Modeling and Hazard Analysis Using Stpa

MODELING AND HAZARD ANALYSIS USING STPA

Takuto Ishimatsu[1], Nancy Leveson[1], John Thomas[1], Masa Katahira[2], Yuko Miyamoto[2], Haruka Nakao[3]

[1]Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA
Email: takuto@mit.edu, leveson@mit.edu, jthomas4@mit.edu
[2]Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba-shi, Ibaraki 305-8505, Japan
Email: katahira.masafumi@jaxa.jp, miyamoto.yuko@jaxa.jp
[3]Japan Manned Space Systems Corporation, Urban Bldg., 1-1-26, Kawaguchi, Tsuchiura, Ibaraki 300-0033, Japan
Email: haruka@jamss.co.jp

ABSTRACT

A joint research project between MIT and JAXA/JAMSS is investigating the application of a new hazard analysis to the system and software in the HTV. Traditional hazard analysis focuses on component failures but software does not fail in this way. Software most often contributes to accidents by commanding the spacecraft into an unsafe state (e.g., turning off the descent engines prematurely) or by not issuing required commands. That makes the standard hazard analysis techniques of limited usefulness on software-intensive systems, which describes most spacecraft built today.

STPA is a new hazard analysis technique based on systems theory rather than reliability theory. It treats safety as a control problem rather than a failure problem. The goal of STPA, which is to create a set of scenarios that can lead to a hazard, is the same as FTA but STPA includes a broader set of potential scenarios including those in which no failures occur but the problems arise due to unsafe and unintended interactions among the system components. STPA also provides more guidance to the analysts that traditional fault tree analysis. Functional control diagrams are used to guide the analysis. In addition, JAXA uses a model-based system engineering development environment (created originally by Leveson and called SpecTRM) which also assists in the hazard analysis.

One of the advantages of STPA is that it can be applied early in the system engineering and development process in a safety-driven design process where hazard analysis drives the design decisions rather than waiting until reviews identify problems that are then costly or difficult to fix. It can also be applied in an after-the-fact analysis and hazard assessment, which is what we did in this case study.

This paper describes the experimental application of STPA to the JAXA HTV in order to determine the feasibility and usefulness of the new hazard analysis technique. Because the HTV was originally developed using fault tree analysis and following the NASA standards for safety-critical systems, the results of our experimental application of STPA can be compared with these more traditional safety engineering approaches in terms of the problems identified and the resources required to use it.

1. INTRODUCTION

Japan Aerospace Exploration Agency (JAXA) develops various types of space systems such as satellites, rockets, and manned systems including the International Space Station (ISS). Needless to say, safety is one of the essential characteristics to be achieved for these space systems. A hazard analysis is one of the most important elements in developing safe space systems. During system design, component failure based analyses, such as FTA and FMEA, are commonly used as hazard analysis methods. However, it is difficult to identify hazard causes that are not related to component failures using FTA/FMEA, which can lead to inadequate investigation for hazards.

Although JAXA has not experienced any critical accidents caused by factors other than component failures so far, JAXA is considering introducing a new hazard analysis methodology, called STAMP/STPA, to avoid future accidents. STAMP/STPA focuses on control problems, not component failures, and it is able to identify hazards that arise due to unsafe and unintended interactions among the system components without component failures.

As a pilot case study, the H-IIB Transfer vehicle (HTV) was chosen as a target system of analysis. The HTV is an unmanned visiting vehicle launched by the H-IIB rocket to carry necessary components and commodities to the ISS. After launch, the HTV performs an automated rendezvous flight to carry cargo to the ISS. Figure 1 depicts the HTV's approach sequence during Proximity

Operations [1]. After approval for final approach is given by NASA's ISS Mission Management Team, the HTV moves from the Approach Initiation (AI) point to the final approach point guided by Relative GPS Navigation and approaches the ISS from the nadir side of the ISS using a laser sensor called the Rendezvous Sensor. Once the HTV reaches a grappling point 10 meters below the ISS, called the Berthing Point, the ISS crew disables the HTV thrusters and then manipulates the Space Station Remote Manipulator System (SSRMS) to capture a Flight Releasable Grapple Fixture (FRGF) of the HTV [2].

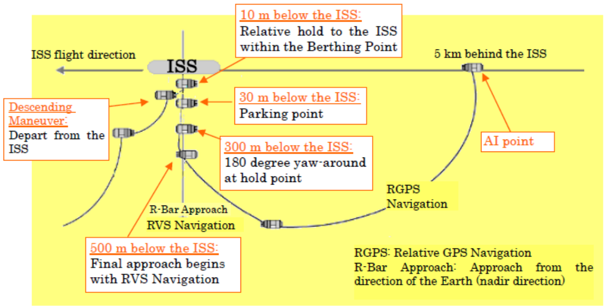

10 m below the ISS:
Relative hold to the ISS within the Berthing Point
ISS flight direction
ISS
5 km behind the ISS
Descending Maneuver:
Depart from the ISS
30 m below the ISS:
Parking point
AI point
300 m below the ISS:
180 degree yaw-around at hold point
RGPS Navigation
R-Bar Approach
RVS Navigation
500 m below the ISS:
Final approach begins with RVS Navigation
RGPS: Relative GPS Navigation
R-Bar Approach: Approach from the direction of the Earth (nadir direction)

Figure 1. HTV Proximity Operations [1].

The first test flight was successfully completed in November, 2009. JAXA plans to start regular cargo transportation to the ISS in 2010.

