Framework for Evaluating Sensor Attack Resistance

Authors: *Koichi Shimizu*, Hisashi Mori*,*
*Ryo Muramatsu*** and *Daisuke Suzuki****

1. Introduction

Autonomous systems such as self-driving cars use various sensors to understand the surrounding environment and control the system accordingly. The AEB-equipped car described in this paper uses radar, cameras, and LiDAR to detect the object in front, measure its position and velocity, and if necessary, warn the driver or automatically apply the brakes. However, there have been many reports of attacks on sensors, and there are also known examples of attacks affecting not only individual sensors but also autonomous systems that use sensors. Therefore, it is essential to ensure the safety of autonomous systems against sensor attacks.

In the case of self-driving cars, it is not realistic to conduct real-world driving tests over tens of billions of kilometers[1] required for safety verification, so evaluations are generally conducted by simulating various driving scenarios. However, the challenge is to narrow down the scenarios to be evaluated from among the countless possible scenarios. Furthermore, there have been few reports on simulators that can evaluate the impact of sensor attacks at the system level.

This paper proposes a framework for exhaustively identifying sensor attack scenarios that should be evaluated using techniques based on STAMP/STPA analysis[2] and evaluating the impact of sensor attacks on autonomous systems through simulation. It also describes the results of developing and evaluating a prototype simulator.

2. Evaluation Framework based on SOTIF

International standards related to automotive safety include the functional safety standard ISO 26262[3] and the SOTIF (Safety Of The Intended Functionality) standard ISO 21448[4]. Functional safety aims for a state in which there is no risk caused by defects such as failures. SOTIF is a relatively new safety concept that supplements functional safety, aiming for a state in which there is no risk due to intended functionality or performance limitations. ISO 21448 takes into consideration sensors that advanced functions such as autonomous driving rely on but does not include security perspectives. ISO/SAE 21434[5] is an international standard for automotive security, but it focuses on security risk management in the development process and does not address specific threats such as sensor attacks.

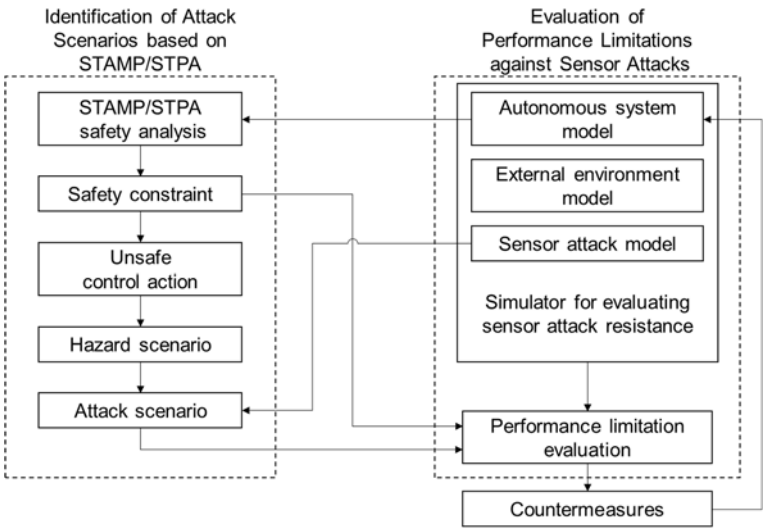

Fig. 1 Framework for evaluating sensor attack resistance

*Information Technology R&D Center **Automotive Electronics Development Center
***Corporate Information Security Div.

ISO 21448 stipulates a process for correcting specifications and improving SOTIF by inputting system specifications and then identifying and evaluating specification deficiencies and performance limitations that may lead to injury or other damage. The process is suitable for evaluating the extent to which an autonomous system that incorporates sensors can withstand sensor attacks and reflecting the evaluation results in sensor design. Therefore, this paper proposes an evaluation framework for sensor attack resistance based on the improvement process of SOTIF (Fig. 1). By identifying scenarios that lead to damage using STAMP/STPA analysis and linking the results to sensor attacks, the sensor attack scenarios that should be considered in evaluating performance limitations are exhaustively identified. When evaluating performance limitations, a simulator for evaluating sensor attack resistance that incorporates sensor attacks and the external environment into the autonomous system is used to evaluate the performance limitations against sensor attacks under each attack scenario. The details of each are described in Chapters 3 and 4.

