Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1429

Published online 25 June 2012 in Wiley Online Library

# On the Reliability Evaluation of Failure Delayed Industrial Systems

Jose Faria*† and Americo Azevedo

**This paper presents an analytical approach for the evaluation of multi-user safety critical systems presenting a failure delayed behavior pattern. As a consequence of a failure event, the performance of these systems worsens progressively due to the internal fault tolerance mechanisms or the complacency of the users regarding the temporary unavailability of the services. A distinctive feature of the approach is the ability to handle stochastic models containing multiple processes with generalized distributions. The approach is based on the determination of analytical expressions to measure reliability, for instance, frequency and probability of failure states, which may be evaluated using general purpose mathematical tools. The paper first reviews other well-established techniques employed in the assessment of non-Markovian systems, particularly those based on stochastic Petri nets. The rationale of the new approach and its fundamental algorithms are presented together with a set of illustrative examples which highlight the strengths of the approach, as well as its limitations. Copyright © 2012 John Wiley & Sons, Ltd.**

**Keywords:** failure delayed systems; reliability analysis; stochastic processes; generalized distributions

## 1. Introduction

This paper presents a systematic approach to support the analysis and design of industrial engineering systems presenting a failure delay behavior pattern, that is, systems whose performance worsens progressively as a consequence of a failure. As will be extensively discussed, these systems contain multiple concurrent processes with generalized distributions that remain active for several consecutive states and are not reinitialized each time a new state is entered. This is a behavioral pattern that corresponds to the execution mechanism with a pre-emptive resume age policy described in[1]. Although there has been significant progress over the past two decades, most of which based on stochastic Petri nets (SPN), such as those reported in [2] and[3], the study of these systems remains a largely open issue in reliability analysis. In fact, existing methods often impose restrictive assumptions on the structure and behavior that limit their practical application to specific classes of systems. The approach presented here aims to further develop this topic, and it may be considered a straightforward alternative to other well-established solutions used to analyze non-Markovian systems, such as those presented in [4,5] and[6], or those based on Monte Carlo simulation techniques, as reported in [7] and[8].

The paper is organized as follows. A definition of failure delayed systems, as they are considered in the context of this paper, will be presented in Section 2. It will also be shown that these systems present a number of structural and behavioral distinctive features, namely the presence of non-exponential concurrent stochastic processes with similar time constants. In these circumstances, as discussed in[9], the reliability and performance indicators become highly sensitive to the shape of the stochastic distributions, and therefore the use of non-Markovian techniques becomes mandatory.

Section 3 presents a review of the existing techniques for non-Markovian systems and discusses their shortcomings regarding the evaluation of failure delayed systems. Then, the fundamental elements of the new approach are introduced in Section 4. In particular, it will be discussed how the inherent features of failure delayed systems may be explored in order to derive a set of effective evaluation algorithms devoted to this class of systems. The algorithm used to determine the frequency of the failure states will be introduced first. Then, the algorithm for the probability of the failure states will be considered. For relatively small models, these expressions for reliability and performance indices may be evaluated directly using general purpose mathematical tools. For larger models, a dedicated tool based on symbolic calculation was developed. The presentation of this tool is beyond the scope of this paper. However, a detailed discussion on its algorithms may be found in[10].

In the final part of the paper, a set of complementary practical examples and numerical results are presented in order to illustrate the practical application and usefulness of the approach. Finally, Sections 6 and 7 summarize the main contributions of the paper and

*INESC TEC - INESC Technology and Science and FEUP - Faculty of Engineering, University of Porto, Rua Roberto Frias, 4200-465 Porto, Portugal*
**Correspondence to: Jose Faria, INESC TEC - INESC Technology and Science and FEUP - Faculty of Engineering, University of Porto, Rua Roberto Frias, 4200-465 Porto, Portugal.*
*†E-mail: jfaria@fe.up.pt*

its content. The application of the method to other engineering domains will also be addressed here. In particular, it will be discussed how the approach can be extended to the analysis of complex business environments where a network of interacting activities and resources provide business services to a number of heterogeneous users.

## 2. Failure delayed systems

In many industrial engineering systems, the users are complacent about a temporary unavailability of the service provided to them. This means that, at first, the disturbances of a system failure are often negligible. However, if the failure persists, the system will enter into successive degraded operation modes. The quality of service will then decay progressively until a successful repair action is undertaken and the system restores its normal operation, or a catastrophic failure occurs.

Three systems presenting this kind of behavior and relating to very different engineering domains are represented in Figure 1. The first example (Figure 1.a) concerns an electrical power system with an alternative power supply that feeds an industrial process; the second example (Figure 1.b) relates to a manufacturing system with intermediate work-in-progress buffers between the cells; the third example (Figure 1.c) looks at the distributed business information system of a large retail company. These systems are analyzed in detail in [11,12] and[13]. For other examples of industrial failure delayed systems, see [14–16] or[17].

In the three models, $s_{up}$ corresponds to the normal operating state of the system, while $s_{fi}$ corresponds to the failure states. The failure, repair and propagation (or delay) processes are represented, respectively, by $p\lambda_i$, $p\mu_i$ and $p\gamma_i$.

According to the model in Figure 1.a, after the main power system fails (process $p\lambda$), the industrial process will be fed by an alternative power supply for a time $T_1$ (process $p\gamma_1$). If the duration of the failure is longer than $T_1$, the industrial process will be halted, but the control systems will continue to be supplied by an uninterruptible power supply (UPS) for a time $T_2$ (process $p\gamma_2$). If the power system restores its normal operation before $T_2$, a warm restart of the industrial process will be possible (process $p\mu_1$). If the autonomy of the UPS is exceeded, a catastrophic failure state will be reached once the materials being processed will suffer irreversible damage, causing a heavy loss (process $p\mu_2$).

In the manufacturing system depicted in Figure 1.b, the cells and plant controllers receive their data from a plant data server. If this server becomes unavailable (process $p\lambda$), the plant will be able to continue producing because the cell and plant level production plans are frozen for some time prior to the physical production (processes $p\gamma_C$ and $p\gamma_P$). However, the plant will enter a sub-optimal mode because it will not be possible to react to production events, such as new urgent orders. If an upstream cell halts its operation, the downstream cells will continue to be fed by the intermediate work in process buffer (processes $p\gamma_B$). The consequences of the failure will propagate downstream only when there is a shortage of products at the output of this buffer. If this production system belongs to a just-in-time supply chain, the severity of the damages is likely to increase dramatically.

Finally, Figure 1.c shows the information system of a business company from the retail sector. End users execute intra and inter-site transactions (which both depend on the availability of a number of remote data servers) and may tolerate a temporary unavailability of the information services. However, this complacency is different regarding intra and inter-sites transactions, and regarding the operations executed in each site (end consumers' point of sales, or logistical support). This behavior is represented in the model by two concurrent failure propagation processes, $p\gamma_1$ and $p\gamma_2$.

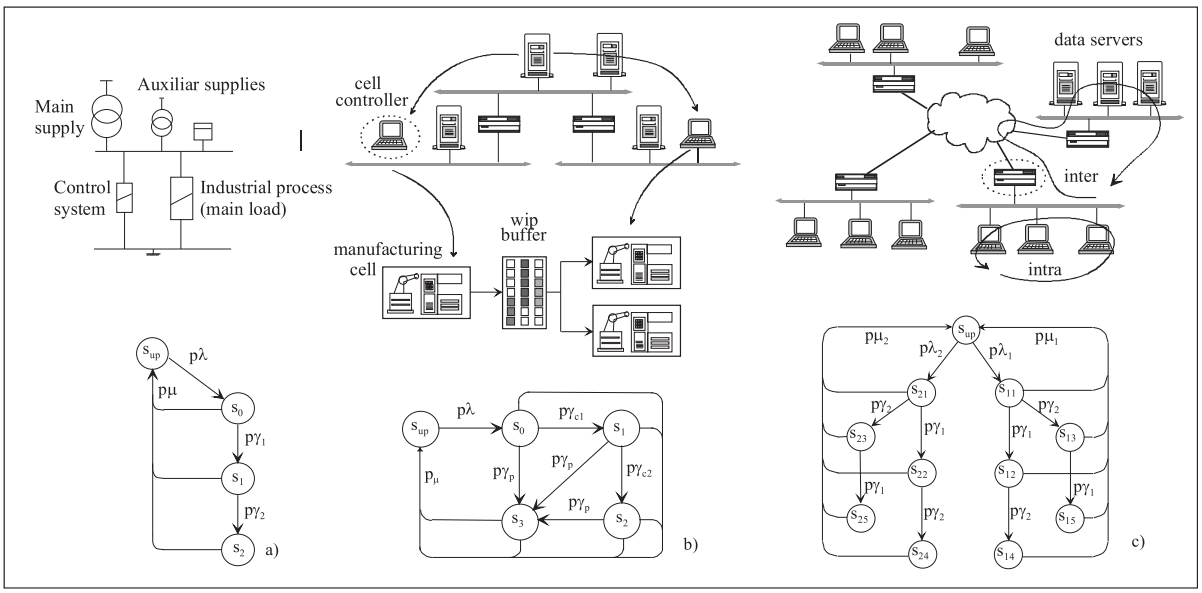


**Figure 1.** Examples of failure delayed systems

These three examples show that a progressive decay of performance after a failure, caused by an internal temporal redundancy mechanism, or by the complacency of the users regarding the temporary unavailability of the services provided to them, is a common behavior pattern in engineering systems.

