

# Software Analysis and Verification

Mini Project

## Expressive Power of a Fragment of WS1S

June 29, 2007

Ghid Maatouk and Vaibhav Rajan

Professor: Viktor Kuncak

# Introduction

We know that weak monadic second-order theory of 1 successor (WS1S) is decidable through automata, that is, the set of satisfying interpretations of a subformula is represented by a finite-state automaton [1, 2]. Decision procedures for WS1S with certain restrictions have been implemented efficiently in MONA using the automata technique. The main problem with this approach is the exponential blow-up in the states of the automaton due to nested quantifiers [3]. We were initially interested in finding another way of deciding WS1S without using the automata. We were interested in investigating whether we can characterize relations in WS1S in a way that helps us eliminate quantifiers and began by looking at a fragment of WS1S. This paper describes the key results that we obtained.

## Mathematical preliminaries

We let  $\mathbb{N}$  denote the set of nonnegative integers  $\{0, 1, 2, \dots\}$ . Monadic second-order logic of 1 successor has the following minimalist syntax:

$$F ::= v \subseteq v \mid s(v, v) \mid F \vee F \mid \neg F \mid \exists v.F$$

where variable  $v$  can denote a first-order variable taking values in  $\mathbb{N}$  or a second-order variable (set variable) taking values in  $2^{\mathbb{N}}$ . We view first-order variables as singleton sets. The successor function  $s(v_1, v_2)$  is defined as the successor relation on integers lifted to singleton sets. We are interested in weak monadic second-order logic of 1 successor over strings. This means that natural numbers are represented as binary strings with a single 1 in the position corresponding to the natural number, and sets are represented as binary strings with 1 in the positions corresponding to the elements of the set. “Weak” means that we are only interested in finite strings.

We define  $\text{WS1S}^R$  to be the same language as that of WS1S with the restriction that there are no free second-order variables. Hence relations on this language can only be relations on first-order variables.

**Definable sets and relations** A subset  $S$  of  $\mathbb{N}^m$  is definable if there exists a formula  $F$  with free variables  $x_1, \dots, x_m$  such that

$$S = \{(d_1, \dots, d_m) \in \mathbb{N}^m \mid F_{x_i:=d_i}(x_1, \dots, x_m) = \text{true}\}.$$

A relation  $R(x_1, \dots, x_m)$  is definable in  $\text{WS1S}^R$  if there exists a formula  $F(x_1, \dots, x_m)$  of  $\text{WS1S}^R$  such that the relation viewed as a set is definable in  $\text{WS1S}^R$ .

**Ultimately periodic sets** Let  $S$  be a subset of  $\mathbb{N}$  with characteristic function  $s : \mathbb{N} \rightarrow \{0, 1\}$  such that

$$s(x) := \begin{cases} 0 & \text{if } x \in S \\ 1 & \text{if } x \notin S \end{cases}$$

**Definition 1** A sequence  $s : \mathbb{N} \rightarrow \{0, 1\}$  is ultimately periodic iff there exists  $n_0 \in \mathbb{N}$  and  $v \in \mathbb{N}$  such that  $\forall n \geq n_0. s_n = s_{n+v}$ .

**Definition 2** A set  $S$  is ultimately periodic iff its characteristic function  $s$  is ultimately periodic.

We define the linear sets  $S(a, b) = \{a + kb \mid k \in \mathbb{N}\}$ .

**Claim 1** A set is ultimately periodic if and only if it is a union of a finite number of sets of the form  $S(a, b)$ .

**Proof of Claim 1**

Consider an ultimately periodic set  $S$  with some  $v$  and  $n_0$ . Consider the set  $I \subseteq \{1, \dots, n_0 - 1\}$  of numbers  $i < n_0$  belonging to  $S$ , and the set  $J \subseteq \{n_0, \dots, n_0 + v - 1\}$  of numbers  $j$  between  $n_0$  and  $n_0 + v - 1$  belonging to  $S$ . Then  $S = \bigcup_{i \in I} \{i + 0k \mid k \in \mathbb{N}\} \cup \bigcup_{j \in J} \{j + vk \mid k \in \mathbb{N}\}$ .

We prove the other side of the equivalence by induction.

Base Case: Clearly  $S(a, b)$  is ultimately periodic ( $n_0 = a, v = b$ ).

