An Agent-based Approach For Safety Analysis of Safety-Critical Organizations

Alexei Sharpanskykh
VU University Amsterdam
sharp@cs.vu.nl

ABSTRACT

Modern safety-critical organizations are characterized by complex, nonlinear dynamics involving many interrelated actors and processes. Safety issues that emerge from these complex dynamics increasingly remain hidden, until an incident or even a serious accident occurs. Traditional safety analysis methods developed long ago for much simpler organizations cannot help identifying, explaining and predicting many safety-related properties of modern organizations. To address this issue, in the paper a formal approach is proposed to establish relations between local dynamics of actors of a complex safety-critical organization and global safety-related properties that emerge from these dynamics. In contrast to the traditional approaches, the organizational dynamics are specified by taking the agent perspective with an organizational layer. The application of the approach is illustrated by a simulation case study, in which spread of safety-critical information in an air navigation service provider is investigated.

Keywords

Safety analysis, multi-agent systems, organization modeling, emergence.

1 INTRODUCTION

In 2002 a tragic accident occurred over the towns of Überlingen and Owingen in Germany (Nunes and Laursen, 2004). Two aircraft collided in mid-air, nobody on board had survived. Directly before the accident the pilots of one aircraft followed instructions of an air traffic controller on duty and disregarded a conflicting advice from their automated collision warning system. On the contrary, the pilots of the second aircraft followed automated instructions of their onboard system. However, this conflict of instructions was not the only cause of the accident. By a deeper investigation (Nunes and Laursen, 2004) several organizational, cognitive and technological factors had been identified, which had led to the accident. Unfortunately many of these factors could not have been determined before the accident because of the lack of appropriate safety analysis tools, well-suited for safety analysis of modern safety-critical organizations (such as air navigation service providers, power plants, railway organizations).

Modern safety-critical organizations are characterized by complex, nonlinear dynamics involving many interrelated actors and processes. Safety issues that emerge from these complex dynamics increasingly remain hidden, until an incident or even a serious accident occurs. Traditional safety analysis methods (Bedford and Cooke, 2001; Eurocontrol, 2004) developed long ago for much simpler organizations cannot help identifying, explaining and predicting many safety-related properties (e.g., safety hazards, issues) of modern organizations. The need for more advanced safety modeling and analysis tools is well recognized in the industry; however the theoretical basis for these tools is largely missing.

To address this issue, in the paper a formal approach is proposed to establish relations between local dynamics of actors of a complex safety-critical organization and safety-related properties that emerge from these dynamics (e.g., safety hazards, safety culture properties, safety requirements). On the one hand, global (or systemic) consequences of local organizational dynamics can be determined using this approach in a bottom-up manner. On the other hand, for emerging safety-related properties major local contributing factors can be identified by the approach applied in a top-down manner. The knowledge about organizational "local-global" relations may also be used for a structured, systematic improvement of the organizational safety.

In contrast to the traditional approaches, the organizational dynamics are specified by taking the agent perspective (Weiss, 1999) with the organizational layer in the proposed approach. From this perspective global organizational properties emerge from local interactions and behavior of autonomous agents representing humans and technical systems situated in the organizational context.

To relate local organizational dynamics to global emergent properties four levels of abstraction are distinguished: internal/cognitive, behavioral, group, and global organizational levels. At each level dynamic properties can be identified. Relations between structures of (adjacent) levels of abstraction can be established by simulation or analytically (e.g., by using techniques from Mathematical Logics, Calculus and Control Theory). In this paper we focus mostly on simulation techniques, however a discussion on analytical tools is provided in Section 3.1.

The application of the approach is illustrated by a simulation case study, in which spread of safety-related information about safety occurrences in an air navigation service provider (ANSP) is investigated. Air traffic controllers in an ANSP are obliged to report safety occurrences observed during air and ground operations. An example of a ground occurrence is 'taxiing aircraft initiates to cross due to misunderstanding in communication'. Knowledge about safety occurrences is particularly useful for timely identification of safety problems in ANSPs. In practice, however, safety occurrences are not always reported. This may create a serious bottleneck in the organizational safety. Empirical evidences exist, which indicate that information about occurrences, which were not reported formally, may spread informally among air traffic controllers working in shifts. Currently, in ANSPs both fixed and variable compositions of shifts of controllers are used. In the simulation study we investigated effects of the shift composition on the completeness of organizational knowledge about safety occurrences. This knowledge comprises formal reports about safety occurrences, as well as knowledge of agents about occurrences, which has been obtained by informal interaction. Two types of ANSPs were considered: an ANSP, which is highly committed to safety and an ANSP with a meager commitment to safety.

The paper is organized as follows. In Section 2 related work is discussed. A novel, agent-based approach for safety analysis is described in Section 3. An agent-based organizational model developed for the case study using the approach is provided in Section 4. Results of the simulation study based on the developed model are discussed in Section 5. Some applications of the proposed approach are discussed in Section 6. The paper ends with conclusions in Section 7.

