Safety Goals in Vehicle Security Analyses – A Method to Assess Malicious Attacks with Safety Impact

David Förster, Claudia Loderhose, Thomas Bruckschlögl, and Franziska Wiemer

Robert Bosch GmbH, Abstatt, Germany {david.foerster, claudia.loderhose, thomas.bruckschloegl, franziska.wiemer}@de.bosch.com
https://www.bosch.com

Abstract. Ensuring safety is the most important objective of security in the automotive domain. However, security analyses often lack systematic input from functional safety. We provide a method for integrating safety goals identified in the Hazard Analysis and Risk Assessment (HARA) from functional safety in a well-established Threat Analysis and Risk Assessment (TARA) for security. Our method treats safety goals as additional security goals and analyzes them in the same way as the other security goals identified by the TARA. By this means, violations of safety goals by a malicious attack are evaluated with respect to their feasibility in terms of attack potential according to Common Criteria. Furthermore, we propose a metric to quantify the security risk with safety impact based on the severity and controllability values from the Automotive Safety Integrity Level (ASIL) ratings done by safety experts in the HARA. We apply our proposal to an Automated Emergency Braking system to demonstrate how it increases the completeness and accuracy of security analyses with respect to vehicle/system safety based on expert safety ratings.

Keywords: Threat Analysis and Risk Assessment · Safety Goals · Safety Security Co-engineering.

1 Introduction

Modern vehicles are an attractive target for hackers. Due to increased connectivity features and complex in-vehicle networks, modern vehicles provide a significant attack surface for remote attacks. With more capable driver assistance systems and upcoming automated driving functions, vehicles are turning into "cyber-physical systems" and the potential impact of a successful attack becomes more and more severe. This has given rise to automotive security as a new discipline.

While drawing heavily from traditional IT security, there are new challenges specific to the automotive domain. In particular, the interaction of modern vehicles with the physical world may lead to disastrous consequences in case of

successful attacks. For several years, safety has been concerned with preventing physical harm due to malfunctions and systematic failures [5] but does not consider malicious attacks to a system.

Traditionally, security Threat Analysis and Risk Assessment (TARA) and safety Hazard Analysis and Risk Assessment (HARA) are done concurrently and independent of each other by the respective experts. Consequently, safety hazards are assessed by safety specialists without taking malicious attacks into account. In the TARA, it is difficult for security experts to accurately assess the safety impact of an attack. Due to their limited safety expertise, they will only be able to provide a very coarse grained assessment of physical harm as a consequence, often using inaccurate worst-case estimations.

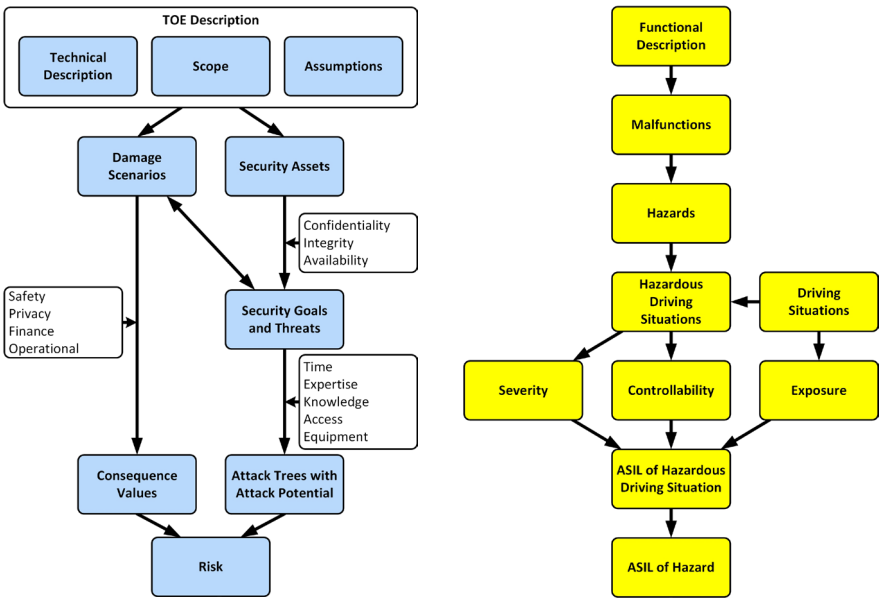
As the basis for an adequate and economical level of security, it is crucial that the safety impact of security threats is accurately reflected in the TARA. Close co-operation and well-defined interfaces between safety and security are mandated by the ISO 26262 [5] and upcoming ISO/SAE 21434 [7] standards but have yet to be broadly established in practice.

