Timed Automata

Rajeev Alur
University of Pennsylvania and Bell Labs
alur@cis.upenn.edu

Abstract

Model checking is emerging as a practical tool for automated debugging of complex reactive systems such as embedded controllers and network protocols (see [CK96] for a survey). In model checking, a high-level description of a system is compared against a logical correctness requirement to discover inconsistencies. Traditional techniques for model checking do not admit an explicit modeling of time, and are thus, unsuitable for analysis of real-time systems whose correctness depends on relative magnitudes of different delays. Consequently, timed automata [AD94] were introduced as a formal notation to model the behavior of real-time systems. Its definition provides a simple, and yet general, way to annotate state-transition graphs with timing constraints using finitely many real-valued clock variables. Automated analysis of timed automata relies on the construction of a finite quotient of the infinite space of clock valuations. Over the years, the formalism has been extensively studied leading to many results establishing connections to circuits and logic, and much progress has been made in developing verification algorithms, heuristics, and tools. This paper provides a survey of theory of timed automata, and their role in specification and verification of real-time systems.

1 Modeling

Transition systems

We model systems by state-transition graphs whose transitions are labeled with event symbols. A transition system $S$ is a tuple $\langle Q, Q^0, \Sigma, \rightarrow \rangle$, where $Q$ is a set of states, $Q^0 \subseteq Q$ is a set of initial states, $\Sigma$ is a set of labels (or events), and $\rightarrow \subseteq Q \times \Sigma \times Q$ is a set of transitions. For a transition $(q, a, q')$ in $\rightarrow$, we write $q \xrightarrow{a} q'$. The system starts in an initial state, and if $q \xrightarrow{a} q'$ then the system can change its state from $q$ to $q'$ on event $a$. We write $q \rightarrow q'$ if $q \xrightarrow{a} q'$ for some label $a$. The state $q'$ is reachable from the state $q$ if $q \rightarrow^* q'$. The state $q$ is a reachable state of the transition system if $q$ is reachable from some initial state.

A complex system can be described as a product of interacting transition systems. Let $S_1 = \langle Q_1, Q^0_1, \Sigma_1, \rightarrow_1 \rangle$ and $S_2 = \langle Q_2, Q^0_2, \Sigma_2, \rightarrow_2 \rangle$ be two transition systems. Then, in the product of $S_1$ and $S_2$, a state is a pair $(q, q')$ with $q \in Q_1$ and $q' \in Q_2$. The transitions of the product are labeled with symbols in $\Sigma_1 \cup \Sigma_2$. For a label $a$, to obtain $a$-labeled transitions of the product, we require each component system with $a$ in its label-set to execute an $a$-labeled transition. Formally, the product, denoted $S_1||S_2$, is the transition system $\langle Q_1 \times Q_2, Q^0_1 \times Q^0_2, \Sigma_1 \cup \Sigma_2, \rightarrow \rangle$ where $(q_1, q_2) \xrightarrow{a} (q'_1, q'_2)$ iff either (i) $a \in \Sigma_1 \cap \Sigma_2$ and $q_1 \xrightarrow{a_1} q'_1$ and $q_2 \xrightarrow{a_2} q'_2$, or (ii) $a \in \Sigma_1 \setminus \Sigma_2$ and $q_1 \xrightarrow{a_1} q'_1$ and $q'_2 = q_2$, or (iii) $a \in \Sigma_2 \setminus \Sigma_1$ and $q_2 \xrightarrow{a_2} q'_2$ and $q'_1 = q_1$. Observe that the symbols that belong to the alphabets of both the automata are used for synchronization. In this definition, synchronization is
blocking, for a common symbol $a$, a component can execute an $a$-labeled switch only if the other
component can also do so.

**Transition systems with timing constraints**

To express system behaviors with timing constraints, we consider finite graphs augmented with a
finite set of (real-valued) clocks. The vertices of the graph are called locations, and edges are called
switches. While switches are instantaneous, time can elapse in a location. A clock can be reset to
zero simultaneously with any switch. At any instant, the reading of a clock equals the time elapsed
since the last time it was reset. With each switch we associate a clock constraint, and require that
the switch may be taken only if the current values of the clocks satisfy this constraint. With each
location we associate a clock constraint called its invariant, and require that time can elapse in a
location only as long as its invariant stays true. Before we define the timed automata formally, let
us consider some examples.

Consider the timed automaton of Figure 1. The initial location is $s$. There is a single clock $x$.
The initial location has no invariant constraint which means that the system can spend arbitrary
amount of time in location $s$. When the system switches to location $s'$ on symbol $a$, the clock $x$
gets reset to 0. While in location $s'$, the value of the clock $x$ shows the time elapsed since the
occurrence of the last switch. The switch from location $s'$ to $s$ is enabled only if this value is
greater than 1. The invariant $x \leq 2$ associated with the location $s'$ specifies the requirement that
the system can stay in location $s'$ for at most 2 units, and a switch must occur before the invariant
is violated. Thus the timing constraint expressed by this automaton is that the delay between $a$
and the following $b$ is always between 1 and 2.

Having multiple clocks allows multiple concurrent delays, as illustrated in Figure 2. The clock $x$
gets set to 0 each time the system switches from $s_0$ to $s_1$ on symbol $a$. The invariant ($x < 1$)
associated with the locations $s_1$ and $s_2$ ensures that $c$-labeled switch from $s_2$ to $s_3$ happens within
time 1 of the preceding $a$. Resetting another independent clock $y$ together with the $b$-labeled switch
from $s_1$ to $s_2$ and checking its value on the $d$-labeled switch from $s_3$ to $s_0$ ensures that the delay between $b$ and the following $d$ is always greater than 2. Since the location $s_3$ has no invariant constraint, the event $d$ can be postponed indefinitely, and is not guaranteed to happen.

Notice that in the above example, to constrain the delay between $a$ and $c$ and between $b$ and $d$ the system does not put any explicit bounds on the time difference between $a$ and the following $b$, or $c$ and the following $d$. This is an important advantage of having multiple clocks which can be set independently of one another. We remark that the clocks of the system do not necessarily correspond to the local clocks of different components in a distributed system. They are fictitious clocks invented to express the timing properties of the system, and all the clocks increase at the uniform rate counting time with respect to a fixed global time frame. Alternatively, we can consider the system to be equipped with a finite number of stop-watches which can be started and checked independently of one another, but all stop-watches refer to the same clock.

**Clock constraints and clock interpretations**

To define timed automata formally, we need to say what type of clock constraints are allowed as invariants and enabling conditions. An atomic constraint compares a clock value with a time constant, and a clock constraint is a conjunction of atomic constraints. Any value from $\mathbb{Q}$, the set of nonnegative rationals, can be used as a time constant. Formally, for a set $X$ of clock variables, the set $\Phi(X)$ of clock constraints $\varphi$ is defined by the grammar

$$\varphi ::= x \leq c \mid c \leq x \mid x < c \mid c < x \mid \varphi_1 \land \varphi_2,$$

where $x$ is a clock in $X$ and $c$ is a constant in $\mathbb{Q}$.

A clock interpretation $\nu$ for a set $X$ of clocks assigns a real value to each clock; that is, it is a mapping from $X$ to the set $\mathbb{R}$ of nonnegative reals. We say that a clock interpretation $\nu$ for $X$ satisfies a clock constraint $\varphi$ over $X$ if $\varphi$ evaluates to true according to the values given by $\nu$. For $\delta \in \mathbb{R}$, $\nu + \delta$ denotes the clock interpretation which maps every clock $x$ to the value $\nu(x) + \delta$. For $Y \subseteq X$, $\nu[Y := 0]$ denotes the clock interpretation for $X$ which assigns 0 to each $x \in Y$, and agrees with $\nu$ over the rest of the clocks.

**Syntax and semantics**

A timed automaton is a tuple $\langle L, L^0, \Sigma, X, I, E \rangle$, where

- $L$ is a finite set of locations,
- $L^0 \subseteq L$ is a set of initial locations,
- $\Sigma$ is a finite set of labels,
- $X$ is a finite set of clocks,
- $I$ is a mapping that labels each location $s$ in $L$ with some clock constraint in $\Phi(X)$, and
- $E \subseteq L \times \Sigma \times 2^X \times \Phi(X) \times L$ is the set of switches. A switch $\langle s, a, \varphi, \lambda, s' \rangle$ represents a transition from location $s$ to location $s'$ on symbol $a$. $\varphi$ is a clock constraint over $X$ that specifies when the switch is enabled, and the set $\lambda \subseteq X$ gives the clocks to be reset with this switch.
The semantics of a timed automaton $A$ is defined by associating a transition system $S_A$ with it. A state of $S_A$ is a pair $(s, \nu)$ such that $s$ is a location of $A$ and $\nu$ is a clock interpretation for $X$ such that $\nu$ satisfies the invariant $I(s)$. The set of all states of $A$ is denoted $Q_A$. A state $(s, \nu)$ is an initial state if $s$ is an initial location of $A$ and $\nu(x) = 0$ for all clocks $x$. There are two types of transitions in $S_A$:

- State can change due to elapse of time: for a state $(s, \nu)$ and a real-valued time increment $\delta \geq 0$, $(s, \nu) \xrightarrow{\delta} (s, \nu + \delta)$ if for all $0 \leq \delta' \leq \delta$, $\nu + \delta'$ satisfies the invariant $I(s)$.

