

## Edge minimality of EDF resource sharing networks

Łukasz Kruk<sup>1</sup> 

Received: 22 November 2016 / Accepted: 12 June 2017 / Published online: 24 June 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** We consider a general real-time network with soft customer deadlines, in which users require service from several shared resources simultaneously. We call the service protocol for such a network edge minimal (locally edge minimal) if it minimizes globally (resp., locally) in time, in a suitable sense, the system resource idleness with respect to customers with lead times not greater than any given threshold value on all the routes of the network. We show that the preemptive Earliest-Deadline-First (EDF) service discipline is edge minimal. Moreover, we characterize the preemptive EDF policy as a protocol making the underlying network locally edge minimal. Our arguments are pathwise, independent on the network topology and requiring very mild assumptions, or even no assumptions, on the model stochastic primitives. Application of our characterization to fluid model analysis for EDF resource sharing networks is also discussed.

**Keywords** Queueing · Deadlines · Resource sharing · EDF · Partial ordering · Minimality

**Mathematics Subject Classification** 60K25 · 68M12 · 60G17 · 68M20 · 90B22

### 1 Introduction

Massoulié and Roberts (1999, 2000) introduced resource sharing networks, also known as bandwidth sharing models, to model congestion control problems for Inter-

---

✉ Łukasz Kruk  
lkrusk@hektor.umcs.lublin.pl

<sup>1</sup> Department of Mathematics, Maria Curie-Skłodowska University,  
Pl. Marii Curie-Skłodowskiej 1, 20-031 Lublin, Poland

net flows. In such systems flows, corresponding to continuous transfers of elastic documents, are being transmitted, requiring simultaneous service at all nodes along their routes. Various service protocols for resource sharing networks have been proposed. The most popular of them are proportional fairness, originating in a work of Kelly (1997), and more general  $\alpha$ -fair policies, introduced by Mo and Walrand (2000). The literature on the latter two disciplines is large. Here we mention only Gromoll and Williams (2009), Kang et al. (2009), Vlasiou et al. (2015), as representative examples of related asymptotic results. Other utility-based protocols were considered, e.g., in Chiang et al. (2006), Key and Massoulié (2006), Ye (2003), Ye et al. (2005). Stability (or the lack thereof) of the size-based Shortest Remaining Processing Time (SRPT) and Least Attained Service first (LAS) policies in resource sharing networks was investigated by Verloop et al. (2005). The interested reader may consult Harrison et al. (2014) for a recent overview of developments in the area of bandwidth sharing models.

In the last two decades we observe increasing importance and popularity of real-time services, in which jobs have individual timing requirements. Examples of such services include various telecommunication networks, manufacturing systems, in which the orders have due dates, tracking systems and real-time vehicle control systems. Another important class of applications arises in medical scheduling problems, like organ allocation or prioritizing admissions to hospital emergency rooms.

In the theory of real-time systems, hard, firm and soft deadlines are distinguished. Each hard deadline has to be met, or a system failure occurs. A firm deadline can be missed, but there is no value in completing a late task. In this paper, we limit our attention to systems with soft deadlines, which permit lateness and use the jobs completed after their deadlines. This setup is very popular in applications, to mention only manufacturing systems or voice and video transmission services, like video-on-demand (VOD), see, e.g., Sivaraman et al. (2001).

Earliest Deadline First (EDF) is a service policy, under which the job with the shortest remaining lead time, i.e., the difference between its deadline and the current time, is selected for service. It is a well-studied protocol, especially in computer and manufacturing systems, see, e.g., Hopp and Spearman (1996), Stankovic et al. (2012). For single server, single customer class queueing systems with customer deadlines, several optimality results for this discipline are known. Liu and Layland (1973) proved optimality of EDF for job loss in a hard deadline environment. Panwar and Towsley (1988) demonstrated that EDF minimizes the fraction of customers who miss their deadlines within the class of preemptive disciplines in a G/M/1 queue. Moyal (2008) showed that in a G/G/1 queue with soft deadlines, non-preemptive EDF minimizes the mean stationary lateness of the customers. In the context of firm deadlines, Panwar and Towsley (1992) showed that EDF minimizes the fraction of reneging customers, not served to completion due to elapsed deadlines, in a G/M/c queue. More recently, Kruk et al. (2011) obtained a similar result for a single server system, in which the amount of reneged work was used as a performance measure.

Counterparts of the above-mentioned EDF optimality results for resource sharing networks are still in short supply. In the case of hard deadlines, Baruah (2006) considered a computing platform consisting of a preemptable processor and several non-preemptable resources, shared by jobs of different tasks. The jobs are scheduled according to the EDF discipline, with access to shared resources arbitrated by the Stack

Resource Policy (SRP). [Baruah \(2006\)](#) showed optimality of the EDF + SRP policy for this platform in the following sense: if a system fails the EDF+SRP schedulability test, then, for some data, it also misses deadlines under any other work-conserving discipline.

[Kruk \(2016\)](#) considered preemptive resource sharing networks with arbitrary topology and soft job deadlines. In order to quantify the ability of such a network to meet the file transmission timing requirements, for each element  $i$  from the set  $\mathbf{I}$  of available routes in this network and each time  $t \geq 0$ , a (random) “cumulative idleness distribution function”  $Y_i(t, \cdot)$  was introduced. Here, for each  $s \in \mathbb{R}$ ,  $Y_i(t, s)$  denotes the cumulative idleness by time  $t$  with regard to transmission of flows with lead times at time  $t$  not greater than  $s$ . The main idea of [Kruk \(2016\)](#) was to find service protocols minimizing the vector of functions  $(Y_i)_{i \in \mathbf{I}}$  or their sum  $\sum_{i \in \mathbf{I}} Y_i$  - equivalently, maximizing the corresponding cumulative transmission times - with respect to the *pointwise* functional inequality. The latter relation is defined as follows: for functions  $f, g$  of two variables  $t, s$ , we have  $f \leq g$  if and only if  $f(t, s) \leq g(t, s)$  for all arguments  $t$  and  $s$ . If such minimizing disciplines exist, they are called *pathwise minimal* or *additively minimal*, respectively. It turns out that without any distributional assumptions on the stochastic model primitives, the EDF resource sharing policy is pathwise minimal ([Kruk 2016, Theorem 1](#)). A partial converse to this result is the fact that in a pathwise minimal resource sharing network, flows on every route are scheduled according to the EDF discipline ([Kruk 2016, Theorem 2](#)). Moreover, under mild distributional assumptions, the EDF resource sharing protocol is additively minimal ([Kruk 2016, Theorem 3](#)).

Taking the results of [Kruk \(2016\)](#) as our starting point, we must admit that pointwise inequality of functions is a very strong relation and hence the notion of a minimal element with respect to this ordering is rather weak. (Somewhat counterintuitively, a minimal element with respect to some partial ordering is also necessarily minimal with respect to any stronger one, but not necessarily with respect to a weaker one, because a weaker relation allows for more comparisons.) Consequently, there are, in general, many minimal, or even additively minimal, service disciplines for a given resource sharing network which differ from the EDF protocol, see Example 5 from [Kruk \(2016\)](#). Therefore, the main idea of this paper is to find a partial ordering on the space  $\mathbb{R}^{\mathbb{R}}$  of real functions on the real line which is weaker than pointwise inequality and induces (in a suitable sense) the ordering on the family of performance processes such that the corresponding minimal elements are as close to the EDF systems as possible. Ideally, we would like to find a relation such that minimality with respect to it characterizes, at least under mild assumptions, the EDF service discipline in a soft deadline environment.

To achieve the goal stated above, we introduce a new partial ordering “ $\ll$ ” on  $\mathbb{R}^{\mathbb{R}}$ . By definition, for  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f \ll g$  if either  $f = g$ , or if  $f$  coincides with  $g$  on some half-line  $(-\infty, c)$  (where  $c$  is assumed to be as large as possible) and  $f \leq g$  in  $[c, c + \epsilon]$  for some  $\epsilon > 0$ . In other words,  $f \ll g$  if the pointwise inequality  $f \leq g$  holds in some, presumably small, neighbourhood of the “left edge” of the set  $\{x \in \mathbb{R} : f(x) \neq g(x)\}$ . Two arbitrary functions  $f, g \in \mathbb{R}^{\mathbb{R}}$  do not have to be comparable with respect to the relation “ $\ll$ ”, because they do not have to coincide on any half-line. However, if they do coincide on some half-line, then the inequality  $f \leq g$

is clearly stronger than  $f \ll g$ . The relation “ $\ll$ ” is well suited for comparing the “cumulative idleness distribution functions” of the form  $f(s) = \sum_{i \in I} Y_i(t, s)$ ,  $s \in \mathbb{R}$ , corresponding to different service disciplines for the same network at a given time  $t \geq 0$ . Moreover, it turns out to be useful in characterizing the EDF discipline, which always works at the “left edge” of the cumulative lead time distribution  $\sum_{i \in I} Z_i(t, \cdot)$  of the flows currently present in the network, and hence on the “left edge” of  $\sum_{i \in I} Y_i(t, \cdot)$ .

We introduce a partial ordering “ $\lll$ ” induced by the relation “ $\ll$ ” on the space of performance processes  $\mathfrak{X}$ , corresponding to service disciplines implemented for a given resource sharing network with soft deadlines, so that, by definition,  $\mathfrak{X}^{(1)} \lll \mathfrak{X}^{(2)}$  if and only if  $\sum_{i \in I} Y_i^{(1)}(t, \cdot) \ll \sum_{i \in I} Y_i^{(2)}(t, \cdot)$  for each  $t \geq 0$ . A discipline corresponding to a minimal element with respect to “ $\lll$ ” is called *edge minimal*. An edge minimal protocol is additively minimal (and hence minimal), but the converse statement is, in general, false. We show the following refinement of Theorem 3 of Kruk (2016): under the same distributional assumptions, as in the latter result, the preemptive EDF resource sharing protocol is edge minimal (Theorem 4). However, even under these assumptions, there may exist other edge minimal disciplines, see Example 3. Moreover, for generally distributed stochastic primitives, an EDF resource sharing network may fail to be edge minimal, see Corollary 1. This is because an EDF scheduler (as, in fact, any other practically implementable protocol) is unable to predict the deadlines and file sizes of future arrivals. Consequently, it makes decisions which are optimal, as far as minimizing the cumulative idleness distribution  $\sum_{i \in I} Y_i(t, \cdot)$  is concerned, conditioning on the currently available information, and hence *optimal locally in time* (more precisely, until the next arrival), but not necessarily on the whole time horizon, on which additional information is being revealed, compare Example 3 of Kruk (2016).

The above discussion suggests an introduction of the following “local” counterpart of the notion of edge minimality. A discipline with cumulative idleness distribution functions  $Y_i$ ,  $i \in I$ , is called *locally edge minimal* if for every time  $t_0 \geq 0$  there exists  $h > 0$  such that replacing this service protocol with another one in the interval  $[t_0, t_0 + h]$  results in the cumulative idleness distribution functions  $Y'_i$  satisfying  $\sum_{i \in I} Y_i(t, \cdot) \ll \sum_{i \in I} Y'_i(t, \cdot)$  for all  $t \in [t_0, t_0 + h]$ . It turns out that under mild distributional assumptions (the same as in Theorem 3 of Kruk (2016) and our Theorem 4), the classes of preemptive EDF resource sharing policies and locally edge minimal service disciplines coincide, see Theorems 5–6. Moreover, in Theorem 7 we show, without any distributional assumptions on the model stochastic primitives, that a service policy is locally edge minimal if and only if it belongs to a class of *strong EDF protocols*. Roughly speaking, an EDF discipline for a resource sharing network is strong EDF if its tie-breaking rule allows for scheduling the transmission of flows with the smallest available lead times on as many routes as possible (see Sect. 5.3 for a precise definition). Hence, even in full generality, local edge minimality is fairly close to being a characterization of the preemptive EDF policy for networks with shared resources.

As in Kruk (2016), our arguments are pathwise, similar in spirit to the proofs of the optimality result in Kruk et al. (2011) and Schrage’s (1968) well known theorem stating that the Shortest Remaining Processing Time (SRPT) discipline minimizes the

queue length in a single-server system. We make no assumptions on the underlying network topology.

In the light of Theorems 5–7, it seems reasonable to use local edge minimality as a definition of the EDF property for fluid models of resource sharing networks. In a strongly related area of “conventional” multiserver, multiclass queueing networks with the EDF protocol, Bramson (2001) defined a fluid model for such a network in terms of a set of suitable fluid model equations. The key one of them, capturing the EDF property of the underlying fluid model, implies that the system is work-conserving, which is true for multiclass queueing networks, but not for resource sharing networks considered in this paper. Moreover, it is not clear to us how to define a counterpart of this equation in our setting. Therefore, in Sect. 6 we suggest to replace it by local edge minimality of a resource sharing fluid model. In fact, our initial motivation for the research presented here and in Kruk (2016) was to find a minimality property of EDF resource sharing networks suitable for a definition of the EDF property in the corresponding fluid models. Once this task is accomplished, we can study the asymptotic behavior of resource sharing networks with the EDF protocol by investigating the corresponding fluid models. We are planning to pursue this idea in a forthcoming article.

This paper is organized as follows. In Sect. 2 we define a stochastic model for an EDF resource sharing network. In Sect. 3 we briefly recall the main notions and results from Kruk (2016). Sections 4 and 5 discuss edge minimality and local edge minimality of EDF resource sharing systems, respectively. In Sect. 6 we sketch an application of the notions of minimality and local edge minimality to the formulation of fluid models for EDF resource sharing networks. Finally, Sect. 7 contains some concluding remarks.

## 1.1 Notation

The following notation will be used throughout. For a finite set  $A$ , let  $|A|$  denote the cardinality of  $A$  and let  $2^A$  denote the family of all the subsets of  $A$ . Let  $\mathbb{R}$  denote the set of real numbers. For  $a, b \in \mathbb{R}$ , we write  $a \vee b$  ( $a \wedge b$ ) for the maximum (minimum) of  $a$  and  $b$ . The minimum taken over the empty set should be interpreted as  $\infty$ . Vector inequalities are to be interpreted componentwise, i.e., for  $a, b \in \mathbb{R}^n$ ,  $a = (a_1, \dots, a_n)$ ,  $b = (b_1, \dots, b_n)$ ,  $a \leq b$  if and only if  $a_i \leq b_i$  for all  $i = 1, \dots, n$ . Functional inequalities are to be interpreted pointwise, i.e., for  $f, g : A \rightarrow \mathbb{R}^n$ , we write  $f \leq g$  if and only if  $f(x) \leq g(x)$  for all  $x \in A$ . By convention, a sum over the empty set of indices equals zero.

The Borel  $\sigma$ -field on  $\mathbb{R}^n$  will be denoted by  $\mathcal{B}(\mathbb{R}^n)$ . For  $B \in \mathcal{B}(\mathbb{R}^n)$ , we denote the indicator of the set  $B$  by  $\mathbb{I}_B$ . The function  $\mathbb{I}_{\mathbb{R}}$  will be denoted by  $\mathbf{1}$ . The set of bounded, continuous real functions on  $\mathbb{R}^n$  will be denoted by  $C_b(\mathbb{R}^n)$ . For a function  $f(x, y)$  of two real variables, let  $d_x f(x, y)$  denote the differential of  $f(x, y)$  with respect to  $x$ , i.e.,  $dg_y(x)$ , where  $g_y(x) = f(x, y)$  is a function of  $x$  depending on a parameter  $y$ .

