

Uniform generation of infinite traces

Samy Abbes

Vincent Jugué

Université Paris Cité
IRIF CNRS UMR 8243
8 place Aurélie Nemours
75013 Paris, France
abbes@irif.fr

Univ Gustave Eiffel,
CNRS, LIGM
F-77454 Marne-la-Vallée,
France
vincent.juge@univ-eiffel.fr

Abstract

We introduce an algorithm for the uniform generation of infinite traces, *i.e.*, infinite words up to commutation of some letters. The algorithm outputs on-the-fly approximations of a theoretical infinite trace, the latter being distributed according to the exact uniform probability measure. The average size of the approximation grows linearly with the time of execution of the algorithm, provided that some—costly—precomputations have been done.

1 Introduction

Trace monoids are models of discrete-event concurrent systems. Consider an alphabet Σ equipped with a binary, symmetric and reflexive relation \mathcal{R} , and let $I = (\Sigma \times \Sigma) \setminus \mathcal{R}$. The *trace monoid* $\mathcal{M} = \mathcal{M}(\Sigma, \mathcal{R})$ is the presented monoid $\langle \Sigma \mid \mathcal{I} \rangle$ where \mathcal{I} is the collection of pairs (ab, ba) for (a, b) ranging over I . Hence, an element in \mathcal{M} , called a *trace*, is the congruence class of some word $x \in \Sigma^*$, and congruent words are obtained from x by successively exchanging the places of contiguous letters a and b such that $(a, b) \in I$. For $(a, b) \in I$, the corresponding elements a and b are therefore commutative in \mathcal{M} , which corresponds, from the systems point of view, to the concurrency of actions represented by a and b . Trace monoids have been ubiquitous in computer science and in combinatorics, since their very first use as models for databases with concurrency [7, 8].

A trace x of a trace monoid \mathcal{M} represents an execution of some concurrent system. It is thus natural to inquire the random generation of traces, and more precisely the random generation of *large* traces—the *length* of a trace is simply the length of the associated congruent words. Given a large integer N , one could turn toward Boltzmann generation techniques to operate the random generation of traces of length N . However, when this is done, this technique is of little help for the generation of traces of a larger length; that would require to start again the procedure back from the beginning. Whereas, when seeking for the simulation of “real-life” executions of a system, there is often little argument for stopping the execution at a particular size.

It is therefore more appealing to design techniques for the random generation of “infinite” executions. The precise target is the following: given a notion of infinite traces and a uniform measure for the space of these infinite traces, we

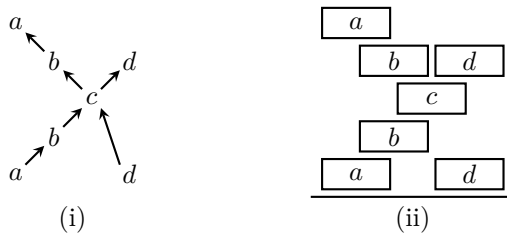


Figure 1: (i) *Hasse diagram of the labeled partial order corresponding to the element $x = a \cdot b \cdot d \cdot c \cdot b \cdot a \cdot d$ of the trace monoid $\mathcal{M}(\Sigma, \mathcal{R})$ with $\Sigma = \{a, b, c, d\}$ and where \mathcal{R} is the reflexive and symmetric closure of $\{(a, b), (b, c), (c, d)\}$.* (ii) *Heap of pieces representing the same element.*

look for an algorithm that produces, for each integer n , a finite random trace of length proportional to n on average, and which coincides with a prefix of a uniformly distributed infinite trace. Our solution makes an extensive use of two tools: some discrete probability distributions on traces of a particular kind, and the combinatorics of so-called pyramidal traces. It applies to any irreducible trace monoid, which extends a work by one of the authors that was restricted to “dimer-like” trace monoids [2].

The problem studied in this paper contrasts with the work in [4] which focuses on two other problems: 1) uniformly generating traces of fixed length; and 2) evaluating the expectation of cost functions on traces of fixed length.

Outline. Section 2 introduces the basic combinatorial and probabilistic material for trace monoids. Section 3 contains the contributions, in the form of random generation algorithms.

2 Trace monoids and probability distributions

Let $\mathcal{M} = \mathcal{M}(\Sigma, \mathcal{R})$ be a trace monoid. It is known that elements of \mathcal{M} can be represented by *heaps* [12], according to a bijective correspondence which we briefly recall now, following the presentation of [10]. As illustrated in Fig. 1 (i), a *heap* is a triple (P, \preceq, ℓ) , where (P, \preceq) is a poset and $\ell : P \rightarrow \Sigma$ is a labeling of P by elements of Σ , satisfying the two following properties: (1) if $x, y \in P$ are such that $\ell(x)\mathcal{R}\ell(y)$, then, $x \preceq y$ or $y \preceq x$; (2) the relation \preceq is the transitive closure of the relations from (1). More precisely, the heap is the equivalence class of (P, \preceq, ℓ) up to isomorphism of labeled partial orders.

