A Measure of Systems Self-Awareness

Maria Mikela Chatzimichailidou
Democritus University of Thrace,
Department of Civil Engineering
mikechat@civil.duth.gr

Stefanos Katsavounis
Democritus University of Thrace,
Department of Production and
Management Engineering
skatsav@pme.duth.gr

Ioannis Dokas
Democritus University of Thrace,
Department of Civil Engineering
idokas@civil.duth.gr

ABSTRACT

In order to be proactive to accidents, there is a need to limit systems' threats and vulnerabilities by being able to perceive and comprehend them as early as possible. Under this notion, the concept of 'risk Situation Awareness provision capability' is introduced, indicating that the elements of a system, tangible or not, have an impact on the enhancement or degradation of the awareness, in reference to its threats and vulnerabilities. As a means of measuring this capability, a methodology, based on existing yet not combined methods, i.e. STPA hazard analysis, EWaSAP early warning sign identification approach, and dissimilarity measures, is offered. This paper looks at analogous SA measurement techniques and finally discusses some limitations and future research directions.

Keywords

Situation Awareness, STPA, EWaSAP, Situation Awareness, Dissimilarity Measures.

INTRODUCTION

Complex socio-technical systems consist of many parts, controlled by human or automated agents, located in different hierarchical levels. In many of such systems, safety is a core value and shared responsibility enforcing the agents that control a part of the socio-technical system to be capable of perceiving and comprehending threats and vulnerabilities as early as possible, as well as projecting what they may bring to the system. That is to say, they should bear risk-focused Situation Awareness (SA). Under this narrative, risk SA provision reflects the emergent and inherent, in accordance with the system design and development specifications, capability of each system part to provide its agent with risk SA (Chatzimichailidou et al., 201x). So far, there is not any method for measuring this capability; this calls for 'something' that, based on the system blueprints, will serve as a means of (a) measuring to what extent the system will be able to perceive, comprehend, and project, i.e. be aware of, the presence of its threats and vulnerabilities and (b) estimate to what extent this self-awareness may affect the system's safety.

To bridge this gap, the paper in hand presents a novel methodology, which brings three already existing approaches together, but in a unique setting. If these methods are applied separately, they lose their functionality as parts of this methodology. The methodology is grounded on a comparison between two design versions of a complex socio-technical system that differ in their composition, i.e. the elements and characteristics that affect differently the risk SA provision capability of the two system versions.

THE CONCEPT OF "RISK SA PROVISION"

To exemplify the concept of “risk SA provision”, consider, the workers of an organisation who receive and disseminate alert emails; in case of a broken window, the eyewitness sends to his co-workers the email: “Broken window on the Ground Floor, Building B”. Apparently, he possesses more information about the incident, such as exact location, extent of the area requiring attention etc. Nevertheless, the receivers of the warning have a fractal picture of the situation, owing to limited access to source data. This reveals that the individual risk SA of the transmitter is more precise and elaborate than the awareness of those not being on the spot. Hence, the formation of the risk SA of the receivers incorporates the following: (a) the alerting email from the transmitter, (b) how and what they understand from the message, and (c) how the relevant information is forwarded from one to the other.

The system elements, such as sensors, communication and alert mechanisms, mental and process models used by the agents for the comprehension of a situation, render the risk SA provision capability dynamic, because it can degrade or be enhanced, by including or excluding system elements.

	System Elements				
System version	Alert email sent/received	Photos from the field	CCTV information	Workers visited the field themselves	Phone calls for verification
Ideal	1	1	1	1	1
Real	1	0	0	0	0
‘Pair’	(1,1) No distance	(1,0) distance	(1,0) distance	(1,0) distance	(1,0) distance

Table 1. ‘Email alert’ example: ideal and real system elements

In the email alert example the elements that dynamically affect this capability would have been: (1) the alert email, (2) photos taken from the field, (3) visual inspection using CCTV system and (4) human senses, e.g. direct view of the field, (5) phone calls between co-workers for verification. In essence, this would have been the ‘ideal’[1] system composition. But in reality, the only available element is the email. In respect to this, Table 1 depicts the ideal and real system versions. When a ‘0’ bit is detected, it signifies the state of ‘not-being-present’ in the real system. This will probably impact on the degradation of the risk SA provision capability, therefore a ‘0’ bit is recognised as an abnormal condition compared to ‘1s’, which are already existing elements.

