Dynamic Allocation of Reusable Resources: Logarithmic Regret in Overloaded Networks

Xinchang Xie and Itai Gurvich
Northwestern University, Kellogg School of Management

Simge Küçükyavuz
Northwestern University, Industrial Engineering and Management Sciences

We study the problem of dynamically allocating reusable resources to customers of n types. There are d pools of resources and a finite number of units from each resource. If a customer request is accepted, the decision maker collects a type-dependent reward and the customer occupies, for a random service time, one unit from each resource in a *set* of these. Upon service completion, these resource units become available for future allocation. This is a loss network: requests that are not accepted leave immediately. The decision maker's objective is to maximize the long-run average reward subject to the resource-capacity constraint.

A natural linear programming (LP) relaxation of the problem serves as an upper bound on the performance of any policy. We identify a condition that generalizes the notion of overload in single-resource networks (i.e., when $d = 1$). The LP guides our construction of a threshold policy. In this policy, the number of thresholds equals the number of resource types (hence does not depend on the number of customer types). These thresholds are applied to a "corrected" headcount process. In the case of a single resource the corrected headcount is the same as headcount: the number of resource units that are occupied. We prove that in overloaded networks, the additive loss (or regret) of this policy, benchmarked against the LP upper bound, is logarithmic in the total arrival volume in the many-customer many-resource-units asymptotic regime. No policy can achieve sub-logarithmic regret. Simulations showcase the performance of the proposed policy.

Key words: sequential resource allocation, regret, linear programming relaxation, loss networks, Lyapunov function

History: This version: November 3, 2023.

1. Introduction

In dynamic resource allocation problems, a decision maker (DM) allocates a finite number of resource units to sequentially arriving customers in order to maximize the revenue collected from accepted customer requests. Customers leave immediately if their request is not accepted.

Dynamic Stochastic Knapsack is a family of resource allocation problems whereby, once a unit of resource is consumed, it cannot be allocated again in the future. These have applications in revenue management where, for example, the resources are seats on a specific flight.

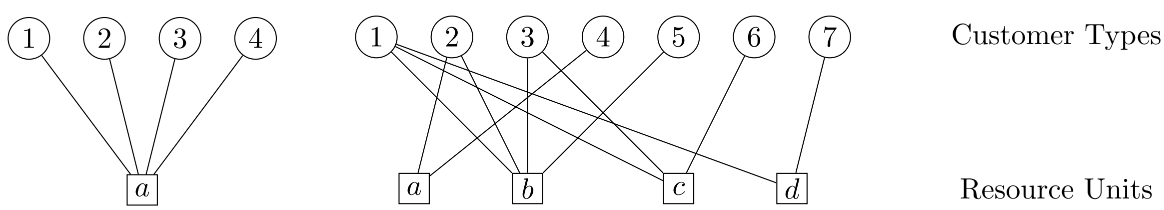
In this paper, we focus on the case where resources are *reusable*. They are used by an accepted request/customer for a (random) duration and are returned to the DM at the conclusion of service (or processing). Rental services (hotels, cars, or cloud computing units) are often modeled as networks of reusable resources. In its fullest (network) generality, our model is standard in the study of telecommunication networks; see Hui (2012). In these networks, a communication (or a call) requires, for its duration, the simultaneous occupation of multiple connected links leading from a source node to a destination node; the links are the resources and the collection required by a call is its route. If one of the links in the route is fully occupied, the call is lost.

Our study was motivated originally by problems of prioritizing the allocation of multiple resources for military uses (Gurvich and Intelligent Automation, 2021). Such simultaneous occupation of multiple resources is also relevant in consulting or IT services. Here, the types of resources would correspond to the different professionals required to deliver the product. A service engagement might require, for example, three database architects, five Java programmers, a network specialist, and so on. These resources must be allocated to the client for the duration of the engagement; see Hu et al. (2010); Cao et al. (2011) for a detailed description of models used for an IBM line of IT services.

The fundamental tradeoff is the same in these applications: accepting a request—and occupying resources—may prevent the acceptance of a later request that requires an overlapping set of resources; one might want to reserve some capacity for highly valuable requests.

In the simplest instance of these problems (see Figure 1(LEFT)) there is a single resource with q units (say q hotel rooms) serving multiple types of customers. Type-i customers arrive according to a Poisson(λ_i) process and request a single unit of the resource. If a type-i request is accepted, the DM collects a reward $r_i > 0$, and a unit of the resource is occupied for an Exp(μ_i) service time after which the resource unit becomes available for future allocations. The objective of the DM is to maximize the expected long-run average reward, subject to the constraint that no more than the q units of the resource can be occupied at any given time.

Figure 1 Two examples of networks: (LEFT) a network with a single resource, and (RIGHT) a network with multiple resources where each type of customer requires the simultaneous possession of multiple resources.



With homogeneous service-times ($\mu_i \equiv \mu$), the optimal policy is a *trunk reservation policy* (Miller, 1969; Lippman, 1975), wherein an arriving type-i customer is accepted if and only if there are more than $R_i \in \mathbb{N} \cup \{\infty\}$ units of resource available; $R_i = 1$ means that type-i customers are accepted whenever there is a resource unit available and $R_i = \infty$ means that these customers' requests are always rejected, regardless of the number of available resource units. Some customer types might have finite thresholds that are strictly greater than 1, these thresholds satisfy a natural monotonicity: the higher the reward that type i brings, the lower the acceptance threshold.

With heterogeneous service times ($\mu_i \neq \mu_j$), the optimal decision is complex and generally depends on the number of customers of each type present in the system, rather than the headcount which tracks only the total number of occupied resource units (Örmeci et al., 2001).

A many-server high-volume asymptotic framework ($\lambda_i^N = N\lambda_i, q^N = Nq$) exposes characteristics of “good” policies. The simplicity of good policies depends on the regime: underloaded, critically loaded, or overloaded; see §1.1. A single-resource corollary of our general result is that, in the overloaded regime, the appealingly simple trunk reservation policy, with a single threshold, has a gap from a deterministic upper bound that is logarithmic in N. This means that no policy, including the possibly complex optimal policy, can improve on this simple prescription by more than $\log N$. In fact, as we will prove, such logarithmic regret is the best one can hope for: no policy can get closer than $\log N$ to the deterministic upper bound; in the overloaded regime, $\log N$ captures precisely the cost of stochasticity.

In the single resource case, regardless of the number of types n, our policy would use a single threshold applied to the class with the lowest value $r_i\mu_i$. We are interested in networks where different request/customer types consume different (subsets of) resources; see Figure 1(RIGHT).

The most intuitive generalization of trunk reservation to d resources suggests assigning # types $\times$ # resources thresholds R_{ij} so that a type-i customer is accepted only if there are more than R_{ij} available units of resource j. If the network is overloaded, we will show, a policy with only d (=# resources) thresholds achieves a logarithmic regret. In Figure 1(RIGHT), requests of type 3 require a unit of resource b and a unit of resource c. Under our policy, a threshold is applied to either b or c but not both: we accept a type-3 request whenever there are units available of resource b (without any requirement on the number of such units) but demand that there are at least an order of $\log N$ “nominal” units available of resource c. *Nominal* here refers to the fact that, instead of applying the threshold directly to the number of busy (or available) units of a resource—the resource’s headcount—we apply it to a *corrected headcount*.

Our proposed policy produces a regret that is logarithmic in N; the formal statement appears in Theorem 4.1 after we introduce the key building blocks of our policy. The key ingredients are:

1. We identify a network version of overload. We establish properties of the linear programming (LP) relaxation of the problem under the network-overload condition. Overload guarantees the existence of a perfect matching between “less-preferred” customer types—as identified by the LP solution—and resources; this matching produces pairs of resources and customer types: for each resource j, there is a single request type i_j coupled with it.

   We show that when this overload condition is violated, the gap of the optimal online policy from the deterministic LP upper bound generally scales like $\sqrt{N}$.

2. We introduce a policy based on *corrected headcount* processes, which are proxies for the true headcounts. The thresholds are applied to these proxies. Informally speaking, the threshold policy based on the corrected headcount guides the allocation to its desired levels as dictated by the LP benchmark.
3. We prove that in overloaded networks, the regret of our policy—the additive reward loss (or approximation gap) of our policy relative to the LP solution—is at most logarithmic in the scaling factor N.
4. We show that this is the best possible: no other policy can get closer than $\log N$ to the LP upper bound. We also prove that state-dependency of the policy is necessary: static policies (e.g., randomized acceptance) induce a larger approximation gap.

1.1. Related Literature

Our work has natural connections to two overlapping streams of literature: queueing theory (loss networks) and revenue management (networks of reusable resources).

Loss Networks. In the queueing literature, loss networks are a class of networks with no buffers. If not accepted, a customer leaves the system immediately instead of being put on hold in a queue. Customers of different types arrive sequentially and request a set of resource units (servers) for a random duration (service time).

The study of loss networks was originally motivated by telecommunication networks and the early focus has been on analyzing the blocking probabilities under specific practical control policies. *Complete sharing*, for example, is a policy that accepts all arriving requests whenever feasible. It induces a stationary distribution that has a so-called product-form structure which renders the blocking-probability calculations tractable; see Kelly (1986) and the comprehensive review Kelly (1991). See also Jung et al. (2019) and the references therein for more recent progress in this area.

The complete sharing policy is not optimal for reward maximization when customers are heterogeneous in the rewards they bring and/or the resources they consume. In the case of a single resource and homogeneous service times, (Miller, 1969; Lippman, 1975) proved that the optimal policy is a *trunk reservation* policy that reserves resource units for the more valuable customers and rejects a less valuable customer if the number of available resource units upon arrival is below the trunk reservation level for that customer. Key (1990) and Reiman (1991) show that the optimal trunk reservation level is logarithmic in the number of servers (resource units).

In the heterogeneous variant of the single-resource model where either different customers request different numbers of resource units, or customers have different service rates, the optimal policy is generally not of a trunk reservation type (Ross and Tsang, 1989; Örmeci et al., 2001; Örmeci and van der Wal, 2006).

Trunk reservation policies are practically appealing and, while not optimal, deliver asymptotically optimal performance in certain conditions. A well-studied limiting regime is one where service rates are fixed while both the arrival rates and the quantities of resource units scale up linearly at the same rate. Puhalskii and Reiman (1998) consider a critically-loaded regime with a single resource type and multiple types of customers with different (i.e., heterogeneous) service times. They prove that the trunk reservation policy is asymptotically optimal (in a central-limit theorem sense) under the requirement that the most valuable requests are also the ones with the longest service time.

Hunt and Kurtz (1994); Bean et al. (1997) prove that, under a large family of policies (that includes trunk reservation), the number of customers in service—scaled in a strong-law scaling—converges to the solution of an integral equation. Subsequent work by Bean et al. (1995) and Hunt and Laws (1997) prove that—with a single resource—a trunk reservation policy is asymptotically optimal in fluid scale (i.e., the optimality gap is $o(N)$ where N is the scaling factor).

Iyengar and Sigman (2004) consider a problem that is more general than the one we study here. The DM not only decides whether to accept a customer but also determines the resource allocation (out of a type-dependent set of such). They devise a control policy that is asymptotically optimal in the approximation ratio sense in the single resource case ($d = 1$). For $d > 1$ they establish a lower bound on the approximation ratio. For our more restricted model, we obtain a logarithmic *additive* gap, a stronger notion that implies, in particular, asymptotic optimality in the approximation-ratio sense.

Pricing, when adjustable, adds a control lever. Paschalidis and Tsitsiklis (2000) study the pricing problem in the case of a single resource and derive structural properties of the optimal policy. They prove that static pricing is asymptotically optimal, in the fluid scaling sense, in the many-server many-customer regime. Paschalidis and Liu (2002) extend this result to the network case of multiple types of resource units (servers) with fixed routing, and show that a static pricing policy remains asymptotically optimal. In our case, prices (and rewards) are fixed.

Revenue Management with Reusable Resources. Our work is closely related to the canonical quantity-based (admission control) network revenue management (NRM) problems; see, for example, Williamson (1992); Gallego and van Ryzin (1997); Reiman and Wang (2008); Jasin and Kumar (2012); Bumpensanti and Wang (2020), as well as the book Talluri and van Ryzin (2004). Motivated by airline revenue management problems, much of this work considers resource units that are either perishable or are allocated at the end of the decision horizon.

The revenue-management literature on reusable resources is more recent and, hence, relatively scarce. Levi and Radovanović (2010) consider the allocation of a single pool of reusable resources to multiple types of customers and devise a *class selection policy* based on the linear programming relaxation of the problem, with guarantees on its approximation ratio. Chen et al. (2017) consider a variant where the customers request the resource units in advance with deterministic starting time and duration. A modified class selection policy, they show, achieves asymptotic optimality in the approximation ratio sense.

Regulation of arrivals through pricing is considered by Xu and Li (2013) who study pricing in a cloud computing platform and characterize structural properties of the optimal policy. Lei and Jasin (2020) study the pricing problem when customers have deterministic advance reservation times and service times. They develop a policy that is based on a deterministic relaxation of the problem and prove an upper bound on its regret. Besbes et al. (2021) consider a pricing problem for a combined objective of revenue, market share, and service level. They prove that a static pricing policy can simultaneously achieve 80% of all three metrics relative to the optimal dynamic programming policy. Jia et al. (2022) tackle the problem of pricing when some of the parameters must be learned.

Recent literature also considers the case where customers make choices and the DM can optimize (dynamically) the assortment an arriving customer sees. Each arriving customer is shown a set of different products and chooses one of them. Owen and Simchi-Levi (2018) devise pricing and assortment optimization policies with provable approximation ratio lower bounds. Rusmevichientong et al. (2020) take a different approach to produce a policy with a provable approximation ratio. Baek and Ma (2019) generalize this guarantee to more general settings in which each product in an assortment consists of multiple types of resource units. Policies for dynamic assortment optimization with reusable resources are developed in Feng et al. (2019); Goyal et al. (2020); Gong et al. (2021) together with approximation ratio guarantees.

In our setting, the assortment per customer type is fixed, but customers do not necessarily consume a single resource unit (a single selection from the assortment); they might consume multiple resource units of different types. Our notion of optimality is stronger than (and implies) approximation-ratio optimality. The policy we propose achieves the optimal approximation gap scaling relative to the linear programming benchmark. The price we "pay" for this optimality is: (i) the knowledge of the arrival rates (at least up to a small perturbation), (ii) a *network overload* condition that is a natural generalization of the single-resource notion. The presence of overload depends both on the graph and the other primitives (arrival rates, service rate, and rewards).

Organization of the Paper

The rest of the paper is organized as follows. Our model is described in Section 2, where we also introduce the deterministic LP relaxation of the problem. In Section 3, we introduce the network-overload condition and study its implications. We construct our corrected-headcount threshold policy and state the main optimality result in Section 4. The main proofs appear in Sections 5 and 6; all lemmas are proved in the appendix. Simulation experiments are reported in Section 7.

Notation. Given a Markov chain $X = (X^t, t \geq 0)$, we let $\mathbb{P}_x\{\mathcal{B}\}$ be the likelihood of the event $\mathcal{B}$ when $X^0 = x$. Similarly, $\mathbb{P}_\Pi\{\mathcal{B}\}$ is the likelihood of that event when $X^0 \sim \Pi$; $\mathbb{E}_x[\cdot]$ and $\mathbb{E}_\Pi[\cdot]$ denote then the corresponding expectations. Throughout the remainder of the paper, for two non-negative sequences a^N, b^N and a non-negative function $h: \mathbb{R}_+ \to \mathbb{R}_+$, $a^N = b^N + \mathcal{O}(h(N))$ means that $|a^N - b^N| = \mathcal{O}(h(N))$; the same is true for the scaling notation $o(\cdot)$ and $\Omega(\cdot)$.

We will be introducing various process variables throughout the paper. For ease of reference, we provide a notation index at the end of the paper.

2. The Model: Dynamic Allocation of Reusable Resources

There are n types of customers labeled by $i \in [n] \equiv \{1, \ldots, n\}$ and d resources labeled by $j \in [d]$. The decision maker has q_j reusable units of resource j.

Customers of type $i \in [n]$ arrive following a Poisson process with rate $\lambda_i > 0$, and request the simultaneous possession of a unit from each of multiple resources. The resource consumption is encoded in an adjacency matrix $A \in \{0,1\}^{d \times n}$ such that $A_{ji} = 1$ if type-i customers require a unit of resource j. We visualize the network topology as in Figure 1(RIGHT). An edge between request type i (a circle) and a resource (a rectangle) j corresponds to $A_{ji} = 1$.

If a type-i customer's request is accepted, the decision maker collects a reward r_i, and the requested resource units are allocated and occupied for an exponentially-distributed amount of time with mean μ_i^{-1}. For example, an accepted request of type 1 in Figure 1(RIGHT) occupies *simultaneously* one unit from each of the resources b, c, and d. These three units are released simultaneously after an exponential amount of time with mean μ_1^{-1}.

For each $i \in [n]$, we denote by $\mathcal{S}(i) := \{j \in [d] : A_{ji} = 1\}$ the set of resources required by type-i customers and for each $j \in [d]$, we denote by $\mathcal{A}(j) := \{i \in [n] : A_{ji} = 1\}$ the set of customer types requiring resource j. To avoid trivialities, we assume that $\mathcal{S}(i), \mathcal{A}(j) \neq \emptyset$ for all $i \in [n], j \in [d]$. A customer's request can be accepted—*accepting it is feasible*—only if all the required resources have available units when the request is made. The decision maker can reject a request even if it is feasible to accept it. Once a customer is accepted, the requested resource units are immediately occupied; these resource units become simultaneously available at the conclusion of the customer's service. *Preemption is not allowed*: once accepted, a request is processed to completion. Arrival processes are assumed independent across customer types, and resource occupation times are independent across customers.

The objective of the decision maker is to maximize the long-run average collected reward subject to the resource constraint. Let $r = (r_1, \ldots, r_n)$ be the reward vector and $D^\pi = (D_1^\pi(t), \ldots, D_n^\pi(t))$

where $D_i^\pi(t)$ is the number of type-i customers accepted by time t under policy π; both column vectors. The performance of the optimal policy is given by

$$\mathcal{R}^* = \sup_\pi \liminf_{t\uparrow\infty} \frac{1}{t}\mathbb{E}\left[r' D^\pi(t)\right],$$

where the sup is taken over all non-anticipating non-preemptive policies.

The optimal policy can be obtained via dynamic programming. The state descriptor X^t is n-dimensional and X_i^t tracks the number of customers of type-i customers in-service at time t. The state space has $\prod_{i=1}^n \min_{j\in\mathcal{S}(i)} q_j$ states which, except the simplest networks, is computationally prohibitive.

Instead, our goals are: (i) to characterize (indirectly) the performance of the optimal policy—specifically its *regret* (or approximation gap): how close its performance is to a natural deterministic upper bound, and (ii) to offer a simple policy that achieves the optimal regret scaling in a high-volume many-server regime.

The linear programming relaxation of the problem is the starting point of our policy design. Given a policy π, let $z_i^\pi = \liminf_{t\uparrow\infty}\frac{1}{t}\mathbb{E}[D_i^\pi(t)]$. Because no more type-$i$ requests can be accepted than those arriving, we must have $z_i^\pi \le \lambda_i$. By Little's law, z_i^π/μ_i is the long-run average number of type-i customers in service (occupying resources) so it must be the case that $\sum_i A_{ji} z_i^\pi/\mu_i \le q_j$, for all $j\in[d]$.

Because each policy π must satisfy these constraints, the linear program below is an upper bound on the long-run average reward of *any* admissible policy:

$$\begin{aligned} \max_{z\in\mathbb{R}^n_+} \ & r'z \\ \text{s.t.}\ & A(z/\mu)\le q, \\ & z\le\lambda. \end{aligned}$$

Here $z/\mu = (z_1/\mu_1,\dots,z_n/\mu_n)$. Changing variables $y\leftarrow z/\mu$ we re-write this as

$$\begin{aligned} \overline{\mathcal{R}}(q,\lambda/\mu) := \max_{y\in\mathbb{R}^n_+} \ & r'_\mu y \\ \text{s.t.}\ & Ay\le q, \\ & y\le \lambda/\mu, \end{aligned} \tag{LP}$$

where $r_\mu = (r_1\mu_1,\dots,r_n\mu_n)$ and λ/μ is the vector with elements $\lambda_i/\mu_i, i\in[n]$.

The value $\overline{\mathcal{R}}(q,\lambda/\mu)$ is an upper bound on the expected long-run average reward, $\mathcal{R}^\pi$, collected by a non-anticipating non-preemptive policy π:

$$\mathcal{R}^\pi \le \mathcal{R}^* \le \overline{\mathcal{R}}(q,\lambda/\mu).$$

The optimal solution y^* of (LP) yields a partition of the customer types: it

- accepts all type-i customers with $y_i^* = \lambda_i/\mu_i$ (the preferred types);
- accepts a fraction of type-i customers with $y_i^* \in (0, \lambda_i/\mu_i)$ (less preferred types); and
- rejects all type-i customers with $y_i^* = 0$ (rejected types).

The groups

$$\mathcal{A}_p := \{i \in [n] : y_i^* = \lambda_i/\mu_i\}, \qquad \text{(preferred)}$$

$$\mathcal{A}_{lp} := \{i \in [n] : y_i^* \in (0, \lambda_i/\mu_i)\}, \qquad \text{(less preferred)}$$

$$\mathcal{A}_0 := \{i \in [n] : y_i^* = 0\}, \qquad \text{(rejected)}$$

form a partition of $[n]$: $\mathcal{A}_p \cup \mathcal{A}_{lp} \cup \mathcal{A}_0 = [n]$.

Example 2.1 (The single resource case) For $d = 1$, (LP) has a single capacity constraint and a "packing" solution. Suppose that types are labeled in decreasing order of $r_i\mu_i$: $r_1\mu_1 > r_2\mu_2 \ldots > r_n\mu_n$ and let $i^* = \max\{i : \sum_{k=1}^{i} \lambda_k/\mu_k \leq q\}$. The optimal solution has $y_l^* = \lambda_l/\mu_l$ for all $l \leq i^*$, $y_{i^*+1}^* = q - \sum_{l=1}^{i^*} \lambda_l/\mu_l$ and $y_l = 0$ otherwise. The optimal value is $\overline{\mathcal{R}}(q, \lambda/\mu) = \sum_{i=1}^{i^*} r_i\lambda_i + r_{i^*+1}\mu_{i^*+1}\left(q - \sum_{l=1}^{i^*} \lambda_l/\mu_l\right)$. ■

We assume for the rest of the paper, and without loss of generality, that $\mathcal{A}_0 = \emptyset$. Our policy does not serve those requests and achieves logarithmic regret relative to the LP-based upper bound.

