A Systems Approach to Safety Engineering

Nancy G. Leveson
Aeronautics and Astronautics
Engineering Systems
MIT
The Problem
Why do we need a new approach?

• New causes of accidents in complex, software-intensive systems
  – Software does not “fail,” it usually issues unsafe commands
  – Role of humans in systems is changing

• Traditional safety engineering approaches were developed for relatively simple electro-mechanical systems

• We need more effective techniques for these new systems and new causes
Accident with No Component Failures
Types of Accidents

• Component Failure Accidents
  – Single or multiple component failures
  – Usually assume random failure

• Component Interaction Accidents
  – Arise in interactions among components
  – Related to interactive complexity and use of computers
  – Level of interactions has reached point where can no longer be thoroughly
    • Planned
    • Understood
    • Anticipated
    • Guarded against
It's still hungry ... and I’ve been stuffing worms into it all day.
So What Do We Need to Do?  
“Engineering a Safer World”

• Expand our accident causation models
• Create new hazard analysis techniques
• Use new system design techniques
  – Safety-driven design
  – Improved system engineering
• Improve accident analysis and learning from events
• Improve control of safety during operations
• Improve management decision-making and safety culture
An Expanded View of Accident Causes
Accident Causality Models

• Explain why accidents occur

• Provide foundations for:
  – Investigating and analyzing cause(s) of accidents
  – Analyzing hazards
  – Designing to prevent future losses
  – Assessing risk

• Current ones are VERY old
  – Created before computers and modern technology
Chain-of-Events Causation Models

- **Assumption**: Accidents are caused by chains of component failures

- Simple, direct relationship between events in chain
  - Ignores non-linear relationships, feedback, etc.

- Events almost always involve component failure, human error, or energy-related event

- Forms the basis for most safety-engineering and reliability engineering analysis:
  - FTA, PRA, FMECA, Event Trees, etc.

and design:
  - e.g., redundancy, over-design, safety margins, ….
Limitations of Chain-of-Events Causation Models

• Oversimplifies causality

• Excludes or does not handle
  – Component interaction accidents (vs. component failure accidents)
  – Indirect or non-linear interactions among events
  – Systemic factors in accidents
  – Human “errors”
  – Migration toward states of increasing risk
Software-Related Accidents

• Are usually caused by flawed requirements
  – Incomplete or wrong assumptions about operation of controlled system or required operation of computer
  – Unhandled controlled-system states and environmental conditions

• Merely trying to get the software “correct” or to make it reliable will not make it safer under these conditions.
Software-Related Accidents (2)

- Software may be highly reliable and “correct” and still be unsafe:
  - Correctly implements requirements but specified behavior unsafe from a system perspective.
  - Requirements do not specify some particular behavior required for system safety (incomplete)
  - Software has unintended (and unsafe) behavior beyond what is specified in requirements.
Operator Error: Old View
(Sidney Dekker, Jens Rasmussen)

• Operator error is cause of incidents and accidents

• So do something about operator involved (suspend, retrain, admonish)

• Or do something about operators in general
  – Marginalize them by putting in more automation
  – Rigidify their work by creating more rules and procedures
Operator Error: New View

- Operator error is a symptom, not a cause
- All behavior affected by context (system) in which occurs
- To do something about error, must look at system in which people work or operate machines:
  - Design of equipment
  - Usefulness of procedures
  - Existence of goal conflicts and production pressures
Adaptation

- Systems are continually changing
- Systems and organizations migrate toward accidents (states of high risk) under cost and productivity pressures in an aggressive, competitive environment [Rasmussen]
- Need to include this in our causality models
STAMP

A new accident causality model using Systems Theory (vs. Reliability Theory)
System’s Theoretic View of Safety

• Safety is a system property
  – It is not a component property
  – It can only be analyzed in the context of the whole

• Accidents arise from
  – Interactions among system components (human, physical, social)
  – That violate the constraints on safe component behavior and interactions

• Losses are the result of complex processes, not simply chains of failure events
STAMP (3)

