An Approach to Design for Safety in Complex Systems

Nicolas Dulac, Nancy Leveson
MIT, Cambridge, MA, U.S.A.
{ndulac, leveson}@mit.edu

Abstract

Most traditional hazard analysis techniques rely on discrete failure events that do not adequately handle software intensive systems or system accidents resulting from dysfunctional interactions between system components. This paper demonstrates a methodology where a hazard analysis based on the STAMP accident model is performed together with the system development process to design for safety in a complex system. Unlike traditional hazard analyses, this approach considers system accidents, organizational factors, and the dynamics of complex systems. The analysis is refined as the system design progresses and produces safety-related information to help systems engineers in making design decisions for complex safety-critical systems. The preliminary design of a Space Shuttle Thermal Tile Processing System is used to demonstrate the approach.

I. Introduction

As the complexity of engineered systems increases, hazard analysis techniques have continued to lag behind the state-of-the-art engineering practice. Traditional event-based analyses consist of identifying the discrete failure events that could lead the system to a hazardous state. These events are usually organized into causal chains or trees. Popular event-based hazard analysis techniques include Fault Tree Analysis (FTA) and Failure Modes and Effects Criticality Analysis (FMECA). Because of their reliance on discrete failure events, neither of these techniques adequately handles software or system accidents that result from dysfunctional interactions between system components. A system accident occurs as a result of unplanned or unexpected interactions between system components. For example, the loss of the Mars Polar Lander occurred because the designers did not take into account a particular interaction between the thruster's software controller and the mechanical leg deployment. When the legs deployed, a spurious signal was interpreted by the controller as a sign that the lander had reached the Martian surface. The controller shut down the thrusters while the lander was still 50 feet above the ground, causing the spacecraft to crash into the surface. Techniques that consider only failure events will miss these types of hazards.

Furthermore, systems evolve in order to accomplish changing objectives and adapt to environmental pressures and disturbances. Often times, accidents in complex systems involve the migration of the system toward an unsafe or unstable state where small deviations can cascade into catastrophes (Rasmussen, 1997). The foundation for an accident is often laid years before. Once the system has reached an unsafe state, a single event may trigger the loss, but if a particular event does not occur, another one will. Capturing the system's dynamics through models helps in understanding this adaptation of complex system. The system design must ensure that the safety constraints continue to be enforced as changes occur throughout the life of the system. Since traditional hazard analysis techniques view systems as static, they inevitably miss hazards resulting from a migration of the system to an unsafe state.

Hazard analyses are often performed on existing systems or very late in the design process when significant design changes are either impossible or too costly to implement. Although it can be useful to perform a hazard analysis at any stage of the system lifecycle, it is believed that most benefits will result from performing a hazard analysis from the very beginning of the system development process, and throughout implementation and operation. The resulting design for safety methodology is an iterative process where the hazard analysis influences design decisions, and is refined as the design evolves and as more information becomes available.

A new hazard analysis based on the STAMP model (Leveson, 2004) defines accidents as the result of inadequate control actions not enforcing necessary constraints on the system design and operation. Using this approach, the role of the system safety engineer is to identify the constraints necessary to maintain safety and to ensure that the constraints will be enforced during the system design and operation. In this paper, STAMP will be used to demonstrate a hazard analysis during the preliminary design of a shuttle Thermal Tile Processing System.

II. A Brief Description of STAMP

STAMP, which stands for Systems Theoretic Accident Modeling and Process, is an approach to accident modeling in which accidents are conceived as resulting not from component failures, but from inadequate control or enforcement of safety-related constraints on the design, development and operation of the system (Leveson, 2004). In other words, the most basic concept in STAMP is not an event, but a constraint. Safety is viewed as a control problem and accidents occur when component failures, external disturbances, and/or dysfunctional interactions among system components are not adequately handled

In the Space Shuttle *Challenger* accident, the O-rings did not adequately control propellant gas release by sealing a tiny gap in the field joint. The control structure of the system itself must be examined to determine why the controls were inadequate to maintain the constraints on safe behavior and why the events occurred. Why were the hot air gases not controlled by the O-rings in the *Challenger* field joints? Why did the designers arrive at an unsafe design? And why were management decisions made to launch despite warnings that it might not be safe to do so?

Preventing accidents requires designing a safety control structure that will enforce the necessary constraints on system development and operations. The scope of the control structure must be large enough as to include all the factors that influence the system's development and operation. Factors such as reduced management oversight or excessive schedule pressure can push a system toward an unsafe state as readily as a faulty piece of hardware. Consequently, the control structure designed must include the technical aspects of the process controlled, as well as the oversight and management factors that affect this process.

