HyperLDLf: a Logic for Checking Properties of Finite Traces Process Logs

Giuseppe De Giacomo, Paolo Felli, Marco Montali, Giuseppe Perelli

1 Sapienza University of Rome
2 Free University of Bozen-Bolzano
degiacomo@diag.uniroma1.it, {montali,pfelli}@inf.unibz.it, perelli@diag.uniroma1.it

Abstract
Temporal logics over finite traces, such as LTLf and its extension LDLf, have been adopted in several areas, including Business Process Management (BPM), to check properties of processes whose executions have an unbounded, but finite, length. These logics express properties of single traces in isolation, however, especially in BPM it is also of interest to express properties over the entire log, i.e., properties that relate multiple traces of the log at once. In the case of infinite-traces, HyperLTL has been proposed to express these “hyper” properties. In this paper, motivated by BPM, we introduce HyperLDLf, a logic that extends LDLf with the hyper features of HyperLTL. We provide a sound, complete, and computationally optimal technique, based on DFAs manipulation, for the model checking problem in the relevant case where the set of traces (i.e., the log) is a regular language. We illustrate how this form of model checking can be used to specify and verify sophisticated properties within BPM and process mining.

1 Introduction
Temporal logics traditionally assume that the traces generated by the dynamic system of interest have infinite length. This does not match a variety of different settings where traces have an unbounded, but finite, length, such as planning [Baier and McIraith, 2006; Baier et al., 2007; Gerevini et al., 2009] and Business Process Management (BPM) [Dumas et al., 2018; van der Aalst and others, 2011].

To specify and reason about temporal properties of finite traces, De Giacomo and Vardi (2013) studied two finite-trace variants of Linear Temporal Logic and Linear Dynamic Logic, namely LTLf and LDLf. These logics turned out to be particularly important in BPM: LTLf has become the formal basis for one of the most studied declarative process modeling notations [Pesic et al., 2007], while LDLf has been employed to realize advanced forms of process monitoring and runtime verification [De Giacomo et al., 2014]. Notably, these logics are not only interesting from the semantical point of view, but also bring an operational advantage over their infinite-trace counterparts, since reasoning can be carried out by directly manipulating finite-state automata, without appealing to automata over infinite objects [De Giacomo and Vardi, 2013; De Giacomo et al., 2014].

LTLf and LDLf express properties of single traces in isolation. However, especially in BPM it is also of interest to express properties that relate multiple traces at once (see, e.g., [van der Aalst et al., 2004; Weidlich et al., 2011; Rinderle-Ma et al., 2016]). Two specific settings arise in this spectrum. In the first setting, such traces belong to the overall, possibly infinite log of traces representing all the valid executions of a given process model. In this case, traces are implicitly represented in the form of a finite-state transition system or a bounded Petri net interpreted under interleaving semantics, which guarantee that the induced log forms a regular language. These logs not only capture any finite concrete process log, but also capture logs that can be inferred through sampling by model-learning and process mining techniques [Vaandrager, 2017; van der Aalst, 2011]. In the second setting, which is the typical one in business process management [van der Aalst and others, 2011], traces belong to a finite event log explicitly storing a set of observed process executions, representing a factual sample of all the traces that belong to an unknown process, which again is assumed to be regular.

In the case of infinite-traces, HyperLTL has been proposed to express “hyper” properties that quantify over multiple traces at once [Clarkson et al., 2014]. In this paper, motivated by BPM we introduce HyperLDLf and its fragment HyperLDLf: two extensions of LDLf and LTLf, respectively, that have the ability to express and verify hyper-properties of finite traces.

To handle the analysis of (finite and infinite) logs, we in particular focus on the very significant task of model checking: the task of verifying HyperLDLf formulas over a regular log, i.e., a (possibly infinite) set of traces forming a regular language. We study the HyperLDLf model checking problem in the key setting where the regular log is compactly described by a deterministic finite automaton. Specifically, we show that model-checking a HyperLDLf formula with k alternations of trace quantifiers is k-EXPTIME-complete. Notably, we show the upper bound by presenting a sound and complete algorithm based on manipulation of regular automata that retains the good, practical computational characteristics already exploited in several reasoning tasks for LTLf/LDLf. Finally, we go back to our original motivation
and discuss how HyperLTL$_J$ and HyperLDL$_J$ can be used to specify and verify relevant, sophisticated properties within BPM and process mining.

