On the Average Complexity of Brzozowski's Algorithm for Deterministic Automata with a Small Number of Final States

Sven de Felice, Cyril Nicaud

On the Average Complexity of Brzozowski's Algorithm for Deterministic Automata with a Small Number of Final States*

Sven De Felice[1] and Cyril Nicaud[2]

[1] LIAFA, Université Paris Diderot - Paris 7 & CNRS UMR 7089, France
Sven.De-Felice@liafa.univ-paris-diderot.fr
[2] LIGM, Université Paris-Est & CNRS UMR 8049, France
nicaud@univ-mlv.fr

Abstract. We analyze the average complexity of Brzozowski's minimization algorithm for distributions of deterministic automata with a small number of final states. We show that, as in the case of the uniform distribution, the average complexity is super-polynomial even if we consider random deterministic automata with only one final state. We therefore go beyond the previous study where the number of final states was linear in the number of states. Our result holds for alphabets with at least 3 letters.

1 Introduction

In this article we continue our investigation of the average complexity of Brzozowski's algorithm [**?**] that was started in [**?**]. Recall that Brzozowski's method is based on the fact that determinizing a trim co-deterministic automaton that recognizes a language $\mathcal{L}$ yields the minimal automaton for $\mathcal{L}$. Hence, starting from an automaton $\mathcal{A}$ that recognizes the language $\mathcal{L}$, one can compute its minimal automaton by first determinizing its reversal, then by determinizing the reversal of the resulting automaton.

This elegant method is not efficient in the worst case, since the first determinization can produce an automaton that has exponentially many states, even if one starts with a deterministic automaton (see [**?**] for a classical example). We are therefore far from the efficient solutions available to minimize deterministic automata, such as Hopcroft's algorithm [**?**], which runs in $\mathcal{O}(n \log n)$ time.

In [**?**] we proved that for the uniform distribution on deterministic and complete automata with n states, or for distributions where each state is final with (fixed) probability $b \in (0, 1)$, the running time of Brzozowski's algorithm is super-polynomial[3] with high probability. One limitation of this result is that under such a distribution, an automaton with n states has around bn final states, for fixed

* This work is supported by the French National Agency (ANR) through ANR-10-LABX-58 and through ANR-2010-BLAN-0204.

[3] Grows quicker than n^d for any positive d.

b, which therefore grows linearly with the number of states. However, in many situations the automata that are built do not have that many final states (see, for instance, Aho-Corasick automaton [**?**], which is used for pattern matching). A natural question is whether this result still holds for automata with, for instance, a fixed number of final states. This is the question we investigate in this article.

The precise definition of a *distribution of automata with a small number of final states* is given in Section 4, but it covers the cases of random size-n automata with just one final state, with $\log n$ final states, or where each state is final with probability $\frac{3}{n}$ or $\frac{2}{\sqrt{n}}$, and so on. It therefore differs significantly from the cases studied in [**?**].

Notice that analyzing distributions of automata with a small number of final states is an up-to-date question in the statistical study of automata. The main results in this field, the average complexity of Moore's algorithm and the asymptotic number of minimal automata, only hold for distributions of automata with "sufficiently many" final states [**?**,**?**,**?**]. Some effort have been undertaken to extend them to, say, automata with only one final state, but with no success so far. To our knowledge, we present in this article the first result of this kind.

We will see that the proof of our main result is not just simply an adaptation of the proof proposed in [**?**] and we will need some deeper understanding of the typical properties of a random automaton. In return, we will establish some new facts that are interesting on their own, and that may be reused for further work on statistical properties of random automata.

The paper is organized as follows. After recalling some basic definitions in Section 2, we briefly revisit the article [**?**] in Section 3 to point out the difficulties encountered when trying to reduce the number of final states. In Section 4 we state our main result and prove it for automata with only one final state in Section 5. In Section 6, we explain how to extend it to get the full proof.

2 Definitions

Let $[n]$ denote the set of integers between 1 and n. If x, y are two real numbers, let $[\![x, y]\!]$ denote the set of integers i such that $x \leq i \leq y$. For any positive integer n, let $\mathfrak{S}_n$ denote the set of all permutations on $[n]$.