Let  $\mathbf{M}_n$  denote the set of finite, nonnegative measures on  $\mathcal{B}(\mathbb{R}^n)$  and let  $\mathbf{M} = \mathbf{M}_1$ . When  $\mu \in \mathbf{M}$  and  $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ , we will simply write  $\mu(a, b)$  and  $\mu(a, b]$  instead of  $\mu((a, b))$  and  $\mu((a, b])$ . For  $\mu \in \mathbf{M}$ , we define  $L_\mu = \sup\{x \in \mathbb{R} : \mu(-\infty, x) = 0\}$ . In particular,  $\mu(\mathbb{R}) = 0$  if and only if  $L_\mu = \infty$ . The set  $\mathbf{M}_n$  is

endowed with the weak topology. That is, for  $\xi_m, \xi \in \mathbf{M}_n$ , we have  $\xi_m \xrightarrow{w} \xi$  if and only if  $\int_{\mathbb{R}^n} g(x) \xi_m(dx) \rightarrow \int_{\mathbb{R}^n} g(x) \xi(dx)$  as  $m \rightarrow \infty$  for all  $g \in \mathbf{C}_b(\mathbb{R}^n)$ . With this topology,  $\mathbf{M}_n$  is a Polish space (see Prohorov 1956). We denote the zero measure in  $\mathbf{M}_n$  by  $\mathbf{0}$  and the measure in  $\mathbf{M}_n$  that puts one unit of mass at a point  $x \in \mathbb{R}^n$  by  $\delta_x$ . For  $(x, y) \in \mathbb{R}^2$ , the measure  $\delta_{(x,y)}^+$  is  $\delta_{(x,y)}$  if  $x > 0$  and  $\mathbf{0}$  otherwise.

All stochastic processes used in this paper are assumed to have paths that are right continuous with finite left limits (r.c.l.l.), unless explicitly stated otherwise. For a Polish space  $\mathcal{S}$ , we denote by  $\mathbf{D}([0, \infty), \mathcal{S})$  the space of r.c.l.l. functions from  $[0, \infty)$  into  $\mathcal{S}$ . For  $x \in \mathbf{D}([0, \infty), \mathbb{R}^n)$  and  $t > 0$ , define  $\Delta x(t) = x(t) - x(t-)$ .

## 2 Stochastic model

### 2.1 Network structure

We consider a network with a finite number of resources (nodes), labelled by  $j = 1, \dots, J$ , and a finite set of routes, labelled by  $i = 1, \dots, I$ . Each route may be identified with a nonempty subset of  $\mathbf{J} = \{1, \dots, J\}$ , interpreted as the set of resources used by this route. Let  $A = [a_{ji}]$  be the  $J \times I$  incidence matrix in which  $a_{ji} = 1$  if resource  $j$  is used by route  $i$  and  $a_{ji} = 0$  otherwise. Let  $\mathbf{I} = \{1, \dots, I\}$ . Then the set  $\mathcal{R}(i)$  of resources used by route  $i$  may be described by the equation  $\mathcal{R}(i) = \{j \in \mathbf{J} : a_{ji} = 1\}$ . Similarly, the set  $\mathcal{F}(j)$  of routes using the resource  $j$  is defined by the equation  $\mathcal{F}(j) = \{i \in \mathbf{I} : a_{ji} = 1\}$ .

By a flow on route  $i$  we mean a continuous transmission of a file through the resources used by this route. We assume that a flow takes simultaneous possession of all the resources on its route during the transmission. For convenience, we also assume that all the resources have a unit service rate.

### 2.2 Stochastic primitives

Let  $(\Omega, \mathcal{A}, \mathbf{P})$  be a probability space on which all the random objects to follow will be defined. The initial condition consists of the nonnegative, integer-valued random variables  $Q_i(0)$ ,  $i \in \mathbf{I}$ , counting the numbers of initial flows on each route at time zero, the strictly positive random initial file sizes of the initial flows  $\tilde{v}_{i,k}$  and their corresponding initial lead times (deadlines)  $\tilde{l}_{i,k}$ , where  $i \in \mathbf{I}$ ,  $k = 1, \dots, Q_i(0)$ . The initial flow with service time  $\tilde{v}_{i,k}$  and deadline  $\tilde{l}_{i,k}$  will be called flow  $k$  on route  $i$ . Let  $Q(0) = (Q_1(0), \dots, Q_I(0))$ .

Let  $N_i(\cdot)$  be the integer-valued exogenous arrival process for the route  $i \in \mathbf{I}$ . For  $t \geq 0$ ,  $N_i(t)$  represents the number of flows arriving to the  $i$ -th route in the time interval  $(0, t]$ . The  $k$ -th arrival modelled by  $N_i(\cdot)$  will be called flow  $Q_i(0) + k$  on route  $i$ . Its arrival time equals  $U_{i,k} = \inf\{t \geq 0 : N_i(t) \geq k\}$ . For  $i \in \mathbf{I}$  and  $t \geq 0$ , let  $A_i(t) = Q_i(0) + N_i(t)$ .

For  $i \in \mathbf{I}$  and  $k \geq 1$ , a random variable  $v_{i,k}$  represents the initial size of the file associated with the  $Q_i(0) + k$ -th flow on route  $i$ , i.e., the cumulative transfer time of

this flow through the network. We assume that for each  $i \in \mathbf{I}$  the random variables  $\{v_{i,k}\}_{k \geq 1}$  are strictly positive.

For  $i \in \mathbf{I}$  and  $k \geq 1$ , a random variable  $l_{i,k}$  represents the initial lead time for the transmission of the file associated with the  $Q_i(0) + k$ -th flow on route  $i$ . Thus, the deadline for the  $Q_i(0) + k$ -th transmission on route  $i$  equals  $U_{i,k} + l_{i,k}$ .

## 2.3 Residual file sizes, lead times

For  $t \geq 0$ ,  $i \in \mathbf{I}$  and  $k \leq A_i(t)$ , let  $v_{i,k}(t)$  denote the residual size of the file (transmission time) of flow  $k$  on route  $i$  at time  $t$ . Thus,  $v_{i,k}(\cdot)$  decreases at rate one during the transmission of the flow  $k$  on route  $i$  and it is constant otherwise.

To determine whether flows meet their timing requirements, one must keep track of each flows's lead time, where

$$\text{lead time} = \text{initial lead time} - \text{time elapsed since arrival}$$

for flows coming to the system after time zero and

$$\text{lead time} = \text{initial lead time} - \text{current time}$$

for initial flows. More formally, let  $t \geq 0$ ,  $i \in \mathbf{I}$  and  $k \leq A_i(t)$ . The lead time at time  $t$  of flow  $k$  on route  $i$  is defined by

$$l_{i,k}(t) = \begin{cases} \tilde{l}_{i,k} - t, & \text{if } k \leq Q_i(0), \\ l_{i,k-Q_i(0)} + U_{i,k-Q_i(0)} - t, & \text{if } k > Q_i(0). \end{cases}$$

We combine the stochastic primitives defined above into the following measure-valued arrival processes: for  $i \in \mathbf{I}$  and  $t \geq 0$ , let

$$\mathcal{A}_i(t) = \sum_{k=1}^{A_i(t)} \delta_{l_{i,k}(t)}, \quad \mathcal{V}_i(t) = \sum_{k=1}^{Q_i(0)} \tilde{v}_{i,k} \delta_{l_{i,k}(t)} + \sum_{k=1}^{N_i(t)} v_{i,k} \delta_{l_{i,Q_i(0)+k}(t)}. \quad (1)$$

Let  $\mathcal{A}(t) = (\mathcal{A}_1(t), \dots, \mathcal{A}_I(t))$  and  $\mathcal{V}(t) = (\mathcal{V}_1(t), \dots, \mathcal{V}_I(t))$ .

For future reference note that for  $0 \leq u \leq t$ ,  $i \in \mathbf{I}$  and  $k \leq A_i(u)$ , we have  $l_{i,k}(t) = l_{i,k}(u) - (t - u)$ , and hence

$$\mathcal{A}_i(u)(-\infty, s + t - u] \leq \mathcal{A}_i(t)(-\infty, s], \quad s \in \mathbb{R}. \quad (2)$$

Indeed, for any  $s \in \mathbb{R}$ ,

$$\mathcal{A}_i(u)(-\infty, s + t - u] = \sum_{k=1}^{A_i(u)} \delta_{l_{i,k}(u)}(-\infty, s + t - u]$$

$$\begin{aligned}
&= \sum_{k=1}^{A_i(u)} \delta_{l_{i,k}(t)+t-u}(-\infty, s+t-u] \\
&= \sum_{k=1}^{A_i(u)} \delta_{l_{i,k}(t)}(-\infty, s] \leq \mathcal{A}_i(t)(-\infty, s].
\end{aligned}$$

## 2.4 Basic performance processes

For  $t \geq 0$ ,  $i \in \mathbf{I}$ , the measure-valued state descriptor for route  $i$  is defined by

$$\mathcal{Z}_i(t) = \sum_{k=1}^{A_i(t)} \delta_{(v_{i,k}(t), l_{i,k}(t))}^+. \quad (3)$$

The random measure  $\mathcal{Z}_i(t) \in \mathbf{M}_2$  puts the unit mass at the point on the plane with the first coordinate equal to the residual transmission time of any flow present on route  $i$  at time  $t$ , and the second coordinate equal to its lead time. State descriptors of this kind were proposed in [Gromoll and Kruck \(2007\)](#) and [Gromoll et al. \(2008\)](#) for processor sharing queues with soft and firm deadlines, respectively. Let  $\mathcal{Z}(t) = (\mathcal{Z}_1(t), \dots, \mathcal{Z}_I(t))$ . It is often convenient to use the following  $\mathbf{M}$ -valued processes, which may be easily retrieved from  $\mathcal{Z}_i(\cdot)$ .

$$\mathcal{Q}_i(t) = \sum_{k=1}^{A_i(t)} \mathbb{I}_{(0,\infty)}(v_{i,k}(t)) \delta_{l_{i,k}(t)}, \quad \mathcal{W}_i(t) = \sum_{k=1}^{A_i(t)} v_{i,k}(t) \delta_{l_{i,k}(t)}. \quad (4)$$

The random measure  $\mathcal{Q}_i(t)$  (resp.  $\mathcal{W}_i(t)$ ) puts a unit of mass (resp., the mass equal to the corresponding residual transmission time) at the lead time of any flow present on route  $i$  at time  $t$ . State descriptors of this kind were originally introduced in [Doytchinov et al. \(2001\)](#) for a G/G/1 EDF queue with soft deadlines. Then  $\mathcal{Q}_i(t) = \langle \mathbf{1}, \mathcal{Q}_i(t) \rangle$  denotes the number of flows on the route  $i \in \mathbf{I}$  at time  $t$ . Let  $\mathcal{Q}(t) = (\mathcal{Q}_1(t), \dots, \mathcal{Q}_I(t))$ ,  $\mathcal{W}(t) = (\mathcal{W}_1(t), \dots, \mathcal{W}_I(t))$  and  $Q(t) = (Q_1(t), \dots, Q_I(t))$ .

The current smallest lead time process for route  $i$  will be denoted by  $C_i(\cdot)$ , where  $C_i(t) = L_{\mathcal{Q}_i(t)}$  for all  $t \geq 0$ .

## 2.5 Service protocol

The network operates under the preeemptive EDF policy, dynamically allocating bandwidth to flows with the shortest remaining lead time. In the case of preemption, we assume preempt-resume and no setup, switchover or other type of overhead. A definition of this protocol for resource sharing networks was given in [Kruk \(2016\)](#). We recall it here, introducing notation which will be used in the sequel.

Let  $t \geq 0$  be such that  $Q(t) \neq 0$  and let  $i_0 \in \mathbf{I}$ ,  $k_0 \leq A_{i_0}(t)$  be such that  $v_{i_0, k_0}(t) > 0$  and  $l_{i_0, k_0}(t)$  is the smallest of the lead times of the flows present in the system at time  $t$ . Here we assume that ties are broken in an arbitrary manner, for

example we may choose the smallest possible pair  $i_0, k_0$  with the required properties, according to the lexicographic order. (There are, of course, many other tie-breaking rules which can be implemented, for instance, a uniform draw between routes and/or flows may be chosen to avoid a bias between routes.) The flow  $k_0$  on route  $i_0$  is chosen for transmission with the maximal (i.e., unit) rate at time  $t$ . Let  $\mathbf{J}_1 = \mathbf{J} \setminus \mathcal{R}(i_0)$ ,  $\mathbf{I}_1 = \{i \in \mathbf{I} : \mathcal{R}(i) \subseteq \mathbf{J}_1\}$ . If  $\sum_{i \in \mathbf{I}_1} Q_i(t) = 0$  (in particular, if  $\mathbf{I}_1 = \emptyset$ ), then the assignment of flows for transmission at time  $t$  is finished, because no more flows can be transmitted at that time. Otherwise let  $i_1 \in \mathbf{I}_1$ ,  $k_1 \leq A_{i_1}(t)$  be such that  $v_{i_1, k_1}(t) > 0$  and  $l_{i_1, k_1}(t)$  is the smallest of the lead times of the flows present in the system at time  $t$  which are on routes belonging to  $\mathbf{I}_1$ . We choose the flow  $k_1$  on route  $i_1$  for transmission with unit rate at time  $t$ . Let  $\mathbf{J}_2 = \mathbf{J}_1 \setminus \mathcal{R}(i_1)$ ,  $\mathbf{I}_2 = \{i \in \mathbf{I} : \mathcal{R}(i) \subseteq \mathbf{J}_2\}$ . If  $\sum_{i \in \mathbf{I}_2} Q_i(t) = 0$  (in particular, if  $\mathbf{I}_2 = \emptyset$ ), we stop, otherwise we continue in this way until, at some step  $n$ , we get  $\sum_{i \in \mathbf{I}_n} Q_i(t) = 0$  and the assignment procedure at time  $t$  stops. This assignment is effective until either one of the ongoing transmissions is finished, or a new flow arrives to the system, when, subject to the same rules, a rearrangement may happen.

We will also consider a generalization of the above protocol, called a *general preemptive EDF policy*, in which the choice of routes for transmission is made according to the above algorithm, but for each  $m = 0, \dots, n - 1$ , the unit transmission rate for route  $i_m$  is split in some (possibly random and/or time-dependent) way between the flows present on this route with lead time  $C_{i_m}(t)$  at time  $t$ . If no two flows on any given route have the same deadline, any general EDF policy coincides with the EDF protocol defined above. However, if, say, two flows on the route  $i_0$  have lead time  $C_{i_0}(t)$  at time  $t$ , a general EDF policy may, for example, transmit both of them simultaneously, with equal rates, so the latter class of policies is broader than the former one.

*Remark 1* The indices  $n, i_0, \dots, i_{n-1}$  and the sets  $\mathbf{I}_1, \dots, \mathbf{I}_n, \mathbf{J}_1, \dots, \mathbf{J}_n$  depend on the time  $t$  (as well as on  $\omega \in \Omega$ ). In what follows, when we want to stress this dependence, we write  $i_0(t)$  instead of  $i_0$ ,  $\mathbf{I}_1(t)$  instead of  $\mathbf{I}_1$ , etc.

We will often compare the performance of the EDF policies defined above with the performance of other service disciplines. In some of them, flows on every route are scheduled for transmission according to a (general) EDF protocol, i.e., in the order of increasing lead times, although the choice of routes on which the transmission takes place does not have to conform to the EDF discipline. An example of such a system is a resource sharing network with fixed priorities of routes, in which flows on every route are served according to EDF.

