Backward reachability analysis for timed automata with data variables

Rebeka Farkas[1,2], Tamás Tóth[1⋆], Ákos Hajdu[1,2], and András Vörös[1,2]

[1] Budapest University of Technology and Economics,
Department of Measurement and Information Systems,
Budapest, Hungary
{farkasr,totht,hajdua,vori}@mit.bme.hu
[2] MTA-BME Lendület Cyber-Physical Systems Research Group

Abstract. Efficient techniques for reachability analysis of timed automata are zone-based methods that explore the reachable state space from the initial state, and SMT-based methods that perform backward search from the target states. It is also possible to perform backward exploration based on zones, but calculating predecessor states for systems with data variables is computationally expensive, prohibiting the successful application of this approach so far. In this paper we overcome this problem by combining the zone-based backward exploration algorithm with the weakest precondition operation for data variables. This combination allows us to handle diagonal constraints efficiently as opposed to zone-based forward search where most approaches require additional operations to ensure correctness. We demonstrate the applicability and compare the efficiency of the algorithm to existing forward exploration approaches by measurements performed on industrial case studies. We show that data variables can be handled by the weakest precondition operation but the large number of target states often prevents successful verification.

Keywords: Timed automata · Reachability analysis · Backward exploration · Weakest precondition.

1 Introduction

The ubiquity of real-time safety-critical systems makes it desirable to model and verify time-dependent behaviour. One of the most common formalisms for this purpose is the timed automaton, introduced by Alur and Dill [1] which extends the finite automaton with real-valued variables called *clock variables* that represent the elapse of time. This makes it a suitable formalism for modelling time-dependent behaviour of systems such as communication protocols [21] and logical circuits with propagation delay [17].

State reachability has a prominent role in verification. Most reachability analysis algorithms perform *forward search*: they check if the set of states reachable

⋆ This work was partially supported by Gedeon Richter's Talentum Foundation (Gyömrői út 19-21, 1103 Budapest, Hungary).

from the initial state contains the target state. *Backward search* can also be used for checking reachability – in this case the algorithm explores set of states from where the target state is reachable and checks if it contains the initial state [18]. The motivation behind backward search is that when the target state is unreachable, the state space explored from the initial state is disjunct from the one explored backwards from the target states and the latter might be smaller.

Reachability for timed automata is decidable, but the complexity is exponential in the number of clock variables [1]. The most efficient techniques for deciding reachability either rely on SMT solvers or abstract state space representations. The key idea of SMT-based techniques is to transform the system into a set of constraints that is satisfiable if the target state is reachable, and then give it to an SMT solver. Although this technique is efficient for other domains and there is an efficient first-order theory for the timed domain, called *difference logic* [9], in practice solver-based techniques resulted more efficient for other timed formalisms, such as the timeout automaton [11]. As for timed automata, the most efficient algorithms rely on the *zone* abstract domain [10] that is usually used for (forward) state space exploration [6].

While it is possible to perform zone-based backward state space exploration of timed automata, the forward approach is preferred. The general reason behind this is that in practice extensions of timed automata are used, such as *timed automata with data variables*, and it is expensive to calculate predecessor states for systems with data variables [8,6]. On the other hand, zone-based backward exploration of timed automata would be desired, because for a class of timed automata called *timed automata with diagonal constraints* backward exploration is the only zone-based method that doesn't introduce an additional exponential factor to the complexity [8,7].

In our research we developed an algorithm for backward reachability analysis of timed automata that combines zone abstraction for clock variables with the *weakest precondition* operation for data variables. Using the operation we could explore the state space of the data variables backwards. We also compared the applicability and the efficiency of zone-based backward and forward exploration by measurements performed on the *XtaBenchmarkSuite* presented in [12].

In this paper we present the algorithm we developed, and prove its correctness. We also present the measurements we have performed and evaluate the results. We show that data variables can be handled by the weakest precondition operation but the large number of target states often prevents successful verification. We propose some possible improvements and other ideas about the applicability of zone-based backward exploration algorithm for the verification of timed automata.

The paper is organized as follows. Section 2 presents the related work, Section 3 provides some theoretical background on the reachability analysis of timed automata, our algorithm is explained in Section 4, Section 5 presents and evaluates the performed measurements, and finally Section 6 concludes the paper.

2 Related work

To the best of our knowledge there is no zone-based backward reachability analysis algorithm for timed automata with data variables in the literature. On the other hand, many zone-based algorithms have been proposed for forward exploration [14] as well as SMT-based methods that perform backward search [19]. Backward exploration is also used for model checking CTL properties.

In [3] the core of the zone-based algorithm and the zone operations – including those we use for backward exploration – are defined. Over the years many optimizations of zone abstraction have been defined. An optimal zone-based abstraction – that is, the weakest abstraction that is still safe – has been presented in [15]. Zone-based techniques perform forward state space exploration in the state space. They explore an *abstract reachability graph* whose nodes contain a zone-based abstraction of the reachable states. In many cases, *lazy abstraction* is used [15].

Usually, SMT-based methods can operate both forwards and backwards, and they can be applied to systems with data variables as well as timed automata with diagonal constraints. However, despite the many advantages, in practice SMT-based methods are less efficient for timed automata, than zone-based techniques. One of the most efficient SMT-based methods for timed automata is presented in [19], where backward exploration is used, however, before exploration the timed automaton is translated to another formalism called *finite state machine with time*.

Backward exploration also has significance at CTL model checking. The tool Kronos [25] performs backward exploration for model checking timed automata against TCTL properties.

