Agent-centred Approach for Assuring Ethics in Dependable Service Systems

Irfan Šljivo, Elena Lisova, Sara Afshar
Mälardalen Real-Time Research Centre, Mälardalen University,
Västerås, Sweden
{irfan.sljivo, elena.lisova, sara.afshar}@mdh.se

Author Copy.
Published in 2017 IEEE 13th World Congress on Services

***Abstract*—As the world enters the information era, more and more dependable services controlling and even making our decisions are moved to the ubiquitous smart devices. While various standards are in place to impose the societal ethical norms on decision-making of those devices, the rights of the individuals to satisfy their own moral norms are not addressed with the same scrutiny. Hence, the right of the individuals to reason on their own and evaluate morality of certain decisions is at stake.**

In this work we propose an agent-centred approach for assuring ethics in dependable technological service systems. We build upon assurance of safety and security and propose the notion of ethics assurance case as a way to assure that individual users have been made aware of all the ethically challenging decisions that might be performed or enabled by the service provider. We propose a framework for identifying and categorising ethically challenging decisions, and documenting the ethics assurance case. We apply the framework on an illustrative example.

Keywords*—*Ethics, assurance case, decision making, critical systems, IoT.

I. Introduction

In the industrial era we have mainly relied on ourselves and services provided by other human controllers. The complexity of the service provider chains was manageable. Nowadays in the information era with the rise of interconnected smart computing devices, such as IoT and Cloud computing, services are increasingly provided through such technological systems (Fig. 1). We wear devices that track our location, conversations, our vital signs, and forward those information to other devices and all that to provide a certain service. We refer to the service provider as the *service system* [1]. Managing and trusting the service provision is becoming more challenging with such complex technological service systems. We increasingly rely on such technological service systems to either make decisions for us or to steer our decisions by providing crucial feedback. Many critical systems such as nuclear, transportation or medical systems have come to depend on such technological service systems. We refer to those as Dependable Technological Service Systems (DTSS). For example, a modern car is a DTSS that facilitates many services through its embedded technology. It has up to 100 embedded computers inside, and its software is designed to make many critical decisions. Many ethical challenges have emerged with the increase of autonomy of such smart devices and our reliance on the decisions made by the software that runs inside of those devices. For example, many safety related ethical issues have been raised regarding the autonomous vehicles, such as "the trolley problem" [2], which raises the issue that the smart vehicles will face ethically challenging situations and they will have to be programmed to make a decision that may be regarded as moral by some, and immoral by others. As these design decisions are highly safety and security related, the ethical challenges arise from the way they are implemented using the current best safety and security practices.

While the DTSS properties such as safety and security are often regulated by domain-specific and property-specific standards, the ethics of the system is conventionally left implicit. The standards usually require either implicitly or explicitly an assurance case to be provided. An assurance case is documented as a clear and well-structured argument to assure a particular property of the system [3]. Such assurance case is then reviewed by the regulation bodies to establish its validity and whether it communicates sufficient confidence that the system exhibits the specific property. The lack of coverage of the ethics in such regulations is due to the fact that different individuals and societies are governed by different ethical principles. Hence, the companies provide a statement with the product regarding certain ethical issues (e.g., privacy) to communicate and assure the user how some of the ethical challenges have been handled within the system. The difference between these two assurances to the regulation body on the one hand, and to the user on the other, is the language and the clarity of the assurance case argument. On the one hand, the argument provided to the regulatory body is required to be clear and well-structured, even a graphical argumentation notations have been established to avoid documenting the assurance case in ambiguous natural language. On the other hand, the assurance provided to the user is mystified in general and written in ambiguous natural language. For example, even in the medical domain where the notion of "informed consent"

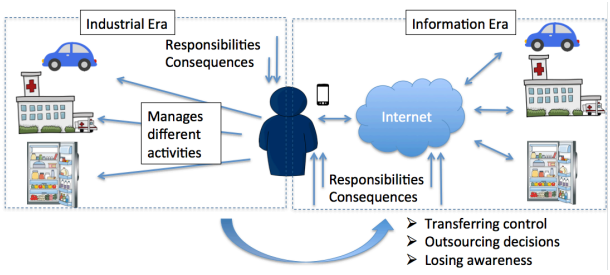

Fig. 1. The world in industrial and information eras.

is used to ensure that the fundamental ethical requirements are satisfied, the ambiguity and clarity of the informed consent texts is questionable [4].

Due to the increase of ethical challenges in DTSS, we argue that both assurances to the user and to the standardisation body should have the same properties, i.e., both should be clear, unambiguous and well-structured. While the assurance to the standardisation body presents how the requirements stated in the corresponding standard are fulfilled, the assurance to the user does not have clearly stated requirements or a standard to follow as they are challenging to standardise. The requirements imposed by the standards are based on the ethical conduct of the society, hence they focus on the general well-being. Those ethical attributes not covered by the standards remain for the user to judge based on their personal moral code. The idea behind the ethics assurance cases is to cover the ethical attributes that remain to the user to judge, which aligns with the agent-centred deontological approach [5] where stress is on the actor's moral rules, as it captures individual responsibility of DTSS users, e.g., when accepting terms and conditions of a service. The approach stresses the set of duties and obligations that can be referred to a contact between a user of a system and the system provider. To document such agent-centred ethics assurance case, we propose structured argument patterns via Goal Structuring Notation (GSN) [6]. GSN representation can help to increase readability and decrees ambiguity of the argument presenting how the different ethical attributes have been addressed in the system.

