9

Infrastructure Risk Analysis and Security

Bilal M. Ayyub
University of Maryland

Massoud Amin
University Minnesota

9.1 Infrastructure Risk Analysis and Management

Bilal M. Ayyub

9.1.1 Introduction

Risk is associated with all projects and business ventures taken by individuals and organizations regardless of their sizes, their natures, and their time and place of execution and utilization. Risk is present in various forms and levels even in small domestic projects such as adding a deck to a residential house, and in large multibillion-dollar projects such as developing and producing a space shuttle. These risks could result in significant budget overruns, delivery delays, failures, financial losses, environmental damages, and even injuries and loss of life. Risks are taken even though they could lead to devastating consequences because of potential benefits, rewards, survival, and future return on investment. The chapter defines and discusses risk and its dimensions, risk analysis, risk management and control, and risk communication.

9.1.2 Risk Terminology

Definitions that are needed for presenting risk-based technology methods and analytical tools are presented in this section.

9.1.2.1 Hazards

A hazard is an act or phenomenon posing potential harm to some person(s) or thing(s), i.e., a source of harm, and its potential consequences. For example, uncontrolled fire is a hazard, water can be a hazard, and strong wind is a hazard. For the hazard to cause harm, it needs to interact with person(s) or thing(s) in a harmful manner. The magnitude of the hazard is the amount of harm that might result, including the seriousness and the exposure levels of people and the environment. Hazards need to be identified and

considered in projects' life cycle analyses because they could pose threats and could lead to project failures.

The interaction between a person (or a system) and a hazard can be voluntary or involuntary. For example, exposing a marine vessel to a sea environment might lead to its interaction with extreme waves in an uncontrollable manner, i.e., an involuntary manner. Although the decision of a navigator of the vessel to go through a storm system that is developing can be viewed as a voluntary act in nature, and might be needed to meet schedule constraints or other constraints, the potential rewards of delivery of shipment or avoidance of delay charges offer an incentive that warrants such an interaction. Other examples can be constructed where individuals interact with hazards for potential financial rewards, fame, and self-fulfillment and satisfaction ranging from investment undertaking to climbing cliffs.

9.1.2.2 Reliability

Reliability can be defined for a system or a component as its ability to fulfill its design functions under designated operating or environmental conditions for a specified time period. This ability is commonly measured using probabilities. Reliability is, therefore, the occurrence probability of the complementary event to failure as provided in the following expression:

$$\text{Reliability} = 1 - \textit{Failure Probability} \tag{9.1}$$

9.1.2.3 Event Consequences

For an event of failure, *consequences* can be defined as the degree of damage or loss from some failure. Each failure of a system has some consequence(s). A failure could cause economic damage, environmental damage, injury, loss of human life, or other possible events. Consequences need to be quantified in terms of failure consequence severities using relative or absolute measures for various consequence types to facilitate risk analysis.

For an event of success, consequences can be defined as the degree of reward or return or benefits from success. Such an event could cause economic outcomes, environmental effects, or other possible events. Consequences need to be quantified using relative or absolute measures for various consequence types to facilitate risk analysis.

9.1.2.4 Risk

The concept of risk can be linked to uncertainties associated with events. Within the context of projects, risk is commonly associated with an uncertain event or condition that, if it occurs, has a positive or a negative effect on a project's objectives.

Risk originates from the Latin term *risicum*, meaning the challenge presented by a barrier reef to a sailor. The Oxford dictionary defines risk as the chance of hazard, bad consequence, loss, etc. Also, risk is the chance of a negative outcome. To measure risk, one must accordingly assess both of its defining components, the chance, its negativity, and potential rewards or benefits. Estimation of risk is usually based on the expected result of the conditional probability of the event occurring times the consequence of the event given that is has occurred.

A risk results from an event or sequence of events called a *scenario*. The event or scenario can be viewed as a cause and, if it occurs, result in consequences with severities. For example, an event or cause may be shortage of personnel needed to perform a task needed to produce a project. The event in this case of personnel shortage for the task will lead to a consequence on the project cost, schedule, and/or quality. The events can reside in the project environment that may contribute to project success or failure, such as project management practices, or external partners or subcontractors.

Risk has certain characteristics that should be used in the risk assessment process. Risk is a characteristic of an uncertain future, and is neither a characteristic of the present nor the past. After uncertainties are resolved and/or the future is attained, the risk becomes nonexistent. Therefore, risks cannot be described for historical events or for events that are currently being realized. Similarly, risks cannot be directly associated with a success. Although risk management through risk mitigation of

selected events could result in project success leading to rewards and benefits, these rewards and benefits cannot be considered as outcomes only of the nonoccurrence of these events associated with the risks. The occurrence of risk events leads to adverse consequences that are clearly associated with their occurrence; however, their nonoccurrences are partial contributors to the project success that lead to rewards and benefits. The credit in the form of rewards and benefits cannot be given solely to the nonoccurrence of these risk events. Some risk assessment literature defines risk to include both potential losses and rewards. They need to be treated separately as (1) risks leading to adverse consequences, and (2) risks that contribute to benefits or rewards in trade-off analyses. An appropriate risk definition is this context is a threat (or opportunity) that could affect adversely (or favorably) achievement of the objectives of a project and its outcomes.

Developing an economic, analytical framework for a decision situation involving risks requires examining the economic and finance environments of a project. This environment could have significant impacts on the occurrence probabilities of events associated with risks. This complexity might be needed for certain projects in order to obtain justifiable and realistic results. The role of such an environment in risk analysis is discussed in subsequent sections.

Formally, risk can be defined as the potential of losses and rewards resulting from an exposure to a hazard or as a result of a risk event. Risk should be based on identified risk events or event scenarios. Risk can be viewed to be a multidimensional quantity that includes event occurrence probability, event occurrence consequences, consequence significance, and the population at risk; however, it is commonly measured as a pair of the probability of occurrence of an event, and the outcomes or consequences associated with the event's occurrence. This pairing can be represented by the following equation:

$$Risk \equiv [(p_1, c_1), (p_2, c_2), ..., (p_i, c_i), ..., (p_n, c_n)], \tag{9.2}$$

where p_i is the occurrence probability of an outcome or event i out of n possible events, and c_i is the occurrence consequences or outcomes of the event. A generalized definition of risk can be expressed as

$$Risk \equiv [(l_1, o_1, u_1, cs_1, po_1), (l_2, o_2, u_2, cs_2, po_2), ..., (l_n, o_n, u_n cs_n po_n)], \tag{9.3}$$

where l is likelihood, o is outcome, u is utility (or significance), cs is causal scenario, po is population affected by the outcome, and n is the number of outcomes. The definition according to Equation 9.3 covers all attributes measured in risk assessment that are described in this chapter, and offers a complete description of risk, from the causing event to the affected population and consequences. The population size effect should be considered in risk studies because society responds differently for risks associated with a large population in comparison to a small population. For example, a fatality rate of 1 in 100,000 per event for an affected population of 10 results in an expected fatality of 10^{-4} per event, whereas the same fatality rate per event for an affected population of 10,000,000 results in an expected fatality of 100 per event. Although the impact of the two scenarios might be the same on the society (same risk value), the total number of fatalities per event/accident is a factor in risk acceptance. Plane travel may be "safer" than for example recreational boating, but 200–300 injuries per accident are less acceptable to society. Therefore, the size of the population at risk and the number of fatalities per event should be considered as factors in setting acceptable risk.

Risk is commonly evaluated as the product of likelihood of occurrence and the impact severity of occurrence of the event:

$$RISK\left(\frac{Consequence}{Time}\right) = LIKELIHOOD\left(\frac{Event}{Time}\right) \times IMPACT\left(\frac{Consequence}{Event}\right) \tag{9.4}$$

In Equation 9.4, the likelihood can also be expressed as a probability. Equation 9.4 presents risk as an expected value of loss or an average loss. A plot of occurrence probabilities and consequences is called a *risk profile* or a *Farmer curve*. An example Farmer curve is given in Figure 9.1 based on a nuclear case

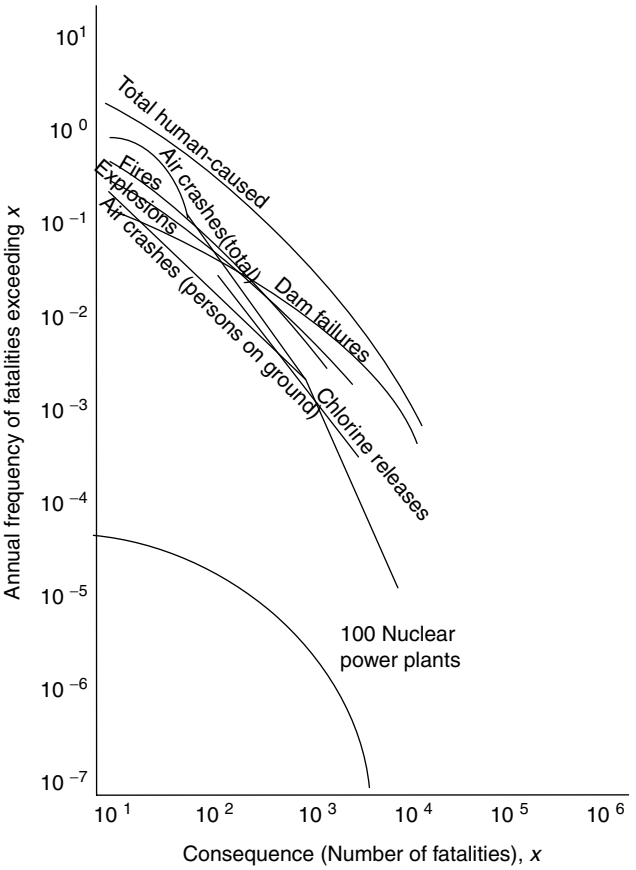

FIGURE 9.1 Example risk profile.

study, provided herein for illustration purposes. It should be noted that the abscissa provides the number of fatalities, and the ordinate provides the annual frequency of exceedence for the corresponding number of fatalities. These curves are sometimes constructed using probabilities instead of frequencies. The curves represent or median average values. Sometimes, bands or ranges are provided to represent uncertainty in these curves. They represent confidence intervals for the average curve or for the risk curve. Figure 9.2 shows examples curves with uncertainty bands. This uncertainty is sometimes called *meta-uncertainty*. A complete treatment of uncertainty analysis is provided by Ayyub and Klir (2006).

The occurrence probability (p) of an outcome (o) can be decomposed into an occurrence probability of an event or threat (t), and the outcome occurrence probability given the occurrence of the event ($o|t$). The occurrence probability of an outcome can be expressed as follows using conditional probability concepts:

$$p(o) = p(t)p(o|t). \tag{9.5}$$

In this context, threat is defined as a hazard or the capability and intention of an adversary to undertake actions that are detrimental to a system or an organization's interest. In this case, threat is a function of only the adversary or competitor, and usually cannot be controlled by the owner or user of the system. However, the adversary's intention to exploit his capability may be encouraged by vulnerability of the system or discouraged by an owner's countermeasures. The probability ($p(o|t)$) can be interpreted as the vulnerability of the system in case of this threat occurrence. Vulnerability is

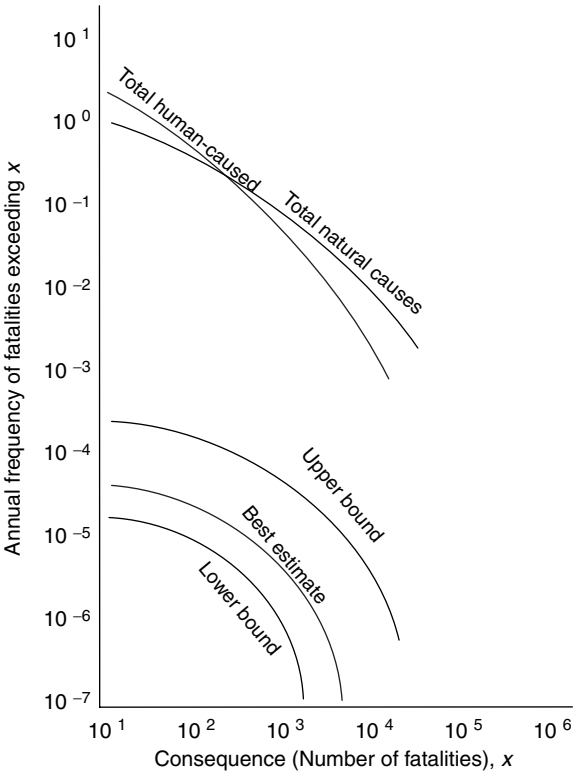

FIGURE 9.2 Uncertain risk profile.

a result of any weakness in the system or countermeasure that can be exploited by an adversary or competitor to cause damage to the system.

9.1.2.5 Performance

The performance of a system or component can be defined as its ability to meet functional requirements. The performance of an item can be described by various elements, including such items as speed, power, reliability, capability, efficiency, and maintainability. The design and operation of the product or system influence performance.

9.1.2.6 Risk-Based Technology

Risk-based technologies (RBT) are methods or tools and processes used to assess and manage the risks of a component or system. RBT methods can be classified into risk management that includes risk assessment/risk analysis and risk control using failure prevention and consequence mitigation, and risk communication as shown in Figure 9.3.

Risk assessment consists of hazard identification, event-probability assessment, and consequence assessment. Risk control requires the definition of acceptable risk and comparative evaluation of options and/or alternatives through monitoring and decision analysis. Risk control also includes failure prevention and consequence mitigation. Risk communication involves perceptions of risk, which depends on the audience targeted, hence, classified into risk communication to the media, the public, and to the engineering community.

9.1.2.7 Safety

Safety can be defined as the judgment of risk acceptability for the system. Safety is a relative term since the decision of risk acceptance may vary depending on the individual making the judgment. Different people

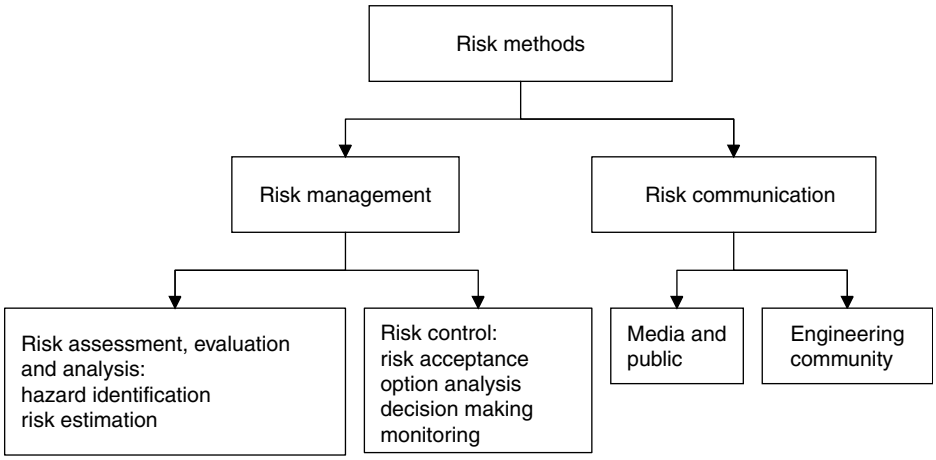

FIGURE 9.3 Risk-based technology methods.

are willing to accept different risks as demonstrated by different factors such as location, method or system type, occupation, and lifestyle. The selection of these different activities demonstrates an individual's safety preference despite a wide range of risk values. Table 9.1 identifies varying annual risks for different activities based on typical exposure times for these activities. Also Figure 9.4 from the Imperial Chemical Industries, Ltd. shows the variation of risk exposure during a typical day that starts by waking up in the morning from sleep and getting ready to go to work, then commuting and working during morning hours, followed by a lunch break, then additional work hours followed by commuting back to having dinner, and round-trip on a motorcycles to a local pub. The ordinate in this figure is the fatal accident frequency rate (FAFR) with a FAFR of 1.0 corresponding to one fatality in 11,415 years, or 87.6 fatalities per one million years. The figure is based on an average number of deaths in 10^8 h of exposure to a particular activity.

Risk perceptions of safety may not reflect the actual level of risk in some activity. Table 9.2 shows the differences in risk perception by three groups of the league of women voters, college students, and experts of 29 risk items. Only the top items are listed in the table. Risk associated with nuclear power was ranked

TABLE 9.1 Relative Risk of Different Activities

Risk of Death	Occupation	Lifestyle	Accidents/Recreation	Environmental Risk
1 in 100	Stunt person			
1 in 1,000	Race car driver Firefighter miner	Smoking (one pack/day)	Skydiving rock climbing snowmobile Canoeing automobile	
1 in 10,000	Farmer police officer	Heavy drinking	All home accidents frequent air travel	
1 in 100,000	Truck driver engineer banker insurance agent	Using contraceptive pills light drinking	Skiing home fire	Substance in drinking water living downstream of a dam
1 in 1,000,000		Diagnostic x-rays smallpox vaccination (per occasion)	Fishing poisoning occasional air travel (one flight per year)	Natural background radiation living at the boundary of a nuclear power
1 in 10,000,000		Eating charcoal-broiled steak (once a week)		Hurricane tornado lightning animal bite or insect sting

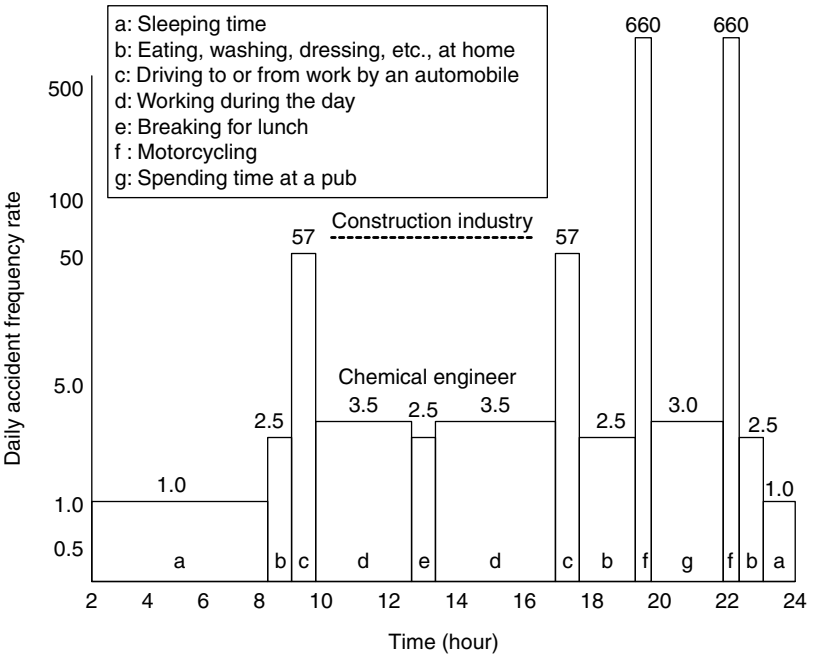

FIGURE 9.4 Daily death risk exposure for a working healthy adult.

as the highest type by women voters and college students, whereas it was placed as the 20th by experts. Experts place motor vehicles as the first risk. Public perception of risk and safety varies by age, gender, education, attitudes, and culture among other factors. Individuals sometimes do not recognize uncertainties associated with risk event or activity causing an unwarranted confidence in an individual's perception of risk or safety. Rare causes of death are often overestimated and common causes of death are often underestimated. Perceived risk is often biased by the familiarity of the hazard. The significance or the impact of safety perceptions stems from that decisions are often made on subjective judgments. If the judgments hold misconceptions about reality, this bias affects the decision. For example, the choice of a transportation mode—train, automobile, motorcycle, bus, bicycle, etc.—results in a decision based on many criteria including such items as cost, speed, convenience, and safety. The weight and evaluation of the decision criteria in selecting a mode of transportation rely on the individual's perception of safety that may deviate sometimes significantly from the actual values of risks. Understanding these differences in risk and safety perceptions is vital to performing risk management decisions and risk communications as provided in subsequent sections on risk management and control.

9.1.3 Risk Assessment

9.1.3.1 Risk Assessment Methodologies

Risk studies require the use of analytical methods at the system level that considers subsystems and components in assessing their failure probabilities and consequences. Systematic, quantitative, qualitative or semiquantitative approaches for assessing the failure probabilities and consequences of engineering systems are used for this purpose. A systematic approach allows an analyst to evaluate expediently and easily complex systems for safety and risk under different operational and extreme conditions. The ability to quantitatively evaluate these systems helps cut the cost of unnecessary and often expensive redesign, repair, strengthening, or replacement of components, subsystems, and systems. The results of risk analysis can also be utilized in decision analysis methods that are based on cost–benefit tradeoffs.

TABLE 9.2 Risk Perception

Activity or Technology	League of Women Voters	College Students	Experts
Nuclear power		1	20
Motor vehicles		5	1
Hand guns		2	4
Smoking		3	2
Motorcycles		6	6
Alcoholic beverages		7	3
General aviation		15	12
Police work		8	17
Pesticides		4	8
Surgery		11	5
Firefighting		10	18
Large construction		14	13
Hunting		18	23
Spray cans		13	25
Mountain climbing		22	28
Bicycles		24	15
Commercial aviation		16	16
Electric (nonnuclear) power		19	9
Swimming		29	10
Contraceptives		9	11
Skiing		25	29
X-rays		17	7
High school or college sports		26	26
Railroads		23	19
Food preservatives		12	14
Food coloring		20	21
Power mowers		28	27
Prescription antibiotics		21	24
Home applications		27	22

Risk assessment is a technical and scientific process by which the risks of a given situation for a system are modeled and quantified. Risk assessment can require and/or provide both qualitative and quantitative data to decision makers for use in risk management.

Risk assessment or risk analysis provides the process for identifying hazards, event-probability assessment, and consequence assessment. The risk assessment process answers three basic questions: (1) What can go wrong? (2) What is the likelihood that it will go wrong? (3) What are the consequences if it does go wrong? Answering these questions requires the utilization of various risk methods as discussed in this chapter.

A risk assessment process should utilize experiences gathered from project personnel including managers, other similar projects and data sources, previous risk assessment models, experiences from other industries and experts, in conjunction with analysis and damage evaluation/prediction tools. A risk assessment process is commonly a part of a risk-based or risk-informed methodology that should be constructed as a synergistic combination of decision models, advanced probabilistic reliability analysis algorithms, failure consequence assessment methods, and conventional performance assessment methodologies that have been employed in related industry for performance evaluation and management. The methodology should realistically account for the various sources and types of uncertainty involved in the decision-making process (Ayyub and McCuen 2003; Ayyub and Klir 2006).