## 2.6 Network equations

For  $i \in \mathbf{I}$ ,  $t \geq 0$  and  $s \in \mathbb{R}$ , let

$$E_i(t, s) = \sum_{k=Q_i(0)+1}^{A_i(t)} \delta_{l_{i,k}(t)}(-\infty, s] = \mathcal{A}_i(t)(-\infty, s] - \mathcal{Q}_i(0)(-\infty, t+s], \quad (5)$$

$$Z_i(t, s) = \mathcal{Q}_i(t)(-\infty, s]. \quad (6)$$

In words,  $E_i(t, s)$  is equal to the number of external arrivals by time  $t$  of flows on route  $i$  with lead times at time  $t$  less than or equal to  $s$  and  $Z_i(t, s)$  is the number of flows on route  $i$  with lead times at time  $t$  less than or equal to  $s$  which are still present in the system at that time. Note that  $Q_i(t) = \lim_{s \rightarrow \infty} Z_i(t, s)$ . Let  $E(t, s) = (E_i(t, s))_{i \in \mathbf{I}}$ ,  $Z(t, s) = (Z_i(t, s))_{i \in \mathbf{I}}$ . Similarly, the vectors  $D(t, s) = (D_i(t, s))_{i \in \mathbf{I}}$ ,  $T(t, s) = (T_i(t, s))_{i \in \mathbf{I}}$  denote the number of departures (i.e., transmission completions) and the cumulative transmission time by time  $t$  corresponding to each route  $i$  of flows with lead times at time  $t$  less than or equal to  $s$ . Note that

$$T_i(t, s) = \mathcal{V}_i(t)(-\infty, s] - \mathcal{W}_i(t)(-\infty, s], \quad i \in \mathbf{I}, \quad t \geq 0, \quad s \in \mathbb{R}. \quad (7)$$

Let

$$Y_i(t, s) = t - T_i(t, s) = t + \mathcal{W}_i(t)(-\infty, s] - \mathcal{V}_i(t)(-\infty, s], \quad i \in \mathbf{I}, \quad (8)$$

denote the cumulative idleness by time  $t$  with regard to transmission of flows on route  $i$  with lead times at time  $t$  less than or equal to  $s$  and let  $Y(t, s) = (Y_i(t, s))_{i \in \mathbf{I}}$ . For  $i \in \mathbf{I}$ ,  $t, t' \geq 0$  and  $s \in \mathbb{R}$ , let  $S_i(t', t, s)$  denote the number of transmission completions of flows on route  $i$  having lead times at time  $t$  less than or equal to  $s$ , by the time the system has spent  $t'$  units of time transmitting these flows. Finally, let

$$\mathfrak{X}(t, s) = (Z(t, s), D(t, s), T(t, s), Y(t, s)), \quad t \geq 0, \quad s \in \mathbb{R}. \quad (9)$$

By definition, all the components of  $\mathfrak{X}$  are nonnegative,  $D(\cdot, s - \cdot)$ ,  $T(\cdot, s - \cdot)$ ,  $Y(\cdot, s - \cdot)$  are nondecreasing in each coordinate and  $D(0, s) = T(0, s) = Y(0, s) = 0$  for  $s \geq 0$ . Moreover, all the coordinates of  $Z(t, \cdot)$  and of the increments  $D(\tilde{t}, \cdot - \tilde{t}) - D(t, \cdot - t)$ ,  $T(\tilde{t}, \cdot - \tilde{t}) - T(t, \cdot - t)$ ,  $Y(\tilde{t}, \cdot - \tilde{t}) - Y(t, \cdot - t)$  are nondecreasing for all  $\tilde{t} \geq t \geq 0$ . The process  $\mathfrak{X}$  satisfies the following *network equations*:

$$Z(t, s) = Z(0, t + s) + E(t, s) - D(t, s), \quad (10)$$

$$D_i(t, s) = S_i(T_i(t, s), t, s), \quad i \in \mathbf{I}, \quad (11)$$

$$T_i(t, s) + Y_i(t, s) = t, \quad i \in \mathbf{I}, \quad (12)$$

$$\sum_{i \in \mathcal{F}(j)} (T_i(\tilde{t}, s - \tilde{t}) - T_i(t, s - t)) \leq \tilde{t} - t, \quad j \in \mathbf{J}, \quad (13)$$

valid for every for  $\tilde{t} \geq t \geq 0$  and  $s \in \mathbb{R}$ . In particular, the equation (13) imposes the resource capacity constraints.

*Remark 2* The Eqs. (5)–(10) imply that for any  $t \geq 0$ , the function  $\mathfrak{X}(t, \cdot)$  is determined by the vectors of random measures  $\mathcal{A}(t)$ ,  $\mathcal{V}(t)$ ,  $\mathcal{Q}(0)$ ,  $\mathcal{Q}(t)$  and  $\mathcal{W}(t)$ . Hence, if the stochastic primitives for a resource sharing network are given, the state descriptor  $\mathcal{Z}(t)$  determines the corresponding function  $\mathfrak{X}(t, \cdot)$  uniquely. The converse statement is, in general, false. More precisely, if there are several flows with the same lead time on a given route, then the observation of the values of  $\mathfrak{X}$  (as opposed to  $\mathcal{Z}$ ) before the departure time of the first of them does not tell us anything about the division of the transmission time among these flows. This is illustrated by the following example.

*Example 1* Let  $I = J = 1$  and suppose that there are two flows with the same transmission time 1 and lead time 1 in the system at time zero. For simplicity, assume that there are no external arrivals. Consider two protocols. The first of them transmits one flow in the time interval  $[0, 1)$  and the other flow in  $[1, 2]$ , always with unit rate, while the second one transmits both of them simultaneously, with equal rates  $1/2$ . For  $t \in [0, 1)$ , the state descriptor corresponding to the first (second) one is  $Z^{(1)}(t) = \delta_{(1-t, 1-t)} + \delta_{(1, 1-t)}$  (resp.  $Z^{(2)}(t) = 2\delta_{(1-t/2, 1-t)}$ ), while the performance processes of the form (9) corresponding to these systems coincide on  $[0, 1)$ .

In what follows, we will see that the performance process  $\mathfrak{X}$  defined by (9), corresponding to an EDF resource sharing network, is efficient in the class of solutions of the corresponding network equations, in the sense of minimizing the associated idleness process  $Y$ . (For precise statements, see Definitions 1–2, 3–4 and 5–7, to follow.) To prove such results rigorously, we must exclude the possibility of “fake transmissions” of flows which are not present in the system during the transmission time. Such “fake transmissions” formally reduce the system idleness, although they make no influence on the real behavior of the network. More precisely, we will only consider systems satisfying the additional equations

$$\int_0^t \mathbb{I}_{[Z_i(u, s-u)=0]} d_u T_i(u, s-u) = 0, \quad i \in \mathbf{I}, \quad (14)$$

valid for all  $t \geq 0, s \in \mathbb{R}$ , and

$$\int_0^t \mathbb{I}_{[Z_i(u, s_2-u)=Z_i(u, s_1-u)]} d_u (T_i(u, s_2-u) - T_i(u, s_1-u)) = 0, \quad (15)$$

valid for all  $t \geq 0, s_1 < s_2, i \in \mathbf{I}$ . By definition, the process  $\mathfrak{X}$  for a general EDF resource sharing network (and, in fact, for any “reasonable” service discipline) satisfies (14)–(15).<sup>1</sup> Using (12), we can rewrite (15) in an equivalent form

$$\int_0^t \mathbb{I}_{[Z_i(u, s_2-u)=Z_i(u, s_1-u)]} d_u (Y_i(u, s_1-u) - Y_i(u, s_2-u)) = 0. \quad (16)$$

The Eqs. (10)–(15), together with the nonnegativity and monotonicity assumptions listed below (9), express general properties of resource sharing networks with all resources having the unit service rate. Note that the departure counting functions  $S_i$  appearing in (11) depend not only on the stochastic primitives, but also on the network service protocol.

<sup>1</sup> It is easy to find a solution  $\mathfrak{X}$  to the network Eqs. (10)–(13), together with the nonnegativity and monotonicity assumptions made below (9), which does *not* satisfy (14)–(15). Consider, for example, the empty G/G/1 system with  $Z = E = D = Y = 0$  and  $S = 0$ . This, of course, does *not* correspond to any “reasonable” queueing system: its interpretation is the system wasting all its power, while serving no one. Clearly, in order to use any performance criterion based on idleness and/or resource utilization, we must exclude such “pathological” objects. To this end, instead of relying only on the reader’s intuitive understanding what a “reasonable” service policy is, we convey some of the intuition behind the notions of file transmission and idleness in a “real” network by means of the additional Eqs. (14)–(15).

### 3 Minimality and additive minimality

In this section we recall the main definitions and results from Kruk (2016) which will be used in the sequel.

#### 3.1 Minimality

**Definition 1** Let  $\mathfrak{X}^{(k)} = (Z^{(k)}, D^{(k)}, T^{(k)}, Y^{(k)}), k = 1, 2$ , be two performance processes of the form (9) for resource sharing networks, having the same incidence matrix  $A$  and the same stochastic primitives (in particular, with  $Z^{(1)}(0, \cdot) = Z^{(2)}(0, \cdot)$  and the same external arrival function  $E$ ), satisfying (10)–(15), together with the non-negativity and monotonicity assumptions made below (9). We write  $\mathfrak{X}^{(1)} \preceq \mathfrak{X}^{(2)}$  if  $Y^{(1)}(\omega) \leq Y^{(2)}(\omega)$  (or, equivalently,  $T^{(1)}(\omega) \geq T^{(2)}(\omega)$ ) for every  $\omega \in \Omega$ .

**Definition 2** A performance process  $\mathfrak{X}$  of the form (9), satisfying (10)–(15), together with the nonnegativity and monotonicity assumptions made below (9), is called pathwise minimal if for any process  $\mathfrak{X}'$  such that  $\mathfrak{X}' \preceq \mathfrak{X}$ , we have  $\mathfrak{X} \preceq \mathfrak{X}'$ .

The relation “ $\preceq$ ” is reflexive and transitive, although it is not necessarily antisymmetric. Indeed, the functions  $S_i$  in the two systems under comparison are, in general different, so (11), together with the identity  $T^{(1)}(\omega) \equiv T^{(2)}(\omega)$ , do not have to imply that  $D^{(1)}(\omega) \equiv D^{(2)}(\omega)$ . However, if we truncate the performance processes of the type (9) to their last two coordinates  $(T(t, s), Y(t, s))$ , then “ $\preceq$ ” is a partial ordering on the set of such pairs and a performance process is pathwise minimal if and only if it is a minimal element relative to this ordering.

*Remark 3* In probability theory, it is customary to treat random variables as equivalence classes of measurable functions which coincide almost surely. Consequently, these objects are defined only up to a set of  $\mathbf{P}$  measure zero. From this point of view, in Definition 1 we should have required that  $Y^{(1)}(\omega) \leq Y^{(2)}(\omega)$  for almost all (instead of all)  $\omega \in \Omega$ . Here we take a slightly different, but equivalent, standpoint, treating every random variable defining the stochastic models under consideration as a representative from the corresponding equivalence class (arbitrarily chosen, but fixed thereafter), with well-defined values for each  $\omega \in \Omega$ . This convention is convenient for making pathwise comparisons for every (instead of almost every) possible scenario, but it does not have any essential influence on all that follows.

**Theorem 1** *The vector  $\mathfrak{X}$  of performance processes given by (9), corresponding to the EDF resource sharing protocol defined in Sect. 2.5, is pathwise minimal.*

The following result is a partial converse of Theorem 1.

**Theorem 2** *Let a performance process  $\mathfrak{X}$  for a resource sharing network be pathwise minimal. Then the flows on each route  $i \in \mathbf{I}$  of this network are scheduled for transmission according to a (general) EDF protocol.*

### 3.2 Additive minimality

**Definition 3** Let  $\mathfrak{X}^{(k)} = (Z^{(k)}, D^{(k)}, T^{(k)}, Y^{(k)}), k = 1, 2$ , be as in Definition 1. We write  $\mathfrak{X}^{(1)} \preccurlyeq \mathfrak{X}^{(2)}$  if  $\sum_{i \in \mathbf{I}} Y_i^{(1)}(\omega) \leq \sum_{i \in \mathbf{I}} Y_i^{(2)}(\omega)$  (or, equivalently,  $\sum_{i \in \mathbf{I}} T_i^{(1)}(\omega) \geq \sum_{i \in \mathbf{I}} T_i^{(2)}(\omega)$ ) for every  $\omega \in \Omega$ .

Clearly, the relation  $\mathfrak{X}' \preceq \mathfrak{X}$  implies  $\mathfrak{X}' \preccurlyeq \mathfrak{X}$ , but the opposite implication is not necessarily true.

**Definition 4** A performance process  $\mathfrak{X}$  of the form (9), satisfying (10)–(15), together with the nonnegativity and monotonicity assumptions made below (9), is called additively minimal if for any process  $\mathfrak{X}'$  such that  $\mathfrak{X}' \preccurlyeq \mathfrak{X}$ , we have  $\mathfrak{X} \preccurlyeq \mathfrak{X}'$ .

Remarks made after Definition 2, up to Remark 3, have obvious counterparts for Definitions 3–4. We also have

*Remark 4* An additively minimal process  $\mathfrak{X}$  is pathwise minimal.

A pathwise minimal process does not have to be additively minimal. In fact, even a preemptive EDF resource sharing network, defined in Sect. 2.5, may fail to be additively minimal, because its tie-breaking rule is not always optimal in this respect. The following theorem shows that in the absence of ties, EDF resource sharing networks are additively minimal.

**Theorem 3** *Assume that the stochastic primitives for a resource sharing network are such that no two flows on different routes have the same deadline. Then the vector  $\mathfrak{X}$  of performance processes given by (9), corresponding to the EDF resource sharing protocol defined in Sect. 2.5, is additively minimal.*

*Remark 5* The “no tie” assumption of Theorem 3 is satisfied if no two initial flows on different routes have the same deadline, the initial condition is independent on other stochastic primitives,  $N_i, i \in \mathbf{I}$ , are mutually independent delayed renewal processes,  $\{l_{i,k}\}_{k \geq 1}, i \in \mathbf{I}$ , form mutually independent i.i.d. sequences, independent on  $N_i, i \in \mathbf{I}$ , and either all the interarrival time distributions, or all the initial lead time distributions are continuous. The latter assumption holds, for example, if the arrival processes  $N_i, i \in \mathbf{I}$ , are Poisson. (From the point of view presented in Remark 3, in this case we have to choose representatives of all the stochastic primitives under consideration in such a way that there is no tie for every  $\omega \in \Omega$ .) Consequently, the assumption of Theorem 3 is rather mild.

A converse of Theorem 3 is, in general, false, even in the absence of ties.

## 4 Edge minimality

### 4.1 Edge comparison for real functions on $\mathbb{R}$

**Definition 5** Let  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  be such that for some  $a \in \mathbb{R}$  we have  $f \equiv g$  on  $(-\infty, a]$  and let  $c = c_{f,g} = \sup\{a \in \mathbb{R} : f(x) = g(x) \forall x \in (-\infty, a]\}$ . We write  $f \ll g$  if either  $c = \infty$  (i.e.,  $f \equiv g$  on  $\mathbb{R}$ ), or  $c < \infty$  and there exists  $\epsilon = \epsilon_{f,g} > 0$  such that  $f \leq g$  on  $[c, c + \epsilon]$ .

*Remark 6* By the definition of  $c = c_{f,g}$ , if  $f \ll g$  and  $c < \infty$ , then for every  $\delta > 0$  there exists a point  $x \in [c, c + (\delta \wedge \epsilon_{f,g})]$  such that  $f(x) < g(x)$ .

**Lemma 1** *The relation “ $\ll$ ” is a partial ordering.*

*Proof* By definition,  $f \ll f$  for every function  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

Suppose that  $f \ll g$ ,  $g \ll f$  and  $f \neq g$ . In this case,  $c = c_{f,g} < \infty$ . Moreover,  $f \leq g$  and  $g \leq f$  on  $[c, c + (\epsilon_{f,g} \wedge \epsilon_{g,f})]$ , so  $f \equiv g$  on the latter interval. This, however, contradicts the definition of  $c_{f,g}$ .

Suppose that  $f \ll g$  and  $g \ll h$ . If either  $f = g$  or  $g = h$ , then clearly  $f \ll h$ , so we may assume that  $c_{f,g} < \infty$  and  $c_{g,h} < \infty$ . If  $c_{f,g} < c_{g,h}$ , then  $c_{f,h} = c_{f,g}$  (see Remark 6) and  $f \leq g \equiv h$  on  $[c_{f,g}, (c_{f,g} + \epsilon_{f,g}) \wedge c_{g,h}]$ , so  $f \ll h$ . Similarly, if  $c_{g,h} < c_{f,g}$ , then  $c_{f,h} = c_{g,h}$  and  $f \equiv g \leq h$  on  $[c_{g,h}, (c_{g,h} + \epsilon_{g,h}) \wedge c_{f,g}]$ , so again  $f \ll h$ . Finally, if  $c_{f,g} = c_{g,h}$ , then  $c_{f,h} = c_{f,g} = c_{g,h}$  and  $f \leq g \leq h$  on  $[c_{f,g}, c_{f,h} + (\epsilon_{f,g} \wedge \epsilon_{g,h})]$ , which yields  $f \ll h$ .  $\square$

The partial ordering “ $\ll$ ” is not linear, even on sets of the form

$$B_{c,d} = \{f \in \mathbb{R}^{\mathbb{R}} : f \equiv d \text{ on } (-\infty, c]\}, \quad c, d \in \mathbb{R}. \quad (17)$$

For example, the functions  $f \equiv 0$  and  $g(x) = x \sin(1/x)$ ,  $x > 0$ ,  $g(x) = 0$ ,  $x \leq 0$ , belong to  $B_{0,0}$ , but neither  $f \ll g$ , nor  $g \ll f$ , holds.

