On Bifix Systems and Generalizations

Jan-Henrik Altenbernd

RWTH Aachen University

Abstract. Motivated by problems in infinite-state verification, we study word rewriting systems that extend mixed prefix/suffix rewriting (short: bifix rewriting). We introduce several types of infix rewriting where infix replacements are subject to the condition that they have to occur next to tag symbols within a given word. Bifix rewriting is covered by the case where tags occur only as end markers. We show results on the reachability relation (or: derivation relation) of such systems depending on the possibility of removing or adding tags. Where possible we strengthen decidability of the derivation relation to the condition that regularity of sets is preserved, resp. that the derivation relation is even rational. Finally, we compare our model to ground tree rewriting systems and exhibit some differences.

1 Introduction

The algorithmic theory of prefix (respectively suffix) rewriting systems on finite words has long been well established, and a number of decision problems over such systems have been proven to be decidable. Such rewriting systems are a general view of pushdown systems, where symbols are pushed onto and removed from the top of a stack.

Büchi showed in [2] that the language derivable from a given word by prefix rewriting is regular (and that a corresponding automaton can be computed). In the theory of infinite-state system verification, the "saturation method" (for the transformation of finite automata) has been applied for this purpose (see e. g. [14, 5, 6]). Caucal [4] showed the stronger result that the derivation relation induced by a prefix rewriting system is a rational relation.

The extension to combined prefix and suffix rewriting goes back to Büchi and Hosken [3]. Karhumäki, Kunc, and Okhotin showed in [9] that when combining prefix and suffix rewriting, the corresponding derivation relation is still rational, and therefore preserves regularity of languages. They extended their work in [8] to rewriting systems with a center marker, simulating two stacks communicating with each other. They singled out a number of cases where universal computation power could already be achieved with very limited communication.

In a more restricted framework, Bouajjani, Müller-Olm and Touili studied dynamic networks of pushdown systems in [1]. Here, a collection of pushdown processes is treated as a word in which a special marker is used to separate the processes. Rewriting of such words is restricted to performing pushdown operations and to creating new processes, where the latter increases the number of markers. It was shown that reachability in this setting is decidable.

In the present paper, we develop a generalised framework of "tagged infix rewriting" which extends some of the cases mentioned above. We clarify the status of the word-to-word reachability relation (or derivation relation) for several types of tagged infix rewriting. More precisely, we determine whether this relation is undecidable, or decidable, or even decidable in two stronger senses: that the relation preserves effectively the regularity of a language, or that the derivation relation itself is rational. (By "effective" preservation of regular languages we mean that from a presentation of L by a finite automaton and from the rewriting system defining the relation R we obtain algorithmically a finite automaton for the image of L under the derivation relation of R.) So the motivation (and contribution) of the paper is twofold: first to push the frontier of decidability further for reachability problems over rewriting systems, and secondly to differentiate clearly between the three levels of decidability proofs mentioned above.

We define a generalisation of mixed prefix/suffix rewriting systems on words by introducing special symbols (*tags* or *markers*) to mark positions in words where rewriting can occur. Typically, a rewriting rule can transform a word $w = w_0\#_1w_1\cdots\#_nw_n$ into a word $w' = w'_0\#_1w'_1\cdots\#_nw'_n$ with $w_i = w'_i$ for all i except for some i_0 where w'_{i_0} is obtained from w_{i_0} by a prefix, suffix, or complete rewriting rule $U \hookrightarrow V$ with regular sets U, V (to be applied to the whole word $u \in U$ between two successive markers, replacing it by some $v \in V$). Thus, arbitrary words in finite sequences can be rewritten independently, extending a case studied in [9]. The variants we consider in this paper deal with the options that markers may be removed or added in the rewriting process. We show that the derivation relation is rational in the basic case mentioned above, where markers are always preserved, and that this fails in general for the other cases. However, we still obtain decidability of the reachability problem in all cases. For applications, our systems are close to models of concurrent processes where states are presented by words between tags, state transitions by local rewriting rules, and e. g. spawning of new processes by the insertion of tags.

The paper is structured as follows: In the subsequent section we summarise technical preliminaries. Section 3 introduces the basic models of bifix systems and its extension tagged infix rewriting, and we show that one obtains different levels of decidability of the derivation relation: We present cases where the derivation relation is not rational but effectively preserves regularity of languages, and where the latter condition fails but the word-to-word reachability problem is still decidable. This refined analysis also exhibits a substantial difference between the two cases of tag insertion and tag removal. The next section is devoted to a comparison of bifix systems and ground tree rewriting systems (and the closely related multi-stack systems).

2 Terminology

Automata and Languages. We use the standard terminology from automata theory and formal language theory (see e. g. [7]). We present nondeterministic finite automata (NFA) in the format $\mathcal{A} = (Q, \Sigma, q_0, \Delta, F)$, where Q is a finite set of states, Σ is a finite alphabet, $q_0 \in Q$ is the initial state, $F \subseteq Q$ is the set of final states, and $\Delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ is a finite set of transitions. We write $\mathcal{A} : p \xrightarrow{w} q$ to denote that there is a w-labelled path from state p to state q in $\mathcal{A}$. $\mathrm{Reg}(\Sigma)$ denotes the class of all regular languages over Σ. We will refer to *normalised NFAs* which have exactly one final state, and in which no incoming respectively outgoing transitions are allowed for the initial respectively final state. A (finite) transducer is an NFA $\mathcal{A} = (Q, \Gamma, q_0, \Delta, F)$, where $\Gamma \subseteq \Sigma^* \times \Sigma^*$ is a finite set of pairs of words over a finite alphabet Σ.