The main contribution of this paper is the definition of an ethics assurance case and a framework for building an ethics assurance case for DTSS. More specifically, we propose a technique to identify which parts of the system are ethically challenging from the agent-centred ethics approach. We focus on the safety and security critical systems and align the technique with the existing safety and security analyses. Then we define assurance case argumentation patterns for presenting how the different identified challenges have been addressed. Finally, we develop an ethics assurance case for an example system by using the proposed ethics assurance case development approach. For example, consider a case of a health monitoring device that analyses users vital signs and stores them on a cloud storage for further usage, which user might not be aware of. To assure that besides the medical challenges covered by the informed consent, other ethical challenges have also been addressed, the ethics assurance case should cover aspects of the informed consent such as its content, the completeness of the information in the content and the readability/clarity of that informed consent form. Evidence that the information consent is sufficiently readable could be obtained through randomised user interviews about their understanding of the informed consent.

The remainder of the paper is organised as follows: In Section II we introduce essential background information. We introduce the framework for identifying and categorising ethically challenging decisions in Section III, and in Section IV we present an argument pattern for structuring the ethics assurance case. In Section V we illustrate the application of the framework for the instantiation of the proposed pattern on an illustrative example. We present discussion in Section VI and related work in Section VII. Finally, in Section VIII we bring conclusions and future work.

II. Background

In this section we briefly introduce the relevant ethics theories. Then, we present the essential information about safety and security analyses. Finally, we introduce the basic notions of assurance case representation via GSN.

A. Normative Ethics Theories and Agent-centred Approach

We consider **ethics** as a set of moral principles that result in decision making. In particular, we consider normative ethics, i.e., the branch of ethics that deals with moral choices and reasoning on which those choices are grounded. Main theories within normative ethics are deontology, consequentialism and virtue ethics [7]. We briefly recall deontology and consequentialism theories, as these two theories are relevant for DTSS.

Deontology implies acting according to one's moral rules. It stresses that choices cannot be justified only by their consequences, but they can be morally prohibited even if the consequences are good. Therefore, within this theory the notion of being right overrides bringing good. It has several sub-streams, e.g., agent-centred and patient-centred deontology. Agent-centred deontology prescribes to derive moral rules from individual perspective, i.e., each person has individual permissions and obligations and makes decisions how to act [8]. In contrast, patient-centred deontology stresses on the rights, e.g., the core right is the one against being used. In this theory rights violations override good consequences.

Consequentialism puts consequences as the main reasoning for decision making process. Generally, consequentialism is agent neutral. One of the main challenges within this direction is to define what is good. Moreover, consequentialism is usually complemented with additional assumptions allowing for prevention of sacrificing for the greater good, e.g., as in Transplant example [9], and make it less demanding, i.e., define within the theory only what is right, not what is wrong.

B. Safety Analysis

Safety is usually defined as *"freedom from unacceptable risk"* [10], where risk relates to the probability and occurrence of harm and its severity. Since it is not practically feasible to achieve absolute safety, i.e., absolute risk-free systems, safety standards define what are the acceptable levels of the remaining risk in the system. To show that the risks have been reduced to acceptable levels, standards usually require implicitly or explicitly that a safety assurance case [3] is provided to assure that the top-level events that lead to unacceptable accidents have been sufficiently addressed. Those top-level events are referred to as **hazards**, and are defined as *"a system state or set of conditions that, together with a set of worst-case environmental conditions, will lead to an accident"* [11]. Different safety analyses are used to examine the hazards and the events that may cause them. For example, Fault Tree Analysis (FTA) is one of the most commonly used techniques for preliminary safety assessment. FTA is a deductive failure analysis technique which focuses on a single undesired event and methodologically determines the causes of that event [12].

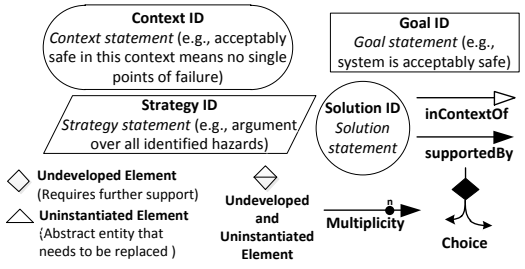

Fig. 2. Overview of the GSN elements

C. Security Analysis and Threats

Security can be defined as a condition that allows *"an enterprise to perform its mission or critical functions despite risks posed by threats"*, where a **threat** can be defined as *"the potential source of an adverse event"* [13]. Threat modeling takes system assets, i.e., values that need to be protected against an adversary, and system entry point, i.e., how it can be accessed, as an input. As an output, threat modeling derives identified high-level threats that should be analyzed [14]. System characterization includes defining scenarios of how it can be used, identification of dependencies and assumptions. One of the formal approaches for threat modeling is based on attack trees [15]. In the Attack Tree Analysis (ATA), the root of the tree is a possible adversary goal and leafs show ways of achieving the goal, i.e., possible threats. Each leaf can be further considered as a sub-goal and has its own leafs connected via OR or AND logical gate. The tree nodes can be assigned with the related cost of goal completion, which is used to calculate the overall cost of the system asset breach. There is a formalised extension of ATA [16] that introduces the notion of an attack suite, i.e., a collection of attacks presented by the tree, and formal language to operate with them and their attributes, e.g., attack cost.

