Branching Bandits and Klimov's Problem: Achievable Region and Side Constraints \(^1\)

Dimitris Bertsimas Ioannis Ch. Paschalidis
John N. Tsitsiklis

Laboratory for Information and Decision Systems
and
Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA 02139

April 1994

\(^1\)Research supported by a Presidential Young Investigator award DDM-9158118 with matching funds from Draper Laboratory, by the Leaders for Manufacturing Program at MIT, and by the ARO under grant DAAL-03-92-G-0115.
Abstract

We consider the average cost branching bandits problem and its special case known as Klimov’s problem. We consider the vector $n$ whose components are the mean number of bandits (or customers) of each type that are present. We characterize fully the achievable region, that is, the set of all possible vectors $n$ that can be obtained by considering all possible policies. While the original description of the achievable region involves exponentially many constraints, we also develop an alternative description that involves only $O(R^2)$ variables and constraints, where $R$ is the number of bandit types (or customer classes). We then consider the problem of minimizing a linear function of $n$ subject to $L$ additional linear constraints on $n$. We show that optimal policies can be obtained by randomizing between $L + 1$ strict priority policies that can be found efficiently (in polynomial time) using linear programming techniques.

Keywords: Stochastic Control, Queueing Systems, Discrete Event Systems.
1 Introduction

Consider a single-server multiclass M/GI/1 queue with Bernoulli feedback. In this context, one wishes to determine a policy that optimizes a linear combination of the mean number of customers of the different classes that are present in the system. This problem was posed and solved by Klimov [Kli74] who established the optimality of strict priority rules. In addition, he developed a fairly simple and efficient one-pass algorithm that determines an optimal priority ordering.

In the branching bandits problem, as defined by Weiss [Wei88], there is again a single server who serves several customer classes and a similar performance criterion. However, at each service completion, the served customer is replaced by a random number of customers of every other class. This model is more general than Klimov's in that the random numbers of new customers need not correspond to Poisson arrival processes.

Both problems can be extended by imposing some additional linear side constraints. For example, we might require that the mean queue length is the same for each customer class. Such side constraints are usually meant to represent fairness constraints.

Much of the work on the branching bandits and Klimov's problems views these problems as extensions of the classical multi-armed bandit problem [Git89, Wal88, Wei88]. In this paper, however, we take a philosophically very different approach. In particular, we consider the vector \( \mathbf{n} \) whose components are the mean number of customers of each type that are present and characterize fully the achievable region, that is, the set of all possible vectors \( \mathbf{n} \) that can be obtained by considering all possible policies. Our characterizations are polyhedral; that is, they are expressed in terms of linear equality and inequality constraints. We are thus able to convert a difficult stochastic control problem to one of optimizing a linear cost function over the achievable region, and this is a linear programming problem. There has already been a fair amount of research on such polyhedral characterizations, which we now discuss.

Gelenbe and Mitrani [GM80] used conservation laws to show that the performance region of a multiclass queue (without feedback) can be described as a polyhedron. Federgruen and Groenvelt [FG88] advanced the theory by observing that in certain special cases of multiclass queues the polyhedron has a special (polymatroidal) structure. Shantikumar and Yao [SY92] generalized the theory further by observing that if a system satisfies conservation laws, then the underlying performance space is necessarily a polymatroid polytope and the optimal policy is a strict priority rule. Closer to the subject of this paper, Tsoucas [Tso91] has derived a characterization of the achievable region for Klimov's problem, but without giving explicit formulae for some of the constants in his characterization. Bertsimas and Niño-Mora [BNM92] generalize the idea of conservation laws and show that for all systems that satisfy these generalized conservation laws (including multiclass queues, usual and branching bandits), the underlying performance space is a polyhedron with very strong
structural properties, called an extended polymatroid in [BGT92]. They also obtained an explicit characterization of the achievable region for Klimov's problem. Finally, the authors, in [BPT92a] and [BPT92b], have used quadratic potential functions to develop conservation laws for general controlled multiclass queueing networks with Poisson arrivals and exponential service times. In the network case, these conservation laws do not provide an exact characterization of the achievable region but lead to bounds for the achievable region which are often quite tight. For the special case of Klimov's problem in which service times are exponential and preemption is allowed, the potential method of [BPT92a] and [BPT92b] was shown to lead to an exact characterization of the achievable region.

Given that the achievable region is a polyhedron, the problem of finding an optimal policy amounts to a linear programming problem. Since it is already known that optimal policies are strict priority rules, it is hardly surprising that the extreme points of the achievable region are the performance vectors of such priority rules. Note that if linear side constraints are imposed, the performance of an optimal policy is still a linear programming problem. In particular, an optimal policy can be expressed as a policy that randomizes between a number of strict priority rules. In addition, the problem of finding the probability with which each particular priority rule is to be used is the same as the problem of expressing an element of a polyhedron as a convex combination of its extreme points. This latter problem can be also solved, in principle, using linear programming techniques.

Unfortunately, the polyhedral characterizations discussed so far involve a number of constraints which is exponential in the number of customer classes. Therefore, even though linear programming problems are solvable in polynomial time, the naive application of the preceding ideas to the side-constrained problem leads to exponential time algorithms. For this reason, we use an alternative method developed by the authors [BPT92b] and Kumar and Kumar [KK92] whereby the achievable region is bounded in terms of a new polyhedron $Q$ that involves a number of variables and constraints which is quadratic in the number of customers. We establish in this paper, that the achievable region is equal to the image of such a polyhedron $Q$ under a linear mapping into a lower-dimensional space. In particular, the side-constrained problem can be now solved efficiently as a linear programming problem involving the polyhedron $Q$. As will be shown later, some of the extreme points of $Q$ do not correspond to strict priority rules. Thus, although, we can express any element of $Q$ as a combination of its extreme points, this does not solve for us the problem we are actually interested in: expressing an element of the achievable region as a combination of its extreme points. Later in this paper, we will manage to develop a polynomial time algorithm for the latter problem; as it turns out, this is much more complicated than it might have appeared at first sight.

We refer briefly to some earlier work on variations of the Klimov problem, involving side constraints. Nain and Ross [NR86] consider a multiclass M/GI/1 queue with a single side constraint and establish that an optimal policy randomizes between two priority policies.
Makowski and Shwartz [MS93] derive similar structural results for the Klimov problem; their methods are easily generalized to the branching bandits model as well. Nevertheless, in the absence of a polyhedral characterization of the achievable region, their methods do not seem to lead to usable algorithms for computing the optimal cost or an optimal policy, especially when more than one side constraints are present.

We wish to summarize at this point the technical contributions of this paper:

1. We derive a “parsimonious” characterization of the achievable region for the branching bandits problem, involving only a quadratic number of variables and constraints. This should be contrasted with all previous work, in which the characterizations involve an exponential number of constraints.

2. We extend the methodology developed in [BPT92a] and then refined in [BPT92b] and [KK92] to characterize the achievable region of stochastic systems with general distributions; earlier work could only handle exponential distributions.

3. We give a polynomial time algorithm to solve the branching bandit problem with side constraints. More generally, we derive a polynomial time algorithm for expressing an element of a polyhedron as a convex combination of its extreme points, when the polyhedron is specified as the projection of a higher-dimensional polyhedron. This algorithm could be of independent interest.

The rest of the paper is organized as follows: In Section 2, we formally define the problem and establish our notation. In Section 3, we characterize the achievable region for the vector $n^+$ of mean queue lengths as observed on a typical service completion time. In Section 4, the same achievable region is described as a projection of a higher-dimensional polyhedron. In Section 5, we provide analogs of the results of Sections 3 and 4, regarding the achievable region for the vector $n$ of mean queue lengths. In Section 6, we discuss how to specialize the results of Section 5 to Klimov’s problem. In Section 7, we bring side constraints into the picture and establish the structure of optimal policies. In addition, we develop a polynomial time algorithm for computing the coefficients needed to specify an optimal policy. Section 8 contains some concluding remarks.

## 2 Problem Formulation

In this section, we define the average cost branching bandits problem, as well as the special case known as Klimov’s problem. We also define our notation and terminology.