Systems are viewed, in this approach, as interrelated components that are kept in a state of dynamic equilibrium by feedback loops of information and control. A system is not treated as a static design, but as a dynamic process that is continually adapting to achieve its ends and to react to changes in itself and its environment. The design must not only enforce appropriate constraints on behavior to ensure safe operation, but it must continue to operate safely as changes and adaptations occur over time.

Instead of viewing accidents as the result of an initiating (root cause) event in a series of events leading to a loss, accidents are viewed as resulting from interactions among components that violate the system safety constraints. The control processes that enforce these constraints must limit system behavior to the safe changes and adaptations implied by the constraints.

System accidents may also involve inadequate coordination among several controllers and decision makers, including unexpected side effects of decisions or actions or conflicting control actions. While decision-makers usually make decisions that are "locally" rational, when taken into the context of the entire systems operation, these decisions and actions may interact in unexpected ways to produce an accident. Accidents are most likely in boundary areas or in overlapping areas of control. In both boundary and overlap areas, the potential for ambiguity and for conflicts among independently made decisions exists.

Unlike traditional hazard analysis techniques, STAMP takes into consideration the technical, organizational, dynamic factors that characterize today's complex systems. STAMP has been applied in accident analysis using as an example a friendly fire Blackhawk helicopter shootdown over Northern Iraq (Leveson, 2002), and a water contamination accident in Ontario, Canada (Leveson, 2003a). STAMP has also been used to perform a hazard analysis on an existing aircraft collision avoidance system (Leveson, 2003b).

III. Thermal Tile Processing System

In this paper, a hazard analysis technique based on STAMP is described using a Space Shuttle Thermal Tile Processing System (TTPS) as an example. The idea for this example comes from an experimental design for a similar system built at CMU (Dowling, 1992). The TTPS is designed to service the thermal protection tiles on the belly of the space shuttle. The tiles require extensive waterproofing because their tendency to absorb large amounts of water creates substantial weight and balance problems that affect the shuttle's launch and orbit capacities. The TTPS must also inspect each tile for potential damage caused during launch, reentry, and/or transport

The thermal tile maintenance activities take place at the Orbiter Processing Facility (OPF) at NASA's Kennedy Space Center. Except for the jack stands that support the orbiters, the floorspace directly beneath the orbiter is initially clear but substantial structure may surround the orbiter. The work areas of the Orbiter Processing Facility can be crowded at times. The TTPS must negotiate jackstands, columns, workstands, and cables as well as hanging cords, clamps and hoses. The layout of the OPF imposes constraints on the size of the TTPS. The TTPS must be smaller than 1.1m by 2.5m, and not more than 1.75m tall in order to clear the structural beams. The TTPS must operate in a crowded environment while being able to reach tiles with a height ranging from 2.9m to 4m. These requirements can be fulfilled by the use of a robot consisting of a mobile base with a manipulator arm that can reach the thermal tiles and that is fitted with the required tile servicing equipment. The next section provides a sample step-by-step walkthrough of a STAMP-based hazard analysis for the TTPS system.

IV. Hazard Analysis using STAMP

The objectives of a STAMP hazard analysis are the same as that of a traditional hazard analysis. The first general goal is to identify the system hazards and the related safety constraints necessary to ensure acceptable risk. The second general goal is to accumulate some information about how the safety constraints may be violated and use this information to eliminate, reduce and control hazards in the system design and operation. A STAMP hazard analysis includes five steps. The first two steps are similar to those performed during a traditional hazard analysis. The later steps either deviate from traditional practice or provide a guiding framework for doing what is traditionally done in an ad hoc manner.

Step 1: Identify the system hazards

Like traditional hazard analysis processes, identifying system hazards is the first step in performing a STAMP analysis. As the design and analysis are performed, the list of hazards may be modified to include new hazards, modify existing hazards, or eliminate irrelevant hazards. The hazard analysis has to be augmented to take into account the new hazards identified.

Given the initial, high-level requirements of a system, a preliminary hazard identification is performed in order to initially identify hazardous situations that could be encountered during the system's operation. A hazard is a state or set of conditions that, together with other conditions in the environment, will lead to an accident, or loss event. Hazards are not failures, but a set of conditions that may lead to a failure. For example, the overheating of the space shuttle Columbia's internal wing structure was a hazard; the Columbia accident was loss of the seven astronauts and the shuttle itself.

