Chapter 3

Safety and Security Are Two Sides of the Same Coin

Nancy Leveson

Abstract Whether safety and security overlap depends on how one defines each of these qualities. Definitions are man-made and the definer can include or exclude anything they want. The question really is what the definitions imply for the solution of the problems being defined and which definitions lead to the most effective achievement of the property or properties being defined. This chapter starts by proposing an inclusive definition that combines safety and security and then discusses the practical implications of this definition for solving our safety and security problems. These implications include (1) recognizing that safety and security are not equivalent to reliability, (2) broadening the focus in security from information security and keeping intruders out, and (3) creating new integrated analysis methods based on system theory.

Keywords Safety · Security · STAMP · STPA

3.1 Definitions Are Boring But Necessary

Definitions of the terms we use are necessary for effective communications. There is no right or wrong definition, only the one we choose to use. If we limit our definition of the terms “safety” and “security”, then we can effectively limit any overlap. Limited definitions, however, may also limit potential solutions to the problems. If we start from more inclusive and practical definitions, then overlap and common approaches to achieving the properties are possible.

Safety has been a part of engineering for at least 100 years and has been a concern to societies for much longer than that. Those in engineering use a precise definition of the term, while others, in social sciences, for example, tend to use much less carefully crafted definitions and sometimes change the definition depending on local context or goals. The definition of safety also differs among industries. Some limit safety and accidents to include only those events that impact on human life and injury.

N. Leveson (✉)

Department of Aeronautics and Astronautics, MIT, Cambridge, USA

e-mail: leveson@mit.edu

© The Author(s) 2020

C. Bieder and K. Pettersen Gould (eds.), *The Coupling of Safety and Security*, SpringerBriefs in Safety Management,
https://doi.org/10.1007/978-3-030-47229-0_3

Commercial aviation has historically defined safety in terms of aircraft hull losses. Some industries, such as nuclear power, which have serious political concerns, have proposed politically useful definitions but ones that are almost useless in engineering design.

The most inclusive definition, started in the U.S. Defense industry after WWII, is the one used in this chapter:

Definition Safety is freedom from accidents (losses).

Definition An accident/mishap is any undesired or unplanned event that results in a loss, as defined by the system stakeholders.

Losses may include loss of human life or injury, equipment or property damage, environmental pollution, mission loss (non-fulfillment of mission), negative business impact (e.g., damage to reputation, product launch delay, legal entanglements), etc. There is nothing in the definition that distinguishes between inadvertent and intentional causes. In fact, the definition does not limit the causes in any way. So security is included in the definition.

The concept of a *hazard* is critical in safety engineering.

Definition A hazard is a system state or set of conditions that, together with some (worst-case) environmental conditions, will lead to a loss.

Note that hazards are defined in safety engineering as states of the system, not the environment. The ultimate goal of safety engineering is to eliminate losses, but some of the conditions that lead to a loss may not be under the control of the designer or operator, i.e., they are outside the boundary of the designed and operated system. So for practical reasons, hazards are defined as system states that the designers and operators never want to occur and thus try to eliminate or, if that is not possible, control. Although the term “hazard” is sometimes loosely used to refer to things outside the system boundaries, such as inclement weather or high mountains in aviation, hazards in safety engineering are limited to system states that are within the control of the system designer. In this example, the hazard is not the inclement weather or mountain, but rather it is the aircraft being negatively impacted by inclement weather or the aircraft violating minimum separation from the mountain. We cannot eliminate the weather or the mountain, but we can control the design and operation of our system to eliminate the threat posed by the weather or mountain. Constraints or controls may involve designing the aircraft to withstand the impact of the weather or it may involve operational controls such as staying clear of the weather or the mountain. Thus, the goal of the designers and operators is to identify the system hazards (defined as under the control of the designers) and eliminate or control them in the design and operation of the system.

In security, the equivalent term for a hazard is a vulnerability, i.e., a weakness in a product that leaves it open to a loss. In the most general sense, security can be defined in terms of the system state being free from threats or vulnerabilities, i.e., potential losses. Here hazard and vulnerability are basically equivalent.

Definition Hazard analysis is the process of identifying the causal scenarios of hazards.

While hazard analysis usually only considers scenarios made up of inadvertent events, including security requires only adding a few extra causal scenarios in the hazard analysis process. This addition will provide all the information needed to prevent losses usually considered as security problems. For example, the cause of an operator doing the wrong thing might be that he or she is inadvertently confused about the state of the system, such as thinking that a valve is already closed and therefore, not closing it when required. That misinformation may result from a sensor failure that provides the wrong information or may result from a hostile actor purposely providing false information. These considerations result in adding during the analysis more paths to get to the hazardous state (which must be dealt with in design or operations), but do not necessarily change the way the designer or operator of the system attempts to prevent that operator unsafe behavior (see the Stuxnet example below).

