Representing decision-makers in SGAM-H: the Smart Grid Architecture Model Extended with the Human Layer*

Adam Szekeres[1] and Einar Snekkenes[1]

Department of Information Security and Communication Technology,
Norwegian University of Science and Technology - NTNU,
Gjøvik, Norway
{adam.szekeres,einar.snekkenes}@ntnu.no

Abstract. The safety and security of critical infrastructures is both a technical and a social issue. However, most risk analysis methods focus predominantly on technical aspects and ignore the impact strategic human decisions have on the behavior of systems. Furthermore, the high degree of complexity and lack of historical data for probability estimations in case of new and emerging systems seriously limit the practical utility of traditional risk analysis methods. The Conflicting Incentives Risk Analysis (CIRA) method concentrates on human decision-makers to address these problems. However, the method's applicability is restricted by the fact that humans are not represented in the Smart Grid Architecture Model (SGAM) which is the industry's most well-known model of the Smart Grid ecosystem. Therefore, the main objective of this paper is to establish a connection between CIRA and SGAM by proposing the SGAM-H, an enhanced version of the original architecture model complemented by the Human Layer. The development and evaluation of the artifact is guided by the Design Science Research methodology. The evaluation presents a working example of applying the CIRA method on a scenario involving intra-organizational risks at a Distribution System Operator. The key benefit of the SGAM-H is that it enables the construction of a common understanding among stakeholders about risks related to key decision-makers, which is a fundamental first step towards forming a more complete picture about potential issues affecting the electric grids of the future.

Keywords: Information security risk analysis · Conflicting Incentives Risk Analysis (CIRA) · Smart Grid Architecture Model (SGAM) · SGAM-H · Human Layer · Stakeholder motivation.

1 Introduction

Nation-wide electrification of industries and societies beginning in the 1880s had tremendous economical and societal benefits [7] and the demand for a stable and reliable supply of electricity has exceeded that for any other forms of energy [28]. A properly

* This work was partially supported by the project IoTSec – Security in IoT for Smart Grids, with number 248113/O70 part of the IKTPLUSS program funded by the Norwegian Research Council

functioning power grid represents an indispensable infrastructure for modern societies, which supports all aspects of life. While demand for electricity will keep rising in the future (e.g., due to increasing electrification of the transportation sector, growing populations, etc.) international directives and regulations have been pushing toward a shift from dependency on fossil and nuclear power sources to more eco-friendly and sustainable renewables. Most renewable power sources (e.g., wind, solar) are intermittent in nature which requires a paradigm shift from centralized large-scale generation models to flexible, distributed and small-scale solutions [11]. At the same time economic constraints make the complete reconstruction of the power grid highly unfeasible. The envisaged solution is encompassed in the concept of the Smart Grid (SG), which aims at solving the challenges of the future by relying on the physical infrastructure of the past with enhancements from novel information and communication technologies. Thus the SG represents a highly complex system with real-time sensing and control capabilities using a bidirectional flow of electricity and information, enabled by the addition of internet of things (IoT) devices at various parts of the grid. Several stakeholders are involved in SG-related activities including: legislators, governmental agencies, standardizing bodies, data protection authorities, organizations focusing on the generation, transmission, distribution of electricity, equipment manufacturers, software and security providers, researchers and consumers [8].

Developments in SGs are driven by a combination of political, economic and ecological motives. Misaligned incentives are unavoidable when the number of interacting stakeholders is considered in a system of such complexity (both technically and socially). Misaligned incentives are particularly prevalent in information systems where those who are responsible for providing security are not the same people who benefit from the protection or suffer when things go wrong. For example, increasing the dependency of critical infrastructures on public information systems (network convergence) can be an efficient short-term cost saving strategy for utility companies, but it increases society's long-term vulnerability, which will ultimately bear the costs [24]. It has been demonstrated that misaligned incentives, negative externalities and moral hazard arise in a variety of settings within the field of information security [1]. The identification and mitigation of such problems is crucial for ensuring the safety and security of societies depending on SGs and other critical infrastructures.

1.1 Conflicting Incentives Risk Analysis (CIRA)

The Conflicting Incentives Risk Analysis (CIRA) method focuses on the motivation of individual stakeholders to define risks. The lack of relevant historical data in case of emerging and dynamic systems creates a significant challenge for traditional (i.e., relying on frequentist probability estimations) risk analysis methods [37]. Furthermore, deliberate human actions due to misalignment of incentives is rarely at the center of risk analysis procedures. CIRA defines risk as the misalignment between stakeholder incentives. The analysis focuses on the *Risk owner's* (i.e., person at risk) exposure to the actions or inactions of several other stakeholders (*Strategy owners*) who are in the position to choose courses of actions [32]. CIRA combines quantitative methods to characterize risks attributed to key decision-makers, therefore, aims at overcoming some of the problems associated with qualitative risk scoring methods [15].

1.2 Smart Grid Architecture Model (SGAM)

The creation of the Smart Grid Architecture Model (SGAM) was motivated by the need to represent stakeholders, applications and systems that will have to achieve efficient interdependent operations in future SGs. To ensure these goals, developers and standardization bodies of the SG need to have a common understanding or shared model about the systems which will be implemented. To capture the EU-specific requirements the SGAM was designed to tackle the complexity by representing systems in a consistent and comprehensive way. It enables standards gap analysis; visualization and assessment of use cases in a technology-neutral way; comparison of different approaches and roadmaps from various viewpoints. Figure 1 presents the SGAM, based on [4]. *Domains* represent the energy conversation chain from generation site to customer premises. *Zones* capture the power system management supported by ICT from the level of processes to markets. *Interoperability layers* represent different levels of abstraction from the physical hardware to business perspectives highlighting the interconnectedness and dependencies between entities.

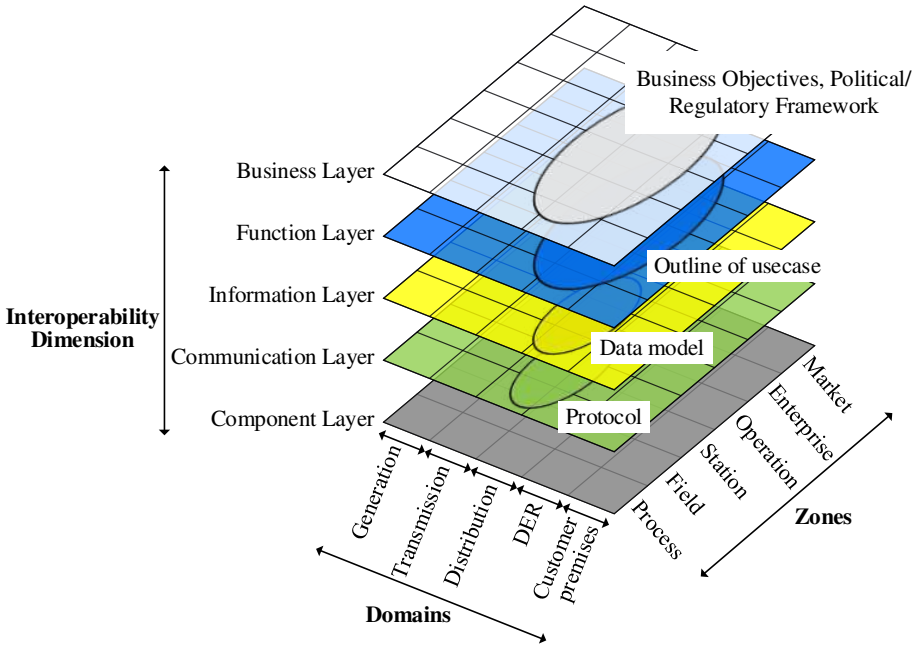

Fig. 1. The Smart Grid Architecture Model (SGAM) based on [4].

