Module 3: Safety Assurance for Autonomous Vehicles

Lesson 1: Safety Assurance for Self-Driving Vehicles

単語：lane splitting, fallback, crashworthiness

Uber Crash

- No real time checks on safety drivers
- After the woman was detected on the road (6 sec before)
  - first classified as unknown object
  - then misclassified as a vehicle
  - then a bicycle
  - in the end, the decision made by the autonomy software was to ignore the detections possibly because they were too unreliable
- 1.3 sec before, Volvo system tried to do emergency braking maneuver
  - Uber had disabled the Volvo system when in autonomous mode (理由：it is not safe to have multiple collision avoidance systems operating simultaneously during testing)

失敗した点：the perception system to correctly identify the pedestrian with a bicycle; the planning system to avoid the detective object even though its class was uncertain; the lack of human or emergency braking backup.

Basic Safety Term

- Safety : absence of unreasonable risk of harm
  - Any riskではない!
- Hazard : potential source of unreasonable risk of harm

Major Hazard Sources

- Mechanical
- Electrical
- Hardware
- Software
- Sensors (perception)
- Behavioral (planning)
- Fallback：縮退運転とは、システムの機能や性能を部分的に停止させた状態で稼働を維持する (driving-task fallback)
- Cyber (cybersecurity)

NHTSA: Safety Framework

- Systems engineering approach to safety
- Autonomy design
  - ODD
  - OEDR : object and event detection and response
  - Fallback
  - Traffic Laws
  - Cybersecurity
  - HMI (whether all sensors are operational, what the current motion plans are, which objects in the environment are affecting our driving behavior…)
- Testing & Crash mitigation
  - Testing: simulation, close track testing, public road driving
  - Crashworthiness : 飛行機や車両など、特にヘリコプターにおいて、衝突の衝撃から乗員の安全性を確保する性能
    - Crashes remain a reality of public road driving and autonomy systems that can minimize crash energy and exceed passenger safety standards in terms of restraints, airbags, and crashworthiness should be the norm.
  - Post crash
  - Data recording
  - Consumer Education

NTSB's Report on the 2018 Uber Crash:

https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

non-mandatory safety guidelines for autonomous cars in the NHTSA - Automated Driving Systems: A Vision for Safety 2.0 report:
https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

Lesson 2: Industry Methods for Safety Assurance and Testing

the safety perspectives of two big names in the industry: Waymo and GM

Waymo

Waymo: Safety Levels

- Behavioral Safety
- Functional Safety (backups and redundancies)
  - even if a fault or failure occurs, the car can switch to a secondary component or a backup process to minimize the severity of failures and return the vehicle to a safe state, continuing the drive if possible.
- Crash Safety
- Operational Safety (interfaces are usable, convenient, intuitive)
- Non-Collision Safety (まだわかっていないけど)
  - system designs that minimize the danger to people that may interact with the system in some way, first responders, mechanics, hardware engineers…

Waymo: Safety Processes

- Identify hazard scenarios & potential mitigations
- Use hazard assessment methods to define safety requirements
  - Preliminary analysis
  - Fault tree (top down)
  - Design Failure Modes & Effects Analysis (bottom up)
    - assess the effects of small subsystem failures on the overall capabilities of the system
- Conduct extensive testing to make sure safety requirements are met

Waymo: Levels of testing to ensure safety (the most publicly visible and well-documented program)

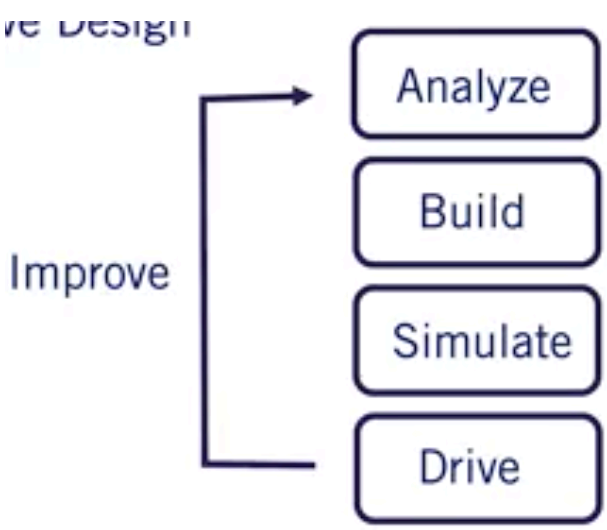
- Simulation testing: on the order of 10 million miles of simulation per day
  - mine all of their on-road experiences for challenging scenarios and perform systematic scenario fuzzing, which changes the position and velocity parameters of other vehicles and pedestrians randomly.
  - useful for finding hard edge cases.
- Closed-course testing:
  - Follow 28 core + 19 additional scenario competencies on private test tracks
  - Focus on 4 most common crashes (84% crashes): rear-end（追突）, intersection, road departure（車線逸脱）, lane change
- Real-world testing
  - increase public confidence in the technology

