Modelling and dependability evaluation of safety systems in control and monitoring applications

Jean Arlat, Karama Kanoun

MODELLING AND DEPENDABILITY EVALUATION OF SAFETY SYSTEMS IN CONTROL AND MONITORING APPLICATIONS

J. Arlat and K. Kanoun

Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse Cedex, France

ABSTRACT

The paper addresses the problem of dependability evaluation of high safety computer systems. Continuously activated control systems and incident-driven monitoring systems that are considered in the paper are characterized in the framework of a general analysis and classification of safety systems. The behavior of these two types of system is desribed by a comprehensive unified model that is useful for (a) the definition of relevant dependability measures and (b) the identification of the main features of the modeling task. Based on this model a general evaluation methodology is presented and applied to the study of simplex and duplex structures of both control and monitoring safety systems. Special emphasis is put on the identification of the tradoff between the safety and availability aspects.

Keywords: safety systems, dependability, modeling, evaluation, Markov process.

INTRODUCTION

Safety systems considered in this paper are computer systems associated with critical applications whose failure may have catastrophic consequences due to human and/or economical reasons. Accordingly, during the design of such systems special emphasis is put on dependability evaluation and validation.

Although this problem of dependability validation covers a wide range of problems and approaches (LAP 85), we focus our presentation on the modeling and evaluation aspects only.

The paper extends and generalizes the work of two previous studies (LAP 80, ARL 85) which concern:
- the definition of a new architecture for the monitoring system of an Extra High Voltage substation for the French electrical network, the overall results obtained constituted a data base for the design of a new system (MED 80),
- the design and validation of a computerized interlocking system for the French railroad (ARL 84).

Section I provides a comprehensive framework for the analysis of safety systems and characterizes the types of safety systems that are considered in the paper: namely, control and monitoring systems. In section II a general model of the behavior of these systems is presented which allows the derivation of relevant dependability measures suitable for their evaluation. Guidelines for the applied methodology of evaluation are also presented and discussed in section III. Section IV is devoted to the dependability evaluation of simplex and duplex safety systems.

I - ANALYSIS OF SAFETY SYSTEMS

A general representation of the types of safety systems considered is given in figure 1 which details the interactions between the different elements and the terminology used throughout the paper.

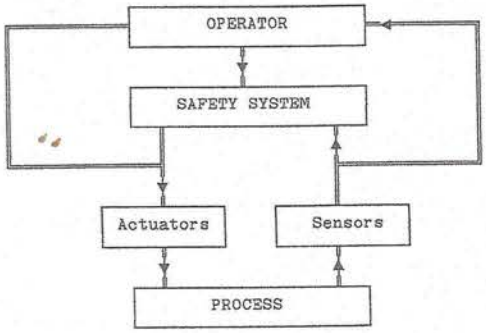


Figure 1

Let us now introduce and discuss here two aspects that we feel important in the study of safety systems; they are related to :
- the response time of the process to an erroneous output from the safety system,
- the ability of the process to exhibit two failure modes: a benign mode, where the system is brought into a prescribed safe state, and a catastrophic mode.

A. Process Response time

Depending on the reaction time of the process to an erroneous order, (i. e. the latency of a process failure before it leads to a catastrophic failure at the end user level) different graded approaches are applicable to ensure safe behavior.

1) Response Time = 0 : the process reacts quickly and leads directly to a catastrophic failure; the safety mechanisms thus have to be included in the safety system in order to reduce (or avoid) any erroneous output to the process.

2) Response Time ≠ 0 : the reaction of the process is not immediate and the deviation of the process is gradual, this latency time allows the application of appropriate safety actions in order to avoid catastrophic failure.

B. Presence of two failure modes

The ability of the process to exhibit two failure modes: a benign mode and a catastrophic mode, depends strongly on the type of application considered. From the functional point of view, this ability depends on (1) the possibility of the system to degrade the quality of service, (2) the existence of a stable safe state.

In the benign failure mode, the functional mission cannot be fully accomplished but safety constraints are preserved; e. g. process activity is frozen for a certain time or the process is put into a safe shut-down state.

C. Safety Systems Classification

According to the previously introduced aspects, safety systems can be classed in four groups as shown in figure 2. In order to be more specific, we will give typical examples of applications with high safety requirements for each group of figure 2.

RESPONSE TIME	BENIGN FAILURE MODE YES	NO
0	GROUP 1	GROUP 3
≠ 0	GROUP 2	GROUP 4

Figure 2

Ground transportation (e.g. railroad traffic control) is an illustration of the types of processes in **group 1**. An erroneous points position cannot be tolerated in the case of dense traffic. However, in case of failure of the control system, it is usually assumed that it is possible to stop the traffic. If this assumption is still valid in most of the practical cases, it has to be tempered by the fact that the capability of providing a freeze state is less and less likely as traffic becomes denser; this is already the case in subway systems.

Most safety systems devoted to industrial processes (chemical processes, nuclear power plants...) may be found in **group 2**. A characteristic example is that of electrical power distribution systems. In this case, most emphasis on the protection is put on the monitoring of the network electrical parameters and although it is economically undesirable, partial deconnection of the network provides a possible stable shutdown state.

Groups 3 and 4 can be characterized by aeronautical applications related to new-generation aircraft with reduced stability margins where the safety of the flight will depend upon active controls derived from computer outputs; it follows that no safe shutdown state is possible in this case. Group 3 corresponds to critical phases of the mission: take-off and landing where no erroneous command can be tolerated. Examples of control systems that belong to these groups are described in (HOP 78, WEN 78).