3 Preliminaries

This section presents the important aspects of modeling timed systems, the extensions of timed automata used in practice, as well as the basics of zone-based state space exploration. The weakest precondition operation is also defined.

3.1 Timed automata

Basic definitions Clock variables (or *clocks*, for short) are introduced to represent the elapse of time. The set of clock variables is denoted by $\mathcal{C}$.

The most commonly used type of predicates over clocks is *clock constraints*.

Definition 1. *A clock constraint is a conjunction of* atomic clock constraints*, that can either be* simple constraints *of the form* $x \sim n$ *or* diagonal constraints *of the form* $x - y \sim n$*, where* $x, y \in \mathcal{C}$*,* $\sim \in \{\leq, <, =, >, \geq\}$ *and* $n \in \mathbb{N}$*. The set of clock constraints is denoted by* $\mathcal{B}(\mathcal{C})$*.*

In other words clock constraints define upper and lower bounds on the values of clocks and the differences of clocks.

A *timed automaton* extends a finite automaton with clock variables.

Definition 2. *A timed automaton $\mathcal{A}$ is a tuple $\langle L, l_0, E\rangle$ where*

- *L is the set of locations (or* control states*),*
- *$l_0 \in L$ is the initial location,*
- *$E \in L \times \mathcal{B}(\mathcal{C}) \times 2^{\mathcal{C}} \times L$ is the set of edges. An edge is defined by the source location, the guard (represented by a clock constraint), the set of clocks to reset and the target location.*

Note: It is usual to allow *invariants* on the locations of the automaton. In this paper we omit them to keep the explanations simple. However, the presented algorithms can be adjusted to handle invariants.

In the literature many algorithms are only defined for a subclass of timed automata called *diagonal-free timed automata*.

Definition 3. *A diagonal-free timed automaton is a timed automaton where all appearing atomic clock constraints are simple constraints.*

It is possible to transform a timed automaton containing diagonal constraints to a diagonal-free timed automaton – i.e. the expressive power of diagonal-free timed automata is the same as that of timed automata –, however, the number of locations in the resulting automaton is exponential in the number of diagonal constraints [7].

Operational semantics The values of the variables at a given moment is described by a valuation.

Definition 4. *Let us denote by* $\mathrm{dom}(V)$ *the set of possible values for a set of variables V, i.e. the* domain *of V. A* valuation *is a function $v : V \to \mathrm{dom}(V)$ that assigns a value for each variable $v \in V$.*

Accordingly, a clock valuation is a function $v^c : \mathcal{C} \to \mathbb{R}_{\geq 0}$. In this paper we denote clock valuations by v^c. Valuations over other types of variables are denoted by v^d. We denote by $v \models c$ if valuation v satisfies a constraint c. The set of all valuations satisfying a constraint c is denoted by $[\![c]\!]$.

Clock variables are initialized to 0 – that is, initially $v_0^c(c) = 0$ for all $c \in \mathcal{C}$ –, and their value are constantly and steadily increasing (at the same pace for all clocks).

The only operation on clock valuations is the operation *reset*, denoted by $[C \leftarrow 0]v^c$ which set the value of a set $C \subseteq \mathcal{C}$ to 0, that is $[C \leftarrow 0]v^c(c) = 0$ for $c \in C$ and $[C \leftarrow 0]v^c(c) = v^c(c)$ otherwise. It is an instantaneous operation, after which the value of the clock will continue to increase.

The state of a timed automaton depends on the current location and the values of the clocks. Formally, A *state* of $\mathcal{A}$ is a pair $\langle l, v^c\rangle$ where $l \in L(\mathcal{A})$ is a location and v^c is the current valuation.

Two kinds of operations are defined. The state $\langle l, v^c\rangle$ has a *discrete transition* to $\langle l', v^c\prime\rangle$ if there is an edge $e = \langle l, g, r, l'\rangle \in E$ in the automaton such that $v^c \models g$ and $v^c\prime = [r \leftarrow 0]v^c$. The state $\langle l, v^c\rangle$ has a *time transition* to $\langle l, v^c\prime\rangle$ if $v^c\prime$ assigns $v^c(c) + d$ for some non-negative d to each $c \in \mathcal{C}$.

Definition 5. *A run of $\mathcal{A}$ is a finite sequence of consecutive transitions $R = \langle l_0, v_0^c\rangle \xrightarrow{t_0} \langle l_0, v_0^{c\prime}\rangle \xrightarrow{d_0} \langle l_1, v_1^c\rangle \xrightarrow{t_1} \langle l_1, v_1^{c\prime}\rangle \ldots \langle l_n, v_n^c\rangle \xrightarrow{t_n} \langle l_n, v_n^{c\prime}\rangle$ where for all $1 \leq i \leq n, t_i$ denotes a time transition and d_i denotes a discrete transition.*

Extensions of timed automata In practice the low descriptive power of the original timed automaton formalism makes it inefficient for modeling systems. Over the years many extensions of the formalism have been invented to overcome this issue. In this paper we are focusing on *timed automata with data variables*.

Data variables are variables whose value can only be changed by discrete transitions, e.g. variables of types *integer, bool*, etc. They can appear in constraints to enable transitions (*data guards*) and can be modified by transitions (*update*). However, clock variables are not allowed to appear in data guards or updates. The extended formalism can be formally defined as follows.

