Fluid Limit for the Machine Repairman Model with Phase-Type Distributions

Laura Aspirot[1], Ernesto Mordecki[1], and Gerardo Rubino[2] **

[1] Universidad de la República, Montevideo, Uruguay,
[2] INRIA, Rennes, France.

We consider the Machine Repairman Model with N working units that break randomly and independently according to a phase-type distribution. Broken units go to one repairman where the repair time also follows a phase-type distribution. We are interested in the behavior of the number of working units when N is large. For this purpose, we explore the fluid limit of this stochastic process appropriately scaled by dividing it by N.

This problem presents two main difficulties: two different time scales and discontinuous transition rates. Different time scales appear because, since there is only one repairman, the phase at the repairman changes at a rate of order N, whereas the total scaled number of working units changes at a rate of order 1. Then, the repairman changes N times faster than, for example, the total number of working units in the system, so in the fluid limit the behavior at the repairman is averaged. In addition transition rates are discontinuous because of idle periods at the repairman, and hinders the limit description by an ODE.

We prove that the multidimensional Markovian process describing the system evolution converges to a deterministic process with piecewise smooth trajectories. We analyze the deterministic system by studying its fixed points, and we find three different behaviors depending only on the expected values of the phase-type distributions involved. We also find that in each case the stationary behavior of the scaled system converges to the unique fixed point that is a global attractor. Proofs rely on martingale theorems, properties of phase-type distributions and on characteristics of piecewise smooth dynamical systems. We also illustrate these results with numerical simulations.

1 Introduction

The Machine Repairman Model. The Machine Repairman Model (MRM) is a basic Markovian queue representing a finite number N of machines that can fail independently and, then, be repaired by a repair facility. The latter, in the basic model, is composed of a single repairing server with a waiting room for failed machines managed in FIFO order, in case the repairing server is busy when units fail. In Kendall's notation, this is the $M/M/1//N$ model, specifying that lifetimes and repair times are exponentially distributed. This model is well known and widely studied in queuing theory and in many applications, as for example

** This paper received partial support from the STIC-AmSud project "AMMA" and project FCE-2-2011-1-6739.

in telecommunications or in reliability. Almost all these studies look at the queue in equilibrium. The model is a precursor of the development of queuing network theory, motivated first in computer science. In particular, Scherr from IBM used it in 1972 for analyzing the S360 OS (see [1]). Many extensions of the basic model have been studied, considering more than one repairing server, different queuing disciplines, and other probability distributions for the life-time or for the repair time. We refer to [2] for further reference on the problem.

Fluid Limits. Fluid limits is a widely developed technique that proves very useful for the study of large Markov systems. Many of these systems, under a suitable scaling, have a deterministic limit given by an ordinary differential equation (ODE). As an example, let us consider the fluid limit for a $M/M/1$ queue [3] with arrival rate λ and service rate μ. Let $X(t)$ be the number of units at time t and let $\widehat{X}^N(t) = X(Nt)/N$ be the scaled number of units. Time is accelerated by a factor N, and the initial state is also scaled by the same factor. If the scaled initial condition converges with N, then the process $\widehat{X}^N$ can be approximated, for large N, by the deterministic solution to $\dot{x} = \lambda - \mu$ if $x > 0$, $\dot{x} = 0$ if $x = 0$. For $\lambda < \mu$ the equation defines a piecewise smooth dynamical system, with a solution for the initial condition $x(0)$ that is smooth on $[0, x(0)/(\mu - \lambda))$ and $(x(0)/(\mu - \lambda), \infty)$. If the initial condition is 0, the solution remains at zero.

Other examples from [3] are the $M/M/\infty$ queue and the $M/M/N/N$ queue. Let λN be the arrival rate and μ the service rate in both cases and let $\widetilde{X}^N$ be the number of units in the system. The scaling is different from the $M/M/1$, as time is not scaled, only the arrival rate is accelerated, and the total service rate scales with the number of units in the system. The scaled number of units $X^N = \widetilde{X}^N/N$ converges to the solution to $\dot{x} = \lambda - \mu x$ for the $M/M/\infty$ queue, and to the solution to $\dot{x} = \lambda - \mu x$, if $x < 1$, $\dot{x} = 0$ if $x = 1$, in the $M/M/N/N$ model. In the last case, we find again a piecewise smooth dynamical system, that converges exponentially fast to $\rho = \lambda/\mu$ if $\rho < 1$ and to 1 if $\rho \geq 1$.

Looking for fluid limits is a suitable approach to repairman problems, as shown in [4], where the MRM model with two repair facilities, studied by Iglehart and Lemoine in [5, 6], is analyzed using these tools. In [5] there are N operating units subject to exponential failures with parameter λ. Failures are of type 1 (resp. 2) with probability p (resp. $1 - p = q$). If failure is of type i ($i = 1, 2$) the unit goes to repair facility i that has s_N^i exponential servers, each one with exponential service rate μ_i. The goal is to study the stationary distribution when $s_i^N \sim N s_i$ as $N \to \infty$, $i = 1, 2$. The behavior of the system is characterized in terms of the parameter set that defines the model. In addition, the case with spares is presented in [6]. The original approach consists in approximating the number of units in each repair facility by binomial random variables, and then proving for them a law of large numbers and a central limit theorem. Kurtz, in [4] studies the same model with a fluid limit approach, proves convergence to a deterministic system, and goes a step beyond, considering the rate of this approximation through a central limit theorem-type result. The same discussion as in [5] in terms of the different parameter set follows from the study of the ODE's fixed point.

Contributions. In this paper we analyze a repairman problem with N working units that break randomly and independently according to a phase-type distribution. Broken units go to one repairman where the repair time also follows a phase-type distribution (that is, a $PH/PH/1//N$ model). We consider a scaled system, dividing the number of broken units and the number of working units in each phase by the total number of units N. The scaled process has a deterministic limit when N goes to infinity. The first problem that the model presents is that there are two time scales: the repairman changes its phase at a rate of order N, whereas the total scaled number of working units changes at a rate of order 1. Another problem is that transition rates are discontinuous because of idle periods at the repairman (this second issue is also present in the models $M/M/1$ and $M/M/N/N$ described above).

In our main result we prove that the scaled Markovian multi-dimensional process describing the system dynamics converges to the solution of an ODE as $N \to \infty$. The convergence is in probability and takes place uniformly in compact time intervals (usually denoted *u.c.p.* convergence), and the deterministic limit, the solution to the ODE, is only picewise smooth. We analyze the properties of this limit, and we prove the convergence in probability of the system in stationary regime to the ODE's fixed point. We also find that this fixed point only depends on the repair time by its mean. As a matter of fact, recall that when in equilibrium, if the repair times are exponentially distributed, the distribution of the number of broken machines has the insensitivity property with respect to the life-time distribution (only the latter's mean appears in the former). For an example of what happens here, see the end of Section 4.

