

Fluid models in performance analysis ^{*}

Marco Gribaudo¹, Miklós Telek²

¹ Dip. di Informatica, Università di Torino, marcog@di.unito.it

² Dept. of Telecom., Technical University of Budapest telek@hitbme.hu

Abstract. Stochastic fluid models have been applied to model and evaluate the performance of many important real systems. The automatic analysis tools to support of fluid models are still not as improved as the ones for discrete state Markov models, but there is a wide range of models which can be effectively described and analyzed with fluid models. Also the model support of hybrid models from various performance evaluation tools improves continuously.

The aim of this work is to summarize the basic concepts and the potential use of Markov fluid models. The factors which determine the limits of solvability of fluid models are also discussed. Practical guidelines can be extracted from these factors to determine the applicability of fluid models in practical modeling examples. The work is supported by an example where Fluid Models, derived from an higher level modeling language (Fluid Stochastic Petri Nets), have been exploited to study the transfer time distribution in Peer-to-Peer file sharing applications.

1 Motivations

1.1 Problems with discrete state models

Even if the conventional performance evaluation techniques are well suited to describe a wide range of real systems, things are not always as easy as they seem. The modeler usually faces several problems when trying to describe a system with a particular formalism, and sometimes these problems make the models extremely hard to handle. Some examples are:

- **State space explosion.** One of the weakest points in performance evaluation is that the complexity of the solution of discrete state models generally grows exponentially with the complexity of the model behavior. Many analysis techniques for most of the formalisms require the generation and the visit of all the possible states that the system may reach. This set of states is called the *state space of the model*, and for many applications it must be stored in the central memory. Since it grows exponentially with the complexity of the model, the size of this set may reach very quickly the storage capacity of the machine. In many situations this problem prevents a well defined model from being solved.

^{*} This work is partially supported by the Italian-Hungarian R&D project 9/2003 and by the OTKA K61709 grant.

- **Inaccurate results.** A model is always a simplification of reality. Some simplifications are motivated by the fact that they cut out some parameters of the model that do not influence the required solutions. Some others are required in order to produce a system that can be analyzed with the tools that a modeler has. These simplifications may not be adequate sometimes and can lead to incorrect or inaccurate results. Many of these simplifications involve the characterization of some stochastic process by a Poisson process and the probability distribution of the time between two events by an exponential distribution.
- **Granularity and sizes.** In many situations the user must deal with a huge number of small elements. Let us consider for example a production line that produces bolts and screws. Thousands of parts will be produced in a very short time. A model that wishes to capture the number of parts produced, must deal with this big number which usually makes the state space explode even faster. Similar problems arise in today's communication systems which deals with a high number of very small data packets.
- **Modeling power limitations.** Sometimes a model depends on some physical quantity such as temperature or power consumption. Those are continuous quantities and they cannot be emulated correctly by discrete states. In these cases, the modeling power of a discrete state model specification language may not be adequate to describe the system.

1.2 Possible solutions

In order to overcome the mentioned limitations, new modeling techniques have been developed. In this paper, we will examine how the previous problems can be attacked using *Hybrid continuous / discrete techniques*. Continuous and hybrid models can in some circumstances solve the above mentioned problems, or give better results than conventional discrete state techniques in terms of computational complexity or accuracy of results. For example, hybrid models may solve the problems in the following way:

- *State space explosion:* Usually hybrid models are analyzed by splitting the discrete state space into a *discrete part*, that takes into account the possible states that the system may reach (by considering only the discrete components of the system) and a *continuous part*. Usually, when solving a hybrid model, only the discrete part of the state space must be memorized explicitly, while the continuous part is expressed as a set of functions or predicates. This greatly reduces the number of states that must be memorized and in some cases may solve the state space explosion problem.
- *Inaccurate results:* When the inaccuracy of a result is caused by the Poisson arrival or the exponential service time distribution, continuous models can be used to overcome this problem. Continuous components can be used to explicitly model the time since an action has been enabled, and can thus be used to model non-Markovian processes and complex memory properties.

- *Granularity and sizes*: When the modeling problems are caused by a variable that has a very small granularity, this variable may be approximated by a continuous quantity. Even if in the real system the variable is actually discrete, usually its continuous approximation can lead to very good results, especially if the changes in the real quantity happens at a constant rate. For example modeling the number of packets contained in a queue of an ATM router or the number of bytes allocated in the central memory of a PC can produce good results .
- *Modeling power limitations*: If the system under study depends on a physical continuous quantity that must be modeled explicitly to capture its real behavior, then hybrid models seem to be the natural solution. For example the instantaneous fuel consumption of a turbine in a power plant can be modeled explicitly by a continuous variable.

2 Formalisms

Continuous quantities have been introduced in performance models in many flavors. Many high-level and low-level performance evaluation formalisms have been developed to deal with continuous quantities. In this work we will consider:

- Reward Models (RM),
- Fluid Models (FM), and
- Fluid Stochastic Petri Nets (FSPN).

2.1 Reward Models

A **Reward Model** is a Markov chain in which each state has associated a positive quantity called *Reward Rate*. A continuous variable, takes into account all the Reward accumulated over a time interval. This quantity grows proportionally with the time spent in a state and with the corresponding reward rate. One of the key aspect of reward models, is that the accumulated reward is unbounded.

The Markov Chain that governs the reward accumulation is called the *underlying Markov Chain*, and is described by a generator matrix \mathbf{Q} , whose element q_{ij} defines the transition from state i to state j , as in any other Markov Chain:

$$q_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P\{S(t + \Delta t) = j | S(t) = i\}}{\Delta t}, \text{ for } i \neq j$$

$$q_{ij} = - \sum_{k \neq i} q_{ik}, \text{ for } i = j,$$

where $S(t)$ is the state of the underlying Markov chain at time t .

The *reward rate* of the state i is denoted by r_i , $r_i \geq 0$. This quantity describes the rate at which the accumulated reward grows when the underlying Markov chain is in state i . Reward rates are collected in a diagonal matrix, \mathbf{R} , whose elements R_{ij} are such that:

$$R_{ij} = 0, \quad \text{for } i \neq j,$$

$$R_{ij} = r_i, \quad \text{for } i = j.$$

We denote with $X(t)$ the total reward accumulated until time t , and we set $X(0) = 0$. If we know the evolution of the underlying Markov chain (that is $S(t)$), then we can compute $X(t)$ as follows:

$$X(t) = \int_0^t r_{S(u)} du.$$

The fact that the reward rates are always positive, and that the accumulated reward is unbounded, greatly simplifies the analytical description of the systems. Many efficient techniques exist in the literature to analyze Reward Models [26, 10, 21, 23, 25].

2.2 Fluid Models

Fluid Models are an extension of Reward Models. Various definitions of fluid models exist, and they will be fully addressed in section 3. Here we will put just a general presentation of the main formalism. As RMs, FMs are characterized by an underlying Markov Chain, defined by matrix \mathbf{Q} , and a reward matrix \mathbf{R} . The main difference with respect to RM, is that in FM the rate associated to each state (called in this case *flow rate or drift*) can be positive, negative or zero. Usually, the accumulated reward is called *Fluid Level*, since the continuous value of the reward can be used to represent the level of fluid contained in a reservoir. The second main difference between FMs and RMs, is that in a FM the fluid level has at least a lower bound at zero, and may also have an upper bound at a fixed positive value.

Even if the differences between FMs and RMs seem to be negligible at first sight, FMs are much more complex to be analyzed. The presence of boundaries and negative rates, imposes the introduction of boundary conditions in the equations that describe the evolution of the system. These conditions reduce the applicability of analytical results, and makes the solution much more complex. However, from a modeling point of view, the introduction of negative rates and bounds allows the characterization of a larger set of interesting systems, which could not be analyzed by simple RMs. FMs can be used to approximate large buffers with continuous quantities, making thus the formalism well suited to analyze high speed communication and production systems.

2.3 Fluid Stochastic Petri Nets

A **Fluid Stochastic Petri Net (FSPN)** is an extension of an ordinary Stochastic Petri Net, capable of incorporating continuous quantities. Other similar extensions with minor differences are: *Continuous Petri Nets* and *Hybrid Petri Nets* [5]. In this work we will not consider such formalisms, and we will present only the basic formalism, intended for stochastic analysis. Several extensions have been considered to allow the description of more complex models, which however can only be solved using simulation [9].

Formally, a FSPN is a tuple $\langle \mathcal{P}, \mathcal{T}, \mathcal{A}, B, F, W, R, M_0 \rangle$, where:

- \mathcal{P} is the set of places, partitioned into a set of discrete places $\mathcal{P}_d = \{p_1, \dots, p_{|\mathcal{P}_d|}\}$ and a set of continuous places $\mathcal{P}_c = \{c_1, \dots, c_{|\mathcal{P}_c|}\}$ (with $\mathcal{P}_d \cap \mathcal{P}_c = \emptyset$ and $\mathcal{P}_d \cup \mathcal{P}_c = \mathcal{P}$). The discrete places may contain tokens (the number of tokens in a discrete place is a natural number), while the marking of a continuous place is a non negative real number that we call the fluid level. In the graphical representation, a discrete place is drawn as a single circle while a continuous place is signified by two concentric circles. The complete state (marking) of a FSPN is described by a pair of vectors $M = (\mathbf{m}, \mathbf{x})$, where the vector \mathbf{m} , of dimension $|\mathcal{P}_d|$ is the marking of the discrete part of the FSPN and the vector \mathbf{x} , of dimension $|\mathcal{P}_c|$, represents the fluid levels in the continuous places (with $x_l \geq 0$ for any $c_l \in \mathcal{P}_c$). We use \mathcal{S} to denote the partly discrete and partly continuous state space. In the following we denote by \mathcal{S}_d and \mathcal{S}_c the discrete and the continuous component of the state space respectively. The marking $M = (\mathbf{m}, \mathbf{x})$ evolves in time. We can imagine the marking M at time τ as the stochastic marking process $\mathcal{M}(\tau) = \{(\mathbf{m}(\tau), \mathbf{x}(\tau)), \tau \geq 0\}$.
- \mathcal{T} is the set of transitions partitioned into a set of stochastically timed transitions \mathcal{T}_e and a set of immediate transitions \mathcal{T}_i (with $\mathcal{T}_e \cap \mathcal{T}_i = \emptyset$ and $\mathcal{T}_e \cup \mathcal{T}_i = \mathcal{T}$). A timed transition $T_j \in \mathcal{T}_e$ is drawn as a rectangle and has an instantaneous firing rate associated with it. An immediate transition $t_h \in \mathcal{T}_i$ is signified by a thin bar and has constant zero firing time.
- \mathcal{A} is the set of arcs partitioned into three subsets: \mathcal{A}_d , \mathcal{A}_h and \mathcal{A}_c . The subset \mathcal{A}_d contains the discrete arcs which can be seen as a function³ $A_d : ((\mathcal{P}_d \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P}_d)) \rightarrow \mathbb{N}$. The arcs \mathcal{A}_d are drawn as single arrows. The subset \mathcal{A}_h contains the inhibitor arcs, $A_h : (\mathcal{P}_d \times \mathcal{T}) \rightarrow \mathbb{N}$. These arcs are drawn with a small circle at the end. The definitions of $\bullet t_j$, t_j^\bullet , and ${}^\circ t_j$ involve only discrete places and are the same as for the standard GSPNs. The subset \mathcal{A}_c define the continuous arcs. These arcs are drawn as double arrows to suggest a pipe. \mathcal{A}_c is a subset of $(\mathcal{P}_c \times \mathcal{T}_e) \cup (\mathcal{T}_e \times \mathcal{P}_c)$, i.e., a continuous arc can connect a fluid place to a timed transition or it can connect a timed transition to a fluid place.
- The function $B : \mathcal{P}_c \rightarrow \mathbb{R}^+ \cup \{\infty\}$ describes the fluid upper boundaries on each continuous place. This boundary has no effect when it is set to infinity. From this it follows that $\forall M = (\mathbf{m}, \mathbf{x}) \in \mathcal{S}$ and $c_l \in \mathcal{P}_c$, $0 \leq x_l \leq B(c_l)$. Each fluid place has an implicit lower boundary at level 0.
- The firing rate function F is defined for timed transitions \mathcal{T}_e so that $F : \mathcal{T}_e \times \mathcal{S} \rightarrow \mathbb{R}^+$. Therefore, a timed transition T_j enabled at time τ in a discrete marking $\mathbf{m}(\tau)$ with fluid level $\mathbf{x}(\tau)$, may fire with rate $F(T_j, \mathbf{m}(\tau), \mathbf{x}(\tau))$, that is:
$$\lim_{\Delta\tau \rightarrow 0} Pr\{T_j \text{ fires in } (\tau, \tau + \Delta\tau) | \mathcal{M}(\tau) = (\mathbf{m}(\tau), \mathbf{x}(\tau))\} = F(T_j, \mathbf{m}, \mathbf{x}) \Delta\tau$$
We also use as a short hand notation $F(T_j, M)$, where $M = (\mathbf{m}, \mathbf{x})$.
- The weight function W is defined for immediate transitions \mathcal{T}_i such that $W : \mathcal{T}_i \times \mathcal{S}_d \rightarrow \mathbb{R}^+$. Note that the firing rates for timed transitions may be dependent both on the discrete and the continuous part of the marking,

³ Note that when the arcs are defined as a function we use uppercase symbols.

while the weights for immediate transitions may only be dependent on the discrete part.

- The function $R : \mathcal{A}_c \times \mathcal{S} \rightarrow \mathbb{R}^+ \cup \{0\}$ is called the *flow rate function* and describes the marking dependent flow of fluid across the input and output continuous arcs connecting timed transitions and continuous places.
- The initial state of the FSPN is denoted by the pair $M_0 = (\mathbf{m}_0, \mathbf{x}_0)$.

