Reasoning About Sequences of Memory States
Rémi Brochenin, Stéphane Demri, Etienne Lozes

To cite this version:
Rémi Brochenin, Stéphane Demri, Etienne Lozes. Reasoning About Sequences of Memory States. Proceedings of the Symposium on Logical Foundations of Computer Science (LFCS07), Sergei Artemov; Anil Nerode, Jul 2007, New York, United States. pp.100-114, 10.1007/978-3-540-72734-7_8. hal-03201337

HAL Id: hal-03201337
https://hal.science/hal-03201337
Submitted on 18 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Reasoning about sequences of memory states*

Rémi Brochenin, Stéphane Demri, and Etienne Lozes

LSV, ENS Cachan, CNRS, INRIA
{brocheni,demri,lozes}@lsv.ens-cachan.fr

Abstract. In order to verify programs with pointer variables, we introduce a temporal logic LTL_{mem} whose underlying assertion language is the quantifier-free fragment of separation logic and the temporal logic on the top of it is the standard linear-time temporal logic LTL. We analyze the complexity of various model-checking and satisfiability problems for LTL_{mem}, considering various fragments of separation logic (including pointer arithmetic), various classes of models (with or without constant heap), and the influence of fixing the initial memory state. We provide a complete picture based on these criteria. Our main decidability result is PSPACE-completeness of the satisfiability problems on the record fragment and on a classical fragment allowing pointer arithmetic. \Sigma_1^0-completeness or \Sigma_1^1-completeness results are established for various problems by reducing standard problems for Minsky machines, and underline the tightness of our decidability results.

1 Introduction

Verification of programs with pointers. Model-checking of infinite-state systems is a very active area of formal verification [BCM01] even though in full generality, simple reachability questions are undecidable. Nevertheless, many classes of infinite-state systems can be analyzed, such as Petri nets, timed automata, etc. Programs with pointer variables suffer the same drawback since reachability problems are also undecidable, see e.g. [BFN04,BBH +06]. It is worth noting that specific properties need to be verified for such programs, such as the existence of memory leaks, memory violation, or shape analysis. Prominent logics for analyzing such programs are Separation Logic [Rey02], pointer assertion logic PAL [JJKS97], TVLA [LAS00] and alias logic [BIL04], to quote a few examples. Temporal Separation Logic: what for? Since [Prie77], temporal logics are also used as languages for formal specification of programs. General and powerful automata-based techniques for verification have been developed, see e.g. [VW94]. On the other hand, Separation Logic is a static logic for program annotation [Rey02], and more recently for symbolic computation [BCO05]. Extending the scope of application of Separation Logic to standard temporal logic-based verification techniques has many potential interests. First, it provides a rich underlying assertion language where properties more complex than accessibility

* Work supported by the RNTL project “AVERILES”. The first author is supported by a fellowship from CNRS/DGA.
can be stated. Second, this probably yields a significant feedback for the purely static Separation Logic extended with general recursion, which has not been very studied up to now. For instance, if we write \(Xx\) to denote the next value of \(x\) (also sometimes written \(x'\)), the formula \((x \leftarrow Xx)U(x \leftarrow \text{null})\), understood on a model with constant heap, characterises the existence of a simple flat list, which is usually written \(\mu L(x). x \leftarrow \text{null} \lor \exists x'. x \leftarrow x' \land L(x')\). Third, temporal logics allow to work in the very convenient framework of "programs-as-formulae" and decision procedures for logical problems can be directly used for program verification. For instance, the previous formula can be seen as a program walking on a list, and more generally programs without destructive updates can be expressed as formulae. Some programs with destructive updates that perform a simple pass on the heap, have an input-output relation that may be described by a formula. For instance, the formula \((x \rightarrow_{\rightarrow} Xx \land Xx \rightarrow_{\rightarrow} x)Ux \rightarrow_{\rightarrow} \text{null}\) expresses broadly that the list in initial heap \(h_0\) is reversed in final heap \(h_1\). Fourth, pointer arithmetic has been poorly studied until now, whereas arithmetical constraints in temporal logics are known to lead to undecidability, see e.g. \([\text{CC}00]\). Actually, there is a growing interest in understanding the interplay of pointer arithmetic, temporal reasoning, and non aliasing properties.

Our contribution. We introduce a linear-time temporal logic \(\text{LTL}^{\text{mem}}\) to specify sequences of memory states with underlying assertion language based on quantifier-free Separation Logic \([\text{Rey}02]\). From a logical perspective, the logic \(\text{LTL}^{\text{mem}}\) can be viewed as a many-dimensional logic \([\text{GKWZ}03]\) since \(\text{LTL}^{\text{mem}}\) contains a temporal dimension and the spatial dimension for memory states. Our logic addresses a very general notion of models, including the aspects of pointer arithmetic and recursive structures with records. We distinguish the satisfiability problems from the model-checking problems, as well as distinct subclasses of interesting programs, like for instance the programs without destructive update. The result that is the most promising for future implementation is the \(\text{PSPACE}\)-completeness of the satisfiability problems \(\text{SAT}(\text{CL})\) and \(\text{SAT}(\text{RF})\) where \(\text{CL}\) is the classical fragment without separation connectives and \(\text{RF}\) is the record fragment with no pointer arithmetic but with separation connectives. This result is very tight, as both propositional \(\text{LTL}\) and static Separation Logic are already \(\text{PSPACE}\)-complete \([\text{SC}85,\text{CYO}01]\). These results are obtained by reduction to the nonemptiness problem for Büchi automata on an alphabet of symbolic memory states obtained by an abstraction that we show sound and complete, see e.g. \([\text{Loz}04,\text{CGH}05]\). Such abstractions are similar to resource graphs from \([\text{GM}05]\). This is a variant of the automata-based approach introduced in \([\text{VW}94]\) for plain \(\text{LTL}\) and further developed with concrete domains of interpretation in \([\text{DD}07]\). Surprisingly, the abstraction method used to establish these results does not scale to the whole logic, due to a subtle interplay between separation connectives and pointer arithmetic. Moreover, we provide new undecidability results for several problems, for instance \(\text{SAT}^{\text{mem}}(\text{LF})\) (satisfiability with constant heap on the list fragment).

