Large Deviations for Polling Systems

Franck Delcoigne, Arnaud de la Fortelle

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Large deviations for polling systems

Franck Delcoigne , Arnaud de La Fortelle

N ° 3892

Mars 2000

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3892--FR+ENG

Large deviations for polling systems

Franck Delcoigne * † , Arnaud de La Fortelle †

Thème 1 — Réseaux et systèmes
Projet MEVAL

Rapport de recherche n° 3892 — Mars 2000 — 43 pages

Abstract: In this paper, we prove a sample path large deviation principle for a rescaled process $n^{-1}Q_{nt}$, where Q_t represents the joint number of clients at time t in a polling system with N nodes, one server and Markovian routing. Our main goal is to identify the rate function. We introduce a so called *empirical generator* consisting of Q_t and of two empirical measures associated with S_t, the position of the server at time t. The analysis relies on a suitable change of measure and on a representation of fluid limits for polling systems.

Key-words: Large deviations, polling system, fluid limits, empirical generator, change of measure, contraction principle, entropy, convex program

(Résumé : tsvp)

The work of the second author is a part of a PhD thesis supported by the École Nationale des Ponts et Chaussées and by the French Ministère de l'Équipement.

* Postal address: Université Paris 10, UFR SEGMI, 200 av. de la République - 92000 Nanterre - FRANCE.
† Postal address: INRIA - Domaine de Voluceau, Rocquencourt - BP 105 - 78153 Le Chesnay Cedex - FRANCE.

Grandes déviations pour les systèmes à polling

Résumé : On établit un principe de grandes déviations pour la suite de processus renormalisés $n^{-1}Q_{nt}$, où Q_t représente le nombre joint de clients à l'instant t dans un système à polling à N files, un serveur et routage Markovien. L'objectif principal de cette étude est d'identifier la fonctionnelle d'action. Une des difficultés propre au système étudié provient du fait que si S_t représente la position du serveur à l'instant t, (Q_t, S_t) est un processus de Markov mais non Q_t. Dans cette perspective, on démontre dans un premier temps des bornes de grandes déviations pour une fonctionnelle appellée *générateur empirique* et qui est composée de $n^{-1}Q_{nt}$ et de deux mesures empiriques liées au déplacement du serveur. Cette étape se fonde sur un changement de mesure approprié et sur la représentation des limites fluides pour les systèmes à polling. Un principe de contraction permet ensuite d'obtenir les bornes de grandes déviations voulues pour $n^{-1}Q_{nt}$.

Mots-clé : Grandes déviations, système à polling, limites fluides, générateur empirique, changement de mesure, contraction, entropie, programme convexe

1 Introduction

Consider a polling system consisting of N nodes attended by a single server and denote by $\mathcal{S} \stackrel{\text{def}}{=} \{1, \ldots, N\}$ the set of nodes. At node i, arrivals of clients form a Poisson process with rate λ_i. Each customer at node i requires service, whose duration is exponentially distributed with parameter μ_i. Let $\rho_i \stackrel{\text{def}}{=} \lambda_i/\mu_i$, the intensity factor at node i. When the server arrives at a busy node, say i, it serves one customer and then moves to some node, chosen via some ergodic routing matrix $\mathbf{P} = (P_{ij})_{i,j\in\mathcal{S}}$ with invariant measure $\eta = (\eta_i)_{i\in\mathcal{S}}$. If it reaches an empty node, then it immediately switches to some other node, still chosen according to $\mathbf{P}$. The switch-over time to go from node i to node j, for $i, j \in \mathcal{S}$, is exponentially distributed with mean τ_{ij}. All stochastic input sequences (inter arrival times, services, switch-over times) are supposed to be mutually independent. When the join number of clients and the position of the server at time 0 are respectively given by $x = (x_1, \ldots, x_N)$ and s, $Q(t, x, s) = \big(q_1(t, x, s), \ldots, q_N(t, x, s)\big)$ and $S(t, x, s)$ represent the join number of clients at each node and the position of the server at time t. As a rule, we shall write $S(t, x, s) = i$ if the server is serving some customer at node i and $S(t, x, s) = ij$ if it is in transit between nodes i and j. Set $\mathcal{S}_0 \stackrel{\text{def}}{=} \mathcal{S} \cup \mathcal{S}^2$, the state space of the server. Then

$$X_{x,s} \stackrel{\text{def}}{=} \Big\{ \big(Q(t, x, s), S(t, x, s)\big),\ t \geq 0 \Big\}$$

is a Markov process with generator R such that

$$Rf(x, s) = \sum_{(y,s')\in\mathbb{Z}_+^N\times\mathcal{S}_0} q(x, s; y, s')\Big(f(y, s') - f(x, s)\Big), \qquad \forall (x, s) \in \mathbb{Z}_+^N \times \mathcal{S}_0,$$

where $f \in \mathcal{B}\big(\mathbb{Z}_+^N \times \mathcal{S}_0\big)$ and

$$q(x, s; y, s') = \begin{cases} \lambda_i, & \text{if} \quad y = x + e_i, s' = s,\ \forall i \in \mathcal{S}, \\ \mu_i P_{ij}, & \text{if} \quad x_i > 0, s = i, y = x - e_i, s' = ij,\ \forall i, j \in \mathcal{S}, \\ \dfrac{1}{\tau_{ji}}, & \text{if} \quad x_i > 0, s = ji, y = x, s' = i,\ \forall i, j \in \mathcal{S}, \\ \dfrac{1}{\tau_{ji}} P_{il}, & \text{if} \quad x_i = 0, s = ji, y = x, s' = il,\ \forall i, j, l \in \mathcal{S}, \\ 0, & \text{otherwise.} \end{cases}$$

Whenever no confusion arises, the initial state (x, s) will be dropped. Let us recall that, if $\{x_n,\ n \geq 0\}$ is a sequence of $\mathbb{Z}_+^N$ such that $|x_n| \to \infty$, then every limiting point in distribution of the sequence of processes $|x_n|^{-1} X_{x_n,s}$ will be called a fluid limit. A precise definition of fluid limits for polling systems are given in subsection 2.2.

Let us introduce now a definition and a notation which will be of constant use in the sequel :

Definition *For every $x = (x_1, \ldots, x_N) \in \mathbb{R}_+^N$, denote by $\Lambda(x)$ the set of indices i such that $x_i > 0$. If Λ is a subset of $\mathcal{S}$, the subset of $\mathbb{R}_+^N$*

$$\{x \in \mathbb{R}_+^N / x_i > 0, \forall i \in \Lambda, x_i = 0, \forall i \in \Lambda^c\}$$

is called face Λ.

- For any set A, A^c will denote its complementary and $\mathbb{1}_{\{A\}}$ its indicator function.
- For any space E, $\mathcal{B}(E)$, $\mathcal{M}(E)$, $\mathcal{P}(E)$, represent respectively the sets of bounded functions on E, of positive measures on E and of probability measures on E.
- $D\left([0, T], \mathbb{R}^N\right)$, the space of right continuous functions $f\ :\ [0, T] \to \mathbb{R}^N$ with left limits, endowed with the Skorokhod metric denoted by d_d ;
- $\mathcal{PL}\left([0, T], \mathbb{R}^N\right)$, the set of piecewise linear functions whose derivative has only finitely many discontinuities ;
- $\mathcal{C}\left([0, T], \mathbb{R}^N\right)$, the space of continuous functions equipped with the metric of the uniform convergence denoted by d_c ;
- $\mathcal{AC}\left([0, T], \mathbb{R}^N\right)$, the space of absolutely continuous functions.
- $\mathcal{L}^p\left([0, T], \mathbb{R}^N\right)$, the space of functions f such that $|f|^p$ is integrable with respect to the Lebesgue measure.

A huge literature has been devoted to the study of polling systems because of their wide range of applicability. In [2, 8, 15], the necessary and sufficient conditions of ergodicity has been established for systems with one or several servers under a rich variety of service policies. However, the problem of determining the invariant measure for such systems is still open. Even, for limited policies, the mean waiting time can be computed only under symmetry assumptions [3]. The reader is referred to [20] for an overview about polling systems. In the present paper, a sample path large deviation principle or a sample path LDP for the rescaled process $n^{-1}Q_{nt}$ is established. This could be a preliminary step in order to obtain large deviations estimates for the stationary distribution. In view of future applications, some particular attention is devoted to the computation of the rate function governing the sample path LDP. All this program falls into the framework of LDP for Markov processes with discontinuous statistics i.e those for which the coefficients of their generator are not spatially continuous.

It seems that one of the first paper dealing with such processes is [14], where large deviations problems for Jackson networks were investigated using partial differential equations techniques. Quite recently, the LDP for a large class of Markov processes with discontinuous statistics has been proved in [11]. Roughly speaking, the authors of [11] express the logarithm of large deviation probabilities as the minimal cost of some stochastic optimal-control problem, and the limit of the optimal cost is shown to exist by means of a sub-additivity argument. However, the rate function is not explicit. Note that in [13], an explicit upper bound of large deviations involving Legendre transforms is proved. While for Jackson networks and some processor sharing models this bound is tight [1], in general the problem of the lower bound remains open. It is worth emphasizing that in our case, Q_t is not a Markov process so that the polling model does not satisfy the assumptions of [11]. Besides, the rate function can be explicitely described in terms of entropy functions.

Untill now, the identification of the rate function has been carried out in some particular cases and usually for low dimensional systems. In [6], using the contraction principle, the exponential decay of the stationary distribution of the waiting time is computed for a two dimensional tandem networks taking advantage that it can be expressed simply as a continuous function of the input processes. It should be noted that in this setting, a sample path LDP for processes with independent increments over infinite intervals of time is needed [7]. General results were obtained in [10, 17] where the LDP has been established for random walks whose generator has a discontinuity along an hyperplane. These results are applied in [17] to compute the exponential decay of the stationary distribution of ergodic random walks in $\mathbb{Z}_+^2$. Nevertheless, in such examples, there are at most two boundaries with codimension one or two where discontinuity arise. Ultimately, the identification of the rate function governing the LDP for Jackson networks has been carried out in [1, 16].

Following [11], in order to get a sample path LDP for polling systems, the main step is to prove the forthcoming large deviations local bounds :

Theorem 1.1 [Local bounds] *Take* $x \in \mathbb{R}_+^N$ *and* $D \in \mathbb{R}^N$ *such that* $D_i = 0$, $\forall i \in \Lambda(x)^c$. *Then, for any* τ *satisfying* $x_i + D_i\tau > 0$, $\forall i \in \Lambda(x)$,

$$\begin{aligned} & \lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\inf_{|y-nx|<\epsilon n}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right] \quad (1.1)\\ = \quad & \lim_{\delta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\sup_{|y-nx|<\epsilon n}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right]\\ \stackrel{\text{def}}{=} \quad & -\tau L(\Lambda(x),D). \end{aligned}$$

Note that for $D_i = 0$, $\forall i \in \Lambda(x)^c$, the conditions $x_i + D_i\tau > 0$, $\forall i \in \Lambda(x)$ is equivalent to $x + Dt$ lies in the face $\Lambda(x)$, for all $t \in [0, \tau]$.

In establishing theorem 1.1, one must know in some sense how the different transition rates have to be modified in order that Q_t follows a given drift D. This means that rather than studying Q_t itself, we focus on the so called empirical generator

$$G_t \stackrel{\text{def}}{=} \left(\frac{N_t}{t}\hat{L}_{N_t}, L_t, \frac{Q_t - Q_0}{t}\right).$$

where

- N_t is the number of jumps of the server till t ;
- S_n is the embedded process of the server just before it jumps. Note that it is not a Markov chain ;
- $\hat{L}_n \stackrel{\text{def}}{=} \dfrac{1}{n}\displaystyle\sum_{i=0}^{n}\delta_{S_i} \in \mathcal{P}(\mathcal{S}_0)$ is the empirical measure of the process $\{S_n,\ n \geq 0\}$;
- $L_t \stackrel{\text{def}}{=} \dfrac{1}{t}\displaystyle\int_0^t \delta_{S_u}du \in \mathcal{P}(\mathcal{S}_0)$ is the empirical measure of the process $\{S_t,\ t \geq 0\}$.

In theorem 4.1, we prove large deviations bounds with a rate function $H(.\|R)$ defined in definition 3.6, for a uniform version of G_t^Λ, where G_t^Λ is the empirical generator associated to a *localized* polling system X^Λ. The transition mechanism describing the evolution of X^Λ is identical to X's except that the components indexed by Λ for X^Λ can be negative. In order to prove this result, we use a change of measure (chosen in a restricted class) which gives rise to a new polling system, for which the fluid limits can be completely characterized (see theorems 2.2, 2.3), and this is a key ingredient in the proof of the lower bound. This ensures that for D as in theorem 1.1, after a suitable change of measure, X lies in any neighborhood of D. This way of reasoning is quite classical and was used among others in [16] for identifying the rate function governing the LDP for Jackson networks. However, owing to the presence of S_t, our approach has much more in common with the method used in [4] to prove the weak Sanov LDP for jump Markov processes in continuous time. Moreover, the function $H(.\|R)$ governing the large deviations bounds for G_t^Λ appears as an entropy and is then easily seen to possess good properties.

At this stage, as an application of a kind of contraction principle, one can get theorem 1.1. Since $L(\Lambda, D)$ is the rate function derived from $H(.\|R)$ by a contraction principle, many properties can be derived without much effort and it turns out that $L(\Lambda, D)$ is solution of a convex program. In general, for Markov processes in $\mathbb{Z}_+^N$, the fluid limits cannot be characterized, even in the case of maximal spatial homogeneity. So, it seems that the sole traditional change of measure would be ineffective for the identification of the rate function in this general setting.

