Evaluation of bus and ring communication topologies for the DELTA-4 distributed fault-tolerant architecture

Karama Kanoun, David Powell

Dependability Evaluation of Bus and Ring Communication Topologies for the Delta-4 Distributed Fault-Tolerant Architecture*

Karama Kanoun
kanoun@laas.fr

David Powell
dpowell@laas.fr

LAAS-CNRS
7, avenue du Colonel Roche, 31077 Toulouse (France)

Abstract*

The Delta-4 distributed fault-tolerant architecture aims to provide dependability in locally-distributed systems that are open to heterogeneity and off-the-shelf hardware. Reliability and availability are provided by means of replicating computation across distinct nodes interconnected by a local area network. The achievable dependability is therefore limited by that of the underlying communication system. This paper reports a study of the dependability of the various communication topologies that can be used in order to construct a Delta-4 system. Single and dual bus and ring configurations are possible (based on 802.4, 802.5 and FDDI standards); the paper gives closed-form expressions for the reliability and availability of each topology when repair is taken into account. It is shown that the dimensioning parameter in the dependability of the communication system is the coverage of the self-checking mechanisms built into the network attachment controllers.

Introduction

The Delta-4 distributed fault-tolerant architecture is the result of a 5-nation 13-partner project carried out in the framework of the European ESPRIT programme [3,17, 20]. The major aim of the architecture is to provide dependability in locally-distributed systems that are open to heterogeneity and off-the-shelf hardware. It is necessary that the users of the Delta-4 system be able to justifiably place their confidence in the architecture. Consequently such an architecture must undergo extensive *validation* both from the *verification* and the *evaluation* viewpoints.

Verification is that part of the validation activity aimed at removal of design and implementation faults. In the Delta-4 project, verification is carried out at two levels:

- verification of the design of the communication protocols to detect and correct the design errors [2],
- verification of the implementation by means of injection of hardware faults to verify the effectiveness of the architecture's self-checking and fault-tolerance mechanisms [1].

Dependability evaluation is that part of the global activity of validation that pertains to fault-forecasting, i.e. the estimation of the presence, the creation and the consequence of faults. In the Delta-4 project, dependability evaluation is also carried out at two levels:

- modelling and evaluation of dependability measures of Delta-4 architecture configurations taking into account the nature of the different elements (e.g. fail-silent or fail-uncontrolled NACs, replication domain of the different components, replication techniques, reconfiguration possibilities, repair policies, …)
- evaluation of software reliability through the application of reliability growth models.

The reported work deals with dependability modelling and evaluation. The objective of this activity is to provide the users with a quantified assessment of the amount of dependability that the architecture provides, i.e. the *degree* by which they can justifiably rely on the architecture. It is very difficult to directly establish correct global models so a progressive method is used. It is necessary to establish a global evaluation strategy in terms of inter-connected sub-models. After the study of the sub-models, it is necessary to aggregate them and study the global model. This organization in sub-models should also give some early feedback about the design of the different components included in the sub-models. We have defined three main objectives associated to three different sub-models:

- modelling and evaluation of the communication system,
- extension of the communication architecture model to include the host-resident management information base,
- establishment of the models of some target applications and evaluation of their dependability in

* This work was partially financed under the CEC ESPRIT programme project n°2252, Delta-4: Definition and Design of an open Dependable Distributed architecture.

order to provide a framework for quantifying the dependability offered by particular configurations of the architecture.

This paper is focussed on the dependability modelling and evaluation of the communication system. Various communication topologies are considered (802.4 token bus, and 802.5 and FDDI token rings); for each of them, a single and a dual configuration is modelled. Two measures of dependability are evaluated: the reliability and the asymptotic unavailability. The aim is to compare these various design solutions, to define some essential parameters and to study their effects on system dependability.

The paper is composed of three sections. In the first section a brief description of the Delta-4 architecture is given. The second section presents the modelling method. The third section is devoted to modelling and evaluation of the different communication systems.

1. Delta-4 architecture

This section recapitulates the basic aspects of the architecture that is to be modelled. More details can be found in [3, 5, 17, 20]. The Delta-4 architecture supports three basic techniques for coordinating replicated computation, called: active, passive and semi-active replication:

- In active replication, all replicates process all input messages concurrently so that their internal states are closely synchronized — in the absence of faults, outputs can be taken from any replicate.
- In passive replication, only one of the replicates (the primary replicate) processes the input messages and provides output messages — in the absence of faults, the other replicates (the standby replicates) do not process input messages and do not produce output messages; their internal sates are however regularly updated by means of checkpoints from the primary replicate.
- Semi-active replication can be viewed as hybrid of both active and passive replication; only one replicate (the leader replicate) processes all input messages and provides output messages — in absence of faults, the other replicates (the following replicates) do not produce output messages; their internal state is updated either by direct processing of input messages or, where appropriate, by means of "mini-checkpoints" from the leader replicate.

Active replicates can be either fail-silent (any output sent by any replicate of the group can be assumed to be a correct value) or fail-uncontrolled (the set of outputs must be considered as a whole so that the value errors and unexpected outputs may be masked). Since only one replicate at a time can send outputs when passive or semi-active replication is employed, these techniques require replicates to be fail-silent. Active replication has the potential disadvantage of requiring computation to be deterministic. When computation cannot be assumed to be *a priori* deterministic in the absence of faults, then passive or semi-active replication must be employed and it is thus necessary to assume that hosts are fail-silent. This is possible when the coverage of the host self-checking mechanisms[1] is commensurate with the dependability objectives of the supported application.

1.1 Hardware

In order to be able to use standard local area networks instead of resorting to the costly interconnection topologies normally required to accommodate arbitrary (or uncontrolled) failures, each host is connected to the underlying local area network via a *fail-silent* network attachment controller or NAC (figure 1).

