The Annals of Applied Probability
1997, Vol. 7, No. 4, 1058–1079

OPTIMIZATION VIA TRUNK RESERVATION IN SINGLE RESOURCE LOSS SYSTEMS UNDER HEAVY TRAFFIC

BY P. J. HUNT AND C. N. LAWS

ABN AMRO Bank and University of Oxford

Trunk reservation is a simple, robust and extremely effective mechanism for controlling loss systems which allows priority to be given to chosen traffic streams. We consider the control of a single resource under a limiting regime in which capacity and arrival rates increase together. We obtain trunk reservation control policies which are asymptotically optimal when calls have differing capacity requirements, holding times, arrival rates and reward rates. The priority levels associated with these trunk reservation policies arise from an attainable bound on the performance of any control policy.

1. Introduction. The use of transmission capacity in telecommunications networks must be carefully controlled in order to ensure that a network's users receive a satisfactory quality of service. The choice of control policy is important because different policies can give markedly different behavior and the performance of individual control strategies may well be counterintuitive [4, 6]. Strategies based on *trunk reservation*, such as those considered in this paper, are simple, robust and yet extremely effective control mechanisms which essentially assign different levels of priority to different types of traffic.

Loss networks are widely used to model systems at which demands arrive and request use of some of the system's capacity. An arriving demand may either be accepted by the system or be rejected and lost from the system. Trunk reservation control policies provide a simple means of deciding when calls should be accepted: a decision at time t depends only on the free capacity of each resource of a network at time t, and not on a detailed description of the entire state of the network. Further, such policies provide a very robust control mechanism: given only that trunk reservation parameters are chosen within some sensible range, the behavior observed is typically rather insensitive to call arrival rates. Our main aim is to show that, for a general model of a single resource, trunk reservation is asymptotically optimal under a limiting regime of interest in applications: namely, a regime in which capacity and arrival rates increase together. Central to our approach is a result of Hunt and Kurtz [12]. The paper of Bean, Gibbens and Zachary [2] is closely related to ours, the main difference being that their work is more concerned with modelling the performance of trunk reservation than with optimization. Throughout our notation follows closely that of [2] and [12].

Received September 1995; revised May 1997.
AMS 1991 *subject classifications.* Primary 60K30; secondary 60K20, 90B22, 93E20.
Key words and phrases. Loss network, optimization, trunk reservation, time-scale separation.

We now describe the single resource which is the focus of this paper. Consider a resource of integer capacity C. Suppose that the finite set I indexes the types of call offered to the resource. For $i \in I$, calls of type i arrive as a Poisson stream of rate κ_i. When a call arrives it is a control decision as to whether the call should be accepted by the resource. An arriving call must be rejected and lost from the system if the resource has insufficient free capacity to accept it, and calls may also be rejected in other situations. A call of type i requires use of A_i units of capacity (A_i a positive integer) and, if accepted, will use this amount of resource for the holding time of the call, after which this capacity is released for use by other calls. The holding time of a type i call is exponentially distributed with mean μ_i^{-1}. All call arrival streams and holding times are independent. An accepted call of type i earns reward at a rate of w_i per unit time in progress. We assume, without loss of generality, that the greatest common divisor of the capacity requirements A_i, $i \in I$, is equal to 1 (this ensures that the free capacity of the resource evolves as an irreducible stochastic process).

A control policy determines, for each arriving call, whether or not to accept the call. The aim is to maximize the expected reward received per unit time. Control policies may, in principle, have full information on all previous call arrivals, departures and acceptance decisions, although they may not have future knowledge, nor do we allow a new arrival to cause a call in progress to be terminated prematurely. The form of a trunk reservation policy is: accept an arriving call of type i if and only if the resource currently has at least r_i units of free capacity. Here r_i is known as the trunk reservation parameter used against calls of type i.

The resource described above corresponds to a loss network of a single link in the terminology of Kelly [16]; see that paper for a wide-ranging review of the modelling and optimization of loss networks. For work on trunk reservation in loss networks, including the robustness and effective performance of trunk reservation policies under varying call arrival rates, see [2], [7], [8], [18], [19], [24] and [25]. For asymptotic optimality of trunk reservation in limiting regimes other than that considered in this paper, see [13], [16] and [19].

Some results about the exact optimality (or not) of trunk reservation policies are available. Suppose that $A_i = A$ and $\mu_i = \mu$ for all i: without loss of generality, $A = 1$ and $\mu = 1$. Then the policy that maximizes the expected reward rate is a trunk reservation policy, with multiple priority levels, as follows [23]. When a type i call arrives, accept it if there are at least r_i units of free capacity, and reject it otherwise. That is, type i calls are accepted subject to a trunk reservation parameter of r_i. If call types are labelled so that $w_1 \geq w_2 \geq \cdots \geq w_I$, where $I = |I|$, then $r_1 \leq r_2 \leq \cdots \leq r_I$, and $r_1 = 1$ (since an optimal policy will clearly accept some calls when one or more units of capacity are free). The optimal choices of trunk reservation parameters r_i, $i \in I$, can be determined from the stationary distribution of the birth–death process describing the number of units of capacity in use, or via policy improvement. Optimal values of trunk reservation parameters are studied in detail in [18] and [28], where various limiting regimes are considered. Lippman [21] has shown that the same form of policy is optimal under a variety of discount and finite-horizon criteria. Also, Nguyen [26] considers a closely related problem with $I = 2$ but where one arrival stream is Poisson and the other is an overflow stream from an $M/M/m/m$ queue: again the optimal policy is of (generalized) trunk reservation form.

The exact optimality of trunk reservation does not extend to networks: see [18] for a counterexample with two resources. Neither does exact optimality extend to the general model of a resource described above: trunk reservation is not necessarily optimal if the capacity requirements A_i depend on i [31], nor if the mean holding times of calls μ_i^{-1} depend on i [19]. However, the case where the values of A_i and μ_i vary with i (perhaps by orders of magnitude) is especially relevant in applications. In the context of multiservice telecommunications networks, the capacity requirements A_i, $i \in I$, are often called *effective bandwidths* [5, 10, 15]: these provide an accurate assessment of the capacity required (at each resource of a network) by calls in order to ensure that certain constraints relating to quality of service are satisfied (e.g., constraints on loss or delay). We examine the behavior of the resource described above under trunk reservation policies, and so our results have applications to systems in which traffic of different types is integrated. In particular, we obtain results on asymptotic behavior and asymptotic optimality: these results are approximate for systems with high capacity and are therefore relevant for modern systems. Although the results do not provide a concrete guideline for setting trunk reservation parameters, they do suggest an insensitivity to the value of such parameters.

The paper is organized as follows. In the following section we describe an upper bound on the expected reward rate of any control policy; in later sections we prove asymptotic optimality by showing that this bound is attained in the limit as capacity and arrival rates increase together. In Section 2 we also note (from [18]) that the upper bound is attained, asymptotically, by a randomized control policy. However, this policy has some undesirable features, in particular a lack of robustness to varying call arrival rates, and so, for the remainder of the paper, we concentrate on trunk reservation strategies. In Section 3 we define a trunk reservation strategy with three priority levels, after partitioning call types into sets of high, medium and low priority types. We then apply a result of Hunt and Kurtz [12] to obtain an asymptotic description of the behavior of this strategy. The call acceptance probabilities appearing in this description can be calculated explicitly (Section 3) and, in Section 4, they play an important role in the proof of asymptotic optimality of our three-level strategy. In Section 4 we also observe that asymptotic optimality can sometimes be achieved with two levels of trunk reservation, but not, in general, with a single level. Section 5 contains generalizations of the results in Sections 3 and 4: it allows call types to be partitioned into arbitrarily many priority levels. In particular we show that a policy that uses a trunk reservation parameter proportional to A_i/w_i is asymptotically optimal, provided that the constant of proportionality is suitably chosen to reflect the size of the resource under consideration. This last result is further evidence of the robustness of trunk reservation policies to call arrival rates: no information about arrival rates is needed in order to specify an asymptotically optimal policy. Finally, we mention briefly some of the difficulties involved when trying to extend our results to networks.

2. A performance bound. We start this section with an upper bound on the performance of the resource described in the Introduction. We restrict attention to control policies which are time-independent and which depend only on the current state of the system, where the state of the system is the number of type i calls in progress, for all $i \in I$. When considering optimal policies (or upper bounds on the performance of any policy), making this restriction and thus considering only stationary, Markov policies loses no generality here, from the general theory of Markov decision processes (see, e.g., [33]).

The following is (a special case of) a simple adaptation of a result in [7] (cf. [18], Lemma 2.1). Here and throughout, unrestricted summations over i range over I. Let $\nu_i = \kappa_i/\mu_i$.