Automata. Let A be a finite alphabet, an *automaton* $\mathcal{A}$ is a tuple (Q, δ, I, F), where Q is its finite set of *states*, $I \subseteq Q$ is its set of *initial states* and $F \subseteq Q$ is its set of *final states*. Its *transition function* δ is a (partial) map from $Q \times A$ to 2^Q. A *transition* of $\mathcal{A}$ is a tuple $(p, a, q) \in Q \times A \times Q$, which we write $p \xrightarrow{a} q$, such that $q \in \delta(p, a)$. The map δ is classically extended by morphism to $Q \times A^*$. We denote by $\mathcal{L}(\mathcal{A})$ the set of words recognized by $\mathcal{A}$. A *deterministic and complete automaton* is an automaton such that $|I| = 1$ and for every $p \in Q$ and $a \in A$, $|\delta(p, a)| = 1$; for such an automaton we consider that δ is a (total) map from $Q \times A^*$ to Q to simplify the notations. A state q is *accessible* when there exists a path from an initial state to q. It is *co-accessible* when there exists a path from

q to a final state. If $\mathcal{A}$ is an automaton, we let $\mathrm{Trim}(\mathcal{A})$ denote the automaton obtained after removing states that are not accessible or not co-accessible.

For any automaton $\mathcal{A} = (Q, \delta, I, F)$, we denote by $\tilde{\mathcal{A}}$ the *reverse* of $\mathcal{A}$, which is the automaton $\tilde{\mathcal{A}} = (Q, \tilde{\delta}, F, I)$, where $p \xrightarrow{a} q$ is a transition of $\tilde{\mathcal{A}}$ if and only if $q \xrightarrow{a} p$ is a transition of $\mathcal{A}$. The automaton $\tilde{\mathcal{A}}$ recognizes the reverse[4] of $\mathcal{L}(\mathcal{A})$. An automaton is *co-deterministic* when its reverse is deterministic.

Recall that the *minimal automaton* of a rational language $\mathcal{L}$ is the smallest deterministic and complete automaton[5] that recognizes $\mathcal{L}$. To each rational language $\mathcal{L}$ corresponds a minimal automaton, which is unique up to isomorphism.

Subset construction and Brzozowski's algorithm. If $\mathcal{A} = (Q, \delta, I, F)$ is a non-deterministic automaton, it is classical that the subset automaton of $\mathcal{A}$ defined by

$$\mathcal{B} = \left(2^Q, \gamma, \{I\}, \{X \in 2^Q \mid F \cap X \neq \emptyset\}\right)$$

is a deterministic automaton that recognizes the same language, where for every $X \in 2^Q$ and every $a \in A$, $\gamma(X, a) = \cup_{p \in X} \delta(p, a)$. This is of course still true if we only take the accessible part of $\mathcal{B}$, and this is not a difficulty when implementing it, since the accessible part of $\mathcal{B}$ can be built on the fly, using the rule for γ in a depth-first traversal of $\mathcal{B}$ starting from I. We denote by $\mathrm{Subset}(\mathcal{A})$ the accessible part of the subset automaton of $\mathcal{A}$.

In [**?**], Brzozowski established the following result:

Theorem 1 (Brzozowski). *If $\mathcal{A}$ is a trim co-deterministic automaton then $\mathit{Subset}(\mathcal{A})$ is the minimal automaton of $\mathcal{L}(\mathcal{A})$.*

This theorem readily yields an algorithm to compute the minimal automaton of the language recognized by an automaton $\mathcal{A}$, based on the subset construction: since $\mathcal{B} = \mathrm{Subset}(\mathrm{Trim}(\tilde{\mathcal{A}}))$ is a deterministic automaton recognizing the mirror of $\mathcal{L}(\mathcal{A})$, then $\mathrm{Subset}(\mathrm{Trim}(\tilde{\mathcal{B}}))$ is the minimal automaton of $\mathcal{L}(\mathcal{A})$.

Mappings. A *mapping* of size n is a total function from $[n]$ to $[n]$. A mapping f can be seen as a directed graph with an edge $i \to j$ whenever $f(i) = j$. Such a graph is a union of cycles of Cayley trees (i.e., rooted labelled trees), as depicted in Fig. 1 (see [**?**] for more information on this graph description). Let f be a size-n mapping. An element $x \in [n]$ is a *cyclic point* of f when there exists an integer $i > 0$ such that $f^i(x) = x$. The *cyclic part* of a mapping f is the permutation obtained when restricting f to its set of cyclic points. The *normalized cyclic part* of f is obtained by relabelling the c cyclic points of f with the elements of $[c]$, while keeping their relative order[6].

