ON/OFF Sources in an Interconnection Network: Performance Analysis when Packets are Routed to the Shortest Queue of two Randomly Selected Nodes

Philippe Jacquet, Nikita Vvedenskaya

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

ON/OFF SOURCES IN AN INTERCONNECTION NETWORK: PERFORMANCE ANALYSIS WHEN PACKETS ARE ROUTED TO THE SHORTEST QUEUE OF TWO RANDOMLY SELECTED NODES

Philippe Jacquet, Nikita Vvedenskaya

N° 3570

THÈME 1

Rapport de recherche

ISSN 0249-6399

ON/OFF SOURCES IN AN INTERCONNECTION NETWORK: PERFORMANCE ANALYSIS WHEN PACKETS ARE ROUTED TO THE SHORTEST QUEUE OF TWO RANDOMLY SELECTED NODES

Philippe Jacquet, Nikita Vvedenskaya

Thème 1 — Réseaux et systèmes
Projet Hipercom

Rapport de recherche n˚3570 — — 36 pages

Abstract: The authors* investigate a system with N servers and with N sources connected with the servers. A sources can be in state "on" or "off". In state "on" the source generates the Poisson flow of packets of rate λ. The service time of a packet is distributed exponentially with mean one. Upon its arrival a packet is directed to the server with the shortest queues of the following two servers: the server where the packet has been generated and another randomly selected server. The queue length probability as $N \to \infty$ is investigated.

Key-words: on/off sources, network, routing, subexponential tails, superexponential tail

(Résumé : tsvp)

* The first author is with Institute for Information transmission Problems, Russian AS, Moscow. The work of the author was partly supported by Russian Fundamental Research Foundation (grant 96-01-0050) and by French-Russian Center of Applied Mathematics (progect 6-97.) and by HIPERCOM project

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : 01 39 63 55 11 - International : +33 1 39 63 55 11
Télécopie : (33) 01 39 63 53 30 - International : +33 1 39 63 53 30

Sources on/off dans un réseau d'interconnexion: analyse des performances quand les paquets sont routés vers la file d'attente la plus courte entre deux serveurs choisis aléatoirement

Résumé : Nous considérons un système constitué de N serveurs connectés à N sources. Une source est alternativement en état "on" et en état "off". À l'état "on" la source crée un flot poissonien de paquets de taux λ par unité de temps. Le temps de service sur un paquet est distribué suivant une loi exponentielle de moyenne un. Lorsqu'un paquet est créé il est dirigé vers la plus courte des file d'attente des deux serveurs suivants: le serveur où le paquet a ́té créé et un autre serveur choisi de manière aléatoire.

Mots-clé : Sources on/off, réseaux, routage, distributions sous-exponentielles, distribution super-exponentielles

1 Introduction

In this paper we investigate the performance of an interconnection network where traffic is generated by "on/off" sources. An "on/off" source is a source which alternates between state "on" and state "off". In state "on" the source generates packets according to a Poisson rate λ. In state "off" the source does not generate any packet. An "on/off" source is a convenient model of the behavior of a user in a data network. Indeed, "on" periods are when the user uploads or downloads files of data, and "off" periods are when the user is inactive regarding network activity: *i.e.* is thinking or locally processing its data.

We model the interconnection network S_N by a set of N nodes. On each node there is a user, a queue and a server. In absence of routing the user directs its packet to the queue of its node. In this paper we will investigate the effect of the following routing protocol:

1. when a user generates a packet, it randomly selects a remote node;
2. if the queue at the remote node is shorter than the local queue, the packet is directed to the queue of the remote node, if the remote queue is greater than the local queue, the packet is directed to the local queue;
3. if both remote and local queues are equal then the packet is randomly directed to the local queue with probability 1/2 or to the remote queue with probability 1/2.

This routing protocol has been proposed and analyzed in [6] (see also [3], [4], [5], [8], [9]) where user traffic model is Poisson of rate λ, $\lambda < 1$, and the service is exponential of rate 1. Under this model it was proved that as $N \to \infty$ the distribution of the queue size Q_R of any random node has super-exponential tail:

$$\Pr\{Q_R > k\} = O(\lambda^{2^k})$$

This result sounds very interesting, since without routing the queues are independent and the distribution of their size Q_U have only exponential tails:

$$\Pr\{Q_U > k\} = O(\lambda^k)\ .$$

The present paper is intended to analyze the same routing protocol under the assumption that user traffic models are independent "on/off" sources and the service is exponential of rate 1. We will prove that as $N \to \infty$ queue size Q_R has still super-exponential tail.

This result also sounds interesting because a large number of the "on/off" sources can sometimes create a cumulated traffic with long term dependencies [11](see also [12]- [16]). These long term dependencies are known to produce queueing with heavy tails. In other words

$$\Pr\{Q_U > k\} = O(k^{-\beta})$$

for some $\beta > 0$.

The system model of [6] is a model of a queueing systems with interacting servers and with dynamic routing of the packets. Mathematical investigation of such systems is a very difficult problem. We are not interested in specific interconnection network between nodes; we just assume that every node are equally visitable, as it could be in a *crossbar* network. In this case it may be attractive to investigate the systems with very large number of servers using the mean field approximation of statistical physics. This approach was used in a number of papers (see, for example, [1], [3]– [9] and the references in [7].) It was shown ([6], [7], [9], [8]) that using the information about a small number of randomly selected queues and directing the packet to a server that is less loaded may considerably reduce the mean lengths of the queues.

In the above mentioned papers the input flow was supposed to be Poisson of constant rate, and, in most cases, with an exponentially distributed packet service time. We are interested in case where the input flow is Poisson with varying rates.

In section 2 we describe the model of the system where all sources have the same transition parameters $\tau^{\pm}$. In section 3 we consider the system of differential equations that describe the system S_N in the limit as $N \to \infty$. Main theorems about the convergence as $N \to \infty$ of S_N to the limit dynamical system are presented in section 4. The system where sources have different transition parameters $\tau^{\pm}$ is considered in section 5. Here the case of infinite number of different $\tau^{\pm}$ is investigated with and without routing.

2 System model

Consider a system S_N that consists in N servers connected with N sources that can be in state "on" or "off". Each source is connected with "its" server and has potential to be connected to any of the other servers. In state "on" the source generates Poisson flow of packets of rate λ. In state "off" the source generates no packet.

The transition from "on" to "off" and from "off" to "on" is Markovian: each source changes "on" to "off" with intensity τ^-, and changes "off" to "on" with intensity τ^+ independently of the mode of other sources.

Upon a packet arrival at a server (connected with its source), an other server is selected (all servers are equally probable.) The packet is directed to the least busy of two servers, where it joins a FCFS queue. (By least busy we mean the server with a shorter queue). If the queue lengths of two servers are equal each server is selected with probability 1/2.

The service time of each message is independent of service times of other messages and is distributed exponentially with mean equal to 1.

The work of S_N can by described by of a Markov chain with the state space represented by the two sequences $\mathbf{u}_N = (\mathbf{u}_N^+ , \mathbf{u}_N^-)$, $\mathbf{u}_N^{\pm} = \{u_{N,k}^{\pm}\}_{k=0}^{\infty}$, $u_{N,k}^{\pm} = r_k^{\pm}/N$. Here r_k^+ is the number of servers in mode "on" with queue length equal to k, and r_k^- is the number of servers in mode "off" with queue length equal to k.

It is useful for us to consider the performance of S_N as a Markov factor-chain with the state space represented by two sequences $\mathbf{U}_N = (\mathbf{U}_N^+ , \mathbf{U}_N^-)$, $\mathbf{U}_N^{\pm} = \{U_{N,k}^{\pm}\}_{k=0}^{\infty}$, $U_{N,k}^{\pm} = R_k^{\pm}/N$. Here R_k^+ is the number of servers in mode "on" with queue length not less then k, and R_k^- is the number of servers in mode "off" with queue length not less then k. Obviously,

$$u^{\pm}_{N,k} = U^{\pm}_{N,k} - U^{\pm}_{N,k+1}.$$

Below we will omit index N and use the notations

$$u_k = u_k^+ + u_k^-, \quad \mathbf{u} = \{u_k\}_{k=0}^{\infty}, \tag{1}$$

$$U_k = U_k^+ + U_k^-, \quad U^{\pm}_{k+1} - U^{\pm}_k = u^{\pm}_k, \quad \mathbf{U} = \{U_k\}_{k=0}^{\infty}. \tag{2}$$

We suppose that the non-overload condition holds

$$\lambda \frac{\tau^+}{\tau^+ + \tau^-} = \lambda_0 < 1. \tag{3}$$

The behavior of the Markov system is described by a system of differential equations of the form $\dot{\mathbf{U}} = A_N f(\mathbf{U})$, where $\dot{\mathbf{U}}$ denotes the first derivative of sequence $\mathbf{U}$ with respect to time. Quantity $A_N f(\mathbf{U})$ is the generating operator of the Markov factor-chain. The generating operator $A_N f(\mathbf{U})$ has the form

$$\begin{array}{rcl} A_N f(\mathbf{U}) & = & N\lambda \sum_{i=1}^{\infty} \Big(f(\mathbf{U} + e_i^+/N) - f(\mathbf{U}) \Big) \frac{U_{i-1}^+ + U_i^+ + U_{i-1} + U_i}{2} \Big(U_{i-1}^+ - U_i^+ \Big) \\ & & + N \sum_{i=1}^{\infty} \Big(f(\mathbf{U} - e_i^+/N) - f(\mathbf{U}) \Big) \Big(U_i^+ - U_{i+1}^+ \Big) \\ & & + N\lambda \sum_{i=1}^{\infty} \Big(f(\mathbf{U} + e_i^-/N) - f(\mathbf{U}) \Big) \frac{U_{i-1}^+ + U_i^+}{2} \Big(U_{i-1}^- - U_i^- \Big) \\ & & + N \sum_{i=1}^{\infty} \Big(f(\mathbf{U} - e_i^-/N) - f(\mathbf{U}) \Big) \Big(U_i^- - U_{i+1}^- \Big) \\ & & + N\tau^+ \sum_{i=0}^{\infty} \Big(f(\mathbf{U} + (e_i^+ - e_i^-)/N) - f(\mathbf{U}) \Big) U_i^- \\ & & + N\tau^- \sum_{i=0}^{\infty} \Big(f(\mathbf{U} - (e_i^+ + e_i^-)/N) - f(\mathbf{U}) \Big) U_i^+ \ , \end{array} \tag{4}$$

Here $e_i^{\pm}$ is a vector $(0, ...0, 1, 0...)$, with i-th coordinate equal to 1 and with all other coordinates equal to 0.

Let us comment on (4). For example, the first RHS term of (4) indicates that a server in mode "on" with a queue of length $i-1$ is selected. That happens in two cases:

1. the server is selected at the first step and at the second step a server with the queue length not shorter then $i-1$ is selected. That happens with probability $\Big(U_{i-1}^+ - U_i^+ \Big) \frac{U_{i-1} + U_i}{2}$.
2. the server is selected at the second step, while at the first step a server in mode "on" with the queue length not shorter then $i-1$ was selected. That happens with probability $\Big(U_{i-1}^+ - U_i^+ \Big) \frac{U_{i-1}^+ + U_i^+}{2}$.

The second term indicates that a service at a server in mode "on" with a queue length i is finished. The last two terms stand for the change of the modes of the servers with the queue lengths equal to i. Other terms can be explained similarly.

We will show that as $N \to \infty$ the performance of S_N can be described by a dynamical systems.

$$\begin{aligned} \dot{U}_i^+(t) &= \lambda(U_{i-1}^+(t) - U_i^+(t))[U_{i-1}^+(t) + U_i^+(t) + U_{i-1}(t) + U_i(t)]/2 \\ &\quad + U_{i+1}^+(t) - U_i^+(t) + \tau^+ U_i^-(t) - \tau^- U_i^+(t), \quad i \geq 1, \end{aligned} \tag{5}$$

$$\begin{aligned} \dot{U}_i^-(t) &= \lambda(U_{i-1}^-(t) - U_i^-(t))[U_{i-1}^+(t) + U_i^+(t)]/2 \\ &\quad + U_{i+1}^-(t) - U_i^-(t) - \tau^+ U_i^-(t) + \tau^- U_i^+(t), \quad i \geq 1, \end{aligned} \tag{6}$$

$$\dot{U}_0^+(t) = \tau^+ U_0^-(t) - \tau^- U_0^+(t), \quad \dot{U}_0^-(t) = -\tau^+ U_0^-(t) + \tau^- U_0^+(t). \tag{7}$$

$$U_0^+(t) + U_0^-(t) \equiv 1, \tag{8}$$

The sum of (5) and (6) gives

$$\dot{U}_i(t) = \lambda\Big(U_{i-1}(t)U_{i-1}^+(t) - U_i(t)U_i^+(t)\Big) + U_{i+1}(t) - U_i(t). \tag{9}$$

Consider the initial-value problem for equations (5)- (9) with the initial conditions at $t = 0$:

$$U_i^\pm(0) = G_i^\pm, \quad G_i^\pm \geq G_{i+1}^\pm, \quad i \geq 0, \quad G_0^+ + G_0^- = 1. \tag{10}$$

Note that the condition (3) can be rewritten in the form

$$\lambda U_0^+(\infty) = \lambda_0 < 1. \tag{11}$$

The solution of (5)- (10) will be sometimes denoted by $\mathbf{U}(t, G)$ to indicate the dependence on initial values G.