Different initial configurations could be chosen that would introduce different hazards. For this example, the initial configuration chosen consists of a robot including a mobile base and a manipulator arm. A preliminary hazard identification is performed in order to initially identify hazardous situations that could be encountered during the system's operation. For the mobile robot configuration of the TTPS, seven system-level hazards have been identified in the preliminary hazard analysis phase:

1- Violation of minimum separation between the mobile base and external objects
2- Robot base becomes unstable
3- Manipulator arm hits an external object
4- Damage to the mobile robot caused by robot components or failure
5- Fire or explosion
6- Contact of human with DMES
7- Inadequate shuttle thermal protection

Once the preliminary hazards have been identified, an approach to design for safety includes several strategies for mitigating the possible effects of the hazards:

a. Complete removal of the hazard from the design,
b. Reduction of the probability that the hazard will occur,
c. Reduction of the hazard's negative impact, and

d. Implementation of contingency plans in the case that the hazard does occur.

Using hazard #6 of the TTPS as an example (Contact of human with DMES), an approach to designing for safety can be described using the following four types of hazard mitigation strategies. (1) The first strategy is to design the system in order to completely remove the possibility of the hazard occurring. This could involve the use of a different waterproofing chemical that is safe for humans. If it is not possible to completely remove the hazard, (2) the second mitigation strategy is to reduce the possibility of the hazard occurrence. This could involve using sensors to ensure that humans are not present in the work area whenever DMES is injected, as well as visual and aural alerts to inform humans to exit the area because dangerous chemicals will be injected. (3) A third strategy is to reduce the negative consequences of the hazard occurring. This could involve a vacuum system that reduces the concentration of DMES in order to lessen the effect of chemical exposure on humans in the work area. (4) A fourth strategy is to implement a set of contingency plans in order to react in an effective and timely manner to a hazard occurrence. This could include planning for emergency response and adequate healthcare in the case of human exposure to DMES.

In this paper, we demonstrate a STAMP hazard analysis using hazard #3 (robot base becomes unstable) of the TTPS system. Depending on the physical environment of the system at the time when the hazard occurs, the consequences could range from damage to the TTPS and/or orbiter to serious injuries and even death. A complete hazard analysis would begin by looking for design concepts that would completely remove the hazard from the system. A possible solution to this particular problem would be to make the robot base so heavy that it cannot become unstable, no matter what the position of the manipulator arm is. This solution has many shortcomings. For example, while a very heavy base could stabilize the robot, it would violate other functional requirements such as the need for workers to be able to manually move the robot out of the way in an emergency situation. A very heavy base could also increase the damage caused if the robot hits an outside object (hazard #1). If this solution was chosen, a tradeoff would have to be made between the probability of the base becoming unstable (hazard #3) and the extend of the negative effects resulting from the robot hitting an outside object (hazard #1). The goal of the hazard analysis is to provide information to help guide systems engineers in evaluating tradeoffs between alternative design solutions. Ideally, the hazard analyst should work hand-in-hand with the systems engineer during design. At the very least, the hazard analysis process should be documented and the information should be made available to the systems engineers involved in the design process.

Since a very heavy robot base is neither desirable, nor practical, an alternative solution would be to make the base long and wide so that the moment created by the operation of the manipulator arm would be compensated by the moments created by base supports that are far from the robot's center of mass. A long and wide base could remove the hazard, but may require a violation of the functional requirements associated with facility layouts. It was previously determined through analysis of the requirements that the length of the robot base could not exceed 2.5m and its width could not exceed 1.1m. Given the maximum extension length of the manipulator arm and the weight of the equipment carried, a simple analysis shows that the length of the robot base is sufficient to prevent any longitudinal instability, but the width of the base is clearly not sufficient to prevent lateral instability. Based on these findings, it is not possible to completely remove the hazard probability solely through the size of the mobile base. Other

solutions could be proposed in order to remove the possibility of the base becoming unstable while meeting functional requirements. One solution would be to include lateral stabilizer legs that are deployed in order to prevent lateral instability whenever the manipulator arm is used.

A successful implementation of the first mitigation strategy completely removes the need for any additional mitigation efforts. If the hazards cannot be designed out of the system, as in the example above, the second, third and fourth mitigation strategies can and should be used in combination to both reduce the probability of the hazard occurring and reduce the negative consequences of the hazard occurrence.

Hazard analyses are performed throughout the system development process. As design decisions are made, the analysis is used to evaluate those decisions and their effects on the system safety. Early in the system design phase, little information is known about the functioning of the system, so the STAMP hazard analysis will be general at first, but will be refined through an iterative process as additional information emerges from the system design activities. Performing hazard analysis when the design is complete is too late to significantly improve safety. Safety must be built in the system from the beginning.

