Large Deviations Problems for Star Networks: the Min Policy Part I: Finite Time

Franck Delcoigne, Arnaud de la Fortelle

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Large deviations problems for star networks: the min policy Part I: Finite time

Franck Delcoigne — Arnaud de La Fortelle

N° 4143

Mars 2001

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4143--FR+ENG

Large deviations problems for star networks: the min policy
Part I: Finite time

Franck Delcoigne* , Arnaud de La Fortelle†

Thème 1 — Réseaux et systèmes
Projet MEVAL

Rapport de recherche n° 4143 — Mars 2001 — 33 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 connections at time t in a star network where the bandwidth is shared between customers according to the so-called min policy. The rate function is computed explicitly. One of the main steps consists in deriving large deviation bounds for an *empirical generator* constructed from the join number of customers and arrivals on each route. The rest of the analysis relies on a suitable change of measure together with a localization procedure.

Key-words: Large deviations, empirical generator, change of measure, contraction principle, entropy, star network, bandwidth sharing, min protocol.

This work is a part of the contract N^o 981B016 between INRIA and France Telecom R&D.

* Postal address: France Telecom R&D - DAC/OAT - 38-40, rue du general de Gaulle - 92794 Issy-les-Moulineaux - FRANCE.

† Postal address: INRIA - Domaine de Voluceau, Rocquencourt - BP 105 - 78153 Le Chesnay Cedex - FRANCE. e-mail address: `Arnaud.De_La_Fortelle@inria.fr`. This research is supported by the French Ministère de l'Équipement.

Problème de grandes déviations pour les réseaux en étoile sous la politique min
Partie I : temps fini

Résumé : Dans cet article, nous prouvons un principe de grandes déviations pour un processus renormalisé $n^{-1}Q_{nt}$, où Q_t représente le nombre de connexions au temps t dans un réseau en étoile où la bande passante est partagée entre les clients selon la politique min. La fonctionnelle d'action est calculée explicitement. Une des principales étapes consiste à dériver des bornes locales de grandes déviations pour un processus appelé générateur empirique, décrivant sur chaque routes le nombre de connexions présentes et le nombre de connexions qui se sont établies entre 0 et t. L'analyse repose sur un changement de mesure adéquat et une procédure de localisation.

Mots-clés : Grandes déviations, générateur empirique, changement de mesure, contraction, entropie, réseau en étoile, partage de bande passante, protocole min.

1 Introduction

The model Consider a star shaped network consisting of N channels linked to the other $N-1$ through a central hub: there are $N(N-1)/2$ routes of length two. In the sequel, the set of channels is denoted by $\mathcal{S} = \{1, \ldots, N\}$ whereas the set of routes is simply the set of ordered two-uples ij[1], $i, j \in \mathcal{S}$. Denote by $q_{ij}(t)$ (resp. $q_i(t)$) the number of calls on route ij (resp. the number of calls involving channel i) at time t. Each channel has a capacity or a bandwidth equal to C_i. Note that $q_i(t) = \sum_j q_{ij}(t)$. Then $Q(t,x) = (q_{ij}(t), i, j \in \mathcal{S})$ represents the state of the network at time t when it starts initially from state x. For the sake of simplicity, we shall sometimes omit x or t when they do not play a role.
Calls arrive on route ij according to a Poisson process of rate λ_{ij}. We shall denote by $\mathcal{R}$ the set of active routes, i.e. with $\lambda_{ij} > 0$. The duration of a call on route ij is supposed to be exponentially distributed with parameter μ_{ij}. Each call on route ij is allocated a portion $\nu_{ij}(x)/x_{ij}$ of the bandwidth when the state of the network is x. Hence a call on route ij is released at rate $\mu_{ij}\nu_{ij}(x)$. There are several possibilities in order to allocate a fair proportion of the bandwidth to customers. A classical one is to choose the coefficients $\nu_{ij}(x)$ according to the max-min fairness allocation. It consists in maximizing the smaller bandwidth dedicated to a call, $\nu_{ij}(x)/x_{ij}$, under the constraints

$$\sum_j \nu_{ij}(x) \le C_i, \qquad \forall i \in \mathcal{S}.$$

This network is proposed as a model for a router where the bandwidth is shared fairly between calls. However, the max-min fairness allocation is not explicit and hard to analyze at first. In order to get a more tractable model, we focus on the min-policy,

$$\nu_{ij}(x) = \begin{cases} x_{ij} \dfrac{C_i}{x_i} \wedge \dfrac{C_j}{x_j}, & \text{if } x_{ij} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

It has been shown in [9] that the system under the max-min fairness allocation is stochastically smaller than the one with the min policy. Hence, the min policy represents a conservative approximation to max-min fairness.

Extensions We present in this paper a sample path large deviation principle for ergodic star shaped networks under the min policy. We would like to emphasize what

[1] For the sake of simplicity, we do not distinguish between ij and ji (i.e. we consider non-oriented routes), but there is no additional difficulty to handle oriented routes.

kind of further results we aim at deducing from this theorem. First, it seems that the optimal paths of large deviation can be calculated, leading to explicit expressions for the asymptotics of stationary distribution (which is not known). This is a performance criteria of practical value: bounds for buffer size could be optimized, or simulation accuracy (trough importance sampling using optimal paths) could be improved.

Besides, the ergodicity of the network is not crucial. Large deviations can be proved for transient networks, at the cost of some more detailed analysis. This is an important feature since it is linked with the study of networks under max-min fair allocation (or similar ones). The reason is that, for an *ergodic* network under max-min fair allocation, when some routes are made saturated, the rest of the routes can behave as a *transient* network, still under max-min fair allocation: the local rate function must include the cost for this transient network to stay near 0. This is to the opposite of our framework, where only ergodic network are considered, for which the cost to stay around 0 is null.

Moreover, the topology of the network can be extended as well as the length of the routes (but not arbitrarily) to include more realistic networks. However, the notation becomes very heavy and our aim is to present tools (extending those developed for polling networks [3]) in a fairly simple way for achieving the above program.

Notation In our settings,

$$Q = \big\{Q(t, x_0),\ t \geq 0\big\}$$

is a Markov process with generator R such that

$$Rf(x) = \sum_{y \in \mathbb{Z}_+^{\mathcal{R}}} q(x,y)\Big(f(y) - f(x)\Big), \qquad \forall x \in \mathbb{Z}_+^{\mathcal{R}}, \quad \forall f \in \mathcal{B}\left(\mathbb{Z}_+^{\mathcal{R}}\right),$$

$$\text{where} \qquad q(x,y) \stackrel{\text{def}}{=} \begin{cases} \lambda_{ij}, & \text{if } y - x = e_{ij}, \\ \mu_{ij}(x) \stackrel{\text{def}}{=} \mu_{ij} x_{ij} \dfrac{C_i}{x_i} \wedge \dfrac{C_j}{x_j}, & \text{if } y - x = -e_{ij}, \\ 0, & \text{otherwise,} \end{cases}$$

using the convention that $0/0 = 0$ (i.e. when $x_{ij} = 0$). Let us recall that it has been shown in [9] that the network is ergodic if, and only if,

$$\sum_j \frac{\lambda_{ij}}{\mu_{ij}} < C_i,\ \forall i \in \mathcal{S}. \tag{1.1}$$

- For any set A, A^c will denote its complementary and $\mathbb{1}_{\{A\}}$ its indicator function;
- for any space E, $\mathcal{B}(E)$, represents the set of bounded functions on E;
- $D\left([0,T],\mathbb{R}_+^{\mathcal{R}}\right)$ is the space of right continuous functions $f\ :\ [0,T]\to\mathbb{R}_+^{\mathcal{R}}$ with left limits, endowed with the Skorokhod metric denoted by d_d;
- $\mathcal{C}\left([0,T],\mathbb{R}_+^{\mathcal{R}}\right)$ is the space of continuous functions equipped with the metric of the uniform convergence denoted by d_c;
- $\mathcal{AC}\left([0,T],\mathbb{R}_+^{\mathcal{R}}\right)$ is the space of absolutely continuous functions;
- $\mathcal{PL}\left([0,T],\mathbb{R}_+^{\mathcal{R}}\right)$ is the set of piecewise linear functions whose derivative has only finitely many discontinuities.

The relation between these sets is: $\mathcal{PL}\subset\mathcal{AC}\subset\mathcal{C}\subset D$.

Definition 1.1 (Face) *For $x\in\mathbb{R}_+^{\mathcal{R}}$, the face $\Lambda(x)$ is defined by:*

$$\Lambda(x)\stackrel{\text{def}}{=}\{ij\in\mathcal{R}:x_{ij}>0\}.$$

By an abuse of notation, the following subset of $\mathbb{R}_+^{\mathcal{R}}$ will be also called face Λ:

$$\left\{y\in\mathbb{R}_+^{\mathcal{R}}:y_{ij}>0,\ \forall ij\in\Lambda,\quad and\quad y_{ij}=0,\ \forall ij\in\Lambda^c\right\}\tag{1.2}$$

From a face Λ, a partition of the routes (see Figure 1) is defined by Λ and

$$\begin{aligned}\Lambda_1&\stackrel{\text{def}}{=}\{ij\in\Lambda^c:\exists k\in\mathcal{S},\ ik\in\Lambda\ or\ jk\in\Lambda\}.\\ \Lambda_2&\stackrel{\text{def}}{=}\{ij\in\Lambda^c:\forall k\in\mathcal{S},\ ik\notin\Lambda\ and\ jk\notin\Lambda\}.\end{aligned}$$

We also define the vector space relative to Λ

$$\mathbb{R}^{\Lambda}\stackrel{\text{def}}{=}\left\{y\in\mathbb{R}^{\mathcal{R}}:y_{ij}=0,\ \forall ij\in\Lambda^c\right\}.$$

Structure of the paper In this paper, we aim at deriving a sample path large deviation principle or LDP for the rescaled process $n^{-1}Q_{nt}$ on finite intervals of time. Our main concern is to identify the rate function. The major difficulty comes from the fact that the coefficients of the generator are not spatially continuous (the service rate $\mu_{ij}(x)$). It seems that one of the first paper dealing with large deviations for processes with discontinuous statistics is [8] where the case of Jackson networks

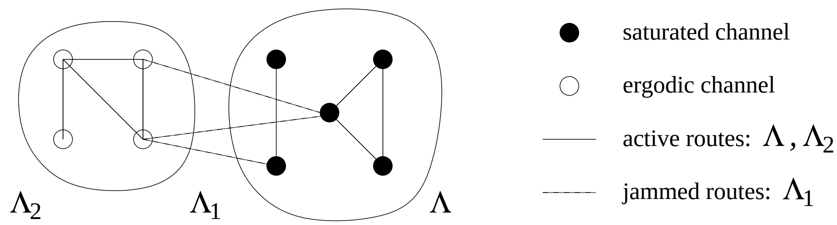

Figure 1: Representation of a star-shaped network: lines symbolize routes using two channels (circles at the ends of the lines). The routes are partitioned into saturated routes (Λ), jammed routes (Λ_1) — the service rate being null on these routes — and ergodic routes (Λ_2).

was investigated using partial differential equations techniques. In [7], a sample path LDP is proved for a wide class of jump Markov processes with discontinuous statistics. However, the methodology of proof uses subadditivity arguments and the rate function is not identified. The identification of the rate function in this general framework is still an open problem when the dimension of the network is arbitrary. General results were obtained in [6, 11] where the LDP has been established. Nevertheless, in such examples, there are at most two boundaries with codimension one or two where discontinuity arises. Using special features of the models and the fact that fluid limits could be completely identified, this program was carried out for example in [1, 3, 10].
In order to establish a sample path LDP, one of the main task is to establish the local linear large deviation bounds of Theorem 1.1.