2 RELATED WORK

Traditionally, safety modeling and analysis are performed in a top-down manner starting from a set of safety issues, such as technical system failures (e.g., 'radar processing failure') and human errors (e.g., 'failure in recognizing deviation in altitude by a pilot').

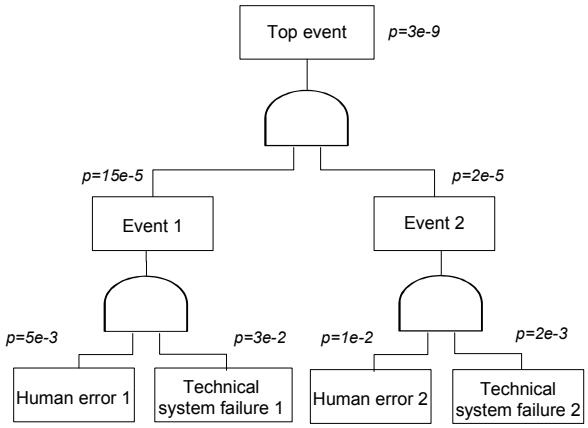


Figure 1. A fault tree

These issues are usually identified by brainstorming of safety engineers with domain experts. Then, for every safety issue temporally ordered sequences of events resulting into the issue (a fault tree, see Figure 1), and resulting from the issue (an event tree) are constructed (Bedford and Cooke, 2001; Eurocontrol, 2004). Fault and event trees are used for sequential (linear) reasoning about direct causes and safety effects of issues. Indirect (emergent) safety properties resulting from complex, nonlinear organizational dependencies and dynamics, as in the Überlingen accident, cannot be revealed and captured by the traditional methods (Hollnagel, 2004). Because of this limitation, some organizational safety issues may remain hidden until an incident or even a serious accident occurs. More advanced analysis techniques based on epidemiological accident models (e.g., Reason's "Swiss cheese" model (Reason, 1997), Bayesian belief networks (Greenberg, Cook, and Harris, 2005)) also suffer from the same limitation.

To address this shortcoming of the traditional approaches, Hollnagel (2004) advocated the systemic view on safety modeling and analysis. From this view safety issues and risks emerge from distributed, dynamic, nonlinear interaction of social and technical components of a system. Most of the few existing formal systemic approaches, STPA (Leveson, 2004) and SoTeRiA (Mohaghegh, Kazemi, and Mosleh, 2009) in particular, take an aggregated view on the organizational dynamics. By doing so, the link to the local behavior of organizational actors is lost, so that the level of analysis is reduced. Also, mapping the actual system structures and processes to the abstract aggregated model variables may be difficult and error-prone. Moreover, local dynamics of interacting actors may result in unexpected emergent global effects in the system, which cannot be analyzed using the aggregate view.

In contrast to these approaches, the TOPAZ safety assessment methodology (Stroeve, Blom, and Bakker, 2009) takes into account the local perspective of separate actors and their behavior. To achieve this, the *agent modeling paradigm* (Weiss, 1999) was used. An agent is an autonomous entity able to reason and interact with the environment. Agents can represent both humans and technical systems. The agent paradigm is particularly suitable for modeling interaction and behavior of actors in complex, highly dynamic organizations. Unfortunately TOPAZ does not address the organizational layer. This layer describes the formal organization, i.e., structures, properties, constraints and norms prescribed by organizational documentation, which are imposed on the organizational agents. An agent may reason about the formal organization and decide to which extent to comply with these prescriptions. TOPAZ also ignores the safety effects of the formal organization.

To address the identified limitations of the existing approaches, a novel approach for safety analysis based on models of agent organizations is proposed in the next section.

3 AN AGENT-BASED MULTI-LEVEL SAFETY ANALYSIS APPROACH

The proposed approach is based on identifying relations between local properties of actors of safety-critical organizations and global safety-related properties emerging from these local dynamics. To specify the local organizational dynamics the agent perspective with the organizational layer is taken. To establish "local-global" relations, four levels of abstraction (as shown in Figure 2) are introduced.

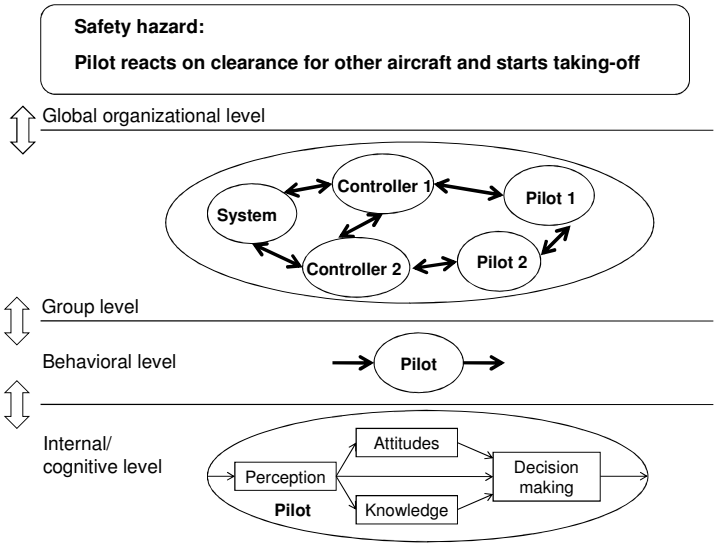


Figure 2. Four levels of abstraction of the dynamics of a safety critical organization in the context of air traffic

At the *Internal/cognitive level* internal properties of an organizational agent representing either a human (e.g., a pilot) or a technical system are specified. Cognitive specifications are particularly suitable to describe the dynamics of a proactive agent with reasoning capabilities about the environment, the formal organization, and its own states. Cognitive specifications may describe a great variety of human processes, e.g., perception, decision making, attention management, workload.