Investigating the properties of solution of (5)- (10) we introduce the space $\overline{\mathcal{U}}$ of sequences $\mathbf{U} = \{U_i^+, U_i^-\}_{i=0}^\infty$:

$$\overline{\mathcal{U}} : \ \{\mathbf{U}, \ \ U_0^+ + U_0^- = 1, \ \ U_i^\pm \geq U_{i+1}^\pm \geq 0, \ \ i = 0, 1, ...\}, \tag{12}$$

with the norm

$$\rho\Big(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}\Big) = \sup_{i \geq 0} \frac{| U_i^{\pm(1)} - U_i^{\pm(2)} |}{i+1}, \quad \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \in \overline{\mathcal{U}},$$

and the spaces

$$\mathcal{U} : \{\mathbf{U} \in \overline{\mathcal{U}}, \ \sum_{i=1}^\infty U_i < \infty\}, \tag{13}$$

$$\mathcal{U}_N : \{\mathbf{U}_N \in \mathcal{U}, \ \ U_{N,k}^\pm = R_k/N\}. \tag{14}$$

The space $\mathcal{U}_N$ is the set of attainable sequences $\mathbf{U}_N$ when N is fixed and t varies.

3 The analysis of the identical sources case

By identical sources we mean sources with same $\tau^{\pm}$ and λ parameters. Our aim is to present the asymptotic analysis of the stationary distribution when $N \to \infty$. We will first focuse on the tail of the stationary distribution. Second we will present an analytical method to obtain a numerical estimate of any coefficient of this limiting distribution.

3.1 The solution properties for some differential equations

To investigate the initial-value problem (5)- (10) we consider first a similar problem for a truncated system.

$$\begin{aligned} \dot{U}_i^+(t) &= \lambda(U_{i-1}^+(t) - U_i^+(t))[U_{i-1}^+(t) + U_i^+(t) + U_{i-1}(t) + U_i(t)]/2 \\ &+U_{i+1}^+(t) - U_i^+(t) + \tau^+ U_i^-(t) - \tau^- U_i^+(t), \quad 1 \le i \le K, \end{aligned} \tag{15}$$

$$\begin{aligned} \dot{U}_i^-(t) &= \lambda(U_{i-1}^-(t) - U_i^-(t))[U_{i-1}^+(t) + U_i^+(t)]/2 \\ &+U_{i+1}^-(t) - U_i^-(t) - \tau^+ U_i^-(t) + \tau^- U_i^+(t), \quad 1 \le i \le K, \end{aligned} \tag{16}$$

$$\begin{aligned} \dot{U}_0^+(t) &= \tau^+ U_0^-(t) - \tau^- U_0^+(t), \\ \dot{U}_0^-(t) &= -\tau^+ U_0^-(t) + \tau^- U_0^+(t), \end{aligned}$$

$$U_0^+(t) + U_0^-(t) \equiv 1. \tag{17}$$

$$\dot{U}_i(t) = \lambda(U_{i-1}(t)U_{i-1}^+(t) - U_i(t)U_i^+(t)) + U_{i+1} - U_i, \quad 1 \le i \le K, \tag{18}$$

$$U_i^{\pm}(0) = G_i^{\pm}, \quad 0 \le i \le K, \quad G_0^+ + G_0^- = 1. \; U_{K+1}^{\pm} = c = \text{const} \ge 0, \tag{19}$$

$$c \le \inf_{t>0}(U_0^+(t), U_0^-(t))$$

Lemma 1 *If initial values of problem (15)- (19) satisfy conditions*

$$U_i^+(0) \ge U_{i+1}^+(0), \quad U_i^-(0) \ge U_{i+1}^-(0), \quad i = 0, 1, ..., K, \tag{20}$$

then the solution of the problem satisfies these conditions for all t.

PROOF. Because of the continuous dependence of a solution on initial values it is sufficient to consider a case where strict inequalities hold in (20) for $i > 0$ and to prove that such inequalities are valid for all t. Suppose that the contrary takes place. Let strict inequalities hold for $t < t_0$, and let some equalities to appear at $t = t_0$. Since $U_0^{\pm}(t) > U_{K+1}^{\pm}(t), \; t \ge 0$, at least one of two cases takes place :

1. there exists i such that $U_{i-1}^+(t_0) > U_i^+(t_0) = U_{i+1}^+(t_0)$ or $U_{i-1}^-(t_0) > U_i^-(t_0) = U_{i+1}^-(t_0)$

2. there exists i such that $U_{i-1}^+(t_0) = U_i^+(t_0) > U_{i+1}^+(t_0)$ or $U_{i-1}^-(t_0) = U_i^-(t_0) > U_{i+1}^-(t_0)$.

Let, for example, the case 1 takes place and let the equality holds for U_i^+, U_{i+1}^+. It follows from (15) that

$$\begin{aligned}\dot{U}_i^+(t_0) &= \lambda(U_{i-1}^+(t) - U_i^+(t))[U_{i-1}^+(t) + U_i^+(t) + U_{i-1}(t) + U_i(t)]/2 + \tau^+ U_i(t) - (\tau^- + \tau^+)U_i^+(t) \\ &\geq \dot{U}_{i+1}(t_0) = U_{i+2}^+(t) - U_{i+1}^+(t) + \tau^+ U_{i+1}(t) - (\tau^- + \tau^+)U_{i+1}^+(t) \ .\end{aligned}$$

These inequalities contradict the assumption $U_i^+(t) > U_{i+1}^+(t)$ for $t < t_0$. Other cases are considered similarly. △

Lemma 2 *Let* $(U_i^{+(1)}(t), U_i^{-(1)}(t))$, $(U_i^{+(2)}(t), U_i^{-(2)}(t))$, $i = 0, ..., K$, *be two solutions of Eqs. (15)- (19). If* $U_i^{\pm(1)}(0) \geq U_i^{\pm(2)}(0)$, $i = 1, ..., K$, $U_0^{\pm(1)}(t) = U_0^{\pm(2)}(t)$, $U_{K+1}^{\pm(1)}(t) \geq U_{K+1}^{\pm(2)}(t)$, $t \geq 0$, *then* $U_i^{\pm(1)}(t) \geq U_i^{\pm(2)}(t)$, $i = 1, ..., K$, *for* $t > 0$.

PROOF. It is sufficient again to consider the case where strict inequalities (for $i > 0$) take place at $t = 0$ and show that $U_i^{\pm(1)}(t) > U_i^{\pm(2)}(t)$, $i = 1, ..., K$, $U_0^{\pm(1)}(t) = U_0^{\pm(2)}(t)$, $t > 0$. The equality for $i = 0$ follows from (7). Suppose that strict inequalities hold for t, $0 \leq t < t_0$ and are broken at $t = t_0$. Let j, $j > 0$, be the largest or the smallest index with $U_j^{+(1)}(t_0) = U_j^{+(2)}(t_0)$ or $U_j^{-(1)}(t_0) = U_j^{-(2)}(t_0)$. Let, for example, the equality take place for $U_j^{+\cdot}$ where j is the largest index with the equality. It follows from (15) that $\dot{U}_j^{+(1)}(t_0) > \dot{U}_j^{+(2)}(t_0)$. This contradicts the assumption that $U_j^{+(1)}(t) > U_j^{+(2)}(t)$ for $t < t_0$. Other cases are considered similarly. △

We turn to the problem (5)- (10)

Lemma 3 *Let* $G = (G^+, G^-) \in \overline{\mathcal{U}}$. *Then there exists in* $\overline{\mathcal{U}}$ *a unique solution of problem This solution can be obtained as a limit as* $K \to \infty$ *of the solutions of (15)– (19) with* $U_{K+1}^{\pm}(t) = 0$.

If $G_i^{\pm(1)}(0) \geq G_i^{\pm(2)}(0)$, $i > 0$, $G_0^{\pm(1)}(0) = G_0^{\pm(2)}(0)$ *then*

$$U_i^{\pm(1)}(t) \geq U_i^{\pm(2)}(t), \ i > 0, \ t \geq 0. \tag{21}$$

PROOF. Denote by $\mathbf{U}^{(K)}(t)$, $K = 2, 3, ...$, the solution of (15)- (19) with $U_i^{\pm(K)}(t) = G_i^{\pm}$, $i = 0, ..., K$. By Lemma 1 $\mathbf{U}_i^{\pm(K)}(t)$ satisfy conditions (12), therefore $U_K^{\pm(K+1)}(t) \geq U_K^{\pm(K)}(t) = 0$. It follows from Lemma 2 that for any fixed t, $t > 0$ and for i, $i < K$, the values $U_i^{\pm(K)}(t)$ do not decrease as K increases, and $U_i^{\pm(K)}(t) \leq 1$. Therefore $\lim_{K\to\infty} U_i^{\pm(K)}(t) = U_i(t)$ exists and $\mathbf{U}(t) = \{(U_i^+(t), U_i^-(t))\}_{i=0}^{\infty} \in \overline{\mathcal{U}}$. Turning from differential equations to the integral ones we confirm that $U_i^{\pm}(t)$ satisfy Eqs. (5)– (9). The uniqueness of this solution in the class of function from $\overline{\mathcal{U}}$ can by proved by the Picard successive approximation method.

Lemma 4 *Let $G = (G^+, G^-) \in \mathcal{U}$ and let the non-overload condition (3) hold . Then* $\mathbf{U} \in \mathcal{U} \ \ \forall t \leq \infty$.

PROOF. Consider

$$\mathbf{V}(\mathbf{U}) = (V^+, V^-), \quad V^\pm = \{V_i^\pm\}_{i=0}^\infty, \quad V_k^\pm = \sum_{i=k}^\infty U_i^\pm, \quad V_k = V_k^+ + V_k^-. \tag{22}$$

If $U(0) \in \mathcal{U}$, then it follows from (9) that

$$\dot{V}_i(t) = \lambda U_{i-1}(t) U_{i-1}^+(t) - U_i(t), \tag{23}$$

$$\dot{V}_1(t) = \lambda U_0(t) U_0^+(t) - U_1(t) = \lambda U_0^+(t) - U_1(t) < \lambda,$$

thus $V_1(t) < \infty$ for $t < \infty$.

