decomposition*, if

$$\omega = u_1 u_2 \cdots = v_1 v_2 \cdots \ ,$$

where $(u_i, v_i) = \beta(a_i)$ for letters a_i with $i \geq 1$.

All the solutions of I are either of the form

(S1) $\omega \in E^\omega(I)$, that is, $\omega = u_1 u_2 \cdots$ for some $u_i \in E_{\min}(I)$.

or of the form

(S2) $\omega' = v\omega$, where $v \in E(I)$ and ω is an infinite solution of I that does not have a prefix from $E(I)$.

In (S2), there are two possible cases for ω:

(S2a) ω has a block decomposition.

(S2b) ω does not have any block decomposition.

Note that there are only finitely many infinite solutions ω of types (S2a) and (S2b), since each such solution begins with a different letter.

The cases where the suffix complexity is zero or the domain alphabet is unary are easy to determine.

Lemma 6. *Let I be a unary instance or a marked instance with $\sigma(I) = 0$. Then all solutions of I are of type (S1), that is, the set of infinite solutions is exactly $E_{\min}(I)^\omega$.*

Proof. The infinite solutions of unary instances are, trivially, of the form $\omega = a^\omega$ for the only letter a of the domain alphabet, and therefore the claim holds for these cases.

Assume then that $\sigma(I) = 0$. Then the images of the letters have length one (see Lemma 4). Hence, there exists an infinite solution $\omega = a_1 a_2 \dots$ if and only if $h(a_i) = a_i = g(a_i)$ for all a_i. Therefore the solutions are of type (S1), and thus in $E_{\min}(I)^\omega$. Of course, all words in $E_{\min}(I)^\omega$ are infinite solutions of the instance I. □

Now we consider the form of the solutions (S2b).

Theorem 3. *Let $I = (h, g)$ be a marked instance of the PCP. The infinite solutions of type (S2b) of I are ultimately periodic.*

Proof. Consider an infinite solution ω of type (S2b). Then

$$\omega = u_1 u_2 \cdots u_n \omega_1 = v_1 v_2 \cdots v_n \omega_2,$$

where $(u_i, v_i) = \beta(a_i)$ for some letters a_i, $1 \leq i \leq n$, and ω_1, ω_2 are infinite words, which do not have a block as a prefix. Note that also $n = 0$ is possible.

It is clear that $h(\omega_1) = g(\omega_2)$, since $h(u_i) = g(v_i)$ for all i. Let b be the first letter of $h(\omega_1)$. Thus, there are no blocks for b in I, and therefore, in the block procedure **BP** for b after some step the overflows appear periodically, that is,

$$\omega_1 = ux^\omega \quad \text{and} \quad \omega_2 = vy^\omega$$

for some words u, x, v and y with $|h(x)| = |g(y)|$. Hence, the infinite solutions are ultimately periodic in this case. □

For type (S2a), we modify first a result from [4].

Lemma 7. *Let $I = (h, g)$ be a marked instance. If an infinite word ω of type (S2a) is a solution of I, then ω' defined by $h'(\omega') = \omega$ is an infinite solution of type (S2a) or (S2b) of the successor $I' = (h', g')$. Here ω and ω' begin with the same letter.*

Proof. Let ω be an infinite solution of type (S2a) of I. Then it has two factorizations,

$$\omega = u_1 u_2 \cdots = v_1 v_2 \cdots , \tag{3}$$

where $(u_i, v_i) = \beta(a_i)$ for the letters a_i. Now, $\omega' = a_1 a_2 \cdots$, and $h'(\omega') = \omega = g'(\omega')$ by the definition of S'. Clearly, $h(a_1) \bowtie g(a_1)$ and $h'(a_1) = u_1 \bowtie g'(a_1) = v_1$. By assumption, $a_1 \leq h(a_1)$, and hence both sides begin with a_1. It remains to observe that ω' is not in $E(I')^\omega$. Indeed, if $x = a_1 a_2 \cdots a_n \in E(h', g')$, for some $n \geq 1$, then, by Lemma 2, $h'(x) = u_1 u_2 \cdots u_n = v_1 v_2 \cdots v_n = g'(x)$, is in $E(h, g)$ contradicting the assumption that ω is of type (S2a). □

By Lemma 7, instead of considering infinite solutions of type (S2a) of I, we can consider the simpler instance I'. The difficulty here is that we do not know in advance, whether a possible solution of I' is of type (S2a) or (S2b).

For type (S2a) solutions we need to study the successor sequence. We shall first prove a simple case, where all instance of the entire successor sequence have an infinite solution of type (S2a).