RELATED WORK

The literature presents several SA measurement techniques mainly based either on individual SA models (e.g. Endsley, 1995b) or on team SA models (e.g. Salas et al., 1995). Salmon et al. (2009) categorises thirty measurement techniques into six individual SA categories: (1) freeze probe techniques, (2) real-time probe techniques, (3) self-rating techniques, (4) observer rating techniques, (5) performance measures, (6) process indices, and three for team-shared SA: (1) team probe-recall techniques, (2) observer rating team SA, and (3) team task performance-based.

Related research works (e.g. Salmon et al., 2009) identify most of the profound defects of the aforementioned techniques, e.g. time-consuming processes, training presupposed, resources required etc. But Chatzimichailidou et al. (2015) detect and group some of the deeper problems that underlie the lack of proper SA measurement techniques in a complex socio-technical context. These are:

(1) Unclear context and definition of system boundaries: The majority of the existing SA measurement techniques were originally developed to measure individual SA in the field of aviation, meaning that they were based on assumptions made in the context of cockpits and air traffic control systems, limiting thus their generality.

(2) SA models depict the individual’s in-the-mind process: Some of the most widely known SA models, such as the three level model (Endsley, 1995) and the perceptual cycle (Neisser, 1976), only illustrate what is going on

---

[1] No methodology can perfectly cover all aspects of complex socio-technical systems. It is possible however to approximate its behaviour and components with system theoretic approaches.

within one's own head, being indifferent to interactions with other humans or artifacts. It is inevitable, though, that the environment affects inner operations.

(3) 'Blurred' perception of what is going to be measured: There is a fuzziness in what characteristic and/or behaviour is about to be measured. Even when theory is articulate, the product of the measurement is different from the pursued objective. For example, according to Salmon et al. (2009) there is an arguable relationship between SA and performance because reactions might be biased, deliberate, or 'by-the-book', not necessarily mirroring 'in-the-head' cognitive operations.

(4) Information as the only factor that determines SA levels: There is an indication that individuals who possess much information perform better, and are aware of more elements in their environment, however, in complex collaborative systems, this is by far simplistic. Information requires filtering and processing and clearly it is not the only component that contributes to SA. It triggers awareness, but does not entirely shape it.

(5) Researchers apply SA measurement techniques when the system is already operating: The known measurement techniques are not regarded as precautionary measures for enhancing and preserving the awareness of system's possible future states. Some require the freeze of operations, while others work in real-time, either comprising self- or hetero-measurement.

(6) All SA measurement techniques arrive at qualitative conclusions: This entails that they are subject to subjective collection and interpretation of data and information. For example, rating scales are a numerical interpretation and estimation of qualitative characteristics that contribute to the shaping of SA, but they do not reach a quantitative conclusion.

(7) The means to implement measurement techniques: Questions, such as in questionnaires, rating scales etc., limit the scope of SA and focus the interest on an individual's opinion and point of view. Consequently, systems' technical parts are underestimated and information is lost. Appropriate and widely understandable wording and question formulation are also important in avoiding misunderstandings or divergence from what the question tries to elicit from the respondents. Moreover, techniques where observes 'draw' the picture of the system, judging by what they see other people do, bear the risk of differently understanding the same situation.

The focus is now on those measurement techniques, which exhibit proximity to the rationale of the methodology prescribed herein, on the grounds that they use a point of reference against which things may be compared and/or assessed. To be more specific, SAGAT (Endsley, 1995b) measures the extent to which human operators are aware of pre-defined elements in relatively stable environments, their properties, and what the potential future states might be (Salmon et al., 2009). With SART (Taylor, 1990) and SARS (Waag and Houck, 1994) observers are asked to recognise benchmarking behaviours (to others or themselves) and measure their performance. SABARS (Matthews and Beal, 2002) is developed to assess the SA of infantry personnel based on their responses to various stimuli and having a list of 28 SA related behaviours in hand. In a nutshell, they identify and 'collect' benchmarking behaviours that are supposed to convey a positive SA-related conclusion (Chatzimichailidou et al., 2015).