Definition 6. *A* timed automaton with data variables *is a tuple $\langle L, l_0, E, v_0^d\rangle$ where*

- *L is the set of locations,*
- *$l_0 \in L$ is the initial location,*
- *$E \in L \times \mathcal{B}(\mathcal{C}) \times \mathcal{P}_V \times 2^{\mathcal{C}} \times \mathcal{U}_V \times L$ is the set of edges, where $\mathcal{P}_V$ denotes the set of first order predicates over a set of data variables V and $U \in \mathcal{U}_V$ is a sequence of assignments (updates) of the form $v \leftarrow t$ where $v \in V$ and t is a term built from variables in V and function symbols interpreted over* $\mathrm{dom}(V)$.
- *v_0^d is the initial data valuation.*

  An edge is defined by the source location, the clock guard, the data guard, the set of clocks to reset, the updates, and the target location.

The possible types of data variables and the expressions appearing in data guards and updates may vary by model checker, however, for efficient verification it is important, that the expressive power of the extended formalism is the same as that of the original timed automaton. In many cases this is achieved by restricting the set of possible values to a finite set (e.g. an interval for integers) – this way, the model can be translated to a (simple) timed automaton by encoding the values of data variables to locations. In this paper we denote the set of general conditions on a set of variables V by $\mathrm{cond}(V)$.

The operational semantics are also modified. A state of the extended formalism $\mathcal{A}$ over a set of clock variables $\mathcal{C}$ and a set of data variables V can be described by a tuple $\langle l, v^d, v^c\rangle$ where $l \in L$ is the current location, $v^d : V \to \mathrm{dom}(V)$ is the current *data valuation* and $v^c : \mathcal{C} \to \mathbb{R}_{\geq 0}$ is the *clock valuation*.

While the operational semantics of time transitions are not affected by data variables, discrete transitions become more complex: data guards have to be satisfied and updates have to be executed. Let us denote by $U(v^d)$ the data valuation after executing the sequence of updates U on a system with the initial data valuation v^d. Formally, the state $\langle l, v^d, v^c\rangle$ has a discrete transition to $\langle l', v^{d\prime}, v^{c\prime}\rangle$ if there is an edge $e = \langle l, g, p, r, U, l'\rangle \in E$ in the automaton such that $v^c \models g$, $v^d \models p$, $v^{c\prime} = [r \leftarrow 0]v^c$ and $v^{d\prime} = U(v^d)$.

Another popular extension of timed automata allows the decomposition of the automaton to a *network* of automata. A network $(\mathcal{A}_1|\ldots|\mathcal{A}_n)$ is a parallel composition of a set of timed automata. Communication is possible by shared variables or handshake synchronization using *synchronization channels*. A network of timed automata can be transformed into a timed automaton by constructing the product of the automata in the network. For formal definition, the interested reader is referred to [2].

3.2 Verification of timed automata

In this section we present the basics of zone-based reachability analysis of timed automata. SMT-based verification is beyond the scope of this paper.

In case of (simple) timed automata the reachability problem can be formalized as follows.

Problem 1. **Input:** A timed automaton $\mathcal{A}$, and a location $l_{\text{trg}} \in L(\mathcal{A})$
Question: Is there a run $\langle l_0, v_0\rangle \rightarrow \langle l_{\text{trg}}, v_n\rangle$ for some v_n in $\mathcal{A}$?

Zone-based verification An introduction to the abstract domain *zone* can be found in [3], along with the operations necessary for verification and an efficient representation, called *difference bound matrix* (DBM).

Definition 7. *A* zone $Z = \{v^c \mid v^c \models g\}$ *is a set of clock valuations satisfying some clock constraint* g.

Many operations are defined on zones. In this paper the following notations are used:

- $[R \leftarrow 0]Z$ denotes the result of operation *reset*(R), $R \subseteq \mathcal{C}$ on a zone Z: $[R \leftarrow 0]Z = \{v^c \mid \exists v^{c\prime} \in Z : v^c = [R \leftarrow 0]v^{c\prime}\}$
- $[R \leftarrow 0]^{-1}Z, R \subseteq \mathcal{C}$ denotes the inverse of the reset operation: $[R \leftarrow 0]^{-1}Z = \{v^c \mid [R \leftarrow 0]v^c \in Z\}$
- $Z^{\uparrow}$ denotes the set of valuations reachable from Z by time transitions: $Z^{\uparrow} = \{v^c + t \mid v^c \in Z, t \in \mathbb{N}\}$, where $v^c + t = v^{c\prime}$ denotes a valuation where $v^{c\prime}(c) = v^c(c) + t$ for all $c \in \mathcal{C}$
- $Z^{\downarrow}$ denotes the set of nonnegative valuations from where Z is reachable by time transitions: $Z^{\downarrow} = \{v^c \mid \exists t \in \mathbb{N} : v^c + t \in Z\}$

The core of zone-based reachability analysis is to construct an *abstract reachability graph*, that contains a location and a zone (or an abstraction of it) in each node.

Definition 8. *A* zone graph *is a finite graph containing* $\langle l, Z\rangle$ *pairs as nodes, where* $l \in L$ *is a location of the automaton and* Z *is a zone.*

A node of the zone graph represents a set of states reachable from the initial state (or a set of states from where the target states are reachable, in case of

backward exploration). Edges between nodes correspond to discrete transitions through which those states are reachable.

The zone graph is constructed similarly to any kind of graph-based state space exploration: starting from the initial node n_0, new nodes are introduced based on the outgoing (or incoming) edges of the automaton with a corresponding edge. If all the states represented by a new node n are also represented by an existing node n' (i.e. n' *covers* n), then instead of adding n to the graph, the corresponding graph edge is introduced pointing to n'. The algorithm terminates when the states represented by some n include the target (or the initial) state (the state is reachable), or when the complete state space is explored and it does not contain the desired state (the state is unreachable).

In case of timed automata this approach can be adjusted as follows.

