Unambiguous Büchi automata

Olivier Carton
Institut Gaspard Monge, CNRS
5, boulevard Descartes
F-77454 Marne-la-Vallée cedex 2
Olivier.Carton@univ-mlv.fr

Max Michel
CNET
38, rue du Général Leclerc
F-92131 Issy-les-Moulineaux
Max.Michel@cnet.francetelecom.fr

February 28, 2001

Abstract

In this paper, we introduce a special class of Büchi automata called unambiguous. In these automata, any infinite word labels exactly one path going infinitely often through final states. The word is recognized by the automaton if this path starts in an initial state. The main result of the paper is that any rational set of infinite words is recognized by such an automaton. We give two proofs of this result. We also provide several related results.

1 Introduction

Automata on infinite words have been introduced by Büchi [Büc62] in order to prove the decidability of the monadic second-order logic of the ordering the natural numbers. Since then, automata on infinite objects have often been used to prove the decidability of numerous problems. From a more practical point of view, they also lead to efficient decision procedures as for temporal logic [SVW87]. Therefore, automata of infinite words or infinite trees are one of the most important ingredients in model checking tools [Var96]. The complementation of automata is then an important issue since the systems are usually modeled by logical formulas which involve the negation operator.

There are several kinds of automata that recognize sets of infinite words. In 1962, Büchi [Büc62] introduced automata on infinite words, now referred to as *Büchi automata*. These automata have initial and final states and a path is successful if it starts at an initial state and goes infinitely often through final states. However, not all rational sets of infinite words are recognized by a deterministic Büchi automaton [Lan69]. Therefore, complementation is a rather difficult operation on Büchi automata [SVW87].

In 1963, Muller [Mul63] introduced automata, now referred to as *Muller automata*, whose accepting condition is a family of accepting subsets of states. A path is then successful if it starts at the unique initial state and if the set of states which occurs infinitely in the path is accepting. A deep result of McNaughton [McN66] shows that any rational set of infinite words is recognized by a deterministic Muller automaton. A deterministic automaton is unambiguous in the following sense. With each word is associated a canonical path which is the unique path starting at the initial state. A word is then accepted iff its canonical path is successful. In a deterministic Muller automaton, the unambiguity is due to the uniqueness of the initial state and to the determinism of the transitions. Independently, the acceptance condition determines if a path is successful or not. The unambiguity of a deterministic Muller automaton makes it easy to complement. It suffices to exchange accepting and non-accepting subsets of states. However, the main drawback of using deterministic Muller automata is that the acceptance condition is much more complicated. It is a family of subsets of states instead of a simple set of final states. There are other kinds of deterministic automata recognizing all rational sets of infinite words like Rabin automata [Rab69], Street automata or parity automata [Mos84]. In all these automata, the acceptance condition is more complicated than a simple set of final states.

In this paper, we introduce a class of Büchi automata in which any infinite word labels exactly one path going infinitely often through final states. A canonical path can then be associated with each infinite word and we call these automata unambiguous. In these automata, the unambiguity is due to the transitions and to the final states whereas the initial states determine if a path is successful. An infinite word is then accepted iff its canonical path starts at an initial state. The main result of the paper is that any rational set of infinite words is recognized by such an automaton. It is proved in [Arn83] that any rational set of infinite words can be described by an unambiguous rational expression. However, an unambiguous automaton cannot be deduced from an unambiguous rational expression. It turns out that these unambiguous Büchi automata are codeterministic, i.e., reverse deterministic. Therefore, the main result is thus the counterpart of McNaughton's result for codeterministic automata. It has already been proved independently in [Mos82] and [BP85] that any rational set of infinite words is recognized by a codeterministic automaton but the construction given in [BP85] does not provide unambiguous automata.

These unambiguous automata are well suited for boolean operations and especially complementation. Indeed, if the set of initial states of an unambiguous automaton is complemented, the automaton recognizes then the complement of the set of infinite words. Therefore, our construction which yields an unambiguous Büchi automaton can be used for complementation. Furthermore, from a Büchi automaton with n states, our construction provides an unambiguous automaton which has at most $(12n)^n$ states. It has already been proved that in the worst case, the complementation of a Büchi automaton needs $2^{O(n \ln n)}$ states [Mic88]. Our construction is thus optimal.

The unambiguous automata introduced in the paper recognize right-infinite words. However, the construction can be adapted to bi-infinite words. Two unambiguous automata on infinite words can be joined to make an unambiguous automaton on bi-infinite words. This leads to an extension of McNaughton's result to the realm of bi-infinite words.

The main result of this paper has been first obtained by the second author and his proof has circulated as a hand-written manuscript among a bunch of people. It was however never published. Later, the first author found a different proof of the same result based on algebraic constructions on semigroups. Both authors have decided to publish their whole work on this subject together.

The paper is organized as follows. Section 2 is devoted to general notation and basic definitions. Background on automata is recalled in this section. In Section 3, unambiguous Büchi automata are defined. The main result (Theorem 7) is stated there. The first properties of these automata are proved in Section 4. Boolean Operations are studied in Section 5. The two proofs of the main result are postponed to Section 6.

2 Preliminaries

The definitions of notions which appear at several places of the paper are gathered in this section. The main purpose of this section is to fix some notation and terminology.

2.1 General notation

Through the paper, the set of integers, the set of positive integers and the set of relative integers are respectively denoted by $\mathbb{N}$, $\mathbb{N}^*$ and $\mathbb{Z}$. The set $\mathbb{N} \cup \{\infty\}$ is denoted by $\overline{\mathbb{N}}$. The usual order on integers is extended to $\overline{\mathbb{N}}$ by setting $n < \infty$ for any integer n. For an integer n, the finite sets $\{0, \dots, n\}$ and $\{1, \dots, n\}$ are respectively denoted by $\mathbb{N}_n$ and $\mathbb{N}_n^*$ whereas the sets $\mathbb{N}_n \cup \{\infty\}$ and $\mathbb{N}_n^* \cup \{\infty\}$ are naturally denoted by $\overline{\mathbb{N}}_n$ and $\overline{\mathbb{N}}_n^*$.

We recall here some notation about preorders and equivalence relations. A *preorder* is a relation which is reflexive and transitive but not necessarily antisymmetric. Let $\prec$ be a preorder on a set E. If the relation $x \prec y$ holds, x is said to be *smaller* than y. It is said to be *strictly smaller* than y if the relation $x \prec y$ holds but the relation $y \prec x$ does not hold. An element x of E is said to be *minimal* (respectively *maximal*) in a subset X of E iff x belongs to X and no element of X is strictly smaller than x (respectively no element is strictly bigger than x). With a preorder $\prec$ is associated a canonical equivalence relation $\sim$ defined by $x \sim y \iff x \prec y$ and $y \prec x$ for any x and y in E. The preorder induces an order on the equivalence classes of $\sim$. This order is also denoted by $\prec$. A preorder $\prec$ on a set E is said to *compatible* with a function f defined on E if for any $x, y \in E$,

$$f(x) = f(y) \iff x \prec y \text{ and } y \prec x$$

If the preorder $\prec$ is compatible with f, the function f naturally defines a function which maps any equivalence class to the value by f of any element of this class. This function is also denoted by f.

We point out that an equivalence relation is a preorder which is equal to its associated equivalence relation. Everything that has been said for preorders also applies to equivalence relations.

2.2 Automata

We recall here some elements of the theory of rational sets of finite and infinite words. For further details on automata and rational sets of finite words, see [Per90] and for background on automata and rational sets of infinite words, see [Tho90]. Let A be a set called an *alphabet* and usually assumed to be finite. We respectively denote by A^* and A^+ the set of finite words and the set of nonempty finite words. The set of right-infinite words, also called ω-words, is denoted by A^ω.

A *Büchi automaton* $\mathcal{A} = (Q, A, E, I, F)$ is a non-deterministic automaton with a set Q of states, subsets $I, F \subset Q$ of *initial* and *final* states and a set $E \subset Q \times A \times Q$ of *transitions*. A transition (p, a, q) of $\mathcal{A}$ is denoted $p \xrightarrow{a} q$. A *path* in $\mathcal{A}$ is an infinite sequence

$$\gamma : q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots$$

of consecutive transitions. The *starting* state of the path is q_0 and the infinite word $\lambda(\gamma) = a_0 a_1 \ldots$ is called the *label* of γ. A path is said to be *initial* if its starting state is initial. It is said to be *final* if it visits infinitely often a final state. It is said to be *successful* if it is initial and final.

As usual, an infinite word is accepted by the automaton if it is the label of a successful path. The set of accepted infinite words is said to be recognized by the automaton and is denoted by $\mathrm{L}(\mathcal{A})$. It is well known that a set of infinite words is rational iff it is recognized by some automaton.

A state of a Büchi automaton $\mathcal{A}$ is said to be *coaccessible* if it is the starting state of a final path. A Büchi automaton is said to be *trim* if all states are coaccessible. Any state which occurs in a final path is coaccessible and thus non-coaccessible states of an automaton can be removed. In the sequel, automata are usually assumed to be trim.

An automaton $\mathcal{A} = (Q, A, E, I, F)$ is said to be *codeterministic* if for any pair $(a, q) \in A \times Q$, there is at most one state $p \in Q$ such that $p \xrightarrow{a} q$ is a transition of $\mathcal{A}$. This unique state p is denoted by $a \cdot q$. This defines a partial left action of A^* on Q. Since a letter a and a state q uniquely determine the transition $a \cdot q \xrightarrow{a} q$, we denote this transition by $a \triangleright q$. Likewise, for a word $w = a_1 \ldots a_n$ and a state q, we denote by $w \triangleright q$ the finite path

$$w \cdot q \xrightarrow{a_1} \cdots \xrightarrow{a_{n-2}} a_{n-1} a_n \cdot q \xrightarrow{a_{n-1}} a_n \cdot q \xrightarrow{a_n} q$$

which is the unique finite path labeled by w which ends in q.

3 Unambiguous automata

In this section, we introduce the concept of unambiguous Büchi automata. We first give the definition and we state one basic property of these automata. We then establish a characterization of these automata. We give some examples and we state the main result.

Definition 1 *A Büchi automaton $\mathcal{A}$ is said to be* unambiguous *(respectively* complete*) iff any infinite word labels at most (respectively at least) one final path in $\mathcal{A}$. When $\mathcal{A}$ is unambiguous, this unique path is denoted by $\gamma(x)$.*

The set of final paths is only determined by the transitions and the final states of $\mathcal{A}$. Thus, the property of being unambiguous or complete does not depend on the set of initial states of $\mathcal{A}$. In the sequel, we will freely say that an automaton $\mathcal{A}$ is unambiguous or complete without specifying its set of initial states.

The definition of the word "complete" we give in this paper is not exactly the usual one. First of all, we should have used the word "cocomplete" since we deal with codeterministic automata and not deterministic ones but we prefer "complete" for simplicity. Next, an automaton is usually said to be complete (i.e., cocomplete) if no transition is missing. Here, we say that an automaton is complete if any word is the label of a final path. For finite words, both notions are equivalent. Any finite word labels a path ending in a final state in a codeterministic automaton iff for any pair $(a, q) \in A \times Q$, there is at least one state $p \in Q$ such that $p \xrightarrow{a} q$ is a transition. However both notions do not coincide anymore for infinite words as it is shown in Example 6. Our definition is actually stronger than the usual one as it will be seen in Proposition 3. In any case, our definition is what is really needed.

In the sequel, we write UBA for Unambiguous Büchi Automaton and CUBA for Complete Unambiguous Büchi Automaton. The following example is the simplest CUBA.

Example 2 The automaton $(\{q\}, A, E, I, \{q\})$ with $E = \{q \xrightarrow{a} q \mid a \in A\}$ is obviously a CUBA. It recognizes the set A^ω of all infinite words if q is initial and recognizes the empty set otherwise. It is called the *trivial* CUBA.

The following proposition states that an UBA must be codeterministic. Such an automaton can be seen as a deterministic automaton which reads infinite words from right to left. It starts at infinity and ends at the beginning of the word. Codeterministic automata on infinite words have already been considered in [Mos82] and [BP85] where is proved in that paper that any rational set of infinite words is recognized by a codeterministic automata. Our main theorem generalizes this results. It states that any rational set of infinite words is recognized by a CUBA.

Proposition 3 *Let $\mathcal{A} = (Q, A, E, I, F)$ be a trim Büchi automaton. If $\mathcal{A}$ is unambiguous, then $\mathcal{A}$ is codeterministic. If $\mathcal{A}$ is complete, then for any pair*

$(a, q) \in A \times Q$, *there is at least one state* $p \in Q$ *such that* $p \xrightarrow{a} q$ *is a transition of* $\mathcal{A}$.

Proof We first prove that the automaton $\mathcal{A}$ is codeterministic. Suppose that there are two transitions $p \xrightarrow{a} q$ and $p' \xrightarrow{a} q$. Since the automaton is assumed to be trim, the state q is the starting state of a final path γ labeled by x. Since both paths $p\gamma$ and $p'\gamma$ are two final paths labeled by ax, we have $p = p'$ and the automaton is codeterministic.

We now prove that the automaton $\mathcal{A}$ is complete. Let q be a state of $\mathcal{A}$ and a be a letter. Since the automaton $\mathcal{A}$ is assumed to be trim, there is a final path γ starting at q labeled by some infinite word x. If the automaton is complete, the infinite word ax labels a final path $\gamma' = q_0 q_1 q_2 \ldots$ in $\mathcal{A}$. Since both paths γ and $\gamma'' = q_1 q_2 q_3 \ldots$ are final paths labeled by x, the states q and q_1 are equal. There is then a transition $q_0 \xrightarrow{a} q$ and the automaton $\mathcal{A}$ is complete. $\square$

If the automaton $\mathcal{A}$ is unambiguous, the transitions define a left action of A^* on Q. Furthermore, if $\mathcal{A}$ is also complete, this action is completely defined. The converse does not hold. It is not true that if for any pair $(a, q) \in A \times Q$ there is exactly one transition $p \xrightarrow{a} q$, then the automaton is unambiguous and complete. This will be shown in Example 6. Proposition 8 provides some additional condition on the automaton to ensure that it is unambiguous and complete.

Before giving some other examples of CUBA, we provide a simple characterization of CUBA which makes it easy to verify that an automaton is unambiguous and complete. This proposition also shows that it can be effectively checked whether a given automaton is unambiguous or complete.

Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi automaton and let q be a state of $\mathcal{A}$. We denote by $\mathcal{A}_q = (Q, A, E, \{q\}, F)$ the new automaton obtained by taking the singleton $\{q\}$ as set of initial states. The set $\mathrm{L}(\mathcal{A}_q)$ is then the set of infinite words labeling a final path starting at state q.

Proposition 4 *Let* $\mathcal{A} = (Q, A, E, I, F)$ *be a Büchi automaton. The automaton* $\mathcal{A}$ *is unambiguous iff the sets* $\mathrm{L}(\mathcal{A}_q)$ *are pairwise disjoint. The automaton* $\mathcal{A}$ *is complete iff* $A^\omega = \bigcup_{q \in Q} \mathrm{L}(\mathcal{A}_q)$

If the automaton $\mathcal{A}$ is trim, each set $\mathrm{L}(\mathcal{A}_q)$ is nonempty. Therefore, the automaton $\mathcal{A}$ is unambiguous and complete iff the family of sets $\mathrm{L}(\mathcal{A}_q)$ for $q \in Q$ is a partition of A^ω.

The previous characterization can be used as follows to check whether a given automaton is unambiguous or complete. Given two distinct states q and q' of $\mathcal{A}$, it can be effectively checked whether the two sets recognized by the automata $\mathcal{A}_q$ and $\mathcal{A}_{q'}$ are disjoint. Furthermore, this test can be performed in polynomial time. The set $\bigcup_{q \in Q} \mathrm{L}(\mathcal{A}_q)$ is recognized by the automaton $\mathcal{A}_Q = (Q, A, E, Q, F)$ all of which states are initial. The inclusion $A^\omega \subset \bigcup_{q \in Q} \mathrm{L}(\mathcal{A}_q)$ holds iff this automaton recognizes A^ω. This can be checked but it does not seem it can be performed in polynomial time.

Proof We first prove that the automaton $\mathcal{A}$ is unambiguous iff the sets $\mathrm{L}(\mathcal{A}_q)$ are pairwise disjoint. It is clear that if $\mathcal{A}$ is unambiguous, then the sets $\mathrm{L}(\mathcal{A}_q)$ are pairwise disjoint. Conversely, suppose that the infinite word x labels two different final paths $\gamma = q_0q_1q_2\dots$ and $\gamma' = q'_0q'_1q'_2\dots$ of $\mathcal{A}$. Let n be the least integer such that $q_n \neq q'_n$. The infinite word $x' = a_na_{n+1}a_{n+2}\dots$ belongs to both $\mathrm{L}(\mathcal{A}_{q_n})$ and $\mathrm{L}(\mathcal{A}_{q'_n})$ and this is a contradiction.

We now prove that the automaton $\mathcal{A}$ is complete iff the inclusion $A^\omega \subset \bigcup_{q\in Q}\mathrm{L}(\mathcal{A}_q)$ holds. If $\mathcal{A}$ is complete, any infinite word labels at least one final path and the inclusion obviously holds. Conversely, if the inclusion holds, any infinite word belongs to $\mathrm{L}(\mathcal{A}_q)$ for some q and labels a final path starting at q. $\square$

We now come to examples. We use Proposition 4 to verify that the following two automata are unambiguous and complete. In the figures, a transition $p \xrightarrow{a} q$ of an automaton is represented by an arrow labeled by a from p to q. Initial states have a small incoming arrow whereas final states are marked by a double circle. A complete but ambiguous Büchi automaton will be given in Example 9.

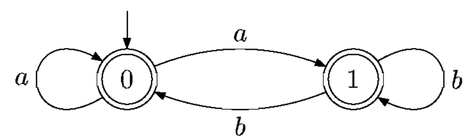

Figure 1: CUBA of Example 5

Example 5 Let A be the alphabet $A = \{a, b\}$ and let $\mathcal{A}$ be the automaton pictured in Figure 1. This automaton is unambiguous and complete since we have $\mathrm{L}(\mathcal{A}_0) = aA^\omega$ and $\mathrm{L}(\mathcal{A}_1) = bA^\omega$. It recognizes the set aA^ω of infinite words beginning with an a.