For systems belonging to groups 3 and 4, the classical reliability measure is applicable. However, for systems of groups 1 and 2, one should note that the presence of a benign failure mode is, most of the time, associated with the property of maintenance of the safety system.

The type of maintainability considered here is the maintenance (or operator) actions that are applied during the mission either (1) on-line, for the redundant parts of the system or (2) off-line, when the process has been put in the benign failure mode. However, no maintenance actions are considered when the process is in the catastrophic failure mode. We will be interested essentially in the impact of the application of off-line maintenance actions in the derivation of relevant measures.

We do not explicitly consider here a specific classification with respect to maintainability since its impact on the dependability measures is closely related to the presence of a benign failure state; it has, however, to be noted that among all application domains that are considered, all are maintainable after a benign interruption except aeronautics.

We concentrate our presentation on the evaluation of the large proportion of safety systems of groups 1 and 2 of figure 2 referenced as control and monitoring systems respectively hereafter.

II - DEPENDABILITY MEASURES

A. General Behavior of Safety Systems

The evaluation is essentially devoted to rate the safety system dependability either a control system or a monitoring system. Figure 3 presents a general model of the behavior for such systems.

Three major state classes may be identified which correspond to:
- **accomplishment (A)**: the safety system accomplishes its tasks in conformity with its specifications,
- **benign interruption (B)**, following a benign failure of the Safety system; it is assumed that, as a consequence, the process is halted,
- **catastrophic interruption (C)**, following a catastrophic failure; although this position constitutes a pessimistic point of view, a catastrophic failure of the process will be assumed whenever the safety system considered exhibits a catastrophic interruption, i. e. it provides erroneous (or no) command to the process.

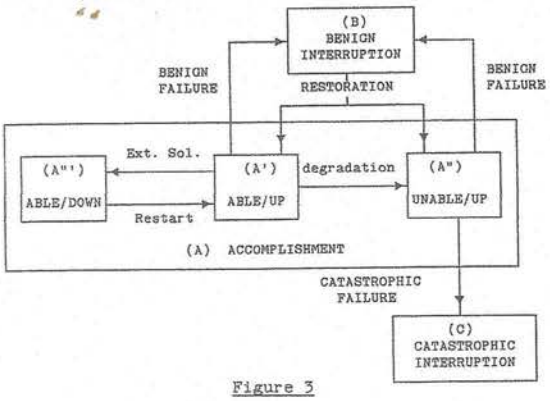


Figure 3

This high level partition is in conformity with the one presented in (ARL 85) which addressed essentially control systems that are continuously exercized. Accordingly, the interaction with the process was not explicitly considered, and emphasis was mainly put on the modeling of the ability of the system to successfully handle solicitations corresponding to error activation at the system level.

We introduce here a further refinement of class A into three sub-classes in order to (1) specify the main features of the modeling task when dealing with specific systems and (2) show how this partition applies to monitoring systems that are basically exercized in response to external solicitations corresponding to the occurrence of incidents in the monitored process.

1) Accomplishment Class

Sub-class A' characterizes the cases when the system is up and able to perform its task: it is either error-free or the latent errors accumulated in it will not hamper correct processing of a subsequent solicitation.

For control systems, solicitations correspond to the activation of an error in the system. This event covers two main types of activations:
- internal activations, with respect to the imbedded error processing mechanisms,
- external activations, with respect to the commands sent to the process.

The underlying stochastic processes model the elapsed times (latencies) between the creation of a latent error and its activation at the considered level; it has to be noted that these processes are inoperant when the system is error-free. These activations will only provoke a transition to the benign interruption class B when the provided redundancy is

no longer sufficient to bring these errors back into a latent state on an on-line basis.

For monitoring systems, one also has to explicitly consider external solicitations induced by incidents in the process. Two types of incidents must be considered:
- T1 incidents: actions carried out by the monitoring system are enough in order to restore the normal working conditions of the monitored process,
- T2 incidents: the monitored process is shut down and manual maintenance intervention is required.

It will be assumed that T2 incidents will lead to an interruption of service of the monitoring system. Although it is down, the monitoring system is still able to accomplish its task; thus, this case has to be distinguished from a service interruption due to the inability of the monitoring system to perform its task. In order to stress this difference we consider that this case is part of the accomplishment class A: it corresponds to sub-class A"'.

Sub-class A" gathers all potentially dangerous states; it corresponds to the cases when, due to the accumulation of specific patterns of latent errors (at system level), the safety system has been significantly degraded and is no longer able to perform its assigned task; when actually solicited (external activation or solicitation), the system enters the catastrophic interruption class. However, a benign failure induced, for example, by further accumulation of errors that will trigger the imbedded error processing mechanisms and/or preventive maintenance actions may still occur and provoke a transition from class A" to the benign interruption class B.

In practice, for the user, the distinction between A' and A" is impossible and then, the discrimination between them is strongly dependent on the error hypotheses that underly the level of redundancy provided for the system and the design of its error processing mechanisms.

Sub-class A"' is empty in the case of control systems since the notion of external solicitation is implicit. The only transitions from class A"' are to class A' following the maintenance of the monitored system and the restart of the system.

2) Benign Interruption Class

Although the benign interruption class B may gather a large number of states that are induced by the difference in the types of (off-line) restoration actions that have to be performed in each case, we do not further decompose this class here. However, it is important to note that, depending on the quality of the restoration, the system may be brought back either into class A' or A".