Definition The goal of safety engineering is to eliminate or control hazard scenarios in design and operations.

The difference between physical security and cybersecurity is irrelevant except that cybersecurity focuses on only one aspect of the system design and thus has a more limited scope. Physical system security now almost always includes software components and thus cybersecurity is usually a component of physical system security.

3.2 Safety and Security Are Not Equal to Reliability

There has been much confusion between safety and reliability, which are two very different qualities.¹ When systems were relatively simple, were made up solely of electromechanical parts, and could be exhaustively analyzed or tested, design errors leading to a loss could be identified and eliminated for the most part before the system was fielded and used: the remaining causes of losses were primarily physical failures. The traditional hazard analysis techniques (which are used to identify the potential causes of the system hazards), such as fault tree analysis, HAZOP (in the chemical industry), event tree analysis (in the nuclear industry), FMECA (failure modes and criticality analysis) all stem from this era, which includes the 1970s and before. For these relatively simple electromechanical systems, reliability of the components was a convenient proxy for safety, as most accidents resulted from component failure. Therefore, the analysis techniques were designed to identify the component failures that can lead to a loss.

¹It is not possible to discuss this confusion in depth in this short paper. The reader is referred to Nancy Leveson, *Engineering a Safer World*, MIT Press, 2012 [1] for an extensive discussion.

Since the introduction of computer controls and software in critical systems starting around 1980, system complexity has been increasing exponentially. The bottom line is that system design errors (i.e., system engineering errors) cannot be eliminated prior to use and are an important cause of accidents in systems today. There is also increased recognition that losses can be related to human factors design, management, operational procedures, regulatory and social factors, and changes within the system or in its environment over time. This is true for both safety and security. System components can be perfectly reliable (they can satisfy their stated requirements and thus do not fail), but accidents can (and often do) occur. Alternatively, system components and indeed the system itself can be unreliable, and the system can still be safe. Defining safety or security in terms of reliability does not work for today's engineered systems. Losses are not prevented by simply preventing system or component failures.

3.3 We Need to Broaden the Focus from Information Security and Keeping Intruders Out

Too often the focus in security, particularly cybersecurity, is on protecting information. But there are important losses that do not involve information that are for the most part being ignored. These losses involve mission assurance. The loss of power production from the electrical grid or a nuclear power plant or the loss of the scientific mission for a spacecraft is just as important (and in some respects more important) than the loss of information. In addition, given that it has proven virtually impossible to keep people out of systems, particularly cyber systems that are connected to the outside world, preventing those with malicious intentions from entering our systems does not appear to be an effective way to solve the security problem.

While intentionality does differ between safety and security, intentionality is not very important when analyzing safety and security and preventing losses. That difference is irrelevant from a safety engineering perspective when the consequences are the same. Whether the explosion of a chemical plant, for example, is the result of an intentional act or an inadvertent one, the result is the same, i.e., harmful to both the system and the environment. Intentionality simply adds some additional causal scenarios to the hazard analysis. The techniques used to identify and prevent causal scenarios for both can be identical.

As an example, consider the Stuxnet worm that targeted the Iranian nuclear program. In this case, the loss was *Damage to the reactor* (specifically, the centrifuges). The hazard/vulnerability was that *the centrifuges are damaged by spinning too fast*. The constraint that needed to be enforced was that *the centrifuges must never spin above a maximum rate*. The hazardous control action that occurred was *issuing an increase speed command when the centrifuges are already spinning at the maximum speed*. One potential causal scenario is that the operator/software controller thought that the centrifuges were spinning at less than the maximum speed. This mistake

could be inadvertent (a human or software error) or (as in this case) deliberate. But no matter which it was, the most effective potential controls for both cases are the same and include such designs as using a mechanical limiter (interlock) to prevent excess spin rate or an analog RPM gauge.

Note that security concerns need not start from outside the system: Security breaches can actually start from inside the system and the results can wreak havoc on the environment.

3.4 More Effective Approaches to Safety and Security Require a Paradigm Change

Finding more effective solutions to safety and security problems requires reconsidering the foundation on which the current solutions rest, that is, the models of causality that we assume are at the root of safety and security problems. Traditionally, accidents or losses are seen as resulting from a chain of failure events, where A fails and causes the failure of B and so on until the loss occurs. This model (called the Domino or, more recently, the Swiss cheese model of accident causation) has been around a very long time, but our engineered systems are very different than those that existed previously. The model no longer accounts for all the causes of accidents today.

To find more effective solutions to safety and security problems requires a paradigm change to a model of causality based on system theory. System Theory arose around the middle of the last century to deal with the increased complexity of the systems we were creating.

