CS 477: Model Checking

Sasa Misailovic

Based on previous slides by Elsa Gunter and Armando Solar-Lezama (used with permission)

University of Illinois at Urbana-Champaign

Another good time for a recap

- Propositional Logic
- Operational Program Semantics
- Dataflow Analysis (CFG + finite-height lattice)
- Abstract Interpretation

 (abstraction/concretization + CFG + infinite-height lattice)
- First order logic, as an engine for solving constraints extracted from Axiomatic program semantics
- Axiomatic Semantics
- Coming up next....

Model Checking Today

Hardware Model Checking - part of the standard toolkit for hardware design

- Intel has used it for production chips since Pentium 4
- For the Intel Core i7, most pre-silicon validation was done through formal methods (i.e. Model Checking + Theorem Proving)
- Many commercial products

Software Model Checking

- Static driver verifier now a commercial Microsoft product
- Java PathFinder used to verify code for mars rover

History of Model Checking

 Clarke and Emerson, "Design and Synthesis of Synchronization Skeletons using branching time temporal logic"

"Proof Construction is Unnecessary in the case of finite state concurrent systems and can be replaced by a model-theoretic approach which will mechanically determine if the system meets a specification expressed in propositional temporal logic"

Obtained Turing Award

Precursors:

- Verification through exhaustive exploration of finite state models: G. V. Bochmann and J. Gecsei, A unified method for the specification and verification of protocols, Proc. IFIP Congress 1977
- Linear Temporal Logic, used for specifying system properties: A. Pnueli, The temporal semantics of concurrent programs. 1977

The model checking approach o (as characterized by Emerson)

- Start with a program that defines a <u>finite state model M</u>
- Search M for patterns that tell you whether a specification ϕ holds
- Pattern specification is flexible
- The method is efficient in the sizes of M and hopefully also $\boldsymbol{\phi}$
- The method is algorithmic

Model Checking

Most generally Model Checking is

- an automated technique, that given
- a finite-state model M of a system
- and a logical property ϕ ,
- checks whether the property holds of model: $M \models \, \phi$?
- or if it fails returns a counter-example (example of failure) – useful for debugging

Basic Notions of Model Theory

When an interpretation I makes S true, we say that I satisfies S

• or that I is a **model** of S (or I ⊨ S)

We are interested in deciding whether for the special case where

- I is a finite-state automaton with specific properties (e.g., Kripke structure or a labeled transition system)
- S is a temporal logic formula

High-level Idea:

- The program will determine the model through the translation to the transition system
- Recall, in axiomatic semantics, the program was a part of the theorem

Kripke Structures as Models

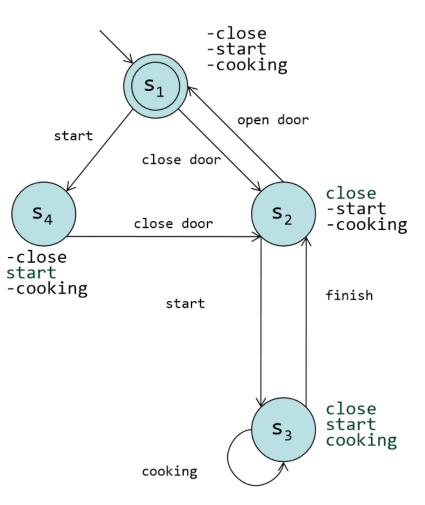
• Kripke structure is a finite size model with labels

For a set AP of atomic propositions, Kripke structure = (S, S₀, R, L)

- S : finite set of states
- $S_0 \subseteq S$: set of initial states
- $R \subseteq S \times S$: transition relation
- L : S → ℘ (AP) : labels *each state* with a set of atomic propositions

Microwave Example

- S = {s1, s2, s3, s4 }
- S₀={s1}
- R = { (s1,s2), (s2,s1), (s1,s4), (s4,s2), (s2,s3), (s3,s2), (s3,s3) }
- L(s1)={-close, -start, -cooking}
- L(s2)={close, -start, -cooking}
- L(s3)={close, start, cooking}
- L(s4)={-close, start, -cooking}

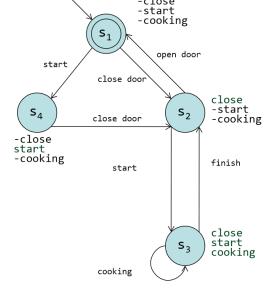


Q: Can the microwave cook with the door open (-close)?

Properties over States

State formula:

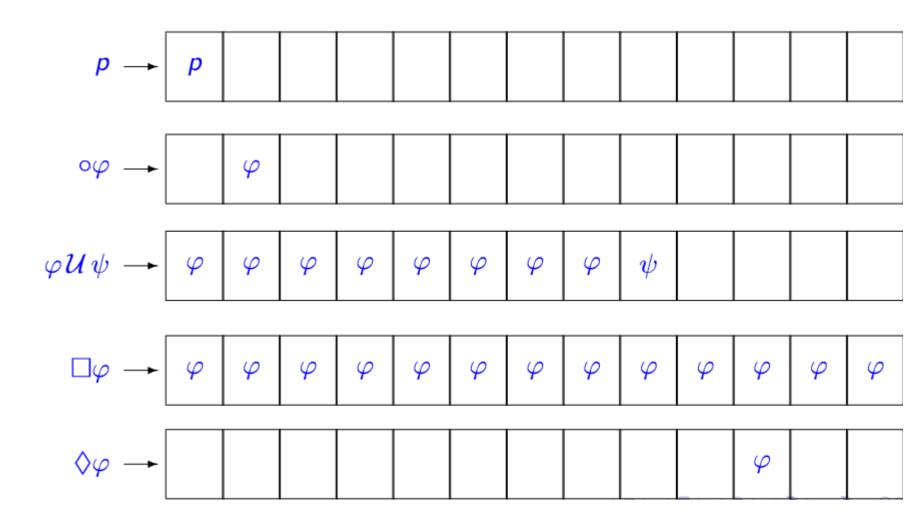
- Can be established as true or false on a given state
- If $p \in AP$ then p is a state formula
- if f and g are state formulas, so are (f and g), (not f), (f or g)
- E.g.: not close and cooking