D. Assurance Case Representation

An assurance case is a generic term for assuring a specific property such as safety or security. For example, the work performed during the development of a safety-critical system such as compilation of all the hazards, the results of their analysis and development of safety measures to address them are all part of the safety assurance case. A safety assurance case is composed of all the work products produced during the development as well as the argument that connects the requirements about the hazards and the evidence supporting those requirements. A safety assurance case is presented in the form of an explained and structured argument supported by evidence to clearly communicate that the system is acceptably safe to operate in a given context [17]. GSN is a graphical argumentation notation that can be used to record and present the main elements of any argument. A subset of the GSN elements used in this paper is shown in Fig. 2. The main purpose of GSN is to show how **goals** (claims about the system), are broken down into **subgoals** and supported by **solutions** (the gathered evidence used to back up the claims). The rationale for decomposing the goals into subgoals is represented by **strategies**, while the clarification of the goals (their scope and domain) is done in the **context** elements. The *inContextOf* relation connects goals, solutions and strategies to clarification elements such as contexts, while *supportedBy* relation connects goals with subgoals, strategies and solutions. The bottom row of the signs is used for building patterns of reusable reasoning. The *undeveloped* element can be attached to goals to imply that they need further development, while *uninstantiated* element can be attached to any goal meaning that there are variables marked with curly brackets that need to be instantiated in the element statement. *Multiplicity* relation is used to mark that a goal can be decomposed to multiple subgoals, while a *choice* is used to specify a conditional element in a pattern.

III. Framework for Ethics Assurance

In this section we present a framework for identifying and categorising ethically sensitive aspects of DTSSs. We first present the rationale of the framework, and then propose a way to identify the ethically sensitive decisions made by the system or other involved stakeholders. Finally, we propose decision categorisations based on the agent performing the decision and user awareness of the decision in order to derive the ethical sensitivity of the considered decisions.

A. The Rationale

As mentioned in Section I, the transfer of control and responsibilities in form of decisions away from the user impairs the user's right of influencing the different decisions. The inability to judge the decisions and their outcomes based on the user's individual moral rules contributes to the egocentrism as the main drawback of the agent-centred ethics approach, as mentioned in Section II-A. The decisions that the system makes may result in consequences that are not necessarily affecting only the system, but also the user of the system and in some cases even the user's environment. Moreover, the decisions are not only transferred from users to DTSSs, but the systems enable many stakeholders to make new decisions based on the outcomes of the provided service.

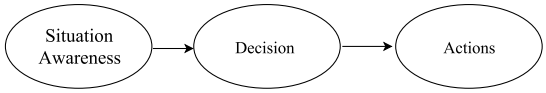

Fig. 3. Situation awareness model

The basic decision-making process (Fig. 3) consists of situation (preconditions) awareness, decision and performance of actions [18]. In information era, the decision-making process concerns the user not only for user's own decisions, but also for decisions of all the stakeholders. For example, a semi-autonomous vehicle makes decisions such as overtaking, braking or disengaging the autonomous mode, which may have direct consequences on the user or the environment. Moreover, it is estimated that one smart vehicle generates around 4 Gigabyte of data per hour. This amount of data stored in a cloud may be accessible to many third party stakeholders that may use the data for different actions such as location tracking or prediction, which may be used against the user's moral code. Hence, preconditions to the decisions of the different stakeholders, the decisions themselves and the consequences in form of actions of those decisions all need to be considered.

We propose a framework (Fig. 4) to assist service providers to identify and analyze the emerging decisions in the information era. Moreover, the purpose of the framework is to build

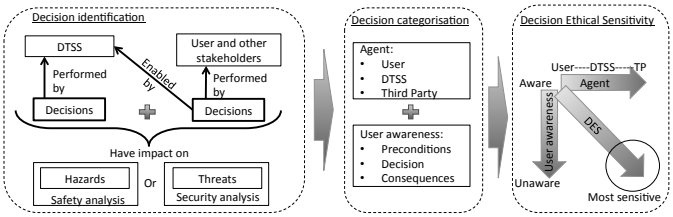

Fig. 4. The framework structure

an ethics assurance case to assure that all relevant information have been clearly communicated to the user so they can form an informed opinion about the ethics of the system. The aim of the framework is not to make moral judgment, but to assist in assuring that enough has been done to allow the users to make their own well-informed moral judgment. The framework consists of three consecutive phases: decision identification, decision categorisation and establishing Decision Ethical Sensitivity (DES). In the first stage we identify the decisions that should be analysed. Then, in the second stage we categorise the decisions based on the agent performing the decision as well as the awareness of the user over the different aspects of the situation awareness model. Finally, in the third stage we establish the sensitivity attributes DES for each decision. We detail the three stages in the reminder of the section.

B. Decision Identification

As depicted in Fig. 4, we primarily consider decisions performed by a DTSS. But since the DTSS in addition acts as an enabler of different decisions, we also consider those decisions performed by the other stakeholders (including the user) that are enabled by the service provision. To narrow down the set of decisions that need to be analysed, we focus on those decisions that are relevant from the perspective of safety or security analysis as for safety and security relevant systems there are already means in place for identifying safety and security relevant decisions.