Figure 1 visually represents the discrete primitives of a FSPN (that are identical to their GSPN counterparts), and Figure 2 shows the continuous primitives of the formalism.

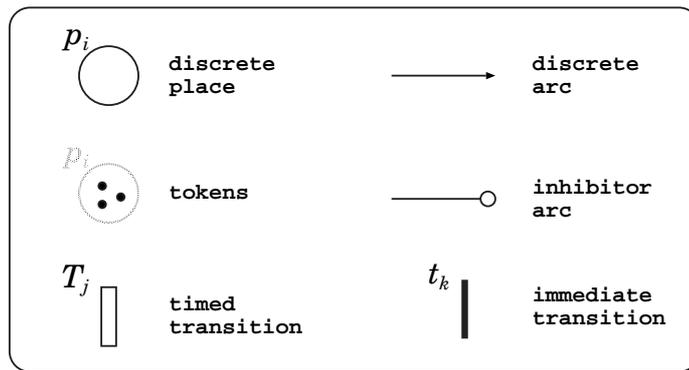


Fig. 1. Discrete primitives

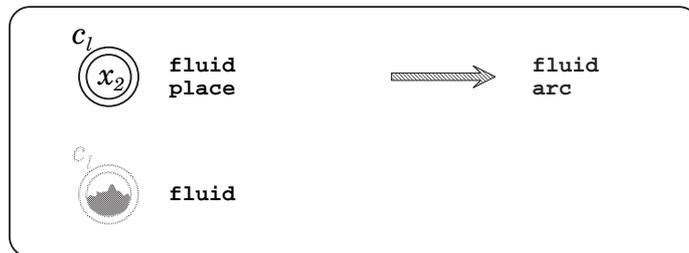


Fig. 2. Continuous primitives

The role of the previous sets and functions will be clarified by providing the enabling and firing rules.

Let us denote by m_i the i -th component of the vector \mathbf{m} , i.e., the number of tokens in place p_i when the discrete marking is \mathbf{m} . We say that a transition $t_j \in \mathcal{T}$ (no matter whether t_j is an immediate or timed transition) has concession

in marking $M = (\mathbf{m}, \mathbf{x})$ iff $\forall p_i \in \bullet t_j, m_i \geq A_d(p_i, t_j)$ and $\forall p_i \in \circ t_j, m_i < A_h(p_i, t_j)$. If an immediate transition has concession in $M = (\mathbf{m}, \mathbf{x})$, it is said to be enabled and the marking is vanishing. Otherwise, the marking is tangible and any timed transition with concession is enabled in it. Note that the previous definition is exactly the one of standard GSPNs [4], i.e., the concession and the enabling conditions depend only on the discrete part of the FSPN. Let $\mathcal{E}(M)$ denote the set of enabled transitions in marking $M = (\mathbf{m}, \mathbf{x})$, we have that $\mathcal{E}(M) = \mathcal{E}(M')$, for any marking $M' = (\mathbf{m}, \mathbf{x}')$.

The stochastic evolution of the discrete part of the FSPN in a tangible marking is governed by a race [3]. In a vanishing marking, instead, the weights are used to decide which transition should fire according to the standard rules for GSPNs [4]. Let us see how enabled transitions may influence the continuous part of the marking. Each continuous arc that connects a fluid place $c_l \in \mathcal{P}_c$ to an enabled timed transition $T_j \in \mathcal{T}_e$ (resp. an enabled transition T_j to a fluid place c_l), causes a “change” in the fluid level of place c_l . Let $\mathcal{M}(\tau)$ be the marking process, i.e., $\mathcal{M}(\tau) = M_i$ if at time τ the marking of the FSPN is $M_i = (\mathbf{m}_i, \mathbf{x}_i)$. Thus, when the FSPN marking is $\mathcal{M}(\tau)$ fluid can leave place $c_l \in \mathcal{P}_c$ along the arc $(c_l, T_j) \in \mathcal{A}_c$ at rate $R((c_l, T_j), \mathcal{M}(\tau))$ and can enter the continuous place c_l at rate $R((T_j, c_l), \mathcal{M}(\tau))$ for each (timed) transition T_j enabled in $\mathcal{M}(\tau)$. The potential rate of change of fluid level for the continuous place c_l in marking $\mathcal{M}(\tau)$ is:

$$r_l^p(\mathcal{M}(\tau)) = \sum_{T_j \in \mathcal{E}(\mathcal{M}(\tau))} R((T_j, c_l), \mathcal{M}(\tau)) - R((c_l, T_j), \mathcal{M}(\tau)).$$

We require that for every discrete marking \mathbf{m} and continuous arc (c_l, T_j) (resp. (T_j, c_l)), that the rate $R((c_l, T_j), (\mathbf{x}, \mathbf{m}))$ (resp. $R((T_j, c_l), (\mathbf{x}, \mathbf{m}))$) be a piecewise continuous function of \mathbf{x} .

Now let $X_l(\tau)$ be the fluid level at time τ in a continuous place $c_l \in \mathcal{P}_c$. The fluid level in each continuous place c_l can never become negative or exceed the bound $B(c_l)$, so the (actual) rate of change over time, τ , when the marking is $\mathcal{M}(\tau)$, is

$$r_l(\mathcal{M}(\tau)) = \frac{dX_l(\tau)}{d\tau} = \begin{cases} r_l^p(\mathcal{M}(\tau)) & \text{if } X_l(\tau) = 0 \text{ and } r_l^p(\mathcal{M}(\tau)) \geq 0 \\ r_l^p(\mathcal{M}(\tau)) & \text{if } X_l(\tau) = B(c_l) \text{ and } r_l^p(\mathcal{M}(\tau)) < 0 \\ 0 & \text{if } X_l(\tau) = 0 \text{ and } r_l^p(\mathcal{M}(\tau)) < 0 \\ 0 & \text{if } X_l(\tau) = B(c_l) \text{ and } r_l^p(\mathcal{M}(\tau)) > 0 \\ r_l^p(\mathcal{M}(\tau)) & \text{if } 0 < X_l(\tau) < B(c_l) \text{ and } r_l^p(\mathcal{M}(\tau^-))r_l^p(\mathcal{M}(\tau^+)) \geq 0 \\ 0 & \text{if } 0 < X_l(\tau) < B(c_l) \text{ and } r_l^p(\mathcal{M}(\tau^-))r_l^p(\mathcal{M}(\tau^+)) < 0. \end{cases} \quad (1)$$

The first two cases of the previous equation concern situations when $X_l(\tau) = 0$ (resp. $X_l(\tau) = B(c_l)$) and the potential rate is $r_l^p(\mathcal{M}(\tau)) \geq 0$ (resp. $r_l^p(\mathcal{M}(\tau)) < 0$). In both cases the actual rate is equal to the potential rate. The third and the fourth cases prevent the fluid level from overcoming the lower and the upper boundaries. The last two cases require a deeper explanation (a reference for a complete discussion of these aspects is [11]). As it has been assumed in [18], the flow rate function $R(\cdot, \cdot)$ is a piecewise continuous function of the continuous part of the marking. The meaning of the last case is that a sign

change (from + to -) in $r_l^p(\mathcal{M}(\tau))$ will “trap” $X_l(\tau)$ in a state in which it will be constant. With this assumption, the analysis of the stochastic process $\mathcal{M}(\tau)$ is simplified (see [11] for a discussion on this type of situation). The fifth case, which is the most common one, accounts for the fact that there is no sign change from + to - in $r_l^p(\mathcal{M}(\tau))$ and hence the actual rate is equal to the potential rate.

Fluid Stochastic Petri Nets are analyzed by transforming them into equivalent Fluid Models. If the FSPN has a single fluid place, then standard FM can be applied. If the FSPN has more than one fluid place, then special FM with multiple continuous variables must be used.

We will begin by describing how to compute the *infinitesimal generator* \mathbf{Q} of the equivalent FM. Since fluid arcs do not change the enabling condition of a transition, standard analysis techniques can be applied to the discrete marking process $\mathbf{m}(\tau)$ [4]. These techniques split the discrete state space into two disjoint subsets; called respectively, the *tangible marking* set and the *vanishing marking* set. Since the process spends no time in vanishing markings, they can be removed and their effect can be included in the transitions between tangible markings. From this point on, we will consider only tangible markings. In GSPNs, the underlying stochastic process is a CTMC, whose infinitesimal generator is a matrix \mathbf{Q} . Each entry q_{ij} represents the rate of transition from a tangible state \mathbf{m}_i to a tangible state \mathbf{m}_j , that is:

$$q_{ij} = \sum_{T_k \in \mathcal{E}(\mathbf{m}_i) \mid \mathbf{m}_i \xrightarrow{T_k} \mathbf{m}_j} F(T_k, \mathbf{m}_i),$$

where $\mathcal{E}(\mathbf{m}_i)$ represents the set of enabled transitions in marking \mathbf{m}_i , and $\mathbf{m}_i \xrightarrow{T_k} \mathbf{m}_j$ means that the firing of transition T_k changes the discrete state of the system from \mathbf{m}_i to \mathbf{m}_j .

In the FSPN model defined in [18], the firing rate of each timed transition can be made dependent on the continuous component of the state. With this extension, the infinitesimal generator matrix must be also dependent on the fluid component of the state, that is $\mathbf{Q}(\mathbf{x}) = \{q_{ij}\}$ where:

$$q_{ij}(\mathbf{x}) = \sum_{T_k \in \mathcal{E}(\mathbf{m}_i) \mid \mathbf{m}_i \xrightarrow{T_k} \mathbf{m}_j} F(T_k, \mathbf{m}_i, \mathbf{x}).$$

The summation considers the transition rates of all the transitions T_k that bring the net from state \mathbf{m}_i to \mathbf{m}_j . In the standard equations that describe a CTMC, the terms on the diagonal of the infinitesimal generator account for the probability of exiting from a state. In this case, we simply define:

$$q_{ii}(\mathbf{x}, \emptyset) = - \sum_{j \neq i} q_{ij}(\mathbf{x}). \quad (2)$$

Matrix $\mathbf{Q}(\mathbf{x})$ is equivalent to the infinitesimal generator of a CTMC, in the sense that each row sum of $\sum_{s \in 2^{\mathcal{P}_c}} Q(\mathbf{x}, s)$, is equal to zero. In other words:

$$\mathbf{Q}(\mathbf{x})\mathbf{1} = \mathbf{0}$$

where $\mathbf{1}$ (respectively $\mathbf{0}$) is a column vector with all the $|\mathcal{S}_d|$ components equal to 1 (resp. 0).

If we have only a single fluid place c_l , the fluid rate matrix $R(\mathbf{x})$, of the underlying fluid model, can be simply computed by defining $r(i, x) = r_l(M)$, where $M = (\mathbf{m}_i, x)$. Then $\mathbf{R}(x) = \text{diag}(r(i, x))$ becomes the diagonal matrix whose components account for the actual flow rate out of the fluid place.

No boundary conditions are needed, since they are included in the definition of the potential flow rate (Equation (1)). Dirac's delta functions in the solution, represent cases where there is a non zero probability of finding the system in a particular marking (both discrete and continuous).

In [16] a new kind of fluid primitive, called Flush-out arcs has been added to the FSPN formalism.

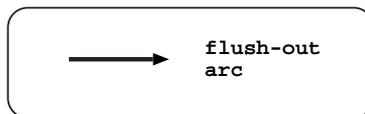


Fig. 3. Flushout arcs

Flush-out (See Figure 3 arcs are special arcs that connect fluid places to timed transition (but not timed transition to fluid places). They are drawn using thick lines. When a transition fires, the places connected with a flush-out arc are emptied in zero time.

Despite their simplicity, Flush-out Arcs can be exploited to obtain many interesting effects, like dropping the content of the transmission buffer. The underlying stochastic model is no longer a standard Fluid Model, but it can be analyzed similarly using appropriate boundary conditions. It has been shown in [16] that FSPNs with flush-out arcs can be used to simulate Non-Markovian Stochastic Petri Nets [13].

3 Analytical Description of Fluid Models

3.1 Introduction to fluid models

Since the behaviour of the considered class of fluid models contain random elements they belong to the large family of stochastic processes. Stochastic processes can be viewed as a set of random variables, which are ordered according to a parameter. In typical engineering applications the parameter represents the time and it takes value either from the natural numbers, $0, 1, 2, \dots$, or from the set of non-negative real numbers. The former case is referred to as discrete time stochastic process and the later as continuous time stochastic process.

Also in typical engineering applications the random variables has the same support set. The characteristics of this support set is the other main feature of the stochastic process. We distinguish the following cases:

- discrete support set, e.g., the number of customers in a queue,

- continuous support set, e.g., the unfinished work in a queue,
- hybrid (continuous and discrete) support set, e.g., the unfinished work and the number of customers.

General continuous and hybrid valued stochastic processes are hard to analyze but, there are special cases which allow the application of simple analysis methods. Focusing on the hybrid valued case the simplest processes are obtained when the continuous part of the model is determined by its discrete part through a very simple function, which is the case with reward models and fluid models.

In both cases a simple function of a discrete state stochastic process governs the evolution of the continuous variable. E.g., the continuous value is increasing or decreasing at a given rate while the discrete value is constant (see Figure 4). In case of reward models this evolution is non-decreasing and unbounded, while in case of fluid models the evolution of the continuous variable is bounded.

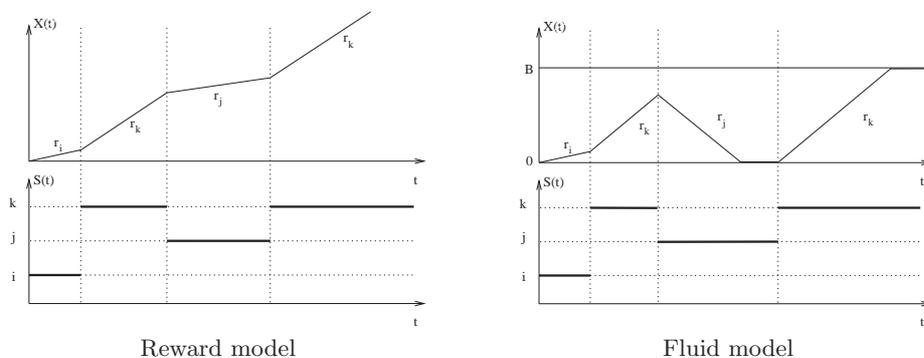


Fig. 4. Evolution of the continuous variable in reward and fluid models.

