The Annals of Applied Probability
2010, Vol. 20, No. 3, 1126–1176
DOI: 10.1214/09-AAP649

NETWORK STABILITY UNDER MAX–MIN FAIR BANDWIDTH SHARING

BY MAURY BRAMSON[1]

University of Minnesota

There has recently been considerable interest in the stability of different fair bandwidth sharing policies for models that arise in the context of Internet congestion control. Here, we consider a connection level model, introduced by Massoulié and Roberts [*Telecommunication Systems* **15** (2000) 185–201], that represents the randomly varying number of flows present in a network. The weighted α-fair and weighted max–min fair bandwidth sharing policies are among important policies that have been studied for this model. Stability results are known in both cases when the interarrival times and service times are exponentially distributed. Partial results for general service times are known for weighted α-fair policies; no such results are known for weighted max–min fair policies. Here, we show that weighted max–min fair policies are stable for subcritical networks with general interarrival and service distributions, provided the latter have $2 + \delta_1$ moments for some $\delta_1 > 0$. Our argument employs an appropriate Lyapunov function for the weighted max–min fair policy.

1. Introduction. We consider a connection level model for Internet congestion control that was first studied by Massoulié and Roberts [9]. This stochastic model represents the randomly varying number of flows in a network for which bandwidth is dynamically shared among flows that correspond to the transfer of documents along specified routes. Standard bandwidth sharing policies are the weighted α-fair, $\alpha \in (0, \infty)$, and the weighted max–min fair policies. An important example of the former is the proportionally fair policy, which corresponds to $\alpha = 1$. The weighted max–min fair policy corresponds to $\alpha = \infty$. These policies allocate service uniformly to documents along a given route, and allocate service amongst different routes in a "fair" manner. A question of considerable interest is when such policies are stable.

De Veciana, Lee and Konstantopoulos [5] studied the stability of weighted max–min fair and proportionally fair policies; Bonald and Massoulié [2] studied the stability of weighted α-fair policies. Both papers assumed exponentially distributed interarrival and service times for documents. The first condition is equivalent to Poisson arrivals, and the second condition corresponds to exponentially distributed document sizes with documents processed at a constant rate. Both papers

Received December 2008; revised July 2009.
[1]Supported in part by NSF Grants DMS-02-26245 and CCF-0729537.
AMS 2000 subject classifications. 60K25, 68M20, 90B15.
Key words and phrases. Bandwidth sharing, max–min fair, stability.

constructed Lyapunov functions which imply the stability of such models when the models are subcritical, that is, the underlying Markov process is positive Harris recurrent when the average load at each link is less than its capacity.

Relatively little is currently known regarding the stability of subcritical networks with general interarrival and service times. Massoulié [8] showed stability for the proportionally fair policy for exponentially distributed interarrival times and general service times that are of phase type. A suitable Lyapunov function was employed to show stability.

The stability problem for bandwidth sharing policies is in certain aspects similar to the analogous problem for multiclass queueing networks. A significant complication that arises in the context of bandwidth sharing policies is the requirement of simultaneous service of documents at all links along a route. This can reduce the efficiency of service, and complicates analysis when the interarrival and service times are not exponentially distributed.

When the interarrival and service times are exponentially distributed, finer results are possible. In Kang et al. [7], a diffusion approximation is established under weighted proportionally fair policies. There and in Gromoll and Williams [6], summaries and a more detailed bibliography are provided for different bandwidth sharing policies, for both exponentially distributed and more general interarrival and service times.

Here, we investigate the behavior of weighted max–min fair policies for subcritical networks whose interarrival and service times have general distributions. We show that such networks are stable, provided that the service distributions have $2 + \delta_1$ moments for some $\delta_1 > 0$. No conclusion is reached when fewer moments exist. As in previous papers on stability, we construct a suitable Lyapunov function. Because of the more general framework here, the Markov process underlying the model will now have a general state space, and will require the machinery associated with positive Harris recurrence.

We next give a more detailed description of the model we consider, after which we state our main results. We then provide some basic motivation behind their proof together with a summary of the remainder of the paper.

Description of the model. In the model we consider, *documents* are assumed to arrive at one of a finite number of *routes* $r \in \mathcal{R}$ according to independent renewal processes, with interarrival times denoted by $\xi_r(1), \xi_r(2), \dots$. Here, $\xi_r(1)$ are the initial residual interarrival times, and are considered part of the initial state. The remaining variables $\xi_r(2), \xi_r(3), \dots$ are assumed to be i.i.d. with mean $1/\nu_r$, $\nu_r > 0$, for each r, with the sequences being independent of one another; ξ_r will denote a random variable with the corresponding distribution. The service times of documents are assumed to be independent of one another and of the interarrival times, and have distribution functions $H_r(\cdot)$ with means $m_r < \infty$. The initial state will include the residual service times of documents initially in the network.

On each route r, there are a finite number of links l, where service is allocated to the documents on the route. For the models considered in [2, 5, 8] and [9], documents on a route r receive service simultaneously at all links l on the route, with all such documents being allocated the same rate of service λ_r at all such links at a given time. Associated with such a network is an *incidence matrix* $A = (A_{l,r})$, $l \in \mathcal{L}$, $r \in \mathcal{R}$, with $A_{l,r} = 1$ if link l lies on route r and $A_{l,r} = 0$ otherwise. When $A_{l,r} = A_{l,r'} = 1$, with $r \neq r'$, the routes r and r' share a common link.

Setting z_r equal to the number of documents on route r, $\Lambda_r = \lambda_r z_r$ denotes the rate of service allocated to the totality of all documents on the route. Each link l is assumed to have a given *bandwidth capacity* $c_l > 0$. A *feasible policy* requires that this capacity not be exceeded, namely

$$\sum_{r \in \mathcal{R}} A_{l,r} \Lambda_r \leq c_l \qquad \text{for all } l \in \mathcal{L}. \tag{1.1}$$

Denoting by $\Lambda = (\Lambda_r)$ and $c = (c_l)$ the corresponding column vectors, this is equivalent to $A\Lambda \leq c$, with the inequality being interpreted coordinatewise.

None of the results in this paper relies on the restriction that either $A_{l,r} = 1$ or $A_{l,r} = 0$. Here, we continue to assume that (1.1) is satisfied, for given A, but with the weaker assumption $A_{l,r} \geq 0$. Under this new setup, each link may be interpreted as belonging to every route. A given link l now allocates the same rate of service Λ_r to each route r, which utilizes this service at rate $A_{l,r}$. For $A_{l,r} \in [0, 1]$, $A_{l,r}$ may be interpreted as the proportion of this potential service that is actually utilized at link l by route r.

The *traffic intensity* $\rho_r = \nu_r m_r$ measures the average rate over time at which work enters a route r. We say a network is *subcritical* if

$$\sum_{r \in \mathcal{R}} A_{l,r} \rho_r < c_l \qquad \text{for all } l \in \mathcal{L}, \tag{1.2}$$

or, in matrix form, $A\rho < c$, where $\rho = (\rho_r)$ is the corresponding column vector. This corresponds to the definition of subcriticality that is employed in the context of queueing networks, where the load at each station (here, load at each link) is strictly less than its capacity. Condition (1.2) is needed for stability. It is assumed in, for example, [2, 5] and [8].

The α-fair and max–min fair policies are examples of feasible policies for which the allocation of service to documents at a given time is determined by the vector $z = (z_r)$; the weighted α-fair and max–min fair policies are defined analogously, but with a weight $w_r > 0$ assigned to route r. We do not define α-fair here, or, in particular, proportionally fair, referring the reader to the previous references. *Weighted max–min fair* (WMMF) is defined as a feasible policy that, at each time, allocates service so that

$$\min_{r \in \mathcal{R}'} \{\lambda_r / w_r\} \text{ is maximized,} \tag{1.3}$$

among nonempty routes $\mathcal{R}'$. That is, the minimum amount of weighted service each document receives is maximized, on r with $z_r > 0$, subject to the constraint (1.1).

As defined above, a WMMF policy always exists, although it need not be unique, since there may be some flexibility in allocating service among those routes where documents are receiving more than the minimal amount of service. Since our results apply to all such policies, we will not bother to select a "best" member that, for instance, maximizes service on the routes that are already receiving more than the minimum service. Such a "best" policy can be obtained by solving a hierarchy of optimization problems, as mentioned above display (2) in [5]. [By employing the convexity that is inherent in the constraint (1.1), it is routine to verify the existence of such policies.]

Since the vector z of documents changes as time evolves, so will the allocation of service. From this point on, we reserve the notation $\lambda_r(t)$ and $\Lambda_r(t)$ for the allocation of service for a WMMF policy at time t. We find it useful to also introduce

$$\lambda^w(t) = \min_{r \in \mathcal{R}'} \{\lambda_r(t)/w_r\} \tag{1.4}$$

with (1.3) in mind. Between arrivals and departures of documents, $\lambda(\cdot) = (\lambda_r(\cdot))$ and $\lambda^w(\cdot)$ will be constant; we specify that they be right continuous with left limits.

The state of the network evolves over time as documents arrive in the network, are served, and then depart. For networks with exponentially distributed interarrival and service times and an assigned policy, $z = (z_r)$ suffices to describe its state. As with queueing networks, one needs to specify the residual interarrival and service times in general. With this in mind, we employ the notation $z_r(B_r)$ to denote the number of documents on route r that have residual service times in $B_r \subseteq \mathbb{R}^+$, and u_r to denote the residual interarrival time for r, with $z(B) = (z_r(B_r))$, $B = (B_r)$, and $u = (u_r)$ denoting the corresponding vectors. Setting

$$x = (z(\cdot), u), \tag{1.5}$$

the state x contains this information. We will employ $X(t)$, $Z(t, \cdot)$ and $U(t)$ for the corresponding random states at time t. The natural metric space S that corresponds to the states x is no longer discrete. We will describe S in more detail in Section 2.

One can specify a Markov process $X(\cdot)$ on S that corresponds to the network with the assigned WMMF policy. The process $X(\cdot)$ is constructed in the same manner as is its analog for a queueing network. More detail is again given in Section 2. We note here that since S is not discrete, the notion of positive recurrence needs to be replaced by that of positive Harris recurrence. When $X(\cdot)$ is positive Harris recurrent, we will say that the network is *stable*.

In order to demonstrate positive Harris recurrence for $X(\cdot)$, we will define, in Section 3, an appropriate nonnegative function, or *norm*, $\|x\|$, for $x \in S$. It is defined in terms of the norms $|x|_L$, $|x|_R$ and $|x|_A$, by

$$\|x\| = |x|_L + |x|_R + |x|_A. \tag{1.6}$$

Without going into detail here, we note that $|x|_L$ and $|x|_R$ are defined from $z(\cdot)$, where $|x|_L$, in essence, measures residual service times smaller than N, for a given large N, $|x|_R$ measures residual service times greater than N, and $|x|_A$ is a function of the largest residual interarrival time. (When a distribution function H_r has a thin enough tail, we actually replace N by a smaller value N_{H_r} that depends on H_r.) As one should expect, as either the total number of documents $\sum_r z_r \to \infty$ or $|u| \to \infty$, then $\|x\| \to \infty$.

Main results. We now state our two main results.

THEOREM 1.1. *Suppose that a subcritical network with a weighted max–min fair policy has interarrival times with finite means and service times with* $2 + \delta_1$ *moments*, $\delta_1 > 0$. *For the norm in* (1.6), *and appropriate* N, L *and* $\varepsilon_1 > 0$,

$$E_x[\|X(N^3)\|] \le (\|x\| \vee L) - \varepsilon_1 N^2 \qquad \textit{for all } x \in S. \tag{1.7}$$

Inequality (1.7) states that, for large $\|x\|$, $X(\cdot)$ has an average negative drift over $[0, N^3]$ that is at least of order $1/N$. This rate will be a consequence of the application of N in the construction of the norm $|x|_L$ that appears in (1.6).

The reader will recognize (1.7) as a version of Foster's criterion. It will imply the positive Harris recurrence of $X(\cdot)$, provided that the states in S communicate with one another in an appropriate sense. Petite sets are typically employed for this purpose; they will be defined in Section 2. A petite set A has the property that each measurable set B is "equally accessible" from all points in A with respect to a given measure.

THEOREM 1.2. *Suppose that a subcritical network with a weighted max–min fair policy has interarrival times with finite means and service times with* $2 + \delta_1$ *moments*, $\delta_1 > 0$. *Also, suppose that* $A_L = \{x : \|x\| \le L\}$ *is petite for each* $L > 0$, *for the norm in* (1.6). *Then*, $X(\cdot)$ *is positive Harris recurrent.*

Theorem 1.2 will follow from Theorem 1.1 by standard reasoning. More detail is given in Section 2.

A standard criterion that ensures the above sets A_L are petite is given by the following two conditions on the interarrival times. The first condition is that the distribution of $\xi_r(2)$ is unbounded for all r, that is,

$$P(\xi_r(2) \ge s) > 0 \qquad \text{for all } s. \tag{1.8}$$

The second condition is that, for some $l_r \in \mathbb{Z}^+$, the $(l_r - 1)$-fold convolution of $\xi_r(2)$ and Lebesque measure are not mutually singular. That is, for some nonnegative $q_r(\cdot)$ with $\int_0^\infty q_r(s)\,ds > 0$,

$$P\big(\xi_r(2) + \cdots + \xi_r(l_r) \in [c, d]\big) \geq \int_c^d q_k(s)\,ds \tag{1.9}$$

for all $c < d$. When the interarrival times are exponentially distributed, both (1.8) and (1.9) are immediate. More detail is given in Section 2.

We therefore have the following corollary of Theorem 1.2.

COROLLARY 1.1. *Suppose that a subcritical network with a weighted max–min fair policy has interarrival times with finite means that satisfy* (1.8) *and* (1.9), *and service times with* $2 + \delta_1$ *moments*, $\delta_1 > 0$. *Then*, $X(\cdot)$ *is positive Harris recurrent.*

Outline of the paper and main ideas. In Section 2, we will provide a brief background of Markov processes and will summarize the construction of the space S and Markov process $X(\cdot)$ described above. We will also provide background that will be employed to derive Theorem 1.2 from Theorem 1.1 and to obtain Corollary 1.1. The machinery for this is standard in the context of queueing networks; we explain there the needed modifications.

The remainder of the paper is devoted to the demonstration of Theorem 1.1. (One minor result, Proposition 3.1, is needed for Theorem 1.2.) In Section 3, we will specify the norms $|\cdot|_L$, $|\cdot|_R$ and $|\cdot|_A$ that define $\|\cdot\|$ in (1.6). Employing bounds on these three norms that will be derived in Sections 4, 5 and 10, we obtain the conclusion (1.7) of Theorem 1.1.

For large $\|x\|$, at least one of the norms $|x|_V$, with V equal to L, R or A, must also be large. When $|x|_V$ is large for given V, it will follow that $E_x[|X(N^3)|_V] - |x|_V$ is sufficiently negative so that (1.7) will hold.

The analysis for $|\cdot|_A$ is straightforward and is done in Section 4. The behavior of $E_x[|X(N^3)|_R] - |x|_R$ is analyzed in Section 5. The remaining five sections are devoted to analyzing $E_x[|X(N^3)|_L] - |x|_L$. In the last two cases, one needs to reason that, in an appropriate sense, the decrease in residual service times of existing documents more than compensates for the increase due to arriving documents, thus producing a net negative drift.

For such an analysis, it makes sense to decompose the process $X(\cdot)$ into processes $\tilde{X}(\cdot)$ and $X^A(\cdot)$, with

$$X(t) = \tilde{X}(t) + X^A(t) \qquad \text{for all } t.$$

The process $\tilde{X}(t)$ is obtained from $X(t)$ by retaining only those documents, the *original documents*, that were initially in the network, and $X^A(t)$ consists of the remaining documents. Neither $\tilde{X}(\cdot)$ nor $X^A(t)$ is Markov. One defines $\tilde{Z}(t, B)$ and $Z^A(t, B)$ analogously to $Z(t, B)$.

Because of the WMMF policy, all documents that remain on a route r, over the time interval $[0, t]$, receive the same service $\Delta_r(t)$, with $\Delta_r(t) = \int_0^t \lambda_r(t')\,dt'$. Consequently,

$$\tilde{Z}_r(t, B) = z_r\big(B + \Delta_r(t)\big) \qquad \text{for } t \geq 0, r \in \mathcal{R}, B \subseteq \mathbb{R}^+. \tag{1.10}$$

The norms $|\cdot|_L$ and $|\cdot|_R$ will be defined so that documents with greater residual service times contribute more heavily to the norms. On account of (1.10), $|\tilde{X}(t)|_L$ and $|\tilde{X}(t)|_R$ will therefore decrease over time; one can also obtain upper bounds on $|X^A(t)|_L$ and $|X^A(t)|_R$. One can use this to obtain a negative net drift on $E_x[|X(N^3)|_L] - |x|_L$ and $E_x[|X(N^3)|_R] - |x|_R$, as mentioned earlier.

Only limited use of inequalities arising from (1.10) is needed in Section 5 for $|\cdot|_R$. More detailed versions are needed for $|\cdot|_L$, which are presented in the first part of Section 6.

In Section 6, we also introduce the sets $\mathcal{A}(t)$, along which we will be able to obtain good pathwise upper bounds on $|X^A(t)|_L$. We show in Section 6, by using elementary large deviation estimates, that the probabilities of the complements $\mathcal{A}(t)^c$ are small enough so that

$$E_x[|X(N^3)|_L - |x|_L; \mathcal{A}(N^3)^c]$$

is negligible with respect to $E_x[|X(N^3)|_L] - |x|_L$.

Sections 7–10 analyze the behavior of $|X(N^3)|_L$ on $\mathcal{A}(N^3)$. Section 7 considers the contribution to $|X(N^3)|_L$ of residual times $s > N_{H_r}$; N_{H_r} was mentioned parenthetically after (1.6) and satisfies $N_{H_r} \leq N$. Sections 8 and 9 consider the contribution to $|X(N^3)|_L$ of residual times $s \leq N_{H_r}$. In Section 8, this is done for $\Delta_r(N^3) > 1/b^3$, for given r, with the constant b introduced in (3.3). Here, service of individual documents is intense enough to provide straightforward upper bounds for $|X(N^3)|_L - |x|_L$.

Section 9 considers the case with $\Delta_r(N^3) \leq 1/b^3$. This is the only place in the paper where the subcriticality of the network is employed; estimation for $|X(N^3)|_L$ must therefore be more precise. The short Section 10 combines the results of Sections 6–9 to give the desired bounds on $E_x[|X(N^3)|_L] - |x|_L$.