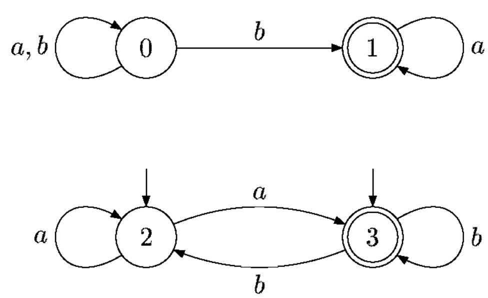

Figure 2: CUBA of Example 6

The following example shows that a CUBA may have several connected components.

Example 6 Let A be the alphabet $A = \{a, b\}$ and let $\mathcal{A}$ be the automaton pictured in Figure 2. It is unambiguous and complete since we have $\mathrm{L}(\mathcal{A}_0) =$

A^*ba^ω, $\mathrm{L}(\mathcal{A}_1) = a^\omega$, $\mathrm{L}(\mathcal{A}_2) = a(A^*b)^\omega$ and $\mathrm{L}(\mathcal{A}_3) = b(A^*b)^\omega$. It recognizes the set $(A^*b)^\omega$ of words having an infinite number of b.

The automaton of the previous example has two connected components. Since it is unambiguous and complete any infinite word labels exactly one final path in this automaton. This final path lies in the first component if the infinite word has finitely many b and it lies in the second component otherwise. This automaton shows that our definition of completeness for an unambiguous Büchi automaton is slightly different from the usual one. Any connected component is complete in the usual sense if it is considered as a whole automaton. For any letter a and any state q of this component there exists exactly one state $p \in Q$ such that $p \xrightarrow{a} q$ is a transition. However, each component is not complete according to our definition since not any infinite word labels a final path in this component.

In the realm of finite words, an automaton is usually made unambiguous by the usual subsets construction [HU79, p. 22]. This construction associates with an automaton $\mathcal{A}$ an equivalent deterministic automaton whose states are subsets of states of $\mathcal{A}$. Since left and right are symmetric for finite words, this construction can be reversed to get a codeterministic automaton which is also equivalent to $\mathcal{A}$. In the case of infinite words, the result of McNaughton [McN66] states that any Büchi automaton is equivalent to a Muller automaton which is deterministic. However, this construction cannot be reversed since infinite words are right-infinite. We have seen in Proposition 3 that a CUBA is codeterministic. The following theorem is the main result of the paper. It states that any rational set of infinite words is recognized by a CUBA. This theorem is thus the counterpart of McNaughton's result for codeterministic automata. Like Muller automata, CUBAs make the complementation very easy to do. This will be shown in Section 5. Thus Theorem 7 leads to a new proof that the class of rational sets of infinite words is closed under complementation.

Theorem 7 *Any rational set of infinite words is recognized by a complete unambiguous Büchi automaton.*

We give two proofs of Theorem 7. However, both proofs are rather long and involve elaborate arguments. They are postponed to Section 6.

4 Properties and characterizations

In this section, we present some additional properties of CUBA. We first give another characterization of CUBA which involves loops going through final states. We present some consequences of this characterization. The characterization of CUBA given in Proposition 4 uses sets of infinite words. The family of sets of infinite words labeling a final path starting in the different states must be a partition of the set A^ω of all infinite words. The following proposition only uses sets of finite words to characterize UBAs and CUBAs.

Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi automaton such that for any pair (a, q) there exists exactly one transition $p \xrightarrow{a} q$. Let S_q be the set of nonempty finite words w such that $w \cdot q = q$ and such that the path $w \triangleright q$ contains a final state. We have then the following proposition.

Proposition 8 *The automaton $\mathcal{A}$ is unambiguous iff the sets S_q are pairwise disjoint. The automaton $\mathcal{A}$ is unambiguous and complete iff the family of sets S_q for $q \in Q$ is a partition of A^+ (up to the fact that some of them might be empty). In this case, the final path labeled by the periodic infinite word w^ω is the path*

$$q \xrightarrow{w} q \xrightarrow{w} q \cdots$$

where q is the unique state such that $w \in S_q$.

The second statement of the proposition says that if the automaton $\mathcal{A}$ is supposed to be unambiguous, it is complete iff the inclusion $A^+ \subset \bigcup_{q \in Q} S_q$ holds. The assumption that the automaton is unambiguous is necessary. As the following example shows, it is not true in general that the automaton is complete iff the inclusion holds.

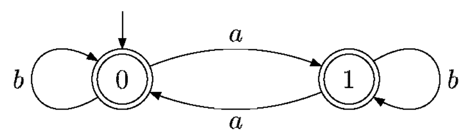

Figure 3: Automaton of Example 9

Example 9 The automaton pictured in Figure 3 is ambiguous since the infinite word b^ω labels two final paths. Since this automaton is deterministic and all states are final, it is complete. However, it is not true that $A^+ \subset \bigcup_{q \in Q} S_q$. Indeed, no loop in this automaton is labeled by the finite word a.

We now come to the proof of Proposition 8.

Proof We first prove that $\mathcal{A}$ is unambiguous iff the sets S_q are pairwise disjoint. If the finite word w belongs to S_q for some state q, the path

$$q \xrightarrow{w} q \xrightarrow{w} q \cdots$$

is a final path labeled by w^ω. This proves that if $\mathcal{A}$ is unambiguous, the sets S_q must be pairwise disjoint. We now suppose that these sets are pairwise disjoint. It is clear that if w belongs to S_q, then w^n belongs to S_q for any integer n. We claim that if the sets S_q are pairwise disjoint, the converse also holds. If w^n belongs to S_q for some integer n, then w belongs to S_q. Indeed, if w^n belongs to S_q, there exists a finite path

$$q \xrightarrow{w} q_1 \xrightarrow{w} \cdots \xrightarrow{w} q_{n-1} \xrightarrow{w} q$$

which contains a final state. Furthermore w^n also belongs to S_{q_k} for $1 \leq k \leq n-1$. Thus, q_k is equal to q for $1 \leq k \leq n-1$ and w belongs to S_q. If the two rational sets $\mathrm{L}(\mathcal{A}_q)$ and $\mathrm{L}(\mathcal{A}_{q'})$ contain a common infinite word x, it may be assumed that x is ultimately periodic and that x is equal to uw^ω for two finite words u and w. Let γ and γ'

$$\gamma : q \xrightarrow{u} q_1 \xrightarrow{w} q_2 \xrightarrow{w} q_3 \cdots$$
$$\gamma' : q' \xrightarrow{u} q'_1 \xrightarrow{w} q'_2 \xrightarrow{w} q'_3 \cdots$$

the two final path labeled by x starting in q and q'. Since the number of state is finite, there exists a state p which appears infinitely often in γ. In particular, there are two integers $i < j$ such that $p = q_i = q_j$ and such that the path $w^{j-i} \triangleright p$ contains a final state. Thus, w^{j-i} belongs to S_p and w belongs also to S_p by the previous remark. Finally, we obtain $q_n = q$ for any positive integer and $q = w \cdot p$. By symmetry, there exists a state p' such that $q'_n = q$ for any positive integer and $q' = w \cdot p'$. Since w belongs to S_p and to $S_{p'}$, we have $p = p'$ and $q = q'$. The automaton $\mathcal{A}$ is thus unambiguous by Proposition 4.

We suppose that $\mathcal{A}$ is unambiguous and we prove that $\mathcal{A}$ is complete iff the inclusion $A^+ \subset \bigcup_{q \in Q} S_q$ holds. If the inclusion holds, any ultimately periodic word uw^ω has a final path which is the path

$$p \xrightarrow{u} q \xrightarrow{w} q \xrightarrow{w} q \cdots$$

where q is a state such that $w \in S_q$ and $p = w \cdot q$. The rational set $\bigcup_{q \in Q} \mathrm{L}(\mathcal{A}_q)$ contains all the ultimately periodic infinite words and is then equal to A^ω. By Proposition 4, the automaton $\mathcal{A}$ is complete. We now suppose that the automaton is complete and we prove the inclusion $A^+ \subset \bigcup_{q \in Q} S_q$. Since the automaton is complete, the infinite word w^ω labels a final path

$$\gamma : q_0 \xrightarrow{w} q_1 \xrightarrow{w} q_2 \cdots$$

Since the number of state is finite, there are two integers $i < j$ such that $q_i = q_j = q$ and such that the path $w^{j-i} \triangleright q$ contains a final state. Thus, w^{j-i} belongs to S_q and w belongs also to S_q. Thus, any finite word w belong to at least one of the S_q. $\Box$

Proposition 8 gives another method to check whether a given Büchi automaton is unambiguous and complete. It must be first verified that for any pair (a, q) there exists exactly one transition $p \xrightarrow{a} q$. Then, it must be checked whether the family of sets S_q for $q \in Q$ forms a partition of A^+. The sets S_q are rational and a codeterministic automaton recognizing S_q can be easily deduced from the automaton $\mathcal{A}$. It is then straightforward to verify that the sets S_q form a partition of A^+.

The last statement of Proposition 8 says that the final path labeled by a periodic word is also periodic. It is worth mentioning that the same result does not hold for deterministic automata. Indeed, we have the following proposition.

Proposition 10 *The only deterministic CUBA is the trivial CUBA (see Example 2) with only one state q and the transitions defined by $q \xrightarrow{a} q$ for any letter a of A.*

Proof If a CUBA $\mathcal{A}$ is deterministic, it is then a permutation automaton in the sense that for any letter a, the transitions labeled by a induce a permutation of the states. It means that for each letter a of A, the left action of a is a permutation of the states. The semigroup generated by the actions of the letter is a finite group, and there is an integer n such that $w^n \cdot q = q$ for any word w and any state q. In particular, we have $a^n \cdot q = q$ for any letter a. We claim that the automaton $\mathcal{A}$ has one final state. Suppose that both states q and q' are final. If q is different from q', there are two different final paths labeled by a^ω. There is then a unique state q_0 which is final and for any letter a, we have $a \cdot q_0 = q_0$. All other states are then non-coaccessible and the automaton $\mathcal{A}$ is the trivial one. □

5 Boolean combinations

There are standard methods to construct Büchi automata recognizing the union $X \cup Y$ and the intersection $X \cap Y$ from two Büchi automata recognizing X and Y. If they are applied to complete and unambiguous automata, the resulting automata are neither complete nor unambiguous. However, they can be slightly modified in order to preserve the unambiguity and the completeness of the automata. In this section, we show how to construct complete and unambiguous automata recognizing the union $X \cup Y$ and the intersection $X \cap Y$ from two complete and unambiguous Büchi automata recognizing X and Y. We also show that if a complete and unambiguous Büchi automaton recognizes a set X, it suffices to change the initial states in order to recognize the complement of X.

5.1 Complement

We begin with complementation which turns out to be a very easy operation for CUBAs. Indeed, it suffices to change the initial states of the automaton to recognize the complement. For a subset J of a set Q, we denote by $Q \setminus J$, the complement of J in Q.

Proposition 11 *Let $\mathcal{A} = (Q, A, E, I, F)$ be a CUBA recognizing a set X of infinite words. The automaton $\mathcal{A}' = (Q, A, E, Q \setminus I, F)$ is unambiguous and complete. It recognizes the complement of X.*

It must be pointed out that it is really necessary for the automaton $\mathcal{A}$ to be unambiguous and complete. Indeed, if $\mathcal{A}$ is ambiguous, it may happen that an infinite word x of X labels a final path starting at an initial state and another final path starting at a non initial state. In this case, the infinite word x is also recognized by the automaton $\mathcal{A}'$. If $\mathcal{A}$ is not complete, some infinite word x

does not label any final path. This infinite word which does not belong to X is not recognized by the automaton $\mathcal{A}'$.

Proof Since the property of being unambiguous or complete does not depend on the initial states of an automaton, the automaton $\mathcal{A}'$ is obviously unambiguous and complete. Furthermore, for any infinite word x, the respective final paths $\gamma(x)$ and $\gamma'(x)$ in $\mathcal{A}$ and $\mathcal{A}'$ are equal. Since a state of Q is initial in $\mathcal{A}$ iff it is not initial in $\mathcal{A}'$, the path $\gamma(x)$ is successful in $\mathcal{A}$ iff the path $\gamma'(x)$ is not successful in $\mathcal{A}'$. The automaton $\mathcal{A}'$ recognizes the complement of X. $\square$

By the previous result, the proof of Theorem 7 also provides a new proof of the fact that the family of rational sets of infinite words is closed under complementation.

5.2 Union and intersection

In this section, we show how CUBAs recognizing the union $X_1 \cup X_2$ and the intersection $X_1 \cap X_2$ can be obtained from CUBAs recognizing X_1 and X_2.

We suppose that the sets X_1 and X_2 are respectively recognized by the CUBA $\mathcal{A}_1 = (Q_1, A, E_1, I_1, F_1)$ and $\mathcal{A}_2 = (Q_2, A, E_2, I_2, F_2)$. We will construct two CUBA $\mathcal{U} = (Q, A, E, I_{\mathcal{U}}, F)$ and $\mathcal{I} = (Q, A, E, I_{\mathcal{I}}, F)$ respectively recognizing the union $X_1 \cup X_2$ and the intersection $X_1 \cap X_2$. Both automata $\mathcal{U}$ and $\mathcal{I}$ share the same states set Q, the same transitions set E and the same set F of final states.

We first describe the states and the transitions of both automata $\mathcal{U}$ and $\mathcal{I}$. These automata are based on the product of the automata $\mathcal{A}_1$ and $\mathcal{A}_2$ but a third component is added. The final states may not appear at the same time in $\mathcal{A}_1$ and $\mathcal{A}_2$. The third component synchronizes the two automata by indicating in which of the two automata comes the first final state. The set Q of states is $Q = Q_1 \times Q_2 \times \{1, 2\}$. Each state is then a triple (q_1, q_2, ε) where q_1 is a state of $\mathcal{A}_1$, q_2 is a state of $\mathcal{A}_2$ and ε is 1 or 2. We define the set E of transitions by giving the left action of a letter a on a state (q_1, q_2, ε) of Q. We define $a \cdot (q_1, q_2, \varepsilon) = (a \cdot q_1, a \cdot q_2, \varepsilon')$ where $a \cdot q_1$ and $a \cdot q_2$ are the respective left actions defined by E_1 and E_2 and where ε' is defined as follows.

$$\varepsilon' = \begin{cases} 1 & \text{if } q_1 \in F_1 \\ 2 & \text{if } q_1 \notin F_1 \text{ and } q_2 \in F_2 \\ \varepsilon & \text{otherwise} \end{cases}$$

This definition is not completely symmetric. When both q_1 and q_2 are final states, we choose to set $\varepsilon' = 1$. We now define the set F of final states as

$$F = \{(q_1, q_2, \varepsilon) \big| q_2 \in F_2 \text{ and } \varepsilon = 1\}.$$

This definition is also non symmetric. The following lemma states the key property of final states.

Lemma 12 *Let $q = (q_1, q_2, 1)$ be a final state and let w be a finite word such that $w \cdot q = q$. The path $w \triangleright q$, which is a loop around q, contains a state $q' = (q'_1, q'_2, \varepsilon')$ such that $q'_1 \in F_1$.*

Proof Assume that $w \triangleright q$ is the path

$$q^0 \xrightarrow{a_1} q^1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q^n$$

where $q^0 = q^n = q$ and where each state $q^i \in Q$ is the triple $(q^i_1, q^i_2, \varepsilon_i)$ for $0 \leq i \leq n$. Let k the greatest integer strictly smaller than n such that $\varepsilon_k = 1$. Such an integer k exists since $\varepsilon_0 = 1$. We claim that $q' = q^{k+1}$ is such that $q'_2 \in F_2$.

- Suppose first that $k = n - 1$. Since q is a final state, the state q_2 satisfies $q_2 \in F_2$. This ensures that $q_1 \in F_1$. Otherwise, the integer ε_k would have been equal to 2.
- Suppose now that $k < n - 1$. By definition of k, we have $\varepsilon_{k+1} = 2$. Since ε_k is different from ε_{k+1} and $\varepsilon_k = 1$, we have $q^{k+1}_1 \in F_1$.

□

Each loop containing a final state contains then two states q and q', possibly equal, such that $q_1 \in F_1$ and $q'_2 \in F_2$. We claim that, for any set I of initial states, the automaton $\mathcal{A} = (Q, A, E, I, F)$ is a CUBA. By definition of the transitions, the automaton $\mathcal{A}$ is codeterministic. In order to prove that this automaton is unambiguous and complete, we define a function π which maps a infinite path γ in $\mathcal{A}$ labeled by an infinite word x into a pair (γ_1, γ_2) of paths in $\mathcal{A}_1$ and $\mathcal{A}_2$ labeled by x. We then prove that the function π is one to one from the set of final paths in $\mathcal{A}$ to the sets of pairs of final paths in $\mathcal{A}_1$ and $\mathcal{A}_2$. Let γ be the path

$$\gamma : q^0 \xrightarrow{a_0} q^1 \xrightarrow{a_1} q^2 \cdots$$

where each state $q^i \in Q$ is the triple $(q^i_1, q^i_2, \varepsilon_i)$ for $i \geq 0$. The define $\pi(\gamma) = (\gamma_1, \gamma_2)$ where the paths γ_1 and γ_2 are defined by

$$\begin{aligned} \gamma_1 :& q^0_1 \xrightarrow{a_0} q^1_1 \xrightarrow{a_1} q^2_1 \cdots \\ \gamma_2 :& q^0_2 \xrightarrow{a_0} q^1_2 \xrightarrow{a_1} q^2_2 \cdots \end{aligned}$$

The following lemma states the main property of the function π.

Lemma 13 *The function π is one to one from the set of final paths in $\mathcal{A}$ to the sets of pairs of final paths in $\mathcal{A}_1$ and $\mathcal{A}_2$.*

Proof We first prove that if γ is a final path, then both γ_1 and γ_2 are respectively final paths in $\mathcal{A}_1$ and $\mathcal{A}_2$. Assume that γ goes infinitely often through a final state $q = (q_1, q_2, 1)$. The fact $q_2 \in F_2$ ensures that γ_2 is a final path. The path γ can be factorized

$$\gamma : q^0 \xrightarrow{w_0} q \xrightarrow{w_1} q \cdots$$

Since for any integer $i > 0$, we have $w_i \cdot q = q$, the path $w_i \triangleright q$ contains, by Lemma 12, a state q' such that $q'_1 \in F_1$. The path γ_1 is also final.