2 Automata and Languages

We recall some basic notions on alternating, nondeterministic, universal, and deterministic finite automata. All these automata capture regular languages but they have different characteristics that we will exploited later [Davis et al., 1994]. For a given set $X$, $B^+(X)$ is the set of positive Boolean formulas over $X$, that is Boolean formulas built from elements of $X$ using $\land$ and $\lor$, where we also allow true and false. A subset $Y \subseteq X$ satisfies a Boolean formula $\chi \in B^+(X)$ if the truth assignment assigning true to the elements in $Y$ and false to the elements in $X \setminus Y$ satisfies $\chi$. Given an alphabet $\Sigma$, a finite word over it is a finite sequence $w = \sigma_0 \cdot \sigma_1 \ldots \sigma_n$ of letters in $\Sigma$. By $|w| = n + 1$ we denote the length of $w$. Moreover, by $\Sigma^*$ we denote the set of all possible finite words over $\Sigma$. An alternating finite automaton (AFA) on finite words is a tuple $A = (\Sigma, Q, q_0, \delta, F)$ where $\Sigma$ is the (finite) input alphabet, $Q$ is a finite set of states with $q_0 \in Q$ being initial state, $\delta : Q \times \Sigma \to B^+(Q)$ is a transition function and $F \subseteq Q$ is a set of final states. Intuitively, the transition function $\delta(q, \sigma)$ describes the possible configurations that $A$ can reach when reading the symbol $\sigma$ while being in state $q$. The automaton $A$ is nondeterministic (NFA), denoted $N$, or universal (UFA), denoted $U$, if $\delta(q, \sigma)$ contains only disjunctions or conjunctions, respectively. Finally, it is deterministic (DFA), denoted $D$, if the transition function maps each pair $(q, \sigma)$ to a single state $q'$.

The concepts of run, accepting run, and accepting word for an automaton $A$ are quite standard and we refer to [Leiss, 1985; Davis et al., 1994] for their formal definition. Let $L(A)$ be the set of words accepted by the automaton $A$. Recall that every automaton $A$ can be turned into an equivalent nondeterministic one $N$, that is, such that $L(A) = L(N)$. Such operation is called nondeterminization and by $\text{Ndet}(A)$ we denote the nondeterministic equivalent automaton. Analogously, we can turn an automaton $A$ into a universal one, denoted $\text{Univ}(A)$. Such operation is called universalization. Moreover, recall that if $A$ is either alternating or universal, $\text{Ndet}(A)$ can be of size exponential w.r.t. $|A|$. Analogously, if $A$ is either alternating or nondeterministic, $\text{Univ}(A)$ can be of size exponential w.r.t. $|A|$ [Chandra et al., 1981; Davis et al., 1994].

3 HyperLDL$_J$ over Finite Traces

We start by defining the hyper-version of LDL$_J$. We refer to [De Giacomo and Vardi, 2013] for details on LDL$_J$ itself.

Definition 1 (HyperLDL$_J$ syntax). For a set $A$ containing atomic propositions $p$ and a set $V$ of trace variables $\pi$, the syntax of HyperLDL$_J$ is the following:

\[
\begin{align*}
\varphi &::= \psi \mid \exists \pi \varphi \mid \forall \pi \varphi \\
\psi &::= tt \mid ff \mid \neg \psi \mid \psi \land \psi \mid \psi \lor \psi \mid (\rho) \psi \mid [\rho] \psi \\
p &::= \phi \mid \psi ? \mid p + \rho \mid p \land \rho \mid p^* \\
\phi &::= p_\pi \mid \neg p_\pi \mid \phi \land \phi \mid \phi \lor \phi
\end{align*}
\]

Observe that the atomic propositions occurring in a HyperLDL$_J$ formula are typed over the trace (variable) on which they are evaluated. This is also the case for the propositional true and false predicate. In particular, we have true$_\pi \equiv p_\pi \lor \neg p_\pi$ and false$_\pi \equiv \neg$true$_\pi$ for every trace variable occurring in a HyperLDL$_J$ formula. Analogously, by true$_P \equiv \bigwedge_{\pi \in P}$true$_\pi$ and false$_P \equiv \neg$true$_P$ we denote the propositional true and false over a set of traces. For convenience we use the abbreviation finished$_\pi \equiv [\text{true}_\pi]t$ to specify that a trace $\pi$ is finished, and last$_\pi \equiv (\text{true}_\pi)\text{finished}_\pi$ to specify the last element of the trace.

A log $\text{Log} \subseteq (2^{A^P})^*$ is a (possibly infinite) subset of traces over the powerset of AP. Note that an element $P$ of $2^{A^P}$ can be regarded also as a truth-assignment of AP, where every element $P$ in $\text{Log}$ takes the value true and every element not in $P$ takes the value false. For a trace $t \in \text{Log}$, $|t|$ denotes the length of $t$. By $t[i]$ we denote the $i$-th element of $t$, with $t[0]$ and $t[|t|-1]$ being the first and last elements, respectively. For every $i, j \in \mathbb{N}$ with $i \leq j$, by $t[i,j]$ we denote the subtrace obtained from $t$ by taking the elements form $t[i]$ to $t[j]-1$. For the case of $|t| \leq i$, the trace $t[i,j] = e$ is the empty trace. For the case $i < |t| \leq j$, the trace $t[i,j] = t[i,|t|-1]$ is the suffix of $t$ starting from the $i$-th element. We also denote the suffix of $t$ from element $i$ with $t[i,\ldots]$. A trace assignment (or simply assignment) $\Pi : V \to \text{Log}$ is a function mapping trace variables $V$ to concrete traces in Log. By $\Pi[\pi \leftarrow t]$ we denote the trace assignment that adheres with $\Pi$, except that $\Pi[\pi \leftarrow t(\pi)] = t$, that is, $t$ is assigned to $\pi$. By $\Pi[i,j]$, for every $i, j \in \mathbb{N}$, we denote the projection of assignment $\Pi$ over the subtraces between $i$ and $j$, i.e., such that $\Pi[i,j](\pi) = \Pi(\pi)[i,j]$ for every $\pi \in V$. By $\Pi[i,\ldots]$, for every $i \in \mathbb{N}$, we denote the projection of $\Pi$ over the suffix from $i$, i.e., such that $\Pi[i,\ldots](\pi) = \Pi(\pi)[i,\ldots]$ for each $\pi \in V$.