In this section, a typical overall methodology is provided in the form of a workflow or block diagram. The various components of the methodology are described in subsequent sections. Figure 9.5 provides an overall description of a methodology for risk-based management of structural systems for the purpose of demonstration. The methodology consists of the following primary steps:

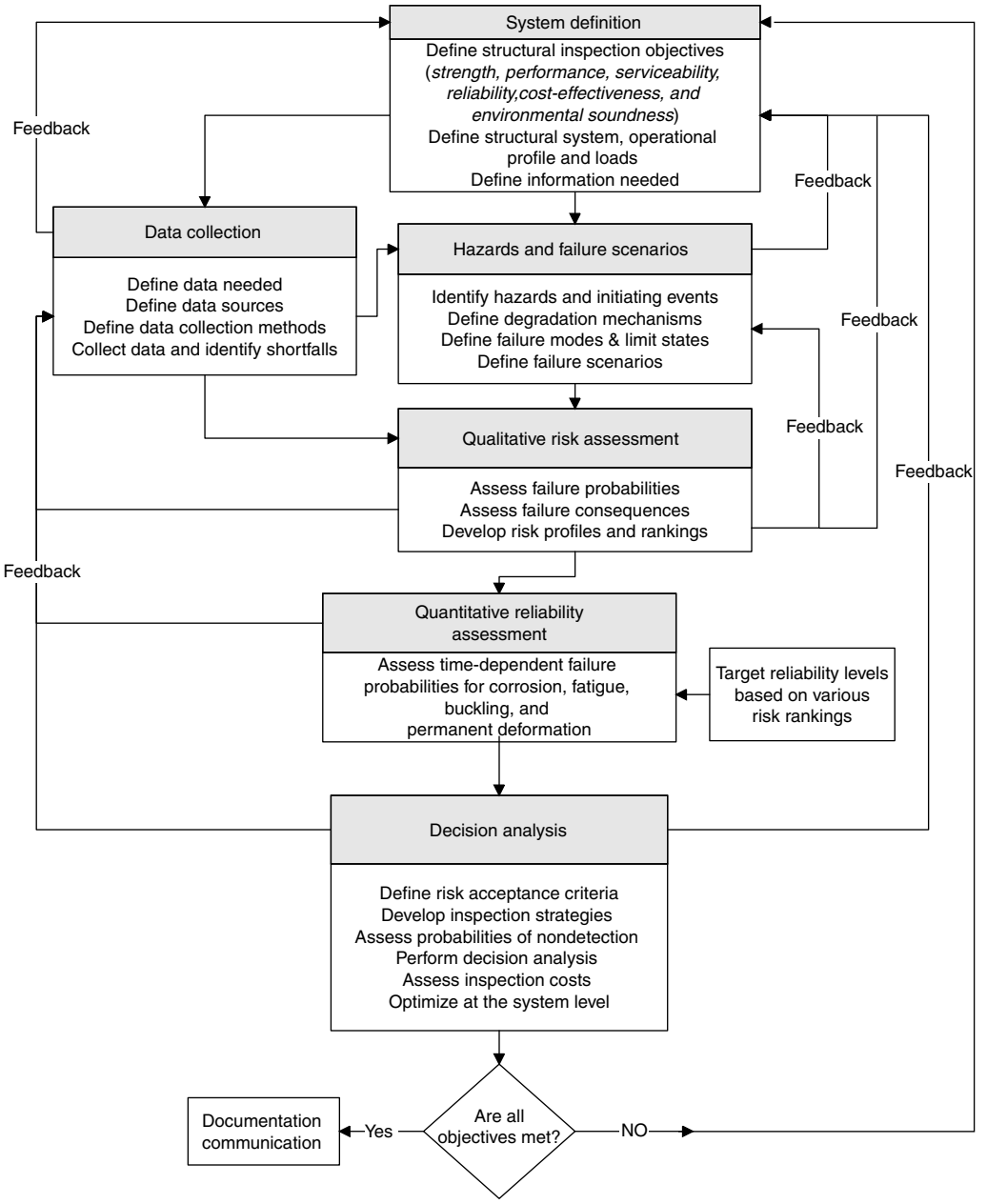

FIGURE 9.5 Methodology for risk-based life cycle management of structural systems.

1. Definition of analysis objectives and systems
2. Hazard analysis, definition of failure scenarios, and hazardous sources and their terms
3. Collection of data in a life cycle framework
4. Qualitative risk assessment
5. Quantitative risk assessment
6. Management of system integrity through failure prevention and consequence mitigation using risk-based decision making.

These steps are briefly described below with additional background materials provided in subsequent sections.

The first step of the methodology is to define the system. This definition should be based on a goal that is broken down into a set of analysis objectives. A system can be defined as an assemblage or combination of elements of various levels and/or details that act together for a specific purpose. Defining the system provides the risk-based methodology with the information it needs to achieve the analysis objectives. The system definition phase of the proposed methodology has four main activities. The activities are to

- Define the goal and objectives of the analysis
- Define the system boundaries
- Define the success criteria in terms of measurable performances
- Collect information for assessing failure likelihood
- Collect information for assessing failure consequences

For example, structural systems require a structural integrity goal that can include objectives stated in terms of strength, performance, serviceability, reliability, cost-effectiveness, and environmental soundness. The objectives can be broken down further to include other structural integrity attributes, such as alignment and water tightness in case of marine vessels. A system can be defined based on a stated set of objectives. The same system can be defined differently depending on these stated objectives. A marine vessel structural system can be considered to contain individual structural elements such as plates, stiffened panels, stiffeners, longitudinals, etc. These elements could be further separated into individual components and/or details. Identifying all of the elements, components and details allows an analysis team to collect the necessary operational, maintenance, and repair information throughout life cycle on each item so that failure rates, repair frequencies, and failure consequences can be estimated. The system definition might need to include nonstructural subsystems and components that would be affected in case of failure. The subsystems and components are needed to assess the consequences.

To understand failure and the consequences of failure, the states of success need to be defined. For the system to be successful, it must be able to perform its designed functions by meeting measurable performance requirements. But the system may be capable of various levels of performance, all of which might not be considered a successful performance. While a marine vessel may be able to get from point A to point B only at a reduced speed due to a fatigue failure that results in excessive vibration at the engine room, its performance would probably not be considered successful. The same concept can be applied to individual elements, components, and details. It is clear from this example that the vessel's success and failure impacts should be based on the overall vessel performance that can easily extend beyond the structural systems.

With the development of the definition of success, one can begin to assess the likelihood of occurrence and causes of failures. Most of the information required to develop an estimate of the likelihood of failure might exist in maintenance and operating histories available on the systems and equipment, and based on judgment and expert opinion. This information might not be readily accessible, and its extraction from its current source might be difficult. Also, assembling it in a manner that is suitable for the risk-based methodology might be a challenge.

Operation, maintenance, engineering, and corporate information on failure history needs to be collected and analyzed for the purpose of assessing the consequences of failures. The consequence information might not be available from the same sources as the information on the failure itself. Typically, there are documentations of repair costs, reinspection or recertification costs, lost person-hours of labor, and possibly even lost opportunity costs due to system failure. Much more difficult to find and assess are costs associated with the effects on other systems, the cost of shifting resources to cover lost production, and things like environmental, safety–loss or public relations costs. These may be attained

through carefully organized discussions and interviews with cognizant personnel including the use of expert-opinion elicitation.

9.1.3.2 Risk Events and Scenarios

To adequately assess all risks associated with a project, the process of identification of risk events and scenarios is an important stage in risk assessment. Risk events and scenarios can be categorized as follows:

- Technical, technological, quality, or performance risks, such as unproven or complex technology, unrealistic performance goals, and changes to the technology used or to the industry standards during the project.
- Project management risks, such as poor allocation of time and resources, inadequate quality of the project plan, and poor use of project management disciplines.
- Organizational risks, such as cost, time, and scope objectives that are internally inconsistent, lack of prioritization of projects, inadequacy or interruption of funding, resource conflicts with other projects in the organization, errors by individuals or by an organization, and inadequate expertise and experience by project personnel.
- External risks, such as shifting legal or regulatory environment, labor issues, changing owner priorities, country risk, and weather.
- Natural hazards, such as earthquakes, floods, strong wind, and waves generally require disaster recovery actions in addition to risk management. Within these categories, several risk types can be identified.

9.1.3.3 Identification of Risk Events and Scenarios

The risk assessment process starts with the question "What can go wrong?" The identification of what can go wrong entails defining hazards, risk events, and risk scenarios. The previous section provided categories of risk events and scenarios. Risk identification involves determining which risks might affect the project and documenting their characteristics. The risk identification generally requires the participation from a project team, risk management team, subject matter experts from other parts of the company, customers, end users, other project managers, stakeholders, and outside experts on as needed basis. Risk identification can be an iterative process. The first iteration may be performed by selected members of the project team, or by the risk management team. The entire project team and primary stakeholders may take a second iteration. To achieve an unbiased analysis, persons who are not involved in the project may perform the final iteration. Risk identification can be a difficult task, because it is often highly subjective, and there are no unerring procedures that may be used to identify risk events and scenarios other than relying heavily on the experience and insight of key project personnel.

The development of the scenarios for risk evaluation can be created deductively (e.g., fault tree) or inductively (e.g., failure mode and effect analysis (FMEA)) as provided in Table 9.3. The table shows methods of multiple uses including likelihood or frequency estimation expressed either deterministically or probabilistically. Also, they can be used to assess varying consequence categories including such items as: economic loss, loss of life, or injuries.

The risk identification process and risk assessment requires the utilization of these formal methods as shown in Table 9.3. These different methods contain similar approaches to answer the basic risk assessment questions; however, some techniques may be more appropriate than others for risk analysis depending on the situation.

9.1.3.4 Risk Breakdown Structure

Risk sources for a project can be organized and structured to provide a standard presentation that would facilitate understanding, communication and management. The previously presented methods can be viewed as simple linear lists of potential sources of risk, providing a set of headings under which risks can

TABLE 9.3 Risk Assessment Methods

Method	Scope
Safety/Review Audit	Identifies equipment conditions or operating procedures that could lead to a casualty or result in property damage or environmental impacts.
Checklist	Ensures that organizations are complying with standard practices.
What-If	Identifies hazards, hazardous situations, or specific accident events that could result in undesirable consequences.
Hazard and Operability Study (HAZOP)	Identifies system deviations and their causes that can lead to undesirable consequences and determine recommended actions to reduce the frequency and/or consequences of the deviations.
Preliminary Hazard Analysis (PrHA)	Identifies and prioritizes hazards leading to undesirable consequences early in the life of a system. It determines recommended actions to reduce the frequency and/or consequences of the prioritized hazards. This is an inductive modeling approach.
Probabilistic Risk Analysis (PRA)	Methodology for quantitative risk assessment developed by the nuclear engineering community for risk assessment. This comprehensive process may use a combination of risk assessment methods.
Failure Modes and Effects Analysis (FMEA)	Identifies the components (equipment) failure modes and the impacts on the surrounding components and the system. This is an inductive modeling approach.
Fault Tree Analysis (FTA)	Identifies combinations of equipment failures and human errors that can result in an accident. This is an deductive modeling approach.
Event Tree Analysis (ETA)	Identifies various sequences of events, both failures and successes that can lead to an accident. This is an inductive modeling approach.
The Delphi Technique	Assists to reach consensus of experts on a subject such as project risk while maintaining anonymity by soliciting ideas about the important project risks that are collected and circulated to the experts for further comment. Consensus on the main project risks may be reached in a few rounds of this process.
Interviewing	Identifies risk events by interviews of experienced project managers or subject matter experts. The interviewees identify risk events based on experience and project information.
Experience-Based Identification	Identifies risk events based on experience including implicit assumptions.
Brainstorming	Identifies risk events using facilitated sessions with stakeholders, project team members, and infrastructure support staff.

be arranged. These lists are sometimes called risk taxonomy. A simple list of risk sources might not provide the richness needed for some decision situations since it only presents a single level of organization. Some applications might require a full hierarchical approach to define the risk sources, with as many levels as are required to provide the necessary understanding of risk exposure. Defining risk sources in such a hierarchical structure is called a risk breakdown structure (RBS). The RBS is defined as a source-oriented grouping of project risks organized to define the total risk exposure of a project of interest. Each descending level represents an increasingly detailed definition of risk sources for the project. The value of the RBS can be in aiding an analyst to understand the risks faced by the project.

An example RBS is provided in Table 9.4. In this example, four risk levels are defined as shown in the table. The project's risks are viewed as level 0. Three types of level 1 risks are provided in the table for the purpose of demonstration. The number of risk sources in each level varies and depends on the application at hand. The subsequent level 2 risks are provided in groups that are detailed further in level 3. The RBS provides a means to systematically and completely identify all relevant risk sources for a project.

The risk breakdown structure should not be treated as a list of independent risk sources since commonly they have interrelations and common risk drivers. Identifying causes behind the risk sources is a key step towards an effective risk management plan including mitigation actions. A process of risk interrelation assessment and root-cause identification can be utilized to potentially lead to identifying credible scenarios that could lead to snowball effects for risk management purposes.

TABLE 9.4 Risk Breakdown Structure for a Project

Level 0	Level 1	Level 2	Level 3
			History, experiences, culture, personnel
		Corporate	Organization structure, stability, communication
			Finances conditions
			Other projects
			M
	Management		History, experiences, culture, personnel
			Contracts and agreements
		Customers &	Requirement definition
		Stakeholders	Finances and credit
			M
			Physical environment
		Natural	Facilities, site, equipment, materials
		environment	Local services
			M
			Political
			Legal, regulatory
Project Risks	External	Cultural	Interest groups
			Society and communities
			M
		Economic	Labor market, conditions, competition
			Financial markets
			M
	Technology	Requirements	Scope and objectives
			Conditions of use, users
			Complexity
			M
		Performance	Technology maturity
			Technology limitations
			New technologies
			New hazards or threats
			M
		Application	Organizational experience
			Personnel skill sets & experience
			Physical resources
			M

9.1.3.5 System Definition for Risk Assessment

Defining the system is an important first step in performing a risk assessment. A system can be defined as a deterministic entity comprising an interacting collection of discrete elements and commonly defined using deterministic models.

The word "deterministic" implies that the system is identifiable and not uncertain in its architecture. The definition of the system is based on analyzing its functional and/or performance requirements. A description of a system may be a combination of functional and physical elements. Usually, functional descriptions are used to identify high information levels on a system. A system may be divided into subsystems that interact. Additional detail leads to a description of the physical elements, components, and various aspects of the system.

The examination of a system needs to be made in a well-organized and repeatable fashion so that risk analysis can be consistently performed, therefore insuring that important elements of a system are defined and extraneous information is omitted. The formation of system boundaries is based upon the objectives of the risk analysis.

The establishment of system boundaries can assist in developing the system definition. The decision on what the system boundary is partially based on what aspects of the system's performance are of concern.

The selection of items to include within the external boundary region is also reliant on the goal of the analysis. Beyond the established system boundary is the external environment of the system.

Boundaries beyond the physical/functional system can also be established. For example, time may also be a boundary since an overall system model may change, as a product is further along in its life cycle. The life cycle of a system is important because some potential hazards can change throughout the life cycle. For example, material failure due to corrosion or fatigue may not be a problem early in the life of a system; however, this may be an important concern later in the life cycle of the system.

Along with identifying the boundaries, it is also important to establish a resolution limit for the system. The selected resolution is important because it limits the detail of the analysis. Providing too little detail might not provide enough information for the problem. Too much information may make the analysis more difficult and costly due to the added complexity. The depth of the system model needs to be sufficient for the specific problem. Resolution is also limited by the feasibility of determining the required information for the specific problem. For failure analysis, the resolution should be to the components level where failure data are available. Further resolution is not necessary and would only complicate the analysis.

The system breakdown structure is the top-down division of a system into subsystems and components. This architecture provides internal boundaries for the system. Often the systems/subsystems are identified as functional requirements that eventually lead to the component level of detail. The functional level of a system identifies the function(s) that must be performed for the operation of the system. Further decomposition of the system into "discrete elements" leads to the physical level of a system definition identifying the hardware within the system. By organizing a system hierarchy using a top-down approach rather than fragmentation of specific systems, a rational, repeatable, and systematic approach to risk analysis can be achieved.

Further system analysis detail is addressed from modeling the system using some of the risk assessment methods described in Table 9.3. These techniques develop processes that can assist in decision making about the system. The logic of modeling based on the interaction of a system's components can be divided into induction and deduction. This difference in the technique of modeling and decision making is significant. Induction logic provides the reasoning of a general conclusion from individual cases. This logic is used when analyzing the effect of a fault or condition on a systems operation. Inductive analysis answers the question, "What are the system states due to some event?" In reliability and risk studies, this "event" is some fault in the system. Several approaches using the inductive approach include: PrHA, FMEA, and ETA. Deductive approaches provide reasoning for a specific conclusion from general conditions. For system analysis this technique attempts to identify what modes of a system/subsystem/component failure can be used to contribute to the failure of the system. This technique answers the question, "How a system state can occur?" Inductive reasoning provides the techniques for FTA or its complement success tree analysis (STA).

9.1.3.6 Selected Risk Assessment Methods

Qualitative versus Quantitative Risk Assessment. The risk assessment methods can be categorized according to how the risk is determined: by quantitative or qualitative analysis. Qualitative risk analysis uses judgment and sometimes "expert" opinion to evaluate the probability and consequence values. This subjective approach may be sufficient to assess the risk of a system, depending on the available resources.

Quantitative analysis relies on probabilistic and statistical methods, and databases that identify numerical probability values and consequence values for risk assessment. This objective approach examines the system in greater detail to assess risks.

The selection of a quantitative or qualitative method depends upon the availability of data for evaluating the hazard and the level of analysis needed to make a confident decision. Qualitative methods offer analyses without detailed information, but the intuitive and subjective processes may result in differences in outcomes by those who use them. Quantitative analysis generally provides a more uniform understanding among different individuals, but requires quality data for accurate results. A combination of both qualitative and quantitative analyses can be used depending on the situation.

Risk assessment requires estimates of the failure likelihood at some identified levels of decision making. The failure likelihood can be estimated in the form of lifetime failure likelihood, annual failure likelihood, mean time between failures, or failure rate. The estimates can be in numeric or nonnumeric form. An example numeric form for an annual failure probability is 0.00015, and for a mean time between failures is 10 years. An example nonnumeric form for "an annual failure likelihood" is large, and for a "mean time between failures" is medium. In the latter nonnumeric form, guidance needs to be provided regarding the meaning of terms such as large, medium, small, very large, very small, etc. The selection of the form should be based on the availability of information, the ability of the personnel providing the needed information to express it in one form or another, and the importance of having numeric vs. nonnumeric information in formulating the final decisions.

The types of failure consequences that should be considered in a study need to be selected. They can include production loss, property damage, environmental damage, and safety–loss in the form of human injury and death. Approximate estimates of failure consequences at the identified levels of decision making need to be determined. The estimates can be in numeric or nonnumeric form. An example numeric form for production loss is 1000 units. An example nonnumeric form for production loss is large. In the latter nonnumeric form, guidance needs to be provided regarding the meaning of terms such as large, medium, small, very large, very small, etc. The selection of the form should be based on the availability of information, the ability of the personnel providing the needed information to express it in one form or another, and the importance of having numeric vs. nonnumeric information in formulating the final decisions.

Risk estimates can be determined as a pair of the likelihood and consequences, and computed as the arithmetic multiplication of the respective failure likelihood and consequences for the equipment, components and details. Alternatively, for all cases, plots of failure likelihood versus consequences can be developed. Then, approximate ranking of them as groups according to risk estimates, failure likelihood, and/or failure consequences can be developed.

Preliminary Hazard Analysis. Preliminary hazard analysis (PrHA) is a common risk-based technology tool with many applications. The general process is shown in Figure 9.6. This technique requires experts to identify and rank the possible accident scenarios that may occur. It is frequently used as a preliminary method to identify and reduce the risks associated with major hazards of a system.

Failure Mode and Effects Analysis. Failure mode and effects analysis (FMEA) is another popular risk-based technology tool as shown in Figure 9.7. This technique has been introduced both in the national

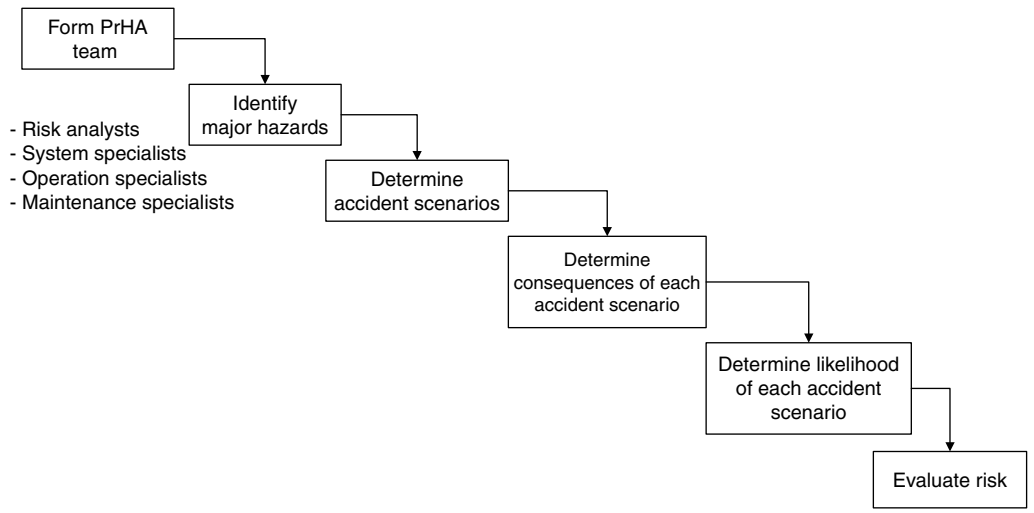

FIGURE 9.6 Preliminary hazard analysis (PrHA) process.

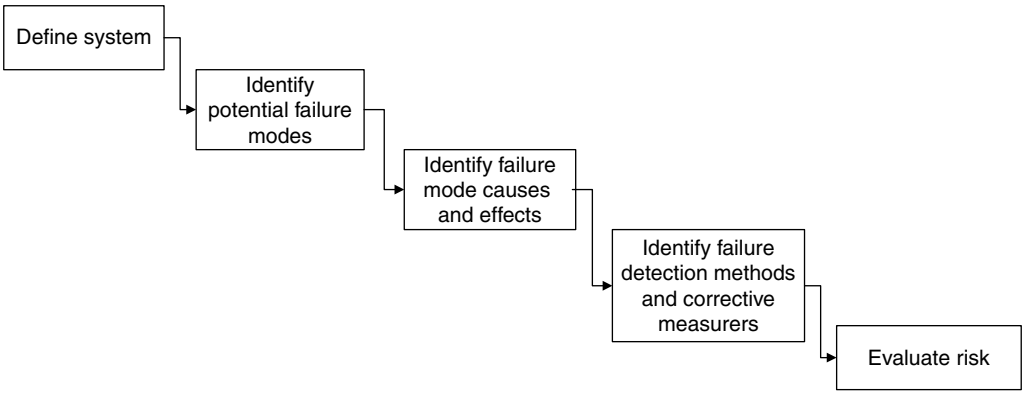

FIGURE 9.7 Failure mode and effects analysis (FMEA) process.

and international regulations for the aerospace (US MIL-STD-1629A), processing plant, and marine industries. The Society of Automotive Engineers, in its recommended practice, introduces two types of FMEA: design and process FMEA. This analysis tool assumes a failure mode occurs in a system/component through some failure mechanism; the effect of this failure on other systems is then evaluated. A risk ranking can be developed for each failure mode for the effect on the overall performance of the system.

The various terms used in FMEA with examples based on the manufacturing of personal flotation devices (PFDs) are provided under subsequent headings to include failure mode, failure effect, severity rating, causes, occurrence rating, controls, detection rating, and risk priority number.

Risk Matrices. Risk can be assessed and presented using matrices for preliminary screening by subjectively estimating probabilities and consequences in a qualitative manner. A risk matrix is a two-dimensional presentation of likelihood and consequences using qualitative metrics for both dimensions. According to this method, risk is characterized by categorizing probability and consequence on the two axes of a matrix. Risk matrices have been used extensively for screening of various risks. They may be used alone or as a first step in a quantitative analysis. Regardless of the approach used, risk analysis should be a dynamic process, i.e., a living process where risk assessments are reexamined and adjusted. Actions or inactions in one area can affect risk in another; therefore continuous updating is necessary.