The analysis of these models shows that failure delayed systems (hereafter referred to as FDS) present a number of common features that have a direct impact on their reliability and performance evaluation. Let us assume that *S* is a repairable failure delayed system and *M* is its behavior model (Figure 2). In this case, the following assumptions regarding *S* and *M* will be considered in the context of this paper:

- *S* provides services to multiple users (for instance, downstream manufacturing cells, electrical consumers or information systems users) each of which presents its own complacency regarding the unavailability of the services of *S*.
- *S* has a normal operating state which is represented in *M* as $s_{up}$.
- In $s_{up}$, one or more failure processes are active. Each one of these processes $p\lambda^{\alpha}$ corresponds to a particular failure mode $\alpha$.
- The execution of a failure process leads *S* from $s_{up}$ to one of the initial failure states where the disturbances for the users will typically be negligible.
- In each failure state, several concurrent delay processes, $p_{\gamma i}$, may be active. Each one of them corresponds to the complacency of a particular type of user regarding the failures of the system.
- The execution of a delay process leads the system to a delayed failure state, such as $s_n^{\alpha}$ with $n \geqslant 0$, where the severity of the damage will increase with *n*.
- In each initial or delayed failure state, a repair process $p_{\mu j}$ may be active. The execution of this process leads the system to the $s_{up}$. In other words, it is assumed that the repair is a regenerative process that completely restores the normal operating conditions (the extension of the model to non-regenerative repair will be discussed in Section 6).
- Failure, delay and repair processes may any type of distribution, deterministic or stochastic.
- When a transition occurs, the other processes that were also active in the initial may be de-activated, reinitialized or remain active (and keeping their firing time). Simultaneously, other repair or delay processes may be activated on the arrival at the new state.

These assumptions are summarized in the meta-model presented in Figure 3. This meta-model represents the generic elements that compose the model *M* of a particular FDS. The model *M* of a particular system *S* will contain one state corresponding to a normal operation, several (1. . .N) initial and delayed failure states and several transitions. Each transition links two states and is fired by the execution of a process. A process represents a physical (for instance, failure or repair) or functional (such as error propagation) mechanism whose execution causes the transition from a previous to a post state. The execution time of a process may be deterministic or stochastic (exponential or non-exponential). Several processes may be simultaneously active in the same state, and the same process may remain active in several states.

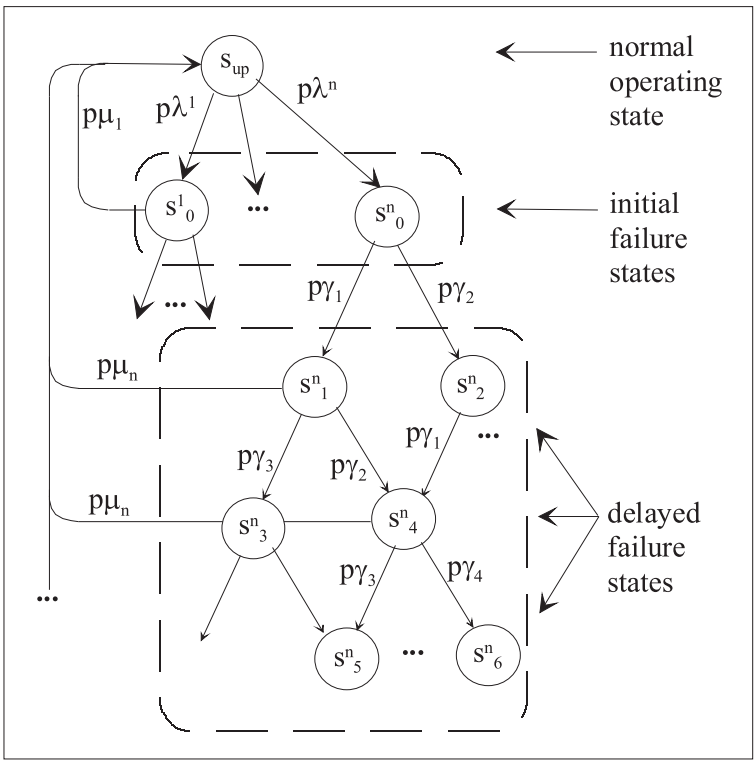


**Figure 2.** Failure delayed system models

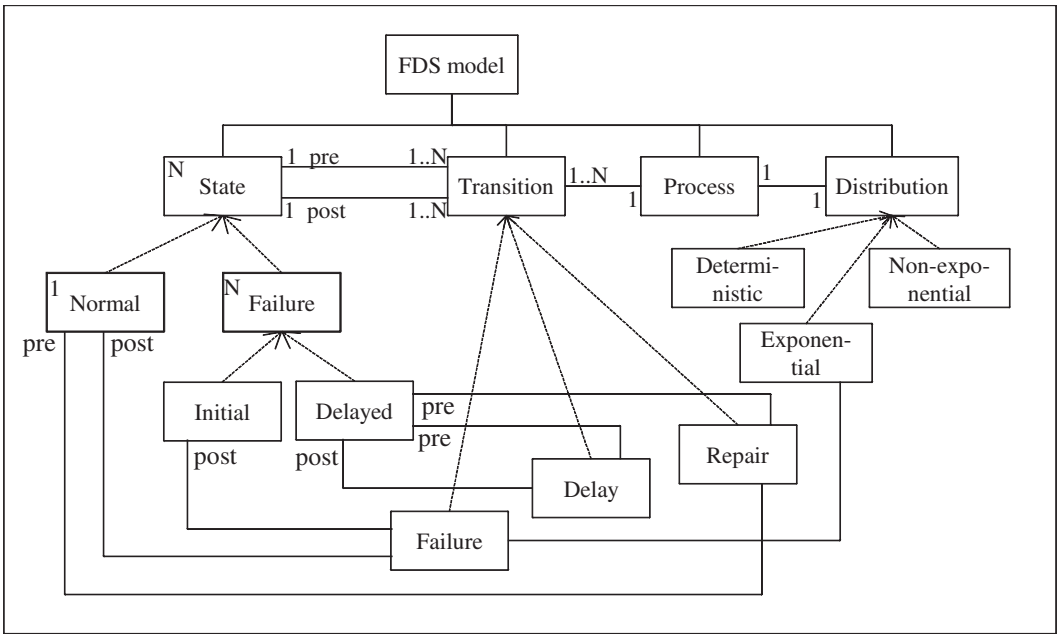


**Figure 3.** Failure delayed systems meta-model

In the normal operating state, only failure processes will be active. In each failure state (either initial or delayed), several delay processes and 0 or 1 repair processes may be active. A failure transition links the normal state to an initial failure state and the process assigned to it is assumed to be exponential. A delay transition links two failure states and a repair transition links a failure state to $s_0$. The processes assigned to these transitions may have any type of distribution.

It is important to emphasize that, according to the previous assumptions, $s_{up}$ and the initial failure states will be regeneration points because, once arrived at one of these states, the future evolution of the system does not depend on its history. On the contrary, the delayed failure states do not have this property because non-exponential delay processes may remain active in several consecutive states without being reinitialized. In these conditions, the model implements an execution mechanism, usually referred to as the pre-emptive resume age policy.[18] The regeneration time points of the underlying regenerative Markov process are the instants of arrival at the normal operating states $s_{up}$, and at the initial failure states (later, it will be shown that it is possible to free the restriction regarding the regeneration of the stochastic process at the initial failure states).

It should also be noted that a fundamental difference between the FDS meta-model and the Petri nets models is the fact that the stochastic distributions are assigned directly to the processes, not to the transitions. This highly simplifies system modeling when a non-exponential process remains active for several consecutive states without being reinitialized (that is, pre-emptive resume policy) as this is often the case with the FDS.

## 3. Review of existing methods

The device of stages is one of the most frequently used techniques for the evaluation of non-Markovian systems. This technique makes it possible to model a large range of experimental probability density functions using a set of additional exponential processes in a sequence or in parallel. For example, a log-normal distribution often found in repair processes may be represented by a combination of a series of states with two states in parallel, as shown in [19] and[20]. First introduced in[21], it has been applied to the evaluation of reliability in fault-tolerant computer systems,[22] and to the analysis of reliability in electrical power systems.[23] An extension of the method has been proposed in [24] to make it possible to assign a memory policy – resampling, enabling and memory, as defined in [3] – to any timed transition. One of the most important features of the method is the possibility of designing automated tools to support its application, as presented in[25]. This tool uses Petri nets as the modeling tool and converts the reachability set of the net into a continuous time Markov chain defined over an extended state space. Although this method is very flexible, it restricts the firing times of the stochastic processes so that they have phase-type distributions.[26] Consequently, the method presents a major limitation when the systems under analysis contain deterministic or quasi-deterministic processes. This happens because the number $n$ of additional states (or stages) required to approximate an experimental distribution rises quadratically with the ratio $\sigma/m$, with $\sigma$ being the standard deviation of the distribution, and $m$ its mean. If the system contains several simultaneously active non-exponential processes (as is often the case with failure degradable systems), there will be an 'explosion' of the number of states in the equivalent Markov chain, as the stages of each process should be combined with those of the concurrent processes.