3) Catastrophic Interruption Class

Equally, we do not further decompose the notion of catastrophic interruption, although, in practice, this notion covers a large spectrum of failure behaviors due to the fact that a catastrophic failure may (1) result from various types of problems, and (2) propagate.

Also, no system restoration is considered following a catastrophic interruption; this results from the fact that the consequences of a catastrophic failure are such that system restoration is much less important than the repair of the consequences (damage to property, law suits, etc.) and the analysis of the causes (board of inquiry, etc.).

B. Definition of the Measures(LAP 85)

The model in figure 3 allows us to define the considered measures. Let z(t) be the state function of the system defined over the set of state classes (A,B,C). In what follows, it will be assumed that $z(t=0) \in A'$ and more specifically that the system is initially error free.

The main quantity of interest when dealing with safety systems, is the time spent in the safe states before catastrophic interruption. Let MST (Mean Safe Time) denote the mean time spent in the safe class S = A U B, and S(t) the safety function, i. e. the probability of avoiding catastrophic interruption, they are defined by:

$$S(t) = \text{Prob}\ (\ z(x) \in (A \cup B)\ ,\ \text{for all } x \in (0,t)\ )$$

$$MST = \int_0^\infty S(t)\ dt.$$

However the very existence of a safety system has a counterpart: it can fail (benign failure) and thus lead to the loss of service. In order to measure the influence of the safety system on the service delivered to the process we introduce another quantity of interest: the time during which the service is interrupted due to the safety system's inability to conduct or monitor the process. Let MBT (Mean Benign interruption Time) denote this mean time and B(t) the point benign interruption function defined as:

$$B(t) = \text{Prob}\ \{z(t) \in B,\ z(x) \in S,\ \text{for all } x \in (0,t)\}$$

$$MBT = \int_0^\infty B(t)\ dt.$$

An average unavailability-like measure can thus be defined that rates the ratio of time spent in state-class B with respect to the time spent in the safe class S, before a catastrophic failure. This measure is called, the unavailability before catastrophic failure (UAC) and is formally defined as:

$$UAC = MBT / MST.$$

Comments about the definition of S(t)

Although states in sub-class A" are potentially dangerous, they are considered as safe states: the safety function as defined here is thus different from the classical one for which potentially dangerous states are not safe. This function is also different from the classical reliability measure for which states in class B are exluded from S.

Tradoff: It is worth noting that when the system is in class B it is safe and it cannot enter the catastrophic state C, so increasing the benign interruption time will improve the time to catastrophic interruption; we see thus the tradoff between the two measures considered.

Both point Safety (S(t)) and unavailability before catastrophic failure (UAC) will be used as complementary evaluation measures for the safety systems under consideration.

III - EVALUATION METHODOLOGY

The main goal of evaluation is to provide objective aids in comparing the merits of various structures of a safety system during the design phase. This comparison is based on the measures defined in the previous paragraph. One tough task is to obtain a model of system behavior that (1) adequately embodies the most significant features of the system and (2) is sufficiently tractable to allow intensive sensitivity analysis of its parameters when comparing different structures.

A. Construction of the Model

In the attempt to derive a tractable yet representative model, the major difficulties are related to:

(1) the choice of the distribution functions of the random variables considered in the model,

(2) the number of accumulated latent errors in sub-class A" that have to be accounted for,

(3) the identification of the most significant variables among those considered in the model.

Two extreme approaches, involving intensive studies based jointly on theoretical results and

numerical analysis of the models with respect to the considered measures can be used in order to validate the model:

- start-small: formulation of simplifying hypotheses leading to a "basic" model which is a posteriori validated by a study of the impact of their relaxation ("extended" model),
- start-big: thorough description of system behavior is considered in a "complete" model that is a posteriori simplified ("simplified" model).

The start-small approach is in particular used to address the problem of the non-exponentially distributed variables: all variables are first assumed to be exponentiallly distributed; serial and parallel extension of the states (COX 68) to which apply the non-exponentially distributed variable(s) is explicitly carried out. Based on previous results (LAP 75), in both cases one supplementary state is sufficient to reveal the impact of the modification of the distribution.

As a consequence, the dependability measures considered here can be obtained using time homogeneous Markov processes. This is particularly useful since, in most of the cases, Markov transition graphs deduced from the general model of figure 2, are generally too complex to allow exact analytical expressions of S(t) and UAC to be formulated; thus numerical solutions are of prime interest.

The start-small approach is also used to study the problem of the number of errors accumulated in sub-class A": the number of accounted latent errors is a priori limited to a low value and its impact is studied either numerically, by progressively increasing the number of states of the model, or analytically using for example results concerning the mergeability property of Markov processes (e.g., see HOW 71; pp. 38-40).

The identification of the most significant variables corresponds essentially to a start-big approach: significance is estimated on the basis of the impact induced, by the modification of the graphs and the values of the variables, on the evaluated measures. This may be obtained from a study based jointly on theoretical results on Markov processes and on an intensive numerical analysis of the models. Two major theoretical results prevail in this study.

1- The strong connectivity of the graph that is induced by the maintainability of the considered systems allows identification of an equivalent catastrophic failure rate q_{eq} which can be directly related to the safety (S(t)) and mean safe time (MST) measures as (see PAG 80; pp. 167-170):

$q_{eq} \approx (MST)^{-1}$
$S(t) \approx Exp\ (-\ q_{eq}\ t)$, i. e. $S(t) \approx Exp\ [-(MST)^{-1}\ t]$

Accordingly, S(t) is entirely specified by the estimation of q_{eq}.