Step 2: Identify system-level safety-related requirements and constraints

Once the system hazards associated with a selected initial design configuration (in this example, a robot base with a manipulator arm) have been identified, associated system level requirements and constraints are determined. For the TTPS robot, the hazard and associated safety constraints are:

Hazard:

3. Robot base becomes unstable

Safety Constraints:

3.1 Manipulator arm must move only when stabilizers are fully deployed
3.2 Stabilizer legs must not be retracted until manipulator arm is fully stowed

The requirements and constraints are derived from an analysis of the potential failure modes, dysfunctional interactions or unhandled environmental conditions in the controlled system that could lead to the hazard.

Step 3: Define the basic system control structure

Defining the control structure of the system is the third step in the STAMP-based hazard analysis. A control structure is a representation of the interactions between the various different components of the system. Control structures are hierarchical, consisting of various levels of control that influence the controlled process at the lowest level of the hierarchy. Information and control actions are passed laterally between components at the same level and vertically between components at different levels. Feedback information about the state of the system is provided

so that components at higher levels in the hierarchy are able to control components at lower levels.

Management and organization components play an integral role in the safety control structure, because they impact controller decisions and actions, which in turn impact the controlled process. The control structure changes over time and consequently the feedback received at various levels of the hierarchy also changes. Depending on the complexity of the system and its social setting, the control structure may need to take into consideration the management and organizational components that affect the system's development and operation.

A candidate control structure for the robot operation is provided in Figure 1. Since the system under consideration is relatively simple, only the operational part of the control structure is depicted. As with other steps in the analysis, the control structure will have to be revisited when more information is available about the system design and its operating environment. Because of the safety-critical nature of the tasks performed by the TTPS (improper tile servicing could results in a mission loss), a design decision is made to include a human operator in the control structure in order to supervise the robot during its operation and to perform critical tasks.

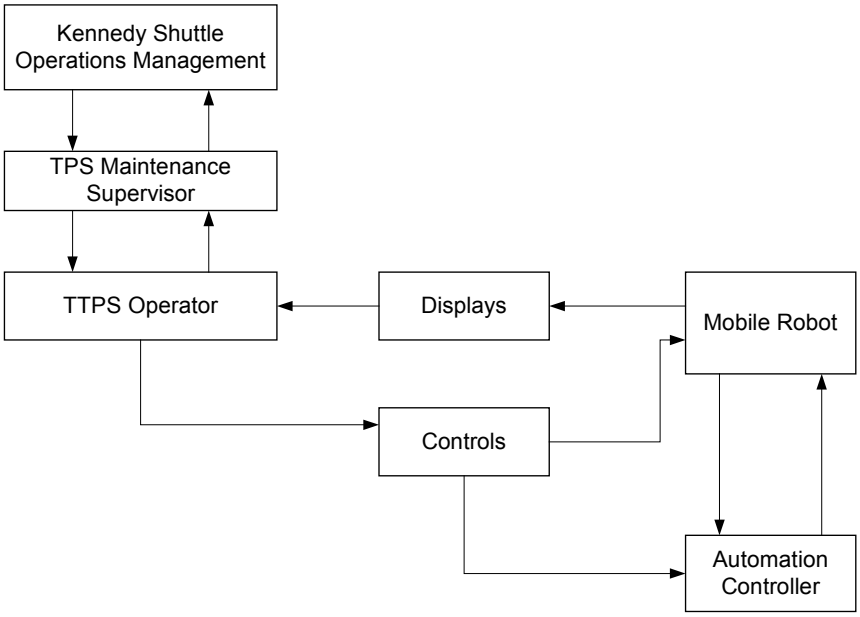

Figure 1: Control structure of the TTPS robot

Step 4: Identify inadequate control actions that could lead to a hazardous system state

The fourth step uses the control structure developed in step 3 to determine how the system could reach a hazardous state. A STAMP model assumes that hazardous states (states that violate safety constraints) result from inadequate control. Consequently, a STAMP hazard analysis starts by identifying these potentially inadequate control actions. Generally, a controller can provide four different types of inadequate control:

1. A required control action is not provided
2. An incorrect or unsafe control action is provided
3. A potentially correct or adequate control action is provided too late or at the wrong time.
4. A correct control action is stopped too soon.

In the context of the TTPS, the safety constraints can be violated if either of four inadequate control actions occurs:

1. The controller does not command a deployment of the stabilizer legs when arm movements are enabled.
2. The controller commands a retraction of the stabilizer legs when the manipulator arm is not stowed.
3. The controller commands a deployment of the stabilizer legs after arm movements are enabled or commands a retraction of the stabilizer legs before the manipulator arm is stowed.
4. The controller commands a retraction of the stabilizer legs but arm movements are enabled before the stabilizer legs are completely extended.