How is it possible to analyse risks arising from human decision-making in a complex system as the SG? Several management failures (management of tree growth, lack of vulnerability and system-health assessment, etc.) contributed to the 2003 Northeast

blackout in the US, affecting 55 million people with an estimated economic impact of $6 billion [25]. Organizations responsible for the development and maintenance of the grid need to have the right incentives in place to achieve their goals at a socially optimal level. Are measures in place to protect the privacy of customers despite increased monitoring capabilities enabled by smart meters and other smart home devices [22]? Does information security contribute to the organizational goals or is it perceived as a impediment to smooth operations [48]? Can the SG fulfill the hopes by providing electricity in a safe, reliable and secure way without significantly increasing society's exposure to new threats [19]?

1.3 Problem statement and motivation

In order to enable the application of the CIRA method on SG use cases, a connection between the models has to be established. Human decision-makers are not represented in the existing SGAM, which may result in ignoring the impact strategic human decisions have on the grid. The SGAM documentation briefly mentions human-aspects: *"The concept of an Actor is very general and can cover People (their roles or jobs), systems, databases, organizations, and devices"* [4]. However some critical distinguishing features justify separating human decision-makers from the Actor concept. Human decision-makers:

- are self-determined (i.e., choosing their own goals [10]);
- have unique motivations, which may not be in alignment with organizational/societal objectives (e.g., principal-agent models [47]);
- are in the unique position to control all other objects (e.g., regulations, business goals, components, etc.) within a system.

Ergo, human decision-makers have distinctive and significant impact on every aspect of the system's behavior which requires the explicit integration of human decision-makers into a reference architecture to provide a more comprehensive model. Furthermore, it is necessary to investigate the CIRA method's adequacy for analysing risks in highly complex emerging systems, where the application of traditional risk analysis methods may be infeasible (due to lack of historical data for probability estimations and unmanageable complexity of information systems).

This paper presents an approach for addressing these gaps in the literature. The paper is structured as follows: Section 2 provides an overview about modifications to the basic SGAM as well as approaches for modeling humans from a broad range of domains. Section 3 describes the Design Science Research Methodology (DSRM) which guided the development and evaluation of the paper's artifact. The artifact is presented and evaluated by a case study throughout Section 4. Section 5 discusses key findings and Section 6 draws conclusions. The paper ends with ideas for further work in Section 7.

2 Related work

This section is divided into two parts. The first part reviews research work which proposes or implements extensions to the generic SGAM to solve specific tasks. A literature search using the search string ("sgam" extend OR extension) appearing anywhere in

the articles was conducted on Google Scholar and articles citing the original publication were screened; other relevant articles were identified among references. Studies describing the application of SGAM were excluded. The second part presents approaches for modeling human behavior across various domains to illustrate design decisions about the models.

2.1 Variants of SGAM

The Information System Architecture for e-Mobility (EM-ISA) is an early SGAM variant focusing on electric vehicle (EV) integration into the grid. The model significantly reduces the number of the domains and zones, then proposes the integration of human-machine interfaces into the model to capture interactions between humans (operators) and objects without further specifying human attributes [35]. The Electric Mobility Architecture Model (EMAM) focuses on EV integration as well. In EMAM, the Generation domain is removed and an electric mobility domain is added to the grid plane, while keeping the rest of the original model unchanged. Recognizing the utility of the SGAM for standardisation purposes, two other reference models were developed following similar architecture engineering principles. While the layers of The Smart City Infrastructure Architecture Model (SCIAM) and the Smart Home Architecture Model (SHAM) are the same as those of SGAM, different domains and zones are introduced which may decrease compatibility between models [46]. SGs may differ between countries, therefore it is important to increase compatibility between various implementations. Two state-of-the-art models (the SGAM from EU and the NISTIR 7628 from U.S.) are combined in order to facilitate security analysis from the beginning of the development process [45]. In addition to the previously described variants two more architecture models are described in [43]. The Home and Building Architecture Model (HBAM) utilizes SGAM's layered approach with different zones and domains introduced to capture relevant concepts within scope of smart homes and buildings. The Reference Architecture Model for Industry 4.0 (RAMI 4.0) is regarded as the most sophisticated derivative of the SGAM containing zones and domains relevant for industrial applications and extending the interoperabilty perspectives with an additional layer. Two more reference models have been developed using the SGAM's design principles. The Reference Architecture Model Automotive (RAMA) represents the life-cycle of connected vehicles and the related information technologies and the Maritime Architecture Framework (MAF) models information exchange between various actors in the maritime domain [44].

2.2 Approaches for modeling humans

Models in general, are abstract representations of a complex entity or phenomenon capturing its most significant aspects for a pre-specified purpose. Analogies, shared features and other similarities between entities play a key role in modelling activities. For example, pigs and other animals can represent humans in medical experiments due to the high number of shared features (in terms of genetics, physiology and anatomy, etc.) [23]. Investigations in road safety require human models which accurately capture the physical properties of real humans in car crash scenarios [2]. Personas or user archetypes

are widely used human models in the software engineering industry. Personas guide the development process by representing future users and their goals in relation to the product [5]. Realism of human models is becoming increasingly important in virtual environments where representations can replace real humans (in communication context [3]) or simulated agents are required to act realistically (in training context [27]). For behavior prediction, a human model must incorporate psychological constructs that are most likely to govern or influence (i.e., mediate and moderate) the behavior of interest. Models reduce real-world complexity, which enables that only a small set of well-defined parameters are required for predictions. The importance of appropriately modeling humans and human behavior has been recognized in a variety of domains. Human performance and mental load models have been developed to represent operator characteristics and to assist the design of human-machine interfaces in the context of industrial control systems [38]. A variety of human behaviors are of interest to the military, therefore a wide range of human models have been developed (at the individual and group level) to support agent-based behavioral simulations [30]. A key challenge is to find the right balance between the model's complexity and its realism [16]. In the context of information security, humans can be represented by a utility function which is the most suitable level of abstraction for game theoretic simulations [20]. People have great impact on the Earth's overall condition, but humans are not yet explicitly represented in Earth system models used for simulating ecological dynamics. The selection of an appropriate human model relies on the modeler's understanding about the strengths and weaknesses of each model [26].

