Can STPA contribute to identify hazards of different natures and improve safety of automated vehicles?

Stephanie Alvarez, Franck Guarnieri & Yves Page
(MINES ParisTech, PSL Research University and RENAULT SAS)
Introduction:

• Recent technologies like ADAS and ITS are enabling the progressive introduction of vehicle automation into the road transport system.

• Motivation: “Vehicle automation will eliminate road crashes due to human driver error (95% of crashes)”.

• What about the changes and new hazards that automation can bring into the system, as experienced in aviation (i.e. HF issues)?
Changes Introduced by VA:
- VA brings changes into the roles and interactions of the VDE.
- VA introduces mixed traffic conditions ranging from no-automation to full-automation.

SAE levels of vehicle automation
Introduction:

- The approaches from road safety were not developed for vehicle automation.
- Such approaches are not meant to deal with VA and the complexity that it brings into a the road transport system.
- They cannot comprehensively identify the hazards that automation introduces.

(Leveson, 2016)

We also need something new!
Research Question:

• We need an approach that can identify the hazards of different natures that come with vehicle automation.

• Can STAMP and STPA identify hazards of different natures for vehicle automation?

• We did an STPA analysis on a case study of VA (Traffic Jam Pilot) and then we evaluated the outcomes of the analysis relative to the natures of identified hazards.
Methodology:

Data Collection

- Renault’s network
- Renault experts
- System description & understanding

Human Driver Model

- Extension of the human controller model
- Human driver error model
- HF concepts associated to automation

System description & understanding

STPA analysis on Traffic Jam Pilot

1. System Engineering foundation
2. STPA tables (UCAs)
3. Causal analysis on 6 UCAs (AD to MD transition)
4. Recommendations

Hazards identified via STPA

Classification of Hazards according to their Nature

- Component failure
- Design errors
- Flawed requirements
- Interactions
Traffic Jam Pilot:

• System that performs longitudinal and lateral control of the vehicle, and monitoring of the driving environment on limited portions of highways and under restricted conditions.
Traffic Jam Pilot: Forecasted end of AD mode

1 min for the driver

Takeover request

Vehicle stops minimal risk maneuver

Standstill

AD mode

10 sec

Countdown for transition

Stopping the car

Manual mode
Traffic Jam Pilot: Quick end of AD mode, type 1

Nominal autonomous mode  | Countdown  | Stopping the car  | Manual mode

AD mode

Notification to driver

Takeover request

Vehicle stops minimal risk maneuver

Standstill

Road Exit
Traffic Jam Pilot: Quick end of AD mode, type 2

AD mode

Notification to driver

HandOver Requested

Nominal autonomous mode

Stopping the car

Manual mode

Vehicule stops minimal risk maneuver

Standstill
The Human Driver Controller Model

Extension of the human controller in STPA (Thornberry 2014)

DREAM (Sagberg 2008)
CREAM (Hollnagel 1998)

Human driver failure model (Van Elslande 1997)
The Human error (Reason 1990)

Human factor issues associated to automation

- Training/operation readiness
- Experience
- Long-time learning
- Behavioral adaptation
- Traffic regulations
- Driver State
- Distraction
- Attentional Resources
- Emotional state
- Motivation
- Driving style
- Risk perception
- Operational culture
- Social context
- Driving context
- Attitudes/Acceptance
- SA
- Trust issues

Human Driver Controller

- Decision-making
- Diagnosis and Planning
- Model of automated process
- Model of controlled process
- Detection and Interpretation

Pedals and steering wheel
Commands
Displays, instrument cluster, auditory and haptic signals, etc.
Navigation system

Sensory perception
Environment
## Accident definitions:

- **[ACC-1]**: People die or get injured from road crashes.
- **[ACC-2]**: Property damage from road crashes.

## Hazard definitions:

<table>
<thead>
<tr>
<th>Hazard definition</th>
<th>Safety Constraints:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>[H-1]</strong>: Vehicle (driven by human, automation or in cooperation) violates minimum safety distance to objects, road users, vehicles, etc.</td>
<td><strong>[SC-1]</strong>: Vehicle must not violate minimum safety distance to objects, road users, vehicles, etc.</td>
</tr>
<tr>
<td><strong>[H-2]</strong>: Vehicle (driven by human, automation or in cooperation) leaves the roadway</td>
<td><strong>[SC-2]</strong>: Vehicle must not leave the roadway.</td>
</tr>
</tbody>
</table>
STPA tables overview:

<table>
<thead>
<tr>
<th></th>
<th>Control Actions</th>
<th>Unsafe Control Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human Driver</strong></td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td><strong>HMI</strong></td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td><strong>Automated Controller</strong></td>
<td>8</td>
<td>25</td>
</tr>
</tbody>
</table>

Causal Analysis on 6 Unsafe Control Actions during AD to MD mode transition.