In the development of the HTV, NASA safety requirements were applied and potential HTV hazards were analyzed using FTA. Emphasis was on the hazards of the integrated operation phase, i.e., from approaching the ISS, berthing to departure, and the results of FTA-based hazard analysis were documented in the Hazard Report (HR). "Collision with the ISS" is one of the catastrophic hazards. Redundant design is used for important safety-related components. In addition, a collision avoidance maneuver is implemented to abort from the ISS collision trajectory if redundant component failure occurs. The HR was reviewed by the NASA Safety Review Board. In the Safety Review, the validity of all the contents of the HR, such as the identified hazard causes, the hazard control used for each hazard cause, the design of the control and the verification method, were reviewed.

NASA-JAXA also analyzed the hazards identified for ISS-HTV integrated operation and documented the results as the Integrated Hazard Analysis (IHA). In accordance with the results of the IHA, NASA-JAXA has defined flight rules for integrated operation. In the first flight, both NASA and JAXA operators adhered to the flight rules and carefully communicated with each other, and the mission was accomplished successfully.

After this first flight, JAXA is continuing HTV operation while including new operators. However, JAXA has not performed hazard analysis focusing on control problems and regular operations. To ensure safe transportation service, we have started analysis using STAMP/STPA to identify whether there are potential hazards caused by unsafe and unintended controls.

2. STAMP/STPA

Current hazard analysis techniques start from a completed design and assume that accidents are caused by component failures. Because the primary cause of accidents in the old systems was component failure, the hazard analysis techniques and safety design techniques focused on identifying critical components and either preventing their failure (increasing component integrity) or providing redundancy to mitigate the effects of their failure.

There are several limitations of these approaches. One of the major problems is that most common hazard analysis techniques such as FTA or FMECA, work on an existing design. Therefore, much of the effort goes into proving that existing designs are safe rather than building designs that are safe from the beginning. But system designs have become so complex that waiting until a design is mature enough to perform a safety analysis on it is impractical. The only practical and cost-effective safe design approach in these systems is to design safety in from the beginning. In safety-driven design, the information needed by the designers to make good decisions is provided to them before they create the design and the analyses are performed in parallel with the design process rather than after it. Because software errors and flawed human decision making do not involve random failures, hazard analysis techniques that only identify such failures will not be effective for them. A new approach to hazard analysis is required, which in turn must rest on an expanded model of accident causality.

Against this background, Leveson developed a new accident model called STAMP (Systems-Theoretic Accident Model and Processes), which has been described in detail elsewhere [2]. The rest of this section describes a new hazard analysis technique, based on STAMP, which is called STPA (STAMP-Based Process Analysis) [3]. An important advantage of this technique is that it can be used to drive the earliest design decisions and then proceed in parallel with ensuring design decisions and design refinement.

In STPA, the system is viewed as a collection of interacting loops of control. The assessment begins with identifying hazards for the system and translating them into top-level system safety constraints. Next, a basic control structure is defined. A control structure diagram depicts the components of the system and the paths of control and feedback. Using the control structure diagram as a guide for conducting the analysis, each control action is assessed for potential contribution to hazards. Identified inadequate control actions are used to refine system safety constraints. Finally, the analyst determines how the potentially hazardous control actions could occur. If the controls in place are inadequate, recommendations should be developed for additional mitigations.

2.1. Review System Hazards and System-Level Safety Constraints

A safety-driven design starts from identifying accidents or unacceptable loss events, such as loss of vehicle, loss of life, or loss of mission or equipment, and then defining the hazardous states in the system that would allow these accidents to occur. The hazards are then translated into safety constraints on the system state and behavior so will prevent the hazardous states from occurring. For example, the translation of hazard to related safety constraint in an automated elevator door controller is simple:

Hazard:
A person is present in the doorway when the door is closing.
Safety Constraint:
An elevator door must not close while anyone is in the doorway.

Although the first step of STPA is similar to that performed in other hazard analysis techniques, the later steps either deviate from traditional practice or provide a rigorous framework for doing what is traditionally done in an ad hoc manner.

2.2. Define Safety Control Structure

Once the hazards to be assessed have been reviewed, the analyst develops a diagram of the safety control structure of the system. Figure 2 shows a generalized safety control structure diagram, which does not represent any one particular system. Each node in the graph is a human or machine component in a socio-technical system. Connecting lines show control actions used to enforce safety constraints on the system and feedback that provides information to the controlling entity. These lines are annotated with a description of the information reported or controls applied.

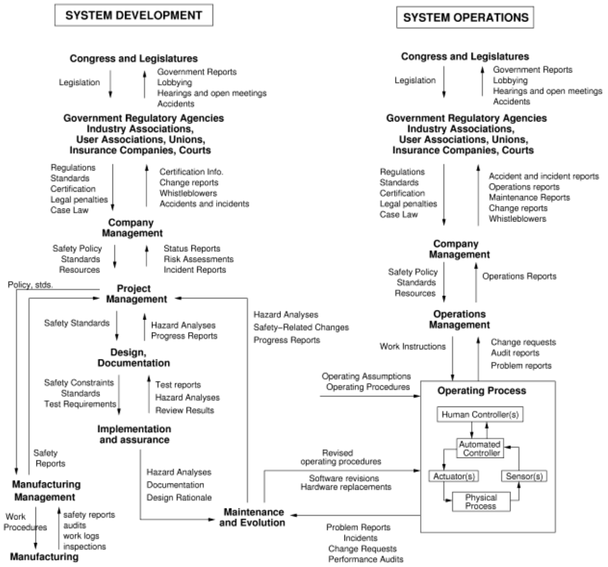


Figure 2. Generalized control structure diagram [3].

