Hazard prevention in mission plans for aerial vehicles based on soft institutions

Hazard prevention in mission plans for aerial vehicles based on soft institutions

Flavio S. Correa da Silva[1], Paul W. H. Chung[2], Marcelo K. Zuffo[1], Petros Papapanagiotou[3], David Robertson[3], and Wamberto Vasconcelos[4] *

[1] University of Sao Paulo, Sao Paulo Brazil
[2] Loughborough University, Loughborough UK
[3] University of Edinburgh, Edinburgh UK
[4] University of Aberdeen, Aberdeen UK

Abstract. Hazard prevention in mission plans requires careful analysis and appropriate tools to support the design of preventive and/or corrective measures. It is most challenging in systems with large sets of states and complex state relations. In the case of sociotechnical systems, hazard prevention becomes even more difficult given that the behaviour of human centric components can at best be partially predictable. In the present article we focus on a specific class of sociotechnical systems – namely air spaces containing pilot controlled as well as autonomous aircrafts – and introduce the notion of *relevant hazards*. We also introduce *soft institutions* as an appropriate basis for analysis, with the aim of addressing relevant hazards. The concept of soft institutions is drawn from specification languages for interaction between agents in multi-agent systems but, in our case, is adapted for use in systems that combine human and automated actors.

Keywords: Safety engineering, hazard prevention, sociotechnical systems, soft institutions.

1 Introduction

Hazard prevention requires the assessment of all possible behaviours of a system so that safety engineers can intervene in the system design to ensure that each behaviour leads to planned, foreseen and safe states [1], providing information support to design preventive and/or corrective measures for each potential hazard.

* This work has been partially supported by FAPESP-Brazil and by the EPSRC-UK. The present article is a revised and extended version of the article *Hazard identification for UAVs based on soft institutions*, by the same authors, presented at the workshop *Coordination, Organisations, Institutions and Norms – AAMAS 2017.* Many important comments and criticisms on early versions of this work have been generously provided by Dr. David Murray-Rust (Edinburgh, UK) and Dr. Amanda Whitbrook (Derby, UK).

Hazard prevention is most challenging in systems with large sets of states and complex state relations, which require careful planning and appropriate tools to generate and analyse potential hazard states, avoiding issues related to undecidability or combinatorial explosion during exhaustive scan of state spaces. In the case of sociotechnical systems, hazard prevention becomes even more difficult given that the behaviour of human centric components can at best be partially predictable.

The concept of sociotechnical systems was coined in the early 50s to analyse the impact of the introduction of novel technologies in coal mining, after the empirical observation that gains in productivity were not uniform in all studied workgroups. Its roots can be traced back to the analysis of the introduction of mechanisation in jute milling in Scotland during the 30s [3, 11]. Sociotechnical systems can be characterised as open asynchronous concurrent systems in which some entities are humans and others are machines. Hence, interactions involving heterogeneous entities are a central concept to design, implement and analyse sociotechnical systems.

In the present article we focus on safety and reliability and, more specifically, on the construction of tools to support systems design based on hazard prevention. Given that it can be impossible or too difficult to fully predict the behaviour of a sociotechnical system as a whole, we introduce the notion of *relevant hazards* to be considered during the design of a system.

In brief, we characterise a well determined subset of the set of all potential hazards for a system and perform backward induction to identify all initial states and chains of events that can lead to them. We then revise the system design in order to identify points in which design interventions can either prevent hazards or inject remedial procedures to be taken in case they occur.

We focus on a specific class of sociotechnical systems for which hazard prevention is particularly relevant – namely, bounded air spaces containing pilot controlled aircrafts as well as unmanned aerial vehicles (UAVs). We introduce a diagrammatic language to support the characterisation of relevant hazards, of sequences of events that can lead to them and of events to which can be associated actions to be kept in store for each relevant hazard.

We also introduce *soft institutions* as an appropriate platform for hazard prevention based on relevant hazards, and illustrate how soft institutions can be used as a formal counterpart to diagrams employed to design a system for safe operations in bounded air spaces in which pilot controlled aircrafts share space with UAVs.

This paper is organised as follows:

- In section 2 we detail a characterisation of sociotechnical systems, highlighting as a relevant special case mission planning for coordinated UAVs with diversified levels of autonomy.
- In section 3 we briefly introduce the main concepts related to hazard prevention and characterise in detail the notion of relevant hazards. We also introduce a diagrammatic language to represent sociotechnical systems aiming specifically at the prevention and analysis of failures.

– In section 4 we illustrate how the proposed diagrammatic language can be used to characterise complex agent interactions in such way that hazard prevention is supported. As a concrete example, we illustrate how it can be used to support the design of missions in bounded air spaces in which pilot controlled aircrafts share space with UAVs.
– In section 5 we introduce the concept of soft institutions, a corresponding computational platform based on this concept and how it can be used as a platform to support hazard prevention for the design of sociotechnical systems.
– Finally, in section 6 we present a brief discussion, conclusions and proposed future work.

2 Sociotechnical systems

A sociotechnical system can be characterised as an open network of heterogeneous interacting entities which can exchange messages and, therefore, coordinate their actions. Some of these entities are engineered and can be programmed to behave according to rules which are explicitly determined and fully understood, even in the cases when they are not fully deterministic; other entities are human centric and therefore their behaviour can, at best, be nudged towards desired patterns of behaviour.

Following Davis et alli [3] we can characterise six facets of sociotechnical systems:

1. **People** characterised as interacting entities who can have different competences, attitudes, skills and interests, based on which they coordinate their actions with other entities as well as are considered by other entities in proposals for coordination and collaboration;
2. **Technologies** and tools that characterise engineered interacting entities which have different capabilities to sense, interpret and act upon the environment, based on which they can engage into interactions;
3. **Processes/procedures** embodied as programs and rules for engineered entities as well as norms, regulative policies, sanctioning and incentive mechanisms to steer people towards expected patterns of behaviour;
4. **Buildings/infrastructure** which characterise environmental resources as well as constraints for interactions;
5. **Goals** and metrics to characterise whether the system as a whole as well as its individual entities are approaching or diverting from goals; and
6. **Culture** which characterises defeasible assumptions and heuristics shared and adopted by groups of entities participating in a sociotechnical system.

Depending on the combination and organisation of these facets, different design strategies for sociotechnical systems are most appropriate and require different strategies for design, implementation and management of sociotechnical systems:

1. **Openness** to admit or dismiss entities: a system can be *closed*, *partially open* or *fully open* to the admission or dismissal of entities. Partially open systems can require certain conditions to be fulfilled in order to admit or dismiss entities from it;
2. **Coordination** levels among entities: a system can be *uncoordinated*, *locally coordinated* or *globally coordinated*. In other words, entities participating in a sociotechnical system can act fully on their own, based on coordination rules involving groups of entities or based on coordination rules that engage the whole system to behave globally as a mechanism;
3. **Heterogeneity** of entities in a system: a system can be comprised *primarily of humans* – thus characterising a social network in which human entities communicate and interact; *primarily of technological entities* – thus characterising a distributed computational system, possibly containing entities whose behaviour is not fully deterministic; or have *varying proportions of humans and technological entities*;
4. **Statefulness**: a sociotechnical system can be *stateless*, i.e. the global state of the system as well as the internal states of entities are static, and therefore do not need to be managed; *globally stateful*, i.e. the global state of the system can change but the internal states of entities are static, and therefore entities can be reactive and their modelling is simplified; or *fully stateful*, i.e. the global state of the system as well as the internal states of entities are dynamic and must be monitored and managed;
5. **Context sensitiveness**: updates in the environment can be *irrelevant or unnoticeable*, in which case context needs not be managed; *dynamic although irrespective of the states of the system*, in which case the system as a whole as well as its components must be able to monitor changes in the environment and to adapt accordingly; and *dynamic and sensitive to system states*, in which case system components must be able to monitor changes in the environment, correlate these changes with their actions and adjust actions to manage the environment while they pursue their goals.