Linear Time Logic Syntax

- $\varphi ::= p \mid (\varphi) \mid \neg \varphi \mid \varphi \land \varphi' \mid \varphi \lor \varphi'$ $\mid \circ \varphi \mid \varphi \cup \varphi' \mid \Box \varphi \mid \diamond \varphi$ or $\mid X \varphi \mid \varphi \cup \varphi' \mid G \varphi \mid F \varphi \leftarrow$ alternative notation
- p a proposition over state variables
- Standard negation, conjunction and disjunction
- $\circ \phi$ "next" (also denoted X ϕ)
- φUφ' "until"
- $\Box \phi$ "box", "always", "forever" (also G ϕ)
- $\diamond \phi$ "diamond", "eventually", "sometime" (also F ϕ)

Intuition



Paths and Path Formulas

Path π : a sequence of connected states: π := s₀, s₁, s₂, ...

Path formulae. Let f be a formula, which

- a state formula p is also a path formula: $\llbracket p \rrbracket(\pi_i) := \llbracket p \rrbracket(s_i)$
- boolean operations on path formulae are path formulae
 e.g. [[f □ g]](π_i) := [[f]](π_i) ∧ [[g]](π_i)
- path quantifiers

 $[\Box] \quad Gf(\pi_{i}) := globally f(\pi_{i}) = \forall k \ge i . [[f]](\pi_{k})$ $[\diamond] \quad Ff(\pi_{i}) := eventually f(\pi_{i}) = \exists k \ge i [[f]](\pi_{k})$ $[\circ] \quad Xf(\pi_{i}) := next f(\pi_{i}) = [[f]](\pi_{i+1})$ $f \cup g(\pi_{i}) := f until g = \exists k \ge i s.t. [[g]](\pi_{k}) and \forall j. i \le j < k \land [[f]](\pi_{j})$ Given a formula f and a path π if $[[f]](\pi)$ is true, then $\pi \models f$

Liveness Vs. Safety

Two very common terms:

Safety:

- Something **bad will never happen**: *G* ¬*bad*
- If it fails to hold, it's easy to produce a witness

Liveness:

- Something good will eventually happen: F good
- What does a witness for this look like? (¬ F good)

Box vs. Diamond

In another notation:

 $\bullet \Diamond p \Leftrightarrow \neg \Box \neg p \qquad \bullet \mathsf{F} p \Leftrightarrow \neg \mathsf{G} \neg p$

• $\Box p \Leftrightarrow \neg \Diamond \neg p$ • $G p \Leftrightarrow \neg G \neg p$

Common Combinations

- p will hold infinitely often:
 - G (F p)
- p will continuously hold from some point on
 F (G p)
- If p happens infinitely often, then so does q
 - (G p) \Rightarrow (G q)

LTL Examples

 If you submit your homework (submit) you eventually get a grade back (grade)

• You should get your grade before you submit the next homework

 If assignment i was submitted before drop date, you should get your grade before drop date

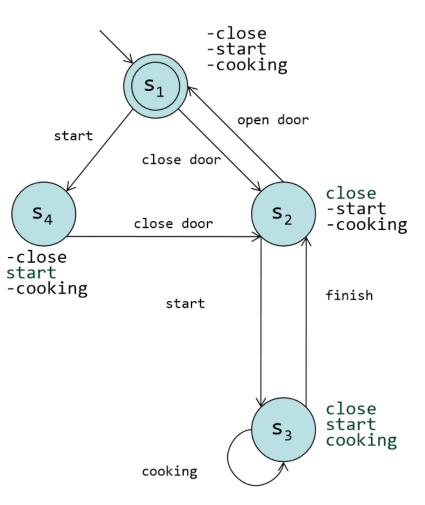
LTL Examples

- If you submit your homework (submit) you eventually get a grade back (grade)
 - G (submit \Rightarrow F grade)
- You should get your grade before you submit the next homework
 - G (submit \Rightarrow X (\neg submit U grade))

- If assignment i was submitted before the drop date, you should get your grade before the drop date
 - (G (submit_i \Rightarrow F dropDate)) \Rightarrow (G (grade_i \Rightarrow F dropDate))
 - and G (submit_i \Rightarrow F grade_i)

Microwave Example

- S = {s1, s2, s3, s4 }
- S₀={s1}
- R = { (s1,s2), (s2,s1), (s1,s4), (s4,s2), (s2,s3), (s3,s2), (s3,s3) }
- L(s1)={-close, -start, -cooking}
- L(s2)={close, -start, -cooking}
- L(s3)={close, start, cooking}
- L(s4)={-close, start, -cooking}



Q: Can the microwave cook with the door open (-close)?