To picture heaps corresponding to traces in \mathcal{M} , one represents elements of Σ as elementary pieces that can be piled up with the following constraints, as illustrated in Fig. 1 (ii): (1) pieces can only be moved vertically; (2) pieces labeled by the same letter move along the same vertical lane; and (3) two pieces labeled by a and b in Σ can be moved independently of each other if and only if $(a, b) \notin \mathcal{R}$.

If Σ' is a subset of Σ , we denote by $\mathcal{M}_{\Sigma'}$ the sub-monoid of $\mathcal{M}(\Sigma, \mathcal{R})$ gen-

erated by Σ' . In particular, $\mathcal{M}(\Sigma, \mathcal{R}) = \mathcal{M}_\Sigma$, a notation that we shall use from now on.

A *clique* of \mathcal{M} is any commutative product $a_1 \cdot \dots \cdot a_k$, where a_1, \dots, a_k are distinct elements of Σ such that $(a_i, a_j) \notin \mathcal{R}$ for all distinct i and j . We denote by \mathcal{C}_Σ the set of cliques of \mathcal{M}_Σ .

Cliques of \mathcal{M}_Σ play an important role for the study of its combinatorics. Indeed, each pair (Σ, \mathcal{R}) is associated with the *Möbius polynomial* $\mu_\Sigma(X)$ and the generating series $G_\Sigma(X)$ defined as in [5] by:

$$\mu_\Sigma(X) = \sum_{\gamma \in \mathcal{C}_\Sigma} (-1)^{|\gamma|} X^{|\gamma|} \quad \text{and} \quad G_\Sigma(X) = \sum_{x \in \mathcal{M}_\Sigma} X^{|x|}.$$

More generally, let U be a subset of Σ . For $x \in \mathcal{M}_\Sigma$, let us write $\max(x) \subseteq U$ if the heap (P, \preceq, ℓ) corresponding to x has the property that all the maximal elements of the poset (P, \preceq) are labeled by elements in U . Let $G_{\Sigma, U}(X)$ be the generating series of the elements of \mathcal{M}_Σ satisfying this property:

$$G_{\Sigma, U}(X) = \sum_{x \in \mathcal{M}_\Sigma : \max(x) \subseteq U} X^{|x|}.$$

Then the following formula holds [12]:

$$G_{\Sigma, U}(X) = \frac{\mu_{\Sigma \setminus U}(X)}{\mu_\Sigma(X)}. \quad (1)$$

If $\Sigma \neq \emptyset$, then $\mu_\Sigma(X)$ has a unique root of smallest modulus in the complex plane [9, 6, 11]. This root, which we denote by p_Σ , is positive real and is at most 1. It coincides with the radius of convergence of the power series $G_{\Sigma, U}(X)$ for any non empty subset U of Σ . Hence, substituting p to X in the above identity provides an equality in \mathbb{R} if $p \in (0, p_\Sigma)$.

As a particular case, obtained for $U = \Sigma$, one has: $G_\Sigma(p) = 1/\mu_\Sigma(p)$, for all $p \in (0, p_\Sigma)$. Consequently, for each $p \in (0, p_\Sigma)$, the following formula defines a probability distribution on the countable set \mathcal{M}_Σ :

$$\forall x \in \mathcal{M}_\Sigma \quad B_{\Sigma, p}(\{x\}) = \mu_\Sigma(p) p^{|x|} \quad (2)$$

Let \leq denote the left-divisibility relation on \mathcal{M}_Σ , defined by $x \leq y$ if and only if $x \cdot z = y$ for some $z \in \mathcal{M}_\Sigma$. For every $x \in \mathcal{M}_\Sigma$, let $\uparrow x = \{y \in \mathcal{M}_\Sigma : x \leq y\}$. Then, for every $p \in (0, p_\Sigma)$, the distribution $B_{\Sigma, p}$ is the unique probability distribution on \mathcal{M}_Σ satisfying the following identities [4] for all $x \in \mathcal{M}_{\Sigma, p}$:

$$B_{\Sigma, p}(\uparrow x) = p^{|x|}. \quad (3)$$

Now, we briefly explain the construction of infinite traces and of the uniform measure on their set. If $x = (x_n)_{n \geq 0}$ and $y = (y_n)_{n \geq 0}$ are two non-decreasing sequences in \mathcal{M}_Σ , we define $x \sqsubseteq y$ whenever, for all $n \geq 0$, there exists $k \geq 0$