- State can change due to a location-switch: for a state $(s, \nu)$ and a switch $(s, a, \varphi, \lambda, s')$ such that $\nu$ satisfies $\varphi$, $(s, \nu) \xrightarrow{\lambda} (s', \nu[\lambda := 0])$.

Thus, $S_A$ is a transition system with label-set $\Sigma \cup \mathbb{R}$. For instance, for the timed automaton of Figure 2, the state-space of the associated transition system is $\{s_0, s_1, s_2, s_3\} \times \mathbb{R}^2$, the label-set is $\{a, b, c, d\} \cup \mathbb{R}$, and some sample transitions are

$$(s_0, 0, 0) \xrightarrow{1/2} (s_0, 1, 2, 1.2) \xrightarrow{a} (s_1, 0, 1, 2) \xrightarrow{0.7} (s_1, 0.7, 1.9) \xrightarrow{b} (s_2, 0.7, 0) \xrightarrow{0.1} (s_2, 0.8, 0.1) \xrightarrow{0.1} (s_2, 0.9, 0.2)$$

Note the time-additivity property of the transition system $S_A$: if $q \xrightarrow{\delta} q'$ and $q' \xrightarrow{\delta'} q''$ then $q \xrightarrow{\delta + \delta'} q''$.

**Remark 1 (Multiple events in zero time)** By our definition, a transition corresponds to a single event or elapse of time, and arbitrary interleaving of the two types of transitions is allowed. Thus, multiple events can occur without any intervening time-elapsed transition. Variants of this scheme are possible (e.g. the original definition in [AD94] requires nonzero time to elapse between two location-switches) and do not affect any of the results.

**Remark 2 (Nonzenoness)** In this introductory survey, we have omitted requirements on the definition necessary for executability. First, when the invariant of a location is violated, some outgoing edge must be enabled. Second, from every reachable state, the automaton should admit the possibility of time to diverge. For example, the automaton should not enforce infinitely many events in a finite interval of time. Automata satisfying this operational requirement are called nonZeno. The interested reader is referred to [AL91, GSSL94, AH97].

**Product construction**

We proceed to define a product construction for timed automata so that a complex system can be defined as a product of component systems. Let $A_1 = \langle L_1, \Sigma_1, X_1, I_1, E_1 \rangle$ and $A_2 = \langle L_2, L_2^0, \Sigma_2, X_2, I_2, E_2 \rangle$ be two timed automata. Assume that the clock sets $X_1$ and $X_2$ are disjoint. Then, the product, denoted $A_1 || A_2$, is the timed automaton $\langle L_1 \times L_2, L_1^0 \times L_2^0, \Sigma_1 \cup \Sigma_2, X_1 \cup X_2, I, E \rangle$, where $I(s_1, s_2) = I(s_1) \land I(s_2)$ and the switches are defined by:

1. for $a \in \Sigma_1 \cap \Sigma_2$, for every $\langle s_1, a, \varphi_1, \lambda_1, s'_1 \rangle$ in $E_1$ and $\langle s_2, a, \varphi_2, \lambda_2, s'_2 \rangle$ in $E_2$, $E$ contains $\langle (s_1, s_2), a, \varphi_1 \land \varphi_2, \lambda_1 \lor \lambda_2, (s'_1, s'_2) \rangle$.
2. for $a \in \Sigma_1 \setminus \Sigma_2$, for every $\langle s, a, \varphi, \lambda, s' \rangle$ in $E_1$ and every $t$ in $L_2$, $E$ contains $\langle (s, t), a, \varphi, \lambda, (s', t) \rangle$.
3. for $a \in \Sigma_2 \setminus \Sigma_1$, for every $\langle s, a, \varphi, \lambda, s' \rangle$ in $E_2$ and every $t$ in $L_1$, $E$ contains $\langle (t, s), a, \varphi, \lambda, (t, s') \rangle$.

Thus, locations of the product are pairs of component-locations, and the invariant of a compound location is the conjunction of the invariants of the component locations. The switches are obtained
by synchronizing the switches with identical labels. The product construction is illustrated in Figure 3.

It is easy to verify that transition system associated with the product is the product of the transition systems of the components: $S_{A_1 \parallel A_2}$ and $S_{A_1} \parallel S_{A_2}$ are isomorphic. This implies that timed transition systems synchronize not only on switches with common labels, but also on amount of time elapsed.

**Remark 3 (Compositionality)** For communication between system components, many competing alternatives to blocking synchronization exist. The choice of synchronization primitives is somewhat orthogonal to the problem of analysis of timing constraints. Examples of formalisms that admit modeling real-time systems include timed transition systems [Ost90, HMP99], timed I/O automata [LA92], process algebras such timed CSP [RR88], and Modecharts [JM87]. The algorithmic techniques for timed automata can be applied to these other models also via compilation. To model open real-time systems (i.e., those interacting with the environment), one needs to make a distinction between which events are controlled by the system and which events are controlled by the environment. Such a compositional framework provides foundations to decompose the analysis problem into simpler problems [MMT91, AH97]. Issues pertaining to the impact of timing on synchronization are studied in [BST98].

**Train-Gate Controller Example**

We consider an example of an automatic controller that opens and closes a gate at a railroad crossing. The system is composed of three components: Train, Gate and Controller as shown in Figure 4. The train communicates with the controller with two events approach and exit. The events in and out mark the events of entry and exit of the train from the railroad crossing. The train is required to send the signal approach at least 2 minutes before it enters the crossing. This requirement is expressed by the guard $x \geq 2$ associated with the event in. Furthermore, we know that the maximum delay between the signals approach and exit is 5 minutes. This requirement is expressed by the invariant $x \leq 5$ in the locations $s_1$, $s_2$, and $s_3$. The gate is open in location $t_0$ and closed in location $t_2$. It communicates with the controller through the signals lower and raise. The events up and down denote the opening and the closing of the gate. The gate responds to the signal lower by closing within 1 minute, and responds to the signal raise within 1 to 2 minutes.
The clock $y$ is used to express these constraints. The controller idle location is $u_0$. Whenever it receives the signal $\text{approach}$ from the train, it responds by sending the signal $\text{lower}$ to the gate. Whenever it receives the signal $\text{exit}$, it responds with a signal $\text{raise}$ to the gate. The response time of the controller to the $\text{approach}$ signal is 1 minute, and to the signal $\text{exit}$ is at most 1 minute. These constraints are expressed using the clock $z$. The entire system $\text{GRC}$ is then $\text{TRAIN} \parallel \text{GATE}$.

The safety correctness requirement for the system is that whenever the train is inside the gate, the gate should be closed. This corresponds to establishing that in the product system $\text{GRC}$, in every reachable state, if the location of $\text{TRAIN}$ is $s_2$ then the location of $\text{GATE}$ should be $t_2$. Observe that such a location is reachable in the product graph. For example, there is an edge from the initial location $(s_0, t_0, u_0)$ to $(s_1, t_0, u_1)$, and from $(s_1, t_0, u_1)$ to $(s_2, t_0, u_1)$, corresponding to the scenario in which the event $\text{approach}$ is immediately followed by the event $\text{in}$. This is because our product is simply a syntactic operation that annotates product locations with conjunctions of invariants, and product edges with conjunctions of enabling conditions, without any analysis. If we consider the timing information, we can establish that the event $\text{approach}$ cannot be immediately followed by the event $\text{in}$: in the location $(s_1, t_0, u_1)$ both clocks $x$ and $z$ have the same value, and hence the event $\text{lower}$ with guard $z = 1$ is guaranteed to precede the event $\text{in}$ with guard $x > 2$. In fact, in the transition system $S_{\text{GRC}}$, no reachable state has $\text{TRAIN}$ in location $s_2$ with $\text{GATE}$ not in location $t_2$. The computational problem in timing verification is to make such deductions by analyzing the timing constraints.

2 Reachability Analysis

A location $s$ of the timed automaton $A$ is said to be reachable if some state $q$ with location component $s$ is a reachable state of the transition system $S_A$. The input to the reachability problem consists of a timed automaton $A$ and a set $L^F \subseteq L$ of target locations of $A$. The reachability problem is to determine whether or not some target location is reachable. Verification of safety requirements
of real-time systems can be formulated as reachability problems for timed automata, as illustrated in the train-gate example. Since the transition system $S_A$ of a timed automaton is infinite, our solution to the reachability problem involves construction of finite quotients.

**Time-abstract transition system**

Let $A$ be a timed automaton. The transition system $S_A$ has infinitely many states and infinitely many symbols. As a first step, we define another transition system whose transitions are labeled only with the symbols in $\Sigma$ by hiding the labels denoting the time increments. For this purpose, whenever $S_A$ contains a $\delta$-labeled transition from state $q$ to state $q'$ and an $\alpha$-labeled transition from $q'$ to state $q''$, we add an $\alpha$-labeled transition from $q$ to $q''$. The resulting transition system is called *time-abstract* as it does not retain information about amount of time elapsed during transitions.

Formally, the time-abstract transition system of a timed automaton $A$, denoted $U_A$, is defined as follows. The state-space of $U_A$ equals the state-space $Q_A$ of $S_A$. The set of initial states of $U_A$ equals the set of initial states of $S_A$. The set of labels of $U_A$ equals the set $\Sigma$ of labels of $A$. The transition relation of $U_A$ is the relation $\Rightarrow$: for states $q$ and $q'$ and a label $\alpha$, $q \Rightarrow q'$ iff there exists a state $q''$ and a time value $\delta \in \mathbb{R}$ such that $q \xrightarrow{\delta} q'' \xrightarrow{\alpha} q'$ holds in the transition system $S_A$.