Microwave Example

Labeled Transition System **Kripke Structure** -close -close -start -start -cooking -cooking S_1 S_1 open door start close door close close -start **s**₂ S_4 -start S_2 S_4 close door -cooking -cooking -close -close start start -cooking finish -cooking start close close start S₃ start Sa cooking cooking cooking

Q: Can the microwave cook with the door open (-close)?

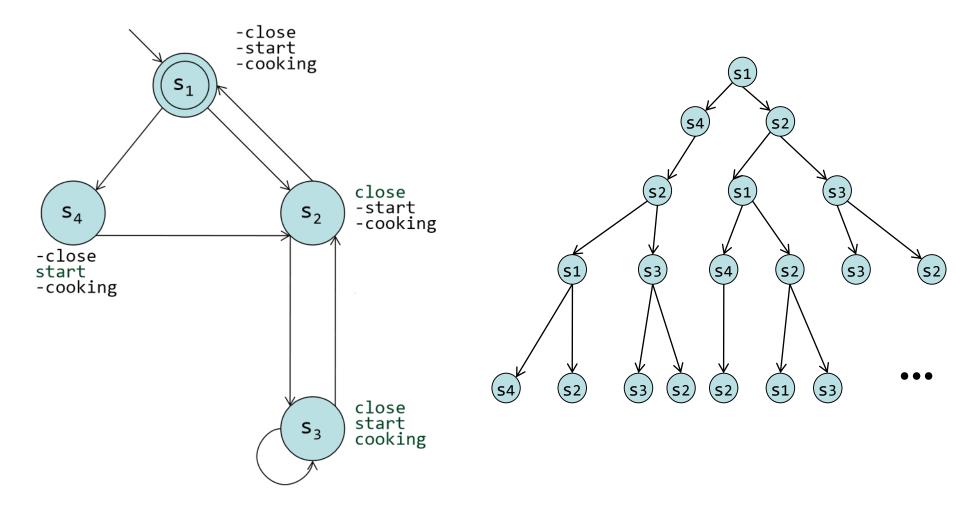
Reminder: Transition System (TS)

Describes potential system behaviors

- TS: Tuple (S, Θ, →): S is set of states, Θ ⊆ S are start states,
 → is a relation of state transitions
 - $\rightarrow \subseteq S \times S$ (we often write $s_1 \rightarrow s_2$ for $(s_1, s_2) \in \cdot \rightarrow \cdot$)
- Labeled TS: $(S, \Theta, \rightarrow, \Lambda)$: Λ is a set of labels
 - $\rightarrow \subseteq S \times \Lambda \times S$
 - we often write $s_1 \xrightarrow{\lambda} s_2$ for $(s_1, \lambda, s_2) \in \cdot \rightarrow \cdot$
- (recall) For atomic propositions set AP, Kripke structure = (S, S₀, R, L)
 - S : finite set of states; $S_0 \subseteq S$: set of initial states
 - $R \subseteq S \times S$: transition relation
 - L : S $\rightarrow \wp$ (AP) : labels *each state* with a set of atomic propositions

Why are Kripke Structures Enough?

• Can still represent all (finite or infinite) traces



Liveness Vs. Safety

- Two terms you are likely to run into:
- Safety:
 - Something bad will never happen: $G \neg bad$
 - If it fails to hold, it's easy to produce a witness
- Liveness:
 - Something good will eventually happen: *F* good
 - What does a witness for this look like?

Automata for LTL properties

- LTL defines properties over a trace
- Given a trace, we want to know whether it satisfies the property
- **Model checking:** Language(Model) ⊆ Language(Formula)
- Problem:
 - we need to build an automata to recognize infinite strings!
 - $\omega Regular$ Languages

Reminder: Finite State Machine (FSM)

- A FSM is a 5-tuple $\langle \Sigma, S, I, \delta, F \rangle$
- Σ is an **alphabet**
- S is a finite set of states
- $I \subseteq S$ is a set of **initial states**
- $\delta \subseteq S \times \Sigma \times S$ is a **transition relation**
- $F \subseteq S$ is a set of **accepting states**

Accepts the word w iff it ends in the accepting state after consuming the word

Buchi Automata

- Similar to a DFA
 - but with a stronger notion of acceptance
- In DFA, you have an accept state
 - when you reach accept state, you are done
 - this means you only accept finite strings
- In Buchi automata you also have accepting states but you only accept strings that visit the accept state infinitely often

(Non-deterministic) Buchi Automata

- A Buchi Automaton is a 5-tuple $\langle \Sigma, S, I, \delta, F \rangle$
 - Σ is an alphabet
 - S is a finite set of states
 - $I \subseteq S$ is a set of initial states
 - $\delta \subseteq S \times \Sigma \times S$ is a transition relation
 - $F \subseteq S$ is a set of accepting states
- <u>Non-deterministic Buchi Automata are not</u> <u>equivalent to deterministic ones</u>