Automata as combinatorial structures. In the sequel, A is always a fixed alphabet with $k \geq 2$ letters. Let $\mathfrak{A}_n$ (or $\mathfrak{A}_n(A)$ when we want to make precise the alphabet) denote the set of all deterministic and complete automata with

[4] If $u = u_0 \cdots u_{n-1}$ is a word of length n, the *reverse* of u is the word $\tilde{u} = u_{n-1} \cdots u_0$.

[5] Minimal automata are not always required to be complete in the literature.

[6] The notion of normalization will be used for other substructures, always for relabelling the atoms with an initial segment of the positive integers, while keeping their relative order.

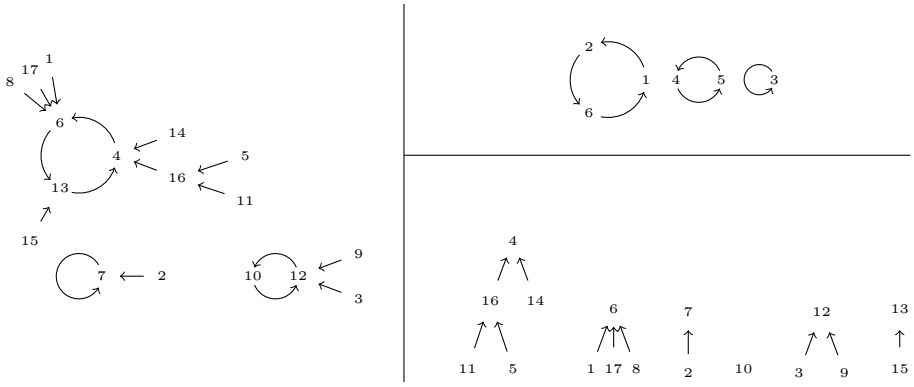

Fig. 1. A mapping of size 17, on the left. On the upper right we have its normalized cyclic part, and on the lower right its Cayley trees (not normalized).

input alphabet A whose set of states is $[n]$ and whose initial state is 1. Such an automaton $\mathcal{A}$ is characterized by the tuple (n, δ, F). A *transition structure* is an automaton without final states, and we denote by $\mathfrak{T}_n$ the set of n-state transition structures with the same label restrictions as for $\mathfrak{A}_n$. If $\mathcal{A} \in \mathfrak{A}_n$, an a-cycle of $\mathcal{A}$ is a cycle of the mapping induced by a, i.e., $p \mapsto \delta(p, a)$. A state of an a-cycle is called an *a-cyclic state*.

Distributions of combinatorial structures. Let E be a set of combinatorial objects with a notion of size such that the set E_n of elements of size n of E is finite for every $n \geq 0$. The *uniform distribution* (which is a slight abuse of notation since there is one distribution for each n) on the set E is defined for any $e \in E_n$ by $\mathbb{P}_n(\{e\}) = \frac{1}{|E_n|}$. The reader is referred to [**?**] for more information on combinatorial probabilistic models.

Probabilities on automata. Let A be an alphabet. We consider two kinds of distribution on size-n deterministic and complete automata. The first one is the *fixed-size distribution* on $\mathfrak{A}_n$ of parameter m. It is the uniform distribution on size-n automata with exactly m states. The parameter m may depend[7] on n; one can for instance consider the fixed-size distribution of parameter $\lfloor \sqrt{n} \rfloor$. The second one is the *p-distribution* on $\mathfrak{A}_n$, where the transition structure of the automaton is chosen uniformly at random and where each state is final with probability p independently; in this model also, p may depend on n, for instance $p = \frac{2}{n}$ yields automata with two final states on average.

Note that the Bernoulli model of parameter b of [**?**] is the same as the p-distribution for $p = b$: it is the case where p does not depend on n.

Some terminology. We consider a (sequence of) distributions on $E = \bigcup_n E_n$. Let P be a property defined on E. We say that P holds *generically* (or *with*

[7] The term "fixed" stands for: for any given n, the number of final states is fixed.

high probability) when the probability it holds tends to 1 as n tends to infinity. We say that P is *visible* (or holds with *visible probability*) when there exists a positive constant C and an integer n_0 such that for every $n \geq n_0$ the probability that P holds on size-n elements is at least C.

In the sequel we will implicitly use that if P and Q are generic then $P \wedge Q$ is generic; if P is generic and Q is visible, then $P \wedge Q$ is visible; if P and Q are visible and independent, then $P \wedge Q$ is visible, and so on.