Control actions may be required to handle component failures, environmental changes, or dysfunctional component interactions. In addition, incorrect control actions may cause dysfunctional interactions between components. The inadequate control actions can be restated as constraints on the behavior of the TTPS controller:

1. The controller must ensure that the stabilizer legs are extended whenever arm movements are enabled.
2. The controller must not command a retraction of the stabilizer legs when the manipulator arm is not stowed.
3. The controller must command a deployment of the stabilizer legs before arm movements are enabled and the controller must not command a retraction of the stabilizer legs before the manipulator arm is stowed.
4. The controller must not enable arm movements before the stabilizer legs are completely extended.

Step 5: Determine how the constraints could be violated and attempt to eliminate, prevent and control them in the system design.

In this step, system safety engineers us a top-down approach to identify scenarios in which the safety constraints could be violated. Viewing the controlled process as a control problem provides guidance for engineers to identify the problematic control actions that could result in constraint violations. At this point, a formal modeling language (e.g. SpecTRM-RL) can be used to provide models for the analysis. A continuous simulation environment can be defined at the beginning of the system design process and maintained in order to evaluate design changes as they are proposed.

The presence of humans in the control structure increases the complexity of the control system. Humans do not always follow the normative procedures assumed by the system designers. Human operators will adapt their behavior to be consistent with the reality of their work

environment and informal work systems will emerge as a more efficient way of attaining the conflicting goals of task performance, schedule pressure and resource scarcity (Dekker, 2002). A STAMP-based analysis starts from the hazard, working backwards to identify which type of deviation could lead to it. Human factors experts and system dynamics models (Sterman, 2000) are used to evaluate the type of deviations that should be expected and their impact on the hazard. Using this approach, the system designers would attempt to eliminate the problem through adequate design changes. If the problem cannot be eliminated, the system should be designed in order to either prevent or reduce the likelihood of the deviation. For humans, mitigation strategies may include training, monitoring of procedure deviations, and developing operator's skills at judging when (and when not) and how to adapt procedures to local circumstances (Dekker, 2002).

For the hardware, software and human components of the system, the information provided by the hazard analysis can be used (1) to guide the test and verification procedures (or training for humans), (2) to change the overall system design to provide protection against the hazard, or (3) to add fault tolerant features to protect against the identified hazardous behavior.

In this step of the hazard analysis, the analyst determines how the potentially hazardous control actions can occur and either eliminates them through system design, or mitigates them through design or constraints on the system's operation. While the analysis should start early in order to eliminate as many hazards as possible, more information will be available in later stages of the design and therefore the analysis should be refined and augmented throughout the system development process.

Step 5a: Create the process models for the system components

Whether trying to control a simple inverted pendulum or the operation of a complex space vehicle, controllers (humans and/or automations) need a model of the system to perform effective control actions. The model of the process (*plant*) may require a complex model with a large number of state variables and transitions (such as that needed for air traffic control). Whether the model is embedded in the control logic of an automated controller or in the mental model of a human controller, it must contain the same type of information: the required relationship among the system variables (the control laws), the current state (the current values of the system variables), and the ways the process can change state. This model is used to determine what control actions are needed, and it is updated through various forms of feedback. When the model of the process does not match the actual controlled process, accidents can result. Accidents, particularly system accidents, often result from inconsistencies between the model of the process used by the controllers (both human and automated) and the actual process state. For example, the software controller "thinks" the spacecraft has reached the Martian surface and shuts down the engine.

Figure 2 presents a generic process control loop. The control loop includes both an automated controller and a human supervisor. At this point of the analysis, if the process controlled is well defined, a formal modeling language such as SpecTRM-RL can be used to provide executable models for the analysis. A continuous simulation environment can be defined at the beginning of

the system design process and maintained in order to evaluate design changes as they are proposed.

