A novel real-time safety level calculation approach based on STPA

Apostolos Zeleskidis[1,*], *Ioannis M.* Dokas[2], and *Basil* Papadopoulos[3]
[1]Undergraduate Student (Democritus University of Thrace, Civil Engineering, Greece)
[2]Assistant Professor (Democritus University of Thrace, Civil Engineering, Greece)
[3]Professor (Democritus University of Thrace, Civil Engineering, Greece)

Abstract. This paper proposes a novel approach to dynamic safety level calculation for safety-critical systems based on the STAMP accident model and the implementation of a mathematical model. The proposed approach utilises (1) an STPA hazard analysis applied to the system in question, (2) system operational data from domain experts regarding process duration and reaction times, and (3) real-time system data. The STPA analysis is transformed into acyclic diagrams that graphically indicate every possible sequence of safety constraint violations that could lead to a loss (path). Based on this diagram the safety level (SL) of a system is defined as SL=$\vec{p}_w$ where $\vec{p}_w$ is the most detrimental to safety path which is active for any possible time value or context in the system's operation. This approach is also demonstrated using as a case study the "classical" Train Door STPA analysis example. This paper aims to introduce a new perspective on the problem of measuring and managing the actual safety level of highly complex socio-technical systems in real time and discusses related limitations and future research opportunities of this approach.

1 Introduction

Several safety indicators are utilised to monitor safety drifts and to assess whether a system maintains its safety within acceptable levels. As a result, numerous categories of safety indicators have been proposed in the literature such as event indicators, barrier indicators, activity indicators, and programming indicators [1,2]. However, the need to "get smarter at predicting the next accident" [3] and the challenge of measuring what the actual safety level of a system is at a certain moment in time at a given context [4] still remains, regardless of the introduction of new accident models and the view of safety as an emergent system property. To address this problem, Leveson [5] proposed that useful leading indicators can be identified based on the assumptions [6] underlying our safety engineering practices rather than on the likelihood of loss events. Thus, there is a need to monitor whether the generated safety assumptions based on which a system was designed, hold during the phase of operations are vulnerable or changed. Chatzimicailidou et al. [7] suggested to measure the gap between system design and operation as a measure for safety and introduced the

* Corresponding author: aposzele@civil.duth.gr

RiskSOAP indicator. With RiskSOAP, one can compare the safety constraints of existing systems compared to their ideal set of safety constraints that derive from their STPA [8] and EWaSAP [9] analysis, and calculate a situational awareness indicator for the system under study.

The problem is that there are no methodologies paired with a mathematical model that can determine dynamically what the safety level is at any certain moment in time for any given context and how much time is left until an accident takes place. To fill this gap, we propose a novel approach including a mathematical model for systems' safety level determination and its dynamic calculation based on the STAMP accident model. The proposed model utilizes the outcomes of STPA hazard analyses which are transformed into diagrams. In these diagrams, each node depicts an STPA based safety constraint, and each path of the acyclic diagram depicts a possible scenario of safety constraints violations that can lead to accidents.

2 STPA

STPA is a hazard analysis technique based on the STAMP accident model. According to STAMP, safety is an emergent property of systems, and accidents can occur not just because of component failures but also due to unsafe interactions among system components that did not fail [8]. That means that the feedback control loops of the system should be designed, developed and operated in a manner that their controllers will not enforce unsafe control actions in any possible operational context or environment due to lack of awareness of the system state (i.e. mental model of the system). This could be achieved by applying the STPA hazard analysis during the life cycle of the system, ideally as early as possible.

The STPA hazard analysis consists of four steps. In the first step, the purpose of the analysis and the boundaries of the system are defined, together with the losses or accidents (A), the system level hazards (H), and the system level safety constraints. During the second step of STPA, a functional diagram known as the safety control structure diagram is developed, where the controllers and the controlled processes of the system are shown, together with the control actions and the feedback each controller is receiving from the process it controls.