In the reachability problem for timed automata, we wish to determine reachability of target locations. Note that a location $s$ of $A$ is reachable iff some state with location $s$ is reachable in the time-abstract transition system $U_A$. It follows that to solve reachability problems, we can consider the time-abstract transition system $U_A$ instead of $S_A$.

**Stable quotients**

While the time-abstract transition system $U_A$ has only finitely many labels, it still has infinitely many states. To address this problem, we consider equivalence relations over the state-space $Q_A$ that group states together. An equivalence relation $\sim$ over the state-space $Q_A$ of a timed automaton $A$ is said to be *stable* iff whenever $q \sim u$ and $q \Rightarrow q'$, there exists a state $u'$ such that $u \Rightarrow u'$ and $q' \sim u'$. In other words, a stable equivalence is a bisimulation of the time-abstract transition system.

Suppose $A$ is a timed automaton, and $\sim$ is a stable partition of $Q_A$. Consider two equivalence classes $\pi$ and $\pi'$ of the relation $\sim$. Due to stability, $\pi$ contains a state $q$ such that $q \Rightarrow q'$ for some $q' \in \pi'$ iff for every $q$ in $\pi$, $q \Rightarrow q'$ for some $q' \in \pi'$. This implies that to solve reachability problems, we need to keep track of only the equivalence class rather than individual states. The *quotient* of $U_A$ with respect to a stable partition $\sim$ is the transition system $[U_A]_\sim$: states of $[U_A]_\sim$ are the equivalence classes of $\sim$, an equivalence class $\pi$ is an initial state of $[U_A]_\sim$ if $\pi$ contains an initial state of $U_A$, the set of labels is $\Sigma$, and $[U_A]_\sim$ contains an $\alpha$-labeled transition from the equivalence class $\pi$ to the class $\pi'$ if for some $q \in \pi$ and $q' \in \pi'$, $q \Rightarrow q'$ holds in $U_A$.

To reduce the reachability problem $(A, L^F)$ to a reachability problem over the quotient with respect to $\sim$, we need to ensure, apart from stability, that $\sim$ does not equate target states with non-target states. An equivalence relation $\sim$ is said to be $L^F$-sensitive, for a set $L^F \subseteq L$ of target locations, if whenever $(s, \nu) \sim (s', \nu')$, either both $s$ and $s'$ belong to $L^F$, or both $s$ and $s'$ do not belong to $L^F$.

**Proposition 1 (Stable quotients)** Let $A$ be a timed automaton, $\sim$ be an equivalence relation over $Q_A$, and $L^F$ be a set of locations of $A$ such that $\sim$ is stable and $L^F$-sensitive. Then, a location in $L^F$ is reachable iff there exists an equivalence class $\pi$ of $\sim$ such that $\pi$ is reachable in the quotient $[U_A]_\sim$ and $\pi$ contains a state whose location is in $L^F$. ■
Consequently, to solve the reachability problem \((A, L^F)\), we search for an equivalence relation \(\sim\) that is stable, \(L^F\)-sensitive, and has only finitely many equivalence classes.

**Restriction to integer constants**

Recall that our definition of timed automata allows clock constraints which involve comparisons with rational constants. Given a timed automaton \(A\), we can multiply each constant by the least common multiple of denominators of all the constants appearing in the clock constraints of \(A\). This transformation leaves the answer to the reachability problem unchanged. Consequently, we can restrict ourselves to timed automata whose clock constraints involve only integer constants. Note that the largest constant in the transformed automaton is bounded by the product of the constants in the original automaton. It follows that the transformation causes at most a quadratic blow-up in the length of the encoding of the clock constraints.

**Region equivalence**

We define an equivalence relation on the state-space of an automaton that equates two states with the same location if they agree on the integral parts of all clock values and on the ordering of the fractional parts of all clock values. The integral parts of the clock values are needed to determine whether or not a particular clock constraint is met, whereas the ordering of the fractional parts is needed to decide which clock will change its integral part first. For example, if two clocks \(x\) and \(y\) are between 0 and 1 in a state, then a transition with clock constraint \((x = 1)\) can be followed by a transition with clock constraint \((y = 1)\), depending on whether or not the current clock values satisfy \((x < y)\). The integral parts of clock values can get arbitrarily large. But if a clock \(x\) is never compared with a constant greater than \(c\), then its actual value, once it exceeds \(c\), is of no consequence in deciding the allowed switches.

Now we formalize this notion. For any \(\delta \in \mathbb{R}\), \(fr(\delta)\) denotes the fractional part of \(\delta\), and \([\delta]\) denotes the integral part of \(\delta\); that is, \(\delta = [\delta] + fr(\delta)\). For each clock \(x \in X\), let \(c_x\) be the largest integer \(c\) such that \(x\) is compared with \(c\) in some clock constraint appearing in an invariant or a guard.

The equivalence relation \(\cong\), called the **region equivalence**, is defined over the set of all clock interpretations for \(X\). For two clock interpretations \(\nu\) and \(\nu'\), \(\nu \cong \nu'\) iff all the following conditions hold:

1. For all \(x \in X\), either \([\nu(x)]\) and \([\nu'(x)]\) are the same, or both \(\nu(x)\) and \(\nu'(x)\) exceed \(c_x\).
2. For all \(x, y \in X\) with \(\nu(x) \leq c_x\) and \(\nu(y) \leq c_y\), \(fr(\nu(x)) \leq fr(\nu(y))\) iff \(fr(\nu'(x)) \leq fr(\nu'(y))\).
3. For all \(x \in X\) with \(\nu(x) \leq c_x\), \(fr(\nu(x)) = 0\) iff \(fr(\nu'(x)) = 0\).

A **clock region** for \(A\) is an equivalence class of clock interpretations induced by \(\cong\). We will use \([\nu]\) to denote the clock region to which \(\nu\) belongs. Each region can be uniquely characterized by a (finite) set of clock constraints it satisfies. For example, consider a clock interpretation \(\nu\) over two clocks with \(\nu(x) = 0.3\) and \(\nu(y) = 0.7\). Every clock interpretation in \([\nu]\) satisfies the constraint \((0 < x < y < 1)\), and we will represent this region by \([0 < x < y < 1]\).

The nature of the equivalence classes can be best understood through an example. Consider a timed transition table with two clocks \(x\) and \(y\) with \(c_x = 2\) and \(c_y = 1\). The clock regions are shown in Figure 5.
Figure 5: Clock regions

Note that there are only a finite number of regions. Also note that for a clock constraint \( \varphi \) of \( A \), if \( \nu \equiv \nu' \) then \( \nu \) satisfies \( \varphi \) iff \( \nu' \) satisfies \( \varphi \). We say that a clock region satisfies a clock constraint \( \varphi \) iff every clock interpretation in the region satisfies \( \varphi \). Each region can be represented by specifying

1. for every clock \( x \), one clock constraint from the set
   \[ \{ x = c \mid c = 0, 1, \ldots, c_x \} \cup \{ c - 1 < x < c \mid c = 1, \ldots, c_x \} \cup \{ x > c_x \}, \]

2. for every pair of clocks \( x \) and \( y \) such that \( c - 1 < x < c \) and \( d - 1 < y < d \) appear in (1) for some \( c, d \), whether \( \text{fr}(x) \) is less than, equal to, or greater than \( \text{fr}(y) \).

By counting the number of possible combinations of equations of the above form, we conclude that the number of clock regions is bounded by \( k! \cdot 2^k \cdot 2^{(2c_x + 2)} \), where \( k \) is the number of clocks. Thus, the number of clock regions is exponential in the encoding of the clock constraints.

Region automaton

Region equivalence relation \( \cong \) over the clock interpretations is extended to an equivalence relation over the state-space by requiring equivalent states to have identical locations and region-equivalent clock interpretations: \( (s, \nu) \cong (s', \nu') \) iff \( s = s' \) and \( \nu \equiv \nu' \). The key property of region equivalence is its stability:

**Proposition 2 (Stability)** Region equivalence \( \cong \) is stable. ■

For intuitive understanding of stability, let us consider the regions of Figure 5. Two states belonging to the same region satisfy the same set of guards of \( A \), and hence, if a switch is possible from one state then the same switch is possible from the other. During the switch some clocks may be reset to 0. The effect of setting the clock \( x \) to 0 is projection onto the \( y \)-axis. Note that equivalent states project to equivalent states. Now consider the evolution of a state due to elapse of time. As time elapses, a state moves along the diagonally upwards direction since both clocks \( x \) and \( y \) increase at the same rate. For a given clock region, the sequence of regions encountered by such a translation along the diagonally upwards direction is the same irrespective of the choice of a state within the region.

The quotient \( [U_A]_{\cong} \) of a timed automaton with respect to the region equivalence is called the *region automaton* of \( A \), and is denoted \( R(A) \). The number of equivalence classes of \( \cong \) is finite, it is stable, and it is \( L^F \)-sensitive irrespective of the choice of the target locations. It follows that to solve the reachability problem \( (A, L^F) \), we can search the finite region automaton \( R(A) \).