such that $x_n \leq y_k$. The relation \sqsubseteq is a preorder relation on the set of non-decreasing sequences. Let \asymp be the equivalence relation defined by $x \asymp y$ if and only if $x \sqsubseteq y$ and $y \sqsubseteq x$. Equivalence classes of non-decreasing sequences modulo \asymp are called *generalized traces*, and their set is denoted by $\overline{\mathcal{M}}_\Sigma$. The set $\overline{\mathcal{M}}_\Sigma$ is equipped with an ordering relation, denoted by \leq , which is the collapse of the preordering relation \sqsubseteq .

The partial order $(\mathcal{M}_\Sigma, \leq)$ is embedded into $(\overline{\mathcal{M}}_\Sigma, \leq)$, by sending an element $x \in \mathcal{M}_\Sigma$ to the equivalence class of the constant sequence $(x_n)_{n \geq 0}$ with $x_n = x$ for all $n \geq 0$. Hence, we identify \mathcal{M}_Σ with its image in $\overline{\mathcal{M}}_\Sigma$, and we put $\partial\mathcal{M}_\Sigma = \overline{\mathcal{M}}_\Sigma \setminus \mathcal{M}_\Sigma$. Elements of $\partial\mathcal{M}_\Sigma$ are called *infinite traces*. Visually, infinite traces can be pictured as heaps obtained as in Fig. 1, but with infinitely many pieces piled up.

For every $x \in \mathcal{M}_\Sigma$, we define the *visual cylinder* of base x as the following subset of $\partial\mathcal{M}_\Sigma$: $\uparrow x = \{\xi \in \partial\mathcal{M}_\Sigma : x \leq \xi\}$.

Via the embedding $\mathcal{M}_\Sigma \rightarrow \overline{\mathcal{M}}_\Sigma$, the family $(B_{\Sigma,p})_{p \in (0,p_\Sigma)}$ can be seen as a family of discrete distributions on the compactification $\overline{\mathcal{M}}_\Sigma$ rather than on \mathcal{M}_Σ . Standard techniques from functional analysis allow to prove the weak convergence of $B_{\Sigma,p}$, when $p \rightarrow p_\Sigma$, toward a probability measure \mathbf{B} on $\partial\mathcal{M}_\Sigma$, characterized by the following Bernoulli-like identities [3, 4] for all $x \in \mathcal{M}_\Sigma$:

$$\mathbf{B}(\uparrow x) = p_\Sigma^{|x|}. \quad (4)$$

Definition 2.1. The probability measure \mathbf{B} on $\partial\mathcal{M}_\Sigma$ is the *uniform measure at infinity*.

So far, we have thus defined a family of probability measures $B_{\Sigma,p}$ on $\overline{\mathcal{M}}_\Sigma$, for p ranging over the open interval $(0, p_\Sigma)$, completed by a probability measure \mathbf{B} . Note the alternative: $B_{\Sigma,p}$ is concentrated on \mathcal{M}_Σ for all $p < p_\Sigma$; whereas \mathbf{B} is concentrated on $\partial\mathcal{M}_\Sigma$.

3 Random generation of traces

For the random generation of infinite traces, we consider as a first task the random generation of finite traces according to a probability distribution $B_{\Sigma,p}$ for $p \in (0, p_\Sigma)$. We target an incremental procedure, where elements of Σ are added one after the other.

We fix an arbitrary element $a_1 \in \Sigma$. The *link* $\mathcal{L}(a_1)$ of a_1 is defined by:

$$\mathcal{L}(a_1) = \{b \in \Sigma : (a_1, b) \in \mathcal{R}\}; \quad (5)$$

the *a_1 -pyramidal elements* of \mathcal{M} are the traces belonging to the following subset:

$$\text{Pyr}_\Sigma(a_1) = \{z \cdot a_1 : z \in \mathcal{M}_{\Sigma \setminus \{a_1\}} \text{ and } \max(z) \subseteq \mathcal{L}(a_1)\}. \quad (6)$$

Hence, a trace $x \in \text{Pyr}_\Sigma(a_1)$ has only one occurrence of a_1 . This occurrence is the only maximal element of the heap representing x , as illustrated in Fig. 2 (i). For a generic trace x , the successive occurrences of a_1 within x are associated with a_1 -pyramidal elements, which yields the following decomposition result.

