22

Diffusion models of leaky bucket and partial buffer sharing policy: transient analysis

T. Czachórski
IITiS PAN, Polish Academy of Sciences
44-100 Gliwice, ul. Bałtycka 5, Poland;
PRiSM, Université de Versailles-St-Quentin
45 Av. des Etats-Unis, 78000 Versailles, France
phone: +4832 317026, fax: +4832 317026
e-mail: tadek@atos.iitis.gliwice.pl

F. Pekergin
LIPN, Université de Paris-Nord
Avenue J. B. Clément, 93430 Villetaneuse, France
phone: + 331 49.40.36.84, fax: +331 48.26.07.12
e-mail: pekergin@ura1507.univ-paris13.fr

Abstract

Leaky bucket and partial buffer sharing policy have already become classic examples of preventive and reactive traffic control functions implemented in ATM networks. They help to respect the negotiated connection parameters, to avoid the congestion and therefore to ensure the guaranteed quality of service. We revisit their performance models with the use of diffusion approximation adopting our previously developed method of transient state analysis and extending it to the case of state-dependent input. This kind of approach gives us an inside look upon the transient behaviour of the traffic. The dynamics of the traffic is displayed and the influence of both mechanisms on the traffic characteristics appears as a function of time. General cell interarrival times and the burstiness of the traffic are represented in a natural way in these models. The diffusion method is a second-order approximation and thus has certain superiority upon the fluid approximation. Both models can be easily implemented in a general queueing network model. Hence, the impact of both mechanisms on the performance of the whole network may be studied. The models may be applied also in cases of very small losses which are difficult to study by simulation.

Keywords

Performance evaluation, ATM networks, diffusion approximation.

1 INTRODUCTION

The presence of Variable Bit Rate sources and the use of statistical multiplexing in ATM networks create the need of traffic control assuring a reasonable compromise between bandwidth efficiency and quality of service. Leaky bucket, introduced almost ten years ago (Akhtar, 1987) and partial buffer sharing policy, see e.g. (Kröner, 1991) have already become classic examples of preventive and reactive traffic control functions implemented in ATM networks.

In the leaky bucket scheme, the cells, before entering the network, must obtain a token. Tokens are generated at constant rate and stocked in a buffer of finite capacity. If there is a token available, an arriving cell consumes it and leaves the bucket. If not, it waits for the token in the cell buffer. The capacity of this buffer is also limited. Tokens and cells arriving to full buffers are lost. The analysis of the leaky bucket performance includes discrete-time Markovian models (Holtzinger, 1992) and fluid approximation (Elwalid, 1991).

Partial buffer sharing policy is a well known space priority mechanism aiming to resolve congestion problems arising at a network node. Cells waiting for transmission towards a specified direction are queued in a buffer. The quality of service demanded by different types of traffic is not the same, hence cell priority may depend on the type of traffic. The cell priority may be also set by an interface (e.g. jumping or sliding window) at a network entrance: there are regular cells admitted to the network on the contract basis and additional ones which may only conditionally enter the network and are discarded when congestion arises. In general, the cells belong to two classes, the first having higher and the second having lower priority. If the number of cells in the buffer is below a defined level, the partial buffer sharing policy allows the arriving cells of both classes to enter the buffer, otherwise only priority ones are stocked and arriving lower class cells are lost. The models of this mechanism are usually based on the discrete time Markov processes (Hébuterne, 1989), (Kröner, 1991), (Meyer, 1993) or simulation.

We revisit the models of the above mechanisms with the use of diffusion approximation, mainly to test the utility of this method as a tool to analyse transient states resulting from traffic control.

Diffusion approximation replaces process $N(t)$, representing the number of customers in a service station by a diffusion process $X(t)$ whose pdf $f(x,t;x_0)$ is defined by the diffusion equation

$$\frac{\partial f(x,t;x_0)}{\partial t} = \frac{\alpha}{2}\frac{\partial^2 f(x,t;x_0)}{\partial x^2} - \beta\frac{\partial f(x,t;x_0)}{\partial x}\,. \tag{1}$$

If we solve this equation with appropriate boundary conditions and diffusion parameters α, β, we obtain an approximation of the queue distribution: $p(n,t;n_0) \approx f(n,t;n_0)$.

Let $A(x)$, $B(x)$ denote the interarrival and service time distributions at a service station. The distributions are general, their means and variances are:

$E[A] = 1/\lambda$, $E[B] = 1/\mu$, $\mathrm{Var}[A] = \sigma_A^2$, $\mathrm{Var}[B] = \sigma_B^2$. Squared coefficients of variation are denoted as $C_A^2 = \sigma_A^2\lambda^2$, $C_B^2 = \sigma_B^2\mu^2$. In the case of $G/G/1$ and $G/G/1/N$ stations the parameters α, β are chosen as $\beta = \lambda - \mu$, $\alpha = \sigma_A^2\lambda^3 + \sigma_B^2\mu^3 = C_A^2\lambda + C_B^2\mu$, see (Newell, 1971).

The choice of boundary conditions is based on the approach proposed by Gelenbe for a finite queue of capacity N in (Gelenbe, 1975) and refined in (Gelenbe, 1995). The diffusion process $X(t)$ is limited by two boundaries situated at $x = 0$ and $x = N$. When the process comes to $x = 0$, it remains there for a time which corresponds to the idle time of the system and then jumps to $x = 1$; when it comes to $x = N$ it stays there for a time during which the queue is full and then jumps to $x = N - 1$. This finite queue model and the steady-state solution of the diffusion equation with the described boundaries were given in (Gelenbe, 1975) and broadly used afterwords. Here, we use transient solution of this model and develop the approach of (Czachórski, 1993, 1994).

2 THE LEAKY BUCKET MODEL

In the leaky bucket model the diffusion process $X(t)$ is defined on the interval $x \in [0, N = B + M]$ where B is the capacity of cell buffer and M is the capacity of token buffer. The current value of the process is defined as $x = b - m + M$, b and m being the current contents of the buffers, Figure 1.

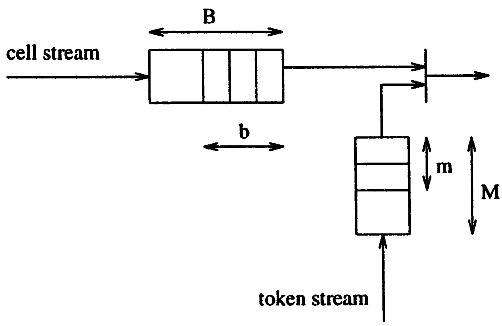


Figure 1 Leaky bucket scheme

Let us suppose that the cell interarrival time distribution has the mean $1/\lambda_c$ and squared coefficient of variation $C_{A_c}^2$. The tokens are generated with constant rate λ_t, hence $C_{A_t}^2 = 0$. Arrival of a cell increases the value of the process and arrival of a token decreases it, therefore we choose the parameters of the diffusion process as:

$$\beta = \lambda_c - \lambda_t\ , \qquad \alpha = \lambda_c C_{A_c}^2\ .$$

The process evolves between two barriers placed at $x = 0$ and at $x = M + B$;

$x = 0$ represents the state where the whole token buffer is occupied and arriving tokens are lost; $x = M + B$ represents the state where the token buffer is empty and the cell buffer is full: arriving cells are lost.

The sojourn time at $x = M + B$ corresponds to the residual token interarrival time, i.e. the time between the moment when the cell buffer becomes full and the moment of the next token arrival. We use here the density of holding time at the upper barrier of $M/D/1/N$ diffusion model as obtained in (Gelenbe, 1995).