Using irreducibility preperties of X, one gets from theorem 1.1 large deviations bounds for the probability that the process drifts along some linear path :

Proposition 1.2 [Linear bounds] *Let $x \in \mathbb{R}_+^N$ and $D \in \mathbb{R}^N$ satisfying $x + D\tau \in \mathbb{R}_+^N$. Then*

$$\begin{aligned}
& \lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\inf_{|y-nx|<\epsilon n}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-nDt|<\delta n\right] \\
= & \lim_{\delta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\sup_{|y-nx|<\epsilon n}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-nDt|<\delta n\right] \\
= & -\tau L\big(\Lambda(D)\cup\Lambda(x), D\big).
\end{aligned}$$

In theorem 1.1, $x + Dt$ is supposed to lie in the face $\Lambda(x)$ for all $t \in [0, \tau]$. In proposition 1.2, it is assumed that $x + Dt$ evolves in the face $\Lambda(D) \cup \Lambda(x)$ only for $t \in]0, \tau[$. In fact, the boundary effects possibly occuring for t in the neighborhood of 0 or τ can be ignored. The derivation of proposition 1.2 from theorem 1.1 was done in [11]. For the sake of completeness, a simple proof of this fact is given in appendix

A. Now, the rate function $I_T(.)$ for the sample path LDP is expressed as

$$I_T(\varphi) = \begin{cases} \displaystyle\int_0^T L\left(\varphi(t), \dot\varphi(t)\right) dt, & \text{if } \varphi \text{ is absolutely continuous,} \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.2}$$

Remark 1 *$I_T(.)$ is defined by all the values $L(x,D)$, $x \in \mathbb{R}_+^N$ and $D \in \mathbb{R}^N$ such that $D_i = 0$,*
$\forall i \in \Lambda(x)^c$. Indeed, assume that for some t, $\varphi_i(t) = 0$ and $\dot\varphi_i(t)$ exists. Since $\varphi_i(t) \le \varphi_i(s)$ for all s, this implies $\dot\varphi_i(t) \le 0$. Then, necessarily $\dot\varphi_i(t) = 0$. Moreover, φ being absolutely continuous, $\dot\varphi_i(t)$ exists for almost all t.

Remark 2 *In theorem 1.4, it is proved that the rate function $I_T(.)$ has the expected form given in (1.2). It is worth noting that in [11], the rate function governing the LDP was defined as the lower semi-continuous regularization of a functional defined on the set of piecewise linear functions. Nevertheless, it was conjectured that the rate function takes a form like in (1.2).*

Using the Markov property and proposition 1.2, one deduces easily the following large deviations bounds for the probability that the process evolves around some piecewise linear function :

Proposition 1.3 [Piecewise linear bounds] *Let $\varphi \in \mathcal{PL}\left([0,T], \mathbb{R}_+^N\right)$ and set $x \stackrel{\text{def}}{=} \varphi(0)$. Then*

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty} \frac{1}{n} \inf_{|y-nx|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[0,nT]} |Q(t,y) - n\varphi(t)| < \delta n\right]$$
$$= \lim_{\delta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty} \frac{1}{n} \sup_{|y-nx|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[0,nT]} |Q(t,y) - n\varphi(t)| < \delta n\right] = -I_T(\varphi),$$

Set

$$Q_{x,s}^n \stackrel{\text{def}}{=} \left\{\frac{1}{n} Q(nt, [nx], s),\ t \ge 0\right\},$$

and

$$\Phi_x(K) = \{\varphi \in D([0,T], \mathbb{R}_+^N) : I_T(\varphi) \le K,\ \varphi(0) = x\}.$$

When proposition 1.3 has been established, using various properties of the rate function $I_T(.)$ and the exponential tightness of $\{Q_{x,s}^n,\ n \ge 1\}$, one can get the sample path LDP :

Theorem 1.4 [Sample path LDP] *The sequence* $\{Q^n_{x,s},\ n \geq 1\}$ *satisfies a LDP in* $D\big([0,T],\mathbb{R}_+^N\big)$ *with good rate function* $I_T(.)$ *: for every* $T > 0$, $x \in \mathbb{R}_+^N$ *and* s,

(*i*) *For any compact set* $C \subset \mathbb{R}_+^N$, $\bigcup_{x\in C} \Phi_x(K)$ *is a compact set of* $\mathcal{C}([0,T],\mathbb{R}_+^N)$.

(*ii*) *for each closed set* F *of* $D\big([0,T],\mathbb{R}_+^N\big)$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[Q^n_{x,s} \in F\right] \leq -\inf\{I_T(\phi),\ \phi \in F, \phi(0) = x\};$$

(*iii*) *for each open set* O *of* $D\big([0,T],\mathbb{R}_+^N\big)$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[Q^n_{x,s} \in O\right] \geq -\inf\{I_T(\phi),\ \phi \in O, \phi(0) = x\}.$$

The organization of the paper is the following one. In section 2, the notion of localized polling system is discussed as well as fluid limits ; the local bounds (1.1) are then restated in a more convenient way using bounds for such systems. In section 3, the empirical generator of localized polling system is introduced as well as the entropy $H(.\|R)$ and the local rate function $L(x,D)$. The connections between empirical generator and fluid limits are discussed together with the properties of $H(.\|R)$ and $L(x,D)$. In section 4, we derive large deviations bounds for the empirical generator of localized polling system. Then, using some kind of contraction principle, the local bounds (1.1) are proved. The purely technical proof of the linear bounds (Proposition 1.2) and piecewise linear bounds (Proposition 1.3) is relegated to appendix A. In section 5, the sample path LDP (Theorem 1.4) is proved.

2 Localized polling systems and fluid limits

2.1 Localized polling systems

The choice of x, D and τ in theorem 1.1 ensures that the bounds (1.1) will not depend on the transition mechanism when some components indexed by $\Lambda(x)$ are null. This allows one to restate (1.1) in terms of bounds for a local model without boundary conditions on the nodes $i \in \Lambda(x)$. More precisely, for each subset Λ of $\mathcal{S}$, let us define the process

$$X^{\Lambda}_{x,s} \stackrel{\text{def}}{=} \left\{\big(Q^{\Lambda}(t,x,s), S^{\Lambda}(t,x,s)\big),\ t \geq 0\right\}$$

where $Q^\Lambda(t,x,s) = \left(q_1^\Lambda(t,x,s),\ldots,q_N^\Lambda(t,x,s)\right)$ and $S^\Lambda(t,x,s)$ represent the join number of clients at each node, the position of the server at time t and where (x,s) stands for the initial state. The transition mechanism defining the evolution of X^Λ is identical to X's except that there is no boundary condition on the nodes $i \in \Lambda$: the components $q_i^\Lambda(t,x,s)$ for $i \in \Lambda$ may be negative. X^Λ is then called a localized polling system. All the notation defined for polling systems in section 1 are adapted in a straightforward way for localized polling systems. Note that if $Q^\Lambda_{\Lambda^c}(t,x,s)$ denotes $\left\{q_i^\Lambda(t,x,s),\ i \in \Lambda^c\right\}$, then

$$\left\{\left(Q^\Lambda_{\Lambda^c}(t,x,s), S^\Lambda(t,x,s)\right),\ t \geq 0\right\}$$

is a Markov process.

Definition *By an abuse of notation, one says that $X^\Lambda_{x,s}$ is ergodic if*

$$\left\{\left(Q^\Lambda_{\Lambda^c}(t,x,s), S^\Lambda(t,x,s)\right),\ t \geq 0\right\}$$

is ergodic.

Take τ satisfying $x_i + D_i\tau > 0,\ \forall i \in \Lambda(x)$. Then for ϵ sufficiently small and all y with $|y - nx| < \epsilon$

$$\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] = \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q^{\Lambda(x)}(t,y) - nx - Dt| < \delta n\right].$$

Since $x_i = 0,\ \forall i \in \Lambda(x)^c$, the distribution of $X^\Lambda_{y,s}$ is invariant with respect to the shift nx. Then, one can restate the local bounds (1.1) in term of local bounds for $X^{\Lambda(x)}_{0,s}$.

Proposition 2.1 *Let $x \in \mathbb{R}_+^N$ and $D \in \mathbb{R}^N$ such that $D_i = 0,\ \forall i \in \Lambda(x)^c$. Then, for τ satisfying $x_i + D_i\tau > 0, \forall i \in \Lambda(x)$,*

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\inf_{|y-nx|<\epsilon}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right] = \lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,0)-Dt|<\delta n\right], \tag{2.1}$$

and

$$\lim_{\delta\to0}\lim_{\epsilon\to0}\limsup_{n\to\infty}\frac{1}{n\tau}\sup_{|y-nx|<\epsilon}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right] =$$

$$\lim_{\delta\to0}\limsup_{n\to\infty}\frac{1}{n\tau}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,0)-Dt|<\delta n\right], \tag{2.2}$$

the limits being independent of τ.

Proof Although lengthy, this step is rather classical and mimics mainly the proof of proposition 4.1 of [16]. It is deferred to section B.

2.2 Fluid limits for localized polling systems

In this section, we recall some results of [2, 8, 15] about the convergence of the sequence $\{Q^n_{q,s}\}$ and the ergodicity of localized polling systems. These facts will be needed when proving the lower bound for the empirical generator. Denote $\tau \stackrel{\text{def}}{=} \sum_{i,j\in\mathcal{S}}\eta_i p_{ij}\tau_{ij}$. Then,

Theorem 2.2 *(i) X^Λ is ergodic if, and only if,*

$$\sum_{k\in\Lambda^c}\rho_k+\frac{\lambda_i}{\eta_i}\left(\tau+\sum_{k\in\Lambda}\frac{\eta_k}{\mu_k}\right)<1,\ \forall i\in\Lambda^c.$$

(ii) For any sequence $\{x_n,\ n\ge1\}$ of $\mathbb{Z}_+^N$ with $|x_n|\to\infty$, it exists a subsequence $\{n_k,\ k\ge1\}$ such that :

(a) in probability, $\lim_{k\to\infty}\frac{1}{|x_{n_k}|}Q^\Lambda(0,x_{n_k},s)=\overline{Q}^\Lambda(0)$;

(b) uniformly over all compact sets, the following limits hold in probability :

$$\lim_{k\to\infty}\left(\frac{N^\Lambda_{n_kt}}{n_k}\hat{L}^\Lambda_{N_{n_kt}},tL^\Lambda_{n_kt},\frac{1}{|x_{n_k}|}Q^\Lambda(n_kt,x_{n_k},s)\right)=\left(\overline{A}^\Lambda_t,\overline{L}^\Lambda_t,\overline{Q}^\Lambda(t)\right),$$

where $\overline{A}_t^\Lambda \in \mathcal{M}(\mathcal{S}_0), t^{-1}\overline{L}_t^\Lambda \in \mathcal{P}(\mathcal{S}_0)$, $\forall t \geq 0$. *Moreover,* $\overline{A}^\Lambda, \overline{L}^\Lambda, \overline{Q}^\Lambda$ *are Lipschitz and satisfy the following set of relations :*

$$\begin{aligned}
\overline{Q}_i^\Lambda(t) &= \overline{Q}_i^\Lambda(0) + \lambda_i t - \overline{A}_t^\Lambda\{i\}, \ \forall i \in \mathcal{S}, && (2.3)\\
\sum_{j\in\mathcal{S}} \overline{A}_t^\Lambda\{ij\} &= \sum_{j\in\mathcal{S}} \overline{A}_t^\Lambda\{ji\}, \ \forall i,j \in \mathcal{S}, && (2.4)\\
\overline{A}_t^\Lambda\{i\} &= \sum_{j\in\mathcal{S}} \overline{A}_t^\Lambda\{ij\}, \ \forall i \in \Lambda, && (2.5)\\
\overline{A}_t^\Lambda\{i\} &\leq \sum_{j\in\mathcal{S}} \overline{A}_t^\Lambda\{ij\}, \ \forall i \in \Lambda^c, && (2.6)\\
\overline{A}_t^\Lambda\{ij\} &= P_{ij} \sum_{k\in\mathcal{S}} \overline{A}_t^\Lambda\{ik\}, \ \forall i,j \in \mathcal{S}, && (2.7)\\
\overline{A}_t^\Lambda\{ij\} &= \frac{1}{\tau_{ij}} \overline{L}_t^\Lambda\{ij\}, \ \forall i,j \in \mathcal{S}, && (2.8)\\
\overline{A}_t^\Lambda\{i\} &= \mu_i \overline{L}_t^\Lambda\{i\}, \ \forall i \in \mathcal{S}. && (2.9)
\end{aligned}$$

Remark 3 *Note that in some sense,* $\overline{A}_t^\Lambda\{s\}$ *and* $\overline{L}_t^\Lambda\{s\}$ *represent respectively the number of visits and the time spent in state* s *by the server till time* t.

Now, let us restate theorem 2.2 for localized polling systems with initial condition $(0, s)$ since by proposition 2.1 this will be sufficient in order to prove the local bounds of large deviation. Moreover, for our purpose, it will be sufficient to assume the stability condition (2.10). It is worth noting that the fluid limits obtained in this case are linear.

