Optimal Transmission Policies for Noisy Channels

Ger Koole
Division of Mathematics and Computer Science
Vrije Universiteit
De Boelelaan 1081a, 1081 HV Amsterdam
THE NETHERLANDS

Zhen Liu
INRIA
Centre Sophia Antipolis
2004 Route des Lucioles, B.P. 93, 06902 Sophia-Antipolis
FRANCE

Rhonda Righter
Department of Operations and Management Information Systems
Santa Clara University
Santa Clara, CA 95053
USA

Operations Research **49**:892–899, 2001

Abstract

We consider transmission policies for multiple users sharing a single wireless link to a base station. The noise, and hence the probability of correct transmission of a packet, depends on the state of the user receiving the packet. The state for each user is independent of the states of the other users and changes according to a two-state (good/bad) Markov chain. The state of a user is observed only when it transmits. We give conditions under which the optimal policy is the myopic policy, in which a packet is transmitted to the user that is most likely to be in the better of the two states. We do this by showing that the optimal value function is marginally linear in each of the users' probabilities of being in the good state.

Our model also may be applied to flexible manufacturing systems with unreliable tools, and networked computer systems.

SCHEDULING, STOCHASTIC DYNAMIC PROGRAMMING, RESTLESS BANDITS, MOBILE COMMUNICATIONS

1 Introduction

We consider transmission policies for communication channels in the presence of noise. Our work is motivated by a wireless LAN model, in which links are subject to errors that tend to be bursty. Messages are divided into fixed length packets, and there is a base station that relays packets between the wireless links and a wired backbone network. Multiple mobile sessions or users share a single wireless link to the base station. Since the wired network is much faster than the wireless links, packets destined for different users are queued at the base station. As users move, the strength of the signal and the effects of fading, shadowing, frequency hopping, and interference vary, so the noise on the link depends on the user receiving or transmitting packets, is independent for different users, and tends to be bursty. Under some protocols, such as the stop-and-wait ARQ (automatic repeat request) protocol, a packet is repeatedly retransmitted until it is received correctly. This results in head-of-line blocking, because during these retransmissions other packets for other users might have been successfully transmitted.

We consider a single shared link for transmitting packets from the base station to the user. Time is slotted so that at most one packet can be transmitted in each time slot. We model the noisiness of the link for the users with the Gilbert model (Gilbert, 1960) of independent identical two-state (good and bad) Markov chains. Such a model appears to be reasonable given earlier empirical studies (Duchamp and Reynolds, 1992). If a packet destined to a particular user is transmitted and the state for that user is good, the packet is transmitted successfully, otherwise it is not. If a user is in the good (bad) state at the beginning of a time slot, it moves to the bad (good) state at the beginning of the next slot with probability p (q). Let $r = 1-p-q$, where we generally assume $r > 0$, i.e., the state process has positive autocorrelation. In other words, the probability that a user is in the good state immediately after a successful transmission, $1-p$, is greater than the corresponding probability immediately after an unsuccessful transmission, q. Indeed, given the burstiness of the noise, it is expected that $p < 1/2$ and $q < 1/2$. If a packet is transmitted to a user in a particular slot, we learn what the user's state was at the beginning of the slot; otherwise we cannot observe the states of the users. We assume that there are an infinite number of packets waiting to be transmitted to each user.

In the classic burst error model, the probability of error is between 0 and 1 when the user is in the bad state. This considerably complicates the analysis, and we do not consider that case here.

Our objective is to maximize the expected discounted number of successful transmissions, where the discount factor is $0 < \alpha < 1$. We show that for most parameter values it is optimal to always transmit to the user that is most likely to be in the good state. We call this the BU (best user) policy. When users are initially ordered according to the current probability that they are in the good state, and $r > 0$, the BU policy is equivalent to PRR (Persistent Round Robin), in which the link is dedicated to each user in a cyclic fashion according to this order, and packets are transmitted to the same user until a packet fails to be transmitted correctly.

If $r < 0$, so the state process has negative autocorrelation, we show that for two users the BU policy is still optimal, though the BU policy no longer corresponds to PRR. Now it is optimal to continue transmitting a packet until it is received successfully, as TCP does. At that point it is optimal to switch to the other user. Finally, if $r = 0$, then all users are equally likely to be in the good state at any time, so all policies are equivalent.

Our model is a type of restless single-armed bandit problem. In standard bandit problems, "arms" that are not pulled (users that are not served) do not change state. For such problems, an index policy is known to be optimal (Gittins, 1989, Gittins and Jones, 1974). Index policies are not generally optimal for restless bandits, in which bandits that are not pulled change state. However, for our system the optimal policy is not only an index policy, it is myopic.

In most prior work it is assumed that that packets arrive over time and the states of users (their connectivities and queue lengths) are known at the beginning of each slot. Tassiulas and Ephremides (1993) show that for slotted systems with Markovian dynamics, the LCQ policy of always serving the longest connected queue maximizes the stability region of the system. For stochastically identical users LCQ also minimizes the delay. Bambos and Michailidis (1995) extend these results to general stationary ergodic arrival and connectivity processes. Bhagwat et al. (1996) use simulation to investigate the effects of various channel state dependent packet (CSDP) scheduling methods. Among the users in the good state, FIFO, round robin, earliest timestamp first, and LCQ policies are studied, and round robin performs best in terms of channel utilization, fairness, and throughput. A model in which the state of the user is not observed at all is also considered, and round robin among all users performs well relative to FIFO. Carr and Hajek (1993) and Tassiulas and Papavassiliou (1995) consider non-slotted, asynchronous systems in which connectivity instances for each queue occur according to random continuous-time processes, and these instances are known at the time they occur. Shakkottai and Srikant (1999) consider a model with known but random connectivities and deadlines for packets. See also Lu, Bharghavan, and Srikant (1999), Altman and Kushner (1999), and Lott and Teneketzis (1998).

