Cybernetics and Systems Analysis, Vol. 30, No. 2, 1994

MATHEMATICAL MODELS AND METHODS OF RISK ASSESSMENT IN ECOLOGICALLY HAZARDOUS INDUSTRIES*

V. S. Mikhalevich, P. S. Knopov, and A. N. Golodnikov

UDC 502.55:621.039.7

INTRODUCTION

Analysis of critical industrial situations leading to accidents or catastrophes has shown that the main factors responsible for accidents include technological inadequacy of ecologically hazardous facilities, equipment design errors, and insufficient preventive maintenance of facilities with an enhanced level of environmental hazard. The scale of the accident after-effects essentially depends on the location of the ecologically hazardous facility, timely development of preventive measures, and prompt implementation of these measures in emergency in compliance with strict deadlines for decision making.

Scientific analysis of the admissible location of ecologically hazardous facilities in a given region requires solving the following problems.

1. Expert evaluation of the safety of the proposed ecologically hazardous technological process, equipment, system, etc., in order to determine the probability of occurrence of accidents and to identify the bottlenecks that have the greatest impact on system safety. Since absolute safety does not exist in nature, an acceptable design is a technological process whose safety is not lower than some threshold level established by system analysis of the consequences of accidents in the given ecologically hazardous facility.

2. Development of an optimal maintenance strategy for the technical equipment in the given ecologically hazardous facility with the object of increasing its safety level.

The maintenance problem is relevant because over time the safety of any system declines as a result of aging and wear-and-tear of the component units. Technical maintenance is intended to prevent failure: it includes testing of subsystems and component units and correction of the detected faults. The quality of technical maintenance that has an effect on the safety level of the system depends on both the schedule and the volume of equipment tests. Optimal planning of system maintenance therefore requires development of a testing schedule with a certain volume of tests so as to maximize the safety level of the given ecologically hazardous facility.

3. Prediction of the consequences of possible accidents in a given ecologically hazardous facility and development of protective measures intended to safeguard the population and the environment from the effect of harmful emissions produced by the accident. The solution of this problem requires modeling the entire propagation chain of pollutants from the source to the human being, allowing for diverse meteorological, physicochemical, and radiological factors, as well as the specifics of the local soil and agricultural products produced in the region. Individual dose and collective, or population, dose are used as measures of the effect of noxious emissions on the population.

* The research was carried out in accordance with the program of the State Committee of Science and Technology of Ukraine.

Translated from Kibernetika i Sistemnyi Analiz, No. 2, pp. 121-139, March-April, 1994. Original article submitted January 12, 1994.

1060-0396/94/3002-0259$12.50

1. FORMULATION OF THE PROBLEM AS A TWO-STAGE STOCHASTIC PROGRAMMING MODEL

All three problems listed above are interconnected, jointly forming a general two-stage stochastic-programming problem. Two-stage stochastic programming problems model a two-stage decision-making process under risk and uncertainty. In the first stage, a strategic decision x is made before a random realization of the state of nature ω becomes known (in our case, this realization is the occurrence of an accident, emission and spread of pollutants in the environment, and harmful effect of pollutants on the population in the affected territory). The strategic decision x involves choosing a particular design of an ecologically hazardous facility, its location, and a technical maintenance strategy for the proposed equipment.

When one of the possible states of nature ω is realized and the degree and scale of environmental pollution become known, we reach the second stage of decision making, namely the choice of a set of protective measures $y(\omega)$ that depend on the realized state of nature ω and are intended to alleviate the damage to public health caused by the noxious pollutants. At this stage, the protective measures are chosen on the basis of predetermined criteria, which were developed previously, before the accident. According to these criteria, the i-th protective measure is implemented when the expected equivalent individual dose $H(\omega)$ exceeds some threshold h_i established from biological risk considerations.

The set of protective measures $y(\omega)$ chosen at the second stage is characterized by its effectiveness, i.e., reduction of the expected individual dose, and its cost $Z_1(y(\omega))$. When the measures $y(\omega)$ have been implemented, the expected equivalent dose is $H(\omega) = f(y, H_0(\omega)$, where $H_0(\omega)$ is the expected equivalent dose without the protective measures. The costs $Z_1(y(\omega))$ include the cost of dosimetric monitoring, evacuation of the population, decontamination activities, etc. In addition to material costs of implementing the preventive measures, we also have to allow for the direct damage $Z_2(\omega)$, which includes the reconstruction costs, the cost of the fixed assets abandoned in the evacuated territory, etc. The total damage after the realization of the state of nature ω is thus $Z(\omega) = Z_1(\omega) + Z_2(\omega)$.

To determine the radiation risk, we need to estimate the expected collective dose, which is given by the formula

$$S(\omega) = \sum_i N_i(\omega) H_i(\omega), \tag{1}$$

where $N_i(\omega)$ is the number of people for which the expected equivalent individual dose is $H_i(\omega)$. The radiation risk R and the expected damage are given by the formulas

$$R = \sum_i S(\omega_i) \cdot p_i, \tag{2}$$

$$Z'' = \sum_i Z(\omega_i) \cdot p_i, \tag{3}$$

where ω_i is a specific realization of the state of nature, p_i is the probability of the realization ω_i.

The set of states of nature is determined by the possible values of the emission parameters in the ecologically hazardous facility (duration of emission, emission strength or volume, chemical composition of the emitted pollutants, emission temperature) and parameters describing weather conditions (direction and velocity of wind, type of precipitation, air temperature). The values of the emission parameters are determined by experts who analyze the outcome of accidents in the first stage of decision making. The probability of a specific combination of parameter values characterizing the emission is therefore equal to the probability of occurrence of an accident with the corresponding values of emission parameters. The probability of values of various weather parameters can be established from long-term series of observations in a network of meteorological stations.