The likelihood metric can be constructed using the categories shown in Table 9.5, whereas the consequences metric can be constructed using the categories shown in Table 9.6 with an example provided in Table 9.7. The consequence categories of Table 9.6 focus on the health and environmental aspects of consequences. The consequence categories of Table 9.7 focus on the economic impact, and should be adjusted to meet specific needs of industry and/or applications. An example risk matrix is shown in Figure 9.8. In the figure, each boxed area is shaded depending on a subjectively assessed risk level. Three risk levels are used herein for illustration purposes: low (L), medium (M), and high (H). Other risk levels may be added using a scale of five levels instead of three levels if needed. These risk levels

TABLE 9.5 Likelihood Categories for a Risk Matrix

Category	Description	Annual Probability Range
A	Likely	>0.1 (1 in 10)
B	Unlikely	>0.01 (1 in 100) but <0.1
C	Very Unlikely	>0.001 (1 in 1,000) but <0.01
D	Doubtful	>0.0001 (1 in 10,000) but <0.001
E	Highly Unlikely	>0.00001 (1 in 100,000) but <0.0001
F	Extremely Unlikely	<0.00001 (1 in 100,000)

TABLE 9.6 Consequence Categories for a Risk Matrix

Category	Description	Examples
I	Catastrophic	Large number of fatalities and/or major long-term environmental impact.
II	Major	Fatalities and/or major short-term environmental impact.
III	Serious	Serious injuries and/or significant environmental impact.
IV	Significant	Minor injuries and/or short-term environmental impact.
V	Minor	First aid injuries only and/or minimal environmental impact.
VI	None	No significant consequence.

are also called *severity factors*. The high (H) level can be considered as unacceptable risk level, the medium (M) level can be treated as either undesirable or as acceptable with review, and the low (L) level can be treated as acceptable without review.

Event Modeling: Event, Success Trees, and Fault Trees. Event modeling is a systematic—and often most complete—way to identify accident scenarios and quantify risk for risk assessment. This risk-based technology tool provides a framework for identifying scenarios to evaluate the performance of a system or component through system modeling. The combination of event tree analysis (ETA), success tree analysis (STA), and fault tree analysis (FTA) can provide a structured analysis to system safety.

Event tree analysis is often used if the successful operation of a component/system depends on a discrete (chronological) set of events. The initiating event is first followed by other events leading to an overall result (consequence). The ability to address a complete set of scenarios is developed because all combinations of both the success and failure of the main events are included in the analysis. The probability of occurrence of the main events of the event tree can be determined using a fault tree or its complement the success tree. The scope of the analysis for event trees and fault trees depends on the objective of the analysis.

Event tree analysis is appropriate if the operation of some system/component depends on a successive group of events. Event trees identify the various combinations of event successes and failures as a result of an initiating event to determine all possible scenarios. The event tree starts with an initiating event followed by some reactionary event. This reaction can either be a success or failure. If the event succeeds, the most commonly used indication is the upward movement of the path branch. A downward branch of the event tree marks the failure of an event. The remaining events are evaluated to determine the different possible scenarios. The scope of the events can be functions/systems that can provide some reduction to the possible hazards from the initiating event. The final outcome of a sequence of events identifies the overall state resulting from the scenario of events. Each path represents a failure scenario with varying levels of probability and risk. Different event trees can be created for different event initiators. Figure 9.9 shows an example event tree for the basic elements of a sprinkler system that might be critical for maintaining the integrity of a marine vessel.

Based on the occurrence of an initiating event, event tree analysis examines possible system outcomes or consequences. This analysis tool is particularly effective in showing interdependence of system components which is important in identifying events, that at first might appear insignificant, but due to

TABLE 9.7 Example Consequence Categories for a Risk Matrix in 2003 Monetary Amounts (US$)

Category	Description	Cost
I	Catastrophic Loss	> $10,000,000,000
II	Major Loss	>$1,000,000,000 but <$10,000,000,000
III	Serious Loss	>$100,000,000 but <$1,000,000,000
IV	Significant Loss	>$10,000,000 but <$100,000,000
V	Minor Loss	>$1,000,000 but <$10,000,000
VI	Insignificant Loss	<$1,000,000

Severity factors. The high (H) level can be considered as unacceptable risk level, the medium (M) level can be treated as either undesirable or as acceptable with review, and the low (L) level can be treated as acceptable without review.

Table 9-5. Likelihood categories for a risk matrix

Category	Description	Annual probability range
A	Likely	≥ 0.1 (1 in 10)
B	Unlikely	≥ 0.01 (1 in 100) but < 0.1
C	Very unlikely	≥ 0.001 (1 in 1,000) but < 0.01
D	Doubtful	≥ 0.0001 (1 in 10,000) but < 0.001
E	Highly unlikely	≥ 0.00001 (1 in 100,000) but < 0.0001
F	Extremely unlikely	< 0.00001 (1 in 100,000)

Table 9-6. consequence categories for a risk matrix

Category	Description	Examples
I	Catastrophic	Large number of fatalities and/or major long-term environmental impact.
II	Major	Fatalities and/or major short-term environmental impact.
III	Serious	Serious injuries and/or significant environmental impact.
IV	Significant	Minor injuries and/or short-term environmental impact.
V	Minor	First aid injuries only and/or minimal environmental impact.
VI	None	No significant consequence.

Table 9-7. Example consequence categories for a risk matrix in 2003 monetary amounts (US$)

Category	Description	Cost
I	Catastrophic loss	≥ $10,000,000,000
II	Major loss	≥ $1,000,000,000 but < $10,000,000,000
III	Serious loss	≥ $100,000,000 but < $1,000,000,000
IV	Significant loss	≥ $10,000,000 but < $100,000,000
V	Minor loss	≥ $1,000,000 but < $10,000,000
VI	Insignificant loss	< $1,000,000

Probability category	A	L	M	M	H	H	H
	B	L	L	M	M	H	H
	C	L	L	L	M	M	H
	D	L	L	L	L	M	M
	E	L	L	L	L	L	M
	F	L	L	L	L	L	L
		VI	V	IV	III	II	I
	Consequence category						

FIGURE 9.8 Example Risk Matrix

the interdependency result in devastating results. Event tree analysis is similar to fault tree analysis because both methods use probabilistic reliability data of the individual components and events along each path to compute the likelihood of each outcome.

A quantitative evaluation of event tree probability values can be used for each event to evaluate the probability of the overall system state. Probability values for the success or failure of the events can be used to identify the probability for a specific event tree sequence. The probabilities of the events in a sequence can be provided as an input to the model or evaluated using fault trees. These probabilities for various events in a sequence can be viewed as conditional probabilities and therefore can be multiplied to obtain the occurrence probability of the sequence. The probabilities of various sequences can be summed up to determine the overall probability of a certain outcome. The addition of consequence evaluation of a scenario allows for generation of a risk value. For example, the occurrence probability of the top branch, i.e., scenario, in Figure 9.9 is computed as the product of the probabilities of the composing events to this scenario, i.e., F ∩PO∩SF∩SS∩FE or (F)(PO)(SF)(SS)(FE) for short.

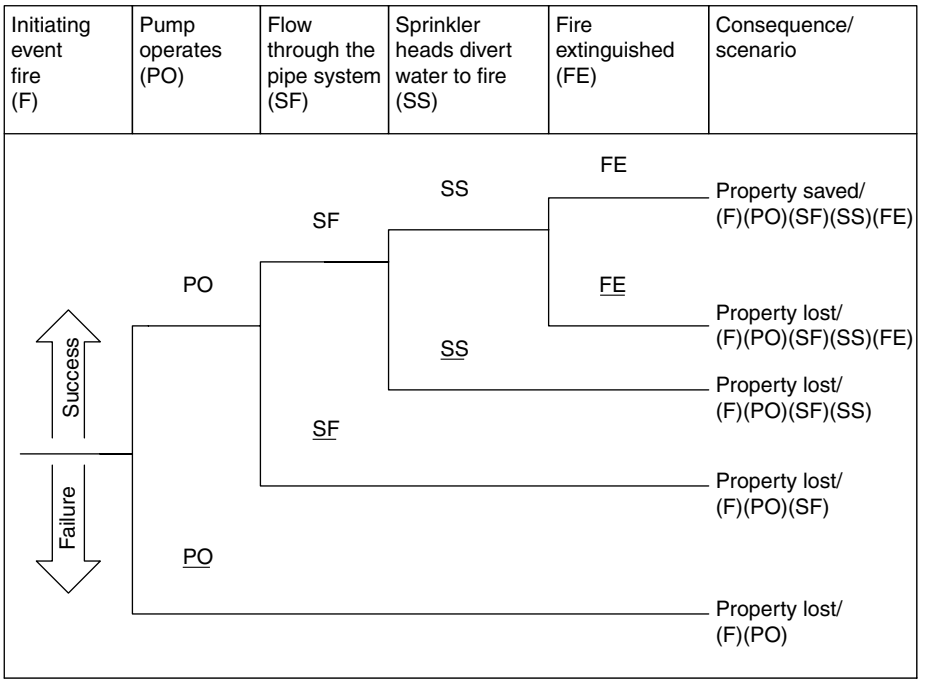

FIGURE 9.9 Event tree example for sprinkler system.

Complex systems are often difficult to visualize and the effect of individual components on the system as a whole is difficult to evaluate without an analytical tool. Two methods of modeling that have greatly improved the ease of assessing system reliability/risk are fault trees (FT) and success trees (ST). A fault tree is a graphical model created by deductive reasoning leading to various combinations of events that lead to the occurrence of some top event failure. A success tree shows the combinations of successful events leading to the success of the top event. A success tree can be produced as the complement (opposite) of the fault tree as illustrated in this section. Fault trees and success trees are used to further analyze the event tree headings (the main events in an event tree) to provide further detail to understand system complexities. In constructing the FT/ST only those failure/success events that are considered significant are modeled. This determination is assisted by defining system boundaries. For example, the event "pump operates (PO)" in Figure 9.9 can be analyzed by developing a top-down logical breakdown of failure or success using fault tress or event trees, respectively.

Fault tree analysis (FTA) starts by defining a top event that is commonly selected as an adverse event. An engineering system can have more than one top event. For example, a ship might have the following top events for the purpose of reliability assessment: power failure, stability failure, mobility failure, or structural failure. Then, each top event needs to be examined using the following logic: in order for the top event to occur, other events must occur. As a result, a set of lower-level events is defined. Also, the form in which these lower-level events are logically connected (i.e., in parallel or in series) needs to be defined. The connectivity of these events is expressed using "AND" or "OR" gates. Lower-level events are classified into the following types:

1. *Basic events*: These events cannot be decomposed further into lower-level events. They are the lowest events that can be obtained. For these events, failure probabilities need be obtained.
2. *Events that can be decomposed further*: These events can be decomposed further to lower levels. Therefore, they should be decomposed until the basic events are obtained.
3. *Undeveloped events*. These events are not basic and can be decomposed further. However, because they are not important, they are not developed further. Usually, the probabilities of these events are

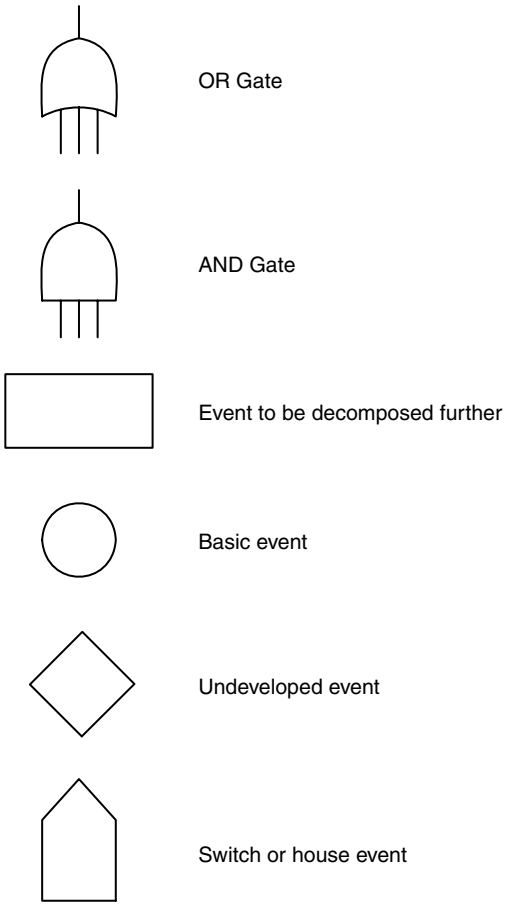

FIGURE 9.10 Symbols used in fault tree analysis.

very small or the effect of their occurrence on the system is negligible, or can be controlled or mediated.

4. *Switch (or house) events.* These events are not random, and can be turned on or off with full control. The symbols shown in Figure 9.10 are used for these events. Also, a continuation symbol is shown that is used to break up a fault tree into several parts for the purpose of fitting it in several pages.

FTA requires the development of a tree-looking diagram for the system that shows failure paths and scenarios that can result in the occurrence of a top event. The construction of the tree should be based on the building blocks and the Boolean logic gates.

The outcome of interest from the fault tree analysis is the occurrence probability of the top event. Because the top event was decomposed into basic events, its occurrence can be stated in the form of "AND" and "OR" of the basic events. The resulting statement can be restated by replacing the "AND" with the intersection of the corresponding basic events, and the "OR" with the union of the corresponding basic events. Then, the occurrence probability of the top event can be computed by evaluating the probabilities of the unions and intersections of the basic events. The dependence between these events also affects the resulting probability of the system.

For large fault trees, the computation of the occurrence probability of the top event can be difficult because of their size. In this case, a more efficient approach is needed for assessing the reliability of a system; such as the minimal cut set approach. According to this approach, each cut set is defined as a set of basic events where the joint occurrence of these basic events results in the occurrence of the top event. A minimal cut set is a cut set with the condition that the nonoccurrence of any one basic event from this set results in the nonoccurrence of the top event. Therefore, a minimal cut set can be viewed as a subsystem in parallel. In general, systems have more than one minimal cut sets. The occurrence of the top event of the system can, therefore, be due to any one of these minimal cut sets. As a result, the system can be viewed as the union of all the minimal cut sets for the system. If probability values are assigned to the cut sets, a probability for the top event can be determined.

A simple example of this type of modeling is shown in Figure 9.11 for a pipe system using a reliability block diagram. If the goal of the system is to maintain water flow from one end of the system to the other, then the individual pipes can be related with a Boolean logic. Both pipe (a) and pipe (d) and pipe (b) or pipe (c) must function for the system to meet its goal as shown in the success tree Figure 9.12a. The compliment of the success tree is the fault tree. The goal of the fault tree model is to construct the logic for system failure as shown in Figure 9.12b. After these tree elements have been defined, possible failure scenarios of a system can be defined.

As previously described, a failure path is often referred to as a *cut set*. One objective of the analysis is to determine the entire minimal cut sets, where a minimal cut set is defined as a failure combination of all

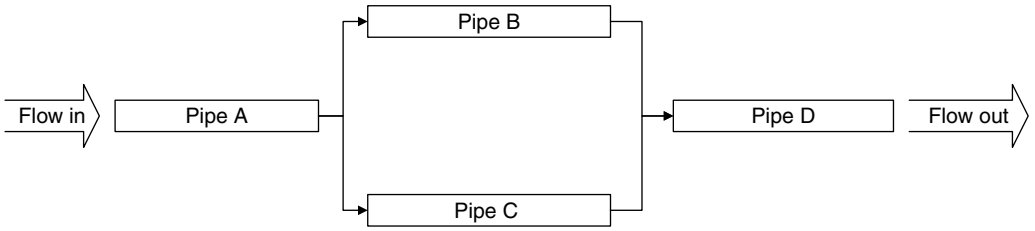

FIGURE 9.11 A reliability block diagram for a piping system.

essential events that can result in the failure top event. A minimal cut set includes in its combination all essential events, i.e., the nonoccurrence of any of these essential events in the combination of a minimal cut set results in the nonoccurrence of the minimal cut set. These failure combinations are used to compute the failure probability of the top event. The concept of the minimal cut sets applies only to the fault trees. A similar concept can be developed in the complementary space of the success trees, and is called the minimal pass set. In this case, a minimal pass set is defined as a survival (or success) combination of all essential success events that can result in success as defined by the top event of the success tree. For the piping example, the minimal cut sets are

$$\text{A} \tag{9.6a}$$

$$\text{D} \tag{9.6b}$$

$$\text{B and C} \tag{9.6c}$$

A minimal cut set includes events that are all necessary for the occurrence of the top event. For example, the following cut set is not a minimal cut set:

$$\text{A and B} \tag{9.7}$$

Example 9.1 Trends in Fault Tree Models and Cut Sets

This example demonstrates how the cut sets can be identified and constructed for different arrangements of OR and AND gates logically defining a top event occurrence. Generally, the number of cut sets increases by increasing the number of OR gates in the tree. For example, Figure 9.13 shows this trend by comparing cases a, b, and d. On the other hand, increasing the number of AND gates results in increasing the number of events included in the cut sets as shown in case c of Figure 9.13.

Common cause scenarios are events or conditions that result in the failure of seemingly separate systems or components. Common cause failures complicate the process of conducting risk analysis because a seemingly redundant system can be rendered ineffective by a common cause failure. For example, an emergency diesel generator fed by the same fuel supply as the main diesel engine will fail with the main diesel generator, if the fuel supply is the root source of the failure. The redundant emergency diesel generator is not truly redundant due to a common cause failure. Another example of common cause events is the failure of two separate but similar pieces of machinery due to a common maintenance problem, two identical pieces of equipment failing due to a common manufacturing defect, or two pieces of equipment failing due to a common environmental condition such as the flooding of a compartment or a fire in the vicinity of both pieces of machinery. A method for calculating the reliability of a system while taking into account common cause effects is the beta-factor model. Other methods include multiple Greek letter model, alpha-factor model, and beta-binomial failure-rate model.

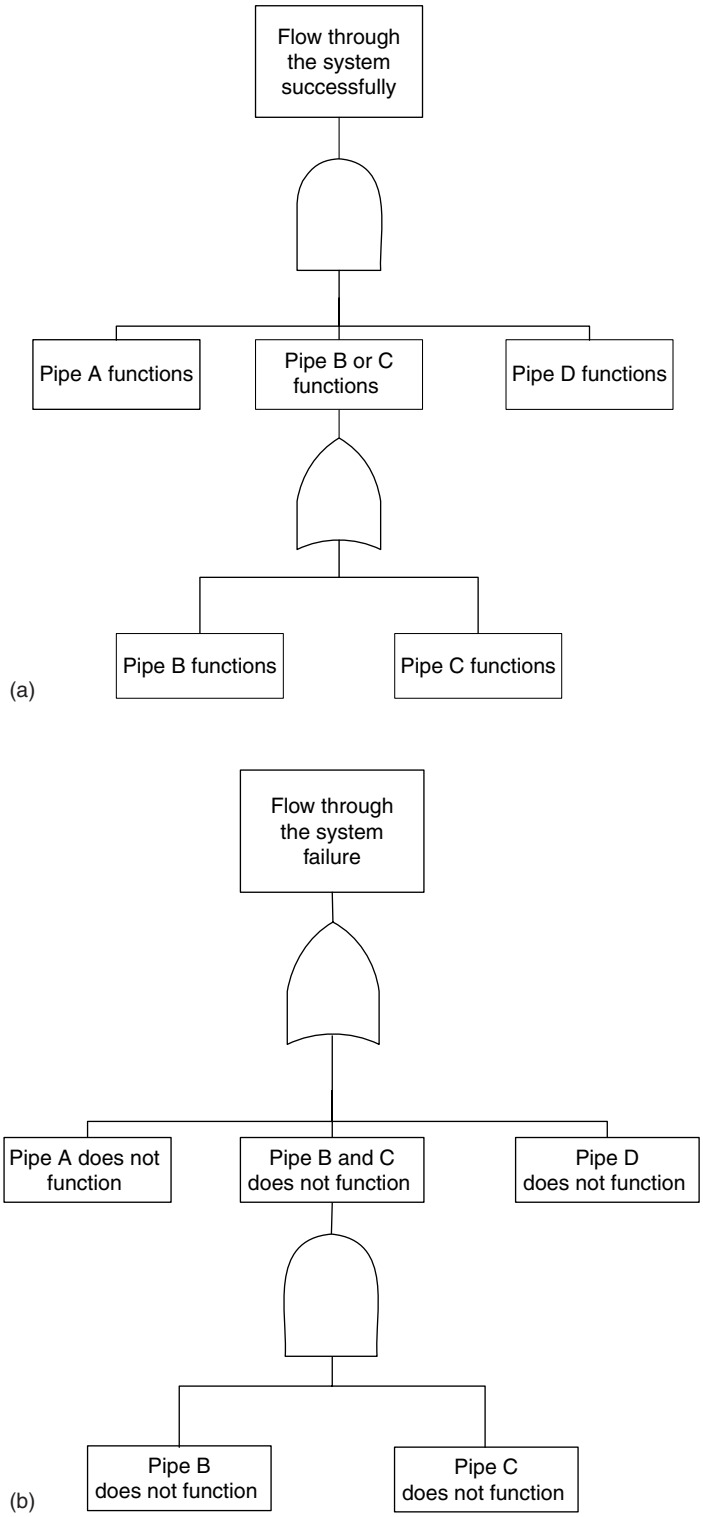

FIGURE 9.12 (a) Success tree for the pipe system example. (b) Fault tree for the pipe system example.

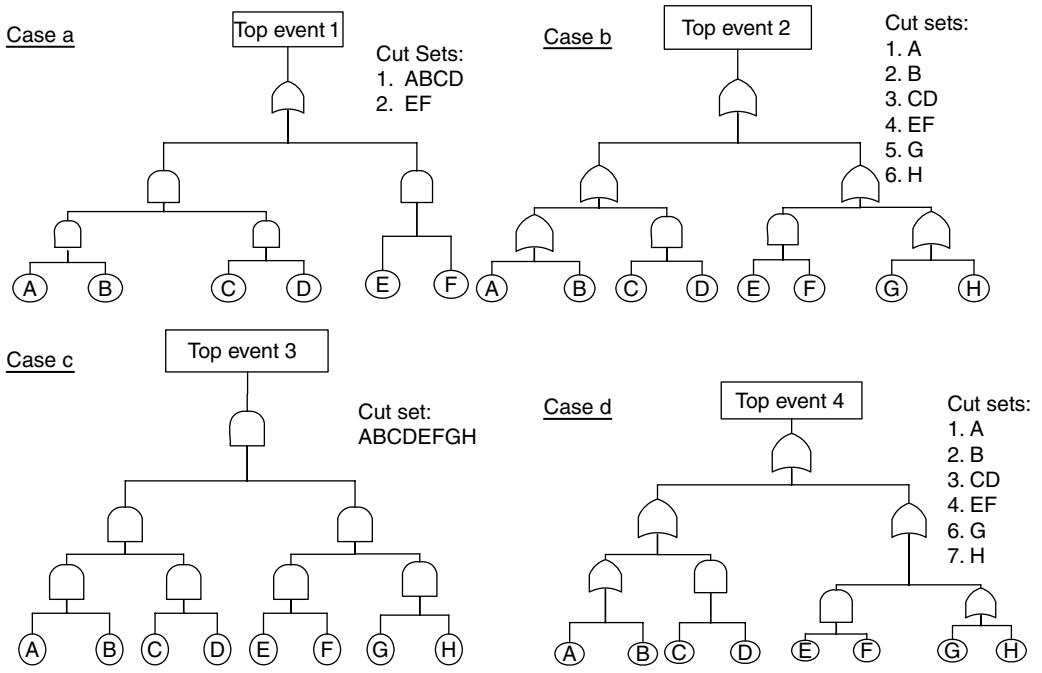

FIGURE 9.13 Trends in fault tree models and cut sets. (From Maryland Emergency Management Agency (MEMA), 2006, State of Maryland Guide for the Protection of Critical Infrastructure and Key Resources for Homeland Security, Volume 1: Critical Asset & Portfolio Risk Assessment (CAPRA) Methodology, Office of Homeland Security, Annapolis, MD.)