As mentioned in Section II, the main goal of safety and security analyses is to identify, analyze and handle the top-level events namely hazards and threats. In general, these analyses aim at identifying which events in and around the system may lead to top-level events in order to ensure that the likelihood of the hazards and threats is reduced to acceptable levels. The different identified events, e.g., in FTA or ATA, are closely related to the different decisions that the system and the different stakeholders may make. Although traditional safety analyses are usually made on the system level, and do not include other stakeholders, there are safety analyses such as Systems-Theoretic Process Analysis (STPA) [19] that consider not only the system at hand, but also other socio-technical agents in the identification of causes that may lead to top-level events. Hence, we scope the set of relevant decisions to consider in our analysis by identifying only those that result in actions considered in either safety or security analyses.

C. Categorising Decisions and Establishing DES

Once we have scoped a set of decisions of interest to analyse, in the second phase we categorise each decision with respect to the agent performing the decision as well as the awareness of the user over the different aspects of the situation awareness model. The following agents in the information world are considered: the DTSS, the user of the DTSS and the third party stakeholders (e.g., other DTSSs, or companies). The different agents performing the decisions imply different levels of control or influence that a user can have on those agents and their decisions. For example, a user has the most influence on her own decisions, then the service provider she is using, and finally the user has the least influence on the third party entities out of her direct reach and influence.

TABLE I. USER AWARENESS

Decision	Preconditions	Actions	Categories
Aware	Aware	Aware	**C0**
Aware	Aware	Not Aware	**C1**
Aware	Not Aware	Aware	**C2**
Aware	Not Aware	Not Aware	**C3**
Not Aware			**C4**

TABLE II. DECISION ETHICAL SENSITIVITY

User	DTSS	Third Party	DES Level
C0			**DES0**
C1	C0		**DES1**
C2	C1	C0	**DES2**
C3	C2	C1	**DES3**
C4	C3	C2	**DES4**
	C4	C3	**DES5**
		C4	**DES6**

Besides the agent making a decision, it is important to understand the user awareness of the different aspects of those decisions. We explore user awareness of the three aspects of the situation awareness model for the case when the user is aware of the decision. If the user is unaware of the decision, then the knowledge of the preconditions and the possible actions cannot help the user to predict the decision itself. Consequently, we explore the awareness of the preconditions and actions for the case when the user is aware of the decision. We categorise the user awareness in five categories from C0 to C4 (Table I). We categorise the case where the user is aware of the decision, its preconditions and the following actions as C0 that represents full user awareness. In contrast to that, we assume full unawareness when the user is unaware of the decision. For example, not knowing that a semi-autonomous vehicle can decide to disengage the pilot on its own falls under the C4 category. Alternatively, while knowing that the vehicle may do so, but not knowing the cases under which it may occur nor which consequences it may lead to, falls under C3 category. If the user is aware that the sensor confusion due to an unknown situation is a precondition for the decision to disengage, and that the decision results in disengaging followed by different sounds and alerts to notify the driver, then we consider that the user is fully aware of the decision and categorise it as C0.

In the third stage of the framework we establish the ethical sensitivity for each of the decisions. We define the Decision Ethical Sensitivity (DES) in terms of the agent and user awareness categorisations. As depicted in Fig. 4, lower user awareness and less influence on the agent lead to higher ethical sensitivity of the decision. We define seven DES levels (Table II), from DES0 – the decisions that are not ethically sensitive, to DES6 – the most ethically sensitive decisions.