Over the past two decades, several evaluation techniques based on SPN modeling have been developed in order to support the reliability analysis and the performance evaluation of complex systems. When SPN were first introduced,[27] all the random variables

associated with the transitions were assumed to be exponentially distributed, so that the evolution of a Petri net could be mapped into a continuous Markov chain. Since then, and in order to broaden the field of application of SPN, several classes of Petri nets incorporating non-exponential features in their definition have been proposed. This is the case of the deterministic and SPN defined in[28], in which a single deterministic transition may exist in each marking. Subsequently, it was observed in [29] that the underlying marking process is a Markov regenerative process. This made it possible for the model to be extended in order to accommodate immediate transitions, exponentially distributed timed transitions and generally distributed timed transitions. However, an important restriction is maintained – that at least one generally distributed timed transition is enabled in each marking. Two evaluation approaches were then developed: one based on the derivation of the time-dependent transition probability matrix in the Laplace transform,[29] and the other based on the supplementary variables method.[30]

The restriction of a single non-exponential process in each marking was subsequently removed by the class of regenerative SPN introduced in [1], through a time discretization approach and an approximation of non-exponential firing times by means of the phase type distribution. A tool implementing this technique is presented in[31]. Another approach is presented in.[32] This approach allows models with concurrently enabled generally distributed transitions to be analyzed using a discrete time approximation of the stochastic behavior of the marking process. However, these two techniques may lead to an explosion of the state space. In order to deal with the state space explosion, [33] presents an efficient algorithm to generate and store the reachability graph based on a symbolical representation of the macro states.

In spite of this progress, several restrictions still apply to the analytical evaluation of non-Markov systems, and no general solution is available. In fact, each extension of the Petri net model normally represents a particular compromise between modeling power and numerical tractability, designed to fit the requirements of a particular class of systems or studies.

This is also the case for the approach that will be presented below once it looks at the common behavioral and structural patterns of FDS systems in order to provide a fully automated procedure that closely fits their characteristics and requirements.

## 4. New approach fundamentals

This section introduces the mathematical foundations to determine the two fundamental performance measures for repairable systems: the *frequencies* of arrival and the *probabilities* of the states of the non-Markovian model *M*. The analytical expressions for the frequencies will be considered first in chapter 4.1. Then, the state probability expressions will be addressed in chapter 4.2. The expressions are obtained using a systematic procedure that takes as input the structure of the model (that is, the model's states and transitions as specified in the FDS meta-model) and the distributions of the additional stochastic processes.

The procedure is based on the notion of *state trajectory*: immediately after a failure event occurs, the system stays at one of the initial failure states. Then, it returns to the normal operating state following one of the several possible trajectories, as shown in Figure 4. A trajectory is an ordered set of failure states $\{s_1^\alpha, s_2^\alpha, s_3^\alpha, \ldots s_n^\alpha\}$ that starts at one of the regenerative initial failure states $s_0^\alpha$.

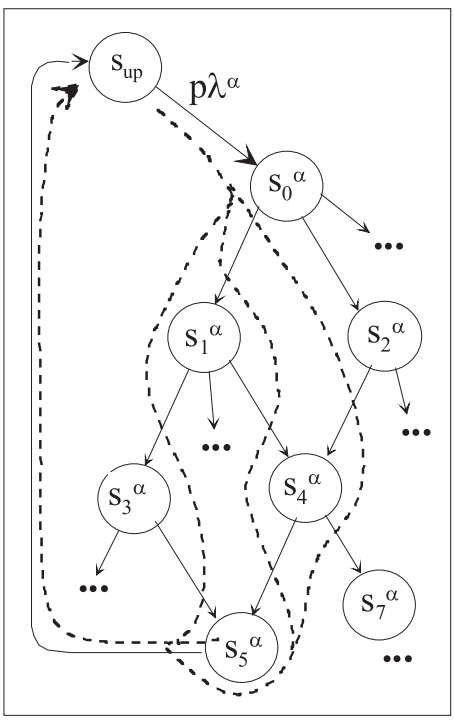


**Figure 4.** Alternative trajectories

For each pair of consecutive states, $s^{\alpha}_{k-1}$ and $s^{\alpha}_{k}$, there is a delay process $p\gamma_i$ in $M$ whose execution causes the transition from $s^{\alpha}_{k-1}$ to $s^{\alpha}_{k}$. For simple models, the trajectories may be determined by a visual inspection of the graph. For larger models, an automatic search algorithm may be useful. The discussion of such algorithms is out of the scope of this paper. However, it can be easily derived from the meta-model in Figure 2.

To present the procedure, the following notation will be adopted:

- $\boldsymbol{\Lambda}_M$ and $\boldsymbol{P}_M$: two vectors such $\boldsymbol{\Lambda}_M(s)$ and $\boldsymbol{P}_M(s)$ contain the frequency and the probability of state $s$, respectively;
- $s_{up}$: the normal operating state,
- $p\lambda^{\alpha}$: the failure process corresponding to failure mode $\alpha$;
- $s^{\alpha}_0$: the initial failure state corresponding to failure mode $\alpha$;
- $p\gamma_i$ and $p\mu_j$: the processes corresponding to the propagation delay $i$ and the repair action $j$, respectively;
- $s^{\alpha}_n$: a delayed failure state subsequent to $s^{\alpha}_0$ ($n \geq 1$);
- $f_p(t)$: the probability density function of process $p$.

### 4.1. *Failure states frequency*

Suppose that $s^{\alpha}_n$ is a failure state whose frequency is to be determined and that $\Psi^{\alpha}_n$ is the set of trajectories starting at $s^{\alpha}_0$ and ending at $s^{\alpha}_n$. The frequency of the failure state $\Lambda(s^{\alpha}_n)$ results from the sum of the frequencies of each trajectory $\psi$ of $\Psi^{\alpha}_n$:

$$\Lambda(s^{\alpha}_n) = \sum_{\psi \in \Psi^{\alpha}_n} \Lambda(\psi) \tag{1}$$

The frequency of each trajectory $\psi$ comes from the product of (i) the frequency of $s^{\alpha}_0$ (note that accordingly to the meta-model of Figure 2, all the trajectories of $\Psi^{\alpha}_n$ have the same initial state) and (ii) the probability that, once arrived at $s^{\alpha}_0$, the system follows $\psi$:

$$\Lambda(s^{\alpha}_n) = \Lambda(s^{\alpha}_0) \sum_{\psi \in \Psi^{\alpha}_n} \mathrm{P}(\psi) \tag{2}$$

The determination of $\mathrm{P}(\psi)$ will be addressed hereafter, whereas that of $\Lambda(s^{\alpha}_0)$ will be addressed in chapter 4.3 because it requires formulae that will be introduced only in 4.2.

*4.1.1. Probability of a trajectory.* The probability of a trajectory comes from the product of the probabilities of each one of its transitions. Consider the following trajectory as an example:

$$\psi = \{s^{\alpha}_0, s^{\alpha}_a, s^{\alpha}_b, \ldots s^{\alpha}_r, s^{\alpha}_s\}$$

Its probability will be:

$$\mathrm{P}(\psi) = \mathrm{P}(s^{\alpha}_0 \rightarrow s^{\alpha}_a) \times \mathrm{P}(s^{\alpha}_a \rightarrow s^{\alpha}_b) \times \ldots \times \mathrm{P}(s^{\alpha}_r \rightarrow s^{\alpha}_s)$$

For the sake of simplicity of the expressions, it will be considered that, within each trajectory, the states are renumbered according to their order, as exemplified in Figure 5 for the three trajectories considered above. As it was assumed that the delay and repair

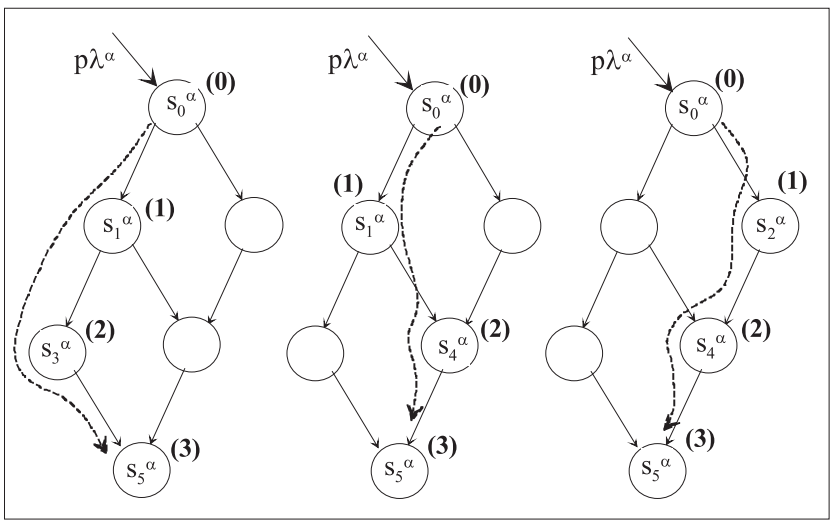


**Figure 5.** Renumbering of the states within each trajectory

processes may present non-exponential distributions and that these processes may remain active for several failure delayed states without being reinitialized, the probability of a transition depends on the instants of the previous transitions, that is, the probability of the transition $s_i \rightarrow s_{i+1}$ depends on the instants of the transitions $s_0 \rightarrow s_1, \ldots$ and $s_{i-1} \rightarrow s_i$.

Therefore, if the random variable $t_i$ represents the time elapsed between the arrival at the initial failure state $s_0$ and the arrival at the ith state $s_i$, the probability of a trajectory leading to the nth state, $s_n$, may be expressed as:

$$P(\psi) = \quad P(s_0 \rightarrow s_1) \times P_{t_1}(s_1 \rightarrow s_2) \times \ldots \times P_{t_1 t_2 \ldots t_{n-1}}(s_{n-1} \rightarrow s_n)$$

or:

$$P(\psi) = \prod_{i=1}^{n} P_{t_1 \ldots t_{i-1}}(s_{i-1} \rightarrow s_i) \tag{3}$$

where $P_{t_1 \ldots t_{i-1}}(s_{i-1} \rightarrow s_i)$ represents the conditional probability of transition from $s_{i-1}$ to $s_i$ given that the previous transitions of $\psi$ have occurred at $t_1..t_{i-1}$. These conditional probabilities may, in turn, be evaluated according to the following expression:

$$P_{t_1, t_2 \ldots t_{i-1}}(s_{i-1} \rightarrow s_i) = \int_0^{\infty} T(t_1) \int_{t_1}^{\infty} T(t_2) \ldots \int_{t_{i-1}}^{\infty} T(t_i)\, dt_i \ldots dt_2 dt_1 \tag{4}$$

where $T(t_i)$ is the density function of the random variable $t_i$. Time $t_i$ depends on the set of stochastic processes that are active in state $s_{i-1}$. If $\boldsymbol{\Omega}_n$ is the set of processes that are active in a state $s_k$, and $p_i$ is the process that causes the transition from $s_{i-1}$ to $s_i$ within the trajectory (Figure 6), then the expression for $T(t_i)$ comes from the product of the density function of this process, $f_{p_i}(t_i)$, and from the probability that the other processes $p$ belonging to set $\Omega_{i-1}$ do not occur before $t_i$ ($p \in \Omega_{i-1}$ and $p \neq p_i$).