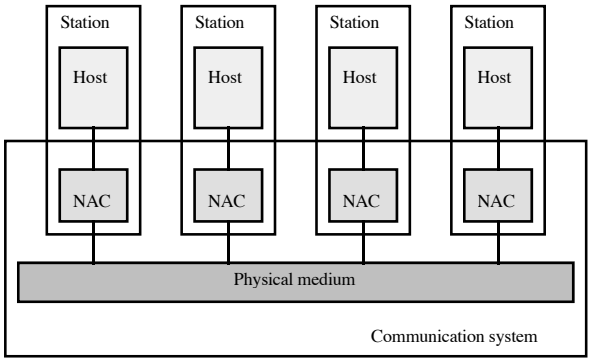


Figure 1: Delta-4 hardware architecture

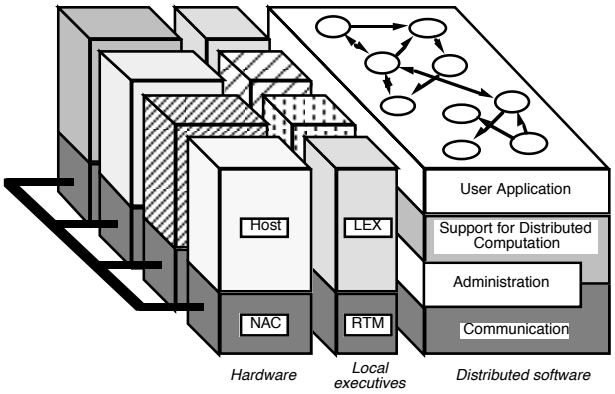
NACs are implemented using hardware self-checking techniques in order to substantiate the fail-silent assumption. The important consequence of this split in failure mode assumptions between host and NAC is that, even when a (fail-uncontrolled) host forwards erroneous data to its NAC, the latter will either process this data in a consistent manner or simply remain silent; the possibility of forwarding the data *inconsistently* to multiple destinations is effectively removed.

At this stage of the project two sorts of NACs are available: type 1 (with *limited* self-checking mechanisms) and type 2 NACs (with *extended* self-checking mechanisms); both of them will be considered for system modelling.

1.2 Software

Figure 2 provides an abstract view of the overall Delta-4 architecture.

[1] Or, equivalently, the conditional probability that, when a host fails, it fails by going silent..



NAC: Network Attachment Controller
LEX: Local EXecutive (of host)
RTM: Real-Time Monitor (of NAC)

Figure 2: Abstract view of the de Delta-4 architecture

- The left-hand "slice" of the diagram recapitulates the hardware architecture discussed in the previous section.
- The middle slice of the figure represents the local executives residing on the host and NAC hardware. The local executives (LEXes) resident on the hosts are shown shaded differently (like the hosts) in order to underline that heterogeneous host hardware and executive software may be accommodated. In practice, the present implementations all use different flavours of UNIX[1].
  The right-hand slice of the figure represents the *distributed Delta-4 software* which can be represented in four parts:
  - the distributed user application software represented as a set of "software components" (logical units of distribution) that communicate by messages (only),
  - the host-resident infrastructure for support of distributed computation,
  - the computation and communication administration software (executing partly on the host computers and partly on the NACs),
  - the communication protocol software (executing on the NACs).

1.3. Communication system

Within the Delta-4 system, it is possible to implement various kinds of local area networks (LANs), they differ by their performance characteristics, their dependability as well as their price [22]. Before system design, the user can choose one of the possible LANs depending on the nature of the application (performance and dependability aspects) and on the price of the LAN.

1 UNIX is a registered trade-mark of AT&T.

In each architecture, every host possesses a NAC that interfaces the host and the underlying media. The couple host-NAC form a node or a station, and the set of the NACs with the underlying media constitute the communication system .

There are thus two essential aspects to be taken into account in the models: the communication topology and the nature of the NACs (type 1 or type 2, cf. section 1.1).

The communication topologies presently taken into account in Delta-4 are: the 802.4 token bus [6], the 802.5 [7], and FDDI token rings [8]. Each of them can use a single (non redundant) medium or duplicated media. These various communication topologies and the associated fault-tolerance strategies are summarized in the following.

802.4 token bus

The Delta-4 implementation of the token bus includes the possibility of dual physical media (figure 3). If an inter node physical link of the bus fails then the entire bus fails due to impedance mismatch[2]. If the station fails in a controlled fashion, then the station is disconnected and the other stations remain connected. The stubs are treated as integral parts of their associated physical media.

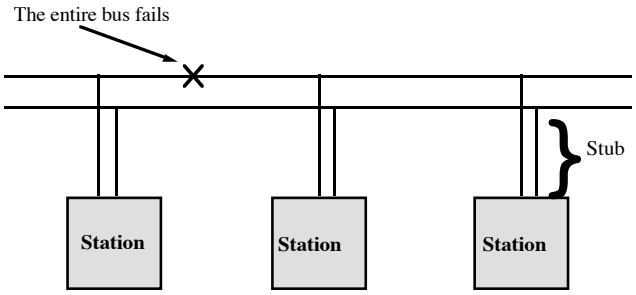


Figure 3: Dual token bus 802.4 architecture

802.5 token ring

Fault-tolerance in the 802.5 token ring is based on two principles: by-pass switches (in wiring concentrators) and dual counter-rotating rings.

Figure 4 illustrates the by-pass switch concept in a single (non-redundant) ring. The switches are electric relays that are physically localized in *wiring concentrators* (WC). The switches can be separated into multiple wiring

2 Partial failures of busses are possible in certain (rare) conditions (shorts/opens inside stub connectors, short message frames over long-haul busses,...). However, we choose not to include this possibility in the dependability models since assuming complete failure leads to a lower bound on the achievable dependability (i.e. the models are pessimistic). Furthermore, exploitation of partial connectivity after a bus failure would entail considerable extra communication protocol complexity.

concentrators with links (of the medium) between each WC and one or more stations connected to each WC. This is illustrated by figure 5.

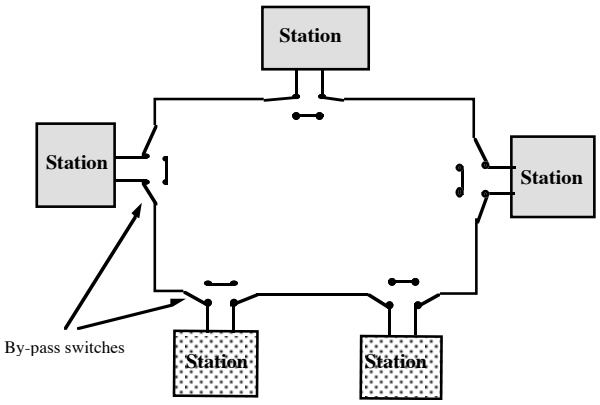


Figure 4: Abstract view of the non-redundant ring.

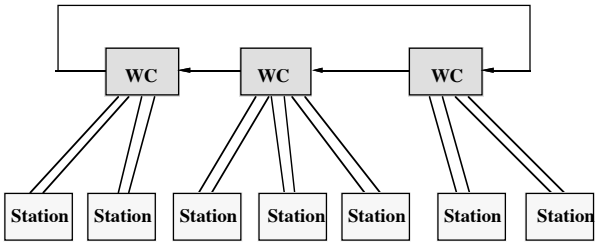


Figure 5: Single ring with multiple wiring concentrators

A failure of an inter-WC link causes the complete failure of the communication system. But a controlled failure between a host and a WC causes the disconnection of the host and the communication system does not fail. In this case there is no reconfiguration (other than the closing of by-pass switches and the possible re-election of an active monitor) and the WCs are passive. They have no autonomous power; each NAC powers the relay by which it is connected. From a reliability viewpoint, some components of the WCs are included in the ring, and the others are included in the connection with the station. The latter can be thus included in the NAC study. The ring itself and the components of the WCs that are included with it constitute the hard-core of the communication system. However, being entirely passive, the failure rates of these components should not be very high.