Relations. Let Σ be a finite alphabet. A relation $R \subseteq \Sigma^* \times \Sigma^*$ is *recognisable* if it is a finite union of products of regular languages over Σ, that is, $R = \bigcup_{i=1}^{n} L_i \times M_i$ for some $n \in \mathbb{N}$ and regular L_i, M_i; when using R as a rewriting system, we write rules in the form $L_i \hookrightarrow M_i$. R is *rational* if it is recognisable by a transducer, i. e. an NFA with transitions labelled by finite subsets of $\Sigma^* \times \Sigma^*$. We then write $R \in \mathrm{Rat}(\Sigma^* \times \Sigma^*)$.

For relations $R, S \subseteq \Sigma^* \times \Sigma^*$, we call $\mathrm{Dom}(R) = \{u \mid \exists v : (u, v) \in R\}$ the *domain* of R, and $\mathrm{Im}(R) = \{v \mid \exists u : (u, v) \in R\}$ the *image* of R. For $L \subseteq \Sigma^*$, we call $R(L) = \{v \mid \exists u \in L : (u, v) \in R\}$ the set derivable from L according to R. We define the concatenation of R and S as $R \cdot S = \{(ux, vy) \mid (u, v) \in R \wedge (x, y) \in S\}$, which we also shorten to RS, if no ambiguity arises, and their composition as $R \circ S = \{(u, w) \mid \exists v : (u, v) \in R \wedge (v, w) \in S\}$.

We call $I = \{(w, w) \mid w \in \Sigma^*\}$ the identity relation on Σ^*. Note that I is rational, but not recognisable. When considering iteration, we have to distinguish two cases. Let $R^* = \bigcup_{n \geq 0} R^n$, where $R^0 = \{(\varepsilon, \varepsilon)\}$, and $R^{n+1} = R^n \cdot R$, and let $R^{\circledast} = \bigcup_{n \geq 0} R^{(n)}$, where $R^{(0)} = I$, and $R^{(n+1)} = R^{(n)} \circ R$.

We recall some basic results about rational relations: $\mathrm{Rat}(\Sigma^* \times \Sigma^*)$ is closed under union, concatenation and the concatenation iteration *. Furthermore, if R is a rational relation, then $R(L)$ is regular for regular L, hence $\mathrm{Dom}(R)$ and $\mathrm{Im}(R)$ are regular. Finally, if R is a rational relation, and S is a recognisable relation, then $R \cap S$ is rational.

Mixed Prefix / Suffix Rewriting Systems. A *mixed prefix/suffix rewriting system* is a tuple $\mathcal{R} = (\Sigma, R, S)$, where Σ is a finite alphabet, and $R, S \subseteq \mathrm{Reg}(\Sigma) \times \mathrm{Reg}(\Sigma)$ are recognisable relations of rewriting rules. We write $w \underset{\mathcal{R}}{\longrightarrow} w'$ if $(w, w') \in (RI \cup IS)$, i. e. R and S are used for prefix respectively suffix rewriting. We denote the derivation relation $\underset{\mathcal{R}}{\longrightarrow}^{\circledast} = (RI \cup IS)^{\circledast}$ by $\mathcal{R}^{\circledast}$.

Proposition 1 ([9]). *The derivation relation $\mathcal{R}^{\circledast}$ of a mixed prefix/suffix rewriting system $\mathcal{R}$ is rational.*

3 Bifix Rewriting Systems and Extensions

As a first and minor extension of mixed prefix/suffix rewriting systems, we introduce *bifix rewriting systems*, which will serve as a basis for further extensions. A bifix rewriting system is a tuple $\mathcal{R} = (\Sigma, R, S, T)$, with Σ, R, S as in the case of mixed prefix/suffix rewriting systems, and where $T \subseteq \mathrm{Reg}(\Sigma) \times \mathrm{Reg}(\Sigma)$ is also a recognisable relation. We write $w \underset{\mathcal{R}}{\longrightarrow} w'$ if $(w, w') \in (RI \cup IS \cup T)$, that is, R and S are used as before, and T is used to rewrite complete words. The other notions carry over.

As a first result, it is easy to see that Proposition 1 holds again:

Proposition 2. *The derivation relation of a bifix rewriting system is rational.*

Proof. We have to show that $W = (RI \cup IS \cup T)^{\circledast}$ is rational. For this, introduce $\# \notin \Sigma$, and consider $U = \#R \cup \#T\#$ and $V = S\#$. Then $\#W\# = (UI \cup IV)^{\circledast} \cap (\#\Sigma^*\# \times \#\Sigma^*\#)$, that is, we use rewriting of complete words with T for prefix rewriting, and we restrict the corresponding derivation relation to pairs of words with $\#$ at the beginning and end only. Since U, V, and $(\#\Sigma^*\# \times \#\Sigma^*\#)$ are recognisable, $(UI \cup IV)^{\circledast}$ is rational by Proposition 1, and it follows that $\{(\#u\#, \#v\#) \mid (u, v) \in W\}$ is rational. Removing the symbols $\#$ preserves this rationality, so W is rational.