3 Result for a large number of final states

In [**?**] we proved that the complexity of Brzozowski's algorithm is generically super-polynomial for the uniform distribution on deterministic complete automata with n states. For this distribution, every state is final with probability $\frac{1}{2}$. Thus if one take such an automaton uniformly at random, with high probability it has around $\frac{n}{2}$ final states. The article also considers the case where the probability of being final is some fixed $b \in (0, 1)$.

In this paper we consider distributions on $\mathfrak{A}_n$ where the typical number of final states can be small, for instance in $o(n)$, and we will try to reuse some ideas from [**?**] to do the analysis. We therefore recall in this section, very briefly, the proof of the following result:

Theorem 2. (De Felice, Nicaud [?]) *Let A be an alphabet with at least 2 letters. If $\mathcal{L}$ is the language recognized by a deterministic and complete n-state automaton over A taken uniformly at random, then generically the minimal automaton of the mirror of $\mathcal{L}$ has a super-polynomial number of states.*

The first observation is the following lemma, which will be used in a slightly modified version in this paper. It is the only result needed from automata theory; the remainder of the proof consists in analyzing the combinatorial structure of the underlying graph of a random automaton. A cycle is *primitive* when the sequence of types of states (final and non-final) in the cycle forms a primitive word.

Lemma 1 ([?]). *Let $\mathcal{A} \in \mathfrak{A}_n$ be a deterministic automaton that contains m primitive a-cycles $\mathcal{C}_1, \ldots, \mathcal{C}_m$ of lengths at least two that are all accessible. The minimal automaton of $\mathcal{L}(\tilde{\mathcal{A}})$ has at least* $\mathrm{lcm}(|\mathcal{C}_1|, \ldots, |\mathcal{C}_m|)$ *states.*

The proof of Theorem 2 is organized as follows:

1. If we look just at the action of a in a random automaton, it is a random mapping from the set of states to itself. Using the classical properties of random mappings [**?**] we get that, with high probability, there are at least $n^{1/3}$ a-cyclic states.
2. The cyclic part of a uniform random mapping behaves likes a uniform random permutation. We therefore want to use a celebrated result of Erdős and P. Turán [**?**] which states that the lcm of the lengths of the cycles of a random permutation[8] is super-polynomial with high probability.

[8] It is exactly the order of the permutation.

3. We prove that the a-cycles of lengths at least $\log n$ are generically primitive and accessible in a random automaton, using properties of random automata established in [**?**].
4. We conclude by proving that even if we remove the cycles of lengths smaller than $\log n$ in Erdős and P. Turán's result, we still have a super-polynomial lcm with high probability.

Now assume that we are considering the uniform distribution on automata with just one final state. It is no longer true that the large cycles are generically primitive: with high probability the final state is not in any a-cycle, which is therefore not primitive. In the sequel we will show how to get around this problem, which has consequences at every step of the proof (in particular we can no longer use the result of Erdős and P. Turán).

4 Main result

A distribution on automata is said to have *a small number of final states* when it is either a fixed size distribution or a p-distribution on size-n automata such that the number of final states is in $[\![1, \frac{n}{2}]\!]$ with visible probability. Our main result is the following:

Theorem 3. *Let A be an alphabet with at least 3 letters. If $\mathcal{L}$ is the language recognized by a random n-state deterministic and complete automaton following a distribution with a small number of final states, then for any $d > 0$, the minimal automaton of the mirror of $\mathcal{L}(\mathcal{A})$ has, with visible probability, more than n^d states.*

Compared to the main result of [**?**] we capture many more distributions on automata, by weakening the statement a bit: it holds for an alphabet of 3 or more letters and it does not hold generically but with positive probability. The latter is unavoidable: as proved in [**?**] there is a linear number of states that are not accessible in a typical random automaton. Thus for the fixed size distribution with one final state, the final state has a positive probability of not being accessible.

The average complexity of Brzozowski's algorithm is a direct consequence of Theorem 3.

Corollary 1. *Let A be an alphabet with at least 3 letters. The average complexity of Brzozowski's algorithm is super-polynomial for distributions with a small number of final states.*

Proof. For any $d > 0$, the expected number of states after the first determinization is at least n^d times the probability that an automaton has at least n^d states after the first determinization. This probability is greater than some positive constant C for n sufficiently large by Theorem 3, concluding the proof. □

Our results hold, for instance, for the p-distributions with $p = \frac{\alpha}{n}$ for some positive real α: there are α final states on average, and it is straightforward to check that it has a small number of final states. They also hold for the fixed-size distribution with $\lfloor\sqrt{n}\rfloor$ final states, since $1 \leq \lfloor\sqrt{n}\rfloor \leq \frac{n}{2}$ for n sufficiently large.