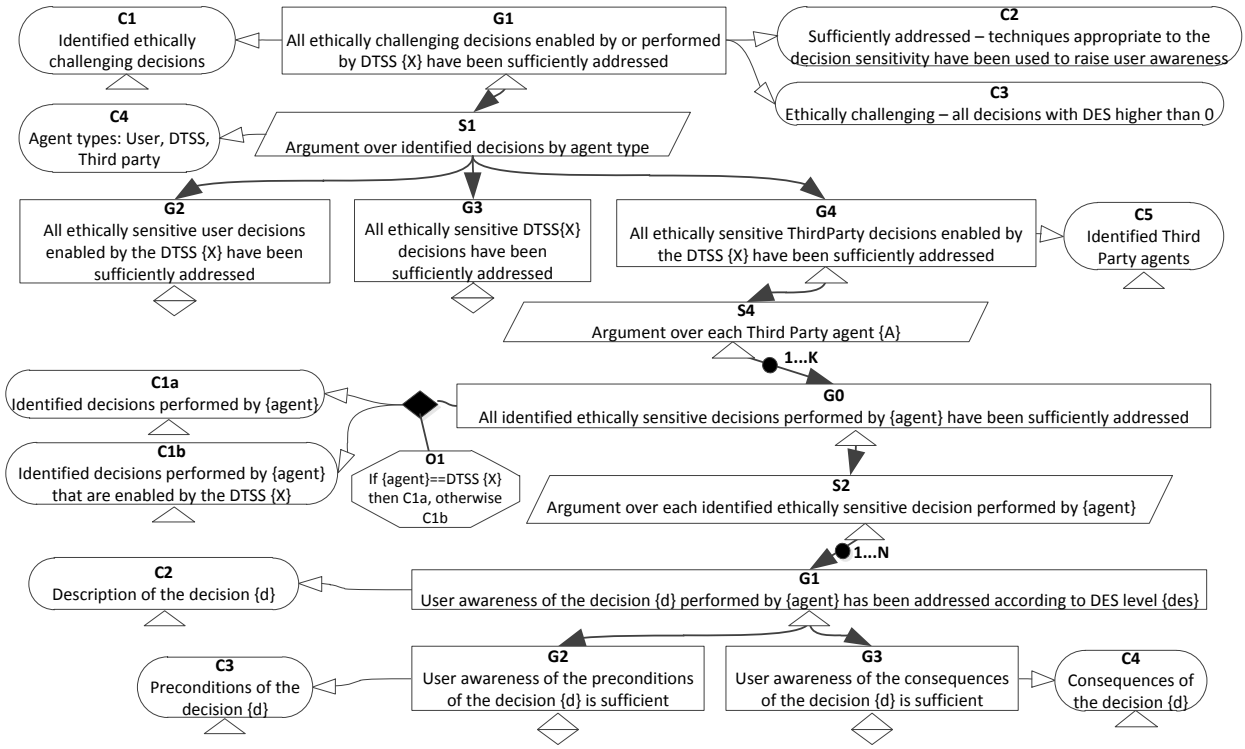

Fig. 5. The ethics assurance argumentation pattern

IV. Ethics Assurance Case

As discussed in Section II-D, the purpose of an assurance case is to build confidence that a certain property of a system has been achieved by presenting evidence in a clear, unambiguous and structured way. Just as safety assurance case assures that the system is acceptably safe, we propose that we build ethics assurance case to assure user awareness of the ethically sensitive aspects in the system. As ethics is a subject of both societal and individual rules, the current standards impose societal ethical norms on different systems, while specific individual moral norms are out of the scope of such standards and are left to the individuals to check their satisfaction. As users are not always given full information to do so, we have proposed a framework in Section III to ensure that individual users of a DTSS have the chance of assessing the ethically sensitive decisions performed by or enabled by the DTSS. We build the ethics assurance case based on the proposed framework. The ethics assurance case arguments for each decision should show that sufficient effort has been invested in ensuring that the user has been made aware of the identified decision. The requirements for establishing the sufficiency of the invested effort are out of the scope of this paper. We assume that techniques with different efficiency of raising awareness can be used for different decisions based on their ethical sensitivity levels.

The ethics assurance argument pattern is presented in Fig. 5. As the top-level goal of an ethics assurance case we aim to assure that all ethically challenging decisions either enabled by or performed by the DTSS have been sufficiently addressed. We clarify in the beginning what do we mean by sufficiently addressed and ethically challenging by the *C2* and *C3* contexts, while the *C1* context points to the identified decisions. To assure that this goal is satisfied, we need to assure awareness of each decision. To make the argument structured and increase its clarity, we break it down with the strategy *S1* to argue over the different agents of the decisions so that we have separate arguments for the decisions of the user, the DTSS, and each of the third party agents.

For each different agent, we further develop the argument pattern on the "decision-level". We list the identified decisions, and then for each decision we assure the claim that user awareness of the decision has been addressed in accordance to its level of ethical sensitivity DES. We break down the argument in two parts (goals G2 and G3) to assure that sufficient steps have been taken to raise awareness of not only the decision but also its preconditions and consequences. To further clarify the argument, we include the three contexts C2-C4 to present the decision, its preconditions and consequences. We envisage that when instantiating this pattern, evidence should be presented to support all three claims related to the awareness of the decision, and its preconditions and consequences.

V. Illustrative Example

In this section we first briefly describe the semi-autonomous vehicle example, and then we illustrate the application of the proposed framework on the example.

A. The Semi-autonomous Vehicle Example Description

We consider a semi-autonomous vehicle as a DTSS, that provides full autonomy on a specific part of the road, e.g., highway. It collects data from different subsystems, e.g., image recognition system, as well as external systems in case of vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communication functionality. There can be different formats

TABLE III. DECISIONS ANALYSIS

Decision	Preconditions	Actions	Agent	User Awareness	DES	Challenging Scenario
To assist the driver with speed control.	— image recognition system detects a speed limitation sign; — AAS is on; — the speed control option is on.	To control the speed of the car	Car/System	C4	DES5	The driver is not fully aware of AAS functionality, i.e., he does not know that the speed of the car can be controlled by the system. There is someone in need of emergency medical care in the car as a passenger and the driver drives to a hospital.
To enable V2I communication.	— the data processed by the road infrastructure is stored; — the location sharing option is switched on.	To broadcast car location.	Driver/User	C3	DES3	There is a saved route of the car and the driver is not aware of its existence.
To process data by using a cloud.	— V2I communication is enabled; — traffic situation with the input from the car location is analysed externally, therefore, data (car position) is shared with the road infrastructure system.	To store data on the server of a third party.	Road Infrastructure/Third Party	C4	DES6	Data is protected, i.e., encrypted, during the transmission, but stored as plain text inside the server. An adversary has access to a server of the cloud, i.e., the route information is available for the adversary.

and interfaces of V2I communication, we assume that there is an option of sharing location, which has predefined value. The data is not only collected by the system, but also processed, and the car makes decisions based on the outcome of its analysis, e.g., take the control over steering wheel. Such system is not isolated and there is data exchange with other systems, e.g., the road infrastructure system, and interactions with a driver, e.g., releasing control of the steering wheel to the driver. Therefore, this system also enables decisions of third parties and the user of the system. Active assistance system (AAS) is a part of the semi-autonomous vehicle. When AAS gets an input from the image recognition system in form of the speed limitation on the current part of the highway, AAS can make a decision to control the car speed. For example, if the car is driving faster, it may reduce the speed to the allowed range.