By (7) we have $lim_{t\to\infty} U_0^+(t) = U_0^+(\infty)$. Let $\lambda_0 < \lambda_1 < 1$, $\lambda_1 U_0^+(\infty) < 1$, and

$$T = \inf_{t \geq 0} \{t : \lambda U_0^+(t) \leq \lambda_1\}.$$

By (7), (11) we have $T < \infty$. Therefore $V_1(T) < \infty$. Further,

$$\dot{V}_i(t) = \lambda U_{i-1}(t) U_{i-1}^+(t) - U_i(t)$$

$$< \lambda_1 U_{i-1}(t) - U_i(t) = \lambda_1 (V_{i-1} - V_i) + V_{i+1} - V_i, \ t > T.$$

It follows from this differential inequality that

$$V_i(t) < V_i^{lin}(t), \quad t > T,$$

where $\mathbf{V}^{lin}(t) = \{V_i^{lin}(t)\}_{i=0}^\infty$ is the solution of a boundary value problem for a system of linear equations:

$$\dot{V}_i^{lin}(t) = \lambda_1 (V_{i-1}^{lin}(t) - V_i^{lin}(t)) + V_{i+1}^{lin}(t) - V_{i-1}^{lin}(t), \ i > 0, \ t \geq T, \tag{24}$$

$$V_0^{lin}(t) - V_1^{lin}(t) = 1, \quad V_i^{lin}(T) = V_i(T), \ i > 0.$$

The solution of (24) with $V_1^{lin}(T)$ bounded is bounded for all $t > T$, thus $\mathbf{U} \in \mathcal{U}$ (See, for example, Lemma 12 of [6].) $\triangle$

Lemma 5 *There exist in $\mathcal{U}$ a stationary solution $\mathbf{U}^{st}$ of problem (5)-(10).*

PROOF Consider the solution of (5)-(10) with the initial data

$$G_0^\pm = \frac{\tau^\pm}{\tau^+ + \tau^-}, \quad G_i^\pm = 0, \ i \geq 1.$$

It follows from Lemma 3 that the values $U_i^\pm(t), \ i \geq 1$, increase as t increases ($U_0^\pm(t) = U_0^\pm(0)$), therefore $\mathbf{U}(t) \to \mathbf{U}^{st}$ as $t \to \infty$, and by Lemma 4 $\mathbf{U}^{st} \in \mathcal{U}$. $\triangle$

Lemma 6 *Let* $(U_i^{+(1)}(t), U_i^{-(1)}(t))$, $(U_i^{+(2)}(t), U_i^{-(2)}(t))$, $i = 0, 1, ...,$ *be two solutions of Eqs. (5)- (10). If* $\mathbf{U}^{\pm(j)}(0) \in \mathcal{U}$, $j = 1, 2$, *then*

$$\lim_{t\to\infty} (U_i^{\pm(1)}(t) - U_i^{\pm(2)}(t)) = 0, \quad i = 1, 2, \ldots.$$

Proof. We use the following equality

$$x^{(1)}y^{(1)} - x^{(2)}y^{(2)} = \frac{1}{2}\Big((x^{(1)} - x^{(2)})y^{(2)} + (y^{(1)} - y^{(2)})x^{(1)} + (x^{(1)} - x^{(2)})y^{(1)} + (y^{(1)} - y^{(2)})x^{(2)}\Big).$$

Denote

$$\Delta_i^{\pm}(t) = U_i^{\pm(1)}(t) - U_i^{\pm(2)}(t).$$

The equations for $\Delta_i^{\pm}(t)$ are (we write $\Delta_i^{\pm}, U_{\cdot}^{\pm\cdot}$ instead of $\Delta_i^{\pm}(t), U_{\cdot}^{\pm\cdot}(t)$.)

$$\dot{\Delta}_i^{+} = -\tau^{-}\Delta_i^{+} + \tau^{+}\Delta_i^{-} + \Delta_{i+1}^{+} - \Delta_i^{+}$$

$$+\frac{\lambda}{2}\Big((\Delta_{i-1}^{+} - \Delta_i^{+})(U_{i-1}^{+(1)} + U_i^{+(1)}) + (\Delta_{i-1}^{+} + \Delta_i^{+})(U_{i-1}^{+(2)} - U_i^{+(2)})$$

$$+(\Delta_{i-1}^{+} - \Delta_i^{+})(U_{i-1}^{+(2)} + U_i^{+(2)}) + (\Delta_{i-1}^{+} + \Delta_i^{+})(U_{i-1}^{+(1)} - U_i^{+(1)})\Big)$$

$$+\frac{\lambda}{4}\Big((\Delta_{i-1}^{+} - \Delta_i^{+})(U_{i-1}^{-(1)} + U_i^{-(1)}) + (\Delta_{i-1}^{-} + \Delta_i^{-})(U_{i-1}^{+(2)} - U_i^{+(2)})$$

$$+(\Delta_{i-1}^{+} - \Delta_i^{+})(U_{i-1}^{-(2)} + U_i^{-(2)}) + (\Delta_{i-1}^{-} + \Delta_i^{-})(U_{i-1}^{+(1)} - U_i^{+(1)})\Big),$$

$$\dot{\Delta}_i^{-}(t) = \tau^{-}\Delta_i^{+} - \tau^{+}\Delta_i^{-} + \Delta_{i+1}^{-} - \Delta_i^{-}$$

$$+\frac{\lambda}{4}\Big((\Delta_{i-1}^{-} - \Delta_i^{-})(U_{i-1}^{+(1)} + U_i^{+(1)}) + (\Delta_{i-1}^{+} + \Delta_i^{+})(U_{i-1}^{-(2)} - U_i^{-(2)})$$

$$+(\Delta_{i-1}^{-} - \Delta_i^{-})(U_{i-1}^{+(2)} + U_i^{+(2)}) + (\Delta_{i-1}^{+} + \Delta_i^{+})(U_{i-1}^{-(1)} - U_i^{-(1)})\Big).$$

First, we prove that $|\Delta_1^{+}| + \Delta_1^{-}| \to 0$ as $t \to \infty$. Let us estimate $\sum_{i=1}^{\infty}(|\Delta_i^{+}(t)| + |\Delta_i^{-}(t)|),\ t > 0$. Consider $(\sum_{i=1}^{\infty}(|\dot{\Delta}_i^{+}| + |\Delta_i^{-}|))$ Following [9] we suppose first that all $\Delta_i^{\pm}$ are of the same sign, say, are positive. Then $\Delta_i^{\pm} = |\Delta_i^{\pm}|$ and

$$\begin{aligned}
\dot{\Big(\sum_{i=1}^{\infty} (|\Delta_i^+| + |\Delta_i^-|) \Big)} \quad = \quad & \lambda |(\Delta_0^+)|(3/2 + |U_1^{-(1)}| + |U_1^{-(2)}|) + |\Delta_0^-|))+ \\
& + \sum_{i=1}^{\infty} |\Delta_i^+| \Big(-1 - \tau^- - \tfrac{\lambda}{2}((U_{i-1}^{+(1)} + U_i^{+(1)}) + (U_{i-1}^{+(2)} + U_i^{+(2)}))+ \\
& +(U_{i-1}^{+(1)} - U_i^{+(1)}) + (U_{i-1}^{+(2)} - U_i^{+(2)})- \\
& -\tfrac{\lambda}{4}((U_{i-1}^{-(1)} + U_i^{-(1)}) + (U_{i-1}^{-(2)} + U_i^{-(2)}))+ \\
& + \Big[1 + \tau^- + \tfrac{\lambda}{2}((U_i^{+(1)} - U_{i+1}^{+(1)}) + (U_i^{+(2)} - U_{i+1}^{+(2)})+ \\
& +(U_i^{+(1)} + U_{i+1}^{+(1)}) + (U_i^{+(2)} + U_{i+1}^{+(2)}))+ \\
& +\tfrac{\lambda}{4}((U_{i-1}^{-(1)} - U_i^{-(1)}) + (U_{i-1}^{-(2)} - U_i^{-(2)}) + (U_i^{-(1)} - U_{i+1}^{-(1)})+ \\
& +(U_i^{-(2)} - U_{i+1}^{-(2)}) + (U_{i-1}^{-(1)} + U_i^{-(1)}) + (U_{i-1}^{-(2)} + U_i^{-(2)}))\Big)+ \\
& +|\Delta_i^-| \Big(-1 - \tau^+ - \tfrac{\lambda}{4}((U_{i-1}^{+(1)} + U_i^{+(1)}) + (U_{i-1}^{+(2)} + U_i^{+(2)})+ \\
& +1 + \tau^+ + \tfrac{\lambda}{4}(+(U_{i-1}^{+(1)} - U_i^{+(1)}) + (U_{i-1}^{+(2)} - U_i^{+(2)}) + (U_i^{+(1)} + U_{i+1}^{+(1)})+ \\
& +(U_i^{+(2)} + U_{i+1}^{+(2)}) + (U_i^{+(1)} - U_{i+1}^{+()}) + (U_i^{+(2)} - U_{i+1}^{+(2)}))\Big) = S
\end{aligned} \tag{25}$$

Here

$$S = \lambda[|\Delta_0^+|(3/2 + |U_1^{-(1)}| + |U_1^{-(2)}|) + |\Delta_0^-|/2)|] - |\Delta_1^+| - |\Delta_1^-|. \tag{26}$$

Note, that all summands in second and fourth square brackets are not negative. If some of the differences $\Delta_i^{\pm}$ are negative then the terms $-|\Delta_1^+| - |\Delta_1^-|$ and the summands in the first and third square brackets do not change (because from $\dot{y} = ay + b$ follows that $\dot{|y|} = a|y| \pm b$), but the terms with $\Delta_0^{\pm}$ and some of the summands in the second and fourth square brackets can change there signs. Because these terms change there signs from '+' to '-' in general case where $\Delta_i^{\pm}$ have different signs we get an inequality

$$S \leq \lambda[|\Delta_0^+|(3/2 + |U_1^{-(1)}| + |U_1^{-(2)}|) + |\Delta_0^-|/2] - |\Delta_1^+| - |\Delta_1^-|. \tag{27}$$

By (7),(9) we have that $|\Delta_0^{\pm}(t)| \to 0$ as $t \to \infty$, If $|\Delta_1^{\pm}|$ do not tend to 0 then by (26) the sum $\sum_{i=1}^{\infty}(|\Delta_i^+| + |\Delta_i^-|)$ will become negative. This is impossible, thus $|\Delta_1^{\pm}(t)| \to 0$ as $t \to \infty$. (Actually the integrals $\int_0^{\infty} |\Delta_1^{\pm}(t)| dt$ have to be bounded).

By induction on i we get that $|\Delta_i^{\pm}(t)| \to 0$ as $t \to \infty$. △

Lemma 7 *Any solution of (5)- (10) with $U^{\pm} \in \mathcal{U}$ tends to the unique stationary solution of (5)- (9).*

Proof The statement follows from Lemma 6. △

Lemma 8 *For the stationary solutions $\mathbf{U}^{st} \in \mathcal{U}$ the values of U_i^{st} decrease superexponentially as $i \to \infty$.*

PROOF For stationary solution we get from equation (23)

$$\lambda U_0^{+st} U_i^{st} \geq \lambda U_i^{st} U_i^{+st} = U_{i+1}^{st},$$

If $\lambda < 1$ the statement is obvious. Let us assume that $\lambda > 1$. As $\lambda U_0^{+st} < 1$ the values U_i decrease as i increases, and for some i_0 we have $(\lambda U_{i_0}^{st})^2 < 1$ and

$$(\lambda U_i^{st})^2 > U_{i+1}^{st}, \quad U_i^{st} \leq (\lambda U_{i_0}^{st})^{2^{i-i_0}}, \quad i > i_0.$$

That shows the superexponential decrease of U_i. △

For the proofs of Lemma 11 and Theorem 2 we need the following lemmas.

Lemma 9 (The Lemma is similar to Lemma 8 of [6] *) For the solutions of (5)- (10), for any T there exists such $C = C(T)$ that*

$$\frac{\partial U_k(t)}{\partial G_i} \leq C(T), \quad \frac{\partial^2 U_k(t)}{\partial G_i \partial G_j} \leq C(T), \quad i, j, k > 0, \quad t \leq T.$$

We omit the proof of this Lemma.

Lemma 10 (Lemma 4 from [6]) *For the solutions of (5)- (10) the estimates*

$$U_k(t) \leq \sum_{i=0}^{k} U_i(0)(\lambda t)^{k-i}/(k-i)!, \quad V_1(U(t)) \leq \exp(\lambda t) V_1(\mathbf{U}(0)).$$

hold.

PROOF Use induction on k. For $k = 0$ the estimates are obvious. Since $U_k^+ \leq 1$ we have by (9) that $\dot{U}_k \leq \lambda U_{k-1}$. This gives us the first inequality. The second inequality follows from the first one. △

3.2 Main theorems

The queueing system S_N determines the Markov process

$$U_N = U_N(t)$$

with state space $\mathbf{U} \in \overline{\mathcal{U}}$ given by the operator (4). Corresponding to the process U_N is the semigroup $T_N = T_N(t)$ on the function space $\mathcal{U}_N$. Namely, if $f : \mathcal{U}_N \to \mathbf{R}^1$, then

$$T_N(t) f(U) = (\mathbf{E} f(U_N(t)) \mid U_N(0) = G), \quad G \in \mathcal{U}_N. \tag{28}$$

Let $L = C(\overline{\mathcal{U}})$ be the Banach space of continuous functions $f : \overline{\mathcal{U}} \to \mathbf{R}^1$ with uniform metric $\| f \| = \max_{U \in \overline{\mathcal{U}}} | f(u) |$, and, similarly, let $L_N = C(\mathcal{U}_N)$. The inclusion $\mathcal{U}_N \subset \overline{\mathcal{U}}$ induces a contraction mapping $\Pi_N : L \to L_N$, $\Pi_N f(U) = f(U), \quad f \in L, \quad U \in \mathcal{U}_N$.