5 Proof of Theorem 3 for automata with one final state

In this section we prove our main theorem for the fixed-size distribution of parameter 1, that is, for the uniform distribution on the sets $\mathfrak{A}_n^1$ of automata with exactly one final state. From now on we are working over the alphabet $A = \{a, b, c\}$, as adding more letters just makes the problem easier.

We start with a generalization of Lemma 1 from [**?**]. Let $\mathcal{A}$ be an automaton with transition function δ and set of final states F, let $\mathcal{C}$ be an a-cycle of $\mathcal{A}$ of length ℓ and let $u \in A^*$. The *u-word of $\mathcal{C}$* is the word $v = v_0 \dots v_{\ell-1}$ of length ℓ on $\{0, 1\}$ defined as follows: let x be the smallest element of $\mathcal{C}$, we set $v_i = 1$ if and only if $\delta(x, a^i u) \in F$, for $i \in \{0, \dots, \ell - 1\}$. This is a generalization of [**?**] where the word associated with $\mathcal{C}$ is exactly the ε-word of $\mathcal{C}$. The cycle $\mathcal{C}$ is *u-primitive* when its u-word is a primitive word.

Lemma 2. *Let u be a word of A^* and let $\mathcal{A} \in \mathfrak{A}_n$ be a deterministic automaton that contains m u-primitive a-cycles $\mathcal{C}_1, \dots, \mathcal{C}_m$ of lengths at least two that are all accessible. The minimal automaton of $\mathcal{L}(\tilde{\mathcal{A}})$ has at least* $\operatorname{lcm}(|\mathcal{C}_1|, \dots, |\mathcal{C}_m|)$ *states.*

Proof. (sketch) This is the same proof as in [**?**], except that we are considering the sets $\delta^{-1}(F, u) \cap \mathcal{C}$ instead of $F \cap \mathcal{C}$, where $\mathcal{C} = \cup_{i=1}^m \mathcal{C}_i$. □

In the sequel we first find a suitable word u, then a collection of a-cycles with good properties, in order to apply Lemma 2.

5.1 Finding $u \in \{b, c\}^*$ such that $\delta_{\mathcal{A}}^{-1}(f, u)$ is sufficiently large

For the first step, we consider letters b and c only and try to build a sufficiently large set of states in the determinization of the mirror of an automaton with one final state. In this section we prove the following result.

Proposition 1. *There exists a constant $w > 0$ such that if we draw an element $\mathcal{A}$ of $\mathfrak{A}_n^1(\{b, c\})$ uniformly at random, there exists a word $u \in \{b, c\}^*$ such that $\delta_{\mathcal{A}}^{-1}(f, u)$ has size between $w\sqrt{n}$ and $w\sqrt{n}\log n$ with visible probability, where f denotes the final state of $\mathcal{A}$.*

Of course, in Proposition 1 the word u depends on $\mathcal{A}$. We need some preliminary results on random mappings to establish the proposition.

Lemma 3. *Generically, a random mapping of $[n]$ has no element with more than* $\log n$ *preimages.*

Lemma 3 is used the following way. If we find a word v such that $\delta_{\mathcal{A}}^{-1}(f, v)$ has size greater than $w\sqrt{n}\log n$, then there exists a prefix u of v such that $w\sqrt{n} \leq \delta_{\mathcal{A}}^{-1}(f, u) \leq w\sqrt{n}\log n$ since the image of a set of states X by a letter in $\tilde{\mathcal{A}}$ generically has size at most $|X| \log n$. Hence, we just have to find a word v such that $\delta_{\mathcal{A}}^{-1}(f, v) \geq w\sqrt{n}$ to conclude the proof.

A set of vertices X of a digraph is *stable* when there is no edge $x \to y$ for $x \in X$ and $y \notin X$.

Lemma 4. *Let $\mathcal{A}$ be an element of $\mathfrak{A}_n^1$ taken uniformly at random, and let G be the digraph induced on $[n]$ by the actions of b and c (there is an edge $x \to y$ if and only if $\delta_{\mathcal{A}}(x,b) = y$ or $\delta_{\mathcal{A}}(x,c) = y$). Generically, G has a unique stable strongly connected component, which has size greater than $\frac{1}{2}n$.*

Proof. (sketch) We first prove that generically there is no stable set of states of size smaller than $\frac{1}{4}n$: we overcount the number of transition structures having a stable subset X of size ℓ by choosing the ℓ states, their images by both letters in X and the images of the other states. This yields an upper bound of $\binom{n}{\ell}\ell^{2\ell}n^{2n-2\ell}$ for the number of such transition structures. Summing for ℓ from 1 to $n/4$ this upper bound is sufficient to prove that it generically does not happen.