THE PROPOSED METHODOLOGY

RATIONALE

The email alert example follows the rationale of the proposed methodology; in the ideal case the system should encapsulate all (5) elements (Table 1), but in reality it only encompasses the email service. Apparently, in line with Table 1, there is 'distance' between the two system versions, which is useful and easy to be specified. But when it comes to multifaceted and complex systems, where analysts need to carefully consider and specify what the system should consist of, structured methods are the only way for a guided analysis.

To give flesh to this methodology three phases are necessary: (1) adopt a hazard analysis and an early warning sign identification approach to define safety requirements and sensory services that the system should ideally incorporate, (2) based on the original system design, identify the elements that are present in the real system, and (3) measure the distance between the ideal and the original system design. To accomplice this, the body of the methodology integrates three methods: (1) the STAMP Based Process Analysis (STPA) (Leveson, 2011), (2) the Early Warning Sign Analysis based on the STPA (EWaSAP) approach (Dokas et al., 2013), and (3) a binary dissimilarity measure.

Before elaborating on the methods, two assumptions should be noted:

1. The awareness of threats and vulnerabilities (i.e. risk SA) enhances safety: Stanton et al., 2001; Fioratou et al., 2010; Naderpour et al., 2014; 2014; Chatzimichailidou et al., 201x support a positive correlation between safety and awareness. Namely, the more aware of its threats and vulnerabilities, the less vulnerable the system is.
2. An ideal, in terms of risk SA provision capability, system design could derive from hazard analyses, because they help designers gather essential system elements and characteristics that ideally should be included into the system design, serving to enhance its preparedness against accidents.

METHODS USED

The embodied methods are briefly[2] presented as follows:

1. STPA identifies, right from the early stages of the system design, inadequate control actions within the control loops of the safety control structure (Figure 1) and examines scenarios or paths to accidents, as well as factors poorly handled by traditional hazard analyses, e.g. software requirements errors, inadequate coordination among multiple controllers, flawed management (Leveson, 2014). It does not generate a probability number related to a hazard, otherwise it has to omit important causal factors that are not stochastic or for which probabilistic information does not exist (Leveson, 2011). Safety is, thusly, treated as a dynamic control problem, rather than a component reliability problem.

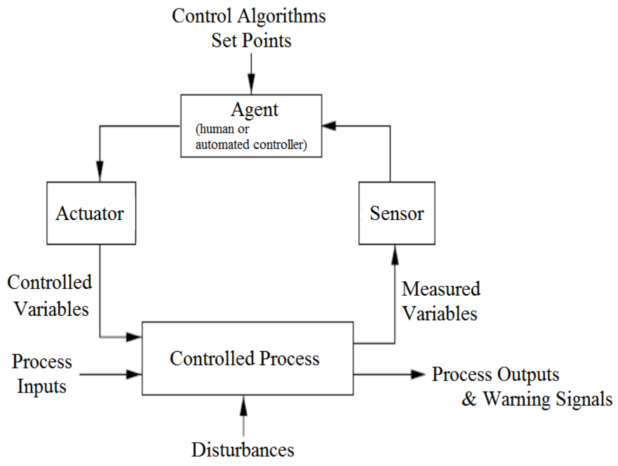

Figure 1. A typical control loop

2. EWaSAP is an add-on to STPA (Dokas et al., 2013) and serves for the identification of early warning signs. It introduces awareness actions that allow controllers to provide warning messages and alerts to other controllers, whenever data indicating the presence of threats or vulnerabilities has been perceived and comprehended.

3. A plethora of distance/dissimilarity measures (Mahmoud et al., 2011; Zhang and Srihary, 2003) detect mismatches between binary data sets. Dissimilarity measures consist of the ‘S00’, ‘S01’, ‘S10’, and ‘S11’ terms, which denote the total number of the corresponding (0,0), (0,1), (1,0), and (1,1) ‘pairs’ of binary integers of the two compared vectors.

Some facts about dissimilarity measures are: (a) The minimum dissimilarity is ‘0’; when dissimilarity tends to ‘1’ then the compared vectors are almost dissimilar. (b) All variables are normalised; [0,1]. (c) Distance is the dual or ‘complementary’ of similarity; d(i,r)=1-s(i,r).

