Formal Methods Module II: Model Checking Ch. 10: **SMT-Based Model Checking**

Roberto Sebastiani

DISI, Università di Trento, Italy - roberto.sebastiani@unitn.it URL:http://disi.unitn.it/rseba/DIDATTICA/fm2021/ Teaching assistant: Giuseppe Spallitta - giuseppe.spallitta@unitn.it

M.S. in Computer Science, Mathematics, & Artificial Intelligence Systems Academic year 2020-2021

last update: Tuesday 1st June, 2021, 11:34

Copyright notice: some material (text, figures) displayed in these slides is courtesy of R. Alur, M. Benerecetti, A. Cimatti, M. Di Natale, P. Pandya, M. Pistore, M. Roveri, C. Tinelli, and S. Tonetta, who detain its copyright. Some exampes displayed in these slides are taken from [Clarke, Grunberg & Peled, "Model Checking", MIT Press], and their copyright is detained by the authors. All the other material is copyrighted by Roberto Sebastiani. Every commercial use of this material is strictly forbidden by the copyright laws without the authorization of the authors. No copy of these slides can be displayed in public without containing this copyright notice.

Motivations & Context

Background



SMT-Based Bounded Model Checking of Timed Systems

- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study



Motivations & Context

Background

3 SMT-Based Bounded Model Checking of Timed Systems

- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study

4 Exercises

Motivations

• Model Checking for Timed Systems:

- relevant improvements and results over the last decades
- historically, "explicit-state" search style, based on DBMs
 - notable examples: Kronos, Uppaal
- More recently, symbolic verification techniques:
 - extensions of decision diagrams
 - CDD, DDD, RED, ...
- Key problem: potential blow up in size
- A more recent and viable alternative to Binary Decision Diagrams: SAT-based MC
 - Bounded Model Checking (BMC), K-induction, IC3/PDR, ...

Context

First Idea: SMT-based BMC of Timed Systems [Audemard et al. 2002], [Sorea, MTCS'02], [Niebert et al.,FTRTFT'02]

Leverage the SAT-based BMC approach to Timed Systems by means of SMT Solvers

Extensions

- SMT eventually applied to other SAT-based MC techniques
 - K-Induction
 - interpolant-based
 - IC3/PDR
- SMT applied to a variety of domains:
 - hybrid systems
 - verification of SW (loop invariants/proof obbligations, ...)
 - hardware verification
- Nowadays SMT leading backend technology for FV

We restrict to BMC for Timed Systems only

Motivations & Context

2 Background



- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study

4 Exercises

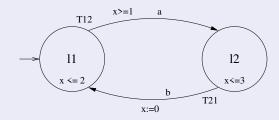
Bounded Model Checking [Biere et al., TACAS'99]

- Given a Kripke Structure *M*, an LTL property *f* and an integer bound *k*, is there an execution path of *M* of length (up to) *k* satisfying *f*? (*M* ⊨_k E*f*)
- Problem converted into the satisfiability of the Boolean formula: $[[M]]_{k}^{f} := I(s^{(0)}) \wedge \bigwedge_{i=0}^{k-1} R(s^{(i)}, s^{(i+1)}) \wedge (\neg L_{k} \wedge [[f]]_{k}^{0}) \vee \bigvee_{l=0}^{k} ({}_{l}L_{k} \wedge {}_{l}[[f]]_{k}^{0})$
 - s.t. ${}_{l}L_{k} \stackrel{\text{def}}{=} R(s^{(k)}, s^{(l)}), \ L_{k} \stackrel{\text{def}}{=} \bigvee_{l=0}^{k} {}_{l}L_{k}$
 - A satisfying assignment represents a satisfying execution path.
 - Test repeated for increasing values of k
 - Incomplete
 - Very effective for debugging, alternative to OBDDs
 - Complemented with K-Induction [Sheeran et al. 2000]
 - Further developments: IC3/PDR [Bradley, VMCAI 2011]