*Remark 7* If  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ , then the inequality  $f \leq g$  does not, in general, imply  $f \ll g$ . Indeed, the latter relation may not make sense, since  $f$  and  $g$  may not agree on any half-line of the form  $(-\infty, a]$ ,  $a \in \mathbb{R}$ . However, if  $f \leq g$  and  $f \equiv g$  on some half-line  $(-\infty, a]$ , then  $f \ll g$ . In particular, the relation “ $\ll$ ” is weaker than “ $\leq$ ” in any set  $B_{c,d}$  of the form (17).

*Remark 8* Let  $\mathcal{G} \subset \mathbb{R}^{\mathbb{R}}$  be a family of functions and let  $f \in \mathcal{G}$ . It is easy to see that  $f$  is a minimal element of  $\mathcal{G}$  with respect to the relation “ $\ll$ ” if and only if for every  $c > 0$ ,  $\epsilon > 0$  and every  $g \in \mathcal{G}$  satisfying  $f \equiv g$  on  $(-\infty, c)$ , the inequality  $g \leq f$  on  $[c, c + \epsilon)$  implies that  $f \equiv g$  on  $(-\infty, c + \epsilon)$ . Hence, in spite of the local nature of the relation “ $\ll$ ”, minimality of  $f$  in  $\mathcal{G}$  with respect to “ $\ll$ ” is actually a *global* property of the function  $f$  if  $\mathcal{G}$  is large enough.

## 4.2 Edge minimality for networks

**Definition 6** Let  $\mathfrak{X}^{(k)} = (Z^{(k)}, D^{(k)}, T^{(k)}, Y^{(k)})$ ,  $k = 1, 2$ , be as in Definition 1. We write  $\mathfrak{X}^{(1)} \lll \mathfrak{X}^{(2)}$  if  $\sum_{i \in \mathbf{I}} Y_i^{(1)}(t, \cdot)(\omega) \lll \sum_{i \in \mathbf{I}} Y_i^{(2)}(t, \cdot)(\omega)$  (or, equivalently,  $\sum_{i \in \mathbf{I}} T_i^{(2)}(t, \cdot)(\omega) \lll \sum_{i \in \mathbf{I}} T_i^{(1)}(t, \cdot)(\omega)$ ) for every  $t \geq 0$  and  $\omega \in \Omega$ .

To see that the relations “ $\lll$ ” in Definition 6 make sense, let  $\mathfrak{X}^{(k)}$ ,  $k = 1, 2$ , be as above and let  $t \geq 0$ ,  $\omega \in \Omega$ . Then, because both systems  $\mathfrak{X}^{(1)}$ ,  $\mathfrak{X}^{(2)}$ , have the same stochastic primitives, the measure-valued arrival processes (1) corresponding to them coincide. Let  $c = c(t, \omega) = \min\{L_{\mathcal{A}_i(t)(\omega)} : i \in \mathbf{I}\} - 1$ . Then  $\mathcal{A}_i(t)(\omega)(-\infty, c] = 0$  for all  $i$ , so for  $k = 1, 2$ ,  $i \in \mathbf{I}$  and  $u \in [0, t]$ ,

$$\begin{aligned} 0 \leq Z_i^{(k)}(u, t + c - u)(\omega) &\leq Z_i^{(k)}(0, t + c)(\omega) + E_i(u, t + c - u)(\omega) \\ &\leq \mathcal{A}_i(u)(\omega)(-\infty, t + c - u] \leq \mathcal{A}_i(t)(\omega)(-\infty, c] = 0, \end{aligned}$$

where the second inequality follows from (10), the third inequality follows from (5)–(6) and the fourth one follows from (2). Hence, for  $k, i, u$  as above,  $Z_i^{(k)}(u, t + c - u)(\omega) = 0$ , so (14) yields

$$\begin{aligned} T_i^{(k)}(t, c)(\omega) &= \int_0^t d_u T_i^{(k)}(u, t + c - u)(\omega) \\ &= \int_0^t \mathbb{I}_{[Z_i^{(k)}(u, c+t-u)(\omega)=0]} d_u T_i^{(k)}(u, c + t - u)(\omega) = 0. \end{aligned}$$

This, together with the monotonicity of  $T_i^{(k)}(t, \cdot)(\omega)$ , implies that

$$\sum_{i \in \mathbf{I}} T_i^{(k)}(t, \cdot)(\omega) \in B_{c,0}, \quad k = 1, 2, \quad (18)$$

so, by (12),

$$\sum_{i \in \mathbf{I}} Y_i^{(k)}(t, \cdot)(\omega) \in B_{c,tI}, \quad k = 1, 2. \quad (19)$$

Analogous statements are valid for two arbitrary performance processes  $\mathfrak{X}^{(1)}, \mathfrak{X}^{(2)}$  of the form (9) for resource sharing networks having the same incidence matrix  $A$  (not necessarily with the same stochastic primitives). Indeed, let  $\mathcal{A}_i^{(k)}, k = 1, 2$ , be the corresponding measure-valued arrival processes. Then the above argument proves (18)–(19) with

$$c = c(t, \omega) = \min\{(L_{\mathcal{A}_i^{(1)}(t)(\omega)} \wedge L_{\mathcal{A}_i^{(2)}(t)(\omega)}) : i \in \mathbf{I}\} - 1.$$

Remark 7 and (18)–(19) show that the relation  $\mathfrak{X}' \preccurlyeq \mathfrak{X}$  implies  $\mathfrak{X}' \lll \mathfrak{X}$ . The opposite implication, however, is not necessarily true.

**Definition 7** A performance process  $\mathfrak{X}$  of the form (9), satisfying (10)–(15), together with the nonnegativity and monotonicity assumptions made below (9), is called edge minimal if for any process  $\mathfrak{X}'$  such that  $\mathfrak{X}' \lll \mathfrak{X}$ , we have  $\mathfrak{X} \lll \mathfrak{X}'$ .

In other words, the process  $\mathfrak{X}$  is edge minimal if the inequality  $\mathfrak{X}' \lll \mathfrak{X}$  for  $\mathfrak{X}' = (Z', D', T', Y')$  implies that  $\sum_{i \in \mathbf{I}} Y_i(\omega) \equiv \sum_{i \in \mathbf{I}} Y'_i(\omega)$  (hence  $\sum_{i \in \mathbf{I}} T_i(\omega) \equiv \sum_{i \in \mathbf{I}} T'_i(\omega)$ ) for each  $\omega \in \Omega$ . Remarks made after Definition 2, up to Remark 3, have obvious counterparts for Definitions 6–7.

In Remark 8, we have observed that minimality with respect to the relation “ $\lll$ ” in a sufficiently large set of “comparison functions” is a global property of the minimizer. Consequently, edge minimality of a performance process  $\mathfrak{X}$  is actually a strong global property of the associated “cumulative idleness distribution function”  $\sum_{i \in \mathbf{I}} Y_i$ . In particular, it implies the minimality of  $\sum_{i \in \mathbf{I}} Y_i$  in the sense of the pointwise functional inequality, as the following remark indicates.

*Remark 9* An edge minimal process  $\mathfrak{X}$  is additively minimal.

*Proof* Suppose that  $\mathfrak{X}$  is edge minimal and let  $\mathfrak{X}' = (Z', D', T', Y')$  be such that  $\mathfrak{X}' \preccurlyeq \mathfrak{X}$ . Then  $\mathfrak{X}' \lll \mathfrak{X}$ , so by edge minimality of  $\mathfrak{X}$ ,  $\sum_{i \in I} Y_i(\omega) \equiv \sum_{i \in I} Y'_i(\omega)$  for each  $\omega \in \Omega$ . This, in turn, clearly implies that  $\mathfrak{X} \preccurlyeq \mathfrak{X}'$ .  $\square$

**Corollary 1** A preemptive EDF resource sharing network may fail to be edge minimal in the presence of ties.

This follows from the above Remark 9 and Example 3 in Kruk (2016), presenting a preemptive EDF resource sharing network which fails to be additively minimal.

An additively minimal process does not have to be edge minimal, as the following example shows.

*Example 2* Let  $I = 3$ ,  $J = 2$ ,  $\mathcal{R}(1) = \{1\}$ ,  $\mathcal{R}(2) = \{2\}$  and  $\mathcal{R}(3) = \{1, 2\}$ . Let  $\mathcal{W}_1(0) = \mathcal{W}_2(0) = \delta_2$  and  $\mathcal{W}_3(0) = \delta_1$ . Assume, for simplicity, that there are no external arrivals. Suppose that the network protocol gives priority to routes 1, 2 and let  $\mathfrak{X}$  be the corresponding performance process in the form (9). Then, at time  $t \in [0, 1)$ , the flows on routes 1, 2 are being transmitted, while the flow on route 3 waits for transmission, so

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} t, & s \geq 2, \\ 3t, & s < 2. \end{cases}$$

For  $t \in [1, 2)$ , the system transmits the flow on route 3, while other routes are empty, and hence

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 2t-1, & s \geq 2, \\ 2t+1, & 1 \leq s < 2, \\ 3t, & s < 1. \end{cases}$$

For  $t \geq 2$  the system is empty and

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 3t-3, & s \geq 2, \\ 3t-1, & 1 \leq s < 2, \\ 3t, & s < 1. \end{cases}$$

Proceeding as in Example 5 from Kruk (2016), we can show that the system  $\mathfrak{X}$  is additively minimal.

Let  $\mathfrak{X}' = (Z', D', T', Y')$  be the performance process describing the same network, with the same stochastic primitives, but using the EDF scheduling policy. For  $t \in [0, 1)$ , the EDF system transmits the flow on route 3, while the flows on other routes wait for transmission, so

$$\sum_{i=1}^3 Y'_i(t, s-t) = \begin{cases} 2t, & s \geq 1, \\ 3t, & s < 1. \end{cases}$$

For  $t \in [1, 2)$ , the EDF system transmits the flows on routes 1, 2, while route 3 is empty, so

$$\sum_{i=1}^3 Y'_i(t, s-t) = \begin{cases} t+1, & s \geq 2, \\ 3t-1, & 1 \leq s < 2, \\ 3t, & s < 1. \end{cases}$$

Finally, for  $t \geq 2$  the EDF system is empty and  $\sum_{i=1}^3 Y'_i(t, \cdot) = \sum_{i=1}^3 Y_i(t, \cdot)$ .

It is easy to check that  $\mathfrak{X}' \lll \mathfrak{X}$ , while clearly the processes  $\sum_{i \in \mathbf{I}} Y'_i$  and  $\sum_{i \in \mathbf{I}} Y_i$  do not coincide. Consequently, the system  $\mathfrak{X}$  is not edge minimal.

**Remark 9** and Example 2 show that the property of edge minimality is stronger than additive minimality. Hence, the following theorem is a refinement of Theorem 3. In spite of this fact, its proof appears to be somewhat shorter and simpler than the proof of Theorem 3 presented in Kruk (2016).

**Theorem 4** *Assume that the stochastic primitives for a resource sharing network are such that no two flows on different routes have the same deadline. Then the vector  $\mathfrak{X}$  of performance processes given by (9), corresponding to a general EDF resource sharing protocol defined in Sect. 2.5, is edge minimal.*

*Proof* Suppose that  $\mathfrak{X}$  is not edge minimal. Let  $\mathfrak{X}' = (Z', D', T', Y')$  be such that  $\mathfrak{X}' \lll \mathfrak{X}$ , but the relation  $\mathfrak{X} \lll \mathfrak{X}'$  does not hold. Fix  $\omega \in \Omega$  such that for some  $t \geq 0$ , the relation  $\sum_{i \in \mathbf{I}} Y_i(t, \cdot)(\omega) \ll \sum_{i \in \mathbf{I}} Y'_i(t, \cdot)(\omega)$  fails. In the remainder of the proof, all the random objects under consideration are evaluated at this  $\omega$ . Let  $Q'(t) = \lim_{s \rightarrow \infty} Z'(t, s)$  for each  $t \geq 0$  and let

$$t_0 = \sup \left\{ \bar{t} \geq 0 : \sum_{i \in \mathbf{I}} Y'_i(t, \cdot) = \sum_{i \in \mathbf{I}} Y_i(t, \cdot) \forall t \in [0, \bar{t}] \right\}. \quad (20)$$

Since  $Y'(0, \cdot) = Y(0, \cdot) = 0$  by assumption, the set in (20) is nonempty. On the other hand,  $t_0 < \infty$  by assumption. Arguing as in the proof of Theorem 3, we get

$$\left( \sum_{i \in \mathbf{I}} T_i, \sum_{i \in \mathbf{I}} Y_i \right) (t, \cdot) = \left( \sum_{i \in \mathbf{I}} T'_i, \sum_{i \in \mathbf{I}} Y'_i \right) (t, \cdot), \quad 0 \leq t \leq t_0. \quad (21)$$

Let

$$t_1 = \inf \{t > t_0 : \Delta N(t) \neq 0 \text{ or } \Delta Q(t) \neq 0 \text{ or } \Delta Q'(t) \neq 0\}. \quad (22)$$

For the rest of the proof we will use the notation introduced in Sect. 2.5.

Let  $\hat{t} \in (t_0, t_1)$  be such that  $f \ll g$  and  $f \neq g$ , where

$$f(s) = \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, s), \quad g(s) = \sum_{i \in \mathbf{I}} Y_i(\hat{t}, s), \quad s \in \mathbb{R}. \quad (23)$$

Let  $c = c_{f,g}$  and  $\epsilon = \epsilon_{f,g}$  be as in Definition 5. By the no-tie assumption, without loss of generality we may assume that

$$\epsilon < \epsilon_0 := \{\min |l_{i,k}(\hat{t}) - l_{i',k'}(\hat{t})| : i, i' \in \mathbf{I}, i \neq i', k \leq A_i(\hat{t}), k' \leq A_{i'}(\hat{t})\}. \quad (24)$$

We claim that  $c$  is the lead time of a flow present at both systems at time  $\hat{t}$ . If  $c$  is not equal to the lead time of any flow present at time  $\hat{t}$  in the system  $\mathfrak{X}$  (hence, by (21)–(22), the relation  $\hat{t} \in (t_0, t_1)$  and the fact that both systems have the same stochastic primitives, also in  $\mathfrak{X}'$ ), we can choose  $\delta > 0$  such that  $Z_i(\hat{t}, c + \delta) = Z_i(\hat{t}, c - \delta)$ ,  $Z'_i(\hat{t}, c + \delta) = Z'_i(\hat{t}, c - \delta)$  for every  $i \in \mathbf{I}$ . Using (22) and the relation  $\hat{t} \in (t_0, t_1)$  again, for all  $i \in \mathbf{I}, t \in [t_0, \hat{t}]$ , we get

$$Z_i(t, c + \delta + \hat{t} - t) = Z_i(t, c - \delta + \hat{t} - t), \quad Z'_i(t, c + \delta + \hat{t} - t) = Z'_i(t, c - \delta + \hat{t} - t). \quad (25)$$

Substituting  $t := \hat{t}, s_1 := c - \delta + \hat{t}, s_2 := c + \delta + \hat{t}$  into (16) and using (25), for every  $i \in \mathbf{I}$  we obtain

$$\begin{aligned} Y_i(\hat{t}, c - \delta) - Y_i(\hat{t}, c + \delta) &= Y_i(t_0, c - \delta + \hat{t} - t_0) - Y_i(t_0, c + \delta + \hat{t} - t_0), \\ Y'_i(\hat{t}, c - \delta) - Y'_i(\hat{t}, c + \delta) &= Y'_i(t_0, c - \delta + \hat{t} - t_0) - Y'_i(t_0, c + \delta + \hat{t} - t_0). \end{aligned}$$

This, together with (21), yields

$$\begin{aligned} &\sum_{i \in \mathbf{I}} Y_i(\hat{t}, c + \delta) - \sum_{i \in \mathbf{I}} Y_i(\hat{t}, c - \delta) \\ &= \sum_{i \in \mathbf{I}} Y_i(t_0, c + \delta + \hat{t} - t_0) - \sum_{i \in \mathbf{I}} Y_i(t_0, c - \delta + \hat{t} - t_0) \\ &= \sum_{i \in \mathbf{I}} Y'_i(t_0, c + \delta + \hat{t} - t_0) - \sum_{i \in \mathbf{I}} Y'_i(t_0, c - \delta + \hat{t} - t_0) \\ &= \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, c + \delta) - \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, c - \delta). \end{aligned} \quad (26)$$

On the other hand, by the definitions of  $f, g$  and  $c$ ,

$$\sum_{i \in \mathbf{I}} Y_i(\hat{t}, c - \delta) = \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, c - \delta). \quad (27)$$

By (23) and (26)–(27), we have  $f(c + \delta) = g(c + \delta)$  for  $\delta > 0$  small enough, which contradicts the definition of  $c$ .