Notation. For the reader's convenience, we list here some of the notation in the paper, part of which has already been employed. We set $\bar{H}_r(s) = 1 - H_r(s)$; quantities such as $\bar{H}_r^*(s)$ and $\bar{\Phi}_r^*(s)$, are defined analogously in terms of $H_r^*(s)$ and $\Phi_r^*(s)$, which will be introduced later on. The term x indicates a state in S and the corresponding term $X(t)$ indicates a random state at time t; $z(\cdot)$ and $Z(t, \cdot)$, and u and $U(t)$ play analogous roles. We will abbreviate $\Delta_r = \Delta_r(N^3)$ and set $i_r(s) = s + \Delta_r$; $i_r(s)$ is the initial residual service time of an original document that has residual service time s at time N^3. We employ $C_1, C_2, \ldots$ and $\varepsilon_1, \varepsilon_2, \ldots$ for different positive constants that appear in our bounds, whose precise values are unimportant. The symbols $\mathbb{Z}^+$ and $\mathbb{R}^+$ denote the positive integers and positive real numbers, and $\mathbb{Z}^{+,0} = \mathbb{Z}^{+} \cup \{0\}$; $\lfloor y \rfloor$ and $\lceil y \rceil$ denote the integer part of $y \in \mathbb{R}^{+}$ and the smallest integer n with $n \geq y$; and $c \vee d$ and $c \wedge d$ denote the greater and smaller value of $c, d \in \mathbb{R}$. The acronyms LHS and RHS will stand for "left-hand side" and "right-hand side" when referring to equations or inequalities. Since the paper is devoted to demonstrating Theorems 1.1 and 1.2, we will implicitly assume that the network under consideration has a WMMF policy, except when stated otherwise, and that the moment conditions on the interarrival and service times given in Theorem 1.1 hold. We assume the network is subcritical only when explicitly stated.

2. Markov process background. In this section, we provide a more detailed description of the construction of the Markov process $X(\cdot)$ that underlies a WMMF network. We then show how Theorem 1.2 and its corollary follow from Theorem 1.1. Analogs of this material for queueing networks are given in Bramson [1]. Because of the similarity of the two settings, we present a summary here and refer the reader to [1] for additional detail.

Construction of the Markov process. As in (1.5), we define the state space S to be the set of pairs $x = (z(\cdot), u)$, where $z(\cdot) = (z_r(\cdot))$ and $z_r(\cdot)$ is a counting measure that maps $B \subseteq \mathbb{R}^{+}$ to $\mathbb{Z}^{+,0}$, and $u = (u_r)$, $r \in \mathcal{R}$, has positive components. Here, $z(\cdot)$ corresponds to the residual service times of documents and u to the residual interarrival times. (One could, as in (4.1) of [1], distinguish documents based on their "age," which is not needed here.)

For the purpose of constructing a metric $d(\cdot,\cdot)$ on S, we assign to each document the pair (r_i, s_i), $i = 1, 2, \ldots,$ where $r_i \in \{1, \ldots, |\mathcal{R}|\}$ denotes its route and $s_i > 0$ its residual service time. Documents are ordered so that $s_1 \leq s_2 \leq \cdots$, with the decision for ties being made based on a given ordering of the routes. When i exceeds the number of documents belonging to x, we assign the value $(r_i, s_i) = (0, 0)$. For $x, x' \in S$, with the coordinates labeled correspondingly, we set

$$d(x, x') = \sum_{i=1}^{\infty} \bigl((|r_i - r_i'| + |s_i - s_i'|) \wedge 1\bigr) + \sum_{r} |u_r - u_r'|. \tag{2.1}$$

One can check that $d(\cdot,\cdot)$ is separable and locally compact. (See page 82 of [1] for details.) We equip S with the standard Borel σ-algebra inherited from $d(\cdot,\cdot)$, which we denote by $\mathscr{S}$. In Proposition 3.1, we will show $|\cdot|_L$, $|\cdot|_R$ and $|\cdot|_A$ are continuous in $d(\cdot,\cdot)$.

The Markov process $X(t) = (Z(t,\cdot), U(t))$ underlying the network, with $Z(t,\cdot)$ and $U(t)$ taking values $z(\cdot)$ and u as above, is defined to be the right continuous process whose evolution is determined by the assigned WMMF policy. Documents are allocated service according to the rates $\lambda_r(\cdot)$, which are constant in between arrivals and departures of documents on routes. Upon an arrival or departure, rates are re-assigned according to the policy. We note that this procedure is not policy specific, and also applies to α-fair policies. By modifying the state space descriptor to contain more information, one could also include more general networks.

Along the lines of page 85 of [1], a filtration $(\mathcal{F}_t)$, $t \in [0, \infty]$, can be assigned to $X(\cdot)$ so that $X(\cdot)$ is Borel right and, in particular, is strong Markov. The processes $X(\cdot)$ fall into the class of piecewise-deterministic Markov processes, for which the reader is referred to Davis [4] for more detail.

Recurrence. The Markov process $X(\cdot)$ is said to be *Harris recurrent* if, for some nontrivial σ-finite measure φ,

$$\varphi(B) > 0 \text{ implies } P_x(\eta_B = \infty) = 1 \qquad \text{for all } x \in S,$$

where $\eta_B = \int_0^\infty 1\{X(t) \in B\}\, dt$. If $X(\cdot)$ is Harris recurrent, it possesses a stationary measure π that is unique up to a constant multiple. When π is finite, $X(\cdot)$ is said to be *positive Harris recurrent*.

A practical condition for determining positive Harris recurrence can be given by using petite sets. A nonempty set $A \in \mathscr{S}$ is said to be *petite* if for some fixed probability measure a on $(0, \infty)$ and some nontrivial measure ν on $(S, \mathscr{S})$,

$$\nu(B) \le \int_0^\infty P^t(x, B) a(dt)$$

for all $x \in A$ and $B \in \mathscr{S}$. Here, $P^t(\cdot, \cdot)$, $t \ge 0$, is the semigroup associated with $X(\cdot)$. As mentioned in the Introduction, a petite set A has the property that each set B is "equally accessible" from all points $x \in A$ with respect to the measure ν. Note that any nonempty measurable subset of a petite set is also petite.

For given $\delta > 0$, set

$$\tau_B(\delta) = \inf\{t \ge \delta : X(t) \in B\}$$

and $\tau_B = \tau_B(0)$. Then, $\tau_B(\delta)$ is a stopping time. Employing petite sets and $\tau_B(\delta)$, one has the following characterization of Harris recurrence and positive Harris recurrence. (The Markov process and state space need to satisfy minimal regularity conditions, as on page 86 of [1].) The criteria are from Meyn and Tweedie [10]; discrete time analogs of the different parts of the proposition have long been known; see, for instance, Nummelin [11] and Orey [12].

THEOREM 2.1. (a) *A Markov process $X(\cdot)$ is Harris recurrent if and only if there exists a closed petite set A with*

$$P_x(\tau_A < \infty) = 1 \qquad \textit{for all } x \in S. \tag{2.2}$$

(b) *Suppose the Markov process $X(\cdot)$ is Harris recurrent. Then, $X(\cdot)$ is positive Harris recurrent if and only if there exists a closed petite set A such that for some $\delta > 0$,*

$$\sup_{x \in A} E_x[\tau_A(\delta)] < \infty. \tag{2.3}$$

One can apply Theorem 2.1, together with a stopping time argument, to show the following version of Foster's criterion. It is contained in Proposition 4.5 in [1].

PROPOSITION 2.1. *Suppose that $X(\cdot)$ is a Markov process, with norm $\|\cdot\|$, such that for some $\varepsilon > 0$, $L > 0$ and $M > 0$,*

$$E_x[\|X(M)\|] \le (\|x\| \vee L) - \varepsilon \qquad \textit{for all } x. \tag{2.4}$$

Then, for $0 < \delta \le M$,

$$E_x[\tau_{A_L}(\delta)] \le \frac{M}{\varepsilon}(\|x\| \vee L) \qquad \textit{for all } x, \tag{2.5}$$

where $A_L = \{x : \|x\| \le L\}$. In particular, if A_L is closed petite, then $X(\cdot)$ is positive Harris recurrent.

Theorem 1.2 and its corollary. Proposition 2.1 and Theorem 1.1 provide the main tools for demonstrating Theorem 1.2. We also require Proposition 3.1, which states that the norm $\|\cdot\|$ in (1.6) is continuous in the metric $d(\cdot,\cdot)$, and hence that $A_L = \{x : \|x\| \le L\}$ is closed for each L. Together, they give a quick proof of the theorem.

PROOF OF THEOREM 1.2. From the conclusion (1.7) in Theorem 1.1, we know that the assumption (2.4) in Proposition 2.1 is satisfied for some L, with $M = N^3$ and $\varepsilon = \varepsilon_1 N^2$. In Theorem 1.2, it is assumed that A_L is petite; by Proposition 3.1, we also know it is closed. So, all of the assumptions in Proposition 2.1 are satisfied, and hence $X(\cdot)$ is positive Harris recurrent. □

Corollary 1.1 follows immediately from Theorem 1.2 and the assertion, before the statement of the corollary, that the sets A_L are petite under the assumptions (1.8) and (1.9). A somewhat stronger version of the analogous assertion for queueing networks is demonstrated in Proposition 4.7 of [1]. (The proposition states that the sets A are uniformly small.) The reasoning is the same in both cases and does not involve the policy of the network. The argument, in essence, requires that one wait long enough for the network to have at least a given positive probability of being empty; this time t does not depend on x for $\|x\| \le L$. One uses (1.8) for this. By using (1.9), one can also show that the joint distribution function of the residual interarrival times has an absolutely continuous component at this time, whose density is bounded away from 0. It will follow that the set A_L is petite with respect to ν, with a chosen as the point mass at t, if ν is concentrated on the empty states, where it is a small enough multiple of $|\mathcal{R}|$-dimensional Lebesque measure restricted to a small cube.

3. Summary of the proof of Theorem 1.1. As mentioned in Section 1, the norm $\|\cdot\|$ in Theorem 1.1 consists of three components, with

$$\|x\| = |x|_L + |x|_R + |x|_A \tag{3.1}$$

for each x. After introducing these components, we will state the bounds associated with each of them that we will need, leaving their proofs to the remaining sections. We then show how Theorem 1.1 follows from these bounds.

Definition of norms. We first define $|x|_L$. This requires a fair amount of notation, which we will introduce shortly. We begin by expressing $|x|_L$ in terms of this notation; when the notation is then specified, we motivate it by referring back to $|x|_L$.

We set $|x|_L = \sup_{r,s} |x|_{r,s}$ for $r \in \mathcal{R}$ and $s > 0$, where

$$|x|_{r,s} = \frac{w_r(1 + as_N)z_r^*(s)}{\nu_r \Gamma(\bar{H}_r^*(s_N))}. \tag{3.2}$$

We need to define the terms $H_r^*(\cdot)$, $z_r^*(\cdot)$, $\Gamma(\cdot)$, a and s_N.

Starting with $H_r^*(\cdot)$ and $z_r^*(\cdot)$, we recall the distribution functions $H_r(\cdot)$ and counting measure $z_r(\cdot)$ from Section 1. In (3.2), we will require their analogs $H_r^*(\cdot)$ and $z_r^*(\cdot)$ to have densities with bounded first derivatives and to be "close" to $H_r(\cdot)$ and $z_r(\cdot)$. For this, we define $H_r^*(\cdot)$ and $z_r^*(\cdot)$ as the convolutions of $H_r(\cdot)$ and of $z_r(\cdot)$ by an appropriate distribution function $\Phi(\cdot)$ with density $\phi(\cdot)$. Setting

$$\phi(s) = \begin{cases} \frac{2}{3}ebe^{-bs}, & \text{for } s > 1/b, \\ \frac{2}{3}b^2 s, & \text{for } s \in (0, 1/b], \\ 0, & \text{for } s \le 0, \end{cases} \tag{3.3}$$

for $b \in \mathbb{Z}^+$ with $b \ge 2$, $\phi(\cdot)$ is the density of $\Phi(s) = \int_{-\infty}^{s} \phi(s')\,ds'$. We note that $\Phi(\cdot)$ has mean at most $2/b$ and that $\phi(\cdot)$ satisfies

$$\phi'(s) \le b^2 \quad \text{and} \quad \phi(s + s')/\phi(s) \ge e^{-bs'} \tag{3.4}$$

for $s, s' > 0$. The above properties and the exponential tail of $\phi(\cdot)$ will be useful later when analyzing $|\cdot|_L$ and $|\cdot|_R$ [as in (3.7), (5.28), (6.39), (6.40) and (9.25)].

Convoluting by $\Phi(\cdot)$, we set

$$\begin{aligned} H_r^*(s) &= (H_r * \Phi)(s) = \int_0^\infty \Phi(s - s')\,dH(s'), \\ z_r^*((0, s]) &= (z_r * \Phi)((0, s]) = \int_0^\infty \Phi(s - s')\,dz_r((0, s']) \end{aligned} \tag{3.5}$$

with $z_r^*(B)$ being defined analogously for $B \subseteq \mathbb{R}^+$. Differentiating both quantities in (3.5), we also set

$$\begin{aligned} h_r^*(s) &= (H_r * \Phi)'(s) = \int_0^\infty \phi(s - s')\,dH(s'), \\ z_r^*(s) &= (z_r * \Phi)'((0, s]) = \int_0^\infty \phi(s - s')\,dz_r((0, s']). \end{aligned} \tag{3.6}$$

Convolution by $\Phi(\cdot)$, as in (3.5), produces a measure $z_r^*(\cdot)$ that approximates $z_r(\cdot)$ and possesses a density.

Since $H_r(\cdot)$ is assumed to have a finite $(2+\delta_1)$th moment for all r, the same is true for $H_r^*(\cdot)$. This implies that for appropriate $C_1 \geq 1$,

$$\bar{H}_r^*(s) \leq \frac{C_1}{(1+s)^{2+\delta_1}} \qquad \text{for all } s > 0 \text{ and } r \in \mathcal{R}, \tag{3.7}$$

for δ_1 chosen as in Theorem 1.1. We assume wlog that $\delta_1 \leq 1$. Since the difference of the means of $H_r^*(\cdot)$ and $H_r(\cdot)$ is at most $2/b$ for each r and $H(\cdot)$ is subcritical, $H^*(\cdot)$ will also be subcritical for large enough b.

We set

$$\Gamma(\sigma) = \sigma + C_2 a \sigma^\gamma \qquad \text{for } \sigma \in [0, 1]. \tag{3.8}$$

We choose $\gamma \in (0, \delta_1/24]$, $C_2 \geq 2C_1/\gamma$ and a small enough so that $a \leq (1/C_2) \wedge 1$ and (9.4) is satisfied. One can think of $\Gamma(\cdot)$ as being almost linear for values of σ that are not too small; the power γ needs to be small in order to be able to bound $|x|_{r,s}$ later on for $H_r^*(s_N)$ small; $\gamma > 0$ is needed so that the integral in (9.7) is finite.

We set $s_N = s \wedge (N_{H_r} + 1)$ for $N \in \mathbb{Z}^+$, where

$$N_{H_r} = (\bar{H}_r^*)^{-1}(1/N^4) \wedge N. \tag{3.9}$$

It follows that

$$1/N^4 \leq \bar{H}_r^*(N_{H_r}) \leq C_1/N^{2+\delta_1}. \tag{3.10}$$

If $H_r^*(\cdot)$ has a relatively fat tail, say $\bar{H}_r^*(s) \sim s^{-3}$, (3.9) implies that $N_{H_r} = N$; otherwise, $N_{H_r} < N$ and $\bar{H}_r^*(N_{H_r}) = 1/N^4$. In either case, it will follow from (3.10) that $\Gamma(\bar{H}_r^*(N_{H_r}))$ is "large enough" for us to adequately bound $|x|_{r,s}$. We will assume that $N \in \mathbb{Z}^+$ is chosen large enough so $N \geq 1/a$ and $N_{H_r} \geq 1$ for all r.

The norm $|\cdot|_L$ has been defined with the following motivation. As the process $X(\cdot)$ evolves, documents arrive at each route, are served, and eventually depart. In Proposition 9.2, we will show that, under certain assumptions for $X(t)$ on $t \in [0, N^3]$, for large enough b,

$$\lambda^w(t) \geq (1+\varepsilon_2)/|x|_L \qquad \text{on } t \in [0, N^3], \tag{3.11}$$

for some $\varepsilon_2 > 0$, because of the subcriticality of $H^*(\cdot)$. Reasoning as below (1.10), this will imply that individual documents receive enough service so that $|X(t)|_L$ decreases on average over $[0, N^3]$. More specifically, the increase in the term $\Gamma(\bar{H}_r^*(s_N))$ in (3.2), after translating s_N according to the service of documents, will compensate for the arrival of new documents. For documents with residual service $s \leq N_{H_r} \leq N+1$ at $t = 0$, the term $1 + as_N$ in (3.2), after translating s_N according to the service of documents, will decrease sufficiently over $[0, N^3]$ to produce the term $-\varepsilon_1 N^2$ in (1.7). For documents with residual service $s > N_{H_r}$, we will instead need to employ the norm $|\cdot|_R$, which we introduce next. (On $(N_{H_r}, N_{H_r}+1]$, the intervals overlap.)

The norm $|\cdot|_R$ in (3.1) is given by

$$|x|_R = M_1 \sum_r \kappa_{N,r} \int_{N_{H_r}}^{\infty} N_r(s) z_r^*(s)\, ds. \tag{3.12}$$

We need to identify the terms $\kappa_{N,r}$, $N_r(\cdot)$ and M_1. We set

$$\kappa_{N,r} = 1/\Gamma(\bar{H}_r^*(N_{H_r})) \tag{3.13}$$

and

$$N_r(s) = \begin{cases} s^2/N, & \text{for } s > N, \\ s, & \text{for } s \le N. \end{cases} \tag{3.14}$$

Later on, we will also employ $\kappa_N \stackrel{\text{def}}{=} \max_r \kappa_{N,r}$. For the term M_1, we will require that

$$M_1 \ge 8C_3 \Big(\max_{r,r'} w_r / w_{r'} \Big), \tag{3.15}$$

where C_3 is chosen as in (3.30).

Since $|\cdot|_R$ is given by a weighted sum of the residual service times of the different documents, it will be easier to work with than $|\cdot|_L$, which is a supremum. For smaller values of s, we required $|\cdot|_L$ because of the nature of the WMMF policy. Because of the bound on $\bar{H}_r^*(\cdot)$ in (3.7), the impact of large residual service times on the evolution of $X(\cdot)$ will typically be small, and so one can employ the "more generous" definition over (N_{H_r}, ∞) given in (3.12).

As we will see in Section 5, we will require the presence of the term $N_r(s)$ in the integrand in (3.12) to ensure that the integral decreases sufficiently rapidly from the service of documents when the integral is large. This will rely on $N_r'(s) \ge 1$ on (N_{H_r}, ∞). For $s > N$, the denominator N in s^2/N is needed so that the expected increase due to arrivals does not dominate the term $-\varepsilon_1 N^2$ in (1.7), which was mentioned in the motivation for the definition of $|\cdot|_L$. This denominator is not needed for $s \in [N_{H_r}, N)$ because (3.9) will guarantee that the integrand is already sufficiently small there. The terms $\kappa_{N,r}$ are needed when we combine the norms $|\cdot|_L$ and $|\cdot|_R$ in $\|\cdot\|$, because of the denominator $\Gamma(\cdot)$ in $|\cdot|_{r,s}$.

The norm $|\cdot|_A$ in (3.1) is needed for the residual interarrival times. It is given by

$$|x|_A = \frac{1}{N} \max_r \theta(u_r), \tag{3.16}$$

where $\theta(y)$, $y > 0$, satisfies the following properties. We assume that $\theta(y) > 0$ for all y and that $\theta(\cdot)$ and $\theta'(\cdot)$ are strictly increasing, with

$$\theta'(y) \to \infty \qquad \text{as } y \to \infty. \tag{3.17}$$

We also assume that

$$\theta(y) \le y^2 \qquad \text{for all } y, \tag{3.18}$$

and that $\theta(\cdot)$ grows sufficiently slowly so that

$$E[\theta(\xi_r)] < \infty \qquad \text{for all } r. \tag{3.19}$$

Since $E[\xi_r] < \infty$, it is possible to specify such $\theta(\cdot)$ that also satisfy the previous two displays.