Related work. Previous temporal logics designed for pointer verification include Evolution Temporal Logic \([\text{YRSW}03]\), based on the three-valued logic abstrac-
tion method that made the success of TVLA [LAS00], and Navigation temporal logic [DKR04], based on a tableau method quite similar to our automaton-based reduction. In these works, the assertion language for states is quite rich, as it includes for instance list predicate, quantification over adresses, and a freshness predicate. Because of this high expressive power, only incomplete abstractions are proposed, whereas we stick to exact methods. More importantly, our work addresses models with constant heaps and pointer arithmetic, which has not been done so far, and leads to a quite different perspective.

Omitted proofs can be found in the report [BDL07].

2 Memory Model and Specification Language

In this section, we introduce a separation logic dealing with pointer arithmetic and record values, and a temporal logic $\text{LTL}^{\text{mem}}$. Unlike BI’s pointer logic from [1001], we allow pointer arithmetic. Model-checking programs with pointer variables over $\text{LTL}^{\text{mem}}$ specifications is our main problem of interest.

2.1 A separation logic with pointer arithmetic

Memory states. Let us introduce our model of memory. It captures features of programs with pointer variables that use pointer arithmetic and records. We assume a countably infinite set $\text{Var}$ of variables (as usual, for a fixed formula we need only a finite amount), and an infinite set $\text{Val}$ of values containing the set $\mathbb{N}$ of naturals, thought as address indexes, and a special value $\text{nil}$. For simplicity, we assume that $\text{Val} = \mathbb{N} \cup \{\text{nil}\}$. In order to model field selectors, we consider some infinite set $\text{Lab}$ of labels. We will usually range over values with $u, v$, over naturals with $i, j$, over labels with $l, r, next, prev$, and over variables with $x, y$. In the remainder, we will assume some fixed injection $(x, i) \in \text{Var} \times \mathbb{N} \rightarrow (x, i) \in \text{Var}$.

We use the notation $E \rightarrow_f F$ for the set of partial functions from $E$ to $F$ of finite domain; and $E \rightarrow_{fin} F$ for the set of partial functions from $E$ to $F$ of finite and nonempty domain. The sets $\mathcal{S}$ of stores and $\mathcal{H}$ of heaps are then defined as follows: $\mathcal{S} \overset{\Delta}{=} \text{Var} \rightarrow \text{Val}$ and $\mathcal{H} \overset{\Delta}{=} \mathbb{N} \rightarrow_{fin} (\text{Lab} \rightarrow_{fin} \text{Val})$. We will range over a store with $s, s'$ and over a heap with $h, h', h_1, h_2$. We call memory state a couple $(s, h) \in \mathcal{S} \times \mathcal{H}$.

We will refer to the domain of a heap $h$ by $\text{dom}(h) \subseteq \mathbb{N}$. Intuitively, in our memory model, each index is thought as an entry point on a record cell containing several fields. Cells are either not allocated, or allocated with some record stored in. In a memory state $(s, h)$, the memory cell at index $i$ is allocated if $i \in \text{dom}(h)$; in this case the stored record is $h(i) = \{l_1 \mapsto v_1, ..., l_n \mapsto v_n\}$.

Note that the size of the information held in a memory cell is not fixed, nor bounded. Our models could be more concrete considering labels as offsets and relying on pointer arithmetic. But for our purpose, it will be convenient to consider pointer arithmetic independently.
Expressions                   State Formulae
\( e ::= x \mid \text{null} \) \hspace{1cm} \( A ::= \pi \)

Atomic formulae
\( \pi ::= e = e' \mid e + i \leftarrow e \)
\( A \wedge B \mid A \rightarrow A \mid \top \mid \bot \) (classical fragment)

Satisfaction
\[ (s, h) \models_{SL} e = e' \] if \( \llbracket e \rrbracket_s = \llbracket e' \rrbracket_s \) with \( \llbracket x \rrbracket_s = s(x) \) and \( \llbracket \text{null} \rrbracket_s = \text{nil} \)
\[ (s, h) \models_{SL} e + i \leftarrow e \] if \( \llbracket e \rrbracket_s \in \mathbb{N} \) and \( \llbracket e \rrbracket_s + i \in \text{dom}(h) \) and \( h(s(x) + i)(l) = \llbracket e \rrbracket_s \)
\[ (s, h) \models_{SL} \text{emp} \] if \( \text{dom}(h) = \emptyset \)
\[ (s, h) \models_{SL} A_1 \ast A_2 \] if \( \exists h_1, h_2 \text{ s.t. } h = h_1 \ast h_2, (s, h_1) \models_{SL} A_1 \text{ and } (s, h_2) \models_{SL} A_2 \)
\[ (s, h) \models_{SL} A' \rightarrow A \] if for all \( h' \), if \( h \perp h' \) and \( (s, h') \models_{SL} A' \) then \( (s, h \ast h') \models_{SL} A \)
\[ (s, h) \models_{SL} A_1 \wedge A_2 \] if \( (s, h) \models_{SL} A_1 \) and \( (s, h) \models_{SL} A_2 \)
\[ (s, h) \models_{SL} A' \rightarrow A \] if \( (s, h) \models_{SL} A' \) implies \( (s, h) \models_{SL} A \)
\[ (s, h) \models_{SL} \bot \] never and \( (s, h) \models_{SL} \top \) always

| Table 1. The syntax and semantics of SL with pointer arithmetic and records |

---

**Separation Logic.** We now introduce the separation logic (SL) on top of which we will define our temporal logic. The syntax of the logic is given in Table 1.

In short, Separation logic is about reasoning on disjoint heaps, and we need to define what we mean by “disjoint heaps” in our model. We choose to allow to reason at the granularity of record cells, so that a record cell cannot be decomposed in disjoint parts. Let \( h_1 \) and \( h_2 \) be two heaps; we say that \( h_1 \) and \( h_2 \) are disjoint, noted \( h_1 \perp h_2 \), if \( \text{dom}(h_1) \cap \text{dom}(h_2) = \emptyset \). The operation \( h_1 \ast h_2 \) is defined for disjoint heaps as the disjoint union of the two partial functions.

Semantics of formulae is defined by the satisfaction relation \( \models_{SL} \) (see Table 1).