We have just checked that  $c$  is the lead time of a flow present at both systems at time  $\hat{t}$ . By Remark 6, there exists  $\delta \in (0, \epsilon_0)$  such that  $f(c + \delta) < g(c + \delta)$ , where  $f, g$  and  $\epsilon_0$  are given by (23)–(24). Because of (21) and properties of a general EDF policy described in Sect. 2.5 (in particular, the fact that it does not change priorities of routes in  $[t_0, t_1]$  and always transmits flows on a chosen route with unit rate), this is possible only if in the time interval  $[t_0, \hat{t}]$  the system  $\mathfrak{X}'$  has transmitted (at least partially) a flow with lead time  $c$  at time  $\hat{t}$ , while the EDF system  $\mathfrak{X}$  has not scheduled any flow with this lead

time for transmission in this time interval. Let  $B = \{j \in \{0, \dots, n-1\} : C_{ij}(\hat{t}) < c\}$ . The subsequent analysis is divided into several cases.

If  $B = \emptyset$ , then

$$C_{i_0}(\hat{t}) \geq c. \quad (28)$$

Since  $c$  is the lead time of a flow present at the EDF system at time  $\hat{t}$ , the strict inequality in (28) contradicts the definition of  $i_0$ . Hence,  $C_{i_0}(\hat{t}) = c$ , so in the time interval  $[t_0, \hat{t}]$  the EDF system is transmitting a flow with lead time  $c$  at time  $\hat{t}$  and we have obtained a contradiction.

In what follows, we assume that  $B \neq \emptyset$ . Let us observe that (21), the fact that both systems have the same stochastic primitives, the no-tie assumption, the definition of  $c$  and the above-mentioned properties of a general EDF policy imply that in the time interval  $[t_0, \hat{t}]$  both systems transmitted with unit rate (or abstained from transmission of) flows on the same routes, provided that their lead times at time  $\hat{t}$  are less than  $c$ . Let  $\bar{j} = \max(B)$ .

Suppose that  $\bar{j} < n-1$ . Then

$$c \leq C_{i_{\bar{j}+1}}(\hat{t}). \quad (29)$$

If equality holds in (29), then we have a contradiction, just as in the case of  $B = \emptyset$ . Otherwise, a flow with lead time  $c$  should be chosen for transmission instead of a flow with lead time  $C_{i_{\bar{j}+1}}(\hat{t})$  at the  $\bar{j} + 1$ -st stage of the EDF scheduling at the time  $t_0$  (and hence at  $\hat{t}$ ), which again leads to a contradiction.

Finally, if  $\bar{j} = n-1$ , then  $C_{i_{n-1}}(\hat{t}) < c$ . Thus, in the time interval  $[t_0, \hat{t}]$ , the system  $\mathfrak{X}'$  transmits, with unit rate, flows on all the routes on which the system  $\mathfrak{X}$  does. Moreover, in this time interval,  $\mathfrak{X}'$  also provides some transmission to an additional flow, with lead time  $c$  at time  $\hat{t}$ . However, (21)–(22), the no-tie assumption and the fact that both systems have the same stochastic primitives, implies that for  $t \in [t_0, t_1)$  and  $i \in \mathbb{I}$ ,  $Q_i(t) > 0$  if and only if  $Q'_i(t) > 0$ . Consequently, it is possible to transmit an additional flow by the system  $\mathfrak{X}$  in the time interval  $[t_0, \hat{t}]$ , without interrupting the active transmission of any other flow in the system, which contradicts the definition of the EDF protocol.  $\square$

A converse of Theorem 4 is, in general, false, even in the absence of ties, as the following example shows.

*Example 3 (Edge minimal “anti-EDF” linear network)* Consider the network with the same topology as in Example 2. Assume that  $\mathcal{W}_1(0) = \delta_1$ ,  $\mathcal{W}_2(0) = \mathbf{0}$  and  $\mathcal{W}_3(0) = \delta_2$ . For simplicity, let us also assume that there is only one external arrival to this system after time zero, namely, an arrival to route 2 at time 1 with  $v_{2,1} = 1$ ,  $l_{2,1} = 2$ . Suppose that the network protocol gives priority to route 3 and let  $\mathfrak{X}$  be the corresponding performance process in the form (9). Then, at time  $t \in [0, 1)$ , the flow on route 3 is being transmitted, while the flow on route 1 waits for transmission, so

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 2t, & s \geq 2, \\ 3t, & s < 2. \end{cases}$$

In the time interval  $[1, 2)$ , the flows on routes 1, 2 are being transmitted, so for  $t \in [1, 2)$ ,

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} t+1, & s \geq 3, \\ 2t, & 2 \leq s < 3, \\ 2t+1, & 1 \leq s < 2, \\ 3t, & s < 1. \end{cases}$$

If a performance process  $\mathfrak{X}' = (Z', D', T', Y')$  is such that  $\mathfrak{X}' \lll \mathfrak{X}$ , then  $f(s) := \sum_{i=1}^3 Y'_i(2, s-2) \lll \sum_{i=1}^3 Y_i(2, s-2) =: g(s)$ . Due to (14)–(15),  $f$  (as well as  $g$ ) is constant on the intervals  $(-\infty, 1)$ ,  $[1, 2)$ ,  $[2, 3)$ ,  $[3, \infty)$  and, moreover,  $f \equiv g \equiv 6$  on  $(-\infty, 1)$ . Hence,  $c = c_{f,g} \in \{1, 2, 3, \infty\}$ . In particular,  $c \geq 1$ , so  $f(s) \leq g(s) = 5$  for  $s \in [1, (1 + \epsilon_{f,g}) \wedge 2)$ . By (12), this means that  $\sum_{i=1}^3 T'_i(2, -1) = 6 - f(1) \geq 1$ , so the flow with lead time  $-1$  at time 2 (i.e., lead time 1 at time 0) on the first route has been fully transmitted by the system  $\mathfrak{X}'$  by time 2. Its transmission time is 1, and hence  $\sum_{i=1}^3 T'_i(2, -1) = 1$ , which implies that  $f(1) = g(1) = 5$  and  $c \geq 2$ . Proceeding similarly, we upgrade  $c \geq 2$  to  $c \geq 3$ , and finally to  $c = \infty$ . Hence  $f \equiv g$ , and consequently  $Q'(2) = 0$ . By the network topology and the equality  $Q_2(1) = W_2(1) = 1$ , this is possible only if the flows on routes 1, 2 are being transmitted with unit rate on the time interval  $[1, 2)$  and the flow on route 3 is being transmitted with unit rate on the time interval  $[0, 1)$ . We have shown that  $\mathfrak{X}' = \mathfrak{X}$  and the system  $\mathfrak{X}$  is edge minimal.

**Corollary 2** *There may be more than one edge minimal service policy for a given network, even if no two flows on different routes have the same deadline. Consequently, edge minimality does not characterize the corresponding service protocol uniquely.*

This follows immediately from Example 3 and the fact that, by Theorem 4, the EDF service protocol is also edge minimal for the network considered in that example.

In the light of non-uniqueness of an edge minimal protocol for a given network, pointed out in Corollary 2, one may wonder whether First-In, First-Out (FIFO), SRPT, proportional fairness, or some other “natural” policy different from EDF is also optimal in this sense. The answer to this question is negative (except for the case of the initial lead times of all the flows being equal, in which FIFO coincides with EDF). Indeed, an edge minimal system  $\mathfrak{X}$  is necessarily additively minimal by Remark 9, and hence it is pathwise minimal by Remark 4. Thus, by Theorem 2, the flows on each route of the network are being transmitted according to a (general) EDF policy. In other words, distinct edge minimal policies for a given network differ only by the transmission rates (priorities) assigned to routes, but on any given route, conditioned on the amount of transmission time available for this route, they prioritize flows according to the EDF protocol. In particular, for a single resource, single customer class network ( $I = J = 1$ ), the notions of edge minimality, additive minimality and pathwise minimality coincide and they are all equivalent to the resource working under a generalized EDF policy (compare Corollary 1 in Kruk 2016).

## 5 Local edge minimality

### 5.1 Definition

In this section, we let  $\mathcal{Z}$  be a state process of the form (3), describing the time evolution of some resource sharing network. Let  $\mathfrak{X}$  denote the corresponding performance process of the form (9), satisfying (10)–(15), together with the nonnegativity and monotonicity assumptions made below (9). We also consider another resource sharing network, having the same incidence matrix and the same stochastic primitives as the former one, with the corresponding state process  $\mathcal{Z}'$  and performance process  $\mathfrak{X}'$ , which is subject to (10)–(15), together with the nonnegativity and monotonicity assumptions made below (9).

**Definition 8** The state process  $\mathcal{Z}$  is called locally edge minimal at a time  $t_0 \geq 0$ , if there exists a strictly positive random variable  $h$  such that for any state process  $\mathcal{Z}'$  as above, satisfying

$$\mathcal{Z}'(t_0)(\omega) = \mathcal{Z}(t_0)(\omega), \quad \omega \in \Omega, \quad (30)$$

we have  $\sum_{i \in \mathbf{I}} Y_i(t, \cdot)(\omega) \ll \sum_{i \in \mathbf{I}} Y'_i(t, \cdot)(\omega)$  (or, equivalently,  $\sum_{i \in \mathbf{I}} T'_i(t, \cdot)(\omega) \ll \sum_{i \in \mathbf{I}} T_i(t, \cdot)(\omega)$ ) for every  $\omega \in \Omega$ ,  $t \in (t_0, t_0 + h(\omega))$ .

The state process  $\mathcal{Z}$  is called locally edge minimal, if it is locally edge minimal at every  $t_0 \geq 0$ .

Intuitively a network described by the state process  $\mathcal{Z}$  is locally edge minimal if, restarted at any time  $t_0$ , it behaves “locally”—on a short time horizon—like an edge minimal network with the same initial state. Note that Definition 8 is formulated in terms of the process  $\mathcal{Z}$ , rather than  $\mathfrak{X}$ . This is because  $\mathcal{Z}(t_0)$  contains detailed information about the state of the system at time  $t_0$  (in particular, about the residual service times of all the flows in the system at this time), which, in general, cannot be inferred from  $\mathfrak{X}(t_0)$ , or even the whole trajectory of the process  $\mathfrak{X}$  on the interval  $[0, t_0]$ , see Remark 2.

For future reference, note that the Eqs. (3)–(4), (8) and (30) imply

$$Y'(t_0, s)(\omega) = Y(t_0, s)(\omega), \quad s \in \mathbb{R}, \quad \omega \in \Omega. \quad (31)$$

In general, the performance process corresponding to a locally edge minimal state process may fail to be additively minimal (and hence, by Remark 9, edge minimal), as the following reworking of Example 3 illustrates.

*Example 4* Consider the network with the same topology as in Examples 2, 3. Let  $\mathcal{W}_1(0) = \mathcal{W}_3(0) = \delta_3$  and  $\mathcal{W}_2(0) = \mathbf{0}$ . For simplicity, assume that there is only one external arrival to this system after time zero, namely, an arrival to route 2 at time 1 with  $v_{2,1} = 1$ ,  $l_{2,1} = 1$ . Consider the network protocol giving priority to routes 1, 2. Denote by  $\mathcal{Z}$  the corresponding state process of the form (3) and by  $\mathfrak{X}$  the corresponding performance process in the form (9), respectively. Then, at time  $t \in [0, 1)$ , the flow on route 1 is being transmitted, while the flow on route 3 waits for transmission, so

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 2t, & s \geq 3, \\ 3t, & s < 3. \end{cases}$$

In the time interval  $[1, 2)$ , the flow on route 2 is being transmitted, while the flow on route 3 still waits for transmission, so for  $t \in [1, 2)$ ,

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 2t, & s \geq 3, \\ 2t+1, & 2 \leq s < 3, \\ 3t, & s < 2. \end{cases}$$

In the time interval  $[2, 3)$ , the flow on route 3 is being transmitted and hence for  $t \in [2, 3)$ ,

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 2t, & s \geq 3, \\ 3t-1, & 2 \leq s < 3, \\ 3t, & s < 2. \end{cases}$$

Finally, for  $t \geq 3$ , the system is empty and hence

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 3t-3, & s \geq 3, \\ 3t-1, & 2 \leq s < 3, \\ 3t, & s < 2. \end{cases}$$

It is not hard to check that the process  $\mathcal{Z}$  is locally edge minimal. Indeed, let  $t_0 \geq 0$  and let the state process  $\mathcal{Z}'$  satisfy (30). For a given  $t \geq 0$ , let

$$f_t(s) = \sum_{i \in \mathbf{I}} Y'_i(t, s-t), \quad g_t(s) = \sum_{i \in \mathbf{I}} Y_i(t, s-t), \quad s \in \mathbb{R}.$$

If  $0 \leq t_0 < 1$ , then at every time  $t \in (t_0, 1)$ , only the flows on routes 1, 3, both with lead time  $3 - t$ , are present in both  $\mathcal{Z}$  and  $\mathcal{Z}'$ . Moreover,  $\mathcal{R}(1) \cap \mathcal{R}(3) \neq \emptyset$ , so the overall rate of transmission of both these flows in the time interval  $[t_0, t]$  in  $\mathcal{Z}'$  is not greater than one. Consequently, either  $f_t(3) = 2t$ , which corresponds to the system  $\mathcal{Z}'$  transmitting these flows with the overall unit rate in  $[t_0, t]$  and results in  $f_t \equiv g_t$ , or  $f_t(3) > 2t$ , implying  $g_t \ll f_t$ ,  $c_{f_t, g_t} = 3$ .

If  $1 \leq t_0 < t < 2$ , then the smallest lead time in both  $\mathcal{Z}$ ,  $\mathcal{Z}'$  at time  $t$  is  $2 - t$ , the lead time of the flow on route 2. Moreover, its transmission excludes the possibility of transmitting the remaining flow, located on route 3 in both systems. Hence,

$$f_t(2) - f_{t_0}(2) \geq 2(t - t_0) = g_t(2) - g_{t_0}(2) = g_t(2) - f_{t_0}(2), \quad (32)$$

where the last equality follows from (31). If equality holds in (32), then the system  $\mathcal{Z}'$  is transmitting the flow on route 2 with unit rate on  $[t_0, t]$ , so  $\mathcal{Z}' \equiv \mathcal{Z}$  on  $[t_0, t]$ , and hence  $f_t \equiv g_t$  by (31). If the inequality (32) is strict, then  $g_t \ll f_t$  and  $c_{f_t, g_t} = 2$ .

The case of  $2 \leq t_0 < t < 3$  is similar to the ones considered above (if fact, simpler, because there is only one flow present in both systems in the time interval  $[t_0, t]$ ).

Finally, for  $3 \leq t_0 < t$ , both systems are empty in  $[t_0, t]$  and hence, by (14), both of them incur the maximal possible idleness in this time interval.

We have checked that the process  $\mathcal{Z}$  is locally edge minimal. (Alternatively, one may deduce this fact from Theorem 7, to follow.) At the same time, the corresponding process  $\mathfrak{X}$  is not additively minimal. Indeed, consider another protocol for the network under consideration, giving priority to route 3. Denote by  $\mathfrak{X}'$  the corresponding performance process in the form (9). Then  $\sum_{i=1}^3 Y'_i(t, \cdot) \equiv \sum_{i=1}^3 Y_i(t, \cdot)$  for  $t \in [0, 1]$ . For  $t \in [1, 2]$ , under the latter discipline, the route 3 is empty and the flows on the remaining routes are being transmitted, so

$$\sum_{i=1}^3 Y'_i(t, s-t) = \begin{cases} t+1, & s \geq 3, \\ 2t+1, & 2 \leq s < 3, \\ 3t, & s < 2. \end{cases}$$

For  $t \geq 2$ , under the second protocol, the system is empty and hence

$$\sum_{i=1}^3 Y'_i(t, s-t) = \begin{cases} 3t-3, & s \geq 3, \\ 3t-1, & 2 \leq s < 3, \\ 3t, & s < 2. \end{cases}$$

Consequently,  $\mathfrak{X}' \preccurlyeq \mathfrak{X}$ , while the relation  $\mathfrak{X} \preccurlyeq \mathfrak{X}'$  is false.