In this paper we propose a method to use safety goals identified in the HARA and their Automotive Safety Integrity Level (ASIL) ratings as input for the TARA. By leveraging safety expert judgments we avoid inaccurate estimations by security experts with limited safety expertise. The proposed approach increases the completeness of security analyses by making sure all safety goals are considered, and we are able to define the safety-related security risk more accurately. By applying our method to an Automated Emergency Braking (AEB) system as a real-world example, we demonstrate how security risks with safety impact can be assessed more accurately.

The remainder of this paper is structured as follows. In Section 2, we review current industry best practice approaches to safety and security. Next, we describe our proposal how safety goals can be integrated into a TARA in Section 3 and introduce a metric for quantifying the safety impact of security threats in Section 4. In Section 5, we show the feasibility and benefit of our approach by applying it to a real-world example. Previous work related to this paper is presented in Section 6. We conclude with a summary and discussion of future work in Section 7.

2 Preliminaries

In this section we describe a Threat Analysis and Risk Assessment (TARA) method based on Common Criteria [6] and EVITA [4], which is commonly used in the automotive industry in the presented or a similar form. It will be the basis for the changes we propose in the next sections. Note that our proposal can also be integrated into other asset-based TARA methods. For readers who may be less familiar with safety processes, we review how safety goals and their ASIL ratings are identified during the Hazard Analysis and Risk Assessment (HARA) according to ISO 26262 [5]. Figure 1 gives an overview over the two methods.



(a) Overview of the TARA workflow (b) Overview of the HARA workflow

Fig. 1: Overview of the TARA and HARA workflows

2.1 Threat Analysis and Risk Assessment

The purpose of a TARA is the identification and prioritization of security requirements during development. It starts with a description of the Target of Evaluation (TOE). The technical description provides information on the TOE's functionality, its interfaces, related signals and communication links, its HW/SW architecture, and its environment. In the scope definition, we state precisely which parts of the TOE and its environment are the subject of the analysis. The assumptions describe what is happening outside our scope. Based on the description, we assess the threats to the TOE in terms of their feasibility and their impact (consequence). See Figure 1a for a high level overview of the process.

We begin with the identification of assets that need to be protected from a security point of view (e.g. sensor signals, Electronic Control Unit (ECU) firmware, logging data, etc). In order to specify the need for protection, we assign relevant security properties (confidentiality, integrity, availability) to each asset. The combination of a security asset and a security property is called security goal (e.g. integrity of firmware). The violation of a security goal is called a threat (e.g. firmware manipulation). Using this terminology, a threat is the collection of all attack paths that lead to the violation of the related security goal.

The feasibility of each attack is quantified by its attack potential according to Common Criteria [6]: The harder an attack is, the higher the attack potential (capabilities of the attacker) required to execute it. To calculate the attack

potential, we construct an attack tree for each threat [11]. The root node of this directed tree represents the successful realization of the threat and the other nodes and their connections model the different attack paths that lead to a violation of the underlying security goal. We assign values for the required attacker capabilities in terms of the required time, expertise, knowledge, access and equipment to each leaf node. Then a specific algorithm propagates these 5-tuples of values through the tree to the root node in a way that the root node receives 5-tuples with minimum sum over the five values and therefore represents the easiest attack path. This minimum sum is the attack potential, denoted $AP(T)$ for a given threat T, which we cluster into different categories $\{Basic, Enhanced\ Basic, Moderate, High, Beyond\ High\}$.

To assess the consequences of the successful execution of an attack, we create a list of damage scenarios (e.g., a crash).[1] For each damage scenario D we assign a consequence value, denoted $cons(D)$, that rates its severity in different categories, e.g., safety, privacy, financial, and operational [4]. We cluster them into the categories $\{Negligable, Moderate, Serious, Severe\}$. Note that these ratings are often done by security experts, who may struggle to provide an accurate rating regarding safety. We denote the set of all threats that can lead to a damage scenario D as $\mathcal{T}(D)$.

Finally, we calculate the security risk of a damage scenario D from its consequence value $cons(D)$ and the threat $T \in \mathcal{T}(D)$ with the lowest attack potential $AP(T)$:

$$risk_{sec}(D) = cons(D) \cdot \max_{T \in \mathcal{T}(D)} \frac{1}{AP(T)} \tag{1}$$

Usually, the risk values $risk_{sec}(D)$ are clustered into categories $\{Low, Medium, High, Very\ high\}$ and Equation (1) then can be represented by a lookup table. If $AP(T) = 0$, the highest risk category is applied.

As the next step after completing the TARA, the risk ratings are used to prioritize mitigation of the identified attacks until an acceptable level of residual risk is achieved.