Part of risk-based decision analysis is pinpointing the system components that result in high-risk scenarios. Commercial system reliability software provides this type of analysis in the form of system reliability sensitivity factors to changes in the underlying component reliability values. In performing risk analysis, it is desirable to assess the importance of events in the model, or the sensitivity of final results to changes in the input failure probabilities for the events. Several sensitivity or importance factors are available and can be used. The most commonly used factors include (1) Fussell–Vesely factor and (2) Birnbaum factor. Also, a weighted combination of these factors can be used as an overall measure.

9.1.3.7 Human-Related Risks

Risk assessment requires the performance analysis of an entire system composed of a diverse group of components. The system definition readily includes the physical components of the system; however, humans are also part of most systems and provide significant contributions to risk. It has been estimated that nearly 90% of the accidents at sea are contributed to human error. The human contribution to risk can be estimated from an understanding of behavioral sciences. Both the "hardware failure" and human error should be addressed in the risk assessment since they both contribute to risks associated with the system. After the human error probabilities are determined, human error/failures are treated in the same fashion as hardware failures in performing risk assessment quantification.

The determination of the human error contribution to risk is determined by human reliability analysis (HRA) tools. HRA is the discipline that enables the analysis and impact of humans on the reliability and safety of systems. Important results of HRA are determining the likelihood of human error as well as ways in which human errors can be reduced. When combined with system risk analysis, HRA methods provide an assessment of the detrimental effects of humans on the performance of the system. Human reliability analysis is generally considered to be composed of three basic steps: error identification, modeling, and quantification.

Other sources of human-related risks are in the form of deliberate sabotage of a system from within a system or as threat from outside the system, such as a computer hacker or a terrorist. The hazard in this case is not simply random but intelligent. The methods introduced in earlier sections might not be fully applicable for this risk type. The threat scenarios to the system in this case have a dynamic nature that are affected by the defense or risk mitigation and management scenarios that would be implemented by an analyst. The use of game theory methods might be needed in this case in combination with other risk analysis and management methods. Game theory is introduced in the last subsection herein.

Human Error Identification. Human errors are unwanted circumstances caused by humans that result in deviations from expected norms that place systems at risk. It is important to identify the relevant errors to make a complete and accurate risk assessment. Human error identification techniques should provide a comprehensive structure for determining significant human errors within a system. Quality HRA allows for accuracy in both the HRA assessment and overall system risk assessment.

Identification of human errors requires knowledge about the interactions of humans with other humans or machines (the physical world). It is the study of these interfaces that allows for the understanding of human errors. Potential sources of information for identifying human error may be determined from task analysis, expert judgment, laboratory studies, simulation, and reports. Human errors may be considered active or latent depending on the time delay between when the error occurs and when the system fails.

January 09, 2006 9–32. It is important to note the distinction between human errors and human factors. Human errors are generally considered separately from human factors that applies information about human behavior, abilities, limitations, and other characteristics to the design of tools, machines, systems tasks, jobs, and environments for productive, safe, comfortable, and effective human use. Human factors are determined from performing descriptive studies for characterizing populations and experimental research. However, human factors analysis may contribute to the human reliability analysis.

Human Error Modeling. After human errors have been identified, they must be represented in a logical and quantifiable framework along with other components that contribute to the risk of the system. This framework can be determined from development of a risk model. Currently, there is no consensus on how to model human reliably. Many of these models utilize human event trees and fault trees to predict human reliability values. The identifications of human failure events can also be identified using failure mode and effects analysis. The human error rate estimates are often based on simulation tests, models, and expert estimation.

Human Error Quantification. Quantification of human error reliability promotes the inclusion of the human element in risk analysis. This is still a developing science requiring understanding of human performance, cognitive processing, and human perceptions. Because an exact model for human cognition has not been developed, much of the current human reliability data relies on accident databases, simulation, and other empirical approaches. Many of the existing data sources were developed for from specific industry data such as nuclear and aviation industries. The application of these data sources for a specific problem should be thoroughly examined prior to application for a specific model. The result of the quantification of human reliability in terms of probability of occurrence is typically called a *human error probability (HEP)*. There are many techniques that have been developed to help predict the HEP values. The technique for human error rate prediction (THERP) is one of the most widely used methods for HEP. This technique is based on data gathered from the nuclear and chemical processing industries. THERP relies on HRA event tree modeling to identify the events of concern. Quantification is performed from data tables of basic HEP for specific tasks that may be modified based on the circumstances affecting performance.

The degree of human reliability is influenced by many factors often called *performance shaping factors (PSFs)*. PSFs are those factors that affect the ability of people to carry out required tasks. For example, the knowledge people have on how to don/activate a personal flotation device (PFD) will affect the performance of this task. Training (another PSF) in donning PFDs can also assist in the ability to perform this task. Another example is the training that is given to passengers on airplanes before takeoff on using seatbelts, emergency breathing devices, and flotation devices. Often, the quantitative estimates

of reliability are generated from a base error rate that is then altered based on the PSFs of the particular circumstances. Internal performance shaping factors are an individual's own attributes (experience, training, skills, abilities, attitudes) that affect the ability of the person to perform certain tasks. External PSFs are the dynamic aspects of situation, tasks, and system that affect the ability to perform certain tasks. Typical external factors include environmental stress factors (such as heat, cold, noise, situational stress, time of day), management, procedures, time limitations, and quality of person–machine interface. With these PSFs, it is easy to see the dynamic nature of HEP evaluation based on the circumstances of the analysis.

Reducing Human Errors. Error reduction is concerned with lowering the likelihood for error in an attempt to reduce risk. The reduction of human errors may be achieved by human factors interventions or by engineering means. Human factors interventions include improving training or improving the human–machine interface (such as alarms, codes, etc.) based on an understanding of the causes of error. Engineering means of error reduction may include automated safety systems or interlocks. The selection of the corrective actions to take can be done through decision analysis considering cost–benefit criteria.

Game Theory for Intelligent Threats. Game theory can be used to model human behavior as a threat to a system. Generally, game theory utilizes mathematics, economics, and the other social and behavioral sciences to model human behavior.

An example of intelligent threats is terrorism and sabotage as an ongoing battle between coordinated opponents representing a two-party game, where each opponent seeks to achieve their own objectives within a system. In the case of terrorism, it is a game of a well-established political system as a government vs. an emerging organization that uses terrorism to achieve partial or complete dominance. Each player in this game seeks a utility, i.e., benefit, that is a function of the desired state of the system. In this case, maintaining system survival is the desired state for the government, whereas the opponent seeks a utility based on the failure state of the system. The government, as an opponent, is engaged in risk mitigation whose actions seek to reduce the threat, reduce the system vulnerability, and/or mitigate the consequences of any successful attacks. The terrorists, as an opponent, can be viewed as the aggressor who strives to alter or damage their opponent's desired system state. This game involves an intelligent threat and is dynamic. The game is ongoing until the probability of a successful disruptive attempt of the aggressor reaches an acceptable level of risk—a stage where risk is considered under control—and the game is brought to an end. Classical game theory can be used in conjunction with probabilistic risk analysis to determine optimal mitigation actions that maximize benefits.

A classical example used to introduce game theory is called the *prisoners' dilemma* and is based on two suspects that are captured near the scene of a crime and are questioned separately by authority such as the police. Each has to choose whether or not to confess and implicate the other. If neither person confesses, then both will serve, for example, one year on a charge of carrying a concealed weapon. If each confesses and implicates the other, both will go to prison for, say, 10 years. However, if one person confesses and implicates the other, and the other person does not confess, the one who has collaborated with the police will go free, while the other person will go to prison for, say, 20 years on the maximum penalty. The strategies in this case are: confess or do not confess. The payoffs, herein *penalties*, are the sentences served. The problem can be expressed compactly in a payoff table of a kind that has become standard in game theory as provided in Table 9.8. The entries of this table mean that each prisoner chooses one of the two strategies, i.e., the first suspect chooses a row and the second suspect chooses a column. The two numbers in each cell of the table provide the outcomes for the two suspects for the corresponding pair of strategies chosen by the suspects as an ordered pair. The number to the left of the comma is the payoff to the person who chooses the rows, i.e., the first suspect, whereas the number to the right of the comma is the payoff to the person who chooses the columns, i.e., the second suspect. Thus reading down the first column, if they both confess, each gets 10 years, but if the second suspect confesses and first suspect does not, the first suspect gets 20 years and second suspect goes free. This example is not a zero-sum game because the payoffs are all losses. However, many problems can be cast with losses (negative numbers) and gains (positive numbers) with a total for each cell in the payoff table. A problem with a payoff table such that the payoffs in each cell add up to zero is called a *zero-sum game*.

TABLE 9.8 Payoff Table in Years for the Prisoners' Dilemma Game

		Second Suspect	
		Confess	Don't Confess
First suspect	Confess	(10, 10)	(0, 20)
	Don't confess	(20, 0)	(1, 1)

The solution to this problem needs to be based on identifying rational strategies that can be based on both persons wanting to minimize the time they spend in jail. One suspect might reason that "two things can happen: either the other suspect confesses or keeps quiet. Suppose the second suspect confesses, I will get 20 years if I don't confess, 10 years if I do; therefore, in this case, it's best to confess. On the other hand, if the other suspect doesn't confess, and I don't either, I get a year; but in that case, if I confess I can go free. Either way, it's best if I confess. Therefore, I'll confess." But the other suspect can and presumably will reason in the same way. In this case, they both confess and go to prison for 10 years each, although if they had acted irrationally, and kept quiet, they each could have gotten off with one year each. The rational strategies of the two suspects have fallen into something called *dominant strategy equilibrium*. The meaning of the term dominant strategy equilibrium requires defining the term dominant strategy that results from an individual player (the suspect, in this case) in a game evaluating separately each of the strategy combinations being faced, and, for each combination, choosing from these strategies the one that gives the greatest payoff. If the same strategy is chosen for each of the different combinations of strategies the player might face, that strategy is called a dominant strategy for that player in that game. The dominant strategy equilibrium occurs if, in a game, each player has a dominant strategy, and each player plays the dominant strategy, then that combination of (dominant) strategies and the corresponding payoffs are said to constitute the dominant strategy equilibrium for that game. In the prisoners' dilemma game, to confess is a dominant strategy, and when both suspects confess, dominant strategy equilibrium is established. The dominant strategy equilibrium is also called *Nash equilibrium*. The definition of Nash equilibrium is a set of strategies with the property that no player can benefit by changing his/her strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute the Nash equilibrium.

The prisoners' dilemma game is based on two strategies per suspect that can be viewed as deterministic in nature, i.e., nonrandom. In general, many games, especially ones permitting repeatability in choosing strategies by players, can be constructed with strategies that have associated probabilities. For example, strategies can be constructed based on probabilities of 0.4 and 0.6 that sum to one. Such strategies with probabilities are called *mixed strategies* as opposed to *pure strategies* that do not involve probabilities of the prisoners' dilemma game. A mixed strategy occurs in a game if a player chooses among two or more strategies at random according to specific probabilities.

In general, gaming could involve more than two players. In the prisoners' dilemma game, a third player that could be identified is the authority and its strategies. The solution might change as a result of adding the strategies of this third player. The use of these concepts in risk analysis and mitigation needs further development and exploration.

Risk Methods for Protecting Infrastructure and Key Resources. The protection of critical infrastructure and key resources (CI/KR) for homeland security requires choosing among a large set of protective, response, and recovery actions for reducing risk to an acceptable level. The selection of investment alternatives for improving asset security and increasing infrastructure resilience depends on two factors: their cost to implement and relative cost-effectiveness. To accomplish this task, the Department of Homeland Security (DHS) has identified risk methods as the primary underlying framework for system evaluations, operational assessments, technology assessments, resource and support analyses, and field operations analyses. According to the draft DHS National Infrastructure Protection Plan, cost–benefit analysis is the hallmark of critical infrastructure protection decision making. Figure 9.14 shows a methodology for informing decisions relating to CI/KR protection

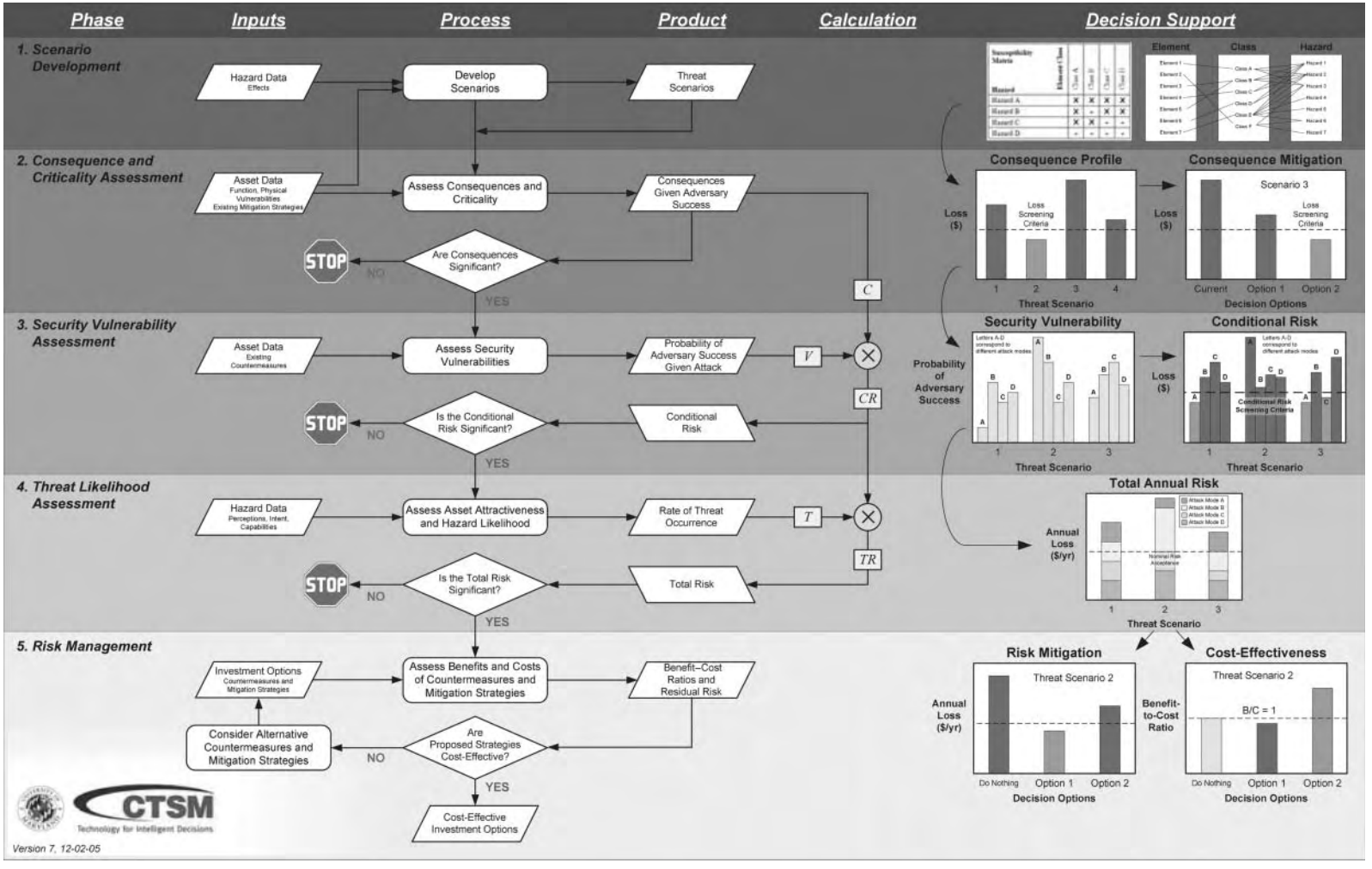

FIGURE 9.14 Critical Asset and Portfolio risk analysis methodology.

(MEMA 2006). It is a quantitative, asset-driven framework for assessing the risk of consequential CI/KR exploitation and disruption due to a variety of plausible threat scenarios and hazard events. This process is divided into five phases as shown in the Figure 9.14. The results from each phase supports decision making of some type, whether it be asset screening and selection based on consequences and their criticality, or countermeasure selection and evaluation based on the results from the security vulnerability assessment phase.

9.1.3.8 Economic and Financial Risks

Economic and financial risks can be grouped into categories that include market risks, credit risks, operation risks, and reputation risks. These four categories are described in subsequent sections.

Market Risks. Governments and corporations operate in economic and financial environments with some levels of uncertainty and instability. A primary contributor to defining this environment is interest rates. Interest rates can have significant impact on the costs of financing a project, and corporate cash flows and asset values. For example, interest rates in the U.S. shot up in 1979 and peaked in 1981, followed by gradual decline with some fluctuations until 2002.

For projects that target global markets, exchange rate instability can be a major risk source. Exchange rates have been volatile ever since the breakdown of the Bretton Woods system to fixed exchange rates in the early 1970s. An example of a bust-up in exchange rates is the fall of the value of the British sterling and Italian lira as a result of the failure of the exchange rate mechanism in September 1992.

Many projects are dependent on availability of venture capital and the stock performance of corporation thereby introducing another risk source related to stock market volatility. Stock prices rose significantly in the inflationary booms of the early 1970s, then fell considerably a little later. They recovered afterward, and fell again in the early 1981. The market rose to a peak until it crashed in 1987, followed by an increase with some swings until reaching a new peak fueled by Internet technologies until its collapse in 2001.

Other contributing factors to economic and finance instability is commodity prices in general and energy prices in particular, primarily crude oil. The hikes in oil prices in the 1973–1974 affected commodity prices greatly and posed series challenges to countries and corporations.

Another contributing source to volatility is derivatives for commodities, foreign currency exchange rates, and stock prices and indices, among others. Derivatives are defined as contracts whose values or payoffs depend on those of other assets, such as the options to buy commodities in the future or options to sell commodities in the future. They offer not only opportunities for hedging positions and managing risks that can be stabilizing, but also speculative opportunities to others that can be destabilizing and a contributor to volatility.

Credit Risks. Credit risks are associated with potential defaults on notes or bonds, as examples, by corporations, including subcontractors. Also, credit risks can be associated with market sentiments that determine a company likelihood of default that could affect its bond rating and ability to purchase money, and maintain projects and operations.

Operational Risks. Operational risks are associated with several sources that include out-of-control operations risk that could occur when a corporate branch undertake significant risk exposure that is not accounted for by a corporate headquarters leading potentially to its collapse, an example being the British Barings Bank that collapsed primarily as a result of its failure to control the market exposure being created within a small overseas branch of the bank.

Another risk source in this category is liquidity risk in which a corporation needing funding more than it can arrange. Also, it could include money transfer risks and agreement breach.

Operational risks include model risks. Model risks are associated with the models and underlying assumptions used to incorrectly value financial instruments and cash flows.

Reputation Risks. The loss of business attributable to decrease in a corporation's reputation can pose another risk source. This risk source can affect its credit rating, ability to maintain clients, workforce, etc. This risk source usually occurs at a slow attrition rate. It can be an outcome of poor management decisions and business practices.

9.1.3.9 Data Needs for Risk Assessment

In risk assessment, the methods of probability theory are used to represent engineering uncertainties. In this context, it refers to event occurrence likelihoods that occur with periodic frequency, such as weather, yet also to conditions that are existent but unknown, such as probability of an extreme wave. It applies to the magnitude of an engineering parameter, yet also to the structure of a model. By contrast, probability is a precise concept. It is a mathematical concept with an explicit definition. The mathematics of probability theory are used to represent uncertainties, despite that those uncertainties are of many forms.

The term *probability* has a precise mathematical definition, but its meaning when applied to the representation of uncertainties is subject to differing interpretations. The frequentist view holds that probability is the propensity of a physical system in a theoretically infinite number of repetitions; i.e., the frequency of occurrence of an outcome in a long series of similar trials (for example, the frequency of a coin landing heads-up in an infinite number of flips is the probability of that event). In contrast, the Bayesian view holds that probability is the rational degree of belief that one holds in the occurrence of an event or the truth of a proposition; probability is manifest in the willingness of an observer to take action upon this belief. This latter view of probability, which has gained wide acceptance in many engineering applications, permits the use of quantified professional judgment in the form of subjective probabilities. Mathematically, such subjective probabilities can be combined or operated on as any other probability.

Data are needed to perform quantitative risk assessment or provide information to support qualitative risk assessment. Information may be available if data have been maintained on a system and components of interest. The relevant information for risk assessment included the possible failures, failure probabilities, failure rates, failure modes, possible causes, and failure consequences. In the case of a new system, data may be used from similar systems if this information is available. Surveys are a common tool used to provide some means of data. Statistical analysis can be used to assess confidence intervals and uncertainties in estimated parameters of interest. Expert judgment may also be used as another source of data (Ayyub 2002). The uncertainty with the quality of the data should be identified to assist in the decision making process.

Data can be classified to including generic and project or plant specific types. Generic data are information from similar systems and components. This information may be the only information available in the initial stages of system design. Therefore, potential differences due to design or uncertainty may result from using generic data on a specific system. Plant-specific data are specific to the system being analyzed. This information is often developed after the operation of a system. Relevant data need to be identified and collected as data collection can be costly. The data collected can then be used to update the risk assessment. Bayesian techniques can be used to combine objective and subjective data.

Data can be classified as failure probability data and failure consequence data. The failure probability data can include failure rates, hazard functions, times between failures, results from reliability studies, and any influencing factors and their effects. Failure consequence data include loss reports, damages, litigation outcomes, repair costs, injuries, and human losses. Also included are influencing factors and effects of failure prevention and consequence mitigation plans. Areas of deficiency in terms of data availability should be identified, and sometimes failure databases need to be constructed. Data deficiency can be used as a basis for data collection and expert-opinion elicitation.

9.1.4 Risk Management and Control

Adding risk control to risk assessment produces risk management. Risk management is the process by which system operators, managers, and owners make safety decisions, regulatory changes, and choose different system configurations based on the data generated in the risk assessment. Risk management involves using information from the previously described risk assessment stage to make educated decisions about system safety. Risk control includes failure prevention and consequence mitigation.

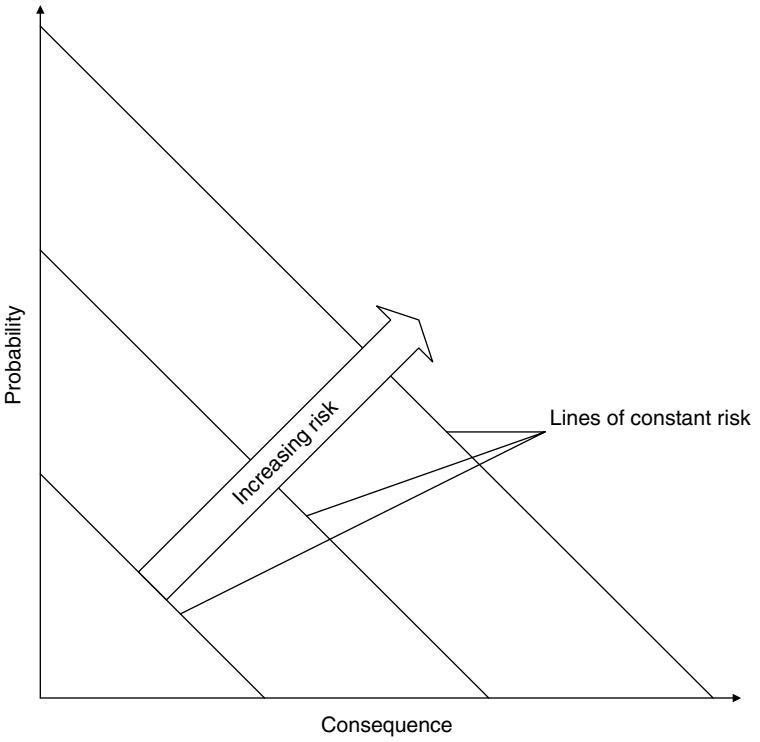

FIGURE 9.15 Risk plot.

