Model Checking

- Motivation for verification
- Computation Tree Logic — syntax and semantics
- Example: mutual exclusion
- A model checking algorithm

Verification Methods

Verification methods may be classified according to the following main criteria:

- **Proof-based vs. model-based** — if a soundness and completeness theorem holds, then:
  - proof = valid formula is true in all models
  - model-based = check satisfiability in one model
- **Degree of automation** — fully automated, partially automated, or manual
- **Full vs. property-verification** — a single property vs. full behavior
- **Domain of application** — hardware or software, sequential or concurrent, reactive or terminating, etc.
- **Pre- vs. post-development**

Kripke Structures

- Classical propositional and predicate calculi use a **unique** universe for interpreting formulas.
- In the 1950s, Kripke introduced a type of semantic models where more (local) universes are possible.
- There is a relation of **accessibility** between these universes and operators to express relationships between such universes, leading to various kinds of **modalities**.
- When such operators are added, one gets **modal logics**. **When** is the parameter that causes the passing from one universe to another; one speaks about **temporal logics**.

Kripke Structures for Programs

Programs (software) fit well in this framework:

- a universe corresponds to a state;
- the accessibility relation is given by the transition from one state to another;
- classic predicate logic may be used to specify relationships between variables in a state.

At this point we are lacking a mechanism to relate these universes (states). A variety of such mechanisms shall be introduced throughout this course.

Views of Time

We have the following characterizations of time:

- **linear** — a chain of time instances, or
- **branching** — several alternative future worlds may be possible at a given point in time;

or

- **discrete** — the time is represented by the set of integers, or
- **continuous** — time is represented by the set of real numbers.

Next, we shall study **Computation Tree Logic (CTL)** which is a type of temporal logic using branching and discrete time.
Model Checking

For any model checking problem (based on CTL, or other logic), we are required to answer the question of whether

$$M, s \models \Phi?$$

where

- $M$ is an appropriate model for the given system, and $s$ is a state of the model;
- $\Phi$ is a CTL formula intended to be satisfied by the system.

The Syntax of CTL

BNF definition of CTL:

$$\Phi ::= \perp | \top | p | (\neg \Phi) | (\Phi \land \Phi) | (\Phi \lor \Phi) | (\Phi \rightarrow \Phi)
| AX \Phi | EX \Phi | A[p U q] | E[p U q] | AG \Phi | EG \Phi | AF \Phi | EF \Phi$$

The new connectives $AX$, $EX$, $AU$, $EU$, $AG$, $EG$, $AF$, and $EF$ are called *temporal connectives*.

Depth and Breadth Quantification

The temporal connectives use two letters:

- $X$ and $U$ to quantify over the breadth in a branching point:
  - $AX$ alternatives in a branching point;
  - $EU$ exists at least one alternative in a branching point
- $G$ and $F$ to quantify along the path:
  - $AG$ all future states on a path (globally);
  - $EF$ there exists at least one future state along the path

Two more operators expressing properties along the paths are used:

- $N$ to refer to the next state in the path (this leads to the discrete feature of the time, and
- $U$ — the until operator.

Examples

1. $EG r$
2. $AG(q \rightarrow EG r)$
3. $A[p U q]$
4. $EF E[r U q]$
5. $A[p U EF r]$
6. $EF EGp \rightarrow AF r$
7. $AG AF r$
10. $AG(p \rightarrow A[p U (\neg p \land A[\neg p U q])])$

Priorities

Convention:

- the unary connectives (including $AX$, $EX$, $AG$, $EG$, $AF$, and $EF$) bind most tightly;
- next come $\land$ and $\lor$;
- lowest priority $\rightarrow$, $AU$ and $EU$.