In the third step, the control actions of each controller are analyzed to examine under which contexts they could lead to the identified losses. The so-called Unsafe Control Actions (UCA) are then translated into safety constraints or safety specifications. In addition, all Control Actions (CA) given by the controller are analysed to determine how they could be unable to affect the controlled process as intented while enforced by the controller within the appropriate context. The fourth step aims at identifying the reasons why unsafe control actions might be enforced in the system. As a result, loss scenarios (S) are created to explain how incorrect feedback, inadequate requirements, design errors, component failures and other factors that could cause unsafe or ineffective control actions (CA) which are grouped and analysed fundementaly as UCAs [8].

This analysis incorporates safety barriers and other similar mechanisms as parts of the system and finds the weaknesses in the complete system together with its defenses.

3 Methodology

The complete approach we suggest is depicted in Figure 1. The proposed model is comprised of two main processes. The results of STPA and EWaSAP analyses as well as time ranges taken from domain experts are used for the first step of the methodology which

is the Model Development process of the system under study which leades to the creation of a mathematical model dedicated to the system.

Then, from real time system data and the operational mode of the system the Safety Level Determination process is initiated. Operational modes are a tool used in complicated systems to distinguish between different modes in a systems operation. For example, an airplane during a typical flight goes through 3 different operational modes: take off, landing and cruising. The goal of the Safety Level Determination process is to dynamically calculate the Safety Level of the system in real time

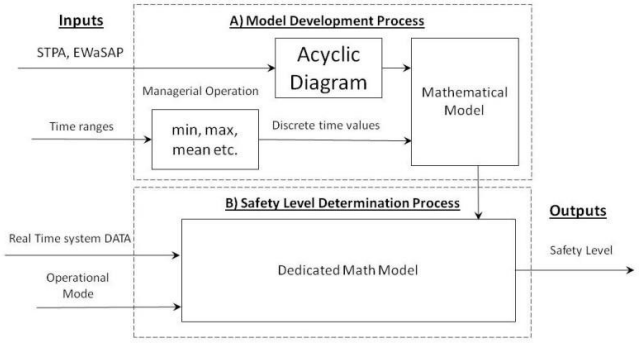

Figure. 1. The proposed approach.

3.1 Model Developments

In more detail, the main goal of the Model Development process is to create and populate an acyclic diagram that represents the system in terms of safety and calibrate the mathematical model to the exact system in study. This is achieved by utilising the following inputs:

1. Safety specifications and constraints obtained from the STPA analysis are turned into nodes of the diagram and are categorised according to their impact on safety. The Categories (levels) are ranked from most to least consequential according to STPA: Level 4 = Accidents, Level 3 = Hazards, Level 2 = Unsafe Control Actions (UCAs) and ineffective Control Actions (CA) and Level 1 = Scenarios. Also, safety constraints are causally linked as shown in the diagram by the connections between the nodes which represent the specifications. These connections are always between nodes of two ascending levels (from level 1 to level 2 or from level 2 to level 3 etc.). See Figure 2 for nodes connections and levels of safety constraints in acyclic diagrams.
2. Time ranges given by domain experts or observed by studying the system regarding the connections of the casually linked safety constraints. These should be distilled in the form of sets (min(t), max(t)) for every connection between two nodes, such as the time it would take from the moment that an Unsafe Control Action is given to the moment that a Hazardous state occurs because of the specific UCA in the system. Each set represents the maximum and minimum amount of time it would take for an active node to activate the node connected with it (Figure 3).
3. The operation used (minimum, maximum, mean average etc.) to discern a single time value $t_{x \to y}$ from the time ranges which represent the flexibility in how the entire operation can be managed. A path is defined as a possible scenario of safety constraints violations that can lead to accidents. It is one of the possible "roads" leading from the lowest level (loss scenarios) to the highest (losses) using the connections between nodes. For instance, the example of Figure 2 has a total of 17 paths and one possible path is S1→ UCA1→ H-2→ A (shown with bold arrows). Each path is comprised of 4 nodes,

one in each level of the tree-like structure. Since a path is comprised of 4 nodes it also has 3 corresponding time values (t1→2, t2→3, t3→4). Paths have two main characteristics: i) Path completeness y which expresses which Level of the acyclic diagram is the highest active node of the path in time t. Path completeness takes values in the discrete set {1,2,3,4}, ii) Time remaining until accident $t+(y)$ which is calculated according to the path completeness and the singular time values assigned to each path and expresses the time which is left for a path to reach the node that represents the occurrence of an accident. An example of path completeness and time remaining until accident is given in Figure 3.