The above properties for $\theta(\cdot)$ will enable us to show that the expected value of $|X(t)|_A$ will decrease over time when $|X(t)|_A$ is large. In particular, because of (3.17) and (3.19), the decrease in $|\cdot|_A$ due to decreasing residual interarrival times will, on the average, dominate the increase in $|\cdot|_A$ due to new interarrival times that occur when a document joins a route. The argument for this is given in Section 4 and is fairly quick. We note that when ξ_r are all exponentially distributed, the term $|\cdot|_A$ may be omitted in the definition of $\|\cdot\|$.

The reader attempting to understand the norm $\|\cdot\|$ should first concentrate on $|\cdot|_L$, which was chosen to accommodate the WMMF policy. When the service distributions $H_r(\cdot)$ all have compact support and the interarrival times are exponentially distributed, one may, in fact, set $\|x\| = |x|_L$ for a large enough choice of N.

We note that the norm $|\cdot|_L$ is not appropriate for weighted α-fair policies. In particular, the supremum and the function $\Gamma(\cdot)$ in its definition are not appropriate factors in this context. On the other hand, $|\cdot|_R$, with suitable M_1, and $|\cdot|_A$ should still be applicable to α-fair policies, provided a suitable replacement of $|\cdot|_L$ can be found.

In order to apply Proposition 2.1 in the proof of Theorem 1.2 in Section 2, we needed to know that the sets $A_L = \{x : \|x\| \le L\}$ are closed. For this, it suffices to show the norm $\|\cdot\|$ is continuous in the metric $d(\cdot,\cdot)$ that is given in (2.1).

PROPOSITION 3.1. *The norm* $\|\cdot\|$ *in* (3.1) *is continuous in the metric* $d(\cdot,\cdot)$ *given by* (2.1).

PROOF. It suffices to show $|\cdot|_L$, $|\cdot|_R$ and $|\cdot|_A$ are each continuous in $d(\cdot,\cdot)$. For $|\cdot|_L$, note that the coefficients of $z_r^*(s)$ in (3.2) are bounded. On the other hand, if $d(x,x') \le \varepsilon < 1$, then one can show, by using the first part of (3.4), that

$$|z_r^*(s) - z_r^{\prime,*}(s)| \le b^2\varepsilon \qquad \text{for all } s \text{ and } r, \tag{3.20}$$

where $z_r^{\prime,*} = (z')_r^*(s)$. It follows from this and (3.2) that $|\cdot|_L$ is in fact Lipschitz in $d(\cdot,\cdot)$.

For $|\cdot|_R$, one can apply both parts of (3.4) to show with a bit of work that, if $d(x,x') \le \varepsilon < 1$ and x has no residual service times greater than M, for given M, then

$$\int_{N_{H_r}}^{\infty} N_r(s)|z_r^*(s) - z_r^{\prime,*}(s)|\,ds \le (M+1)^2 b^2\varepsilon + (1 - e^{-b\varepsilon})|x|_R \tag{3.21}$$

for all r. Since the coefficients of $\int_{N_{H_r}}^{\infty} N_r(s) z_r^*(s)\,ds$ in $|x|_R$ are bounded and the RHS of (3.21) goes to 0 as $\varepsilon \to 0$, the continuity of $|\cdot|_R$ follows.

Since $\theta'(u_r)$ is bounded for bounded values of u_r, $|\cdot|_A$ is also continuous. □

In addition to the norms in (3.1), we will employ the following norms in showing Theorem 1.1:

$$|x| = \sum_r z_r(\mathbb{R}^+) = \sum_r z_r^*(\mathbb{R}^+) \tag{3.22}$$

and

$$|x|_K = \sum_r \kappa_{N,r} z_r^*((N_{H_r}, \infty)). \tag{3.23}$$

Although we will not employ them in this section, we also introduce the norms

$$|x|_1 = \sum_r z_r^*((0, N_{H_r}]), \qquad |x|_2 = \sum_r z_r^*((N_{H_r}, \infty)) \tag{3.24}$$

and

$$|x|_S = |x|_L + \max_r \frac{w_r}{\rho_r} z_r^*((N_{H_r}, \infty)). \tag{3.25}$$

It obviously follows from (3.22) and (3.24) that $|x| = |x|_1 + |x|_2$. The norm $|\cdot|_S$ will be employed in Proposition 9.2 to derive the bound given in (3.11).

Bounds on $|\cdot|_L, |\cdot|_R$ *and* $|\cdot|_A$. In order to derive (1.7), we need bounds on $|\cdot|_L$, $|\cdot|_R$ and $|\cdot|_A$ as the process $X(t)$ evolves from $t = 0$ to $t = N^3$. We first need to specify the term L appearing in (1.7). We choose l_1 large enough so that

$$\theta'(l_1/2) \geq M_1 N \tag{3.26}$$

and, for all r,

$$E[\theta(\xi_r); \xi_r > l_1/2] \leq (1/|\mathcal{R}|) P(\xi_r > N^3). \tag{3.27}$$

We set

$$L_1 = \frac{1}{N}\theta(l_1) \tag{3.28}$$

and

$$L = 6(\kappa_N^2 N^{17} \vee L_1). \tag{3.29}$$

For $|\cdot|_L$, we employ the bound from Proposition 10.2 that, for large enough N and b, small enough a, and appropriate C_3 and $\varepsilon_3 > 0$,

$$\begin{aligned} &E_x[|X(N^3)|_L] - |x|_L \\ &\qquad \leq C_3 N^3 \cdot 1\{|x| \leq N^6\} + [C_3(|x|_K/|x|) N^3 - \varepsilon_3 N^2] \cdot 1\{|x| > N^6\} \end{aligned} \tag{3.30}$$

for all x. The precise value of ε_3 is not important; in Proposition 10.2, it is given by $\frac{1}{4}\min_r w_r$. We assume wlog that $\varepsilon_3 \le C_3$.

For $|\cdot|_R$, we employ the bound from Proposition 5.1 that, for given $\varepsilon_4 > 0$, large enough N, and $M_2 = \frac{1}{8}(1 \wedge \min_l c_l)(\min_{r,r'}(w_r/w_{r'}))M_1 \ge C_3$,

$$\begin{aligned} E_x[|X(N^3)|_R] - |x|_R \\ \le \varepsilon_4 N^2 - M_2(|x|_K/|x|)N^3 \cdot 1\{|x| > N^6\} \\ - \kappa_N N^4 \cdot 1\{|x|_R > \kappa_N^2 N^{17}, |x| \le N^6\} \end{aligned} \tag{3.31}$$

for all x. We will later choose ε_4 small with respect to ε_3; the constant C_3 is chosen as in (3.15) and (3.30).

For $|\cdot|_A$, we will show in Proposition 4.1 and Proposition 4.2 that, for this choice of ε_4 and large enough N,

$$E_x[|X(N^3)|_A] - |x|_A \le \varepsilon_4 N^2 - M_1 N^3 \cdot 1\{|x|_A > L/6\} \tag{3.32}$$

for all x.

Derivation of (1.7) *from* (3.30), (3.31) *and* (3.32). We now derive (1.7) from these three bounds. Adding the RHS of (3.30), (3.31) and (3.32), one obtains, for large enough N and b, and small enough a,

$$E_x[\|X(N^3)\|] - \|x\| \le 2C_3 N^3 \tag{3.33}$$

for all x. We next consider the behavior of the LHS of (3.33) for $\|x\| > L/2$, where L is given by (3.29). This condition implies that either $|x|_L > \kappa_N^2 N^{17}$, $|x|_R > \kappa_N^2 N^{17}$ or $|x|_A \ge L/6$.

Suppose first that $|x|_L > \kappa_N^2 N^{17}$. We note that if $|x| \le N^6$, then

$$|x|_L \le C_4 N^8$$

for some constant C_4. This bound follows from the definition of $|x|_L$ in (3.2), together with the bounds $z_r^*(s) \le 12b^2|x|$ for all s, $s_N \le N$, and $\Gamma(\bar{H}_r^*(s_N)) \ge C_5/N$, for some $C_5 > 0$ [which follows from (3.10) and $\gamma \le 1/4$]. Therefore, if $|x|_L > \kappa_N^2 N^{17}$ and N is large enough so that $\kappa_N \ge 1$, one must have $|x| > N^6$.

On the other hand, it follows from (3.31) that, on $|x| > N^6$,

$$E_x[|X(N^3)|_R] - |x|_R \le \varepsilon_4 N^2 - M_2(|x|_K/|x|)N^3. \tag{3.34}$$

Adding the terms corresponding to $|x| > N^6$ in (3.30) and (3.32) to this implies that, for $|x| > N^6$, and hence for $|x|_L > \kappa_N^2 N^{17}$,

$$\begin{aligned} E_x[\|X(N^3)\|] - \|x\| &\le (2\varepsilon_4 - \varepsilon_3)N^2 + (C_3 - M_2)(|x|_K/|x|)N^3 \\ &\le -\varepsilon_1 N^2, \end{aligned} \tag{3.35}$$

where the latter inequality follows for $\varepsilon_4 \le \varepsilon_3/3$ and $\varepsilon_1 \stackrel{\text{def}}{=} \varepsilon_3/3$, since $M_2 \ge C_3$.

Suppose next that $|x|_R > \kappa_N^2 N^{17}$ and $|x| \le N^6$. Adding up the corresponding terms from (3.30), (3.31) and (3.32) implies that the LHS of (3.35) is at most

$$2\varepsilon_4 N^2 + C_3 N^3 - \kappa_N N^4 \le -\varepsilon_1 N^3 \tag{3.36}$$

for large N, which is better than the bound in (3.35).

Suppose finally that $|x|_A \ge L/6$. We need to consider only the case $|x| \le N^6$, since $|x| > N^6$ is covered by (3.35). In this case, it follows from (3.30), (3.31) and (3.32) that the LHS of (3.35) is at most

$$(3C_3 - M_1)N^3 \le -\varepsilon_1 N^3, \tag{3.37}$$

since $M_1 \ge 4C_3$.

Together, (3.35), (3.36) and (3.37) imply that, for large enough N and b, and small enough a,

$$E_x[\|X(N^3)\|] - \|x\| \le -\varepsilon_1 N^2 \tag{3.38}$$

for all $\|x\| > L/2$. Since for large N,

$$L - L/2 \ge 2C_3 N^3 + \varepsilon_1 N^2,$$

(1.7) follows easily form (3.33) and (3.38).

4. Upper bounds on $E_x[|X(N^3)|_A]$. In this section, we will demonstrate the inequality (3.32) for the upper bounds on $E_x[|X(N^3)|_A] - |x|_A$. In Proposition 4.1, we obtain the first term on the RHS of (3.32); this holds for all x. We then obtain a better bound in Proposition 4.2, which is valid on $|x| \ge L/6$. Both parts require just standard techniques.

The first bound employs the following elementary inequality on the residual interarrival times at time N^3:

$$|X(N^3)|_A \le |x|_A \vee \frac{1}{N} \max\{\theta(\xi_r(k)) : r \in \mathcal{R}, k \in [2, A_r(N^3) + 1]\}. \tag{4.1}$$

Here and in later sections, $A_r(t)$ denotes the cumulative number of arrivals at the route r by time t; $A(t)$ will denote the corresponding vector. The inequality $k \le A_r(t) + 1$ implies that the interarrival epoch associated with $\xi_r(k)$ has already begun by time t. Recall that $\xi_r(1)$ is the initial residual time at route r and $\xi_r(2), \xi_r(3), \dots$ are i.i.d. random variables, and $\theta(\cdot)$ satisfies (3.16)–(3.19).

PROPOSITION 4.1. *For any $\varepsilon > 0$ and large enough N, not depending on x,*

$$E_x[|X(N^3)|_A] - |x|_A \le \varepsilon N^2. \tag{4.2}$$

PROOF. By (3.19), $E[\theta(\xi_r)] < \infty$ for all r. One can therefore show with some estimation that, for each r,

$$\frac{1}{t} E_x\Big[\max_{k \in [2, A_r(t)+1]} \theta(\xi_r(k))\Big] \to 0, \tag{4.3}$$

uniformly in x as $t \to \infty$. For fixed x, (4.3) follows immediately from (4.83) of [1]; since $A_r(t)$ decreases when $\xi_r(1)$ increases, this limit is uniform in x.

Inequality (4.2) follows immediately from (4.1) and (4.3), with $t = N^3$. □

We proceed to Proposition 4.2. For the proposition, it will be useful to decompose $|X(t)|_A - |x|_A$ as

$$|X(t)|_A - |x|_A = I_A(t) - D_A(t), \tag{4.4}$$

where $I_A(t)$ and $D_A(t)$ are the nondecreasing functions corresponding to the cumulative increase and decrease of $|X(\cdot)|_A$ up to time t. That is, $I_A(0) = D_A(0) = 0$, with $I_A(t)$ being the jump process, with

$$I_A(t) - I_A(t-) = |X(t)|_A - |X(t-)|_A$$

and $D'_A(t)$ being the rate of decrease of $|X(t)|_A$ at other times. We note that $D_A(t)$ is locally Lipschitz, with $D'_A(t)$ defined except at arrivals. In particular, since $U'_r(t) = -1$ except at arrivals,

$$D'_A(t) = \frac{1}{N} \max_r \theta'(U_r(t)) \qquad \text{almost everywhere.} \tag{4.5}$$

We recall the definitions for l_1, L_1 and M_1 in (3.26)–(3.28) and (3.15).

PROPOSITION 4.2. *Suppose that $|x|_A \ge L/6$. Then, for large enough N not depending on x,*

$$E_x[|X(N^3)|_A] - |x|_A \le 1 - M_1 N^3 \le -M_1 N^3/2. \tag{4.6}$$

PROOF. We first show that

$$D_A(N^3) \ge M_1 N^3. \tag{4.7}$$

Since $|x|_A \ge L/6 = \kappa_N^2 N^{17} \vee L_1$ and $\theta(y) \le y^2$ for all y, one has, for $N \ge 2$, that $\max_r u_r \ge N^8 \vee l_1$. So, for all $t \in [0, N^3]$,

$$\max_r u_r - \max_r U_r(t) \le N^3 \le \frac{1}{2} \max_r u_r. \tag{4.8}$$

Consequently, for all $t \in [0, N^3]$,

$$\max_r U_r(t) \ge \frac{1}{2} \max_r u_r \ge N^3 \vee \frac{1}{2} l_1. \tag{4.9}$$

Moreover, by (3.26) and (4.5), for $\max_r U_r(t) \ge \frac{1}{2} l_1$, $D'_A(t) \ge M_1$ almost everywhere. Together with (4.9), this implies $D'_A(t) \ge M_1$ almost everywhere on $[0, N^3]$, and hence (4.7) holds.

On account of (4.7), in order to show (4.6), it suffices to show

$$E_x[I_A(N^3)] \le 1 \tag{4.10}$$

for large N. To obtain (4.10), we first note that, for each r, there cannot be more than one interarrival time occurring over $(0, N^3]$ with value greater than N^3. Moreover, because of (4.9), only interarrival times with value at least $N^3 \vee (l_1/2)$ can contribute to $I_A(N^3)$. The expectation of $\theta(\xi_r)$, for ξ_r conditioned on being greater than N^3 and restricted to being greater than $l_1/2$, is

$$E[\theta(\xi_r); \xi_r > l_1/2]/P(\xi_r > N^3). \tag{4.11}$$

(If ξ_r is bounded above by N^3, set the ratio equal to 0.) It follows that, for any x,

$$E_x[I_A(N^3)] \le \frac{1}{N} \sum_r E[\theta(\xi_r); \xi_r > l_1/2]/P(\xi_r > N^3). \tag{4.12}$$

By (3.27), the RHS of (4.12) is at most $1/N$, which implies (4.10). □

5. Upper bounds on $E_x[|X(N^3)|_R]$. In this section, we will demonstrate the following proposition for the upper bounds on $E_x[|X(N^3)|_R] - |x|_R$, where $|\cdot|_R$ is the norm introduced in (3.12).

PROPOSITION 5.1. *For given $\varepsilon > 0$, large enough N and all x,*

$$\begin{aligned} E_x[|X(N^3)|_R] - |x|_R &\le \varepsilon N^2 - M_2(|x|_K/|x|)N^3 \cdot 1\{|x| > N^6\} \\ &\qquad - \kappa_N N^4 \cdot 1\{|x|_R > \kappa_N^2 N^{17}, |x| \le N^6\}, \end{aligned} \tag{5.1}$$

where M_2 is specified before (3.31).

The bound (5.1) implies (3.31), which was employed in Section 3, together with bounds on $E_x[|X(N^3)|_L]$ and $E_x[|X(N^3)|_A]$, to obtain (1.7) of Theorem 1.1. The bound on $E_x[|X(N^3)|_A]$ was derived relatively quickly, whereas the bound on $E_x[|X(N^3)|_L]$ will require substantial estimation and will be derived in Sections 6–10. The bound on $E_x[|X(N^3)|_R]$ that is given here will require a moderate amount of work.

In order to show Proposition 5.1, it will be useful to rewrite $|X(t)|_R - |x|_R$ as

$$|X(t)|_R - |x|_R = I_R(t) - D_R(t), \tag{5.2}$$

where $I_A(t)$ and $D_A(t)$ are the nondecreasing functions corresponding to the cumulative increase and decrease of $|X(\cdot)|_R$ up to time t. A similar decomposition was used in Section 4 for $|X(t)|_A$. Here, $I_R(0) = D_R(0) = 0$, with $I_R(t)$ being the jump process with

$$I_R(t) - I_R(t-) = |X(t)|_R - |X(t-)|_R.$$

One can check that $D_R(\cdot)$ is continuous except when a document departs from a route. Its derivative is defined almost everywhere, being defined except at the arrival or departure of a document. Since $D_R(\cdot)$ is nondecreasing,

$$D_R(t_2) - D_R(t_1) \ge \int_{t_1}^{t_2} D'_R(t)\,dt \qquad \text{for } t_1 \le t_2.$$

It is easy to obtain a suitable upper bound on $E_x[I_R(N^3)]$; a suitable lower bound on $E_x[D_R(N^3)]$ requires more effort. We therefore first demonstrate Proposition 5.2, which analyzes $E_x[I_R(N^3)]$.

As in Section 4, $A_r(t)$ denotes the cumulative number of arrivals at route r by time t. It follows from elementary renewal theory that, for appropriate C_6 and $t \geq 1$,

$$E_x[A_r(t)] \leq C_6 t \qquad \text{for each } r \tag{5.3}$$

(see, e.g., [3], page 136). Since large residual interarrival times can only delay arrivals, the bound is uniform in x.