Dual ring without WCs

In the case of the non-redundant ring, the physical medium of the ring is a hard-core. It must not fail or the complete communication system fails. To solve this problem, a dual counter-rotating ring can be used. When a station or a link of the physical medium fails there is a reconfiguration of the network; the incriminated station is disconnected and the dual ring becomes a single ring.

Figure 6 illustrates the dual ring 802.5 architecture and the reconfiguration of the ring when a station fails. A similar reconfiguration occurs when a link between a pair of stations fails, but in the latter case the ring is reconfigured without any station loss.

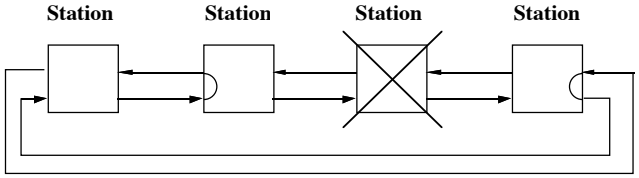


Figure 6: Dual ring - example of reconfiguration by counter-rotating ring

By-passing and counter-rotating rings

Figure 7 illustrates the 802.5 architecture in the case of a combination of by-passing and counter-rotating rings.

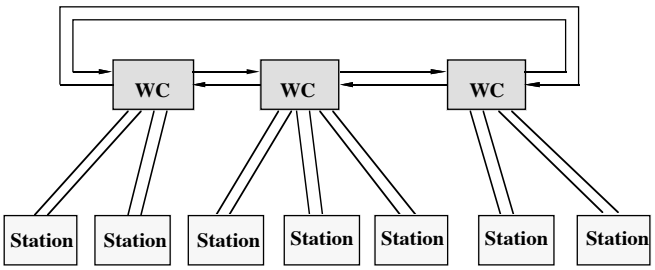


Figure 7: Dual token ring physical architecture

These two techniques are complementary and their combination gives the facilities and the advantages of both. Note however, that when a dual ring is used, the WCs are active because of the implied necessity for "intelligent" reconfiguration. System reconfiguration possibilities are summarized in figure 8.

Event	Treatment
Failure of a link in the dual ring	Counter-rotating ring
Failure of a link between a NAC and its WC or failure of a NAC	NAC by-passing by the WC.
Failure of a WC	Local by-passing of the WC or use of counter-rotating ring

Figure 8: Dual token ring reconfiguration possibilities

FDDI token ring

An FDDI ring has the same basic topology as an 802.5 ring but the physical links of the ring are optical fibres and the relays of the WCs may be electrical or optical. The optical relay solution is more difficult (from a

dependability analysis viewpoint) and in a first step we will assume the electrical relay solution. In this case, the FDDI model differs from the 802.5 model only by the values of the model parameters.

In section 3, single and dual bus and ring topologies are studied without further reference to particular standards for each topology.

2. Modelling and evaluation method

Several methods for dependability evaluation can be distinguished: reliability block diagrams, fault-tree (or event-tree) analysis and state diagrams. The main advantages of the latter are:

- their ability to account for the stochastic dependencies which result for instance from maintenance and solicitation processes, or from simultaneous consideration of several classes of faults,
- various dependability measures can be derived from the same model.

A state diagram is a graph in which the nodes represent the states of the system and the edges the elementary events leading to system transition from one state to another. The system model may be viewed as a representation of (i) the modifications of the system structure resulting from the events likely to affect system dependability (fault-error-failure process, maintenance actions) and of (ii) other events of interest (e.g. solicitation process corresponding to user requests). When the elementary events can be considered as exponentially distributed (constant failure rates) the state diagram corresponds to a time-homogeneous Markov chain. Markov modelling is well adapted to dependability evaluations in which different possible structures are compared during the design phase (or during operational life if the architecture of the system allows this possibility) in order to select the "best" one.

Regarding the validity of the failure rate constancy assumption, the following situations have to be distinguished:

- it is a realistic assumption for accidental events such as physical failures,
- it has been shown that it is a good assumption for maintenance process if the mean maintenance duration is small compared to the mean time to failure [12],
- for the other cases the method of stages (fictitious states) [4, 21] has to be used to "simulate" non-constant failure rates.

Our recommendation (which has been put into practice for several years in the dependability group at LAAS) is the following (see, for example [12, 14]:

- consider all the hazard rates as constant and derive a Markov chain,
- perform sensitivity studies using the device of stages for those rates which are considered to be non-constant, starting with one fictitious state for each non constant rate and stopping when addition of more states is of non-perceptible influence.

This recommendation stems from the following facts:

- a model is always an approximation of the real word and this approximation has to be globally consistent,
- when modelling phenomena stochastically, the first moment generally determines the order of magnitude, the further moments bring in refinements; an exponential distribution can be seen as the distribution corresponding to the knowledge of the first moment only.

2.1. Measures of dependability

When evaluating the dependability of a communication system interconnecting more than just two stations, one is immediately faced with the problem of defining a suitable measure of dependability and/or of how the operational states or *up-states* of the communication are defined. For instance, the up-states could be defined to be those in which k out of the total number of stations can communicate. Alternatively, a weighted dependability measure that encompasses all degraded states of the communication system could be used (see, for example, the *cooperability* measure defined in [19]).

Since the final aim of modelling the Delta-4 fault-tolerant architecture is to quantify the dependability of an application composed of redundant software components executing on different hosts (cf. section 1), it was desired that the communication system dependability be defined in such a way that it can be incorporated into the overall application dependability measure as a series "hardcore" term. It was therefore felt that a suitable definition of the up-states of the communication system is that set of states in which all operational stations can communicate, as used in [16]. However, unlike [16], we are interested in evaluating the dependability of the communication system in a *maintainable* environment, i.e. taking repair operations into account. Two measures of interest are taken into account [13]: reliability and availability.

2.2. Model processing

Dependability can be evaluated from the Markov chain either directly — when the models are not very complex — or using a dependability evaluation tool such as the SURF program [15] when the models are more complex. Tools do not usually give analytical expressions of the

different measures due to the complexity of these models. However, when the non failed states of a Markov chain constitute an irreducible set (i.e. the graph associated with the non absorbing states is strongly connected), it can be shown that the absorption process is asymptotically a homogeneous Poisson process; approximate expressions of the measures of dependability can be obtained through the "equivalent" failure rate [18].