General Encoding for LTL Formulae

f	$[[f]]_k^i$	$I[[f]]_k^i$
р	p ⁽ⁱ⁾	$p^{(i)}$
$\neg p$	$\neg p^{(i)}$	$\neg p^{(i)}$
$h \wedge g$	$[[h]]_k^i \wedge \ [[g]]_k^i$	$I[[h]]_k^i \wedge I[[g]]_k^i$
$h \lor g$	$[[h]]_k^i \vee \ [[g]]_k^i$	$I[[h]]_k^i \vee I[[g]]_k^i$
Xg	$[[g]]_{k}^{i+1}$ if $i < k$	$\int_{k} [[g]]_{k}^{i+1}$ if $i < k$
Ng	\perp otherwise.	$[[g]]_k^l$ otherwise.
Gg	\perp	$\bigwedge_{j=\min(i,l)}^{k} I[[g]]_{k}^{j}$
Fg	$\bigvee_{j=i}^{k} [[g]]_{k}^{j}$	$\bigvee_{j=\min(i,l)}^{k} I[[g]]_{k}^{j}$
h U g	$\bigvee_{j=i}^k \left(\left[[g] \right]_k^j \wedge \bigwedge_{n=i}^{j-1} \left[[h] \right]_k^n \right)$	$\bigvee_{j=i}^k \left(I[[g]]_k^j \wedge \bigwedge_{n=i}^{j-1} I[[h]]_k^n \right) \lor$
		$\left \bigvee_{j=l}^{i-1} \left(\int_{\mathbb{R}} [[g]]_k^j \wedge \bigwedge_{n=i}^k \int_{n=i}^{n} \int_{\mathbb{R}} [[h]]_k^n \wedge \bigwedge_{n=l}^{j-1} \int_{\mathbb{R}} [[h]]_k^n \right) \right $
h R g	$\bigvee_{j=i}^k \left(\left[[h] \right]_k^j \land \bigwedge_{n=i}^j \left[[g] \right]_k^n \right)$	$\bigwedge_{j=\min(i,l)}^k [[g]]_k^j \lor$
		$\bigvee_{j=i}^{k} \left({}_{I}[[h]]_{k}^{j} \wedge \bigwedge_{n=i}^{j} {}_{I}[[g]]_{k}^{n} \right) \vee$
		$\bigvee_{j=l}^{i-1} \left(I[[h]]_k^j \wedge \bigwedge_{n=l}^k I[[g]]_k^n \wedge \bigwedge_{n=l}^j I[[g]]_k^n \right)$

Timed Automata [Alur and Dill, TCS'94; Alur, CAV'99]



- Clocks: real variables (ex. x)
- Locations:
 - label: (ex. *l*₁),
 - invariants: (conjunctive) constraints on clocks values (ex. $x \le 2$)
- Switches:
 - event labels (ex. a),
 - clock constraints (ex. $x \ge 1$),
 - reset statements (ex. x := 0)
- Time elapse: all clocks are increased by the same amount

\mathcal{LRA} -Formulae

[Audemard et al., CADE'02]; [Sorea, MTCS'02]; [Niebert et al., FTRTFT'02]

- *LRA*-formulae are Boolean combinations of
 - Boolean variables and
 - linear constraints over real variables (equalities and differences)

• e.g., $(x - 2 \cdot y \ge 4) \land ((x = y) \lor \neg A)$

- An interpretation ${\cal I}$ for a ${\cal LRA}$ formula assigns
 - truth values to Boolean variables
 - real values to numerical variables and constants

• e.g., I(x) = 3, I(y) = -1, $I(A) = \bot$

I satisfies a *LRA*-formula φ, written "*I* ⊨ φ", iff
 I(φ) evaluates to true under the standard semantics of Boolean
 and mathematical operators.