---

[2] Due to space-saving reasons, the following references are suggested for further reading: Leveson, 2011; Dokas et al., 2013; Zhang and Srihary, 2003

According to the rationale, ‘dissimilarity’ measures are more preferable than ‘similarity’ ones, thus the Rogers-Tanimoto dissimilarity measure is selected as the only dissimilarity measure that gives weight to the total number of ‘(1,0)’ pairs; that is, ‘S10’ multiplied by two (Eq.1).

$$Rogers - Tanimoto\ dissimilarity = \frac{2S10+2S01}{S11+S00+2S10+2S01} \qquad (1)$$

THE STEPS OF THE METHODOLOGY

Figure 2 provides an overview of the three main phases that make up the total body of the step-wise methodology.

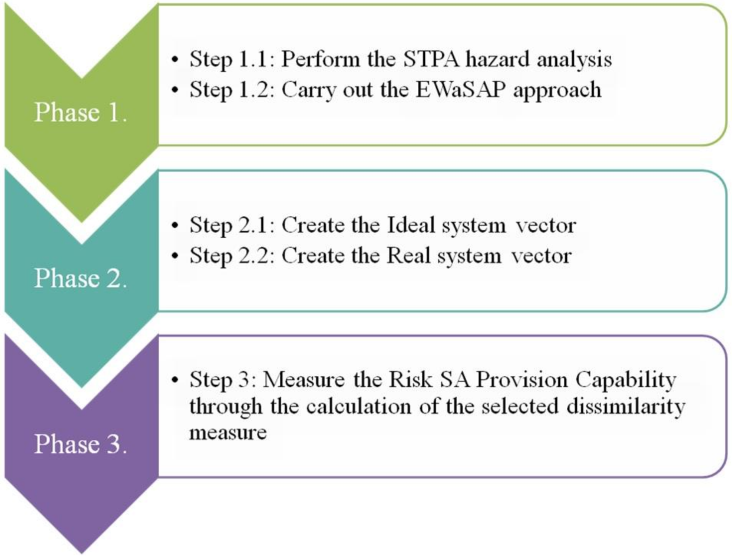

Figure 2. The step-wise methodology

By assuming no limitations on resources, in *Phase 1*, STPA (*Step 1.1*) establishes safety requirements to define the ideal design of the system. Internal sensory services to capture early warning signs are determined by EWaSAP in *Step 1.2*. These safety requirements and sensor characteristics are combined to form mental models integrating ‘if-then’ rules, for both human and automated controllers.

In *Phase 2* one can create the ideal system vector (*Step 2.1*) consisting of qualitative values, i.e. safety requirements and sensory services determined in *Phase 1*. The real system vector is built (*Step 2.2*) by tabulating (a) all elements that exist in the real system, as is, and (b) those from *Phase 1* that may be either present or absent. Then, all elements of both vectors have to be transformed to quantitative ones, i.e. take binary values. These two vectors are the input to the dissimilarity measure.

In *Phase 3*, *Step 3* one has to apply the selected dissimilarity measure, meaning that the two vectors are brought together for comparison in order to measure their distance. To calculate the value of the indicator, the ‘S10’, ‘S01’, ‘S11’, and ‘S00’ terms (Eq.4.1) have to be substituted. The emanating value is a quantitative expression of the system’s inherent risk SA provision capability.

DISCUSSION AND CONCLUDING REMARKS

SA measurement techniques (a) embark on a direct measurement of SA, which is not the question asked here at all, and (b) appreciate a small portion of system elements, mainly human ones. The new methodology departs from the notion that a system consists of fixed elements, it is harmonised, though, with the idea that it is feasible to assess and improve their utility and role in enhancing or causing the degradation of the risk SA provision capability.

As regards the methods adopted, STPA was selected as an orderly analysis of assessing safety risks arising from the integration of the system elements (Leveson, 2011). STPA is reinforced by EWaSAP, which identifies sensors characteristics and defines awareness actions, being responsible for the provision of early warnings and alert messages. Scholars however, are free to apply other hazard analysis and early warning sign identification technique, in *Step 1.1* and *Step 1.2*, respectively. Similarly, in *Step 3*, they can choose a suitable dissimilarity measure, depending on the problem statement.

