PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 30, Fasc. 2 (2010), pp. 289–315

INVARIANT STATES FOR FLUID MODELS OF EDF NETWORKS: NONLINEAR LIFTING MAP

BY

ŁUKASZ KRUK (LUBLIN)

Abstract. We study fluid models of an open, subcritical multiclass queueing network with the earliest-deadline-first (EDF) service discipline and we provide a characterization of the corresponding invariant manifold. We show that the invariant states exhibit nonlinear state space collapse. Consequences of these findings for diffusion limits for EDF queueing networks are also discussed.

2000 AMS Mathematics Subject Classification: Primary: 60K25, 90B10; Secondary: 68M20.

Key words and phrases: Multiclass queueing networks, fluid models, deadlines, invariant states, state space collapse, lifting map.

1. INTRODUCTION

The earliest-deadline-first (EDF) discipline, also called earliest-due-date-first-served (EDDFS), is the rule where each customer has a deadline, assigned upon arrival at the network, and the customer with the earliest deadline is selected for service at each station of the network. It is a well-studied protocol, especially in computer and manufacturing systems; see, e.g., [12], [19]. In spite of theoretical and practical importance of stochastic EDF queueing networks, there are still few mathematically rigorous results for such systems. In [4], it was shown that fluid limits of the performance processes for a non-preemptive strictly subcritical EDF network satisfy the first-in-system-first-out (FISFO) fluid model equations. It was also proved in this paper that a sufficiently rich class of FISFO fluid models is stable. This, by a variation of Theorem 4.2 of [6], implies stability of the network under consideration. This stability result was extended in [13] to preemptive strictly subcritical EDF networks with fixed customer routes and to a broad class of networks (including preemptive and non-preemptive EDF) with impatient customers. In [7], a diffusion approximation for measure-valued state descriptors of a preemptive EDF GI/G/1 queue was provided. Recently, a similar diffusion approximation for a preemptive EDF GI/G/1 queue with reneging was found [15]. The results of [7] have been generalized in [22] to preemptive EDF feedforward networks. A further generalization to the case of acyclic networks, with or without preemption, was given in [17]. However, the latter result rests on a strong assumption implying the existence of a heavy traffic limit for the corresponding real-valued workload process. Currently, we are able to verify this assumption only in a number of special cases. To summarize, although a lot is known for a single class, single server EDF system, it seems that satisfactory asymptotic theory for multiclass EDF queueing networks with feedback is still to be developed.

In the late 90's, a new modular approach to diffusion approximations for open multiclass queueing systems was developed [1]–[3], [20], [21]. This approach may be summarized informally as follows. Consider a multiclass queueing network with servers indexed by $j = 1, \dots, J$ and customer classes indexed by $k = 1, \dots, K$. First, we define *fluid models* of the queueing network under consideration and identify their *invariant states*. It is crucial for the subsequent analysis that there exists a *linear lifting map* $\Delta\colon \mathbb{R}_+^J \to \mathbb{R}_+^K$ mapping the workload vector $W(\infty) = \big(W_j(\infty)\big)_{j=1,\dots,J}$ to the corresponding queue length vector $Z(\infty) = \big(Z_k(\infty)\big)_{k=1,\dots,K}$ in an invariant state of the fluid model, i.e., in any such state,

$$Z(\infty) = \Delta W(\infty). \tag{1.1}$$

In fact, the structure of Δ in known examples is rather simple, since for every k there exists a number δ_k such that $Z_k(\infty) = \delta_k W_j(\infty)$, where station j serves class k customers. The next step is to show that fluid models of the network under consideration are *asymptotically stable*, i.e., they converge to invariant states: $\big(W(t), Z(t)\big) \to \big(W(\infty), Z(\infty)\big)$, $t \to \infty$, and, consequently, in such fluid models

$$Z(t) \approx \Delta W(t), \quad t \gg 1. \tag{1.2}$$

Next, we analyze a sequence of queueing networks, indexed by a parameter r, which is *asymptotically critical*, i.e., the traffic intensity ρ_j^r at station j in the r-th system converges to 1 for each j. The consequence of (1.2) is that for large r,

$$\overline{Z}^r(t) \approx \Delta \overline{W}^r(t), \quad t \gg 1, \tag{1.3}$$

where $\overline{Z}^r$ ($\overline{W}^r$) is the fluid-scaled queue length (workload) vector in the r-th system. This follows from the fact that fluid models (Z, W) approximate the paths $(\overline{Z}^r, \overline{W}^r)$ as r gets large. The relation (1.3) is used to show that under appropriate initial conditions and model parameters a *multiplicative state space collapse* holds:

$$\|\widehat{Z}^r(\cdot) - \Delta \widehat{W}^r(\cdot)\|_T / \|\widehat{W}^r(\cdot) \vee 1\|_T \xrightarrow{P} 0 \tag{1.4}$$

for every $T > 0$, where $\widehat{Z}^r$ ($\widehat{W}^r$) is the diffusion-scaled queue length (workload) vector in the r-th system, $\|\cdot\|_T$ is the supremum norm on $[0, T]$ and $\xrightarrow{P}$ denotes convergence in probability. The condition (1.4), together with an invariance principle for semimartingale reflected Brownian motions [20], is used to prove that

$$(\widehat{W}^r, \widehat{Z}^r) \Rightarrow (\tilde{W}, \tilde{Z}), \tag{1.5}$$

where $\Rightarrow$ denotes weak convergence, $\tilde{W}$ is a Brownian motion reflected obliquely in an orthant, with appropriate drift θ, diffusion matrix Γ and reflection matrix R, and $\tilde{Z} = \Delta \tilde{W}$. The relation (1.5) will be referred to as a *conventional heavy traffic limit theorem* [10]. We would like to stress that the vector θ and the matrices Γ and R determining the reflected Brownian motion W depend on the matrix Δ.

This approach was used in [3] and [21] to obtain new heavy traffic limit theorems for FIFO (first-in first-out) networks of Kelly type and for HLPPS (head-of-the-line proportional processor sharing) networks. See also [5] for other applications of this technique. Let us note that the approach described above was originally developed for head-of-the-line (HL) service disciplines. Recal that an HL service discipline requires that service within each class is on the FIFO basis and the proportions of each server's time devoted to various customer classes are constant between changes of the arrival and departure processes. (Some measurability and nonanticipativity conditions are also required, see [21] for details.) However, a modification of this technique was later used to obtain a heavy traffic limit for a GI/G/1 processor sharing queue, which is clearly not HL [9], [18], [8].

Bramson ([4], p. 81) and Williams (private communication) posed a question whether the above modular approach can be applied, at least in some situations, to EDF queueing networks. This paper is the first step in answering this question. We introduce fluid models of EDF queueing networks and we characterize the associated invariant manifold, which turns out to have a much more complicated structure than its known counterparts, e.g., for FIFO or HLPPS networks. In particular, in the EDF case, the lifting map Δ in (1.1) is *nonlinear*. Thus, while it is plausible that an appropriate extension of the techniques from [1]–[3], [20], [21] to the EDF case can be made, we do not expect such an extension to be straightforward. In fact, in the light of the above-mentioned dependence of θ, Γ and R on Δ, it is reasonable to conjecture that the limiting heavy traffic workload distribution in the EDF case (if it exists) is, in general, a reflected diffusion with state-dependent coefficients and the reflection direction not necessarily constant on a given face of the nonnegative orthant. In other words, some EDF networks may exhibit *unconventional heavy traffic behavior*. In a subsequent paper [14] we verify this conjecture, providing an example of a simple feedforward FISFO (a special case of EDF) queueing network with asymptotically stable fluid model and unconventional heavy traffic diffusion approximation. The analysis in [14] depends heavily on the main results of this paper, in particular on nonlinearity of the mapping Δ. The example from [14] shows that the heavy traffic conjecture from [10], stating that a FIFO network has a conventional heavy traffic approximation if and only if its fluid model is asymptotically stable, cannot be generalized to other service disciplines, even in the case of feedforward networks with the HL property. Let us also mention that while an example of a *closed* network with somewhat related unconventional heavy traffic limit is known [11], we are not aware of an open network with protocol other than EDF, which gives rise to an unconventional heavy traffic limit.

This paper is organized as follows. Section 2 presents the basic notation and definitions. It also describes EDF queueing networks and their fluid models. Section 3 presents the main results of the paper and provides an example of nonlinearity of the mapping Δ. In Section 4 we introduce the notions of the frontier and the null set of an invariant EDF fluid model, the former one being an analog of a frontier for an EDF queueing system introduced in [7]. We also prove several properties of invariant states involving these notions. The main results of the paper are shown in Section 5.

2. EDF NETWORKS AND THEIR FLUID MODELS

2.1. Basic notation. The following notation will be used throughout the paper. Let $\mathbb{N} = \{1, 2, \ldots\}$, let $\mathbb{R}$ be the set of real numbers, and $\mathbb{R}_+ = [0, \infty)$. The Borel σ-field on $\mathbb{R}$ will be denoted by $\mathcal{B}(\mathbb{R})$. For $n \in \mathbb{N}$, put $\mathbb{R}_+^n$ for $(\mathbb{R}_+)^n$, the positive orthant in $\mathbb{R}^n$. For $a, b \in \mathbb{R}$, we write $a \vee b$ for the maximum of a and b, $a \wedge b$ for the minimum of a and b, and a^+ for $a \vee 0$. For $a, b \in \mathbb{R}^n$, $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$, the vector $(a_1 b_1, \ldots, a_n b_n)$ will be denoted by $a \circ b$. All vectors in the paper are to be interpreted as column vectors. For a finite set B, $|B|$ denotes the cardinality of B. Denote by $\mathcal{M}$ the set of all finite, nonnegative measures on $\mathcal{B}(\mathbb{R})$ with the Lévy–Prokhorov metric

$$\rho(\mu, \nu) = \inf\{\epsilon > 0 : \mu(B) \leqslant \nu(B^\epsilon) + \epsilon, \nu(B) \leqslant \mu(B^\epsilon) + \epsilon \text{ for all } B \in \mathcal{B}(\mathbb{R})\}$$

for $\mu, \nu \in \mathcal{M}$, where

$$B^\epsilon = \{x \in \mathbb{R} : \inf_{y \in B} |x - y| \leqslant \epsilon\}.$$

It is known that ρ is a complete metric on $\mathcal{M}$ inducing the weak topology. Finally, let $\rho_n = \rho \times \ldots \times \rho$ be the product metric on the n-fold Cartesian product $\mathcal{M}^n$.

2.2. EDF networks. This paper contains a characterization of the invariant manifold for a family of fluid models corresponding to queueing networks with EDF service discipline. To motivate the introduction of these fluid models, we first provide a brief description of the corresponding queueing networks.

We consider a network consisting of J single server stations, indexed by $j = 1, \ldots, J$. The network is populated by K customer classes, indexed by $k = 1, \ldots, K$. There is a stationary external arrival process with rate α_k associated with each class k. In particular, if $\alpha_k = 0$, there are no external arrivals to class k. We put $\alpha = (\alpha_1, \ldots, \alpha_K)$. A customer of class k receives service at a unique station j, written $k \in \mathcal{C}(j)$ or $j = s(k)$. Let m_k be the mean service time for the class k and let $m = (m_1, \ldots, m_K)$. Upon being served at station j, a customer of class k immediately becomes a customer of class l with probability p_{kl}, independently of its past history. Thus, the probability that a customer of class k leaves the network after completion of service equals $1 - \sum_{l=1}^{K} p_{kl}$. The routing matrix $P = (p_{kl})$ is assumed to be transient, i.e., such that the matrix

$$Q \triangleq (I - P')^{-1} = I + P' + (P')^2 + \ldots \tag{2.1}$$

exists, where $'$ denotes the transpose. We define the *total arrival rate* vector

$$\lambda = (\lambda_1, \ldots, \lambda_K) = Q\alpha. \tag{2.2}$$

Without loss of generality we assume that $\lambda_k > 0$ for each k. Next, we define the *traffic intensity* at station j as

$$\rho_j = \sum_{k \in \mathcal{C}(j)} m_k \lambda_k. \tag{2.3}$$

When $\rho_j \leqslant 1$ ($\rho_j < 1$, $\rho_j = 1$) for each j, the network is called *subcritical* (*strictly subcritical*, *critical*). Class k customers entering the network have *initial lead times* with cumulative distribution function (c.d.f.) G_k. For notational convenience, we define G_k for every $k = 1, \ldots, K$, including classes with no external arrival streams. For k such that $\alpha_k = 0$, G_k may be chosen in an arbitrary way and this choice does not affect any further considerations. We put $G = (G_1, \ldots, G_K)$. To simplify the presentation, we assume that

$$y_k^* \triangleq \sup\{y \in \mathbb{R} : G_k(y) < 1\} < \infty \tag{2.4}$$

for $k = 1, \ldots, K$, but all the results of this paper and their proofs remain valid, after appropriate modifications, in a more general case in which for $k = 1, \ldots, K$,

$$\int_0^\infty \big(1 - G_k(y)\big)\, dy < \infty. \tag{2.5}$$

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

lead time = initial lead time − time elapsed since arrival

for customers coming to the system after time zero and

lead time = initial lead time − current time

for *initial customers*, i.e., those who are present in the network at time zero.