Examples (2)

The parse tree of the formula

$$A[AX \neg p U (EF (p \land q) - p)]$$

$$A$$

$$AX$$

$$EX$$

$$EU$$

$$\neg$$

$$\land$$

$$p$$

$$q$$

$$A$$

$$A[p U p]$$

$$A[p U (\neg p \land A[p U p])]$$

$$AG(p \rightarrow A[p U (\neg p \land A[p U p])])$$

$$AG(q \rightarrow EG r)$$

$$AG(q \rightarrow AD r)$$

$$AG(q \rightarrow AF r)$$
CTL Semantics

A model $M = (S, \rightarrow, L)$ for CTL consists of:

- a set of states $S$
- a binary relation $\rightarrow$ on $S$ such that for every $s \in S$ there exists $s' \in S$ with $s \rightarrow s'$
- a labeling function $L: S \rightarrow 2^{\text{Atoms}}$

The intuition is that $L$ says which atoms are true in a state and $\rightarrow$ describes how the systems move from state to state.

Graphical description:

$$S = \{s_0, s_1, s_2\}$$
$$L(s_0) = \{p\}$$
$$L(s_1) = \{q\}$$
$$L(s_2) = \{r\}$$

The Satisfaction Relation

Let $M = (S, \rightarrow, L)$ be a model for CTL, $s \in S$, and $\phi$ a CTL formula. The satisfaction relation $M, s \models \phi$ is inductively defined by:

1. $M, s \models \top$ and $M, s \not\models \bot$, for all $s \in S$;
2. $M, s \models p$ iff $p \in L(s)$;
3. $M, s \models \neg \phi$ iff $M, s \not\models \phi$;
4. $M, s \models \phi \land \psi$ iff $M, s \models \phi$ and $M, s \models \psi$;
5. $M, s \models \phi \lor \psi$ iff $M, s \models \phi$ or $M, s \models \psi$;
6. $M, s \models \phi \rightarrow \psi$ iff $M, s \not\models \phi$ or $M, s \models \psi$;
7. $M, s \models \Box \phi$ iff for all $s'$ such that $s \rightarrow s'$ we have $M, s' \models \phi$;
8. $M, s \models \Diamond \phi$ iff for some $s'$ such that $s \rightarrow s'$ we have $M, s' \models \phi$.

The Satisfaction Relation (2)

9. $M, s \models \forall x. \phi$ iff for all paths $s' = s_1 \rightarrow s_2 \rightarrow \ldots$, we have $M, s_i \models \phi$, for all $i$;
10. $M, s \models \exists x. \phi$ iff there exists a path $s = s_1 \rightarrow s_2 \rightarrow \ldots$ such that $M, s_i \models \phi$, for some $i$;
11. $M, s \models \mathbf{A} \phi$ iff for all paths $s = s_1 \rightarrow s_2 \rightarrow \ldots$, we have $M, s_i \models \phi$, for some $i$;
12. $M, s \models \mathbf{E} \phi$ iff there exists a path $s = s_1 \rightarrow s_2 \rightarrow \ldots$ such that $M, s_i \models \phi$, for all $i$;
13. $M, s \models \Box (\phi \lor \psi)$ iff for all paths $s = s_1 \rightarrow s_2 \rightarrow \ldots$, there exists an $i$ such that $M, s_i \models \psi$ and $M, s_j \models \phi$ for all $j < i$;
14. $M, s \models \Diamond (\phi \lor \psi)$ iff there exists a path $s = s_1 \rightarrow s_2 \rightarrow \ldots$ and an $i$ such that $M, s_i \models \psi$ and $M, s_j \models \phi$ for all $j < i$.

Comments

- Notice that “the future” is the reflexive-transitive closure $\rightarrow^*$ of the (direct) accessibility relation $\rightarrow$.
- In common words:
  - the future contains the present
  - a future of a future of $s$ is a future of $s$.
- By unfolding [unwinding] the graph of a CTL model one gets an infinite tree, whence ‘computation tree logic’.

Unfolding

ACTL graph and its unfolding.