## 5.2 The case of no ties between routes

If the stochastic primitives for a resource sharing network are such that no two flows on different routes have the same deadline (which will be referred to as “no ties between routes”), the class of locally edge minimal service protocols coincides with the class of general EDF resource sharing policies. This equivalence is established by the following Theorems 5 and 6.

**Theorem 5** *Under the assumption of no ties between routes, the state process  $\mathcal{Z}$  given by (3), corresponding to a general EDF resource sharing protocol defined in Sect. 2.5, is locally edge minimal.*

*Proof* Fix  $t_0 \geq 0$ . We will show that  $\mathcal{Z}$  is locally edge minimal at  $t_0$ . Let a process  $\mathcal{Z}'$ , subject to the assumptions made at the beginning of this section, be such that (30) holds. Fix  $\omega \in \Omega$ . In the remainder of the proof, all the random objects under consideration are evaluated at this  $\omega$ . We will use the notation introduced in Sect. 2.5. Put

$$h = \min\{v_{i,k}(t_0) : i \in \mathbf{I}, k \leq A_i(t_0), v_{i,k}(t_0) > 0\} \wedge \inf\{u > 0 : \Delta N(t_0 + u) \neq 0\}. \quad (33)$$

By definition,  $h > 0$  and  $h$  is defined solely in terms of the process  $\mathcal{Z}$  (i.e., it does not depend on the choice of  $\mathcal{Z}'$ ). Moreover, in both  $\mathcal{Z}$  and  $\mathcal{Z}'$  we have the same flows (no arrivals/departures) in the time interval  $(t_0, t_0 + h)$ .

Let  $\hat{t} \in (t_0, t_0 + h)$ , let the functions  $f, g$  be defined by (23) and let  $c = c_{f,g}$  be as in Definition 5. Our aim is to show that  $g \ll f$ .

If  $c = \infty$ , then  $f \equiv g$ , so  $g \ll f$ .

Assume that  $c < \infty$ . Proceeding as in the proof of Theorem 4 (with  $t_0 + h$  instead of  $t_1$ ), we check that  $c$  is the lead time of a flow present at both systems at time  $\hat{t}$ . In particular, by the definition of  $i_0 = i_0(\hat{t})$ , we have  $c \geq C_{i_0}(\hat{t})$ .

If  $c = C_{i_0}(\hat{t})$ , then in the time interval  $[t_0, \hat{t}]$  the EDF system is transmitting, at unit rate, flows with lead time  $c$  at time  $\hat{t}$ . Moreover, by the no-tie assumption, there is only one route with flows having this lead time, present in  $\mathcal{Z}$  (and hence in  $\mathcal{Z}'$ ) at time  $\hat{t}$ . Thus, by (14), (31) and (13), for any  $\epsilon > 0$  satisfying (24) and any  $s \in [c, c + \epsilon]$ ,

$$\begin{aligned} g(s) &= \sum_{i \in \mathbf{I}} Y_i(\hat{t}, s) = \sum_{i \in \mathbf{I}} Y_i(t_0, s + (\hat{t} - t_0)) + (I - 1)(\hat{t} - t_0) \\ &= \sum_{i \in \mathbf{I}} Y'_i(t_0, s + (\hat{t} - t_0)) + (I - 1)(\hat{t} - t_0) \\ &\leq \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, s) = f(s). \end{aligned}$$

This implies that  $g \ll f$ .

As in the proof of Theorem 4, we observe that in the time interval  $[t_0, \hat{t}]$  both systems transmit with unit rate (or abstain from transmission of) flows on the same routes, provided that their lead times at time  $\hat{t}$  are less than  $c$ . Consequently, if  $c > C_{i_0}(\hat{t})$ , then only flows on routes belonging to  $\{i_0\} \cup \mathbf{I}_1(\hat{t})$  are being transmitted by  $\mathcal{Z}'$  in  $[t_0, \hat{t}]$ , and hence  $c \geq C_{i_1}(\hat{t})$ , where  $i_1 = i_1(\hat{t})$ .

Proceeding similarly, after  $n - 1$  such steps, we get either  $g \ll f$ , or  $c > C_{i_{n-1}}(\hat{t})$ . In the latter case, the system  $\mathcal{Z}'$  transmits, with unit rate, flows on all the routes on which the system  $\mathcal{Z}$  does. Moreover,  $\mathcal{Z}'$  also transmits an additional flow, with lead time  $c$ , at least on some part of the time interval  $[t_0, \hat{t}]$ . This, however, contradicts the definition of the EDF protocol, as in the last paragraph of the proof of Theorem 4.  $\square$

A modification of the above argument may be used to show the following converse to Theorem 5.

**Theorem 6** *Let  $\mathcal{Z}$  be a locally edge minimal state process of a resource sharing network with stochastic primitives having no ties between routes. Then flows in the system  $\mathcal{Z}$  are being transmitted under a general EDF service protocol.*

*Proof* Fix  $t_0 \geq 0$  and let  $h > 0$  be as in Definition 8. Decreasing  $h$ , if necessary, we may assume that  $h$  is not greater than the right-hand side of (33). Consequently, there are no arrivals or departures in the time interval  $(t_0, t_0 + h)$  under any service policy which satisfies the resource capacity constraints (13) and agrees with the one used in the system  $\mathcal{Z}$  on  $[0, t_0]$ . We will show that in the time interval  $[t_0, t_0 + h]$  flows are being scheduled for transmission according to a general EDF discipline.

Let  $\mathcal{Z}'$  be a state process with the same stochastic primitives as  $\mathcal{Z}$ , corresponding to the service discipline which emulates the one used in  $\mathcal{Z}$  up to the time  $t_0$  and uses the preemptive EDF protocol thereafter. Fix  $\omega \in \Omega$ . In what follows, the random objects under consideration are evaluated at this  $\omega$ .

Let  $\hat{t} \in (t_0, t_0 + h)$  and let the functions  $f, g$  be defined by (23). By local edge minimality of  $\mathcal{Z}$ , we have  $g \ll f$ . On the other hand, proceeding as in the proof

of Theorem 5, we get  $f \ll g$ , so  $f \equiv g$ . This, however, is possible only if the service protocol used by  $\mathcal{Z}$  in the time interval  $[t_0, \hat{t}]$  coincides with a general EDF policy. Indeed, under the assumption of no ties between routes, (30) and the identity  $\sum_{i \in I} Y'_i(\hat{t}, \cdot) = \sum_{i \in I} Y_i(\hat{t}, \cdot)$  (and hence, by (12),  $\sum_{i \in I} T'_i(\hat{t}, \cdot) = \sum_{i \in I} T_i(\hat{t}, \cdot)$ ) for each  $\hat{t} \in (t_0, t_0 + h)$ , imply that in the time interval  $[t_0, t_0 + h]$ , transmission in both systems takes place on the same routes and, moreover, on any of these routes, the flows with the same (i.e., the earliest) deadline are being transmitted.  $\square$

**Corollary 3** *An edge minimal resource sharing network may fail to be locally edge minimal.*

This follows immediately from Example 3 and Theorem 6.

**Corollary 4** *Let  $\mathcal{Z}$  be a locally edge minimal state process of a resource sharing network with stochastic primitives having no ties between routes. Then the corresponding performance process  $\mathfrak{X}$  is edge minimal.*

This follows immediately from Theorems 4 and 6. Example 4 shows that the assumption of no ties between routes in Corollary 4 cannot be dropped.

### 5.3 General case

If ties between lead times of flows on different routes are allowed, the situation is much more complicated. In particular, a characterization of general EDF service protocols by local edge minimality (and vice versa) is no longer available. The following example shows that in the presence of ties, Theorem 5 is, in general, false.

*Example 5* Let the network topology be the same as in Examples 2–3. Assume that  $\mathcal{W}_1(0) = \mathcal{W}_2(0) = \mathcal{W}_3(0) = \delta_1$ . For simplicity, let us also assume that there are no external arrivals. Let  $\mathcal{Z}$  ( $\mathcal{Z}'$ ) be the state process in the form (3) for this network in which priority is given to route 3 (resp., routes 1, 2). Both the above service policies conform to the definition of the preemptive EDF protocol given in Sect. 2.5. For  $t \in (0, 1]$ , we have

$$\sum_{i=1}^3 Y_i(t, s-t) = \begin{cases} 2t, & s \geq 1, \\ 3t, & s < 1. \end{cases}$$

while

$$\sum_{i=1}^3 Y'_i(t, s-t) = \begin{cases} t, & s \geq 1, \\ 3t, & s < 1. \end{cases}$$

Consequently,  $\sum_{i=1}^3 Y'_i(t, \cdot) \ll \sum_{i=1}^3 Y_i(t, \cdot)$ ,  $\sum_{i=1}^3 Y'_i(t, \cdot) \neq \sum_{i=1}^3 Y_i(t, \cdot)$ , and hence  $\mathcal{Z}$  is not locally edge minimal at 0.

Furthermore, there may be many different service protocols for a given network which are indistinguishable in terms of the corresponding cumulative idleness processes  $\sum_{i \in \mathbf{I}} Y_i$ , and hence they are “equally good” with respect to any optimality criterion based on these processes, like additive minimality or (local) edge minimality. The following example illustrates this point.

*Example 6* Let  $I = J = 4$ ,  $\mathcal{R}(1) = \{1, 2\}$ ,  $\mathcal{R}(2) = \{3, 4\}$ ,  $\mathcal{R}(3) = \{1, 3\}$ , and  $\mathcal{R}(4) = \{2, 4\}$ . Put  $\mathcal{W}_1(0) = \mathcal{W}_2(0) = \mathcal{W}_3(0) = \mathcal{W}_4(0) = \delta_1$ . As in Example 5, let us also assume that there are no external arrivals. Let  $\mathcal{Z}$  ( $\mathcal{Z}'$ ) be the state process in the form (3) for this network in which priority is given to routes 1, 2 (resp., routes 3, 4). Both the above service policies belong to the class of preemptive EDF protocols defined in Sect. 2.5. Then for  $t \in [0, 2]$ ,

$$\sum_{i=1}^4 Y_i(t, s-t) = \sum_{i=1}^4 Y'_i(t, s-t) = \begin{cases} 2t, & s \geq 1, \\ 4t, & s < 1, \end{cases}$$

while for  $t > 2$ ,

$$\sum_{i=1}^4 Y_i(t, s-t) = \sum_{i=1}^4 Y'_i(t, s-t) = \begin{cases} 4t - 4, & s \geq 1, \\ 4t, & s < 1. \end{cases}$$

Furthermore, consider the service policy which is a “mixture” of the above two protocols in the following sense: for  $t \in [0, 1/2] \cup [1, 3/2]$ , the priority is given to the routes 1, 2, while in the time intervals  $[1/2, 1], [3/2, 2]$ , the routes 3, 4 get preferential treatment. (This is not an EDF policy in the sense of Sect. 2.5, but it belongs to a more general class of EDF policies, to be defined later.) Let  $\mathcal{Z}''$  be the corresponding state process. Then  $\sum_{i=1}^4 Y''_i \equiv \sum_{i=1}^4 Y_i \equiv \sum_{i=1}^4 Y'_i$ . It may be checked that all the three above processes are locally edge minimal.

As Example 5 suggests, in order to provide counterparts of Theorems 5 and 6 in the general situation, we have to identify a suitable “subclass” of the family of (general) EDF protocols, coinciding with locally edge minimal policies. Moreover, as Example 6 indicates, such a “subclass” must be closed under appropriate “mixtures in time” of suitable service disciplines. The (somewhat involved) construction of this “subclass” is described in the sequel.

Fix  $\omega \in \Omega$ . In what follows, the random objects under consideration are evaluated at this  $\omega$ . Let  $t \geq 0$  be such that  $Q(t) \neq 0$ . We will identify the set of routes on which the transmission, with unit rate, should hold at time  $t$ .

Let  $\mathcal{P}(t) \subset 2^{\mathbf{I}}$  be the family of subsets of  $\mathbf{I}$  of the form  $\{i_0, \dots, i_{n-1}\}$  constructed as all possible outcomes of the EDF scheduling policy described in Sect. 2.5. If there are no ties between routes, as in Sect. 5.2, then  $|\mathcal{P}(t)| = 1$ , but otherwise we may have  $|\mathcal{P}(t)| \geq 2$ , e.g., in Example 5,  $\mathcal{P}(0) = \{\{1, 2\}, \{3\}\}$ . We will choose a subset  $\mathcal{P}_s(t) \subseteq \mathcal{P}(t)$  corresponding to performance processes which are locally edge minimal at  $t$ . (In our Example 5,  $\mathcal{P}_s(0) = \{\{1, 2\}\}$ .)

Let  $C^{(1)}(t) = \min\{C_i(t) : i \in \mathbf{I}\}$ . By the definition of the EDF policy, for every  $B \in \mathcal{P}(t)$ , we have  $\min\{C_i(t) : i \in B\} = C^{(1)}(t)$ . Let

$$\begin{aligned} m^{(1)}(t) &= \max\{|\{i \in B : C_i(t) = C^{(1)}(t)\}| : B \in \mathcal{P}(t)\}, \\ \mathcal{P}^{(1)}(t) &= \{B \in \mathcal{P}(t) : |\{i \in B : C_i(t) = C^{(1)}(t)\}| = m^{(1)}(t)\}. \end{aligned}$$

In words, the elements of  $\mathcal{P}^{(1)}(t)$  correspond to these EDF scheduling policies at time  $t$ , which have active transmission of flows with the smallest deadline in the system at time  $t$  on as many routes, as possible. For instance, in Example 5,  $\mathcal{P}^{(1)}(0) = \{\{1, 2\}\}$ . If  $|\mathcal{P}^{(1)}(t)| = 1$ , we put  $\mathcal{P}_s(t) = \mathcal{P}^{(1)}(t)$ , otherwise we take the second step as follows.

For  $k = 1, 2, \dots$ , in the  $k + 1$ -st step of our search procedure, if  $\max\{C_i(t) : i \in B\} = C^{(k)}(t)$  for every  $B \in \mathcal{P}^{(k)}(t)$ , let  $\mathcal{P}_s(t) = \mathcal{P}^{(k)}(t)$ . Otherwise, let

$$\begin{aligned} C^{(k+1)}(t) &= \min_{B \in \mathcal{P}^{(k)}(t)} \min\{C_i(t) : i \in B, C_i(t) > C^{(k)}(t)\}, \\ m^{(k+1)}(t) &= \max\{|\{i \in B : C_i(t) = C^{(k+1)}(t)\}| : B \in \mathcal{P}^{(k)}(t)\}, \\ \mathcal{P}^{(k+1)}(t) &= \{B \in \mathcal{P}^{(k)}(t) : |\{i \in B : C_i(t) = C^{(k+1)}(t)\}| = m^{(k+1)}(t)\}. \end{aligned}$$

In words, the elements of  $\mathcal{P}^{(k+1)}(t)$  correspond to these scheduling policies in  $\mathcal{P}^{(k)}(t)$  at time  $t$ , which have active transmission of flows with the smallest available lead time greater than  $C^{(k)}(t)$  at time  $t$  on as many routes, as possible. If  $|\mathcal{P}^{(k+1)}(t)| = 1$ , we take  $\mathcal{P}_s(t) = \mathcal{P}^{(k+1)}(t)$ . Otherwise, we take the next (i.e.,  $k + 2$ -nd) step and continue the procedure in this way until we finally choose the set  $\mathcal{P}_s(t)$ . In what follows, the index  $k$  such that  $\mathcal{P}_s(t) = \mathcal{P}^{(k)}(t)$  will be denoted by  $k_s = k_s(t)$ .

By construction, between the arrivals and departures of flows, the quantities  $C^{(k)}(\cdot)$  decrease at unit rate, while  $m^{(k)}(\cdot)$  and the sets  $\mathcal{P}^{(k)}(\cdot)$  remain constant. Consequently,  $k_s(\cdot)$  and the sets  $\mathcal{P}_s(\cdot)$  do not change in these time intervals, either. To fix ideas, for  $t \geq 0$  such that  $Q(t) = 0$ , we additionally put  $\mathcal{P}_s(t) = \{\emptyset\}$ .