In our model a user's state or connectivity is observed only upon packet transmission to the user, and at other times we know only the probability that the user is in the good state, based on the time of the last transmission to the user and whether it was successful or not. Choi, Wasserman, and Stark (1999) and Wasserman and Lennon Olsen (1998) consider a similar model for state observation. In the first paper there is a single user, and the problem is to decide whether to attempt or suspend transmission to trade off energy efficiency and utilization. In the latter paper, each user has its own link or server, and the problem is to decide the power levels for transmission, taking into consideration interference among users.

Our model also applies to flexible manufacturing systems in which a single machine must do multiple tasks requiring different tools. The tools may break down or be unavailable at random times (e.g., they may be shared with a higher priority machine), and the machine (or worker) learns the state of a tool when it tries to use the tool. Multi-tasking computer systems with shared resources have similar dynamics. In this case, a computer may attempt to access a data base or a printer, say, that may be in use by another computer on the network.

2 Two Users

We first consider the case when there are only two users. For this case, we can completely specify the optimal value function, as well as the optimal policy, and the BU policy is optimal for all parameter values, though in this section we assume $r = 1 - p - q \geq 0$. As we do not observe the actual state of the users, we take as *information state* (Kumar and Varaiya, 1986, Ch. 6) of the system the probabilities that each of the users is in the good state. Let $\pi = (\pi_1, \pi_2)$ represent these probabilities, where we assume $\pi_1 \geq \pi_2$ without loss of generality.

If we transmit a packet to user 1 we receive a reward π_1 (the probability of a successful transmission) and the system moves from state (π_1, π_2) to $(1-p, \pi_2(1-p)+(1-\pi_2)q) = (q+r, q+r\pi_2)$ with probability π_1 (the system was in the good state at the beginning of the slot), and to $(q + r\pi_2, q)$ with probability $1 - \pi_1$ (the system was in the bad state). Note that $q + r = 1 - p$; we prefer to express everything in terms of q and r. The same transition mechanism holds for transmission to user 2. We therefore have the following dynamic programming equation, where $V(\pi_1, \pi_2)$ is the optimal value function for the infinite horizon problem with discount rate α.

$$\begin{aligned} V(\pi_1, \pi_2) \quad = \quad & \max\Big\{\pi_1 + \alpha\pi_1 V(q + r, q + r\pi_2) + \alpha(1 - \pi_1)V(q + r\pi_2, q); \\ & \pi_2 + \alpha\pi_2 V(q + r, q + r\pi_1) + \alpha(1 - \pi_2)V(q + r\pi_1, q)\Big\}. \end{aligned}$$

Note that $q \leq q + r\pi \leq q + r$ for all $0 \leq \pi \leq 1$, since we assume $r \geq 0$, and by definition $r \leq 1$. Thus, under the BU policy if a user has a successful transmission in a time slot, we continue to permit that user to transmit, otherwise we switch to the other user. That is, we alternate between the two users, starting with the user with the higher probability of being in the good state (user 1), and letting each user transmit continuously until it fails to transmit successfully, and then switching to the other user. This we call PRR (persistent round robin). We therefore have the following.

Proposition 2.1 *When $r > 0$, the BU policy corresponds to PRR.*

Theorem 2.2 *For two users, when $r > 0$, the optimal policy is the BU policy, and for all π, $V(q + r, q + r\pi) = a + b\pi$ and $V(q + r\pi, q) = c + d\pi$, where*

$$a = \frac{(q + r - \alpha r + \alpha r q)(1 - \alpha r q - \alpha r^2)}{(1 - \alpha)\gamma}$$

$$b = \frac{(1 - q - r)\alpha r^2}{\gamma}$$

$$c = \frac{q(1 + \alpha r)(1 - \alpha r q - \alpha r^2)}{(1 - \alpha)\gamma}$$

$$d = \frac{r(1 - \alpha r q - \alpha r^2)}{\gamma}$$

and $\gamma = (1 - \alpha r^2)(1 - \alpha r)$. Also,

$$V(\pi_1, \pi_2) = \pi_1 + \alpha[c + d(\pi_1 + \pi_2) + (b - d)\pi_1\pi_2].$$

Proof. Since we have a discounted dynamic programming problem, we need only show that the BU policy with the value function given above satisfies the dynamic programming recursion. See, for example, Ross (1983). Substituting our expressions for $V(q+r, q+r\pi)$ and $V(q+r\pi, q)$ into the general recursion, we have, for $\pi_1 \geq \pi_2$,

$$\begin{aligned}
V(\pi_1, \pi_2) = \max \Big\{ &\pi_1 + \alpha\pi_1 V(q + r, q + r\pi_2) + \alpha(1 - \pi_1)V(q + r\pi_2, q); \\
&\pi_2 + \alpha\pi_2 V(q + r, q + r\pi_1) + \alpha(1 - \pi_2)V(q + r\pi_1, q) \Big\} \\
= \max \Big\{ &\pi_1 + \alpha\pi_1(a + b\pi_2) + \alpha(1 - \pi_1)(c + d\pi_2); \\
&\pi_2 + \alpha\pi_2(a + b\pi_2) + \alpha(1 - \pi_2)(c + d\pi_1) \Big\} \\
= \max \Big\{ &\pi_1 + \alpha[c + (a - c)\pi_1 + d\pi_2 + (b - d)\pi_1\pi_2]; \\
&\pi_2 + \alpha[c + (a - c)\pi_2 + d\pi_1 + (b - d)\pi_1\pi_2] \Big\} \\
= \max \Big\{ &\pi_1 + \alpha[c + d\pi_1 + d\pi_2 + (b - d)\pi_1\pi_2]; \\
&\pi_2 + \alpha[c + d\pi_2 + d\pi_1 + (b - d)\pi_1\pi_2] \Big\} \\
= &\pi_1 + \alpha[c + d(\pi_1 + \pi_2) + (b - d)\pi_1\pi_2]
\end{aligned}$$

where we use the easily checked fact that $d = a - c$ for a, c, and d as given in the theorem. Now we need only show that a, b, c, and d are as given in the theorem under the BU policy.