It is proven in [**?**] that in a random transition structure with n states on a two-letter alphabet, the accessible part is generically of size greater than $\frac{1}{2}n$. If we have a transition structure with a stable strongly connected component $\mathcal{C}$ of size between $\frac{1}{4}n$ and $\frac{1}{2}n$, then by symmetry, the initial state is in $\mathcal{C}$ with probability at least $\frac{1}{4}$. But in such a case, the accessible part has size at most $\frac{1}{2}n$, which generically cannot happen according to [**?**]. This concludes the proof, as there can be at most one component of size greater $\frac{1}{2}n$. □

Remark 1. If an automaton has a unique stable strongly connected component $\mathcal{C}$, then for every state q there exists a path from q to any state of $\mathcal{C}$, as one can see on the acyclic graph of strongly connected components. In particular, $\mathcal{C}$ is necessarily accessible.

Recall that a mapping f on $[n]$ can be seen as a union of cycles of Cayley trees. Define the *largest tree* of f as its largest Cayley tree, taking the tree with the smallest root label if there are several trees with the maximum number of nodes. In a transition structure or in an automaton, the largest b-tree is the largest tree for the mapping associated with the action of b. Our next lemma states that the largest b-tree of a random structure behaves like a uniform random tree (when there is only one tree of maximum size). Thus we can use classical results on random mappings and Cayley trees to estimate the typical width of such a tree.

Lemma 5. *Let t and n be two integers such that $1 \leq \frac{n}{2} < t \leq n$ and let $\mathfrak{M}_n^{(t)}$ denote the set of mapping on $[n]$ whose largest tree has t nodes. The normalized largest tree of a uniform element of $\mathfrak{M}_n^{(t)}$ is distributed as a uniform random Cayley tree with t nodes.*

The following result is the mix between a classical result on the largest tree in a random mapping [**?**] and the analysis of the width of a random Cayley tree done in [**?**].

Theorem 4 (Kolčin, Chassaing and Marckert). *There exist two positive constants w and C such that the probability that the largest tree of a random mapping on $[n]$ has width at least $w\sqrt{n}$ is greater than C, for n sufficiently large.*

We can now give the proof of Proposition 1.

Proof. (of Proposition 1) By Theorem 4, there is a b-tree T of width at least $w\sqrt{n}$ in a random size-n transition structure with input alphabet $A' = \{b, c\}$ with positive probability. Moreover, by Lemma 4 such a transition structure generically has only one stable strongly connected component $\mathcal{C}$, which contains more than $\frac{1}{2}n$ states. Hence if we add a final state uniformly at random, it is in $\mathcal{C}$ with probability at least $\frac{1}{2}$. As stated in Remark 1, if the final state is in the unique stable strongly connected component, then there exists a word v that labels a path from the root of T to f. Consider the word $v' = \tilde{v}b^i$, where i is the layer of T with the maximal number of nodes (the level that gives its width). Then $\delta^{-1}(f, v')$ contains all the states of the ith layer of T, and it therefore contains at least $w\sqrt{n}$ elements.

By Lemma 3, every state has generically less than $\log n$ preimages. Thus if $|\delta^{-1}(f, v')| \geq w\sqrt{n}$, then there exists a prefix u of v' such that $\delta^{-1}(f, u)$ contains between $w\sqrt{n}$ and $w\sqrt{n}\log n$ elements, concluding the proof. □

5.2 Finding a good collection of a-cycles

We now switch to a-cycles and consider only the action of the letter a (the actions of the three letters are independent). Recall that conditioned by its number of cyclic points m, the cyclic permutation of a uniform random mapping is a uniform permutation of $\mathfrak{S}_m$. We first establish some properties of random permutations.

If σ is a permutation of $[n]$, its *sequence of cycles* is the ordered sequence of its cycles $(\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m)$, where the cycles are ordered by their smallest element. If $(\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m)$ is the sequence of cycles of σ and $d \leq m$, the *d first cycles of σ* are the cycles $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_d$. Let $L_d(\sigma) = (|\mathcal{C}_1|, \ldots, |\mathcal{C}_d|)$ denote the d first cycles of σ, when σ has at least d cycles and let $L_d(\sigma) = \perp$ otherwise.

