Principles of Safe Autonomy
ECE 498 SM
Lecture 2: System Safety

Professors: Sayan Mitra
Graduate Teaching Assistants: Yangge Li and Minghao Jiang
Plan for today

- Key concepts in assuring safety
  - Models, assumptions, requirements, invariants, counter-examples
- What would it take to assure safety of an autonomous system?
  1. Create a *model* of a simple safety scenario
  2. Identify the *requirements* and *assumptions*
  3. Analyze model to show that it meets the requirements under the assumptions
A “simple” safety scenario

A car moving down a straight road has to detect any pedestrian in front of it and stop before it collides.

Automatic Emergency Braking

Not a trivial requirement
MP0: Simulate model for testing
Our model

Parameters: $D_{\text{sense}}, a, b, x_0, x_2, v_0, \ldots$

State: $x_1, x_2, v_1, \ldots$; $d \triangleq x_2 - x_1$

Initial condition:

$x_1 = x_0$, $x_2 = x_2(0)$, $v_1 = v_0$

State transition rule

If $d < D_{\text{sense}}$

$v_1 = v_1 - ab$

Else

$v_1 = v_1$

Sampling time

$x_1 = x_1 + v_1 \cdot \Delta$  $\Delta = 1$
A model as a program

1 \text{SimpleCar}(D_{sense}, v_0, x_{10}, x_{20}, a_b), \ x_{20} > x_{10}
\textbf{initially:} \ x_1 = x_{10}, v_1 = v_0, x_2 = x_{20}, v_2 = 0
3 \quad s = 0, \ timer = 0
\textbf{if} d \leq D_{sense}
5 \quad s = 1
\quad \textbf{if} v_1 \geq a_b
7 \quad v_1 = v_1 - a_b
\quad \text{timer} = \text{timer} + 1
9 \quad \textbf{else}
\quad v_1 = 0
11 \quad x_1 = x_1 + v_1
Behaviors of the system model

An execution of the model captures a single run or behavior

An execution $\alpha$ is a sequence $x(0), x(1), \ldots$ such that

- $x(0)$ satisfies the initially clause, and
- for each $t$, $x(t)$ goes to or transitions to $x(t + 1)$ by executing SimpleCar

$x(t)$ is the complete state of the model at time $t$; $x(t)$. $v_1$ is the velocity at time $t$
“All models are wrong, some are useful.”
Wrong and useless models

FIGURE 4. A turkey using “evidence”; unaware of Thanksgiving, it is making “rigorous” future projections based on the past. Credit: George Nasr
Identifying requirements: Define safety

- A *requirement* is a precise statement about what the behaviors of the system should and should not do.

- An *invariant* is a requirement that something *always* holds. Examples:
  - “Car always remains far from the pedestrian”
  - “Drones never cross over to above 400ft in the airspace”
  - “A fully attentive safety driver should always be present during autonomy experiments”
How to prove that our model satisfies the requirement?

- An invariant is a requirement that something always holds.
  Examples:
  - “Car always remains far from the pedestrian”
  - Invariant 1. For all $x_{10}, x_{20}, v_0, D_{\text{sense}}, a_b$ and for all $t \ x(t).d > 0$

- Does this invariant hold? Why or why not?
  
  \[
  D_{\text{sense}} = x_{20} - x_{10} - \varepsilon
  \]
  \[
  v_0 = x_{20} - x_{10} + \varepsilon
  \]

- A counter-example is an execution that violates a requirement

- We will need to add some assumptions on model parameters
  $(x_{10}, x_{20}, v_0, D_{\text{sense}}, a_b)$ for Invariant 1 to hold (Homework)
Another invariant

▶ Invariant 2. \( \text{timer} + \frac{v_1}{a_b} \leq \frac{v_0}{a_b} \)

▶ Invariant 2. For all \( x_{10}, x_{20}, v_0, D_{\text{sense}}, a_b \) and for all \( t \),

\[
x(t). \text{timer} + \frac{x(t).v_1}{a_b} \leq \frac{v_0}{a_b}
\]

▶ How can we prove this invariant?
Invariant 2. For all \( x_{10}, x_{20}, v_0, D_{sense}, a_b \) and for all \( t \),
\[
\begin{align*}
\mathbf{x}(t) \cdot \text{timer} + \frac{x(t) \cdot v_1}{a_b} \leq \frac{v_0}{a_b} \\
\end{align*}
\]
Proof. Fix arbitrary \( x_{10}, x_{20}, v_0, D_{sense}, a_b \).