We say that $f \in L$ depends only on k first variables, if for any $U^{\pm(1)}, U^{\pm(2)} \in \overline{\mathcal{U}}$ from $U_i^{\pm(1)} = U_i^{\pm(2)},\ i = 1, ..., k,$ it follows that $f(U^{\pm(1)}) = f(U^{\pm(2)})$.

Proposition 1 (Proposition 2 of [6]) *The set of functions from L that depends on finite number of variables is dense in L.*

We omit the proof of this proposition.

The generator A_N of semigroup $T_N = T_N(t)$ is defined by formula (4).

A mapping $G \to U(t, G)$, where $U(t, G)$ is a solution of (5)- (10), defines in space $\overline{\mathcal{U}}$ a dynamical system with a corresponding semigroup $T = T(t)$ that acts in the space L. If $f \in L,\ G \in \overline{\mathcal{U}}$, then

$$T(t)f(G) = f(U(t, G)). \tag{29}$$

The semigroups T_N and T are strongly continuous and contractive (see, for example, ([2], ch. 1. n^0 1.1]).

Denote by A the generator of the semigroup T, and denote by $\mathcal{D}(A)$ the domain of its definition. It follows from (29) that if f is a function from L that has partial derivatives $\partial f / \partial U_i^{\pm} \in L$, and if $\sup_i \| \partial f / \partial U_i^{\pm} \| < \infty$, then $f \in \mathcal{D}(A)$ and

$$Af(U) = \sum_{i=1}^{\infty} \Big(\frac{\partial f}{\partial U_i^+} \dot{U}_i^+ + \frac{\partial f}{\partial U_i^-} \dot{U}_i^- \Big). \tag{30}$$

Here $\dot{U}_i^{\pm}$ is defined by (5)– (8).

Denote by D the set of all functions $f \in L$ that have derivatives $\partial f / \partial U_i^{\pm}$ and $\partial^2 f / \partial U_i \partial U_k$, for which there exists $C = C(f) < \infty$ such that

$$\sup_{i,U} | \frac{\partial f(U)}{\partial U_i^{\pm}} | \le C, \ \sup_{i,k,U} | \frac{\partial^2 f(U)}{\partial U_i^{\pm} \partial U_k^{\pm}} | \le C. \tag{31}$$

Lemma 11 (Lemma 15 from [6]) *The set D is a core for the operator A*

PROOF It is obvious that D is dense in L and $D \subset \mathcal{D}(A)$. Let D_0 be the set of functions from D that depend only on a finite number of variables. It follows from Proposition 1 that D_0 is dense in L. Therefore it is sufficient to show that for any t the operator $T(t)$ does not lead D_0 out of D. Select an arbitrary function $F \in D_0$ and let $f(G) = F(U(t, G)),\ G \in \overline{\mathcal{U}}$. It follows from Lemma 9 that f has partial derivatives $\partial f / \partial G_i, \partial^2 f / \partial G_i \partial G_k$ that satisfy conditions (31) for $t > 0$. Therefore $f \in D$. △

Theorem 1 *Let f be any continuous vector-function $f : \overline{\mathcal{U}} \to \mathbf{R}^1$. Then for all $t \geq 0$*

$$\lim_{N\to\infty} \sup_{G\in\mathcal{U}_N} \mid T_N(t)f(G) - f(U(t,G)) \mid= 0. \tag{32}$$

The convergence in (32) is uniform with respect to t on any finite interval of t.

PROOF. We use the known result that connects the convergence of semigroups with the convergence of their generators [2](ch. 1, Theorem 6.1]). By this result and by Lemma 11 it is sufficient to show that for any function $f \in D$

$$\lim_{N\to\infty} \sup_{U\in\mathcal{U}_N} \mid A_N f(U) - Af(U) \mid= 0. \tag{33}$$

Consider, for example, the terms of the first sum at the right sides of (4) and (5). We have for $i > 0$

$$N\Big(f(\mathbf{U}^+ + e_i^+/N) - f(\mathbf{U}^+) - \frac{\partial f(U)}{\partial U_i^+}\Big) = \frac{\vartheta_{1,i}}{N}\left(\frac{\partial^2 f(U+\vartheta_{2,i}e_k/N)}{\partial U_i^2}\right) \leq \frac{C}{N},$$

where $0 < \vartheta_{j,i} < 1,\ j = 1, 2$. Similar estimates are valid also for other terms in formulas (5)–(7). Therefore, after summation

$$\mid A_N f(U) - Af(U) \mid\leq \frac{C}{N}\sum_{i=1}^{\infty}\Big(4(U_i - U_{i+1}) + 2V_1(U)(\tau^+ + \tau^-)\Big).$$

That gives (33). △

Proposition 2 *Under the condition (3) for any $\mathbf{U}_N(t) = \{U_{k,N}(t)\}_{k=0}^{\infty}$ with $U_N(0) = G \in \mathcal{U}_N$ there exist such $T,\ N_0$ and $\lambda_1, \lambda_0 < \lambda_1 < 1$ that for $N \geq N_0,\ t \geq T$*

$$\lambda \mathbf{E}(U_{k,N}^+(t)U_{k,N}(t)) < \lambda_1 \mathbf{E}U_{k,N}(t). \tag{34}$$

PROOF. We have $U_{k,N}^+(t) \leq U_{0,N}^+ \leq 1$ and the variance of $U_{0,N}^+(t)$ tends to 0 as $N \to \infty$. Therefore for any $\epsilon > 0,\ \delta > 0$ there exist such $N_0,\ T$ that $\Pr(|U_{0,N}^+(t) - U_{0,N}^+(\infty)| > \delta) < \epsilon$ for $N > N_0,\ t > T$, and

$$\mathbf{E}(U_{k,N}^+(t)U_{k,N}(t)) \leq \mathbf{E}(U_{0,N}^+(t)U_{k,N}(t)|\ |U_{0,N}^+(t) - U_{0,N}^+(\infty)| < \delta) + \epsilon.$$

The needed estimate follows from this inequality. △

Lemma 12 *a) Let $U \in \mathcal{U}_N$, $N > N_0$, $V_k = V_k(U)$, where V_k is given by (22). Set*

$$W_{k,N}(t,U) = T_N(t)V_k(U), \tag{35}$$

and let $\mathbf{V}^{lin}(t) = \{V_k^{lin}(t)\}_{k=0}^{\infty}$ be a solution of (24) for $t > T$, $V_k^{lin}(T) = T_N(T)V_k$. Then

$$W_{k,N}(t,U) \leq V_k^{lin}(t), \quad k = 1,2,..., \quad t \geq T, \tag{36}$$

where T is defined in Proposition 2.

b) Let $U(t)$ be a solution of problem (5)- (10), $W_k(t) = V_k(U(t))$, $k = 0,1,...$, and let $V^{lin}(t)$ be a solution of (24) for $t > T_1$, $V_k^{lin}(T_1) = V_k(U(T_1))$, where T_1 if defined by the condition: $\lambda U^+(t) \leq \lambda_1$, $t \geq T_1$. Then

$$W_k(t) \leq V_k^{lin}(t), \quad k = 0,1,..., \quad t \geq T_1. \tag{37}$$

PROOF. a) Using formula (4), find the action of operator A_N on function $f_k^{\pm}(U) = U_k^{\pm}$, $f_k(U) = U_k$.

$$A_N f_k(U) = \lambda(U_{k-1}^+ U_{k-1} - U_k^+ U_k) + U_{k+1} - U_k, \quad k = 1,2,...$$

Since $V_k = \sum_{i=k}^{\infty} f_i$, we get

$$A_N V_k(U) = \lambda U_{k-1}^+ (U_{k-1} - U_k \tag{38}$$

It follows from (38, (34), and from Preposition 2 that

$$\dot{W}_{k,N}(t,U) \leq \lambda_1(W_{k-1,N}(t,U) - W_{k,N}(t,U)) + W_{k+1,N}(t,U) - W_{k,N}(t,U), \ t > T.$$

Besides, $W_{0,N}(t,U) - W_{1,N}(t,U) = T_N(t)(V_0(U) - V_1(U)) = 1$ and $W_{k,N}(T,U) = V_k(U(T))$. Therefore, $W_{k,N}$ is upperbounded by the solution of (24).

b) is proved in similar way. $\triangle$

Theorem 1 states the convergence of the mean value of any continuous function $\mathbf{E}f(U_N(t))$ to the function $f(U(t))$. In the next theorem we prove that $\mathbf{E}V(U_N(t))$ converges to $V(U(t))$ for any $t < \infty$. This fact needs a proof because $V(U)$ is not a continuous function on $U \in \overline{\mathcal{U}}$. Note, that $NV(U_N)$ is equal to the number of packets in the system, and that $V(U_N)$ is the mean number of packets per server.

Theorem 2 *Let $G \in \mathcal{U}$, and let the sequences $G_N = \{G_{Nj}\}_{j=0}^{\infty} \in \mathcal{U}_N$, $N = 1,2,...$, be such that $G_N \to G$ as $N \to \infty$. Let series $V_1(G_N) = \sum_{k=1}^{\infty} G_{Nk}$ converges to $V_1(G)$ uniformly with respect to N. Then*

$$\lim_{N\to\infty} T_N(t)V_1(G_N) = V_1(U(t,G)), \quad t \geq 0. \tag{39}$$

The convergence in (39) is uniform with respect to t on any finite interval of t.

PROOF. Fix $T_0 > 0$. Let $U(0) = G_N$, $W_{k,N}(t) = W_{k,N}(t, U)$ (see (36).) Represent functions $V_1(U(t,G))$ and $W_{1,N}(t)$ in the form

$$V_1(U(t,g)) = \sum_{k=1}^{n-1} U_k(t,G) + V_n(U(t,G)), \quad W_{1,N}(t) = \sum_{k=1}^{n-1} T_N(t) f_k(G_N) + W_{n,N}(t),$$

where $f_k(U) = U_k$. For any $n < \infty$ the needed convergence of sums $\sum_{k=1}^{n-1} \cdot$ is given by Theorem 1. To prove the theorem it is sufficient to show that for any $\epsilon > 0$ one can find a large n such that for all $t < T_0$ and for any N

$$V_n(U(t,G)) < \epsilon, \quad W_{n,N}(t) < \epsilon. \tag{40}$$

For the solution $U(t,G)$ of (5)– (10) and therefore for the first inequality in (40) the needed estimates are given by Lemma 9. The second inequality follows from Lemmas 10 and 11. $\triangle$

In Theorems 1, 2 the convergense of functions on $U_N(t)$ to functions on $U(t)$, $t < \infty$, was shown. In the next theorem we show the convergence of stationary distributions, namely, so to say, the convergence at $t = \infty$.

Theorem 3 *Let the condition (3) be valid. Then*
a) the process U_N, $N > N_0$ is ergodic, i.e. there exists an unique stationary probability measure such that for any initial distribution the distribution at a moment t converges to this measure as $t \to \infty$. Let $\mathbf{E}_N$ be the mean value with respect to the stationary measure of the process U_N . Then $\mathbf{E}_N V_1 < \infty$.
b) there exists on the set $\mathcal{U}$ a unique probability measure Π that is invariant with respect to the dynamic system $G \to U(t;G)$, $G \in \mathcal{U}$. This measure is concentrated at the fixed point $\Pi = \{\Pi_i\}_{i=0}^{\infty}$ of the dynamical system.
c)

$$\lim_{N\to\infty} \mathbf{E}_N U_i = \Pi_i,) \tag{41}$$

where Π_i decrease superexponentially with respect to i as $i \to \infty$.

