MATHEMATICS OF OPERATIONS RESEARCH
Vol. xx, No. x, Xxxxxxx 201x, pp. xxx–xxx
ISSN 0364-765X | EISSN 1526-5471 |1x|xx0x|0xxx

DOI 10.1287/moor.xxxx.xxxx

Large Deviations Of Max-Weight Scheduling Policies On Convex Rate Regions

Vijay G. Subramanian
Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland
email: vijay.subramanian@nuim.ie http://hamilton.ie/vsubramanian

We consider a single server discrete-time system with K users where the server picks operating points from a compact, convex and coordinate convex set in $\Re^K_+$. For this system we analyse the performance of a stabilising policy that at any given time picks operating points from the allowed rate region that maximise a weighted sum of rates, where the weights depend upon the workloads of the users. Assuming a Large Deviations Principle (LDP) for the arrival processes in the Skorohod space of functions that are right-continuous with left-hand limits, we establish an LDP for the workload process using a generalised version of the *contraction principle* to derive the corresponding rate function. With the LDP result available we then analyse the tail probabilities of the workloads under different buffering scenarios.

Key words: Queueing theory; scheduling; max-weight policies; convex rate regions; large deviations; maximal monotone maps; Skorohod problem.

MSC2000 Subject Classification: Primary: 60F10, 60K25; Secondary: 90B22, 90B36.

OR/MS subject classification: Primary: Queues; Secondary: Limit theorems.

History: Received: November 18, 2009; Revised: July 19, 2010.

1. Introduction In this paper we consider a multi-class discrete-time queueing system where the server is allowed to pick operating points from within a compact, convex and coordinate convex set. Our motivation for considering compact, convex and coordinate convex rate-regions arises from information theoretic analysis of multi-user channels as described in Cover and Thomas [11, Chapter 14] like the multiple-access channel (MAC) or the broadcast channel (BC). Such models are particularly useful for modelling wireless systems. To operate near the capacity boundary of these channels sufficiently long code-words need to be used, which naturally leads to a discrete-time operation: pick a time long enough such that at all operating points one can choose long enough code-words so that the probability of error of decoding the code-words is small enough, and then schedule at the granularity of the chosen time-interval. For a simple class of such systems that includes the traditional single-server queue, where the rate regions are simplexes, a class of policies called the maximum weighted queue-length first policies were proposed in the context of wireless networks by Tassiulas and Ephremides [36] and switches by McKeown et al. [24]. Under fairly general conditions it was shown by Tassiulas and Ephremides [36], Andrews et al. [1], and McKeown et al. [24] that these policies are stabilising, i.e., if the average arrival rate vectors are strictly within the capacity region (in a manner to be defined later on), then the underlying Markov processes are positive recurrent (as described in Chen and Yao [10] and Foss and Konstantopoulos [19]). For a network of nodes with fixed routes for each flow, a generalisation of the maximum weighted queue-length first policy was presented by Maglaras and Van Mieghem [23] where the flow with the largest weighted sum queue-length is given service over its entire route was again shown to be stabilising. A related class of policies that use the age of the head-of-the-line packet instead of queue-length have been shown to exhibit optimal performance in a Large Deviations sense over a general class of work-conserving stationary scheduling policies for a single node by Stolyar and Ramanan [34] and for a network of nodes by Stolyar [35] with fixed routes for each flow, in all cases the rate regions for the nodes were simplexes. In Stolyar and Ramanan [34] and Stolyar [35] the queueing processes are embedded into the space of right continuous functions with left hand limits on the real line endowed with the topology of uniform convergence on compact sets. The authors then analyse the behaviour of the (stationary) maximum weighted end-to-end delay. They provide a large deviations upper-bound for the largest weighted delay first scheduling policy but only a large deviations like lower bound using inner measures over a general class of work-conserving stationary scheduling policies since the (stationary) maximum weighted delay need not be measurable for all the policies considered. For the largest weighted delay first scheduling policy, the lower bound is exactly a large deviations lower bound. The work in [34] also extends the large deviations analysis to the maximum weighted queue-length of a policy that serves the queue with the largest weighted queue-length, thus including serve-the-longest-queue as a special case. In the context of large deviations analysis of scheduling policies recent work by Venkataraman and Lin [39] shows that for a class of scheduling policies a large deviations principle can be proved for a specific one-dimensional functional of the queue-lengths, namely the Lyapunov function associated with each scheduling policy.

The work in this paper is along the lines of the buffer overflow problem described in the work by Bertsimas et al. [3], we consider a specific (parametric) policy and analyse its performance. Instead of just considering simplex rate-regions and two-users as in Bertsimas et al. [3] we analyse a larger class of compact, convex and coordinate convex rate-regions for an arbitrary but finite number of users. For these rate-regions the maximum weighted queue-length first policies can be generalised to choosing an operating point that maximises (over the rate-region) the weighted sum of rates. In keeping with the original policy we term these policies as Max-Weight policies. We prove an LDP result in the Skorohod space (as described in Billingsley [4], Ethier and Kurtz [17], and Jacod and Shiryaev [21]) of functions that are right-continuous with left-hand limits. Our method of proof follows the steps laid out in Puhalskii [27], Puhalskii[29], and Puhalskii and Vladimirov [30]. With the LDP result available we can then analyse the tail behaviour of the workloads under different buffering scenarios by an application of the *contraction principle* (see Dembo and Zeitouni [12]). Other related work is by Shwartz and Weiss [33, Sections 15.6-15.9, pp. 443–454] where the authors prove a Large Deviations Principle (LDP) (see Dembo and Zeitouni [12]) for a continuous-time problem where there are two queues and the server serves the longest queue (at a fixed rate). Another related work in the continuous-time context is in Dupuis et al. [16] where the authors extend the result from Shwartz and Weiss [33] to the case of many users for a (general) simplex rate-region where the weighted longest queue is served. They use the structural properties of Max-Weight policies to show that stability conditions are automatically implied in the formulation of a general Large Deviations upper bound; one of our results, namely, Theorem 3.1, can be viewed in the same vein.

The paper is organised as follows. In Section 2 we briefly describe the theory in Puhalski [27], Puhalskii[29], and Puhalskii and Vladimirov [30] that is needed to prove our result. Our main results, the LDP result and applications of it, are presented in Section 3. The analysis proceeds by proving (in Section 4) certain properties of a deterministic problem that emerges from the limiting procedure used to prove the LDP result. In Section 5 we apply the results to three two-user rate-regions: an elliptical rate-region, a Gaussian broadcast channel, a symmetrical multiple-access channel. We conclude in Section 6.

2. Background Material

Here we attempt to collect together, in brief, the mathematical background material necessary to understand and prove our result; since our paraphrasing of the material will necessarily be restrictive, for a cogent, detailed and more general explanation the reader is referred to Puhalskii [27] (and Dembo and Zeitouni [12, Chapter 4]), to Puhalskii [29] and Puhalskii and Vladimirov [30] for other worked examples of the method of proof, and to other references in this section. For the sake of consistency as much as possible we will use notation similar to Puhalski [27], Puhalskii[29], and Puhalskii and Vladimirov [30].

Let $\mathfrak{E}$ be a metric space with metric $\rho_E(\cdot,\cdot)$. A function Π from $\mathfrak{P}(\mathfrak{E})$ (the power set of $\mathfrak{E}$) to $[0,1]$ is called an idempotent probability (see Puhalskii [27, Definition 1.1.1, pp. 5–6]) if $\Pi(\emptyset)=0$, $\Pi(E)=\sup_{\mathfrak{e}\in E}\Pi(\{\mathfrak{e}\}),\ E\subseteq\mathfrak{E}$ and $\Pi(\mathfrak{E})=1$, and the pair $(\mathfrak{E},\Pi)$ is called an idempotent probability space. For ease of notation we will denote $\Pi(\mathfrak{e})=\Pi(\{\mathfrak{e}\})$ for $\mathfrak{e}\in\mathfrak{E}$. A property $\mathcal{P}(\mathfrak{e}),\ \mathfrak{e}\in\mathfrak{E}$ about the elements of $\mathfrak{E}$ is defined to hold $\Pi-a.e.$ if $\Pi(\{\mathfrak{e}:\mathcal{P}(\mathfrak{e})\text{ does not hold}\})=0$. A function f from a set $\mathfrak{E}$ equipped with idempotent probability Π to a set $\mathfrak{E}'$ is called an idempotent variable. The idempotent distribution of an idempotent variable f is defined as the set function $\Pi(f^{-1}(E'))=\Pi(f\in E'),\ E'\in\mathfrak{E}'$. Let $\mathcal{F}$ be a collection of subsets of $\mathfrak{E}$ that contains the null set $\emptyset$. Then Π is termed an $\mathcal{F}$-idempotent probability measure (see Puhalskii [27, Definition 1.1.1, pp. 5–6]) if $\Pi(\inf_n F_n)=\inf_n\Pi(F_n)$ for every decreasing sequence of elements of $\mathcal{F}$. From now onwards unless stated otherwise we will take $\mathcal{F}_C$ to be the set of all closed sets of $\mathfrak{E}$. Define an idempotent probability measure Π to be tight if for every $\epsilon>0$, there exists a compact $\mathit{\Gamma}\subseteq\mathfrak{E}$ such that $\Pi(\mathfrak{E}\setminus\mathit{\Gamma})\leq\epsilon$. A tight $\mathcal{F}_C$-idempotent probability measure is called a deviability. Using Puhalskii [27, Lemma 1.7.4, pp. 51–52] an alternate characterisation of a deviability is a function Π such that the sets $\{\mathfrak{e} \in \mathfrak{E} : \Pi(\mathfrak{e}) \geq \gamma\}$ are compact for all $\gamma \in (0,1]$. One defines another function I from $\mathfrak{E}$ to $[0,+\infty]$ that is deemed an action functional (or good rate function) if the sets $\{\mathfrak{e} \in \mathfrak{E} : I(\mathfrak{e}) \leq x\}$ are compact for $x \in \Re_+$ and $\inf_{\mathfrak{e}\in\mathfrak{E}} I(\mathfrak{e}) = 0$ (termed lower compact). It is immediate that Π is a deviability if and only if $I(\mathfrak{e}) = -\log \Pi(\mathfrak{e})$ is an action functional. If f a mapping from $\mathfrak{E}$ to another metric space $\mathfrak{E}'$ is continuous on the sets $\{\mathfrak{e} \in \mathfrak{E} : \Pi(\mathfrak{e}) \geq \gamma\}$ for $\gamma \in (0,1]$, then $\Pi(f^{-1}(\cdot))$ is a deviability on $\mathfrak{E}'$. We define f to be a Luzin idempotent variable if $\Pi(f^{-1}(\cdot))$ is a deviability on $\mathfrak{E}'$.

Let $\{\mathbb{P}_n, n \in \mathbb{N}\}$ be a sequence of probability measures on $\mathfrak{E}$ endowed with the Borel σ-algebra, and let Π be a deviability on $\mathfrak{E}$. Let $\{m_n,\ n \in \mathbb{N}\}$ be a sequence with $m_n \to +\infty$ as $n \to +\infty$. The sequence $\{\mathbb{P}_n, n \in \mathbb{N}\}$ large deviation converges (LD converges, in short) at rate m_n to Π as $n \to +\infty$ if the inequalities $\limsup_{n\to+\infty} \mathbb{P}_n(F)^{1/m_n} \leq \Pi(F)$ and $\liminf_{n\to+\infty} \mathbb{P}_n(G)^{1/m_n} \geq \Pi(G)$ hold for all closed sets F and open sets G, respectively. Note that this is an equivalent means of describing a large deviations principle for scale m_n with good rate function $I(e) = -\log \Pi(e),\ e \in \mathfrak{E}$, given the close association of deviabilities and action functionals. Equivalent definitions that make an association with convergence of measures in the traditional sense can be found in Puhalskii [27, Theorem 3.1.3, pp.254–255]. As with the traditional weak convergence of measures the LD convergence result is shown in two steps: first, by claiming the existence of limit points by proving relative (sequential) compactness of the set of measures using some notion of tightness; and second, by demonstrating uniqueness of the limit point. A deviability Π is said to be an LD limit point of the sequence $\{\mathbb{P}_n, n \in \mathbb{N}\}$ for rate m_n if each subsequence $\{\mathbb{P}_{n_t},\ t \in \mathbb{N}\}$ contains a further subsequence $\{\mathbb{P}_{n_{t_u}},\ u \in \mathbb{N}\}$ that LD converges to Π at rate $m_{n_{t_u}}$ as $u \to +\infty$. The notion of tightness from weak convergence of measures theory translates to the notion of exponential tightness that holds as follows: the sequence $\{\mathbb{P}_n, n \in \mathbb{N}\}$ is exponentially tight on order m_n, if for arbitrary $\epsilon > 0$ there exists a compact set $\varGamma \subseteq \mathfrak{E}$ such that $\limsup_{n\to+\infty} \mathbb{P}_n(\mathfrak{E} \setminus \varGamma)^{1/m_n} < \epsilon$. From Puhalskii [27, Theorem 3.1.19, pp.262–263] exponential tightness implies LD relative compactness of a sequence of measures, and therefore existence of limit points; furthermore, the limit points are all deviabilities. The LD convergence of probability measures can also be stated as the LD convergence in distribution of the associated random variables. A sequence of random variables $\{\mathfrak{X}_n,\ n \in \mathbb{N}\}$ defined on probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, respectively, and assuming values in $\mathfrak{E}$ LD converges at rate m_n as $n \to +\infty$ to a Luzin idempotent variable $\mathfrak{X}$ defined on idempotent probability space (Υ, Π) and assuming values in $\mathfrak{E}$, if the sequence of probability laws of $\mathfrak{X}_n$ LD converges to the idempotent distribution of $\mathfrak{X}$ at rate m_n. If the convergence of probability measures is to a deviability, then we get LD convergence in the canonical setting. The role of the *continuous mapping principle* in preserving convergence is played by the *contraction principle* (see Dembo and Zeitouni [12, Theorem 4.2.1, pp. 126–127] and Puhalskii [27, Theorem 3.1.14 and Corollary 3.1.15 pp. 261–262]) whereby if a sequence of random variables $\mathfrak{X}_n$ LD converge in distribution to $\mathfrak{X}$, and f is a $(\Pi-a.e.)$ continuous function from $\mathfrak{E}$ to another metric space $\mathfrak{E}'$, then the sequence $f(\mathfrak{X}_n)$ LD converges in distribution to $f(\mathfrak{X})$ in $\mathfrak{E}'$.

In many cases of interest the LD limit points belong to a subset $\mathfrak{E}_0$ of $\mathfrak{E}$. Equip $\mathfrak{E}_0$ with the relative topology. Then we say that an idempotent probability is supported by $\mathfrak{E}_0$ if $\Pi(\mathfrak{E} \setminus \mathfrak{E}_0) = 0$. The sequence $\{\mathbb{P}_n, n \in \mathbb{N}\}$ is termed $\mathfrak{E}_0$-exponentially tight if it is exponentially tight and every LD accumulation point Π is supported by $\mathfrak{E}_0$. Important here from the point of view of the contraction principle is the definition of $\mathfrak{E}_0$-closed and $\mathfrak{E}_0$-open sets as well as $\mathfrak{E}_0$-continuous functions. From Puhalskii [27, Definition 1.9.9, pg. 68], we have the following [1]: 1) a set $F \subset \mathfrak{E}$ is $\mathfrak{E}_0$-closed if it contains all its accumulation points that are in $\mathfrak{E}_0$, i.e., $\mathrm{cl}(F) \cap \mathfrak{E}_0 \subset F$; and 2) a set $G \subset \mathfrak{E}$ is $\mathfrak{E}_0$-open if every point of $G \cap \mathfrak{E}_0$ is an interior point of G, i.e., $G \cap \mathfrak{E}_0 \subset \mathrm{int}(G)$. Additionally, from Puhalskii [27, Definition 1.9.11, pg. 69] we deem a function $f : \mathfrak{E} \mapsto \Re$ to be $\mathfrak{E}_0$-continuous if $h^{-1}(G)$ is $\mathfrak{E}_0$-open for each open $G \subset \Re$, and alternatively, $\mathfrak{E}_0$-continuity of f is equivalent to $\lim_{n\to\infty} f(\mathfrak{x}_n) = f(\mathfrak{x})$ for every sequence $\{\mathfrak{x}_n\}_{n=1}^{\infty} \subset \mathfrak{E}$ such that $\lim_{n\to\infty} \mathfrak{x}_n = \mathfrak{x} \in \mathfrak{E}_0$. From Puhalskii [27, Corollary 3.1.9, pp.257–258] LD convergence of sequence $\{\mathbb{P}_n, n \in \mathbb{N}\}$ to Π that is supported by $\mathfrak{E}_0$ only needs to be checked for all $\mathfrak{E}_0$-open and $\mathfrak{E}_0$-closed Borel measurable subsets of $\mathfrak{E}$. Then the *contraction principle* (see Garcia [20] and Puhalskii [27, Corollary 3.1.22, pg. 264]) applies to Borel-measurable but $\mathfrak{E}_0$-continuous functions.

For the results of this paper the space $\mathfrak{E}$ will be $\mathbb{D}(\mathbb{X}) := \mathbb{D}([0,1];\mathbb{X})$ the space of $\mathbb{X}$-valued, right-continuous with left-hand limits functions $\mathbf{x} = (\mathbf{x}(t),\ t \in [0,1])$ where $\mathbb{X}$ is a complete separable metric space. Our results carry over if the setting was, instead, $\mathbb{D}([0,T];\mathbb{X})$ for fixed T with $0 < T < +\infty$.

[1] From Puhalskii [27, Remark 1.9.10, pg. 68] we note that both the interior and closure operation are taken in $\mathfrak{E}$.

Equipping $\mathbb{D}(\mathbb{X})$ with the Skorohod J_1-topology (see Billingsley [4], Ethier and Kurtz [17], and Jacod and Shiryaev [21]) and metrising it with the Skorohod-Prohorov-Lindvall metric (see Billingsley [4], Ethier and Kurtz [17], and Jacod and Shiryaev [21]) we get a complete separable metric space. Following Billingsley [4, Chapter 3, Section 12] and Ethier and Kurtz [17, Sections 3.5 and 3.6] let $\mathfrak{F}$ be the collection of (strictly) increasing functions $f = \{f(t) \in [0,1],\ t \in [0,1]\}$ mapping $[0,1]$ onto $[0,1]$; in particular, $f(0) = 0$, $\lim_{t\to 1} f(t) = 1$, and f is continuous. Let $\tilde{\mathfrak{F}}$ be the set of Lipschitz continuous functions $f \in \mathfrak{F}$ such that

$$\gamma(f) := \sup_{1\geq t>s\geq 0} \left|\log \frac{f(t)-f(s)}{t-s}\right| < +\infty. \tag{1}$$

For two functions $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathbb{D}(\mathbb{X})$ the Skorohod-Prohorov-Lindvall metric is given by

$$\rho_{J_1}(\mathbf{x}_1, \mathbf{x}_2) := \inf_{f\in\tilde{\mathfrak{F}}} \left\{ \max\left(\gamma(f), \sup_{t\in[0,1]} \rho_X\left(\mathbf{x}_1(t), \mathbf{x}_2\big(f(t)\big)\right)\right)\right\},$$

where $\rho_X(\cdot,\cdot)$ is the metric on $\mathbb{X}$.

The (closed) subset $\mathfrak{E}_0$ of $\mathbb{D}(\mathbb{X})$ that we will be dealing with in this paper is the set of all continuous functions $\mathbb{C}(\mathbb{X}) := \mathbb{C}([0,1];\mathbb{X})$ with the induced topology, which is the uniform topology with metric $\rho_C(\mathbf{x}_1, \mathbf{x}_2)$ for two functions $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathbb{C}(\mathbb{X})$ given by

$$\rho_C(\mathbf{x}_1, \mathbf{x}_2) := \sup_{t\in[0,1]} \rho_X\Big(\mathbf{x}_1(t), \mathbf{x}_2(t)\Big). \tag{2}$$

For $\mathbf{x} \in \mathbb{X}$ define $\mathbb{C}_{\mathbf{x}}(\mathbb{X}) := \{\mathbf{a} \in \mathbb{C}(\mathbb{X}) : \mathbf{a}(0) = \mathbf{x}\}$, which is a closed subset of $\mathbb{C}(\mathbb{X})$. We could prove our results for $\mathbb{D}(\mathbb{X})$ with the uniform topology. However, we prefer using the Skorohod J_1-topology since $\mathbb{D}(\mathbb{X})$ is then a complete separable metric space where by using Puhalskii [27, Theorem 3.1.28, pg. 268], exponential tightness is equivalent to LD relative (sequential) compactness.

Since $\mathbb{C}(\mathbb{X})$ is a closed subset of $\mathbb{D}(\mathbb{X})$ using Puhalskii [27, Corollaries 1.7.12 on pg. 54 and 1.8.7 on pg. 62] we do not distinguish between deviabilities on $\mathbb{C}(\mathbb{X})$ and deviabilities on $\mathbb{D}(\mathbb{X})$ that are supported by $\mathbb{C}(\mathbb{X})$. From Puhalskii [27, Remark 3.2.4 and Theorem 3.2.3, pg. 278] (and Ethier and Kurtz [17, Sections 3.5 and 3.6]) we note that a sequence of processes $\mathbf{X}^N$ with trajectories in $\mathbb{D}(\mathbb{X})$ is $\mathbb{C}(\mathbb{X})$-exponentially tight on order N if and only if the two statements below hold, namely,

(i) (*Exponential tightness of random variables*) for every $t \in [0,1]$, $\mathbf{X}^N(t)$ is exponentially tight, i.e.,

$$\inf_{\Gamma\in\mathbf{\Gamma}} \limsup_{N\to+\infty} \mathbb{P}\big(\mathbf{X}^N(t) \in \mathbb{X}\setminus\Gamma\big)^{1/N} = 0, \tag{3}$$

where $\mathbf{\Gamma}$ is the set of compact subsets of $\mathbb{X}$; and

(ii) (*Continuous limit points*) for every $T \in (0,1]$, $\epsilon > 0$ the following holds

$$\lim_{\delta\to 0} \limsup_{N\to+\infty} \mathbb{P}\left(\sup_{\substack{s,t\in[0,T]:\\ |s-t|\leq\delta}} \rho_X\big(\mathbf{X}^N(t), \mathbf{X}^N(s)\big) > \epsilon \right)^{1/N} = 0. \tag{4}$$

The above condition can also be stated as

$$\lim_{\delta\to 0} \limsup_{N\to+\infty} \sup_{t\in[0,T]} \mathbb{P}\left(\sup_{\substack{s\in[0,T]:\\ |s-t|\leq\delta}} \rho_X\big(\mathbf{X}^N(t), \mathbf{X}^N(s)\big) > \epsilon \right)^{1/N} = 0, \tag{5}$$

which is many times easier to verify.

Instead if the $\mathbf{X}^N$ have trajectories in $\mathbb{D}(\Re^K)$, then by Puhalskii [27, Theorem 3.2.3, pg. 278] $\mathbb{C}(\Re^K)$-exponential tightness on order N holds if and only if

$$\lim_{L\to+\infty} \limsup_{N\to+\infty} P(\|\mathbf{X}^N(0)\| > L)^{1/N} = 0; \text{ and} \tag{6}$$

$$\lim_{\delta\to 0} \limsup_{N\to+\infty} P\left(\sup_{s,t\in[0,T]:\ |s-t|\leq\delta} \|\mathbf{X}^N(s) - \mathbf{X}^N(t)\| > \epsilon \right)^{1/N} = 0,\ T \in (0,1],\ \epsilon > 0. \tag{7}$$

Similar to condition (4), the condition in (7) can also be stated as

$$\lim_{\delta\to 0}\limsup_{N\to+\infty}\sup_{t\in[0,T]} P\left(\sup_{s\in[0,T]:\,|s-t|\leq\delta}\|\mathbf{X}^N(s)-\mathbf{X}^N(t)\|>\epsilon\right)^{1/N}=0,\ T\in(0,1],\ \epsilon>0. \tag{8}$$

Alternate characterisations of $\mathbb{C}(\mathbb{X})$-exponential tightness can be found in Feng and Kurtz [18, Section 4.4, pp. 65–67] and Puhalskii [28, Theorem 2.5, pg. 15].