Theorem 2.3 *Assume the following stability condition :*

$$\sum_{k\in\Lambda^c} \rho_k + \frac{\lambda_i}{\eta_i}\left(\tau + \sum_{k\in\Lambda} \frac{\eta_k}{\mu_k}\right) \leq 1, \ \forall i \in \Lambda^c. \tag{2.10}$$

Uniformly over all compact sets, the following limits hold in probability :

$$\lim_{n\to\infty} \left(\frac{N_{nt}^\Lambda}{n} \hat{L}_{N_{nt}}^\Lambda, tL_{nt}^\Lambda, \frac{1}{n} Q^\Lambda(nt, 0, s) \right) = \big(tA, t\pi, tD\big),$$

where $A \in \mathcal{M}(\mathcal{S}_0)$, $\pi \in \mathcal{P}(\mathcal{S}_0)$ *and* $D \in \mathbb{R}^N$. *Moreover,*

$$a_s = \begin{cases} \lambda_i, & \textit{if} \quad s = i \in \Lambda^c, \\ C\eta_i, & \textit{if} \quad s = i \in \Lambda, \\ C\eta_i P_{ij}, & \textit{if} \quad s = ij, \ \forall i,j \in \mathcal{S}, \end{cases} \tag{2.11}$$

$$\pi_s = \begin{cases} \dfrac{a_i}{\mu_i}, & \textit{if } s = i \in \mathcal{S}, \\ a_{ij}\tau_{ij}, & \textit{if } s = ij, \ \forall i,j \in \mathcal{S}. \end{cases} \tag{2.12}$$

$$D_i = \lambda_i - a_i, \quad \forall i \in \mathcal{S}, \tag{2.13}$$

$$\textit{where} \qquad C = \frac{1 - \sum_{k \in \Lambda^c} \rho_k}{\tau + \sum_{k \in \Lambda} \frac{\eta_k}{\mu_k}}.$$

Proof : Let us briefly sketch the proof. Put

$$\overline{W}_t^\Lambda = \sum_{k \in \Lambda^c} \frac{\overline{Q}_k^\Lambda(t)}{\mu_k}.$$

Note that $\overline{W}_t^\Lambda \geq 0$ and $\overline{W}_0^\Lambda = 0$. Moreover, $\overline{W}^\Lambda$ is Lipschitz and so absolutely continuous. Then, as in lemma 1 of [15] or lemma 3.1 of [8], using (2.10) one can prove that at all regular point the derivative of $\overline{W}_t$ is negative. Hence $\overline{W}_t = 0$, $\forall t$. This implies that $\overline{Q}_k^\Lambda(t) = 0$, $\forall k \in \Lambda^c$, $\forall t \geq 0$, and by (2.3)

$$\overline{A}_t^\Lambda\{i\} = \lambda_i t, \ \forall i \in \Lambda^c. \tag{2.14}$$

Moreover, (2.4) ensures that it exists C_t such that

$$\sum_{j \in \mathcal{S}} \overline{A}_t^\Lambda\{ij\} = C_t \eta_i, \ \forall i \in \mathcal{S}.$$

Hence, by (2.7),

$$\overline{A}_t^\Lambda\{ij\} = C_t \eta_i P_{ij}, \ \forall i,j \in \mathcal{S}. \tag{2.15}$$

Hence, using (2.14), (2.15) together with (2.5), (2.8) and (2.9), one has

$$t^{-1}\overline{L}_t^\Lambda\{s\} = \begin{cases} \rho_i, & \text{if} \quad s = i \in \Lambda^c, \\ \dfrac{C_t}{t}\dfrac{\eta_i}{\mu_i}, & \text{if} \quad s = i \in \Lambda, \\ \dfrac{C_t}{t}\eta_i P_{ij}\tau_{ij}, & \text{if} \quad s = ij, \ \forall i,j \in \mathcal{S}, \end{cases}$$

$t^{-1}\overline{L}_t^\Lambda$ being a probability for all t,

$$C_t = t \frac{1 - \sum_{k \in \Lambda^c} \rho_k}{\tau + \sum_{k \in \Lambda} \frac{\eta_k}{\mu_k}}.$$

Now, the other relations can be derived directly. ■

3 Generators and entropy

The purpose of this section is to define a new representation of the classical generator, which is a kind of description for the long run behavior. We shall also call it generator. It fits very well with fluid limits and large deviations and is easily measured. Moreover it is a natural object inasmuch as operations and properties can be naturally defined for these objects. This section is a generalization for polling systems of the generators introduced in [4] for Markov processes.
The entropy is here a meaningful function which has the classical properties that are expected for such an object, e.g. positivity, continuity and convexity.

3.1 Empirical generators

The empirical generator is the central object of this paper, from which most of the properties are derived. It describes rather naturally the measured behavior of a chain.

Definition 3.1 *Let $X^\Lambda = \left\{Q^\Lambda(t,x,s), S^\Lambda(t,x,s)\right\}_{t\geq 0}$ be a localized polling system. Define S_n^Λ to be the embedded process of the server just before it jumps*[1]*, and,*

- N_t^Λ*, the number of jumps of the server till t ;*
- $\hat{L}_n^\Lambda \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{S_i^\Lambda} \in \mathcal{P}(\mathcal{S}_0)$*, the empirical measure of the process $\{S_n^\Lambda,\ n \geq 1\}$;*
- $L_t^\Lambda \stackrel{\text{def}}{=} \frac{1}{t}\int_0^t \delta_{S_u^\Lambda} du \in \mathcal{P}(\mathcal{S}_0)$*, the empirical measure of the process $\{S_t^\Lambda,\ t \geq 0\}$;*

[1]Note that S_n^Λ is not a Markov chain, because the state of the queues is missing.

- $G_t^\Lambda \stackrel{\text{def}}{=} \left(\frac{N_t^\Lambda}{t}\hat{L}^\Lambda_{N_t^\Lambda}, L^\Lambda_{N_t^\Lambda}, \frac{Q_t^\Lambda - Q_0^\Lambda}{t}\right)$, *the empirical generator associated to the polling system* X^Λ.

We shall denote by $\mathcal{G} \stackrel{\text{def}}{=} \mathcal{M}(\mathcal{S}_0) \times \mathcal{P}(\mathcal{S}_0) \times \mathbb{R}^N$ *the space of empirical generators. Let* $\Gamma \subset \mathcal{S}^2$ *be a subgraph on the nodes* $\mathcal{S}$*, then the space of empirical generators with transitions in* Γ *is denoted by* $\mathcal{G}(\Gamma) \stackrel{\text{def}}{=} \mathcal{M}(\mathcal{S} \cup \Gamma) \times \mathcal{P}(\mathcal{S} \cup \Gamma) \times \mathbb{R}^N$.

For the sake of brevity we shall drop the Λ for the measures N_t^Λ, $\hat{L}_n^\Lambda$, L_t^Λ and G_t^Λ.

3.2 Generators

We shall need the notion of balanced measure in order to describe the transitions of the polling system. It is a well-known object that appears in LDP for discrete time Markov chains [5].

Definition 3.2 (balanced measures) *Let* E *be countable. The set* $\mathcal{M}_s(E^2)$ *of balanced measures is the set of finite positive measures on* E^2*, with both identical 1–dimensional projections*

$$A \in \mathcal{M}_s(E^2) \iff A(E,.) = A(.,E) \in \mathcal{M}(E). \tag{3.1}$$

Let Γ *be a graph on* E. $\mathcal{M}_s(\Gamma) \subset \mathcal{M}_s(E^2)$ *is the set of balanced measures with support included in* Γ.

For $A \in \mathcal{M}_s(E^2)$, let us put

- $a_{ij} \stackrel{\text{def}}{=} A(\{(i,j)\})$ the 2–dimensional law;
- $a_i \stackrel{\text{def}}{=} A(\{i\} \times E)$, the unique 1–dimensional projection;
- $A_{ij} \stackrel{\text{def}}{=} A(\{(i,j)\} | \{i\} \times E)$, the conditional law, and hence $a_{ij} = a_i A_{ij}$.

Note that we shall use the identity $A_{ij} = B_{ij}$ for $A, B \in \mathcal{M}_s(E^2)$. This does not mean equality $A = B$ for the two-dimensional law, but only for the conditional law. Now, the analysis of polling systems has some specific features that must be reflected in the mathematical structure. The balance equation (3.1) just states that, in a Markov chain, a process enters a state as often as it exits it. In a polling system, the server goes to a queue as many times as it goes from that queue, but it stops and serves less times.

Definition 3.3 *For $A \in \mathcal{M}(\mathcal{S}_0)$, let $\overline{a}_i = \sum_{j\in\mathcal{S}} a_{ij}$ for all i in $\mathcal{S}$. Denote $\mathcal{M}_s(\mathcal{S}_0)$ the set of elements $A \in \mathcal{M}(\mathcal{S}_0)$ satisfying*

$$\sum_{j\in\mathcal{S}} a_{ji} = \sum_{j\in\mathcal{S}} a_{ij} \stackrel{\text{def}}{=} \overline{a}_i, \quad \forall i \in \mathcal{S}, \tag{3.2}$$

$$a_i \leq \overline{a}_i, \qquad \forall i \in \mathcal{S}. \tag{3.3}$$

For all subset Λ of $\mathcal{S}$, denote $\mathcal{M}_s^\Lambda(\mathcal{S}_0)$ the set of elements $A \in \mathcal{M}(\mathcal{S}_0)$ satisfying (3.2) for all $i \in \mathcal{S}$, (3.3) for all $i \in \Lambda^c$, and the equality in (3.3) for all $i \in \Lambda$.

The equality in (3.3) means that, when the server arrives at a saturated queue, then it serves a client. Therefore $\mathcal{M}_s(\mathcal{S}_0) = \mathcal{M}_s^\emptyset(\mathcal{S}_0)$. Better, when the constraints (3.3) become equalities, i.e. when Λ increases, the set $\mathcal{M}_s^\Lambda(\mathcal{S}_0)$ decreases:

$$\Lambda \subset \Lambda' \Longrightarrow \mathcal{M}_s^{\Lambda'}(\mathcal{S}_0) \subset \mathcal{M}_s^\Lambda(\mathcal{S}_0). \tag{3.4}$$

Note that, by (3.2), the matrix $\overline{a}_{ij} \stackrel{\text{def}}{=} a_{ij}$ associated to A is a balanced measure on $\mathcal{S}^2$. Its conditional law is $\overline{A}_{ij} \stackrel{\text{def}}{=} a_{ij}/\overline{a}_i$.

Definition 3.4 (generators) *For all subset Λ of $\mathcal{S}$, the set $\mathcal{G}_s^\Lambda$ of generators is defined by*

$$\mathcal{G}_s^\Lambda \stackrel{\text{def}}{=} \left\{(A,\pi,D) \in \mathcal{M}_s^\Lambda(\mathcal{S}_0) \times \mathcal{P}(\mathcal{S}_0) \times \mathbb{R}^\Lambda \;:\; a_i + D_i \geq 0,\ \forall i \in \Lambda\right\},$$

where we denote by $\mathbb{R}^\Lambda$, for the sake of simplicity, the subspace of $D \in \mathbb{R}^N$ with $D_i = 0$ for $i \in \Lambda^c$.
The characteristics of a generator such as irreducibility or the support refer to the routing part A of the generator. Let Γ be a subgraph of $\mathcal{S}^2$, then $\mathcal{G}_s^\Lambda(\Gamma) \subset \mathcal{G}_s^\Lambda$ is the set of generators with support included in Γ.
We shall denote by l the overall mean intensity of server jumps, $l = \sum_{s\in\mathcal{S}_0} a_s$.

Note that $\mathcal{G}_s^\Lambda$ does not decrease with Λ as $\mathcal{M}_s^\Lambda$. Indeed the drift part $\mathbb{R}^\Lambda$ increases with Λ.

Proposition 3.1 *There is a one-to-one mapping between the irreducible generators of $\mathcal{G}_s^\Lambda$ and the stable irreducible localized polling systems. (A,π,D) corresponds to the fluid limit of a localized polling system with parameters $(\tilde\lambda_i, \tilde\mu_i, \tilde\tau_{ij}, \tilde P_{ij})$, i.e. equations (2.11)–(2.13), when $(\tilde\lambda_i, \tilde\mu_i, \tilde\tau_{ij}, \tilde P_{ij})$ are derived from (A,π,D) by equations (3.5)–(3.8).*

Proof : Let $G = (A, \pi, D) \in \mathcal{G}_s^\Lambda$ be an irreducible generator. It is associated a unique localized polling system $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})_{i,j\in\mathcal{S}}$ by the relations

$$\begin{aligned} \tilde{\lambda}_i &= a_i + D_i, && (3.5)\\ \tilde{\mu}_i &= \frac{a_i}{\pi_i}, && (3.6)\\ \tilde{\tau}_{ij} &= \frac{\pi_{ij}}{a_{ij}}, && (3.7)\\ \tilde{P}_{ij} &= \overline{A}_{ij} = \frac{a_{ij}}{\bar{a}_i}. && (3.8) \end{aligned}$$

It must be now proved that $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})_{i,j\in\mathcal{S}}$ verifies the stability condition (2.10). By (3.2) and (3.8), $\bar{a}_i$ is an invariant measure for $\tilde{P}$. For G is irreducible, so is $\tilde{P}$, thus $\bar{a}_i$ is the unique invariant measure, up to a multiplicative constant. The invariant probability of $\tilde{P}$ is then $\tilde{\eta}_i = C\bar{a}_i$. Since $G \in \mathcal{G}_s^\Lambda$, the condition $a_i \leq \bar{a}_i$ for $i \in \Lambda^c$ can be written

$$\sum_{k\in\Lambda^c} \pi_k + \frac{a_i}{\bar{a}_i}\left(\sum_{k,l\in\mathcal{S}} \pi_{kl} + \sum_{k\in\Lambda} \pi_k\right) \leq 1, \qquad \forall i \in \Lambda^c. \tag{3.9}$$

Since $D_i = 0$ for $i \in \Lambda^c$, and $a_i = \bar{a}_i$ for $i \in \Lambda$, (3.9) is equivalent to (2.10). Therefore $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})_{i,j\in\mathcal{S}}$ is a stable localized polling system.
Reciprocally, let $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})_{i,j\in\mathcal{S}}$ be a stable irreducible localized polling system. The relations (2.11)–(2.13) define a unique triple (A, π, D) such that A verifies the balance relation (3.2) and that $a_i = \bar{a}_i$ for $i \in \Lambda$, by (2.4) and (2.5). The stability condition $a_i \leq \bar{a}_i$ for $i \in \Lambda^c$ is immediately derived from the fluid limit equation (2.6). Therefore $A \in \mathcal{M}_s^\Lambda$. Since $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})_{i,j\in\mathcal{S}}$ is stable, $D_i = 0$ for $i \in \Lambda^c$, hence $(A, \pi, D) \in \mathcal{G}_s^\Lambda$.
We have constructed two mappings between $\mathcal{G}_s^\Lambda$ and $\mathcal{P}^\Lambda$, mainly verifying that the stability conditions are the same. We have now to prove that these mappings are inverse of each other. Fix (A, π, D), then makes the correspondence

$$(A, \pi, D) \to (\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij}) \to (\tilde{A}, \tilde{\pi}, \tilde{D}).$$

Remind that we proved just above that $\tilde{\eta}_i = C\bar{a}_i$; now, by (2.11), $\tilde{a}_i = \tilde{C}\tilde{\eta}_i$ for $i \in \Lambda$. Calculations similar to (3.9) yield that $C\tilde{C} = 1$, then it is straight to prove $\tilde{A} = A$, $\tilde{\pi} = \pi$ and $\tilde{D} = D$. The bijection is proved. ∎

The triple (A, π, D) of a generator is easily divided into a server part (A, π) and the drift part D. The server part a_s is interpreted as the mean number of passages through s per unit time, the stationary distribution π and the routing $\overline{A}$.