Customers are served at each station according to the EDF discipline. That is, the customer with the shortest remaining lead time, regardless of class, is selected for service at each station. Late customers (customers with negative lead times) stay in the system until served to completion. Two types of EDF network protocols may be considered. In the *preemptive* case preemption occurs when a customer more urgent than the customer in service arrives (we assume preempt-resume). In EDF networks *without preemption* customer service continues until he is served to completion, even if a more urgent customer enters the station.

Probability and Mathematical Statistics 30/2010, z. 2

2.3. EDF fluid models. *Fluid models* are deterministic, continuous analogs of queueing networks, in which individual customers are replaced by a divisible commodity (fluid) of K types or classes, indexed by $k = 1, \ldots, K$, which change as the fluid moves between stations $j = 1, \ldots, J$ until it leaves the system. In analogy with customers of queueing networks described above, class k fluid arrives exogeneously to a unique station $j = s(k)$ with rate α_k and initial lead time distribution G_k, it is processed at $s(k)$ with mean service time m_k and changes class to l with transition probability p_{kl} after service completion. As in the case of queueing networks, we say that a fluid model is *subcritical* (*strictly subcritical*, *critical*) if $\rho_j \leqslant 1$ ($\rho_j < 1$, $\rho_j = 1$) for each j, where ρ_j are given by (2.3). Fluid models are defined rigorously in terms of the appropriate *fluid model equations*.

Fluid models for EDF queueing networks consist of the six-tuples of vectors

$$\mathfrak{X}(t,s) = \big(Z(t,s), W(t,s), A(t,s), D(t,s), T(t,s), Y(t,s)\big), \tag{2.6}$$

where $t \geqslant 0$, $s \in \mathbb{R}$, the vectors $Z(t,s), W(t,s), A(t,s), D(t,s), T(t,s)$ are indexed by $k = 1, \ldots, K$ and the vector $Y(t,s)$ is indexed by $j = 1, \ldots, J$. Here $Z_k(t,s)$ denotes the amount of class k fluid with lead times less than or equal to s at time t and $W_k(t,s)$ represents the workload for station $s(k)$ associated with this fluid, i.e., the amount of time necessary for the server $s(k)$ to process it to completion (provided that the station devotes all its capacity to it, without processing any other fluids at the same time). The quantity $A_k(t,s)$ ($D_k(t,s)$) denotes the amount of fluid with lead times at time t less than or equal to s which has arrived at (departed from) class k by time t and $T_k(t,s)$ represents the amount of work devoted to this fluid by server $s(k)$ by time t. Finally, $Y_j(t,s)$ denotes the cumulative idleness by time t at station j with regard to service of fluids with lead times at time t less than or equal to s. The vectors defining $\mathfrak{X}$ are the continuous analogs of the corresponding quantities in the EDF queueing network described in Section 2.2. We assume that all the components of $\mathfrak{X}$ are continuous and nonnegative, with $A(\cdot, s-\cdot)$, $D(\cdot, s-\cdot)$, $T(\cdot, s-\cdot)$, $Y(\cdot, s-\cdot)$ nondecreasing in each coordinate, $A(0,s) = D(0,s) = T(0,s) = 0$ and $Y(0,s) = 0$ for $s \in \mathbb{R}$. We also assume that every coordinate of $A(t,\cdot)$, $D(t,\cdot)$, $T(t,\cdot)$, $-Y(t,\cdot)$, $Z(t,\cdot)$ and $W(t,\cdot)$ is nondecreasing for $t \geqslant 0$. The *EDF fluid model equations*, defining the model, are:

$$A(t,s) = \alpha \circ \textstyle\int_0^t G(s+\eta)\,d\eta + P'D(t,s), \tag{2.7}$$

$$Z(t,s) = Z(0,t+s) + A(t,s) - D(t,s), \tag{2.8}$$

$$T(t,s) = m \circ D(t,s), \tag{2.9}$$

$$\textstyle\sum_{k\in\mathcal{C}(j)} T_k(t,s) + Y_j(t,s) = t, \tag{2.10}$$

$$Y_j(t,s-t) \text{ can only increase in } t \text{ if } \textstyle\sum_{k\in\mathcal{C}(j)} Z_k(t,s-t) = 0, \tag{2.11}$$

$$W(t,s) = m \circ Z(t,s), \tag{2.12}$$

where $t \geqslant 0$, $s \in \mathbb{R}$, $j = 1, \ldots, J$. A system (2.6) satisfying the equations (2.7)–(2.12) will be called an *EDF fluid model*. The terms α, m, P and G are the model data, given in advance. By (2.7), the external arrival process to class k has the form

$$\alpha_k \int_0^t G_k(s+\eta)\, d\eta. \tag{2.13}$$

This can be explained as follows. A customer who has entered the network through the buffer k at time $\xi \in (0,t]$ and has lead time not greater than s at time t had lead time upon arrival not greater than $s+t-\xi$. The rate with which such customers enter the system is $\alpha_k G_k(s+t-\xi)$, so the total number of such arrivals in the interval $[0,t]$ is $\alpha_k \int_0^t G_k(s+t-\eta)\, d\eta$, which is the same as (2.13). The equations (2.7), (2.8) and (2.10) hold regardless of the service protocol under consideration. The equation (2.11) characterizes the EDF service discipline. The equations (2.9) and (2.12) imply that the number of partially served customers in the pre-limit EDF system is negligible under fluid scaling. This is obvious in the non-preemptive case, but it requires a proof for preemptive EDF networks; see Corollary 3.8 in [7], Corollary 4.8 in [22], Corollary 4.7 in [17] and Proposition 6.1 in [13] for the corresponding arguments in some special cases.

In what follows, we shall make the following assumption on the initial condition, which is compatible with the definition of y_k^*.

ASSUMPTION 2.1. We have

$$Z(0, \max_{k:\alpha_k>0} y_k^*) = \lim_{s\to\infty} Z(0,s). \tag{2.14}$$

In particular, under the assumptions (2.4) and (2.14), the support of the measure with the distribution function $Z_k(0,\cdot)$ is bounded above for each k. Of course, if (2.4) does not hold for some k with $\alpha_k > 0$, (2.14) can be regarded as a tautology.

An important special case of the EDF fluid model equations may be obtained by putting $G_k(y) = \mathbb{I}_{[0,\infty)}(y)$ for each k, so that (2.7) simplifies to

$$A(t,s) = \alpha\big(t + (s\wedge 0)\big)^+ + P'D(t,s). \tag{2.15}$$

The equations (2.8)–(2.12), (2.15) will be referred to as the *FISFO fluid model equations*. If we change the coordinates (t,s) to $(t,\tilde{s})$, where $\tilde{s} = s-t$, in (2.8)–(2.11), (2.15), we obtain the FISFO fluid model equations introduced in [4]:

$$\overline{A}(t,\tilde{s}) = \alpha(t\wedge\tilde{s}) + P'\overline{D}(t,\tilde{s}), \tag{2.16}$$

$$\overline{Z}(t,\tilde{s}) = \overline{Z}(0,\tilde{s}) + \overline{A}(t,\tilde{s}) - \overline{D}(t,\tilde{s}), \tag{2.17}$$

$$\overline{D}_k(t,\tilde{s}) = \overline{T}_k(t,\tilde{s})/m_k \text{ for } k = 1,\ldots,K, \tag{2.18}$$

$$\textstyle\sum_{k\in\mathcal{C}(j)} \overline{T}_k(t,\tilde{s}) + \overline{Y}_j(t,\tilde{s}) = t \text{ for } j = 1,\ldots,J, \tag{2.19}$$

$$\overline{Y}_j(t,\tilde{s}) \text{ can only increase in } t \text{ when } \textstyle\sum_{k\in\mathcal{C}(j)} \overline{Z}_k(t,\tilde{s}) = 0, \tag{2.20}$$

for $t, \tilde{s} \geqslant 0$. In (2.16)–(2.20), the coordinate $\tilde{s}$ represents the arrival times of customers (fluids) to the network, rather than their lead times. However, it seems that for the sake of the characterization of the corresponding invariant manifold, the equations (2.8)–(2.12), (2.15) (and, more generally, (2.7)–(2.12)) are more convenient than (2.16)–(2.20), see Remark 3.1 to follow.

If we take *fluid limits*, i.e., the limits of sample paths along subsequences under scaling which is linear in both time and space (called *fluid* or *hydrodynamic scaling*) obtained from a *single* EDF network, then the initial lead time distributions disappear in the limit, giving rise to the FISFO fluid models. This has been proved in [4] in the case of no preemption and in [13] for preemptive EDF networks with fixed customer routes. In this paper, we chose to consider more general EDF fluid models satisfying (2.7)–(2.12) with nontrivial lead time distributions G_k, which may be useful in the asymptotic analysis of a *sequence* of EDF networks where the initial lead time distributions dilate with the same rate as the space scaling parameter. Thanks to such lead time scaling, employed, e.g., in [7], [15]–[17], [22], the customer lead times are "realistic", i.e., they are of the same order as the queue lengths and the sojourn times, so a typical customer has a reasonable chance to have his job done on time. In contrast, fluid and diffusion scaling of a critical network without scaling the corresponding lead times leads to a model in which most (if not all) incoming customers are late in (or sufficiently close to) the limit, which is not a desirable feature for a real-time queueing model. Also, the characterization of the invariant manifold, which is the topic of this paper, is not substantially more difficult in the case of general initial lead time distributions G_k.