Using the well-known characterisation that a sequence $\{\mathbf{x}_n\}_{n=1}^\infty\subset\mathbb{D}(\mathbb{X})$ converging to $\mathbf{x}\in\mathbb{C}(\mathbb{X})$ in the Skorohod J_1-topology also converges in the local uniform topology on $\mathbb{D}(\mathbb{X})$ (c.f., Ethier and Kurtz [17, Lemma 10.1& Theorem 10.2, pg. 148]), $\mathbb{C}(\mathbb{X})$-open and $\mathbb{C}(\mathbb{X})$-closed sets as well as $\mathbb{C}(\mathbb{X})$-continuous functions are also characterised using the local uniform topology on $\mathbb{D}(\mathbb{X})$. Thus, we can then use Puhalskii [27, Corollary 3.1.9, pp. 257–258 and Corollary 3.1.22, pg. 264] to characterise Large Deviations inequalities for these objects. We will use this property many times over.

For some results we will use $\mathbb{X}=\Re_+^K,\ K\in\mathbb{N}$ and for others we will use $\mathbb{X}=\Re_+^K\times\mathcal{M}(\mathcal{R}(L)),\ K,\ L\in\mathbb{N}$ where $\mathcal{M}(\mathcal{R}(L))$ is the set of all finite non-negative Borel measures on $\mathcal{R}(L)$ a convex compact set in $\Re_+^L$. The set of all finite non-negative Borel measures $\mathcal{M}(\mathfrak{E})$ on a complete separable metric space $\mathfrak{E}$, is a complete separable metric space with the Lévy-Prohorov metric and the topology of weak convergence. The Lévy-Prohorov metric (in the symmetric form, see Billingsley [4], Ethier and Kurtz [17], and Dembo and Zeitouni [12, Theorem D.8, pp. 355–356]) $\rho_P(\nu_1,\nu_2)$ for two measures $\nu_1,\ \nu_2\in\mathcal{M}(\mathfrak{E})$ is given by the following

$$\rho_P(\nu_1,\nu_2):=\inf\bigl\{\epsilon>0:\nu_1(C)\leq\nu_2(C^\epsilon)+\epsilon \text{ and } \nu_2(C)\leq\nu_1(C^\epsilon)+\epsilon\ \forall\ C\in\mathcal{B}(\mathfrak{E})\bigr\}, \tag{9}$$

where $\mathcal{B}(\mathfrak{E})$ is the Borel σ-algebra on $\mathfrak{E}$, and $C^\epsilon:=\{\mathfrak{e}\in\mathfrak{E}:\inf_{\mathfrak{e}_1\in C}\rho_E(\mathfrak{e},\mathfrak{e}_1)<\epsilon\}$. In practice, it is sufficient to consider only $\mathcal{F}_C$ instead of $\mathcal{B}(\mathfrak{E})$ in the definition in (9). For $t\geq 0$ define $\mathcal{M}_t(\mathfrak{E}):=\{\nu\in\mathcal{M}(\mathfrak{E}):\nu(\mathfrak{E})\leq t\}$ to be the set of (non-negative) finite measures assigning a measure at most t to $\mathfrak{E}$, and $\mathcal{M}^t(\mathfrak{E}):=\{\nu\in\mathcal{M}(\mathfrak{E}):\nu(\mathfrak{E})=t\}$ to be the set of (non-negative) finite measures assigning a measure exactly t to $\mathfrak{E}$. Then $\mathcal{M}_t(\mathfrak{E})$ and $\mathcal{M}^t(\mathfrak{E})$ are compact if and only if $\mathfrak{E}$ is compact (see Doob [14, Section VIII.5, pg. 132] and Dembo and Zeitouni [12, Theorem D.8, pp. 355–356]). The topology of weak convergence also results by using the Kantorovich-Wasserstein metric (see Dembo and Zajic [13, Lemma A.1, pg. 222], Dembo and Zeitouni [12, Theorem D.8, pp. 355–356] and Dudley [15]). Let $\mathfrak{C}^b(\mathfrak{E})$ denote the set of bounded continuous functions on $\mathfrak{E}$ that take values in $\Re$. Then the Kantorovich-Wasserstein metric $\rho_{KL}(\nu_1,\nu_2)$ for two measures $\nu_1,\nu_2\in\mathcal{M}(\mathfrak{E})$ is constructed using bounded and Lipschitz continuous functions on $\mathfrak{E}$, and is given by

$$\rho_{KL}(\nu_1,\nu_2):=\sup\left\{\left|\int_{\mathfrak{E}}fd\nu_1-\int_{\mathfrak{E}}fd\nu_2\right|:f\in\mathfrak{C}^b(\mathfrak{E}),\|f\|_\infty+\|f\|_L\leq 1\right\}, \tag{10}$$

where $\|f\|_\infty=\sup_{\mathfrak{e}\in\mathfrak{E}}f(\mathfrak{e})$ and $\|f\|_L:=\sup_{\{\mathfrak{e}_1,\mathfrak{e}_2\in\mathfrak{E}:\mathfrak{e}_1\neq\mathfrak{e}_2\}}\frac{|f(\mathfrak{e}_1)-f(\mathfrak{e}_2)|}{\rho_E(\mathfrak{e}_1,\mathfrak{e}_2)}$ is the Lipschitz constant of f. We will use the Kantorovich-Wasserstein metric instead of the Lévy-Prohorov metric to prove (4) in our analysis. In our analysis $\mathfrak{E}$ will be a compact convex subset of $\Re_+^L$. In this setting weak convergence of finite Borel measures on $\mathfrak{E}$ is the same as weak* convergence in a Banach space since the space of finite Borel measures on $\mathfrak{E}$ is the dual space of $\mathfrak{C}^b(\mathfrak{E})$.

Following Dembo and Zajic [13] for non-decreasing functions Φ in $\mathbb{D}(\mathcal{M}(\mathfrak{E}))$, i.e., functions such that $\Phi(t)-\Phi(u)\in\mathcal{M}(\mathfrak{E})$ for all $t\geq u$ with $t,u\in[0,1]$, we say that the right weak derivative exists at $t\in[0,1)$ if $\frac{\Phi(t+\epsilon)-\Phi(t)}{\epsilon}\in\mathcal{M}(\mathfrak{E})$ weakly converges as $\epsilon\to 0$. Similarly the left weak derivative exists at $t\in(0,1]$ if $\frac{\Phi(t)-\Phi(t-\epsilon)}{\epsilon}\in\mathcal{M}(\mathfrak{E})$ weakly converges as $\epsilon\to 0$. For $t\in(0,1)$ if both the right and left weak derivative exist, then we deem Φ to be weakly differentiable at t with the limit denoted as $\dot{\Phi}(t)$. For absolutely continuous Φ we have $\Phi(t)-\Phi(u)=\int_u^t\dot{\Phi}(\tau)d\tau$ where the integral is interpreted set-wise, i.e., for $C\in\mathcal{B}(\mathfrak{E})$ we have $\bigl(\Phi(t)-\Phi(u)\bigr)\Bigl(C\Bigr)=\int_u^t\dot{\Phi}(\tau)\bigl(C\bigr)d\tau$.

We make use of the terminology "strong" and "weak" in defining solutions (see Brézis [7, pg. 64]) to differential inclusions. For absolutely continuous function $\mathbf{a}\in\mathbb{C}(\Re^L)$ and set-valued function $H(\cdot)$ such that $H(\mathbf{x})\subseteq\Re^L$ for $\mathbf{x}\in\Re^L$ with domain $\mathrm{D}(H):=\{\mathbf{x}\in\Re^L:H(\mathbf{x})\neq\emptyset\}$, we say that $\mathbf{w}\in\mathbb{C}(\Re^L)$ is a strong solution to differential inclusion

$$\dot{\mathbf{w}}(t)\in\dot{\mathbf{a}}(t)+H(\mathbf{w}(t))\quad t\in[0,1], \tag{11}$$

if $\mathbf{w}$ is absolutely continuous with $\mathbf{w}(t) \in \mathrm{Dom}(H)\ \forall t \in [0,1]$ and satisfies

$$\dot{\mathbf{w}}(t) \in \dot{\mathbf{a}}(t) + H(\mathbf{w}(t)) \text{ for } a.e.\ t \in (0,1).$$

Note that the absolute continuity of $\mathbf{a}$ automatically posits the $(a.e.)$ existence of the integrable function $\dot{\mathbf{a}} \in \mathcal{L}^1([0,1];\Re^L)$. One defines a function $\mathbf{w} \in \mathbb{C}(\Re^L)$ to be a weak solution to differential inclusion (11) for input $\mathbf{a} \in \mathbb{C}(\Re^L)$, if there exist a sequence of absolutely continuous functions $\{\mathbf{a}_N \in \mathbb{C}(\Re^L)\}_{N=1}^{+\infty}$ and a sequence $\{\mathbf{w}_N \in \mathbb{C}(\Re^L)\}_{N=1}^{+\infty}$ such that each $\mathbf{w}_N$ is a strong solution of the differential inclusion

$$\dot{\mathbf{w}}_N(t) \in \dot{\mathbf{a}}_N(t) + H(\mathbf{w}_N(t)) \quad t \in [0,1], \tag{12}$$

and $\dot{\mathbf{a}}_N \Rightarrow_{N\to+\infty} \dot{\mathbf{a}}$ in $\mathcal{L}^1([0,1];\Re^L)$ and $\mathbf{w}_N \Rightarrow_{N\to+\infty} \mathbf{w}$ uniformly in $\mathbb{C}(\Re^L)$.

Finally if a sequence of random variables $\{\mathbf{X}^N,\ N \in \mathbb{N}\}$ defined on a complete probability space $(\Omega,\mathcal{F},\mathbb{P})$ and assuming values in $\Re^K$ converges in probability to $\mathbf{x} \in \Re^K$ such that $\lim_{N\to+\infty}\mathbb{P}(\|\mathbf{X}_N - \mathbf{x}\| > \epsilon)^{1/N} = 0$ for all $\epsilon > 0$, then we deem the sequence as converging super-exponentially in probability at rate N and write $\mathbf{X}_N \xrightarrow{\mathbb{P}^{1/N}} \mathbf{x}$.

3. Model And A Fluid Limit We consider a discrete-time queueing system with one server that can pick operating points from a set $\mathcal{R}(K)$ that is a compact, with non-empty interior $\mathrm{int}(\mathcal{R}(K))$, and convex subset of $\Re_+^K$ that includes the origin. We also assume that $\mathcal{R}(K)$ is coordinate-convex, i.e., if $\mathbf{r} \in \mathcal{R}(K)$, then $\hat{\mathbf{r}} \in \mathcal{R}(K)$ for all $\mathbf{0} \le \hat{\mathbf{r}} \le \mathbf{r}$ where the inequalities hold coordinate-wise [2]. Since $\mathcal{R}(K)$ is compact there exists a $r_{\max} < +\infty$ such that for every $\mathbf{r} \in \mathcal{R}(K)$ we have $r^k \le r_{\max}$ for all $k = 1,2,\ldots,K$. For user k we assume an arrival process of work brought into the system given by a sequence $\{A_m^k\}_{m=0}^{+\infty}$ where $A_m^k \in \Re_+$ is the work brought in at time m into the queue of user k. For $-1 \le m_1 \le m_2$ integers define $A^k(m_1,m_2] := \sum_{m=m_1+1}^{m_2} A_m^k$ which is the total amount of work to arrive for user k after time slot m_1 and until time-slot m_2. If $m_1 \ge m_2$, then we interpret $A^k(m_1,m_2]$ to be 0. Let the unfinished work in user k's queue at time $m \ge 0$ be W_m^k. Then work at time $m+1$ in the k^{th} user's queue is given by Lindley's recursion

$$\begin{aligned} W_{m+1}^k = \max(A_m^k, W_m^k + A_m^k - r_m^k) &= \max\left(0, W_m^k - r_m^k\right) + A_m^k \\ &:= \left(W_m^k - r_m^k\right)_+ + A^k(m-1,m] \end{aligned} \tag{13}$$

where $\mathbf{r}_m \in \mathcal{R}(K)$ is the operating point chosen at time m. Note that we choose to wait for at least one slot before serving newly arrived work. Allowing the server to work on newly arrived work immediately does not change the results.

Our scheduling policy will be to choose a rate vector that maximises a (dynamic) weighted sum of rates over this rate region, i.e.,

$$\forall\, m \ge 1 \quad \mathbf{r}_m \in \arg\max_{\mathbf{r}\in\mathcal{R}(K)} \langle \boldsymbol{\alpha}_m, \mathbf{r}\rangle$$

with components r_m^k where $\boldsymbol{\alpha}_m$ is given by

$$\alpha_m^k = \tilde{\beta}^k W_m^k \quad \text{s.t. } \sum_{k=1}^K \tilde{\beta}^k = 1,\ \tilde{\beta}_k > 0\ \forall k,$$

and where $< \cdot,\cdot >$ is the standard inner product in $\Re^K$. Note that $\boldsymbol{\alpha}_m$ is the Hadamard/Schur product of $\tilde{\boldsymbol{\beta}}$ and $\mathbf{W}_m$ and we will write this as $\boldsymbol{\alpha}_m = \tilde{\boldsymbol{\beta}} \circ \mathbf{W}_m$. Define the following (set-valued) functions for $\mathbf{x} \in \Re_+^K$

$$H(\mathbf{x}) := \arg\max_{\mathbf{r}\in\mathcal{R}(K)} \langle \mathbf{x}, \mathbf{r}\rangle\,, \tag{14}$$

$$\tilde{H}(\mathbf{x}) := H(\tilde{\boldsymbol{\beta}} \circ \mathbf{x}). \tag{15}$$

We will fix on a specific solution in case there is more than one maximiser. For a closed convex set $\mathcal{S} \subseteq \Re^K$ define the projection of element $x \in \Re^K$ to be the unique element $\mathbf{x}^* \in \mathcal{S}$ that solves

$$\min_{\mathbf{y}\in\mathcal{S}} \|\mathbf{x} - \mathbf{y}\|^2,$$

[2] Henceforth, unless specified otherwise, we assume that all vector inequalities hold coordinate-wise.

where $\|\cdot\|$ is the Euclidean norm given by $\sqrt{<\mathbf{x},\mathbf{x}>}$ for $\mathbf{x} \in \Re_+^K$. We define the function from $\mathbf{x}$ to $\mathbf{x}^*$ for a given $\mathcal{S}$ to be $\mathrm{Proj}_{\mathcal{S}}(\mathbf{x})$. For every $\mathbf{x} \in \Re_+^K$ is clear that $\tilde{H}(\mathbf{x})$ is a closed and convex set. Then the specific operating point that we choose at time m is given by

$$\mathbf{r}_m = \mathrm{Proj}_{\tilde{H}(\mathbf{W}_m)}(\mathbf{0}).$$

Based upon the above definition we call operating point $\mathbf{r}_m$ the minimum norm solution.

Using the operating point $\mathbf{r}_m$ at time m we get

$$\begin{aligned} W_{m+1}^k &= \left(W_m^k - r_m^k\right)_+ + A^k(m-1,m] \\ &= W_m^k - \min(W_m^k, r_m^k) + A^k(m-1,m] \\ &= W_0^k - \sum_{l=0}^{m} S_l^k + A^k(-1,m] \\ &= W_0^k - S^k(-1,m] + A^k(-1,m], \end{aligned} \tag{16}$$

where for $m \geq 1$ we define $S_m^k := \min(W_m^k, r_m^k) \leq r_{\max}$, which is the amount of work from the queue of user k served at time m. Coordinate convexity ensures that for every $\mathbf{r} \in \mathcal{R}(K)$ every point $\min(\mathbf{x},\mathbf{r})$ (with the minimum taken along every coordinate) belongs to $\mathcal{R}(K)$ as $\mathbf{x}$ is allowed to vary in $\Re_+^K$. Before proceeding it is necessary to point out that our description of the Max-Weight policy is such that it is neither work-conserving nor rate-maximising. This can be best appreciated for a simplex rate-region: if the chosen queue does not have enough work, then the excess work is not applied to another set of queues that may have some work, i.e., the server does not try to empty the system as much as possible.

To aid in the proof of the result we will define a few more quantities. Denote by $R^k(m_1,m_2] = \sum_{l=m_1+1}^{m_2} r_l^k$ the total amount of service given to user k from m_1+1 until m_2. As discussed earlier not all of this service is used since the queues might not contain enough work to be served. Therefore we define $Y^k(m_1,m_2] := \sum_{l=m_1+1}^{m_2}(r_l^k - W_l^k)_+$ to be total amount of service not utilised by user k from m_1+1 until m_2. Note that this is analogous to the server idle time (see Chen and Yao [10]) in a continuous-time queueing system. It is obvious from the definitions that $S^k(m_1,m_2] = R^k(m_1,m_2] - Y^k(m_1,m_2]$ where all three terms are non-negative. The form above is not very conducive to analysis. Therefore we modify it for $m \geq 1$ as follows:

$$\begin{aligned} Y^k(m-1,m] &= (r_m^k - W_m^k)_+ = \max(0, r_m^k - W_m^k) \\ &= \max\left(0, \min(r_m^k + r_{m-1}^k - A_{m-1}^k - W_{m-1}^k, r_m^k - A_{m-1}^k)\right) \\ &= \max\Bigg(0, \min\Bigg(R^k(-1,m] - A^k(-1,m-1] - W_0^k, \\ &\qquad\qquad \min_{1\leq i\leq m}\Big(R^k(i-1,m] - A^k(i-2,m-1]\Big)\Bigg)\Bigg) \\ &= \max\Bigg(0, R^k(-1,m) - A^k(-1,m-1] - W_0^k \\ &\qquad - \max\Bigg(0, \max_{1\leq i\leq m}\Big(R^k(-1,i-1] - A^k(-1,i-2] - W_0^k\Big)\Bigg)\Bigg) \\ &= \max\Bigg(0, \max_{1\leq i\leq m+1}\Big(R^k(-1,i-1] - A^k(-1,i-2] - W_0^k\Big)\Bigg) \\ &\quad - \max\Bigg(0, \max_{1\leq i\leq m}\Big(R^k(-1,i-1] - A^k(-1,i-2] - W_0^k\Big)\Bigg) \end{aligned}$$

where we use the relationship in (13) many times over to unravel the recursive definition. Therefore we have

$$\begin{aligned} Y^k(-1,m] &= \max\Bigg(0, \max_{0\leq i\leq m}\Big(R^k(-1,i] - A^k(-1,i-1] - W_0^k\Big)\Bigg) \\ &= \max\Bigg(W_0^k, \max_{0\leq i\leq m}\Big(R^k(-1,i] - A^k(-1,i-1]\Big)\Bigg) - W_0^k. \end{aligned} \tag{17}$$

Using this we can also write the following

$$W_m^k = \begin{cases} W_0^k & \text{if } m = 0; \\ \max\Bigl(W_0^k + A^k(-1, m-1] - R^k(-1, m-1], & \\ \qquad A^k(-1, m-1] - R^k(-1, m-1] - & \text{otherwise.} \\ \qquad\quad \min_{0 \le i \le m-1}\Bigl(A^k(-1, i-1] - R^k(-1, i]\Bigr)\Bigr) & \end{cases} \tag{18}$$

Assume that we are given a sequence $\{\mathbf{W}_0^N\}_{N\in\mathbb{N}}$ taking values in $\Re_+^K$ that accounts for the vector of initial work in the system. We then embed the sequences $\{A^k(-1, m]\}$, $\{R^k(-1, m]\}$, $\{Y^k(-1, m]\}$, $\{S^k(-1, m]\}$ and $\{W_m^k\}$ into functions in $\mathbb{D}(\Re_+^K)$ by defining (scaling both space and time) for $t \in [0, 1]$ the following: $A^{k,N}(t) := \frac{A^k(-1, \lfloor Nt \rfloor]}{N}$, $R^{k,N}(t) := \frac{R^k(-1, \lfloor Nt \rfloor]}{N}$, $Y^{k,N}(t) := \frac{Y^k(-1, \lfloor Nt \rfloor]}{N}$, $S^{k,N}(t) := \frac{S^{k,N}(-1, \lfloor Nt \rfloor]}{N}$ and $W^{k,N}(t) := \frac{W_{\lfloor Nt \rfloor}^{k,N}}{N}$ for $N \in \mathbb{N}$ where $\lfloor t \rfloor$ is the largest integer less than or equal to t. The index N in $R^{k,N}$, $Y^{k,N}$, $S^{k,N}$ and $W^{k,N}$ takes into account the different initial workload vectors given by $\mathbf{W}_0^N$. Denote the vector quantities by $\mathbf{A}^N(t)$, $\mathbf{R}^N(t)$, $\mathbf{Y}^N(t)$, $\mathbf{S}^N(t)$ and $\mathbf{W}^N(t)$, respectively. Also define the processes $\mathbf{A}^N := (\mathbf{A}^N(t),\ t \in [0, 1])$, $\mathbf{R}^N := (\mathbf{R}^N(t),\ t \in [0, 1])$, $\mathbf{Y}^N := (\mathbf{Y}^N(t),\ t \in [0, 1])$, $\mathbf{S}^N := (\mathbf{S}^N(t),\ t \in [0, 1])$, and $\mathbf{W}^N := (\mathbf{W}^N(t),\ t \in [0, 1])$. The workload arrivals sequence $\{\mathbf{A}_m\}_{m\in\mathbb{N}}$ and the initial workload vector sequence $\{\mathbf{W}_0^N\}_{N\in\mathbb{N}}$ are assumed to be defined on a common complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Construct the random empirical measure $\Psi(-1, m](\cdot) := \sum_{l=0}^m \boldsymbol{\delta}_{\mathbf{r}_l}(\cdot)$ where $\boldsymbol{\delta}_{\mathbf{x}}(\cdot)$ is the Dirac measure concentrated at $\mathbf{x}$, i.e., for all $C \in \mathcal{B}(\Re_+^K)$ we have

$$\boldsymbol{\delta}_{\mathbf{x}}(C) = \begin{cases} 1 & \text{if } \mathbf{x} \in C; \\ 0 & \text{otherwise.} \end{cases}$$

Define the scaled empirical measure process $\Psi^N(t) := \frac{\Psi^N(-1, \lfloor Nt \rfloor]}{N}$ for $t \in [0, 1]$. Again the index N accounts for the different initial workload vector. Let $\mathcal{M}(\mathcal{R}(K))$ be the set of finite (non-negative) Borel measures on $\mathcal{R}(K)$; when endowed with the topology of weak convergence of measures generated by the Kantorovich-Wasserstein metric, $\mathcal{M}(\mathcal{R}(K))$ is a complete separable metric space. Then the processes Ψ^N take values in $\mathbb{D}(\mathcal{M}(\mathcal{R}(K)))$ again with the Skorohod J_1 topology (see Ethier and Kurtz [17]). In fact for every $t \in [0, 1]$ we have $\Psi^N(t) \in \mathcal{M}_{t+1}(\mathcal{R}(K)) = \{\nu \in \mathcal{M}(\mathcal{R}(K)) :\ \nu(\mathcal{R}(K)) \le t + 1\}$ where $\mathcal{M}_{t+1}(\mathcal{R}(K))$ is a compact subset of $\mathcal{M}(\mathcal{R}(K))$. For a Borel measurable function f from $\mathcal{R}(K)$ to $\Re$ denote the integral (if it exists) with respect to a measure $\nu \in \mathcal{M}(\mathcal{R}(K))$ by $\int_{\mathcal{R}(K)} f d\nu$; this is a random variable taking values in $\Re$ that we denote as $< \nu, f >$.