2.3. Identify Potentially Inadequate Control Actions

After the system control structure has been defined, the next step is to determine how the controlled system can get into a hazardous state. A hazardous state is a state that violates the safety constraints that are defined for the system. STPA views hazardous states as a result of ineffective control. Therefore, the assessment proceeds by identifying potentially inadequate control actions. Inadequate controls fall into the following four general categories:

1. A required control action to maintain safety is not provided.
2. An incorrect or unsafe control action is provided that induces a loss.
3. A potentially correct or adequate control action is provided too early, too late, or out of sequence.
4. A correct control action is stopped too soon.

Control actions may be required to handle component failures, environmental disturbances, or dysfunctional interactions among the components. Incorrect or unsafe control actions may also cause dysfunctional behavior or interactions among components. Note that these inadequate control actions may or may not be present in the actual system. These are hypotheses that must be confirmed or rejected based on investigation into the behavior of the system as it has been designed and built. To ensure a complete assessment, each control action must be investigated in turn.

2.4. Determine How Potentially Inadequate Control Actions Could Manifest in the System and Develop Mitigations

The previous step of the assessment will yield a set of potentially inadequate control actions. If present in the system, these inadequate control actions will provide a means for the system to enter a hazardous state. In this step of the assessment, the analyst determines how the potentially hazardous control actions can occur.

STPA works on functional control diagrams and is guided by a set of generic control loop flaws. Because accidents result from inadequate control and enforcement of safety constraints, the process that leads to accidents can be understood in terms of flaws in the system development and system operations control structures in place during design, implementation, manufacturing, and operation. These flaws can be classified and used during accident analysis to assist in identifying all the factors involved in the accident or during hazard analysis and other accident prevention activities. Figure 3 shows the causal factors leading to hazards.

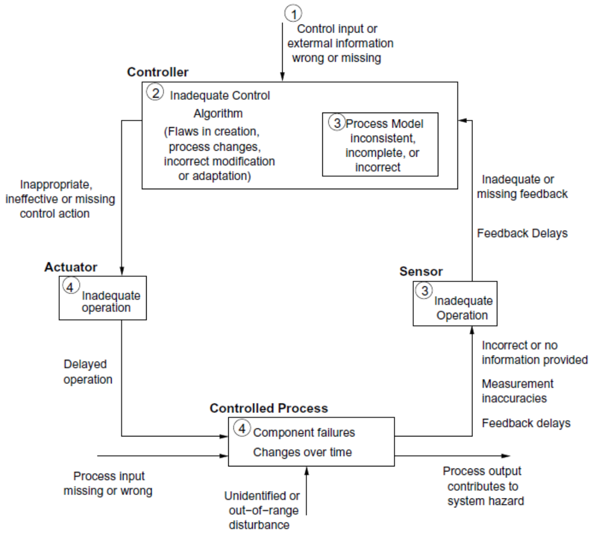

1
Control input or external information wrong or missing
Controller
2 Inadequate Control Algorithm (Flaws in creation, process changes, incorrect modification or adaptation)
3 Process Model inconsistent, incomplete, or incorrect
Inappropriate, ineffective or missing control action
Inadequate or missing feedback
Feedback Delays
Actuator
4 Inadequate operation
Sensor
3 Inadequate Operation
Delayed operation
Incorrect or no information provided
Measurement inaccuracies
Feedback delays
Controlled Process
4 Component failures
Changes over time
Process input missing or wrong
Unidentified or out-of-range disturbance
Process output contributes to system hazard

Figure 3. Causal factors leading to hazards [3].

In each control loop at each level of the socio-technical control structure, unsafe behavior results from either a missing or an inadequate constraint on the process at the lower level or inadequate enforcement of the constraint leading to its violation. Because each component of the control loop may contribute to inadequate control, classification starts by examining each of the general control loop components and evaluating their potential contribution: (1) the controller may issue inadequate or inappropriate control actions, including inadequate handling of failures or disturbances in the physical process, (2) control actions may be inadequately executed, or (3) there may be missing or inadequate feedback. These same general factors apply at each level of the socio-technical safety control structure, but the interpretations (applications) of the factor at each level may differ. For all of the factors, at any point in the control loop where a human or organization is involved, it is necessary to evaluate the context in which decisions are made in order to understand the types and reasons for potentially unsafe decisions to be made and to design controls or mitigation measures for them. Note that accidents caused by basic component failures are included here.

3. STPA CASE STUDY OF HTV CAPTURE OPERATION

An STPA case study of HTV capture operation was conducted. While, as discussed in the previous section, one of the important advantages of STPA is that it can be applied as soon as the high-level system accidents and hazards are known in the early stage of development, it is also interesting to apply STPA to the existing system and to compare the results with the past hazard reports.

This section provides an overview of the HTV proximity operations that we focused on in our case study and then presents the results according to the steps described in the previous section.

3.1. Overview of HTV Proximity Operations

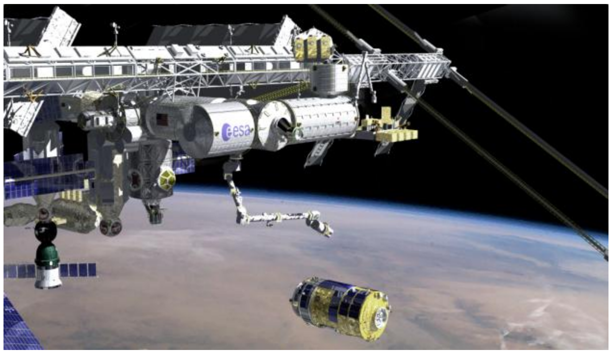


Figure 4. HTV approach to the ISS [1].