The behavioral properties of an agent, described at the *Behavioral level*, emerge from its cognitive dynamics. For example, a pilot may perform taking-off action as the result of his or her (complex) internal decision making. A behavioral property is a correlation of some temporal complexity between the agent's input and output states. Behavioral properties concern communication of an agent with other agents (e.g., interaction of an

air traffic controller with a pilot), and interaction of an agent with the environment by observation and actions (e.g., observation of an air traffic controller of aircraft on a radar). The dynamics of simple agents (e.g., stimulus-response agents) can often be specified directly at the behavioral level, without any cognitive specification.

To specify the local organizational dynamics at the Internal/cognitive and Behavioral levels, a formal, generic framework for modeling agent organizations developed previously is used (Sharpanskykh, 2008). An application of this framework in the context of the case study considered is provided in Section 4.

Based on behavioral properties of individual agents, properties of groups of agents may be determined, considered at the *Group level*. For example, for a shift of air traffic controllers such properties may describe aspects of the shift's performance, shared situation awareness, safety issues arising from interaction between shift members.

Safety-related properties emerge at the *Global organizational level* from interaction between behavioral and group properties. The following types of safety-related properties can be distinguished:

- *safety hazards*, which are organizational states that may result into incidents or accidents. For example, 'a pilot reacts on clearance given by a controller for other aircraft and starts taking off'.

- *safety culture issues*, which describe organizational bottlenecks that may have a negative impact on safety. For example, 'air traffic controllers have to break the rules to cope with the workload'.

- *safety requirements*, which are properties that need to be achieved or maintained to ensure appropriate safety levels in an organization. For example, 'to ensure timely investigation of safety occurrences in the organization'.

Relations between levels of abstraction can be established either by simulation or analytically. In the former case, a simulation model of a lower level of abstraction should be provided. Then, by simulation, properties of a higher level of abstraction are determined. For the example provided in Figure 2, a simulation model may be defined at the Behavioral level comprising behavioral properties of pilots, air traffic controllers, and technical systems, which describe taxiing and taking off operations. Then, by simulation the occurrence of safety hazards at the Global organizational level, such as 'pilot reacts on clearance for other aircraft and starts taking off', can be established. To specify simulation models a logic-based language called LEADSTO is used (Bosse et al, 2007). Dynamics in LEADSTO is represented as evolution of states over time. A state is characterised by a set of properties that hold at a certain point in time. LEADSTO enables modeling of direct temporal dependencies between two state properties in successive states, also called *dynamic properties*. For performing simulation based on LEADSTO specifications a dedicated automated tool exists. The simulation-based approach for establishing interlevel relation is further elaborated in the context of the case study in Sections 4 and 5.

Since a specification of local organizational dynamics may have a high complexity, establishing relations between levels of abstraction by analytical means is a challenging task. However it is necessary to address this challenge, since, in contrast to the simulation-based approach, knowledge about interlevel relations obtained analytically is general and can be applied for analysis of a wide range of safety-critical organizations. Previously, formal techniques based on logical stratification (Sharpanskykh and Treur, 2012) and interpretation mappings (Sharpanskykh and Treur, 2010) were developed to relate cognitive structures of an agent to the agent's behavior. Since analytical techniques are involved and rather technical, in this paper we focus on the simulation-based technique. The interested reader may refer to (Sharpanskykh and Treur, 2010, 2012) for an analytical approach.

4 AGENT-BASED ORGANIZATIONAL MODEL

For modeling organizational dynamics the organization modeling framework from (Sharpanskykh, 2008) was used. In this framework organizations are considered from different perspectives (views). *Process-oriented view* describes workflows as well as static structures of tasks and resources. *Performance-oriented view* describes a goal structure, a performance indicators structure, and relations between them as well as relations between goals and tasks, performance indicators and processes, goals and roles or agents. *Organization-oriented view* defines the organizational roles, each associated with a set of tasks and characterized by authority and responsibility relations on tasks, resources and information. Commitment, obligation and power relations and sets of competences required for agent allocation to roles are also defined. *Agent-oriented view* identifies different types of agents with their capabilities and behavior. The framework does not prescribe any specific architecture or approach for modeling of agents.

Furthermore, this framework describes a sequence of organization design steps, an overview of which is provided in Table 1. In the following it is discussed how a simulation model for the case study was constructed

using the framework along the design steps from Table 1. All properties in the simulation model were specified using the LEADSTO language.