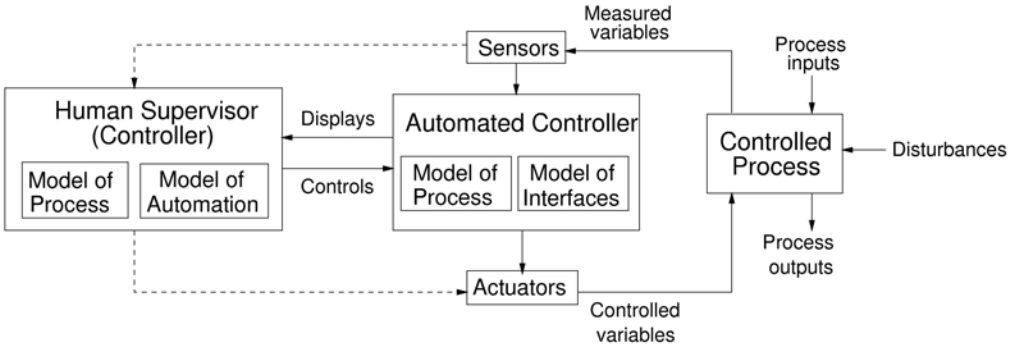

Figure 2. A Generic Process Control Loop

The control structure of the TTSP provided in Figure 1 is augmented with the model of the process under control. Figure 3 shows an overview of the process model for the robot controller. Additions to the process models are required as more information is made available, and as other hazards are considered. The displays, controls, and human operator must also be modeled and include the operator's model of both the automation behavior and the controlled process.

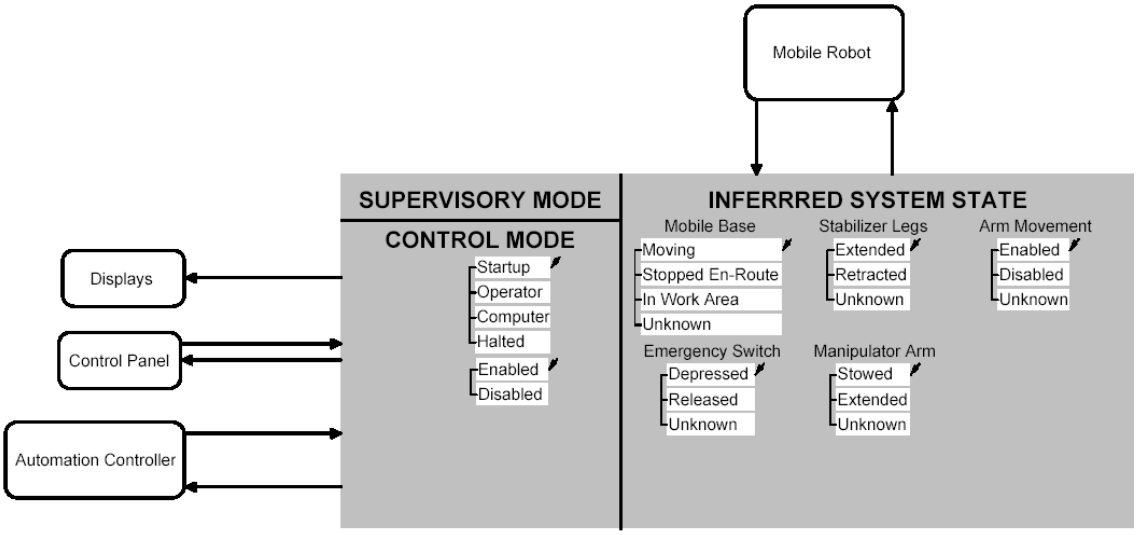

Figure 3: Process model for the TTPS robot controller

Step 5b: For each of the inadequate control actions identified, examine the control loop to determine if it could cause the inadequate control action.

A STAMP hazard analysis uses the formal model of the system in the context of its feedback control loop structure, as well as a set of generic control loop flaws. If accidents are considered as the result of inadequate control and enforcement of safety constraints, then the process that leads to accidents can be understood in terms of flaws in the system development and operation control structure in place during design, implementation and operation of the system. Figure 4 shows the general classification of control flaws that can lead to hazards. These flaws can be used during hazard analysis to identify the required safety constraints.

In each control loop, at each level of the socio-technical control structure, hazardous behavior results from inadequate enforcement of constraints on the process controlled at the level below. Since each component of the control structure may contribute to the inadequate enforcement of safety constraints, a STAMP hazard analysis starts by an examination of each component of the control loop and an evaluation of their potential contribution based on the control flaws classification. A control loop representation of the TTPS robot operation is provided in Figure 5. Using the controls flaw classification (Figure 4) and the TTPS control loop (Figure 5), the following sections provide examples of typical control flaws that could occur during the robot operation.