The semantics of a HyperLDL$_J$ formula over a set Log of traces, a trace assignment $\Pi$, and a natural number $n \in \mathbb{N}$ is given as follows:

- **Log.** $\Pi[i] = \exists \pi \varphi$ if there exists $t \in \text{Log}$ such that $\Pi[\pi \leftarrow t]$ such that $\varphi$.
- **Log.** $\Pi[i] = \forall \pi \varphi$ if for each $t \in \text{Log}$ it holds that $\Pi[\pi \leftarrow t]$ such that $\varphi$.
- **Log.** $\Pi[i] = tt$.
- **Log.** $\Pi[i] \equiv ff$.
- **Log.** $\Pi[i] \equiv \neg \psi$ if $\text{Log.} \Pi[i] \equiv \psi$.
- **Log.** $\Pi[i] = \psi_1 \land \psi_2$ if both $\text{Log.} \Pi[i] \equiv \psi_1$ and $\text{Log.} \Pi[i] \equiv \psi_2$.
- **Log.** $\Pi[i] = \psi_1 \lor \psi_2$ if either $\text{Log.} \Pi[i] \equiv \psi_1$ or $\text{Log.} \Pi[i] \equiv \psi_2$.
- **Log.** $\Pi[i] \equiv (\rho) \psi$ if there exists $j \geq i$ such that $(i,j) \in R_\Pi(\rho)$ and $\Pi[j] \equiv \psi$.
- **Log.** $\Pi[i] \equiv [\rho] \psi$ if for each $j \geq i$ such that $(i,j) \in R_\Pi(\rho)$ it holds that $\Pi[j] \equiv \psi$.

where the relation $R_\Pi(\rho)$ contains all pairs of indexes $(i,j)$ so that $\Pi(\pi)[i,j]$ conforms to $\rho$ for each $\pi$ that appears as an index of a proposition in $\rho$, defined as follows:

- $(i,j) \in R_\Pi(p_\pi)$ if $j = i + 1$, $\Pi(\pi)[i,j] \neq \epsilon$, and $p \in \Pi(\pi)[i]$;
- $(i,j) \in R_\Pi(-p_\pi)$ if $j = i + 1$, $\Pi(\pi)[i,j] \neq \epsilon$, and $p \notin \Pi(\pi)[i]$;
- \((i, j) \in R_H(\psi_1 \land \psi_2)\) if \((i, j) \in R_H(\psi_1)\) and \((i, j) \in R_H(\psi_2)\);
- \((i, j) \in R_H(\psi_1 \lor \psi_2)\) if \((i, j) \in R_H(\psi_1)\) or \((i, j) \in R_H(\psi_2)\);
- \((i, j) \in R_H(\psi?)\) if \(j = i\) and \(\log, \Pi, j \models \psi\);
- \((i, j) \in R_H(\rho_1 + \rho_2)\) if either \((i, j) \in R_H(\rho_1)\) or \((i, j) \in R_H(\rho_2)\);
- \((i, j) \in R_H(\rho_1; \rho_2)\) if there exists \(k \in [i, j]\) such that \((i, k) \in R_H(\rho_1)\) and \((k, j) \in R_H(\rho_2)\);
- \((i, j) \in R_H(\rho^*)\) if \(j = i\) or there exists \(k \in [i+1, j]\) such that \((i, k) \in R_H(\rho)\) and \((k, j) \in R_H(\rho^*)\).

By \(\log, \Pi \models \varphi\) we denote the fact that \(\log, \Pi, 0 \models \varphi\).

A variable \(\pi\) in \(\varphi\) is free if it occurs out of the scope of either \(\exists \pi\) or \(\forall \pi\). By \(\text{free}(\varphi)\) we denote the set of free variables in \(\varphi\). Formula \(\varphi\) is closed if \(\text{free}(\varphi) = \emptyset\). Observe that the semantics of a closed formula \(\varphi\) replaces all the traces in the assignment \(\Pi\), which then becomes irrelevant. Thus, we can denote by \(\log \models \varphi\) the fact that \(\log, \Pi, 0 \models \varphi\) for some \(\Pi\).

Note how the semantics allows to consider the evolution of multiple traces at once, in a synchronous fashion. Indeed, \(R_H(\rho)\) includes all the pairs \((i, j)\) so that the synchronous execution described by \(\Pi(\pi)[i, j]\) conforms to \(\rho\), for each \(\pi\) appearing in the regular expression. As a result, while evaluating formulas of the form \(\rho \psi\) or \(\rho \psi\), all the traces in \(\log\) that are assigned through \(\Pi\) to trace variables appearing in \(\rho\) are synchronously executed.

Similarly to HyperLDL, we can define the hyper-version of \(\text{LTDL}\) [De Giacomo and Vardi, 2013].