Definition 1. *Markov property:* A stochastic process is said to enjoy the Markov property at time t , when the future evolution of the process is independent of its past and depend only on the value of the random variable at time t .

Those stochastic processes which enjoy the Markov property at all time are referred to as Markov processes. Continuous time Markov chains, Markov reward models, Markov fluid models are examples of Markov processes. In this chapter we focus on these Markovian cases.

3.2 Classification of fluid models

The following features of fluid models are used for classification:

- Buffer size:
It is commonly assumed that the minimal buffer level is 0. This way the

size of the buffer determines the maximal buffer content. The two main cases are **finite buffer** and **infinite buffer**. In case of an infinite buffer the continuous quantity is only lower bounded at zero and in case of a finite buffer the continuous quantity is lower bounded at zero and upper bounded at B (Figure 5).

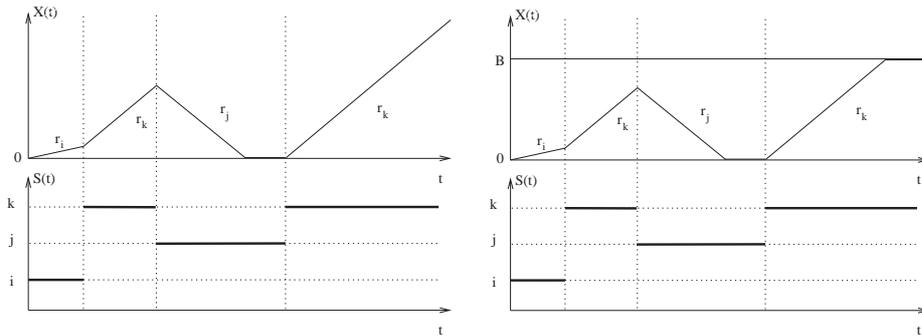


Fig. 5. Infinite and finite buffer fluid models

– Evolution of the continuous variable:

The evolution of the continuous variable depends on the value of the discrete variable, but this dependence can be of two kinds. The case when the continuous variable deterministically increases/decreases as long as the discrete variable is constant, is referred to **first order** fluid model (Figure 6). The case when the increment of the continuous variable during a period when the discrete variable is constant is a normal distributed random quantity is referred to **second order** fluid model (Figure 6).

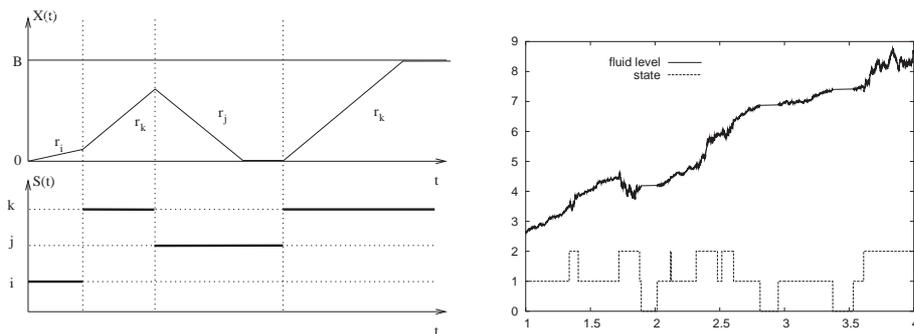


Fig. 6. First and second order fluid models

Second order fluid models can be interpreted as the limiting process of a two dimensional random walk according to Figure 7, where the horizontal dimension represents the discrete variable and the vertical dimension represents the buffer level.

In this model the probabilities of the vertical state transitions determine the mean fluid increase for each value of the discrete variable. It can be seen that the fluid level can increase and also decrease in each column.

Reducing the time step and the granularity of the fluid level of this model to zero results in the second order fluid flow model.

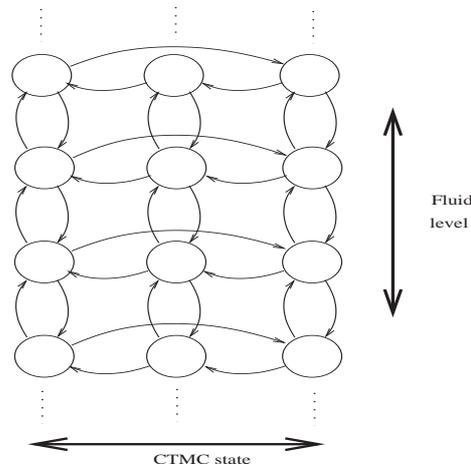


Fig. 7. Interpretation of second order fluid models

- Effect of the buffer content on the discrete variable:
 With this respect there are two main cases. The evolution of the discrete variable can be independent of or can depend on the instantaneous value of the fluid variable. In case of a Markov fluid model, the first means that the discrete part of the model is an “independent” CTMC which modulates the fluid accumulation. In the later case there is a mutual dependence of the continuous and the discrete part of the model and it is not possible to analyze the discrete variable in isolation. The first case is also referred to as space **inhomogeneous model**, since the generator matrix of the discrete variable is constant, i.e., independent of the fluid level, while the second case is also referred to as **fluid level dependent** model.
- Behaviour of the second order model at the boundaries:
 In case of first order fluid models the model behaviour is quite well defined when the fluid level reaches a boundary. When the fluid level reaches the lower boundary (empty buffer) the system must be in a state with a negative

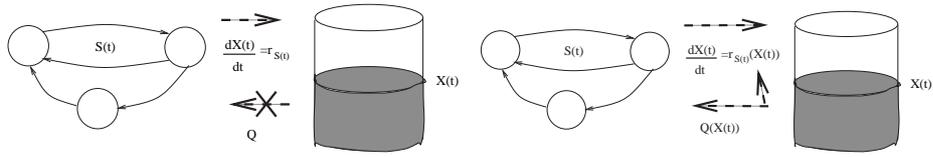


Fig. 8. Markov fluid models with independent and dependent discrete parts

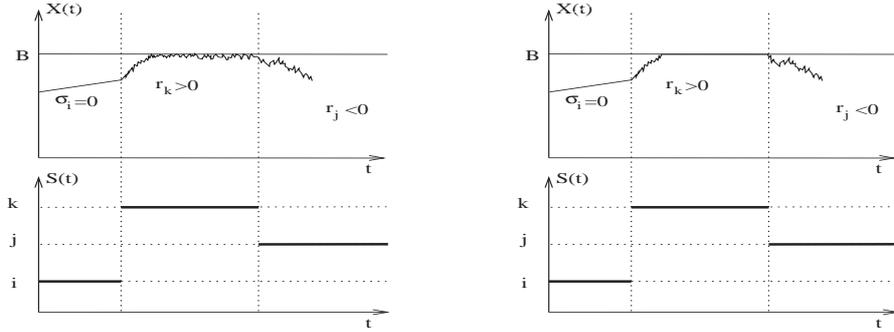


Fig. 9. Second order fluid models with reflecting and absorbing boundary behaviour

fluid rate. After reaching the lower boundary (in a state with a negative fluid rate), the buffer remains empty as long as a transition to a state with positive fluid rate takes place. The system behaviour at the upper boundary (if any) is similar.

The behaviour of second order fluid models is more complex at the boundaries. In this case we might assume a deterministic and a stochastic behaviour depending on the behaviour of the modeled system.

The deterministic boundary behaviour of second order models is more or less identical with the described boundary behaviour of the first order model. The only difference is that the fluid level can become zero also in states with non-negative drift and positive variance. This case is referred to as **absorbing boundary**, since the fluid level gets identical with the boundary for a positive amount of time (Figure 9).

The stochastic boundary behaviour of second order models is similar to the general evolution of these models between the boundaries, where the fluid level alternates randomly all over the time and does not remain constant for a non-zero time period with positive probability. In this case, the fluid level process is reflected as soon as it reaches a boundary, and this way it always remain between the boundaries with probability 1. This case is referred to as **reflecting boundary**, since the fluid level gets reflected at the boundary (Figure 9).

These boundary behaviours can be interpreted using the same random walk approximation as we used for the interpretation of the second order fluid

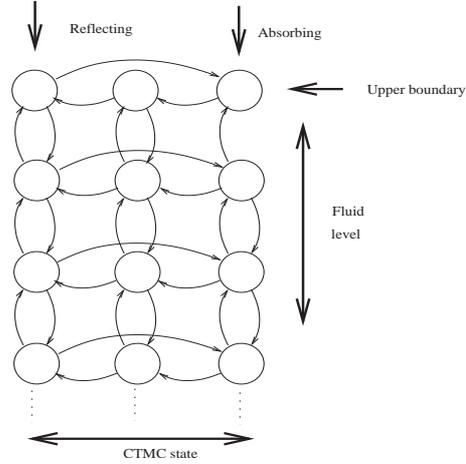


Fig. 10. Interpretation of the boundary behaviour

model. Those states exhibit absorbing boundary behaviour, where there is no vertical state transition out of a boundary state, i.e., the discrete variable must change its value to leave the given boundary state. (Figure 10). In contrast, those states exhibit reflecting behaviour, where there is a vertical transition out of the boundary state, i.e., the fluid level is alternating all over the time.

3.3 Transient behaviour of the fluid level process

In this section we study the transient behaviour of the fluid level process according to a simple to more complex approach. The simplest case is the transient analysis of *first order, infinite buffer, homogeneous* Markov fluid models. Later on we extend this model with finite buffer, second order fluid accumulation and fluid level dependency.

First order, infinite buffer, homogeneous Markov fluid models During an infinitesimally short period of time, $(t, t + \Delta)$, while the continuous variable is i , or equivalently we also say that the (discrete part of the) system is in state i , $(S(\tau) = i, \forall \tau \in (t, t + \Delta))$, the fluid level $(X(t))$ increases at rate r_i when $X(t) > 0$:

$$X(t + \Delta) - X(t) = r_i \Delta$$

that is

$$\frac{d}{dt} X(t) = r_i \quad \text{if } S(t) = i, X(t) > 0.$$

When $X(t) = 0$ the fluid level can not decrease:

$$\frac{d}{dt} X(t) = \max(r_i, 0) \quad \text{if } S(t) = i, X(t) = 0.$$

That is

$$\frac{d}{dt}X(t) = \begin{cases} r_{S(t)} & \text{if } X(t) > 0, \\ \max(r_{S(t)}, 0) & \text{if } X(t) = 0, \end{cases}$$

where $r_{S(t)}$ denotes the fluid rate in the actual discrete state of the process.

First order, finite buffer, homogeneous Markov fluid models When $X(t) = B$ the fluid level can not increase:

$$\frac{d}{dt}X(t) = \min(r_i, 0), \quad \text{if } S(t) = i, X(t) = B.$$

That is

$$\frac{d}{dt}X(t) = \begin{cases} r_{S(t)}, & \text{if } X(t) > 0, \\ \max(r_{S(t)}, 0), & \text{if } X(t) = 0, \\ \min(r_{S(t)}, 0), & \text{if } X(t) = B. \end{cases}$$

Second order, infinite buffer, homogeneous Markov fluid models with reflecting barrier During a sojourn in the discrete state i ($S(t) = i$) in the sufficiently small $(t, t + \Delta)$ interval the distribution of the fluid increment $(X(t + \Delta) - X(t))$ is normal distributed with mean $r_i \Delta$ and variance $\sigma_i^2 \Delta$:

$$X(t + \Delta) - X(t) = \mathcal{N}(r_i \Delta, \sigma_i^2 \Delta), \quad \text{if } S(u) = i, u \in (t, t + \Delta), X(t) > 0,$$

where $\mathcal{N}(r_i \Delta, \sigma_i^2 \Delta)$ denotes a normal distributed random variable with mean $r_i \Delta$ and variance $\sigma_i^2 \Delta$.

If at $X(t) = 0$ the fluid process is reflected immediately in state i , it means that the time spent at the boundary has a 0 measure, and so the probability of staying at the boundary is

$$Pr(X(t) = 0, S(t) = i) = 0.$$

Second order, infinite buffer, homogeneous Markov fluid models with absorbing barrier Between the boundaries the evolution of the process is the same as before.

When the fluid level decreases to zero in an absorbing barrier state, i , the fluid process gets stopped and the fluid level remains zero for a positive amount of time. Due to this behaviour

$$Pr(X(t) = 0, S(t) = i) > 0.$$

On the other hand, due to the absorbing property of the boundary the probability that the fluid level is close to the boundary in an absorbing state is very low,

$$\lim_{\Delta \rightarrow 0} \frac{Pr(0 < X(t) < \Delta, S(t) = i)}{\Delta} = 0.$$

Inhomogeneous (fluid level dependent), first order, infinite buffer Markov fluid models The evolution of the fluid level differs due to the fluid level dependency of the fluid rate $r_i(x)$, where x is the fluid level. This way the fluid level changes as:

$$\frac{d}{dt}X(t) = \begin{cases} r_{S(t)}(X(t)), & \text{if } X(t) > 0, \\ \max(r_{S(t)}(X(t)), 0), & \text{if } X(t) = 0. \end{cases}$$

The evolution of the discrete part also depends on the fluid level. The detailed discussion of this dependence is delayed to the next subsection.