In the simulation air traffic controllers work in 4 shifts. At step 1, the generic air traffic controller role (ATCO) was identified, instantiated into 12 role instances. The generic Shift role comprises 3 ATCO instances (see Figure 3). At step 2, two-way interaction links between the role instances within each shift were specified. Through these links role instances influence the each other's attitude towards reporting and communicate information about the observed safety occurrences. Furthermore, interaction links between each ATCO role instance and the environment were specified. Through these links information about safety occurrences that take place in the environment is propagated to the role instances.

Design step	View			
	Organization	Performance	Process	Agent
1. Identification of organizational roles	x			
2. Identification of interactions between roles and with the environment	x			
3. Identification of requirements for roles	x			x
4. Identification of organizational performance indicators and goals		x		
5. Identification of resources			x	
6. Identification of organizational tasks and workflows, and relations between tasks, resources and goals		x	x	
7. Identification of authority relations	x		x	
8. Identification of characteristics (skills, psychological and cognitive characteristics) of agents		x		x
9. Identification of goals and needs of agents				x
10. Identification of commitments, obligations and responsibilities of agents				x
11. Identification of attitudes and beliefs of agents				x
12. Identification of relations between agents (e.g., interaction, trust and informal power relations) and informal structures of agents	x			x
13. Allocation principles of agents to organizational roles				x
14. Specification of the environmental dynamics				x

Table 1. Overview of steps in organizational modeling and their relation with the views considered

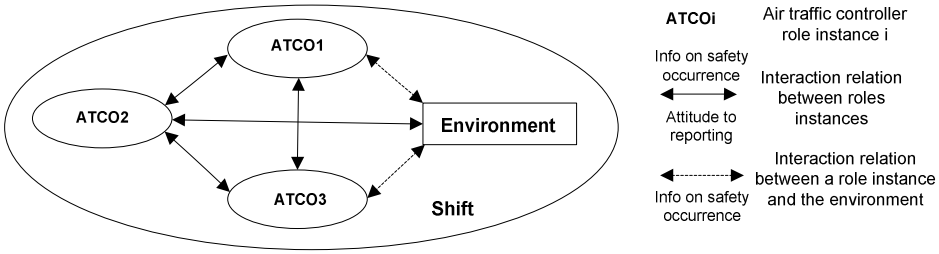


Figure 3. The roles and interaction relations between them within Shift role

At step 3, it is assumed that the controller agents allocated to the ATCO role instances in the simulation have all capabilities and skills required by the ATCO job prescription.

Steps 4 and 5 are not relevant for the case study.

At step 6 the following workflow is specified: A shift consists of three sessions. The duration of each session is 1 hour. If during a session an ATCO observes a safety occurrence, it makes the decision whether or not to create a formal report about this occurrence. After each session an obligatory break follows, which lasts for 1 hour.

Formal and informal power relations identified at steps 7 and 12 were combined by introducing the social contagion strength parameter between ATCOs *i* and *j*: $\gamma_{ij} \in [0,1]$. This parameter can be interpreted as the degree of influence of *i* on *j*. Air traffic controller supervisors of the shifts were not modeled as separate roles, but rather as ATCOs with a high social contagion strength towards other ATCOs of the shift.

It is generally accepted in Social Science that people influence each other more if they have similar states (e.g., similar opinions on some subject or similar interests) (Pinder, 1997). Following this principle, in the context of the case study the contagion strength $\gamma_{AB}(t)$ were made dependent on the closeness of the controller A and B's attitudes to reporting, i.e., $| attrep_A(t) - attrep_B(t) |$: the closer the attitudes of the controllers to reporting, the higher the contagion strength between them. To bound $\gamma_{AB}(t)$ to interval $[0,1]$, the logistic function is used:

$$\gamma_{AB}(t) = 1/(1+e^{-\omega 1 | attrep_A(t) - attrep_B(t)| + \omega 2}), \quad (1)$$

where $\omega 1$ is the degree of steepness, and $\omega 2$ is the threshold of the function. In the simulation $\omega 1 = 7$ and $\omega 2 = 5$.

Steps 8-10 were not considered explicitly in the study.

At step 11, the attitudes of the controller agents to safety occurrence reporting ($attrep_a(t) \in [0,1]$), and their beliefs about safety occurrences were specified. Previous studies (Stroeve, Sharpanskykh, and Kirwan, 2011) showed that this attitude depends greatly on the quality of the ANSP's safety culture. In the organizations with good safety culture the attitude to reporting of controllers is normally high. This type of organization is represented by ANSP1 in the simulation study. In ANSP1 the attitudes of the controller agents are drawn from the uniformly distributed range *[0.7, 1]*. In the ANSPs with meager safety culture the attitudes to reporting vary greatly between low and high (Stroeve, Sharpanskykh, and Kirwan, 2011). This type of organization is represented by ANSP2 in the simulation study. In ANSP2 the attitudes of the controller agents are drawn from the uniformly distributed range *[0.3, 1]*.