Also the following lemma was proved in [4]. We give the proof of this result, since it will be used in the following.

Lemma 8. *There exists an infinite solution of type (S2a) for all I_i starting with a letter b if and only if $h_i(b) \bowtie g_i(b)$ for all $i \geq 0$.*

Proof. It is clear that the existence of type (S2a) solution implies the latter condition of the claim. For the converse, assume that $h_i(b) \bowtie g_i(b)$ for all i. By Lemma 5, the words

$$x_i = h_1 \cdots h_{i-1}h_i(b) \quad \text{and} \quad y_i = g_1 \cdots g_{i-1}g_i(b) \tag{4}$$

are comparable and they both begin with b by the assumption. Let $z_i = x_i \wedge y_i$, the longest common prefix of x_i and y_i. Obviously, z_i is either x_i or y_i. Since $b \leq h_i(b)$, we have $h_i(b) = bx_i$ for a word x_i, and hence $h_{i-1}h_i(b) = h_{i-1}(b)h_{i-1}(x_i)$, and so $h_{i-1}(b) \leq h_{i-1}h_i(b)$. Therefore, $z_i \leq z_{i+1}$ for all i. Now the word $\omega = \lim_{i\to\infty} z_i$ is an infinite solution of I_0 starting with b, since, by Lemma 5, $h_0h_1 \cdots h_i(b)$ and $g_0g_1 \cdots g_i(b)$ are comparable for all i. The claim now follows inductively for all $i \geq 0$. □

We prove our main result for the remaining case (S2a).

Theorem 4. *Let I be a marked instance of the PCP. Type (S2a) solutions of I are either ultimately periodic or of the form $f(\omega)$, where f is a morphism and ω is a fixed point of the D0L system with a single letter axiom.*

Proof. Let I_i be the successor sequence for the marked instance (h_0, g_0), where $I = I_0$ and $I_i = (h_i, g_i)$ for $h_i, g_i \colon A_i^* \to A_{i-1}^*$. Assume that the instance $I = (h_0, g_0)$ has an infinite solution ω_0 beginning with a letter a and having a block decomposition as in type (S2a).

By the construction of the successor sequence, see Lemma 2, we have a sequence ω_i, $i \geq 0$, where ω_i is an infinite solution of the instance I_i such that $\omega_{i-1} = h_i(\omega_i)$ for all $i \geq 1$. As at the end of the proof of Lemma 7, we can show that if $\omega_i \in E(h_i, g_i)^\omega$, then also $\omega_{i-1} \in E(h_{i-1}, g_{i-1})^\omega$. By Lemma 6, it follows that the successor sequence is cycling, that is, satisfies (C3).

Assume that n_0 and $d \geq 1$ are such that $I_j = I_{j+d}$ for all $j \geq n_0$.

We have two cases according to whether the letter a disappears or remains in the successor sequence. Note that if there exists an infinite solution starting with a, then $h_i(a) \bowtie g_i(a)$ as long as the letter appears in the successor sequence.

Assume first that a disappears from the successor sequence after the instance I_j, that is, there is a block for a in I_j but not in I_{j+1}. By Lemma 7, there is an instance I_i, with $i \leq j$, that has a solution of type (S2b) beginning with a. Hence the type changes from (S2b) to (S2a) at some step of the sequence. By Theorem 3, the instance I_i has an ultimately periodic solution ω' (starting with a). Therefore, by Lemma 5, the word

$$\omega = h_1h_2 \cdots h_i(\omega')$$

is an infinite solution of I and it is ultimately periodic, since it is a morphic image of an ultimately periodic word.

Assume then that the letter a remains in the successor sequence. Since there is an infinite solution of I beginning with a, we have $h(a) \bowtie g(a)$ and, since $a \leq h(a)$ and also $a \leq g(a)$, we obtain that $h_i(a) \bowtie g_i(a)$ for all $i \geq 0$. Therefore, by Lemma 8, there exists an infinite solution of type (S2a) beginning with a for all instances in the successor sequence.

Assume that the sequence is such that $I_i = (h_i, g_i)$ and for all $i \geq n$, $I_i = I_{i+d}$ as in the above. As in the proof of Lemma 8, we may determine an infinite solution of I by

$$h_1 h_2 \cdots h_{n-1}(h_n h_{n+1} \cdots h_{n+d-1})^i(a),$$

where i tends to infinity. Define then

$$f = h_1 h_2 \cdots h_{n-1} \quad \text{and} \quad h = h_n h_{n+1} \cdots h_{n+d-1}.$$

Now, f is a morphism, and (h, a) is a D0L system with a fixed point $\omega' = \lim_{n\to\infty} h^n(a)$. Hence $\omega = f(\omega')$, and this proves the claim. □

By the previous theorems, we have immediately our main theorem for marked instances.