The aim of this technique is to transform the initial Markov chain to a reduced Markov chain made up of two states: the non failed-state and the failed state, the asymptotic transition rate, λ_C, from the non failed-state to the failed state is called the equivalent failure rate. This transition rate is obtained directly from the initial chain and is given by:

$$\lambda_C = \sum_{\substack{\text{paths from} \\ \text{initial state (I)} \\ \text{to failed state}}} \left\{ \frac{\prod \text{transition rates of the considered path}}{\prod_{\substack{\text{states in path} \\ \text{(except I)}}} [\sum \text{output rates of the considered state}]} \right\} \quad (1)$$

Reliability is then given by:

$$R(t) = \exp(-\lambda_C t) \quad (2)$$

and the asymptotic unavailability is equal to:

$$UA = \frac{\lambda_C}{\mu} \quad (3)$$

where μ is the repair rate from the failed state.

Since the Delta-4 system is intended for applications in which repair is possible, the associated chains are generally strongly connected so this approach will be adopted: the different sub-systems will be evaluated through their equivalent failure rates. For sake of simplicity, the equivalent failure rate of the system will be termed the system failure rate.

3. Communication system modelling and evaluation

This section details the models and the results concerning communication system dependability. For each architecture, a Markov model taking into account the various events leading to system failure or system reconfiguration is first given; expression of the equivalent failure rate and unavailability are then derived and numerical processing is finally carried out. More details can be found in [10].

3.1. NAC modelling

Even though there are two basic sorts of NACs: fail-uncontrolled NAC is, in reality, a NAC with limited self-checking and fail-silent NACs, in reality, a NAC with extended self-checking; modelling will be carried out in the same manner.

The two sorts of NACs (type 1 and type 2, cf. section 1.1) differ by the extent of their self-checking techniques and thus have different parameters (failure rates and coverage factors) and different advantages. The choice of one sort of NAC or the other depends on the global architecture, and their effect is described by the model through the numerical values of their parameters.

The coverage factor of the NAC is defined as the probability of correct passivation of the NAC after its failure.

Let λ_{Ni} and p_{Ni} denote respectively the failure rate and the coverage factor for the type i NAC (i=1 or 2). We should thus have $p_{N2} > p_{N1}$ but inevitably $\lambda_{N1} < \lambda_{N2}$. From a practical viewpoint, we will consider only one pair of parameters for the NACs (p_N, λ_N), and a sensitivity study with respect to these parameters will be carried out.

3.2. Single bus

The communication system is based on a single bus on which n stations are connected. It is assumed that any failure of the bus leads to communication system failure.

The maintenance policy is as follows:

- a covered failure (of the NAC or of the medium) does not affect service delivery, moreover the repair of such a failure does not need service interruption,
- after a non-covered failure of the NAC or of the medium service delivery is interrupted, repair of all the failed elements is carried out before service is resumed.

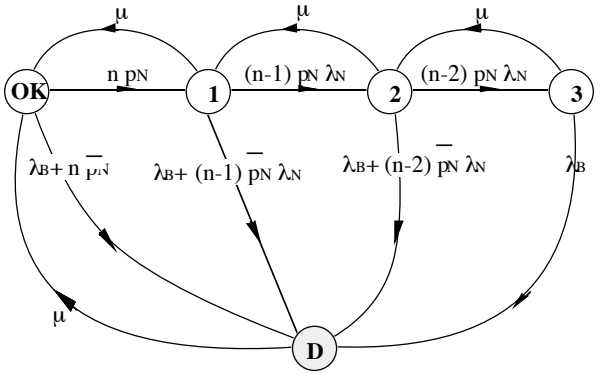
The Markov chain is given in figure 9 where only three consecutive covered NAC failures are considered and with:

- λ_N: the failure rate of the NAC,
- p_N: the coverage factor of a NAC and $\overline{p}_N = 1 - p_N$,
- n: the number of stations (of NACs here),

- λ_B: the failure rate of the bus,
- μ: the repair rate.

The communication system failure rate derived from this model is:

$$\lambda_C = \lambda_B + n\cdot \overline{p}_N\cdot\lambda_N + n\cdot p_N\cdot\frac{\lambda_N}{\mu}\cdot[\lambda_B+(n-1)\ \overline{p}_N\cdot\lambda_N]$$
$$+n\cdot(n1)\cdot\left(p_N\cdot\frac{\lambda_N}{\mu}\right)^2\cdot[\lambda_B+(n-2)\cdot\ \overline{p}_N\cdot\lambda_N]$$
$$+ n\cdot(n-1)\cdot(n-2)\cdot\left(p_N\cdot\frac{\lambda_N}{\mu}\right)^3\lambda_B \qquad (4)$$



State	Signification
OK	Initial state
1	Covered failure of a NAC
2	Covered failures of two NACs
3	Covered failures of three NACs
D	System failure due either to the failure of the bus or to a non-covered failure of a NAC

Figure 9: Markov model of the communication system with a single bus.

Using the fact that $\lambda_B/\mu<<1$ and $\lambda_N/\mu<<1$ and considering only the first and second order terms leads to: $\lambda_C \approx \lambda_B + n$

$$\overline{p}_N\lambda_N + n\cdot p_N\cdot\frac{\lambda_N}{\mu}\cdot[\lambda_B+(n-1)\ \overline{p}_N\cdot\lambda_N] \qquad (5)$$

These expressions show the prime importance of the coverage factor of the NAC on system dependability. It can also be seen that both the failure rate and the unavailability are limited by the failure rate and the unavailability of the bus; this is not surprising since from dependability viewpoint, the bus is in series with the NACs.

Numerical processing

At the moment, "reasonable" figures have been chosen and the values of these parameters will be used until such time as real values become available. The following parameter values are considered:

- λ_N: a value of 10^{-4} / h has been taken as a reference (i.e. 1 failure per year),
- n: the number of stations — fixed (arbitrarily) at 15,
- λ_B: a value of 2 10^{-5} / h has been taken; this corresponds to 1 failure per 5 years,
- μ: a mean repair duration of 2 hours has been adopted, leading to $\mu = 0.5$ /h.

Figure 10 plots λ_C/λ_N versus p_N and figure 11 gives some values of λ_C and of the unavailability, UA, expressed in hours per year. These figures confirm the fact that system dependability is limited by the dependability of the bus.