2.4. Paths and multi-indices. Let

$$\mathbf{K} = \big\{(k_1, \ldots, k_n) : n \in \mathbb{N}, k_1, \ldots, k_n \in \{1, \ldots, K\}, \alpha_{k_1} p_{k_1 k_2} \ldots p_{k_{n-1} k_n} > 0\big\},$$

where $p_{k_1 k_2} \ldots p_{k_{n-1} k_n}$ should be interpreted as 1 for $n = 1$. The elements of $\mathbf{K}$ will be called *multi-indices*. They represent paths of finite length which are being followed with positive probability by customers (fluids) since their arrival to the network. For $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbf{K}$, let $p_{\mathbf{k}} = p_{k_1 k_2} \ldots p_{k_{n-1} k_n}$, $\alpha_{\mathbf{k}} = \alpha_{k_1} p_{\mathbf{k}}$ and $\rho_{\mathbf{k}} = m_{k_n} \alpha_{\mathbf{k}}$. Also, for $\mathbf{k}$ as above, let $b(\mathbf{k}) = k_1$ and $e(\mathbf{k}) = k_n$ be the beginning and the end of the path $\mathbf{k}$, respectively, and let $\mathcal{S}(\mathbf{k}) = \{s(k_1), \ldots, s(k_{n-1})\}$, which should be interpreted as $\emptyset$ if $n = 1$. In other words, $\mathcal{S}(\mathbf{k})$ is the set of servers encountered by customers following the path $\mathbf{k}$ before they become class $e(\mathbf{k})$ customers. For $\mathbf{k} \in \mathbf{K}$, $k \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, J\}$, we write $\mathbf{k} \in \widetilde{\mathcal{C}}(k)$ if $e(\mathbf{k}) = k$ and $\mathbf{k} \in \overline{\mathcal{C}}(j)$ if $e(\mathbf{k}) \in \mathcal{C}(j)$. Let $\mathcal{K}_0(j) = \{(k) : k \in \mathcal{C}(j), \alpha_k > 0\}$ be the set of multi-indices of length one corresponding to customer classes that enter station $j \in \{1, \ldots, J\}$ from outside the system and let $\mathcal{J}_0 = \big\{j \in \{1, \ldots, J\} : \mathcal{K}_0(j) \neq \emptyset\big\}$ be the set of stations that serve as the entry point for at least one external arrival process. For $m = 1, \ldots, J-1$ and $j_1, \ldots, j_m, j \in \{1, \ldots, J\}$, let $\mathcal{K}_m^{j_1, \ldots, j_m}(j) = \big\{\mathbf{k} \in \mathbf{K} \colon \mathbf{k} \in \overline{\mathcal{C}}(j), \mathcal{S}(\mathbf{k}) \subseteq \{j_1, \ldots, j_m\}\big\}$ be the set of paths of customers eventually visiting station j who visit only stations in the set $\{j_1, \ldots, j_m\}$

before arriving at j. We say that it is the set of paths which reach station j through $\{j_1, \ldots, j_m\}$. Finally, for $m = 1, \ldots, J-1$ and $j_1, \ldots, j_m \in \{1, \ldots, J\}$, let

$$\mathcal{J}_m^{j_1,\ldots,j_m} = \{j \in \{1, \ldots, J\} : \mathcal{K}_m^{j_1,\ldots,j_m}(j) \neq \emptyset\} \setminus \{j_1, \ldots, j_m\}$$

be the set of stations not in the set $\{j_1, \ldots, j_m\}$ that can be reached through $\{j_1, \ldots, j_m\}$.

DEFINITION 2.1. We say that a network is *connected* if for every two stations j and j', $j \neq j'$, there exist $n \in \mathbb{N}$, $j_0, \ldots, j_n \in \{1, \ldots, J\}$ such that $j_0 = j$, $j_n = j'$, and for each $i = 1, \ldots, n$, there exist $k_1^i, \ldots, k_{m_i}^i \in \{1, \ldots, K\}$ such that $p_{k_1^i k_2^i} \ldots p_{k_{m_i-1}^i k_{m_i}^i} > 0$ and either $k_1^i \in \mathcal{C}(j_{i-1})$, $k_{m_i}^i \in \mathcal{C}(j_i)$, or $k_1^i \in \mathcal{C}(j_i)$, $k_{m_i}^i \in \mathcal{C}(j_{i-1})$.

In other words, a network is *conected* if for any two stations there is a way of reaching one station from the other by following fragments of paths indexed by elements of $\mathbf{K}$, not necessarily in the forward direction. In what follows, we fix an EDF queueing network under consideration (which, for the sake of construction of the corresponding fluid models, is completely determined by α, m, P, G and the sets $\mathcal{C}(j)$, $j = 1, \ldots, J$) and assume that it is connected.

2.5. Invariant states.

DEFINITION 2.2. An EDF fluid model of the form (2.6) is called *invariant* if for all $t \geqslant 0$ and $s \in \mathbb{R}$ we have

$$Z(t, s) = Z(0, s).$$

In other words, an EDF fluid model is *invariant* if the customer instantaneous lead time profiles $Z(t, \cdot)$ (and thus, by (2.12), $W(t, \cdot)$) do not change with time. It is easy to see that invariant EDF fluid models with $\rho_j > 1$ for some j do not exist. Thus, in the remainder of the paper we assume that the network under consideration is subcritical. We shall almost entirely focus on *invariant* EDF fluid models. The coordinate t in $Z(t, s)$ and $W(t, s)$ of such models will usually be skipped.

DEFINITION 2.3. A vector $L = (L_k)_{k=1,\ldots,K}$ of cumulative distribution functions (c.d.f.'s) of finite nonnegative measures on $\mathbb{R}$ is called an *invariant state* of an EDF fluid model if there exists an invariant EDF fluid model of the form (2.6) such that for every $s \in \mathbb{R}$,

$$L(s) = Z(s).$$

The set of all invariant states of EDF fluid models will be called the *invariant manifold* and denoted by $\mathfrak{S}$.

Our first observation is that an invariant state completely determines the corresponding invariant EDF fluid model. This implies that we can think of $\mathfrak{S}$ as of the family of invariant fluid models.

PROPOSITION 2.1. *Let $\mathfrak{X}$ be an invariant EDF fluid model of the form* (2.6). *Then for $t \geqslant 0$, $s \in \mathbb{R}$ and $j = 1, \ldots, J$,*

$$A(t,s) = Q\Big(\alpha \circ \int_0^t G(s+\eta)\, d\eta + P'\big(Z(t+s) - Z(s)\big)\Big), \tag{2.21}$$

$$D(t,s) = Q\big(\alpha \circ \int_0^t G(s+\eta)\, d\eta + Z(t+s) - Z(s)\big), \tag{2.22}$$

$$Y_j(t,s) = t - \sum_{k\in\mathcal{C}(j)} m_k \big\{Q\big(\alpha \circ \int_0^t G(s+\eta) d\eta + Z(t+s) - Z(s)\big)\big\}_k. \tag{2.23}$$

In particular, the invariant state $Z(\cdot)$ uniquely determines $\mathfrak{X}$.

P r o o f. Fix $t \geqslant 0$, $s \in \mathbb{R}$. By (2.8) and the invariance of $\mathfrak{X}$, we have

$$D(t,s) = Z(t+s) - Z(s) + A(t,s). \tag{2.24}$$

Substituting (2.24) into (2.7), we get

$$(I - P')A(t,s) = \alpha \circ \int_0^t G(s+\eta)\, d\eta + P'\big(Z(t+s) - Z(s)\big). \tag{2.25}$$

Multiplying (2.25) from the left by Q and using (2.1), we obtain (2.21). Putting (2.21) into (2.24) and applying the fact that, by (2.1), $I + QP' = Q$, we get (2.22). Substituting (2.22) into (2.9) and using (2.10), we have (2.23). The last claim follows immediately from (2.9), (2.12) and (2.21)–(2.23). ∎

3. MAIN RESULTS

Let $J_1 = |\{j : \rho_j = 1\}|$ and let $J_2 = J - J_1$. In particular, the network is critical (strictly subcritical) if $J_1 = J$ ($J_1 = 0$). Without loss of generality we can assume that $\rho_1 = \ldots = \rho_{J_1} = 1$, $\rho_{J_1+1} < 1, \ldots, \rho_J < 1$. For $\mathbf{k} \in \mathbf{K}$, let $\mathcal{S}_c(\mathbf{k}) = \mathcal{S}(\mathbf{k}) \cap \{1, \ldots, J_1\}$, i.e., $\mathcal{S}_c(\mathbf{k})$ is the set of *critical* servers encountered by customers following the path $\mathbf{k}$ before they become class $e(\mathbf{k})$ customers. In the case of $J_1 \geqslant 1$ we define Π as the set of all permutations $\pi = (\pi_1, \ldots, \pi_{J_1})$ of $1, \ldots, J_1$ such that $\pi_m \in \mathcal{J}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}$, $m = 1, \ldots, J_1$, i.e., the station π_m can be reached through $\{\pi_1, \ldots, \pi_{m-1}, J_1+1, \ldots, J\}$. For $\pi \in \Pi$, we set

$$D^\pi = \{y \in \mathbb{R}^{J_1} : y_{\pi_1} \geqslant \ldots \geqslant y_{\pi_{J_1}},\ y_{\pi_m} \leqslant \max_{\mathbf{k}\in\mathcal{K}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}(\pi_m)} y^*_{b(\mathbf{k})}\ \forall m\}.$$

Let $D = \bigcup_{\pi\in\Pi} D^\pi$. For $k = 1, \ldots, K$ and $y \in \mathbb{R}$, let us define

$$H_k(y) = \int_y^\infty \big(1 - G_k(\eta)\big) d\eta. \tag{3.1}$$

Each function H_k is finite by (2.4) (or (2.5)). We put $H = (H_1, \ldots, H_K)$. Next, we define the mapping $\Phi = (\Phi_1, \ldots, \Phi_{J_1}) : D \to \mathbb{R}_+^{J_1}$ by the formula

$$\Phi_j(y_1, \ldots, y_{J_1}) = \sum_{\mathbf{k} \in \overline{\mathcal{C}}(j)} \rho_{\mathbf{k}} [H_{b(\mathbf{k})}(y_j) - H_{b(\mathbf{k})}(\min_{i \in \mathcal{S}_c(\mathbf{k})} y_i)]^+, \tag{3.2}$$

$j = 1, \ldots, J_1$. Here and elsewhere in this paper, the minimum (maximum) taken over the empty set should be interpreted as ∞ $(-\infty)$. Note that the series (3.2) converges. Indeed, for any j, by the definition of $\rho_{\mathbf{k}}$ and (2.1)–(2.3), we have

$$\begin{aligned} \Phi_j(y_1, \ldots, y_{J_1}) &\leqslant \sum_{\mathbf{k} \in \overline{\mathcal{C}}(j)} \rho_{\mathbf{k}} \max_{k'=1,\ldots,K} H_{k'}(y_j) \\ &\leqslant \sum_{k \in \mathcal{C}(j)} m_k \big\{ \big(I + P' + (P')^2 + \ldots \big) \alpha \big\}_k \max_{k'=1,\ldots,K} H_{k'}(y_j) \\ &= \rho_j \max_{k'=1,\ldots,K} H_{k'}(y_j). \end{aligned} \tag{3.3}$$

PROPOSITION 3.1. *The function Φ is a homeomorphism of D onto $\mathbb{R}_+^{J_1}$.*

The proof of this proposition and an explicit algorithm for inverting Φ may be obtained by an appropriate extension of the proof of Proposition 5.5 in [17].

The main result of this paper is the following theorem.

THEOREM 3.1 (Characterization of the invariant manifold). *For $J_1 \geqslant 1$ and $F = (F_j)_{j=1,\ldots,J_1} \in \mathbb{R}^{J_1}$, let $Z^F = \big(Z_k^F(\cdot)\big)_{k=1,\ldots,K}$, where for $s \in \mathbb{R}$,*

$$Z_k^F(s) = \begin{cases} 0, & s(k) > J_1, \\ \sum_{\mathbf{k} \in \tilde{\mathcal{C}}(k)} \alpha_{\mathbf{k}} [H_{b(\mathbf{k})}(F_{s(k)}) - H_{b(\mathbf{k})}(s \wedge \min_{i \in \mathcal{S}_c(\mathbf{k})} F_i)]^+, & s(k) \leqslant J_1. \end{cases} \tag{3.4}$$

Also, let $\mathfrak{X}(F)$ be the vector of functions of the form (2.6) *with $Z(t,s) = Z^F(s)$ for all $t \geqslant 0$, $s \in \mathbb{R}$, and satisfying* (2.9), (2.12) *and* (2.21)–(2.23). *Let $\mathfrak{S}^D = \{Z^F : F \in D\}$ if $J_1 \geqslant 1$ and let $\mathfrak{S}^D = \{\mathbf{0}\}$ otherwise, where $\mathbf{0}$ is the vector of K distribution functions corresponding to the zero measure: $\mathbf{0}_k \equiv 0$, $k = 1, \ldots, K$. Then $\mathfrak{S} = \mathfrak{S}^D$.*

Intuitively, this result may be explained as follows. As it should be expected, the invariant queue length (workload) vanishes at strictly subcritical stations. For a critical station j, F_j is the lead time of the fluid currently in service at j. By the fact that the EDF service discipline is used, there is no mass with lead times smaller than F_j at this station. We will later show that all the fluid with lead times greater than F_j which visits server j along its route is currently either at j, or upstream (see (4.4) of Proposition 4.1). Consequently, for $\mathbf{k} \in \overline{\mathcal{C}}(j)$, the fluid following the path $\mathbf{k}$ which is currently at station j must have lead times not greater than $\min_{i \in \mathcal{S}_c(\mathbf{k})} F_i$, since otherwise it has not left one of the upstream critical stations along the path $\mathbf{k}$. In particular, in spite of the assumed Markovian routing between customer classes, the invariant states form a complex, highly non-Markovian structure, which is not observed in the case of FIFO, HLPPS or other previously investigated service protocols. Another consequence of the above fact is that the network topology implies an ordering of F_j, $j = 1, \ldots, J_1$. For example, if $\overline{\mathcal{C}}(j) = \{\mathbf{k}\}$, where $\mathbf{k} = (k_1, k_2)$, $j' = s(k_1) \neq j = s(k_2)$, then $F_j \leqslant F_{j'}$, since the fluid currently served at j must have left the station j' by this time. In general, such ordering relations are much more complicated and, together with the condition (2.4), they imply that $F = (F_j)_{j=1,\ldots,J_1} \in D$. The appearance of $H_{b(\mathbf{k})}$ in the term corresponding to $\mathbf{k}$ in (3.4) is not surprising, because $H_{b(\mathbf{k})}$ corresponds (up to a multiplicative factor) to the stationary excess distribution associated with the lead time distribution $G_{b(\mathbf{k})}$ of the fluid following the path $\mathbf{k}$.