Let j be the parameter index (1 is a parameter characterizing the type of the accident, 2 is wind direction, 3 is wind velocity, 4 is type of precipitation, 5 is air temperature). Suppose that parameter i takes n_i possible values, r_j^i is the j-th value, and p_j^i is the probability that parameter i takes the value r_j^i, $j = 1, \ldots, n_i$, $i = 1, \ldots, 5$. Then the set Ω of states of nature is defined by all possible combinations $\omega = (r_{j_1}^1, r_{j_2}^2, r_{j_3}^3, r_{j_4}^4, r_{j_5}^5)$. If we assume that the parameters are mutually independent, then the probability of the state of nature $\omega = (r_{j_1}^1, r_{j_2}^2, r_{j_3}^3, r_{j_4}^4, r_{j_5}^5)$, is given by the formula

$$p(\omega) = \prod_{i=1}^{5} p_{j_i}^{i}. \tag{4}$$

The two-stage stochastic programming problem seeks a decision $x \in X$ which chooses a particular design of the ecologically hazardous facility, its specific location, and technical maintenance parameters so that the radiation risk is below some prespecified threshold value r_0,

$$R(x) \leq r_0, \tag{5}$$

while minimizing the sum of material costs and the expected material damage:

$$Z'(x) + Z''(x) \to \min. \tag{6}$$

Here Z'(x) is the maintenance cost corresponding to the decision x; Z"(x) is the expected damage determined from formula (3); X is the feasible solution set generated by alternative designs of the ecologically hazardous facility, alternative plant locations, and admissible maintenance parameters.

Let us consider in more detail each of the subproblems, focusing specifically on the case of a nuclear power station.

2. ESTIMATING ACCIDENT PROBABILITY

The probability of an accident in a nuclear power station can be estimated by two approaches.

The first approach collects and analyzes statistical data on operating failures in the nuclear power station. Analysis of the statistical material produces an estimate of the probability of various accidents. This approach to accident probability estimation suffers from an obvious shortcoming. Since the nuclear power station is a complex technical system, consisting of many interconnected subsystems and units which have so far functioned for a relatively short time, most of the potential accidents still have not occurred in practice, and those that have occurred constitute an insufficient statistical sample for reliable estimation of accident probabilities.

The second approach simulates the nuclear power station as a complex system of interconnected units, such that the failure of each unit has a certain impact on the occurrence of accidents in the nuclear power station. With this approach, the accident probability of the system is computed from the failure probabilities of the component units. The algorithm that computes the probability of occurrence of accidents is determined by the logical structure of the system. Therefore, in the second approach, the reliability of the estimated accident probability depends on the completeness and adequacy of the modeling of the nuclear power station by a complex system of interconnected units and also on the reliability of the estimated failure probabilities of the component units. The determination of failure probabilities of the different units in the nuclear power station is a much simpler task than statistical estimation of accident probability for the nuclear power station as a whole. In what follows, we adopt the second approach, which is based on the fault-tree method.

The fault tree [1-6] is a graphical model of various combinations of parallel and sequential faults which produce an undesirable event (failure of a subsystem or accident in the nuclear power station). Faults may be caused by technological malfunctions, errors of the service personnel, and also damaging effects of the external environment. The fault tree depicts the logical interconnection between events leading to the occurrence of the undesirable event, which is represented by the top event in the fault tree. The logical interconnection between events is expressed by logical elements, which either allow or prevent logical transmission of faults from lower-lying to upper-lying units. Lower-lying events are the inputs of a logical element, and an upper-lying event is an output from the logical element. In complex systems, we normally use the logical elements AND and OR. The logical element OR is used in the fault tree to depict the fact that the event on the output occurs only if at least one of the input events has occurred; the logical element AND signifies that the output event occurs when all the input events have occurred.

The application of the fault-tree method usually starts with the identification of some undesirable event associated with the system. When the top event has been chosen, the system is analyzed in order to elucidate the causes that produce the top event. This process elucidates other events linked by the logical elements AND and OR. The analysis is carried out sequentially

until we reach events that cannot be decomposed or that may be left intact without a noticeable effect. These events are called primary or elementary. Figure 1 is a graphical representation of the fault tree.

The event Q is the top event in the fault tree, the events A, C, D are the elementary events. The events C and D are inputs of the logical element AND, and the event B is its output. This logical element represents the following logical connection between the lower-lying events C and D and the upper-lying event B: the event B occurs when both events C and D have occurred. The events A and B are similarly the input of the logical element OR, and the event Q occurs when at least one of the two events, A or B, has occurred.

Using the operations of Boolean algebra, we can represent fault trees by analytical expressions. As noted previously, AND and OR are the main logical elements in the graphical representation of the fault tree. The logical elements correspond to the operations of "logical multiplication" and "logical addition" in Boolean algebra. Since the logical elements AND and OR and their Boolean analogs perform identical logical operations on the set of events, there is a one-to-one correspondence between graphical and algebraic representations of the fault tree. For instance, in the fault-tree schema shown in Fig. 1, the event Q occurs when at least one of the lower-lying events, A or B, has occurred. This fault-tree fragment is equivalent to the Boolean expression

$$Q = A + B. \tag{7}$$

The event B occurs when two events, C and D, occur simultaneously. This fault-tree fragment is equivalent to the Boolean expression

$$B = C \cdot D. \tag{8}$$

Thus, functions of Boolean algebra can be used to write out a system of Boolean equations, which constitute an analytical representation of the fault tree.

The graphical representation of the fault tree is visually convenient for the user, but it is not particularly suitable for computer-based analysis. The analytical representation is more appropriate for this purpose.