Related Work. Fluid limits, density dependent population process, approximation by differential equations of Markov chains, are all widely developed objects. As a general reference we refer to the monograph by Ethier and Kurtz [7] and references therein. The main approach there consists in a random change of time that allows to write the original Markov chain as a sum of independent unit Poisson processes evaluated at random times. Darling and Norris [8] present a survey about approximation of Markov chains by differential equations with an approach based on martingales. However, [8] does not deal with discontinuous transition rates. We refer to the books by Shwartz and Weiss [9], and Robert [3] for extensive analysis of the $M/M/1$ and the $M/M/\infty$ queues, including deterministic limits, asymptotic distributions and large deviations results. In particular, in [9] the discontinuous transition rates and different time scales are considered. The latter situation is also considered in [10] and [11]. We mostly follow the approach of [12] to deal with discontinuous transition rates, which considers hybrid limits for continuous time Markov chains with discontinuous transition rates, with examples in queuing and epidemic models. Discontinuous transition rates are also studied in [13, 14]. Paper [13] analyzes queuing networks with batch services and batch arrivals, that lead to fluid limits represented by ODEs with discontinuous right hand sides. Paper [14] models optical packet switches, where the queuing model lead to ODEs with discontinuous right hand sides, and where they consider both exponential and phase-type distributions for

packet lengths. Convergence to the fixed points is studied in several works (e.g. [9, 13, 15]). However there are counterexamples where there is no convergence of invariant distributions to fixed points [15]. There are general results with quite strong hypotheses as in [16], where reversibility is assumed in order to prove convergence to the fixed point.

Organization of the Paper. In Section 2 we present our model. In Section 3 we compute the drift, and we describe the ODE that defines the fluid limit. In Section 4 we state our main results about convergence and description of the fluid limit. In Section 5 we show several numerical examples that illustrates the results and we conclude in Section 6. Proofs are provided in Section 7.

2 Model

We consider N identical units that work independently, as part of some system, that randomly fail and that get repaired. Broken units go to a repairman with one server, where the repair time is a random variable with a phase-type distribution. After being repaired units start working again. A given unit's life-times are independent identically distributed random variables also with phase-type distribution. We want to describe the number of working units in each phase before failure and the number of broken units in the system. We consider the system for large N, with the repair time scaled by N. This means that the repair time per unit decreases as N increases. We describe the limit behavior of the system when N goes to infinity. The assumption of phase-type distributions allows to represent a wide variety of systems, as phase-type distributions well approximate many positive distributions, allowing, at the same time, to exploit properties of exponential distributions and Markov structure. Concerning the repairing facility, we consider a single server with the service time also scaled according with the number of units, and we find a different behavior that for the model scaled both in the number of units and the number of servers.

Phase-Type Distributions. A phase-type distribution with k phases is the distribution of the time to absorption in a finite Markov chain with $k+1$ states, where one state is absorbing and the remaining k states are transient. With an appropriate numbering of the states, the transient Markov chain has infinitesimal generator $\widehat{M} = \left(\begin{array}{c|c} M & m \\ \hline 0 & 0 \end{array}\right)$, where M is a $k \times k$ matrix, and $m = -M\mathbb{1}$, with $\mathbb{1}$ the column vector of ones in $\mathbb{R}^k$. The initial distribution for the transient Markov chain is a column vector $(r, 0) \in \mathbb{R}^{k+1}$, where r is the initial distribution among the transient states. We represent this phase-type distribution by (k, r, M). We refer to [17] for further background about phase-type distributions.

Variables. We describe the distributions and variables involved in the model. All vectors are column vectors.

Repair time. The repair time follows a phase-type distribution (m, p, NA), with m phases, matrix NA (where A is a fixed matrix and N is the scaling factor) and initial distribution p. We denote $Na = N(a_1, \ldots, a_m) = -NA\mathbb{1}$.

Life-time. The life-time for each unit is phase-type (n, q, B), with n phases, matrix B and initial distribution q. We denote $b = (b_1, \ldots, b_n) = -B\mathbb{1}$.

Working units. $\widetilde{X}_i^N(t)$ is the number of units working in phase i at time t, for $i = 1, \ldots, n$, and $\widetilde{X}^N = (\widetilde{X}_1^N, \ldots, \widetilde{X}_n^N)$.

Repairman state. $\widetilde{Z}_i^N(t)$ is number of units being repaired in phase i for $i = 1, \ldots, m$ ($\widetilde{Z}_i^N(t)$ is zero or one), and $\widetilde{Z}^N = (\widetilde{Z}_1^N, \ldots, \widetilde{Z}_m^N)$.

Waiting queue. $\widetilde{Y}^N(t)$ is the number of broken units waiting to be repaired.

Scaling. We consider the scaling: $X^N = \dfrac{1}{N}\widetilde{Y}^N,\ Y^N = \dfrac{1}{N}\widetilde{Y}^N,\ Z^N = \dfrac{1}{N}\widetilde{Z}^N$.

Note that $\mathbb{1}^T\widetilde{Z}^N(t) + \chi_{\{\mathbb{1}^T\widetilde{X}^N(t)=N\}} = 1$, where $\chi_{\mathcal{P}}$ is the indicator function of the predicate $\mathcal{P}$. That means that units are all working, or there is one unit being repaired at the server. In addition, $\mathbb{1}^T\widetilde{X}^N(t) + \widetilde{Y}^N(t) + \sum_{i=1}^m \widetilde{Z}_i^N(t) = N$.

Model Dynamics. $\widetilde{U}^N = \left(\widetilde{X}_1^N, \ldots, \widetilde{X}_n^N, \widetilde{Y}^N, \widetilde{Z}_1^N, \ldots, \widetilde{Z}_m^N\right)$ is a Markov chain.

We denote by $e^i \in \mathbb{R}^{n+m+1}$ the vector $e^i = (e_1^i, \ldots, e_{n+m+1}^i)$ with $e_i^i = 1$ and $e_j^i = 0$ for $i \neq j$, $i, j = 1, \ldots, n+m+1$. We describe the possible transitions for this Markov chain from a vector $\widetilde{u}$ in the state space, with its corresponding transition rates. The vector $\widetilde{u} = (\widetilde{x}, \widetilde{y}, \widetilde{z})$ with $\widetilde{x} = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$, $\widetilde{z} = (\widetilde{z}_1, \ldots, \widetilde{z}_m)$, with $\widetilde{x}_i \in \{0, 1, \ldots, N\}$ for all $i = 1, \ldots n$, $\widetilde{y} \in \{0, 1, \ldots, N\}$ and $\widetilde{z}_i \in \{0, 1\}$ for all $i = 1, \ldots m$.