PROPOSITION 5.2. *For given $\varepsilon > 0$ and large enough N,*

$$E_x[I_R(N^3)] \leq \varepsilon N^2 \qquad \textit{for all } x. \tag{5.4}$$

PROOF. It follows from (3.12) that the expected increase in $I_R(\cdot)$, due to a document that arrives at route r, is

$$M_1 \kappa_{N,r} \int_{N_{H_r}}^{\infty} N_r(s) h_r^*(s)\, ds.$$

Since the number of arriving documents by time N^3 and their initial service times are independent, it follows that

$$E_x[I_R(N^3)] = \left(M_1 \kappa_{N,r} \int_{N_{H_r}}^{\infty} N_r(s) h_r^*(s)\, ds \right) E_x[A_r(N^3)]. \tag{5.5}$$

In order to bound the first term on the RHS of (5.5), we decompose the integral there into $\int_{N_{H_r}}^{N} + \int_{N}^{\infty}$. When $N \geq N_{H_r}$, one has, by (3.9) and (3.13),

$$\begin{aligned} \kappa_{N,r} \int_{N_{H_r}}^{N} N_r(s) h_r^*(s)\, ds &= \frac{1}{\Gamma(1/N^4)} \int_{N_{H_r}}^{N} s h_r^*(s)\, ds \\ &\leq \frac{N}{\Gamma(1/N^4)} \bar{H}_r^*(N_{H_r}) \leq (N^3 \Gamma(1/N^4))^{-1}. \end{aligned} \tag{5.6}$$

This is, for large enough N, at most $1/N^2$, because of the small power γ in the definition of $\Gamma(\cdot)$. Also,

$$\begin{aligned} \kappa_{N,r} \int_{N}^{\infty} N_r(s) h_r^*(s)\, ds &\leq \frac{1}{N\Gamma(1/N^4)} \int_{N}^{\infty} s^2 h_r^*(s)\, ds \\ &\leq \frac{1}{N^{1+\delta_1/2}\Gamma(1/N^4)} \int_{N}^{\infty} s^{2+\delta_1/2} h_r^*(s)\, ds \\ &\leq \frac{C_7}{N^{1+\delta_1/2}\Gamma(1/N^4)} \end{aligned} \tag{5.7}$$

for appropriate C_7, with the last inequality holding because of (3.7). Since $\gamma \le \delta_1/24$, this is, for large N, at most $1/N^{1+\delta_1/4}$. Together, the bounds for the two integrals imply that, for large enough N,

$$M_1\kappa_{N,r}\int_{N_{H_r}}^{\infty} N_r(s)h_r^*(s)\,ds \le 2/N^{1+\delta_1/4}. \tag{5.8}$$

Application of (5.8) and (5.3) to (5.5), with $t = N^3$ in (5.3), implies (5.4). □

We now derive a lower bound on $E_x[D_R(N^3)]$. As in (5.1), we need to consider two separate cases, depending on whether $|x| > N^6$ or both $|x|_R > \kappa_N^2 N^{17}$ and $|x| \le N^6$ hold. In both cases, we will employ the following lemma. Recall that $M_2 = \frac{1}{8}C_8 M_1$, with $C_8 = (1 \wedge \min_l c_l)(\min_{r,r'}(w_r/w_{r'}))$.

LEMMA 5.1. (a) *For all* t,

$$D_R(t) \ge M_1\big(|x|_K - |X(t)|_K\big). \tag{5.9}$$

(b) *For almost all* t,

$$\begin{aligned} D_R'(t) &\ge \frac{C_8 M_1}{|X(t)|}\sum_r \kappa_{N,r}\int_{N_{H_r}}^{\infty}\Big(\frac{s}{N}\vee 1\Big)Z_r^*(t,s)\,ds \\ &\ge 8M_2|X(t)|_K/|X(t)|. \end{aligned} \tag{5.10}$$

PROOF. We first show (a). Recall that $\tilde{X}(\cdot)$ is the stochastic process constructed from $X(\cdot)$ in Section 1, where service of documents is pathwise identical to $X(\cdot)$, but where the arrival of documents is suppressed. One can check that, for all t and ω,

$$|\tilde{X}(t)|_K \le |X(t)|_K \tag{5.11}$$

and

$$D_R(t) \ge |x|_R - |\tilde{X}(t)|_R. \tag{5.12}$$

Inequality (5.11) follows immediately from $\tilde{Z}^*(t,B) \le Z^*(t,B)$ for all $B \subseteq \mathbb{R}^+$. For (5.12), note that the LHS gives the cumulative decrease of $|X(\cdot)|_R$ over $[0,t]$ due to the service of all documents, whereas the RHS gives the decrease due to service of only the original documents while ignoring the decrease due to service of new documents.

On account of (5.11) and (5.12), to show (5.9) it suffices to show

$$|x|_R - |\tilde{X}(t)|_R \ge M_1\big(|x|_K - |\tilde{X}(t)|_K\big). \tag{5.13}$$

Substituting in the definition of $|\cdot|_R$ given by (3.12) and integrating by parts on the LHS of (5.13) gives

$$\begin{aligned} &M_1 \sum_r \kappa_{N,r} N_r(N_{H_r})\big(z_r^*((N_{H_r},\infty)) - \tilde{Z}_r^*(t,(N_{H_r},\infty))\big) \\ &\qquad + M_1 \sum_r \kappa_{N,r} \int_{N_{H_r}}^{\infty} N_r'(s)\big(z_r^*((s,\infty)) - \tilde{Z}_r^*(t,(s,\infty))\big)\,ds. \end{aligned} \tag{5.14}$$

It follows from (3.14) and $N_{H_r} \geq 1$ that $N_r(N_{H_r}) \geq 1$ and that $N_r'(s) \geq 1$ for all s. Consequently, (5.14) is at least

$$\begin{aligned} &M_1 \sum_r \kappa_{N,r}\big(z_r^*((N_{H_r},\infty)) - \tilde{Z}_r^*(t,(N_{H_r},\infty))\big) \\ &\qquad = M_1\big(|x|_K - |\tilde{X}(t)|_K\big), \end{aligned}$$

which implies (5.13).

For (b), we first note that because of the weighted max–min fair protocol and (1.1), the rate at which each document is served is at least

$$\Big(\min_l c_l\Big)\Big(\min_{r,r'}(w_r/w_r')\Big)/|X(t)|. \tag{5.15}$$

Moreover, the rate of decrease of $|X(t)|_R$ per unit service of each document on route r is at least

$$\begin{aligned} M_1\kappa_{N,r}\int_{N_{H_r}}^{\infty} N_r'(s) Z_r^*(t,s)\,ds &\geq M_1\kappa_{N,r}\int_{N_{H_r}}^{\infty}\Big(\frac{s}{N}\vee 1\Big) Z_r^*(t,s)\,ds \\ &\geq M_1\kappa_{N,r} Z_r^*(t,(N_{H_r},\infty)). \end{aligned} \tag{5.16}$$

Summing (5.16) over r and multiplying by (5.15) gives each of the bounds in (5.10). $\square$

We first derive a lower bound on $E_x[D_R(N^3)]$ in the case where $|x| > N^6$.

PROPOSITION 5.3. *For large enough N and all $|x| > N^6$,*

$$E_x[D_R(N^3)] \geq M_2(|x|_K/|x|)N^3. \tag{5.17}$$

PROOF. We restrict our attention to the set

$$B_1 = \{\omega : |X(t)| \leq |x| + N^6 \text{ for all } t \in [0, N^3]\}.$$

By applying Markov's inequality to inequality (5.3) with $t = N^3$, one has that, for large N,

$$P_x\Big(\sum_r A_r(N^3) > N^6\Big) \leq \frac{C_6}{N^3}|\mathcal{R}| \leq \frac{1}{2} \tag{5.18}$$

for all x. Consequently,

$$P(B_1) \geq 1/2. \tag{5.19}$$

This bound does not depend on $|x|$.

We now consider two cases, depending on whether the set

$$B_2 = \{\omega : |X(t)|_K > \tfrac{1}{2}|x|_K \text{ for all } t \in [0, N^3]\}$$

occurs. Since $|x| > N^6$, it follows from the second half of (5.10) that, for all $t \in [0, N^3]$,

$$D_R'(t) \geq 2M_2|x|_K/|x|$$

on $B_1 \cap B_2$. Consequently, on $B_1 \cap B_2$,

$$D_R(N^3) \geq 2M_2(|x|_K/|x|)N^3. \tag{5.20}$$

On the other hand, on $B_1 \cap B_2^c$,

$$|x|_K - |X(\tau)|_K \geq \tfrac{1}{2}|x|_K \tag{5.21}$$

for some (random) $\tau \in [0, N^3]$. By (5.9),

$$D_R(t) \geq M_1(|x|_K - |X(t)|_K)$$

for all t. Together with (5.21), this implies that

$$D_R(N^3) \geq D_R(\tau) \geq \tfrac{1}{2}M_1|x|_K \geq 2M_2(|x|_K/|x|)N^6, \tag{5.22}$$

where $|x| > N^6$ was used in the last inequality.

Together, (5.20) and (5.22) imply that, on B_1,

$$D_R(N^3) \geq 2M_2(|x|_K/|x|)N^3.$$

Inequality (5.17) follows from this and (5.19). □

We now derive a lower bound on $E_x[D_R(N^3)]$ in the case where $|x|_R > \kappa_N^2 N^{17}$ and $|x| \leq N^6$ both hold. We note that, starting from (5.26), the argument relies on the discreteness of documents. If one wishes to employ a fluid limit based argument rather than the discrete setting employed in this paper, different reasoning will be required at this point; it is not obvious how one would proceed.

PROPOSITION 5.4. *For large enough N,*

$$E_x[D_R(N^3)] \geq \kappa_N N^4 \tag{5.23}$$

for all $|x|_R > \kappa_N^2 N^{17}$ and $|x| \leq N^6$.

PROOF. As in the proof of Proposition 5.3, we restrict attention to the set B_1 defined there. The bound $P(B_1) \geq 1/2$ in (5.19) continues to hold here. In our present setting, since $|x| \leq N^6$, $\omega \in B_1$ implies that

$$|X(t)| \leq 2N^6 \qquad \text{for all } t \in [0, N^3].$$

We also consider two cases, depending on whether

$$B_3 = \{\omega : |X(t)|_R > \tfrac{1}{2}\kappa_N^2 N^{17} \text{ for all } t \in [0, N^3]\}$$

occurs.

The case B_3^c is almost immediate. It follows from (5.2) that, for large enough N and for some $\tau \in (0, N^3]$,

$$D_R(N^3) \geq D_R(\tau) \geq |x|_R - |X(\tau)|_R \geq \tfrac{1}{2}\kappa_N^2 N^{17} > 2\kappa_N N^4 \tag{5.24}$$

for $\omega \in B_3^c$.

The case B_3 requires some work. We first note that, by the first part of (5.10),

$$\begin{aligned} D_R'(t) &\geq \frac{C_8 M_1}{|X(t)|} \sum_r \kappa_{N,r} \biggl(\int_{N_{H_r}}^{\infty} \Bigl(\frac{s}{N} \vee 1 \Bigr) Z_r^*(t,s)\,ds \biggr) \\ &\geq \frac{C_8 M_1}{2N^6} \sum_r \kappa_{N,r} \biggl(\int_{N_{H_r}}^{\infty} \Bigl(\frac{s}{N} \vee 1 \Bigr) Z_r^*(t,s)\,ds \biggr), \end{aligned} \tag{5.25}$$

when $\omega \in B_1$.

We will truncate the second integral in (5.25) in order to be able to introduce an additional factor s into the integrand. We first note that, since $\Phi(0) = 0$, if a document with residual service time at least s is present at time t on some route r, then, for large N,

$$|X(t)|_R \geq M_1 \kappa_{N,r} s^2/N \geq M_1 s^2/N. \tag{5.26}$$

Hence, there are no documents with residual service time

$$s > s_1 \stackrel{\text{def}}{=} \bigl((N/M_1)|X(t)|_R\bigr)^{1/2}. \tag{5.27}$$

It follows that, for appropriate $C_9 > 0$, (5.25) is at least

$$\begin{aligned} &\frac{C_8 M_1}{2N^6} \sum_r \kappa_{N,r} \biggl(\int_{N_{H_r}}^{s_1+1} \Bigl(\frac{s}{N} \vee 1 \Bigr) Z_r^*(t,s)\,ds \biggr) \\ &\quad \geq \frac{C_8 M_1^{3/2}}{4N^{13/2}|X(t)|_R^{1/2}} \sum_r \kappa_{N,r} \biggl(\int_{N_{H_r}}^{s_1+1} N_r(s) Z_r^*(t,s)\,ds \biggr) \\ &\quad \geq \frac{2C_9 M_1^{3/2}}{N^{13/2}|X(t)|_R^{1/2}} \sum_r \kappa_{N,r} \biggl(\int_{N_{H_r}}^{\infty} N_r(s) Z_r^*(t,s)\,ds \biggr) \\ &\quad = \frac{2C_9 M_1^{1/2}}{N^{13/2}} |X(t)|_R^{1/2} \geq C_9 M_1^{1/2} \kappa_N N^2 \end{aligned} \tag{5.28}$$

for all $t \in [0, N^3]$. The exponential tail of $\Phi(\cdot)$ is used in the last inequality; the equality relies on $\omega \in B_3$.

Employing the bound on $D'_R(t)$ obtained from (5.25) and (5.28), and integrating over $t \in [0, N^3]$, it follows that, for large N,

$$D_R(N^3) \geq C_9 M_1^{1/2} \kappa_N N^5 > 2\kappa_N N^4$$

on $B_1 \cap B_3$. Together with (5.24), this implies that $D_R(N^3) > 2\kappa_N N^4$ on B_1. Inequality (5.23) follows from this and $P(B_1) \geq 1/2$. □

Proposition 5.1 follows immediately from (5.2) and Propositions 5.2, 5.3 and 5.4.

6. Upper bounds on $E_x[|X(N^3)|_L]$: Basic layout and bounds on exceptional sets. In this section, we begin our investigation of upper bounds on $E_x[|X(N^3)|_L] - |x|_L$. Since these bounds will require us to examine a number of subcases in Sections 6–9, we will only arrive at the desired bounds in Section 10. In the current section, we first state certain elementary inequalities, mostly involving $|\cdot|_{r,s}$, that will be useful later on. We then define the "good" sets $\mathcal{A}(\cdot)$ of realizations of $X(\cdot)$ to which our bounds in Sections 7–9 will apply. The remainder of the section is spent demonstrating Proposition 6.1, which gives an upper bound on $E_x[|X(t)|_L - |x|_L; \mathcal{A}(t)^c]$, where $\mathcal{A}(t)^c$ is the small exceptional set.

Elementary inequalities. Here we state a number of elementary inequalities that will be useful later on. Let $z_i(\cdot)$, $i = 1, 2, 3$, denote configurations of particles on $\mathbb{R}^+$, with $z_i(B)$ denoting the number of particles (or documents) in $B \subseteq \mathbb{R}^+$. If one assumes

$$z_3(B) = z_1(B) + z_2(B) \qquad \text{for all } B \subseteq \mathbb{R}^+, \tag{6.1}$$

it follows that

$$z_3^*(B) = z_1^*(B) + z_2^*(B) \qquad \text{for all } B \subseteq \mathbb{R}^+, \tag{6.2}$$

where $z_i^*(B)$ is defined analogously to $z_r^*(B)$ below (3.5), with convolution being with respect to $\phi(\cdot)$. Several elementary equalities follow from (6.2), including

$$|x_3|_{r,s} = |x_1|_{r,s} + |x_2|_{r,s} \qquad \text{for all } r \in \mathcal{R} \text{ and } s > 0, \tag{6.3}$$

where x_i are states in the metric space S introduced in Section 2 for which the analog of (6.1) is satisfied for each r and $|\cdot|_{r,s}$ is given by (3.2).

Recall that $\tilde{X}(\cdot)$ and $X^A(\cdot)$ are the processes constructed from $X(\cdot)$ that were introduced in Section 1, where service of each document is pathwise identical to $X(\cdot)$, but where, for $\tilde{X}(\cdot)$, the arrival of documents is suppressed and, for $X^A(\cdot)$, only new documents are included. One has

$$Z(t, B) = \tilde{Z}(t, B) + Z^A(t, B) \qquad \text{for } t \geq 0 \text{ and } B \subseteq \mathbb{R}^+,$$

where the processes $Z(\cdot)$, $\tilde{Z}(\cdot)$, and $Z^A(\cdot)$ correspond to $X(\cdot)$, $\tilde{X}(\cdot)$ and $X^A(\cdot)$. From (6.2),

$$Z^*(t, B) = \tilde{Z}^*(t, B) + Z^{A,*}(t, B) \qquad \text{for } t \geq 0 \text{ and } B \subseteq \mathbb{R}^+, \tag{6.4}$$

and from (6.3),

$$|X(t)|_{r,s} = |\tilde{X}(t)|_{r,s} + |X^A(t)|_{r,s} \qquad \text{for } t \geq 0, r \in \mathcal{R}, s > 0. \tag{6.5}$$

Another elementary equality involving $X(\cdot)$ is given by

$$\tilde{Z}_r(t, B) = z_r\bigl(B + \Delta_r(t)\bigr) \qquad \text{for } t \geq 0, r \in \mathcal{R}, B \subseteq \mathbb{R}^+, \tag{6.6}$$

where, we recall, $\Delta_r(t)$ is the translation that gives the amount of service an original document that has not yet completed service has received by time t. The equality relies on all documents on a given route r receiving equal service at each time. [If $\tilde{Z}_r(t, \mathbb{R}^+) = 0$, set $\Delta_r(t) = \infty$ and $z_r(\mathbb{R}^+ + \infty) = 0$.] From (6.6), one obtains

$$\tilde{Z}_r^*(t, B) \leq z_r^*\bigl(B + \Delta_r(t)\bigr) \qquad \text{for } t \geq 0, r \in \mathcal{R}, B \subseteq \mathbb{R}^+; \tag{6.7}$$

the inequality arises from the possibility that original documents have completed service by time t.

A consequence of (3.2) and (6.7) is that

$$|\tilde{X}(t)|_{r,s} \leq |x|_{r,s+\Delta_r(t)} \qquad \text{for } t \geq 0, r \in \mathcal{R}, s > 0. \tag{6.8}$$

Combining (6.5) and (6.8) produces

$$|X(t)|_{r,s} \leq |x|_{r,s+\Delta_r(t)} + |X^A(t)|_{r,s} \qquad \text{for } t \geq 0, r \in \mathcal{R}, s > 0; \tag{6.9}$$

taking the supremum over all r and s therefore gives

$$|X(t)|_L \leq |x|_L + |X^A(t)|_L \qquad \text{for all } t \geq 0. \tag{6.10}$$

Application of (6.7) also implies

$$\tilde{Z}_r^*(t, s) \leq z_r^*\bigl(s + \Delta_r(t)\bigr) \qquad \text{for } t \geq 0, r \in \mathcal{R}, B \subseteq \mathbb{R}^+, \tag{6.11}$$

and application of (6.7), together with (6.4), implies that

$$|X(t)|_2 \leq |x|_2 + |X^A(t)|_2 \qquad \text{for } t \geq 0, \tag{6.12}$$

where $|\cdot|_2$ is given in (3.24). The term on the LHS of (5.16) can also be derived using (6.11).