In the present work we are specifically interested in bounded air spaces in which pilot controlled aircrafts share space with UAVs. In this scenario, a system is typically:

1. **Partially open**, as aircrafts are allowed in and out of the air space provided that well specified rules and norms are followed;
2. **Locally coordinated**, as entities communicate and coordinate their actions following strict protocols which induce a hierarchy of control;
3. **Heterogeneous**, as we are considering autonomous vehicles interacting with pilot controlled vehicles and control systems comprised by sensors and actuators as well as human operators;
4. **Fully stateful**, as the states of individual entities – especially engineered entities – must be stored and managed in order to manage the whole system, particularly with respect to hazard prevention and engineering;
5. **Sensitive to system states** and changes resulting from external factors as well as from consequences of state updates of entities.

Our focus in the present article is on hazard prevention during system design. We are interested in structuring the interactions among entities in this scenario in such way that all relevant hazards are taken into account and design decisions are made in order to avoid failures or to build readiness to fix them in case they occur.

3 Hazard prevention based on relevant hazards

We adopt the simplifying assumption that all participating entities have been admitted to the system by following the interaction protocols that characterise it. Entities which do not follow certified interaction protocols are considered as external entities which can influence but are not part of the system and, therefore, are not subject to design decisions related to it.

We also assume that the behaviour of an entity can be completely described by the interactions in which it is prepared to participate. The internal functioning of any entity is not taken into account explicitly. This way, human centered entities can be considered uniformly together with complex engineered entities, and entities can be described using different levels of abstraction, according to the level of detail used to specify each interaction protocol under consideration.

Two fundamental strategies can be considered for hazard prevention during systems design [6]:

1. **Avoiding that things go wrong**, i.e. anticipating hazards and their corresponding causes, to allow system re-design in order to prevent those causes to occur, and
2. **Ensuring that things go right**, i.e. identifying hazards and their corresponding causes, and then looking ahead to events that can be a consequence of those hazards, so that corrective measures can be included in the system for each of the considered failures and/or their causes.

We focus on a subset of the set of *all* hazards, which are considered to be the *relevant* ones, which are in fact the ones we are able to advance during synthesis and scrutiny of a system design. The design of complex systems that are resilient to failures must combine these two strategies in such way that all relevant hazards are considered.

In summary, our proposed strategy for hazard prevention during the design of a sociotechnical system is based on the principles outlined in Figure 1.

In order to support this strategy, we introduce a simple diagrammatic language to abstract entities in a sociotechnical system based on interaction protocols. The proposed language is presented in Figure 2.

Each element in the proposed language can be represented using standardised notation as presented in Figures 3 and 4. Our purpose while designing this language was to make it as simple and compact as possible, as well as easy to translate as declarative executable specifications using the existing infrastructure based on *soft institutions*, as detailed in section 5.

1. *System entities are uniformly abstracted as components capable of reacting to incoming messages from other, interacting components. Their reactions correspond to*
   (a) *triggering internal, encapsulated behaviours which are influenced by environmental events,*
   (b) *updating internal states, and*
   (c) *interfacing with well specified interaction protocols which can generate outgoing messages to other components.*
2. *General system states and behaviour can be characterised by published states, messages and interaction protocols used by system entities.*
3. *Hazard prevention can be performed based on general system states.*
4. *Hazard prevention based on relevant hazards corresponds to the prevention of a set of system states which are considered hazards, prevention of events that can lead to these states, and prevention of events that can result from hazards.*

Fig. 1. *Principles for hazard prevention*

In Figure 3 we depict an entity which can participate in several contexts and assume several states within each of these contexts. For each state there are several interaction protocols which can be triggered by the entity. Some protocols have hand-offs in different contexts and/or states. Interaction protocols are portrayed as graphs inside white rectangles and hand-offs are represented as dashed arrows connecting graphs.

In Figure 4 we depict all possible types of actions that can belong to an interaction protocol.

As a brief example to illustrate the use of the diagrams, we feature in Figure 5 two entities – namely, a UAV and the Air Traffic Control (ATC) – during a simple interaction[5]. In this interaction, if necessary the UAV refuels and then it asks for permission to take-off. The ATC confirms the permission to take-off, and then the UAV changes state from *standing* to *taxiing*.

Hazard prevention can raise the possibility that the message from the ATC never gets to the UAV. Backward reasoning could suggest that the exchange of messages between the UAV and the ATC should contain additional steps, so that the UAV would acknowledge receipt of the message and the ATC would not stop sending copies of the permission to take-off until receiving an acknowledgment. Forward reasoning could suggest the inclusion of a time-out sensing operation as part of the interaction protocol for the UAV in *standing* state, to prevent the UAV from staying idle in case the message from the ATC never arrives. Both strategies could be combined in order to design a system that is resilient to failures.

[5] A detailed example is presented in section 4.

- *System entities are represented as boxes. Each box is labeled by a unique ID that identifies the corresponding entity.*
- *Inside an entity box we can have another boxes representing the set of contexts into which the entity can enter.*
- *Inside a context box we can have another boxes representing the set of states admitted for the entity in that context.*
- *Inside a state box we can have another boxes representing the set of interaction protocols allowed for an entity in a given context and state. An interaction protocol can make an entity change context and/or state. In this case, the interaction protocol has a hand-off in a different context and/or state.*
- *Inside an interaction protocol we have a directed graph of actions, in which nodes represent individual actions and edges characterise the order in which actions must occur in the interaction protocol. Every graph of actions has a root node which determines the first action to be performed, followed by its successor nodes in sequence. A branch represents a committed choice. A confluence represents a continuation that can be performed once at least one of the conflating branches succeeds. Hence, a graph of actions is a concise representation for a collection of alternative chains of actions that comprise an interaction protocol. An action can correspond to (1) querying the knowledge base of an entity, (2) performing a sensor-based operation in the environment, based on which the entity captures information from the environment, (3) receiving a message from another entity. Incoming messages must be sent by a specific entity in a given context and state, (4) updating a statement in the knowledge base of an entity, (5) performing an actuator-based operation in the environment, based on which the entity performs actions upon the environment, (6) sending a message to another entity. Outgoing messages must be addressed to a specific entity in a given context and state, or (7) changing context and/or state of the entity, in which case a hand-off of the interaction protocol in a different context and/or state is triggered.*
  Actions containing queries to the knowledge base, sensor-based operations and receipt of messages are called in-actions while actions corresponding to updates in the knowledge base, actuator-based operations, remittance of messages and change of context and/or state are called out-actions. Sequences of in-actions can work as preconditions for individual out-actions to occur.