Induction: The union of an ultimately periodic set  $(n_{old}, v_{old})$  and a set of the form  $S(a, b)$  is an ultimately periodic set with  $n_0 = \max(n_{old}, a)$  and  $v = b \times v_{old}$ .  $\square$

## Characterizing Relations in $\text{WS1S}^R$

The aim of this project is to characterize the relations (on first-order variables) definable in  $\text{WS1S}^R$ . We first prove a useful intermediary result, then use it to characterize unary and binary relations in  $\text{WS1S}^R$ .

## Length of words of regular languages

Let  $L$  be a regular language over an alphabet  $\Sigma$  and  $M = \{|w|, w \in L\}$  be the set of lengths of words in  $L$ .

**Claim 2**  $M$  is a union of linear sets  $S(a, b)$  for some finite number of tuples  $(a, b)$ .

We say that a set has property  $F$  if it is a union of finitely many sets  $S(a, b)$  for some tuples  $(a, b)$ . We say a language  $L$  (or the corresponding regular expression) has property  $P$  if its set of lengths  $M$  has property  $F$ . Hence, we need to prove that any regular language has property  $P$ .

First we state and prove the following lemma.

**Lemma 1** *The set  $S = \{a + bk_1 + ck_2, k_1, k_2 \in \mathbb{N}\}$ , satisfies property  $F$ .*

### Proof

Given the set  $S = \{a + bk_1 + ck_2, k_1, k_2 \in \mathbb{N}\}$ , consider  $\gcd(b, c)$ . The following two cases are mutually disjoint and exhaustive:

1.  $\gcd(b, c) = 1$
2.  $\gcd(b, c) = d, d \neq 1$ .

We analyze each case separately.

**Case 1:**  $\gcd(b, c) = 1$ . Consider the numbers contained in the set  $\{bk_1 + ck_2\}$ . Any such number is congruent to  $ck_2 \pmod{b}$ . Letting  $k_2$  range over  $\{0, \dots, b-1\}$ , we obtain for each value of  $k_2$  a distinct congruence class  $\pmod{b}$ :  $[0], [c], [2c], \dots, [(b-1)c]$ .

The congruence classes are distinct by the following argument: assume there exist  $j$  and  $l$  such that  $0 \leq j, l < b$  and  $jc = lc \pmod{b}$ . Then  $jc + mb = lc + nb$  for some  $m, n$ . Hence  $(l-j)c = (m-n)b$ . Since  $b$  divides  $(m-n)b$ , it must divide  $(l-j)c$ . But  $\gcd(b, c) = 1$  hence  $b$  must divide  $l-j$ . But  $l-j < b$ , hence we must have  $l-j = 0$ .

Since there are  $b$  distinct congruence classes  $\pmod{b}$ , the set  $\{bk_1 + ck_2\}$  contains all natural numbers beyond  $n_0 = (b-1)(c-1)$ . This is because of the following theorem from elementary number theory: If  $\gcd(b, c) = 1$  and  $n \geq (b-1)(c-1)$ , then  $bx + cy = n$  has a non-negative solution, that is, one in which both  $x$  and  $y$  are non-negative integers. There are also finitely many

numbers below  $n_0$  that are contained in the set  $\{bk_1 + ck_2\}$  (For example, 0 is in the set).

Hence, the set  $\{a + bk_1 + ck_2\}$  is equal to the union of a finite number of constants (less than  $a + n_0$ ) and the set  $\{a + n_0 + k\}$ . Each constant  $c$  can be represented in the set  $\{c + 0.k, k \in \mathbb{N}\}$ . In other words, if  $\gcd(b, c) = 1$ , then the set  $S = \{a + bk_1 + ck_2, k_1, k_2 \in \mathbb{N}\}$  satisfies property  $F$ .

**Case 2:**  $\gcd(b, c) = d, d \neq 1$ . Any number of the form  $bk_1 + ck_2$  can be written as  $d(mk_1 + nk_2)$  where  $b = md, c = nd$  and  $\gcd(m, n) = 1$ . From Case 1, we know that the set  $T = \{mk_1 + nk_2, k_1, k_2 \in \mathbb{N}\}$  satisfies property  $F$ . The set  $T' = \{d(mk_1 + nk_2), k_1, k_2 \in \mathbb{N}\}$  can be obtained from  $T$  by multiplying the constant factors ( $a$  and  $b$ ) in each element of  $T$  by  $d$ . So  $T'$  satisfies property  $F$ .

By Case 1 and Case 2, if  $\gcd(b, c) = d$  then the set  $S = \{a + bk_1 + ck_2, k_1, k_2 \in \mathbb{N}\}$  satisfies property  $F$ .  $\square$

Now we can prove the claim.

## Proof of Claim 2

We prove the claim by induction on the structure of the regular expression representing the language.