We now prove that for any pair (γ_1, γ_2), there exists a unique final path such that γ such that $\pi(\gamma) = (\gamma_1, \gamma_2)$. Assume that γ_1 and γ_2 are the paths

$$\begin{aligned} \gamma_1 :& q_1^0 \xrightarrow{a_0} q_1^1 \xrightarrow{a_1} q_1^2 \cdots \\ \gamma_2 :& q_2^0 \xrightarrow{a_0} q_2^1 \xrightarrow{a_1} q_2^2 \cdots \end{aligned}$$

By definition of π, a path γ such that $\pi(\gamma) = (\gamma_1, \gamma_2)$ must be equal to a path of the form

$$\gamma : q^0 \xrightarrow{a_0} q^1 \xrightarrow{a_1} q^2 \cdots$$

where each state $q^i \in Q$ is the triple $(q_1^i, q_2^i, \varepsilon_i)$ for $i \geq 0$. For any integer i, let j the least integer strictly greater than i such that $q_1^j \in F_1$ or $q_2^j \in F_2$. This integer always exists since γ_1 and γ_2 are final paths. By definition of the left action, the integer ε_i is equal to 1 if $q_1^j \in F_1$ and is equal to 2 otherwise. This proves that there exists at most one path γ such that $\pi(\gamma) = (\gamma_1, \gamma_2)$.

We now claim that the path γ defined as above is final. Since the path γ_2 is final, there is an infinite sequence $k_1, k_2, \ldots$ of integers such that $q_2^{k_i} \in F_2$ for $i > 0$. Since the path γ_2 is also final, there is an infinite number of integers k_i such that there is an integer $k_i < l_i \leq k_{i+1}$ such that $q_1^{l_i} \in F_1$. For all these integers k_i, we have $\varepsilon_{k_i} = 1$. The path γ is then final. □

We finally prove that the automaton $\mathcal{A} = (Q, A, E, I, F)$ is unambiguous and complete. Let x be an infinite word and let γ_1 and γ_2 be the final paths labeled by x in $\mathcal{A}_1$ and $\mathcal{A}_2$. By Lemma 13, there is a final path γ labeled by x in $\mathcal{A}$. Suppose now that there are two paths γ and γ' labeled by x in $\mathcal{A}$. By Lemma 13, both pairs $\pi(\gamma)$ and $\pi(\gamma')$ are formed of final paths in $\mathcal{A}_1$ and $\mathcal{A}_2$. Since both automata $\mathcal{A}_1$ and $\mathcal{A}_2$ are unambiguous, we get $\pi(\gamma) = \pi(\gamma')$. By lemma 13, the function π is one to one and $\gamma = \gamma'$. This proves that the automaton $\mathcal{A}$ is unambiguous.

Let x be an infinite word and let γ, γ_1 and γ_2 the respective final paths labeled by x in $\mathcal{A}$, $\mathcal{A}_1$ and $\mathcal{A}_2$. We have proved that $\pi(\gamma) = (\gamma_1, \gamma_2)$. If q_1 and q_2 are the respective starting states of γ_1 and γ_2, the starting state of γ is then equal to (q_1, q_2, ε) with $\varepsilon \in \{1, 2\}$. We thus define the sets $I_\mathcal{U}$ and $I_\mathcal{I}$ of initial states of the automata $\mathcal{U}$ and $\mathcal{I}$ as follows.

$$\begin{aligned} I_\mathcal{U} &= \big\{(q_1, q_2, \varepsilon) \big| (q_1 \in I_1 \text{ or } q_2 \in I_2) \text{ and } \varepsilon \in \{1, 2\}\big\} \\ I_\mathcal{I} &= \big\{(q_1, q_2, \varepsilon) \big| q_1 \in I_1 \text{ and } q_2 \in I_2 \text{ and } \varepsilon \in \{1, 2\}\big\} \end{aligned}$$

From these definitions, it is clear that both automata $\mathcal{U}$ and $\mathcal{I}$ are unambiguous and complete and that they respectively recognize $X_1 \cup X_2$ and $X_1 \cap X_2$.

6 Proofs

In this section, we give two proofs of the main theorem which states that any rational set of infinite words is recognized by a CUBA. Although the two proofs share a common part, they are of different nature. The first proof uses graphs whereas the second one uses semigroups. Both proofs are constructive. They can be used to compute a CUBA recognizing a given rational set X. However, both proofs do not need the same input. The first one needs a Büchi automaton recognizing the set X whereas the second one needs a morphism from A^+ into a finite semigroups which recognizes the set X.

The proofs are organized as follows. A first part is devoted to a reduction of the problem. We introduce a generalization of Büchi automata called *generalized Büchi automata*. These automata are more convenient for constructing unambiguous automata. However, we show that these automata are equivalent to usual Büchi automata. From an unambiguous and complete generalized Büchi automaton can be deduced an unambiguous and complete usual Büchi automaton. In a second part, we use graphs to construct a complete and unambiguous generalized Büchi automaton recognizing a set of infinite words. In a third part, we use semigroups to construct a complete and unambiguous generalized Büchi automaton recognizing a set of infinite words. Thus, the second and third parts are independent from one another but both use the first part.

6.1 Generalized Büchi automata

In a Büchi automaton, the set of final paths is the set of paths which go infinitely often through final states. In a generalized Büchi automaton, the set of final paths is given by a family of subsets of transitions.

Definition 14 *A* generalized Büchi automaton $\mathcal{A} = (Q, A, E, I, \mathcal{F})$ *is an automaton where* $\mathcal{F} = \{F_1, \ldots, F_m\}$ *is a family of subsets of transitions. A path* γ *is said to be* final *if for any* F_k *in* $\mathcal{F}$*, there is a transition of* F_k *which appears infinitely often in* γ*. The automaton* $\mathcal{A}$ *is said to be* unambiguous *(respectively* complete*) iff any infinite word labels at most (respectively at least) one final path in* $\mathcal{A}$*.*

In the sequel, we write GBA for Generalized Büchi Automaton. We respectively write UGBA and CUGBA instead of Unambiguous GBA and Complete Unambiguous GBA.

We point out that usual Büchi automata are a particular case of GBA. Indeed, any Büchi automaton $\mathcal{A} = (Q, A, E, I, F)$ is equivalent to the generalized Büchi automaton $(Q, A, E, I, \mathcal{F})$ where the family $\mathcal{F}$ only contains one set F_1 of transitions. This set F_1 is equal to the set of transitions $p \xrightarrow{a} q$ such that p is final.

It is well known that any generalized Büchi automaton can be effectively transformed into an equivalent Büchi automaton. However, the standard constructions do not preserve the unambiguity and the completeness of the automaton. Therefore, we present another construction.

Proposition 15 *Let $\mathcal{A} = (Q, A, E, I, \mathcal{F})$ be a generalized Büchi automaton such that $|Q| = n$ and $|\mathcal{F}| = m$. There is an equivalent Büchi automaton $\mathcal{A}' = (Q', A, E', I', F')$ such that $|Q'| \le 2^m n$. Furthermore, if $\mathcal{A}$ is unambiguous (respectively complete), then $\mathcal{A}'$ is also unambiguous (respectively complete).*

The standard methods to transform a generalized Büchi automaton into a Büchi automaton usually yield an automaton with mn states instead of $2^m n$. However, this exponential blow up cannot be avoided if the unambiguity and the completeness are preserved as it is shown in Proposition 18.

The proof of the previous proposition is based on the following two lemmas. The first lemma states that any generalized Büchi automaton is equivalent to another one which has one less subset of transitions in the acceptance condition. The second lemma states that for any generalized Büchi automaton with only one set of transitions in its acceptance condition, there is an equivalent Büchi automaton. Both constructions preserve the unambiguity and the completeness of the automaton.

Lemma 16 *Let $\mathcal{A} = (Q, A, E, I, \mathcal{F})$ be a generalized Büchi automaton such that $|Q| = n$ and $|\mathcal{F}| = m \ge 2$. There is an equivalent generalized Büchi automaton $\mathcal{A}' = (Q', A, E', I', \mathcal{F}')$ such that $|Q'| \le 2n$ and $|\mathcal{F}| = m - 1$. Furthermore, if $\mathcal{A}$ is unambiguous (respectively complete), then $\mathcal{A}'$ is also unambiguous (respectively complete).*

Proof We describe here the construction. Let $\mathcal{F}$ be the family $\{F_1, \ldots, F_m\}$. The general idea is to add some memory to the states which allows us to replace the two sets F_1 and F_2 of transitions by a single set F'_2. The generalized Büchi automaton $\mathcal{A}' = (Q', A, E', I', \mathcal{F}')$ where $\mathcal{F}' = \{F'_2, \ldots, F'_m\}$ is defined as follows. The state set of $\mathcal{A}'$ is $Q \times \{1, 2\}$ and the set I' of initial states is $I \times \{1, 2\}$. There is a transition $(q, \varepsilon) \xrightarrow{a} (q', \varepsilon')$ if there is a transition $q \xrightarrow{a} q'$ in $\mathcal{A}$ and if ε and ε' satisfy

$$\varepsilon = \begin{cases} 1 & \text{if } q \xrightarrow{a} q' \in F_1 \\ 2 & \text{if } q \xrightarrow{a} q' \in F_2 \setminus F_1 \\ \varepsilon' & \text{otherwise} \end{cases}$$

The set F'_k of transitions are then defined by

$$\begin{aligned} F'_2 &= \{(q, \varepsilon) \xrightarrow{a} (q', \varepsilon') \mid q \xrightarrow{a} q' \in F_2 \text{ and } \varepsilon' = 1\} \\ F'_k &= \{(q, \varepsilon) \xrightarrow{a} (q', \varepsilon') \mid q \xrightarrow{a} q' \in F_k\} \quad \text{for } k \ge 3. \end{aligned}$$

□

Lemma 17 *Let $\mathcal{A} = (Q, A, E, I, \mathcal{F})$ be a generalized Büchi automaton such that $|Q| = n$ and $\mathcal{F} = \{F\}$. There is an equivalent Büchi automaton $\mathcal{A}' =$*

(Q', A, E', I', F') such that $|Q'| \leq 2n$. Furthermore, if $\mathcal{A}$ is unambiguous (respectively complete), then $\mathcal{A}'$ is also unambiguous (respectively complete).

Proof Let $\mathcal{F}$ be the family $\{F\}$ where F is a subset of transitions of $\mathcal{A}$. The Büchi automaton $\mathcal{A}' = (Q', A, E', I', F')$ is defined as follows. The state set of $\mathcal{A}'$ is $Q \times \{0,1\}$ and the set I' of initial states is $I \times \{0,1\}$. There is a transition $(q, \varepsilon) \xrightarrow{a} (q', \varepsilon')$ if there is a transition $q \xrightarrow{a} q'$ in $\mathcal{A}$. The boolean value ε is equal to 1 iff the transition $q \xrightarrow{a} q'$ belongs to F and to 0 otherwise. The set F' of final states is $Q \times \{1\}$. □

The following proposition shows that the exponential blow up of the number of states cannot be avoided.

Proposition 18 *Let A_n be the alphabet $\{a_0, \ldots, a_n\}$ and let X_n be the set $(A^*a_1A^* \ldots A^*a_n)^\omega$. For any integer n, the set X_n is recognized by a generalized Büchi automaton with one state but any UBA recognizing the set X_n has at least $\binom{n}{n/2}^{\frac{1}{2}}$ states.*

By Stirling's formula, $\binom{2n}{n}^{\frac{1}{2}}$ is equivalent to $2^n(n\pi)^{-\frac{1}{4}}$ and grows exponentially with n. It may be remarked that the set X_n can be recognized by a deterministic Muller automaton with n states whereas any unambiguous Büchi automaton recognizing X_n has approximately 2^n states. The proof of the proposition is based on the following lemma. This lemma is interesting by itself. It is quite general and can be used in several cases to lower bound the number of states of an UBA.

Lemma 19 *Let $\mathcal{A}$ be an UBA with m states recognizing a set X. Suppose that the n pairs (v_k, w_k) of finite words are such that the infinite word $(v_k w_l)^\omega$ belongs to X iff $k = l$. Then the number m of states of $\mathcal{A}$ satisfies $m^2 \geq n$.*

Proof Since the automaton $\mathcal{A}$ is unambiguous, the final path labeled by $(v_k w_k)^\omega$ is, by Proposition 8, periodic. For any k, the word $v_k w_k$ labels a loop

$$p_k \xrightarrow{v_k} q_k \xrightarrow{w_k} p_k$$

which contains a final state. We claim that if k is different from l then the pair (p_k, q_k) is different from the pair (p_l, q_l). Indeed, if $p_k = p_l$ and $q_k = q_l$, we have another loop

$$p_k \xrightarrow{v_l} q_k \xrightarrow{w_l} p_k$$

which also contains a final state. By symmetry, we can suppose that the finite path $p_k \xrightarrow{v_l} q_k$ contains a final state. We consider the loop

$$p_k \xrightarrow{v_l} q_k \xrightarrow{w_k} p_k$$

which contains a final states. Since the infinite word $(v_k w_k)^\omega$ belongs to X, the state p_k is initial. The infinite word $(v_k w_l)^\omega$ belongs also to X and this is a contradiction. This proves that if k is different from l then the pair (p_k, q_k) is different from the pair (p_l, q_l). The inequality $n \leq m^2$ follows easily. $\square$

Actually, it can be seen that $p_k \neq q_k$ for any k. Thus, we get the slightly better inequality $n \leq m(m-1)$. We now prove the proposition using the previous lemma.

Proof Let P be a subset of $\mathbb{N}_{2n}$ containing n elements. Let $\{i_1, \ldots, i_n\}$ be the elements of P and let $\{j_1, \ldots, j_n\}$ be the elements of the complement of P. We define the words v_P and w_P by $v_P = a_{i_1} \ldots a_{i_n}$ and $w_P = a_{j_1} \ldots a_{j_n}$. The number of subsets P of size n is $\binom{2n}{n}$. Furthermore, the infinite word $(v_P w_{P'})^\omega$ belongs to X_{2n} iff $P = P'$. Applying the previous lemma, we get the result. $\square$

6.2 Proof with graphs

In this section, we give a first proof of Theorem 7. This proof is elementary and needs no particular background. From a Büchi automaton $\mathcal{A}$ recognizing a set X, this proof constructs a CUGBA $\hat{\mathcal{A}}$ recognizing X. Using the result of section 6.1, this automaton can be transformed into a CUBA recognizing X.

This proof is organized as follows. In Section 6.2.1, we define two operators on graphs. The two operators are used in Section 6.2.2 to define a particular function h associated with a graph. The automaton $\hat{\mathcal{A}}$ is described in Section 6.2.3. We prove in Section 6.2.4 that if $\mathcal{A}$ has m states, then $\hat{\mathcal{A}}$ has at most $(3m)^m$ states.

6.2.1 Operators int and att

In this section, we consider oriented graphs and we introduce notions that will be useful here. An oriented graph is a pair (V, E) where V is the set of vertices and the set E of edges is a subset of $V \times V$. The notion of path we have already defined for automata is of course valid for graphs. We use the same notation for edges in a graph as for transitions in an automaton. We write $x \to y$ to mean that (x, y) is an edge. A path in a graph is a sequence of consecutive edges. A path γ of length n is thus a sequence of n edges

$$x_0 \to x_1 \to \cdots \to x_n.$$

In this case, we write $x_0 \to^n x_n$ and we often identify the path with the sequence $x_0, \ldots, x_n$ of vertices appearing along the path. In this framework, an automaton is considered as a labeled graph. The set of transitions of the automaton is then seen as a family $\{\xrightarrow{a} \mid a \in A\}$ of relations indexed by the letters of the alphabet.

Let $G = (V, E)$ be a graph. For a subset X of vertices and any integer n, we denote by $\mathrm{s}^n(X)$ the set of vertices that can be reached from a vertex x of X by a path of length n. More formally, we have $\mathrm{s}^n(X) = \{y \mid \exists x \in X \;\; x \to^n y\}$. We respectively denote by $\mathrm{s}^*(X)$ and $\mathrm{s}^+(X)$ the sets $\bigcup_{n \geq 0} \mathrm{s}^n(X)$ and $\bigcup_{n \geq 1} \mathrm{s}^n(X)$. For a vertex x, we write $\mathrm{s}^n(x)$ instead of $\mathrm{s}^n(\{x\})$. An element of $\mathrm{s}^*(x)$ is called a *successor* of x whereas an element of $\mathrm{s}(x)$ is called an *immediate successor* of x.

We define the notions we need in a general framework since almost everything holds without any particular assumptions. However, we always suppose that for any vertex x of the graph, the set $\mathrm{s}(x)$ of immediate successors of x is finite. This assumption is not always necessary but it is often needed.

We now define two operators int and att which map subsets of vertices to subsets of vertices. These two operators are idempotent and increasing. For a set X of vertices, we define the interior of X as follows.

Definition 20 *Let $G = (V, E)$ be a graph and let X be a subset of vertices. The* interior *of X is $\mathrm{int}(X) = \{x \mid \mathrm{s}^*(x) \subset X\}$.*

The interior of X is the set of vertices x such that all its successors belong to X. If x belongs to the interior of X, all path starting in x remain in X. A subset X of vertices is called a *trap* iff $\mathrm{int}(X) = X$.

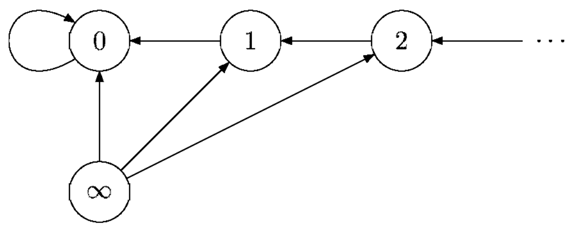

Figure 4: Graph of Example 21

Example 21 Let G be the graph $G = (\overline{\mathbb{N}}, E)$ where $E = \{0 \to 0\} \cup \{n+1 \to n \mid n \in \mathbb{N}\} \cup \{\infty \to n \mid n \in \mathbb{N}\}$. This graph is pictured in Figure 4. If X is the subset $\mathbb{N} \setminus \{1\}$, then $\mathrm{int}(X) = \{0\}$.

The following properties of the operator int are straightforward. We just mention them without proof.

Proposition 22 *The operator* int *satisfies the following properties.*

1. $\mathrm{int}(X) \subset X$;
2. $X \subset Y$ *implies* $\mathrm{int}(X) \subset \mathrm{int}(Y)$;
3. $X = \mathrm{int}(X) \iff (x \in X \implies \mathrm{s}(x) \subset X)$;
4. $\mathrm{int}(\mathrm{int}(X)) = \mathrm{int}(X)$ *and* $\mathrm{int}(X)$ *is the biggest trap contained in* X.