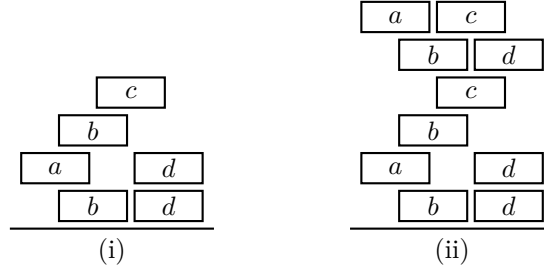


Figure 2: (i) Heap representing the c -pyramidal element $b \cdot a \cdot b \cdot d \cdot d \cdot c$ in the trace monoid \mathcal{M}_Σ , where (Σ, \mathcal{R}) is as in Figure 1. (ii) An element $x \in \mathcal{M}_\Sigma$ whose decomposition through c -pyramidal traces is $(babddc) \cdot (bdc) \cdot a$.

Proposition 3.1. *Let \mathcal{M}_Σ be a trace monoid, let $a_1 \in \Sigma$, and let $x \in \mathcal{M}_\Sigma$. There exists a unique integer $k \geq 0$, given by $k = |x|_{a_1}$ (the number of occurrences of a_1 in x), and a unique tuple $(u_0, \dots, u_{k-1}, u_k)$ such that:*

$$(1) \ u_0, \dots, u_{k-1} \in \text{Pyr}_\Sigma(a_1), \quad (2) \ u_k \in \mathcal{M}_{\Sigma \setminus \{a_1\}}, \quad (3) \ x = u_0 \cdot \dots \cdot u_k.$$

For example, if $a_1 = c$, the element $x = b \cdot a \cdot b \cdot d \cdot d \cdot c \cdot b \cdot d \cdot a \cdot c$, represented in Fig. 2 (ii), is decomposed as the product $u_0 \cdot u_1 \cdot u_2$ of a_1 -pyramidal elements given by $u_0 = b \cdot a \cdot b \cdot d \cdot d \cdot c$, $u_1 = b \cdot d \cdot c$ and $u_2 = a$.

Based on the decomposition of traces given by Proposition 3.1 on the one hand, and on the inversion formula (1) on the other hand, we obtain the following generation procedure.

Theorem 3.2. *Let \mathcal{M}_Σ be a trace monoid and let $p \in (0, p_\Sigma)$. Assume that, for all subsets X of Σ , the real $\mu_X(p)$ has been precomputed. Then, Algorithm 1 is a random recursive algorithm which, provided with the input (Σ, Σ) , outputs an element $\xi \in \mathcal{M}_\Sigma$ distributed according to $B_{\Sigma, p}$.*

We assume that every function call, variable assignation and multiplication in \mathcal{M}_Σ takes a constant number of steps, and that the call to a routine outputting a random integer X takes a number of steps bounded by X . Then, Algorithm 1 outputs an element ξ of \mathcal{M}_Σ in $\mathcal{O}(|\Sigma|(|\xi| + 1))$ steps.

From now on, we assume that the monoid \mathcal{M}_Σ is *irreducible*, meaning that the graph (Σ, \mathcal{R}) is connected.

It must be noted that a naive approach, consisting for example in concatenating *i.i.d.* samples of elements in \mathcal{M}_Σ distributed according to some distribution $B_{\Sigma, p}$ with $p < p_\Sigma$, would not yield infinite traces uniformly distributed in general: see a counter-example in [2, § 6.1.2]. Instead, we make use of an extension of Proposition 3.1 for infinite traces and obtain the following result.

Theorem 3.3. *Let \mathcal{M}_Σ be an irreducible trace monoid and let $a_1 \in \Sigma$. Then, $p_\Sigma < p_{\Sigma \setminus \{a_1\}}$ and, if we use p_Σ in place of p , Algorithm 1 is well-defined for all inputs (S, T) such that $S \subseteq \Sigma \setminus \{a_1\}$.*

Algorithm 1 Outputs $\xi \in \mathcal{M}_S$ distributed according to $B_{S,p}(\cdot \mid \max(\xi) \subseteq T)$

Require: Real parameter $p \in (0, p_S)$, Subsets S and T of Σ

```

1: if  $S \cap T = \emptyset$  then
2:   return  $\mathbf{e}$  ▷  $\mathbf{e}$  is the unit element of the monoid
3: else
4:   choose  $a_1 \in S \cap T$ 
5:    $r \leftarrow 1 - \mu_S(p) / \mu_{S \setminus \{a_1\}}(p)$ 
6:    $K \leftarrow \mathcal{G}(r)$  ▷ Random integer with a geometric law
7:    $\xi \leftarrow \mathbf{e}$  ▷ Initialization with the unit element of the monoid
8:   for  $i = 0$  to  $K - 1$  do
9:      $v \leftarrow$  output of Algorithm 1 on input  $(p, S \setminus \{a_1\}, \mathcal{L}(a_1))$ 
10:     $\xi \leftarrow \xi \cdot v \cdot a_1$ 
11:    $u \leftarrow$  output of Algorithm 1 on input  $(p, S \setminus \{a_1\}, T)$ 
12:    $\xi \leftarrow \xi \cdot u$ 
13: return  $\xi$ 