*Base Case* The length of a single character in  $\Sigma$  is 1. Thus the length of any word corresponding to a single character belongs to the set  $\{1 + 0.k, k \in \mathbb{N}\}$ ; hence any such word has property  $P$ .

*Inductive Step* We prove that the operations union, concatenation and Kleene closure of two regular expressions satisfying property  $P$  yield a regular expression satisfying property  $P$ .

1. **Union** Let  $r_1$  and  $r_2$  be two regular expressions satisfying property  $P$ . Then, their corresponding sets of lengths  $M_1$  and  $M_2$  satisfy property  $F$ . The set of lengths  $M$  of  $r = r_1 \cup r_2$  is simply  $M = M_1 \cup M_2$ . Hence  $M$  satisfies property  $F$  and thus  $r$  satisfies property  $P$ .
2. **Concatenation** Let  $r_1$  and  $r_2$  be two regular expressions satisfying property  $P$ . Then their corresponding sets of lengths  $M_1$  and  $M_2$  satisfy property  $F$ . The set of lengths of  $r = r_1.r_2$  is the finite set  $Q = \{t_1 + t_2, t_1 \in M_1, t_2 \in M_2\}$ . This set contains elements of the form

$a_1 + bk_1 + a_2 + ck_2 = a + bk_1 + ck_2$ . By Lemma 1,  $Q = \{a + bk_1 + ck_2\}$  satisfies property  $F$ . Thus,  $r$  satisfies property  $P$ .

3. **Kleene Closure** Let  $r$  be a regular expression satisfying property  $P$ . The corresponding set of lengths  $M$  satisfies property  $F$ .  $M = \bigcup_{i \in I} m_i$  where  $m_i$  is a linear  $S(a, b)$  and  $I$  is a finite subset of  $\mathbb{N}$ . The regular expression  $r^*$  can contain any number of repetitions (including zero) of any of the sub-regular expressions. Hence, the possible lengths of words in  $r^*$  is given by  $M^* = \sum_{i \in I} m_i k_i$ ,  $k_i \in \mathbb{N}$ . Each term of this summation is of the form  $(a_i + kb_i)k_i$ ,  $k, k_i \in \mathbb{N}$ .

Consider one such term:  $(a + kb)k_1$ . For different values of  $k$ , we get terms of the form:  $ak_1, (a + b)k_1, (a + 2b)k_1, (a + 3b)k_1, \dots$ . Any number of the form  $(a + nb)k$  can also be obtained from  $ak_1 + (a + b)k_2 = a(k_1 + k_2) + bk_2$  by choosing the constants appropriately:  $k = k_1 + k_2, nk = k_2$ . Thus, the term  $(a + kb)k_1$  can be written as  $ak_1 + (a + b)k_2$ . However, this is true if  $k_1$  and  $k_2$  can also take negative values.

As an example, consider the regular expression  $r = (aaa(aaaaa)^*)$ . The set of lengths of words corresponding to its language is  $3 + 5k, k \in \mathbb{N}$ . The corresponding set for  $r^*$  is a union of the set  $\{3k_1 + 8k_2, k_1, k_2 \in \mathbb{N}\}$  and  $\{13\}$ .

So  $M^* = \sum_{i \in I} a_i k_i + c_i k_j$  where  $c_i = a_i + b_i$  and  $k_1, k_2 \in \mathbb{N}$ . From Lemma 1, we know that the set  $\{a_i k_i + c_i k_j\}$  satisfies property  $F$ . Thus,  $M^*$  is the sum of finitely many sets, each of them satisfying property  $F$ . We can repeatedly combine the terms of all these sets and by the same argument as in the case of concatenation, we know that the resulting set also satisfies property  $F$ . We have neglected the values obtained due to negative values of the variables  $k_1$  and  $k_2$ . However, note that the set  $\{a_i k_i + c_i k_j\}$  contains all natural numbers beyond  $(a_i - 1)(c_i - 1)$  and there are finitely many numbers between zero and  $(a_i - 1)(c_i - 1)$  that are in  $M^*$ . Hence,  $M^*$  satisfies property  $F$ .

Since any regular expression can be constructed only by the application of the above three steps (union, concatenation and Kleene closure), the language corresponding to any regular expression satisfies property  $P$ .  $\square$

## Characterizing unary relations in $\text{WS1S}^R$

We consider the special case of a unary relation on natural numbers corresponding to a formula  $F(x)$  with only one free variable  $x$ . This formula defines a set  $S = \{x \mid F(x)\}$ .