• E.g., $\mathcal{I}((x - 2 \cdot y \ge 4) \land ((x = y) \lor \neg A)) = \top$

Bottom level: a T-Solver for sets of LRA constraints

- E.g. $\{..., z_1 x_1 \le 6, z_2 x_2 \ge 8, x_1 = x_2, z_1 = z_2, ...\} \Longrightarrow unsat.$
- Combination of symbolic and numerical algorithms (equivalence class building, Belman-Ford, Simplex)

• Top level: a CDCL procedure for propositional satisfiability

- mathematical predicates treated as propositional atoms
- invokes \mathcal{T} -Solver on every assignment found
- used as an enumerator of assignments
- lots of enhancements

(see chapter on SMT)

Motivations & Context

Background



SMT-Based Bounded Model Checking of Timed Systems

- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study

4) Exercises

Motivations & Context

Background

SMT-Based Bounded Model Checking of Timed Systems
 Basic Ideas

- Basic Encoding
- Improved & Extended Encoding
- A Case-Study

4 Exercises

SMT-Based BMC for Timed Systems



- based on automata-theoretic approach for LTL
- [Niebert et al., FTRTFT'02]: encoding into \mathcal{DL}
 - limited to reachability

Disclaimer

These slides are adapted from [Audemard et al. FORTE'02]:

G. Audemard, A. Cimatti, A. Kornilowicz, R. Sebastiani Bounded Model Checking for Timed Systems, proc. FORTE 2002, Springer freely available as http://eprints.biblio.unitn.it/124/

Basic ingredients:

- An extension of propositional logic expressive enough to represent timed information: "*LRA*-formulae"
- A SMT(\mathcal{LRA}) solver for deciding \mathcal{LRA} -formulae \implies e.g., the MATHSAT solver
- An encoding from timed BMC problems into \mathcal{LRA} -formulae
 - \mathcal{LRA} -satisfiable iff an execution path within the bound exists

Motivations & Context

Background



SMT-Based Bounded Model Checking of Timed Systems

- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study

4) Exercises

Given a timed automaton A and a LTL formula f:

• The encoding [[*A*, *f*]]_{*k*} is obtained following the same schema as in propositional BMC:

$$[[A, f]]_{k} := I(s^{(0)}) \wedge \bigwedge_{i=0}^{k-1} R(s^{(i)}, s^{(i+1)}) \wedge (\neg L_{k} \wedge [[f]]_{k}^{0}) \vee \bigvee_{l=0}^{k} ({}_{l}L_{k} \wedge {}_{l}[[f]]_{k}^{0})$$

- $[[M, f]]_k$ is a \mathcal{LRA} -formula, where
 - Boolean variables encode the discrete part of the state of the automaton
 - constraints on real variables represent the temporal part of the state

Encoding: Boolean Variables

- Locations: an array \underline{I} of $n \stackrel{\text{def}}{=} \lceil log_2(|L|) \rceil$ Boolean variables
 - *I_i* holds iff the system is in the location *I_i*
 - ex: " $\neg l_i[3] \land l_i[2] \land \neg l_i[1] \land l_i[0]$ " means "the system is in location l_3 "
 - " $(\underline{l_i} = \overline{l_j})$ " stands for " $\bigwedge_n (\underline{l_i}[\overline{n}] \leftrightarrow l_j[n])$ ",
 - "primed" variables l_i' to represent location after transition
- Events: for each event $a \in \Sigma$, a Boolean variable <u>a</u>
 - <u>a</u> holds iff the system executes a switch with event a.
- Switches: for each switch ⟨*I_i*, *a*, φ, λ, *I_j*⟩ ∈ *E*, a Boolean variable *T*,
 - T holds iff the system executes the corresponding switch
- Time elapse and null transitions: two variables T_{δ} and T_{null}^{J}
 - *T*_δ holds iff time elapses by some δ > 0
 - T_{null}^{j} holds if and only A_{j} does nothing (specific for automaton A_{j})