PROOF
a) The process U_N is a Markov process, with a denumerable number of states $\mathcal{U}_N$, and all states are attainable. Therefore it is sufficient to show that for a function $V_1 : \mathcal{U} \to [0, \infty)$ the following statements are true:
1. Any subset of $U \in \mathcal{U}_N$ with bounded $V_1(U)$ is bounded.
2. For any $U \in \mathcal{U}_N$

$$\sup_{t\geq 0} T_N(t) V_1(U) < \infty. \tag{42}$$

Statement 1 is easily checked directly.
Statement 2 follows from Lemma 12.

b) is proved in Lemmas 5– 8.

c) Let μ_N be an invariant measure of the process U_N. The set $\overline{\mathcal{U}}$ is compact, therefore the set of probability measures on $\overline{\mathcal{U}}$ is also compact with respect to the weak convergence. It follows from Theorem 1 that any measure μ that is a fixed point for a sequence of measures μ_N, is invariant under the semigroup T. Therefore, by item b), it is sufficient for the proof of (41) to show that for the measure μ

$$\mu(\mathcal{U}) = 1. \tag{43}$$

By item a) $\mathbf{E}_N V_i < \infty,\ i = 1, 2, ...$ Therefore, in accordance to (34) for stationary distribution,

$$\lambda_1 \mathbf{E}_N U_{k-1} \geq \lambda \mathbf{E}_N (U_{k-1}^+ U_{k-1}) = \mathbf{E}_N U_k,\ \ k = 1, 2, ...$$

It follows from this inequality that $\mathbf{E}U_k \leq \lambda_1^k$ where $\mathbf{E}$ is the mean value with respect to the measure μ, and $\mathbf{E}V_1 < \infty$. That gives (43). The superexponential decrease was shown in Lemma 8. $\triangle$

3.3 Numerical estimate of the stationary distribution coefficients

It is possible to derive exact estimate of the probability distribution of the limiting stationary process. Indeed, some algebra gives:

$$\begin{array}{lcl} U_{k+1}^+ & = & \frac{\lambda}{2}(2U_k^+ + U_k^-)U_k^+ + U_1^+ - A_k \\ U_{k+1}^- & = & \frac{\lambda}{2}U_k^- U_k^+ - U_1^+ + A_k \end{array} \tag{44}$$

with

$$A_k = \sum_{i=1}^{i=k} \frac{\lambda}{2}(U_{i-1}^+ U_i^- - U_{i-1}^- U_i^+) - \tau^+ U_i^+ + \tau^- U_i^- \ .$$

Therefore provided that U_1^+ is known (and $U_1^- = \frac{\lambda\tau^+}{\tau^+ + \tau^-} - U_1^+$), it is possible to recursively derive on integer k every value for $U_k^\pm$. In this case every $U_k^\pm$ is formally a function $U_k^\pm(U_1^+)$ of U_1^+. Indeed $U_k^\pm$ are polynomials of degree $2^k - 1$ of U_1^+. In order to get rid of unknown U_1^+ it suffices to remark that U_1^+ is the only positive value such that

$$\lim_{k \to \infty} U_k^\pm(U_1^+) = 0 \tag{45}$$

Therefore U_1^+ is the only limit of the roots of $U_k^\pm$ in the interval $[0, 1]$. Furthermore since we know that $U_k^\pm$ is bounded by a super-exponential bound it is easy to get very acurate estimate of U_1^+ via the analysis of the neighborhood of the roots of the $U_k^\pm(x)$. Knowing an estimate of U_1^+ the estimates of the $U_k^\pm$ follow.

For example figure 1 display the values of U_1^+ versus λ computed for $(\tau^+, \tau^-) = (0.7, 0.3)$.

Knowing U_1^+ we can compute the sequence $U_k^\pm$. For example, for $\lambda = 2$ we obtain the following table:

k	0	1	2	3	4	5
U_k^+	0.3	0.227870	0.122257	0.033256	0.002326	0.000011
U_k^-	0.7	0.372130	0.151187	0.033605	0.002121	0.000010

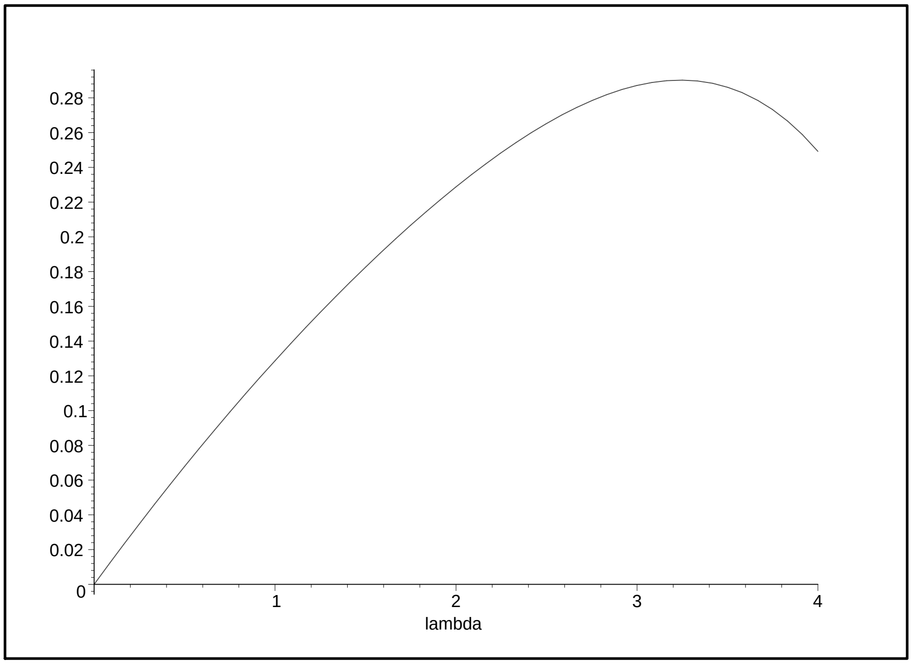

Figure 1: Quantity U_1^+ versus λ obtained via analytical model. The knowledge of U_1^+ allows the complete knowledge of the queue size distribution

4 Groups of servers with different $\tau^{\pm}$

4.1 Groups of servers without routing

In [11] inspired from [10] it is proven the following

Theorem 4 *The size Q of a queue under a single "on/off" source such that $\frac{\lambda\tau^+}{\tau^++\tau^-} < 1$, satisfies the following distribution:*

$$\Pr\{Q > k\} = \lambda \frac{\tau^+}{\tau^+ + \tau^-} \rho^{1-k} \tag{46}$$

with ρ being the largest root of equation $((z-1)+z\tau^+)((z-1)(1-z\lambda)+z\tau^-)-z^2\tau^+\tau^- = 0$.

If $\lambda > 1$ but still with $\frac{\lambda\tau^-}{\tau^++\tau^-} < 1$, which is now set as hypothesis, we have

$$\rho = \frac{1}{2}\left(\frac{1+\lambda+\tau^- + \tau^+ + \sqrt{(\lambda-1)^2 + 2(\tau^- + \tau^+) + 2(\tau^- - \tau^+)\lambda + (\tau^- + \tau^+)^2}}{\lambda + \lambda\tau^+}\right) \tag{47}$$

Consider a system with J groups of servers with different $\tau_j^{\pm}$, $j = 1, ..., J$. When a user from group j generates a packet it randomly selects a note from the j group.

Let $R_j = R_j^{(J)}$ denotes the number of nodes in the jth group ($\sum_j R_j = N$.) For every node in jth group we have $\tau^{\pm} = \tau_j^{\pm}$. Denote ρ_j the quantity ρ computed with quantities $\tau_j^{\pm}$. We therefore have

$$\Pr\{Q_U > k\} = \sum_j \frac{R_j}{N} \lambda \frac{\tau_j^-}{\tau_j^- + \tau_j^+} \rho_j^{1-k} \tag{48}$$

The aim of this section is to exhibit a sequence of quantities $(R_j, \tau_j^{\pm})$ such that the queue size distribution without routing shows polynomial (heavy) tails and such that, of course, queue size with routing has super-exponential tail.

For this end we introduce self-similar "on/off" sources where the pairs $\tau_j^{\pm}$ are proportional to a same vector $\tau^{\pm}$, *i.e.* $(\tau_j^+, \tau_j^-) = (\varepsilon_j\tau^+, \varepsilon_j\tau^-)$ for some ε_j. More precisely we select $R_j = \frac{j^{-\beta}}{\zeta(\beta)}N$ and $\varepsilon_j = j^{-\beta}$ for some $\beta > 1$. Function $\zeta(s) = \sum_j j^{-s}$, or Riemann *zeta* function, is a convergent series for $\Re(s) > 1$.

Theorem 5 *The distribution of the queue size Q_U without routing when groups setting satisfy $R_j = Nj^{-\beta}/\zeta(\beta)$ and $\varepsilon_j = j^{-\beta}$, for an arbitrary $\beta > 1$, satisfies the asymptotic expansion.*

$$\Pr\{Q_U > k\} = \mu k^{-1+1/\beta} + O(k^{-2+1/\beta}) \tag{49}$$

with

$$\mu = (\frac{\lambda\tau^+}{\tau^+ + \tau^-})\frac{\Gamma(1-1/\beta)}{\beta\zeta(\beta)}((\frac{\tau^-}{\lambda-1} - \tau^+))^{-1+1/\beta} \tag{50}$$

Proof: Let denote $f(x)$ the function $\sum_j \varepsilon_j \lambda \frac{\tau^+}{\tau^+ + \tau^-}\rho_j^{1-x}$, therefore $\Pr\{Q_U > k\} = f(k)$. Since $f(x)$ is an harmonic sum, as explained in [10], it is useful to make use of the Mellin transform: $f^*(s) = \int_0^\infty f(x)x^{s-1}dx$ and of its inverse:

$$f(x) = \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds \tag{51}$$

where c belongs to the definition set of $f^*(s)$. We get

$$f^*(s) = g(s)\Gamma(s) \tag{52}$$

where

$$g(s) = \sum_j (\log\rho_j)^{-s}\varepsilon_j\lambda\frac{\tau^+}{\tau^+ + \tau^-} \tag{53}$$

and $\Gamma(s)$ is Euler *Gamma* function.

Since the ε_j tend to zero, we have the asymptotic expansion

$$\begin{aligned}\rho_j &= 1 + (\frac{\tau^-}{\lambda-1} - \tau^+)\varepsilon_j + O(\varepsilon_j^2) \\ (\log\rho_j)^{-s} &= ((\frac{\tau^-}{\lambda-1} - \tau^+)^{-s}\varepsilon_j^{-s}(1 + sO(\varepsilon_j))\end{aligned}$$

Therefore

$$g(s) = (\frac{\lambda\tau^+}{\tau^+ + \tau^-})\frac{1}{\zeta(\beta)}((\frac{\tau^-}{\lambda-1} - \tau^+)^{-s}\zeta((1-s)\beta)) + O(\zeta(-s\Re((2-s)\beta)) \tag{54}$$

which basically means that $g(s)$ and $f^*(s)$ are defined for $\Re(s) \in]0, 1-\frac{1}{\beta}[$. Excepted the pole with residue $1/\beta$ of $\zeta((1-s)\beta)$ at $s = 1 - 1/\beta$, $f^*(s)$ can be analytically continued for $\Re(s) \in]0, 2-\frac{1}{\beta}[$. Since the expansion of ρ_j can be arbitrarily continued in power of ε_j, the error term on the right-hand side can be expanded into a series of factors $\zeta((k-s)\beta)$ for $k \geq 2$. Therefore $g(s)$ can be analytically contined on $\Re(s) \in]0, +\infty$ with poles on $s = k - 1/\beta$ for $k \geq 1$.

Using the inverse Mellin transform it comes:

$$f(x) = \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} g(s)\Gamma(s)x^{-s}ds \tag{55}$$

for $c \in]0, 1-1/\beta[$. By moving the integration line on the right of the first pole of $g(s)$ at $s = 1 - 1/\beta$, and applying residue theorems:

$$f(x) = (\frac{\lambda\tau^+}{\tau^+ + \tau^-})\frac{\Gamma(1-1/\beta)}{\beta\zeta(\beta)}((\frac{\tau^-}{\lambda - 1} - \tau^+)x)^{-1+1/\beta} + \frac{1}{2i\pi}\int_{c_2 - i\infty}^{c_2 + i\infty} g(s)\Gamma(s)x^{-s}ds \quad (56)$$

for any $c_2 \in]1-1/\beta, 2-1/\beta[$. The integral on the right-hand side is of order $O(x^{-c_2})$ which can be arbitrarily close to order $O(x^{-2+1/\beta})$. By applying a second time the residu theorem on the second pole of $g(s)$ at $s = 2 - 1/\beta$ we obtain an exact determination of the error term which is $O(x^{-2+1/\beta})$. $\triangle$

4.2 Groups of servers with routing

Consider a queueing system S_{NJ} with N nodes. On each node there is a user, a server and a FCFS queu. As in the last subsection there are J groups of servers, each with its parameters $\tau_j^{\pm,J}$, $j = 1, ..., J$, $N \gg J$, of Markovian transition from state "on" to state "off". In state "on" the user generates the Poisson flow of packets of rate λ. The service time of a packet is distributed exponentially with meat one.