A new, more inclusive model of accident causality based on system theory is STAMP (System-Theoretic Accident Model and Processes) [1]. Instead of treating accidents as simply the result of chains of failure events, STAMP treats safety and security as a dynamic control problem where the goal is to enforce constraints on the behavior of the system as a whole, including individual component behavior as well as the interactions among the system components. In the Stuxnet example, the required system constraint was to control the rotational speed of the centrifuges to reduce wear. Other example constraints might be that minimum separation is maintained between aircraft and automobiles, that chemicals or radiation is never released from a plant, that workers must not be exposed to workplace hazards, or that a bomb must never be detonated without positive action by an authorized person. STAMP basically extends the traditional causality model to include more than just failures.

STAMP is just a theoretical model. On top of that model, a variety of new (and more powerful) tools can be created. CAST (Causal Analysis based on System Theory) can be used for analyzing the cause of losses that have already occurred. The causes may involve both unintentional and intentional actions. Security-related losses have been analyzed using CAST.

A second tool, STPA (System-Theoretic Process Analysis), can be used to identify the potential causes of losses that have not yet occurred but could in the future, i.e., to perform hazard analysis by identifying loss scenarios [2]. The potential causes of future accidents identified by STPA provide information about design and operation that system designers and operators can use to eliminate or control the identified causal scenarios.

To give the reader some feeling for what is produced by STPA and how safety and security are handled in an integrated manner, consider an aircraft ground braking system. The system-level deceleration hazards might be defined as:

- H-4.1 Deceleration is insufficient upon landing, rejected takeoff, or during taxiing
- H-4.2 Asymmetric deceleration maneuvers aircraft toward other objects
- H-4.3 Deceleration occurs after V1 point during takeoff.

The V1 point is the point where braking during takeoff is dangerous and it is safer to continue the takeoff than to abort it.

STPA is performed on a functional model of the system. An example is shown in Fig. 3.1 where the Flight Crew (humans) control the Brake System Control Unit (BSCU), which is composed of an autobrake controller and a hydraulic controller, both of which will be composed of a significant amount of software in today’s aircraft. The BSCU controls the Hydraulic Controller, which actually provides the physical commands to the aircraft wheel brakes. The Flight Crew can also send commands directly to the hydraulic braking system to decelerate the aircraft.

Fig. 3.1 Functional control structure of the wheel braking system

STPA is performed on this control structure. The analysis starts in the same way for both safety and security, i.e., nothing additional is needed to handle security until the end of the process. First, the potential unsafe/insecure control actions are identified. A small example is shown in Table 3.1 for the BSCU *Brake* control action. The table contains the conditions under which this control action could lead to a system hazard (H-4.1 in this partial example). Table 3.2 shows that the same process can be performed for the humans in the system, they are treated the same as any system component.

The next step is to identify the scenarios that can lead to these unsafe control actions. The scenarios will include the normal failure scenarios identified by the traditional hazard analysis techniques such as FTA, FMECA, HAZOP, but almost always more than they produce. UCA-1 in Table 3.1 is “BSCU Autobrake does not provide the Brake control action during landing roll when the BSCU is armed [H-4.1]”. Because pilots may be busy during touchdown, this example braking system allows them to set an automatic braking action (autobrake) to brake after touchdown occurs. UCA-1 in Table 3.1 is that the autobrake does not activate when it has been set. The hazard analysis goal then is to identify the reasons why this unsafe control

Table 3.1 Examples of unsafe control actions for the BSCU (partial example)

Control action	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long
Brake	UCA-1: BSCU Autobrake does not provide the Brake control action during landing roll when the BSCU is armed [H-4.1]	UCA-2: BSCU Autobrake provides Brake control action during a normal takeoff [H-4.3, H-4.6] UCA-5: BSCU Autobrake provides Brake control action with an insufficient level of braking during landing roll [H-4.1] UCA-6: BSCU Autobrake provides Brake control action with directional or asymmetrical braking during landing roll [H-4.1, H-4.2]	UCA-3: BSCU Autobrake provides the Brake control action too late (>TBD seconds) after touchdown [H-4.1]	UCA-4: BSCU Autobrake stops providing the Brake control action too early (before TBD taxi speed attained) when aircraft lands [H-4.1]

Table 3.2 Example unsafe control actions for the flight crew (partial example)

Control action	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long
Power Off BSCU	UCA-1: Crew does not provide BSCU Power Off when abnormal WBS behavior occurs [H-4.1, H-4.4, H-7]	UCA-2: Crew provides BSCU Power Off when Anti-Skid functionality is needed and WBS is functioning normally [H-4.1, H-7]	Crew powers off BSCU too early before Autobrake or Anti-Skid behavior is completed when it is needed [H-4.1, H-7]	N/A

action could occur. The scenarios can be used to create safety/security requirements and to design the scenarios out of the system.

