Chapter 4

Definitions for System Safety Analysis

4.1 Reliability and Safety

Reliability and Failure We have talked about failure, and inferring from failure of a system to failure of parts. But the failure of a system to fulfil its function, and the success of a system in filling its function, are not directly related to safety. If we install our LAN server in a fireproof room, and there are no essential functions of the company which depend on the computer functioning, then whether my LAN server fulfils its function most of the time or hardly at all is not a safety matter. *Reliability* is the property of a system whereby it fulfils its function. A firearm may reliably fire when the trigger is pulled; but if it's loaded and a child is playing with it, and there is no safety catch, it may reliably fire and kill someone.

Safety and Accidents The property of a system whereby it does not produce or encourage accidents is known as *safety*. An *accident* is taken to be any undesired or unwanted (but not necessarily unexpected) behavior. Definitions are taken from [Lev95]. This means that an accident can be almost anything you want it to be. Usually, we are concerned whether the operation of a system will kill or injure humans or other animals, but little in safety engineering techniques actually depends on whether this particular unwanted behavior is what one is considering to be an accident.

Reliability and Safety are Related However, situations such as just mentioned can be moderated by the introduction of safety mechanisms. For example, a trigger lock, which prevents the firearm being fired by anyone other than the keyholder. In order for the device to continue to function safely in these circumstances, the safety mechanism must be reliable. This is the most frequent

connection between safety and reliability: safety is assured through the reliable operation of certain mechanisms.

Safety Mechanisms Safety is, roughly speaking, the *absence* of certain kinds of problems. Often, this absence is assured, or we attempt to assure it, through the *presence* of specific mechanisms, which are intended to inhibit rare but possible unsafe system behaviors. These systems must function reliably in order to ensure safety. But they are hardly ever used; just on the rare occasions when there would be a safety problem which triggers their operation. It is notoriously hard to ensure the reliable operation of a mechanism which is rarely used. Ensuring the reliability of safety mechanisms is often a much harder engineering problem than redesigning a system to avoid the potential safety problem without the use of specific mechanisms.

4.2 Definitions of Safety Concepts

Terminology Leveson notes that terminology in system safety has not always been used consistently [Lev95, p171]. She gives a series of definitions of such terms as *reliability*, *failure*, *error*, *accident*, *incident*, *hazard*, *risk* and *safety* [Lev95, Chapter 9: Terminology], which attempts to do the most justice to the engineering definitions, and is the result of considerable research into the engineering literature over a number of years. These definitions indeed seem to be amongst the most precise in the literature.

Reliability Leveson defines [Lev95, p172]:

> **Reliability** is the probability that a piece of equipment or component [of a system] will perform its intended function satisfactorily for a prescribed time and under stipulated environmental conditions.

Failure [Lev95, p172]:

> **Failure** is the nonperformance or inability of the system or component to perform its intended function for a specified time under specified environmental conditions.

Accidents and Safety [Lev95, pp172,181]:

> An **accident** is an undesired and unplanned (but not necessarily unexpected) event that results in (at least) a specified level of loss. [....]
> **Safety** is freedom from accidents or losses.

In order to use these definitions, one has to specify what one considers to be losses (and their levels). Such losses are often specified as numbers of deaths or injuries, financial losses to concerned parties, damage to the natural environment, and so forth. Typically, there is considerable agreement on what is to be considered a 'loss' (for example, deaths, injuries, money, damage), and how the levels are measured (mostly by numbers; more generally on ordinal or ratio scales [KLST71]). Leveson notes that this is stipulatory: it is up to us to specify what we consider a loss and what levels constitute an accident.

Accidents and the System Boundary There is nothing in the definition of accident concerning the system boundary; we may presume that many accidents involving both system and environment occur. Examples could be: the airplane crumples and dismembers, because the mountain rose through the cloud to smite it. When dealing with teleological systems, we may be presumed to be able to exercise more control over the constitution and behavior of the system than we may over the environment. We shall see that, depending on the openness of the system and various other factors, accidents may depend more or less on the interaction of the system with its environment.

System Contributions to an Accident The aircraft can be engineered to predict the looming presence of the mountain and fly above it; it is considerably harder to move the mountain out of the way of the encounter. Accordingly, we shall wish to speak about the part of the system that contributes to an accident, even though given favorable environmental conditions the accident will not occur: if the aircraft flies at or above a (true) 30,000ft (above mean sea level, MSL) altitude, there will be no mountain for it to encounter; if it flies through the Himalayas below 28,000ft MSL, there are some places it cannot fly without meeting an obstacle. Accordingly, we can distinguish airspace including an altitude of less than 28,000ft MSL over the Himalayas as hazardous, potentially leading to an controlled-flight-into-terrain (CFIT) accident, and other airspace as non-hazardous. The property of being hazardous or not has thereby been ascribed to the airspace, that is, part of the environment. However, there is a corresponding pair of properties of the aircraft, namely *being in/out of hazardous airspace*. One may wonder after considering this example whether hazards can be always be described either through environmental properties or through system properties, as desired. If so, there are reasons to classify system states and not environment states as hazards, namely that one brings them into the domain in which control and redesign can be exercise if necessary. But we shall see later that system and environmental hazard states are not always dual in this manner.

Hazard, Severity, and Risk The following definitions are said to be standard in U.S. System Safety engineering [Lev95, pp177-9]:

> *A **hazard** is a state or set of conditions of a system (or an object) that, together with other conditions in the environment of the system (or object), will lead inevitably to an accident (loss event). [....] A hazard is defined with respect to the environment of the system or component. [....] What constitutes a hazard depends upon where the boundaries of the system are drawn. [....] A hazard has two important characteristics: (1) **severity** (sometimes called **damage**) and (2) **likelihood** of occurrence. Hazard **severity** is defined as the worst possible accident that could result from the hazard given the environment in its most unfavorable state. [....] The combination of severity and likelihood of occurrence is often called the **hazard level**. [....] **Risk** is the hazard level combined with (1) the likelihood of the hazard leading to an accident (sometimes called **danger**) and (2) hazard exposure or duration (sometimes called **latency**).*

So a hazard, flying under 28,000ft MSL, in combination with other conditions in the environment (doing so in a particular direction in a particular geographical location, so that impact cannot be avoided) will inevitably lead to an accident (loss of airplane and death or injury of occupants) that may be more or less severe, depending on how many people on board there are, how expensive the aircraft is, what environmental damage is sustained, and so on. We shall later call this notion of hazard *Hazard-1*, to distinguish it from three other useful formulations of the concept.

The Concept of Hazard Partitions States It is important to note that this concept of hazard divides states of the system into two classes, consisting respectively of those states in which the aircraft is flying at an altitude greater than that of the obstructions in the vicinity; and of those in which the aircraft is flying at or below that altitude. The first category of states will not (because they cannot) lead to a CFIT accident, and states in the second category allow the potential for that kind of accident. Accordingly, the states in the second category are hazard states for CFIT, and those in the first category are not.

To take another example: an aircraft flying through cloud with the potential for embedded thunderstorms actually encounters one. The hazard consists in flying through cloud with embedded thunderstorms (rather than flying clear of such weather); the severity is loss of the aircraft and occupants; the 'most unfavorable state' of the environment is a thunderstorm of sufficient power to upset the aircraft and cause breakup under aerodynamic loads; the danger is how likely one is to fly through such a thunderstorm while flying through the stormclouds; and the duration is the length of time one flies through the stormclouds. One could presumably measure the relevant probabilities (likelihood and danger) by measuring the spatial distribution of thunderstorms in stormclouds of the given type, and the frequency of severe ones. All well and good. But do these concepts

work generally?