- The initial node can be formalized as $\langle l_0(\mathcal{A}), \{v_0^c\}^{\uparrow}\rangle$ in case of forward exploration (it contains the states reachable from the initial state by time transitions) and $\langle l_{\text{trg}}, [\![true]\!]\rangle$ in case of backward exploration.
- The postimage of a zone based on an edge can be calculated by determining the maximal subzone that satisfies guard and resetting the given clocks in the subzone. Time transitions are applied after the post (or pre) image calculation. (Formally, $\text{Post}_{g,R}(Z) = ([R \leftarrow 0](Z \cap [\![g]\!]))^{\uparrow}$ and accordingly $\text{Pre}_{g,R}(Z) = ([R \leftarrow 0]^{-1} Z \cap [\![g]\!])^{\downarrow}$.)
- Node $\langle l, Z\rangle$ is covered by $\langle l', Z'\rangle$ if $l = l'$ and $Z \subseteq Z'$.
- The state is reachable if a node $\langle l_{\text{trg}}, Z \neq \emptyset\rangle$ (or $\langle l_0, Z\rangle$ where $v_0 \in Z$, in case of backward exploration) is reached.

In case of backward exploration the graph explored with the presented approach is always finite, but this is not true for forward exploration, since using loop-edges it is possible to create a timed automaton with a run of ever-increasing clock valuations [3]. To ensure the termination of the algorithm, extrapolation was introduced, however, later it turned out that in case of diagonal constraints extrapolation may introduce false positive states to the state space [4].

Many approaches have been introduced to ensure correctness, including the operation *split* that splits the zone to extrapolate to smaller zones so that the extrapolation keeps the zone safe [3], and CEGAR-based methods where they explore the state space with the original overapproximating algorithm and if they find a spurious trace they refine the state space or the automaton [8], but the complexity of these methods is exponential in the number of diagonal constraints.

Due to the difficult nature of calculations with diagonal constraints, and the fact that in practice they are infrequently used, in the recent years the research has been focused on optimizing zone abstractions for forward exploration of diagonal-free timed automata.

Verification of extended timed automata The reachability problem can be adjusted for the extended formalisms. For instance, in case of timed automata with data variables the target states can also depend on the data valuation.

Problem 2. **Input:** A timed automaton with data variables $\mathcal{A}$, a location $l_{\text{trg}} \in L(\mathcal{A})$, and a first order predicate P_{trg} over the data variables of $\mathcal{A}$
Question: Is there a run from the initial state $\langle l_0, v_0^d, v_0^c \rangle$ to some $\langle l, v_n^d, v_n^c \rangle$ where $v_n^d \models P$?

In case of networks of timed automata, instead of a single control location, the state of the system depends on a *configuration* – that is, a set of locations containing the current location of each automaton. However, usually reachability criteria only determine the location for *some* of the automata (e.g. collision detection only requires two out of an arbitrary number of stations to be transmitting).

The extended reachability problems can be transformed to the original reachability problem over (simple) timed automata, however, in practice it is more efficient to apply the algorithms on timed automata to the extended formalisms. For example, in case of forward exploration calculating the data variables is straightforward based on the updates. On the other hand, in case of backward exploration there can be more than one predecessor states. Since it is complicated to calculate the possible previous values of data variables, in case of extended timed automata forward exploration is preferred [6,8,5].

3.3 Weakest precondition

In our research we used the weakest precondition operation to calculate the preimage of data variables. The weakest precondition can be defined many ways. Here, we define it over predicates and updates.

Definition 9. *Given a sequence of updates $U \in \mathcal{U}_V$ and a predicate $P \in \mathcal{P}_V$ describing a postcondition, the* weakest precondition*, denoted by $wp(P, U)$ is the predicate describing the weakest constraint on the initial state ensuring that the execution of U terminates in a state satisfying P:*

$$wp : \mathcal{P}_V \times \mathcal{U}_V \to \mathcal{P}_V \text{ such that } [\![wp(P,U)]\!] = \{v^d : V \to \text{dom}(V) \mid U(v^d) \models P\}.$$

The weakest precondition of a sequence of updates can be calculated by iterating backwards over the assignment statements and calculating the weakest precondition one by one. Formally if $U = [v_1 \leftarrow t_1; v_2 \leftarrow t_2]$, then $wp(P, U) = wp(wp(v_2 \leftarrow t_2, P), v_1 \leftarrow t_1)$.

The weakest precondition of a predicate $P \in \mathcal{P}_V$ based on an update $v \leftarrow t$ can be calculated by replacing all occurrences of v in P with t.

Example 1. Let $y < 5$ be the postcondition of the assignment $y \leftarrow x + 3$. Replacing y with $x + 3$ yields $x + 3 < 5$ that can be automatically simplified to $x < 2$. Therefore if $U(v^d) = v^d\prime$, where $v^d\prime(y) = v^d(x) + 3$ and $v^d\prime(v) = v^d(v), v \neq y$, then $wp(y < 5, U) = x < 2$. In other words in order to ensure $y < 5$ after assigning $x + 3$ to y, initially $x < 2$ must hold.

In practice the calculation of the weakest precondition is only efficient if the sequences of assignments can be kept small and the assignments simple. For more information on the weakest precondition operation the reader is referred to [16].

4 Backward exploration using weakest precondition

In this section we present our algorithm and demonstrate it on an example. We also provide a proof of its correctness.

4.1 Algorithm

Our approach starts from the target states and builds an abstract reachability graph where the nodes are of the form $\langle l, P, Z\rangle$ where l is a location, P is a predicate (over the data variables), and Z is a zone. For each incoming discrete transition the preimage Z' of Z is calculated as described in Section 3.2, and if $Z' \neq \emptyset$ the preimage of P is calculated using the weakest precondition operation.

