Fundamenta Informaticae 110 (2011) 1–4
DOI 10.3233/FI-2011-531
IOS Press

The Unique Decipherability in the Monoid of Regular Languages is Undecidable[*]

Juhani Karhumäki and Aleksi Saarela[†]
Department of Mathematics and Turku Centre for Computer Science TUCS
University of Turku, FI-20014 Turku, Finland
{*karhumak,amsaar*}*@utu.fi*

Abstract. We show by a simple reduction that the unique decipherability problem in the language monoid of regular languages over a non-unary alphabet is undecidable.

1. Introduction

The unique decipherability problem in a monoid $\mathcal{M}$ asks whether a given finite subset M of $\mathcal{M}$ is a free generating set of the submonoid of $\mathcal{M}$ it generates. The problem is very natural in the theory of information transmission. It was probably first encountered when asking whether or not a finite encoding $i \mapsto w_i$ can be uniquely decoded, that is whether a finite set of words $\{w_i \mid i \in I\}$ is a free generating set of a submonoid of a free monoid. An affirmative answer to this problem was given in a paper known as a source of the classical *Sardinas-Patterson algorithm* [8]. This algorithm extends straightforwardly to regular languages, see, e.g., Section I.3 in [1].

There are several options to try to extend the above problem. One such direction is to consider instead of the freeness of a finitely generated subsemigroup of Σ^*, the isomorphism of two such semigroups. This reveals some interesting phenomena. First of all the problem remains decidable, see [2], but the proof relies on something surprising, namely systems of equations over free semigroups and their compactness properties. Even more interestingly this approach does not extend to subsemigroups generated by regular languages – in fact, the decidability of their isomorphism is an open problem. Another interesting feature here is that when moving from subsemigroups of a free semigroup to more general

[*] Supported by the Academy of Finland under grant 121419

[†] Address for correspondence: Department of Mathematics and Turku Centre for Computer Science TUCS, University of Turku, FI-20014 Turku, Finland

semigroups the isomorphism problem becomes undecidable. For example, for finitely generated multiplicative semigroups of 3×3 matrices over natural numbers the isomorphism, or even the freeness, problem is undecidable, see, e.g., [6] or [5] as a survey of these problems. These problems remain undecidable for upper triangular 3×3 matrices.

Another way of trying to extend the freeness problem is to consider some other semigroups. The monoids of languages constitute such examples related to automata and formal language theory. Recently such an attempt was made in [3]. It was shown that the unique decipherability in the monoid of unary languages is decidable in the case of finite languages, as well as in the case of regular, that is ultimately periodic languages. Later in [7] a characterization of unary finite languages possessing the unique decipherability property was given, thus sharpening the above result of [3]. Also a simple case of non-unary languages was settled affirmatively in [3]: if there is a letter that appears exactly once in every word of every language, then the problem is decidable.

As far as we know, the general problem is very much untouched. This note is a contribution to that. We show that, given a finite collection of regular languages, it is undecidable whether it is a free generating set in the monoid of languages. Our result is based on another undecidability result in [4] which states that the unique decipherability problem is undecidable in the trace monoid $\{a, b\}^* \times \{c, d\}^*$.

2. The Result

The *unique decipherability problem* in a monoid $\mathcal{M}$ asks whether a given finite subset $M = \{a_1, \dots, a_n\} \subset \mathcal{M}$ is a free generating set of the submonoid M^*. In other words, it asks whether every element of $\mathcal{M}$ that has a representation as a product of the elements $a_1, \dots, a_n$ has a unique such representation.

We fix two disjoint binary alphabets $\Sigma_1 = \{a, b\}$ and $\Sigma_2 = \{c, d\}$. We will use the following lemma that was proved in [4].

Lemma 2.1. The unique decipherability problem is undecidable in the trace monoid $\Sigma_1^* \times \Sigma_2^*$.

We will show that the monoid $\Sigma_1^* \times \Sigma_2^*$ can be effectively embedded in the monoid of regular languages over a non-unary alphabet.