GM

leading position in self-driving development

- Address all 12 elements of NHTSA Safety Framework
- Iterative Design
- Control car production! (Waymo relies on OEMs to design its vehicles and only discuss mechanical and electrical hazards related to its autonomy hardware. GM manufactures their cars entirely themselves and so can enforce a more integrated design with consistent quality standards throughout the self-driving hardware.) Waymoより強いポイント
- Safety through Comprehensive Risk Management and Deep Integration

→ identify and address risks, validate solutions.
→ prioritize elimination of risks, not just mitigation.



ITERATIVE DESIGN

-All hardware, software systems meet standards

GM: Safety Processes（Waymoと同じだ）
-Deductive Analysis
-fault tree analysis
-Inductive Analysis
-Design & Process FMEA
-Exploratory Analysis
-HAZOP: Hazard & Operability Study

GM: Safety Thresholds
-Fail safes: There is redundant functionality (second controllers, backup systems etc) such that even if primary systems fail, the vehicle can stop normally
-SOTIF: All critical functionalities are evaluated for unpredictable scenarios

GM: Testing
-Performance testing at different levels
-Requirements validation of components, levels
- Fault injection testing of safety critical functionality
- Intrusive testing such as electromagnetic interference, etc
- Durability testing and simulation based testing

Is it really possible to truly precisely assess whether an autonomous car is safe? Or at least safer than a human driver?

Analytical vs Data Driven assessment

Analytical vs Data Driven: Definitions
- Analytical Safety: Ensure the system works in theory and meets safety requirements found by hazard assessment
  - example: Space Shuttle programのanalytical failure rates
  - results need to be validated through experience (data driven).
- Data driven safety: Safety guarantee due to the fact that the system has performed autonomously without fail on the roads for a very large number of kms.
  - these desired failure rates can be tied to human level driving performance.

Are autonomous cars safer?
- driving is still dangerous by human standards
- Car accidents are mostly caused due to human errors (90%) (NHTSA report, 2015)
  - a lack of judgement, a failure of human perception
- humans are also extremely good at driving, and indeed, the entire driving environment has been designed based on human perception and planning abilities.
- In US, on average
  - 1 fatal collision per 146 million km
  - 1 injury collision per 2.1 million km
  - ~ 1 collision per 400,000 km
- Disengagement: when either this autonomy software requests the driver to take over control or the safety driver feels the need to intervene.
- In 2017, Waymo had
  - Driven 563,000 km autonomously in California
  - 63 disengagements
    - unwanted vehicle maneuvers
    - perception discrepancy
    - hardware issue
    - software issue

        - behavior predictions
        - reckless road user (1 case)
    - 1 disengagement every 9,000 km➝12,500km
- In 2017, GM had
    - Driven 210,000 km
    - 105 disengagements
    - 1 disengagement every 2,000 km➝8,300km

What are some components of closed course testing emphasized by Waymo?
“Testing actual autonomous sensors and software products” should be selected.
“Testing vehicle behavior on actual roads” should not be selected. Real road testing is performed after the closed course testing.

Waymo Safety Report
https://waymo.com/safety/

GM Safety Report
https://www.gm.com/content/dam/company/docs/us/en/gmcom/gmsafetyreport.pdf

あとUber, FordのSafety Report

Lesson 3: Safety Frameworks for Self-Driving

今参加している自動運転システムの安全性評価はどうやってされている?
Generic Safety Frameworks:
- Fault Trees
- Failure modes and effects analysis (FMEA)
- Hazard and operability analysis (HAZOP)

Autonomous/Automotive Safety Frameworks
- Functional safety (FUSA), Safety of Intended Functionality (SOTIF)