3.4 Transient description of fluid models

The aim of this section is to derive partial differential equations representing the evolution of Markov fluid models in time. To this end we introduce the following notations:

- $\pi_i(t) = Pr(S(t) = i)$ – state probability,
- $u_i(t) = Pr(X(t) = B, S(t) = i)$ – buffer full probability,
- $\ell_i(t) = Pr(X(t) = 0, S(t) = i)$ – buffer empty probability,
- $p_i(t, x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} Pr(x < X(t) < x + \Delta, S(t) = i)$ – fluid density.

Based on these notations the total probability law for time t gives

$$\pi_i(t) = \ell_i(t) + u_i(t) + \int_x p_i(t, x) dx. \quad (3)$$

In the simplest considered case the evolution of the discrete variable is independent of the fluid level, hence the discrete part of the model is a continuous time Markov chains (CTMC).

Continuous time Markov chains We start with the transient behaviour of CTMCs and based on that we extend the analysis to the various fluid models. The transient behaviour of CTMCs is characterized by the transition rates that are determined by the transition probabilities as follows

$$\lim_{\Delta \rightarrow 0} \frac{Pr(S(t + \Delta) = j | S(t) = i)}{\Delta} = q_{ij}. \quad (4)$$

The commonly applied forward argument to analyze Markovian stochastic processes is based on the analysis of the short term behaviour of the process. We can say that a CTMC which is in state i at time t can do 3 different things in the $(t, t + \Delta)$ interval:

- no state transition:

The process can remain in state i during the whole period with probability $1 - \sum_{j, j \neq i} q_{ij} \Delta + \sigma(\Delta)$, where $q_{ij} \Delta$ is the probability of a state transition from state i to j in the $(t, t + \Delta)$ interval and $\sigma(\Delta)$ is a small error term that

quickly vanishes as Δ tends to zero, i.e., $\lim_{\Delta \rightarrow 0} \sigma(\Delta)/\Delta = 0$. Introducing the notation, $q_{ii} = -\sum_{j,j \neq i} q_{ij}$, we can write the probability of no state transition in the $(t, t + \Delta)$ interval as $1 + q_{ii}\Delta + \sigma(\Delta)$. Note that q_{ii} is negative.

– one state transition:

The process can have a state transition from i to j during the $(t, t + \Delta)$ interval with probability $q_{ji}\Delta + \sigma(\Delta)$.

– more than one state transitions:

The probability of having more than one state transitions in a short interval is quickly vanishes as Δ tends to zero, it is $\sigma(\Delta)$.

Based on these three options we can evaluate the probability of being in state i at time $t + \Delta$ ($\pi_i(t + \Delta)$) as a function of the probability of being in the various states at time t ($\pi_j(t)$):

$$\pi_i(t + \Delta) = \left(1 + q_{ii}\Delta + \sigma(\Delta)\right)\pi_i(t) + \sum_{j,j \neq i} \left(q_{ji}\Delta + \sigma(\Delta)\right)\pi_j(t) + \sigma(\Delta).$$

A $\sigma(\Delta)$ function multiplied with a bounded function ($0 \leq \pi_i(t) \leq 1$) remains to be a $\sigma(\Delta)$ function as well as the finite sum of such functions. Using this we can rearrange the expression to

$$\pi_i(t + \Delta) - \pi_i(t) = q_{ii}\Delta\pi_i(t) + \sum_{j,j \neq i} q_{ji}\Delta\pi_j(t) + \sigma(\Delta) = \sum_j q_{ji}\Delta\pi_j(t) + \sigma(\Delta).$$

Dividing both sides by Δ and making the $\Delta \rightarrow 0$ limit we have

$$\frac{\pi_i(t + \Delta) - \pi_i(t)}{\Delta} = \sum_j q_{ji}\pi_j(t) + \frac{\sigma(\Delta)}{\Delta}$$

and

$$\frac{d\pi_i(t)}{dt} = \sum_j \pi_j(t)q_{ji}. \quad (5)$$

(5) is the differential equation describing the transient behaviour of CTMCs. We apply the same forward approach to evaluate the transient behaviour of Markov fluid models.

First order, infinite buffer, homogeneous Markov fluid models We perform the same analysis for the fluid density using these 3 possible events during the $(t, t + \Delta)$ interval. For simplicity we neglect the unnecessary $\sigma(\Delta)$ terms.

If $S(t + \Delta) = i$, then during the $(t, t + \Delta)$ interval the CTMC

– stays in i and increases the fluid level with $r_i\Delta$ with probability $1 + q_{ii}\Delta$,

- moves from k to i and changes the fluid level with $\mathcal{O}(\Delta)$ with probability $q_{ki}\Delta$, where $\mathcal{O}(\Delta)$ is a function which vanishes as Δ tends to zero, i.e., $\lim_{\Delta \rightarrow 0} \mathcal{O}(\Delta) = 0$ (in this particular case the change of the fluid level is between $r_i\Delta$ and $r_j\Delta$),
- has more than 1 state transition with probability $\sigma(\Delta)$.

Considering these three cases we can express the fluid density at time $t + \Delta$ as a function of the fluid density at time t :

$$p_i(t + \Delta, x) = (1 + q_{ii}\Delta) p_i(t, x - r_i\Delta) + \sum_{k \in \mathcal{S}, k \neq i} q_{ki}\Delta p_k(t, x - \mathcal{O}(\Delta)) + \sigma(\Delta).$$

Rearranging the terms, dividing both sides by Δ and making the $\Delta \rightarrow 0$ limit gives

$$p_i(t + \Delta, x) - p_i(t, x - r_i\Delta) = \sum_{k \in \mathcal{S}} q_{ki}\Delta p_k(t, x - \mathcal{O}(\Delta)) + \sigma(\Delta),$$

$$\frac{p_i(t + \Delta, x) - p_i(t, x)}{\Delta} + r_i \frac{p_i(t, x) - p_i(t, x - r_i\Delta)}{r_i\Delta} = \sum_{k \in \mathcal{S}} q_{ki} p_k(t, x - \mathcal{O}(\Delta)) + \frac{\sigma(\Delta)}{\Delta},$$

$$\boxed{\frac{\partial}{\partial t} p_i(t, x) + r_i \frac{\partial}{\partial x} p_i(t, x) = \sum_{k \in \mathcal{S}} q_{ki} p_k(t, x)}. \quad (6)$$

(6) is the basic partial differential equation describing the transient behaviour of Markov fluid models. Indeed this equation describes the process behaviour during the period while the fluid level is between the boundaries.

The model behaviour at the boundaries can be obtained by the same forward argument. If $r_i > 0$, then the fluid level increases in state i , which means that the buffer cannot be empty in state i , i.e., $l_i(t) = Pr(X(t) = 0, S(t) = i) = 0$.

If $r_i \leq 0$, we can consider the same 3 cases for the $(t, t + \Delta)$ interval:

- If there is no state transition the fluid level is zero at $t + \Delta$ if it was zero at t or if it was between 0 and $-r_i\Delta$ at t .
- If there is one state transition in the interval, the fluid level was zero at t or if it was between 0 and $\mathcal{O}(\Delta)$ (between $-r_i\Delta$ and $-r_k\Delta$) at t .
- The case of having more than one state transitions in the interval, is treated in the same way as before.

$$\begin{aligned}
\ell_i(t + \Delta) = & \\
& (1 + q_{ii}\Delta) \left(\ell_i(t) + \underbrace{\int_0^{-r_i\Delta} p_i(t, x) dx}_{*} \right) + \\
& \sum_{k \in \mathcal{S}, k \neq i} q_{ki}\Delta \left(\ell_k(t) + \underbrace{\int_0^{\mathcal{O}(\Delta)} p_k(t, x) dx}_{\mathcal{O}(\Delta)} \right) + \\
& \sigma(\Delta) .
\end{aligned}$$

When $x \leq -r_i\Delta$, then using the first elements of the Taylor series of $p_i(t, x)$, we have

$$p_i(t, x) = p_i(t, 0) + xp'_i(t, 0) + \sigma(\Delta) ,$$

and substituting it into the previous expression we obtain

$$\begin{aligned}
* &= \int_0^{-r_i\Delta} p_i(t, x) dx \\
&= \int_0^{-r_i\Delta} p_i(t, 0) dx + \int_0^{-r_i\Delta} xp'_i(t, 0) dx + \int_0^{-r_i\Delta} \sigma(\Delta) dx \\
&= -r_i\Delta p_i(t, 0) + \underbrace{\frac{(-r_i\Delta)^2}{2} p'_i(t, 0)}_{\sigma(\Delta)} + \underbrace{(-r_i\Delta) \sigma(\Delta)}_{\sigma(\Delta)} .
\end{aligned}$$

From which we can calculate the differential equation for the empty buffer probability using the same steps as before:

$$\begin{aligned}
\ell_i(t + \Delta) = & (1 + q_{ii}\Delta) \left(\ell_i(t) - r_i\Delta p_i(t, 0) + \sigma(\Delta) \right) + \\
& \sum_{k \in \mathcal{S}, k \neq i} q_{ki}\Delta \left(\ell_k(t) + \mathcal{O}(\Delta) \right) + \sigma(\Delta) ,
\end{aligned}$$

$$\begin{aligned}
\ell_i(t + \Delta) - \ell_i(t) = & q_{ii}\Delta \ell_i(t) - r_i\Delta p_i(t, 0) + \\
& \sum_{k \in \mathcal{S}, k \neq i} q_{ki}\Delta \left(\ell_k(t) + \mathcal{O}(\Delta) \right) + \sigma(\Delta) ,
\end{aligned}$$

$$\begin{aligned}
\frac{\ell_i(t + \Delta) - \ell_i(t)}{\Delta} = & \\
& - r_i p_i(t, 0) + \sum_{k \in \mathcal{S}} q_{ki} \left(\ell_k(t) + \mathcal{O}(\Delta) \right) + \frac{\sigma(\Delta)}{\Delta} ,
\end{aligned}$$

$$\boxed{\frac{d}{dt}\ell_i(t) = -r_i p_i(t,0) + \sum_{k \in \mathcal{S}} q_{ki} \ell_k(t)} . \quad (7)$$

Having these expressions we can conclude the transient description of first order, infinite buffer, homogeneous Markov fluid models. The fluid density is governed by (6) while the empty buffer probability is $\ell_i(t) = 0$ if $r_i > 0$ and (7) if $r_i \leq 0$. There is no simple symbolic solution to this set of differential equations. When the initial condition of the fluid model is known, it can be solved using numerical techniques. The solution has to fulfill the following equations:

$$\int_0^\infty p_i(t,x)dx + \ell_i(t) = \pi_i(t) . \quad (8)$$

$$\pi_i(t) = \pi_i(0)e^{Qt}, \quad (9)$$

where (8) is the special form of (3) for infinite buffer model and (9) is the solution of (5).

First order, finite buffer, homogeneous behaviour The presence of an upper boundary at B does not change the transient description a lot. It leaves the behaviour at the lower boundary, (7), unchanged, it reduces the validity of (6) to $0 < x < B$ and it introduces a differential equation, very similar to (7) for the upper boundary. That is, $u_i(t) = 0$ if $r_i < 0$ and

$$\frac{d}{dt}u_i(t) = r_i p_i(t,B) + \sum_{k \in \mathcal{S}} q_{ki} u_k(t), \quad (10)$$

if $r_i \geq 0$. (10) is obtained in the same way as (7).

Second order, infinite buffer, homogeneous behaviour. The case of second order Markov fluid model can be analyzed using the same method based on the short term behaviour of the Markov model. We derive the fluid density at time $t + \Delta$ based on the fluid density at time t :

- If there is no state transition in the $(t, t + \Delta)$ interval we need to evaluate a convolution with respect to the pdf of the normal distributed amount of fluid accumulated over the $(t, t + \Delta)$ interval, $f_{\mathcal{N}(\Delta r_i, \Delta \sigma_i^2)}(u)$. For simplicity we set the limits of this integration to $-\infty$ and ∞ . It is to avoid the introduction of additional vanishing error terms.
- The analysis of the case with one state transition is also simplified. A convolution with finally vanishing terms should be taken into consideration in a more detailed analysis. The analysis of the term without state transition indicates how the term with one state transition vanishes, but we do not detail this point here.

$$p_i(t + \Delta, x) = (1 + q_{ii}\Delta) \underbrace{\int_{-\infty}^{\infty} p_i(t, x - u) f_{\mathcal{N}(\Delta r_i, \Delta \sigma_i^2)}(u) du}_{**} + \sum_{k \in \mathcal{S}, k \neq i} q_{ki} \Delta p_k(t, x - \mathcal{O}(\Delta)) + \sigma(\Delta)$$

To obtain the under braced term we use the Taylor expansion again, but now with 3 terms:

$$p_i(t, x - u) = p_i(t, x) - up'_i(t, x) + \frac{u^2}{2} p''_i(t, x) + \mathcal{O}(u)^3.$$

Based on this expansion we have:

$$** = p_i(t, x) \underbrace{\int_{-\infty}^{\infty} f_{\mathcal{N}(\Delta r_i, \Delta \sigma_i^2)}(u) du}_1 - p'_i(t, x) \underbrace{\int_{-\infty}^{\infty} u f_{\mathcal{N}(\Delta r_i, \Delta \sigma_i^2)}(u) du}_{\Delta r_i} + p''_i(t, x) \underbrace{\int_{-\infty}^{\infty} \frac{u^2}{2} f_{\mathcal{N}(\Delta r_i, \Delta \sigma_i^2)}(u) du}_{\Delta^2 r_i^2 + \Delta \sigma_i^2 / 2 = \Delta \sigma_i^2 / 2 + \sigma(\Delta)} + \underbrace{\int_{-\infty}^{\infty} \mathcal{O}(u)^3 f_{\mathcal{N}(\Delta r_i, \Delta \sigma_i^2)}(u) du}_{\mathcal{O}(\Delta)^2 = \sigma(\Delta)}.$$

Back substituting this results and performing the same steps as before we obtain

$$p_i(t + \Delta, x) = (1 + q_{ii}\Delta) \left(p_i(t, x) - p'_i(t, x) \Delta r_i + p''_i(t, x) \Delta \sigma_i^2 / 2 \right) + \sum_{k \in \mathcal{S}, k \neq i} q_{ki} \Delta p_k(t, x - \mathcal{O}(\Delta)) + \sigma(\Delta),$$

$$p_i(t + \Delta, x) - p_i(t, x) = q_{ii} \Delta p_i(t, x) - p'_i(t, x) \Delta r_i + p''_i(t, x) \Delta \sigma_i^2 / 2 + \sum_{k \in \mathcal{S}, k \neq i} q_{ki} \Delta p_k(t, x - \mathcal{O}(\Delta)) + \sigma(\Delta),$$

$$\boxed{\frac{\partial}{\partial t} p_i(t, x) + \frac{\partial}{\partial x} p_i(t, x) r_i - \frac{\partial^2}{\partial x^2} p_i(t, x) \frac{\sigma_i^2}{2} = \sum_{k \in \mathcal{S}} q_{ki} p_k(t, x).} \quad (11)$$

(11) also justifies the name of this fluid models. In this case not only the first derivative of the fluid density, but also the second one appear in the partial differential equation describing the transient behaviour of the model.

