# 1 Combining Narrative-Based Representations and Model-Checking for Robust Supervised Autonomy

## 1.1 Objectives and Expected Significance

## 1.2 Technical Approach

### 1.2.1 Narrative

The explicit representation of narrative in emerging systems (http://www.cc.gatech.edu/ riedl/pubs/aimag.pdf) produce interactive systems that are engaging to humans and compatible with the ways humans understand and predict complex behaviors. Moreover, a narrative framework is a natural way to consider how the activities of one agent constrain or influence the activities of another, allowing a formalism for analyzing how different roles affect a mission. Importantly, these explicit representations are compatible with a suite of algorithms from computer science.

**Mike: Heres what is said in the project description.**

At this time, NASA is specifically seeking techniques that can handle the stochastic and complex nature of many elements (including automated systems, humans, environment conditions) in the NAS, especially the unpredictable nature of human responses and the safety aspects involved in trying to game the system. NASA is seeking analysis techniques that complement traditional human factor studies by addressing off-nominal conditions and coping with variability.

Discuss how narrative is a useful metaphor for describing things that are inherently based in a network. Swarming and social analytics-based techniques to the NAS problem assume scales that are too large for the types of networks involved, and deterministic centralized controllers assume scales that are too small. Pose the problem of designing and managing a networked team with information passing among agents. State that the objective is to produce dynamics that induce acceptable attractors in the system, where the attractors are an efficient distribution over space and time of vehicles and operator workload.

Discuss how individuals in the distributed network will not be able to (nor interested in) knowing everything about the entire network, but rather need an appropriate level of abstraction for things happening beyond their immediate neighbors in the network. Narrative is the way that we create this abstraction, including forming a probabilistic mapping between a humans or autonomous agents actions and consequences for other actors in the narrative.

Describe what can be done with deterministic approaches and how non-deterministic approaches allow us more flexibility in narrative abstraction provided that we know how to act within these narratives. Then, introduce the principles of persistence as a means for exerting appropriate influence for important needs, and negotiation as a means for reaching Pareto efficient solutions given the rules of networked interactions.

Present the model as one with underlying game theory principles encoded in as a mechanism design problem, state the limits of algorithmic approaches to solving this problem, and then introduce the need for appropriate (narrative) metaphors that will allow humans to form accurate and useful mental models for making choices within the system. These same metaphors will serve as the basis for helping autonomy to know what (and at what level of abstraction) to communicate with humans in the system.

### 1.2.2 Formal Modeling

Three elements of of narrative-based systems enable robust human-machine interaction: First, there exist good techniques for representing narratives, including automata-based representations for encoding operator intent. Second, a narrative can be represented as a trajectory through a state space of possible situations. A mission is a planned trajectory through a state space, with agent choices and environment input alternative trajectories. Third, narrative-based systems support flexible allocations of authority and autonomy, thus allowing distributed humans and machines to robustly perform a mission.

We propose two classes of models. First, we propose to use (deterministic) timed automata to explicitly represent the set of possible behaviors of agents in the system. This builds on our previous work (Modeling UASs for Role Fusion and Human Machine Interface Optimization; Gledhill, Mercer, & Goodrich, Proc of IEEE Conf on Sys, Man, and Cybernetics, 2013). These timed automata are appropriate for human-machine interaction because they implicitly represent the set of afforded behaviors of the team.

Second, we propose to use Markov chains to represent likely outcomes of agent behaviors. Systems that operate in the real environment with real humans must be robust to deviations from a deterministic plan. Probabilistic models allow us to quantify performance bounds as a function of level of uncertainty and to inform a human operator of the level of persistence required to produce a robust outcome (Abstraction and Persistence: Macro-Level Guarantees of Collective Bio-Inspired Teams under Human Supervision; Goodrich & Mercer, Proc of Infotech@Aerospace, 2012).

Model checking is particularly effective in isolating violations of system-level properties. A user posing "what-if" scenarios is able to assess possible outcomes and quickly isolate trajectories that enter high-risk, high workload, or failure situations. The proposed verification approaches will (a) use symbolic execution and SMT technology to manage state explosion and (b) leverage advances in probabilistic analysis to express bounds on high-risk trajectories. This yields the ability to detect and predict problems, with guidance on how to mediate these problems.

**Eric, I think that the key elements for your writing derive from the following excerpts from the call:**

The focus should be on human-human as well as human-automation interactions and the exploration of off-nominal scenarios that can emerge from these interactions. It is important to take into account the stochastic elements of the NAS. For example, the analyses of interest could include the following, non-exhaustive list of off-nominal scenarios:

- Misunderstandings due to mixed equipage
- Sudden deviations from planned trajectories due to exceptional circumstances such as unexpected pilot actions or unexpected critical hardware or software failures
- Safety issues resulting from overly competitive behaviors or from gamesmanship.