Define the mapping $\Psi = (\Psi_1, \ldots, \Psi_K) : D \to \mathbb{R}_+^K$ by

$$\Psi_k(y_1, \ldots, y_{J_1}) = \sum_{\mathbf{k} \in \tilde{\mathcal{C}}(k)} \alpha_{\mathbf{k}} [H_{b(\mathbf{k})}(y_{s(k)}) - H_{b(\mathbf{k})}(\min_{i \in \mathcal{S}_c(\mathbf{k})} y_i)]^+, \tag{3.5}$$

if $s(k) \leqslant J_1$ and $\Psi_k \equiv 0$ otherwise. An argument similar to (3.3) shows that the series in (3.5) and the analogous series in the definition of Z^F converge.

COROLLARY 3.1 (Lifting map). *For $w = (w_j)_{j=1,\ldots,J_1} \in \mathbb{R}_+^{J_1}$, $\mathfrak{X}\big(\Phi^{-1}(w)\big)$ is the unique invariant EDF fluid model with workload $w_j = \lim_{s\to\infty} \sum_{k \in \mathcal{C}(j)} W_k(s)$ at station j for each $j = 1, \ldots, J_1$. In particular, the map $\Delta : \mathbb{R}_+^{J_1} \to \mathbb{R}_+^K$ given by the formula*

$$\Delta = \Psi \circ \Phi^{-1} \tag{3.6}$$

maps the workload vector w to the corresponding queue length vector

$$z = (z_k)_{k=1,\ldots,K}, \quad z_k = \lim_{s\to\infty} Z_k(s),$$

in the invariant EDF fluid model $\mathfrak{X}\big(\Phi^{-1}(w)\big)$.

REMARK 3.1. It is easy to check that if we define an invariant FISFO fluid model by the equations (2.16)–(2.20) (instead of (2.8)–(2.12), (2.15)), requiring that $\overline{Z}(t, \tilde{s}) = \overline{Z}(0, \tilde{s})$ for all $t, \tilde{s} \geqslant 0$, then the corresponding invariant manifold consists of a single state $\mathbf{0}$.

3.1. Example. Perhaps the most striking consequence of Theorem 3.1 is that the mapping $\Delta = (\Delta_1, \ldots, \Delta_K)$ in Corollary 3.1 is, in general, nonlinear, even in the acyclic FISFO case. Note that while general EDF networks with Markovian routing are not necessarily HL, all FISFO networks with fixed customer routes enjoy the HL property.

EXAMPLE 3.1. Let us consider a critical FISFO fluid model with $J = J_1 = 2$, $K = 4$, $s(1) = s(4) = 1$, $s(2) = s(3) = 2$, $\alpha_1 > 0$, $\alpha_3 > 0$, $\alpha_2 = \alpha_4 = 0$ and

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We have $\Pi = \{(1,2),(2,1)\}$, $D = (-\infty, 0] \times (-\infty, 0]$, $H_1(y) = H_3(y) = (-y)^+$ and the mapping $\Phi : D \to \mathbb{R}_+^2$ is given by

$$\Phi_1(y_1, y_2) = -\rho_{(1)} y_1 + \rho_{(3,4)} [y_2 - y_1]^+, \;\; \Phi_2(y_1, y_2) = \rho_{(1,2)} [y_1 - y_2]^+ - \rho_{(3)} y_2.$$

One may check by direct computation that for $(w_1, w_2) \in \mathbb{R}_+^2$ such that $w_2 \geqslant \rho_{(3)} w_1 / \rho_{(1)}$, we have

$$\Delta_1(w_1, w_2) = w_1 / m_1, \quad \Delta_2(w_1, w_2) = \alpha_1 (w_2 - \rho_{(3)} w_1 / \rho_{(1)}),$$
$$\Delta_3(w_1, w_2) = \alpha_3 (m_2 w_1 / m_1 + w_2), \quad \Delta_4(w_1, w_2) = 0,$$

and for $(w_1, w_2) \in \mathbb{R}_+^2$ such that $w_2 < \rho_{(3)} w_1 / \rho_{(1)}$,

$$\Delta_1(w_1, w_2) = \alpha_1 (w_1 + m_4 w_2 / m_3), \quad \Delta_2(w_1, w_2) = 0,$$
$$\Delta_3(w_1, w_2) = w_2 / m_3, \quad \Delta_4(w_1, w_2) = \alpha_3 (w_1 - \rho_{(1)} w_2 / \rho_{(3)}).$$

In particular, Δ is nonlinear.

REMARK 3.2. It is easy to see that for every FISFO network, Φ and Δ are *piecewise linear*, i.e., the domain of each of these maps can be decomposed by a finite number of hyperplanes into disjoint sets such that the corresponding mapping is linear on each of these sets. In fact, for Φ (Δ), these sets coincide with D^π ($\Phi(D^\pi)$), $\pi \in \Pi$. In the general EDF case, neither Φ nor Δ is piecewise linear.

4. FRONTIERS AND NULL SETS

In this section we define the frontiers and the null set of an invariant state (Definitions 4.1 and 4.2) and we use these notions to show several properties of invariant EDF fluid models. Proposition 4.1 shows that all the incoming fluid with lead times greater than the frontier at station j and visiting this station along its route is still either at j or upstream. It also implies Theorem 3.1 for strictly subcritical networks (Corollary 4.1) and shows that invariant EDF fluid models are, in some sense, maximal (compare (4.3) with (4.4), see also (4.6)). Proposition 4.2 shows that the frontiers and the null set determine the corresponding invariant EDF fluid model uniquely. In Proposition 4.3 we show that the vector of the frontiers corresponding to an invariant EDF fluid model with nonzero total mass at every critical station belongs to the set D. Finally, Lemma 4.2 establishes partial ordering and an upper bound for the frontiers.

DEFINITION 4.1. Let $\mathfrak{X}$ be an EDF fluid model of the form (2.6). Let $t \geqslant 0$ and let $j \in \{1, \ldots, J\}$ be such that $\sum_{k \in \mathcal{C}(j)} Z_k(t, \cdot)$ is not identically equal to zero. The quantity $F_j(t) = \inf\left\{s \in \mathbb{R} : \sum_{k \in \mathcal{C}(j)} Z_k(t, s) > 0\right\}$ will be called the *frontier* at station j at time t.

In other words, $F_j(t)$ is the left endpoint of the support of the measure with the distribution function $\sum_{k \in \mathcal{C}(j)} Z_k(t, \cdot)$. By the continuity of Z,

$$\sum_{k \in \mathcal{C}(j)} Z_k\big(t, F_j(t)\big) = 0. \tag{4.1}$$

If the EDF fluid model $\mathfrak{X}$ is invariant (which will usually be assumed in the sequel), then $F_j(t)$ is constant in t. In this case, the variable t in $F_j(t)$ will be skipped.

DEFINITION 4.2. Let $\mathfrak{X}$ be an invariant EDF fluid model of the form (2.6). The set $\mathcal{N}(\mathfrak{X}) = \left\{j \in \{1, \ldots, J\} : \sum_{k \in \mathcal{C}(j)} Z_k(\cdot) \equiv 0\right\}$ will be called the *null set* of $\mathfrak{X}$.

In what follows, we will sometimes write $\mathcal{N}$ instead of $\mathcal{N}(\mathfrak{X})$, when it is clear from the context what fluid model $\mathfrak{X}$ we have in mind. Note that for an invariant state $\mathfrak{X}$, F_j is defined iff $j \notin \mathcal{N}$. The following lemma is a restatement of (2.11) for an invariant EDF fluid model.

LEMMA 4.1. *Let $\mathfrak{X}$ be an invariant EDF fluid model of the form* (2.6). *For every $j \notin \mathcal{N}$ and every $t \geqslant 0$, $s \in \mathbb{R}$ such that $s - t > F_j$, we have*

$$\begin{aligned} \sum_{k \in \mathcal{C}(j)} m_k \big\{Q\big(Z(s-t) - Z(s-t-dt)\big)\big\}_k \\ = \Big(1 - \sum_{k \in \mathcal{C}(j)} m_k \big\{Q\big(\alpha \circ G(s-t)\big)\big\}_k\Big) dt. \end{aligned} \tag{4.2}$$

Proof. By (2.23), we have

$$Y_j(t, s-t) = t - \sum_{k \in \mathcal{C}(j)} m_k \Big\{Q\Big(\alpha \circ \int_0^t G(s-\eta)d\eta + Z(s) - Z(s-t)\Big)\Big\}_k.$$

By (2.11), for $s - t > F_j$, $Y_j(t, s-t)$ does not grow in t, which implies (4.2). ∎

In the sequel, we will need the following refinement of Lemma 4.1.

PROPOSITION 4.1. *Let $\mathfrak{X}$ be an EDF fluid model of the form* (2.6). *For every $s_1 < s_2$ and for t large enough*,

$$Q\big(Z(t, s_2) - Z(t, s_1)\big) \leqslant \lambda(s_2 - s_1) - Q\Big(\alpha \circ \int_{s_1}^{s_2} G(\eta)d\eta\Big). \tag{4.3}$$

Moreover, if the EDF fluid model $\mathfrak{X}$ is invariant, then $\{j : \rho_j < 1\} \subseteq \mathcal{N}$ *and for every* $j \notin \mathcal{N}$, $k \in \mathcal{C}(j)$ *and* $s_2 > s_1 \geqslant F_j$,

$$\{Q\big(Z(s_2) - Z(s_1)\big)\}_k = \{Q\big[\alpha \circ \big(H(s_1) - H(s_2)\big)\big]\}_k. \tag{4.4}$$

P r o o f. By (2.7) and (2.8), we have

$$Z(t,s) = Z(0,t+s) + \alpha \circ \int_0^t G(s+\eta)d\eta - (I-P')D(t,s),$$

which, by (2.1), implies

$$QZ(t,s) = QZ(0,t+s) + Q\big(\alpha \circ \int_0^t G(s+\eta)d\eta\big) - D(t,s).$$

Thus, by (2.14), for fixed $s_1 \leqslant s_2$ and $t \geqslant 0$ large enough,

$$\begin{aligned} (4.5)\quad & Q\big(Z(t,s_2) - Z(t,s_1)\big) \\ &= Q\Big(\alpha \circ \big(\int_{s_1+t}^{s_2+t} G(\eta)d\eta - \int_{s_1}^{s_2} G(\eta)d\eta\big)\Big) - \big(D(t,s_2) - D(t,s_1)\big) \\ &\leqslant Q\alpha(s_2 - s_1) - Q\big(\alpha \circ \int_{s_1}^{s_2} G(\eta)d\eta\big), \end{aligned}$$

which, by (2.2), implies (4.3). Now assume that the EDF fluid model $\mathfrak{X}$ is invariant, $j \notin \mathcal{N}$ and $s_2 > s_1 \geqslant F_j$. We have

$$\begin{aligned} s_2 - s_1 - \sum_{k \in \mathcal{C}(j)} m_k \{Q\big(\alpha \circ \int_{s_1}^{s_2} G(\eta)d\eta\big)\}_k &= \sum_{k \in \mathcal{C}(j)} m_k \{Q\big(Z(s_2) - Z(s_1)\big)\}_k \\ &\leqslant \rho_j(s_2 - s_1) - \sum_{k \in \mathcal{C}(j)} m_k \{Q\big(\alpha \circ \int_{s_1}^{s_2} G(\eta)d\eta\big)\}_k, \end{aligned}$$

where the equality follows from (4.2) and the inequality from (4.3) and (2.3). Since $\rho_j \leqslant 1$, this is possible only if $\rho_j = 1$ and equality holds on every coordinate $k \in \mathcal{C}(j)$ of (4.3), i.e., for every $k \in \mathcal{C}(j)$,

$$\begin{aligned} (4.6)\qquad \{Q\big(Z(s_2) - Z(s_1)\big)\}_k &= \lambda_k(s_2 - s_1) - \{Q\big(\alpha \circ \int_{s_1}^{s_2} G(\eta)d\eta\big)\}_k \\ &= \Big\{Q\Big(\alpha \circ \int_{s_1}^{s_2} \big(1 - G(\eta)\big)d\eta\Big)\Big\}_k \\ &= \{Q\big[\alpha \circ \big(H(s_1) - H(s_2)\big)\big]\}_k, \end{aligned}$$

where (2.2) and (3.1) were used for the second and third equation, respectively. ∎

An immediate consequence of Proposition 4.1 is that Theorem 3.1 holds in the strictly subcritical case.