Boundary condition with reflecting barrier The boundary condition of second order Markov fluid models depends on the type of the boundary. In case of reflecting barriers the probability of empty buffer is zero, $\ell_i(t) = 0$ and the

initial value of the fluid density can be computed based on (3) using (11) and (5).

Since the buffer is infinite buffer and $\ell_i(t) = 0$, we have

$$\int_0^{\infty} p_i(t, x) dx = \pi_i(t) .$$

Taking the derivatives of both side with respect to t results

$$\int_{x=0}^{\infty} \frac{\partial}{\partial t} p_i(t, x) dx = \frac{\partial}{\partial t} \pi_i(t)$$

Substituting (11) into the left and (5) into the right hand side we have

$$\int_{x=0}^{\infty} -\frac{\partial p_i(t, x)}{\partial x} r_i + \frac{\partial^2 p_i(t, x)}{\partial x^2} \frac{\sigma_i^2}{2} + \sum_{k \in \mathcal{S}} q_{ki} p_k(t, x) dx = \sum_{k \in \mathcal{S}} q_{ki} \pi_i(t),$$

from which we obtain the boundary condition as

$$-r_i \underbrace{\left[p_i(t, x) \right]_{x=0}^{\infty}}_{-p_i(t,0)} + \frac{\sigma_i^2}{2} \underbrace{\left[p_i'(t, x) \right]_{x=0}^{\infty}}_{-p_i'(t,0)} + \sum_{k \in \mathcal{S}} q_{ki} \underbrace{\int_{x=0}^{\infty} p_k(t, x) dx}_{\pi_i(t)} = \sum_{k \in \mathcal{S}} q_{ki} \pi_i(t) ,$$

$$\boxed{r_i p_i(t, 0) - \frac{\sigma_i^2}{2} p_i'(t, 0) = 0} \quad (12)$$

Fluid level dependent model behaviour Up to now we considered fluid models where between the boundaries the fluid level does not effect the evolution of the system. It is not always the case in practice and the presented analytical description of fluid models allows to integrate fluid level dependence into the transient description in a simple way.

As a consequence we assumed that the transition rate of the discrete part of the process, q_{ij} , the mean and the variance of the fluid changing rate, r_i and σ_i , respectively, are independent of the current fluid level. When these quantities depend on the fluid level we have the following model behaviour.

$$\lim_{\Delta \rightarrow 0} \frac{Pr(S(t + \Delta) = j | S(t) = i, X(t))}{\Delta} = q_{ij}(X(t)) .$$

When the first order model stays in state i during the $(t, t + \Delta)$ interval and the fluid level is between the boundaries

$$X(t + \Delta) - X(t) = r_i(X(t))\Delta + \sigma(\Delta),$$

and when the second order model does the same

$$X(t + \Delta) - X(t) = \mathcal{N}(r_i(X(t))\Delta, \sigma_i^2(X(t))\Delta) + \sigma(\Delta).$$

Formally it is easy to incorporate fluid level dependency into all previous equations by making the transition rates of the discrete part, the mean and the variance of the fluid changing rate depend on the fluid level, i.e., $q_{ij}(x)$, $r_i(x)$ and $\sigma_i(x)$, respectively. This way, e. g., (6) becomes

$$\frac{\partial}{\partial t} p_i(t, x) + r_i(x) \frac{\partial}{\partial x} p_i(t, x) = \sum_{k \in \mathcal{S}} q_{ki}(x) p_k(t, x),$$

and the associated boundary equation, (7) becomes, if $r_i(0) < 0$ (and $r_i(x)$ is continuous):

$$\frac{d}{dt} \ell_i(t) = -r_i(0) p_i(t, 0) + \sum_{k \in \mathcal{S}} q_{ki}(0) \ell_k(t).$$

General case We summarize the results by providing the most general equation and present the ways to simplify it in case of special fluid models.

First of all we compose vector equations out of the set of scalar equations presented before. Let $p(t, x) = \{p_i(t, x)\}$, $\ell(t) = \{\ell_i(t)\}$ and $u(t) = \{u_i(t)\}$ be the row vectors of fluid densities, empty buffer probabilities and buffer full probabilities respectively, further more let $\mathbf{Q}(x) = \{q_{ij}(x)\}$, $\mathbf{R}(x) = \text{Diag}\langle r_i(x) \rangle$ and $\mathbf{S}(x) = \text{Diag}\langle \frac{\sigma_i^2(x)}{2} \rangle$ be the generator matrix of the discrete variable, the diagonal matrix of the mean fluid rates and the diagonal matrix of the variance parameter of the fluid process.

The most general equations are obtained with **second order**, **finite buffer**, **fluid level dependent** fluid models, where we do not define the boundary behaviour yet:

$$\begin{aligned} \frac{\partial p(t, x)}{\partial t} + \frac{\partial p(t, x)}{\partial x} \mathbf{R}(x) - \frac{\partial^2 p(t, x)}{\partial x^2} \mathbf{S}(x) &= p(t, x) \mathbf{Q}(x), \\ p(t, 0) \mathbf{R}(0) - p'(t, 0) \mathbf{S}(0) &= \ell(t) \mathbf{Q}(0), \\ -p(t, B) \mathbf{R}(B) + p'(t, B) \mathbf{S}(B) &= u(t) \mathbf{Q}(B), \end{aligned} \quad (13)$$

These general equations simplify as follows, according to the boundary behaviour of the model:

- if $\sigma_i = 0$ and $r_i(x)$ is positive and continuous around zero then $\ell_i(t) = 0$, if $\sigma_i = 0$ and $r_i(x)$ is negative and continuous around B then $u_i(t) = 0$.
- if $\sigma_i > 0$ and the lower boundary is reflecting in state i then $\ell_i(t) = 0$, if $\sigma_i > 0$ and the upper boundary is reflecting in state i then $u_i(t) = 0$.
- if $\sigma_i > 0$ and the lower boundary is absorbing in state i then $p_i(t, 0) = 0$, if $\sigma_i > 0$ and the upper boundary is absorbing in state i then $p_i(t, B) = 0$.

The special cases of this general case are:

- the first order model: green parts vanish ,
- the infinite buffer model: blue equation vanishes ,
- the fluid level independent model: $\mathbf{Q}(x), \mathbf{R}(x), \mathbf{S}(x)$ become $\mathbf{Q}, \mathbf{R}, \mathbf{S}$.

Normalizing condition: In case of transient analysis the set of differential equations is accompanied with an initial condition that defines the normalization of the model. Indeed the initial condition should fulfill the normalizing condition

$$\int_0^B p(0, x) dx \mathbb{1} + \ell(0) \mathbb{1} + u(0) \mathbb{1} = 1.$$

The set of differential equations we presented in this section preserves the probability, which means that if the initial condition satisfies the normalizing condition, then for all $t > 0$ the following normalizing condition holds

$$\int_0^B p(t, x) dx \mathbb{1} + \ell(t) \mathbb{1} + u(t) \mathbb{1} = 1.$$

3.5 Stationary description of fluid models

The presented transient description of fluid models describes also the time limiting behaviour of these models, but as it is common with several other stochastic models, the direct stationary analysis is more efficient when only the stationary behaviour is of interest. The general approach to obtain the stationary description of fluid models is to make the t goes to infinity limit in the transient equations.

Two main questions has to be considered during this transition. If the transient functions tend to stationary values, and if this value is unique, i.e., independent of the initial condition in the sense that it converges to the same limit starting from any valid initial condition.

The typical behaviour of the above differential equations is that the solution either converges to a finite value or diverges, but does not exhibit strange behaviours like cyclic alternation, etc. Finite buffer models usually converge. To decide if an infinite buffer model converges to a proper stationary distribution we need the stability property.

Definition 2. A fluid model is said to be stable, if for $\forall x \in \mathbb{R}^+, \forall i \in \mathcal{S}$ the time to empty the buffer

$$T_i^E(x) = \min_{t>0} (X(t) = 0 | X(0) = x, S(0) = i)$$

has a finite mean (i.e., $E(T_i^E(x)) < \infty$).

Stable infinite buffer models usually converge. It is easy to decide if a model is stable in case of fluid level independent Markov fluid models. The condition of stability is

$$\sum_{i \in \mathcal{S}} \pi_i r_i < 0,$$

where π_i is the stationary distribution of the discrete part of the model. The stability of fluid level dependent Markov fluid models is more complex to decide. It requires the solution of the differential equations describing the stationary behaviour of the process.

To decide if the stationary behaviour is unique we need the ergodic property.

Definition 3. A fluid model is said to be ergodic, if for $\forall x, y \in \mathbb{R}^+, \forall i, j \in \mathcal{S}$ the transition time

$$T_{i,j}(x, y) = \min_{t > 0} (X(t) = y, S(t) = j | X(0) = x, S(0) = i)$$

has a finite mean (i.e., $E(T) < \infty$).

The stationary behaviour of ergodic fluid models is independent of the initial condition.

Stationary equations Assuming the following limits exists, describe a proper distribution and independent of the initial condition we present the stationary equations obtained from the transient ones.

- $\pi_i = \lim_{t \rightarrow \infty} Pr(S(t) = i)$ – state probability,
- $u_i = \lim_{t \rightarrow \infty} Pr(X(t) = B, S(t) = i)$ – buffer full probability,
- $\ell_i = \lim_{t \rightarrow \infty} Pr(X(t) = 0, S(t) = i)$ – buffer empty probability,
- $p_i(x) = \lim_{t \rightarrow \infty} \lim_{\Delta \rightarrow 0} 1/\Delta Pr(x < X(t) < x + \Delta, S(t) = i)$ – fluid density,
- $F_i(x) = \lim_{t \rightarrow \infty} Pr(X(t) < x, S(t) = i)$ – fluid distribution.

The stationary counterpart of (13) can be obtained by making the t goes to infinity limit on both sides of the equations:

$$\begin{aligned} p'(x) \mathbf{R}(x) - p''(x) \mathbf{S}(x) &= p(x) \mathbf{Q}(x), \\ p(0) \mathbf{R}(0) - p'(0) \mathbf{S}(0) &= \ell \mathbf{Q}(0), \\ -p(B) \mathbf{R}(B) + p'(B) \mathbf{S}(B) &= u \mathbf{Q}(B), \end{aligned} \tag{14}$$

where the fluid rate and the boundary conditions determine the following variables:

- if $\sigma_i = 0$ and $r_i(x)$ is positive and continuous around zero then $\ell_i = 0$, if $\sigma_i = 0$ and $r_i(x)$ is negative and continuous around B then $u_i = 0$,

| | transient | stationary |
|------------------------|-----------|---------------|
| differential equations | [8] | [14] |
| spectral decomposition | + | [20, 1, 7, 6] |
| randomization | [28] | [30, 29] |
| transform domain | [27] | + |
| Markov regenerative | [2] | + |
| matrix exponent | + | [15] |

Table 1. Solution methods for Markov fluid models

- if $\sigma_i > 0$ and the lower boundary is reflecting in state i then $\ell_i = 0$, if $\sigma_i > 0$ and the upper boundary is reflecting in state i then $u_i = 0$,
- if $\sigma_i > 0$ and the lower boundary is absorbing in state i then $p_i(0) = 0$, if $\sigma_i > 0$ and the upper boundary is absorbing in state i then $p_i(B) = 0$.

Normalizing condition: In case when the stationary solution is computed based on (14), we cannot utilize the information about the initial condition of the model, but the solution must fulfill the normalizing condition:

$$\int_0^B p(x) dx \mathbf{1} + \ell \mathbf{1} + u \mathbf{1} = 1.$$

4 Solution methods

There are several different ways to evaluate Markov fluid models. They differ in their applied analysis approach, provided results and applicability. It is possible to obtain symbolic solution for rather small models (Markov fluid models with less than 5 discrete states), but for larger models the application of numerical methods is feasible only. Table 1 presents a summary of some potential approaches and classifies some research papers according to their applied approaches. In this section we summarize some analysis approaches, but we do not intend to provide a complete view.

Table 1 indicates that all of the mentioned solution methods are applicable to both, the transient and the stationary analysis, but in a different way. In case of transient analysis we have a set of partial differential equations (13), a set of boundary conditions, and a set of explicit initial conditions. Starting from this initial condition it is possible to evaluate the model behaviour using a forward analysis approach. In case of stationary analysis we have a set of ordinary differential equations (14), a set of boundary conditions, and a normalizing conditions. A difficulty of the stationary analysis with respect to the transient one is that normalizing condition does not provide an explicit expression to start the solution from. Apart of this the transient analysis is more complex than the stationary one, since we have one variable (t) more in the transient case.

In the rest of this section we summarize the main ideas of some selected solution methods.