The HTV operations can be divided into the following phases [1]:

1. Launch
2. Rendezvous flight to the ISS
3. Berthing to the ISS
4. Docked operations
5. Undock/Departure from the ISS
6. Reentry

Among the above six phases, we decided to focus on the berthing phase in our case study because this phase includes HTV capture by the SSRMS (Figure 4). Because the ISS, which protects the lives of the crew, can be damaged in this operation, capture is one of the most critical operations of the mission.

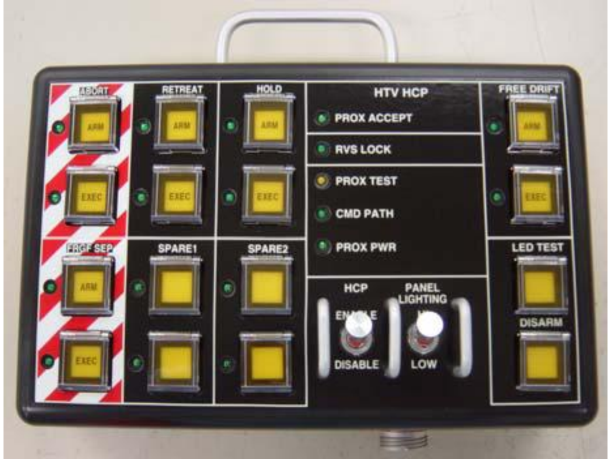


Figure 5. Hardware Command Panel (HCP) [1].

The ISS crew can control the HTV during its final approach to the ISS using the Hardware Command Panel (HCP) in Figure 5 in case of emergency. The key functions of the HCP include:

- <u>ABORT</u>
- <u>RETREAT</u>
  Retreat to 30 m or 100 m below the ISS
- <u>HOLD</u>
  Hold the approach
- <u>FREE DRIFT</u>
  Disable the HTV thrusters
- <u>FRGF SEP</u>
  Separate the FRGF

3.2. System-Level Hazards

During the proximity operation, the most catastrophic accident is obviously an HTV collision with the ISS. It might not only result in loss of the HTV mission, but could also lead to damage to the ISS modules or the SSRMS. Even if collision does not occur, the loss of the HTV mission (e.g. an unintended abort) will waste a great deal of money. Therefore, the hazard analysis must identify all the potential hazards unintendedly caused by capture operation that would allow the collision or loss of mission to occur.

3.3. Basic Control Structure and Event/Command Sequence during Capture Phase

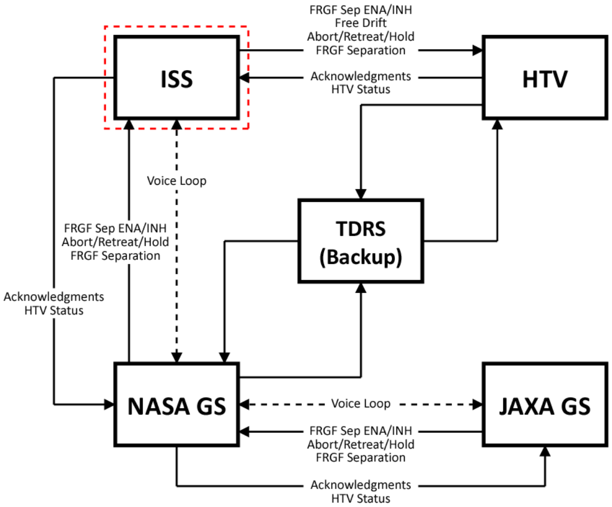


Figure 6. Control Structure Diagram – Level 0.

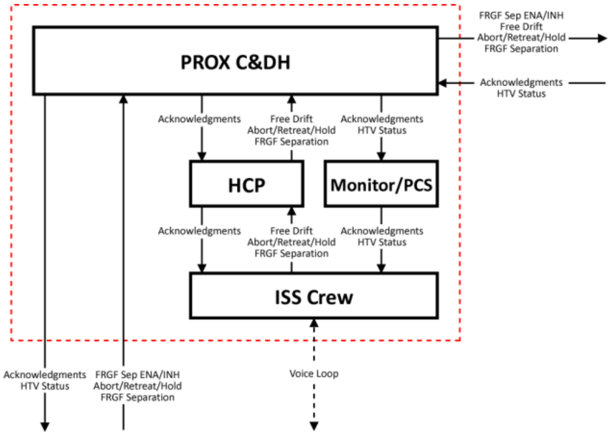


Figure 7. ISS Control Structure – Level 1.

Table 1. Command Sequence during Capture Phase.

#	Event/Command	from	to	Description
1	FRGF Sep ENA	JAXA GS	HTV	Enable FRGF separation right in case of emergency
2	Free Drift (Deactivation)	ISS (Crew)	HTV	Transition from "Capture Point Hold Mode" to "Free Drift Mode" to disable the HTV guidance and control functions
C	Capture	ISS (Crew)	HTV	Manipulate the SSRMS to capture FRGF of the HTV
3	FRGF Sep INH	JAXA GS	HTV	Inhibit FRGF separation to prevent an unintended separation after the capture

Once the operation phase to be focused on was determined, we defined the basic control structure. To limit the complexity of the diagram, we split the structure into two levels of abstraction. Figure 6 shows a level-0 control structure diagram for the HTV capture operation.