If the cell stream is Poisson, the pdf $l_0(x)$ of the sojourn time at $x = 0$ is defined by the density of cell interarrival time; otherwise we take this density as an approximation of $l_0(x)$. Note that the sojourn times in boundaries are defined here by the densities $l_0(t)$, $l_N(t)$ and **are not restricted to exponential distributions**.

The values $x > M$ of the process correspond to states where cells are waiting for tokens, the value $x - M$ determines in this case the number of cells in the buffer; $x < M$ means that there are tokens waiting for cells and the value $M - x$ corresponds to the number of tokens in the buffer. Probability of b cells in the buffer at time t is defined by $f(M + b, t)$; probability of the empty cell buffer is given by $p_t(t) = p_0(t) + \int_0^M f(x,t)dx$. Probability of m tokens in the buffer is given by $f(M - m, t)$ and probability of empty token buffer is determined by $\int_M^{M+B} f(x,t)dx + p_N(t)$ where $p_0(t) = Pr[X(t) = 0]$, $p_N(t) = Pr[X(t) = N]$.

The service time is constant, hence the density function of the cell waiting time for tokens (response time of leaky bucket) may be estimated as $r(x,t) = \lambda_t f(\lambda_t x + M, t)$.

To obtain the transient solution $f(x,t)$ we follow the approach which we proposed previously in (Czachórski, 1993, 1994). Its main idea is to express $f(x,t;x_0)$ with the use of a superposition

$$\begin{aligned} f(x,t;x_0) \quad = \quad & \phi(x,t;x_0) + \int_0^t g_1(\tau)\phi(x,t-\tau;1)d\tau + \\ & \int_0^t g_{N-1}(\tau)\phi(x,t-\tau;N-1)d\tau \end{aligned} \tag{2}$$

of densities $\phi(x,t;x_0)$ of diffusion process with another kind of boundary conditions: absorbing barriers placed at $x = 0$ and $x = N$; the process bounded by these barriers is finished when it comes to one of them. The densities $\phi(x,t;x_0)$ are easier to obtain than $f(x,t;x_0)$ and their form is known, see e.g. (Cox, 1965).

In practice, we obtain the Laplace transform of $f(x,t;x_0)$ and invert it numerically. Hence, we obtain transient $f(x,t;\psi)$ and steady-state $f(x)$ distributions of the diffusion process for $0 \le x \le M + B$. This gives us the distribution of the number of tokens and cells in the leaky bucket, the response time distribution, the loss probabilities, the properties of the output stream, etc. The capacities of cell and token buffers may be null, so we are able to consider a number of leaky bucket variants.

The output process of the leaky bucket is the same as the cell input process provided, with probability $p_t(t)$, that there are tokens available and it is the same as token input process with probability $1 - p_T(t)$ that tokens are not available; at the time moment t the pdf $d(x)$ of interdeparture times in the output stream is

$$d(x,t) = p_t(t)a(x,t) + [1 - p_t(x,t)]\delta(x - \frac{1}{\lambda_t}), \qquad (3)$$

where $a(x,t)$ is the time-dependent pdf of cell interarrival times distribution. Eq. (3) gives us the mean value and squared coefficient of interdeparture times distribution, i.e. whole information needed to incorporate one or multiple leaky-bucket stations (for example a cascade of leaky-buckets with different parameters) in the diffusion queueing network model of $G/G/1$ or $G/G/1/N$ stations.

Numerical example. At $t = 0$ the cell buffer is empty and the token buffer contains $M(0)$ tokens. The tokens are generated regularly each time-unit. The cell arrival stream is Poisson; the mean interarrival time is 0.5 time-unit for $0 \leq t < 100$ and 1.5 time units for $t \geq 100$, i.e. there is a traffic wave exceeding the accorded level during the first 100 units and then the traffic goes down below this level.

The buffer capacities are $B = M = 100$. Figure 2 displays the diffusion and simulation results concerning the output stream of leaky bucket for the initial number of tokens $M(0) = 0$, 50 and 100. The output dynamics given by simulation and by difusion model are very similar. Simulation results are obtained as a mean of 100 000 independent runs. If there is no tokens at the beginning, the cell stream is immediately cut to the level of token intensity (one cell per time unit), the excess of cells is stocked in the cell buffer and transmitted later, when $t > 100$ and input rate becomes smaller. If there are tokens in the token buffer, a part (for $M(0) = 50$) or almost totality (for $M(0) = 100$) of the traffic wave may enter into the network.

Figure 3 presents the comparison of mean number of cells in the cell buffer as a function of time, for different initial content of the token buffer $M(0) = 0$, 50 and 100, obtained by diffusion and simulation models. In Figure 5 the distributions of cell buffer contents obtained by simulation and by diffusion are presented for $t = 100$, i.e. at the end of high traffic period, when the congestion is the biggest. We see that although the mean queue length is below the buffer capacity, the probability that the buffer is full is important (≈ 0.4). Note that we could not obtain this result with the use of fluid approximation even if the mean number of cells in the buffer predicted by diffusion and fluid approximations were similar. Figures 4, 6, 7 present distribution of the number of waiting cells in different moments of the considered period. Diffusion results are compared with simulation.

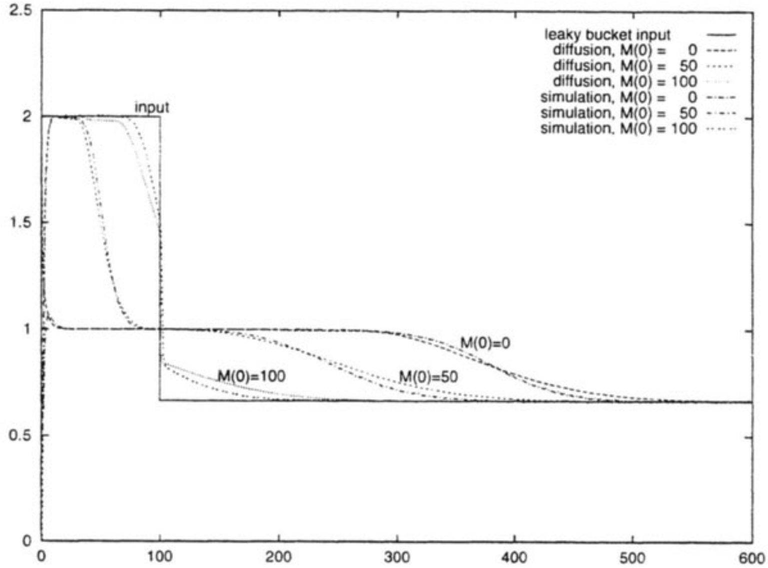


Figure 2 The input and output of leaky bucket as a function of time — the stream intensities for the initial number of tokens $M(0) = 0$, 50 and 100; diffusion and simulation results

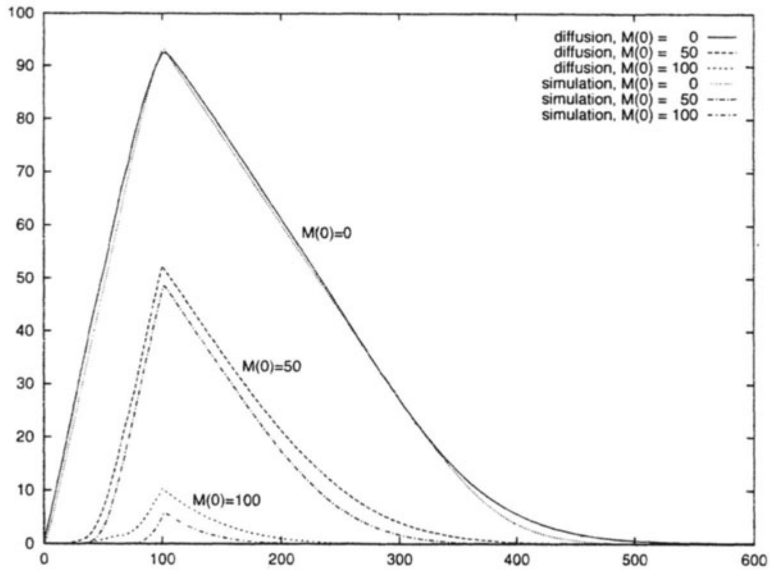


Figure 3 Mean number of cells in the cell buffer as a function of time, $M(0) =$ 0, 50 and 100; diffusion and simulation results