4.1 Transient solution methods

Numerical solution of differential equations Chen et al. proposed a discretization based numerical technique to evaluate the transient behaviour of fluid models [8]. The main strength of their approach is that that all mentioned model behaviour can be analyzed with it. Indeed this is the only approach for the transient analysis of fluid level dependent models. The proposed approach starts from the initial condition, and computes (approximates) the evolution of the fluid distribution step-by-step in Δ long time intervals at some fluid levels based on the differential equations and the boundary condition.

Randomization Randomization is an effective numerical analysis approach that is widely used for the transient analysis of CTMCs, i.e., for the numerical solution of (5). It is numerically stable procedure where the convex combination of probabilities (non-negative numbers less or equal to one) are computed. The procedure is based on a symbolic solution of (5). Sericola extended this technique to the transient analysis of first order, infinite buffer, fluid level independent Markov fluid models [28]. Indeed he provided a symbolic solution of (6) in the following form:

$$F_i^c(t, x) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \sum_{k=0}^n \binom{n}{k} x_j^k (1 - x_j)^{n-k} b_i^{(j)}(n, k),$$

where $F_i^c(t, x) = Pr(X(t) > x, S(t) = i)$, $x_j = \frac{x - r_{j-1}^+ t}{r_j t - r_{j-1}^+ t}$ if $x \in [r_{j-1}^+ t, r_j t)$, and $b_i^{(j)}(n, k)$ is defined by initial value and a simple recursion.

The main properties of this randomization based solution method are as follows:

- the expression with the given recursive formulas is a solution of the differential equation,
- the initial value of $b_i^{(j)}(n, k)$ is set to fulfill the boundary condition,
- due to the fact that $0 \leq x_j \leq 1$ we have the same numerical stability properties as for the transient analysis of CTMCs:
 - convex combination of non-negative numbers are computed, and hence the floating point errors has a limited effect and it does not cause problems like “ringing” (change of sign),
- the initial fluid level must be $X(0) = 0$ (extension to $X(0) > 0$ and to finite buffer is not available).

Markov regenerative approach Ahn and Ramaswami recommended to divide the transient analysis of first order, infinite buffer, fluid level independent Markov fluid models into periods according to the busy/idle state of the buffer [2]. When T_i is the beginning of the i th busy (non-empty) period of the fluid buffer then the $(S(t_i), T_i)$ pairs form a Markov renewal sequence. The analysis of a busy and an idle cycle, i.e., a (T_{i-1}, T_i) interval, is divided into two parts. The

idle period is easier to analyze. Its length is phase type distributer. The analysis of the busy period is more complex, but Ahn and Ramaswami recognized the similarities between fluid and queueing models and provided a solution method based on Matrix analytic technique.

Transform domain description Ren and Kobayashi proposed a solution technique based on the Laplace transform domain description first order, infinite buffer, fluid level independent Markov fluid models [27]. The Laplace transform of (6) is

$$p^{**}(s, v) = (\underbrace{p^*(0, v)}_{\text{initial condition}} + \underbrace{p^*(s, 0)}_{\text{unknown}} \mathbf{R}) (s\mathbf{I} + v\mathbf{R} - \mathbf{Q})^{-1}.$$

where $p^{**}(s, v)$ must be analytical. Since $p^*(0, v)$ is known from the initial condition $p^*(s, 0)$ is set to eliminates the roots of $\det(s\mathbf{I} + v\mathbf{R} - \mathbf{Q})$.

This approach provides a closed form solution also for the case of initially non-empty buffer ($X(0) > 0$), but its applicability is limited to small models (less than 5 discrete states) since it is based on complex symbolic functional analysis.

4.2 Stationary solution methods

Spectral decomposition One of the first papers on the application of Markov fluid models for modeling of telecommunication systems [7] already applied the spectral decomposition method for the solution of the obtained model. Later on Kulkarni presented a survey on spectral decomposition based analysis of first order, infinite and finite buffer, fluid level independent Markov fluid models [20].

To present these results we need the following notations. The set of discrete states are partitioned as follows:

- \mathcal{S}^+ : $i \in \mathcal{S}^+$ iff $\sigma_i > 0$ – second order states,
- \mathcal{S}^0 : $i \in \mathcal{S}^0$ iff $r_i = 0$ and $\sigma_i = 0$, – zero states,
- \mathcal{S}^{0+} : $i \in \mathcal{S}^{0+}$ iff $r_i > 0$ and $\sigma_i = 0$, – positive first order states,
- \mathcal{S}^{0-} : $i \in \mathcal{S}^{0-}$ iff $r_i < 0$ and $\sigma_i = 0$, – negative first order states,
- $\mathcal{S}^* = \mathcal{S}^{0-} \cup \mathcal{S}^{0+}$, – first order states.

The general form of the solution of the differential equation $p'(x)\mathbf{R} = p(x)\mathbf{Q}$ is

$$p(x) = e^{\lambda x} \phi,$$

where ϕ is a row vector. Substituting this solution into the differential equation we get the characteristic equation:

$$\phi(\lambda\mathbf{R} - \mathbf{Q}) = 0,$$

whose solutions are obtained at

$$\det(\lambda\mathbf{R} - \mathbf{Q}) = 0.$$

The characteristic equation has $|\mathcal{S}^{0+}| + |\mathcal{S}^{0-}|$ solutions, with $|\mathcal{S}^{0+}|$ negative eigenvalues, 1 zero eigenvalue, and $|\mathcal{S}^{0-}| - 1$ positive eigenvalues. Having these eigenvalues and eigenvectors the solution is

$$p(x) = \sum_{j=1}^{|\mathcal{S}^{0+}|+|\mathcal{S}^{0-}|} a_j e^{\lambda_j x} \phi_j,$$

and the a_j coefficients are set to fulfill the boundary and normalizing conditions. In the *infinite buffer* case these conditions are:

- $p(0) \mathbf{R} = \ell \mathbf{Q}$,
- $\ell_i = 0$ if $r_i > 0$, and
- $\int_0^\infty p_i(x) dx + \ell_i = \pi_i$.

From which $a_j = 0$ for $\lambda_j > 0$ and the a_j coefficients for $\lambda_j < 0$ are obtained from the solution of the linear system of equations determined by the conditions of infinite buffer.

In the *finite buffer* case these conditions are:

- $p(0) \mathbf{R} = \ell \mathbf{Q}$, $p(B) \mathbf{R} = u \mathbf{Q}$,
- $\ell_i = 0$ if $r_i > 0$, $u_i = 0$ if $r_i < 0$, and
- $\int_0^\infty p_i(x) dx + \ell_i + u_i = \pi_i$.

From which all a_j coefficients are obtained from the linear system of equations determined by the conditions of infinite buffer.

The result on the sign of the eigenvalues has the following consequences:

- If $|\mathcal{S}^{0-}| > 1$ and the buffer is infinite then there is at least one positive eigenvalue, which needs to be excluded from the solution (if the fluid model is stable). The exclusion of the positive eigenvalue makes the spectral decomposition necessary.
- If $|\mathcal{S}^{0-}| = 1$ and the buffer is infinite, then all eigenvalues are non-positive and there is no need to exclude any eigenvalue from the solution.
- If the buffer is finite all eigenvalues plays role in the solution, i.e., there is no need for special treatment of the positive eigenvalues.

Matrix exponent An algebraic approach was proposed by Gribaudo and German to solve the set of equations given for first order, finite buffer, fluid level independent Markov fluid models [15]. Assuming that $|\mathcal{S}^0| = 0$ and $\mathcal{S} = \mathcal{S}^*$ they introduced $v = \ell + u$, \mathbf{Q}^- , \mathbf{Q}^+ , where $q_{ij}^- = q_{ij}$ if $i \in \mathcal{S}^-$ and otherwise $q_{ij}^- = 0$. With these notations the set of equations becomes:

$$\begin{aligned} \frac{\partial p(x)}{\partial x} \mathbf{R} = p(x) \mathbf{Q} &\longrightarrow p(B) = p(0) e^{\mathbf{Q} \mathbf{R}^{-1} B} = p(0) \Phi, \\ p(0) \mathbf{R} = v \mathbf{Q}^- &\longrightarrow p(0) = v \mathbf{Q}^- \mathbf{R}^{-1}, \\ -p(B) \mathbf{R} = v \mathbf{Q}^+ &\longrightarrow \boxed{v(\mathbf{Q}^- \mathbf{R}^{-1} \Phi \mathbf{R} + \mathbf{Q}^+) = 0}, \end{aligned}$$

where the equation in the box is linear for the unknown element of vector v . The normalizing condition of this equation is

$$\ell \mathbb{I} + u \mathbb{I} + p(0) \underbrace{\int_0^B e^{\mathbf{QR}^{-1}x} dx}_{\Psi} \mathbb{I} = \boxed{v(\mathbf{I} + \mathbf{Q}^{-1}\mathbf{R}^{-1}\Psi)\mathbb{I} = 1}.$$

Relation of spectral decomposition and matrix exponent: With some rearrangement of the spectral solution we can show that the above presented 2 solutions are identical. Suppose that $|\mathcal{S}^0| = 0$ and $\mathcal{S} = \mathcal{S}^*$ the characteristic equation is $\phi(\lambda\mathbf{I} - \mathbf{QR}^{-1}) = 0$, and the spectral solution is $p(x) = \sum_{j=1}^{|\mathcal{S}|} a_j e^{\lambda_j x} \phi_j$, where λ_j and ϕ_j are the eigenvalues and the left eigenvector of matrix \mathbf{QR}^{-1} .

Introducing vector $a = \{a_j\}$ and matrix $\mathbf{B} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{|\mathcal{S}^*|} \end{pmatrix}$, the spectral solution can be rewritten as

$$\begin{aligned} p(x) &= \sum_{j=1}^{|\mathcal{S}|} a_j e^{\lambda_j x} \phi_j = a \text{Diag}\langle e^{\lambda_i x} \rangle \mathbf{B} \\ &= \underbrace{a \mathbf{B}}_{p(0)} \underbrace{\mathbf{B}^{-1} \text{Diag}\langle e^{\lambda_i x} \rangle \mathbf{B}}_{e^{\mathbf{QR}^{-1}x}}, \end{aligned}$$

which is the matrix exponential form used in [15].

Spectral decomposition of second order models The spectral decomposition based analysis of second order, infinite and finite buffer, fluid level independent Markov fluid models is presented by Karandikar and Kulkarni in [19]. In this case the differential equation has the form $p'(x) \mathbf{R} - p''(x) \mathbf{S} = p(x) \mathbf{Q}$. The general form of the solution if this equation is the same as in the first order case, $p(x) = e^{\lambda x} \phi$, but back substituting this solution we a different characteristic equation:

$$\phi(\lambda \mathbf{R} - \lambda^2 \mathbf{S} - \mathbf{Q}) = 0.$$

This characteristic equation has $2|\mathcal{S}^+| + |\mathcal{S}^*|$ solutions, with $|\mathcal{S}^+| + |\mathcal{S}^{0+}|$ negative eigenvalues, 1 zero eigenvalue, and $|\mathcal{S}^+| + |\mathcal{S}^{0-}| - 1$ positive eigenvalues. The final form of the solution is

$$p(x) = \sum_{j=1}^{2|\mathcal{S}^+| + |\mathcal{S}^*|} a_j e^{\lambda_j x} \phi_j,$$

and the a_j coefficients are set to fulfill the boundary and normalizing conditions.

A transformation of the quadratic equation to a linear one To avoid handling quadratic equations several authors recommended to transform the system into a linear one with enlarged size, e.g., [6]. In case of second order, infinite and infinite buffer, fluid level independent models with $|\mathcal{S}^0| = |\mathcal{S}^*| = 0$ and $\mathcal{S} = \mathcal{S}^+$, this transformation is based on the following representation of the differential equation

$$\begin{aligned} \frac{d}{dx} p(x) \mathbf{R} - \frac{d}{dx} p'(x) \mathbf{S} &= p(x) \mathbf{Q} , \\ \frac{d}{dx} p(x) \mathbf{I} &= p'(x) \mathbf{I} . \end{aligned}$$

This equations form a vector equation of double size

$$\frac{d}{dx} \begin{bmatrix} p(x) \\ p'(x) \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{I} \\ -\mathbf{S} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} p(x) \\ p'(x) \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} .$$

Introducing $\hat{p}(x) = \begin{bmatrix} p(x) \\ p'(x) \end{bmatrix}$, $\hat{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \mathbf{I} \\ -\mathbf{S} & \mathbf{0} \end{bmatrix}$, and $\hat{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, we obtain a first order differential equation

$$\frac{d}{dx} \hat{p}(x) \hat{\mathbf{R}} = \hat{p}(x) \hat{\mathbf{Q}} ,$$

whose solution is

$$\hat{p}(B) = \hat{p}(0) e^{\hat{\mathbf{Q}} \hat{\mathbf{R}}^{-1} B} .$$

Randomization One of the simplest stationary solution methods is based on randomization. It is applicable for first order, infinite [30] or finite [29] buffer, fluid level independent Markov fluid models. A symbolic solution of the differential equation is

$$F_i(x) = \sum_{n=0}^{\infty} e^{-\lambda t/r} \frac{(\lambda t/r)^n}{n!} b_i(n)$$

where $r = \min(r_i | r_i > 0)$ and $b_i(n)$ is defined by simple recursion such that the boundary conditions are fulfilled. Similar to other randomization based methods the numerical procedure computes convex combination of non-negative numbers, which ensures nice numerical properties.

The main limitation of these randomization based methods is that $|\mathcal{S}^{0-}|$ must be 1. Extension to the $|\mathcal{S}^{0-}| > 1$ case is not known.

Numerical solution of differential equations Unfortunately non of the above stationary analysis methods is applicable with fluid level dependent models. The only approach that is applicable for fluid level dependent cases is based on the numerical solution of the

$$\mathbf{M}'(x) \mathbf{R}(x) - \mathbf{M}''(x) \mathbf{S}(x) = \mathbf{M}(x) \mathbf{Q}(x) \quad (15)$$

differential equation with initial condition $\mathbf{M}(0) = \mathbf{I}$, as it is proposed by Gribaudo et al. in [14].