COROLLARY 4.1. *If the network is strictly subcritical, then* $\mathfrak{S} = \{\mathbf{0}\}$.

Indeed, in this case, by Proposition 4.1, for any invariant EDF fluid model $\mathfrak{X}$, $\mathcal{N} = \{1, \ldots, J\}$, which implies that $Z \equiv 0$.

We have thus characterized the invariant manifold for a strictly subcritical network. From now on, we assume that $J_1 \geqslant 1$.

PROPOSITION 4.2. *The set* $\mathcal{N} = \mathcal{N}(\mathfrak{X})$ *and the frontiers* $F_j, j \in \{1, \ldots, J\} \setminus \mathcal{N}$, *determine the corresponding invariant EDF fluid model* $\mathfrak{X}$ *uniquely.*

Proof. By Proposition 2.1, it suffices to show that the coordinate $Z(\cdot)$ of $\mathfrak{X}$ is determined uniquely by $\mathcal{N}$ and (F_j). Let $n = |\mathcal{N}|$. If $n = J$, then $Z \equiv \mathbf{0}$ and there is nothing to prove. Assume that $n < J$. Without loss of generality we can also assume that $\{1, \ldots, J\} \setminus \mathcal{N} = \{1, \ldots, J-n\}$. Let $\pi = (\pi_1, \ldots, \pi_{J-n})$ be a permutation of $1, \ldots, J-n$ such that

$$F_{\pi_1} \geqslant F_{\pi_2} \geqslant \ldots \geqslant F_{\pi_{J-n}}. \tag{4.7}$$

We proceed by induction. Let $F_{\pi_0} = \infty$ and assume for some $i \in \{1, \ldots, J-n\}$, the increments of Z are determined uniquely by $\mathcal{N}$ and (F_j) on the interval $[F_{\pi_{i-1}}, \infty)$ (for $i = 1$, the hypothesis is vacuous). Our aim is to show that they are also determined uniquely by this data on $[F_{\pi_i}, F_{\pi_{i-1}})$, and thus on $[F_{\pi_i}, \infty)$.

Let $\mathfrak{K}^{(i)} = \bigcup_{j=i}^{J-n} \mathcal{C}(\pi_j)$ and let

$$\mathfrak{L}^{(i)} = \{1, \ldots, K\} \setminus \mathfrak{K}^{(i)} = \bigcup_{j \in \mathcal{N}} \mathcal{C}(j) \cup \bigcup_{j=1}^{i-1} \mathcal{C}(\pi_j).$$

By the definitions of $\mathcal{N}$ and F_j, together with (4.7), we have

$$Z_k(F_{\pi_{i-1}}) - Z_k(F_{\pi_i}) = 0, \quad k \in \mathfrak{L}^{(i)}. \tag{4.8}$$

Thus, for $k \in \mathfrak{K}^{(i)}$ and $s_1, s_2 \in [F_{\pi_i}, F_{\pi_{i-1}})$, $s_1 \leqslant s_2$,

$$\big\{Q\big(Z(s_2) - Z(s_1)\big)\big\}_k = \big\{Q^{(i)}\big(Z^{(i)}(s_2) - Z^{(i)}(s_1)\big)\big\}_k, \tag{4.9}$$

where $Q^{(i)} = (Q_{kl})_{k,l \in \mathfrak{K}^{(i)}}$ and $Z^{(i)} = (Z_k)_{k \in \mathfrak{K}^{(i)}}$. Let $P^{(i)} = (p_{kl}^{(i)})_{k,l \in \mathfrak{K}^{(i)}}$, where

$$p_{kl}^{(i)} = p_{kl} + \sum_{n=1}^{\infty} \sum_{k_1, \ldots, k_n \in \mathfrak{L}^{(i)}} p_{kk_1} p_{k_1 k_2} \ldots p_{k_n l}$$

is the probability that a class k customer moves to class l, visiting only stations from $\mathcal{N} \cup \{\pi_1, \ldots, \pi_{i-1}\}$ along the way. It is easy to see that $Q^{(i)} = I + (P^{(i)})' +$

$\big((P^{(i)})'\big)^2+\ldots$, so $Q^{(i)}$ is invertible and $(Q^{(i)})^{-1}=I-(P^{(i)})'$. By (4.9) and (4.4), for $F_{\pi_i}\leqslant s_1<s_2<F_{\pi_{i-1}}$,

$$Q^{(i)}\big(Z^{(i)}(s_2)-Z^{(i)}(s_1)\big)=\Big(\big\{Q\big[\alpha\circ\big(H(s_1)-H(s_2)\big)\big]\big\}_k\Big)_{k\in\mathfrak{K}^{(i)}}. \tag{4.10}$$

Multiplying (4.10) from the left by $(Q^{(i)})^{-1}$ we get the increments of $Z^{(i)}$ (and thus, by (4.8), of Z) on the interval $[F_{\pi_i},F_{\pi_{i-1}})$. This ends the inductive proof, establishing uniqueness of the increments of Z with given $\mathcal{N}$ and (F_j) on $[F_{\pi_{J-n}},\infty)$. By (4.7) and the definition of F_j, $Z\equiv 0$ on $(-\infty,F_{\pi_{J-n}})$. ■

PROPOSITION 4.3. *Let $\mathfrak{X}$ be an invariant EDF fluid model with null set $\mathcal{N}(\mathfrak{X})=\{J_1+1,\ldots,J\}$ and frontiers $F=(F_j)_{j=1,\ldots,J_1}$. Then $F\in D$.*

P r o o f. We will first construct a permutation $\pi\in\Pi$ such that

$$F_{\pi_1}\geqslant F_{\pi_2}\geqslant\ldots\geqslant F_{\pi_{J_1}}. \tag{4.11}$$

Let $B_1=\{1,\ldots,J_1\}$ and $I_1=\{j\in B_1: F_j=\max_{i\in B_1}F_i\}$. We will show that

$$I_1\subseteq\mathcal{J}_{J_2}^{J_1+1,\ldots,J}. \tag{4.12}$$

Let $j\in I_1$ and let $k\in\mathcal{C}(j)$. By (2.7) and (2.24), for $t\geqslant 0$ and $s\in\mathbb{R}$ we have

$$\begin{aligned}A_k(t,s)&=\alpha_k\int_0^t G_k(s+\eta)d\eta+\sum_{k_1=1}^K p_{k_1k}D_{k_1}(t,s)\\&=\alpha_k\int_0^t G_k(s+\eta)d\eta+\sum_{k_1=1}^K p_{k_1k}\big(Z_{k_1}(t+s)-Z_{k_1}(s)+A_{k_1}(t,s)\big).\end{aligned}$$

Using (2.7) and (2.24) once again, we get

$$\begin{aligned}A_k(t,s)&=\sum_{k_1=1}^K p_{k_1k}\big(Z_{k_1}(t+s)-Z_{k_1}(s)+\alpha_{k_1}\int_0^t G_{k_1}(s+\eta)d\eta\big)\\&\qquad+\alpha_k\int_0^t G_k(s+\eta)d\eta+\sum_{k_1,k_2=1}^K p_{k_2k_1}p_{k_1k}D_{k_2}(t,s)\\&=\sum_{k_1=1}^K p_{k_1k}\big(Z_{k_1}(t+s)-Z_{k_1}(s)+\alpha_{k_1}\int_0^t G_{k_1}(s+\eta)d\eta\big)\\&+\alpha_k\int_0^t G_k(s+\eta)d\eta+\sum_{k_1,k_2=1}^K p_{k_2k_1}p_{k_1k}\big(Z_{k_2}(t+s)-Z_{k_2}(s)+A_{k_2}(t,s)\big).\end{aligned}$$

Using (2.7), (2.24) again and iterating, we obtain

$$(4.13)\quad A_k(t,s)=\alpha_k\int_0^t G_k(s+\eta)d\eta$$
$$+\sum_{n=1}^{\infty}\sum_{k_1,\ldots,k_n=1}^{K} p_{k_nk_{n-1}}\ldots p_{k_1k}\big(Z_{k_n}(t+s)-Z_{k_n}(s)+\alpha_{k_n}\int_0^t G_{k_n}(s+\eta)d\eta\big).$$

The series in (4.13) converges by (2.1). Note that (4.13) may be rewritten as

$$(4.14)\quad A_k(t,s)=\alpha_k\int_0^t G_k(s+\eta)d\eta$$
$$+\sum_{s(k_1)\in B_1} p_{k_1k}\big\{Q\big(Z(t+s)-Z(s)+\alpha\circ\int_0^t G(s+\eta)d\eta\big)\big\}_{k_1}$$
$$+\sum_{n=2}^{\infty}\sum_{s(k_1),\ldots,s(k_{n-1})\notin B_1}\sum_{s(k_n)\in B_1} p_{k_nk_{n-1}}\ldots p_{k_1k}$$
$$\times\big\{Q\big(Z(t+s)-Z(s)+\alpha\circ\int_0^t G(s+\eta)d\eta\big)\big\}_{k_n}+$$
$$\sum_{n=1}^{\infty}\sum_{s(k_1),\ldots,s(k_n)\notin B_1} p_{k_nk_{n-1}}\ldots p_{k_1k}\big(Z_{k_n}(t+s)-Z_{k_n}(s)+\alpha_{k_n}\int_0^t G_{k_n}(s+\eta)d\eta\big).$$

By Proposition 4.1, $Z_{k_n}\equiv 0$ for $s(k_n)\notin B_1$. This, together with (2.4), (2.14) and (4.4), implies that for $s>F_j=\max_{i\in B_1}F_i$ and t large enough,

$$(4.15)\quad A_k(t,s)-A_k(t,F_j)$$
$$=\sum_{s(k_1)\in B_1} p_{k_1k}\big\{Q\big(Z(F_j)-Z(s)+\alpha\circ[H(F_j)-H(s)]\big)\big\}_{k_1}$$
$$+\alpha_k\big(s-F_j-\int_{F_j}^s G_k(\eta)d\eta\big)+\sum_{n=2}^{\infty}\sum_{s(k_1),\ldots,s(k_{n-1})\notin B_1}\sum_{s(k_n)\in B_1} p_{k_nk_{n-1}}\ldots p_{k_1k}$$
$$\times\big\{Q\big(Z(F_j)-Z(s)+\alpha\circ[H(F_j)-H(s)]\big)\big\}_{k_n}$$
$$+\sum_{n=1}^{\infty}\sum_{s(k_1),\ldots,s(k_n)\notin B_1}\alpha_{k_n}p_{k_nk_{n-1}}\ldots p_{k_1k}\big(s-F_j-\int_{F_j}^s G_{k_n}(\eta)d\eta\big)$$
$$=\alpha_k\big(s-F_j-\int_{F_j}^s G_k(\eta)d\eta\big)$$
$$+\sum_{n=1}^{\infty}\sum_{s(k_1),\ldots,s(k_n)\notin B_1}\alpha_{k_n}p_{k_nk_{n-1}}\ldots p_{k_1k}\big(s-F_j-\int_{F_j}^s G_{k_n}(\eta)d\eta\big).$$

Suppose that (4.12) is false. Let $j \in I_1 \setminus \mathcal{J}_{J_2}^{J_1+1,\ldots,J}$. Then for every $k \in \mathcal{C}(j)$ (4.15) vanishes. Indeed, the fact that $j \notin \mathcal{J}_{J_2}^{J_1+1,\ldots,J}$ implies that $\alpha_k = 0$ and, for $s(k_1), \ldots, s(k_n) \notin B_1$, $\alpha_{k_n} p_{k_n k_{n-1}} \ldots p_{k_1 k} = 0$. By (2.8), (2.4), (2.14) and (4.15), for s and t as above,

$$\begin{aligned} 0 \leqslant Z_k(s) - Z_k(F_j) &= A_k(t,s) - A_k(t,F_j) - \big(D_k(t,s) - D_k(t,F_j)\big) \\ &\leqslant A_k(t,s) - A_k(t,F_j) = 0. \end{aligned}$$

Thus, by (4.1), $\sum_{k\in\mathcal{C}(j)} Z_k(s) = \sum_{k\in\mathcal{C}(j)} \big(Z_k(s) - Z_k(F_j)\big) = 0$ for $s > F_j$, which contradicts the definition of F_j. We have proved (4.12).