Theorem 1. *Let $I = (h, g)$ be a marked instance of the Post Correspondence Problem. Then the infinite solutions of I form a set*

$$E^\omega(I) = E^\omega_{\min} \cup E^*_{\min}\,(P \cup F)\,,$$

where P is a finite set of ultimately periodic words, and F is a finite set of morphic images of fixed points of D0L systems.

Example 1. Let $I_0 = (h_0, g_0)$, where $h_0, g_0\colon \{a, b, c, d\}^* \to \{a, b, c, d\}^*$, defined in the following table:

	a	b	c	d
h	a	baa	c	dd
g	baa	aa	caa	ddd

Now, there is an infinite solution beginning with d, namely d^ω. The successor morphisms are all listed in the following table.

	a	b	c	d
h_1	aa	b	caa	ddd
g_1	b	a	c	dd
h_2	a	b	c	dd
g_2	bb	a	cbb	ddd
h_3	a	bb	cbb	ddd
g_3	b	a	c	dd
h_4	a	b	c	dd
g_4	b	aa	caa	ddd

In this example, we have $I_5 = I_1$. We see that there are infinite solutions of type (S2a) beginning with c and with d. These can be defined by D0L systems, but the solution for d is periodic: d^ω.

The infinite solution starting with c is the fixed point of the D0L system (h, c), where $h = h_1 h_2 h_3 h_4$ is defined by

	a	b	c	d
h	aa	bb	$caabb$	d^{12}

Now one can see that the fixed point is

$$caabbaaaabbbbaaaaaaaabbbbbbbb\cdots = ca^2b^2a^4b^4a^8b^8\cdots,$$

i.e., a sequence counting powers of two. This sequence is clearly nonregular in the language theoretic sense.

4 Binary infinite PCP

It was proved in [7] that the existence of an infinite solution is decidable for the binary instances. We state the reduction introduced in [7], where a given (nonperiodic) binary instance I is transformed to an equivalent instance of the marked PCP with three letters. Therefore, also the form of the solution of the binary infinite PCP follows from the study in the previous section.

In this section we assume that the domains and the ranges of the morphisms equal to the binary set $\{0,1\}$. Therefore the instances are of the form $I = (h, g)$ such that $h, g\colon \{0,1\}^* \to \{0,1\}^*$.

Theorem 5. *Let $I = (h, g)$ be a binary instance of the PCP that has at least one infinite solution.*

(i) If both h and g are periodic, then the set of infinite solutions of I is $\{0,1\}^\omega$.

(ii) If exactly one of the morphisms h or g is periodic, then I has a unique solution that is ultimately periodic.

Proof. Let $I = (h, g)$ be such that $h(0), h(1) \in v^*$ for a primitive word $v \in \{0,1\}^+$. Clearly, $\omega \in \{0,1\}^\omega$ is an infinite solution of the instance I if and only if $g(\omega) = v^\omega$.

(i) Assume first that also g is periodic, say $g(0), g(1) \in u^*$ for a primitive word $u \in \{0,1\}^+$. Now, there exists an infinite solution of I if and only if $v^\omega = u^\omega$, which means that $u^{|v|} = v^{|u|}$ and thus $u = v$, since each word has a unique primitive root. Therefore, the set of infinite solutions is exactly $\{0,1\}^\omega$.

(ii) Assume then that g is nonperiodic, and therefore also nonerasing. We prove that there is exactly one infinite solution. Assume on the contrary that there are two different infinite solutions ω_1' and ω_2'. Let $u = \omega_1' \wedge \omega_2'$ so that $\omega_1' = u\omega_1$ and $\omega_2' = u\omega_2$. Hence we have $g(\omega_1) = v^\omega = g(\omega_2)$. By the defect theorem for infinite words, see, e.g., [1, Theorem 6.2.4], $g(0)$ and $g(1)$ are powers of a common word; a contradiction.

We still need to prove that the unique solution is ultimately periodic. Assume that ω is a solution. Since $h(\omega) = v^\omega$, there are prefixes u and uv of ω such that $g(u) = v^r x$ and $g(uv) = v^s x$ for some $r < s$ and a word x. In this case, $g(v) = yv^i x$ for some $i \geq 0$ and a word y with $v = xy$. It follows that $\omega = uv^\omega$ and thus it is ultimately periodic. □

Next we consider the nonperiodic cases. Note that in the binary case, nonperiodic morphisms are nonerasing.