2.2 Hazard Analysis and Risk Assessment

HARA is a method to systematically identify and rate hazards on vehicle-level caused by malfunctions. The rating is based on three factors: the severity of the hazard in terms of physical harm, the exposure of the vehicle to a relevant driving situation and the controllability of the hazard during this driving situation by the driver or other traffic participants. For each combination of a hazard and a related driving situation these three values yield the ASIL rating. The ASIL of the hazard is the maximum ASIL rating over all relevant driving situations. The higher the ASIL rating, the more assurance must be provided during development that the hazard will not occur.

[1] Identification of threats and damage scenario is an iterative process: Damage scenarios can be derived from threats, and threats can be derived from damage scenarios.

Formally, let H be a hazard related to a malfunction of our object of analysis. We identify by $\mathcal{S}(H)$ the set of all driving situations that are considered to be relevant when H occurs. In the following, we call all combinations (H, S) with $S \in \mathcal{S}(H)$ hazardous driving situations. Let $e(S)$ be the value that represents the exposure of the vehicle to a driving situation S. Further let $c(H, S)$ denote the controllability and $s(H, S)$ the severity of the hazardous driving situation (H, S). Table 1 provides the corresponding criteria and the values for $e(S)$ [2], $c(S)$ and $s(S)$ as defined in ISO 26262 [5].

For a hazardous driving situation (H, S) the ASIL is defined as follows:

$$\mathrm{ASIL}(H,S) = \begin{cases} \mathrm{QM} & \text{, if } s(H,S)+c(H,S)+e(S) \leq 6 \\ \mathrm{A} & \text{, if } s(H,S)+c(H,S)+e(S) = 7 \\ \mathrm{B} & \text{, if } s(H,S)+c(H,S)+e(S) = 8 \\ \mathrm{C} & \text{, if } s(H,S)+c(H,S)+e(S) = 9 \\ \mathrm{D} & \text{, if } s(H,S)+c(H,S)+e(S) = 10 \end{cases}. \tag{2}$$

Now we define the ASIL of a hazard H as

$$\mathrm{ASIL}(H) = \max_{S \in \mathcal{S}(H)} \mathrm{ASIL}(H,S). \tag{3}$$

For each hazard we define a safety goal as the avoidance of the hazard. The ASIL of the hazard is assigned as an attribute to the corresponding safety goal.

Consider for example a brake ECU with Electronic Stability Program (ESP) functionality. It has the hazard H = "unintended asymmetric braking". The corresponding safety goal is "avoid unintended asymmetric braking". Here most of all driving situations are relevant. For example S = "Driving on a country road or a highway" is relevant and has $e(S) = 4$. In this driving situation H is

[2] The value $e(S) = 0$ is defined as "incredible" but no ASIL is derived. Therefore, it is not further relevant for this paper.

Table 1: Exposure, Controllability and Severity Rating

Exposure e	**1**	**2**	**3**	**4**
	Very low probability	Low probability	Medium probability	High probability
Controllability c	**0**	**1**	**2**	**3**
	Controllable in general	Simply controllable	Normally controllable	Difficult to control or uncontrollable
Severity s	**0**	**1**	**2**	**3**
	No injuries	Light and moderate injuries	Severe and life-threatening injuries (survival possible)	Life-threatening injuries (survival uncertain), fatal injuries

difficult to control ($c(S) = 3$). Furthermore, (H, S) = "unintended asymmetric braking while driving on a country road or highway" is very likely to result in a crash with fatal injuries ($s(H, S) = 3$). Hence $ASIL(H, S) = \mathrm{D}$ is the ASIL of the hazardous driving situation and also the ASIL of H.

The safety risk of a hazard is a combination of its ASIL and the probability of an error occurring that leads to a malfunction that results in the hazard. Therefore, the safety risk can be modeled as

$$risk_{\text{safe}}(H) = \max_{S \in \mathcal{S}(H)} s(H, S) \cdot c(S, H) \cdot e(S) \cdot p(H, S) \tag{4}$$

for the occurrence probability $p(H, S)$ of a malfunction that causes H during S.

As the occurrence probabilities are generally hard to evaluate, the ASIL itself remains a key component in safety considerations and is used to determine the required process rigor and technical measures.

3 Safety goals as input to the TARA

As described in Section 2, the methodical approaches of a HARA from the safety domain and a TARA from the security domain are very similar to each other. Both first identify hazards (safety) and threats (security) and afterwards quantify the associated risk by assessing the likelihood of occurrence[3] and the severity of the potential consequence. In this section we present our approach for identification of safety related threats and in the next section we describe how to quantify the identified threats according to their safety impact.

A crucial step in the threat analysis for security is the identification of assets that need protection. One approach is to break the TOE down into its technical parts. In case of a single ECU, the assets could be its interfaces, volatile and non-volatile memory, and various embedded software parts. In case of a driver assistance system, the assets could be the different ECUs that are part of the system and the data stored and exchanged by them. However, in this technical approach, it is very hard to accurately assess the safety impact if one of the assets is compromised. How to determine the safety impact if an attacker can manipulate a certain network connection or the software on a certain ECU?