THEOREM 2.1. *Under any policy, the expected reward received per unit time is bounded above by the value attained in the maximum flow problem,*

$$(MF) \quad \text{maximize} \quad \sum_i w_i x_i, \tag{2.1a}$$

$$\text{subject to} \quad \sum_i A_i x_i \le C, \tag{2.1b}$$

$$0 \le x_i \le \nu_i, \qquad i \in I. \tag{2.1c}$$

An interpretation of this result is as follows. Let x_i be the average number of type i calls in progress under a policy. Then $\sum_i w_i x_i$ is the expected reward rate. Constraint (2.1b) must be satisfied since the average amount of capacity in use cannot exceed the capacity C of the resource. Also, the number of type i calls in progress is dominated by the number of customers in an $M/M/\infty$ queue with arrival and service rates κ_i and μ_i, since this queue corresponds to the number of type i calls that would be in progress if all type i calls were accepted. Hence we obtain (2.1c).

For networks of resources with alternative routing, more refined bounds than that arising from the natural generalization of problem MF are obtained by Kelly [17] and investigated in [9]. Under the limiting regime that we consider, the bound given by problem MF is tight.

The regime we consider is that studied in detail by Kelly [14, 16] and termed the heavy traffic limit by Whitt [32]. Consider a sequence of systems indexed by N. In the Nth system replace C by $C(N)$, κ_i by $\kappa_i(N)$ and μ_i by $\mu_i(N)$, where

$$C(N)/N \to C, \qquad \kappa_i(N)/N \to \kappa_i, \qquad \mu_i(N) \to \mu_i, \tag{2.2}$$

and all limits are as $N \to \infty$. Let the random variable $n_i^{\pi}(N)$ denote the number of type i calls in progress in equilibrium when the Nth system is operated under a control policy π. (Here and below, a policy π may depend on N, though we do not indicate this explicitly.) Also define the expected reward rate when the Nth system is operated under policy π by $R^{\pi}(N) = \mathbf{E}[\sum_i w_i n_i^{\pi}(N)]$. If MF$(N)$ denotes problem MF but with C and ν_i replaced by $C(N)$ and $\nu_i(N) = \kappa_i(N)/\mu_i(N)$, then applying Theorem 2.1 gives $R^{\pi}(N) \le \sum_i w_i \bar{x}_i(N)$, where $\bar{x}(N) = (\bar{x}_i(N),\ i \in I)$ is any optimal solution of MF(N). The normalization considered throughout the paper (cf. [14, 16]) is that obtained by dividing the reward rate of the Nth system by N (or, equivalently, dividing the number of calls in progress of each type by N). On this scale, the bound on the reward rate of the Nth system is

$$\frac{R^{\pi}(N)}{N} \le \sum_i w_i \frac{\bar{x}_i(N)}{N}. \tag{2.3}$$

The solution $\bar{x}(N)$ of MF(N) is not unique in general [since MF(N) is a linear program] and $\lim_{N\to\infty} \bar{x}_i(N)/N$ may not exist. However, since $\bar{x}(N)/N$ solves problem MF but with C and ν_i replaced by $C(N)/N$ and $\nu_i(N)/N$, we can (and do) choose $\bar{x}(N)$ so that $\bar{x}_i(N)/N$ converges to a limit as $N \to \infty$ for all $i \in I$. [This follows because (2.2) ensures that $C(N)/N$ and $\nu_i(N)/N$, for $i \in I$, have limits as $N \to \infty$.] Further, if we define $\bar{x}_i = \lim_{N\to\infty} \bar{x}_i(N)/N$, then $\bar{x} = (\bar{x}_i,\ i \in I)$ is optimal for MF. So taking limits in (2.3) gives

$$\limsup_{N\to\infty} \frac{R^{\pi}(N)}{N} \le \sum_i w_i \bar{x}_i. \tag{2.4}$$

Hence the optimal value of MF provides a bound on the normalized reward rate of any control policy, in the limit $N \to \infty$.

Now fix $\bar{x}$ a solution to MF, let $p_i = \bar{x}_i/\nu_i$ and consider the following randomized control policy. When a type i call arrives, reject the call with probability $1 - p_i$; otherwise accept the call if there is sufficient free capacity to carry it, and reject it if not. The following result, a special case of Theorem 2.1 of [18], shows that this policy is asymptotically optimal in that it attains bound (2.4).

THEOREM 2.2. *The above call acceptance policy (policy π_0, say) is asymptotically optimal; that is,*

$$\lim_{N\to\infty} \frac{R^{\pi_0}(N)}{N} = \sum_i w_i \bar{x}_i.$$

Interpret ν_i as the average amount of type i traffic available. Then from Theorems 2.1 and 2.2, if $\sum_i A_i \nu_i \le C$, we can ensure, asymptotically, that $\bar{x}_i = \nu_i$ for all i; that is, all available traffic is accepted by the resource. So the most interesting case is that where $\sum_i A_i \nu_i > C$, when some traffic must be rejected; in [2] the term "heavy traffic" refers to this case.

Although policy π_0 is asymptotically optimal, the policy has undesirable features when considering applications. To evaluate probabilities p_i, problem MF must be solved. This requires a large amount of information, such as call arrival rates and mean holding times, to be known or estimated. Most importantly, arrival rates κ_i will not be known in practice and would need to be estimated. Key [18] illustrates how such policies based on incorrect values of κ_i can give very poor performance. Further, other policies which are asymptotically optimal may well give higher reward rates for more realistic capacities and arrival rates. Trunk reservation policies, which we now consider, have such advantages and overcome the above difficulties.

3. Heavy traffic behavior. In this section we consider the behavior of our resource under a trunk reservation control policy. For networks of resources, the asymptotic behavior was first discussed heuristically by Kelly [14] (for the case of no trunk reservation) and is justified by Hunt and Kurtz [12]. See also [2] for careful discussion of the asymptotic behavior of a single resource with fixed trunk reservation parameters: in addition, work in [1] develops results of [12] further.

Let $\bar{x}$ be some fixed solution of problem MF. Define the sets H, M and L by

$$H = \{i \in I : \bar{x}_i = \nu_i\},$$
$$M = \{i \in I : 0 < \bar{x}_i < \nu_i\},$$
$$L = \{i \in I : \bar{x}_i = 0\}.$$

Interpret H, M and L as partitioning the set of call types I into sets of high, medium and low priority types. Observe that one, but not both, of sets H and M may be empty; independently of this, L may be empty.

Suppose that the Nth system operates according to the following three-level trunk reservation strategy. Calls of type $i \in H$ are accepted subject to a trunk reservation parameter of $r_1 = \max_{i \in H} A_i$; calls of type $i \in M$ are accepted subject to a trunk reservation parameter of $r_2(N) = \log N$; and calls of type $i \in L$ are accepted subject to a trunk reservation parameter of $r_3(N) = 2 \log N$. An application of a result of [12] allows us to obtain a description of the asymptotic behavior under this control policy. In Section 4 we show that this policy is asymptotically optimal.

REMARKS. (i) If $r_2(N)$ and $r_3(N)$ are fixed as N increases, then bound (2.4) will not be attained in general, as positive proportions of high, medium and low priority calls will be rejected (cf. [12]).

(ii) Observe that r_1 does not vary with N and that its effect is to accept a high priority call only when there is sufficient free capacity to accept a call of any type $i \in H$. So call acceptance probabilities are equal for all types $i \in H$, and this ensures that these probabilities can be calculated explicitly (see below). Similarly, acceptance probabilities are equal (and can be calculated explicitly) for all types $i \in M$, and also for all types $i \in L$.

(iii) The choices $r_2(N) = \log N$ and $r_3(N) = 2\log N$ are not the only ones suitable for our purpose. Any choices such that $r_2(N)$, $r_3(N)$ and $r_3(N) - r_2(N)$ increase without bound and such that these three quantities are $o(N)$ as $N \to \infty$ would suffice [$r_2(N) = o(N)$ is also necessary, although it is not necessary that $r_3(N) = o(N)$]. The results below would all follow, as the result of [12] that we apply is unaltered in such cases. We choose $\log N$ for concreteness and by comparison with [18] and [28], where optimal trunk reservation parameters are shown to be $O(\log N)$ in several cases.

Let $n_i^N(t)$ be the number of type i calls in progress at time t in the Nth system under the above trunk reservation strategy. Let $x^N(t) = (x_i^N(t), i \in I)$, where $x_i^N(t) = n_i^N(t)/N$. Then $x^N(\cdot)$ is a Markov process with state space $X = \{x: \sum_i A_i x_i \le C,\ x_i \ge 0 \text{ for all } i \in I\}$. Let $\Rightarrow$ denote weak convergence and assume that $x^N(0) \Rightarrow x(0)$ as $N \to \infty$. Then apply Theorem 3 of [12] and its extension to the case where trunk reservation parameters vary with N ([12], Section 3): the sequence of processes $\{x^N(\cdot)\}$ is relatively compact, and the limit $x(\cdot)$ of any convergent subsequence satisfies

$$x_i(t) = x_i(0) + \int_0^t v_i(x(u))\, du, \tag{3.1}$$

where

$$v_i(x) = \kappa_i P_i(x) - \mu_i x_i \tag{3.2}$$

and where the call acceptance probabilities $P_i(x)$ are still to be defined.