2- The difference in orders of magnitude of the considered processes (e.g., see LAP 76), for example, restoration of the safety system with respect to the fault process and incident processing or process maintenance with respect to the solicitation process, allows handling of further simplifications on the graph and/or closed-form approximate expressions of the dependability measures.

These simplified expressions help to identify the most significant parameters and thus reduce the transition graphs and can be validated by comparison with numerical results obtained from the SURF program (COS 81).

B. Parameters of the model

We define here the various parameters (processes and associated factors) that are considered.

As usually done, it will be assumed that the fault occurrence process is exponentially distributed. Let q' denote the failure rate of a hypothetical unit without any error detection mechanism, provision of these mechanisms will increase the failure rate (due to the presence of extra material) by a multiplicative factor denoted b : the actual failure rate of one unit is thus q = b q'.

d will denote the rate of activation of the error processing mechanisms in the safety system (i.e. 1/d is the detection latency). A detection efficiency (coverage)p (BOU 69, ARN 72) will be associated with this process. No a priori information exists concerning the distribution of this process; however, if d is significantly greater than the other rates, the exponential assumption is valid (LAP 81).

Let 1/a be the mean error latency time at the system level, i.e. the time elapsed before the activation by the process of a latent error in A' or A". We do not have any information concerning the corresponding latency rate. The impact of a non exponential distribution has been studied using the start-small approach in (ARL 85) and it results that the use of an exponential distribution is sufficiently representative.

The mean restoration time is denoted 1/m; some previous studies (APO 77, HEL 80, LAP 80) have shown the influence of imperfect repair, this has been accounted for by means of parameter r, the restoration efficiency.

For the monitoring systems, external solicitations correspond to accidental anomalies in the process and thus the exponential assumption for the interval between incident occurrence seems reasonable, this rate will be denoted s.

Incident processing time will be denoted 1/g, h represents the proportion of type T1 incidents and v the maintenance rate of the process following a type T2 incident. When processing an incident, the environmental constraints can be (depending on the type of monitored system) much more severe than in the absense of incidents; this leads to the introduction of a failure rate increase factor k during incident processing which can vary, according to HDBKS standard, from 1 to 50.

Remark: If x denotes an efficiency or a coverage factor, its complement 1-x will be noted x.

The order of magnitude of the different rates is given in figure 4.

Rates	Label	Range	Typical value (h^{-1})
Fault occurrence	q'	one fault every year	10^{-4}
Error activation	a	$100 \leq a/q' \leq 10000$	$10^{-2} \leq a \leq 1$
Error detection	d	$10 \leq d/a \leq 1000$	$10^{+2} \leq d \leq 10^{+4}$
Restoration	m	$100 \leq m/q' \leq 10000$	$10^{-2} \leq m \leq 1$
Incident occurrence	s	$1 \leq s/q' \leq 100$	$10^{-4} \leq s \leq 10^{-2}$
Incident processing	g	$g/s \geq 10$	$10^{-1} \leq g \leq 10^{+4}$
Process maintenance	v	$10 \leq v/s \leq 100000$	$10^{-2} \leq d \leq 1$

Figure 4

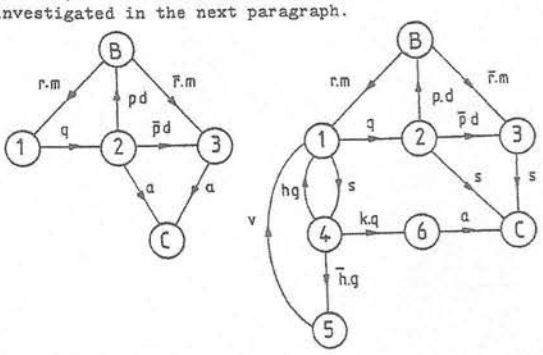
IV. EVALUATION OF SIMPLEX AND DUPLEX SAFETY SYSTEMS

A. Simplex Safety Systems

1) The Detection Process

The detection process is characterized both by the error detection latency (rate d) and the associated efficiency (p). Transition graphs a and b of figur 5 respectively present the models for the control and

monitoring safety systems deduced from the general model of figure 3 when only one fault occurrence is accounted for in the system. This choice has been made in order to enhance the impact of the detection process; the relaxation of this assumption will be investigated in the next paragraph.



a) Control System b) Monitoring System

Figure 5

Based on the procedure presented in section III, approximate closed-form expressions have been derived for the measures considered. The table of figure 6 gives the value of $(MST)^{-1}$ and UAC.

	CONTROL SYSTEM	MONITORING SYSTEM
1/MST	$q\left(\frac{a}{d} + \bar{p} + p\,\bar{r}\right)$	$q\left(\frac{s}{d} + \bar{p} + p\,\bar{r} + k\,\frac{s}{g}\right)$
UAC	$\frac{p\,q}{m}\left(1 - \frac{a}{d}\right)$	$\frac{p\,q}{m}\left(1 - \frac{s}{d}\right)$

Figure 6

On top of accounting for the orders of magnitude of the parameters of the model (figure 4), the approximations given in figure 6 have been derived considering that:

- for a control system: $a/d \ll 1$,
- for a monitoring system : $s/d \ll 1$,

which correspond to fairly natural conditions for a safety system if one wants to minimize the risk of latent errors.

These results show that the relative value of a and d (resp. s and d) which represent the rates of the error activation (resp. of the external solicitation) and of the detection process, has no impact on UAC under the above assumptions for both types of safety systems.