If $p$ is one of the processes of $\Omega_{i-1}$ whose density function is $f_p(t)$ and became active at a previous instant $t_p^0$, then the density function for the execution time of this process (knowing that it became active at $t_p^0$ and it is still active at $t_{i-1}$) is:

$$f_p'(t) = \frac{f_p\left(t - t_p^0\right)}{1 - \int_{t_p^0}^{t_{i-1}} f_p\left(\tau - t_p^0\right) d\tau}, \quad t > t_{i-1}$$

where $\tau$ is an auxiliary variable with local scope. Therefore, it comes for $T(t_i)$:

$$T(t_i) = \frac{f_{p_i}\left(t_i - t_{p_i}^0\right)}{1 - \int_{t_{p_i}^0}^{t_{i-1}} f_{p_i}\left(\tau - t_{p_i}^0\right) d\tau} \left( \prod_{\substack{p \in \Omega_i \\ p \neq p_i}} \frac{\int_{t_i}^{\infty} f_p\left(\tau' - t_p^0\right)}{1 - \int_{t_p^0}^{t_{i-1}} f_p\left(\tau - t_p^0\right) d\tau)} d\tau' \right) \tag{5}$$

where:

- $t_p^0$ is the instant at which process $p$ is activated. This will always coincide with one of the random variables $t_j$, with $j < i - 1$;
- $\frac{f_{p_i}\left(t_i - t_{p_i}^0\right)}{1 - \int_{t_{p_i}^0}^{t_{i-1}} f_{p_i}\left(\tau - t_{p_i}^0\right) d\tau}$ represents the density function of the instant of transition from $s_{i-1}$ to $s_i$ due to $p_i$;
- $\frac{\int_{t_i}^{\infty} f_p\left(\tau' - t_p^0\right)}{1 - \int_{t_p^0}^{t_{i-1}} f_p\left(\tau - t_p^0\right) d\tau)} d\tau'$ represents the probability that another process $p$ of $\Omega_{i-1}$ does not occur before $p_i$.

Now, combining (3), (4) and (5), the expression for the probability of the trajectory $\psi$ may be obtained from:

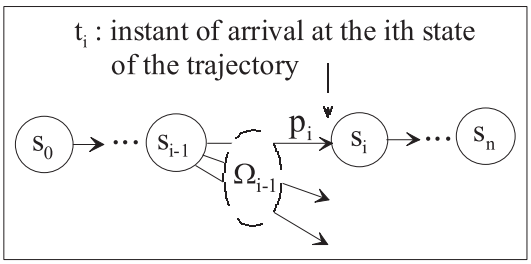


**Figure 6.** Arrival at the ith state of the trajectory

$$P(\psi) = \int_0^{\infty} T(t_1) \int_{t_1}^{\infty} T(t_2) \,.. \int_{t_n-1}^{\infty} T(t_n)\, dt_n \,..dt_2\, dt_1 \tag{6}$$

If a process $p$ stays active from state $s_k$ (that is, $t_p^0 = t_k$) to state $s_m$, its density function will take part in the expressions $T(t_j)$ for $k \le j \le m$. Therefore, the contribution of $p$ to $P(\psi)$ will be:

$$\frac{\int_{t_{k+1}}^{\infty} f_p\left(\tau' - t_k\right) d\tau'}{1} \times \frac{\int_{t_{k+2}}^{\infty} f_p\left(\tau' - t_k\right) d\tau'}{1 - \int_{t_k}^{t_{k+1}} f_p(\tau - t_k) d\tau} \,.. \frac{\int_{t_{m+1}}^{\infty} f_p(\tau - t_k)\, d\tau'}{1 - \int_{t_k}^{t_m} f_p(\tau - t_k) d\tau}$$

Since $\int_{t_{l+1}}^{\infty} f_p(\tau - t_k) d\tau$ equals $(1 - \int_{t_l}^{t_{l+1}} f_p(\tau - t_k) d\tau)$, the global contribution of $p$ to $T(t_k)$ will be equivalent to $\int_{t_{m+1}}^{\infty} f_p(\tau - t_k) d\tau$.

This means that if a process is active from $s_k$ to $s_m$, it is possible to consider the contribution of $p$ to $T(t_j)$ only at state $s_m$. This fact leads to a significant simplification of the density functions:

$$T(t_i) = f_{p_i}\left(t_i - t_{p_i}^0\right) \left( \prod_{\substack{p \in \Omega_{i-1} \\ p \notin \Omega_i \\ p \neq p_i}} \int_{t_i}^{\infty} f_p\left(\tau - t_p^0\right) d\tau \right) \tag{7}$$

The illustrative example presented below will help clarify this step.

*4.1.1. Illustrative example: state frequency.* To illustrate the practical application of the procedure presented above, consider state $s_4^1$ in the model of Figure 7.a.

Three trajectories lead to this state:

$$\Psi^1 = \{s_0^1 \to s_1^1 \to s_4^1\}; \Psi^2 = \{s_0^1 \to s_2^1 \to s_3^1 \to s_4^1\} \text{ and } \Psi^3 = \{s_0^1 \to s_1^1 \to s_3^1 \to s_4^1\}$$

For the first trajectory $\psi^1$ (Figure 7.b), the probabilities of the two transitions are:

$$\mathrm{P}\left(s_0^1 \to s_1^1\right) = \int_0^{\infty} p_{\gamma 1}(t_1) \int_{t_1}^{\infty} p_{\gamma 2}(\tau) d\tau \;\; dt_1$$

$$\mathrm{P}\left(s_1^1 \to s_4^1\right) = \int_{t_1}^{\infty} p_{\gamma 3}(t_2) \int_{t_2}^{\infty} \frac{p_{\gamma 2}(\tau) d\tau}{1 - \int_0^{t_1} p_{\gamma 2}(\tau) d\tau} \int_{t_2}^{\infty} p_{\mu}(\tau - t_1) d\tau \;\; dt_2$$

Thus, the probability of the trajectory is:

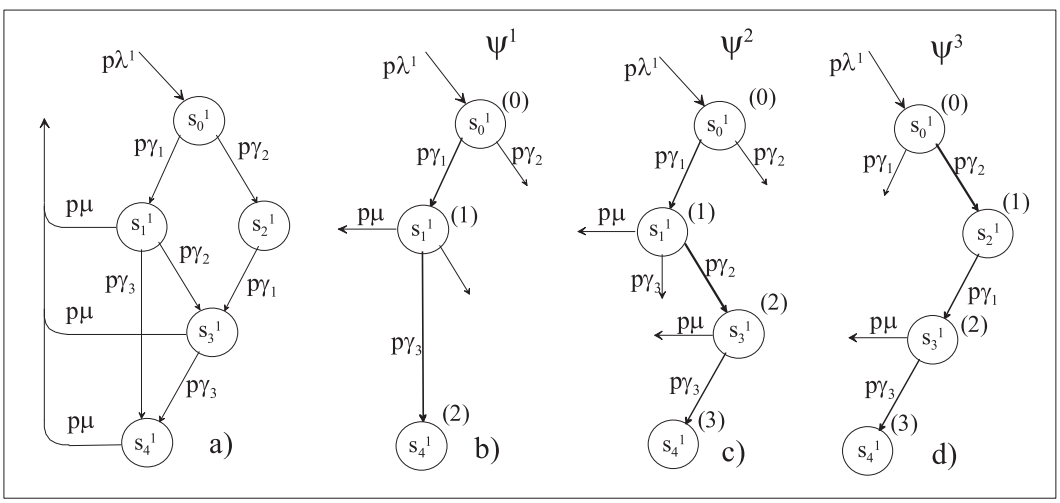


**Figure 7.** Illustrative example

$$P(\Psi^1) = \int_0^\infty p_{\gamma 1}(t_1) \int_{t_1}^\infty p_{\gamma 3}(t_2) \int_{t_2}^\infty p_{\gamma 2}(\tau) d\tau \int_{t_2}^\infty p_\mu(\tau - t_1) d\tau \; dt_2 dt_1$$

The corresponding expressions for the second trajectory (Figure 7.c) are:

$$P(s_0^1 \to s_2^1) = \int_0^\infty p_{\gamma 1}(t_1) \int_{t_1}^\infty p_{\gamma 2}(\tau) d\tau \; dt_1$$

$$P(s_2^1 \to s_3^1) = \int_{t_1}^\infty \frac{p_{\gamma 2}(t_2)}{1 - \int_0^{t_1} p_{\gamma 2}(\tau) d\tau} \int_{t_2}^\infty p_{\gamma 3}(\tau - t_1) d\tau \int_{t_2}^\infty p_\mu(\tau - t_1) d\tau \; dt_2$$

$$P(s_3^1 \to s_4^1) = \int_{t_2}^\infty \frac{p_{\gamma 3}(t_3 - t_1)}{1 - \int_{t_1}^{t_2} p_{\gamma 3}(\tau - t_1) d\tau} \int_{t_3}^\infty \frac{p_\mu(\tau' - t_1) d\tau'}{1 - \int_{t_1}^{t_2} p_\mu(\tau - t_1) d\tau} \; dt_3$$

$$P(\Psi^2) = \int_0^\infty p_{\gamma 1}(t_1) \int_{t_1}^\infty p_{\gamma 2}(t_2) \int_{t_2}^\infty p_{\gamma 3}(t_3 - t_1) dt \int_{t_3}^\infty p_\mu(\tau - t_1) d\tau \; dt_3 dt_2 dt_1$$