The dynamics of the limit process $x(\cdot)$, given by (3.1), are straightforward away from the boundary set $B = \{x \in X: \sum_i A_i x_i = C\}$. If $x \notin B$, then $P_i(x) = 1$ for all $i \in I$: for such x the rate of change x_i is $v_i(x) = \kappa_i - \mu_i x_i$, where the κ_i term is due to type i arrivals (which are all accepted since $x \notin B$) and the $-\mu_i x_i$ term is due to type i departures.

Define the free capacity at time t by $m_1^N(t) = C(N) - \sum_i A_i n_i^N(t)$. Also define $m_2^N(t) = m_1^N(t) - \log N$ and $m_3^N(t) = m_1^N(t) - 2\log N$. In terms of these three processes, our trunk reservation strategy is as follows: at time t, accept an arrival of type $i \in H$ if $m_1^N(t) \ge r_1$; accept an arrival of type $i \in M$ if $m_2^N(t) \ge 0$; and accept an arrival of type $i \in L$ if $m_3^N(t) \ge 0$. Markov processes closely related to these three processes play an important role in defining probabilities $P_i(x)$, but first we describe the link between $m_2^N(\cdot)$ and $x(\cdot)$. There are very similar links between $m_1^N(\cdot)$, $m_3^N(\cdot)$ and $x(\cdot)$. In the case of no medium or low priority call types ($M = L = \varnothing$), the link between $m_1^N(\cdot)$ and $x(\cdot)$ is a special case of discussion in [2]; see also [20].

By considering the states in which calls of different priorities are rejected, we find that the transition rates of $m_2^N(\cdot)$, at time t, are given by

$$m_2^N \to \begin{cases} m_2^N - A_i, & \text{at rate } \kappa_i(N) I\{m_2^N \ge -\log N + r_1\}, \text{ if } i \in H, \\ m_2^N - A_i, & \text{at rate } \kappa_i(N) I\{m_2^N \ge 0\}, \text{ if } i \in M, \\ m_2^N - A_i, & \text{at rate } \kappa_i(N) I\{m_2^N \ge \log N\}, \text{ if } i \in L, \\ m_2^N + A_i, & \text{at rate } N\mu_i(N) x_i^N(t), \text{ for each } i \in I, \end{cases}$$

where $I\{\cdot\}$ denotes an indicator function. All of these transition rates are $O(N)$, and so $m_2^N(\cdot)$ evolves rapidly. In contrast, over time periods of $o(1)$, $x^N(\cdot)$ remains approximately constant (owing to the normalization). Now define the set

$$X_2 = \left\{ x \in X\colon \sum_{i \in H} A_i \kappa_i < \sum_{i \in I} A_i \mu_i x_i < \sum_{i \in H \cup M} A_i \kappa_i \right\},$$

suppose that $x(t) = x \in X_2 \cap B$ and consider the Nth system over the short time interval $[t, t + \varepsilon]$. From the definition of X_2 and the $O(N)$ transition rates, $m_2^N(\cdot)$ will satisfy $m_2^N(\cdot) < \log N$ for almost all of this interval [since $|m_2^N(\cdot)|$ has strictly negative drift when $x^N(\cdot) \in X_2$ and $m_2^N(\cdot) \neq 0$], and the number of calls of type $i \in L$ accepted over the interval will be $o(N)$ so that the fraction of such calls accepted is $o(1)$. In contrast, the fraction of type $i \in H$ calls accepted will be $1 - o(1)$ since $m_2^N(\cdot)$ will satisfy $m_2^N(\cdot) \geq -\log N + r_1$ for almost all of the interval $[t, t + \varepsilon]$.

Now, in order to consider calls of type $i \in M$, for each $x \in X_2 \cap B$ let π_x^2 be the stationary distribution of the Markov process $m_2(\cdot)$ on $\mathbb{Z}$ whose transition rates are given by

$$(3.3) \qquad m_2 \to \begin{cases} m_2 - A_i, & \text{at rate } \kappa_i, \text{ if } i \in H, \\ m_2 - A_i, & \text{at rate } \kappa_i I\{m_2 \geq 0\}, \text{ if } i \in M, \\ m_2 + A_i, & \text{at rate } \mu_i x_i, \text{ for each } i \in I. \end{cases}$$

(The condition $x \in X_2 \cap B$ ensures that the stationary distribution π_x^2 exists; see Theorem 3.1 below.) Observe the close relation of this process to $m_2^N(\cdot)$: transition rates have been normalized, and indicator functions involving $\log N$ replaced by 0 or 1 (as suggested by the above discussion). For finite N, a medium priority call is accepted at time t if and only if $m_2^N(t) \geq 0$. Owing to the fast evolution of $m_2^N(\cdot)$ and the condition $x \in X_2 \cap B$, $m_2^N(\cdot)$ behaves approximately as a Markov process with stationary distribution π_x^2 over the interval $[t, t + \varepsilon]$. Thus, in the Nth system, the proportion of medium priority calls accepted over $[t, t + \varepsilon]$ is approximately $\sum_{j \geq 0} \pi_x^2(j)$.

The above discussion is made rigorous by the results of [12]. For $x \in X_2 \cap B$, as in definitions (3.4)–(3.6) below, the call acceptance probability $P_i(x)$ appearing in (3.2) is 1, $\sum_{j \geq 0} \pi_x^2(j)$ or 0, according as $i \in H$, M or L. In the limit $N \to \infty$ there is a separation of time scales: on the time-scale of the limit process $x(\cdot)$, the free capacity process moves infinitely fast.

We now define two further stationary distributions which play important roles in determining the acceptance probabilities of high and low priority calls on other parts of the boundary B. Let

$$X_1 = \left\{ x \in X\colon \sum_{i \in I} A_i \mu_i x_i < \sum_{i \in H} A_i \kappa_i \right\},$$

$$X_3 = \left\{ x \in X\colon \sum_{i \in H \cup M} A_i \kappa_i < \sum_{i \in I} A_i \mu_i x_i < \sum_{i \in I} A_i \kappa_i \right\}.$$

For $x \in X_1 \cap B$, define π_x^1 to be the stationary distribution of the Markov process on $\mathbb{Z}_+$ whose transition rates are given by

$$m_1 \to \begin{cases} m_1 - A_i, & \text{at rate } \kappa_i I\{m_1 \ge r_1\}, \text{ if } i \in H, \\ m_1 + A_i, & \text{at rate } \mu_i x_i, \text{ for each } i \in I. \end{cases}$$

For $x \in X_3 \cap B$, define π_x^3 to be the stationary distribution of the Markov process on $\mathbb{Z}$ whose transition rates are given by

$$m_3 \to \begin{cases} m_3 - A_i, & \text{at rate } \kappa_i, \text{ if } i \in H \cup M, \\ m_3 - A_i, & \text{at rate } \kappa_i I\{m_3 \ge 0\}, \text{ if } i \in L, \\ m_3 + A_i, & \text{at rate } \mu_i x_i, \text{ for each } i \in I. \end{cases}$$

For use below, also define the sets

$$X_2^- = \left\{ x \in X: \sum_{i \in I} A_i \mu_i x_i \le \sum_{i \in H} A_i \kappa_i \right\},$$

$$X_3^- = \left\{ x \in X: \sum_{i \in I} A_i \mu_i x_i \le \sum_{i \in H \cup M} A_i \kappa_i \right\}.$$

For each $x \in X$ and each $i \in I$, $P_i(x)$ denotes the acceptance probability of type i calls when $x(t) = x$, and is defined as follows:

$$(3.4) \qquad i \in H, \qquad P_i(x) = \begin{cases} \sum_{j \ge r_1} \pi_x^1(j), & \text{if } x \in X_1 \cap B, \\ 1, & \text{otherwise;} \end{cases}$$

$$(3.5) \qquad i \in M, \qquad P_i(x) = \begin{cases} 0, & \text{if } x \in X_2^- \cap B, \\ \sum_{j \ge 0} \pi_x^2(j), & \text{if } x \in X_2 \cap B, \\ 1, & \text{otherwise;} \end{cases}$$

$$(3.6) \qquad i \in L, \qquad P_i(x) = \begin{cases} 0, & \text{if } x \in X_3^- \cap B, \\ \sum_{j \ge 0} \pi_x^3(j), & \text{if } x \in X_3 \cap B, \\ 1, & \text{otherwise.} \end{cases}$$

For $x \in X_2^- \cap B$ the Markov process on $\mathbb{Z}$ making transitions according to (3.3) has nonpositive drift in all states: no stationary distribution exists and $P_i(x) = 0$ for $i \in M$. [For such situations, compactification of the state space of $m_2(\cdot)$ to $\mathbb{Z} \cup \{-\infty\} \cup \{+\infty\}$ is important; see [12] for details.] Similarly, for $x \in X_3^- \cap B$ we have $P_i(x) = 0$ for $i \in L$.