The Meaning of $\mathbf{E} \phi$, $\mathbf{G} \phi$, $\mathbf{A} \phi$, and $\mathbf{F} \phi$
Until

$p \lor q$ in a linear time model [or on a path in CTL].
The formula $p \lor q$ holds in $s_3$, but not in $s_0$ (we suppose $p$
holds only in the indicated states)

Examples

$M_{s_0} = p \land q$ - Yes
$M_{s_0} = \neg r$ - Yes
$M_{s_0} = T$ - Yes
$M_{s_0} = \text{EX} (q \land r)$ - Yes
$M_{s_0} = \neg \text{AX} (q \land r)$ - Yes
$M_{s_0} = \neg \text{EF} (p \land r)$ - Yes
$M_{s_0} = \text{EG} r$ - No
$M_{s_2} = \text{EG} r$ - Yes
$M_{s_2} = \text{AG} r$ - Yes
$M_{s_0} = \text{AF} r$ - Yes
$M_{s_0} = \mathcal{E}(p \land q) \land r$ - Yes
$M_{s_0} = \mathcal{A}(p \lor r)$ - Yes

Practical Patterns of Specifications

Examples of practically relevant properties that may be checked:

- it is possible to reach a state where $\text{started}$ holds, but $\text{ready}$ not:
  $\mathcal{E}(\text{started} \land \neg \text{ready})$

- for any state, if a $\text{request}$ occurs, then it will eventually be acknowledged:
  $\mathcal{A}(\neg \text{request} \rightarrow \text{AF acknowledged})$

- a certain process is enabled infinitely often on every computation path:
  $\mathcal{A}(\forall \text{ enabled})$

- whatever happens, a certain process will eventually be permanently deadlocked:
  $\mathcal{A}(\forall \text{ deadlocked})$

Useful Equivalences

**Definition:** Two CTL formulas $\phi$ and $\psi$ are *semantically equivalent*, denoted $\phi \equiv \psi$ if any state in any model that satisfies one of them also satisfies the other.

Useful equivalences:

1. $\neg \text{AF} \phi \equiv \text{EG} \neg \phi$
2. $\neg \text{EF} \phi \equiv \text{AG} \neg \phi$
3. $\neg \text{AX} \phi \equiv \text{EX} \neg \phi$
4. $\text{AF} \phi \equiv \text{A}[\text{T U } \phi]$
5. $\text{EF} \phi \equiv \text{E}[\text{T U } \phi]$

Adequate Sets of Temporal Connectives

**Corollary (adequate sets of temporal connectives):** The following sets of connectives are adequate for CTL (in the sense that each CTL formula may be transformed into an equivalent one using only those connectives):

- $\text{AU}$, $\text{EU}$ and $\text{EX}$
- $\text{EX}$, $\text{EU}$ and $\text{EX}$ (hint for proof: $\mathcal{A} [\phi \lor \psi] \equiv \neg \mathcal{E} [\neg \psi \lor (\neg \phi \land \neg \psi)] \lor \text{EG} \neg \psi)$
- $\text{AG}$, $\text{AU}$ and $\text{AX}$
- $\text{AF}$, $\text{EU}$ and $\text{AX}$
**Mutual Exclusion (1)**

**Goal:** to develop protocols for accessing some critical sections such that only one process can be in its critical section at a time. A collection of desirable properties is:

- **Safety:** the protocol allows only one process to be in its critical section at any time.
- **Liveness:** whenever any process wants to enter its critical section, it will eventually be permitted to do so.
- **Non-blocking:** a process can always request to enter its critical section.
- **No strict sequencing:** Processes need not enter their critical section in strict sequence.

**Mutual Exclusion (2)**

A simple model MUT1:

![Diagram of MUT1 model]

The system consists of two processes $P_1$ and $P_2$, each making a loop $a \rightarrow b \rightarrow c \rightarrow \ldots$ (noncritical $\rightarrow$ trying $\rightarrow$ critical $\rightarrow \ldots$). The system's behaviour is the product (interleaving) of the behaviours of $P_1$ and $P_2$, but the state $(c_1, c_2)$ is excluded.