We propose a different approach: A system's safe operation is a fundamental asset that needs protection. More specifically, the safety goals identified in the HARA must be protected, not only against hazards coming from systematic faults and random hardware faults but also against malicious attacks.

Therefore, we define each safety goal as an asset in the threat analysis and add "fulfillment"[4] as a corresponding new security property to the existing CIA triad (confidentiality, integrity, availability). The rest of the analysis remains the same: Every possible violation of security goals (including the ones derived

[3] The attack potential is negatively correlated to the likelihood of an attack to occur.

[4] To minimize the changes to existing processes and tooling, it is also possible to speak of *integrity* as the safety goal's security property that needs protection instead of introducing *fulfillment* as a new property.

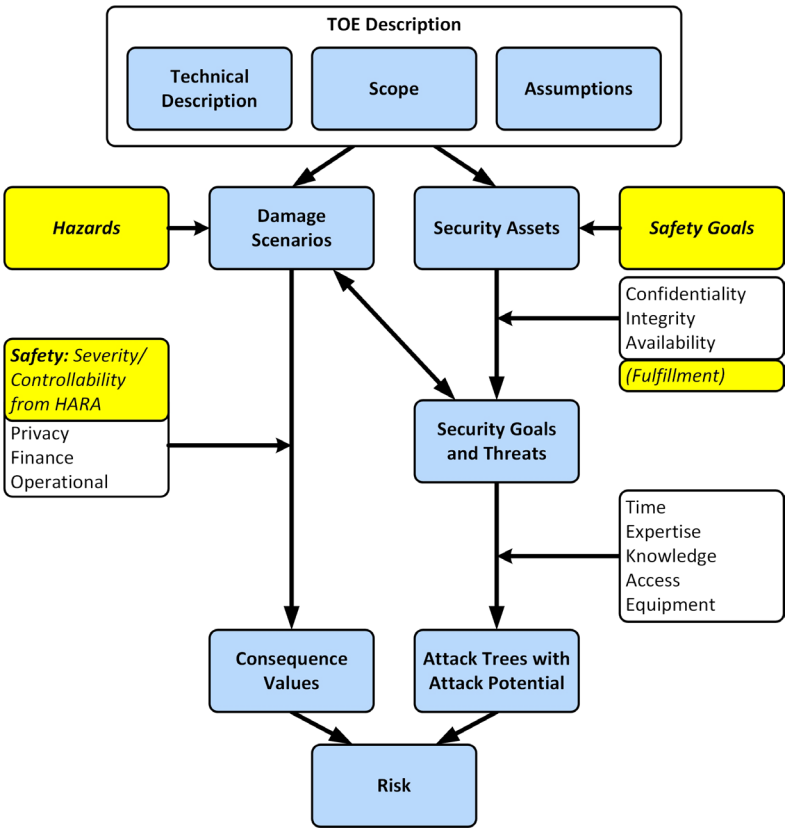


Fig. 2: The TARA methodology with integrated safety elements

from safety goals) results in a threat and is assessed by creating a corresponding attack tree. For a safety-related threat T, the corresponding damage scenario is the hazard H associated with the safety goal the threat was derived from. We discuss in more detail how to use the safety goal's ASIL rating for consequence rating in the next section. Figure 2 gives a high-level overview of the additional safety input to the security analysis and Table 2 shows how elements from the HARA are re-interpreted as elements in the TARA.

As safety goals are based on functions and malfunctions, the analysis of the corresponding threats must also be done on the functional level. Therefore, we propose to derive attack trees from the functional architecture, which lists all involved hardware components, e.g., sensors, computational units, actuators, all the required signals, and all data transferred within the in-vehicle network. As additional input relevant information about non-functional communication within the system, e.g., for arbitration or synchronization, shall also be considered.

Table 2: Elements from safety HARA as input to security TARA

Safety Element	**TARA Element**
Safety Goal	→ Security Asset
Fulfillment of Safety Goal	→ Security Goal
Violation of the Safety Goal	→ Security Threat
Hazard	→ Damage Scenario
Severity and controllability from Safety Evaluation (ASIL)	→ Consequence rating

Based on this input, an attack tree will have a branch for each top-level threat that can violate the corresponding safety goal. Let us consider the example for a functional architecture for a radar-based Automated Emergency Braking system displayed in Figure 3 and the safety goal "Avoid unintended or too high vehicle deceleration". The safety goal can be violated by a number of ways: Manipulate the radar ECU to report false obstacles, manipulate the brake ECU to suddenly apply the brakes, or spoof the network signals from radar to brakes. Based on these, three subtrees are created for the threats as shown in Figure 5a. It is easy to see that existing attack trees that were created in a traditional security analysis, e.g., for accessing internal memory, modifying a component's firmware or accessing debug interfaces, can be re-used and plugged into the safety goal based analysis as subtrees.