3.1 Tagged Infix Rewriting Systems

Let Σ be a finite alphabet. We will use a finite set M of *tags* (or *markers*) with $M \cap \Sigma = \emptyset$ to mark positions in a finite word where rewriting can occur. Given Σ and M, let $P_{\Sigma,M} := M\Sigma^* \cup \Sigma^* M \cup M\Sigma^* M$ denote the set of all words over $\Sigma \cup M$ that contain at least one marker from M, but only at the beginning and/or end.

A *tagged infix rewriting system* (TIRS) is a structure $\mathcal{R} = (\Sigma, M, R)$ with disjoint finite alphabets Σ and M and a relation $R \subseteq P_{\Sigma,M} \times P_{\Sigma,M}$ which is a finite union of

$$\begin{array}{l} \textit{prefix rules} \text{ of the form } \#U \hookrightarrow \#V \text{ (denoting } \#U \times \#V\text{)}, \\ \textit{suffix rules} \text{ of the form } U\$ \hookrightarrow V\$\,, \text{ and} \\ \textit{bifix rules} \text{ of the form } \#U\$ \hookrightarrow \#V\$\,, \end{array} \qquad (1)$$

where $U, V \in \mathrm{Reg}(\Sigma)$ and $\#, \$ \in M$. Note that when using R to rewrite a word w over $\Sigma \cup M$, all tags in w are preserved, and none are added. We write $xuy \underset{\mathcal{R}}{\longrightarrow} xvy$ if $(u, v) \in R$ and $x, y \in (\Sigma \cup M)^*$, and we denote $\underset{\mathcal{R}}{\longrightarrow}^{\circledast}$ by $\mathcal{R}^{\circledast}$.

As a first example, consider $\mathcal{R} = (\{a, b, c\}, \{\#\}, R)$ with the following set R of rules:

	$\#\# \hookrightarrow \#acb\#$		(bifix rule),
	$\#a \hookrightarrow \#aa$	$\#a^+cb \hookrightarrow \#b$	(prefix rules),
	$b\# \hookrightarrow bb\#$	$acb^+\# \hookrightarrow a\#$	(suffix rules).

Then $\mathcal{R}^{\circledast}(\{\#\#\}) = \#a^+cb^+\# \ \cup\ \#a^*\# \ \cup\ \#b^*\#$.

As a second example, note that the infinite grid can be generated with the simple TIRS $(\{a, b\}, \{\#\}, \{\# \hookrightarrow a\#, \# \hookrightarrow \#b\})$, starting with marker $\#$:

$$\begin{array}{ccccccc}
\# & \to & \#b & \to & \#bb & \to & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \\
a\# & \to & a\#b & \to & a\#bb & \to & \cdots \\
\downarrow & & \downarrow & & \downarrow & & \\
aa\# & \to & aa\#b & \to & aa\#bb & \to & \cdots \\
\downarrow & & \downarrow & & \downarrow & &
\end{array}$$

Since the monadic second-order logic (MSO) of the infinite grid is undecidable (see e. g. [15]), we can immediately conclude the following.

Proposition 3. *The MSO theory of graphs generated by TIRSs is undecidable.*

It is well known that prefix (resp. suffix) and mixed prefix/suffix rewriting systems preserve regularity ([4, 9]), that is, given such a system $\mathcal{R}$ and a regular set L, the set derivable from L according to $\mathcal{R}$ is again regular. It has also been shown that the derivation relation $\mathcal{R}^{\circledast}$ of such systems is rational. We show in the following that these results carry over to tagged infix rewriting systems.

Theorem 4. *The derivation relations of TIRSs are rational.*

Proof. Let $\mathcal{R} = (\Sigma, M, R)$ be a TIRS. We construct an NFA $\mathcal{A}_{\mathcal{R}} = (Q, \Gamma, q_0, \Delta, \{q_f\})$ whose edges are labelled with rational relations (i. e. Γ is a finite set of rational relations), such that $L(\mathcal{A}_{\mathcal{R}}) = \mathcal{R}^{\circledast}$. Since we know that every finite concatenation of rational relations is again rational, every path in $\mathcal{A}_{\mathcal{R}}$ from q_0 to q_f is labelled with a rational relation.

It is important to note that markers are preserved in the derivation process. Thus, the derivation relation is a concatenation of derivation relations of rewriting that occurs before the first marker (see (i) below), after the last marker (ii), or between two markers (iii), which are basically mixed prefix/suffix rewriting derivations.