For our convergence proofs we will be considering processes $(\mathbf{A}^N, \mathbf{R}^N, \mathbf{Y}^N, \mathbf{S}^N, \Psi^N, \mathbf{W}^N)$ taking values in the Skorohod space $\mathbb{D}(\Re_+^K \times \Re_+^K \times \Re_+^K \times \Re_+^K \times \mathcal{M}(\mathcal{R}(K)) \times \Re_+^K)$; we denote the complete separable metric space $\Re_+^K \times \Re_+^K \times \Re_+^K \times \Re_+^K \times \mathcal{M}(\mathcal{R}(K)) \times \Re_+^K$ by $\mathbb{X}$. Denote the Euclidean metric on $\Re_+^K$ by ρ_E and the Kantorovich-Wasserstein metric on $\mathcal{M}(\mathcal{R}(K))$ by ρ_{KL}. Then for two elements $(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1),\ (\mathbf{a}_2, \boldsymbol{\gamma}_2, \boldsymbol{\eta}_2, \mathbf{s}_2, \Phi_2, \mathbf{w}_2) \in \mathbb{X}$ the distance between the two elements is given by the following metric

$$\begin{aligned} &\rho_X\bigl((\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1), (\mathbf{a}_2, \boldsymbol{\gamma}_2, \boldsymbol{\eta}_2, \mathbf{s}_2, \Phi_2, \mathbf{w}_2)\bigr) := \\ &\quad \max(\rho_E(\mathbf{a}_1, \mathbf{a}_2), \rho_E(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2), \rho_E(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2), \rho_E(\mathbf{s}_1, \mathbf{s}_2), \rho_{KL}(\Phi_1, \Phi_2), \rho_E(\mathbf{w}_1, \mathbf{w}_2)). \end{aligned}$$

The topology that results from this metric on $\mathbb{X}$ is the product topology.

To prove the required large deviations result we will follow the programme outlined in Puhalskii [27], Puhalskii [29], and Puhalskii and Vladimirov [30]. Loosely speaking, we first show in Theorem 3.1 that the sequence of measures on a metric space $\mathfrak{E}$ is large deviations relatively compact using exponential tightness. In proving relatively compactness we will prove that all the limit points are supported on $\mathfrak{E}_0$ a closed subset of $\mathfrak{E}$. The limit points are determined by weak solutions to idempotent equations the properties of which are characterised by taking large deviation limits of the stochastic equations that determine the behaviour of the original system. Compactness of the rate-regions plays an important role in proving exponential tightness and characterising the large deviations limit points. Using all the characterised properties of the idempotent equations we show the existence of unique weak solutions to the idempotent equations in Theorem 4.1 leading to an LDP result in Theorem 3.2. In effect this step can also be referred to as establishing uniqueness in idempotent distribution (in analogy to weak convergence of measures) and is accomplished by proving uniqueness of trajectories. The convexity of the rate-regions and the nature of the scheduling policy (maximising a linear functional over a convex set) play an important part in not only proving the existence and uniqueness of solutions to the idempotent equations but also in providing a simple expression for the solution; coordinate convexity is then key in obtaining a further simplified expression for the solution. The LDP result then follows directly from the uniqueness of trajectories proved in Theorem 4.1 by using exactly the same argument as in the proof of Lemma 3.1 in Puhalskii and Vladimirov [30]; interested readers are also referred to the proof of Theorem 3.1 in Puhalskii [29].

Assume we are given a function $\chi^{\mathbf{A}} : \Re_+^K \to [0,+\infty]$ that attains zero at some $\boldsymbol{\mu} \in \Re_+^K$ such that $\boldsymbol{\mu} < \boldsymbol{\lambda}^*$ with $\boldsymbol{\lambda}^* \in \mathcal{R}(K)$ where the inequality holds coordinate-wise. This is used to define function $\mathbf{I}^{\mathbf{A}} : \mathbb{D}(\Re_+^K) \to \Re_+$ by

$$\mathbf{I}^{\mathbf{A}}(\mathbf{a}) = \int_0^1 \chi^{\mathbf{A}}\left(\dot{\mathbf{a}}(t)\right) dt$$

if the function $\mathbf{a} = \big(\mathbf{a}(t),\ t \in [0,1]\big) \in \mathbb{D}(\Re_+^K)$ is absolutely continuous such that $\mathbf{a}(0) = \mathbf{0}$, and where $\dot{\mathbf{a}} \in \mathcal{L}^1\big([0,1];\Re^K\big)$ is the (Lebesgue) *a.e.* derivative of $\mathbf{a}$ taking values in $\Re_+^K$; $\mathbf{I}^A(\mathbf{a})$ is defined to be equal to $+\infty$ if the function $\mathbf{a} \in \mathbb{D}(\Re_+^K)$ does not have the above properties. For $\mathbf{x} \in \Re_+^K$ we define by $\mathbb{AC}_{\mathbf{x}}$ the set of absolutely continuous functions in $\mathbb{D}(\Re_+^K)$ such that for $\mathbf{a} \in \mathbb{AC}_{\mathbf{x}}$ we have $\mathbf{a}(0) = \mathbf{x}$, and where $\dot{\mathbf{a}} \in \mathcal{L}^1([0,1];\Re^K)$ takes values in $\Re_+^K$. It is assumed that $\mathbf{I}^{\mathbf{A}}(\cdot)$ is an action functional on $\mathbb{D}(\Re_+^K)$ which in turn implies that $\chi^{\mathbf{A}}(\cdot)$ is an action functional on $\Re_+^K$. The function $\chi^{\mathbf{A}}(\cdot)$ being convex and lower compact is sufficient (see Dembo and Zajic [13, Lemma 8, pg. 203]) for $\mathbf{I}^{\mathbf{A}}(\cdot)$ to be an action functional.

We assume that the arrival process of workloads is such that $\{\mathbf{A}^N,\ N \in \mathbb{N}\}$ satisfies an LDP at rate N as $N \to +\infty$ in the space $\mathbb{D}(\Re_+^K)$ with action functional $\mathbf{I}^{\mathbf{A}}(\cdot)$. If the arrival process of workloads is such that $\{\mathbf{A}_m\}_{m\geq 0}$ is a sequence of *i.i.d.* $\Re_+^K$ valued random variables with $\mathbb{E}\left(e^{<\mathbf{x},\mathbf{A}_0>}\right) < +\infty$ for all $\mathbf{x} \in \Re^K$ where $< \cdot,\cdot >$ is the inner product on $\Re^K$, then by the Borovkov-Mogulskii theorem (see Puhalskii [28, Theorem 2.15, pg. 25], Borovkov [5], Mogulskii [25, 26]) we have $\{\mathbf{A}^N,\ N \in \mathbb{N}\}$ satisfying an LDP at rate N as $N \to +\infty$ in the space $\mathbb{D}(\Re_+^K)$ with action functional $\mathbf{I}^{\mathbf{A}}(\cdot)$ with

$$\chi^{\mathbf{A}}(\mathbf{x}) = \sup_{\mathbf{y}\in\Re^K} \left(\big\langle \mathbf{y},\mathbf{x} \big\rangle - \log \mathbb{E}\Big(e^{\big\langle \mathbf{y},\mathbf{A}_0 \big\rangle}\Big) \right) \quad \forall\, \mathbf{x} \in \Re_+^K. \tag{19}$$

See Dembo and Zajic [13, Theorem 5, pg. 216] for general mixing conditions on a stationary sequence $\{\mathbf{A}_m\}$ so that $\{\mathbf{A}^N,\ N \in \mathbb{N}\}$ satisfies an LDP at rate N as $N \to +\infty$ in the space $\mathbb{D}(\Re_+^K)$ with action functional $\mathbf{I}^{\mathbf{A}}(\cdot)$.

Under the above assumptions about the arrival processes we now prove the $\mathbb{C}(\mathbb{X})$-exponential tightness of the sequence $(\mathbf{A}^N, \mathbf{R}^N, \mathbf{Y}^N, \mathbf{S}^N, \Psi^N, \mathbf{W}^N)$.

Theorem 3.1 *Assume that the arrival process is such that the sequence $\{\mathbf{A}^N,\ N \in \mathbb{N}\}$ satisfies an LDP at rate N as $N \to +\infty$ in the space $\mathbb{D}(\Re_+^K)$ with action functional $\mathbf{I}^{\mathbf{A}}(\cdot)$. Also assume*[3] *that $\frac{\mathbf{W}_0^N}{N} \xrightarrow{\mathbb{P}^{1/N}} \mathbf{w}(0)$. Then the sequence $(\mathbf{A}^N, \mathbf{R}^N, \mathbf{Y}^N, \mathbf{S}^N, \Psi^N, \mathbf{W}^N)$ is $\mathbb{C}(\mathbb{X})$-exponentially tight on order N in $\mathbb{D}(\mathbb{X})$.*

If an idempotent process $(\mathbf{a}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \mathbf{s}, \Phi, \mathbf{w})$ defined on an indempotent probability space (Υ, Π) and having trajectories in $\mathbb{C}(\mathbb{X})$ is a limit point of $(\mathbf{A}^N, \mathbf{R}^N, \mathbf{Y}^N, \mathbf{S}^N, \Psi^N, \mathbf{W}^N)$ for LD convergence in distribution at rate N, then the following properties hold for any limit point $(\mathbf{a}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \mathbf{s}, \Phi, \mathbf{w})$:

(i) the function $\mathbf{a}$ is ($\Pi-$a.e.) component-wise non-negative and non-decreasing, and absolutely continuous with $\mathbf{a}(0) = \mathbf{0}$;

(ii) the function $\boldsymbol{\gamma}$ is ($\Pi-$a.e.) component-wise non-negative and non-decreasing, and Lipschitz continuous with $\boldsymbol{\gamma}(0) = \mathbf{0}$;

(iii) the function $\boldsymbol{\eta}$ is ($\Pi-$a.e.) component-wise non-negative and non-decreasing, and Lipschitz continuous with $\boldsymbol{\eta}(0) = \mathbf{0}$;

[3] This can be relaxed with a suitable LDP assumption - LDP with good rate function, for example.

(iv) *the function* $\mathbf{s}$ *is* ($\Pi-$*a.e.) component-wise non-negative and non-decreasing, and Lipschitz continuous with* $\mathbf{s}(0) = \mathbf{0}$*;*

(v) *the measure-valued function* Φ *is* ($\Pi-$*a.e.) such that* $\Phi(t) \in \mathcal{M}^t(\mathcal{R}(K))$ *for all* $t \in [0,1]$*, absolutely continuous with respect to the total variation norm (see Yosida [42, pg. 35-38, 118-119]) such that* $\Phi(t) - \Phi(u) \in \mathcal{M}^{t-u}(\mathcal{R}(K))$ *for all* $1 \geq t \geq u \geq 0$ *with* $\Phi(0)(\mathcal{R}(K)) = 0$*, and possesses a weak derivative* $\dot{\Phi}(t)$ *for almost every* $t \in [0,1]$*. The following inequality holds* ($\Pi-$*a.e.) for all* $t, u \in [0,1]$ *with* $t \geq u$

$$s^k(t) - s^k(u) \leq \big\langle \Phi(t) - \Phi(u), e_k \big\rangle, \tag{20}$$

where $e_k : \mathcal{R}(K) \to \Re_+$ *is the* k^{th}*-coordinate projection operator such that* $e_k(\mathbf{r}) = r_k$*;*

(vi) *the function* $\mathbf{w}$ *is* ($\Pi-$*a.e.) component-wise non-negative, and absolutely continuous with* $\mathbf{w}(0)$ *given such that*

$$\dot{\mathbf{w}}(t) = \dot{\mathbf{a}}(t) - \dot{\mathbf{s}}(t) \quad \textit{for (Lebesgue) a.e. } t \in [0,1]; \tag{21}$$

(vii) *if* $w^k(t) > 0$ *for* $t \in [t_1, t_2]$ *for* $t_1,\ t_2 \in [0,1]$*, then it follows that*

$$s^k(t) - s^k(u) = \big\langle \Phi(t) - \Phi(u), e_k \big\rangle \ \forall\, t \geq u \ \textit{with } t,\ u\ \in [t_1, t_2]; \tag{22}$$

(viii) ($\Pi-$*a.e.) for (Lebesgue) almost every* $t \in [0,1]$*,*

$$\dot{\Phi}(t)\left(\mathcal{R}(K) \setminus \tilde{H}(\mathbf{w}(t))\right) = 0; \ \textit{and} \tag{23}$$

(ix) ($\Pi-$*a.e.) for every* $k = \{1, 2, \ldots, K\}$ *we have*

$$\begin{aligned} \eta^k(t) &= \max\!\left(0, \sup_{0 \leq s \leq t}\left(\gamma^k(s) - a^k(s)\right) - w^k(0)\right) \\ &= \max\!\left(w^k(0), \sup_{0 \leq s \leq t}\left(\gamma^k(s) - a^k(s)\right)\right) - w^k(0). \end{aligned} \tag{24}$$

and for (Lebesgue) almost every $t \in [0,1]$ *it follows that*

$$\dot{\eta}^k(t) = \begin{cases} \left(\dot{\gamma}^k(t) - \dot{a}^k(t)\right)_+ & \textit{if } w^k(t) = 0; \\ 0 & \textit{otherwise.} \end{cases} \tag{25}$$

Thus $\Pi-$*a.e. every limit point is an absolutely continuous solution (strong solution) of the following differential inclusion for (Lebesgue) almost all* $t \in [0,1]$*:*

$$\dot{\mathbf{w}}(t) - \dot{\boldsymbol{\eta}}(t) \in \dot{\mathbf{a}}(t) - \tilde{H}(\mathbf{w}(t)) \tag{26}$$

with $\mathbf{w}(0)$ *the initial condition such that* $\dot{\boldsymbol{\eta}}(t) \geq \mathbf{0}$ *and* $\dot{\eta}^k(t) w^k(t) = 0$ *for all* $k = 1, 2, \ldots, K$*.*

Proof. Refer to Appendix A. □

We will show the existence of solutions of (26), and show uniqueness of the solution and other properties as a consequence of Theorem 4.1 in Section 4.

Remarks:

(i) The deterministic problem (26) is an instantiation of the Skorohod problem (see Chen and Yao [10, Chapter 7]), and in that setting one would term $\mathbf{w}$ the *reflected* process and $\boldsymbol{\eta}$ the *regulator*, respectively. In the language of Atar et al. [2] we are seeking solutions to a *constrained discontinuous media problem* where the domain $\mathcal{G}$ is $\Re_+^K$, the constraint vector field $\mathcal{D}$ are the normal directions of constraint on the boundary of $\mathcal{G}$ and the velocity vector field is determined by the maximal monotone map $\tilde{H}(\cdot)$ that has domain $\mathcal{G}$.

(ii) If we make the additional assumption that $\chi^{\mathbf{A}}(\boldsymbol{\nu}) > 0$ for all $\boldsymbol{\nu} \in \Re_+^K \setminus \{\boldsymbol{\mu}\}$ and $\{\mathbf{A}_m\}_{m \geq 0}$ being stationary, then we can construct a (regular) fluid limit (a functional strong law of large numbers result) that will obey an relationship similar to (26) given by absolutely continuous solutions to

$$\dot{\mathbf{w}}(t) - \dot{\boldsymbol{\eta}}(t) \in \boldsymbol{\mu} - \tilde{H}(\mathbf{w}(t)) \tag{27}$$

with $\mathbf{w}(0)$ the initial condition such that $\sum_{k=1}^{K} w^k(0) = 1$. Now one can easily argue for stability (existence of stationary regime and stationary distribution, see Andrews et al. [1], Chen and Yao [10], and Foss and Konstantopoulos [19]) by using a quadratic Lyapunov function $V(t) := \frac{\|\sqrt{\boldsymbol{\beta}} \circ \mathbf{w}(t)\|^2}{2}$. The details are very similar to those in Andrews et al. [1] and are skipped for brevity. However, it is worth noting that one needs to use the property that for every $k \in \{1, 2, \ldots, K\}$ we have $w^k(t)\dot{\eta}^k(t) = 0$. Uniform integrability (see Billingsley [4, Equation 3.15, pg. 31]) of the sequence $\{\mathbf{A}^N(t)\}$ (for some $t > 0$) would need to be proved for the result to hold. Under the conditions for the Borovkov-Mogulskii theorem, namely, $\{\mathbf{A}_m\}_{m \geq 0}$ an *i.i.d.* sequence with $\mathbb{E}(e^{\langle \mathbf{x}, \mathbf{A}_0 \rangle}) < +\infty$ for all $\mathbf{x} \in \Re^K$, we can use Hölder's inequality to prove the uniform integrability requirement using a construction described in Botvich and Duffield [6]. The proof is as follows. We have for $x > 0$

$$\begin{aligned}\mathbb{E}\left(e^{x\frac{A^k(-1,\lfloor Nt \rfloor]}{N}}\right) &= \mathbb{E}\left(\prod_{i=0}^{\lfloor Nt \rfloor} e^{\frac{x}{N}A_i^k}\right) \\ &\leq \prod_{i=0}^{\lfloor Nt \rfloor} \mathbb{E}(e^{\frac{x(1+\lfloor Nt \rfloor)}{N}A_i^k})^{\frac{1}{1+\lfloor Nt \rfloor}} \\ &= \mathbb{E}(e^{\frac{x(1+\lfloor Nt \rfloor)}{N}A_0^k}) \leq \mathbb{E}(e^{x(1+t)A_0^k}) < +\infty.\end{aligned} \tag{28}$$

Note that this bound only uses stationarity, and hence, is applicable under the conditions of Dembo and Zajic [13, Theorem 5, pg. 216]. For a real-valued non-negative random variable X with distribution $\mathbb{P}$ using $x \leq ye^{x-y}$ for $x \geq y$ for all $y \geq 1$ we have the following bound

$$\int_y^{+\infty} x d\mathbb{P}(x) \leq ye^{-y} \int_y^{+\infty} e^x d\mathbb{P}(x) \leq ye^{-y}\mathbb{E}(e^X).$$

Using this bound and the bound in (28) developed using Hölder's inequality, uniform integrability follows.

Now we state our main result.

Theorem 3.2 *Assume that the sequence of arrival processes* $\{\mathbf{A}^N,\ N \in \mathbb{N}\}$ *satisfy an LDP at rate* N *as* $N \to +\infty$ *in the space* $\mathbb{D}(\Re_+^K)$ *with action functional* $\mathbf{I}^{\mathbf{A}}(\cdot)$*. Also assume that* $\frac{\mathbf{W}^N(0)}{N} \xrightarrow{\mathbb{P}^{1/N}} \mathbf{w}(0)$*. Then the sequence* $\mathbf{W}^N$ *obeys an LDP for scale* N *in the Skorohod space* $\mathbb{D}(\Re_+^K)$ *with action functional* $\mathbf{I}_{\mathbf{w}(0)}^{\mathbf{W}}(\cdot)$ *given in* (50)*.*

We defer the proof and the identification of the action functional to Section 4. We can immediately write down a corollary to Theorem 3.2 that considers many applications of the result.

Corollary 3.1 *Under the conditions of Theorem 3.2 with* $\mathbf{w}(0) = \mathbf{0}$*, for* $\mathbf{x} \in \Re_+^K$*,* $x \in \Re_+$ *and* $t \in [0, 1]$ *we have*

$$\limsup_{N \to +\infty} \frac{\log\bigl(\mathbb{P}(\mathbf{W}(\lfloor Nt \rfloor) \geq N\mathbf{x})\bigr)}{N} \leq - \inf_{\mathbf{y} \in \Re_+^K : \mathbf{y} \geq \mathbf{x}} \mathbf{J}(\mathbf{y}, t), \tag{29}$$

$$\liminf_{N \to +\infty} \frac{\log\bigl(\mathbb{P}(\mathbf{W}(\lfloor Nt \rfloor) > N\mathbf{x})\bigr)}{N} \geq - \inf_{\mathbf{y} \in \Re_+^K : \mathbf{y} > \mathbf{x}} \mathbf{J}(\mathbf{y}, t), , \tag{30}$$

where

$$\mathbf{J}(\mathbf{x}, t) := \inf_{\mathbf{w} \in \mathbb{C}_{\mathbf{0}}(\Re_+^K) : \mathbf{w}(t) = \mathbf{x}} \mathbf{I}_{\mathbf{0}}^{\mathbf{W}}(\mathbf{w}). \tag{31}$$

Furthermore, we also have

$$\limsup_{N\to+\infty}\frac{\log\big(\mathbb{P}(\max_{k=1,2,\ldots,K}W^k(\lfloor Nt\rfloor)\geq Nx)\big)}{N}\leq-\min_{k=1,2,\ldots,K}\inf_{\mathbf{y}\in\Re_+^K:y^k\geq x}\mathbf{J}(\mathbf{y},t),\tag{32}$$

$$\liminf_{N\to+\infty}\frac{\log\big(\mathbb{P}(\max_{k=1,2,\ldots,K}W^k(\lfloor Nt\rfloor)>Nx)\big)}{N}\geq-\min_{k=1,2,\ldots,K}\inf_{\mathbf{y}\in\Re_+^K:y^k>x}\mathbf{J}(\mathbf{y},t).\tag{33}$$

$$\limsup_{N\to+\infty}\frac{\log\big(\mathbb{P}(\sum_{k=1,2,\ldots,K}W^k(\lfloor Nt\rfloor)\geq Nx)\big)}{N}\leq-\inf_{\mathbf{y}\in\Re_+^K:\sum_{k=1}^K y^k\geq x}\mathbf{J}(\mathbf{y},t),\tag{34}$$

$$\liminf_{N\to+\infty}\frac{\log\big(\mathbb{P}(\sum_{k=1,2,\ldots,K}W^k(\lfloor Nt\rfloor)>Nx)\big)}{N}\geq-\inf_{\mathbf{y}\in\Re_+^K:\sum_{k=1}^K y^k>x}\mathbf{J}(\mathbf{y},t).\tag{35}$$

$$\limsup_{N\to+\infty}\frac{\log\big(\mathbb{P}(\max_{k=1,2,\ldots,K}\sup_{t\in[0,1]}W^k(\lfloor Nt\rfloor)\geq Nx)\big)}{N}\leq-\min_{k=1,2,\ldots,K}\inf_{\mathbf{y}\in\Re_+^K:y^k\geq x}\inf_{t\in[0,1]}\mathbf{J}(\mathbf{y},t),\tag{36}$$

$$\liminf_{N\to+\infty}\frac{\log\big(\mathbb{P}(\max_{k=1,2,\ldots,K}\sup_{t\in[0,1]}W^k(\lfloor Nt\rfloor)>Nx)\big)}{N}\geq-\min_{k=1,2,\ldots,K}\inf_{\mathbf{y}\in\Re_+^K:y^k>x}\inf_{t\in[0,1]}\mathbf{J}(\mathbf{y},t).\tag{37}$$

$$\limsup_{N\to+\infty}\frac{\log\big(\mathbb{P}(\sup_{t\in[0,1]}\sum_{k=1,2,\ldots,K}W^k(\lfloor Nt\rfloor)\geq Nx)\big)}{N}\leq-\inf_{\mathbf{y}\in\Re_+^K:\sum_{k=1}^K y^k\geq x}\inf_{t\in[0,1]}\mathbf{J}(\mathbf{y},t),\tag{38}$$

$$\liminf_{N\to+\infty}\frac{\log\big(\mathbb{P}(\sup_{t\in[0,1]}\sum_{k=1,2,\ldots,K}W^k(\lfloor Nt\rfloor)>Nx)\big)}{N}\geq-\inf_{\mathbf{y}\in\Re_+^K:\sum_{k=1}^K y^k>x}\inf_{t\in[0,1]}\mathbf{J}(\mathbf{y},t).\tag{39}$$

Proof. Since the coordinate projection map $\pi_t:\mathbb{D}(\mathbb{X})\to\Re_+^K$ for $t\in\Re_+$ given by $\pi_t\big((\mathbf{a},\boldsymbol{\gamma},\boldsymbol{\eta},\mathbf{s},\Phi,\mathbf{w})\big)=\mathbf{w}(t)$ for $(\mathbf{a},\boldsymbol{\gamma},\boldsymbol{\eta},\mathbf{s},\Phi,\mathbf{w})\in\mathbb{D}(\mathbb{X})$ is $\mathbb{C}(\mathbb{X})$-continuous and Borel measurable with the Skorohod J_1 topology, the results in (29) and (30), and the expression for the rate function in (31) follow from a generalised version of the *contraction principle* in Puhalskii [27, Corollary 3.1.22, pg. 264]. Now the results in (32) and (33) follow by an application of the *Principle of the Largest Term* (see Dembo and Zeitouni [12, Lemma 1.2.15, pg. 7] and Puhalskii [27, Definition 1.1.1 and Lemma 1.1.4, pp. 5–6] or equivalently the same generalised *contraction principle* from Puhalskii [27, Corollary 3.1.22, pg. 264]). Since the function that maps $\mathbf{y}\in\Re_+^K$ to $\sum_{k=1}^K y^k$ (in $\Re_+$) is continuous, another application of the generalised *contraction principle* from Puhalskii [27, Corollary 3.1.22, pg. 264] yields (34) and (35). Similarly (36), (37), (38) and (39) follow from the $\mathbb{C}(\mathbb{X})$-continuity (and Borel measurability with the Skorohod J_1 topology) of $\sup_{t\in[0,1]}w_t^k,\forall k\in\{1,\ldots,K\}$ for $(\mathbf{a},\boldsymbol{\gamma},\boldsymbol{\eta},\mathbf{s},\Phi,\mathbf{w})\in\mathbb{D}(\mathbb{X})$. □

The restriction $\mathbf{w}(0)=\mathbf{0}$ is only to simplify our further characterisation of $\mathbf{J}(\mathbf{x},t)$. Note that (36) and (37) are useful in calculating the tail probabilities of the workload when each user has a buffer to itself and (38) and (39) are useful when there is a shared buffer.