Formulae \( A \) of SL are called state formulae. The size of the state formula \( A \), written \(|A|\), is the length of the string \( A \) for some reasonably succinct encoding of variables and integers (binary representation). We will use the map \(|.|\) for other syntactic objects such as LTL mem formulae. A formula \( A \ast B \) with the separation conjunction states that \( A \) holds on some portion of the memory heap and \( B \) holds on a disjoint portion. A formula \( A \rightarrow B \) states that the current heap, when extended with any disjoint heap verifying \( A \), will verify \( B \). Boolean operators are understood as usual. In the remainder, we focus on several specific fragments of this separation logic. We say that a formula is in the record fragment (RF) if all subformulae \( x + i \leftarrow e \) use \( i = 0 \). In that case, we write \( x \leftarrow e \). We say that a formula is in the classical fragment (CL) if it does not use the connectives \( *, \rightarrow \). Finally, we say that a formula is in the list fragment (LF) if it is in the classical fragment and all subformulae \( x + i \leftarrow e \) use \( i = 0 \) and \( l = \text{next} \), and we may simply write \( x \leftarrow e \). Clearly, the classical and record fragments are incomparable, while the list fragment is included in both of them.

Let us illustrate the expressive power of SL on examples. The formula \( \neg \text{emp} \ast \neg \text{emp} \) means that at least two memory cells are allocated. The formula \( x \leftarrow e \),
defined as \( \neg (\neg \text{emp} \land \neg \text{emp}) \land x \xrightarrow{t} e \), is the local version of \( x \xrightarrow{t} e \) : \( s, h \models_{\text{SL}} x \xrightarrow{t} e \) iff \( \text{dom}(h) = \{s(x)\} \) and \( h(s(x))(l) = [e] \). The formula \( x \xrightarrow{t} \text{null} \xrightarrow{\bot} \) is satisfied at \( (s_0, h_0) \) whenever there is no heap \( h_1 \) with \( h_1 \perp h_0 \) that allocates the variable \( x \) to \text{null} on \( l \) field, that is \( x \) is allocated in \( h_0 \).

\( A \) is valid iff for every memory state \( (s, h) \), we have \( (s, h) \models_{\text{SL}} A \) (written \( \models_{\text{SL}} A \)). Satisfiability is defined dually.

**Proposition 1.** The model-checking, satisfiability and validity problems for SL are \( \text{PSPACE-complete} \).

PSPACE-hardness results are consequences of [CYO01, Sect. 5.2]. The PSPACE upper bound for model-checking for SL is obtained by reduction to model-checking for RF that is shown in PSPACE thanks to forthcoming Lemma 2. Satisfiability for SL is reduced to model-checking for SL thanks to a small memory state property: every satisfiable state formulae \( A \) can be satisfied by a memory state that can be encoded in polynomial size in \( A \).

### 2.2 Temporal extension

**Memory states sequences** Models of the logic \( \text{LTL}^{\text{mem}} \) are \( \omega \)-sequences of memory states, that is elements in \( (S \times H)^\omega \) and they are understood as infinite computations of programs with pointer variables. In order to analyze computations from programs without destructive update, we shall also consider models with constant heap, that is elements in \( S^\omega \times H \).

The logic \( \text{LTL}^{\text{mem}} \). Formulae of \( \text{LTL}^{\text{mem}} \) are defined in Table 2. Atomic formulae of \( \text{LTL}^{\text{mem}} \) are state formulae from SL except that variables can be prefixed by the symbol “\( X \)”. For instance, \( Xx \) is interpreted by the value of \( x \) at the next memory state. We use the notation \( \underbrace{X \cdots X}_t x \) for \( \underbrace{X \cdots X}_t x \) (but keep in mind that encoding \( X^t x \) requires memory space in \( O(i) \)). The temporal operators are the standard next-time operator \( X \) and until operator \( U \) present in LTL, see e.g. [SCS85]. The satisfaction relation \( \rho, t \models \phi \) where \( \rho \) is a model of \( \text{LTL}^{\text{mem}} \), \( t \in \mathbb{N} \) and \( \phi \) is a formula is also defined in Table 2. We use standard abbreviations such as \( F\phi, G\phi \ldots \). We freely use propositional variables \( p, q \), having in mind that the propositional variable \( p \) should be understood as \( x_p = x \uparrow \) for some fixed extra variables \( x_p, x_q, \ldots, x_{\uparrow} \).

Given a fragment \( \text{Frag} \) of SL, \( \text{LTL}^{\text{mem}}(\text{Frag}) \) is the restriction of \( \text{LTL}^{\text{mem}} \) to formulae in which occur only state formulae built over \( \text{Frag} \) (with extended variables \( X'x \)), and we write \( \text{SAT}(\text{Frag}) \) to denote the satisfiability problem for \( \text{LTL}^{\text{mem}}(\text{Frag}) \): given a temporal formula \( \phi \) in \( \text{LTL}^{\text{mem}}(\text{Frag}) \), is there a model \( \rho \) such that \( \rho, 0 \models \phi ? \) The variant problem in which we require that the model has a constant heap [resp. that the initial memory state is fixed, say \( (s, h) \)] is denoted by \( \text{SAT}^c(\text{Frag}) \) [resp. \( \text{SAT}^c_{\text{init}}(\text{Frag}) \)]. The problem \( \text{SAT}^{c}_{\text{init}}(\text{Frag}) \) is defined analogously.
Enriched expressions  \( \eta ::= x \mid X\eta \mid \text{null} \)

Atomic formulae  \( \pi ::= \eta = \eta' \mid \eta + i \lhd \eta' \)

State formulae  \( \mathcal{A} ::= \pi \mid \text{emp} \mid \mathcal{A} \ast \mathcal{B} \mid \mathcal{A} \rightarrow \mathcal{B} \mid \mathcal{A} \land \mathcal{B} \mid \mathcal{A} \rightarrow \mathcal{B} \mid \bot \)

Temporal formulae  \( \phi ::= \mathcal{A} \mid X\phi \mid \phi \lor \phi' \mid \neg \phi \)

Semantics

\( \rho, t \models X\phi \) if \( \rho, t + 1 \models \phi \).

\( \rho, t \models \phi \lor \psi \) if \( \rho, t \models \phi \) and \( \rho, t \models \psi \).

\( \rho, t \models \neg \phi \) if \( \rho, t \not\models \phi \).

\( \rho, t \models \mathcal{A} \) if \( s'_t, h_t \models_{\text{SL}} \mathcal{A}[X^k x \leftarrow (x, k)] \) where \( \rho = (s_t, h_t)_{t \geq 0} \) and \( s'_t \) is defined by \( s'_t((x, k)) = s_{t+k}(x) \).