When G is not irreducible, there might exist several stationary distributions which determine several triples. However, if $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})$ and one stationary distribution η for the routing $\tilde{P}$ are known, (A, π, D) is uniquely determined. When $\overline{a}_i = 0$, the mapping is ill-defined but it does not lead to undetermined forms if a stationary distribution π_i is defined. Actually the routing is not well defined, but it does not matter as long as it concerns the entropy function.

Topological considerations : The mean of two generators with weight $\alpha > 0$ and $\beta > 0$ such that $\alpha + \beta = 1$ is defined by $G = \alpha G' + \beta G''$ with

$$\begin{cases} a_s = \alpha a'_s + \beta a''_s, \\ \pi_s = \alpha \pi'_s + \beta \pi''_s, \\ D_i = \alpha D'_i + \beta D''_i, \end{cases} \qquad \text{or} \qquad \begin{cases} \pi_s = \alpha \pi'_s + \beta \pi''_s, \\ \pi_i \tilde{\mu}_i = \alpha \pi'_i \tilde{\mu}'_i + \beta \pi''_i \tilde{\mu}''_i, \\ \pi_{ij}/\tilde{\tau}_{ij} = \alpha \pi'_{ij}/\tilde{\tau}'_{ij} + \beta \pi''_{ij}/\tilde{\tau}''_{ij}, \\ \tilde{\lambda}_i = \alpha \tilde{\lambda}'_i + \beta \tilde{\lambda}''_i, \end{cases}$$

for all $s \in \mathcal{S}_0$ and $i, j \in \mathcal{S}$. Note that all equations are similar, e.g. $\pi_i \tilde{\mu}_i = a_i$, and are defined as additions on flows.

The distance is defined by

$$\|G - G'\|_\infty \stackrel{\text{def}}{=} \max_{s \in \mathcal{S}_0,\ i \in \mathcal{S}} \left\{ |a_s - a'_s|, |\pi_s - \pi'_s|, |D_i - D'_i| \right\}. \tag{3.10}$$

and the topology of $\mathcal{G}_s^\Lambda$ is induced by this distance. A bit different distance is defined for the second representation,

$$\|G - G'\|_\infty \stackrel{\text{def}}{=} \max_{s \in \mathcal{S}_0,\ i,j \in \mathcal{S}} \left\{ |\pi_s - \pi'_s|, \left|\tilde{\lambda}_i - \tilde{\lambda}'_i\right|, |\pi_i \tilde{\mu}_i - \pi'_i \tilde{\mu}'_i|, \left| \frac{\pi_{ij}}{\tilde{\tau}_{ij}} - \frac{\pi'_{ij}}{\tilde{\tau}_{ij}} \right| \right\},$$

but both distances are equivalent since the spaces are finite-dimensional. It is now clear that either representations can be used indifferently, both sharing the same properties.

♦♦♦

Generators and empirical generators : Note that the set of empirical generators $\mathcal{G}$ is not included in any $\mathcal{G}_s^\Lambda$ since the measure $\hat{L}_{N_t}$ is *not necessarily balanced*. Indeed there is generally two states where it is not balanced: if X begins in i and if it is in $j \neq i$ at time t, the server exits i one more time than it enters and the balance equation (3.2) is not satisfied for i. This is also true for the end point j, but for none of the other states. Finally this effects is bounded by $2/t$ for $t^{-1} N_t \hat{L}_{N_t}$ so that it vanishes when the time tends to infinity.

For the purpose of large deviations, bounds should be obtained for open and closed sets in $\mathcal{G}$. Since the distance between $\mathcal{G}$ and $\mathcal{G}_s^\Lambda$ tends to 0, the entropy has only to be defined on $\mathcal{G}_s^\Lambda$ and is infinite otherwise. Therefore the study of large deviations has only to be done locally around $\mathcal{G}_s^\Lambda$. The distance and the topology in $\mathcal{G}$ are defined similarly as in $\mathcal{G}_s^\Lambda$, by (3.10).

♦♦♦

3.3 Entropy

Definition 3.5 *The entropy in discrete time for a balanced measure A with respect to a stochastic matrix $\mathbf{P}$ is defined by*

$$H_d(A\|P) \stackrel{\text{def}}{=} \sum_{i,j\in E} a_{ij} \log\left(\frac{A_{ij}}{P_{ij}}\right), \tag{3.11}$$

with the usual convention $0\log(0/0) = 0, \quad 1/0 = +\infty$ *and* $0\log 0 = 0$.
The continuous-time relative entropy of Poisson processes of intensities ν and λ is defined by

$$I_p(\nu\|\lambda) \stackrel{\text{def}}{=} \nu \log\frac{\nu}{\lambda} - \nu + \lambda \tag{3.12}$$

We recall the properties of H_d and I_p, when E is finite. We shall also denote by $\mathbf{P}$ the graph of $\mathbf{P}$, i.e. all transitions such that $P_{ij} > 0$.

- $H_d(.\|P)$ is positive, it is null if, and only if $A_{ij} = P_{ij}$;
- $H_d(.\|P)$ is 1-homogeneous and convex; if P is irreducible, $H_d(.\|P)$ is strictly convex, i.e. $H(A\|P) = H(B\|P)$ if, and only if, $A_{ij} = B_{ij}$;
- $H_d(.\|P)$ is finite and continuous on $\mathcal{M}_s(\mathbf{P})$, infinite otherwise; it has compact level sets.
- $I_p(\nu\|\lambda)$ is positive, 1-homogeneous, continuous and strictly convex w.r.t. the pair (ν,λ) whenever $\lambda > 0$;
- $I_p(\nu\|\lambda)$ is null if, and only if, $\nu = \lambda$; for all λ, it has compact level sets in ν.

Definition 3.6 (relative entropy) *Let $R = (\lambda_i, \mu_i, \tau_{ij}, P_{ij})$ denotes the generator of the polling system, $G = (A, \pi, D) \in \mathcal{G}_s^\Lambda$ be a generator and $(\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})$ its* *representation as a polling system. The relative entropy of G with respect to R is*[2]

$$H(G\|R) \stackrel{\text{def}}{=} \sum_{i\in\mathcal{S}} \Big(I_p(\tilde{\lambda}_i\|\lambda_i) + \pi_i I_p(\tilde{\mu}_i\|\mu_i) \Big) + \sum_{i,j\in\mathcal{S}} \pi_{ij} I_p(\tilde{\tau}_{ij}^{-1}\|\tau_{ij}^{-1}) + H_d(\overline{A}\|P). \tag{3.13}$$

This entropy has an easy interpretation in terms of information theory. The relative entropy can be defined as the *mean information gain.* $H(.\|R)$ is decomposed in (3.13) as the sum of the information gain for the arrivals $I_p(\tilde{\lambda}_i\|\lambda_i)$, the information gain for the service time $I_p(\tilde{\mu}_i\|\mu_i)$ multiplied by the time π_i spent at each queue, similarly the information gain for the transit time $\pi_{ij} I_p(\tilde{\tau}_{ij}^{-1}\|\tau_{ij}^{-1})$ and the information gain for the routing $H_d(\overline{A}\|P)$. For this last term, it can be decomposed ($H_d(A\|P)$ is 1-homogeneous w.r.t. A) into $\overline{a}$ which is the mean number of routing decisions by unit time, and $H_d(\overline{A}/\overline{a}\|P)$ which is the information gain for each routing decision. This natural interpretation leads to natural properties.

Proposition 3.2 *The relative entropy $H(.\|R)$ is positive, finite, continuous and strictly convex on $\mathcal{G}_s^\Lambda(\mathbf{P})$; it is infinite otherwise; it is null if, and only if, $G = R$. It has compact level sets.*

Proof : All those properties are immediately derived from the same properties of H_d and I_p, under the condition that μ_i, λ_i and τ_{ij} are strictly positive and finite, and that $\mathbf{P}$ is irreducible, conditions that are naturally satisfied here. ■

This is an easy but useful lemma for the LDP.

Lemma 3.3 *The set of irreducible generators is a dense subset of $\mathcal{G}_s^\Lambda(\mathbf{P})$.*

Proof : Let $G \in \mathcal{G}_s^\Lambda(\mathbf{P})$. For $\varepsilon > 0$, $(1-\varepsilon)G + \varepsilon R$ is irreducible because R is irreducible. It converges to G when ε tends to 0, hence the lemma. ■

3.4 Properties of the local functional $L(x, D)$

Definition 3.7 *The rate function $L(\Lambda, D)$ is defined by*

$$L(\Lambda, D) \stackrel{\text{def}}{=} \inf_{G\in f_\Lambda^{-1}(D)} H(G\|R), \quad \forall D \in \mathbb{R}^\Lambda, \tag{3.14}$$

where $f_\Lambda : \mathcal{G}_s^\Lambda \mapsto \mathbb{R}^\Lambda$ is the projection $f_\Lambda(G) = D$. For all $x \in \mathbb{R}_+^N$, $L(x, D)$ is defined to be $L(\Lambda(x), D)$.

[2] Note that the relative entropy $H(G\|R)$ is independent of Λ.

Note that $L(\Lambda, D)$ is a rate function derived by the contraction of $H(.\|R)$, which is a good rate function (see Proposition 3.2). Even though the corresponding LDP are not proved, since f_Λ is linear, L inherits lots of good properties of H (see [4]).

Proposition 3.4 *The rate function $L(\Lambda, D)$ is positive, finite, continuous and strictly convex with respect to $D \in \mathbb{R}^\Lambda$; it is null if, and only if, D is the drift of the localized polling system; it has compact level sets. Moreover, the infimum $G \in \mathcal{G}_s^\Lambda$ such that $H(G\|R) = L(\Lambda, D)$ is reached at a unique point $G(D) \in \mathcal{G}_s^\Lambda(\mathbf{P})$, and $G(D)$ is a continuous function of D.*

Proof : Since f_Λ is linear and $H(.\|R)$ is strictly convex, by proposition 4.8 in [4], $L(\Lambda, .)$ is convex. The compactness of the level sets is immediately derived from the compactness of the level sets of $H(.\|R)$.
Since $H(.\|R)$ is strictly convex, since its level sets are compact and since $L(\Lambda, D)$ is the solution of a convex program, the infimum $G(D)$ is reached at a unique point, which is obviously in $\mathcal{G}_s^\Lambda(\mathbf{P})$ where H is finite.
Let $D_n \stackrel{n\to\infty}{\longrightarrow} D$. Since $H(.\|R)$ has compact level sets and since it is continuous on $\mathcal{G}_s^\Lambda(\mathbf{P})$, it is easily seen that the sequence $\{G(D_n),\ n \geq 1\}$ is relatively compact. Let $\{G(D_{n_k}),\ k \geq 1\}$ be a subsequence such that $G(D_{n_k}) \stackrel{k\to\infty}{\longrightarrow} G$. $H(.\|R)$ being continuous,

$$L(\Lambda, D_{n_k}) = H(G(D_{n_k})\|R) \stackrel{k\to\infty}{\longrightarrow} H(G\|R).$$

By definition of L, $H(G\|R) \geq L(\Lambda, D)$. Now $G(D) = (A, \pi, D)$, and consider $G_{n_k} = (A, \pi, D_{n_k})$. Since H is continuous and since G_{n_k} tends to $G(D)$,

$$L(\Lambda, D) = H(G(D)\|R) = \lim_{k\to\infty} H(G_{n_k}\|R) \geq \lim_{k\to\infty} H(G(D_{n_k})\|R) = H(G\|R).$$

Then, $H(G\|R) = H(G(D)\|R)$, and by the uniqueness of the infimum, $G = G(D)$. So, $G(D_n) \stackrel{n\to\infty}{\longrightarrow} G(D)$. Finally $G(D)$ is continuous and so is $L(\Lambda, D)$. ■

Remark : We shall extend definition 3.7 for all $D \in \mathbb{R}^N$ by extending the validity of the equation (3.14). First, extend the set $\mathcal{G}_s^\Lambda$ of generators to

$$\overline{\mathcal{G}}_s^\Lambda \stackrel{\text{def}}{=} \left\{(A, \pi, D) \in \mathcal{M}_s^\Lambda(\mathcal{S}_0) \times \mathcal{P}(\mathcal{S}_0) \times \mathbb{R}^N \ : \ a_i + D_i \geq 0,\ \forall i\right\}.$$

This actually means that all localized generators are considered, not only the stable ones. The projection $\bar{f}_\Lambda :\ \overline{\mathcal{G}}_s^\Lambda \mapsto \mathbb{R}^N$ is the analog of f_Λ, and now

$$L(\Lambda, D) \stackrel{\text{def}}{=} \inf_{G \in \bar{f}_\Lambda^{-1}(D)} H(G\|R)$$

is defined for all $D \in \mathbb{R}^N$. Of course the definitions coincide for $D \in \mathbb{R}^\Lambda$.