The assignment of transmission rates for the routes  $i \in \mathbf{I}$  (and hence the corresponding service policy) is called *strong EDF consistent* if

$$T(t, \infty) := \lim_{s \uparrow \infty} T(t, s) = \int_0^t r(u) du, \quad t \geq 0, \quad (34)$$

where the transmission rate process  $r(\cdot)$  is such that for almost every  $t \geq 0$ , with respect to the Lebesgue measure,

$$r(t) = (\mathbb{I}_{B(t)}(i))_{i \in \mathbf{I}} \quad \text{for some } B(t) \in \mathcal{P}_s(t). \quad (35)$$

We assume that the sample paths of  $r$  are Lebesgue measurable, so that the integral in (34) makes sense, but we do not require them to be r.c.l.l..

A strong EDF consistent policy such that the set  $B(t)$  in (35) remains constant between the arrivals and departures of flows and, moreover, for each  $i \in B(t)$ , a single flow  $k$  on route  $i$  such that  $l_{i,k}(t) = C_i(t)$  is scheduled for transmission, with unit rate, at time  $t$ , will be called a *strong preemptive EDF protocol*. Clearly, any strong EDF

policy is an EDF policy in the sense introduced in Sect. 2.5, but the opposite implication is, in general, false. For instance, both the service protocols from Example 5 are EDF, while only the second one is a strong EDF policy. However, under the assumption of no ties between routes, the family  $\mathcal{P}_s(t) = \mathcal{P}(t)$  contains a single element and hence the classes of EDF policies and strong EDF policies coincide.

More generally, if a strong EDF consistent policy is such that for almost every  $t \geq 0$  (with respect to the Lebesgue measure) and for each  $i \in B(t)$ , the unit transmission rate for route  $i$  is split in some (possibly random and/or time-dependent) way between the flows present on this route with lead time  $C_i(t)$  at time  $t$ , we have a *general strong preemptive EDF protocol*. If the set  $B(t)$  in (35) remains constant between the arrivals and departures of flows, then the corresponding general strong EDF policy is a general EDF protocol in the sense of Sect. 2.5. If, moreover, no two flows on any given route have the same deadline, the corresponding general strong EDF policy is a strong EDF protocol, as defined above.

The following result generalizes Theorems 5–6.

**Theorem 7** *Let  $\mathcal{Z}$  denote a state process of the form (3), with the corresponding performance process  $\mathfrak{X}$  in the form (9) satisfying (10)–(15), together with the nonnegativity and monotonicity assumptions made below (9). Then  $\mathcal{Z}$  is locally edge minimal if and only if the flows in  $\mathcal{Z}$  are being transmitted under a general strong EDF service protocol.*

The proof of Theorem 7 uses the ideas from the proofs of Theorems 5–6, but it is unavoidably more complicated.

*Proof of Theorem 7* Let  $\mathcal{Z}$  be the state process corresponding to a general strong EDF service protocol. Fix  $t_0 \geq 0$ . We will show that  $\mathcal{Z}$  is locally edge minimal at  $t_0$ . Let a process  $\mathcal{Z}'$ , subject to the assumptions made at the beginning of this section, be such that (30) holds. Fix  $\omega \in \Omega$ . In the remainder of the proof, all the random objects under consideration are evaluated at this  $\omega$ . Let  $h$  be given by (33). By definition,  $h > 0$  and in both  $\mathcal{Z}$  and  $\mathcal{Z}'$  we have the same flows in the time interval  $(t_0, t_0 + h)$ .

Let  $\hat{t} \in (t_0, t_0 + h)$ , let the functions  $f, g$  be defined by (23) and let  $c = c_{f,g}$  be as in Definition 5. Our aim is to show that  $g \ll f$ .

If  $c = \infty$ , then  $f \equiv g$ , so  $g \ll f$ .

Assume that  $c < \infty$ . Proceeding as in the proof of Theorem 4 (with  $t_0 + h$  instead of  $t_1$ ), we check that  $c$  is the lead time of a flow present at both systems at time  $\hat{t}$ . In particular, by the definition of  $C^{(1)}(\hat{t})$ , we have  $c \geq C^{(1)}(\hat{t})$ .

If  $c = C^{(1)}(\hat{t})$ , then in the time interval  $[t_0, \hat{t}]$  the system  $\mathcal{Z}$  is transmitting, at unit rate, flows with lead time  $c$  at time  $\hat{t}$  on  $m^{(1)} := m^{(1)}(\hat{t})$  routes. Let

$$\epsilon = \min\{|l_{i,k}(\hat{t}) - c| : i \in \mathbf{I}, k \leq A_i(\hat{t}), v_{i,k}(\hat{t}) > 0, l_{i,k}(\hat{t}) \neq c\}. \quad (36)$$

By (31), for any  $s \in [c, c + \epsilon]$ ,

$$\begin{aligned} g(s) &= \sum_{i \in \mathbf{I}} Y_i(\hat{t}, s) = \sum_{i \in \mathbf{I}} Y_i(t_0, s + (\hat{t} - t_0)) + (I - m^{(1)})(\hat{t} - t_0) \\ &= \sum_{i \in \mathbf{I}} Y'_i(t_0, s + (\hat{t} - t_0)) + (I - m^{(1)})(\hat{t} - t_0). \end{aligned} \quad (37)$$

By the definition of  $m^{(1)}$ , at any time  $t \in (t_0, t_0 + h)$ , it is impossible to schedule the transmission of flows with lead time  $c$  at time  $\hat{t}$  on more than  $m^{(1)}$  routes, so for  $s$  as above, we have

$$\sum_{i \in \mathbf{I}} Y'_i(t_0, s + (\hat{t} - t_0)) + (I - m^{(1)})(\hat{t} - t_0) \leq \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, s) = f(s).$$

This, together with (37), shows that  $g(s) \leq f(s)$  for  $s \in [c, c + \epsilon]$ , and consequently  $g \ll f$ .

From now on, let us suppose that  $c > C^{(1)}(\hat{t})$ . Let  $s = C^{(1)}(\hat{t})$ . By the definition of  $c$ , we have  $f(s) = g(s)$ , so (37) and the definition of  $C^{(1)}(\hat{t})$  imply that in the time interval  $[t_0, \hat{t}]$  the system  $\mathcal{Z}'$  is transmitting flows with lead time  $s$  at time  $\hat{t}$  on at least  $m^{(1)}$  routes. Recalling that it is impossible to schedule the transmission of flows with lead time  $s$  at time  $\hat{t}$  on more than  $m^{(1)}$  routes in this time interval, we conclude from (37) that at almost every time  $t \in [t_0, \hat{t}]$  (with respect to the Lebesgue measure), the transmission of these flows by the system  $\mathcal{Z}'$  is active on exactly  $m^{(1)}$  routes, with unit rate on each of them.

Let  $p = \max\{p' : C^{(p')}(\hat{t}) < c\}$ . By assumption,  $p \geq 1$ . By definition of a general strong EDF protocol, for  $p' = 1, \dots, p$ , at almost every time  $t \in [t_0, \hat{t}]$  (with respect to the Lebesgue measure), the system  $\mathcal{Z}$  is transmitting flows with lead time  $C^{(p')}(\hat{t})$  at time  $\hat{t}$  on  $m^{(p')}$  routes, with unit rate on each of them. Using induction on  $p' = 1, \dots, p$  and arguing similarly as in the last paragraph, we show that the same statement is valid for the system  $\mathcal{Z}'$ . (We have just proved the base case and the proof of the induction step is similar.)

We first consider the case in which  $p < k_s$ . Then  $c \leq C^{(p+1)}(\hat{t})$ . If  $c < C^{(p+1)}(\hat{t})$ , then, by the definition of a strong EDF protocol, the system  $\mathcal{Z}$  is not transmitting any flow with lead time  $c$  at time  $\hat{t}$  in the time interval  $[t_0, \hat{t}]$ . In the same time interval, the system  $\mathcal{Z}'$  has to transmit such a flow (at least partially), to avoid a contradiction with the definition of  $c$ . This, however, contradicts the definition of  $C^{(p+1)}(\hat{t})$ , since it is possible to transmit, with rate one, flows with lead time  $C^{(p')}(\hat{t})$  at time  $\hat{t}$  on  $m^{(p')}$  routes,  $p' = 1, \dots, p$ , and transmit a flow with lead time smaller than  $C^{(p+1)}(\hat{t})$  at the same time. This shows that  $c = C^{(p+1)}(\hat{t})$ . Consequently, for any  $s \in [c, c + \epsilon]$ , where  $\epsilon$  is given by (36), we have

$$\begin{aligned} g(s) &= \sum_{i \in \mathbf{I}} Y_i(\hat{t}, s) = \sum_{i \in \mathbf{I}} Y_i(t_0, s + (\hat{t} - t_0)) + \left( I - \sum_{p'=1}^{p+1} m^{(p')} \right) (\hat{t} - t_0) \\ &= \sum_{i \in \mathbf{I}} Y'_i(t_0, s + (\hat{t} - t_0)) + \left( I - \sum_{p'=1}^{p+1} m^{(p')} \right) (\hat{t} - t_0) \\ &\leq \sum_{i \in \mathbf{I}} Y'_i(\hat{t}, s) = f(s), \end{aligned}$$

where the second equality follows from the definition of a strong EDF protocol, the third one follows from (31), while the inequality follows from the last paragraph and the definitions of  $C^{(p+1)}(\hat{t})$ ,  $m^{(p+1)}$ . We have obtained the relation  $g \ll f$  again.

Finally, if  $p = k_s$ , then  $c > C^{(k_s)}(\hat{t})$ . Arguing as at the beginning of the last paragraph, we conclude that this implies a transmission of a flow with lead time  $c$  at time  $\hat{t}$  in the time interval  $[t_0, \hat{t}]$  by the system  $\mathcal{Z}'$  which, in turn, contradicts the definition of  $k_s$  in the description of a strong EDF scheduling algorithm.

Summarizing the above considerations, in every possible case we have  $g \ll f$ , and hence the state process  $\mathcal{Z}$  is locally edge minimal at  $t_0$ .

Conversely, if the state process  $\mathcal{Z}$  is locally edge minimal, then, proceeding as in the proof of Theorem 6, we may check that the corresponding service protocol agrees with a general strong EDF policy.  $\square$

**Theorem 8** *Let  $\mathcal{Z}$  be a locally edge minimal state process of the form (3). Then the corresponding performance process  $\mathfrak{X}$  is pathwise minimal.*

*Proof* By Theorem 7, the flows in  $\mathcal{Z}$  are being transmitted according to a general strong EDF service protocol. If, moreover, it is an EDF policy in the sense of Sect. 2.5, then it is pathwise minimal by Theorem 1. Let us recall, however, that there exist more general locally edge minimal protocols, for example changing transmission rates between the arrivals and departures of flows (see Example 6), so an additional argument is necessary in order to prove their pathwise minimality.

Suppose that  $\mathfrak{X}$  is not pathwise minimal. Let  $\mathfrak{X}' = (Z', D', T', Y')$  be such that  $\mathfrak{X}' \preceq \mathfrak{X}$  and  $\mathfrak{X} \not\preceq \mathfrak{X}'$ . Fix  $\omega \in \Omega$  such that  $Y'(\omega) \leq Y(\omega)$ , but the relation  $Y(\omega) \leq Y'(\omega)$  is false. For the remainder of the proof, all the random objects under consideration are evaluated at this  $\omega$ . Let

$$t_0 = \sup\{\bar{t} \geq 0 : Y'(\bar{t}, \cdot) = Y(\bar{t}, \cdot) \forall t \in [0, \bar{t}]\}. \quad (38)$$

Since  $Y'(0, \cdot) = Y(0, \cdot) = 0$  by assumption, the set in (38) is nonempty. On the other hand,  $t_0 < \infty$  by assumption. Arguing as in the proof of Theorem 1 (which may be found on pp. 268–270 of Kruskal 2016), we get

$$(T, Y)(t, \cdot) = (T', Y')(t, \cdot), \quad 0 \leq t \leq t_0. \quad (39)$$

Let  $Q'(t) = \lim_{s \rightarrow \infty} Z'(t, s)$  for each  $t \geq 0$  and let  $t_1$  be given by (22).

By (39), we have

$$T_i(t_0, s) - T_i(t_0, s-) = T'_i(t_0, s) - T'_i(t_0, s-), \quad i \in \mathbf{I}, s \in \mathbb{R}. \quad (40)$$

Since the arrival processes for the both systems under consideration are the same, for any  $i \in \mathbf{I}$  and  $s \in \mathbb{R}$ , the transmission time sufficient to empty the  $i$ -th route from the flows with lead time  $s$  at time  $t_0$  in one of them is also sufficient to accomplish the same task in the other one. This fact, together with (40), implies that for every  $i \in \mathbf{I}$  and  $s \in \mathbb{R}$

$$Z_i(t_0, s) > Z_i(t_0, s-) \Leftrightarrow Z'_i(t_0, s) > Z'_i(t_0, s-). \quad (41)$$

Consequently, for  $t \in [t_0, t_1]$ , the quantities  $k_s(t)$ ,  $m^{(k)}(t)$ ,  $C^{(k)}(t)$  and the sets  $\mathcal{P}^{(k)}(t)$  are the same for the both systems under consideration.

Let  $\hat{t} \in (t_0, t_1)$  be such that

$$Y'(\hat{t}, \cdot) \neq Y(\hat{t}, \cdot) \quad (42)$$

(the existence of such  $\hat{t}$  follows from the definition of  $t_0$ ). Note that

$$Y'(\hat{t}, \cdot) \leq Y(\hat{t}, \cdot). \quad (43)$$

Let the functions  $f, g$  be defined by (23) and let  $c = c_{f,g}$  be as in Definition 5. By (42)–(43), we have  $f \leq g$  and  $f \neq g$ , so  $c < \infty$ . Using (41) and arguing as in the proof of Theorem 4, we check that  $c$  is the lead time of a flow present at both systems at time  $\hat{t}$ .

Proceeding as in the proof of Theorem 7, we show that for  $\epsilon$  given by (36), we have  $g(s) \leq f(s)$  for every  $s \in [c, c + \epsilon]$ . Consequently,  $f \equiv g$  on  $(-\infty, c + \epsilon)$ , which contradicts the definition of  $c$ .  $\square$

## 6 Prospective application: fluid models and fluid limits

This section is devoted to a somewhat informal discussion of a prospective application of notions of minimality and local edge minimality to defining fluid models and fluid limits for EDF networks with resource sharing.

For this section, for each  $i \in \mathbf{I}$ , we make the following additional distributional assumptions.  $N_i(\cdot)$  is a delayed renewal process with rate  $\alpha_i \in (0, \infty)$ . The random variables  $\{v_{i,k}\}_{k \geq 1}$  form an independent and identically distributed (i.i.d.) sequence with finite mean  $m_i$ . Finally, the random variables  $\{l_{i,k}\}_{k \geq 1}$  are nonnegative and they form an i.i.d. sequence with a finite first moment. Let  $\alpha = (\alpha_1, \dots, \alpha_I)$ .

In our context, fluid models are deterministic, continuous analogs of resource sharing networks, in which individual flows are replaced by a divisible commodity (fluid), moving along  $I$  routes with  $J$  resources (nodes). They usually arise from formal functional law of large numbers approximations of the corresponding stochastic flow level models. In the existing literature, fluid models are defined rigorously in terms of the corresponding fluid model equations.