The algorithm performs graph-based state space exploration as described in Section 3.2. Here we explain the operations specific to our algorithm.

Initialization The initial node can be formalized as $\langle l_{\text{trg}}, P_{\text{trg}}, [\![true]\!]\rangle$.

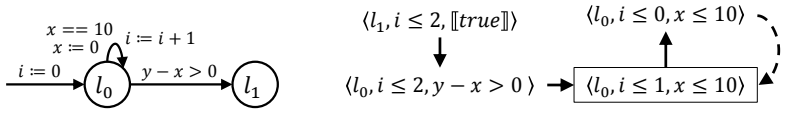
Preimage computation For a node $\langle l, P, Z\rangle$ and incoming edge $\langle l', g, p, r, U, l\rangle$ the preimage $\langle l', P', Z'\rangle$ is calculated as follows.

- The preimage of the location is the source location l' of the edge.
- The preimage of the zone can be calculated as presented in Section 3.2: $Z' = \text{Pre}_{g,R}(Z)$.
- P' is only calculated if $Z' \neq \emptyset$. In this case $P' = wp(P, U) \wedge p \wedge \text{cond}(V)$. It is important to check if P' is satisfiable.

Checking coverage A node $\langle l, P, Z\rangle$ is covered by another node $\langle l^c, P^c, Z^c\rangle$ if $l = l^c$,$Z \subseteq Z^c$ and $P \implies P^c$. There are many ways to check implication. In our implementation we use a solver to determine whether $P \wedge \neg P^c$ is unsatisfiable.

Termination The algorithm terminates if there are no unexplored nodes in the current graph or if it reached the initial state. The set of states represented by a node $n = \langle l, P, Z\rangle$ contain the initial state if $l = l_0$, $v_0^d \models P$, and $v_0 \in Z$.

4.2 Example



(a) Automaton (b) Explored graph

Fig. 1: Example

This section demonstrates the operation of the proposed algorithm on the automaton presented in Figure 1a and the target location l_1 with predicate $i \leq 2$.

In this automaton (which is a modified version of the one presented in [3]), there are two clock variables x and y and a data variable i that serves as a loop counter. The property to decide is whether l_1 can be reached by taking the loop transition less than 2 times.

Initially, clocks x and y are both initialized to 0, which means $x = y$ will hold as long as the system stays in l_0. When the loop edge is taken (that can first happen when $x = y = 10$) x is reset but the value of y continues to increase from the same value – that is, taking the loop edge means increasing $y - x$ by 10. The algorithm operates as follows.

The initial node is $n_0 = \langle l_1, i \leq 2, [\![true]\!] \rangle$. The only incoming edge does not change the data valuation, but has a clock guard. Therefore the preimage of the node is $n_1 = \langle l_0, i \leq 2, y - x > 0 \rangle$.

The preimage of the zone based on the loop edge is $\mathrm{Pre}_{x==10,\{x\}}(y-x > 0) = = ([\{x\} \leftarrow 0]^{-1}(y - x > 0) \cap [\![x == 10]\!])^{\downarrow} = (y > 0 \cap [\![x == 10]\!])^{\downarrow} = x \leq 10$, and $wp(i \leq 2, i := i + 1) = i \leq 1$. The created node is $n_2 = \langle l_0, i \leq 1, x \leq 10 \rangle$. Notice, $i = 0 \models i \leq 1$ and $x = y = 0 \models x \leq 10$: n_2 represents the initial state, the property is evaluated to *true*.

The algorithm terminates, however, in order to demonstrate the operations of the algorithm here we continue to explore the state space. The preimage based on the loop edge is $n_3 = \langle l_0, i \leq 0, x \leq 10 \rangle$. Since $i \leq 0 \implies i \leq 1$, n_2 covers n_3. This means all the states represented by n_3 are also represented by n_2, so if n_3 is reachable from a state, n_2 is also reachable. Because of the coverage, node n_3 does not need to be expanded further.

The explored graph can be seen in Figure 1b.

4.3 Proof of correctness

Claim: $\langle l_0, v_0^d, v_0^c \rangle$ is represented by one of the nodes of the abstract reachability graph *iff* there is a run $s_0 = \langle l_0, v_0^d, v_0^c \rangle \to \cdots \to s'_n = \langle l_n, v_n^d, v_n^c \prime \rangle$ of $\mathcal{A}$ where $l_n = l_{\mathrm{trg}}$ is the target location.

The proof is based on the fact, the values of clock and data variables are independent of each other. Therefore, the computation of zones and (data) predicates are also independent and if both of the calculations are correct, the complete algorithm is also correct. Zone calculations are proven to be correct, and the weakest precondition operation is *defined* to be correct.

Proof: ($\Leftarrow$) s_n is represented by the initial node. Suppose a state of the run s'_k is represented by some $n_k = \langle C_k, P_k, Z_k \rangle$ in this case n_k also represents s_k since the $v_k^c \in \{v_k^c \prime\}^{\downarrow}$.

Let $e_{k-1} = \langle l_{k-1}, g, p, r, U, l_k \rangle$ be the edge corresponding to the discrete transition $s_{k-1} \to s_k$.

In this case the following claims hold.

- $v_{k-1}^c \in Pre_{g,r}(\{v_k^c\})$, since the zone operations are correct

– $v^d_{k-1} \models wp(\bigwedge_{v \in V} v = v^d_k(v), U)$, since the weakest precondition was defined like that, therefore
– $v^d_{k-1} \models wp(P, U)$ for any P where $v^d_k \models P$
– $v^d_{k-1} \models p$ because the transition is enabled
– $v^d_{k-1} \models \text{cond}(V)$

Therefore, there is also an edge of the graph representing e from n_k to some other node n_{k-1} that represents s_{k-1} and – by induction – there is also a node representing s_0.