It is composed of 5 major components; ISS, HTV, NASA ground station, JAXA ground station, and Tracking and Data Relay Satellite (TDRS) as backup for communication. Figure 7 shows a level-1 ISS control structure. Major components inside the ISS include the Proximity Communication Command and Data Handling (PROX C&DH) system, the Hardware Command Panel (HCP), visual monitors/Portable Computer System (PCS), and the ISS crew. Connecting lines between those components show control actions, information, and acknowledgments (feedback). There is also a voice loop connection between the ISS crew, NASA ground station, and JAXA ground station.

In order to annotate some of those connecting lines with command actions, we reviewed the nominal command sequence during the capture phase. Table 1 lists selected command actions around the time of the capture. After the HTV has reached the Berthing Point or Capture Point, the JAXA ground station sends an FRGF Separation ENABLE command, which enables FRGF separation in case of an emergency. The ISS crew then sends a Free Drift command using HCP to disable the HTV guidance and control functions. If the capture is started without this deactivation, the contact with the robotic arm could be interpreted as a disturbance by an external force, which would trigger an automatic attitude control action. Once the HTV is deactivated, the ISS crew has to manipulate the SSRMS to grapple the HTV as promptly as possible. After the successful capture, the JAXA ground station issues an FRGF Separation INHIBIT command to the HTV to prevent an unintended separation. These four events are the critical proximate events of the capture phase.

Table 2. Potentially Hazardous Commands/Events during the Capture Phase.

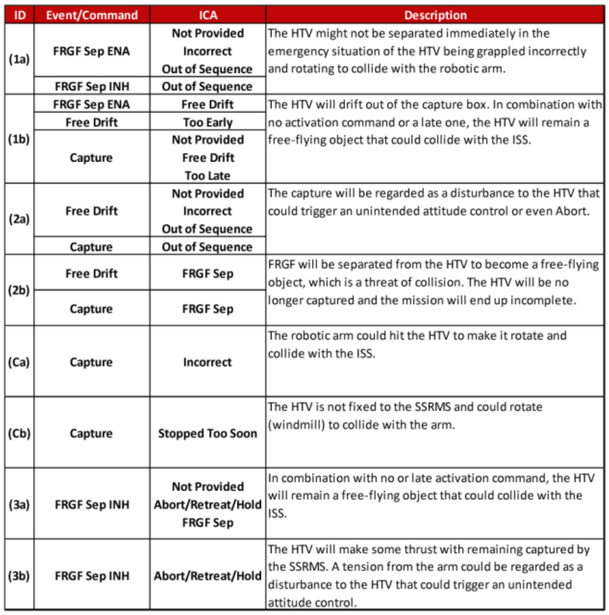
		Category 1	Category 2				Category 3			Category 4
#	Event/Command	Not Provided	Incorrect Provided				Provided			Stopped Too Soon
				Abort/Retreat/Hold (unintended)	Free Drift (unintended)	FRGF Sep (unintended)	Too Early	Too Late	Out of Sequence	
1	FRGF Sep ENA	If this is not detected and capture is started, (1a) the HTV might not be separated immediately in the emergency situation of the HTV being grappled incorrectly and rotating to collide with the robotic arm.	If it is not detected that FRGF Sep INH is provided instead of ENA and capture is started, (1a).	If an Abort/Retreat/Hold command is provided, an uninteded Abort/Retreat/Hold will start processing, which is not hazardous. But the mission will end up incomplete or the capture process will have to be started over.	If Deactivation command is provided instead of FRGF Sep ENA, the mode transition is too early and (1b) the HTV will drift out of the capture box. In combination with no activation command or a late one, the HTV will remain a free-flying object that could collide with the ISS.	If FRGF Separation comnand is provided, nothing will happen since FRGF separation remains inhibited.	If FRGF Sep ENA is provided too early, it will just increase the time during which the HTV accepts an FRGF separation, which will then contribute to increasing the possibility of an unintended FRGF separation by crew error.	If FRGF Sep ENA is provided too late, it will only delay the capture process.	If FRGF Sep ENA is provided out of sequence with capture, (1a).	N/A
2	Free Drift (Deactivation)	If this is not detected and capture is started, (2a) the capture will be regarded as a disturbance to the HTV that could trigger an unintended attitude control or even Abort.	If the HTV fails to transition to the Free Drift Mode and capture is started, (2a).	If an Abort/Retreat/Hold command is provided instead of Deacticvation, an uninteded Abort/Retreat/Hold will start processing, which is not hazardous. But the mission will end up incomplete or the capture process will have to be started over.		Since FRGF separation has been enabled, if FRGF Separation command is provided instead of Deactivation, (2b) FRGF will be separated from the HTV to become a free-flying object, which is a threat of collision. The HTV will be no longer captured and the mission will end up incomplete.	If Deactivation is provided too early and capture is not started immediately enough, (1b).	If Deactivation is provided too late, it will delay the capture process. Since FRGF separation has been enabled, it will contribute to increasing the possibility of an unintended FRGF separation by crew error.	If Deactivation is provided out of sequence with capture, (2a).	N/A
C	Capture	If capture is not performed, (1b).	If the crew makes an operational mistake of the SSRMS, (Ca) the robotic arm could hit the HTV to make it rotate and collide with the ISS.	If an Abort/Retreat/Hold command is provided, an uninteded Abort/Retreat/Hold will start processing, which is not hazardous. But the mission will end up incomplete or the capture process will have to be started over.	Since the HTV has already been in the Free Drift Mode, nothing will happen. In combination with no or late capture, (1b).	Since FRGF separation has been enabled, if FRGF Separation command is provided, (2b).	Since the HTV has already been in the Free Drift Mode, a too early capture is nothing but good.	If capture is performed too late, (1b).	If capture is performed out of sequence with Deactivation, (2a).	If capture is stopped halfway and incomplete, (Cb) the HTV is not fixed to the SSRMS and could rotate (windmill) to collide with the arm.
3	FRGF Sep INH	If FRGF Sep INH is not provided, the HTV is left capable of FRGF separation. An unintended FRGF separation after the successful capture could occur. (3a) In combination with no or late activation command, the HTV will remain a free-flying object that could collide with the ISS.	If FRGF Sep ENA is provided instead of INH, the HTV is left capable of FRGF separation. An unintended FRGF separation after the successful capture could occur. (3a).	If an Abort/Retreat/Hold command is provided, an unintended Abort/Retreat/Hold will start processing. If FRGF separation is provided while RVFS fails to return its mode back to CP Hold Mode, (3a). If RVFS returns its mode to CP Hold Mode while FRGF separation is not provided, (3b) the HTV will make some thrust with remaining captured by the SSRMS. A tension from the arm could be regarded as a disturbance to the HTV that could trigger an unintended attitude control.	Since the HTV has already been in the Free Drift Mode and captured by the SSRMS, nothing will happen.	Since FRGF separation still remains enabled, if FRGF Separation command is provided, the HTV will be separated from the SSRMS. (3a).	Since capture has already been successfully completed, a too early FRGF Sep INH is nothing but good.	If FRGF Sep INH is provided too late, it will just increase the time during which the HTV accepts an FRGF separation, which will then contribute to increasing the possibility of an unintended FRGF separation by crew error.	If FRGF Sep INH is out of sequence with capture, (1a).	N/A