3.1 Polytope rate-regions If, in addition, the rate region $\mathcal{R}(K)$ is a polytope, then one can use the method of types (see Dembo and Zeitouni [12, Section 2.1.1]) to cast the result of Theorem 3.1 in a simpler setting.

Let the extreme points of $\mathcal{R}(K)$ be $\{\mathbf{r}_p\}_{\{p=1,2,\ldots,P\}}$ with $\mathbf{r}_1$ equalling the all zero vector in $\Re^K$; in other words, we include the origin in the set of extreme points. Since our policy either chooses the extreme points or chooses a minimum norm solution from the convex hull of a subset of extreme points we can build up a bigger finite set of operating points $\{\mathbf{r}_q\}_{\{q=1,2,\ldots,Q\}}$ with $Q\leq 1+2^{P-1}$ such that the mapping from $\mathbf{x}$ to the operating points is single-valued. This we do by looking at the minimum norm solution corresponding to the largest subset $\tilde{P}(\mathbf{x})$ of $\{1,2,\ldots,P\}$ such that $\mathbf{r}_p\in\arg\max_{\mathbf{r}\in\mathcal{R}(K)}\langle\mathbf{x},\mathbf{r}\rangle$ for every $p\in\tilde{P}$ for given $\mathbf{x}\in\Re^K$. Denote the label of operating point corresponding to point $\mathbf{x}\in\Re^K$ by $\tilde{q}(\mathbf{x})$. Again we assume that $\mathbf{r}_1$ equals the all zero vector in $\Re^K$.

For $q\in\{1,2,\ldots,Q\}$ define ψ_{qm} as follows

$$\psi_{qm}=\begin{cases}1 & \text{if } q=\tilde{q}(\mathbf{W}_m)\\ 0 & \text{otherwise}\end{cases},$$

then $\sum_{l=0}^m\psi_{ql}$ counts the number of times until time $m+1$ that operating point q is picked. Using the sequence $\{\psi_{qm}\}$ we can rewrite the queueing equation (16) as follows

$$\begin{aligned} W^k_{m+1} &= W^k_0 - S^k(-1,m] + A^k(-1,m] \\ &= W^k_0 - \sum_{q=1}^{Q}\sum_{l=0}^{m} \psi_{ql} \min(W^k_l, r^k_q) + A^k(-1,m]. \end{aligned} \tag{40}$$

Define for every $q = 1, 2, \ldots, Q$ the number of times in $[0, \lfloor t \rfloor]$ that operating point q is chosen to be $\psi_q(-1, \lfloor t \rfloor] := \sum_{l=0}^{\lfloor t \rfloor} \psi_{ql}$. For scale N and starting workload vector $\mathbf{W}^N_0$, define $\psi^N_q(t) := \frac{\psi^N_q(-1, \lfloor Nt \rfloor]}{N}$ (in vector notation $\boldsymbol{\psi}^N(t) = \big(\psi^N_1(t), \psi^N_2(t), \ldots, \psi^N_Q(t)\big)$); denote the process by $\boldsymbol{\psi}^N := (\boldsymbol{\psi}^N(t),\ t \in [0,1])$.

Here we would define $\mathbb{X} = \Re^K_+ \times \Re^K_+ \times \Re^K_+ \times \Re^K_+ \times \Re^Q_+ \times \Re^K_+$. Then we would show $\mathbb{C}(\mathbb{X})$-exponential tightness on order N in $\mathbb{D}(\mathbb{X})$ of sequence $\Big(\mathbf{A}^N, \mathbf{R}^N, \mathbf{Y}^N, \mathbf{S}^N, \boldsymbol{\psi}^N, \mathbf{W}^N\Big)$. Then each limit point $(\mathbf{a}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \mathbf{s}, \boldsymbol{\phi}, \mathbf{w})$ in $\mathbb{C}(\mathbb{X})$ would satisfy the following (modified) properties $\Pi-$*a.e.*:

(v) the function $\boldsymbol{\phi}$ is component-wise non-negative and non-decreasing, and Lipschitz continuous with $\boldsymbol{\phi}(0) = \mathbf{0}$. The following inequality holds for all $1 \leq t \geq u \geq 0$

$$s^k(t) - s^k(u) \leq \sum_{q=2}^{Q} (\phi_q(t) - \phi_q(u)) r^k_q; \tag{41}$$

(vii) if $w^k(t) > 0$ for $t \in [t_1, t_2]$ for $t_1 > t_2 \in [0,1]$, then it follows that

$$s^k(t) - s^k(u) = \sum_{q=2}^{Q} (\phi_q(t) - \phi_q(u)) r^k_q \quad \forall\, t \geq u \text{ with } t,\ u\ \in [t_1, t_2];\ \text{and} \tag{42}$$

(viii) for any regular point $t \in [0,1]$ and for a chosen operating $\hat{q} \in \{1, 2, \ldots, Q\}$, if

$$\langle \boldsymbol{\beta} \circ \mathbf{w}(t), \mathbf{r}_{\hat{q}} \rangle \quad < \max_{q=1,2,\ldots,Q} \langle \boldsymbol{\beta} \circ \mathbf{w}(t), \mathbf{r}_q \rangle\,, \tag{43}$$

then $\frac{d\phi_{\hat{q}}(t)}{dt} = 0$.

4. Analysis of Fluid Limit

Before addressing the main result of this section we state and prove a few preliminary results that will be key for the analysis of the fluid limit. We are interested in the properties of $H(\mathbf{x}) = \arg\max_{\mathbf{r} \in \mathcal{R}(K)} \langle \mathbf{x}, \mathbf{r} \rangle$ and $\tilde{H}(\mathbf{x}) = \arg\max_{\mathbf{r} \in \mathcal{R}(K)} \langle \boldsymbol{\beta} \circ \mathbf{x}, \mathbf{r} \rangle$ for $\mathbf{x} \in \Re^K$.

Let $\mathfrak{X}$ be a Hilbert space. A set-valued map $\mathcal{H}$ from $\mathfrak{X}$ to $\mathfrak{P}(\mathfrak{X})$ (the power set of $\mathfrak{X}$) with domain $\mathrm{Dom}(\mathcal{H})$ is *monotone* (see Brézis [7, Definition 2.1, pg. 20] and Rockafellar and Wets [32, Chapter 12]) if and only if

$$\forall \mathfrak{x}_1,\ \mathfrak{x}_2 \in \mathrm{Dom}(F),\ \ \forall \mathfrak{v}_i \in \mathcal{H}(\mathfrak{x}_i),\ \ i = 1, 2, \quad \langle \mathfrak{v}_1 - \mathfrak{v}_2, \mathfrak{x}_1 - \mathfrak{x}_2 \rangle \geq 0, \tag{44}$$

where $< \cdot, \cdot >$ is the inner-product on $\mathfrak{X}$. A *monotone* set-valued map $\mathcal{H}$ is *maximal* (see Brézis [7, Definition 2.2, pg. 22] and Rockafellar and Wets [32, Chapter 12]) if there is no other monotone set-valued map $\tilde{\mathcal{H}}$ whose graph strictly contains the graph of $\mathcal{H}$. The reader is referred to Brézis [7], Browder [8], Rockafellar [31], Rockafellar and Wets [32] for the properties of monotone maps, maximal monotone maps and their connections to convex analysis, functional analysis and semigroups of non-expansive maps.

Lemma 4.1 *$H(\mathbf{x}) = \arg\max_{\mathbf{r} \in \mathcal{R}(K)} \langle \mathbf{x}, \mathbf{r} \rangle$ and $\tilde{H}(\mathbf{x}) = \arg\max_{\mathbf{r} \in \mathcal{R}(K)} \langle \boldsymbol{\beta} \circ \mathbf{x}, \mathbf{r} \rangle$ for $\mathbf{x} \in \Re^K$ are maximal monotone maps from $\Re^K$ to $\mathcal{R}(K) \subset \Re^K$.*

Proof. Since $\mathcal{R}(K)$ is a proper, closed convex set, from Rockafellar [31, Theorem 23.6, Corollary 23.5.1 and Corollary 23.5.3, pp. 218–219] the definition of $H(\mathbf{x})$ characterises all the subgradients[4] of a proper, lower-semicontinuous and convex function $\max_{\mathbf{r} \in \mathcal{R}(K)} \langle \mathbf{x}, \mathbf{r} \rangle$ from $\Re^K$ to $\Re \cup \{+\infty\}$. Now using

[4] Following Rockafellar [31, pg. 214] for a convex function $F(\mathbf{x}) : \Re^K \to \Re$ with domain $\mathcal{G} \subseteq \Re^K$ (a convex set), a vector $\tilde{\mathbf{x}} \in \Re^K$ is a *subgradient* of $F(\mathbf{x})$ at $\mathbf{x}$ if

$$F(\mathbf{y}) \geq F(\mathbf{x}) + \langle \tilde{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle, \quad \forall\, \mathbf{y} \in \mathcal{G},$$

where $\langle \cdot, \cdot \rangle$ is the inner product in $\Re^K$. We denote the set of subgradients of $F(\mathbf{x})$ at $\mathbf{x}$ by $\partial F(\mathbf{x})$.

Brézis [7, Example 2.3.4, pg. 25], Rockafellar [31, Corollary 31.5.2, pg. 340] and Rockafellar and Wets [32, Theorem 12.17, pg. 542] we can assert that $H(\mathbf{x})$ is a *maximal monotone* map. The exact same proof applies to $\tilde{H}(\mathbf{x})$ too. □

As described in Section 3 each policy in the class of Max-Weight scheduling policies can be associated with a unique vector $\tilde{\boldsymbol{\beta}} \in \Re_+^K$ such that $\tilde{\boldsymbol{\beta}} > \mathbf{0}$ and $\sum_{k=1}^K \tilde{\beta}^k = 1$. We now demonstrate that the performance of a Max-Weight policy with a given $\tilde{\boldsymbol{\beta}}$ can be quantified by analysing the performance of a Max-Weight policy with weights $\boldsymbol{\beta} = \underbrace{\left(\frac{1}{K}, \frac{1}{K}, \ldots, \frac{1}{K}\right)}_{K \text{ times}}$ but with a new rate-region that is a scaled version of the original rate region. Consider the set-valued map $\tilde{H}(\mathbf{x}) = \arg\max_{\mathbf{r} \in \mathcal{R}(K)} \left\langle \tilde{\boldsymbol{\beta}} \circ \mathbf{x}, \mathbf{r} \right\rangle$ for all $\mathbf{x} \in \Re_+^K$. Then the differential inclusion (26) that we need to analyse is $\dot{\mathbf{w}}(t) - \dot{\boldsymbol{\eta}}(t) \in \dot{\mathbf{a}}(t) - \tilde{H}(\mathbf{w}(t))$ where $\mathbf{a} \in \mathbb{AC}_\mathbf{0}$. Now it is clear that

$$\arg\max_{\mathbf{r} \in \mathcal{R}(K)} \left\langle \tilde{\boldsymbol{\beta}} \circ \mathbf{x}, \mathbf{r} \right\rangle = \arg\max_{\mathbf{r} \in \mathcal{R}(K)} \left\langle \sqrt{\tilde{\boldsymbol{\beta}}} \circ \mathbf{x}, \sqrt{\tilde{\boldsymbol{\beta}}} \circ \mathbf{r} \right\rangle,$$

where $\sqrt{\tilde{\boldsymbol{\beta}}}$ is defined by taking square-root coordinate-wise. Define

$$\mathcal{R}_{\sqrt{\tilde{\boldsymbol{\beta}}}}(K) := \left\{ \mathbf{r} \in \Re^K : \frac{1}{\sqrt{\tilde{\boldsymbol{\beta}}}} \circ \mathbf{r} \in \mathcal{R}(K) \right\},$$

which is a scaled version of $\mathcal{R}(K)$, and again both convex, coordinate convex and compact, and

$$\hat{H}_{\sqrt{\tilde{\boldsymbol{\beta}}}}(\mathbf{x}) := \arg\max_{\mathbf{r} \in \mathcal{R}_{\sqrt{\tilde{\boldsymbol{\beta}}}}(K)} \langle \mathbf{x}, \mathbf{r} \rangle,$$

which is equivalent to $\left\{ \mathbf{r} \in \Re^K : \frac{1}{\sqrt{\tilde{\boldsymbol{\beta}}}} \circ \mathbf{r} \in \tilde{H}(\mathbf{x}) \right\}$, i.e., a scaling of $\tilde{H}(\mathbf{x})$. Therefore we can now claim the following equivalence

$$\dot{\mathbf{w}}(t) - \dot{\boldsymbol{\eta}}(t) \in \dot{\mathbf{a}}(t) - \tilde{H}(\mathbf{w}(t)) = \dot{\mathbf{a}}(t) - \frac{1}{\sqrt{\tilde{\boldsymbol{\beta}}}} \circ \hat{H}_{\sqrt{\tilde{\boldsymbol{\beta}}}}\left(\sqrt{\tilde{\boldsymbol{\beta}}} \circ \mathbf{w}(t)\right)$$

$$\Updownarrow$$

$$\sqrt{\tilde{\boldsymbol{\beta}}} \circ \dot{\mathbf{w}}(t) - \sqrt{\tilde{\boldsymbol{\beta}}} \circ \dot{\boldsymbol{\eta}}(t) \in \sqrt{\tilde{\boldsymbol{\beta}}} \circ \dot{\mathbf{a}}(t) - \hat{H}_{\sqrt{\tilde{\boldsymbol{\beta}}}}\left(\sqrt{\tilde{\boldsymbol{\beta}}} \circ \mathbf{w}(t)\right).$$

Therefore setting $\tilde{\mathbf{w}}(t) := \sqrt{\tilde{\boldsymbol{\beta}}} \circ \mathbf{w}(t)$, $\tilde{\boldsymbol{\eta}}(t) = \sqrt{\tilde{\boldsymbol{\beta}}} \circ \boldsymbol{\eta}(t)$ and $\tilde{\mathbf{a}}(t) := \sqrt{\tilde{\boldsymbol{\beta}}} \circ \mathbf{a}(t)$ it suffices to analyse the following differential inclusion with

$$\dot{\tilde{\mathbf{w}}}(t) - \dot{\tilde{\boldsymbol{\eta}}}(t) \in \dot{\tilde{\mathbf{a}}}(t) - \arg\max_{\mathbf{r} \in \mathcal{R}_{\sqrt{\tilde{\boldsymbol{\beta}}}}(K)} \langle \tilde{\mathbf{w}}(t), \mathbf{r} \rangle = \dot{\tilde{\mathbf{a}}}(t) - \hat{H}_{\sqrt{\tilde{\boldsymbol{\beta}}}}(\tilde{\mathbf{w}}(t)). \tag{45}$$

Note that Lemma 4.1 still applies and both the underlying domain $\mathcal{G} = \Re_+^K$ and the constraint vector field $\mathcal{D}$ do not change. Thus, without loss of generality, from now onwards we assume that $\boldsymbol{\beta} = \underbrace{\left(\frac{1}{K}, \frac{1}{K}, \ldots, \frac{1}{K}\right)}_{K \text{ times}}$ and analyse the solutions of the following differential inclusion

$$\forall t \in [0,1] \quad \dot{\mathbf{w}}(t) - \dot{\boldsymbol{\eta}}(t) \in \dot{\mathbf{a}}(t) - H(\mathbf{w}(t)), \tag{46}$$

where $\mathbf{a} \in \mathbb{AC}_\mathbf{0}$, $\mathbf{w}(0) \in \Re^K$ is given, $H(\cdot) : \Re^K \to \mathfrak{P}\left(\Re^K\right)$ is a *maximal monotone* set-valued map with closed domain $\mathrm{Dom}(H)$ possessing a non-empty interior and $\dot{\boldsymbol{\eta}}(t)$ takes values in $-\partial\tilde{\delta}(\mathbf{w}(t)|\mathrm{Dom}(H))$ where

$$\tilde{\delta}(\mathbf{x}|\mathrm{Dom}(H)) = \begin{cases} 0 & \text{if } \mathbf{x} \in \mathrm{Dom}(H) \\ +\infty & \text{otherwise} \end{cases}$$

is the (convex) indicator function of $\mathrm{Dom}(H)$ and $\partial\tilde{\delta}(\mathbf{x}|\mathrm{Dom}(H))$ is the set of all subgradients of $\tilde{\delta}(\mathbf{w}(t)|\mathrm{Dom}(H))$. In other words, $\dot{\boldsymbol{\eta}}(t)$ takes values among the normal directions of constraint on the boundary of $\mathrm{Dom}(H)$ depending upon $\mathbf{w}(t)$.

Instead of just seeking a solution to (46) when $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$, in Cépa [9] (weak) solutions for $\mathbf{a} \in \mathbb{C}_{\mathbf{w}(0)}(\Re_+^K)$ were analysed. Using the results of Cépa [9] we then have the following theorem where we define $\mathrm{cl}(\mathcal{S})$ to be the closure of set $\mathcal{S}$ and the (*a.e.*) right derivative at time $t \in [0,1]$ of an absolutely continuous function $\mathbf{w} \in \mathbb{AC}$ to be $\frac{d^+\mathbf{w}}{dt}(t)$.

Theorem 4.1 *Let $\mathcal{H}$ be a maximal monotone map such that its domain $\mathrm{Dom}(\mathcal{H})$ has a non-empty interior $\mathrm{int}(\mathrm{Dom}(\mathcal{H}))$. If $\mathbf{w}(0) \in \mathrm{cl}(\mathrm{Dom}(\mathcal{H}))$, then for $\mathbf{a} \in \mathbb{C}_{\mathbf{0}}(\Re_+^K)$ there exists a unique (weak) solution of*

$$\dot{\mathbf{w}}(t) \in \dot{\mathbf{a}}(t) - \mathcal{H}(\mathbf{w}(t)) \quad \forall t \in [0,1], \tag{47}$$

with $\mathbf{w} \in \mathbb{C}_{\mathbf{w}(0)}(\Re_+^K)$ taking values in $\mathrm{cl}(\mathrm{Dom}(\mathcal{H}))$. The map from $\mathbf{w}(0) + \mathbf{a}$ to $\mathbf{w}$ is continuous with the uniform topology on $\mathbb{C}(\Re_+^K)$. If, in addition, $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$, then $\mathbf{x} \in \mathbb{AC}_{\mathbf{w}(0)}$ (and is a strong solution) with right derivative $\frac{d^+\mathbf{w}}{dt}$ given by

$$\frac{d^+\mathbf{w}}{dt}(t) = \dot{\mathbf{a}}(t) - \mathrm{Proj}_{\mathcal{H}(\mathbf{w}(t))}\left(\dot{\mathbf{a}}(t)\right). \tag{48}$$

Proof. The first part is sufficient to prove Theorem 3.2 and follows from Cépa [9, Theorem 3.2]. However, the second part aids explicit calculations and follows from a combination of Brézis[7, Theorems 3.4 and 3.5, and Proposition 3.8]. Also see Browder [8, Theorems 9.25 and 9.26]. □

Remarks:

(i) For a comprehensive survey of the Skorohod problem and the state of the art, the reader is referred to Atar et al. [2].

(ii) From the above discussion it is also clear that the results of Cépa [9] answer in the affirmative questions about the existence and uniqueness of fluid limits (e.g. [41]) such as (27) in the context of Max-Weight scheduling for the single-server set-up that we have assumed. Additionally, following (48) the solution is such that the right derivative is given by

$$\frac{d^+\mathbf{w}}{dt}(t) = \boldsymbol{\mu} - \mathrm{Proj}_{\mathcal{H}(\mathbf{w}(t))}\left(\boldsymbol{\mu}\right). \tag{49}$$

Now we spell out the details of the proof of Theorem 3.2.

Proof of Theorem 3.2. First note that $H(\cdot)$ is *maximal monotone* with $\mathrm{Dom}(H) = \Re_+^K$. Additionally, from Rockafellar [31, pp. 215–216, 226] and Brézis [7, Example 2.3.4, pg. 25], Rockafellar [31, Corollary 31.5.2, pg. 340] and Rockafellar and Wets [32, Theorem 12.17, pg. 542] we have $\partial\tilde{\delta}(\cdot|\mathrm{Dom}(H))$ also being a maximal monotone map, again with domain $\Re_+^K$. Therefore using Brézis[7, Corollary 2.7] we have $H(\cdot) + \partial\tilde{\delta}(\cdot|\mathrm{Dom}(H))$ also being a maximal monotone map[5]. If we take the map $\mathcal{H}(\cdot)$ to be $H(\cdot) + \partial\tilde{\delta}(\cdot|\mathrm{Dom}(H))$, then we can apply Theorem 4.1. Now if one considers the function in $\mathbb{C}_{\mathbf{w}(0)}(\Re_+^K)$ given by $\hat{\mathbf{a}} = \mathbf{w}(0) + \mathbf{a}$ for $\mathbf{a} \in \mathbb{C}_{\mathbf{0}}(\Re_+^K)$, then using the results of Cépa [9] summarised in Theorem 4.1, one can show the existence of a unique continuous (weak) solution to (46) for continuous input $\hat{\mathbf{a}}$ such that the map from $\hat{\mathbf{a}}$ to $\mathbf{w}$ is continuous with the uniform topology on $\mathbb{C}(\Re_+^K)$. With the uniqueness of trajectories for every $\mathbf{a} \in \mathbb{C}_{\mathbf{0}}(\Re_+^K)$, we can now follow the argument of the proof from Puhalskii and Vladimirov [30, Lemma 3.1] to prove the LD convergence result; since the argument is exactly the same we do not reproduce it here. Owing to relative compactness (established using exponential tightness) one gets LD convergence along subsequences. Uniqueness of trajectories then allows one to show that the limit along every subsequence has to be the same, which automatically yields convergence.