We have

$$\begin{aligned}
a + b\pi &= V(q + r, q + r\pi) \\
&= q + r + \alpha(q + r)(V(q + r, q + r(q + r\pi)) + \alpha(1 - q - r)(V(q + r(q + r\pi), q) \\
&= q + r + \alpha(q + r)(a + b(q + r\pi)) + \alpha(1 - q - r)(c + d(q + r\pi)) \\
c + d\pi &= V(q + r\pi, q) \\
&= q + r\pi + \alpha(q + r\pi)V(q + r, q + rq) + \alpha(1 - q - r\pi)V(q + rq, q) \\
&= q + r\pi + \alpha(q + r\pi)(a + bq) + \alpha(1 - q - r\pi)(c + dq).
\end{aligned}$$

As these inequalities hold for each value of π, we obtain, by considering the constant and first order terms:

$$\begin{aligned}
a &= q + r + \alpha(q + r)(a + bq) + \alpha(1 - q - r)(c + dq) \\
b &= \alpha(q + r)br + \alpha(1 - q - r)dr \\
c &= q + \alpha q(a + bq) + \alpha(1 - q)(c + dq) \\
d &= r + \alpha r(a + bq) - \alpha r(c + dq).
\end{aligned}$$

Using Maple to solve this system, we get the expressions of the theorem. □

3 Multiple Users

Now consider the general case of $n \geq 2$ users. Let $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$, where π_i is the probability that user i is in the good state, and let $\pi^i = (\pi_1, \pi_2, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n)$. If a user i with probability π_i of being in the good state is not sent a packet, its updated probability of being in the good state becomes $\pi_i(1 - p) + (1 - \pi_i)q = q + r\pi_i$ at the next time period. Therefore let $u(\pi_i) = q + r\pi_i$ be the updating function and let $u(\pi) = (q + r\pi_1, \ldots, q + r\pi_n)$. Also let $u^1(\pi) = u(\pi)$ and $u^n(\pi) = u(u^{n-1}(\pi))$, for all $n > 1$. Recall that PRR first orders the users according to their initial probability of being in the good state, so $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$, and then serves them cyclically in this order starting with user 1, where each user is served until the first time a packet for the user is unsuccessfully transmitted.

Proposition 3.1 *When $r > 0$, PRR is equivalent to the BU policy.*

Proof. Assume $r > 0$. Under both policies, the next packet will be transmitted to user 1. If the packet is successfully (resp. unsuccessfully) transmitted, the new state of user 1 in the next slot becomes $q + r$ (resp. q). Since $q + r \geq u(\pi) \geq q$ for all π, if the packet was successful (unsuccessful) user 1 will become the highest (lowest) priority user in the next slot under the

BU policy. Note that if $\pi_i \geq \pi_j$, then $u(\pi_i) \geq u(\pi_j)$, so all unserved users maintain their relative priorities under the BU policy. Hence, if the first packet is unsuccessful, the following packet will be transmitted to user 2, otherwise it will be transmitted to user 1. The same argument applies for each slot, and the BU policy agrees with PRR. □

Note that under PRR, after all users have been served at least once, the state of the system at time t (just before the transmission at time t) can be characterized by $(s, \mathbf{k})$, where s and $\mathbf{k}$ are defined as follows. Let n be the user that was served at time $t-1$. Then $s \in \{G, B\}$ is defined as the "good" or "bad" state of user n at time $t-1$, $\mathbf{k} = (k_1, \ldots, k_{n-1})$, and $k_i \geq 2$ is the time since the last (unsuccessful) transmission for user i. At time t we have $\pi_n(G) = 1 - p = q + r$, $\pi_n(B) = q$, and for $1 \leq i < n$:

$$\pi_i = q\frac{1 - r^{k_i}}{1 - r}.$$

Of course, if $k_i > k_j$, then $\pi_i > \pi_j$, again confirming that PRR agrees with the BU policy.

We require the following condition to show that the BU policy is optimal for $n > 2$ users.

$$\textbf{(A)} \qquad \alpha \leq \frac{2 - p - \sqrt{4p + p^2}}{(2 - 4p)(1 - p - q)}.$$

Note that the right-hand-side of the above inequality approaches 1 as p and q approach 0. It is also increasing in q, so that (A) holds if

$$\alpha \leq \frac{2 - p - \sqrt{4p + p^2}}{(2 - 4p)(1 - p)}.$$

It is easy to see that

$$\frac{2 - p - \sqrt{4p + p^2}}{(2 - 4p)(1 - p)} \geq 1$$

if $p \geq 1 - 1/\sqrt{2} = .293$ or $p = 0$. Therefore, a sufficient condition for (A) to be true is $p \geq .293$ or $p = 0$. Also, for $0 \leq p \leq 1$, the right hand side of the inequality in condition (A) is convex in p and is minimized when $p = .872$ for $q = 0$. Thus, another sufficient condition for (A) is $\alpha \leq .872$.

To get another sufficient condition when $p < .293$ or $\alpha > .872$ let us rewrite condition (A) as

$$\alpha r \leq \frac{2 - p - \sqrt{4p + p^2}}{(2 - 4p)}.$$

The right hand side is decreasing in p, so for $p < .293$ the right hand side is at least .707. Thus, (A) will hold if $\alpha r < .707$, ie., $p+q > 1-.707/\alpha$. This implies the stronger sufficient condition: $p+q \geq .293$.

Condition (A) is a sufficient condition for the optimality of the BU policy, but there is no reason to suppose it is necessary. Indeed, it is not necessary when there are only two users. However, with more than two users, we have been unable to prove that the BU policy holds when (A) does not hold, nor have we found a counterexample for which the BU policy does not hold when (A) does not hold.

We define the Generalized Persistent Round Robin (GPRR) policy as the policy that serves the users cyclically in the order $1, 2, \ldots, n$, where each user is served until the first time a packet for the user is unsuccessfully transmitted, and where the initial order is arbitrary. When $\pi_1 \geq \cdots \geq \pi_n$ then GPRR corresponds to PRR. Let $V(\pi)$ be the optimal value function, and let $W(\pi)$ be the value function under GPRR.