We now give a definition of the operator *att*. This definition may seem a bit complicated. In the sequel, we will only be interested in $\mathrm{att}(X)$ when the set X is a trap. In this case, we will give a simpler definition in Proposition 26. However we define $\mathrm{att}(X)$ in the general case since all properties remain valid.

Definition 23 *Let $G = (V, E)$ be a graph and let X be a subset of vertices. The* attractor *of X is* $\text{att}(X) = \bigcup_{n \geq 0} X_n$ *where $X_0 = X$ and $X_{n+1} = X_n \cup \{x \mid \text{s}(x) \subset X_n\}$.*

A vertex x belongs to X_n iff any path of length n and starting in x meets X. Therefore, a vertex x belongs to $\text{att}(X)$ if there is an integer n such that any path of length n and starting in x meets X. This implies that any infinite path starting in x meets X. A subset X of vertices is called an *attractor* iff $\text{att}(X) = X$. The following properties of the operator att are elementary and are given without proof.

Proposition 24 *The operator* att *satisfies the following properties*

1. $X \subset \text{att}(X)$;
2. $X \subset Y$ *implies* $\text{att}(X) \subset \text{att}(Y)$;
3. $X = \text{att}(X) \iff (\text{s}(x) \subset X \implies x \in X)$;
4. $\text{att}(\text{att}(X)) = \text{att}(X)$ *and* $\text{att}(X)$ *is the smallest attractor containing* X.

It must pointed out that the operator att is not idempotent if it is not assumed that the number of immediate successors of any vertex is finite. The following example shows that this hypothesis is really necessary.

Example 25 Consider again the graph of Example 21. Let X be the subset $\{0\}$. We have $X_n = \{0, \ldots, n\}$ and $\text{att}(X) = \mathbb{N}$. However, we have $\text{att}(\text{att}(X)) = \overline{\mathbb{N}}$.

The following propositions gives a simpler definition of $\text{att}(X)$ when the set X is a trap.

Proposition 26 *Let $G = (V, E)$ be a graph and let X be a subset of vertices. If X is a trap, then* $\text{att}(X)$ *is equal to* $\{x \mid \exists n \;\; \text{s}^n(x) \subset X\}$ *and* $\text{att}(X)$ *is also a trap.*

Proof Since X is a trap, the inclusion $\text{s}^n(x) \subset X$ for some integer n implies the inclusion $\text{s}^m(x) \subset X$ for any integer $m \geq n$. The attractor $\text{att}(X)$ is equal to $\bigcup_{n \geq 0} X_n$. An induction on n shows that $X_n = \{x \mid \text{s}^n(x) \subset X\}$ and this proves that $\text{att}(X) = \{x \mid \exists n \;\; \text{s}^n(x) \subset X\}$.

If $x \in \text{att}(X)$, then $\text{s}^n(x) \subset X$ for some integer n. For any $y \in \text{s}(x)$, we have $\text{s}^{n-1}(y) \subset \text{s}^n(x) \subset X$. This proves that $\text{s}(x) \subset \text{att}(X)$ and that $\text{att}(X)$ is a trap by Proposition 22(3). □

It is also true that if X is an attractor, then $\text{int}(X)$ is also an attractor but this result is not needed in the sequel. However, the two operators int and att do not commute.

Some relationship between logic and the operators we have just defined should be enlightened. The interior of X is equal to the set $\{x \mid \text{s}^*(x) \subset X\}$ which can be written $\{x \mid \forall n \;\; \text{s}^n(x) \subset X\}$. If the set X is a trap, the attractor of X is equal to $\{x \mid \exists n \;\; \text{s}^n(x) \subset X\}$. Therefore, the operator int is a kind of "universal" operator whereas the operator att is a kind of "existential" operator.

6.2.2 Function h

In this section, we consider graphs with a subset F of distinguished vertices. The vertices of F are called *final*. Through this section, the complement of F is denoted by $\overline{F}$. Given a graph and a fixed subset F of final vertices, we define a function h from V to $\overline{\mathbb{N}}$ canonically associated with F.

Definition 27 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. We define the sets $(X_k)_{k \geq -1}$ of vertices by $X_{-1} = \varnothing$ and*

$$X_k = \begin{cases} \mathrm{att}(X_{k-1}) & \text{if } k \text{ is odd} \\ \mathrm{int}(X_{k-1} \cup \overline{F}) & \text{if } k \text{ is even} \end{cases}$$

The function h is defined by $h(x) = \min\{k \geq 0 \mid x \in X_k\}$ if x belongs to X_k for some k and $h(x) = \infty$ otherwise.

The definition of the sets X_k uses alternately the operators int and att like an alternation of quantifiers. We illustrate this definition with the following examples. In the figure, the final vertices are, as final states of automata, marked by a double circle.

Example 28 If no vertex is final, i.e. $F = \varnothing$, then $h(x)$ is equal to 0 for any vertex x. Indeed, the set $X_0 = \mathrm{int}(\overline{F})$ is equal to V.

Example 29 If all vertices are final, i.e. $F = V$, then $h(x)$ is either equal to 1 or to ∞. Indeed, the set X_0 is empty and for any positive integer k, the set X_k is equal to X_1 .

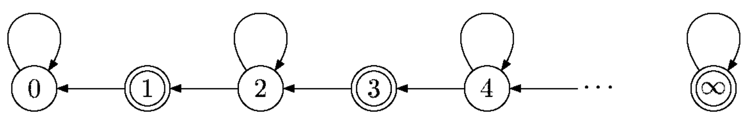

Figure 5: Graph of Example 30

Example 30 Let G be the graph $G = (\overline{\mathbb{N}}, E)$ where $E = \{n + 1 \to n \mid n \in \mathbb{N}\} \cup \{n \to n \mid n \in 2\mathbb{N}\} \cup \{\infty \to \infty\}$. This graph is pictured in Figure 5. Let $F = 2\mathbb{N} + 1 \cup \{\infty\}$ be the set containing the odd integers and ∞. For any integer k, the set X_k is equal to $\{0, \ldots, k\}$. Thus, for any vertex $x \in \overline{\mathbb{N}}$, we have $h(x) = x$.

Before characterizing the function h, we need some notation and definitions. If f is a function from the set V of vertices of a graph to $\overline{\mathbb{N}}$, we define a function $\vec{f}$ from V to $\overline{\mathbb{N}}$ by $\vec{f}(x) = \max\{f(y) \mid x \to y\}$ if $\mathrm{s}(x)$ is nonempty and by $\vec{f}(x) = 0$ otherwise. The notation $\vec{f}$ suggests that $\vec{f}(x)$ is equal to the maximal value of $f(y)$ for $x \to y$. When a function f from V to $\overline{\mathbb{N}}$ is fixed, we define for any vertex x and any $k \in \overline{\mathbb{N}}$ the set $\mathrm{s}^n_k(x)$ as $\mathrm{s}^n(x) \cap f^{-1}(k)$. A vertex y belongs to $\mathrm{s}^n_k(x)$ if y can be reached from x by a path of length n and if $f(y) = k$. This notation is naturally extended to $\mathrm{s}^*_k(x)$ and $\mathrm{s}^+_k(x)$ which respectively denote $\mathrm{s}^*(x) \cap f^{-1}(k)$ and $\mathrm{s}^+(x) \cap f^{-1}(k)$. We now define a relation between functions from V to $\overline{\mathbb{N}}$.

Definition 31 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. For two functions f and f' from V to $\overline{\mathbb{N}}$, we write $f' \to f$ if for any vertex x*

$$f'(x) = \begin{cases} \vec{f}(x) + 1 & \text{if } \vec{f}(x) \text{ is even and } x \in F \\ \vec{f}(x) & \text{otherwise.} \end{cases}$$

This relation is denoted by the symbol $\to$ to enlighten that it depends on the edges of the graphs. Furthermore, it will be used later to define transitions in an automaton. The relation $f' \to f$ is defined for a specific subset F of final vertices although it is omitted in the notation. It will be always clear from the context which subset F is referenced. It should be noticed that for a given function f, there always exists exactly one function f' such that $f' \to f$ as it is suggested by the notation. Thus, the relation $\to$ can be seen as an operator which maps f to f'.

We now define a family of relations. For each integer n, we define a relation e^n between a vertex and two functions f and f' from V to $\overline{\mathbb{N}}$. This relation will be mainly useful when the relation $f' \to f$ holds but this is not needed for the definition.

Definition 32 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. Let f and f' be two functions from V to $\overline{\mathbb{N}}$ and let x be a vertex. For any integer n we define a relation e^n as follows.*

$$\begin{array}{lll}
\mathrm{e}^n(f', x, f) \text{ always holds} & & \text{if } f'(x) = 0 \\
\mathrm{e}^n(f', x, f) & \stackrel{\text{def}}{\iff} \mathrm{s}^n(x) \cap f^{-1}(f'(x)) = \varnothing & \text{if } f'(x) \text{ is odd} \\
\mathrm{e}^n(f', x, f) & \stackrel{\text{def}}{\iff} \mathrm{s}^n(x) \cap f^{-1}(f'(x) - 1) \neq \varnothing & \text{if } f'(x) \text{ is even and } f'(x) \geq 2 \\
\mathrm{e}^n(f', x, f) & \stackrel{\text{def}}{\iff} \mathrm{s}^n(x) \cap f^{-1}(f'(x)) \cap F \neq \varnothing & \text{if } f'(x) = \infty
\end{array}$$

When $f' = f$, we write $\mathrm{e}^n(f, x)$ instead of $\mathrm{e}^n(f', x, f)$.

The relation $\mathrm{e}^n(f', x, f)$ holds if some special event happens for the successors of x. This event depends on the parity of $f'(x)$. If $f'(x)$ is odd, all successors $y \in \mathrm{s}^n(x)$ must satisfy $f(y) \neq f'(x)$. If $f'(x)$ is even, at least one successor $y \in \mathrm{s}^n(x)$ must satisfy $f(y) = f'(x) - 1$. If $f'(x) = \infty$, at least one successor $y \in \mathrm{s}^n(x)$ must be final and must satisfy $f(y) = f'(x)$. We point out that if $f'(x)$ is odd, the event is "universal" in the sense that all the successor of $\mathrm{s}^n(x)$ must satisfy some property. On the contrary, if $f'(x)$ is even or if $f'(x) = \infty$, the event is "existential" in the sense that one successor of $\mathrm{s}^n(x)$ must satisfy some property.

It has always been supposed that in the graph the number of immediate successors of any vertex x is finite. The characterization of the function h we give in the following theorem needs a stronger property of the graph that we now introduce.

Definition 33 (Bounded graph) *A graph $G = (V, E)$ is said to be* bounded by M *if for any vertex x and any integer n the number of vertices in $\mathrm{s}^n(x)$ is bounded by M. A graph is said to be* bounded *if it is bounded by M for some constant M.*

For instance, the graph of Example 30 in not bounded. The following theorem provides a characterization of the function h associated with the subset of a graph, when the graph is bounded.

Theorem 34 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. The function h canonically associated with F satisfies*

$$h \to h. \tag{1}$$

If the graph G is bounded, the function h is the only function from V to $\overline{\mathbb{N}}$ satisfying Equation (1) such that for any vertex x there exists a positive integer n such that $\mathrm{e}^n(h, x)$ holds.

Equation (1) is a local property since it gives a relation between the value of h for x and the values of h for the immediate successors of x. The value of $h(x)$ only depends on the values of $h(y)$ for $y \in \mathrm{s}(x)$ and on the membership of x to the set F. Since function h satisfies Equation (1), it is clear that $h(y) \leq h(x)$ for any $y \in \mathrm{s}(x)$. This immediately implies that $h(y) \leq h(x)$ for any $y \in \mathrm{s}^*(x)$. This means that the value of h decreases along paths.

Equation (1) does not fully characterizes the function h canonically associated with F. There may exist other functions satisfying this equation as we will see in Example 35. An additional property is needed to completely characterize the function h. For any vertex, the relation $\mathrm{e}^n(f, x)$ must hold for some integer n. In the sequel, we say that a function f satisfies the property PO (respectively PE and PI) if for any vertex such that $f(x)$ is Odd (respectively Even and equal to Infinity) there exists an integer n such that $\mathrm{e}^n(f, x)$.

Before proceeding to the proof of Theorem 34, the following example shows that the additional property is really necessary.

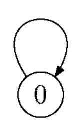

Figure 6: Graph of Example 35

Example 35 Let G be the graph $G = (\{0\}, \{0 \to 0\})$ pictured in Figure 6. If the vertex is not final, any function f from $\{0\}$ to $\overline{\mathbb{N}}$ satisfies Equation (1). If the vertex is final, a function f satisfies Equation (1) if $f(0)$ is odd or equal to ∞. In both cases, the function h is not the only function satisfying Equation (1).

We now come to the proof of Theorem 34. The proof is divided in several lemmas. The proof of the theorem is organized as follows. We first verify that the function h satisfies Equation (1) and Properties PO, PE and PI. Then we prove that any function g also satisfying these properties is equal to h.

Lemma 36 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. The function h canonically associated with F satisfies Equation (1).*

Proof We claim that, for any integer k, the set X_k introduced in Definition 27 is a trap and that $X_k \subset X_{k+1}$. If k is even, $X_k = \text{int}(X_{k-1} \cup \overline{F})$ is a trap by Proposition 22(4). If k is odd, $X_k = \text{att}(X_{k-1})$ is a trap by Proposition 26. The inclusion $X_{k-1} \subset X_k$ follows directly from Proposition 24(1). If k is even, we have $X_{k-1} = \text{int}(X_{k-1}) \subset \text{int}(X_{k-1} \cup \overline{F}) = X_k$ by Proposition 22(2). Since each X_k is a trap, the inequality $\vec{h}(x) \leq h(x)$ holds.

If $\text{s}(x)$ is empty, the vertex x belongs to X_0 if $x \notin F$ and to X_1 otherwise. Equation (1) holds since $\vec{h}(x) = 0$.

We now suppose that $\text{s}(x)$ is nonempty. If $\vec{h}(x) = \infty$, the equality $h(x) = \vec{h}(x)$ holds and Equation (1) is satisfied. We now suppose that $k = \vec{h}(x)$ is an integer. If k is odd, then X_k is an attractor and x belongs to X_k since $\text{s}(x) \subset X_k$. Thus, we have $h(x) = k$. We now suppose that k is even and that $X_k = \text{int}(X_{k-1} \cup \overline{F})$. If x does not belong to F, then x belongs to X_k by definition of int and $h(x) = \vec{h}(x)$. It remains the case when $x \in F$. Since $X_{k-1} \subset X_k \subset X_{k-1} \cup \overline{F}$, x cannot belong to X_k but it belongs to $X_{k+1} = \text{att}(X_k)$ since $\text{s}(x) \subset X_k$. In this case, we have $h(x) = \vec{h}(x) + 1$. Finally, Equation (1) is always satisfied. $\square$

The following lemma states that if a function f satisfies Equation (1), the image by f of a final vertex cannot be even. This is in particular true for the function h but this lemma will be also applied to other functions in the proof that h is the unique function satisfying Equation (1) and the additional property.

Lemma 37 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. If a function f from V to $\overline{\mathbb{N}}$ satisfies Equation (1), the sets $\{x \mid f(x)$ is even$\}$ and F are disjoint.*

Proof If $\vec{f}(x)$ is odd, one necessarily has $f(x) = \vec{f}(x)$ and $f(x)$ is odd. If $\vec{f}(x)$ is even and $x \in F$, one has $f(x) = 1 + \vec{f}(x)$ and $f(x)$ is also odd. Thus, if $f(x)$ is even, we have $x \notin F$ and the claim is proved. $\square$

We now prove that the function h satisfies properties PO and PE.

Lemma 38 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. The function h canonically associated with F satisfies properties PO and PE.*

Proof If $k = h(x)$ is odd, x belongs to $X_k = \mathrm{att}(X_{k-1})$. Since X_{k-1} is a trap, there exists, by Proposition 26, an integer n such that $\mathrm{s}^n(x) \subset X_{k-1}$. Thus, the set $\mathrm{s}^n_{h(x)}(x)$ is empty and property PO is satisfied.

If $k = h(x)$ is even and greater than 2, x belongs to $X_k = \mathrm{int}(X_{k-1} \cup \overline{F})$ and one has $\mathrm{s}^*(x) \subset X_{k-1} \cup \overline{F}$. If there does exist any $y \in \mathrm{s}^+(x)$ such that $h(y) = k - 1$, one has $\mathrm{s}^*(x) \subset X_{k-2} \cup \overline{F}$. Since k is even, one has the inclusion $X_{k-2} \cup \overline{F} \subset X_{k-3} \cup \overline{F}$ by Lemma 37. Finally one gets $\mathrm{s}^*(x) \subset X_{k-3} \cup \overline{F}$ and x belongs to $X_{k-2} = \mathrm{int}(X_{k-3} \cup \overline{F})$. This is a contradiction and we have proved that there exists $y \in \mathrm{s}^+(x)$ such that $h(y) = k - 1$. Thus, the set $\mathrm{s}^+_{h(x)-1}(x)$ is nonempty and property PE is satisfied. □

The following lemma states that Equation (1) and properties PO and PE implies that a vertex has successors y such that $f(y) = k$ for any $1 \leq k \leq f(x)$. It actually gives a result which is slightly more precise. The result holds for the function h since we have already proved that it satisfies Equation (1) and properties PO and PE.

Lemma 39 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. Let f be a function satisfying Equation (1) and Properties PO and PE. Let x be a vertex such that $f(x)$ is finite. For any integer $1 \leq k \leq f(x)$, $\mathrm{s}^*_k(x) \neq \varnothing$. More precisely, if $1 \leq k \leq f(x) - 1$ is odd, $\mathrm{s}^n_k(x) \cap F \neq \varnothing$ for an infinite number of integers n and if $2 \leq k \leq f(x)$ is even, $\mathrm{s}^n_k(x) \neq \varnothing$ for almost any integer n.*

Proof We first prove that if $f(x)$ is odd, the set $\mathrm{s}^*_{f(x)}(x) \cap F$ is nonempty and if furthermore $f(x) \geq 3$, the set $\mathrm{s}^*_{f(x)-1}(x)$ is also nonempty. By property PO, there exists a least integer n such that the set $\mathrm{s}^n_{f(x)}$ is empty. By definition of n, there is a vertex $y \in \mathrm{s}^{n-1}(x)$ such that $f(y) = f(x)$. We claim that y belongs to F. Since $\mathrm{s}(y) \subset \mathrm{s}^n(x)$, the set $\mathrm{s}_{f(y)}(y)$ is also empty and the inequality $f(y) > \vec{f}(y)$ holds. Equation (1) then implies that $f(y) = 1 + \vec{f}(y)$ and that $y \in F$. Thus, y belongs to $\mathrm{s}^*_{f(x)}(x) \cap F$. If $f(x) \geq 3$, the set $\mathrm{s}(y)$ cannot be empty and the set $\mathrm{s}_{f(y)-1}(y)$ is nonempty. The inclusion $\mathrm{s}(y) \subset \mathrm{s}^*(x)$ implies that $\mathrm{s}^*_{f(x)-1}(x)$ is nonempty.