**Table 2.** The syntax and semantics of LTL\(^{\text{mem}}\)

2.3 Programs with pointer variables

In this section, we define the model-checking problems for programs with pointer variables over LTL\(^{\text{mem}}\) specifications. The set \( I \) of instructions used in the programs is defined by the grammar below:

\[
\text{instr} ::= x := y \mid \text{skip} \\
\hspace{1cm} | x := y \rightarrow l \mid x := y \rightarrow l := y \mid x := \text{cons}(l_1 : x_1, \ldots, l_k : x_k) \mid \text{free } x \\
\hspace{1cm} | x := y[i] \mid x[i] := y \mid x = \text{malloc}(i) \mid \text{free } x, i
\]

The denotational semantics of an instruction \( \text{instr} \) is defined as a partial function \( \llbracket \text{instr} \rrbracket : \mathcal{S} \times \mathcal{H} \rightarrow \mathcal{S} \times \mathcal{H} \), undefined when the instruction would cause a memory violation. We list in Table 3 the formal denotational semantics of our instruction set. Boolean combinations of equalities between expressions are called guards and its set is denoted by \( G \). A program is defined as a triple \( (Q, \delta, q_I) \) such that \( Q \) is a finite set of control states, \( q_I \) is the initial state and \( \delta \) is the transition relation, a subset of \( Q \times G \times I \times Q \). We use \( q \xrightarrow{g, \text{instr}} q' \) to denote a transition. We say that a program is without destructive update if transitions are labeled only with instructions of the form \( x := y, x := y \rightarrow l, \) and \( x := y[i] \). We write \( P \) to denote the set of programs and \( P^{\text{ct}} \) to denote the set of programs without destructive update.

A program is a finite object whose interpretation can be viewed as an infinite-state system. More precisely, given a program \( p = (Q, \delta, q_I) \), the transition system \( \mathcal{S}_p = (S, \rightarrow) \) is defined as follows: \( S = Q \times (\mathcal{S} \times \mathcal{H}) \) (set of configurations) and \( (q, (s, h)) \rightarrow (q', (s', h')) \) if there is a transition \( q \xrightarrow{g, \text{instr}} q' \in \delta \) such that \( (s, h) = g \) and \( (s', h') = \llbracket \text{instr} \rrbracket (s, h) \). Note that \( \mathcal{S}_p \) is not necessarily linear. A computation (or execution) of \( p \) is defined as an infinite path in \( \mathcal{S}_p \) starting with control state \( q_I \). Computations of \( p \) can be viewed as LTL\(^{\text{mem}}\) models, using
\[ \begin{align*}
\text{[ } x := y \text{ ] } (s, h) & \quad \overset{\text{def}}{=} (s[x \mapsto s(y)], h). \\
\text{[ } x := \text{l } \rightarrow \text{l } \text{ ] } (s, h \ast \{i \mapsto \{l \mapsto v, \ldots\}\}) & \quad \overset{\text{def}}{=} (s[x \mapsto v], h \ast \{i \mapsto \{l \mapsto v, \ldots\}\}) \\
\text{with } s(y) = i \\
\text{[ } x \rightarrow \text{l } := y \text{ ] } (s, h \ast \{i \mapsto \{l \mapsto v, \ldots\}\}) & \quad \overset{\text{def}}{=} (s, h \ast \{i \mapsto \{l \mapsto s(y), \ldots\}\}) \\
\text{with } s(x) = i \\
\text{[ } x := \text{cons}(l_1 : x_1, \ldots, l_k : x_k) \text{ ] } (s, h) & \quad \overset{\text{def}}{=} (s[x \mapsto i], h \ast \{i \mapsto \{l_1 \mapsto s(x_1), \ldots, l_k \mapsto s(x_k)\}\}) \\
& \quad \text{with } i \not\in \text{dom}(h) \\
\text{[ free } x, l \text{ ] } (s, h \ast \{i \mapsto \{l \mapsto v, \ldots\}\}) & \quad \overset{\text{def}}{=} (s, h \ast \{i \mapsto \{\ldots\}\}) \\
& \quad \text{with } s(x) = i \\
\text{[ skip ] } (s, h) & \quad \overset{\text{def}}{=} (s, h) \\
\text{[ } x := y[i] \text{ ] } (s, h \ast \{i + i' \mapsto \{\text{next } \mapsto v\}\}) & \quad \overset{\text{def}}{=} (s[x \mapsto v], h \ast \{i \mapsto \{\text{next } \mapsto v\}\}) \\
& \quad \text{with } s(y) = i' \\
\text{[ } x[i] := y \text{ ] } (s, h \ast \{i' + i \mapsto \{\text{next } \mapsto v\}\}) & \quad \overset{\text{def}}{=} (s, h \ast \{i + i' \mapsto \{\text{next } \mapsto s(y)\}\}) \\
& \quad \text{with } s(x) = i' \\
\text{[ } x := \text{malloc}(i) \text{ ] } (s, h) & \quad \overset{\text{def}}{=} (s[x \mapsto i'], h \ast \{i' \mapsto \{\text{next } \mapsto \text{nil}\}\}) \\
& \quad \ast \ldots \ast \{i' + i \mapsto \{\text{next } \mapsto \text{nil}\}\} \\
& \quad \text{with } i', \ldots, i' + i \not\in \text{dom}(h) \\
\text{[ free } x, i \text{ ] } (s, h \ast \{i' + i \mapsto f\}) & \quad \overset{\text{def}}{=} (s, h) \text{ with } s(x) = i'
\end{align*} \]

Table 3. Semantics for instructions
constant heap. Intuitively, we express the control of the program with propositional variables, and define a formula that encode the transitions. To do so, we translate instructions of the form $x := y$ into $Xx = y$, $x := y \rightarrow l$ into $y \leftarrow Xx$, and $x := y[i]$ into $y + i \leftarrow Xx$. Guards are translated accordingly. As a consequence, the following result can be derived:

**Lemma 1.** Let $\text{Frag}$ be a fragment among $\text{SL}$, $\text{CL}$, $\text{RF}$, or $\text{LF}$. There is a logspace reduction from $\text{MC}^{ct}(\text{Frag})$ to $\text{SAT}^{ct}(\text{Frag})$ (resp. from $\text{MC}^{ct}_{init}(\text{Frag})$ to $\text{SAT}^{ct}_{init}(\text{Frag})$).