Consider the timed automaton \( A_0 \) shown in Figure 6. The alphabet is \( \{a, b, c, d\} \). The corresponding region automaton \( R(A_0) \) is shown in Figure 7. Only the regions reachable from the initial region
Figure 6: The automaton $A_0$

Figure 7: The region automaton $R(A_0)$
\{s_0, [x = y = 0]\} are shown. Note that \(c_x = 1\) and \(c_y = 1\). The timing constraints of the automaton ensure that the switch from \(s_2\) to \(s_3\) is never taken. The only reachable region with state component \(s_2\) satisfies the constraints \([y = 1, x > 1]\), and this region has no outgoing edges. Thus the region automaton helps us in concluding that no switches can follow a \(b\)-labeled switch.

### Complexity of reachability

Suppose every constant in the clock constraints of \(A\) is bounded by \(c\). Then, the region automaton \(R(A)\) has at most \(n \cdot k! \cdot 2^k \cdot (2c + 2)^k\) vertices, which is \(n \cdot 2^{O(k \log (kc))}\). To solve the reachability problem \((A, L^F)\), we need to determine if the region automaton \(R(A)\) has a reachable region with location in \(L^F\). Such a search can be performed in time linear in the number of vertices and edges of the region automaton. Thus, the complexity of solving the reachability problem \((A, L^F)\) is linear in the number of locations of \(A\), exponential in the number of clocks, and exponential in the encoding of the constants. Technically, the reachability problem is \(\text{PSPACE}\)-complete. For membership in \(\text{PSPACE}\), note that the reachability problem reduces to a search in an exponential automaton. \(\text{PSPACE}\)-hardness is established in [AD94]. In fact, in [CY91], it is established that both sources of complexity, the number of clocks and the magnitudes of the constants, render \(\text{PSPACE}\)-hardness independently of each other.

**Theorem 3 (Complexity of Reachability)** Let \(A\) be a timed automaton with \(n\) locations and \(k\) clocks, and suppose every constant in the clock constraints of \(A\) is bounded by \(c\). Then, the reachability problem \((A, L^F)\) can be solved in time \(n \cdot 2^{O(k \log (kc))}\). The reachability problem \((A, L^F)\) is \(\text{PSPACE}\)-complete. □

In practice, the input automaton for the reachability problem is a product of component automata. Thus, given component automata \(A_i\), the solution to the reachability problem requires searching the region automaton \(R(\prod_i A_i)\). Observe that, while both the product construction and the region-automaton construction involve an exponential blow-up, the region automaton \(R(\prod_i A_i)\) is only singly-exponential in the descriptions of the component automata.

**Remark 4 (Choice of timing constraints and decidability)** The clock constraints in the enabling conditions and invariants of a timed automaton compare clocks with constants. Such constraints allow us to express (constant) lower and upper bounds on delays. For any generalization of the constraints, our analysis technique breaks down. In fact, if we allow constraints of the form \(x = 2y\) (a special case of linear constraints over clocks), then the reachability problem becomes undecidable [AD94]. □

### Implementation

As noted earlier, the input to a verification problem consists of a set of component automata \(A_i\), and the solution demands searching the region automaton \(R(\prod_i A_i)\). Let us first consider an enumerative implementation. A state of the region automaton \(R(\prod_i A_i)\) consists of locations of all the component automata, the integer parts of all the clocks, and the ordering of the fractional parts of all the clocks. Both the product construction and the region automaton construction is done on-the-fly, exploring successors of states of \(R(\prod_i A_i)\) only as needed.

Verification problems can be solved efficiently using symbolic search based on binary decision diagrams [BDD+92, McM93]. We sketch implementation of timed \(\text{COSPAN}\) [AK96] to illustrate the application of \(\text{BDD}\)-based search to timed reachability. The original \(\text{COSPAN}\) search engine can
take an input file containing descriptions of coordinating automata described in the language S/R, and perform a symbolic search based on BDDs with many optimizations. In timed version, the input language for describing automata admits annotations to express timing constraints. Suppose the input program $P$ consists of a collection of coordinating timed automata $A_i$. A preprocessor generates a new program $P'$. The automata in $P'$ are described in original S/R without any timing annotations. For each $A_i$, let $A'_i$ be the automaton without any timing annotations. The program $P'$ consists of automata $A'_i$, together with the description of a monitor automaton $A_R$. Suppose $||i:A_i$ has $k$ clocks, and all the constants are bounded by $c$. The automaton $A_R$ has $2k$ variables; $k$ variables ranging over $0..c$ that keep track of the integral parts of the clocks, and $k$ variables ranging over $1..k$ that give the ordering of the fractional parts. The update rules for these variables refer to the state-variables of the component automata. Searching the region automaton of $||i:A_i$ is semantically equivalent to searching the product of $||i:A_i$ with $A_R$. The details of the construction of the monitor automaton $A_R$ are quite intricate, and heavily dependent on the language S/R. However, let us emphasize that the preprocessing step does not involve search, and can be performed in linear time. Following the preprocessing step, the search engine of Cospan is used to perform the search on the input program $P'$ using BDDs.

**Zone automata**

Let us revisit the region automaton of Figure 7. The initial region contains a single state $(s_0, x = y = 0)$, and has three successors with location $s_1$ corresponding to the clock regions $[y = 0 \land x < 1]$, $[y = 0 \land x = 1]$, and $[y = 0 \land x > 1]$. One strategy is to collapse all these three regions together to obtain the union $[y = 0]$. Such convex unions of clock regions are called clock zones. A clock zone is a set of clock interpretations described by conjunction of constraints each of which puts a lower or upper bound on a clock or on difference of two clocks. Formally, the set of clock zones is generated by the grammar

$$\varphi ::= x < c \mid x \leq c \mid c < x \mid c \leq x \mid x - y < c \mid x - y \leq c \mid \varphi_1 \land \varphi_2.$$  

For a clock zone $\varphi$, the set of clock interpretations satisfying $\varphi$ will also be denoted $\varphi$. If $A$ has $k$ clocks, then the set $\varphi$ is a convex set in the $k$-dimensional euclidean space. Observe the following properties of clock zones.

1. Every clock region is a clock zone.
2. For a clock region $\pi$ and a clock zone $\varphi$, either $\pi$ is entirely contained in $\varphi$, or has an empty intersection with $\varphi$.
3. The intersection of two clock zones is a clock zone.
4. Every clock constraint used in the invariants of locations and guards of switches is a clock zone (follows from 1 and 3).
5. For two clock zones $\varphi$ and $\psi$, if the union $\varphi \cup \psi$ is convex, then it is a clock zone. It follows that the set of clock zones coincides with the set of convex unions of regions.

The reachability analysis based on clock zones uses the following three operations on zones.

- For two clock zones $\varphi$ and $\psi$, the intersection of the two zones is a zone, denoted $\varphi \land \psi$. 

12
• For a clock zone $\varphi$, let $\varphi \uparrow$ denote the set of clock interpretations $\nu + \delta$ for $\nu \in \varphi$ and $\delta \in \mathbb{R}$. Thus, $\varphi \uparrow$ denotes the set of clock interpretations obtained by letting time elapse from some clock interpretation in $\varphi$. The set of clock zones is closed under this operation: for a clock zone $\varphi$, the set $\varphi \uparrow$ is a clock zone.

• For a subset $\lambda$ of clocks and a clock zone $\varphi$, let $\varphi[\lambda := 0]$ denote the set of clock interpretations $\nu[\lambda := 0]$ for $\nu \in \varphi$. Verify that $\nu[\lambda := 0]$ is a clock zone.

A **zone** is a pair $(s, \varphi)$ for a location $s$ and a clock zone $\varphi$. We build a transition system whose states are zones. Consider a zone $(s, \varphi)$ and a switch $e = (s, a, \psi, \lambda, s')$ of $A$. Let $\text{suc}(\varphi, e)$ be the set of clock interpretations $\nu'$ such that for some $\nu \in \varphi$, the state $(s', \nu')$ can be reached from the state $(s, \nu)$ by letting time elapse and executing the switch $e$. That is, the set $(s', \text{suc}(\varphi, e))$ describes the successors of the zone $(s, \varphi)$ under the switch $e$. To obtain the set $\text{suc}(\varphi, e)$, we (i) intersect $\varphi$ with the invariant of $s$, (ii) let time elapse using $\uparrow$, (iii) take intersection with the invariant of $s$, (iv) take intersection with the guard $\psi$ of $e$, and (v) reset the clocks in $\lambda$. The first and third steps ensure that the invariant is satisfied during elapse of time (since the invariant is convex, it suffices to ensure that the start and the final states satisfy the invariant). Thus,

$$\text{suc}(\varphi, e) = \left((\varphi \land I(s)) \uparrow\right) \land \text{inv}(s) \land \psi)[\lambda := 0]$$

Thus, the key property of the clock zones is closure under successors with respect to switches.

**Proposition 4 (Zone Successor)** For a clock zone $\varphi$ and a switch $e$ of a timed automaton $A$, the set $\text{suc}(\varphi, e)$ of clock interpretations is a clock zone.

A zone automaton is obtained by adding edges between zones $(s, \varphi)$ and $(s', \text{suc}(\varphi, e))$. For a timed automaton $A$, the **zone automaton** $Z(A)$ is a transition system: states of $Z(A)$ are zones of $A$, for every initial location $s$ of $A$, the zone $(s, [X := 0])$ is an initial location of $Z(A)$, and for every switch $e = (s, a, \psi, \lambda, s')$ of $A$ and every clock zone $\varphi$, there is a transition $((s, \varphi), a, (s', \text{suc}(\varphi, e)))$.