The notion of minimum-cut set plays a major role in fault-tree analysis. The minimum-cut set is formally defined as a combination of the least number of elementary events that are sufficient for the top event to occur. This combination is called "minimal" among all combinations that lead to the occurrence of the top event: if one of the elementary events is removed from the minimal combination, the top event in the fault tree will not occur. The fault tree shown in Fig. 1 has two minimum-cut sets: {A} and {C, D}.

Any fault tree has finitely many minimum-cut sets. A minimum-cut set consisting of a single elementary events is called a minimum cut of first order; that consisting of two elementary events is called a minimum cut of second order, and so on. Algebraic transformation of the system of Boolean equations representing the fault tree with the aid of the rules of Boolean algebra may be applied to express the top event in terms of the minimum-cut sets. Each fault tree contains k minimum-cut sets $M_1, M_2, \ldots, M_k$; the set M_i consists of n_i elementary events $x_1, x_2, \ldots, x_n$, $i = 1, \ldots, k$. Then the Boolean equations for the top event T in the fault tree are representable in the form

$$\begin{aligned} T &= M_1 + M_2 + \ldots + M_k, \\ M_1 &= x_1 \cdot x_2 \cdot \ldots \cdot x_{n_1}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ M_k &= x_1 \cdot x_2 \cdot \ldots \cdot x_{n_k}. \end{aligned} \tag{9}$$

Using the above Boolean equations, we can write a formula for the probability of occurrence of the top event:

$$\begin{aligned} p(T) = p(M_1 + M_2 + \ldots + M_k) &= \sum_{i=1}^{k} p(M_i) - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} p(M_i \cdot M_j) \\ + \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{l=j+1}^{k} p(M_i \cdot M_j \cdot M_l) &+ \ldots + (-1)^k p(M_1 \cdot M_2 \cdot \ldots \cdot M_k), \end{aligned} \tag{10}$$

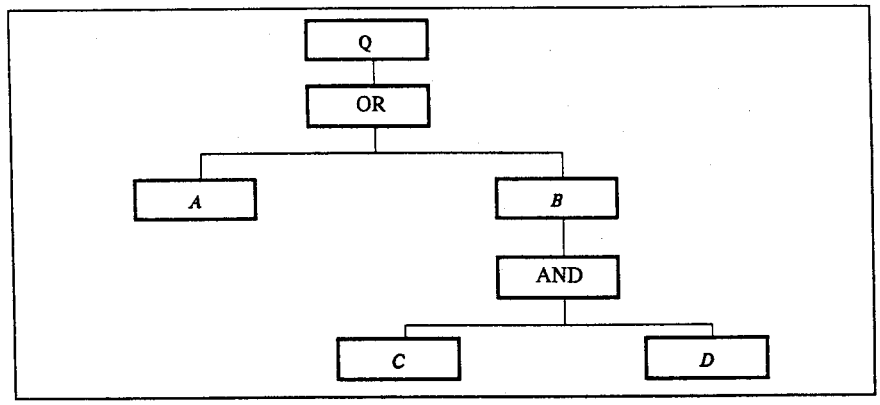

Fig. 1

where p(T) is the probability that the top event T occurs; $p(M_1 + M_2 + \ldots + M_k)$ is the probability that at least one of the events $M_1, M_2, \ldots, M_k$ occurs; $p(M_i \cdot M_j)$ is the probability that both events M_i and M_j occur simultaneously; $p(M_1 \cdot M_2 \cdot \ldots \cdot M_k)$ is the probability that k events $M_1, M_2, \ldots, M_k$ occur simultaneously.

Since the event M_i occurs when n_i elementary events $x_1, x_2, \ldots, x_{n_i}$ occur simultaneously, we obtain, assuming that all the elementary events from the minimum-cut set M_i are independent,

$$p(M_i) = \prod_{j=1}^{n_i} p(x_j). \tag{11}$$

We see from formula (11) that the fault probability associated with a minimum cut exponentially decreases with the increase of the order of the cut. For instance, if the fault probability of a unit is of the order of 10^{-3}, then the fault probability of a first-order cut is of the order of 10^{-3}, the fault probability of a second-order cut is of the order of 10^{-6}, of a third-order cut 10^{-9}, and so on. The probabilities $p(M_i)$ are thus very small, and in formula (10) we can ignore the possibility of simultaneous occurrence of two or more minimum-cut sets M_i. Then we obtain an approximate formula for the probability of occurrence of the top event in the form

$$p(T) = \sum_{i=1}^{k} p(M_i). \tag{12}$$

For the tree shown in Fig. 1, the top event Q is expressible in terms of the minimum cuts in the following form:

$$Q = M_1 + M_2; M_1 = A; M_2 = C \cdot D. \tag{13}$$

The probability of occurrence of the top event p(Q) is computed by the formula

$$p(Q) = p(M_1) + p(M_2) = p(A) + p(C \cdot D) = p(A) + p(C) \cdot p(D). \tag{14}$$

Formulas (9), (11), and (12) suggest a technique for solving the first of our problems: estimating the safety level (the accident probability) for the proposed nuclear power station project. The solution of this problem requires

a) analyzing the systems of the nuclear power station;

b) determining the sequence of faults that lead to accidents in the nuclear power station;

c) estimating the fault probabilities of the component units in the complex system that represents the nuclear power station;

d) computing the probabilities of occurrence of accidents in the nuclear power station based on system logic and fault probabilities of the component units.

Once all minimum-cut sets have been identified, we can determine the accident probability using formula (12).

Note that the minimum-cut sets obtained in this way contain more information about the design of nuclear power station systems than what is needed for estimating the accident probability. The minimum-cut sets can be used to analyze the sensitivity of overall system safety to faults in component units. The determination of sensitivity involves estimating the effect on overall

system safety of changes in system design, reliability of the component units in the system, and also the time moments when the units are tested and repaired in the process of technical maintenance of the nuclear power station.