Theorem 1.1 *Assume that Q is ergodic and let $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$. Then, writing $\lim_{\tau,\delta,\epsilon\to 0}$ for $\lim_{\tau\to 0}\lim_{\delta\to 0}\lim_{\epsilon\to 0}$,*

$$\lim_{\tau,\delta,\epsilon\to 0} \inf_{|y-nx|<\epsilon n} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \tag{1.3}$$
$$= \lim_{\tau,\delta,\epsilon\to 0} \sup_{|y-nx|<\epsilon n} \limsup_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right].$$

Moreover, if a face Λ and a drift $D \in \mathbb{R}^\Lambda$ are fixed, then the preceding limit in τ is uniform w.r.t to x in compact sets of Λ (see Definition 1.1). The common value of *these limits is denoted by* $-L(x,D)$ *and*

$$L(x,D) = \sum_{ij\in\Lambda(x)\cup\Lambda_1(x)} l\big(D_{ij}\|\lambda_{ij},\mu_{ij}(x)\big), \tag{1.4}$$

where

$$l(D\|\lambda,\mu) \stackrel{\text{def}}{=} D\log\left(\frac{D+\sqrt{D^2+4\lambda\mu}}{2\lambda}\right) + \lambda + \mu - \sqrt{D^2+4\lambda\mu} \geq 0 \tag{1.5}$$

stands for the cost that a M/M/1 queue with parameters λ *and* μ*, starting far from the origin, follows the drift* D *(see [13], for example).*

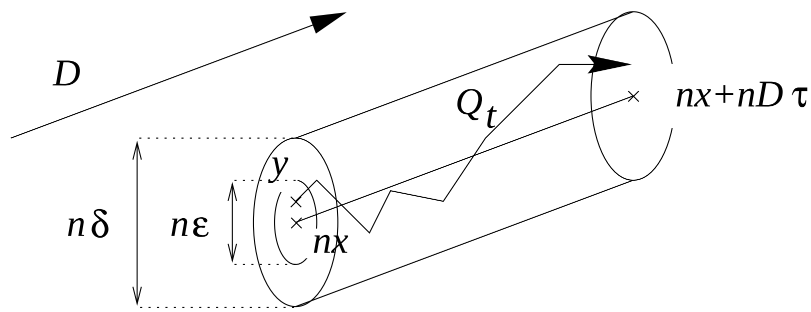

Figure 2: Structure of the local linear bounds of Theorem 1.1. $L(x,D)$ is the cost per unit time for the path $Q(t,y)$ (starting near nx) to stay in the neighborhood of $nx+Dt$ over the time $t\in[0,n\tau]$.

Let us explain briefly the meaning of the different terms appearing in $L(x,D)$ (see equation (1.4)). Owing to the fact that the service rate $\mu_{ij}(x)$ tends to 0 when x_{ij} becomes null while $x_i \vee x_j$ remains strictly positive, the arrivals must be cut on the routes belonging to $\Lambda_1(x)$ in order to keep these routes in a neighborhood of 0. The cost to do this is $\sum_{ij\in\Lambda_1(x)}\lambda_{ij}$. Since the arrivals are cut on the routes belonging to $\Lambda_1(x)$, the routes belonging to $\Lambda_2(x)$ are isolated from the rest of the network (see Figure 1) and so by (1.1) this set of routes behaves as an ergodic star network (with $\mathcal{R}=\Lambda_2(x)$) since Q is ergodic by assumption. Hence the cost for these components to stay in a neighborhood of 0 is null. Now locally, the routes belonging to $\Lambda(x)$ behaves as a set of independent M/M/1 queues with arrival and service rates λ_{ij} and $\mu_{ij}(x)$. The first term in $L(x,D)$ represents the cost that this set of queues follows

the prescribed drift D. The proof is done introducing a functional so called empirical generator consisting of Q_t and of the join number of arrivals on routes belonging to $\Lambda(x) \cup \Lambda_1(x)$. Large deviation bounds are first derived for this functional. Then Theorem 1.1 is obtained by means of an adaptation of the contraction principle.
Theorem 1.1 states large deviation bounds for ergodic networks. However, at the expense of cumbersome notation, it is possible to compute these bounds directly without ergodicity assumption introducing a more detailed empirical generator. For the ease of the exposition, the study was first performed for ergodic systems. We show now how one can compute in general L(x,D). The discussion after Theorem 1.1 explains with hands that locally in time and space the set of routes belonging to $\Lambda_2^c(x)$ behaves as an ergodic star network under the min policy and that this set of routes is separated from the rest of the network. Moreover when Q is ergodic, this sub-network is ergodic whereas when Q is not ergodic, it evolves as a possibly transient star network under the min policy. Hence the main difficulty to overcome is to compute the cost for an arbitrary star network under the min policy to stay in a neighborhood of 0.

Proposition 1.2 *Let Q be not necessarily ergodic. Then, for all $\tau \geq 0$*

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

The common value of these limits is denoted by $-L(0,0)$ and

$$L(0,0) = \inf_{\nu\in V} \sum_{ij\in\mathcal{R}} \left(\sqrt{\lambda_{ij}} - \sqrt{\mu_{ij}\nu_{ij}}\right)^2 = \inf_{\nu\in V} \sum_{ij\in\mathcal{R}} l\big(0\|\lambda_{ij}, \mu_{ij}\nu_{ij}\big) \tag{1.6}$$

where $l(.\|.,.)$ is defined in (1.5) and the set V by

$$V \stackrel{\text{def}}{=} \Big\{\nu \in \mathbb{R}_+^{\mathcal{R}} : \sum_{j\in\mathcal{S}} \nu_{ij} \leq C_i, \quad \forall i \in \mathcal{S}\Big\}. \tag{1.7}$$

Note that Proposition 1.2 is a bit stronger that the equality (1.3) of Theorem 1.1 applied to $x = D = 0$ since the time τ is not necessarily short. Besides, the rate function $L(0,0)$ is not explicit, but is an algorithmically fairly simple problem since it is a convex program in $\sqrt{\nu_{ij}}$.
Taking into account Proposition 1.2, one gets the following expression for $L(x,D)$ for a network without ergodicity condition:

Theorem 1.3 *Let Q be not necessarily ergodic, $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$. Then Theorem 1.1 is valid with the extended local rate function*

$$L(x,D) = \sum_{ij\in\Lambda(x)\cup\Lambda_1(x)} l\big(D_{ij}\|\lambda_{ij},\mu_{ij}(x)\big) \;+\; \inf_{\nu\in V} \sum_{ij\in\Lambda_2(x)} l\big(0\|\lambda_{ij},\mu_{ij}\nu_{ij}\big) \tag{1.8}$$

where V is defined in (1.7).

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

$$I_T(\varphi) \stackrel{\text{def}}{=} \begin{cases} \displaystyle\int_0^T L\left(\varphi(t),\dot\varphi(t)\right)dt, & \text{if } \varphi \text{ is absolutely continuous,} \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.9}$$

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

♦♦♦

Introduce the scaled process

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

and define the level set

$$\Phi_x(K) \stackrel{\text{def}}{=} \left\{\varphi \in D\left([0,T],\mathbb{R}_+^{\mathcal{R}}\right)\ :\ I_T(\varphi)\le K,\ \varphi(0)=x\right\}. \tag{1.10}$$

The main result of the paper if the following one:

Theorem 1.4 (Sample path LDP) *Assume Q is ergodic. The sequence $\{Q_x^n,\ n\ge 1\}$ satisfies a LDP in $D\left([0,T],\mathbb{R}_+^{\mathcal{R}}\right)$ with good rate function I_T: for every $T>0$, $x\in\mathbb{R}_+^{\mathcal{R}}$,*

(i) for $C\subset\mathbb{R}_+^{\mathcal{R}}$ compact, $\displaystyle\bigcup_{x\in C}\Phi_x(K)$ is compact in $\mathcal{C}([0,T],\mathbb{R}_+^{\mathcal{R}})$;

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

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

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

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

The organization of the paper is the following one. In Section 2, we introduce the localized empirical generator and study various properties of the functionals $l(.,.)$ and $I_T(.)$. In Section 3, large deviation bounds are obtained for the localized empirical generator from which Theorem 1.1 is derived using an adaptation of the contraction principle. The proof of Theorem 1.4 is discussed in Section 4. Some technicalities are postponed to Appendices A and B. Finally, we turn to the proof of Proposition 1.2 and Theorem 1.3 in Appendix C.