($\Rightarrow$) Let node $n_0 = \langle l_0, P_0, Z_0 \rangle$ be a node of the graph, where l_0 is the initial location $v^c_0 \in Z_0$ and $v^0_d \models P_0$. If a node n of the abstract reachability graph is not the initial node, then it was created along with an incoming edge $e(n)$. Find the path $n_i \xrightarrow{e(n_{i-1})} n_{i-1} \dots n_1 \xrightarrow{e(n_0)} n_0$ where n_i is the initial node. Because of the way the edges are chosen, all states represented by a node n_k are a preimage of some state represented by n_{k+1} (other incoming edges might have been introduced based on coverage and in that case, the preimage might be just a subset of the states represented by n_{k+1}). Therefore for all states s_k represented by n_k there is a state s_{k+1} represented by n_{k+1} such that s_{k+1} is reachable from s_k by the edge corresponding to $e(n_k)$. By induction, n_i represents a state reachable from s_0.

5 Comparison of backward and forward exploration

5.1 Measurements

To analyse the applicability and the efficiency of the algorithm we have implemented it in the Theta verification framework [22] and run measurements on the XtaBenchmarkSuite[3], that is a set of test cases for benchmarking timed automaton verification algorithms (we have presented an earlier version of the benchmark suite in [12], but it has been improved since). To compare the algorithm to forward exploration we also run measurements on the algorithm presented in [15].

The benchmark suite contains *networks of timed automata with data variables*. As mentioned in Section 3.2, for some inputs the properties only define the target location for a few of the automata in the network. In these cases, the possible target configurations were manually collected – all possible combinations of the locations of the remaining automata were included. In many cases, this resulted in a set of target configurations too large for efficient verification. However, by observing the models, we could eliminate many of the configurations. To demonstrate the significance of describing a precise property, we run the backward exploration algorithm on both the original and the modified property. These modifications did not affect the forward exploration algorithms.

The measurements were executed on a virtual 64 bit Windows 7 operating system with a 2 core CPU (2.50 GHz) with 1,5 GB of memory. Each algorithm

[3] https://github.com//farkasrebus/XtaBenchmarkSuite

was run 10 times on each input, the longest and the shortest was eliminated and the average of the remaining runtimes was taken. The timeout was 10 minutes.

5.2 Evaluation

The XtaBenchmarkSuite consists of more models than those mentioned here, but we have excluded those that contain a model element that is not yet supported by Theta (e.g. broadcast channels) and those where the property to check is not a reachability property (e.g. liveness properties) or only constrains the data variables. We have also excluded the models where both of the studied algorithms timed out.

Table 1 describes the results of the benchmarks on diagonal-free timed automata. The column *Name* contains the name of the model in the XtaBenchmarkSuite and in case of scalable models, the value of the parameter. The next two columns contain the number of target configurations as described by the original and – when we managed to refine it – the precise property. The remaining columns contain the runtime and the number of explored nodes by the algorithms. The columns denoted by LU correspond to the algorithm presented in [15], while the remaining columns correspond to the execution of our backward exploration algorithm on the model with the original and the modified property.

Table 2 describes the results of the benchmarks on timed automata with diagonal constraints. We did not manage to find an industrial case study containing diagonal constraints, thus we use the models we found in the literature as examples. Model *SPLIT* was introduced in [3] to demonstrate the incorrectness of the forward exploration algorithm and *DIAG* was used to demonstrate techniques describes in [8]. We also run the algorithms on modified versions: we added new edges to these examples to create cyclic models.

Weakest precondition One of our research goals was to find out if the weakest precondition operation can be used to efficiently calculate data variables during backward exploration of timed automata. Only a few models of the benchmark suite contain data variables (protocols at most one per process, and inputs of category *system* have integer variables, but their set of possible values is rather small). A possible difficulty would have occurred if the abstract reachability graph were deeper, but this was not the case as we used breadth-first search. Out of these models *ENGINE* had the deepest reachability graph with a depth of 63.

Furthermore, most algorithms calculate the values of data variables explicitly. Weakest precondition makes it possible to use predicate abstraction on the data variables. This also shows that the weakest precondition is useful for backward exploration of timed automata.

Property definition One of the motivations for using backward exploration is that the state space explored from the target states might be smaller than

Table 1: Results on diagonal-free systems

Name	Target configurations		LU		BW			
	original	precise			original		precise	
			time (ms)	nodes	time (ms)	nodes	time (ms)	nodes
EXSITH	1	-	1	4	-	-	2	5
LATCH	1	-	2	7	-	-	1	2
SRLATCH	4	1	0	6	5	28	4	16
ANDOR	4	1	11	9	42	39	13	21
ENGINE	288	-	35	674	3731	8783	-	-
SIMOP	8960	128	72	325	9820	26271	800	1517
MALER	1	-	553	7157	-	-	134127	53519
MUTEX	5	-	1248	5563	[TO]	-	-	-
CRITICAL2	56	1	135	112	1631	1914	18	22
CRITICAL3	784	8	148	1632	[TO]	-	599	811
CRITICAL4	10976	64	770	25202	[TO]	-	233114	51488
CSMA2	4	-	23	18	16	26	-	-
CSMA3	16	-	29	71	247	338	-	-
CSMA4	64	-	55	262	16547	11306	-	-
CSMA5	256	-	177	855	[TO]	-	-	-
FISCHER2	1	-	14	18	19	11	-	-
FISCHER3	4	-	11	65	93	78	-	-
FISCHER4	16	-	59	220	476	594	-	-
FISCHER5	64	-	79	727	10221	10426	-	-
FISCHER6	256	-	366	2378	96102	93735	-	-
FISCHER7	124	-	1913	7737	[TO]	-	-	-
LYNCH2	1	-	20	38	18	33	-	-
LYNCH3	9	-	32	125	656	821	-	-
LYNCH4	81	-	47	380	35344	27523	-	-
TRAIN2	64	-	14	44	156	380	-	-
TRAIN3	256	-	37	165	694	1330	-	-
TRAIN4	1024	-	315	1123	6886	10290	-	-
TRAIN5	4096	-	3874	6488	66837	79892	-	-
TRAIN6	16384	-	103455	60379	[TO]	-	-	-