Fig. 2. *Diagrammatic language to represent entities in sociotechnical systems*

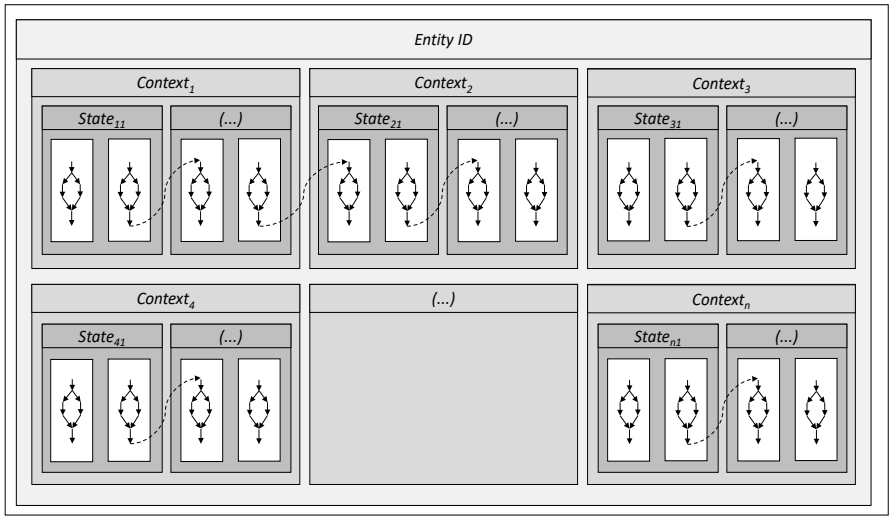

Fig. 3. Representation of an entity in the diagrammatic language

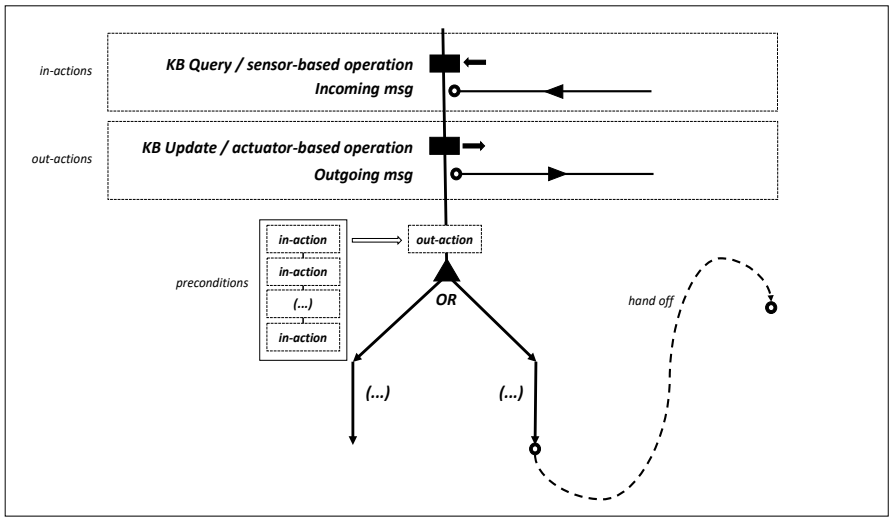

Fig. 4. Representation of actions in interaction protocols in the diagrammatic language

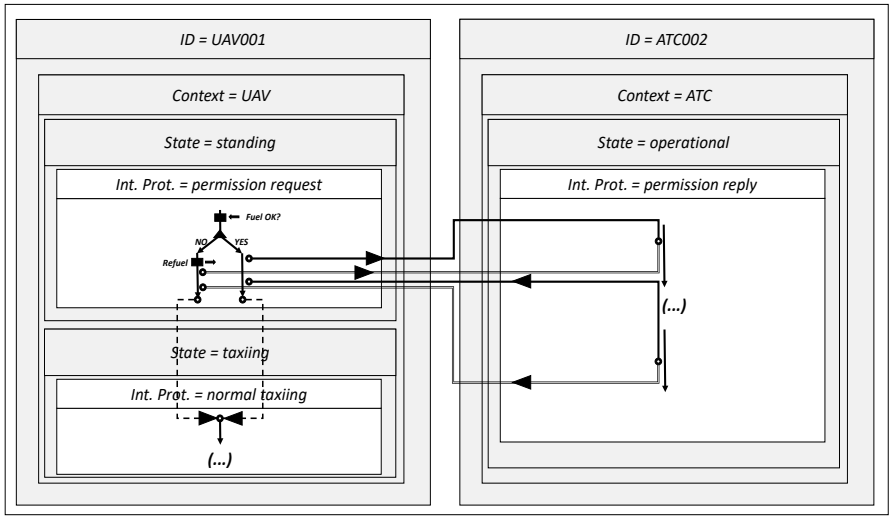

Fig. 5. Message exchanges between an autonomous UAV and an ATC

Our purpose in building this diagrammatic language has been to support system designers activities with a clear and intuitive pictorial language capable of exposing hazards in a system which can then be considered accordingly.

In the next section we present a detailed example in which a UAV is followed from standing off-lane through flying to landing. We use this example to illustrate how the proposed diagrammatic language can be used to represent complex systems in operation and how it can be used to identify hazards and help in the refinement of system design to provide appropriate care to potential hazards.

4 An illustrative example

In order to show how the proposed diagrammatic language can be used for hazard prevention, we consider a slightly more sophisticated example in which a complete mission for a UAV is depicted and analysed. This mission corresponds to a complete flight – from standing off-lane through flying to landing – and requires interactions involving the UAV and an ATC. The number of states through which the UAV passes is seven: *Standing*, *Taxiing*, *Take-off*, *Initial climb*, *En route*, *Approach* and *Landing*.

The diagrams corresponding to each state are depicted in Figures 6 to 12.

In Figure 6 the entity UAV001 is initially switched off and off-lane. It is assumed that it is listening to the appropriate channel for messages to receive a message requiring it to start the engine, which takes entity UAV001 to the context of UAV and standing state. The message triggers the interaction protocol depicted in Figure 6. When it receives a message to start the engine, it updates

the knowledge base and performs the action of starting the engine. It then queries the knowledge base to check whether the engine has started. If there is a failure, then it tries again to start the engine, otherwise it updates the knowledge base and checks fuel level and systems. If there is a problem, then it stops the engine and tries to start again, otherwise it updates the knowledge base and hands off control to an interaction protocol in Taxiing state.

The proposed strategies for hazard prevention and prevention/recovery have resulted in the loops back to the engine start message, together with the action to stop the engine in case fuel and system messages indicate that the UAV is not ready for flying.

In Figure 7 we have two entities, resp. UAV001 and ATC001. UAV001 stays in the context of UAV but now moves to taxiing state. ATC001 assumes context ATC and state to authorise taxiing towards take-off.

The interaction protocol for UAV001 in context UAV and taxiing state is slightly more complex than the protocol for standing state. UAV001 sends a message to an entity that is available in the context of ATC. In our example, ATC001 receives this message and replies back with either *take-off OK* or *take-off denied*. If take-off is denied, then UAV001 loops back and re-sends the message, until take-off is OK. When take-off is OK, then UAV001 checks whether power back is required. In case it is, then it performs appropriate operations and checks again. When power back is not required, then it finally performs taxiing and hands off control to an interaction protocol in Take-off state.

In Figure 8, UAV001 moves to take-off state and requests authorisation to take-off. If ATC001 authorises take-off, then UAV001 performs fuel and systems verification. If there is something wrong, then take-off is aborted and a new authorisation is requested; if verification succeeds then UAV001 proceeds to take-off. If ATC001 does not authorise take-off, then UAV001 checks its knowledge base to decide whether to hold take-off or to give up. If decision is to hold take-off, then a new authorisation is requested, otherwise mission is aborted.