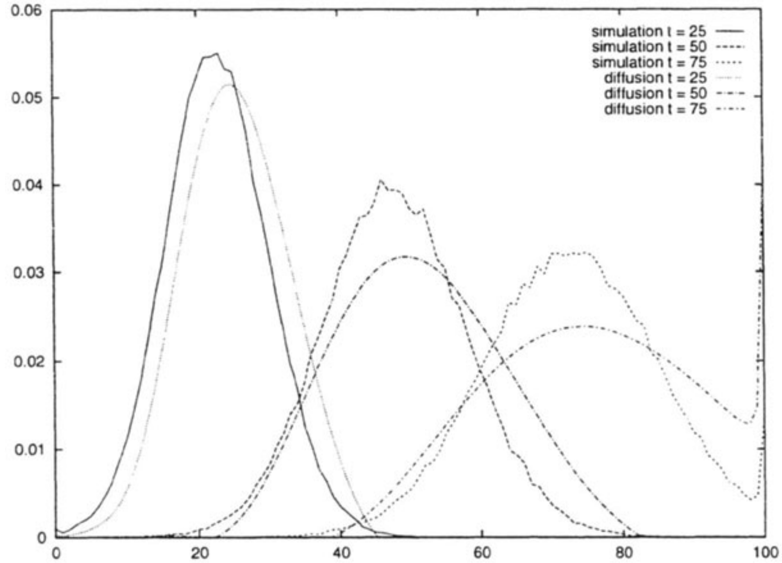


Figure 4 Density of the number of cells during high source activity period, $t = 25, 50, 75$, $M(0) = 0$; diffusion and simulation results

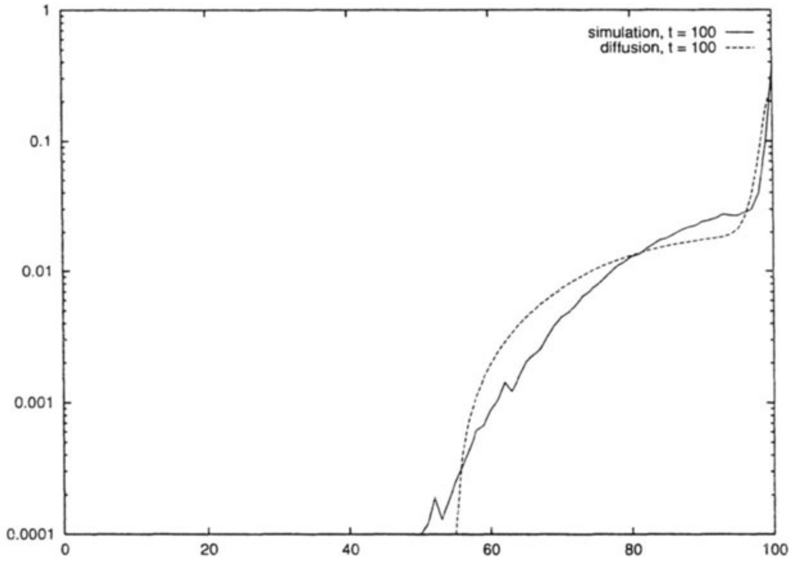


Figure 5 Density of the number of cells at the end of high source activity period, $t = 100$, $M(0) = 0$; diffusion and simulation results

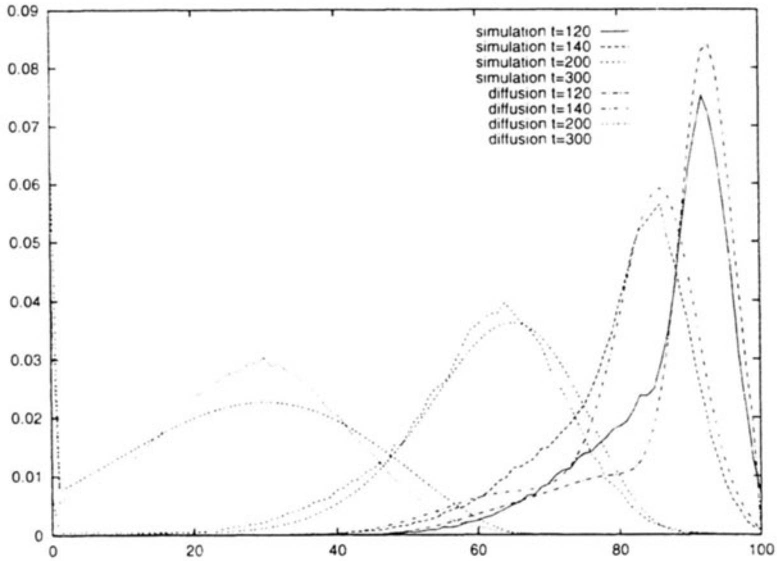


Figure 6 Density of the number of cells at the beginning of low source activity period, $t = 120, 140, 200, 300$, $M(0) = 0$; diffusion and simulation results

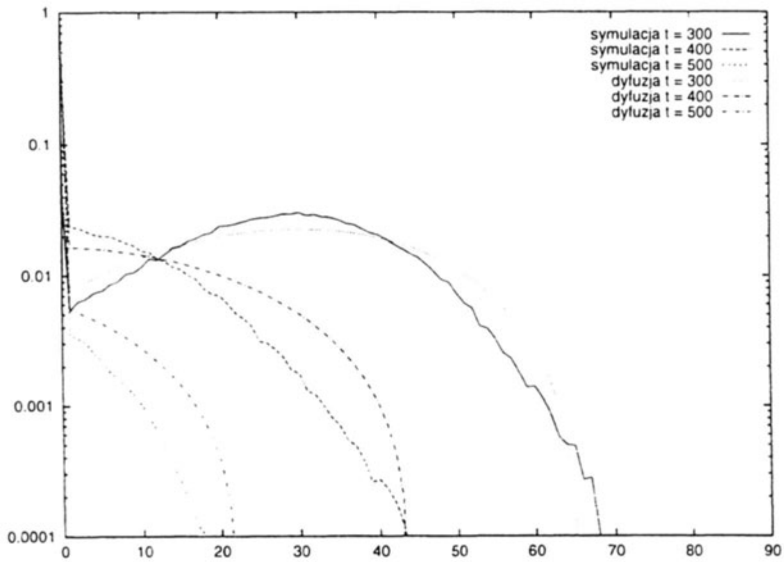


Figure 7 Density of the number of cells at the end of low source activity period, $t = 300, 400, 500$, $M(0) = 0$; diffusion and simulation results

3 PARTIAL BUFFER SHARING POLICY

In a node with partial buffer sharing policy the diffusion process represents the content of the cell buffer. The process is determined on the interval $x \in [0, N]$ where N is the buffer capacity. When the number of cells is equal or greater than the threshold N_1 $(N_1 < N)$, only priority cells are admitted and ordinary ones are lost. Diffusion process represents the number cells of both classes, hence its parameters depend on their input and service parameters which are different for $x \le N_1$ and $x > N_1$:

$$\beta(x) = \begin{cases} \beta_1 = \lambda^{(1)} + \lambda^{(2)} - \mu & \text{for} \quad 0 < x \le N_1 , \\ \beta_2 = \lambda^{(1)} - \mu & \text{for} \quad N_1 < x < N \end{cases} \tag{4}$$

and

$$\alpha(x) = \begin{cases} \alpha_1 = \lambda^{(1)} C_A^{(1)^2} + \lambda^{(2)} C_A^{(2)^2} + \mu C_B^2 & \text{for} \quad 0 < x \le N_1 , \\ \alpha_2 = \lambda^{(1)} C_A^{(1)^2} + \mu C_B^2 & \text{for} \quad N_1 < x < N. \end{cases} \tag{5}$$

We assume constant service time, hence $C_B^2 = 0$.