One checks this extension preserves all properties proved here, particularly the lower semi-continuity and the convexity, which is convenient for the proof of the sample path LDP. Moreover, as underlined in remark 1 p. 8, these values are almost never used, so that finally this extension does not modify the rate function I_T.

♦♦♦

Lemma 3.5 *$L(\Lambda, D)$ is increasing w.r.t. Λ. $L(x, D)$ is jointly lower semi-continuous w.r.t. x and D.*

Proof : By Definition 3.7, $H(L\|R)$ does not depend of Λ and the domain $f_\Lambda^{-1}(D)$ where the infimum is taken decreases w.r.t. Λ; therefore $L(\Lambda, D)$ is increasing w.r.t. Λ. Actually $(A, \pi, D) \in f_\Lambda^{-1}(D)$ if, and only if $\pi \in \mathcal{P}(\mathcal{S}_0)$ and $A \in \mathcal{M}_s^\Lambda(\mathcal{S}_0)$ with $a_i + D_i \geq 0$ if $i \in \Lambda$, so that, by (3.4), it decreases w.r.t. Λ.
Let (x_n, D_n) tends to (x, D). Since all strictly positive coordinates of x have to be strictly positive for n large enough,

$$\exists n_0, \quad \Lambda(x) \subset \Lambda(x_n), \qquad \forall n \geq n_0.$$

Therefore, using both the continuity w.r.t. D and the growth of $L(., D)$,

$$\begin{aligned} &\liminf_{n\to\infty} \big(L(x_n, D_n) - L(x, D)\big) \\ \geq\ & \lim_{n\to\infty} \inf_{\Lambda\supset\Lambda(x)} \big(L(\Lambda, D_n) - L(\Lambda, D)\big) + \liminf_{n\to\infty} \big(L(x_n, D) - L(x, D)\big) \geq 0. \end{aligned}$$

Hence the lower semi-continuity with respect to both variables x and D. ■

Lemma 3.6 *There exists $M \in \mathbb{R}$ such that,*

$$L(x, D) \geq \frac{1}{2}\|D\| \log \|D\|, \qquad \forall x \in \mathbb{R}_+^N,\ \forall \|D\| \geq M. \tag{3.15}$$

Proof : Assume that $|D_j| = \|D\| \stackrel{\text{def}}{=} \max_i |D_i|$. From the definitions of $L(x, D)$ (Definition 3.7), of $H(G\|R)$ (Definition 3.6) and from the positivity of I_p and H_d,

$$L(x, D) \geq \inf_{L\in f_\Lambda^{-1}(D)} \big(I_p(a_j + D_j\|\lambda_j) + I_p(a_j\|\pi_j\mu_j)\big).$$

But, $I_P(x\|y)$ is increasing in x and decreasing in y for $x \geq y$ and necessarily $a_j \geq -D_j$, so that,

$$L(x,D) \geq \min\Big\{I_p(|D_j|\|\lambda_j)\ ,\ I_p(|D_j|\|\mu_j)\Big\}, \quad \forall |D_j| \geq \max\{\lambda_j, \mu_j\}.$$

Now $I_P(x\|y) \sim x\log x$ when x tends to infinity, so that there exists $M(y)$ such that

$$I_p(x\|y) \geq \frac{1}{2} x \log x, \quad \forall x \geq M(y).$$

Therefore (3.15) holds with

$$M = \max_{i\in\mathcal{S}} \big\{M(\lambda_i), M(\mu_i), \lambda_i, \mu_i\big\}.$$

■

4 Large deviations for the empirical generator

It is to be stressed that the following theorem 4.1 relies on the one-to-one mapping between empirical generators and polling systems expressed in proposition 3.1, which itself depends on the linear description of the fluid limit when the polling system starts from an empty state (see theorem 2.3). Thus, we shall only consider throughout this section that the queues are initially empty[3] ($Q^\Lambda(0) = 0$). Now, by proposition 2.1, the problem of large deviations has been reduced precisely to this case.
We shall prove in this section a LDP for the combination of the empirical generator *and* the uniform variable

$$\left\{\sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right\},$$

defined for any $D \in \mathbb{R}^\Lambda$. We shall also define the projection $f_\Lambda : \mathcal{G}_s^\Lambda \mapsto \mathbb{R}^\Lambda$ by $f_\Lambda(G) = D$, and the ball $B(G,r)$ of center G and radius r. Note that $f_\Lambda(B(G,r)) = B(D,r)$, the second ball being in $\mathbb{R}^N$, since the distance is a max.

[3]For this reason, the initial condition $(0,s)$ in $Q^\Lambda(t,0,s)$ will be dropped.

Theorem 4.1 (Generator's local bounds) *Let* Λ *be a face and* $G = (\pi, A, D) \in \mathcal{G}_s^\Lambda$. *Then,*

$$\lim_{\delta \to 0} \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in B(G,\delta),\ \sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right] \tag{4.1}$$

$$= \lim_{\delta \to 0} \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in B(G,\delta),\ \sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right] \tag{4.2}$$

$$= -H(G\|R).$$

The proof will be done in four steps. First, a change of measure, so as to treat only neighborhoods of a generator G; second the lower and upper local LDP bounds for particular generators; third a continuity argument, in order to extend the bounds to all generators; fourth the exponential tightness of the measures.

4.1 Exponential change of measure

Fix $G \in \mathcal{G}_s^\Lambda$. In this section the bijection of Proposition 3.1 is used to describe G indifferently as $G = (A, \pi, D)$ or as $G = (\tilde{\lambda}_i, \tilde{\mu}_i, \tilde{\tau}_{ij}, \tilde{P}_{ij})$. In the next section, fluid limit results will be used. Since these results are restricted to irreducible polling systems, the generator G will henceforth be assumed to be irreducible. Secondly, in order for the mapping h_2 to be properly defined (see below), $\tilde{P}_{ij}$ is assumed to be null when P_{ij} is, i.e. $G \in \mathcal{G}_s^\Lambda(\mathbf{P})$.
The first step is to amount the problem to neighborhoods of G. For this purpose we introduce a change of measure. Consider a localized polling system X^Λ with generator $R = (\lambda_i, \mu_i, \tau_{ij}, P_{ij})$. Then define

- the vector $\alpha \in \mathbb{R}^N$ by $\alpha_i \stackrel{\text{def}}{=} \log \frac{\tilde{\lambda}_i}{\lambda_i}$, for all $i \in \mathcal{S}$;
- the mapping $h_1 : \mathcal{S}_0 \mapsto \mathbb{R}$ by $\begin{cases} h_1(i) \stackrel{\text{def}}{=} \alpha_i + \log \frac{\tilde{\mu}_i}{\mu_i}, \\ h_1(ij) \stackrel{\text{def}}{=} \log \frac{\tau_{ij}}{\tilde{\tau}_{ij}}, \end{cases}$ for all $i, j \in \mathcal{S}$;
- the mapping $h_2 : \mathcal{S}_0 \mapsto \mathbb{R}$ by $\begin{cases} h_2(ij) \stackrel{\text{def}}{=} \log \frac{\tilde{P}_{ij}}{P_{ij}}, \\ h_2(i) \stackrel{\text{def}}{=} 0, \end{cases}$ for all $i, j \in \mathcal{S}$;

- the compensator $K : \mathcal{S}_0 \mapsto \mathbb{R}$ by $\begin{cases} K(i) \stackrel{\text{def}}{=} \sum_{k\in\mathcal{S}} (\tilde{\lambda}_k - \lambda_k) + \tilde{\mu}_i - \mu_i, \\ K(ij) \stackrel{\text{def}}{=} \sum_{k\in\mathcal{S}} (\tilde{\lambda}_k - \lambda_k) + \tilde{\tau}_{ij}^{-1} - \tau_{ij}^{-1}, \end{cases}$

  $\forall i, j \in \mathcal{S}$;

- and the process

$$\mathcal{M}_t^\Lambda \stackrel{\text{def}}{=} \exp\left\{ < \alpha, Q^\Lambda(t,x,s) - x > + \sum_{i=0}^{N_t^\Lambda - 1} h_1(S_i^\Lambda) + h_2(S_{i+1}^\Lambda) - \int_0^t K(S_v^\Lambda)\,\mathrm{d}v \right\}.$$

Since K has been exactly defined so that[4]

$$K(s) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\exp\left\{ < \alpha, Q^\Lambda(t,x,s) - x > + \sum_{i=0}^{N_t^\Lambda - 1} h_1(S_i^\Lambda) + h_2(S_{i+1}^\Lambda) \right\} \right]_{t=0},$$

it is easily checked that the derivative of $\mathbb{E}\left[\mathcal{M}_t^\Lambda\right]$ at $t = 0$ is null. Then using the Markov property, one can get that the derivative is null for all $t \geq 0$, so that $\mathbb{E}\left[\mathcal{M}_t^\Lambda\right] = 1$. Using again the Markov property, this proves that $\mathbb{E}\left[\mathcal{M}_t^\Lambda | \mathcal{F}_s\right] = \mathcal{M}_s^\Lambda$, for all $t \geq s \geq 0$, hence $\{\mathcal{M}_t^\Lambda,\ t \geq 0\}$ is a martingale w.r.t. the natural filtration $\mathcal{F}_t$. Then define a new probability measure by

$$\tilde{\mathbb{P}}\left[B\right] \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{B\}}\mathcal{M}_t^\Lambda\right], \quad \forall B \in \mathcal{F}_t.$$

It is a matter of routine to show that under $\tilde{\mathbb{P}}$, X^Λ is again a Markov process. Moreover if $q^\Lambda(x,s;y,s')$ denotes the intensity of jump from (x,s) to (y,s') of X^Λ under $\mathbb{P}$, then the generator X^Λ under $\tilde{\mathbb{P}}$ is given by

$$\sum_{y,s'} q^\Lambda(x,s;y,s') e^{<\alpha, y-x> + \mathbb{1}_{\{s\neq s'\}}(h_1(s)+h_2(s'))} \big(f(y,s') - f(x,s)\big).$$

Hence under $\tilde{\mathbb{P}}$, X^Λ depicts the evolution of a localized polling system without boundary conditions on the nodes belonging to Λ. The arrival rate at node i is $\lambda_i e^{\alpha_i} = \tilde{\lambda}_i$ whereas the service rate is equal to $\mu_i e^{-\alpha_i + h_1(i)} = \tilde{\mu}_i$. The intensity of switch-over time and the probability of routing between i and j are respectively given by $\tau_{ij}^{-1} e^{h_1(ij)} = \tilde{\tau}_{ij}^{-1}$ and $P_{ij} e^{h_2(ij)} = \tilde{P}_{ij}$.
This is the change of measure we wanted.

[4] Note that the derivative is independent of Λ and x, so that they are dropped.

4.2 Upper and lower bounds

The ball $B(G,\delta)$ is precisely defined by

$$B(G,\delta) \stackrel{\text{def}}{=} \Big\{ G' \in \mathcal{G}(\mathbf{P}) : \ \|G'-G\|_\infty < \delta \Big\}.$$

Define the continuous mapping $\phi : \mathcal{G}(\mathbf{P}) \mapsto \mathbb{R}$ by

$$\phi(G') \stackrel{\text{def}}{=} < \alpha, D' > + \sum_{s\in\mathcal{S}_0} a'_s h_1(s) + \sum_{i,j\in\mathcal{S}} a'_{ij} h_2(ij) - \sum_{s\in\mathcal{S}_0} \pi'_s K(s). \tag{4.3}$$

Simple manipulations using (3.5)–(3.8) yield $\phi(G) = H(G\|R)$. Note that ϕ is finite everywhere as soon as $H(G\|R)$ is finite, since then there are no infinite terms in (4.3). The change of measure has been done so that

$$\log \mathcal{M}_t^\Lambda = t\phi(G_t).$$

For the sake of brevity, we shall denote by $E_t(G,\delta)$ the event

$$E_t(G,\delta) \stackrel{\text{def}}{=} \left\{ G_t \in B(G,\delta), \ \sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta \right\}.$$

Applying the change of measure and the previous relation yields

$$\begin{aligned} \frac{1}{t}\log \mathbb{P}\left[E_t(G,\delta)\right] &= \frac{1}{t}\log \tilde{\mathbb{E}}\left[\left(\mathcal{M}_t^\Lambda\right)^{-1} \mathbb{1}_{E_t(G,\delta)}\right] \\ \geq \ & - \sup_{G'\in B(G,\delta)} \phi(G') + \frac{1}{t}\log \tilde{\mathbb{P}}\left[E_t(G,\delta)\right]. \end{aligned} \tag{4.4}$$

Now, it has been shown through fluid limit that, under $\tilde{\mathbb{P}}$, G_t converges to G in probability and $Q^\Lambda(s,0)$ converges to sD uniformly for $s \in [0,t]$, so that the probability of the event $E_t(G,\delta)$ converges to 1. Moreover, by the continuity of ϕ, the first term of (4.4) tends to $\phi(G) = H(G\|R)$, hence the local lower bound,

$$\lim_{\delta\to 0} \liminf_{t\to\infty} \frac{1}{t}\log \mathbb{P}\left[E_t(G,\delta)\right] \geq -H(G\|R). \tag{4.5}$$

Then, reversing the inequality in (4.4) yields

$$\frac{1}{t}\log \mathbb{P}\left[E_t(G,\delta)\right] \leq - \inf_{G'\in B(G,\delta)} \phi(G') + \frac{1}{t}\log \tilde{\mathbb{P}}\left[E_t(G,\delta)\right],$$

which turns to the local upper bound by the same argument,

$$\lim_{\delta\to 0}\limsup_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[\tilde{\mathbb{P}}\left[E_t(G,\delta)\right]\right] \le -H(G\|R). \tag{4.6}$$

Now the irreducibility assumption will be relaxed. First note that the upper bound (4.6) is obtained by bounding the probability $\tilde{\mathbb{P}}[E_t(G,\delta)]$ by 1, thus it is valid for reducible generators in $\mathcal{G}_s^\Lambda(\mathbf{P})$.