3 Decidable Satisfiability Problems by Abstracting Computations

In this section we establish the PSPACE-completeness of the problems $\text{SAT}(\text{CL})$ and $\text{SAT}(\text{RF})$. To do so, we abstract memory states whose size is a priori unbounded by finite symbolic memory states. As usual, temporal infinity in models is handled by Büchi automata recognizing $\omega$-sequences. We propose below an abstraction that is correct for CL (allowing pointer arithmetic) and for RF (allowing all operators from Separation Logic) taken separately but that is not exact for the full language SL.

3.1 Syntactic measures

The main approach to get decision procedures to verify infinite-state systems consists in introducing a symbolic representation for infinite sets of configurations. The symbolic representation defined below plays a similar role and has similarities with symbolic heaps for Separation Logic in Smallfoot [BCO05]. Let us start by some useful definitions. Following [Loz04], we introduce the set of *test formulae* that are formulae from SL of the forms below:

- $\text{alloc } x \equiv f(x \stackrel{\text{next}}{\to} \text{null}) \land \bot$ ($x$ is allocated).
- $\text{size} \geq k \equiv f \stackrel{k}{\text{times}} \neg \text{emp} \land \neg \text{emp}$ (at least $k$ indices are allocated).
- $e + i \leftarrow e, e = e'$.

Given a formula $\phi$ of $\text{LTL}^{mem}$, we define its measure $\mu_{\phi}$ understood as some pieces of information about the syntactic resources involved in $\phi$. Indeed, forthcoming symbolic states are finite objects parameterized by such syntactic measures.

For a state formula $\mathcal{A}$ of $\text{LTL}^{mem}$, the size of memory examined by $\mathcal{A}$, written $w_{\mathcal{A}}$, is inductively defined as follows: $w_{\mathcal{A}}$ is $1$ for atomic formulae, $\max\{w_{\mathcal{A}_1}, w_{\mathcal{A}_2}\}$ for $\mathcal{A}_1 \land \mathcal{A}_2$ or $\mathcal{A}_1 \rightarrow \mathcal{A}_2$ or $\mathcal{A}_1 = \mathcal{A}_2$, and $w_{\mathcal{A}_1} + w_{\mathcal{A}_2}$ for $\mathcal{A}_1 \ast \mathcal{A}_2$. Observe that $w_{\mathcal{A}} \leq |\mathcal{A}|$. Other simple sets about the syntactic resources of $\mathcal{A}$ need to be defined; $\text{Lab}_{\mathcal{A}}$ is the set of labels from Lab occurring in $\mathcal{A}$, $\text{Var}_{\mathcal{A}}$
is the set of variables from Var occurring in $A$, $\epsilon_A$ is the set of natural numbers $i$ such that $e + i \rightarrow e'$ occurs in $A$ and $m_A$ is the maximal $k$ such that $X^k x$ occurs in $A$ for some variable $x$. A measure is defined as an element of $\mathbb{N} \times P_f(\mathbb{N}) \times \mathbb{N} \times P_f(\text{Lab}) \times P_f(\text{Var})$ where $P_f(X)$ denotes the set of finite subsets of some set $X$. The set of measures has a natural lattice structure for the pointwise order, noted below $\mu \leq \mu'$. We also write $\mu[w \leftarrow 0]$ to denote the measure $\mu$ except that $w = 0$.

The measure for $A$, written $\mu_A$, is the tuple $(m_A, \epsilon_A, w_A, \text{Lab}_A, \text{Var}_A)$. The measure of some formula $\phi$ of $\text{LTL}^{\text{mem}}$, written $\mu_\phi$, is $\sup\{\mu_A : A \text{ occurs in } \phi\}$.

**Definition 1.** Given a measure $\mu = (m, \epsilon, w, X, Y)$, we write $T_\mu$ to denote the finite set of test formulae $\psi$ of the grammar:

$$
e := \langle x, u \rangle \mid \text{null} \quad f := e + i \quad \psi := f \rightarrow e \mid \text{alloc } f \mid e = e' \mid \text{size } \geq k$$

with $u \leq m$, $i \in \epsilon$, $l \in X$, $k < w$ and $x \in Y$.

Observe that the cardinal of $T_{\mu_{\phi}}$ is polynomial in $|\phi|$. Given a measure $\mu = (m, \epsilon, w, X, Y)$ and a memory state $(s, h)$, we write $\text{Abs}_\mu(s, h) = \{A \in T_\mu : (s, h) \models_{\text{SL}} A\}$ to denote the abstraction of $(s, h)$ wrt $\mu$. Given a measure $\mu$ and two memory states $(s, h)$ and $(s', h')$, we write $(s, h) \simeq_\mu (s', h')$ if $\text{Abs}_\mu(s, h) = \text{Abs}_\mu(s', h')$, that is, formulae in $T_\mu$ cannot distinguish the two memory states.

Lemma 2 below states that our abstraction is correct for CL and RF.

**Lemma 2.** Let $(s, h)$ and $(s', h')$ be two memory states such that $(s, h) \simeq_\mu (s', h')$ [resp. $(s, h) \simeq_\mu (s', h') \mid w - 0$] if any state formula $A$ such that $\mu_A \leq \mu$ and $A$ belongs to $T_{\mu_{\phi}}$ [resp. $T_{\mu_{\phi}} \mid w - 0$], we have $(s, h) \models_{\text{SL}} A$ iff $(s', h') \models_{\text{SL}} A$.

Note that we can extend this result to the whole SL by considering test formulae of the form $e + i = e' + j$.

### 3.2 Symbolic models

We write $\Sigma_\mu$ to denote the powerset of $T_\mu$; $\Sigma_\mu$ is thought as an alphabet, and elements $a \in \Sigma_\mu$ are called letters. A symbolic model wrt $\mu$ is defined as an infinite sequence $\sigma \in \Sigma_\mu^\omega$. Symbolic models are abstractions of models from $\text{LTL}^{\text{mem}}$: given a model $\rho : \mathbb{N} \rightarrow S \times H$ and a measure $\mu$, we write $\text{Abs}_\mu(\rho) : \mathbb{N} \rightarrow \Sigma_\mu$ to denote the symbolic model wrt $\mu$ such that for any $t$, $\text{Abs}_\mu(\rho)(\langle t \rangle_{\sigma}) = \{A \in T_\mu : \rho, t \models A \mid A(\langle x, u \rangle \leftarrow X^w x)\}$.