Let there be given a set $\mathcal{R}_0 = \{0, 1, 2, \ldots, R\}$ of $R + 1$ customer classes and a single server who keeps serving available customers. We assume that there is always an available customer. At any service completion time, the server chooses a customer, say of class $i$, to
serve next. The duration of that customer’s service is a positive random variable $T_i$. At the time of the service completion, the customer just served disappears and is replaced by $N_{i0}, N_{i1}, \ldots, N_{iR}$, customers of classes 0, 1, \ldots, $R$, respectively, with each $N_{ij}$ a nonnegative integer random variable. For any $i \in R_0$, we assume that the joint distribution of the random variables $(T_i, N_{i1}, \ldots, N_{iR})$ is given and is the same each time a class $i$ customer is served. We also assume that the realizations of the random vector $(T_i, N_{i1}, \ldots, N_{iR})$ corresponding to services of different customers (of the same or of different classes) are statistically independent.

The model just described assumes that the service of a customer cannot be interrupted, which means that we are only considering non-preemptive policies. Finally, we assume that $N_{00}$ is equal to 1, with probability 1, and that $N_{i0} = 0$ for every $i \neq 0$. Thus, if we start with a single customer of class 0, there will always be exactly one such customer; in particular, our assumption that there is always an available customer is satisfied.

We now define Klimov’s problem and then argue that it is a special case of the branching bandits model. We have a single server who serves customers belonging to a set $\mathcal{R} = \{1, \ldots, R\}$ of different customer classes. Customers of each class $i \in \mathcal{R}$ arrive in the system according to an independent Poisson process with rate $\lambda_i$ and require a random service time with mean $m_i$ and second moment $\sigma_i^2$. The service times of the customers of each class are independent and identically distributed. Service times of customers of different classes are independent. Finally, service times are independent of the arrival process. Upon service completion, a class $i$ customer is fed back to the system as a class $j$ customer, with probability $p_{ij}$, or leaves the system, with probability $p_{i0} = 1 - \sum_{j=1}^{R} p_{ij}$. We assume again that service is non-preemptive. At any service completion time, the server can choose an available customer, if any, to be served next. It can also decide to stay idle. If it decides to stay idle, it is natural to stay idle until the “state” of the system changes and this can only happen if there is a new arrival. We therefore impose the additional assumption that an idle period can only be terminated by a new arrival.

We now indicate how Klimov’s model can be obtained as a special case of our variant of the branching bandits model. We identify idling in Klimov’s problem with serving a class 0 customer in the branching bandits model. Since idling is supposed to last until the next arrival, $T_0$ has an exponential distribution with mean $\lambda = \lambda_1 + \cdots + \lambda_R$. In addition, the vector $(N_{01}, \ldots, N_{0R})$ is the $j$th unit vector with probability $\lambda_j / \lambda$. (This is the probability that the arriving customer that interrupts the idling period is of class $j$.) We also let $N_{00} = 1$ and $N_{i0} = 0$ for $i \neq 0$. If a class $i$ customer is served, the mean service time is $E[T_i] = m_i$ and the second moment is $\sigma_i^2$. Finally, $N_{ij}$, for $i, j \neq 0$, is equal to the number of class $j$ Poisson arrivals during the service time $T_i$, to which number we must add 1 if the customer served was transformed to a class $j$ customer. In particular, we have

$$E[N_{ij}] = m_i \lambda_j + p_{ij}, \quad i, j = 1, \ldots, R,$$ (1)
\[ E[N_{ij}^2] = \lambda_j^2 \sigma_i^2 + m_i \lambda_j + p_{ij} + 2m_i \lambda_j p_{ij}, \quad i, j = 1, \ldots, R. \]  

(2)

(In deriving the last formula, we have used the fact that the second moment of the number of Poisson arrivals with rate \( \lambda_j \), during the service time \( T_i \), is \( \lambda_j^2 \sigma_i^2 + m_i \lambda_j \).)

Let \( N_r(t) \) be the number of class \( r \) customers present in the system at time \( t \), assumed to be a right-continuous function of time. In particular, if \( r \) is a service completion time, then \( N_r(\tau) \) refers to the number of customers right after the service completion. The vector \( N(t) = (N_1(t), \ldots, N_R(t)) \) will be called the state of the system at time \( t \). (By our assumptions, \( N_0(t) \) is the same for all times, and, therefore, does not need to be included in the state vector.) Finally, let \( \{\tau_k\} \) be the sequence of service completion times.

**Definition 2.1**  
a) We say that a policy gives priority to class \( i \) over class \( j \) if there is zero probability of choosing a class \( j \) customer to serve while class \( i \) customers are available.  
b) We say that a policy is non-idling if it gives priority to class \( i \) over class 0, for all \( i \neq 0 \).  
c) For any subset \( S \) of \( \{1, \ldots, R\} \), we say that a policy is an \( S \)-priority if it gives priority to class \( i \) over class \( j \) for every \( i \in S \) and every \( j \notin S \).  
d) We say that a policy is a priority if it is non-idling and there exists some ordering \( (i_1, i_2, \ldots, i_R) \) of the set \( \{1, \ldots, R\} \) such that the policy gives priority to class \( i_k \) over class \( i_{k+1} \), for \( k = 1, \ldots, R - 1 \).  

**Assumption A**  
a) The \( R \times R \) matrix \( N \) with entries \( E[N_{ij}] \), \( i, j = 1, \ldots, R \), has spectral radius smaller than 1.  
b) The random variables \( N_{ij} \) and \( T_i \) are of exponential type for every \( i \) and \( j \); that is, there exists some \( \lambda > 0 \) such that \( E[e^{\lambda N_{ij}}] < \infty \) and \( E[e^{\lambda T_i}] < \infty \).  

Part (b) of the above assumption is much stronger than needed, but we introduce it in order to avoid certain technical digressions. In the last section of the paper, we comment on how this assumption can be relaxed.

Assumption A guarantees that the stochastic process \( N(\tau_k) \) is "stable" under all non-idling policies [BNM92]. For a self-contained proof, let \( w = (w_1, \ldots, w_R) \) be a positive vector and \( \delta \) be a positive scalar satisfying

\[ \sum_{j=1}^{R} E[N_{ij}] w_j \leq w_i - \delta, \quad i = 1, \ldots, R. \]

[Such a vector exists by the Perron–Frobenius theorem and Assumption A(a).] It follows that for every nonidling policy and for every time \( \tau_k \) for which \( N(\tau_k) \neq 0 \), we have

\[ \sum_{i=1}^{R} E[N_{i}(\tau_{k+1}) | N(\tau_k)] w_i \leq \sum_{i=1}^{R} N_{i}(\tau_k) w_i - \delta. \]
Thus, $\sum_{i=1}^{R} N_i(\tau_k) w_i$ has negative drift away from the origin. In particular, if $N(\tau_k)$ is a Markov chain under the policy under consideration (in which case, we say that the policy is Markovian), this Markov chain is geometrically ergodic [Haj82] and all of its moments are finite under the corresponding ergodic distribution.

Let $\Pi^+$ be the set of all stationary policies that result into a discrete time stochastic process $\{N(\tau_k)\}_{k=-\infty}^{\infty}$ with a unique stationary distribution satisfying $E[N_i^2(\tau_k)] < \infty$ for all $i \in \{1, \ldots, R\}$. According to the preceding discussion, $\Pi^+$ contains all nonidling stationary Markovian policies. For any policy $\pi \in \Pi^+$, let $n_i^+$ be the expectation of $N_i(\tau_k)$ under the corresponding stationary distribution. Let $n^+ = (n_1^+, \ldots, n_R^+)$. Let $X^+$ (respectively, $X_{ni}^+$) be the set of all vectors $n^+$ that can be obtained by considering different policies in $\Pi^+$ (respectively, non-idling policies in $\Pi^+$). We will refer to $X^+$ (respectively, $X_{ni}^+$) as the achievable region for $n^+$ under all (respectively, non-idling) policies. A complete characterization of $X^+$ and $X_{ni}^+$ is obtained in the next section.