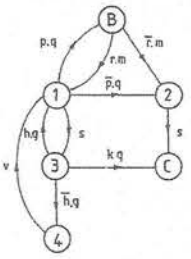
However, in the case of $(MST)^{-1}$, it appears that the ratio a/d (resp. s/d) acts at the same level as the complement of the efficiencies p and r. Accordingly, the impact of the rate of the detection process would be negligible only when :

- for a control system: $\frac{a}{d} \ll \bar{p} + p\,\bar{r}$,
- for a monitoring system: $\frac{s}{d} \ll \bar{p} + p\,\bar{r} + k\,\frac{s}{g}$.

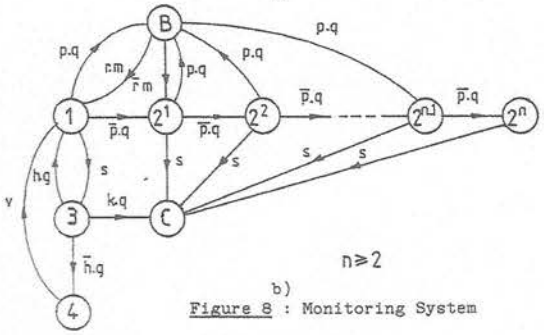
Due to the respective orders of magnitude, the most difficult condition to fulfill is that of the control system case. This point is of prime importance with respect to the implementation of recovery and restoration strategies when redundancy is provided at the system level. These aspects will be investigated in paragraph B, by considering a duplex structure. Figure 7 shows the respective influence of p and a/d, which confirms the opposite role of d and $\bar{p}$ on safety.

In the sequel, we will consider only systems for which these conditions are verified and thus, only the efficiency aspect of the detection process will be accounted for. Also, due to the respective orders of magnitude of a and (k q), state 6 in figure 5 can be deleted.

Accordingly, the model of the monitoring system of figure 5-b can be simplified as shown in figure 8-a.



a)



b)

Figure 8 : Monitoring System

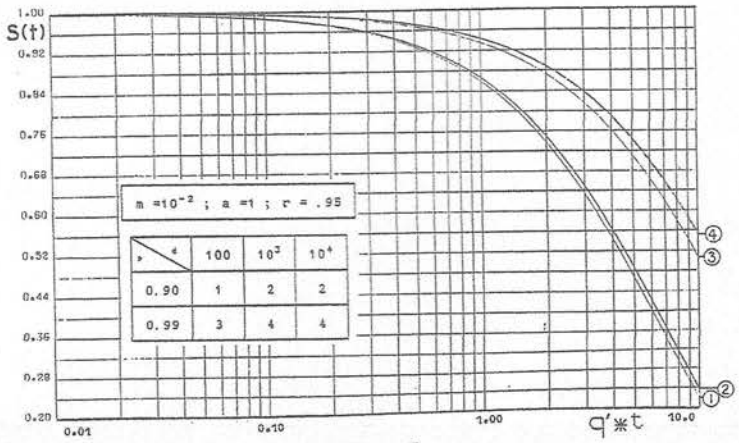


Figure 7

For sake of conciseness, we do not show explicitely here the simplified model for the control system since, due to the fact that the error activation process is inoperant when the system is error free (state 1), this graph represents a subgraph of the monitoring system graph (figure 8-a), where states 3 and 4 are deleted and rate s (incident occurrence) is changed to rate a (error activation).

2) Accumulation of faults in Sub-Class A"

When one assumes that n faults may be accumulated, the associated model for the monitoring system is shown in figure 8-b. The case n=1 is equivalent to the models of figure 8-a. The corresponding model for the control system is easily deduced as previously stated. The approximate values for the considered measures in the cases when n=1 and n≥2 are given in the tables of figure 9.

		CONTROL SYSTEM	MONITORING SYSTEM
n=1	$(MST)^{-1}$	$q\,(\bar{p} + p\,\bar{r})$	$q\,(\bar{p} + p\,\bar{r} + k\,\frac{s}{g})$
	UAC	$\frac{p\,q}{m}\,[\,1 - \frac{q}{a}\,(\bar{p} + p\,\bar{r})\,]$	$\frac{p\,q}{m}\,[\,1 - \frac{q}{s}\,(\bar{p} + p\,\bar{r})\,]$
n≥2	$(MST)^{-1}$	$q\,(\bar{p} + p\,\bar{r})(\,1 - p\,\frac{q}{a+q}\,)$	$q[(\bar{p} + p\,\bar{r})(1 - p\,\frac{q}{s+q}) + k\,\frac{s}{g}]$
	UAC	$\frac{p\,q}{m}\,[1 + \frac{q^2}{a(a+q)}\,(\bar{p} + p\,\bar{r})]^{-1}$	$\frac{p\,q}{m}\,[1 + \frac{q^2}{s(s+q)}\,(\bar{p} + p\,\bar{r})]^{-1}$

Figure 9

Both MST and UAC measures increase when n is varied from 1 to 2. Different consequences may be identified according to the type of safety system considered:

- control system: due to the fact that a >> q, the impact of n is negligible and it is sufficient to consider n=1 in subsequent studies,
- monitoring system: in this case, the impact of n is function of the relative values of s and q:
  . if s>>q, it is sufficient to consider that n=1,
  . if s and q are of the same order of magnitude, it is necessary to investigate further the case n≥2.

Figure 10 shows the variation of the safety S(t) for the monitoring system when the number of faults accumulated is modified. It appears that:

- a significant variation is observed for the considered values of the parameters when n is varied from 1 to 2,
- no variation is observed when considering n>2.