2 Localized empirical generator, entropy and the rate function

2.1 Localized empirical generator

Take $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$. We are interested in computing large deviations bounds of the form (1.3) (i.e. linear bounds as presented in Figure 2). In order to prove Theorem 1.1, we introduce a functional which allows one to measure how the different arrival rates should be modified in order that the rescaled process Q_x^n follows a prescribed drift D. Moreover, the explanation exposed just after the statement of Theorem 1.1 suggests that the transition rates of routes indexed by $\Lambda_2(x)$ should not be modified and so it is useless to measure the arrivals on routes belonging to $\Lambda_2(x)$. Let us introduce the localized empirical generator at point x as well as suitable state spaces associated to this process:

Definition 2.1 (Localized empirical generators) *Let Λ be a face and denote*

- *$A_{ij}(t)$, the number of arrivals on route ij till t;*
- *the restriction $A^\Lambda(t) \stackrel{\text{def}}{=} \left(A_{ij}(t),\ ij \in \Lambda \cup \Lambda_1\right)$;*
- *$G_t^\Lambda = \left(\frac{1}{t} A^\Lambda(t), \frac{Q_t - Q_0}{t}\right)$, the localized empirical generator on the face Λ.*

The set Γ^Λ of localized empirical generators is the set of elements (A^Λ, D) with $D \in \mathbb{R}^{\mathcal{R}}$ satisfying

$$\begin{array}{lll} (i) & a_{ij} \geq 0, & \forall ij \in \Lambda \cup \Lambda_1, \\ (ii) & a_{ij} - D_{ij} \geq 0, & \forall ij \in \Lambda \cup \Lambda_1. \end{array} \tag{2.1}$$

The space Γ^Λ *is equipped with the distance* d *defined by*

$$d(G,G') \stackrel{\text{def}}{=} \sum_{ij\in\Lambda\cup\Lambda_1} |a_{ij}-a'_{ij}| + \sum_{ij\in\mathcal{R}} |D_{ij}-D'_{ij}|, \qquad \forall G,G'\in\Gamma^\Lambda.$$

The inequalities (i) and (ii) in (2.1) refer respectively to the mean number of arrivals a_{ij} and to the mean number of deconnections per unit time $a_{ij}-D_{ij}$ being positive. Since it is difficult to analyze at first the behavior of $Q(t)$ as in (1.3), we shall first establish large deviation bounds for the event

$$E^{(n)}_{\tau,\delta,y}(x,G) \stackrel{\text{def}}{=} \left\{ G^{\Lambda(x)}_{n\tau} \in B(G,\delta),\ \sup_{t\in[0,n\tau]} |Q(t,y)-nx-Dt| < \delta n \right\} \tag{2.2}$$

where $B(G,\delta)$ is the ball of center G and radius δ (within the metric space $(\Gamma^{\Lambda(x)},d)$). As it will emerge, strong constraints must be imposed on G in order that the event $E^{(n)}_{\tau,\delta,y}(x,G)$ occurs at a large deviation scale. More precisely, the arrivals must be cut on routes belonging to $\Lambda_1(x)$:

Lemma 2.1 *Take* $x\in\mathbb{R}_+^{\mathcal{R}}$ *and* $G=(A,D)\in\Gamma^{\Lambda(x)}$ *such that* $D\in\mathbb{R}^{\Lambda(x)}$*. If there exist* m *and* p *such that*

$$x_m = 0, \quad \textit{and} \quad x_p > 0, \quad \textit{and} \quad a_{pm} > 0,$$

then the event $E^{(n)}_{\tau,\delta,y}(x,G)$ *almost never occurs at a large deviation scale, i.e:*

$$\lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y-nx|<\epsilon n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] = -\infty. \tag{2.3}$$

Proof : In fact, on $E^{(n)}_{\tau,\delta,y}(x,G)$ the service rate on route pm tends to 0 when the different limits are taken. Since on $E^{(n)}_{\tau,\delta,y}(x,G)$, the arrival process is not cut on route pm, the cost to keep the component pm of the rescaled process near 0 is infinite. Details are provided in Appendix A. ■

Lemma 2.1 states that in order to prove large deviation bounds for the localized empirical generator, it will be sufficient to deal with the following subspace of $\Gamma^{\Lambda(x)}$:

Definition 2.2 $\mathcal{G}^\Lambda$ *denotes the set of localized empirical generators* (A^Λ, D) *such that:*

$$\begin{array}{ll} \textit{(i)} & D\in\mathbb{R}^\Lambda, \\ \textit{(ii)} & a_{ij}=0 \ \textit{and}\ a_{ij}-D_{ij}\geq 0, \quad \forall ij\in\Lambda_1, \\ \textit{(iii)} & a_{ij}>0 \ \textit{and}\ a_{ij}-D_{ij}>0, \quad \forall ij\in\Lambda. \end{array} \tag{2.4}$$

In this setting $\overline{\mathcal{G}}^{\Lambda}$ will represent the closure of $\mathcal{G}^{\Lambda}$.

Owing to Lemma 2.1, it is sufficient to deal with empirical generators satisfying (ii). In order to prove the large deviation local bounds, it will be sufficient to deal with empirical generators such that arrival and service rates are not cut, for $ij \in \Lambda(x)$, hence condition (iii). A simple continuity argument will allow to extend the bounds obtained for $G \in \mathcal{G}^{\Lambda(x)}$ to $G \in \overline{\mathcal{G}}^{\Lambda(x)}$.

2.2 Correspondence between localized empirical generators and star networks

Let $G = (A, D) \in \overline{\mathcal{G}}^{\Lambda}$ be a localized empirical generator. It is associated a unique localized star network $\left(\tilde{\lambda}_{ij}, \tilde{\mu}_{ij}(y), y \in \mathbb{R}_+^{\mathcal{R}}\right)$ by the relations:

$$\begin{array}{llll} \text{(i)} & \tilde{\lambda}_{ij} = a_{ij}, & \forall ij \in \Lambda \cup \Lambda_1 & \\ \text{(ii)} & \tilde{\lambda}_{ij} = \lambda_{ij}, & \forall ij \in \Lambda_2, & \\ \text{(iii)} & \tilde{\mu}_{ij}(y) = \tilde{\lambda}_{ij} - D_{ij}, & \forall ij \in \Lambda \cup \Lambda_1,\ \forall y \in \mathbb{R}_+^{\mathcal{R}} & \\ \text{(iv)} & \tilde{\mu}_{ij}(y) = \mu_{ij}(y), & \forall ij \in \Lambda_2,\ \forall y \in \mathbb{R}_+^{\mathcal{R}}. & \end{array} \tag{2.5}$$

Let us describe the behaviour of this network when it starts from x (with $\Lambda = \Lambda(x)$). In this case, the routes belonging to Λ_2 behave as a star network of the type presently studied and the parameters of the routes belonging to this set are left unchanged. Moreover, they are independent from the rest of the network. Indeed, if $ij \in \Lambda_2$ then $x_{ik} = 0$ for all k such that $ik \notin \Lambda_2$ (actually $ik \in \Lambda_1$, see Figure 1), hence the constraints imposed on G insures that $\tilde{\lambda}_{ik} = 0$. Hence

$$\mu_{ij}(Q(s)) = Q_{ij}(s) \frac{\mu_i}{\sum_{ik \in \Lambda_2^c} Q_{ik}(s)} \wedge \frac{\mu_j}{\sum_{jk \in \Lambda_2^c} Q_{jk}(s)}, \ \forall ij \in \Lambda_2^c(x)$$

proving the asserted independence. Moreover, the network consisting of the routes belonging to Λ_2 is ergodic when the initial network is. Indeed, for all ergodic channel i (see Figure 1),

$$\sum_{j:\ ij \in \Lambda_2} \frac{\tilde{\lambda}_{ij}}{\tilde{\mu}_{ij}} = \sum_{j:\ ij \in \Lambda_2} \frac{\lambda_{ij}}{\mu_{ij}} \leq \sum_{j \in \mathcal{S}} \frac{\lambda_{ij}}{\mu_i} < C_i.$$

Besides, routes belonging to Λ behave like independent M/M/1 queues up to the initial conditions whereas the routes indexed by Λ_1 remain null. Now, the parameters have been chosen so that:

Lemma 2.2 *Assume that Q is ergodic. Let $x \in \mathbb{R}_+^{\mathcal{R}}$, $G = (A, D) \in \mathcal{G}^{\Lambda(x)}$ a localized empirical generator and denote $\widetilde{\mathbb{P}}$ the law of its associated star network. Then, for all τ,*

$$\lim_{\delta,\epsilon \to 0} \inf_{|y-nx|<\epsilon n} \liminf_{n\to\infty} \widetilde{\mathbb{P}} \left[E_{\tau,y}^{(n)}(x, G) \cap \left\{ A_{ij}(n\tau) = 0, \ \forall ij \in \Lambda_1(x) \right\} \right] = 1.$$

Proof : The proof is omitted: it is a classical fluid limit. ■

2.3 Entropy

Definition 2.3 *Let $x \in \mathbb{R}_+^{\mathcal{R}}$, $R(x) = (\lambda_{ij}, \mu_{ij}(x))$ denotes the generator of the star network at x, $G = (A, D) \in \mathcal{G}^{\Lambda(x)}$ be a localized generator and $\left(\tilde{\lambda}_{ij}, \tilde{\mu}_{ij}(y), y \in \mathbb{R}_+^{\mathcal{R}}\right)$ its representation as a star network. The relative entropy of G with respect to $R(x)$ is*

$$H(G\|R(x)) = \sum_{ij \in \Lambda(x)\cup\Lambda_1(x)} I_p\left(\tilde{\lambda}_{ij}\|\lambda_{ij}\right) + I_p\left(\tilde{\mu}_{ij}\|\mu_{ij}(x)\right)$$

where $I_p(\nu\|\lambda)$ is the relative entropy of Poisson processes of intensities ν and λ defined by

$$I_p(\nu\|\lambda) \stackrel{\text{def}}{=} \nu \log \frac{\nu}{\lambda} - \nu + \lambda \tag{2.6}$$

with the convention $\frac{0}{0} = 0$ and $0 \log 0 = 0$.

The entropy has an easy interpretation in terms of information theory: it can be defined as the *mean information gain*. $H(.\|R)$ is decomposed as the sum of the information gain for the arrivals $I_p\left(\tilde{\lambda}_{ij}\|\lambda_{ij}\right)$, the information gain for the service time $I_p\left(\tilde{\mu}_{ij}\|\mu_{ij}(x)\right)$. Note that $\lambda_{ij} = I_p(0\|\lambda_{ij})$ and that $\mu_{ij}(x) = \tilde{\mu}_{ij} = 0$ for all $ij \in \Lambda_1(x)$. Hence $I_p\left(\tilde{\mu}_{ij}\|\mu_{ij}(x)\right) = 0$ for all $ij \in \Lambda_1(x)$ and these terms do not appear in the expression of the entropy.

Lemma 2.3 *For fixed x, the entropy $H(.\|R(x))$ is continuous on $\overline{\mathcal{G}}^{\Lambda(x)}$.*

Proof : It is an easy consequence of the expression (2.6). ■

2.4 The local rate function $L(x, D)$

Definition 2.4 *The local rate function $L(x, D)$ is defined by*

$$L(x, D) \stackrel{\text{def}}{=} \inf_{G \in f_{\Lambda(x)}^{-1}(D)} H(G \| R(x)), \qquad \forall D \in \mathbb{R}^{\Lambda(x)}, \tag{2.7}$$

where $f_{\Lambda(x)} : \mathcal{G}^{\Lambda(x)} \mapsto \mathbb{R}^{\Lambda(x)}$ is the projection $f_{\Lambda(x)}(G) = D$.