Similarly, for the third trajectory (figure 7.d):

$$P(\Psi^3) = \int_0^\infty p_{\gamma 2}(t_1) \int_{t1}^\infty p_{\gamma 1}(t_2) \int_{t_2}^\infty p_{\gamma 3}(t_3 - t_1) dt \int_{t_\infty}^\infty p_\mu(\tau - t_2) d\tau \; dt_3 dt_2 dt_1$$

### 4.2. *Failure state probability*

In this chapter, the procedure introduced in 4.1 will be extended in order to address the probability of the failure states. As before, let us assume that $s_n^\alpha$ is a failure state of a model $M$, $\Psi_n^\alpha$ is the set of trajectories leading to $s_n^\alpha$ and $P(\psi)$ is the probability of the trajectory $\psi$. In these conditions, the probability of $s_n^\alpha$ may be obtained from:

$$P(s_n^\alpha) = \Lambda(s_0^\alpha) \sum_{\psi \in \Psi_n^\alpha} P(\psi) \times \overline{t_{n,\alpha}} \tag{8}$$

where the new term $\overline{t_{n,\alpha}}$ represents the mean sojourn time of $s_n^\alpha$ within $\psi$, that is, the time elapsed between the arrival at $s_n^\alpha$ and the departure from this state when the system follows trajectory $\psi$.

If $p$ is a process of $\Omega_n$, the mean sojourn time in state $s_n^\alpha$ when the transition to $s_{n+1}^\alpha$ is caused by $p$ results from the product of (i) the mean execution time of $p$ and (ii) the probability that the other processes of $\boldsymbol{\Omega}_n$ do not occur before $p$. This is represented by:

$$\int_{t_n}^\infty (t_{n+1} - t_n) f_p(t_{n+1} - t_{0p}) \left( \prod_{\substack{p' \in \Omega_n \\ p' \neq p}} \int_{t_{n+1}}^\infty f_{p'}\left(t' - t_{0p'}\right) dt' \right) dt_{n+1}$$

As the output transition from state $s_n^\alpha$ may be caused by any of the processes belonging to $\boldsymbol{\Omega}_n$, the total sojourn time in this state may be obtained from:

$$\overline{t_{n,\alpha}} = \sum_{p \in \Omega_n} \int_{t_n}^\infty (t_{n+1} - t_n) f_p\left(t_{n+1} - t_p^0\right) \left( \prod_{\substack{p' \in \Omega_n \\ p' \neq p}} \int_{t_{n+1}}^\infty f_{p'}\left(\tau - t_{p'}^0\right) d\tau \right) dt_{n+1} \tag{9}$$

The expression of $\overline{t_{n,\alpha}}$ depends on the instants of the previous transitions of $\psi$ (due to the instants of activation $t_p^0$ and $t_{p'}^0$ of the processes belonging to $\boldsymbol{\Omega}_n$. Therefore, this expression should be combined with the probability of $\psi$ (6), yielding:

$$P(s_n^\alpha) = \Lambda(s_0^\alpha) \sum_{\psi \in \Psi_n^\alpha} \int_0^\infty T(t_1) \ldots \int_{t_{n-1}}^\infty T(t_n) \times$$

$$\times \left[ \sum_{p \in \Omega_n} \int_{t_n}^\infty (t_{n+1} - t_n) f_p\left(t_{n+1} - t_p^0\right) \left( \prod_{\substack{p' \in \Omega_n \\ p' \neq p}} \int_{t_{n+1}}^\infty f_{p'}\left(\tau - t_{p'}^0\right) d\tau \right) dt_{n+1} \right] dt_{n-1} \ldots dt_1 \tag{10}$$

The expressions for the state probabilities (as the previous expressions for the state frequencies) depend on the frequency of arrival at the initial failure state, $\Lambda(s_0^\alpha)$, which will be addressed in chapter 4.3, after the following illustrative example.

*4.2.1. Illustrative example: state probability.* For the three trajectories considered before (Figure 7), the mean time that the system will remain in state $s_5$ is:

$$\overline{t_{4,1}^\psi} = \int_{t_2}^\infty (t_3 - t_2) \frac{p_\mu(t_3 - t_1)}{1 - \int_{t_1}^{t_2} p_\mu(\tau - t_1) d\tau} \; dt_3$$

$$\overline{t_{4,2}^\psi} = \int_{t_3}^\infty (t_4 - t_3) \frac{p_\mu(t_4 - t_2)}{1 - \int_{t_2}^{t_3} p_\mu(\tau - t_2) d\tau} \; dt_4$$

$$\overline{t_{4,3}^\psi} = \int_{t_3}^\infty (t_4 - t_3) \frac{p_\mu(t_4 - t_1)}{1 - \int_{t_1}^{t_3} p_\mu(\tau - t_1) d\tau} dt_4$$

The combination of these expressions with those regarding the probability of the trajectories presented in 4.1.2 yields:

$$P(\psi^1) \times \overline{t_{4,1}^\psi} = \int_0^\infty p_{\gamma 1}(t_1) \int_{t1}^\infty p_{\gamma 3}(t_2) \int_{t_2}^\infty p_{\gamma 2}(\tau) d\tau \int_{t_2}^\infty (t_3 - t_2) p_\mu(t_3 - t_1) \; dt_3 dt_2 dt_1$$

$$P(\psi^2) \times \overline{t_{4,2}^\psi} = \int_0^\infty p_{\gamma 1}(t_1) \int_{t_1}^\infty p_{\gamma 2}(t_2) \int_{t_2}^\infty p_{\gamma 3}(t_3 - t_1) dt \int_{t_3}^\infty (t_4 - t_3) p_\mu(t_4 - t_2) dt_4 dt_3 dt_2 dt_1$$

$$P(\psi^3) \times \overline{t_{4,3}^\psi} = \int_0^\infty p_{\gamma 2}(t_1) \int_{t_1}^\infty p_{\gamma 1}(t) \int_{t_2}^\infty p_{\gamma 3}(t_3 - t_1) dt \int_{t_3}^\infty (t_4 - t_3) p_\mu(t_4 - t_1) dt_4 dt_3 dt_2 dt_1$$

As a final example, the determination of the probability of state $s_1^1$ should be considered in which several processes are active. The probability of the single transition that leads to this state is:

$$\mathrm{P}(\psi) = \int_0^\infty p_{\gamma 1}(t_1) \int_{t_1}^\infty p_{\gamma 2}(t) dt \; dt_1$$

The mean time spent in the state for the three output processes, $p_{\gamma 2}$, $p_{\gamma 3}$ and $p_\mu$, are:

$$\text{for } p_{\gamma 2}: \int_{t_1}^\infty (t_2 - t_1) \frac{p_{\gamma 2}(t_2)}{1 - \int_0^{t_1} p_{\gamma 2}(\tau) d\tau} \int_{t_2}^\infty p_{\gamma 3}(\tau - t_1) d\tau \int_{t_2}^\infty p_\mu(\tau - t_1) \, d\tau \; dt_2$$

$$\text{for } p_{\gamma 3}: \int_{t_1}^\infty (t_2 - t_1) p_{\gamma 3}(t_2 - t_1) \int_{t_2}^\infty \frac{p_{\gamma 2}(\tau') d\tau'}{1 - \int_0^{t_1} p_{\gamma 2}(\tau) d\tau} \int_{t_2}^\infty p_\mu(\tau - t_1) \, d\tau \; dt_2$$

$$\text{for } p_\mu: \int_{t_1}^\infty (t_2 - t_1) p_\mu(t_2 - t_1) \int_{t_2}^\infty \frac{p_{\gamma 2}(\tau') d\tau'}{1 - \int_0^{t_1} p_{\gamma 2}(\tau) d\tau} \int_{t_2}^\infty p_{\gamma 3}(\tau - t_1) \, d\tau \; dt_2$$

The combination of these expressions yields:

$$\begin{aligned} P(\psi) \times \overline{t_{1,1}^\psi} = {} & \int_0^\infty p_{\gamma 1}(t_1) [\int_{t_1}^\infty (t_2 - t_1) p_{\gamma 2}(t_2) \int_{t2}^\infty p_{\gamma 3}(\tau - t_1) d\tau \int_{t2}^\infty p_\mu(\tau - t_1) \, d\tau \; dt_2 + \\ & + \int_{t_1}^\infty (t_2 - t_1) p_{\gamma 3}(t_2 - t_1) \int_{t_2}^\infty p_{\gamma 2}(\tau) d\tau \int_{t_2}^\infty p_\mu(\tau - t_1) \, d\tau \; dt_2 + \\ & + \int_{t_1}^\infty (t_2 - t_1) p_\mu(t_2 - t_1) \int_{t_2}^\infty p_{\gamma 2}(\tau) d\tau \int_{t_2}^\infty p_{\gamma 3}(\tau - t_1) \, d\tau \; dt_2] \, dt_1 \end{aligned}$$

### 4.3. *Initial failure state frequencies*

Depending on the distributions of the failure and the repair processes, four situations regarding the determination of frequencies for the initial failure states have to be considered:

- exponential failure processes and a common repair process
- exponential failure processes and several repair processes
- non-exponential failure processes and a common repair process
- non-exponential failure processes and several repair processes

Hereafter, only the situations corresponding to exponential failure processes, which are the most common ones, will be considered. The analysis of the other two situations is addressed in annex.