As a consequence, in what follows, modeling of the redundant structures will be carried out considering that n = 1, for a control system, and n = 2 for a monitoring system.

The other major results that can be extracted from the approximate closed-form expressions of figure 9 concern:

- the prominent role of the efficiency of the detection process for both measures in all cases: $(MST)^{-1}$ is proportional to p, while UAC is proportionnal to p; this quantifies the tradeoff already mentioned at the end of the paragraph B of section II,
- the fact that the impact of the restoration process on safety is restricted to its efficiency r, rather than on its rate m.
- for the monitoring system, the impact of a fault occurring during incident processing is of influence only when the quantity (k s/g) is comparable to $\bar{p}$ and $\bar{r}$.

All these points have been confirmed by numerical results obtained with the SURF program (MED 80); however, due to the space limitation, they are not recalled in the paper.

B. Duplex Safety Systems

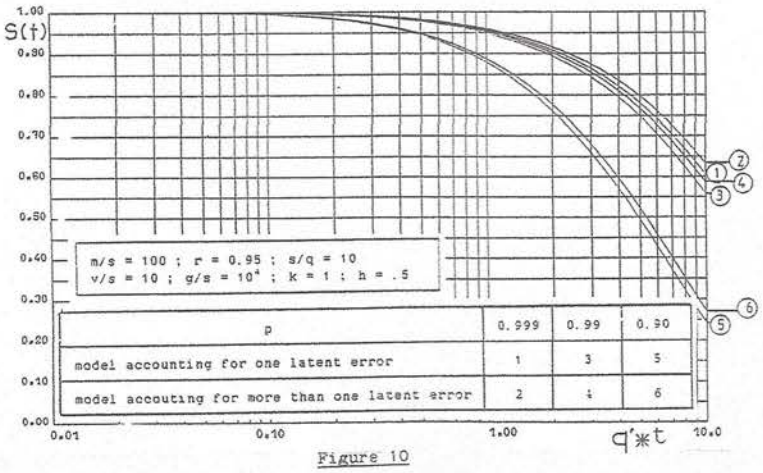
As a consequence of the discussion introduced in paragraph A.1 of this section, in the case of a duplex system, the provided redundancy will be used to improve the detection process (latency and efficiency) by comparing the results of the two units.

For the control system, in case of identified discrepancy, the system is preferably safely shut down and accordingly, no on-line restoration will be carried out (detection latency may be of the same order of magnitude as the activation latency in the case of the simplex structure). Redundancy is provided only to improve the detection process.

On the contrary, for the monitoring system, redundancy is used to provide fault-tolerance. Emphasis will be put on the recovery strategy consisting of the degradation of the system in a simplex structure; thus, on-line restoration will be considered.

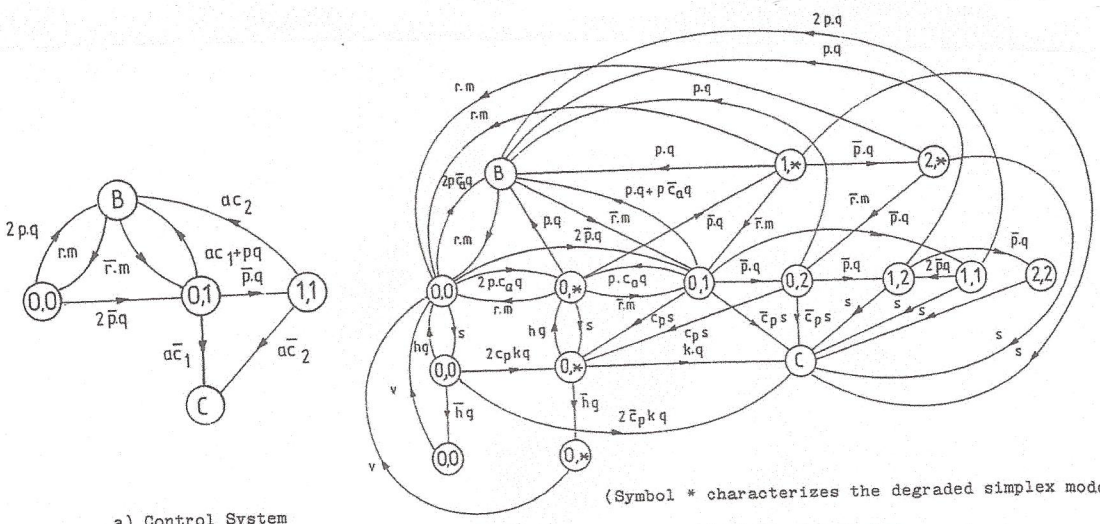
The introduction of the above different characteristics concerning the restoration strategy is due to the fact that, in practice, the control system is continuously activated, while the monitoring system is driven by isolated external solicitations. Based on these hypotheses, the models of the duplex structures considered for the control and monitoring systems are shown in figures 11 a and b respectively.

Two supplementary parameters have been added for each system. For the control system, c_1 and c_2 denote the efficiency of the comparison when respectively, only one or both units are faulty; as the system is continuously activated, failure of the comparison leads to the catastrophic interruption



p	0.999	0.99	0.90
model accounting for one latent error	1	3	5
model accouting for more than one latent error	2	4	6

Figure 10



a) Control System

(Symbol * characterizes the degraded simplex mode)

b) Monitoring System

Figure 11 (Label i = # of latent errors in one unit.)

state. For the monitoring system, c_a and c_p denote the coverage factor (detection and recovery efficiencies), respectively in the absence or presence of an incident; failure of these actions leads respectively to the benign and the catastrophic interruption states.