In each shift controllers influence each other's attitudes to reporting through a social contagion process (Deffuant et al., 2001):

$$attrep_A(t+\Delta t) = attrep_A(t) + \eta_A \, \delta_A(t) \Delta t, \quad (2)$$

where η_A is an agent-dependent parameter within the range [*0,1*], which determines how fast agent *A* adjusts to the opinion of other agents. This parameter reflects the idea from the Lewin's change management model (Lewin, 1951) that opinion change is not an event, but rather a process, which may run at different rates depending on personality characteristics of agents, obtained training, etc. The influence of η_A on the simulation results is investigated in section 5. The change of the reporting attitude $\delta_A(t)$ from *(2)* is defined by:

$$\delta_A(t) = \sum_{shift(A) \setminus A} \gamma_{BA} \, [attrep_B(t) - attrep_A(t)] \; / \; \sum_{shift(A) \setminus A} \gamma_{BA}, \quad (3)$$

where *shift(A)* is the set of all controller agents that belong to the shift of controller agent *A*, γ_{BA} is the social contagion strength from *B* to *A* defined by *(1)*. The intuition behind the formulae (2) and (3) is that the agent changes its attitude to reporting gradually in the direction of the opinions of other agents with whom it interacts. The higher γ_{BA} value, the more influence the agent B's opinion has on the agent A's opinion. It is assumed that controller agent *A* reports an observed safety occurrence only when $attrep_A(t) > rth$, where *rth* is the reporting threshold. The influence of *rth* on the simulation results is investigated in section 5. All safety occurrences observed by a controller agent are stored in the form of agent's beliefs.

At step 12, to model informal interaction by which controllers exchange knowledge about safety occurrences, the Burt's social contagion theory was used (Burt, 1987). According to this theory, the intensity of informal communication between actors in an organization is influenced positively by the following factors: (1) similarity of the communication patterns of the actors; (2) equality of the statuses of the actors in the organization; (3) physical possibilities to communicate; (4) similarity of states of the actors. In a shift of air traffic controllers the factors (1)-(3) are well-pronounced. Since the contagion strength parameter γ_{AB} depends on the degree of similarity of states of agents *A* and *B*, (4) is expressed by γ_{AB}. It is assumed that controller agents exchange information about occurrences during breaks, only when their contagion strength to each other is sufficiently high, i.e., $\forall A,B \; such \; that \; B \in shift(A): \; \gamma_{AB} > csth \; \& \; \gamma_{BA} > csth$, where *csth* is the contagion strength threshold. The influence of *csth* on the simulation results is investigated in section 5.

At step 13, two mechanisms of allocation of 12 agents to the ATCO role instances of the organization were introduced, which are often employed in real ANSPs. The first mechanism is based on fixed shifts of air traffic

controllers: the same agents are allocated to the ATCO role instances in each shift throughout the whole simulation. The second mechanism is based on a variable shift composition: every simulation day (12 hours) the controller agents are randomly allocated to the ATCO roles in 4 shifts. The effects of these two allocation mechanisms on the global organizational dynamics are discussed in section 5.

At step 14, the environmental dynamics is specified by a random process with safety occurrences taking place at each time point (1 hour in the simulation model) with probability 0.05.

5 SIMULATION STUDY

In the simulation study we investigated how the composition of shifts (i.e., fixed vs. variable allocation of controller agents) in ANSP1 (with good safety culture) and ANSP2 (with meager safety culture) affects safety-related properties of the shifts and global organizational safety-related properties. The following properties of the shifts were considered:

SP1(S, t): The reporting ratio of safety occurrences of shift S:

$SP1(S, t) = \sum_{a \in S_t} OR_a(t) / \sum_{a \in S_t} OO_a(t)$,

where set S_t contains all names of air traffic controller agents that belong to shift S at time point t; $OR_a(t)$ is the number of occurrences formally reported by controller a in the time interval $[0, t]$, $OO_a(t)$ is the number of occurrences observed by controller a in the time interval $[0, t]$.

SP2(S, t): The completeness of knowledge about safety occurrences possessed by shift S:

$SP2(S, t) = |\cup_{a \in S_t} OK_t^a| / \sum_{a \in CONTROLLER} OO_a(t)$,

where set S_t contains all names of air traffic controller agents that belong to shift S at time point t; $CONTROLLER$ is the set of all names of the air traffic controller agents; OK_t^a is the set of the names of occurrences known to controller a at time point t, $OO_a(t)$ is the number of occurrences observed by controller a in the time interval $[0, t]$.