The action functional is also immediate from the argument of the proof from Puhalskii and Vladimirov [30, Lemma 3.1]. Define the composite map from $\mathbf{a} \in \mathbb{C}_{\mathbf{0}}(\Re_+^K)$ to $\mathbf{w}$ through $\mathbf{w}(0) + \mathbf{a}$ to be $\mathcal{T}$. Let $\mathcal{T}_A$ be the image under $\mathcal{T}$ of absolutely continuous functions $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$. Note that $\mathcal{T}_A$ is a subset of the set of absolutely continuous functions from $[0,1]$ with initial value $\mathbf{w}(0)$. Then the action functional for the LDP result $\mathbf{I}^{\mathbf{W}}_{\mathbf{w}(0)}(\cdot)$ is given as follows: if $\mathbf{w} \in \mathcal{T}_A$, then

$$\mathbf{I}^{\mathbf{W}}_{\mathbf{w}(0)}(\mathbf{w}) = \mathbf{I}^{\mathbf{A}}\left(\mathcal{T}^{-1}(\mathbf{w})\right) = \inf_{\mathbf{a} \in \mathbb{AC}_{\mathbf{0}} : \mathcal{T}(\mathbf{a}) = \mathbf{w}} \mathbf{I}^{\mathbf{A}}(\mathbf{a}), \tag{50}$$

and for every other $\mathbf{w} \in \mathbb{C}(\Re_+^K)$ we set $\mathbf{I}^{\mathbf{W}}_{\mathbf{w}(0)}(\mathbf{w})$ to $+\infty$. □

[5] $H(\cdot) + \partial\tilde{\delta}(\cdot|\mathrm{Dom}(H))$ at $\mathbf{x}$ is the set $\{\mathbf{y} \in \Re^k : \mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 \text{ s.t. } \mathbf{y}_1 \in H(\mathbf{x}) \text{ and } \mathbf{y}_2 \in \partial\tilde{\delta}(\mathbf{x}|\mathrm{Dom}(H))\}$.

Remark: Following the proof from Puhalskii [29, Theorem 3.1] it is sufficient to prove the existence of unique solutions to (46) using results from Brézis [7] for $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$ such that $\mathbf{I}^{\mathbf{A}}(\mathbf{a}) < +\infty$, for the LDP result to hold or to an even smaller set of functions that determine the rate function. The characterisation in (48) would also still apply. However, we feel that using the results of Cépa [9] provides us with a more complete characterisation.

For $\mathbf{x} \in \Re_+^K$ let $\mathcal{K}'(\mathbf{x}) := \{k : x_k = 0\}$ be the set of coordinates that are zero. Then using the coordinate convexity of $\mathcal{R}(K)$ it is immediate that $H(\mathbf{x})$ is coordinate convex in each of the coordinates in $\mathcal{K}'(\mathbf{x})$. From this observation and noting that $\partial\tilde{\delta}(\mathbf{x}|\Re_+^K) = \{\mathbf{y} \in \Re^K : y_k \leq 0 \ \forall\ k \in \mathcal{K}'(\mathbf{x}) \text{ and } y_k = 0\ \forall k \in \{1, 2, \ldots, K\} \setminus \mathcal{K}'(\mathbf{x})\}$ we obtain the following lemma.

Lemma 4.2 *For every* $\mathbf{x} \in \Re_+^K$ *and* $\mathbf{a} \in \Re_+^K$ *we have*

$$\mathrm{Proj}_{H(\mathbf{x})+\partial\tilde{\delta}(\mathbf{x}|\Re_+^K)}(\mathbf{a}) = \mathrm{Proj}_{H(\mathbf{x})}(\mathbf{a}). \tag{51}$$

Proof. Let $\mathbf{y}^* := \mathrm{Proj}_{H(\mathbf{x})}(\mathbf{a}) \in H(\mathbf{x})$. We know that $\mathbf{y}^*$ is the unique element in $H(\mathbf{x})$ such that

$$\langle \mathbf{y}^* - \mathbf{a}, \mathbf{y} - \mathbf{y}^* \rangle \geq 0 \qquad \forall\ \mathbf{y} \in H(\mathbf{x}). \tag{52}$$

Fix $k \in \mathcal{K}'(\mathbf{x})$. Since $H(\mathbf{x})$ is coordinate convex along k, we have $\tilde{\mathbf{y}} \in H(\mathbf{x})$ where given $\epsilon \in [0,1)$ we set

$$\forall k' \in \{1, 2, \ldots, K\} \quad \tilde{y}_{k'} = \begin{cases} y_{k'}^* & \text{if } k' \neq k; \\ \epsilon y_{k'}^* & \text{otherwise.} \end{cases}$$

Therefore, substituting $\tilde{\mathbf{y}}$ in (52) we get

$$(y_k^* - a_k)y_k^* \leq 0. \tag{53}$$

This holds for every $k \in \mathcal{K}'(\mathbf{x})$.

Since $\mathcal{R}(K)$ is coordinate convex with non-empty interior, it is easy to argue using (53) the following for any given $\mathbf{x} \in \Re_+^K$:

(i) if $a_{k'} = 0$ for some $k' \in \mathcal{K}'(\mathbf{x})$, then $y_{k'}^* = 0$;

(ii) if we have $y_k^* = 0$ for some $k \in \mathcal{K}'(\mathbf{x})$ for which $a_k > 0$, then for all $\mathbf{y} \in H(\mathbf{x})$ we must have $y_k = 0$; and

(iii) if we have a $\mathbf{y} \in H(\mathbf{x})$ such that $y_k > 0$ for some $k \in \mathcal{K}'(\mathbf{x})$ for which $a_k > 0$, then $y_k^* > 0$ and $y_k^* \leq a_k$.

Therefore in all cases we have $y_k^* \leq a_k$ for all $k \in \mathcal{K}'(\mathbf{x})$.

Now any $\mathbf{y}' \in H(\mathbf{x}) + \partial\tilde{\delta}(\mathbf{x}|\Re_+^K)$ is such that we can write $\mathbf{y}' = \tilde{\mathbf{y}} + \hat{\mathbf{y}}$ where $\tilde{\mathbf{y}} \in H(\mathbf{x})$ and $\hat{\mathbf{y}} \in \partial\tilde{\delta}(\mathbf{x}|\Re_+^K)$. Such a decomposition need not be unique and we do not need uniqueness for the proof. Note that $\hat{y}_k = 0$ for all $k \in \{1, 2, \ldots, K\} \setminus \mathcal{K}'(\mathbf{x})$ and $\hat{y}_k \leq 0$ for $k \in \mathcal{K}'(\mathbf{x})$. Therefore we have

$$\begin{aligned} \langle \mathbf{y}^* - \mathbf{a}, \mathbf{y}' - \mathbf{y}^* \rangle &= \langle \mathbf{y}^* - \mathbf{a}, \tilde{\mathbf{y}} - \mathbf{y}^* \rangle + \sum_{k \in \mathcal{K}'(\mathbf{x})} (y_k^* - a_k)\hat{y}_k \\ &\geq \langle \mathbf{y}^* - \mathbf{a}, \tilde{\mathbf{y}} - \mathbf{y}^* \rangle \geq 0, \end{aligned}$$

where the penultimate inequality follows because $y_k^* - a_k \leq 0$ and $\hat{y}_k \leq 0$ for all $k \in \mathcal{K}'(\mathbf{x})$. This completes the proof. □

An elementary but extremely useful conclusion from Lemma 4.2 is the following Corollary.

Corollary 4.1 *If* $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$*, then* $\mathbf{w} \in \mathbb{AC}_{\mathbf{w}(0)}$ *such that* $\mathbf{w} = \mathcal{T}(\mathbf{a})$ *has right derivative* $\frac{d^+\mathbf{w}}{dt}$ *given by*

$$\begin{aligned} \frac{d^+\mathbf{w}}{dt}(t) &= \dot{\mathbf{a}}(t) - \mathrm{Proj}_{H(\mathbf{w}(t))+\partial\tilde{\delta}(\mathbf{w}(t)|\Re_+^K)}\left(\dot{\mathbf{a}}(t)\right) \\ &= \dot{\mathbf{a}}(t) - \mathrm{Proj}_{H(\mathbf{w}(t))}\left(\dot{\mathbf{a}}(t)\right) \end{aligned} \tag{54}$$

Applying this to the fluid limit says that the fluid limit (27) and (49) has an even simpler characterisation where the right derivative is given by

$$\frac{d^+\mathbf{w}}{dt}(t) = \boldsymbol{\mu} - \mathrm{Proj}_{H(\mathbf{w}(t))}\left(\boldsymbol{\mu}\right)$$

Now that the underlying rate function $\mathbf{I}^{\mathbf{W}}_{\mathbf{w}(0)}(\cdot)$ has been specified, we set out to derive an alternate expression for $\mathbf{J}(\mathbf{x}, t)$ (see (31)) that converts the calculus of variations problem to a finite-dimensional optimisation; note that $\mathbf{w}(0) = \mathbf{0}$ in the definition of $\mathbf{J}(\mathbf{x}, t)$. In the process we will show that for determining the rate function, it suffices to consider piece-wise linear functions (illustrated in Figure 1) $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$ determined by two parameters $u \in [0, t]$ for $t \in (0, 1]$ and $\boldsymbol{\lambda} \in \Re^K_+ \setminus \mathcal{R}(K)$ such that for $v \in [0, 1]$ we have

$$\mathbf{a}(v) = \begin{cases} \boldsymbol{\mu} v & \text{if } v \in [0, u]; \\ \boldsymbol{\lambda}(v - u) + \boldsymbol{\mu} u & \text{if } v \in [u, t]; \\ \boldsymbol{\mu}(v - t + u) + \boldsymbol{\lambda}(t - u) & \text{if } v \in [t, 1]. \end{cases} \tag{55}$$

This is proved in the following Lemma.

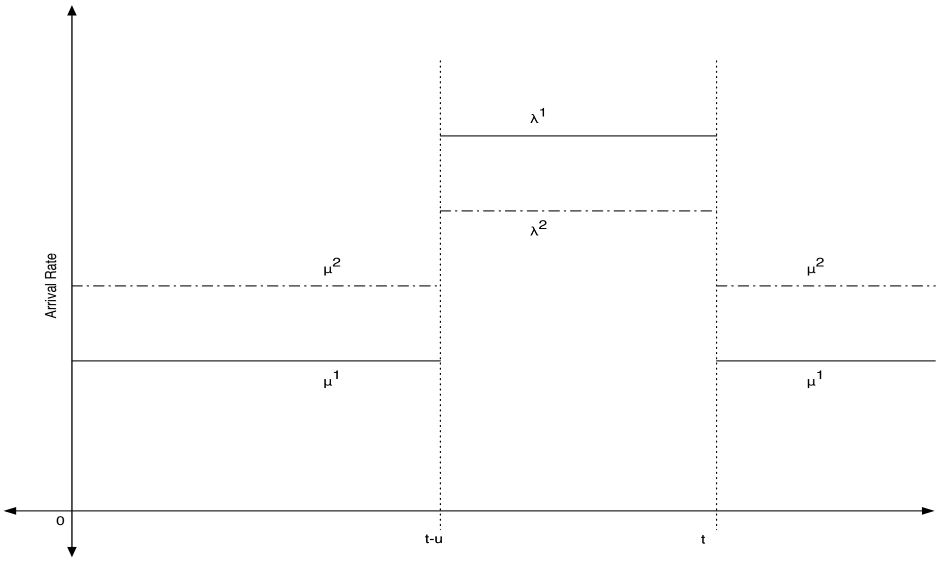

Figure 1: Typical element of class of $\dot{\mathbf{a}}$ considered for optimisation.

Lemma 4.3 *If $\chi^{\mathbf{A}}(\mathbf{x}) : \Re^K_+ \to \Re_+$ is convex with $\chi^{\mathbf{A}}(\boldsymbol{\mu}) = 0$ for some $\boldsymbol{\mu} \in \Re^K_+$ with $\boldsymbol{\mu} \in \mathcal{R}(K)$ such that there exists $\boldsymbol{\lambda}^* \in \mathcal{R}(K)$ with $\boldsymbol{\mu} < \boldsymbol{\lambda}^*$, then for $\mathbf{x} \in \Re^K_+$ and $t \in (0, 1]$ we have*

$$\mathbf{J}(\mathbf{x}, t) = \begin{cases} \inf_{u \in (0,t]} u \inf_{\boldsymbol{\lambda} \in \arg\max_{\mathbf{r} \in \mathcal{R}} <\mathbf{x}, \mathbf{r}>} \chi^{\mathbf{A}}\left(\frac{\mathbf{x}}{u} + \boldsymbol{\lambda}\right) & \textit{if } \mathbf{x} \neq \mathbf{0}; \\ 0 & \textit{otherwise.} \end{cases} \tag{56}$$

Proof. We can rewrite $\mathbf{J}(\mathbf{x}, t)$ for $\mathbf{x} \in \Re_+$ and $t \in (0, 1]$ as

$$\mathbf{J}(\mathbf{x}, t) = \inf_{\mathbf{a} \in \mathbb{AC}_{\mathbf{0}} : \mathcal{T}(\mathbf{a})(t) = \mathbf{x}} \int_0^1 \chi^{\mathbf{A}}(\dot{\mathbf{a}}(u)) du \tag{57}$$

Since we are only interested in the behaviour of the workload in $[0, t]$, it suffices to consider $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$ such that $\mathcal{T}(\mathbf{a})(t) = \mathbf{x}$ and $\dot{\mathbf{a}}(u) = \boldsymbol{\mu}$ for all $u \in (t, 1]$; note that $\int_t^1 \chi^{\mathbf{A}}(\dot{\mathbf{a}}(u))du$ will be 0 with this restriction, and since $\boldsymbol{\mu}$ is strictly inside $\mathcal{R}(K)$ (as given by the assumptions above), after a finite time (depending on $\mathbf{x}$) $\mathbf{w}$ will reduce to $\mathbf{0}$ and remain there [6].

[6] As mentioned in Section 3 this can be shown using Lyapunov function $V(t) = \frac{1}{2}\|\mathbf{w}(t)\|^2$.

For a given $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$ define $u^* = \sup\{u \in [0,t] : \mathbf{w}(u) = \mathbf{0}\}$, i.e., the last time $\mathbf{w}$ is zero in all coordinates before time t. Since $\int_0^{u^*} \chi^{\mathbf{A}}(\dot{\mathbf{a}}(u))du \geq 0$ and $\mathbf{w}(u^*) = \mathbf{0}$ (by continuity) we can reduce the cost by setting $\dot{\mathbf{a}}(u) = \boldsymbol{\mu}$ for all $u \in [0, u^*]$ all the while ensuring $\mathbf{w}(u^*) = \mathbf{0}$. Thus for $\mathbf{x} \neq \mathbf{0}$ it suffices to solve the following calculus of variations problem for $u \in (0, t]$

$$\inf_{\substack{\mathbf{a} \in \mathbb{AC}_{\mathbf{0}} : \mathcal{T}(\mathbf{a})(u) = \mathbf{x}, \\ \mathcal{T}(\mathbf{a})(v) \neq \mathbf{0} \ \forall\ 0 < v \leq u}} \int_0^u \chi^{\mathbf{A}}(\dot{\mathbf{a}}(v))dv \tag{58}$$

From Corollary 4.1 any $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$ that achieves $\mathcal{T}(\mathbf{a})(u) = \mathbf{x}$ and, in particular, any $\mathbf{a} \in \mathbb{AC}_{\mathbf{0}}$ such that $\mathcal{T}(\mathbf{a})(v) \neq \mathbf{0}\ \forall\ 0 < v \leq u$, satisfies the following equation

$$\mathbf{x} = \mathbf{a}(u) - \int_0^u \mathrm{Proj}_{H(\mathbf{w}(v))}(\dot{\mathbf{a}}(v))dv. \tag{59}$$

Now using Jensen's inequality (see Rockafellar [31]) and (59) we have

$$\begin{aligned}\int_0^u \chi^{\mathbf{A}}(\dot{\mathbf{a}}(v))dv &\geq u\, \chi^{\mathbf{A}}\left(\frac{1}{u}\int_0^u \dot{\mathbf{a}}(v)dv\right) = u\, \chi^{\mathbf{A}}\left(\frac{\mathbf{a}(u)}{u}\right) \\ &= u\, \chi^{\mathbf{A}}\left(\frac{\mathbf{x}}{u} + \frac{\int_0^u \mathrm{Proj}_{H(\mathbf{w}(v))}(\dot{\mathbf{a}}(v))dv}{u}\right).\end{aligned}$$

If one can now find a constant $\boldsymbol{\lambda} \in \Re_+^K$ such that $\frac{\mathbf{x}}{u} + \mathrm{Proj}_{H\left(\frac{\mathbf{x}v}{u}\right)}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}$ for all $v \in [0, u]$, then it suffices to optimise over the possible values $(\Lambda(\mathbf{x}, u))$ of the constant $\boldsymbol{\lambda}$ to solve the calculus of variations problem in (58). The answer to (58) would then be $u \inf_{\boldsymbol{\lambda} \in \Lambda(\mathbf{x},u)} \chi^{\mathbf{A}}(\boldsymbol{\lambda})$ with $\mathbf{w}(v) = \frac{\mathbf{x}v}{u}$ for $v \in [0, u]$, and

$$\mathbf{J}(\mathbf{x}, t) = \inf_{u \in (0,t]} u \inf_{\boldsymbol{\lambda} \in \Lambda(\mathbf{x},u)} \chi^{\mathbf{A}}(\boldsymbol{\lambda}). \tag{60}$$

The workload trajectories (up to time t) that result from the considered class of arrivals is illustrated in Figure 2 where for the sake of illustration, for $v \in [0, t]$ we take $\mathbf{a}(v) = \boldsymbol{\mu} \min(v, t-u) + \boldsymbol{\lambda}(v - t + u)_+$ for some $\boldsymbol{\lambda} \in \Re_+^K \setminus \mathcal{R}(K)$.

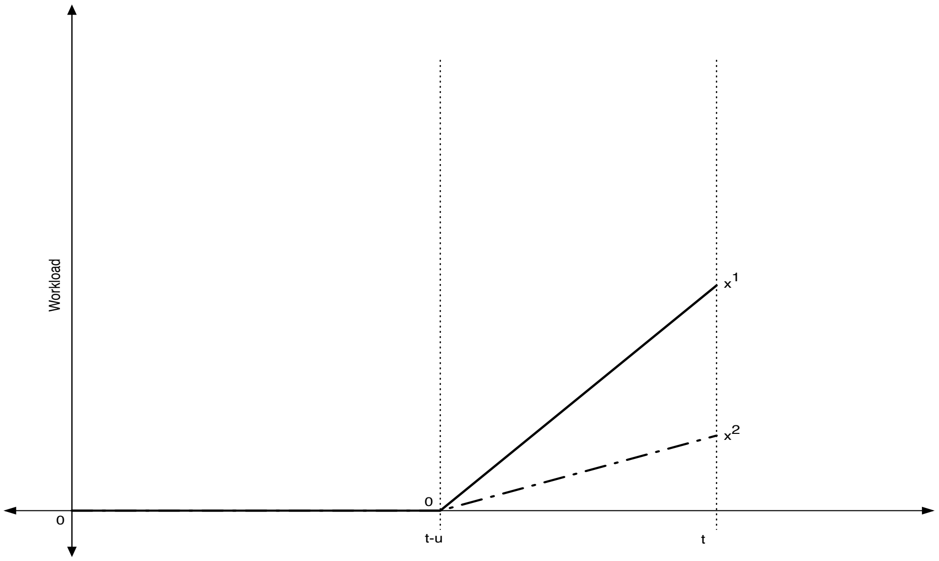

Figure 2: Typical workload trajectory until time t for the analysed class of input functions $\mathbf{a}$.

Since we are maximising a linear functional it is clear that $H\left(\frac{\mathbf{x}v}{u}\right) = H(\mathbf{x})$ for $v \in (0, u]$ and $H\left(\frac{\mathbf{x}v}{u}\right) = \mathcal{R}(K)$ for $v = 0$. Thus we need to solve for $\boldsymbol{\lambda}$ that solves both

$$\frac{\mathbf{x}}{u} + \mathrm{Proj}_{H(\mathbf{x})}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}, \text{ and} \tag{61}$$

$$\frac{\mathbf{x}}{u} + \mathrm{Proj}_{\mathcal{R}(K)}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}. \tag{62}$$

Pick $\mathbf{r}^*, \mathbf{r} \in \mathcal{R}(K)$, then

$$\left\|\frac{\mathbf{x}}{u} + \mathbf{r}^* - \mathbf{r}\right\|^2 = \|\mathbf{r}^* - \mathbf{r}\|^2 + \left\|\frac{\mathbf{x}}{u}\right\|^2 + 2\left\langle \frac{\mathbf{x}}{u}, \mathbf{r}^*\right\rangle - 2\left\langle \frac{\mathbf{x}}{u}, \mathbf{r}\right\rangle.$$

Therefore

$$\begin{aligned}\min_{\mathbf{r}\in\mathcal{R}(K)} \left\|\frac{\mathbf{x}}{u} + \mathbf{r}^* - \mathbf{r}\right\|^2 &\geq \left\|\frac{\mathbf{x}}{u}\right\|^2 + 2\left\langle \frac{\mathbf{x}}{u}, \mathbf{r}^*\right\rangle + \min_{\mathbf{r}\in\mathcal{R}(K)} \|\mathbf{r}^* - \mathbf{r}\|^2 - 2\max_{\mathbf{r}\in\mathcal{R}(K)} \left\langle \frac{\mathbf{x}}{u}, \mathbf{r}\right\rangle \\ &= \left\|\frac{\mathbf{x}}{u}\right\|^2 + 2\left\langle \frac{\mathbf{x}}{u}, \mathbf{r}^*\right\rangle - 2\max_{\mathbf{r}\in\mathcal{R}(K)} \left\langle \frac{\mathbf{x}}{u}, \mathbf{r}\right\rangle,\end{aligned}$$

since $\mathbf{r}^* \in \mathcal{R}(K)$. Now $\min_{\mathbf{r}\in\mathcal{R}(K)} \left\|\frac{\mathbf{x}}{u} + \mathbf{r}^* - \mathbf{r}\right\|^2 \leq \left\|\frac{\mathbf{x}}{u}\right\|^2$ since we can set $\mathbf{r} = \mathbf{r}^*$. From this it follows that if $\mathbf{r}^* \in H(\mathbf{x})$, then $\mathbf{r}^* = \mathrm{Proj}_{\mathcal{R}(K)}\left(\frac{\mathbf{x}}{u} + \mathbf{r}^*\right)$.

If $\mathbf{r}^* = \mathrm{Proj}_{\mathcal{R}(K)}\left(\frac{\mathbf{x}}{u} + \mathbf{r}^*\right)$, then $2\left\langle \mathbf{r}^* - \frac{\mathbf{x}}{u} - \mathbf{r}^*, \mathbf{r}^* - \mathbf{r}\right\rangle \leq 0$ for all $\mathbf{r} \in \mathcal{R}(K)$, which is another way of saying that $\mathbf{r}^* \in H(\mathbf{x})$.

For $\mathbf{r}^* \in H(\mathbf{x})$ we also have

$$\begin{aligned}\min_{\mathbf{r}\in H(\mathbf{x})} \left\|\frac{\mathbf{x}}{u} + \mathbf{r}^* - \mathbf{r}\right\|^2 &= \left\|\frac{\mathbf{x}}{u}\right\|^2 + \min_{\mathbf{r}\in H(\mathbf{x})} \|\mathbf{r}^* - \mathbf{r}\|^2 \qquad \left(\text{Since } \left\langle \frac{\mathbf{x}}{u}, \mathbf{r}^*\right\rangle = \left\langle \frac{\mathbf{x}}{u}, \mathbf{r}\right\rangle\right) \\ &= \left\|\frac{\mathbf{x}}{u}\right\|^2,\end{aligned}$$

and $\mathbf{r}^* = \mathrm{Proj}_{H(\mathbf{x})}\left(\frac{\mathbf{x}}{u} + \mathbf{r}^*\right)$

Thus any $\boldsymbol{\lambda} \in \frac{\mathbf{x}}{u} + H(\mathbf{x})$ solves both (61) and (62), and there are no other solutions, i.e., $\Lambda(\mathbf{x}, u) = \frac{\mathbf{x}}{u} + H(\mathbf{x})$ and the result holds. □

Remarks:

(i) Lemma 4.3 informs us that for a given $\mathcal{R}(K)$ to quantify the performance one needs to characterise $H(\mathbf{x})$ for all $\mathbf{x} \in \Re^K_+$. Note that this is equivalent to a complete characterisation of the boundary of $\mathcal{R}(K)$.