*4.2.1. Exponential failure processes and common repair process.* This is the simplest and most common situation found in practical applications regarding industrial FDS: the failure processes present exponential distributions; the repair processes are enabled immediately after the occurrence of the failures; and they remain active until the system re-enters the normal operating state $s_{up}$.

In this case, the set of failure states corresponding to a particular failure mode may be grouped into a single macro state because all of them share the same repair process (figure 8). The mean sojourn time in the macro state corresponding to failure mode $\alpha$ is:

$$\overline{t_{S,\alpha}} = \int_0^\infty \mathrm{t}\ \mathrm{f}_{\mu^\alpha}(\mathrm{t})\ \mathrm{dt}$$

where $f_{\mu^\alpha}(t)$ is the density function of the repair process corresponding to failure mode $\alpha$. Once the failure rates $\lambda_\alpha$ are constant and the state probabilities verify:

$$P(s_{\mathrm{up}}) + \sum_{s \in F_M} P(s) = 1$$

where $\boldsymbol{F}_M$ is the set of failure states of $M$, the probability of the normal operating state may be obtained from:

$$P(\mathrm{s_{up}}) = \frac{1}{1 + \sum_\alpha \lambda^\alpha \int_0^\infty \mathrm{t}\ \mathrm{f}_{\mu^\alpha}(\mathrm{t})\mathrm{dt}} \tag{12}$$

Now, the frequency of the initial failure state corresponding to a particular failure mode $\alpha$ may be readily obtained from:

$$\Lambda(s_0^\alpha) = \lambda^\alpha P(s_{up}) \tag{13}$$

*4.2.2. Exponential failure processes and several repair processes.* Figure 9 provides an example of a model where the set of states corresponding to the same failure mode present different repair processes: in $s_0$, there is no active repair process; after $s_1$ is achieved, the repair process $p\mu_1$ is activated; if the catastrophic failure state $s_4$ is achieved, a different repair process $p\mu_2$ has to be performed in order to restore the normal operation of the system. In this situation, it is no longer possible to consider the mean sojourn time in the macro states as before. Instead, the sojourn time in each failure state should be considered. According to (8), the probability of a failure state $s$ may be obtained from:

$$\mathrm{P}(s) = \mathrm{P}(s_{up})\,\lambda \sum_{\psi \in \Psi_s} P(\psi) \times \overline{t_s^\psi} \tag{14}$$

where $\lambda$ is the failure process corresponding to the failure mode that leads to $s$. Therefore, the state probability may be obtained from the following set of equations:

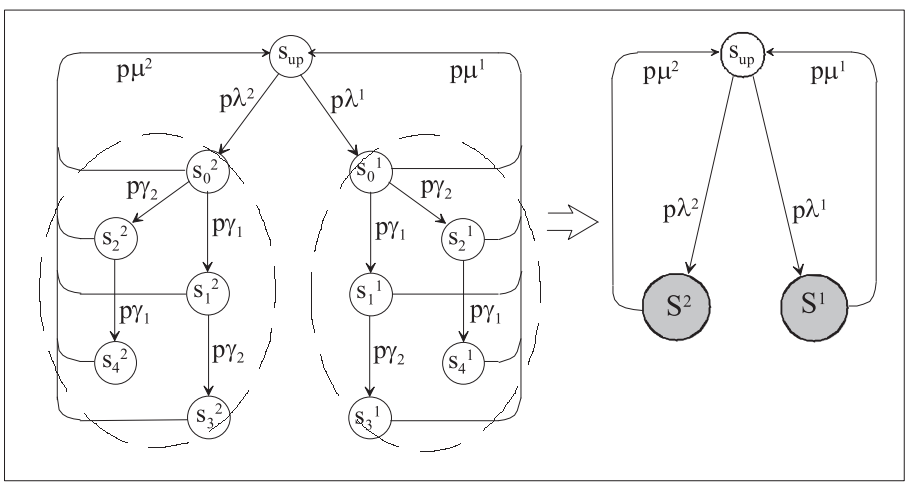


**Figure 8.** Macro-failure states

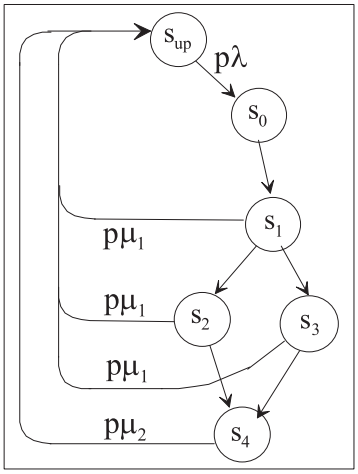


**Figure 9.** Several repair processes

$$P(s_{up}) + \sum_{s \in F_M} P(s) = 1 \tag{15}$$

$$P(s) = P(s_{up})\,\lambda \sum_{\psi \in \Psi_s} P(\psi) \times \overline{t_s^{\psi}}$$

### 4.3. *Evaluating reliability in mission systems*

The expressions obtained in the previous chapters are related to the steady-state values of the frequency and probability of the failure states. Here, the evaluation of the probability of arrival, at a particular failure state within a specified time frame, will be addressed. This is particularly useful in the context of *mission systems*, which are expected to perform a function for a limited period of time. For these systems, the fundamental reliability measure is the probability of arrival at a catastrophic failure state $s$ before a specified period of time $\Delta$ has elapsed since the beginning of the mission, $\Lambda_M^T(s, \Delta)$.

To obtain the analytical expressions for this reliability measure, two main adjustments to the procedures presented above are required: the transitions from the regenerative state $s_{up}$ to the initial failure states should be included in the trajectories; and the upper limit of the integrals should be bounded to $\Delta$. Let us suppose that $s_n^{\alpha}$ is a catastrophic failure state subsequent to failure mode $\alpha$, and $\Psi_n^{\alpha}$ denotes the set of trajectories that lead from $s_{up}$ to $s_n^{\alpha}$. Then, the probability $\boldsymbol{P}(s_n^{\alpha},\Delta)$ of arrival at $s_n^{\alpha}$ before $\Delta$ is:

$$\mathbf{P}(s_n^{\alpha}, \Delta) = \sum_{\psi \in \Psi_n^{\alpha}} P(\psi, \Delta) \tag{16}$$

where the probability of arriving at $s_n^{\alpha}$ at $t < \Delta$ following trajectory $\psi$, i.e. $\boldsymbol{P}(\psi,\Delta)$ may be directly obtained from a set of expressions similar to those of chapter 4.1, but now the upper integration limits are bounded to $\Delta$:

$$\mathrm{P}(\psi, \Delta) = \int_0^{\Delta} T(t_1) \int_{t_1}^{\Delta} T(t_2) \,..\int_{t_n-1}^{\Delta} T(t_n)\, dt_n \,..dt_2\, dt_1 \tag{17}$$

$$T(t_i) = f_{p_i}\left(t_i - t_{p_i}^0\right) \left( \prod_{\substack{p \in \Omega_{i-1} \\ p \notin \Omega_i \\ p \neq p_i}} \int_{t_i}^{\infty} f_p\left(\tau - t_p^0\right) d\tau \right) \tag{18}$$

Now, suppose that FC is the set of catastrophic failure states in a non-repairable mission system $S$. The reliability of $S$ for a mission with duration $\Delta$ may be obtained from:

$$R(\Delta) = 1 - \sum_{s \in FC} \sum_{\psi \in \psi^s} P\left(\psi^s, \Delta\right) \tag{19}$$

## 5. Application example

This section presents several results regarding the evaluation of the model represented in Figure 10. It is assumed that $s_4$ is a catastrophic failure state and that its probability and frequency are to be evaluated. The analytical expressions for these two measures have already been introduced in subchapters 4.1.2 and 4.2.1.

Two scenarios will be considered here for illustrative purposes:

- all the processes present exponential distributions (scenario 1)
- the repair and delay processes present third-order Erlang distributions (scenario 2)

For the sake of simplicity, it will also be assumed that the three delay processes are identical and that their mean $\overline{m_{p\gamma}}$ is 3 h. For the mean of the repair processes, several values will be considered ranging from $\overline{m_{p\mu}}/2$ to $4\,\overline{m_{p\gamma}}$. For the failure rate, a typical value of $10^{-3}\,h^{-1}$ is assumed.

Figures 11.a and 11.b represent the evolution of the probability and of the frequency of the catastrophic failure state with the ratio $\rho = \overline{m_{p\mu}}/\overline{m_{p\gamma}}$, for the two scenarios.