The number of servers in jth group is $R_j^{(J)}$, $\sum_{j=1}^J R_j^{(J)} = N$, $R_j^{(J)}$, $R^{(J)}$ are the same as in the last subsection.

The routing protocol:

1. when a user from jth group generates a packet it randomly selects a remote note, all N nodes are i.i.e. distributed;
2. if the queue at the remote node is shorter than the local queue in jth group, the packet is directed to the queue of the remote node, if the remote queue is greater than the local queue, the packet is directed to the local queue;
3. if both remote and local queues are equal then the packet is randomly directed to the local queue with probability 1/2 or to the remote queue with probability 1/2.

Let $r_{j,k}^{\pm,J}$ be the number of servers from the jth group in modes "on" and "off" where the queue length is equal to k, and let $R_{j,k}^{\pm,J}$ be the number of servers from the jth group where the queue length is not less then k, $R_j^{(J)} = R_{j,0}^{+,J} + R_{j,0}^{-,J}$. Denote

$$U_{j,k}^{\pm,J} = \sum_{i=0}^{k} r_{j,i}^{\pm,J}/N = R_{j,k}^{\pm,J}, \quad (57)$$

$$U_k^{\pm,J} = \sum_{j=1}^{J} U_{j,k}^{\pm,J}, \quad (58)$$

$$U_{j,k}^{J} = U_{j,k}^{+,J} + U_{j,k}^{-,J}. \tag{59}$$

Further we are interested in case $J \to \infty$ and in the limit case $J = \infty$.

In case of finite N, J the system S_{NJ} can be described in terms of a Markov chain. The generating operator for this chain can be easily written, its form is similar to (4) and we do not present it here.

In the limit $N = \infty, J \le \infty$ we get the differential equations for $\mathbf{U}_{i,j}^{\pm},\ j = 1, ...J$. This equations are (to ease the notations the upper index 'J' is omitted):

$$\begin{aligned} \dot{U}_{j,i}^{+}(t) &= \lambda(U_{j,i-1}^{+}(t) - U_{j,i}^{+}(t))[U_{i-1}^{+}(t) + U_{i}^{+}(t) + U_{i-1}(t) + U_{i}(t)]/2 + \\ &+ U_{j,i+1}^{+}(t) - U_{j,i}^{+}(t) + \tau_j^{+} U_{j,i}^{-}(t) - \tau_j^{-} U_{j,i}^{+}(t), \quad i \ge 1, \end{aligned} \tag{60}$$

$$\begin{aligned} \dot{U}_{j,i}^{-}(t) &= \lambda(U_{j,i-1}^{-}(t) - U_{j,i}^{-}(t))(U_{i-1}^{+}(t) + U_{i}^{+}(t))/2 + U_{j,i+1}^{-}(t) - U_{j,i}^{-}(t) - \\ &- \tau_j^{+} U_{j,i}^{-}(t) + \tau_j^{-} U_{j,i}^{+}(t), \quad i \ge 1, \end{aligned} \tag{61}$$

$$\dot{U}_{j,0}^{+}(t) = \tau_j^{+} U_{j,0}^{-}(t) - \tau_j^{-} U_{j,0}^{+}(t), \quad \dot{U}_{j,0}^{-}(t) = -\tau_j^{+} U_{j,0}^{-}(t) + \tau_j^{-} U_{j,0}^{+}(t). \tag{62}$$

$$U_{j,0}^{+}(t) + U_{j,0}^{-}(t) \equiv R_j/N, \quad \sum_{j=1}^{J} (U_{j,0}^{+}(t) + U_{j,0}^{-}(t)) = 1, \tag{63}$$

$$\dot{U}_i = \lambda \Big(U_{i-1}(t) U_{i-1}^{+}(t) - U_i(t) U_i^{+}(t) \Big) + U_{i+1}(t) - U_i(t), \tag{64}$$

$$U_{j,i}^{\pm}(0) = G_{j,i}^{\pm}, \quad G_{j,i}^{\pm} \ge G_{j,i+1}^{\pm} \ge 0, \quad i \ge 0, \quad \sum_{j} (G_{j,0}^{+} + G_{j,0}^{-}) = 1. \tag{65}$$

We suppose the nonoverload condition

$$\lambda \sum_{j=1}^{J} \frac{R_j^{(J)} \tau_j^{+}}{N(\tau_j^{-} + \tau_j^{-})} = \lambda \sum_{j=1}^{J} U_j^{+}(\infty) < \lambda_0 < 1. \tag{66}$$

Note that equation (64) is the same as (9). This is an important property of the system 60, it helps to prove the statements similar to the ones proved for the system (5)-(10).

We introduce the space $\overline{\mathcal{U}}$ of sequences $\mathbf{U} = \{U_{j,i}^{+}, U_{j,i}^{-}\}_{j=1,i=0}^{J,\infty}$:

$$\overline{\mathcal{U}} : \ \{\mathbf{U}, \ \ 1 \ge U_{j,i}^{+} \ge U_{j,i+1}^{+} \ge 0, \ \ 1 \ge U_{j,i}^{-} \ge U_{j,i+1}^{-} \ge 0, \ \ j = 1, \ldots, J, \ i = 0, 1, ...\}, \tag{67}$$

with the norm

$$\rho\Big(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}\Big) = \sup_{1 \le j \le J, \ i \ge 0} \frac{| U_{j,i}^{\pm(1)} - U_{j,i}^{\pm(2)} |}{j(i+1)}, \quad \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \in \overline{\mathcal{U}},$$

and the spaces

$$\mathcal{U} : \{\mathbf{U} \in \overline{\mathcal{U}}, \ \sum_{j=1}^{J} \sum_{i=1}^{\infty} U_{j,i} < \infty\}, \tag{68}$$

$$\mathcal{U}_{NJ}: \{\mathbf{U}_{NJ} \in \mathcal{U}, \quad U^{\pm,J}_{Nj,k} = R^{\pm,J}_{j,k}/N\}, \tag{69}$$

It is not difficult to prove that in case $J < \infty$ all statements of the sections 2 and 3 are valid.

Here we will be mostly interested in the case where $J \to \infty, \quad N \to \infty, N \gg J$.

Suppose that

1. as $N \to \infty, J \to \infty$ for any $j,\ j < \infty$, the ratio of servers of jth group, $U^J_{j,0} = R^{(J)}_j/N$, converge to $U^{(\infty)}_{j,0}(= \lim_{J\to\infty, N\to\infty} R^{(J)}_j/N$ uniformly with respect to N, J.
2. as $J \to \infty$ the parameters $\tau^{\pm}_j = \tau^{\pm,J}_j$ converge to $\tau^{\pm,\infty}_j$ uniformly with respect to N, J. (In particular, it may happen that $\tau^{\pm,\infty}_j \to 0$ as $j \to \infty$.)
3. the nonoverload condition takes place

$$\lim_{J\to\infty, N\to\infty} \lambda \sum_{j=1}^{J} \frac{R^{(J)}\tau^{-,J}_j}{N(\tau^{-,J}_j + \tau^{+,J}_j)} = \lambda \sum_{j=1}^{\infty} U^{+,\infty}_{j,0}(\infty) = \lambda_0 < 1. \tag{70}$$

4. the sequences $\mathbf{G}^{\pm(J)}_N = \{\mathbf{G}^{\pm(J)}_{Njk}\}^{J,\infty}_{j=1,k=0} \in \mathcal{U}$, are such that $G^{\pm(J)}_{Njk} \to G^{\pm(\infty)}_{jk}$ as $J \to \infty,\ N \to \infty$ uniformly with respect to J, N, and $\mathbf{G}^{\pm(\infty)} \in \mathcal{U}$.

Proposition 3 *There exsist such $J_0 < \infty,\ N_0 < \infty,\ T < \infty$ and $\lambda_1,\ \lambda_0 < \lambda_1 < 1$ that $\forall$ $J > J_0,\ N > N_0,\ t > T$*

$$\lambda\Big(\sum_{j=1}^{J_0} U^{+,J}_{j,0}(t) + \sum_{j=J_0+1}^{J} U^{t}_{j,0}(\infty)\Big) < \lambda_1. \tag{71}$$

Further we list the statements similar to the statements of sections 3, 4.

Lemma 13 *Let $J \le \infty,\quad \mathbf{G} = \mathbf{G}^\infty = (G^+, G^-) \in \overline{\mathcal{U}}$. Then there exists in $\overline{\mathcal{U}}$ a unique solution $\mathbf{U}(t) = \mathbf{U}(t, G)$ of problem (60)- (65). If $G^{\pm(1)}_{j,i}(0) \ge G^{\pm(2)}_{j,i}(0)\ \ i > 0,\ G^{\pm(1)}_0(0) = G^{\pm(2)}_0(0)$ then*

$$U^{\pm(1)}_{j,i}(t) \ge U^{\pm(2)}_{j,i}(t),\ i > 0,\ t \ge 0. \tag{72}$$

Lemma 14 *Let $J \le \infty,\ \mathbf{G}^\infty = (G^{+,\infty}, G^{-,\infty}) \in \mathcal{U}$ and let the non-overload condition (71) hold. Then a solution of problem (60) – (65) $\mathbf{U} \in \mathcal{U},\ \ \forall t \le \infty$.*

PROOF.(Compare with the proof of Lemma 4.) Consider

$$\mathbf{V}(\mathbf{U}) = (V^+, V^-), \quad V_k^{\pm} = \sum_{j=1}^{\infty}\sum_{i=k}^{\infty} U_{j,i}^{\pm}, \quad V_k = V_k^+ + V_k^-. \tag{73}$$

If $U(0) \in \mathcal{U}$, then

$$\dot{V}_1(t) = \lambda U_0(t) U_0^+(t) - U_1(t) = \lambda U_0^+(t) - U_1(t) < \lambda,$$

thus $V_1(t) < \infty$ for $t < \infty$. Let λ_2 be such that $\lambda_1 < \lambda_2 < 1$ and let

$$T = (T(\lambda_2)) = \inf_{t \geq 0}\{t : \lambda U_0^+(t) \leq \lambda_2\}.$$

Because of (71) we have $T = T(J_0) < \infty$ and therefore $V_1(T) < \infty$. Further, $V_1(t),\ t > T$ is upperbounded by the bounded solution of a boundary value problem

$$\dot{V}_i^{lin}(t) = \lambda_2(V_{i-1}^{lin}(t) - V_i^{lin}(t)) + V_{i+1}^{lin}(t) - V_{i-1}^{lin}(t),\ i > 0,\ t \geq T, \tag{74}$$

$$V_0^{lin}(t) - V_1^{lin}(t) = 1, \quad V_i^{lin}(T) = V_i(T),\ i > 0.\ \triangle \tag{75}$$

Lemma 15 *Let $J \leq \infty$. There exist a stationary solution $|U^{st} \in \mathcal{U}$ for the differential exuations (ref(5,j))- (65).*

Lemma 16 *Let $J \leq \infty$, $(U_i^{+(1)}(t), U_i^{-(1)}(t))$, $(U_i^{+(2)}(t), U_i^{-(2)}(t))$, $i = 0, 1, ...,$ be two solutions of Eqs. (60)- (65). If $\mathbf{U}^{\pm(r)}(0) \in \mathcal{U}$, $r = 1, 2$, then*

$$\lim_{t\to\infty}(U_i^{\pm(1)}(t) - U_i^{\pm(2)}(t)) = 0, \quad i = 1, ...$$

Lemma 17 *Any solution of (60)-(65) with $\mathbf{U}^{\pm} \in \mathcal{U}$, $j \leq \infty$ tends to the stationary solution of (60)-(65).*

Lemma 18 *The values of stationary solution $U_i^{st} = \sum_{j=1}^{\infty} U_{j,i}^{st}$ of (60)- (64) decrease superexponentially as $i \to \infty$.*

Lemma 19 *For the solutions of (60)- (65) the estimates*

$$U_k(t) \leq \sum_{i=0}^{k} U_i(0)(\lambda t)^{k-i}/(k-i)!, \quad V_1(U(t)) \leq \exp(\lambda t) V_1(\mathbf{U}(0)).$$

hold.