We have to show that \( \forall t \in \mathbb{N}, \mathbf{x}(t) \cdot \text{timer} + \frac{x(t) \cdot v_1}{a_b} \leq \frac{v_0}{a_b} \)

Use induction!

Recall, to show \( \forall t \in \mathbb{N}, P(t) \) it suffices to show that
\begin{itemize}
  \item \( P(0) \) and \( P(t) \Rightarrow P(t + 1) \)
  \item Here, \( P(\mathbf{x}(0)) \) and \( P(\mathbf{x}(t)) \Rightarrow P(\mathbf{x}(t + 1)) \)
  \item \( P(\mathbf{x}(0)) \) and \( P(\mathbf{x}(t)) \Rightarrow P(\text{SimpleCar}(\mathbf{x}(t))) \)
\end{itemize}
Writing the model more explicitly

1 SimpleCar($D_{\text{sense}}$, $v_0$, $x_{10}$, $x_{20}$, $a_b$), $x_{20} > x_{10}$

initially: $x_1 = x_{10}$, $v_1 = v_0$, $x_2 = x_{20}$, $v_2 = 0$

3 $s = 0$, $\text{timer} = 0$

if $d \leq D_{\text{sense}}$

5 $s = 1$

if $v_1 \geq a_b$

7 $v_1 = v_1 - a_b$

$\text{timer} = \text{timer} + 1$

9 else

$v_1 = 0$

11 $x_1 = x_1 + v_1$

---

1 SimpleCar($D_{\text{sense}}$, $v_0$, $x_{10}$, $x_{20}$, $a_b$), $x_{20} > x_{10}$

initially: $x_1(0) = x_{10}$, $v_1(0) = v_0$, $x_2(0) = x_{20}$, $v_2(0) = 0$

3 $s(0) = 0$, $\text{timer}(0) = 0$

$d(t) = x_2(t) - x_1(t)$

5 if $d(t) \leq D_{\text{sense}}$

7 $s(t + 1) = 1$

9 $v_1(t + 1) = v_1(t) - a_b$

else

$v_1(t + 1) = 0$

$\text{timer}(t + 1) = \text{timer}(t)$

13 else

$s(t + 1) = 0$

$\text{timer}(t + 1) = \text{timer}(t)$

15 $v_1(t + 1) = v_1(t)$

$\text{timer}(t + 1) = \text{timer}(t)$

if $x_1(t + 1) = x_1(t) + v_1(t)$
Proof (continued)

- **Invariant 2.** For all $x_{10}, x_{20}, v_0, D_{sense}, a_b$ and for all $t$, 
  \[ x(t) \cdot \text{timer} + \frac{x(t) \cdot v_1}{a_b} \leq \frac{v_0}{a_b} \]

- **Proof (continued).**

- **Base case.** $P(x(0))$
  \[ x(0) \cdot \text{timer} + \frac{x(0) \cdot v_1}{a_b} \]
  \[ = 0 + \frac{v_0}{a_b} \leq \frac{v_0}{a_b} \]

- **Induction.** $P(x(t)) \Rightarrow P(\text{SimpleCar}(x(t)))$
  Three cases to consider
  - $d > D_{sense}$
  - $d \leq D_{sense} \land v_1 \geq a_b$
  - $d \leq D_{sense} \land v_1 < a_b$

```
1 \text{SimpleCar}(D_{sense}, v_0, x_{10}, x_{20}, a_b, x_{20} > x_{10})
   \text{initially: } x_1(0) = x_{10}, v_1(0) = v_0, x_2(0) = x_{20}, v_2(0) = v_0
3 \ s(0) = 0, timer(0) = 0
5 \ d(t) = x_2(t) - x_1(t)
   \text{if } d(t) \leq D_{sense}
7 \ if v_1(t) \geq a_b
9 \ \{ \begin{array}{l}
   v_1(t + 1) = v_1(t) - a_b \\
   \text{timer}(t + 1) = \text{timer}(t) + 1
   \end{array}
   \text{else}
\{ \begin{array}{l}
   v_1(t + 1) = 0 \\
   \text{timer}(t + 1) = \text{timer}(t)
   \end{array}
\text{else}
13 \ \{ \begin{array}{l}
   s(t + 1) = 0 \\
   v_1(t + 1) = v_1(t) \\
   \text{timer}(t + 1) = \text{timer}(t)
   \end{array}
15 \ x_1(t + 1) = x_1(t) + v_1(t)
```
Proof (continued)

- **Invariant 2.** For all $x_{10}, x_{20}, v_0, D_{\text{sense}}, a_b$ and for all $t$,
  \[ x(t).\text{timer} + \frac{x(t).v_1}{a_b} \leq v_0/a_b \]

- Proof (continued).