The units are ranked qualitatively in the following way. The most significant from the point of view of their contribution to overall system failure are the first-order minimum cuts, then come the second-order minimum cuts, and so on. Faults corresponding to elementary events included in first-order minimum-cut sets have the greatest impact on the safety of the entire system. Designers of nuclear power stations must focus their attention specifically on these elementary events, and technical maintenance must primarily aim to neutralize these events. If there are no first-order minimum-cut sets, the most significant from the point of view of overall system safety are the elementary events in second-order minimum-cut sets, and so on.

Information about minimum cuts can be utilized directly to check the criteria that must be satisfied by the system design. For instance, if the design criterion is to ensure that failure of one of the component units does not lead to failure of the entire system, then this requirement is equivalent to the condition that the system fault tree should not contain first-order minimum-cut sets.

3. PLANNING THE OPTIMAL MAINTENANCE STRATEGY

So far we have ignored the time dependence of the probability of failure of the component units. The time dependence of the unit failure probabilities is essential for the solution of the second problem – planning the optimal maintenance strategy of an ecologically hazardous facility with the object of increasing its level of safety. In the static model of Sec. 2, the fault probability of a single component unit is taken equal to the time-averaged point values of its fault probabilities. Then the probability of the top event is computed from formulas (9), (11), (12). In the dynamic model (see [6-12]) considered in this section, the unit fault probabilities are a function of time. This model allows for the fact that the main system responsible for the safety of the nuclear power station (the protection system) can function in two modes – standby (when there is no danger of a nuclear accident) and active (when such a danger appears). The arrival of danger signals is viewed in what follows as a stream of requests for the protection system to be switched from standby to active state. After the arrival of such a request, the protection system should switch from standby to active or operating mode. The time during which the system is in an active state will be called the active or operating period. The entire life of a protection system thus consists of two phases: the time during which the system is in standby and the time during which the system is active, i.e., executes the functions for which it is designed. The probability of system failure at each time instant is calculated by combining the unavailability of the system in the standby state (using the fault tree for the standby phase) and its failure probability in the active mode (using the fault tree for the operating phase). The averaged probability of system failure is computed in this case by integrating the point values of system unavailability during the standby mode.

Note that the average system failure probabilities calculated for the static and the dynamic model are not necessarily equal. This is due to the fact that the point probability of system failure is determined by a particular set of minimum cuts and is equal to the product of the fault probabilities corresponding to the elementary events in the appropriate set of minimum cuts. As we know, the mean of a product is not equal to the product of the means. If the arrivals of requests to activate various units in the system were independent at different time instants, then the mean unavailability of the system could be determined by averaging the unavailability of each unit independently of all other units. If a request is received for simultaneous activation of all units, averaging can be carried out only on the level of the entire system.

If the arrival rate of requests and the point value of system unavailability are functions of time, the probability that an accident occurs during the time T is also a function of time. It is only when the arrival rate of requests is constant, i.e., the time during which the power station functions in a steady-state mode, that we may treat the accident probability as proportional to the mean unavailability. In general, with an arbitrary time dependence of the arrival rate of requests, correct estimation of the accident probability requires knowledge of the time dependence of system unavailability.

As we have noted previously, the reliability function of a system in standby mode can be split into two consecutive periods or phases: the standby phase and the active phase. In each phase, the system is expected to perform specific tasks. In the standby phase, the protection system should be available for immediate activation once the triggering events occur; in the active phase the system must function as planned for a prescribed time after switching from the standby mode. The system fulfills its functions if and only if it performs these two tasks. Thus, the probability of failure of the entire system equals the unavailability of the system to start functioning when a request arrives plus the probability that, having successfully switched

to the active mode, it will not be able to function normally during the prescribed time. The probability of system failure depends on the following factors: logical configuration of the component units, reliability characteristics of all units, and maintenance procedures in the standby phase and the operating phase. The fault tree for the standby phase, where the top event is defined as inability of the system to switch to active mode in response to an arriving request, can be essentially different from the fault tree that describes system failure after successful transition from the standby phase to the operating phase.

After the arrival of the request, the system configuration changes and during the operating period the fault regimes of the units may differ from the fault regimes of the same units in the standby phase. Therefore, to determine the accident probability, we need to construct separate fault trees for the standby phase and the operating phase.

Let us now consider in more detail the reliability characteristics of the system in the standby mode [6-12]. The reliability of the system in standby mode is characterized by point system unavailability, mean system unavailability, and contributions to mean system unavailability associated with testing, repair, and failure of the entire system.

The point unavailability Q(t) is defined as the probability that the system cannot perform its function at time t, in response to the occurrence of events which the system was designed to neutralize. The time dependence of the point unavailability of the entire system is affected by the point unavailability q(t) of each unit and on the logical interconnection of the units. To determine the time dependence of system unavailability, we need to compute the unavailability of each unit at discrete times with a certain interval and then use the logical interconnection between unit faults, as described by the fault tree, to compute the unavailability of the system at these time points by the formula

$$Q(t) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} q(M_{ij}), \tag{15}$$

where M_{ij} is the i-th minimum-cut set of j-th order, $q(M_{ij})$ is the contribution to system unavailability associated with the events in M_{ij}.

The mean system unavailability is the averaged probability that the system is unavailable at any given time instant t_0:

$$\overline{Q} = \frac{1}{T} \int_0^T Q(t)dt, \tag{16}$$

where T is the time interval during which the system is in standby.

Given the dependence Q(t), we can compute the effect of maintenance strategy on system unavailability. The maintenance strategy is defined by the following parameters. Testing of system components starts at time instants $t_1, t_2, \ldots, t_n$ and takes the time $\tau_1, \tau_2, \ldots, \tau_n$. If a fault is detected during the testing of a component, the repair starts immediately when the test ends and it takes the time T_{R_i}.