2.3 Summary of related work

The reviewed literature demonstrates the SGAM's acceptance among practitioners and researchers and presents several domain- or task-specific variants inspired by the original model. However, the representation of human decision-makers is lacking, which impedes the efficient application of CIRA on SG scenarios. The broad overview on the literature of human modeling approaches highlights that models should be developed according to relevant design considerations (e.g., specifying the model's content in relation to the behavior of interest, complexity-realism trade off, etc.).

3 Methodology

This study is based on the design science research (DSR) paradigm, which provides an organizing framework for the development of purposeful artifacts to solve a specific problem [14]. The DSR methodology defines three cycles which interact with each other during task execution [13]. The *design cycle* represents the core activities (development and evaluation of the artifact in an iterative process) which is embedded in a broader context. The design cycle receives input from two sources. The *relevance cycle* refers to the interaction between the environment (where problems and needs for a new solution arise) and the design cycle (produces solutions). Artifacts from the design cycle are fed back to the environment through the relevance cycle and the artifacts are applied in the context where they were intended to function. Interaction of the design

cycle with the supporting knowledge-base defines the *rigor cycle* which provides the necessary tools, methodologies, theories for the development and evaluation of the artifact. Information flows in both directions between the rigor and design cycles as well, thus new knowledge and experience resulting from the construction of the artifact are recorded in the knowledge-base using the most suitable format (presentation, tutorial, academic paper, etc.).

The relevance cycle serves as a starting point for any DSR activity by specifying the context and problems in the domain (i.e., requirements), that the artifact should solve. Furthermore, it defines evaluation criteria for testing the artifact's utility within the environment. The need to represent human stakeholders within the SG has been arising from interactions with other stakeholders (students, conference and project participants). Difficulty of creating a common understanding among stakeholders about CIRA's applicability and relevance was identified as a major barrier to the method's acceptance and adoption. Thus, a more efficient method of conveying meaning was set as a requirement. The second step focuses on the identification of suitable theories, frameworks to meet requirements. Therefore, the rigor cycle was used for the identification of existing frameworks by reviewing the relevant literature, which resulted in identifying the SGAM as an ideal candidate requiring customization. The development activity within the design cycle was used to extract key concepts from CIRA and to create visual representations of its abstract concepts. An important design consideration was to keep a high degree of compatibly with the original SGAM version, therefore an extension is proposed: the SGAM-H enhanced by a Human Layer and its necessary components. The artifact model was built from scratch in Microsoft Visio, to ensure reusability and mutability (the Visio-based templates reported in [34] were not available online). The final step within the design cycle is the evaluation of the artifact which is achieved through a hypothetical case study (qualitative, descriptive method) demonstrating how key CIRA concepts are mapped onto the Human Layer and how it conveys meaning. The artifact is evaluated in terms of its efficacy, ease of use, completeness and homomorphism (i.e., correspondence with another model) [31].

4 Human Layer

This section presents the Human Layer as an extension of the SGAM, giving rise to the SGAM-H. The Human Layer's basic elements for constructing and representing the context of risk analysis are introduced. Next, the artifact's efficacy is demonstrated on a hypothetical case study which applies the CIRA method on a SG scenario focusing on risks experienced by the CEO of a Distribution System Operator (DSO). Several aspects of the case study were inspired by media reports [36] and analyses of real-world incidents [25] accompanied by relevant scientific literature [6] in order to increase its realism. Finally, the artifact is evaluated along the previously identified criteria.

Figure 2 presents the Human Layer placed on top of the business layer of the original SGAM. This implementation enables the representation of human stakeholders with their relevant attributes on the architecture model and emphasizes the critical role that strategic human decisions can have on various aspects of SGs.

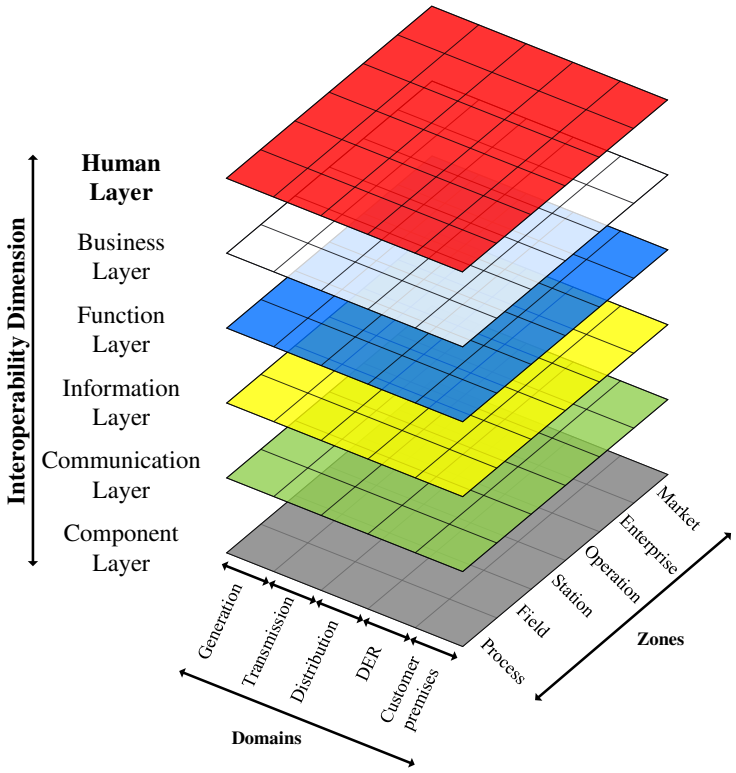

Fig. 2. SGAM-H including the Human Layer.

Figure 3 presents the stakeholder models; components to represent human attributes and other elements of the layer to capture key concepts of CIRA. Two types of stakeholder classes are distinguished by color and related captions: human models in blue represent the risk owner, human models in white represent the class of strategy owners. Post-analysis states are distinguished by a tag above the models to display the risks explicitly (i.e., consequences for the risk owner, incentives for the strategy owner). The sign (+/-) represents the direction of utility change following strategy execution. Furthermore, incentives are marked with red fill color on the strategy owner figures. The height of the red coloring from the bottom of the figure matches with the magnitude of the incentive (i.e., an incentive of 50 produces a red fill color up to 50% of the figure's height). Strategy owners' profile information is captured in brackets, to record the information used for the construction of motivational profiles before the analysis. Stakeholders are linked to other entities (e.g., physical hardware, organizations, etc.) by dashed lines. Strategies are represented by continuous lines ending in an arrow, directed from the strategy owner to the risk owner.