B. Building the Ethics Assurance Case

To illustrate our approach we identify several decisions performed by different agents and make assumptions about possible user awareness of the decisions. Moreover, for each decision we consider a scenario, in which it can lead to an ethically challenging situation that also relates to possible hazards or threats. The examples of the decisions are summarised in Table III. Each decision is assigned with a user awareness category and DES, furthermore part of the assurance argument for one of the decisions is presented. Since complex systems are characterised by chains of decisions leading to a specific set of actions, we consider decisions that precede actions immediately without decisions in between, further preceding decisions are a part of preconditions.

The first considered decision is *to assist the driver with speed control*. The decision to take control over the car is performed by the AAS and is considered within the framework as it relates to the hazard of a user loosing control of the car. The assumption is that the driver can know that there is an image recognition system and that the AAS is enabled. However the driver may not be aware of AAS full functionality, i.e., that AAS can take over control of the speed. Thus, the user does not know all preconditions and consequences. We categorise the user awareness of this decision performed by the AAS as C4. Hence, according to the Table II, its ethical sensitivity is DES5.

This decision is additionally illustrated by a part of the assurance argument presented in Fig. 6. The argument presents an evidence to assure that sufficient effort has been made to raise awareness of this decision according to its DES level. User awareness of the decision itself should be increased by explicit description of the whole AAS functionality in the user manual. It also should be communicated additionally by a seller, as it is critical information and relates to safety functionality. Moreover, sticker on the dashboard should notify a driver about embedded functionality, as usually the car is driven not only by the person who bought it. Awareness of preconditions and consequences should be increased by addressing it in the user manual and a seller guidebook as well. Knowledge of preconditions may increase predictability of the instance when decision is done. Knowledge of consequences may bring understanding of functionality to the user, i.e., make the awareness of the decision complete.

The second considered decision from Table III, *to enable V2I communication*, is made by the user. However, it is enabled by the system providing the functionality. The decision is investigated as it relates to threats for user data confidentiality. We assume that the driver may not know about predefined setting for the position sharing along with the existence of this option. The driver may also be unaware that the data is not only processed, but stored as well. As a consequence, the unawareness of the stored routes can cause privacy challenges for the DTSS. The last presented decision *to process data by using a cloud* is performed by the road infrastructure system, i.e., by a third party, but enabled by the car. We consider this decision as it can lead to threats to user data confidentiality. We assume that the V2I communication is enabled and data, e.g., car location, needs to be analysed externally. As described in Table III, the user data can be exposed under an assumption of cloud service provider security breaches.

VI. DISCUSSION

As discussed in Section I, standardisation bodies assess only a portion of ethically challenging aspects of DTSSs, while many are left for the user to assess. The problem we are addressing in the paper is that the assurance information for the standardisation body is clear, unambiguous and well-structured, while the information for the user rarely comes with the same properties. To address the problem we have proposed the notion of ethics assurance case to structure the information of interest for the user, and assure that sufficient efforts are made to raise user awareness. We understand that it is not practical to expect from an average user to understand all the technical aspects of certain complex decisions. But this does not mean efforts should not be made to make it easier