Then the contribution of testing to mean system unavailability is given by the formula

$$\overline{Q}_{test} = \frac{1}{T} \sum_{i=1}^{n} \int_{t_i}^{t_i+\tau_i} Q(t)dt, \tag{17}$$

where $(t_i, t_i + \tau_i)$ is the i-th interval in which at least one unit is being tested. The contribution of unit repairs to average system unavailability is given by the formula

$$\overline{Q}_{rep} = \frac{1}{T} \sum_{i=1}^{n} \int_{t_i+\tau_i}^{t_i+\tau_i+T_{R_i}} Q(t)dt, \tag{18}$$

where $(t_i + \tau_i, t_i + \tau_i + T_{R_i})$ is the i-th interval in which none of the units is being tested and at least one unit is being repaired.

The contribution to system unavailability associated with failure of the system at time instants when it is neither being tested nor being repaired is given by the formula

$$Q_{sys} = \overline{Q} - \overline{Q}_{test} - \overline{Q}_{rep}. \tag{19}$$

We can use formulas (17), (18) to analyze the effect of maintenance parameters on the unavailability of the entire system.

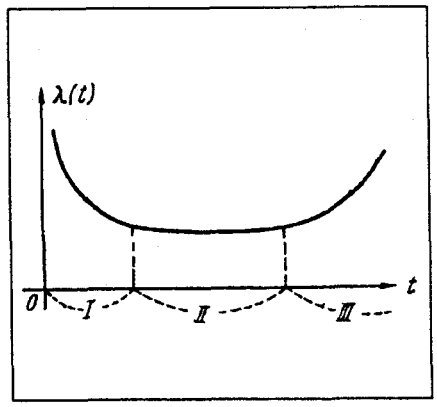

Fig. 2

We see from formula (15) that in order to determine the time dependence of system unavailability Q(t), we need to find the time dependence of unavailability q(t) of a single unit. The unavailability q(t) of a unit in standby mode is the probability that the unit will not be able to switch to active mode at time instant t when a request for such a transition arrives.

The form of the curve q(t) depends on various fault mechanisms and on the maintenance parameters for the particular unit. The reliability of a unit is affected by the following factors:

- fault elimination procedure;
- fault detection efficiency;
- maintenance program.

Depending on the reliability characteristics of the units in standby mode and on the maintenance strategy, we distinguish four types of unit fault models:

- unit with time-independent unavailability;
- nonrepairable unit;
- continuously monitored unit;
- periodically tested unit.

Faults may depend either on time or on arriving requests. The frequency of time-dependent faults increases with time, and the frequency of faults that depend on arriving requests increases with the number of arrivals. Arrival-dependent faults can be modeled by a unit with time-independent unavailability. The last three unit fault models are time-dependent. The nonrepairable unit model can be used to model units that are not repaired after they fail in the standby mode. Unit elements that can be repaired in standby mode are modeled either by a continuously monitored units or by a periodically tested unit. The continuously monitored unit model describes faults which are detected instantaneously and eliminated immediately. The periodically tested unit model describes faults which can be detected and repaired only during scheduled maintenance, which is performed at regular time intervals.

Both models (continuously monitored unit and periodically tested unit) allow for fault detection efficiency: some faults remain undetected during testing, and such units are not repaired.

The periodically tested unit model includes an option for disconnecting the unit (with finite probability) and allows test-induced faults. Test-induced faults occur in the process of testing. Such faults are detected immediately and are repaired after the test ends.

To compute the unavailability q(t) both in the continuously monitored unit model and in the periodically tested unit model, we have to introduce time-independent unavailability modeling service personnel errors.

Both detected and undetected faults are modeled by the fault rate function $\lambda(t)$ of a nonrepairable unit. In general, this function has the form shown in Fig. 2. The time axis is divided into three intervals: I – the run-in or "burn-in" interval; II – the normal operation interval; III – the aging interval. The run-in interval corresponds to the early stage in the life of the component, when latent manufacturing defects are gradually revealed. The normal operation interval is characterized by a constant fault rate function; random behavior of faults is usually observed in this period. Finally, the aging interval describes component faults associated with irreversible physico-chemical effects, which are responsible for the aging of the element. In this time period, the fault rate function $\lambda(t)$ is increasing with time.

For normal operation

$$\lambda(t) = \lambda = \text{const} \tag{20}$$

and the fault probability is given by

$$F(t) = 1 - \exp(-\lambda t). \tag{21}$$

With an exponential fault distribution, the unavailability of a nonrepairable component in a given time interval (t, t + τ) is independent of past history and depends only on the interval length τ.

A generalization of the exponential fault distribution is the Weibull distribution, which contains an additional parameter β. For the Weibull fault distribution, the fault rate function is given by the formula

$$\lambda(t) = \lambda\beta t^{\beta - 1}, \tag{22}$$

where λ, $\beta > 0$.

The values of the parameters λ and β can be fitted so that the Weibull distribution function corresponds to the curve sections shown in Fig. 2. The exponential fault rate is a particular case of the Weibull distribution for $\beta = 1$.

In what follows we assume that the fault rate follows the Weibull law. If a unit did not fail in the time interval (t_r, t), then the probability of failure in the interval (t, t + dt) is

$$\lambda(t)dt = \beta\lambda(t - t_r)^{\beta - 1} dt. \tag{23}$$

The parameter β characterizes the run-in section ($\beta < 1$) or the aging section ($\beta > 1$). The scalar parameter λ determines the absolute fault rate. The time t_r defines the last time instant before the time t when the reliability characteristic of the unit is the same as at time t = 0 (the time when the unit was connected to the system).

The relationships between the Weibull parameters λ, β and the mean time to failure $\overline{T}$ are given by

$$\overline{T} = \Gamma[(\beta + 1) / \beta]\lambda^{-(1/\beta)}, \tag{24}$$

$$\lambda = [\Gamma((\beta + 1) / \beta) / \overline{T}]^{\beta}, \tag{25}$$

where Γ is the gamma-function.