(ii) It is also worth noting that $\mathbf{J}(\mathbf{x}, t)$ need not be convex in its arguments. We will explicitly show this later on using some examples.

(iii) The starting point being $\mathbf{0}$ was critically exploited in the characterisation of the critical sample-paths. However, it is possible to argue with polytope rate-regions that one only needs to consider piece-wise linear sample-paths. The reason for this is that using Corollary 4.1 it is clear that under Max-Weight policies the state-space can be partitioned into a finite set of cones where the scheduling policy is fixed. Now for any sample-path Jensen's inequality can be used to show that in every interval where the workload remains within each of these cones, the cost of the path can be bettered by choosing an arrival rate vector that is constant in this interval. This then parallels Dupuis et al. [16, Lemma 4.1].

Using the expression for $\mathbf{J}(\mathbf{x}, t)$ we can now write down simpler expressions for (36), (37), (38) and (39). First consider (36) and without loss of generality let $k = 1$. Then we have the following

$$\begin{aligned}\inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq x} \inf_{t\in(0,1]} \mathbf{J}(\mathbf{y}, t) &= \inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq x} \inf_{t\in(0,1]} \inf_{u\in(0,t]} u \inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}\left(\frac{\mathbf{y}}{u} + \boldsymbol{\lambda}\right) \\ &= \inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq 1} \inf_{t\in(0,1]} \inf_{u\in(0,t]} u \inf_{\boldsymbol{\lambda}\in H(\mathbf{y}x)} \chi^{\mathbf{A}}\left(\frac{\mathbf{y}x}{u} + \boldsymbol{\lambda}\right) \\ &= \inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq 1} \inf_{t\in(0,1]} x \inf_{v\in(0,t/x]} v \inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}\left(\frac{\mathbf{y}}{v} + \boldsymbol{\lambda}\right) \\ &= x \inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq 1} \inf_{t\in(0,1]} \inf_{z\geq x/t} \frac{\inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}(\mathbf{y}z + \boldsymbol{\lambda})}{z} \\ &= x \inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq 1} \inf_{z\geq x} \frac{\inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}(\mathbf{y}z + \boldsymbol{\lambda})}{z} \\ &= x \inf_{z\geq x} \frac{\inf_{\mathbf{y}\in\Re^K_+ : y^1 \geq 1} \inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}(\mathbf{y}z + \boldsymbol{\lambda})}{z}. \qquad (63)\end{aligned}$$

The rest of the terms in (36) have a similar form. Using a similar logic we have one of the terms of (37) given by

$$\inf_{\mathbf{y}\in\Re^K_+ : y^1 > x} \inf_{t\in(0,1]} \mathbf{J}(\mathbf{y},t) = x \inf_{z\geq x} \frac{\inf_{\mathbf{y}\in\Re^K_+ : y^1>1} \inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}\left(\mathbf{y}z+\boldsymbol{\lambda}\right)}{z}. \tag{64}$$

Similar logic applied to (38) and (39) yields

$$\inf_{\mathbf{y}\in\Re^K_+ : \sum_{k=1}^K y^k \geq x} \inf_{t\in(0,1]} \mathbf{J}(\mathbf{y},t) = x \inf_{z\geq x} \frac{\inf_{\mathbf{y}\in\Re^K_+ : \sum_{k=1}^K y^k \geq 1} \inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}\left(\mathbf{y}z+\boldsymbol{\lambda}\right)}{z}; \text{ and} \tag{65}$$

$$\inf_{\mathbf{y}\in\Re^K_+ : \sum_{k=1}^K y^k > x} \inf_{t\in(0,1]} \mathbf{J}(\mathbf{y},t) = x \inf_{z\geq x} \frac{\inf_{\mathbf{y}\in\Re^K_+ : \sum_{k=1}^K y^k > 1} \inf_{\boldsymbol{\lambda}\in H(\mathbf{y})} \chi^{\mathbf{A}}\left(\mathbf{y}z+\boldsymbol{\lambda}\right)}{z}. \tag{66}$$

These expressions can be simplified further with additional assumptions on the arrival processes (such as independence) and the rate-region $\mathcal{R}(K)$.

5. Examples We will conclude by presenting three example rate-regions to show how the analysis developed above applies. The first is a two-user elliptical rate-region. The last two examples are obtained from Information Theory (see Cover and Thomas [11]). The first of these considers a two-user Gaussian broadcast channel and the second a symmetrical two-user multiple-access channel. For the remainder of this section we will set the scheduling weight vector $\boldsymbol{\beta}$ to $(1/2, 1/2)$. Note from the analysis in Section 4 that other values of $\boldsymbol{\beta}$ can be analysed by modifying the parameters of $\mathcal{R}(2)$.

5.1 Example I: A Two-User Queue With An Elliptical Rate-Region Consider a specific enunciation of our model with two users such that the rate region $\mathcal{R}(2)$ is a quadrant of an ellipse with parameters $r^M, r^m > 0$, i.e.,

$$\mathcal{R}(2) = \left\{ (r^1, r^2) \in \Re^2_+ : \left(\frac{r^1}{r^M}\right)^2 + \left(\frac{r^2}{r^m}\right)^2 \leq 1 \right\}. \tag{67}$$

Now let us solve the scheduling policy generation problem, namely, $H(\mathbf{x}) = \arg\max_{\mathbf{r}\in\mathcal{R}(2)} < \mathbf{x}, \mathbf{r} >$ for $\mathbf{x} \in \Re^2_+$. We will only consider the case of at least one coordinate being positive because $H(\mathbf{0}) = \mathcal{R}(2)$. The optimal solution is the unique point $(\tilde{r}^1, \tilde{r}^2) \in \tilde{\mathcal{R}}(2)$ that satisfies

$$\frac{\tilde{r}^1}{r^M} = \frac{r^M x^1}{\sqrt{(r^M x^1)^2 + (r^m x^2)^2}} \quad \text{and} \quad \frac{\tilde{r}^2}{r^m} = \frac{r^m x^2}{\sqrt{(r^M x^1)^2 + (r^m x^2)^2}}, \tag{68}$$

This can be derived easily using Lagrange multipliers. The rate-region and the solution of the scheduling problem are illustrated in Figure 3.

The expression for $\mathbf{J}(\mathbf{x}, t)$ with $\mathbf{x} \neq \mathbf{0}$ for this example is given by

$$\mathbf{J}(\mathbf{x},t) = \inf_{u\in(0,t]} u\, \chi^{\mathbf{A}} \left(\left(\frac{x^1}{u} + r^M \frac{r^M x^1}{\sqrt{(r^M x^1)^2 + (r^m x^2)^2}}, \frac{x^2}{u} + r^m \frac{r^m x^2}{\sqrt{(r^M x^1)^2 + (r^m x^2)^2}} \right) \right). \tag{69}$$

5.2 Example II: Two-user Gaussian Broadcast Channel The broadcast channel (see Cover and Thomas [11, Section 14.6]) models a communication system where there is one transmitter and multiple receivers who can all listen to the transmitter. The capacity region (in natural units, i.e., nats) of a two user Gaussian broadcast channel (see Cover and Thomas [11, Section 14.6]) is determined by two parameters (signal to noise ratios) $P_1 > P_2 > 0$ and is given as follows:

$$\mathcal{R}(2) = \bigcup_{\gamma\in[0,1]} \left\{ (r^1, r^2) \in \Re_+2 : r^1 \leq \frac{1}{2}\log\left(1+\gamma P_1\right),\ r^2 \leq \frac{1}{2}\log\left(\frac{1+P_2}{1+\gamma P_2}\right) \right\}. \tag{70}$$

If $P_1 = P_2 > 0$, then one gets a simplex.

The scheduling rule $H(x^1, x^2)$ with at least one coordinate positive is then given by $\left(\frac{1}{2}\log(1+\gamma^* P_1), \frac{1}{2}\log\left(\frac{1+P_2}{1+\gamma^* P_2}\right)\right)$ with

$$\gamma^* = \begin{cases} 1 & \text{if } x^1\left(1+\frac{1}{P_2}\right) \geq x^2\left(1+\frac{1}{P_1}\right); \\ 0 & \text{if } x^1 P_1 \leq x^2 P_2; \\ \frac{\frac{x^1}{P_2} - \frac{x^2}{P_1}}{x^2 - x^1} & \text{otherwise.} \end{cases} \tag{71}$$

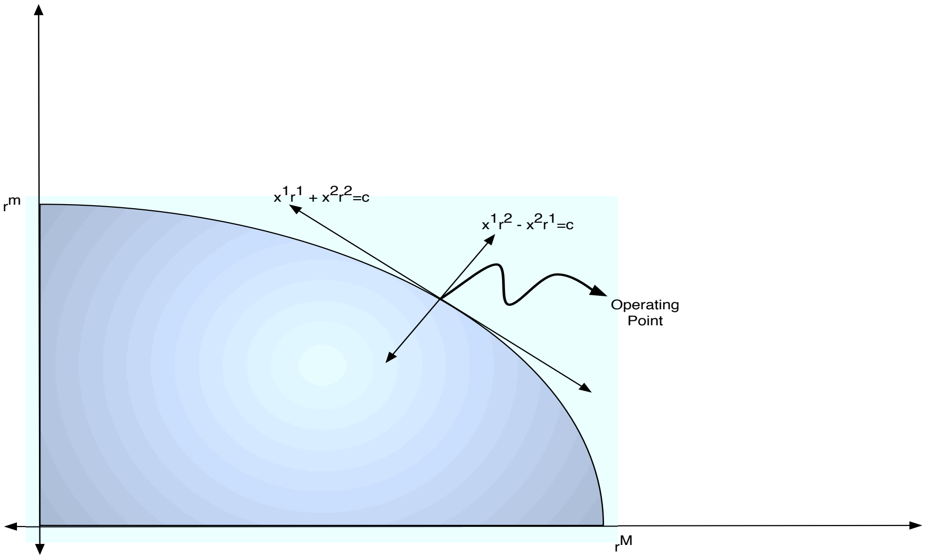

Figure 3: Illustration of an elliptical rate-region with the solution of the scheduling problem shown.

Again $H(0,0) = \mathcal{R}(2)$.

From this exercise we can now write down expressions for $\mathbf{J}(\mathbf{x}, t)$ for $\mathbf{x} \neq \mathbf{0}$ as follows:

$$\mathbf{J}(\mathbf{x}, t) = \begin{cases} \inf_{u\in(0,t]} u\, \chi^{\mathbf{A}} \left(\frac{x^1}{u} + \frac{1}{2}\log(1+P_1), \frac{x^2}{u}\right) & \text{if } x^1\left(1+\frac{1}{P_2}\right) \geq x^2\left(1+\frac{1}{P_1}\right); \\ \inf_{u\in(0,t]} u\, \chi^{\mathbf{A}} \left(\frac{x^1}{u}, \frac{x^2}{u} + \frac{1}{2}\log(1+P_2)\right) & \text{if } x^1P_1 \leq x^2P_2; \\ \inf_{u\in(0,t]} u\, \chi^{\mathbf{A}} \left(\begin{array}{c} \frac{x^1}{u} + \frac{1}{2}\log\left(\frac{x^1(P_1-P_2)}{P_2(x^2-x^1)}\right), \\ \frac{x^2}{u} + \frac{1}{2}\log\left(\frac{(1+P_2)P_1(x^2-x^1)}{x^2(P_1-P_2)}\right) \end{array}\right) & \text{otherwise.} \end{cases} \tag{72}$$

5.3 Example III: Centralised Multiple Access Channel Consider the rate region $\mathcal{R}(K)$ to be the capacity region of a K-user Multiple-Access Channel (MAC) (see Cover and Thomas [11, Section 14.3]) and Tse and Hanly [38]. At the beginning of every transmission interval each of the users communicate their queue-lengths to a centralised scheduler that then determines the operating point to be used. It was shown in Tse and Hanly [38, Lemma 3.4] that $\mathcal{R}(K)$ is a polymatroid (see Chen and Yao [10, Section 11.1]) where the rank-function is given by conditional mutual information terms. As show in Tse et al. [37] this rate region is also applicable in the asymptotic regime of (very) high signal-to-noise ratio of the multiple-input, multiple-output multiple-access channel with fading such that the receiver has perfect channel information and the transmitters have no channel information. Such a model [37] exhibits a nice trade-off between diversity and multiplexing that was used to provide performance bounds based upon the tails of the queue-lengths for Max-Weight type scheduling algorithms in Kittipiyakul and Javidi [22]. In Kittipiyakul and Javidi [22] the two-user case was analysed completely when the rate-region is a simplex, and bounds were presented when the rate-region is a symmetric polymatroid. The analysis presented here can be used to improve upon the bounds of Kittipiyakul and Javidi [22] in the general case.

Define $\mathcal{K} = \{1, 2, \ldots, K\}$ and suppose that we are given a function $f : \mathfrak{P}(\mathcal{K}) \to \Re_+$ from the power set of $\mathcal{K}$ to the (non-negative) real line. Then the polytope

$$\mathcal{R}_f(K) := \left\{\mathbf{r} \in \Re_+^K : \sum_{i\in\mathcal{J}} r^i \leq f(\mathcal{J}),\ \mathcal{J} \subseteq \mathcal{K}\right\} \tag{73}$$

is a polymatroid (see Chen and Yao [10, Section 11.1]) if the function f satisfies the following properties:

(i) (*normalised*) $f(\emptyset) = 0$;

(ii) (*increasing*) if $\mathcal{J}_1 \subseteq \mathcal{J}_2 \subseteq \mathcal{K}$, then $f(\mathcal{J}_1) \leq f(\mathcal{J}_2)$; and

(iii) (*submodular*) if $\mathcal{J}_1, \mathcal{J}_2 \subseteq \mathcal{K}$, then $f(\mathcal{J}_1) + f(\mathcal{J}_2) \geq f(\mathcal{J}_1 \cup \mathcal{J}_2) + f(\mathcal{J}_1 \cap \mathcal{J}_2)$.

A function f with these properties is called a rank function. Let π be a permutation of $\mathcal{K}$, then the vector $\mathbf{r}_\pi$ defined by

$$\begin{aligned} r_\pi^{\pi(1)} &= f(\{\pi(1)\}) \\ r_\pi^{\pi(2)} &= f(\{\pi(1),\pi(2)\}) - f(\{\pi(1)\}) \\ &\vdots \\ r_\pi^{\pi(K)} &= f(\{\pi(1),\pi(2),\dots,\pi(K)\}) - f(\{\pi(1),\pi(2),\dots,\pi(K-1)\}) \end{aligned}$$

belongs to $\mathcal{R}_f(K)$ for all permutations π. Along with $\mathbf{0}$ the points $\mathbf{r}_\pi$ are the extreme points of $\mathcal{R}_f(K)$. Also for any pair of sets $\mathcal{J}_1 \subset \mathcal{J}_2 \subseteq \mathcal{K}$, there exists a point $\mathbf{r} \in \mathcal{R}_f(K)$ such that

$$\sum_{i\in\mathcal{J}_1} r^i = f(\mathcal{J}_1) \text{ and } \sum_{i\in\mathcal{J}_2} r^i = f(\mathcal{J}_2).$$

Maximising a linear functional ($< \mathbf{x}, \mathbf{r} >$) over a polymatroid is very easy (see Chen and Yao [10, Section 11.1.2]) and is given by the following:

(i) without loss of generality assume that $x^k \geq 0$ for all $k \in \mathcal{K}$. Otherwise we simply set the corresponding $r^k = 0$ for the optimal solution;

(ii) let π be a permutation of $\mathcal{K}$ such that the weights are in decreasing order, i.e.,

$$x_{\pi(1)} \geq x_{\pi(2)} \geq \cdots \geq x_{\pi(K)} \geq 0, \tag{74}$$

then $\mathbf{r}_\pi$ is an optimal solution; and

(iii) the set of optimal solutions is the convex hull of $\mathbf{r}_\pi$ for all permutations of $\mathcal{K}$ that yield the ordering in (74).

With $\mathcal{R}_f(K)$ as the rate-region for our system, this completely specifies $H(\mathbf{x})$ for all $\mathbf{x} \in \Re_+^K$. To understand this better we will look at a class of two-user channels.

5.3.1 Symmetric Two-User Case Let $K = 2$ and given two parameters $r^M > r^m > 0$ define the rank function f as follows:

$$f(\mathcal{J}) := \begin{cases} 0 & \text{if } \mathcal{J} = \emptyset; \\ r^M & \text{if } \mathcal{J} = \{1\}; \\ r^M & \text{if } \mathcal{J} = \{2\}; \\ r^M + r^m & \text{if } \mathcal{J} = \{1,2\}. \end{cases} \tag{75}$$

Then our rate region is $\mathcal{R}_f(2)$. The edge cases for the parameters r^M, r^m do not give any new insights: if $r^M = r^m > 0$, then $\mathcal{R}_f(2)$ is a square; and if $r^M > r^m = 0$, then $\mathcal{R}_f(2)$ is a simplex. An example of this rate-region is show in Figure 4.

We now solve for $H(\mathbf{x}) = \arg\max_{\mathbf{r}\in\mathcal{R}_f(2)} < \mathbf{x}, \mathbf{r} >$ for $\mathbf{w} \in \Re_+^2$. We need to partition $\Re_+^2$ into 6 regions; these and the corresponding sets $H(\mathbf{x})$ are:

(i) Region $A = \{\mathbf{0}\}$. Here it is clear that $H(\mathbf{0}) = \mathcal{R}(2)$;

(ii) Region $B = \{x^1 > 0,\ x^2 = 0\}$. Then $H(\mathbf{x}) = \{\tilde{r}^1 = r^M,\ \tilde{r}^2 \in [0, r^m]\}$;

(iii) Region $C = \{x^1 > x^2 > 0\}$. Then $H(\mathbf{x}) = \{\tilde{r}^1 = r^M,\ \tilde{r}^2 = r^m\}$;

(iv) Region $D = \{x^1 = x^2 > 0\}$. Then $H(\mathbf{x}) = \{(\tilde{r}^1, \tilde{r}^2) \in [r^m, r^M]^2 :\ \tilde{r}^1 + \tilde{r}^2 = r^M + r^m\}$;

(v) Region $E = \{x^2 > x^1 > 0\}$. Then $H(\mathbf{x}) = \{\tilde{r}^1 = r^m,\ \tilde{r}^2 = r^M\}$; and

(vi) Region $F = \{x^2 > 0,\ x^1 = 0\}$. Then $H(\mathbf{x}) = \{\tilde{r}^1 \in [0, r^m],\ \tilde{r}^2 = r^M\}$.

The different scheduling regions are shown in Figure 5.

Using the above we can write down expressions for $\mathbf{J}(\mathbf{x}, t)$ as follows:

(i) if $\mathbf{x} \in B$, then

$$\mathbf{J}(\mathbf{x}, t) = \inf_{u\in(0,t]} u \inf_{r^2\in[0,r^m]} \chi^{\mathbf{A}}\left(\frac{x^1}{u} + r^M, r^2\right); \tag{76}$$

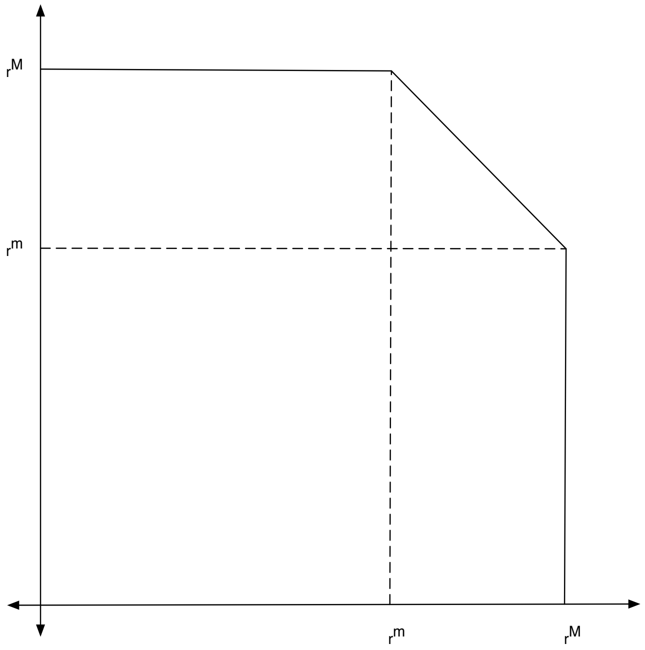

Figure 4: An example of a symmetrical two-user polymatroidal rate-region.

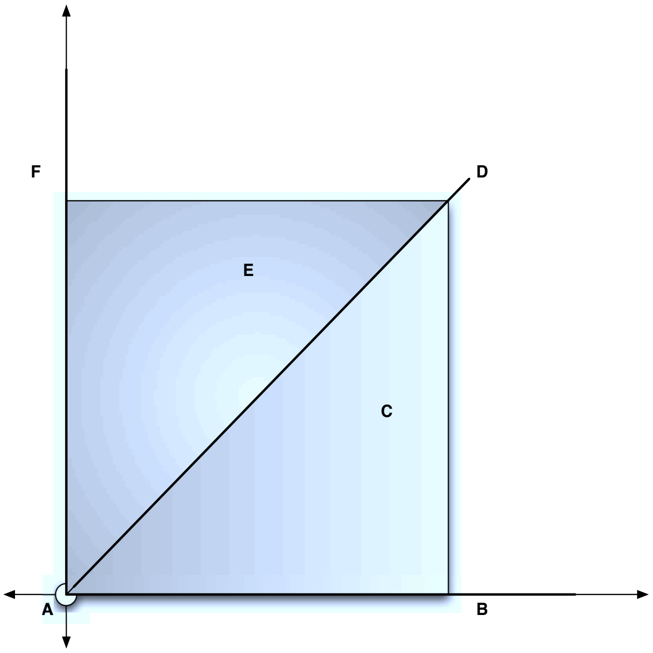

Figure 5: Partitioning of $\Re_+^2$ into regions where the same scheduling action results.

(ii) if $\mathbf{x} \in C$, then

$$\mathbf{J}(\mathbf{x}, t) = \inf_{u \in (0,t]} u\, \chi^{\mathbf{A}} \left(\frac{x^1}{u} + r^M, \frac{x^2}{u} + r^m \right); \tag{77}$$

(iii) if $\mathbf{x} \in D$, then

$$\mathbf{J}(\mathbf{x}, t) = \inf_{u \in (0,t]} u \inf_{\substack{(r^1,r^2) \in [r^m, r^M]^2: \\ r^1 + r^2 = r^M + r^m}} \chi^{\mathbf{A}} \left(\frac{x^1}{u} + r^1, \frac{x^2}{u} + r^2 \right); \tag{78}$$

(iv) if $\mathbf{x} \in E$, then

$$\mathbf{J}(\mathbf{x}, t) = \inf_{u \in (0,t]} u\, \chi^{\mathbf{A}} \left(\frac{x^1}{u} + r^m, \frac{x^2}{u} + r^M \right); \tag{79}$$

(v) if $\mathbf{x} \in F$, then

$$\mathbf{J}(\mathbf{x}, t) = \inf_{u \in (0,t]} u \inf_{r^1 \in [0, r^m]} \chi^{\mathbf{A}} \left(r^1, \frac{x^2}{u} + r^M \right). \tag{80}$$

5.4 Numerical and Simulation Results Here we demonstrate how our results can be used to predict certain tail-exponents. To validate our results we compare the numerical results to ones obtained using simulations. For simplicity we restrict our attention to two-user scenarios. We assume that the arrival processes are independent across users and across time-slots. In fact, the work brought in per slot for user 1 is assumed to be Bernoulli with mean 0.4955 and that for user 2 is assumed to be Bernoulli with mean 0.4. We consider three rate regions: circular, symmetric simplex and symmetric MAC. The rate-regions are assumed to be such that $(0.4955, 0.4)$ is strictly in the interior so as to ensure stability. The simplex rate region is such that x-asymptote and y-asymptote are $(0.9, 0)$ and $(0, 0.9)$, respectively. The MAC rate-region is obtained using the rank function from (75) where $r^M = 0.5$ and $r^m = 0.4$. Note that this corresponds to the simplex rate region being restricted to a maximum of 0.5 in either coordinate. For the circular rate region we set $r^M = r^m = \sqrt{0.5^2 + 0.4^2} = \sqrt{0.41} \approx 0.6403$ in (67). These are illustrated in Figure 6 where we also show the location of the average arrival rate vector.