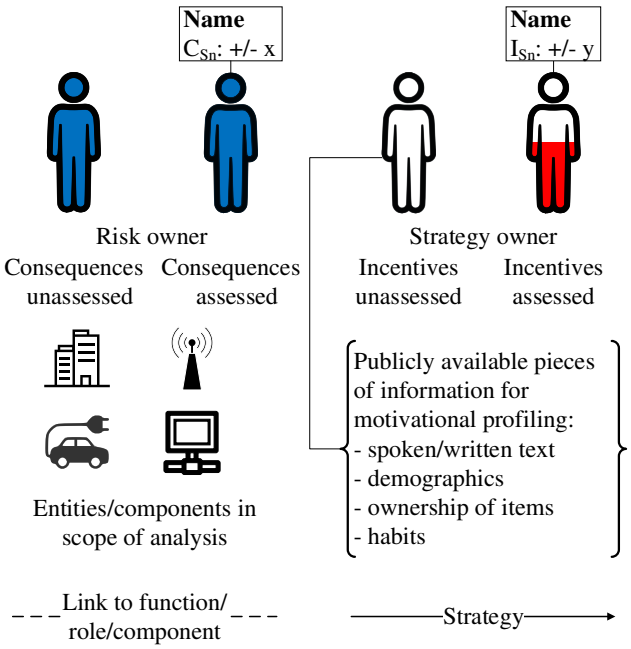

Fig. 3. Components of the Human Layer.

4.1 Case study: DSO risks

This sub-section demonstrates the use of the SGAM-H through a case study in which the CIRA method is applied to a scenario focusing on the risks faced by the organizational leader of a DSO, since the organization has a critical role in the SG ecosystem. Numbering of the subsequent paragraphs follows the steps of the CIRA procedure based on [32].

1. Identification of the risk owner The risk owner is the CEO of a DSO, who is interested in intra-organizational risks which may interfere with the objectives of the organization.

2. Identification of the risk owner's key utility factors The key utility factors (UFs) were identified by relying on the Balanced Scorecard (BSC) method, which was designed to aid managers in evaluating and measuring organizational performance through a set of measures linked to organizational objectives [18]. Four perspectives are distinguished by the BSC method: *Financial*, *Customers and stakeholders*, *Learning and growth* and *Internal business processes*. The method enables the development of key performance indicators at various levels (departments, individuals) to achieve better organizational performance. Since utility companies such as DSOs operate as natural

monopolies due to high infrastructural costs, their operations differ from purely for-profit organizations. In the not-for-profit sector, the financial perspective is often seen as a constraint rather than an objective, which requires different priorities [21]. Some work has been done to adapt the BSC to the specific needs of utility companies [17,33]. Table 1 presents the risk owner's key utility factors derived from the BSC perspectives.

Table 1. Key utility factors of the CEO.

BSC perspectives	**Utility factors**
Financial	Revenue
Customers and stakeholders	Customer privacy
	Contribution to public welfare
Learning and growth	Innovation
Internal business processes	Relationship with regulators

3-5. Identification of strategies that may influence the risk owner's utility factors; Identification of roles and named strategy owners which can execute the strategies Steps 3-5. of the procedure are summarized in Table 2. For each utility factor an appropriate strategy was identified by considering key processes and functions at a DSO. The identification of roles and strategy owners is aided by the organizational chart which allocates the responsibilities and tasks to various roles occupied by actual persons. The scenario description for each person illustrates motivational factors at play regarding the dilemmas they face in a given situation.

Table 2. The risk owners' utility factors (UFs); strategies that impact the risk owner's utility factors; roles and individuals.

Affected UFs	**Strategy**	**Role**	**Person**
Customer privacy	Help a friend (S_1)	Dispatcher	Sigurd
Contribution to public welfare	Fix street lights (S_2)	Operations manager	Emma
Innovation	Recruit research applicants (S_3)	Head of R&D	Hanne
Relationship with regulators	Support system integration (S_4)	CISO	Henry

Sigurd works as a dispatcher at the organization. He is approached by his best friend who suspects that his wife is cheating on him and asks Sigurd to monitor the detailed electricity consumption of their holiday house which he thinks is used as a hideout by her. He has access to the relevant data, and thinks he can fulfil the request without getting into trouble. The legal and financial implications of a privacy breach are of key

interest to the risk owner. Emma is responsible for distributing tasks efficiently within her team of technicians working in the field. Citizens are complaining about faulty street lights and dangerously dark streets. She has to decide how to allocate tasks within the team based on existing efficiency measures in place. Hanne works at the R&D department developing new services for customers. Students with novel ideas apply to get work experience at the organization, but she perceives recruitment and training of students as a nuisance since student projects rarely get converted into successful products. She has to decide whether increasing the number of student projects (to fulfill an important societal role) worth lowering her performance indicators. Henry believes that the new agenda to harmonize all data acquisition systems at the organization would create a singularity threat and he believes in security through diversity. He has the final word regarding the new system's implementation in the project.

6. Identification of the strategy owners' utility factors For each strategy owner two types of utility factors are distinguished. Work-related factors are derived from the BSC method's perspectives. Personal utility factors are represented by basic human values [40]. Table 3 presents the key utility factors for each strategy owner.

Table 3. Work-related and personal utility factors for each strategy owner.

Strategy owner	Utility factors						
	Work-related (associated with role)	Personal					
Sigurd	Percentage of successfully located faults and dispatched repair teams within time frame (%)	ST	OC	CO	HE	SE	
Emma	Percentage of reconnected electricity customers within time frame (%)						
Hanne	New services ready for market (%)						
Henry	Percentage of resolved cyber-incidents within a time frame (%)						
Note. ST: self-transcendence, OC: openness to change, CO: conservation, HE: hedonism, SE: self-enhancement.							

7. Operationalization of utility factors To operationalize the utility factors, existing work on DSO-specific KPIs was surveyed [6,12] as well as relevant regulations (GDPR [9]). KILE (quality-adjusted revenue frames for energy not delivered) represents customers' costs for interruptions, and is a form of revenue reduction due to interruptions, which aims at incentivizing utility companies to maintain operational reliability [29]. Utility factors capturing personal motivations were operationalized in previous work as publicly observable pieces of information, for the construction of motivational profiles [39,41,40]. Table 4 presents how each utility factor is operationalized.

Table 4. Utility factors operationalized.

Role	Type of utility factor	Utility factor	Operationalized as
Risk owner	Professional	Revenue	R = Revenue cap - KILE (CENS) [29]
		Customer's data privacy (%)	CDP = 1 - (privacy-related penalties/privacy breach cap (0.04*annual turnover)) [9]
		Contribution to public welfare (%)	PW = resolved public complaints within 1 month / all complaints in a period
		Innovation (%)	INN = number of established research collaborations with universities / number of applications from students
		Relationship with regulators (%)	REG = number of reports accepted without modification / all reports submitted
Strategy owner		Percentage of successfully located faults and dispatched repair teams within time frame (%)	TDISP = number of successful responses within 30 mins / all trouble calls received
		Percentage of reconnected electricity customers within time frame (%)	TREST = number of successfully reconnected customers within 24 hours / number of customers assigned without electricity supply
		New services ready for market (%)	MARK = new market ready-services / all R&D projects initiated
		Percentage of resolved cyber- incidents within time frame (%)	CYINC = successfully mitigated cyber-incidents within 12 hours / all reported
	Personal	Self-transcendence	Publicly available pieces of information for psychological profiling: text analysis [39], demographic features [41], item ownership and habits [40].
		Openness to change	
		Conservation	
		Hedonism	
		Self-enhancement	