With an exponential fault rate $\lambda(t) = \lambda = \text{const}$, $\beta = 1$, and we have

$$\overline{T} = \frac{1}{\lambda}. \tag{26}$$

Let us now consider each of the fault models assuming a Weibull fault rate.

Time-Independent Unavailability Model

The model is described by the relationship

$$q(t) = g_1 = \text{const}. \tag{27}$$

Here g_1 is the probability that the particular fault occurs at the instant when the request arrives. As we have noted previously, this fault model can be used to model the probability of stress-induced service personnel errors.

Nonrepairable Unit Model

The unavailability of a nonrepairable unit in standby mode is computed from the formula

$$q(t) = 1 - \exp(-\lambda t^{\beta}), \tag{28}$$

where λ and β are the Weibull parameters. For large t ($t \gg \bar{T}$), $q(t) \to 1$, and for small t ($t \ll \bar{T}$)

$$q(t) \approx \lambda t^{\beta}. \tag{29}$$

The mean unavailability $\bar{q}(T)$ is thus computed from the formula

$$\bar{q}(T) = \frac{1}{T}\int_0^T \lambda t^{\beta} dt = \lambda T^{\beta} / (\beta + 1). \tag{30}$$

Thus, if the time that the system spends in standby mode is much shorter than the mean unit life T, the mean unavailability of a nonrepairable unit depends on the time that the system remains in the standby mode. If $t \gg \bar{T}$, the mean unavailability of a nonrepairable unit does not depend on the time that the system remains in standby mode and goes asymptotically to 1.

Continuously Monitored Unit

This model makes it possible to compute the unavailability of units in which faults are detected as soon as they appear. The fault is repaired immediately after it is detected. The fault rate follows the Weibull distribution and the repair time is constant. Two types of recovery of a faulty unit are possible:

– the unit is restored to its initial state as observed at the time when the unit was connected to the system (total recovery);

– after failure the unit is restored to a state in which the fault rate is the same as directly prior to failure (partial recovery).

The unavailability of a continuously monitored unit depends on the following parameters:

– the Weibull scalar parameter λ;

– the Weibull parameter β;

– the mean repair time T_R;

– monitoring efficiency p (the detection probability of a fault in a continuously monitored unit);

– the probability g_1 that a fault occurs simultaneously with the arrival of the request (this fault is associated with factors that are different from those incorporated in time-dependent fault models);

– the recovery type of a faulty element – total or partial recovery.

For very small t ($t \ll \bar{T}$), when the probability of several simultaneous faults is negligible, the unavailability of a continuously monitored unit for the two recovery types is approximately given by the formula

$$q(t) = \lambda\beta T_R t^{\beta - 1}, \tag{31}$$

and the mean unavailability for $T \ll \bar{T}$ is calculated from the formula

$$\bar{q}(t) = \frac{1}{T}\int_0^T \lambda\beta T_R t^{\beta - 1} dt = \lambda T_R T^{\beta - 1}. \tag{32}$$

Note that if the appearance of several faults is little probable, the recovery type does not play a special role. The recovery type of the faulty unit becomes significant for large t ($t \gg \bar{T}$).

In case of total recovery of a faulty unit for $t \gg \bar{T}$, the point unavailability is approximated by the value q_∞, which is calculated from the formula

$$q_\infty = T_R / (T_R + \bar{T}). \tag{33}$$

If there is a probability p that some faults in a continuously monitored unit remain undetected, then the corresponding part of faults can be treated as nonrepairable faults. Then for these faults the fault rate is $\lambda_1 = p\lambda$, and for the remaining faults in a continuously monitored unit $\lambda_2 = (1 - p)\lambda$. Thus, the complete expression for unavailability allowing for the probability g_1 can be written in the form

$$q(t) = g_1 + (1 - g_1)[g_2(t) + (1 - g_2(t))g_3(t)], \tag{34}$$

where $g_3(t)$ is the unavailability associated with detected faults ($\lambda = \lambda_2$) and $g_2(t)$ is the unavailability associated with faults that are not detected at time t ($\lambda = \lambda_1$).

Periodically Tested Units

Protection systems normally use periodically tested units, which remain in standby mode until there is a danger of an accident. Detailed modeling of point unavailabilities of periodically tested units must allow for the following factors: the effect of disconnecting the unit from the system for testing and repair; faults associated with testing; imperfect fault detection during testing; service personnel errors. The fault rate in periodically tested units is modeled as a function that depends on time and on the number of unit testing operations.

To describe the reliability of periodically tested units, we need to specify some new parameters in addition to the Weibull parameters λ and β. These new parameters include T_1 and T_2, where T_1 is the time from the beginning of the standby phase to the first test and T_2 is the time between two successive tests. We also need to specify the testing time τ and the repair time T_R, the probability that a fault remains undetected p, the constant unavailability component g_1, and other parameters.

The time τ to run scheduled tests and perform technical maintenance of a unit includes the actual testing time during which the component is not available for its intended use and the repair time prescribed by the maintenance program. This parameter does not include the time to repair a suddenly failing component. For instance, suppose that a unit is tested every month. A current test detects a deviation of the technical state of the unit. This deviation does not interfere with the unit's ability to perform its prescribed functions, but in the future may lead to failure of the unit. The maintenance program in this case prescribes some minimum work to correct the observed deviation, followed by repeated testing. Since the unit was nonfaulty during testing, this maintenance program must be included in the value of the parameter τ.