**Definition 2 (Hyper\(\text{LTDL}\) syntax).** For a set \(\text{AP}\) containing atomic propositions \(p\) and a set \(V\) of trace variables \(\pi\), the syntax of Hyper\(\text{LTDL}\) is the following:

\[
\varphi ::= \psi \mid \exists \pi \varphi \mid \forall \pi \varphi
\]

where

\[
\psi ::= p \mid \neg \psi \mid \psi \land \psi \mid \psi \lor \psi \mid X \psi \mid X \psi \mid X \psi \mid X \psi \mid F \psi \mid G \psi
\]

In fact, HyperLDL can express the classic linear-temporal logic operators over finite traces such as next \((X)\) and until \((U)\) and so to have Hyper\(\text{LTDL}\) defined as a syntactic fragment of Hyper\(\text{LDL}\). Indeed, the next and weak next operators can be encoded in Hyper\(\text{LDL}\) by \(X \psi = (\text{true}(\text{free}(\varphi))) \psi\) and \(\neg X \psi = \neg \text{next} \psi\), respectively. Note that the propositional true in the diamond operator refers only to the traces that are free in \(\psi\). That is, we do not impose any condition on the rest of the traces, which may or may not terminate (or may have terminated already) at the current execution step. Analogously, for the until operator, we have that \(\psi_1 \psi_2 = (\text{true}(\text{free}(\varphi))) \psi_1 \psi_2\). Note that in this case it is sufficient to impose the propositional true only on the traces occurring in \(\psi_2\) since, for those occurring in \(\psi_1\), this is automatically imposed by evaluating \(\psi_1\) at each step of the execution specified in the regular expression of the diamond operator. Clearly, we can define the finally and globally operators as usual as \(F \psi = \text{true}(\text{free}(\varphi)) U \psi\) and \(G \psi = \neg F \psi\). These shortcut definitions for the temporal operators automatically induce the semantics of Hyper\(\text{LTDL}\).

Note that, we can impose two traces to be of equal length by \(|\pi_1| = |\pi_2| = G(\text{last}_{\pi_1} \leftrightarrow \text{last}_{\pi_2})\). Analogously, we can say that \(\pi_1\) is not longer than \(\pi_2\) by \(|\pi_1| \leq |\pi_2| = (\neg \text{last}_{\pi_2}) U \text{last}_{\pi_1}\). Finally, we can also compare the content of traces. For example, we can say that \(\pi_1\) and \(\pi_2\) evaluate a subset \(L \subseteq \text{AP}\) in the same way: \(\text{equal}_L(\pi_1, \pi_2) = |\pi_1| = |\pi_2| \land \bigwedge_{p \in L} (p_{\pi_1} \leftrightarrow p_{\pi_2})\).

### 4 Model Checking HyperLDL

We are interested in model checking HyperLDL properties over a log. In particular we focus on a quite general class of logs: those that are regular, i.e., the set of traces in the log can be recognized by a deterministic finite automaton, which we call the Log DFA \(D\), i.e., \(\log = L(D)\). We denote \(D, \Pi, i \models \varphi\) the fact that \(L(D), \Pi, i \models \varphi\), and by \(D \models \varphi\) the fact that \(D, \Pi, 0 \models \varphi\) for some trace assignment \(\Pi\). Therefore, model checking in our context is defined as follows.

**Definition 3.** For a given Log DFA \(D\) and a closed HyperLDL formula \(\varphi\), the model-checking problem is the problem of deciding whether \(D \models \varphi\).

We solve the model-checking problem by reducing it to the emptiness problem of a suitably defined regular automaton. Consider a quantifier free HyperLDL formula \(\psi\) and let \(V = \{\pi_1, \ldots, \pi_n\}\) be free(\(\psi\)). Thus, \(\psi\) can be regarded as an LDL formula over \(\bigcup_{p \in \text{AP}} \Pi(p)\), where \(\Pi(p)\) denotes the alphabet LDL relativized with variable \(p\), i.e., \(\Pi(p) = \{p_\pi : p \in \text{AP}\}\). However, we need to consider its semantics in the context of HyperLDL, which provides, at once, traces of different length. In order to use the LDL automata construction, we restrict to trace assignments that map every variable in traces all of the same length. To do this, consider the function tail mapping a trace assignment \(\Pi\) to another \(\Pi(\pi)\) over the same set \(V\), in which each \(\Pi(\pi)\) is replaced with \(\text{tail}(\Pi(\pi)) = \Pi(\pi) \cdot \text{end}_\pi^{\text{length}(\Pi(\pi)) - \#(\pi) + 1}\). Intuitively, every trace in the codomain of \(\Pi\) is appended with a tail of symbols \(\text{end}_\pi\) in a way that each of them has the same length.

By \(\Theta_\pi = 2^\text{AP}\) we denote the set of possible truth-assignments over the set \(\text{AP}\) and by \(\Theta_\pi^{\text{end}} = \Theta_\pi \cup \{\text{end}_\pi\}\) the set of assignments augmented with the corresponding \(\text{end}_\pi\) symbol. Observe that the symbol \(\text{end}_\pi\) differs from the empty assignment, setting every propositional variable false, but rather mean that the trace is finished. Indeed, it will hold that both \(\text{end}_\pi^{\text{free}} \neq p_\pi^{\text{free}}\) and \(\text{end}_\pi \neq p_\pi^{\text{free}}\).