For simplicity in notation, π in $W(\pi)$ will be shift-rotated each time GPRR switches the user. Thus, GPRR will always serve the user whose state is represented by the first component of the state vector in $W(\pi)$. Let $A(\pi^i) = W(u(1, \pi^i))$ and $B(\pi^i) = W(u(\pi^i, 0))$, so

$$W(\pi) := \pi_1 + \alpha\pi_1 A(\pi^1) + \alpha(1-\pi_1)B(\pi^1).$$

Theorem 3.2 *Suppose that $r > 0$ and condition (A) holds. Then, for all i, there exist positive increasing functions $f_i(\pi^i)$ and $g_i(\pi^i)$ such that*

$$W(\pi) = f_i(\pi^i) + g_i(\pi^i)\pi_i.$$

Moreover, if $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$, then $V(\pi) = W(\pi)$. Thus the BU policy is optimal.

Proof. We will use value iteration, i.e., induction on a finite time horizon. Let a superscript of N denote the value functions, etc., when there are N periods to go. Also, let $W_i^N(\pi) := W^N(\pi_i, \pi^i)$. Therefore, by definition, $W_1^N(\pi) = W^N(\pi)$ and

$$W_i^N(\pi) := \pi_i + \alpha\pi_i A^{N-1}(\pi^i) + \alpha(1-\pi_i)B^{N-1}(\pi^i).$$

We show by induction the following six relations.

(i) If $\pi_i \geq \pi_{i+1}$ then $W_i^N(\pi) \geq W_{i+1}^N(\pi)$.

(ii) If $\pi_i \geq \pi_{i+1}$ then $W^N(\pi) \geq W^N(\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \pi_i, \pi_{i+2}, \ldots, \pi_n)$.

(iii) For all i, there exist positive increasing functions $f_i^N(\pi^i)$ and $g_i^N(\pi^i)$ such that

$$W^N(\pi) = f_i^N(\pi^i) + g_i^N(\pi^i)\pi_i.$$

(iv) For all π (of dimension $n-1$), $A^N(\pi) - B^N(\pi) \leq r/(1-\alpha r)$.

(v) For all π (of dimension $n-2$),

$$W^N(q+r, q+r, \pi) - W^N(q+r, q, \pi) \leq 1/\alpha.$$

(vi) $V^N(\pi) = W^N(\pi)$ if $\pi_1 \geq \cdots \geq \pi_n$.

Note that relation (vi) implies that the BU policy is optimal, by proposition 3.1.

Statements (i)-(vi) are trivially true for $N = 0$ since $V^0(\pi) = W^0(\pi) = 0$. Suppose they hold for N and consider $N+1$.

Proof of (i): Let $\pi^{i,i+1} := (\pi_1, \ldots, \pi_{i-1}, \pi_{i+2}, \ldots, \pi_n)$. From the induction hypothesis for (iii),

$$A^N(\pi^i) = W^N(u(1,\pi^i)) = f_{i+1}^N(u(1,\pi^{i,i+1})) + g_{i+1}^N(u(1,\pi^{i,i+1}))(q + r\pi_{i+1})$$

and we have a similar equation for $B^N(\pi^i)$. Therefore, we have that for some $a_{i+1}^N(\pi^{i,i+1})$, $b_{i+1}^N(\pi^{i,i+1})$, $c_{i+1}^N(\pi^{i,i+1})$, and $d_{i+1}^N(\pi^{i,i+1})$,

$$\begin{aligned}
A^N(\pi^i) &= a_{i+1}^N(\pi^{i,i+1}) + b_{i+1}^N(\pi^{i,i+1})\pi_{i+1} \\
B^N(\pi^i) &= c_{i+1}^N(\pi^{i,i+1}) + d_{i+1}^N(\pi^{i,i+1})\pi_{i+1} \\
A^N(\pi^{i+1}) &= a_{i+1}^N(\pi^{i,i+1}) + b_{i+1}^N(\pi^{i,i+1})\pi_i \\
B^N(\pi^{i+1}) &= c_{i+1}^N(\pi^{i,i+1}) + d_{i+1}^N(\pi^{i,i+1})\pi_i.
\end{aligned}$$

For example, $a_{i+1}^N(\pi^{i,i+1}) = f_{i+1}^N(u(1,\pi^{i,i+1})) + qg_{i+1}(u(1,\pi^{i,i+1}))$. Therefore,

$$\begin{aligned}
W_i^{N+1}(\pi) &= \pi_i + \alpha\pi_i A^N(\pi^i) + \alpha(1-\pi_i)B^N(\pi^i) \\
&= \pi_i + \alpha\pi_i[a_{i+1}^N(\pi^{i,i+1}) + b_{i+1}^N(\pi^{i,i+1})\pi_{i+1}] + \alpha(1-\pi_i)[c_{i+1}^N(\pi^{i,i+1}) + d_{i+1}^N(\pi^{i,i+1})\pi_{i+1}] \\
W_{i+1}^{N+1}(\pi) &= \pi_{i+1} + \alpha\pi_{i+1}[a_{i+1}^N(\pi^{i,i+1}) + b_{i+1}^N(\pi^{i,i+1})\pi_i] + \alpha(1-\pi_{i+1})[c_{i+1}^N(\pi^{i,i+1}) + d_{i+1}^N(\pi^{i,i+1})\pi_i]
\end{aligned}$$

so

$$W_i^{N+1}(\pi) - W_{i+1}^{N+1}(\pi) = (\pi_i - \pi_{i+1})\left[1 + \alpha\left(a_{i+1}^N(\pi^{i,i+1}) - c_{i+1}^N(\pi^{i,i+1}) - d_{i+1}^N(\pi^{i,i+1})\right)\right].$$