- A working unit in phase i changes to phase j. For $i, j = 1, \ldots, n$, transition $e^j - e^i$ occurs with rate $b_{ij}\widetilde{x}_i$.
- A working unit in phase i breaks and goes to the buffer. The unit goes to the buffer because there is one unit in service. For $i = 1, \ldots, n$, transition $e^{n+1} - e^i$ occurs at rate $b_i\widetilde{x}_i\chi_{\{\mathbb{1}^T\widetilde{x}<N\}}$.
- A working unit in phase i breaks and starts being repaired. The unit starts being repaired because the repairman is idle, at phase j. For $i = 1, \ldots, n$, $j = 1, \ldots, m$, transition $e^{n+1+j} - e^i$ occurs at rate $b_i p_j \widetilde{x}_i \chi_{\{\mathbb{1}^T\widetilde{x}=N\}}$.
- A unit that is being repaired in phase i changes to phase j. For $i, j = 1, \ldots, m$, transition $e^{n+1+j} - e^{n+1+i}$ occurs at rate $Na_{ij}\widetilde{z}_i$.
- A unit that is being repaired in phase i ends its service and starts working at phase j with the buffer empty. If the buffer is empty, nobody starts being served and for $j = 1, \ldots, n$, $i = 1, \ldots, m$, the transition $e^j - e^{n+1+i}$ occurs at rate $Na_i q_j \widetilde{z}_i \chi_{\{\widetilde{y}=0\}}$.
- A unit that is being repaired in phase i ends its service and starts working at phase j with nonempty buffer. If the buffer is nonempty a unit in the buffer starts being served in phase k at the same time, then for $j = 1, \ldots, n$ and $i, k = 1, \ldots, m$, the transition $e^j + e^{n+1+k} - e^{n+1} - e^{n+1+i}$ occurs at rate $Na_i q_j p_k \widetilde{z}_i \chi_{\{\widetilde{y}>0\}}$.

3 Drift Computation and Description of the Limit

In order to understand an summarize the dynamics of the stochastic process we compute the drift for our model. We compute it and we analyze the behavior of the ODE that will define the limit. For this purpose we also present a brief description of ODEs with discontinuous right hand sides.

Let us recall that for a Markov chain $V \in \mathbb{R}^d$, with transition rates $r_v(x)$ from x to $x+v$, the drift is defined by $\beta(x) = \sum_v v r_v(x)$, where the sum is in all possibles values of v. One possible representation of a Markov chain is in terms of the drift, where in a general way $V(t) = V(0) + \int_0^t \beta(V(s))ds + M(t)$, with $M(t)$ a martingale. One approach to establish a fluid limit is to exploit this decomposition for the scaled process, to prove that there is a deterministic limit for the integral term and to prove that the martingale term converges to 0.

We write down the drift β of the scaled process $U^N = \widetilde{U}^N/N$, evaluated at $u = (x, y, z)$ with $x = (x_1, \ldots, x_n)$, $z = (z_1, \ldots, z_m)$, with $\widetilde{x}_i \in \{0, 1/N, \ldots, 1\}$ for all $i = 1, \ldots n$, $\widetilde{y} \in \{0, 1/N, \ldots, 1\}$ and $\widetilde{z}_i \in \{0, 1/N\}$ for all $i = 1, \ldots m$. Let $\beta = (\beta_1, \ldots, \beta_{n+m+1})$. For $i = 1, \ldots, n$ we have the following equations:

$$\beta_i(u) = \sum_{j=1}^{n} b_{ji} x_j + q_i \sum_{j=1}^{m} a_j N z_j.$$

Let us call β_X the first n coordinates of the drift. In matrix notation:

$$\beta_X(u) = B^T x + a^T N z q.$$

For $i = n+1$ the drift equation (the $(n+1)$th coordinate of the drift) is:

$$\beta_{n+1}(u) = \sum_{j=1}^{n} b_j x_j \chi_{\{\mathbb{1}^T x < 1\}} - \sum_{i=1}^{m} a_i N z_i \chi_{\{y>0\}}$$

and in matrix notation (we also call this coordinate β_Y):

$$\beta_Y(u) = b^T x \chi_{\{\mathbb{1}^T x < 1\}} - a^T N z \chi_{\{y>0\}}.$$

For $k = n+1+i$, with $i = 1, \ldots, m$ we have:

$$\beta_k(u) = p_i \chi_{\{\mathbb{1}^T x = 1\}} \sum_{j=1}^{n} b_j x_j + \sum_{j=1}^{m} N z_j \left(a_{ji} + p_i \chi_{\{y>0\}} a_j\right).$$

In matrix notation (we call these coordinates of the drift β_Z) we have:

$$\beta_Z(u) = b^T x \chi_{\{\mathbb{1}^T x = 1\}} p + A^T N z + a^T N z \chi_{\{y>0\}} p.$$

We call the drift $\beta(u) = \beta(x, y, z)$.

$$\beta_X(x, y, z) = B^T x + a^T N z q, \tag{1}$$

$$\beta_Y(x, y, z) = b^T x \chi_{\{\mathbb{1}^T x < 1\}} - a^T N z \chi_{\{y>0\}}, \tag{2}$$

$$\beta_Z(x, y, z) = b^T x \chi_{\{\mathbb{1}^T x = 1\}} p + A^T N z + a^T N z \chi_{\{y>0\}} p. \tag{3}$$

These equations suggest the ODE that should verify the deterministic limits (x, y) of (X^N, Y^N), if they exist. However, the drift depends on the values of Nz. The process $\widetilde{Z}^N$ varies at a rate of order N whereas the processes X^N and Y^N vary at a rate of order 1. So, we can assume that when N goes to infinity and for a fixed time the process $\widetilde{Z}^N$ has reached its stationary regime and then the limit of the last m coordinates of the drift is negligible. With this argument the candidate to the ODE defining the fluid limit is obtained by replacing in equations (1) and (2) Nz by $\widetilde{z}$, the solution to the n-dimensional equation

$$b^T x \chi_{\{\mathbb{1}^T x=1\}} p + A^T \widetilde{z} + a^T \widetilde{z} \chi_{\{y>0\}} p = 0.$$