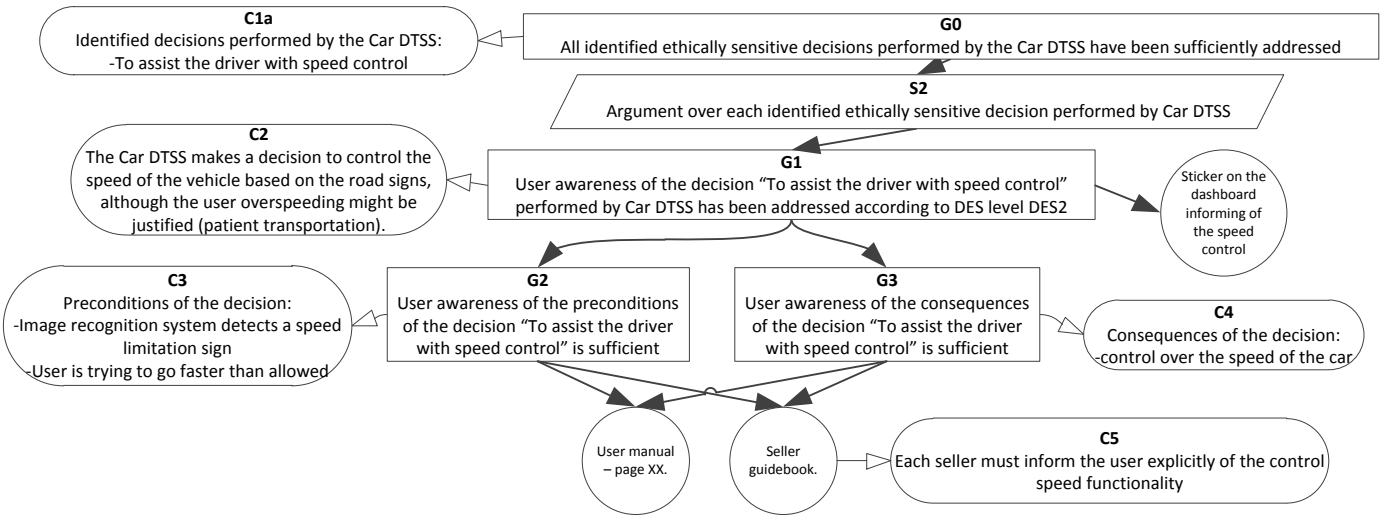

Fig. 6. Speed control assist decision argument

for the users to perform their assessment. To support freedom of choice of consumers, the society should ensure that the consumers have the needed information to make the choice. We envisage that ethics assurance case would not only be intended for the end-user, but also for certification organisations that should assess whether sufficient effort was made to raise user awareness of all the ethically challenging decisions. The very existence of an ethics assurance case for a system is an important step in ensuring sufficient user awareness efforts.

Hans Jonas in his work "The Imperative of Responsibility" expresses his fear of the power of technology and calls for reevaluating the traditional ethical assumptions [20]. This especially applies for the todays technology-driven consumer society, where consumers are given much of the responsibility without full awareness of the consequences of the services provided or they enabled by the ubiquitous technological service providers. Due to the egocentrism of the consumer society, we put agent-centred deontology as a base for our framework, so that ethical aspects of the system could be investigated from the user perspective. We deem that a way to address the inherent egocentrism is by providing clear, unambiguous and well-structured information about the different aspects of technology-enabled decisions the user is responsible for. Although raising awareness does not necessarily influence every users choice of using a product or not, we assume that supporting access to information for those that are interested in knowing the consequences of using a DTSS is an important step in addressing the issue of responsibility.

VII. Related Work

Nowadays, technologies are developing with an alarmingly fast speed and ethics behind them has difficulties of keeping up with the pace. Therefore, there are many works targeting ethics aspects and challenges for the emerging Information systems (IS). Discourse or argumentation ethics can be applied to tackle ethics challenges of ISs [21]. We in our work focused on agent-based approach that relates more to DTSSs. However, the approach of applying discourse ethics, e.g., universalisation, can be used to develop techniques for awareness enhancement. Assigning responsibilities for safety-critical systems with different categories of its autonomy addresses ethical challenges of cyber-physical systems [22]. Our framework can be applied to systems with different levels of autonomy as long as the system-related decisions can be identified. It would be interesting to see how our approach can be used to complement responsibility assignment, as we consider a service provider being responsible for the awareness enhancement without detailing the responsibilities further. The importance of users becoming aware of ethical challenges and their connection to different technologies was investigated before [23]. However, we make one step further proposing a way to assure the awareness. Another approach that targets ethical analysis of emerging technologies [24] includes several ethical levels, namely: technology, artifact and application. Each level has its objects capable of raising ethical issues. Such categorisation can be used to extend our framework and identification of DTSS-related decisions.

An approach to formal verification of ethical choices for autonomous system [25] considers ethical side of decisions made by autonomous systems. The approach formulates a theoretical framework that allows to constrain behavior of an autonomous system according to a predefined paradigm. In contrast to this approach, we do not consider how ethical are particular decisions. Instead, in our framework we aim to make a system user aware of the decisions and support the user's own assessment of how ethical are certain decisions. One of the ethical challenges in robotics is that a robot does not reflex on ethical aspects of its actions [26]. In our framework we solve it by increasing the user awareness of the logic behind the provided services, so that the user can assess the DTSS in accordance with her own ethical code. Another approach [27] focuses on technology and service transparency in how the data is managed and decisions are made for the users in IoT systems. It considers a configurable policy-based framework that targets to provide a higher level of control for users over their interactions with IoT systems. While such framework should be implemented at the design phase of the system, our framework also can be considered at the design phase, e.g., decision identification part of it, however the main target of our framework is to assure user awareness level for DTSSs once their architecture is designed

and the related safety and security analyses conducted. One more approach [28] to tackle ethical challenges of autonomous systems focuses on shaping the behavior of such autonomous system according to the controlled ethical paradigm, e.g., by using an ethical governor, a system component, that evaluates any system response relating to lethal consequences before the response. While the described approaches relate to tools for establishing ethics in the system, our framework complements them by introducing ethics assurance cases to clearly and unambiguously present how the ethical challenges have been addressed in the considered DTSS.

VIII. Conclusions

The current standards for safety and security critical systems leave ethics implicit and users are often not made aware of all the information needed to form an informed opinion about ethics of the systems. In this paper we advocate for the need of ethics assurance cases as a way to ensure that sufficient efforts have been made to inform the users of the ethically challenging decisions that a dependable technological service system they are using might be performing or enabling. One of the main challenges in building such an ethics assurance case is the identification of ethically challenging decisions. We focus on the dependability technological systems as they have established analyses for identifying critical decisions in the system that may lead to different types of harm. We build upon those analyses and propose to identify ethically challenging decisions as well. Moreover, we propose a way to establish the ethical sensitivity of those decisions based on the estimate of the average user awareness of the different aspects of the decision. We use the identified ethically challenging decisions and their ethical sensitivity as the basis for building an ethics assurance case. We have illustrated on an example how to build an ethics assurance case. Since complex systems may be associated with many ethically challenging decisions that can be usually represented with a decision chain, we have considered only those decisions preceding the challenging actions. Any preceding decisions in the decision chain are considered as a part of the final decision precondition.