In the nontrivial cases we can, in fact, calculate the probabilities $P_i(x)$ explicitly, essentially by considering drifts. For example, for $x \in X_2 \cap B$, $m_2(\cdot)$ has constant negative drift on $[0, \infty)$ and constant positive drift on $(-\infty, 0)$: we require the equilibrium probability that $m_2(\cdot) \ge 0$, which is determined by these two drifts. Determining probabilities $P_i(x)$ for the case of a network of resources is considerably more difficult (see [11], [12] and comments in Section 5.3).

THEOREM 3.1. *For $i \in H$ and for each $x \in X_1 \cap B$,*

$$P_i(x) = \frac{\sum_{i \in I} A_i \mu_i x_i}{\sum_{i \in H} A_i \kappa_i}.$$

For $i \in M$ and for each $x \in X_2 \cap B$,

$$P_i(x) = \frac{\sum_{i \in I} A_i \mu_i x_i - \sum_{i \in H} A_i \kappa_i}{\sum_{i \in M} A_i \kappa_i}.$$

For $i \in L$ and for each $x \in X_3 \cap B$,

$$P_i(x) = \frac{\sum_{i \in I} A_i \mu_i x_i - \sum_{i \in H \cup M} A_i \kappa_i}{\sum_{i \in L} A_i \kappa_i}.$$

PROOF. We obtain $P_i(x)$ for $i \in M$ and $x \in X_2 \cap B$. The other cases follow similarly.

The conditions of Theorem 7.1 of [22] (extended, in the obvious manner, to a state space of $\mathbb{Z}$) are satisfied by the test function $V(j) = z^{|j|}$ for $j \in \mathbb{Z}$, with $z > 1$ and sufficiently close to 1. So, applying that theorem, $m_2(\cdot)$ is exponentially ergodic in the terminology of [22]. In particular, π_x^2 exists and $|\mathbf{E}[m_2(t)]| \to |\sum_{j \in \mathbb{Z}} j\pi_x^2(j)| < \infty$ as $t \to \infty$.

Now, by Dynkin's formula (e.g., [29], page 254), we have

$$\begin{aligned} \mathbf{E}[m_2(t)] = m_2(0) + \mathbf{E}\bigg[\int_0^t \bigg\{ \sum_{i \in I} A_i \mu_i \kappa_i - \sum_{i \in H} A_i \kappa_i \\ - I\{m_2(u) \geq 0\} \sum_{i \in M} A_i \kappa_i \bigg\} \, du \bigg]. \end{aligned} \tag{3.7}$$

Now divide by t and take limits as $t \to \infty$: the left hand side of (3.7) has limit zero since, from above, $\mathbf{E}[m_2(t)]$ has a finite limit. Hence we have

$$0 = \sum_{i \in I} A_i \mu_i \kappa_i - \sum_{i \in H} A_i \kappa_i - \bigg(\sum_{i \in M} A_i \kappa_i \bigg) \lim_{t \to \infty} \frac{1}{t} \mathbf{E}\bigg[\int_0^t I\{m_2(u) \geq 0\} \, du \bigg]. \tag{3.8}$$

The limit on the right-hand side of (3.8), the limiting expected proportion of time that $m_2(\cdot) \geq 0$, is simply $\sum_{j \geq 0} \pi_x^2(j)$ (e.g., [30], page 98). So for $i \in M$, rearranging (3.8) and using (3.5), we have

$$P_i(x) = \frac{\sum_{i \in I} A_i \mu_i x_i - \sum_{i \in H} A_i \kappa_i}{\sum_{i \in M} A_i \kappa_i},$$

as required. □

4. Asymptotic optimality. In this section we prove that our trunk reservation strategy is asymptotically optimal. First, we consider the behavior of the limit process $x(\cdot)$ as $t \to \infty$, for all possible starting points $x(0) \in X$.

If $f(\cdot)$ and $g(\cdot)$ are defined on $[0, \infty)$ (or if g is a constant), write $f \succcurlyeq g$ if $\liminf_{t \to \infty} f(t) \geq \liminf_{t \to \infty} g(t)$ [or if $\liminf_{t \to \infty} f(t) \geq g$]. Also write $a \wedge b = \min(a, b)$. Let $\mu_{\min} = \min_{i \in I} \mu_i$.

LEMMA 4.1. *All trajectories of the limit process $x(\cdot)$ are such that $x_i \succcurlyeq \nu_i$ for all $i \in H$.*

PROOF. *Step* 1. We suppose that $x_i \succcurlyeq a\nu_i$ for all $i \in H$ and some $a \in [0,1]$, and deduce that $x_i \succcurlyeq f(a)\nu_i$ for all $i \in H$, where

$$f(a) = \left(a + \left(C - a\sum_{i\in H} A_i\nu_i\right)\frac{\mu_{\min}}{\sum_{i\in H} A_i\kappa_i}\right) \wedge 1.$$

We have

$$\begin{aligned}\sum_{i\in I} A_i\mu_i x_i &= a\sum_{i\in H} A_i\kappa_i + \sum_{i\in H} A_i\mu_i(x_i - a\nu_i) + \sum_{i\in M\cup L} A_i\mu_i x_i\\ &\succcurlyeq a\sum_{i\in H} A_i\kappa_i + \mu_{\min}\left(\sum_{i\in I} A_i x_i - a\sum_{i\in H} A_i\nu_i\right),\end{aligned}$$

using the assumption that $x_i \succcurlyeq a\nu_i$ for $i \in H$, and so

$$\frac{\sum_{i\in I} A_i\mu_i x_i}{\sum_{i\in H} A_i\kappa_i} \succcurlyeq a + \left(\sum_{i\in I} A_i x_i - a\sum_{i\in H} A_i\nu_i\right)\frac{\mu_{\min}}{\sum_{i\in H} A_i\kappa_i}. \tag{4.1}$$

Now combining (4.1) with the expressions for $P_i(x)$, $i \in H$, given by (3.4) and Theorem 3.1, and using the fact that $\sum_{i\in I} A_i x_i = C$ when $x \in X_1 \cap B$, we have $P_i(x) \succcurlyeq f(a)$ for $i \in H$, where $f(a)$ is as defined above. Hence given $\varepsilon > 0$ there exists T such that, for all $i \in H$, $x_i(t) \ge y_i(t)$ for all $t \ge T$, where $y_i(T) = x_i(T)$ and, for $t \ge T$,

$$\frac{dy_i}{dt} = \kappa_i(f(a) - \varepsilon) - \mu_i y_i.$$

Solving this equation, $y_i(t) \to (f(a) - \varepsilon)\kappa_i/\mu_i$ as $t \to \infty$, and since ε is arbitrary we have $x_i \succcurlyeq f(a)\kappa_i/\mu_i = f(a)\nu_i$, as required for Step 1.

Step 2. Now fix a value of $i \in H$. Trivially $x_i \succcurlyeq 0$ and hence, repeatedly applying Step 1, $x_i \succcurlyeq a_n\nu_i$ for all n, where $a_0 = 0$ and $a_n = f(a_{n-1})$ for $n \ge 1$. For our purposes, the three important properties of f are as follows:

(i) $f(0) > 0$;
(ii) $f(1) = 1$ since, from the definition of H in terms of a solution of problem MF, $C - \sum_{i\in H} A_i\nu_i \ge 0$;
(iii) $f(a)$ is linear in a.

So (i)–(iii) imply that the values a_n form a nondecreasing sequence and that a_n converges to the unique solution of $f(a) = a$ as $n \to \infty$. From (ii) this solution is $a = 1$, and hence $x_i \succcurlyeq (\lim_{n\to\infty} a_n)\nu_i = \nu_i$, as required. □

The following result is closely related to Lemma 4.1, but refers to medium (rather than high) priority call types. Write $a^+ = \max(a, 0)$.

LEMMA 4.2. *All trajectories of the limit process $x(\cdot)$ are such that $x_i \succcurlyeq b^*\nu_i$ for all $i \in M$, where $b^* = (C - \sum_{i\in H} A_i\nu_i)/\sum_{i\in M} A_i\nu_i$.*

PROOF. Suppose that $M \neq \emptyset$, for otherwise there is nothing to prove.