It appears that $L(x, D)$ is the cost for a set of M/M/1 independent queues indexed by $\Lambda(x) \cup \Lambda_1(x)$ to follow the prescribed drift D when the queues are far from all boundaries[2]. The λ_{ij} represent the intensities of arrivals to these queues whereas the intensity of departures of the queues belonging to the set $\Lambda(x)$ [resp. $\Lambda_1(x)$] are given by $\mu_{ij}(x)$ [resp. 0]. A simple computation yields

$$l(D \| \lambda, \mu) \stackrel{\text{def}}{=} D \log \left(\frac{D + \sqrt{D^2 + 4\lambda\mu}}{2\lambda} \right) + \lambda + \mu - \sqrt{D^2 + 4\lambda\mu} \geq 0$$

for the cost that a M/M/1 queue with parameters λ and μ follows the drift D (see [13], for example). Using this remark and the identity $l(0 \| \lambda, 0) = \lambda$, one can deduce the explicit representation (1.4) for $L(x, D)$ (which is equal to (2.8) under the constraint $D_{ij} = \mu_{ij}(x) = 0$ for $ij \in \Lambda_1(x)$).
In equations (1.4) and (2.7), $L(x, D)$ is only defined for $D \in \mathbb{R}^{\Lambda(x)}$. In order to study the properties of the rate function $I_T(.)$, it is convenient to extend the definition of $L(x, D)$ for all D such that $D_{ij} \geq 0$ for all $ij \in \Lambda^c(x)$ by

$$L(x, D) \stackrel{\text{def}}{=} \sum_{ij \in \Lambda(x) \cup \Lambda_1(x)} l\big(D_{ij} \| \lambda_{ij}, \mu_{ij}(x)\big). \tag{2.8}$$

In the rest of this section, a number of useful properties of $L(x, D)$ are derived. Note that $L(x, D)$ is a rate function derived by the contraction of $H(.\|R(x)$ and that these properties could be derived taking advantage of this fact. Nevertheless, in this case it is simpler to use (2.8).

Proposition 2.4 *The local rate function $L(x, D)$ possesses the following properties.*

(*i*) *It is positive, finite, strictly convex and continuous with respect to D such that $D_{ij} \geq 0$ for all $ij \in \Lambda(x)$. It has compact level sets;*

[2] Or equivalently this is the cost for a set of independent random walk in $\mathbb{Z}$ to follow the drift D.

(*ii*) *there exists* $M \in \mathbb{R}$ *such that,*

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

(*iii*) *for a fixed* D *and a prescribed face* Λ, $L(x,D)$ *is continuous for* $x \in \Lambda$ *(see equation (1.2));*

(*iv*) $L(x,D)$ *is jointly lower semi-continuous w.r.t.* x *and* D.

Proof : Properties (i) and (ii) are obvious from (2.8).

(iii) is clear from (2.8) noting that the functions $\mu_{ij}(x)$, $ij \in \Lambda$, are continuous for x belonging to the face Λ. Moreover, $\Lambda_1(x) = \Lambda_1$ is constant for $x \in \Lambda$.

Let $(x^{(n)}, D^{(n)})$ tends to (x,D). First, it is clear that for n large enough, $\Lambda(x) \subset \Lambda(x^{(n)})$ and also $\Lambda(x) \cup \Lambda_1(x) \subset \Lambda(x^{(n)}) \cup \Lambda_1(x^{(n)})$. Hence, since l is positive, for sufficiently large n,

$$L(x^{(n)}, D^{(n)}) \geq \sum_{ij \in \Lambda(x) \cup \Lambda_1(x)} l\big(D_{ij}^{(n)} \| \lambda_{ij}, \mu_{ij}(x^{(n)})\big). \tag{2.9}$$

Now, $\lambda_{ij} > 0$ (since $ij \in \mathcal{R}$) so that $l(.\|\lambda_{ij}, .)$ is continuous. Moreover $\mu_{ij}(x^{(n)}) \to \mu_{ij}(x)$, $\forall ij \in \Lambda(x) \cup \Lambda_1(x)$. Therefore the right part of (2.9) converges to $L(x,D)$ and the lower semi-continuity (iv) is proved. ■

2.5 The sample path rate function $I_T(.)$

In this section, we verify that the rate function $I_T(.)$ (see definitions (1.9) and 1.10)) possesses the usual properties.

Proposition 2.5 *The rate function* $I_T(.)$ *possesses the following properties.*

(*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* $\Big(D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big), d_d\Big)$;

(*iii*) *for* $C \subset \mathbb{R}_+^{\mathcal{R}}$ *compact,* $\bigcup_{x \in C} \Phi_x(K)$ *is compact in* $\mathcal{C}([0,T],\mathbb{R}_+^{\mathcal{R}})$*;*

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

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

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

Proof : One proves (i) using Proposition 2.4 (ii) in a way similar to [13, Lemma 5.18].

In order to prove the lower semi-continuity of $I_T(.)$, (i) shows it is sufficient to consider sequences of absolutely continuous functions. Since on $\mathcal{C}\big([0,T],\mathbb{R}_+^{\mathcal{R}}\big)$, the metrics d_c and d_d are equivalent, one can use d_c. Now, using Proposition 2.4 (ii), the fact that $L(x,D)$ is lower semi-continuous in (x,D) and convex with respect to D by Proposition 2.4, (ii) is proved by means of [12, Theorem 3 of Section 9.1.4].

(iii) is a consequence of (i) and (ii) (see [13, Proposition 5.46]).

The proof of (iv) is a simple adaptation of [3, Proposition 5.1 (iv)]. ∎

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

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

3 Large deviations bounds for the localized empirical generator

In this section, we aim at proving the following theorem:

Theorem 3.1 *Let* $x \in \mathbb{R}_+^{\mathcal{R}}$ *and* $G = (A,D) \in \overline{\mathcal{G}}^{\Lambda(x)}$ *be a localized generator. Then*

$$\begin{aligned} -H\big(G \| R(x)\big) &= \lim_{\tau,\delta,\epsilon\to 0} \inf_{|y-nx|<\epsilon n} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] \\ &= \lim_{\tau,\delta,\epsilon\to 0} \sup_{|y-nx|<\epsilon n} \limsup_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right], \end{aligned}$$

where $E^{(n)}_{\tau,\delta,y}(x,G)$ *is the event defined in (2.2). Moreover, if a face* Λ *and a drift* $D \in \mathbb{R}^\Lambda$ *are fixed, then the preceding limit in* τ *is uniform w.r.t to* x *in compact sets of* Λ *(see Definition 1.1).*

3.1 An exponential change of measure

Fix an empirical generator $G = (A, D) \equiv (\tilde{\lambda}_{ij}, \tilde{\mu}_{ij}(y), y \in \mathbb{R}_+^{\mathcal{R}}) \in \overline{\mathcal{G}}^{\Lambda(x)}$. Let us describe how $\widetilde{\mathbb{P}}$ can be obtained from $\mathbb{P}$. For, denote by

- N_t, the number of jumps of the process till t.
- $Q(k) = \{Q_{ij}(k),\ i, j \in \mathcal{S}\}$, the embedded Markov chain[3] at time $k \in \mathbb{N}$.

Define

- the mapping $h : \mathbb{Z}_+^{\mathcal{R}} \times \mathbb{Z}_+^{\mathcal{R}} \mapsto \mathbb{R}$ by

$$h(x,y) \stackrel{\text{def}}{=} \begin{cases} \log \dfrac{\tilde{\lambda}_{ij}}{\lambda_{ij}} & \text{if } y - x = e_{ij} \text{ and } \tilde{\lambda}_{ij} > 0, \\[2ex] \log \dfrac{\tilde{\mu}_{ij}(x)}{\mu_{ij}(x)} & \text{if } y - x = -e_{ij} \text{ and } \tilde{\mu}_{ij}(x) > 0, \\[2ex] 0 & \text{otherwise.} \end{cases}$$

- the compensator $K : \mathbb{Z}_+^{\mathcal{R}} \mapsto \mathbb{R}$ by

$$\begin{aligned} K(x) &\stackrel{\text{def}}{=} \sum_{y \in \mathbb{Z}_+^{\mathcal{R}}} q(x,y)\Big(e^{h(x,y)} - 1\Big) \\ &= \sum_{ij \in \mathcal{R}} \Big(\tilde{\lambda}_{ij} - \lambda_{ij}\Big) + \sum_{ij \in \mathcal{R}} \Big(\tilde{\mu}_{ij}(x) - \mu_{ij}(x)\Big). \end{aligned} \tag{3.1}$$

- and the process

$$\mathcal{M}_t \stackrel{\text{def}}{=} \exp\left\{ \sum_{k=0}^{N_t - 1} h\Big(Q(k), Q(k+1)\Big) - \int_0^t K\big(Q(s)\big)\, \mathrm{d}v \right\}.$$

Note that the compensator is always bounded, so that $\mathcal{M}_t$ takes only finite values. Since K has been exactly defined so that[4]

$$K(x) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\exp\left\{ \sum_{k=0}^{N_t - 1} h\Big(Q(k,x), Q(k+1,x)\Big) \right\}\right]_{t=0},$$

[3] We shall distinguish between discrete and continuous time by using k for discrete and s or t for continuous time.

[4] Note that the derivative is independent of Λ, so that it is dropped.

it is easily checked that the derivative of $\mathbb{E}[\mathcal{M}_t]$ 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 $\mathbb{E}[\mathcal{M}_t] = 1$. Using again the Markov property, this proves that $\mathbb{E}[\mathcal{M}_t | \mathcal{F}_s] = \mathcal{M}_s$, for all $t \geq s \geq 0$, hence $\{\mathcal{M}_t,\ t \geq 0\}$ is a martingale w.r.t. the natural filtration $\mathcal{F}_t$.

Then define a new probability measure by

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

It is a matter of routine to show that under $\widetilde{\mathbb{P}}$, X is again a Markov process. In fact, under $\widetilde{\mathbb{P}}$, the system behaves like a star network where the arrival and the service rates at node ij are respectively given by $\tilde{\lambda}_{ij}$ and $\tilde{\mu}_{ij}(y)$ (whence the notation).

> **Remark :** The probability measure $\mathbb{P}$ is not necessarily absolutely continuous with respect to $\widetilde{\mathbb{P}}$. This is the case for instance if for some $ij \in \mathcal{R}$, $\tilde{\lambda}_{ij} = 0$ (whereas $\lambda_{ij} > 0$).