The performance vector $n^+$ refers to the average number of customers of each class that are present in the system at a typical completion time. Alternatively, we may be interested in $n$, the steady-state mean of $N(t)$. We let $\Pi$ be the set of all stationary policies that result into a continuous time stochastic process $\{N(t)\}_{t=-\infty}^{\infty}$ with a unique stationary distribution satisfying $E[N_i^2(t)] < \infty$ for all $i \in \{1, \ldots, R\}$. Under Assumption A, every nonidling policy can be shown to belong to $\Pi$. The achievable region for $n$ under policies in $\Pi$ (respectively, under non-idling policies in $\Pi$) is denoted by $X$ (respectively, by $X_{ni}$). These regions are studied in Section 5.

The following table provides a brief summary of our notation.

| $n^+$ | vector of average number of customers at service completions |
| $n$ | vector of steady-state mean number of customers |
| $\rho^+$ | vector of traffic intensities at service completions |
| $\rho$ | vector of steady-state traffic intensities |
| $X^+$ (resp. $X_{ni}^+$) | achievable region for $n^+$ (resp. under nonidling policies) |
| $X$ (resp. $X_{ni}$) | achievable region for $n$ (resp. under nonidling policies) |
| $P^+$ (resp. $P_{ni}^+$) | exponential in size characterization of $X^+$ (resp. $X_{ni}^+$) |
| $P$ (resp. $P_{ni}$) | exponential in size characterization of $X$ (resp. $X_{ni}$) |
| $U^+$ (resp. $U_{ni}^+$) | polynomial in size characterization of $X^+$ (resp. $X_{ni}^+$) |
| $U$ (resp. $U_{ni}$) | polynomial in size characterization of $X$ (resp. $X_{ni}$) |

Table 1: Notation summary.
3 Derivation of the Achievable Region for $n^+$

The line of development in this section is as follows. We first derive a set of linear inequalities that have to be satisfied by the vector $n^+$ under every policy. These constraints define a polyhedron and we show that its extreme points are the vectors $n^+$ corresponding to priority policies. We then conclude that the achievable region is equal to this polyhedron.

We start with a few definitions. We use $X_i(t)$ to denote the indicator function of the event that at time $t$ the server is serving a customer of class $i$. We assume that $X_i(t)$ is a right-continuous function of time so that $X_i(t_k)$ is 1 if at time $t_k$ a class $i$ customer starts being served. For any policy in $\Pi^+$, we let

$$\rho_i^+ = E[X_i(t_k)],$$

where the expectation is taken with respect to the stationary distribution. The next lemma states that $\rho_i^+$ is the same for all policies. The proof, as well the proofs of several other results, relies on the following formula that describes the evolution of the system:

$$N_i(t_{k+1}) = N_i(t_k) + \sum_{j=0}^{R} X_j(t_k)(N_{ji} - \delta_{ij}),$$

where $\delta_{ij}$ is the Kronecker delta.  

**Lemma 3.1** The value of $\rho_i^+$ is the same for all policies in $\Pi^+$ and can be obtained as the unique solution of the system of equations

$$\sum_{j=0}^{R} \rho_j^+ E[N_{ji}] = \rho_i^+, \quad i = 1, \ldots, R,$$

and

$$\sum_{i=0}^{R} \rho_i^+ = 1.$$  

**Proof:** Fix a policy in $\Pi^+$. By taking expectations of both sides of Eq. (3) with respect to the stationary distribution, we obtain Eq. (4). Equation (5) follows from the definition of $\rho_i^+$.

Let $\rho = (\rho_1^+, \ldots, \rho_R^+)$ and let $u = (E[N_{01}], \ldots, E[N_{0R}])$. Then, Eq. (4) can be rewritten as

$$\rho'N + \rho_0u' = \rho'.$$  

\footnote{Strictly speaking, we should have used a notation like $N_{ji}(\tau_k)$ instead of simply $N_{ji}$, to indicate the fact that $N_{ji}$ is selected independently after each service completion of a class $j$ customer.}
Because of Assumption A(a), the matrix \( I - N \) is invertible and \((I - N)^{-1} = I + N + N^2 + \cdots \) is a nonnegative matrix. We therefore have \( \rho' = \rho_0 w'(I - N)^{-1} = \rho_0 w' \), where \( w' \) is the nonnegative row vector \( w'(I - N)^{-1} \). Equation (5) can then be used to determine \( \rho_0 \) uniquely.

For the remainder of the paper, we impose the following assumption which is meant to exclude certain degenerate cases.

**Assumption B** For every class \( i \in \{0, 1, \ldots, R\} \), we have \( \rho_i^+ > 0 \).

Under Assumption A, the system is stable and we are guaranteed that \( \rho_0^+ > 0 \). We then see that Assumption B is guaranteed to hold if the vector \( u \) is nonzero and the matrix \( I + N + N^2 + \cdots \) is positive.

Let \( S \) be some nonempty subset of \( \{1, \ldots, R\} \). We define a set of parameters \( f^+_S, i \in S \), by means of the system of equations

\[
1 + \sum_{i \in S} E[N_{ji}] f^+_S = f^+_S, \quad j \in S. 
\]

Notice that this is a linear system of the form \((I - A)\xi = e\), where \( e \) is a vector with all entries equal to 1. Here \( A \) is a square submatrix of the nonnegative matrix \( N \) which has been assumed to have spectral radius less than 1. It follows that the spectral radius of \( A \) is also less than 1, \( I - A \) is invertible and \((I - A)^{-1} = I + A + A^2 + \cdots \) is a nonnegative matrix. This establishes that the coefficients \( f^+_S \) are uniquely defined and are nonnegative. We then use Eq. (6) once more to conclude that the coefficients \( f^+_S \) are in fact positive.

We note that \( f^+_S \) can be interpreted as the expected number of customers served under an \( S \)-priority policy until we run out of customers whose class belongs to \( S \), and if we started with a single customer of class \( j \).

**Theorem 3.2** For every nonempty subset \( S \) of \( R = \{1, \ldots, R\} \), and any policy in \( \Pi^+ \), we have

\[
\sum_{i \in S} f^+_S n^+_i \geq G^+(S),
\]

where

\[
G^+(S) = \frac{1}{2} \sum_{j=0}^{R} \rho_j^+ E[(\sum_{i \in S} f^+_S (N_{ji} - \delta_{ij}))^2].
\]

The inequality (7) holds with equality if and only if we have an \( S \)-priority policy.

**Proof**: Let \( R_S(t) = \sum_{i \in S} f^+_S N_i(t) \). We use Eq. (3) and obtain

\[
R_S(\tau_{k+1}) = R_S(\tau_k) + \sum_{i \in S} f^+_S \sum_{j=0}^{R} \chi_j(\tau_k) (N_{ji} - \delta_{ij})
\]

The equation in the paper is:

\[
R_S(\tau_{k+1}) = R_S(\tau_k) + \sum_{i \in S} f^+_S \sum_{j=0}^{R} \chi_j(\tau_k) (N_{ji} - \delta_{ij})
\]

The next equation is:

\[
G^+(S) = \frac{1}{2} \sum_{j=0}^{R} \rho_j^+ E[(\sum_{i \in S} f^+_S (N_{ji} - \delta_{ij}))^2].
\]
\[ R \equiv R_S(\tau_k) + \sum_{j=0}^{R} \chi_j(\tau_k) \sum_{i \in S} f^+_{Si}(N_{ji} - \delta_{ij}). \]  

(8)

We square both sides of (8), use the fact \( \chi_i(\tau_k)\chi_j(\tau_k) = \delta_{ij} \), and take expectations with respect to the stationary distribution corresponding to the policy under consideration. Using also the fact \( E[R^2_S(\tau_{k+1})] = E[R^2_S(\tau_k)] \), we obtain

\[ 2E[R_S(\tau_k) \sum_{j=0}^{R} \chi_j(\tau_k) \sum_{i \in S} f^+_{Si}(N_{ji} - \delta_{ij})] + \sum_{j=0}^{R} \rho_j^2 E[\sum_{i \in S} (f^+_{Si}(N_{ji} - \delta_{ij}))^2] = 0. \]  

(9)

Notice that the second term in the left-hand side of Eq. (9) is \( 2G^+(S) \), by definition. We now have,

\[
E[R_S(\tau_k)] \geq E[R_S(\tau_k) \sum_{j \in S} \chi_j(\tau_k)] \\
= -E[R_S(\tau_k) \sum_{j \in S} \chi_j(\tau_k) \sum_{i \in S} f^+_{Si}(N_{ji} - \delta_{ij})] \\
= G^+(S) + E[R_S(\tau_k) \sum_{j \not\in S} \chi_j(\tau_k) \sum_{i \in S} f^+_{Si}(N_{ji} - \delta_{ij})] \\
= G^+(S) + E[R_S(\tau_k) \sum_{j \not\in S} \chi_j(\tau_k) \sum_{i \in S} f^+_{Si}N_{ji}] \\
\geq G^+(S). 
\]

The first inequality follows from \( \sum_{j \in S} \chi_j(\tau_k) \leq 1 \); the first equality from Eq. (9) and (6); the second equality from Eq. (9); the third equality because \( i \in S \) and \( j \not\in S \) imply \( \delta_{ij} = 0 \).