The sets $\mathcal{A}(t)$. In this subsection, we define the random set $\mathcal{A}(t)$, which is a function of $X(t')$, for $t' \in [0,t]$. In Sections 7–10, we will establish upper bounds on $|X(N^3)|_{r,s}$ for $\omega \in \mathcal{A}(N^3)$; the exceptional small set $\mathcal{A}(N^3)^c$ will be treated in the next subsection. The set $\mathcal{A}(t)$ will be a "good" set in the sense that the number of arrivals over $[0,t]$, for given t, is restricted by upper bounds, which will enable us to show that $|X(\cdot)|_L$ decreases in an appropriate manner.

The set $\mathcal{A}(t)$ is given by $\mathcal{A}(t) = \mathcal{A}_1(t) \cap \mathcal{A}_2(t)$, with

$$\mathcal{A}_i(t) = \bigcap_{r,j} \mathcal{A}_{i,r,j}(t) \qquad \text{for } i = 1, 2, \tag{6.13}$$

where $\mathcal{A}_{i,r,j}(t)$ specify upper bounds on the numbers of weighted arrivals of documents with different service times. To define $\mathcal{A}_{i,r,j}(t)$, we denote by $v_0, v_1, \dots, v_J$ the increasing sequence with

$$v_{j+1} = v_j + 1/b^3 \qquad \text{for } j = 0, \dots, J-1, \tag{6.14}$$

with $v_0 = 0$ and $v_J = N + 1$, and where b is as in (3.3). Note that it follows from the second half of (3.4) that, for $b \geq 2$,

$$\bar{H}_r^*(v_{j+1})/\bar{H}_r^*(v_j) \geq 1/2 \qquad \text{for all } r \text{ and } j. \tag{6.15}$$

We also denote by $S_r^1(k)$, $k = 1, \dots, A_r(t)$, the service time of the kth arrival at route r, where $A_r(t)$ is the cumulative number of arrivals at r by time t.

We set, for $r \in \mathcal{R}$ and $j = 0, \dots, J$,

$$\mathcal{A}_{1,r,j}(t) = \left\{\omega : \sum_{k=1}^{A_r(t)} \bar{\Phi}\big(v_j - S_r^1(k)\big) \leq 2\nu_r\big(\bar{H}_r^*(v_j)t \vee t^\eta\big)\right\}. \tag{6.16}$$

Here, we assume $\eta \in (0, 1/12]$, and, as elsewhere, we set $\bar{H}_r(\cdot) = 1 - H_r(\cdot)$ and $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$. One has, as a special case of (6.16), that

$$A_r(t) \leq 2\nu_r t \qquad \text{on } \mathcal{A}_{1,r,0}(t). \tag{6.17}$$

Since

$$E\big[\bar{\Phi}\big(v_j - S_r^1(k)\big)\big] = \int_0^\infty \bar{\Phi}(v_j - s)\, dH_r(s) = \bar{H}_r^*(v_j) \tag{6.18}$$

and $A_r(t) \sim \nu_r t$ for large t, the probability of the complement $\mathcal{A}_{1,r,j}(t)^c$ can be bounded above by using standard large derivation estimates. The term t^η is included on the RHS of (6.16) so that, when $\bar{H}_r^*(v_j)$ is small, the probability of the event remains small.

We also set, for $r \in \mathcal{R}$ and $j = 0, \dots, J$,

$$\mathcal{A}_{2,r,j}(t) = \left\{\omega : \sum_{k=1}^{A_r(t)} \phi\big(v_j - S_r^1(k)\big) \leq (1 + \varepsilon_5)\nu_r\big(h_r^*(v_j)t \vee t^\eta\big)\right\}, \tag{6.19}$$

where $\varepsilon_5 > 0$. Analogous to (6.18), one has

$$E\big[\phi\big(v_j - S_j^1(k)\big)\big] = \int_0^\infty \phi(v_j - s)\,dH_r(s) = h_r^*(v_j). \tag{6.20}$$

The probabilities $P_x(\mathcal{A}_{2,r,j}(t)^c)$ will satisfy large deviation bounds as well. The constant ε_5 here will later be required to satisfy $\varepsilon_5 \le \varepsilon_7/4$, where ε_7 is specified in (9.1) and measures how subcritical the network is. In (6.16), we only need to employ the constant 2, rather than $1+\varepsilon_5$ as in (6.19), because (6.16) will be applied to the right tail of $\bar{H}_r^*(\cdot)$, rather than the "main body" of $H_r^*(\cdot)$, as will (6.19).

Upper bounds on $\mathcal{A}(t)^c$. The main result in this last subsection is the following proposition.

PROPOSITION 6.1. *For large enough t,*

$$E_x[|X(t)|_L - |x|_L; \mathcal{A}(t)^c] \le N^3 e^{-C_{10}t^\eta} \tag{6.21}$$

for all N, x and appropriate $C_{10} > 0$.

Proposition 6.1 gives strong bounds on the growth of $|X(t)|_L$ on $\mathcal{A}(t)^c$. This behavior is primarily due to the small probability $P_x(\mathcal{A}(t)^c)$, which is given in the next proposition.

PROPOSITION 6.2. *For large enough t,*

$$P_x(\mathcal{A}(t)^c) \le N e^{-C_{11}t^\eta} \tag{6.22}$$

for all N, x and appropriate $C_{11} > 0$.

The interarrival times are assumed to be independent, and large initial residual interarrival times only delay future arrivals. The initial state x will therefore not affect the bounds in (6.21) and (6.22). Note that only the arrival process $A(\cdot)$ is relevant for the bounds in (6.22).

Proposition 6.2 will serve as the main step in demonstrating Proposition 6.1; it will also be used along with Proposition 6.1 in Section 10. When we apply (6.21) and (6.22) there, we will set $t = N^3$ and so the factors N^3 and N can be absorbed into the corresponding exponentials. We note that C_{10} and C_{11} in (6.21) and (6.22), and the bound on t depend on our choices of ε_5 and b, and on ν_r and w_r.

In order to show Proposition 6.2, we will employ elementary large deviation estimates, which are given in the following two lemmas.

LEMMA 6.1. *Let $W(1), W(2), \dots$ denote nonnegative i.i.d. random variables with mean $\beta < \infty$. Then, for each $\varepsilon > 0$, there exists $C_{12} > 0$, so that*

$$P\left(\sum_{k=1}^n W(k) \le (1-\varepsilon)\beta n\right) \le e^{-C_{12}n}. \tag{6.23}$$

When the support of $W(1)$ *is contained in* $[0, 1]$ *and* $\varepsilon \in (0, 1]$,

$$P\left(\sum_{k=1}^{n} W(k) \geq (1+\varepsilon)\beta n\right) \leq e^{-C_{13}\varepsilon^2\beta n}, \tag{6.24}$$

where $C_{13} > 0$ *does not depend on the distribution of* $W(1)$ *or on* ε.

PROOF. Both (6.23) and (6.24) are elementary large deviation bounds. We summarize the argument for (6.24); (6.23) can be shown directly or by applying (6.24) after truncating $W(k)$.

As usual, one employs the moment generating function

$$\psi_\theta(n) = E\big[e^{\theta\sum_{k=1}^{n}(W(k)-\beta)}\big] \qquad \text{for } \theta > 0. \tag{6.25}$$

By expanding the exponential for $n = 1$, it follows that for appropriate $C_{14} \geq 1$ and for $\theta \in (0, 1]$,

$$\psi_\theta(1) \leq 1 + C_{14}\beta\theta^2, \tag{6.26}$$

and hence

$$\psi_\theta(n) \leq (1 + C_{14}\beta\theta^2)^n \leq e^{C_{14}\beta\theta^2 n}. \tag{6.27}$$

By applying Markov's inequality and setting $\theta = \varepsilon/2C_{14}$, it follows that the LHS of (6.24) is at most

$$e^{-\varepsilon\beta\theta n}\psi_\theta(n) \leq e^{-\varepsilon^2\beta n/4C_{14}} \leq e^{-C_{13}\varepsilon^2\beta n} \tag{6.28}$$

for $C_{13} = 1/4C_{14}$, as desired. $\square$

Let $W(1), W(2), \ldots$ denote the successive interarrival times for a renewal process (with delay), with $A(t) = \max\{n : \sum_{k=1}^{n} W(k) \leq t\}$ denoting the number of renewals by time t. Here, $W(2), W(3), \ldots$ are i.i.d., with $W(1)$ being the residual interarrival time. We also introduce i.i.d. random variables $Y(1), Y(2), \ldots,$ with $Y(1) \in [0, 1]$ that are defined on the same space as $W(k)$. Set $E[W(2)] = \beta > 0$ and $E[Y(1)] = m$.

LEMMA 6.2. *Let* $W(1), W(2), \ldots$ *and* $Y(1), Y(2), \ldots$ *be as above. Then, for given* $\varepsilon \in (0, 1]$ *and large* t,

$$P\left(\sum_{k=1}^{A(t)} Y(k) > (1+\varepsilon)\beta^{-1}mt\right) \leq e^{-C_{15}mt}, \tag{6.29}$$

where $C_{15} > 0$ *does not depend on the distribution of* $Y(1)$.

PROOF. $\{A(t) \geq n\}$ is contained in the event $\{\sum_{k=1}^{n} W(k) \leq t\}$. Consequently, by (6.23) of Lemma 6.1, substitution of $\varepsilon/3$ for ε there implies that, for $n(t) = \lceil (1-\varepsilon/3)^{-1}\beta^{-1}t \rceil$,

$$P\bigl(A(t) > n(t)\bigr) \leq P\left(\sum_{k=2}^{n(t)+1} W(k) \leq t\right) \leq e^{-C_{16}t} \tag{6.30}$$

for appropriate $C_{16} > 0$ and large t (which may depend on ε and the distribution of W).

We next consider the set where $A(t) \leq n(t)$. It follows from (6.24) of Lemma 6.1 that

$$\begin{aligned} &P\left(\sum_{k=1}^{A(t)} Y(k) > (1+\varepsilon)\beta^{-1}mt;\, A(t) \leq n(t)\right) \\ &\quad \leq P\left(\sum_{k=1}^{n(t)} Y(k) > (1+\varepsilon)\beta^{-1}mt\right) \leq e^{-C_{13}\varepsilon^2\beta^{-1}mt/9}. \end{aligned} \tag{6.31}$$

Inequality (6.29) follows from (6.30) and (6.31). □

We now employ Lemma 6.2 to prove Proposition 6.2.

PROOF OF PROPOSITION 6.2. We first note that since $\mathcal{A}(t) = \mathcal{A}_1(t) \cap \mathcal{A}_2(t)$, with $\mathcal{A}_i(t) = \bigcap_{r\in\mathcal{R}} \bigcap_{j=0}^{J} \mathcal{A}_{i,r,j}(t)$, where $J = b^3(N+1)+1 \leq 2b^3N$, it suffices to show that for each $\mathcal{A}_{i,r,j}(t)$,

$$P_x(\mathcal{A}_{i,r,j}(t)^c) \leq e^{-C_{17}t^{\eta}} \tag{6.32}$$

for $t \geq t_0$, for some fixed t_0 and appropriate $C_{17} > 0$.

We consider the case where $i = 1$. Denote by $W(1), W(2), \dots$ the interarrival times of documents on route r and set $Y(k) = \bar{\Phi}(v_j - S_r^1(k))$. Then, $Y(k)$ are i.i.d. random variables and, except for $W(1)$, so are $W(k)$. One has

$$\beta \stackrel{\text{def}}{=} E[W(2)] = \nu_r^{-1} \quad \text{and} \quad m \stackrel{\text{def}}{=} E[Y(1)] = \bar{H}_r^*(v_j) \tag{6.33}$$

with the last equality following from (6.18).

We break the problem into two cases, depending on whether or not $\bar{H}_r^*(v_j) \geq t^{\eta-1}$, in each case applying Lemma 6.2, with $\varepsilon = 1$. Under $\bar{H}_r^*(v_j) \geq t^{\eta-1}$, one has

$$P_x(\mathcal{A}_{1,r,j}(t)^c) = P_x\left(\sum_{k=1}^{A_r(t)} \bar{\Phi}\bigl(v_j - S_r^1(k)\bigr) > 2\nu_r \bar{H}_r^*(v_j)t\right) \leq e^{-C_{15}t^{\eta}} \tag{6.34}$$

for large t and $C_{15} > 0$ as in the lemma, where neither depends on the particular value of $\bar{H}_r^*(v_j)$.

For $\bar{H}_r^*(v_j) < t^{\eta-1}$, we replace the random variables defined above (6.33) by i.i.d. random variables $Y'(k) \in (0,1]$, with $Y'(k) \geq Y(k)$ and $E[Y'(k)] = t^{\eta-1}$. Then, again applying Lemma 6.2, but this time to $Y'(k)$, $k = 1, 2, \dots,$

$$P_x(\mathcal{A}_{1,r,j}(t)^c) \leq P_x\left(\sum_{k=1}^{A_r(t)} Y'(k) > 2\nu_r t^\eta\right) \leq e^{-C_{15}t^\eta} \tag{6.35}$$

as before. Together with (6.34), this implies (6.32) for $i = 1$, with $C_{17} = C_{15}$.

The reasoning for (6.32) when $i = 2$ is the same, except that one now sets $Y(k) = \phi(v_j - S_r^1(k))$, from which one obtains

$$m \stackrel{\text{def}}{=} E[Y(1)] = h_r^*(v_j). \tag{6.36}$$

Also, the coefficient 2 on the RHS of (6.16) is replaced by the coefficient $1 + \varepsilon_5$ in (6.19). Setting $\varepsilon = \varepsilon_5$ in Lemma 6.2, one obtains

$$P_x\left(\sum_{k=1}^{A_r(t)} \phi\big(v_j - S_r^1(k)\big) > (1+\varepsilon_5)\nu_r\big(h_r^*(v_j) \vee t^\eta\big)\right) \leq e^{-C_{15}t^\eta} \tag{6.37}$$

for large t and appropriate $C_{15} > 0$, chosen as in the lemma. Setting $C_{17} = C_{15}$, one obtains (6.32) for $i = 2$ as well. $\square$

Setting $|A(t)| = \sum_r A_r(t)$, where $A_r(t)$ is the number of arrivals at each route by time t, it follows from elementary renewal theory that for appropriate C_{18} and $t \geq 1$,

$$E[|A(t)|^2] \leq C_{18}t^2 \tag{6.38}$$

(see, e.g., [3], page 136). Inequality (6.38) is not difficult to show by applying a standard truncation argument.

Here and later on, we will also use the two inequalities

$$z_r^*(s) \leq b z_r^*((s,\infty)) \qquad \text{for all } s > 0, \tag{6.39}$$

and

$$\Gamma\big(\bar{H}_r^*(N_{H_r}+1)\big)/\,\Gamma(\bar{H}_r^*(N_{H_r})) \geq e^{-b}, \tag{6.40}$$

which follow from the definition of $\phi(\cdot)$ and the second inequality in (3.4). Employing Proposition 6.2 and these inequalities, we now demonstrate Proposition 6.1.

PROOF OF PROPOSITION 6.1. By Hölder's inequality,

$$E_x[|X(t)|_L - |x|_L; \mathcal{A}(t)^c] \leq \sqrt{P_x(\mathcal{A}(t)^c)}\sqrt{E_x\big[\big(|X(t)|_L - |x|_L\big)^2\big]}. \tag{6.41}$$

Also, by Proposition 6.2, one has

$$\sqrt{P_x(\mathcal{A}(t)^c)} \leq \sqrt{N}e^{-C_{11}t^{\eta/2}} \tag{6.42}$$

for all N, x and appropriate $C_{11} > 0$. So it remains to bound the expectation on the RHS of (6.41).

It follows from the definitions of $|\cdot|_L$, $\phi(\cdot)$ and $\Gamma(\cdot)$, and from (3.10), (6.39) and (6.40), that

$$|x'|_L \le \left(\sup_r \frac{w_r}{\nu_r}\right)\frac{2be^b(1+aN)|x'|}{\Gamma(1/N^4)} \le C_{19}N^2|x'| \tag{6.43}$$

for all $x' \in S$ and appropriate C_{19}. So application of (6.10), together with (6.43) for $x' = X^A(t)$, implies that

$$\begin{aligned} E_x\big[\big(|X(t)|_L - |x|_L\big)^2\big] &\le E_x[|X^A(t)|_L^2] \le C_{19}^2 N^4 E_x[|X^A(t)|^2] \\ &\le C_{19}^2 N^4 E_x[|A(t)|^2]. \end{aligned}$$

Together with (6.38), this implies

$$\sqrt{E_x\big[\big(|X(t)|_L - |x|_L\big)^2\big]} \le C_{20}N^2 t \tag{6.44}$$

for appropriate C_{20} and large t.

Substitution of (6.42) and (6.44) into (6.41) implies that for large enough t, (6.21) holds, as desired. $\square$

7. Upper bounds on $|X(N^3)|_{r,s}$ for $s > N_{H_r}$. In Section 6, we obtained upper bounds on $E_x[|X(N^3)|_L - |x|_L; \mathcal{A}(N^3)^c]$; we still need to analyze the behavior of $|X(N^3)|_L - |x|_L$ on $\mathcal{A}(N^3)$. For this, we analyze $|X(N^3)|_{r,s}$ for several cases that depend on whether or not $|x| > N^6$ and $s > N_{H_r}$.

In this section, we consider the case where $|x| > N^6$ and $s > N_{H_r}$, which is the simplest case. The main result here is the following proposition. Recall that $|x|_2$ is defined in (3.24).

PROPOSITION 7.1. *For given $\varepsilon_3 > 0$, large enough N, and $|x| > N^6$ and $|x|_2/|x| \le 1/2$,*

$$\begin{aligned} &E_x\Big[\sup_{r,s>N_{H_r}} |X(N^3)|_{r,s} - |x|_L; G\Big] \\ &\qquad \le C_3(|x|_K/|x|)N^3 + \varepsilon_3 N^2\left(\frac{1}{2} - P(G)\right) \end{aligned} \tag{7.1}$$

for all measurable sets G, with $C_3 > 0$ not depending on N, G or x.

In the proof of Proposition 10.1, we will employ Proposition 7.1 by setting

$$G = \mathcal{A}(N^3) \cap \Big\{\omega : |X(N^3)|_L = \sup_{r,s>N_{H_r}} |X(N^3)|_{r,s}\Big\}. \tag{7.2}$$

Much of the work needed to demonstrate Proposition 7.1 is done in the following proposition. We recall that $i_r(s) = s + \Delta_r$, where $\Delta_r = \Delta_r(N^3)$.

PROPOSITION 7.2. *For given $\varepsilon > 0$, large enough N and all x,*

$$E_x\Big[\sup_{r,s>N_{H_r}} \{|X(N^3)|_{r,s} - |x|_{r,i_r(s)}\}; G\Big] \le C_{21}\varepsilon N^2 \tag{7.3}$$

for all measurable sets G, with C_{21} not depending on ε, N, G or x.

PROOF. We will instead show that

$$E_x\Big[\sup_{r,s>N_{H_r}} |X^A(N^3)|_{r,s}\Big] \le C_{21}\varepsilon N^2. \tag{7.4}$$

Inequality (7.3) follows immediately from this and inequality (6.9) since $|X^A(N^3)|_{r,s} \ge 0$.