The high-volume many-server regime. We study reward maximization in a standard (e.g., Puhalskii and Reiman (1998); Hunt and Kurtz (1994); Hunt and Laws (1997)) high-volume and many-server regime. The customer arrival rates and the number of resource units scale at the same rate:

$$\lambda_i^N = N\lambda_i, \; i \in [n], \qquad \text{and} \qquad q_j^N = Nq_j, \; j \in [d],$$

where $\lambda_i, i \in [n]$ and $q_j, j \in [d]$ are strictly positive.

With this scaling, $\overline{\mathcal{R}}(q^N, \lambda^N/\mu) = N\overline{\mathcal{R}}(q, \lambda/\mu)$ so that

$$\mathcal{R}^{\pi,N} \leq \overline{\mathcal{R}}(q^N, \lambda^N/\mu) = N\overline{\mathcal{R}}(q, \lambda/\mu),$$

for any policy π in the N^{th} network. We add the superscript N, to denote quantities for the N^{th} network; $\mathcal{R}^{*,N}$, for example, is the reward collected by the optimal policy in the N^{th} network.

Remark 2.1 (An offline upper bound) The number of type-i customers in the system at a given time t, it is easily seen, is bounded from above by that number in an infinite server queue with arrival rate λ_i and service rate μ_i; this number is distributed, in steady state, as a Poisson random variable with mean λ_i^N/μ_i. Then $\mathcal{R}^{*,N} \leq \mathbb{E}[\mathcal{R}(q^N, Y^N)] \leq \overline{\mathcal{R}}(q^N, \lambda^N/\mu)$. where

$$\overline{\mathcal{R}}(q^N, Y^N) := \begin{cases} \max\limits_{y \in \mathbb{R}^n_+} r'_\mu y \\ \text{ s.t. } Ay \leq q^N, \\ \qquad y \leq Y^N, \end{cases} \tag{1}$$

with Y_i^N a Poisson random variable with mean λ_i^N/μ_i and $Y^N = (Y_i^N, i \in [n])$. It is easy to show that, under Assumption 3.1, $\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathbb{E}[\overline{\mathcal{R}}(q^N, Y^N)] = \mathcal{O}(1)$ so that the offline static upper is as crude as the deterministic upper bound. Nevertheless, it is useful in showing that, when our overload assumption is violated, the gap of the optimal policy from the deterministic LP upper bound can be substantial; see Lemma 3.1. ■

3. Overloaded Networks

In the single resource case, a resource is, intuitively speaking, overloaded if there is more demand than the server can handle: $q < \sum_{i \in [n]} \lambda_i/\mu_i$. We require, in addition, that $y_{i^*+1} > 0$, which makes the LP non-degenerate; recall Example 2.1. The following condition is a network generalization.

Assumption 3.1 (Network overload.) *The linear program* (LP) *has a unique and non-degenerate solution and at least one resource constraint is, at optimality, tight.*

Two implications of this assumption justify referring to networks that satisfy this assumption as overloaded. The obvious one is that there is a resource constraint held at equality. But this is not all. Because of the uniqueness and non-degeneracy of the primal, the dual has a unique solution. The unique solution pair (of the primal and the dual) must satisfy strict complementarity (see, e.g., Theorem 10.7 of Vanderbei, 1998). In turn, the dual variables of all *binding resource* constraints are strictly positive: increasing the capacity of any of these binding resource constraints will lead to an increase in objective function value. This means that the resource constraints held at equality are binding in a strong sense.

Non-degeneracy implies dual uniqueness. The uniqueness of the dual variables corresponding to the demand constraints is, in fact, necessary for a logarithmic regret. For the following recall hat $\mathcal{R}^{*,N} \leq \overline{\mathcal{R}}(q^N, \lambda^N/\mu)$.

Lemma 3.1 (Necessity of dual uniqueness) *Suppose that the dual to* (LP) *has two optimal solutions that differ in the shadow prices of the demand constraints. Then,*

$$\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathcal{R}^{*,N} = \Omega(\sqrt{N}).$$

The intuition here is simple and best understood through the offline upper bound (1). This upper bound is a stochastic perturbation of the deterministic one (where the perturbations are centered on the demand and service times, in turn on the maximal mean occupancy λ/μ). When the shadow prices of the demand constraints are not unique, perturbations of the demand in different directions have different effects on the dual (and hence primal) objective function value. The perturbation of the Poisson right-hand side Y^N in (1) around its mean *is symmetric*, and of the order of $\sqrt{N}$ and, because of the dual non-uniqueness, the effects of these stochastic perturbations on the objective do not "average out". Thus, $\overline{\mathcal{R}}(q^N,\lambda^N/\mu)-\mathcal{R}^{*,N}\geq\overline{\mathcal{R}}(q^N,\lambda^N/\mu)-\mathbb{E}[\overline{\mathcal{R}}(q^N,Y^N)]=\Omega(\sqrt{N})$.

Assumption 3.1 allows for non-binding resource constraints at optimality, i.e., for the existence of "underutilized" resources that have strictly positive slack. These resources and their constraints can be removed from (LP) without affecting the optimal solution and its value. For the remainder of the paper, *we assume that all d resource constraints are binding.* We re-visit this simplification upon the conclusion of the proofs (see Remark 6.1).

Lemma 3.2 *Suppose that Assumption 3.1 holds, then $|\mathcal{A}_{lp}|=d$. Moreover, the sub-matrix of A that has only the columns for $i\in\mathcal{A}_{lp}$ is full rank.*

The optimization problem (LP) is defined on a bipartite graph, e.g., Figure 1(RIGHT) where customer types are on one side of the partition and resources are on the other side. There is an edge between customer type $i\in[n]$ and resource $j\in[d]$ if $A_{ji}=1$.

Definition 3.1 (The lp-residual graph) *The lp-residual graph, $\mathcal{G}_{lp}$, is the graph obtained by removing all preferred types $i\in\mathcal{A}_p$ (and the edges that connect them to resources).*

It follows from Lemma 3.2 that the lp-residual graph is bipartite with d vertices in each of its constituent sets: d customer types and d resources. A perfect matching in a graph is a set of edges such that each vertex is incident to exactly one edge. The incidence matrix of the lp-residual graph is the sub-matrix of A that has only the columns for $i\in\mathcal{A}_{lp}$; by Lemma 3.2 it is full rank. This guarantees the existence of a perfect matching in the lp-residual graph (see, e.g., Theorem 7.3 of Motwani and Raghavan, 1995; Tutte, 1947).

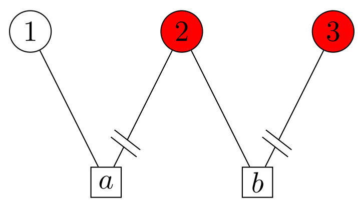
If there exist multiple perfect matchings in the residual graph (see e.g., Figure 4) we pick one arbitrarily. *For each resource j, we write i_j for the (less-preferred) request type that is matched with resource j in this perfect matching. For each type $i\in\mathcal{A}_{lp}$, j_i is the resource matched to i.*

We will be using three network examples throughout this paper. Examples 3.1 and 3.2 have a unique perfect matching in the residual graph. The former is the simplest possible example with more than one resource and is useful to illustrate basic constructions and develop intuition. The latter is a more elaborate example and is a useful test for the numerical performance of our proposed algorithm. Example 3.3 has multiple perfect matchings. The *corrected headcount process* that we introduce later, and that plays a role in our algorithm, takes on a less intuitive form in this case.

Example 3.1 Consider the network with two types of resource units $\{a,b\}$, and three types of customers $\{1,2,3\}$ in Figure 2. Type-2 customers request a unit of both type-a and type-b resources, type-1 customers request a unit of type-a resource, and type-3 customers request a unit of type-b resource.

We set the resource units to $q=(7,6)$ and the customer type parameters to $\lambda=(3,2,5)$, $\mu^{-1}=(2,1,3)$ $(\lambda/\mu=(6,2,15))$. The reward vector is $r=(5,1,2)$ so that $r_\mu=(5/2,1,2/3)$. The (LP) has the unique non-degenerate solution $y^*=(6,1,5)$—$y_1^*=\lambda_1/\mu_1$, $y_2^*\in(0,\lambda_2/\mu_2)$, $y_3^*\in(0,\lambda_3/\mu_3)$—and the dual variables for the resource constraints are $\alpha^*=(1/3,2/3)>0$.

Figure 2 An example with three customer types and two resource types. Less preferred types are marked in red. Edges corresponding to a type and its paired resource are marked with ||.



In this case, types 2 and 3 are the less-preferred types ($\mathcal{A}_{lp}=\{2,3\}$), and are colored in red in Figure 2. The residual graph contains types 2 and 3 and both resources. The unique perfect matching is $\{(2,a),(3,b)\}$ so that $i_a=2, i_b=3$. ■

Example 3.2 Consider the network with 7 customer types and 4 resource types in Figure 3(LEFT). We set the resource units to $q=(11,19,14,7)$ and the customer type parameters to $\lambda=(2,3,5,1,6,2,3)$, $\mu^{-1}=(1,3,2,3,5,4,2)$, so that $\lambda/\mu=(2,9,10,3,30,8,6)$. Finally we take the reward $r=(2,1,3,5,1,6,5)$ so that $r_\mu=(2,1/3,3/2,1\,2/3,1/5,3/2,5/2)$.

The LP has the unique non-degenerate solution $y^*=(1,8,5,3,5,8,6)$, with the resource-constraint dual variables equal to $\alpha^*=(2/15,1/5,13/10,1/2)$. The unique perfect matching in the residual graph is $\{(1,d),(2,a),(3,c),(5,b)\}$. ■

Example 3.3 Consider the network with 6 customer types and 3 resource types in Figure 4(LEFT). We set the resource units to $q=(2,2,2)$ and the customer type parameters to $\lambda=(2,2,2,1/2,1/3,1/4)$, $\mu^{-1}=(1,1,1,2,3,4)$, so that $\lambda/\mu=(2,2,2,1,1,1)$. Finally we take the reward $r=(1,1,1,4,6,8)$ so that $r_\mu=(1,1,1,2,2,2)$. The LP has the unique non-degenerate solution $y^*=(1,1,1,1,1,1)$ and the dual variables for the resource constraints are $\alpha^*=(1/2,1/2,1/2)$. This network has two perfect matchings: one is $(1,a)$,$(2,b)$ and $(3,c)$ and the other is $(1,b)$, $(2,c)$ and $(3,a)$. ■

Figure 3 An example with 7 customer types and 4 resource types. (LEFT) Less preferred types are marked in red. Edges corresponding to a type and its paired resource are marked with ||. (RIGHT) The *lp*-residual graph.

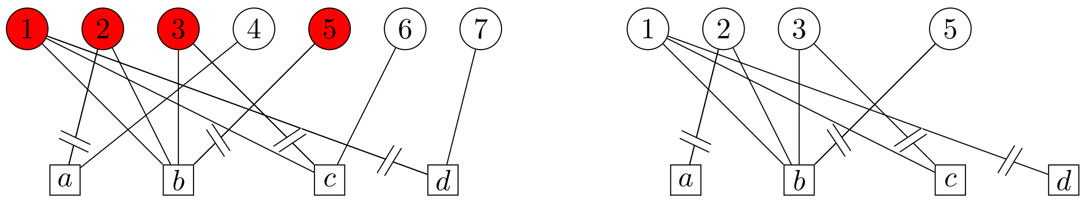
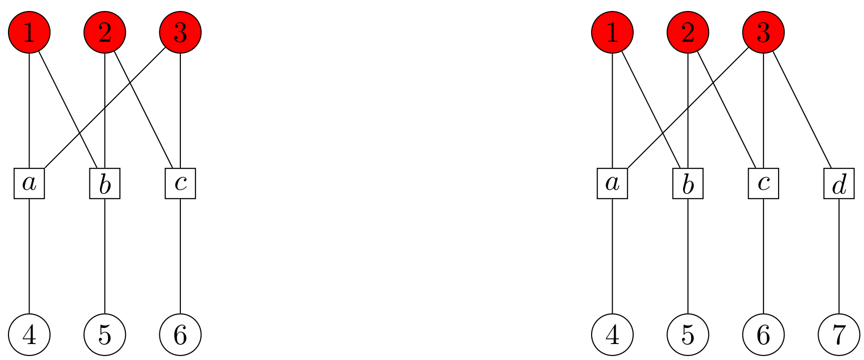


Figure 4 Example with either multiple perfect matchings or none. In the network on the left, $\lambda = (2,2,2,1/2,1/3,1/4)$, $\mu = (1,1,1,1/2,1/3,1/4)$, $r = (1,1,1,4,6,8)$ and $q = (2,2,2)$. The LP solution satisfies Assumption 3.1 and the less-preferred types are 1, 2, and 3. The *lp*-residual graph has two perfect matchings: $\{(1,a),(2,b),(3,c)\}$ and $\{(1,b),(2,c),(3,a)\}$. On the right is an expanded network with the addition of a type 7 and a resource d that have $\lambda_7 = \mu_7 = 1$, $r_7 = 10$, $q_d = 2$; all other parameters remain the same. In this network, there is no perfect matching. The LP for this network has a degenerate LP solution where the slack variable for the binding resource d is zero. For visibility, we draw the preferred types at the bottom of the graph.



The target allocation levels. Let A_{lp} be the incidence matrix of the *lp*-residual graph. This is the square $d \times d$ sub-matrix of A that has all d rows and the d columns corresponding to the less-preferred types $i \in \mathcal{A}_{lp}$; $\mathcal{A}_{lp}$ is a non-singular matrix by Lemma 3.2. Similarly, let A_p be the matrix with the columns corresponding to the preferred types. The residual resource-j capacity, after allocation to the preferred types, is $q_j^N - \sum_{i \in \mathcal{A}_p} A_{ji}\lambda_i^N/\mu_i$ or, in vector form, $q^N - A_p(\lambda^N/\mu)_{\mathcal{A}_p}$. The optimal deterministic allocation to the less-preferred types is then

$$Ny^*_{\mathcal{A}_{lp}} = A_{lp}^{-1}\left(q^N - A_p(\lambda^N/\mu)_{\mathcal{A}_p}\right). \tag{2}$$

The solution to (LP) is non-degenerate so that for a small perturbation ζ of $(\lambda^N/\mu)_{\mathcal{A}_p}$, we have the same optimal basis and, in turn,

$$x^*(\zeta) := A_{lp}^{-1}\left(q^N - A_p\zeta\right) \tag{3}$$

is the optimal allocation of residual capacity among less-preferred types given a load ζ from the preferred types. Using (2), we write this *target allocation* as

$$x^*(\zeta) = y^*_{\mathcal{A}_{lp}} N + A_{lp}^{-1} A_p[(\lambda^N/\mu)_{\mathcal{A}_p} - \zeta]; \tag{4}$$

because $|\mathcal{A}_{lp}| = d$, $x^*(\zeta)$ is a d-dimensional vector.

We use $x^*(\zeta)$ as a definition regardless of whether or not ζ is a sufficiently small perturbation of $(\lambda_l^N/\mu)_{\mathcal{A}_p}$.

4. The Corrected-Headcount Threshold (CHT) Policy and the Regret Bound

For each $i \in [n]$, we use X_i^t to denote the number of type-i customers occupying resource unit(s) at time t. We let

$$\Sigma_j^t = \sum_{i \in \mathcal{A}(j)} X_i^t = \sum_{i \in \mathcal{A}} A_{ji} X_i^t, \qquad \text{(resource } j \text{ headcount)}$$

be the number of customers that are occupying resource-j units at time t.

Our policy uses, dynamically, the target allocation levels (3). At time t, and with $X^t_{\mathcal{A}_p}$ being the real-time allocation to preferred types, the target allocation to less-preferred type i is

$$X_i^{*,t} := x_i^*(X^t_{\mathcal{A}_p}); \tag{5}$$

it is a random quantity that depends on the real-time count of preferred-type customers in service.

Recall that $i_j \in [n]$ is the customer type matched to resource $j \in [d]$ under the chosen perfect matching. The corrected-headcount threshold policy (CHT, see Algorithm 1) is applied to *corrected* headcount processes

$$\Sigma_j^{*,t} := \Sigma_j^t + \sum_{i \in \mathcal{A}_{lp} \setminus i_j} A_{ji}(X_i^{*,t} - X_i^t),$$

where the *correction* process

$$Z_j^{*,t} = \sum_{i \in \mathcal{A}_{lp} \setminus i_j} A_{ji}(X_i^{*,t} - X_i^t),$$

captures the difference between targeted occupancy levels and actual levels for less preferred types.

A threshold on the corrected headcount process at resource j_i (the resource matched with $i \in \mathcal{A}_{lp}$) determines whether a type-i request can be accepted or not. Thresholds are placed only on the edges (i, j_i) for $i \in \mathcal{A}_{lp}$; these are the edges marked with || in Figures 2 and 3. There is one threshold per $i \in \mathcal{A}_{lp}$ and hence, by Lemma 3.2, a total of $d = |\mathcal{A}_{lp}|$ thresholds.

Conceptually, the use of the corrected headcount process helps network alignment. Resource j_i will be more conservative in accepting type i if types $k \in \mathcal{A}_{lp}(j_i) \setminus i$ are significantly below their targeted levels and, in doing so, will be reserving capacity for their arrivals.

Algorithm 1 Corrected-Headcount Threshold Policy (**CHT**)

Require: Threshold constants δ_i for every less-preferred customer type $i \in \mathcal{A}_{lp}$.

1: Accept an arriving request of type $i \in \mathcal{A}_p$ (preferred types) whenever feasible.

2: Accept an arriving type $i \in \mathcal{A}_{lp}$ at time t if and only if it is feasible ($\Sigma_j^t < q_j^N$ for all $j \in \mathcal{S}(i)$), and there are—in terms of the corrected headcount—more than $R_i := \delta_i \log N$ units of resource j_i available:

$$q_{j_i}^N - \Sigma_{j_i}^{*,t} \geq R_i = \delta_i \log N.$$

The vector $\delta = (\delta_1, \ldots, \delta_d)$ is not arbitrary. These must be chosen to be sufficiently large (but independent of N). We provide some guidance on the choice of thresholds in §B of the Appendix.

Given the identification of the threshold placement, a more natural threshold policy would be one that uses the headcount Σ_j^t instead of its corrected counterpart $\Sigma_j^{*,t}$, but the corrected headcount has significant mathematical benefits. Except X_i^t, all less-preferred types requiring j_i appear in the corrected headcount only through their targeted amounts. We further discuss the mathematical implications of the policy design in Section 5 and compare CHT and the natural threshold policy in our numerical experiments in Section 7; see also Example 4.1.

Example 4.1 Let's spell out the policy implementation in Examples 3.1, 3.2, and 3.3. In Example 3.1, the unique perfect matching is $\{(2,a),(3,b)\}$. The policy is spelled out in Figure 5. In Example 3.2, the unique perfect matching is $\{(1,d),(2,a),(3,c),(5,b)\}$, and Figure 6 spells out the policy for this network. In Example 3.3, there are two perfect matchings; we use the perfect matching $\{(1,a),(2,b),(3,c)\}$ and the policy is spelled out in Figure 7.

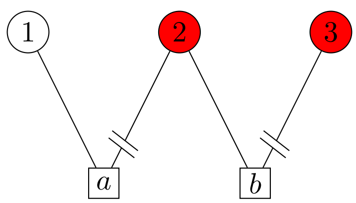
It is useful to use the simple Example 3.1 to highlight a benefit of the corrected headcount process. Suppose that, at a time t, $X_2^t + X_3^t \leq q_b^N - \delta_3 \log N$, but $X_2^t < X_2^{*,t}$ and $X_2^{*,t} + X_3^t > \delta_3 \log N$. In this state, the "standard" threshold policy, acting on the true headcount, accepts type-3 requests. CHT, in contrast, is more conservative in accepting type-3, effectively reserving capacity for type-2 so that it can be brought back to its targeted level $X_2^{*,t}$. While the regular threshold policy might eventually "correct itself", CHT acts forcefully to align allocation levels with their targeted values.

■

Under CHT, $X^t \in \mathbb{N}_+^n$ forms an n-dimensional continuous-time Markov chain with a finite state space. It is easy to check that the chain is irreducible[1] and, as a consequence, conclude that

[1] Take states $x, y \in \mathcal{X}$ where $\mathcal{X} = \mathcal{X}^N \equiv \{x \in \mathbb{N}^n : Ax \leq q^N\} \cap \{x \in \mathbb{N}^n : x_i \leq q_{j_i}^N - \delta_i \log N, \text{ for all } i \in \mathcal{A}_{lp}\}$ is the state space of the chain. Then, there is a path from x to 0, through x_i service completions of type i. There is then a path from 0 to y, through y_i arrivals of type i.

Figure 5 Policy implementation for Example 3.1.

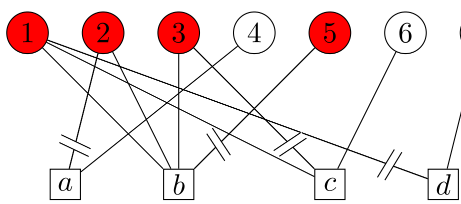


Accept a request of type

- 1 when feasible
- 2 when feasible and $q_a - \Sigma_a^{*,t} > \delta_2 \log N$ with
$$\Sigma_a^{*,t} = \Sigma_a^t = X_1^t + X_2^t$$
- 3 when feasible and $q_b - \Sigma_b^{*,t} > \delta_3 \log N$ where
$$\Sigma_b^{*,t} = \Sigma_b^t + X_2^{*,t} - X_2^t = X_3^t + X_2^{*,t}$$

The target allocation for type 2 is $X_2^{*,t} = q_a^N - X_1^t$.

Figure 6 Policy implementation for Example 3.2.