Recall the automaton $A_0$ and its region automaton $R(A_0)$ of Figure 7. The reachable part of the zone automaton $Z(A_0)$ is shown in Figure 8. Note that, unlike the region automaton, in the zone automaton, each vertex has at most one successor per input symbol, and the number of vertices of $Z(A_0)$ is less than the number of vertices of $R(A_0)$.
Difference-bound matrices

Clock zones can be efficiently represented using matrices [Dil89]. Suppose the timed automaton $A$ has $k$ clocks, $x_1, \ldots, x_k$. Then a clock zone is represented by a $(k+1) \times (k+1)$ matrix $D$. For each $i$, the entry $D_{0i}$ gives an upper bound on the clock $x_i$, and the entry $D_{0i}$ gives a lower bound on the clock $x_i$. For every pair $i, j$, the entry $D_{ij}$ gives an upper bound on the difference of the clocks $x_i$ and $x_j$. To distinguish between a strict and a nonstrict bound (i.e., to distinguish between constraints such as $x < 2$ and $x \leq 2$), and allow for the possibility of absence of a bound, define the bounds-domain $\mathbb{D}$ to be $\mathbb{Z} \times \{0, 1\} \cup \{\infty\}$. The constant $\infty$ denotes the absence of a bound, the bound $(c, 1)$, for $c \in \mathbb{Z}$, denotes the nonstrict bound $\leq c$, and the bound $(c, 0)$ denotes the strict bound $< c$. A difference-bound matrix (DBM) $D$ is a $(k+1) \times (k+1)$ matrix $D$ whose entries are elements from $\mathbb{D}$. A clock interpretation $\nu$ satisfies a DBM $D$ iff for all $1 \leq i \leq k$, $x_i \leq D_{0i}$ and $-x_i \leq D_{0i}$, and for all $1 \leq i, j \leq k$, $x_i - x_j \leq D_{ij}$. Observe that every DBM represents a clock zone, and every clock zone is represented by some DBM.

As an example, consider the clock zone

$$(0 \leq x_1 < 2) \land (0 < x_2 < 1) \land (x_1 - x_2 \geq 0)$$

can be represented by the matrix $D$

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\infty$</td>
<td>(0,1)</td>
<td>(0,0)</td>
</tr>
<tr>
<td>1</td>
<td>(2,0)</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>2</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>$\infty$</td>
</tr>
</tbody>
</table>

The DBM representation of a clock zone is not unique. In our example, there are many implied constraints that are not reflected in the matrix $D$. For instance, since $x_1 < 2$, we can conclude $x_1 - x_2 < 2$. Thus, the entry $D_{12}$ can be updated from $\infty$ to $(2,0)$ without changing the set of satisfying clock interpretations. The following DBM $D'$ is equivalent to $D$:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,0)</td>
</tr>
<tr>
<td>1</td>
<td>(2,0)</td>
<td>(0,1)</td>
<td>(2,0)</td>
</tr>
<tr>
<td>2</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(0,1)</td>
</tr>
</tbody>
</table>

The matrix $D'$ is obtained from the matrix by “tightening” all the constraints. Such a tightening is obtained by observing that sum of the upper bounds on the clock differences $x_i - x_j$ and $x_j - x_i$ is an upper bound on the difference $x_i - x_i$. Matrices like $D'$ with tightest possible constraints are called canonical. To formalize this notion, we extend the operation of addition over integers to define addition over $\mathbb{D}$: adding $\infty$ to any element gives $\infty$, and $(i, j) + (i', j') = (i + i', j \land j')$. Similarly, we extend the comparison relation $<$ over integers to allow comparison of elements in $\mathbb{D}$: $(i, j) < \infty$, and $(i, j) < (i', j')$ iff $i < i'$, or $i = i'$ and $j < j'$.

The DBM $D$ is satisfiable if it represents a nonempty clock zone. It can be established that $D$ is unsatisfiable iff there exists a sequence $0 \leq i_1, i_2, \ldots, i_l \leq k$ of indices such that $D_{i_1i_2} + D_{i_2i_3} + \cdots + D_{i_li} < (0,1)$. The DBM $D$ is said to be canonical iff for all $0 \leq i, j, l \leq k$, $D_{il} \leq D_{ij} + D_{jl}$. Every satisfiable DBM has an equivalent canonical DBM. We use canonical DBMs to represent clock zones. Given a DBM, using classical algorithms for computing all-pairs shortest paths, we check whether the DBM is satisfiable, and if so, convert it into a canonical form. Two canonical DBMs $D$
and $D'$ are equivalent iff $D_{ij} = D'_{ij}$ for all $0 \leq i, j \leq k$. This test can be used during the search to determine if a zone has been visited earlier.

To implement reachability algorithm based on zones, given a DBM representation of a clock zone $\varphi$ and a switch $e$, we need to compute the successor zone $\text{succ}(\varphi, e)$. For this purpose, we show how to obtain conjunction of two zones, and the representation of the zones $\psi \uparrow$ and $\psi[\lambda := 0]$ from the representation of $\psi$.

- **Intersection:** Consider two canonical DBMs $D$ and $D'$. Define the DBM $D''$ such that $D''_{ij} = \min(D_{ij}, D'_{ij})$. Verify that $D''$, represents the intersection of the zones of $D$ and $D'$. The matrix $D''$ need not be in canonical form, and we need to check if it is satisfiable, and if so, reduce it to canonical form.

- **Time-elapsed:** Consider a DBM $D$, and suppose we wish to compute the matrix representation of $D \uparrow$. The set $D \uparrow$ contains all clock interpretations obtained from clock interpretations in $D$ by letting time elapse. Note that the upper bounds on individual clocks are no longer valid when time elapses, while the lower bounds on individual clocks and the upper bounds on clock differences stay unchanged due to elapse of time. Consequently, given $D$, we update, for each $1 \leq i \leq k$, the entry $D_{i0}$ to $\infty$, and canonicalize the resulting matrix to obtain $D \uparrow$.

- **Projection:** Let us consider the effect of resetting a set $\lambda$ of clocks to 0. Given a matrix $D$, to obtain $D[\lambda := 0]$, (i) for $x_i \in \lambda$, update $D_{i0}$ and $D_{0i}$ to $(0, 1)$, (ii) for $x_i, x_j \in \lambda$, update $D_{ij}$ to $(0, 1)$, and (iii) for $x_i \in \lambda$ and $x_j \not\in \lambda$, update $D_{ij}$ to $D_{0j}$ and $D_{ji}$ to $D_{j0}$. As usual, the resulting matrix needs to be made canonical.

Theoretically, the number of zones is exponential in the number of regions, and thus, the zone automaton may be exponentially bigger than the region automaton. However, in practice, the zone automaton has fewer reachable vertices, and thus, leads to an improved performance. Furthermore, while the number of clock regions grows with the magnitudes of the constants used in the clock constraints, experience indicates that the number of reachable zones is relatively insensitive to the magnitudes of constants. As in case of region automata, the zone automaton is constructed on-the-fly.

**Remark 5 (Dense vs discrete time)** Our choice of time domain is $\mathbb{R}$, the set of nonnegative real numbers. Alternatively, we could choose $\mathbb{Q}$, the set of rational numbers, and all of the results stay unchanged. The key property of the time domain, in our context, is its denseness, which implies that arbitrarily many events can happen at different times in any interval of nonzero length. On the other hand, if we choose $\mathbb{N}$, the set of nonnegative integers, to model time, we have a discrete-time model, and the flavor of the analysis problems changes quite a bit. In the dense-time model, reachability for timed automata is $\text{PSPACE}$, while universality is undecidable; in the discrete-time case, reachability for timed automata is still $\text{PSPACE}$, while universality is $\text{EXPSPACE}$. We believe that discrete-time models, while appropriate for scheduling applications, are inappropriate for modeling asynchronous applications such as asynchronous circuits. For verification of real-time systems using discrete-time models, see, for instance, [EMSS90, CC94]. In [HMP92], it is established that under certain restrictions the timed reachability problem has the same answer irrespective of whether the time domain is $\mathbb{N}$ or $\mathbb{R}$.

**Remark 6 (Minimization)** Suppose we wish to explicitly construct a representation of the state-space of a timed automaton. Then, instead of building the region or the zone automaton, we can employ a minimization algorithm that constructs the coarsest stable refinement of a given initial partition by refining it as needed [ACH+92, YL93, KL94, TY96].
3 Automata-Theoretic Verification

In the last section, we studied solutions to the reachability problem for timed automata. This is adequate to check safety properties of real-time systems. To verify liveness properties such as “if a request occurs infinitely often, so does the response” we need to consider nonterminating, infinite, executions. Specification and verification of both safety and liveness properties can be formulated in a uniform and elegant way using an automata-theoretic approach.

3.1 Verification via Automata Emptiness

In the linear time model, it is assumed that an execution can be completely modeled as a sequence of states or system events, called a trace. The behavior of the system is a set of such traces. Since a set of sequences is a formal language, this leads naturally to the use of automata for the specification and verification of systems. The more familiar definition of a formal language is as a set of finite words over some given (finite) alphabet. As opposed to this, an \( \omega \)-language consists of infinite words. Thus an \( \omega \)-language over a finite alphabet \( \Sigma \) is a subset of \( \Sigma^\omega \) — the set of all infinite words over \( \Sigma \). \( \omega \)-automata provide a finite representation for certain types of \( \omega \)-languages.