In particular this solicitation seeks techniques that can handle stochastic events and emerging behaviors. The techniques of interest include:

- Game theory: in particular for modeling human aspects involved in decision making between humans, but in the context of interaction with some automation.
- Bayesian analysis: for modeling probability propagation throughout a graph of modeling interactions between human and automated agents.
- Learning and coordination in multi-agent systems: for capturing the evolution of human behaviors over time and especially the learning of specific situations and the optimization or simplification of procedures in air traffic that can lead to confusion for other agents

The idea would be to state what can and cant be done using existing techniques in model-checking, leading to the idea of extramodel information and the need for constrained negotiation. State what is unique about our approach.

#### 1.2.3 Summaries and Abstractions

Narrative-based representations provide a comprehensible and flexible way to explore agent behaviors and interactions with their environment over time and from different perspectives of inquiry. We propose to use narrative-based performance summaries, building on work by the proposers for remote supervision of NASA planetary robots (NASA Phase II STTR. Anytime Summarization for Remote Robot Operations, Schreckenghost/TRACLabs). These summaries orient personnel quickly about the performance of remote agents performing complex tasks with variable levels of autonomy. Summary measures identify what progress the agents have made and, when progress is impeded, indicate what went wrong. Trending measures determine how well agent assets and resources are being utilized, identify opportunities to improve agent performance, and detect impacts to performance resulting from degraded agent capabilities. Key to our approach is the ability to summarize important differences between actual performance and performance expectations.

For the NAS we will investigate performance-based abstractions to support narrative interaction about the consequences of anomalous events involving Unmanned Aircraft Systems (UAS) and the actions taken in response to these events. Anomalous events with consequences can include (1) system failure, (2) human error, (3) automation error, (4) human threat (such as terrorism), (5) environmental threat (e.g., weather), and (6) infrastructure problem (e.g., unavailable runway). The abstraction of consequences can be summative, describing how we have reached the current situation, or projective, predicting what is expected should mitigating actions be taken. These abstractions will support Air Traffic Controllers (ATCs) and pilots in choosing both what action to take and when to take action. These abstractions also will define the concepts of discourse and the importance of these concepts when the UAS (both piloted and autonomous) interacts with the ATC and pilot. Because a single anomalous event can affect multiple UASs and crewed aircraft, it is necessary to support abstraction for different participant roles (e.g., the UASs, the ATCs, and the pilots). It also is necessary to support abstraction for different levels of UAS autonomy, because these differences in autonomy change the roles and responsibilities of both the pilot and the UAS.

Performance measures will abstract consequences for each UAS (referred to as agent below) as the impacts of anomalous events and their mitigating actions from the following perspectives:

- Mission/flight Success: Are critical milestones in the flight plan achieved as expected? If not, how significant is the deviation?
- Margin or Buffer: Are margins in altitude, vehicle separation, and fuel level decreasing at an exceptional rate?
- Utilization of Airspace and Runways: Does the density and separation of agents and other vehicles in the airspace approach acceptable limits? Are exclusion zones in the airspace threatened? Does time between takeoff and landing for each runway approach safety limits?
- Vehicle Functionality and Integrity: Do events or actions affect critical functionality needed to follow the plan? If so, are alternative capabilities available?
- Human Safety: Do events or actions increase risk of collision in air or on landing? Because of shared airspace, runways, and other airport facilities and services, events affecting one agent can have consequences on other agents. We will propagate the impacts of events and mitigating actions affecting one agent to other agents in the airspace or airports.

We will compare the performance of each agent to expected or desired performance (e.g., flight plan, vehicle separation, fuel level, etc.). These expectations can change as the situation changes. When performance deviates from expectations, the consequences of the deviation and their importance will be predicted. These consequences can include delayed landing, increased backlog at airport, and increased threat of in-air collision, or unsafe takeoff or landing. We also will support predicting performance improvements when mitigating action is taken. Algorithms for abstracting consequences using performance measures will combine data about active flights with model-based predictions and expectations based on acceptable flight performance.

### 1.2.4 Motivating Example

An Unmanned Aircraft Systems (UAS) moving along the periphery of an airport suddenly deviates into the controlled airspace. This might be due to human error, automation error, or system failure. The ATC informs the UAS it is in violation of the airspace and directs it to resume nominal flight path. Concurrently the ATC must assess the immediate impact of this event on all aircraft in the area near the UAS. This includes such impacts as potential separation violations and delays that could adversely reduce fuel margins. If the UAS cannot resume its nominal flight path, the ATC must now recommend mitigating actions to the UAS. These recommendations must take into account any loss of functionality sustained by the UAS. The impact of these mitigating actions must be assessed on both the UAS and the other aircraft in the area. These impacts must be projected forward in time and through the airspace. The ATC also must work with other ATCs to assess the impact of the anomalous event and mitigating actions on the aircraft managed by them. Tools and techniques are needed to help ATCs understand and predict UAS behaviors and their impacts on each other and on crewed aircraft. These techniques should support the evaluation and comparison of alternative actions taken to mitigate adverse impacts.

## 1.3 Impact

## 1.4 Relevance to NASA

## 1.5 General Work Plan

### 1.5.1 Collaborative Plans

### 1.5.2 Milestones and Accomplishments

## 1.6 Data Sharing Plan

### 1.7 References and Citations

## References

### 1.8 Biographical Sketches

## 1.9 Current and Pending Support