♦♦♦

3.2 Proof of the upper bound of Theorem 3.1

Since $\mathbb{P}$ is not necessarily absolutely continuous with respect to $\widetilde{\mathbb{P}}$, in order to prove the upper bound, we introduce a sequence of change of measure $\{\widetilde{\mathbb{P}}^{(\eta)},\ \eta > 0\}$ such that

$$\tilde{\lambda}_{ij}^{(\eta)} > 0 \quad \text{and} \quad \lim_{\eta \to 0} \tilde{\lambda}_{ij}^{(\eta)} = \tilde{\lambda}_{ij}, \qquad \forall ij \in \Lambda(x) \cup \Lambda_1(x),$$
$$\tilde{\mu}_{ij}^{(\eta)} > 0 \quad \text{and} \quad \lim_{\eta \to 0} \tilde{\mu}_{ij}^{(\eta)} = \tilde{\mu}_{ij}(x), \qquad \forall ij \in \Lambda(x).$$

In this setting, $\{\mathcal{M}_t^{(\eta)},\ t \geq 0\}$ is the martingale defining $\widetilde{\mathbb{P}}^{(\eta)}$ with respect to $\mathbb{P}$ and $h^{(\eta)}(x, y)$ and $K^{(\eta)}(x)$ are the functions used to defined $\mathcal{M}_t^{(\eta)}$ according to Section 3.1. Now, $\widetilde{\mathbb{P}}^{(\eta)}$ and $\mathbb{P}$ are mutually absolutely continuous and,

$$\mathbb{P}\left[E_{\tau,\delta,y}^{(n)}(x, G)\right] = \widetilde{\mathbb{E}}^{(\eta)}\left[\mathbb{1}_{\{E_{\tau,\delta,y}^{(n)}(x,G)\}} \left(\mathcal{M}_{n\tau}^{(\eta)}\right)^{-1}\right]. \tag{3.2}$$

Let us majorize $\left(\mathcal{M}_{n\tau}^{(\eta)}\right)^{-1}$ on $E_{\tau,\delta,y}^{(n)}(x, G)$ when $|y - nx| < \delta n$. First, recalling $\tilde{\lambda}_{ij} = \lambda_{ij}$ for $ij \in \Lambda_2$ and $\tilde{\mu}_{ij}(y) = \mu_{ij}(y)$ for $ij \in \Lambda_1 \cup \Lambda_2$ and $y \in \mathbb{R}_+^{\mathcal{R}}$, one has the following bounds:

$$
\begin{aligned}
& -\sum_{k=0}^{N_{n\tau}-1} h^{(\eta)}\big(Q(k),Q(k+1)\big) \qquad (3.3)\\
\leq &\quad -n\tau\left(\sum_{ij\in\Lambda(x)} \tilde{\mu}_{ij}\log\frac{\tilde{\mu}_{ij}^{(\eta)}}{\sup\limits_{s\in[0,n\tau]}\mu_{ij}(Q(s))} + \sum_{ij\in\Lambda(x)\cup\Lambda_1(x)} \tilde{\lambda}_{ij}\log\frac{\tilde{\lambda}_{ij}^{(\eta)}}{\lambda_{ij}}\right)\\
& \quad +n\tau\delta\left(\sum_{ij\in\Lambda(x)}\left|\log\frac{\tilde{\mu}_{ij}^{(\eta)}}{\inf\limits_{s\in[0,n\tau]}\mu_{ij}(Q(s))}\right| + \sum_{ij\in\Lambda(x)\cup\Lambda_1(x)}\left|\log\frac{\tilde{\lambda}_{ij}^{(\eta)}}{\lambda_{ij}}\right|\right)
\end{aligned}
$$

Moreover the compensator K is bounded in 3.1 by

$$
\begin{aligned}
& \int_0^{n\tau} K^{(\eta)}\big(Q(s)\big)\,\mathrm{d}s \qquad (3.4)\\
\leq &\quad n\tau\sum_{ij\in\Lambda(x)\cup\Lambda_1(x)}\big(\tilde{\lambda}_{ij}^{(\eta)}-\lambda_{ij}\big) + n\tau\sum_{ij\in\Lambda(x)}\left(\tilde{\mu}_{ij}^{(\eta)} - \inf_{s\in[0,n\tau]}\mu_{ij}\big(Q(s)\big)\right).
\end{aligned}
$$

Besides, on $E_{\tau,\delta,y}^{(n)}(x,G)$, for $ij\in\Lambda(x)$ one has

$$
\begin{aligned}
\mu_{ij}(Q(s,y)) &\geq \frac{C_i(x_{ij}-\delta+sD_{ij})}{(x_i+\delta+sD_i)^+}\wedge\frac{C_j(x_{ij}-\delta+sD_{ij})}{(x_j+\delta+sD_j)^+}\\
\mu_{ij}(Q(s,y)) &\leq \frac{C_i(x_{ij}+\delta+sD_{ij})}{(x_i-\delta+sD_i)^+}\wedge\frac{C_j(x_{ij}+\delta+sD_{ij})}{(x_j-\delta+sD_j)^+}
\end{aligned}
$$

Hence, on $E_{\tau,\delta,y}^{(n)}(x,G)$, we have for $ij\in\Lambda(x)$

$$
\begin{aligned}
0<\mu_{ij}(x) &= \lim_{\tau,\delta,\epsilon\to 0}\ \inf_{|y-nx|<\epsilon n}\ \liminf_{n\to\infty}\ \inf_{s\in[0,n\tau]}\mu_{ij}(Q(s,y)) \qquad (3.5)\\
&= \lim_{\tau,\delta,\epsilon\to 0}\ \sup_{|y-nx|<\epsilon n}\ \limsup_{n\to\infty}\ \sup_{s\in[0,n\tau]}\mu_{ij}(Q(s,y)).
\end{aligned}
$$

Finally, bounding $\mathcal{M}_{n\tau}^{(\eta)}$ using (3.3), (3.4) and (3.5), majorizing $\mathbb{I}_{\{E_{\tau,\delta,y}^{(n)}(x,G)\}}$ by 1 and taking into account the order in which the different limits are taken, the representation formula (3.2) yields

$$
\begin{aligned}
& \lim_{\tau,\delta,\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n\tau}\sup_{|y-nx|<\epsilon n}\log\mathbb{P}\left[E_{\tau,\delta,y}^{(n)}(x,G)\right]\\
\leq &\quad -\sum_{ij\in\Lambda(x)\cup\Lambda_1(x)}\tilde{\lambda}_{ij}\log\frac{\tilde{\lambda}_{ij}^{(\eta)}}{\lambda_{ij}} - \tilde{\lambda}_{ij}^{(\eta)} + \lambda_{ij} - \sum_{ij\in\Lambda(x)}\tilde{\mu}_{ij}^{(\eta)}\log\frac{\tilde{\mu}_{ij}^{(\eta)}}{\mu_{ij}(x)} - \tilde{\mu}_{ij}^{(\eta)} + \mu_{ij}(x).
\end{aligned}
$$

The proof of the upper bound is concluded letting η tends to 0.

3.3 Proof of the lower bound of Theorem 3.1

Take $G \in \mathcal{G}^{\Lambda(x)}$ and denote the event (appearing in Lemma 2.2)

$$F^{(n)}_{\tau,\delta,y}(x,G) \stackrel{\text{def}}{=} E^{(n)}_{\tau,y}(x,G) \cap \{A_{ij}(n\tau)=0,\ \forall ij \in \Lambda_1(x)\}.$$

Although $\mathbb{P}$ is not absolutely continuous w.r.t $\widetilde{\mathbb{P}}$, by definition of $\mathcal{G}^{\Lambda(x)}$, $\tilde{\lambda}_{ij} > 0$ and $\tilde{\mu}_{ij} > 0$, $\forall ij \in \Lambda(x)$ so that $\mathbb{P}$ is absolutely continuous w.r.t $\widetilde{\mathbb{P}}$ on $F^{(n)}_{\tau,\delta,y}(x,D)$ and

$$\mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] \geq \mathbb{P}\left[F^{(n)}_{\tau,\delta,y}(x,G)\right] \geq \inf_{\omega\in F^{(n)}_{\tau,\delta,y}(x,D)} \mathcal{M}^{-1}_{n\tau}(\omega)\widetilde{\mathbb{P}}\left[F^{(n)}_{\tau,\delta,y}(x,G)\right].$$

By Lemma 2.2, $\widetilde{\mathbb{P}}\left[F^{(n)}_{\tau,\delta,y}(x,G)\right]$ tends to 1. Therefore, reversing the inequalities obtained for the upper bound yields

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \inf_{|y-nx|<\delta n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] \\
\geq\ & -\sum_{ij\in\Lambda(x)\cup\Lambda_1(x)} \tilde{\lambda}_{ij}\log\frac{\tilde{\lambda}_{ij}}{\lambda_{ij}} - \tilde{\lambda}_{ij} + \lambda_{ij} - \sum_{ij\in\Lambda(x)} \tilde{\mu}_{ij}\log\frac{\tilde{\mu}_{ij}}{\mu_{ij}(x)} - \tilde{\mu}_{ij} + \mu_{ij}(x).
\end{aligned}$$

This concludes the proof of the lower bound when $G \in \mathcal{G}^{\Lambda(x)}$. Consider $G \in \overline{\mathcal{G}}^{\Lambda(x)}$. For any $\delta > 0$, there exists $G' \in \mathcal{G}^{\Lambda(x)}$ and $\delta' > 0$ such that $B(G',\delta') \subset B(G,\delta)$. Hence

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \inf_{|y-nx|<\delta n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] \\
\geq\ & \lim_{\tau,\delta',\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \inf_{|y-nx|<\delta' n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta',y}(x,G')\right] = -H(G'\|R(x)).
\end{aligned}$$

Since this is true for any $G' \in \mathcal{G}^{\Lambda(x)}$ arbitrary closed to G, by continuity of the entropy (see Lemma 2.3), the lower bound of Theorem 3.1 is proved for any $G \in \overline{\mathcal{G}}^{\Lambda(x)}$.

The uniformity of the limit stated in Theorem 3.1 is easily checked. Nonetheless, this uniformity is clear as far as x evolves on compact sets of some face Λ. Indeed, if x_{ij} goes to 0 for some $ij \in \Lambda$, then $\mu_{ij}(x)$ possibly vanishes and difficulties can appear (see Lemma B.1).