Accept a request of type

- 4, 6, 7 when feasible
- 1 when feasible and $q_d - \Sigma_d^{*,t} > \delta_1 \log N$ with
$$\Sigma_d^{*,t} = \Sigma_d^t = X_7^t + X_1^t$$
- 2 when feasible and $q_a - \Sigma_a^{*,t} > \delta_2 \log N$ where
$$\Sigma_a^{*,t} = \Sigma_a^t = X_4^t + X_2^t$$
- 3 when feasible and $q_c - \Sigma_c^{*,t} > \delta_3 \log N$ where
$$\Sigma_c^{*,t} = \Sigma_c^t + X_1^{*,t} - X_1^t = X_6^t + X_3^t + X_1^{*,t}$$
- 5 when feasible and $q_b - \Sigma_b^{*,t} > \delta_5 \log N$ where
$$\Sigma_b^{*,t} = \Sigma_b^t + \sum_{i=1}^{3}(X_i^{*,t} - X_i^t) = X_5^t + \sum_{i=1}^{3} X_i^{*,t}$$

The target allocations for types 1, 2, and 3 are given by

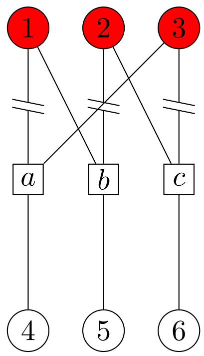
$$X_1^{*,t} = q_d^N - X_7^t,\ X_2^{*,t} = q_a - X_4^t,\ \text{and}\ X_3^{*,t} = q_c^N - X_6^t - X_1^{*,t}.$$

there exists a unique stationary distribution Π which is also the steady-state distribution. When considering stationary variables, we omit the time superscript t. By Little's law, the number of type-i requests accepted per unit of time in stationarity equals $\mu_i \mathbb{E}_\Pi[X_i]$ so the performance of CHT is $\mathcal{R}^{CHT} = r_\mu' \mathbb{E}_\Pi[X] = \sum_i r_i \mu_i \mathbb{E}_\Pi[X_i]$.

Theorem 4.1 (Main result) *Suppose that Assumption 3.1 holds. Let* Π *be the stationary distribution induced by* CHT*. Then,*

$$\mathcal{R}^{*,N} - r_\mu' \mathbb{E}_\Pi[X] \le \overline{\mathcal{R}}(q^N, \lambda^N/\mu) - r_\mu' \mathbb{E}_\Pi[X] = \mathcal{O}(\log N).$$

Figure 7 Policy implementation for Example 3.3.



Accept a request of type
- $4, 5, 6$ when feasible
- 1 when feasible and $q_a - \Sigma_a^{*,t} > \delta_1 \log N$ with
$$\Sigma_a^{*,t} = X_1^t + X_3^{*,t} + X_4^t$$
- 2 when feasible and $q_b - \Sigma_b^{*,t} > \delta_2 \log N$ where
$$\Sigma_b^{*,t} = X_2^t + X_1^{*,t} + X_5^t$$
- 3 when feasible and $q_c - \Sigma_c^{*,t} > \delta_3 \log N$ where
$$\Sigma_c^{*,t} = X_3^t + X_2^{*,t} + X_6^t$$

The target allocations for types 1, 2, and 3 are given by

$$X_1^{*,t} = \frac{1}{2}(q_a^N + q_b^N - q_c^N) - \frac{1}{2}(X_4^t + X_5^t - X_6^t),$$
$$X_2^{*,t} = \frac{1}{2}(-q_a^N + q_b^N + q_c^N) - \frac{1}{2}(-X_4^t + X_5^t + X_6^t), \text{ and}$$
$$X_3^{*,t} = \frac{1}{2}(q_a^N - q_b^N + q_c^N) - \frac{1}{2}(X_4^t - X_5^t + X_6^t).$$

Moreover, for any network that satisfies Assumption 3.1 with $\mathcal{A}_p \neq \emptyset$, and any family of admissible policies $(\pi_N, N \in \mathbb{N}_+)$,

$$\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathcal{R}^{\pi_N, N} = \Omega(\log N).$$

Remark 4.1 (The case $d = 1$) If there is a single resource, then by Lemma 3.2, $|\mathcal{A}_{lp}| = 1$ (a single less-preferred type). In this case, the corrected headcount process is the same as the headcount itself: $\Sigma^{*,t} = \Sigma^t = \sum_{i \in \mathcal{A}_p} X_i^t + X_{i_0}^t$, where i_0 is the unique less preferred type. Our policy reduces in this case to the simple threshold policy that:

1. accepts any request of types $i \in \mathcal{A}_p$ whenever there are units of the resource available, and
2. accepts a request of type i_0 arriving at time t, if $q^N - \Sigma^t \geq \delta_i \log N$.

It follows from Theorem 4.1 that the approximation-gap of standard trunk reservation with a logarithmic threshold is logarithmic in N. For the special case with equal service rates ($\mu_i \equiv \mu$), the logarithmic lower bound is implied by the literature. Reiman (1991); Morrison (2010) proved that the *optimal policy's* threshold for the class with the smallest reward is logarithmic in N. It then follows, from arguments similar to our proofs, that the approximation gap of the optimal policy in the overloaded regime is logarithmic in N; no policy can do better than the optimal one. In this paper, we prove that this lower bound is generally true in overloaded networks. ■

Remark 4.2 (Why a logarithmic threshold?) Having the thresholds scale logarithmically in N, is the sufficient (and necessary) "capacity reservation" to protect the preferred-type requests.

For intuition, it suffices to consider the case of a single resource ($d=1$) and with a common service rate across types ($\mu_i \equiv \mu$). When the number of occupied servers is greater $q^N - R$, only customers of types $\mathcal{A}_p$ are accepted. In these states, the headcount goes up at rate $\sum_{i\in\mathcal{A}_p}\lambda_i^N$ and down, at approximately, rate μq^N. Approximately, the headcount behaves in these states like an $M/M/1$ queue with utilization

$$\rho^- := \frac{\sum_{i\in\mathcal{A}_p}\lambda_i^N}{\mu q^N} < 1.$$

It is the non-degeneracy assumption that guarantees that $\rho^- < 1$. The likelihood that all servers are busy—at which point preferred customers will be rejected—is roughly ρ^R. Choosing R to be of the order of $\log N$, guarantees that this probability is of the order of $1/N$. A threshold that is an order of magnitude smaller would not suffice; a constant threshold (i.e., one that does not scale with N) in particular would result in a non-negligible fraction of preferred customers being rejected. ∎

Remark 4.3 (the lower bound when $\mathcal{A}_p = \emptyset$) For the lower bound in Theorem 4.1 we assume that $\mathcal{A}_p \neq \emptyset$. This is generally necessary. To see this, consider a single resource ($d=1$) and suppose that types are labeled in decreasing order of $r_i\mu_i$: $r_1\mu_1 > r_2\mu_2 \ldots > r_n\mu_n$. Recalling the structure of (LP)'s solution in Example 2.1, we note that if $\mathcal{A}_p = \emptyset$, then that $y_1^* = q < \lambda_1/\mu_1$, $y_i^* = 0$ for all $i \neq 1$, and $\overline{\mathcal{R}}(q, \lambda/\mu) = r_1\mu_1 q$.

The policy, that accepts a type-1 request whenever there is capacity available and rejects all other requests, achieves constant regret. Indeed, under this policy X_1^t has the law of a single-class $M/M/q^N/q^N$ queue with arrival rates λ_1^N, service rate μ_1 with load $\rho = \frac{\lambda_1^N}{\mu_1 q^N} = \frac{\lambda_1}{\mu_1}/q > 1$. It is a simple fact that this "overloaded" single-class loss queue has $\mathbb{E}_\Pi[q^N - X_1] = \mathcal{O}(1)$ where Π is the steady-state distribution. In turn, $r_1\mu_1\mathbb{E}_\Pi[X_1] = r_1\mu_1 q^N + \mathcal{O}(1) = \overline{\mathcal{R}}(q^N, \lambda^N/\mu) + \mathcal{O}(1) = \overline{\mathcal{R}}(q^N, \lambda^N/\mu) + o(\log N)$. ∎

Remark 4.4 (state-dependence in the policy is necessary) Our proposed policy is state-dependent, meaning that the decision to accept or reject a request depends on the state of the system (beyond the mere availability of resource units). In Section C of the Appendix we prove that state-dependence is necessary: there is no static randomization policy that achieves the logarithmic regret. ∎

5. Proof of the upper bound in Theorem 4.1

Recall that $\overline{\mathcal{R}}(q^N, \lambda^N/\mu) = r_\mu' N y^*$ and $\mathcal{R}^{CHT} = r_\mu'\mathbb{E}_\Pi[X]$, so that

$$\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathcal{R}^{CHT} = r_\mu'(Ny^* - \mathbb{E}_\Pi[X]),$$

where $\mathbb{E}_{\Pi}[X] = (\mathbb{E}_{\Pi}[X_1], \ldots, \mathbb{E}_{\Pi}[X_n])'$. Accordingly, we focus our analysis on bounding the gap $Ny_i^* - \mathbb{E}_{\Pi}[X_i]$.

The vector $\mathbb{E}_{\Pi}[X]$ is a non-trivial entity to study directly because of the dependencies between customer types. In the network of Figure 3, whether a type-3 request can be accepted depends on both the number of available units of resource c and the number of available units of resource b, which is in turn affected by the number of types 1 and 2 customers in the system.

Instead of studying X^t directly, we introduce a network where the resource constraints are relaxed in a way that facilitates analysis but, at the same time, has a provably close performance to that of CHT in the original network.

A network with relaxed constraints

The network, operated by CHT, has the state space

$$\mathcal{X} = \mathcal{X}^N \equiv \left\{x \in \mathbb{N}^n : Ax \leq q^N\right\} \cap \{x \in \mathbb{N}^n : x_i \leq q_{j_i}^N - \delta_i \log N, \text{ for all } i \in \mathcal{A}_{lp}\}. \tag{6}$$

Because no type-i ($i \in \mathcal{A}_{lp}$) requests are accepted when more than $q_{j_i}^N - \delta_i \log N$ units of resource j_i are occupied, there can be no more than this number of these in the system; this is captured by the second set in the intersection that defines $\mathcal{X}$. In the auxiliary network, we remove all but these constraints; it has the state space

$$\widetilde{\mathcal{X}} = \widetilde{\mathcal{X}}^N \equiv \left\{x \in \mathbb{N}^n : x_i \leq q_{j_i}^N - \delta_i \log N \text{ for all } i \in \mathcal{A}_{lp}\right\}. \tag{7}$$

In the auxiliary system, there is an infinite number of units of each resource, and access of types $i \in \mathcal{A}_{lp}$ is restricted only by the threshold on the headcount of resource j_i (and not by the occupancy of any other resource $j \neq j_i$). Access to types $i \in \mathcal{A}_p$ is not restricted at all. This localizes admission decisions. Whether or not we accept a type-i request depends only on the corrected headcount of resource j_i, not that of any other resource.

Auxiliary system notation: We superscript with $\sim$ processes associated with the auxiliary system: $\widetilde{X}_i^t$ counts the number of type-i customers in the relaxed network under the policy $\widetilde{\pi}$ in Algorithm 2, $\widetilde{\Sigma}_j^t$ is the resource-j headcount process in the auxiliary system, $\widetilde{Z}_j^{*,t}$ is the correction process, and $\widetilde{\Sigma}_j^{*,t}$ is the corrected headcount process. We use $\widetilde{\Pi}$ for the steady-state distribution in the auxiliary system using the policy $\widetilde{\pi}$; $\widetilde{\Pi}(\mathcal{B})$ for $\mathcal{B} \subseteq \widetilde{\mathcal{X}}$ is the probability that this distribution assigns to the set $\mathcal{B}$ of states. The existence of this distribution is established in Proposition 5.1.

How the relaxation supports the analysis of CHT? In Example 4.1, we illustrated how CHT, in contrast to the "regular" threshold policy (one that acts directly on the true, rather than corrected headcount), acts forcefully to bring the allocations toward their optimally targeted levels. The mathematical benefit of the corrected headcount becomes evident in the relaxed network.

Algorithm 2 The policy $\widetilde{\pi}$ in the relaxed network

Require: Threshold constants δ_i for every less-preferred customer type $i \in \mathcal{A}_{lp}$ (same as those of CHT).

1: Accept all preferred customer types $i \in \mathcal{A}_p$ (even if it violates any resource constraint).

2: Accept an arriving type $i \in \mathcal{A}_{lp}$ if and only if the corrected headcount, $\widetilde{\Sigma}^{*,t}_{j_i} = \widetilde{\Sigma}^t_{j_i} + \widetilde{Z}^{*,t}_{j_i}$, with resource j_i has

$$q^N_{j_i} - \widetilde{\Sigma}^{*,t}_{j_i} \geq R_i = \delta_i \log N.$$

Take the network in Figure 3 and recall the policy implementation as spelled out in Figure 6. In the relaxed network, requests of preferred types, in particular of type 7, face no resource constraints — all are accepted and, in turn, their number-in-system evolves like an infinite server queue and is straightforward to analyze; the same is true for type 6. Consider now type-3 requests; these are matched with resource c ($j_i = c$). We note two things: (i) in the auxiliary system, their acceptance/rejection does not depend on the real-time headcount of resource b which they also require, and (ii) the corrected headcount of their matched resource c is given by $\widetilde{\Sigma}^*_c = \widetilde{X}^t_3 + \widetilde{X}^t_6 + \widetilde{X}^{*,t}_1$. Here, less-preferred type 1 appears only through its targeted value $\widetilde{X}^*_1 = q^N_d - \widetilde{X}^t_7$ where $\widetilde{X}^t_7$, as we just explained, is straightforward to analyze; so is $\widetilde{X}^t_6$.

In this way, the analysis can be *localized to each pair of resource j and its coupled request type i_j*. Through this localized analysis we show that, *in the relaxed network*,

$$(A\widetilde{X})_j = \widetilde{\Sigma}_j \approx \widetilde{\Sigma}^*_j \approx q^N_j - \delta_{i_j} \log N \leq q^N_j,$$

so that, while the auxiliary network has the large state space $\widetilde{\mathcal{X}}$, it effectively remains within the smaller state space $\mathcal{X}$ of the original network. Because the policies CHT and $\tilde{\pi}$ take the same actions in states within $\mathcal{X}$ the auxiliary network and policy—while easier to analyze—capture (approximately) the behavior of the original network and CHT.

Crucially, the auxiliary network is hardly a relaxation from a performance perspective. The reward collected by CHT in the original network and of $\widetilde{\pi}$ in the relaxed network are, as we will prove, close to each other. In this way, we are able to bound the performance of CHT in the network by analyzing a simpler one.

Proof Steps

Step 1. In the *relaxed* network, the gap between the performance of $\widetilde{\pi}$ and the (LP) optimal value is logarithmic in N. That is,

$$r'_\mu \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}] = \overline{\mathcal{R}}(q^N, \lambda^N/\mu) + \mathcal{O}(\log N).$$

Step 2. In the *original* network, the Markov chain X^t, induced by CHT, has a $1/N$-mixing time that is polynomial in N. This would imply that for all time t greater than $t_0 = t_0(N)$ that is polynomial in N, and regardless of the initial (at $t = 0$) state x, the distribution of X^t is close to the steady-state distribution Π,

$$r'_\mu \mathbb{E}_x[X^{t_0}] = r'_\mu \mathbb{E}_\Pi[X] + \mathcal{O}(1), \text{ for all } t \geq t_0.$$

Step 3. "Late decoupling": X^t and $\widetilde{X}^t$, suitably initialized (at $t = 0$)[2] with the stationary distribution $\widetilde{\Pi}$ induced by $\widetilde{\pi}$, decouple later than the mixing time of X^t. That is, with high probability, $X^t = \widetilde{X}^t$ until a time $t_1 > t_0$ (with t_0 as in the previous item).

At time t_0—where X^t is close to its stationary distribution Π—it is also equal to $\widetilde{X}^t$ which has (and is initialized at time $t = 0$ with) the stationary distribution $\widetilde{\Pi}$ induced by $\widetilde{\pi}$. This implies that the two stationary distributions—of CHT and of $\widetilde{\pi}$ are close—and so are their collected rewards:

$$\mathcal{R}^{CHT} = r'_\mu \mathbb{E}_\Pi[X] \overset{\text{Step 2}}{\approx} r'_\mu \mathbb{E}_{\widetilde{\Pi}}[X^{t_0}] \overset{\text{Step 3}}{\approx} r'_\mu \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}^{t_0}] \overset{\text{Stationarity}}{=} r'_\mu \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}] \overset{\text{Step 1}}{\approx} \overline{\mathcal{R}}(q^N, \lambda^N/\mu) + O(\log N).$$

Sections 5.1-5.3 formalize steps 1 through 3. These are combined in Section 5.4 to obtain the upper bound in Theorem 4.1.

Throughout the proofs, we use $\eta_l, l = 1, 2, \ldots$ to denote strictly positive constants that do not depend on N. The exact values of these constants might change from one proof to the next.

5.1. Step 1: The Performance of the Relaxed Policy $\widetilde{\pi}$

We use symbols such as $\widetilde{X}, \widetilde{\Sigma}$ to denote quantities associated with the relaxed network and the relaxed policy $\widetilde{\pi}$. Recall that $\mathcal{X}$ and $\widetilde{\mathcal{X}}$ are the state space of CHT and the relaxed policy, $\widetilde{\pi}$ respectively in (6) and (7).

Proposition 5.1 shows that the stationary reward collected by the relaxed policy $\widetilde{\pi}$ is logarithmically close to the optimal value $\overline{\mathcal{R}}$ from (LP). The second statement of the proposition stipulates that, while allowing for a larger state space, the process $\widetilde{X}^t$ spends most of its time in the smaller state space $\mathcal{X}$ of X^t.

Proposition 5.1 (Performance of $\widetilde{\pi}$.) *The Markov chain $\widetilde{X}^t$ has a unique stationary distribution $\widetilde{\Pi}$ that satisfies*

$$r'_\mu \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}] = \overline{\mathcal{R}}(q^N, \lambda^N/\mu) + \mathcal{O}(\log N) \qquad \text{and} \qquad \widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X}) = \mathcal{O}\left(N^{-m\delta_{min}}\right),$$

where $\delta_{min} = \min_i \delta_i$ and $m > 0$ does not depend on δ and N.

[2] Because $\widetilde{X}^t$ has a strictly larger state space than X^t the latter cannot be strictly speaking initialized with $\widetilde{\Pi}$; this initialization is formalized in Section 5.4.

We use two preliminary lemmas in the proof of Proposition 5.1. We start with the observation that, for $i \in \mathcal{A}_p$, $\widetilde{X}_i^t$ has the law of an $M/M/\infty$ queue and, in particular, its stationary distribution is Poisson with mean ${}^{\lambda_i^N}/{}_{\mu_i}$; the probability bound in the next lemma follows from standard concentration results for Poisson random variables.

Lemma 5.1 (Concentration of preferred types) *The Markov chain $\widetilde{X}^t$ has a unique stationary distribution $\widetilde{\Pi}$ and, for every fixed $\epsilon > 0$,*

$$\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{X} \notin \Omega_\epsilon^N\right\} \leq m_1 e^{-m_2 N}, \text{ and } \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}_i] = {}^{\lambda_i^N}/{}_{\mu_i}, \text{ for } i \in \mathcal{A}_p,$$

where

$$\Omega_\epsilon^N := \left\{x \in \widetilde{\mathcal{X}} : \left| x_i - {}^{\lambda_i^N}/{}_{\mu_i} \right| \leq \epsilon N \text{ for all } i \in \mathcal{A}_p\right\}, \tag{8}$$

and $m_1, m_2 > 0$ may depend on ϵ but not on N.

The (random) target allocations $(\widetilde{X}_i^*, i \in \mathcal{A}_{lp})$, defined in terms of $\widetilde{X}_{\mathcal{A}_p}$, inherit properties from this lemma. By construction, $\widetilde{X}^* = y^* N + A_{lp}^{-1} A_p[(\lambda^N/\mu)_{\mathcal{A}_p} - \widetilde{X}_{\mathcal{A}_p}]$, so by Lemma 5.1,

$$\mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}_i^*] = y_i^* N. \tag{9}$$

For each $j \in [d]$, define the function

$$\begin{aligned} \sigma_j(x) &= \sum_{k \in \mathcal{A}(j)} x_k + \sum_{i \in \mathcal{A}_{lp} \setminus i_j} A_{ji}(x_i^* - x_i) \\ &= x_{i_j} + \sum_{l \in \mathcal{A}_p(j)} x_l + \sum_{l \in \mathcal{A}_{lp}(j) \setminus i_j} x_l^*, \end{aligned} \tag{10}$$

where we recall $x_{\mathcal{A}_{lp}}^*(x) = y_{\mathcal{A}_{lp}}^* N + A_{lp}^{-1} A_p[(\lambda^N/\mu)_{\mathcal{A}_p} - x_{\mathcal{A}_p}]$; see equation (4). In particular, $\widetilde{\Sigma}_j^{*,t} = \sigma_j(\widetilde{X}^t)$. We define also the functions

$$f_i(x) = |\sigma_{j_i}(x) - \widehat{q}_{j_i}^N|, \text{ and } g_i^\theta(x) = \exp(\theta f_i(x)), \tag{11}$$

where we use the shorthand notation

$$\widehat{q}_j^N = q_j^N - \delta_{i_j} \log N.$$

Lemma 5.2 *There exists $\theta > 0$ (not depending on N) such that for all $i \in \mathcal{A}_{lp}$*

$$\sum_{x \in \widetilde{\mathcal{X}}} \widetilde{\Pi}(x) |\mathcal{Q}(x,x)| |g_i^\theta(x)| < \infty,$$

where $\mathcal{Q}$ is the transition-rate matrix of the Markov chain $\widetilde{X}^t$. In turn,

$$\mathbb{E}_{\widetilde{\Pi}}[(\mathcal{Q} g_i^\theta)(\widetilde{X})] = 0.$$

Lemma 5.3 *Under the assumptions of Theorem 4.1 there exists* $\theta, m_1, m_2 > 0$ *(that do not depend on* N *but may depend on* ϵ*) such that*

$$(\mathcal{Q}g_i^\theta)(x) \le -2m_1 N g_i^\theta(x) + m_2 N, \text{ for all } x \in \Omega_\epsilon^N.$$

Lemma 5.3 shows that the process $g(\widetilde{X}^t)$ has a centering drift property; when $g(\widetilde{X}^t)$ is large (meaning $\sigma_{j_i}(\widetilde{X}^t)$ is far from $\widehat{q}_{j_i}^N$) it decreases in expectation ($\sigma_{j_i}(\widetilde{X}^t)$ is pushed back towards $\widehat{q}_{j_i}^N$).