Risk management requires the optimal allocation of available resources in support of group goals. Therefore, it requires the definition of acceptable risk, and comparative evaluation of options and/or alternatives for decision making. The goals of risk management are to reduce risk to an acceptable level and/or prioritize resources based on comparative analysis. Risk reduction is accomplished by preventing an unfavorable scenario, reducing the frequency, and/or reducing the consequence. A graph showing the risk relationship is shown in Figure 9.15 as linear contours of constant risk, although due to risk aversion these lines are commonly estimated as nonlinear curves and should be treated as nonlinear curves. Moreover, the vertical axis is termed as probability whereas it is commonly expressed as an annual exceedence probability or frequency as shown in Figure 9.1. In cases involving qualitative assessment, a matrix presentation can be used as shown in Figure 9.8. The figure shows probability categories, severity categories, and risk ratings. A project's base value is commonly assumed as zero. Each risk rating value requires a different mitigation plan.

9.1.4.1 Risk Acceptance

Risk acceptance constitutes a definition of safety as discussed in previous sections. Therefore, risk acceptance is considered a complex and controversial subject that is often subject to debate. The determination of acceptable levels of risk is important to determine the risk performance a system needs to achieve to be considered safe. If a system has a risk value above the risk acceptance level, actions should be taken to address safety concerns and improve the system through risk reduction measures. One difficulty with this process is defining acceptable safety levels for activities, industries, structures, etc. Because the acceptance of risk depends upon society's perceptions, the acceptance criteria do not depend on the risk value alone. This section describes several methods that have been developed to assist in determining acceptable risk values as summarized in Table 9.9.

Risk managers make decisions based on risk assessment and other considerations including economical, political, environmental, legal, reliability, producibility, safety, and other factors. The answer to the question "How safe is safe enough?" is difficult and constantly changing due to different perceptions and understandings of risk. To determine "acceptable risk," managers need to analyze alternatives for the best choice. In some industries, an acceptable risk has been defined by consensus. For example, the U.S. Nuclear Regulatory Commission requires that reactors be designed such that the probability of a large radioactive release to the environment from a reactor incident shall be less than 1×10^{-6} per year. Risk levels for certain carcinogens and pollutants have also been given acceptable concentration levels based on some assessment of acceptable risk. However, risk acceptance for many other activities are not stated.

For example, qualitative implications for risk acceptance are identified in the several existing maritime regulations. The International Maritime Organization High Speed Craft Code and the U.S. Coast Guard Navigation and Vessel Inspection Circular (NVIC) 5–93 for passenger submersible guidance both state that if the end effect is hazardous or catastrophic, a backup system and a corrective operating procedure is required. These references also state that a single failure must not result in a catastrophic event, unless the likelihood is extremely remote.

Often the level of risk acceptance with various activities is implied. Society has reacted to risks through the developed level of balance between risk and potential benefits. Measuring this balance of accepted safety levels for various risks provides a means for assessing society values. These threshold values of acceptable risk depend on a variety of issues including the activity type, industry, and users, and the society as a whole.

Target risk or reliability levels are required for developing procedures and rules for ship structures. For example, the selected reliability levels determine the probability of failure of structural components. The following three methods were used to select target reliability values:

1. Agreeing upon a reasonable value in cases of novel structures without prior history
2. Calibrating reliability levels implied in currently successfully used design codes
3. Choosing target reliability level that minimizes total expected costs over the service life of the structure for dealing with design for which failure results in only economic losses and consequences

The first approach can be based on expert-opinion elicitation. The second approach, called *code calibration*, is the most commonly used approach as it provides the means to build on previous experiences. For example, rules provided by classification and industry societies can be used to determine

TABLE 9.9 Methods for Determining Risk Acceptance

Risk Acceptance Method	Summary
Risk conversion factors	This method addresses the attitudes of the public about risk through comparisons of risk categories. It also provides an estimate for converting risk acceptance values between different risk categories
Farmer's curve	It provides an estimated curve for cumulative probability risk profile for certain consequences (e.g., deaths). It demonstrates graphical regions of risk acceptance/nonacceptance
Revealed preferences	Through comparisons of risk and benefit for different activities, this method categorizes society preferences for voluntary and involuntary exposure to risk.
Evaluation of magnitude of consequences	This technique compares the probability of risks to the consequence magnitude for different industries to determine acceptable risk levels based on consequence
Risk effectiveness	It provides a ratio for the comparison of cost to the magnitude of risk reduction. Using cost–benefit decision criteria, a risk reduction effort should not be pursued if the costs outweigh the benefits. This may not coincide with society values about safety
Risk comparison	The risk acceptance method provides a comparison between various activities, industries, etc., and is best suited to comparing risks of the same type

TABLE 9.10 Risk Conversion Values for Different Risk Factors

Risk Factors	Risk Conversion (RF) Factor	Computed RF Value*
Origin	Natural/human made	20
Severity	Ordinary/catastrophic	30
Volition	Voluntary/involuntary	100
Effect	Delayed/immediate	30
Controllability	Controlled/uncontrolled	5 to 10
Familiarity	Old/new	10
Necessity	Necessary/luxury	1
Costs	Monetary/nonmonetary	NA
Origin	Industrial/ Regulatory	NA
Media	Low profile/ high profile	NA

* NA, not available.

the implied reliability and risk levels in respective rules and codes, then target risk levels can be set in a consistent manner, and new rules and codes can be developed to produce future designs and vessels that are of similar levels that offer reliability and/or risk consistency. The third approach can be based on economic and trade-off analysis. In subsequent sections, the methods of Table 9.9 for determining risk acceptance are discussed.

Risk Conversion Factors. Analysis of risks shows that there are different taxonomies that demonstrate the different risk categories, often called *risk factors*. These categories can be used to analyze risks on a dichotomous scale comparing risks that invoke the same perceptions in society. For example, the severity category may be used to describe both ordinary and catastrophic events. Grouping events that could be classified as ordinary and comparing the distribution of risk to a similar grouping of catastrophic categories yields a ratio describing the degree of risk acceptance of ordinary events as compared to catastrophic events. The comparison of various categories determined the risk conversion values as provided in Table 9.10. These factors are useful in comparing the risk acceptance for different activities, industries, etc. By computing the acceptable risk in one activity, an estimate of acceptable risk in other activities can be calculated based on the risk conversion factors. A comparison of several common risks based on origin and volition is shown in Table 9.11.

Farmer's Curve. The Farmer's curve is graph of the cumulative probability vs. consequence for some activity, industry, or design, as shown in Figure 9.1 and Figure 9.2. This curve introduces a probabilistic approach in determining acceptable safety limits. Probability (or frequency) and consequence values are calculated for each level of risk generating a curve that is unique to hazard of concern. The area to the right (outside) of the curve is generally considered unacceptable because the probability and consequence values are higher than the average value delineated by the curve. The area to the left (inside) of the curve is considered acceptable because probability and consequence values are less than the estimated valve of the curve.

TABLE 9.11 Classification of Common Risks

		Voluntary		Involuntary	
Source	Size	Immediate	Delayed	Immediate	Delayed
Human Made	Catastrophic	Aviation		Dam failure building fire nuclear accident	Pollution building fire
	Ordinary	Sports boating automobiles	Smoking occupation carcinogens	Homicide	
Natural	Catastrophic			Earthquakes hurricanes tornadoes epidemics	
	Ordinary			Lighting animal bites	Disease

Method of Revealed Preferences. The method of revealed preferences provides a comparison of risk versus benefit and categorization for different risk types. The basis for this relationship is that risks are not taken unless there is some form of benefit. Benefit may be monetary or some other item of worth such as pleasure. The different risk types are for the risk category of voluntary versus involuntary actions as shown in Figure 9.16.

Magnitudes of Risk Consequence. Another factor affecting the acceptance of risk is the magnitude of consequence of the event that can result from some failure. In general, the larger the consequence, the less the likelihood that this event may occur. This technique has been used in several industries to demonstrate the location of the industry within societies' risk acceptance levels based on consequence magnitude as shown in Figure 9.17. Further evaluation has resulted in several estimates for the relationship.

January 09, 2006 9–45 between the accepted probability of failure and the magnitude of consequence for failure as provided by Allen in 1981 and called herein the CIRIA (Construction Industry Research and Information Association) equation:

$$P_f = 10^{-4}\frac{KT}{n} \tag{9.8}$$

where T is the life of the structure, K is a factor regarding the redundancy of the structure, and n is the number of people exposed to risk. Another estimate is Allen's equation that is given by:

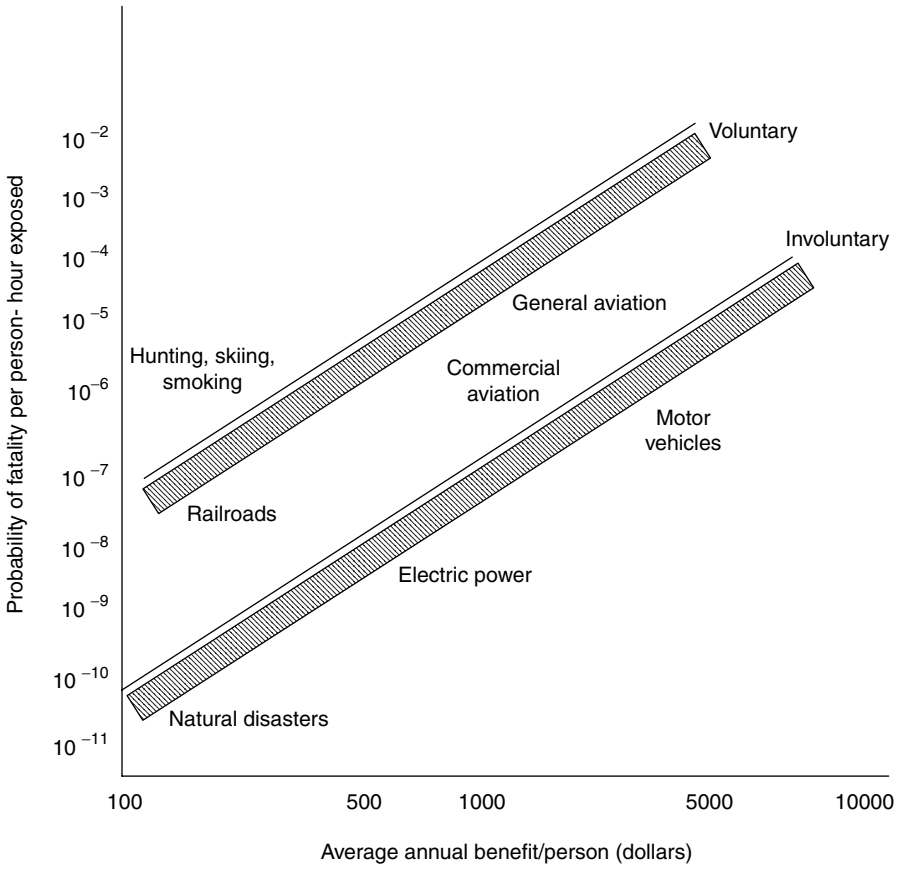

FIGURE 9.16 Accepted rist of voluntary and involuntary activities.

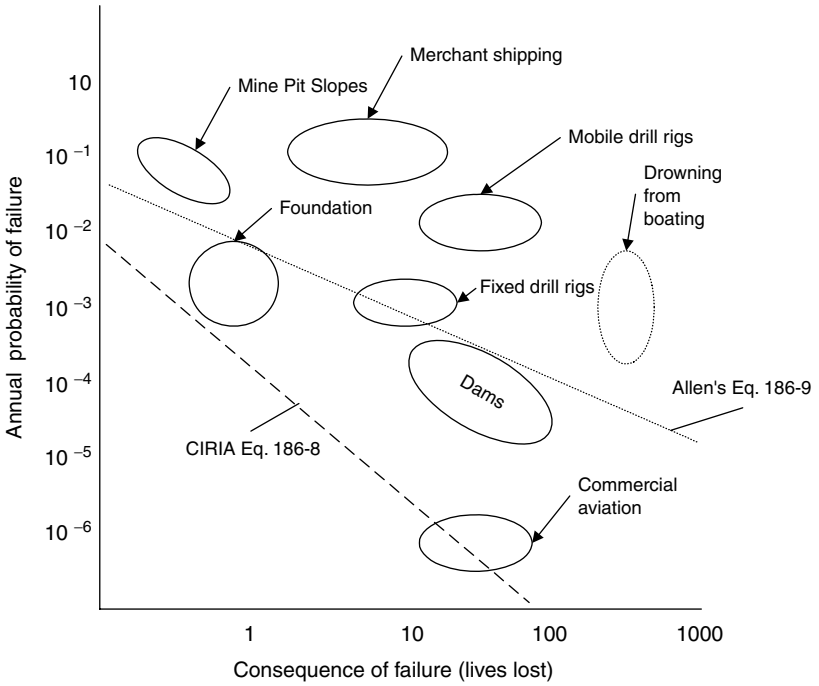

FIGURE 9.17 Comparison of risk and control costs.

$$P_f = 10^{-5}\frac{TA}{W\sqrt{n}} \tag{9.9}$$

where T is the life of the structure, n is the number of persons exposed to risk, and A and W are factors regarding the type and redundancy of the structure. Equation 9.8 offers a lower bound, whereas Equation 9.9 offers a middle line.

Risk Reduction Cost-Effectiveness Ratio. Another measuring tool to assess risk acceptance is the determination of risk reduction effectiveness:

$$Risk\ Reduction\ Effectiveness = \frac{Cost}{\Delta\ Risk} \tag{9.10}$$

where the cost should be attributed to risk reduction, and $\Delta Risk$ is the level of risk reduction as follows:

$$\Delta\ Risk = (Risk\ before\ mitigation\ action) - (Risk\ after\ mitigation\ action) \tag{9.11}$$

The difference in Equation 9.11 is also called the *benefit attributed to a risk reduction action*. Risk effectiveness can be used to compare several risk reduction efforts. The initiative with the smallest risk effectiveness provides the most benefit for the cost. Therefore, this measurement may be used to help determine an acceptable level of risk. The inverse of this relationship may also be expressed as cost-effectiveness.

Risk Comparisons. This technique uses the frequency of severe incidents to directly compare risks between various areas of interest to assist in justifying risk acceptance. Risks can be presented in different ways that can impact how the data are used for decisions. Often values of risk are manipulated in different forms for comparison reasons as demonstrated in Table 9.12. Comparison of risk values should be taken in the context of the values' origin and uncertainties involved.

TABLE 9.12 Ways to Identify Risk of Death

Ways to Identify Risk of Death	Summary
Number of fatalities	This measure shows the impact in terms of the number of fatalities on society. Comparison of these values is cautioned since the number of persons exposed to the particular risk may vary. Also, the time spent performing the activity may vary. Different risk category types should also be considered to compare fatality rates
Annual mortality rate/individual	This measure shows the mortality risk normalized by the exposed population. This measure adds additional information about the number of exposed persons; however, the measure does not include the time spent on the activity.
Annual mortality	This measure provides the most complete risk value since the risk is normalized by the exposed population and the duration of the exposure.
Loss of life exposure (LLE)	This measure converts a risk into a reduction in the expected life of an individual. It provides a good means of communicating risks beyond probability values.
Odds	This measure is a layman format for communicating probability, for example, 1 in 4.

This technique is most effective for comparing risks that invoke the same human perceptions and consequence categories. Comparing risks of different categories is cautioned because the differences between risk and perceived safety may not provide an objective analysis of risk acceptance. The use of risk conversion factors may assist in transforming different risk categories. Conservative guidelines for determining risk acceptance criteria can be established for voluntary risks to the public from the involuntary risk of natural causes.

9.1.4.2 Rankings Based on Risk Results

Another tool for risk management is the development of risk ranking. The elements of a system within the objective of analysis can be analyzed for risk and consequently ranked. This relative ranking may be based on the failure probabilities, failure consequences, risks, or other alternatives with concern towards risk. Generally, risk items ranked highly should be given high levels of priority; however, risk management decisions may consider other factors such as costs, benefits, and effectiveness of risk reduction measures. The risk ranking results may be presented graphically as needed.

9.1.4.3 Decision Analysis

Decision analysis provides a means for systematically dealing with complex problems to arrive at a decision. Information is gathered in a structured manner to provide the best answer to the problem. A decision generally deals with three elements: alternatives, consequences, and preferences. The alternatives are the possible choices for consideration. The consequences are the potential outcomes of a decision. Decision analysis provides methods for quantifying preference tradeoffs for performance along multiple decision attributes while taking into account risk objectives. Decision attributes are the performance scales that measure the degree to which objectives are satisfied. For example, one possible attribute is reducing lives lost for the objective of increasing safety. Additional examples of objectives may include minimize the cost, maximize utility, maximize reliability, and maximize profit. The decision outcomes may be affected by uncertainty; however, the goal is to choose the best alternative with the proper consideration of uncertainty. The analytical depth and rigor for decision analysis depends on the desired detail in making the decision. Cost–benefit analysis, decision trees, influence diagrams and the analytic hierarchy process are some of the tools to assist in decision analysis. Also, decision analysis should consider constraints, such as availability of system for inspection, availability of inspectors, preference of certain inspectors, and availability of inspection equipment.

9.1.4.4 Cost–Benefit Analysis

Risk managers commonly weigh various factors including cost and risk. The analysis of three different alternatives is shown graphically in Figure 9.18 as an example. The graph shows that alternative (C) is the best choice because the level of risk and cost is less than alternatives (A) and (B). However, if the only alternatives were A and B, the decision would be more difficult. Alternative (A) has higher cost and lower risk than alternative (B); alternative (B) has higher risk but lower cost than alternative (A). A risk manager needs to weigh the importance of risk and cost in making this decision and availability of resources, and make use of risk-based decision analysis.

Risk–benefit analysis can also be used for risk management. Economic efficiency is important to determine the most effective means of expending resources. At some point, the costs for risk reduction do not provide adequate benefit. This process compares the costs and risk to determine where the optimal risk value is on a cost basis. This optimal value occurs, as shown in Figure 9.19, when costs to control risk are equal to the risk cost due to the consequence (loss). Investing resources to reduce low risks below this equilibrium point is not providing a financial benefit. This technique may be used when cost values can be attributed to risks. This analysis might be difficult to perform for certain risk such as risk to human health and environmental risks because the monetary values are difficult to estimate for human life and the environment.

The present value of incremental costs and benefits can be assessed and compared among alternatives that are available for risk mitigation or system design. Several methods are available to determine which, if any, option is most worth pursuing. In some cases, no alternative will generate a net benefit relative to the base case. Such a finding would be used to argue for pursuit of the base case scenario. The following are the most widely used present value comparison methods: (1) net present value (NPV), (2) benefit-cost ratio, (3) internal rate of return, and (4) payback period. The net present value (NPV) method requires that each alternative need to meet the following criteria to warrant investment of funds: (1) having a positive NPV; and (2) having the highest NPV of all alternatives considered. The first condition insures that the alternative is worth undertaking relative to the base case, e.g., it contributes more in incremental benefits than it absorbs in incremental costs. The second condition insures that maximum benefits are obtained in a situation of unrestricted access to capital funds. The NPV can be calculated as follows:

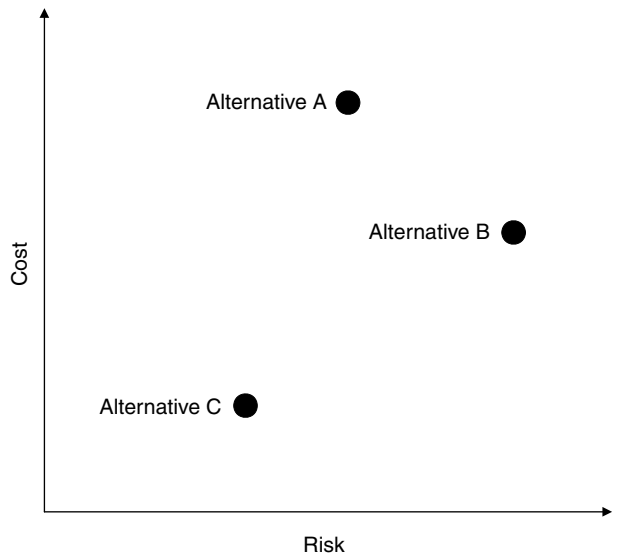

FIGURE 9.18 Risk benefit for three alternatives. (From CERT.)

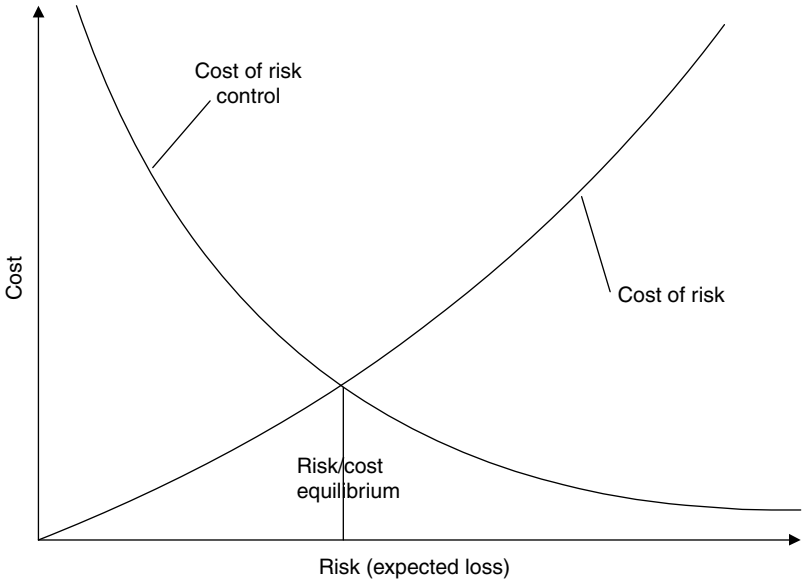

FIGURE 9.19 Comparison of risk and control costs. (From CIAO 2001.)

$$\text{NPV} = \sum_{t=0}^{k} \frac{(B-C)_t}{(1+r)^t} = \sum_{t=0}^{k} \frac{B_t}{(1+r)^t} - \sum_{t=0}^{k} \frac{C_t}{(1+r)^t} \tag{9.12}$$

where B is future annual benefits in constant dollars, C is future annual costs inconstant dollars, r is annual real discount rate, k is number of years from the base year over which the project will be evaluated, and t is an index running from 0 to k representing the year under consideration.