For the analysis to be most useful, the item definition from the safety analysis and the TOE from the security analysis must be on the same level of abstraction (e.g., a single ECU or an automated driving system consisting of several ECUs).

The formal description of the mapping from safety to security elements is as follows: Let $\mathcal{D}_{\text{Security}}$ be the set of all damage scenarios of the TOE regarding traditional security analysis and let $\mathcal{D}_{\text{Safety}}$ be the set newly added damage scenarios that represent violation of a safety goal. Then $\mathcal{D} = \mathcal{D}_{\text{Security}} \cup \mathcal{D}_{\text{Safety}}$ is the set of all relevant damage scenarios for the security analysis. $\mathcal{D}_{\text{Security}}$ and $\mathcal{D}_{\text{Safety}}$ will not be disjoint if safety-related threats have been considered in the security analysis before adding the safety goals. In this case, safety related damage scenarios within $\mathcal{D}_{\text{Security}}$ can be removed as they are usually based on worst case estimations and therefore lack accuracy. In contrast to that, all safety relevant damage scenarios are systematically covered by $\mathcal{D}_{\text{Safety}}$.

When the TOE is a complex system, there can be a large number of hazards and derived damage scenarios. To reduce the number of safety-related security risks, the hazards can be grouped such that a damage scenario includes multiple hazards. In this case we define $D = \{H_1, \ldots, H_\ell\}$.

4 Quantifying the safety impact of a security threat

In Section 3, we described how for each safety goal a corresponding threat (of violating this safety goal) is created. In this section, we propose a method to

determine the consequence rating based on the ASIL rating of the affected safety goal. Our objective is to avoid having security experts perform this task, which they may have limited expertise for and which has already been performed by safety experts during their HARA.

As described in Section 2.2, a safety goal's ASIL rating consists of the *severity*, *controllability*, and *exposure* ratings. We discuss each of the ratings and their applicability for quantifying a malicious attack.

Severity There can be little argument that the physical harm, expressed by severity, is independent of whether it was caused by a malfunction or by a malicious attack.

Controllability Similarly, the driver's ability to control unintended system behavior is independent of its cause. Obviously, for automated driving functions that do not require driver supervision no controllability must be assumed, both from a safety and a security point of view.

Some argue that exposure should not be considered because an attacker may take additional actions to divert the driver, such as switching on wipers or turning up the radio. However, in our opinion, simultaneous attacks to several parts of the system are a more complex issue that need to be investigated independently from the employed metric.

Exposure The underlying assumption from safety, that a risk rating can be reduced if a hazard only occurs in certain driving situations, does not apply to security [3]. A malicious attacker can pick a time for his attack when the target vehicle is exposed (i.e, the attack will result in a safety goal violation). Therefore, this rating cannot be used to quantify the safety impact of a malicious attack.

Therefore, quantifying the consequence of a damage scenario $D \in \mathcal{D}_{\text{Safety}}$ should be based on the severity and the controllability values of the related safety goals and should not consider their exposure values. We propose a mapping from *severity* and *controllability* to a four step consequence scale in Table 3. Based on the proposed quantification we define for each safety related damage scenario $D \in \mathcal{D}_{\text{Safety}}$ the consequence value $cons(D)$ based on all relevant hazardous driving situations (H, S).

Table 3: Mapping severity and controllability values from a safety goal's ASIL rating to the consequence of a malicious attack violating the safety goal

	c = 0	**c = 1**	**c = 2**	**c = 3**
s = 0	Negligible	Negligible	Negligible	Negligible
s = 1	Negligible	Moderate	Moderate	Serious
s = 2	Negligible	Moderate	Serious	Serious
s = 3	Negligible	Serious	Serious	Severe

For each hazard H and each driving situation $S \in \mathcal{S}(H)$, we define the consequence value as follows (cf. Table 3).

$$cons(H,S) = \begin{cases} Negligible \text{ , if } s(H,S) \cdot c(H,S) = 0 \\ Moderate \text{ , if } 1 \leq s(H,S) \cdot c(H,S) \leq 2 \\ Serious \text{ , if } 3 \leq s(H,S) \cdot c(H,S) \leq 6 \\ Severe \text{ , if } 7 \leq s(H,S) \cdot c(H,S) = 9 \end{cases} \tag{5}$$

For a damage scenario $D = H \in \mathcal{D}_{\text{Safety}}$, this yields the consequence value

$$cons(D) = \max_{S \in \mathcal{S}(H)} cons(H,S). \tag{6}$$

According to Equation (1), we obtain the following security risk of D. Again, if $AP(T) = 0$, the highest risk category is applied.

$$risk_{sec}(D) = risk_{sec}(H) = \max_{S \in \mathcal{S}(H)} cons(H,S) \cdot \frac{1}{AP(T)} \tag{7}$$

where T is the threat related to H.