1 Inadequate Control Actions (enforcement of constraints)

1.1 Design of control algorithm (process) does not enforce constraints

1.2 Process models inconsistent, incomplete, or incorrect

1.2.1 Flaw(s) in creation process or updating process (e.g. asynchronous evolution)

1.2.2 Inadequate or Missing Feedback

- Not provided in system design
- Communication flaw
- Inadequate sensor operation (incorrect or no information provided)

1.2.3 Time lags and measurement inaccuracies not accounted for

1.3 Inadequate coordination among controllers and decision makers (e.g. boundary and overlap areas)

2 Inadequate Execution of Control Action

2.1 Communication flaw

2.2 Inadequate "actuator" operation

2.3 Time Lag

Figure 4: Classification of Control Flaws

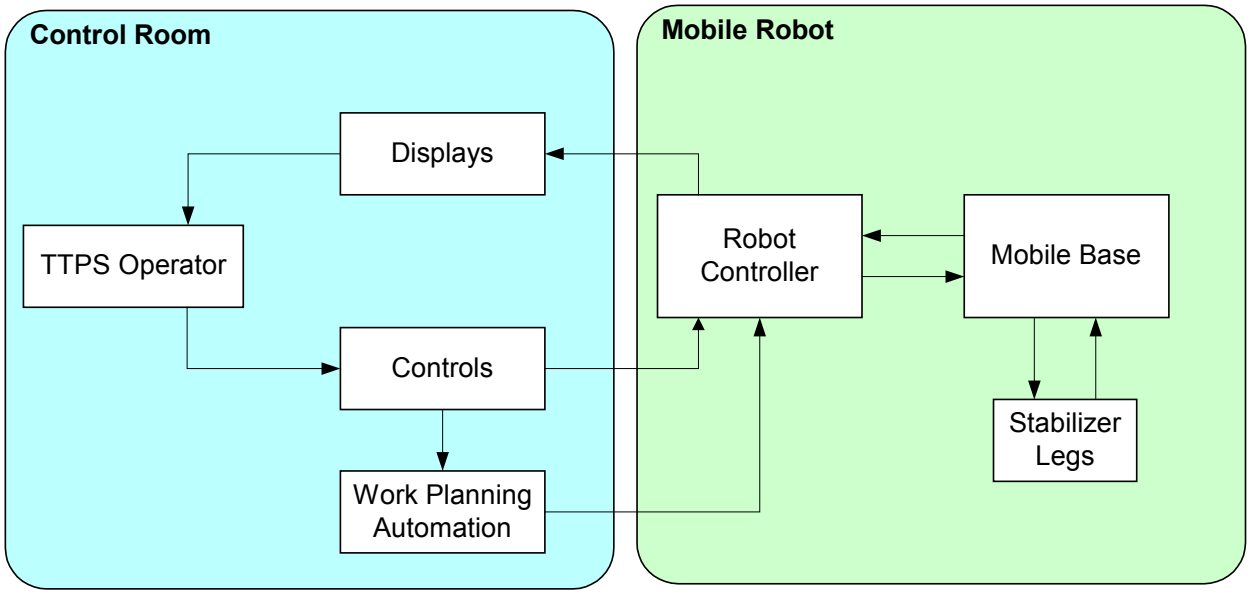

Figure 5: Control Loop of the TTPS Robot Operation

1. Inadequate Control Actions (Enforcement of Constraints) - Inadequate enforcement of safety constraints can occur if the control actions do not properly enforce safety constraints. Control actions may not properly enforce safety constraints if: (1) flawed control algorithms are used, (2) inconsistent or inadequate process models are used by the control algorithms, or (3) inadequate coordination occurs between multiple controllers and decision makers.

1.1 Inadequate Control Algorithms - Formal model simulations and analysis can be used to verify that control algorithms enforce safety constraints. The readability of the formal models also allow for human review. For example, examining the logic that describes the behavior of the robot controller can help identify situations in which the stabilizer legs are *retracted* when arm movements are *enabled*, which constitutes a violation of a safety constraint. In this case, the logic of the arm movement enabling is investigated to ensure that it cannot occur when the stabilizer legs are not extended. Another example associated with inadequate control action #4 (arm movements are enabled before the stabilizer legs are *completely* extended) involves the timing of the transition behavior when the stabilizer legs are extended. The state value of the Stabilizer Legs state variable should transition from retracted to extended only after the legs are physically extended and locked in place. If the state variable transitions before the legs are completely extended, the safety constraint may be violated.

1.2 Inconsistent Process Models - Inconsistencies may occur between the model of the process used by the controllers and the actual process state. For example, if an external object prevents the complete extension of the stabilizer legs, MAPS may still "think" the stabilizer legs are extended because the extensions motors have been powered up. Subsequent movements of the manipulator arm would then violate the safety constraints. Accidents often occur during initial system startup and when restarting the system after a temporary shutdown. For example, the requirements for the mobile robot specify that it must be possible to move the mobile base out of the way in case of an emergency. If the mobile robot goes through an emergency shutdown while servicing the tiles, the stabilizer legs may have to be manually retracted in order to move the robot out of the way. When the robot is restarted, the controller may assume that the

stabilizer legs are still extended and arm movements may be commanded that would violate the safety constraints.