In Figure 9, UAV001 performs the transition from take-off to climb, which is itself a transition state towards en route state.

In Figure 10, UAV001 moves to en route state and maintains communication with ATC001 anytime it requests change in cruise level, until it identifies it is time to start descent. When this situation arises, then UAV001 requests permission to start descent. When ATC001 grants permission for descent then UAV001 performs descent and state moves to approach.

In Figure 11, UAV001 moves to approach and maintains communication with ATC001 to request permission to start approach for landing. In case meteorological conditions are not adequate, permission is denied and, depending on what conditions are occurring, appropriate measures are taken before a second attempt to start approach for landing is started. In case meteorological conditions are fine, permission is granted and approach is started. In case some operation does not succeed during approach, UAV001 goes to circling and approach is restarted, otherwise approach is finalised and the entity moves to landing, which is the final state in this mission.

Finally, in Figure 12, UAV moves to landing and attempts to perform landing. If it succeeds, then it goes to taxiing and switches off engines, otherwise it takes-off again.

A design tool to support hazard prevention in these terms must allow the representation of complex systems based on this vocabulary, and the exhaustive simulation of interactions involving entities in a system once an event (or set of events) is highlighted. In the next section we introduce *soft institutions* as an appropriate platform to build one such tool.

5 Soft institutions

We argue that soft institutions can be used as a tool to design and implement sociotechnical systems which is particularly useful for hazard prevention, given that a translation from the diagrammatic language presented in the previous sections to interactions protocols in a soft institution is immediate.

Soft institutions generalise the concept of electronic institutions [4, 5, 10] to provide means to model complex systems comprised by human as well as engineered peers [7]. They have been proposed as an appropriate platform to design and implement sociotechnical systems [2].

Electronic institutions are a powerful framework to build systems comprised by multiple entities based on the principle that the global behaviour of a complex system can be managed by the establishment of norms, rewards for entities that abide by these norms and sanctions for those who challenge them. In order for an entity to participate in an electronic institution, it must be prepared to respond to norms, rewards and sanctions, as well as interact with other participating entities.

Norms, rewards and sanctions in an electronic institution form a *normative system* which should be flexible in order to adjust to the observed behaviour of participating entities in an institution. The normative system dictates the way entities should behave in order to be allowed into an electronic institution and an entity (or organisation comprised by entities) must comply with the normative system in order to be able to request participation in an electronic institution.

Technological entities can be designed and built to comply with normative systems and, therefore, participate in electronic institutions. Human entities, however, may feel uncomfortable to need to learn and then to be submissive to third party rules as a prerequisite to join into a network of peers.

Soft institutions, in contrast, allow entities to act freely and adjust their behaviour in a minimalist way to be able to join into local interaction protocols. Instead of having a centralised control around the normative system (as is the case with electronic institutions), soft institutions have a decentralised, possibly asynchronous control, centered on entities which choose to interact according to available protocols. This way, the barrier to enter a soft institution is significantly lower for humans, hence an interaction platform based on soft institutions can be more appealing to human entities than one based on electronic institutions,