By Lemma 3.3, the set of irreducible generators is dense in $\mathcal{G}_s^\Lambda(\mathbf{P})$. For any $G \in \mathcal{G}_s^\Lambda(\mathbf{P})$ and any $\delta > 0$, it is possible to find an irreducible $G' \in \mathcal{G}_s^\Lambda(\mathbf{P})$ such that the distance with G is less than δ. Hence there exists $\delta' > 0$ such that $B(G',\delta') \subset B(G,\delta)$. Therefore

$$\liminf_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[E_t(G,\delta)\right] \ge \liminf_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[E_t(G',\delta')\right] \ge -H(G'\|R).$$

Since this is true for any G' close to G, by continuity of H on $\mathcal{G}_s^\Lambda(\mathbf{P})$, (4.5) is also true on $\mathcal{G}_s^\Lambda(\mathbf{P})$ for reducible generator.

If $G \notin \mathcal{G}_s^\Lambda(\mathbf{P})$, by Proposition 3.2, $H(G\|P) = \infty$. Obviously (4.5) is valid. When G does not belong to $\mathcal{G}_s^\Lambda(\mathbf{P})$ which is closed, there exists $\delta > 0$ such that $B(G,\delta)$ is entirely outside $\mathcal{G}_s^\Lambda(\mathbf{P})$. The probability for such an event is null. Hence the bound (4.6) is true for all generators.

4.3 Exponential tightness

We shall prove here that the measures are exponentially tight, that is, for all $C \in \mathbb{R}$, there exists a compact set $K \subset \mathcal{G}$ such that

$$\limsup_{t\to\infty}\frac{1}{t}\log \mathbb{P}\left[G_t \in K^c\right] \le -C. \tag{4.7}$$

For a Poisson process with intensity bounded by r, the measured intensity N_t/t is bounded by,

$$\mathbb{P}\left[N_t/t > l_0\right] = e^{-rt}\sum_{n>l_0 t}\frac{(rt)^n}{n!} \le e^{-rt}\frac{(rt)^{l_0 t}}{(l_0 t)!}\frac{l_0}{l_0 - r}.$$

Using Stirling's expansion, the logarithm of the right part is shown to be about $-tl_0\log l_0$ when t and l_0 are large, which yields the exponential tightness.

Now, the intensity of jumps for the server and for the arrivals is bounded by

$$r \stackrel{\text{def}}{=} \max_{i,j\in\mathcal{S}} \left\{ \lambda_i, \mu_i, \frac{1}{\tau_{ij}} \right\}.$$

Therefore, taking l_0 big enough, $K \stackrel{\text{def}}{=} \{(A, \pi, D) : l \leq l_0\}$, where $l = \sum_{s\in\mathcal{S}_0} a_s$, is a compact set verifying the bound (4.7).

The exponential tightness allows classically to extend a weak LDP to a full LDP. We did not prove an LDP here, but the exponential tightness is required for the kind of contraction that follows.

4.4 Contraction

Since $f_\Lambda(G) = D$ is a continuous mapping from $\mathcal{G}_s^\Lambda$ into $\mathbb{R}^\Lambda$, the contraction principle applied to the local bounds (4.1)–(4.2) suggests the following contraction.

Theorem 4.2 (Local bounds) *Let Λ be a face and $D \in \mathbb{R}^\Lambda$. Then,*

$$\lim_{\delta\to 0} \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P} \left[\sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta \right] \tag{4.8}$$

$$= \lim_{\delta\to 0} \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P} \left[\sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta \right] = -L(\Lambda, D), \tag{4.9}$$

where $L(\Lambda, .)$ is the good rate function

$$L(\Lambda, D) \stackrel{\text{def}}{=} \inf_{G\in f_\Lambda^{-1}(D)} H(G\|R), \quad \forall D \in \mathbb{R}^\Lambda. \tag{4.10}$$

Proof : The sketch of the proof is exactly the same as a proof of the contraction principle, except that it is restricted around $\mathcal{G}_s^\Lambda$. First note that,

$$f_\Lambda^{-1}(B(D, \delta)) = \bigcup_{G\in f_\Lambda^{-1}(D)} B(G, \delta). \tag{4.11}$$

The lower bound is checked with a sequence $G(n)$ such that $H(G(n)\|R)$ converges to $L(\Lambda, D)$.

For the upper bound, consider the right part of (4.11). From this set, it can be extracted by the exponential tightness a compact set $K \subset \mathcal{G}$, so that

$$\begin{aligned} &\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in K,\ \sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right] \\ =\ &\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[\sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right]. \end{aligned}$$

Fix $\varepsilon > 0$. By (4.2), for all $G \in \mathcal{G}_s^\Lambda$, there exists $\delta > 0$ such that

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in B(G,\delta),\ \sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right] \leq -H(G\|R) + \varepsilon.$$

Since $K \cap f_\Lambda^{-1}(D)$ is compact, it can be covered by a finite number of balls, as in (4.11). Moreover there exists δ_0 such that

$$K \cap f_\Lambda^{-1}(B(D,\delta_0)) \subset \bigcup_{i=1}^n B(G(i),\delta(i)) \quad \text{with } f_\Lambda(G_i) = D.$$

But $B(D,\delta)$ is decreasing with δ so that the previous equation is valid for $\delta \leq \delta_0$. Finally

$$\begin{aligned} &\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left[G_t \in K,\ \sup_{s\in[0,t]} |Q^\Lambda(s) - sD| < t\delta\right] \\ \leq\ &-\min_i H(G(i)\|R) + \varepsilon \leq -L(\Lambda,D) + \varepsilon. \end{aligned}$$

This is valid for all $\varepsilon > 0$ and $\delta < \delta_0$, hence the upper bound. Since (4.8) is less than (4.9), they are both equal to $-L(\Lambda, D)$. The proof is completed. ■

5 The sample path large deviations principle

First, we verify that the rate function $I_T(.)$ possesses the usual properties. For, recall that

$$\Phi_x(K) = \{\varphi \in D([0,T],\mathbb{R}_+^N) : I_T(\varphi) \leq K, \varphi(0) = x\}.$$

Then

Proposition 5.1 *(i) Assume $I_T(\varphi) \leq K$ for some K. Then, for all $\epsilon > 0$, there exists $\delta > 0$ independent of φ such that for any collection of non overlapping intervals $[t_j, t_{j+1}]$ in $[0,T]$ with $\sum_j t_{j+1} - t_j = \delta$,*

$$\sum_j |\varphi(t_{j+1}) - \varphi(t_j)| \leq \epsilon.$$

(ii) $I_T(.)$ is lower semi-continuous in $\Big(D\big([0,T], \mathbb{R}_+^N\big), d_d\Big)$.

(iii) For any compact set $C \subset \mathbb{R}_+^N$, $\cup_{x \in C} \Phi_x(K)$ is a compact set of $\mathcal{C}([0,T], \mathbb{R}_+^N)$.

(iv) Let $\varphi \in \mathcal{AC}\big([0,T], \mathbb{R}_+^N\big)$ with $I_T(\varphi) < \infty$. Then, for all $\epsilon > 0$, it exists $\varphi_\epsilon \in \mathcal{PL}\big([0,T], \mathbb{R}_+^N\big)$ such that:

(a) $d_c(\varphi_\epsilon, \varphi) \leq \epsilon$,

(b) $I_T(\varphi_\epsilon) \leq I_T(\varphi) + \epsilon$.

Proof : As in lemma 5.18 of [19], one can prove (i) using lemma 3.6.

By (i), in order to prove the lower semi-continuity of $I_T(.)$, it is sufficient to consider sequences of absolutely continuous functions. Since on $\mathcal{C}\big([0,T], \mathbb{R}_+^N\big)$, the metrics d_c and d_d are equivalent, one can use d_c. Now, using lemma 3.6, since $L(x,D)$ is lower semi-continuous in (x,D) and convex with respect to D (see lemma 3.5 and proposition 3.4), theorem 3 of section 9.1.4 of [18] yields (i).

(iii) is a consequence of (i) and (ii) (see proposition 5.46 of [19]).

For (iv), similar results were proved in lemma 7.5.4 of [12] in the case of a random walk with discontinuity along an hyperplane. Nevertheless, a different proof is provided in section C. ■

Remark 4 *(ii) and (iv) imply that for $\varphi \in \mathcal{AC}\big([0,T], \mathbb{R}_+^N\big)$ with $I_T(\varphi) < \infty$, it exists a sequence $\{\varphi_n,\ n \geq 1\}$ with $\varphi_n \in \mathcal{PL}\big([0,T], \mathbb{R}_+^N\big)$ for all n and satisfying*

$$\lim_{n\to\infty} d_c(\varphi_n, \varphi) = 0 \text{ and } \lim_{n\to\infty} I_T(\varphi_n) = I_T(\varphi).$$

In order to prove theorem 1.4, the piecewise linear interpolation of $Q^n_{x,s}$ denoted $Z^n_{x,s}$ is needed. For, n being fixed, put

$$t_j^n = j\frac{T}{n},\ j = 0, 1, \dots, n.$$

Then $Z^n_{x,s}$ is defined by

$$Z^n_{x,s}(t) = \left(\frac{n}{T}t - j\right)Q^n_{x,s}(t^n_{j+1}) + \left((j+1) - \frac{n}{T}t\right)Q^n_{x,s}(t^n_j),\ \forall t \in [t^n_j, t^n_{j+1}].$$

The following property is needed in proving the upper bound of theorem 1.4.

Proposition 5.2 *(i) For all $\delta > 0$, uniformly in $x \in \mathbb{R}^N_+$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\sup_{t\in[0,nT]} |Z^n_{x,s}(t) - Q^n_{x,s}(t)| \geq \delta\right] = -\infty.$$

(ii) Fix C a compact set of $\mathbb{R}^N_+$. For each $0 < M < \infty$, there is a compact $\mathcal{K}$ of $\mathcal{C}\left([0,T], \mathbb{R}^N_+\right)$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[Z^n_{x,s} \in \mathcal{K}\right] \leq -M,\ \forall x \in C.$$

Proof : This is a straight adaptation of lemmas 2.4, 2.5 of [13]. ∎

Proof of Theorem 1.4 Although technical, the extension of proposition 1.3 to theorem 1.4 is rather standard using propositions 5.1 and 5.2. The lower bound is obtained using proposition 1.3 and remark 4. The upper bound is derived via propositions 1.3, 5.1 (iii) and 5.2. The reader is referred to section 5 of [11] for details. ∎

Appendix

A From local bounds to piecewise linear functions

In this section, we shall use strong properties of irreducibility of the Markov process X. Denote

$$p = \min_{s,s',y,y'} \frac{q(y,s;y',s')}{\sum_{y'',s''} q(y,s;y'',s'')} > 0.$$

Let $x \in \mathbb{Z}_+^N$. $x + e_i$ can be reached with probability larger than p in one step from x. $\mathbf{P}$ being irreducible, $x - e_i$ can be reached with probability larger than p^J in less than J steps from x. Moreover $J \leq 3N$. Then

Lemma A.1 *For all $x, y \in \mathbb{Z}_+^N$, there exists some path in $\mathbb{Z}_+^N$,*

$$i_1 = x, \dots, i_k = y$$

with length less than $J|x-y|$, i.e $k \leq J|x-y|$, and the embedded Markov chain associated to X follows this path with probability larger than $p^{J|x-y|}$. Since the jumps of X are bounded by one, the path can be chosen so that

$$|i_j - x| \leq |x-y| + J, \ \forall j = 1, \dots, k.$$

The following notation will be used in this section :

$$q_* = \min_{s,y} \sum_{y'',s''} q(y,s;y'',s'') \text{ and } q^* = \max_{s,y} \sum_{y'',s''} q(y,s;y'',s'').$$

Lemma A.2 *Let $0 \leq t_0 \leq t_1$. If $|y_0 - y_1| \leq C$, then for all x with $0 \leq x \leq \dfrac{t_1 - t_0}{J|y_1 - y_0|}$,*

$$\begin{aligned} &\mathbb{P}\left[Q(t_0) = y_0, Q(t_1) = y_1, \sup_{t \in [t_0,t_1]} |Q(t) - y_0| \leq C + J \right] \\ \geq\ & p^{J|y_1 - y_0|} \left(1 - e^{-q_* x}\right)^{J|y_1 - y_0|} e^{-q^*(t_1 - t_0)}. \end{aligned}$$

Proof : It is sufficient to exhibit some particular event which has the desirable properties. By lemma A.1, it exists some path in $\mathbb{Z}_+^N$,

$$i_1 = y_0, \dots, i_k = y_1$$

with length less than $J|y_0 - y_1|$ and the embedded Markov chain associated to X follows this path with probability larger than $p^{J|y_0-y_1|}$. Moreover

$$|i_j - y_0| \leq |y_1 - y_0| + J \leq C + J, \ \forall j = 1, \dots, k.$$

Now, the process is forced to stay in each state i_j a time less than x. This is done with probability larger than $\left(1 - e^{-q_* x}\right)^{J|y_1-y_0|}$ since there are at most $J|y_0 - y_1|$ jumps. When the process has reached y_1, it is forced to stay in y_1 a time larger than $t_1 - t_0$. This is done with probability larger than $e^{-q^*(t_1-t_0)}$. Hence, the probability of such event for X is larger than $p^{J|y_1-y_0|}\left(1 - e^{-q_* x}\right)^{J|y_1-y_0|} e^{-q^*(t_1-t_0)}$. This achieves the proof. ■

Using the preceding lemma, one can deduce some bound on the probability that X stays in some tube centered around some affine function.