We can therefore construct $\mathcal{A}$ as follows: For $\#, \$ \in M$, let $R_{\#} = \{(u, v) \mid (u\#, v\#) \in R\}$, ${}_{\#}R = \{(u, v) \mid (\#u, \#v) \in R\}$, and ${}_{\#}R_{\$} = \{(u, v) \mid (\#u\$, \#v\$) \in R\}$. We choose $Q = \{q_0, q_f\} \cup \{s_m, t_m \mid m \in M\}$, that is, we take one source state s_m and one target state t_m for every marker m, and we set Δ to be the following set of edges labelled with relations:

$$\begin{aligned}
\Delta = \{(s_m, \{m\} \times \{m\}, t_m) \mid m \in M\} \;\cup\; \{(q_0, I, q_f)\} & \\
\cup \{(q_0, (IR_m)^{\circledast}, s_m) \mid m \in M\} & \qquad \text{(i)} \\
\cup \{(t_m, ({}_mRI)^{\circledast}, q_f) \mid m \in M\} & \qquad \text{(ii)} \\
\cup \{(t_m, ({}_mRI \cup IR_{m'} \cup {}_mR_{m'})^{\circledast}, s_{m'}) \mid m, m' \in M\} \;. & \qquad \text{(iii)}
\end{aligned}$$

We know that $\{m\} \times \{m\}$, $(IR_m)^{\circledast}$, $({}_mRI)^{\circledast}$, and I are rational, and by Proposition 2 the same holds for $({}_mRI \cup IR_{m'} \cup {}_mR_{m'})^{\circledast}$. ☐

We can immediately deduce that TIRSs effectively preserve regularity.

3.2 Extending TIRSs by Removing Tags

We consider an extension of TIRSs where removing tags is allowed, thereby breaking up the preservation of markers. We will see that in this case some effective reachability analysis is still possible.

A *TIRS with tag-removing rules* is a structure $\mathcal{R} = (\Sigma, M, R)$ with disjoint finite alphabets Σ and M as before and a relation $R \subseteq P_{\Sigma,M} \times (P_{\Sigma,M} \cup \Sigma^*)$ containing rules of the basic form (1) and also rules of the forms $\#U \hookrightarrow V$, $U\$ \hookrightarrow V$, $\#U\$ \hookrightarrow \#V$, $\#U\$ \hookrightarrow V\$$, and $\#U\$ \hookrightarrow V$, where $U, V \in \mathrm{Reg}(\Sigma)$ and $\#, \$ \in M$. We show that in this case the derivation relation is not rational in general, but that regularity is still preserved (the latter result involving a nontrivial saturation construction).

Proposition 5. *Derivation relations of TIRSs with tag-removing rules are not rational in general.*

Proof. Consider $\mathcal{R} = (\{a, b\}, \{\#\}, R)$, where R contains only the rules $\#a \hookrightarrow b$ and $b\# \hookrightarrow a$. Then $\mathrm{Dom}(\mathcal{R}^{\circledast} \cap (\#^*a\#^* \times \{a\})) = \{\#^n a \#^n \mid n \geq 0\}$ is not regular, and so $\mathcal{R}^{\circledast}$ is not rational. □

Before showing that such systems still preserve regularity, we need to introduce some more terminology. We call an NFA $\mathcal{A} = (Q, \Sigma \cup M, q_0, \Delta, F)$ *unravelled* if it satisfies the following conditions:

1. for every $q \in Q$: $|\{(q, m, p) \in \Delta \mid m \in M\}| \cdot |\{(p, m, q) \in \Delta \mid m \in M\}| = 0$; that is, every state is the source or the target state of transitions labelled with markers (or none of the above), but not both at the same time;
2. for every $m \in M$ and $(q, m, q') \in \Delta$: $|\{(q, a, r) \in \Delta \mid a \in \Sigma \cup M \cup \{\varepsilon\}\}| = 1$ and $|\{(r, a, q') \in \Delta \mid a \in \Sigma \cup M \cup \{\varepsilon\}\}| = 1$; that is, every source state of a marker transition has no other outgoing transitions, and every target state of a marker transition has no other incoming transitions.

Lemma 6. *For every NFA $\mathcal{A}$ over an alphabet $\Sigma \cup M$ one can effectively construct an unravelled NFA $\mathcal{A}'$ with $L(\mathcal{A}) = L(\mathcal{A}')$.*

Proof. Let $\mathcal{A} = (Q, \Sigma \cup M, q_0, \Delta, F)$ be an NFA. Construct $\mathcal{A}' = (Q', \Sigma \cup M, q_0', \Delta', F')$ with

- $Q' := \{q_0'\} \cup \{(\overline{p}, a, q), (p, a, \overline{q}) \mid (p, a, q) \in \Delta\}$,
- $F' := \{(p, a, \overline{q}) \mid (p, a, q) \in \Delta, q \in F\} \cup \{q_0' \mid q_0 \in F\}$, and
- $\Delta' := \{(q_0', \varepsilon, (\overline{q_0}, a, q)) \mid (q_0, a, q) \in \Delta\}$
  $\cup \{((\overline{p}, a, q), a, (p, a, \overline{q})) \mid (p, a, q) \in \Delta\}$
  $\cup \{((p, a, \overline{q}), \varepsilon, (\overline{q}, b, r)) \mid (p, a, q), (q, b, r) \in \Delta\}$.