Step 1. We suppose that $x_i \succcurlyeq b\nu_i$ for all $i \in M$ and some $b \in [0,1)$, and deduce that $x_i \succcurlyeq g(b)\nu_i$ for all $i \in M$, where

$$g(b) = \left(b + \left(C - \sum_{i \in H} A_i \nu_i - b \sum_{i \in M} A_i \nu_i\right) \frac{\mu_{\min}}{\sum_{i \in M} A_i \kappa_i}\right)^+ \wedge 1.$$

We have

$$\begin{aligned}\sum_{i \in I} A_i \mu_i x_i = {} & \sum_{i \in H} A_i \kappa_i + b \sum_{i \in M} A_i \kappa_i + \sum_{i \in H} A_i \mu_i (x_i - \nu_i) \\ & + \sum_{i \in M} A_i \mu_i (x_i - b\nu_i) + \sum_{i \in L} A_i \mu_i x_i \\ \succcurlyeq {} & \sum_{i \in H} A_i \kappa_i + b \sum_{i \in M} A_i \kappa_i + \mu_{\min}\left(\sum_{i \in I} A_i x_i - \sum_{i \in H} A_i \nu_i - b \sum_{i \in M} A_i \nu_i\right),\end{aligned}$$

using Lemma 4.1 and the assumption that $x_i \succcurlyeq b\nu_i$ for $i \in M$, and so

$$\begin{aligned}&\frac{\sum_{i \in I} A_i \mu_i x_i - \sum_{i \in H} A_i \kappa_i}{\sum_{i \in M} A_i \kappa_i} \\ &\qquad \succcurlyeq b + \left(\sum_{i \in I} A_i x_i - \sum_{i \in H} A_i \nu_i - b \sum_{i \in M} A_i \nu_i\right) \frac{\mu_{\min}}{\sum_{i \in M} A_i \kappa_i}.\end{aligned} \tag{4.2}$$

Now combining (4.2) with the expressions for $P_i(x)$, $i \in M$, given by (3.5) and Theorem 3.1, and using the fact that $\sum_{i \in I} A_i x_i = C$ when $x \in X_2 \cap B$, we have $P_i(x) \succcurlyeq g(b)$ for $i \in M$, where $g(b)$ is as defined above. Now arguing exactly as in the second part of Step 1 in the proof of Lemma 4.1, we obtain $x_i \succcurlyeq g(b)\nu_i$, as required.

Step 2. Now fix a value of $i \in M$. Trivially $x_i \succcurlyeq 0$ and hence, repeatedly applying Step 1, $x_i \succcurlyeq b_n \nu_i$ for all n, where $b_0 = 0$ and $b_n = g(b_{n-1})$ for $n \geq 1$. The three important properties of g are:

(i) $g(0) > 0$ since, from the definitions of H and M in terms of a solution of problem MF, $C - \sum_{i \in H} A_i \nu_i > 0$ as $M \neq \emptyset$;
(ii) $g(1) < 1$ since, from problem MF, $C - \sum_{i \in H \cup M} A_i \nu_i < 0$ as $M \neq \emptyset$;
(iii) $g(b)$ is linear in b.

So (i)–(iii) imply that the b_n form a nondecreasing sequence and that b_n converges to the unique solution b^* of $g(b) = b$ as $n \to \infty$ [and b^* is in $(0,1)$]. Solving $g(b) = b$, we find that b^* is as given above. Hence $x_i \succcurlyeq (\lim_{n \to \infty} b_n)\nu_i = b^* \nu_i$, as required. □

In order to prove asymptotic optimality, the following theorem gives the important result needed about the behavior of our trunk reservation strategy as $t \to \infty$.

THEOREM 4.3. *All trajectories of the limit process* $x(\cdot)$ *are such that the limit* $x_i^* = \lim_{t\to\infty} x_i(t)$ *exists, for all* $i \in I$, *and is given by*

$$x_i^* = \begin{cases} \nu_i, & \text{if } i \in H,\\ b^*\nu_i, & \text{if } i \in M,\\ 0, & \text{if } i \in L.\end{cases}$$

In particular, x^* *is the unique fixed point of equations* (3.1) *when the call acceptance probabilities* $P_i(x)$ *are given by* (3.4)–(3.6) *and Theorem* 3.1.

PROOF. *Step* 1. The limit process $x(\cdot)$ satisfies (3.1) where $v_i(x)$ is given in (3.2). Now, since $P_i(x) \le 1$ for all x, $v_i(x) \le \kappa_i - \mu_i x_i < 0$ when $x_i > \nu_i$. So it is immediate that $\limsup_{t\to\infty} x_i(t) \le \nu_i$. Combining this fact with Lemma 4.1, $x_i^* = \lim_{t\to\infty} x_i(t)$ exists for $i \in H$ and $x_i^* = \nu_i$ for such i.

Step 2(i). *The case* $M = \emptyset$. In this case, considering the definitions of sets H, M and L via problem MF, we find that either $L = \emptyset$ or $\sum_{i\in H} A_i\nu_i = C$ (or both). If $L = \emptyset$, there is nothing more to prove. If $\sum_{i\in H} A_i\nu_i = C$, then, using Step 1, $\lim_{t\to\infty} \sum_{i\in H} A_i x_i(t) = C$. Hence $\lim_{t\to\infty} \sum_{i\in L} A_i x_i(t)$ exists and equals 0, otherwise the capacity constraint $\sum_{i\in I} A_i x_i(t) \le C$ would be violated for some t. Then, since $x_i(t) \ge 0$, $\lim_{t\to\infty} x_i(t)$ exists and equals 0 for all $i \in L$.

Step 2(ii). *The case* $M \ne \emptyset$. From Lemmas 4.1 and 4.2 we have $x_i \succcurlyeq \nu_i$ for $i \in H$, and $x_i \succcurlyeq b^*\nu_i$ for $i \in M$. Multiplying these relations by A_i and summing over i gives

$$(4.3)\qquad \sum_{i\in H\cup M} A_i x_i \succcurlyeq \sum_{i\in H} A_i\nu_i + b^* \sum_{i\in M} A_i\nu_i = C,$$

where the equality follows by using the expression for b^* given in Lemma 4.2. Combining (4.3), Step 1 and Lemma 4.2, $\lim_{t\to\infty} x_i(t)$ exists and equals $b^*\nu_i$ for $i \in M$, otherwise the capacity constraint $\sum_{i\in I} A_i x_i(t) \le C$ would be violated for some t. Then considering the capacity constraint once more, $\lim_{t\to\infty} x_i(t)$ exists and equals 0 for $i \in L$. □

REMARK. This result is closely related to Theorem 3.4 of Bean, Gibbens and Zachary [2]. When $A_i = 1$ for all $i \in I$, they show existence of a unique fixed point of their limit process, and convergence of all trajectories to that point under the additional condition $|I| = 2$. Their limit process is slightly different from ours, in general: all trunk reservation parameters in [2] are fixed independent of N (corresponding to $M = L = \emptyset$) but otherwise arbitrary, whereas our fixed trunk reservation parameters (those for types $i \in H$) are equal. These equal trunk reservation parameters allow us to calculate explicitly the call acceptance probabilities (Theorem 3.1), and the form of these probabilities plays an important role in the proof of Theorem 4.3 (via Lemmas 4.1 and 4.2). Although the sets H, M and L here are defined in terms of a solution of problem MF, Theorem 5.1 shows convergence of all trajectories of the limit process to the unique fixed point for any number of priority levels, chosen arbitrarily, when trunk reservation parameters vary suitably with N.

An immediate corollary of Theorem 4.3 is as follows.

COROLLARY 4.4. *The unique invariant distribution of the limit process* $x(\cdot)$ *is the distribution concentrated on* x^*.

Although sets H, M and L were chosen according to solution $\bar{x}$ of problem MF, the limit vector $x^* = (x_i^*,\ i \in I)$ is not necessarily $\bar{x}$. If $|M| \le 1$, then $x^* = \bar{x}$. If $|M| \ge 2$, then x_i^* and $\bar{x}_i$ are different in general, for $i \in M$, owing to the fact that solutions of MF are not unique. However, the following result shows that x^* also solves MF.