Steady state solution. Let $f_1(x)$ and $f_2(x)$ denote the pdf function of the diffusion process in intervals $x \in (0, N_1]$ and $x \in [N_1, N)$. We suppose that

- $\lim_{x \to 0} \; f_1(x,t;x_0) = \lim_{x \to N} \; f_2(x,t;x_0) = 0$,
- $f_1(x)$ and $f_2(x)$ functions have the same value at the point N_1: $f_1(N_1) = f_2(N_1)$,
- there is no probability mass flow within the interval $x \in (1, N-1)$: $\frac{\alpha_n}{2} \frac{d f_n(x)}{dx} - \beta_n f_n(x) = 0$ for $x \in (1, N_1), n = 1$ and $x \in (N_1, N-1), n = 2$

and we obtain the solution of diffusion equations:

$$f_1(x) = \begin{cases} \dfrac{[\lambda^{(1)} + \lambda^{(2)}]p_0}{-\beta_1} (1 - e^{z_1 x}) & \text{for} \quad 0 < x \le 1 , \\ \dfrac{[\lambda^{(1)} + \lambda^{(2)}]p_0}{-\beta_1} (1 - e^{z_1}) e^{z_1 (x-1)} & \text{for} \quad 1 \le x \le N_1 , \end{cases}$$

$$f_2(x) = \begin{cases} f_1(N_1) e^{z_2 (x - N_1)} & \text{for} \quad N_1 \le x \le N - 1 , \\ \dfrac{\mu p_N}{\beta_2} \left[1 - e^{z_2 (x-N)}\right] & \text{for} \quad N - 1 \le x < N , \end{cases} \tag{6}$$

where $z_n = \dfrac{2\beta_n}{\alpha_n}$, $n = 1, 2$. Probabilities p_0, p_N are obtained with the use of normalization condition. The loss ratio $L^{(1)}$ is expressed by the probability p_N,

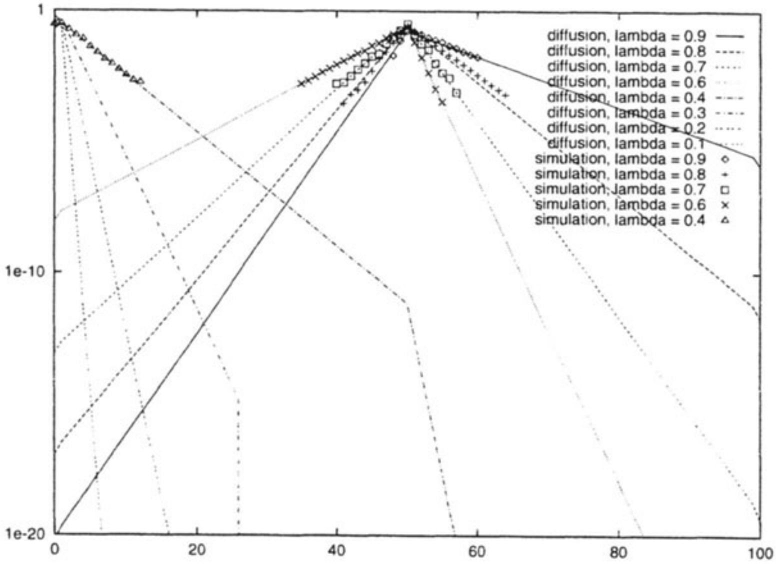


Figure 8 Steady state distribution of the number of cells for traffic densities $\lambda^{(1)} = \lambda^{(2)} = \lambda = 0.1 - 0.9$; diffusion and simulation results

the loss ratio $L^{(2)}$ is determined by the probability $P[x > N_1] = \int_{N_1}^{N} f_2(x)dx + p_N$.

Numerical example. Figure 8 presents the steady sate distribution given by Eqs. (6) of the number of cells present in a station. The buffer length is $N = 100$, the threshold value is $N_1 = 50$. Some of the values are compared with simulation histograms which we were able to obtain only for relatively large values of probabilities. In Figure 9 the probabilities that the buffer is full and that the threshold is attained are compared with class 2 loss probabilities obtained by simulation. Once again, only relatively big values of losses could be obtained by simulation.

Transient solution. The transient solution which we obtain below for a diffusion process with coefficients $\alpha(x)$ $\beta(x)$ depending on its value is, as far as we know, a novelty on theoretical plan of diffusion models. It makes use of the balance equations for probability flows crossing the barrier situated at the boundary between the intervals with different diffusion coefficients, i.e. at $x = N_1$. Let us consider two separate diffusion processes $X_1(t)$, $X_2(t)$:

$X_1(t)$ is defined on the interval $x \in (0, N_1)$. At $x = 0$ there is a barrier with sojourn times defined by a pdf $l_0(t)$ and instantaneous returns to the point $x = 1$. At $x = N_1$ an absorbing barrier is placed. Denote by $\gamma_{N1}^{L}(t)$ the pdf that the process enters the absorbing barrier at $x = N_1$. The process is reinitiated at $x = N_1 - \varepsilon$ with a density $g_{N_1-\varepsilon}(t)$.

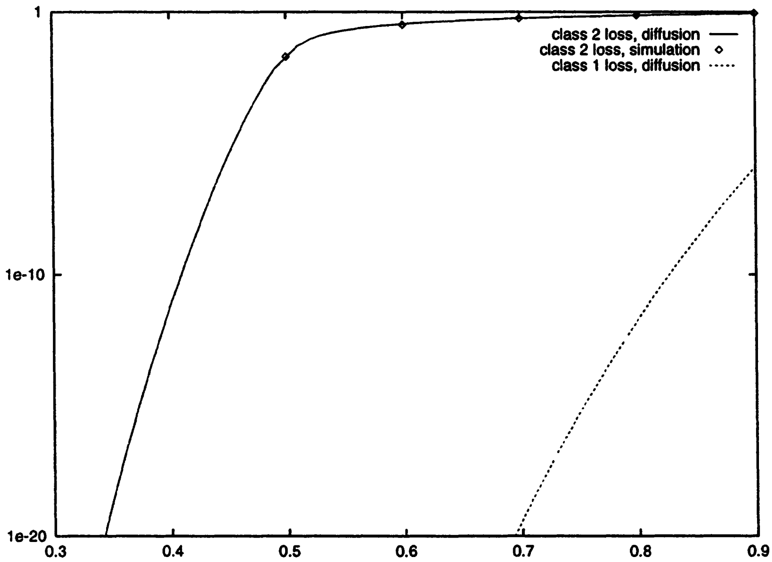


Figure 9 Probability that the threshold is attained and that the buffer is full given by diffusion model; probability of class 2 losses as a function of $\lambda^{(1)} = \lambda^{(2)} = \lambda = 0.3 - 0.9$, diffusion and simulation results

$X_2(t)$ is defined on the interval $x \in (N_1, N)$. It is limited by an absorbing barrier at $x = N_1$ and by a barrier with instantaneous returns at $x = N$. The sojourn time at this barrier is defined by a pdf $l_N(t)$ and the returns are performed to $x = N - 1$. The process is reinitiated at $x = N_1 + \varepsilon$ with a density $g_{N_1+\varepsilon}(t)$. Denote by $\gamma^R_{N1}(t)$ the pdf that the process $X_2(t)$ enters the absorbing barrier at $x = N_1$.
The interaction between two processes is given by equations

$$g_{N_1+\varepsilon}(t) = \gamma^L_{N_1}(t) \quad \text{and} \quad g_{N_1-\varepsilon}(t) = \gamma^R_{N_1}(t),$$

i.e. the probability density that one process enters to its absorbing barrier is equal to the density of reinitialization of the other process in the vicinity of the barrier.