Then $L(\mathcal{A}') = L(\mathcal{A})$, and $\mathcal{A}'$ is unravelled.

A state $(\overline{p}, a, q)$ in $\mathcal{A}'$ symbolizes that p is the current state and (p, a, q) the next transition to be taken in a run of $\mathcal{A}$; $(p, a, \overline{q})$ denotes that q is the current state and (p, a, q) is the last transition used in a run of $\mathcal{A}$. After every such step, a transition of the form $((p, a, \overline{q}), \varepsilon, (\overline{q}, b, r))$ allows us to guess the next transition taken in a run of $\mathcal{A}$ (in this case (q, b, r)). We omit the details of the correctness proof due to space restrictions. □

The notion of unravelled NFA is important for the following theorem.

Theorem 7. *TIRSs with tag-removing rules effectively preserve regularity.*

Proof. Let $\mathcal{R} = (\Sigma, M, R)$ be a TIRS with tag-removing rules, and let $\mathcal{A} = (Q, \Sigma \cup M, q_0, \Delta, F)$ be an unravelled NFA with $L(\mathcal{A}) = L$. We provide an algorithm that constructs an NFA $\mathcal{A}'$ from $\mathcal{A}$ such that $L(\mathcal{A}') = \mathcal{R}^{\circledast}(L)$. For this, we first extend an initial automaton $\mathcal{A}_0 = (Q_0, \Sigma \cup M, q_0, \Delta_0, F)$ with $Q_0 := Q$ and $\Delta_0 := \Delta$ as follows.

We have to capture derivation at and between all possible combinations of markers, possibly involving the deletion of markers. If, for instance, there is a rule $\#U \hookrightarrow \#V$ in R, then it may be applied at different positions of the marker $\#$ in $\mathcal{A}$, and we thus have to distinguish between these applications to avoid side effects. Therefore, we add normalised NFAs for all $(p, m, q), (p', m', q') \in \Delta$ with $m, m' \in M$, taking disjoint copies for different applications of rules inside the given automaton:

- for every prefix rule of the form $mU \hookrightarrow mV$ or $mU \hookrightarrow V$ in R, we add $\mathcal{A}_{(p,q,V)} = (Q_{(p,q,V)}, \Sigma, s_{(p,q,V)}, \Delta_{(p,q,V)}, \{t_{(p,q,V)}\})$ with $L(\mathcal{A}_{(p,q,V)}) = V$; we set $Q_0 := Q_0 \cup Q_{(p,q,V)}$ and $\Delta_0 := \Delta_0 \cup \Delta_{(p,q,V)}$, and we add $(q, \varepsilon, s_{(p,q,V)})$ (resp. $(p, \varepsilon, s_{(p,q,V)})$) to Δ_0;
- for every suffix rule of the form $Um' \hookrightarrow Vm'$ or $Um' \hookrightarrow V$ in R, we add $\mathcal{A}_{[p',q',V]} = (Q_{[p',q',V]}, \Sigma, s_{[p',q',V]}, \Delta_{[p',q',V]}, \{t_{[p',q',V]}\})$ with $L(\mathcal{A}_{[p',q',V]}) = V$; we set $Q_0 := Q_0 \cup Q_{[p',q',V]}$ and $\Delta_0 := \Delta_0 \cup \Delta_{[p',q',V]}$, and we add $(t_{[p',q',V]}, \varepsilon, p')$ (resp. $(t_{[p',q',V]}, \varepsilon, q')$) to Δ_0;
- for every bifix rule of the form $mUm' \hookrightarrow mVm'$, $mUm' \hookrightarrow mV$, $mUm' \hookrightarrow Vm'$, or $mUm' \hookrightarrow V$ in R, we add $\mathcal{A}_{(p,q,p',q',V)} = (Q_{(p,q,p',q',V)}, \Sigma, s_{(p,q,p',q',V)}, \Delta_{(p,q,p',q',V)}, \{t_{(p,q,p',q',V)}\})$ with $L(\mathcal{A}_{(p,q,p',q',V)}) = V$; we set $Q_0 := Q_0 \cup Q_{(p,q,p',q',V)}$ and $\Delta_0 := \Delta_0 \cup \Delta_{(p,q,p',q',V)}$, and we add $(q, \varepsilon, s_{(p,q,p',q',V)})$ in the first two cases resp. $(p, \varepsilon, s_{(p,q,p',q',V)})$ in the last two cases to Δ_0.

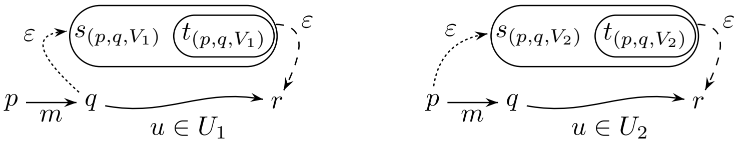
For the automaton $\mathcal{A}_0$ generated this way, we have $L(\mathcal{A}_0) = L(\mathcal{A})$.