- Scenario 1: UCA-1 could occur if the BSCU incorrectly believes the aircraft has already come to a stop. One possible reason for this flawed belief is that the received feedback momentarily indicates zero speed during landing roll. The received feedback may momentarily indicate zero speed during anti-skid operation, even though the aircraft is not stopped.
- Scenario 2: The BSCU is armed and the aircraft begins the landing roll. The BSCU does not provide the brake control action because the BSCU incorrectly believes the aircraft is in the air and has not touched down. This flawed belief will occur if the touchdown indication is not received upon touchdown. The touchdown indication may not be received when needed if any of the following occur:
 - Wheels hydroplane due to a wet runway (insufficient wheel speed),
 - Wheel speed or weight on wheel feedback is delayed due to the filtering used,
 - Conflicting air/ground indications due to crosswind landing,
 - Failure of wheel speed sensors,
 - Failure of air/ground switches,
 - Etc.

As a result, insufficient deceleration may be provided upon landing [H4.1].

To include causes related to security, only one additional possibility needs to be considered: identify how the scenarios, for example, the specified feedback and other information, could be affected by an adversary. More specifically, how could feedback and other information be injected, spoofed, tampered, intercepted, or disclosed to an adversary? The following causes might be added to the scenario above to include security:

- Adversary spoofs feedback indicating insufficient wheel speed
- Wheel speed is delayed due to adversary performing a DoS (Denial of Service) attack
- Correct wheel speed feedback is intercepted and blocked by an adversary
- Adversary disables power to the wheel speed sensors.

Scenarios must also be created for the situations where a correct and safe control action is provided but it is not executed. In our example, the BSCU sends the brake command but the brakes are not applied. Some example scenarios for this case are as follows:

- Scenario 3: The BSCU sends a Brake command, but the brakes are not applied because the wheel braking system was previously commanded into an alternate braking mode (bypassing the BSCU). As a result, insufficient deceleration may be provided upon landing [H-4.1].
- Scenario 4: The BSCU sends Brake command, but the brakes are not applied due to insufficient hydraulic pressure (pump failure, hydraulic leak, etc.). As a result, insufficient deceleration may be provided upon landing [H-4.1].
- Scenario 5: The BSCU sends Brake command, the brakes are applied, but the aircraft does not decelerate due to a wet runway (wheels hydroplane). As a result, insufficient deceleration may be provided upon landing [H-4.1].

Again, to include security issues, the same additional possibilities need to be considered, i.e., identify how adversaries can interact with the control process to cause the unsafe control actions. For example,

- Scenario 6: The BSCU sends Brake command, but the brakes are not applied because an adversary injected a command that puts the wheel braking system into an alternate braking mode. As a result, insufficient deceleration may be provided upon landing [H-4.1].

STPA can handle humans in the same way it handles hardware and software. Table 3.2 shows an example of the crew responsibility to power off the BSCU. As one simple example,

Crew-UCA-1	Crew does not provide BSCU Power Off when abnormal WBS behavior occurs [H-4.1, H-4.4].
Scenario 1 for Crew-UCA-1	Abnormal WBS behavior occurs and a BSCU fault indication is provided to the crew. The crew does not power off the BSCU [Crew-UCA-1] because the operating procedures did not specify that the crew must power off the BSCU upon receiving a BSCU fault indication.

Sophisticated human factors considerations can be included here, but this topic is beyond the scope of this chapter.

3.5 What Can We Conclude from This Argument?

Safety and security can be considered using a common approach and integrated analysis process if safety and security are defined appropriately [3]. The definitions commonly used in the defense industry provide this facility. Other limitations in how we handle these properties also need to be removed to accelerate success in achieving these two properties, which are really just two sides of the same coin:

- Safety analysis needs to be extended beyond reliability analysis,
- Security has to be broadened beyond the current limited focus on information security and keeping intruders out, and
- A paradigm change is needed to go beyond accidents considered to be a chain of failure events and basing our hazard analysis techniques on reliability theory to one where our accident causality models and hazard analysis techniques are based on system theory.

Will these changes provide greater success? The system-theoretic approach to safety engineering and the related integrated approach to safety and security have been experimentally compared with current approaches many times and empirically compared by companies on their own systems. In all comparisons (now numbering a hundred or so), the system-theoretic and integrated approaches have proven superior to the traditional approaches. They are currently being used on critical systems around the world and in almost every industry, but particularly in automobiles and aviation where autonomy is advancing quickly.

References

1. N.G. Leveson, *Engineering a Safer World* (MIT Press, 2012)
2. N.G. Leveson, J.P. Thomas, *STPA Handbook* (2018) http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf. Accessed 22 Sept 2018
3. W. Young, N.G. Leveson, An integrated approach to safety and security based on systems theory. *Commun. ACM* **57**(2), 31–35 (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.