Thus we must show that $c_{i+1}^N(\pi^{i,i+1}) + d_{i+1}^N(\pi^{i,i+1}) - a_{i+1}^N(\pi^{i,i+1}) \le 1/\alpha$. Note that since $A^N(\pi^i) = a_{i+1}^N(\pi^{i,i+1}) + b_{i+1}^N(\pi^{i,i+1})\pi_{i+1}$, we have

$$\begin{aligned} a_{i+1}^N(\pi^{i,i+1}) &= A^N(\pi_1, \pi_2, \ldots, \pi_{i-1}, 0, \pi_{i+2}, \ldots, \pi_n) \\ &= W^N(u(1, \pi_1, \pi_2, \ldots, \pi_{i-1}, 0, \pi_{i+2}, \ldots, \pi_n)) \\ &= W^N(q+r, q+r\pi_1, \ldots, q+r\pi_{i-1}, q, q+r\pi_{i+2}, \ldots, q+r\pi_n) \end{aligned}$$

and since $B^N(\pi^i) = c_{i+1}^N(\pi^{i,i+1}) + d_{i+1}^N(\pi^{i,i+1})\pi_{i+1}$,

$$\begin{aligned} c_i^N(\pi^{i,i+1}) + d_i^N(\pi^{i,i+1}) &= B^N(\pi_1, \pi_2, \ldots, \pi_{i-1}, 1, \pi_{i+2}, \ldots, \pi_n) \\ &= W^N(u(\pi_1, \pi_2, \ldots, \pi_{i-1}, 1, \pi_{i+2}, \ldots, \pi_n, 0)) \\ &= W^N(q+r\pi_1, \ldots, q+r\pi_{i-1}, q+r, q+r\pi_{i+2}, \ldots, q+r\pi_n, q). \end{aligned}$$

Therefore, by the induction assumption on (ii),

$$\begin{aligned} a_i^N(\pi^{i,i+1}) &= W^N(q+r, q+r\pi_1, \ldots, q+r\pi_{i-1}, q, q+r\pi_{i+2}, \ldots, q+r\pi_n) \\ &\ge W^N(q+r, q, q+r\pi_1, \ldots, q+r\pi_{i-1}, q+r\pi_{i+2}, \ldots, q+r\pi_n) \\ &= W^N(q+r, q, u(\pi^{i,i+1})) \end{aligned}$$

Also, since from (iii) $W^N(\pi)$ is increasing in all of its arguments, and from (ii),

$$\begin{aligned} c_i^N(\pi^{i,i+1}) + d_i^N(\pi^{i,i+1}) &= W^N(q+r\pi_1, \ldots, q+r\pi_{i-1}, q+r, q+r\pi_{i+2}, \ldots, q+r\pi_n, q) \\ &\le W^N(q+r\pi_1, \ldots, q+r\pi_{i-1}, q+r, q+r\pi_{i+2}, \ldots, q+r\pi_n, q+r) \\ &\le W^N(q+r, q+r, q+r\pi_1, \ldots, q+r\pi_{i-1}, q+r\pi_{i+2}, \ldots, q+r\pi_n) \\ &= W^N(q+r, q+r, u(\pi^{i,i+1})). \end{aligned}$$

We therefore have $c_i^N(\pi^{i,i+1}) + d_i^N(\pi^{i,i+1}) - a_i^N(\pi^{i,i+1}) \le 1/\alpha$ by the induction hypothesis for (v).

Proof of (ii): Now let us show that for all i, if $\pi_i \ge \pi_{i+1}$ then

$$W^{N+1}(\pi) \ge W^{N+1}(\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \pi_i, \pi_{i+2}, \ldots, \pi_n).$$

First suppose $i \ge 2$. Then, by definition,

$$\begin{aligned} W^{N+1}(\pi) &= W_1^{N+1}(\pi) = \pi_1 + \alpha\pi_1 A^N(\pi^1) + \alpha(1-\pi_1)B^N(\pi^1) \\ &= \pi_1 + \alpha\pi_1 W^N(u(1, \pi^1)) + \alpha(1-\pi_1)W^N(u(\pi^1, 0)) \\ &\ge \pi_1 + \alpha\pi_1 W^N(u(1, \pi_2, \ldots, \pi_{i-1}, \pi_{i+1}, \pi_i, \ldots, \pi_n)) + \alpha(1-\pi_1)W^N(u(\pi_2, \ldots, \pi_{i-1}, \pi_{i+1}, \pi_i, \ldots, \pi_n \\ &= W_1^{N+1}(\pi_1, \pi_2, \ldots, \pi_{i-1}, \pi_{i+1}, \pi_i, \ldots, \pi_n), \end{aligned}$$

where the inequality follows from the induction assumption for (ii). Now suppose $i = 1$. Then

$$W^{N+1}(\pi) = W_1^{N+1}(\pi) \geq W^{N+1}(\pi_2, \pi_1, \pi_3, \ldots, \pi_n) = W_2^{N+1}(\pi)$$

where the inequality follows from (i).

Proof of (iii): Now let us show that for all i,

$$W^{N+1}(\pi) = f_i^{N+1}(\pi^i) + g_i^{N+1}(\pi^i)\pi_i$$

where $f_i^{N+1}(\pi^i)$ and $g_i^{N+1}(\pi^i)$ are increasing and positive. We have

$$W^{N+1}(\pi) = \pi_1 + \alpha\pi_1 A^N(\pi^1) + \alpha(1-\pi_1)B^N(\pi^1),$$

which is clearly linear in π_1. Also, as we argued earlier, from the induction hypothesis for (iii), for any $i > 1$, $A^N(\pi^1) = a_i^N(\pi^{1,i}) + b_i^N(\pi^{1,i})\pi_i$ and $B^N(\pi^1) = c_i^N(\pi^{1,i}) + d_i^N(\pi^{1,i})\pi_i$, where $a_i^N(\cdot)$, $b_i^N(\cdot)$, $c_i^N(\cdot)$, and $d_i^N(\cdot)$ are increasing and positive. The result follows.