To show (7.4), we first note that for all r and s,

$$|X^A(N^3)|_{r,s} \le C_{22}\kappa_{N,r} N_{H_r} Z_r^{A,*}(N^3, s) \tag{7.5}$$

for appropriate C_{22}, where $Z_r^{A,*}(N^3, s) \stackrel{\text{def}}{=} (Z^A)_r^*(N^3, s)$. The inequality uses (3.2) and (6.40). On $s > N_{H_r}$, the RHS of (7.5) is at most

$$C_{22}b\kappa_{N,r} N_{H_r} Z_r^{A,*}(N^3, (s,\infty)) \le C_{22}b\kappa_{N,r} \int_{N_{H_r}}^{\infty} N_r(s') Z_r^{A,*}(N^3, s')\, ds' \tag{7.6}$$

on account of (6.39) and $N_r(s) \ge s$.

On the other hand, by (3.12), the RHS of (7.6) is at most

$$(C_{22}b/M_1)|X^A(N^3)|_R \le C_{21} I_R(N^3), \tag{7.7}$$

where $C_{21} \stackrel{\text{def}}{=} C_{22}b/M_1$ and $I_R(\cdot)$ is as in Section 5. Putting (7.5)–(7.7) together, it follows that, for large N,

$$\sup_{r,s>N_{H_r}} |X^A(N^3)|_{r,s} \le C_{21} I_R(N^3). \tag{7.8}$$

Also, by Proposition 5.2, for given ε, one has that for large enough N,

$$E_x[I_R(N^3)] \le \varepsilon N^2 \qquad \text{for all } x. \tag{7.9}$$

Taking expectations in (7.8) and applying (7.9) implies (7.4). □

In order to demonstrate Proposition 7.1, we need Lemma 7.1, which bounds $|x|_K$ from below in terms of $|x|$ when $(\sup_{r,s\ge N_{H_r}} |x|_{r,s})/|x|_L$ is not small. For the lemma, we require the inequality

$$z_r^*((0, N_{H_r}]) \le C_{23}|x|_L \qquad \text{for } r \in \mathcal{R}, \tag{7.10}$$

for appropriate C_{23}. This is a weaker version of (9.8), which we prove in Lemma 9.2. [Equation (7.10) does not require any additional assumptions on a or b, unlike (9.8).]

If one supposes that $|x|_2 \le |x|/2$, it then follows easily by summing (7.10) over r that

$$|x|_L \ge C_{24}|x| \tag{7.11}$$

for $C_{24} = 1/2C_{23}|\mathcal{R}|$. This inequality will be used in Proposition 7.1 and will also be used in Sections 8 and 9.

LEMMA 7.1. *Suppose that, for some r_0 and $s_0 \ge N_{H_{r_0}}$,*

$$|x|_{r_0,s_0} \ge |x|_L/2. \tag{7.12}$$

Then, for appropriate $\varepsilon_6 > 0$ not depending on N,

$$|x|_K \ge \varepsilon_6|x|/N. \tag{7.13}$$

PROOF. Applying (7.10), and then substituting (7.12) into (3.2), one obtains for given r that

$$\begin{aligned} z_r^*((0, N_{H_r}]) &\le C_{25}Nz_{r_0}^*(s_0)/\Gamma\big(\bar{H}_{r_0}^*(N_{H_{r_0}}+1)\big) \\ &\le C_{25}be^bN\kappa_{N,r_0}z_{r_0}^*((N_{H_{r_0}}, \infty)) \end{aligned} \tag{7.14}$$

for appropriate $C_{25} > 0$, where the second inequality employs the assumption $s_0 \ge N_{H_{r_0}}$, together with (6.39) and (6.40). Addition of $z_r^*((N_{H_r}, \infty))$ to both sides of (7.14) gives

$$\begin{aligned} z_r^*(\mathbb{R}^+) &\le z_r^*((N_{H_r}, \infty)) + C_{25}be^bN\kappa_{N,r_0}z_{r_0}^*((N_{H_{r_0}}, \infty)) \\ &\le (1 + C_{25}be^b)N\sum_{r'}\kappa_{N,r'}z_{r'}^*((N_{H_{r'}}, \infty)). \end{aligned}$$

Summing over r then implies

$$|x| \le \varepsilon_6^{-1}N|x|_K \tag{7.15}$$

with $\varepsilon_6 = [|\mathcal{R}|(1 + C_{25}be^b)]^{-1}$. $\square$

We now apply Proposition 7.2, together with Lemma 7.1 and (7.11), to demonstrate Proposition 7.1.

PROOF OF PROPOSITION 7.1. Suppose first that $|x|_{r_0,s_0} > |x|_L/2$ for some r_0 and $s_0 > N_{H_{r_0}}$. Choosing $\varepsilon > 0$ and C_{21} as in Proposition 7.2, with ε small enough so $\varepsilon < \varepsilon_3/C_{21}$ for given $\varepsilon_3 > 0$, it follows from the proposition and Lemma 7.1 that for large N and any G, the LHS of (7.1) is at most

$$\begin{aligned} C_{21}\varepsilon N^2 \le \varepsilon_3N^2 &\le 2\varepsilon_3\varepsilon_6^{-1}(|x|_K/|x|)N^3 - \varepsilon_3N^2 \\ &\le C_3(|x|_K/|x|)N^3 - \varepsilon_3N^2, \end{aligned} \tag{7.16}$$

if C_3 is chosen to be at least $2\varepsilon_3\varepsilon_6^{-1}$, where ε_6 is as in the lemma. This is at most the RHS of (7.1).

Suppose, on the other hand, that $|x|_{r,s} \le |x|_L/2$ for all $s > N_{H_r}$ and r. Under $|x| > N^6$ and $|x|_2 \le |x|/2$, it follows from (7.11) that $|x|_L \ge C_{24}N^6$. Hence,

$$\sup_{r,s>N_{H_r}} |x|_{r,s} - |x|_L \le -\frac{1}{2}C_{24}N^6. \tag{7.17}$$

Since $i_r(s) \ge s > N_{H_r}$, it follows from Proposition 7.2 and (7.17) that the LHS of (7.1) is at most

$$\begin{aligned} C_{21}\varepsilon N^2 + \sup_{r,s>N_{H_r}} |x|_{r,s} - |x|_L &\le C_{21}\varepsilon N^2 - \frac{1}{2}C_{24}N^6 P(G) \\ &\le \varepsilon_3 N^2\left(\frac{1}{2} - P(G)\right) \end{aligned} \tag{7.18}$$

for large N, if we choose $\varepsilon \le \varepsilon_3/2C_{21}$. This is at most the RHS of (7.1), which completes the proof. □

8. Pathwise upper bounds on $|X(N^3)|_{r,s}$ for $s \le N_{H_r}$ and $\Delta_r > 1/b^3$. In the previous section, we analyzed the behavior of $|X(N^3)|_{r,s} - |x|_L$ for $s > N_{H_r}$. When $s \le N_{H_r}$, we analyze the cases where $\Delta_r \le 1/b^3$ and $\Delta_r > 1/b^3$ separately. The latter case is quicker and we do it in this section, postponing the case $\Delta_r \le 1/b^3$ until Section 9. For both cases, we will require certain pathwise upper bounds on $|X^A(N^3)|_{r,s}$ that hold on $\mathcal{A}_1(N^3)$, which are given in Proposition 8.1. We begin the section with these bounds.

Upper bounds on $|X^A(N^3)|_{r,s}$ on $\mathcal{A}_1(N^3)$. In order to derive bounds on $|X^A(N^3)|_{r,s}$, we first require bounds on $Z_r^{A,*}(\cdot,\cdot)$ that measure how quickly documents with the corresponding service times enter a route r up to a given time. In Lemma 8.1, we provide uniform bounds on $Z_r^{A,*}(t,s)$ for $t \in [0, N^3]$ and $\omega \in \mathcal{A}_1(N^3)$. As in previous sections, $S_r^1(k)$, $k = 1, \dots, A_r(t)$, denotes the positions of the arrivals of documents up to time t. We also denote here by $S_r^2(t,k)$ the amount of service such a document has received by time t; $S_r^1(k) - S_r^2(t,k)$ is therefore the residual service time of the kth document at time t.

LEMMA 8.1. *Suppose $\omega \in \mathcal{A}_1(N^3)$ for some N. Then, for all $r, s \in [0, N+1]$ and $t \in [0, N^3]$,*

$$Z_r^{A,*}(t,s) \le 4b\nu_r\big(\bar{H}_r^*(s)N^3 \vee N^{3\eta}\big). \tag{8.1}$$

If instead $s > N+1$, then

$$Z_r^{A,*}(t,s) \le 2b\nu_r\big(\bar{H}_r^*(N+1)N^3 \vee N^{3\eta}\big). \tag{8.2}$$

PROOF. For all $r, s \in [0, N+1]$ and $t \in [0, N^3]$,

$$
\begin{aligned}
Z_r^{A,*}(t,s) &= \sum_{k=1}^{A_r(t)} \phi\bigl(s - S_r^1(k) + S_r^2(t,k)\bigr) \\
&\le \sum_{k=1}^{A_r(N^3)} \sup_{s' \in [0,\infty)} \phi\bigl(s - S_r^1(k) + s'\bigr) \le b \sum_{k=1}^{A_r(N^3)} \bar{\Phi}\bigl(s - S_r^1(k)\bigr)
\end{aligned}
\tag{8.3}
$$

with the latter inequality employing $\phi(s) \le b\bar{\Phi}(s)$ and the monotonicity of $\bar{\Phi}(\cdot)$. Letting j_0 denote the largest j with $v_j \le s$, the last term in (8.3) is at most

$$
b \sum_{k=1}^{A_r(N^3)} \bar{\Phi}\bigl(v_{j_0} - S_r^1(k)\bigr) \le 2b\nu_r\bigl(\bar{H}_r^*(v_{j_0})N^3 \vee N^{3\eta}\bigr)
\tag{8.4}
$$

on $\mathcal{A}_1(N^3)$. The inequality in (8.2) follows from this, with $j_0 = J$. The inequality in (8.1) follows by applying (6.15) to the RHS of (8.4). □

We now derive uniform upper bounds on $|X^A(t)|_{r,s}$ for $t \in [0, N^3]$ and $\omega \in \mathcal{A}_1(N^3)$. In applications, we will be primarily interested in the behavior at $t = N^3$.

PROPOSITION 8.1. *Suppose $\omega \in \mathcal{A}_1(N^3)$ for some N. Then, for all r and s,*

$$
|X^A(t)|_{r,s} \le C_{26}N^3 \qquad \text{for } t \in [0, N^3] \text{ and all } x,
\tag{8.5}
$$

for appropriate C_{26} not depending on x, N, ω, r or s. In particular,

$$
|X(t)|_L - |x|_L \le C_{26}N^3 \qquad \text{for } t \in [0, N^3] \text{ and all } x.
\tag{8.6}
$$

PROOF. By (6.10),

$$
|X(t)|_L - |x|_L \le |X^A(t)|_L \qquad \text{for all } t,
\tag{8.7}
$$

and so (8.6) follows immediately from (8.5).

We now investigate $|X^A(t)|_{r,s}$. From (3.2) and Lemma 8.1, it follows that, for $t \in [0, N^3]$,

$$
\begin{aligned}
|X^A(t)|_{r,s} &= \frac{w_r(1 + as_N)Z_r^{A,*}(t,s)}{\nu_r \Gamma(\bar{H}_r^*(s_N))} \\
&\le 4bw_rN^3(1 + as_N)\bar{H}_r^*(s_N)/\Gamma(\bar{H}_r^*(s_N)) \\
&\quad + 4bw_rN^{3\eta}(1 + as_N)/\Gamma(\bar{H}_r^*(s_N)).
\end{aligned}
\tag{8.8}
$$

We proceed to analyze the two terms on the RHS of (8.8).

It follows from the definition of $\Gamma(\cdot)$ in (3.8) that, for all s,

$$
(1 + as_N)\bar{H}_r^*(s_N)/\Gamma(\bar{H}_r^*(s_N)) \le (1 + as_N)(\bar{H}_r^*(s_N))^{1-\gamma}/aC_2.
\tag{8.9}
$$

Since by assumption, $\bar{H}_r^*(\cdot)$ has more than two moments and $\gamma \le 1/2$, the RHS of (8.9) goes to 0 as $s_N \to \infty$. Hence, it is bounded for all s_N, which implies that the first term on the RHS of (8.8) is bounded above by $C_{27}N^3$, for some C_{27} not depending on t, r or s.

On the other hand, for all s,

$$\begin{aligned}(1+as_N)/\Gamma(\bar{H}_r^*(s_N)) &\le \big(1+a(N+1)\big)\bar{H}_r^*(N_{H_r}+1)^{-\gamma}/aC_2\\ &\le \big(1+a(N+1)\big)(e^bN^4)^{\gamma}/C_2a.\end{aligned}\tag{8.10}$$

Since $\gamma \le 1/4$, $\eta \le 1/3$ and $aN \ge 1$, the latter term on the RHS of (8.8) is bounded above by $C_{28}N^2$, for some C_{28} not depending on t, r or s.

The above bounds for the two terms on the RHS of (8.8) sum to $(C_{27}+C_{28})N^3$. Setting $C_{26} = C_{27} + C_{28}$, this implies (8.5). □

Upper bounds on $|X(N^3)|_{r,s}$ for $s \le N_{H_r}$ and $\Delta_r > 1/b^3$. Proposition 8.2 gives an upper bound on $|X(N^3)|_{r,s} - |x|_L$ when $s \le N_{H_r}$ and $\Delta_r > 1/b^3$. The proof, which employs Proposition 8.1, is quick.

PROPOSITION 8.2. *Suppose that $|x| > N^6$, with $|x|_2/|x| \le 1/2$. Then, for large enough N,*

$$\sup_{\Delta_r>1/b^3}\ \sup_{s\le N_{H_r}} |X(N^3)|_{r,s} - |x|_L \le -N^4 \qquad \textit{for all } \omega \in \mathcal{A}_1(N^3),\tag{8.11}$$

where N does not depend on x or ω.

PROOF. For each r and s,

$$\begin{aligned}|X(N^3)|_{r,s} - |x|_L &= |X^A(N^3)|_{r,s} - \big(|x|_L - |\tilde{X}(N^3)|_{r,s}\big)\\ &\le C_{26}N^3 - \big(|x|_L - |\tilde{X}(N^3)|_{r,s}\big)\end{aligned}\tag{8.12}$$

with the last line following from Proposition 8.1. We consider two cases, depending on whether $|x|_{r,i_r(s)} > |x|_L/2$ for given r and s.

Suppose first that $|x|_{r,i_r(s)} > |x|_L/2$, with $s \le N_{H_r}$ and $|x|_2 \le |x|/2$. One has

$$|x|_{r,i_r(s)} - |\tilde{X}(N^3)|_{r,s} \ge \frac{w_r}{\nu_r}\cdot\frac{a}{b^3}\cdot\frac{z_r^*(i_r(s))}{\Gamma(\bar{H}_r^*(i_r(s)_N))}.\tag{8.13}$$

To see this, one applies (6.7) to the definition of $|x|_{r,s}$ in (3.2), noting that since $s \le N_{H_r}$,

$$i_r(s)_N - s_N = i_r(s)\wedge(N_{H_r}+1) - s \ge \Delta_r \wedge 1 > 1/b^3,\tag{8.14}$$

and that $\Gamma(\bar{H}_r^*(i_r(s)_N)) \le \Gamma(\bar{H}_r^*(s))$. On account of (3.2) and $|x|_{r,i_r(s)} > |x|_L/2$, one obtains, from the RHS of (8.13),

$$\begin{aligned}&\frac{a}{b^3}\cdot\left(\frac{w_r z_r^*(i_r(s))}{\nu_r\Gamma(\bar{H}_r^*(i_r(s)_N))}\Big/|x|_{r,i_r(s)}\right)\cdot\frac{|x|_{r,i_r(s)}}{|x|_L}\cdot|x|_L\\ &\qquad\ge \frac{a}{b^3}\cdot\big(1+ai_r(s)_N\big)^{-1}\cdot\frac{1}{2}\cdot|x|_L.\end{aligned}$$

Because of $|x|_2 \le |x|/2$, (7.11), $i_r(s)_N \le N$, $|x| > N^6$ and $aN \ge 1$, this is at most $C_{29}N^5$, where $C_{29} > 0$ does not depend on N, x or ω. It follows from (8.13) and the succeeding inequalities that

$$|x|_L - |\tilde{X}(N^3)|_{r,s} \ge |x|_{r,i_r(s)} - |\tilde{X}(N^3)|_{r,s} \ge C_{29}N^5. \tag{8.15}$$

Together with (8.12), this gives the RHS of (8.11).

Suppose, on the other hand, that $|x|_{r,i_r(s)} \le |x|_L/2$, with $|x|_2 \le |x|/2$. Then, by (7.11) and (6.8), the RHS of (8.12) is at most

$$\begin{aligned} &C_{26}N^3 - \tfrac{1}{2}|x|_L - \big(|x|_{r,i_r(s)} - |\tilde{X}(N^3)|_{r,s}\big) \\ &\quad \le C_{26}N^3 - \tfrac{1}{2}C_{24}N^6 \le -N^5 \end{aligned} \tag{8.16}$$

for large N. This implies (8.11) for $|x|_{r,i_r(s)} \le |x|_L/2$, and hence completes the proof. $\square$

9. Pathwise upper bounds on $|X(N^3)|_{r,s}$ for $s \le N_{H_r}$ and $\Delta_r \le 1/b^3$. In Sections 7 and 8, we analyzed the behavior of $|X(N^3)|_{r,s} - |x|_L$ for $s > N_{H_r}$, and for $s \le N_{H_r}$ with $\Delta_r > 1/b^3$. There remains the case $s \le N_{H_r}$ with $\Delta_r \le 1/b^3$, which is the subject of this section. This is, in essence, the "main case" one needs to show in order to establish the stability of the network since the other cases dealt with less sensitive behavior and did not employ the subcriticality of the system that was given in (1.2). The same was also true for the computations of the $|\cdot|_A$ and $|\cdot|_R$ norms in Sections 4 and 5.

Section 9 consists of three subsections. First, in Proposition 9.2, we give lower bounds on the minimal service rates $\lambda^w(\cdot)$ of documents in terms of the norm $|\cdot|_L$. In the next subsection, we begin our analysis of $|X(N^3)|_{r,s}$ for $s \le N_{H_r}$ and $\Delta_r \le 1/b^3$. We decompose $|X(N^3)|_{r,s} - |x|_{r,i_r(s)}$ into several parts that are easier to analyze. In Proposition 9.3, we then obtain upper bounds on the factor $Z_r^*(N^3, s) - z_r^*(i_r(s))$ of one of the parts. In the third subsection, we do a detailed analysis of the decomposition from the previous subsection, which also employs the bounds on $\lambda^w(\cdot)$ from the first subsection. From this, we obtain in Proposition 9.5 the desired bound on $|X(N^3)|_{r,s} - |x|_L$. We note that, whereas in Section 8, our results pertained to $\omega \in \mathcal{A}_1(N^3)$, starting from the second subsection here, we require $\omega \in \mathcal{A}_2(t)$. Our final results on $|X(N^3)|_{r,s} - |x|_L$, for $s \le N_{H_r}$, will therefore be valid on $\mathcal{A}(N^3) = \mathcal{A}_1(N^3) \cap \mathcal{A}_2(N^3)$.