In case several hazards are grouped into one damage scenario $D = \{H_1, \ldots, H_\ell\} \in \mathcal{D}_{\text{Safety}}$ as mentioned at the end of Section 3, the security risk is calculated as follows

$$risk_{sec}(D) = \max_{i=1,\ldots,\ell} \left(\max_{S \in \mathcal{S}(H_i)} cons(H_i,S) \right) \cdot \frac{1}{AP(T_i)} \tag{8}$$

where T_i is the threat related to H_i.

5 Evaluation

We evaluate our proposal by comparing it to the traditional TARA approach described in Section 2.1 using AEB as a simple but practical, real-world example.

5.1 AEB functional example

AEB supports the driver in critical traffic situations in which immediate strong braking is required to prevent an accident. The simplified model of an AEB system consists of a sensor to detect obstacles in the front of the vehicle, usually a radar, and a brake system to perform the necessary deceleration. Figure 3 shows the functional flow of the AEB within the two systems being connected via an in-vehicle network. If the sensor detects an obstacle in front and the driver does not react to this obstacle, the sensor sends a signal to the brake actuator to perform an emergency braking. Thus the feature can reduce the vehicle's velocity in case of an impact or bring the vehicle to full stop avoiding the accident.

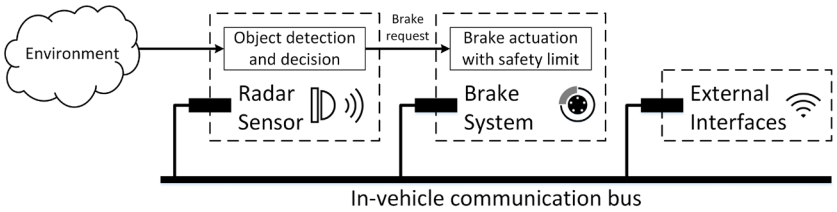


Fig. 3: A simplified system architecture and functional flow of an AEB function

5.2 Assumptions and attack model

For the evaluation we assume that a telematics unit is also attached to the communication bus. An attacker may be able to compromise this telematics unit through a remote connection and by this gain access to the in-vehicle network. While AEB interventions can prevent harm as described above, a false positive activation that triggers strong braking with not apparent reason can potentially cause an accident. In this example, we assume that the target of the attacker is to manipulate the AEB-related components and interfaces to cause harm to the driver.

5.3 Traditional security analysis approach

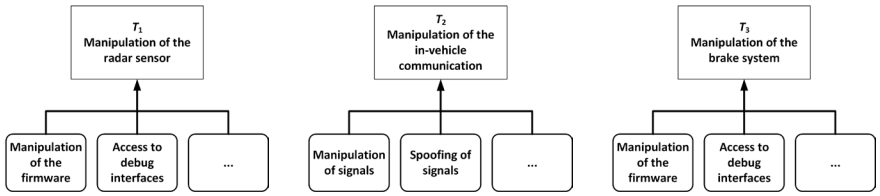
While performing the TARA for an AEB function, one step is the identification of damage scenarios and their consequences. In this example, the unintended behavior of the vehicle is an executed AEB with heavy breaking, which is potentially leading to a rear-end collision with a following vehicle. Therefore, the identified damage scenario is a *crash*. Without further insights into the safety concept, the worst-case estimation for the consequence rating is *severe* (the highest value). Based on the system architecture, the involved components and communication infrastructure, we identify three threats that potentially lead to the identified damage scenario:

T_1 Manipulate the radar ECU to report false obstacles
T_2 Spoof/manipulate the network signal from radar to brakes communication.
T_3 Manipulate the brake ECU by hacking into it and changing its software to suddenly apply the brakes

For the sake of simplicity, we assume that the attack potential for all three threats is rated *High*. Figure 4 shows the three corresponding simplified trees for manipulating the components or the in-vehicle communication.

Calculating the risk rating, we find that all components and the in-vehicle network must be secured to reduce the risk to an acceptable level.

Risk rating for damage scenario "Crash": *High*, based on consequence rating *Severe* and minimum attack potential *High*.



(a) T_1 Manipulation of the radar sensor

(b) T_2 Manipulation of in-vehicle communication

(c) T_3 Manipulation of the brake system

Fig. 4: Simplified attack trees for manipulation of components and in-vehicle communication

5.4 Safety goal based security analysis approach

Let us now consider the safety goal based approach. Based on the HARA, the safety experts derived two distinct safety goals that depend on the safety measure of putting a limiter onto the brake ECU that will limit the maximum deceleration that can be applied during an AEB:

SG_1 Avoid unintended or too high vehicle deceleration *within* function limits
SG_2 Avoid unintended or too high vehicle deceleration *beyond* function limits

with related hazards H_1, H_2.