Lemma 20 *For the solutions of (60)- (65) for any T there exists such $C = C(T)$ that*

$$\frac{\partial U_k(t)}{\partial G_i} \leq C(T), \quad \frac{\partial^2 U_k(t)}{\partial G_i \partial G_j} \leq C(T), \quad i,j,k > 0, \quad t \leq T.$$

For finite N, J the system S_{NJ} determines the Markov process $U_{NJ} = U_{NJ}(t)$. Corresponding to this process is the semigroup $T_{NJ} = T_{NJ}(t)$, namely, if $f : \mathcal{U}_{NJ} \to \mathbf{R}^1$ then

$$T_{NJ}(t) f(U) = (\mathbf{E} f(U_{NJ}(t)) \mid U_{NJ}(0) = G), \quad G \in \mathcal{U}_{NJ}.$$

Proposition 4 *Under the condition (71) for any $\mathbf{U}^J(0) \in \overline{\mathcal{U}}$ there exist such T, N_0 and $\lambda_2, \lambda_1 < \lambda_2 < 1$ that for $N \geq N_0$, $t \geq T$*

$$\lambda \mathbf{E}(U_{k,N}^{+J}(t) U_{k,N}^J(t)) < \lambda_2 \mathbf{E} U_{k,N}^J(t). \tag{76}$$

Lemma 21 *a) Let $U \in \mathcal{U}_{NJ}$, $N > N_0$, $J > J_0$, $V_k = V_k(U)$. Set*

$$W_{k,N}^{(J)}(t,U) = T_{NJ}(t) V_k(U), \tag{77}$$

and let $\mathbf{V}^{lin}(t) = \{V_k^{lin}(t)\}_{k=0}^{\infty}$ be a solution of (74) for $t > T$, $V_k^{lin}(T) = T_N(T) U_k$, where T is defined by Proposition **??**. *Then*

$$W_{k,N}^{(J)}(t,U) \leq V_k^{lin}(t), \quad k = 1, 2, ..., \quad t \geq T, \tag{78}$$

where T is defined in Proposition 4.

b) Let $U(t)$ be a solution of problem (60)- (65), $W_k(t) = V_k(U(t))$, $k = 0, 1, ...,$ and let $V^{lin}(t)$ be a solution of (74) for $t > T_2$, $V_k^{lin}(T_1) = V_k(U(T_1))$, where T_2 if defined by the condition (76). Then

$$W_k(t) \leq V_k^{lin}(t), \quad k = 0, 1, ..., \quad t \geq T_1. \tag{79}$$

Theorem 6 *Let $N, J < \infty$ and let f be any continuous vector-function $f : \overline{\mathcal{U}} \to \mathbf{R}^1$. Then for all $t \geq 0$*

$$\lim_{N\to\infty,\ J\to\infty} \sup_{G\in\mathcal{U}_{NJ}} \mid T_{NJ}(t)f(G) - f(U(t,G)) \mid = 0. \tag{80}$$

The convergence in (80) is uniform with respect to t on any finite interval of t.

PROOF follows the proof of Theorem 1.

Theorem 7 *Let $N, J < \infty$. Let series $V_1(G_N^{(J)}) = \sum_{j=1}^{J}\sum_{k=1}^{N} G_{Nk}^{(J)}$ converges to $V_1(G^\infty)$ uniformly with respect to N, J. Then*

$$\lim_{N\to\infty} T_{NJ}(t)V_1(G_N^{(J)}) = V_1(U(t,G)), \quad t \geq 0. \tag{81}$$

The convergence in (81) is uniform with respect to t on any finite interval of t.

Theorem 8 *Let the condition (71) be valid. Then*

a) the process U_{NJ}, $N > N_0, J > J_0$, $N, J < \infty$, is ergodic, i.e. there exists an unique stationary probability measure such that for any initial distribution the distribution at a moment t converges to this measure as $t \to \infty$. Let $\mathbf{E}_{NJ}$ be the mean value with respect to the stationary measure of the process U_{NJ} . Then $\mathbf{E}_{NJ}V_1 < \infty$.

b) for $N, J = \infty$ there exists on the set $\mathcal{U}$ a unique probability measure Π that is invariant with respect to the dynamic system $G \to U(t;G)$, $G \in \mathcal{U}$. This measure is concentrated at the fixed point $\Pi = \{\Pi_k\}_{k=0}^{\infty}$ of the dynamical system.

c)

$$\lim_{N\to\infty,\ J\to\infty} \mathbf{E}_{NJ}U_i = \Pi_i,) \tag{82}$$

where Π_i decrease superexponentially with respect to i as $i \to \infty$.

Below we present the simulation results to compare the performance of the systems with and without routing. The simulations demonstrate that in case of finite N the tails of queue length distribution in cases of routing are really much "lighter" then in cases without routing. The simulation is done on rather long period of time (800,000 time units), and queue length are sampled 800 times during this period. We first present the case where all sources are identical, *i.e.* carry the same on/off parameters (τ^+, τ^-) and λ, for instance respectively $(0.7, 0.3)$ and 2.0. Figure 2 and 3 respectively give the queue length distribution in the case respectively without routing and with routing when $N = 5$. Notice the dramatic change in tail distribution significant even with such a small value of N. In the case without routing we have compared the actual results obtained by simulation and the analytical model discribed in theorem 4.

Figure 4 displays the queue size distribution when sources are identical and $N = 50$, and routing is applied. Notice that there is no need to display the case without routing since it will be exactly the same as with $N = 5$ (figure 2): the queues don't interact therefore it does not make any change to simulate 50 instead of 5.

Figure 5 is interesting because it shows the same results of figure 4 but compared to the analytical model explained in the subsection 3.3. The matching is very good since it holds within few 0.001 range unit, better than what could be expected from 800 sampling (over 100 sources).

Next figures are related to the interesting case of several group of servers with different (τ^+, τ^-) vectors (but with still same peak rate $\lambda = 2.0$). Figure 6 displays the vectors (τ^+, τ^-) versus source index. We choose the case where sources are self-similar, *i.e.* vectors (τ^+, τ^-) are colinear, for instance $(\tau^+, \tau^-) = (0.7\varepsilon_j, 0.3\varepsilon_j)$ with $\varepsilon_j = j^{-\beta}$, j being the index of the group. The number of sources in group j is the closest integer to $Nj^{-\beta}/\zeta(\beta)$. We choose $\beta = 2$, (and therefore $\zeta(\beta) = \frac{\pi^2}{6}$.

Figure 7 displays the queue size distribution in the case of self-similar sources with $N = 100$ and without routing. With the actual results we also display the results obtained by the analytical model of theorem 4. The simulation confirms the polynomial tail behaviour of the queue length without routing.

Figure 8 shows the queue size distribution in the case of self-similar sources with $N = 100$ and with routing. The dramatic change from sub-exponential to super-exponential tail is evident from figure 7.

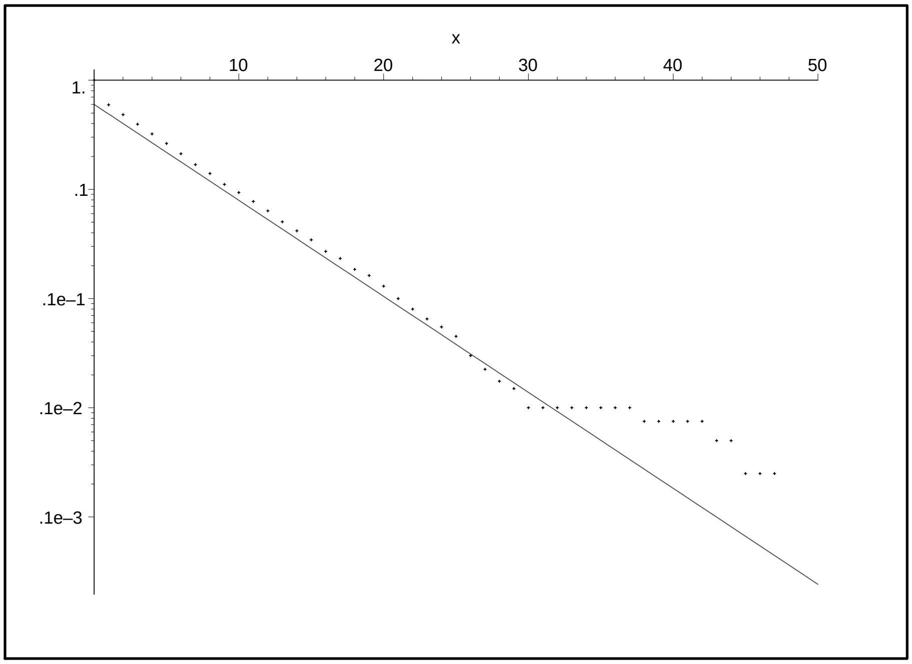

Figure 2: Queue length distribution: No routing case $N = 5$, $\lambda = 2$, same $(\tau^+, \tau^-) = (0.7, 0.3)$ on/off parameter, dots obtained by simulation, continuous line obtained via the analytical model

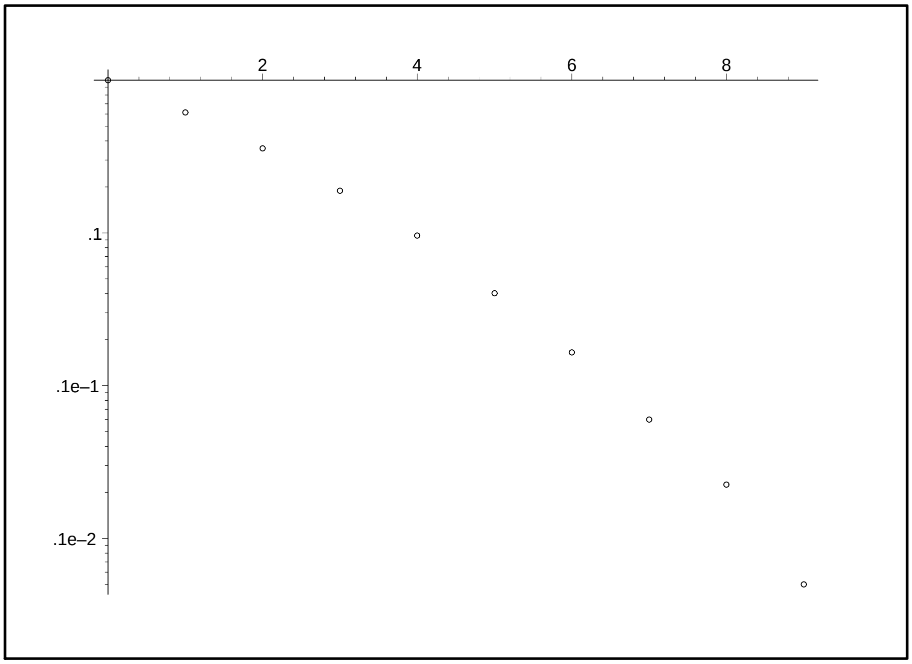

Figure 3: Queue length distribution: With routing case $N = 5$, $\lambda = 2$, same $(\tau^+, \tau^-) = (0.7, 0.3)$ on/off parameter

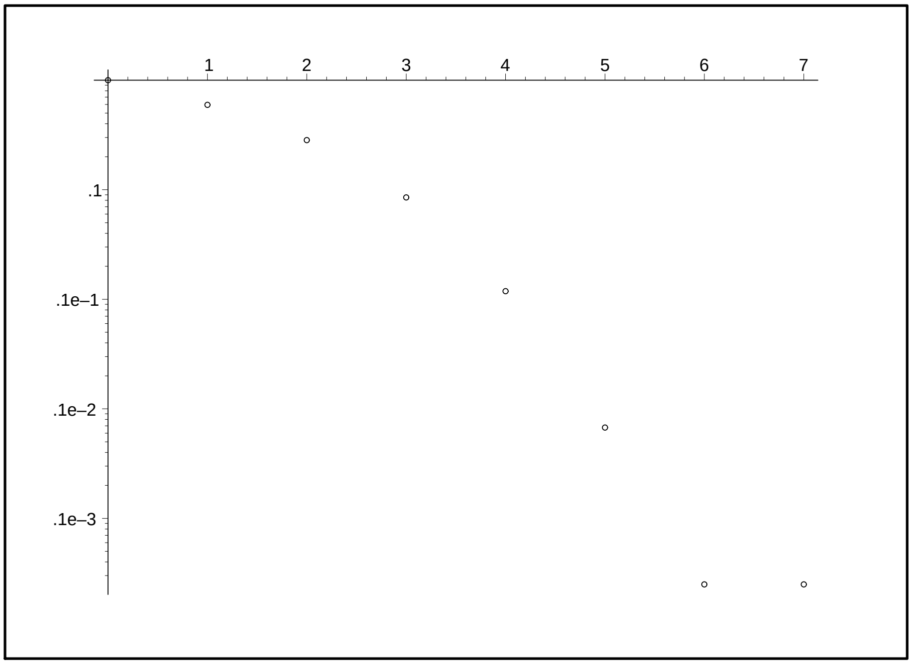

Figure 4: Queue length distribution: With routing case $N = 50$, $\lambda = 2$, same $(\tau^+, \tau^-) = (0.7, 0.3)$ on/off parameter, dots obtained via simulation