To a letter $a$, we associate the formula $A_a = \bigwedge_{A \in [a]} A \land \bigwedge_{A \notin [a]} \neg A$. For $\sigma$ a symbolic model, and $\phi$ a formula such that $\mu_\phi \leq \mu$, we define the symbolic satisfaction relation $\sigma, t \models_\mu \phi$ as satisfaction for models except for the clause about atomic subformulae that becomes; $\sigma, t \models_\mu A$ iff $\models_{\text{SL}} A_{\sigma(t)} \Rightarrow A[X^w x \leftarrow (x, u)]$.

We write $L^\mu(\phi)$ to denote the set of symbolic models $\sigma$ wrt $\mu$ such that $\sigma, 0 \models_\mu \phi$. As a corollary of Lemma 2, we have a soundness result for our abstraction;
Proposition 2. Let φ be a formula of LTL_{mem}^w(RF) [resp. of LTL_{mem}^w(CL)] and μφ ≤ μ. For any model ρ, we have that ρ ⊨ φ iff Abs_μ(ρ) ⊨ φ [resp. Abs_μ([w − 0])(ρ) ⊨ φ].

Note that Abs_μ is not surjective; we note L^ω sat the set of symbolic models wrt μ that are abstractions of some model of LTL_{mem}. Consequently, φ in LTL_{mem}^w(RF) is satisfiable iff L^ω sat(φ) ∩ L^ω sat is nonempty.

3.3 ω-regularity and PSPACE upper bound

In order to show that SAT(RF) and SAT(CL) are in PSPACE we shall explain why testing the nonemptiness of L^ω sat(φ) ∩ L^ω sat can be done in PSPACE. Below we treat explicitly the case for RF. For CL, replace every occurrence of μφ by μφ[w − 0].

To do so, we show that each language can be recognized by an exponential-size Büchi automaton satisfying the good properties to establish the PSPACE upper bound. If A is a Büchi automaton, we note L(A) the language recognized by A. Following [WV94,DDO07], let A be the generalized Büchi automaton defined by the structure (Σ, Q, δ, I, F) s.t.:

- Q is the set of so-called atoms of φ, that are sets of temporal formulae included in the so-called closure set cl(φ) (see [VW94]), I = {X ∈ Q : φ ∈ X}, and Σ = Σ_μ.
- X ↛ Y iff 1. for every atomic formula A of X, |=_SL A_μ ⇒ A[X^u x ← ⟨x, u⟩].
- 2. for every Xφ' ∈ cl(φ), Xφ' ∈ X iff φ' ∈ Y.
- Let {φ_1 Uφ'_{i+1},...,φ_n Uφ'_{i+n}} be the set of until formulae in cl(φ). We pose F = {F_1,...,F_n} where F_i = {X ∈ Q : φ_i Uφ'_{i} \not∈ X or φ'_i \not∈ X} for i ∈ {1,...,n}.

Let A^μ_φ be the Büchi automaton equivalent to the generalized Büchi automaton A. It is easy to observe that A^μ_φ has an exponential amount of states in the size of φ and its transition relation can be checked in polynomial space in the size of φ. Moreover,

Lemma 3. Let φ in LTL_{mem}^w(RF) [resp. LTL_{mem}^w(CL)] and μ ≥ μφ [resp. μ ≥ μφ[w − 0]]. Then, L(A^μ_φ) = L^μ(φ).

We can also build a Büchi automaton A^μ sat such that L(A^μ sat) = L^μ sat. A^μ sat is defined as (Σ, Q, δ, I, F), where Σ = Σ_μ, Q = Σ_μ, F = I = Q and a a' a'' iff:

1. A_μ, A_μ a are satisfiable, and a = a',
2. for every formula ⟨x, u⟩ = ⟨x', u'⟩ ∈ T_μ with u, u' ≥ 1, ⟨x, u⟩ = ⟨x', u'⟩ ∈ a iff ⟨x, u − 1⟩ = ⟨x', u' − 1⟩ ∈ a''.

If μ = μφ, then A^μ sat is of exponential-size in |φ| and the transition relation can be checked in polynomial space in |φ|. More importantly, this automaton recognizes satisfiable symbolic models.

Lemma 4. Let φ in LTL_{mem}^w(RF) [resp. LTL_{mem}^w(CL)] and μ = μφ [resp. μ = μφ[w − 0]]. Then, L(A^μ sat) = L^μ sat.
This lemma is essential and it is not possible to extend it to the whole logic LTL\textsuperscript{mem} even by allowing test formulae of the form $x + i = y + j$ since we would need automata with counters. Now, we can state our main complexity result.

**Theorem 1.** SAT(RF) and SAT(CL) are PSPACE-complete.

**Proof.** (sketch) The lower bound is from LTL [EC85]. Let $\phi$ be an instance formula of SAT(RF) (for SAT(CL) replace below $\mu_\phi$ by $\mu_\phi[w \leftarrow 0]$). As seen earlier, $\phi$ is satisfiable iff $L^{\mu_\phi}(\phi) \cap L^{\mu_\phi}_{sat}$ is nonempty. Hence, $\phi$ is satisfiable iff $L(A^{\mu_\phi}_\phi) \cap L(A^{\mu_\phi}_{sat}) \neq \emptyset$. The intersection automaton is of exponential size in the size of $\phi$ and can be checked nonempty by a nondeterministic on-the-fly algorithm. Since nonemptiness problem for Büchi automata is NLOGSPACE-complete and the transition relation in the intersection automaton can be checked in polynomial space in $|\phi|$, we obtain a nondeterministic polynomial space algorithm for testing satisfiability of $\phi$. By Savitch’s theorem, we get the PSPACE upper bound. \hfill \square

### 3.4 Other problems in PSPACE

Let Frag be either the classical fragment or the record fragment. Lemma 1 provides a reduction from MC\textsuperscript{ct} \textsuperscript{init}(Frag) to SAT\textsuperscript{ct} \textsuperscript{init}(Frag) based on a program-as-formula encoding. As we will see now, we may also reduce SAT\textsuperscript{ct} \textsuperscript{init}(Frag) to SAT(Frag) internalizing an approximation of the initial memory state whose logical language cannot distinguish from the initial memory state. As a consequence, the PSPACE upper bound for SAT(Frag) entails the PSPACE upper bound for both SAT\textsuperscript{ct} \textsuperscript{init}(Frag) and MC\textsuperscript{ct} \textsuperscript{init}(Frag).