Proof of Proposition 5.1. We first show that, for each i,

$$\mathbb{P}_{\widetilde{\Pi}}\left\{|\widetilde{X}_i^* - \widetilde{X}_i| \ge \delta_i \log N + x\right\} \le \eta_1 e^{-\eta_2 x}, \text{ and } \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}_i] = y_i^* N + \mathcal{O}(\log N). \tag{12}$$

Fix $i \in \mathcal{A}_{lp}, \theta > 0$ and let $f(x)$ and $g(x)$ (with the corresponding subscript/superscript dropped) be as in (11) for this fixed i, θ. Recall that $\mathcal{Q}$ stands for the infinitesimal generator (the transition-rate matrix) of the n-dimensional process $\widetilde{X}^t$. By Lemma 5.2

$$0 = \mathbb{E}_{\widetilde{\Pi}}\left[(\mathcal{Q}g)(\widetilde{X})\right] = \mathbb{E}_{\widetilde{\Pi}}\left[(\mathcal{Q}g)(X)\mathbb{1}\left\{\widetilde{X} \in \Omega_\epsilon^N\right\}\right] + \mathbb{E}_{\widetilde{\Pi}}\left[(\mathcal{Q}g)(\widetilde{X})\mathbb{1}\left\{\widetilde{X} \notin \Omega_\epsilon^N\right\}\right]. \tag{13}$$

Lemma 5.3 immediately implies that

$$\mathbb{E}_{\widetilde{\Pi}}\left[(\mathcal{Q}g)(\widetilde{X})\mathbb{1}\left\{\widetilde{X} \in \Omega_\epsilon^N\right\}\right] \le -\eta_1 N \mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\mathbb{1}\left\{\widetilde{X} \in \Omega_\epsilon^N\right\}\right] + \eta_2 N, \tag{14}$$

and we turn to consider the second term on the right-hand side of (13). By Hölder's inequality

$$\mathbb{E}_{\widetilde{\Pi}}\left[|(\mathcal{Q}g)(\widetilde{X})|\mathbb{1}\left\{\widetilde{X} \notin \Omega_\epsilon^N\right\}\right] \le \sqrt{\mathbb{E}_{\widetilde{\Pi}}\left[((\mathcal{Q}g)(\widetilde{X}))^2\right]}\sqrt{\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{X} \notin \Omega_\epsilon^N\right\}}.$$

The transition-rate matrix satisfies $|\mathcal{Q}(x,y)| \le \sum_{l\in[n]}(\lambda_l^N + \mu_l x_l)$ and $\mathcal{Q}(x,y) \neq 0$ only for $y = x$ or $|y - x| = e_l$ for some $l \in [n]$. Recalling (10), and that $x_i^*(\zeta) = y_i^* N + \sum_{l\in\mathcal{A}_p} \alpha_i^l(\zeta_l - \lambda_l^N/\mu_l)$, where $\alpha_i^l = [A_{lp}^{-1}A_p]_{i,l},\ i \in \mathcal{A}_{lp}, l \in \mathcal{A}_p$, we have that

$$\sigma_{j_i}(x) \le dq_{max}N + (1 \vee \bar{\alpha}) \sum_{k\in\mathcal{A}_p} x_l, \text{ and } |\sigma_{j_i}(x + e_l) - \sigma_{j_i}(x)| \le 1 \vee \bar{\alpha},$$

where $\bar{\alpha} = \max_{i,l}|\alpha_{i,l}|$, and we use the fact that $x_i \le q_{\max}N$ for all $i \in \mathcal{A}_{lp}$, with $q_{\max} := \max_j q_j$. Combining these we have that

$$\begin{aligned} |(\mathcal{Q}g)(x)| &\le \sum_y |Q(x,y)||g(y)| \le e^{\theta(\sigma_{j_i}(x)+1\vee\bar{\alpha})} \sum_{l\in[n]}(\lambda_l^N + \mu_l x_l) \\ &\le \eta_3 e^{\theta(dq_{\max}N + (1\vee\bar{\alpha})\sum_{l\in\mathcal{A}_p} x_l + 1\vee\bar{\alpha})}(N + \sum_{l\in\mathcal{A}_p} x_l). \end{aligned}$$

In particular,

$$((\mathcal{Q}g)(x))^2 \leq \eta_4 e^{2\theta(dq_{\max}N+(1\vee\bar{\alpha})\sum_{l\in\mathcal{A}_p}x_l+1\vee\bar{\alpha})}(N^2+\sum_{l\in\mathcal{A}_p}x_l^2).$$

We recall that under the steady-state distribution $\widetilde{\Pi}$, $\widetilde{X}_l \sim \text{Poisson}(\lambda_l/\mu_l)$, $l\in\mathcal{A}_p$ (and $\widetilde{X}_l, l\in\mathcal{A}_p$ are independent). Using Hölder's inequality we have for θ small enough (not dependent on N),

$$\mathbb{E}_{\widetilde{\Pi}}\left[\sum_{l\in\mathcal{A}_p}(\widetilde{X}_l)^2 e^{2\theta\sum_{l\in\mathcal{A}_p}\widetilde{X}_l}\right] \leq \eta_4 N^2 e^{2\theta\kappa N},$$

where $\kappa=\sum_{l\in\mathcal{A}_p}\frac{\lambda_l}{\mu_l}$ (recall $\lambda_l^N=\lambda_l N$). Choosing θ such that $2\theta\kappa\leq m_2/2$ (with m_2 as in Lemma 5.1), we have

$$\mathbb{E}_{\widetilde{\Pi}}\left[|(\mathcal{Q}g)(\widetilde{X})|\mathbb{1}\left\{\widetilde{X}\notin\Omega_\epsilon^N\right\}\right] \leq \sqrt{\mathbb{E}_{\widetilde{\Pi}}\left[((\mathcal{Q}g)(\widetilde{X}))^2\right]}\sqrt{\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{X}\notin\Omega_\epsilon^N\right\}} \leq \eta_5. \tag{15}$$

Plugging (14) and (15) into (13) we have that

$$-\eta_5 \leq \mathbb{E}_{\widetilde{\Pi}}\left[(\mathcal{Q}g)(\widetilde{X})\mathbb{1}\left\{\widetilde{X}\in\Omega_\epsilon^N\right\}\right] \leq -\eta_6 N\mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\mathbb{1}\left\{\widetilde{X}\in\Omega_\epsilon^N(i)\right\}\right]+\eta_7 N,$$

so that

$$\mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\mathbb{1}\left\{\widetilde{X}\in\Omega_\epsilon^N(i)\right\}\right] \leq \frac{\eta_5+\eta_7 N}{\eta_6 N} \leq \eta_8 = \mathcal{O}(1).$$

Recall that $g(x)=e^{\theta|\sigma_{j_i}(x)-\widehat{q}_{j_i}^N|}\leq e^{\theta(\sigma_{j_i}(x)+q_{j_i}N)}$. Following a similar argument to those leading to (15) we can conclude that $\mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\mathbb{1}\left\{\widetilde{X}\notin\Omega_\epsilon^N(i)\right\}\right]\leq\eta_9$ so that

$$\mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\right] = \mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\mathbb{1}\left\{\widetilde{X}\in\Omega_\epsilon^N(i)\right\}\right]+\mathbb{E}_{\widetilde{\Pi}}\left[g(\widetilde{X})\mathbb{1}\left\{\widetilde{X}\notin\Omega_\epsilon^N(i)\right\}\right] \leq \eta_{10}.$$

By Markov's inequality

$$\mathbb{P}_{\widetilde{\Pi}}\{|\sigma_{j_i}(\widetilde{X})-\widehat{q}_{j_i}^N|\geq x\} \leq \mathbb{E}_{\widetilde{\Pi}}[g(\widetilde{X})]e^{-\theta x} \leq \eta_{10}e^{-\theta x}. \tag{16}$$

By the definition of $\widetilde{X}_k^*$ (recall (3) and (5)) $\sum_{k\in\mathcal{A}_p(j_i)}X_k^t+\sum_{k\in\mathcal{A}_{lp}(j_i)}X_k^{*,t}=q_{j_i}^N$ so that

$$\begin{aligned}
\widetilde{X}_i^* &= q_{j_i}^N - \sum_{k\in\mathcal{A}_p(j_i)}\widetilde{X}_k - \sum_{k\in\mathcal{A}_{lp}(j_i)\setminus i}\widetilde{X}_k^* \\
&= q_{j_i}^N - \sigma_{j_i}(\widetilde{X}) + \widetilde{X}_i \\
&= \widehat{q}_{j_i}^N + \delta_i\log N - \sigma_{j_i}(\widetilde{X}) + \widetilde{X}_i,
\end{aligned}$$

and $|\widetilde{X}_i-\widetilde{X}_i^*+\delta_i\log N| = |\sigma_{j_i}(\widetilde{X})-\widehat{q}_{j_i}^N|$ and

$$\mathbb{P}_{\widetilde{\Pi}}\left\{\left|\widetilde{X}_i-\widetilde{X}_i^*+\delta_i\log N\right|\geq x\right\} = \mathbb{P}_{\widetilde{\Pi}}\{|\sigma_{j_i}(\widetilde{X})-\widehat{q}_{j_i}^N|\geq x\} \leq \eta_{11}e^{-\theta x}, \tag{17}$$

which implies

$$\mathbb{P}_{\widetilde{\Pi}}\left\{\left|\widetilde{X}_i-\widetilde{X}_i^*\right|\geq\delta_i\log N+x\right\}\leq\eta_{12}e^{-\theta x}. \tag{18}$$

In turn,

$$\mathbb{E}_{\widetilde{\Pi}}\left[\left|\widetilde{X}_i-\widetilde{X}_i^*\right|\right]\leq\eta_{13}\log N. \tag{19}$$

By Jensen's inequality we also have that

$$\mathbb{E}_{\widetilde{\Pi}}\left[f(\widetilde{X})\right]=\mathbb{E}_{\widetilde{\Pi}}\left[\left|\sigma_{j_i}(\widetilde{X})-\widehat{q}_{j_i}^N\right|\right]\leq\log\mathbb{E}_{\widetilde{\Pi}}[g(\widetilde{X})]\leq\log\eta_{14}.$$

By (9), $\mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}_i^*]=y_i^*N$ and we conclude from (19) that

$$\mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}_i]=y_i^*N+\mathcal{O}(\log N).$$

Repeating this argument for all $i\in\mathcal{A}_{lp}$ we have

$$\mathbb{E}_{\widetilde{\Pi}}\left[\widetilde{X}_i\right]=y_i^*N+\mathcal{O}(\log N), i\in[n],$$

with $y_i^*=\lambda_i/\mu_i$ for $i\in\mathcal{A}_p$, and we arrive at the first statement of the proposition, namely that

$$r'_\mu\mathbb{E}_{\widetilde{\Pi}}\left[\widetilde{X}\right]=\overline{\mathcal{R}}\left(q^N,\lambda^N/\mu\right)+\mathcal{O}(\log N).$$

We turn to the second statement of the proposition, namely that $\widetilde{\Pi}(\widetilde{\mathcal{X}}\backslash\mathcal{X})=\mathcal{O}\left(N^{-\eta\delta_{min}}\right)$. Recall that the true (not corrected) headcount satisfies

$$\begin{aligned}\widetilde{\Sigma}_{j_i}&=\sigma_{j_i}(\widetilde{X})-\widetilde{\mathcal{Z}}_{j_i}^*=\sigma_{j_i}(\widetilde{X})-\sum_{k\in\mathcal{A}_{lp}(j_i)\backslash i}(\widetilde{X}_k^*-\widetilde{X}_k)\\&=\sigma_{j_i}(\widetilde{X})-\sum_{k\in\mathcal{A}_{lp}(j_i)\backslash i}\delta_k\log N-\sum_{k\in\mathcal{A}_{lp}(j_i)\backslash i}(\widetilde{X}_k^*-\widetilde{X}_k-\delta_k\log N),\end{aligned}$$

Let $\bar{\delta}_{j_i}=\sum_{k\in\mathcal{A}_{lp}(j_i)\backslash i}\delta_k$. Then,

$$|\widetilde{\Sigma}_{j_i}-\sigma_{j_i}(\widetilde{X})+\bar{\delta}_{j_i}\log N|\leq\sum_{i\in\mathcal{A}_{lp}}|\widetilde{X}_i^*-\widetilde{X}_i-\delta_i\log N|.$$

Using (17), a union bound gives

$$\mathbb{P}_{\widetilde{\Pi}}\{|\widetilde{\Sigma}_{j_i}-\sigma_{j_i}(\widetilde{X})+\bar{\delta}_{j_i}\log N|>x\}\leq\eta_{15}e^{-\eta_{16}x}.$$

Note that $\widetilde{X}\notin\mathcal{X}$ if and only if there exists i such that $\widetilde{\Sigma}_{j_i}>\widehat{q}_{j_i}^N+\delta_i\log N=q_{j_i}^N$. Also,

$$\begin{aligned}\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{\Sigma}_{j_i}>\widehat{q}_{j_i}^N+\delta_i\log N\right\}\leq&\mathbb{P}_{\widetilde{\Pi}}\left\{|\widetilde{\Sigma}_{j_i}-\sigma_{j_i}(\widetilde{X})+\bar{\delta}_{j_i}\log N|>\frac{\delta_i}{2}\log N\right\}\\&+\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{\Sigma}_{j_i}>\widehat{q}_{j_i}^N+\delta_i\log N,|\widetilde{\Sigma}_{j_i}-\sigma_{j_i}(\widetilde{X})+\bar{\delta}_{j_i}\log N|\leq\frac{\delta_i}{2}\log N\right\}.\end{aligned}$$

On the event $\{|\widetilde{\Sigma}_{j_i} - \sigma_{j_i}(\widetilde{X}) + \bar{\delta}_{j_i}\log N| \le \frac{\delta_i}{2}\log N\}$, $\widetilde{\Sigma}_{j_i} > \widehat{q}_{j_i}^N + \delta_i \log N$ implies that $\sigma_{j_i}(\widetilde{X}) > \widehat{q}_{j_i}^N + \bar{\delta}_{j_i}\log N + \frac{\delta_i}{2}\log N$. We then have, using (16), that

$$\begin{aligned}\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{X}\notin\mathcal{X}\right\} &\le \sum_{i\in\mathcal{A}_{lp}}\mathbb{P}_{\widetilde{\Pi}}\left\{\widetilde{\Sigma}_{j_i} > \widehat{q}_{j_i}^N + \delta_i\log N\right\}\\ &\le \sum_{i\in\mathcal{A}_{lp}}\mathbb{P}_{\widetilde{\Pi}}\left\{|\widetilde{\Sigma}_{j_i} - \sigma_{j_i}(\widetilde{X}) + \bar{\delta}_{j_i}\log N| > \frac{\delta_i}{2}\log N\right\}\\ &\quad + \mathbb{P}_{\widetilde{\Pi}}\left\{\sigma_{j_i}(\widetilde{X}) > \widehat{q}_{j_i}^N + \bar{\delta}_{j_i}\log N + \frac{\delta_i}{2}\log N\right\} \le \eta_{17}N^{-\eta_{18}\underline{\delta}},\end{aligned}$$

with $\underline{\delta} = \min_i \delta_i$. □

5.2. Step 2: The Mixing Time Bound of X^t under CHT

In this second step, we return to the original (vs. the auxiliary) Markov chain X^t—on the state space $\mathcal{X}$, and induced by CHT—and study its mixing time. Throughout this subsection, X^t refers to this chain.

The mixing time of a continuous-time Markov chain M^t on a finite state space $\mathcal{M}$ and that has a steady-state distribution Π is defined as

$$\tau_{mix}(\epsilon) = \min\left\{t>0: \left|\frac{h_t(x,y)-\Pi(y)}{\Pi(y)}\right| \le \epsilon \text{ for all } x,y\in\mathcal{M}\right\},$$

where $h_t(x,y) = \mathbb{P}_x\{M^t = y\}$ is the transition kernel and it is given by the matrix exponential $h_t = e^{t\mathcal{Q}}$, where $\mathcal{Q}$ is the transition-rate matrix.

Lemma 5.4 (Morris and Peres (2005, Theorem 13).) *Let M^t be a Markov chain on a finite state space $\mathcal{M}$, with generator matrix $\mathcal{Q} = (\mathcal{Q}_{xy})_{x,y\in\mathcal{M}}$, and steady-state distribution Π. Define the uniformization constant $\mathbb{U} \equiv \max_{x\in\mathcal{M}}\sum_{y\neq x}\mathcal{Q}_{xy}$ and the (discrete-time) transition probability matrix $P = \mathbb{U}^{-1}\mathcal{Q} + I$. Define also the conductance of P as*

$$\Phi = \min_{S\subset\mathcal{M},\Pi(S)\le 1/2}\frac{\sum_{x\in S,y\in S^c}P_{xy}}{\Pi(S)}.$$

Then the mixing time of M^t satisfies

$$\tau_{mix}(\epsilon) \le \mathbb{U}\left\lceil 1 + \frac{8}{\Phi^2}\log\frac{1}{\epsilon\Pi_*}\right\rceil, \tag{20}$$

where $\Pi_* = \min_{x\in\mathcal{X}}\Pi(x)$.

For the uniformization constant for X^t (operated under CHT) we take

$$\mathbb{U} = N\sum_{i\in[n]}\lambda_i + \max_{x\in\mathcal{X}}\sum_{i\in[n]}\mu_i x_i. \tag{21}$$

Lemma 5.5 *Let* Π *be the stationary distribution of* X^t*, let* $\mathcal{Q}$ *be its transition rate matrix and* $P = \mathbb{U}^{-1}\mathcal{Q} + I$ *with* $\mathbb{U}$ *in* (21)*. There then exist constants* $\bar{c}, K > 0$ *(not depending on* N*) such that*

$$\Pi_* = \min_{x\in\mathcal{X}} \Pi(x) \geq (\bar{c}N)^{-KN}, \tag{22}$$

and

$$\Phi = \min_{S\subset\mathcal{X},\Pi(S)\leq 1/2} \frac{\sum_{x\in S, y\in S^c} P_{xy}}{\Pi(S)} \geq \frac{1}{\bar{c}N}. \tag{23}$$

Armed with this lemma, we bound the mixing time of the Markov chain X^t in terms of the scaling factor N. The bound is given in the following proposition.

Proposition 5.2 (Mixing time bound for X^t) *The mixing time of* X^t *satisfies*

$$\tau_{mix}\left(N^{-1}\right) = \mathcal{O}\left(N^4 \log N\right).$$

Proof: Under CHT, $X^t = x$ is subject to the resource constraint $Ax \leq q^N$. Because $\mathcal{S}(i) \neq \emptyset, i \in [n]$, $x_i \leq \max_{j\in[d]} q_j^N$, so that the uniformization constant satisfies

$$\mathbb{U} = N\sum_{i\in[n]} \lambda_i + \max_{x\in\mathcal{X}} \sum_{i\in[n]} \mu_i x_i \leq \sum_{i\in[n]} (\lambda_i + \mu_i q_{\max})N = \mathcal{O}(N),$$

where $q_{\max} = \max_{j\in[d]} q_j$. Taking $\epsilon = 1/N$ in Lemma 5.4 and plugging there also the lower bounds for Φ and Π^* from Lemma 5.5 we obtain

$$\tau_{mix}(N^{-1}) \leq \mathbb{U}\left\lceil 1 + \frac{8}{(\bar{c}N)^{-2}} \log \frac{1}{N^{-1}(\bar{c}N)^{-KN}} \right\rceil = \mathcal{O}\left(N^4 \log N\right),$$

as needed. □

5.3. Step 3: The Decoupling Time of X^t and $\widetilde{X}^t$

The two chains X^t and $\widetilde{X}^t$ are easily constructed on the same sample space so that, having the same initial state at $t = 0$, $X^t = \widetilde{X}^t$ up to the time at which $\widetilde{X}^t$ leaves $\mathcal{X}$. We will bound the tail probability of this exit time.

Proposition 5.3 (Tail-probability bound.) *The time that* $\widetilde{X}^t$ *exits* $\mathcal{X}$

$$\widetilde{\tau} := \min\left\{t \geq 0 : \widetilde{X}^t \in \widetilde{\mathcal{X}} \backslash \mathcal{X}\right\}$$

satisfies

$$\mathbb{P}_{\widetilde{\Pi}}\{\widetilde{\tau} \leq t\} \leq m_1 N t \left(\widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X}) + \exp(-m_2 N)\right),$$

for some constants $m_1, m_2 > 0$ *that do not depend on* N*.*

Proof: We note that $\widetilde{\tau} \leq t$ if and only if the number of visits of $\widetilde{X}^t$ to the set $\mathcal{E} := \widetilde{\mathcal{X}} \backslash \mathcal{X}$ by time t is greater than, or equal to, 1. For any $x \in \widetilde{\mathcal{X}}$, let $N^t(x)$ be the number of visits of $\widetilde{X}^t$ to that state by time t. For a subset $\mathcal{C} \subseteq \widetilde{\mathcal{X}}$, let $N^t(\mathcal{C}) = \sum_{x \in \mathcal{A}} N^t(x)$. Then,

$$\{\widetilde{\tau} \leq t\} = \{N^t(\mathcal{E}) \geq 1\}.$$

By Markov Inequality,

$$\mathbb{P}_{\widetilde{\Pi}}\{\widetilde{\tau} \leq t\} = \mathbb{P}_{\widetilde{\Pi}}\{N^t(\mathcal{E}) \geq 1\} \leq \mathbb{E}_{\widetilde{\Pi}}[N^t(\mathcal{E})] = \sum_{x \in \mathcal{E}} \mathbb{E}_{\widetilde{\Pi}}[N^t(x)].$$

It is a basic fact of continuous-time Markov chains that $\mathbb{E}_{\widetilde{\Pi}}[N^t(x)] = \mathfrak{e}(x)\widetilde{\Pi}(x)t$ where $\mathfrak{e}(x)$ is the total exit rate from state x.