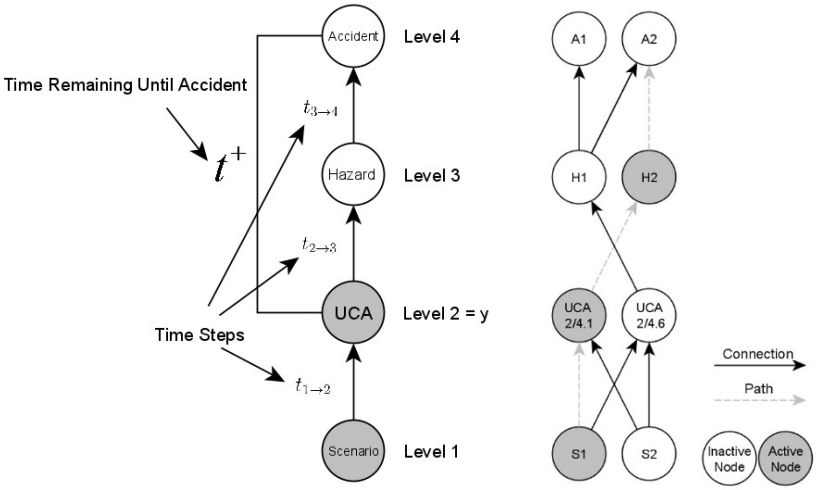

Figure. 2. (Right) Example of an acyclic diagram and representations of active and inactive nodes, paths, and node connections.

Figure. 3. (Left) Example path, completeness levels and the corresponding time values.

3.2 Safety Level Determination

The goal of the Safety Level Determination process is the actual real-time safety monitoring of the system during its operation. The data needed are:

1. The Operational Mode of the system. For instance, referring to a plane system during its take-off mode, out of the total set of the acyclic diagram paths that can be produced from the plane system, the paths which will be computed during the Operational System process of the methodology will be only those belonging to the take-off mode of plane operations.
2. System specific data according to the STPA analysis. This data must be able to show if any node or safety constraint of the diagram has activated or not for any time *t*. Activated means the safety constraint was violated in the specific time value monitored.

Using the mathematical model and the active nodes, the most detrimental to safety system path is calculated.

3.3 The mathematical model

The mathematical model defines the safety level of the system in a specific time t as $SL = \vec{p}_w(t) = \max(\vec{\wp})$, where $\vec{\wp}$ is the ordered set of all paths of the system derived by STPA, $\vec{p}_w(t)$ indicates the path or set of paths that for time t are considered the "worst" among all the paths of the diagram concerning their completeness level (i.e. how "close" to the loss node is the last active node of this path), and the time left until the accident (i.e. the time until the loss node is activated meaning an accident has taken place in the system) graphically represented in Figure 3.
Mathematically paths are defined as follows:
For path $\vec{p} = (y, t^+_{\vec{p}}(y)) \in Y \times \mathbb{R}^+$ where:
y the path completeness for p in time t, y ∈ {0,1,2,3,4}
$t^+_{\vec{p}}(y)$:{0,1,2,3,4}→*[0,∞)* is the time remaining until the accident of path p is activated and is calculated according to the following function:

$$t^+_{\vec{p}}(y) = \begin{cases} t^{\vec{p}}_{1\to2} + t^{\vec{p}}_{2\to3} + t^{\vec{p}}_{3\to4} & y \in \{0,1\} \\ t^{\vec{p}}_{2\to3} + t^{\vec{p}}_{3\to4} & y = 2 \\ t^{\vec{p}}_{3\to4} & y = 3 \\ 0 & y = 4 \end{cases} \quad (1)$$