3.4. Identification of Hazardous Control Behavior

For each command, the conditions under which the command could lead to a system hazard were identified using the four general categories of inadequate control actions: "Not Provided when it should be," "Incorrectly Provided," "Provided Too Early, Too Late, or Out of Sequence," and "Stopped Too Soon." Table 2 shows various hazardous behaviors identified. Each cell in the table describes what could happen if each command is executed inadequately. Because, except for the nominal commands shown in Table 1, the ISS crew can control the HTV by using HCP to issue ABORT, RETREAT, HOLD, or FRGF SEP, the “Incorrectly Provided” category included unintended Abort/Retreat/Hold, Free Drift, and FRGF separation.

It was found that some cells converge to the same or similar hazard and that some cells do not lead to a hazardous state. We identified a total of eight types of hazardous control behaviors, which are underlined in Table 2. Each hazard was assigned an identifier from (1a) through (3b). A hazard (1a) at the top left corner, for instance, is that if capture is started without detecting that an FRGF separation ENABLE command has not been provided, the HTV might not be separated immediately in the emergency situation of the HTV being grappled incorrectly and rotating to collide with the robotic arm. Table 3 summarizes the eight hazardous control behaviors.

Table 3. Hazardous Control Behavior and the Potential Result.



ID	Event/Command	ICA	Description
(1a)	FRGF Sep ENA	Not Provided Incorrect Out of Sequence	The HTV might not be separated immediately in the emergency situation of the HTV being grappled incorrectly and rotating to collide with the robotic arm.
	FRGF Sep INH	Out of Sequence	
(1b)	FRGF Sep ENA	Free Drift	The HTV will drift out of the capture box. In combination with no activation command or a late one, the HTV will remain a free-flying object that could collide with the ISS.
	Free Drift	Too Early	
	Capture	Not Provided Free Drift Too Late	
(2a)	Free Drift	Not Provided Incorrect Out of Sequence	The capture will be regarded as a disturbance to the HTV that could trigger an unintended attitude control or even Abort.
	Capture	Out of Sequence	
(2b)	Free Drift	FRGF Sep	FRGF will be separated from the HTV to become a free-flying object, which is a threat of collision. The HTV will be no longer captured and the mission will end up incomplete.
	Capture	FRGF Sep	
(Ca)	Capture	Incorrect	The robotic arm could hit the HTV to make it rotate and collide with the ISS.
(Cb)	Capture	Stopped Too Soon	The HTV is not fixed to the SSRMS and could rotate (windmill) to collide with the arm.
(3a)	FRGF Sep INH	Not Provided Abort/Retreat/Hold FRGF Sep	In combination with no or late activation command, the HTV will remain a free-flying object that could collide with the ISS.
(3b)	FRGF Sep INH	Abort/Retreat/Hold	The HTV will make some thrust with remaining captured by the SSRMS. A tension from the arm could be regarded as a disturbance to the HTV that could trigger an unintended attitude control.

3.5. Identification of Causes of the Hazardous Control Behavior

While some hazards can be designed out of the system without knowing all of the potential scenarios that can lead to the hazard, more information about causality is often very useful.

After the hazardous control behavior has been identified, design features are used to eliminate or control it or, if the system design already exists, the design is analyzed to determine if the potentially hazardous behavior has been eliminated or controlled. Accomplishing this goal may require more information about the cause of the behavior and this information is identified using the fourth step of STPA. The control structure diagram is evaluated using the potential control flaws in Figure 3. For an example of this step of the analysis, we selected a hazardous control behavior (1b).