By \(\Theta_\pi^{\text{end}} = \bigcup_{p \in V} \Theta_\pi^{\text{end}}\) we denote the union set of assignments over the atomic propositions typed with \(\pi \in V\). Note that the sets of truth-assignments are pairwise disjoint. Therefore, a truth-assignment in \(\Theta_\pi^{\text{end}}\) can be uniquely identified with a list of assignments in \(\Theta_\pi^{\text{end}}\), one for each \(\pi \in V\). At this point, every trace assignment \(\Pi(\pi)\) of this form can be regarded as a single trace over the set \(\Theta_\pi^{\text{end}}\), to which a quantifier-free HyperLDL formula can be interpreted as an LDL formula. We have the following proposition.

**Proposition 1.** For a given quantifier-free HyperLDL formula \(\varphi\) and a trace assignment \(\Pi\), it holds that \(\Pi \models \varphi\) if and only if \(\text{tail}(\Pi) \models \varphi\).

Hence, without loss of generality, we assume the trace assignments to map traces of the same length and then to regard them as a single trace over \(\Theta_\pi^{\text{end}}\). This allows us to apply the...
Proposition 2. For every quantifier free HyperLDL$_f$ formula $\psi$ there exists an alternating finite automaton $A_\psi$ of size polynomial in $|\psi|$ such that, for every trace assignment $\Pi$, it holds that $\Pi \models \psi$ if and only if $\Pi \in \mathcal{L}(A_\psi)$.

Next we focus on regular logs. First, recall that the set Log of valid traces over AP is described in terms of a DFA $D = \langle \Sigma, Q, q_0, \delta, F \rangle$ with $\Sigma = 2^\mathcal{P}$, that is Log = $\mathcal{L}(D)$. In our model-checking technique, we may need to read traces that are tied with the end symbol a finite but unbounded number of times. This is done by amending the automaton and defining $D^\text{tail} = \langle \Sigma', Q', q_0, \delta', F' \rangle$, where $\Sigma' = \Sigma \cup \{\text{end}\}, Q' = Q \cup \{\text{acc}, \text{rej}\}, F' = F \cup \{\text{acc}\}$ and $\delta'$ extends $\delta$ as follows:

\begin{itemize}
  \item $\delta'(q, \text{end}) = \begin{cases} \text{acc} & \text{if } q \in F \\
  \text{rej} & \text{otherwise} \end{cases}$
  \item $\delta'(\text{acc}, \sigma) = \begin{cases} \text{acc} & \text{if } \sigma = \text{end} \\
  \text{rej} & \text{otherwise} \end{cases}$
  \item $\delta'(\text{rej}, \sigma) = \text{rej}$.
\end{itemize}

Intuitively, $D^\text{tail}$ is as $D$ until the end of trace is reached. From that point on, it accepts only if both the trace is accepted by $D$ and the rest is made by an arbitrarily long sequence of end symbols. The following trivially holds.

Proposition 3. For every DFA $D$, the automaton $D^\text{tail}$ is such that $\mathcal{L}(D^\text{tail}) = \{w; \text{end}^n : w \in \mathcal{L}(D) \text{ and } n \in \mathbb{N}\}$.

Finally we deal with the quantifiers of HyperLDL$_f$. To do so we make use of the classic notions of existential and universal projection (see e.g. [David E. Muller and Paul E. Schupp, 1995]).

Definition 4 (Automata projection). For a given automaton $A = \langle \Sigma, Q, q_0, \delta, F \rangle$ such that $\Sigma = 2^X$ and $P \subseteq X$:

- If $A$ is nondeterministic, then the existential projection of $A$ over $P$ is the automaton $A^\exists_P = \langle \Sigma, Q, q_0, \delta^\exists_P, F \rangle$ with $\delta^\exists_P(q, \sigma) = \bigvee_{p \in \mathcal{P}} \delta(q, \sigma_X \cap P \cup \sigma_P)$.
- If $A$ is universal, then the universal projection of $A$ over $P$ is the automaton $A^\forall_P = \langle \Sigma, Q, q_0, \delta^\forall_P, F \rangle$ where $\delta^\forall_P(q, \sigma) = \bigwedge_{p \in \mathcal{P}} \delta(q, \sigma_X \cap P \cup \sigma_P)$.

Observe that, by using the Ndet and Univ, we can compute the existential and universal projections of any kind of automaton $A$. However this might require an exponential blow-up in its size, as nondeterminization and universalization are involved. For simplicity, we use the notation $A^\exists_P$, in place of $(\text{Ndet}(A))^\exists_P$, to denote the existential projection for every automaton. Analogously, $A^\forall_P$ denotes the universal projection for every automaton. Intuitively, the existential projection $A^\exists_P$ accepts words $w$ for which there exists a replacement of the variables in $P$ that make the resulting word $w'$ accepted by $A$. The universal projection, instead, $A^\forall_P$ accepts words $w$ for which every possible replacement of the variable evaluation in $P$ is a word $w'$ accepted by $A$. This is formalized in the following theorem which adapts a key result in [David E. Muller and Paul E. Schupp, 1995] for infinite words to finite ones.