Proof of Theorem 1.1 Now Theorem 3.1 implies the large deviations local bounds of Theorem 1.1. Moreover, if a face Λ and a drift $D \in \mathbb{R}^{\Lambda}$ are fixed, then the limits in (1.3) in τ are uniform w.r.t to x in compact sets of Λ. The proof relies on a simple adaptation of the contraction principle, similarly to the proof of [3, Theorem 4.2]. Details are omitted. ■

4 The sample path large deviations principle

Using Markov property, Theorem 1.1 and the continuity of $L(x, D)$ with respect to $x \in \Lambda(D)$ for fixed D, we start proving large deviations bounds for the probability that the process stays near some linear path.

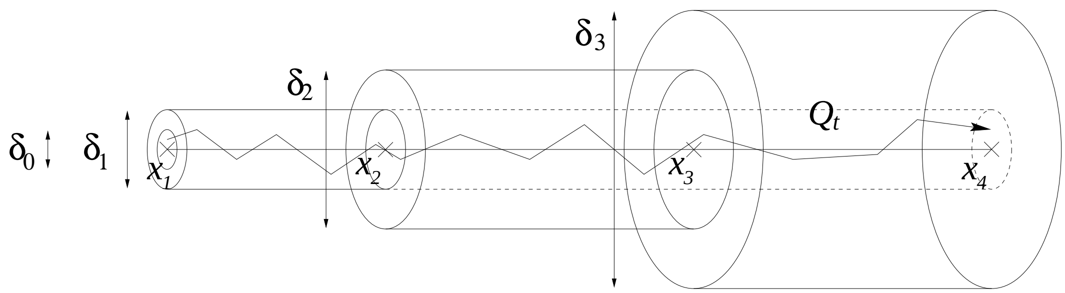

Figure 3: The structure of the bounds for the proof of Theorem 4.1.

Proposition 4.1 (Linear bounds) *Let $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\mathcal{R}}$ satisfying $x + DT \in \mathbb{R}_+^{\mathcal{R}}$. Denote φ the function such that $\varphi(t) = x + Dt$ for all $t \in [0, T]$. Then*

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

Proof : First, note that $\Lambda(\varphi(t)) = \Lambda(D)$ for all $t \in]0, T[$. The bounds obtained in Theorem 1.1 can be used in all intervals of the form $]\eta, T - \eta[$ where η is a positive constant.

Upper bound Fix η and denote by $\{t_i\}$ the partition of $[\eta, T-\eta]$ such that $t_i \stackrel{\text{def}}{=} \eta + i\Delta T$ where $\Delta T \stackrel{\text{def}}{=} (T-2\eta)/K$ and $K \in \mathbb{N}$ is fixed for a while; let $x_i \stackrel{\text{def}}{=} \varphi(t_i)$. Introduce a sequence of numbers $0 < \delta_0 < \ldots < \delta_K$. Then using the Markov property at times t_i (as suggested in Figure 3), we get:

$$\begin{aligned}
&\sup_{|y-nx|<\delta_0 n} \log \mathbb{P}\left[\sup_{t\in[0,T]} |Q(t,y) - n\varphi(t)| < \delta_1 n\right] \\
\leq\ &\sum_{i=0}^{K-1} \sup_{|y-nx_i|<\delta_i n} \log \mathbb{P}\left[\sup_{t\in[0,\Delta T]} |Q(t,y) - nx_i - nDt| < \delta_{i+1} n\right].
\end{aligned}$$

For the sake of brevity we introduce the notation

$$\bar{I}_\varphi(t_0,t_1) \stackrel{\text{def}}{=} \lim_{\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \sup_{|y-n\varphi(t_0)|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[t_0,t_1]} |Q(t,y) - n\varphi(t)| < \delta n\right]$$

which, combined with the previous bounds yield:

$$\bar{I}_\varphi(0,T) \leq \sum_{i=0}^{K-1} \bar{I}_\varphi(t_i, t_{i+1}).$$

Fix $\gamma > 0$. The linear path $\varphi(t)$ pertaining to a compact set of $\Lambda(D)$ for all $t \in]\eta, T-\eta[$, Theorem 1.1 (applied for each $\bar{I}(t_i, t_{i+1})$) ensures that there exists K_γ such that for all $K \geq K_\gamma$

$$\bar{I}_\varphi(0,T) \leq -\Delta T \sum_{i=0}^{K-1} \Big(L\big(x_i, D\big) - \gamma\Big).$$

Letting first K tends to infinity and then γ to 0, one obtains using the continuity of $L(.,D)$ (see Proposition 2.4 (iii))

$$\bar{I}_\varphi(0,T) \leq -\int_\eta^{T-\eta} L\big(\varphi(t), \dot{\varphi}(t)\big)\, \mathrm{d}t.$$

The functional $I_T(\varphi)$ being finite, the upper bound follows letting η tends to 0.

Lower bound As for the upper bound, the bounds obtained in Theorem 1.1 are used in an interval of the form $]\eta, T-\eta[$. Using additionally Lemma B.1 dealing with the intervals $[0,\eta]$ and $[T-\eta, T]$ when η tends to 0, the proof of the lower bound is similar to the upper bound one. Details are omitted. ■

From linear paths to LDP The sample path local bounds of Theorem 1.4 are now proved for linear paths (Proposition 4.1). There are some steps to reach the LDP, which we outline here.

First, the local bounds are extended to piecewise linear paths (i.e. for φ belonging to $\mathcal{PL}\big([0,T],\mathbb{R}_+^{\mathcal{R}}\big)$). Using the Markov property and bounds as described in Figure 3, the proof looks very much like that of Proposition 4.1.

Second, the local bounds are extended to absolutely continuous paths (i.e. for φ belonging to $\mathcal{AC}\big([0,T],\mathbb{R}_+^{\mathcal{R}}\big)$) with finite entropy, using the remark of p. 16 and the properties of I_T.

The next step is to prove the exponential tightness of $\big\{n^{-1}Q(nt,[nx]),\ n\geq 1\big\}$ over finite interval of time (uniformly for x belonging to a compact set). Finally Theorem 1.4 is proved. These last two steps use various properties of the rate function I_T and Proposition 2.5. The reader is referred to [7, Section 5] for details.

Appendix A Proof of Lemma 2.1

Lemma 2.1 *Take $x\in\mathbb{R}_+^{\mathcal{R}}$ and $G=(A,D)\in\Gamma^{\Lambda(x)}$ such that $D\in\mathbb{R}^{\Lambda(x)}$ when $x_{ij}=0$. If there exist m and p such that $x_m=0$, $x_p>0$ and $a_{pm}>0$, then the event $E^{(n)}_{\tau,\delta,y}(x,G)$ almost never occurs at a large deviation scale (see equation (2.3)).*

Proof : The proof uses a change of measure as described in Section 3.1, which we keep the notation of. Let the parameters $\tilde\lambda_{ij}$ and $\tilde\mu_{ij}(y)$ be chosen so that

$$\begin{aligned}\tilde\lambda_{ij}&>0, &&\forall ij,\\ \tilde\mu_{ij}(y)=\tilde\mu_{ij}&>0, &&\forall ij,\ \forall y\in\mathbb{R}_+^{\mathcal{R}},\\ D_{ij}&=\tilde\lambda_{ij}-\tilde\mu_{ij}, &&\forall ij.\end{aligned}$$

The probabilities $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are mutually absolutely continuous and so that the representation formula (3.2) is valid with $\mathcal{M}_{n\tau}$. Let us majorize $\mathcal{M}_{n\tau}^{-1}$ more explicitly on $E^{(n)}_{\tau,y}(x,D)$, on which the following rough bounds hold.

$$\begin{aligned}&-\sum_{k=0}^{N_{n\tau}-1} h\big(Q(k),Q(k+1)\big)\\ \leq\ & n\tau\sum_{ij\in\mathcal{R}}(\tilde\lambda_{ij}+\delta)\left|\log\frac{\tilde\lambda_{ij}}{\lambda_{ij}}\right|+(\tilde\mu_{ij}+\delta)\left|\log\frac{\tilde\mu_{ij}}{\mu_{ij}(C_i\wedge C_j)}\right|.\end{aligned}$$

For the term due to services on route pm, remembering that $\tilde{\mu}_{pm} > 0$, $x_{pm} = 0$ (since $x_m = 0$) and so $D_{pm} \geq 0$, one has the following bound at least when δ and τ are sufficiently small (since $x_p > 0$),

$$\begin{aligned} & - \sum_{k=0}^{N_{n\tau}-1} h(Q(k), Q(k+1))) \mathbb{I}_{\{Q(k+1)-Q(k)=-e_{pm}\}} \\ \leq \quad & -n\tau(\tilde{\mu}_{pm} - \delta) \log \frac{\tilde{\mu}_{pm}(nx_p - |D_p|\tau n - \delta n)}{\mu_{pm} C_p(\tau n D_{pm} + \delta n)}. \end{aligned}$$

Now, note that all quantities in the compensator $K(x)$ (see (3.1)) are bounded w.r.t. x, hence there exists $K_+ < \infty$ such that $K(x) \leq K_+$ for all $x \in \mathbb{R}_+^{\mathcal{R}}$. Combining the previous inequalities yields

$$\begin{aligned} \frac{1}{n\tau} \log \mathbb{P}\left[E_{\tau,\delta,y}^{(n)}(x,D)\right] &= \frac{1}{n\tau} \log \widetilde{\mathbb{E}}\left[\mathbb{I}_{\{E_{\tau,\delta,y}^{(n)}(x,D)\}} \mathcal{M}_{n\tau}^{-1}\right] \\ \leq \quad & \sum_{ij} (\tilde{\lambda}_{ij} + \delta) \left| \log \frac{\tilde{\lambda}_{ij}}{\lambda_{ij}} \right| + \sum_{ij \neq pm} (\tilde{\mu}_{ij} + \delta) \left| \log \frac{\tilde{\mu}_{ij}}{\mu_{ij}(C_i \wedge C_j)} \right| \\ & -(\tilde{\mu}_{pm} - \delta) \log \frac{\tilde{\mu}_{pm}(x_p - |D_p|\tau - \delta)}{\mu_{pm} C_p(\tau D_{pm} + \delta)} + K_+. \end{aligned}$$

Taking into account the order in which the limits are taken and using the fact that $x_p > 0$, $\tilde{\mu}_{pm} > 0$, one can conclude since the right hand side tends to $-\infty$. ■

Appendix B Irreducibility matters

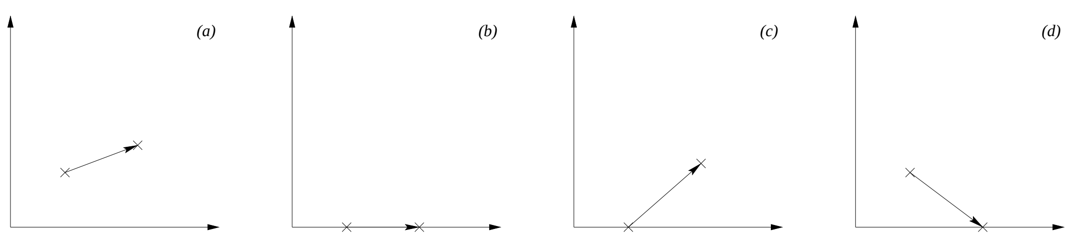

Figure 4: Different cases for the Lemma B.1.