We now prove that $\mathrm{s}^*_k(x)$ is nonempty for any integer $1 \leq k \leq f(x)$. The proof is by induction on the difference $f(x) - k$. If $k = f(x)$, it is clear that $\mathrm{s}^*_k(x)$ is nonempty since $x \in \mathrm{s}^*(x)$. We now assume that $k \geq 2$ and that $y \in \mathrm{s}^*_k(x)$ and we claim that $\mathrm{s}^*_{k-1}(x)$ is nonempty. If k is odd, the set $\mathrm{s}^*_{f(y)-1}(y)$ is nonempty and the inclusion $\mathrm{s}^*(y) \subset \mathrm{s}^*(x)$ implies that $\mathrm{s}^*_{k-1}(x)$ is nonempty. If k is even, the set $\mathrm{s}^*_{f(y)-1}(y)$ is nonempty by property PE and $\mathrm{s}^*_{k-1}(x)$ is also nonempty.

We now prove that if k is even, the set $\mathrm{s}^n_k(x)$ is non empty for almost any integer n. Since the set $\mathrm{s}^*_k(x)$ is nonempty, it suffices to prove the result when $k = f(x)$. We claim that if $f(x)$ is even, $\mathrm{s}^n_{f(x)}(x)$ is nonempty for any integer n. The proof is by induction on the integer n. The result obviously holds when

$n = 0$. Let y be a vertex of $\mathrm{s}^n(x)$ such that $f(y) = f(x)$. Since the function f satisfies Equation (1) and since $f(y)$ is even, one has $\vec{f}(y) = f(y)$. Thus, there exists a vertex $z \in \mathrm{s}(y)$ such that $f(z) = f(x)$. Thus, the set $\mathrm{s}^{n+1}_{f(x)}(x)$ is nonempty.

We finally prove that if k is odd and $1 \leq k \leq f(x) - 1$, the set $\mathrm{s}^n_k(x) \cap F$ is nonempty for infinitely many integers n. Since the set $\mathrm{s}^n_{k+1}(x)$ is nonempty for almost any integer n, the set $\mathrm{s}^n_k(x)$ is, by property PE, nonempty for infinitely many integers n. It has already been proved that if $f(y)$ is odd, the set $\mathrm{s}^*_{f(y)}(y) \cap F$ is nonempty. Thus, the set $\mathrm{s}^n_k(x) \cap F$ is also nonempty for infinitely many integers n. □

Up to now, the hypothesis that the graph is bounded has not been used. The following lemma states that if the underlying graph is bounded, any function satisfying Equation (1) is also bounded.

Lemma 40 *Let $G = (V, E)$ be a graph bounded by M and F be the subset of final vertices. Let f be a function satisfying Equation (1) and Properties PO and PE. Either $f(x) = \infty$ or $f(x) \leq 2M - 1$.*

Proof Let x be a vertex such that $f(x)$ is finite. By Lemma 39, there exists an integer m such that for any even integer $2 \leq k \leq f(x)$, and any integer n larger than m, the set $\mathrm{s}^n_k(x)$ is nonempty. Furthermore, since $\mathrm{s}^m_2(x)$ is nonempty, the set $\mathrm{s}^n_1(x)$ is, by Lemma 39, nonempty for some integer n greater than m. Finally, the set $\mathrm{s}^n(x)$ contains at least $\lfloor f(x)/2 \rfloor + 1$ elements. Since the graph is bounded by M, one has $\lfloor f(x)/2 \rfloor \leq M - 1$ and $f(x) \leq 2M - 1$. □

We are now able to prove that the function h canonically associated with F satisfies Property PI when the underlying graph is bounded.

Lemma 41 *Let $G = (V, E)$ be a bounded graph and let F be the subset of final vertices. The function h canonically associated with F satisfies property PI.*

Proof Let x be a vertex such that $h(x) = \infty$. By Lemma 40, for any vertex $y \in \mathrm{s}^+(x)$, one has either $h(y) = \infty$ or $h(y) \leq 2M - 1$. If $\mathrm{s}^+(x) \subset X_{2M-1} \cup \overline{F}$ then any immediate successor y of x belongs to $X_{2M} = \mathrm{int}(X_{2M-1} \cup \overline{F})$ and x belongs to $X_{2M+1} = \mathrm{att}(X_{2M})$. This is a contradiction with $h(x) = \infty$. There exists $y \in \mathrm{s}^+(x)$ such that $h(y) = \infty$ and $y \in F$. Thus, $\mathrm{s}^+_\infty(x) \cap F$ is nonempty and property PI is satisfied. □

The following example shows that the assumption that the graph is bounded is necessary.

Example 42 Let G be the graph $(\mathbb{Z}, E)$ where $E = \{n \to n - 1 \mid n \in \mathbb{Z}^*\} \cup \{-n \to n \mid n \in \mathbb{N}^*\} \cup \{n \to n \mid n \in 2\mathbb{N}\}$. This graph is pictured in Figure 7. It may be easily verified that this graph is not bounded. Let F be the subset $F = 2\mathbb{N} + 1$ of odd positive integers. It can be verified that $h(n) = n$ if $n \geq 0$ and that $h(n) = \infty$ if $n < 0$. Although $h(-1) = \infty$, there does not exist

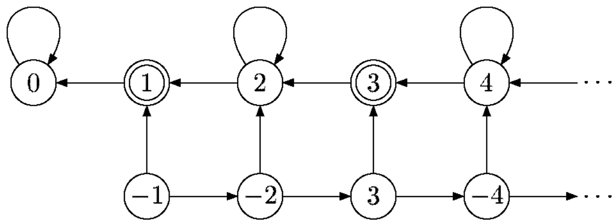

Figure 7: Graph of Example 42

$y \in \mathrm{s}^*(-1)$ such that $h(y) = \infty$ and $y \in F$. Thus, property PI is not satisfied by the function h.

The following lemma gives a characterization of the vertices x such that $f(x) = 0$ for a function f satisfying Equation (1) and Properties PO, PE and PI. This characterization is independent of the function f.

Lemma 43 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. Let f be a function satisfying Equation (1) and Properties PO, PE and PI. For any vertex x, $f(x) = 0$ holds if and only if $\mathrm{s}^*(x) \cap F = \varnothing$.*

Proof If $f(x) = 0$, Equation (1) implies that $f(y) = 0$ for any $y \in \mathrm{s}^*(x)$. By Lemma 37, the equality $\mathrm{s}^*(x) \cap F = \varnothing$ holds.

If $f(x) = \infty$, the set $\mathrm{s}^+_\infty(x) \cap F$ is nonempty by property PI and the equality $\mathrm{s}^*(x) \cap F = \varnothing$ does not hold.

If $1 \le f(x) < \infty$, the set $\mathrm{s}^*_1(x) \cap F$ is nonempty by Lemma 39 is nonempty and the equality $\mathrm{s}^*(x) \cap F = \varnothing$ does not hold. $\square$

The following lemma gives a characterization of the vertices x such that $f(x) = \infty$ for a function f satisfying Equation (1) and Properties PO, PE and PI. This characterization is independent of the function f.

Lemma 44 *Let $G = (V, E)$ be a bounded graph and let F be the subset of final vertices. Let f be a function satisfying Equation (1) and Properties PO, PE and PI. A vertex x satisfies $f(x) = \infty$ if and only if x is the starting vertex of a path going infinitely often through F.*

Proof Since h satisfies the property PI, it is clear that any vertex x satisfying $f(x) = \infty$ is the starting vertex of a path going infinitely often through F.

Conversely, we claim that almost all vertices of a path starting at a vertex x such that $h(x)$ is finite are in $\overline{F}$. The proof is by induction on $f(x)$. If $f(x) = 0$, one also has $f(y) = 0$ for any vertex $y \in \mathrm{s}^*(x)$ since the function h satisfies Equation (1). By Lemma 37, all vertices of a path starting at x are in $\overline{F}$.

If $f(x)$ is odd, there exists, by Property PO, an integer n such that $\mathrm{s}^n_{f(x)}(x) = \varnothing$. Thus, any path of length greater than n has a vertex y such that $f(y) < f(x)$. By the induction hypothesis almost all vertices of the path are in $\overline{F}$.

Suppose that $f(x)$ is even. If γ is a path starting at x, either there exists a vertex y of γ such that $f(y) < f(x)$ or all vertices y of γ satisfy $f(y) = f(x)$. In the former case, the induction hypothesis applies and the result is true. In the latter case, Lemma 37 implies that all vertices of the path are in $\overline{F}$. $\square$

Up to now, we have proved that the function h satisfies Equation (1) and Properties PO, PE and PI. The following lemma states that the function h is actually the only function from V to $\overline{\mathbb{N}}$ satisfying these properties. It completes the proof of Theorem 34.

Lemma 45 *Let $G = (V, E)$ be a bounded graph and let F be the subset of final vertices. The function h canonically associated with F is the unique function satisfying Equation (1) and Properties PO, PE and PI.*

Proof By Lemmas 36, 38 and 41, the function h canonically associated with F satisfies Equation (1) and Properties PO, PE and PI.

We now prove that h is the only function satisfying these properties. Let f and g be two functions from V to $\overline{\mathbb{N}}$ satisfying Equation (1) and Properties PO, PE and PI. We claim that $f = g$.

By Lemma 44, both sets $f^{-1}(\infty)$ and $g^{-1}(\infty)$ are equal to the sets of vertices x such that a path going infinitely often through F starts at x. Thus, one has $f(x) = \infty$ if and only if $g(x) = \infty$.

We now prove that if $f(x)$ or $g(x)$ is finite, the equality $f(x) = g(x)$ holds. The proof is by induction on $\min(f(x), g(x))$. By Lemma 43, one has $f(x) = 0$ if and only if $g(x) = 0$.

We now assume that if $f(x) < m$ or $g(x) < m$, the equality $f(x) = g(x)$ holds and we prove that it also holds if $\min(f(x), g(x)) = m$. By symmetry, it can be assumed that $f(x) \leq g(x)$ and that $f(x) = m$.

If $f(x)$ is odd, there exists, by Property PO, an integer n such that for any $y \in \mathrm{s}^n(x)$, one has $f(y) < f(x)$. If $f(x) < g(x)$, there exists, by Lemma 39, another integer $n' > n$ such that, there exists $y' \in \mathrm{s}^{n'}(x)$ such that $g(y') = h(x) + 1$. Since $n' > n$, one has $f(y') < m$ and this is a contradiction with the induction hypothesis, since $f(y') < m < g(y')$.

We now suppose that $f(x)$ is even. Since $f(x)+1$ is odd and since $f(x)+1 \leq g(x)$, there exists, by Lemma 39, $y \in \mathrm{s}^*(x) \cap F$ such that $g(y) = f(x) + 1$. Since y belongs to $\mathrm{s}^*(x)$, one has $f(y) \leq f(x)$ and since $x \in F$, one has $f(y) < f(x)$. This is a contradiction with the induction hypothesis, since $f(y) < m < g(y)$. $\square$

For a function f, we define a preorder $\prec_f$ which compares vertices according to the number of steps n needed such that $\mathrm{e}^n(f, x)$.

Definition 46 *Let $G = (V, E)$ be a graph and F be the subset of final vertices. Let f be a function from V to $\overline{\mathbb{N}}$. For any vertices x and y, the preorder $\prec_f$ on V is defined by*

$$x \prec_f y \overset{\text{def}}{\iff} f(x) = f(y) \text{ and } \hat{f}(x) \leq \hat{f}(y).$$

where $\hat{f}(x)$ is the least positive integer n such that $\mathrm{e}^n(f,x)$ if such an integer exists and $\hat{f}(x) = \infty$ otherwise.

The preorder $\prec_f$ may not be an order since it may happen that $f(x) = f(y)$ and $\hat{f}(x) = \hat{f}(y)$ for two distinct vertices x and y. By definition, the preorder $\prec_f$ is compatible with the function f.

Definition 47 *Let $G = (V,E)$ be a graph and F be the subset of final vertices. Let f and f' be two functions from V to $\overline{\mathbb{N}}$ such that $f' \to f$ and let $\prec$ be a preorder compatible with f. For a vertex x, we denote by $\vec{x}$ the class of a maximal (respectively minimal) element of the set $\{y \mid x \to y \text{ and } f(y) = f'(x)\}$ if $f'(x)$ is odd (respectively $f'(x)$ is even or equal to ∞).*

It may happen that the set $\{y \mid x \to y \text{ and } f(y) = f'(x)\}$ is empty and the class $\vec{x}$ is then undefined. However, this set is not empty if $\mathrm{e}(f',x,f)$ does not hold. It will be always the case in the sequel. Since the preorder $\prec$ is compatible with f, the class of an element y is always contained in the set $\{z \mid f(z) = f(y)\}$. Thus, the class $\vec{x}$ is always contained in the set $\{y \mid f(y) = f'(x)\}$.

Definition 48 *Let $G = (V,E)$ be a graph and F be the subset of final vertices. Let f and f' be two functions from V to $\overline{\mathbb{N}}$ and let $\prec$ and $\prec'$ be two preorders respectively compatible with f and f'. We write $(f',\prec') \to (f,\prec)$ iff $f' \to f$ and if for any vertices x and y such that $f'(x) = f'(y)$*

$$x \prec' y \iff \mathrm{e}(f',x,f) \vee (\neg\, \mathrm{e}(f',x,f) \wedge \neg\, \mathrm{e}(f',y,f) \wedge \vec{x} \prec \vec{y})$$

Since $f'(x) = f'(y)$ both classes $\vec{x}$ and $\vec{y}$ are contained in the set $\{z \mid f(z) = f'(x)\}$. These classes are then comparable for the preorder $\prec$ which is compatible with f. The preorder $\prec'$ can be described in the following way. Since it is compatible with f', it suffices to describe its restriction to $f'^{-1}(k)$ for any $k \in \overline{\mathbb{N}}$. Let k be a fixed element of $\overline{\mathbb{N}}$. Any element $x \in f'^{-1}(k)$ such that $\mathrm{e}(f',x,f)$ is smaller than any other element. Thus, if at least one element satisfies $\mathrm{e}(f',x,f)$, the smallest class of $f'^{-1}(k)$ is formed of the elements such that $\mathrm{e}(f',x,f)$. Other elements are compared with respect to their class $\vec{x}$. If x and y are equivalent for $\prec'$, then either both relations $\mathrm{e}(f',x,f)$ and $\mathrm{e}(f',y,f)$ hold or none of them holds. In the latter case, the classes $\vec{x}$ and $\vec{y}$ are equal.

It should be noticed that for a given pair $(f,\prec)$, there always is exactly one pair $(f',\prec')$ such that $(f',\prec') \to (f,\prec)$. Indeed, it has already been mentioned that there is exactly one function f' such that $f' \to f$ and the additional condition on $\prec'$ determines it.

Theorem 49 *Let $G = (V,E)$ be a graph and F be the subset of final vertices. Let h be the function canonically associated with F. We then have*

$$(h,\prec_h) \to (h,\prec_h)$$

The proof of the theorem is based on the following technical lemma. This lemma states that if a function f satisfies Equation (1), the function $\hat{f}$ also satisfies some local properties.

Lemma 50 *If the function f satisfies $f \to f$, the function $\hat{f}$ used in the Definition 46 satisfies the following property. For any vertex x, if $\mathrm{e}(f, x)$ holds then $\hat{f}(x) = 1$ and otherwise $\hat{f}(x)$ equals*

$$\begin{array}{ll} 1 + \max\{\hat{f}(y) \mid x \to y \text{ and } f(y) = f(x)\} & \text{if } f(x) \text{ is odd} \\ 1 + \min\{\hat{f}(y) \mid x \to y \text{ and } f(y) = f(x)\} & \text{if } f(x) \text{ is even or } f(x) = \infty \end{array}$$

It may be remarked that the relation satisfied by $\hat{f}$ involves a maximum if $f(x)$ is odd and a minimum if $f(x)$ is even or $f(x) = \infty$. This is due to the fact that $\mathrm{e}^n(f, x)$ is a universal event in the former case whereas it is an existential one in the latter case.

Proof We first suppose that $k = f(x)$ is odd. If $\hat{f}(x) > 1$, the set $\mathrm{s}_k(x)$ is nonempty and the set $\mathrm{s}_k^{n+1}(x)$ is equal to $\mathrm{s}_k^n(\mathrm{s}_k(x))$ and it is empty iff $\mathrm{s}_k^n(y)$ is empty for any $y \in \mathrm{s}_k(x)$. The two other cases when $f(x)$ is even or $f(x) = \infty$ are very similar. □

We now complete the proof of Theorem 49.

Proof Theorem 34 already states that the function h associated with F satisfies $h \to h$. It remains to show that the preorders $\prec$ and $\prec'$ fulfill the additional conditions. The rest of the proof follows easily from Lemma 50 and from the definitions of $\prec_h$ and $\vec{x}$. □

6.2.3 Automaton $\hat{\mathcal{A}}$

In this section, we describe the CUGBA $\hat{\mathcal{A}}$ associated with a Büchi automaton $\mathcal{A}$. This section is organized as follows. We first define a family of graphs $G_{\mathcal{A}}(x)$. We use these graphs and the notions introduced in Section 6.2.2 to describe the automaton $\hat{\mathcal{A}}$. We finally prove that this automaton is really unambiguous and complete and that it recognizes the same set as $\mathcal{A}$.

Let $\mathcal{A}$ be a Büchi automaton with m states. With any infinite word x, we associate a graph $G_{\mathcal{A}}(x)$. This graph is an unfolding of the automaton $\mathcal{A}$ and summarizes all the paths labeled by x in $\mathcal{A}$.

Definition 51 *Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi automaton and $x = a_0a_1a_2\ldots$ be an infinite word over A. The graph $G_{\mathcal{A}}(x)$ is the graph (V, E') where the set of vertices is $V = Q \times \mathbb{N}$ and the set of edges is*

$$E' = \{(p, n) \to (q, n+1) \mid n \in \mathbb{N} \text{ and } p \xrightarrow{a_n} q \text{ in } \mathcal{A}\}.$$

The set of final vertices of $G_{\mathcal{A}}(x)$ is $F \times \mathbb{N}$.