Recalling that $\widetilde{\mathcal{X}} = \left\{x \in \mathbb{N}^n : x_i \leq q_{j_i}^N - \delta_i \log N \text{ for all } i \in \mathcal{A}_{lp}\right\}$, we have

$$\mathfrak{e}(x) \leq \sum_{i \in [n]} \lambda_i^N + \sum_{i \in \mathcal{A}_p} \mu_i x_i + \sum_{i \in \mathcal{A}_{lp}} \mu_i q_{j_i}^N \leq \eta_1 N + \bar{\mu} \sum_{i \in \mathcal{A}_p} x_i, \; x \in \widetilde{\mathcal{X}},$$

where $\bar{\mu} := \max_i \mu_i$. Defining

$$\mathcal{B} := \left\{ x \in \widetilde{\mathcal{X}} : \sum_{i \in \mathcal{A}_p} x_i \leq 2 \sum_{i \in \mathcal{A}_p} \lambda_i^N / \mu_i \right\},$$

we have

$$\mathfrak{e}(x) \leq \eta_1 N + 2\bar{\mu} \sum_{i \in \mathcal{A}_p} \lambda_i^N / \mu_i \leq \eta_2 N, \; x \in \mathcal{E} \cap \mathcal{B}.$$

Then,

$$\begin{aligned} \mathbb{P}_{\widetilde{\Pi}}\{\widetilde{\tau} \leq t\} &\leq \sum_{x \in \mathcal{E}} \mathbb{E}_{\widetilde{\Pi}}[N^t(x)] = \sum_{x \in \mathcal{E} \cap \mathcal{B}} \mathfrak{e}(x)\widetilde{\Pi}(x)t + \sum_{x \in \mathcal{E} \backslash \mathcal{B}} \mathfrak{e}(x)\widetilde{\Pi}(x)t \\ &\leq \eta_3 N \widetilde{\Pi}(\mathcal{E})t + \eta_4 (N + \sum_{x \in \mathcal{E} \backslash \mathcal{B}} |x_{\mathcal{A}_p}| \widetilde{\Pi}(x)t) \\ &= \eta_3 N \widetilde{\Pi}(\mathcal{E})t + \eta_4 t \mathbb{E}_{\widetilde{\Pi}} \left[(N + |\widetilde{X}_{\mathcal{A}_p}|) \mathbb{1} \left\{ |\widetilde{X}_{\mathcal{A}_p}| > 2|(\lambda^N/\mu)_{\mathcal{A}_p}| \right\} \right], \end{aligned} \tag{24}$$

where $|\cdot|$ is the L_1 norm. We turn to bound the second term in the last row. Under $\widetilde{\Pi}$, recall, $(\widetilde{X}_i, i \in \mathcal{A}_p)$ are independent Poisson random variables with mean λ_i^N/μ_i for $i \in \mathcal{A}_p$. In turn, $|\widetilde{X}_{A_p}| \sim Poisson\Big((\lambda^N/\mu)_{\mathcal{A}_p}\Big)$. By Hölder's inequality

$$\begin{aligned} \mathbb{E}_{\widetilde{\Pi}} \left[(N + |\widetilde{X}_{\mathcal{A}_p}|) \mathbb{1} \left\{ |\widetilde{X}_{\mathcal{A}_p}| > 2|(\lambda^N/\mu)_{\mathcal{A}_p}| \right\} \right] &\leq \sqrt{\mathbb{E}_{\widetilde{\Pi}} \left[(N + |\widetilde{X}_{\mathcal{A}_p}|)^2 \right]} \sqrt{\mathbb{P}_{\widetilde{\Pi}}\{ |\widetilde{X}_{\mathcal{A}_p}| > 2|(\lambda^N/\mu)_{\mathcal{A}_p}| \}} \\ &\leq \eta_4 N e^{-\eta_5 |(\lambda^N/\mu)_{\mathcal{A}_p}|} \leq \eta_6 e^{-\eta_7 N}, \end{aligned} \tag{25}$$

where we used the second moment of the Poisson random variable $\mathbb{E}[|\widetilde{X}_{\mathcal{A}_p}|^2] = |(\lambda^N/\mu)_{\mathcal{A}_p}| + |(\lambda^N/\mu)_{\mathcal{A}_p}|^2$, and standard tail bounds for the Poisson distribution.

Plugging (25) into (24) we conclude that

$$\mathbb{P}_{\widetilde{\Pi}}\{\widetilde{\tau} \leq t\} \leq \eta_8 N t(\widetilde{\Pi}(\mathcal{E}) + e^{-\eta_9 N}),$$

as required. □

5.4. Combining the Steps

The process X^t has the state space

$$\begin{aligned}\mathcal{X} = \left\{x \in \mathbb{N}^n : Ax \leq q^N\right\} \cap \left\{x \in \mathbb{N}^n : x_i \leq q_{j_i}^N - \delta_i \log N \text{ for all } i \in \mathcal{A}_{lp}\right\} \\ \subset \widetilde{\mathcal{X}} = \left\{x \in \mathbb{N}^n : x_i \leq q_{j_i}^N - \delta_i \log N \text{ for all } i \in \mathcal{A}_{lp}\right\}.\end{aligned}$$

It cannot be initialized with the stationary distribution of $\widetilde{X}$ as the latter may assign positive probabilities to states outside of $\mathcal{X}$. To that end, fix $x_0 \in \mathcal{X}$ and define the distribution Π^0 on $\mathcal{X}$ by

$$\Pi^0(x) = \begin{cases} \widetilde{\Pi}(x) & \text{if } x \neq x_0, \\ \widetilde{\Pi}(x_0) + \widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X}) & \text{if } x = x_0. \end{cases}$$

Then, $\mathbb{E}_{\Pi^0}[X^t] = \sum_{x \in \mathcal{X}} \widetilde{\Pi}(x)\mathbb{E}_x[X^t] + \widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X})\mathbb{E}_{x_0}[X^t]$. We define $\mathbb{E}_{\widetilde{\Pi}}[X^t] := \mathbb{E}_{\Pi^0}[X^t]$.

For any $t > 0$,

$$|\mathbb{E}_{\widetilde{\Pi}}[r'_\mu \widetilde{X}^t] - \mathbb{E}_{\Pi}[r'_\mu X^t]| \leq |\mathbb{E}_{\widetilde{\Pi}}[r'_\mu \widetilde{X}^t] - \mathbb{E}_{\widetilde{\Pi}}[r_\mu X^t]| + |\mathbb{E}_{\widetilde{\Pi}}[r'_\mu X^t] - \mathbb{E}_{\Pi}[r'_\mu X^t]|. \tag{26}$$

Recall

$$\widetilde{\tau} := \min\left\{t \geq 0 : \widetilde{X}^t \in \widetilde{\mathcal{X}} \backslash \mathcal{X}\right\}.$$

Because X^t and $\widetilde{X}^t$ have the same transition rates inside $\mathcal{X}$, it is straightforward to build them on the same probability space so that $X^t = \widetilde{X}^t$ up to $\widetilde{\tau}$. We can then write, for any $t \geq 0$,

$$|\mathbb{E}_{\widetilde{\Pi}}[r'_\mu \widetilde{X}^t] - \mathbb{E}_{\widetilde{\Pi}}[r'_\mu X^t]| = |\mathbb{E}_{\widetilde{\Pi}}[r'_\mu(\widetilde{X}^t - X^t)\mathbb{1}\{\widetilde{\tau} \leq t\}]| \leq \sqrt{\mathbb{E}_{\widetilde{\Pi}}[(r'_\mu(\widetilde{X}^t - X^t))^2]}\sqrt{\mathbb{P}\{\widetilde{\tau} \leq t\}}.$$

Because $X^t \in \mathcal{X}$, $\mathbb{E}_x[(r'_\mu X^t)^2] \leq \eta_1 N^2$ for any initial state $x \in \mathcal{X}$. Because, $\widetilde{X}_i^t \leq q_{\max} N$ for all $i \in \mathcal{A}_{lp}$ and, in steady-state, $\widetilde{X}_i^t \sim \text{Poisson}(\lambda_i^n/\mu_i)$ for all $i \in \mathcal{A}_p$, we also have that $\mathbb{E}_{\widetilde{\Pi}}[(r'_\mu \widetilde{X}^t)^2] \leq \eta_2 N^2$. Thus, $\mathbb{E}_{\widetilde{\Pi}}[(r'_\mu(\widetilde{X}^t - X^t))^2] \leq \eta_3 N^2$. Finally, setting $t = \tau(N^{-1})$ we have, by Propositions 5.2 and 5.3, that

$$\mathbb{P}\{\widetilde{\tau} \leq t\} \leq \eta_4 N t\left(\widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X}) + \exp(-\eta_4 N)\right) \leq \eta_5 N^5 \log N \left(\widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X}) + \exp(-\eta_6 N)\right),$$

where for the last inequality we plug $t = \tau(N^{-1}) = \mathcal{O}(N^4 \log N)$. By Proposition 5.1, $\widetilde{\Pi}\left(\widetilde{\mathcal{X}} \backslash \mathcal{X}\right) = \mathcal{O}(N^{-m\delta_{min}})$. We can choose the threshold-coefficients $\delta_i, i \in \mathcal{A}_{lp}$ large enough so that $m\delta_{min} \geq 8$ in which case

$$\begin{aligned}|r'_\mu \mathbb{E}_{\widetilde{\Pi}}[\widetilde{X}^t] - r'_\mu \mathbb{E}_{\widetilde{\Pi}}[X^t]| &= |\mathbb{E}_{\widetilde{\Pi}}[r'_\mu(\widetilde{X}^t - X^t)\mathbb{1}\{\widetilde{\tau} \leq t\}]| \\ &\leq \eta_5 N^7 \log N \left(\widetilde{\Pi}(\widetilde{\mathcal{X}} \backslash \mathcal{X}) + \exp(-\eta_6 N)\right) = \mathcal{O}(1),\end{aligned}$$

which bounds the first summand in (26). For the second summand, we have, by definition, at $t=\tau(N^{-1})$ that $|h_t(x,y)-\Pi(y)|\leq \frac{1}{N}\Pi(y)$ for all $x\in\mathcal{X}$. In particular, at this t,

$$|r_\mu'\mathbb{E}_x[X^t]-r_\mu'\mathbb{E}_\Pi[X]|\leq \frac{1}{N}r_\mu'\mathbb{E}_\Pi[X]=\mathcal{O}(1),$$

where the last equality follows noting that $r_\mu' x\leq (r_\mu)_{\max}q_{\max}dN=\mathcal{O}(N)$ for all $x\in\mathcal{X}$. In turn,

$$|\mathbb{E}_{\widetilde{\Pi}}[r_\mu' X^t]-\mathbb{E}_\Pi[r_\mu X^t]|=|\sum_{x\in\mathcal{X}}\widetilde{\Pi}(x)(\mathbb{E}_x[r_\mu' X^t]-\mathbb{E}_\Pi[r_\mu' X^t])|\leq \max_{x\in\mathcal{X}}|r_\mu'\mathbb{E}_x[X^t]-r_\mu'\mathbb{E}_\Pi[X]|=\mathcal{O}(1).$$

This concludes the proof. □

6. Proof of the lower bound in Theorem 4.1

We note that, because (i) the state space is finite and the action sets—corresponding to admission probabilities—compact, and (ii) the MDP is unichain[3] we can assume, without loss of generality, that—for each N—there exists a stationary deterministic optimal policy; see (see Theorem 11.4.6 of Puterman, 2014). We let $\pi^{*,N}$ be the optimal policy in the N^{th} system, $\Pi^{*,N}$ be the steady-state distribution under this policy, and $\mathcal{R}^{*,N}$ be the optimal reward.

First, in Lemma 6.1, we establish some properties that must be satisfied for sub-logarithmic regret. Specifically, for the optimal policy to have $o(\log N)$ regret it must (i) keep all, up to $o(\log N)$, resource units busy in stationarity, and (ii) it must accept all, up to $o(\log N)$, customers of types $i\in\mathcal{A}_p$.

In step 2, we use Markov Chain analysis to show that a policy that satisfies properties (i) and (ii) must have a stationary distribution that assigns a non-negligible probability to states where all servers are busy. In those states all, in particular preferred, customers are rejected. Because that probability is non-negligible, we will have more $\Omega(\log N)$ customers rejected of types $i\in\mathcal{A}_p$, contradicting the $o(\log N)$ gap.

Lemma 6.1 *Suppose that*

$$\overline{\mathcal{R}}(q^N,\lambda^N/\mu)-\mathcal{R}^{*,N}=o(\log N). \tag{27}$$

Then, $x_i^N=\mathbb{E}_{\Pi^{*,N}}[X_i]$ *must satisfy*

$$Ax^N=q^N-o(\log N), \tag{28}$$

and

$$x_i^N=\frac{\lambda_i^N}{\mu_i}-o(\log N),\ \ i\in\mathcal{A}_p. \tag{29}$$

[3] Because the state space is finite, under any deterministic stationary policy, there is at least one recurrent class. Suppose there are two (or more) recurrent classes. Because from any state $x\in\mathcal{X}$ there is a path–through consecutive service completions—to $0\in\mathcal{X}$ (the empty state), 0 must be part of any recurring class. Hence, there can be at most one recurring class.

Recall that $\Sigma_j = (AX^N)_j$ is the headcount in station j. Markov's inequality applied to (28), gives that, for all $j \in [d]$,

$$\mathbb{P}_{\Pi^{*N}}\left\{q_j^N - \Sigma_j > \kappa \log N\right\} \leq \frac{o(\log N)}{\kappa \log N} \to 0, \text{ as } N \uparrow \infty. \tag{30}$$

We will prove that (30) implies, in fact, that (29) is violated and, in turn, a contradiction to (27).

For the rest of this proof, we fix a resource $j \in [d]$ whose resource constraint is binding and for which there exists a request $i \in \mathcal{A}_p$ with $A_{ji} = 1$; such j and i exist by assumption. We denote this type by i_0 and drop the subscript j from all notations. Let $s(x) = \sum_i A_{ji} x_i$ be the total resource-j headcount in state $x \in \mathcal{X}$. Fix $\kappa > 0$, let $K_N = \lceil \kappa \log N \rceil$ and, for $l = 0, 1, \ldots, K_N$ consider the sets

$$\mathcal{B}(l) := \{x \in \mathcal{X} : q^N - s(x) \leq K_N - l\} = \{x \in \mathcal{X} : s(x) \geq q^N - K_N + l\},$$

and define $\mathcal{B}(K_N + 1) := \emptyset$. Then, $\mathcal{B}(K_N) \subseteq \mathcal{B}(l) \subseteq \mathcal{B}(0)$ for all $l \in [K_N]$. For $l = 0, 1, \ldots, K_N$ define also

$$\mathcal{B}_=(l) := \mathcal{B}(l) \backslash \mathcal{B}(l+1) = \{x \in \mathcal{X} : s(x) = q^N - K_N + l\}.$$

Notice that $\mathcal{B}(0) = \cup_{l=0}^{K_N} \mathcal{B}_=(l)$.

Let $\lambda_\Sigma = \sum_i \lambda_i$, $\bar{\mu} = \max_i \mu_i$ and $\underline{\mu} = \min_i \mu_i$. Transitions in the Markov chain X^t are either of the form $x \to x + e_k$ (accepting an arriving type-k request) and $x \to x - e_k$ (a type-k service completion). For any set $\mathcal{C} \subseteq \mathcal{X}$ and its complement $\mathcal{C}^c = \mathcal{X} \backslash \mathcal{C}$

$$\sum_{x \in \mathcal{C}, z \in \mathcal{C}^c} \Pi(x) Q(x,z) = \sum_{z \in \mathcal{C}^c, x \in \mathcal{C}} \Pi(z) Q(z,x);$$

see e.g., Exercise 5.34 in Ross (1996). Take these sets to be $\mathcal{C} = \mathcal{B}(l)$ and $\mathcal{C}^c = \mathcal{B}(l)^c$. For $x \in \mathcal{B}(l), z \in \mathcal{B}(l)^c$, $s(x) \geq q^N - K_N + l$ and $s(z) < q^N - K_N + l$. If $Q(x,z) > 0$ there must exist $k \in [n]$ such that $z = x - e_k$ and $x \in \mathcal{B}_=(l), z \in \mathcal{B}_=(l-1)$. Similarly, for $z \in \mathcal{B}(l)^c, x \in \mathcal{B}(l)$ with $Q(z,x) > 0$ there must exist $k \in [n]$ such that $x = z + e_k$ and that $z \in \mathcal{B}_=(l-1), x \in \mathcal{B}_=(l)$.

Recall that $i_0 \in \mathcal{A}_p$ is such that $A_{ji_0} = 1$. By (29), $x_{i_0} = \mathbb{E}_{\Pi^*,N}[X_{i_0}] = \lambda_{i_0}^N / \mu_{i_0} - o(\log N)$, implying that

$$\sum_{x \in \mathcal{X}: Q(x, x+e_{i_0}) = 0} \Pi(x) = o(\log N / N).$$

Otherwise, if $\sum_{x \in \mathcal{X}: Q(x,x+e_{i_0})=0} \Pi(x) = \Omega(\log N / N)$, PASTA implies that $\Omega(\lambda_{i_0}^N \log N / N) = \Omega(\log N)$ type-i_0 customers arrive in those states x and are rejected, alternatively, at most $\lambda_{i_0}^N - \Omega(\log N)$ are accepted. By Little's law, we would then have $\mathbb{E}_{\Pi^*,N}[X_i] = \frac{1}{\mu_i}(\lambda_{i_0}^N - \Omega(\log N)) = \frac{\lambda_{i_0}^N}{\mu_i} - \Omega(\log N)$.

In turn,

$$\sum_{z\in\mathcal{B}(l)^c,x\in\mathcal{B}(l)}\Pi(z)Q(z,x)\geq\sum_{x\in\mathcal{B}_=(l-1),z=x+e_{i_0}}\Pi(z)Q(z,x)=\lambda_{i_0}^N\left(\sum_{z\in\mathcal{B}_=(l-1),Q(z,z+e_{i_0})>0}\Pi(z)\right)$$
$$=\lambda_{i_0}^N\left(\sum_{z\in\mathcal{B}_=(l-1)}\Pi(z)-\sum_{z\in\mathcal{B}_=(l-1),Q(z,z+e_{i_0})=0}\Pi(z)\right)$$
$$=\lambda_{i_0}^N\Big(\Pi(\mathcal{B}_=(l-1))-o(\log N/N)\Big). \tag{31}$$

Recalling that transition from $\mathcal{B}(l)$ to $\mathcal{B}(l)^c$ must be of the form $z=x-e_k$ we have

$$\sum_{x\in\mathcal{B}(l),z\in\mathcal{B}(l)^c}\Pi(x)Q(x,z)\leq\bar{\mu}qN.$$

Then,

$$\lambda_{i_0}^N\Big(\Pi(\mathcal{B}_=(l-1))-o(\log N/N)\Big)\leq\sum_{z\in\mathcal{B}(l)^c,x\in\mathcal{B}(l)}\Pi(z)Q(z,x)=\sum_{x\in\mathcal{B}(l),z\in\mathcal{B}(l)^c}\Pi(x)Q(x,z)\leq\Pi(\mathcal{B}_=(l))\bar{\mu}qN,$$

where the first inequality follows from (31). Recalling that $\lambda_{i_0}^N=N\lambda_{i_0}$, we have

$$\lambda_{i_0}\Big(\Pi(\mathcal{B}_=(l-1))-o(\log N/N)\Big)\leq\Pi(\mathcal{B}_=(l))\bar{\mu}q. \tag{32}$$

Applying this recursively we have the lower bound

$$\Pi(\mathcal{B}_=(l))\geq\delta^l\Pi(\mathcal{B}_=(0))-\frac{1}{1-\delta}\times o(\log N/N)=\Pi(\mathcal{B}_=(0))-o(\log N/N), \tag{33}$$

where $\delta:=\lambda_{i_0}/\bar{\mu}q<1$.

We next obtain an upper bound. Because transitions from $\mathcal{B}(l)^c$ to $\mathcal{B}(l)$ must be of the form $x=z+e_k$, the rate of these transition is bounded by $\sum_i\lambda_i^N=N\sum_i\lambda_i=N\lambda_\Sigma$, that is $\lambda_\Sigma N\Pi(\mathcal{B}_=(l))\geq\sum_{z\in\mathcal{B}(l)^c,x\in\mathcal{B}(l)}\Pi(z)Q(z,x)$. For all $l=\{0,1,\dots,K_N\}$, we have that $|q^N-s(x)|\leq K_N$ and, in particular, that $s(x)\geq q^N/2=qN/2$ for all N sufficiently large. Thus, $Q(x,z)\geq\underline{\mu}qN/2$ and we have

$$\bar{\lambda}N\Pi(\mathcal{B}_=(l-1))\geq\sum_{z\in\mathcal{B}(l)^c,x\in\mathcal{B}(l)}\Pi(z)Q(z,x)=\sum_{x\in\mathcal{B}(l),z\in\mathcal{B}(l)^c}\Pi(x)Q(x,z)\geq\Pi(\mathcal{B}_=(l))\underline{\mu}qN/2. \tag{34}$$

We then have the upper bound

$$\Pi(\mathcal{B}_=(l))\leq\bar{\delta}^l\Pi(\mathcal{B}_=(0)), \tag{35}$$

where $\bar{\delta}:=\lambda_\Sigma/\underline{\mu}q/2>1>\lambda_{i_0}/\bar{\mu}q=\delta$. Here we recall that the resource constraint for j is binding so that $\lambda_\Sigma/\underline{\mu}\geq\sum_i\lambda_i/\mu_i\geq\sum_i A_{ji}y_i^*=q$.