In the proof of LDP bounds for linear paths (Proposition 4.1), we can use the local bounds of Theorem 1.1 since it is required that D belongs to $\mathbb{R}^{\Lambda(x)}$ (hence cases (a)

and (b) in Figure 4 fit with Theorem 1.1). The beginning or the end of a *linear* path are the two sole points where this is not necessarily true (see cases (c) and (d)). Nevertheless Lemma B.1 proves that there is no extra cost for entering or leaving a face (at a large deviation scale). Hence there is no irreducibility[5] problem (at a large deviation scale) with the star network.
For the sake of brevity we shall denote

$$E^{(n)}_{\eta,\delta,y}(x,D) \stackrel{\text{def}}{=} \left\{ \sup_{t\in[0,n\eta]} |Q(t,y)-nx-Dt| < \delta n \right\},$$

Lemma B.1 *Let $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\mathcal{R}}$ satisfying $x+DT \in \mathbb{R}_+^{\mathcal{R}}$. Denote by φ the function such that $\varphi(t) = x + Dt$ for all $t \in [0,T]$. Then*

$$\lim_{\eta,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \inf_{|y-nx|<\epsilon n} \log \mathbb{P}\left[E^{(n)}_{\eta,\delta,y}(x,D)\right] = 0, \tag{B.1}$$

$$\lim_{\eta,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \inf_{|y-n\varphi(T-\eta)|<\epsilon n} \log \mathbb{P}\left[E^{(n)}_{\eta,\delta,y}(\varphi(T-\eta),D)\right] = 0. \tag{B.2}$$

Proof : Tracing back where the limitation $D \in \mathbb{R}^{\Lambda(x)}$ comes from leads to (3.3): it is necessary that $\mu_{ij}(Q(s))$ be bounded away from 0, uniformly over $[0,n\tau]$. Since the service rate is only changed for $ij \in \Lambda(x)$, equation (3.5) fix the problem, hence the limitation.
This analysis allows to solve the case (c) in Figure 4) which only happens when *exiting* a face (i.e. $D_{ij} > 0$). It is easily checked (using a change of measure as in the proof of Theorem 3.1) that one can modify λ_{ij} and $\mu_{ij}(x)$ so that the path follows a drift D. The cost is *less* than cutting the service to 0 (for a cost $I_p(0\|\mu_{ij}(x)) = \mu_{ij}(x) < \infty$) and putting the arrival rate to D_{ij} (for a cost $I_p(D_{ij}\|\lambda_{ij}) < \infty$ because $\lambda_{ij} > 0$). Since the finite cost is multiplied by a time η decreasing to 0, we get equation (B.1) (case (d) does not apply to (B.1)).

The case (d) is much more difficult to handle. Note that $D_{ij} < 0$ and $x_{ij} + D_{ij}T = 0$, hence $\mu_{ij}(x+DT) = 0$. A rough analysis similar to the case (c) shows that *it is necessary to increase the service rate to at least $-D_{ij} > 0$*. But the cost per unit time to do that is $I_p(-D_{ij}\|0) = \infty$ in the neighborhood of $x+DT$. So it is necessary to get more precise bounds.
Denote $x' \stackrel{\text{def}}{=} \varphi(T-\eta)$. First we shall only focus on the set $\mathcal{R}'$ of routes ij for which $x'_{ij} > 0$ and $x'_{ij} + \eta T = 0$. In the other cases, the above discussion shows that the cost to follow D_{ij} is finite (say bounded by C).

[5]This feature is not obvious, since the intensities $\mu_{ij}(x)$ are not bounded away from 0).

We shall denote by $F^{(n)}_{\eta,\delta,y}(x,D)$ the event when $q_{ij}(t,y)$ jumps downwards once, and only once, over the time interval $\left[k|D_{ij}^{-1}|,(k+1)|D_{ij}^{-1}|\right)$, for all $k \leq n\eta|D_{ij}|$ and all $ij \in \mathcal{R}'$. When n is large, it is easily checked that $F^{(n)}_{\eta,\delta,y}(x,D)$ is included in $E^{(n)}_{\eta,\delta,y}(x,D)$. Without loss of generality, assume also $nx'_{ij}+n\epsilon/2 < y_{ij} < nx'_{ij}+n\epsilon$. In this case, the $q_{ij}(t,y)$, for $ij \in \mathcal{R}'$ never reach 0 over $[0,n\eta]$. This assumption is necessary so as to focus on $F^{(n)}_{\eta,\delta,y}(x,D)$, since q_{ij} can not jump downward if it reach 0: it protects $F^{(n)}_{\eta,\delta,y}(x,D)$ from being the empty set.
When $|y-nx'|<\epsilon n$, on $A^{(n)}_{\eta,\delta,y}(x,D)$ (hence on $F^{(n)}_{\eta,\delta,y}(x,D)$), there exists a constant K_{ij} independent of n, η, δ, ϵ satisfying

$$\mu_{ij}(Q(s)) \geq \mu_{ij} q_{ij}(s) \frac{K_{ij}}{n} \ . \tag{B.3}$$

Now we introduce the change of measure:

$$\tilde{\lambda}_{ij} \stackrel{\text{def}}{=} 0, \quad \text{and} \quad \tilde{\mu}_{ij}(z) \stackrel{\text{def}}{=} \mu_{ij} K_{ij} \quad \forall z \in \mathbb{R}_+^{\mathcal{R}},\ \forall ij \in \mathcal{R}',$$

with the same K_{ij} as in (B.3). Since all intensities λ_{ij}, $\mu_{ij}(z)$, $\tilde{\lambda}_{ij}$ and $\tilde{\mu}_{ij}(z)$ are bounded, the compensator $K(x)$ in (3.1) is bounded, say by K_+. The new probability measure (defined as in Section 3.1) $\widetilde{\mathbb{P}}$ is absolutely continuous w.r.t $\mathbb{P}$ so that

$$\mathbb{P}\left[E^{(n)}_{\eta,\delta,y}(x,D)\right] \geq e^{-Cn\eta}\widetilde{\mathbb{E}}\left[\mathcal{M}_{n\eta}^{-1}\mathbb{1}_{\{F^{(n)}_{\eta,\delta,y}(x,D)\}}\right]. \tag{B.4}$$

Moreover, the martingale $\mathcal{M}_{n\eta}^{-1}$ is bounded on $F^{(n)}_{\eta,\delta,y}(x,D)$ by

$$\log \mathcal{M}_{n\eta} \leq n\eta K_+ + \sum_{k=0}^{N_{n\eta}-1}\sum_{ij\in\mathcal{R}'} \log\left(\frac{\tilde{\mu}_{ij}}{\mu_{ij}(Q(k))}\right)\mathbb{1}_{\{q_{ij}(k+1)-q_{ij}(k)=-1\}}. \tag{B.5}$$

We used the fact that there are only downward jumps on $F^{(n)}_{\eta,\delta,y}(x,D)$, so that the arrival rates do not appear. In (B.4), $Q(k)$ is the state of the process after the k-th jump (counting *all* jumps): for the coordinate ij we denote by T_k^{ij} the time of k-th jump. On $F^{(n)}_{\eta,\delta,y}(x,D)$, the embedded chain is deterministic *coordinate by coordinate*, i.e. $q_{ij}(T_k^{ij}) = y_{ij}-k$. This means that

$$\frac{\tilde{\mu}_{ij}}{\mu_{ij}(Q(T_k^{ij}))} \leq \frac{n\tilde{\mu}_{ij}}{\mu_{ij}q_{ij}(T_k^{ij})K_{ij}} \leq \frac{\tilde{\mu}_{ij}}{\mu_{ij}K_{ij}(x'_{ij}+\epsilon/2-k/n)}, \tag{B.6}$$

using successively the bound (B.3) and the assumption $y_{ij} > nx'_{ij} + n\epsilon/2$. Inequalities (B.6) yield, on $F^{(n)}_{\eta,\delta,y}(x, D)$.

$$\begin{aligned} & n^{-1} \sum_{k=0}^{N_{n\eta}-1} \sum_{ij\in\mathcal{R}'} \log\left(\frac{\tilde{\mu}_{ij}}{\mu_{ij}(Q(k))}\right) \mathbb{1}_{\{q_{ij}(k+1)-q_{ij}(k)=-1\}} \\ \leq \quad & -n^{-1} \sum_{k=0}^{n\eta|D_{ij}|} \log \frac{\tilde{\mu}_{ij}(x'_{ij} + \epsilon/2 - k/n)}{\mu_{ij} K_{ij}} \\ \overset{n\to\infty}{\longrightarrow} \quad & -\int_0^{\eta|D_{ij}|} \log \frac{\tilde{\mu}_{ij}(x'_{ij} + \epsilon/2 - u)}{\mu_{ij} K_{ij}} \, du. \end{aligned} \tag{B.7}$$

Finally, using (B.5) and (B.7), the bound (B.4) is transformed into

$$\begin{aligned} & \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[E^{(n)}_{\eta,\delta,y}(x, D)\right] \\ \geq \quad & -(C + K_+)\eta + \sum_{ij\in\mathcal{R}'} \int_0^{\eta|D_{ij}|} \log \frac{\tilde{\mu}_{ij}(x'_{ij} + \epsilon/2 - u)}{\mu_{ij} K_{ij}} \, du. \end{aligned}$$

The bound (B.2) follows easily and thus Lemma B.1 is proved. ∎

Appendix C The transient case

C.1 Proof of Proposition 1.2

The main point is to prove Proposition 1.2. As in the ergodic case, the proof relies on four steps: the introduction of a suitable empirical generator, the association of a star network to each empirical generator, the proof of large deviation bounds for empirical generator and finally the proof of Proposition 1.2 using an adaptation of the contraction principle.