For the sketch of the correctness proof later on, let Q_i denote the set of all initial states of the NFAs added for suffix rules, and let Q_f denote the set of all final states of the NFAs added for prefix and bifix rules.

After these preparatory steps, we now repeat the following saturation steps until no more transitions can be added, starting with $k = 0$:

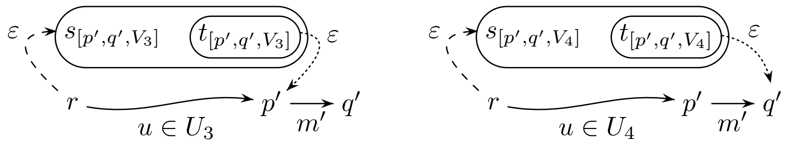
1. If there are $(p, m, q) \in \Delta$, $r \in Q_0$, a prefix rule of the form $mU \hookrightarrow mV$ or $mU \hookrightarrow V$ in R, and a path $\mathcal{A}_k : q \xrightarrow{u} r$ for some $u \in U$, then we add the transition $(t_{(p,q,V)}, \varepsilon, r)$ to Δ_k to obtain $\mathcal{A}_{k+1}$, and we set $k := k + 1$.

   The following illustrates this for rules $mU_1 \hookrightarrow mV_1$ and $mU_2 \hookrightarrow V_2$ and a path $p \xrightarrow{m} q \xrightarrow{u} r$. The dotted lines denote the transitions added in the preparatory steps, while the dashed lines show the ε-transitions added in the saturation steps.



2. If there are $(p', m', q') \in \Delta$, $r \in Q_0$, a suffix rule of the form $Um' \hookrightarrow Vm'$ or $Um' \hookrightarrow V$ in R, and a path $\mathcal{A}_k : r \xrightarrow{u} p'$ for some $u \in U$, then we add the transition $(r, \varepsilon, s_{[p',q',V]})$ to Δ_k to obtain $\mathcal{A}_{k+1}$, and we set $k := k+1$.

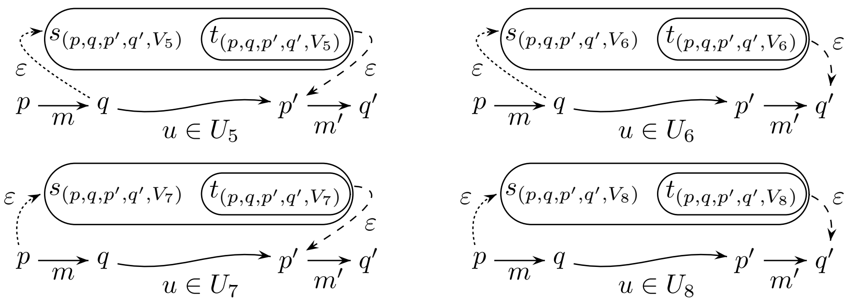
   The following illustrates this for rules $U_3m' \hookrightarrow V_3m'$ and $U_4m' \hookrightarrow V_4$ and a path $r \xrightarrow{u} p' \xrightarrow{m'} q'$.



3. If there are $(p, m, q), (p', m', q') \in \Delta$ and a path $\mathcal{A}_k : q \xrightarrow{u} p'$ for some $u \in U$ for a bifix rule of the form
   (a) $mUm' \hookrightarrow mVm'$ or $mUm' \hookrightarrow Vm'$ in R, then we add the transition $(t_{(p,q,p',q',V)}, \varepsilon, p')$ to Δ_k;
   (b) $mUm' \hookrightarrow mV$ or $mUm' \hookrightarrow V$ in R, then we add the transition $(t_{(p,q,p',q',V)}, \varepsilon, q')$ to Δ_k;

   we obtain $\mathcal{A}_{k+1}$, and we set $k := k+1$.

   The case of bifix rules of the form $mU_5m' \hookrightarrow mV_5m'$, $mU_6m' \hookrightarrow mV_6$, $mU_7m' \hookrightarrow V_7m'$, and $mU_8m' \hookrightarrow V_8$ is basically a combination of cases 1. and 2. above.



After saturating $\mathcal{A}_0$ this way, we set $\mathcal{A}' := \mathcal{A}_k$, thereby obtaining the desired automaton with $L(\mathcal{A}') = \mathcal{R}^{\circledast}(L)$. Since only finitely many transitions can be added in the saturation steps, the algorithm terminates.

For the completeness of the algorithm, we can show by induction on n that if $z \underset{\mathcal{R}}{\longrightarrow}^{(n)} w$ for some $z \in L(\mathcal{A})$, then there is a path $\mathcal{A}' : q_0 \xrightarrow{w} F$. For the soundness, we can show that if there is a path $\mathcal{A}' : q_0 \xrightarrow{w} F$, then $w \in \mathcal{R}^{\circledast}(L(\mathcal{A}))$. This follows directly from the more general claim

$$\mathcal{A}' : p \xrightarrow{w} q \text{ with } p \in Q \cup Q_i \wedge q \in Q \cup Q_f \;\Rightarrow\; \exists w' : w' \underset{\mathcal{R}}{\longrightarrow}^{\circledast} w \wedge \mathcal{A}_0 : p \xrightarrow{w'} q \;\;.$$

For $p = q_0$ and $q \in F$ this yields the original claim. Note that we are using Q (states of the original automaton $\mathcal{A}$) in the claim, not Q_0. We omit the proof details due to space restrictions. $\square$

3.3 Extending TIRSs by Adding Tags

We extend our basic model such that R allows rules of the forms $\#U \hookrightarrow \#V$, $U\# \hookrightarrow V\#$, and $\#U\$ \hookrightarrow \#V\$$, where $U \subseteq \Sigma^*$ and $V \subseteq (\Sigma \cup M)^*$ are regular sets. This means that the right hand sides of rules may contain new tags, thereby allowing tags to be added when rewriting words.