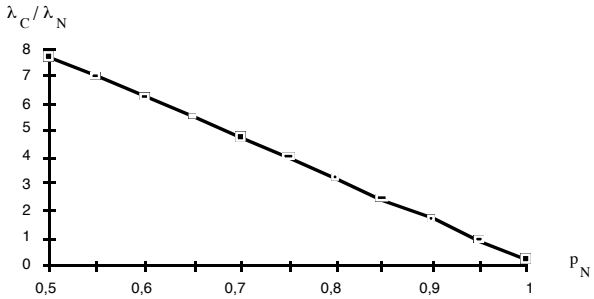


Figure 10: Numerical application for the single bus[1]

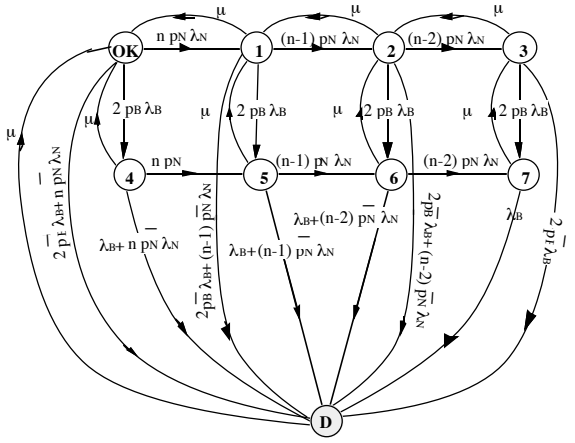
p_N	0.7	0.75	0.8	0.85	0.9
λ_C	4.7E-04	4.0E-04	3.2E-04	2.5E-04	1.7E-04
UA	8.24	6.93	5.61	4.30	2.98
p_N	0.95	0.99	0.999	0.9999	1
λ_C	9.5E-05	3.5E-05	2.2E-05	2.0E-05	2.0E-05
UA	1.67	0.61	0.377	0.353	0.351

Figure 11: λ_C (in failures/hour) and UA (in hours per year) versus p_N for the single bus[4]

3.3. Dual bus

The same architecture as previously is considered but the single bus is replaced by a dual bus. A coverage factor of the bus, p_B, has to be introduced for this architecture which is defined as the probability of correct reconfiguration of the dual bus into a single bus in case of failure of one of the buses. The repair policy is also the same as previously. However, in case of one or several covered NAC failures, followed by a covered failure of the medium, repair priority is given to the medium. Figure 12 gives the corresponding Markov model.

[1] Figures 10 and 11 also apply to the single ring; see section 3.4 below.



State	Signification
OK	Initial state.
1	Covered failure of a NAC, the medium is OK
2	Covered failure of 2 NACs, the medium is OK
3	Covered failure of 3 NACs, the medium is OK
4	Covered failure of the medium: it is no longer fault-tolerant
5	Covered failure of the medium and of a NAC
6	Covered failure of the medium and of 2 NACs
7	Covered failure of the medium and of 3 NACs.
D	System failure.

Figure 12: Markov model of the system with a dual bus

The failure rate is given by:

$$\lambda_C = 2\,\bar{p}_B\cdot\lambda_B + n\cdot\bar{p}_N\cdot\lambda_N + 2\,p_B\cdot\frac{\lambda_B}{\mu}\cdot[\,\lambda_B + n\cdot\bar{p}_N\cdot\lambda_N]$$

$$+ n\cdot p_N\cdot\frac{\lambda_N}{\mu}\cdot[\,2\,\bar{p}_B\cdot\lambda_B + (n-1)\cdot\bar{p}_N\cdot\lambda_N]$$

+

n

.

Erreur!

$$+ 6\cdot p_B\cdot\frac{\lambda_B}{\mu}\cdot n\cdot(n-1)\cdot\left(p_N\cdot\frac{\lambda_N}{\mu}\right)^2\cdot[\,\lambda_B + (n-2)\cdot\bar{p}_N\cdot\lambda_N]$$

$$+ 8\,n\cdot(n-1)\cdot(n-2)\cdot\left(p_N\cdot\frac{\lambda_N}{\mu}\right)^3\cdot\lambda_B\cdot\left[\bar{p}_B + p_B\cdot\frac{\lambda_B}{\mu}\right] \quad (6)$$

Using the fact that $\lambda_B/\mu \ll 1$ and $\lambda_N/\mu \ll 1$ and considering only the first and second order terms leads to:

$$\lambda_C \approx 2\,\bar{p}_B\lambda_B + n\,\bar{p}_N\lambda_N + 2\,n\,\frac{\lambda_N}{\mu}\,\lambda_B\,[\,p_N\bar{p}_B + p_B\,\bar{p}_N]$$

$$+ n(n-1)p_N.\bar{p}_N\,\frac{\lambda_N}{\mu}\,\lambda_N \quad (7)$$

It can be noted that, for the dual bus, dependability is directly related to the failure rate of non-covered failures of the bus and of the NACs. The coverage factors are thus of prime importance.

For numerical processing, the same parameter values are used. Figure 13 plots several curves of λ_C/λ_N versus p_N for $0.65 \le p_B \le 1$; even though these curves lead one to think that p_B has no influence due to the linear scale, it has a small influence which is masked by the high influence of p_N.

Since p_B has less influence than p_N it is more important to improve p_N than p_B to enhance system dependability. This is due to the fact that $n\cdot\lambda_N$ is higher than λ_B, and thus has more influence on the probability of system failure. p_B has more influence for high values of the NAC coverage factor, p_N.

It can also be noticed that the failure rate of the system can be lower than the failure rate of the bus for high coverage factors of the NAC only.

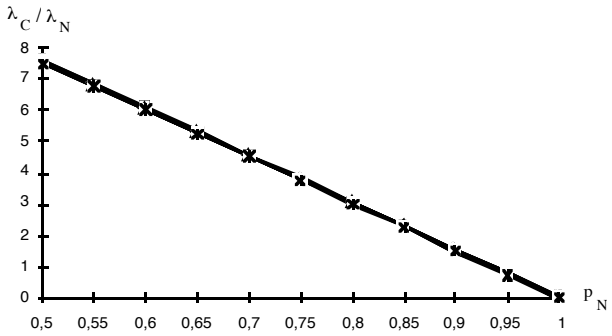


Figure 13:Numerical application for the dual bus

Figure 14 gives system unavailability versus p_N for p_B =0.9; for instance, when p_N increases from 0.85 to 0.95 unavailability is divided by 3.

p_N	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
λ_C	4.5E-04	3.8E-04	3.0E-04	2.3E-04	1.5E-04
UA	7.96	6.64	5.33	4.01	2.70
p_N	**0.95**	**0.99**	**0.999**	**0.9999**	**1**
λ_C	7.9E-05	1.9E-05	5.5E-06	4.2E-06	4.0E-06
UA	1.39	0.33	0.096	0.073	0.070

Figure 14: λ_C (in failures/hour) and UA (in hours per year) versus p_N for the dual bus

Comparison of figures 11 and 14 shows the slight improvement in system availability; for instance, system unavailability is 1.67 hours/year for the single bus with a NAC coverage factor $p_N = 0.95$, it is equal to 1.39 hours for the dual bus, i.e. doubling the bus leads to only a 17 minutes/year gain in availability. Note, however, that this improvement is closely related to the failure rate of the bus, λ_B (set equal to $2\ 10^{-5}$ / h).