Theorem 1. For a given automaton $A$ over the alphabet $2^X$ and a subset $P \subseteq X$, the two following propositions hold:

1. For every word $w \in (2^X)^*$, $w \in \mathcal{L}(A^\exists_P)$ iff there exists a word $w'$ s.t. $|w'| = |w|$, $w'(i)_X \cap P = w(i)_X \cap P$ for every $i < |w|$, and $w' \in \mathcal{L}(A)$.
2. For every word $w \in (2^X)^*$, $w \in \mathcal{L}(A^\forall_P)$ iff for every word $w'$ such that $|w'| = |w|$ and $w'(i)_X \cap P = w(i)_X \cap P$ for every $i < |w|$, it holds that $w' \in \mathcal{L}(A)$.

For a given DFA $D$ reading words over an alphabet $\Sigma = 2^X$, the $\pi$-typification of $D$, denoted $D_\pi$, is obtained by renaming every proposition $p$ in AP to $p_\pi$. By $\mathcal{A}_{\pi} = \{p_\pi : p \in \mathcal{A}\}$ we denote the set of typed variables from AP.

We are now ready to present our algorithm that, given a Log DFA $D$ and a HyperLDL$_f$ formula $\varphi$, constructs an automaton $\text{Automaton}(D, \varphi)$ whose nonemptiness problem is equivalent to check whether $D \models \varphi$. We first define Algorithm 1 for generating the automaton $\text{Automaton}(\cdot, \cdot)$.

Algorithm 1 Automata construction Automa$\text{ton}(\cdot, \cdot)$. 

\begin{algorithm}
\caption{Automa$\text{ton}(\cdot, \cdot)$}
\textbf{Input:} HyperLDL$_f$ formula $\varphi$ and Log DFA $D$.
\textbf{Output:} The automaton $\text{Automa}$t$\text{ton}(D, \varphi)$.

\begin{algorithmic}
    \State \textbf{if} $\varphi$ is quantifier free \textbf{then}
    \State \hspace{1em} return $\mathcal{A}_{\pi} \cap \bigcap_{\pi\in \text{free}(\varphi)} D^\text{tail}_{\pi}$
    \State \textbf{if} $\exists \pi \varphi$ \textbf{then}
    \State \hspace{1em} return (Automaton($D, \varphi'$))$^\exists_{\mathcal{A}_{\pi}}$
    \State \textbf{if} $\forall \pi \varphi$ \textbf{then}
    \State \hspace{1em} return (Automaton($D, \varphi'$))$^\forall_{\mathcal{A}_{\pi}}$
\end{algorithmic}
\end{algorithm}

Algorithm 1 recursively scans the formula on a top-to-bottom fashion, handling its quantifiers one by one. At each recursive call, it removes the outermost quantifier of $\varphi$ and computes the automata projection that corresponds to the quantification modality. For an existential quantification it applies an existential projection, whereas for a universal quantification, it applies a universal one. Algorithm 1 makes use of a polynomial operation: the typification of the alphabet.

Theorem 2. For every HyperLDL$_f$ formula $\varphi$ and a Log DFA $D$ it holds that $\mathcal{L}(\text{Automaton}(D, \varphi)) = \{\Pi : V \to \mathcal{L}(D) : D, \Pi, 0 \models \varphi\}$.

Proof sketch. By induction on the structure of $\varphi$, the base case handles the quantifier free formulas, for which the algorithm returns the DFA construction of $\mathcal{A}_{\varphi}$ intersected with a suitable number of copies of $D^\text{tail}_{\pi}$ to make sure the selected traces are in the log. The inductive cases handle the quantifiers by means of automata projection. The correctness of such procedure follows from Theorem 1.

We can now solve model-checking of $D$ and $\varphi$ by checking the nonemptiness of $\text{Automaton}(D, \varphi)$, see Algorithm 2.

The correctness of Algorithm 2 is a direct consequence of Theorem 2. Algorithm 2 gives us an upper-bound on the computational complexity of the problem.
Theorem 3. The model checking problem for a HyperLDLₐ formula φ with quantifier alternation depth k over a Log DFA D can be solved in k-EXPSPACE in both φ and D.

Proof sketch. For quantified free formulas φ, Algorithm 2 checks the nonemptiness of a polynomial sized alternating automaton, whose complexity is PSPACE. For all the other cases, at each automaton projection, the size of the automaton increases by one exponential, thus requiring an extra exponent in space complexity to check for its nonemptiness.

This bound is tight since we prove a matching lower-bound for the model-checking of both HyperLDLₐ and HyperLTLₐ.

Theorem 4. The model checking problem for a HyperLTLₐ (and so HyperLDLₐ) formula φ with quantifier alternation depth k over a Log DFA D is k-EXPSPACE-hard.

Proof sketch. Through an encoding from Turing machines. The proof adapts for proving the lower-bound of the satisfiability problem for QPTL in [Sistla et al., 1987].