We consider the set of all the paths $\vec{\wp} = \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n\}$ where n is the number of paths of the complete system.
In the set $\vec{\wp}$ we define the following ordering relation, according to which $\max(\vec{\wp})$ is determined:

$$\forall \vec{p}_1 \left(y_{\vec{p}_1}, t^+_{\vec{p}_1}\left(y_{\vec{p}_1}\right)\right), \vec{p}_2 \left(y_{\vec{p}_2}, t^+_{\vec{p}_2}\left(y_{\vec{p}_2}\right)\right) \in \vec{\wp}$$
$$\vec{p}_1 \geq \vec{p}_2 \Leftrightarrow t^+_{\vec{p}_1} < t^+_{\vec{p}_2} \vee \left(t^+_{\vec{p}_1} = t^+_{\vec{p}_2} \wedge y_{\vec{p}_1} \geq y_{\vec{p}_2}\right) \quad (2)$$

4 Case study

The fictitious system of the train door STPA analysis example [10] will be used to demonstrate our approach based on a scenario of operation to demonstrate the methodology's real-time capabilities. The train door system monitors and controls the process of the typical train door. This process concerns the safe boarding, departing and travel of train passengers. The system is comprised of a door controller, the door actuator, various sensors and the physical door itself. Figure 4 depicts the safety control structure of the system. The STPA analysis applied to this system identified 91 safety requirements in total. After removing the repeated specifications which were produced during the last step of the STPA analysis (i.e., from the translation of the unsafe control actions scenarios) the analysis defined the following: 1 accident, 3 hazards, 14 unsafe control actions, 4 improperly executed control actions and 36 safety scenarios (sample in Table 1). This system is used purely for demonstrative purposes and the complete analysis is not of much importance other than the example.

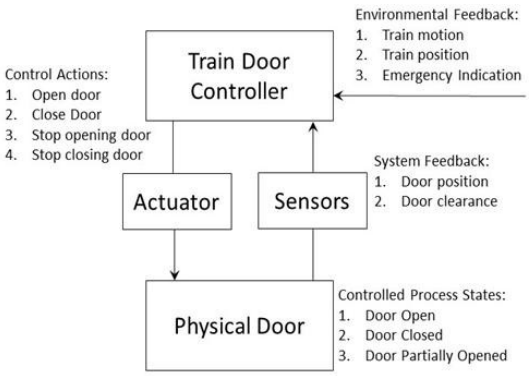

Figure. 4. Train Door System Safety Control Structure

Table 1. A subset of STPA Produced Safety Specifications of the Train Door System.

#	Symbols	Description	Connections
1	A	Passenger loss of life or injury.	-
2	H-1	The door is open when the train is moving fast.	A
3	H-2	A person is unable to leave the train (by the door) in case of emergency.	A
4	H-3	A passenger is caught by the door and the train is moving.	A
5	UCA5	Train Door Controller provides open door while the train is in motion.	H-1
6	CA3	Train Door Controller is unable to provide close door while the train is moving, and the door isn't closed.	H-1
7	S05	Small objects stuck in door movement rail.	CA1, CA3
8	S24	Evacuation mode misfire.	UCA5

4.1 Results

The scenario unfolds as follows. The train is stopped at the station and the door rail is filled with dirt, resulting in the door being unable to close. Then the train starts for the next station but while the command to close the door is issued by the door controller the actuator is unable to execute it, leaving the door open while the train is moving. A passenger then sees the door open and the train moving and scared presses the emergency button. This causes the system to change contexts of operation fundamentally changing the safety priorities of the system.

4.1.1 Model Development process

1. Creation of acyclic diagram. The safety constraints from the STPA analysis were translated into an acyclic diagram (Figure 5). The diagram is comprised of 143 distinct paths. In our example, only 2 of these paths are studied (Figure 6).