The set $\left\{l\in\{0,\dots,K_N-1\}:\sum_{x\in\mathcal{B}(l),z\in\mathcal{B}(l+1)}Q(x,z)>0\right\}$ is non-empty. Indeed, if it were empty (meaning that no requests are accepted when there q^N-K_N or more units of server j busy), then we would have that $(Ax^N)_j\leq q^N-K_N$ which, because $K_N=\Omega(\log N)$, would contradict (28). Let

$$l^*:=\max\left\{l\in\{0,\dots,K_N-1\}:\sum_{x\in\mathcal{B}(l),z\in\mathcal{B}(l+1)}Q(x,z)>0\right\}.$$

Recall that $\mathcal{B}(0) = \{x \in \mathcal{X} : s(x) \geq q^N - K_N\} = \cup_{l=0}^{K_N} \mathcal{B}_=(l)$. By (30), $\Pi(\mathcal{B}(0)) = \mathbb{P}_{\Pi^*,N}\{s(X) \geq q^N - K_N\} \geq \frac{1}{2}$ for all large enough N. Using the upper bound (35), and recalling that $\bar{\delta} > 1$, we have that

$$\frac{1}{2} \leq \Pi(\mathcal{B}(0)) = \sum_{l=0} \Pi(\mathcal{B}_=(l) \leq K_N \Pi(\mathcal{B}_=(0))\bar{\delta}^{K_N}.$$

In turn,

$$\Pi(\mathcal{B}_=(0)) \geq \frac{2}{K_N \bar{\delta}^{K_N}}.$$

Using now the lower bound (33), and recalling that $K_N = \lceil \kappa \log N \rfloor$ for some $\kappa > 0$, we have

$$\Pi(\mathcal{B}_=(l)) \geq \frac{2}{K_N}\left(\frac{\delta}{\bar{\delta}}\right)^{K_N} - o(\log N/N) \geq \frac{2}{\kappa \log N + 1}\left(\frac{\delta}{\bar{\delta}}\right)^{\kappa \log N + 1} - o(\log N/N).$$

Letting $\beta = -\log(\delta/\bar{\delta})$, we then have, for all $l \in [l^* + 1]$,

$$\Pi(\mathcal{B}_=(l)) \geq 2e^{-\beta}\frac{1}{N^{\beta\kappa}(\kappa \log N + 1)} - o(\log N/N) \geq 2e^{-\beta}\frac{1}{N^{\kappa(\beta+1)}} - o(\log N/N),$$

where in the last inequality we use the fact that, for fixed κ and all N large, $N^\kappa \geq \kappa \log N$. Choosing $\kappa = \frac{1}{2(\beta+1)}$ we then have for all such l (including $l^* + 1$) that

$$\Pi(\mathcal{B}_=(l)) \geq 2e^{-\beta}\frac{1}{\sqrt{N}} - o(\log N/N).$$

By definition, no arrivals of type $i_0 \in \mathcal{A}_p$, are accepted in states $x \in \mathcal{B}_=(l^* + 1)$. In turn, $\{x \in \mathcal{X} : Q(x, x + e_{i_0}) = 0\} \supseteq \mathcal{B}_=(l^* + 1)$. Using PASTA we then have that the number of type-i_0 customers rejected per unit of time is

$$\lambda_{i_0}^N \sum_{x:Q(x,x+e_{i_0})=0} \Pi(x) \geq \lambda_{i_0}^N \Pi(\mathcal{B}_=(l^* + 1)) \geq \lambda_{i_0}^N \left(2e^{-\beta}\frac{1}{\sqrt{N}} - o(\log N/N)\right) = \Omega\left(\sqrt{N}\right).$$

The long-run number of type i_0 customers in system, $x_i^N = \mathbb{E}_{\Pi^*,N}[X_i]$, then satisfies

$$x_i^N = \frac{\lambda_{i_0}^N}{\mu_{i_0}} - \Omega(\sqrt{N}),$$

contradicting (29) and, in turn, the optimality of the policy. □

Remark 6.1 *Resources with strictly positive slack capacity.* We recall that Assumption 3.1 allows for the resources to have strictly positive slack in the optimal solution. Thus, far we have made the simplifying assumption that there are no such resources. Upon conclusion of the proofs, we are now in the position to explain why this restriction comes at no loss of generality. First, we note that these resources would not appear in the definition of the relaxed network or the relaxed policy and hence play no role in its analysis. They could play a role in the mixing-time result. These, it is evident, will also not change the mixing time result for the original network operated under CHT. In turn, these resources have no impact on the proof of the upper bound. For the lower bound, the assumption $\mathcal{A}_p \neq \emptyset$ in Theorem 4.1 has to be strengthened to the requirement that $\mathcal{A}_p(j) \neq \emptyset$ for at least one j whose capacity constraint is binding in the optimal solution to (LP).

■

7. Simulations

In this section, we present simulation results for the networks in Figures 2, 3, and 4. In the simulations, we scale both the customer arrival rates $\lambda_i^N = N\lambda_i$ for all $i \in [n]$ and the resource units $q_j^N = Nq_j$ for all $j \in [d]$ with $N \in \{200, 300, \ldots, 1500\}$. For each value of N, we run one long sample path and report the long-run average after dropping a warm-up period. Taking into account that mixing time increases with the network scale, we use a horizon length that scales up with N^2.

Network 1. This is the network in Example 3.1 and we use the same parameters. The less-preferred customer types $\mathcal{A}_{lp} = \{2, 3\}$ are those colored in red in Figure 2. The parameters for this particular setting are given by

$$\lambda = (3, 2, 5), \quad \mu^{-1} = (2, 1, 3), \quad q = (7, 6), \quad r = (5, 1, 2).$$

We choose the coefficients of CHT as $\delta_a = \delta_b = 20$.

Network 2. This is the network in Example 3.2 and we use the same parameters. The less-preferred customer types $\mathcal{A}_{lp} = \{1, 2, 3, 5\}$ are those colored in red in Figure 3. The parameters for this particular setting are given by

$$\lambda = (2, 3, 5, 1, 6, 2, 3), \quad \mu^{-1} = (1, 3, 2, 3, 5, 4, 2), \quad q = (11, 19, 14, 7), \quad r = (2, 1, 3, 5, 1, 6, 5).$$

We choose the coefficients of CHT as $\delta_1 = \delta_2 = \delta_3 = \delta_5 = 20$.

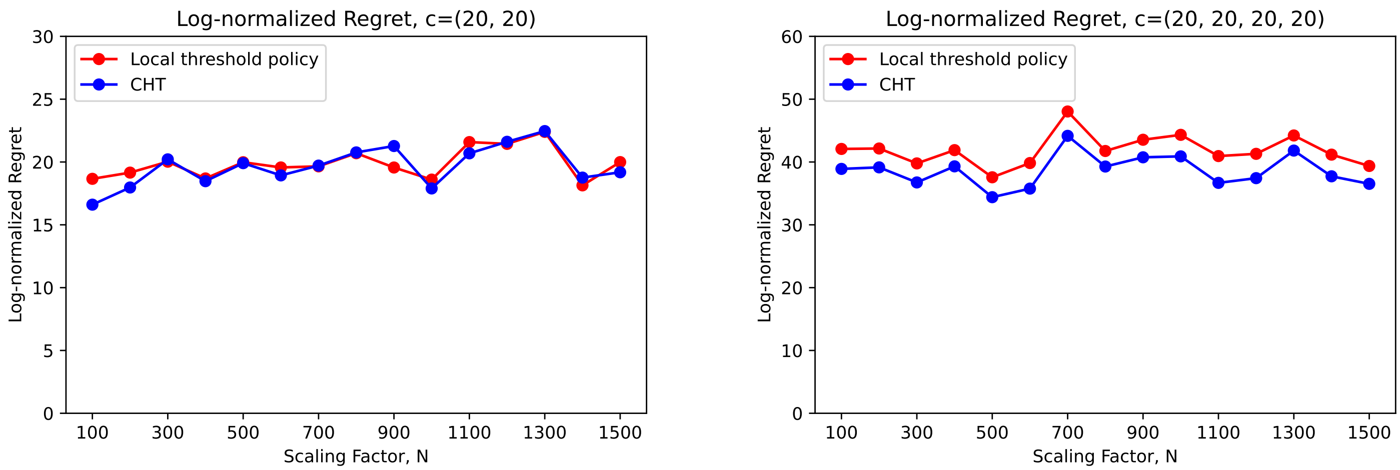
Figure 8 depicts the log-normalized regret of CHT as a function of the scaling factor N for Network 1 and Network 2. The plot suggests that the log-normalized regret does not grow with the scaling factor N, which echoes our results. For each of the networks we also plot the log-normalized regret under a threshold policy with the same d logarithmic thresholds but where these are applied to the true headcount Σ^t instead of the corrected one $\Sigma^{*,t}$; in the network of Figure 3 type 3 is accepted whenever feasible and $q_c^N - \Sigma_c^t \geq \delta_3 \log N$. The "natural" threshold policy—with the threshold identified, still, based on the perfect matching—seems to perform as well as CHT. These two networks, however, have a special property — their LP-residual network graph has a *unique* perfect matching.

Network 3. Interestingly, we see similar performance in the network of Example 3.3. Here, the parameters are

$$\lambda = (2, 2, 2, 1/2, 1/3, 1/4), \quad \mu^{-1} = (1, 1, 1, 2, 3, 4), \quad q = (2, 2, 2), \quad r = (1, 1, 1, 4, 6, 8).$$

We use the perfect matching $(1, a), (2, b), (3, c)$ for the threshold assignment, accepting a type-1, for example, only if the headcount of resource a is smaller than $q_a^N - \delta_1 \log N$. In this network, the two policies—CHT and the regular (using the true headcount) threshold policy—*no longer*

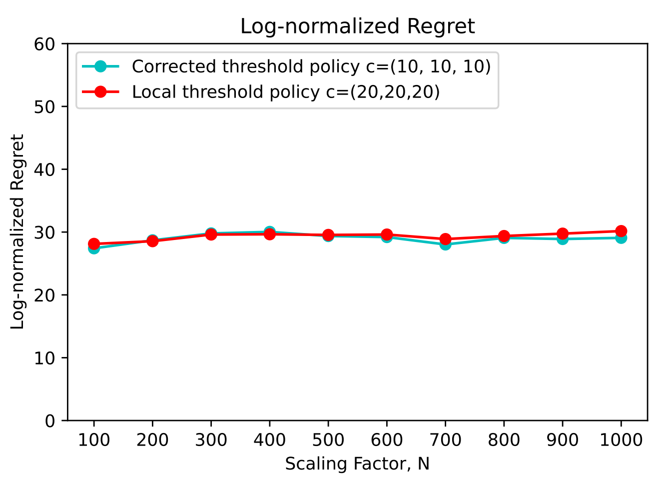
Figure 8 Log-normalized regret for the networks in Example 3.1(LEFT) and in Example 3.2(RIGHT)



Notes. The plot displays the (simulated) regret of CHT normalized by the natural log of the scaling factor N in the setting of Example 3.1 and Example 3.2 respectively. Specifically, we simulate the regret—the difference between the LP upper bound and the long-run average reward—in a long time horizon with the warm-up period being dropped off. In both networks, the local threshold policy is the one that uses the true, instead of the corrected, headcount.

show identical performance. They both have logarithmic regret, but the exact constant is different if the same thresholds are used. However, with these specific parameters, the same performance that a threshold δ achieves in the regular threshold policy is achieved by $\delta/2$ under CHT. This is shown in Figure 9 which depicts the log-normalized regret under the two policies. We tried different parameter combinations for this network and all showed similarly stable and good performance.

Figure 9 Log-normalized regret for the network in Figure 4(LEFT)



Finally, we conducted a sensitivity analysis for the setting of Network 1. In Table 1, we report CHT's stationary rewards for $N = 1000$ and different values of the two thresholds (coefficients) δ_a, δ_b. The important threshold seems to be δ_a which protects the preferred type 1: the reward is

Table 1 Performance of CHT with Different Coefficients–Example 1. The numbers are rounded to the nearest integer.

δ_a ($R_a = \delta_a \log N$) \ δ_b ($R_b = \delta_b \log N$)	0.5	1	2	3	5	10	15	20
0.5	18727	18672	18659	18655	18646	18623	18600	18577
1	19076	19055	19049	19043	19033	19010	18987	18964
2	19299	19293	19289	19285	19276	19253	19230	19207
3	19327	19326	19322	19318	19308	19286	19262	19239
5	19328	19328	19323	19318	19309	19285	19263	19239
10	19319	19315	19310	19305	19296	19272	19250	19227
15	19306	19303	19299	19293	19284	19260	19237	19213
20	19296	19294	19287	19282	19273	19250	19228	19204

highest when this threshold is large and is less sensitive to how big the threshold is for resource b. It is an interesting question, left for future work, to understand the dependence of the required threshold magnitude on the location of a resource in the network.

In Section B of the appendix, we provide a simple heuristic to guide the choice of the threshold coefficients, accompanied by numerical evidence.

8. Closing Remarks

In this paper, we study a dynamic resource allocation problem with multiple types of customers and multiple types of reusable resource units. Under an overload condition of the associated LP of the problem, we devise a threshold policy—with the number of thresholds equal to the number of resources—and show that its regret is at most logarithmic in the problem size in the many-customer many-resource regime. The thresholds are applied to a corrected headcount process at each resource.

The study of networks with simultaneous resource possession (as the ones we study here) inevitably leads to questions about the relationship between performance and the underlying combinatorial/graph structure; see also Gurvich and Van Mieghem (2015); Dawande et al. (2021).

Our solution in this paper is based strongly on the existence, in overloaded networks, of a perfect matching in the residual graph. The policy we propose is a centralized policy where, to make an acceptance decision, one must calculate the targeted levels for less-preferred types which may require the knowledge of the full vector $X^t_{\mathcal{A}_p}$ of preferred customers in the system. Our numerical experiments in §7 suggest that a decentralized threshold policy—where acceptance decisions are made locally at each resource based on its true (instead of corrected) headcount—achieves, as well, logarithmic performance, but proving this seems challenging. Our proof for CHT relied on the fact that, in the auxiliary system, each resource and its "matched" less-preferred type can be studied in isolation. A general analysis of the decentralized policy might necessitate a complicated Lyapunov function that captures the interaction between resources. We conjecture, however, that logarithmic regret of the decentralized policy is provable, using decomposition arguments, for networks with a *unique* perfect matching such as those in Examples 3.1 and 3.2.

It is important that we benchmarked our policy against an LP upper bound (a deterministic counterpart). In the dynamic stochastic knapsack setting (finite horizon with non-reusable resources), the offline decision maker—one that sees the future realization of demand—was used as a benchmark; see Arlotto and Gurvich (2019); Vera and Banerjee (2021). The offline objective value is a tighter upper than the LP. In that setting, the offline decision maker solves an LP with a random right-hand side. In the case of reusable resources, the offline problem—where arrivals and service times are known to the decision maker—is a complicated dynamic program and, hence, difficult to use as a benchmark. We use the LP, instead. It is natural to ask whether (i) one can do better relative to the offline upper bound, and (ii) the non-degeneracy assumption can be removed if a tighter offline bound is used as a benchmark. The answer to both of these was answered affirmatively in the dynamic stochastic knapsack setting. It is also worth noting that one can have multiple offline problems. One simple offline version is the one in equation (1) that we used in the proof of Lemma 3.1; it is reminiscent of the one used in the stochastic knapsack setting. That crude offline suffices to prove that, with degeneracy, the gap from the LP is $\mathcal{O}(\sqrt{N})$, but–precisely because it does not capture the state constraints under non-preemption—it is too loose. Indeed—with non-degeneracy—this offline is $\mathcal{O}(1)$ from the LP, while we prove that no online policy can achieve sub-logarithmic regret.

The development of tighter offline benchmarks that allow to go beyond non-degeneracy, remains desirable, especially given the strong results in the non-reusable case; see Vera et al. (2021).

There are several natural extensions to what we proved here. In our model, a request requires one unit of each of a set of resources. A natural extension would model requests where the number of units is itself heterogeneous and random. The second question pertains to the regret when the network itself (the number of types and number of resources) scales, rather than the number of units of each of (a finite set of) resources. A challenge here is to model the sequence of (growing) networks. The literature on flexibility might provide some clues here; see for instance Tsitsiklis and Xu (2017); Van der Boor et al. (2018).

Answering these questions would give us a more complete picture of dynamic resource allocation with reusable resources.

Acknowledgments

We thank the AE and three referees for their constructive comments.

Notation

For easy reference, below is a list of notations that we use in the paper.

	Problem setup
$i \in [n]$ and $j \in [d]$	The set of customer types and resource types respectively
$\lambda_i^N = N\lambda_i, q_j^N = Nq_j$	Arrival rates and available resource quantities, where N is the scaling factor
$\mathcal{A}(j)$	The set of customer types that request type-j resource unit
$\mathcal{S}(i)$	The set of resource types requested by type-i customers
	The LP solution under the overload condition
$\mathcal{R}^*$ and $\overline{\mathcal{R}}(q, \lambda/\mu)$	The optimal value of DP and of the LP relaxation
$[n] = \mathcal{A}_p \cup \mathcal{A}_{lp}$	Partition of the customer base: preferred types and less-preferred types
$\mathcal{A}_p(j) = \mathcal{A}_p \cap \mathcal{A}(j)$	The set of preferred customer types that request type-j resource unit
$\mathcal{A}_{lp}(j) = \mathcal{A}_{lp} \cap \mathcal{A}(j)$	The set of preferred customer types that request type-j resource unit
	CHT
$\delta_i, i \in \mathcal{A}_{lp}$	Threshold coefficient for each less-preferred customer type
X^t	The number of customers in service customers at time t
$x_i^*(X^t)$	the targeted number of type-i ($i \in \mathcal{A}_{lp}$) customers at time t
$\Sigma_j^t = \sum_{i \in \mathcal{A}(j)} X_i^t$	Number of customers using type-j resource
$\Sigma_j^{*,t} = \sigma_j(X^t)$	the corrected resource-j headcount process
	Markov chain analysis
$\mathcal{Q}$	The infinitesimal generator (of the relaxed policy $\widetilde{\pi}$)
CHT, $\widetilde{\pi}$	The corrected headcount threshold policy and the relaxed policy
$X^t, \widetilde{X}^t$	The number of under-service customers under the two policies at time t
$\Pi, \widetilde{\Pi}$	The stationary distribution of X^t and $\widetilde{X}^t$
$\mathcal{X}, \widetilde{\mathcal{X}}$	the state space of the two policies
η	absolute constant (that does not depend on N)

Appendix

A. Proofs of Lemmas

A.1. Proofs for lemmas in Section 3

Proof of Lemma 3.2. The feasible region of (LP) is the polyhedron

$$\{y \in \mathbb{R}^n : Ay \leq q \text{ and } 0 \leq y \leq \lambda/\mu\}.$$

Due to our assumed strict complementary slackness, all d resource constraints are tight (i.e., hold at equality). Furthermore, since the LP solution is, by assumption, non-degenerate, and there are n variables, there are exactly n linearly independent tight constraints at the (extreme point) optimal solution. Thus, there remain exactly $n-d$ tight constraints among the demand constraints $0 \leq y \leq \lambda/\mu$, i.e., exactly $n-d$ variables with y_i^* that is equal to either 0 or λ_i/μ_i. As a result, there are exactly d variables with $0 < y_i^* < \lambda_i/\mu_i$ (corresponding to the less preferred types).

For the second statement of the lemma, consider the *residual* optimization problem for types $i \in \mathcal{A}_{lp}$. Specifically, let $r_{\mathcal{A}_{lp}}$ and $y_{\mathcal{A}_{lp}}$ be the suitable sub-vectors of r and y, and $A_{\mathcal{A}_{lp}}$ be the sub-matrix of A that has only the columns for $i \in \mathcal{A}_{lp}$. Since $|\mathcal{A}_{lp}| = d$, this is a $d \times d$ matrix. The residual LP is then given by

$$\begin{aligned} &\max\ (r_\mu)'_{\mathcal{A}_{lp}} y_{\mathcal{A}_{lp}} \\ &\text{s. t.}\ \ A_{\mathcal{A}_{lp}} y_{\mathcal{A}_{lp}} \leq \widetilde{q}, \\ &\qquad 0 \leq y_{\mathcal{A}_{lp}} \leq (\lambda/\mu)_{\mathcal{A}_{lp}}, \end{aligned}$$

where $\tilde{q} = q - A_{\mathcal{A}_p}(\lambda/\mu)_{\mathcal{A}_p}$ is the residual capacity after allocation to the preferred types. This residual optimization problem has the unique (sub) solution $y^*_{\mathcal{A}_{lp}}$. The residual LP has $2d$ constraints. Because $|\mathcal{A}_{lp}| = d$ and because $y_i^* \in (0, \lambda_i/\mu_i)$ for all $i \in \mathcal{A}_{lp}$ the optimal basis has all the decision variables as well as the slack variables for the demand constraint. The slacks for the capacity constraints are zero-valued and non-basic. We conclude that: (i) the solution for this residual problem is, as well, unique and non-degenerate, and (ii) the basis matrix is $\mathcal{A}_{lp}$ itself and has linearly independent rows. □

Proof of Lemma 3.1. The proof is an adaptation of the proof of Proposition 1 in Vera and Banerjee (2021). The difference here is that the right-hand side corresponds to the stationary count of requests instead of the total number of arrivals.

Fix a stationary policy and let X_i^t be the number of type-i customers in the system. Then, a simple coupling argument shows that $X_1^t, \ldots, X_n^t$ is bounded component-wise from above by $Y_1^t, \ldots, Y_n^t$ where Y_i^t is the number of type-i customers in an infinite server queue that accepts all arriving customers. The variables $Y_1^t, \ldots, Y_n^t$ are independent and as $t \uparrow \infty$, Y_i^t converges to a Poisson random variable with mean λ_i^N/μ_i.

In turn, an upper bound on the performance of any stationary policy is given by the expectation applied to (LP), where the deterministic demand constraints λ^N/μ are replaced with the random vector $Y^N = (Y_i^N, i \in [n])$, i.e., let

$$\overline{\mathcal{R}}(q^N, Y^N) := \begin{cases} \max_{y \in \mathbb{R}_+^n} r_\mu' y \\ \text{s.t. } Ay \le q^N, \\ \qquad y \le Y^N. \end{cases}$$

Then, by Jensen's inequality

$$\mathbb{E}[\overline{\mathcal{R}}(q^N, Y^N)] \le \mathcal{R}(q^N, \lambda^N/\mu) = \begin{cases} \max_{y \in \mathbb{R}_+^n} r_\mu' y \\ \text{s.t. } Ay \le q^N, \\ \qquad y \le \lambda^N/\mu. \end{cases}$$

The dual to the problem on the right-hand side (which is the deterministic upper bound) is given by

$$\begin{aligned} D[\lambda^N/\mu] := \min\ & \alpha' q^N + \beta'(\lambda^N/\mu) \\ \text{s. t. } & \alpha' A + \beta \ge r_\mu, \\ & \alpha \in \mathbb{R}_+^d, \beta \in \mathbb{R}_+^n. \end{aligned}$$

Notice that because $q^N = qN$ and $\lambda^N = N\lambda$, the dual solution does not depend on N. Let $\beta = \beta^1 - \beta^2$ and $\alpha = \alpha^1 - \alpha^2$. By the assumption of this lemma $\beta \neq 0$.