The benefit of a risk mitigation action can be assessed as follows:

$$\text{Benefit} = \text{unmitigated risk} - \text{mitigated risk}. \tag{9.13}$$

The cost in Equation 9.13 is the cost of the mitigation action. The benefit minus the cost of mitigation can be used to justify the allocation of resources. The benefit-to-cost ratio can be computed, and may also be helpful in decision making. The benefit-to-cost ratio (B/C) can be computed as

$$\text{Benefit-to-Cost Ratio } (B/C) = \frac{\text{Benefit}}{\text{Cost}} = \frac{\text{Unmitigated Risk} - \text{Mitigated Risk}}{\text{Cost of Mitigation Action}} \tag{9.14}$$

Cost of mitigation action ratios greater than one are desirable. In general, the larger the ratio, the better the mitigation action.

Accounting for the time value of money would require defining the benefit–cost ratio as the present value of benefits divided by the present value of costs. The benefit–cost ratio can be calculated

as follows:

$$B/C = \frac{\sum_{t=0}^{k} \frac{B_t}{(1+r)^t}}{\sum_{t=0}^{k} \frac{C_t}{(1+r)^t}} \tag{9.15}$$

where *Bt* is future annual benefits in constant dollars, *Ct* is future annual costs inconstant dollars, *r* is annual real discount rate, and *t* is an index running from 0 to k representing the year under consideration. A proposed activity with a *B*/*C* ratio of discounted benefits to costs of 1 or more is expected to return at least as much in benefits as it costs to undertake, indicating that the activity is worth undertaking.

The internal rate of return (*IRR*) is defined as the discount rate that makes the present value of the stream of expected benefits in excess of expected costs zero. In other words, it is the highest discount rate at which the project will not have a negative *NPV*. To apply the *IRR* criterion, it is necessary to compute the IRR and then compare it with a base rate of, say, a 7% discount rate. If the real *IRR* is less than 7%, the project would be worth undertaking relative to the base case. The *IRR* method is effective in deciding whether or not a project is superior to the base case; however it is difficult to utilize it for ranking projects and deciding among mutually exclusive alternatives. Project rankings established by the *IRR* method might be inconsistent with those of the *NPV* criterion. Moreover, a project might have more than one *IRR* value, particularly when a project entails major final costs, such as cleanup costs. Solutions to these limitations exist in capital budgeting procedures and practices that are often complicated or difficult to employ in practice and present opportunities for error.

The payback period measures the number of years required for net undiscounted benefits to recover the initial investment in a project. This evaluation method favors projects with near-term and more certain benefits, and fails to consider benefits beyond the payback period. The method does not provide information on whether an investment is worth undertaking in the first place.

The previous models for cost–benefit analysis presented in this section do not account for the full probabilistic characteristics of *B* and *C* in their treatment. Concepts from reliability assessment 4 can be used for this purpose. Assuming *B* and *C* to normally distributed, a benefit–cost index ($\beta_{B/C}$) can be defined as follows:

$$\beta_{B/C} = \frac{\mu_B - \mu_C}{\sqrt{\sigma_B^2 + \sigma_C^2}} \tag{9.16}$$

where μ and σ are the mean and standard deviation. The failure probability can be computed as

$$P_{f,B/C} = P(C > B) = 1 - \Phi(\beta) \tag{9.17}$$

In the case of lognormally distributed *B* and *C*, the benefit–cost index ($\beta_{B/C}$) can be computed as

$$\beta_{B/C} = \frac{\ln\left(\frac{\mu_B}{\mu_C}\sqrt{\frac{\delta_C^2+1}{\delta_B^2+1}}\right)}{\sqrt{\ln[(\delta_B^2+1)(\delta_C^2+1)]}} \tag{9.18}$$

where δ is the coefficient of variation. Equation (9.18) also holds for the case of lognormally distributed *B* and *C*. In the case of mixed distributions or cases involving basic random variables of *B* and *C*, the advanced second moment method or simulation method can be used. In cases where benefit is computed as revenue minus cost, benefit might be correlated with cost requiring the use of other methods.

Example 9.2 Protection of Critical Infrastructure

This example is used to illustrate the cost of cost–benefit analysis using a simplified decision situation. As an illustration, assume that there is a 0.01 probability of an attack on a facility containing hazardous material during the next year. If the attack occurs, the probability of a serious release to the public is 0.01 with a total consequence of $100B. The total consequence of an unsuccessful attack is negligible. The unmitigated risk can therefore be computed as

$$\textit{Unmitigated Risk} = 0.01(\$0.01)(100\text{B}) = \ \$\ 10\text{M}.$$

If armed guards are deployed at each facility, the probability of attack can be reduced to 0.001 and the probability of serious release if an attack occurs can be reduced to 0.001. The cost of the guards for all plants is assumed to be \$100 M per year. The mitigated risk can therefore be computed as

$$\textit{Mitigated Risk} = 0.001(0.001)(\$100\text{B}) = \$0.10\text{M}.$$

The benefit in this case is

$$\textit{Benefit} = \$10\text{M} - \$0.1\text{M}\ \textit{or}\ \sim \$10\text{M}.$$

The benefit-to-cost ratio is about 0.1. Therefore, the \$100 M cost might be difficult to justify.

9.1.4.5 Risk Mitigation

A risk mitigation strategy can be presented from a financial point of view. Risk mitigation in this context can be defined as an action to either reduce the probability of an adverse event occurring or to reduce the adverse consequences if it does occur. This definition captures the essence of an effective management process of risk. If implemented correctly a successful risk mitigation strategy should reduce any adverse (or downside) variations in the financial returns from a project, which are usually measured by either (1) the net present value (NPV) defined as the difference between the present value of the cash flows generated by a project and its capital cost and calculated as part of the process of assessing and appraising investments, or (2) the internal rate of return (IRR) defined as the return that can be earned on the capital invested in the project, i.e., the discount rate that gives an NPV of zero, in the form of the rate that is equivalent to the yield on the investment.

Risk mitigation involves direct costs like increased capital expenditure or the payment of insurance premiums; hence might reduce the average overall financial returns from a project. This reduction is often a perfectly acceptable outcome, given the risk aversion of many investors and lenders. A risk mitigation strategy is the replacement of an uncertain and volatile future with one where there is less exposure to adverse risks and so less variability in the return, although the expected NPV or IRR may be reduced. These two aspects are not necessarily mutually exclusive. Increasing risk efficiency by simultaneously improving the expected NPV or IRR and simultaneously reducing the adverse volatility is sometimes possible and should be sought. Risk mitigation should cover all phases of a project from inception to closedown or disposal.

Four primary ways are available to deal with risk within the context of a risk management strategy as follows:

- Risk reduction or elimination
- Risk transfer, e.g., to a contractor or an insurance company
- Risk avoidance
- Risk absorbance or pooling

These four methods are described in subsequent sections.

Risk Reduction or Elimination. Risk reduction or elimination is often the most fruitful form for exploration. For example, could a design of a system be amended so as to reduce or eliminate either the probability of occurrence of a particular risk event or the adverse consequences if it occurs? Alternatively, could the risks be reduced or eliminated by retaining the same design but using different materials or a different method of assembly? Other possible risk mitigation options in this category include as examples: a better labor relations policy to minimize the risk of stoppages, training of staff to avoid hazards, better site security to avoid theft and vandalism, a preliminary investigation of possible site pollution, advance ordering of key components, noise abetment measures, good signposting, and liaisons with the local community.

Risk Transfer. A general principle of an effective risk management strategy is that commercial risks in projects and other business ventures should be borne wherever possible by the party that is best able to manage them, and thus mitigate the risks. Contracts and financial agreements are the principal forms to transfer risks. Companies specializing in risk transfer can be consulted that could appropriately meet the needs of a project. Risks can be transferred alternately to an insurance company which, in return for a payment (i.e., premium) linked to the probability of occurrence and severity associated with the risk, is obliged by the contract to offer compensation to the party affected by the risk. Insurance coverage can range from straight insurance for expensive risks with a low probability, such as fire, through performance bonds, which ensure that the project will be completed if the contractor defaults, to sophisticated financial derivatives such as hedge contracts to avoid such risks as unanticipated losses in foreign exchange markets.

Risk Avoidance. A most intuitive way of avoiding a risk is to avoid undertaking the project in a way that involves that risk. For example, if the objective is to generate electricity but a nuclear power source, although cost-efficient, is considered to have a high risk due to potentially catastrophic consequences, even after taking all reasonable precautions, the practical solution is to turn to other forms of fuel to avoid that risk. Another example would be the risk that a particularly small contractor would go bankrupt. In this case, the risk could be avoided by using a well-established contractor for that particular job.

Risk Absorbance and Pooling. Cases where risks cannot, or cannot economically, be eliminated, transferred, or avoided, they must be absorbed if the project is to proceed. Normally, a sufficient margin in the project's finances needs to be created to cover the risk event should it occur. However, it is not always essential for one party alone to bear all these absorbed risks. Risks can be reduced through pooling possibly through participation in a consortium of contractors, when two or more parties are able to exercise partial control over the incidence and impact of risk. Joint ventures and partnerships are other examples of organizational forms for pooling risks.

Uncertainty Characterization. Risk can be mitigated through proper uncertainty characterization. The presence of improperly characterized uncertainty could lead to higher adverse event occurrence likelihood and consequences. Also, it could result in increasing estimated cost margins as a means of compensation. Therefore, risk can be reduced by a proper characterization of uncertainty. The uncertainty characterization can be achieved through data collection and knowledge construction.

9.1.5 Risk Communication

Risk communication can be defined as an interactive process of exchange of information and opinion among stakeholders such as individuals, groups, and institutions. It often involves multiple messages about the nature of risk or expressing concerns, opinions, or reactions to risk managers or to legal and institutional arrangements for risk management. Risk communication greatly affects risk acceptance and defines the acceptance criteria for safety.

Risk communication provides the vital link between the risk assessors, risk managers, and the public to understand risk. However, this does not necessarily mean that risk communication will always lead to agreement among different parties. An accurate perception of risk provides for rational decision making. The Titanic was deemed an unsinkable ship, yet was lost on its maiden voyage. Space shuttle flights were perceived to be safe enough for civilian travel until the Space Shuttle Challenger disaster. These disasters obviously had risks that were not perceived as significant until after the disaster. Risk communication is a dynamic process that must be considered prior to management decisions.

The communication process deals with technical information about controversial issues. Therefore, it needs to be skillfully performed by risk managers and communicators who might be viewed as adversaries to the public. Risk communication between risk assessors and risk managers is necessary to effectively apply risk assessments in decision making. Risk managers must participate in determining the criteria for determining what risk is acceptable and unacceptable. This communication between the risk managers and risk assessors is necessary for a better understanding of risk analysis in making decisions.

Risk communication also provides the means for risk managers to gain acceptance and understanding by the public. Risk managers need to go beyond the risk assessment results and consider other factors in making decisions. One of these concerns is politics, which is largely influenced by the public. Risk managers often fail to convince the public that risks can be kept to acceptable levels. Problems with this are shown by the public's perception of toxic waste disposal and nuclear power plant operation safety. As a result of the public's perceived fear, risk managers may make decisions that are conservative to appease the public.

The value of risk calculated from risk assessment is not the only consideration for risk managers. All risks are not created equal and society has established risk preferences based on public preferences. Decision makers should take these preferences into consideration when making decisions concerning risk.

To establish a means of comparing risks based on the society preferences, risk conversion factors (RCF) may be used. The RCF expresses the relative importance of different attributes concerning risk. An example of possible risk conversion factors is shown in Table 9.10. These values were determined by inferences of public preferences from statistical data with the consequence of death considered.

For example, the voluntary and involuntary classification depends on whether the events leading to the risk are under the control of the persons at risk or not, respectively. Society, in general, accepts a higher level of voluntary risk than involuntary risk by an estimated factor of 100. Therefore, an individual will accept a voluntary risk that is 100 times greater than an involuntary risk.

The process of risk communication can be enhanced and improved in three aspects: (1) the process, (2) the message, and (3) the consumers. The risk assessment and management process needs to have clear goals with openness, balance, and competence. The contents of the message should account for audience orientation and uncertainty, provide risk comparison, and be complete. There is a need for consumer's guides that introduce risks associated with a specific technology, the process of risk assessment and management, acceptable risk, decision making, uncertainty, costs and benefits, and feedback mechanisms. Improving risk literacy of consumers is an essential component of the risk communication process.

The USACE has a 1992 Engineering Pamphlet (EP) on risk communication (EP 1110-2-8). The following are guiding considerations in communicating risk:

- Risk communication must be free of jargon
- Consensus of expert needs to be established
- Materials must be cited, and their sources must be credible
- Materials must be tailored to audience
- The information must be personalized to the extent possible
- Motivation discussion should stress a positive approach and the likelihood of success
- Risk data must be presented in a meaningful manner

9.2 Electricity Infrastructure Security

Massoud Amin

9.2.1 Introduction

The massive power outages in the United States, Canada, U.K. and Italy in 2003 underscored electricity infrastructure's vulnerabilities (President's Commission on Critical Infrastructure Protection 1997; Amin 2000, 2001, 2002a, 2002b, 2003; U.S. Department of Energy 2002; Energy Information Administration 2003; North American Electric Reliability Council; EPRI 2003). The North American power network may be considered to be the largest and most complex machine in the world; its transmission lines connect all the electric generation and distribution on the continent. In that respect, it exemplifies many of the complexities of electric power infrastructure and how technological innovation combined with efficient markets and enabling policies can address them. This network represents an enormous investment,

including more than 15,000 generators in 10,000 power plants, and hundreds of thousands of miles of transmission lines and distribution networks, whose estimated worth is over US$800 billion. In 2000, transmission and distribution was valued at US$358 billion (EPRI 1999, 2000, 2001, 2003a, 2003b; Hauer and Dagle 1999; Energy Information Administration 2003; North American Electric Reliability Council).

Through the North American electricity infrastructure, every user, producer, distributor, and broker of electricity buys and sells, competes, and cooperates in an "electric enterprise." Every industry, every business, every store, and every home is a participant, active or passive, in this continent-scale conglomerate. Over the last decade and during the next few years, the electric enterprise will undergo dramatic transformation as its key participants—the traditional electric utilities—respond to deregulation, competition, renewable energy portfolio standards, tightening environmental/land-use restrictions, and other global trends.

However, this network has evolved without formal analysis of the system-wide implications of this evolution, including its diminished transmission and generation shock absorber capacity under the forces of deregulation, the digital economy, and interaction with other infrastructures. Only recently, with the advent of deregulation, unbundling, and competition in the electric power industry, has the possibility of power delivery beyond neighboring areas become a key design and engineering consideration, yet the existing grid is still expected to handle a growing volume and variety of long-distance, bulk power transfers. To meet the needs of a pervasively digital world that relies on microprocessor-based devices in vehicles, homes, offices, and industrial facilities, grid congestion and atypical power flows are increasing, as are customer reliability expectations.

The vulnerability of the energy infrastructure to natural disasters and physical attacks has long been recongnized, but this vulnerability has significantly increased in recent years because (1) relatively little infrastructure (especially transmission lines) has been added to handle increased demand with an adequate safety "cushion" (EPRI 2001, NERC 2001) and (2) terrorists might in fact be planning attacks on the system (EPRI 2001, EEI 2003, DOE and DHS 2006).

Ongoing efforts at NERC, DOE, and EPRI—including EPRI's Enterprise Information Security Program and Infrastructure Security Initiative—have highlighted utility-specific threats and the technologies to counteract them. As an example, the Infrastructure Security Assessment, developed by EPRI in response to the September 11, 2001, terrorist attacks, discusses some of the specific threats and recommends counter measures; as noted in these and other reports (EPRI 2001, 2004, 2005, EEI 2003, NERC, 2001, 2004, 2006, DOE and DHS 2006), the existing power delivery system is vulnerable to natural disasters and intentional attack. Regarding the latter, a successful terrorist attempt to disrupt the power delivery system could have adverse effects on national security, the economy, and the lives of every citizen.

The existing power delivery system is vulnerable to natural disasters and intentional attack. Regarding the latter, a successful terrorist attempt to disrupt the power delivery system could have adverse effects on national security, the economy, and the lives of every citizen.

Both the importance and difficulty of protecting power systems have long been recognized. In 1990, the Office of Technology Assessment (OTA) of the U.S. Congress issued a detailed report, *Physical Vulnerability of the Electric System to Natural Disasters and Sabotage,* concluding: "Terrorists could emulate acts of sabotage in several other countries and destroy critical [power system] components, incapacitating large segments of a transmission network for months. Some of these components are vulnerable to saboteurs with explosives or just high-powered rifles." The report also documented the potential cost of widespread outages, estimating them to be in the range of $1 to $5/kWh of disrupted service, depending on the length of outage, the types of customers affected, and a variety of other factors, such as the unintended release of pollutants/wastewater and the corresponding environmental cost, social unrest, damage from looting and arson as components of the total cost.

The reality of a coordinated attack has raised the issue of security to be considered along with power systems' reliability, which posits more random and independent failures. The system's vulnerability to natural disasters and physical attacks has long been recognized, but this vulnerability has significantly increased in recent years, in part because the system is operating closer to its capacity and in part because terrorist attacks are no longer hypothetical.

The situation has become even more complex because accounting for all critical assets includes thousand of transformer, line reactors, series capacitors, and transmission lines. Protection of ALL the widely diverse and dispersed assets is impractical because there are so many involved:

- Over 230,000 miles of HV lines (230 kV and above)
- Over 6,644 transformers in the Eastern Interconnection
- Over 6,000 HV transformers in the North American Interconnection (of which less than 300 are critical assets)
- Control centers
- Interdependence with gas pipelines
- Compressor stations
- Dams
- Rail lines
- Telecommunication equipment (monitoring and control of the system)

In this section, the security, agility, and robustness/survivability of large-scale power delivery infrastructure that face new threats and unanticipated conditions is presented. In addition, focus is placed, in part, on challenges associated with development of a smart self-healing electric power grid for enhanced system security, reliability, efficiency, and quality.

As an example, trends show that worldwide cyber attacks are on the rise; the number of documented attacks and intrusions has been rising very rapidly in recent years. Due to the increasingly sophisticated nature and speed of some malicious code, intrusions, and denial-of-service attacks, human response may be inadequate. Some trends and documented nonsensitive modes of attack are shown in Figure 9.20 and Figure 9.21.

The vulnerability of the energy infrastructure to natural disasters and physical attacks has long been recognized, but this vulnerability has significantly increased in recent years because relatively little infrastructure (especially transmission lines) has been added to handle increased demand with an adequate safety "cushion."[1]

Secure and reliable operation of these systems is fundamental to national and international economy, security and quality of life. Their very interconnectedness makes them more vulnerable to global disruption, initiated locally by material failure, natural calamities, intentional attack, or human error.

In addition to the security problems associated with electric transmission systems additional concerns arise with nuclear power systems. Nuclear power serves the United States as well as many countries around the world as a fairly reliable source of electricity that will not contribute to global warming producing carbon dioxide. In that sense it is a "clean" source of electricity. But there are ways in which nuclear power can be a hazardous technology.

There are two pressing safety concerns with nuclear power in this age of terrorism. The first has to do with the spent fuel storage areas located outside and near the reactors. These spent fuel "swimming pools" are laden with a far higher inventory of radioactive material than is present in the reactor itself and are far less protected from attack than the reactor. If a "9/11" plane had hit the Indian Point Nuclear Plant spent fuel storage area (about 35 miles from New York City), it is possible that millions of people would have had to be evacuated to escape radiation and that parts of New York would have had to be abandoned. There might of course be solutions to avoid this catastrophic scenario, but they would add to the cost of nuclear power.

The other long-term safety concern is the eventual transportation of thousands of high-level nuclear waste canisters across open roads of the United States (through 43 states) on their way to a final resting place, e.g., Yucca Mountain. These shipments would be enticing targets for terrorists. Technology may be able to ameliorate this type of accident, but it is currently not in place.

[1]NERC 2001, "Reliability Assessment 2001–2010"

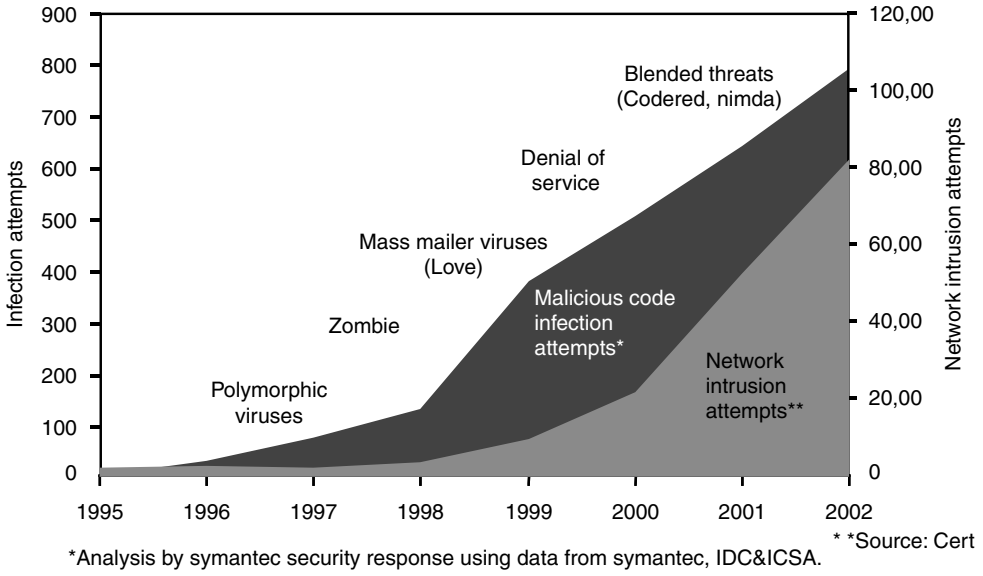

FIGURE 9.20 Attack trends: Documented network intrusion and infection attempts. (From CERT.)

9.2.2 The Electricity Enterprise: Today and Tomorrow

The North American power network's transmission lines connect all generation and distribution on the continent to form a vertically integrated hierarchical network. The question is raised as to whether there is a unifying paradigm for the simulation, analysis, and optimization of time-critical operations (both financial transactions and actual physical control) in these multiscale, multicomponent, and distributed systems. In addition, mathematical models of interactive networks are typically vague (or may not even exist); moreover, existing and classical methods of solution are either unavailable, or are not

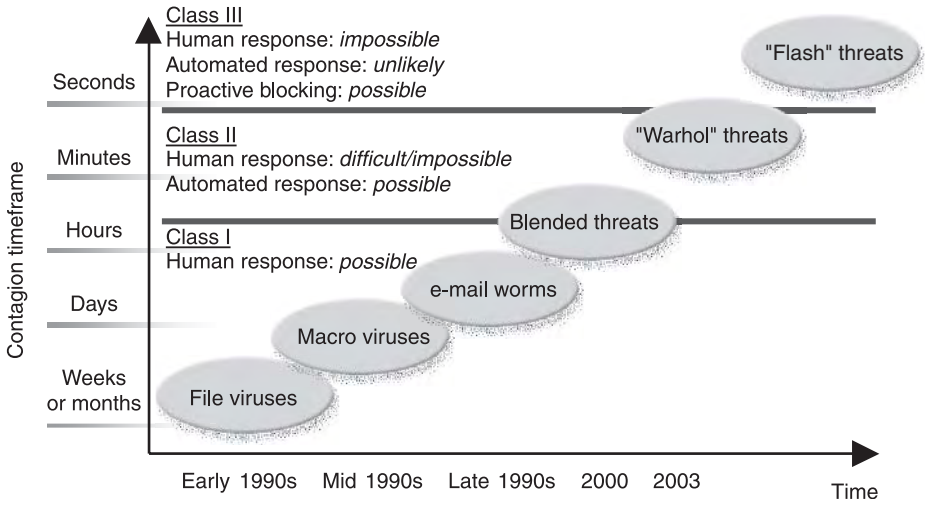

FIGURE 9.21 Malicious codes' threat evolution: Increased sophistication and much faster than a few years ago. (From CIAO 2001.)

sufficiently powerful. For the most part, no present methodologies are suitable for understanding their behavior.