Let $n_1 = |I_1|$ and let $\pi_1, \ldots, \pi_{n_1}$ be such that $I_1 = \{\pi_1, \ldots, \pi_{n_1}\}$. Then, by (4.12), for $m = 1, \ldots, n_1$, $\pi_m \in \mathcal{J}_{J_2}^{J_1+1,\ldots,J} \subseteq \mathcal{J}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}$ and $F_{\pi_1} = \ldots = F_{\pi_{n_1}} > \max_{i\notin\{\pi_1,\ldots,\pi_{n_1}\}} F_i$. If $n_1 = J_1$, the permutation $\pi \in \Pi$ satisfying (4.11) has been found, otherwise we continue our construction inductively as follows.

Assume that for some $n_k \in \{1, \ldots, J_1 - 1\}$, the numbers $\pi_1, \ldots, \pi_{n_k}$ have been chosen so that

$$\pi_m \in \mathcal{J}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}, \quad m = 1, \ldots, n_k, \tag{4.16}$$

$$F_{\pi_1} \geqslant \ldots \geqslant F_{\pi_{n_k}} > \max_{i\notin\{\pi_1,\ldots,\pi_{n_k}\}} F_i. \tag{4.17}$$

(We have just proved the validity of this assumption for $k = 1$.) Let

$$B_{k+1} = B_1 \setminus \{\pi_1, \ldots, \pi_{n_k}\}, \quad I_{k+1} = \{j \in B_{k+1} : F_j = \max_{i\in B_{k+1}} F_i\}.$$

We want to get

$$I_{k+1} \subseteq \mathcal{J}_{J_2+n_k}^{\pi_1,\ldots,\pi_{n_k},J_1+1,\ldots,J}. \tag{4.18}$$

The argument is similar to the one given above, with I_1, $\mathcal{J}_{J_2}^{J_1+1,\ldots,J}$ and B_1 replaced by I_{k+1}, $\mathcal{J}_{J_2+n_k}^{\pi_1,\ldots,\pi_{n_k},J_1+1,\ldots,J}$ and B_{k+1}, respectively. The only difference is that, in general, the terms $Z_{k_n}(t+s)$, $Z_{k_n}(s)$ with $s(k_n) \in \{\pi_1, \ldots, \pi_{n_k}\}$ in (4.14) (with B_1 replaced by B_{k+1}) do not vanish. However, for $s(k_n) \in \{\pi_1, \ldots, \pi_{n_k}\}$ and $s \in (F_j, F_{\pi_{n_k}})$, by (4.17) and the definition of the frontier, $Z_{k_n}(s) = Z_{k_n}(F_j) = 0$. For t large enough, it follows that $Z_{k_n}(t+s) = Z_{k_n}(t+F_j)$ by (2.14). Hence, for $s \in (F_j, F_{\pi_{n_k}})$, (4.15) (with B_1 replaced by B_{k+1}) holds and the argument goes through as in the case of $k = 1$. Let $n_{k+1} = n_k + |I_{k+1}|$ and choose $\pi_{n_k+1}, \ldots, \pi_{n_{k+1}}$ so that $I_{k+1} = \{\pi_{n_k+1}, \ldots, \pi_{n_{k+1}}\}$. By (4.18), the condition (4.16) holds with k replaced by $k+1$. The definition of the set I_{k+1} implies that (4.17) also holds with k replaced by $k+1$. The inductive step is complete.

When the above construction terminates, we have a permutation $\pi \in \Pi$ satisfying (4.11). We will show that $F \in D^{\pi}$. Suppose that $F \notin D^{\pi}$. By (4.11) and the definition of D^{π}, this implies that for some $m \in \{1, \ldots, J_1\}$,

$$F_{\pi_m} > \max_{\mathbf{k} \in \mathcal{K}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}(\pi_m)} y^*_{b(\mathbf{k})}. \tag{4.19}$$

Let $B = \{\pi_1, \ldots, \pi_{m-1}, J_1+1, \ldots, J\}$, $k \in \mathcal{C}(\pi_m)$ and let $s > F_{\pi_m}$. The relation (4.11), together with the reasoning similar to that in (4.13)–(4.15), implies that for t large enough,

$$\begin{aligned} 0 &\leqslant A_k(t,s) - A_k(t, F_{\pi_m}) \\ &= \sum_{\mathbf{k} \in \mathcal{K}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}(\pi_m)} \alpha_{\mathbf{k}} \int_0^t \big(G_{b(\mathbf{k})}(s+\eta) - G_{b(\mathbf{k})}(F_{\pi_m}+\eta)\big) d\eta \\ &\quad + \sum_{\mathbf{k} \in \mathcal{K}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}(\pi_m) \setminus \{(k)\}} p_{\mathbf{k}} \big(Z_{b(\mathbf{k})}(F_{\pi_m}) - Z_{b(\mathbf{k})}(s)\big) \\ &\leqslant 0, \end{aligned} \tag{4.20}$$

since $G_{b(\mathbf{k})}(s+\eta) = G_{b(\mathbf{k})}(F_{\pi_m}+\eta) = 1$ for $\mathbf{k} \in \mathcal{K}_{J_2+m-1}^{\pi_1,\ldots,\pi_{m-1},J_1+1,\ldots,J}(\pi_m)$ and $\eta \geqslant 0$ by (4.19), so the integral in (4.20) vanishes. Thus,

$$A_k(t,s) - A_k(t, F_{\pi_m}) = 0,$$

which, as in the proof of (4.11), implies that $\sum_{k \in \mathcal{C}(\pi_m)} Z_k(s) = 0$, and this contradicts the definition of F_{π_m}. ■

LEMMA 4.2. *Let $\mathfrak{X}$ be an invariant EDF fluid model with $\mathcal{N} = \mathcal{N}(\mathfrak{X})$ and let $F = (F_j)_{j \in \{1,\ldots,J\} \setminus \mathcal{N}}$ be the corresponding frontiers. Then for $j \in \{1, \ldots, J\} \setminus \mathcal{N}$,*

$$F_j < \sup_{\mathbf{k} \in \overline{\mathcal{C}}(j)} \big(y^*_{b(\mathbf{k})} \wedge \min_{i \in \mathcal{S}(\mathbf{k}) \setminus \mathcal{N}} F_i\big). \tag{4.21}$$

Proof. Suppose that for some $j \in \{1, \ldots, J\} \setminus \mathcal{N}$ the inequality (4.21) is false, i.e.,

$$y^*_{b(\mathbf{k})} \leqslant F_j \quad \text{or} \quad \min_{i \in \mathcal{S}(\mathbf{k}) \setminus \mathcal{N}} F_i \leqslant F_j \tag{4.22}$$

for every $\mathbf{k} \in \overline{\mathcal{C}}(j)$. Let $B = \big\{i \in \{1, \ldots, J\} \setminus \mathcal{N} \colon F_i \leqslant F_j\big\}$, $k \in \mathcal{C}(j)$ and $s > F_j$.

Using (4.13) and proceeding as in (4.14) and (4.15), we get, for t large enough,

$$
\begin{aligned}
(4.23)\quad & A_k(t,s) - A_k(t,F_j) \\
= & \sum_{s(k_1)\in B} p_{k_1 k}\big\{Q\big(Z(F_j) - Z(s) + \alpha\circ[H(F_j) - H(s)]\big)\big\}_{k_1} \\
& + \alpha_k\Big(s - F_j - \int_{F_j}^{s} G_k(\eta)d\eta\Big) + \sum_{n=2}^{\infty}\ \sum_{s(k_1),\dots,s(k_{n-1})\notin B}\ \sum_{s(k_n)\in B} p_{k_n k_{n-1}}\dots p_{k_1 k} \\
& \qquad\qquad \times \big\{Q\big(Z(F_j) - Z(s) + \alpha\circ[H(F_j) - H(s)]\big)\big\}_{k_n} \\
& + \sum_{n=1}^{\infty}\ \sum_{s(k_1),\dots,s(k_n)\notin B} \alpha_{k_n} p_{k_n k_{n-1}}\dots p_{k_1 k}\Big(s - F_j - \int_{F_j}^{s} G_{k_n}(\eta)d\eta\Big) \\
& + \sum_{n=1}^{\infty}\ \sum_{s(k_1),\dots,s(k_n)\notin B} p_{k_n k_{n-1}}\dots p_{k_1 k}\big(Z_{k_n}(F_j) - Z_{k_n}(s)\big).
\end{aligned}
$$

The first two terms on the right-hand side of (4.23) vanish by (4.4). The next two terms vanish because, by (4.22), $\alpha_k > 0$ implies $y_k^* \leqslant F_j$ and for $k_1,\dots,k_n$ such that $\alpha_{k_n} p_{k_n k_{n-1}}\dots p_{k_1 k} > 0$ and $s(k_1),\dots,s(k_n)\notin B$, $y^*_{b(\mathbf{k})} \leqslant F_j$. Finally, the fifth term is nonpositive. Thus, $A_k(t,s) - A_k(t,F_j) = 0$, which, as in the proof of Proposition 4.3, leads to a contradiction. ∎

5. PROOFS OF THE MAIN RESULTS

In this section we prove Theorem 3.1 and Corollary 3.1. First (Definition 5.1), we introduce a family of (vectors of) c.d.f.s $Z^{F,\mathcal{A}}$ which belong to the invariant manifold (Lemma 5.1) and, moreover, any invariant state with frontiers F and null set $\mathcal{N}$ can be represented as $Z^{F,\mathcal{N}}$ (Corollary 5.2). The proofs of Theorem 3.1 and Corollary 3.1 follow.

DEFINITION 5.1. Let $\mathcal{A}$ be a set such that $\{J_1+1,\dots,J\}\subseteq\mathcal{A}\subseteq\{1,\dots,J\}$, let $\mathcal{A}^c = \{1,\dots,J\}\setminus\mathcal{A}$ and let $F = (F_j)_{j\in\mathcal{A}^c}\in\mathbb{R}^{|\mathcal{A}^c|}$. We define

$$Z^{F,\mathcal{A}} = \big(Z_k^{F,\mathcal{A}}(\cdot)\big)_{k=1,\dots,K},$$

where for $s(k)\in\mathcal{A}$ we have $Z_k^{F,\mathcal{A}}(s)\equiv 0$ and for $s(k)\in\mathcal{A}^c$,

$$Z_k^{F,\mathcal{A}}(s) = \sum_{\mathbf{k}\in\tilde{\mathcal{C}}(k)} \alpha_{\mathbf{k}}[H_{b(\mathbf{k})}(F_{s(k)}) - H_{b(\mathbf{k})}(s\wedge \min_{i\in\mathcal{S}(\mathbf{k})\cap\mathcal{A}^c} F_i)]^+.$$

Also, let $\mathfrak{X}(F,\mathcal{A})$ be the vector of functions of the form (2.6) with $Z(t,s) = Z^{F,\mathcal{A}}(s)$ for all $t\geqslant 0$, $s\in\mathbb{R}$, and satisfying (2.9), (2.12), (2.21)–(2.23).