At the global organizational level two properties related to the completeness of knowledge possessed by the organization are considered. The first property **OP1(t)** indicates *the completeness of formal knowledge* (i.e., in form of reports) possessed by the organization at time t. It is defined similarly to the shift property SP1(S, t):

$OP1(t) = \sum_{a \in CONTROLLER} OR_a(t) / \sum_{a \in CONTROLLER} OO_a(t)$

The second property **OP2(t)** reflects *the completeness of the overall organizational knowledge*, gained by the controller agents both formally and informally. It is defined similarly to the shift property SP2(s, t):

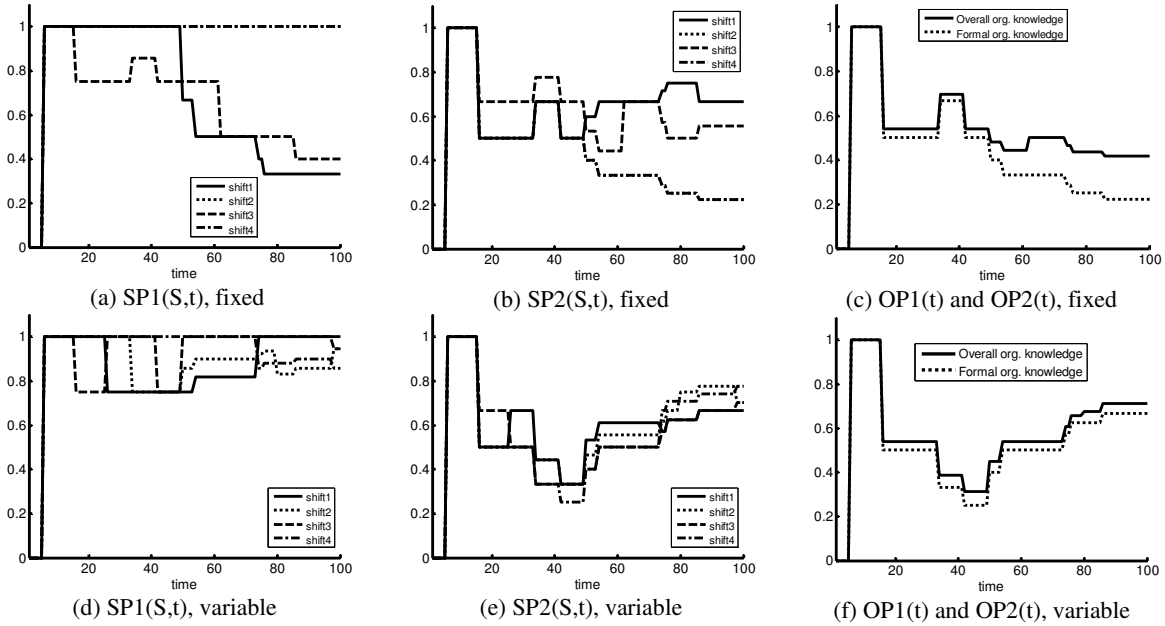
$OP2(t) = |\cup_{a \in CONTROLLER} OK_t^a| / \sum_{a \in CONTROLLER} OO_a(t)$

Both shift properties and global organizational properties were evaluated by simulation based on the local agent-based organizational model described in Section 4. In the simulation 4 setups were considered: (1) ANSP1 with a fixed composition of shifts; (2) ANSP1 with a variable composition of shifts; (3) ANSP2 with a fixed composition of shifts; (4) ANSP2 with a variable composition of shifts. Every setup was simulated 1000 times. In Figure 4 the graphs for the shift properties SP1(S,t) and SP2(S,t), and organizational properties OP1(t) and OP2(t) for one run of the setups (3) and (4) are provided. The same agents and the environmental dynamics were used for both setups. The mean and variance values for the global organizational properties OP1(t) and OP2(t) for all simulation setups over 1000 runs are provided in Table 2.

As one can see from Figure 4, a variable composition of shifts decreases knowledge differences between the shifts. Furthermore, by a rotation of agents, knowledge about occurrences is propagated through social contagion in each shift, resulting in a higher degree of completeness of the overall organizational knowledge (compare (c) with (f) in Figure 4). Also, according to the simulation results, a variable composition of shifts hinders the formation of a joint negative attitude to occurrence reporting in a shift, reflected by SP1(S,t). Specifically, in Figure 4(a) the agents in fixed shifts 1 and 3 are gradually decreasing their willingness to report, whereas the attitudes to reporting in fixed shifts 2 and 4 remain highly positive. When a variable composition of shifts is introduced in such an organization (Figure 4(d)), the attitudes to reporting in all shifts remain highly positive. This result was confirmed by observations of management in an existing Western European air navigation service provider, in which we conducted an interview.

As can be seen from Table 2, also for ANSP1 a variable composition of shifts results into a higher than with fixed shifts completeness of the formal and overall organizational knowledge. Furthermore, according to the simulation results, the positive effects of the variable composition of shifts are significantly lower for ANSPs

with good safety culture (ANSP1) than for ANSPs with meager safety culture (ANSP2). This result still needs to be validated.