8. Weighing of utility factors Table 5 presents each utility factor's contribution to the person's overall utility. For the purpose of demonstration, the CEO's overall utility is entirely composed of work-related utility factors. Employees on the other hand, derive utility from other factors which are not directly linked to their professional role (i.e., human values). Work-life balance is represented by the global ratio between work-related and personal utility factors. Weights (w) of the personal utility factors capture the relative importance of basic human values for the subject. Thus, weights are inferred from psychological profiles based on various publicly available pieces of information (e.g., demographics [41], texts produced by the subject [39], evidence of past choices reflecting value trade-offs, habits [40]). Various metrics have been used for quantifying the accuracy/uncertainty of the inferred profiles: R^2 - coefficient of determination (range: 0.19-0.39), PI - prediction interval (Mean: 0.077, SD: 0.794), Pearson correlation coefficients between predicted and ground-truth scores (range: 0.34-0.52) [40]. All the weights sum to 1 for each stakeholder.

Table 5. Weighing of utility factors.

CEO	w	Sigurd	w	Emma	w	Hanne	w	Henry	w
Revenue	0.300	Percentage of successfully located faults and dispatched repair teams within time frame (%)	0.25	Percentage of reconnected electricity customers within time frame (%)	0.30	New services ready for market (%)	0.35	Percentage of resolved cyber-incidents within time frame (%)	0.40
Customer's data privacy (%)	0.175	Self-transcendence	0.18	Self-transcendence	0.12	Self-transcendence	0.10	Self-transcendence	0.11
Contribution to public welfare (%)	0.175	Openness to change	0.14	Openness to change	0.20	Openness to change	0.20	Openness to change	0.10
Innovation (%)	0.175	Conservation	0.17	Conservation	0.09	Conservation	0.05	Conservation	0.18
Relationship with regulators (%)	0.175	Hedonism	0.16	Hedonism	0.12	Hedonism	0.16	Hedonism	0.06
		Self-enhancement	0.10	Self-enhancement	0.17	Self-enhancement	0.14	Self-enhancement	0.15

9. Determination of each strategy's impact on the utility factors Each strategy owner's decision-making process is modeled in Table 6 with the decisions' impact on the risk owner's utility factors. For simplicity each strategy's influence is limited to a maximum of two utility factors. Real-world choices are determined by the complex trade-offs between utility factors as perceived by the stakeholders in a choice situation (i.e., dilemma). Personal features (represented by the weights of each utility factor) interact with salient features of the immediate situation (i.e., initial and final values-capturing states as opposed to traits). Decisions are motivated/demotivated by the overall gains/losses expected from the execution of a strategy. The decision-making process is modeled as $C = f(P \times S)$, where C is a choice, P refers to personal features and S captures situational features. The formula may include the accuracies with which an analyst can assess the relevant person-situation interactions. The results of the context establishment are depicted on the SGAM-H in Figure 4.

Table 6. Impact of the strategies on utility factors.

				Final values after strategy execution			
				A	B	C	D
	Utility factors	**Weights**	**Initial Value**	Help a friend (S_1)	Fix street lights (S_2)	Recruit research applicants (S_3)	Support system integration (S_4)
CEO	Revenue	0.3	50	50	48	53	55
	Customer's data privacy (%)	0.175	50	15	50	50	50
	Contribution to public welfare (%)	0.175	50	50	60	50	50
	Innovation (%)	0.175	50	50	50	65	50
	Relationship with regulators (%)	0.175	50	50	50	50	90
Sigurd	Percentage of successfully located faults and dispatched repair teams within time frame (%)	0.25	50	50			
	Self-transcendence	0.18	20	90			
	Openness to change	0.14	50	50			
	Conservation	0.17	50	50			
	Hedonism	0.16	50	50			
	Self-enhancement	0.1	50	50			
Emma	Percentage of reconnected customers within time frame (%)	0.3	90		30		
	Self-transcendence	0.12	50		50		
	Openness to change	0.2	50		50		
	Conservation	0.09	50		50		
	Hedonism	0.12	50		50		
	Self-enhancement	0.17	50		50		
Hanne	New services ready for market (%)	0.35	50			10	
	Self-transcendence	0.1	50			50	
	Openness to change	0.2	50			50	
	Conservation	0.05	50			50	
	Hedonism	0.16	50			20	
	Self-enhancement	0.14	50			50	
Henry	Percentage of resolved cyber-incidents within time frame (%)	0.4	60				30
	Self-transcendence	0.11	50				50
	Openness to change	0.1	50				50
	Conservation	0.18	50				40
	Hedonism	0.06	50				50
	Self-enhancement	0.15	50				50

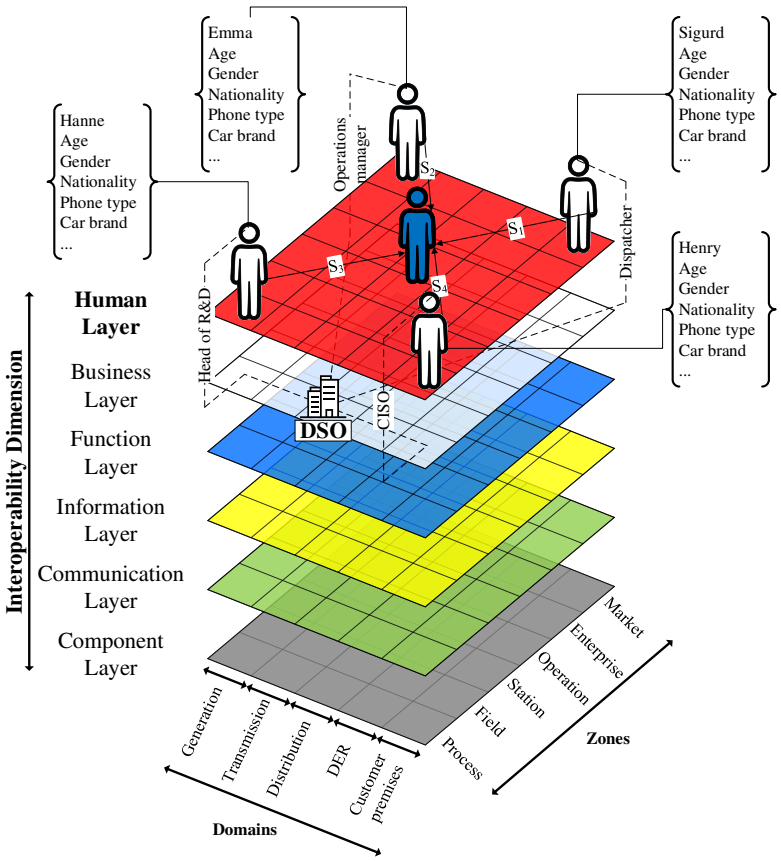

Fig. 4. Summary of context establishment on the SGAM-H.