Lemma A.3 *Let $0 \leq t_0 \leq t_1$, $x_0 \in \mathbb{R}_+^N$ and $D \in \mathbb{R}^N$ satisfying $x_0 + D(t - t_0) \in \mathbb{R}_+^N$, $\forall t$ with $t_0 \leq t \leq t_1$. Denote φ, the function defined by $\varphi(t) = x_0 + D(t - t_0)$. Set $x_1 \stackrel{\text{def}}{=} \varphi(t_1)$. Assume*

$$|x_0 - y_0| \wedge |x_1 - y_1| < |D|(t_1 - t_0) < \frac{C}{3}.$$

Then

$$\begin{aligned} & \mathbb{P}\left[Q(t_0) = y_0, Q(t_1) = y_1, \sup_{t \in [t_0, t_1]} |Q(t) - \varphi(t)| \leq 2C + J\right] \\ \geq \ & p^{3|D|J(t_1-t_0)}\left(1 - e^{-\frac{q_*}{3|D|J}}\right)^{3|D|J(t_1-t_0)} e^{-q^*(t_1-t_0)} \stackrel{\text{def}}{=} \psi(D)^{t_1-t_0}. \end{aligned}$$

Proof : We have

$$|y_1 - y_0| < 3|D|(t_1 - t_0) < C. \tag{A.1}$$

If $0 \leq x \leq \dfrac{t_1 - t_0}{J|y_1 - y_0|}$, lemma A.2 and the triangular inequality imply

$$\begin{aligned} & \mathbb{P}\left[Q(t_0) = y_0, Q(t_1) = y_1, \sup_{t \in [t_0, t_1]} |Q(t) - \varphi(t)| \leq 2C + J\right] \\ \geq \ & p^{J|y_1-y_0|}\left(1 - e^{-q_* x}\right)^{J|y_1-y_0|} e^{-q^*(t_1-t_0)}, \\ \geq \ & p^{3|D|J(t_1-t_0)}\left(1 - e^{-q_* x}\right)^{3|D|J(t_1-t_0)} e^{-q^*(t_1-t_0)}, \end{aligned}$$

where the last inequality follows by (A.1) and since $0 < p < 1$ and $0 < 1 - e^{-q_* x} < 1$. The proof is concluded remarking that (A.1) implies $\dfrac{1}{3|D|J} \le \dfrac{t_1 - t_0}{J|y_1 - y_0|}$. ■

Proof of proposition 1.2 If $x + Dt \in \Lambda(D)$, $\forall t \in [0, \tau]$, then the result is nothing than theorem 1.1. Nevertheless, it is known only that $x + Dt \in \Lambda(D)$ in the open interval $]0, \tau[$. In order to circumvent this difficulty, fix Δ positive with $2\Delta < \tau$. In $[n\Delta, n(\tau - \Delta)]$, we will apply theorem 1.1 and then let Δ tends to 0. More precisely, if $\delta' > \delta$, then

$$
\begin{aligned}
& \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
\le\ & \mathbb{P}\left[\sup_{t\in[0,n\Delta]} |Q(t,y) - nx - Dt| < \delta n, \sup_{t\in[n\Delta,n(\tau-\Delta)]} |Q(t,y) - nx - Dt| < \delta' n\right].
\end{aligned}
$$

Letting n tends to infinity and then successively ϵ, δ, δ' to 0, theorem 1.1 yield

$$
\begin{aligned}
& \lim_{\delta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\sup_{|y-nx|<\epsilon n}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
\le\ & -(\tau - 2\Delta)L\big(\Lambda(D), D\big).
\end{aligned}
$$

Letting Δ tends to 0, one obtains

$$
\begin{aligned}
& \lim_{\delta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\sup_{|y-nx|<\epsilon n}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
.\ \le\ & -\tau L\big(\Lambda(D), D\big)
\end{aligned}
\tag{A.2}
$$

Let us prove

$$
\begin{aligned}
& \lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\inf_{|y-nx|<\epsilon n}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
\ge\ & -\tau L\big(\Lambda(D), D\big).
\end{aligned}
\tag{A.3}
$$

In $[n\Delta, n(\tau - \Delta)]$, we will apply theorem 1.1 and in $[0, n\Delta] \cup [n(\tau - \Delta), n\tau]$, lemma A.3. For, put $2C_n + J = \delta n$. Moreover, taking into account the order in which the limits are taken, assume without loss of generality that $J \le \dfrac{\delta n}{2}$. In order to use lemma A.3, $[0, n\Delta]$ will be divided in K_δ intervals, $\left[\frac{(i-1)\Delta n}{K_\delta}, \frac{i\Delta n}{K_\delta}\right]$, $i = 1, \dots, K_\delta$, with $K_\delta = \left[\frac{12\Delta|D|}{\delta}\right] + 1$. We apply lemma A.3 on each interval $\left[\frac{(i-1)\Delta n}{K_\delta}, \frac{i\Delta n}{K_\delta}\right]$ with $2C_n + J$ and $n\Delta/K_\delta$ replacing respectively $2C + J$ and $t_1 - t_0$. Note that

$$|D|\frac{n\Delta}{K_\delta} \le \frac{\delta n}{12} \le \frac{C_n}{3}.$$

Now, each process $\{Q(nt, y, s),\ t \ge 0\}$ is forced to be at time $\frac{i\Delta n}{K_\delta}$ at some point $x_i^{(N)}$, $i = 0, \dots, K_\delta$. $x_i^{(N)}$ is chosen in such a way that

$$\begin{aligned} x_0^{(N)} &= y, \\ \left|x_i^{(N)} - x - D\frac{i\Delta n}{K_\delta}\right| &\le N,\ i = 1, \dots, K_\delta. \end{aligned}$$

Taking into account the order in which the limits are taken, assume

$$\epsilon n \le |D|\frac{n\Delta}{K_\delta} \le \frac{\delta n}{12} \le \frac{C_n}{3}.$$

Then, lemma A.3 and the Markov property give

$$\begin{aligned} &\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\ \ge\ &\psi(D)^{\Delta n}\mathbb{P}\left[|Q(\Delta n) - x + D\Delta n| < N, \sup_{t\in[n\Delta,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right]. \end{aligned}$$

Let $\delta' < \delta$ satisfies

$$\delta' n \le |D|\frac{n\Delta}{K_\delta} \le \frac{\delta n}{12} \le \frac{C_n}{3}.$$

Again lemma A.3 and the Markov property give

$$\begin{aligned} &\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\ \ge\ &\psi(D)^{2\Delta n}\mathbb{P}\left[|Q(\Delta n) - x + D\Delta n| < N, \sup_{t\in[n\Delta,n(\tau-\Delta)]} |Q(t,y) - nx - Dt| < \delta' n\right]. \end{aligned}$$

Letting n tends to infinity and then successively ϵ, δ', δ to 0, one gets using Theorem 1.1

$$\begin{aligned} &\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\inf_{|y-nx|<\epsilon n}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right] \\ \geq\ & 2\Delta\log\psi(D)-L(\Lambda(D),D)(\tau-2\Delta). \end{aligned}$$

Let Δ tends to 0 in order to obtain (A). Then (A.2), (A) give the result. ■

Proof of proposition 1.3 This can be derived in a straight way using proposition 1.2 and the Markov property. ■

B Proof of proposition 2.1

Let us prove (2.1). Take τ satisfying $x_i + D_i\tau > 0$, $\forall i \in \Lambda(x)$. Then for ϵ sufficiently small and all y with $|y - nx| < \epsilon$

$$\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right]=\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,y)-nx-Dt|<\delta n\right].$$

Then, note that

$$\begin{aligned} &\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\inf_{|y-nx|<n\epsilon}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,y)-nx-Dt|<\delta n\right] \\ =\ &\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\inf_{|y-nx|<n\epsilon}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,y)-nx-Dt|<\delta n\right], \\ =\ &\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\inf_{|y|<n\epsilon}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,y)-Dt|<\delta n\right], \end{aligned}$$

where the last relation holds because the distribution of $X^{\Lambda}_{y,s}$ is invariant with respect to the shift nx since $x_i = 0$, $\forall i \in \Lambda(x)^c$. Hence, it is sufficient to prove that

$$\begin{aligned} &\lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\inf_{|y|<n\epsilon}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,y)-Dt|<\delta n\right] \\ &\geq \lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\log \mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q^{\Lambda(x)}(t,0)-Dt|<\delta n\right]. \end{aligned}$$

Taking into account the order in which the limits are taken, assume without loss of generality that $\epsilon n + J \le \frac{\delta n}{2}$. Hence, by lemma A.2

$$\inf_{|y|<n\epsilon} \mathbb{P}\left[Q^{\Lambda}(\epsilon n, y, s) = 0, \sup_{t\in[0,\epsilon n]} |Q^{\Lambda}(t,y,s)| \le \frac{n\delta}{2}\right] \tag{B.1}$$
$$\ge \; p^{J\epsilon n}\Big(1 - e^{-\frac{q_*}{J}}\Big)^{J\epsilon n} e^{-q^*\epsilon n} \stackrel{\text{def}}{=} \tilde{\gamma}^{\epsilon n}.$$

Finally, if $|n\epsilon D| < n\delta/2$, this being in no way a restriction, (B.1) and the Markov property yield

$$\begin{aligned}
&\inf_{|y|<n\epsilon} \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q^{\Lambda}(t,y) - Dt| < \delta n\right] \\
&\quad\ge \inf_{|y|<n\epsilon} \mathbb{P}\left[Q^{\Lambda}(\epsilon n, y, s) = 0, \sup_{t\in[0,\epsilon n]} |Q^{\Lambda}(t,y,s)| \le \frac{n\delta}{2}\right] \\
&\qquad\qquad \times \mathbb{P}\left[\sup_{t\in[0,n(\tau-\epsilon)]} |Q^{\Lambda}(t,0,s) - Dt - D\epsilon n| \le n\delta\right], \\
&\quad\ge \tilde{\gamma}^{\epsilon n}\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q^{\Lambda}(t,0,s) - Dt| \le \frac{n\delta}{2}\right].
\end{aligned}$$

Taking the different limits, one can get (2.1). The proof of (2.2) is similar and is therefore omitted. ■

C Proof of proposition 5.1 (iv)

First, one can assume that $|\dot{\varphi}(t)| \le K$, a.e, for some K. Indeed,

Lemma C.1 *Let $\varphi \in \mathcal{AC}\big([0,T], \mathbb{R}_+^N\big)$ with $I_T(\varphi) < \infty$. Then, for all $\epsilon > 0$, it exists $K_\epsilon > 0$, $\varphi_\epsilon \in \mathcal{AC}\big([0,T], \mathbb{R}_+^N\big)$ such that:*

(i) $|\dot{\varphi}_\epsilon(t)| \le K_\epsilon$, *a.e,*

(ii) $d_c(\varphi_\epsilon, \varphi) \le \epsilon$,

(iii) $I_T(\varphi_\epsilon) \le I_T(\varphi) + \epsilon$.

Proof : This can be done by introducing a time change as in lemma 6.5.3 of [12]. The proof is omitted. ■

In order to prove the assertion, it is sufficient to construct a sequence $\{\psi^{(n)}, n \geq 1\}$ such that
$\psi^{(n)} \in \mathcal{PL}\big([0,T], \mathbb{R}_+^N\big)$ and with the following properties :

- (i) $\lim_{n\to\infty} d_c(\psi^{(n)}, \varphi) = 0$,
- (ii) $\displaystyle\lim_{n\to\infty} \int_0^T \mathbb{I}_{\{\Lambda(\psi^{(n)}(t)) \neq \Lambda(\varphi(t))\}} dt = 0$,
- (iii) $|\dot{\psi}^{(n)}(t)| \leq K$, a.e.
- (iv) $\lim_{n\to\infty} \dot{\psi}^{(n)}(t) = \dot{\varphi}(t)$, a.e.