Table 2: Results on diagonal timed automata

Name	LU				BW	
	Transformation time (ms)	Algorithm time (ms)	Total time (ms)	Nodes	Time (ms)	Nodes
SPLIT (original)	4	4	8	9	3	4
SPLIT (modified)	27	16	43	10	4	4
DIAG (original)	14	6	20	22	8	8
DIAG (modified)	32	22	54	24	14	8

the reachable state space. However, in practice the properties describing the target states are kept simple, only mentioning a few locations and variables of the system. While this helps understandability, it also increases the size of the explored state space.

Clocks and data variables can be handled with symbolic methods, but locations have to be explored explicitly. For instance, in case of mutual exclusion protocols (that appears to be the most common area of applicability for timed automata) the target state is any state where at least two processes are in the critical section. In case of a system of n similar processes containing k locations there are k^{n-2} possible control vectors that satisfy this property (on the other hand, this could be the reason for the depths of the reachability graphs).

We believe this issue could be overcome *in many cases*, such as the *CRITICAL* example, where the target location is only given for the *Prod cell* automata, however it is easy to determine the current location of the corresponding *Arbiter* automata. Therefore, we believe that some initial static analysis of the automaton and the constraints (e.g. the *forward* exploration of the location graph) could lower the number of initial states.

As it is seen in Table 1 adding details to the property increases the efficiency of backward exploration.

Timed automata with diagonal constraints One of the key strengths of backward exploration for timed automata is that it is *correct* for timed automata with diagonal constraints. The transformation of a timed automaton containing diagonal constraints to a diagonal-free automaton takes time and increases the size of the automaton and thus makes verification less efficient–as demonstrated in Table 2.

We were only able to perform these measurements on small examples from the literature, since *there are no public industrial case studies of timed automata with diagonal constraints*. We believe this is due to the fact that most analysis tools do not support diagonal constraints and because of this the case studies are stored in diagonal-free form.

General conclusions In case of timed automata with diagonal constraints, backward exploration performs better on the small examples, however, we could not find an industrial example with diagonal constraints to support this claim.

In case of diagonal-free timed automata forward exploration performs better in most cases. However, for many small models backward exploration explored a smaller state space, e.g. the *LATCH* circuit and protocols with a small number of participants.

Forward and backward exploration are complementary techniques. Theta is able to perform both, so it is possible chose the more appropriate one for a certain model.

5.3 Possible improvements

One of the reasons why forward exploration performs better on the case studies is that there are many optimizations for zone-based forward search. It would be desired to apply these optimizations for backward exploration as well.

Backward exploration can also be combined with other techniques to improve the efficiency of existing algorithms. A common application for backward exploration is to run it in parallel with forward exploration – i.e. *two-way search*. This way, if the target state is reachable, the explored state space becomes smaller than a single search from any direction. The two algorithms can also increase each others efficiency by exchanging information on the explored state space [24]. Two-way search for timed automata appears to be an area worth exploring.

6 Conclusions and future work

In this paper we have presented an algorithm that uses zone-based backward exploration for the reachability analysis of timed automata. Unlike zone-based forward exploration, this algorithm is also applicable for timed automata with diagonal constraints. We have extended the algorithm to handle complex data: the weakest precondition operation is suitable for calculating the values of data variables. The combination of the two algorithms resulted and efficient procedure for timed systems with data.

XtaBenchmarkSuite is a benchmark suite consisting of timed automata including small examples and industrial case studies. XtaBenchmarkSuite was extended in this paper to evaluate the various timed algorithms more rigorously. Measurements on the XtaBenchmarkSuite demonstrated the efficiency and the applicability of our new approach. We have shown that while the algorithm performs better on timed automata with diagonal constraints than forward exploration, some improvements and optimizations are necessary in order to be applicable for industrial case studies. Our results provide a step towards the efficient verification of real-time systems.

In the future we will improve the backward exploration algorithm to solve complex industrial case studies. We will analyse the applicability of the improvements of the forward exploration algorithm presented in [15,23] to our backward exploration approach. We will also combine backward and forward search algorithms – as it was investigated that two-way search is efficient for SAT-based algorithms in [24]. We will also investigate the combination of forward and backward algorithms in the CEGAR framework as it was discussed in [20].

In order to further investigate the strengths and weaknesses of the timed algorithms, we will extend the XtaBenchmarkSuite with more industrial case studies (e.g. the FlexRay protocol [13]) and perform an exhaustive benchmark on the existing approaches. Based on the results a – possibly automatic – process can be derived for deciding which algorithm is expected to be the most efficient for verifying a given system. Diagonal constraints represent one extension of timed automata, which proved to be useful for a certain class of models. In the future, we have to analyse other extensions from the verification point of view.