No.	Accident	Hazard	Safety constraint
1	Unnecessary sudden deceleration injures the occupants of the ego vehicle or causes a rear-end collision from the following vehicle.	Deceleration with AEB is performed under the condition that the ego vehicle will not collide (leading to injury of ego vehicle occupants and rear-end collision from the following vehicle).	Do not decelerate the ego vehicle with AEB under the condition that the vehicle will not collide.
2	The ego vehicle collides with an obstacle, injuring its occupants and other traffic participants.	Colliding without deceleration with AEB (vehicle occupants and other traffic participants are seriously injured).	When the object is less than a certain distance, apply the brakes to decelerate/stop the ego vehicle within a defined time.

Required control action: brake control command ← Control structure (Fig. 3)

Unsafe control action: The brake control command is not transmitted to the vehicle, and the vehicle will collide without AEB activating.

Hazard scenario	Sensor	Sensor attack	Attack scenario
Due to a sensor failure, the AEB ECU misjudges that the object does not exist even though it does exist, so the AEB ECU does not transmit the brake control command and the AEB does not actuate.	Radar	Electronic jamming	Irradiating the radar with attack radio waves causes transmission of radar information that failed to detect the object. As a result, the AEB ECU does not transmit a brake control command and the AEB does not actuate.

STAMP/STPA analysis → Attack scenario identification by linking the sensor attack

ECU: Electronic Control Unit

Fig. 2 Example of attack scenario identification for AEB-equipped car

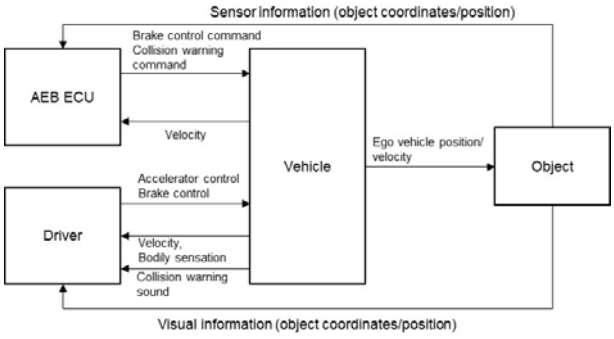

Fig. 3 Control structure of AEB-equipped car

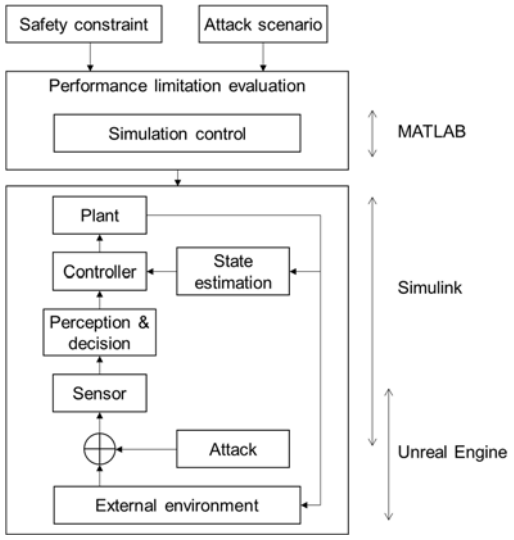

Fig. 4 Simulator for evaluating sensor attack resistance

Table 1 Supported sensor attacks

Sensor	Obstruction	Deception
Radar	Electronic jamming	Range deception Velocity deception Spoofing attack
Camera	Adversarial patch against single object detection Adversarial patch against all object detection Falsification of road markings *1	Projection of false pedestrians or vehicles *1
LiDAR	Light injection Light absorption	Spoofing attack

[1] Not supported in simulator

3. Identification of Attack Scenarios based on STAMP/STPA Analysis

Damage to be prevented, such as injury and economic loss, varies depending on the system, and is referred to as an accident in this paper. In addition, a system state in which an accident is latent is referred to as a hazard. Fault Tree Analysis (FTA), Failure Mode and Effect Analysis (FMEA), STAMP/STPA, etc. are existing techniques for identifying hazard scenarios in which systems succumb to hazards. FTA and FMEA mainly focus on hazards caused by component failures, whereas STAMP/STPA assumes that hazards can occur due to unintended interactions between components even if there are no component failures. In this paper, STAMP/STPA is adopted because the main focus is not on failures but on the interactions between AEB control and the vehicle based on sensor information.