The set of vertices of $G_{\mathcal{A}}(x)$ is $Q \times \mathbb{N}$ and the subset of final vertices is equal to $F \times \mathbb{N}$. Both sets do not depend on x. A vertex (q, n) is final if the state q is a final state of $\mathcal{A}$. We illustrate this definition with the following example.

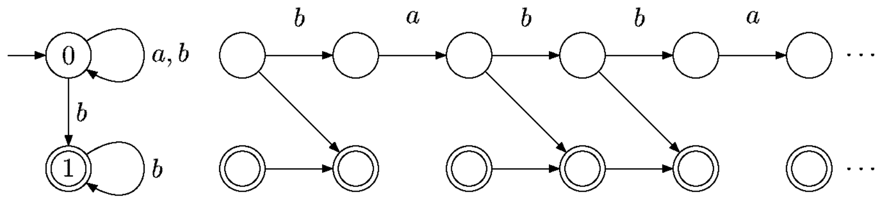

Figure 8: Automaton $\mathcal{A}$ and graph $G_{\mathcal{A}}(babba^\omega)$

Example 52 Let $\mathcal{A}$ be the Büchi automaton $(\{0,1\}, A, E, \{0\}, \{1\})$ and let x be the ultimately periodic infinite word $babba^\omega$. The automaton $\mathcal{A}$ and the graph $G_{\mathcal{A}}(x)$ are pictured in Figure 8.

The edges of $G_{\mathcal{A}}(x)$ depend on the infinite word x. The edges of the graph $G_{\mathcal{A}}(x)$ are defined in such a way that there is correspondence between infinite paths in $G_{\mathcal{A}}(x)$ and paths in $\mathcal{A}$ labeled by suffixes of x. Any infinite path γ of $G_{\mathcal{A}}(x)$ starting in some vertex (q_0, l) is of the form $(q_0, l)(q_1, l+1)(q_2, l+2)\dots$ where $q_n \xrightarrow{a_{l+n}} q_{n+1}$ is a transition of $\mathcal{A}$ for any integer n. Thus, the path $q_0q_1q_2\dots$ is a path of $\mathcal{A}$ labeled by the suffix $x_l = a_la_{l+1}a_{l+2}\dots$ of x. Conversely, any path $q_0q_1q_2\dots$ labeled by some suffix x_l of x gives the path $(q_0, l)(q_1, l+1)(q_2, l+2)\dots$ in $G_{\mathcal{A}}(x)$. Furthermore, since the set of final vertices is $F \times \mathbb{N}$, the path $q_0q_1q_2\dots$ is a final path of $\mathcal{A}$ iff the corresponding path in $G_{\mathcal{A}}(x)$ goes infinitely often through final vertices.

The graph $G_{\mathcal{A}}(x)$ is bounded by the number m of states of $\mathcal{A}$. Indeed, for any vertex $x = (q, l)$, the set $\mathrm{s}^n(x)$ is a subset of $Q \times \{l+n\}$ and contains at most m elements. For an infinite word x , we denote by h_x the function from V to $\overline{\mathbb{N}}$ given by Definition 27 and we denote by $\prec_x$ the preorder $\prec_{h_x}$ given by Definition 46. Since the graph $G_{\mathcal{A}}(x)$ is bounded, Theorem 34 is relevant for the function h_x. Furthermore, by Lemma 40, the function h_x is actually a function from V to $\overline{\mathbb{N}}_{2m-1}$.

Since the set of vertices of $G_{\mathcal{A}}(x)$ is $V = Q \times \mathbb{N}$, any function h from V to $\overline{\mathbb{N}}$ canonically defines a sequence $(h_n)_{n\geq 0}$ of functions from Q to $\overline{\mathbb{N}}$ given by $h_n(q) = h(q, n)$. Likewise, any preorder $\prec$ on V defines a sequence $(\prec_n)_{n\geq 0}$ of preorders on Q given by $q \prec_n q' \iff (q, n) \prec (q', n)$. Furthermore, if the preorder $\prec$ is compatible with the function h, each preorder $\prec_n$ is compatible with the function h_n. Furthermore, the function h satisfies $h \to h$ iff $h_n \xrightarrow{a_n} h_{n+1}$ holds for any integer n. Indeed, the immediate successors of the states of $Q \times \{n\}$ belong to $Q \times \{n+1\}$. If the pair $(h, \prec)$ satisfies $(h, \prec) \to (h, \prec)$, then $(h_n, \prec_n) \xrightarrow{a_n} (h_{n+1}, \prec_{n+1})$ holds for any integer n. Since the preorders $\prec_n$ do not fully determine the preorder $\prec$, the converse does not necessarily hold but this does not matter for our concern.

In particular the function h_x defines a sequence of pairs $(h_{x,n}, \prec_{x,n})$ where each $h_{x,n}$ is a function from Q to $\overline{\mathbb{N}}_{2m-1}$ and $\prec_{x,n}$ is a preorder on Q compatible with $h_{x,n}$. Furthermore, by Theorem 49, the relation $(h_{x,n}, \prec_{x,n}) \xrightarrow{a_n}$

$(h_{x,n+1}, \prec_{x,n+1})$ holds for any integer n.

In the sequel we use the pairs $(h, \prec)$ as states of the automaton $\hat{\mathcal{A}}$. We now come to the definition of an automaton $\hat{\mathcal{A}}$ constructed from an automaton $\mathcal{A}$. Let $\mathcal{A}$ be a Büchi automaton (Q, A, E, I, F) with m states. We define a general Büchi automaton $\hat{\mathcal{A}} = (\hat{Q}, A, \hat{E}, \hat{I}, \mathcal{F})$.

The states set $\hat{Q}$ of $\mathcal{A}$ is the set of pairs $(h, \prec)$ where h is a function from Q to $\overline{\mathbb{N}}_{2m-1}$ and $\prec$ is a preorder on Q compatible with h. The set $\hat{I}$ of initial states is the set of pairs $(h, \prec)$ such that $h(q) = \infty$ for some initial state $q \in I$ of $\mathcal{A}$.

We now describe the transitions of the automaton $\hat{\mathcal{A}}$. Let $p = (h, \prec)$ and $p' = (h', \prec')$ be two states of $\hat{\mathcal{A}}$. Then $p' \xrightarrow{a} p$ is a transition of $\hat{\mathcal{A}}$ iff

$$(h', \prec') \xrightarrow{a} (h, \prec)$$

The family $\mathcal{F}$ contains $2m$ subsets of transitions called $F_1, \ldots, F_{2m-1}$ and F_∞. A transition $\tau = (h', \prec') \xrightarrow{a} (h, \prec)$ belongs to the set F_k for $k \in \overline{\mathbb{N}}^*_{2m-1}$ if $h'^{-1}(k)$ does not contain a state q such that $\mathrm{e}(h', q, h)$ does not hold and such that the class $\vec{q}$ is the maximal class of $h^{-1}(k)$. We also point out that if the set $h'^{-1}(k)$ is empty, then τ belongs to F_k.

We illustrate the construction with the two following examples. In the previous construction, some pairs $(h, \prec)$ are not coaccessible states and can be thus removed without changing the behavior of the automaton. In the Figure 10 and 12, only the coaccessible states are shown. If the automaton $\mathcal{A}$ has one state, i.e. $Q = \{1\}$, preorder on Q must be trivial. It suffices to label a state of $\hat{\mathcal{A}}$ by the value $h(1)$. If the automaton $\mathcal{A}$ has two states, i.e. $Q = \{1, 2\}$, either $h(1) \neq h(2)$ or $h(1) = h(2)$. In the former case, the preorder $\prec$ must be trivial and it suffices to label a state of $\hat{\mathcal{A}}$ by the values $h(1), h(2)$. In the latter case, either the states 1 and 2 are equivalent or $1 \prec 2$ or $1 \succ 2$. A state of $\hat{\mathcal{A}}$ is then labeled by the values $h(1), h(2)$ with an additional symbol from $\{\sim, \prec, \succ\}$ in between the two values $h(1)$ and $h(2)$. In the Figure 10 and 12, a transition $\tau = (h', \prec') \xrightarrow{a} (h, \prec)$ is drown by an arrow from $(h', \prec')$ to $(h, \prec)$. This arrow is labeled by $a/k_1 \ldots k_r$ where $k_1, \ldots, k_r$ are the elements of $\overline{\mathbb{N}}^*_{2m-1}$ such that the transition $(h', \prec') \xrightarrow{a} (h, \prec)$ belongs to F_{k_j}.

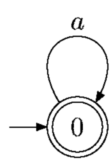

Figure 9: Automaton $\mathcal{A}$ of Example 53

Example 53 Let $\mathcal{A}$ be the Büchi automaton pictured in Figure 9 over the alphabet $\{a, b\}$. This automaton recognizes the set a^ω. It is unambiguous but not complete. The infinite word a^ω is the only word labeling a final path in $\mathcal{A}$. The automaton $\hat{\mathcal{A}}$ is shown in Figure 10. Since all states of $\mathcal{A}$ are final, $h_x(q)$ is either equal to 1 or to ∞ (see Example 29).

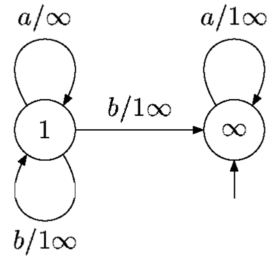

Figure 10: Automaton $\hat{\mathcal{A}}$ of Example 53

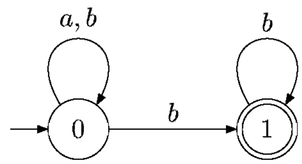

Figure 11: Automaton $\mathcal{A}$ of Example 54

Example 54 Let $\mathcal{A}$ be the automaton pictured in Figure 11 over the alphabet $\{a, b\}$. This automaton recognizes the set $A^* b^\omega$ of infinite words having finitely many a. It is neither unambiguous or complete. The infinite word b^ω labels several final paths whereas the infinite word a^ω labels no final path in $\mathcal{A}$. The automaton $\hat{\mathcal{A}}$ is shown in Figure 12. A state of $\hat{\mathcal{A}}$ is a pair $(h, \prec)$ where h is a function from $\{1, 2\}$ to $\overline{\mathbb{N}}_1$. Each state is labeled by the values $h(1), h(2)$. If $h(1) = h(2)$, we write $\sim$, $\prec$ or $\succ$ in between $h(1)$ and $h(2)$ to describe the preorder $\prec$.

The link between the automata $\mathcal{A}$ and $\hat{\mathcal{A}}$ is the following. For any infinite word x, the graph $G_{\mathcal{A}}(x)$, is bounded by the number m of states of $\mathcal{A}$. By Lemma 40, the function h associated with $G_{\mathcal{A}}(x)$ by definition 27 is actually a function from V to $\overline{\mathbb{N}}_{2m-1}$. Thus, the sequence of pairs $(h_n, \prec_n)$ defined by the function h associated with $G_{\mathcal{A}}(x)$ and its preorder $\prec_h$, gives a sequence of states of the automaton $\hat{\mathcal{A}}$.

Theorem 55 *For any Büchi automaton $\mathcal{A}$, the automaton $\hat{\mathcal{A}}$ recognizes the same set as $\mathcal{A}$ and it is unambiguous and complete. For any infinite word x, the final path of $\hat{\mathcal{A}}$ labeled by x is the path*

$$\gamma_x = (h_{x,0}, \prec_{x,0})(h_{x,1}, \prec_{x,1})(h_{x,2}, \prec_{x,2}) \dots$$

Proof Let $x = a_0 a_1 a_2 \dots$ be an infinite word. We first prove that γ_x is actually a final path of $\hat{\mathcal{A}}$ labeled by x. Since the pair $(h_x, \prec_x)$ satisfies $(h_x, \prec_x) \to (h_x, \prec_x)$, it is clear that for any integer n

$$\tau_{x,n} = (h_{x,n}, \prec_{x,n}) \xrightarrow{a_n} (h_{x,n+1} \prec_{x,n+1}).$$

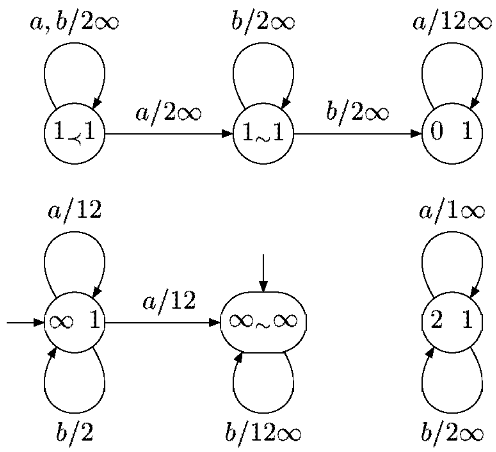

Figure 12: Automaton $\hat{\mathcal{A}}$ of Example 54

is a transition of $\hat{\mathcal{A}}$. Thus, γ_x is a path in $\hat{\mathcal{A}}$. It remains to show that this path is really final. We claim that for any $k \in \overline{\mathbb{N}}^*_{2m-1}$, the transition $\tau_{x,n}$ belongs to F_k for an infinite number of integers n. Let k be a fixed element of $\overline{\mathbb{N}}^*_{2m-1}$. Let g be the function from $\mathbb{N}$ to $\mathbb{N}$ defined by $g(n) = \max\{\hat{h}_{x,n}(q) \mid h_{x,n}(q) = k\}$ if the set $h^{-1}_{x,n}(k)$ is nonempty and $g(n) = 1$ otherwise. By Lemma 50, the inequality $g(n) \leq g(n+1) + 1$ holds for any integer n. We claim that if $\tau_{x,n}$ does not belong to F_k, the equality $g(n) = g(n+1) + 1$ holds. This implies that $\tau_{x,n}$ must belong to F_k for an infinite number of integers n. Otherwise, the equality $g(n+1) = g(n) - 1$ holds for any integer greater than some fixed integer n_0 and g cannot be a function from $\mathbb{N}$ to $\mathbb{N}$. If $\tau_{x,n}$ does not belong to F_k, there exists a state q such that $\mathrm{e}(h_{x,n}, q, h_{x,n+1})$ does not hold and such that $\vec{q}$ is maximal with respect to $\prec_{x,n+1}$. Thus, $\hat{h}_{x,n+1}(\vec{q})$ is equal to $g(n+1)$ and by Lemma 50, $\hat{h}_{x,n}(q) = \hat{h}_{x,n+1}(\vec{q}) + 1$. We finally get $g(n) = g(n+1) + 1$.

It remains to prove that γ_x is actually the unique final path in $\hat{\mathcal{A}}$ labeled by x. For this purpose, we need the following technical lemma. For any element q of Q, we denote by $\mathrm{up}_\prec(q)$ the number of equivalence classes greater with respect to $\prec$ than the equivalence class of q. We have then the following result.

Lemma 56 *Let $\tau = (h', \prec') \xrightarrow{a} (h, \prec)$ be a transition of $\hat{\mathcal{A}}$. Let q be a state of Q such that $\mathrm{e}(h', q, h)$ does not hold and $k = h(q)$. We then have*

$$\mathrm{up}_{\prec'}(q) + \delta \leq \mathrm{up}_\prec(\vec{q})$$

where δ equals 1 if $k \in \mu(\tau)$ and 0 otherwise.

Proof Since $\mathrm{e}(h', q, h)$ does not hold, the relation $\mathrm{e}(h', q', h)$ does not hold for any state q' such that $q \prec' q'$. Let $m = \mathrm{up}_{\prec'}(q)$. There exists then m states

$q_1, \ldots, q_m$ such that $q \prec' q_1$, $q_i \prec' q_{i+1}$ but $q_{i+1} \not\prec' q_i$. It is straightforward to verify that $\vec{q} \prec \vec{q}_1$, $\vec{q}_i \prec \vec{q}_{i+1}$ but $\vec{q}_{i+1} \not\prec \vec{q}_i$. Thus, the inequality $\mathrm{up}_{\prec'}(q) \leq \mathrm{up}_{\prec}(\vec{q})$ holds. Furthermore, if τ belongs to F_k, then $\vec{q}_i$ is not the maximal class of $h^{-1}(k)$ for any i. There is then one more class greater than the class of $\vec{q}$. Finally, the inequality $\mathrm{up}_{\prec'}(q) + \delta \leq \mathrm{up}_{\prec}(\vec{q})$ holds. □

We now prove that γ_x is the unique final path in $\hat{\mathcal{A}}$ labeled by x. Let $\gamma = (h_0, \prec_0)(h_1, \prec_1)(h_2, \prec_2)\ldots$ be a final path labeled by x. Let h be the function from V to $\overline{\mathbb{N}}$ defined by $h(q,n) = h_n(q)$. We claim that $h = h_x$. Since $h_n \to h_{n+1}$ for any integer n, the function h satisfies $h \to h$. By Theorem 34, it is sufficient to prove that for any vertex $v = (q,n)$ of $G_{\mathcal{A}}(x)$, there exists an integer p such that $\mathrm{e}^p(h,v,h)$ holds. Let k be $h_n(q)$. We define a sequence $(c_n)_{n\geq 0}$ of classes of states in the following way. The class c_0 is the class of q with respect to $\prec_n$ and $c_{l+1} = \vec{c}_l$ with respect to the transition $(h_{n+l}, \prec_{n+l}) \xrightarrow{a_{n+l}} (h_{n+l+1}, \prec_{n+l+1})$. By Lemma 56, the integer $\mathrm{up}_{\prec_{n+l}}(c_l)$ increases infinitely often with l unless there exists an integer p such that $\mathrm{e}(h, c_p, h)$. Thus, We have $\mathrm{e}^p(h,q,h)$ and $h = h_x$. It is also straightforward to show that $q \prec_n q'$ iff $q \prec_{x,n} q'$. We finally have proved that $\gamma = \gamma_x$. □

6.2.4 Number of states of $\hat{\mathcal{A}}$

In this section, we study the size of the automaton $\hat{\mathcal{A}}$. Since this automaton is codeterministic, the number of transitions is equal to the number of states multiplied by the cardinality of the alphabet A. It is then sufficient to upper bound the number of states. We have the following proposition.

Proposition 57 *If the automaton $\mathcal{A}$ has m states, the automaton $\hat{\mathcal{A}}$ has at most $(3m)^m$ states.*

The proof of the proposition is based on the following lemma.