2. Time ranges. The time values for the connections of the safety constraints were arbitrarily assigned (sample in Table 2).
3. Selection of operation. Because in this system small changes can lead to big variations in the safety level the MINIMUM operation was used to discern the single time values (Table 2).

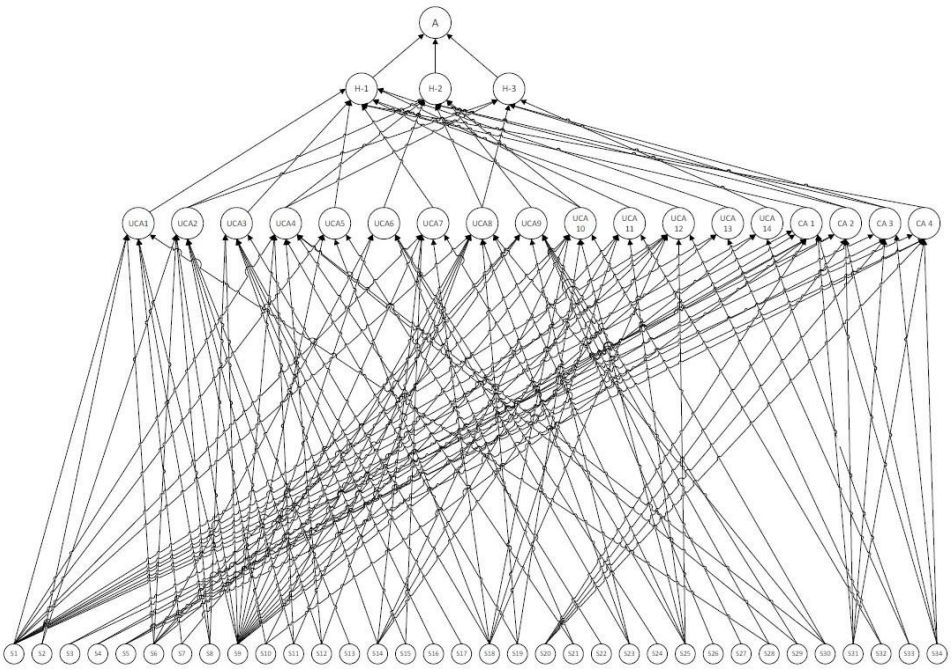

Figure 5. The complete acyclic diagram of the train door system.

Table 2. Time ranges and single time values.

Connection	Time range	Time value with minimum operation
H-1→A	[60,120]	MIN[60,120] = 60
CA3→H-1	[2,6]	MIN[2,6] = 2
UCA5→H-1	[10,30]	MIN[10,30] = 10
S05→CA3	[90,300]	MIN[90,300] = 90
S24→UCA5	[8,20]	MIN[8,20] = 8

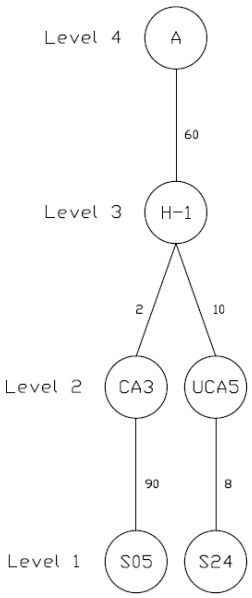

Figure 6. The acyclic diagram for the train door case study.