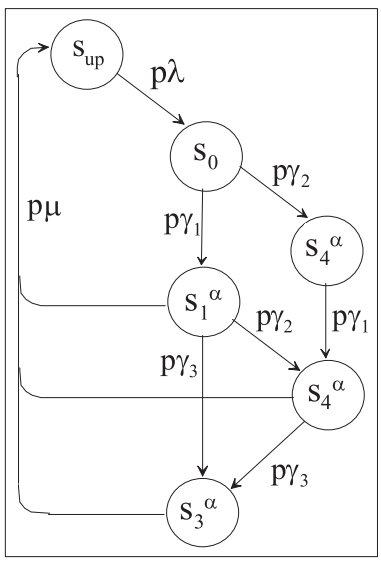


**Figure 10.** Application example

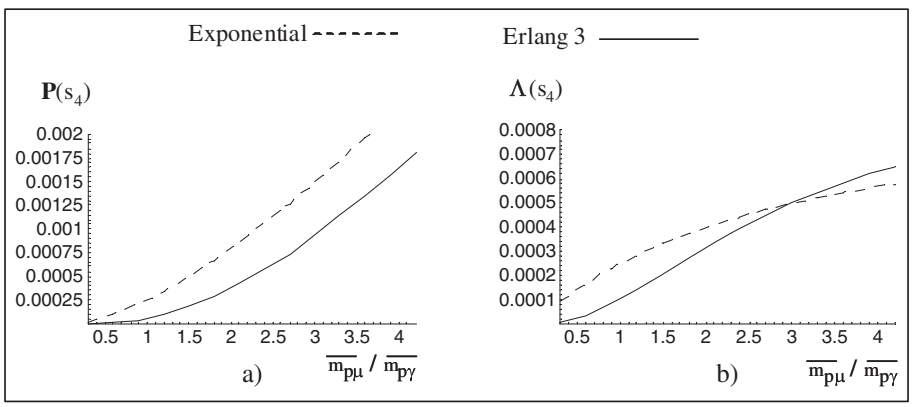


**Figure 11.** Numerical results

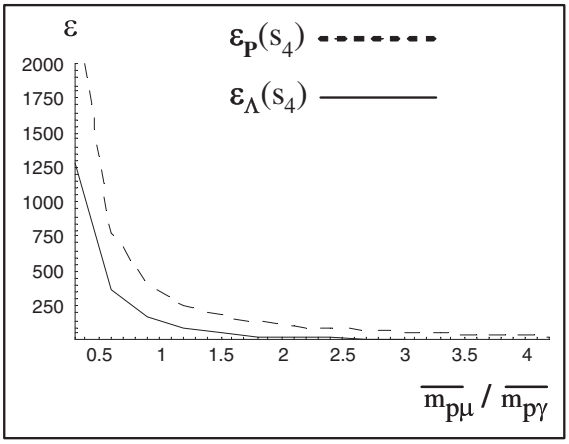


**Figure 12.** Markov assumption error

Figure 12 provides another important result. It shows the error that will be introduced in the evaluation of a system presenting the non-Markovian behavior corresponding to scenario 2, using the Markovian model of scenario 1 (frequent in reliability analyses). The error $\varepsilon$ in a reliability measurement $R$ is calculated from:

$$\varepsilon_R = \frac{R_1 - R_2}{R_2}$$

where $R_1$ and $R_2$ are the values corresponding to the two scenarios. These results reinforce the idea that when a model contains concurrent processes with non-exponential distributions, the use of non-Markovian techniques is mandatory. In fact, even with this simple system, the error may be higher than 1000%.

## 6. Discusson

The analysis of systems with a non-Markovian behavior is an intensely discussed problem, and, in recent years, several methods have been developed targeted at more and more restricted classes of stochastic models and seeking to improve the performance of the algorithms and the computational power required.

The paper has presented an approach for the evaluation of reliability and performance in systems containing a Markov regenerative state (corresponding to a normal operation) and multiple concurrent processes with generalized distributions. The approach is mainly targeted at the evaluation of industrial engineering systems which, as discussed in Section 2, typically contain a large number of components and provide services to multiple users. As the users normally tolerate a temporary unavailability of services, these systems will present a failure delayed behavior pattern.

A well-established analytical solution for the transient and steady-state evaluation of regenerative Markov systems is described in [1]. This solution allows immediate, exponentially distributed and generally distributed timed transitions to be considered. However, all the non-exponential processes should be enabled at the same instant. As it has been shown, the approach presented here does not impose this important restriction.

Other approaches to the evaluation of non-Markovian systems require the use of additional variables, whose number increases quickly when the model contains several concurrent processes with narrow hyper-exponential distributions – deterministic or quasi-deterministic processes –, as is the case of the device of stages.[21] In these conditions, the approach presented here may offer a more straightforward solution. In fact, the analytical expressions for the relevant reliability measurements may be obtained using a systematic procedure taken directly from the structure of the model and the distributions of the stochastic processes. No auxiliary variables are required, and the expressions may be evaluated using general purpose mathematical tools.

Another important characteristic of the new approach is the fact that the analytical expressions for the reliability measurements are insensitive to the actual shape of the distributions of the stochastic processes. Therefore, in the evaluation of systems containing deterministic or quasi-deterministic processes, this approach may offer a more effective solution.

The paper has primarily focused on the steady-state evaluation (probability and frequency of the failure states) of ergodic repairable systems. However, as it was shown in chapter 4.4, the approach may be extended in order to address the evaluation of reliability in transient conditions, as required in the analysis of mission-oriented systems.

In the paper, it was also assumed that the failure and the repair processes affect the global state of the system as a whole. In fact, it was implicit in the meta-model of Figure 2 that (i) the execution of any one of the repair processes lead the system to the regenerative state $s_{up}$, (ii) the failure processes are active solely in $s_{up}$ and (iii) each failure process corresponds to an exclusive failure mode.

However, the approach can also be extended in order to address the evaluation of multi-component systems with partial failure and repair processes, as is the case of redundant structures. Assessing these systems can be difficult due to the fact that the set of failure states do not form an acyclic graph. As a consequence, the same failure state may be visited several times between the instant when a failure occurs, before the system returns to the regenerative state $s_{up}$, so that the number of states within a trajectory is no longer bounded. A possible approach to overcome this difficulty consists of replacing the state-diagram with a state-tree, so that each passage in a non-regenerative state is considered a different state within the trajectory. If it can be assumed that the mean-time-to-failure is much longer than the mean-time-to-repair (as is usually the case), the probabilities of the additional states will decrease quickly. Therefore, for any arbitrary small value of the error tolerated in the evaluation, it will always be possible to truncate the tree and limit the number of states in each trajectory.

The material presented in the paper is mainly focused on industrial engineering systems, but a similar approach may also be applied to other classes of systems. In particular, complex business environments are now being studied in which a network of interacting activities and resources provides business services to a number of heterogeneous users. In normal operating conditions, each activity operates at an expected performance level. However, due to failures or other disturbing events such as demand fluctuations, the performance level of one or more activities may decay, disturbing the downstream activities and the services provided to the users.

The propagation of failures throughout the activity network will follow two main patterns: a logistical patter, which is the horizontal propagation between activities linked by a producer/consumer relationship, and a management pattern, which is a vertical propagation between operational and supervision activities. These propagation processes will normally involve propagation delays, and therefore the behavior of these activity networks will also follow a failure delayed pattern. The extension of the algorithms presented in the paper is now being investigated in order to cope with this new application domain.

## 7. Conclusion

The context, aims and organization of this paper were presented in Section 1. In the first part of Section 2, the concept of failure delayed system was introduced, and three examples of this class of systems for different engineering domains were discussed. In the second part of this section, the meta-model specifying the structure of the FDS models was discussed.

In Section 3, the existing techniques to assess non-Markovian systems were reviewed, along with their shortcomings. Then, the fundamentals of the new approach were introduced in Section 4. Illustrative examples were provided for the evaluation of the failure state probabilities and frequencies. In the final part of this section, the assessment of mission critical systems was also discussed.

Section 5 presents some numerical results that reinforce the idea that the usual assumption in reliability analyses, according to which all the stochastic processes are exponentially distributed, leads to very significant calculation errors in the case of failure delayed systems. Therefore, the use of non-Markovian techniques becomes mandatory.

Finally, Section 6 discusses the strengths and weaknesses of the approach presented in the paper and shows how it can be extended to multi-component systems.

## References

1. Puliafito A, Scarpa M, Trivedi K. "Petri nets with k simultaneously enabled generally distributed timed transitions", *Performance Evaluation*, Feb 1998; **32**(1):1–34.
2. Balbo G. "Introduction to generalized stochastic Petri nets, Formal Methods for Performance Evaluation", 7th International School on Formal Methods for the Design of Computer, Communication and Software Systems, SFM 2007. Advanced Lecture (Lecture Notes in Computer Science) 2007; **4486**:83–131.
3. Bobbio A, Telek M. "Non-exponential stochastic Petri nets: an overview of methods and techniques", *Computer Systems Science and Engineering*, Nov 1998; **13**(6):339–351.
4. Carneiro J, Ferrarini L. "Reliability analysis of power system based on generalized stochastic Petri nets", Proceedings of the 10th International Conference on Probablistic Methods Applied to Power Systems (PMAPS 2008) 2008; 6.
5. Haiyan Z, Shengqiang L. "Modeling and analysis of reverse supply chain based on generalized stochastic Petri nets", 2009 International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII 2009) 2009; 437–40.
6. Zhan H, Gu J. "Study of the normal generalized stochastic Petri nets and its application in testing system", IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 06CH37714C) 2006; 6.
7. Billinton R, Li W. *Reliability Assessment of Electrical Power Systems Using Monte Carlo Methods*, Plenum Press, New York, 1994.
8. Windebank E. A Monte Carlo simulation method versus a general analytical method for determining reliability measures of repairable systems. *Reliability Engineering* 1983; **5**(2):73–81.
9. Nunes E, Faria JA, Matos MA. "*A comparative analysis of dependability assessment methodologies*" Proceedings of the λμ13 ESREL Conference, Lyon, France, May 2002.
10. Faria JA. "Dependability modeling, analysis and evaluation of industrial information systems", Ph.D. Thesis, Porto University, Portugal 1997 (in Portuguese).
11. Faria JA, Matos MA. "*A new approach for reliability analysis of non-Markovian systems*", Proceedings of the Probability Methods Applications to Power Systems Conference 2000, Funchal, Portugal 2000.
12. Faria JA, Nunes E, Matos MA. "Optimal dimensioning of work-in-process buffers", Proceedings of the International Conference on Industrial Engineering and Production, Portugal, May 2003.
13. Faria JA and Matos MA. "Availability Analysis and Design of Business Information Systems", *International Journal of Business and Information* 2006; **1**(1).