Note: also for events, switches&transitions it is possible to use arrays of Boolean variables of size $\lceil log_2(|\Sigma|) \rceil$, $\lceil log_2(|E|+2) \rceil$ respectively

Encoding: Clock Values and Constraints

- Clocks values x are "normalized" wrt absolute time (x z):
 - a clock value is written as difference x z
 - z represents (the negation of) the absolute time
 - "offset" variable x represents (the negation of) the absolute time when the clock was reset
- Clock constraints reduce to $(x z \bowtie c)$, $\bowtie \in \{\le, \ge, <, >\}$, $c \in \mathbb{Z}$
- Clock reset conditions reduce to (x := z)
- Clock equalities like $(x_k = x_l)$ reduce to $(x_k z_k = x_l z_l)$
 - appear only in loops
 - only place where full \mathcal{LRA} is needed (rather than \mathcal{DL})
 - \implies for invariant checking (no loops) \mathcal{DL} suffices
- Encoding the effect of transitions:
 - with a time elapse transition

• z' < z, and x' = x

- otherwise:
 - z' = z, absolute time does not elapse
 - x' = z', if the clock is reset
 - x' = x, if the clock is not reset

Encoding: Initial Conditions

Initial condition I(s):

• Initially, the automaton is in an initial location:

• Initially, clocks have a null value:

$$\bigwedge_{x\in X} (x=z)$$

 $I_i \in L^0$

 I_i

Transition relation R(s, s'): Invariants

• Always, being in a location implies the corresponding constraints: $\bigwedge (l_i \rightarrow \bigwedge \psi)$.

 $\bigwedge_{l_i \in L} (\underline{l_i} \to \bigwedge_{\psi \in I(l_i)} \psi),$

Encoding: Transitions

Transition relation
$$T(s, s')$$
:
• Switches:

$$\bigwedge_{T \stackrel{\text{def}}{=} \langle l_i, a, \varphi, \lambda, l_j \rangle \in E} T \rightarrow \left(\underbrace{l_i \land \underline{a} \land \varphi \land \underline{l'_i} \land \bigwedge_{x \in \lambda} (x' = z') \land \bigwedge_{x \notin \lambda} (x' = x) \land (z' = z)}_{x \notin \lambda} \right)$$

• Time elapse:

$$\mathcal{T}_{\delta}
ightarrow \left((z'-z<0) \land (\underline{l'}=\underline{l}) \land \bigwedge_{x\in X} (x'=x) \land \bigwedge_{a\in \Sigma} \neg \underline{a}
ight)$$

• Null transition:

$$T^{j}_{null} \to \left((z'=z) \land (\underline{l}'=\underline{l}) \land \bigwedge_{x \in X} (x'=x) \land \bigwedge_{a \in \Sigma} \neg \underline{a} \right)$$

Encoding: Relations between Transitions

Mutual exclusion between events: $\bigwedge \quad (\neg \underline{a}_k \lor \neg \underline{a}_r)$ $a_k, a_r \in \Sigma, a_k \neq a_r$ At least one transition takes place: $T_{null}^{J} \vee T_{\delta} \vee \bigvee T$ T∈F Mutual exclusion between transitions: $\land (\neg T_k \lor \neg T_r)$ $T_k, T_r \in E \cup \{T_{pull}^j\} \cup \{T_\delta\}, T_k \neq T_r$

If events and transitions are encoded via arrays of Booleans, mutual exclusion constraints are not needed

Automata Product Construction

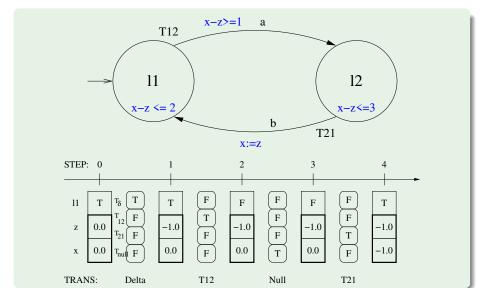
- The encoding is compositional wrt. product of automata
- The encoding of $A = A_1 ||A_2|$ is given by the conjunction of the encodings of A_1 and A_2 , plus a few extra axioms
- Mutual exclusion between events that are local