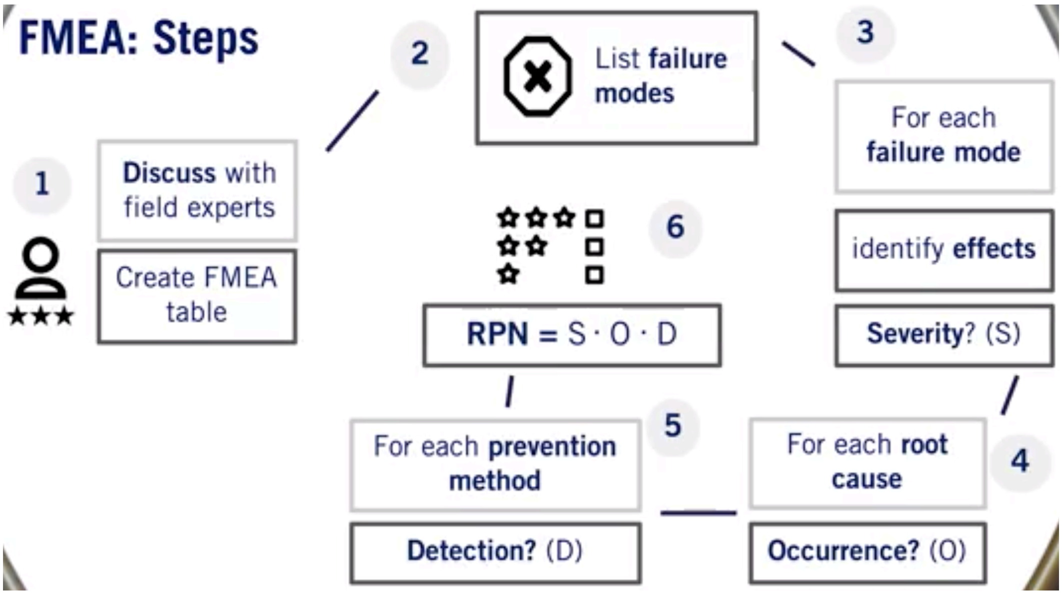
Fault Tree Analysis
- Assign probabilities to fault “leaves”
    - assign probabilities of occurrence per hour or per mile of operation
- Use logic gates to construct failure tree
    - OR : p(A) + p(B)
    - AND : p(A)*p(B)

Failure Mode and Effects Analysis (FMEA)

FMEA: Idea
- Categorize failure modes by priority
    - How serious are their effects?
    - How frequently do they happen?
    - How easily can they be detected?
- Eliminate or reduce failures, starting with top priority

FMEA: Steps
construct a table of all possible risky situations
1. Discuss with field experts, Create FMEA table
2. question the purpose of the system and list all failure possibilities
3. For each failure mode, identify effects and Severity (S, 1~10)
4. For each consequence, identify the possible root causes; for each root cause, assign Occurrence (O, 1~10)
5. identify all the ways in which the failure mode can be detected by operator, maintenance, inspection, or a fault detection system. overall mode detection likelihood before it can causes an effect, Detection (D, 1~10, 10: impossible to detect)
6. risk priority number: RPN = S * O * D

FMEA: Example



FMEA: Steps

1

Discuss with field experts

Create FMEA table

2

List failure modes

3

For each failure mode

identify effects

Severity? (S)

4

For each root cause

Occurrence? (O)

5

For each prevention method

Detection? (D)

6

RPN = S · O · D

FMEA: FAILURE MODES AND EFFECTS ANALYSIS

the vehicle has driven onto a gravel patch that appears in its test area due to road construction, which leads to controller instability
System encounters gravel, controller failure

- Severity: physical crash (S=10)
- Occurrence: whenever construction encountered, out of ODD, so somewhat likely (O=4)
- Detection: this problem is not currently detectable as the road surface texture is not actively monitored during operation of our autonomy software (D=10)
- risk priority number : 400

Other failure modes, for example:

- Sign perception failure (RPN=100)
- GPS synchronization failure (RPN=300)
- Vehicle motion prediction failure (RPN=150)

HAZOP

- HAZOP is more of a qualitative process as compared to FMEA (quantitative).
- HAZOP is often used earlier in the design process to guide the conceptual design phase.

qualitative brainstorming process, needs imagination
a simplified ongoing FMEA brainstorming approach
FMEA (quantitive)
Uses guide words (HAZOPのkey addition) to trigger brainstorming (not, more, less, early, late etc.)

Automotive Safety Frameworks
ISO 26262 - Functional Safety Standard（システムのみと関係ある）

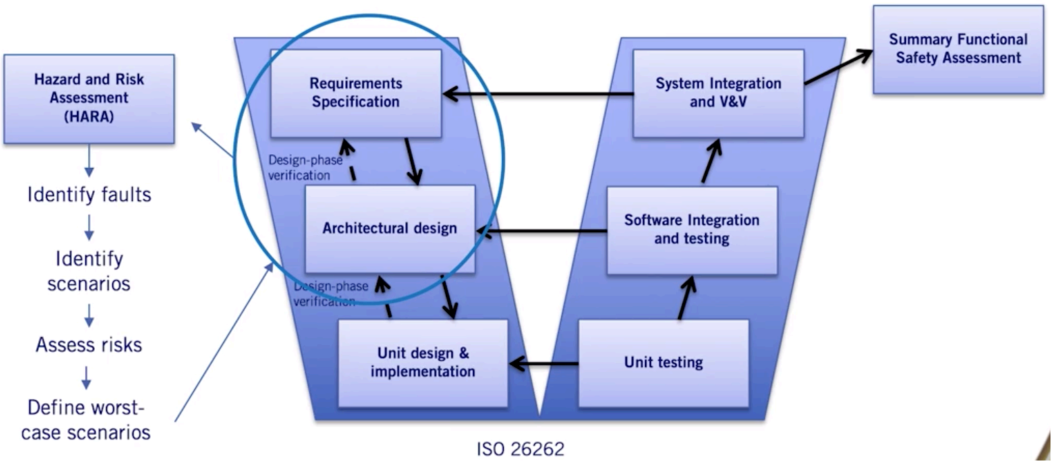
- Functional Safety
  - safety due to absence of unreasonable risk
  - only concerned about malfunctioning system