Lower bounds on $\lambda^w(\cdot)$. In order to demonstrate the stability of the network, its subcriticality needs to be employed at some point. With this in mind, we choose $\varepsilon_7 \in (0, 1]$ small enough so that

$$(1+\varepsilon_7)^2 \sum_{r\in\mathcal{R}} A_{l,r}\rho_r \le c_l \qquad \text{for all } l, \tag{9.1}$$

which is possible because of (1.2). We henceforth assume $\varepsilon_5 \le \varepsilon_7/4$, where ε_5 was employed in (6.19) in the definition of $\mathcal{A}_2(\cdot)$.

The main results in this subsection are Propositions 9.1 and 9.2. Proposition 9.1 gives a lower bound on $\lambda^w(t)$ in terms of $|X(t)|_S$; Proposition 9.2, under additional assumptions, gives the bound in terms of $|x|_L$.

PROPOSITION 9.1. *Assume* (9.1) *holds for some* $\varepsilon_7 > 0$. *Then*, *for large enough* b *and small enough* a,

$$\lambda^w(t) \geq (1+\varepsilon_7)/|X(t)|_S \tag{9.2}$$

for almost all t.

In this and the previous subsection, we need to employ certain properties of $\Gamma(\bar{H}_r^*(\cdot))$, which appears in the denominator in (3.2). In Lemma 9.1, we state two such properties; the first is employed for Lemma 9.3 and the second is employed for Lemma 9.2. Recall that m_r is the mean of $H_r(\cdot)$.

LEMMA 9.1. *For* $\Gamma(\cdot)$ *as defined in* (3.8),

$$\Gamma'(\bar{H}_r^*(s)) \geq 1 + as \qquad \text{for all } r \text{ and } s. \tag{9.3}$$

Moreover, *for large enough* b *and small enough* a,

$$\int_0^\infty \frac{\Gamma(\bar{H}_r^*(s))}{1+as}\,ds \leq (1+\varepsilon_7)m_r \tag{9.4}$$

for $\varepsilon_7 > 0$ *satisfying* (9.1).

PROOF. By (3.8) and then (3.7), one has, for all r and s,

$$\begin{aligned} \Gamma'(\bar{H}_r^*(s)) &= 1 + C_2\gamma a(\bar{H}_r^*(s))^{\gamma-1} \\ &\geq 1 + C_2 C_1^{\gamma-1}\gamma a(1+s)^{(1-\gamma)(2+\delta_1)} \geq 1 + as, \end{aligned} \tag{9.5}$$

where the last inequality uses $\gamma \leq 1/2$ and $C_2 \geq {C_1}^{(1-\gamma)/\gamma}$. This implies (9.3).

For (9.4), we note from (3.8) and (3.7) that

$$\int_0^\infty \frac{\Gamma(\bar{H}_r^*(s))}{1+as}\,ds \leq \int_0^\infty \bar{H}_r^*(s)\,ds + C_1^2 a\int_0^\infty (1+s)^{-2\gamma}(1+as)^{-1}\,ds. \tag{9.6}$$

The constant b can be chosen large enough so the first term on the RHS of (9.6) is at most $(1+\varepsilon_7/2)m_r$. Also, by choosing $a > 0$ small enough, since the second term can be chosen as close to 0 as desired, by monotone convergence,

$$a\int_0^\infty (1+s)^{-2\gamma}(1+as)^{-1}\,ds = \int_0^\infty (1+s)^{-(1+2\gamma)}\frac{1+s}{1/a+s}\,ds \to 0 \tag{9.7}$$

as $a \searrow 0$. So, for large enough b and small enough a, (9.4) holds. □

By employing (9.4), we obtain upper bounds for $z_r^*((0, N_{H_r}])$ and $z_r^*(\mathbb{R}^+)$ in terms of $|x|_L$ and $|x|_S$. Inequality (9.9) will be crucial for Proposition 9.1.

LEMMA 9.2. *For large enough b and small enough a,*

$$z_r^*((0, N_{H_r}]) \le (1+\varepsilon_7) w_r^{-1} \rho_r |x|_L \tag{9.8}$$

and

$$z_r^*(\mathbb{R}^+) \le (1+\varepsilon_7) w_r^{-1} \rho_r |x|_S \tag{9.9}$$

for all N and r, where $\varepsilon_7 > 0$ is as in (9.1).

PROOF. We note that by (3.2),

$$z_r^*((0, N_{H_r}]) = \int_0^{N_{H_r}} z_r^*(s)\, ds \le w_r^{-1} \nu_r |x|_L \int_0^{N_{H_r}} \frac{\Gamma(\bar{H}_r^*(s))}{1+as}\, ds. \tag{9.10}$$

By (9.4), for large enough b and small enough a, the last term in (9.10) is at most

$$(1+\varepsilon_7) w_r^{-1} \nu_r m_r |x|_L = (1+\varepsilon_7) w_r^{-1} \rho_r |x|_L \tag{9.11}$$

for all N and r, which implies (9.8). It follows from (9.8) and the definition of $|\cdot|_S$ in (3.25) that

$$\begin{aligned} z_r^*(\mathbb{R}^+) &\le (1+\varepsilon_7) w_r^{-1} \rho_r \left[|x|_L + \frac{w_r}{\rho_r} z_r^*((N_{H_r}, \infty)) \right] \\ &\le (1+\varepsilon_7) w_r^{-1} \rho_r |x|_S, \end{aligned}$$

which implies (9.9). □

A weaker version of the bound (9.8) was used in (7.10), where the RHS of (9.8) was replaced by $C_{23}|x|_L$, and no additional assumptions on b and a were required. This follows by noting that the second term on the RHS of (9.6) does not depend on a (since $a \le 1$).

We now demonstrate Proposition 9.1.

PROOF OF PROPOSITION 9.1. On account of (9.1), a feasible protocol is given by assigning service to each nonempty route r at rate $\Lambda_{r,F} \stackrel{\text{def}}{=} (1+\varepsilon_7)^2 \rho_r$. By (9.9), the rate at which each document is served is

$$\lambda_{r,F} = \frac{(1+\varepsilon_7)^2 \rho_r}{Z_r(t, \mathbb{R}^+)} = \frac{(1+\varepsilon_7)^2 \rho_r}{Z_r^*(t, \mathbb{R}^+)} \ge \frac{(1+\varepsilon_7) w_r}{|X(t)|_S} \tag{9.12}$$

at almost all times t. It follows from this and the definition of the weighted max–min fair protocol that

$$\lambda^w(t) = \min_{r \in \mathcal{R}'} \frac{\lambda_r(t)}{w_r} \ge \min_{r \in \mathcal{R}'} \frac{\lambda_{r,F}}{w_r} \ge \frac{(1+\varepsilon_7)}{|X(t)|_S} \qquad \text{for almost all } t,$$

which implies (9.2). □

We apply Proposition 9.1 to derive the following lower bound of $\lambda_r(t)$ on $[0, N^3]$. We note that, by (8.6) of Proposition 8.1 and (7.11), for $\omega \in \mathcal{A}_1(N^3)$, $|x| > N^6$ and $|x|_2 \le |x|/2$,

$$|X(t)|_L \le |x|_L + C_{26}N^3 \le (1+\varepsilon)|x|_L \tag{9.13}$$

holds for given $\varepsilon > 0$ and large enough N. In the proposition, we will use

$$\varepsilon_8 \stackrel{\text{def}}{=} \left[\frac{C_{24}}{8}\left(\max_r \frac{w_r}{\rho_r}\right)^{-1}\varepsilon_7\right] \wedge \frac{1}{2}.$$

PROPOSITION 9.2. *Suppose that* (9.1) *holds for some* $\varepsilon_7 \in (0, 1]$, *and that* $|x| > N^6$, *with* $|x|_2 \le \varepsilon_8|x|$. *Then, for large enough* N *and* b, *and small enough* a,

$$\lambda^w(t) \ge (1+\varepsilon_7/2)/|x|_L \tag{9.14}$$

for almost all $t \in [0, N^3]$ *on* $\omega \in \mathcal{A}_1(N^3)$.

PROOF. It follows from Proposition 9.1 that

$$\lambda^w(t) \ge (1+\varepsilon_7)/|X(t)|_S \qquad \text{almost everywhere,} \tag{9.15}$$

for large enough b and small enough a. On the other hand, it follows from (3.25), (9.13), (6.12) and (6.17) that, since $|x| > N^6$ and $|x|_2 \le \varepsilon_8|x|$,

$$\begin{aligned} |X(t)|_S &\le |X(t)|_L + \left(\max_r \frac{w_r}{\rho_r}\right) Z_r^*(t, (N_{H_r}, \infty)) \\ &\le (1+\varepsilon)|x|_L + \left(\max_r \frac{w_r}{\rho_r}\right)\left[|x|_2 + 2\left(\max_r \nu_r\right)N^3\right] \end{aligned} \tag{9.16}$$

holds for given $\varepsilon > 0$ and large enough N, for all $\omega \in \mathcal{A}_1(N^3)$ and $t \in [0, N^3]$. Applying $|x| > N^6$, $|x|_2 \le \varepsilon_8|x|$ and (7.11) to the RHS of (9.16) implies that it is at most

$$\left(1+\varepsilon+\frac{\varepsilon_7}{8}\right)|x|_L + 2\left(\max_r \frac{w_r}{\rho_r}\right)\left(\max_r \nu_r\right)|x|_L/C_{24}N^2.$$

Consequently, for small enough $\varepsilon > 0$,

$$|X(t)|_S \le (1+\varepsilon_7/4)|x|_L \qquad \text{for all } t \in [0, N^3]. \tag{9.17}$$

Together with (9.15), this implies (9.14). □

Decomposition of $|X(N^3)|_{r,s} - |x|_{r,i_r(s)}$. In this short subsection, we decompose $|X(N^3)|_{r,s} - |x|_{r,i_r(s)}$ into several parts, one of which contains the factor $Z_r^*(N^3, s) - z_r^*(i_r(s))$. In Proposition 9.3, we then obtain upper bounds on this factor. In this and the remaining subsection, the estimates need to be more precise than in previous sections in order to make use of the subcriticality of $X(\cdot)$.

The decomposition that was referred to above is given by

$$
\begin{aligned}
&|X(N^3)|_{r,s} - |x|_{r,i_r(s)} \\
&\quad = \frac{w_r(1+as)(Z_r^*(N^3,s) - z_r^*(i_r(s)))}{\nu_r\Gamma(\sigma_r)} \\
&\qquad - |x|_{r,i_r(s)}\frac{1+as}{1+ai_r(s)}\frac{\Gamma(\sigma_r)-\Gamma(\sigma_r')}{\Gamma(\sigma_r)} - \frac{aw_r\Delta_r z_r^*(i_r(s))}{\nu_r\Gamma(\sigma_r')},
\end{aligned}
\tag{9.18}
$$

and holds for $s \le N_{H_r}$ and $\Delta_r \le 1/b^3$. It will be employed in Corollary 9.1. Here and later on, we abbreviate, setting $\sigma_r = \bar{H}_r^*(s)$ and $\sigma_r' = \bar{H}_r^*(i_r(s))$. [One can check that (9.18) holds as given, without employing either s_N or $i_r(s)_N$, as in (3.2), since $i_r(s) = s + \Delta_r \le N_{H_r} + 1$, and hence $s_N = s$ and $i_r(s)_N = i_r(s)$.]

To apply the bound (6.19) on $\omega \in \mathcal{A}_2(N^3)$ and derive an upper bound on $Z_r^*(N^3,s) - z_r^*(i_r(s))$, we need to select a v_j from among $v_0,\dots,v_J$, as given by (6.14). For this, we denote by $v(s)$ the value v_j with

$$
v_j \in [i_r(s), i_r(s) + 1/b^3). \tag{9.19}
$$

Under $s \le N_{H_r}$ and $\Delta_r \le 1/b^3$, such a $v(s)$ exists.

PROPOSITION 9.3. *Suppose $\omega \in \mathcal{A}_2(N^3)$, for some N and b, with b as in* (3.3). *Then*,

$$
Z_r^*(N^3,s) - z_r^*(i_r(s)) \le (1+\varepsilon_5)(1+4/b^2)\nu_r[h_r^*(v(s))N^3 \vee N^{3\eta}] \tag{9.20}
$$

for all r and s with $\Delta_r \le 1/b^3$ and $s \le N_{H_r}$, where $\varepsilon_5 > 0$ is as in (6.19) *and $v(s)$ is given by* (9.19).

PROOF. By (6.7), the LHS of (9.20) is at most $Z_r^{A,*}(N^3,s)$. For $s \le N_{H_r}$, this equals

$$
\begin{aligned}
\sum_{k=1}^{A_r(N^3)} \phi\big(s - S_r^1(k) + S_r^2(N^3,k)\big) &\le e^{2/b^2}\sum_{k=1}^{A_r(N^3)} \phi\big(v(s) - S_r^1(k)\big) \\
&\le (1+4/b^2)\sum_{k=1}^{A_r(N^3)} \phi\big(v(s) - S_r^1(k)\big).
\end{aligned}
\tag{9.21}
$$

To see (9.21), we note that since $S_r^2(N^3,k) \le \Delta_r \le 1/b^3$,

$$
v_j - S_r^1(k) \in [s - S_r^1(k) + S_r^2(N^3,k), s - S_r^1(k) + S_r^2(N^3,k) + 2/b^3]. \tag{9.22}
$$

Together with the second half of (3.4), this implies the first inequality. The second inequality follows by expanding e^{2/b^2}. Since $\omega \in \mathcal{A}_2(N^3)$, the RHS of (9.20) then follows by applying (6.19). $\square$

In the next subsection, we will also employ the following bound on $h_r^*(s_2) - h_r^*(s_1)$ for $s_1 \le s_2$.

PROPOSITION 9.4. *For any* r, $s_1 \le s_2$ *and* b,

$$h_r^*(s_2) - h_r^*(s_1) \le eb^2(s_2 - s_1)\bar{H}_r^*(s_1). \tag{9.23}$$

PROOF. Since $h_r^*(s) = \int_0^\infty \phi(s - s')\,dH_r(s')$ for each s, the LHS of (9.23) equals

$$\int_0^\infty (\phi(s_2 - s') - \phi(s_1 - s'))\,dH_r(s'). \tag{9.24}$$

By the first part of (3.4) and the definition of $\phi(\cdot)$, $\phi'(s) \le b^2$ for all s and $\phi(\cdot)$ is decreasing on $[1/b, \infty)$. So, (9.24) is at most

$$\begin{aligned}\int_0^\infty b^2(s_2 - s_1)1\{s' > s_1 - 1/b\}\,dH(s') &\le b^2(s_2 - s_1)\bar{H}_r(s_1 - 1/b)\\ &\le b^2(s_2 - s_1)\bar{H}_r^*(s_1 - 1/b)\\ &\le eb^2(s_2 - s_1)\bar{H}_r^*(s_1).\end{aligned} \tag{9.25}$$

□

Upper bounds on $|X(N^3)|_{r,s}$. In this subsection, we employ the previous two subsections to obtain upper bounds on $|X(N^3)|_{r,s} - |x|_L$ for $\omega \in \mathcal{A}(N^3)$, when $s \le N_{H_r}$ and $\Delta_r \le 1/b^3$. Our main result is the following proposition. As elsewhere in this paper, we are assuming that $aN \ge 1$.

PROPOSITION 9.5. *Suppose that* (9.1) *holds for some* $\varepsilon_7 \in [0, 1]$ *and that* $|x| > N^6$, *with* $|x|_2 \le \varepsilon_8 |x|$, *where* ε_8 *is specified below* (9.13). *Then*, *for large enough* N *and* b, *and small enough* a,

$$|X(N^3)|_{r,s} - |x|_L \le -\tfrac{1}{2} w_r N^2 \tag{9.26}$$

for $\omega \in \mathcal{A}(N^3)$, *and all* r *and* s *with* $\Delta_r \le 1/b^3$ *and* $s \le N_{H_r}$.

Our main step in demonstrating Proposition 9.5 will be to demonstrate the following proposition.

PROPOSITION 9.6. *Under the same assumptions as in Proposition* 9.5,

$$\begin{aligned}\frac{w_r(1 + as)(Z_r^*(N^3, s) - z_r^*(i_r(s)))}{v_r \Gamma(\sigma_r)} &\le |x|_L \cdot \frac{1 + as}{1 + ai_r(s)} \cdot \frac{\Gamma(\sigma_r) - \Gamma(\sigma_r')}{\Gamma(\sigma_r)}\\ &\quad + \frac{C_{30} w_r N^3}{ab(1 + as)} + C_{31} w_r N^{3/2}\end{aligned} \tag{9.27}$$

for appropriate C_{30} *and* C_{31} *not depending on* w, N, a, b, r *or* s.

In order to demonstrate Proposition 9.6, we note that, on account of Proposition 9.3, the LHS of (9.27) is, under the assumptions for the latter proposition, at most

$$\begin{aligned}
&d_r(s)\big(h_r^*(v(s))N^3 \vee N^{3\eta}\big) \\
&\qquad \le d_r(s)\Big(\inf_{s'\in[s,i_r(s)]} h_r^*(s')\Big)N^3 \\
&\qquad\qquad + d_r(s)\Big(h_r^*(v(s)) - \inf_{s'\in[s,i_r(s)]} h_r^*(s')\Big)N^3 + d_r(s)N^{3\eta},
\end{aligned} \tag{9.28}$$

where

$$d_r(s) = (1+\varepsilon_5)(1+4/b^2)w_r(1+as)/\Gamma(\sigma_r).$$

We will show in Lemmas 9.3, 9.4 and 9.5 that each of the three terms on the RHS of (9.28) is bounded above by the corresponding term on the RHS of (9.27). Proposition 9.6 then follows.

We first show Lemma 9.3, which applies to the first term on the RHS of (9.28), and should be thought of as the "main term" there.