It turns out that regularity is not preserved with this extension, and thus also the derivation relation is not rational in general. In view of Theorem 7, this illustrates well that the two cases of removing and of adding tags behave differently with respect to preservation of regularity.

Proposition 8. *TIRSs with tag-adding rules do not preserve regularity.*

Proof. Consider $\mathcal{R} = (\{a\}, \{\#\}, R)$, where R contains only the rule $\#a \hookrightarrow \#\#a\#$. Then $\mathcal{R}^{\circledast}(\{\#a\#\}) = \{\#^n a \#^n \mid n > 0\}$ is not regular. $\square$

However, we still keep decidability of the word-to-word reachability problem.

Theorem 9. *The word-to-word reachability problem for TIRSs with tag-adding rules is decidable.*

Proof. Let $\mathcal{R} = (\Sigma, M, R)$ be a TIRS with tag-adding rules, and let $u, v \in (\Sigma \cup M)^*$. Let $|w|_M$ denote the number of markers of M in w. If $|u|_M > |v|_M$, then clearly v is not reachable from u. Otherwise, a maximum of $n := |v|_M - |u|_M$ rewriting steps that add tags will suffice to obtain v from u, if at all possible. Let R_0 denote the set of rules of R that do not add tags, and let $R_1 = R \setminus R_0$. Similarly, let $\mathcal{R}_0 = (\Sigma, M, R_0)$ and $\mathcal{R}_1 = (\Sigma, M, R_1)$. Then we have to iterate the following at most n times to decide whether v is reachable from u, starting with $i = 0$ and $U_0 = \{u\}$:

1. Set $i := i + 1$, and compute $U_i' := \xrightarrow[\mathcal{R}_0]{}^{\circledast} (U_{i-1})$ and $U_i := \xrightarrow[\mathcal{R}_1]{} (U_i')$;
2. If $v \in U_i$, then v is reachable from u, else if $i = n$, then v is not reachable from u.

With the algorithm of Theorem 7, we can compute an NFA recognizing U_i' in every step, starting from an unravelled NFA recognizing U_{i-1}. Then, since $\xrightarrow[\mathcal{R}_1]{}$ is rational, U_i is also effectively regular. $\square$

3.4 Remarks on Further Extensions

There are several natural ways how the basic model of TIRS may be extended further. For instance, one may allow tag-removing and tag-adding rules at the same time, or rules might be allowed to rename the tags that are involved in a rewriting step. It is not difficult to see that these models allow to transfer information across tags in either direction, which makes it possible to move markers arbitrarily and thus to apply rewriting rules at any position within a word. Therefore, these models are Turing powerful, and all interesting properties over such systems are undecidable.

Another interesting extension is to allow information transfer across tags in only one direction, e. g. by allowing rules of the form $u\# \hookrightarrow \#v$. In [8], Karhumäki et al. distinguished the cases of *controlled* or *uncontrolled* transfer. In the controlled case, a connection of the u's and v's is allowed, that is, the word to be removed to the left of the marker # can determine the word to be added to the right of #. In the uncontrolled case, no such connection is allowed, that is, the words to be removed and added are chosen independently. Karhumäki et al. showed that the language derivable from a regular initial set $L \subseteq \Sigma^*\#\Sigma^*$ is context-free in the case of uncontrolled transfer. For the controlled case however, they showed that even finite relations for the transfer suffice to obtain computational universality.

4 Comparison with Ground Tree Rewriting

Ground tree rewriting systems (GTRSs) have been studied intensively in [11]. They allow to substitute subtrees of finite ranked trees by other finite trees according to given rules. In this section we give a comparison with bifix rewriting systems.

Ranked trees are finite ordered trees over some ranked alphabet A which determines the labels and numbers of successors of nodes in a tree. T_A denotes the set of all finite trees over a given ranked alphabet A. A GTRS is a structure $\mathcal{R} = (A, \Sigma, R, t_{\text{in}})$, where A is a ranked alphabet, Σ is an alphabet to label rewriting rules, R is a finite set of rewriting rules of the form $s \overset{\sigma}{\hookrightarrow} s'$, where $\sigma \in \Sigma$ and $s, s' \in T_A$, and $t_{\text{in}} \in T_A$ is the initial tree.

Intuitively, a rule $s \overset{\sigma}{\hookrightarrow} s'$ may be applied to a tree $t \in T_A$ if s is a proper subtree of t. Applying the rule yields a tree that is obtained from t by replacing one occurrence of the subtree s by s'.