The solution is composed by the following steps:

- Numerically solve the matrix function $\mathbf{M}(x)$ based on the differential equation (15)
- calculate the unknowns $(p(0), p(B), \ell, u)$ based on the boundary conditions, the normalizing condition and

$$p(B) = p(0) \mathbf{M}(B)$$

The major limitation of this approach is that it is limited to finite buffer models.

5 Application

Fluid Models and FSPNs have been successfully used in the literature to study several interesting systems. Here we present how Fluid Stochastic Petri Nets have been used in [22] to compute the transfer time distribution of resource in a Peer-to-Peer file sharing application.

File transfer using Peer-to-Peer file sharing applications is usually divided into two steps: resource search and resource download. Depending on the file size and its popularity, either of the two phases can become the bottleneck. In this section we describe both the location and download phases of a generic Peer-to-Peer file sharing application using a fluid model. We propose a model that allows the computation of the transfer time distribution, and that it is capable of considering some advanced characteristic such as parallel downloads and on-off peer behavior. These features, although quite common in the real applications, have not been considered in previous models proposed in the literature. Model parameters reflect network, application, resource and user characteristics, and can be tuned to analyze a large number of different real implementations.

Peer-to-Peer Model The proposed fluid model for the estimation of the transfer time distribution in P2P file sharing applications will be described using the Fluid Stochastic Petri Net (FSPN) formalism [18, 17]. Table 2 reports the other notations derived from the reference.

Table 2. Model Notations

| Notation | Description | Range |
|---------------|--|---|
| \mathcal{B} | Set of bandwidths | {14.4, 28.8, 33.6, 56, 64 128, DSL, Cable, T1, T3} |
| N | Number of peers holding the resource | \mathcal{N} |
| SB | Server bandwidth | \mathcal{B} |
| CB | Tagged client bandwidth | \mathcal{B} |
| S | Resource size | \mathcal{N} |
| $K(b)$ | Max. number of concurrent peers | $\mathcal{B} \rightarrow \mathcal{N}$ |
| LT | Average number of requests of uploads | \mathcal{N} |
| W | Bandwidth dependent weight | $\mathcal{B} \rightarrow [0, 1]$ |
| $L(b)$ | LT as function of the peer bandwidth | $LT * W(b)$ |

The FSPN basic Model The basic model [12] computes the transfer time distribution of a resource of size S downloaded by a client with a bandwidth CB , from a server with bandwidth SB . It neglects both the search and the queueing phase, and download interruptions. That model is defined by means of a FSPN. The main assumption in the basic model, is that the session time of concurrent peers is described by an Hyperexponential distribution (with parameters α , μ_1 and μ_2), and that the interarrival time of concurrent downloaders is approximated by an exponential distribution (whose parameter $L(cb)$ is bandwidth dependent). The maximum number of concurrent downloads from a server is limited by a bandwidth dependent parameter $K(sb)$. Moreover, the server bandwidth is equally shared among the concurrent downloaders. For a discussion on the validity of these assumptions, please refer to [12].

Using these assumptions, the available bandwidth at the client can be computed as a function of the number of concurrent peers. In particular, if we call I_j the total number of concurrent peers in a discrete state of FSPN model, then the available bandwidth is equal to:

$$f(I_j) = \min\left(\frac{sb}{I_j + 1}, cb\right) . \quad (16)$$

The FSPN model is analyzed by solving the system of partial derivative differential equations that describes its underlying stochastic process. From the solution to these equations the probability density $\bar{\pi}(\tau, x)$ of the fluid level at a given time instant τ can be directly computed. $\bar{\pi}(\tau, x)$ corresponds to the probability density that the number of bytes downloaded at time τ is equal to x . By integrating this quantity, the probability distribution that a file of size s can be downloaded in less than t can be computed:

$$F_t(t|s) = \int_s^\infty \bar{\pi}(\tau, x) dx \Big|_{\tau=t} . \quad (17)$$

Modeling the search time, queueing time and peer unavailability

Search time is conditioned by many factors such as the popularity of the resource, protocol characteristics, the participation level of the user and the number of neighbor peers. After the searching phase the client selects peers from which get the resource. Queueing time is the time spent before a selected server serves the client request. It also depends on many factors, as the number of concurrent downloads allowed, the bandwidth of the server and the number of concurrent clients, the protocol, and the participation level of clients.

Creating a detailed model to consider all these aspects would be too complex. Instead we simplify the model by considering *the aggregate search plus queueing time perceived by a client*. That is, we suppose that we could compute the distribution $QS(\tau)$ of the time required from the start of the search to the start of the actual download of a resource. This seems to be a quite strong assumption, but we will prove, at the end of this section, that despite its simplicity, the proposed model is able to get most of the qualitative features that characterize parallel download in peer to peer applications.

Figure 11 represents the extension of the model proposed [12]. The arrival of a new concurrent download is modelled by transition `request_arrival`. The session length distribution is modelled by the sub-net composed by places `CHOICE`, `STAGE_1`, `STAGE_2`, `END_SERVICE` and transitions `choose_1`, `choose_2`, `terminate_service`, `service_1`, `service_2`. Their parameters are directly mapped to the parameters of the distributions outlined in Section 5. The maximum number of concurrent downloads is determined by the initial marking of place `AVAILABLE`, and is set according to parameter $K(sb)$. The amount of byte transferred is modelled by fluid place `TRANSFERRED` and fluid transition `transfer`. The value of parameter I_j corresponds to the sum of the marking of places `STAGE_1` and `STAGE_2`. The search and queueing phases are represented by the generic firing time transition `TON`, with distribution ϕ_{on} .

Due to the active/non-active peer dynamics the server may become unavailable and then its service is stopped. When failures occur, the client starts a new search of the same resource, and then it continues the download (likely from another peer), after experiencing a new queueing time. The failure of server is represented in the model by generic firing time transition `TOFF`, with firing time distribution ϕ_{off} .

Place `SandQ` represents the search and queueing phases, and place `TRANS` the resource transfer phase.

As reported in [12], special care should be used to compute the initial distribution of the number of concurrent peers at the server. In this case, the initial state of the places representing the concurrent peers at the server, should be determined at the time when the actual transfer starts, i.e. at the firing of transition `TON`. When transition `TON` fires, it should set the number of tokens in places `AVAILABLE`, `STAGE_1` and `STAGE_2` according to the initial distribution, determined following the technique proposed in [12]. The setting of the initial state is achieved by an appropriate set of immediate transitions, weighted according to the initial state distribution. In order to simplify Figure 11, this sub-net has been removed and has been represented by the gray arrow labeled with *Set Initial state*. Similarly, when the server experience a failure, all the places of the sub-model representing its state must be emptied. This also can be achieved by an appropriate set of immediate transition, which has been represented in Figure 11 by the gray arrow labeled with *Clear state*.

In this model, the popularity of the resource is considered when determining the rate of transition `TON`. A very popular resource will have a shorter search and queueing time, since will be available from more peers. A rare resource will instead have a very high searching and queueing time.

Considering the parallel download from multiple sources

The model that represents parallel download from multiple servers can be obtained by repeating H times the sub-models of Figure 11 representing the server and the search-queueing state, where H corresponds to the maximum number of parallel downloads. This is represented in Figure 12. Note that the H sub-models representing the H servers, share the same resource download buffer, modeled by fluid place `TRANSFERRED`. In this case, the rate at which the file

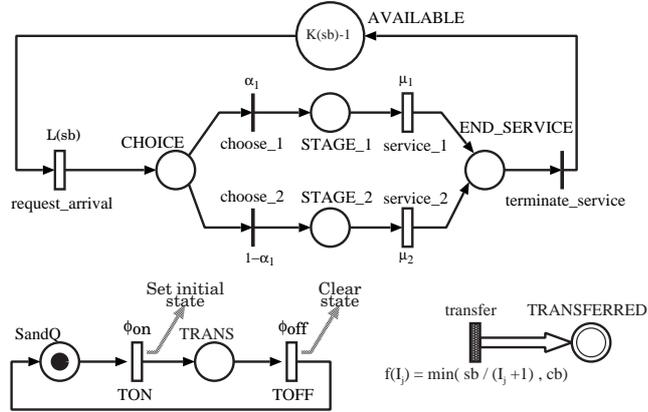


Fig. 11. FSPN model representation of an unreliable server with search and queueing phases.

is downloaded, is expressed as the minimum between the client bandwidth cb , and the sum of the download rate from each server that is active in that time instant, that is:

$$f(I_j) = \min \left(\sum_{k=1}^H \mathcal{I}(\#_{\text{TRANS}_k} = 1) \frac{sb_k}{I_{jk} + 1}, cb \right) \quad (18)$$

where $\mathcal{I}(\#_{\text{TRANS}_k})$ is an indicator function that returns 1 if the the number of tokens in place TRANS of the submodel representing the k -th server is equal to 1 (i.e. active download), zero otherwise. $I_{jk} + 1$ represents the sum of the tokens in places STAGE_1 and STAGE_2 for each tangible (discrete) state \mathbf{m}_j of the k -th server, i.e. the number of requests that interfere on that server with the tagged client service.

Despite the symmetries, the sub-models are not independent, since they are coupled by the fluid buffer TRANSFERRED . Moreover the relation that governs the rate of the growth of the fluid place (Equation 18) is non-linear, due to the presence of the $\min(\cdot)$ function. This prevents to apply a solution technique that analyzes each server separately, and combine them afterward.

| Service parameters | |
|--------------------|-------|
| μ_1 | 0.001 |
| μ_2 | 0.1 |
| α_1 | 0.6 |
| α_2 | 0.4 |
| Arrival rate | |
| LT | 0.01 |

Table 3. Model parameters used for experiments

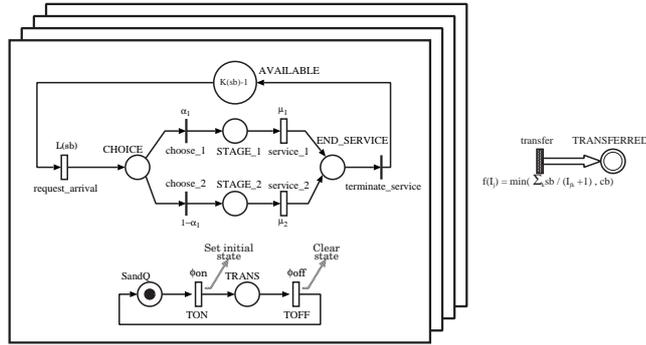


Fig. 12. FSPN model for multiple servers download.

| Resource Size | Average Bandwidth (Kbit/sec) | | |
|---------------|------------------------------|-------------------------|-------------------------|
| | Session Time 1000 sec | Session Time 20 sec. | Session Time 10 sec. |
| 512 KB | 24.32 | 10.4 | 6.56 |
| 4 MB | 59.68 | 14.08 | 7.84 |
| 10 MB | 68.48 | 6.64 | 3.12 |

Table 4. Downloading bandwidth versus session duration and resource size

Experiments The proposed models, despite their simplifying assumptions, can describe the qualitative behavior of real peer to peer systems. In both cases, we present our analysis only for the case when all peers have the same bandwidth connection (in particular, we consider 640 MB/s DSL technology). We also approximate both search and queue time distribution and the server failure distribution with exponential distributions. For this reason, in the following we will use parameters ϕ_{on} and ϕ_{off} to indicate the rate of the corresponding exponential distribution.

Extensive validation of results for our modeling technique shares the same difficulty of previous works on analytical models for P2P systems. It is a difficult task since existing measurement based studies have not focused on characterizing the duration of the transfer phase. Although it might be possible to validate our model through detailed simulations of realistic P2P file sharing applications it would have a prohibitive programming and computational cost. Nevertheless, we performed simple validations by comparing model results in selected cases where theoretical results are known or can be exactly computed. In particular, we compared model results with the ideal case where there is no competition for the server bandwidth and the transfer is only conditioned by the minimum bandwidth between server and clients. In these cases we found a perfect agreement between the model predictions and the theoretical results. It is a safety check that allow us to know that at least in the deterministic case, without

concurrent operations, model result is identical to the expected one (that is the ratio between the resource size and the minimum bandwidth among client and server ones). Moreover, results presented in Table 5 are partially supported by the measurement study presented in [24]. In particular, in [24] it is shown that the average download speed is $30KB/sec$ that in the case of a $4MB$ resource corresponds to an average transfer time of $133seconds$. This average is comparable with most of average values, referring to different number of sources, shown in Table 5.

A first intuitive result (see Table 4) shows that the transfer time increases with the increasing of unavailability rate. However, we must point out that this effect heavily depends on the resource size. We thus perform an analysis with respect to the resource size, in particular, we look at average bandwidth experienced during the file transfer as function of the failure rate. We keep the searching-queueing rate constant to 0.01: this means that client wait a mean of 100 seconds to find a new connection. We vary the failure rate in order to get server sessions of 10, 20, and 1000 seconds. The number of concurrent peers on the server, $K(sb)-1$ (minus one takes into account the tagged client), is set to 3. In this analysis we does not consider parallel downloads. FSPN model parameters used in this experiment are reported in Table 3 while results are shown in Table 4. The index we use to evaluate the performance is the average bandwidth experienced to complete the transfer of the resource. It has been computed as the ratio between the resource size and the average of the time transfer. It is interesting to note that bigger resources suffer significantly from servers failure. For instance, in the case of a 10 MBytes resource, the bandwidth falls down when the failure rate is 0.05 and 0.1 (that is session time of 10 and 20 seconds). Instead in the case of a 512 KByte resource, the penalty introduced by the failure of the server is less significative. This is due to fact that, on the average, the resource can be completely transferred before the server fails, despite shorter server session.