LEMMA 9.3. *Under the same assumptions as in Proposition 9.5,*

$$d_r(s)\Big(\inf_{s'\in[s,i_r(s)]} h_r^*(s')\Big)N^3 \le |x|_L \frac{1+as}{1+ai_r(s)} \cdot \frac{\Gamma(\sigma_r)-\Gamma(\sigma_r')}{\Gamma(\sigma_r)}. \tag{9.29}$$

PROOF. It follows from Proposition 9.2 that

$$\lambda^w(t) \ge (1+\varepsilon_7/2)/|x|_L \qquad \text{for almost all } t\in[0,N^3], \tag{9.30}$$

for large enough N and b, and small enough a, and therefore

$$\Delta_r \ge (1+\varepsilon_7/2)w_r N^3/|x|_L \qquad \text{for all } r. \tag{9.31}$$

Consequently, the LHS of (9.29) is at most

$$\begin{aligned}
&d_r(s)|x|_L\Big(\inf_{s'\in[s,i_r(s)]} h_r^*(s')\Big)\Delta_r/w_r(1+\varepsilon_7/2) \\
&\qquad \le d_r(s)|x|_L\big(\bar{H}_r^*(s) - \bar{H}_r^*(i_r(s))\big)/w_r(1+\varepsilon_7/2).
\end{aligned} \tag{9.32}$$

This last quantity can be rewritten as

$$\begin{aligned}
&\frac{(1+\varepsilon_5)(1+4/b^2)}{1+\varepsilon_7/2}\cdot|x|_L\cdot\frac{1+as}{1+ai_r(s)}\cdot\frac{\Gamma(\sigma_r)-\Gamma(\sigma_r')}{\Gamma(\sigma_r)} \\
&\qquad \times \frac{1+ai_r(s)}{(\Gamma(\sigma_r)-\Gamma(\sigma_r'))/(\sigma_r-\sigma_r')}.
\end{aligned} \tag{9.33}$$

We proceed to bound the components of (9.33). Since $\varepsilon_5 \le \varepsilon_7/4$, one has for large enough b, depending on ε_7, that

$$\frac{(1+\varepsilon_5)(1+4/b^2)}{1+\varepsilon_7/2} \le (1+1/b^2)^{-1}. \tag{9.34}$$

Since $\Gamma(\cdot)$ is concave and $\sigma_r > \sigma_r'$,

$$\frac{\Gamma(\sigma_r)-\Gamma(\sigma_r')}{\sigma_r-\sigma_r'} \ge \Gamma'(\sigma_r) \ge 1+as, \tag{9.35}$$

with the second inequality holding on account of (9.3). So the last term in (9.33) is at most

$$\frac{1+ai_r(s)}{1+as} = 1+\frac{a\Delta_r}{1+as} \le 1+1/b^3, \tag{9.36}$$

where the inequality uses $\Delta_r \le 1/b^3$. Consequently, (9.33) is, for large b, at most

$$(1+1/b^2)^{-1}(1+1/b^3)|x|_L \frac{1+as}{1+ai_r(s)} \cdot \frac{\Gamma(\sigma_r)-\Gamma(\sigma_r')}{\Gamma(\sigma_r)},$$

which is at most as large as the RHS of (9.29). This implies the lemma. $\square$

We next demonstrate Lemma 9.4, which applies to the second term on the RHS of (9.28).

LEMMA 9.4. *For all r and s with $\Delta_r \le 1/b^3$ and $s \le N_{H_r}$,*

$$d_r(s)\Big(h_r^*(v(s)) - \inf_{s' \in [s, i_r(s)]} h_r^*(s')\Big)N^3 \le \frac{C_{30} w_r N^3}{ab(1+as)} \tag{9.37}$$

for appropriate C_{30} not depending on w, N, a, b, r or s.

PROOF. Since $v(s) - s \le 2/b^3$, it follows from Proposition 9.4 that the LHS of (9.37) is at most

$$(1+\varepsilon_5)(1+4/b^2)\frac{2eb^2}{b^3} w_r N^3 \bar{H}_r^*(s)\frac{(1+as)}{\Gamma(\sigma_r)}. \tag{9.38}$$

On account of (3.8), since $\gamma \le \delta_1/4$, $b \ge 2$ and $\varepsilon_5 \le 1$, this is at most

$$\begin{aligned} \frac{24w_r}{C_2ab} N^3 (\bar{H}_r^*(s))^{1-\gamma}(1+as) &\le \frac{24C_1w_r}{C_2ab} N^3(1+as)^{1-(1-\gamma)(2+\delta_1)} \\ &\le \frac{24C_1w_rN^3}{C_2ab(1+as)}. \end{aligned} \tag{9.39}$$

Recall that C_1 and C_2 do not depend on w, N, a, b, r or s. The RHS of (9.37) follows from this last term by setting $C_{30} = 24C_1/C_2$. $\square$

We now demonstrate Lemma 9.5, which applies to the third term on the RHS of (9.28).

LEMMA 9.5. *For all $s \le N_{H_r}$,*

$$d_r(s)N^{3\eta} \le C_{31} w_r N^{3/2} \tag{9.40}$$

for appropriate C_{31} not depending on w, N, a, b, r or s.

PROOF. Since $s \le N_{H_r} \le N$, $\gamma \le 1/24$, $\eta \le 1/12$, $b \ge 2$ and $\varepsilon_5 \le 1$, it follows from (3.8) and (3.10) that the LHS of (9.40) is at most

$$\frac{4w_r N^{3\eta}(1+aN)}{C_2 a \bar{H}_r^*(N_{H_r})^\gamma} \le \frac{4}{C_2 a} w_r N^{1/2}(1+aN). \tag{9.41}$$

Since $aN \ge 1$, this is at most $8w_r N^{3/2}/C_2$, which gives the RHS of (9.40) for $C_{31} = 8/C_2$. $\square$

Proposition 9.6 follows by applying Lemmas 9.3, 9.4 and 9.5 to (9.28).

We will apply the following corollary of the proposition to Proposition 9.5. The corollary combines the inequality (9.27) with (9.18).

COROLLARY 9.1. *Under the same assumptions as in Propositions 9.5 and 9.6,*

$$|X(N^3)|_{r,s} - |x|_L \le \frac{C_{30} w_r N^3}{ab(1+as)} + C_{31} w_r N^{3/2} - \frac{a w_r \Delta_r z_r^*(i_r(s))}{\nu_r \Gamma(\sigma_r')} \tag{9.42}$$

for appropriate C_{30} and C_{31} not depending on w, N, a, b, r or s.

PROOF. The first term on the RHS of (9.27) of Proposition 9.6 is at most

$$|x|_{r,i_r(s)} \frac{1+as}{1+ai_r(s)} \cdot \frac{\Gamma(\sigma_r) - \Gamma(\sigma_r')}{\Gamma(\sigma_r)} + |x|_L - |x|_{r,i_r(s)} \tag{9.43}$$

since the coefficients of $|x|_{r,i_r(s)}$ in the first term in (9.43) are at most 1. Substituting (9.43) into (9.27) and then applying the resulting inequality to the RHS of (9.18), we note that the term on the LHS of (9.27) is the first term on the RHS of (9.18) and the first term in (9.43) is the negative of the second term on the RHS of (9.18). After the resulting cancellation, the last two terms on the RHS of (9.27), together with the last term on the RHS of (9.18), give the RHS of (9.42). $\square$

In order to show Proposition 9.5, we will need a lower bound on the last term on the RHS of (9.42) and an upper bound on each of the first two terms. In the following lemma, we obtain the former. Note that the assumptions in the lemma are those of Proposition 9.2, with the additional assumption that

$$|x|_{r,i_r(s)} \ge |x|_L/(1+\varepsilon_7/2) \qquad \text{for some } s \le N_{H_r}. \tag{9.44}$$

LEMMA 9.6. *Suppose that* (9.1) *holds for some* $\varepsilon_7 \in (0,1]$, *that* $|x| > N^6$ *with* $|x|_2 \le \varepsilon_8|x|$, *and that* (9.44) *is satisfied for a given* s. *Then, for large enough* N *and* b, *and small enough* a,

$$\frac{\Delta_r z_r^*(i_r(s))}{\nu_r\Gamma(\sigma_r')} \ge \frac{N^3}{1+ai_r(s)} \qquad \text{on } \omega \in \mathcal{A}_1(N^3). \tag{9.45}$$

PROOF. By Proposition 9.2,

$$\lambda^w(t) \ge (1+\varepsilon_7/2)/|x|_L \qquad \text{for almost all } t \in [0,N^3]. \tag{9.46}$$

Consequently,

$$\Delta_r \ge (1+\varepsilon_7/2)w_r N^3/|x|_L \qquad \text{for all } r. \tag{9.47}$$

It follows from (9.47), (3.2) and (9.44) that the LHS of (9.45) is at least

$$\frac{(1+\varepsilon_7/2)w_r N^3 z_r^*(i_r(s))}{\nu_r\Gamma(\sigma_r')|x|_L} = \frac{(1+\varepsilon_7/2)N^3|x|_{r,i_r(s)}}{(1+ai_r(s))|x|_L} \ge \frac{N^3}{1+ai_r(s)}. \qquad \square \tag{9.48}$$

We now apply Corollary 9.1 and Lemma 9.6 to demonstrate Proposition 9.5.

PROOF OF PROPOSITION 9.5. We will consider two cases for a given $s \le N_{H_r}$, depending on whether (9.44) holds. Suppose it does. Then, by Lemma 9.6,

$$\frac{aw_r\Delta_r z_r^*(i_r(s))}{\nu_r\Gamma(\sigma_r')} \ge \frac{aw_r N^3}{1+ai_r(s)}, \tag{9.49}$$

which is a lower bound for the third term on the RHS of (9.42).

On the other hand, if one chooses $b \ge 8C_{30}/a^2$, then, since $a \le 1$ and $\Delta_r \le 1/b^3 \le 1$, the first term on the RHS of (9.42) satisfies

$$\frac{C_{30}w_r N^3}{ab(1+as)} \le \frac{aw_r N^3}{4(1+ai_r(s))}, \tag{9.50}$$

which is 1/4 of the RHS of (9.49). Since $s \le N_{H_r}$, $i_r(s) \le N+1$. So, the sum of the first and third terms on the RHS of (9.42) is, for large N, at most

$$-\frac{3aw_r N^3}{4(1+ai_r(s))} \le -\frac{5}{8}w_r N^2. \tag{9.51}$$

The second term on the RHS of (9.42) satisfies

$$C_{31}w_r N^{3/2} \le \tfrac{1}{8}w_r N^2$$

for large N. Combining this with (9.51), one obtains from Corollary 9.1 that

$$|X(N^3)|_{r,s} - |x|_L \le -\tfrac{1}{2}w_r N^2,$$

which implies (9.26) under (9.44).

When (9.44) fails for s, one has, for large N,

$$
\begin{aligned}
|X(N^3)|_{r,s} - |x|_L &= \big(|X(N^3)|_{r,s} - |x|_{r,i_r(s)}\big) - \big(|x|_L - |x|_{r,i_r(s)}\big) \\
&\le |X(N^3)|_{r,s} - |x|_{r,i_r(s)} - \tfrac{1}{4}\varepsilon_7|x|_L \\
&\le |X^A(N^3)|_{r,s} - \tfrac{1}{4}C_{24}\varepsilon_7|x| \\
&\le C_{26}N^3 - \tfrac{1}{4}C_{24}\varepsilon_7 N^6 \le -N^5,
\end{aligned}
\tag{9.52}
$$

where, in the second inequality, we applied (6.9) and (7.11), and in the third inequality, we applied (8.5) of Proposition 8.1 and $|x| > N^6$. This implies (9.26) when (9.44) fails. □

10. Conclusion: Upper bounds on $E_x[|X(N^3)|_L]$. In the preceding four sections, we obtained upper bounds on

$$|X(N^3)|_{r,s} - |x|_L \quad \text{and} \quad E_x[|X(N^3)|_L - |x|_L; \mathcal{A}(N^3)^c]$$

under various assumptions. In Propositions 6.1 and 6.2 we showed that $P_x(\mathcal{A}(N^3)^c)$ and the corresponding expectation $E_x[|X(N^3)|_L; \mathcal{A}(N^3)^c]$ are small. In Proposition 7.1, we showed that the expected value of $|X(N^3)|_{r,s} - |x|_L$ is small for $s > N_{H_r}$. In Sections 8 and 9, we obtained pathwise estimates on $\mathcal{A}(N^3)$ when $s \le N_{H_r}$, depending on whether $\Delta_r > 1/b^3$ or $\Delta_r \le 1/b^3$. Proposition 8.2 gives an upper bound in the former subcase and Proposition 9.5 gives an upper bound in the latter subcase. Except for Propositions 6.1 and 6.2, we assumed that $|x| > N^6$; for the different results, we also required various side conditions.

We tie these results together in Proposition 10.2 to obtain inequality (3.30) that was cited earlier. We do this in several steps, first combining the results for $s \le N_{H_r}$, then combining these with Proposition 7.1 for $s > N_{H_r}$, and lastly including the bound from Proposition 6.1 on $\mathcal{A}(N^3)^c$. The first two steps are done in Proposition 10.1. As elsewhere in the paper, $aN \ge 1$ is assumed.

PROPOSITION 10.1. *Suppose that* (9.1) *holds for some $\varepsilon_7 \in (0, 1]$ and that $|x| > N^6$, with $|x|_2 \le \varepsilon_8|x|$, where ε_8 is specified below* (9.13). *Then, for large enough N and b, and small enough a,*

$$|X(N^3)|_{r,s} - |x|_L \le -\tfrac{1}{2}w_r N^2 \tag{10.1}$$

for $\omega \in \mathcal{A}(N^3)$, all r, and s with $s \le N_{H_r}$. Moreover, for large enough N and b, and small enough a,

$$
\begin{aligned}
&E_x[|X(N^3)|_L - |x|_L; \mathcal{A}(N^3)] \\
&\quad \le C_3(|x|_K/|x|)N^3 - \left(\frac{1}{4}\min_r w_r\right)N^2\big(2P_x(\mathcal{A}(N^3)) - 1\big).
\end{aligned}
\tag{10.2}
$$

Note that the assumptions for the first half of Proposition 10.1 are the same as for Proposition 9.5, except that the restriction that $\Delta_r \le 1/b^3$ has been removed.

PROOF OF PROPOSITION 10.1. Inequality (9.26) in Proposition 9.5 covers the case where $\Delta_r \le 1/b^3$; (8.11) of Proposition 8.2 covers the case where $\Delta_r > 1/b^3$. Together, they imply (10.1).

In order to demonstrate (10.2), we partition $\mathcal{A}(N^3)$ into $G \cup H$, with

$$G = \Big\{\omega : |X(N^3)|_L = \sup_{r,s > N_{H_r}} |X(N^3)|_{r,s}\Big\}.$$

Applying Proposition 7.1 to this G, with $\varepsilon_3 = (\min_r w_r)/2$, and applying (10.1) on H, it follows that the LHS of (10.2) equals

$$\begin{aligned}
&E_x\Big[\sup_{r,s > N_{H_r}} |X(N^3)|_{r,s} - |x|_L; G\Big] + E_x\Big[\sup_{r,s \le N_{H_r}} |X(N^3)|_{r,s} - |x|_L; H\Big] \\
&\quad \le C_3(|x|_K/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^2\big(2P_x(G) + 2P_x(H) - 1\big) \qquad (10.3)\\
&\quad = C_3(|x|_K/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^2\big(2P_x(\mathcal{A}(N^3)) - 1\big).
\end{aligned}$$

This implies (10.2). □

We now obtain our desired result, Proposition 10.2, which gives upper bounds on $E_x[|X(N^3)|_L] - |x|_L$. The first part of the proposition applies to all x; the second part requires that $|x| > N^6$.

PROPOSITION 10.2. *Suppose that* (9.1) *holds for some* $\varepsilon_7 \in (0, 1]$.

(a) *For large enough* N,

$$E_x[|X(N^3)|_L] - |x|_L \le C_3 N^3 \qquad \textit{for all } x. \tag{10.4}$$

(b) *For* $|x| > N^6$, *large enough* N *and* b, *and small enough* a,

$$E_x[|X(N^3)|_L] - |x|_L \le C_3(|x|_K/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^2. \tag{10.5}$$

In both parts, C_3 *is an appropriate constant that does not depend on* x *or* N.

PROOF. We first show (a). By (6.10) and (8.6) of Proposition 8.1,

$$|X(N^3)|_L - |x|_L \le C_{26}N^3 \tag{10.6}$$

for all $\omega \in \mathcal{A}_1(N^3)$ and appropriate $C_{26} > 0$ not depending on x, N, or ω. Together with Proposition 6.1, this implies

$$\begin{aligned}
&E_x[|X(N^3)|_L] - |x|_L \\
&\quad = E_x[|X(N^3)|_L; \mathcal{A}(N^3)] + E_x[|X(N^3)|_L; \mathcal{A}(N^3)^c] - |x|_L \qquad (10.7)\\
&\quad \le C_{26}N^3 + N^3 e^{-C_{10}N^{3\eta}} \le 2C_{26}N^3
\end{aligned}$$

for large enough N. For $C_3 \geq 2C_{26}$, this implies (10.4).

For (b), we suppose first that $|x|_2 \leq \varepsilon_8|x|$, where ε_8 is given below (9.13). Then, (10.2) of Proposition 10.1, together with Propositions 6.1 and 6.2, implies that the LHS of (10.5) is equal to

$$\begin{aligned}
&E_x[|X(N^3)|_L; \mathcal{A}(N^3)] + E_x[|X(N^3)|_L; \mathcal{A}(N^3)^c] - |x|_L \\
&\quad \leq C_3(|x|_K/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^2\big(2P_x(\mathcal{A}(N^3)) - 1\big) \\
&\qquad + N^3 e^{-C_{10}N^{3\eta}} \\
&\quad \leq C_3(|x|_K/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^2
\end{aligned} \tag{10.8}$$

for large N and b, and small a. This implies (10.5) for $|x|_2 \leq \varepsilon_8|x|$.

Assume now that $|x|_2 > \varepsilon_8|x|$. Choosing $C_3 \geq (2C_{26} + \frac{1}{4}\min_r w_r)/\varepsilon_8$, it follows from (10.7) that, for large N,

$$\begin{aligned}
E_x[|X(N^3)|_L] - |x|_L &\leq \Big(C_3\varepsilon_8 - \frac{1}{4}\min_r w_r\Big)N^3 \\
&\leq C_3(|x|_2/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^3 \\
&\leq C_3(|x|_K/|x|)N^3 - \Big(\frac{1}{4}\min_r w_r\Big)N^3.
\end{aligned}$$

This implies (10.5) for $|x|_2 > \varepsilon_8|x|$. □

Acknowledgments. The author thanks the referees for a detailed reading of the paper and for helpful comments.

REFERENCES

[1] BRAMSON, M. (2008). *Stability of Queueing Networks. Lecture Notes in Math.* **1950**. Springer, Berlin. MR2445100

[2] BONALD, T. and MASSOULIÉ, L. (2001). Impact of fairness on Internet performance. In *Proceedings of ACM Sigmetrics* 82–91. ACM, New York.

[3] CHUNG, K. L. (1985). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

[4] DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

[5] DE VECIANA, G., LEE, T. J. and KONSTANTOPOULOS, T. (2001). Stability and performance analysis of networks supporting elastic services. *IEEE/ACM Transactions on Networking* **9** 2–14.

[6] GROMOLL, H. C. and WILLIAMS, R. J. (2009). Fluid limits for networks with bandwidth sharing and general document size distributions. *Ann. Appl. Probab.* **19** 243–280. MR2498678

[7] KANG, W. N., KELLY, F. P., LEE, N. H. and WILLIAMS, R. J. (2009). State space collapse and diffusion approximation for a network operating under a fair bandwidth sharing policy. *Ann. Appl. Probab.* **19** 1719–1780.

[8] MASSOULIÉ, L. (2007). Structural properties of proportional fairness: Stability and insensitivity. *Ann. Appl. Probab.* **17** 809–839. MR2326233

[9] MASSOULIÉ, L. and ROBERTS, J. (2000). Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems* **15** 185–201.

[10] MEYN, S. P. and TWEEDIE, R. L. (1993). Generalized resolvents and Harris recurrence of Markov processes. In *Doeblin and Modern Probability* (*Blaubeuren*, 1991). *Contemp. Math.* **149** 227–250. Amer. Math. Soc., Providence, RI. MR1229967

[11] NUMMELIN, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. *Cambridge Tracts in Mathematics* **83**. Cambridge Univ. Press, Cambridge. MR776608

[12] OREY, S. (1971). *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities*. Van Nostrand-Reinhold, London. MR0324774

SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
TWIN CITIES CAMPUS
INSTITUTE OF TECHNOLOGY
127 VINCENT HALL
206 CHURCH STREET S.E.
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: bramson@math.umn.edu