Basic examples

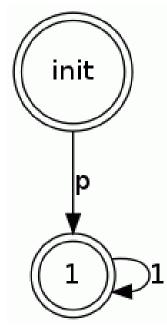
• p • F p

• p U q

• G p

Basic examples

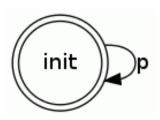
• p



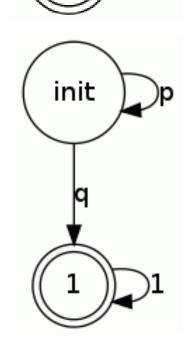
• F p

• p U q

• G p

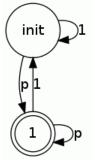


Graphs from: http://www.lsv.fr/~gastin/ltl2ba/index.php

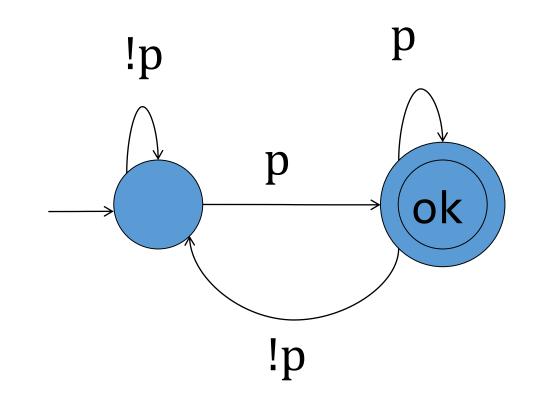


init

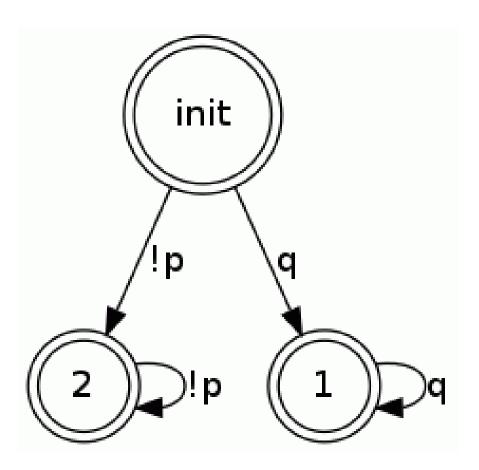
• G F p



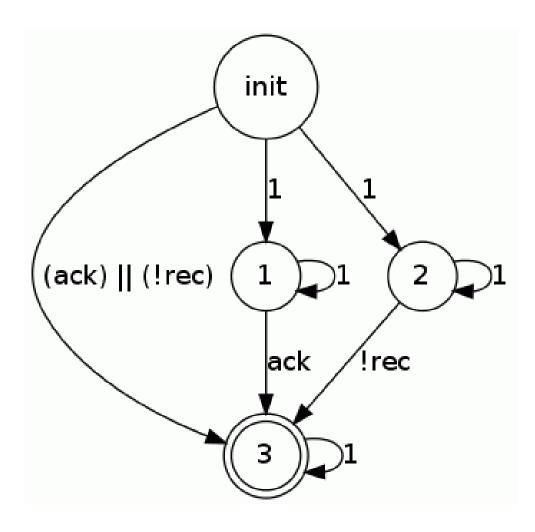
• G F p



• F p => G q



• G rec \Rightarrow F ack



CS 477: Model Checking

Sasa Misailovic

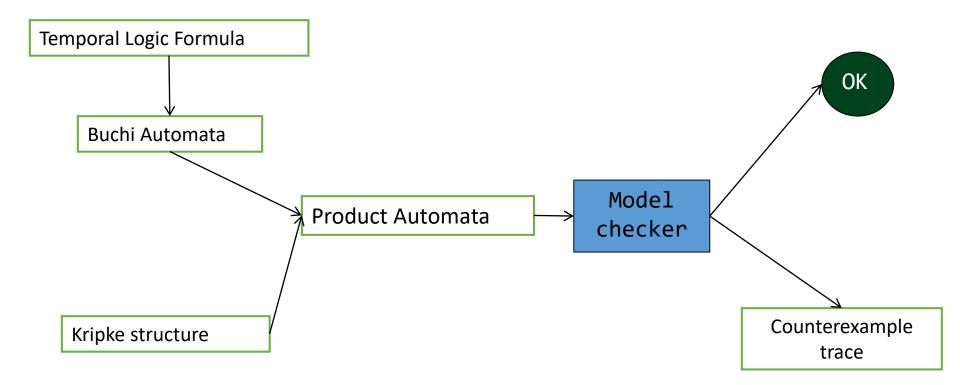
Based on previous slides by Elsa Gunter and Armando Solar-Lezama (used with permission)

University of Illinois at Urbana-Champaign

From LTL to automata

- Any LTL formula can be expressed as a nondeterministic Buchi automata (NBA)
- But the construction of the automata is complicated: exponential on the size of the formula
- See Vardi and Wolper, *Reasoning about infinite computations*, 1983.
- To visualize the formula: <u>http://www.lsv.fr/~gastin/ltl2ba/index.php</u>

Explicit State Model checking The Basic Strategy



A Bit About Complexity

- Satisfiability of a LTL formula: PSPACE-hard
- There is an algorithm that can solve the problem in $M \models F$ in O($|M| 2^{|F|}$)
- LTL and a fragment of FOL can express the same class of languages (infinite word languages)
 - But the ways of expressing the properties are different
 - For the detailed treatment, see Mahesh's notes: <u>https://courses.engr.illinois.edu/cs498mv/fa2018/LTL.pdf</u>

Proof System (Informational)

- First: Extend all rules of Propositional Logic to LTL
- Second Step: Add one more rule $\begin{array}{c} G \phi \\ \phi \end{array}$ Gen
- Third Step: Add a collection of axioms (a sufficient set of 8 exists)