• Systems can be viewed as hierarchical control structures
  – Systems are viewed as interrelated components kept in a state of dynamic equilibrium by feedback loops of information and control
  – Controllers imposes constraints upon the activity at a lower level of the hierarchy: safety constraints

• A change in emphasis:
  “prevent failures”
  ↓
  “enforce safety constraints on system behavior”
Control processes operate between levels of control

Controller

Model of Process

Control Actions

Feedback

Controlled Process

Process models must contain:
- Required relationship among process variables
- Current state (values of process variables)
- The ways the process can change state
Relationship Between Safety and Process Models

- Accidents occur when models do not match process and
  - Required control commands are not given
  - Incorrect (unsafe) ones are given
  - Correct commands given at wrong time (too early, too late)
  - Control stops too soon

Explains software errors, human errors, component interaction accidents …
HAZARD: ITP and Reference Aircraft violate minimum separation standard

Pilot

- Wrong interpretation of ITP requirements/procedures
- Incorrect input into aircraft controls (e.g. give too much throttle, or out of order - climb before accelerate)
- Wait too long to execute procedure

Command issued but not received by engines, wing flaps, etc.

Data inconsistency between ADS-B, ATC, instrumentation and pilot experience

Inaccurate feedback about relative aircraft position

Actuator Failure

Aircraft controls delay maneuver instigation or scramble order of operations

Sensor Failure

ITP & Ref Aircraft attitude not detected or detection/update is delayed

ITP Aircraft too close to Reference Aircraft
STAMP: System Theoretic Accident Model and Processes

• Treat safety as a dynamic control problem rather than a component failure problem.
  – O-ring did not control propellant gas release by sealing gap in field joint of Challenger Space Shuttle
  – Software did not adequately control descent speed of Mars Polar Lander
  – Temperature in batch reactor not adequately controlled in system design
  – Public health system did not adequately control contamination of the milk supply with melamine
  – Financial system did not adequately control the use of financial instruments

• Events are the result of the inadequate control
  – Result from lack of enforcement of safety constraints in system design and operations
Summary: Accident Causality

- Accidents occur when
  1. Control structure or control actions do not enforce safety constraints
     a. Unhandled environmental disturbances or conditions
     b. Unhandled or uncontrolled component failures
     c. Dysfunctional (unsafe) interactions among components
  2. Control actions inadequately coordinated among multiple controllers
  3. Control structure degrades over time (slow migration of control structure toward a state of high-risk)
     - Need to control and detect this migration
Uncoordinated “Control Agents”

“UNSAFE STATE”
BOTH TCAS and ATC provide uncoordinated & independent instructions

Control Agent (TCAS)

Instructions

Instructions

No Coordination

Control Agent (ATC)

Instructions

Instructions
Uses for STAMP

• Create new, more powerful hazard analysis techniques (STPA)
• Safety-driven design (physical, operational, organizational)
• More comprehensive accident/incident investigation and root cause analysis
• Organizational and cultural risk analysis
  – Identifying physical and project risks
  – Defining safety metrics and performance audits
  – Designing and evaluating potential policy and structural improvements
  – Identifying leading indicators of increasing risk
• New holistic approaches to security
Improved Hazard Analysis
STPA: System Theoretic Process Analysis

A new hazard analysis technique that:

• Identifies safety constraints (system and component requirements)

• Identifies causes of hazards including
  – Hardware failure
  – Unsafe software behavior
  – Unsafe human action and decision-making
  – Organizational and Management structure flaws
  – Operational flaws
  – Safety Culture
Controller

1. Control input or external information wrong or missing

2. Inadequate Control Algorithm
   (Flaws in creation, process changes, incorrect modification or adaptation)

3. Process Model inconsistent, incomplete, or incorrect

Actuator

4. Inadequate operation

Sensor

3. Inadequate Operation

5. Missing or wrong communication with another controller

Controlled Process

4. Component failures
   Changes over time

Inappropriate, ineffective or missing control action

Delayed operation

Inadequate or missing feedback

Feedback Delays

Incorrect or no information provided

Measurement inaccuracies

Feedback delays

Process input missing or wrong

Unidentified or out-of-range disturbance

Process output contributes to system hazard
Ballistic Missile Defense System (BMDS) Non-Advocate Safety Assessment using STPA