(a) SP1(S,t), fixed (b) SP2(S,t), fixed (c) OP1(t) and OP2(t), fixed

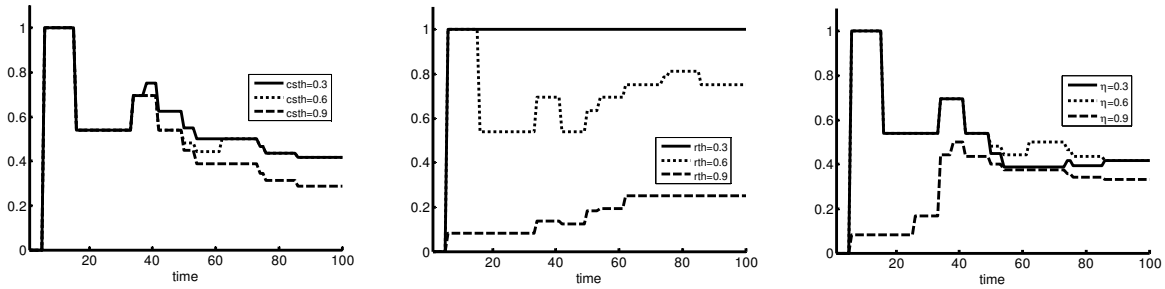
(d) SP1(S,t), variable (e) SP2(S,t), variable (f) OP1(t) and OP2(t), variable

Figure 4. The graphs for shift properties SP1(S,t) and SP2(S,t) and organizational properties OP1(t) and OP2(t) for fixed and variable shifts of ANSP2

	Formal org. knowledge (OP1)		Overall org. knowledge (OP2)	
	Fixed	**Variable**	**Fixed**	**Variable**
ANSP1	0.92 (3e-28)	0.95 (1e-3)	0.92 (2e-28)	0.99 (7e-3)
ANSP2	0.22 (2e-30)	0.42 (3e-3)	0.46 (5e-30)	0.62 (8e-3)

Table 2. The mean and variance (in brackets) values of the degree of completeness of formal and overall organizational knowledge about safety occurrences at the last simulation time step in 4 simulation setups simulated 1000 times each.

In the following the sensitivity of the simulation results to parameter settings is explored. A change of the value of the contagion strength threshold *csth* influences the completeness of the overall organizational knowledge insignificantly (Figure 5(a)). At the same time, even a small variation in the reporting threshold *rth* causes a large change in the knowledge completeness (Figure 5(b)). Also, parameter η has a significant influence on the simulation results (Figure 5(c)). In the simulation study, to enable comparison of the simulation results along different setups, all three parameters were kept constant in all simulation trials.



(a) (b) (c)

Figure 5. Sensitivity of the overall organizational knowledge (OP2) in ANSP2 with fixed composition of shifts to settings of parameters *csth, rth, and* η

6 APPLICATIONS

The application of the proposed approach is not limited to air traffic organizations only. It can be applied for analysis of dynamics of safety-critical organizations in other domains too (such as power plants, railway

systems, incident management organizations). The organization modeling framework, on which the approach is based, is generic, and addresses all most essential aspects of safety-critical organizations, including performance, processes and resources, interaction and power relations between organizational actors.

The main group of the potential users of the proposed approach are organizational policy makers and safety managers. Using the proposed approach, they may be able to get a better picture of the organizational dynamics, to predict organizational weaknesses and their effects, and to analyze safety incidents and their causes in hindsight. Furthermore, using the approach they may be able to initiate more informed organizational safety improvement arrangements, which could first be tested by computer simulation. In the following, several possible implications of the approach are discussed in more detail.

Quantitative estimation of safety risks of operations performed in safety-critical organizations taking into account cognitive, social, technical, and institutional factors.

A risk is a combination of the frequency and severity of an undesirable state or event that occurs in an organization. Risks are considered to be quantities that emerge from "distributed, dynamic, nonlinear interaction of social and technical components of a system" (Hollnagel, 2004). Using the proposed approach, emerging risks of operations may be calculated by relating them to local properties of agents performing these operations in the organizational context. These calculated risks may be presented to organizational policy makers to improve their awareness about hidden organizational hazards.

Identifying previously unknown safety issues in safety-critical organizations.

For this the proposed approach can be applied in a bottom-up manner. Starting from local properties of individual agents, properties of groups of interacting agents can be determined. From these properties diverse global safety effects can be inferred using the proposed approach. For example, in such a way properties emerging from complex interactions of different parties (e.g., police, ambulance, authorities) in an incident management organization can be determined.

Identifying the most influential organizational factors responsible for particular safety-related properties (e.g., issues) of organizations.

For this the proposed approach can be applied in a top-down manner. For a global safety-related property related structures of properties of the adjacent lower level of abstraction can be identified. Then, these structures can be further related to structures of other levels until the lowest level of abstraction is reached. In such a way, the most influential organizational factors causing the emergence of the safety-related property can be identified. Based on these factors safety improvement options that involve organizational change may be formulated. To evaluate the effects of these options the proposed approach can be used as well.

Determining consequences of organizational changes for the organizational safety.

To evaluate the global safety effects of local organizational changes the proposed approach can be applied in a bottom-up manner. Using the approach changes in local properties of an organization can be propagated to higher levels of abstraction, until the global level is reached.

Refinement of safety requirements.

To determine how to implement global organizational safety requirements set by formal regulatory bodies or by management, the proposed approach can be applied in a top-down manner. Safety requirements can be seen as global, emergent safety properties of an organization, which can be related by the proposed approach to local properties of agents that need to be ensured in the organization.