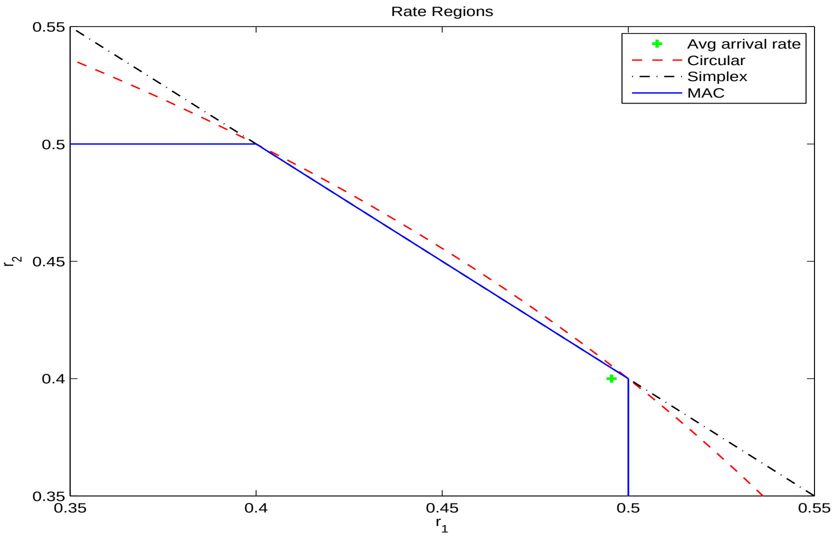

Figure 6: Rate regions used for the numerical analysis and simulations.

The numerical calculations involve evaluation of (56) for each of the rate regions and for every chosen value of $\mathbf{x}$ where the local action functional $\chi^{\mathbf{A}}(\cdot)$ is constructed using the rate function of a Bernoulli random variable and is given by

$$\chi^{\mathbf{A}}(\mathbf{x}) = \begin{cases} x_1 \log\left(\frac{x_1}{0.4955}\right) + (1 - x_1) \log\left(\frac{1-x_1}{0.5045}\right) + x_2 \log\left(\frac{x_2}{0.4}\right) + (1 - x_2) \log\left(\frac{1-x_2}{0.6}\right), & \text{if } \mathbf{x} \in [0,1]^2; \\ +\infty, & \text{otherwise,} \end{cases}$$

where $0\log(0)$ is assumed to be 0 by continuity. The rate functions for the simplex rate region, the MAC rate region and the circular rate region are shown in Figures 7, 8 and 9, respectively. Note that the polytope nature of the simplex and MAC rate regions results in the discontinuity of the rate functions. In the case of the simplex rate region it is clear that the workloads being equal is much more likely than any other configuration.

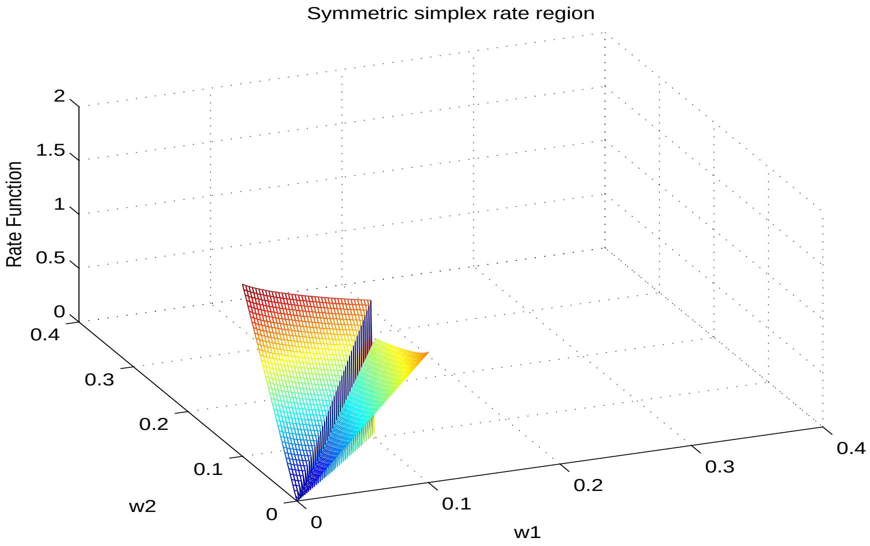

Figure 7: Rate function of the simplex rate region.

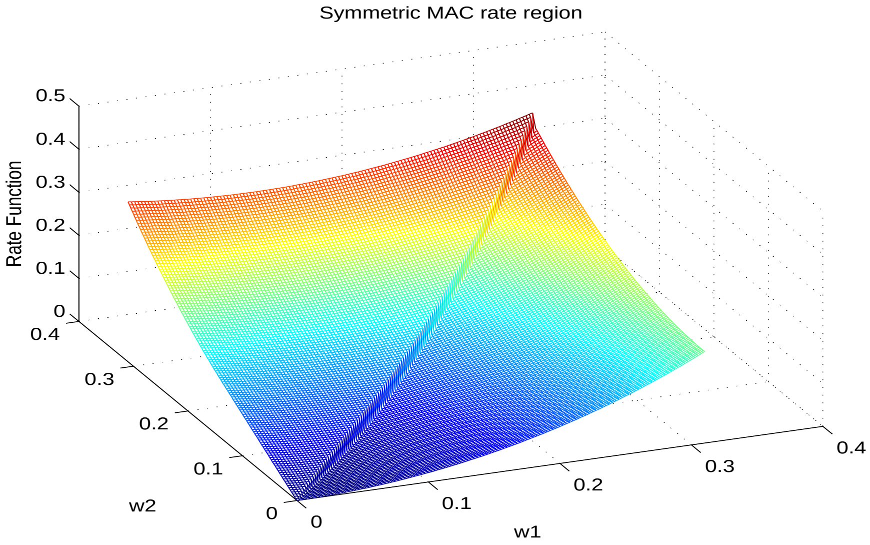

Figure 8: Rate function of the MAC rate region.

We also compare the numerical results with empirical calculations of the negative of the tail asymptote obtained from simulations; all for the case of $\boldsymbol{\beta} = \mathbf{1}$. For this purpose we set $N = 100$ for the computation of the fluid limit. Thereafter, starting from $\mathbf{0}$ we run 10^9 trials in order to estimate the probabilities in (29) for many values of $\mathbf{x}$. Note that the smallest value of the tail asymptote that we can estimate is $\log(10^{-9})/100 \approx -0.2072$. The comparison of the numbers obtained from numerical optimisation and simulations for the quantities in (29) for the simplex rate region, the MAC rate region and the circular rate region are shown in Figures 10, 11 and 12, respectively. For ease of comparison we only present

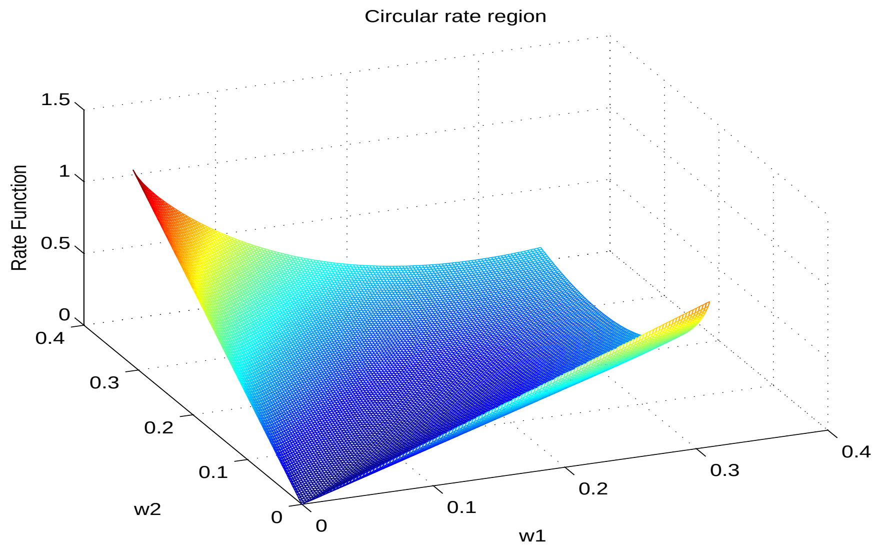

Figure 9: Rate function of the circular rate region.

three cases on the graphs: $w_2 = 0$ which is tagged as $\mathrm{w1}_{\mathrm{emp/calc}}$; $w_1 = 0$ which is tagged as $\mathrm{w2}_{\mathrm{emp/calc}}$; and $w_1 = w_2$ which is tagged as $\mathrm{diag}_{\mathrm{emp/calc}}$. In all cases 'emp' stands for the numbers obtained using simulations while 'calc' stands for the numbers using numerical techniques. In the case of the simplex rate region the numerical results seems to indicate that the preferred way to achieve a large workload is to increase both workloads such that they are equal. However, in the case of the MAC rate-region or the circular rate-region, the preferred way to achieve a large workload only for the first user is different from the preferred way to achieve a large workload only for the second user; the latter seems to follow a trajectory above where both workloads are increased simultaneously in a way such that they remain equal.

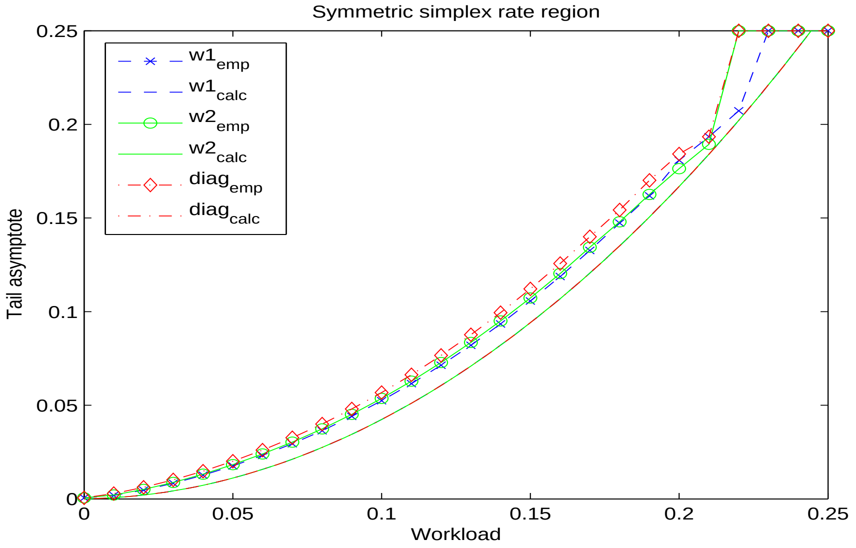

Figure 10: Comparison of numerical results with simulations for the symmetric simplex rate-region.

6. Conclusions In this paper we proved an LDP for Max-Weight scheduling where the server can choose rate-points from within a compact, convex and coordinate-convex set. Since the rate-points chosen by the scheduling policy are closely related to the set of sub-gradients of a convex function that

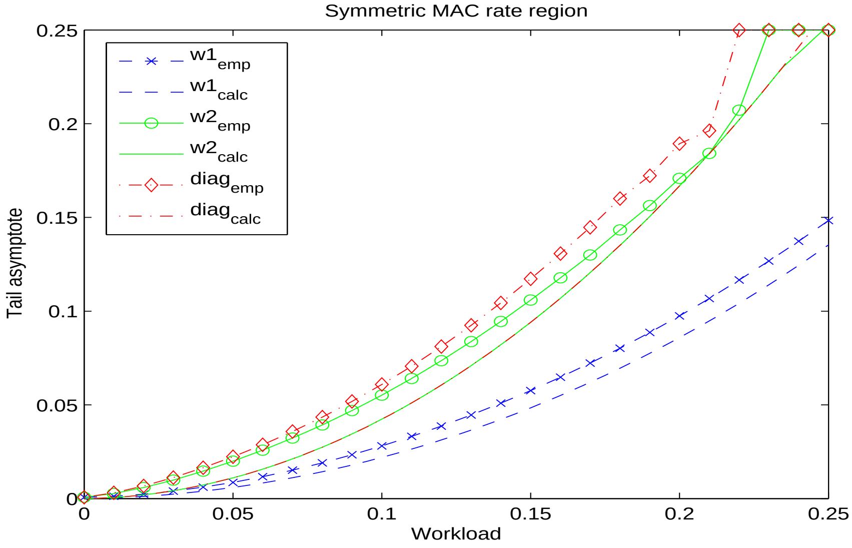

Figure 11: Comparison of numerical results with simulations for the symmetric simplex rate-region.

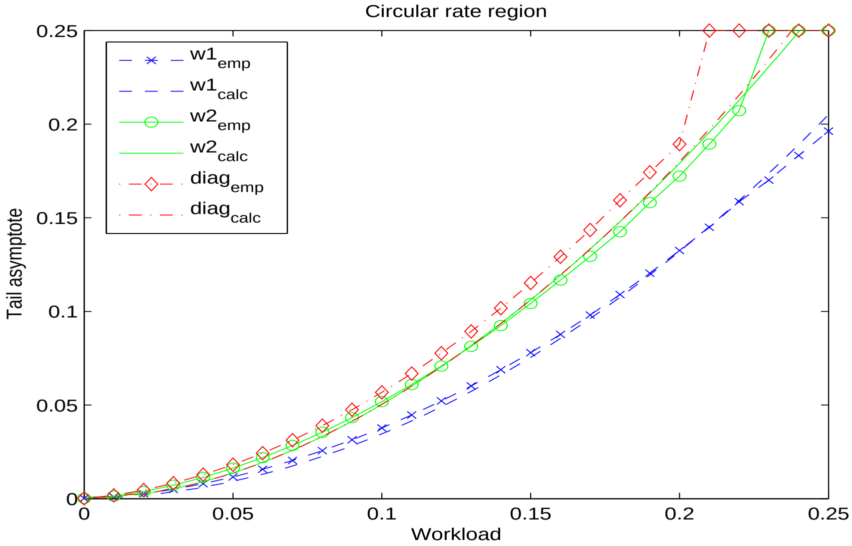

Figure 12: Comparison of numerical results with simulations for the symmetric simplex rate-region.

is constructed using the rate-region, compactness and convexity of the rate-region is sufficient to prove the LDP and identify the rate-function. Coordinate-convexity of the rate-region which is a natural assumption for queueing systems, was then used to simplify the rate-function.

Appendix A. Proofs PROOF OF THEOREM 3.1. From Section 2 we need to prove (3) and (4) to show $\mathbb{C}(\mathbb{X})$-exponential tightness of the net process $\mathbf{X}^N := \left(\mathbf{A}^N, \mathbf{R}^N, \mathbf{Y}^N, \mathbf{S}^N, \Psi^N, \mathbf{W}^N\right)$. Now it is clear using metric $\rho_X(\cdot,\cdot)$ that it suffices to demonstrate both the statements (3) and (4) for $\mathbf{A}^N$, $\mathbf{R}^N$, $\mathbf{Y}^N$, $\mathbf{S}^N$, Ψ^N and $\mathbf{W}^N$ separately so as to prove the result for $\mathbf{X}^N$. Equivalently, the same result follows from the *contraction principle* (see Puhalskii [27, Corollary 3.2.7, pg. 283]).

Since both these properties for $\mathbf{A}^N$ are a consequence of the assumptions on the arrival process there is nothing to prove there.

Proof of exponential tightness (3)
Fix $t \in [0,1]$. From the compactness of $\mathcal{R}(K)$ it is clear that

$$\|\mathbf{S}^N(t)\| \leq K r_{\max} \frac{\lfloor Nt \rfloor + 1}{N} \leq K r_{\max}(t+1) \leq 2K r_{\max},$$

which immediately yields exponential tightness. The same bound can be used to prove exponential tightness of $\mathbf{R}^N$ and $\mathbf{Y}^N$. Similarly, for every $t \in [0,1]$ the exponential tightness of the sequence of measures $\Psi^N(t)$ is a direct consequence of $\mathcal{M}_{t+1}(\mathcal{R}(K))$ being a compact set. Finally we have the following bound

$$\mathbf{W}^N(t) \leq \mathbf{W}^N(0) + \mathbf{A}^N(t),$$

and the exponential tightness of $\mathbf{W}^N$ follows (using the super-exponential convergence of $\frac{\mathbf{W}_0^N}{N}$ to $\mathbf{w}(0)$).

Proof of continuous limits points (4)
Fix $T \in (0,1]$, $\epsilon > 0$, $u \in [0,T)$ and $t \in (u, \min(T, u+\delta)]$. We have

$$\begin{aligned}\|\mathbf{S}^N(t) - \mathbf{S}^N(u)\| = \frac{\|\mathbf{S}^N(\lfloor Nu \rfloor, \lfloor Nt \rfloor]\|}{N} &\leq K r_{\max} \frac{\lfloor Nt \rfloor - \lfloor Nu \rfloor}{N} \\ &\leq K r_{\max}(t - u + \frac{1}{N}) \leq K r_{\max}(\delta + \frac{1}{N}).\end{aligned} \tag{81}$$

Therefore (4) holds for $\mathbf{S}^N$. The very same logic can be used to prove that (4) holds for $\mathbf{R}^N$ and $\mathbf{Y}^N$.

For any Borel set $C \in \mathcal{R}(K)$ we have

$$\begin{aligned}\Psi^N(u)(C) &= \frac{1}{N}\sum_{i=0}^{\lfloor Nu \rfloor} \boldsymbol{\delta}_{\mathbf{r}_i^N}(C) \\ &\leq \frac{1}{N}\sum_{i=0}^{\lfloor Nt \rfloor} \boldsymbol{\delta}_{\mathbf{r}_i^N}(C) = \Psi^N(t)(C) \\ &= \frac{1}{N}\sum_{i=0}^{\lfloor Nu \rfloor} \boldsymbol{\delta}_{\mathbf{r}_i^N}(C) + \frac{1}{N}\sum_{i=\lfloor Nu \rfloor+1}^{\lfloor Nt \rfloor} \boldsymbol{\delta}_{\mathbf{r}_i^N}(C) \\ &\leq \Psi^N(u)(C) + \frac{\lfloor Nt \rfloor - \lfloor Nu \rfloor}{N} \\ &\leq \Psi^N(u)(C) + \delta + \frac{1}{N}.\end{aligned} \tag{82}$$

Using the relationships in (82) and upper bounding the Kantorovich-Wasserstein metric using only bounded continuous functions we can assert (4) for Ψ^N. Since (4) also holds for $\mathbf{A}^N$ and $\mathbf{S}^N$, from (16) and the super-exponential convergence of $\frac{\mathbf{W}_0^N}{N}$ to $\mathbf{w}(0)$ we can assert (4) for $\mathbf{W}^N$ too.

Since we have LD relative compactness there exist limit points $(\mathbf{a}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \mathbf{s}, \Phi, \mathbf{w})$ taking values in $\mathbb{C}(\mathbb{X})$ which are obtained as limits along subsequence $\{N_n\}_{n=1}^{+\infty}$ with $\lim_{n\to+\infty} N_n = +\infty$. We will now prove the required properties of the limit points. Note once again that (i) follows from the assumptions, and there is nothing to prove.

Proof of properties (ii), (iii), (iv), (v) and (vi)

For given $1 \geq t > u \geq 0$ and $\epsilon > 0$ consider the set $C_\epsilon := \{(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1) \in \mathbb{D}(\mathbb{X}) : \|\mathbf{s}_1(t) - \mathbf{s}_1(u)\| \leq Kr_{max}(t-u) + \epsilon\}$. Note that the set C_ϵ is $\mathbb{C}(\mathbb{X})$-closed. Define $O_\epsilon := \mathbb{C}(\mathbb{X}) \setminus C_\epsilon$ which is $\mathbb{C}(\mathbb{X})$-open. Note that we are interested in $\Pi(C)$ and $\Pi(O)$ where $C := \cap_{\epsilon>0} C_\epsilon$ (C_ϵs decrease as $\epsilon \to 0$) and $O := \cup_{\epsilon>0} O_\epsilon$ (O_ϵs increase as $\epsilon \to 0$). Since O_ϵ is $\mathbb{C}(\mathbb{X})$-open, by Puhalskii [27, Corollary 3.1.9, pp. 257–258] we have $\liminf_{n\to+\infty} \mathbb{P}\left(\mathbf{X}^{N_n} \in O_\epsilon\right)^{1/N_n} \geq \Pi(O_\epsilon)$ and by (81) for every $\epsilon > 0$ and for all n such that $N_n > \lfloor \frac{Kr_{\max}}{\epsilon} \rfloor$ we have $\mathbb{P}\left(\mathbf{X}^{N_n} \in O_\epsilon\right) = 0$. Thus, $\Pi(O_\epsilon) = 0$ and therefore $\Pi(O) = \sup_{\epsilon>0} \Pi(O_\epsilon) = 0$. Therefore $\Pi-a.e.$ we have $\mathbf{s}$ also being Lipschitz.

The relationships in (82) are also sufficient to show for $1 \geq t > u \geq 0$ that $\Phi(t) - \Phi(u) \in \mathcal{M}^{t-u}(\mathcal{R}(K))$ and that the total variation norm (see Yosida [42, pg. 35-38, 118-119]) of $\Phi(t) - \Phi(u)$ is $t - u$. Since $\mathcal{M}^1(\mathcal{R}(K))$ is compact we can follow the second part of the proof of Dembo and Zajic [13, Lemma 4, pg. 198-200] to construct $\dot{\Phi}(t) \in \mathcal{M}^1(\mathcal{R}(K))$ for almost every $t \geq 0$. Using ideas similar to those mentioned above we can also show that for every limit point $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and $\mathbf{s}$ are component-wise non-decreasing, and $\boldsymbol{\gamma}(0) = \boldsymbol{\eta}(0) = \mathbf{s}(0) = \mathbf{0}$ and $\Phi(0)(\mathcal{R}(K)) = 0$.

Using the assumptions for $\frac{\mathbf{W}_0^{N_n}}{N_n}$ and results for $\mathbf{S}^{N_n}$ and $\mathbf{A}^{N_n}$ and (16) we now prove results for $\mathbf{W}^{N_n}$. It is clear from the convergence of each term in (16) (and using the $\mathbb{C}(\mathbb{X})$-continuity of addition and subtraction) that every limit point satisfies the following for $k = 1, 2, \ldots, K$ and for all $t \in [0, 1]$

$$w^k(t) = w^k(0) + a^k(t) - s^k(t), \tag{83}$$

for absolutely continuous $a^k(t)$ and where $s^k(t)$ is a limit point of $S^{k,N}(t)$. Note that this is the same as (21). Since $\mathbf{a}$ is absolutely continuous and $\mathbf{s}$ is Lipschitz continuous, we have $\mathbf{w}$ being absolutely continuous for every limit point. The (component-wise) non-negativity of $\mathbf{w}$ is proved using the non-negativity of the original sequence $\mathbf{W}^{N_n}$ just as illustrated above.