**Claim 3** *Set  $S$  is ultimately periodic.*

### Proof of Claim 3

Given formula  $F(x)$ , there exists an automaton  $A$  with input alphabet  $\{0, 1\}$  which accepts the string  $0^{x-1}10^*$  iff  $F(x)$  is true. We define the automaton  $A^*$  on  $\{0, 1\}$  which accepts the string  $0^{x-1}1$  iff  $F(x)$  is true. Note that  $x = |0^{x-1}1|$ .

Also note that the language corresponding to  $0^{x-1}1$  such that  $F(x)$  is true is regular iff the language corresponding to  $0^{x-1}10^*$  such that  $F(x)$  is true is regular. Indeed, if  $0^{x-1}1$  is regular then  $0^{x-1}10^*$  is the concatenation of the regular expressions  $0^{x-1}1$  and  $0^*$  and hence is regular. Conversely, if  $0^{x-1}10^*$  is regular, then the language corresponding to  $0^{x-1}1$  is the intersection of the languages corresponding to the regular expressions  $0^{x-1}10^*$  and  $0^*1$ , and hence is regular.

Furthermore,  $A^*$  accepts string  $0^{x-1}1$  iff  $A$  accepts strings  $0^{x-1}10^*$ .

Consider the language  $L$  of  $A$ . By Claim 2, the set of lengths  $M_L$  of words in  $L$  is a finite union of linear sets  $S_i(a, b)$ . For all  $i$ , the set  $S_i(a, b)$  is the set of lengths of a subset of the words accepted by  $A$ , all of the form  $0^{x-1}1$ . Since  $x = |0^{x-1}1|$ ,  $S_i(a, b)$  is a subset of the integers  $x$  such that  $0^{x-1}1$  is accepted by  $A$ , i.e. a subset of  $S$ . Furthermore, the finite disjoint union  $\bigcup_i S_i(a, b) = M_L$  contains exactly all lengths of words accepted by  $A$ , hence  $\bigcup_i S_i(a, b)$  contains exactly all integers  $x$  such that  $0^{x-1}1$  is accepted by  $A$ , i.e.  $\bigcup_i S_i(a, b) = S$ .

$S$  is thus a finite union of linear sets  $S_i(a, b)$ . Hence by definition,  $S$  is ultimately periodic.  $\square$

## Characterizing binary relations in $\text{WS1S}^R$

Consider a binary relation  $R(x, y)$  in  $\text{WS1S}^R$  and the corresponding formula  $F(x, y)$  with two free variables  $x$  and  $y$ .  $F(x, y)$  defines a set  $Q = \{(x, y) \mid F(x, y)\}$ . Elements of the set  $Q$  are recognized by an automaton  $A$  with parallel inputs and input alphabet  $\Sigma = \left\{ \binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1} \right\}$ .

The language  $L$  of  $A$  is a subset of the regular language corresponding to  $\binom{0}{0}^* \binom{1}{0} \binom{0}{0}^* \binom{0}{1} \binom{0}{0}^* \cup \binom{0}{0}^* \binom{0}{1} \binom{0}{0}^* \binom{1}{0} \binom{0}{0}^* \cup \binom{0}{0}^* \binom{1}{1} \binom{0}{0}^*$ . Thus the input string to the automaton is of the form  $\binom{0}{0}^{k_1} \binom{1}{0} \binom{0}{0}^{k_2} \binom{0}{1} \binom{0}{0}^{k_3}$  or  $\binom{0}{0}^{k_4} \binom{0}{1} \binom{0}{0}^{k_5} \binom{1}{0} \binom{0}{0}^{k_6}$  or  $\binom{0}{0}^{k_7} \binom{1}{1} \binom{0}{0}^{k_8}$ .

**Claim 4** *If a pair  $(x, y)$  belongs to some set  $Q = \{(x, y) \mid F(x, y)\}$  for some  $F$  over  $WS1S^R$  then the sets  $\{k_i\}_i$  of possible values of the exponents  $k_i$  are ultimately periodic.*

#### Proof of Claim 4

Let  $(x, y)$  belong to  $Q = \{(x, y) \mid F(x, y)\}$ . There exists an automaton  $M$  with parallel inputs which accepts the corresponding input string.

The input string can be of one of the following forms:

1.  $\binom{0}{0}^{k_1} \binom{1}{0} \binom{0}{0}^{k_2} \binom{0}{1} \binom{0}{0}^{k_3}$
2.  $\binom{0}{0}^{k_4} \binom{0}{1} \binom{0}{0}^{k_5} \binom{1}{0} \binom{0}{0}^{k_6}$
3.  $\binom{0}{0}^{k_7} \binom{1}{1} \binom{0}{0}^{k_8}$ .