- A layered defense to defeat all ranges of threats in all phases of flight (boost, mid-course, and terminal)

- Made up of many existing systems (BMDS Element)
  - Early warning radars
  - Aegis
  - Ground-Based Midcourse Defense (GMD)
  - Command and Control Battle Management and Communications (C2BMC)
  - Others

- MDA used STPA to evaluate (prior to deployment and test) the residual safety risk of inadvertent launch
Safety Control Structure Diagram for FMIS
Results

• Deployment and testing of BMDS held up for 6 months because so many scenarios identified for inadvertent launch.

• In many of these scenarios:
  – All components were operating exactly as intended
    • Missing cases in software, obscure timing interactions
    • Could not be found by fault trees or other standard techniques
  – Complexity of component interactions led to unanticipated system behavior

• STPA also identified component failures that could cause inadvertent launch (most analysis techniques consider only these failure events)
Results (2)

- A lot of money spent to fix problems found
- Now being used proactively for upgrades
  - As changes are made to the system, the differences are assessed by updating the control structure diagrams and assessment analysis templates.
HTV

• Joint research project between MIT and JAXA to determine feasibility and usefulness of STPA for JAXA projects

• Comparison between STPA and FTA for HTV
Comparison between STPA and FTA

- ISS component failures
- Crew mistakes in operation
- Crew process model inconsistent

- Activation missing/inappropriate
- Activation delayed

- HTV component failures
- HTV state changes over time
- Out-of-range radio disturbance
- Physical disturbance

- $t_x$ feedback missing/inadequate
- $t_x$ feedback delayed
- $t_x$ feedback incorrect
- Flight Mode feedback missing/inadequate
- Flight Mode feedback incorrect
- Visual Monitoring feedback missing/inadequate

- Wrong information/directive from JAXA/NASA GS

Identified by both (STPA and FTA)
Identified by STPA only
Additional Uses for STAMP

• Safety-driven design
  – Instead of designing and then performing hazard analysis, use STPA to guide design process

• Integration of safety into system engineering
  – “Intent specifications”
  – Embed safety analysis and information within system engineering process and specifications

• CAST: Improved accident investigation and causal analysis

• Safety during operations

• Improved management decision-making and safety culture
Does it work? Is it practical?

Technical

• Safety analysis of new missile defense system (MDA)
• Safety-driven design of new JPL outer planets explorer
• Safety analysis of the JAXA HTV (unmanned cargo spacecraft to ISS)
• Incorporating risk into early trade studies (NASA Constellation)
• Orion spacecraft (Space Shuttle replacement)
• Automobiles (Intelligent Cruise Control)
• NextGen (for NASA, just starting)
• Accident/incident analysis (aircraft, petrochemical plants, air traffic control, railway accident, …)
Does it work? Is it practical?

Social and Managerial

• Analysis of the management structure of the space shuttle program (post-Columbia)

• Risk management in the development of NASA’s new manned space program (Constellation)

• NASA Mission control — re-planning and changing mission control procedures safely

• Food safety

• Safety in pharmaceutical drug development

• Risk analysis of outpatient GI surgery at Beth Israel Deaconess Hospital

• Analysis and prevention of corporate fraud
Conclusions

• Traditional safety engineering techniques are based on assumptions no longer true for the systems we are building

• Trying to add software and human error to traditional accident models and techniques is hopeless

• A new, more sophisticated causality model is needed to handle the new causes of accidents and the complexity in our modern systems

• Using STAMP, we can create much more powerful and effective safety engineering tools and techniques
Nancy Leveson, *Engineering a Safer World*

http://sunnyday.mit.edu/safer-world

(to be published by MIT Press, 2011)