It is easy to realize the infinite $\mathbb{N} \times \mathbb{N}$ grid by a GTRS (using a tree of two unary branches of lengths i, j to represent vertex (i, j)). Hence the MSO theory of a GTRS graph is in general undecidable. As shown in [12], even the "universal reachability problem" ("Does every path from v reach a vertex in a regular tree set T?") is undecidable. On the other hand, as also shown in [12], the first-order theory with reachability (short: FO(R) theory) of a GTRS graph is decidable. In the FO(R) theory, the graph signature is extended by a symbol for the closure E^* of the edge relation E.

For bifix rewriting systems, the undecidability result on universal reachability is easily transferred from GTRSs. The proof for GTRSs only uses trees with two unary branches (for the representation of the left and right inscriptions of a Turing tape); in bifix rewriting systems, one simply combines the two branches into a single word with a separator between the left and right part.

It is remarkable that a converse simulation cannot work. This is clarified by the following result:

Theorem 10. *The FO(R) theory of a mixed prefix/suffix rewriting system is in general undecidable.*

For the proof, we remark that for the bifix rewriting system with rules $\Sigma \hookrightarrow \varepsilon$ for both prefix and suffix rewriting, the transitive closure gives the infix relation. As proved by Kuske [10], the first-order theory of Σ^* with the infix relation is undecidable.

This result shows that there is an essential difference between

- the "multiple stack" model that is inherent in ground tree rewriting (when a collection of unary branches is used as a list of stacks, with leaves as the top symbols of stacks), and
- the bifix rewriting model, where two stacks are easily simulated, but where an internal information flow between the two sides is possible.

5 Conclusion

We have introduced a general form of "tagged" rewriting system which extends the mixed prefix/suffix rewriting as studied in [3, 9], and where reachability (or the derivation relation) is decidable. We studied systematically the effects of removing and adding tags and showed that these cases are not dual. At the same time, we exhibited examples which separate decidability proofs by preservation of regularity, by rationality, or just by recursiveness of the derivation relation.

Many questions arise from these results in infinite-state system verification, where the universe of words with the tagged infix rewriting relation is considered as an infinite transition graph. For example, it should be investigated which logics admit an algorithmic solution of the model-checking problem over tagged infix rewriting graphs (see e. g. [13]). Another field of study is the definition of natural extended models where the derivation relation is no more rational, but still decidable.

Acknowledgement. I thank Didier Caucal, Christof Löding, and Wolfgang Thomas for their support and fruitful discussions, and anonymous reviewers for their helpful remarks.

References

1. Ahmed Bouajjani, Markus Mueller-Olm, and Tayssir Touili. Regular symbolic analysis of dynamic networks of pushdown systems. In *Proc. of the 16th CONCUR*, volume 3653 of *LNCS*, pages 473–487, 2005.
2. Richard Büchi. Regular canonical systems. *Archiv für Mathematische Logik und Grundlagenforschung*, 6:91–111, 1964.

3. Richard Büchi and William H. Hosken. Canonical systems which produce periodic sets. *Mathematical Systems Theory*, 4(1):81–90, 1970.
4. Didier Caucal. On the regular structure of prefix rewriting. In *Proc. of the 15th CAAP*, volume 431 of *LNCS*, pages 87–102, Copenhagen, 1990.
5. Jean-Luc Coquidé, Max Dauchet, Rémi Gilleron, and Sándor Vágvölgyi. Bottom-up tree pushdown automata: classification and connection with rewrite systems. *Theoretical Computer Science*, 127:69–98, 1994.
6. Javier Esparza, David Hansel, Peter Rossmanith, and Stefan Schwoon. Efficient algorithms for model checking pushdown systems. Technical Report TUM-I0002, Techn. Universität München, Institut für Informatik, 2000.
7. John Hopcroft, Rajeev Motwani, and Jeffrey Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison Wesley, 2000.
8. Juhani Karhumäki, Michal Kunc, and Alexander Okhotin. Communication of two stacks and rewriting. In *Proc. of the 33rd ICALP*, volume 4052 of *LNCS*, pages 468–479, 2006.
9. Juhani Karhumäki, Michal Kunc, and Alexander Okhotin. Computing by commuting. *Theoretical Computer Science*, 356(1-2):200–211, 2006.
10. Dietrich Kuske. Theories of orders on the set of words. *Theoretical Informatics and Applications*, 40:53–74, 2006.
11. Christof Löding. *Infinite Graphs Generated by Tree Rewriting*. Doctoral thesis, RWTH Aachen University, 2003.
12. Christof Löding. Reachability problems on regular ground tree rewriting graphs. *Theory of Computing Systems*, 39(2):347–383, 2006.
13. Richard Mayr. Process rewrite systems. *Information and Computation*, 156(1-2):264–286, 2000.
14. Kai Salomaa. Deterministic tree pushdown automata and monadic tree rewriting systems. *Journal of Computer and System Sciences*, 37:367–394, 1988.
15. Wolfgang Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B: Formal Models and Semantics, pages 133–192. Elsevier, Amsterdam, 1990.