Solving the last equation (multiplying by $\mathbb{1}^T$, by $\mathbb{1}^T \left(A^T\right)^{-1}$, and using the relationship $\mathbb{1}^T \widetilde{z} = \chi_{\{\mathbb{1}^T x<1\}}$) we obtain $a^T \widetilde{z} = \mu \chi_{\{\mathbb{1}^T x<1\}}$, with $1/\mu = -\mathbb{1}^T \left(A^T\right)^{-1} p$, the mean time before absorption for the transient Markov chain defining the phase-type repair time distribution. We refer to [17] for properties of phase-type distributions. As we want to obtain an ODE for x, the candidate to ODE's vector field is $F(x) = B^T x + \mu \chi_{\{\mathbb{1}^T x<1\}} q$. We observe that the equation $\dot{x} = B^T x + \mu q$ is valid when $\mathbb{1}^T x < 1$, or, in the border $\mathbb{1}^T x = 1$, when the vector field $B^T x + \mu q$ points towards the region $\mathbb{1}^T x < 1$, that is $\mathbb{1}^T (B^T x + \mu q) < 0$. Using that $B\mathbb{1} = -b$ the condition is $b^T x > \mu$. When $\mathbb{1}^T x = 1$ and $b^T x \leq \mu$ the equation presents what is called *sliding motion*. We follow the presentation of this topic in [12]. What happens is that the deterministic system has trajectories in the border surface $\mathbb{1}^T x = 1$. The vector field that drives the equation in the border is $G(x)$, where $\mathbb{1}^T G(x) = 0$ and $G(x)$ is a linear combination of $B^T x + \mu q$ and $B^T x$ (the vectors fields corresponding to the drift in the interior and in the border). Then $G(x) = (1 - \phi(x))(B^T x + \mu q) + \phi(x) B^T x$ with $\mathbb{1}^T G(x) = 0$ that leads to $\phi(x) = 1 - b^T x/\mu$ and then, computing $G(x)$,

$$\dot{x} = \begin{cases} B^T x + \mu q, & \text{if } b^T x > \mu \text{ or } \mathbb{1}^T x < 1, \\ B^T x + b^T x q, & \text{if } b^T x \leq \mu \text{ and } \mathbb{1}^T x = 1, \end{cases} \tag{4}$$

4 Main Results

In this section we state our main results. Proofs are presented in Section 7. First we show that the scaled stochastic process $\left(X^N, Y^N\right)$ converges to the deterministic piecewise smooth dynamical system (x, y). Processes $\left(X^N, Y^N\right)$ and (x, y) are multidimensional (they live in $\mathbb{R}^{n+1}$), as the number of phases for working units is n. Convergence is in probability, uniformly in compact time intervals (*u.c.p.* convergence). From the calculus of the drift for the stochastic processes in Section 3 we have that the limit processes is driven by the vector field $B^T x + \mu q$ in the interior $\mathbb{1}^T x < 1$ and by the vector field $B^T x$ in the border $\mathbb{1}^T x = 1$. Very close to the border $\mathbb{1}^T x = 1$, when $b^T x \leq \mu$ the vector field $B^T x + \mu q$ points outside the region $\mathbb{1}^T x < 1$, but if we consider the vector field induced by transitions in the border, when $b^T x \leq \mu$ we should have a vector field $B^T x$ that points towards the region $\mathbb{1}^T x < 1$. Because of this, the

processes driven by those vector fields present a sliding motion, that means that, when $b^T x \leq \mu$, the trajectory remains in the border $\mathbb{1}^T x = 1$ driven by a linear combination of both fields. So, we must first define the piecewise smooth dynamical system (x, y), where $y = 1 - \mathbb{1}^T x$ and x is the solution (in the sense of Filippov) of the differential equation with discontinuous right hand side (4).

Lemma 1. *The differential equation (4) has a unique solution for each initial condition x_0, with $\mathbb{1}^T x_0 \leq 1$.*

Once our limit candidate is defined, we state the following theorem.

Theorem 1. *Let $\lim_{N\to\infty} X^N(0) = x_0$ in probability, with x_0 deterministic. Then for all $T > 0$*

$$\lim_{N\to\infty} \sup_{[0,T]} \left\| \left(X^N(t), Y^N(t)\right) - (x(t), y(t)) \right\| = 0,$$

in probability. The process (x, y) is defined by $y = 1 - \mathbb{1}^T x$ and x the solution to equation (4) with initial condition x_0.

The main contribution here is the study of different time scales, one at the repairman and one for life-times, and the averaging phenomena at the repairman. We treat this problem, together with the problem of discontinuous transition rates. The problem of different time scales has been addressed in other contexts (e.g. [11, 10]).

Let us study the behavior of the system defined by equation (4) by studying its fixed points. We observe that $1/\mu$ is the mean of the phase-type distribution (m, p, A). We define $1/\lambda = -\mathbb{1}^T \left(B^T\right)^{-1} q$, the expected value of a phase-type distribution (n, q, B), and

$$\rho = \frac{\mu}{\lambda}. \tag{5}$$

We identify three different behaviors, that we call (using the same definitions that in [3] for the $M/M/N/N$ queue) sub-critical when $\rho < 1$, critical when $\rho = 1$ and super-critical when $\rho > 1$.

Sub-critical case, $\rho < 1$. The mean repair time per unit $1/\mu$ is greater than the mean life-time, so we find an equilibrium with a positive number of broken units in the system. When we compute the fixed points in equation (4) the fixed point is an interior point in $\mathbb{1}^T x \leq 1$, and it is a global attractor.

Super-critical case, $\rho > 1$. When $\rho > 1$, intuitively the repairman is more effective, and we have an equilibrium with all the units (in the deterministic approximation) working. The fixed point is also a global attractor, and it is in the border $\mathbb{1}^T x = 1$.

Critical case, $\rho = 1$. In this case the fixed points for the equation in the interior and in the border coincide, giving a fixed point in the border that is a global attractor.

We state these results in the following lemma.

Lemma 2. *There are three different behaviors for equation (4):*

- $\rho < 1$ *(sub-critical). There is a unique fixed point* x^* *that is a global attractor and verifies* $\mathbb{1}^T x^* = \rho < 1$*:*

$$x^* = -\mu (B^T)^{-1} q, \tag{6}$$

- $\rho > 1$ *(super-critical). There is a unique fixed point* x^* *that is a global attractor and verifies* $\mathbb{1}^T x^* = 1$ *and* $b^T x^* < \mu$*:*

$$x^* = -\lambda (B^T)^{-1} q, \tag{7}$$

- $\rho = 1$ *(critical). There is unique fixed point* x^* *given by equations (6) or (7). It is a global attractor and verifies* $\mathbb{1}^T x^* = 1$ *and* $b^T x^* = \mu = \lambda$*.*

Theorem 2. *The system in stationary regime* $X^N(\infty)$ *converges in probability to* x^**, when* $N \to \infty$*, where* x^* *is the unique fixed point of equation (4).*

We observe that the fixed point depends only on the mean repair-time but on matrix B. That means that different life-time distribution with the same mean lead to different stationary behaviors.