```

Moreover, Algorithm 2 is a random endless algorithm that outputs at its k^{th} loop an element $\xi_k \in \mathcal{M}_\Sigma$ with the following properties:

1. $(\xi_k)_{k \geq 1}$ is a non-decreasing sequence.
2. The element $\xi = \bigvee_{k \geq 1} \xi_k$, i.e., the equivalence class of $(\xi_k)_{k \geq 1}$ for the relation \asymp , is an infinite trace distributed according to the uniform measure \mathbf{B} .
3. Under the same assumptions as in Theorem 3.2, the first k loops require the execution of $\mathcal{O}(|\Sigma| |\xi_k|)$ steps overall, and the average and minimal sizes of ξ_k are linear in k . Hence, the algorithm produces on average a constant number of additional elements of Σ by unit of time.

Algorithm 2 Outputs approximation of $\xi \in \partial \mathcal{M}_\Sigma$ distributed according to \mathbf{B}

Require: — ▷ No input

```

1:  $\xi \leftarrow \mathbf{e}$  ▷ Initialization with the unit element of the monoid
2: repeat forever:
3:    $v \leftarrow$  output of Algorithm 1 on input  $(p_\Sigma, \Sigma \setminus \{a_1\}, \mathcal{L}(a_1))$ 
4:    $\xi \leftarrow \xi \cdot v \cdot a_1$ 
5:   output  $\xi$  ▷ Writes on a register

```

One might be concerned by the fact that the sequence $(\xi_k)_{k \geq 1}$, output of Algorithm 2, has a particular “shape”, since it is the concatenation of a_1 -pyramidal traces. For instance, if one wishes to use it for parametric estimation or to sample some statistics on traces, the result could *a priori* depend on the choice of a_1 . But asymptotically, for a large class of statistics, the result will not depend on the choice of a_1 ; a precise justification of this fact can be found in [1]. For instance, the density of appearance of an arbitrary letter in an infinite trace can be approximated in this way.

References

- [1] S. Abbes. A cut-invariant law of large numbers for random heaps. *Journal of Theoretical Probability*, 30(4):1692–1725, 2017.
- [2] S. Abbes. Synchronization of Bernoulli sequences on shared letters. *Information and Computation*, 255(1):1–26, 2017.
- [3] S. Abbes and J. Mairesse. Uniform and Bernoulli measures on the boundary of trace monoids. *Journal of Combinatorial Theory, Series A*, 135:201–236, 2015.
- [4] S. Abbes and J. Mairesse. Uniform generation in trace monoids. In G. Italiano, G. Pighizzini, and D. Sannella, editors, *Mathematical Foundations of Computer Science 2015 (MFCS 2015), part 1*, volume 9234 of *LNCS*, pages 63–75. Springer, 2015.
- [5] P. Cartier and D. Foata. *Problèmes combinatoires de commutation et réarrangements*, volume 85 of *Lecture Notes in Mathematics*. Springer, 1969.
- [6] P. Csikvári. Note on the smallest root of the independence polynomial. *Combinatorics, Probability and Computing*, 22(1):1–8, 2013.
- [7] V. Diekert. *Combinatorics on traces*, volume 454 of *LNCS*. Springer, 1990.
- [8] V. Diekert and G. Rozenberg, editors. *The Book of Traces*. World Scientific, 1995.
- [9] M. Goldwurm and M. Santini. Clique polynomials have a unique root of smallest modulus. *Information Processing Letters*, 75(3):127–132, 2000.
- [10] C. Krattenthaler. The theory of heaps and the Cartier-Foata monoid. In *Electronic reedition of “Problèmes combinatoires de commutation et réarrangements”*, pages 63–73. EMIS, 2006.
- [11] D. Krob, J. Mairesse, and I. Michos. Computing the average parallelism in trace monoids. *Discrete Mathematics*, 273:131–162, 2003.
- [12] X. Viennot. Heaps of pieces, i: basic definitions and combinatorial lemmas. In G. Labelle and P. Leroux, editors, *Combinatoire énumérative*, volume 1234 of *Lecture Notes in Mathematics*, pages 321–350. Springer, 1986.