Notice that the equality \( E[R_S(\tau_k)] = G^+(S) \) is obtained if and only if

\[ R_S(\tau_k) \sum_{j \not\in S} \chi_j(\tau_k) = 0, \quad \text{w.p.1;} \]

equivalently, if and only if \( N_i(\tau_k)\chi_j(\tau_k) = 0 \) for all \( i \in S \) and \( j \not\in S \). This is equivalent to having an \( S \)-priority policy.

Notice that non-idling policies are the same as \( \mathcal{R} \)-priority policies. It follows that the inequality \( \sum_{i=1}^{R} f^+_{R_i} n_i^+ \geq G^+(\mathcal{R}) \) becomes an equality if and only if the policy is non-idling.

Theorem 3.2 provides us with \( 2^R - 1 \) linear inequality constraints on the vector \( n^+ \), one for each nonempty subset of \( \{1, \ldots, R\} \). These inequality constraints define a polyhedron in \( R \)-dimensional space, which we will denote by \( P^+ \). Let us also define \( P^+_{ni} \) as the subset of \( P^+ \) on which the equality \( \sum_{i=1}^{R} f^+_{R_i} n_i^+ = G^+(\mathcal{R}) \) holds. (Note that \( P^+_{ni} \) is a bounded polyhedron while \( P^+ \) is unbounded.) Theorem 3.2 establishes that \( X^+ \subset P^+ \) and \( X^+_{ni} \subset P^+_{ni} \). We wish to show that \( X^+ = P^+ \) and \( X^+_{ni} = P^+_{ni} \); that is, that we have a complete characterization of the achievable region for the branching bandits problem under general (or non-idling,
respectively) policies. Our first step is to characterize the extreme points of $P_{n_i}^+$. 

**Theorem 3.3** A vector is an extreme point of the set $P_{n_i}^+$ if and only if it is equal to the performance vector $n^+$ corresponding to some priority policy.

**Proof:** Given a set of inequality constraints that define a polyhedron, we say that a constraint is active at those points at which it is satisfied with equality. Recall that an element of a polyhedron in $\mathbb{R}^R$ is an extreme point if and only if there are $R$ linearly independent constraints that are active at that point.

Consider the priority policy corresponding to the ordering $(1, 2, \ldots, R)$. This policy is an $S$-priority for every set $S$ of the form $\{1, \ldots, i\}$ and the inequality $\sum_{i \in S} f_{S_i}^+ n_i^+ \geq G^+(S)$ is satisfied with equality for every such $S$. Notice that the $R$ equalities thus obtained form a triangular system of equations and are therefore linearly independent. It follows that the vector $n^+$ is an extreme point of $P_{n_i}^+$. The same argument applies to any other priority policy.

In order to show that every extreme point corresponds to a priority policy, we observe that $P_{n_i}^+$ satisfies the definition of an extended polymatroid and the result follows from Theorem 1 of [BNM92]. We provide here an alternative self-contained proof.

Let us introduce the additional assumption that under any policy and for any $i, j$, there is a positive probability that customers of classes $i$ and $j$ may coexist. Consider an extreme point of $P_{n_i}^+$ that corresponds to some priority policy, say the priority policy corresponding to the ordering $(1, 2, \ldots, R)$. Theorem 3.2 implies that the constraints $\sum_{i \in S} f_{S_i}^+ n_i^+ = G^+(S)$ are active, for every $S$ of the form $S = \{1, \ldots, i\}$. If there are more than $R$ active constraints at $n^+$, we must also have $\sum_{i \in S} f_{S_i}^+ n_i^+ = G^+(S)$ for some $S \subset \{1, \ldots, R\}$ which is not of this form; in particular, there exist $i$ and $j$ such that $i < j$, $i \notin S$ and $j \in S$. Thus, $j$ must have priority over $i$. On the other hand, since $i < j$, $i$ must also have priority over $j$. This can only happen if customers of classes $i$ and $j$ can never coexist under the priority policy under consideration, which contradicts our earlier assumption. We conclude that there are exactly $R$ active constraints at every extreme point corresponding to a priority policy.

We say that two extreme points are adjacent if there are $R - 1$ constraints that are active at both points. Since the constraint corresponding to $S = \{1, \ldots, R\}$ is satisfied at all points, it follows that an extreme point can have up to $R - 1$ adjacent extreme points. We say that two priority policies are adjacent if one can be obtained from the other by interchanging the order of two classes that are ordered consecutively. [For example, the priority ordering $(1, 2, 3, 4)$ is adjacent to $(1, 3, 2, 4)$ but is not adjacent to $(1, 3, 4, 2)$.] It is seen that for adjacent priority policies there are $R - 1$ common active constraints and therefore the corresponding extreme points are adjacent. We conclude that if we have an extreme point that corresponds to a priority policy, all of its $R - 1$ adjacent extreme points correspond to priority policies. It is well known that if we keep moving from an extreme
point of a bounded polyhedron to an adjacent extreme point, every extreme point can be reached. Therefore, all extreme points of \( P_{n_i}^+ \) correspond to priority policies.

The general case in which two customer types might have zero probability of coexisting can be viewed as the limit of a sequence of problems in which the probability of coexisting is positive and tends to zero. The result then follows from a continuity argument whose details are omitted. ■

**Corollary 3.4** There holds \( X_{n_i}^+ = P_{n_i}^+ \).

**Proof**: From Theorem 3.2, we have \( X_{n_i}^+ \subseteq P_{n_i}^+ \). Consider a collection of priority policies \( \pi^1, \ldots, \pi^K \) whose performance vectors are \( x^1, \ldots, x^K \). Consider also a policy that at the beginning of every busy period \(^2\) decides with probability \( p \) that policy \( \pi^i \) will be followed for the entire duration of the busy period. It is then easily seen that this is a non-idling policy and its performance vector is \( \sum_{i=1}^K p \cdot x^i \). This establishes that every element of \( P_{n_i}^+ \) is the performance vector of some non-idling policy of this type. ■

We note that in the preceding proof, a value of \( K \) larger than \( R + 1 \) is never needed, by virtue of Caratheodory's theorem.

We now turn our attention to policies that are not necessarily non-idling. We first extend Theorem 3.3.