14. Wang W, Banjevic D, Pecht M. A multi-component and multi-failure mode inspection model based on the delay time concept, *Reliability Engineering and System Safety*, Aug. 2010; **95**(8):912–20.
15. Mourani I, Hennequin S, Xie X. "Simulation-based optimization of a single-stage failure-prone manufacturing system with transportation delay", *International Journal of Production Economics*, March 2008; **112**(1):26–36.
16. Limnios N. "Failure Delay Systems Reliability Modelling", in *Systems Reliability Assessment*, Edited by Colombo AG, Saiz de Bustamante A, ECSC Brussels, 1990.
17. Wu L. "Operational models for the evaluation of degradable computing systems", *Performance Evaluation Review (USA)*, Winter 1982–1983; **11** (4):179–85.
18. Bobbio A. "Stochastic Reward Models in Performance/Reliability Analysis", *Journal on Communications*, January 1992; **XLIII**:27–35.
19. Singh C and Billinton R. *System Reliability Modelling and Evaluation*, Hutchinson & Co. London, UK, 1977.
20. Pages A and Gondran M. *Fiabilité des systèmes*. Eyrolles, France, 1980.
21. Cox DR and Miller HD. *The Theory of Stochastic Processes*, Chapman and Hall, London, UK, 1965.
22. Laprie JC. "Prévision de la Sûreté de Fonctionnement et Architecture des Structures Numériques Temps Réel Réparables", Ph.D Thesis, Université Paul Sabatier, Toulouse, France 1975.
23. Singh C, Billinton R, Lee SY. "The method of stages for non-Markov models", *IEEE Transactions on Reliability* 1977; **R-26**(2):135-7.
24. Haverkort BR and Trivedi KS. "Specification techniques for markov reward models", *Discrete Event ynamic Systems: Theory and Applications*, Jul, 1993; **3**(2–3):219.
25. Cumani A. "*ESP – a package for the evaluation of stochastic Petri nets with phase type distributed transition times*", Proceedings of the International Workshop Timed Petri Nets, Torino, Italy, 1985; 144–151.
26. Neuts MF. *Matrix Geometric Solutions in Stochastic Models*, John Hopkins University Press, Baltimore USA, 1981.
27. Molloy M. "Performance analysis using stochastic Petri nets", *IEEE Transactions on Computers*, September 1982; **C-31**(9):913–17.
28. Marsan MA and Chiola G. "On Petri nets with deterministic and exponentially distributed firing times", *Lecture Notes in Computer Science*, Springer Verlag, 1987; **266**:132–245.
29. Choi H, Kulkarni VG and Trivedi KS. "Markov regenerative stochastic Petri nets", *Performance Evaluation*, May, 1994; **20**(1–3):337–357.
30. German R and Lindemann C. "Analysis of stochastic Petri nets by the method of supplementary variables", *Performance Evaluation*, May, 1994; **20**(1–3):317–335.
31. Scarpa M, Distefano S, Puliafito A, "A parallel approach for the solution of non-Markovian Petri nets", *Recent Advances in Parallel Virtual Machine and Message Passing Interface, Lecture Notes in Computer Science*, Springer Verlag Berlin Heidelberg, 2003; **2840**:196–203.
32. Horvath A, Puliafito A, Scarpa M, Telek M. "Analysis and evaluation of non-Markovian Stochastic Petri Nets", *Computer Performance Evaluation, Proceedings Lectures notes in Computer Science* 2000; **1786**:171–187.
33. Scarpa M, Longo F. "Applying symbolic technques to the representation of non-markovian models with continuous PH distributions", 6th European Performance Engineering Workshop, Springer Verlag LNCS, 2009 Feb 1998; **5652**:44–58.

## Annex Initial states frequency for non-exponential failures

In chapter 4.3, the procedure to determine the frequencies of the initial failure states was discussed considering exponential failure processes. This annex presents the procedures for the non-exponential failure processes. As before, two situations will be considered: a single repair process common to all the states below each failure process, and different repair processes for each state.

## A1. Non-exponential failure processes and common repair process

In this situation, the failure states may be grouped into macro states, as in 4.3.1. However, the frequencies of the initial failure states cannot be obtained directly from (13) because the failure rates are not constant. Therefore, a different approach, based on the frequency and on the sojourn time of state $s_{up}$, $\Lambda(s_{up})$ and $\overline{t_{s_{up}}}$, will be adopted. The probability of $s_{up}$ may be obtained from:

$$P(s_{up}) = \Lambda\left(s_{up}\right)\overline{t_{s_{up}}}$$

If $\Omega_{up}$ is the set of failure processes that are active in $s_{up}$ and $f_{p\lambda}(t)$ is the non-exponential density function of the failure process $p_\lambda$ ($p_\lambda \in \Omega_{up}$), then the probability of a failure occurring at an instant $t$ due to $p\lambda$ is:

$$f_{p\lambda}(t) \prod_{\substack{p\lambda' \in \Omega_{up} \\ p\lambda' \neq p\lambda}} f_{p\lambda'}(t)\,dt \tag{A.1}$$

Therefore, the mean sojourn time in $s_{up}$ when the set of failure processes is considered may then be obtained from:

$$\overline{t_{s_{up}}} = \sum_{p\lambda \in \Omega_{up}} \int_0^\infty t f_{p\lambda}(t) \prod_{\substack{p\lambda' \in \Omega_{up} \\ p\lambda' \neq p\lambda}} f_{p\lambda'}(t)\,dt \tag{A.2}$$

The frequency of arrival at the macro state $S^\alpha$ results from the product of $\Lambda(s_{up})$ and the probability of transition from $s_{up}$ to $s_0^\alpha$:

$$\Lambda(s_0^\alpha) = \Lambda(s_{up})P(s_{up} \to s_0^\alpha) \tag{A.3}$$

If $p\lambda^\alpha$ is the failure process that causes this transition, then:

$$\Lambda(s_0^{\alpha}) = \Lambda(s_{up}) \int_0^{\infty} f_{p\lambda^{\alpha}}(t) \prod_{\substack{p\lambda' \in \Omega_{up} \\ p\lambda' \neq p\lambda^{\alpha}}} \int_t^{\infty} f_{p\lambda'}\left(t'\right) dt' dt \tag{A.4}$$

The probability of each macro state may be obtained from the product of its frequency and the mean sojourn time, which is:

$$P(S_0^{\alpha}) = \Lambda(s_0^{\alpha}) \int_0^{\infty} t f_{\mu^{\alpha}}(t) dt \tag{A.5}$$

The combination of (1) and (4) allows the frequency of arrival to $s_{up}$ to be obtained from:

$$\Lambda(s_{up}) = \frac{1}{\overline{t_{s_{up}}} + \sum_{p^{\lambda\alpha} \in \Omega_{up}} \left( \int_0^{\infty} f_{p\lambda^{\alpha}}(t) \prod_{\substack{p\lambda' \in \Omega_{up} \\ p\lambda' \neq p\lambda^{\alpha}}} \int_t^{\infty} f_{p\lambda'}(t') dt' dt \right) \int_0^{\infty} t f_{p\mu^{\alpha}}(t) dt} \tag{A.6}$$

Once $\Lambda(s_{up})$ is known, the frequencies of arrival at the initial failure states may then be readily evaluated using expression (13).

## A2. Non-exponential failure processes and several repair process

In this case, the set of equations considered in chapter 4.3.2 may also be employed but, now, $\boldsymbol{P}(s_{up})$ has to be replaced by $\Lambda(s_{up})\ \overline{t_{s_{up}}}$:

$$\Lambda(s_{up})\ \overline{t_{s_{up}}} + \sum_{s \in F_S} P(s) = 1 \tag{A.7}$$

$$P(s) = \Lambda(s_{up}) P(s_{up} \rightarrow s_0^{\alpha}) \sum_{\psi \in \Psi_s} P(\psi)\ \overline{t_s^{\psi}}, \quad \text{for } s \in F_M$$

The combination of these two expressions makes it possible to obtain $\boldsymbol{\Lambda}(s_{up})$:

$$\Lambda(s_{up}) = \frac{1}{\overline{t_{s_{up}}} \sum_{s \in F_M} \sum_{\psi \in \Psi_s} P(s_{up} \rightarrow s_0^{\alpha}) P(\psi)\ \overline{t_s^{\psi}}} \tag{A.8}$$

The frequencies of arrival at the initial failure states may be obtained from expression (2), once $\boldsymbol{\Lambda}(s_{up})$ is known.

*Authors' biographies*

**Jose Faria** holds a PhD in Electrical Engineering from the Faculty of Engineering of the University of Porto (FEUP) (1997). The PhD focused on reliability modeling, analysis and evaluation of the integrated management information system of a large European car manufacturer. Currently, he is Assistant Professor at FEUP and Researcher at INESC Tec, an applied private non-profit research Institute. His research interests include quality and reliability management systems, in particular, reliability analysis methods and tools for non-Markovian systems.

**Américo Azevedo** is Associate Professor in the Department of Industrial Engineering and Management of FEUP and a Research Manager at INESC Tec. He holds a degree in Electrical and Computers Engineering (1988) and a PhD in Production Planning (2000), both from University of Porto. He managed several R&D project funded by the European Union and Portuguese public and private institutions. He has been also reviewer and evaluator of several international R&D industrial projects and member of several scientific program committees. His research interests are in the domain of operations management and industrial processes engineering.