As safety experts have already evaluated the potential hazards and their consequences, the damage scenario *crash* is no longer needed and can be replaced by the two damage scenarios directly related to the safety goals and hazards from the HARA. The consequence rating for those scenarios can be derived from the ASIL ratings that are taken from the HARA as discussed in Section 4:

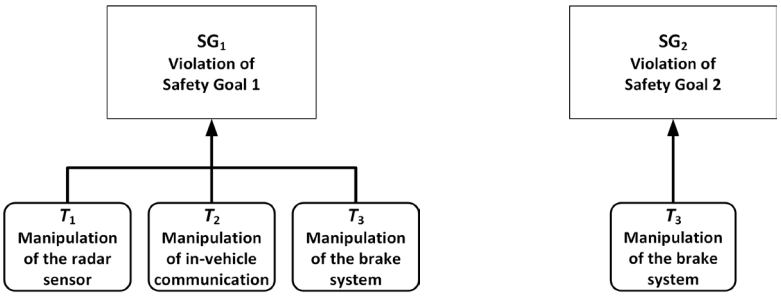
H_1 severity $s = 2$ and controllability $c = 3$ results in $cons(H_1) = serious$
H_2 severity $s = 3$ and controllability $c = 3$ results in $cons(H_2) = severe$

Assessing the threats to each of the safety goals, we find that all of the threats T_1, T_2, T_3 identified in the previous section can violate SG_1. However, only T_3 can violate SG_2. We can re-use the trees as subtrees as shown in Figure 5.

Calculating the resulting risks, we find that one of the risks (involving only the brakes) is still rated *High* while the other one (involving all ECUs) is only rated *Medium*. Consequently, the brake ECU still requires the same protection as in the previous analysis, but the radar ECU and the communication link require less protection to reduce the risk to an acceptable level.

Risk rating for Damage Scenario 1 (Avoid unintended or too high vehicle deceleration *within* function limits): $Medium$, based on consequence rating $Serious$ and minimum attack potential $High$.

Risk rating for Damage Scenario 2 (Avoid unintended or too high vehicle deceleration *beyond* function limits): $High$, based on consequence rating $Severe$ and minimum attack potential $High$.



(a) Violation of SG_1 leading to H_1 (b) Violation of SG_2 leading to H_2

Fig. 5: Simplified attack trees for the violation of safety goals based on the functional flow and involved components and communication

5.5 Comparison of the two approaches

As we can see in this example, the proposed approach to integrate safety goals into the TARA provides valuable input for a more accurate risk assessment with respect to the safety impact of malicious attacks. In the traditional TARA scenario, all involved components require the same level of protection based on a rather abstract damage scenario "crash". Taking the safety goals, we see that the radar ECU and the communication link potentially require less protection than the brake ECU based on the lower rating for the risks that apply to them.

6 Related work

Safety and security co-engineering goes back as far as 1999, where Eames and Moffett discuss the integration of safety and security requirements in the context of an air traffic control system [2]. They identify the risk of conflicting requirements if security and safety are treated separately and propose to harmonize (but not unify) the two disciplines in order to avoid conflicting requirements.

Burton et al., in contrast, propose to combine safety and security into an integrated "safety + security" analysis methodology [1]. We intentional share their approach of considering malicious violation of safety goals as security threats. However, we recognize that a complete unification of security and safety is a big step and may not be practical due to different skills and mindsets in the respective disciplines. Our approach is more lightweight and emphasizes information exchange between safety and security at well-defined interfaces while leaving existing processes and organizations unchanged.

Glas et al. discuss the challenges associated with combining safety and security [3]. They discuss conflicting objectives as well as potential synergies. In particular, they raise the question whether ASIL ratings can be used for assessing security threats, which we propose a solution for.

There are various proposals to augment methods from safety with security aspects: Schmittner et al. extend Failure Mode and Effects Analysis (FMEA)

to Failure Mode, Vulnerabilities and Effects Analysis (FMVEA) [10], Steiner and Liggesmeyer consider security aspects in Component Fault Tree Analysis (CFT) [12], and Macher et al. augment HARA with security aspect to Security-Aware Hazard Analysis and Risk Assessment (SAHARA) [8]. All of these proposals aim to provide additional input from security to safety whereas our approach addresses the completeness of a security analysis.

Considering safety aspects in security analyses has been proposed before. The EVITA method [4], which is widely used in the automotive domain, considers severity and controllability to quantify security threats. In his Master thesis, Winsen proposes to re-use the controllability and severity ratings from safety quantification of security threats in his proposed "composite threat model" [13]. While these approaches are similar to ours in some aspects (we build upon the EVITA methodology), they do not explicitly use safety goals to ensure completeness of a security threat analysis, which is one of our major contributions.