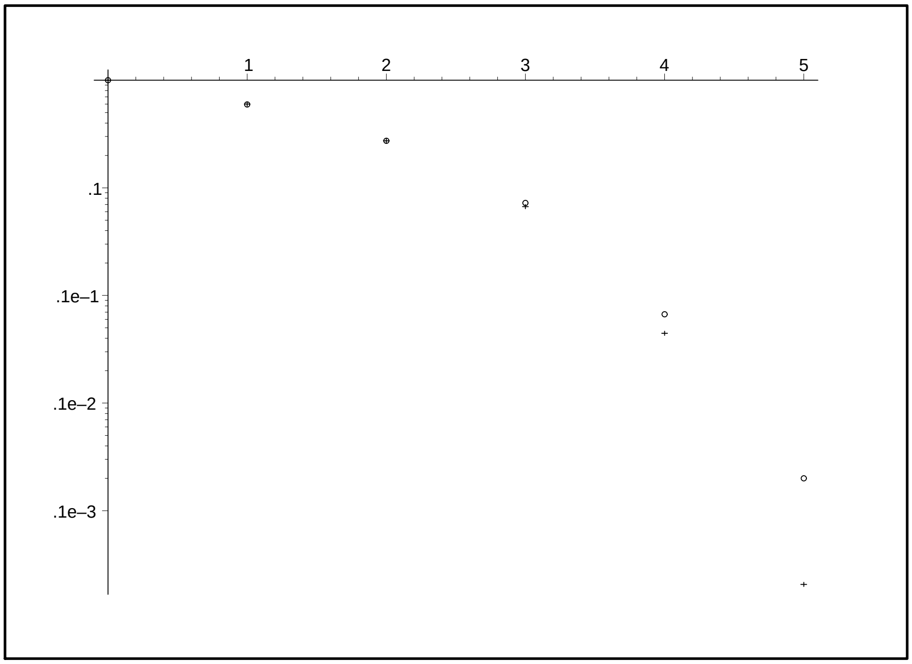

Figure 5: Queue length distribution: With routing case $N = 50$, $\lambda = 2$, same $(\tau^+, \tau^-) = (0.7, 0.3)$ on/off parameter, dots obtained via simulation, crosses obtained via an analytical model

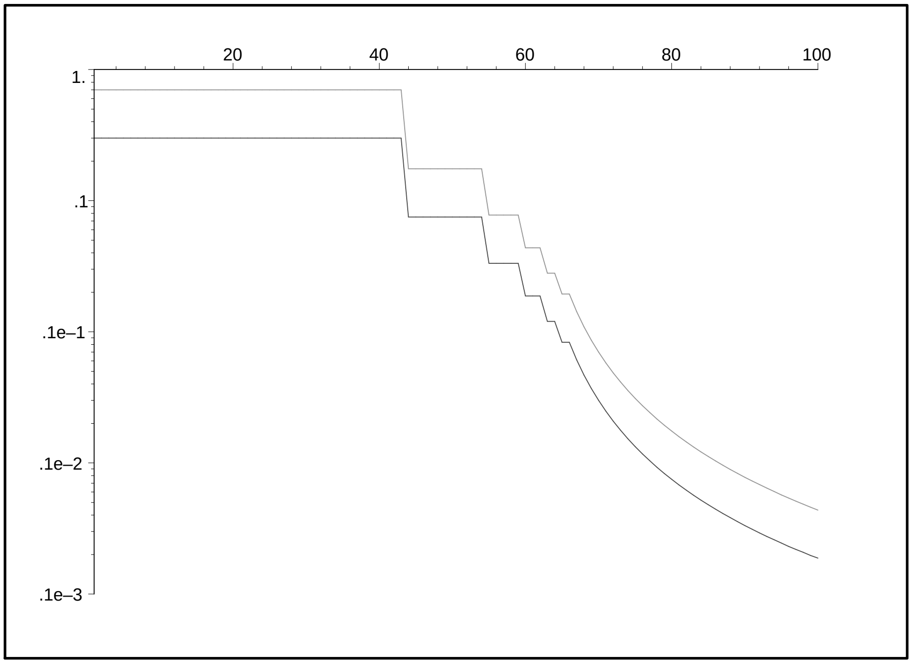

Figure 6: distribution of τ^+ (grey) and τ^- versus server: self-similar group of on/off sources, $N = 100$, $\lambda = 2$, same $\varepsilon_j = j^{-\beta}$ with $\beta = 2$.

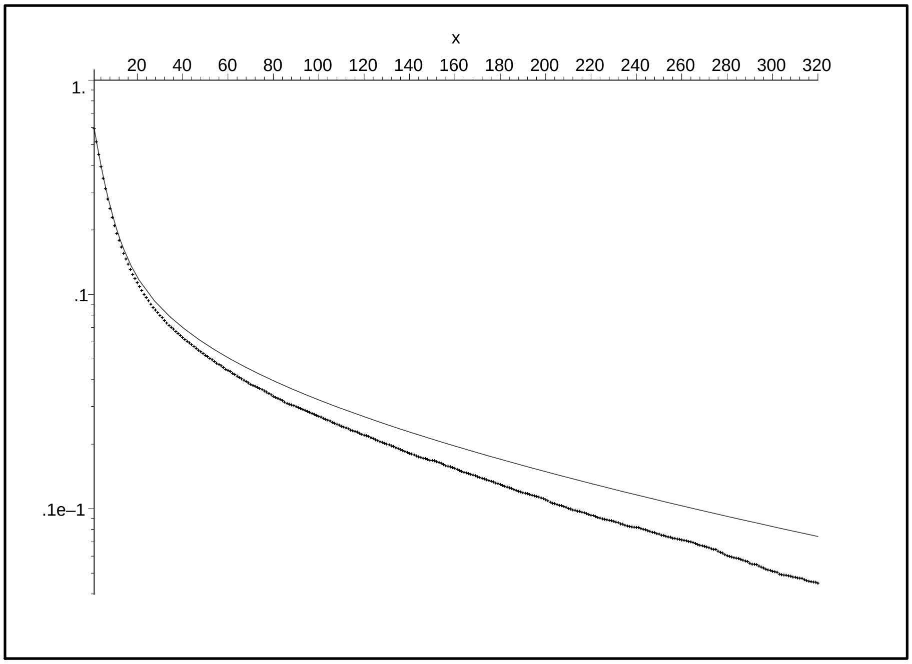

Figure 7: Queue length distribution: No routing case $N = 100$, $\lambda = 2$, self-similar sources $\beta = 2$, dots obtained by simulation, continuous line obtained via the analytical model

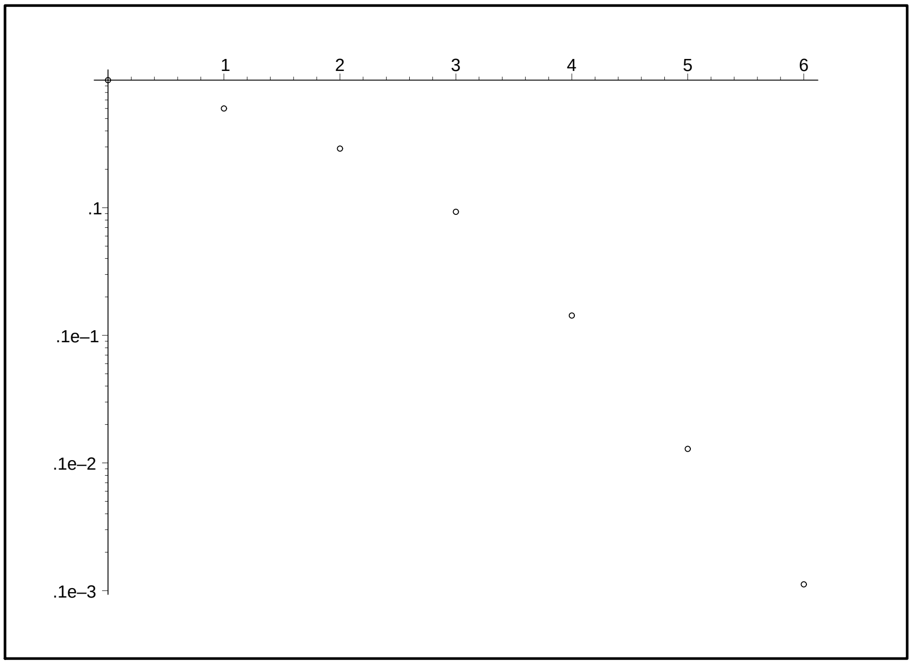

Figure 8: Queue length distribution: With routing case $N = 100$, $\lambda = 2$, self-similar sources $\beta = 2$, dots obtained by simulation

References

[1] Dobrushin, R. L. Switching Networks, Gibbson Fields – Interconnections, *Proc. of the 1-st World Congress of the Bernoulli Society*, Tashkent (1986),VNS sci. press, Utrecht, 1987, **1**, pp. 377–393 (1987).

[2] Ethierand, S. N., and Kurtz T. G. Markov Processes Characterization and Convergence,*N.Y. etc: John Willey and Stons*, (1986).

[3] Vvedenskaya, N. D. Distribution of Message Delay in a Network with many multiple routes," *Proc. 16-th Symposium on Inf. Theory in Benelux*, Enschede, (1995), pp.49-51.

[4] Vvedenskaya, N. D. The Delay in a Network with many Multiple Routes and Feedback, *Proc. 7-th Joint Swedish–Russian Int. Workshop Inf. Theory*, St. Petersburg (1995),pp. 243-245).

[5] Vvedenskaya, N. D.Large queueing system where selection of the shortest of several queues is allowed, In *S'eminaire sur la m'ecanique statistique des grands r'eseaux* (INRIA Roconquecourt, France, 1996), pp. 19-21.

[6] Vvedenskaya, N. D., Dobrushin, R. L., and Karpelevich F. I., Queueing Systems with Selection of the Shortest of Two Queues: an Asymptotic Approach, *Problems of Information Transmission*, 1996, v.32,No 1, pp. 15-27

[7] Vvedenskaya, N. D., and Suhov, Yu. M. Doblushin's Mean-Field Approximation for a Queue with Dynamic Routing,*Rapport de Recherche No 3328*, INRIA, Versaille-Roquencourt, 1997, pp 1-37, and *Markov Processes and Related Fields*, 1997, v.3, No 4, pp

[8] Turner, R. E. The Effect of Increasing Routing Choice on Resource Pooling, *Probability in the Engineering and Informational Sciences*, 1998, v.10, No 1.

[9] Mitzenmacher, M. The Power of Two Choices in Randomized Load Balancing, *PhD thesis*, University of California, Berkeley, 1996.

[10] P. Jacquet, Analytic information theory in service of queueing with aggregated exponential on/off arrivals, *in proc. 35th annual Allerton conf. on Communication, Control and Computing*, pp. 242-251, 1997.

[11] P. Jacquet, Long term dependences and heavy tails in traffics and queues generated by memoriless on/off sources in series, 1998, INRIA Research Report, RR-3516.

[12] Tsybakov, B. S., and Georganas, N. D.On Self-Similar Traffic in ATM Queues: Definitions,Overflow Probability Bound,and Cell Delay Distribution, *IEEE/ACM Trans. on Networking*,1997,v.5,No.3, pp.397-409.

[13] Tsybakov, B. S., and Georganas, N. D. Overflow probability in ATM queue with self-similar input traffic,*ICC'97, IEEE Interna- tional conference on communications (Montreal, Canada),Conference Record*, 1997, v.2., pp. 822-826 .

[14] Tsybakov B. S. and Georganas N. D. A model of self-similar communications-network traffic,*Proc.of International Conference, Distributed Computer Communication Networks (DCCN'97)*, Tel-Aviv (Israel), 1997, pp.212-217.

[15] Tsybakov B. S. and Georganas N. D.,Self-similar traffic: Upper bounds to buffer-overflow probability in ATM Queu,*Proc.the Canadian conference on Broadband Research (Ottawa, Canada)*, 1997, pp. 137-148.

[16] Tsybakov, B. S., and Georganas, N. D., On Self-Similar Traffic and upper bounds to buffer-overflow probability in ATM queue, *Performance Evaluation*, 1997, v.32, No.4, pp.365-385.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399