We observe that our technique is based on automata constructions involving projection (existential projection) and determinization and complementation (universal projection). For different operations are computationally dominated by determinization (projection is polynomial as is complementation after determinization). In fact the alternation of quantifiers is directly related to the number of required determinization steps (after projection). This is what makes the tower of exponential directly related to quantifier alternations. On the other hand, there is evidence that determinization in the case of regular nondeterministic automata, while worst case exponential, is often polynomial in practice for minimized DFAs [Tabakov and Vardi, 2005] and this often gives rise to scalable algorithms in spite of the worst case computational complexity [Zhu et al., 2017; Zhu et al., 2020; Tabajara and Vardi, 2020]. For this reason Algorithm 1 and 2 are indeed promising also in practice.

5 HyperLDLₐ and BPM

We illustrate how HyperLDLₐ can be used to capture a variety of interesting properties within Business Process Management (BPM). An operational process consists of a collection of activities that are executed in coordination [Dumas et al., 2018]. The process is instantiated multiple times, and each instance represents the execution of the process on a specific, so-called case object (an order, a claim, ...), moving it from the initial state of the process to one of its final states. This happens through the application of finitely many activities to the case object, in agreement with the ordering constraints imposed by the process control-flow. This setting perfectly matches HyperLDLₐ, thanks to the ability of the logic to express hyper-properties over the finite traces induced by multiple process instances.

We fix a set A of activities. Given a trace π over A, we assume that at each step, at most one activity is executed. The case where no activity is executed, which we represent for convenience using formula $\text{nop}_π = \bigwedge_{a \in A} \neg a^π$, represents an idle step in the execution. With this notions at hand, we concentrate on three relevant aspects in the BPM spectrum.

Process Mining is a family of techniques used to get insights about operational processes, and to improve processes based on event data obtained from actual executions [van der Aalst and others, 2011]. Process mining techniques range from the discovery of process models from event data to checking whether an event log conforms to a process model (either manually crafted or automatically discovered from data). To do so, they need to inspect and relate different traces that are contained in the event log or that can be generated by the process model (two notions that in our setting homogeneously map to the HyperLDLₐ notion of regular log). Specifically, several process mining techniques do not directly work with raw traces, but rather use all the traces at hand to establish behavioral relations between activities, in turn using them as a succinct, abstract representation.

As a representative example, we consider here the reference ordering relations introduced in one of the seminal works on process discovery [van der Aalst et al., 2004]. Such ordering relations capture whether, according to the log, two activities directly follow each other, are mutually exclusive, or concurrent. For example, a log $\text{Log}$ indicates that activities a and b directly follow each other, that is, are in a strict sequencing (written $a \rightarrow_{\text{Log}} b$ in [van der Aalst et al., 2004]), if: (i) there exists a trace in $\text{Log}$ where a is followed by b without any other activity in between; (ii) b in none of the traces from $\text{Log}$ the converse happens, that is, b is followed by a without any other activity in between. While these two properties separately quantify over traces, they have to holds simultaneously in $\text{Log}$ to infer that a and b are in a strict sequencing, hence calling for the hyper-features of HyperLDLₐ. Indeed, we can check whether $a \rightarrow_{\text{Log}} b$ by:

$$\text{Log} \models \exists π_1 \forall π_2 ((\text{true}_{π_1})^* ; a_{π_1} ; (\text{nop}_{π_1})^* ; b_{π_1}) \text{tt} \land [(\text{true}_{π_2})^* ; b_{π_2} ; (\text{nop}_{π_2})^* ; a_{π_2}) \text{ff}$$

The other ordering relations can be formalized analogously.

Instance-Spanning Constraints are behavioral constraints that span across multiple instances of a given process, and thus inherently require to related multiple traces at once. In [Rinderle-Ma et al., 2016], a multitude of instance-spanning constraints is collected from different application scenarios, from logistics to healthcare and security. Interestingly, previous formalization attempts in this spectrum either failed or required to combine all distinct traces in a single stream equipping events with trace identifiers [Rinderle-Ma et al., 2016], an approach that is not applicable when the (possibly infinite) log to be considered is generated by a process model. Differently from those attempts, thanks to its hyper-
features HyperLDL$_f$ has the direct ability to elegantly capture instance-spanning constraints.

Consider for example an activity propagation constraint dictating that whenever an activity occurs in a trace $\pi$ of the log, then $b$ has to occur later on in all traces $\pi'$ related to $\pi$. By assuming that the fact that $\pi$ is related to $\pi'$ is expressed through a domain-specific HyperLDL$_f$ formula $rel(\pi_1, \pi_2)$ having $\pi$ and $\pi'$ as free variables, we can formalize in HyperLDL$_f$ this activity propagation constraint as: $\forall \pi_1, \pi_2 \ rel(\pi_1, \pi_2) \rightarrow G(\alpha_{\pi_1} \rightarrow XF b_{\pi_2})$. This formula can be instantiated to capture that whenever a package shipped via truck is detected as contaminated, then all the packages traveling in the same truck need to be later inspected.

**Alternative behaviors in a process.** We finally consider the use of HyperLDL$_f$ to express sophisticated properties on alternative behaviors exhibited by a process that cannot be captured using branching-time logics. To illustrate this, we first partition the activities $A$ into observable activities $A_0$ (public activities executed by or visible to external actors), and unobservable activities $A_u$ (private activities). We want to express a variety of outcome property capturing that whenever the process can produce a trace $\pi_1$ satisfying a desired behavioral pattern (a regular expression $\rho$) and leading to a given outcome (a final, observable activity), then it is also able to produce a different trace $\pi_2$ that (i) mimics $\pi_1$ by exhibiting the same observable behavior, and (ii) satisfies the same desired pattern $\rho$ leading to an alternative outcome.