Lemma 6. *Let d be a positive integer. For any $(\ell_1, \ldots, \ell_d) \in [\![\frac{n}{3d}, \frac{n}{2d}]\!]^d$, the following lower bound holds for n sufficiently large:*

$$\mathbb{P}\big(L_d = (\ell_1, \ldots, \ell_d)\big) \geq \frac{1}{n^d}.$$

We now turn our attention to the lcm of the first d cycles of a random permutation, and establish the following proposition.

Proposition 2. *Let $(x_1, \ldots, x_d)$ be a uniform element of $[\![\frac{n}{3d}, \frac{n}{2d}]\!]^d$. There exists a constant $\lambda > 0$ such that $\operatorname{lcm}(x_1, \ldots, x_d) \geq \lambda n^d$ with visible probability.*

Let $\mathrm{Cycle}_d(n)$ be the set of permutations σ of $[n]$ such that $L_d(\sigma) \in [\![\frac{n}{3d}, \frac{n}{2d}]\!]^d$ and $\operatorname{lcm}(\ell_1, \ldots, \ell_d) \geq \lambda n^d$, with $L_d(\sigma) = (\ell_1, \ldots, \ell_d)$. We use the λ of Proposition 2.

If we take a permutation σ uniformly at random, conditioned by $L_d(\sigma) \in [\![\frac{n}{3d}, \frac{n}{2d}]\!]^d$, the vector $L_d(\sigma)$ is not uniformly distributed in $[\![\frac{n}{3d}, \frac{n}{2d}]\!]^d$. However, we can control the lack of uniformity and use Proposition 2 to obtain sufficiently many permutations such that the lcm of their d first cycles is large enough.

Lemma 7. *For any positive integer d, a uniform random permutation of $[n]$ is in $\mathrm{Cycle}_d(n)$ with visible probability.*

In [?], having generically more than $n^{1/3}$ a-cyclic states was enough to implies the desired result for the uniform distribution. Here we need something more precise. In the next lemma we show that with positive probability, the number of a-cyclic states is in $\Theta(\sqrt{n})$.

Lemma 8. *The cyclic part of a uniform random mapping of size n has size in $[\![\sqrt{n}, 2\sqrt{n}]\!]$ with visible probability.*

Proof. (sketch) We rely on tools from analytic combinatorics [?] applied to the decomposition of a mapping into a union of cycles of Cayley trees, as in [?]. We obtain that the expected number of cyclic points is asymptotically equivalent to $\sqrt{\frac{\pi n}{2}}$, which is already in [?], and that the standard deviation is asymptotically equivalent to $\sqrt{\frac{(4-\pi)n}{2}}$. The result follows by Chebyshev's inequality. □

At this point, we know that with visible probability, the cyclic permutation of the action of a is in $\mathrm{Cycle}_d(i)$ for $i \in [\![\sqrt{n}, 2\sqrt{n}]\!]$. To complete the proof, we need to verify that they are accessible (which is easy since large a-cycles are generically accessible) and that they are sufficiently often u-primitive for the u of Proposition 1.

5.3 Completing the proof

We will use the following lemma to establish the primitivity of the first d a-cycles:

Lemma 9. *Let $n \geq 2$ be an integer and let $i \in [\![1, n-1]\!]$. For the uniform distribution on binary words of length n having i occurrences of the letter 0, the probability that a word is not primitive is smaller than $\frac{2}{n}$.*

We therefore have two independent random sets, $\delta^{-1}(f, u)$ and the union of the first d a-cycles, and are interested in their intersection. The two following lemmas establish that this intersection is not trivial with visible probability. Together with Lemma 9, this will ensure that these a-cycles are u-primitive with positive probability.

Lemma 10. *Let α and β be two positive real numbers. Let X be a subset of $[n]$ of size $\lceil \alpha\sqrt{n} \rceil$ and let Y be a uniform random subset of $[n]$ of size $\lceil \beta\sqrt{n} \rceil$. For every integer $j \geq 0$, there exists a positive constant M_j such that $|X \cap Y| = j$ with probability at least M_j, for n sufficiently large.*

Lemma 11. *Let α be a positive real number. Let X be a subset of $[n]$ of size $m = \lceil \alpha\sqrt{n} \rceil$ and let Y be a uniform random subset of $[n]$ of size m' with $1 \leq m' < \frac{n}{2}$. The probability that $X \subseteq Y$ is smaller than $B\sqrt{n}2^{-\alpha\sqrt{n}}$ for some positive constant B and for n sufficiently large.*

We can now establish the proof of Theorem 3 for automata with one final state as follows. For every $x \in [\![\sqrt{n}, 2\sqrt{n}]\!]$, let $E(x)$ denote the set of mappings of size n whose cyclic part σ has size x and belongs to $\mathrm{Cycle}_d(x)$. By Lemma 8 and Lemma 7, a random mapping is in $\cup_{x \in [\![\sqrt{n}, 2\sqrt{n}]\!]} E(x)$ with visible probability.