An \( \omega \)-automaton is essentially the same as a nondeterministic finite-state automaton, but with the acceptance condition modified suitably so as to handle infinite input words. We refer the reader to [Tho90] for a summary of the theory of \( \omega \)-regular languages, and to [VW86, Kur94] for its application to verification.

Timed languages

To introduce time in the automata-theoretic framework, we begin by defining timed words—indefinite sequences in which a real-valued time of occurrence is associated with each symbol. A time sequence \( \tau = \tau_1 \tau_2 \cdots \) is an infinite sequence of time values \( \tau_i \in \mathbb{R} \), satisfying the following constraints:

1. Monotonicity: \( \tau \) increases monotonically; that is, \( \tau_i \leq \tau_{i+1} \) for all \( i \geq 1 \).

2. Progress (divergence): For every \( t \in \mathbb{R} \), there is some \( i \geq 1 \) such that \( \tau_i > t \).

A timed word over an alphabet \( \Sigma \) is a pair \( (\sigma, \tau) \) consisting of an infinite word \( \sigma = \sigma_1 \sigma_2 \cdots \) over \( \Sigma \) and a time sequence \( \tau \). A timed language \( T \) over \( \Sigma \) is a set of timed words over \( \Sigma \). If each symbol \( \sigma_i \) is interpreted to denote an event occurrence then the corresponding component \( \tau_i \) is interpreted as the time of occurrence of \( \sigma_i \). The progress requirement ensures that we disallow infinitely many events to occur within a finite interval of time. For example, define a timed language \( T_1 \) to consist of all timed words in which \( a \) and \( b \) alternate, and for the successive pairs of \( a \) and \( b \), the time difference between \( a \) and \( b \) keeps increasing:

\[
T_1 = \{((ab)^\omega, \tau) \mid \text{for } i \geq 1, \ ((\tau_{2i} - \tau_{2i-1}) < (\tau_{2i+2} - \tau_{2i+1}))\}.
\]

The language-theoretic operations such as intersection, union, complementation are defined for timed languages as usual. In addition, we define the Utime operation which discards the time values associated with the symbols, that is, it considers the projection of a timed word \( (\sigma, \tau) \) on the first component: for a timed language \( T \) over \( \Sigma \), Utime(\( T \)) is the \( \omega \)-language consisting of words \( \sigma \) such that \( (\sigma, \tau) \in T \) for some time sequence \( \tau \). For instance, Utime(\( T_1 \)) consists of the single word \( (ab)^\omega \).
Timed Büchi automata

In Section 1, we studied timed transition tables as an operational model for real-time systems. The same definition can be used as an acceptor that reads timed words. Let $A$ be a timed automaton over the alphabet $\Sigma$. The automaton starts in an initial state. To read a timed word $(\sigma, \tau)$, at every step $i \geq 1$, the automaton must let time elapse equal to the difference $\tau_i - \tau_{i-1}$ (let $\tau_0 = 0$), and then execute some $\sigma_i$-labeled location-switch. For instance, the timed automaton of Figure 1 can successfully read a timed word $(\sigma, \tau)$ if $\sigma = (ab)^\omega$ and $1 \leq \tau_{2i} - \tau_{2i-1} \leq 2$ for all $i \geq 1$. Formally, the run of the automaton $A$ over the timed word $(\sigma, \tau)$ is an infinite sequence $r$:

$$q_0 \overset{\tau_1}{\rightarrow} q_1 \overset{\sigma_1}{\rightarrow} q_1 \overset{\tau_2}{\rightarrow} q_2 \overset{\sigma_2}{\rightarrow} q_2 \overset{\tau_3}{\rightarrow} q_3 \overset{\sigma_3}{\rightarrow} q_3 \rightarrow \cdots \overset{\tau_{n-1}}{\rightarrow} q_n \overset{\sigma_n}{\rightarrow} q_n \rightarrow \cdots$$

The run $r$ is initialized if $q_0$ is an initial state of $A$. The set $\inf(r) \subseteq L$ consists of those locations $s$ of $A$ for which $s$ is the location-component of $q_i$ for infinitely many indices $i$.

Different types of timed $\omega$-automata can be defined by adding acceptance conditions to the definition of timed automata. We will use Büchi acceptance. A Büchi condition consists of a set of locations, and requires accepting runs to visit one of these locations infinitely often. Büchi conditions are useful to specify weak-fairness requirement for resolution of choice in modeling of nondeterministic systems.

A timed Büchi automaton (TBA) consists of a timed automaton $A = (L, L^0, \Sigma, X, I, E)$ and a set $L^F \subseteq L$ of accepting locations of $A$. A run $r$ of a TBA over a timed word $(\sigma, \tau)$ is an accepting run iff the intersection $\inf(r) \cap L^F$ is nonempty. For a TBA $A$, the language $T(A)$ of timed words it accepts is defined to be the set

$$\{(\sigma, \tau) \mid A \text{ has an accepting initialized run over } (\sigma, \tau)\}.$$

For example, consider the automaton of Figure 2, and add the Büchi acceptance condition $\{s_0\}$. The accepting condition requires the location $s_0$ to be visited infinitely often. This rules out runs that loop at $s_3$ forever.

**Remark 7 (Implicit liveness)** Implicit in the definition of a timed automaton and its timed language, there are already two forms of liveness. First, since we require timed words to contain infinitely many symbols, every run must contain infinitely many switches. For instance, the automaton of Figure 1 cannot stay in the initial location $s$ forever even though its invariant permits so (if we interpret $a$ as input, we would like to allow the possibility of $a$ never occurring, which can be modeled by adding a self-loop on location $s$ labeled with an idling event). Second, the invariants of locations, together with the divergence of time sequences, ensure progress. For instance, the automaton of Figure 1 cannot stay in the location $s'$ forever even if add a self-loop on $s'$ labeled with an idling event. In fact, one can argue that adding explicit Büchi conditions, while useful for nondeterministic systems, is not needed for timed systems. However, let us note that accepting conditions are useful to specify requirements, and thus, timed Büchi automata provide a uniform framework for both system and its specification.

The product construction for timed automata is modified to obtain a product construction for timed Büchi automata so that a complex system can be defined as a product of TBAs. Let $A_1$ and $A_2$ be two timed Büchi automata. The location of the product, besides the locations of the component automata, contains a counter for cycling through the accepting sets of the component automata. The counter ranges over $\{0,1,2\}$, and is initially 0. It is updated from 0 to 1 when $A_1$
executes a switch to some location in its accepting set $L_1^F$, from 1 to 2 when $A_2$ executes a switch to some location in its accepting set $L_2^F$, and is reset to 0 on the subsequent switch. If the counter equals 2 infinitely often, we can conclude that both automata execute switches to their respective accepting locations infinitely often. The details can be found in [AD94].

**Specification and verification**

In the automata-theoretic approach, a system is modeled as a timed Büchi automaton $A$, where $A$ is typically a product of automata modeling system components. Verification corresponds to asking questions regarding the timed language $T(A)$. For instance, in the railroad controller example, $T(\text{GRC})$ consists of timed words over the alphabet \{approach, exit, in, out, lower, raise\}. Suppose we wish to prove the **bounded response property** that when the gate goes down, it is guaranteed to open within 10 units. This can be viewed as a requirement of the timed language $\tilde{T}(\text{GRC})$: in every timed word, the event down must be followed by an event up within 10 units. This property can be specified by a timed Büchi automaton. For ease of analysis, we require the designer to specify the complement of the desired property. Thus, the specification would be a TBA whose timed language is the set of timed words that **violate** the requirement. The specification, then, consists timed words $(\bar{\sigma}, \overline{\tau})$ such that for some $i$, $\sigma_i = \text{down}$, and for all $j > i$ such that $\tau_j - \tau_i < 10$, $\sigma_j \neq \text{up}$. The automaton corresponding to this specification is shown in Figure 9. The specification automaton starts in the initial location $s_0$. In the locations $s_0$ and $s_2$, it can synchronize with other components on any event, but in location $s_1$ it does not allow the event up. The accepting location is $s_2$, and a run leading to $s_2$ must contain the event down within 10 units. It follows that the language of the product of GRC with the specification automaton is empty iff GRC satisfies the bounded response property.

In summary, the input to the **automata-theoretic timing verification problem** consists of a collection of timed Büchi automata $A_i$ modeling the system components and the specification TBA $A_S$ that accepts undesirable timed words. The verification problem is to determine whether the intersection of the timed languages $T(\|iA_i)$ and $T(A_S)$ is empty.