Arguing as in Vera and Banerjee (2021) we conclude that

$$\begin{aligned} \overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathbb{E}[\overline{\mathcal{R}}(q^N, Y^N)] &= \mathbb{E}[(\lambda^N/\mu - Y)'\beta \mathbb{1}\left\{(\lambda^N/\mu - Y^N)'\beta > 0\right\}] \\ &= \sqrt{N}\mathbb{E}\left[\frac{1}{\sqrt{N}}(\lambda^N/\mu - Y^N)'\beta \mathbb{1}\left\{\frac{1}{\sqrt{N}}(\lambda^N/\mu - Y^N)'\beta > 0\right\}\right]. \end{aligned}$$

Recall that $Y_i \sim Poisson(\lambda_i^N/\mu_i)$. By standard results, $\frac{1}{\sqrt{N}}(\lambda^N/\mu - Y)'\beta \Rightarrow \widehat{W}$ where $\widehat{W}$ is a normal random variable with 0 mean and variance $(\lambda/\mu)'\beta$. By Fatou's lemma

$$\lim_{n \uparrow \infty} \mathbb{E}\left[\frac{1}{\sqrt{N}}(\lambda^N/\mu - Y^N)'\beta \mathbb{1}\left\{\frac{1}{\sqrt{N}}(\lambda^N/\mu - Y^N)'\beta > 0\right\}\right] \ge \mathbb{E}\left[\widehat{W}\mathbb{1}\left\{\widehat{W} > 0\right\}\right] > 0,$$

where the last inequality is a basic property of the normal random variable. We conclude that

$$\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathbb{E}[\overline{\mathcal{R}}(q^N, Y^N)] = \Omega(\sqrt{N}).$$

□

A.2. Proofs for lemmas in Section 5

Proof of Lemma 5.1. The Markov chain $(\widetilde{X}^t, t \geq 0)$ is easily verified to be irreducible relative to the state space

$$\widetilde{\mathcal{X}} = \{x \in \mathbb{Z}_+^n : x_i \leq q_{j_i}^N - \delta_i \log N, \text{for all } i \in \mathcal{A}_{lp}\}.$$

For each $i \in \mathcal{A}_{lp}$, $x_i \leq q_{j_i} - \delta_i \log N$. For $i \in \mathcal{A}_p$, $\widetilde{X}_i^t$ follows the law of an infinite server queue (independently of X_j^t, $j \neq i$) so we have that $\lim_{t\uparrow\infty} \mathbb{E}_x[X_i^t] = \lambda_i^N/\mu_i$. We then have that for any initial state x

$$\limsup_{t\uparrow\infty} \sum_{i\in[n]} \mathbb{E}_x[X_i^t] < \infty.$$

By Markov's inequality, there exists K such that $\liminf_{t\geq 0} \mathbb{P}\{\sum_{i\in[n]} X_i^t \leq K\} \geq \frac{1}{2}$. For an irreducible chain on $\mathbb{R}_+^n$ we can have only one of two outcomes: either $\lim_{t\uparrow\infty} \mathbb{P}_x\{\sum_{i\in[n]} X_i^t = j\} = 0$ for all j or the chain is positive recurrent and has a unique stationary distribution (see Corollary 4.7 in Asmussen (2003)). We conclude that the chain is positively recurrent.

The concentration bound for types $i \in \mathcal{A}_p$ is standard and follows immediately from the fact that the stationary distribution for these types is (independently of everything else) Poisson with mean $\lambda_i^N/\mu_i = \lambda_i N/\mu_i$. For a Poisson random variable Z with mean ν, it is known that

$$\mathbb{P}\{Z \geq \nu + x\} \leq \exp\left(-\frac{x^2}{(2(\nu + x/3)}\right).$$

Plugging $\nu = \lambda_i N/\mu_i$ and $x = \epsilon N$, and applying a union bound, we obtain the bound in the lemma with $m_1 = 2(n-d)$ and $m_2 = \epsilon^2/(2\lambda_i/\mu_i + 2\epsilon/3)$. □

Proof of Lemma 5.2. We will use the following result for Markov chains; see, for example, Proposition 3 in Glynn and Zeevi (2008).

Lemma A.1 *Suppose that $Z = (Z^t : t \geq 0)$ is a non-explosive continuous-time Markov chain on a state space $\mathcal{Z}$ with rate matrix $\mathcal{Q}$ and a stationary distribution Π. Then, for any function $g : \mathcal{Z} \to \mathbb{R}$ for which $\sum_{z\in\mathcal{Z}} \Pi(z)|\mathcal{Q}(z,z)||g(z)| < \infty$,*

$$\mathbb{E}_\Pi[(\mathcal{Q}g)(Z)] = 0.$$

Trivially,

$$|\mathcal{Q}(x,x)| \leq \sum_{i\in[n]} (\lambda_i^N + \mu_i x_i) \leq \eta_1 N + \eta_2 \sum_{i\in\mathcal{A}_p} x_i.$$

Here we used that fact that for $i \in \mathcal{A}_{lp}$, $x_i \le q_{j_i}^N = q_{j_i}N$. Recall that $g_i^\theta(x) = e^{\theta|\sigma_{j_i}(x) - \widehat{q}_{j_i}^N|}$. Recalling also the definition of $\sigma_j(x)$ in (10), we have that

$$\begin{aligned}
|\sigma_{j_i}(x) - \widehat{q}_{j_i}^N| &\le \sigma_{j_i}(x) + \widehat{q}_{j_i}^N \\
&\le x_i + \sum_{k\in\mathcal{A}_p} x_k + \sum_{k\in\mathcal{A}_{lp}} x_k^*(x) + \widehat{q}_{j_i}^N \\
&\le x_i + \sum_{k\in\mathcal{A}_p} x_k + \sum_{k\in\mathcal{A}_{lp}} |(A_{lp}^{-1}\left(q^N - A_p x_{\mathcal{A}_p}\right))_k| + \widehat{q}_{j_i}^N \\
&\le x_i + \eta_1 \sum_{k\in\mathcal{A}_p} x_k + \eta_2 N \\
&\le \eta_3\Big(N + \sum_{k\in\mathcal{A}_p} x_k\Big),
\end{aligned}$$

where we used the fact that $x^*(\zeta) := A_{lp}^{-1}\left(q^N - A_p\zeta\right)$ (see (3)). In the last inequality, we use again the fact that $x_i \le q_{j_i}^N = q_{j_i}N$ for $i \in \mathcal{A}_{lp}$. Thus,

$$\begin{aligned}
\sum_x \widetilde{\Pi}(x)|Q(x,x)|g_i^\theta(x) = \mathbb{E}_{\widetilde{\Pi}}[|Q(\widetilde{X},X)|g_i^\theta(\widetilde{X})] &\le \eta_5 \mathbb{E}_{\widetilde{\Pi}}\left[\Big|N + \sum_{k\in\mathcal{A}_p}\widetilde{X}_k\Big| e^{\eta_6\theta(N+\sum_{k\in\mathcal{A}_p}\widetilde{X}_k)}\right] \\
&\le \eta_5 \sqrt{\mathbb{E}_{\widetilde{\Pi}}\Big[(N + \sum_{k\in\mathcal{A}_p}\widetilde{X}_k))^2\Big]}\sqrt{\mathbb{E}_{\widetilde{\Pi}}\left[e^{2\eta_6\theta(N+\sum_{k\in\mathcal{A}_p}\widetilde{X}_k)}\right]},
\end{aligned}$$

where the last step is Hölder's inequality. For $k \in \mathcal{A}_p$, $\widetilde{X}_k$ is, under $\widetilde{\Pi}$, a Poisson random variable with mean λ_k^N/μ_k and $\widetilde{X}_i, i \in \mathcal{A}_p$ are independent random variables; the random variable $\sum_{k\in\mathcal{A}_p}\widetilde{X}_k$ is, under $\widetilde{\Pi}$, a Poisson random variable with mean $\sum_{k\in\mathcal{A}_p}\lambda_k^N/\mu_k$. In particular, $\mathbb{E}_{\widetilde{\Pi}}[(\sum_{k\in\mathcal{A}_p}\widetilde{X}_k)^2] < \infty$ and, for all small enough θ, $\mathbb{E}_{\widetilde{\Pi}}[e^{\eta_6\theta\sum_{k\in\mathcal{A}_p}\widetilde{X}_k}] < \infty$. We conclude that $\sum_x \widetilde{\Pi}(x)|Q(x,x)|g_i^\theta(x) < \infty$ as required. The conclusion then follows from Lemma A.1 above. $\square$

Proof of Lemma 5.3. Recalling (4) we write, for $i \in \mathcal{A}_{lp}$,

$$x_i^*(\zeta) = y_i^* N + \sum_{l\in\mathcal{A}_p} \alpha_i^l(\zeta_l - \lambda_l/\mu_l),$$

where $\alpha_i^l = [A_{lp}^{-1}A_p]_{i,l},\ i \in \mathcal{A}_{lp}, l \in \mathcal{A}_p$, are real numbers.

Recall the expression for $\sigma_j(x)$ on the second row of (10) and that, given i, $g_i^\theta(x) = e^{\theta|\sigma_{j_i}(x) - \widehat{q}_{j_i}^N|}$. We fix i and θ and drop both, writing $g(x)$ instead of $g_i^\theta(x)$. We have

$$(\mathcal{Q}g)(x) = \begin{cases}
\lambda_i^N(g(x+e_i) - g(x)) + \mu_i x_i(g(x-e_i) - g(x)) \\
\quad + \sum_{l\in\mathcal{A}_p(j_i)} \lambda_l^N(g(x+e_l) - g(x)) + \sum_{l\in\mathcal{A}_p(j_i)} \mu_l x_l(g(x-e_l) - g(x)) \\
\quad + \sum_{l\in\mathcal{A}_p} \lambda_l^N(g(x+\alpha_i^l e_l) - g(x)) + \sum_{l\in\mathcal{A}_p} \mu_l x_l(g(x-\alpha_i^l e_l) - g(x)) & \text{if } \sigma_{j_i}(x) \le \widehat{q}_{j_i}^N \\
\mu_i x_i(g(x-e_i) - g(x)) \\
\quad + \sum_{l\in\mathcal{A}_p(j_i)} \lambda_l^N(g(x+e_l) - g(x)) + \sum_{l\in\mathcal{A}_p(j_i)} \mu_l x_l(g(x-e_l) - g(x)) \\
\quad + \sum_{l\in\mathcal{A}_p} \lambda_l^N(g(x+\alpha_i^l e_l) - g(x)) + \sum_{l\in\mathcal{A}_p} \mu_l x_l(g(x-\alpha_i^l e_l) - g(x)) & \text{if } \sigma_{j_i}(x) > \widehat{q}_{j_i}^N
\end{cases}$$

where e_l stands for the vector in $\mathbb{R}^n$ with 1 in the l^{th} coordinate and 0 everywhere else.

Let

$$\beta(x) := \frac{1}{g(x)}\left(\sum_{l\in\mathcal{A}_p}\lambda_l^N(g(x+\alpha_i^l e_l)-g(x)) + \sum_{l\in\mathcal{A}_p}\mu_l x_l(g(x-\alpha_i^l e_l)-g(x))\right).$$

Then,

$$|\beta(x)| \le 2\sum_{l\in\mathcal{A}_p}\max\{|\lambda_l^N e^{|\alpha_i^l\theta|}-\mu_l x_l|, |\lambda_l^N-\mu_l x_l e^{|\alpha_i^l\theta|}|\}. \tag{36}$$

Expanding on $\mathcal{Q}g$ we have

$$(\mathcal{Q}g)(x) = \begin{cases} -(1-e^{-\theta})\left(\sum_{l\in\mathcal{A}_p(j_i)\cup\{i\}}(\lambda_l - e^{\theta}\mu_l x_l)\right)\exp(\theta f(x)) + \beta(x)\exp(\theta f(x)) & \text{if } \sigma_{j_i}(x) < \widehat{q}_{j_i}^N \\ \sum_{l\in\mathcal{A}_p(j_i)\cup\{i\}}(\lambda_l^N + x_l\mu_l)(e^{\theta}-1) + \beta(x) & \text{if } \sigma_{j_i}(x) = \widehat{q}_{j_i}^N \\ -(1-e^{-\theta})\left(x_i\mu_i + \sum_{l\in\mathcal{A}_p(j_i)}(x_l\mu_l - e^{\theta}\lambda_l^N)\right)\exp(\theta f(x)) + \beta(x)\exp(\theta f(x)) & \text{if } \sigma_{j_i}(x) > \widehat{q}_{j_i}^N. \end{cases} \tag{37}$$

We used in this derivation the fact that, when $\sigma_{j_i}(x) = \widehat{q}_{j_i}^n$, $f(x)=0$ and $g(x)=\exp(\theta f(x))=1$.

Choose $\epsilon = (1+\|A_{lp}^{-1}\|)^{-1}\min_{i\in\mathcal{A}_{lp}} y_i^*/(16n) \le \min_{i\in\mathcal{A}_{lp}} y_i^*/(16n)$. For $x\in\Omega_\epsilon^N$, choosing $\theta>0$ small enough (dependent on ϵ and $\bar{\alpha} = \max_{i,l}|\alpha_{i,l}|$), we have

$$\sum_{k\in\mathcal{A}_p}|e^{\theta}\lambda_k^N - \mu_k x_k| + \sum_{k\in\mathcal{A}_p}|\lambda_k^N - e^{\theta}\mu_k x_k| \le N\min_{i\in\mathcal{A}_{lp}} y_i^*/8,$$

as well as

$$\beta(x) \le N\min_{i\in\mathcal{A}_{lp}} y_i^*/8. \tag{38}$$

Consider now the three cases $\sigma_{j_i}(x) > \widehat{q}_{j_i}^N$, $\sigma_{j_i}(x) < \widehat{q}_{j_i}^N$, and $\sigma_{j_i}(x) = \widehat{q}_{j_i}^N$.

(i) $\sigma_{j_i}(x) > \widehat{q}_{j_i}^N$: For $x\in\Omega_\epsilon^N$ with $\sigma_{j_i}(x) > \widehat{q}_{j_i}^N$,

$$\begin{aligned} x_i &= \sigma_{j_i}(x) - \sum_{k\in\mathcal{A}_p(j_i)} x_k - \sum_{k\in\mathcal{A}_{lp}(j_i)\setminus\{i\}} x_k^* \\ &\ge \widehat{q}_{j_i}^N - \sum_{k\in\mathcal{A}_p(j_i)} x_k - \sum_{k\in\mathcal{A}_{lp}(j_i)\setminus\{i\}} x_k^*. \end{aligned}$$

Recall that $x^*(\zeta) = y_{\mathcal{A}_{lp}}^* N + A_{lp}^{-1}\mathcal{A}_p[(\lambda^N/\mu)_{\mathcal{A}_p} - \zeta]$. Given $x\in\Omega_\epsilon^N$, take $\zeta = x_{\mathcal{A}_p}$. Then, for $x\in\Omega_\epsilon^N$, $\|x_{\mathcal{A}_p} - (\lambda^N/\mu)_{\mathcal{A}_p}\| \le n\epsilon N$ and $\|x^*(x_{\mathcal{A}_p}) - y_{\mathcal{A}_{lp}}^* N\| \le \|A_{lp}^{-1}\| n\epsilon N$. Because

$$q_{j_i}^N = \sum_{i\in\mathcal{A}(j_i)} y_i^* N = \sum_{i\in\mathcal{A}_p(j_i)}\lambda_i/\mu_i + \sum_{i\in\mathcal{A}_{lp}(j_i)} y_i^* N,$$

we then have that

$$\begin{aligned} x_i &\ge \widehat{q}_{j_i}^N - \sum_{k\in\mathcal{A}_p(j_i)}\lambda_k^N/\mu_k - [A_{lp}^{-1}A_p(\lambda^N/\mu)_{\mathcal{A}_p}]_i - (1+\|A_{lp}^{-1}\|)n\epsilon N \\ &= y_i^* N - \delta_i \log N - (1+\|A_{lp}^{-1}\|)n\epsilon N. \end{aligned}$$

Recall that, by choice, $(1+\|A_{lp}^{-1}\|)n\epsilon = \min_{i\in\mathcal{A}_{lp}} y_i^*/16$. Taking $\theta>0$ small enough, the term in parentheses in the last row of (37) satisfies, for all N large,

$$x_i\mu_i + \sum_{k\in\mathcal{A}_p(j_i)} (x_k\mu_k - e^{\theta}\lambda_k^N) \geq \mu_i y_i^* N - \mu_i(1+\|A_{lp}^{-1}\|)n\epsilon N \geq \mu_i(y_i^*/4)N.$$

Using (38) we finally have, choosing θ smaller if needed, that

$$-(1-e^{-\theta})\left(x_i\mu_i + \sum_{k\in\mathcal{A}_p(j_i)} (x_k\mu_k - e^{\theta}\lambda_k^N)\right) + \beta(x) \leq -(1-e^{-\theta})\mu_i y_i^* N/4 + |\beta(x)| \leq -\eta_1 N.$$

(ii) $\sigma_{j_i}(x) < \widehat{q}_{j_i}^N$: For $x\in\Omega_\epsilon^N$ with $\sigma_{j_i}(x)<\widehat{q}_{j_i}^N$, we have, following similar reasoning, that

$$-\left(1-e^{-\theta}\right)\left(\sum_{l\in\mathcal{A}_p(j_i)\cup\{i\}} (\lambda_l - e^{\theta}\mu_l x_l)\right) + \beta(x) \leq -\eta_2 N$$

(iii) $\sigma_{j_i}(x) = \widehat{q}_{j_i}^N$: Because $x_i \leq \widehat{q}_{j_i}^N$ for $x\in\widetilde{\mathcal{X}}$ and $x_k \leq \lambda_k^N/\mu_k + \epsilon N$ for all $x\in\Omega_\epsilon^N, k\in\mathcal{A}_p$, we have that

$$|Qg(x)| \leq (e^{\theta}+1)\sum_{k\in\mathcal{A}_p(j_i)\cup\{i\}} (\lambda_k^N + x_k\mu_k) + \beta(x) \leq \eta_3 N.$$

Overall, we have the existence of $\eta_4, \eta_5 > 0$ such that

$$(\mathcal{Q}g)(x) \leq -2\eta_4 N g(x) + \eta_5 N, \text{ for } x\in\Omega_\epsilon^N,$$

as stated. □

Proof of Lemma 5.4. This is a direct corollary of Theorem 13 of Morris and Peres (2005). Equation (46) there states that

$$\tau(\epsilon) \leq \int_{4\Pi_*}^{4/\epsilon} \frac{8\,du}{u\Phi^2(u)}.$$

where $\Phi(\cdot)$ is what they call the conductance profile. If we can identify a uniform lower bound $\Phi(u)\geq\Phi>0$ for all $u\in[4\Pi^*, 4/\epsilon]$, then

$$\tau(\epsilon) \leq \frac{8}{\Phi^2}\int_{4\Pi_*}^{4/\epsilon} \frac{8\,du}{u} = \frac{8}{\Phi^2}\log u\Big|_{4\Pi_*}^{4/\epsilon} = \frac{8}{\Phi^2}\log\frac{1}{\epsilon\Pi_*}.$$

Our Lemma 5.4 finally provides the uniform lower bound on the conductance profile $\Phi(u)$. □

Proof of Lemma 5.5. We first prove (22). Recall that the transition probability matrix P is induced by the generator matrix through $P = \mathbb{U}^{-1}\mathcal{Q} + I$, where

$$\mathbb{U} = N\sum_{i\in[n]}\lambda_i + \max_{x\in\mathcal{X}}\sum_{i\in[n]}\mu_i x_i \leq N\left(\sum_{i\in[n]}\lambda_i + \sum_{j\in[d]} q_j \max_{i\in\mathcal{A}(j)}\mu_i\right)$$

is the uniformization constant. Define

$$\bar{c} := \frac{\sum_{i\in[n]} \lambda_i + \sum_{j\in[d]} q_j \max_{i\in\mathcal{A}(j)} \mu_i}{\min\left\{\min_{i\in[n]} \mu_i, \min_{i\in[n]} \lambda_i\right\}}.$$

Transitions between state $x \in \mathcal{X}$ and $y \in \mathcal{X}$ occur either through arrivals (at rate at least $\min_i \lambda_i^N = N \min_i \lambda_i$), or through service departures at minimal rate of $\min_i \mu_i$. We then have that

$$\min_{x\neq y\in\mathcal{X}:P_{xy}>0} P_{xy} \geq \frac{\min\left\{\min_{i\in[n]} \mu_i, N\min_{i\in[n]} \lambda_i\right\}}{\mathbb{U}} \geq \frac{1}{\bar{c}N}. \tag{39}$$

There exist constants $K_1, K_2 > 0$ that do not depend on N such that

$$K_1 N \leq \max_{x,y\in\mathcal{X}} \min\{n \geq 1 : P_{xy}^n > 0\} \leq K_2 N.$$

For the upper bound, take states $x, y \in \mathcal{X}$. Then, there is a path from x to 0, through x_i service completions of type i. There is then a path from 0 to y, through y_i arrivals of type i. The total length of this path is $\sum_i x_i + \sum_i y_i \leq 2\max_j q_j^N$ and we can take $K_2 = 2\max_j q_j$. For the lower bound, take $x = 0 \in \mathbb{R}^n$ and $y = \lfloor y^* N \rfloor = (\lfloor y_1^* N\rfloor, \ldots, \lfloor y_n^* N\rfloor)$. Then, it takes at least $\sum_i \lfloor y_i^* N\rfloor$ arrivals to transition from x to y and, for all N large enough, we can take $K_1 = \frac{1}{2}\sum_i \frac{1}{2} y_i^*$.

As a result, for every $x, y \in \mathcal{X}$, we have using (39) that

$$P_{xy}^{K_2 N} \geq (\bar{c}N)^{-K_2 N}.$$

For all $m = 0, 1, 2, \ldots$, we also have

$$P^{m+K_2 N}(x,y) = \sum_{z\in\mathcal{X}} P^m(x,z) P^{K_2 N}(z,y) \geq (\bar{c}N)^{-K_2 N} \sum_{z\in\mathcal{X}} P^m(x,z) = (\bar{c}N)^{-K_2 N}.$$

In particular, regardless of the initial state x,

$$\Pi(y) = \lim_{m\to\infty} P^{m+KN}(x,y) \geq (\bar{c}N)^{-KN} \qquad \text{for all } y \in \mathcal{X}.$$

Here we used the fact that the continuous-time chain and its discrete (uniformized) counterpart have the same stationary distribution.