References

1. Alur, R., Dill, D.L.: The theory of timed automata. In: Proceedings of the Real-Time: Theory in Practice, REX Workshop. pp. 45–73. Springer (1991). https://doi.org/10.1007/BFb0031987
2. Bengtsson, J., Larsen, K.G., Larsson, F., Pettersson, P., Yi, W.: UPPAAL - a tool suite for automatic verification of real-time systems. In: DIMACS/SYCON 1995. pp. 232–243. Springer (1995). https://doi.org/10.1007/BFb0020949
3. Bengtsson, J., Yi, W.: Timed automata: Semantics, algorithms and tools. In: ACPN 2003. pp. 87–124. Springer (2003). https://doi.org/10.1007/978-3-540-27755-2_3
4. Bouyer, P.: Untameable timed automata! In: STACS 2003. pp. 620–631. Springer (2003). https://doi.org/10.1007/3-540-36494-3_54
5. Bouyer, P.: Forward analysis of updatable timed automata. Formal Methods in System Design **24**(3), 281–320 (2004). https://doi.org/10.1023/B:FORM.0000026093.21513.31
6. Bouyer, P.: An introduction to timed automata. ETR 2005 pp. 25–52 (2005)
7. Bouyer, P., Chevalier, F.: On conciseness of extensions of timed automata. Journal of Automata, Languages and Combinatorics **10**(4), 393–405 (2005)
8. Bouyer, P., Laroussinie, F., Reynier, P.: Diagonal constraints in timed automata: Forward analysis of timed systems. In: FORMATS 2005. pp. 112–126. Springer (2005). https://doi.org/10.1007/11603009_10
9. Bradley, A.R., Manna, Z.: The calculus of computation - decision procedures with applications to verification. Springer (2007). https://doi.org/10.1007/978-3-540-74113-8
10. Dill, D.L.: Timing assumptions and verification of finite-state concurrent systems. In: Proceedings of the International Workshop on Automatic Verification Methods for Finite State Systems. pp. 197–212. Springer (1989). https://doi.org/10.1007/3-540-52148-8_17
11. Dutertre, B., Sorea, M., et al.: Timed systems in SAL. Tech. rep., SRI International, Computer Science Laboratory (2004)
12. Farkas, R., Bergmann, G.: Towards reliable benchmarks of timed automata. In: Proceedings of the 25th PhD Mini-Symposium. pp. 20–23. Budapest University of Technology and Economics, Department of Measurement and Information Systems (2018)
13. Gerke, M., Ehlers, R., Finkbeiner, B., Peter, H.J.: Model checking the flexray physical layer protocol. In: FMICS 2010. LNCS, vol. 6371, pp. "132–147". Springer (2010). https://doi.org/10.1016/j.entcs.2005.04.014
14. Henzinger, T.A., Nicollin, X., Sifakis, J., Yovine, S.: Symbolic model checking for real-time systems. In: LICS 1992. pp. 394–406. IEEE (1992). https://doi.org/10.1109/LICS.1992.185551
15. Herbreteau, F., Srivathsan, B., Walukiewicz, I.: Lazy abstractions for timed automata. In: CAV 2013. pp. 990–1005. Springer (2013). https://doi.org/10.1007/978-3-642-39799-8_71
16. Leino, K.R.M.: Efficient weakest preconditions. Inf. Process. Lett. **93**(6), 281–288 (2005). https://doi.org/10.1016/j.ipl.2004.10.015
17. Maler, O., Pnueli, A.: Timing analysis of asynchronous circuits using timed automata. In: CHARME 1995. pp. 189–205. Springer (1995). https://doi.org/10.1007/3-540-60385-9_12

18. Mitchell, I.M.: Comparing forward and backward reachability as tools for safety analysis. In: HSCC 2007. pp. 428–443. Springer (2007). https://doi.org/10.1007/978-3-540-71493-4_34
19. Morbé, G., Pigorsch, F., Scholl, C.: Fully symbolic model checking for timed automata. In: CAV 2011. pp. 616–632. Springer (2011). https://doi.org/10.1007/978-3-642-22110-1_50
20. Ranzato, F., Rossi-Doria, O., Tapparo, F.: A forward-backward abstraction refinement algorithm. In: VMCAI 2008. pp. 248–262. Springer (2008). https://doi.org/10.1007/978-3-540-78163-9_22
21. Ravn, A.P., Srba, J., Vighio, S.: A formal analysis of the web services atomic transaction protocol with UPPAAL. In: ISoLA 2010. pp. 579–593. Springer (2010). https://doi.org/10.1007/978-3-642-16558-0_47
22. Tóth, T., Hajdu, A., Vörös, A., Micskei, Z., Majzik, I.: Theta: a framework for abstraction refinement-based model checking. In: FMCAD 2017. pp. 176–179. IEEE (2017). https://doi.org/10.23919/FMCAD.2017.8102257
23. Tóth, T., Majzik, I.: Lazy reachability checking for timed automata using interpolants. In: Formal Modelling and Analysis of Timed Systems, LNCS, vol. 10419, pp. 264–280. Springer (2017). https://doi.org/10.1007/978-3-319-65765-3_15
24. Vizel, Y., Grumberg, O., Shoham, S.: Intertwined forward-backward reachability analysis using interpolants. In: TACAS 2013. pp. 308–323. Springer (2013). https://doi.org/10.1007/978-3-642-36742-7_22
25. Yovine, S.: KRONOS: A verification tool for real-time systems. International Journal on Software Tools for Technology Transfer **1**(1-2), 123–133 (1997). https://doi.org/10.1007/s100090050009