The mean duration of unscheduled repairs T_R includes the mean time of repairs which are performed when a fault is detected in the unit. The length of unscheduled repairs includes the total time from the instant when the fault is detected to the instant when the repaired unit passes the repeated test and is reconnected to the system. This parameter does not include the normal maintenance time (when the unit does not fail), which is incorporated in the parameter τ. It is assumed that during the entire time of unscheduled repairs the unit is disconnected from the system. In modeling a periodically tested unit, we accordingly allow for partial unavailability of the unit by reducing the estimate of the parameter T_R.

There are three types of recovery for a periodically tested unit.

1. The unit fault rate is completely restored during both testing and repair.
2. The unit fault rate is only partially restored during both testing and repair.
3. The unit is partially restored during testing and completely restored during repair. If a fault is detected during periodic testing, the complete recovery point is when the repair work is finished. This model is often used to describe aging components, which are replaced with new components when a fault is detected.

The behavior of the unavailability curve q(t) of a periodically tested unit is determined by the parameters considered above. If a periodically tested unit fails in the time interval between two successive tests, the fault remains undetected until the next testing phase begins, and the unit remains unavailable until the repairs are completed.

The time dependence of unavailability of a periodically tested unit is a periodic curve of period T_2. On the n-th interval (n-th testing starts at time t_n, and it may be followed by repairs during the time T_R), the unavailability q(t) consists of the following components.

1. The unavailability component associated with testing (unavailability during the testing time τ):

$$q(t) = q_1^n \quad \text{for} \quad t_n < t \leq t_n + \tau. \tag{35}$$

2. The unavailability component associated with repairs (unavailability for the duration of repairs T_R):

$$q(t) = q_2^n \quad \text{for} \quad t_n + \tau < t \leq t_n + \tau + T_R . \tag{36}$$

3. The unavailability component associated with failure of the unit (unavailability during the time interval between two successive tests):

$$q(t) = q_3^n(t) \quad \text{for} \quad t_n + \tau + T_R < t \leq t_{n+1} .$$

We assume that q_1^n and q_2^n are constant on the corresponding time intervals.
Unavailability of the unit immediately before the n-th test is

$$Q_n = q_3^n(t_n). \tag{37}$$

Reliability Characteristics of the System in Operating Mode

The unreliability of the system $G(t_0, t_0 + T)$ is defined as the probability that the system cannot start performing its functions at the instant t_0 when the danger of an accident is sensed or, having successfully started working, it fails in the time interval $(t_0, t_0 + T)$, where T is the system operating time (necessary to eliminate the danger of an accident). According to this definition, the expression for the probability of system failure has the form

$$G(t_0, t_0 + T) = Q(t_0) + [1 - Q(t_0)]F(t_0, t_0 + T), \tag{38}$$

where $Q(t_0)$ is the unavailability of the system in the standby mode at time t_0, when a request to activate the system arrives; $F(t_0 + T, t_0)$ is the probability that the system, having successfully switched to active state, will not be able to function normally during the scheduled time T.

Suppose that the system switched to active mode at time t_0. The mean number of system failures in the time interval $(t_0, t_0 + T)$ is given by

$$M(t_0, t_0 + T) = \int_{t_0}^{t_0 + T} m(u)du, \tag{39}$$

where m(u) is the recovery density [8], equal to the mean number of faults in unit time at the instant u. From the definition of the mean, we can write

$$M(t_0, t_0 + T) = \sum_{i=1}^{\infty} i \cdot p_i, \tag{40}$$

where p_i is the probability that there will be i system failures in the time interval $(t_0, t_0 + T)$. If all the units in the system are nonrepairable, we have

$$M(t_0, t_0 + T) = p_1 . \tag{41}$$

Then the probability $F(t_0 + T, t_0)$ that failure occurs in the time interval $(t_0, t_0 + T)$ after the system starts operating at t_0 is given by the formula

$$F(t_0, t_0 + T) = M(t_0, t_o + T). \tag{42}$$

Thus, if the system consists of nonrepairable units, the probability that failure occurs in the time interval $(t_0, t_0 + T)$ during system operation is computed by integrating the system recovery density m(t). In this case, the system recovery density m(t) is equal to the failure density f(t) and we have the equality

$$m(t)dt = f(t)dt, \tag{43}$$

where f(t)dt is the probability that the system fails in the interval (t, t + dt) given that no failures occurred prior to the instant t.

If the system consists of repairable components, at most one fault may occur on a sufficiently small time interval (t, t + dt) and we have

$$M(t, t + dt) = p_1, \tag{44}$$

where p_1 is the probability that there is one system failure in the interval (t, t + dt).

On the other hand,

$$M(t, t + dt) = m(t)dt. \tag{45}$$

Thus, if the system consists of repairable units, the mean number of faults in a sufficiently small time interval also can be interpreted as the probability of system failure in this interval.

If the time T is not sufficiently small, then contrary to the case of a system with nonrepairable components, the integral

$$M(t_0, t_0 + T) = \int_{t_0}^{t_0 + T} m(t)dt, \tag{46}$$

strictly speaking, should be interpreted as the mean number of faults, and not as the probability of system failure. However, for high-reliability systems, $M(t_0, t_0 + T)$ is a sufficiently good approximation of the probability $F(t_0, t_0 + T)$ that the system will fail on the interval $(t_0, t_0 + T)$. Therefore,

$$F(t_0, t_0 + T) = M(t_0, t_0 + T). \tag{47}$$

Thus, if some components are repairable, the exact value of system failure probability on the time interval $(t_0, t_0 + T)$, which is fairly difficult to compute, can be replaced with $M(t_0, t_0 + T)$ as a sufficiently accurate approximation of the estimated probability.

It follows from formulas (46), (47) that to compute the failure probability of a system in the active phase on an arbitrary time interval $(t_0, t_0 + T)$, we need to find point values of the system recovery density m(t).