Intuitively, by “mimicking” we mean that two traces are equal once all unobservable activities are projected out. Additionally, observable activities are not required to happen at the same time: an observable activity happening in either trace is replicated by the other with an arbitrary delay by adding unobservable activities in between. Until this happens, however, no other observable activity can be executed in either trace. Note that the role of mimicking and of being mimicked can be exchanged among $\pi_1$ and $\pi_2$ at each observable activity: it can happen that an observable activity of $\pi_1$ may be mimicked by $\pi_2$ with a certain delay, and that the next observable activity of $\pi_2$ is then mimicked by $\pi_1$ with some other delay. This requirement cannot be captured by a simple branching-time formula but is indeed a hyper property that requires the ability to compare observable projections of traces.

This can be used, for example, to check that the process is so that whenever a customer properly follows a sequence of public steps required by a payment system, if the payment is ultimately rejected then there is an alternative execution, with the same public steps but with possibly different unobservable ones (in type or number), where the payment is accepted.

To formalize this, we assume that in each step at most one task is executed. Let $\rho_{\pi}$ be a regular expression over the set $A$ in $\pi$, and $a, b \in A_u$ respectively denote the actual and alternative outcomes. Then property can be captured as:

$$\forall \pi_1 \exists \pi_2 (\rho_{\pi_1} a \rightarrow ((\rho_{\pi_2}) b \land \text{mimics}(\pi_1, \pi_2)))$$

where \text{mimics}(\pi_1, \pi_2) formalizes that $\pi_2$ mimics $\pi_1$. Since $\pi_1$ and $\pi_2$ may have a different length, we write $\text{mimics}(\pi_1, \pi_2) = |\pi_1| \leq |\pi_2| \land \text{mimicsObs}(\pi_1, \pi_2) \lor |\pi_1| > |\pi_2| \land \text{mimicsObs}(\pi_2, \pi_1)$, and then formalize this notion of mimicking via \text{mimicsObs}(\pi_s, \pi_t), taking advantage from the fact that $\pi_s$ is of equal or shorter length than $\pi_t$. To do so, we define a support formula: given a set $T$ of trace variables, \( u_T = \bigwedge_{\pi \in T} \ (\neg \text{nop}_\pi \lor (\bigvee_{a \in A_u} \text{nop}_a)) \) captures an unobservable step jointly performed in all traces from $T$. Using this, we write $\varphi_{\text{skip}} = (\text{un}(\pi_s, \pi_t))^*$ to indicate an arbitrary sequence of possibly different unobservable steps simultaneously performed by $\pi_s$ and $\pi_t$. As a result, $\text{mimicsObs}(\pi_s, \pi_t) = (\varphi_{\text{skip}}; \text{nop}_{\pi_s} \land \text{nop}_{\pi_t} \land \text{un}(\pi_s, \pi_t))$. This formula can be instantiated to capture that whenever a package shipped via truck is detected as contaminated, then all the packages traveling in the same truck need to be later inspected.

6 Related Work on Infinite Traces

To deal with multiple traces at once, the specific extension of LTL, called HyperLTL, has been originally introduced in the infinite traces setting. HyperLTL supports explicit quantification over traces, as well as hyper-properties that inspect and relate multiple traces at once [Clarkson et al., 2014]. In [Clarkson et al., 2014], HyperLTL and HyperCTL* are introduced as extensions of LTL and CTL* to reason about a set of (infinite) traces all at once. A hyper extension of PDL, namely HyperPDL-$\Delta$, is considered in [Gutsfeld et al., 2020]. The model-checking and satisfiability problems for HyperLTL and these extensions have been investigated and solved [Clarkson et al., 2014, 2015; Mascle and Zimmermann, 2020]. Their solutions are based on automata on infinite objects (traces), and provide a similar complexity hierarchy as the one in Theorem 3. Our technique avoids the manipulation of automata on infinite objects, sidestepping the difficulties of complementation of such automata, in favor of standard automata for regular languages, which have better computational characteristics in practice, see the discussion at the end of Section 4.

7 Conclusion

We have introduced HyperLDL$_f$ a logic for expressing hyper properties over logs, which are of particular interest, e.g., in BPM. Specifically, we have proposed a technique for model-checking of regular logs that takes full advantage of the relationship between LTL$_f$/LDL$_f$ and regular automata. For this reason, such a technique promises to be easily implementable and effective in practice in spite of the high worst-case complexity of the problem. It is also of interest to study other problems such as synthesis of a log satisfying certain HyperLDL$_f$ properties, which is related to satisfiability, or synthesis of controller generating a log with desired HyperLDL$_f$ properties, cfr. [Pnueli and Rosner, 1989].

Acknowledgments

Research partially supported by the ERC Advanced Grant ‘WhiteMech’ (No. 834228) and by the EU ICT-48 2020 project ‘TAILOR’ (No. 952215).
References