5 Numerical examples

We consider the repairman problem with $N = 100$ units, for different phase type time distributions and for different values of ρ. The parameters are defined in Section 2 and ρ is defined in equation (5). As initial condition we fix the total number of working units and sample the number in each phase according to the phase type initial distribution. We show the scaled number of working units in each phase. We illustrate the convergence to the ODE's fixed point x^* and the sliding motion. The parameters for each example (figure) are given in Table 1.

Exponencial life-time and hypoexponential repair time. In Figure 1 we consider exponential life-time and hypoexponential (sum of independent exponentials). We represent for each parameter set the evolution with time of the stochastic process X^N and we show the convergence to the fixed point x^*.

Hypoexponential life and repair time, sliding motion. In Figure 2 we consider two phases both in the life and repair times (both distributions are hypoexponential). At the left we represent, for each parameter set, the evolution with time of the stochastic processes X_1^N and X_2^N and the ODE's solution (x_1, x_2). We also show the convergence to the fixed point $x^* = (x_1^*, x_2^*)$. At the right we represent, for each parameter set, the trajectory of process X^N, the trajectory of the ODE's solution x and the fixed point. Depending on the initial condition, we find sliding motion (as shown in the bottom figures), but, as we have the same parameters, both trajectories converge to the fixed point.

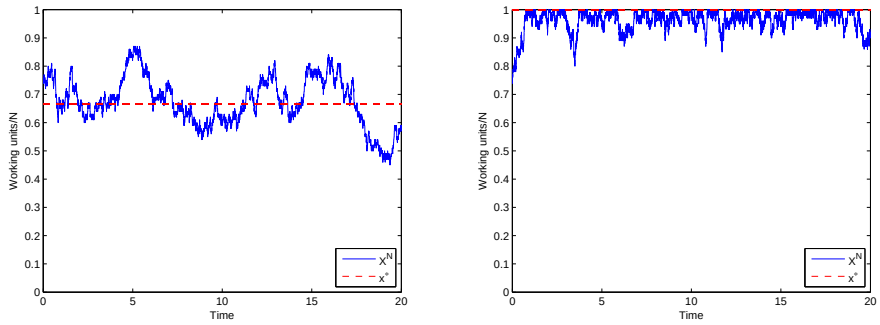


Fig. 1. Exponential life-time and hypoexponential repair time. In the right we show the sliding motion.

Table 1. Parameters for Figures 1, 2 and 3.

	Fig.1(left)	Fig.1(right)	Fig.2(top)	Fig.2(bottom)	Fig.3(top)	Fig.3(bottom)
A	$\begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}$	$\begin{pmatrix} -2 & 2 \\ 0 & -3 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}$	-0.5	-1.5
B	-1	-1	$\begin{pmatrix} -2 & 2 \\ 0 & -3 \end{pmatrix}$	$\begin{pmatrix} -2 & 2 \\ 0 & -3 \end{pmatrix}$	$\begin{pmatrix} -3 & 3 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & -5 \end{pmatrix}$	$\begin{pmatrix} -3 & 3 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & -5 \end{pmatrix}$
p	$(1,0)$	$(1,0)$	$(1,0)$	$(1,0)$	1	1
q	1	1	$(1,0)$	$(1,0)$	$(1,0,0)$	$(1,0,0)$
$X^N(0)$	$0.75N$	$0.75N$	$0.75N$	N	$0.75N$	$0.75N$
ρ	0.6667	1.2	0.5556	0.5556	0.5167	1.55

Hypoexponential life-time with three phases and exponential repair time. In Figure 3 we consider three phases (hypoexponential) in the life-time and exponential repair time. At the left we represent the evolution with time of the stochastic processes X_1^N, X_2^N, X_3^N and we show the convergence to the fixed point $x^* = (x_1^*, x_2^*, x_3^*)$. At the right we represent the trajectory of process X^N and the fixed point. The example at the top has $\rho < 1$. The example at the bottom has $\rho > 1$, so we find sliding motion in the plane $\mathbb{1}^T x = 1$ and the fixed point is also in the same plane.

6 Conclusions

In this paper we find a deterministic fluid limit for a Machine Repairman Model with phase type distributions. The fluid limit is a deterministic process with

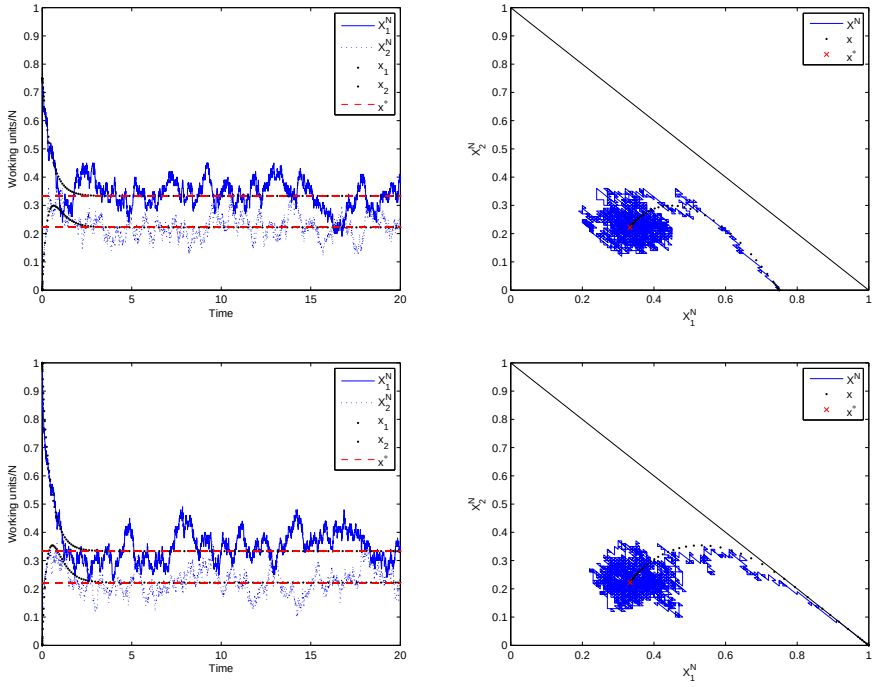


Fig. 2. Hypoexponential life and repair time (parameters in Table 1). The bottom figure shows the sliding motion.