**Theorem 3.5** The polyhedra \( P_{n_i}^+ \) and \( P^+ \) have the same set of extreme points.

**Proof**: At any extreme point of \( P_{n_i}^+ \) there are \( R \) linearly independent active constraints and therefore we also have an extreme point of \( P^+ \). We now prove the converse. If \( P^+ \) has more extreme points than \( P_{n_i}^+ \), then there are two adjacent extreme points of \( P^+ \) such that one, call it \( z \), is an extreme point of \( P_{ni}^+ \) and the other, call it \( y \), is not. Assume for simplicity that \( z \) is associated to the priority ordering \( (1, 2, \ldots, R) \). From the point \( z \), we can move to an adjacent extreme point (along an edge) if exactly one of the active constraints becomes inactive. If any constraint other than the constraint \( \sum_{i=1}^R f_{i}^+ n_i^+ \geq G^+(\mathcal{R}) \) becomes inactive, we end up at another extreme point of \( P_{ni}^+ \). Therefore, in order to reach \( y \), the constraint \( \sum_{i=1}^R f_{i}^+ n_i^+ \geq G^+(\mathcal{R}) \) must become inactive. Recall, that the active constraints at the point \( z \) form a triangular system of equations. Therefore, by making the constraint \( \sum_{i=1}^R f_{i}^+ n_i^+ \geq G^+(\mathcal{R}) \) inactive, the variable \( n_R^+ \) becomes free. The value of that variable can be increased without limit without violating any of the constraints associated with \( P^+ \). This means that the corresponding edge that starts at \( z \) does not end at another extreme point. ■

\(^2\) A busy period starts at a moment where a zero state vector becomes nonzero; it ends at the first time that the state becomes again zero.
We will next characterize the points that lie on infinite edges of $P^+$. We first need to define a set of policies pertinent to this problem. Consider an ordering $\sigma$ of the classes $1, \ldots, R$, and relabel the classes such that $\sigma = (1, 2, \ldots, R)$. Let $\pi(p)$ be the policy under which:

a. Class $i$ always has priority over class $j$, if $i < j \leq R$.

b. The policy never idles when there are available customers of some class $i < R$.

c. Whenever all available customers are of class $R$, there is a constant probability $p$ of idling.\(^3\)

We refer to all such policies as *almost-priority* policies.

Recall that the vector $n^+$ associated to a priority policy can be obtained by solving a triangular system of linear equations. We will now describe a procedure for determining the vector $n^+$ associated with an almost-priority policy. Let us consider the almost priority policy $\pi(p)$ associated with the ordering $(1, \ldots, R)$. Under this policy, each time that there are only customers of class $R$ available, we will have

$N_i(\tau_{k+1}) = N_i(\tau_k) + (1 - \chi)(N_{Ri} - \delta_{Ri}) + \chi N_{0i}, \quad i = 1, \ldots, R,$

where $\chi$ is a binary random variable which is independent of everything else and is equal to 1 with probability $p$. Equivalently,

$N_i(\tau_{k+1}) = N_i(\tau_k) + (1 - \chi)N_{Ri} + \chi(N_{0i} + \delta_{Ri}) - \delta_{Ri}, \quad i = 1, \ldots, R.$

This implies that under policy $\pi(p)$, the system evolves exactly the same as if there were no idling and $N_{Ri}$ were replaced by $\tilde{N}_{Ri} = (1 - \chi)N_{Ri} + \chi(N_{0i} + \delta_{Ri})$, for $i = 1, \ldots, R$. Therefore, the vector $n^+$ associated with an almost priority policy can be found by evaluating the vector $n^+$ associated with a priority policy in a new branching bandits problem with a different distribution for the random variables $N_{Ri}, i = 1, \ldots, R$. In the new branching bandits problem, the matrix $N$ is replaced by a matrix $\tilde{N}(p)$ that differs from $N$ only at the last row; in particular, the $(R, j)$th entry of $\tilde{N}(p)$ is equal to $(1 - p)E[N_{Rj}] + pE[N_{0j}] + p\delta_{Rj}$.

Let us define $p^* = \rho_0^+/(\rho_0^+ + \rho_R^+)$, where the coefficients $\rho_i^+$ are those corresponding to the original matrix $N$, as in Lemma 3.1. We then have the following result.

**Lemma 3.6** The spectral radius of $\tilde{N}(p)$ is less than 1 for $p < p^*$ and equal to 1 for $p = p^*$.

---

\(^3\)Note that this is the same as the Markovian policy that uses the priority ordering $(1, 2, \ldots, R, 0)$ with probability $1 - p$ and the priority ordering $(1, 2, \ldots, R - 1, 0, R)$ with priority $p$. 

12
Proof: We start from the fact that the coefficients $\rho_i^+$ satisfy Eq. (4), use the definitions of $\tilde{N}(p)$ and $p^*$, and do some straightforward algebra to verify that the vector $(\rho_1^+, \rho_2^+, \ldots, \rho_{R-1}^+, \rho_0^* + \rho_R^*)$ is a left eigenvector of $\tilde{N}(p^*)$, with eigenvalue 1. In addition, notice that the determinant of $I - \tilde{N}(p)$ is affine in $p$. Therefore, for every $p \neq p^*$, the determinant of $I - \tilde{N}(p)$ is nonzero and the spectral radius of $\tilde{N}(p)$ is different from 1. Since the spectral radius is less than 1 for $p = 0$ (Assumption A), a continuity argument implies the same for all values of $p$ between 0 and $p^*$. ■

Under the almost-priority policy $\pi(p)$, the values of $\rho_i^+$ and $n_i^+$ remain the same for $i = 1, \ldots, R - 1$. It remains to determine how $\rho_R^+$ and $n_R^+$ vary with $p$ and we will be using the notation $\rho_R^+(p)$ and $n_R^+(p)$, in order to make this dependence explicit. In addition, we let $f_{Ri}^+(p)$, $i = 1, \ldots, R$, stand for the unique solution of Eq. (6) when $N_{ij}$ is replaced by $\tilde{N}_{ij}(p)$ and when $S$ is equal to $R = \{1, \ldots, R\}$. Using Cramer's rule, we see that $f_{Ri}^+(p)$ is the ratio of two affine functions of $p$, the denominator being the determinant of $I - \tilde{N}(p)$. Since the latter determinant becomes zero when $p = p^*$, we conclude that the denominator can be taken to be $p^* - p$.

We also note that $(1 - p)\rho_R^+(p) = \rho_R^+$. (Intuitively, this expresses the fact that a fraction $1 - p$ of all class $R$ services in the modified model corresponds to class $R$ services in the original model.) Concerning $n_R^+(p)$, it can be determined from the relation

$$\sum_{i=1}^{R} f_{Ri}^+(p)n_i^+(p) = \frac{1}{2} \sum_{j=0}^{R} \rho_j^+(p)E[(\sum_{i=1}^{R} f_{Ri}^+(p)(\tilde{N}_{ji} - \delta_{ij}))^2].$$

Using our earlier discussion on the dependence of $\rho_i^+(p)$ and $f_{Ri}^+(p)$ on $p$, we conclude that $n_R^+(p)$ is a rational function of $p$ with a term of the form $p^* - p$ appearing in the denominator. This implies that $n_R^+(p)$ tends to infinity as $p$ increases to $p^*$. In addition, $p$ can be determined from $n_R^+(p)$ by solving a polynomial equation in $p$.

We summarize this discussion in the following theorem.

**Theorem 3.7** Any point on an infinite edge of $P^+$ is the performance vector of some almost-priority policy. In addition, the value of $p$ that corresponds to any given point can be determined by solving a polynomial equation.

Using the same argument as in the proof of Corollary 3.4, we conclude:

**Corollary 3.8** There holds $X^+ = P^+$. 

13
4 A Parsimonious Representation of the Achievable Region

The polyhedra $P^+$ and $P^+_{ni}$ provide an exact representation of the achievable regions $X^+$ and $X^+_{ni}$, respectively. Their drawback is that they are specified in terms of an exponential number of constraints. In this section, we use the approach of [BPT92b] and [KK92], to obtain an equivalent but more compact representation. This new representation involves $R(R + 1)$ variables but only $O(R^2)$ linear constraints.

The achievable region will be represented in terms of the auxiliary variables

$$I_{ji} = E[\chi_j(\tau_k)N_i(\tau_k)], \quad i = 1, \ldots, R, \quad j = 0, \ldots, R. \quad (10)$$

Let $I$ stand for the $R(R + 1)$-dimensional vector with components $I_{ij}$. Notice that $I_{ji} = 0$ if and only if $N_i(\tau_k) > 0$ implies $\chi_j(\tau_k) = 0$; that is, if and only if class $i$ has priority over class $j$. In particular, a policy is nonidling if and only if $I_{0i} = 0$ for all $i \neq 0$.