Proof of (iv): We now show that for all π (of dimension $n-1$), $A^{N+1}(\pi) - B^{N+1}(\pi) \leq r/(1-\alpha r)$. We have, using the induction hypotheses for (ii) and (iv) for the two inequalities below, respectively,

$$\begin{aligned}
A^{N+1}(\pi) - B^{N+1}(\pi) &= W^{N+1}(q+r, u(\pi)) - W^{N+1}(u(\pi), q) \\
&\leq W^{N+1}(q+r, u(\pi)) - W^{N+1}(q, u(\pi)) \\
&= q + r + (q+r)\alpha A^N(u^2(\pi)) + (1-q-r)\alpha B^N(u^2(\pi)) \\
&\quad - \left[q + q\alpha A^N(u^2(\pi)) + (1-q)\alpha B^N(u^2(\pi))\right] \\
&= r\left[1 + \alpha\left(A^N(u^2(\pi)) - B^N(u^2(\pi))\right)\right] \\
&\leq r\left(1 + \alpha r/(1-\alpha r)\right) = r/(1-\alpha r).
\end{aligned}$$

Proof of (v): Now we show that for all π (of dimension $n-2$),

$$\Delta := W^{N+1}(q+r, q+r, \pi) - W^{N+1}(q+r, q, \pi) \leq 1/\alpha.$$

Note that

$$\begin{aligned}
W^{N+1}(q+r, q+r, \pi) &= q + r + (q+r)\alpha W^N(q+r, u(q+r), u(\pi)) \\
&\quad + (1-q-r)\alpha W^N(u(q+r), u(\pi), q)
\end{aligned}$$

so

$$\begin{aligned}
\Delta &= (q+r)\alpha\left[W^N(q+r,u(q+r),u(\pi)) - W^N(q+r,u(q),u(\pi))\right] \\
&\quad +(1-q-r)\alpha\left[W^N(u(q+r),u(\pi),q) - W^N(u(q),u(\pi),q)\right] \\
&=: (q+r)\alpha\Delta_1 + (1-q-r)\alpha\Delta_2
\end{aligned}$$

From (iii) we have

$$\begin{aligned}
W^N(q+r,u(q+r),u(\pi)) &= f_2^N(q+r,u(\pi)) + g_2^N(q+r,u(\pi))u(q+r) \\
&= f_2^N(q+r,u(\pi)) + g_2^N(q+r,u(\pi))(q+r(q+r))
\end{aligned}$$

so

$$\begin{aligned}
\Delta_1 &= g_2^N(q+r,u(\pi))\,[q+r(q+r)-(q+rq)] = g_2^N(q+r,u(\pi))r^2 \\
&= r\left[g_2^N(q+r,u(\pi))r\right] \\
&= r\left[W^N(q+r,q+r,u(\pi)) - W^N(q+r,q,u(\pi))\right] \\
&\leq r/\alpha
\end{aligned}$$

where the last inequality follows from the induction hypothesis for (v). We also have

$$\begin{aligned}
W^N(u(q+r),u(\pi),q) &= q+r(q+r)+(q+rq+r^2)\alpha A^{N-1}(u^2(\pi),u(q)) \\
&\quad +(1-q-rq-r^2)\alpha B^{N-1}(u^2(\pi),u(q))
\end{aligned}$$

so

$$\begin{aligned}
\Delta_2 &= r^2\left[1+\alpha\left(A^{N-1}(u^2(\pi),u(q)) - B^{N-1}(u^2(\pi),u(q))\right)\right] \\
&\leq r^2\left[1+\frac{\alpha r}{1-\alpha r}\right] = \frac{r^2}{1-\alpha r}
\end{aligned}$$

where the inequality follows from the induction hypothesis for (iv). We therefore have

$$\begin{aligned}
\Delta &= (q+r)\alpha\Delta_1 + (1-q-r)\alpha\Delta_2 \\
&\leq (q+r)r + (1-q-r)\frac{\alpha r^2}{1-\alpha r} \\
&= (1-p)r + p\frac{\alpha r^2}{1-\alpha r}.
\end{aligned}$$

Thus, $\Delta \leq 1/\alpha$ if

$$(1-p)\alpha r + p\frac{\alpha^2 r^2}{1-\alpha r} \leq 1,$$

which is equivalent to

$$\alpha r \leq \frac{2-p-\sqrt{4p+p^2}}{2-4p}.$$

This is just condition (A).

Proof of (vi): Suppose $\pi_1 \geq \cdots \geq \pi_n$. We now show that $V^{N+1}(\pi) = W^{N+1}(\pi)$. We have the following, where the first equation follows from the definition of $V^{N+1}(\pi)$, the second follows from the induction hypothesis for (vi), the third from the definitions of $A^N(\pi)$ and $B^N(\pi)$, and the fourth from the induction hypothesis for (i).

$$\begin{aligned}
V^{N+1}(\pi) &= \max_i \pi_i + \alpha\pi_i V^N(q+r, u(\pi^i)) + \alpha(1-\pi_i)V^N(u(\pi^i), q) \\
&= \max_i \pi_i + \alpha\pi_i W^N(q+r, u(\pi^i)) + \alpha(1-\pi_i)W^N(u(\pi^i), q) \\
&= \max_i \pi_i + \alpha\pi_i A^N(\pi^i) + \alpha(1-\pi_i)B^N(\pi^i) \\
&= \pi_1 + \alpha\pi_1 A^N(\pi^1) + \alpha(1-\pi_1)B^N(\pi^1) = W^{N+1}(\pi).
\end{aligned}$$

□

Another sufficient condition for the BU policy to be optimal, which is not included in condition (A), is $q = 0$ (and $r = 1-p > 0$ still). This means that a user, once it is in the bad state, never leaves it, so $u(q) = u(0) = 0$. It is intuitively clear that a GPRR policy must be optimal. We have the following.

Theorem 3.3 *Suppose that $r > 0$ and $q = 0$. Then $V(\pi) = W(\pi)$ and the BU policy is optimal.*

Proof. Again we use induction on a finite time horizon. Let a superscript of N denote the value functions, etc., when there are N periods to go. Suppose that the theorem holds for N or fewer periods to go, and consider $N+1$. (The result is trivially true for $N = 0$ since $V^0(\pi) = W^0(\pi) = 0$.)