- **Base case. $P(x(0))$**
  \[ x(0).\text{timer} + \frac{x(0).v_1}{a_b} = 0 + \frac{v_0}{a_b} \leq \frac{v_0}{a_b} \]

- **Induction. Assume $P(x(t))$, i.e., $x(t).\text{timer} + \frac{x(t).v_1}{a_b} \leq v_0/a_b$**

- **Three cases to consider**
  \[ d > D_{\text{sense}}: \]
  \[ x(t+1).\text{timer} + \frac{x(t+1).v_1}{a_b} = x(t).\text{timer} + \frac{x(t).v_1}{a_b} \leq v_0/a_b \]
  \[ d \leq D_{\text{sense}} \land v_1 \geq a_b \]
  \[ x(t+1).\text{timer} + \frac{x(t+1).v_1}{a_b} = x(t).\text{timer} + 1 + \frac{x(t+1)(v_1-a_b)}{a_b} \leq v_0/a_b \]
  \[ d \leq D_{\text{sense}} \land v_1 < a_b \]
  \[ x(t+1).\text{timer} + \frac{x(t+1).v_1}{a_b} = x(t).\text{timer} + 0 \leq v_0/a_b \]

- **SimpleCar($D_{\text{sense}}, v_0, x_{10}, x_{20}, a_b, x_{20} > x_{10}$, initially: $x_1(0) = x_{10}, v_1(0) = v_0, x_2(0) = x_{20}, v_2(0) = 0$, $s(0) = 0$, timer(0) = 0)**
  \[ d(t) = x_2(t) - x_1(t) \]

- **if** $d(t) \leq D_{\text{sense}}$
  \[ s(t+1) = 1 \]

- **else if** $v_1(t) \geq a_b$
  \[ v_1(t+1) = v_1(t) - a_b \]
  \[ \text{timer}(t+1) = \text{timer}(t) + 1 \]

- **else**
  \[ s(t+1) = 0 \]
  \[ v_1(t+1) = v_1(t) \]
  \[ \text{timer}(t+1) = \text{timer}(t) \]

- **else**
  \[ x_1(t+1) = x_1(t) + v_1(t) \]
Remarks and takeaway messages from the exercise

- **Invariant 2** takes us close to proving safety of our model (Invariant 1)
- We will need to **add assumptions** on the model to complete the proof (Homework)
- The proof by induction shows a property of **all behaviors of our model**
- The proof is conceptually simple, but can quickly get tedious and error prone
  - Verification tools like Z3, Dafny, PVS, CoQ, automate this
  - More on this later in the course
Proving safety (next time)

let us try Kahoot!
What are some of the baked-in assumptions in our model?

1 SimpleCar($D_{\text{sense}}, v_0, x_{10}, x_{20}, a_b$), $x_{20} > x_{10}$

initially: $x_1 = x_{10}, v_1 = v_0, x_2 = x_{20}, v_2 = 0$

3 $s = 0$, timer $= 0$

if $d \leq D_{\text{sense}}$

5 $s = 1$

if $v_1 \geq a_b$

7 $v_1 = v_1 - a_b$

    timer $= timer + 1$

9 else

11 $x_1 = x_1 + v_1$
Baked-in Assumptions in our scenario

- Perception.
  - Sensor detects obstacle iff distance $d \leq D_{\text{sense}}$
  - very idealized
  - Pedestrian is known to be moving with constant velocity from initial position. This will be used in the safety analysis, but not in the vehicle's automatic braking algorithm

- No sensing-computation-actuation delay.
  - The time step in which $d \leq D_{\text{sense}}$ becomes smaller is exactly when the velocity starts to decrease
Baked-in Assumptions (continued)

- Mechanical or Dynamical assumptions
  - Vehicle and pedestrian moving in 1-D lane.
  - Does not go backwards.
  - Perfect discrete kinematic model for velocity and acceleration.

- Nature of time
  - Discrete steps. Each execution of the above function models advancement of time by 1 step. If 1 step = 1 second, 
    \[ x_1(t + 1) = x_1(t) + v_1(t). 1 \]
    - We cannot talk about what happens between \([t, t+1]\)
  - Atomic steps. 1 step = complete (atomic) execution of the program.
    - We cannot directly talk about the states visited after partial execution of program
Summary

- An example of an inductive proof for safety verification of a discrete time model
  - Discrete time model: states, initial states, transition function
  - Requirements, invariants, e.g., safety
  - Counter-examples

- Detailed discussion of baked-in *assumptions* and discovered assumptions