Proof of (20)

First we show that $< \Psi^{N_n}(t), e_k >$ LD converges to $< \Phi(t), e_k >$ for every $t \in [0, 1]$. Since $\Pi-a.e.$ the limit points are in $\mathbb{C}(\mathbb{X})$ and since the projection operator $(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1) \to \Phi_1(t)$ is $\mathbb{C}(\mathbb{X})$-continuous (and Borel measurable under the Skorohod J_1 topology) we get LD convergence of $\Psi^{N_n}(t)$ to $\Phi(t)$ in $\mathcal{M}(\mathcal{R}(K))$ by an application of the *contraction principle* (see Puhalskii [27, Corollary 3.1.22, pg. 264]). Since e_k is a (bounded) continuous function from a compact set $\mathcal{R}(K)$ to $\Re_+$, the LD convergence of $< \Psi^{N_n}(t), e_k >$ to $< \Phi(t), e_k >$ follows from Puhalskii [27, Corollary 3.1.9, pp. 257–258].

Since no more than $r_{\max}$ can be served from any user no matter which operating point is chosen we claim that for all $1 \geq t \geq u \geq 0$

$$s^k(t) - s^k(u) \leq \left\langle \Phi(t) - \Phi(u), e_k \right\rangle \quad (\Pi-a.e.). \tag{84}$$

The proof of this follows from the observation that for any subsequence $\left(\mathbf{A}^{N_n}, \mathbf{R}^{N_n}, \mathbf{Y}^{N_n}, \mathbf{S}^{N_n}, \Psi^{N_n}, \mathbf{W}^{N_n}\right)$ that LD converges to $(\mathbf{a}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \mathbf{s}, \Phi, \mathbf{w})$ we have

$$\begin{aligned} 0 = \liminf_{n\to+\infty} \mathbb{P}\left(\left(S^{k,N_n}(t) - S^{k,N_n}(u)\right) - \left\langle \Psi^{N_n}(t) - \Psi^{N_n}(u), e_k \right\rangle > 0\right)^{1/N_n} \\ \geq \Pi\left(\left(s^k(t) - s^k(u)\right) - \left\langle \Phi(t) - \Phi(u), e_k \right\rangle > 0\right), \end{aligned}$$

which implies the result in (84) since

$$\begin{aligned} &\Pi\left(\cup_{1\geq t>u\geq 0} \left\{s^k(t) - s^k(u) - \left\langle \Phi(t) - \Phi(u), e_k \right\rangle > 0\right\}\right) \\ &= \sup_{1\geq t>u\geq 0} \Pi\left(s^k(t) - s^k(u) - \left\langle \Phi(t) - \Phi(u), e_k \right\rangle > 0\right) = 0. \end{aligned}$$

Above we used the fact that $\{(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1) \in \mathbb{D}(\mathbb{X}) : (s_1^k(t) - s_1^k(u)) - \left\langle \Phi_1(t) - \Phi_1(u), e_k \right\rangle > 0\}$ is $\mathbb{C}(\mathbb{X})$-open.

Proof of property (vii)

Without loss of generality it suffices to prove this for $k = 1$. We first rephrase the statement that needs to be proved. Define $\mathbb{C}(\mathbb{X})$-open set $O := \{(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1) \in \mathbb{D}(\mathbb{X}) : \min_{t\in[t_1,t_2]} w^1(t) > 0\}$, $\hat{E}_{t,u} := \{(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1) \in \mathbb{D}(\mathbb{X}) : t \geq u, \ s^1(t) - s^1(u) = \left\langle \Phi(t) - \Phi(u), e_1 \right\rangle\}$, and $\tilde{E} := \{(\mathbf{a}_1, \boldsymbol{\gamma}_1, \boldsymbol{\eta}_1, \mathbf{s}_1, \Phi_1, \mathbf{w}_1) \in \mathbb{D}(\mathbb{X}) : s^1(t) - s^1(u) = \left\langle \Phi(t) - \Phi(u), e_1 \right\rangle, \ \forall t \geq u, \ t, u \in [t_1, t_2]\} =$

$\cap_{t\geq u,\, t,u\in[t_1,t_2]}\hat{E}_{t,u}$. Now $\hat{E}_{t_2,t_1}\subseteq\hat{E}_{t,u}$ for all $t\geq u$ with $t,u\in[t_1,t_2]$ from which it follows that $\tilde{E}=\hat{E}_{t_2,t_1}$. Also define $\tilde{E}^c$ and $\hat{E}^c_{t,u}$ to be the complement of $\tilde{E}$ and $\hat{E}_{t,u}$ in $\mathbb{D}(\mathbb{X})$, respectively. Then we need to prove that $\Pi(O\cap\tilde{E}^c)=0$, in other words, $\Pi(O\cap\hat{E}^c_{t_2,t_1})=0$ holds. For ease of notation from now onwards we set $E:=\hat{E}_{t_2,t_1}$ with E^c being the complement. For future use we note that E is a $\mathbb{C}(\mathbb{X})$-closed set.

For $\epsilon>0$ define (increasing in ϵ as $\epsilon\to 0$) $\mathbb{C}(\mathbb{X})$-open sets $O_\epsilon:=\{(\mathbf{a}_1,\boldsymbol{\gamma}_1,\boldsymbol{\eta}_1,\mathbf{s}_1,\Phi_1,\mathfrak{w}_1)\in\mathbb{D}(\mathbb{X}):\min_{t\in[t_1,t_2]}w^1(t)>\epsilon\}$; it is clear that $O:=\cup_{\epsilon>0}O_\epsilon$. Therefore we have $\Pi(O)=\sup_{\epsilon>0}\Pi(O_\epsilon)$ and $\Pi(O\cap E^c)=\sup_{\epsilon>0}\Pi(O_\epsilon\cap E^c)$. Therefore we will prove that $\Pi(O_\epsilon\cap E^c)=0$ for all $\epsilon>0$. Also define sets $O_{(n),\epsilon}:=\left\{\min_{t\in[t_1,t_2]}\tilde{W}^{1,N_n}(t)>\epsilon\right\}\subseteq\mathbb{D}(\mathbb{X})$ and $E_n:=\left\{\tilde{S}^{1,N_n}(t_2)-\tilde{S}^{1,N_n}(t_1)=\left\langle\Psi^{N_n}(t_2)-\Psi^{N_n}(t_1),e_1\right\rangle\right\}$ (with $E_n\subseteq\mathbb{D}(\mathbb{X})$). Now $E_n\supseteq\left\{\min_{t\in[t_1,t_2]}\tilde{W}^{k,N_n}(t)>\frac{r_{\max}}{N_n}\right\}$ since any service will be at the allocated rate if there is sufficient work to be done. If $N_n\geq\lfloor\frac{r_{\max}}{\epsilon}\rfloor+1$, then $E_n\supseteq O_{(n),\epsilon}$. Let E^c_n be the complement of E_n with respect to $\mathbb{D}(\mathbb{X})$. Therefore, for n large enough (such that $N_n\geq\lfloor\frac{r_{\max}}{\epsilon}\rfloor+1$) we get $E^c_n\cap O_{(n),\epsilon}=\emptyset$, and therefore $\mathbb{P}(E^c_n\cap O_{(n),\epsilon})=0$. The result follows since we can derive the following

$$\Pi(O_\epsilon\cap E^c)\leq\liminf_{n\to+\infty}\mathbb{P}(O_{(n),\epsilon}\cap E^c_n)^{1/N_n}=0\qquad\qquad(O_\epsilon\cap E^c\text{ is }\mathbb{C}(\mathbb{X})\text{-open}).$$

Proof of property (viii)

If $\mathbf{w}(t)=\mathbf{0}$, then the result trivially holds since $\tilde{H}(\mathbf{0})=\mathcal{R}(K)$.

Following Billingsley [4, pg. 8] for $\epsilon>0$ define functions f^ϵ,g^ϵ from $\mathcal{R}(K)\times\Re^K_+$ to $[0,1]$ as follows

$$f^\epsilon(\tilde{\mathbf{r}},\mathbf{x})=\left(1-\frac{\max_{\mathbf{r}\in\mathcal{R}(K)}\left\langle\mathbf{r},\boldsymbol{\beta}\circ\mathbf{x}\right\rangle-\left\langle\tilde{\mathbf{r}},\boldsymbol{\beta}\circ\mathbf{x}\right\rangle}{\epsilon}\right)_+,\tag{85}$$

and $g^\epsilon(\tilde{\mathbf{r}},\mathbf{x})=1-f^\epsilon(\tilde{\mathbf{r}},\mathbf{x})$. From the definition it is clear that $f^\epsilon(\tilde{\mathbf{r}},\mathbf{x})=1$ if and only if $\tilde{\mathbf{r}}\in\tilde{H}(\mathbf{x})$, and $f^\epsilon(\tilde{\mathbf{r}},\mathbf{x})=0$ if and only if $\tilde{\mathbf{r}}\in F^\epsilon:=\{\mathbf{r}\in\mathcal{R}(K):\left\langle\mathbf{r},\boldsymbol{\beta}\circ\mathbf{x}\right\rangle\leq\max_{\hat{\mathbf{r}}\in\mathcal{R}(K)}\left\langle\hat{\mathbf{r}},\boldsymbol{\beta}\circ\mathbf{x}\right\rangle-\epsilon\}$. As ϵ decreases to 0, F^ϵ increases to (open) $\mathcal{R}(K)\setminus\tilde{H}(\mathbf{x})$, and $g^\epsilon(\tilde{\mathbf{r}},\mathbf{x})$ converges to $1_{\{\mathcal{R}(K)\setminus\tilde{H}(\mathbf{x})\}}(\tilde{\mathbf{r}})$ yielding the indicator function of $\mathcal{R}(K)\setminus\tilde{H}(\mathbf{x})$. Both $f^\epsilon(\cdot,\cdot)$ and $g^\epsilon(\cdot,\cdot)$ are continuous functions. In fact, it is easy to prove that

$$|f^\epsilon(\tilde{\mathbf{r}}_1,\mathbf{x}_1)-f^\epsilon(\tilde{\mathbf{r}}_2,\mathbf{x}_2)|\leq\frac{2Kr_{\max}}{\epsilon}\|\boldsymbol{\beta}\circ(\mathbf{x}_1-\mathbf{x}_2)\|+\frac{\left\|\boldsymbol{\beta}\circ(\frac{\mathbf{x}_1+\mathbf{x}_2}{2})\right\|}{\epsilon}\|\tilde{\mathbf{r}}_1-\tilde{\mathbf{r}}_2\|.$$

Given $1\geq t>u\geq 0$ for element $(\mathbf{a},\boldsymbol{\gamma},\boldsymbol{\eta},\mathbf{s},\Phi,\mathbf{w})\in\mathbb{D}(\mathbb{X})$ define the following functional

$$h^\epsilon_{u,t}(\mathbf{w}):=\int_u^t\int_{\mathcal{R}(K)}g^\epsilon(\tilde{\mathbf{r}},\mathbf{w}_v)d\Phi(v)(\tilde{\mathbf{r}}).\tag{86}$$

For every $\epsilon>0$ we have $\int_{\mathcal{R}(K)}g^\epsilon(\tilde{\mathbf{r}},\mathbf{x})d\nu(\tilde{\mathbf{r}})$ being a bounded and continuous function on $(\mathbf{x},\nu)\in\Re^K_+\times\mathcal{M}(\mathcal{R}(K))$ (and also on $\mathbb{X}$), and hence by Ethier and Kurtz [17, Problem 13, pg. 151] and continuity of integrals (see Whitt [40, Theorem 11.5.1, pg. 383]), $h^\epsilon_{u,t}(\mathbf{w})$ is a bounded and continuous function on $\mathbb{D}(\mathbb{X})$. Thus, we can directly apply the definition of LD convergence (see Puhalskii [27, Definition 3.1.1, pp. 253–254]) by going back to the LD converging sub-sequence. Using the property that our scheduling policy enforces $\mathbf{r}_m\in\tilde{H}(\mathbf{W}_m)$ for all $m\geq 0$, it is easy to argue that $\Pi-a.e.$ we have $h^\epsilon_{u,t}(\mathbf{w})=0$ for all $\epsilon>0$.

If we further assume the absolute continuity and Lipschitz continuity assumptions from the already established *properties (i), (ii), (iii), (iv), (v)* and *(vi)*, then using Fubini's Theorem we can write the following alternate expression for $h^\epsilon_{u,t}(\mathbf{w})$, namely,

$$h^\epsilon_{u,t}(\mathbf{w})=\int_u^t\left(\int_{\mathcal{R}(K)}g^\epsilon(\tilde{\mathbf{r}},\mathbf{w}_v)d\dot{\Phi}(v)(\tilde{\mathbf{r}})\right)dv.\tag{87}$$

Under exactly these assumptions we can further prove (by the bounded convergence theorem) that $h^\epsilon_{u,t}(\mathbf{w})$ converges to $h_{u,t}(\mathbf{w})=\int_u^t\dot{\Phi}(v)\left(\mathcal{R}(K)\setminus\tilde{H}(\mathbf{w}_v)\right)dv$ as $\epsilon\downarrow 0$. Now $h^\epsilon_{u,t}(\mathbf{w})=0$ for all limit points, implying that $h_{u,t}(\mathbf{w}) = 0$. Therefore, we have for (Lebesgue) almost all $t \in [0,1]$ the result that $\dot{\Phi}(t)\Big(\mathcal{R}(K) \setminus \tilde{H}(\mathbf{w}_t)\Big) = 0$. Note that this is another way of saying that for every limit point $\boldsymbol{\gamma}$ it holds that $\dot{\boldsymbol{\gamma}}(t) \in \tilde{H}(\mathbf{w}(t))$ for (Lebesgue) almost all $t \in [0,1]$.

Proof of property (ix)

First we fix $k \in \{1,2,\ldots,K\}$. From (17) we can write $\tilde{Y}^{k,N}(t)$ as follows

$$\tilde{Y}^{k,N}(t) = \max\Bigg(0, \sup_{0\le s\le t}\Big(\tilde{R}^{k,N}(s) - \tilde{A}^{k,N}(s-1/N)\Big) - \tilde{W}^{k,N}(0)\Bigg).$$

Since $\{\tilde{A}^{k,N}(s-1/N)\}_{N=1}^{\infty}$ is exponentially equivalent (see Dembo and Zeitouni [12, Defn. 4.2.10 & Theorem 4.2.13, pg. 130]) to $\{\tilde{A}^{k,N}(s)\}_{N=1}^{\infty}$ which satisfies an LDP with a good rate function and since function $\hat{\eta}(t) := \max\Big(0, \sup_{0\le s\le t}\Big(\hat{\gamma}(s) - \hat{a}(s)\Big) - \hat{w}(0)\Big)$ is Borel measurable under the Skorohod J_1 topology (including the subtraction operation) and continuous under the local uniform topology (see Billingsley [4], Ethier and Kurtz [17] or Whitt [40, Chapter 13]) on $\mathbb{D}(\mathbb{X})$ (and hence, $\mathbb{C}(\mathbb{X})$-continuous), by an invocation of the *contraction principle* (see Puhalskii [27, Corollary 3.1.22, pg. 264]) we get that every limit point should satisfy the following

$$\begin{aligned}\eta^k(t) &= \max\Bigg(0, \sup_{0\le s\le t}\Big(\gamma^k(s) - a^k(s)\Big) - w^k(0)\Bigg)\\ &= \max\Bigg(w^k(0), \sup_{0\le s\le t}\Big(\gamma^k(s) - a^k(s)\Big)\Bigg) - w^k(0).\end{aligned}$$

Since $w^k(t) = w^k(0) + a^k(t) - \gamma^k(t) + \eta^k(t)$ we also obtain

$$w^k(t) = \max\Bigg(a^k(t) - \gamma^k(t) + w^k(0), a^k(t) - \gamma^k(t) - \inf_{0\le s\le t}\Big(a^k(s) - \gamma^k(s)\Big)\Bigg)$$

Now it is clear that $\eta^k(t)$ is non-negative, non-decreasing such that

$$\dot{\eta}^k(t) = \begin{cases}\Big(\dot{\gamma}^k(t) - \dot{a}^k(t)\Big)_+ & \text{if } \gamma^k(t) - a^k(t) = \sup_{0\le s\le t}\Big(\gamma^k(s) - a^k(s)\Big) \ \&\ \gamma^k(t) - a^k(t) \ge w^k(0);\\ 0 & \text{otherwise.}\end{cases}$$

Note that the condition for non-trivial derivative of $\eta^k(t)$ is equivalent to $w^k(t) = 0$; in other words, $\eta^k(t)$ can only increase when $w^k(t) = 0$. Also note that $\gamma^k(t) - a^k(t) = \sup_{0\le s\le t}\Big(\gamma^k(s) - a^k(s)\Big)$ directly implies that $\dot{\gamma}^k(t) \ge \dot{a}^k(t)$ but nevertheless we choose to emphasise the positive part in the formula above.

Thus every limit point is an absolutely continuous solution of the following differential inclusion for all $t \in [0,1]$:

$$\dot{\mathbf{w}}(t) - \dot{\boldsymbol{\eta}}(t) \in \dot{\mathbf{a}}(t) - \tilde{H}(\mathbf{w}(t))$$

with $\mathbf{w}(0)$ the initial condition such that $\dot{\boldsymbol{\eta}}(t) \ge \mathbf{0}$ and $\dot{\eta}^k(t)w^k(t) = 0$ for all $k = 1,2,\ldots,K$. □

Acknowledgments. This work was supported by Science Foundation of Ireland grants 03/IN3/I396 and 07/IN.1/I901. A preliminary version of this work was presented at the 2008 Information Theory and Applications Workshop at the University of California, San Diego, Jan 27th-Feb 1st, 2008.

The author would like to express his gratitude to Prof. Anatolli Puhalskii for his detailed comments and his suggestions to improve the presentation of a previous version of this paper and for his encouragement. The author is indebted to Dr. Ken Duffy for detailed suggestions and comments on a previous version of the paper that were critical in expanding its scope and improving its presentation. The author would also like to thank the anonymous reviewers and editor for pointing out errors, omissions and stylistic faults in an earlier version of this paper. Thanks are also due to Dr. Somsak Kittipiyakul for a detailed reading of the paper and for pointing out important typographical errors. Finally, many thanks are owed to Dr. David Malone for help with the simulations, and Dr. Helena Šmigoc and Prof. Barak Pearlmutter for aiding the author in obtaining a copy of Brézis[7].

References

[1] M. Andrews, K. K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting. Scheduling in a queueing system with asynchronously varying service rates. *Probability in Engineering and Informational Sciences*, 18(2):191–217, 2004.

[2] R. Atar, A. Budhiraja, and K. Ramanan. Deterministic and stochastic differential inclusions with multiple surfaces of discontinuity. *Probab. Theory Related Fields*, January 2008. Published online.

[3] D. Bertsimas, I.Ch. Paschalidis, and J.N. Tsitsiklis. Asymptotic buffer overflow probabilities in multi-class multiplexers: An optimal control approach. *IEEE Trans. on Automatic Control*, 43:315–335, March 1998.

[4] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[5] A. A. Borovkov. Boundary value problems for random walks and large deviations in function spaces. *Theory Probab. Appl.*, 12(4):635–654, 1967.

[6] D. D. Botvich and N. G. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems Theory Appl.*, 20(3-4):293–320, 1995.

[7] H. Brézis. *Opérateurs maximaux monotones et semigroups de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[8] Felix E. Browder. Nonlinear operators and nonlinear equations of evolution in Banach spaces. In *Nonlinear functional analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 2, Chicago, Ill., 1968)*, pages 1–308. Amer. Math. Soc., Providence, R.I., 1976.

[9] Emmanuel Cépa. Problème de Skorohod multivoque. *Ann. Probab.*, 26(2):500–532, 1998.

[10] H. Chen and D. D. Yao. *Fundamentals of Queueing Networks*. Springer, New York, 2001.

[11] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[12] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications, 2nd Ed.* Springer, 1998.

[13] Amir Dembo and Tim Zajic. Large deviations: From empirical mean and measure to partial sums process. *Stochastic Process. Appl.*, 57(2):191–224, 1995.

[14] J. L. Doob. *Measure theory*, volume 143 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.

[15] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002.

[16] Paul Dupuis, Kevin Leder, and Hui Wang. On the large deviations properties of the weighted-serve-the-longest-queue policy. In *In and out of equilibrium. 2*, volume 60 of *Progr. Probab.*, pages 229–256. Birkhäuser, Basel, 2008.

[17] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[18] Jin Feng and Thomas G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[19] S. Foss and T. Konstantopoulos. An overview of some stochastic stability methods. *Journal of the Operations Research Society of Japan*, 47(4):275–303, 2004.

[20] J. Garcia. An extension of the contraction principle. *J. Theoret. Probab.*, 17(2):403–434, 2004.

[21] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[22] S. Kittipiyakul and T. Javidi. Optimal operating point for MIMO multiple access channel with bursty traffic. *IEEE Transactions on Wireless Communication*, 6:4464–4474, Dec 2007.

[23] Constantinos Maglaras and Jan A. Van Mieghem. Queueing systems with leadtime constraints: A fluid-model approach for admission and sequencing control. *European J. Oper. Res.*, 167(1):179–207, 2005.

[24] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand. Achieving 100% throughput in an input-queued switch. *IEEE Trans. on Comm.*, 47(8):1260–1267, Aug. 1999.

[25] A. A. Mogulʹskiĭ. Large deviations for the trajectories of multidimensional random walks. *Theory Probab. Appl.*, 21(2):309–323, 1976.

[26] A. A. Mogulʹskiĭ. Large deviations for processes with independent increments. *Ann. Probab.*, 21(1):202–215, 1993.

[27] A. A. Puhalskii. *Large deviations and idempotent probability*, volume 119 of *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics*. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[28] A. A. Puhalskii. Large deviations for stochastic processes. Notes for LMS/EPSRC Short Course: Stochastic Stability, Large Deviations and Coupling Methods, Heriot-Watt University, Edinburgh, Sept. 2006.

[29] A. A. Puhalskii. The action functional for the Jackson network. *Markov Processes and Related Fields*, 13(1):99–136, 2007.

[30] A. A. Puhalskii and A. A. Vladimirov. A large deviation principle for Join the Shortest Queue. *Mathematics of Operations Research*, 32:700–710, 2007.

[31] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[32] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

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

[34] A. L. Stolyar and K. Ramanan. Largest weighted delay first scheduling: Large deviations and optimality. *The Annals of Applied Probability*, 11(1):1–48, 2001.

[35] Alexander L. Stolyar. Control of end-to-end delay tails in a multiclass network: LWDF discipline optimality. *Ann. Appl. Probab.*, 13(3):1151–1206, 2003.

[36] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37(12):1936–1948, 1992.

[37] D. N. C. Tse, P. Viswanath, and L. Zheng. Diversity-multiplexing tradeoff in multiple-access channels. *IEEE Trans. Information Theory*, 50(9):1859–1874, 2004.

[38] David N. C. Tse and Stephen V. Hanly. Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. Inform. Theory*, 44(7):2796–2815, 1998.

[39] V. J. Venkataramanan and X. Lin. On wireless scheduling algorithms for minimizing the queue-overflow probability. *IEEE/ACM Transactions on Networking*, 2010. Accepted.
`http://cobweb.ecn.purdue.edu/~linx/paper/ton08-ldp.pdf`.

[40] Ward Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.

[41] Damon Wischik. Tutorial on queueing theory for switched networks. ICMS workshop for young researchers, on stochastic processes in communication networks, Edinburgh, UK, June 2010.
`http://www.cs.ucl.ac.uk/staff/d.wischik/Talks/netsched.html`.

[42] Kōsaku Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, reprint of the sixth (1980) edition, 1995.