4.1.2 Safety Level Determination

1. Operational modes. The system is divided into two operational modes: the train stopped at a station, the train travelling between stations, but for the sake of simplicity in the example takes place only when the Travelling between stations Operational mode is active.
2. System monitoring data. The system specific data needed for the nodes of Figure 6 are: the door's state (open, closed), the train's velocity, doorway clearance, evacuation state, if the door movement rail is filled with small objects and evacuation mode misfire.
3. Mathematical model. In Figure 6 the sample of the acyclic diagram that will be used for the demonstration of the suggested approach is shown. This diagram contains 2 distinct system paths ($\vec{p}_1$: S05→CA3→H-1→A and $\vec{p}_2$: S24→UCA5→H-1→A). According to equations (3), (4) the values for the time remaining until accident for these two paths are the following:

$$t^{+}_{\vec{p}_1}(y) = \begin{cases} t^{\vec{p}_1}_{1\to2} + t^{\vec{p}_1}_{2\to3} + t^{\vec{p}_1}_{3\to4} = 90 + 2 + 60 = 152 & y \in \{0,1\} \\ t^{\vec{p}_1}_{2\to3} + t^{\vec{p}_1}_{3\to4} = 2 + 60 = 62 & y = 2 \\ t^{\vec{p}_1}_{3\to4} = 60 & y = 3 \\ 0 & y = 4 \end{cases} \tag{3}$$

$$t^{+}_{\vec{p}_2}(y) = \begin{cases} t^{\vec{p}_2}_{1\to2} + t^{\vec{p}_2}_{2\to3} + t^{\vec{p}_2}_{3\to4} = 8 + 10 + 60 = 78 & y \in \{0,1\} \\ t^{\vec{p}_2}_{2\to3} + t^{\vec{p}_2}_{3\to4} = 10 + 60 = 70 & y = 2 \\ t^{\vec{p}_2}_{3\to4} = 60 & y = 3 \\ 0 & y = 4 \end{cases} \tag{4}$$

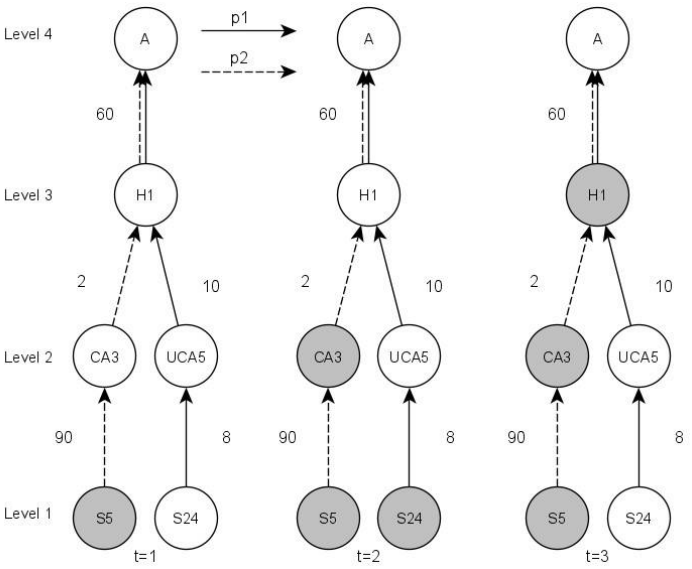

Figure 7. Example scenario of operation with acyclic diagrams for every time moment.

As Figure 7 shows, in the first time moment S05 (Table 1) is active this activates $\vec{\mathbf{p}}_1$ with values $\vec{\mathbf{p}}_1$= (1,152) and $\vec{\mathbf{p}}_2$= (0,78), so $\vec{\mathbf{p}}_w{}^{t=1}$=$\vec{\mathbf{p}}_2$ =(0,78), with the context of the system being that the train is stationary on the platform. In the second time moment the context changes to the train starting moving and gathering speed, with this change in context (that a computerized system can monitor with various sensors) the active nodes become S05 and CA3, S24 is active because a passenger has pressed the emergency button due to seeing the open door, these in turn activate $\vec{\mathbf{p}}_2$ with value $\vec{\mathbf{p}}_2$ =(1,78) and change $\vec{\mathbf{p}}_1$=(2,62), so $\vec{\mathbf{p}}_w{}^{t=2}$=$\vec{\mathbf{p}}_1$=(2,62). In the final time moment the train reaches cruising speed making H-1 active and in a few moments will start decelerating because of the emergency button being pressed this results in the two paths having the same value $\vec{\mathbf{p}}_1$= $\vec{\mathbf{p}}_2$ =(3,60), so $\vec{\mathbf{p}}_w{}^{t=3}$ = $\vec{\mathbf{p}}_1$= $\vec{\mathbf{p}}_2$ =(3,60).