**Proposition 3.** The problems SAT\textsuperscript{ct} \textsuperscript{init}(RF), MC\textsuperscript{ct} \textsuperscript{init}(RF), SAT\textsuperscript{ct} \textsuperscript{init}(CL) and MC\textsuperscript{ct} \textsuperscript{init}(CL) are PSPACE-complete.

**Proof.** By Lemma 1 and since SAT\textsuperscript{ct} \textsuperscript{init}(RF) is known to be PSPACE-hard, it remains to establish the PSPACE upper bound for SAT\textsuperscript{ct} \textsuperscript{init}(RF).

Given a formula $\phi$ and an initial memory state $(s,h)$, we shall build in polynomial-time a formula $\phi_{s,h}^{ct}$ in SAT(RF) such that $\phi$ is satisfiable in a model with initial memory state $(s,h)$ and constant heap $\phi_{s,h}^{ct}$ is satisfiable by a general model. Since we have shown that SAT(RF) is in PSPACE, this guarantees that SAT\textsuperscript{ct} \textsuperscript{init}(RF) is in PSPACE. The idea of the proof is to internalize the initial memory state and the fact that the heap is constant in the logic SAT(RF). Actually, one cannot exactly express that the heap is constant (see details below) but the approximation we use will be sufficient for our purpose.

Apart from the variables of $\phi$, the formula $\phi_{s,h}^{ct}$ is built over additional variables in $V = \{x_i : i \in \text{dom}(h) \cup \text{Im}(s)\} \cup \{x_{i,l} : i \in \text{dom}(h), l \in \text{dom}(h(i))\}$. The formula $\phi_{s,h}^{ct}$ of the form $G(\psi_1 \land \psi_2 \land \psi_3) \land \psi_4 \land \psi_5$, where the subformulae are defined as follows.

- $\psi_1$ states that the heap is almost equal to $h$ since we cannot forbid in the logical language additional labels (\text{dom}(h) = \{i_1, \ldots, i_k\}): 
  
  $$\psi_1 = \left( \bigwedge_{l \in \text{dom}(h(i_1))} x_{i_1} \leftrightarrow x_{i_1,l} \right) \land \ldots \land \left( \bigwedge_{l \in \text{dom}(h(i_k))} x_{i_k} \leftrightarrow x_{i_1,l} \right).$$

11
$\psi_2$ states which variables are equal and which ones are not, depending on
the initial memory state. By way of example, for $i \neq j \in \text{dom}(h)$, a conjunct
of $\psi$ is $x_i \neq x_j$. Similarly, if $h(i)(l) = j$ and $j \in \text{dom}(h)$ then $x_{i,l} = x_j$ is a
conjunct of $\psi$. Details are omitted.

$\psi_3$ states that the auxiliary variables remain constant: $\bigwedge_{x \in V} x = Xx$.

The formula $\psi'$ is obtained from $\phi$ by replacing each occurrence of $x \mapsto e$ by

$$x \mapsto e \land \bigwedge_{i \in \text{dom}(h), j \notin \text{dom}(h(i))} x \neq x_i.$$

The additional conjunct is useful because our logical language cannot state
that a label is not in the domain of some allocated address.

$\psi_s$ states constraints about the initial store $s$: $\psi_s \triangleq \bigwedge_{x \in \phi} x = x_s(x)$.

It is then easy to check that $\phi$ is satisfiable in a model with initial memory
state $(s, h)$ with constant heap if $\phi^{s,h}_{ct}$ is satisfiable by a general model.

As far as the results for the classical fragment are concerned, by Lemma 1, there is a logspace reduction from $\text{MC}^{ct}_{\text{init}}(\text{CL})$ to $\text{SAT}^{ct}_{\text{init}}(\text{CL})$ and as done above
one can reduce $\text{SAT}^{ct}_{\text{init}}(\text{CL})$ to $\text{SAT}(\text{CL})$. \hfill \Box

4 Undecidability Results

In this section, we show several undecidability results by using reduction from
problems for Minsky machines. So, first, we recall that a Minsky machine $M$
consists of two counters $C_1$ and $C_2$, and a sequence of $n \geq 1$ instructions, each
of which may increment or decrement one of the counters, or jump conditionally
upon the counters being zero. The $l$th instruction ($l$ is its location counter) has one of the following forms either “$l$: $C_i := C_i + 1$ ; goto $l''$ or “$l$: if $C_i = 0$
then goto $l'$ else $C_i := C_i - 1$; goto $l''$”. In a nondeterministic machine, after an
incrementation or a decrementation a nondeterministic choice of the form “goto
$l_1$ or goto $l_2$” is performed.

The configurations of $M$ are triples $(l, c_1, c_2)$, where $1 \leq l \leq n$, $c_1 \geq 0$, and
$c_2 \geq 0$ are the current values of the location counter and the two counters $C_1$
and $C_2$, respectively. The consecution relation on configurations is defined in
the obvious way. A computation of $M$ is a sequence of related configurations,
starting with the initial configuration $(1, 0, 0)$.

Different encodings of counters are used here. For instance, in [BFLS06],
a counter $C$ with value $n$ is represented by a list of length $n$ pointed to by a $x$
dedicated to $C$. The same idea is used in the proof of Proposition 4 below. In order
to show undecidability of $\text{SAT}(\text{SL})$ we alternatively encode counters by relying
on pointer arithmetic and properties of heaps. Programs without destructive
updates can simulate finite computations of Minsky machines by guessing at the
start of the computation the maximal value of counters (encoded by a list of the
length of the maximal value). As a consequence,
Proposition 4. SAT\textsuperscript{ct}(LF) and MC\textsuperscript{ct}(LF) are Σ\textsubscript{1}\textsuperscript{0}-complete.

By contrast, programs with destructive update can work with unbounded heaps, and by using the representation of counters as above, they can faithfully simulate a Minsky machine even if an empty heap is the initial heap. Because LTL can express repeated accessibility, Σ\textsubscript{1}\textsuperscript{1}-hardness can be obtained.