ISO/PAR 21448.1 - Safety of Intended Functionality

ISO 26262 defines 4 Automotive Safety Integrity Levels (ASIL)

- ASIL-D most stringent, ASIL-A least stringent

Functional Safety Process
V-shape


Hazard and Risk Assessment (HARA)
Identify faults
Identify scenarios
Assess risks
Define worst-case scenarios
Requirements Specification
Design-phase verification
Architectural design
Design-phase verification
Unit design & implementation
System Integration and V&V
Software Integration and testing
Unit testing
Summary Functional Safety Assessment
ISO 26262

FUNCTIONAL SAFETY PROCESS

Vの左側Requirements SpecificationとArchitectural designの段階で、HARA（Hazard and Risk Assessment）が行う。

- Identify faults (FMEA or HAZOP)
- Identify scenarios (ODD)
- Assess risks
- Define worst-case scenarios

Main idea behind functional safety: focus on worst-case requirements and then implement hardware and software that can at least handle these worst-case requirements.

Safety of the Intended Functionality (SOTIF)
Failures due to performance limitations and misuse

- Sensor limitations
- Algorithm failures / insufficiencies
- User misuse - overload, confusion, overconfidence

Designed for level 0-2 autonomy
Extension of FuSa

- V-shaped process
- Employ HARA

FMEA (failure mode and effects analysis):
https://asq.org/quality-resources/fmea

Function Safety for Road Vehicles: ISO 26262-1:2018（システム、hardware and softwareに関する安全）
https://www.iso.org/standard/68383.html

Road Vehicles - Safety of the intended functionality: ISO/PAS 21448:2019（limitations and misuseに関する安全）
https://www.iso.org/standard/70939.html

RAND Corporation Report on Driving to Safety
How many miles of driving would it take to demonstrate autonomous vehicle reliability?

https://www.rand.org/pubs/research_reports/RR1478.html

Address hazards in:
- design
- implementation
- operation

Distinction between:
- Validation testing (Are we building the right product?)妥当性確認する
  - ensure that the product actually meets the user's needs and that the specifications were correct in the first place
- Verification testing (Are we building the product right?)検証する
  - ensure that the product is being built according to the requirements and design specifications

Example, Difference from Aerospace:
In aerospace you have a pilot who is highly trained and essentially the automation does all the easy medium staff, pilot is still there for all the hard staff. In self-driving case, we want to essentially automate everything.

The Trolley Problem: ある人を助けるために他の人を犠牲にするのは許されるか?

Many of us would have known family members or people who've had an accident whilst driving a car. That's a bad thing but after that accident, your neighbor doesn't necessarily get a better driver.
But with driverless cars in 20 years time, I'll be able to get into driverless cars that have the benefit of all the miles the other driverless cars have ever driven, and we exactly don't have that with humans.
So there's so many reasons why we want to fix transports in cities and when you're in a plane and you look down out of a window, look how much of the infrastructure that we have built on our nations is because we want to move stuff.

テストメモ:
1. the most accurate and complete definition of risk in terms of self-driving vehicles?
   1. 僕の答えは：risk is a probability or threat of damage, injury, liability, loss, or any other negative occurrence that is caused by external or internal factors.
   2. 復習：Basic Safety Terms
      1. Harm: physical harm to a living thing
      2. Risk: probability that an event occurs, combined with the severity of the harm that the event can cause. (Lesson 1: Safety Assurance for Self-Driving Vehicles)
2. What kind of safety system is described by the following definition? This system can be analyzed to define quantifiable safety performance based on critical assessment of various scenarios.
   1. 復習：By analytical safety assessment, we mean that the system can be analyzed to define quantifiable safety performance or failure rates based on critical assessment of hazards and scenarios.
   2. If the overall system failure rates can be determined through analysis, it can provide strong guidance on which aspects of the system are the biggest contributors to overall safety.
3. the most accurate and complete definition of functional safety in terms of self-driving vehicles?
   1. 復習：Functional Safety or FUSA, is the absence of unreasonable risk from malfunctioning behavior caused by failures of the hardware and software in a car, or unintended behaviors arising with respect to its intended design. (Lesson 3: Safety Frameworks for Self-Driving)