Figure 2 shows an example of attack scenario identification for an AEB-equipped car. First, as a rough structure for realizing safety in systems, we define accidents, hazards, and safety constraints for controlling hazards. An example of the definitions is shown at the top of Fig. 2. In addition, the components (subsystems and devices) involved in the realization of safety constraints and the interactions between the components (control actions and feedback data) are analyzed and organized into a control structure as shown in Fig. 3. Next, from the control structure, the control actions required to enforce the safety constraints are identified and unsafe control actions leading to hazards are identified. Then, the hazard scenario leading to hazards is identified for each unsafe control action.

Finally, the sensor attack scenario is identified by linking the sensor failure in the hazard scenario with the sensor attack that caused it. The purposes of sensor attacks are classified into two types: obstruction and deception when detecting objects, and the sensor attacks include currently known attacks on radar, cameras, and LiDAR (Table 1).

In the analysis results of Fig. 2, “Brake control command” is identified as a control action necessary for implementing the safety constraint, “When the object is less than a certain distance, apply the brakes to decelerate/stop the ego vehicle within a defined time,” and an unsafe control action leading to hazards is identified, “The brake control command is not transmitted to the vehicle, and the vehicle will collide without AEB activating,” thereby identifying the corresponding hazard scenario and attack scenario.

4. Evaluation of Performance Limitations against Sensor Attacks

4.1 Simulator for evaluating sensor attack resistance

Figure 4 shows the structure of the simulator for evaluating sensor attack resistance. The prototype in this paper implements MathWorks’ MATLAB/Simulink *1, which is widely used in model-based design, as a platform, and models related to the external environment are implemented using Epic Games’ Unreal Engine *2. It is assumed that models developed in each domain will be used for the plants, controllers, state estimation, perception & decision, and sensors that make up the autonomous system. This time, we combined sample models provided by MathWorks to create a model of an AEB-equipped car based on radar, cameras, and LiDAR. The external environment model makes it possible to evaluate the impact of various sensor attacks on a running vehicle by covering the driving scenarios of the AEB test specified in the Japan and European NCAP (New Car Assessment Programme) and combining them with the exhaustive attack scenarios identified in the STAMP/STPA analysis. The attack model is unique to Mitsubishi Electric, and the supported sensor attacks are shown in Table 1.

[1] MATLAB and Simulink are registered trademarks of The MathWorks, Inc.

[2] Unreal Engine is a registered trademark of Epic Games, Inc.

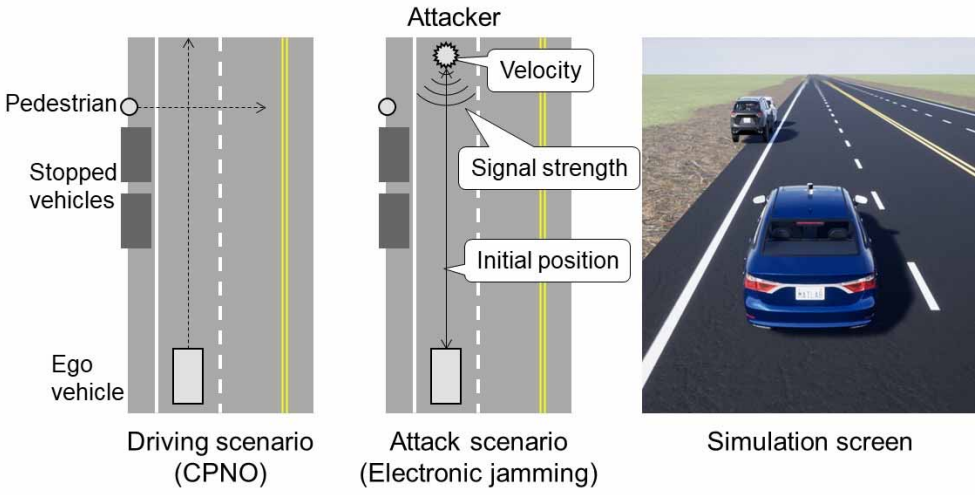

Fig. 5 Example of electronic jamming in CPNO test scenario

4.2 Example of simulation evaluation

An example of evaluating sensor attack resistance is shown for the M-out-of-N algorithm, which is a process for eliminating false detections in object detection with sensors. This algorithm confirms detection when the same object is detected at least M times out of the results of N consecutive times of detection.