As previously for the leaky bucket model, the pdfs $f_1(x,t;\psi_1)$, $f_2(x,t;\psi_2)$ of both processes are expressed by the pdfs $\phi_1(x,t;x_0)$, $\phi_2(x,t;x_0)$ of processes with two absorbing barriers: at $x = 0$, $x = N_1$ for the first process and at $x = N_1$, $x = N$ for the second:

$$\begin{aligned} f_1(x,t;\psi_1) &= \phi_1(x,t;\psi_1) + \int_0^t g_1(\tau)\phi_1(x,t-\tau;1)d\tau + \\ &+ \int_0^t g_{N_1-\varepsilon}(\tau)\phi_1(x,t-\tau;N_1-\varepsilon)d\tau\,, \end{aligned} \tag{7}$$

$$
\begin{aligned}
f_2(x,t;\psi_2) &= \phi_2(x,t;\psi_2) + \int_0^t g_{N-1}(\tau)\phi_2(x,t-\tau;N-1)d\tau + \\
&\quad + \int_0^t g_{N_1+\varepsilon}(\tau)\phi_2(x,t-\tau;N_1+\varepsilon)d\tau \,. \qquad (8)
\end{aligned}
$$

In order to use this solution, we need to determine the densities $g_1(t)$, $g_{N_1-\varepsilon}(t)$, $g_{N_1+\varepsilon}(t)$, $g_{N-1}(\tau)$. The equations of probability flows are

$$
\begin{aligned}
\gamma_0(t) &= p_0(0)\delta(t) + \gamma_{\psi_1,0}(t) + \int_0^t g_1(\tau)\gamma_{1,0}(t-\tau)d\tau + \\
&\quad \int_0^t g_{N_1-\varepsilon}(\tau)\gamma_{N_1-\varepsilon,0}(t-\tau)d\tau \,, \\
\gamma_{N1}^L(t) &= \gamma_{\psi_1,N_1}(t) + \int_0^t g_1(\tau)\gamma_{1,N_1}(t-\tau)d\tau + \\
&\quad \int_0^t g_{N_1-\varepsilon}(\tau)\gamma_{N_1-\varepsilon,N_1}(t-\tau)d\tau \,, \\
\gamma_N(t) &= p_N(0)\delta(t) + \gamma_{\psi_2,N}(t) + \int_0^t g_{N_1+\varepsilon}(\tau)\gamma_{N_1+\varepsilon,N}(t-\tau)d\tau + \\
&\quad \int_0^t g_{N-1}(\tau)\gamma_{N-1,N}(t-\tau)d\tau \,, \\
\gamma_{N1}^R(t) &= \gamma_{\psi_2,N1}(t) + \int_0^t g_{N_1+\varepsilon}(\tau)\gamma_{N_1+\varepsilon,N_1}(t-\tau)d\tau + \\
&\quad \int_0^t g_{N-1}(\tau)\gamma_{N-1,N_1}(t-\tau)d\tau \qquad (9)
\end{aligned}
$$

and

$$
\begin{aligned}
g_1(\tau) &= \int_0^\tau \gamma_0(t)l_0(\tau-t)dt \,, & g_{N_1+\varepsilon}(t) &= \gamma_{N_1}^L(t) \,, \\
g_{N-1}(\tau) &= \int_0^\tau \gamma_N(t)l_N(\tau-t)dt \,, & g_{N_1-\varepsilon}(t) &= \gamma_{N_1}^R(t) \,. \qquad (10)
\end{aligned}
$$

Equations (9) and (10) form a set of eight equations with eight unknown functions. When we transform these equations with the use of Laplace transform, the convolutions of density functions become products of transforms and we have a set of linear equations where the unknown variables are: $\bar{g}_1(s)$, $\bar{g}_{N_1-\varepsilon}(s)$, $\bar{g}_{N_1+\varepsilon}(s)$, $\bar{g}_{N-1}(s)$, $\bar{\gamma}_0(s)$, $\bar{\gamma}_N(s)$, $\bar{\gamma}_{N_1-\varepsilon}(s)$, $\bar{\gamma}_{N_1+\varepsilon}(s)$. They may be expressed by all other functions, that means $\bar{\gamma}_{\psi_1,0}(s)$, $\bar{\gamma}_{\psi_1,N_1}(s)$, $\bar{\gamma}_{1,0}(s)$, $\bar{\gamma}_{1,N_1}(s)$, $\bar{\gamma}_{N_1-\varepsilon,0}(s)$, $\bar{\gamma}_{N_1-\varepsilon,N_1}(s)$, $\bar{\gamma}_{\psi_2,N1}(s)$, $\bar{\gamma}_{\psi_2,N}(s)$, $\bar{\gamma}_{N_1+\varepsilon,N_1}(s)$, $\bar{\gamma}_{N_1+\varepsilon,N}(s)$, $\bar{\gamma}_{N-1,N_1}(s)$, $\bar{\gamma}_{N-1,N}(s)$ which are already determined with the use of functions $\phi_1(x,t;x_0)$, $\phi_2(x,t;x_0)$. This way we obtain the functions $\bar{g}_1(s)$, $\bar{g}_{N_1-\varepsilon}(s)$, $\bar{g}_{N_1+\varepsilon}(s)$, $\bar{g}_{N-1}(s)$ and use them in the pdfs (7), (8). The time-domain originals $f_1(x,t;\psi_1)$, $f_2(x,t;\psi_2)$ are obtained numerically (Stehfest, 1970) from their transforms. The density of the whole process is

$$f(x,t;\psi) = \begin{cases} f_1(x,t;\psi_1) & \text{for} \quad 0 < x < N_1 \, , \\ f_2(x,t;\psi_2) & \text{for} \quad N_1 < x < N \, . \end{cases}$$

This algorithm can be extended to include several zones with different diffusion parameters. These zones will be bounded by barriers similar to the one which we have placed at $x = N_1$.

To see the evolution of the number of cells belonging to a class, we have to consider the composition of input and output streams. Let us denote by $p^{(i)}(t)$ probability that a cell arriving at time t belongs to class i.

$$p^{(i)}(t) = \frac{\lambda_{\text{eff}}^{(i)}(t)}{\lambda_{\text{eff}}^{(1)}(t) + \lambda_{\text{eff}}^{(2)}(t)} \, , \tag{11}$$

where

$$\lambda_{\text{eff}}^{(1)}(t) = \lambda^{(1)}(t)[1 - p_N(t)], \qquad \lambda_{\text{eff}}^{(2)}(t) = \lambda^{(2)}(t)[1 - p_{n \ge N_1}(t)] \tag{12}$$

and $p_{n \ge N_1}(t)$ is probability that the buffer space accessible for class 2 cells is full and these cells are rejected. We try to reflect the mutual influence of both classes in effective parameters of their service and then analyze the class behaviour independently. We know the distribution of the total number $n(t)$ of cells in the buffer at time t. Among those cells there are $n^{(2)}(t)$ class 2 cells. Let us denote by $\nu(t)$, $0 \le \nu(t) \le n$, the number of class 1 cells gathered at the end of the buffer behind the last class 2 cell, seen at time t by the arriving new class 2 cell. As the service time is equal to one time unit, the effective service time for the arriving class 2 cell is $1 + \nu$. If $n^{(2)}(t) > 0$ then

$$P[\nu = i \mid n(t) = n, n^{(2)}(t) = n^{(2)}] = \frac{C_{n-i-1}^{n^{(2)}-1}}{C_n^{n^{(2)}}} \, , \qquad 0 \le \nu \le n - n^{(2)}, \tag{13}$$

where

$$C_m^l = \begin{cases} \dfrac{m!}{l!(m-l)!} & \text{for } m \ge l \ge 0 \, , \\ 0 & \text{otherwise} \, . \end{cases}$$