The unavailability of the communication system with a single and a dual bus, versus the failure rate of the bus (λ_B) and for $p_N = 0.95$ and 1, is given in figure 15. When the coverage factors are less than 1, duplication does not necessarily lead to an appreciable improvement to system dependability, depending on the value of the bus failure rate. For instance, for $p_N = 0.95$, the improvement is noticeable (on the logarithmic scale of figure 15) only for $\lambda_B \geq 1\ 10^{-4}$ / h. This is due to the fact that when λ_B is low, dependability is conditioned by the NAC failures.

When the coverage factor of the NAC is equal to 1, duplication is worthwhile, for example:

- for $\lambda_B = 10^{-5}$/h, duplication of the bus decreases UA from 0.2 (11mn/year) to 1mn/year,
- for $\lambda_B = 10^{-4}$/h, unavailability is decreased from 1 h 45 mn to 11 mn / year,

which is a significant improvement.

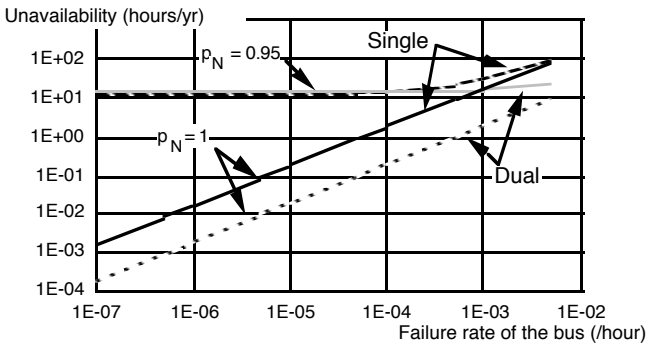


Figure 15: Communication system unavailability for the single and the dual bus, with $p_N = 0.95$ and 1.

3.4. Single ring

We will consider n stations connected to a single ring and we will apply the same repair policy as previously.The associated Markov model is the same as for the bus replacing λ_B by λ_R where the failure rate of the ring λ_R is given by: $\lambda_R = m \lambda$

Erreur!. In the single ring, the wiring concentrators are entirely passive and their failure rate can be neglected[1].
Using the same approach as for the bus leads to:

$$\lambda_C \approx \lambda_R + n\,\overline{p}_N\,\lambda_N + n \cdot p_N \cdot \frac{\lambda_N}{\mu} \cdot [\,\lambda_R + (n-1)\,\overline{p}_N \cdot \lambda_N] \quad (8)$$

Considering the following parameter values:

- n: fixed (arbitrarily) at 15 as for the bus,
- λ_R: the failure rate of the ring: it has been taken equal to $\lambda_B = 2\ 10^{-5}$ / h
- μ: the repair rate, a mean repair duration of 2 hours has also been adopted.

leads to numerical results which are identical to those of the single bus: figures 13 and 14 also apply in this case replacing λ_B by λ_R.

3.5. Dual ring

The system is made up of N wiring concentrators and n stations. It is assumed that m stations are connected to each wiring concentrator which means that all the concentrators have the same influence from the dependability viewpoint.

Two more coverage factor have to be introduced:

- $\overline{p}_L$, the coverage factor of a link in the ring, defined as the probability of correct switching on the non failed ring in case of a failure of a link,
- $\overline{p}_{WC}$, the coverage factor of a WC, defined as the probability of correct passivation of the WC by counter-rotating the ring in case of a WC failure.

As far as concerns the maintenance policy, the same assumptions as for the bus and the single ring are considered; however, for the dual ring, the wiring concentrators have the highest repair priority. These assumptions are recalled hereafter:

- a covered failure (of the NAC or of the medium) does not affect service delivery; moreover, the repair of such a failure does not need service interruption,
- after a non-covered failure (of the NAC or of the medium), service delivery is interrupted; repair of all the failed elements is carried out before service is resumed,
- in case of one or several covered NAC failures, followed by a covered failure of the medium, repair priority is given to the medium,
- the WCs have the highest repair priority.

[1] In fact, the by-pass switches within a wiring concentrator are powered by their corresponding stations and can be included in the station failure rate, together with the failure rate of the link between the wiring concentrator and the station.

Figure 16 gives the corresponding Markov model and figure 17 gives the different notations for the dual ring model. The significance of the different states is given in figure 18.

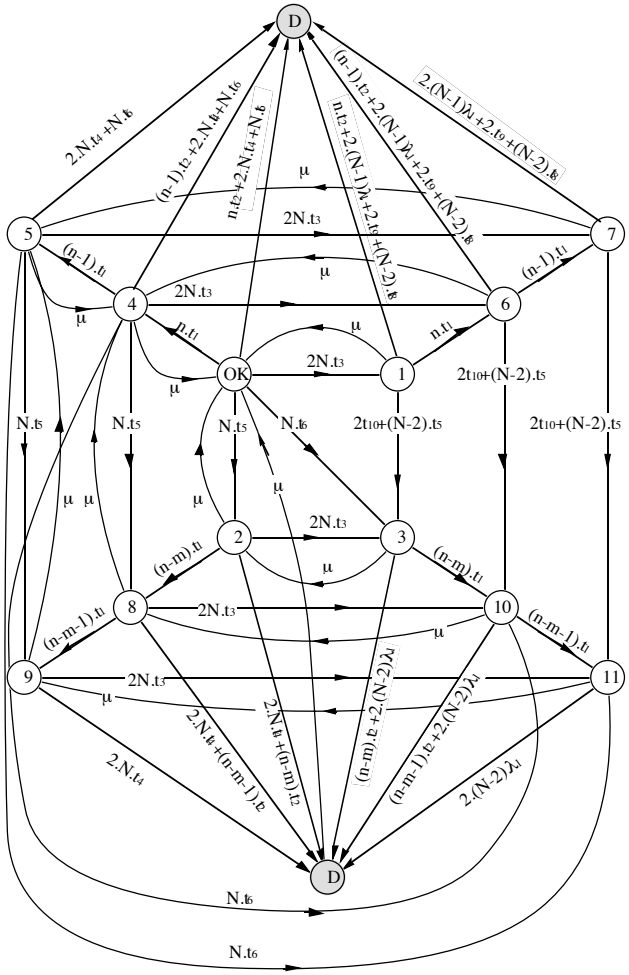


Figure 16: Markov model for the dual ring

Name	Associated failure rate
t1	$p_N .\lambda_N$
t2	$(1-p_N).\lambda_N$
t3	$p_L .\lambda_L$
t4	$(1-p_L).\lambda_L$
t5	$p_{WC} .\lambda_{WC}$
t6	$(1-p_{WC}).p_{WC}'.\lambda_{WC}$
t7	$(1-p_{WC}).(1-p_{WC}').\lambda_{WC}$
t8	$(1-p_{WC}).\lambda_{WC}$
t9	$(1-p_{WC}').\lambda_{WC}$
t10	$p_{WC}'.\lambda_{WC}$