Indeed, if these assumptions are fulfilled, then

$$\begin{aligned} |I_T(\psi^{(n)}) - I_T(\varphi)| \quad &\leq \quad \sup_{\Lambda} \int_0^T |L(\Lambda, \dot{\psi}^{(n)}(t)) - L(\Lambda, \dot{\varphi}(t))| dt \\ &+ \quad 2 \sup_{x \leq K} \sup_{\Lambda} \big(L(\Lambda, x)\big) \int_0^T \mathbb{I}_{\{\Lambda(\psi^{(n)}(t)) \neq \Lambda(\varphi(t))\}} dt. \end{aligned}$$

Since $L(\Lambda, x)$ is continuous with respect to x, the first term tends to 0 using properties (iii), (iv) and dominated convergence theorem. The second one tends to 0 by (ii).

In a first step, one constructs a sequence $\{\varphi^{(n)}, n \geq 1\}$ with $\varphi^{(n)} \in \mathcal{PL}\big([0,T], \mathbb{R}_+^N\big)$ satisfying (i), (ii), (iii). As it will emerge, one can prove that $\dot{\varphi}^{(n)}$ tends to $\dot{\varphi}$ only for the weak topology in $\mathcal{L}^1\big([0,T], \mathbb{R}^N\big)$. In a second step, the use of Mazur's theorem, see corollary 14 of [9] page 422, allows one to construct from $\{\varphi^{(n)}, n \geq 1\}$ a sequence $\{\psi^{(n)}, n \geq 1\}$ satisfying (i) to (iv). More precisely :

Lemma C.2 *It exists $\{\varphi^{(n)}, n \geq 1\}$ with $\varphi^{(n)} \in \mathcal{PL}\big([0,T], \mathbb{R}_+^N\big)$ satisfying :*

- (i) $\lim_{n\to\infty} d_c(\varphi^{(n)}, \varphi) = 0$.
- (ii) *The sequence $\{\varphi^{(n)}, n \geq 1\}$ satisfies*

$$\lim_{n\to\infty}\uparrow \big\{t \in [0,T] : \varphi_k^{(n)}(t) > 0\big\} = \big\{t : \varphi_k(t) > 0\big\}, \quad \forall k \in \mathcal{S}.$$

As a consequence

$$\begin{aligned} & \lim_{n\to\infty}\!\downarrow \{t \in [0,T] : \Lambda(\varphi(t)) \neq \Lambda(\varphi^{(n)}(t))\} \\ = & \lim_{n\to\infty}\!\downarrow \bigcup_{k\in\mathcal{S}} \Big(\{t : \varphi_k(t) > 0\} / \{t : \varphi_k^{(n)}(t) > 0\} \Big) \end{aligned}$$

and

$$\lim_{n\to\infty} \int_0^T \mathbb{I}_{\{\Lambda(\varphi(t))\neq\Lambda(\varphi^{(n)}(t))\}} dt = 0.$$

(iii) $|\dot{\varphi}_k^{(n)}(t)| \leq K$*, a.e,* $\forall k \in \mathcal{S}$*.*

(iv) $\dot{\varphi}^{(n)}$ *tends to* $\dot{\varphi}$ *for the weak topology in* $\mathcal{L}^1\big([0,T],\mathbb{R}^N\big)$*.*

Proof : $\{\varphi^{(n)}, n \geq 1\}$ is constructed component by component. φ being continuous, $O_k \stackrel{\text{def}}{=} \{t : \varphi_k(t) > 0\}$ is open and can be written as a countable union of disjoint open intervals,

$$O_k = \sum_{i\geq 0} \Big(\big] a_k(i), b_k(i) \big[\Big),$$

with $\varphi_k\big(a_k(i)\big) = \varphi_k\big(b_k(i)\big) = 0$. In the following, $O_k^{(n)}$ will stand for the partial sum $\sum_{i=0}^n \Big(\big] a_k(i), b_k(i) \big[\Big)$. Denote

$$\eta_k = \int_0^T \mathbb{I}_{\{O_k\}}(t)dt \quad \text{and} \quad \eta_k^{(n)} = \int_0^T \mathbb{I}_{\{O_k^{(n)}\}}(t)dt.$$

In order to define $\varphi^{(n)}$ in $O_k^{(n)}$, take some grid $\{t_k^{(n)}(i)\}_i$ in $O_k^{(n)}$. Then in $O_k^{(n)}$, $\varphi^{(n)}$ is defined as the interpolation of φ with respect to the grid $\{t_k^{(n)}(i)\}_i$. Moreover, owing to the continuity of φ, it is possible to choose the grid in such a way that

$$|\varphi_k^{(n)}(t) - \varphi_k(t)| \leq \frac{1}{n}, \quad \forall t \in O_k(n).$$

For $t \notin O_k^{(n)}$, put $\varphi_k^{(n)}(t) = 0$. Besides, since $\varphi_k\big(a_k(i)\big) = 0$ and $|\dot{\varphi}_k(t)| \leq K$ a.e, for $t \in O_k / O_k^{(n)}$,

$$\begin{aligned} |\varphi_k^{(n)}(t) - \varphi_k(t)| &= |\varphi_k(t)| \leq \sum_{i>n} \mathbb{I}_{\{t\in]a_k(i),b_k(i)[\}} \Big(\varphi_k\big(a_k(i)\big) + \int_{a_k(i)}^t \dot{\varphi}_k(u)du \Big), \\ &\leq K \int_0^T \mathbb{I}_{\{O_k/O_k^{(n)}\}}(u)du = K(\eta_k - \eta_k^{(n)}), \end{aligned}$$

where $\eta_k - \eta_k^{(n)}$ goes to 0 as n tends to infinity. Hence, the preceding discussion implies (i).

On a other hand, it is clear by construction that $\left\{t \in [0,T] : \varphi_k^{(n)}(t) > 0\right\} = O_k^{(n)} \uparrow O_k$ with n. Hence (ii) follows.

Since $\varphi_k^{(n)}$ is the interpolation of φ_k and $|\dot\varphi_k|$ is bounded by K a.e, one has (iii).

Since $|\dot\varphi^{(n)}(t)| \leq K$ a.e, for all n and $L(\Lambda, D)$ is continuous with respect to D, $I_T(\varphi^{(n)})$ is uniformly bounded. Using lemma 3.6, it is clear that

$$\lim_{r\to\infty} \sup_n \int_0^T \mathbb{1}_{\{|\dot\varphi^{(n)}(t)|\geq r\}} |\dot\varphi^{(n)}(t)| dt = 0.$$

Hence, quite classically, one obtains

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \int_0^T \mathbb{1}_{\{A\}}(t)|\dot\varphi^{(n)}(t)|dt < \epsilon, \forall n \text{ if } \int_0^T \mathbb{1}_{\{A\}}(t)dt \leq \delta.$$

Then, Dunford-Pettis theorem (see corollary 11 page 294 of [9]) implies that $\{\dot\varphi^{(n)}, n \geq 1\}$ is relatively compact for the weak topology in $\mathcal{L}^1\big([0,T], \mathbb{R}^N\big)$. Let f be a limiting point of $\{\dot\varphi^{(n)}, n \geq 1\}$. First,

$$\lim_{n\to\infty} \varphi^{(n)}(t_2) - \varphi^{(n)}(t_1) = \lim_{n\to\infty} \int_{t_1}^{t_2} \dot\varphi^{(n)}(s)ds = \int_{t_1}^{t_2} f(s)ds, \forall t_1, t_2.$$

Since $\lim_{n\to\infty} d_c(\varphi^{(n)}, \varphi) = 0$, one has

$$\varphi(t_2) = \varphi(t_1) + \int_{t_1}^{t_2} f(s)ds, \forall t_1, t_2.$$

Radon Nikodym's theorem ensures that $f(t) = \dot\varphi(t)$ a.e. So $\{\dot\varphi^{(n)}, n \geq 1\}$ converges to $\dot\varphi$ for the weak topology in $\mathcal{L}^1\big([0,T], \mathbb{R}^N\big)$. ■

Now, let us construct $\{\psi^{(n)}, n \geq 1\}$ from $\{\varphi^{(n)}, n \geq 1\}$. Since $\{\dot\varphi^{(n)}, n \geq 1\}$ tends to $\dot\varphi$ for the weak topology in $\mathcal{L}^1\big([0,T], \mathbb{R}^N\big)$, Mazur's theorem implies that it exists a sequence $\{f^{(n)}, n \geq 1\}$, each $f^{(n)}$ being a convex combination of the $\dot\varphi^{(n)}$ such that $\{f^{(n)}, n \geq 1\}$ converges strongly to $\dot\varphi$ in $\mathcal{L}^1\big([0,T], \mathbb{R}^N\big)$. Denote

$$f^{(n)} = \sum_{k=1}^{s(n)} \alpha_{nk} \dot\varphi^{(m_{nk})} \text{ with } \sum_{k=1}^{s(n)} \alpha_{nk} = 1, \alpha_{nk} \geq 0.$$

Set

$$\psi^{(n)} = \sum_{k=1}^{s(n)} \alpha_{nk} \varphi^{(m_{nk})}, \forall n.$$

$\psi^{(n)} \in \mathcal{PL}\big([0,T], \mathbb{R}_+^N\big)$ and clearly $\dot{\psi}^{(n)} = f^{(n)}$. Hence, possibly extracting some subsequence, one can suppose that $\lim_{n\to\infty} \dot{\psi}^{(n)} = \dot{\varphi}$ a.e. Without loss of generality, one can assume that $\lim_{n\to\infty} \min_k m_{nk} = \infty$. Then $\lim_{n\to\infty} d_c(\psi^{(n)}, \varphi) = 0$. Moreover, it is clear that $|\dot{\psi}^{(n)}(t)| \le K$, a.e. Finally, if one denotes $m_n^* = \max_k m_{nk}$, using (ii) of lemma C.2, one has

$$\big\{t \in [0,T] : \Lambda(\psi^{(n)}(t)) \ne \Lambda(\varphi(t))\big\} = \big\{t \in [0,T] : \Lambda(\varphi^{(m_n^*)}(t)) \ne \Lambda(\varphi(t))\big\}.$$

Hence, again by (ii) of lemma C.2,

$$\lim_{n\to\infty} \int_0^T \mathbb{1}_{\{\Lambda(\psi^{(n)}(t)) \ne \Lambda(\varphi(t))\}} \, dt = 0.$$

The construction of $\{\psi^{(n)}, n \ge 1\}$ is achieved. ■

References

[1] Atar, R., and Dupuis, P. Large deviations and queueing networks: methods for rate function identification. *Stochastic Process. Appl. 84*, 2 (1999), 255–296.

[2] Borovkov, A. A., and Schassberger, R. Ergodicity of a polling network. *Stochastic Process. Appl. 50*, 2 (1994), 253–262.

[3] Boxma, O., and Weststrate, J. Waiting time in polling systems with markovian server routing. In *Messung, Modellierung und Bewertung von Rechensysteme* (1989), Springer-Verlag, pp. 89–104.

[4] de La Fortelle, A. Large deviation principle for markov chains in continuous time. Tech. Rep. 3877, INRIA, 2000.

[5] de La Fortelle, A., and Fayolle, G. Large deviation principle for markov chains in discrete time. Tech. Rep. 3791, INRIA, 1999.

[6] Dobrushin, R. L., and Pechersky, E. A. Large deviations for tandem queueing systems. *J. Appl. Math. Stochastic Anal. 7*, 3 (1994), 301–330.

[7] Dobrushin, R. L., and Pechersky, E. A. Large deviations for random processes with independent increments on infinite intervals. In *Probability theory and mathematical statistics (St. Petersburg, 1993)*. Gordon and Breach, Amsterdam, 1996, pp. 41–74.

[8] Down, D. On the stability of polling models with multiple servers. *J. Appl. Probab. 35*, 4 (1998), 925–935.

[9] Dunford, N., and Schwartz, J. T. *Linear Operators. I. General Theory.* Interscience Publishers, Inc., New York, 1958. With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7.

[10] Dupuis, P., and Ellis, R. S. Large deviations for Markov processes with discontinuous statistics. II. Random walks. *Probab. Theory Related Fields 91*, 2 (1992), 153–194.

[11] Dupuis, P., and Ellis, R. S. The large deviation principle for a general class of queueing systems. I. *Trans. Amer. Math. Soc. 347*, 8 (1995), 2689–2751.

[12] Dupuis, P., and Ellis, R. S. *A weak convergence approach to the theory of large deviations.* John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[13] Dupuis, P., Ellis, R. S., and Weiss, A. Large deviations for Markov processes with discontinuous statistics. I. General upper bounds. *Ann. Probab. 19*, 3 (1991), 1280–1297.

[14] Dupuis, P., Ishii, H., and Soner, H. M. A viscosity solution approach to the asymptotic analysis of queueing systems. *Ann. Probab. 18*, 1 (1990), 226–255.

[15] Fricker, C., and Jaibi, M. Stability of multi-server polling models. Tech. Rep. 3347, INRIA, 1998.

[16] Ignatiouk-Robert, I. Large deviations of jackson networks. Tech. Rep. 14/99, Université de Cergy-Pontoise, 1999.

[17] Ignatyuk, I. A., Malyshev, V. A., and Shcherbakov, V. V. The influence of boundaries in problems on large deviations. *Uspekhi Mat. Nauk 49*, 2(296) (1994), 43–102.

[18] Ioffe, A. D., and Tihomirov, V. M. *Theory of extremal problems.* North-Holland Publishing Co., Amsterdam, 1979. Translated from the Russian by Karol Makowski.

[19] Shwartz, A., and Weiss, A. *Large deviations for performance analysis.* Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[20] Takagi, H. Queuing analysis of polling models. *ACM Comput. Surveys 20*, 1 (1988), 5–28.

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