Lemma 58 *Let Q be a finite set with m elements. The number of pairs $(f, \prec)$ where f is a function from Q to $\mathbb{N}_n$ and $\prec$ is a preorder on Q compatible with f is less than $(m+n)^m$.*

Proof We define a one to one mapping which maps any pair $(f, \prec)$ to a function $g_{f,\prec}$ from Q to $\mathbb{N}_{m+n-1}$. We denote by $<$ the preorder on Q defined by

$$q < q' \overset{\text{def}}{\iff} f(q) < f(q') \text{ or } (f(q) = f(q') \text{ and } q \prec q')$$

It is straightforward to check that the relation $<$ is a total preorder on Q. This preorder induces then a total order on its equivalence classes. For $q \in Q$, we denote by $\mathrm{do}_<(q)$ the number of equivalence classes of $<$ which are strictly smaller than the class of q. Since every equivalence class contains at least an element, the integer $\mathrm{do}_<(q)$ satisfies $\mathrm{do}_<(q) \leq m-1$. The function $g_{f,\prec}$ is then defined by

$$g_{f,\prec}(q) = f(q) + \mathrm{do}_<(q)$$

Since $\text{do}_<(q) \leq m-1$, the function $g_{f,\prec}$ is really a function from Q to $\mathbb{N}_{m+n-1}$. We claim that the mapping which maps $(f, \prec)$ to $g_{f,\prec}$ is one to one. It can be verified that

$$f(q) = |\{k < g_{f,\prec}(q) \mid g_{f,\prec}^{-1}(k) = \varnothing\}|$$
$$q \prec q' \iff f(q) = f(q') \text{ and } g_{f,\prec}(q) \leq g_{f,\prec}(q')$$

Thus, the function $g_{f,\prec}$ determines the pair $(f, \prec)$. Since the number of functions from Q to $\mathbb{N}_{n+m-1}$ is $(m+n)^m$, the number of pairs $(f, \prec)$ is less than $(m+n)^m$. □

The states of the automaton $\hat{\mathcal{A}}$ are pairs $(f, \prec)$ where f is a function from Q to $\overline{\mathbb{N}}_{2m-1}$ and $\prec$ is a preorder compatible with f. The number of such states is equal to the number of pairs $(f, \prec)$ where f is a function from Q to $\mathbb{N}_{2m}$ and $\prec$ is a preorder compatible with f. By the previous lemma, this number is less than $(3m)^m$.

The acceptance condition of the generalized Büchi automaton $\hat{\mathcal{A}}$ contains $2m$ subsets of transitions. By Proposition 15, this generalized Büchi automaton can be transformed into a CUBA with at most $(3m)^m 2^{2m} = (12m)^m$ states.

6.3 Proof with semigroups

In this section, we give a second proof of Theorem 7 using semigroups. No deep result on semigroups is involved but it is assumed that the reader is familiar with usual definitions like Green's relations. From a semigroup S recognizing a set X of infinite words, we construct a CUGBA $\mathcal{A}_S$ recognizing X. The acceptance condition of this CUGBA contains one subset of transitions. By Proposition 15, this CUGBA $\mathcal{A}_S$ can be easily transformed into a CUBA recognizing X.

This proof is organized as follows. In section 6.3.1, is recalled how semigroups can be used to recognize sets of infinite words. The chains expansion of a semigroup is detailed in Section 6.3.2. Finally, the construction of the automaton $\mathcal{A}_S$ is described in Section 6.3.3

6.3.1 Recognition of sets of infinite words by semigroups

In this section, we recall some needed material on the connections between rational sets of infinite words and semigroups. For further details, see [PP95]. We assume some familiarity with the basic notions of semigroup theory. We use the notation of [How95] for all undefined notions in semigroup theory.

The following theorem is special case of Ramsey's theorem.

Theorem 59 (Ramsey) *Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. For every infinite word $x \in A^\omega$ there is a pair (s, e) of elements of S such that $s = se$ and $e^2 = e$, and a factorization $x = x_0 x_1 x_2 \ldots$ of finite words such that $\varphi(x_0) = s$ and $\varphi(x_n) = e$ for $n > 0$.*

This motivates the following definition. A *linked pair* of a finite semigroup S is a pair (s, e) of elements of S such that $s = se$ and $e^2 = e$. A factorization $x = x_0x_1x_2\dots$ such that $\varphi(x_0) = s$ and $\varphi(x_n) = e$ for $n > 0$ for a linked pair (s, e) is called a *ramseyan factorization* with respect to the morphism φ and the linked pair is said to be associated with the factorization. For a given infinite word x, there may be several linked pairs (s, e) associated with different factorizations. Two linked pairs (s_1, e_1) and (s_2, e_2) are *conjugate* if there exists $x_1, x_2 \in S^1$ such that $s_1x_1 = s_2$, $x_1x_2 = e_1$ and $x_2x_1 = e_2$. The definition also implies $s_2x_2 = s_1x_1x_2 = s_1$ and the relation is therefore symmetrical. We have the following result (see [PP95]).

Proposition 60 *Conjugacy is an equivalence relation among linked pairs of a semigroup S. Moreover, if $\varphi : A^+ \to S$ is a semigroup morphism from A^+ onto S, two linked pairs (s_1, e_1) and (s_2, e_2) are conjugate iff*

$$\varphi^{-1}(s_1)\varphi^{-1}(e_1)^\omega \cap \varphi^{-1}(s_2)\varphi^{-1}(e_2)^\omega \neq \varnothing$$

In the sequel, the conjugacy class of the linked pair (s, e) is denoted by $[s, e]$ and the set of conjugacy classes of linked pairs of the semigroup S is denoted by $\widetilde{S}$. Any morphism $\varphi : A^+ \to S$ from A^+ into a finite semigroup S can be extended to a map from A^ω to $\widetilde{S}$. This extension is also denoted by φ and is defined by $\varphi(x) = [s, e]$ where (s, e) is any linked pair associated with a ramseyan factorization of x with respect to φ. This definition is consistent since if two linked pairs (s_1, e_1) and (s_2, e_2) are associated with two factorizations of x, there are conjugate and the equality $[s_1, e_1] = [s_2, e_2]$ therefore holds.

The definition of conjugacy implies that two conjugate linked pairs (s_1, e_1) and (s_2, e_2) satisfy $s_1 \mathcal{R}\, s_2$ and $e_1 \mathcal{D}\, e_2$. This is then consistent to define the $\mathcal{R}$-class and the $\mathcal{D}$-class of $[s, e]$ as the $\mathcal{R}$-class of s and the $\mathcal{D}$-class of e.

The following two examples present two semigroups S_1 and S_2 and give their respective linked pairs. We will use later these two semigroups to illustrate several notions.

Example 61 Let S_1 be the two-elements semigroup $\{e, f\}$ where the product is given by the equation $xy = x$ for $x, y \in S_1$. The semigroup S_1 is said to be left-zero and is often called $B(2, 1)$ in the literature. Any pair of this semigroup is linked. It may be easily verified that $[e, f] = [e, e]$ and $[f, e] = [f, f]$ and that the two pairs (e, e) and (f, f) are not conjugate. Thus, the set $\widetilde{S_1}$ is equal to $\{[e, e], [f, f]\}$.

Example 62 Let S_2 be the two-elements semigroup $\{0, 1\}$ with the usual multiplication of integers. This semigroup is usually called U_1 in the literature. The linked pairs of this semigroup are $(1, 1)$, $(0, 1)$ and $(0, 0)$. No linked pair is conjugate with another one. Thus, the set $\widetilde{S_2}$ is equal to $\{[1, 1], [0, 1], [0, 0]\}$.

The following theorem states that any rational set of infinite words is recognized by a morphism into a finite semigroup.

Theorem 63 *Let $X \subset A^\omega$ be a rational set of infinite words. There exists a morphism $\varphi : A^+ \to S$ from A^+ into a finite semigroup S and a subset $P \subset \widetilde{S}$ such that $X = \varphi^{-1}(P)$.*

For a proof, see [PP95]. There is natural left action of A^* on $\widetilde{S}$ defined for any $a \in A$ and $[s, e] \in \widetilde{S}$ by

$$a \cdot [s, e] = [\varphi(a)s, e]$$

An easy verification shows that if the linked pairs (s_1, e_1) and (s_2, e_2) are conjugate, then the linked pairs (ts_1, e_1) and (ts_2, e_2) are also conjugate for any $t \in S$. The left action is thus well defined. Furthermore, this left action is compatible with the concept of extension of morphism in the following sense.

Proposition 64 *Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. For any word u and any infinite word x,*

$$u \cdot \varphi(x) = \varphi(ux)$$

We now examine more closely the linked pairs of a $\mathcal{D}$-class. We have the following lemma which characterizes the conjugacy relation for the linked pairs of a $\mathcal{D}$-class.

Lemma 65 *Let D be a $\mathcal{D}$-class of a finite semigroup S and let (s_1, e_1) and (s_2, e_2) be two linked pairs such that $s_1, e_1, s_2, e_2 \in D$. The linked pairs (s_1, e_1) and (s_2, e_2) are conjugate iff $s_1 \mathcal{R} s_2$.*

Proof We already have pointed out that the definition of conjugacy implies that $s_1 \mathcal{R} s_2$. Suppose conversely that $s_1 \mathcal{R} s_2$. Since the idempotent e_1 belongs to D, this $\mathcal{D}$-class is regular and the $\mathcal{R}$-class of s_1 and s_2 contains an idempotent e. We claim that $[s_1, e_1] = [s_2, e_2] = [e, e]$. Since $s_1 e_1 = s_1$, the relation $s_1 \mathcal{L} e_1$ holds. Since s_1 belongs to $R_e \cap L_{e_1}$, there is an element $t_1 \in R_{e_1} \cap L_e$ such that $s_1 t_1 = e$ and $t_1 s_1 = e_1$. This proves the equality $[s_1, e_1] = [e, e]$. In the same way, we have $[s_2, e_2] = [e, e]$. □

The set of conjugacy classes of linked pairs of a $\mathcal{D}$-class D can be then identified with the set of $\mathcal{R}$-classes of D.

6.3.2 Chains expansion

In this section, we describe a construction which is close to the Rhodes expansion of a semigroup. For further details on semigroup theory, the reader is referred to [How95, Pin86]. Concerning the Rhodes expansion, see [Til74, Bir84]. We warn the reader that the Rhodes expansion is usually constructed with the $\mathcal{L}$-order when we construct it with the $\mathcal{R}$-order. This is due to the fact that infinite words are right-infinite.

A finite sequence $(s_1, \ldots, s_n)$ of elements of a semigroup S such is called an *$\mathcal{R}$-chain* if it is decreasing for the $\mathcal{R}$-order, i.e. $s_1 \geq_{\mathcal{R}} \cdots \geq_{\mathcal{R}} s_n$. An $\mathcal{R}$-chain is said to be *strict* if it is strictly decreasing for the $\mathcal{R}$-order, i.e. $s_1 >_{\mathcal{R}} \cdots >_{\mathcal{R}} s_n$. The set of strict $\mathcal{R}$-chains of a semigroup S is denoted by $\widehat{S}$. If the semigroup is finite, this set is finite since the length of a strict $\mathcal{R}$-chain is less than the cardinality of the semigroup. In the sequel, we write chain instead of $\mathcal{R}$-chain.

Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. Our aim is to associate a strict chain of $\widehat{S}$ with any infinite word. We start by associating a strict chain of $\widehat{S}$ with any finite word and we then extend this concept to infinite words. Let $w = a_1 \dots a_n$ be a word of A^+ and let $w_i = a_1 \dots a_i$ be its prefix of length i for $1 \leq i \leq n$. Let $s_i = \varphi(w_i)$ be the image of the prefix w_i by φ. The sequence $(s_1, \dots , s_n)$ is then a chain of S. However, this chain might not be strict. In order to associate with w a strict chain, we define a reduction ρ which maps any chain to a strict chain obtained by suppressing some elements. This reduction is inductively defined as follows

$$\rho(s_1) = (s_1)$$

$$\rho(s_1, \dots , s_n) = \begin{cases} \rho(s_1, \dots , s_{n-1}) & \text{if } s_n \mathrel{\mathcal{R}} s_{n-1} \\ (\rho(s_1, \dots , s_{n-1}), s_n) & \text{if } s_n >_{\mathcal{R}} s_{n-1} \end{cases}$$

It may be easily checked that $\rho(s_1, \dots , s_n)$ is a strict chain and that this chain is in fact a subsequence of the sequence $(s_1, \dots , s_n)$. There are integers $i_0, \dots , i_k$ such that $\rho(s_1, \dots , s_n) = (s_{i_0}, \dots , s_{i_k})$. Furthermore, the integer i_0 is equal to 1 whereas the integers $i_1, \dots , i_k$ are the positions where the chain $(s_1, \dots , s_n)$ is strictly decreasing. Indeed, it can be checked that the set $\{i_1, \dots , i_k\}$ is exactly the set of integers i such that $s_{i-1} >_{\mathcal{R}} s_i$. The reduction ρ removes from the chain any element which is $\mathcal{R}$-equivalent to its left neighbour. The first element of the chains always remains in the reduction whereas the last element may disappear if it is $\mathcal{R}$-equivalent to the last but one. Our reduction differs from the one usually used for the Rhodes expansion of a semigroup. The usual Rhodes expansion removes any element which is $\mathcal{R}$-equivalent to its right neighbour and the result would be $(s_{i_1-1}, \dots , s_{i_k-1}, s_n)$.

With any finite word $w = a_1 \dots a_n$, we associate the strict chain $\widehat{\varphi}(w) = \rho(s_1, \dots , s_n)$ where $s_i = \varphi(a_1 \dots a_i)$ for $1 \leq i \leq n$. We point out that the map $\widehat{\varphi}$ is not a morphism. The reduction ρ is illustrated in the following example.

Example 66 In Figure 13 is pictured an $\mathcal{R}$-chain of length 8. The set I is equal to $\{3, 4, 7\}$ and the reduced chain $\rho(s_1, \dots , s_8)$ is (s_1, s_3, s_4, s_7). These elements are marked in black in the figure.

The following two examples give the set of strict chains of the semigroups S_1 and S_2 already presented in Examples 61 and 62.

Example 67 The semigroup S_1 of Example 61 has only one $\mathcal{L}$-class which is $\mathcal{R}$-trivial. The only $\mathcal{R}$-chains are the sequences on length 1. Thus, the set $\widehat{S}_1$ is equal to $\{(e), (f)\}$.

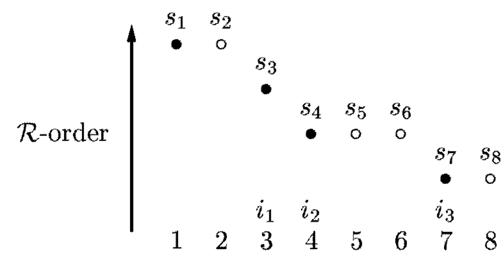

Figure 13: An $\mathcal{R}$-chain of length 8

Example 68 The strict $\mathcal{R}$-chains of the semigroup S_2 of Example 62 are the sequences (1), (0) and $(1, 0)$.

We define a left action of A^* on $\widehat{S}$ by

$$a \cdot (s_1, \dots, s_n) = \rho(\varphi(a), \varphi(a)s_1, \dots, \varphi(a)s_n)$$

The map $\widehat{\varphi}$ is not a morphism but the left action of A^* on $\widehat{S}$ is compatible with $\widehat{\varphi}$ in the following sense.

Proposition 69 *Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. For any words u and w of A^+, we have*

$$u \cdot \widehat{\varphi}(w) = \widehat{\varphi}(uw).$$

Proof We only prove the result when the word u is a single letter. The whole result follows by induction on the length of u. Assume the word u is the letter a. Let $w = a_1 \dots a_n$ and let $w_i = a_1 \dots a_i$ be its prefix of length i for $1 \leq i \leq n$. Let $s_i = \varphi(w_i)$ be the image of the prefix w_i by φ. We note $s'_0 = \varphi(a)$ and $s'_i = \varphi(a)s_i$ for $1 \leq i \leq n$. The elements $(s'_0, \dots, s'_n)$ are thus the images of the prefixes of the word aw. Assume $\widehat{\varphi}(w) = (s_{i_0}, \dots, s_{i_k})$ where $i_0 = 1$. Since $s_j \mathrel{\mathcal{R}} s_{i_k}$ for $i_k \leq j \leq i_{k+1} - 1$, we have $s'_j \mathrel{\mathcal{R}} s'_{i_k}$ for $i_k \leq j \leq i_{k+1} - 1$. We then have

$$\begin{aligned} a \cdot \widehat{\varphi}(w) &= \rho(s'_0, s'_{i_0}, \dots, s'_{i_k}) \\ &= \rho(s'_0, s'_1, \dots, s'_n) \\ &= \widehat{\varphi}(aw) \end{aligned}$$

□

We now distinguish some of the transitions $a \triangleright (s_1, \dots, s_n)$ defined by this left action. These transitions will play an important role in the definition of the automaton $\mathcal{A}_S$. We note $s'_0 = \varphi(a)$ and $s'_i = \varphi(a)s_i$ for $1 \leq i \leq n$. The transition $a \triangleright (s_1, \dots, s_n)$ is said to be *cutting* iff $s'_{n-1} \mathrel{\mathcal{R}} s'_n$. In this case, the strict chain $\rho(s'_0, \dots, s'_n)$ is, by definition of ρ, equal to $\rho(s'_0, \dots, s'_{n-1})$ and the last element s'_n of the chain is cut off. The following proposition states the key property of cutting transitions.

Proposition 70 *Let $w \in A^+$ be a finite word and let $(s_1, \dots, s_n)$ be any strict chain of $\widehat{S}$. If the path $w \triangleright (s_1, \dots, s_n)$ contains more than n cutting transitions, the strict chain $w \cdot (s_1, \dots, s_n)$ is then equal to $\widehat{\varphi}(w)$.*

Proof We claim that, for any word u, if the chain $u \cdot (s_1, \dots, s_n)$ is equal to $(t_1, \dots, t_m)$, there is an integer $1 \le k \le m$ such that the chain $(t_1, \dots, t_k)$ is equal to $\widehat{\varphi}(u)$. Furthermore, if this integer k is strictly smaller than m, the difference $m - k$ is bounded by the difference $n - c$ where c is the number of cutting transitions in the path $u \triangleright (s_1, \dots, s_n)$. If we prove this assumption, we prove the result. Indeed, when the number c of cutting transitions is greater than n, the integer k must be equal to m and $\widehat{\varphi}(w)$ is equal to $w \cdot (s_1, \dots, s_n)$.