Figure 17: Notations for the dual ring model

St.	Signification
OK	Initial state
1	Covered failure of a link in the ring: the dual ring is transformed into a single ring.
2	Failure of a WC covered by by-passing: the medium is still dual.
3	Failure of a WC covered by counter-rotating ring: the medium is reconfigured as a single ring.
4	Covered failure of a NAC.
5	Covered failure of two NACs.
6	Covered failure of a NAC and covered failure of a link in the ring.
7	Covered failure of two NACs and covered failure of a link in the ring.
8	Covered failure of a NAC and covered failure of a WC by by-passing: the medium is still dual.
9	Covered failure of two NACs and covered failure of a WC by by-passing: the medium is still dual.
10	Covered failure of a NAC and covered failure of a WC by counter-rotating ring: the medium is no longer dual.
11	Covered failure of two NACs and covered failure of a WC by counter-rotating ring: the medium is no longer dual.
D	System failure.

Figure 18: States of the dual ring model

Using the fact that $\lambda_R/\mu << 1$ and $\lambda_N/\mu << 1$ and considering only the first order terms leads to the following :

$$\lambda_C \approx 2\ \overline{p}_L\ \lambda_R + n\ \overline{p}_N\ \lambda_N + N\ \overline{p}_{WC}\ p'_{WC}\ \lambda_{WC} \qquad (9)$$

The main problem concerns the numerical value of λ_{WC}, the failure rate of a WC in the dual ring; since the WCs have to participate "intelligently" in system reconfiguration they must be active devices and their failure rate should be about the same as the failure rate of a NAC (it has been taken in fact equal to λ_N in this study).

Figure 19 plots several curves of λ_C / λ_N versus p_N , for $0.65 \le p_L \le 1$; n=15; N=5; m=3 ; $\lambda_{WC} = \lambda_N = 10^{-4}$ / h $\lambda_R = 2\ 10^{-5}$/h; $p_{WC} = p_{WC'} = 0.9$; $\mu = 0.5$ /h.

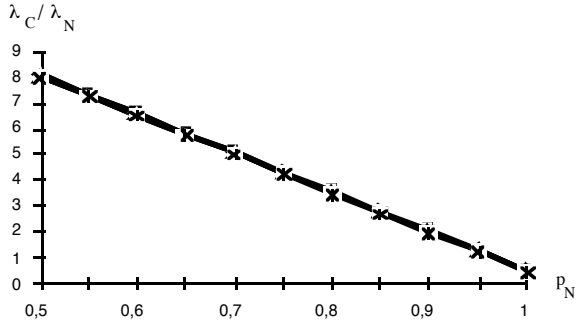


Figure 19: Numerical application for the dual ring

As for the bus, the coverage factor of a NAC, p_N , has more influence than p_L , the coverage factor of a link. Figure 20 gives λ_C and system unavailability, UA, (in hours per year) for different values of p_N with $p_L = 0.9$.

p_N	0.7	0.75	0.8	0.85	0.9
λ_C	5.0E-04	4.2E-04	3.5E-04	2.7E-04	2.0E-04
UA	8.75	7.43	6.12	4.80	3.49
p_N	0.95	0.99	0.999	0.9999	1
λ_C	1.2E-04	6.4E-05	5.1E-05	4.9E-05	4.9E-05
UA	2.17	1.12	0.885	0.862	0.859

Figure 20: λ_C (in failures/hour) and UA (in hours per year) versus p_N for the dual ring

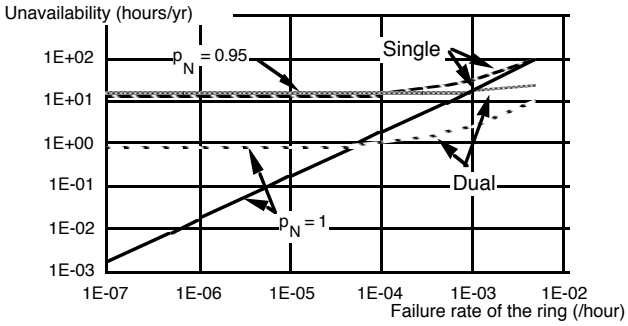


Figure 21: Communication system unavailability for the single and the dual ring, with $p_N = 0.95$ and 1.

With the chosen parameter values, figures 11 and 20 show that doubling the ring leads to a *decrease* in system availability. The curves of figure 21 (for which $p_L = 0.95$ and $p_{WC} = p'_{WC} = 0.9$) show unavailability versus the failure rate of the ring for $p_N = p_L = 0.95$ and $p_N = p_L = 1$ respectively; duplication only leads to an increase in dependability when $\lambda_R < 5\ 10^{-5}$ / h even with a perfect coverage of the NACs, $p_N = 1$, otherwise a single ring seems better (for the considered parameter values).

Considering $p_N = 1$, duplication of the ring acts as follows:

- for $\lambda_R = 10^{-5}$ / h, it increases the unavailability from 11 mn to 48 mn / year,
- for $\lambda_R = 10^{-4}$ / h, it decreases the unavailability from 1 h 45 mn to 58 mn / year.

3.6 Comparison

From a dependability viewpoint, it is very difficult to give an order of preference of the various communication systems considered for Delta-4. Assuming the same failure rate for the bus and for the ring leads to the same expression of the equivalent failure rate and for the unavailability of the single medium. Figures 15 and 21 show that for reasonable failure rates ($\lambda_B \approx \lambda_R << \lambda_N$) dependability measures are independent of this failure rate. Which means that the single bus and the single ring are equivalent.

With the value taken for the mean repair time ($1/\mu$) (2 hours, which is relatively low), the dominant source of undependability is lack of coverage rather than exhaustion of redundancy; i.e. the initial terms in the availability expressions dominate and the unavailability is therefore directly proportional to the mean repair time ($1/\mu$).

In the case the ring, duplication of the medium can actually deteriorate the dependability measures depending in the parameter values. The results enable the different architectures to be compared according to the various parameters in order to make a tradeoff and to select the most suitable architecture. For instance, for the considered values, the dual bus seems more interesting than the dual ring for $\lambda_B < 4\ 10^{-3}$ / h; however, the value of the failure rate of the wiring concentrator is of prime importance: a lower value of λ_{WC} (e. g. passive WCs) acts in favour of the dual ring.