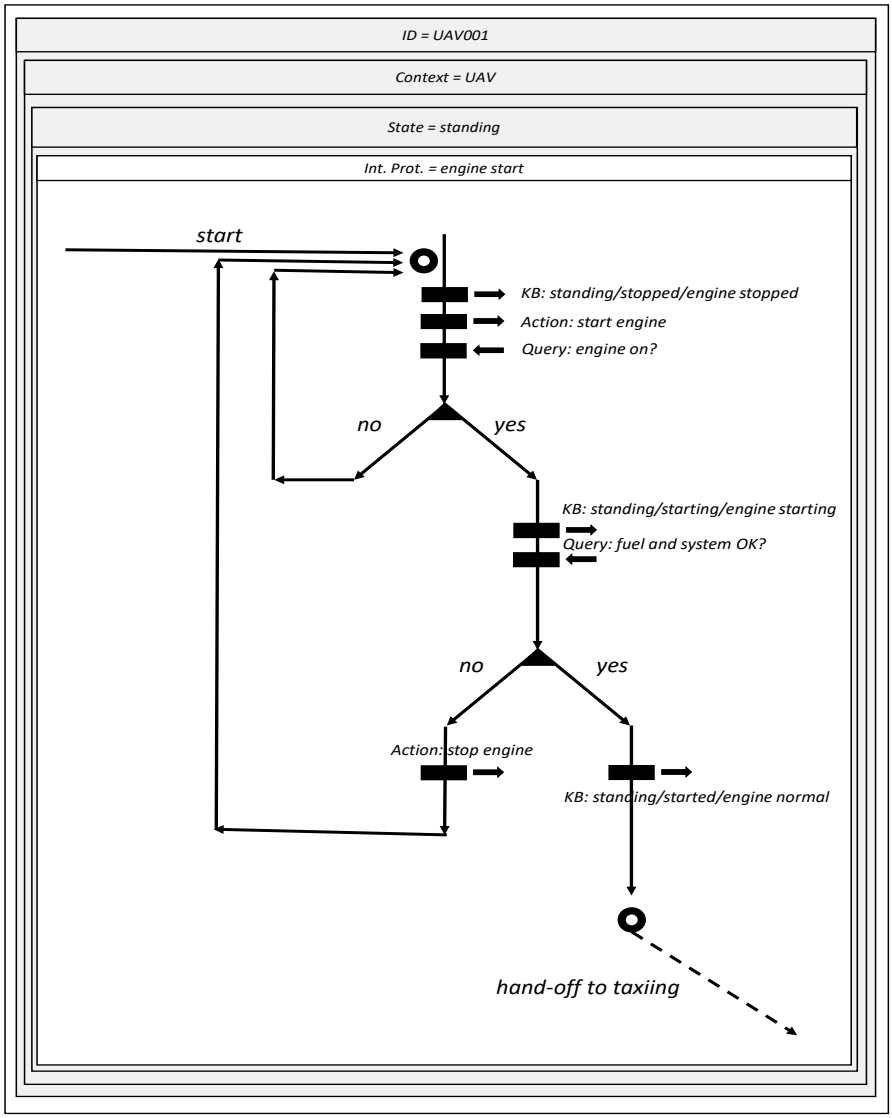

Fig. 6. Interaction protocol for entity in context of UAV and state as Standing

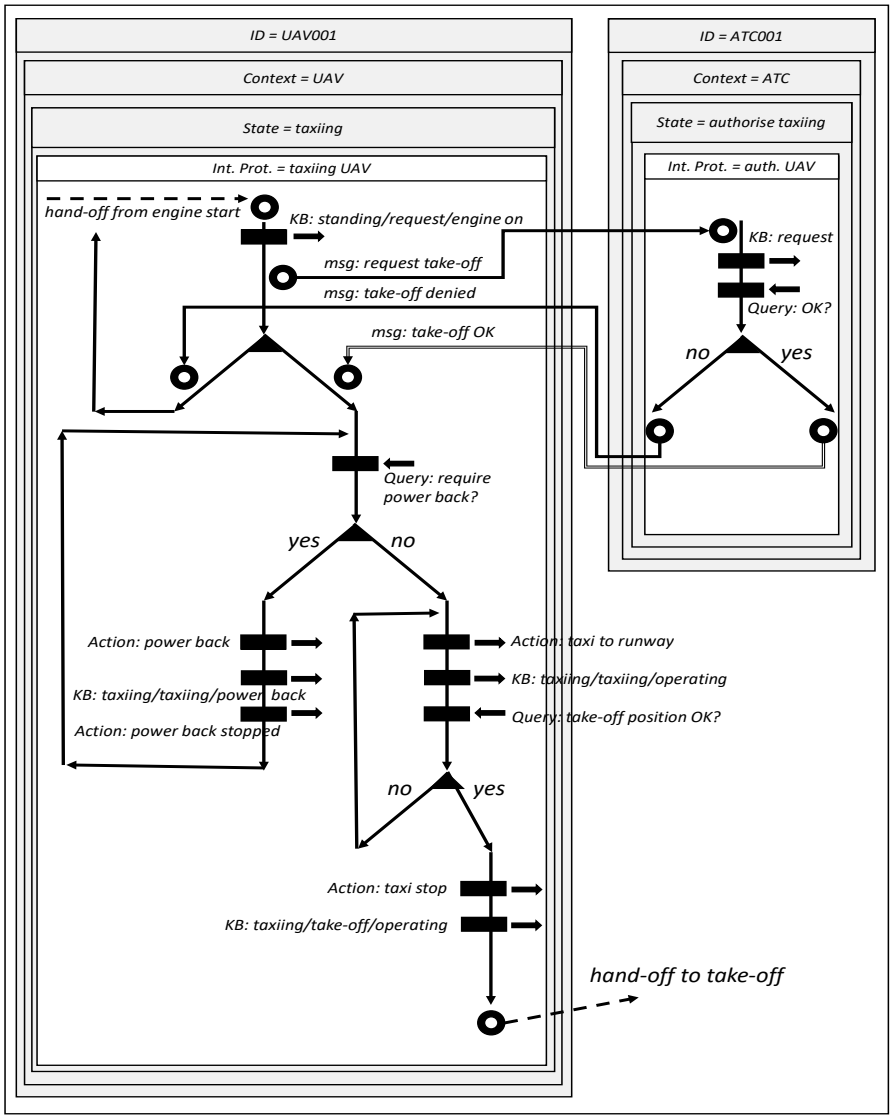

Fig. 7. Interaction protocol for entities as UAV (Taxiing) and ATC (Auth. Taxiing)

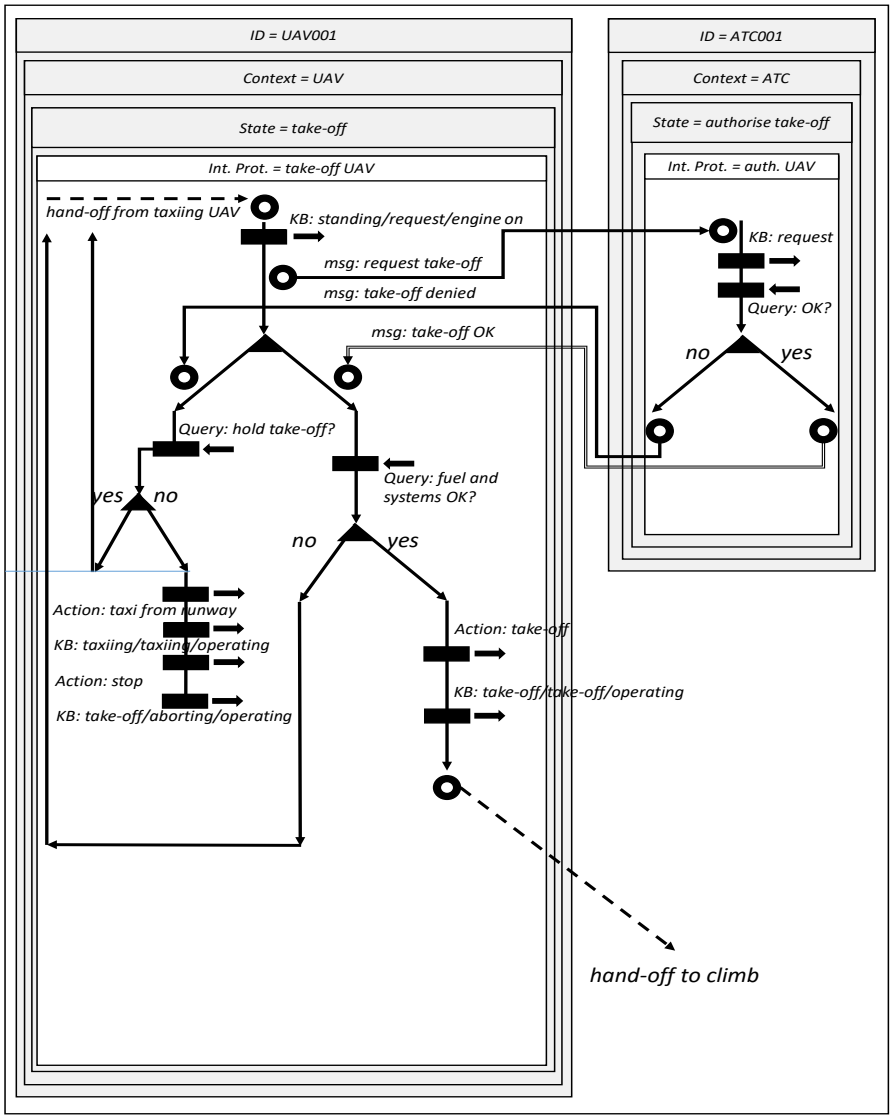

Fig. 8. Interaction protocol for entities as UAV (Take-off) and ATC (Auth. Take-off)