Forcing system activity:

$$\bigvee_{j=0}^{N-1} \neg T_{null}^{j}$$

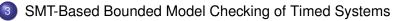
- one distinct T_{null}^{j} for each automaton A_{j}
- *T*_δ is common to all automata *A_j*

A Simple Example



Motivations & Context

Background



- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study

4) Exercises

Encoding: Extension

Adding Global Variables

Dealing with some global variable v on discrete domain:

• A switch
$$T \stackrel{\text{\tiny def}}{=} \langle I_i, a, \varphi, \lambda, I_j \rangle$$
 can

- be subject to a condition $\psi(\mathbf{v})$
- \implies add $T \rightarrow \psi(v)$
 - assign v to some value n or keep its value
- \implies add $T \rightarrow (v' = n)$ or add $T \rightarrow (v' = v)$
- T_{δ} mantains the value of v:
 - \implies add $T_{\delta} \rightarrow (v' = v)$
- T_{null}^{j} imposes no constraint on v:
 - \implies add nothing (for A_j)

MATHSAT: Optimizations

Customization of MATHSAT

 Limit Boolean variable-selection heuristic to pick transition variables, in forward order

Encoding: Optimizations

Boolean Propagation of Math Constraints: Idea: add small and mathematically-obvious lemmas

- \implies force assignments by unit-propagation,
- \implies saves calls to the \mathcal{T} -Solvers

Encoding Variants

Shortening counter-examples:

- Collapsing consequent time elapsing transitions:
 - $s \stackrel{\delta}{\longmapsto} s, s \stackrel{\delta'}{\longmapsto} s$ reduced to $s \stackrel{\delta+\delta'}{\longmapsto} s$
 - add $\neg T_{\delta} \lor \neg T'_{\delta}$ to transition relation R(s, s')
 - ⇒ implements the notion of "non-Zeno-ness" (see previous chapter)
- Allow multiple parallel transitions
 - remove mutex between labels local to processes
 - \implies allows a form of parallel progression

Remark: may change the notion of "next step"

Encoding Variants (cont.)

A limited form of symmetry reduction

If N automata are symmetric (frequent with protocol verification):

- Intuition: restrict executions s.t.
 - At step 0 only A₀ can move
 - At step 1 only A_0, A_1 can move
 - At step 2 only A_0, A_1, A_2 can move

• for step
$$i < N - 1$$
,

we drop the disjunct $\neg T_{null}^{i+1}$ (*i*) $\lor \ldots \lor \neg T_{null}^{N-1}$ (*i*):

set
$$\bigvee_{j=0}^{\min(i,N-1)} \neg T_{null}^{j(i)}$$
 rather than $\bigvee_{j=0}^{N-1} \neg T_{null}^{j(i)}$

- → drops "symmetric" executions
- \implies reduces the search space of a up to $2^{N(N-1)/2}$ factor!

Motivations & Context

Background



SMT-Based Bounded Model Checking of Timed Systems

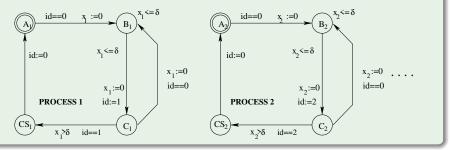
- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study



A Case-study: Fischer's Protocol

A Mutual-Exclusion Real-Time Protocol

- N identical processes accessing one critical section
- shared variable $id \in \{0, 1, 2, ..., N\}$: process identifier (0: none)
 - when entering wait state C_j, agent A_j writes its code on id
 - if id = j after δ , then A_j can enter the critical session
- Two properties under test
 - Reachability: **EF** $\bigwedge_i P_i . C$ (reached in N+1 steps)
 - Fairness: $\mathbf{E} \neg (\mathbf{GFP}_i.B \rightarrow \mathbf{GFP}_i.CS)$ (reached in N+5 steps)



Fischer's protocol: (cont.)