**Theorem 4.1** For every policy in $\Pi^+$, the vector $I$ belongs to the polyhedron $Q^+$ defined as the set of all nonnegative vectors $z$ with components $z_{ji}$, $j = 0, \ldots, R$, $i = 1, \ldots, R$, that satisfy the following linear equality constraints:

$$\sum_{j=0}^{R} \rho_j^+ E[(N_{ji} - \delta_{ji})^2] + 2 \sum_{j=0}^{R} z_{ji} E[N_{ji} - \delta_{ji}] = 0, \quad i = 1, \ldots, R, \quad (11)$$

and

$$\sum_{j=0}^{R} z_{jr} E[N_{jr'} - \delta_{jr'}] + \sum_{j=0}^{R} z_{jr} E[N_{jr} - \delta_{jr}] + \sum_{j=0}^{R} \rho_j^+ E[(N_{jr} - \delta_{jr})(N_{jr'} - \delta_{jr'})] = 0, \quad r, r' = 1, \ldots, R. \quad (12)$$

**Proof:** Consider the evolution equation (3). We square both sides, use the fact $\chi_i(\tau_k)\chi_j(\tau_k) = \delta_{ij}$, and take expectations with respect to the stationary distribution corresponding to the policy under consideration. Using also the fact $E[N_i(\tau_k+1)] = E[N_i(\tau_k)]$, we obtain Eq. (11). (In the derivation of this formula we have also used the independence of $N_{ij}$ from the state of the system.)

To derive Eq. (12) we use Eq. (3) to derive a recursion for $N_r(\tau_{k+1})N_{r'}(\tau_{k+1})$ and proceed similarly. \[ \square \]

Note that for every policy in $\Pi^+$, we have

$$n_i^+ = \sum_{j=0}^{R} I_{ji}, \quad i = 1, \ldots, R.$$
In particular, \( n^+ \) belongs to the set \( U^+ \) defined by

\[
U^+ = \{ x \geq 0 \mid \exists z \in Q^+ \text{ such that } x_i = \sum_{j=0}^{R} z_{ji} \}. 
\]

The set \( U^+ \) is the image of the polyhedron \( Q^+ \) under a particular linear mapping. Therefore, \( U^+ \) is also a polyhedron.

We have already shown that the achievable region \( X^+ \) is contained in \( U^+ \). It has been shown in [BPT92b], in much greater generality, that the use of auxiliary variables as in the proof of Theorem 4.1, always provides a smaller polyhedron than the one obtained using the method of the preceding section; thus, \( X^+ \subset U^+ \subset P^+ \). Since we have shown earlier that \( X^+ = P^+ \), we have the following main result.

**Theorem 4.2** There holds \( P^+ = U^+ = X^+ \).

Theorem 4.2 states that the achievable region \( X^+ \) is the image of the polyhedron \( Q^+ \). Given that \( Q^+ \) involves a much smaller (quadratic instead of exponential) number of constraints, this representation is much more suitable for the development of efficient algorithms.

A natural question to raise at this point is the following: is it true that every element of \( Q^+ \) is equal to the vector \( I \) associated to some policy in \( \Pi^+ \)? Interestingly enough, the answer is negative, as explained in the Appendix. In other words, the set \( Q^+ \) is larger than the achievable region for the vector \( I \), even though its image is exactly equal to the achievable region for the vector \( n^+ \). In particular, not every extreme point of \( Q^+ \) can be associated with an extreme point of \( P^+ \) and a priority policy.

If we are interested in nonidling policies, the preceding results are modified as follows. Notice that a policy is nonidling if and only if \( I_0 = 0 \) for all \( i \neq 0 \). We define \( Q_{ni}^+ \) as the subset of \( Q^+ \) in which the additional constraints \( x_{0i} = 0 \) hold for \( i = 1, \ldots, R \). By using the same reasoning as before, we conclude that \( X_{ni}^+ = U_{ni}^+ = P_{ni}^+ \).

## 5 Achievable Region for the Mean Queue Lengths

In this section we characterize the achievable region \( X \) (respectively, \( X_{ni} \)) for the vector \( n \) of mean queue lengths, under policies in \( \Pi \) (respectively, under non-idling policies in \( \Pi \)). In fact, we obtain two different characterizations which are similar to the characterizations of \( X^+ \) in terms of the polyhedra \( P^+ \) and \( Q^+ \).

We first establish a connection between the steady-state mean number of customers \( n_i \) and the mean number \( n_i^+ \) of customers at a typical service completion time. Let us denote by \( m_j \) the expectation of the service time \( T_j \) for a customer of class \( j \in \{0, \ldots, R\} \).

15
Lemma 5.1 For any policy in II, and for any \( i \in \{1, \ldots, R\} \), we have

\[
n_i = \frac{\sum_{j=0}^{R} m_j I_j}{\sum_{j=0}^{R} m_j \rho_j^+}.
\]

Proof: The general formula for passing from a Palm distribution to a stationary distribution (see, e.g., p. 226 of [Wal88]) states that \( n_i \), the steady-state mean of \( N_i(t) \), is given by

\[
n_i = \frac{E[\int_{\tau_k}^{\tau_{k+1}} N_i(\sigma) \, d\sigma]}{E[\tau_{k+1} - \tau_k]},
\]

where the expectations are taken with respect to the stationary distribution of the discrete-time Markov chain \( N(\tau_k) \). We have \( N_i(\sigma) = N_i(\tau_k) \) for \( \sigma \in [\tau_k, \tau_{k+1}) \), which leaves us with

\[
E[(\tau_{k+1} - \tau_k)N_i(\tau_k)]
\]

Note that \( E[\tau_{k+1} - \tau_k] = \sum_{j=0}^{R} m_j \rho_j^+ \). Furthermore,

\[
E[(\tau_{k+1} - \tau_k)N_i(\tau_k)] = \sum_{j=0}^{R} E[(\tau_{k+1} - \tau_k)N_i(\tau_k)\chi_j(\tau_k)] = \sum_{j=0}^{R} m_j I_{ji},
\]

which completes the proof.

We now define a polyhedron \( U \) as the image of \( Q^+ \) under the linear mapping suggested by Lemma 5.1. That is,

\[
U = \left\{ x \geq 0 \mid \exists \ z \in Q^+ \text{ such that } x_i = \frac{\sum_{j=0}^{R} m_j z_{ji}}{\sum_{j=0}^{R} m_j \rho_j^+} \right\}.
\]

If we are interested in nonidling policies only, we define \( U_{ni} \) similarly, except that \( Q^+ \) is replaced by \( Q^+_{ni} \). Theorem 4.1 and Lemma 5.1 readily imply that the achievable region \( X \) (respectively, \( X_{ni} \)) is contained in \( U \) (respectively, \( U_{ni} \)). We intend to show that \( U = X \) and \( U_{ni} = X_{ni} \). Our first step in this direction is to derive polyhedra \( P \) and \( P_{ni} \) with structure similar to the polyhedra \( P^+ \) and \( P_{ni}^+ \) that were derived in Section 3.

Let \( S \) be a nonempty subset of \( \{1, \ldots, R\} \). We define a set of parameters \( f_{Si}, i \in S \), by means of the system of equations

\[
m_j + \sum_{i \in S} E[N_{ji}] f_{Si} = f_{Sj}, \quad \forall j \in S.
\] (13)

This system of equations has a unique solution, which is positive, for the same reasons that were given when the coefficients \( f_{Si}^+ \) were defined.
**Theorem 5.2** For every nonempty subset $S$ of $\mathcal{R} = \{1, \ldots, R\}$, and any policy in $\Pi$, we have

$$\sum_{i \in S} f_{Si} n_i \geq G(S),$$  \hspace{1cm} (14)

where

$$G(S) = \frac{1}{2} \rho_j^+ E[\{\sum_{r \in S} f_{Sr} (N_{jr} - \delta_{jr})\}^2] \sum_{w=0}^{R} m_w \rho_w^+.$$

The inequality (14) holds with equality if and only if we have an $S$-priority policy.