Let us fix some mapping f_a of $E(x)$ for the action of a, and let σ_a be its cyclic part. Let $S_1, S_2, \ldots, S_d$ be arbitrary subsets of size $m = \lceil \frac{\sqrt{n}}{3d} \rceil$ of the first d cycles of σ_a, and let $S = \cup_{i=1}^{d} S_i$. Since the actions of b and c are independent of the action of a, by Proposition 2 there exists a word u such that $Y = \delta^{-1}(f, u)$ has size in $[\![w\sqrt{n}, w\sqrt{n}\log n]\!]$ with positive probability. Let Y' be a uniform subset of Y of size $y = \lceil w\sqrt{n} \rceil$. By symmetry, the set Y' is a uniform random subset of size y of $[n]$. Therefore by Lemma 10, with positive probability we have $|S \cap Y'| = d$, and a direct computation shows that this implies that, with positive probability, $|S_i \cap Y'| = 1$ for every $i \in [d]$. Moreover, since $|Y| \leq \frac{n}{2}$, by Lemma 11, the probability that at least one S_i is a subset of Y is smaller than $dB\sqrt{n}2^{-\sqrt{n}}$. Hence, with visible probability, the intersection of Y and S_i is non-trivial for every $i \in [d]$, and so is the intersection of Y and the first d cycles of σ_a (since they contain S_i). Hence, by Lemma 9, there exists a constant $M > 0$ such that the first d cycles are u-primitive with probability at least M for n sufficiently large; and importantly, the value of M is the same for any $x \in [\![\sqrt{n}, 2\sqrt{n}]\!]$ and any $y \in [\![w\sqrt{n}, w\sqrt{n}\log n]\!]$. Therefore, if we sum the contributions for all x and y with the good properties, we get that with visible probability the first d a-cycles are u-primitive (for some word u) and therefore that the lcm of their lengths is at least $\lambda n^{d/2}$.

But the first d a-cycle have lengths greater than $\frac{\sqrt{n}}{2d}$ and are therefore generically accessible (this is Proposition 1 of [**?**]). This concludes the proof by Lemma 2: by choosing $d = 2d' + 1$ there are more than $n^{d'}$ states in the first determinization step of Brzozowski's algorithm with visible probability. □

6 General case

The proof for a general distribution with a small number of final states is not difficult once we have establish the result for the uniform distribution on automata with one final state. We consider two cases depending on whether the automaton has between 1 and $w\sqrt{n}$ final states or between $w\sqrt{n}$ and $\frac{n}{2}$ final states.

For the first case, we select one of the final states f and apply the same construction as in Section 5. With visible probability, we therefore obtain a word u such that $\delta^{-1}(f, u)$ has size at least $w\sqrt{n}$, and therefore $\delta^{-1}(f, u)$ also has size at least $w\sqrt{n}$. Hence, by Lemma 3, there generically exists a prefix u' of u such that $\delta^{-1}(f, u) \in [\![w\sqrt{n}, w\sqrt{n}\log n]\!]$, and we can continue the proof as in Section 5.

The second case is easier. We do not need to build the word u since F is already large enough to apply Lemma 10 and still small enough to apply Lemma 11. The general statement of Theorem 3 follows. □

A natural question is whether the average super-polynomial complexity of Brzozowski's algorithm still holds for alphabets with two letters. The proof of this paper relies on the fact that we built $u \in \{b, c\}^*$ and the a-cycles independently, so that we can apply Lemma 9, Lemma 10 and Lemma 11. If u uses the letter a, we need a more complicated proof that takes the dependency into account, which

is usually difficult. Thereforem the best way is probably to find a completely different approach.

Acknowledgment. We would like to thanks Jean-François Marckert for his patient explanation of his result on the width of random Cayley trees [**?**]. We also thank the referees for providing helpful comments, which helped to improve the quality of this article.