LEMMA 4.5. *The vector* x^* *is an optimal solution of problem MF.*

PROOF. The dual D of MF is the problem

$$\begin{aligned} \text{(D)} \quad & \text{minimize} \quad Cy_0 + \sum_i \nu_i y_i \\ & \text{subject to} \quad A_i y_0 + y_i \ge w_i, \qquad i \in I, \\ & \qquad\qquad\quad y_0 \ge 0, \qquad y_i \ge 0, \qquad i \in I. \end{aligned}$$

The corresponding complementary slackness conditions are

$$(A_i y_0 + y_i - w_i) x_i = 0, \qquad i \in I, \tag{4.4}$$

$$y_0\Big(C - \sum_i A_i x_i\Big) = 0, \tag{4.5}$$

$$y_i(\nu_i - x_i) = 0, \qquad i \in I. \tag{4.6}$$

Let $\bar{y} = (\bar{y}_0; \bar{y}_i,\ i \in I)$ be an optimal solution of D. Then x^* is feasible for MF, $\bar{y}$ is feasible for D and it is straightforward to check that $(x^*, \bar{y})$ satisfy the complementary slackness conditions (4.4)–(4.6). Hence x^* is an optimal solution of MF. □

We can now complete the proof that bound (2.4) is attained by our trunk reservation strategy.

THEOREM 4.6. *The three-level trunk reservation strategy (policy* π_1, *say) is asymptotically optimal; that is,*

$$\lim_{N \to \infty} \frac{R^{\pi_1}(N)}{N} = \sum_i w_i \bar{x}_i.$$

PROOF. Consider the Nth system operating under policy π_1 and, as in Section 2, let the random variable $n_i^{\pi_1}(N)$ denote the number of type i calls in progress in equilibrium when the Nth system is operated under policy π_1.

Now, for all N,

$$0 \le n_i^{\pi_1}(N)/N \le C, \qquad i \in I, \tag{4.7}$$

and hence the sequence of random vectors $(n_i^{\pi_1}(N)/N,\ i \in I)$ is tight (e.g., [3], page 392). So there is a subsequence (N_k), say, along which, for all $i \in I$, $n_i^{\pi_1}(N_k)/N_k$ converges weakly as $k \to \infty$. Since the N_kth system is stationary, the limit process is also, and hence $n_i^{\pi_1}(N_k)/N_k \Rightarrow x_i^*$ as $k \to \infty$, by Corollary 4.4. As these limits (of x_i^*) are independent of the subsequence chosen, the whole sequence converges; that is,

$$n_i^{\pi_1}(N)/N \Rightarrow x_i^* \quad \text{as } N \to \infty,\ i \in I. \tag{4.8}$$

Hence

$$\begin{aligned}
\lim_{N\to\infty} \frac{R^{\pi_1}(N)}{N} &= \sum_i w_i \lim_{N\to\infty} \mathbf{E}\left[\frac{n_i^{\pi_1}(N)}{N}\right] \\
&= \sum_i w_i x_i^* \\
&= \sum_i w_i \bar{x}_i,
\end{aligned}$$

where the second equality follows from (4.7) and (4.8), and the third from Lemma 4.5. □

REMARKS. (i) Asymptotically and in equilibrium, the proportion of type i calls accepted is 1, b^* or 0 according as i is in H, M or L. In a finite-sized system, the effects of trunk reservation are less pronounced (cf. numerical results in [2]). For example, in a system corresponding to a finite value of N, the proportion of high priority calls accepted would be large, but not equal to 1; and the proportion of low priority calls accepted would be smaller, but not 0.

(ii) Some evidence of the robustness of trunk reservation can be seen here. Our trunk reservation strategy is defined in terms of sets H, M and L, which, in turn, are determined from a solution $\bar{x}$ of problem MF. Whenever the solution of MF is nonunique (the general case), the particular optimal solution $\bar{x}$ used to define H, M and L is unimportant: we will always have an asymptotically optimal strategy [although the proportions of accepted calls will differ if H, M and L differ; see remark (i) above]. In particular, the exact values of arrival rates κ_i are not required in order to specify an asymptotically optimal policy: rather, only sufficient information as required to determine suitable sets H, M and L is needed. We see further and stronger evidence of the robustness of trunk reservation in the following section.

Theorem 4.6 shows that a policy using three trunk reservation parameters is asymptotically optimal. In Section 5, asymptotically optimal policies which use more than three trunk reservation parameters are obtained. Are there trunk reservation policies that are asymptotically optimal and that use fewer than three trunk reservation parameters? The answer is no for just one parameter. For then there is no way to give three different levels of priority according as call types are in H, M or L: one parameter allows only "higher" priority calls (accepted whenever possible, so subject only to the total capacity constraint of the resource) and "lower" priority calls (accepted subject to a trunk reservation criterion being satisfied). However, this suggests that two parameters may be sufficient: use the two trunk reservation parameters against medium and low priority calls, and accept high priority calls whenever possible. Indeed, if $A_i = A$ for all $i \in H$ (and some A), then policy π_1 above accepts high priority calls whenever possible; so in this case π_1 is really an asymptotically optimal policy which uses two trunk reservation parameters, against medium and low priority calls. Our reason for taking $r_1 = \max_{i \in H} A_i$ was to ensure that, for $i \in H$, call acceptance probabilities $P_i(x)$ were equal and could be calculated explicitly. Suppose values of A_i, for $i \in H$, differ and that we alter π_1 by accepting high priority calls whenever possible. This new strategy, with two trunk reservation parameters, may be asymptotically optimal but it is not clear whether our methods of proof could be adapted, as they rely on the calculated form of $P_i(x)$, for $i \in H$.

5. Generalizations. We now examine the behavior of variants of the trunk reservation strategy considered in the previous two sections. We obtain further results on equilibrium behavior under trunk reservation, and asymptotically optimal policies which require no knowledge of arrival rates κ_i.

5.1. *Multiple priority levels.* We now allow any number of priority levels, where the priority level assigned to a call type is chosen arbitrarily. Suppose the nonempty sets $I_1, I_2, \dots, I_K$ partition the set of call types I, and suppose that the Nth system operates as follows. For $i \in I_1$, accept calls of type i subject to a trunk reservation parameter of $r_1 = \max_{i \in I_1} A_i$; for $i \in I_k$, where $2 \le k \le K$, accept calls of type i subject to a trunk reservation parameter of $r_k(N) = (k-1)\log N$. This generalizes the previous strategy: the number of priority levels K is arbitrary, and priority levels need not be chosen according to a solution of MF.

REMARK. As before, there is some flexibility in our choice of trunk reservation parameters: the results below hold if, for $2 \le k \le K$, the parameters $r_k(N)$ are such that $r_k(N)$ and $r_k(N) - r_{k-1}(N)$ are $o(N)$ and increase without bound as $N \to \infty$. Further, they hold with r_1 any fixed value of at least $\max_{i \in I_1} A_i$, or with $r_1 = r_1(N)$ dependent on N and such that $r_1(N)$ and $r_2(N) - r_1(N)$ are $o(N)$ and increase without bound as $N \to \infty$.

Let $x^N(t) = (x_i^N(t),\ i \in I)$ be the Markov process, with state space X, defined by $x_i^N(t) = n_i^N(t)/N$, where $n_i^N(t)$ is the number of type i calls in progress at time t in the Nth system under the above multilevel trunk reservation strategy. Then, as in Section 3, assuming $x^N(0) \Rightarrow x(0)$ as $N \to \infty$, the sequence of processes $\{x^N(\cdot)\}$ is relatively compact and the limit $x(\cdot)$ of any convergent subsequence satisfies (3.1) and (3.2). The only way that the description of the limit process $x(\cdot)$ differs from that in Section 3 is through the call acceptance probabilities $P_i(x)$ on certain parts of the boundary B of X. As in Section 3, these probabilities are determined by stationary distributions of certain Markov processes and have similar interpretations in terms of call acceptances in the Nth system. Since the definitions below are closely related to those in Section 3, and reduce to them in the case $K = 3$, we simply give the values of $P_i(x)$ and omit further details.

Let

$$X_k = \left\{ x \in X \colon \sum_{j=1}^{k-1} \sum_{i \in I_j} A_i \kappa_i < \sum_{i \in I} A_i \mu_i x_i < \sum_{j=1}^{k} \sum_{i \in I_j} A_i \kappa_i \right\}, \qquad 1 \le k \le K,$$

$$X_k^- = \left\{ x \in X \colon \sum_{i \in I} A_i \mu_i x_i \le \sum_{j=1}^{k-1} \sum_{i \in I_j} A_i \kappa_i \right\}, \qquad 2 \le k \le K,$$

where, here and throughout, an empty sum is taken to be 0. The call acceptance probabilities $P_i(x)$ are defined as follows:

$$i \in I_1, \qquad P_i(x) = \begin{cases} p_1(x), & \text{if } x \in X_1 \cap B, \\ 1, & \text{otherwise,} \end{cases}$$

and, for $k = 2, 3, \dots, K$,

$$i \in I_k, \qquad P_i(x) = \begin{cases} 0, & \text{if } x \in X_k^- \cap B, \\ p_k(x), & \text{if } x \in X_k \cap B, \\ 1, & \text{otherwise,} \end{cases}$$

where

$$p_k(x) = \frac{\sum_{i \in I} A_i \mu_i x_i - \sum_{j=1}^{k-1} \sum_{i \in I_j} A_i \kappa_i}{\sum_{i \in I_k} A_i \kappa_i}, \qquad 1 \le k \le K.$$

Define $k^* = K + 1$ if $\sum_{i \in I} A_i \nu_i \le C$ and otherwise define k^* to be the unique $k \in \{1, 2, \dots, K\}$ satisfying

$$\sum_{j=1}^{k-1} \sum_{i \in I_j} A_i \nu_i \le C < \sum_{j=1}^{k} \sum_{i \in I_j} A_i \nu_i.$$

The following result (a generalization of Theorem 4.3) is the crucial step in enabling us to describe, asymptotically, the equilibrium behavior of the above trunk reservation strategy.