Focusing the decision identification phase on safety and security narrows the scope of the possibly ethically challenging decisions that we can identify. It would be interesting to explore widening the base for decision identification. Moreover, more concrete and structured techniques are needed for deriving the decisions from the existing analyses. An interesting future work would be to map the possible awareness techniques with the different decision ethical sensitivity levels. Such mapping would ease determining whether enough has been done to raise awareness over different decisions.

Acknowledgements

Irfan Šljivo is supported by the Swedish Foundation for Strategic Research (SSF) via the project Factories in the Cloud. Elena Lisova is supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n607727. Sara Afshar is supported by the SSF via the project PRESS. We would also like to thank Gordana Dodig-Crnkovic for all the support through her Professional Ethics course.

References

[1] P. P. Maglio and J. Spohrer, "Fundamentals of service science," *Journal of the Academy of Marketing Science*, vol. 36, no. 1, pp. 18–20, 2008.

[2] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars?," *arXiv preprint arXiv:1510.03346*, 2015.

[3] R. Bloomfield and P. Bishop, "Safety and assurance cases: Past, present and possible future–an Adelard perspective," in *Making Systems Safer*, pp. 51–67, Springer, 2010.

[4] M. Paasche-Orlow, H. Taylor, and F. Brancati, "Readability standards for informed-consent forms as compared with actual readability," *New England journal of medicine*, vol. 348, no. 8, pp. 721–726, 2003.

[5] D. R. Forsyth, "A taxonomy of ethical ideologies," *Journal of Personality and Social Psychology*, vol. 39, no. 1, pp. 175–184, 1980.

[6] "GSN Community Standard Version 1," tech. rep., Origin Consulting (York) Limited, November 2011.

[7] L. Alexander and M. Moore, "Deontological ethics," in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), 2015.

[8] L. A. Alexander, "Scheffler on the independence of agent-centered preogatives from agent-centered restrictions," *The Journal of Philosophy*, vol. 84, no. 5, pp. 277–283, 1987.

[9] J. J. Thomson, "Killing, letting die, and the trolley problem," *The Monist*, vol. 59, no. 2, pp. 204–217, 1976.

[10] CENELEC, *IEC 61508: Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems. Part 4: Definitions and abbreviations*. UK Ministry of Defence, 2007.

[11] N. G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press, 2011.

[12] SAE and EUROCAE, *ED-135/ARP-4761: Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment*. Society of Automotive Engineers, 1996.

[13] R. Kissel, *Glossary of key information security terms*. U.S. Dept. of Commerce, National Institute of Standards and Technology, 2006.

[14] F. Swiderski and W. Snyder, *Threat Modeling*. Redmond, WA, USA: Microsoft Press, 2004.

[15] B. Schneier, "Attack trees," *Dr. Dobb's Journal*, vol. 24, no. 12, pp. 21–29, 1999.

[16] S. Mauw and M. Oostdijk, "Foundations of attack trees," in *Proceedings of the 8th International Conference on Information Security and Cryptology*, ICISC'05, pp. 186–198, Springer, 2006.

[17] T. Kelly, *Arguing Safety — A Systematic Approach to Managing Safety Cases*. PhD thesis, University of York, York, UK, 1998.

[18] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors The Journal of the Human Factors and Ergonomics Society (HUM FACTORS)*, vol. 37, no. 1, pp. 32–64, 1995.

[19] T. Ishimatsu *et al.*, "Hazard analysis of complex spacecraft using systems-theoretic process analysis," *Journal of Spacecraft and Rockets*, vol. 51, no. 2, pp. 509–522, 2014.

[20] H. Jonas, *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago press, 1985.

[21] J. Mingers and G. Walsham, "Toward ethical information systems: The contribution of discourse ethics," *MIS Quarterly*, vol. 34, no. 4, pp. 833–854, 2010.

[22] A. Thekkilakattil and G. Dodig-Crnkovic, "Ethics aspects of embedded and cyber-physical systems," in *The 39th Annual International Computers, Software & Applications Conference*, July 2015.

[23] D. Popescul and M. Georgescu, "Internet of things some ethical issues," *The USV Annals of Economics and Public Administration*, vol. 13, no. 2(18), pp. 210–216, 2013.

[24] P. A. E. Brey, "Anticipating ethical issues in emerging it," *Ethics and Information Technology*, vol. 14, no. 4, pp. 305–317, 2012.

[25] L. Dennis, M. Fisher, M. Slavkovik, and M. Webster, "Formal verification of ethical choices in autonomous systems," *Robotics and Autonomous Systems*, vol. 77, pp. 1 – 14, 2016.

[26] B. C. Stahl and M. Coeckelbergh, "Ethics of healthcare robotics: Towards responsible research and innovation," *Robotics and Autonomous Systems*, vol. 86, pp. 152 – 161, 2016.

[27] G. Baldini, M. Botterman, R. Neisse, and M. Tallacchini, "Ethical design in the internet of things," *Science and Engineering Ethics*, 2016.

[28] R. C. Arkin, P. Ulam, and A. R. Wagner, "Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception," *Proc. of the IEEE*, vol. 100, pp. 571–589, March 2012.