The deterministic analogs of the network Eqs. (10)–(13) are the following fluid model equations:

$$\bar{Z}(t, s) = \bar{Z}(0, t + s) + \alpha(t + (s \wedge 0))^+ - \bar{D}(t, s), \quad (44)$$

$$\bar{D}_i(t, s) = \bar{T}_i(t, s)/m_i, \quad i \in \mathbf{I}, \quad (45)$$

$$\bar{T}_i(t, s) + \bar{Y}_i(t, s) = t, \quad i \in \mathbf{I}, \quad (46)$$

$$\sum_{i \in \mathcal{F}(j)} (\bar{T}_i(\tilde{t}, s - \tilde{t}) - \bar{T}_i(t, s - t)) \leq \tilde{t} - t, \quad j \in \mathbf{J}, \quad (47)$$

where  $\tilde{t} \geq t \geq 0$ ,  $s \in \mathbb{R}$ . Roughly speaking, the Eqs. (44)–(47) have been obtained from (10)–(13) by replacing the underlying random quantities by their means. This

replacement is justified in the limit, under suitable fluid scaling, by appropriate functional laws of large numbers. Note that we do not include counterparts of (14)–(15) in our fluid model equations. In fact, such conditions are, in general, false for fluid limits, since there may be a small number of waiting flows in the pre-limit network for extended periods of time, keeping the system busy, but asymptotically vanishing under fluid scaling. This is the case, for example, in a G/G/1 queue (which is a special case of our network, obtained by letting  $I = J = 1$ ) with  $\rho := \alpha m = 1$  and zero initial condition. On the other hand, the conditions (14)–(15) were introduced to address the problem of “fake transmissions” in the pre-limit system, which does not exist in the fluid model. Indeed, by (44)–(45), any increase in the transmission time process  $\bar{T}$  results in a proportional increase in  $\bar{D}$ , and hence a decrease of  $\bar{Z}$ .

In analogy with the processes  $Z, D, T, Y$ , we assume that  $\bar{D}(\cdot, s - \cdot)$  (and hence, by (45),  $\bar{T}(\cdot, s - \cdot)$ ) is nondecreasing in each coordinate. This, together with (46)–(47), implies that  $\bar{Y}(\cdot, s - \cdot)$  is nondecreasing in each coordinate. We also assume that  $\bar{D}(0, s) = \bar{T}(0, s) = 0$  and  $\bar{Y}(0, s) = 0$  for  $s \in \mathbb{R}$ . Similarly, we assume that all the coordinates of  $\bar{Z}(t, \cdot)$  and of the increments  $\bar{D}(\tilde{t}, \cdot - \tilde{t}) - \bar{D}(t, \cdot - t)$  are nondecreasing for all  $\tilde{t} \geq t \geq 0$ . Consequently, by (45)–(46), all the coordinates of  $\bar{T}(\tilde{t}, \cdot - \tilde{t}) - \bar{T}(t, \cdot - t)$  and  $\bar{Y}(\tilde{t}, \cdot - \tilde{t}) - \bar{Y}(t, \cdot - t)$  are nondecreasing for all  $\tilde{t} \geq t \geq 0$ .

A system

$$\bar{\mathfrak{X}}(t, s) = (\bar{Z}(t, s), \bar{D}(t, s), \bar{T}(t, s), \bar{Y}(t, s)), \quad t \geq 0, \quad s \in \mathbb{R}, \quad (48)$$

with continuous, nonnegative components, satisfying the Eqs. (44)–(47), together with the above assumptions, will be called a *fluid model for the resource sharing network* under consideration (or just a *resource sharing fluid model*). The terms  $\alpha, m_i$  and the network topology, described by the matrix  $A$ , are the fluid model data, given in advance.

It is instructive to compare our Eqs. (44)–(47) with the FIFSO (First-In-System, First-Out, the limiting case of EDF under suitable scaling) fluid model equations for “conventional” multiclass, multiserver queueing networks, introduced in (3.7)–(3.11) of Bramson (2001). (In Bramson (2001), the second time coordinate of a performance process denotes the arrival time of a job/customer mass to the system rather than its lead time. After a conversion to the coordinates used in this paper, the fluid model equations from Bramson (2001) take a slightly different form, compare the Eqs. (2.8)–(2.11), (2.15) in Kruk (2010).) Our Eqs. (44)–(47) may be regarded as counterparts of the Eqs. (3.7)–(3.10) from Bramson (2001) in the setting of resource sharing networks. All of them express basic properties of fluid models for general systems and they are by no means specific to the EDF service discipline. To capture the EDF property in a fluid model of a multiclass queueing network, Bramson (2001) introduced an additional equation (3.11). Its counterpart, in our setting, would read

$$\int_0^t \sum_{i \in \mathcal{F}(j)} \bar{Z}_i(u, s - u) d_u \bar{Y}_j(u, s - u) = 0, \quad t \geq 0, \quad s \in \mathbb{R}, \quad j \in \mathbf{J}, \quad (49)$$

with the corresponding pre-limit version

$$\int_0^t \sum_{i \in \mathcal{F}(j)} Z_i(u, s-u) d_u \tilde{Y}_j(u, s-u) = 0, \quad t \geq 0, \quad s \in \mathbb{R}, \quad j \in \mathbf{J}, \quad (50)$$

where

$$\tilde{Y}_j(t, s) = t - \sum_{i \in \mathcal{F}(j)} T_i(t, s) = \sum_{i \in \mathcal{F}(j)} Y_i(t, s) - (|\mathcal{F}(j)| - 1)t \quad (51)$$

is the cumulative idleness of the resource  $j$  on the time interval  $[0, t]$  related to flows with lead times at time  $t$  not greater than  $s$  and

$$\bar{Y}_j(t, s) = t - \sum_{i \in \mathcal{F}(j)} \bar{T}_i(t, s) = \sum_{i \in \mathcal{F}(j)} \bar{Y}_i(t, s) - (|\mathcal{F}(j)| - 1)t,$$

is its fluid counterpart. However, the equation (50) is, in general, *false* for EDF resource sharing networks. Consider, for example, the performance process  $\mathfrak{X}$  from Example 4, which is consistent with the EDF protocol, but

$$\int_0^1 \sum_{i \in \mathcal{F}(2)} Z_i(u, 3-u) d_u \tilde{Y}_2(u, 3-u) = Y_2(1, 2) + Y_3(1, 2) - 1 = 1.$$

Similarly, it may be shown that the equation (49) is, in general, false for fluid models of EDF resource sharing networks. This is not surprising, since the Eqs. (50), (49) imply the non-idling (i.e., work conserving) property of the underlying system and its fluid model, respectively, which hold for multiclass queueing networks, but they are well known to be false for resource sharing networks, see, e.g., Kruck (2016), Example 2. In fact, it is not clear to us how to find suitable counterparts of the integral Eqs. (50), (49), characterizing the EDF service protocol in preemptive networks with resource sharing and their fluid models. Below we propose to overcome this difficulty by formulating the EDF property of fluid models in terms of minimality and local edge minimality, rather than additional fluid model equations analogous to (49).

Motivated by the developments of Sects. 3 and 5, we introduce notions of minimality and local edge minimality for a resource sharing fluid model, which are even more natural in this context, in which the random processes  $E, S$ , related to the stochastic primitives, have been replaced by simpler deterministic ones.

**Definition 9** Let  $\bar{\mathfrak{X}}^{(k)} = (\bar{Z}^{(k)}, \bar{D}^{(k)}, \bar{T}^{(k)}, \bar{Y}^{(k)})$ ,  $k = 1, 2$ , be two fluid models for resource sharing networks, with the same fluid model data  $\alpha, m_i, A$ , and the same initial distribution  $\bar{Z}^{(1)}(0, \cdot) = \bar{Z}^{(2)}(0, \cdot)$ . We write  $\bar{\mathfrak{X}}^{(1)} \preceq \bar{\mathfrak{X}}^{(2)}$  if  $\bar{Y}^{(1)} \leq \bar{Y}^{(2)}$  (equivalently,  $\bar{T}^{(1)} \geq \bar{T}^{(2)}, \bar{D}^{(1)} \geq \bar{D}^{(2)}$  or  $\bar{Z}^{(1)} \leq \bar{Z}^{(2)}$ ).

The relation “ $\preceq$ ” is a partial ordering in the class of fluid models for a given network.

**Definition 10** A fluid model  $\bar{\mathfrak{X}}$  of the form (48) for a resource sharing network is called minimal if for any fluid model  $\bar{\mathfrak{X}}'$  such that  $\bar{\mathfrak{X}}' \preceq \bar{\mathfrak{X}}$ , we have  $\bar{\mathfrak{X}}' = \bar{\mathfrak{X}}$ .

Intuitively, minimal fluid models correspond to sample paths of fluid limits for resource sharing networks in which flows on each route are scheduled according to a general EDF protocol, see Theorems 1 and 2.

**Definition 11** A fluid model  $\bar{\mathfrak{X}}$  of the form (48) for a resource sharing network with  $\sum_{i \in \mathbf{I}} \bar{Z}_i(0, \cdot) \in B_{c,0}$  for some  $c \in \mathbb{R}$  is called locally edge minimal at a time  $t_0 \geq 0$  if there exists  $h > 0$  such that for any fluid model  $\bar{\mathfrak{X}}'$ , with the same data  $\alpha, m_i, A$  and the same initial state  $\bar{Z}'(0, \cdot) = \bar{Z}(0, \cdot)$ , satisfying

$$\bar{\mathfrak{X}}'(t_0, \cdot) = \bar{\mathfrak{X}}(t_0, \cdot), \quad (52)$$

we have  $\sum_{i \in \mathbf{I}} \bar{Y}_i(t, \cdot) \ll \sum_{i \in \mathbf{I}} \bar{Y}'_i(t, \cdot)$  (equivalently,  $\sum_{i \in \mathbf{I}} \bar{T}'_i(t, \cdot) \ll \sum_{i \in \mathbf{I}} \bar{T}_i(t, \cdot)$ ,  $\sum_{i \in \mathbf{I}} \bar{D}'_i(t, \cdot) \ll \sum_{i \in \mathbf{I}} \bar{D}_i(t, \cdot)$  or  $\sum_{i \in \mathbf{I}} \bar{Z}_i(t, \cdot) \ll \sum_{i \in \mathbf{I}} \bar{Z}'_i(t, \cdot)$ ) for every  $t \in (t_0, t_0 + h)$ .

The fluid model  $\bar{\mathfrak{X}}$  is called locally edge minimal, if it is locally edge minimal at every  $t_0 \geq 0$ .

*Remark 10* By the fluid model Eqs. (44)–(47), the condition (52) is equivalent to each of the following four conditions:  $\bar{Z}'(t_0, \cdot) = \bar{Z}(t_0, \cdot)$ ,  $\bar{D}'(t_0, \cdot) = \bar{D}(t_0, \cdot)$ ,  $\bar{T}'(t_0, \cdot) = \bar{T}(t_0, \cdot)$ ,  $\bar{Y}'(t_0, \cdot) = \bar{Y}(t_0, \cdot)$ .

It is not hard to verify that a locally edge minimal fluid model is minimal. The following definition is motivated by Theorems 5–7.

**Definition 12** A locally edge minimal resource sharing fluid model is called an EDF fluid model for a resource sharing network (or just an EDF resource sharing fluid model).

It is natural to expect that under fluid scaling and suitable assumptions on the stochastic primitives, the performance processes of EDF resource sharing networks converge to a fluid limit with sample paths being (a.s.) locally edge minimal resource sharing fluid model solutions. If this is indeed the case, then asymptotic properties of EDF networks with resource sharing, like stability or convergence to the invariant manifold for large times in the critical case, may be inferred from the asymptotic behavior of the corresponding EDF resource sharing fluid models. We hope to address these issues in a forthcoming paper.

## 7 Concluding remarks

The notions of edge minimality and local edge minimality, introduced in Sections 4–5, are based on comparing idleness in transmissions of flows on distinct routes for different service disciplines. It also makes sense to make similar comparisons on the level of the system resources, using the processes  $\tilde{Y}_j$  defined by (51) instead of  $Y_i$ . This leads to somewhat different, but strongly related, concepts and results (compare Kruk (2016), Section 5). We omit the details.

**Acknowledgements** The author thanks the associate editor and the referees for their helpful suggestions for improving the paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Baruah SK (2006) Resource sharing in EDF-scheduled systems: a closer look. In: Proceedings of the 27th IEEE international real-time systems symposium (RTSS'06). IEEE Computer Society, Los Alamos, CA
- Bramson M (2001) Stability of earliest-due-date, first-served queueing networks. Queueing Syst Theory Appl 39:79–102
- Chiang M, Shah D, Tang A (2006) Stochastic stability under fair bandwidth allocation: general file size distribution. In: Proceedings of the 44th Allerton Conference, pp. 899–908
- Doytchinov B, Lehoczky JP, Shreve SE (2001) Real-time queues in heavy traffic with earliest-deadline-first queue discipline. Ann Appl Probab 11:332–379
- Gromoll HC, Kruk Ł (2007) Heavy traffic limit for a processor sharing queue with soft deadlines. Ann Appl Probab 17:1049–1101
- Gromoll HC, Robert P, Zwart B (2008) Fluid limits for processor sharing queues with impatience. Math Oper Res 33:375–402
- Gromoll HC, Williams RJ (2009) Fluid limits for networks with bandwidth sharing and general document size distribution. Ann Appl Probab 19:243–280
- Harrison JM, Mandayam C, Shah D, Yang Y (2014) Resource sharing networks: overview and an open problem. Stoch Syst 4:524–555
- Hopp W, Spearman M (1996) Factory physics: foundations of manufacturing management. Irwin, Chicago
- Kang WN, Kelly FP, Lee NH, Williams RJ (2009) State space collapse and diffusion approximation for a network operating under a fair bandwidth sharing policy. Ann Appl Probab 19:1719–1780
- Kelly FP (1997) Charging and rate control for elastic traffic. Eur Trans Telecom 8:33–37
- Key P, Massoulié L (2006) Fluid models of integrated traffic and multipath routing. Queueing Syst 53:85–98
- Kruk Ł (2010) Invariant states for fluid models of EDF networks: nonlinear lifting map. Probab Math Stat 30:289–315
- Kruk Ł (2016) Minimality of EDF networks with resource sharing. Math Meth Oper Res 84:259–283
- Kruk Ł, Lehoczky JP, Ramanan K, Shreve SE (2011) Heavy traffic analysis for EDF queues with reneging. Ann Appl Probab 21:484–545
- Liu CL, Layland JW (1973) Scheduling algorithms for multiprogramming in a hard real-time environment. J Assoc Comput Mach 20(1):40–61
- Massoulié L, Roberts J (1999) Bandwidth sharing: objectives and algorithms. In: Proceedings of the INFO-COM 2009, eighteenth annual joint conference of the IEEE computer and communications societies, vol 3, pp 1395–1403, IEEE
- Massoulié L, Roberts J (2000) Bandwidth sharing and admission control for elastic traffic. Telecommun Syst 15:185–201
- Mo J, Walrand J (2000) Fair end-to-end window-based congestion control. IEEE/ACM Trans Netw 8:556–567
- Moyal P (2008) Convex comparison of service disciplines in real time queues. Oper Res Lett 36(4):496–499
- Panwar SS, Towsley D (1988) On the optimality of the STE rule for multiple server queues that serve customers with deadlines. Technical Report 88-81, Department of Computer and Information Science, University Massachusetts, Amherst
- Panwar SS, Towsley D (1992) Optimality of the stochastic earliest deadline policy for the G/M/c queue serving customers with deadlines. In: Second ORSA telecommunications conference. ORSA (Operations Research Society of America), Baltimore, MD
- Prohorov YuV (1956) Convergence of random processes and limit theorems in probability theory. Theory Probab Appl 1:157–214
- Schrage LE (1968) A proof of the optimality of the shortest remaining processing time discipline. Oper Res 16:687–690

- Sivaraman V, Chiussi FM, Gerla M (2001) End-to-end statistical delay service under GPS and EDF scheduling: a comparison study. In: Proceedings of the INFOCOM 2001, twentieth annual joint conference of the IEEE computer and communications societies, vol 2, pp 1113–1122, IEEE
- Stankovic JA, Spuri M, Ramamritham K, Buttazzo GC (2012) Deadline scheduling for real-time systems: EDF and related algorithms. Springer, Berlin
- Verloop M, Borst S, Núñez-Queija R (2005) Stability of size-based scheduling disciplines in resource-sharing networks. Perform Eval 62:247–262
- Vlasiou M, Zhang J, Zwart B (2015) Insensitivity of proportional fairness in critically loaded bandwidth sharing networks. arXiv:1411.4841
- Ye H-Q (2003) Stability of data networks under an optimization-based bandwidth allocation. IEEE Trans Autom Control 48:1238–1242
- Ye H-Q, Ou J, Yuan X-M (2005) Stability of data networks: Stationary and bursty models. Oper Res 53:107–125