Hazard (1b):
The HTV will drift out of the capture box. In combination with no activation command or a late one, the HTV will remain a free-flying object that could collide with the ISS.
Safety Constraint (1b):
The ISS crew must activate the HTV as soon as possible after drift out.

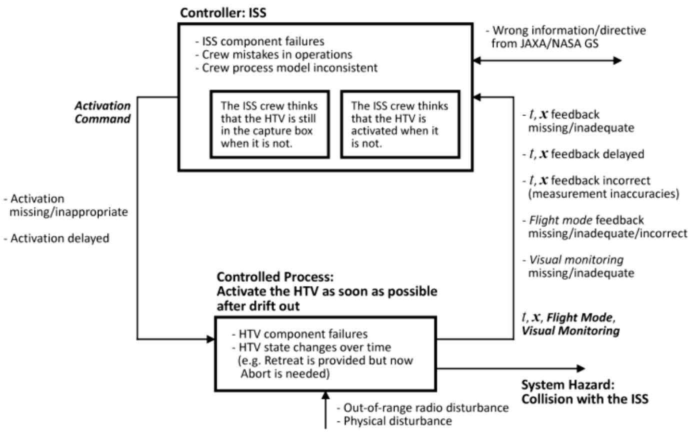

Controller: ISS
- ISS component failures
- Crew mistakes in operations
- Crew process model inconsistent
The ISS crew thinks that the HTV is still in the capture box when it is not.
The ISS crew thinks that the HTV is activated when it is not.
- Wrong information/directive from JAXA/NASA GS
Activation Command
- Activation missing/inappropriate
- Activation delayed
- t, x feedback missing/inadequate
- t, x feedback delayed
- t, x feedback incorrect (measurement inaccuracies)
- Flight mode feedback missing/inadequate/incorrect
- Visual monitoring missing/inadequate
t, x, Flight Mode, Visual Monitoring
Controlled Process: Activate the HTV as soon as possible after drift out
- HTV component failures
- HTV state changes over time (e.g. Retreat is provided but now Abort is needed)
System Hazard: Collision with the ISS
- Out-of-range radio disturbance
- Physical disturbance

Figure 8. Causal factors leading to hazardous control behavior (1b).

Figure 8 shows the causal factors leading to a hazardous control behavior (1b), where t and $\boldsymbol{x}$ denote the time elapsed since the HTV is deactivated and the HTV’s state vector, respectively. As required by the HTV flight rules, the ISS crew must capture the HTV within 99 seconds from deactivation; otherwise the HTV must be activated again. In addition, if the ISS crew confirms by the state vector feedback or visual monitoring that the HTV drifts out of the capture box, the HTV must be activated again. Therefore, t, $\boldsymbol{x}$, the HTV *Flight Mode* (activated or deactivated), and *Visual Monitoring* are the critical information for the crew to make an appropriate decision. If either of them is missing or inadequate, the crew must send an *Activation Command* to the HTV. For each of those causal factors identified in Figure 8, examples of hazardous scenarios that could lead to collision with the ISS are listed as follows:

- ISS component failures
  Due to an ISS component failure, the *Activation Command* might not be processed although the crew is trying to issue it.
- Crew mistakes in operation
  The ISS crew might make a careless error to issue a wrong command.
- Crew process model inconsistent
  Due to a freezing of the visual monitor, the ISS crew might think that the HTV is still in the capture box when it has already drifted out, which would delay the *Activation Command* by the crew.
  Due to an incorrect *Flight Mode* feedback, the crew might think that the HTV is activated when it is not and therefore the crew might not issue the *Activation Command*.
- Activation missing/inappropriate
  The *Activation Command* might be corrupted during transmission and the crew must reissue it, which would delay the activation of the HTV.
- Activation delayed
  The *Activation Command* could be delayed during transmission, which would then delay the activation of the HTV.
- HTV component failures
  Due to an HTV component failure, the HTV might not execute the activation although the HTV has received it.
- HTV state changes over time
  Due to a change in the HTV's position relative to the ISS while the crew was trying to issue a Hold command, the HTV might now need an Abort command instead of Hold to escape in a safe trajectory.
- Out-of-range radio disturbance
  A radio disturbance could interfere with the *Activation Command* coming in.
- Physical disturbance
  Physical disturbance by the robotic arm could accelerate the change in HTV's attitude and the activation by the crew might not be in time.
- *t*, ***x*** feedback missing/inadequate
  Due to a missing ***x*** during transmission, the ISS crew might be confused and issue the *Activation Command* too late.
- *t*, ***x*** feedback delayed
  An ***x*** feedback could be delayed during transmission and arrive too late for the crew to issue an Abort command.
- *t*, ***x*** feedback incorrect
  An ***x*** feedback could be incorrect due to measurement inaccuracies as if the HTV was still in the capture box and the crew might not issue the *Activation Command*.
- *Flight Mode* feedback missing/inadequate
  A *Flight Mode* feedback might not be received and the crew might be confused and issue the *Activation Command* too late.
- *Flight Mode* feedback incorrect
  A *Flight Mode* feedback might be incorrect and the crew might think that the HTV is activated when it is not and therefore might not issue the *Activation Command*.
- *Visual Monitoring* missing/inadequate
  A freezing of visual monitor might delay the *Activation Command* by the crew.
- Wrong information/directive from JAXA/NASA GS
  Judging from delayed information, the JAXA GS might tell the ISS crew to capture the HTV when the crew should now issue an Abort command, which could confuse the crew.