piecewise smooth trajectories, that presents three different behaviors depending only on the expected values of the phase type distributions involved. The stationary behavior of the scaled system converges to a fixed point that only depends on the mean time between failures and on the mean repair time at the repairman. There are characteristics of the system that hinder us from using classical results from fluid limits or density dependent population processes, as the presence of two different time scales (one for the working units and one for the repairman) and discontinuities in the transition rates due to idle time at the repairman. Concerning the different time scales we find an averaging result, where the phase type distribution at the repairman is represented in the limit only by its mean value. The behavior of the system due to idle times is similar to the behavior of the $M/M/N/N$ queue if we consider exponential distributions instead of phase type ones. The phase type distribution at the failure time adds more dimensions to the problem. This leads to different behaviors for the deterministic system than for the $M/M/N/N$ queue.

Another situation to be addressed in future work is to consider a general distribution at the repairman, as with this scaling it seems to be some insensitivity property where in the limit the behavior depends on the repair time only by its mean. However, the variability at life-time leads to different fluid limits so

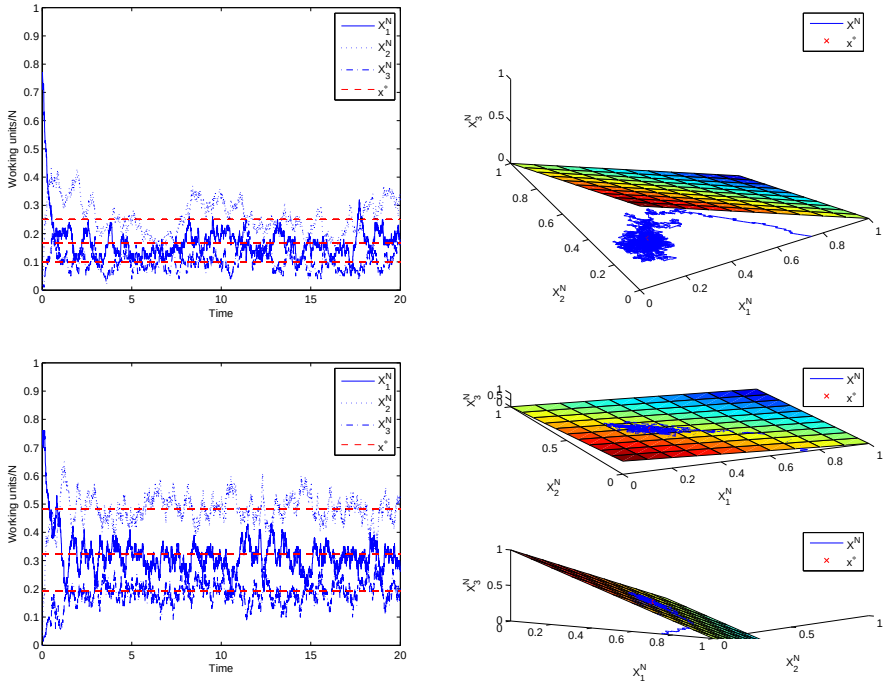


Fig. 3. Hypoexponential life-time and exponential repair times (parameters in Table 1). The three bottom figures shows the sliding motion and the fixed point in the plane $\mathbb{1}^T x = 1$ at the right.

another possible topic of study is the impact of the life-time distribution on performance measures. Last, we also intend to analyze the asymptotic distribution of the difference between the scaled system and its deterministic limit.

7 Proofs

In what follows we give some proofs of results stated in section 4. First we address the existence of solutions to equation (4) and we prove Lemma 1. For a differential equation $\dot{x} = F(x)$, with F discontinuous, solutions are defined in the set of absolutely continuous functions, instead of differentiable functions as in the classical case. The ODE is defined in the region $\{\mathbb{1}^T x \leq 1\}$ by $B^T x + \mu q$ in the region $\{\mathbb{1}^T x < 1\}$ and $B^T x$ in the region $\{\mathbb{1}^T x = 1\}$. In order to consider the framework of differential equations with discontinuous right hand sides, we extend the definition of the ODE. We define the equation by two continuous fields, $F_1(x) = B^T x + \mu q$ in the region $R_1 = \{\mathbb{1}^T x < 1\}$, $F_2(x) = B^T x$ in the region $R_2 = \{\mathbb{1}^T x > 1\}$ with a region $H = \{x \in \mathbb{R}^n : \ \mathbb{1}^T x = 1\}$ where the field is discontinuous. The field $B^T x$ always point towards R_1. The field $B^T x + \mu q$ points toward R_1 when $b^T x > \mu$ and point towards R_2 when $b^T x < \mu$. We find

that the equation presents *transversal crossing* in H for $b^T x > \mu$ and a *stable sliding motion* in H for $b^T x < \mu$ (defined as in [12]). Transversal crossing occurs when both vector fields point towards R_1 and trajectories from R_2 crosses H. If trajectories start in R_1 or in H, with $b^T x > \mu$ they go into R_1. Stable sliding motion occurs as F_1 points towards R_2 and F_2 points towards F_1 (in H, for $b^T x < \mu$). The ODE has trajectories in the border surface H. The vector field that drives the equation is $G(x)$, with $G = F_1$ in the interior and in the border, when $b^T x < \mu$, $G(x)$ verifies $\mathbb{1}^T G(x) = 0$ and it is a linear combination of $F_1(x) = B^T x + \mu q$ and $F_2(x) = B^T x$, leading to equation (4). When $b^T x + \mu$, $B^T x + \mu q$ is tangential to H, whereas $B^T x$ points toward R_1, so the trajectories go into R_1. It is called *first order exit condition* of sliding motion.

Proof (Lemma 1). From [12], in order to prove the existence of solutions we need to verify that the field F defined as $F_1(x) = B^T x + \mu q$ in R_1 and $F_2(x) = B^T x$ in R_2 is continuous in each closure $\bar{R}_1$ and $\bar{R}_2$. In addition, if we consider the normal vector to H, $\mathbb{1}$, we can verify that (except in the region $\{b^T x = \mu\}$) we have $\mathbb{1}^T F_1(x) > 0$ and $\mathbb{1}^T F_2(x) < 0$. These conditions mean that there is a stable sliding motion, where the solution belongs to H, driven by G, the linear combination of F_1 and F_2. In addition, unique solutions are also defined for initial conditions in H. □

Now we address the proof of Theorem 1. We recall the definitions of the drift and the vector fields. Let $U^N = \left(X^N, Y^N, Z^N\right)$ and $u = (x, y, 0)$. We have that $(x(t), y(t)) = (x_0, y_0) + \int_0^t G(x(s), y(s))\mathrm{d}s$.