C.1.1 Empirical generator

Let us introduce a functional which allows one to measure how the different arrival and service rates should be modified so that the rescaled process Q^n_x stays in a neighborhood of 0. This process is a bit different than the one defined in the ergodic case (see Definition 2.1). It takes into account the sole case $x = D = 0$, but in the transient case.

Definition C.1 *The empirical generator G_t is the functional defined by*

$$G_t \stackrel{\text{def}}{=} \left(\frac{1}{t}A(t), \frac{1}{t}\int_0^t \nu\big(Q(s)\big)ds\right).$$

where $\nu(x) \stackrel{\text{def}}{=} (\nu_{ij}(x),\ i,j \in \mathcal{S})$. The set Γ of empirical generators is $\mathbb{R}_+^{\mathcal{R}} \times V$; its elements will be denoted by $G = (A, \nu)$. It is equipped with the distance d defined by

$$d(G,G') \stackrel{\text{def}}{=} \sum_{ij\in\mathcal{R}} |a_{ij} - a'_{ij}| + \sum_{ij\in\mathcal{R}} |\nu_{ij} - \nu'_{ij}|, \qquad \forall G, G' \in \Gamma.$$

In order to prove Proposition 1.2, in a first and main step, large deviation bounds are established for the event (similarly to (2.2))

$$E^{(n)}_{\tau,\delta,y}(G) \stackrel{\text{def}}{=} \left\{G_{n\tau} \in B(G,\delta),\ \sup_{t\in[0,n\tau]} |Q(t,y)| < \delta n\right\} \tag{C.1}$$

where $B(G,\delta)$ is the ball of center G and radius δ. Roughly speaking, when $\nu_{pm} = 0$ the service rate are cut on route pm and so some constraints must be imposed on A. More precisely

Lemma C.1 *Take $G = (A,\nu) \in \Gamma$. If there exist m and p such that*

$$\nu_{pm} = 0, \quad \text{and} \quad a_{pm} > 0,$$

then $E^{(n)}_{\tau,\delta,y}(G)$ almost never occurs at a large deviation scale, i.e:

$$\lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y|<\epsilon n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(G)\right] = -\infty.$$

Proof : It is enough to select a change of measure where the transition rates are changed only on route pm. Choose $\tilde{\lambda}_{pm}$ and $\tilde{\mu}_{pm}(y)$

$$\begin{aligned} \tilde{\lambda}_{pm} &= a_{pm}, \\ \tilde{\mu}_{pm}(y) &= \tilde{\mu}_{pm} y_{pm} \frac{C_p}{y_p} \wedge \frac{C_m}{y_m},\ \forall y \in \mathbb{R}_+^{\mathcal{R}}, \end{aligned}$$

where $\tilde{\mu}_{pm}$ is an arbitrary strictly positive number. Then $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are mutually absolutely continuous and

$$\lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y|<\epsilon n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(G)\right] \le -a_{pm}\Big(\log\frac{a_{pm}}{\lambda_{pm}} + \log\frac{\tilde{\mu}_{pm}}{\mu_{pm}}\Big).$$

The proof is concluded letting $\tilde{\mu}_{pm}$ tends to $+\infty$. ■

By Lemma C.1, it is enough to deal with the following subspace of Γ.

Definition C.2 *$\mathcal{G}$ denotes the set of empirical generator (A,ν) such that*

$$\begin{array}{ll} (i) & a_{ij}=0 \quad \text{when} \quad \nu_{ij}=0, \\ (ii) & \sum_j \nu_{ij} < C_i, \quad \forall i. \end{array}$$

$\overline{\mathcal{G}}$ stands for the closure of $\mathcal{G}$.

C.1.2 Correspondence between empirical generators and star networks

Let $G=(A,\nu)\in\overline{\mathcal{G}}$. It is associated arrival and departure rates:

$$\begin{aligned} \tilde{\lambda}_{ij} &\stackrel{\text{def}}{=} a_{ij}, \qquad \forall ij \in \mathcal{R} \\ \tilde{\mu}_{ij}(y) &\stackrel{\text{def}}{=} \tilde{\mu}_{ij} y_{ij} \frac{C_i}{y_i} \wedge \frac{C_j}{y_j} \mathbb{I}_{\{y_{ij}>0\}}, \quad \forall ij\in\mathcal{R},\ \forall y \in \mathbb{R}_+^{\mathcal{R}}, \\ \text{where} \qquad \tilde{\mu}_{ij} &\stackrel{\text{def}}{=} \begin{cases} \frac{\tilde{\lambda}_{ij}}{\nu_{ij}} & \forall ij \text{ such that } \nu_{ij}>0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Then $(\tilde{\lambda}_{ij}, \tilde{\mu}_{ij}(y),\ y\in\mathbb{R}_+^{\mathcal{R}})$ simply describes a star network under the min policy where the arrivals intensity and the duration of calls on route ij are respectively given by $\tilde{\lambda}_{ij}$ and $\tilde{\mu}_{ij}$.
Similarly to Lemma 2.2, we now prove:

Lemma C.2 *Let $G=(A,\nu)\in\mathcal{G}$ and denote $\widetilde{\mathbb{P}}$ the law of its associated star network. Then Q is ergodic under $\widetilde{\mathbb{P}}$. Besides, for all τ,*

$$\lim_{\delta,\epsilon\to 0} \inf_{|y|<\epsilon n} \liminf_{n\to\infty} \widetilde{\mathbb{P}}\left[E^{(n)}_{\tau,\delta,y}\right] = 1. \tag{C.2}$$

Proof : Since $G\in\mathcal{G}$, the ergodicity condition (1.1) are easily checked for $(\tilde{\lambda}_{ij},\tilde{\mu}_{ij})$, so that Q is ergodic under $\widetilde{\mathbb{P}}$. Moreover a straight application of the ergodic theorem yields

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \nu_{ij}\big(Q(s)\big)ds = \frac{\tilde{\lambda}_{ij}}{\tilde{\mu}_{ij}} = \nu_{ij}, \quad \forall ij. \tag{C.3}$$

Equation (C.2) is thus just a statement about fluid limits. ■

C.1.3 Entropy and local bounds

Definition C.3 (entropy) *Let $G = (A, D) \in \overline{\mathcal{G}}$ be an empirical generator and $\left(\tilde{\lambda}_{ij}, \tilde{\mu}_{ij}\right)$ its representation as a star network. The relative entropy of G with respect to R, the generator of the initial star network is*

$$H(G\|R) = \sum_{ij\in\mathcal{R}} \Big(I_p\left(\tilde{\lambda}_{ij}\|\lambda_{ij}\right) + I_p\left(\tilde{\lambda}_{ij}\|\nu_{ij}\mu_{ij}\right) \Big)$$

where $I_p(\nu\|\lambda)$ is the relative entropy of Poisson processes of intensities ν and λ defined by

$$I_p(\nu\|\lambda) \stackrel{\text{def}}{=} \nu \log\frac{\nu}{\lambda} - \nu + \lambda \qquad \textit{with the convention } \frac{0}{0} = 0 \quad \textit{and} \quad 0\log 0 = 0.$$

We are ready to state large deviation bounds for the empirical generator

Proposition C.3 *Let $G = (A, \nu) \in \overline{\mathcal{G}}$ be an empirical generator. Then*

$$\begin{aligned} -H\big(G\|R\big) &= \lim_{\delta,\epsilon\to 0}\ \inf_{|y|<\epsilon n}\ \liminf_{n\to\infty} \frac{1}{n\tau}\log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(G)\right] \\ &= \lim_{\delta,\epsilon\to 0}\ \sup_{|y|<\epsilon n}\ \limsup_{n\to\infty} \frac{1}{n\tau}\log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(G)\right], \end{aligned}$$

where $E^{(n)}_{\tau,\delta,y}(G)$ is the event defined in (C.1).

Proof : The proof is similar to the one of Theorem 3.1 and will not be repeated. Note simply that the lower bound is first proved for $G \in \mathcal{G}$ using in particular Lemma C.2. It is then extended to all $G \in \overline{\mathcal{G}}$ using the continuity of the entropy H.
■

Proof of Proposition 1.2 It is derived from Proposition C.3 and from Lemma C.1 by means of an adaptation of the contraction principle as in [3, Theorem 4.2]. Details are similar to the proof of Theorem 1.1 and thus omitted. Note that

$$L(0,0) = \inf_{G\in\overline{\mathcal{G}}} H(G\|R)$$

Taking $G = (A, \nu) \in \overline{\mathcal{G}}$ and minimizing w.r.t A yields (1.6). ■

C.2 Proof of Theorem 1.3

In Proposition 1.2, we have discussed how to treat the main difference with the ergodic case, that is the calculation of the cost for the routes belonging to $\Lambda_2(x)$ to stay in a neighborhood of 0. Indeed, using the same change of measure defined by $\widetilde{\mathbb{P}}$ in Section 2.2 as well as the asserted independence between routes belonging to $\Lambda_2(x)$ and the rest of the network under $\widetilde{\mathbb{P}}$ when the network starts from x, Theorem 1.1 and Proposition 1.2 yield

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

Using the same type of arguments as in Lemma B.1, one can prove that

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

and

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

This concludes the proof. ■

Remark : At the expense of heavier notation, this theorem could have been derived at once as in Section 3 studying the following more detailed empirical generator

$$L_t = \left(\frac{1}{t}A(t), \frac{1}{t}\int_0^t \nu_{\Lambda_2(x)}\big(Q(s)\big)ds, \frac{1}{t}Q_t\right)$$

where $\nu_{\Lambda_2(x)} = (\nu_{ij}(y), ij \in \Lambda_2(x), y \in \mathbb{R}_+^{\mathcal{R}})$.

♦♦♦

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] de La Fortelle, A. Large deviation principle for markov chains in continuous time. Tech. Rep. 3877, INRIA, 2000. To appear in *Problems of Information Transmission*.

[3] Delcoigne, F., and de La Fortelle, A. Large deviations for polling systems. Tech. Rep. 3892, INRIA, 2000. submitted to *Queueing Systems*.

[4] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications*, second ed. Springer-Verlag, New York, 1998.

[5] Deuschel, J.-D., and Stroock, D. W. *Large deviations*, vol. 137 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, 1989.

[6] 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.

[7] 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.

[8] 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.

[9] Fayolle, G., de La Fortelle, A., Lasgouttes, J. M., Massoulie, L., and Roberts, J. Best-effort networks : Modeling and performance analysis via large networks asymptotics. In *Proceedings of IEEE INFOCOM'01* (2001).

[10] Ignatiouk-Robert, I. Large deviations of Jackson networks. *Ann. Appl. Probab. 10*, 3 (2000), 962–1001.

[11] 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.

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

[13] 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.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

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