Proposition 5. The problems MC(LF) and MC\textsubscript{init}(LF) are Σ\textsubscript{1}\textsuperscript{1}-complete.

Let us briefly explain how to encode incrementation and decrementation with separating connectives and pointer arithmetic. Observe that expressions of the form \(x = y + 1\) are not allowed in the logical language. We solve this point in two different ways: using non-aliasing expressed by the separating conjunction, and using the precise pointing assertion \(x \xrightarrow{\text{next}} \eta\) stating that the heap contains only one cell, in conjunction with the \(\rightarrow\) operator.

\[
\begin{align*}
\phi^*_{x++} &= (Xx \xrightarrow{\text{next}} \text{null} \land x + 1 \xrightarrow{\text{next}} \text{null}) \land \neg(Xx \xrightarrow{\text{next}} \text{null} \ast x + 1 \xrightarrow{\text{next}} \text{null}) \\
\phi^*_{x--} &= (Xx + 1 \xrightarrow{\text{next}} \text{null} \land x \xrightarrow{\text{next}} \text{null}) \land \neg(Xx + 1 \xrightarrow{\text{next}} \text{null} \ast x \xrightarrow{\text{next}} \text{null}) \\
\phi^*_{x-} &= \text{emp} \\n\phi^*_{x+} &= \text{emp} \land ((Xx \xrightarrow{\text{next}} \text{null}) \rightarrow x + 1 \xrightarrow{\text{next}} \text{null}) \\
\phi^*_{x-} &= \text{emp} \land ((x \xrightarrow{\text{next}} \text{null}) \rightarrow Xx + 1 \xrightarrow{\text{next}} \text{null})
\end{align*}
\]

The formulae based on the separating conjunction correctly express incrementation and decrementation when the cells at index \(x, x + 1, x - 1\) are allocated, whereas formulae based on the operator \(\rightarrow\) do not need the same assumption.

Let SAT\textsuperscript{\Sigma\textsubscript{1}\textsuperscript{1}+}(SL) be any satisfiability problem among the four variants.

Proposition 6. SAT\textsuperscript{\Sigma\textsubscript{1}\textsuperscript{1}+}(SL) is Σ\textsubscript{1}\textsuperscript{1}-complete.

Proof. We reduce the recurrence problem for nondeterministic Minsky machines [AH94] to SAT\textsuperscript{\Sigma\textsubscript{1}\textsuperscript{1}+}(SL). Let \(\phi_0\) be the formula \(G(\text{emp} \land \bigwedge_{t=1}^k(x_t \neq \text{null})\). Incrementation and decrementation are performed thanks to \(\phi^*_{x++}\) and \(\phi^*_{x--}\), respectively. For any model \(\rho\) such that \(\rho, 0 \models \phi_0\), and for any \(t\), we have \(\rho, t \models \phi^*_{x++}\) iff \(s_t(x_i) + 1 = s_{t+1}(x_i)\). Hence, we have a means to encode incrementation. Similarly, \(\rho, t \models \phi^*_{x--}\) and \(s_t(x_i) > 0\) iff \(s_t(x_i) - 1 = s_{t+1}(x_i)\).

The fact that a counter does not change is encoded by \(x_i \neq \text{null}\). Given that \(\phi_1 = G(x_{\text{zero}} = Xx_{\text{zero}} \land x_{\text{zero}} \neq \text{null})\) holds, zero tests are encoded by \(x_i = x_{\text{zero}}\).

Given a nondeterministic Minsky machine \(M\), we write \(\psi_l\) to denote the formula encoding instruction \(l\). For instance for the instruction “l: if \(C_1 = 0\) then goto \(l'\) else \(C_1 := C_1 - 1\); goto \(l'\) or goto \(l_2\)” \(\psi_l\) is equal to \(G((l \land x_1 \neq x_{\text{zero}}) \Rightarrow (x_2 = Xx_2 \land (Xl_1' \lor Xl_2') \land \phi^*_{x--})) \land G((l \land x_1 = x_{\text{zero}}) \Rightarrow (x_1 = Xx_1 \land x_2 = Xx_2 \land Xl'))\). Hence, \((x_1 = x_2 = x_{\text{zero}}) \land \phi_0 \land \phi_1 \land \bigwedge_t \psi_l \land \text{GF}\) is satisfiable iff \(M\) has a computation with location counter \(n\) repeated infinitely often.

Proposition 7. The problem SAT(SL \setminus \{\rightarrow\}) is Σ\textsubscript{1}\textsuperscript{1}+ complete.

The proof of Proposition 7 is similar to the proof of Theorem 6 except that incrementation and decrementation are performed with the formulae \(\phi^*_{x++}\) and \(\phi^*_{x--}\), respectively.
5 Conclusion

In the paper, we have introduced a temporal logic LTL\textsuperscript{mem} for which assertion language is quantifier-free separation logic. Figure 1 contains a summary of the complexity results about satisfiability and model-checking problems for the fragments LF, CL and RF. $\Sigma_1^0$-completeness results for SAT\textsuperscript{cl}(SL), SAT(SL \setminus \{\rightarrow\}) and MC(LF) can be found in Propositions 6, 7, and 5, respectively. A thin and straight [resp. bold and curved] arrow between a source problem and a target problem means that the upper [resp. lower] bound for the target problem is shown thanks to the upper [resp. lower] bound for the source problem.

![Complexity diagram](https://example.com/diagram.png)

Fig. 1. Complexity of reasoning tasks with LTL\textsuperscript{mem}

Finally, extending LTL\textsuperscript{mem} with a special propositional variable heap$^=\mathbin{\,}$stating that the current heap is equal to the next one, can lead to undecidability (look at the problems of the form SAT$^\mathsf{cl}_{\mathsf{init}}$(Frag)). However, it is open whether satisfiability becomes decidable if we restrict the interplay between the “until” operator $\mathsf{U}$ and heap$^=\mathbin{\,}$, for instance to forbid subformulae of the form $\mathsf{G}\text{heap}^=\mathbin{\,}$ with positive polarity.

References


S. Bardin, A. Finkel, E. Lozes, and A. Sangnier. From pointer systems to counter systems using shape analysis. 5th International Workshop on Automated Verification of Infinite-State Systems (AVIS'06), 2006.