THEOREM 5.1. *Under the multilevel trunk reservation strategy, the limit* $x_i^* = \lim_{t \to \infty} x_i(t)$ *exists for all* $i \in I$ *and is given by*

$$x_i^* = \begin{cases} \nu_i, & \text{if } i \in I_k,\ k < k^*, \\ p^* \nu_i, & \text{if } i \in I_{k^*}, \\ 0, & \text{if } i \in I_k,\ k > k^*, \end{cases}$$

where $p^* \in [0,1)$ *is defined by*

$$p^* = \frac{C - \sum_{k<k^*}\sum_{i\in I_k} A_i\nu_i}{\sum_{i\in I_{k^*}} A_i\nu_i}.$$

PROOF. The proof is in three steps, which closely follow proofs in Section 4.

Step 1. We first show that $x_i \succcurlyeq \nu_i$ for all $i \in I_k$, $k < k^*$, by induction on k.

The result holds for $k = 1$ by applying Lemma 4.1 with $H = I_1$. So fix $k \le k^* - 1$, and suppose that $x_i \succcurlyeq \nu_i$ for all $i \in I_j$ and all $j < k$. Suppose also that $x_i \succcurlyeq a\nu_i$ for all $i \in I_k$ and some $a \in [0,1]$. Then, following the proof of Lemma 4.2 with $H = \bigcup_{j<k} I_j$ and $M = I_k$, we obtain $P_i(x) \succcurlyeq f(a)$ and then $x_i \succcurlyeq f(a)\nu_i$, where

$$f(a) = \left(a + \left(C - \sum_{j<k}\sum_{i\in I_j} A_i\kappa_i - a\sum_{i\in I_k} A_i\kappa_i\right)\frac{\mu_{\min}}{\sum_{i\in I_k} A_i\kappa_i}\right)^+ \wedge 1.$$

Hence, as before, $x_i \succcurlyeq a_n\nu_i$ for all $i \in I_k$ and all n, where $a_0 = 0$ and $a_n = f(a_{n-1})$. Now f has properties (i)–(iii) in the proof of Lemma 4.1 [where (ii) follows since $k < k^*$]. So $a_n \to 1$ as $n \to \infty$ as in the proof of Lemma 4.1, hence $x_i \succcurlyeq \nu_i$ for $i \in I_k$, and the induction is complete.

Step 2. Next, when $k^* \le K$, we show $x_i \succcurlyeq p^*\nu_i$ for $i \in I_{k^*}$.

When $p^* = 0$ there is nothing to prove. If $p^* > 0$, use the conclusion of Step 1 and follow exactly the proof of Lemma 4.2, but with $H = \bigcup_{k<k^*} I_k$ and $M = I_{k^*}$. Then the definition of k^* and the fact that $p^* > 0$ give the equivalent of properties (i) and (ii) in the proof of Lemma 4.2, and hence the result follows.

Step 3(i). *The case* $k^* = K + 1$. Exactly as in the proof of Step 1 of Theorem 4.3, $\limsup_{t\to\infty} x_i(t) \le \nu_i$. Combining this with the conclusion of Step 1, $x_i^* = \lim_{t\to\infty} x_i(t)$ exists and $x_i^* = \nu_i$ for all $i \in I$. Hence the theorem holds in this case.

Step 3(ii). *The case* $k^* \le K$. Using the results of Steps 1 and 2,

$$\sum_{k=1}^{k^*}\sum_{i\in I_k} A_i x_i \succcurlyeq \sum_{k<k^*}\sum_{i\in I_k} A_i\nu_i + p^*\sum_{i\in I_{k^*}} A_i\nu_i = C,$$

where the equality follows from the definition of p^*. Hence, as in the proof of Theorem 4.3, the conclusion of this theorem holds for $i \in I_k$ and $k \le k^*$, and then for all i and k. □

Let π_2 denote the multilevel trunk reservation strategy and $n_i^{\pi_2}(N)$ the number of type i calls in progress in equilibrium in the Nth system under policy π_2. Then, as in the proof of Theorem 4.6, $\mathbf{E}[n_i^{\pi_2}(N)/N] \to x_i^*$ as $N \to \infty$ and the limiting normalized reward rate under π_2 is given by $\lim_{N\to\infty} R^{\pi_2}(N)/N = \sum_i w_i x_i^*$, where $R^{\pi_2}(N) = \mathbf{E}[w_i n_i^{\pi_2}(N)]$. So, asymptotically and in equilibrium, the proportion of calls accepted at priority level k is 1, p^* or 0 according as $k < k^*$, $k = k^*$ or $k > k^*$. As before, the effect of trunk reservation is most pronounced in the limit $N \to \infty$, with proportions 0 and 1 of calls of certain types being accepted; in any finite system, the effects of trunk reservation are less extreme.

5.2. *Asymptotic optimality.* We now show, by linking problem MF with the results just obtained, that there is an easily defined trunk reservation strategy which is asymptotically optimal and which does not depend on arrival rates κ_i.

Consider problem MF under the substitution $z_i = A_i x_i$: this yields the problem

$$\text{(5.1a)} \qquad \text{maximize} \quad \sum_i \frac{w_i}{A_i} z_i,$$

$$\text{(5.1b)} \qquad \text{subject to} \quad \sum_i z_i \le C,$$

$$\text{(5.1c)} \qquad 0 \le z_i \le A_i \nu_i, \qquad i \in I.$$

Suppose that $I = \{1, 2, \dots, I\}$ and assume

$$\text{(5.2)} \qquad \frac{w_1}{A_1} > \frac{w_2}{A_2} > \cdots > \frac{w_I}{A_I}.$$

From (5.1b), each z_i contributes equally to the new capacity constraint and so w_i/A_i can be regarded as a priority indicator for calls of type i, with large values of w_i/A_i indicating high priority. Thus the optimal solution of problem (5.1) can be constructed as follows. First, take z_1 as large as possible; that is, $z_1 = C \wedge A_1 \nu_1$. Then take z_2 as large as possible subject to (5.1b), (5.1c), and the above choice of z_1; that is, $z_2 = (C - z_1) \wedge A_2 \nu_2$; and so on for z_i, $i \ge 3$. Hence the optimal solution of problem MF [which is unique under condition (5.2)] is as follows.

Define $i^* = I + 1$ if $\sum_i A_i \nu_i \le C$ and otherwise define i^* to be the unique $i \in \{1, 2, \dots, I\}$ for which

$$\sum_{j=1}^{i-1} A_j \nu_j \le C < \sum_{j=1}^{i} A_j \nu_j.$$

Then the optimal solution of MF is $\bar{x} = (\bar{x}_i,\ i \in I)$, where

$$\text{(5.3)} \qquad \bar{x}_i = \begin{cases} \nu_i, & \text{if } i < i^*, \\ \bar{p}\nu_i, & \text{if } i = i^*, \\ 0, & \text{if } i > i^*, \end{cases}$$

where $\bar{p} \in [0, 1)$ is defined by

$$\text{(5.4)} \qquad \bar{p} = \frac{C - \sum_{i < i^*} A_i \nu_i}{A_{i^*} \nu_{i^*}}.$$

Consider the trunk reservation strategy of the previous section with $I_k = \{k\}$ for $1 \le k \le I$. From Theorem 5.1 and the remarks following it, the limiting normalized reward rate of this I-level trunk reservation strategy is

$\sum_i w_i \bar{x}_i$. Hence this strategy attains equality in bound (2.4) and therefore is asymptotically optimal.

REMARK. Assumption (5.2) is not without loss of generality as $\max_{i \in I} w_i/A_i$ may be achieved by several values of i and the maximizing values of i may have different values of A_i.

For the remainder of the section we do not assume that (5.2) holds. Consider the following multilevel trunk reservation strategy: accept calls of type i subject to a trunk reservation parameter of $r_i(N) = (A_i/w_i)\log N$. This strategy corresponds to that considered in Section 5.1, with $I_1 = \varnothing$ and with I_k defined inductively in terms of $I_1, I_2, \dots, I_{k-1}$ by

$$I_k = \left\{ i \in I : \frac{w_i}{A_i} = \max_{i' \notin U_{k-1}} \frac{w_{i'}}{A_{i'}} \right\}, \tag{5.5}$$

where $U_{k-1} = \bigcup_{j=1}^{k-1} I_j$. Observe that the trunk reservation parameter used against type i calls is inversely related to the priority indicator w_i/A_i: this is simply because calls have a high priority if they are subject to a small trunk reservation parameter, and vice versa.