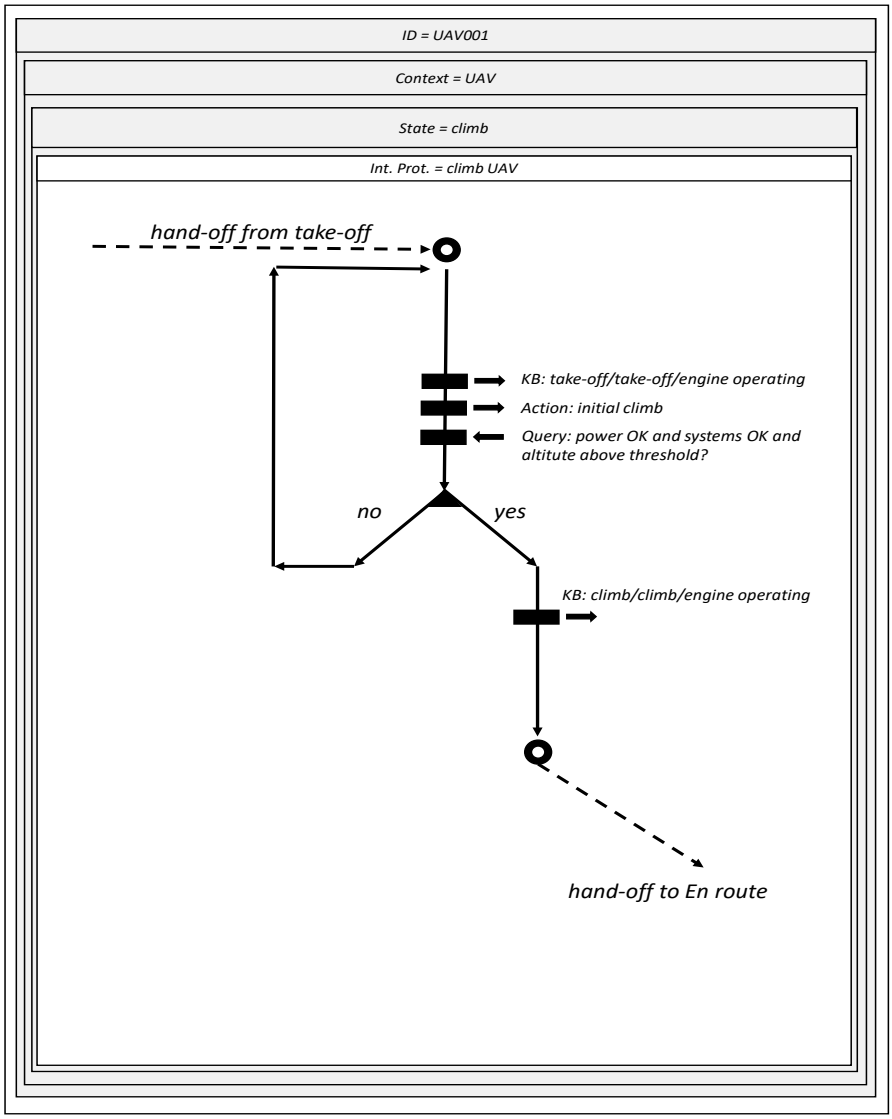

Fig. 9. Interaction protocol for entity as UAV (Climb)

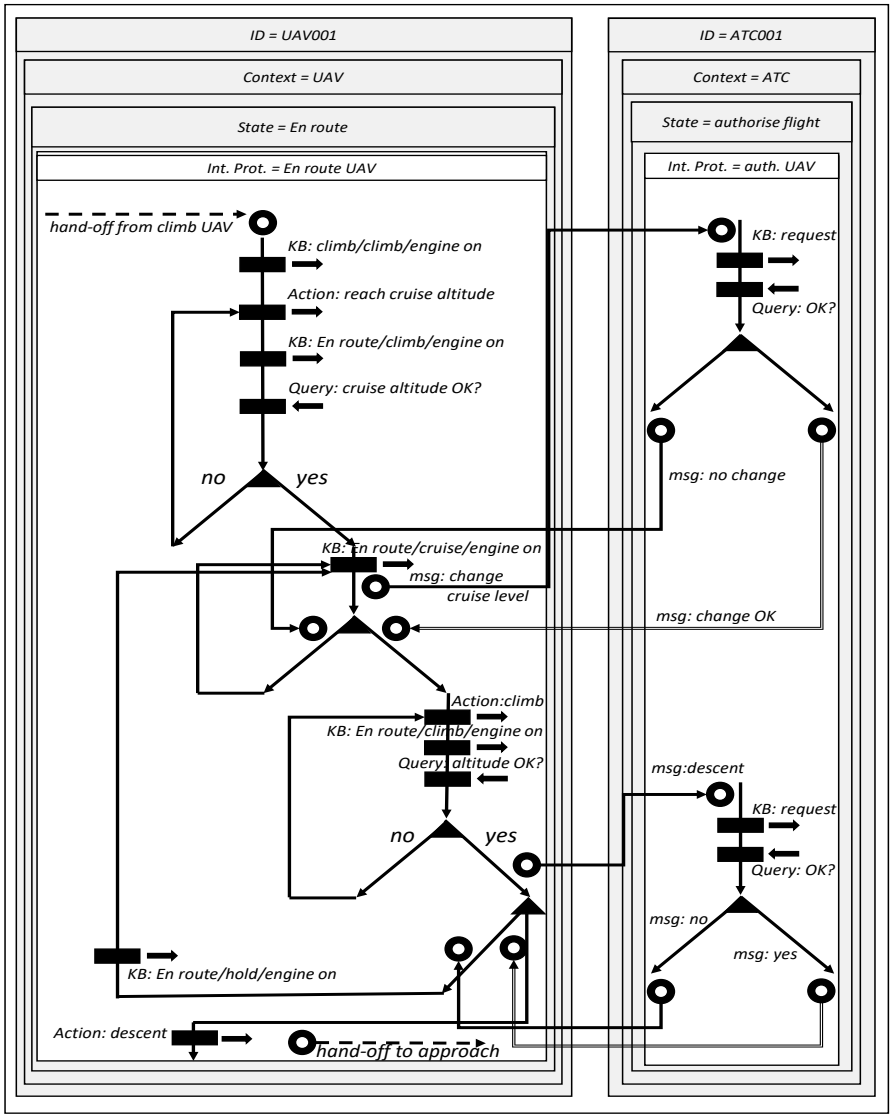

Fig. 10. Interaction protocol for entities as UAV (En route) and ATC (Auth. change level and Auth. descent)

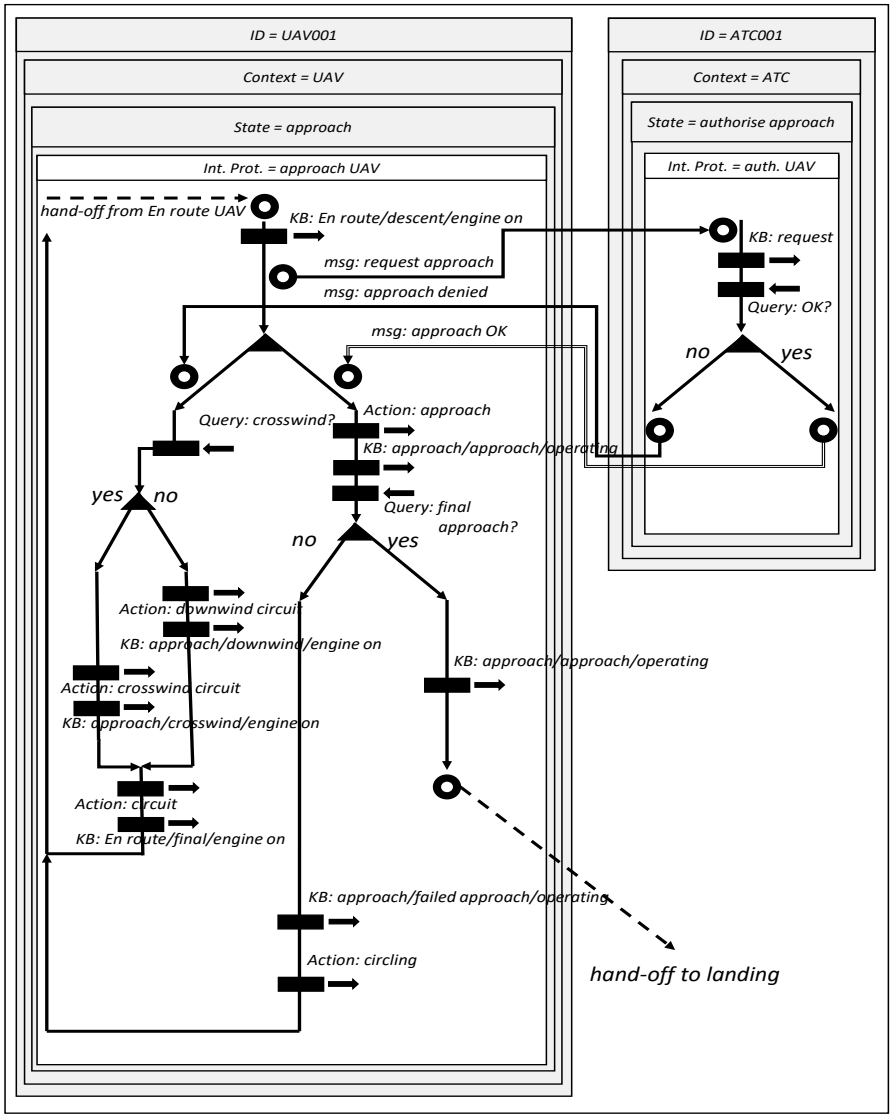

Fig. 11. Interaction protocol for entities as UAV (Approach) and ATC (Auth. approach)