Case 1: We want to show that the sets  $\{k_1\}$ ,  $\{k_2\}$  and  $\{k_3\}$  are ultimately periodic. By claim 2, it is sufficient to show that the languages  $\binom{0}{0}^{k_i}$ ,  $i = 1, 2, 3$  generated by the unary alphabet  $\binom{0}{0}$  are regular. We show that  $\binom{0}{0}^{k_2}$  is regular by constructing an accepting automaton,  $A = (\Sigma, \delta, s, f, S)$ . Let  $M = (\Sigma', \delta', s', f', S')$  be the automaton that accepts the original input string. Let  $R$  be the accepting run in  $M$ . In  $R$ , there exist unique states  $m$  and  $n$  such that  $\delta'(m, \binom{1}{0}) = n$ . Define  $s = n$ . Similarly there exist unique  $g, h$  such that  $\delta'(g, \binom{0}{1}) = h$ . Define  $f = g$ .  $S$  contains all the states of  $S'$  that are in  $R$  between  $n$  and  $g$ . And,  $\delta$  is the restriction of  $\delta'$  to the transitions in  $R$  between  $n$  and  $g$ .  $\Sigma = \{\binom{0}{0}\}$ .

$A$  accepts  $\binom{0}{0}^{k_2}$  because  $M$  accepts the original input string. We thus have an automaton which accepts language  $\binom{0}{0}^{k_2}$ , hence this language is regular.

We can similarly construct automata for the languages  $\binom{0}{0}^{k_1}$  and  $\binom{0}{0}^{k_3}$ . In the former case, the start state corresponds to  $s'$  and the final state is the state (in  $R$ ) from which there is a transition on consuming  $\binom{1}{0}$ . In the latter



case, the final state corresponds to  $f'$  and the start state is that state in  $R$  that is reached after consuming  $\binom{0}{1}$ .

The construction of the automata for cases 2 and 3 is along the same lines. Hence we have that if a pair  $(x, y)$  belongs to some set  $Q = \{(x, y) \mid F(x, y)\}$  for some  $F$  over  $\text{WS1S}^R$ , then the sets of values of the exponents in the corresponding input string are ultimately periodic.

**Example** Consider the relation  $R = \{(x, y) \mid x \equiv y \pmod{3}\}$ . Then there exists an automaton  $A$  which accepts input strings of the form  $\binom{0}{0}^{k_1} \binom{1}{0} \binom{0}{0}^{3k_2} \binom{0}{1} \binom{0}{0}^{k_3}$  and  $\binom{0}{0}^{k_4} \binom{0}{1} \binom{0}{0}^{3k_5} \binom{1}{0} \binom{0}{0}^{k_6}$  and  $\binom{0}{0}^{k_7} \binom{1}{1} \binom{0}{0}^{k_8}$ , where  $k_i \in \mathbb{N}$  for all  $i$ . The sets  $\{k_1\}$ ,  $\{3k_2\}$ , etc. are clearly ultimately periodic.

### Generalization to $\mathbb{N}^m$

The above reasoning can be extended to the general  $m$ -ary case. Now the input string has at most  $m$  non-zero elements and at most  $m + 1$  expressions

of the form  $\binom{0}{\vdots}^{k_i}$ . Note that there are finitely many possible formats for

the input string. If a tuple  $(x_1, x_2, \dots, x_m)$  belongs to an  $m$ -ary relation, then the set of possible values of each exponent  $k_i$  is ultimately periodic.

## Conclusion

In the course of this project we proved an interesting result about the lengths of words of a regular language. We used this result to characterize relations in the language  $\text{WS1S}^R$ . We hope that this is a step in the direction of our original aim: characterizing relations in  $\text{WS1S}$ .

# Bibliography

- [1] Julius Richard Buchi. *On a decision method in restricted second-order arithmetic*. Proc. Internat. Cong. on Logic, Methodol., and Philos. of Sci. Stanford University Press, 1960.
- [2] Calvin C. Elgot. *Decision problems of Finite automata design and related arithmetics*. Transactions of the American Mathematical Society, 98:21-52, 1961.
- [3] Nils Klarlund. *MONA Version 1.4 User Manual*. BRICS, Department of Computer Science, University of Aarhus, 2001.
- [4] Veronique Bruyere, Georges Hansel, Christian Michaux, and Roger Villemaire. *Logic and  $p$ -recognizable sets of integers* Bull. Belg. Math. Soc. 1, 191-238, 1994.