10. Utility estimation Each stakeholder's overall utility is calculated in Table 7 before and after strategy execution. The weighted sum of each utility factor produces the overall utilities according to the Multi Attribute Utility Theory used in CIRA [32].

Table 7. Utility estimation.

	Utility				
	Initial	**Final**			
Stakeholders		Help a friend (S_1)	Fix street lights (S_2)	Recruit research applicants (S_3)	Support system integration (S_4)
CEO	50	43.875	51.15	53.525	58.5
Sigurd	44.6	57.2			
Emma	62		44		
Hanne	50			31.2	
Henry	54				40.2

11. Calculation of incentives Differences in terms of the overall utilities before and after strategy execution are presented in Table 8. Stakeholders prefer options that increase their utility to options that decrease it, therefore options with positive contribution are selected, whereas options which provide disutility are avoided.

Table 8. Change in utilities.

Stakeholders	**Change in utilities (incentives)**			
	Help a friend (S_1)	Fix street lights (S_2)	Recruit research applicants (S_3)	Support system integration (S_4)
CEO	-6.125	1.15	3.525	8.5
Sigurd	12.6			
Emma		-18		
Hanne			-18.8	
Henry				-13.8

12. Determination of risks Risks are expressed and presented to the CEO as incentive-consequence (I-C) pairs in Table 9. Incentives represent the strength of motivation for each strategy owner to select/avoid the related option, consequences capture the risk to the risk owner. Risks that are characterized by a positive incentive and a negative consequence are threat risks. Negative incentive and positive consequence pairs represent opportunity risks, which would be desirable for the risk owner but the strategy owner would have to take a loss to provide the benefit. The assessed risks are shown on the Human Layer in Figure 5.

Table 9. Risks experienced by the CEO.

Strategy	**Incentive**	**Consequence**
Help a friend (S_1)	12.6	-6.125
Fix street lights (S_2)	-18	1.15
Recruit research applicants (S_3)	-18.8	3.525
Support system integration (S_4)	-13.8	8.5

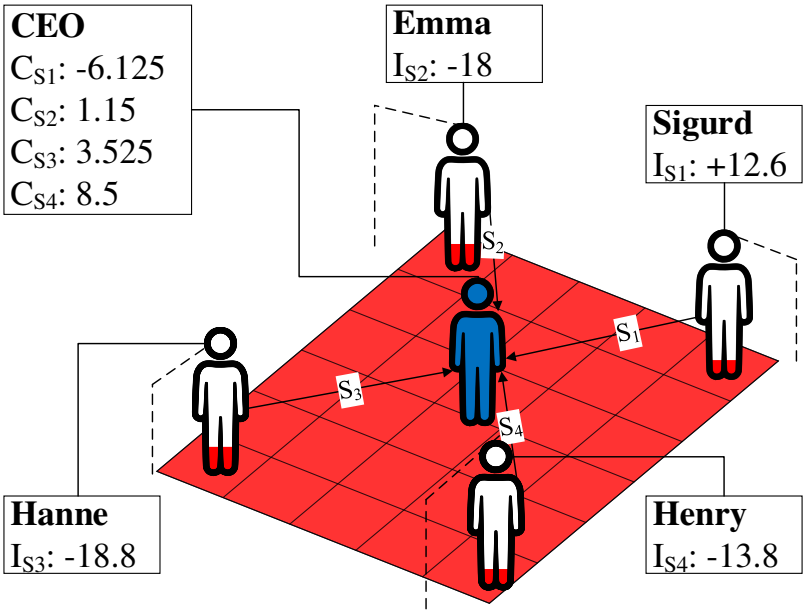

Fig. 5. Risk representation on the Human Layer.

13. Risk evaluation The CEO has to subjectively evaluate whether the risks are above or below the acceptability threshold. Risk that are below the acceptance level may not require further action and may only be monitored (e.g., fixing the street lights, recruit students). Risks that are above the threshold require risk treatment. It should be noted that this demonstration relies on crisp numbers, which do not capture appropriately the accuracies/uncertainties associated with each measurement along the chain of inference. Thus, to draw a more accurate picture for real-world applications it is important to understand how errors propagate. According to [42] the error in a quantity which is derived from other quantities (each measured with some uncertainty) is calculated as:

$$\begin{aligned}
(\text{Measured value of})\, x &= x_{\text{best}} \pm \delta x, \\
x_{\text{best}} &= \text{best estimate for}\, x, \\
\delta x &= \text{uncertainty or error in measurement}, \\
\frac{\delta x}{x_{\text{best}}} &= \text{fractional uncertainty}.
\end{aligned}$$

Since C (choice) is calculated as the product of P and S, the relative error of C can be calculated as the sum of fractional uncertainties in quadrature assuming independent random errors as follows:

$$\frac{\delta C}{C} = \sqrt{\left(\frac{\delta P}{P}\right)^2 + \left(\frac{\delta S}{S}\right)^2}$$

The resulting relative error can be converted into absolute error, and used to compute $C \pm \delta C$ which more accurately captures it's uncertainty.

14. Risk treatment Strategy 1 and 4, are above the risk acceptance threshold, therefore certain incentive modifications are necessary to make the options more (for opportunity risks) or less (for threat risks) desirable for the strategy owners. A risk mitigation for S_1 would be to increase personal accountability in case of privacy violations to make the option less desirable for the strategy owner. Mitigation of S_4 involves the adjustment of the relevant KPI which focuses exclusively on cyber-incident response times by the inclusion of a cross-departmental rating system linked to bonuses which measures co-operation between departments. This can provide incentives to seek mutually beneficial outcomes. The need for alignment between departments requires novel metrics both at the micro and macro levels within the organization.

4.2 Evaluation of the Human Layer

The artifact is qualitatively evaluated across the following criteria by its developers (i.e., internal evaluation by two people): efficacy, ease of use, completeness and homomorphism adhering to the definitions in [31]. A five point grading scale (5-excellent, 4-good, 3-satisfactory, 2-sufficient, 1-unsatisfactory) is used for describing the extent to which the artifact fulfills the evaluation criteria. Efficacy is rated 5 since it successfully establishes a connection between SGAM and CIRA by representing human stakeholder models, thus addressing the identified gap in the literature. Ease of use is rated 3, since the development and construction of the models from scratch required significant effort initially in terms of time spent (several days). After the basic models have been established and with subsequent reuse of the artifacts (i.e., iterative adjustments and updates applied to the models as the case study was developing which involved the identification of relevant literature, extraction of key concepts and customization of the metrics, etc.) it was possible to reduce the effort significantly (below 1 hour for each iteration). Completeness is rated 5 since it captures all the relevant elements and relationships between elements identified in CIRA. Homomorphism refers to the correspondence with a reference model (i.e., original SGAM) and is rated 4 since the extension does not interfere with the original model's structure but further adjustments may be necessary to ensure full, unambiguous compatibility with SGAM objects.