Most P2P file sharing applications (e.g., eDonkey, BitTorrent, etc.) allow parallel downloads. The model presented in Figure 12 represents this feature. Client peer downloads from multiple sources and gets better performance when the number of source increases as shown in Fig. 13. This experiment refers to the transfer of a 4 MByte file, with searching-queueing rate equal to 0.01 and failure rate equal to 0.001, the number of concurrent peers on each server, $K(sb) - 1$ (minus one takes into account the tagged client), is set to 3. However, improvements in performance are limited by the client download bandwidth; i.e. when the total bandwidth provided by multiple servers exceeds the maximum client download bandwidth, the speed at which the file is transferred remains constant, despite the growth in the number of sources. This is shown in table 5, where the mean and quantiles of the transfer time distribution related to a 4 MBytes resource are reported as function of the number of sources. In this case parameters are: searching-queueing rate equal to 0.01 and failure rate equal to 0.001, the number of concurrent peers on each server, $K(sb) - 1$, is set to 1. We can note that the improvement in transfer performance become less significative as the

number of sources increases (since they saturate the client downloading bandwidth). When sources become 9 the time required to transfer the file remains constant. This insight may provide suggestions for the application design. E.g., let suppose that the application protocol is able to monitor the client bandwidth status. If it detects that the client is the bottleneck, then it can avoid to add new (parallel) sources. Their contribute, that should not be exploited in order to improve tagged transfer performance, could be exploited to improve the system service capacity for other peers.

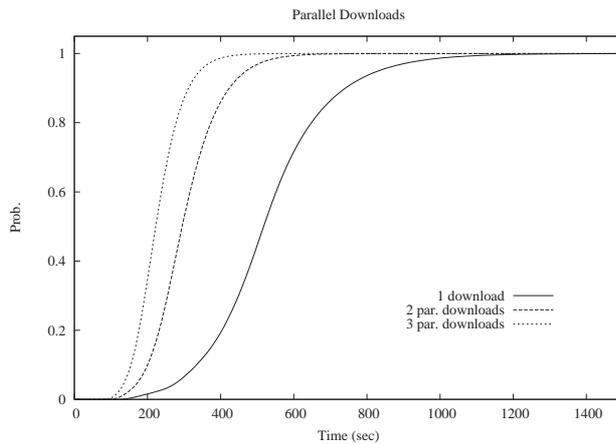


Fig. 13. Improvement provided by parallel downloads

| Number of Sources | Transfer Time (sec.) | | |
|-------------------|----------------------|---------------------------|---------------------------|
| | Mean | 50 th quantile | 90 th quantile |
| 3 | 178 | 170 | 240 |
| 4 | 148 | 140 | 200 |
| 5 | 131 | 130 | 170 |
| 6 | 119 | 120 | 160 |
| 7 | 110 | 110 | 148 |
| 8 | 104 | 100 | 130 |
| 9 | 104 | 100 | 130 |

Table 5. Transfer time as function of the number of sources

It is interesting to see how the benefit derived from the use of parallel download depends on the size of the resource. Consider the case in which downloading session does not suffer from the servers failures (i.e. the failure rate is very low). We set searching-queueing rate much bigger than the failure one, respectively 0.1 and 0.001. The number of concurrent peers on each server, $K(sb) - 1$, is set to 1. The study has been done for 512 KBytes, 4 and 10 MBytes resource sizes

and for a number of parallel downloads that grows from 1 up to 6, as shown In Fig. 14. Small resources take less benefits from parallel downloading, since the downloading time is shorter than the time required by the searching and queuing phase to start a parallel download from another source. For bigger resources instead, the downloading time is reduced significantly with the increases in the number possible download source. These improvements are however limited by the client bandwidth, as shown in the previous example. This can be seen for the 4 and the 10 MBytes cases, when the number of sources increase from 5 to 6.

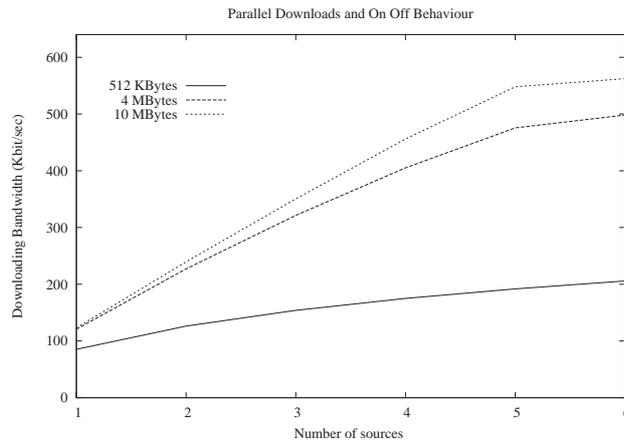


Fig. 14. Benefits of parallel downloading for different resource sizes

| Distribution | Transfer Time (sec.) | | |
|---------------------|----------------------|---------------------------|---------------------------|
| | Mean | 90 th quantile | 95 th quantile |
| Hypo-exponential | 448 | 580 | 630 |
| Exponential | 376.31 | 575 | 660 |
| Hyper-exponential 1 | 306.34 | 495 | 605 |
| Hyper-exponential 2 | 297.35 | 440 | 510 |

Table 6. Modeling searching and queuing phases with different distributions.

In order to describe different system scenario we also approximate the searching and queuing rate with different distributions. All previous results refer to the exponential case. In addition we model the searching and queuing phases with Hyper-exponential and Hypo-exponential distributions. Results are reported in Table 6; Figure 15 refers to the transfer of a 4MB file with 3 parallel downloads and a session mean time of 15 minutes. In all cases the mean time spent in the searching/queuing phase is 5 minutes. In the case "Hyper-exponential

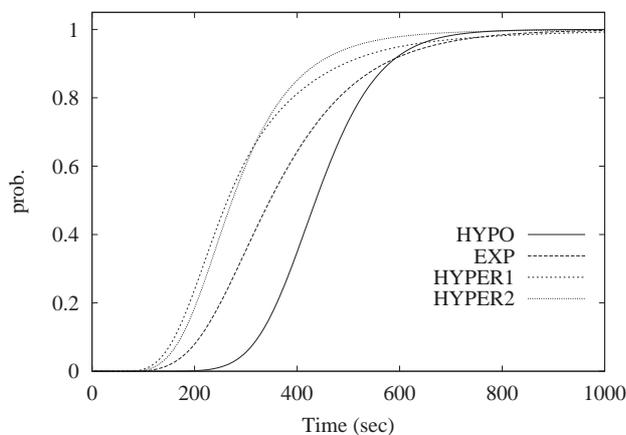


Fig. 15. Transfer time distribution with different searching-queueing rate distributions

1” the mean time spent by the client is 10 minutes with a probability of 44% and 1 minute with a probability of 56%. In this case faster searching/queueing phases are favorite, indeed transfer time is shorter than in the case ”Exponential”. Shorter searching/queueing phases are even more favorite in the case ”Hyper-exponential 2”: 3.45 minutes with probability 80%, and 10 minutes with probability 20%. This setting results in faster transfers, as reported in Table 6. The choice of the Hyper-exponential distribution can be useful for describing different scenario where shorter searching/queueing phases model popular resource transfers and longer ones model rare resource transfers. The Hypo-exponential distribution can be used to model rates when the approximation should be more deterministic. In this case, the Hypo-exponential case corresponds to a 5 stages Erlang distribution. Even if the goal of this work is not to compare different approximations, it shows that the proposed model can be considered a flexible tool for evaluating P2P applications performance.

6 Conclusions

Stochastic models with continuous variables (Reward models, Fluid models and FSPNs) often allows proper modeling of real systems. Their analysis is a more complex than the ones with only discrete variables, but feasible for a wide class of models. The analytical description of Markov fluid models and a set of solution techniques have been introduced. The presented application examples demonstrate the potential use of fluid models in performance analysis.

References

1. M. Agapie and K. Sohraby. Algorithmic solution to second order fluid flow. In *Proc. of IEEE Infocom*, Anchorage, Alaska, Usa, Apr 2001.

2. Soohan Ahn and V. Ramaswami. Matrix-geometric algorithms for stochastic fluid flows. In *SMCtools '06: Proceeding from the 2006 workshop on Tools for solving structured Markov chains*, page 11, New York, NY, USA, 2006. ACM Press.
3. M. Ajmone Marsan, G. Balbo, A. Bobbio, G. Chiola, G. Conte, and A. Cumani. The effect of execution policies on the semantics and analysis of stochastic Petri nets. *IEEE Transactions on Software Engineering*, 15(7):832–846, July 1989.
4. M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, 1995.
5. H. Alla and R. David. Continuous and Hybrid Petri Nets. *Journal of Systems Circuits and Computers*, 8(1), Feb 1998.
6. Eu-Jin Ang and Javier Barria. The markov modulated regulated brownian motion: A second-order fluid flow model of a finite buffer. *Queueing Systems*, 35:263–287, 2000.
7. D. Anick, D. Mitra, and M. M. Sondhi. Stochastic Theory of a Data-Handling System. *Bell Sys. Thech. J*, 61(8):1871–1894, Oct 1982.
8. D.-Y. Chen, Y. Hong, and K. S. Trivedi. Second order stochastic fluid flow models with fluid dependent flow rates. *Performance Evaluation*, 49(1-4):341–358, 2002.
9. G. Ciardo, D. M. Nicol, and K. S. Trivedi. Discrete-event Simulation of Fluid Stochastic Petri Nets. *IEEE Transactions on Software Engineering*, 2(25):207–217, 1999.
10. E. de Souza e Silva and R. Gail. An algorithm to calculate transient distributions of cumulative rate and impulse based reward. *Commun. in Statist. – Stochastic Models*, 14(3):509–536, 1998.
11. A. I. Elwalid and D. Mitra. Statistical Multiplexing with Loss Priorities in Rate-Based Congestion Control of High-Speed Networks. *IEEE Transaction on Communications*, 42(11):2989–3002, November 1994.
12. R. Gaeta, M. Gribaudo, D. Manini, and M. Sereno. Analysis of resource transfers in peer-to-peer file sharing applications using fluid models. *Perform. Eval.*, 63(3):149–174, 2006.
13. R. German. *Performance Analysis of Communication Systems: Modeling with Non-Markovian Stochastic Petri Nets*. John Wiley & Sons, 2000.
14. R. German, M. Gribaudo, G. Horváth, and M. Telek. Stationary analysis of FSPNs with mutually dependent discrete and continuous parts. In *International Conference on Petri Net Performance Models – PNPM 2003*, pages 30–39, Urbana, IL, USA, Sept 2003. IEEE CS Press.
15. M. Gribaudo and R. German. Numerical solution of bounded fluid models using matrix exponentiation. In *Proc. 11th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB)*, Aachen, Germany, Sep 2001. VDE Verlag.
16. M. Gribaudo, M. Sereno, and A. Bobbio. Fluid Stochastic Petri Nets: An Extended Formalism to Include non-Markovian Models. In *Proc. 8th Intern. Workshop on Petri Nets and Performance Models*, Zaragoza, Spain, Sep 1999. IEEE-CS Press.
17. M. Gribaudo, M. Sereno, A. Bobbio, and A. Horvath. Fluid Stochastic Petri Nets augmented with Flush-out arcs: Modelling and Analysis. *Discrete Event Dynamic Systems*, 11(1 & 2), 2001.
18. G. Horton, V. G. Kulkarni, D. M. Nicol, and K. S. Trivedi. Fluid stochastic Petri Nets: Theory, Application, and Solution Techniques. *European Journal of Operations Research*, 105(1):184–201, Feb 1998.
19. R. L. Karandikar and V.G. Kulkarni. Second-order fluid flow models: reflected brownian motion in a random environment. *Operations Research*, 43:77–88, 1995.

20. V. G. Kulkarni. Fluid models for single buffer systems. In J. H. Dshalalow, editor, *Models and Applications in Science and Engineering*, Frontiers in Queueing, pages 321–338. CRC Press, 1997.
21. V.G. Kulkarni, V.F. Nicola, and K. Trivedi. Effects of checkpointing and queueing on program performance. *Stochastic models*, 4(6):615–648, 1990.
22. D. Manini and M. Gribaudo. Modelling search, availability, and parallel download in p2p file sharing applications with fluid model. *Proceedings of the 14th International Conference on Advanced Computing and Communication ADCOM*, 2006.
23. V.F. Nicola, R. Martini, and P.F. Chimento. The completion time of a job in a failure environment and partial loss of work. In *2nd Int. Conf. on Mathematical Methods in Reliability (MMR'2000)*, pages 813–816, Bordeaux, France, July 2000.
24. J.A. Pouwelse, P. Garbacki, D.H.G. Epema, and H.J. Sips. A measurement study of the bittorrent peer-to-peer file-sharing system. *Proc. 19th IEEE Annual Computer Communications Workshop*, 2004.
25. M. A. Qureshi and W. H. Sanders. Reward model solution methods with impulse and rate rewards: An algorithm and numerical results. *Performance Evaluation*, pages 413–436, 1994.
26. A. Reibman, R. Smith, and K.S. Trivedi. Markov and Markov reward model transient analysis: an overview of numerical approaches. *European Journal of Operational Research*, 40:257–267, 1989.
27. Qiang Ren and Hisashi Kobayashi. Transient solutions for the buffer behavior in statistical multiplexing. *Perform. Eval.*, 23(1):65–87, 1995.
28. B. Sericola. Transient analysis of stochastic fluid models. *Perform. Eval.*, 32(4):245–263, 1998.
29. Bruno Sericola. A finite buffer fluid queue driven by a markovian queue. *Queueing Syst. Theory Appl.*, 38(2):213–220, 2001.
30. Bruno Sericola and Bruno Tuffin. A fluid queue driven by a markovian queue. *Queueing Syst. Theory Appl.*, 31(3-4):253–264, 1999.