A1:
$$G \varphi \Leftrightarrow \neg (F (\neg \varphi))$$

A2: $G (\varphi \Rightarrow \psi) \Rightarrow (G\varphi \Rightarrow G\psi)$
A3: $G \varphi \Rightarrow (\varphi \land X G \varphi)$
A4: $X \neg \varphi \Leftrightarrow \neg X\varphi$
A5: $X (\varphi \Rightarrow \psi) \Rightarrow (X \varphi \Rightarrow X \psi)$
A6: $G (\varphi \Rightarrow X \varphi) \Rightarrow (\varphi \Rightarrow G \varphi)$
A7: $\varphi \cup \psi \Leftrightarrow (\varphi \land \psi) \lor (\varphi \land X (\varphi \lor \psi)$
A8: $\varphi \cup \psi \Rightarrow F \psi$

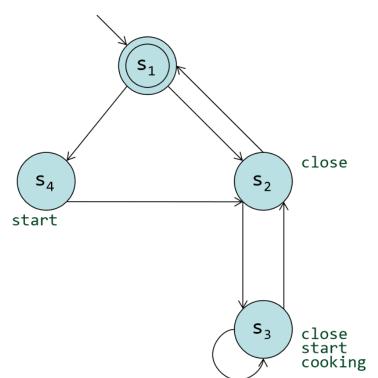
• Result: a sound and relatively complete proof system

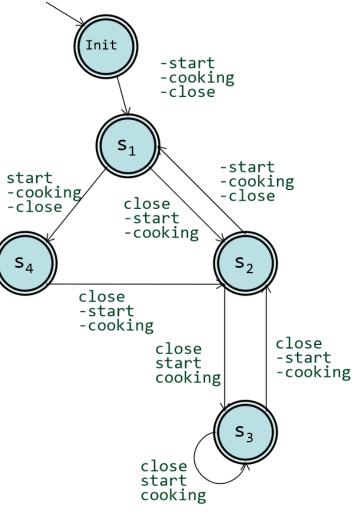
Buchi Automaton from Kripke Structure

- Given a Kripke structure:
 - $-M = (S, S_0, R, L)$
- Construct a Buchi Automaton
 - (Σ, SU {Init}, {Init}, T*, SU {Init})
 - T* is defined s.t.
 - T*(s, σ , s') iff R(s, s') and $\sigma \in L(s')$
 - T*(Init, σ ,s) iff s \in S0 and $\sigma \in$ L(s)

Buchi Automaton from Kripke Structure

- (Σ, S U {Init}, {Init}, T, S U {Init})
- T is defined s.t.
 - T(s, σ , s') iff R(s, s') and $\sigma \in L(s')$
 - T(Init, σ ,s) iff s \in S0 and $\sigma \in$ L(s)





Negated Property

- Given a good property P, you can define a bad property P'
- If the system has a trace that satisfies P', then it is buggy.
- Example
 - Good property: G(req \Rightarrow F ack)
 - Bad property: F (req ∧ (G ¬ack))
- We are going to ask whether M satisfies P'
 If it does, then we found a bug

Computing the Product Automaton

• Given Buchi automata A and B' - A = (Σ , S_A, T_A, {Init_A}, S_A)

- B' = (
$$\Sigma$$
 , S_B, T_B, {Init_B}, F')

- A x B' = (Σ , S_A x S_B, T, {(Init_A, Init_B)}, F)

• Where

- T((s₁,s₂), σ , (s₁', s₂')) iff T_A(s₁, σ , s₁') and T_B(s₂, σ , s₂')

$$-(s_1,s_2) \in F \text{ iff } s_2 \in F'$$

Check if a state is visited infinitely often

- Check for a cycle with an accepting state
- Cycle must be reachable from the initial state

Simple algorithm

- Do a depth-first search (DFS) to find an accepting state
- Do a DFS from that accepting state to see if it can reach itself

From Programs to Models

- Recall operational semantics
- Programs may have an infinite set of states (loops, recursion)
- To get a finite model, bound the number of iterations

Next...

- Continue the discussion from programs to models
- Following Jhala and Majumdar, "Software Model Checking" survey (most of text and examples from there)

Model Checking Concurrent Systems

- Based on Baier and Katoen Book
- Also find more on concurrent processes & SPIN in Moshe Vardi's online chapter: <u>https://cnx.org/contents/zVdF_TJw@3.4:13_CJz0Y</u> <u>@12/Concurrent-Programming-and-Verification-Outline</u>

Reminder: Transition System (TS)

Describes potential system behaviors

- TS: Tuple (S, Θ, →): S is set of states, Θ ⊆ S are start states,
 → is a relation of state transitions
 - $\rightarrow \subseteq S \times S$ (we often write $s_1 \rightarrow s_2$ for $(s_1, s_2) \in \cdot \rightarrow \cdot$)

- Labeled TS*: $(S, A, \Theta, \rightarrow, \Lambda)$: A is a set of actions, Λ is a set of node labels
 - $\rightarrow \subseteq S \times A \times S$
 - we often write $s_1 \xrightarrow{a} s_2$ for $(s_1, a, s_2) \in \xrightarrow{\cdot}$

*one of the versions

Parallel Systems

TS = TS1 || TS2

• Operator || denotes that the two transition systems execute in parallel

Concurrency and Interleaving

- Interleaving widely adopted model for concurrent systems
- Create the global system state that comprises of individual parallel components
- Pure interleaving: non-deterministic choice between the transitions of concurrent processes
- E.g., for two processes P, Q some legal interleavings:

Scheduler

- Decides on the interleavings
- The interleaving model of concurrency abstracts over the choice of real-world scheduler or performance of the processes
- It models all possible interactions
- The number of possible interleavings often leads to a combinatorial explosion

Interleaving of Transition System

- Let $TS_1 = (S_1, A_1, \Theta_1, \rightarrow_1, \Lambda_1)$ and $TS_2 = (S_2, A_2, \Theta_2, \rightarrow_2, \Lambda_2)$ be two labeled transition systems
- The interleaving transition system TS = TS1 || TS2

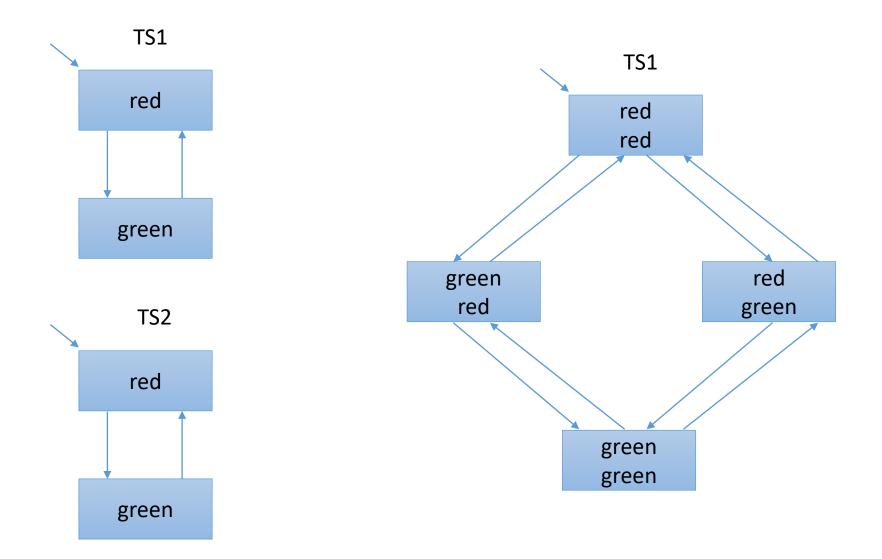
$$TS = (S_1 \times S_2, A_1 \cup A_2, \Theta_1 \times \Theta_2, \rightarrow , \Lambda)$$

where

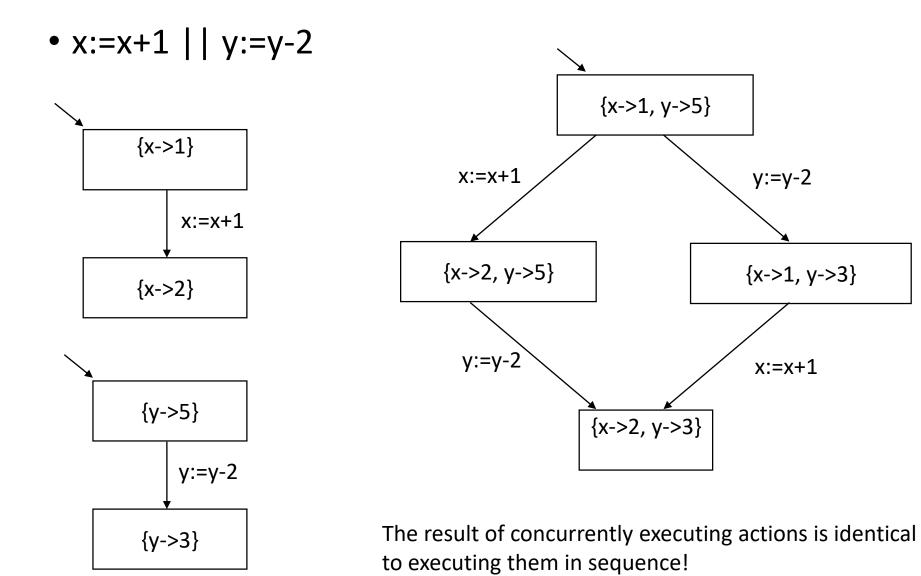
$$\frac{s_1 \stackrel{a}{\rightarrow}_1 s'_1}{(s_1, s_2) \stackrel{a}{\rightarrow} (s'_1, s_2)} \qquad \frac{s_2 \stackrel{a}{\rightarrow}_2 s'_2}{(s_1, s_2) \stackrel{a}{\rightarrow} (s_1, s'_2)}$$