5 Discussion

Critical infrastructures designed and built in the previous century are becoming more autonomous and interconnected by the inclusion of IoT devices. Modernisation is driven by a variety of economical, political and ecological motives. Increasing dependency on ICT gives rise to previously unimaginable risks which may endanger the safety, security and privacy of societies at scale. High levels of complexity and lack of historical data about system behavior represent great practical impediments for traditional risk analysis methods. The CIRA method proposes a solution to these problems by focusing on the behavior of fundamental components of any modern system: key decision-makers.

Human decision-makers are not appropriately represented on the most well-established model of the SG (SGAM) which may lead to under-recognition of people's influence on the SG. Consequently, risk analyses may exclusively focus on technical aspects and miss the point, that technology is under the control of human decision-makers with unique motivations. In order to address this imbalance between perspectives, and to enable the creation of a common understanding about the human aspects, this paper proposed the SGAM-H with the Human Layer on top of the SGAM interoperability layers. The extension aimed at keeping compatibility with the original model to a maximum to increase chances of adoption. The extension's efficacy was demonstrated through a case study which applies the CIRA method to a DSO scenario. The case study was inspired by real-world incidents and presented the application of metrics developed for real-world organizations to ensure its realism. The case study presented one threat risk and three opportunity risks to demonstrate the method's applicability. Since the concept of threat risk is more similar to the traditional concept of risk (i.e., an event with negative consequences), the demonstration served the purpose of providing more details about the concept of opportunity risk which has received relatively less attention previously. The artifact has been evaluated along several criteria, thus completing an iteration within the DSR methodology's design cycle. The evaluation has also uncovered some limitations: lack of formal integration of the decision-maker models (and attributes) into existing SGAM models using the Unified Modeling Language (UML); the case study used for demonstration is hypothetical, since access to real-world organizations is limited; the internal, qualitative evaluation represents a weak form of evaluation.

6 Conclusions

The key contributions of this work are as follows: proposal of the SGAM-H augmenting the original SGAM with the Human Layer to create a common understanding among stakeholders operating in the SG ecosystem about the importance of focusing on human-related risks, and to improve risk communication when the CIRA method is applied to SG scenarios. Furthermore, the study contributes by presenting a fully worked-out example of CIRA's application, which may help students and practitioners in better understanding the method's procedures. Recent developments regarding CIRA have been incorporated into the case study (e.g., use of BSC method, operationalization of motivational profiles, differentiation between various aspects of utility, propagation of errors, risk treatment options) and the artifact is evaluated to identify its strengths and weaknesses.

7 Further work

This study focuses on intra-organizational risks where the CEO is assumed to have the capability to mitigate the identified risks. However, the connection with the other SGAM-layers ensures that relevant stakeholders can be identified from any layer. Stakeholders from other organizations could be identified and elevated from the business layer to analyze inter-organizational risks. Owners of information or physical assets could be identified and elevated to the Human Layer, where the existing connections

between assets are inherited by the stakeholders, enabling the identification and specification of strategies that are at the disposal of the strategy owners. This procedure could be a significant step towards replacing the analyst's intuition for strategy identification (step 3). Development of new tools would be required to increase the usability of the Human Layer (e.g., inclusion of interactive functionality would improve user-experience and risk communication capabilities). Furthermore, scalability could be improved by additional software support to enable the representation of more stakeholders on the Human Layer. Simulation-based analyses could be conducted by a more completely populated SGAM model in which the effects of strategic decisions could propagate through the system to simulate and analyze the reactions of other entities (e.g., customers, competitors). Finally, the evaluation can be improved by using more rigorous quantitative evaluation methods, independent of the developers of the artifact (external evaluation). Field experiments with practitioners, or students require the creation of training materials, while application to real-world cases and expert evaluations can be useful to assess user acceptance. It should be investigated how the general idea of a Human Layer can be applied to other domains (e.g., e-health, transportation domains, etc.) to improve understanding about deliberate human behavior and information security risks.

Acknowledgements

We would like to thank the four anonymous reviewers whose comments helped to improve the quality of the paper.

References

1. Anderson, R., Moore, T.: Information security: where computer science, economics and psychology meet. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **367**(1898), 2717–2727 (2009)
2. Behr, M., Arnoux, P.J., Serre, T., Bidal, S., Kang, H., Thollon, L., Cavallero, C., Kayvantash, K., Brunet, C.: A human model for road safety: from geometrical acquisition to model validation with radioss. Computer Methods in Biomechanics & Biomedical Engineering **6**(4), 263–273 (2003)
3. Capin, T.K., Noser, H., Thalmann, D., Pandzic, I.S., Thalmann, N.M.: Virtual human representation and communication in vlnet. IEEE Computer Graphics and Applications **17**(2), 42–53 (1997)
4. CEN-CENELEC-ETSI Smart Grid Coordination Group: Smart grid reference architecture (2012)
5. Cooper, A.: The inmates are running the asylum. Macmillan (1996)
6. Delgado, I., Aguado, I.: Report on common KPIs D1.4 r2. Project Demonstration 646531, The UPGRID Consortium, Brussels (2016), http://upgrid.eu/wp-content/uploads/2018/01/151104_UPGRID_WP1_D14_KPIs_v14_final.pdf, [Online; accessed 15. Apr. 2020]
7. Devine, W.D.: From shafts to wires: Historical perspective on electrification. The Journal of Economic History **43**(2), 347–372 (1983)