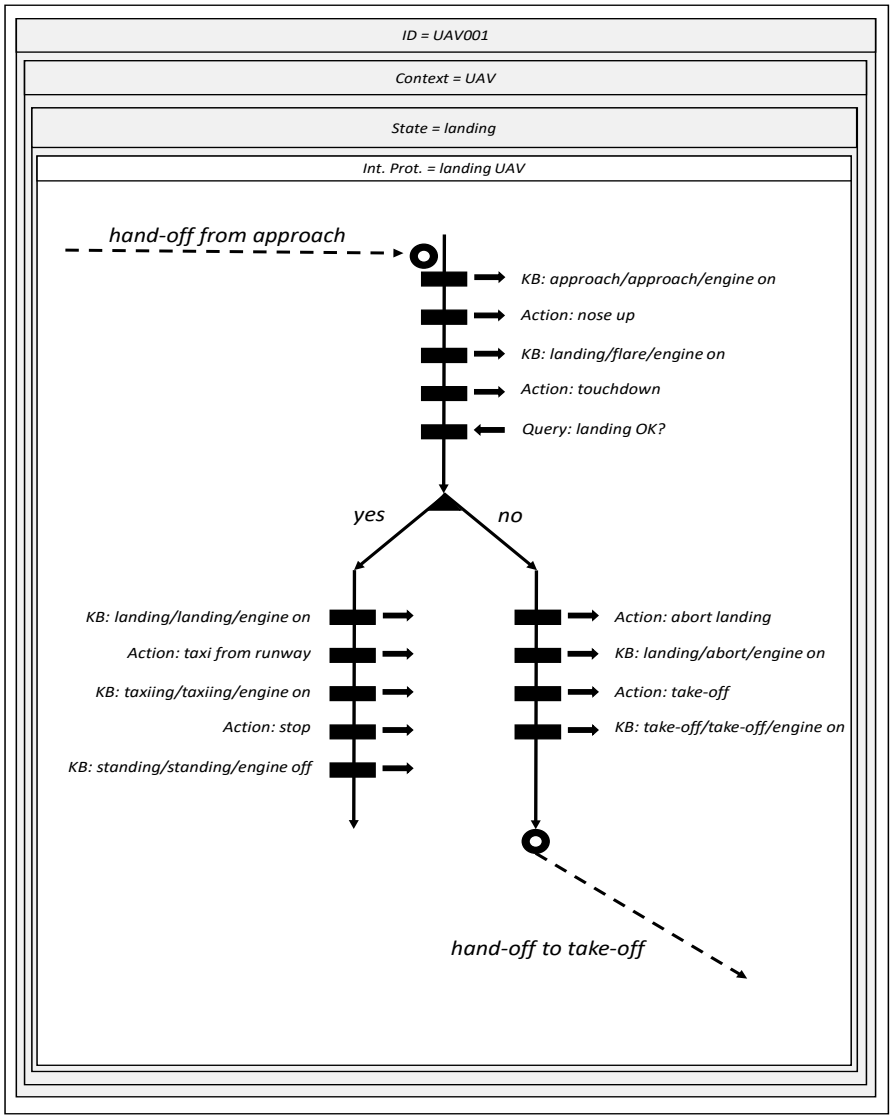

Fig. 12. Interaction protocol for entitiy as UAV (Landing)

at the cost of only being able to have partial control over design, operation and management of a system based on soft institutions.

From the perspective of hazard prevention, soft institutions are a good modeling language for complex sociotechnical systems, well aligned with the strategy for hazard prevention proposed in section 3. Soft institutions also consider as a basic principle that a full account of all states of the systems being modeled is not feasible, hence hazard prevention can only – and at best – be based on relevant hazards as characterised in section 3.

Soft institutions are organised in four layers:

1. **The entity controlled layer**: this layer caters for individual capabilities and actions corresponding to each entity. Entities can be human individuals (e.g. pilots and flight controllers), technological entities (e.g. aircrafts, sensing and communicating devices), or organisations constituted of other entities (e.g. teams of aircrafts flying in formation, teams of controllers);
2. **The communications layer**: this layer comprises the infrastructure and processing power to manage message exchanges between entities. In principle, messaging is peer-to-peer with unique addressing. Additional message control structures can be built using the entity controlled and the communications layer.
3. **The coordination layer**: this layer consists of social norms that constrain and regulate interactions among selected peers (e.g. rules to enter a controlled air space, navigate in it, interact with other entities and leave the air space).
4. **The environment**: this layer comprises all other phenomena that can influence the behaviour and state of the soft institution.

We assume a language $\mathcal{L}$ used to describe facts and computational expressions. The language consists of three constructs:

1. **Terms**: correspond to constant or atomic expressions of different types;
2. **Variables**: are uniquely identified strings to which different values can be assigned;
3. **Functions**: are collections of mappings from tuples of terms to terms.

Value assignments to variables are expressed as substitutions σ of the form $\{x_1 \mapsto c_1, \cdots, x_n \mapsto c_n\}$, which denote that the construct c_i is assigned to the variable x_i. A substitution application function $\hat{\sigma}$ is applied to whole constructs, producing a new construct in which every variable in a construct c that is present in a substitution σ is replaced by the corresponding construct. For example, if $\sigma = \{x_1 \mapsto y, x_2 \mapsto 5\}$ and $c = (x_1 + x_2 + x_3)$, then $\hat{\sigma}c = (y + 5 + x_3)$.

Using substitutions we can naturally define unification ($\stackrel{\hat{\sigma}}{=}$). A substitution application $\hat{\sigma}$ unifies two constructs c_1 and c_2 if the application of $\hat{\sigma}$ to both constructs yields the same result, i.e. $c_1 \stackrel{\hat{\sigma}}{=} c_2$ *iff* $\hat{\sigma}c_1 = \hat{\sigma}c_2$.

Each entity maintains a personal knowledge base that comprises its beliefs, opinions, individual goals, actual knowledge, reasoning capabilities, actions etc. It is assumed, as a design principle, that entities do not have access to each

others' personal knowledge bases. It is also assumed, however, that each entity participating in a soft institution maintains a part of its knowledge base stored as a collection of $\mathcal{L}$ constructs, which we here name *institutional knowledge base*, and which are updated and consulted using two operators:

1. $\mathbb{A}(c)$: this operator updates a fact c (*KB Update* in Figure 4). Depending on specific institutions being designed, an update may correspond to inserting, actual updating or deleting information from the institutional knowledge base;
2. $\mathbb{K}(c, \hat{\sigma})$: this operator consults the institutional knowledge base (*KB Query* in Figure 4). Similar to the $\mathbb{A}$ operator, variations on the semantics of the $\mathbb{K}$ operator can be used for different soft institutions. Essentially, $\mathbb{K}(c, \hat{\sigma})$ checks whether the construct c belongs to the institutional knowledge base of an entity; if it does, then it is retrieved from the knowledge base, and the substitution $\hat{\sigma}$ is used to build the construct $\hat{\sigma}(c)$.