and $\Lambda((s_1, s_2)) = \Lambda_1(s_1) \cup \Lambda_2(s_2)$

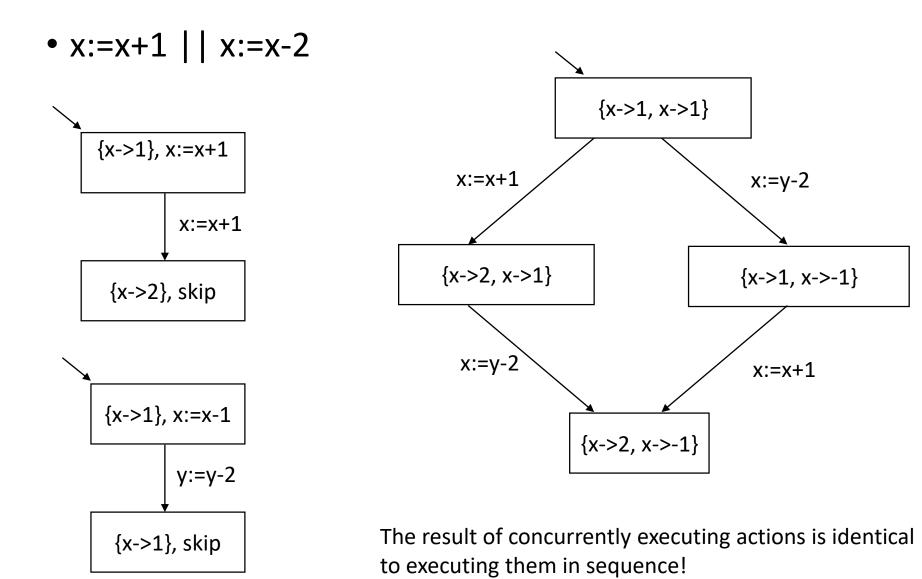
Example: Traffic Lights



How about programs? No sharing data!

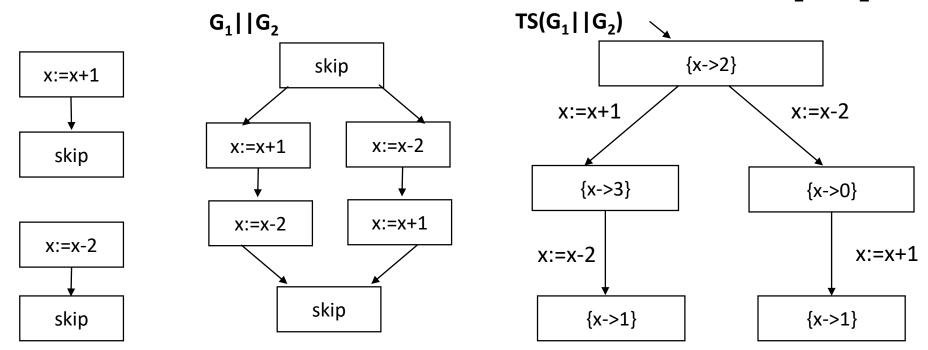


Communicate via Shared Variables



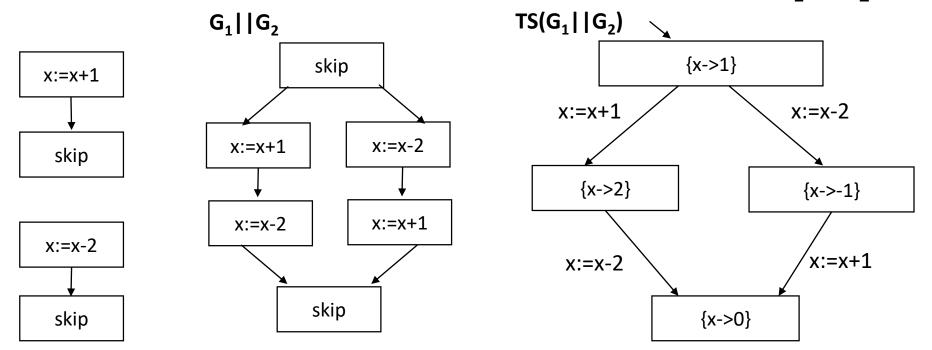
Communicate via Shared Variables

- Instead of defining || on transition systems, define it on the control flow graphs, G₁ and G₂
- The two graphs have a set of shared variables X' The other variables are private to G₁ and G₂
- Then obtain the underlying transition system $TS(G_1||G_2)$



Communicate via Shared Variables

- Instead of defining || on transition systems, define it on the control flow graphs, G₁ and G₂
- The two graphs have a set of shared variables X' The other variables are private to G₁ and G₂
- Then obtain the underlying transition system $TS(G_1||G_2)$



Communicate via Channels

- Asynchronous communication
- Extend the language with send (Expr, ProcessID) and x:=recv(ProcessID)
- Example:

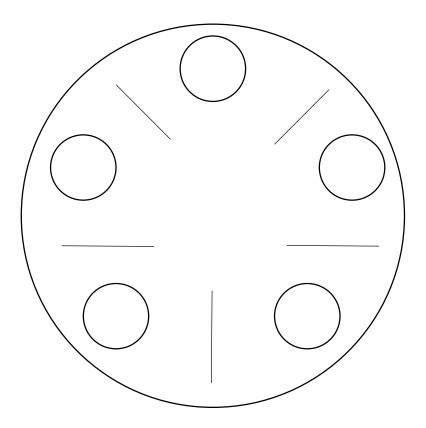
[send (1+1, #1); x:=recv(#1)]_{#0}
|| [x:=recv(#0); x:=x+1; send(x, #0)]_{#1}

- Each process state has a buffer (FIFO queue) with incoming messages from other messages
- To build the transition system, take the || of CFGs (since the channels are analogous to shared variables between processes) but all other variables are local
- And then convert to its corresponding transition system.