Experience, qualification, and a multidisciplinary team with shared mental models, are significant prerequisites for applying STPA and EWaSAP and determine 'how much' ideal a system version is. Another limitation is the overabundance of dissimilarity measures. One measure, for instance, may indicate better risk SA provision capability, whilst, for the same system, another measure may give much worse results, practically constituting an interference. Besides, binary-based indicators disregard values between '0' and '1'. On account of this, fuzzy logic along with continuous variables are to be investigated. Another limitation acknowledged, but left to be dealt with in the future, is that the elements of the control loop (Figure 1) are treated as equivalent to the risk SA provision capability enhancement or degradation. Yet scholars may argue that, for them, specific system elements may impact differently on the risk SA provision capability.

To conclude, it seems reasonable to utilise hazard analysis techniques to identify the possible system hazards and be prepared to respond to and recover from them. A benefit of this methodology is its utility as a criterion for system selection, or a decision-making tool over alternative system designs. It could also support decision-making about design improvements, by 'lessening' the dissimilarity between the compared units. Various dissimilarity measures were tested across different case studies (e.g. ACROBOTER, Überlingen mid-air collision[3]) and the corresponding preliminary results uphold the proposal of this research work. As far as future work is concerned, the application of this methodology to various systems will guarantee its soundness, generality, practicality, as well as its usefulness and could be validated by experts who are going to adopt this methodology. Apparently, benefits as well as limitations are expected to emerge from future case studies.

REFERENCES

1. Chatzimichailidou, M. M., Stanton, N. and Dokas, I. (201x). The Concept of Risk Situation Awareness Provision: Towards a New Approach for Assessing the DSA about the Threats and Vulnerabilities of Complex Socio-technical Systems. Safety Science. *(accepted upon revision)*
2. Chatzimichailidou, M. M., Protopapas, A., and Dokas, I. M. (2015). Seven Issues on Distributed Situation Awareness Measurement in Complex Socio-technical Systems. *In Complex Systems Design & Management* (pp. 105-117). Springer International Publishing.
3. Dokas, I. M., Feehan, J., and Imran, S. (2013). EWaSAP: An early warning sign identification approach based on a systemic hazard analysis. *Safety science*, *58*, 11-26.
4. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64.
5. Endsley, M. R. (1995b). Direct measurement of situation awareness in simulations of dynamic systems: Validity and use of SAGAT. *Experimental analysis and measurement of situation awareness*, 107-113.
6. Fioratou, E., Flin, R., Glavin, R., and Patey, R. (2010). Beyond monitoring: distributed situation awareness in anaesthesia. *British journal of anaesthesia*, 105(1), 83-90.
7. Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*. Mit Press.
8. Leveson, N. (2014). A Systems Approach to Risk Management Through Leading Safety Indicators. *Reliability Engineering & System Safety*.
9. Mahmoud, S. M., Lotfi, A., and Langensiepen, C. (2011). Abnormal behaviours identification for an elder's life activities using dissimilarity measurements. In *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments* (p. 25). ACM.
10. Matthews, M. D., and Beal, S. A. (2002). *Assessing situation awareness in field training exercises*. Military Academy West Point NY Office of Military Psychology and Leadership.

---

[3] Due to space-saving reasons, initial conclusions from these two case studies can be presented in the conference.

11. Naderpour, M., Lu, J., and Zhang, G. (2014). A situation risk awareness approach for process systems safety. *Safety Science*, 64, 173-189.
12. Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. WH Freeman/Times Books/Henry Holt & Co.
13. Salas, E., Prince, C., Baker, D. P., and Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 123-136.
14. Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., and Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*,*39*(3), 490-500.
15. Stanton, N. A., Chambers, P. R. G., and Piggott, J. (2001). Situational awareness and safety. *Safety Science*, 39(3), 189-204.
16. Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53)*.
17. Waag, W. L., and Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation, space, and environmental medicine*.
18. Zhang, B., and Srihari, S. N. (2003). Properties of binary vector dissimilarity measures. In *Proc. JCIS Int'l Conf. Computer Vision, Pattern Recognition, and Image Processing*, (Vol. 1).
19.