Not specifically considering the automotive domain, Piètre-Cambacédès and Bouissou provide an extensive survey of risk identification techniques for safety and security, similarities and differences in respective definition of risk, and applicability of safety methods to security and vice-versa [9].

7 Conclusion

The requirement for safety/security co-engineering is widely recognized throughout the automotive industry, covered by a large body of literature, and mandated by both safety and upcoming security standards. Yet, the disciplines are handled by separate communities with their own mindsets, executing their own processes with limited interaction and sometimes conflicting objectives. While some argue that ultimately security and safety will merge, there is still a long way to go.

In this paper, we propose concrete improvements to increase the completeness and accuracy of a security TARA based on readily available safety artifacts. Our proposal can be implemented within existing organizations, processes and tools. By information exchange between safety and security through the well-defined interface of the HARA, we provide a low entry barrier for safety and security collaboration. The proposed extensions to the TARA methodology can be covered within the established EVITA/Common Criteria-based frameworks: Including safety goals in the list of security assets only adds an additional input to several existing ones. Using severity and controllability values from a safety goal's ASIL rating to quantify the consequence of a malicious violation of this safety goal can be done by mapping them to an existing consequence scale.

While we certainly envision and welcome more elaborate collaboration between safety and security, our lightweight proposal provides a simple and straightforward entry for safety and security co-engineering. It delivers immediate benefits in the form of increased completeness and accuracy of security analyses by leveraging safety expert judgments and existing work products.

In the future, we plan to explore in more detail how to consider advanced concepts from safety, such as ASIL decomposition and redundancy, in security

analyses. While redundancy can provide protection against sporadic or systematic failures, it does not per se provide protection against malicious attacks but, on the contrary, may increase the attack surface. Furthermore, we plan to explore how our proposal can be applied on different abstraction level, e.g. vehicle level, domain level, ECU level. Taking as input requirements from the respective level of the safety requirements trace, we expect that this will enable modular analyses and help master complex systems.

This work is co-founded by the German Federal Ministry of Education and Research in the project SecForCARs with funding number 16KIS0792.

References

1. Burton, S., Likkei, J., Vembar, P., Wolf, M.: Automotive functional safety = Safety + Security. In: Proceedings of the First International Conference on Security of Internet of Things. pp. 150–159. ACM (2012)
2. Eames, D.P., Moffett, J.: The integration of safety and security requirements. In: International Conference on Computer Safety, Reliability, and Security. pp. 468–480. Springer (1999)
3. Glas, B., Gebauer, C., Hänger, J., Heyl, A., Klarmann, J., Kriso, S., Vembar, P., Wörz, P.: Automotive safety and security integration challenges. Automotive-Safety & Security (2015)
4. Henniger, O., Apvrille, L., Fuchs, A., Roudier, Y., Ruddle, A., Weyl, B.: Security requirements for automotive on-board networks. In: 9th International Conference on Intelligent Transport Systems Telecommunications. pp. 641–646. IEEE (2009)
5. International Organization for Standardization: ISO 26262:2018 "Road vehicles – Functional safety", International Standard.
6. International Organization for Standardization: ISO/IEC 15408:2009 "Information technology – Security techniques – Evaluation criteria for IT security", International Standard.
7. International Organization for Standardization: ISO/SAE 21434 "Road vehicles – Cybersecurity engineering", International Standard, working draft.
8. Macher, G., Sporer, H., Berlach, R., Armengaud, E., Kreiner, C.: Sahara: A security-aware hazard and risk analysis method. In: Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition. pp. 621–624. EDA Consortium (2015)
9. Piètre-Cambacédès, L., Bouissou, M.: Cross-fertilization between safety and security engineering. Reliability Engineering & System Safety **110**, 110–126 (2013)
10. Schmittner, C., Gruber, T., Puschner, P., Schoitsch, E.: Security application of failure mode and effect analysis (FMEA). In: International Conference on Computer Safety, Reliability, and Security. pp. 310–325. Springer (2014)
11. Schneier, B.: Attack trees. Dr. Dobb's journal **24**(12), 21–29 (1999)
12. Steiner, M., Liggesmeyer, P.: Combination of safety and security analysis – Finding security problems that threaten the safety of a system. In: Proceedings of Workshop DECS (Dependable Embedded and Cyber-Physical Systems) of the 32nd International Conference on Computer Safety, Reliability and Security (2013)
13. Winsen, S.: Threat Modelling for Future Vehicles: On Identifying and Analysing Threats for Future Autonomous and Connected Vehicles. Master's thesis, University of Twente (2017)