Another important dimension is the effect of deregulation and economic factors on a particular infrastructure. Although other and more populous countries, such as China and India, will have greater potential electricity markets and demands, the United States is presently the largest national market for electric power. Its electric utilities have been mostly privately owned, vertically integrated, and locally regulated. National regulations in areas of safety, pollution, and network reliability also constrain their operations to a degree, but local regulatory bodies, mostly at the state level, have set their prices and their return on investment, and have controlled their investment decisions while protecting them from outside competition. That situation is now rapidly changing; state regulators are moving toward permitting and encouraging a competitive market in electric power.

The electric power grid was historically operated by separate utilities, each independent in its own control area and regulated by local bodies, to deliver bulk power from generation to load areas reliably and economically. As a noncompetitive, regulated monopoly, emphasis was on reliability (and security) at the expense of economy. Competition and deregulation have created multiple energy producers that must share the same regulated energy delivery network. Traditionally, new delivery capacity would be added to handle load increases, but because of the current difficulty in obtaining permits and the uncertainty about achieving an adequate rate of return on investment, total circuit miles added annually are declining while total demand for delivery resources continues to grow. In recent years, the "shock absorbers" have been shrinking; e.g., during the 1990s actual demand in the United States increased some 35%, while capacity increased only 18%, the most visible parts of a larger and growing U.S. energy crisis that is the result of years of inadequate investments in the infrastructure. According to EPRI analyses, since 1995 to the present the amortization/depreciation rate exceeds utility construction expenditures (Figure 9.22).

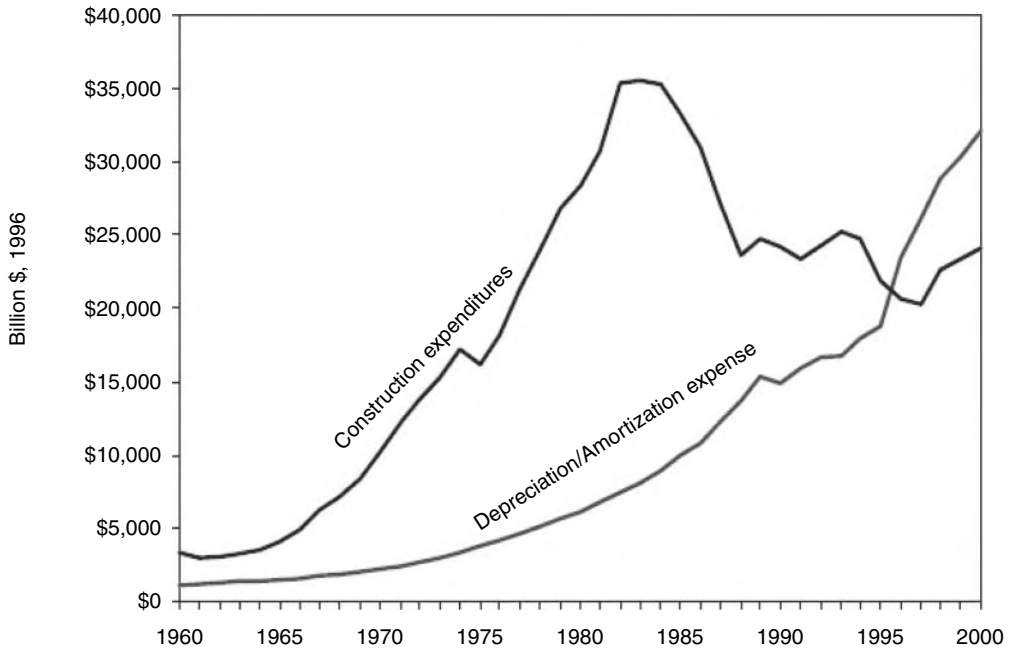

FIGURE 9.22 Since the "crossover" point in about 1995 utility construction expenditures have lagged behind asset depreciation. This has resulted in a mode of operation of the system analogous to "harvesting the farm far more rapidly then planting new seeds." (Data provided by EEI and graph courtesy of EPRI.)

As a result of these "diminished shock absorbers," the network is becoming increasingly stressed, and whether the carrying capacity or safety margin will exist to support anticipated demand is in question. The complex systems used to relieve bottlenecks and clear disturbances during periods of peak demand are at great risk to serious disruption, creating a critical need for technological improvements.

9.2.3 Reliability Issues

Several cascading failures during the past 40 years spotlighted the need to understand the complex phenomena associated with power network systems and the development of emergency controls and restoration. Widespread outages and huge price spikes during the past few years raised public concern about grid reliability at the national level (President's Commission on Critical Infrastructure Protection, 1997; Hauer and Dagle 1999; U.S. Department of Energy 2002; Energy Information Administration, 2003; House Committee on Energy and Commerce, 2003; North American Electric Reliability Council). According to data from the North American Electric Reliability Council (NERC) and analyses from the Electric Power Research Institute (EPRI), average outages from 1984 to the present have affected nearly 700,000 customers per event annually. Smaller outages occur much more frequently and affect tens to hundreds of thousands of customers every few weeks or months, whereas larger outages occur every two to nine years and affect millions. Much larger outages affect seven million or more customers per event each decade. These analyses are based on data collected for the U.S. Department of Energy (DOE), which requires electric utilities to report system emergencies that include electric service interruptions, voltage reductions, acts of sabotage, unusual occurrences that can affect the reliability of bulk power delivery systems, and fuel problems (North American Electric Reliability Council; Hauer and Dagle 1999; Amin 2000; 2001, 2002b, 2003; Energy Information Administration 2003; Amin 2005).

Coupling these analyses with diminished infrastructure investments, and noting that the crossover point for the utility construction investment vs. depreciation occurred in 1995 (Figure 9.20), the number and frequency of major outages along with the number of customers affected during the decade 1991–2000 were analyzed. The data were split into the two time periods: 1991–1995 and 1996–2000 (Figure 19.21). Based on EPRI's analyses (Amin 2003; EPRI 2003) of data in NERC's Disturbance Analysis Working Group (DAWG) database (Amin 2003; Energy Information Administration 2003; North American Electric Reliability Council), 41% more outages affected 50,000 or more consumers in the second half of the 1990s than in the first half (58 outages in 1996–2000 versus 41 outages in 1991–1995). The average outage affected 15% more consumers from 1996 to 2000 than from 1991 to 1995 (average size per event was 409,854 customers affected in the second half of the decade versus 355,204 in the first half of the decade). In addition, there were 76 outages of size 100 MW or more in the second half of the decade, compared to 66 such occurrences in the first half. During the same period, the average lost load caused by an outage increased by 34%, from 798 MW from 1991 to 1995 to 1067 MW from 1996 to 2000 (Figure 9.23) (Amin, 2003; Energy Information Administration, 2003; North American Electric Reliability Council; EPRI 2003).

Electric power utilities typically own and operate at least parts of their own telecommunications systems which often consist of backbone fiber-optic or microwave connecting major substations, with spurs to smaller sites. Increased use of electronic automation raises significant issues regarding the adequacy of operational security. As is true of other critical infrastructures, increased use of automated technologies raises significant security issues, however:

- Reduced personnel at remote sites makes the sites more vulnerable to hostile threats;
- Interconnecting automation and control systems with public data networks makes them accessible to individuals and organizations, from any worldwide location using an inexpensive computer and a modem; and
- Use of networked electronic systems for metering, scheduling, trading, or e-commerce imposes numerous financial risks associated with network failures.

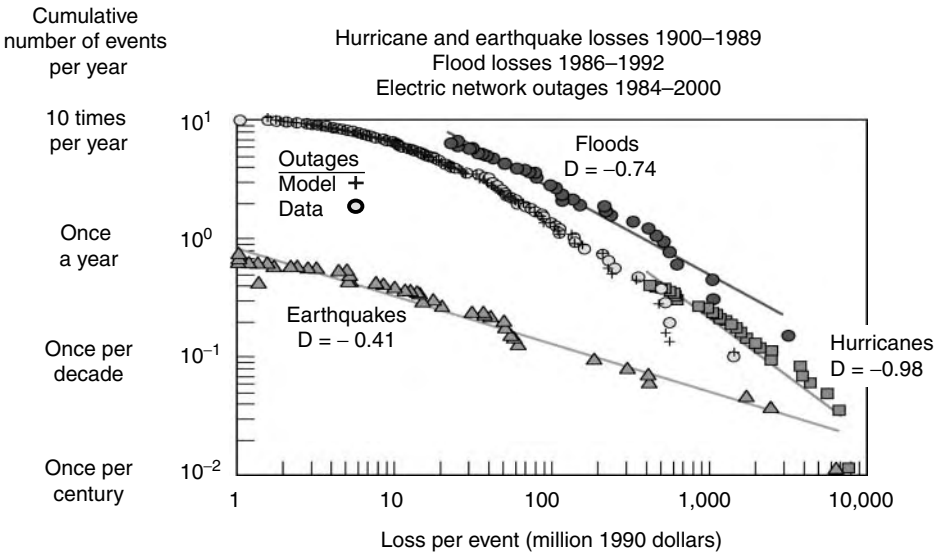

FIGURE 9.23 Understanding complex systems and global dynamics. Economic losses from disasters were found to follow a power law distribution—for hurricanes, floods, earthquakes, and even electrical outages. Fundamental power law distributions also were found for forest fires, Internet congestion, and other systems. CIN/SI results such as these translate in new approaches for optimizing complex systems in terms of productivity and robustness to disaster. Our goal is to move the power outage curve down toward the origin, i.e., to make outages less frequent and with smaller impact on customers. (From The EPRI/DoD Complex Interactive Networks/Systems Initiative [CIN/SI].)

In what follows, a brief overview of some key areas is given and selected security aspects of operational systems are presented, without discussing potentially sensitive material; these aspects include:

- Operational systems rely very heavily on the exchange of information amongst disparate systems
- Utilities rely on very extensive private and leased telecommunication systems
- Networking of these systems is expanding rapidly
- This networking is expanding beyond utility doors, to encompass other utilities, corporations, and customers
- Standard communication protocols and integration techniques are a MUST, despite the increased security risks
- Increased security concerns in the aftermath of tragic events of September 11, 2001
- Deregulation is increasing the incentives for unauthorized access to information

9.2.4 Infrastructures Under Threat

The terrorist attacks of September 11 have exposed critical vulnerabilities in America's essential infrastructures: Never again can the security of these fundamental systems be taken for granted. Electric power systems constitute the fundamental infrastructure of modern society. A successful terrorist attempt to disrupt electricity supplies could have devastating effects on national security, the economy, and the lives of every citizen. Yet power systems have widely dispersed assets that can never be absolutely defended against a determined attack.

Because critical infrastructures touch us all, the growing potential for infrastructure problems stems from multiple sources. These sources include system complexity, deregulation, economic effects, power market impacts, terrorism, and human error. The existing power system is also vulnerable to natural disasters and intentional attacks. Ongoing efforts at NERC, DOE, and EPRI—including EPRI's

Enterprise Information Security Program and Infrastructure Security Initiative—have highlighted utility-specific threats and the technologies to counteract them. Regarding the latter, a November 2001 EPRI assessment developed in response to the September 11, 2001 attacks highlights three different kinds of potential threats to the U.S. electricity infrastructure (Amin, 2001, 2002a, 2003, EPRI 2001):

- Attacks upon the power system. In this case, the electricity infrastructure itself is the primary target, with ripple effects, in terms of outages, extending into the customer base. The point of attack could be a single component, such as a critical substation, or a transmission tower. However, there could also be a simultaneous, multipronged attack intended to bring down the entire grid in a region of the United States. Similarly, the attack could target electricity markets, which, because of their transitional status, are highly vulnerable.
- Attacks by the power system. In this case, the ultimate target is the population, using parts of the electricity infrastructure as a weapon. Power plant cooling towers, for example, could be used to disperse chemical or biological agents.
- Attacks through the power system. In this case, the target is the civil infrastructure. Utility networks include multiple conduits for attack, including lines, pipes, underground cables, tunnels and sewers. An electromagnetic pulse, for example, could be coupled through the grid to with the intention of damaging computer and/or telecommunications infrastructure.

Protection against all three modes of attack thus represents a potentially critical "showstopper" for realizing a digital economy and enabling customer-managed service networks. Indeed, if infrastructure security is not ensured, even maintaining current levels of productivity and service will be jeopardized. Conversely, deploying some of the advanced technologies needed to enhance security will have a positive effect on efforts to improve grid reliability and coordinate power system operations with those of other infrastructures.

Therefore, the imperative for enhancing security in the electric power system has reached a new level that demands industry attention. To address this imperative in a logical and deliberate way, understanding is required of what is involved and how to measure the current and future levels of secure performance.

The technologies that support the operational control of electrical networks range from energy management systems (EMSs) to remote field devices. Critical systems include:

- Energy management system (EMS): The objective of the EMS is to manage the production, purchasing, transmission, distribution, and sale of electrical energy in the power system at a minimal cost with respect to safety and reliability. Management of the real-time operation of an electric power system is a complex task requiring interaction of human operators, computer systems, communications networks, and real-time data-gathering devices in power plants and substations. An EMS consists of computers, display devices, software, communication channels, and remote terminal units (RTUs) that are connected to other remote terminal units, control actuators, and transducers in power plants and substations. The main tasks that an EMS performs have to do with generator control and scheduling, network analysis, and operator training. Control of generation requires that the EMS maintain system frequency and tie-line flows while economically dispatching each generating unit. Management of the transmission network requires that the EMS monitor up to thousands of telemetered values, estimate the electrical state of the network, and inform the operator of the best strategy to handle potential outages that could result in an overload or voltage limit violation. EMSs can have real-time two-way communication links between substations, power plants, independent system operators, and other utility EMSs.
- Supervisory control and data acquisition (SCADA) system: A SCADA system supports operator control of remote (or local) equipment, such as opening or closing a breaker. A SCADA system provides three critical functions in the operation of an electric power system: data acquisition,

supervisory control, and alarm display. It consists of one or more computers with appropriate applications software connected by a communications system to a number of RTUs placed at various locations to collect data, perform intelligent control of electrical system devices, and report results back to an EMS. SCADAs can also be used for similar applications in natural gas pipeline transmission and distribution applications. A SCADA can have real-time communication links with one or more EMSs and hundreds of substations.

- Remote terminal units (RTUs): RTUs are special-purpose microprocessor-based computers that contain analog-to-digital converters (ADCs) and digital-to-analog converters (DACs), digital inputs for status, and digital output for control. There are transmission substation RTUs and distribution automation (DA) RTUs. Transmission substation RTUs are deployed at substation and generation facilities, where a large number of status and control points are required. DA RTUs are used to:
- Control air switches and Var-compensation capacitor banks on utility poles
- Control pad-mounted switches
- Monitor and automate feeders
- Monitor and control underground networks
- Monitor, control, and automate smaller distribution substations

RTUs are also used as indicated above in natural gas transmission and distribution. RTUs can be configured and interrogated using telecommunication technologies. They can have hundreds of real-time communication links with other substations, EMSs, and power plants.

- Programmable logic controllers (PLCs): PLCs have been used extensively in manufacturing and process industries for many years and are now being used to implement relay and control systems in substations. PLCs have extended I/O systems similar to transmission substation RTUs. The control outputs can be controlled by software residing in the PLC and via remote commands from a SCADA system. The PLC user can make changes in the software stored in EEPROM without making any major hardware or software changes. In some applications, PLCs with RTU-reporting capability may have advantages over conventional RTUs. PLCs are also used in many power plant and refinery applications. They were originally designed for use in discrete applications like coal handling. They are now being used in continuous control applications such as feedwater control. PLCs can have many real-time communication links inside and outside substations or plants.
- Protective relays: Protective relays are designed to respond to system faults and short circuits. When faults occur, the relays must signal the appropriate circuit breakers to trip and isolate the faulted equipment. Distribution system relaying must be coordinated with fuses and reclosures for faults while ignoring cold load pickup, capacitor bank switching, and transformer energization. Transmission line relaying must locate and isolate a fault with sufficient speed to preserve stability, reduce fault damage, and minimize the impact on the power system. Certain types of "smart" protective relays can be configured and interrogated using telecommunication technologies.
- Automated metering: Automated metering is designed to upload residential and/or commercial gas and/or electric meter data. These data can then be automatically downloaded to a PC or other device and transmitted to a central collection point. With this technology, real-time communication links exist outside the utility infrastructure.
- Plant-distributed control systems (DCSs): DCSs are plant-wide control systems that can be used for control and/or data acquisition. The I/O count can be higher than 20,000 data points. Often, the DCS is used as the plant data highway for communication to and from intelligent field devices, other control systems (such as PLCs), RTUs, and even the corporate data network for enterprise resource planning (ERP) applications. The DCS traditionally has used a proprietary operating system. Newer versions are moving toward open systems such as Windows NT,

Sun Solaris, and so on. DCS technology has been developed with operating efficiency and user configurability as drivers, rather than system security. Additionally, technologies have been developed that allow remote access, usually via PC, to view and potentially reconfigure the operating parameters.

- Field devices: Examples of field devices are process instrumentation such as pressure and temperature sensors and chemical analyzers. Other standard types of field devices include electric actuators. Intelligent field devices include electronics to enable field configuration, upload of calibration data, and so on. These devices can be configured off-line. They also can have real-time communication links between plant control systems, maintenance management systems, stand-alone PCs, and other devices inside and outside the facility.

Security of these cyber and communication networks is fundamental to the reliable operation of the grid. As power systems rely more heavily on computerized communications and control, system security has become increasingly dependent on protecting the integrity of the associated information systems. Part of the problem is that existing control systems, which were originally designed for use with proprietary, standalone communication networks, were later connected to the Internet (because of its productivity advantages and lower costs), but without adding the technology needed to make them secure. Communication of critical business information and controlled sharing of that information are essential parts of all business operations and processes.

As the deregulation of the energy industry unfolds, information security will become more important. For the energy-related industries, the need to balance the apparently mutually exclusive goals of operating system flexibility with the need for security will need to be addressed from a business perspective. Key electric energy operational systems depend on real-time communication links (both internal and external to the enterprise). The functional diversity of these organizations has resulted in a need for these key systems to be designed with a focus on open systems that are user configurable to enable integration with other systems (both internal and external to the enterprise). In many cases, these systems can be reconfigured using telecommunication technologies. In nearly all cases, the systems dynamically exchange data in real time. This results in a need for highly reliable, secure control and information management systems.

Power plant DCS systems produce information necessary for dispatch and control. This requires real-time information flow between the power plant and the utility's control center, system dispatch center, regulatory authorities, and so on. A power plant operating as part of a large wholesale power network may have links to an independent system operator, a power pool, and so on. As the generation business moves more and more into market-driven competitive operation, both data integrity and confidentiality will become major concerns for the operating organizations.

Any telecommunication link that is even partially outside the control of the organization that owns and operates power plants, SCADA systems, or EMSs represents a potentially insecure pathway into the business operations of the company as well as a threat to the grid itself. The interdependency analyses done by most companies during Y2K preparations have identified these links and the system's vulnerability to their failures. Thus they provide an excellent reference point for a cyber-vulnerability analysis.

In particular, monitoring and control of the overall grid system is a major challenge. Existing communication and information system architectures lack coordination among various operational components, which usually is the cause for the unchecked development of problems and delayed system restoration. Like any complex dynamic infrastructure system, the electricity grid has many layers and is vulnerable to many different types of disturbances. While strong centralized control is essential to reliable operations, this requires multiple, high-data-rate, two-way communication links, a powerful central computing facility, and an elaborate operation control center, all of which are especially vulnerable when they are needed most—during serious system stresses or power disruptions. For deeper protection, intelligent distributed control is also required, which would enable parts of the network to remain operational and even automatically reconfigure in the event of local failures or threats of failure.

9.2.5 The Dilemma: Security and Quality Needs

The specter of terrorism raises a profound dilemma for the electric power industry: How to make the electricity infrastructure more secure without compromising the productivity advantages inherent in today's complex, highly interconnected electric networks? Resolving this dilemma will require both short-term and long-term technology development and deployment, affecting some of the fundamental characteristics of today's power systems:

- Centralization/decentralization of control. For several years, there has been a trend toward centralizing control of electric power systems. Emergence of regional transmission organizations (RTOs) as agents of wide-area control, for example, offers the promise of greatly increased efficiency and improved customer service. But if terrorists can exploit the weaknesses of centralized control, security would seem to demand that smaller, local systems become the system configuration of choice. In fact, strength and resilience in the face of attack will increasingly rely upon the ability to bridge simultaneous top-down and bottom-up decision making in real time.
- Increasing complexity. The North American electric power system has been called the "most complex machine ever built." System integration helps move power more efficiently over long distances and provides redundancy to ensure reliable service, but it also makes the system more complex and harder to operate. In response, new mathematical approaches are needed to simplify the operation of complex power systems and to make them more robust in the face of natural or manmade interruptions.
- Dependence on Internet communications. Today's power systems could not operate without tightly knit communications capability—ranging from high-speed data transfer among control centers to interpretation of intermittent signals from remote sensors. Because of the vulnerability of Internet communications, however, protection of the electricity supply system requires new technology to enhance the security of power system command, control, and communications, including both hardware and software.
- Accessibility and vulnerability. Because power systems are so widely dispersed and relatively accessible, they are particularly vulnerable to attack. Although "hardening" of some key components, such as power plants and critical substations, is certainly desirable, it is simply not feasible or economic to provide comprehensive physical protection to all components. Probabilistic assessments can offer strategic guidance on where and how to deploy security resources to greatest advantage.

A survey of electric utilities revealed real concerns about grid and communications security. Figure 9.24 ranks the perceived threats to utility control centers. The most likely threats were bypassing controls, integrity violations, and authorization violations. Concern about the potential threats generally increased as the size of the utility (peak load) increased.

The system's equipment and facilities are dispersed throughout the North American continent which complicates protection of the system from a determined terrorist attack. In addition, another complexity—the power delivery systems' physical vulnerabilities and susceptibility to disruptions in computer networks and communication systems—must also be considered. For example, terrorists might exploit the increasingly centralized control of the power delivery system to magnify the effects of a localized attack. Because many consumers have become more dependent on electronic systems that are sensitive to power disturbances, an attack that leads to even a momentary interruption of power can be costly. A 20-minute outage at an integrated circuit fabrication plant, for example, could cost US$30 million.

Despite increasing concerns in these areas as well as the continuing erosion of reserve margins, the infrastructures of today do generally function well. The electricity is on, the phones work, traffic flows nearly all of the time. But more and more, the traditional level of performance is no longer good enough;

more robust infrastructures are needed for the "digital society" envisioned for tomorrow. For example in the electric power area, there is a need for an increase in reliability from today's average of about 99.9% (approximately 8 hours of outage per year) to 99.9999% (about 32 seconds outage per year) or even 99.9999999% (one outage lasting less than a single AC cycle per year). Such near-perfect power is needed today for error-free operation of the microprocessor chips finding their way into just about everything, including billions of embedded applications.

Fortunately, the core technologies needed to strategically enhance system security are the same as those needed to resolve other areas of system vulnerability, as identified in the *Electricity Technology Roadmap*. These result from open access, exponential growth in power transactions, and the reliability needed to serve a digital society.