5 Future Work and Concluding Remarks

By using this methodology, it is possible to identify the worst path to violate safety of a system with many possible paths that can lead to accidents. This can be useful to help safety managers and staff understand how system information and changes affects, safety and enable operators to react in timely manner to neutralize the completion of the path and hence the occurrence of an accident. The aim of this new approach is to enhance the capabilities of safety management systems towards preventing accidents and any types of losses by calculating in real time the safety levels of safety critical processes. This is feasible by projecting enriched information in a simple to understand form (second or minutes, instead of safety indexes) in real time as well as the acyclic diagrams, which can help managers understand how certain system interactions effect safety in their systems. This can be seen in the case study by observing that initially $\vec{\mathbf{p}}_1$ seems more severe, if CA3 occurs the system gets much closer to the accident state from $\vec{\mathbf{p}}_1$, and after the system reaches the point of Hazards the priority should be to stop the train and not meddle with the two different control systems S5→CA3, S24→UCA5.

It is the first time where a mathematical model is presented to address the problem of determining dynamically what the safety level is at a certain moment in time at a given context and calculate how much time is left for an accident to happen. This approach is also based on the results of an STPA analysis. This is another novel aspect of this paper since there are no previous concepts on real time safety level calculation based on the STAMP accident model.

A prominent limitation however of this novel approach is the fact that the mathematical model it is based on does not cope with uncertainty. It is planned, however, to address this problem in future work by enhancing the mathematical model with the use of fuzzy sets theory. An extension of the acyclic diagrams is also planned to take place in future work such that their acyclic nature would be nullified when system defences are in place. Finally, the creation of an operational system is planned where the approach will be installed into a real system to calculate its safety level in real time.

References

1. Øien K., (2010). Remote Operation in Environmentally Sensitive Areas; Development of Early Warning Indicators. *2nd iNTeg-Risk Conference, 14 – 18 June, Stuttgart, Germany.*
2. Thieme C. A., Utne I. B., (2017). Safety performance monitoring of autonomous marine systems. Reliability Engineering & System Safety. 159, March 2017, Pages 264-275.
3. Dekker S., (2006) *Resilience Engineering: Chronicling the Emergence of Confused Consensus*. Chapter 7 in Hollnagel, E., Woods, D. D. & Leveson, N. (Eds.), *Resilience Engineering: Concepts and Precepts*. Aldershot, UK: Ashgate, 2006.
4. Knegtering B., Pasman H., (2013). The safety barometer: How safe is my plant today? Is instantaneously measuring safety level utopia or realizable? *Journal of Loss Prevention in the Process Industries*. **26**(4), 821-829.
5. Leveson N., (2015). A systems approach to risk management through leading safety indicators, *Reliability Engineering and System Safety.* **136**, 17–34.
6. Karanikas, N. (2018). Documentation of Assumptions and System Vulnerability Monitoring: the Case of System Theoretic Process Analysis (STPA), *Proceedings of the 5th STAMP European Workshop*, 14-15 September 2017, Reykjavik University, Iceland, *International Journal of Safety Science*, 2(1), pp. 84-93
7. Chatzimichailidou M. M., Karanikas N., Dokas I. M., (2016). Measuring Safety Through the Distance Between Systems States with RiskSOAP Indicator, *Journal of Safety Studies*, 2 (2), 0-5.
8. Leveson N., Thomas J. P., (2018). STPA Handbook [on line], [Viewed 29 November 2018] Available from: http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf
9. Dokas I. M., J. Feehan and Imran S., (2013). EWaSAP: An Early Warning Identification Approach Based on a Systemic Hazard Analysis, *Safety Science*. **58**, pp 11-26
10. Leveson, N., (2011). Engineering a safer world: Systems thinking applied to safety. *MIT press*.