If $n^{(2)} = 0$ then $P(\nu = i \mid n) = \delta_{i,n}$ where $\delta_{i,n}$ is the Kroenecker symbol. We determine the probability of $\nu = i$ when $n(t) = n$, for all possible $n^{(2)}$:

$$P[\nu = i \mid n(t) = n] =$$

$$= (1 - p^{(2)})^i \delta_{i,n} + \sum_{n^{(2)}=1}^{\min\{n-i, N_1-1\}} P[n^{(2)} \mid n] \; P[\nu = i \mid n, n^{(2)}] =$$

$$= (1 - p^{(2)})^i \delta_{i,n} + \sum_{n^{(2)}=1}^{\min\{n-i, N_1-1\}} C_{n-i-1}^{n^{(2)}-1} p^{(2)^{n^{(2)}}} (1 - p^{(2)})^{n-n^{(2)}} \tag{14}$$

and the distribution of ν is

$$P[\nu = i] = (1-p^{(2)})^i P[n(t)=i] + \sum_{l=0}^{N} P[n(t)=l] \sum_{m=1}^{\min\{l-i,N_1-1\}} C_{l-i-1}^{m-1} p^{(2)^m} (1-p^{(2)})^{l-m}. \quad (15)$$

Now we are able to determine the mean and squared coefficient of variation of the random variable $B^{(2)}$ representing the effective service time for class 2 cells, $B^{(2)}(t) = 1 + \nu(t)$:

$$E[B^{(2)}(t)] = 1 + E[\nu(t)],$$
$$C_B^{(2)^2}(t) = \frac{E[B^{(2)^2}(t)]}{E[B^{(2)}(t)]^2} - 1 = \frac{E[\nu^2(t)] - E[\nu(t)]^2}{(E[\nu(t)]+1)^2}. \quad (16)$$

The coefficient $C_B^{(2)^2}$ is given by Eq. (16); $C_B^{(1)^2}$ is deduced on the similar principle and $C_A^{(1)^2}$, $C_A^{(2)^2}$ are also deduced from the input streams.

The changes in the intensities of the input rates at the instant t influence the output with a delay of $n(t)$. The service times of $\nu(t)$ class 1 cells which are at the end of the queue are considered as a part of the service of the arriving class 2 cell. The change of the input at t is taken into account in the service time $\mu^{(2)}(t+n(t)-\nu(t))$. On the other hand, the tile composition of the queue does not depend only on the input composition $p(t)$ but on its evolution since the last class 2 cell arrival moment. It is not easy to determine in transient analysis the delay with which the input changes act on the $\nu(t)$. As a rough approximation, we considered a delay equal to $n(t)/2$. This choice permits us to deal with sudden falls of the class 2 input rate. Although this method captures the dynamics of the second class cell number, further efforts seem to be necessary to obtain more general characterization of time dependent queue composition.

The output stream characteristics may be also presented by the equation used to describe the interdeparture times pdf at the $G/G/1$ station (Gelenbe, 1976):

$$d^{(i)}(x) = \varrho^{(i)} b(x) + (1-\varrho^{(i)}) a^{(i)}(x) * b^{(i)}(x), \quad (17)$$

which gives us the squared coefficient of variation of interevent times in the output stream

$$C_D^{(i)^2} = C_A^{(i)^2}(1-\varrho^{(i)}) + \varrho^{(i)^2} C_B^{(i)^2} + \varrho^{(i)}(1-\varrho^{(i)}). \quad (18)$$

The transient solution (7), (8) assumes constant diffusion parameters, therefore for the use of it we fixe the values of α, β during the intervals of the length of one time unit. The solution obtained at the end of an interval is used as the initial condition, i.e. functions ψ_1, ψ_2 in Eqs. (7)–(9) determining the solution in the next interval with new values of α and β.

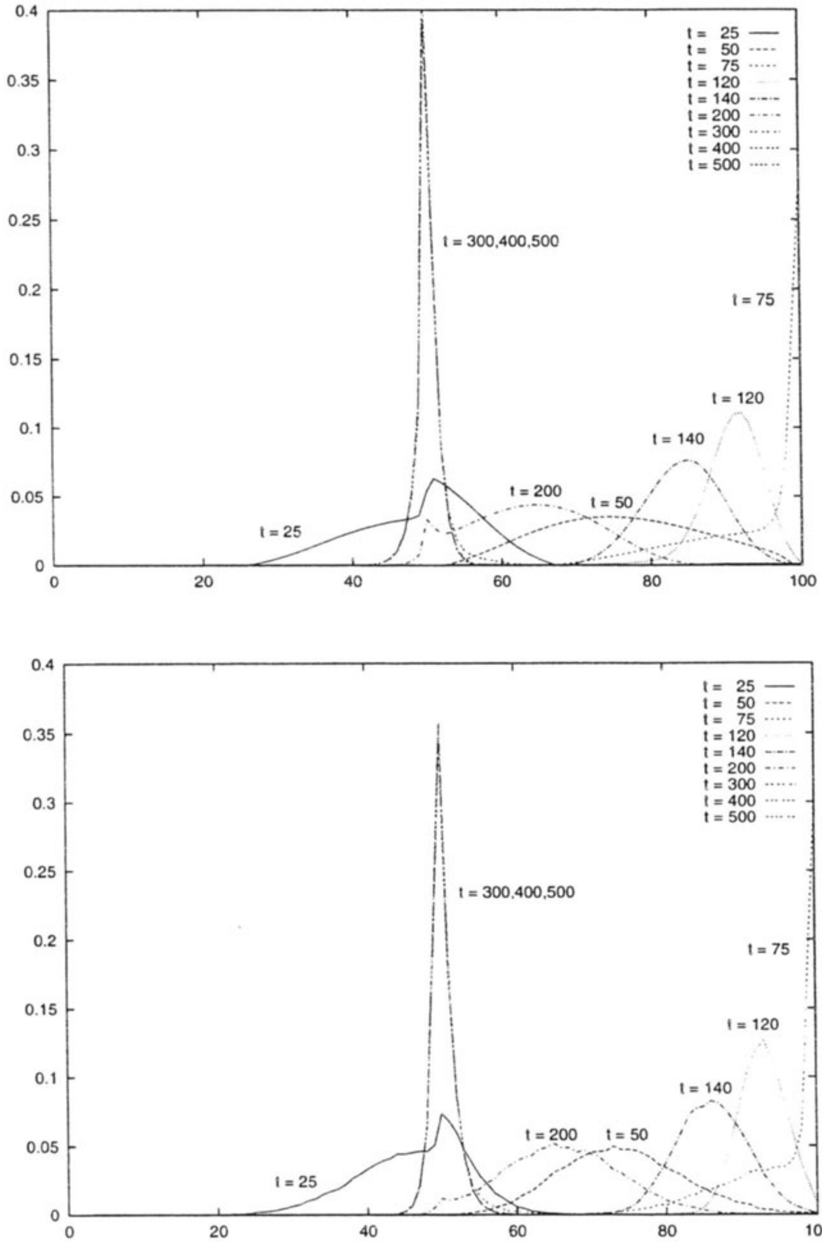
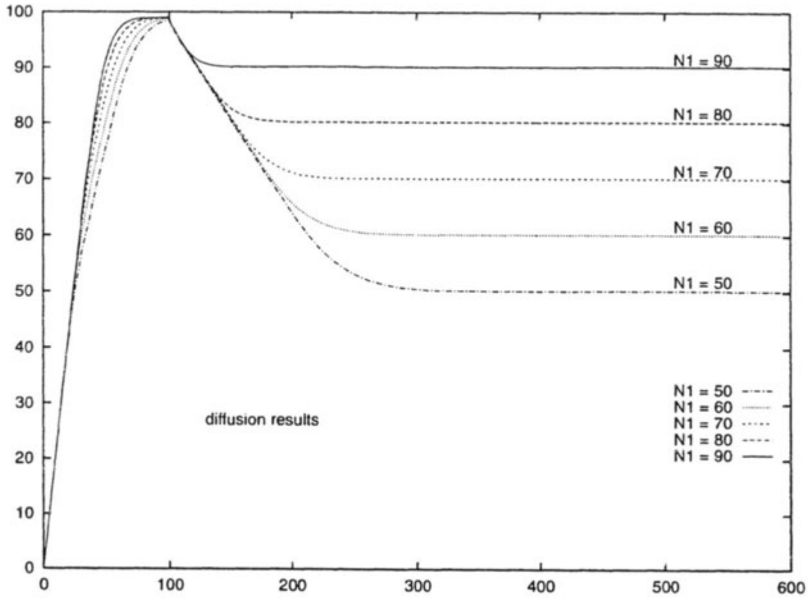