Conclusion

In this paper, we have considered the dependability of the communication system of the Delta-4 system. We have derived basic communication models with four possible communication topologies: single and dual bus, and single and dual ring. The main results concern the derivation of analytical expressions of the failure rate and unavailability of the various communication systems.

These models are parametric; parameter values have been chosen so as to form a coherent set in order to compare the different architectures and to obtain a first estimate of the measures of dependability. Evaluation enabled identification of the most critical parameters and showed

that it is not possible to draw any definite conclusion as to a dependability ordering of the various topologies since the results are closely related to the values of these critical parameters. The estimated values need to be replaced by "real" values issued from field data, either by direct evaluation or by measurement. In particular, fault-injection can be used for measuring coverage factors (see for example, [1]). However it is shown that — whatever the architecture — the coverage factor of the NAC is of prime importance; it is thus worthwhile to put emphasis on this coverage (i.e. self-checking mechanisms) during development.

It is shown that for the single media configurations the equivalent failures rates are limited by the failure rate of the medium and non-covered failures of the NACs and that, for the dual media configurations, they are directly related to the failure rate of the non-covered failures only.

This work constitutes a first step in the validation by means of dependability evaluation. Software aspects will be considered in the next step. Collection of software reliability data is undertaken for some components of the software and software dependability will be evaluated using reliability growth models in the same manner as in [9, 11].

Modelling is to be extended to the complete hardware and software of a Delta-4 system. This activity will also be carried out progressively, considering first two target applications before extending to more general architectures.

References

[1] J. Arlat, M. Aguera, Y. Crouzet, J. Fabre, E. Martins, D. Powell, "Experimental Evaluation of the Fault Tolerance of an Atomic Multicast Protocol", *IEEE Transactions on Reliability,* 39 (4), pp. 455-467, October 1990 (Special Issue on Experimental Evaluation of Computer Reliability).

[2] M. Baptista, S. Graf, J.-L. Richier, L. Rodrigues, C. Rodriguez, P. Verissimo, J. Voiron, "Formal Specification and Verification of a Network Independent Atomic Multicast Protocol", in *3rd. Int. Conf. on Formal Description Techniques (FORTE'90),* (J. Quemada, J. Manas, E. Vazquez, Eds.), (North-Holland), 1990.

[3] P.A. Barrett, A.M. Hilborne, P.G. Bond, P. Verissimo, L. Rodrigues, N.A. Speirs, "The Delta-4 XPA Extra Performance Architecture", in *Proc. 20th Int. Symp. on Fault-Tolerant Computing Systems (FTCS-20),* (IEEE), pp.481-488, Newcastle upon Tyne, UK, June 1990.

[4] D.R. Cox, H.D. Miller, "*The Theory of Stochastic Processes*", London: Methuen (1968).

[5] *Delta-4 Architecture Guide*, The Delta-4 Project Consortium, (Editor D.Powell), Delta-4 Document n°G90.050/I1/R, December 1990.

[6] "*Token-passing Bus Access Method and Physical Layer Specifications*", ISO DIS 8802/4, December, 1984.

[7] "*Token Ring Access Method and Physical Layer Specifications*", ISO DP 8802/5, March, 1985.

[8] "*Fiber Distributed Data Interface*", ISO DP 9314 (ANSI Standard X3.139), 1987.

[9] K. Kanoun, M.R. Bastos, J. Moreira de Souza, "A Method for Software Reliability Analysis and Prediction: Application to the TROPICO-R Switching System", *IEEE Trans. on Software Engineering*, Vol 17, no 4, April 1991, 334-344.

[10] K. Kanoun, D. Powell, "*Dependability Evaluation Report LA2 - Communication System*", Delta- 4 reference R90.197//I1/P, LAAS Report 91.013.

[11] K. Kanoun, T. Sabourin, "Software Dependability of a Telephone Switching System", In *Proc. 17th Int. Symp. on Fault Tolerant Computing (FTCS-17)*, (IEEE), Pittsburgh, USA, June 1987, pp.236-241.

[12] J.C. Laprie, "Reliability and Availability of Repairable Systems", In *Proc. 5th Int. Symp. on Fault Tolerant Computing (FTCS-5)*, (IEEE), Paris, France, June 1975, pp.87-92.

[13] J.C.Laprie, "Dependability: A Unifying Concept for Reliable Computing and Fault Tolerance", in "*Dependability and Resilient Systems*", (T.Anderson, Ed.), Blackwell Scientific Publications, pp.1-28, 1989.

[14] J.C. Laprie, C. Béounes, M. Kaâniche, K. Kanoun, "The Transformation Approach to the Modeling and Evaluation of the Reliability and Availability Growth of Systems in Operation", In *Proc. 20th. Int. Symp. on Fault Tolerant Computing (FTCS-20)*, (IEEE), Newcastle upon Tyne, UK, July 1990, pp.364-371.

[15] J.C. Laprie, A. Costes, C. Landrault, "Parametric Analysis of 2-unit Redundant Computer Systems with Corrective and Preventive Maintenance", *IEEE Trans. on Reliability, R-30*, 2 (June 1981), pp.139-144.

[16] N. Lin, C.B. Silio Jr., "Ring Network Reliability - the Probability that all Operative Nodes can Communicate", in *Proc. 8th. Symp. on Reliable Distributed Systems (SRDS-8),* (IEEE), pp.64-71, Seattle, WA., USA, October 1989.

[17] D. Powell, G. Bonn, D. Seaton, P. Verissimo, F. Waeselynck, "The Delta-4 Approach to Dependability in Open Distributed Computing Systems", in *Proc. 18th Int. Symp. on Fault-Tolerant Computing Systems (FTCS-18),* (IEEE), pp.246-251, Tokyo, Japan, June 1988.

[18] A. Pagès, M. Gondran, "*System Reliability*", Eyrolles, Paris, 1980, in French.

[19] D. Powell, "Dependability Evaluation of Communication Support Systems for Local Area Distributed Computing", in *Proc. 12th. Int. Symposium on Fault-Tolerance Computing (FTCS-12),* (IEEE), pp.259-266, Santa Monica, CA., USA, June 1982.

[20] N.A. Speirs, P.A. Barrett, "Using Passive Replicates in Delta-4 to provide Dependable Distributed Computing", in *Proc. 19th Int. Symp. on Fault-Tolerant Computing Systems (FTCS-19),* (IEEE), pp.184-190, Chicago, MI, U.S.A, June 1989

[21] C. Singh, R. Billington, S.Y. Lee, "The Method of Stages for Non-Markov Models ", *IEEE Trans. on Reliability*, *R-26*, 2 (June 1977), pp.135-137.

[22] F.Venin, K.Kanoun, D.Powell, "*Dependability Evaluation Report LA1*", LAAS Report LAAS n°89-410, Delta-4 Report n°R89.137/11/P, December 1989.