For a word $w = a_1 \dots a_n$, let $\mathrm{sw}(w)$ be the set of all (scattered) subwords of w, that is

$$\mathrm{sw}(w) = \{a_{i_1} \dots a_{i_k} \mid 1 \leq i_1 < \cdots < i_k \leq n\}.$$

Let

$$X = (\Sigma_1 \cup \Sigma_2)^+ \smallsetminus \Sigma_1^+ \smallsetminus \Sigma_2^+$$

be the set of those words that contain letters from both Σ_1 and Σ_2. For all pairs of words $(u, t) \in \Sigma_1^* \times \Sigma_2^*$ we define a regular language

$$L(u, t) = \mathrm{sw}(u) \cup \mathrm{sw}(t) \cup X \tag{1}$$

over $\Sigma_1 \cup \Sigma_2$.

Lemma 2.2. The mapping L defined by (1) is an injective morphism.

Proof:
First we show that if $u, u' \in \Sigma_1^*$ and $t, t' \in \Sigma_2^*$, then

$$L(u,t)L(u',t') = L(uu',tt').$$

Because $X \subset L(u,t)$ and $1 \in L(u',t')$, the set X is a subset of $L(u,t)L(u',t')$. Of the words in Σ_1^*, the language $L(u,t)L(u',t')$ contains exactly those that can be written as xy, where $x \in \mathrm{sw}(u)$ and $y \in \mathrm{sw}(u')$. These words form the set $\mathrm{sw}(uu')$. Similarly, $L(u,t)L(u',t') \cap \Sigma_2^* = \mathrm{sw}(tt')$. Thus

$$L(u,t)L(u',t') = \mathrm{sw}(uu') \cup \mathrm{sw}(tt') \cup X = L(uu',tt'),$$

and L is a morphism.

Next we show that if $u, u' \in \Sigma_1^*$ and $t, t' \in \Sigma_2^*$ and $L(u,t) = L(u',t')$, then $u = u'$ and $t = t'$. The longest words of Σ_1^* and Σ_2^* in $L(u,t)$ are u and t, and the longest words of Σ_1^* and Σ_2^* in $L(u',t')$ are u' and t'. Thus if $L(u,t) = L(u',t')$, then $u = u'$ and $t = t'$ and L is injective. □

Theorem 2.1. The unique decipherability problem is undecidable in the monoid of regular languages over a non-unary alphabet.

Proof:
For the alphabet $\Sigma_1 \cup \Sigma_2$, this follows from Lemma 2.2. This alphabet has four letters, but $(\Sigma_1 \cup \Sigma_2)^*$ can be embedded in $\{a, b\}^*$, so the undecidability holds already for a binary alphabet. □

The question of the decidability of the unique decipherability problem in the monoid of finite languages is also interesting. It is noteworthy that in [4] everything is finite: the input is a finite collection of elements of the monoid $\Sigma_1^* \times \Sigma_2^*$. On the other hand, we obtain our result only for regular subsets of $(\Sigma_1 \cup \Sigma_2)^*$: the languages $L(u,t)$ contain an infinite regular part X, which is essential in the proof of Lemma 2.2. Actually, we do not know whether our result extends to finite collections of finite languages, so it remains an open question whether the problem is undecidable already in the monoid of finite languages.

References

[1] Jean Berstel and Dominique Perrin. *Theory of Codes*. Academic Press, 1985.

[2] Christian Choffrut, Tero Harju, and Juhani Karhumäki. A note on decidability questions on presentations of word semigroups. *Theoret. Comput. Sci.*, 183(1):83–92, 1997.

[3] Christian Choffrut and Juhani Karhumäki. Unique decipherability in the monoid of languages: an application of rational relations. *Theory Comput. Syst.* To appear in special issue of CSR.

[4] Marek Chrobak and Wojciech Rytter. Unique decipherability for partially commutative alphabet. *Fund. Inform.*, 10(3):323–336, 1986.

[5] Tero Harju and Juhani Karhumäki. Morphisms. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 439–510. Springer-Verlag, 1997.

[6] David A. Klarner, Jean-Camille Birget, and Wade Satterfield. On the undecidability of the freeness of integer matrix semigroups. *Internat. J. Algebra Comput.*, 1(2):223–226, 1991.

[7] Aleksi Saarela. Unique decipherability in the additive monoid of sets of numbers. *RAIRO Inform. Theor. Appl.* To appear.

[8] August Albert Sardinas and George W. Patterson. A necessary and sufficient condition for unique decomposition of coded messages. In *IRE Intern. Conv. Rec. 8 (1953)*, pages 104–108. Chapman and Hall, 1953.