We prove the assumption by induction on the length of u. Let a be a letter. Assume that the path $u \triangleright (s_1, \dots, s_n)$ contains c cutting transitions and that $u \cdot (s_1, \dots, s_n)$ is equal to $(t_1, \dots, t_m)$. Assume also that $\widehat{\varphi}(u)$ is equal to $(t_1, \dots, t_k)$ where the integer k is such that $k = m$ or $m - k$ is bounded by $n - c$.

If $k = m$, the chain $(t_1, \dots, t_m)$ equals $\widehat{\varphi}(u)$ and by Proposition 69, the chain $a \cdot (t_1, \dots, t_m) = au \cdot (s_1, \dots, s_n)$ equals $\widehat{\varphi}(au)$.

We now suppose that $k < m$. We note $t'_0 = \varphi(a)$ and $t'_i = \varphi(a)t_i$ for $1 \le i \le m$. By definition of the left action, we have

$$a \cdot (t_1, \dots, t_m) = \rho(t'_0, t'_1, \dots, t'_k, \dots, t'_m).$$

Since $\widehat{\varphi}(u) = (t_1, \dots, t_k)$, we have $\widehat{\varphi}(au) = \rho(t'_0, t'_1, \dots, t'_k)$ by Proposition 69. Assume that $\rho(t'_0, t'_1, \dots, t'_k)$ is equal to the chain $(t'_{i_1}, \dots, t'_{i_{k'}})$ where $i_1 = 0$. The first k' elements of the chain $\rho(t'_0, t'_1, \dots, t'_k, \dots, t'_m)$ are then $(t'_{i_1}, \dots, t'_{i_{k'}})$. We can then write $\rho(t'_0, t'_1, \dots, t'_k, \dots, t'_m) = (t'_{i_1}, \dots, t'_{i_{k'}}, \dots, t'_{i_{m'}})$. The other elements $(t'_{i_{k'}+1}, \dots, t'_{i_{m'}})$ come from the reduction of the chain $(t'_{k+1}, \dots, t'_m)$ This ensures that $m' - k' \le m - k$. If the transition $a \triangleright (t_1, \dots, t_m)$ is not cutting, the number c' of cutting transitions in the whole path $au \triangleright (s_1, \dots, s_n)$ is equal to c and we have $m' - k' \le n - c'$. If this transition is cutting, we have $c' = c + 1$. In this case, we have $t'_{m-1} \mathcal{R}\, t'_m$ and the last element t'_m is cut off by the reduction. This ensures that $m' - k' \le m - k - 1 \le n - c'$. □

We already have defined the strict sequence $\widehat{\varphi}(w)$ for a finite word $w \in A^+$. We now explain how this concept can be extended to infinite words. Let $x = a_1a_2a_3 \dots$ be an infinite word and let $x_n = a_1 \dots a_n$ be its prefix of length n. We define the sequence s_i of S by $s_i = \varphi(x_i)$. Since the semigroup S is finite the sequence s_i can not decrease infinitely often for the $\mathcal{R}$-order. There is then an integer n_0 such that $s_i \mathcal{R}\, s_j$ for $i, j \ge n_0$. We have then $\widehat{\varphi}(x_i) = \widehat{\varphi}(x_j)$ for $i, j \ge n_0$. We define $\widehat{\varphi}(x)$ as $\widehat{\varphi}(x_i)$ for any $i \ge n_0$.

The result of Proposition 69 can be extended to infinite words.

Proposition 71 *Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. For any word u of A^+ and any infinite word x of A^ω, we have*

$$u \cdot \widehat{\varphi}(x) = \widehat{\varphi}(ux).$$

Proof It suffices to apply Proposition 69 to a long enough prefix of x. □

6.3.3 Automaton $\mathcal{A}_S$

In this section, we describe a CUGBA recognizing a set X. Let X be a set of infinite words and let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S such that $X = \varphi^{-1}(P)$ where P is a subset of $\widetilde{S}$.

We now construct the GBA $\mathcal{A}_S = (Q, A, E, I, \mathcal{F})$. The set Q of states is a subset of $\widetilde{S} \times \widehat{S}$ defined by

$$Q = \left\{ \left([s,e],(s_1,\dots,s_n)\right) \middle| s \mathrel{\mathcal{R}} s_n \right\}.$$

The set I of initial states is defined by

$$I = \left\{ \left([s,e],(s_1,\dots,s_n)\right) \in Q \middle| [s,e] \in P \right\}.$$

The set E of transitions of the automaton $\mathcal{A}$ is naturally defined by the actions of A^* on $\widetilde{S}$ and $\widehat{S}$.

$$E = \left\{ \left(a \cdot [s,e], a \cdot (s_1,\dots,s_n)\right) \xrightarrow{a} \left([s,e],(s_1,\dots,s_n)\right) \right\}.$$

The automaton $\mathcal{A}$ is then codeterministic and complete. The family $\mathcal{F}$ actually contains one set F of final transitions. This set F is the set of transitions such that $\varphi(a)s \mathrel{\mathcal{D}} e$ and such that

$$a \cdot (s_1,\dots,s_n) \xrightarrow{a} (s_1,\dots,s_n)$$

is cutting. We illustrate the construction with the two following examples.

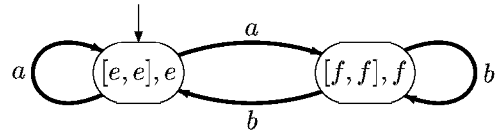

Figure 14: Automaton $\mathcal{A}_{S_1}$ of Example 72

Example 72 Let $A = \{a, b\}$ be the alphabet. The set $X = aA^\omega$ of infinite words beginning with an a is recognized by the morphism $\varphi : A^+ \to S_1$ defined by $\varphi(a) = e$ and $\varphi(b) = f$. See Example 61 for the linked pairs of S_1 and see also Example 67 for the $\mathcal{R}$-chains of this semigroup. The automaton $\mathcal{A}_{S_1}$ is pictured in Figure 14. All its transitions are final.

Example 73 Let A be the alphabet $\{a, b\}$. The set $X = (A^*b)^\omega$ of infinite words having an infinite number of b is recognized by the morphism $\varphi : A^+ \to S_2$ defined by $\varphi(a) = 1$ and $\varphi(b) = 0$. See Example 62 for the linked pairs of S_2 and see also Example 68 for the $\mathcal{R}$-chains of this semigroup. The automaton $\mathcal{A}_{S_2}$ is pictured in Figure 15. The final transitions are drawn with a wider line.

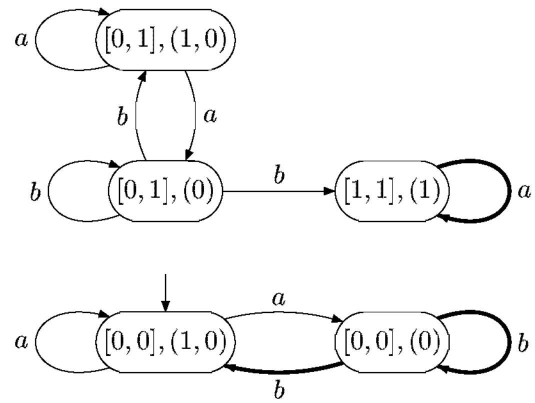

Figure 15: Automaton $\mathcal{A}_{S_2}$ of Example 73

We now prove that the automaton $\mathcal{A}_S$ we have defined is really a CUGBA recognizing the set X. We first prove that any infinite words labels at least one final path in $\mathcal{A}_S$. Then we prove that any infinite word labels at most one final path in $\mathcal{A}_S$.

Let $x = a_0a_1a_2\ldots$ be an infinite word. Let $x_i = a_ia_{i+1}a_{i+2}$ be the suffix of x starting at a_i. Let q_i be the pair $\big(\varphi(x_i), \widehat{\varphi}(x_i)\big)$. Propositions 64 and 71 ensure that the path γ

$$\gamma : q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots$$

is a path in the automaton $\mathcal{A}_S$.

Lemma 74 *The path γ is final.*

Proof Let $x = u_0u_1u_2\ldots$ be a ramseyan factorization associated with a linked pair (s, e). Then for each $i > 0$, $\varphi(u_iu_{i+1}\ldots) = [e, e]$. Fix some $i \geq 2$ and let n_i be the length of the word $u_0u_1\ldots u_i$. The state q_{n_i} is equal to $\big([s, e], (s_1, \ldots, s_n)\big)$ where $(s_1, \ldots, s_n) = \widehat{\varphi}(u_{i+1}u_{i+2}\ldots)$. In particular, $s_n \mathrel{\mathcal{R}} e$ and hence $es_n = s_n$. Suppose first that $n \geq 2$. Then $\varphi(u_i)s_{n-1} = es_{n-1}$ and the relation $\varphi(u_i)s_{n-1} >_{\mathcal{R}} \varphi(u_i)s_n$ does not hold. If $n = 1$, the same argument works by replacing s_{n-1} by 1. It follows that the path of label u_i from $q_{n_{i-1}}$ to q_{n_i} contains a cutting transition. Furthermore, for any j greater than the length of u_0, the state q_j is a state $\big([s', e'], (s'_1, \ldots, s'_m)\big)$ with $s'_m \mathrel{\mathcal{R}} s_n$ and $e' \mathrel{\mathcal{D}} e$. Therefore, the path from $q_{n_{i-1}}$ to q_{n_i} contains a final transition. This proves that the path γ is final. □

We now claim that the final path labeled by an infinite word is unique. The following lemma states that the starting state of a final path labeled by a given word x is completely determined by x. By Proposition 4, this proves that the automaton is unambiguous. By definition of the set I of initial states, it also proves that it recognizes the set X

Lemma 75 *If the infinite word x is the label of final path in $\mathcal{A}_S$, this final path starts at state $\big(\varphi(x), \widehat{\varphi}(x)\big)$.*

Proof Let γ be a final path labeled by x. We prove that the starting state of this path γ must be the pair $\big(\varphi(x), \widehat{\varphi}(x)\big)$. Since the path γ is final, some final transition occurs infinitely often along this path. The infinite word x can be then factorized $x = u_0u_1u_2\dots$ such that the first transition of the finite path labeled by u_i is a transition $q \xrightarrow{a} p$ which is final. Let q be the state $\big([s,e],(s_1,\dots,s_n)\big)$ and let x' be the suffix $u_1u_2u_3\dots$ of x. By Proposition 64 and 71, it suffices to prove that $\varphi(x') = [s,e]$ and $\widehat{\varphi}(x') = (s_1,\dots,s_n)$ in order to prove the result for x.

For i greater than n, the finite path $u_1\dots u_i \triangleright q$ contains more than n cutting transitions. By Proposition 70, this ensures that $\widehat{\varphi}(u_1\dots u_i) = (s_1,\dots,s_n)$. Since $\widehat{\varphi}(x')$ is equal to $\widehat{\varphi}(u)$ for any long enough prefix u of x', $\widehat{\varphi}(x')$ is also equal to $(s_1,\dots,s_n)$.

We now prove that $\varphi(x') = [s,e]$. By definition of the final transitions, we have $s \mathrel{\mathcal{D}} e$. Let D be the common $\mathcal{D}$-class of s, s_n and e. Suppose that $\varphi(x_1) = [s',e']$. Since $\widehat{\varphi}(x')$ equals $(s_1,\dots,s_n)$, we have $s' \mathrel{\mathcal{R}} s_n$ and $s' \mathrel{\mathcal{R}} s$. Since each path $u_i \triangleright q$ contains more than n cutting transitions, this ensures that $\widehat{\varphi}(u_i) = (s_1,\dots,s_n)$ and that $\varphi(u_i) \mathrel{\mathcal{R}} s_n$. The idempotent e' is then $\mathcal{J}$-below the $\mathcal{D}$-class D. Conversely, the infinite word x' cannot have a factor w such that $\varphi(w)$ is strictly $\mathcal{J}$-below the $\mathcal{D}$-class D since $\varphi(u_1\dots u_i)s = s$ and $s \in D$. This proves that $e' \in D$ and by Lemma 65, we have $[s,e] = [s',e']$. $\square$

By the two previous lemmas, the automaton $\mathcal{A}_S$ based on a semigroup S is unambiguous and complete and it recognizes the set X of infinite words. We now analyze how the size of the automaton depends on the size of the semigroup S. We then give an upper bound of the size of the automaton $\mathcal{A}_S$ and we show, by the study of particular cases, that this upper bound is tight.

Proposition 76 *Let n be the cardinality of a semigroup S. The number of states of the automaton $\mathcal{A}_S$ is bounded by $n^2 2^n$.*

Proof The number of linked pairs of S is clearly bounded by n^2. Thus, the cardinality of the set $\widetilde{S}$ is less than n^2. A chain $(s_1,\dots,s_k)$ of S is uniquely determined by the subset $\{s_1,\dots,s_k\}$ of S. The size of $\widehat{S}$ is then bounded by the number 2^n of subsets of S. Combining both results, we obtain that the number of states of the constructed automaton is bounded by $n^2 2^n$. $\square$

Actually, this upper bound could be slightly improved. It can be proved that the cardinality of the set $\widetilde{S}$ is bounded by $n(n+1)/2$. However, the exponential blow up cannot be avoided. By proposition 15, the automaton $\mathcal{A}_S$ can be transformed into a CUBA with $n^2 2^{n+1}$ states recognizing X.

The following family of examples shows that this upper bound is tight. For any integer n, there exists a semigroup S_n of size n such that the automaton $\mathcal{A}_{S_n}$ has at least $(n-1)2^n$ states.

Example 77 Let S_n be the semigroup $\{e_1, \dots, e_n\}$ with the product defined by $e_i e_j = e_{\max(i,j)}$. A pair (e_k, e_l) is linked iff $k \geq l$ and $[e_k, e_l] = [e_{k'}, e_{l'}]$ iff $k = k'$ and $l = l'$. A sequence $(e_{i_1}, \dots, e_{i_m})$ is a strict $\mathcal{R}$-chain iff $i_1 > \cdots > i_m$. The set of states of the automaton $\mathcal{A}_{S_n}$ is

$$Q_n = \left\{ \left([e_k, e_l], (e_{i_1}, \dots, e_{i_m})\right) \middle| k = i_m \text{ and } k \geq l \text{ and } i_1 > \cdots > i_m \right\}.$$

and its cardinality is $\sum_{k=1}^{n} k 2^{k-1} = (n-1)2^n + 1$. Indeed, for each integer k, the integer l can have k different values and there are 2^{k-1} different strict chains such that $i_m = k$.

Conclusion

As a conclusion, we discuss some applications of unambiguous Büchi automata and we mention some open problems. This paper deals with right-infinite but the main result also applies to left-infinite words. Two-sided infinite words, that is, words of $A^{\mathbb{Z}}$ can also be considered. Using simultaneously two unambiguous automata for right-infinite words and for left-infinite words, one gets an unambiguous automaton for tow-sided infinite words. This leads to a McNaughton's like theorem for two-sided infinite words. Since unambiguous Büchi automata are codeterministic, they naturally define right-sequential functions if their transitions are equipped with output. A well-known theorem states that any rational function of finite words is the composition of a left-sequential and a right-sequential function. Adapting the first proof of the main result, an analogous theorem can be proved for infinite words. Since this paper is already rather long, these two applications will be developed in forthcoming papers.

Propositions 4 and 8 give two characterizations of complete unambiguous Büchi automata. Furthermore, they shows that it can be effectively checked whether a given automaton is unambiguous and complete. However, it is not clear whether this test can be performed in polynomial time.

Acknowledgment

The authors would like to thank Dominique Perrin, Jean-Éric Pin and Pascal Weil for helpful discussions and suggestions.

References

[Arn83] André Arnold. Rational omega-languages are non-ambiguous. *Theoretical Computer Science*, 26(1-2):221–223, 1983.

[Bir84] Jean-Camille Birget. Iteration of expansions – unambiguous expansions. *J. Pure and Applied Algebra*, 34:1–55, 1984.

[BP85] Danièle Beauquier and Dominique Perrin. Codeterministic automata on infinite words. *Information Processing Letters*, 20:95–98, 1985.

[Büc62] J. Richard Büchi. On a decision method in the restricted second-order arithmetic. In *Proc. Int. Congress Logic, Methodology and Philosophy of science, Berkeley 1960*, pages 1–11. Stanford University Press, 1962.

[How95] John M. Howie. *Fundamentals of Semigroup Theory.* Oxford University Press, 1995.

[HU79] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, 1979.

[Lan69] Lawrence H. Landweber. Decision problems for ω-automata. *Math. Systems Theory*, 3:376–384, 1969.

[McN66] Robert McNaughton. Testing and generating infinite sequences by a finite automaton. *Inform. Control*, 9:521–530, 1966.

[Mic88] Max Michel. Complementation is more difficult with automata on infinite words. Manuscript, 1988.

[Mos82] A. W. Mostowski. Determinancy of sinking automata on infinite trees and inequalities between various Rabin's pair indices. *Information Processing Letters*, 15(4):159–163, 1982.

[Mos84] A. W. Mostowski. Regular expressions for infinite trees and a standard form for automata. In A. Skowron, editor, *Computation theory*, volume 208 of *Lect. Notes in Comput. Sci.*, pages 157–168. Springer-Verlag, Berlin, 1984.

[Mul63] David Muller. Infinite sequences and finite machines. In Proc. of Fourth Annual IEEE Symp., editor, *Switching Theory and Logical Design*, pages 3–16, 1963.

[Per90] Dominique Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 1. Elsevier, 1990.

[Pin86] Jean-Éric Pin. *Varieties of formal languages.* North Oxford, London and Plenum, New-York, 1986. (Traduction de *Variétés de langages formels*).

[PP95] Dominique Perrin and Jean-Éric Pin. Semigroups and automata on infinite words. In J. Fountain and V. A. R. Gould, editors, *NATO Advanced Study Institute Semigroups, Formal Languages and Groups*, pages 49–72. Kluwer academic publishers, 1995.

[Rab69] Micheal Ozer Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.

[SVW87] A. Prasad Sistla, Moshe Y. Vardi, and Pierre Wolper. The complementation problem for Büchi automata and applications to temporal logic. *Theoret. Comput. Sci.*, 49:217–237, 1987.

[Tho90] Wolfgang Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 4. Elsevier, 1990.

[Til74] B. Tilson. On the complexity of finite semigroups. *J. Pure and Applied Algebra*, 5:187–208, 1974.

[Var96] Moshe Y. Vardi. An automata-theoretic approach to linear temporal logic. In *Logics for Concurrency: Structure versus Automata*, volume 1043 of *Lect. Notes in Comput. Sci.*, pages 238–266. Springer-Verlag, 1996.