The institutional knowledge base contains a set of ground terms $\mathcal{R} = \{R_1, \cdots, R_m\}$ which represent a set of contexts available to the entity. Contexts are parameterised by states, so that e.g. R_{i/s_j} refers to state s_j in context R_i. It also contains a set of constructs $\mathcal{PROT}$ using the syntax specified in the following paragraphs, which characterise interaction protocols available to the entity given a context and a state.

Given an implementation of a platform for soft institutions, contexts and states are the means for an entity to enter a soft institution: an entity can pick a context and then a state from $\mathcal{R}$, which become the institutional context and state of the entity and grant the entity the right to engage into interactions using an appropriate protocol available in $\mathcal{PROT}$. Contexts and states can be retrieved and updated using the $\mathbb{A}$ and $\mathbb{K}$ operators.

Messages are passed from entity to entity via the communications layer. To each entity is assigned a unique ID, and messages depend upon contexts and states to be properly treated. A message M is assumed to have the format $M = \langle R_{send}, gT, R_{rec}, ID_{other} \rangle$, where R_{send} is the context/state that the sending entity must necessarily hold when the message is sent; gT is a ground term which corresponds to the content of the message; R_{rec} is the context/state that the receiving entity must hold in order for the message to be received; ID_{other} is the ID of the "other" entity: it is the ID of the receiver when a message is being sent and the ID of the sender when a message is being received.

The institutional knowledge base also contains two constructs that represent the state of the entity with respect to the soft institution:

1. *Comm* stores the status of communications. It contains the entity ID and two message queues containing incoming and outgoing messages respectively.
2. *Coord* stores the status of coordination. It contains the list of contexts and states already held by the entity including the current context/state as head of the list, the protocol being followed, the stage of execution of the current protocol and the set of variable assignments/substitutions.

Protocols are defined as a variation and extension of the *Lightweight Coordination Calculus (LCC)* [9] according to the specification presented in Figure 13. Carefully crafted sets of protocols embedded into appropriate states and contexts can implement sophisticated patterns of interaction, servicing large and complex sociotechnical systems. Interaction protocols work as support services for entities to engage into well regulated and carefully designed interactions, but they are not mandatory and they do not necessarily cover all aspects of all interactions that connect entities participating in the same sociotechnical system. System modeling based on soft institutions can be used to highlight facets of a system that are considered most relevant. For hazard prevention, relevant hazards can be characterised in detail and simulations can be performed, so that forward and backward reasoning can be performed and the design of a system can be refined and improved towards resilience with respect to failures.

– *A protocol is a list of clauses. A clause defines a script to be followed in order for an interaction to take place. Clauses have the format* $cl(R, [c_1, \cdots, c_r]) ::= Def$ *where* $R \in \mathcal{R}$ *is a context parameterised by a state,* $c_1, \cdots, c_r$ *are optional parameters and* Def *is the body of the clause:*

$$\begin{aligned} Def &:= Closed \mid Out \mid Out \leftarrow [In_1, \cdots, In_s] \mid Def \ \textbf{\textit{then}} \ Def \mid Def \ \textbf{\textit{or}} \ Def \\ In_i &:= rec(Msg) \mid cond(c) \\ Out &:= Null \mid snd(Msg) \mid chR(R', [c'_1, \cdots, c'_{r'}]) \mid \mathbb{A}(c) \end{aligned}$$

– *Closed concludes an interaction.*
– *Out is an output action:*
  - *Null is an empty action that does nothing.*
  - *snd(Msg) sends message Msg to another entity.*
  - $chR(R', [c'_1, \cdots, c'_{r'}])$ *either changes the context of the entity during the execution of a clause or changes the state of the entity within the same context.*
  - $\mathbb{A}(c)$ *updates the construct c into the institutional knowledge base.*
– $Out \leftarrow [In_1, \cdots, In_s]$ *performs a list of input actions and then performs an output action. An input action* In_i *is one of the following alternatives:*
  - *rec(Msg) receives a message Msg from another entity.*
  - *cond*(c) *checks whether there is a construct* c' *in the institutional knowledge base and a substitution* $\hat{\sigma}$ *such that* $\mathbb{K}(c', \hat{\sigma}) = c$*. The construct c is a condition which can be satisfied if the answer is positive.*
– ***then*** *is a connective that represents sequential and, i.e. it joins two computational steps in sequence.*
– ***or*** *is a connective that represents non-deterministic choice between two computational steps.*

Fig. 13. *Protocols in LCC*

6 Conclusion and future work

In this work we have considered hazard prevention during the design of systems for flight control of autonomous UAVs, based on a diagrammatic language that can be translated to protocols in *soft institutions*.

Implementations of platforms for soft institutions have already been presented elsewhere [7], and frameworks for formal verification of interaction protocols with respect to desired properties have also been developed [8]. In future work, we plan to employ these systems as a platform to support the activities of safety engineers during the design of complex systems, by providing them with tools to identify potential relevant hazards.

References

1. F. Belmonte, W. Schon, L. Heurley, and R. Capel. Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: An application to railway traffic supervision. *Reliability Engineering and System Safety*, 96:237–249, 2011.
2. F. S. Correa da Silva, P. Papapanagiotou, D. Murray-Rust, and D. Robertson. Soft institutions – a platform to design and implement sociotechnical systems (submitted). In *20th International Conference on Knowledge Engineering and Knowledge Management*, Italy, 2016.
3. M. C. Davis, R. Challenger, D. N. W. Jayewardene, and C. W. Clegg. Advancing socio-technical systems thinking: a call for bravery. *Applied Ergonomics*, 45:171–180, 2014.
4. M. Esteva, J. A. Rodriguez-Aguilar, C. Sierra, P. Garcia, and J. L. Arcos. On the formal specification of electronic institutions. In *Agent mediated electronic commerce*, pages 126–147. Springer, 2001.
5. M. Esteva and C. Sierra. *Electronic Institutions: from specification to development.* Consell Superior d'Investigacions Científiques, Institut d'Investigació en Intel· ligència Artificial, 2003.
6. E. Hollnagel. A tale of two safeties. *Nuclear Safety and Simulation*, 2013.
7. D. Murray-Rust, P. Papapanagiotou, and D. Robertson. Softening electronic institutions to support natural interaction. *Human Computation*, 2(2), 2015.
8. P. Papapanagiotou, D. Murray-Rust, and D. Robertson. Evolution of the lightweight coordination calculus using formal analysis. *Personal communication*, 2016.
9. D. Robertson. *Multi-agent coordination as distributed logic programming*, pages 416–430. Proceedings 20th International Conference on Logic Programming – Springer LNCS 3132. 2004.
10. C. Sierra, J. A. Rodriguez-Aguilar, P. Noriega, M. Esteva, and J. L. Arcos. Engineering multi-agent systems as electronic institutions. *European Journal for the Informatics Professional*, 4(4):33–39, 2004.
11. E. Trist. The evolution of socio-technical systems. *Occasional paper*, 2:1981, 1981.