8. Dragomir, D., Nölle, C., Stomff, S.: Stakeholders' Requirements Analysis Report - D3.1. Project Demonstration 318782, STARGRID project, Brussels (2013), http://stargrid.eu/downloads/2014/07/STARGRID_Stakeholders-Report_D3.1_v1.0_2013_10_11.pdf, [Online; accessed 15. Apr. 2020]
9. European Parliament, Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council (GDPR). Official Journal of the European Union (2016), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679#d1e40-1-1, [Online; accessed 15. Apr. 2020]
10. Gagné, M., Deci, E.L.: Self-determination theory and work motivation. Journal of Organizational behavior **26**(4), 331–362 (2005)
11. Gungor, V.C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., Hancke, G.P.: Smart grid technologies: Communication technologies and standards. IEEE transactions on Industrial informatics **7**(4), 529–539 (2011)
12. Harder, W.J.: Key Performance Indicators for Smart Grids. Master's thesis, University of Twente, 7522 Enschede (7 2017)
13. Hevner, A.R.: A three cycle view of design science research. Scandinavian journal of information systems **19**(2), 4 (2007)
14. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. Management Information Systems Quarterly **28**(1), 75–106 (2004)
15. Hubbard, D., Evans, D.: Problems with scoring methods and ordinal scales in risk assessment. IBM Journal of Research and Development **54**(3), 2–1 (2010)
16. Hudlicka, E., Zacharias, G., Psotka, J.: Increasing realism of human agents by modeling individual differences: Methodology, architecture, and testbed. In: Simulating human agents, American association for artificial intelligence fall 2000 symposium series. pp. 53–59 (2000)
17. Jürgensen, J.H., Nordström, L., Hilber, P.: A scorecard approach to track reliability performance of distribution system operators. In: 23rd International Conference on Electricity Distribution-CIRED Lyon, 15-18 June 2015. CIRED-Congrès International des Réseaux Electriques de Distribution (2015)
18. Kaplan, R.S., Norton, D.P.: Putting the balanced scorecard to work. The economic impact of knowledge **27**(4), 315–324 (1998)
19. Lee, R., Assante, M., Conway, T.: Analysis of the cyber attack on the Ukrainian power grid, Defense Use Case. Electricity Information Sharing and Analysis Center (E-ISAC) **388** (2016)
20. Liu, P., Zang, W., Yu, M.: Incentive-based modeling and inference of attacker intent, objectives, and strategies. ACM Transactions on Information and System Security (TISSEC) **8**(1), 78–118 (2005)
21. Martello, M., Watson, J.G., Fischer, M.J.: Implementing a balanced scorecard in a not-for-profit organization. Journal of Business & Economics Research (JBER) **6**(9), 67–80 (2008)
22. McKenna, E., Richardson, I., Thomson, M.: Smart meter data: Balancing consumer privacy concerns with legitimate applications. Energy Policy **41**, 807–814 (2012)
23. Meurens, F., Summerfield, A., Nauwynck, H., Saif, L., Gerdts, V.: The pig: a model for human infectious diseases. Trends in microbiology **20**(1), 50–57 (2012)
24. Moore, T.: The economics of cybersecurity: Principles and policy options. International Journal of Critical Infrastructure Protection **3**(3-4), 103–117 (2010)
25. Muir, A., Lopatto, J.: Final report on the august 14, 2003 blackout in the united states and canada: causes and recommendations. US–Canada Power System Outage Task Force, Canada (2004)
26. Müller-Hansen, F., Schlüter, M., Mäs, M., Donges, J.F., Kolb, J.J., Thonicke, K., Heitzig, J.: Towards representing human behavior and decision making in earth system models-an overview of techniques and approaches. Earth System Dynamics **8** (2017)

27. Musharraf, M., Khan, F., Veitch, B.: Validating human behavior representation model of general personnel during offshore emergency situations. Fire technology **55**(2), 643–665 (2019)
28. National Research Council: Electricity in Economic Growth. The National Academies Press, Washington, DC (1986). https://doi.org/10.17226/900, https://www.nap.edu/catalog/900/electricity-in-economic-growth
29. NVE: KILE – kvalitetsjusterte inntektsrammer ved ikke levert energi (10 2019), https://www.nve.no/reguleringsmyndigheten/okonomisk-regulering-av-nettselskap/om-den-okonomiske-reguleringen/kile-kvalitetsjusterte-inntektsrammer-ved-ikke-levert-energi/, [Online; accessed 15. Apr. 2020]
30. Pew, R.W., Mavor, A.S. (eds.): Representing Human Behavior in Military Simulations: Interim Report. The National Academies Press, Washington, DC (1997). https://doi.org/10.17226/5714, https://www.nap.edu/catalog/5714/representing-human-behavior-in-military-simulations-interim-report
31. Prat, N., Comyn-Wattiau, I., Akoka, J.: A taxonomy of evaluation methods for information systems artifacts. Journal of Management Information Systems **32**(3), 229–267 (2015)
32. Rajbhandari, L., Snekkenes, E.: Using the conflicting incentives risk analysis method. In: IFIP International Information Security Conference. pp. 315–329. Springer (2013)
33. Sánchez-Ortiz, J., García-Valderrama, T., Rodríguez-Cornejo, V.: Towards a balanced scorecard in regulated companies: a study of the spanish electricity sector. The Electricity Journal **29**(9), 36–43 (2016)
34. Santodomingo, R., Uslar, M., Gottschlak, M., Goering, A., Nordstrom, L., Valdenmaiier, G.: The discern tool support for knowledge sharing in large smart grid projects. CIRED Workshop (2016)
35. Schuh, G., Fluhr, J., Birkmeier, M., Sund, M.: Information system architecture for the interaction of electric vehicles with the power grid. In: 2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC). pp. 821–825. IEEE (2013)
36. Selyukh, A.: NSA staff used spy tools on spouses, ex-lovers: watchdog. U.S (Sep 2013), https://www.reuters.com/article/us-usa-surveil-lance-watchdog/nsa-staff-used-spy-tools-on-spouses-ex-lovers-watchdog-idUSBRE98Q14G20130927
37. Snekkenes, E.: Position paper: Privacy risk analysis is about understanding conflicting incentives. In: IFIP Working Conference on Policies and Research in Identity Management. pp. 100–103. Springer (2013)
38. Stassen, H.G., Johannsen, G., Moray, N.: Internal representation, internal model, human performance model and mental workload. Automatica **26**(4), 811–820 (1988)
39. Szekeres, A., Snekkenes, E.A.: Predicting ceo misbehavior from observables: Comparative evaluation of two major personality models. In: E-Business and Telecommunications. ICETE 2018. Communications in Computer and Information Science, vol. 1118, pp. 135–158. Springer, Cham (2019)
40. Szekeres, A., Snekkenes, E.A.: Construction of human motivational profiles by observation for risk analysis. IEEE Access **8**, 45096–45107 (2020)
41. Szekeres, A., Wasnik, P.S., Snekkenes, E.A.: Using demographic features for the prediction of basic human values underlying stakeholder motivation. In: Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 2: ICEIS,. pp. 377–389. INSTICC, SciTePress (2019)
42. Taylor, J.R.: An introduction to error analysis: The study of uncertainties in physical measurements. University Science Books, Sausalito, California (1997)
43. Uslar, M., Engel, D.: Towards generic domain reference designation: How to learn from smart grid interoperability. DA-Ch Energieinformatik **1**, 1–6 (2015)

44. Uslar, M., Rohjans, S., Neureiter, C., Pröstl Andrén, F., Velasquez, J., Steinbrink, C., Efthymiou, V., Migliavacca, G., Horsmanheimo, S., Brunner, H., et al.: Applying the smart grid architecture model for designing and validating system-of-systems in the power and energy domain: A european perspective. Energies **12**(2), 258 (2019)
45. Uslar, M., Rosinger, C., Schlegel, S.: Security by Design for the Smart Grid: Combining the SGAM and NISTIR 7628. In: 2014 IEEE 38th International Computer Software and Applications Conference Workshops. pp. 110–115. IEEE (2014)
46. Uslar, M., Trefke, J.: Applying the Smart Grid Architecture Model SGAM to the EV Domain. In: EnviroInfo. pp. 821–826 (2014)
47. Waterman, R.W., Meier, K.J.: Principal-agent models: an expansion? Journal of public administration research and theory **8**(2), 173–202 (1998)
48. Weishäupl, E., Yasasin, E., Schryen, G.: Information security investments: An exploratory multiple case study on decision-making, evaluation and learning. Computers & Security **77**, 807–823 (2018)