Figure 10 Distribution of the number of all cells in the buffer for several time moments $t = 25 - 500$; buffer size $N = 100$, threshold $N_1 = 50$; simulation (above) and diffusion (below) results



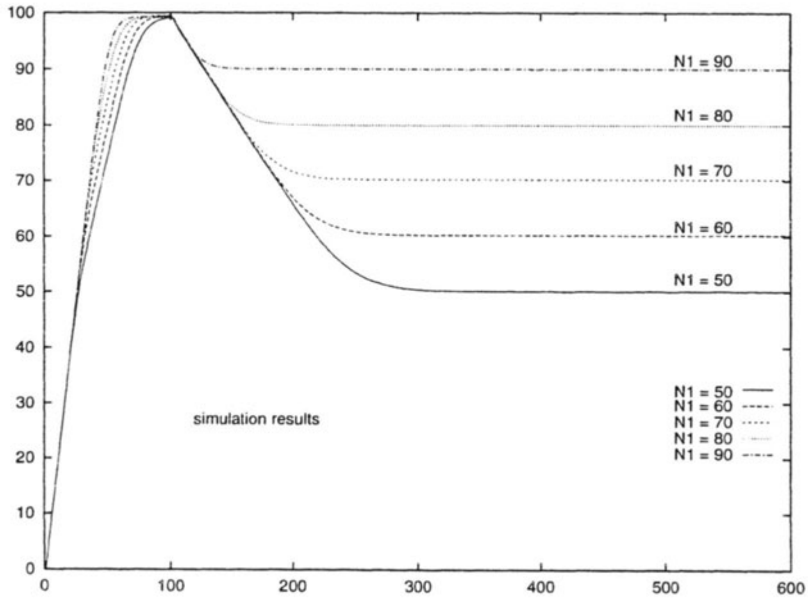


Figure 11 Mean number of cells as a function of time, parametrized by the value $N_1 = 50, 60, 70, 80, 90$ of the threshold, simulation (above) and diffusion (below) results

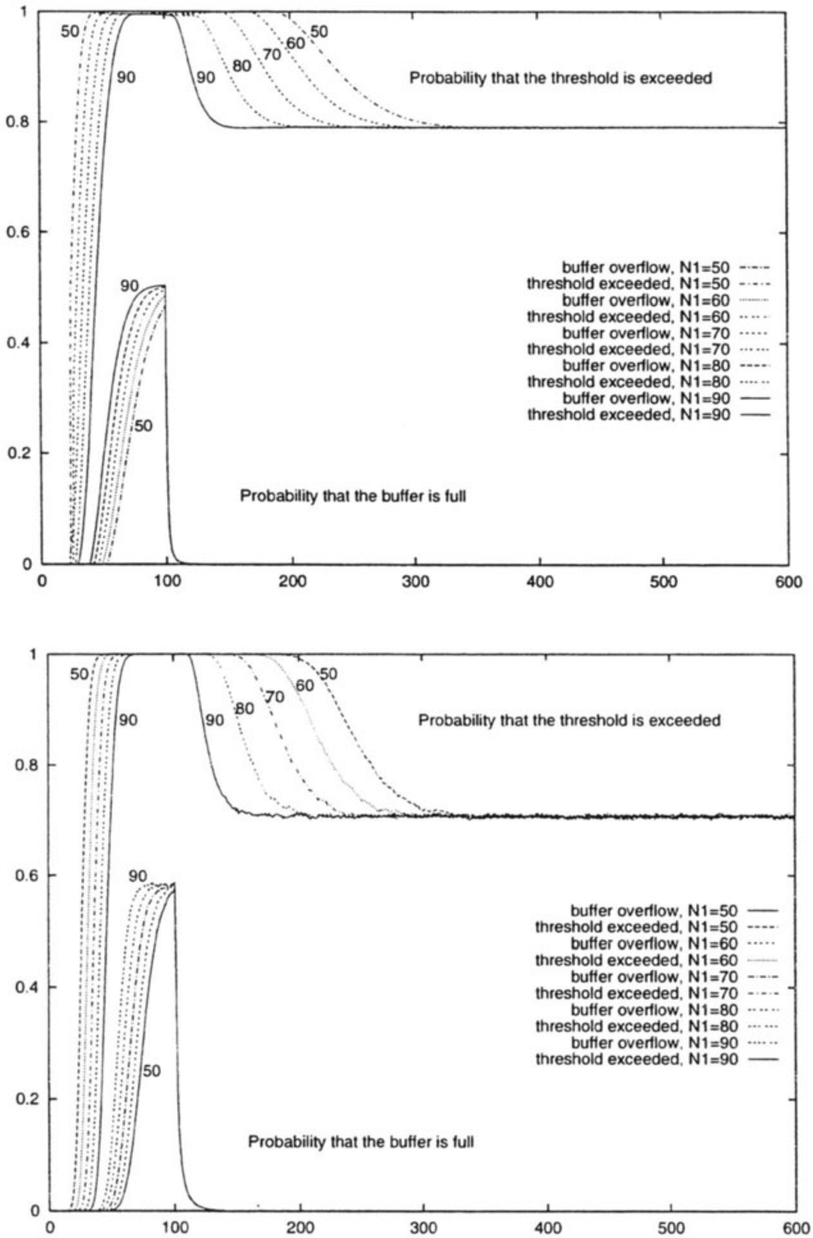


Figure 12 Probability that the buffer of length $N = 100$ is full (priority cells are lost) and that the threshold is exceeded (ordinary cells are lost) as a function of time, parametrized by the threshold value $N_1 = 50, 60, 70, 80, 90$; simulation (above) and diffusion (below) results