Takeover Request
STPA outputs: Automated controller

UCA-1: The automated controller does not send takeover request when AD mode conditions are no longer met.

Scenario-1: The radar provides inaccurate measurements for object detection and consequently automation is not aware that some of the AD mode conditions are not met.

Recommendations: Sensor characterization and testing to assure accurate measurements, adequate operation and calibration; design strategy to detect inaccurate measurements.

Scenario-2: Automation is not aware of pedestrians because its process model does not consider that there may be pedestrians on highways.

Recommendations: Include a pedestrian model in the process model and test pedestrian detection; Review AD mode conditions and design assumptions.
Automated Controller

- Human driver model
- Process model
- Machine sensors: inadequate operation

Vehicle

Environment

Hazard Natures:
- Design errors
- Flawed requirements
- Component failures
- Interactions

Verified route portion

Send Takeover request

External info wrong or missing

Hazard Nature Classification: Automated Controller
STPA outputs: HMI controller

UCA-3: The HMI controller does not provide “display takeover request” when the automated controller sends the request

**Scenario-1:** The control algorithm does not send the command to display takeover request because the algorithm does not update its state.

**Recommendations:** Perform an STPA on software requirements.

**Scenario-2:** The takeover request is not displayed on the screen because there is a screen (or component) malfunction.

**Recommendations:** Hazard analysis and reliability analysis on the screen and other components of the HMI.
Hazard Nature Classification: HMI Controller

Hazard Natures:
- Flawed requirements
- Component failures
- Interactions

Human Driver Controller

HMI

Commands

Navi. System

Automated Controller

Define Route

Activate AD mode
Validate takeover req. Deactivate AD mode

Displays, auditory & haptic signals, etc.

Control algorithm

Route

Display Takeover request

Send Takeover request

Human perception
STPA outputs: Human Driver

UCA-5: The driver does not provide “validate takeover request” when the HMI displays takeover request.

Scenario-1: The driver does not perceive/hear/feel the takeover request because he is distracted watching a movie on his tablet.

Recommendations:
• Support (when possible) LoB activities via embedded screens.
• Design and test a HMI with salient, intuitive and consistent feedback.
• Provide training & accurate information to the driver before driving an AV.
• Design a minimal risk condition strategy in case the driver does not validate the request.
STPA outputs: Human Driver

UCA-6: The driver provides “validate takeover request” when he is not ready to resume manual driving.

Scenario-1: The driver validates the takeover request immediately instead of preparing for takeover (driving position, hands on, feet on, mind on) because he thinks that it is what he is supposed to do.

Recommendations:
• Driver training
• Design a HMI that “suggests” different steps to get ready before validation (i.e. position, eyes on the road, etc.)
• Include sensors that check if the driver is “ready” (driver monitoring)
• Design a validation “button” that limits unintended validation. (i.e. two buttons).
• Reassure the driver via the HMI that it is safer to let the vehicle enter minimal risk condition than to validate takeover when he is not ready to resume manual driving.
Hazard Nature Classification: Human Driver Controller

- Decision-making flaw
- Diagnosis flaw
- Process model flaw
- Detection flaw

Sensory perception: Perception flaw

- Validate takeover req.

HMI
- Commands
- Displays, auditory, haptic signals, etc.

Hazard Nature:
- Design errors
- Flawed requirements
- Component failures
- Human driver behavior
- Interactions

Human Driver Controller

Flowchart illustrating the classification of hazard nature for the Human Driver Controller, with various flaws in decision-making, diagnosis, process modeling, and detection. The sensory perception flaw is highlighted, along with validation of takeover requests and user interface commands.
Conclusions:

• Our first “attempt” with STPA was successful; we were able to identify many hazards for our study case of vehicle automation (even at a microscopic level).

• STPA was very easy to apply and to follow.

• The hardest part is understanding the theory and the paradigm shift upon which STAMP is based.
Conclusions:

• STPA allows to analyze the human factor, technical factors (incl. automation) and their interactions within the same frame.

• **STAMP and STPA enable to identify hazards of different natures associated to vehicle automation (Great candidate for vehicle automation).**

• The recommendations generated thanks to STPA target the design of the automated controller and the HMI, but there are some recommendations that target factors outside our system scope:
  - Driver training and certification, traffic rules, road verification, etc.
Perspectives

• Extending the system boundaries and the scope of the analysis to include controllers at higher levels of the socio-technical system in order to “control”:
  • Driver training and certification (professional drivers and non-professional drivers), Road traffic rules, road verification criteria, etc.
• Including the interactions with other road users at the system operation level.