Qualitatively, the accuracy of detection results increases as M is increased and brought closer to N, but the accuracy becomes more susceptible to noise. Since (M, N) is considered a design parameter determined by the performance requirements that the system must meet, the safety of the system against sensor attacks should also be considered.

Using Japan's NCAP CPNO (Car-to-Pedestrian Nearside Obstructed) as an example of a driving scenario, the attack scenario by electronic jamming of radar, which is identified in Fig. 2, is superimposed on the driving scenario (Fig. 5). The position of the attacker is fixed in front of the ego vehicle, and the initial position and signal strength are varied to evaluate the safety of the system. This evaluation is performed for different (M, N) and the results are compared. Figure 6 shows the evaluation results for (M, N) = (2, 2) and (9, 12). First, both results show reasonable results that the closer the attacker is to the vehicle and the stronger the signal strength of the attacker, the more likely the vehicle is to collide. Next, regarding sensor design parameters, it is confirmed that (M, N) = (2, 2) is generally safer but there are cases where the brakes are applied too early, that is, the brakes are applied at a timing when there is no risk of collision with a pedestrian. This indicates that the other safety constraint shown in Fig. 2, "Do not decelerate the ego vehicle with AEB under the condition that the vehicle will not collide" should be considered.

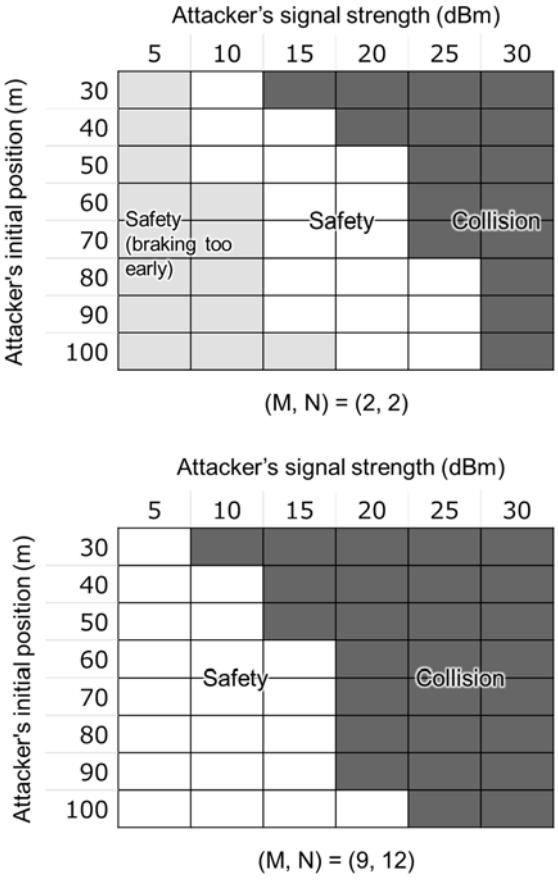

Fig. 6 Evaluation example with respect to attack parameters

5. Conclusion

In this paper, we proposed a framework for evaluating the sensor attack resistance of autonomous systems, and by taking an AEB-equipped car based on radar, cameras, and LiDAR as an example, we identified a sensor attack scenario through STAMP/STPA analysis and showed an example of evaluation using a prototype simulator. As autonomous systems become more complex, the interconnections between different simulators will be important in the future. Therefore, we will consider modularization using techniques such as Functional Mock-up Unit (FMU) to enable the developed sensor attack model to be used for other simulators.

This work is partially based on results obtained from the project (JPNP16007) commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

(1) Nidhi, K., Paddock, S.M.: Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?, RAND Corporation (2016)
(2) Nancy, L., John, T.: STPA HANDBOOK (2018)
(3) ISO, Road vehicles - Functional safety. Standard ISO 26262: 2018 (2018)
(4) ISO, Road vehicles - Safety of the intended functionality. Standard ISO/FDIS 21448: 2022 (2022)
(5) ISO, Road vehicles - Cybersecurity engineering. Standard ISO/SAE 21434:2021(E) (2021)