As we have noted previously, the point system recovery density m(t) can be interpreted as the density of system failure probability at the instant t. Similarly to the point unavailability of the system in the standby phase, this probability depends on the system structure and on the characteristics of unit faults. Assume that we have constructed the fault tree for a system in the operating mode and determined the minimum cuts. The top event in this fault tree is identified with failure of the entire system. For the top event to occur in the time interval (t, t + dt), it is necessary that none of the minimum-cut sets occurs up to the time t and then one or several minimum-cut sets occur in the interval (t, t + dt).

Ignoring the probability of simultaneous occurrence of two or more minimum cuts in a small time interval (t, t + dt), we obtain that the probability of system failure in the interval (t, t + dt) can be computed as the sum of the probabilities of all minimum cuts:

$$m(t)dt = \sum_{i=1}^{N} m_{c_i}(t)dt, \tag{48}$$

where $m_{c_i}(t)dt$ is the probability that the i-th minimum cut occurs in the interval (t, t + dt). A minimum cut occurs in the interval (t, t + dt) if all the units forming the minimum cut, except one, are unavailable at the instant t and the one nonfaulty unit fails in the time interval (t, t + dt). If we assume that units fail independently of one another, then $m_{c_i}(t)$ is given by the formula

$$\begin{aligned} m_{c_i}(t)dt &= p_2(t)p_3(t)\ldots p_{n_i}(t)m_1(t)dt \\ &+ p_1(t)p_3(t)\ldots p_{n_i}(t)m_2(t)dt + p_1(t)p_2(t)\ldots p_{n_i}(t)m_3(t)dt \\ &\ldots + p_1(t)p_2(t)\ldots p_{n_i-1}(t)m_{n_i}(t)dt, \end{aligned} \tag{49}$$

where $p_1(t)$, ..., $p_{n_i}(t)$, $m_1(t)$, ..., $m_{n_i}(t)$ are the point unavailabilities of the units forming the i-th minimum cut in the operating mode and their recovery densities at time t; n_i is the order of the i-th minimum cut. Reducing the right- and left-hand side of Eq. (48) by dt, we obtain an approximate expression for the recovery density of the system in operating mode:

$$m(t) = \sum_{i=1}^{N} m_{c_i}(t), \tag{50}$$

where N is the number of minimum cuts in the fault tree.

Thus, the probability of an accident in a nuclear power station is proportional to the probability that the protection system does not start functioning at the instant t_0 when the danger of an accident is sensed or, having successfully started functioning at t_0, fails in the time interval (t_0, t_0 + T). To compute the time dependence of the accident probability we have to compute from formulas (38), (46)-(49) the point unavailabilities recovery densities of the units in the operating mode.

Having determined the time dependence of system unreliability $G(t_0, t_0 + T)$, we can compute the reliability characteristics of the system that determine the effect of technical maintenance parameters on system safety. These characteristics include:

– the mean unreliability of the system in the time interval T, when the system is in the standby phase:

$$\bar{G} = \frac{1}{T}\int_0^T G(u, u + T)du; \tag{51}$$

– the contribution to mean system unreliability during the time interval when the component units are being tested:

$$\bar{G}_{\text{test}} = \frac{1}{T}\sum_{i=1}^{k}\int_{t_i}^{t_i+\tau_i} G(u, u + T)du, \tag{52}$$

where (t_i, $t_i + \tau_i$) is the i-th interval in the standby phase, when at least one unit is being tested;

– the contribution to mean system unreliability in the time interval when the faulty units are being repaired:

$$\bar{G}_{\text{rep}} = \frac{1}{T}\sum_{i=1}^{k}\int_{t_i+\tau_i}^{t_i+\tau_i+T_{R_i}} G(u,u + T)du, \tag{53}$$

where ($t_i + \tau_i$, $t_i + \tau_i + T_{R_i}$) is the i-th interval in the standby phase when none of the units is being tested and at least one is being repaired.

Using these system reliability characteristics, we can vary the maintenance parameters until optimal values are found that maximize the system safety level.

REFERENCES

1. R. E. Barlow and F. Proschan, Statistical Theory of Reliability and Life Testing, Holt, Reinhart and Winston, New York (1974).
2. J. B. Fussell, "A formal methodology for fault tree construction," Nucl. Sci. Eng., **52**, 421-432 (1973).
3. J. B. Fussell, "Fault tree analysis – concepts and techniques," Proc. NATO Adv. Study Inst., Generic Techniques of System Reliability Assessment, Liverpool, England (1973).
4. Z. W. Kaufman, D. Grouchco, and R. Cruon, Mathematical Models of the Study of the Reliability of Systems, Academic Press, New York (1977).
5. E. J. Henley and H. Kumamoto, Reliability of Technical Systems and Risk Assessment [Russian translation], Mashinostroenie, Moscow (1984).
6. N. H. Roberts, W. E. Vesely, D. F. Haasl, and F. F. Goldberg, Fault Tree Handbook, US Nuclear Regulatory Commission, NUREG-0492 (1979).

7. W. E. Vesely, "A time-dependent methodology for fault-tree evaluation," Nucl. Eng. Design, **13**, 337-360 (Apr. 1970).
8. B. V. Gnedenko, Yu. K. Belyaev, and A. D. Solov'ev, Mathematical Methods in Reliability Theory [in Russian], Nauka, Moscow (1965).
9. S. Garribba, G. Reina, and G. Volta, "Availability of repairable units when failure and restoration rates age in real time," IEEE Trans. Reliab., **R-25**, No. 2, 88-94 (1976).
10. W. L. Smith and M. R. Leadbetter, "On the renewal function for the Weibull distribution," Technometrics, **5**, No. 3, 393-396 (1963).
11. J. B. Fussel and G. R. Burdick, in: Proc. Int. Conf. on Nucl. Systems Reliability and Risk Assessment, Gatlinburg, Tennessee (1975).
12. A. E. Green and A. J. Bourne, Reliability Technology, London (1972).