All the factors above could lead to no activation or a late one after drift out of the capture box, which would contribute to collision with the ISS. One of the features of STPA can be seen in crew process model inconsistency. If the HTV was designed such that it could send back the *Flight Mode* before it really was activated, an inconsistency could result. This kind of hazard cause must be identified in the early stage of development and eliminated by the design.

The design of the HTV must be evaluated with respect to each of these potential causal factors of hazard (1b) to determine whether the design prevents it or whether preventive or mitigation measures must be added to the design. If the analysis is done early in the design process, a design can be created that mitigates the potential causal factors from the start.

4. EVALUATION

4.1. Objective and Procedure

The feasibility of applying STPA to the HTV was demonstrated by the application itself. In order to determine the usefulness of STPA as a method for HTV hazard analysis, hazardous commands/events that were identified by STPA were compared with the existing FTA results. Through this comparison task, we wanted to answer the following two questions.

Q1: Do the hazardous scenarios (causes) identified by STPA cover the causes identified in the HTV fault tree?
Q2: Are additional causes found by STPA that are not in the fault trees?

In the comparison, the fault tree branches for the capture phase were compared with the STPA analysis for the same phase. We mapped the STPA hazardous scenarios to the FT branches and identified differences.

4.2. Results and Discussions

The results from the comparison analysis answering the above two questions were:
Q1: We found that causal factors identified by STPA included all the hazard causes of the fault tree.
Q2: There were causal factors that were identified by STPA only.

Causal factors that were identified by both and causal factors identified by STPA only are shown below.

Identified by both STPA and FTA

Controller:

- ISS component failures

Activation Command:

- Activation missing/inappropriate

Controlled Process:

- HTV component failures
- HTV state changes over time
- Physical disturbance

Identified by STPA only

Controller:

- Crew mistakes in operation
- Crew process model inconsistent

Activation Command:

- Activation delayed

Controlled Process:

- Out-of-range radio disturbance

Acknowledgment of Control Action:

- *t*, ***x*** feedback missing/inadequate/delayed
- *t*, ***x*** feedback incorrect
- *FM* feedback missing/inadequate/incorrect
- *VM* missing/inadequate
  (*FM*: *Flight Mode, VM*: *Visual Monitoring*)

Other Controllers:

- Wrong information/directive from JAXA/NASA GS

We found that causal factors other than component failures such as process model inconsistency, causal factors with regard to "delay of command," "delay of feedback," and "acknowledgment of control action" are not identified by FTA. As HTV's specific causal factors that are identified in Section 3.5 show that these causal factors are caused by control flaws in the control loop involving total system integration among ISS, HTV, and NASA/JAXA GS.

Most of basic events that had been identified by FTA in HTV hazard analysis are events that occur by coincidence such as component failures. Causal factors in the control loop in the control structure diagram that were identified by STPA were not found by FTA.

This result shows that there are several causal factors that are not identified by FTA. However, these causal factors are considered and controlled in the HTV's design and operation.

5. CONCLUSION & FUTURE WORK

For the experimental application of STPA to the HTV, hazardous behaviours and causes for violation of the safety requirements were identified. We selected a hazardous control behavior (1b) and compared the STPA results with the existing Hazard Reports for HTV. From the comparison, we found that STPA identified the failures in the existing fault tree analysis but STPA also identified additional causal factors which had not been identified in the existing FTA.

For case (1b), the potential feasibility and benefit of using STPA for a safety critical space system was demonstrated. We are continuing the analysis of other cases and examining additional potential benefits of STPA.

We also need to confirm that the causal factors found by STPA only in case (1b) have already been considered in the current HTV design and operation manual. Those causal factors must also be documented to prevent future modifications in HTV updates from violating safety constraints.

This paper evaluated the feasibility and usefulness of STPA for system safety analysis, especially for early system design phase. STPA makes it possible to identify safety requirements and safety constraints of the system before the detailed design starts and therefore without knowing the details of the component design and failure modes. Beyond this, STPA identifies additional scenarios not considered in fault trees, i.e., not involving component failures. These additional hazardous scenarios must be eliminated or controlled according to the system design. It also enables the application of STPA from concept design to detailed design and implementation step by step.

Several issues were identified with respect to how to model and analyze real systems. To deal with these difficulties, we are now extending the scope and depth of the analysis and trying to identify systematic procedures for STPA. In addition, we will begin studying how to apply STPA to space development in practice from the perspectives of safety-driven system design and design verification.

6. ACKNOWLEDGMENTS

The authors would like to thank Assoc. Prof. Seiko Shirasaka at Keio University and Margaret Stringfellow at MIT for their support of this study.

This research was performed at the Massachusetts Institute of Technology and the Japan Aerospace Exploration Agency under a contract with the Japan Manned Space Systems Corporation.

7. REFERENCES

1. Japan Aerospace Exploration Agency, "HTV-1 Mission Press Kit," September 2009.
2. Tsukui, J., Hotta, S, Imada, T., Yamanaka, K., and Kasai, T., "Automatic Rendezvous to the International Space Station," *Proceedings of the 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, i-SAIRAS 2003, Nara, Japan, May 19-23, 2003.
3. Leveson, Nancy G., "A New Accident Model for Engineering Safer Systems," *Safety Science*, Vol. 42, No. 4, pp. 237-270, April 2004.
4. Leveson, Nancy G., "Software Challenges in Achieving Space Safety," *Journal of the British Interplanetary Society* (JBIS), Volume 62, 2009.