Figure 12 compares the approximate expressions that are obtained for both types of safety system in the case of the simplex and duplex structures. Results for the simplex structure are derived from figure 9 with n=1 for the control system and n≥2 for the monitoring system.

		CONTROL SYSTEM	MONITORING SYSTEM
SIMPLEX	$(MST)^{-1}$	$q(\bar{p}+p\bar{r})$	$q[(\bar{p}+p\bar{r})(1-p\frac{q}{s+q})+k\frac{s}{\varepsilon}]$
	UAC	$\frac{pq}{m}[1-\frac{q}{a}(\bar{p}+p\bar{r})]$	$\frac{pq}{m}[1+\frac{q^2}{s(s+q)}(\bar{p}+p\bar{r})]^{-1}$
DUPLEX	$(MST)^{-1}$	$2q[(\bar{p}+p\bar{r})(\bar{c}_1+\bar{p}\bar{c}_2\frac{q}{a})]\frac{a}{a+q}$	$2q[(\bar{p}+p\bar{r})\frac{2q}{s+2q}+k\frac{s}{\varepsilon}]\bar{c}_p$
	UAC	$\frac{2pq}{m}(1+\bar{p}[\frac{c_1}{p}+\frac{q}{a}(1+\frac{pc_2}{p})])\frac{a}{a+q}$	$\frac{2pq}{m}[\bar{c}_a+(\bar{p}+c_a\bar{r}p)\frac{q}{s+2q}]$

The difference in the restoration strategies considered in each case is confirmed by the results obtained when considering the variation of the UAC measure between the simplex and duplex structures:

- for the control system (off-line restoration), we observe an increase essentially due to the increase of the fault occurrence rate of the system,

- for the monitoring system (on-line restoration of the redundant unit), as could be expected, UAC is proportional to c_a and thus, significant improvement may result for realistic values of the coverage efficiency in absence of incidents.

It has to be noted that the penalization obtained for the control system may be viewed as a pessimistic limiting case (c_a = 0) of the recovery efficiency.

On the other hand, no major difference is observed between the two types of systems with respect to the safety aspect. It appears that in both cases of duplex structures, the safety measure is strongly dependent on the efficiencies c_1 and c_a respectively; however, it can be expected that safety would be significantly improved for usual values of these efficiencies when compared with the expressions obtained for the simplex structures.

Figure 12

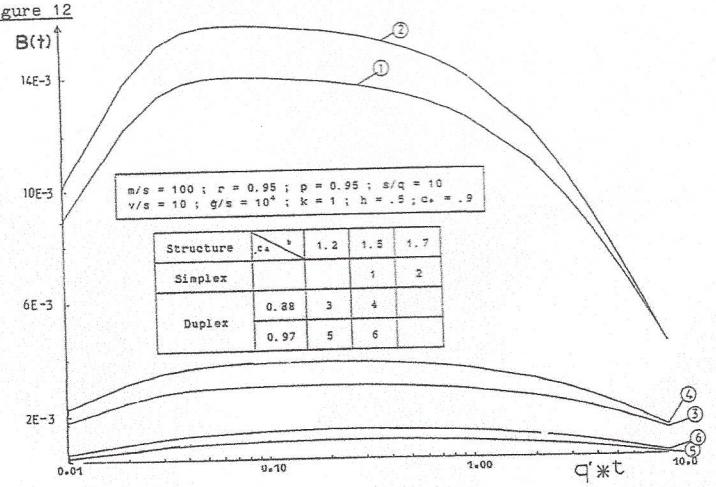


Figure 13

Although the results shown in Figure 12 provide a practical comparison basis, it has been found that the derivation of valid approximate expressions for UAC is a difficult task when the connectivity and the size of the model increases and thus numerical evaluation is required to complement these results.

Moreover, is it is worth noting that in practice the size and complexity of the computing units (and thus their associated fault occurrence rate) may vary when the structure of the system is changed. This point may be easily investigated by varying the ratio b between the actual fault rate q and that of a hypothetical unit q' (q = b q'). As an example, figure 13 shows the impact of this variation when computing B(t) for the duplex monitoring system. The values chosen for b reflect the fact that it is likely that the redundancy will be used to improve the detection process at the system level thus relaxing somewhat the constraints (speed and efficiency) imposed on the intra-unit detection process.

The above results have shown that uniform improvement of performance with respect to both measures may be obtained for monitoring systems when increasing the level of redundancy at the system level; this result has been verified as well for a triplex structure (LAP 80). On the contrary, a tradeoff appears in the case of control systems with respect to safety and availability between simplex and duplex structures. A better solution would be in this case to use a bi-duplex structure, as considered in (ARL 85), in order to allow actual on-line restoration and thus decrease the benign interruption time.

CONCLUSIONS

This paper was devoted to the study of two categories of safety systems:
- control systems: continuously activated,
- monitoring systems: driven by external solicitations due to incidents in the process.

Besides this discrepency they have a common characteristic: in case of a failure in the safety system, the process can be put into a prescribed safe shut-down state.

Two dependability measures have been defined in order to rate both the safety aspects and the impact of the safety system on the service delivered to the process (benign interruption).

Similar but specific models have been derived for each category and a closed-form expression of the dependability measures has allowed the identification of the more sensitive processes for each.