**Mutual Exclusion (4)**

A second model MUT2:

![Diagram of MUT2 model]

**Mutual Exclusion (5)**

- This model is obtained by splitting state $s_3$ of MUT1 in two different states $s_3$ and $s_8$.
- By splitting $s_3$ into two states we are able to identify which process was the first asking to access its critical section; if $P_1$ was the first, the resulting state is $s_3$, otherwise $s_8$.
- **Fact:** All four properties (i.e., formulas $\phi_1$ to $\phi_4$) are valid in MUT2.
Model Checking Algorithms

Recall that the problem to be solved by model checking is

(A) Given a model $M$, a CTL formula $\phi$, and a state $s$, does $M, s \models \phi$ hold?

where

- $M$ is a model of the system and $s$ is a state of the model;
- $\phi$ is a CTL formula intended to be satisfied by the system

What is the model used for the system? Typically:

"A system is represented by a finite transition system (usually, a large labeled directed graph, often with millions of states)"

The infinite trees obtained by unfolding such graphs are useful to develop an intuition of the reasoning process, but not to be used on our finite computers.

Reduced Set of CTL Connectives

We are starting with a version using the following reduced set of CTL connectives:

$\Gamma = \{ \land, \lor, \sim, AF, EF, EX, E \land, E \lor \}$

where:

- $\land, \lor, \sim$ and $\land$ are used for the propositional part
- $AF, EF,$ and $EX$ are used for the temporal part

Hence, there is a preprocessing procedure to:

1. check the CTL syntax correctness of the given formula $\phi$
2. translate it into a formula $\text{TRANSLATE}(\phi)$ written with connectives $\Gamma$ only.

In the sequel, we suppose $\phi$ to be in $\Gamma$-format.

The Labeling Algorithm

The idea of this algorithm is to:

- decompose formula $\phi$ into pieces (sub-formulas) and apply a structural induction to label the graph with sub-formulas of $\phi$ (the intuition is that a formula that labels a state is true in that state)
- for each such sub-formula, parse the graph to infer the truth in a state according to the meaning of the connectives and the truth values of its sub-formulas

In 2, one may need to know the values of sub-formulas in possibly many different states; this is the case for temporal operators, but not for the propositional ones.

The Labeling Algorithm (2)

Input: a CTL model $M = (S, \rightarrow, L)$ and a CTL formula $\phi$

(in $\Gamma$-format)

Output: the set of states of $M$ which satisfy $\phi$

1. $\bot$: no states are labeled with $\bot$
2. $p$: label with $p$ all states $s$ such that $p \in L(s)$
3. $\sim \phi_1$: label $s$ with $\sim \phi_1$ if $s$ is not already labeled with $\phi_1$
4. $\phi_1 \land \phi_2$: label $s$ with $\phi_1 \land \phi_2$ if $s$ is already labeled both with $\phi_1$ and $\phi_2$
5. $EX \phi_1$: label $s$ with $EX \phi_1$ if one of its successors is already labeled with $\phi_1$

The Labeling Algorithm (3)

6. $AF \phi_1$:
   a. (initial marking) label any $s$ with $AF \phi_1$ if $s$ is already labeled with $\phi_1$
   b. (repeated marking) label any $s$ with $AF \phi_1$ if all successor states of $s$ are already labeled with $AF \phi_1$
   c. repeat (2) until there are no change

7. $E[\phi_1 U \phi_2]$
   a. (initial marking) label any $s$ with $E[\phi_1 U \phi_2]$ if $s$ is already labeled with $\phi_2$
   b. (repeated marking) label any $s$ with $E[\phi_1 U \phi_2]$ if $s$ is already labeled with $\phi_2$ and at least one of its successor states is already labeled with $E[\phi_1 U \phi_2]$
   c. repeat (2) until there is no change
A rough analysis of the algorithm shows that it has the worse time complexity 
\[ O(k \cdot m \cdot (m+n)) \]
where \( k \) is the number of connectives of the formula, \( m \) is the number of the states of the model, and \( n \) is the number of the transitions of the model.