The North American electric power system needs a comprehensive strategy to prepare for the diverse threats posed by terrorism. Such a strategy should both increase protection of vital industry assets and ensure the public that they are well protected. A number of actions will need to be considered in formulating an overall security strategy:

- The grid must be made secure from cascading damage.
- Pathways for environmental attack must be sealed off.
- Conduits for attack must be monitored, sealed off and "sectionalized" under attack conditions.
- Critical controls and communications must be made secure from penetration by hackers and terrorists.
- Greater intelligence must be built into the grid to provide flexibility and adaptability under attack conditions, including automatic reconfiguration.
- Ongoing security assessments, including the use of game theory to develop potential attack scenarios, will be needed to ensure that the power industry can stay ahead of changing vulnerabilities.

The dispersed nature of the power delivery system's equipment and facilities complicates the protection of the system from a determined attack. Furthermore, both physical vulnerabilities and susceptibility of power delivery systems to disruptions in computer networks and communication systems must be considered. For example, terrorists might exploit the increasingly centralized control of the power delivery system to magnify the effects of a localized attack. Because many consumers have become more dependent on electronic systems that are sensitive to power disturbances, an attack that leads to even a momentary interruption of power can be costly.

9.2.5.1 Human Performance

Because humans interact with these infrastructures as managers, operators, and users, human performance plays an important role in their efficiency and security. In many complex networks, the human participants themselves are both the most susceptible to failure and the most adaptable in the management of recovery. Modeling and simulating these networks, especially their economic and financial aspects, will require modeling the bounded rationality of actual human thinking, unlike that of a hypothetical "expert" human as in most applications of artificial intelligence (AI). Even more directly, most of these networks require some human intervention for their routine control and especially when they are exhibiting anomalous behavior that may suggest actual or incipient failure.

Operators and maintenance personnel are obviously "inside" these networks and can have direct, real-time effects on them. But the users of a telecommunication, transportation, electric power, or pipeline system also affect the behavior of those systems, often without conscious intent. The amounts, and often the nature, of the demands put on the network can be the immediate cause of conflict, diminished performance and even collapse. Reflected harmonics from one user's machinery degrade power quality for all. Long transmissions from a few users create Internet congestion. Simultaneous lawn watering drops the water pressure for everyone. In a very real sense, no one is "outside" the infrastructure.

Given that there is some automatic way to detect actual or immanent local failures, the obvious next step is to warn the operators. Unfortunately, the operators are usually busy with other tasks, sometimes even responding to previous warnings. In the worst case, the detected failure sets off a multitude of almost simultaneous alarms as it begins to cascade through the system, and, before the operators can determine the real source of the problem, the whole network has shut itself down automatically.

Unfortunately, humans have cognitive limitations that can cause them to make serious mistakes when they are interrupted. In recent years, a number of systems have been designed that allow users to delegate tasks to intelligent software assistants ("softbots") that operate in the background, handling routine tasks and informing the operators in accordance with some protocol that establishes the level of their delegated authority to act independently. In this arrangement, the operator becomes a supervisor, who must either cede almost all authority to subordinates or be subject to interruption by them. At present, this is a very limited understanding of how to design user interfaces to accommodate interruption.

9.2.5.2 Broader Technical Issues

In response to the above challenges, several enabling technologies and advances are/will be available that can provide necessary capabilities when combined in an overall system design. Among them are the following:

- Flexible AC transmission system (FACTS) devices, which are high-voltage thyristor-based electronic controllers that increase the power capacity of transmission lines and have already been deployed in several high-value applications. At peak demand, up to 50% more power can be controlled through existing lines.
- Fault current limiters (FCLs), which absorb the shock of short circuits for a few cycles to provide adequate time for a breaker to trip. It is noteworthy that preliminary results of the August 14 outage show that FCLs could have served as large electrical "shock absorbers" to limit the size of blackouts.
- Wide-area measurement systems (WAMS), which integrate advanced sensors with satellite communication and time stamping using global positioning systems (GPS) to detect and report angle swings and other transmission system changes.
- Innovations in materials science and processing, including high-temperature superconducting (HTS) cables, oxide-power-in-tube technology for HTS wire, and advanced silicon devices and wide bandgap semiconductors for power electronics.
- Distributed resources such as small combustion turbines, solid oxide and other fuel cells, photovoltaics, superconducting magnetic energy storage (SMES), transportable battery energy storage systems (TBESS), etc.
- Information systems and online data processing tools such as the open-access same-time information system (OASIS), and transfer capability evaluation (TRACE) software, which determines the total transfer capability for each transmission path posted on the OASIS network, while taking into account the thermal, voltage, and interface limits.
- Monitoring and use of IT: Wide-area measurement/management systems (WAMS), open-access same-time information system (OASIS), supervisory control and data acquisisiton (SCADA) systems, energy management systems (EMS).
- Analysis tools: Several software systems for dynamic security assessment of large/wide-area networks augmented with market/risk assessment.
- Intelligent electronic devices with security provisions bulit in- combining sensors, computers, telecommunication units, and actuators; integrated sensor; two-way communication; "intelligent agent" functions: assessment, decision, learning; actuation, enabled by advances in several areas including semiconductors and resource-constrained encryption.

However, if most of the above technologies are developed, still the overall systems' control will remain a major challenge. This is a rich area for research and development of such tools, as well as to address systems and infrastructure integration issues of their deployment in the overall network, especially now because of

increased competition, the demand for advanced technology to gain an advantage, and the challenge of providing the reliability and quality consumers demand.

9.2.5.3 Western States' Power Crises: A Brief Overview of Lessons Learned

An example of "urgent" opportunities is within the now seemingly calm California energy markets; the undercurrents that led to huge price spikes and considerable customer pain in recent years are yet to be fully addressed and alleviated. Such "perfect storms" may appear once again during another cycle of California economic recovery and growth. The California power crisis in 2000 was only the most visible parts of a larger and growing U.S. energy crisis that is the result of years of inadequate investments in the infrastructure.

For example, at the root of the California crisis was declining investment in infrastructure components that led to a fundamental imbalance between growing demand for power and an almost stagnant supply. The imbalance had been brewing for many years and is prevalent throughout the nation (see EPRI's Western States Power Crises White Paper; http://www.epri.com/WesternStatesPowerCrisisSynthesis.pdf).

California is a good downside example of a societal testbed for the ways that seemingly "good" theories can fail in the real world. For example, inefficient markets provide inadequate incentives for infrastructure investment:

- Boom–bust cycle may be taking shape in generation investment
- Transmission investment running at one-half of 1975 levels
- Congestion in transmission network is rising, as indicated by increase of number of transmission loading relief (TLR) during the last three years.

The cost of market failure can be also very high; as indicated by the exercise of market power in California during summer of 2000 which cost consumers $4 billion initially, while the ongoing intermediate loss to businesses may well be considerably higher.

More specifically regarding the electricity under investments and persisting undercurrents, very specific "investments" by the state were made, on the order of $10 billion, paid to subsidize (hold down) electricity prices, and to bail out bankrupt companies through long-term noncompetitive contracts that did not address the undercurrents and shortcomings of the earlier policies.

To address these issues there are both tactical as well as strategic needs; for example, the so-called "low-hanging fruits" to improve transmission networks include:

- Deploy existing technologies to improve use of already in place transmission assets (e.g., FACTS, dynamic thermal circuit rating, and energy storage–peak shaving technologies). For example, through the integration of load management technologies shaving nearly 5,000 MW which amounts to about 10% of total demand, combined with a more precise control enabled by the use of FACTS devices, which enable nearly 50% more transfer capability over existing transmission lines.
- Develop and deploy new technologies to improve transmission reliability and throughput (e.g., low sag composite conductors, high-temperature superconducting cables, extra high-voltage AC and DC transmission systems, hierarchical control systems)
- Improve real-time control of network via monitoring and data analysis of dynamic transmission conditions
- Develop and deploy self-healing grid tools to adaptively respond to overload and emergency conditions
- Digital control of the power delivery network (reliability, security, and power quality).
- Integrated electricity and communications for the user
- Transformation of the meter into a two-way energy/information portal
- Integration of distributed energy resource into the network

- The complex grid can operate successfully if technology is deployed and operated in an integrated manner (there is no "silver bullet")

In addition, longer-term strategic considerations must be addressed; they include:

- Greater fuel diversity including renewable energy technologies—regional and national priorities
- Risk assessment of long-term U.S. reliance—analysis of the value of risk management through fuel diversity
- Introduce time-varying prices and competitive market dynamics for all customers
- Create a planning process and in silico testing of designs, devices, and power markets
- Model market efficiencies, environmental constraints, and renewables
- Develop advanced EM threat detection, shielding, and surge suppression capabilities
- Develop the tools and procedures to ensure a robust and secure marketplace for electricity
- Develop the portfolio of advanced power generation technologies to assure energy security
- Transmission network expansion and RTOs, e.g., would an RTO, compliment a competitive wholesale power market and result in a sustainable and robust system? How large should they be?
- Comprehensive architecture for power supply and delivery infrastructure that anticipates rapidly escalating demands of digital society
- Enable self-healing power delivery infrastructure
- Significant investment in R&D, transmission, generation, and conservation resources are needed
- Incentives for technology innovation and accountability for R&D
- Revitalize the national public/private electricity infrastructure partnership needed to fund the "self-healing grid" deployment
- The "law of unintended consequences" should be considered in crafting any solution

Having discussed the above technology-intensive "push," the fact that adoption of new technologies often creates equally new markets must also be considered. For example, wireless communication creates the market of spectrum, and broadband technologies create the market of bandwidth. Reduced regulation of major industries has required new markets wherever the infrastructure is congested: airlines compete for landing rights, power generators for transmission rights, oil, and gas producers for pipeline capacity.

From a national perspective, a key grand challenge is how to redesign, retrofit, and upgrade the nearly 200,000 miles of electromechanically controlled system into a smart self-healing grid that is driven by a well-designed market approach?

In addressing this challenge, as technology progresses, and the economy becomes increasingly dependent on markets, infrastructures such as electric power, oil/gas/water pipelines, telecommunications, financial, and transportation networks becomes increasingly critical and complex. In particular, since it began in 1882, electric power has grown to become a major industry essential to a modern economy

Over the past two decades, governments around the globe have introduced increasing amounts of competition into network industries. With the advent of restructuring in the electric power industry, the onset of a historical transformation of the energy infrastructure in the context of global trends is underway:

- Increasing electricity demand as a consequence of economic and population growth
- Technological innovations in power generation, delivery, control and communications
- Increasing public acceptance of market mechanisms
- Growing public concerns about environmental quality and depletion of exhaustible resources

Services previously supplied by vertically integrated, regulated monopolies are now provided by multiple firms. The transition to competition has fundamentally altered important aspects of the engineering and economics of production. This presents unique opportunities and challenges. Clearly, this change will have far-reaching implications for the future development of the electricity industry. More fundamentally, as we

look beyond the horizon, this change will further power the information revolution and increasing global interdependence. The long-term socioeconomic impacts of such a transformation will be huge, and the tasks are just as daunting, going well beyond the boundary of existing knowledge.

To meet such a challenge, collaborative research between engineers and economists is critical to provide a holistic and robust basis that will support the design and management of complex technological and economic systems in the long term. The electric power industry offers an immediate opportunity for launching such research, as new ways are being sought to improve the efficiency of electricity markets while maintaining the reliability of the network. Complexity of the electric power grid combined with ever more intricate interactions with markets offers a plethora of new and exciting research opportunities.

In what follows we provide our vision and approach to enabling a smart self-healing electric power system that can respond to a broad array of destabilizers.

9.2.5.4 How to Make an Electric Power System Smart?

To add intelligence to an electric power transmission system, independent processors in each component and each substation and power plant are needed. These processors must have a robust operating system and be able to act as independent agents that can communicate with and cooperate with other forming a large distributed computing platform. Each agent must be connected to sensors associated with its own component or its own substation so that it can assess its own operating conditions and report them to its neighboring agents via the communications paths. Thus for example, a processor associated with a circuit breaker would have the ability to communicate with sensors built into the breaker and communicate those sensor values using high-bandwidth fiber communications connected to other such processor agents.

9.2.5.5 Complex System Failure

Beyond the human dimension, there is a strategic need to understand the societal consequences of infrastructure failure risks along with benefits of various tiers of increased reliability. From an infrastructure interdependency perspective, power, telecommunications, banking and finance, transportation and distribution, and other infrastructures are becoming more and more congested, and are increasingly vulnerable to failures cascading through and between them. A key concern is the avoidance of widespread network failure due to cascading and interactive effects. Moreover, interdependence is only one of several characteristics that challenge the control and reliable operation of these networks. Other factors that place increased stress on the power grid include dependencies on adjacent power grids (increasing because of deregulation), telecommunications, markets, and computer networks. Furthermore, reliable electric service is critically dependent on the whole grid's ability to respond to changed conditions instantaneously.

More specifically, secure and reliable operation of critical infrastructures poses significant theoretical and practical challenges in analysis, modeling, simulation, prediction, control, and optimization. To address these challenges, a research initiative—the EPRI/DOD Complex Interactive Networks/Systems Initiative (CIN/SI)—was undertaken during 1998–2001 to enable critical infrastructures to adapt to a broad array of potential disturbances, including terrorist attacks, natural disasters, and equipment failures.

The CIN/SI overcame the longstanding problems of complexity, analysis, and management for large interconnected systems—and systems of systems—by opening up new concepts and techniques. Dynamical systems, statistical physics, information and communication science, and computational complexity were extended to provide practical tools for measuring and modeling the power grid, cell phone networks, Internet, and other complex systems. For the first time, global dynamics for such systems can be understood fundamentally (Figure 9.25).

As an example, related to numerous major outages, narrowly programmed protection devices have contributed to worsening the severity and impact of the outage, typically performing a simple on/off logic which locally acts as preprogrammed while destabilizing a larger regional interconnection.

From a broader perspective, any critical national infrastructure typically has many layers and decision-making units and is vulnerable to various types of disturbances. Effective, intelligent, distributed control is required that would enable parts of the constituent networks to remain operational and even automatically reconfigure in the event of local failures or threats of failure. In any situation subject to rapid changes,

completely centralized control requires multiple, high-data-rate, two-way communication links, a powerful central computing facility, and an elaborate operations control center. But all of these are liable to disruption at the very time when they are most needed (i.e., when the system is stressed by natural disasters, purposeful attack, or unusually high demand).

When failures occur at various locations in such a network, the whole system breaks into isolated "islands," each of which must then fend for itself. With the intelligence distributed, and the components acting as independent agents, those in each island have the ability to reorganize themselves and make efficient use of whatever local resources remain to them in ways consonant with the established global goals to minimize adverse impact on the overall network. Local controllers will guide the isolated areas to operate independently while preparing them to rejoin the network, without creating unacceptable local conditions either during or after the transition. A network of local controllers can act as a parallel, distributed computer, communicating via microwaves, optical cables, or the power lines themselves, and intelligently limiting their messages to only that information necessary to achieve global optimization and facilitate recovery after failure.

9.2.6 Conclusions: Toward a Secure and Efficient Infrastructure

How to control a heterogeneous, widely dispersed, yet globally interconnected system is a serious technological problem in any case. It is even more complex and difficult to control it for optimal efficiency and maximum benefit to the ultimate consumers while still allowing all its business components to compete fairly and freely. A similar need exists for other infrastructures, where future advanced systems are predicated on the near-perfect functioning of today's electricity, communications, transportation, and financial services.

From a strategic R&D viewpoint, agility and robustness/survivability of large-scale dynamic networks that face new and unanticipated operating conditions will be presented. A major challenge is posed by the lack of a unified mathematical framework with robust tools for modeling, simulation, control and optimization of time-critical operations in complex multicomponent and multiscaled networks.

Given the state of art in electricity infrastructure security and control as indicated in this chapter, creating a smart grid with self-healing capabilities is no longer a distant dream; we have made considerable progress. But considerable technical challenges as well as several economic and policy issues remain to be addressed

Acknowledgments

Most of the material and findings presented here were developed while the author was at the Electric Power Research Institute (EPRI) in Palo Alto, California. The author is grateful for EPRI's support and feedback from numerous colleagues at EPRI, universities, industry, and government agencies.

References

Amin, M. 2000. Toward self-healing infrastructure systems, *IEEE Computer Magazine*, 33 (8), 44–53.

Amin, M. 2001. Toward self-healing energy infrastructure systems, *IEEE Computer Applications in Power*, 14 (1), 20–28.

Amin, M. 2002a. Security challenges for the electricity infrastructure, *IEEE Computer Magazine*, April 8–10.

Amin, M. 2002b. *Special Issue* of *IEEE Control Systems Magazine on Control of Complex Networks*, December, Vol. 21, No. 6.

Amin, M. 2002c. *Special Issue* of *IEEE Control Systems Magazine on Control of Complex Networks*, February, Vol. 22, No. 1.

Amin, M. 2003. North America's electricity infrastructure: Are we ready for more perfect storms? *IEEE Security and Privacy Magazine*, 1 (5), 19–25.

Amin, M., *Special Issue of the Proceedings of the IEEE on Energy Infrastructure Defense Systems*. in press.

Amin, M., Gerhart, V., and Rodin, E. Y. 1997. System identification via artificial neural networks: Application to on-line aircraft parameter estimation, *Proceedings of AIAA/SAE 1997 World Aviation Congress*, p. 22, Anaheim, CA.

Ang, A. H-S. and Tang, W. H. 1990. *Decision, Risk, and Reliability, Volume 2 of Probability Concepts in Engineering Planning and Design*, Wiley, New York.

Ayyub, B. M. 2002. *Elicitation of Expert Opinions for Uncertainty and Risks*, CRC Press, Boca Raton, FL.

Ayyub, B. M. 2003. *Risk Analysis in Engineering and Economics*, Chapman and Hall/CRC Press, Boca Raton, FL.

Ayyub, B. M. and Klir, G. J. 2006. *Uncertainty Modeling and Analysis for Engineers and Scientists*, Chapman and Hall/CRC Press, Boca Raton, FL.

Ayyub, B. M. and McCuen, R. 2003. *Probability, Statistics and Reliability for Engineers and Scientists*, *2nd Ed.*, Chapman and Hall/CRC Press, Boca Raton, FL.

Dy Liacco, T. E. 1967. The adaptive reliability control system, *IEEE Transactions on Power Apparatus and Systems*, May, 517–561.

Energy Information Administration (EIA). 2006. *Annual Energy Outlook 2006 with projections to 2030*, U.S. Department of Energy, Washington, DC, February 2006, http://www.eia.doe.gov/oiaf/aeo/figure_3.html

EPRI. 1999. Electricity technology roadmap: 1999 summary and synthesis report, Palo Alto, CA, July 1999, http://www.epri.com/corporate/discover_epri/roadmap/CI-112677-V1_all.pdf

EPRI. 2000. Communication security assessment for the United States electric utility infrastructure, EPRI Report 1001174, EPRI, Palo Alto, CA.

EPRI. 2001. *Electricity Infrastructure Security Assessment*, *Vol. I–II*, EPRI, Palo Alto, CA.

EPRI. 2003. Complex interactive networks/systems initiative: Final summary report—Overview and summary final report for joint EPRI and US Department of Defense University Research Initiative, EPRI, Palo Alto, CA.

EPRI. 2003. *Electricity Technology Roadmap Report*, EPRI, Palo Alto, CA, July 2003, http://www.epri.com/corporate/discover_epri/roadmap/index.html

Edison Electric Institute (EEI), 2003. Critical infrastructure protection, http://www.eei.org/industry_issues/energy_infrastructure/critical_infrastructure_protection/#1

EPRI. 2004. *Supervisory Control and Data Acquisition (SCADA) Systems Security Guide*, EPRI, Rep. 1002604, Palo Alto, CA.

EPRI. 2005. Guideline for securing control system & corporate network interfaces, EPRI, Rep. 1010714, Palo Alto, CA.

Fink, L. H. and Carlsen, K. 1978. Operating under stress and strain, *IEEE Spectrum,* March, 48–53.

Gellings, C. W. and Yeager, K. E. 2004. Transforming the electric infrastructure. *Physics Today,* December, Vol. 57, 45–51.

Hauer, F. F. and Dagle, J. E. 1999. *Review of Recent Reliability Issues and System Events*, Consortium for Electric Reliability Technology Solutions, Transmission Reliability Program, Office of Power Technologies, U.S. Department of Energy (DOE), Washington, DC, August 30.

House Committee on Energy and Commerce. 2003. Blackout 2003: How did it happen and why?, 108th Congress House Hearings, the U.S. Government Printing Office Via GPO Access (DOCID: F: 89467.wais), Washington, DC, September 3, 2003, http://energycommerce.house.gov/108/Hearings/09032003 hearing 1061/print.htm

Kumamoto, H. and Henley, E. J. 1996. *Probabilistic Risk Assessment and Management for Engineers and Scientists*, *2nd Ed.*, IEEE Press, New York.

Kundur, P. 1994. *Power System stability and control, EPRI Power System Engineering Series*, McGraw-Hill, New York.

Maryland Emergency Management Agency (MEMA). 2006. State of Maryland guide for the protection of critical infrastructure and key resources for homeland security, Volume 1: Critical asset & portfolio risk assessment (CAPRA) methodology, Office of Homeland Security, Annapolis, MD.

Modarres, M. 1993. *What every engineer should know about Reliability and Analysis*, Marcel Dekker, New York.

Modarres, M., Kaminskiy, M., and Krivstov, V. 1999. *Reliability Engineering and Risk Analysis: A Practical Guide*, Marcel Dekker, New York.

National Science Foundation, Division of Science Resources Statistics. 2003. Research and development in industry: 2000. NSF 03-318, NSF, Arlington, VA, http://www.nsf.gov/sbe/srs/nsf03318/pdf/taba19.pdf

North American Electric Reliability Council (NERC), 2006. Disturbance Analysis Working Group (DAWG) database (1984–2002), Princeton, NJ, Available: http://www.nerc.com/~dawg

North American Electric Reliability Council (NERC), 2001. *Reliability Assessment 2001-2010*, Princeton, NJ, Available: http://www.nerc.com/~filez/rasreports.html

North American Electric Reliability Council (NERC), 2004. *NERC security guidelines for the electricity sector*, Princeton, NJ, Available: http://www.esisac.com/library-guidelines.htm

North American Electric Reliability Council (NERC), 2006. *Top 10 Vulnerabilities of control system and Their Associated mitigations-2006*, Control Systems Security Working Group, U.S. Department of Energy, National SCADA Test Bed Program, Princeton, NJ, March 16.

President's Commission on Critical Infrastructure Protection. 1997. Critical foundations: Protecting America's infrastructures, http://www.ciao.ncr.gov

Samotyj, M., Gellings, C., and Amin, M. 2003. Power system infrastructure for a digital society: creating the new frontiers, *Proceedings of the GIGRE/IEEE-PES Symposium on Quality and Security of Electric Power Delivery*, p. 10, Montreal, October 7–10.

Silberman, S. 2001. The energy web, *Wired*, 9 (7), 116.

Starr, C., and Amin, M. 2003. Global transition dynamics unfolding the full social implications of national decision pathways, http://cdtlnet.cdtl.umn.edu/Amin/GlobalTransition.pdf

US Department of Energy. 2002. National transmission grid study, http://tis.eh.doe.gov/ntgs/gridstudy/main_screen.pdf

US Department of Homeland Security (DHS), 2006. The National Infrastructure Protection Plan (NIPP),Washington, DC, June, Available: http://www.dhs.gov/nipp

US DOE and the US DHS, 2006. Roadmap to Secure Control Systems in the Energy Sector, Office of Energy Delivery and Energy Reliability, 58 pages, Washington, DC, January.