LEMMA 5.1. *For any F and $\mathcal{A}$ as above, $Z^{F,\mathcal{A}} \in \mathfrak{S}$. Moreover, for $j \in \mathcal{A}^c$, $j \notin \mathcal{N}\big(\mathfrak{X}(F,\mathcal{A})\big)$ if and only if*

$$F_j < \sup_{\mathbf{k}\in\overline{\mathcal{C}}(j)} \big(y^*_{b(\mathbf{k})} \wedge \min_{i\in\mathcal{S}(\mathbf{k})\cap\mathcal{A}^c} F_i\big), \tag{5.1}$$

and (5.1) *implies that F_j is the frontier of $\mathfrak{X}(F,\mathcal{A})$ at station j.*

P r o o f. By definition, $\mathfrak{X}(F,\mathcal{A})$ satisfies (2.9) and (2.12). It is easy to check that it is also continuous and satisfies (2.7), (2.8) and (2.10). Thus, to prove that $Z^{F,\mathcal{A}} \in \mathfrak{S}$, it suffices to show that $\mathfrak{X}(F,\mathcal{A})$ has the required nonnegativity and monotonicity properties, and satisfies (2.11). We will first check that $D_k(\cdot, s-\cdot)$ and $D_k(t,\cdot)$ are nondecreasing for each $k = 1,\ldots,K$. By (2.22), together with Definition 5.1,

$$\begin{aligned} D_k(t,s-t) &= \sum_{l=1}^{K} Q_{kl}\alpha_l \int_{s-t}^{s} G_l(\eta)d\eta \\ &+ \sum_{l:s(l)\in\mathcal{A}^c} Q_{kl} \sum_{\mathbf{l}\in\tilde{\mathcal{C}}(l)} \alpha_{\mathbf{l}}\big[H_{b(\mathbf{l})}\big((s-t)\vee F_{s(l)}\big) - H_{b(\mathbf{l})}\big(s\wedge \min_{i\in\mathcal{S}(\mathbf{l})\cap\mathcal{A}^c} F_i\big)\big]^+. \end{aligned} \tag{5.2}$$

By (2.1),

$$\sum_{l=1}^{K} Q_{kl}\alpha_l \int_{s-t}^{s} G_l(\eta)d\eta = \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1}p_{k_1k_2}\ldots p_{k_nk} \int_{s-t}^{s} G_{k_1}(\eta)d\eta, \tag{5.3}$$

where for $n=0$, $\alpha_{k_1}p_{k_1k_2}\ldots p_{k_nk}$ and G_{k_1} should be understood as α_k and G_k, respectively. Using (2.1) again and changing the order of summation, we get

$$\begin{aligned} &\sum_{l:s(l)\in\mathcal{A}^c} Q_{kl} \sum_{\mathbf{l}\in\tilde{\mathcal{C}}(l)} \alpha_{\mathbf{l}}\big[H_{b(\mathbf{l})}\big((s-t)\vee F_{s(l)}\big) - H_{b(\mathbf{l})}\big(s\wedge \min_{i\in\mathcal{S}(\mathbf{l})\cap\mathcal{A}^c} F_i\big)\big]^+ \\ &= \mathbb{I}_{\{s(k)\in\mathcal{A}^c\}} \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1}p_{k_1k_2}\ldots p_{k_nk} \\ &\qquad \times \big[H_{k_1}\big((s-t)\vee F_{s(k)}\big) - H_{k_1}\big(s\wedge \min_{i\in\{s(k_1),\ldots,s(k_n)\}\cap\mathcal{A}^c} F_i\big)\big]^+ \\ &\quad + \sum_{l_1:s(l_1)\in\mathcal{A}^c} \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1}p_{k_1k_2}\ldots p_{k_nl_1}p_{l_1k} \\ &\qquad \times \big[H_{k_1}\big((s-t)\vee F_{s(l_1)}\big) - H_{k_1}\big(s\wedge \min_{i\in\{s(k_1),\ldots,s(k_n)\}\cap\mathcal{A}^c} F_i\big)\big]^+ \\ &\quad + \sum_{l_1:s(l_1)\in\mathcal{A}^c} \sum_{l_2=1}^{K} \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1}p_{k_1k_2}\ldots p_{k_nl_1}p_{l_1l_2}p_{l_2k} \\ &\quad \times \big[H_{k_1}\big((s-t)\vee F_{s(l_1)}\big) - H_{k_1}\big(s\wedge \min_{i\in\{s(k_1),\ldots,s(k_n)\}\cap\mathcal{A}^c} F_i\big)\big]^+ + \ldots \end{aligned} \tag{5.4}$$

$$= \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1} p_{k_1 k_2} \cdots p_{k_n k}$$

$$\times \big\{ \mathbb{I}_{\{s(k)\in\mathcal{A}^c\}} \big[H_{k_1}\big((s-t)\vee F_{s(k)}\big) - H_{k_1}(s \wedge \min_{i\in\{s(k_1),\ldots,s(k_n)\}\cap\mathcal{A}^c} F_i)\big]^+$$

$$+ \mathbb{I}_{\{s(k_n)\in\mathcal{A}^c\}} \big[H_{k_1}\big((s-t)\vee F_{s(k_n)}\big) - H_{k_1}(s \wedge \min_{i\in\{s(k_1),\ldots,s(k_{n-1})\}\cap\mathcal{A}^c} F_i)\big]^+$$

$$+ \ldots + \mathbb{I}_{\{s(k_1)\in\mathcal{A}^c\}} \big[H_{k_1}\big((s-t)\vee F_{s(k_1)}\big) - H_{k_1}(s)\big]^+ \big\}.$$

Fix n, $k_1,\ldots,k_n$. If $\{s(k_1),\ldots,s(k_n),s(k)\}\cap\mathcal{A}^c = \emptyset$, then the corresponding sum in the curled brackets in (5.4) vanishes. If $\{s(k_1),\ldots,s(k_n),s(k)\}\cap\mathcal{A}^c \neq \emptyset$, let $\big(s(k_1),\ldots,s(k_n),s(k)\big)\cap\mathcal{A}^c = (j_1,\ldots,j_m)$, i.e., let $(j_1,\ldots,j_m)$ be the subsequence of the sequence $\big(s(k_1),\ldots,s(k_n),s(k)\big)$ obtained by deleting the servers belonging to $\mathcal{A}$. For $A,B\in\mathcal{B}(\mathbb{R})$, let $\mu_B^{k_1}(A) = \int_{A\cap B}\big(1-G_{k_1}(\eta)\big)d\eta$. Then, by (3.1), the sum in the curled brackets in (5.4) corresponding to n, $k_1,\ldots,k_n$ can be rewritten as

(5.5)
$$(\mu^{k_1}_{[F_{j_m},\min_{i=1,\ldots,m-1}F_{j_i}]} + \mu^{k_1}_{[F_{j_{m-1}},\min_{i=1,\ldots,m-2}F_{j_i}]} + \ldots + \mu^{k_1}_{[F_{j_1},\infty)})[s-t,s]$$
$$= \mu^{k_1}_{[\min_{i=1,\ldots,m}F_{j_i},\infty)}[s-t,s] = \int\limits_{(s-t)\vee\min_{i=1,\ldots,m}F_{j_i}}^{s\vee\min_{i=1,\ldots,m}F_{j_i}} \big(1-G_{k_1}(\eta)\big)d\eta,$$

where the first equality can be verified by induction. The equations (5.2)–(5.5) yield

$$(5.6)\quad D_k(t,s-t) = \sum_{n=0}^{\infty}\sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1}p_{k_1k_2}\cdots p_{k_nk}\big\{\int\limits_{s-t}^{s} G_{k_1}(\eta)d\eta$$
$$+ \mathbb{I}_{\{\{s(k_1),\ldots,s(k_n),s(k)\}\cap\mathcal{A}^c\neq\emptyset\}} \int\limits_{(s-t)\vee\min_{j\in\{s(k_1),\ldots,s(k_n),s(k)\}\cap\mathcal{A}^c}F_j}^{s\vee\min_{j\in\{s(k_1),\ldots,s(k_n),s(k)\}\cap\mathcal{A}^c}F_j} \big(1-G_{k_1}(\eta)\big)d\eta\big\}.$$

It is easy to check that for each n, $k_1,\ldots,k_n$, the corresponding sum in the curled brackets in (5.6) is nondecreasing in t and s. Thus, $D_k(t,s-t)$ is nondecreasing in both t and s. Consequently, by (2.7) and (2.9), all the coordinate functions of $A(t,s-t)$ and $T(t,s-t)$ are nondecreasing in both t and s. By (2.10), $Y_j(t,s)$ is nonincreasing in s for all j. The fact that $Z(t,s) = Z^{F,\mathcal{A}}(s)$ and $W(t,s)$ are nonnegative and nondecreasing in s follows from Definition 5.1 and (2.12). Also, $D(0,s)=0$ for all s by (5.2). Thus, $A(0,s)=0$, $T(0,s)=0$ and $Y(0,s)=0$ for all s by (2.7), (2.9) and (2.10). This, together with the monotonicity of $A(t,s-t)$, $D(t,s-t)$, $T(t,s-t)$ in t, shows that $A(t,s)\geqslant A(0,s-t)=0$, and similarly $D(t,s)\geqslant 0$, $T(t,s)\geqslant 0$ for all $t\geqslant 0$, $s\in\mathbb{R}$. Let $j=1,\ldots,J$. From (2.1) we have

$$\rho_j = \sum_{k\in\mathcal{C}(j)} m_k (Q\alpha)_k = \sum_{k\in\mathcal{C}(j)} m_k \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1} p_{k_1k_2}\ldots p_{k_nk}.$$

This, together with (2.9), (2.10) and (5.6), implies that

$$(5.7)\quad Y_j(t,s-t) = (1-\rho_j)t + \rho_j t - \sum_{k\in\mathcal{C}(j)} m_k D_k(t,s-t)$$

$$= (1-\rho_j)t + \sum_{k\in\mathcal{C}(j)} m_k \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1} p_{k_1k_2}\ldots p_{k_nk} \Big\{ \int_{s-t}^{s} \big(1-G_{k_1}(\eta)\big)d\eta$$

$$- \mathbb{I}_{\{\{s(k_1),\ldots,s(k_n),j\}\cap\mathcal{A}^c\neq\emptyset\}} \int_{(s-t)\vee\min_{l\in\{s(k_1),\ldots,s(k_n),j\}\cap\mathcal{A}^c} F_l}^{s\vee\min_{l\in\{s(k_1),\ldots,s(k_n),j\}\cap\mathcal{A}^c} F_l} \big(1-G_{k_1}(\eta)\big)d\eta\Big\}$$

$$= (1-\rho_j)t + \sum_{k\in\mathcal{C}(j)} m_k \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n=1}^{K} \alpha_{k_1} p_{k_1k_2}\ldots p_{k_nk}$$

$$\times \int_{(s-t)\wedge\min_{l\in\{s(k_1),\ldots,s(k_n),j\}\cap\mathcal{A}^c} F_l}^{s\wedge\min_{l\in\{s(k_1),\ldots,s(k_n),j\}\cap\mathcal{A}^c} F_l} \big(1-G_{k_1}(\eta)\big)d\eta$$

(recall that, by convention, the minimum taken over $\emptyset$ equals ∞). In particular, $Y_j(t,s-t)$ is nondecreasing in t, and thus $Y_j(t,s) \geqslant Y_j(0,s+t) = 0$ for every $t \geqslant 0$ and $s\in\mathbb{R}$.

We now turn to the verification of (2.11). For $j\in\mathcal{A}$ there is nothing to prove. Assume that $j\in\mathcal{A}^c$, $t\geqslant 0$, $s\in\mathbb{R}$ and $\sum_{k\in\mathcal{C}(j)} Z_k^{F,\mathcal{A}}(s-t) > 0$. Thus, $s-t > F_j$, so for every $n, k_1,\ldots,k_n$ we have $s-t > \min_{l\in\{s(k_1),\ldots,s(k_n),j\}\cap\mathcal{A}^c} F_l$. Moreover, $\rho_j = 1$, since $\mathcal{A}^c \subseteq \{1,\ldots,J_1\}$. Therefore, (5.7) implies that $Y_j(t,s-t) = 0$ and (2.11) holds. We have proved that $Z^{F,\mathcal{A}} \in \mathfrak{S}$.