**Proof:** Consider a policy $\pi \in \Pi$ and a subset $S$ of $\mathcal{R}$. Then, the vector $I_i$, with components $I_{ij}$ satisfies Eqs. (11) and (12). We multiply Eq. (12) by $f_{Sr} f_{Sr'}$ and sum over all $r, r' \in S$ such that $r > r'$. We then obtain

$$\sum_{j=0}^{R} \left( \sum_{r' \in S} f_{Sr'} E[N_{jr'} - \delta_{jr'}] \sum_{r \in S} f_{Sr} I_{jr} + \sum_{r \in S} f_{Sr} E[N_{jr} - \delta_{jr}] \sum_{r' \in S} f_{Sr'} I_{jr'} \right) = 0. \hspace{1cm} (15)$$

We also multiply Eq. (11) by $f_{Sr}'$, and sum over all $r' \in S$ to obtain

$$\sum_{j=0}^{R} \left( \sum_{r' \in S} f_{Sr}^2 E[N_{jr'} - \delta_{jr'}] I_{jr'} + \frac{1}{2} \rho_j^+ \sum_{r' \in S} f_{Sr'}^2 E[(N_{jr'} - \delta_{jr'})^2] \right) = 0. \hspace{1cm} (16)$$

We interchange $r$ and $r'$ in the second term of (15) and add the result to (16) to obtain

$$\sum_{j=0}^{R} \left( \sum_{r' \in S} f_{Sr} E[N_{jr'} - \delta_{jr'}] \sum_{r \in S} f_{Sr} I_{jr} + \frac{1}{2} \rho_j^+ \sum_{r' \in S} f_{Sr'}^2 E[(N_{jr'} - \delta_{jr'})^2] \right) = 0,$$

which yields

$$\sum_{j \in S} \sum_{r' \in S} f_{Sr} E[N_{jr'} - \delta_{jr'}] \sum_{r \in S} f_{Sr} I_{jr} + \sum_{j \in S} \sum_{r \in S} f_{Sr} E[N_{jr'} - \delta_{jr'}] \sum_{r \in S} f_{Sr} I_{jr} + A G(S) = 0,$$

where $A$ is defined by $A = \sum_{w=0}^{R} m_w \rho_w^+$. Using Eq. (13), we obtain

$$\sum_{r \in S} f_{Sr} \sum_{j \in S} m_j I_{jr} = \sum_{j \in S} \sum_{r \in S} f_{Sr} E[N_{jr'} - \delta_{jr'}] \sum_{r \in S} f_{Sr} I_{jr} + A G(S) \hspace{1cm} (17)$$

We now recall Lemma 5.1 and observe that

$$\sum_{j \in S} m_j I_{jr} \leq A n_r. \hspace{1cm} (18)$$

17
Thus, we obtain
\[ \sum_{r \in S} f_{S_r} n_r \geq \frac{1}{A} \sum_{j \not\in S} \sum_{r' \in S} f_{S_r} E[N_{j_{r'}} - \delta_{j_{r'}}] \sum_{r \in S} f_{S_r} I_{jr} + G(S) \geq G(S) \] (19)

because \( I_{jr}, f_{S_r} \) are non-negative and \( \delta_{j_{r'}} = 0 \) for \( j \not\in S \) and \( r' \in S \). It is easily checked that the inequalities in (19) hold with equality if and only if \( I_{jr} = 0 \) for \( j \not\in S \) and \( r \in S \); that is, if and only if the policy under consideration is an \( S \)-priority. \( \blacksquare \)

Since non-idling policies are the same as \( R \)-priority policies, the inequality
\[ \sum_{i \in R} f_{R_i} n_i \geq G(R) \]
becomes an equality if and only if the policy is non-idling. Theorem 5.2 provides us with \( 2^R - 1 \) linear inequality constraints on the vector \( n = (n_1, \ldots, n) \). These constraints define a polyhedron in \( R \)-dimensional space which we denote by \( P \). We also define \( P_{ni} \) to be the subset of \( P \) where the equality \( \sum_{i \in R} f_{R_i} n_i = G(R) \) holds. Theorem 5.2 asserts that \( X_{ni} \subset P_{ni} \) and \( X \subset P \).

The following is our main result.

**Theorem 5.3**

a) A vector is an extreme point of the set \( P_{ni} \) if and only if it is equal to the performance vector \( n \) corresponding to a priority policy.
b) The polyhedra \( P \) and \( P_{ni} \) have the same set of extreme points.
c) Any point on an infinite edge of \( P \) is the performance vector of some almost-priority policy.
d) There holds \( P = U = X \) and \( P_{ni} = U_{ni} = X_{ni} \).

**Proof:** (Outline) The proof of parts (a), (b) and (c) is identical to the proof of Thms. 3.3, 3.5 and 3.7, respectively.

Recall, that we have already shown that \( X \subset U \). Furthermore, in the course of the proof of Thm. 5.2, we showed that every element of \( U \) belongs to \( P \). Therefore, we have \( X \subset U \subset P \) and \( X_{ni} \subset U_{ni} \subset P_{ni} \). On the other hand, part (a) of this theorem states that the extreme points of \( P_{ni} \) belong to \( X_{ni} \), and it follows that \( X_{ni} = P_{ni} \). Similarly, parts (b)–(c) of this theorem imply that \( X = P \). \( \blacksquare \)

### 6 Klimov's Problem Revisited

In the branching bandits problem, the vector \( N(t) \) changes only at service completion times. In contrast, in Klimov's problem, external arrivals are Poisson and will generically occur during a service interval. This makes no difference if we are only watching the system at service completion times. In particular, all of the results in Sections 3 and 4 can be specialized to Klimov's problem by using Eqs. (1) and (2).
Let us now consider the mean number of class $i$ customers present in the system at some typical time $t$. This is equal to the mean number $n_i$, as determined from the branching bandits model, plus the expected number $a_i$ of class $i$ customers that have arrived since the last service completion, which occurred at some time $\tau$. We have

$$a_i = \sum_{j=0}^{R} \Pr(\chi_j(t) = 1) \lambda_i E[t - \tau \mid \chi_j(t) = 1].$$

Notice that

$$Pr(\chi_j(t) = 1) = \frac{m_j \rho_j^+}{\sum_{k=0}^{R} m_k \rho_k^+}.$$ 

In addition, $E[t - \tau \mid \chi_j(t) = 1] = \sigma^2_j / 2m_j$ and this determines $a_i$ completely. Notice that $a_i$ is the same for all policies in $\Pi$.

7 Branching Bandits with Side Constraints

In this section, we consider the branching bandits problem, in the presence of additional linear constraints on the vector $n$ of mean queue lengths. Let these side-constraints be of the form $An \geq b$, where $A$ is a matrix of dimensions $L \times R$. To keep the discussion simple, we only consider nonidling policies. In view of our characterization of the achievable region (Theorem 5.3), the cost of an optimal policy obeying the side-constraints can be found by solving the linear programming problem

$$\min c'x$$

subject to

$$x \in P_{ni}$$

$$Ax \geq b$$

We assume that this problem has a feasible solution.

The linear programming problem (20) is hard to solve because the polyhedron $P_{ni}$ is described by an exponential number of constraints. However, we recall that we have available a parsimonious representation of $P_{ni}$ of the form (Theorem 5.3)

$$P_{ni} = U_{ni} = \{ Fz \mid z \in Q_{ni}^+ \},$$

where $Q_{ni}^+$ is a polyhedron described in terms of a quadratic number of variables and constraints and where $F$ is a known linear mapping. It follows that problem (20) is equivalent to the linear programming problem

$$\min c'x$$

subject to

$$z = Fz$$

19
which is polynomial time solvable because it only has polynomial number of variables and constraints. We thus assume that we have computed, in polynomial time, an optimal solution \( x^* \) of problem (20).