**Untiming**

Consider the timing verification problem $(\|iA_i, A_S)$. The verification problem corresponds to checking whether the timed language of the product of $\|iA_i$ with $A_S$ is empty. Since we already know how to construct products of TBAs, it suffices to find a decision procedure to check emptiness of the timed language of a single timed Büchi automaton. Consider a TBA $A$. First, observe that $T(A)$ is empty iff Untime($T(A)$) is empty.
For a timed automaton $A$, its region automaton can be used to recognize $Untime(T(A))$. For this purpose, let us establish a correspondence between the runs of $A$ and the runs (i.e., infinite paths) of $R(A)$. For a run $r$ of $A$ of the form

$$q_0 \xrightarrow{\sigma_1} q_1 \xrightarrow{\sigma_2} q_2 \xrightarrow{\sigma_3} q_3 \rightarrow \cdots \xrightarrow{\sigma_{n-1}} q_{n-1} \xrightarrow{\sigma_n} q_n \rightarrow \cdots$$

define its projection $[r]$ to be the sequence

$$[q_0] \xrightarrow{\sigma_1} [q_1] \xrightarrow{\sigma_2} [q_2] \xrightarrow{\sigma_3} [q_3] \rightarrow \cdots \xrightarrow{\sigma_n} [q_n] \rightarrow \cdots$$

where, for a state $q$, $[q]$ denotes the region it belongs to. From the definition of the edge relation for $R(A)$, it follows that $[r]$ is a run of $R(A)$. Since time progresses without bound along $r$, every clock $x \in X$ is either reset infinitely often, or from a certain time onwards it increases without bound. Hence, for all $x \in X$, for infinitely many $i \geq 0$, $[q_i]$ satisfies the clock constraint $(x = 0) \lor (x > c_x)$. This prompts the following definition: an infinite sequence $\pi_0 \pi_1 \ldots$ of regions is progressive iff for each clock $x \in X$, there are infinitely many $i \geq 0$ such that $\pi_i$ satisfies $(x = 0) \lor (x > c_x)$. The correspondence between the runs of $A$ and the runs of $R(A)$ can be made precise now (see [AD94] for a proof):

**Proposition 5** An infinite sequence $r$ of regions is a progressive run of the region automaton $R(A)$ over a word $\overline{w}$ if and only if there exists a time sequence $\overline{r}$ and a run $r'$ of $A$ over $(\overline{w}, \overline{r})$ such that $r$ equals $[r']$.

Consider the region automaton $R(A_0)$ of Figure 7. Every run $r$ of $R(A_0)$ has a suffix of one of the following four forms: (i) the automaton cycles between the regions $(s_1, [y = 0 \land x \in [0, 1]])$ and $(s_2, [0 < y < x < 1])$, (ii) the automaton stays in the region $(s_3, [0 < y < 1 < x])$ using the self-loop, (iii) the automaton stays in the region $(s_4, [y = 0, x > 1])$ using the self-loop, or (iv) the automaton stays in the region $(s_3, [x > 1, y > 1])$. Only the case (iv) corresponds to the progressive runs. For runs of type (i), even though $y$ gets reset infinitely often, the value of $x$ is always less than 1. For runs of type (ii) or (iii), even though the value of $x$ is not bounded, the clock $y$ is reset only finitely often, and yet, its value is bounded. Thus every progressive run of $A_0$ corresponds to a run of $R(A_0)$ of type (iv).

Consequently, we need to add acceptance conditions to the region automaton so that only progressive runs satisfy the accepting conditions, and an accepting location of $A$ is visited infinitely often. For the region automaton $R(A_0)$ of Figure 7, since all states of $A_0$ are accepting, from the description of the progressive runs, it follows that $R(A_0)$ can be changed to a Büchi automaton by choosing the accepting set to consist of a single region $(s_3, [x > 1, y > 1])$. Consequently, $Untime(T(A_0))$ equals $(ac)^{d_c} d_e$. This leads to the main theorem for timed automata: the untimed language of a timed automaton is $\omega$-regular.

**Proposition 6 (From timed regular to $\omega$-regular)** Given a $\text{TBA}_A$ $A$, there exists a Büchi automaton over $\Sigma$ which accepts $Untime(T(A))$.

**Solving timing verification**

For a timed Büchi automaton $A$, the language $T(A)$ is nonempty if and only if the region automaton $R(A)$ has a reachable cycle that contains an accepting location of $A$, and for every clock $x \in X$, contains a region satisfying $(x = 0)$ or a region satisfying $(x > c_x)$. Such a search can be performed in time linear in the size of the region automaton using, for instance, the nested depth-first search.
algorithm of [CVWY92] to test Büchi emptiness. The complexity analysis of the timing verification problem is quite similar to the complexity analysis of the reachability problem. In particular, the complexity is linear in the number of locations of the product of all the automata, exponential in the total number of clocks, and exponential in the encoding of the clock constraints.

**Theorem 7 (Complexity of timing verification)** Let $(||i|A_i, A_S)$ be an instance of the timing verification problem such that there are $m$ component automata $A_i$. Suppose each of the $(m + 1)$ automata has at most $n$ locations and $k$ clocks, and suppose every constant in the clock constraints of all the automata is bounded by $c$. Then, the timing verification problem $(||i|A_i, A_S)$ can be solved in time $n^{m+1} \cdot 2^{O(km \log (kcm))}$. The timing verification problem $(||i|A_i, A_S)$ is PSPACE-complete.

The optimizations considered for the implementation of the solution to the reachability problem apply to the timing verification problem also. In particular, the construction of the region automaton of the product of the input automata is done on-the-fly, and the search can be done symbolically by introducing additional variables encoding clock constraints defining regions.

### 3.2 Theory of Timed Languages

In analogy with the class of languages accepted by Büchi automata, we call the class of timed languages accepted by TBAs timed regular languages: a timed language $T$ is a **timed regular language** iff $T = T(A)$ for some TBA $A$. There is a well developed theory of timed regular languages. Here, we explain sample results, and their connection to analysis of real-time systems.

#### Closure properties

Since timed automata can have multiple initial locations, the class of timed regular languages is closed under union. The product construction for TBAs can be used to define intersection of timed regular languages, and thus, the class of timed regular languages is closed under intersection. However, the class is not closed under complementation. The language accepted by the automaton of Figure 10 over $\{a\}$ is

$$\{(a^i, \tau) \mid \text{for some } 1 \leq i < j, \ (\tau_j = \tau_i + 1)\}.$$

The complement of this language cannot be characterized using a TBA. The complement needs to make sure that no pair of $a$’s is separated by distance 1. Since there is no bound on the number of $a$’s that can happen in a time period of length 1, keeping track of the times of all the $a$’s within the past 1 time unit would require an unbounded number of clocks.
**Proposition 8 (Closure properties)** The class of timed regular languages is closed under union and intersection, but not under complementation. ■

The nonclosure under complementation depends on the denseness of the time domain.

**Remark 8 (Alternative characterizations of timed regularity)** The class of $\omega$-regular languages is quite robust, and has many alternative characterizations (e.g. Büchi automata, Streett automata, the monadic second-order language S1S, $\omega$-regular expressions). Obtaining equivalent characterizations of the class of timed regular languages seems much harder. Replacing Büchi accepting condition with the more general forms such as Streett acceptance or Muller acceptance, does not add expressiveness (i.e. the class of timed languages definable by timed Muller automata equals timed regular languages). Attempts have been made to define equivalent formulations using timed version of S1S [WIl94] and timed regular expressions [AMC97]. The effect of allowing $\epsilon$-labeled (silent) switches, and related operations such as hiding is studied in [DGP97]. ■

**Decision problems**

As noted earlier, determining the emptiness of the timed language of a timed automaton reduces to searching for cycles in the region automaton, and is $\text{PSPACE}$-complete. Let us now consider the language inclusion problem for timed automata, that is, given two TBA $A_1$ and $A_2$, we wish to determine if $T(A_1)$ is a subset of $T(A_2)$. When the specification automaton describes the desirable behaviors, then the verification problem corresponds to language inclusion. Let us note that for ordinary automata (or for Büchi automata) to test whether the language of one automaton is contained in another, we test the emptiness of the language of the product of the first automaton with the complement of the latter. This strategy cannot be used for TBAs, as it is not possible to automatically complement a TBA. In fact, there is no algorithm for testing whether the language of one TBA is contained in another. This is because the language inclusion problem, and even the weaker universality problem (i.e. whether a timed automaton accepts all timed words over its alphabet), is undecidable (see [AD94] for a proof via reduction from the halting problem of 2-counter machines).

**Theorem 9 (Decision problems)** The emptiness problem for timed Büchi automata is $\text{PSPACE}$-complete. The universality, language-inclusion, and language-equivalence problems for timed Büchi automata are undecidable. ■

**Remark 9 (Finite timed words)** Instead of considering languages of infinite timed words, we can consider language of finite timed words. Such finitary timed languages are adequate for reasoning about safety properties of real-time systems. If both the system and the specification are defined as finitary timed languages, then verification corresponds to checking reachability, rather than cycle detection, in the region automaton of the product. It should be noted, however, that negative results concerning languages of timed automata, such as $\text{PSPACE}$-hardness of emptiness and undecidability of universality, continue to hold in the finitary version. ■

**Remark 10 (Timed bismulation and timed simulation)** While timed language equivalence is undecidable, stronger equivalences such as timed bisimulation and timed simulation are decidable. For a timed automaton $\mathcal{A}$, a timed bisimulation is an equivalence relation $\sim$ on the state-space $Q_\mathcal{A}$ such that whenever $q_1 \sim q_2$, (1) if $q_1 \xrightarrow{a} q'_1$ for $a \in \Sigma \cup \mathbb{R}$, then there exists $q'_2$ with $q_2 \xrightarrow{a} q'_2$ and $q'_1 \sim q'_2$, and (2) if $q_2 \xrightarrow{\epsilon} q'_2$ for $a \in \Sigma \cup \mathbb{R}$, then there exists $q'_1$ with $q_1 \xrightarrow{\epsilon} q'_1$ and $q'_1 \sim q'_2$. While the number of equivalence classes of the maximal timed bisimulation relation is infinite,
the problem of deciding whether there exists a timed bisimulation that relates two specified initial states is, surprisingly, decidable \cite{C92} (the algorithm involves analysis of the region automaton of the product space $Q(A) \times Q(A)$). The same proof technique is useful to obtain algorithms for checking existence of timed simulation \cite{TAKB96} (timed simulation relations are useful for establishing refinement between descriptions at different levels of abstractions). The complexity of deciding timed (bi)simulation is \textsc{ExpTime}. A hierarchy of approximations to timed bisimulation relation can be defined on the basis of the number of clocks that an observer must use to distinguish between two timed automata \cite{ACH94}. The impact of the precision of the observer’s clocks on the distinguishing ability is studied in \cite{LW93}.