The second claim of the lemma follows easily from the definition of the model $\mathfrak{X}(F,\mathcal{A})$. ∎

COROLLARY 5.1. $\mathfrak{S}^D \subseteq \mathfrak{S}$.

This follows from Lemma 5.1, together with the fact that for the set $\mathcal{A} = \{J_1+1,\ldots,J\}$, $Z^{F,\mathcal{A}} = Z^F$ and $\mathfrak{X}(F,\mathcal{A}) = \mathfrak{X}(F)$.

COROLLARY 5.2. *Let $\mathfrak{X}$ be an invariant EDF fluid model with $\mathcal{N} = \mathcal{N}(\mathfrak{X})$ and let $F = (F_j)_{j\notin\mathcal{N}}$ be the corresponding frontier. Then $\mathfrak{X} = \mathfrak{X}(F,\mathcal{N})$.*

P r o o f. By Lemmas 4.2 and 5.1, the invariant EDF fluid models $\mathfrak{X}$, $\mathfrak{X}(F,\mathcal{N})$ have the same frontiers $(F_j)_{j\notin\mathcal{N}}$ and the same null set $\mathcal{N}$. Thus, the equality $\mathfrak{X} = \mathfrak{X}(F,\mathcal{N})$ follows from Proposition 4.2. ∎

P r o o f o f T h e o r e m 3.1. In the strictly subcritical case, Theorem 3.1 follows immediately from Corollary 4.1. Assume that the network is not strictly subcritical. By Corollary 5.1, it suffices to prove that $\mathfrak{S} \subseteq \mathfrak{S}^D$. Let $\mathfrak{X}$ be an invariant EDF fluid model of the form (2.6), let $\mathcal{N} = \mathcal{N}(\mathfrak{X})$ and let $F = (F_j)_{j \notin \mathcal{N}}$ be the corresponding frontier. If $\mathcal{N} = \{J_1 + 1, \ldots, J\}$, then $Z \in \mathfrak{S}^D$ by Proposition 4.3 and Corollary 5.2. Assume that $\mathcal{N}_0 = \mathcal{N} \setminus \{J_1 + 1, \ldots, J\} \neq \emptyset$. Let $j_0 = \min\{j : j \in \mathcal{N}_0\}$, $\mathcal{N}_1 = \mathcal{N} \setminus \{j_0\}$, and let

$$F_{j_0} = \inf\{a \in \mathbb{R} : Z^{F(a),\mathcal{N}_1} = Z\}, \tag{5.8}$$

where $F(a)_j = F_j$ for $j \notin \mathcal{N}$ and $F(a)_{j_0} = a$. The set in (5.8) is nonempty since, by Lemma 5.1, for $a_0 = \max_{k=1,\ldots,K} y_k^*$, $\mathcal{N}\big(\mathfrak{X}\big(F(a_0), \mathcal{N}_1\big)\big) = \mathcal{N}$ and the frontiers of $\mathfrak{X}$ and $\mathfrak{X}\big(F(a_0), \mathcal{N}_1\big)$ coincide, so $Z^{F(a_0),\mathcal{N}_1} = Z$ by Proposition 4.2. It is also bounded below, because Lemma 5.1 implies that $j_0 \notin \mathcal{N}\big(\mathfrak{X}\big(F(a_0), \mathcal{N}_1\big)\big)$ for $a < (\min_{k=1,\ldots,K} y_k^*) \wedge (\min_{i \notin \mathcal{N}} F_i)$. From the definition of $Z^{F,\mathcal{A}}$ it is easy to see that the infimum in (5.8) is attained, i.e., $Z^{F(F_{j_0}),\mathcal{N}_1} = Z$. Thus, by Lemma 5.1, for $j \notin \mathcal{N}$,

$$F_j < \sup_{\mathbf{k} \in \overline{\mathcal{C}}(j)} (y_{b(\mathbf{k})}^* \wedge \min_{i \in \mathcal{S}(\mathbf{k}) \setminus \mathcal{N}_1} F_i).$$

Let $\epsilon > 0$ be so small that for $j \neq \mathcal{N}$,

$$F_j^\epsilon < \sup_{\mathbf{k} \in \overline{\mathcal{C}}(j)} (y_{b(\mathbf{k})}^* \wedge \min_{i \in \mathcal{S}(\mathbf{k}) \setminus \mathcal{N}_1} F_i^\epsilon), \tag{5.9}$$

where $F_j^\epsilon = F_j$ for $j \notin \mathcal{N}$ and $F_{j_0}^\epsilon = F_{j_0} - \epsilon$. Let $F^\epsilon = (F_j^\epsilon)_{j \notin \mathcal{N}_1}$ and let $\mathfrak{X}^{(1)} = \mathfrak{X}(F^\epsilon, \mathcal{N}_1)$, $Z^{(1)} = Z^{F^\epsilon,\mathcal{N}_1}$. By the definition of $\mathfrak{X}^{(1)}$, together with (5.9) and Lemma 5.1, we have $\mathcal{N}_1 \subseteq \mathcal{N}(\mathfrak{X}^{(1)}) \subseteq \mathcal{N}$ and the frontier of the invariant EDF fluid model $\mathfrak{X}^{(1)}$ at station $j \notin \mathcal{N}$ equals $F_j^\epsilon = F_j$. If $\mathcal{N}(\mathfrak{X}^{(1)}) = \mathcal{N}$, then, by Proposition 4.2, $\mathfrak{X}^{(1)} = \mathfrak{X}$, which contradicts the definition (5.8) of F_{j_0}. Thus, $\mathcal{N}(\mathfrak{X}^{(1)}) = \mathcal{N}_1$. If $\mathcal{N}_1 = \{J_1 + 1, \ldots, J\}$, then $Z^{(1)} \in \mathfrak{S}^D$ by Proposition 4.3 and Corollary 5.2. If this is not the case, we repeat the above construction with $\mathfrak{X}^{(1)}$ instead of $\mathfrak{X}$ and iterate until we get $\mathfrak{X}^{(n)}$ with $Z^{(n)} \in \mathfrak{S}^D$. Moreover, it is easy to see that for any $\eta > 0$ we have $\rho_K(Z, Z^{(n)}) \leqslant \eta$ if we take ϵ small enough in all the steps of the iteration. Therefore, $\rho(Z, \mathfrak{S}^D) \leqslant \eta$ for every $\eta > 0$, so $Z \in \overline{\mathfrak{S}^D}$. However, it is easy to check that $\mathfrak{S}^D$ is closed in $\mathcal{M}^K$, and thus $Z \in \mathfrak{S}^D$. ■

P r o o f o f C o r o l l a r y 3.1. Let $\mathfrak{X}$ be an invariant EDF fluid model of the form (2.6). By Theorem 3.1, there exists $F \in D$ such that $\mathfrak{X} = \mathfrak{X}(F)$. In particular, for $k = 1, \ldots, K$,

$$z_k = \Psi_k(F). \tag{5.10}$$

By (2.12) and (5.10), for $j = 1, \ldots, J_1$,

$$w_j = \sum_{k \in \mathcal{C}(j)} m_k z_k = \sum_{k \in \mathcal{C}(j)} m_k \Psi_k(F) = \Phi_j(F),$$

i.e., $w = \Phi(F)$. Thus, by Proposition 3.1,

$$F = \Phi^{-1}(w), \tag{5.11}$$

so, by (5.10), $z = \Psi \circ \Phi^{-1}(w)$. Uniqueness of the invariant EDF fluid model with workload w_j at station $j = 1, \ldots, J_1$ follows from (5.11) and the fact that, by Theorem 3.1, $F \in D$ determines the corresponding invariant state $\mathfrak{X}(F)$ uniquely. ■

REFERENCES

[1] M. Bramson, *Convergence to equilibria for fluid models of FIFO queueing networks*, Queueing Syst. 22 (1996), pp. 5–45.

[2] M. Bramson, *Convergence to equilibria for fluid models of head-of-the-line proportional processor sharing queueing networks*, Queueing Syst. 23 (1996), pp. 1–26.

[3] M. Bramson, *State space collapse with application to heavy traffic limits for multiclass queueing networks*, Queueing Syst. 30 (1998), pp. 89–148.

[4] M. Bramson, *Stability of earliest-due-date, first-served queueing networks*, Queueing Syst. 39 (2001), pp. 79–102.

[5] M. Bramson and J. G. Dai, *Heavy traffic limit theorems for some queueing networks*, Ann. Appl. Probab. 11 (2001), pp. 49–90.

[6] J. G. Dai, *On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models*, Ann. Appl. Probab. 5 (1995), pp. 49–77.

[7] B. Doytchinov, J. P. Lehoczky and S. E. Shreve, *Real-time queues in heavy traffic with earliest-deadline-first queue discipline*, Ann. Appl. Probab. 11 (2001), pp. 332–379.

[8] H. C. Gromoll, *Diffusion approximation for a processor sharing queue in heavy traffic*, Ann. Appl. Probab. 14 (2004), pp. 555–611.

[9] H. C. Gromoll, A. L. Puha and R. J. Williams, *The fluid limit of a heavily loaded processor sharing queue*, Ann. Appl. Probab. 12 (2002), pp. 797–859.

[10] J. M. Harrison, *Balanced fluid models of multiclass queueing networks: a heavy traffic conjecture*, in: *Stochastic Networks*, IMA Vol. Math. Appl. 71, Springer, New York 1995, pp. 1–20.

[11] J. M. Harrison and R. J. Williams, *A multiclass closed queueing network with unconventional heavy traffic behavior*, Ann. Appl. Probab. 6 (1996), pp. 1–47.

[12] W. Hopp and M. Spearman, *Factory Physics: Foundations of Manufacturing Management*, Irwin, Chicago 1996.

[13] Ł. Kruk, *Stability of two families of real-time queueing networks*, Probab. Math. Statist. 28 (2008), pp. 179–202.

[14] Ł. Kruk, *An open queueing network with asymptotically stable fluid model and unconventional heavy traffic behavior*, preprint.

[15] Ł. Kruk, J. P. Lehoczky, K. Ramanan and S. E. Shreve, *Heavy traffic analysis for EDF queues with reneging*, Ann. Appl. Probab., to appear.

[16] Ł. Kruk, J. P. Lehoczky, S. E. Shreve and S.-N. Yeung, *Multiple-input heavy-traffic real-time queues*, Ann. Appl. Probab. 13 (2003), pp. 54–99.

[17] Ł. Kruk, J. P. Lehoczky, S. E. Shreve and S.-N. Yeung, *Earliest-deadline-first service in heavy traffic acyclic networks*, Ann. Appl. Probab. 14 (2004), pp. 1306–1352.

[18] A. L. Puha and R. J. Williams, *Invariant states and rates of convergence for a critical fluid model of a processor sharing queue*, Ann. Appl. Probab. 14 (2004), pp. 517–554.

[19] J. A. Stankovic, M. Spuri, K. Ramamritham and G. C. Buttazzo, *Deadline Scheduling for Real-Time Systems*, Springer, 1998.
[20] R. J. Williams, *An invariance principle for semimartingale reflecting Brownian motions in an orthant*, Queueing Syst. 30 (1998), pp. 5–25.
[21] R. J. Williams, *Diffusion approximations for open multiclass queueing networks*: *sufficient conditions involving state space collapse*, Queueing Syst. 30 (1998), pp. 27–88.
[22] S.-N. Yeung and J. P. Lehoczky, *Real-time queueing networks in heavy traffic with EDF and FIFO queue discipline*, working paper, 2001, Department of Statistics, Carnegie Mellon University.

Institute of Mathematics
Maria Curie-Skłodowska University
pl. Marii Curie-Skłodowskiej 1
20-031 Lublin, Poland
E-mail: lkruk@hektor.umcs.lublin.pl

Received on 9.3.2009;
revised version on 18.6.2010