Deadlock: Dining Philosophers

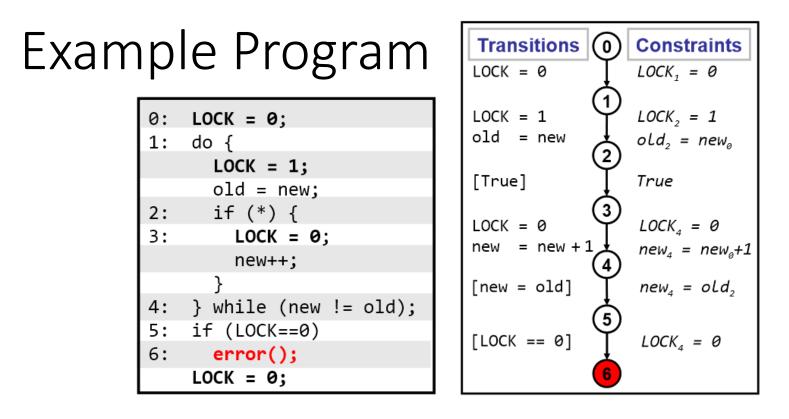
- A state of the system in which it cannot make a transition to perform a useful action
- Typically occurs when processes mutually wait for the others to proceed (e.g., release or aquire resources)
- Freedom of deadlocks is a weak form of liveness property
- If deadlocked, the program cannot reach the accepting state





Abstraction Refinement*

- When abstract analysis produces a counterexample, it can be
 - genuine, i.e., can be reproduced on the concrete program,
 - spurious, i.e., does not correspond to a real computation but arises due to imprecisions in the analysis
- If analysis imprecise, use the counterexample to refine it (CEGAR): use a new abstract domain that can represent strictly more sets of concrete program states [Ball and Rajamani 2000b; Clarke et al. 2000; Saidi 2000]



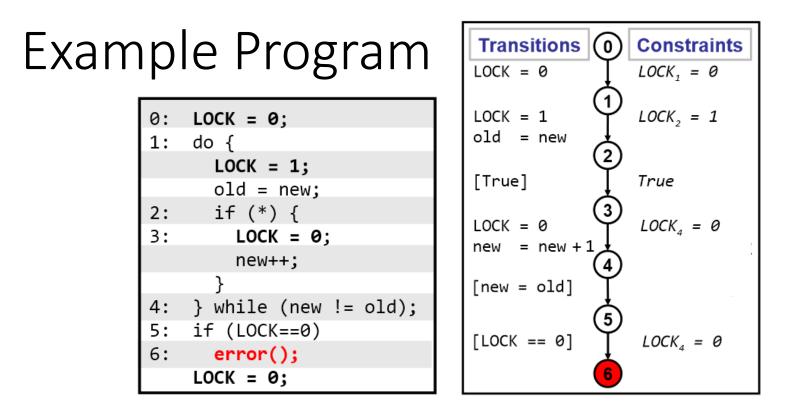




• "The abstraction of the set of reachable set of states:

(LOCK = $1 \land \text{new} = \text{old}$) V (LOCK = $0 \land \text{new} != \text{old}$) captures the intuition that at line 4, either the lock is acquired and new is equal to old, or the lock is not acquired and the value of new is different from old (in fact, new = old + 1)"

Abstraction insufficient if it tracks only the predicates LOCK = 0 and LOCK = 1







• "The abstraction of the set of reachable set of states:

(LOCK = $1 \land \text{new} = \text{old}$) V (LOCK = $0 \land \text{new} != \text{old}$) captures the intuition that at line 4, either the lock is acquired and new is equal to old, or the lock is not acquired and the value of new is different from old (in fact, new = old + 1)"

Abstraction insufficient if it tracks only the predicates LOCK = 0 and LOCK = 1

CEGAR Overview

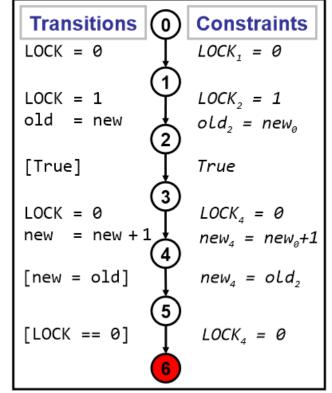
- Input to CEGAR algorithm: a path in the control flow graph that represents a possible counterexample produced by abstract reachability analysis.
- Step #1: constructs the <u>trace formula</u> from the path, such that the formula is satisfiable iff the path is executable by the concrete program.
- Step #2: a solver checks for satisfiability of the trace formula:
 - Yes -- the path is reported as a concrete counterexample to the property.
 - No -- the proof of unsatisfiability is mined for new predicates that can rule out the current counterexample when the abstract domain is augmented with these predicates.
- The CEGAR loop makes progress by eliminating at least one counterexample in each step.

Step 1: Trace Formula

• To convert an abstract counterexample into a trace formula, rename the state variables at each transition of the counterexample and conjoin the resulting transition to get

$$\bigwedge_{i=0}^{n-1} \rho_i(X_i, X_{i+1})$$

- Equivalent to the bounded model checking formula for the unrolled version of the program corresponding to the path, and also to the weakest precondition wp (unrolled_program, true)
- To avoid constraints of the form x_{i+1} = x_i for each x not modified by an operation, convert the path to static single-assignment (SSA) form
 - Thus: LOCK₁, LOCK₂, new₁, new₂



Step 2: Refinement (Syntax Based)

- Strategy: find an unsatisfiable core of atomic predicates appearing in the formula, whose conjunction is inconsistent.
- The trace formula is unsatisfiable as it contains the conjunction of

 $old_2 = new_0$, $new_4 = new_0 + 1$, $new_4 = old_2$

• Refinement: drop the subscripts [Henzinger et al. 2002] and the new predicates are

old = new, new = new + 1, new = old

 hen this predicate is added to the set of predicates, the resulting set

{LOCK = 0, LOCK = 1, new = old} suffices to refute the counterexample, i.e. the path is not a counterexample in the abstract model generated from these predicates.