7 CONCLUSIONS

In this paper a formal agent-based approach for safety analysis of safety-critical organizations is proposed. The approach is based on distinguishing four levels of abstraction in a system specification, and establishing relations between these levels. The approach was illustrated by a simulation case study, in which spread of information about safety occurrences in an ANSP was investigated. The local dynamics of the ANSP was specified by an agent-based organizational model at the internal/cognitive and behavioral levels. Then, by simulation, relations were established between this local model and properties of groups (shifts) and global organizational properties, which characterize the completeness of knowledge about safety occurrences. In particular, it was established by simulation that a variable composition of shifts hinders the formation of the joint negative attitude to occurrence reporting in an ANSP and increases the completeness of the overall organizational knowledge. This result was confirmed in interviews with management of an existing ANSP. The interviews were also useful for the model development: they provided insights in the implementation of the

ANSP's safety occurrence assessment procedures and related informal aspects. More details revealed in the interviews with management are discussed in (Stroeve, Sharpanskykh, and Kirwan, 2011). Another result obtained by simulation is that a variable composition of shifts has a greater positive effect in the ANSPs with meager safety culture than in the ANSPs with good safety culture. This result still requires validation.

The proposed approach can be applied to a wide range of safety-critical organizations (such as power plants, railway systems, incident management organizations), and is aimed to enhance the existing safety modelling and analysis practices significantly.

ACKNOWLEDGMENTS

This research is supported by the Dutch Technology Foundation STW, which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs.

REFERENCES

1. Bedford, T., and Cooke, R. (2001) *Probabilistic Risk Analysis: Foundations and Methods* Cambridge, Cambridge University Press.
2. Bosse, T., Jonker, C.M., Meij, L. van der, and Treur, J. (2007) A Language and Environment for Analysis of Dynamics by Simulation. *International Journal of Artificial Intelligence Tools*, vol. 16, pp. 435-464.
3. Burt R.S. (1987) Social Contagion and Innovation: Cohesion Versus Structural Equivalence Source: The *American Journal of Sociology*, 92, 6, pp. 1287-1335.
4. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G. (2001) Mixing beliefs among interacting agents *Advances in Complex Systems*, 3, pp.87-98.
5. Greenberg, R., Cook, S.C., Harris, D. (2005). A civil aviation safety assessment model using a Bayesian belief network (BNN). The Aeronautical Journal, 2005, pp. 557–568.
6. Eurocontrol. (2004). *Air navigation system safety assessment methodology*. SAF.ET1.ST03.1000-MAN-01, edition 2.0, 2004.
7. Hollnagel, E. (2004) *Barriers and accident prevention*. Aldershot, UK: Ashgate Publishing.
8. Leveson N (2004) A new accident model for engineering safer systems. *Safety Science* 42, pp. 237-270.
9. Lewin, K. (1951) *Field theory in social science; selected theoretical papers.* D. Cartwright (ed.). New York: Harper & Row.
10. Mohaghegh Z, Kazemi R, and Mosleh A. (2009) Incorporating organizational factors into probabilistic risk assessment (PRA) of complex socio-technical systems: A hybrid technique formalization. Reliability *Engineering and System Safety* 94, pp. 1000-1018.
11. Nunes, A, and Laursen, T. (2004). Identifying the factors that contributed to the Ueberlingen midair collision: implications for overall system safety. *Proceedings of the 48th Annual Chapter Meeting of the Human Factors and Ergonomics Society*, New Orleans, LA, USA.
12. Pinder C.C.(1998) *Work motivation in organizational behavior* Upper Saddle River, NJ: Prentice-Hall.
13. Reason, J. (1997) *Managing the Risk of Organizational Accidents* Ashgate, Aldershot, England.
14. Stroeve, S.H., Blom, H.A.P., Bakker, G.J. (2009) Systemic accident risk assessment in air traffic by Monte Carlo simulation *Safety Science*, 47, pp. 238-449.
15. Weiss, G. (ed.). (1999) *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence* MIT Press, MA.
16. Sharpanskykh, A. (2008) *On Computer-Aided Methods for Modeling and Analysis of Organizations* PhD thesis, VU Amsterdam.
17. Sharpanskykh, A., and Treur, J. (2010) Abstraction Relations Between Internal and Behavioural Agent Models for Collective Decision Making. In: Pan, J.-S., Chen, S.-M., and Kowalczyk, R. (eds.), *Proceedings of the Second Int. Conference on Computational Collective Intelligence*, Springer, LNAI6421, pp. 39-53.
18. Sharpanskykh, A., and Treur, J. (2012) Abstraction Relations Between Internal and Behavioural Agent Models for Collective Decision Making. *Web Intelligence and Agent Systems Journal* (in press)
19. Stroeve, S., Sharpanskykh, A., and Kirwan, B. (2011) Agent-based organizational modelling for analysis of safety culture at an air navigation service provider *Reliability Engineering & System Safety*, 96, pp. 515-533.