Some results are common to control and monitoring systems:
- prominent importance of the detection mechanisms,
- in the restoration process, repair efficiency has to be emphasized more than repair duration,

and some others are more specific:
- for control systems, the detection latency has to be reduced in order to enhance safety,
- for monitoring systems:
  . two latent errors in the system have to be accounted for,
  . failure in the process is more likely to occur due to latent errors developed in the absence of incidents than due to errors occurring during incident processing.

In view of these results we considered duplex structures where the redundancy is used, (i) as a means for detection in control sytems, (ii) for fault-tolerance in monitoring systems where on-line repair can be carried out.

Comparative evaluation of the dependability measures showed that the duplex structure constitutes a good tradeoff between safety and benign interruption time in the case of monitoring systems, and that for control systems this tradeoff is not achieved.

ACKNOWLEDGEMENT

The authors wish to thank Jean-Claude Laprie for his encouragement and inciting remarks during the gestation of this paper.

REFERENCES

APO 77 G.E.APOSTOLAKIS and P.P.BANSAL, "Effects of Human Errors on the Availability of Periodically Inspected Redundant Systems", IEEE Transactions on Reliability, vol. R-25, Aug. 1977, pp.220-225.

ARL 84 J.ARLAT, J.P.BLANQUART and J.C.LAPRIE, "On the Certification of Computer Systems: The EVE Project - Application to the Computerized Interlocking System", Proc. 4th Int. Conf. Reliability and Maintainability, Ferros-Guirec and Tregastel, France, May 1984, pp. 650-656, in French.

ARL 85 J.ARLAT and J.C.LAPRIE, "On the Dependability Evaluation of High Safety Systems", Proc. 15th Int. Symp. Fault-Tolerant Computing, Ann Arbor, Michigan, USA, June 1985, pp. 318-323; an extended version "Dependability Evaluation of Maintainable high Safety Control Systems" is available as LAAS-Research Report n° 85-198, August 1985.

ARN 72 T.F.ARNOLD, "The Concept of Coverage and its Effect on the Reliability Model of a Repairable System" Proc. 2nd Int. Symp. Fault-Tolerant Computing, Newton, Mass., USA, June 1972, pp. 200-204.

BOU 69 W.G.BOURRICIUS, W.C.CARTER and P.R.SCHNEIDER, "Reliability Modeling Techniques for Self-Repairing Computer Systems" Proc. 12th ACM National Conf., August 1969, pp. 295-309.

COS 81 A.COSTES, J.E.DOUCET, C.LANDRAULT and J.C.LAPRIE, "SURF: A Program for Dependability Evaluation of Complex Fault-Tolerant Computing Systems" Proc. 11th Int. Symp. Fault-Tolerant Computing, Portland, Maine, June 1981, pp. 72-78.

COX 68 R.E.COX and H.D.MILLER, The Theory of Stochastic Processes, Methuen, London, England, 1968.

HEL 80 B.E.HELVIC, "Periodic Maintenance on the Effect of Imperfectness", Proc. 10th Int. Symp. Fault-Tolerant Computing, Kyoto, Japan, October 1980, pp. 204-206.

HOW 71 R.A.HOWARD, Dynamic probabilistic systems, Volume 1: Markov Models, John Willey & Sons, Inc. New York, 1971.

HOP 78 A.L.HOPKINS, T.BASIL SMITH, and J.H.LALA, "FTMP - A Highly Reliable Fault-Tolerant Multiprocessor for Aircraft", Proc. IEEE, VOL. 66, No 10, October 1978, pp. 1221-1239.

LAP 75 J.C.LAPRIE, "Prediction of the Dependability and Architecture of Maintainable Real Time Digital Structures", State Thesis Diss., Paul Sabatier University, Toulouse, France, June 1975, in French.

LAP 76 J.C.LAPRIE, "On Reliability Prediction of Repairable Redundant Digital Structures", IEEE Transactions on Reliability, vol. R-25, October 1976, pp. 256-258.

LAP 80 J.C.LAPRIE and K.MEDHAFFER-KANOUN, "Dependability Modeling of Safety Systems", Proc. 10th Int. Symp. Fault-Tolerant Computing, Kyoto, Japan, October 1980, pp. 245-250; an extended version appeared in Microelectronics and Reliability, vol. 22, no 5, 1982, pp. 997-1026.

LAP 81 J.C.LAPRIE, A.COSTES and C.LANDRAULT, "Parametric Analysis of 2-Unit Redundant Computer Systems with Corrective and Preventive Maintenance", IEEE Transactions on Reliability, vol. R-30, June 1981, pp. 139-144.

LAP 85 J.C.LAPRIE, "Dependable Computing and Fault-Tolerance: Concepts and Terminology", Proc. 15th Int. Symp. Fault-Tolerant Computing, Ann Arbor, Michigan, June 1985, pp. 2-11.

MED 80 K.MEDHAFFER-KANOUN, "Dependability Evaluation of Safety Systems, Application to the Control System of Extra-High-Voltage substations", Docteur Ingénieur Thesis, Toulouse National Polytechnic Institute, July 1980, in French.

PAG 80 A.PAGES and M.GONDRAN, Systems Reliability, Eyrolles, Paris, France, 1980, in French.

WEN 78 J.H.WENSLEY, L.LAMPORT, J.GOLDBERG, M.W.GREEN, K.L.LEVITT, P.M.MELLIAR-SMITH, R.E.SHOSTAK and C.B.WEINSTOCK, "SIFT: Design and Analysis of a Fault-Tolerant Computer for Aircraft Control", Proc. IEEE, VOL.66, Oct. 78, pp.1240-1255.