**Deterministic automata**

The language inclusion problem is solvable if we use deterministic TBA as specification automata. Recall that in the untimed case a deterministic automaton has a single initial location, and from each location, given the next input symbol, the next location is uniquely determined. We want a similar criterion for determinism for timed automata: given a location and the next input symbol along with its time of occurrence, the state after the next switch should be uniquely determined. So we allow multiple switches starting at the same location with the same symbol, but require their clock constraints to be mutually exclusive so that at any time only one of these switches is enabled. A timed automaton is called \textit{deterministic} iff

1. it has only one initial location, and

2. for all $s \in L$, for all $a \in \Sigma$, for every pair of switches of the form $\langle s, a, \varphi_1, -, - \rangle$ and $\langle s, a, \varphi_2, -, - \rangle$, the clock constraints $\varphi_1$ and $\varphi_2$ are mutually exclusive (i.e., $\varphi_1 \land \varphi_2$ is unsatisfiable).

The automaton of Figure 11 is deterministic. The two $b$-labeled switches out of the location $s_1$ have mutually exclusive enabling conditions. By declaring the location $s_2$ to be the only accepting state, we can ensure that in every accepted timed word $((ab)^{\omega}, \tau)$, for infinitely many $i \geq 1$, $\tau_{2i} - \tau_{2i-1} \leq 1$.

Deterministic timed automata can be easily complemented because a deterministic timed automaton has at most one run over a given timed word. The algorithm for checking emptiness can be used to test whether the language of one TBA is included in the language of a deterministic TBA. More details regarding deterministic TBA can be found in \cite{AD94}. It is worth noting that there is no deterministic timed Büchi automaton whose language equals the language of the nondeterministic automaton of Figure 10.
Remark 11 (Event-clock automata) In [AFH97], a determinizable class of timed automata is obtained by restricting the use of clocks. The clocks of an event-clock automaton have a fixed, predefined association with the symbols of the input alphabet. The event-recording clock of the input symbol $a$ is a history variable whose value always equals the time of the last occurrence of $a$ relative to the current time; the event-predicting clock of $a$ is a prophecy variable whose value always equals the time of the next occurrence of $a$ relative to the current time (if no such occurrence exists, then the clock value is undefined). Thus, unlike a timed automaton, an event-clock automaton does not control the reassignments of its clocks, and, at each input symbol, all clock values of the automaton are determined solely by the input word. This property allows the determination of event-clock automata, which, in turn, leads to a complementation procedure. Indeed, the class of event-clock automata is closed under all boolean operations (timed automata are not closed under complement), and the language-inclusion problem is decidable (PSPACE-complete) for event-clock automata.

4 Tools and Applications

A variety of tools exist for specification and verification of real-time systems. We briefly discuss three that are most closely related to the approach discussed in this paper.

Timed COSPAN

The tool COSPAN (see [HHK96] for an overview) is an automata-based modeling and analysis tool developed at Bell Labs. The real-time extension of COSPAN is described in [AK96]. The analysis of timing constraints can be done either using the region automaton or using the zone automaton, and the search can be performed either by an on-the-fly enumerative routine or by a BDD-based symbolic routine. Two approximations are supported as heuristic improvements. First, instead of analyzing all the timing constraints at once, they are added incrementally, as needed, in an automatic way guided by the results of the previous iterations [AIKY95]. Second, the underlying continuous semantics is approximated, in a conservative way, by the integers, and the former is used only when necessary. A recent version of timed COSPAN supports compositional refinement checking of timed automata using homomorphisms [TAKB96].

KRONOS

The tool KRONOS, developed at VERIMAG, supports analysis of a set of communicating timed automata [DOTY96]. The analysis can be performed by searching the zone automaton by an enumerative or a symbolic routine. The tool supports model checking of the branching real-time temporal logic TCTL, and interfaces to a variety of process-algebraic notations such as ET-LOTOS. Additional heuristics include a variety of minimization algorithms to compute reduced state-space [TY96]. KRONOS is available publicly at http://www.imag.fr/VERIMAG/PEOPLE/Sergio.Yovine/kronos/.

UPPAAL

The UPPAAL toolkit is developed in collaboration between Aalborg University, Denmark and Uppsala University, Sweden [LPY97]. Safety and bounded liveness properties of communicating timed automata are checked by an on-the-fly reachability analysis of the zone automaton.
Compositional techniques are used to reduce the search space [LPY95]. The tool supports a well-developed graphical user interface and features such as simulation. UPPAAL is available publicly at http://www.docs.uu.se/docs/rtmv/uppaal/.

Applications

The methodology described in this paper is suitable for finding logical errors in communication protocols and asynchronous circuits. Examples of analyzed protocols include Philips audio transmission protocol, carrier-sense multiple-access with collision detection, and Bang-Engsen audio/video protocol (a detailed description of these and other case studies can be obtained from the homepages of Kronos or UPPAAL). The application of COSPAN to verification of the asynchronous communication on the STARI chip is reported in [TB97], and to a scheduling problem in telecommunication software is reported in [AJKO97].

5 Discussion

We conclude by brief descriptions of related topics.

Linear real-time temporal logics Linear temporal logic (LTL) [Pn77] is a popular formalism for writing requirements regarding computations of reactive systems. A variety of real-time extensions of LTL have been proposed for writing requirements of real-time systems [Ost90, Koy90, AH94, AFH96]. In particular, the real-time temporal logic Metric Interval Temporal Logic (MITL) admits temporal connectives such as always, eventually, and until, subscripted with intervals. A typical bounded-response requirement that “every request\( p \) must be followed by a response\( q \) within 3 time units” is expressed by the MITL formula

\[
\square(p \rightarrow \diamond_{\leq 3} q).
\]

To verify whether a real-time system modeled as a timed automaton\( A \) satisfies its specification given as a MITL formula\( \phi \), the model checking algorithm constructs a timed automaton\( A_{\neg \phi} \) that accepts all timed words that violate\( \phi \), and checks whether the product of\( A \) with\( A_{\neg \phi} \) has a nonempty language [AFH96]. The definition of MITL requires the subscripting intervals to be nonsingular. In fact, admitting singular intervals as subscript (e.g. formulas of the form\( \square(p \rightarrow \diamond_{= 1} q) \)) makes translation from MITL to timed automata impossible, and the satisfiability and model checking problems for the resulting logic are undecidable.

Branching real-time temporal logics Many tools for symbolic model checking employ the branching-time logic CTL [CE81, Q582] as a specification language. The real-time logic Timed Computation Tree Logic (TCTL) [ACD93] allows temporal connectives of CTL to be subscripted with intervals. For instance, the bounded response property that “every request\( p \) must be followed by a response\( q \) within 3 time units” is expressed by the TCTL formula

\[
\forall \square(p \rightarrow \forall \diamond_{\leq 3} q).
\]

It turns out that two states that are region-equivalent satisfy the same set of TCTL-formulas. Consequently, given a timed automaton\( A \) and a TCTL-formula\( \varphi \), the computation the set of states of\( A \) that satisfy\( \varphi \), can be performed by a labeling algorithm that labels the vertices of the region automaton\( R(A) \) with subformulas of\( \varphi \) starting with innermost subformulas [ACD93]. Alternatively, the symbolic model checking procedure computes the set of states
satisfying each subformula by a fixpoint routine that manipulates boolean combinations of zone constraints [HNSY94].

**Probabilistic models** Probabilistic extensions of timed automata allow modeling constraints such as “the delay between the input event $a$ and the output event $b$ is distributed uniformly between 1 to 2 seconds” (cf. [ACD91]). With introduction of probabilities, the semantics of the verification question changes. Given a probabilistic timed automaton $A$ and a specific automaton $A_S$ that accepts the undesirable behaviors, verification corresponds to establishing that the probability that the run of the system $A$ generates a word accepted by $A_S$ is zero. A modification of the cycle detection algorithm on the region automaton of the product of $A$ and $A_S$ can solve this problem [ACD91].

**Hybrid systems** The model of timed automata has been extended so that continuous variables other than clocks, such as temperature and imperfect clocks, can be modeled. *Hybrid automata* are useful in modeling discrete controllers embedded within continuously changing environment. Verifying correctness of hybrid automata is computationally more expensive than of timed automata, but in simple cases, such as the railroad controller, it allows reasoning with parametric bounds. We refer the reader to [ACH+95] for an introduction to hybrid automata, and to [HHW95] for an introduction to the verifier HYTech.

**Acknowledgements**

My research on timed automata has been in collaboration with Costas Courcoubetis, David Dill, Tom Henzinger, Bob Kurshan, and many others.

**References**