We turn to (23). By definition, the conductance is given by

$$\Phi = \min_{\mathcal{S}\subseteq\mathcal{X}:\Pi(S)\leq 1/2} \frac{\sum_{x\in S, y\in S^c} \Pi(x) P_{xy}}{\Pi(S)}.$$

Transitions in the Markov chain X^t are such that, for $y \neq x$, $P_{xy} > 0$ if and only if $\sum_{i\in[n]} |x_i - y_i| = 1$. Evidently, each state x has at least one such neighbor, so we know from (39) that

$$\sum_{x\in S, y\in S^c} \Pi(x) P_{xy} = \sum_{x\in S} \Pi(x) \sum_{y\in S^c} P_{xy} \geq \frac{\Pi(S)}{\bar{c}N}.$$

As a result, the conductance is lower bounded by $\Phi \geq \frac{1}{\bar{c}N}$. □

A.3. Proofs for lemmas in Section 6

Proof of Lemma 6.1. We first prove that, if (27) holds then (28) is satisfied. Indeed, if $Ax \leq q^N - \kappa \log N$ for all sufficiently large N and some $\kappa > 0$, then x_i is feasible for (LP) with the right-hand side modified to the smaller capacity $q^N - \kappa \log N = N[q - \kappa(\log N)/N]$. Let γ be the dual value of the capacity constraint in (LP). By Assumption 3.1 the primal has a unique non-degenerate solution and, in turn, so does the dual. In turn, for all large enough N, γ is the same for the right-hand sides q and $q - \kappa(\log N)/N$. We would then have that $r'_\mu x \leq \overline{\mathcal{R}}(q^N - \kappa \log N, \lambda^N/\mu) \leq \overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \gamma\kappa \log N$, contradicting (27).

Further, since there always exists a solution that satisfies strict complementarity the (unique) primal and dual solution must satisfy strict complementarity. In particular, the dual variables for the demand constraint for $i \in \mathcal{A}_p$ are strictly positive. A similar argument to the one above implies that the optimal policy must have (29), for otherwise

$$\text{if } x_i \leq \lambda_i^N/\mu_i - \kappa \log N, \text{ then } r'_\mu x \leq \overline{\mathcal{R}}(q^N, \lambda^N - e_i \kappa \log N) \leq \overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \gamma\kappa \log N,$$

for some new $\gamma > 0$.

We conclude, then, that if (27) holds, then the sequence of optimal policies must satisfy both (28) and (29). □

B. The choice of thresholds

We consider here $d = 1$ (a single resource) and provide heuristic guidance towards the choice of the threshold coefficient δ applied to the less-preferred types. The premise here is that the choice of thresholds seeks to strike a balance between the reward collected from less-preferred types and the possible loss of reward from the rejection of preferred customers. The tradeoff is non-linear: decreasing the coefficient has (approximately) a linear effect on the reward from less-preferred customers but a non-linear effect on the probability of rejecting preferred customers.

The simple calculation below is grounded in strong intuition, and the numerical evidence supports this intuition.

Construction. When all (except at most $\log N$) servers are busy approximately $y_i^* N$ of them are occupied with type-i customers. The total departure rate is then

$$\sum_i \mu_i y_i^* N = \sum_{l=1}^{i^*} \lambda_l + \mu_{i^*+1}(q^N - \sum_{l=1}^{i^*} \lambda_l/\mu_l).$$

The input rate when above the threshold (so that type $i^* + 1$ is not accepted) is $\sum_{l=1}^{i^*} \lambda_l$. Heuristically then, when above the threshold, the total number in the system behaves like an $M/M/1$ queue with utilization

$$\rho = \frac{\sum_{l=1}^{i^*} \lambda_l^N}{\sum_{l=1}^{i^*} \lambda_l^N + \mu_{i^*+1} y_{i^*+1} N} < 1.$$

Approximately $\sum_{l=1}^{n}\mu_l y_l^* N$ customers are served per unit of time out of a total of $\sum_{l=1}^{n}\lambda_l^N$ arrivals. The fraction of customers being blocked is then approximately

$$p_{block} := \frac{\sum_{l=1}^{n}\lambda_l - \sum_{l=1}^{n}\mu_l y_l^* N}{\sum_{l=1}^{n}\lambda_l^N},$$

which is approximately the likelihood that all $q^N - K$ servers are busy, where K is the threshold. Thus, the probability that *all* servers are busy–at which point preferred requests are also blocked is approximately

$$p_{block}\times\rho^K,$$

when the threshold is K. Then, we are (heuristically) looking for K that minimizes

$$(\sum_{l=1}^{i^*} r_l\lambda_l^N)p_{block}\rho^K + r_{i^*+1}\mu_{i^*+1}K.$$

This is a convex function of K. Let $\beta = \log(\frac{1}{\rho})$. Provided that $\beta\sum_{l=1}^{i^*} r_l\lambda_l^N > r_{i^*+1}\mu_{i^*+1}$ (which holds for all N large enough), we have the optimal solution

$$K = \frac{1}{\beta}\log\left(\frac{\beta p_{block}\sum_{l=1}^{i^*} r_l\lambda_l^N}{r_{i^*+1}\mu_{i^*+1}}\right). \tag{40}$$

Because $\lambda_l^N = N\lambda_l$, we have that

$$K = \frac{1}{\beta}\log N + \Gamma,$$

where Γ is the constant $\frac{1}{\beta}\log\left(\frac{\beta p_{block}\sum_{l=1}^{i^*} r_l\lambda_l}{r_{i^*+1}\mu_{i^*+1}}\right)$ that does not scale with N.

Notice that the closer that ρ is to 1, the closer β is to 0 and the larger the coefficient of $\log N$ has to be.

As a sanity check we note that this matches the exact asymptotic coefficient derived in Morrison (2010) for the model with two classes that have equal service rates ($\mu_1 = \mu_2$); see Theorem 1 there; our β is the same as $-\log\sigma$ there.

For a numerical illustration, we consider the system with three types and the following parameters $\lambda = (1,1,1)$, $\mu = (1,1/2,1/3)$, $r = (1,1,1)$ and $q = 4$. The optimal solution to the LP has

$$y^* = (1,2,1), \overline{\mathcal{R}}(q,\lambda/\mu) = 7/3.$$

For these parameters we get $K = 30.05$ so that $\delta = 30.05/\log N = 4.84$) for $N = 500$ (corresponding and $K = 34.55$ for $N = 1000$ (so that $\delta = 34.55/\log N = 5$). In Figure 10[TOP] we plot the log-regret as a function of $K/\log N$ (a proxy for the threshold coefficient). For both $N = 500$ and $N = 1000$, we see that the best coefficient in the simulation is 4 (one below our recommendation), but 5 produces close performance. Indeed, all of $3, 4$, and 5 produce a similar performance.

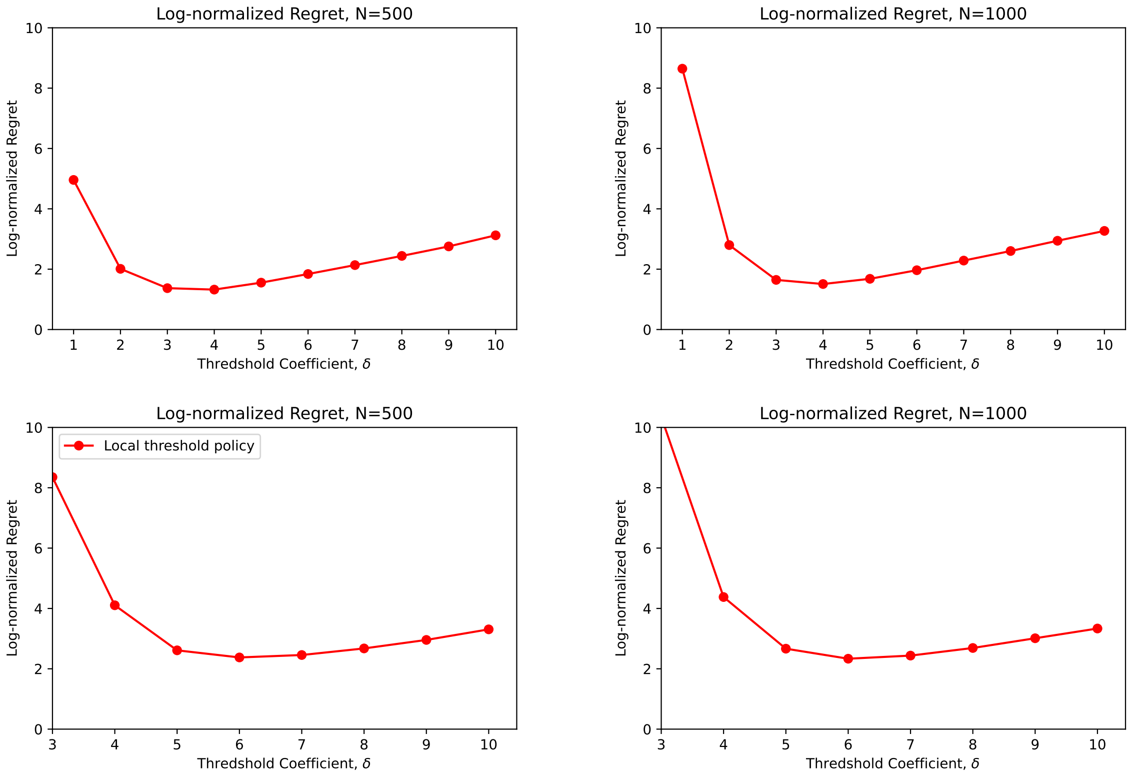


Figure 10 Log-regret as a function of the threshold coefficient δ for the reward vector $r = (1,1,1)$ (TOP) and $r = (10,5,1)$ (BOTTOM)

To show that the threshold and its performance respond as expected to changes in rewards, we repeat the same experiment, but now with the reward vector $r = (10,5,1)$. The LP solution has the same decision values and the larger objective value $46/3$. The thresholds, as expected, increase because the preferred types are more valuable. Specifically, for $N = 500$, the optimal threshold in (40) is $K = 43.12$ corresponding to a coefficient of approximately $\delta = 6.94$. Again, our heuristic would have regret very similar to the simulation-optimal one, which is 7; see Figure 10[BOTTOM]. This is true also for $N = 1000$ where our heuristic recommends $\delta = 6.9$.

C. The insufficiency of static policies

We consider the case of a single resource ($d = 1$). Our goal here is to show that logarithmic regret is not achievable with a static policy. In other words, for logarithmic regret, the accept/reject decision must be state-dependent. A state *independent* policy is characterized by a vector $p = (p_1, \ldots, p_n)$ of probabilities (i.e., $p_i \in [0,1]$). When a type-i request arrives, it is accepted with probability p_i if there are units of the resource available and it is declined otherwise. We denote by $\pi_S(p)$ the static policy using p.

Lemma C.1 *Suppose that $d=1$ (single resource) and that Assumption 3.1 holds. Then, for any sequence of vectors $p^N \in [0,1]^n$,*

$$\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathcal{R}^{\pi_S(p^N)} = \Omega(\sqrt{N}). \tag{41}$$

Proof: Fix the sequence $\{p^N\}$ and let X_i^N be the steady-state number of type-i customers under p^N (and with arrival rate λ^N and q^N resource units). We let $\Pi_S(p^N)$ be the stationary distribution under the policy $\pi_S(p^N)$. The existence of a steady state under such a policy is straightforward. Under such a static policy, the headcount $\Sigma^N = \sum_{i\in[n]} X_i^N$ in the system has the stationary law of an $M/G/q^N/q^N$ with arrival rate $\sum_{i\in[n]} \lambda_i^N p_i^N$, q^N servers, and the service time distribution is a mixture of exponential distributions with mean

$$\frac{1}{\bar{\mu}^N} := \frac{\sum_{i\in[n]} \frac{\lambda_i}{\mu_i} p_i^N}{\sum_{i\in[n]} \lambda_i p_i^N}.$$

Requests can be accepted only when the headcount is strictly below q^N. By PASTA, type-i requests have the (stationary) admission probability $p_i^N \mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N < q^N\}$. By Little's law the number of type-i customers in service is then $x_i^N = \mathbb{E}_{\Pi_S(p^N)}[X_i^N] = \frac{\lambda_i^N}{\mu_i} p_i^N \mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N < q^N\}$. By non-degeneracy we must have that either

$$x_i^N = \frac{\lambda_i^N}{\mu_i} + o(\sqrt{N}),\ i \in \mathcal{A}_p, \tag{42}$$

or that (41) holds; this is argued as in the proof of Lemma 6.1. Assume, then, that (42) holds. It must then be that $p_i^N \mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N < q^N\} = 1 - o(1/\sqrt{N})$ and, in particular, that both

$$1 - p_i^N = o(1/\sqrt{N}),\ i \in \mathcal{A}_p, \text{ and } \mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N = q^N\} = 1 - \mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N < q^N\} = o(1/\sqrt{N}).$$

Consider a sequence of $M/G/q^N/q^N$ queues with total arrival rate $\bar{\lambda}^N = \sum_{i\in[n]} \lambda_i^N p_i^N$, mean service time $1/\bar{\mu}^N$. Then, it is known (e.g., Janssen et al. (2008)) that

$$\mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N = q^N\} = o(1/\sqrt{N}) \text{ if and only if } \bar{\lambda}^N/\bar{\mu}^N = q^N - \Omega(\sqrt{N}).$$

In turn,

$$\sum_{i\in[n]} x_i^N = \sum_{i\in[n]} \frac{\lambda_i^N}{\mu_i} p_i^N \mathbb{P}_{\Pi_S(p^N)}\{\Sigma^N < q^N\} \le \sum_{i\in[n]} \frac{\lambda_i^N}{\mu_i} p_i^N \le q^N - \Omega(\sqrt{N}).$$

By non-degeneracy, and following the argument in the proof of Lemma 6.1, we then have

$$\overline{\mathcal{R}}(q^N, \lambda^N/\mu) - \mathcal{R}^{\pi_S(p^N)} = \Omega(\sqrt{N}),$$

as stated. □

References

Arlotto, A. and Gurvich, I. (2019). Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*, 9(3):231–260.

Asmussen, S. (2003). *Applied Probability and Queues*, volume 2. Springer.

Baek, J. and Ma, W. (2019). Bifurcating constraints to improve approximation ratios for network revenue management with reusable resources. *Available at SSRN 3482457*.

Bean, N. G., Gibbens, R. J., and Zachary, S. (1995). Asymptotic analysis of single resource loss systems in heavy traffic, with applications to integrated networks. *Adv. in Appl. Probab.*, 27(1):273–292.

Bean, N. G., Gibbens, R. J., and Zachary, S. (1997). Dynamic and equilibrium behavior of controlled loss networks. *Ann. Appl. Probab.*, 7(4):873–885.

Besbes, O., Elmachtoub, A. N., and Sun, Y. (2021). Static pricing: Universal guarantees for reusable resources. *Operations Res.*, Articles in Advance.

Bumpensanti, P. and Wang, H. (2020). A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Sci.*, 66(7):2993–3009.

Cao, H., Hu, J., Jiang, C., Kumar, T., Li, T.-H., Liu, Y., Lu, Y., Mahatma, S., Mojsilović, A., Sharma, M., et al. (2011). Onthemark: Integrated stochastic resource planning of human capital supply chains. *Interfaces*, 41(5):414–435.

Chen, Y., Levi, R., and Shi, C. (2017). Revenue management of reusable resources with advanced reservations. *Prod. Oper. Manag.*, 26(5):836–859.

Dawande, M., Feng, Z., and Janakiraman, G. (2021). On the structure of bottlenecks in processes. *Management Science*, 67(6):3853–3870.

Feng, Y., Niazadeh, R., and Saberi, A. (2019). Linear programming based online policies for real-time assortment of reusable resources. *Available at SSRN 3421227*.

Gallego, G. and van Ryzin, G. (1997). A multiproduct dynamic pricing problem and its applications to network yield management. *Oper. Res.*, 45(1):24–41.

Glynn, P. and Zeevi, A. (2008). Bounding stationary expectations of Markov processes. In *Markov processes and related topics: A Festschrift for Thomas G. Kurtz. Selected papers of the conference, Madison, WI, USA, July*, volume 4, pages 195–214.

Gong, X.-Y., Goyal, V., Iyengar, G. N., Simchi-Levi, D., Udwani, R., and Wang, S. (2021). Online assortment optimization with reusable resources. *Management Sci.*, Articles in Advance.

Goyal, V., Iyengar, G., and Udwani, R. (2020). Online allocation of reusable resources: Achieving optimal competitive ratio. *arXiv preprint arXiv:2002.02430*.

Gurvich, I. and Intelligent Automation (2018-2021). Dragons – dynamic resource allocation gains for operational networked sharing, department of defense (army) sttr a18b-t007.

Gurvich, I. and Van Mieghem, J. A. (2015). Collaboration and multitasking in networks: Architectures, bottlenecks, and capacity. *Manufacturing & Service Operations Management*, 17(1):16–33.

Hu, J., Lu, Y., Mojsilović, A., Sharma, M., and Squillante, M. S. (2010). Performance management of it services delivery. *ACM SIGMETRICS Performance Evaluation Review*, 37(4):50–57.

Hui, J. Y. (2012). *Switching and traffic theory for integrated broadband networks*, volume 91. Springer Science & Business Media.

Hunt, P. J. and Kurtz, T. G. (1994). Large loss networks. *Stochastic Process. Appl.*, 53(2):363–378.

Hunt, P. J. and Laws, C. N. (1997). Optimization via trunk reservation in single resource loss systems under heavy traffic. *Ann. Appl. Probab.*, 7(4):1058–1079.

Iyengar, G. and Sigman, K. (2004). Exponential penalty function control of loss networks. *Ann. Appl. Probab.*, 14(4):1698–1740.

Janssen, A. J. E. M., Van Leeuwaarden, J., and Zwart, B. (2008). Gaussian expansions and bounds for the poisson distribution applied to the erlang b formula. *Advances in Applied Probability*, 40(1):122–143.

Jasin, S. and Kumar, S. (2012). A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Math. Oper. Res.*, 37(2):313–345.

Jia, H., Shi, C., and Shen, S. (2022). Online learning and pricing for service systems with reusable resources. *Operations Research*.

Jung, K., Lu, Y., Shah, D. D., Sharma, M., and Squillante, M. (2019). Revisiting stochastic loss networks: Structures and approximations. *Mathematics of Operations Research*, 44(3):890–918.

Kelly, F. P. (1986). Blocking probabilities in large circuit-switched networks. *Adv. in Appl. Probab.*, 18(2):473–505.

Kelly, F. P. (1991). Loss networks. *Ann. Appl. Probab.*, 1(3):319–378.

Key, P. B. (1990). Optimal control and trunk reservation in loss networks. *Probability in the Engineering and Informational Sciences*, 4(2):203–242.

Lei, Y. M. and Jasin, S. (2020). Real-time dynamic pricing for revenue management with reusable resources, advance reservation, and deterministic service time requirements. *Operations Res.*, 68(3):676–685.

Levi, R. and Radovanović, A. (2010). Provably near-optimal LP-based policies for revenue management in systems with reusable resources. *Operations Res.*, 58(2):503–507.

Lippman, S. A. (1975). Applying a new device in the optimization of exponential queuing systems. *Operations Res.*, 23(4):687–710.

Miller, B. L. (1969). A queueing reward system with several customer classes. *Management Sci.*, 16(3):234–245.

Morris, B. and Peres, Y. (2005). Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields*, 133(2):245–266.

Morrison, J. A. (2010). Optimal trunk reservation for an overloaded link. *Oper. Res. Lett.*, 38(6):499–501.

Motwani, R. and Raghavan, P. (1995). *Randomized Algorithms*. Cambridge University Press.

Örmeci, E. L., Burnetas, A., and van der Wal, J. (2001). Admission policies for a two class loss system. *Stoch. Models*, 17(4):513–539.

Örmeci, E. L. and van der Wal, J. (2006). Admission policies for a two class loss system with general interarrival times. *Stoch. Models*, 22(1):37–53.

Owen, Z. and Simchi-Levi, D. (2018). Price and assortment optimization for reusable resources. *Available at SSRN 3070625*.

Paschalidis, I. and Liu, Y. (2002). Pricing in multiservice loss networks: static pricing, asymptotic optimality and demand substitution effects. *IEEE/ACM Transactions on Networking*, 10(3):425–438.

Paschalidis, I. and Tsitsiklis, J. (2000). Congestion-dependent pricing of network services. *IEEE/ACM Transactions on Networking*, 8(2):171–184.

Puhalskii, A. A. and Reiman, M. I. (1998). A critically loaded multirate link with trunk reservation. *Queueing Syst.*, 28(1-3):157–190.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Reiman, M. I. (1991). Optimal trunk reservation for a critically loaded link. In Jensen, A. and Iverson, V. B., editors, *Teletraffic and Datatraffic: In a Period of Change*, pages 247–252. North Holland.

Reiman, M. I. and Wang, Q. (2008). An asymptotically optimal policy for a quantity-based network revenue management problem. *Math. Oper. Res.*, 33(2):257–282.

Ross, K. and Tsang, D. (1989). Optimal circuit access policies in an ISDN environment: a Markov decision approach. *IEEE Transactions on Communications*, 37(9):934–939.

Ross, S. (1996). *Stochastic Processes*. John Wiley & Sons.

Rusmevichientong, P., Sumida, M., and Topaloglu, H. (2020). Dynamic assortment optimization for reusable products with random usage durations. *Management Sci.*, 66(7):2820–2844.

Talluri, K. T. and van Ryzin, G. J. (2004). *The Theory and Practice of Revenue Management*. International Series in Operations Research & Management Science, 68. Kluwer Academic Publishers, Boston, MA.

Tsitsiklis, J. N. and Xu, K. (2017). Flexible queueing architectures. *Operations Research*, 65(5):1398–1413.

Tutte, W. T. (1947). The factorization of linear graphs. *Journal of the London Mathematical Society*, s1-22(2):107–111.

Van der Boor, M., Borst, S. C., Van Leeuwaarden, J. S., and Mukherjee, D. (2018). Scalable load balancing in networked systems: Universality properties and stochastic coupling methods. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3893–3923. World Scientific.

Vanderbei, R. J. (1998). *Linear programming - foundations and extensions*, volume 4 of *Kluwer international series in operations research and management service*. Kluwer.

Vera, A. and Banerjee, S. (2021). The bayesian prophet: A low-regret framework for online decision making. *Management Science*, 67(3):1368–1391.

Vera, A., Banerjee, S., and Gurvich, I. (2021). Online allocation and pricing: Constant regret via bellman inequalities. *Operations Research*, 69(3):821–840.

Williamson, E. L. (1992). *Airline network seat inventory control: Methodologies and revenue impacts*. PhD thesis, Massachusetts Institute of Technology.

Xu, H. and Li, B. (2013). Dynamic cloud pricing for revenue maximization. *IEEE Transactions on Cloud Computing*, 1(2):158–171.