REMARKS. (i) With priority levels given by (5.5), even the highest priority calls (those of types $i \in I_2$) are subject to a trunk reservation parameter of $O(\log N)$. However, if we alter $r_i(N)$ so that $r_i(N) = r$ for $i \in I_2$ and all N, where $r \geq \max_{i \in I_2} A_i$, then the asymptotic behavior of the policy is unaltered, by the results in Section 5.1. In particular, the policy remains asymptotically optimal (see Theorem 5.2).

(ii) If condition (5.2) holds, the asymptotic behavior of this multilevel trunk reservation strategy is exactly the same as that of the I-level strategy above. The new policy allows for the general case where several call types have the same priority level.

Let π_3 denote this multilevel trunk reservation strategy. We can now apply the results obtained in Section 5.1. We have $\mathbf{E}[n_i^{\pi_3}(N)/N] \to x_i^*$ as $N \to \infty$, and the limiting normalized reward rate under π_3 is given by $\lim_{N \to \infty} R^{\pi_3}(N)/N = \sum_i w_i x_i^*$, where x^* is as in Theorem 5.1 with priority levels given by (5.5). When (5.2) holds, x^* reduces to the $\bar{x}$ defined in (5.3) above. In fact, x^* is a natural generalization of this $\bar{x}$ to the case where values of w_i/A_i may be equal. Indeed, it is straightforward to see that x^* is an optimal solution of problem MF. Thus bound (2.4) is attained by policy π_3, and hence π_3 is an asymptotically optimal policy which is independent of the arrival rates κ_i.

THEOREM 5.2. *The multilevel trunk reservation policy π_3, which uses a trunk reservation parameter of $(A_i/w_i)\log N$ against calls of type i, is asymptotically optimal* [*i.e., it attains bound* (2.4) *with* $\pi = \pi_3$].

5.3. *Networks.* Extending our results to networks of resources, where calls request use of capacity from several resources simultaneously, appears difficult. For the single resource case the acceptance probabilities $P_i(\cdot)$ in (3.2) were determined explicitly (Theorem 3.1). For the network case the limit process again satisfies equations of the form (3.1) (see [12]), although the corresponding $P_i(\cdot)$ are unknown in general. In fact, call acceptance probabilities at time t need not be determined by $x(t)$ under general control strategies: Hunt [11] gives examples and illustrates some of the subtle issues involved here, and Bean, Gibbens and Zachary [1] establish further results on identifying the dynamics of the limit process. However, suppose call acceptance probabilities are approximated rather than calculated exactly: Notebaert [27] uses the assumption that resources in a network are independently above/below the high/medium/low trunk reservation levels (cf. [14, 32]), and this allows some further progress. Although this approximation is known to be incorrect, asymptotically, when trunk reservation parameters are fixed [12], it would be interesting to know if it is valid when trunk reservation parameters vary with N, since this issue is closely related to the asymptotic optimality, or not, of trunk reservation in general loss networks.

Acknowledgment. We thank Bruce Hajek for several helpful conversations and especially for providing the method of proof of Lemmas 4.1 and 4.2.

REFERENCES

[1] BEAN, N. G., GIBBENS, R. J. and ZACHARY, S. (1994). Dynamic and equilibrium behaviour of controlled loss networks. Research Report 94-31, Statistical Laboratory, Univ. Cambridge.

[2] 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** 273–292.

[3] BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

[4] COHEN, J. E. and KELLY, F. P. (1990). A paradox of congestion in a queueing network. *J. Appl. Probab.* **27** 730–734.

[5] GIBBENS, R. J. and HUNT, P. J. (1991). Effective bandwidths for the multi-type UAS channel. *Queueing Systems Theory Appl.* **9** 17–28.

[6] GIBBENS, R. J., HUNT, P. J. and KELLY, F. P. (1990). Bistability in communication networks. In *Disorder in Physical Systems* (G. R. Grimmett and D. J. A. Welsh, eds.) 113–127. Oxford Univ. Press.

[7] GIBBENS, R. J. and KELLY, F. P. (1990). Dynamic routing in fully connected networks. *IMA J. Math. Control Inform.* **7** 77–111.

[8] GIBBENS, R. J., KELLY, F. P. and TURNER, S. R. E. (1993). Dynamic routing in multiparented networks. *IEEE/ACM Transactions on Networking* **1** 261–270.

[9] GIBBENS, R. J. and REICHL, P. C. (1995). Performance bounds applied to loss networks. In *Complex Stochastic Systems and Engineering* (D. M. Titterington, ed.) 267–279. Oxford Univ. Press.

[10] HUI, J. Y. (1988). Resource allocation for broadband networks. *IEEE J. Selected Areas in Communications* **6** 1598–1608.

[11] HUNT, P. J. (1995). Pathological behaviour in loss networks. *J. Appl. Prob.* **32** 519–533.

[12] HUNT, P. J. and KURTZ, T. G. (1994). Large loss networks. *Stochastic Process. Appl.* **53** 363–378.

[13] HUNT, P. J. and LAWS, C. N. (1993). Asymptotically optimal loss network control. *Math. Oper. Res.* **18** 880–900.

[14] KELLY, F. P. (1986). Blocking probabilities in large circuit-switched networks. *Adv. in Appl. Probab.* **18** 473–505.

[15] KELLY, F. P. (1991). Effective bandwidths at multi-class queues. *Queueing Systems Theory Appl.* **9** 5–16.

[16] KELLY, F. P. (1991). Loss networks. *Ann. Appl. Prob.* **1** 319–378.

[17] KELLY, F. P. (1994). Bounds on the performance of dynamic routing schemes for highly connected networks. *Math. Oper. Res.* **19** 1–20.

[18] KEY, P. B. (1990). Optimal control and trunk reservation in loss networks. *Probab. Engrg. Inform. Sci.* **4** 203–242.

[19] KEY, P. B. (1994). Some control issues in telecommunications networks. In *Probability, Statistics and Optimisation* (F. P. Kelly, ed.) 383–395. Wiley, New York.

[20] LAWS, C. N. (1995). On trunk reservation in loss networks. In *Stochastic Networks* (F. P. Kelly and R. J. Williams, eds.) 187–198. Springer, New York.

[21] LIPPMAN, S. A. (1975). Applying a new device in the optimization of exponential queueing systems. *Oper. Res.* **23** 686–710.

[22] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548.

[23] MILLER, B. L. (1969). A queueing reward system with several customer classes. *Management Science* **16** 234–245.

[24] MITRA, D. and GIBBENS, R. J. (1992). State-dependent routing on symmetric loss networks with trunk reservations, II: asymptotics, optimal design. *Ann. Oper. Res.* **35** 3–30.

[25] MITRA, D., GIBBENS, R. J. and HUANG, B. D. (1993). State-dependent routing on symmetric loss networks with trunk reservations, I. *IEEE Transactions on Communications* **41** 400–411.

[26] NGUYEN, V. (1991). On the optimality of trunk reservation in overflow processes. *Probab. Engrg. Inform. Sci.* **5** 369–390.

[27] NOTEBAERT, N. (1992). Asymptotical convergence of a trunk reservation policy. Technical report, Statistical Laboratory, Univ. Cambridge and Dépt. de Mathématiques Appliquées, Ecole Polytechnique.

[28] REIMAN, M. I. (1991). Optimal trunk reservation for a critically loaded link. In *Proceedings of the 13th International Teletraffic Congress* (A. Jensen and V. B. Iversen, eds.) 247–252. North-Holland, Amsterdam.

[29] ROGERS, L. C. G. and WILLIAMS, D. (1994). *Diffusions, Markov Processes, and Martingales*, 2nd ed. Wiley, New York.

[30] ROSS, S. M. (1970). *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco.

[31] ROSS, K. W. and TSANG, D. H. K. (1989). Optimal circuit access policies in an ISDN environment: a Markov decision approach. *IEEE Transactions on Communications* **37** 934–939.

[32] WHITT, W. (1985). Blocking when service is required from several facilities simultaneously. *AT & T Technical Journal* **64** 1807–1856.

[33] WHITTLE, P. (1983). *Optimization over Time* **2**. Wiley, New York.

ABN AMRO BANK
199 BISHOPSGATE
LONDON EC2M 3TY
UNITED KINGDOM

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD OX1 3TG
UNITED KINGDOM
EMAIL: laws@stats.ox.ac.uk