Next, we express \( x^* \) as a convex combination of at most \( R + 1 \) extreme points of \( P_{ni} \). This is always possible, by Carathéodory’s theorem. (Later in this section, we show that this can be accomplished in polynomial time.) Let \( u^1, \ldots, u^{R+1} \) be these extreme points. Consider the problem

\[
\begin{align*}
\text{minimize} & \quad \sum_{j=1}^{R+1} \lambda_j (c' u^j) \\
\text{subject to} & \quad \sum_{j=1}^{R+1} \lambda_j = 1 \\
& \quad \sum_{j=1}^{R+1} \lambda_j (A u^j) \geq b \\
& \quad \lambda_j \geq 0
\end{align*}
\]

Since there is a feasible solution of this problem for which \( x^* = \sum_j \lambda_j u^j \), the optimal cost is the same as in problem (20) and any optimal solution of the new problem is also an optimal solution of the original problem (20). Consider an optimal basic feasible solution of the new problem, that is, at least \( R + 1 \) constraints are satisfied with equality. (Such an optimal basic feasible solution can be found in polynomial time because we have \( O(R) \) variables and constraints.) In particular, at least \( R + 1 - L - 1 \) of the constraints \( \lambda_j \geq 0 \) must be satisfied with equality, which means that at most \( L + 1 \) of the variables \( \lambda_j \) are positive. Thus, an optimal solution of the original side-constrained problem (20) can be expressed as a convex combination of no more than \( L + 1 \) extreme points of \( P_{ni} \). Equivalently, an optimal policy can be obtained by randomizing between no more than \( L + 1 \) priority policies. We summarize this discussion in the following theorem.

**Theorem 7.1** If the side-constrained problem (20) is feasible, then there exists an optimal policy which at the beginning of each busy period selects one of \( L + 1 \) priority policies, according to some fixed probabilities, and follows this policy throughout that busy period. Furthermore, such a policy can be found in polynomial time.

The only part of the proof of Theorem 7.1 that we have not yet presented is the fact that once an optimal solution \( x^* \) is available, it can be expressed as a convex combination of extreme points \( u^1, \ldots, u^{R+1} \) of \( P_{ni} \), in polynomial time. We now show how this can be accomplished.
Let $u^1$ be an extreme point of $P_{ni}$. Such an extreme point can be found by choosing an arbitrary priority policy and evaluating its performance vector. If $x^* = u^1$, we are done. If not, let us consider the line from $u^1$ to $x^*$ and let us consider the point at which this line exits the feasible set $P_{ni}$. Such a point exists because $P_{ni}$ is bounded and can be found by solving the linear programming problem

\[
\begin{align*}
\text{maximize} & \quad t \\
\text{subject to} & \quad x = x^* + t(x^* - u^1) \\
& \quad z = Fz \\
& \quad z \in Q_{ni}^+
\end{align*}
\]

This linear programming problem can be solved in polynomial time. Let $z^1$ be its optimal solution, easily seen to be unique and different from $u^1$. The point $z^1$ lies on the boundary of $P_{ni}$ and in particular, it must lie on a facet of $P_{ni}$. Furthermore, since $z^1 \neq u^1$, there exists a facet of $P_{ni}$ such that $z^1$ lies on that facet but $u^1$ does not. We will now proceed to find such a facet.

One way of finding a facet of $P_{ni}$ with the desired properties is to check each one of the constraints

\[\sum_{i \in S} f_{S;ni} \geq G(S)\]

that define $P_{ni}$ to see whether they are satisfied by $z^1$ and $u^1$. However, this would take exponential time because there are exponentially many such constraints and a different approach is needed.

Consider the related to (21) linear programming problem.

\[
\begin{align*}
\text{maximize} & \quad t \\
\text{subject to} & \quad z = z^* + t(z^* - u^1) \\
& \quad z = Fz \\
& \quad z \in Q_{ni}^+
\end{align*}
\]

Let us view the optimal solution $z^1$ of the linear programming problem (22) as a function of $z^*$ and let us consider small perturbations of $z^*$. Using the sensitivity analysis of linear programming, and in the absence of degeneracy, $z^1$ is locally a linear function of $z^*$ and this linear function can be found very easily, e.g. from the final simplex tableau. The range of this function is the desired facet. We omit the discussion of the case where $z^1$ is a degenerate optimal solution; it can be handled by a somewhat more complicated variant of the above outlined approach. It is not hard to verify that all of the above can be accomplished in polynomial time.

Once we have found a facet of $P_{ni}$ to which $z^1$ belongs, we now proceed to express $z^1$. 

21
as a convex combination of \( R \) extreme points of that facet. This is a problem of the same
type as the one we were trying to solve but in one dimension less. We thus have a recursive
algorithm, consisting of \( R \) stages. Each stage only takes polynomial time and the desired
result has been established.

8 Concluding Remarks

We have presented a generalization of the potential function method developed in [BPT92b]
to describe the achievable region of stochastic systems with exponential distributions to
systems with general distributions. A challenging open question is to extend the method
further to queueing networks with general distributions.

Our main result in the paper is a polynomial reformulation of the branching bandit
problem. An exponential characterization of the achievable region has been known partial-
tly through the work of Tsoucas [Tso91] and explicitly through the work of Bertsimas
and Niño-Mora [BNM92]. In particular the achievable region is characterized as an ex-
tended polymatroid. This raises the question whether an arbitrary extended polymatroid
is always a projection of a higher dimensional polyhedron involving a polynomial number
of variables and constraints. Since polymatroids and extended polymatroids appear in sev-
eral applications in combinatorial optimization such a reformulation will be very useful for
combinatorial problems with side constraints.

We finally indicate how to relax Assumption A(b) which required the probability dis-
tributions of the random variables of interest \((N_{ij} \text{ and } T_i)\) to be of exponential type. Let
us only assume that each \( T_i \) has finite mean and each \( N_{ij} \) has finite mean and variance. If
these random variables are not of exponential type, let us approximate them by random
variables of exponential type with the same means and variances and let us take the limit as
this approximation becomes better and better. For each approximant, the results we have
proved establish that the achievable region will be the same; this is because the constraints
that define the achievable region only depend on the means and variances of \( N_{ij} \) and the
mean of \( T_i \). Taking the limit, and using a continuity argument, the same achievable region
is obtained in the limit.

References

eties and optimization. Research report, IBM Research Division, T.J Watson


Appendix

We show here that not every point in the polyhedron $Q_{ni}^+$ is equal to the vector $I$ associated to some nonidling policy in $\Pi^+$. Consider a problem in which $R = 3$ and suppose that, there is a positive probability that customers of all three classes may coexist, no matter what policy is used. (For this, it is sufficient to assume that $E[No_1No_2No_3] > 0$. The polyhedron $Q_{ni}^+$ is described in terms of 9 variables $z_{ij}$, $i, j = 1, 2, 3$, and 6 constraints. If we impose the additional constraints $z_{21} = 0$, $z_{31} = 0$ and $z_{32} = 0$, we obtain an extreme point $z^*$ of $Q_{ni}^+$. This extreme point is in fact the vector $I$ associated with the priority policy corresponding to the ordering $(1, 2, 3)$. Let us now consider the following policy. We follow the priority ordering $(1, 2, 3)$ except that whenever $N_2 = 0$, class 3 gets priority over class 1. With this policy, we will still have $z_{21} = 0$ and $z_{32} = 0$ but $z_{31}$ will be positive. This shows that the set of points $\{z \in Q_{ni}^+ | z_{21} = 0, z_{32} = 0\}$ is an edge of $Q_{ni}^+$. Given that $Q_{ni}^+$ is bounded, if we start at $z^*$ and move that edge, we must eventually hit another extreme point. At that extreme point, at least one of the variables $z_{11}, z_{22}, z_{33}, z_{12}, z_{23},$ or $z_{13}$ is equal to zero. We will argue such a vector cannot be the vector $I$ corresponding to a policy. Indeed, if $z_{12} = 0$, then the extreme point can only be achieved by a policy that simultaneously satisfies $I_{21} = 0$ and $I_{12} = 0$. Such a policy must give priority to class 1 over class 2 and to class 2 over class 1, which is impossible given our assumption that customers of these two classes will sometimes coexist. If $z_{23} = 0$, the extreme point is not achievable for similar reasons. If $z_{13} = 0$, the extreme point can only be achieved by a policy that satisfies $I_{21} = 0$, $I_{32} = 0$, and $I_{13} = 0$. Such a policy would reach an impasse at times when customers of all three classes are present. Finally note that $I_{ii} > 0$ for every policy because otherwise class $i$ customers would be never served. Thus, extreme points at which $z_{ii} = 0$ for some $i$ are not achievable either and this concludes the argument.