1.3 Inadequate Coordination Among Controllers or Decision Makers - System accidents often result from inadequate coordination among multiple controllers (humans and/or automated) or other decision-makers. Communication flaws play an important role in such inadequate coordination. Accidents are most likely to occur in boundary areas or in overlap areas where two or more controllers control the same process (Leplat, 1987). In boundary and overlap control areas, there is potential for ambiguity, confusion, and/or conflict among independent decisions that could lead the system into a hazardous state. The control structure of the TTPS robot clearly shows that overlapping control may occur between the automated robot controller and the human operator. The human operator is required to handle unexpected situations and has the power to override the automated controller when such situations arise. However, overlapping control of the mobile robot could allow the system to reach a hazardous state. For example, if the robot is servicing tiles in the *computer* mode of operation, a conflict could occur if the human operator inadvertently sends a command for retracting the stabilizer legs. Such additional information would allow the identification of additional safety constraints such as "When operating in computer mode, all operator inputs must be ignored except for emergency stop and explicit commands to switch to Operator Mode."

2. Inadequate Execution of Control Actions - Another way for safety constraints to be violated in the controlled process is if a failure occurs in the execution or transmission of control commands. The control action can be inadequate if: (1) there is a communication flaw to the actuator, (2) the actuator does not perform correctly, or (3) there is significant time lag between the control action signal and its execution. This classification of inadequate execution of control actions applies to automated and human controllers. Clearly, component failures can prevent the adequate execution of control actions. Component failures must be taken into account while designing the system in order to ensure that they cannot result in a violation of safety constraint. In the TTPS example, an actuator failure will not result in the violation of a safety constraint unless the process models become inconsistent as a result of the failure (e.g. the stabilizer legs do not extend because of an actuator failure and arm movements are enabled.).

V. Conclusion

This article described an approach to design for safety for complex systems that involves a hazard analysis process starting early in the system development phase and evolving throughout the system lifecycle. The approach is based on a new model of accident causation called STAMP that views safety as a dynamic control problem. In this context, the hazard analysis is used to provide the safety related information required by systems engineers to make tradeoff decisions during the design process. The technique was illustrated using a safety-critical system of a scope that allowed a walkthrough of the hazard analysis methodology. A single hazard was considered in this paper, and the analysis performed was nowhere near exhaustive. The example used did not include a complex socio-technical control structure. As such, complex modeling of the adaptation mechanisms of the socio-technical control structure was not necessary as it would be in a larger, more complex system. Nevertheless, the analysis offered a demonstration of how the STAMP accident modeling framework can be used to design for safety in complex systems.

VI. Acknowledgements

This work was partially supported by the NASA Engineering for Complex Systems program (NAG2-1843) and NSF ITR grant (CCR-0085829).

VII. References

Dekker, S., *The Field Guide to human Error Investigation*, Ashgate Publishing Limited, 2002.

Dowling, K. *et al,*. "A Mobile Robot System for Ground Servicing Operations on the Space Shuttle", Proceedings of Cooperative Intelligent Robots in Space, SPIE OE/Technology, Boston, MA, 1992.

Leplat, J., Occupational Accident Research and Systems Approach, Jens Rasmussen, Keith Duncan, and Jacques Leplat (eds.) *New Technology and Human Error*, John Wiley & Sons, New York, 1987.

Leveson, N.G., "The Analysis of a Friendly Fire Accident using a System Model of Accidents", International Conference of the System Safety Society, Denver, 2002.

Leveson, N.G., Daouk, M., Dulac, N., and Marais, K., "Applying STAMP in Accident Analysis", Workshop on the Investigation and Reporting of Accidents, 2003(a).

Leveson, N.G., "A New Approach to Hazard Analysis for Complex Systems", International Conference of the System Safety Society, 2003.

Leveson, N.G., "A New Accident Model for Engineering Safer Systems", *Safety Science*, Vol. 42, No 4., 2004.

Rasmussen, J., "Risk Management in a Dynamic Society: A Modelling Problem", *Safety Science*, vol. 27, No. 2/3, 1997.

Sterman, J.D. *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw-Hill/Irwin, 2000.

Biography

Nicolas Dulac is a Ph.D. student at MIT in the Department of Aeronautics and Astronautics. His research interests include system engineering, system safety and information visualization.

Dr. Nancy Leveson is Professor of Aeronautics and Astronautics and also Professor of Engineering Systems at MIT. She is a member of the National Academy of Engineering. Her research interests include system engineering, system safety, human-computer interaction and software engineering.