Exercise:

- Why is **EF** $\bigwedge_i P_i \cdot C$ reached in N+1 steps?
- Why is $\mathbf{E} \neg (\mathbf{GFP}_i.B \rightarrow \mathbf{GFP}_i.CS)$ reached in N+5 steps?

(See [Audemard et al, FORTE'02] for the solution.)

Fischer's protocol: (reachability)

 $M \models_k \mathbf{EF} \bigwedge_i P_i.C$

	Матн	SAT	Матн	SAT,Sym	DD)D	Upf	PAL	KRO	NOS	Re	D	Red,	Sym
N	Time	Size	Time	Size	Time	Size	Time	Size	Time	Size	Time	Size	Time	Size
3	0.05	2.9	0.04	2.9	0.11	106	0.01	1.7	0.01	0.8	0.23	2.0	0.19	2.0
4	0.09	3.0	0.08	3.0	0.14	106	0.02	1.9	0.02	2.2	1.00	2.1	0.70	2.1
5	0.20	3.2	0.16	3.2	0.24	106	0.21	1.9	0.09	19	3.70	2.2	2.00	2.4
6	0.60	3.7			0.47	106	3.44		0.39		12.00			
7	3.20	4.2	0.36		1.30	106	153			MEM	38	4.0		
8	29	4.9			3.96	106	TIME				121	7.6	-	
9	343	5.9	0.75	5.9		106					416			
	3331	6.5	-	6.5		106					1382	39	90	
11	TIME		1.39	7.0		106					TIME		157	38
12			1.89	7.5		MEM							266	
13			2.44	8.2									439	
14			3.24	8.9									709	155
15			4.11	9.7									1118	
16			5.10	10.7									1717	342
17			6.30	11.7									2582	492
18			8.00	12.9									TIME	
19			9.50	14.2										

(MATHSAT times are sum of all instances up to *k*)

Fischer's protocol (liveness violation)

$M \models_k \mathbf{E} \neg (\mathbf{GFP}_i.B \rightarrow \mathbf{GFP}_i.CS)$

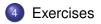
	MATHSAT						MATHSAT with Boehm heuristic					
$k \setminus N$	2	3	4	5	6	2	3	4	5	6		
2	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02		
3	0.01	0.02	0.01	0.01	0.03	0.01	0.01	0.02	0.03	0.04		
4	0.01	0.02	0.02	0.02	0.04	0.01	0.02	0.04	0.07	0.17		
5	0.02	0.03	0.05	0.09	0.18	0.01	0.03	0.09	0.30	1.16		
6	0.03	0.10	0.21	0.54	1.35	0.02	0.07	0.31	1.52	7.74		
7	0.04	0.26	0.97	3.20	9.83	0.02	0.18	1.19	7.14	45.00		
8		0.65	4.80	19.72	70.70		0.06	4.70	33.50	242.00		
9			5.55	112.17	478.00			0.61	165.90	1348.00		
10				303.17	3086.00				9.92	7824.00		
11					5002.00					252.00		
Σ	0.12	1.08	11.62	438.93	8648.15	0.07	0.37	6.98	218.40	9720.13		

Motivations & Context

Background

3 SMT-Based Bounded Model Checking of Timed Systems

- Basic Ideas
- Basic Encoding
- Improved & Extended Encoding
- A Case-Study



Proposed Exercise

Proposed Exercise

- Consider the Train-gate-controller example from [Alur CAV'99] (see previous chapter)
 - Encode the Initial state formula
 - Encode the transition relation
 - Encode the BMC problem for the formula ${f G}(s_2 o t_2)$
- As above, reducing the delay time for the controller from 1 to 0.5
 - what happens?
 - In how many steps?
- Encode the above into MathSAT