Proof (Theorem 1). First, we observe that $\mathbb{1}^T X^N(t) + Y^N(t) + \mathbb{1}^T Z^N(t) = 1$, and that $\lim_{N\to+\infty} \sup_{[0,T]} Z^N(t) = 0$. Then, in order to prove that

$$\lim_{N\to+\infty} \sup_{[0,T]} \left\| \left(X^N(t), Y^N(t)\right) - (x(t), y(t)) \right\| = 0$$

in probability, we only need to prove that $\lim_{N\to+\infty} \sup_{[0,T]} \left\| X^N(t) - x(t) \right\| = 0$ in probability. In the proof of this theorem we follow the approach of [18].

$$\sup_{[0,T]} \left\| X^N(t) - x(t) \right\| \leq \left\| X^N(0) - x(0) \right\| \tag{8}$$

$$+ \sup_{[0,T]} \left\| X^N(t) - X^N(0) - \int_0^t \beta_X \left(U^N(s)\right) \mathrm{d}s \right\| \tag{9}$$

$$+ \sup_{[0,T]} \left\| \int_0^t \beta_X \left(U^N(s)\right) \mathrm{d}s - \int_0^t F\left(X^N(s)\right) \mathrm{d}s \right\| \tag{10}$$

$$+ \sup_{[0,T]} \left\| \int_0^t F\left(X^N(s)\right) \mathrm{d}s - \int_0^t G\left(X^N(s)\right) \mathrm{d}s \right\| \tag{11}$$

$$+ \sup_{[0,T]} \left\| \int_0^t G\left(X^N(s)\right) \mathrm{d}s - \int_0^t G\left(x(s)\right) \mathrm{d}s \right\| \tag{12}$$

We want to prove that (9), (10) and (11) converge to 0 in probability. So, provided that the initial condition $X^N(0)$ converges to $x(0)$, we have that, with probability that tends to 1 with N,

$$\sup_{[0,T]} \left\| X^N(t) - x(t) \right\| \leq \varepsilon + \sup_{[0,T]} \left\| \int_0^t G\left(X^N(s)\right) \mathrm{d}s - \int_0^t G\left(x(s)\right) \mathrm{d}s \right\|.$$

Using that G is piecewise linear and Gronwall inequality, we obtain the bound $\sup_{[0,T]} \left\| X^N(t) - x(t) \right\| \leq \varepsilon e^{KT}$, which leads to $\sup_{[0,T]} \left\| X^N(t) - x(t) \right\| \to 0$ in probability. We study the convergence of (9), (10) and (11). To show the convergence of (9), we first notice that

$$(9) \leq \left\| U^N(t) - U^N(0) - \int_0^t \beta\left(U^N(s)\right) \mathrm{d}s \right\|.$$

Convergence follows from the representation of the process as the initial condition plus the integral of the drift plus a martingale term. The martingale term goes to 0 with N because of the scaling. Let us define for a Markov chain $V \in \mathbb{R}^d$, with transition rates $r_v(x)$ from x to $x+v$, $\alpha(x) = \sum_v \|v\|^2 r_v(x)$. Let us also call α the corresponding object for U^N. Convergence of (9) can be then proved using Proposition 8.7 in [8], that states that

$$E\left(\sup_{[0,T]} \left\| U^N(t) - U^N(0) - \int_0^t \beta\left(U^N(s)\right) \mathrm{d}s \right\|^2 \right) \leq 4 \int_0^T \alpha\left(U^N(s)\right) \mathrm{d}s.$$

As for our scaling $\sup_{[0,T]} \alpha\left(U^N(t)\right) \sim \mathcal{O}(1/N)$, convergence holds.

To prove that (10) converges to 0 in probability we consider the last m coordinates of (9), corresponding to the phases at the repairman. As we have proved that (9) converges to 0 in probability, we conclude that

$$\int_0^t \beta_Z\left(U^N(s)\right) \mathrm{d}s = \int_0^t \left(b^T X^N(s) \chi_{\{\mathbb{1}^T X^N(s)=1\}} p + A^T \widetilde{Z}^N(s) + a^T \widetilde{Z}^N(s) \chi_{\{Y^N(s)>0\}} p \right) \mathrm{d}s$$

converges to 0 in probability. Multiplying by $\mu \mathbb{1}^T (A^T)^{-1}$,

$$\int_0^t \left(-b^T X^N(s) \chi_{\{\mathbb{1}^T X^N(s)=1\}} + \mu \chi_{\{\mathbb{1}^T X^N(s)<1\}} - a^T \widetilde{Z}^N(s) \chi_{\{Y^N(s)>0\}} \right) \mathrm{d}s \quad (13)$$

goes to 0 in probability. In addition, $\mathbb{1}^T \beta_X + \beta_Y + \mathbb{1}^T \beta_Z = 0$. Then, as $\int_0^t \beta_Z\left(U^N(s)\right) \mathrm{d}s$ converges to 0 in probability, $\int_0^t \left(\mathbb{1}^T \beta_X\left(U^N(s)\right) + \beta_Y\left(U^N(s)\right) \right) \mathrm{d}s$ also converges to 0 in probability and it is equal to

$$\int_0^t \left(-b^t X^N(s) \chi_{\{\mathbb{1}^T X^N(s)=1\}} + a^T \widetilde{Z}^N(s) \chi_{\{Y^N(s)=0\}} \right) \mathrm{d}s. \quad (14)$$

Then, considering the sum of equations (13) and (14), we obtain

$$\lim_{N\to+\infty} (10) = \lim_{N\to+\infty} \sup_{[0,T]} \int_0^t \left(a^T \widetilde{Z}^N(s) q - \mu \chi_{\{\mathbb{1}^T X^N(s)<1\}} q \right) \mathrm{d}s = 0.$$

The convergence of (11) can be proved by approximating the continuous process in the border by a discrete process with the same jumps. As our model verifies the hypotheses of [12], the same proof that in Lemma 3 of [18] holds. □

Proof (Lemma 2). First, we compute the fixed points of both fields: $B^T x + \mu q$ in $\mathbb{R}^n$, and $(B^T + qb^T)x$ in $\{\mathbb{1}^T x = 1\}$. We will exploit the linearity of the field in each region where it is continuous. Then we discuss in terms of ρ.