Let $h\colon \{0,1\}^* \to \{0,1\}^*$ be a nonperiodic morphism. Denote $h^{(0)} = h$. For a word w, denote by w_i the prefix of w of length i. Define a morphisms $h^{(i)}$ in a following way: For a letter $x \in \{0,1\}$, let j be such that $j \equiv i \pmod{|h(x)|}$ with $0 \le j < |h(x)|$, and let

$$h^{(i)}(x) = (h(x)_j)^{-1}\big(h(x)h(x)_j\big).$$

In other words, the morphisms $h^{(i)}$ are cyclic shifts of the original morphism h. Denote by $z_h = h(01) \wedge h(10)$ the longest common prefix of the images of 01 and 10.

The following lemma is a combined from the results in [2] and [7]. (In [7], see Lemma 4 and its proof.) In this lemma # is a new letter.

Lemma 9. *Let $m = |z_h|$ and $n = |z_g|$ and assume $m \ge n$. Then the morphisms $h^{(m)}$ and $g^{(n)}$ are marked and the instance (h, g) of the binary PCP has an infinite solution ω if and only if the marked instance (h', g') has the infinite solution $\#\omega$, where*

$$\begin{aligned} h'(\#) &= \# z_g^{-1} z_h\,, \qquad & h'(a) &= h^{(m)}(a),\\ g'(\#) &= \#\,, & g'(a) &= g^{(n)}(a). \end{aligned}$$

for both $a \in \{0,1\}$.

Note that we can have that $z_g = z_h$ in which case in the new instance (h', g') we can safely remove the special symbol #.

By Lemma 9 and the results in the previous section, we have

Corollary 1. *Let I be a binary instance of the PCP that has an infinite solution. Then there is a finite set K of words such that each infinite solution ω is either (i) ultimately periodic, or (ii) $\omega \in K^*\omega'$, where ω' is a morphic images of the fixed points of D0L systems, or (iii) $\omega \in K^\omega$, where K^ω is a set of infinite solutions.*

Next example shows that the D0L case is possible also for binary instances.

Example 2. Consider the binary instance (h, g) defined by

	a	b
h	a	baa
g	aab	aa

Now, $z_h = \varepsilon$ and $z_g = aa$. By Lemma 9, the instance (h, g) has an infinite solution if and only if the instance (h', g') has a solution beginning with #, where

	a	b	$\#$
h'	a	baa	$\#$
g'	baa	aa	$\#aa$

Note that (h', g') is the same instance as in Example 1, if we replace c with # and omit the useless letter d. Therefore, by Example 1, (h, g) has an infinite solution $a^2b^2a^4b^4a^8b^8\cdots$.

References

[1] V. Bruyère. Codes, Chapter 8 in "Algebraic Combinatorics on Words´´ (M. Lothaire) [11], pp. 197–229.

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

[3] V. Halava. *The Post Correspondence Problem for Marked Morphisms.* PhD thesis, Department of Math, Univ. of Turku. TUCS Dissertations no. 37, 2002.

[4] V. Halava and T. Harju. Infinite solutions of the marked Post Correspondence Problem. In J. Karhumäki W. Brauer, H. Ehrig and A. Salomaa, editors, *Formal and Natural Computing*, volume 2300 of *Lecture Notes in Comput. Sci.*, pages 57–68. Springer-Verlag, 2002.

[5] V. Halava, T. Harju, and M. Hirvensalo. Generalized Post correspondence problem for marked morphisms. *Internat. J. Algebra Comput.*, 10(6):757–772, 2000.

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

[7] V. Halava, T. Harju, and J. Karhumäki. Decidability of the binary infinite Post Correspondence Problem. *Discrete Appl. Math.*, 130:521–526, 2003.

[8] V. Halava, M. Hirvensalo, and R. de Wolf. Marked PCP is decidable. *Theoret. Comput. Sci.*, 255(1-2):193–204, 2001.

[9] L. Kari, G. Rozenberg, and A. Salomaa. L systems. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages, Vol. 1*, pages 253–328. Springer, Berlin, 1997.

[10] M. Lothaire, *Combinatorics on Words*, Addison-Wesley, 1983.

[11] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.

[12] Y. Matiyasevich and G. Sénizergues. Decision problems for semi-Thue systems with a few rules. In *Proceedings, 11*$^{\text{th}}$ *Annual IEEE Symposium on Logic in Computer Science*, pages 523–531, New Brunswick, New Jersey, 27–30 July 1996. IEEE Computer Society Press.

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

[14] K. Ruohonen. Reversible machines and Post's correspondence problem for biprefix morphisms. *Elektron. Informationsverarb. Kybernet. (EIK)*, 21(12):579–595, 1985.