Let us relabel the users so that $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$. Suppose the optimal policy, call it ρ, transmits to user i instead of user 1 at time 0, where $\pi_1 > \pi_i$. By the induction hypothesis, ρ will continue to transmit to user i until the first time the transmission is unsuccessful (at time T_1, say), it will then transmit to user 1 until the first unsuccessful transmission to user 1 (at time $T_1 + T_2$, say), and then continue with PRR. Consider an alternative policy, ρ', that, starting at time 0, transmits to user 1 until the first unsuccessful transmission (at time T_1' say), then transmits to user i until the first unsuccessful transmission (at time $T_1' + T_2'$, say), and then continues with PRR. We will show that the expected discounted number of successful transmissions is greater under ρ' than under ρ.

Let us couple the success or failure of the first packet to be transmitted to users 1 and i under the two policies as follows.

With probability $\pi_1\pi_i$ the initial transmissions to both users 1 and i under both policies are successful. Given that the first transmission is successful, the distributions of T_1-1, T_2-1, $T_1'-1$, $T_2'-1$, are all geometric with parameter $q+r$. Thus, we can couple the times so that $T_1'=T_1$ and $T_2'=T_2$. The policies will be identical after time T_1+T_2, and the returns will be the same under both policies. In particular, the return will be

$$ER_1 + R_2 + \alpha^{T_1+T_2} V^{N+1-T_1-T_2}(u^{T_1+T_2}(\pi^{1,i}), 0, 0)$$

where R_1 (R_2) is the total expected discounted number of successes from time 0 to time $\min\{T_1, N\}$ (from time $\min\{T_1, N\}$ to time $\min\{T_1+T_2, N\}$), and $V^k(\pi) := 0$ for all $k \leq 0$.

With probability $(1-\pi_1)(1-\pi_i)$ both initial transmissions under both policies are unsuccessful, so $T_1' = T_1 = T_2' = T_2 = 1$. The policies will be identical after time $T_1+T_2=2$, and the return under both policies will be

$$E\left[\alpha^2 V^{N-1}(u^2(\pi^{1,i}), 0, 0)\right].$$

With probability $\pi_i(1-\pi_1)$, under both policies, the transmission at time 0 is successful, and the transmission at time $T_1+1=T_1'+1$ is unsuccessful. The policies will be identical after time T_1+1, and the returns under both policies will be

$$E\left[R_1 + \alpha^{T_1+1} V^{N-T_1}(u^{T_1+1}(\pi^{1,i}), 0, 0)\right].$$

With probability $\pi_i(1-\pi_1)$, under both policies, the transmission at time 0 is unsuccessful, and the transmission at time $T_1=T_1'=1$ is successful. The policies will be identical after time $1+T_2=1+T_2'$, and the returns under both policies will be

$$E\left[R_2 + \alpha^{T_2+1} V^{N-T_2}(u^{T_2+1}(\pi^{1,i}), 0, 0)\right].$$

Finally, with probability $\pi_1-\pi_i$, under policy ρ the transmission at time 0 is unsuccessful and the transmission at time $T_1=1$ is successful, and under policy ρ' the transmission at time 0 is successful, and the transmission at time $T_1'+1$ is unsuccessful. Letting $T_1'=T_2$ and

$T_1 = T_2' = 1$, the returns under policies ρ and ρ' are, respectively

$$\begin{aligned} U &\leq E\left[\alpha R_1 + \alpha^{T_1'+1} V^{N-T_1'}(u^{T_1'+1}(\pi^{1,i}), 0, 0)\right] \\ U' &= E\left[R_1 + \alpha^{T_1'+1} V^{N-T_1'}(u^{T_1'+1}(\pi^{1,i}), 0, 0)\right] \end{aligned}$$

where R_1 is the total expected discounted number of successes from time 0 to time $\min\{T_1', N\}$ under policy ρ', and αR_1 is an upper bound on the total expected discounted number of successes from time 1 to time $\min\{1+T_1', N\}$ under policy ρ. (The bound is achieved when $N \geq 1+T_1'$.) We therefore have $U' \geq U$. □

Note that the proof above does not work when $q > 0$ for the last case, in which under policy ρ the transmission at time 0 is unsuccessful and the transmission at time $T_1 = 1$ is successful, and under policy ρ' the transmission at time 0 is successful, and the transmission at time $T_1'+1$ is unsuccessful. In this case

$$\begin{aligned} U &\leq E\left[\alpha R_1 + \alpha^{T_1'+1} V^{N-T_1'}(u^{T_1'+1}(\pi^{1,i}), u^{T_1'+1}(q), q)\right] \\ U' &= E\left[R_1 + \alpha^{T_1'+1} V^{N-T_1'}(u^{T_1'+1}(\pi^{1,i}), u^{1}(q), q)\right] \end{aligned}$$

and $u^{T_1'+1}(q) > u^1(q)$.

4 The Case of $r \leq 0$

Now we consider the effect of relaxing the assumption of $r = 1 - p - q > 0$. If $r = 0$, this means that the probability that a user is in the good state immediately after a successful transmission, $1-p$, is the same as the corresponding probability immediately after an unsuccessful transmission, q. In other words, the probability that a user is in the good state is $q = 1-p$ at all times, regardless of whether or not it receives a transmission and regardless of the success of any transmissions that it receives. In this case, all the users are always stochastically identical, and all policies are equivalent in terms of the discounted number of successful transmissions.

Suppose $r = 1-p-q < 0$, i.e., $1-p < q$. This means that a user that is more likely to be in the good state in a particular time slot if it was in the bad state in the last time slot than if it was in the good state in the last time slot. This is certainly less realistic than the case $r > 0$, but could be caused by aggressive corrective action in the bad state. We have the same updating function, $u(\pi) = q + r\pi$, where $u(\pi)$ is the probability that a user is in the good state in the next time slot given that it is not served in the current slot and its current probability of being in the good state is π. Now $q > u(\pi) > q+r$, so, under the BU policy, if a transmission to a user is unsuccessful, that user should receive the next transmission. That is, the base station continues to transmit a packet to a user until the transmission is successful, and then a different user is served. This agrees with stop-and-wait ARQ. Also, if $\pi_i > \pi_j$ then $u(\pi_i) < u(\pi_j)$, so the priorities of unserved users under the BU policy is reversed in each slot. The state of the system at time t (just before the transmission at time t) under the BU policy can be characterized by $(s, \mathbf{k})$, where s and $\mathbf{k}$ are defined as follows. Let n be the user that was served at time $t-1$. Then $s \in \{G, B\}$ is the state of user n at time $t-1$, $\mathbf{k} = (k_1, \ldots, k_{n-1})$, and $k_i \geq 2$ is the time since the last (successful) transmission for user i. We have $\pi_n(G) = 1 - p = q + r$, $\pi_n(B) = q$, and