The fixed point for $B^T x + \mu q$ is $x_1^* = -\mu(B^T)^{-1}q$. We recall that B^T is regular due to the properties of phase type distributions. In addition, the eigenvalues of B^T are all negative (since B^T has the same eigenvalues than B). Matrix B has a negative diagonal, and the sum of all non diagonal entries per row (that are all positive) is less than or equal to the absolute value of the diagonal element. This is because the sum of each row of $\widehat{B}$ is 0 and the last column (that does not belong to B) has non-negative entries. Then, considering the field in $\mathbb{R}^n$, we have that $x_1^* = -\mu(B^T)^{-1}q$ is a global attractor. In addition, as $\mathbb{1}^T x_1^* = \rho$, we have that x_1^* is an interior point of R_1 iff $\rho < 1$, x_1^* is an interior point of R_2 iff $\rho > 1$, and $x_1^* \in H \cap \{b^T x = \mu\}$ iff $\rho = 1$. We observe that when $\rho \geq 1$, as x_1^* is the unique fixed point of $B^T x + \mu q$, and the vector field points outside R_1, the solution of $\dot{x} = B^T x + \mu q$ is pushed towards H.

Now we consider the fixed point of $(B^T + qb^T)x$ in $\{\mathbb{1}^T x = 1\}$. As $b = -B\mathbb{1}$, we have that $(B^T + qb^T)x = 0$ iff $(I - q\mathbb{1}^T)B^T x = 0$, where I is the identity matrix. As B^T is invertible, if v is an eigenvector of $q\mathbb{1}^T$ with eigenvalue 1, $x_1^* = (B^T)^{-1}v$ is a fixed point. Matrix $q\mathbb{1}^T$ has eigenvalues 0 and 1 and, as $q\mathbb{1}^T$ has range 1, the dimensions of the corresponding eigenspaces are respectively $n-1$ and 1. Then there is a one-dimensional space with eigenvalue 1, that intersects $\{\mathbb{1}^T x = 1\}$, giving the fixed point $x_2^* = -\lambda(B^T)^{-1}q$. Since $\mathbb{1}^T x_2^* = 1$, we have that $x_2^* \in H$. In addition $x_2^* \in H \cap \{b^T x < \mu\}$ iff $\rho > 1$, $x_2^* \in H \cap \{b^T x > \mu\}$ iff $\rho < 1$ and $x_2^* \in H \cap \{b^T x = \mu\}$ iff $\rho = 1$. We also have that, restricted to H, x_2^* is a global attractor, so in the case of $\rho \leq 1$, the solution in H is pushed to the region $\{b^t x > \mu\}$, where the solution is again driven by the field $B^T x + \mu q$, so the solution that starts in H does not remain in H for $\rho < 1$. □

Proof (Theorem 2). To prove the convergence in stationary regime, we use the results in [14]. In Theorem 5 in [14] it is proved convergence in stationary regime to the fixed point for piecewise smooth dynamical systems. This is an extension of a general result in [15]. The hypotheses needed are that the fixed point is a unique global attractor and regularity assumptions for the trajectories. We use Lemma 2 to characterize the fixed point. are verified in our case, where solutions are piecewise linear and the discontinuity surface H is a hyperplane. Then applying Theorem 5 in [14] we obtain Theorem 2. □

References

1. Lavenberg, S.: Computer performance modeling handbook. Notes and reports in computer science and applied mathematics. Academic Press (1983)

2. Haque, L., Armstrong, M.J.: A survey of the machine interference problem. European Journal of Operational Research **179**(2) (2007) 469–482
3. Robert, P.: Stochastic Networks and Queues. Stochastic Modelling and Applied Probability Series. Springer-Verlag, New York (2003)
4. Kurtz, T.G.: Representation and approximation of counting processes. In: Advances in filtering and optimal stochastic control (Cocoyoc, 1982). Volume 42 of Lecture Notes in Control and Inform. Sci. Springer, Berlin (1982) 177–191
5. Iglehart, D.L., Lemoine, A.J.: Approximations for the repairman problem with two repair facilities, I: No spares. Advances in Applied Probability **5**(3) (1973) pp. 595–613
6. Iglehart, D.L., Lemoine, A.J.: Approximations for the repairman problem with two repair facilities, II: Spares. Advances in Appl. Probability **6** (1974) 147–158
7. Ethier, S.N., Kurtz, T.G.: Markov processes: Characterization and convergence. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., New York (1986)
8. Darling, R.W.R., Norris, J.R.: Differential equation approximations for Markov chains. Probab. Surv. **5** (2008) 37–79
9. Shwartz, A., Weiss, A.: Large deviations for performance analysis. Stochastic Modeling Series. Chapman & Hall, London (1995)
10. Ayesta, U., Erausquin, M., Jonckheere, M., Verloop, I.M.: Scheduling in a random environment: Stability and asymptotic optimality. IEEE/ACM Trans. Netw. **21**(1) (2013) 258–271
11. Ball, K., Kurtz, T.G., Popovic, L., Rempala, G.: Asymptotic analysis of multiscale approximations to reaction networks. The Annals of Applied Probability **16**(4) (2005) 1925–1961
12. Bortolussi, L.: Hybrid limits of continuous time markov chains. In: Proceedings of Eighth International Conference on Quantitative Evaluation of Systems (QEST), 2011. (sept. 2011) 3 –12
13. Bortolussi, L., Tribastone, M.: Fluid limits of queueing networks with batches. In: Proceedings of the third joint WOSP/SIPEW international conference on Performance Engineering. ICPE '12, New York, NY, USA, ACM (2012) 45–56
14. Houdt, B.V., Bortolussi, L.: Fluid limit of an asynchronous optical packet switch with shared per link full range wavelength conversion. In: SIGMETRICS'12. (2012) 113–124
15. Benaïm, M., Le Boudec, J.Y.: A class of mean field interaction models for computer and communication systems. Perform. Eval. **65**(11-12) (2008) 823–838
16. Le Boudec, J.Y.: The Stationary Behaviour of Fluid Limits of Reversible Processes is Concentrated on Stationary Points. Technical report (2010)
17. Asmussen, S., Albrecher, H.: Ruin probabilities. Second edn. Advanced Series on Statistical Science & Applied Probability, 14. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2010)
18. Bortolussi, L.: Supplementary material of Hybrid Limits of Continuous Time Markov Chains, http://www.dmi.units.it/∼bortolu/files/qest2011supp.pdf (2011) [Online; accesed 15-March-2013].