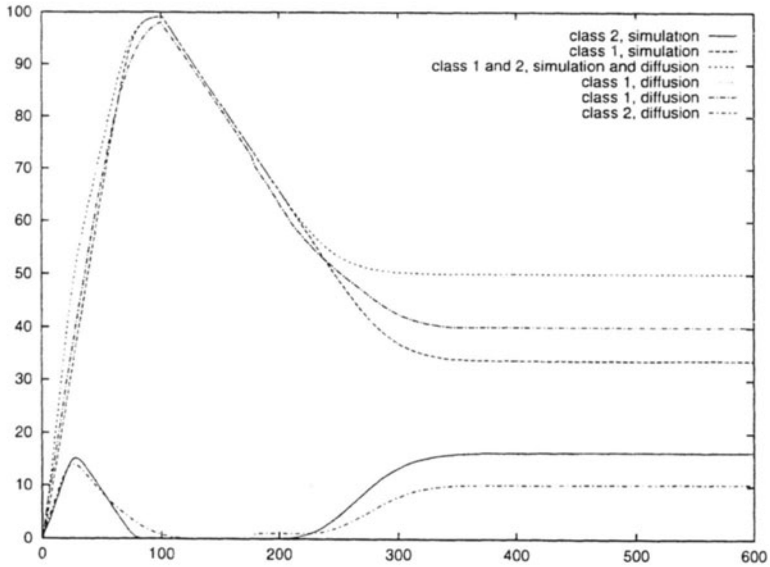


Figure 13 Mean value of class 1 and class 2 cells as a function of time; $N = 100$, $N_1 = 50$

Numerical example. Let us suppose that at the beginning the buffer is empty and during the interval $t \in [0, 100]$ the input stream of priority cells has ratio $\lambda^{(1)} = 2$ cells per time unit and the one of low priority cells $\lambda^{(2)} = 1$; for $t > 100$ the ratio of high priority cells is $\lambda^{(1)} = 0.6667$, the ratio of low priority cells does not change. The service time is constant and equal one time unit. The buffer length is $N = 100$, the value of threshold varies between $N_1 = 50$ and $N_1 = 90$. The value of ε in Eqs. (7)–(9) was chosen $\varepsilon = 0.1$. In Figure 10 the distributions of the number of cells at the buffer obtained by simulation and by diffusion model for chosen time moments are compared. Diffusion and simulation results are placed in separate figures to preserve their legibility. The shape of curves given by two models is very similar. At the end of second period ($t = 400, 500, 600$) the steady state distribution is attained.

Figure 11 displays the mean number of cells in the buffer as a function of time. During the first 100 time units the congestion is clearly visible, the buffer quickly becomes saturated; during the second period the queue is also overcrowded, probability that the threshold is exceeded is near 0.7 but owing to the buffer sharing policy the probability that the buffer is inaccessible for priority cells remains negligible – Figure 12. The threshold value N_1 is a parameter of displayed curves. If N_1 increases the mean values of low priority cells increases (they have more space in the buffer, hence less of them is rejected) and the number of priority cells increases too (as there is more class 2 cells in the queue, class 1 cells wait longer).

Figure 13 displays the mean number of high and low priority cells given by the approach we have described above using Eqs. (11)–(16) and compared with simulation results. We see that the steady state mean value of class 2 cells is underestimated (because of overestimation of class 2 losses by diffusion approximation seen in Figure 12) but the dynamics of class 2 cells vanishing from the queue during heavy saturation periods is well captured.

Some numerical problems were encountered when computing expressions of $\phi(x,t,\psi)$ and $\gamma_0(t)$ for very small values of $\lambda^{(1)}_{\text{eff}}(t)$, $\mu^{(2)}_{\text{eff}}(t)$ and forced us to very careful programming.

4 CONCLUSIONS

This article is sequel to authors' previous studies applying the diffusion approximation in analysis of phenomena related to modern telecommunication networks: push-out policy (Czachórski, 1992), dynamics of flow changes along virtual path (Czachórski, 1994), jitter and flow synchronization (Czachórski, 1995), feed-back traffic control using explicit congestion notification (Atmaca, 1995). It confirms their conviction that the diffusion approximation is a useful tool to solve queueing models, in spite of some drawbacks (the method errors are not negligible and, when transient states are considered, the time needed to develop necessary software, to overcome related numerical problems and to perform calculations increases significantly with the complexity of models). The advantage of diffusion approximation lies in its flexibility to be adapted to various queueing disciplines, in its ease to develop queueing network models and to include customer classes with general interarrival and service time distributions and, especially, in its possibilities to deal with transient states. Both studied models give the characteristics of the output streams, therefore leaky buckets and nodes with buffer sharing policy can be included in a general queueing network diffusion model. Such a model was formulated for an arbitrary topology open network of G/G/1 and G/G/1/N stations in (Gelenbe, 1976) and adapted to transient analysis in (Duda, 1986), (Czachórski, 1994).

5 REFERENCES

Akhtar, S. (1987) Congestion control in a Fast Packet Switching Network, Master's Thesis, Washington University.

Atmaca, T., Czachórski, T. and Pekergin, F. (1995) A Diffusion Model of the Dynamic Effects of Closed-Loop Feedback Control Mechanisms in ATM Networks, in *Proceedings of 3rd IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, Ilkley.

Cox, D.R. and Miller, H.D. (1965) The Theory of Stochastic Processes. Methuen, London.

Czachórski T., Fourneau J-M., Pekergin F. (1992) Diffusion model of a push-out buffer management policy, in *Proceedings of IEEE INFOCOM*, Florence.

Czachórski T. (1993) A method to solve diffusion equation with instantaneous return processes acting as boundary conditions. *Bulletin of Polish Academy of Sciences, Technical Sciences*, **41**, 421-427.

Czachórski, T., Fourneau, J-M. and Pekergin, F. (1994) Diffusion Models to Study Nonstationary Traffic and Cell Loss in ATM Networks, in *ACM 2nd Workshop on ATM Networks*, Bradford.

Czachórski, T., Fourneau, J-M. and Kloul L. (1995) Diffusion Approximation to Study the Flow Synchronization in ATM Networks, in *Proceedings of 3rd IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, Ilkley.

Duda A., (1986) Diffusion Approximations for Time-Dependent Queueing Systems. *IEEE Journal on Selected Areas in Communications*, **SAC-4**, 6.

Elwalid, A.I. and Mitra D., (1991) Stochastic Fluid Models in the Analysis of Access Regulation in High Speed Networks, in *Procedings of GLOBECOM*, Geneva.

Gelenbe, E. (1975) On Approximate Computer Systems Models. *Journal of ACM* **22**, 261-273.

Gelenbe, E. and Pujolle, G. (1976) The Behaviour of a Single Queue in a General Queueing Network. *Acta Informatica*, **7**, 123-136.

Gelenbe, E., Mang, X. and Feng, Y. (1996) A diffusion cell loss estimate for ATM with multiclass bursty traffic, in *Performance Modelling and Evaluation of ATM Networks, Vol. 2* (ed. D.D. Kouvatsos) Chapman and Hall, London.

Hébuterne G. and Gravey A. (1989) A Space Priority Queueing Mechanism for Multiplexing ATM Channels, in *Proceedings of ITC Specialist Seminar*, Adelaide.

Holtsinger, D.S. (1992) Performance Analysis of Leaky Bucket Policing Mechanisms, Ph.D. thesis, Dept. of Electrical and Computer Engineering, North Carolina State University.

Kröner, H., Hébuterne, G., Boyer, P. and Gravey, A. (1991) Priority management in ATM switching nodes. *IEEE Journal on Selected Areas in Communications*, **9**, 418-427.

Meyer J.F., Montagna, S. and Paglino, R. (1993) Dimentioning of an ATM switch with shared buffer and threshold priority. *Computer Networks and ISDN Systems*, **26**, 95-108.

Newell, G.F. (1971) Applications of Queueing Theory. Chapman and Hall, London.

Örs, T. and Jones, S.P.W. (1995) Performance Optimization of ATM Input Control using Multiple Leaky-Buckets, in *Proceedings of ACM 3rd Workshop on ATM Networks*, Ilkley.

Stehfest H. (1970) Algorithm 368: Numeric Inversion of Laplace Transform. *Communications of ACM*, **13**, 47-49.