$$\pi_i = q\frac{1 - r^{k_i}}{1 - r} + r^{k_i}.$$

Let us order the $n-1$ non-transmitting users so that $k_1 \geq k_2 \geq \cdots \geq k_{n-1}$. Then either $\pi_{n-1} = \min_{1 \leq i \leq n-1} \pi_i$ (when k_{n-1} is odd) or $\pi_{n-1} = \max_{1 \leq i \leq n-1} \pi_i$ (when k_{n-1} is even). Thus, under the BU policy, if a transmission to a user (call it user $n-1$) is successful, then in the next slot that user has lowest priority and another user (call it user n) receives a transmission. While user n is being served (the transmissions continue to be unsuccessful), user $n-1$ alternates between being the highest and being the lowest priority user under the BU policy. Therefore, if user n is successful, then user $n-1$ should be served in the next slot as long as k_{n-1} is even. If k_{n-1} is odd, so $\pi_{n-1} = \min_{1 \leq i \leq n-1} \pi_i$, then $\pi_{n-2} = \max_{1 \leq i \leq n-1} \pi_i$ if k_{n-2} is even, and user $n-2$ should be served. In general, user i^* will have highest priority among unserved users, where i^* is the largest integer such that k_{i^*} is even. If k_i is odd for all i, then $i^* = 1$.

We have not been able to show that the BU policy is optimal for $r < 0$ when there are an arbitrary number of users, but when there are only two users the same proof as that for 2.2 shows that the BU policy is in fact optimal, and we can fully characterize the optimal value function. Note that when there are only two users, the BU policy is to transmit to each user until the first successful transmission, and then switch to the other user.

Theorem 4.1 *For two users, the optimal policy is the BU policy, and for all* π, $V(q + r\pi, q + r) = a + b\pi$ *and* $V(q, q + r\pi) = c + d\pi$, *where*

$$\begin{aligned}
a &= \frac{q\left(1 - \alpha r + \alpha q r - \alpha^2 r^2 + \alpha^2 r^3 + \alpha^2 q r^2\right)}{(1-\alpha)\gamma} \\
b &= \frac{r(1 - \alpha r + \alpha q r)}{\gamma} \\
c &= \frac{q(1 - \alpha r + \alpha r q - \alpha r^2 + \alpha^2 r^2 q + \alpha^2 r^3)}{(1-\alpha)\gamma} \\
d &= \frac{\alpha q r^2}{\gamma}
\end{aligned}$$

and $\gamma = (1 - \alpha r^2)(1 - \alpha r)$. Also, for $\pi_1 \geq \pi_2$,

$$V(\pi_1, \pi_2) = \pi_1 + \alpha[c + d(\pi_1 + \pi_2) + (b - d)\pi_1\pi_2].$$

5 Acknowledgements

The clarity of the presentation benefitted greatly from the comments of three anonymous referees.

References

[1] Altman, E., and H. J. Kushner (1999). Optimal scheduling of transmission opportunities in heavy traffic with applications to satellite and mobile radio systems. Preprint.

[2] Bambos, N., and G. Michailidis (1995). On the stationary dynamics of parallel queues with random server connectivities. *Proc. of the 34th Conf. on Dec. and Cont.* 3638–3643.

[3] Bhagwat, P., P. Bhattacharya, A. Krishna, and S.K. Tripathi (1996). Enhancing throughput over wireless LANs using channel state dependent packet scheduling. *IEEE INFOCOM* 1133–1140. IEEE Press.

[4] Carr, M., and B. Hajek (1993). Scheduling with asynchronous service opportunities with applications to multiple satellite systems. *IEEE Trans. Aut. Cont.* 38: 1820–1832.

[5] Choi, J. D, K. M. Wasserman, and W. E. Stark (1999). Effect of channel memory on retransmission protocols for low energy wireless data communications. *Proc. Int'l. Conf. Comm. (ICC'99).*

[6] Duchamp, D., and N. Reynolds (1992). Measured performance of a wireless LAN. *17th Conf. on Local Computer Networks*, 494–499. IEEE Press.

[7] Gilbert, E. N. (1960). Capacity of a burst-noise channel. *Bell Systems Technical Journal.* 39: 1253-1266.

[8] Gittins, J.C. (1989). *Multi-armed Bandit Allocation Indices.* J. Wiley and Sons, New York.

[9] Gittins, J.C., and D.M. Jones (1974). A dynamic allocation index for the sequential design of experiments, In J. Gani et al. (eds.), *Progress in Statistics*, 241–266. North Holland, Amsterdam.

[10] Kumar, P.R., and P. Varaiya (1986). *Stochastic Systems.* Prentice-Hall.

[11] Lott, C., and D. Teneketzis (1998). Multiserver scheduling on priority queues with varying connectivity. Presentation at INFORMS meeting, Oct. 27.

[12] Lu, S., V. Bharghavan, and R. Srikant (1999). Fair scheduling in wireless packet networks. Preprint.

[13] Ross, S.M. (1983). *Introduction to Stochastic Dynamic Programming.* Academic Press, New York.

[14] Shakkottai, S., and R. Srikant (1999). Scheduling real-time traffic with deadlines over a wireless channel. *ACM WOWMOM*, to appear.

[15] Tassiulas, L., and A. Ephremides (1993). Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory.* 39: 466–478.

[16] Tassiulas, L., and S. Papavassiliou (1995). Optimal anticipative scheduling with asynchronous transmission opportunities. *IEEE Trans. Aut. Cont.* 40: 2052–2062.

[17] Wasserman, K. M., and T. Lennon Olsen (1998). On mutually interfering parallel servers subject to external disturbances. Preprint.