### Directly

**EG may be handled directly as follows:**

1. \( 6^* \text{EG} \phi_1 \): 
   - 0. label all states \( s \) with \( \text{EG} \phi_1 \)
   - 1. (initial marking) if \( \phi_1 \) does not hold in \( s \) then delete the label \( \text{EG} \phi_1 \)
   - 2. (repeated marking) delete the label \( \text{EG} \phi_1 \) from any state \( s \) if none of its successor states is labeled with \( \text{EG} \phi_1 \)
   - 3. repeat (2) until there are no change

This different approach is based on the following greatest fixed-point characterization of \( \text{EG} \):

\[ \text{EG} \phi \equiv \phi \lor \text{EX} \text{EG} \phi \]

### An Improved Variant

- Use \( \text{EX}, \text{EU}, \) and \( \text{EG} \) instead of \( \text{EX}, \text{EU}, \) and \( \text{AF} \)
- Handle \( \text{EX} \) and \( \text{EU} \) as before using backwards breadth-first search
- for \( \text{EG} \phi \\
  - restrict to states satisfying \( \phi \)
  - find SCCs (maximally strongly connected components; these are maximal regions such that any vertex is connected to any other vertex in the region)
  - use backwards breadth-first searching on the restricted graph to find any state that can reach an SCC

Complexity is reduced to \( O(k \cdot (m+n)) \) \( \{k,m,n\} \) as before.

### Pseudo-Code

**function** \( \text{SAT}() \):

// preconditions: \( \phi \) is an arbitrary CTL formula /\n// postconditions: \( \text{SAT}(\phi) \) returns the set of states satisfying \( \phi \) /\nbegin function

case
  \( \phi \) is \( T \): \( \text{return} S \)
  \( \phi \) is \( \bot \): \( \text{return} \emptyset \)
  \( \phi \) is atomic formula: \( \text{return} \{ s \in S \mid \phi(s) \} \)
  \( \neg \phi_1 \): \( \text{return} S \setminus \text{SAT}(\phi_1) \)
  \( \phi_1 \land \phi_2 \): \( \text{return} \text{SAT}(\phi_1) \cap \text{SAT}(\phi_2) \)
  \( \phi_1 \lor \phi_2 \): \( \text{return} \text{SAT}(\phi_1) \cup \text{SAT}(\phi_2) \)
  \( \phi_1 \rightarrow \phi_2 \): \( \text{return} \text{SAT}(\neg \phi_1 \lor \phi_2) \)
end case
end function
function SAT_{\exists x} (\phi):
/* pre: \phi is an arbitrary CTL formula */
/* post: SAT_{\exists x} (\phi) returns the set of states satisfying \exists x \phi */
local var X, Y
begin
    X := SAT(\phi);
    Y := \{s_0 \in S | s_0 \rightarrow s_1 \text{ for some } s_1 \in X\};
    return Y
end

function SAT_{\forall \psi} (\phi, \psi) :
/* pre: \phi is an arbitrary CTL formula */
/* post: SAT_{\forall \psi} (\phi, \psi) returns the set of states satisfying E[\phi \forall \psi] */
local var W, X, Y
begin
    W := SAT(\psi);
    X := S;
    Y := SAT(\psi);
    repeat until X = Y
        begin
            X := Y;
            Y := Y \cup (W \cap \{s \in S | \exists s' : s \rightarrow s' \text{ and } s' \in Y\});
        end
    return Y
end

The State Explosion Problem

- the labeling algorithm is quite efficient [linear in the size of the model]
- ... but the model itself may be large, exponential in the number of the components (running in parallel 10 threads each of them having 10 states results in a system with 10^10 = 10,000,000,000 states!)
- the tendency of the state space to become very large is commonly referred to as the state explosion problem
- the state explosion problem is mainly unsolved - no general solution is known at the moment

Dealing With the State Explosion Code

The problem is generally unsolved. The following techniques were developed to overcome it in certain particular cases:

1. efficient data structures - e.g., ordered binary decision diagrams (OBDDs) are used to represent sets of states, not individual states
2. abstraction - one may abstract away variables in the model that are not relevant for the formula being checked
3. partial order reduction - different runnings may be equivalent as far as the formula to be checked is concerned; partial order reduction checks one trace from such a class only
4. induction - this technique is used when a large number of processes is considered
5. composition - try to split the problem in small parts to be separately checked